Personalising Patient Internet Searching Using Electronic Patient Records

by

Asma Ali S. Al-Busaidi

This thesis is submitted in partial fulfilment
of the requirement for the degree of
Doctor of Philosophy
in
Computer Science

School of Computer Science
Cardiff University

February 2008

DECLARATION

This Work has not previously been accepted in substance for any degrees and is not concurrently submitted in candidature for any degree.

Signed ..Asma....Al-Busaidi....(candidate)

Date ...30/04/2008............

STATEMENT 1

This Thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ..Asma....Al-Busaidi..(candidate)

Date ...30/04/2008........

STATEMENT 2

This Thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ..Asma.....Al-Busaidi.....(candidate)

Date ...30/04/2008........

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..Asma......Al-Busaidi..(candidate)

Date ...30/04/2008........

Dedicated to Patients…

Acknowledgement

I start by thanking Allah (God) Almighty for all His favours on me and for giving me the patience and strength to complete this work.

Next, I would like to thank Professor Alex Gray, my first supervisor, for his enormous support and help throughout this research. I would like to thank him for his expert knowledge, helpful feedback, and close attention to the difficulties I faced. I would like also to thank Professor Nick Fiddian, my second supervisor, for his support, valuable feedback and careful readings of joint papers.

On the personal level, I would like first to thank my parents for their love and prayers that bless my life. May Allah (God) Almighty rest their souls in peace. Special thanks to all my family in Oman for all their love, prayers and support. Special thanks also go to my friends in the UK: Masooma Al-Balushi, Widad Rakhyoot and Sahar Evans, for their thoughts and support.

I would like also to thank my colleagues at Cardiff School of Computer Science for their friendship, assistance and feedback on this research: Dr Omnia Allam, Dr Mikhaila Burgess, Dr Wendy Ivins, Dr Rashid Al-Ali, Dr Fahad Al-Wasil, Badr Al-Daihani, Vikas Deora, Ben Sissons, Yahya Ibrahim and R. Sundar. I would like also to thank my colleagues, the Omani students in Cardiff, (and their families) for their friendship and support.

I would like also to thank the Velindre NHS Trust staff for their cooperation and valuable discussion: Dr David Murray, Dr Anthony Bater, Valerie Aston, Hazel Bailey, Wendy Jones, Carmen Cule, Joyce Butters, and Rosie Roberts.

Special thanks also go to Mrs Margaret Evans and Mrs Helen Williams for their kindness and assistance with my character reference letters.

Abstract

The research reported in this thesis addresses a patient's information requirements when searching the Internet for health information. A patient's lack of information about his/her health condition and its care is officially acknowledged and traditional patient information sources do not address today's patient information needs. Internet health information resources have become the foremost health information platform. However, patient Internet searching is currently manual, uncustomised and hindered by health information vocabulary and quality challenges. Patient access to quality Internet health information is currently ensured through national health gateways, medical search engines, third-party accredited search engines and charity health websites. However, such resources are generic, i.e. do not cater for a patient particular information needs.

In this study, we propose personalising patient Internet searching by enabling a patient's access to their Electronic Patient Records (EPRs) and using this EPR data in Internet information searching. The feasibility of patient access to EPRs has recently been promoted by national health information programmes. Very recently, in the literature, there are reports about pilot studies on personal Health Record (PHR) systems that offer a patient online access to their medical records and related health information. However, the extensive literature searching shows no reports about patient-personalised search engines, within the reported PHR prototypes, that utilise a patient's own data to personalise the search features for a patient especially with regard to health information vocabulary needs.

The thesis presents a novel approach to personalising patient information searching based on linking EPR data with relevant Internet Information resources, integrating medical and lay perspectives in a diagnosis vocabulary that distinguishes between medical and lay information needs, and accommodating a variable perspective on online information quality.

To demonstrate our research work, we have implemented a prototype online patient personal health information system, known as the Patient Health Base (PHB) that offers a patient a Summary Medical Record (SMR) and a Personal Internet Search (PerIS) service. PerIS addresses patient Internet search challenges identified in the project.

Evaluation of PerIS's approach to improving a patient's medical Internet searching demonstrated improvements in terms of search capabilities, focusing techniques and results. This research explored a new direction for patient Internet searching and foresees a great potential for further customising Internet information searching for patients, families and the public as a whole.

Contents

List of Figures

List of Abbreviations

AHA	American Heart Association
AHIMA	American Health Information Management Association
AMA	American Medical Association
API	Application Programming Interface
ASTM	American Society for Testing and Materials
CANISC	CAncer Network Information System Cymru
CDC	Centres for Disease Control
CDM	Canonical Data Model
CfH	Connecting for Health
CGS	Customised Google Search
CHV	Consumer Health Vocabulary
CIS	Cancer Information Strategy
CIU	Clinical Information Unit
CMP	Cancer Management Plan
CPT	Current Procedural Terminology
CSE	Customised Search Engine
CST	Cancer management plan Search Topics
CT	Concept Thesaurus
CWS	Charity Websites Search
DB	Database
DBA	Database Administrator
DBMS	Database Management System
DC	Dublin Core
DDE	Diagnosis Data Extractor
DITV	Digital Interactive TV
DOH	Department of Health
DRG	Diagnosis Related Group
DSM	Diagnostic and Statistical Manual of Mental Disorders
DST	Diagnosis Search Topics
DTE	Diagnosis Term Enricher

DW	Data Warehouse
EAI	Enterprise Application Integration
EHR	Electronic Health Record
EMR	Electronic Medical Record
EPR	Electronic Patient Record
ETL	Extract, Transform, Load
FA	Federation Administrator
FDA	Food and Drug Administration
FDBS	Federated Database System
FS	Full semantic Search
FS	Federation System
FWS	Favorite Websites Search
GALEN	Generalized Architecture for Languages, Encyclopaedias and Nomenclatures
GL	Gateway Link
GMN	Gabrrieli Medical Nomenclature
GP	General Practitioner
GS	Generic Search
GUI	Graphical User Interface
GwS	Gateway Search
HIQUA	Health Information Query Assistant
HL7	Health Level 7
HON	Health On the Net
HONCode	HON Code of conduct
HSR	Health Search Refinement
HTML	Hyper Text Markup Language
HTW	Hospital Trusted Websites
HWS	Hospital Websites Search
ICANN	Internet Corporation for Assigned Names and Numbers
ICD-9	International Classification of Diseases – Ninth Revision
ICD-9-CM	International Classification of Disease – Ninth Revision – Clinical Modification
ICDC	International Centre for Digital Content
ICPC	International Classification of Primary care
ICR	Integrated Care Record
IHC	Informing Health Care

IHR	Individual Health Record
ISCO	Information System for Clinical Organisation
JDBC	JAVA Database Connectivity
JSP	JAVA Server Page
LEAF	Layman Education and Activation Form
LS	Lay Search
MDTs	Multi-Disciplinary Teams
MeSH	Medical Subject Headings
ML	Macmillan List
MS	Medical Search
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NIH	National Institute of Health
NLM	National Library of Medicine
NPfIT	National Program for Information Technology
NS	Normal Search
ODBC	Open Database Connectivity
OMNI	Organising Medical Networked Information
PCASSO	Patient-Centred Access to Secure Systems Online
PDE	Patient Data Extractor
PDO	Patient Diagnosis Ontology
PDOS	Patient Diagnosis Ontology Server
PEC	Plain English Campaign
PerIS	Personal Internet Search
PFW	Patient Favorite Websites
PHB	Patient Health Base
PHBM	Patient Health Base Manager
PHIV	Patient Health Information Vocabulary
PHR	Personal Health Record
PSTC	Personalised Search Topic Constructor
RCV2	Read Codes Version 2
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SMC	Search Mode Controller
SMR	Summary Medical Record
SNOMED	Systemized NOmenclature of MEDicine

SS	Semantic Search
STM	Search Tool Manager
STR	Search Topic Refinement (or Search Topic Refiner)
STW	Staff Trusted Websites
SW	Semantic Web
SWS	Specific Website Search
TaO	TAMBIS Ontology
TL	Truste List
TOC	Table of Content
TST	Treatment Search Topics
UL	URAC List
UMLS	Unified Medical Language System
URAC	Utilization Review Accreditation Commission
URL	Uniform Resource Locator
VGS	Velindre Google Search
WGS	Web Google Search
WHO	World Health Organisation
XML	eXtensible Markup Language

CHAPTER 1

Introduction

1.1 Research Background

Information is an essential requirement for patients in order to manage various aspects of their healthcare, educate them about their health condition, and enable them to make informed decisions about their treatment [37, 103, 134]. An informed patient is more likely to be satisfied and respond better to the therapy [262, 281]. Several strategies have been outlined to ensure the availability and quality of information for patients during their healthcare journey [15, 37, 103, 115, 134].

Conventionally, patients obtain health information from several sources. There is the traditional face-to-face consultation with healthcare providers. More extensive background information is disseminated to patients through publications and the media [284]. Patients often report difficulties memorising and comprehending verbal explanations and indicate problems communicating with healthcare professionals [133, 195, 197, 231, 261, 275, 277] and receiving advice. Evidence suggests that patients only retain 10% of the information they receive at consultation [133]. In addition, public information sources address the average patient [195, 284] and do not give personalised advice.

Recently, numerous mechanisms have been implemented to offer supplementary authoritative information such as text messaging health tips or reminders [1, 245], copying clinical letters to patients [21], information prescription initiative [133] and health information systems in KIOSK (e.g. NHS Direct Kiosk [107]), digital TV (e.g. NHSDirect Digital TV [106]) and online (e.g. NHS Direct Online [108]) format.

Electronic patient information systems have claimed to have advantages over conventional sources as they deliver better information for patients focused on their circumstances [276]. Nonetheless, such systems must address the actual patient's needs to avoid failure [188, 262, 284] and without access to detailed information about the patient this is hard to achieve. According to Van't Riet et al [284]:

> "There are not many documented success stories about patient information systems. A core issue for such systems is their 'usability', which includes the extent to which the system takes the actual user's needs and capacities into account."

Tailoring to individual needs is referred to as personalisation. Hence, personalised patient information systems [175, 187, 225, 268, 286] emerged to cater for a patient's personal needs and should take account of their medical condition and health information requirements. However, patient information needs are perceived differently among the various health information stakeholders, such as patients, health professionals, and lay groups. In addition, patients often complain that their information needs are not sufficiently met [133, 293].

Healthcare professionals usually adopt a precautionary approach to patient information delivery. This includes authenticating external health information sources, limiting patient's online access to the medical records in a controlled way, and necessitating professional supervision over the patient's interpretation of both medical and external health information sources [268]. The professionals' attitude towards the use of such patient information is referred to by Dixon-Woods [195] as the paternalistic biomedical Patient-Education approach. This approach perceives patients as passive and incompetent when handling health information, and sees information as a way of correcting a patients' perception.

As a result, professionally-driven personalised patient information systems offer restricted access to the patient's clinical data and provide scientific medical knowledge which can be ambiguous to some patients as it is described in medical terminology. Additionally, it can be limited in scope, as it does not take into account a patient's particular information needs. Furthermore, the suggested health topics may not relate exactly to the patient's particular condition. Hence, identifying information on a patient specific condition is a lengthy process for a patient that requires browsing many related topics.

However, patients and lay groups call for a Patient-Empowerment approach [195] that promotes a patient perspective. It sees information as a way of empowering patients to make informed decisions and argues that patients may have greater capacity for self-control and information handling than is recognized by health professionals [195]. While recognising patient difficulty in understanding and interpreting information, the Patient-Empowerment approach calls for tools to be provided which aid a patient's understanding and use of information.

A move towards patient-oriented services and patient empowerment is now supported by the recent change in the healthcare delivery model from a disease-centred approach to a patient-centred approach [160], that tailors healthcare provision according to individual patient's needs. In addition, healthcare organisations are currently revolutionising their information infrastructure to a model based on rich information sharing and patient empowerment [105, 109, 257]. A key element of this on-going dramatic change in the healthcare sector is the development of the single integrated Electronic Patient Record (EPR) and a summary record for a patient's access known, in literature, as the Personal Health Record (PHR) [113, 117, 118, 151, 194, 199, 257, 285]. In Wales, EPR and PHR technologies are part of the NHS Wales Informing Healthcare (IHC) Programme[1] [109], that is developing the single integrated EPR for clinical use and the Individual Health Record (IHR) [71] for patient and legitimate professional's use. Patients will eventually access their IHR via a Web portal known as "My Health Online" [95]. The aim is to improve health information provision and services for a patient and allow for a better patient participation in healthcare. Thus, there are various moves towards a potential and increasing role for the patient in healthcare provision. This comes at a time when secure information and communication technologies are more available and are having a growing role in life.

The Internet has become the leading information technology and has changed the way people deal with health issues [188]. It offers patients unprecedented selective [292] and anonymous [207] access to a wealth of health information and services at their own convenience. In addition, the Internet spans a wide range of information providers and covers medical, cultural, and social information aspects in various vocabularies and languages. Surveys report high public access to Internet health information estimated as 80% in the USA [229], 27% in the UK [75] and 66% in Wales [145].

Despite its growing role, the Internet environment is inherently general and uncontrolled [188, 262]. Patients, using it, are presented with a vast amount of

[1] "Informing Healthcare is a National Programme to develop new methods, tools and information technologies to transform health services for the people of Wales." [72].

information that they need to assess for quality, interest, and relevance to their needs. In addition, Internet and medical vocabulary pose a challenge for a patient to express the correct form of a medical term and differentiate between related words in a search result. Nonetheless, future online health information content is likely to accommodate more patient-friendly information due to the emerging patient-centeredness approach, the Plain English Campaign (PEC)[2] [120] and numerous charity health websites offering simplified patient-oriented health information (e.g. cancerbackup.org.uk). Online health gateways[3] offer evaluated but generic health information mostly in medical terminologies that need to be simplified for patient understanding and according to a patient's particular needs.

In addition, different Web search mechanisms have emerged to facilitate and lessen the technicalities of online information search. Studies indicate that the majority of patients start their online session at a search engine [145, 206] and that online health searches mostly relate to specific illness and treatment [145, 229]. However, none of the existing Web search engines utilise the patient's own medical information from EPRs to personalise and focus online searching for a patient. EPRs model the actual and current patient health condition and can signify a patient's basic health information requirements during their health journey. Thus, merging EPRs and search engine technologies can result in a personalised patient online search experience.

This research primarily addresses the patient community at the Velindre NHS Trust[4]. It is concerned with improving a patient's access to online health information resources. Patient clinical information is recorded and managed by the Information System for Clinical Organisation (ISCO) [162]. ISCO is a health record system for cancer patients used across Wales. Currently, patients have no access to their ISCO medical records and search Internet information resources manually. However, in a lifelong health condition such as cancer, patients expect and demand easy and timely access to personal medical information and relevant educational health information.

---

[2] The Plain English Campaign (PEC) – a UK-based initiative aimed at establishing clear and understandable information from the first reading. PEC addresses lengthy sentences and technical jargons.

[3] A gateway is a website that acts as point of access or interface to one or more information networks.

[4] The South East Wales Cancer Centre, situated in Cardiff, United Kingdom.

The research is in line with the Welsh Informing Healthcare (IHC)[5] strategy [109] towards patient accessible personal health records and the growing use of the Internet for patient care. It investigates an approach which enables patients to access their basic medical information and utilise such information in personalising, guiding and enrching their online search experience.

1.2 Research Motivations

This research is based on the current status of the Internet as a secure and unlimited hub for health information, the increasing role of patients in the healthcare delivery model, the changing patient information needs and the health organisation's and professional's concerns about quality of Internet information sources.

By and large, the research is motivated by challenges underlying patients' health information acquisition, namely:

- Patients' lack of information [133] and demand for easy access to personal medical information [257]. Currently, Velindre NHS Trust lacks a patient accessible information resource that enables patients to view, comprehend and manage their medical problems [268].
- The information offered in leaflets and brochures is either general or limited [195, 241, 268].
- The diversity of patient health problems that a health information system needs to take account of when retrieving appropriate information within the needs identified in [268].
- Limitations of health information received from healthcare professionals [133, 134, 261, 268, 277], health publications and the media [195, 241, 268].
- The barriers that prevent patient's information needs being met during consultation, such as limited consultation time, and the patient's misunderstanding and anxiety [261, 268, 277].

[5] "Informing Healthcare is a National Programme to develop new methods, tools and information technologies to transform health services for the people of Wales" [72].

- The Web offers patients massive amounts of general but uncontrolled information of varying quality and reliability [188, 284].

- Electronic patient information systems are not tailored towards a patient's specific medical problems [276, 284].

- Different terminologies are used by different online health information sources, health professionals and the patients themselves some are technical others are lay [277].

- The health professional's requirement to guide a patient's access to reliable external health information [268].

- The patient's requirement for various medical, health and social information that relate to their medical condition [188, 229, 241].

- The patient's requirement to receive related health information in various vocabularies and languages that aids their understanding [164, 184].

- The patient's requirement to receive both generic and condition-specific health information [164].

1.3 Research Scope

This study addresses patient information needs when a patient searches the Internet for information related to their health. The Internet has great potential for patient care [134]. However, current online search is laborious and hindered by challenges related to information overload [185], quality [188, 262] and medical vocabulary [232].

The study sees a great potential in EPRs to define a basic patient personal information model that can be utilised and further enriched for personalising and simplifying a patient's online search experience. Modelling patient information needs based on a patient's own EPR and coupling this model with a variety of Internet information sources can help simplify, personalise and focus a patient's access to relevant Internet health information sources [153].

A review of the literature shows no previous attempts to personalise online search for a patient based on EPR or PHR data.

In line with the recent changes in national health information infrastructures towards Web-based PHR, this study presents a novel approach to improving patient online search experience that utilises a patient's own medical data to personalise the online search information topics and focus the search results for a patient. The study sees personalisation as the theme to unlocking many patient online search challenges and proposes to link EPR data to Internet search engines and health gateways [154]. In addition, EPR can bridge patient and professional perspectives with respect to health information vocabulary and quality.

In this study, we investigated building a Web-based patient personal health information system that enables a patient to access essential personal medical information and utilise this data to offer a patient a rich and personalised online search experience. In particular, the study investigates:

- Patient's EPR clinical data that a patient wishes to view and further explore in online search.
- Patient health information requirements during online searching.
- Personalising patient online search topics based on the patient's own diagnoses and treatment details.
- Utilising the patient's individual EPR data and the underlying medical encoding schemes to personalise, explain, and enrich health information vocabulary for a patient.
- A mechanism that delivers valid lay diagnosis terminology.
- Implementing a hospital-trusted websites list that guides patients to accredited online health information.
- Reducing a patient's online search technicalities by mediating the online search process.

While this study, technically, investigates personalising online information search for cancer patients at the Velindre NHS trust, our proposed approach is generic and disease-independent and can be applied to any patient community.

1.4 Hypothesis and Aims

The hypothesis that this thesis demonstrates is:

"Linking integrated Electronic Patient Records (EPRs) with Internet information sources enriches the patient Internet search environment and leads to an improved patient Internet search system when compared to traditional patient Internet searching."

To demonstrate the hypothesis, we aim to meet three objectives:

1. Personalising patient Internet information searching based on the patient's own medical information and health information requirements.
2. Simplifying and enriching a patient's medical search information vocabulary by use of a rich personal health information vocabulary utilising clinical data and the underlying data semantics, i.e., terminological relationships (e.g. synonyms, hierarchies).
3. Guiding a patient to quality Internet health information.

Objective one is concerned with the utilisation of a patient's own EPR data to personalise his/her Internet search features and requirements. Objective two addresses EPR medical vocabulary challenges. While Internet health information is written in mixed medical and lay vocabulary, EPR clinical data is usually described using a medical vocabulary defined by clinical terminology systems that could be incomprehensible to the less highly educated patient. Furthermore, patients indicate variable needs with regard to health information vocabulary in that skilled patients show interest in exploring medical literature using medical terms . This could be difficult for the average patient. Thus, linking EPR medical data to Internet search engines requires enhancing and simplifying EPR medical terms for a patient and enriching them with similar medical and lay terms in order to recover more of the related online health information and improve and focus search results for a patient.

Objective three ensures a patient access to safe and trusted health information as Internet health information comes from various sources some of which could be uncontrolled and unverified and may contain information which could harm patient care. Hence, a patient needs to be guided to "good quality" health websites and search engines, while not being prevented from accessing other sites if they want to.

1.5 Research Achievements

The importance of this research lies in its novel approach to personalising patient Internet searching by:

1) linking Internet search engines to a patient's own EPR data and related medical and lay terminology,

2) Utilising this data to guide a patient's search by enhancing information terminology and taking account of quality requirements to focus search results for a patient.

The integration of EPRs with Internet information sources offers a rich environment for addressing patient Internet search challenges and focusing the patient's online search experience using personalised search ideas and search tools. The thesis demonstrates the feasibility of building a patient Personal Internet Search (PerIS) service (see Section 8.32) within an official online patient health record system, so that healthcare officials' feedback and expert knowledge can be accommodated and communicated to the patient.

The thesis offers additional contributions to the Healthcare Informatics[6] field through:

1. A thorough and fresh exploration of: current patient information sources' limitations, patient Internet search challenges, and the landscape of the emerging health information infrastructures and programmes (see Chapter 2).

2. An exploration of patient clinical data that can be used to focus a patient's medical Internet search using EPR data and the underlying medical encoding terminology (see Section 2.7.4).

3. Extending the notion of the emerging PHR technology to incorporate a patient-personalised Internet search facility. The study developed a PHR prototype, referred to as the Patient Health Base (PHB) (see Chapter 7), that offers an online patient interface to the ISCO patient database. PHB offers patient-personalised services including SMR and personalised-search, i.e.

[6] Healthcare Informatics is a "field of study concerned with the broad range of issues in the management and use of biomedical information, including medical computing and the study of the nature of medical information itself." [267].

PerIS tool, and a mechanism for building a patient's Favorite Websites from a trusted online health websites list.

4. Building a Patient-oriented Diagnosis Ontology (PDO) that integrates diagnosis terminology from the medical and lay perspectives and accommodates patient information needs. PDO offers diagnosis term synonyms and hierarchies in medical and lay terminology (see Sections 6.4, 6.6 and 7.3.3).

5. A mediator architecture for linking (i.e. integrating) EPRs with relevant Internet information sources that alleviates patients as end global users from having to directly query the patient database or various Internet search tools. We present a mediator-based data-level integration architecture that links EPR medical knowledge to relevant Internet information sources. Our mediated architecture offers a set of tools for simplifying and enriching EPR medical terminology, constructing patient-personalised search ideas from a patient's own EPR data, and offering a single point of access to a set of key health gateways and patient-customised search engines (see Chapter 5 and Section 7.2).

6. Developing a generic mechanism for creating medical-to-lay diagnosis terminology based on a generic Concept Thesaurus (CT) system managed by an information specialist (see Section 5.4.1).

7. Offering a dual normal and semantic Internet search through which a patient can select no semantic features, specific or full semantics. For instance, a patient can select normal search only or a semantic search based on lay vocabulary or medical vocabulary or generic terms or all terms (see Section 7.4.3.4.5).

8. Developing a mechanism that aids hospital staff build a trusted website list using lists of third-party accredited health websites and patient-researched health websites (see Section 7.4.2).

9. Offering a patient a variable perspective of online information quality covering key health gateways, hospital-trusted websites, non-official charity health websites and a patient's own Favorite websites (see Section 7.4.3.4.4).

10. Offering patients and hospital staff a platform to share and communicate interesting health websites. A Hospital-trusted website's list is offered to individual patients in the patient interface, whereas a patients' Favorite website's list is communicated to hospital staff members through the staff interface, to utilise in building their own trusted websites lists (see Sections 7.4.2 and 7.4.3.3).

1.6 Organisation of the Thesis

This section presents an overview of the thesis organisation. The first chapter presented an introduction to the research by covering background on the research problem, research motivations, research scope, the research hypothesis, research objectives and main contributions of the thesis.

Chapter 2: *Patient Information Provision: Background*

This chapter explores the area of patient information provision, investigates a patient's information needs, examines the challenges underlying patient Internet medical search and explores the prospect of a patient's access to electronic medical records.

Chapter 3: *Data Integration and Semantic Interoperability*

This chapter reviews data integration challenges and approaches and examines Web data integration effects.

Chapter 4: Research Approach to Requirement Analysis

This chapter explores the system development methodology, the requirement analysis and elicitation approach and presents a defined account of the domain problems and system requirements.

Chapter 5: *The PHB Integration Architecture*

This chapter discusses the integration architecture used to integrate a patient's medical records with Internet health information sources.

Chapter 6: *The PHB Design Principles*

This chapter discusses the PHB design principles and logical foundations.

Chapter 7: *The PHB Prototype System*

This chapter explores the PHB prototype system's architecture, operations and implementation.

Chapter 8: *Research Evaluation*

This chapter evaluates the implemented PHB's functionality against the research hypothesis and objectives.

Chapter 9: *Conclusions and Future Work*

This chapter draws research conclusions and identifies future work.

CHAPTER 2

Patient Information Provision: Background

"The challenge for the NHS is to harness the information revolution and use it to benefit patients". The British Prime Minister addressing the All our Tomorrows Conference at Earls Court on 2 July 1998 [227].

2.1 Introduction

Chapter 1 introduced the research problem addressed in this thesis, namely, the patient's lack of health information and impediments in the patient Internet search process. We highlighted the challenges that patients experience, when receiving information from healthcare professionals or accessing supplementary information sources. We identified the extensive growth of Internet health information and the emerging changes in mainstream healthcare information infrastructures. We propose a new approach to improving patient information provision by enabling patient access to essential personal medical information in the patient's EPR and the linking of such medical information to Internet searching to guide a patient to relevant and good quality online health information sources.

This chapter reviews the landscape of patient information provision in official and public health information platforms to examine the underlying challenges that hinder patient access to medical and relevant health information and motivated this research. We start by highlighting the significance of information for a patient (Section 2.2). This is followed by exploring the limitations of common patient information sources (Section 2.3). As we propose linking patient EPR to the Internet, we examine the challenges and mechanisms concerning a patient's access to Internet health information (Section 2.4), especially with regard to information quality (Section 2.4.3) and vocabulary (Section 2.4.5).

Section 2.5 explores the prospect of the emerging PHR technology that enables a patient access to personal medical information and better participation in healthcare. Section 2.6 describes the Information System for Clinical Organisation (ISCO) that hosts the patient database utilised in this study. Section 2.7 investigates a patient's health information needs in order to personalise and focus health information searching for a patient. Finally, Section 2.8 concludes the chapter with a summary emphasizing the research problem investigated in this thesis and indicating the research approach to tackle it.

2.2 Significance of Information for Patients

Information is central to a patient's healthcare and wellbeing. As Jeff David [192] asserts: "Healthcare is information". Patients demand information to understand their medical condition, decide on appropriate treatment, get a second opinion, prepare for appointments, manage their emotions and social life, seek financial help, and enjoy a safe lifestyle [154, 182, 268]. Information can have direct and indirect benefits for a patient [134] such as:

- Reducing anxiety [171].
- Increasing a patient's confidence [293].
- Improving a patient's satisfaction with his/her care and adjustment to the diagnosis [275].
- Helping a patient to ask what they need [293].
- Some knowledge about a condition is beneficial in the time-constrained consultation [293].
- Enabling informed participation in decision-making [176].
- Enhancing compliance and satisfaction with treatment [197].
- Enhancing a sense of control [197].
- Management of chronic conditions [197].
- Helping them adjust to life changes.
- Key in achieving optimum health and well-being.
- Improving communication with healthcare providers [257].
- Can substitute for a doctor visit [293].

2.3 Patient Information Sources

Generally-speaking, patient information sources can be classified into interpersonal sources and mass media sources [241]. Interpersonal sources involve personal communication between a patient and the information provider. These include (but are not limited to):

- Healthcare Professionals (HCPs) (e.g. doctors, nurses) [241].
- Professions allied to medicine, such as physiotherapy [241].
- Paid online professionals [206].
- Voluntary organisations [241].
- Personalised patient information systems [284].
- Organisations in other countries [241].
- Self-help and Support groups [206].
- Other patients [206].
- Family and friends [176].
- Citizen's Advice Bureau [241].
- Clergies [241].
- Helpline [241].
- Email
- Cellular phone text messaging.
- Internet sources

On the other hand, mass media information sources disseminate general knowledge to the public that addresses the average patient. They include (but are not limited to):

- Hospital written information (e.g. leaflets, brochures, flyers, posters) [195, 241].
- Text-books [176].
- Medical journals, magazines and newspapers [176].
- TV, Radio, and Video tapes [241].
- Libraries [293].
- Electronic information systems [284].

- Digital Interactive TV (DITV)[7] [293].
- Touch-Screen Kiosks.[8]
- The Internet [261].
- WebTV.
- Cellular phones.

Interpersonal information sources are necessary to patients as they address the health condition and information needs of individual patients. A 2002 survey showed patients prefer interpersonal sources of information over written information [241]. However, interpersonal information sources are sometimes limited, inconvenient or inflexible for some patients due to time, geographical situations, personal reasons (e.g. embarrassment), and ethnic barriers. Furthermore, patients indicate they have problems memorizing or comprehending verbal explanations.

On the other hand, public information sources offer patients convenient, unlimited and diverse knowledge. However, public health knowledge is usually general and uncustomised, i.e., it will not usually address a patient's particular health problem and information requirements. Thus, the onus is on patients to explore multiple sources and locate the relevant desired information if it is available. In addition, a patient is responsible for assessing and interpreting the located information which is usually not a straightforward task. The following subsections review patient information provision by healthcare professionals, via printed media, and digital information sources which are the common means of accessing health information.

2.3.1 Healthcare Professionals

Traditionally, healthcare professionals communicate with a patient face-to-face, delivering patient-personalised information that is explained within the patient's medical context and understanding. In addition, healthcare professionals may communicate with patients by phone, letters, email or text messaging.

[7] Digital Interactive TV: Offers on-demand textual and audio/video information services via a digital TV interactive service or the Internet.

[8] Touch-Screen Kiosk: A standalone terminal that inputs and displays information via a touch-screen without the use of a mouse or keyboard.

There is a consensus among patients that information imparted by healthcare professionals is safe and trusted [231, 268]. For this reason, patients highly rank doctors and nurses as preferable information sources [231, 241, 261]. In fact, patients seek health professionals support at difficult times such as at diagnosis to get comforting and reassuring information, and when deciding on appropriate treatment to get authoritative and clear advice [231].

Nonetheless, information imparted by healthcare professionals can be incomprehensible to some patients either because it is stated verbally, possibly using scientific medical terms, or due to the situation in which it is received, e.g. when coping with shock relating to the first news of a health problem [268]. Availability of healthcare professionals also affects the amount of information patients get. A 1991 survey [229] found that more than half of the US population were dissatisfied with the availability of their doctors and the duration of meetings with their doctors.

Poor communication between a patient and healthcare providers is widely reported and can be attributed to:

- Limited consultation time [197].
- Patient's anxiety [261].
- Difficulty remembering information provided during consultation [197].
- Difficulty understanding physicians [261].
- Difficulty expressing feelings [261].
- Difficulty asking physicians questions [261].
- Cultural or language difficulties [197].
- Practitioners' failure to listen and respond to a patient's concerns [197].
- Doctors do not have all the answers [277].
- A patient's dissatisfaction with the given advice or diagnosis [277].
- Embarrassment when discussing sensitive topics [275].
- Healthcare officials lack of knowledge, experience and resources to provide support [275].
- Financial hardship induced in seeking consultation [231].
- Lack of a holistic approach: Doctors are not trained in counselling, may give outdated information or focus on certain conditions [231].

- Some health professionals are reluctant to give patients information as it may cause anxiety [231].

While a healthcare professional's role is vital, patients also need supplementary information sources [103]. Evidence suggests that patients only retain 10% of the information imparted during a consultation [133]. Accordingly, healthcare systems have adopted additional techniques to supplement communication between patients and healthcare professionals, such as:

- *Patient Information Pack [134]*: Each patient is given a folder of papers containing minimally, nationally-produced printed information, locally-produced printed information, and individual patient information on diagnosis, treatments, appointments, tests and key contact information.
- *Patient Information and Support Centre [134]* : Offers a visible contact point for people seeking information on health problems.
- *Web-based Messaging between Patients and Healthcare Providers [216]:* Enables a patient to exchange electronic messages with healthcare officials.
- *NHS Medical Information Systems (e.g. NHSOnline, accessed by the Internet (e.g. NHSOnline [110] and NHS24 [102]) or via Digital TV).*
- *The Copying Letters to Patients Initiative [21]*: An attempt to improve communication between patients and HCPs. Letters communicated between clinicians about an individual patient's care are copied to the patient. The aim is to keep a patient up-to-date with their diagnosis and treatment. Feedback back from patients shows appreciation of this approach [22].
- *Information Prescription Initiative [7, 112, 133]:* A new initiative by the UK Department of Health (DOH), launched in October 2006, where doctors prescribe information resources to patients that explain a patient's medical condition. Many health professionals in the UK and the US already offer "information prescriptions" to patients [257]. Moreover, the Centre for Information Therapy [16] offers online patient-tailored evidence-based health Information.
- *Patient Access to summary medical records [69, 216] commonly referred to as PHR [194, 257, 270]*: A relatively new but promising technology that

offers a patient electronic access to personal medical information (see Section 2.5).

With the adoption of the Internet in healthcare, new requirements have emerged for coordinating patient health information. This was also found in the US National Library of Medicine (NLM) initiative for establishing new healthcare professions: the Informaticist and the Informationist [250]:

- **Informaticist:** A health professional (e.g. nurse, physician, public health practitioner, librarian, computer scientist) who is trained in an interdisciplinary program of health sciences informatics.
- **Informationist:** An information specialist with training and experience in medical or biological sciences and in information sciences.

In addition, the ongoing shift in healthcare systems towards personal health records and Web-based communication with patients is more likely to open up a larger role for healthcare professionals in a patient's healthcare. This research anticipates advanced roles for healthcare professionals in supporting patient care especially with regard to the issues addressed by this research such as the construction of hospital-trusted health websites and lay patient information vocabulary.

2.3.2 Printed Health Information

Generally-speaking, patients find written information easier to comprehend than vocal information that is imparted during a consultation. Printed information acts as a source that a patient can revisit and share with family and friends at all times, especially after coming to terms with their diagnosis or health problems. 86% of the public show preference for written information at consultation [231]. Printed information available to patients comes in various forms:

- Hospital printed information, including (but not limited to):
  - Patient individual sheets [184, 263]: Documents that describe suggested treatments.
  - Leaflets, brochures, flyers, and posters.
  - Patient Information Pack [134] (see Section 2.3.1).
  - Healthcare letters copied to patients [21] (see Section 2.3.1).

- Medical and scientific prose publications such as textbooks, medical journals, magazines, and newspapers.

Hospital printed information can contain patient-specific (or tailored) information such as individual treatment sheets, healthcare letters or a patient information pack. However, other forms of printed information usually deliver common knowledge. A survey by Lewington and Farmer [231] showed the patients' preference for written information of an authoritative source. Printed information, though trusted, can bring some limitations, such as:

- It covers introductory knowledge on medical problems that could be too general for a patient to inspect and filter.
- It uses scientific or medical language that could be incomprehensible to the average educated patients.
- The leaflets are not detailed enough – they need to cover side-effects, descriptions of procedures and treatments [231].

With the increasing availability and use of the Internet in official healthcare, authoritative printed medical and health information resources are likely to be accessible online. Roberts [263] suggested incorporating a patient's individual treatment sheets in an official patient-personalised information system for faster and convenient access.

2.3.3 Media Health Information

Like printed information, media health information – delivered via various technologies (e.g. TV, radio, video) – offer introductory educational knowledge to the population. Additionally, they offer free authoritative health knowledge that is accessible from homes and can be watched with the family in a relaxing atmosphere. The media is a convenient means of raising awareness and educating the public about health problems and how to prevent them. However, information disseminated by the media is limited as it does not cover all aspects of concern to a patient (e.g. side-effects, details about procedures and treatment). A patient states:

> "if the TV could offer me information adequately, then I would go for the TV first" [231].

2.3.4 Digital Health Information

The recent advances in information and communication technologies have largely facilitated communication and information exchange among people. Digital health information denotes health information that is delivered via digital platforms (e.g. Internet, Web-messaging, Digital Interactive TVs (DITVs), Web TV, Touch-screen information kiosks, Cellular phones):

- *Web Health Information*: encompasses all textual, audio and video health information sources accessible via the Internet (e.g. NHS Direct Online). The Internet is a borderless and open information platform. It is, often, used as the first port of information [231] to access extensive knowledge. However, patients need to be guided about how to identify, assess and interpret relevant health information [188, 197] (see Section 2.4).

- *Web*-Messaging: enables interactive exchange of information between patients and their healthcare providers (e.g. MyChart [216]).

- *Web TV and Digital Interactive TVs:* offer a convenient way of delivering on-demand textual or audio-visual health information services via a digital TV or the Internet. The NHS Direct digital [106], for instance, offers general health and lifestyle information topics and information resources that a person can access via digital TV interactive services or the Internet. Some of these services can be customised to people according to their postcode (e.g. finding a local dentist).

- *Touch*-Screen Health Information Kiosks (e.g. *"InTouch With Health" [76], NHS Direct [107]*): offer a convenient means of communicating with people who do not have access to information at home or work without the need for supervision. It can be located in public areas such as patient waiting areas in a surgery or hospital to educate patients about their health problems and consultations.

- *Mobile Phone Text Messaging:* utilises Mobile phones as the leading communication technology to improve health information delivery for a patient. Patients receive text messages containing health information such as healthcare tips, safety messages, appointment or daily treatment reminders [1, 245]. A pilot project [1] to remind patients about their appointments proved to be time and cost effective for the NHS, and resulted in a 30% drop in the patients who missed their appointments.

- *The Patient Health Smart Card [254]:* A patient photo identification card with embedded chip that can store patient information. The card is held by the patient and is routinely updated by healthcare professionals with legitimate access to patient care. The Patient Health Smart Card stores key patient information such as demographics, allergies, current medications, laboratory results etc. The card can be read with a device attached to a computer. The technology of smart cards offers a cheap and flexible alternative to PHR technology (see Section 2.5). It ensures access only by patients and legitimate healthcare officials. Patients say they can carry it at all times and they no longer have privacy concerns. However, there is a concern that smart cards may get lost by patients [30].

Due to its increasing impact in today's life, the digital media can play a major complementary role in delivering authoritative health information for a patient, thus, backing up communication between patients and healthcare officials. The Internet excels as being a massive and leading means of disseminating information, capable of delivering unlimited and extensive information in an anonymous, convenient and borderless manner.

2.4 Patient Internet Access

> "There is some excellent information on health care and medicine available on the Internet, but the problem is how to find it." The British Library [46].

The Internet has become the biggest and foremost medical library in the world. The immediacy and ubiquity of the Internet makes it easily accessible to a multitude of patients, especially in remote places and disadvantaged communities [171]. The wide range of Internet technologies (e.g. the Web, Blogs, Email, discussion groups, news groups, video conferencing and digital TV) offers various platforms to support and maximise healthcare support for a patient. In addition, the Internet offers a scope for modernising mainstream health services and improving communication with patients especially for information sharing and delivery, e.g. supporting online patient-oriented health services (such as online prescription renewal, appointment booking and email reminders) and a secure patient online access to medical records.

The integration of the Internet in healthcare is widely recognised [103, 171, 211, 280] and appreciated by both officials [103, 133, 134, 164] and the public [75, 207, 229, 231]. The Internet promises numerous benefits for patient care, namely:

- *Access to extensive information on healthcare* [103].
- *Anonymity of access*: The Internet enables a patient to get answers to sensitive (or embarrassing) issues without involving other people.
- *Availability of information*: The ubiquity, convenience and immediacy of the Internet enable patient access to information whenever and wherever they need it. This is vital as patients have difficulty remembering information [171].
- *Culture of partnership*: By involving patients in their healthcare as partners rather than receivers [171, 178], it allows incorporation of the patient's perspective in decision making with regard to treatment, information delivery method and customised services' content.
- *Uniting patients* [171]: Bringing patients together with the same condition.
- *Active communication and information sharing*: Patients can communicate with other patients and health officials through email, discussion and support groups [103].
- *Cost reduction*: Communication through Internet services (e.g. Web, Email) can reduce cost for both patients and professionals [171].
- *Customisation*: Tailoring information for individual patients by allowing a patient access to their medical records [171].
- *Modernising health services*: Early patient access to Internet health information encouraged professionals to do the same [280]. In addition, the inclusion of the Internet in healthcare provision led to dramatic changes in healthcare system operations such as integrated records and online patient services.

On the other hand, the literature identifies limitations and challenges concerning the use of the Internet for patient care, namely:

- *Inaccessibility:* The Internet could be inaccessible to some patients who either can not afford it, lack the skills needed to use it [171] or do not want to use it [75], especially older age groups. This could create information inequality and widen the information gap among patients [171].
- *Information overload and lack of organisation*: The proliferation of Internet information and the unorganised nature of the Internet make it difficult and frustrating to locate the exact information sought especially for a patient who may be stressed [46, 171].
- *Multiple vocabularies*: As the Internet is open globally, it hosts different perceptions and vocabularies from different communities. This makes it difficult to locate relevant information due to the related but different vocabulary being used. The literature describes numerous techniques to help alleviate this problem (see Section 2.4.5).
- *Cultural drawbacks*: The Internet has little recognition for cultural diversity and is mostly available in English [171] which makes it less advantageous to under-represented and/or non-English communities.
- *Ethical drawbacks*: such as forged identities, dishonesty and lack of confidentiality due to technology limitations affect its use [171].
- *Unreliability*: Concerns due to the lack of standardization, difficulty in judging quality, differentiating between the notions of education and promotion, and lack of guarantees as to appropriateness currency or information timeliness [171].
- Excessive Information: The extensive Internet information means that often a patient gets a large amount of information that does not apply to his/her situation.
- The threat to human communication that is fundamental and supportive to primary care [178] by replacing it with impersonal online communication.

The Internet offers a complementary and modern means for supporting healthcare information that is much needed in today's life. It is not meant to replace the personal communication between a patient and healthcare providers. Internet access has recently been boosted by innovations in information and communication technologies, especially broadband technology and state-of-the-art security schemes. Surveys [75, 145, 229] report an increase in the Internet population (see Figure 2.1). Online health information is very much appreciated by the public and patients. After surveying the Internet for information on her condition, a patient with restless leg syndrome told her doctor:

> "This is me! If I didn't have access to the Internet, then I would never find out about my true condition." [293]

Feature	**US Pew Internet 2006 [229]**	**UK National Statistics 2006 [75]**	**Welsh Informing Healthcare 2006 [145]**
Internet Access			
Internet population	70%	57% [75], 60% [138]	78%
Broadband Internet access	86% (at home)	69%	71% (at home)
Internet access from home	-	85%	64%
Internet Health Information Access			
Online health information seekers	80%	27%	66%
Internet Access Method			
Search engine	66%	92% (for general search in 2005)	81%
Health-related website	27%	-	12%
Health Information Type			
Seek information on illness or condition	64%	-	54%
By Health Status			
Internet users not in good health	-	-	72%
Internet users in good health	-	-	85%
By Age			
Internet users of age 65+	68%	15%	45%
Internet users of age <=64	80%	71 %	70%

Figure 2.1: Internet Access Statistics in the UK and USA (Adapted from [75, 138, 145, 229])

The 2006 Pew Internet Report [229] points out that Internet health information access by the respondents in the USA has been stable over the last five years at 80% [229]. Similarly, a survey by the Welsh Informing Healthcare (IHC) program [145] revealed that 84% of the (1002) respondents have access to a computer and 78% have Internet access. However, the UK National Statistics Office reported a low access of about 27% for online health information across the UK in 2006. Mair and Kierans [234] attributed the low access by the UK public to online health information is due to the UK, unlike the USA, having a well-established national health network with specialised NHS health information services. However, this suggestion is refuted by the IHC survey as only 12% access the NHS website in Wales in 2006 compared to 81% using search engines.

The IHC survey [145] indicates that the "lack of interest" is the major reason (53%) for not accessing the Internet while cost (7%) and security (2%) were considered minor barriers. In addition, 66% used the Internet to obtain health information, with the most common reason (54%) being to explain an illness or a condition. While it is feared the Internet health information could be disadvantageous to elderly patients, due to low Internet access rate of users over 65 (see Figure 2.1), the 2005 National Statistics survey showed that 41% of the people aged 65+ shop online [234]. Furthermore, 24% of non-Internet users ask someone else to access online health information on their behalf [145]. In addition, the Wanless Report [287] warns that the aging UK population will put a strain on the NHS, and advocates adopting self-healthcare schemes [293], a move that the Internet can easily support.

As far as readability is concerned, the majority (50%) of health seekers found the Internet quite easy to use but only 34% found it very easy to use [145]. The most important criteria for selecting health information is that it should be "clear and easy to understand" (74%) and accurate (73%) while only 49% preferred official health information [145]. According to Theodosiou and Green [280], some patients attempt to avoid mainstream health websites and instead seek lay websites "either because they are wary of the motives of mainstream medicine or because they are searching for information that will be easier to read" [280].

Absolute security is not guaranteed on the Internet [193]. However, currently, Internet access is regulated by state-of-the-art security measures and privacy controls [117] which is reflected in the minimal security concern (2%) among health seekers [145]. In addition, the growth of online mainstream health gateways and charity websites offers reasonably high quality health information for a patient. However, identifying valid and relevant Internet health information is still a problem [46] due to the lack of organisation, information overload, quality and readability of the information. Lack of organisation and information overload is addressed in literature through techniques such as websites categorisation, subject directories, subject search engines, and tailoring information to a patient by allowing access to a patient's medical records [154, 171].

Online information quality is examined in Sections 2.4.3 and 2.4.4 whereas issues concerning information readability and vocabulary are explored in Section 2.4.5.

2.4.1 Patient Activities on the Internet

A patient using the Internet for healthcare information is referred to in the literature by several terms such as e-Patient [206], Online Patient and a more general term Health Seeker [229]. The process of utilising the Internet for medicine and healthcare resulted in movements known as e-health, e-healthcare, e-medicine, telehealth, and telemedicine. Definitions of such terminologies are somewhat ambiguous [262], but illustrate the diversity of terms for concepts in English.

The Ferguson Report [206] describes a ten-level activity schema that e-patients do online that can be further summarised into five activities:

- Searching for health information.
- Communicating online via e-Mail or discussion groups.
- Joining health research activities of shared concerns.
- Using online Medical Guidance Systems[9].
- Using paid services of online medical advisors and consultants.

Additional online patient activities are anticipated in the emerging national health information infrastructure (see Section 2.5) which is developing a Web-based lifelong medical record system for patients to access. Initial prototypes demonstrate online patient activities such as:

- Viewing (and possibly annotating) essential medical information (e.g. condition, medications, allergies, appointments, tests).
- Requesting appointments and referrals.
- Renewing prescriptions.
- Maintaining diaries.

Thus, there is a diversity of health activities undertaken by people on the Web.

2.4.2 The Internet in Official Healthcare

The early Internet health information was published by independent organisations and individuals, offering healthcare information and products of variable quality and safety [280]. The mainstream healthcare providers were initially slow to

[9] A Medical Guidance System is a system that uses computing power to help e-patients make good medical decisions [206].

endorse Internet technology [280]. This could be due to the global and uncontrolled nature of the Internet that raises safety and security concerns regarding a patient's access to false or invalid information. In addition, the use of computers and the Internet in healthcare was seen by professionals as an extra load on people working in an already demanding profession [178].

However, the proliferation of health information on the Internet and the large increase in the number of patients accessing online health information impacted on the healthcare system as patients started taking online information to consultation rooms to discuss it with the clinicians [280]. Theodosiou and Green [280] suggest educating people in the medical sector about Internet access through debates, undergraduate syllabus and professionals' awareness of online health information in specialised areas.

The increasing recognition of the Internet by official healthcare providers is, recently, outlined in a number of official documents [103, 134, 172, 227]. Initial attempts to integrate the Internet in official healthcare provision resulted in the development of health organisation websites. Subsequently, official health gateways were established such as NHS Direct Online, Dipex and the National Electronic Library of Medicine that offer educational information for clinicians, patients and the public. Significant health information is also accessible through voluntary organisations such as charity websites (e.g. Cancerbackup.org.uk). As the Internet continues to revolutionise the healthcare system, new information infrastructures are being developed to enable a patient online access to personal health information and services (see Section 2.5).

2.4.3 Internet Health Information Quality

There is a major concern about the quality of Internet information especially when it relates to health issues [121, 201, 211, 262]. Misinformation can have adverse implications on patients, especially with life-threatening conditions such as cancer [202], namely:

- An action that disrupts the official treatment (e.g. taking unprescribed substance/drug) [202].

- Abandoning a high-quality healthcare provider to pursue ineffective therapy [202].

- Using limited consultation time unproductively [202] discussing extensive online health information some of which is irrelevant.
- Delaying consulting a doctor.
- Losing trust in healthcare provider [202].

Even high-quality Internet health information can be unintentionally harmful when it is out-of-context, outdated, unavailable, inaccurate, incomplete, biased and difficult to understand [262]. Health organizations such as the American Medical Association (AMA) called for Internet regulation [280] to safeguard patients against invalid and harmful information. However, Eysenbach [200] deems Internet regulation as unrealistic as it contradicts the very global and open nature of the Internet. In addition, websites appear and disappear constantly and unpredictably. On the other hand, Wotton [295] argues that restricting and regulating health information accessible to a patient underestimates the patient's potential as a health consumer and contributor in treatment by failure "to distinguish between quality of information and quality of knowledge" [295].

Quality experts have set up four pillars for ensuring patients can access high-quality Internet health information [202]:

- Educating consumers on how to locate and judge "good quality information".
- Encouraging health information publishers to adhere to ethical standards and codes of conducts, i.e., self-regulating and self-labelling.
- Applying independent third-party evaluation of health information and making it available.
- Enforcing existing legislation in the case of false or harmful information.

The literature [188, 201, 262, 280] discusses different techniques for regulating and evaluating Internet health information, namely:

- **Recommended Principles and Codes of Conducts [188, 280]**: A set of rules used by health information authors to ensure that their website content adheres to ethical principles (e.g. [23, 34, 35, 45, 123, 143]).
- **Explicit Evaluation Tools [188, 280]:** A set of criteria (or questions) that can be used by both users and authors to assess the quality of the health information (e.g. [29, 94, 100]).

- **Third-party Certification [262]:** Websites wanting to indicate the quality of their information content to users can use quality seals (or trustmarks) issued by third-party accreditation organisations after the evaluation of the website. A quality seal indicates the adherence of a website to the quality criteria set up by the accreditation organisation. However, third-party certification requires continuous monitoring. Among the organisations offering quality seals are the Health On the Net (HON) Foundation [51], Trust-e [135], and URAC [139]).
  - **HON [51]:** Launched in 1996, HON is a free self-certification organisation based in Geneva, Switzerland. HON offers the HONCode accreditation seal [62] to sites adhering to its accreditation criteria, which is based on the principles summarized by Risk and Dzenowagis [262]. Additionally, HON offers Web search services including the HONCode search engine and MedHunt [88]. The HONCode engine searches over 5000 HON accredited websites in 70 countries [63]. MedHunt is a search engine provided by HON that searches only for sites that are relevant to the health and medical fields, as well as the HON database for medical sites, for hospitals, and support groups conforming to the HON Code of Conduct for health Internet sites [89].
  - **The Utilization Review Accreditation Commission (URAC) [139]:** Commonly known as the American Accreditation HealthCare Commission, URAC aims to improve the quality and accountability of health care organisations using utilisation review programs. URAC has over 16 accreditation and certification programs. The "Health Web Site" accreditation seal assures that a company's Web site is trustworthy and meets URAC's quality standards. URAC certification is first performed by URAC accreditation staff and then by the URAC Accreditation Committee and Executive Committee.
  - **Trust-e [135]:** Based on privacy for personal information on the Internet, Trust-e offers several accreditation seals (e.g. Email Privacy Seal, E-Health Seal). The "E-Health" seal is awarded to companies that meet strict standards of online health information privacy, reinforce a trusting relationship with consumers, and submit to Website reviews.
- **Metadata-Based, Semantic Web Evaluation Tools [188, 201]:** In order to automate the discovery of trustworthy information resources, the Semantic Web

(SW) [141] approach is used to build knowledge-based evaluation tools that specify the evaluation of an information resource using a metadata vocabulary (a common vocabulary) that can be read and interpreted by computers. Such evaluation tools aim to enable an information consumer, using a browser (or client-side application), to filter their information requirements by utilising the metadata describing the information resource. Examples of Metadata-based evaluation tools include (but are not limited to) MedCERTAIN [202], its successor MedCIRCLE [201] and QUATRO [235]).

In addition, online access to quality health information is ensured through mainstream medical websites, dedicated medical search engines (e.g. HON MedHunt), health gateways (e.g. MedlinePlus) [188] and health charity websites (e.g. Cancerbackup). Furthermore, the World Health Organisation (WHO) proposed, in 2000, to set up a new domain ".health" for approved health websites. The proposal was rejected by the Internet Corporation for Assigned Names and Numbers (ICANN) as this would give WHO control over Internet health information [280].

2.4.4 Internet Search Mechanisms

The Internet has led to the introduction of numerous means for providing and accessing health information: The Web, newsgroups, Email, support groups, discussion forums and Blogs. The Web constitutes the largest Internet information service ever seen. A patient can access Web health information by several means, namely:

- **Search engines (e.g. Google):** offer open access but unverified results.
- **Subject directories (e.g. Yahoo):** categorise websites according to search topics for fast access.
- **Health gateways (e.g. MedlinePlus)**: aimed mostly at professionals but ensure high quality information.
- **Charity Websites (e.g. Cancerbackup):** voluntary organisations offering independent and largely easy to understand health information for patients and family.

Though retrieving unverified information, search engines constitute the foremost means for accessing Internet health information according to surveys by the Pew Internet Project [229] and the Welsh Informing Healthcare Programme [145]. Health gateways offer quality information that is largely aimed at professionals. However some patients deemed gateways like MedlinePlus credible to access [4]. Charity websites offer non-official specialised knowledge that is often offered in a patient-friendly vocabulary conforming to the Plain English Campaign [120].

As most people seek online health information using search engines, finding relevant and trustworthy online health information becomes a difficult task [191]. The following subsections explore three types of Web information access mechanisms utilised in this study: search engines, health gateways and charity websites.

2.4.4.1 Web Search Engines

A Web search engine is an automated program (also called robot, spider, worm) that constantly indexes Web resources and allows searching of its index [158]. A review of literature reveals five categories of Web search engines that can be used for health information seeking, namely:

- **General Purpose Search Engine:** a free-text search engine that indexes and searches the entire Web (e.g. Google, Yahoo, Lycos). It takes a search phrase and returns Web documents containing keywords of the search text. This study utilises the Google search engine as one of the simplest and popular search engines [128]. Its Web crawler employs the "PageRank" technique that ranks search results according to the number of websites that link to them [121]. Thus, the first 10 Google search results represent the most referenced pages for the search. Google crawler is also used by the Yahoo search engine [121]. According to Al-Ubaydli [158] Google "can provide, quickly enough, an answer that is good enough". [158] summarises Google features that can be used to improve access to clinical and health information on the Web.
- **NicheSearch [232] (Or Dedicated Medical Search Engine [280]):** involves selected Web resources targeted for a particular audience (e.g. Intute: Health and Life Sciences: Medicine (formerly known as OMNI) [77] for professionals and researchers, and WebMD [142] for health consumers). While such search services ensure quality, they constrain knowledge and are likely to contain high-

level medical vocabulary even if intended for patients and the laypeople (e.g. WebMD). Additional medical search engines include (but are not limited to) HON MedHunt [88], Kosmix [83], Healthline [55], and Mammahealth [86]). This study utilises the MedHunt search engine [88] (see Section 2.4.3) due to HON's credibility.

- **MetaSearch**: represents intelligent health information retrieval systems (e.g. HealthCyberMap [223]) that encode concepts within the health information resources to identify the semantics (or meanings) of terms within the raw text. It makes use of terminological and ontological relationships to label and tag Web resources in order to establish a relationship between tagged Web resources and facilitate the retrieval of related resources. While this approach is convenient for a limited collection of resources, it is unrealistic when considering the entire Web.

- **Semantic Search Engine [232, 280]:** a promising but largely untapped technology based on the concept of the Semantic Web (SW) [141]. SW creates a Web environment where Web content is meaningful to computers. This automates the processing and interpretation of Web information by software agents. For the healthcare domain, such a feature would enable the possibility of identifying and mapping between a consumer's lay vocabulary and a provider's clinical vocabulary. It utilises a thesaurus or ontology to interpret and reformulate the user query in terms of the words held in the thesaurus or ontology conceptual knowledge. Unlike general purpose search engines, semantic search engines are domain-specific (e.g. medicine, law). Most semantic medical information retrieval systems are aimed at clinicians (e.g. MELISA [147], HealthCyberMap [223]), utilise a collection of medical resources, and employ a medical vocabulary (e.g. ICD-9 [73], SNOMED [127]). While clinical information systems ensure high-quality information, their information is largely professional-oriented and the underlying vocabulary could be difficult for some patients or laypeople to manipulate. Woods [294] describes problems in some of the thesaurus systems such as UMLS Metathesaurus (such as lack of terms, lack of synonyms, and multiple terms for the same concept). Westberg and Miller [290] attribute the failure of clinical information systems to the fact that users use improper search terms and fail to select relevant data. They argue that "more user-friendly applications would allow for greater and more relevant retrieval" [290].

There is very little literature on patient-oriented semantic search engines. McRoy et al [238] describe an ontology-based information system for patient education called the Layman Education and Activation Form (LEAF). LEAF analyses a patient's medical history form and uses a natural language dialogue with patients. LEAF offers a patient relevant and personalised medical information based on their medical history.

- **Peer Mediated/Peer Validated (PMPV) [232]:** These are specialised search engines that rely on intermediary intervention assisted by computer technologies. Health information seekers submit their queries to a self-service mediated search engine using email or the Web and subsequently receive relevant medical resources. Query formulation and processing is performed by information specialists. This is a paid service that is welcomed by relatively advanced users due to its convenience, time saving and retrieval of specialist knowledge (e.g. ILIAD [67]).

2.4.4.2 Health Gateways

Health gateways offer free access to a catalogue of official medical and health information. A health gateway can be searched by browsing subject categories or through a keyword search. This study utilises the following health gateways:

- **NHS Direct On-Line [108]:** A UK government health gateway offering access to high quality information and details of NHS services. It is geared "to enable patients to make decisions about their healthcare and that of their families" [108]. NHS Direct Online is accessible via a keyword search engine or by browsing topics using an A-Z index. Apparently, NHS Direct Online retrieves related resources using a different vocabulary but using exact phrases. For instance, searching for phrases such as "cancer of the kidney", "renal cancer" and "renal carcinoma" goes to the NHS Direct Online Health encyclopaedia topic "Cancer of the kidney".

- **Cancer Specialist Library [12]:** A comprehensive evidence-based specialist cancer information resource developed to support health professionals but it also welcomes patients, families, carers and the general public. It is part of the National Library for Health (NLH) specialist libraries for cancer. It offers a free-text search service.

- **MedlinePlus [91]:** A US government Web-based service that brings authoritative information from the US National Library of Medicine (NLM), the National Institute of Health (NIH), and other government agencies and health-related organizations. MedlinePlus can be beneficial for all types of users seeking health information. It offers easy access to medical journal articles and extensive information on drugs, a medical encyclopaedia, and latest health news. MedlinePlus health topics can be searched using a free-text search facility or by browsing an A-Z of health topics categorised using MeSH terms.
- **The Cochrane Library [19]:** An international organization that offers up-to-date evidence-based information about the effects of healthcare in order to inform decision-making. It addresses different categories of users: clinicians, policy makers, researchers and patients. The library can be searched using either simple search terms or MeSH Terms.

2.4.4.3 Health Charity Websites

These are voluntary health organisations offering advice and information to patients and their carers that are both independent and patient-focused. Health charity websites are usually disease-specific (e.g. Cancerbackup [13], British Heart Foundation [9]). The Cancer Information Strategy (CIS) [134] of NICE [144] advocates partnership with charity websites to benefit from their specialist knowledge and expertise. In addition, Roberts [263] recommended patients' access to Cancerbackup – a leading cancer charity website.

This study utilises a list of generic and cancer charity websites that is recommended by the Macmillan Cancer Support [85] and is used at the Patient Information Centre at Velindre NHS Trust to guide patients to key health websites.

2.4.5 Health Information Vocabulary

Despite the growing use of the Internet for health information, much of online health information is written by professionals using medical terms. Patients, often, find it difficult to understand clinical terms and explanations, and interpret them differently according to their personal cultures, experiences, education level and understanding [274]. A study by the US Institute of Medicine [50] found that almost half of American adults have a problem understanding health Information.

In addition, Ownby [251] evaluated 60 health websites on "depression" and found their content to have language above the average reading level [239].

However, websites conforming to the UK-based Plain English Campaign (PEC) guidelines [120] offer clearer Web information that ensures its users have clear content. This process requires constant review. The current PEC list of websites holding the Crystal Mark includes (but is not limited to) the NHS England Connecting for Health (CfH) [105], cancerhelp.com and best-treatments.com.

In addition, the adoption of a patient-centred approach to healthcare in recent years has led professionals to use a language that can be understood by patients [249]. Nonetheless, Butters [184] indicated that some patients complain about the simplicity of patient health information and show preference for reading medical and scientific health information. This view is also shared by the Patient Empowerment movement [195] that sees restricting health information for a patient underestimates the real potential of patients as equal partners in healthcare. Figure 2.2 describes patients' view on the use of health terminology during consultation as reported in [249].

Terminology Type	Patient's View on Health Terminology Usage
Medical	• Medical terminology indicates that the problem is taken more seriously. • Patient would be allowed time off work. • Problem has a definite cause. • Patient feels more confident in the doctor. • Patient is more satisfied with their visit to the doctor. • Patient feels more frightened or anxious. • Patient shows greater understanding of the problem.
Lay	• Lay terminology implies that patients can take care of themselves. • Problem would not last very long. • Problem brought on by the patient.

Figure 2.2: Patients' View on the Use of Health Terminology [249]

With regard to vocabulary, the Internet brings multiple user levels which include information providers, consumers and system designers [232]. Also, health information is consumed by different users. Miller et al [239] describe three health information consumer categories:

- *Patients*: people with minimal familiarity with medical text,
- *Professionals:* people with medical training, and

- *Novice health learners:* people with the desire to learn medical terminologies from educational material.

However, a patient can also be a professional or novice health learner. Thus, a better categorisation could be into professionals, laypeople and novice health learner, where laypeople are people with minimal medical training or knowledge.

Medical terms are usually drawn from a diverse collection of medical terminologies, due to the lack of a standardized medical vocabulary (Figure 2.3 shows some of these sources). However, Lorence and Spink [232] argue that standardized terminology remains uncertain in a Web environment, due to the dynamic and global nature of the Internet.

Acronym	Description
CPT	Current Procedural Terminology [24]
DRG	Diagnosis Related Group [27]
DSM	Diagnostic and Statistical Manual of Mental Disorders [28]
GMN	Gabrrieli Medical Nomenclature [177]
GALEN	Generalized Architecture for Languages, Encyclopaedias and Nomenclatures [296]
ICD-9-CM	International Classification of Disease – Ninth Edition – Clinical Modification [73]
ICPC	International Classification of Primary Care [74]
MeSH	Medical Subject Headings [90]
RC	Read Codes [247]
SNOMED	Systemized Nomenclature of Medicine [127]
UMLS	Unified Medical Language System [137]

Figure 2.3: Major Medical Classification Systems (Adapted from [157])

Generally-speaking, use of a medical vocabulary serves the needs of healthcare professionals but imposes challenges to patients and laypeople using these information sources, as it is hard for them to:

- Locate the desired health information.
- Understand professionally-written health information.
- Express the correct medical term when formulating a query [232].
- Estimate the semantic relationships among similar or related terms in a search result, i.e., how semantically close these terms are [232]. For instance, a patient needs to understand that "renal cancer" is synonymous to "kidney cancer" which is more specific than "urological cancer".

- Differentiate terms from multiple medical encoding systems, due to the lack of a standard medical terminology.

Lerner et al [230] examined a patient's understanding of common medical terms and found that 79% of participants could not recognise analogous terms, such as 'bleeding' versus 'hemorrhage'. He concludes that "medical terminology is often poorly understood by young, urban and poorly educated patients" [230]. Such difficulties in understanding medical terms and expressions are also experienced by elderly patients [293].

Ogden et al [249] indicate that terminology challenges are major problem when it comes to explaining a diagnosis. Due to a lack of medical knowledge, patients and laypeople tend to use popular lay terminology to express a concept of illness or subject of interest [232]. However, lay terminology can only identify part of the desired health information and may lead to misleading or irrelevant information [232]. There are a number of techniques to address the challenges set by medical terminologies, namely:

a. Establishing term definitions for medical terms [232]. However, it is difficult to represent such definitions in a formal form which allows automatic translation. This means, translation must involve human intervention which must be minimized [232].

b. Guiding patients and laypeople to sections (or categories) of interest [240] possibly by use of lay labels. This can reduce the search time but may mean the information needed is not included or properly labelled.

c. Development of an intermediate terminology layer that maps between laypeople and professionals' terms [239, 271, 282]. In building this mapping the professionals' terminologies are usually drawn from known medical encoding systems (see Figure 2.3). However, the laypeople's terminology has to be constructed before the mapping is built, and this can be done by:

    - Laypeople (or their representatives) specify terms they use to describe various medical details. This can be established beforehand as in [240] or at the query formulation time as in [149, 238].
    - A patient information expert builds a consumer-friendly vocabulary from the laypeople's perspective as in [239].
    - A combination of laypeople and information specialists interact to create the list.

Mapping between medical and lay terminologies is widely used (e.g. [149, 238, 240, 298, 299]). The intermediary layer approach offers a method to establish a rich vocabulary which can support patients and laypeople in query formulation and enhance their understanding of medical knowledge. Lay vocabulary is usually used to help a patient formulate queries, while the medical vocabulary is used to identify medical Web content. There are no efforts yet to combine both lay and medical vocabulary for a patient to view and utilise in query formulation. A study by Abidi et al [149] demonstrated that reformulating lay queries into MeSH-based queries resulted in a less effective search results.

In addition, lay terminologies and expressions stem from a patient's medical condition and vary according to cultural and personal context. Zeng et al [297] note that patients use specific terms that relate to disease, syndromes or body parts. Hence, establishing a personal health vocabulary can better reflect a user's preferred health terminology. Tseab and Soergela [282] point out that personal health vocabulary is still a recent topic. The following list gives a sample of terminology-based patient-oriented information retrieval systems:

- **The Layman Education and Activation Form (LEAF) [238]:** an ontology-based information system for patient education. Patients fill a Web-based medical history form specifying their health problems. LEAF analyses the form and extracts terms describing medical details and uses an ontology to retrieve medical and health information related to these details. For instance if a patient specifies a health condition like 'heart disease' or 'estrogen', LEAF returns the same material for both terms.

- **HealthCyberMap [223]:** indexes and stores Web medical resources in a database. HealthCyberMap resources are described in a resource metadata format based on the Dublin Core (DC) [31]. The DC subject field is described in clinical encoding such as ICD-9-CM [73]. HealthCyberMap is searched using ICD-9-CM terms. In addition, its Problem-Knowledge Coupling service [222] links HealthCyberMap medical resources and medical records by employing the same clinical encoding or different encodings with reliable mapping facilities.

- **Consumer Health Vocabulary (CHV) [299]:** contains terms commonly used by a well-defined consumer group to express related terms. However, such a common terminology may not be adequate or detailed enough to express a patient's diverse health problems and health information requirements.

- **Table of Content (TOC) [240]:** combines the resource categorization and terminology mapping techniques. It categorises health information resources that are extracted from the WebMD [142] website using lay or consumer-friendly labels. It scans resources for medical terms that will be extracted and mapped to UMLS semantic types and groups. The UMLS group labels are then translated to consumer-friendly labels by a health information expert but from the layperson point of view. This approach could be beneficial for accessing selected health websites. However, it will be very costly to extract and map terms from the entire Web. In addition, Woods [294] notes problems and inconsistencies in the UMLS Metathesaurus such as the lack of terms, lack of synonyms, and multiple terms for the same concept which will affect this approach.

- **The Health Information Query Assistant (HIQuA) [60, 298]:** helps users "form better, longer, and more precise medical queries for submission to search engines" [60]. Upon entering a text, HIQuA offers users suggestions for completing their query and sending it to any search engine which appears on a list. However, HIQuA expects a patient to know how to express suitable lay or medical term(s) which is not the case with all patients.

- **Reformulating Health Consumer's Free-text Queries to MeSH [149]:** suggests reformulating consumer health queries (lay terminology) to standard medical terminology, such as MeSH, in order to increase the overall effectiveness of the search and improve the retrieval of relevant health information. Both original lay and reformulated MeSH queries are then executed on Google. However, the mean R-Precision[10] of the original lay queries was significantly higher than that of the reformulated query. As pointed out by Woods [294], these findings coincide with Westberg's findings about significant failure in retrieval when using terminologies from clinical codes (e.g. UMLS, MeSH, Read Codes and ICD), to capture substantial clinical content. This shows these codes are not used in the

[10] R-Precision: measures retrieved text relevance. "R-precision is the precision at R where R is the number of relevant documents in the collection for the query. An R-precision of 1.0 is equivalent to perfect relevance ranking and perfect recall" [132].

medical literature available on the Internet. However, integrating the medical terminology with lay terminology and additional popular medical terminology as suggested by our study could offer a more effective solution.

There are reports in the literature of terminology-based information retrieval systems aimed at clinicians (e.g. MELISA [147] and CMIT [258]). These clinical information systems incorporate generic medical encoding schemes and shared domain ontologies that are difficult for a patient to comprehend. In addition, browsing a generic thesaurus or ontology can be a difficult and lengthy process for novice or inexperienced users [294]. Slaughter et al [269] note a problem in utilising clinical-based medical terminology systems (e.g. UMLS) by patients and suggests expanding such systems with terms used by patients to describe their health condition [239]. In addition, Lowe sees searching within a terminology system such as MeSH as a challenge and suggests developing tools to assist potential users find appropriate terms [294]. Also, terminology-based techniques are introduced into the functionalities of Web search engines (e.g. MetaSearch and Semantic Search Engines (Section 2.4.4.1)) to improve health and medical information expressiveness and facilitate mappings between medical and laypeople terminologies. However, semantic search engines, such as HealthCyberMap [223] or LEAF [238], employ medical vocabulary schemes or ontologies (e.g. SNOMED). Also, a lay information search requirement is not appropriately supported, i.e., they lack lay-aware ontologies or medical classification systems. For instance, LEAF [238] maps both medical and lay terminology to the same information resources. Thus, a patient-oriented semantic search engine must distinguish between medical and lay patient information requirements by incorporating lay-aware medical classification systems or ontologies.

2.5 Personal Health Records (PHRs)

The patient is the core of a healthcare system [192]. However, the patient's role as a potential partner in healthcare has long been underutilised and underestimated [257]. A patient's welfare is very much dependent on timely access to essential health information at times of need. This section explores the essence of ongoing work in the development of PHR technology as the basis of the movement in Healthcare Informatics solution [109, 257] towards patient empowerment.

Traditionally, a patient is not permitted direct access to his/her electronic personal medical records. In the UK, for instance, a patient (or a patient's representative) can request a copy of a patient's medical information by an application to health authorities [44]. Subsequently, the 1998 Data Protection Act [26] granted a patient the right to apply to view or receive copies of paper and electronic personal medical records [44].

Currently, the health sector is undergoing dramatic changes in its health information infrastructure that promise to revolutionise the medical profession and the relationship between patients and their healthcare providers [257]. The recently adopted NHS health information strategies [69, 103, 114, 134] are centred on two main themes [192]:

a) **Rich information sharing:** by integrating information throughout the healthcare sector levels. Such integration would eliminate duplicate and inconsistent data and ensure timely and efficient access to consistent information at legitimate points of need. Information will eventually be accessible to individuals who need it through the integrated EPR [47, 134, 192, 254].

b) **Patient empowerment:** by offering a patient the means of timely access to essential personal health information and the exchange of information with health officials involved in their care. The aim is to encourage a patient to be responsible for and an active partner in the management of their healthcare [257]. A patient will be able to access and annotate personal health information through a summary health record that stores selected health information from the integrated EPR. This summary record has several names in the literature but it is commonly known as PHR [192, 194, 257, 270].

This radical shift in the emerging NHS information infrastructure is driven by:

a. A recognition of the underlying problems in traditional health information infrastructure [69, 103, 105, 114, 133, 134], such as the patient's lack of information and difficulty in communicating with healthcare providers [195], poor and inaccessible clinical data [193], the fragmented and unwieldy nature of paper medical records [257] and the under-utilisation of patients as potential partners in their healthcare [195, 257].

b. The patients' demand for better access to personal health information [257] and for an active role in the decision making of their healthcare [71, 195].

c. Moves in the health sector [193] to a patient-centred [118, 160] (or focussed [114]) approach, that focuses healthcare services around individuals receiving care [160] and adopts a patient's empowerment and involvement as a strategy in healthcare [253].

d. Dramatic advances in information and communication technologies [193] (e.g. Internet, Web, Email, Digital TV, Mobile phones) that are revolutionising today's information exchange and communication methods.

e. The experiences gained from a number of similar individual, national and international health information projects [193].

2.5.1 What is a Personal Health Record (PHR)?

The US Department of Health and Human Services [48] defines a personal health record (PHR) as "An electronic application through which individuals can maintain and manage their health information (and that of others for whom they are authorized) in a private, secure, and confidential environment". On the other hand, The Markle Foundation Connecting for Health (CfH) [87] program defines PHR as an "Internet-based set of tools that allows people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it" [257].

As a consumer-oriented health record, PHR offers a comprehensive and convenient means of keeping accessible personal health information. PHR is different from clinical medical records in that it can capture information from both clinical records and patients [118]. Generally-speaking, there are three types of patient medical records:

a. **Organisation-specific Electronic Medical Record (EMR)** [118, 134, 257]**:** This is a patient's medical record within a specific health organisation (e.g. GP, hospital) which is only accessed by local clinicians. Although comprehensive, an EMR poses interoperability and sharing challenges [270] when used in a wider domain.

b. **A Single Integrated Clinical EPR [134, 160]:** This is a single, common, multi-provider, integrated electronic record that is shared across participating health organisations and accessed only by authorised clinicians (e.g. ISCO/CaNISC

System (Section 2.6)). The aim is to improve quality, safety and efficiency of healthcare services [236] and so achieve better health for patients. The single record benefits the entire health stakeholders (e.g. clinicians, managers and patients) [134].

c. **A Patient-Held Personal Health Record (PHR) [134, 236, 285]:** This is a set of tools that grant a patient direct and electronic access to essential information about their healthcare such as health problems, allergies, appointments and medications. The aim is to "empower individuals by giving them the opportunity to take responsibility for their own health and to access the information they want" [71].

The concept of a personal health record system is not new [285]. Some countries have an infant's health card or booklet to record early health data. In addition, some individuals maintain copies of their medical health information in paper folders that they keep at home and update regularly. Electronic personal health records first emerged in 1995 using commercial software [117, 285] (e.g. PCASSO (Patient-Centred Access to Secure Systems Online) [116]). Electronic personal health record systems are either computer-based (e.g. [14, 56]) or Web-based (e.g. [57, 119]) [2].

Three types of personal health records have been identified in recent years [117]:

- A provider-owned and provider-maintained summary of clinically relevant health information made available to patients (e.g. MyChart [216]).

- A patient-owned software program that lets individuals enter, organize, retrieve and update their own health information regularly [117, 199, 270]. It captures the patient's concerns, problems, symptoms, emergency contact information, etc [117]. However, the majority of patients do not update their personal health records regularly [285].

- A portable and interoperable digital file that stores selected clinically relevant health data. Portable PHRs are stored on devices that can be easily plugged into a computer such as smart cards, personal digital assistants, cellular phones and USB-compatible memory devices [117].

[117] lists four types of PHR platforms:

- EPR-linked PHRs (e.g. MyChart).

- Password-protected Web-based applications.
- USB-based tools (e.g. E-HealthKEY, CapMed Personal HealthKey).
- CD-ROM (e.g. CapMed Personal Health Record).

Early PHR prototypes had varying capabilities. Core sets of PHR attributes and functions are described by the American Health Information Management Association (AHIMA) [2] and the Institute of Medicine of the National Academies [82]. PHR Standards are being formulated to ensure the interoperability, safety, security and quality of exchanged healthcare information. Competing standards in the area of EPR and PHR development [61, 288] include (but are not limited to) the American Society for Testing and Materials (ASTM) [3] and the Health Level 7 (HL7) [49].

2.5.2 PHR Benefits

The PHR technology offers new opportunities for involving patients in healthcare. They offer numerous benefits for patients, clinicians and the whole NHS [47]. PHR benefits to patients may include (but are not limited to):

- Providing patients with a view of their medical record (s) [265].
- Improving patient confidentiality [47].
- Increasing safety (e.g. drug transcribing error reduction and minimising adverse drug reactions) [47].
- Capturing information from patients about their needs and preferences (e.g. wheelchair access or organ donation).
- Improving communication between a patient and clinicians in a uni- or bi-directional way [117, 216].
- Widening the range of patient information through linkage to quality information such as specialist networks and NHS Online services [114].
- Empowering patients to be involved directly in healthcare [114, 253].
- Automating healthcare services [47] (e.g. online prescription renewal, appointments, medication reminders).

- Allowing for customisation of healthcare services for individual patients (e.g. choosing preferred referral hospital, specifying nearest pharmacy for receiving prescriptions).
- Supporting a consumer-focused, patient-focused health delivery model [117].
- Promoting preventive self-care [117].
- Supporting self-care of chronic diseases [117].
- Helping improve health data validity and quality control [117].
- Improving patient satisfaction and health [117].
- Supporting patient safety initiatives [117].
- Supporting patient and health services mobility and shared care [117].
- Providing ready access to patient data in an emergency [117].
- Supporting shared care within a fragmented health service [117].
- Providing content to help populate a life-long EPR [117].

2.5.3 PHR Limitations

PHR is a relatively new, though growing technology [194]. Pilot studies [216] raised some concerns about incorrect, incomplete or missing medical information. In a survey by Hassol et al [216], patients reported missing prescriptions and some outdated prescriptions appearing as active in their electronic PHRs. Furthermore, some patients still fear privacy risks in enabling online access to medical records [257]. Such concerns will remain a challenge and be issues to be addressed by health information technology solutions.

2.5.4 Patient's View of PHR

Pilot studies [216] and surveys [257] show that the majority of patients are positive about the use of personal health records. 59% of online health users showed interest in a universal medical record [265], while 70% believed personal health records would improve the quality of healthcare [257]. A survey by the Foundation for Accountability (FACCT) indicated that "70 percent of on-line Americans are interested in the benefits of using one or more aspects of an electronic personal medical record" [20]. In addition, 80% of patients said they could understand their medical information and test results [216], although this depends on a patient's education level.

2.5.5 Summary of PHR

PHR technology is still at an early stage of development, and so is limited but growing [194]. As a Web portal, PHR offers great opportunities for improving health services delivery for a patient through linkage to Internet technologies (e.g. Email, Digital TV, Search Engines). Early prototypes incorporated access to personal health information and personalised health services. A sample of PHR projects is presented in Appendix A. Figure 2.4 analyzes and compares the underlying functionalities of PHR prototypes, highlighted in Appendix A.

The use of PHR for patient education is part of the NHS Scotland "Patient Focus and Public Involvement" Plan [114] and is incorporated in many PHR prototypes (e.g. miHealth [225], iHealthRecord [65] and MyChart at the UT SouthWestern Medical Center [98]). The miHealth's "miInformation" educational service offers categories of information on "breast cancer" that a patient can browse. A review of the literature shows no reports about linking PHR systems with Internet search engines. Our proposal to link electronic medical records to Web search engines is reported in Al-Busaidi et al [155].

Linking patient medical records to the Google search engine has recently been proposed in a new project by Burgess [182], in order to focus online search results on a patient's needs. The project is at an early stage. The first stage attempts to focus Google search results for a patient based on a patient's specified health problem and selected generic information types. This project will eventually be linked to the single patient record in Wales [182]. Burgess proposes a Patient Health Gateway (PHG) that can be linked to the future patient record in Wales. PHG does not represent a PHR system per se. Figure 2.5 analyses and compares the features of Burgess's approach to the approach undertaken by this study.

Potential PHR Functionality	NHS England Summary Care Record [104, 130]	NHS Wales Individual Health Record [71]	NHS Scotland National Integrated Care Record [99]	miHealth [225]	US iHealthRecord [65, 66, 257]	Geisinger System MyChart [38, 216]
Access to essential personal health information (e.g. health problems, medications, allergies)	√	√	√	√	√	√
Access to appointments	√	√	–	√	√	√
Email	√	√	–		√	√
Prescription renewal	√	√	–	–	–	√
Request referral	√	–	–	–	–	√
Nominating pharmacies	√	–	–	–	–	–
Nominating referral hospital and clinics	√	–	–	–	–	–
Diary/Mood journal	√	–	√	√	–	–
Important contacts	√	–	–	√	–	–
Medication reminder via Email	√	–	–		√	–
Information from health events (Discharge, Operation letters)	–	√	–	–	–	–
Care relationships	–	√	–	–	–	–
Wallet emergency card	–	–	–	–	√	–
Linkage to online NHS Services	–	–	√	√	–	–
Education programs	–	–	√	√	√	
Web search	–	–	–	–	–	–
Personalised Web search	–	–	–	–	–	–
Digital TV	–	–	–	–	–	–
Information prescription/ Recommending health websites	–	–	–	–	–	–

Figure 2.4: A Comparison of Various PHR Prototypes' Capabilities

Feature	Patient Health Base (PHB)	Patient Health Gateway (PHG) [182]
Personal Health Record (PHR) Capabilities:		
	a. Access to Personal Medical Data: diagnoses, treatments, cancer management. b. Personalised online search. c. Rich diagnosis thesaurus. d. Hospital recommended websites. e. Patient Favorite websites.	To be linked to future EPR/PHR in Wales.
Personalised Online Search:		
Search Information Topic	Selected from a patient's diagnoses, treatment or cancer management plan	Assumes a search on health condition that is entered by patient.
Search Refinements (Health Information Types)	Selected from an extensive search refinement list investigated from literature and interviews with information staff.	Given as multiple choices of health topics to define information requirement.
Underlying Search Engines	An array of search tools including Google search engine, medical search engines, HON accredited search engine, hospital-trusted websites, patient Favorite websites and specific website search.	Only Google.
Online Information Quality	Wide-ranging search tools such as accredited search engine (e.g. HON), national health gateways, hospital-trusted websites, patient-Favorite Websites and Google.	Relies on Google's PageRank technique.
Search Information Terminology	Builds and utilises a Patient-oriented Diagnosis Ontology (PDO) that encodes diagnosis information types sought by patients and integrating diagnosis information from the medical and lay perspectives. This is utilised to explain, focus and enrich medical knowledge relating to a patient's particular diagnosis. There are options for normal or semantic search.	Not addressed.
Search Results Relevance and Focusing	Search result focusing is influenced by focused patient-personalised search topics drawn from patient EPR, diagnosis related terms, and search refinements. However, we are not addressing prioritising search results. This is because our approach is linking to an extensive list of search engines.	Free-text search on a health condition specified by a patient. Additional information requirements are used to prioritise Google search results for a patient.

Figure 2.5: Comparing Burgess's Patient Health Gateway (PHG) to our Patient Health Base (PHB)

2.6 The Information System for Clinical Organisation (ISCO)

This section describes the Information System for Clinical Organisation (ISCO) [162], the patient information system utilised in this study. ISCO was first developed in 1991 by the Velindre NHS Trust to meet the information requirements of the Trust's oncology service and to collect clinical management information and perform analysis on it. Since 1993, ISCO has been used as a case record system by Velindre Hospital to record and update clinical information throughout the patient's treatment journey from the time of the first diagnosis at the hospital.

In 1998, there was a major development of ISCO in the Cancer Network Information System Cymru (CaNISC) project to "pilot prospective collection of the all Wales cancer datasets by MDTs across Wales" [162]. The Velindre NHS Trust still uses the name ISCO for historical reasons for this enhanced system. ISCO is commonly known outside the Velindre NHS Trust as CaNISC [122]. Currently there is no access for patients to the ISCO/CaNISC System. However, ISCO/CaNISC acts as a model for the anticipated Individual Health Record (IHR) [243] that will ultimately be accessed by patients in Wales using the online gateway "My Health Online" [95]. Thus, the current trend towards patient empowerment and involvement in healthcare will include the prospect for patient access to the ISCO/CaNISC system. This study adopts this new perspective and approach to patient healthcare.

2.7 Investigating Patient Information Needs

> "... Individuals' perceptions of their needs may differ from those of the professional. Good communication between professionals and patients is especially important." - Calman-Hine Report6, para 3.1(iv) [11].

Information is paramount to patients especially when experiencing acute illness or stress. However, patients, often, complain that their information needs are not sufficiently met [293]. Mostly, patients seek information on their health problems and medications. In this study, we investigate how to utilise a patient's own medical data in EPR to build an extensible and enriched patient health information model to utilise in Internet medical searches. Such a personal information model can help personalise and focus the Internet search process for a patient by linking this model to Internet search engines and key health websites.

A patient's medical and health information needs are investigated from several perspectives, namely:

a) Reviewing of literature and publications on patient information needs and Internet access,

b) Attending conferences on patient information and online search,

c) Analysing ISCO EPRs and the underlying encoding schemes, and

d) Interviewing healthcare professionals and patient information specialists.

Our exploration into patient information needs covers four themes:

- **Generic Information Needs** (Section 2.7.1). What types of information do patients usually seek? What are the problems or regulations related to patient information?

- **Internet Access and Medical Online Search**: What types of information are often sought by a patient online? (Section 2.7.2).

- **Health Information Terminologies**: What types of terminologies does a patient need to use or understand when dealing with Internet health information? (Section 2.7.3).

- **Essential Personal Medical Information**: What types of personal medical data do patients usually seek and wish to further explore for education and decision-making purposes? (Section 2.7.4).

2.7.1 Generic Patient Information Needs

Generic patient information needs are investigated through interviews with health information officials, review of literature and publications on patient information and attending conferences on patient information and Internet access.

We interviewed healthcare professionals in the Clinical Information Unit, Patient Information Centre, Radiotherapy Unit and Chemotherapy Unit at the Velindre NHS Trust. The aim of these interviews was to explore health topics often requested by patients and investigate problems reported by patients and regulations concerning patient information.

We were not able to interview patients or patients' representatives at the prototype stages for anonymity reasons. However, we reviewed patients' perspective from a literature survey of patient preferences regarding information needs. This perspective is incorporated in the design of the Patient Diagnosis Ontology (PDO) (see Section 6.4.1) and the patient Personal Internet Search (PerIS) system (see Section 6.3).

Aston [164], a patient information manager, indicated that patients usually require general information as well as information on their particular treatments. This information is mostly concerned with chemotherapy (50%) and radiotherapy (50%) treatments. In addition, patients seek information to double check their diagnosis and suggested treatment, and that they comply with the NICE [144] guidelines. Moreover, patients seek information on their tumour marker, prognostic indicator, complementary therapy and vitamins [164].

Allam [161], emphasized patients' concern about side effects of treatment on cure rate and life span. Roberts [263] pointed out that patients ask about complementary therapy, support information, health management after treatment and outside the clinic. She further explained that patients at diagnosis time ask about basic information such as treatment options, clinical trials and drugs, while at treatment time they ask about treatment procedures and type of visit (e.g. in-patient, out-patient).

Additionally, we investigated a patient's recorded treatment within radiotherapy and chemotherapy departments. Details about a patient's proposed treatment are not known until the patient visits the designated department, where he/she receives general information about his/her problem and individualized information about the proposed treatment [184, 263]. For instance, a chemotherapy treatment sheet describes a patient's treatment plan, combination of drugs, and schedule of treatments [263]. The patient treatment sheet is not fully recorded in the ISCO patient database. ISCO stores a summary of a patient's various treatments that is intended for clinical use. Nonetheless, ISCO medical data such as radiotherapy machine and chemotherapy drug used were identified as useful information [184, 263] for further exploration by a patient. According to Butters [184], side effects vary from one patient to another. In a radiotherapy treatment, this could be determined by the site of the treatment (e.g. brain, abdomen), drug dose, machine, and reaction of the patient to the drug.

A survey by Bilodeau and Degner [176] of women with breast cancer identified the stage of disease, treatment options, and likelihood of cure as the preferred information. Patient and family (or carer) information needs were also surveyed in [261]. Generally, a patient's information needs vary according to the current stage of the cancer journey. At diagnosis, patients seek re-assuring information about the likelihood of cure, treatment options and stage of disease. At treatment, they enquire about treatment procedures such as investigative tests and recurrence of disease. At post-diagnosis stage, patients focus on self-care and social concerns such as self-care behaviours and risks to family members. Nonetheless, information on likelihood of cure is required by patients at all stages of their cancer journey. Family members seek information on the disease, its diagnosis, prognosis, treatments, side-effects of treatments, and expected course of the recovery and prevention of recurrence.

The Cancer Information Strategy (CIS) [134] asserts that patients desire detailed information on their health problems and expect support in understanding this information and deciding on appropriate treatment:

> "Many cancer patients want to be informed about their condition, prognosis (outlook) and treatment options, and to be supported in making decisions about their own care........The majority of patients want detailed information to enable them to be actively involved in decisions about their treatment." Cancer Information Strategy, Section 2.4-2.5 [134].

Additionally, Aston [164] emphasised the following considerations associated with patient information delivery, namely:

- Patients require clear and understandable health information. Patient information is typically audited for the Plain English Campaign (PEC) [120] criteria to ensure that patients receive clear and easily readable information. Health information resources evaluated for PEC receive the Crystal Mark. Documents describing patient information need to be audited for plain English. At Velindre NHS Trust, this process is performed by a reading panel, including patients [164]. Velindre NHS Trust website [140] has been given the PEC crystal mark for English Clarity [164].

- Patient information should also take into account cultural differences and preferable language of minor ethnic communities.

Figure 2.6 enumerates a list of health topics often sought by patients as identified in this study.

Health Topics
Tumour marker, prognostic indicator, complementary therapy, treatment side effects, treatment procedures, likelihood of cure, cure rate, expected life span, complementary therapy, support, treatment options, clinical trials, drugs, procedures, diagnosis, prognosis, recovery, prevention, recurrence, diet, nutrition, vitamins, exercise, fitness, prescription, drugs, alternative treatments, alternative medicine, self-care, health management, radiotherapy machine, chemotherapy drug, site of chemotherapy treatment.

Figure 2.6: Health Topics Often Sought by Patients According to Literature and Interviews

2.7.2 Internet Access and Medical Online Search

Patient Internet access is widely acknowledged in literature and by most professionals interviewed in this study. However, Butters [184], a patient radiotherapy nurse, raised concern about customising Internet information for a patient as patient information requirements are greatly influenced by health condition, treatment and a patient's reaction to treatment. In addition, Butters [184] advocates the use of hospital information as opposed to Internet information as it is more reliable. Several information needs are noted with regard to patients' Internet access, namely:

- Patients should be allowed to view the websites they desire whether local or international [164].
- Information should not be restricted or classified for patients, though some guidance to key health websites and organizations could be useful [164].
- Some patients have indicated a preference for touch screen interfaces to access Internet-based information and services [164].
- Roberts [263] advocates the use of key Internet health charity websites such as Cancerbackup.org.uk in educating patients about health problems and treatments.
- Patients should access local and international Internet resources equally. However, identifying these resources clearly to patients will assist them in

clarifying the Web information and judging its applicability to their healthcare system [263].

The majority of patients access online health information using search engines [206]. Health information searching is regarded by some patients as a sub-optimal process [232]. Medical student Internet users have found health information search via general search engines more successful than using medicine specific search engines [232]. Most online users search for information on health problems (63%) and treatment procedures (47%) [229]. Figure 2.7 shows the Pew Internet Project statistics on health topics sought online [229].

Health Topic	**2002 (%)**	**2004 (%)**	**2006 (%)**
Specific disease or medical problem	63	66	64
Certain medical treatment or procedures	47	51	51
Diet, nutrition, vitamins, or nutritional supplements	44	51	49
Exercise or fitness	36	42	44
Prescription or over-the-counter drugs	34	40	37
A particular doctor or hospital	21	28	29
Health insurance	25	31	28
Alternative treatments or medicines	28	30	27
Depression, anxiety, stress, or mental health issues	21	23	22
Environmental health hazards	17	18	22
Immunizations or vaccinations	13	16	16
Dental health information	*	*	15
Medicare or Medicaid	9	11	13
Sexual health information	10	11	11
How to quit smoking	6	7	9
Problems with drugs or alcohol	8	8	8

*: *Question was not asked in this survey*

Figure 2.7: Internet Users Searching for Health Topics [229]

2.7.3 Health Information Terminologies

Most online health information is published by professionals who normally use medical terminology while the majority of online users, including patients, are from the lay public. A number of online user and patient needs are noted in this regard:

- Online users report problems expressing the correct medical term [232].
- Lay terminologies can lead to part of the information required being found, but can lead to misleading information [232].
- Patients need to distinguish between similar, related terms and/or specific/generic terms [232].

- Highly educated patients complain about receiving health information in a simple form and demand health information in medical or scientific form [184].
- The same health information can be expressed in multiple medical and lay terminologies.
- Mapping between different medical terminologies can prove problematic due to the lack of corresponding concepts or to the use of different representations [294].

2.7.4 Investigating a Patient's Essential Medical Information in ISCO

Following the generic exploration of patient information requirements, electronic medical records were analysed. This process was conducted alongside interviewing the health information staff and ISCO database developers. The interviews with database developers were aimed at locating and interpreting a patient's medical data in ISCO, whereas the interviews with patient information staff and nurses were aimed at investigating the usefulness of the extracted medical information.

Clinical medical records are typically designed for clinical use and often deemed incomprehensible and therefore unsuitable for patient access. However, a patient's EPR serves as a basic source of health topics pertinent to a patient's own medical condition. We have explored ISCO EPRs for clinical data on diagnoses, treatment and cancer management plan, all of which could be of interest to patients and meaningful in online searching. Such information can benefit patients in two ways:-

1. Enabling a patient to view and comprehend their medical details.
2. Helping to customise and focus educational information and online searching for a patient.

While this research is applicable to any patient community or health condition, our exploration into a patient's medical information needs is related to cancer patients registered in the ISCO system. Our study utilises an anonymised version of the ISCO patient database obtained from the Velindre NHS Trust – Clinical Information Unit (CIU). Three types of patient medical data were explored:

- **Diagnoses**: these help a patient view and comprehend their health problems.

- **Treatment Episodes**: these help a patient view and relate to previous treatments.

- **Cancer Management Plan**: these help a patient view and investigate proposed treatment.

At the time of this investigation, ISCO records four types of cancer treatment:

- **Chemotherapy**: treatment of disease by means of chemicals [111].
- **Radiotherapy**: treatment of disease by ionising radiation [111].
- **Surgery**: treatment by operation.
- **Palliative care**: treatment aimed at relieving symptoms and pain rather than affecting a cure [111].

In addition, we explored the ISCO clinical data on the cancer management plan and identified the information that was identified as useful for a patient's understanding. The selection of medical data within these categories is first determined by how much medical information is available in the ISCO patient records. Secondly, it depends on the meaningfulness and usefulness of this information for patients and its appropriateness for Web search. The selection was discussed with the ISCO database team [167, 173] and subsequently verified with patient information staff [164] and specialist nurses [184, 263]. The patient's medical data on technical procedures and/or instruments was ignored as it was regarded as less significant to the patient's education process. A summary of the extracted medical data from a patient's record is given in Figure 2.8. This information forms the basis for a patient model of information that will be utilised in personalising and enriching a patient's medical knowledge and online searching.

Category	**Details**
Diagnosis	Diagnosis name
Chemotherapy	Treatment place, treatment type and drug name
Radiotherapy	Treatment intent, treatment site and machine name
Surgery	Treatment intent and anaesthetic information
Palliative Care	Care type and care aim
Cancer Management Plan	Plan intent, modality and modality order

Figure 2.8: Summary of Extracted ISCO Patient's Medical Data

2.8 Summary

Information is central and significant to patient healthcare. However, there is a consensus that patients lack this information. This has recently been acknowledged by the UK Department of Health (DOH) with evidence suggesting that patients only retain 10% of the information imparted at a consultation. This chapter reviewed traditional and current means of informing patients and outlined the underlying limitations.

In addition, the chapter explored the potential for patients to manage their own healthcare as advocated by the patient-empowerment discourse, which sees information as a way of empowering patients and calls for more participation of patients in their own healthcare. This approach has recently been adopted by official healthcare policies through a revolutionary approach to healthcare based on rich-information sharing, patient-centeredness and patient empowerment. A key component of this new healthcare approach is the PHR that enables a patient to access personal medical information and supports better communication between a patient and healthcare providers. Velindre NHS Trust currently does not support patient access to electronic medical records. However, PHR technology is already part of the Welsh Informing Healthcare (IHC) programme to develop an Individual Health Record (IHR) for the people in Wales, which will eventually be accessed through the Web portal My Health Online. A review of sample PHR projects describes numerous patient-personalised services but does not describe any personalised Internet search tools within the PHR framework that links PHR (or EPRs) to Internet search engines.

This chapter examined the Internet potential, as the leading information technology, for patient healthcare, and noted there was a high but sub-optimal and uncustomised access to online health information sources. In addition, a patient's access to Internet healthcare information is uncustomised and hindered by information quality and vocabulary challenges.

The chapter reviewed techniques addressing these challenges and examined individual patient information requirements as the key to addressing these challenges. The chapter investigated patient information needs for online health information search. Studies have indicated that patients often search for health information on their health problems and medications. Hence, the chapter further argues that enabling a patient to access medical records, utilising the patient medical knowledge domain and linking EPRs to key health gateways and trusted health websites can help overcome these online search challenges and simplify, focus and personalise a patient's online search experience. Chapter 4 further analyses the research problem in terms of domain problems, stakeholders' needs and proposed system features.

As we propose integrating EPRs data with relevant Internet information sources, Chapter 3 reviews approaches to data integration whereas Chapter 5 presents our approach to integrating EPRs data with Internet information sources.

CHAPTER 3

Data Integration and Semantic Interoperability

3.1 Introduction

Chapter 2 investigated the research problem from the healthcare perspective highlighting a patient's lack of information and challenges in patient Internet searching. Accordingly, this study proposes customising a patient Internet search by allowing patient access to essential medical information in EPRs and linking EPR data to relevant Internet health information resources, according to patient information needs and preferences. From a computing perspective, this linkage constitutes an integration task between the patient database and relevant Web documents. The ISCO database is a relational DBMS whereas Web documents can be simple HTML files or Web interfaces to information systems from various organisations.

This chapter explores the data integration environment. Sections 3.2 and 3.3 explain the notions of data integration and semantic interoperability respectively. Section 3.4 examines challenges inherent to the integration process. Section 3.5 classifies common data integration architectures. Section 3.6 highlights issues and decisions to be considered in prospective integration systems. Section 3.7 reviews ontologies while Section 3.8 concludes the chapter.

3.2 What is Data Integration?

Information systems belonging to different organisations are naturally autonomous[11] and heterogeneous, as they are developed independently. The same

[11] Autonomous information systems assume an organisation has control over their data and operations.

or perhaps related information may exist in various sources. An attempt to combine data from such sources requires resolving the inherent heterogeneity at various levels. In addition, it could involve duplicate, overlapped or inconsistent data. Thus, combining data from disparate systems in a way that conciliates the inherent heterogeneity and presents users with a homogenous and uniform view is termed *Data Integration [213].*

Web-based data integration is closely related to traditional data integration of autonomous and heterogeneous information systems [220] because it involves data belonging to independent organisations which could be modelled and implemented differently. Therefore, Web data integration is discussed throughout this chapter using the concept of data integration in heterogeneous and autonomous environment. However, the anonymous and dynamic nature of the Internet brings additional challenges to the integration process (see Section 3.4) that might influence the integration architecture choice. Web data integration is commonly used to solve problems relating to answering queries rather than transaction between sources [220].

Closely related to Data Integration is the concept of *Interoperation* or *Interoperability.* Mostly, the term Interoperation implies the effective exchange of data and operations between different data sources. Elmagarmid et al [198] describe the term Interoperability as "the ability to request and receive services between interoperating systems and use each others' functionality". Interoperability involves only data sources that model similar or related information [198]. Minimal interoperability can be achieved when a system periodically sends data to another system [198]. However, Wiederhold [291] discusses the term Interoperation in terms of virtual integration that does not store data at the integration level. An opposite concept is the *Materialised* or *Data Warehouse (DW)* integration that stores both data and its descriptions (i.e. meta-data) at the integration system. Integration approaches are further discussed in Section 3.5.

3.3 What is Semantic Interoperability?

The problem of semantic interoperability emphasizes the difficulty in integrating resources that were developed using different vocabularies and different views (or perspectives) on the data [217]. Ozsu and Valduriez [252] define *Semantic*

Interoperability as "the process by which information from participating databases can be conceptually integrated to form a single cohesive definition of the data held in multiple databases" [278]. Thus, semantic interoperability is crucial for the effective integration and usability of distributed information systems [148].

Accordingly, integration solutions towards semantic interoperability should allow for both the semantic (i.e. meaning) and the structural (i.e. representation) integration of the data belonging to heterogeneous data sources. Thus, the integration process can be viewed as a requirement for semantic interoperability [278] or a result of interoperation among data sources [291].

3.4 Data Integration Challenges

Conventionally, integrating autonomous and heterogeneous data sources poses challenges in the following dimensions:

a. **Autonomy:** this is concerned with the distribution of control [174] over the data and operations of the individual data sources. Local users versus global users are competing for resources at individual sites [208, 214]. [174, 266] describe four types of autonomy: design autonomy, participation autonomy, communication autonomy, and execution autonomy. These are mostly concerned with federated systems [174].

b. **Heterogeneity**: implies differences or dissimilarity among peer data sources at various levels of abstractions (e.g. system, data model, data semantics). It can be broadly classified into system heterogeneity and semantic heterogeneity [209]:

   - **System heterogeneity**: caused by differences in hardware (e.g. platform, OS, communication protocols) and software (e.g. data model, DBMS, query language) used by different data sources. The hardware heterogeneity can be resolved using gateways and middleware technologies [190] whereas the software heterogeneity is overcome using translators or wrappers (e.g. JDBC/ODBC) [224].

   - **Semantic heterogeneity**: arises from different modelling of the same real world objects and results in variations of concepts, terminologies and structure among various data sources. Examples of semantic heterogeneity include the use of different terms to refer to the same concept (i.e. synonym problem), and the use of the same term to refer to different concepts (i.e. homonym problem) [256]. Furthermore, data represented in different data sources could be interrelated or overlapped. For instance, Website A may use the term "kidney cancer" to represent the same concept that Website B refers to as "renal cancer", yielding a synonym problem. Similarly, Website C may offer more generic information on "renal cancer" that is described using the term "urological cancer". This yields a generalisation/specialisation problem or relationship. Semantic heterogeneity is recently addressed using shared ontologies[12] (see Section 3.7) that specify the terminology used by the problem domain [256]. However, shared ontologies could be complex and might not reflect the requirements or the terminology of the end user. Alternatively, ontologies could be used to define the conceptual view or terminology used by the end user or application domain [190].

c. **Duplicate and inconsistent information**: combining data from more than one source may result in retrieving duplicate or inconsistent data. The integrator has to resolve this issue when retrieving partial results from individual data sources.

d. **Volatility of the data source**: data sources may come and go (e.g. due to migration). Thus, loss of existing data source should not affect the representation of the global view or knowledge [215]. The implementation of the integrated system needs to isolate the implementation and terminologies of the individual data sources from that of the user interface or knowledge domain.

e. **Evolvability of the individual data sources or the global interface system**: changes at individual data sources should not affect the global interface or conception (i.e. how users formulate their requests). Analogously, changes at the global domain level to accommodate new user requirements should not impose changes to the implementation of the individual data sources.

[12] Theoretically, the term ontology is defined as "explicit specification of shared conceptualisation in a domain of interest" [204]

f. **Scalability of the integration system**: the integration system should have the flexibility to accommodate additional data sources without affecting the integration framework or architecture.

In addition, Web data integration introduces new challenges to the traditional data integration process:

a. Information about the meta-data of Web data sources is not accessible. This makes it difficult to make exact comparisons between elements belonging to different data sources [220].

b. Difficulty in identifying or eliciting the intended semantics of the data due to the inaccessibility of both data source metadata and DBAs.

c. Large number of data sources that grow exponentially with the Internet [220].

d. Frequent changes in a data source's content and layout (i.e. presentation) [220].

e. Web queries might involve data sources in multiple domains [220].

f. Users and possibly applications have no control over data sources.

g. Web data carries multiple providers' perceptions and gets interpreted according to multiple users' perceptions.

h. The average Web user lacks database and integration skills which makes it difficult for him/her to perform the technical integration task without help.

3.5 Data Integration Architectures

The integration architecture describes the proposed infrastructure of the integration system, strategies within the system and the communication mechanisms with the participating information sources and the end user. The selection of the architecture is central to the integration process [220]. Literature [159, 174, 198, 224, 266, 278] describes several classifications of interoperability and data integration architectures. In this thesis, we classify integration architectures according to three dimensions:

- ***Level of Abstraction*** **[198]***:* denotes the level at which the integration (or interoperation) occurs [198] (see Section 3.5.1).

- ***Integration Mechanism/Method/Strategy:*** concerns the rules of a data source's participation and the integrated view's generation mechanism (see Section 3.5.2).

- ***Data Management:*** signifies the mechanism by which the integration system services end user queries (see Section 3.5.3).

3.5.1 Abstraction-Level-Based Integration Architectures

In the context of heterogeneous databases, four levels of abstractions are noted; User View Level [198], Conceptual Schema Level [198], Data Level [198] and Behaviour (or Method) Level [159]. Additional integration levels pertinent to Business and IT application research areas are emphasized in [8, 168, 279]. This section highlights five integration levels of abstraction that are deemed useful for the research investigated in this thesis, namely:

a. ***Schema- (or Structure) Level Integration [159, 174, 198, 224, 266, 278]:*** This is the classical level of integration in multidatabase systems. It involves the schema objects of participating data sources and results in the generation of an integrated schema from local data sources. Schema-level integration requires access to data source descriptions (or meta-data) and is performed by specialists, often called integrators.

   Nonetheless, schema-level integration is not feasible for integration tasks aimed at novice Web users such as patients or involving highly autonomous Web data sources such as health gateways and search engines for two reasons:

   - Web data source meta-data is inaccessible to Web users.
   - Web users are not skilled in data source schema manipulation. Rather, they are mostly concerned with a data source's content and possibly its presentation style.

   However, schema-level integration methods and techniques (e.g. loosely-coupled integration and text-matching) can be applicable at other integration levels.

b. ***Data- (or Instance-) Level Integration [8, 168, 198, 220]:*** Several definitions are reported in the literature:

- ***Data-Level Integration [8]:*** is "a data to data integration".

- ***Data Level Integration* [198]*:*** "relies on actual data values to achieve integration".

- ***Data-Centric Integration [186]***: is "the automation and integration of data flows that are exchanged between ISs".

Furthermore, in the business domain, data-level integration is described as ETL (Extract, transform, Load) [8] since data extracted from one data source, might get transformed before loading it into another data source. Data-level integration can be used to create mappings between data exchanged among disparate systems in order to facilitate the exchange and manipulation of the exchanged data [279]. Such an integration type is often used in accounting and EPR systems [8]. Two main issues need to be addressed by Data-level integration systems [198], namely:

- Identifying data about the same real-world entity in participating data sources.

- Resolving differences in data values that represent the same real-world entity.

Recent solutions to data-level integration often utilise IT technologies for the automatic extraction, formatting and mapping of data from various applications [279]. [25] describers EAI[13] tools for accomplishing data-level integration such as JDBC/ODBC-based queries and add/delete/update triggers. Themistocleous and Corbitt [279] discuss more advanced EAI technologies to support data-centric integration such as message brokers and adapters. Accordingly, data-level integration offers a more appropriate type of integration in a Web environment, where a data source's schema is inaccessible and end users are usually novice Web users.

c. ***Conceptual-Level Integration:*** supports a higher level of integration. It uses concepts to model the data exchanged among disparate data sources. This level of integration is key to achieving semantic interoperability. It requires

[13] EAI (Enterprise Application Integration): "The use of middleware to integrate the application programs, databases, and legacy systems involved in an organisation's critical businesses processes" [32].

knowledge about data semantics in the local data sources. However, data semantics may or may not be explicitly presented. Thus, data semantics can be inaccessible, implicit or explicit.

- ***Inaccessible Data Semantics:*** published Web information does not convey the intended data semantics. In such a case, Web users need to use their own knowledge and skills in interpreting the data based on data labels and context. Thus, lexical matching techniques can be used to identify the same labels. However, additional work is needed to identify similar or related data labels.

- ***Implicit Data Semantics:*** occurs when data sources lack mechanisms for specifying, storing, and ultimately accessing data semantics, as in relational databases. In such a case, the data semantics can only be manually recovered from the data source administrator and available documentation, or inferred from the names and values of the entities and attributes [283] possibly using text-matching techniques as in [196].

- ***Explicit Data Semantics:*** recent research into the semantic Web [141] advocates the representation of data semantics using some formalism [273] so it can be recovered by prospective users and applications. Data semantics can be explicitly specified using a conceptual structure such as meta-models [224], thesauri [146] or ontologies [204] (see Section 3.7). The conceptual structure models knowledge about a particular domain. Integration system users formulate their queries using terms and concepts specified at the conceptual layer. In this manner, users do not have to be concerned with the terms or concepts used in local data sources.

In the traditional Web, data semantics (or conceptualisation) is inherently variant and inaccessible. This puts the onus on the Web users to interpret Web data according to their perspectives and capabilities. However, recent research efforts (e.g. SHOE [218]), within the Semantic Web framework, advocate the annotation of webpages with a shared ontology, so that webpage content is interpreted according to the semantics specified by the associated ontology. Nonetheless, this approach is not widely adopted by Web authors in the current Web environment. In addition, Web users can not adhere to shared ontologies as they may not reflect their information requirements or own conceptualisation.

Hence, Web data integration solutions need to cope with variant Web data sources' semantics, and address it in terms of the Web user data semantic requirements. Data semantics is classically addressed in terms of strategies which aid the identification of semantically-related objects (see Section 3.6.4).

d. ***User-Level Integration [166]:*** allows for accommodating user requirements and needs in the integration process and its deliverables. An example of such an integration approach is an application that aggregates data from disparate sources and delivers data and results to an integrating user interface or Web portal in a manner that is personalised to the user's needs [166]. This could be achieved with respect to the content, presentation or semantics (e.g. terminology).

3.5.2 Integration-Method-Based Integration Architectures

This category of integration architectures is pertinent to the integration rules and operations. Particularly, it concerns the transparency of the integration process to either participating data sources or intended integration system's users. The selected architecture needs to address the following issues:

1. Are data sources or integration users aware of the integration process?
2. How skilled are the integration users?
3. Who performs the integration?

Common approaches at this level are the *Federated*, *Unfederated* and *Mediator* architectures. The Unfederated architecture differs from the federated architecture in that it does not support local users at the local data sources [174]. This makes this type of integration an inappropriate choice for the integration of legacy databases such as the ISCO patient database. Such systems are autonomous and need to continue with their local users and operations after and during the integration process. The federated architecture is the most reported architecture in literature. It is usually associated with Schema-level integration. The study reported in this thesis does not address schema-level integration, due to the inaccessibility of a Web data source's schema. However, schema-level integration strategies can be applied at other integration levels (e.g. data-level integration). An extensive review of schema-level integration architectures is given in [159, 196, 219, 224, 266, 278]. Sections 3.5.2.1 and 3.5.2.2 summarize the federated and the mediator integration architectures.

3.5.2.1 The Federated Architecture

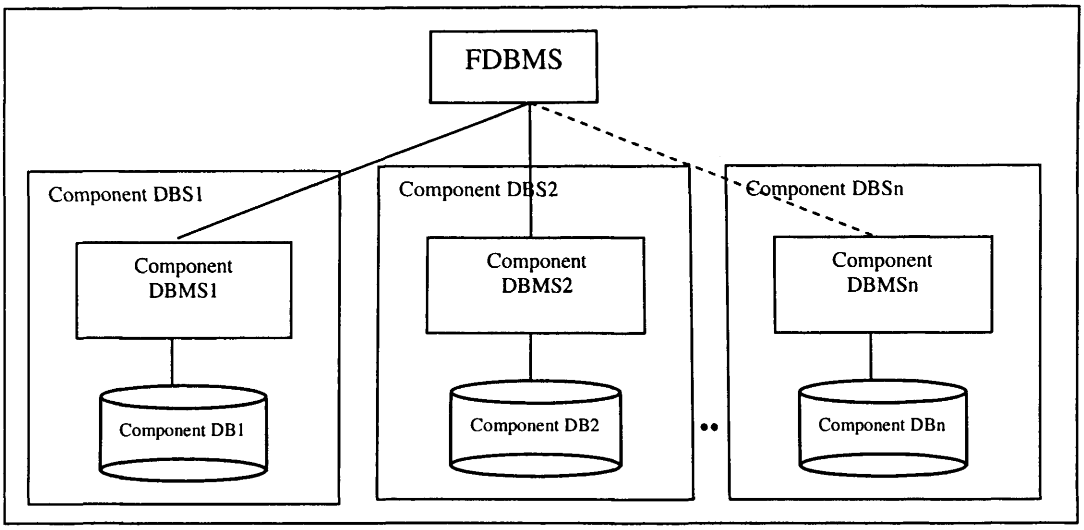


Figure 3.1: An FDBS and Its Components (Based on [266])

According to Sheth and Larson [266], the terms *federated database system* and *federated* architecture (Figure 3.1) were first introduced by Heimbigner and McLeod [219] to mean a "collection of components to unite loosely-coupled federation in order to share and exchange information". A federated database system (FDBS) refers to integrating autonomous and distributed component databases. However, a federated architecture may incorporate other system types (e.g. Geographical Information Systems (GIS), and/or Expert Systems (ETS)) [289]. The federated database system presents an alternate approach to distributed data management that does not hinder the functionalities and the applications of the constituent systems. Pre-existing legacy systems can safely adopt a federated architecture by adding a software layer above their existing DBMSs [266]. The federated approach is traditionally associated with multidatabase systems (MDBS) [252]. The mechanism provided for the federated architecture must balance two conflicting requirements: maintain as much autonomy as possible while achieving a reasonable degree of information sharing [219].

Classically, a federated system can be classified as a loosely-coupled or a tightly-coupled federated system. A tightly-coupled federated system (Figure 3.2) allows the creation of one or more global federated schemas from participating component databases. This is a very complex operation. The creation and maintenance of the federated schema is fully controlled by the Federation Administrator (FA) and achieved by negotiations between FA and component DBAs [198]. A tightly-coupled FDBS is termed a *single* federation, if it allows the creation and management of only one global federated schema. Single federation helps in maintaining uniformity in the semantic interpretation of the integrated data. On the other hand, a tightly-coupled FDBS is said to have *multiple* federations if it allows the creation and management of multiple federated schemas. Having multiple federated schemas allows multiple integrations of component DBSs for different groups of users. A tightly-coupled federation can be achieved based on the *reference architecture* [266] or by means of the *participation* schemas described in [174]. The latter is similar to export schemas in [219].

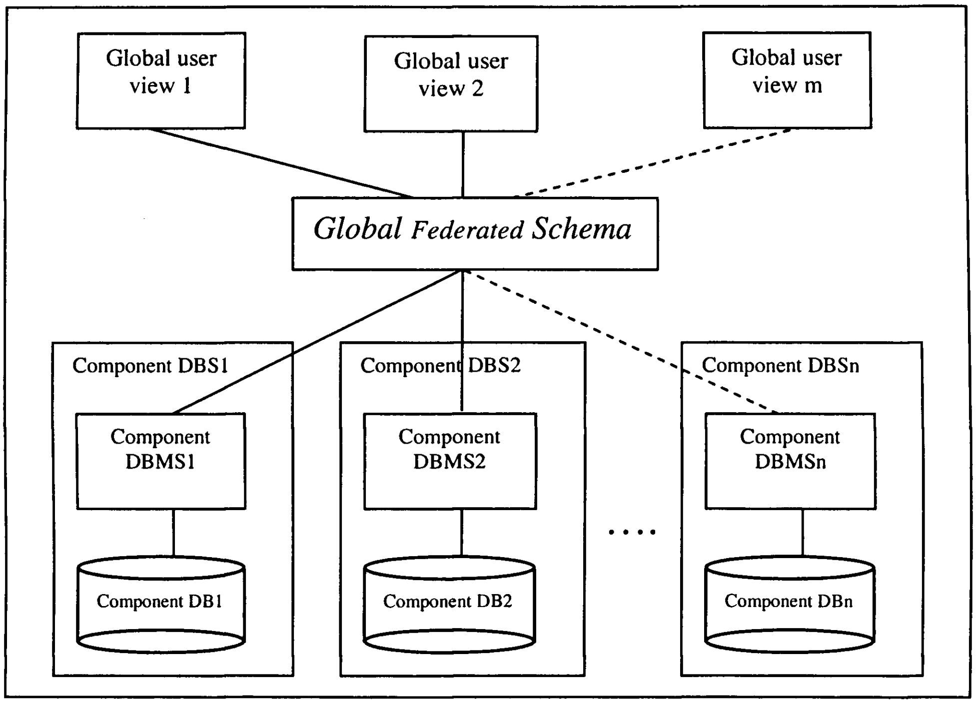


Figure 3.2: A Tightly-Coupled FDBS and Its Components (Based on [174, 266])

Tightly-coupled federation systems are static and predefined. This creates a problem for evolution of the data source content or user requirements. Hence, they are suitable for less evolving and small-scale integration systems. Furthermore, scalability is an issue as adding new data sources to the federation requires changes to the federated schema.

In contrast, a loosely-coupled federated system (also known as interoperable database system [174] (Figure 3.3) is distinguished by the lack of a global federated schema. Global users are responsible for the creation and maintenance of their federated schemas [198]. This allows users to define the federated schema that best meets their needs. Hence, federation users should be knowledgeable in exploring the structure and content of relevant data sources. Loosely-coupled federated integration can be achieved by means of export/import schemas [219] or a multidatabase language [196].

Loosely-coupled federations are dynamic as they can be easily created or dropped. They assume highly autonomous read-only databases and do not support view updates [198]. This makes loosely-coupled federated architecture a potential architecture for read-only Web-based integration endeavourers.

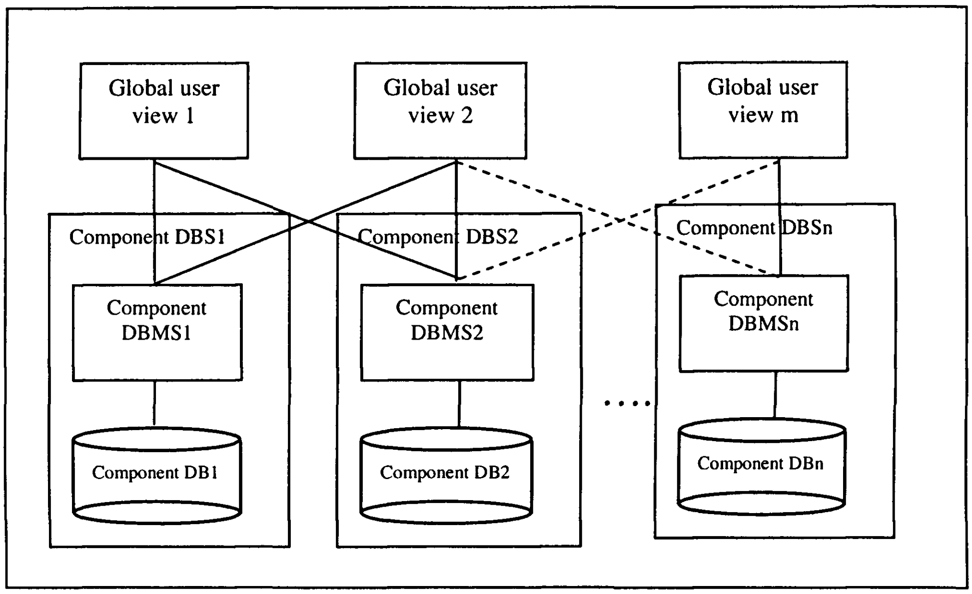


Figure 3.3: A Loosely-Coupled FDBS and Its Components (Based on [174, 266])

3.5.2.2 The Mediator Architecture

The mediator approach (Figure 3.4) emerged to address Internet data integration challenges [252]. Web-based data integration differs from traditional multidatabase integration in the following aspects:

- The large number of data sources creates a problem for view generation and conflict resolution [252].
- Web data source content is very dynamic which impacts on the integrated view [252].
- Different data structure capabilities ranging from structured relational data to text files [252].

- Data can be semi-structured or unstructured which offers no information to the integration process [252].
- The novice Web users, who can only read data, but have difficulty analyzing and/or constructing schemas or proper conceptualization of that data.

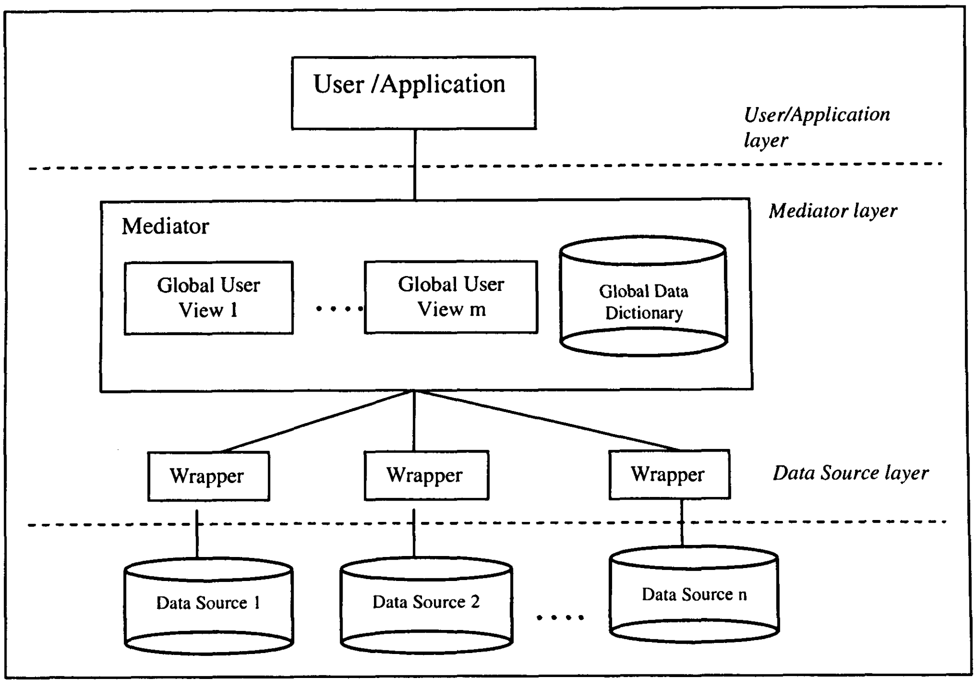


Figure 3.4: Mediator-Wrapper Architecture (Based on [252])

Hence, the federated architecture is extended with two components wrappers and mediators when used in Web data source integration [252]. Wrappers address the variations in data source capabilities and present the integration system with a uniform interface[14] to the data source. On the other hand, mediators attempt to separate the implementation and technical details that are pertinent to the federation from that of the participating data sources. It arranges these details into three separate layers:

- The data source layer: stores information and elements of the participating data sources,
- The mediation layer: stores information and elements of the federation, and
- The user/application layer is an external layer.

[14] Standard set of capabilities for accessing the data sources.

The objective is to allow users in the external layer to access the participating data sources transparently through the mediation layer. Hence, the mediator architecture offers a transparent and uniform view to the shared data of the system [252].

The mediator is in charge of providing the capabilities needed by the integration system for processing global user queries. It mediates between global users and individual data sources via associated wrappers. It interacts with the data sources via wrappers and handles a user query by splitting it into sub-queries, sending the sub-queries to appropriate wrappers, and integrating the outputs from the wrappers before returning the final answer to the user. It is the job of the mediator to find relevant sources to answer the query among the various data sources and to obtain the answer to the query from them [278].

Busse et al [183] advocate the use of a mediator-based architecture for read-only integration systems that involve structured as well as unstructured data sources. Mediators can create materialized (e.g. [300]) or virtual (e.g. [278]) integrated views. Virtual mediators adopt the query-driven data management approach that only presents a virtual integrated view without storing data at the mediation layer. Hence, mediators with virtual integrated views offer the most suitable architecture for Web data integration [220], due to the frequent changes in most Web data sources and layouts.

3.5.3 Data-Management-Based Integration Architectures

This type of architectures is concerned with data management in the integration system and how global user requests are serviced. Two common approaches:

1) ***Data Warehousing Approach [208, 214, 242]:*** services requests using an additional repository, called a *Data Warehouse (DW)*, in the integration system. This technique resembles the materialisation approach that stores data in an integrated view. The problem with this approach is to keep DW up-to-date with changes at the underlying data sources. The cost of propagating changes to the integration system level (or to the data warehouse) is expensive for frequently changing data sources [291] such as a patient database or Web data sources.

2) ***Query-driven approach [159, 196, 224, 278]***: propagates requests to individual data sources based on virtual integrated views, i.e., no data is stored in the

integration system. This technique is dynamic and can cater for changes in the data sources or the user requirements. A query-driven approach is recommended when data freshness is critical or when it is impossible to load the entire data from the sources for processing [150]. This makes it a potential approach to use when linking highly dynamic data sources such as Web data sources and patient databases.

3.6 Building Web-based Integration System: Consideration Issues?

Web-based integration stresses the separation between the global user view and the data sources' views and seeks a unified access to the constituent data sources. Hence, three levels of abstractions (Figure 3.5) are crucial for successful Web-based integration, namely:

1. ***User-level***: depicts a user's view of the problem, and uses either the user's and/or domain concepts, and terminologies to formulate requests against the integration system.
2. ***Source-level***: denotes the individual data sources' views, interfaces, and implementations which are shielded by the integration system from the user.
3. ***Integration-(or Middleware) level:*** represents the integration system. From a user perspective, the integration system is the system that services their requests transparently. However, from the integrator perspective, the integration system is responsible for providing a number of more defined and specific capabilities that support the transparent access sought by global users.

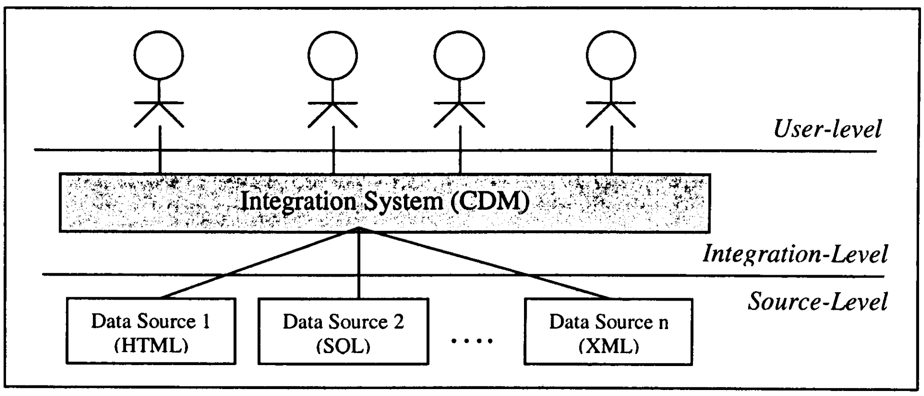


Figure 3.5: High-Level Three Layer Integration Architecture

When designing an integration architecture, prospective integration systems need to address the following:

a) *Integration level(s) of abstraction germane to the problem domain as explored in Section 3.5.1:* this is determined by the nature of the integration environment especially concerning structure, semantics, data sources' accessibility and the user's skills, perspective and requirements.

b) *The integration method (or architecture), as discussed in Section 3.5.2:* this choice is influenced by the data source's size, evolution, volatility and system scalability.

c) *Data management mechanism as discussed in Section 3.5.3:* this selection is based on the data freshness requirement, evolution rate and update cost.

d) *The common data model [220]* – often referred to as the Canonical Data Model (CDM) (see Section 3.6.1).

e) *Integration tasks [224, 278] (see Section 3.6.2).*

f) *Resource Discovery [203, 224] (see Section 3.6.3).*

g) *Semantically-related objects' identification mechanism [220] (see Section 3.6.4).*

h) *Source Mapping and Wrapper Construction [220] (see Section 3.6.5).*

3.6.1 Canonical Data Model (CDM)

The CDM is necessary for unifying data representations among the participating information sources. As far as semantic interoperability is concerned, the selected CDM must be able to capture both the structure and the semantics of the data. Traditional data models (e.g. relational) are inadequate for capturing and representing the full data semantics because in principle they were intended more for organizing and storing the data rather than for organizing its meta-data.

Recent research (e.g. [224]) in the field of semantic interoperability and data integration advocates the use of a *meta-model* as a proper CDM for interoperable information systems. Rasmussen [259] defines a meta-model as a model representing the structure and semantics of a particular set of models. Meta-model systems [224, 278] often employ a shared conceptualization such as a thesaurus or an ontology (see Section 3.7). The CDM needs to be expressive enough to accommodate such semantics, but simple and efficient [220].

3.6.2 Integrations Tasks

Studies [159, 224, 278] offer different classifications of key integration tasks towards semantic interoperability between disparate information sources. However, inherent integration tasks are influenced by the abstraction level at which the integration occurs (e.g. Schema-level, Data-level). Nonetheless, there are two common phases within the integration process: the translation phase, and the integration phase.

The translation phase attempts to unify the structural and possibly semantic knowledge about the shared (or integrated) data elements in various data sources. Hence, the translation phase might include an enrichment phase [224, 278] to unify and/or upgrade semantic knowledge of data sources. Thus, the translation phase could involve two steps:

a. ***Structural (or representational) transformation:*** unifying structural or representational knowledge of data in different data sources using a common CDM.

b. ***Semantic enrichment:*** unifying and upgrading semantic knowledge about data in disparate data sources.

In a federated architecture, the translation phase is achieved by translating the export schema into the structural and possibly semantic representation of the federation CDM. However, in the mediator architecture, the translation phase is taken over by associated wrappers. A schema translation phase is common to most integration levels as it denotes the representation used by the integration system using a given CDM, as illustrated in Figure 3.5.

The integration phase is concerned with the different steps required to integrate semantically-related objects. For the schema-level integration, the integration phase can be further subdivided into four steps, based on [224, 278]:

a. ***Information and Resource discovery:*** identifying information (or schema objects) to be shared and locating information that is of interest to the users (see Section 3.6.3).

b. **Information focusing:** identifying a subset of schema objects from the selected information sources that are relevant to the current user information requirement.

c. ***Detection of interschema knowledge and resolution of schematic differences:*** detecting semantic relationships between the relevant schema objects in various data sources (see Section 3.6.4) so that they can be integrated correctly and meaningfully, and then resolving their schematic (e.g. naming, models, representational) differences.

d. ***Generation of global views:*** providing and applying proper integration operators or linkage among semantically-related schema objects. This results in an integrated global view above local data sources.

Nonetheless, schema-level integration steps can be applied at other levels of abstraction (e.g. data-level integration). Elmagarmid et al [198] describe two main issues to the data-level integration: *Entity-identification*[15] and *Attribute-value conflicts*[16] [198]. These are covered by Step c of the schema-level integration phase.

However, user-level Web-based data integration needs to address strategies for resource discovery and information focusing in order to identify information sources that are relevant to the user information requirement. In addition, the integration system should have the facilities to assist users make proper mappings between semantically-related data objects in disparate data sources.

3.6.3 Resource Discovery and Information Focusing

The discovery process stresses how information of interest is identified for sharing and access. In federated systems, shared information usually models specific domain knowledge and might overlap with a local data source's data and concepts [203]. Hence, a conceptual model is key to the resource discovery process. Such a model can take part in negotiation with and be customised to users. In a database environment, the user and data source DBA can negotiate the shared information by identifying data and concepts modelled by the DB and of interest to the user.

[15] Entity-identification: "How does one identify representations of same real-world entity in different databases" [198].

[16] Attribute-value conflicts: "How does one deal with differences in data values among attributes that represent the same real-world entity" [198].

However, in a Web environment, identifying data sources of relevant data and concepts is not an easy task. Information Retrieval (IR) techniques and Web search engines can assist in retrieving specific information whereas a conceptual model can also be used to enrich the search. A popular resource discovery mechanism is Dublin Core (DC) [31]. However, DC addresses the structure of the data source rather than its conceptual knowledge. DC fails to identify semantic relationships between terms (e.g. hierarchies).

3.6.4 Semantically-Related Objects Identification

The identification of semantically-related objects is central to an effective and meaningful integration solution. Typically, different data sources might model the same real-world objects differently using different structures or terminologies. Data semantics is addressed in traditional database integration using five strategies:

- ***Ignoring data semantics at the integration level [148]:*** this leaves the burden of creating the search to the skills of the user to identify similar or relevant terminology. This can be an ineffective technique for Web data integration, as many Web users especially patients are unskilled in terminology mapping as discussed in Chapter 2.

- ***Using lexical-matching techniques and heuristics***: detects similarity or equivalence between objects, based on object and attribute names. These methods are common in IR systems. However, lexical matching fails to identify relationships between terms such as synonyms and hierarchical relationships.

- ***Utilising existing semantic structure modelling domain knowledge (e.g. Thesaurus):*** The scope of generic models (e.g. GALEN [260], WordNet [146], Read Codes [247]) could be less expressive (e.g. WordNet lacks hierarchical relationships) or too broad to model the given community or user information requirements, i.e., they fail to accommodate the user's requirements (e.g. medical versus lay terminology). Furthermore, it might require annotating of the local data sources with shared conceptualisation. For instance, HealthCyberMap [223] requires annotation of Web pages' Subject mark-up (or section) with ICD-9 [73] terms.

- ***Utilising a shared ontology (e.g. TAMBIS [170] Ontology (TaO) [169], Concept-layer in*** **[224]*)*:** this strategy implies a shared knowledge model that is constructed from the underlying data sources. In addition, data sources negotiate a shared agreement on the data semantics [148] and need to adhere to these semantics through an annotation process. However, shared conceptualisation is impractical in open access and read-only Web applications such as search applications for the following reasons:

  a) Constructing shared semantics from open Web data sources is not possible due to the large number of resources, the volatility of Web data sources, variations in and inaccessibility of data semantics or DBAs, and variations in Web user semantics (e.g. terminology) requirements.

  b) Web information providers may not be interested in annotating their Web pages and storing the shared conceptualisation.

  c) Shared semantics being newly constructed or generic can not cater for changing end user information requirements especially if it is designed by a small group and used by large groups.

Hence, utilising existing (or constructing a new) shared conceptualisation for Web-based data integration seems impractical. Instead, a user-oriented conceptualisation that emphasizes local application and user's perspectives of the domain knowledge is vital. Hieu [220] advocates the use of textual matching techniques from IR for Web-based integration. This can be assisted by a user-oriented semantic model to identify semantically relevant terms that are of interest to the user.

3.6.5 Source Mapping and Wrapper Construction

Source mapping offers a mechanism to link semantically-related objects. In traditional database integration, schemas are analysed and compared for similar objects. Semantically-related objects are integrated using some integration operators as in [159, 196, 224, 278]. However, in Web-based data integration, the data source schema is inaccessible. Hence, text-matching and constraint-based techniques can be used to map between schema or instance objects [220]. Source mapping is less problematic in data-level integration as it only affects the names (or labels) and terminologies of the data values. In fact, data-level integration is, in essence, a mechanism for establishing mappings between exchanged data in disparate systems [279].

Wrappers are used by the integration system for transparent access to local data sources. Relational databases are usually accessed by JDBC/ODBC wrappers. However, building a wrapper for a Web data source is problematic due to the frequent changes in structure and layout it undergoes. In addition, the high number of data sources makes it difficult to construct wrappers manually [220]. Web data source wrappers need to be as automated as possible [220]. Approaches to wrapper construction are discussed in [226]. This research addresses Web wrappers based on website document search capability rather than page layout due to the frequent layout changes, and using a search engine's APIs such as Google API [42].

3.7 Ontologies

Formally, an ontology is a statement of a logical theory [212]. It is defined as explicit formal specifications of the terms in the domain and relations among them [212]. An ontology is richer in representing data than a database schema. Typically, a database schema is concerned with the organization of the data within a database, and represents the structure of the data whereas an ontology is concerned with the understanding of the data [213], and represents the meanings of the data so that inconsistent interpretations or meanings of terms between different data sources is removed or minimized.

The ontology takes the form of a graph or hierarchy of concepts. Ontologies are expressed using some formalism (e.g. RDF [228], OIL [205], LOOM [255], SHOE [218]). These technologies vary in their representation format, expressiveness and reasoning capabilities.

An ontology-based integration architecture defines a conceptual layer to homogenise the semantics and terminology of the underlying data sources. Users formulate their queries using terms and concepts specified at this conceptual layer. This means users do not have to be concerned with the terms or concepts used in local data sources. The conceptual layer knowledge can be constructed in three ways by:

a. Constructing a shared ontology from the knowledge modelled by the underlying data sources, as in TAMBIS [170]. However, the construction and maintenance of such an ontology will be a huge task.

b. Employing an existing generic shared ontology (e.g. GALEN as in [278]) to unify the semantics or terminology of the local data sources.

c. Developing a user or application ontology that defines the terminology of a user or group of users. This could be based on a generic or domain knowledge model.

However, shared ontologies assume a consensus (or a shared agreement) on the meaning of the data and the terms that describe it. The shared ontology integration solution can be useful when multiple parties share the same view of the data and accordingly agree on using a common representation. This can occur among communities belonging to the same organization or among multiple organizations sharing the same goal and approach towards the shared data.

However, shared semantic models are inappropriate for disjoint communities as in the Web environment. Hence, the emphasis should be on the application and user perspectives and information requirements. Thus, a user-customised semantic model is vital for successful Web-based integration.

3.8 Summary

This chapter offered a background on the research problem from the computing perspective. We reviewed the area of data integration covering core concepts, inherent challenges, key issues and potential architectures. In addition, we offered a classification of data integration architectures based on three dimensions: integration level of abstraction, integration method and data management mechanism. This chapter has set out the directions for determining the architecture of the integration system created in this project and presented in chapter 5.

CHAPTER 4

Research Approach to Requirement Analysis

4.1 Introduction

Chapter 2 offered a background investigation of the research problem undertaken in this study. This chapter defines the problem and solution domains in terms of the proposed system requirements and features. A Requirement is a description of what relevant stakeholders want from the system [165]. It represents a feature of the proposed system desired by a stakeholder. This chapter describes a research approach to system development and requirements analysis.

Section 4.2 explains how this research is initiated. Section 4.3 explores the system development methodology. Section 4.4 discusses the system investigation. Section 4.5 examines a requirement elicitation and analysis process, and Section 4.6 concludes the chapter.

4.2 Project Initiation

At first, this research explored an approach to integrating data from relational databases with semistructured Web data. We sought an integration solution that adopted the Web user perspective of data semantics [152]. However, as the Web spans multiple perspectives due to different user communities, there was a need to decide on a specific user community as recommended to the author at BNCOD21 [10]. Hence, this research explored several user communities and a decision was made to investigate the research idea in the Health Informatics domain. Specifically, we focused on investigating an approach to customise patient access to relevant Web information based on a patient perspective. Accordingly, the research investigated the use of EPR, as a basic patient personal information model for focusing relevant Web information. Previous research at the Cardiff School of Computer Science [268] developed a Web portal to the Velindre ISCO patient database that offered out-patients information on their appointments and additional information resources from a database. Our research builds up on Sissons's patient information system [268] by exploiting EPR data to customise Web searching for patients. Hence, our research is concerned with investigating if EPR can be used as a basic (conceptual) data model for customising and improving patient Internet searching.

4.3 System Development Methodology

A system development methodology describes a framework for structuring, planning and controlling an information system development process [126]. Several approaches [126, 165] to system development are introduced which suit various system development considerations. This research adopts an incremental evolutionary development methodology, based on three development methodologies: the waterfall methodology, the prototyping methodology and the incremental methodology.

4.3.1 Waterfall Methodology

The waterfall methodology offers a linear and staged approach to system development [165]. It structures the system development process over separate sequential stages. Each stage has a firm goal and start and end points. Such a well-defined structured methodology supports planning, and produces excellent documentation [272] and helps measure progress [126]. However, it freezes the requirement elicitation stage in the early stages and thus makes it difficult to respond to changes in requirements later. In a real-world project, it is not possible to elicit complete and correct requirements until clients are given the opportunity to experience proposals through a trial version of the system [181].

4.3.2 Prototyping Methodology

A prototyping methodology develops the system as a series of prototypes. It offers an iterative approach over the system development time scale. Each prototype is used to elicit additional information requirements to further enhance the system. This feature allows for accommodating changes in requirements. In addition, it enables users to try the system while it is being developed. The prototyping methodology can lead to a very satisfactory system. However, iterative processes can add to the project budget and schedules [126]. In addition, for a limited time scale projects, the prototyping methodology can be very difficult to manage [268].

4.3.3 The Incremental Methodology

The incremental approach combines the waterfall and the prototyping iterative methodologies [126]. It breaks the project into smaller segments where each segment gets fully developed possibly using a mini-waterfall model [126]. The incremental approach prioritizes requirements, where the requirement of prime priority is delivered first and those of low priority are delivered in subsequent prototypes. This approach can break a long development time into smaller more manageable time units. In addition, it allows accommodation of evolving requirements [165] and incorporates knowledge gained from earlier segments development [126]. However, it is impractical when it is impossible to separate the system requirements into separate segments [268].

4.3.4 The Adopted Methodology

This research combined the incremental, prototyping and waterfall methodologies. The system requirements are split into separate segments. An initial perception of the application domain identified the four major requirement segments of the problem domain, as:

- Patients need online access to personal medical information held in their EPRs.
- Patient Internet search needs personalisation based on the content of EPRs.
- Addressing the Internet information quality issue.
- Addressing and enriching medical information vocabulary.

The solution system is implemented by evolving prototypes. Each prototype implements one or more requirement segments and follows the traditional waterfall model. System prototypes enable further requirements' elicitation for developing other requirement segments and subsequent prototypes. In the time-period of this research, three system prototypes were developed.

This approach was adopted for the following considerations:

a. Time constraints on the project, necessitated the development of an initial basic prototype to investigate and demonstrate the feasibility of developing the proposed system.

b. No prior system exists that offers online patient access to ISCO EPRs or personalises a patient's Internet search based on ISCO EPRs. Hence, prototypes are essential for developing a basic system that allows a stakeholder's view of proposed system features (or operations) and thus utilising prototypes as a technique for verifying or eliciting more correct user and system requirements.

c. User requirements (e.g. Internet information quality, search personalisation, vocabulary enrichment) are relatively disjoint; hence, they can be segmented in order to allow the development of various segments separately.

d. The system is developed within a new study that investigates solutions for problems inherent in the research area. Hence, an optimal solution is not a requirement for this study. Rather, the study aims to offer a core of practical functionality that demonstrates the feasibility of a solution system that can be further enhanced in future prototypes or studies.

4.4 System Investigation

The problem domain investigation is complicated by a number of issues, namely:

a. The lack of an existing information system or prior investigation study that personalises patient Internet searching based on EPR data and patient information requirements. of the research problem.

b. Patient-centred approach to healthcare is still an emerging domain, i.e. the new role of a patient in healthcare is not well-defined.

c. EPRs clinical data is aimed at professionals. Hence, identifying clinical data that is useful and meaningful for patients was not straightforward.

d. No official and adequate documentation is available on the ISCO system.

e. Conflicting and debatable patient and hospital requirements regarding Internet information quality.

f. Accommodating the various professional perspectives on Internet information quality is hard.

g. The author has no training in the medical domain.

Hence, a thorough investigation of the patient information sources, strategies, programmes and needs was crucial to obtaining an adequate knowledge and understanding of the problem domain to aid the system requirement analysis process. This investigation is presented in Chapter two.

Our initial main concern was how to improve patient access to relevant health information. A thorough exploration of traditional and recent patient information sources is given in Section 2.3. Internet information sources offer the foremost and greatest potential for improving patient access to relevant health information for the following reasons:

- The extensive and wide-ranging health information topics available online.
- The availability of key online patient and professional oriented health information resources.
- The mature security level of Internet-based access.
- The recent radical embracement of the Internet in national healthcare through the shared (integrated) EPR and the emerging PHR strategies (e.g. NHS Wales IHC [109], NHS England CfH [105]).

Hence, this research further focused on improving and customising patient access to relevant Internet information. As we propose to improve patient Internet searching through personalisation and based on a patient's own EPR data, Sections 4.4.1 and 4.4.2 summarise our investigation findings regarding a patient's EPR access and Internet searching respectively.

4.4.1 Patient EPR Access

At the time of this investigation, a patient in Wales can access his/her EPR data by making an official request. However, the newly developed IHC's health information infrastructure through IHC strategy [109] offers a prospect for direct patient online access to EPRs. EPRs offer up-to-date patient personal medical information, and a basic patient information model that can indicate likely topics of interest to a patient. However, EPRs are usually modelled for clinicians and utilise clinical terminology. Nonetheless, our exploration into the patient database (i.e. ISCO) EPRs identified a set of clinical data (see Figure 2.8) that was deemed useful for patient Internet searching. However, less highly educated patients might find it difficult to understand and manipulate the medical terminology used for EPR data. Therefore, there is a need to explain medical terms for a patient.

4.4.2 Patient Internet Searching

A patient can access the Internet as long as they have a computer and a network, either from the hospital patient information centre or elsewhere. The Internet offers extensive, wide-ranging and up-to-date health information through several mechanisms (e.g. general-purpose search engines, authenticated search engines, national health gateways, medical search engines, and charity (patient-oriented) websites). However, patient Internet searching is hindered by the following:

1. **Inaccessibility of patient personal medical information** that a patient needs to utilise in medical Internet search. Studies [145, 188, 229] indicate that patient Internet medical search is closely related to their health problems. However, a patient usually lacks access to his/her EPR and has difficulty retaining and memorising verbal information imparted at a consultation. Thus, a patient is challenged to utilise their own knowledge regarding their medical information correctly during Internet searching. In the light of inaccessible official electronic medical records, some patients tend to maintain private paper or electronic records of their health condition details. However, such an unofficial record is difficult to maintain comprehensively, correctly and up-to-date. This situation can complicate the patient online search experience.

2. **Variant patient search information requirements:** There are several information types that a patient may want to look up but often they have difficulty locating them and/or stating them correctly.

3. **Generic health websites list, utilised by ISCO patients:** Patients at the Velindre NHS Trust utilise a generic list of key Internet health information websites, in paper format, that needs to be automated and further customised for a patient according to their health condition by the staff.

4. **Laborious, manual and generic nature of patient Internet search:** A patient can access Internet health information by typing a website URL, browsing subject headings or by using a search engine. Some of the drawbacks are: Subject headings may not contain or indicate their relationship to patient information requirement, and typing URLs can be error-prone especially if a patient is reading from a list. The use of a search engine is the foremost patient Internet access mechanism [145, 206]. However, it can be frustrating and time-consuming for a patient due to the lack of search topics (or ideas), manual entry, and the search having a large quantity of results that are difficult to judge for relevance and quality.

5. **Wide-ranging and disparate nature of Internet health information tools:** Such as search engines (e.g. generic, medical, mediated (see Section 2.4.4.1) search engines)), health gateways and charity websites.

6. **Internet information quality:** Due to the global and open nature of the Internet, Internet information is uncontrolled, difficult to judge and could harm or damage patient care. The following problems are noted in this regard:

    - Generic search does not indicate trusted websites to patients.

    - No authoritative feedback from healthcare providers is given to guide patient Internet access.

    - Healthcare providers do not take advantage of the considerable Internet information research patients usually conduct prior to a consultation session. Thus, patients and healthcare providers do not share and communicate trusted Internet information resources.

7. **Health information vocabulary:** Due to the lack of accessible personal medical information and the mostly scientific or official nature of Internet health information resources, a patient can experience the following information vocabulary challenges when searching:

   - Difficulty in expressing the correct medical term describing their sought information.
   - Difficulty in formulating proper lay terms. Studies indicate that patient-specified lay terms do not lead to successful search results [232].
   - Difficulty in identifying related vocabulary (e.g. synonyms, hierarchies).

8. **Internet information overload:** Denotes the increase in information volume to a limit that makes it difficult for users to assimilate. In a Web environment, information overload can be due to the large number of Internet health information resources. It is exacerbated by the large number of Internet search results and the lack of tools to process and compare them [70] and identify relevant and safe ones. Thus having to deal with too much information can be very stressful especially for patients.

9. **Information pollution:** This is similar to information overload in that it addresses irrelevant details that a user needs to navigate before hitting useful information. However, information pollution could occur at a fine-grained level, at the content level in a document or a phrase while information overload is mostly referred to as a volume problem. The literature discusses different perceptions on information pollution:

   - Information pollution implies misinformation [179, 248].
   - Information pollution implies too much and unorganised information [244], i.e., information overload.
   - Information pollution implies worthless details [246]. This perception is similar to the problem targeted by the Plain English Campaign (PEC) [120]. However, a review of literature shows no link as yet established between the two terms.

10. **Lack of Internet information coordination and sharing between patients and professionals:** Patients surfing the Internet for health information usually take their Internet information resources to their healthcare providers

during consultation. This might disrupt the vital consultation time. In addition, studies [280] indicate that healthcare officials are usually less familiar with Internet health information than patients, and need to be updated with Internet health information resources. Ethically, healthcare providers should be concerned about patient access to Internet information [189, 264, 268]. Health information strategies [15, 37, 103, 115, 134] accentuate the availability of trusted Internet health information resources. In addition, healthcare providers need to recognise and support patients' partnership in their own healthcare and endorse their efforts and contributions.

4.5 Requirement Analysis

The Requirement Analysis process denotes the process of understanding and defining stakeholders' needs for the proposed system. The Requirement Analysis stage includes five distinct steps, based on [125]:

a. ***Setting System Boundaries:*** identifies how the proposed system integrates with the business logic and what will be its scope and limitation.

b. ***Identifying Stakeholders:*** identifies the groups of people who are directly or indirectly impacted by the proposed system.

c. ***Requirement Elicitation:*** describes the types of information gathered from various stakeholders and the mechanisms used.

d. ***Requirement Analysis:*** analyses the gathered information and identifies various stakeholders' information needs.

e. ***Requirement Specification:*** specifies the identified stakeholder's information needs in a well-defined and unambiguous manner.

4.5.1 Setting the Boundaries of the Proposed System

This research is conducted as an investigational study to explore the idea of building a patient-customised Internet search based on a patient's personal medical information. As a first study of this problem in the Health Informatics domain to our knowledge, the main research concern is on two major issues:

a. Identifying major challenges (or problems) relating to a patient's EPR access and customising a patient's Internet medical search as discussed in Sections 4.4.1 and 4.4.2.

b. Identifying and implementing core functionality for resolving such challenges.

Due to the lack of an existing similar system, the proposed system represents the first Web-based interface to the patient database utilised in this study, i.e. ISCO, that delivers the sought functionality. Hence, the study is not seeking an optimal solution system, but identifying inherent challenges and core functionality needed to address such challenges.

4.5.2 Identifying Stakeholders

Stakeholders are different groups of people with special interest and perspective on the problem domain. An investigation of the problem domain indicates three potential stakeholder groups, namely:

a. Patients registered in the ISCO system.
b. Hospital staff interested in indicating trusted health websites to patients.
c. An information specialist with two major tasks:
    - Identifying third-party accredited health websites for assisting hospital staff build a list of trusted websites.
    - Defining medical-to-lay terminology mappings and verifying the generated Patient Health Information Vocabulary (PHIV).

4.5.3 Requirement Elicitation

The Requirement Elicitation attempts to capture the necessary information about the problem domain from user perspectives, before specifying and developing a solution system. The information required covers different aspects:

- The problems to be solved [180].
- Different stakeholders' needs [237].
- Proposed system's expected features [237].

- Constraints on the creation or behaviour of the proposed system [180].

The level of detail required is determined by the complexity of the inherent problems and possibly conflicting stakeholders needs [125]. In this study, requirements' elicitation and analysis is intertwined with various system prototype developments. The information needed was elicited using the following techniques:

a. Background reading.
b. Document inspection including:
    - ISCO tables document prepared by Sissons [268].
    - Patient Casenote screen in ISCO Staff interface – describes a summary of patient information – obtained from Jones [221].
c. One-to-one interviews with information staff [164], radiotherapy patient information nurse [184], chemotherapy patient information nurse [263], ISCO DBA [173] and several ISCO team members (e.g. [167, 221]).
d. Exploring ISCO data dictionary.
e. Attending a session on the ISCO system delivered by Jones [221].
f. Attending a workshop on patient Internet access [131].
g. Feedback from conferences on published papers [153, 154].
h. Proposed system prototypes.
i. Discussion with supervisor [210].
j. Discussion with Medical Doctors [156, 161] and a medical student [163].

The project started by investigating an approach to guide patients to relevant Web health information based on their personal medical information. First, the ISCO patient database was analysed to identify and extract essential EPR data on a patient's diagnosis and treatment as these are the focussed health information types sought by patients. However, as the first ISCO version utilised by this study lacked treatment data, the project, initially, investigated ISCO diagnosis data and linking it to relevant Internet information sources. ISCO encodes patient diagnosis information using Read Codes that can be decoded from Read Code Terms stored by ISCO. Therefore, the first prototype presented a patient with his/her own diagnosis information and a patient Personal Internet Search (PerIS) facility that searches for Internet information sources relevant to the patient's diagnosis.

Subsequently, a fuller anonymised version of the ISCO database was downloaded by Velindre CIU that contains information on various treatment episodes and a cancer management plan. This is followed by a series of one-to-one interviews with several ISCO developers to identify ISCO data describing different patient treatment episodes and the proposed patient cancer management plan. Further interviews were conducted with a patient information specialist [164] and nurses [184, 263] to verify the meaningfulness and the usefulness of the extracted ISCO data for patients and to elicit further information types often sought by patients. Aston [164] recommended background reading and provided articles on key studies and surveys on patient information needs and patient Internet access.

Accordingly, the second prototype utilised a patient treatment and the cancer management plan data as personalised search ideas in PerIS. Additionally, the second prototype addressed Internet information quality by incorporating search tools in PerIS that indicate key and hospital trusted health information websites to patients. Internet information quality seals and accredited websites were elicited from Aston [164], literature readings and a workshop on patient Internet access [131]. The gathered information was discussed with the project principal supervisor Prof. W.A. Gray [210]. Subsequent feedback was obtained from ISHIMR06 conference, based on [153]. Lastly, a third prototype was developed to address two issues:

- Medical term enrichment.
- A suggestion by ISHIMR06 conference to feedback to professionals about the patient's chosen websites.

4.5.4 Requirement Analysis

This section analyses the elicited information to identify inherent problems, different stakeholders' needs, the features the new system should offer and any constraints imposed on the solution system.

Based on the challenges pertinent to a patient's EPR access and Internet medical searching, described in Section 4.4, a precise record of problems inherent in the problem domain is established (see Appendix B.1). Stakeholders needs reflect the problems in the problem domain. An analysis of these problems in terms of the stakeholders' perspectives allows us to define the stakeholders' needs from the proposed system (see Appendix B.2). Stakeholders needs form the functional requirements of the proposed system.

Subsequently, a Solution Domain [237] needs to be analysed and specified that maps to the Problem Domain. The Solution Domain consists of the features that the proposed system should provide and any constraints imposed by the application technology environment on the proposed system. A feature is a service that the proposed system provides to fulfil the stakeholders' needs(s) [237], whereas a constraint is a condition or restriction that must be satisfied [233] by the proposed system. The constraints constitute the non-functional requirements of the proposed system. The transition from Problem Domain to Solution Domain supports the traceability of proposed system features to the corresponding needs [237]. Appendixes B.3 and B.4 illustrate the features and constraints (i.e. non-functional requirements) of the proposed system respectively.

4.6 Summary

This chapter presented a research approach to system development and requirement analysis. The study adopted an incremental evolutionary development methodology to enable evolving system development and eliciting further requirements through different system prototypes. The initial system prototype offered basic functionality but demonstrated the feasibility of extending ISCO to patients. The chapter has defined a clear account of domain problems, stakeholders' needs and proposed system features and constraints.

CHAPTER 5

The Patient Health Base (PHB) Integration Architecture

5.1 Introduction

Chapter 3 reviewed potential architectures for data integration and semantic interoperability. This chapter presents the Patient Health Base (PHB) system integration architecture. Chapter 4 outlined two features of our proposed PHB system that highlight integration problems, namely:

1. *Personal Internet Search (PerIS) (see F7, Appendix B.3)*

2. *Patient Health Information Vocabulary (PHIV) (see F2, Appendix B.3)*

The patient Personal Internet Search (PerIS) system links, i.e. integrates, data from EPRs to relevant Web information. Patient EPR data resides in the ISCO system which is a relational database system whereas Web information may come from standalone Web documents or Web portals linking to legacy databases. PerIS, as a Web-based system, needs to deliver the integrated information to patients in a Web format that is understood by Web browsers.

As EPRs are typically described in medical terminology, the Patient Health Information Vocabulary (PHIV) aims to explain medical terminology and enrich PerIS search results for a patient. Based on patient information needs (see N24 and N25, Appendix B.2), we advocate a patient-oriented health information vocabulary model that covers terminology features that are of interest to patients. In addition, PHIV defines and integrates medical and lay health information vocabulary so it can be correctly utilised by patients. The medical terms come from the patient database whereas the lay terms are defined by a patient information specialist from the lay perspective, and stored in a Concept Thesaurus (CT) database.

In this study, the PHIV integration process generates an integrated conceptual view that will be utilised by the PerIS integration system (see Figure 5.1) to homogenise and enrich health information terminology for PerIS users.

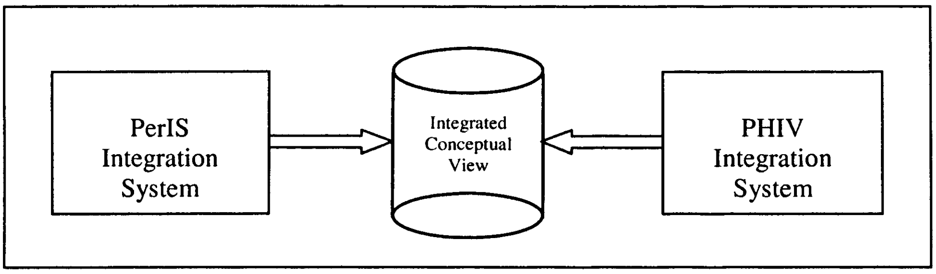


Figure 5.1: Two Parts PHB Integration Architecture

Section 5.2 discusses the research approach to building an integrated patient-oriented conceptual model whereas Section 5.3 describes the research approach to integrating a patient's ISCO data with relevant Internet information sources.

5.2 PHIV Integration Approach

5.2.1 Motivation

Recent data integration approaches employ conceptual structures such as meta-models, ontologies or thesauri, to unify different perceptions on data belonging to autonomous information sources. Common conceptual integration approaches employ two conceptual model types:-

- A generic domain knowledge model.
- A shared conceptual model that is constructed from concepts modelled by the participating data sources.

In the healthcare domain, generic knowledge models (e.g. GALEN, MeSH, Read Codes) are normally aimed at professionals, and employ medical terminology. In addition, they model the needs of common user communities. This makes it difficult to tailor it for a specific user community especially patients. On the other hand, shared knowledge models require access to the semantic knowledge of the participating data sources and achieving consensus (or shared agreement) on the shared conceptualisation and how to use it.

In this research, the proposed patient search information system (i.e. PerIS) is a Web-based integration system that integrates data from patient medical records with relevant Internet information sources. The integrated data can be described using terminology specified by different parties, such as:

- The patient database.
- Many Web information sources, each having an individual perception on the exchanged information.
- Different patients have different ways of describing health information possibly according to educational and/or cultural factors.

In the current Web infrastructure, Web information sources only deliver data and do not allow users (or applications) to retrieve the intended data semantics nor the data source structure. A minimal conceptual knowledge can be recovered from data labels using lexical matching techniques. Very recently, the Semantic Web initiative advocated exchanging data semantics with the data. However, this approach is still an emerging one and applied to certain Web-based projects but not the current entire Web. Hence, building a shared conceptual model from Web information sources seems impractical at the moment.

Accordingly, in this research, we choose to build a patient-oriented conceptual model that models a patient health information vocabulary according to patient information needs. Two distinct terminology perceptions are noted by patients:

- **Medical perspective:** describes health information using medical and/or scientific terminology. While the medical perspective is usually utilised by professionals and the patient database, it is demanded by highly educated patients as well (see N18, Appendix B.2).
- **Lay perspective:** expresses health information using a simplified terminology that can be understood by laypeople. This perspective is usually advocated for and sought by novice patients (see N19, Appendix B.2).

Hence, this research sets two aims to building a Patient-oriented Health Information Vocabulary (PHIV):

a. Integrating medical and lay perspectives on health information vocabulary.
b. Accommodating patient information needs.

Our PHIV conceptual model is restricted to patient diagnosis concepts for the following reasons:

- Time constraint on this research study.
- Diagnosis information is the most commonly sought information by patients [229].
- Patient terminology challenges are paramount when expressing diagnosis information [232].

Consequently, our discussion will focus on integrating diagnosis information from the medical and lay perspectives according to patient information needs. The resulting diagnosis knowledge model is referred to as the Patient Diagnosis Ontology (PDO). PDO integrates three perspectives on diagnosis concepts:

a. The patient medical community perspective as modelled by the ISCO patient database and auxiliary medical classification models utilised by ISCO.
b. The lay perspective that describes a patient's lay literature diagnosis terminology. In this research, we choose to enable a patient information specialist to specify the lay diagnosis terminology so it can be correctly mapped to the medical terminology and be effectively used by patients. Studies indicate that patient specified lay terminology is usually inaccurate and leads to misleading search results [232]. The information specialist is assisted by a Concept Thesaurus (CT) interface to specify the medical-to-lay term mappings.
c. The patient perspective in terms of information vocabulary needs (see N18 – N23, Appendix B.2).

Accordingly, PDO denotes an integrated conceptual view above the medical and lay diagnosis terminologies modelled by ISCO and the Concept Thesaurus (CT) databases respectively. Section 5.2.2 describes the research approach to constructing PDO as an integrated diagnosis conceptual model.

5.2.2 PDO Integration Approach

PDO denotes an integrated conceptual view above the medical and lay diagnosis terminologies modelled by ISCO and the Concept Thesaurus (CT) databases respectively. Two component DBs are involved in this integration:

- ISCO DB delivers diagnosis medical terminology.
- CT DB defines additional term mappings not covered in the ISCO DB, especially medical-to-lay term mappings.

The aim of this integration task is to create an integrated view of medical and lay diagnosis concepts that can be unambiguously understood and utilised by patients. It establishes proper mappings of the patient diagnosis in similar medical, lay and generic terms and defines proper mappings between the three term categories. Hence, each diagnosis concept is associated with medical synonyms, lay synonyms and generic synonyms.

The ISCO DB stores medical diagnosis description in Read Code clinical terms. Additional medical synonym descriptions are recorded in the ISCO Keyv2 table. Moreover, diagnosis concept generic terms are stored in the ISCO Classification table. However, ISCO lacks lay descriptions for its Read Code diagnosis concepts. Hence, CT was created to deliver basic medical-to-lay term mappings. Then, an algorithm was defined to create lay descriptions of ISCO medical diagnosis concepts based on CT medical-to-lay mappings. Generated lay diagnosis descriptions were then linked to an ISCO Read Code diagnosis concept.

PDO is created using a data-level tightly-coupled federated integration approach (see Figure 5.2):

- It is a federation for two reasons:
    a. It is built by the cooperation of the participating DBAs; ISCO DBA and CT DBA (information staff).
    b. ISCO schema is disclosed to the Federation Administrator (FA) to investigate ISCO diagnosis information that can be represented in the PDO at the federation level.

- Tightly-coupled as the generated federated schema is fully specified and controlled by the FA who establishes the mappings between various data objects.

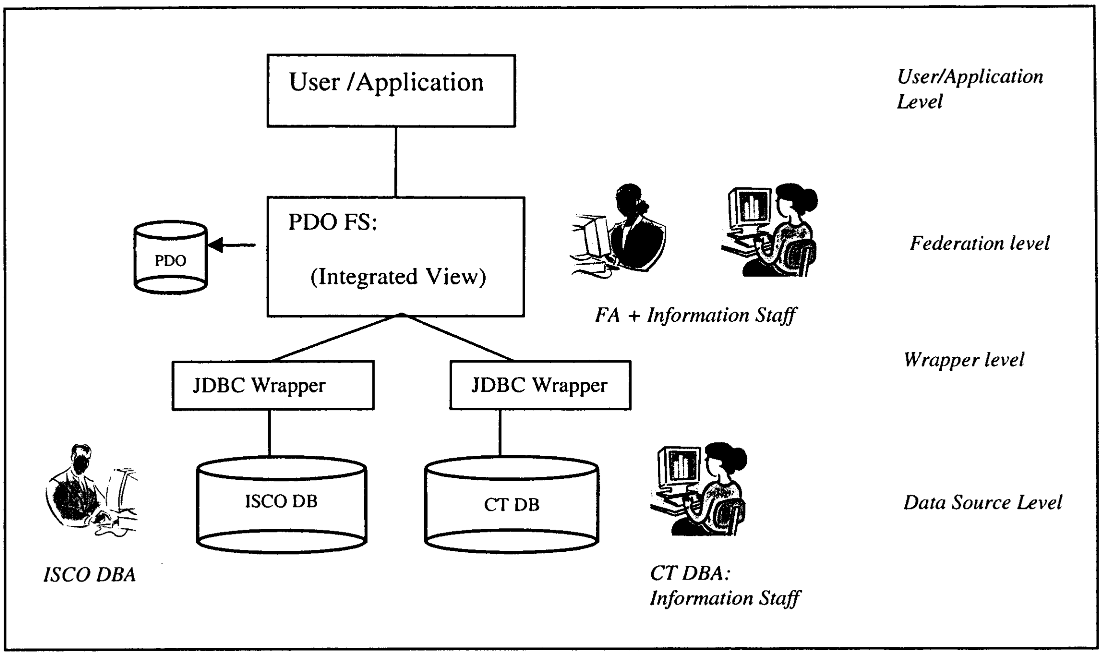


Figure 5.2: PDO Data-level Tightly-Coupled Federated Architecture

In fact, a large part of the PDO integrated view is already established in the ISCO DB. ISCO records specific/generic diagnosis classes, medical diagnosis synonyms, and generic diagnosis concepts in Read Codes and ICD-9 medical classifications. PDO extends the mappings defined in ISCO with medical/lay mappings, similar lay terms, additional similar medical terms and additional similar generic terms not defined by the ISCO DB. This approach can cater for the diagnosis terminology commonly used in both medical and lay patient information literature and not used by ISCO. The PDO FS offers a generic capability to create such mappings.

5.2.2.1 Joining the Federation

Classically, information providers indicate various information that global users or applications may wish to access:

- **ISCO DB:** PDO FS extracts every diagnosis concept and its generic class from the ISCO Classification table, and diagnosis medical term synonyms from the ISCO Keyv2 table. This is investigated and negotiated with the ISCO DBA [173].

- **CT DB:** PDO FS extracts the entire CT data that is created only for utilisation by FS services.

5.2.2.2 Generating the Integrated/Federated Schema

Typically, the federation system restructures the shared or exported information according to the global users' information requirements. In this integration problem, the resulting integrated schema is restructured according to a patient's health information vocabulary needs as investigated in Chapters 2 and 4.

This integration problem seeks one integrated schema that defines a patient health information vocabulary describing the patient's diagnosis information that we refer to as the Patient Diagnosis Ontology (PDO). For a given diagnosis concept, PDO defines a set of its medical term synonyms, a set of its lay term synonyms and a set of its generic term synonyms (see Figure 5.3).

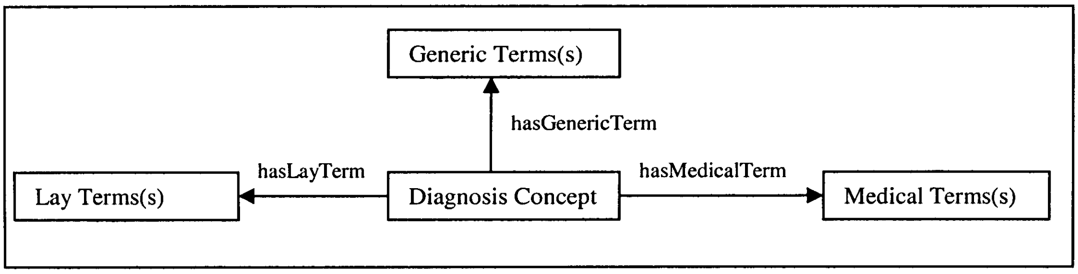


Figure 5.3: PDO Structure (Integrated Schema)

The PDO integrated schema is represented as an RDF model. It is materialised with instances based on the extracted ISCO and CT data and stored in RDF/XML format. ISCO DB provides a basic set of medical and generic diagnosis term synonyms. A set of corresponding lay term synonyms and additional medical synonyms are generated by the two algorithms employed by the federation system:

- **Lay diagnosis descriptions algorithm (see Appendix C.1):** constructs a set of lay synonyms corresponding to the ISCO diagnosis concept medical synonyms using the CT DB medical to lay term mappings.

- **Additional medical diagnosis synonyms algorithm (see Appendix C.2):** utilises CT to construct common diagnosis medical term synonyms not defined in the ISCO system.

FS algorithms use a text-matching technique to detect similar terms defined in ISCO and CT diagnosis term descriptions.

5.2.2.3 Verifying the Integrated Schema Mappings

The Federation system assigns to an information staff member, a task to manage the instances of the integrated schema, i.e. PDO, in a GUI screen. For a given diagnosis concept, the information staff can view its medical synonyms, lay synonyms and generic terms. In addition, s/he can delete incorrectly constructed medical, lay or generic diagnosis terms and/or add proper ones.

5.2.2.4 Integrated Schema (or PDO) Evolution

PDO evolution is determined by evolution in global user requirements and/or the participating data sources. As this integration process is geared towards patient health information vocabulary requirements as investigated in this study, we do not expect immediate evolution in these requirements.

However, evolution in a data source's data is accommodated by propagating the changes to the materialised schema. The information staff can refresh PDO (or the materialised schema) for any changes occurring in CT or ISCO diagnosis concepts' classification data.

5.2.2.5 Benefits of this Approach

The benefits are, it:

1. Allows specialists to define correct mappings of health information vocabulary, so it can be correctly used and interpreted by patients who are non-specialist novice global users.

2. Accommodates health information terminology from both the medical and lay perspectives.

3. Employs information staff as part of the federation system to deliver to a patient an information specialist view on medical and lay health information vocabulary that are common in literature, and the terminology applicable to the patient domain.

5.2.2.6 Limitations of this Approach

The limitations are:

1. It accommodates only the health information vocabulary defined in the ISCO system or added by the information staff.

2. Creating CT terms and verifying the mappings can be a lengthy and troublesome task for the information specialist. However, since PDO is a single integration system, these should not require many changes in the future. In addition, PDO constitutes a small-scale integration task as it only covers diagnosis terminology. Hence, the verification time may not be too long.

5.3 PerIS Integration Approach

The patient Personal Internet Search (PerIS) system builds patient personalised search topics based on a patient's own EPR so it can retrieve Internet information sources that are relevant to the patient's condition. The PerIS integration problem is characterised as follows:

a. Web data sources are highly autonomous and do not enable access to their structure and functionality. Hence, structure-level integration is not feasible for this integration problem.

b. There is a need to resolve EPR medical vocabulary for patients and identify relevant terminology describing patient health information that might be used by the patient database, Web data sources or patients themselves. Hence, a conceptual integration level needs to be incorporated.

c. Different patients have different search information requirements, i.e., different ways of linking sought health information to relevant Web documents. For instance, some patients may wish to search for information described in medical terminology using professional-oriented or medical health gateways. Other patients may wish to search for health information described in lay terminology using charity health websites. In addition, the same patient may want to search for the same lay health information using hospital-trusted websites. Hence, a patient can specify different ways of linking or integrating ISCO-based search information with relevant Internet information sources. This indicates a loosely-coupled integration procedure where the end-user selects the integration units and the way of linking (or mapping) them.

d. Web data sources are not aware of this integration process. Hence, this integration problem can not adopt a federation architecture as local data sources can not negotiate the shared or exchanged data.

e. Patients as the users of PerIS require fresh data from either ISCO or Web data sources. Hence, we are seeking a query-driven rather than a materialised (or data warehouse) integration system.

f. Patients are neither skilled in querying a database structure and data, nor in identifying and querying relevant Internet information sources. Hence, there is a need to mediate the PerIS integration process for a patient to alleviate a patient from the underlying data sources' technicalities.

5.3.1 PerIS Integration Architecture

PerIS represents an integration system that maps data from a patient database to their relevant Web documents. Hence, this task constitutes a data-level integration that interoperates between the ISCO database and Web documents at the data-level. PerIS is based on a loosely-coupled mediated architecture (Figure 5.4) with a conceptual structure to enrich diagnosis search information.

As a Mediation System (MS), PerIS mediates between a patient as a global user and the underlying data sources. First, it interoperates with the ISCO database to extract patient personal medical information. Diagnosis-based search information is enriched with relevant terminology from the conceptual structure – PDO. PerIS queries Internet information sources for information relevant to the patient search query.

The PerIS integration architecture consists of four layers:

- **Information Source Layer:** provides the information to be accessed by global users. It consists of the ISCO patient database and Web information sources.

- **Wrapper Layer:** comprises interfaces (or API) that enables PerIS to communicate with ISCO DB and Web information sources.
- **Mediator Layer:** handles and coordinates the interactions between global users, the underlying data sources and various components utilised by PerIS. PerIS incorporates a set of focusing techniques that enable patients as global users to select the information units of interest to their current search information requirements (see Chapter 7).
- **User/Application Layer:** consists of patients as global users.

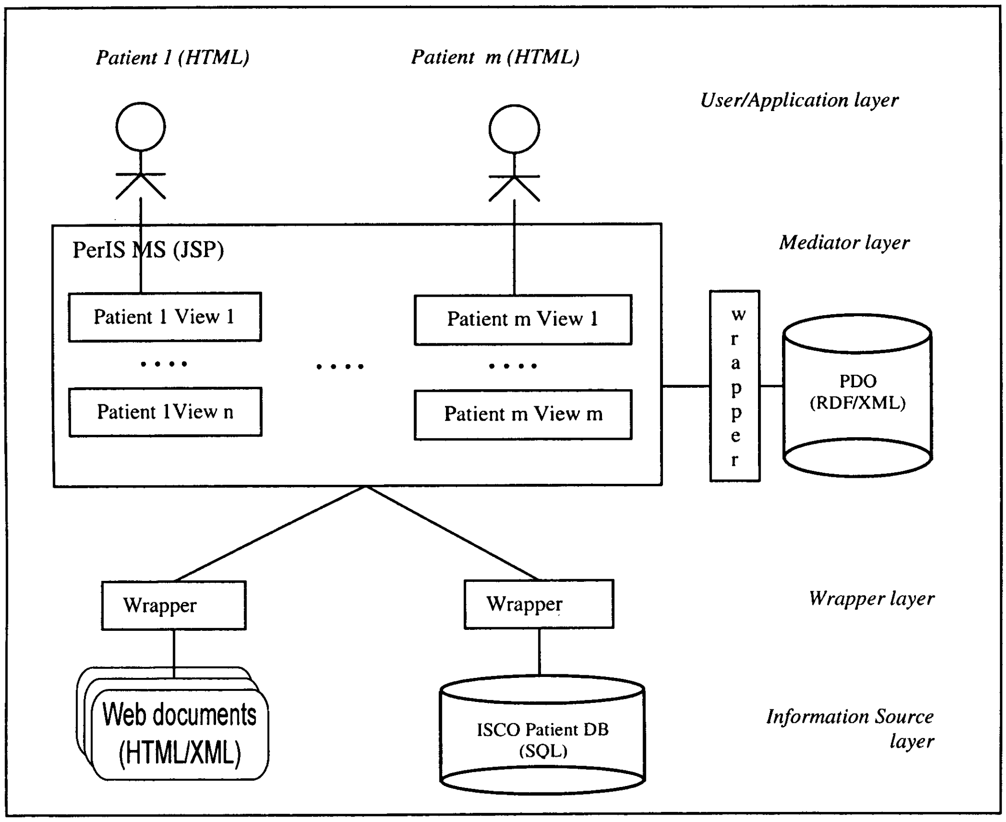


Figure 5.4: PerIS Loosely-Coupled Mediator Architecture

The PerIS mediation system allows each patient to construct one or more virtual integrated views over ISCO and the Web that integrates ISCO data with relevant Web documents in different ways.

PerIS is a virtual system, i.e., it does not store data at the mediation level. Rather, it is a query-driven system that propagates queries to relevant data sources: ISCO, PDO and various Internet search tools. A patient specifies his/her search query using an HTML form within a Web browser. PerIS analyses the patient request and breaks it into subqueries that are executed against relevant information sources.

Results are reconstructed and returned to the patient in an HTML format indicating relevant Web document title and link.

5.3.2 PerIS Wrappers

PerIS interacts with the underlying information sources through a set of wrappers. ISCO DB is a relational database managed by Windows 2000 server. PerIS communicates with ISCO DB through the JDBC wrapper technology.

Internet information sources are usually queried (or searched) by two means:

- A website document search service.
- A generic Web search engine (e.g. Google).

In this study, we utilise the Google API that enables the Google engine to search from application programs. Google API is also used to construct patient-customised search engines that search specific websites, as discussed in Chapter 7. Additionally, an internal wrapper based on the website search document service is implemented to enable the search of key health gateways and charity websites. Furthermore, PerIS communicates with PDO via the JENA OntologyAPI that enables querying the PDO as an RDF model.

5.3.3 PerIS CDM

As a data-level integration system, PerIS is concerned with the representation of the integrated data. Data are extracted from the underlying data sources using JAVA API wrappers (e.g. JDBC, GoogleAPI, JENA OntologyAPI), and hence are represented in the JAVA language. The PerIS mediation system is implemented as a suite of JAVA Server Pages (JSPs) running on a TOMCAT Web server. JSP enables the representation of data in JAVA and Web format (e.g. HTML). This allows global users to communicate with the PerIS integration system using HTML forms in a Web browser.

PerIS employs a semantic enrichment to patient diagnosis information using a PDO conceptual structure to homogenise and unify patient perception on the diagnosis terminology that is extracted from various ISCO and Web data sources.

5.3.4 PerIS Integration Tasks

A typical integration exercise consists of the following steps:

- **Resource discovery:** locating data objects that are relevant to the user information requirement.
- **Information focusing:** selecting a subset of the relevant data objects that are pertinent to the current user information requirement.
- **Detection of semantic similarity:** identifying similar and related data objects in the underlying data sources.
- **Generation of global views:** mapping between similar or related data objects in various data sources.

The following sections illustrate PerIS's approach to implementing such tasks.

5.3.4.1 PerIS Resource Discovery

Resource discovery is concerned with identifying information satisfying global user information requirements. In a mediated architecture, the mediation system is responsible for identifying such information.

As a patient-customised Internet search system, PerIS anticipates and incorporates mechanisms that enable the discovery of relevant information resources. PerIS incorporates several techniques to assist resource discovery:

- It utilises a patient's personal medical information to ensure the discovery of information resources that cover essential information on a patient's health condition.
- It incorporates a patient-oriented conceptual model to ensure the discovery of Internet information sources that are of interest to patients and relevant to the patient search query.
- It incorporates a list of search refinement topics that assist in discovering information currently sought by patients.
- It accommodates a wide-range of Internet search tools to enable patients to discover key Web documents that are relevant to the current patient search information requirement.

5.3.4.2 PerIS Information Focusing

Information focusing techniques allow a user to select specific subsets of relevant data objects. PerIS incorporates a variety of search information focusing techniques that allow patients to select given relevant information, namely;

a. Patient personalised search ideas on treatments, diagnosis and cancer management plan.

b. PDO enables a patient to focus the search on a given diagnosis term or a given term category (e.g. medical terms only, lay terms only or generic terms).

c. Search refinements that focus the search topic to a given information type (e.g. risk factor).

d. Search tool: PerIS incorporates an array of Web search tools that restrict patient Web search, to a given type of information websites.

e. Search domain: PerIS allows the search of a single website, group of websites or the entire Web.

f. Search language: focuses the generic Web search to websites using the patient's preferred language.

PerIS focusing techniques are described in more details in Chapter 7.

5.3.4.3 Detection of Semantic Similarity in PerIS

As a data-level integration system, PerIS maps between similar data objects in the interoperating data sources. Data-level integration systems typically employ text matching techniques to identify the shared data. PerIS uses Web and document search mechanisms to locate documents containing terms of the patient sought information.

In addition, PerIS utilises PDO to detect similar or related data objects. PerIS employs PDO to enrich search results with additional Web documents containing similar or related terms.

5.3.4.4 PerIS Global View Generation

The mapping of patient search information to relevant Internet information sources constitutes a virtual integrated global view above the ISCO DB and the Web. These mappings are achieved by PerIS search options and influenced by various PerIS information focusing techniques.

5.3.5 Benefits of PerIS Integration Approach

The benefits are:

a. PerIS positions patients as global users with a simplified and unified interface above the ISCO DB and Internet information sources. Patients are not required to query ISCO or individual Internet information sources separately using local query methods.

b. The PerIS mediator saves a patient from exploring individual information sources to identify search information integration units.

c. PerIS incorporates a conceptual integration level to enable the identification of diagnosis relevant terminology and the utilisation of a patient preferred PDO terminology.

d. Incorporating a rich functionality that assists effective focusing of search information requirements.

e. PerIS offers a dynamic and flexible means of integrating personal medical information with different sets of Internet information sources.

f. As a query-driven system, PerIS offers patient's access to fresh and up-to-date information from the ISCO DB and relevant Internet information sources.

5.4 Summary

This chapter discussed the Patient Health Base (PHB) system integration architecture. Our approach to personalising patient Internet searching incorporates two integration problems that were addressed in this chapter, namely:

- Establishing a materialised integrated conceptual view on the medical and lay vocabulary perspectives according to patient vocabulary information requirements.
- Establishing a virtual integration system that enables patients to build several virtual integrated views above their EPR data and Internet information sources. Each patient integrated view defines a way of mapping between patient personal health information and relevant Internet information sources as desired by a patient.

CHAPTER 6

PHB Design Principles

6.1 Introduction

Chapter 4 analysed the PHB system requirements and presented a solution system in terms of the anticipated system features. This chapter discusses PHB's logical foundations and how the solution system features are incorporated in the PHB design and mapped to PHB components.

Section 6.2 discusses PHB logical foundations and components. Section 6.3 presents PerIS design assumptions. Section 6.4 examines the PDO's logical foundations. Section 6.5 explains the CT design. And finally, Section 6.7 summarises the chapter.

6.2 PHB Logical Foundations

The Patient Health Base (PHB) is an online patient personal health information system that addresses meeting patient Internet search information needs. PHB's design is based on three principles:

a. Enabling a patient access to EPR personal medical information.

b. Integrating EPR with Internet search tools to personalise a patient's Internet search.

c. Adopting a PHR framework (see Section 2.5) in order to utilise an EPR system's existing functionality and expert knowledge, and to facilitate communication between patients and professionals.

The first principle offers a patient online access to a summary medical record that contains essential personal medical information from the patient's integrated medical record (i.e. ISCO), so that a patient can view and revisit when required. Patient access to EPR was recently promoted by the emerging national health information programmes (e.g. CfH [105], IHC [109]). The second principle enables the utilisation of EPRs to focus and customise patient Internet medical search processes according to the patient's condition. This enabled the development of the Patient Personal Internet Search (PerIS) tool. The adoption of the PHR technology can benefit patient Internet medical search in three ways:

1. Typically, PHR provides online patient-personalised services. This enables PerIS as part of the PHR system to utilise other PHR patient services. For instance, a patient's Favorite Websites list could be used by PerIS as a patient-customised search engine.

2. As an official health record system, PHR allows the utilisation of EPR medical knowledge and its associated classification systems.

3. PHR offers a means of communication between patients and professionals regarding health information vocabulary or trusted health websites which can be fed to PerIS.

The PHB system is designed as a PHR framework with patient-personalised services including a summary medical record, that we call SMR, and a patient Personal Internet Search (PerIS) facility. Additional components are needed to support PHR functionality as explored in Section 6.2.1.

6.2.1 PHB Components

The PHB system is designed as a set of patient-personalised services that are supported by staff services and system tools. The patient-personalised services are geared towards and accessed only by individual patients using a secure PHB patient interface. The staff services deliver essential data and functionality required by patient services especially regarding hospital-trusted health websites and health information vocabulary. The system tools coordinate the execution of patient and staff services. Throughout this thesis we refer to both a patient/staff service and a system tool by the term *component*.

Chapter 4 outlined solution system features that correspond to various stakeholders needs (see Appendix B.3). Each PHB system component covers one or more of the solution system features. Figure 6.1 illustrates anticipated PHB components and their mapping to the solution system features.

PHB Component	Maps to Feature (s) (see Appendix B.3)
The Patient Interface	F5
The Staff Interface	F12
The Information Staff Interface	F32
Login Authentication	F3
Summary Medical Record (SMR)	F1, F4
Hospital-trusted Websites (HTW)	F8
Patient Favorite Websites (PFW)	F11
Personal Internet Search (PerIS)	F6, F7, F9, F10, F14, F24 – F28, F30, F31
Patient Diagnosis Ontology (PDO)	F2, F17 – F 23
Concept Thesaurus (CT)	F32
Search Topic Refinements (STR)	F29
Third-Party Accredited Websites	F15, F16
Gateway Links	F32
Staff Trusted Websites (STW)	F13

Figure 6.1: Mapping of PHB Components to Solution System Features

A complete description of PHB functionality and operations is given in Chapter 7. However, PerIS and PDO represent major PHB components and, hence, are further explored in Sections 6.3 and 6.4 respectively.

6.3 PerIS Design Assumptions

The Personal Internet Search (PerIS) system is a patient-personalised Internet search service based on a patient's own EPR data. PerIS's design and operations logic is based on the following assumptions:

- PerIS is patient-oriented, i.e., it accommodates patient information needs at conceptual, logical and data levels.
    - *Conceptual level:* PerIS vocabulary should cater for terminology demanded by and of interest to patients.
    - *Logical level:* PerIS functionality and services should be sought by a patient.
    - *Data level:* PerIS suggested search topics and search results should be relevant to the patient's condition.

- PerIS is patient-customised, i.e., it customises the search features for a patient. For instance, it incorporates patient-customised search ideas that are drawn from a patient's own medical EPR data. In addition, it includes patient-customised search tools such as a customised Hospital-trusted websites search tool and a patient Favorite Websites search tool.

- PerIS assumes patient difficulty in expressing proper medical or lay terminology on their health problems. Hence, it provides a patient with pre-defined search ideas based on EPR data. A patient can select from EPR data search topics or enter a new search term.

- PerIS search term enrichment is applied only to the Diagnosis term category due to the fact that the PDO conceptual model established by this study and utilised by PerIS covers only diagnosis concepts (see Section 6.4)

- PerIS assumes a patient needs health Search Topic Refinements (STR) in order to further narrow the search. In this study, we determined a list of potential health STR information often sought by patients from patient information literature and through interview with patient information staff.

- PerIS assumes a patient needs to access key health gateways and authenticated health websites.

- PerIS assumes a hospital needs to offer a trusted websites list to patients that are recommended by hospital staff members, and to enable a patient search such a list.

- PerIS assumes a patient needs to search a patient's preferable health websites list.

- PerIS assumes a patient needs to search patient-oriented websites (e.g. charity websites).

- PerIS assumes a patient needs to search a single website or specific Web domain (e.g. UK only websites), or the entire Web.

- PerIS assumes a patient needs to conduct a Web search using a preferable search information language.

- PerIS assumes a patient needs to perform a normal keyword search or a semantic search with varying granularity (e.g. medical term search only, lay search term only).

6.4 PDO Logical Foundations

In this study, an ontology technique is used to create the Patient Diagnosis Ontology (PDO) to be utilised by PHB's SMR and PerIS services. The ontology component is geared towards improving a patient's diagnosis vocabulary. Our choice to focus on diagnosis terminology is informed by the fact that terminology challenges are an important factor and become more difficult when it comes to explaining or expressing a diagnosis [249]. In addition, online health information searches are mostly concerned with health problems and specific conditions [145, 229].

PDO represents a patient-oriented diagnosis terminology model that assists patients understand and relate terminologies describing their medical problems as indicated by the diagnoses. PDO differs from existing diagnosis classification systems in that:

- It utilises EPR medical classification system & additional medical terminology used by the patient database,
- It accommodates a patient's information vocabulary requirements and
- It integrates diagnosis concepts from the medical and lay perspectives.

6.4.1 PDO Design

Our approach to building the PDO utilises the EPRs clinical data and its associated medical terminology schemes to form a core model for a patient's personal medical vocabulary. The PDO conceptual design is based on the patient Internet health information terminology requirements determined in Section 2.7.3 namely:

- The need to express the correct form of a particular medical term [232].
- The need to receive health information in medical and lay terminologies [184].
- The need to distinguish between similar (i.e. synonyms) and related (e.g. specific/generic) terms in a search result [232].

- The need to access general (or generic) health information on their health problems [164].
- The need to distinguish between specific and generic health terminology.

Accordingly, the PDO design defines three term forms (categories) of a given diagnosis concept, namely:

- ***Medical Term***: this denotes the medical form of a given patient diagnosis. It assists patients to express the correct medical form of a diagnosis pertinent to their health problems. Medical diagnosis terms correspond to specific medical diagnosis terms.
- ***Lay Term***: this represents a simple English term corresponding to a given medical diagnosis term. It assists a patient in understanding a particular medical diagnosis term. It is also used to identify health information expressed in lay language.
- ***Generic Term***: this specifies a generic form of a given diagnosis. It assists a patient relate a particular diagnosis to a diagnosis category and distinguish between specific and generic diagnosis terms.

The Patient Diagnosis Ontology assists a patient to relate diagnosis term forms, and identify their relevant instances. For example, a patient diagnosed with "malignant neoplasm of stomach" can express such a diagnosis in lay terminology as "cancer of stomach" or "stomach cancer" and relate it to the generic cancer type "gastrointestinal cancer". In addition, a patient can identify similar medical terms for that same diagnosis (e.g. carcinoma of stomach, gastric neoplasm, gastric carcinoma, stomach neoplasm). This rich terminology can benefit a patient in several ways, namely:

- Improve the information vocabulary for patients and enrich their medical knowledge.
- Facilitate term expression when formulating online queries on medical problems in both medical and lay language.
- Assist a patient to distinguish synonyms and specific/generic terminologies in a search result.

In principle, PDO incorporates four distinct diagnosis terminology perceptions (i.e. perspectives):

- ISCO's underlying medical classification system – Read Codes Version 2 (RCV2).

- The conceptual understanding of the ISCO Database Administrator (DBA) in specifying Read Code medical term synonyms (in the ISCO Keyv2 table) and diagnosis hierarchical classes (in the ISCO Classification table).

- A lay vocabulary perspective specified by a member of the patient information staff using a Concept Thesaurus interface (see Section 6.5).

- Patients' perspectives in terms of patient health information vocabulary requirements incorporated in PDO design.

This study offers a new approach to building a patient-oriented diagnosis terminology system that combines medical classifications and expert knowledge in both medical and lay domains. In addition, PDO is accessed by a patient in a customised manner through linkage to his/her own EPR. PDO is used by the PHB system in two functionalities:

1. *The Patient Diagnosis webpage*: this enables a patient to view the diagnosis information in either medical or lay terminology

2. *Patient Personal Internet Search (PerIS) system*: this formulates search information topics describing a patient's particular diagnoses in lay, medical and generic terms. In addition, the search facility offers either a full semantic search, medical term only search, lay term only search or generic term only search.

Thus, PDO conceptual knowledge improves a patient's understanding of medical diagnosis terminologies and facilitates and enriches the online search experience for a patient.

6.4.2 PDO Instances

The utilisation of EPR in this study is key to the personalisation of the health information vocabulary for a patient. In addition, the medical knowledge encoded within EPR facilitates access to medical terminology describing the patient diagnosis and its encoded semantics and associations. EPR medical terminology is more likely to be the terminology used in imparting medical and health information to patients and accordingly the medical terminology that a patient attempts to use when conducting an online medical search. Hence, the extensional design of the PDO is substantially based on ISCO's EPRs underlying medical encoding system (Read Codes Version 2) and additional medical terminologies specified by the ISCO DBA. In addition, a role of patient information staff is incorporated to define mappings between medical and lay terminology (e.g. "malignant neoplasm" versus "cancer") and additional mappings between medical terminologies (e.g. "malignant neoplasm" versus "carcinoma").

6.4.3 138ISCO-based PDO Data

The utilised version of the ISCO system encodes EPR data using Read Code values (e.g. B11.. – see Figure 6.2).

Read_code	Term_30	Term_60	Term_198
B11..	Malig neop of stomach	Malignant neoplasm of stomach	NULL

Figure 6.2: ISCO Corev2 Read Code Data for Diagnosis "stomach cancer"

ISCO records have three types of information on the Read Code terms that are utilised by this study.

- Unique Read Code values and their term descriptions in table Corev2.
- Read Code term medical synonyms in the table Keyv2.
- Read Code term classes and subclasses in the table Classification. For a given Read Code term, the ISCO Classification table stores a class and subclass value in multiple classification encoding systems (e.g. ICD-9). ISCO Classification table's classes can be thought of as an ontology in relational form. These are used for internal ISCO aggregation operations [173]. However, most ISCO Classification table records are described in RCV2 and ICD-9. In addition, ICD-9 concept classes are described in a clearer language when compared with Read Code (RCV2) descriptions of classes (see Figure 6.3).

Coding Scheme	Classification	Class	Subclass	Concept Code
RCV2	MAJORSITE	Cancer: Upper Gastrointestinal	Malig neop of stomach	B11%
ICD-9	ICD DIAGNOSES	Cancer Diagnoses	Cancer of digestive organs	B11%

Figure 6.3: ISCO Classification Table Data for Diagnosis "stomach cancer"

Thus, our approach to designing PDO instances began by utilising diagnosis descriptions in RCV2 and _RCV2 classes representing Read Code version2 and ICD-9 respectively, to formulate basic specific and generic diagnosis terminologies. As shown in Figure 6.4, the RCV2 class value denotes the diagnosis concept generic term "Cancer: Upper Gastrointestinal" while the RCV2 subclass value describes its specific term "malig neop of stomach". A full term "malignant neoplasm of stomach" can be obtained from the Read Code ISCO table Corev2. The _RCV2 (ICD-9) class shows one description for all diagnosis concepts "Cancer Diagnosis" which is not informative and is therefore excluded. However, the _RCV2 (ICD-9) subclass value "cancer of digestive organs" represents a generic term for "stomach cancer" and offers a more understandable diagnosis description that could be useful for a patient to view and utilise in their online search operations.

The existing mapping method between ISCO Classification RCV2 and ICD-9 classes seems inexact with regard to diagnosis concept meaning as illustrated in Figure 6.4. For instance, ICD-9 subclass values show one description "cancer of digestive organs" for different Read Code diagnosis concepts indicated by RCV2 subclasses.

Different ways of utilising Read Code and ICD-9 class values in our PDO design have been investigated. Initially, ICD-9 classes were intended to offer a lay description of medical diagnosis generic terms but this did not apply to some diagnosis concepts as illustrated in Figure 6.4. For instance, both "Cancer: Upper Gastrointestinal " and "Cancer: Colorectal " have the ICD-9 subclass value "Cancer of digestive organs".

Concept Code	RCV2 Class	RCV2 Subclass	ICD-9 Class	ICD-9 Subclass
B11%	Cancer: Upper Gastrointestinal	Malig neop of stomach	Cancer Diagnoses	Cancer of digestive organs
B13%	Cancer: Colorectal	Malig neop of colon	Cancer Diagnoses	Cancer of digestive organs
B17%	Cancer: Upper Gastrointestinal	Malig neop of pancreas	Cancer Diagnoses	Cancer of digestive organs

Figure 6.4: ISCO Classification Table's Mappings between RCV2 and ICD-9

Therefore, the intensional design of the Patient Diagnosis Ontology (PDO) is based on the following assumptions:

a. The ICD-9 class value does not add useful information to a diagnosis concept and therefore it is excluded from the design of the Patient Diagnosis Ontology (PDO).

b. The RCV2 subclass values indicate a specific medical diagnosis term and RCV2 class values indicate a corresponding generic term.

c. The ICD-9 subclass values offer a lay diagnosis generic term.

d. The ISCO Corev2 table provides the full specific diagnosis term.

e. The ISCO Keyv2 table offers synonyms of a specific medical diagnosis term.

f. The need to incorporate a mechanism to specify a lay diagnosis term corresponding to the RCV2 medical diagnosis term. This is to be delivered by a Concept Thesaurus (CT) managed by a member of the patient information staff who can define proper mappings between medical and lay terms.

The ISCO-based PDO instances were discussed with the ISCO DBA [173] to ensure the correct meanings of this data and its relevance to the concept classes covered by PDO. PDO's lay diagnosis instances and additional medical instances are computed based on CT data (see Section 6.6).

Figure 6.5 illustrates the extensional design of the Patient Diagnosis Ontology (PDO) – how PDO class instances are computed. Each diagnosis concept used by the ISCO Classification table will have a Medical Term class, a Lay Term Class and a Generic Term Class. The instances (i.e. data) stored in these classes represents synonym values and are constructed as follows:

- **Diagnosis Medical Term Synonyms:** three ways:-
  - By retrieving Classification table RCV2 subclass full term value using ISCO Classification and Corev2 tables
  - By retrieving RCV2 term, synonyms from the ISCO Keyv2 table prepared by the ISCO DBA.
  - By constructing new medical diagnosis synonyms using CT medical term synonyms mappings.

- **Diagnosis Generic Term Synonyms:** By retrieving Classification table RCV2 class term and ICD-9 subclass term.

- **Diagnosis Lay Term Synonyms:** By constructing new lay diagnosis synonyms using CT lay term synonyms mappings.

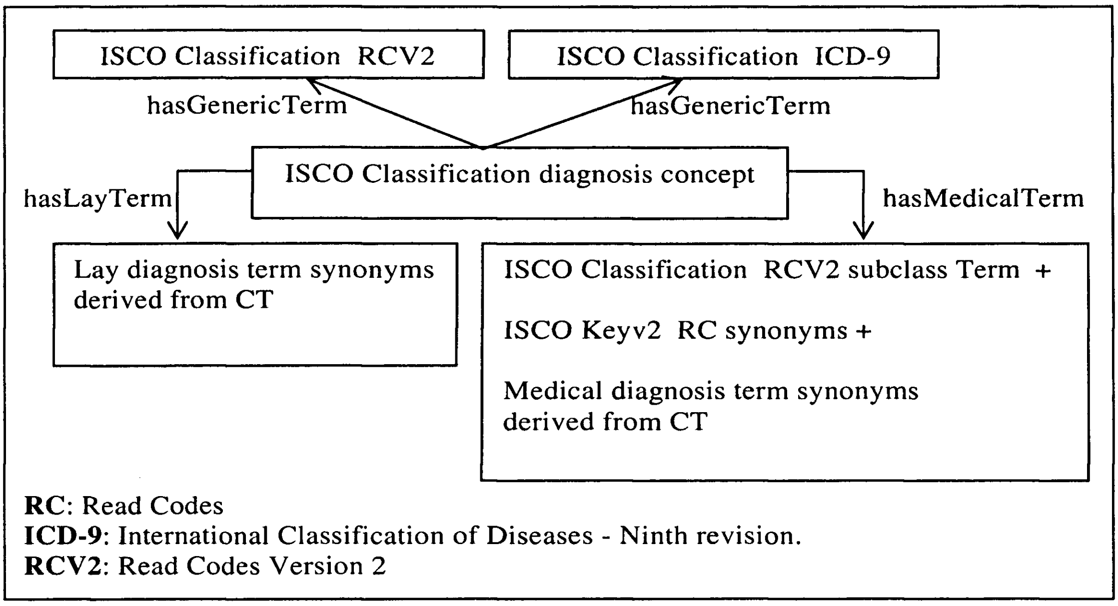


Figure 6.5: PDO Instances Design

6.5 CT Design

The Concept Thesaurus (CT) represents a mechanism through which an information staff member with knowledge in patient information literature and terminology defines medical-to-lay term mappings (e.g. "malignant neoplasm" versus "cancer"). A medical term signifies a scientific term used in the medical domain. There could be different forms of a medical term used by different medical encoding systems or health communities. On the other hand, a lay term denotes a clear simple English description (label) of a medical term that can be clearly understood by patients and laypeople and is commonly used in both official and lay health literature. The purpose of the lay term is to explain the medical term for a patient and aid the discovery of health information written in simple English that can be easily understood by patients. Hence, the lay term has to be defined accurately by a lay information expert rather than by patients themselves as other studies [232] indicate that patient lay terms can retrieve misleading or irrelevant information.

Our choice to define a new CT technique that is managed by a patient information staff rather than using an existing thesaurus is based on the following reasons:

a. Generic thesaurus (e.g. WordNet) may not cover all medical terminology.

b. Lack of a thesaurus or a medical classification system that distinguishes medical and lay term synonyms.

c. The patient information literature uses special patient information terminology that is known to the patient information community. That is, not all medical or lay health information vocabulary is commonly used in patient information literature.

Our CT technique offers a generic method to constructing diagnosis lay terminology based on medical-to-lay term mappings. CT defines three types of term associations:

- Medical-to-lay term mapping (e.g. malignant neoplasm versus cancer).
- Medical term synonyms (e.g. malignant neoplasm, carcinoma).
- Lay term synonyms (e.g. tummy, belly).

CT mappings will be used by PDO to compute lay diagnosis terms. In addition, for a given concept, CT defines similar medical terms denoting medical term synonyms and similar lay terms denoting lay term synonyms. Medical term synonyms are used to derive additional diagnosis medical synonyms not covered in the ISCO database. Lay term synonyms allow the derivation of potential diagnosis lay synonyms. Figure 6.6 shows a CT data sample defined by the author and verified by a medical student [163] to ensure the correctness of these terms and their proper mappings. This is because the author is not very familiar with medical terminology.

Concept	**Medical Term (s)**	**Lay Term (s)**
malignant neoplasm	malignant neoplasm carcinoma	cancer
neoplasm	neoplasm tumor tumour	
renal	renal kidney	
stomach	stomach gastric	belly tummy
uterus	uterus endometrial	womb

Figure 6.6: Concept Thesaurus Data Sample

6.6 Utilising CT Term Mappings in PDO

The PHB system uses CT to generate two diagnosis term types:

- Lay diagnosis terms.
- Additional medical diagnosis term synonyms not defined in the ISCO database.

PHB incorporates two algorithms to generate new lay diagnosis terms and additional medical diagnosis terms based on CT term mappings. Lay diagnosis terms are generated based on the ISCO Read Code diagnosis concept and CT medical to lay term mappings (see Section 6.6.1). Additional medical diagnosis term synonyms are generated based on medical terms extracted from the ISCO DB and CT medical term synonyms (see Section 6.6.2). PDO instances are checked by an information staff member who can delete improper (or insignificant) PDO terms and add proper ones.

6.6.1 Building Diagnosis Lay Terms using CT

Creating lay diagnosis terms is not an easy task as we need to ensure that the whole diagnosis medical description appears in simple English lay terms. The optimal generic approach is to replace every medical term in a diagnosis medical description with a lay term. This requires defining a lay term for every medical term. For instance, the Read Code term "malignant neoplasm of uterus" can be mapped to the lay terms "cancer of womb" or "womb cancer" based on the CT sample in Figure 6.6.

However, there could be many lay terms for a medical term. For instance, the term "stomach" is widely used as a medical term in the medical literature. It appears as the Read Code term "malignant neoplasm of stomach". The term "stomach" could be described in a lay language – informally - using the words "tummy" or "belly". Thus, when constructing a diagnosis lay term, several lay forms need to be considered. However, we need to further investigate which lay terms are more significant or preferable. This can be investigated in a future study. In this study, this task is allocated to an information staff who can define popular and/or significant lay terminology.

Appendix C.1 illustrates the algorithm used for constructing diagnosis lay terms. The current algorithm allows the definition of several lay terms for a medical concept. It constructs diagnosis lay terms by checking if a medical CT term exists in the Read Code diagnosis concept. It, then, creates a new diagnosis lay term for every CT lay term corresponding to and replacing the found CT medical term.

The current algorithm is not an optimal one as it replaces one medical term in a Read Code diagnosis description with a lay term at a time. For instance, the diagnosis "malignant neoplasm of stomach" has two medical terms that have corresponding lay terms in our CT, shown in Figure 6.7.

Medical term	**Lay term**
malignant neoplasm	Cancer
stomach	Belly
stomach	Tummy

Figure 6.7: CT Data on Read Code Term "malignant neoplasm of stomach"

The current algorithm replaces one term at a time, thus, producing the following diagnosis lay synonyms:

- cancer of stomach
- malignant neoplasm of belly
- malignant neoplasm of tummy

However, the optimal result is:

- cancer of stomach
- cancer of belly
- cancer of tummy

Utilising the fact that the medical term "malignant neoplasm" is common to all Read Code diagnosis concepts used by the ISCO Classification table, we can map this term to lay term "cancer" for all existing diagnoses automatically at the code level. This creates a base term for the diagnosis lay term (e.g. cancer of stomach, cancer of uterus, cancer of oropharynx) that can be passed to our algorithm. The second medical term can be replaced using our CT medical to lay mappings. For instance, besides the first (base) lay term "cancer of stomach", the diagnosis "malignant neoplasm of stomach" will yield two lay synonyms from CT mappings: "cancer of belly" and "cancer of tummy".

This approach improves our algorithm to give an optimal result for most diagnosis terms stored in ISCO. Due to the time limitation on this project, an optimal solution to map every medical term in a diagnosis description to a lay term is left as future work discussed in Chapter 9. It is also worth noting that some well-known diagnosis terms contain both medical and lay terms. For example, the term "gastric cancer" is a popular term in medical literature that returned 6,980,000 search results in Google[17].

6.6.2 Building Additional Diagnosis Medical Terms using CT

An additional diagnosis medical term is computed from a Concept Thesaurus (CT) medical term synonyms of a concept by creating new diagnosis terms replacing every medical term found in the diagnosis term with CT's medical terms. This is achieved using the algorithm in Appendix C.2.

6.7 Summary

This chapter discussed PHB design principles and components and how they link to form the solution system features explored in the requirement analysis phase. In addition, it explored the design and logical foundations of two major PHB components, i.e., PerIS and PDO. Chapter 7 fully discusses the PHB architecture, components and operations.

[17] This Google search is performed on 07/04/07 at 20:02.

CHAPTER 7

The PHB Prototype System

7.1 Introduction

In this research study, we have implemented the Patient Health Base (PHB) prototype system to demonstrate the feasibility of personalising and improving patient Internet medical search by integrating data from patient EPR to relevant Internet information sources and providing the functionality needed to address patient Internet medical search challenges.

This chapter explores the PHB prototype system architecture, implementation and operations. The PHB architecture is presented in Section 7.2 showing three major parts; user components (Section 7.3), system components (Section 7.4) and the GUI (Section 7.5). Section 7.6 discusses the PHB software and language implementation choices whereas Section 7.7 summarises the chapter.

7.2 The Architecture

The PHB prototype system represents a PHR prototype system offering patient-oriented services through an integrated staff and patient environment. It comprises patient and hospital staff services. Staff services provide essential features required by patient user services. PerIS is the PHB's central user service aimed at addressing patient Internet search challenges and improving patient Internet medical searching. Other user services are designed to aid PerIS functionality. Our PHB prototype system architecture (Figure 7.1) consists of two component types:

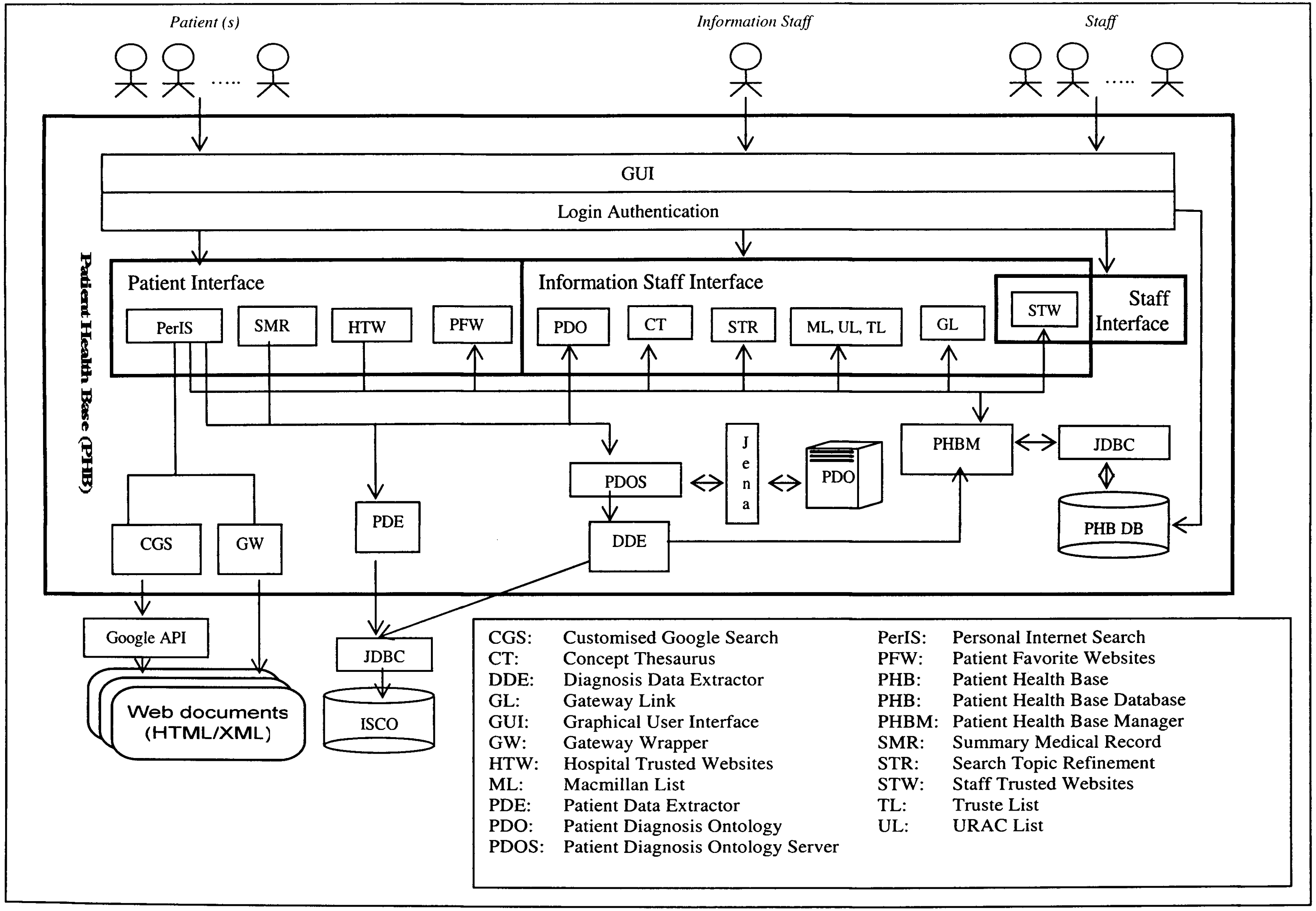


Figure 7.1: PHB Architecture

a. **User Components:** represent the back-end of global users' services and operate in three user interfaces:

1. ***The Patient Interface*:** accessed by individual patients and manages patient-personalised services.

2. ***The Staff Interface:*** accessed by individual hospital staff members involved in patientcare who wish to offer patients a trusted websites list. It administers the construction and update of a Staff Trusted Websites (STW) list.

3. ***The Information Staff Interface:*** geared for the information staff – a specialist in patient information literature – who can aid the identification of key accredited health websites and the construction of the lay health information vocabulary. It manages a collection of PHB services concerning health information vocabulary, accredited websites and system updates.

b. System Components: coordinates interactions between various PHB users, user components and the underlying data sources. Eight system components are incorporated in the PHB functionality:

- ***Patient Data Extractor (PDE):*** coordinates ISCO patient data extraction.
- ***Diagnosis Data Extractor (DDE):*** handles the extraction of diagnosis concepts' data from the ISCO database and the Concept Thesaurus (CT) data from the PHB database.
- ***Patient Diagnosis Ontology Server (PDOS):*** manages PDO construction and manipulation.
- ***Patient Health Base Manager (PHBM):*** administers PHB database operations.
- ***Customised Google Search (CGS):*** customises Google Search based on Google API.
- ***Gateway Wrapper (GW):*** interfaces with individual health gateway search services.

- ***Login Authentication:*** authenticates PHB user's access information.
- ***GUI:*** provides front-end Web interfaces for PHB user's access.

Furthermore, the PHB architecture incorporates two internal data sources:

- ***The Patient Health Base (PHB) database:*** stores data generated and/or required by PHB operations.
- ***The Patient Diagnosis Ontology (PDO) File:*** stores PDO ontology model object for future query and update operations.

7.3 System Components

This section describes the operations defined by different PHB system components. The GUI component, however, is discussed in Section 7.5 due to its extensive operations.

7.3.1 Patient Data Extractor (PDE)

The Patient Data Extractor (PDE) component handles ISCO patient data extraction using JDBC technology [79] and defines the following operations:

- **Establishing connection to ISCO DB:** this requires passing connection parameters such as ISCO DB URL, ISCO User ID and password.
- **Getting patient diagnosis data:** this executes SQL queries that retrieve the full Read Code term of all patient diagnoses concepts recorded in ISCO.
- **Getting patient treatment episodes data:** this queries various ISCO data and codes tables concerning radiotherapy, chemotherapy, surgery and palliative care treatment.
- **Getting patient cancer management plan (CMP) data:** this queries ISCO CMP tables.

7.3.2 Diagnosis Data Extractor (DDE)

The Diagnosis Data Extractor (DDE) component manages ISCO diagnosis concepts' data extraction that is needed for PDO instances and covers three main operations:

- **get_Diagnosis_Classes:** this queries the ISCO classification table for diagnosis concepts and classes. According to Bater [173], the ISCO classification table records all diagnosis concepts describing patient diagnoses. For a given diagnosis concept, this operation establishes two diagnosis data types:
    - Diagnosis medical term based on table classification Read Code subclass value and its full Read Code term from table corev2.
    - Diagnosis Read Code term superclass value. Additionally, we retrieve diagnosis ICD-9 class value, available in ISCO classification table, which represents a more understandable diagnosis superclass term description.
- **get_Diagnosis_Medical_Terms:** this queries ISCO table Keyv2 to retrieve term synonyms recorded for every diagnosis concept in table Classification. In addition, it queries the PHB database tables concept and medical data, recorded using the CT interface, to establish additional diagnosis medical terms based on the algorithm shown in Appendix C.2.
- **create_Diagnosis_Lay_Terms:** this queries the PHB database Concept and English tables to construct diagnosis lay term descriptions for every diagnosis medical term using the algorithm given in Appendix C.1.

7.3.3 Patient Diagnosis Ontology Server (PDOS)

The Patient Diagnosis Ontology Server (PDOS) establishes and manages the Patient Diagnosis Ontology (PDO) as an RDF model using Jena[18] Ontology [80]. Jena [81] is selected for its ability to construct ontologies dynamically. In this research, we implement PDO as a simple diagnosis ontology data model that stores diagnosis

[18] Jena [81] is a JAVA RDF API that enables the construction and query of an RDF model from within JAVA programs.

data from ISCO diagnosis and CT data. The PDOS component covers five major operation types:

- **create_PDO:** this consists of the following steps:
  - Create a Jena Ontology model object using the package ModelFactoty. createOntologyModel().
  - Populate PDO with RDF resources denoting different diagnosis terms retrieved by the DDE component and linked using the following relationships:
    - **RDFS.subClassOf:** denotes an is_A relationship and links RCV2 and ICD_9 diagnosis superclass terms retrieved from the ISCO classification table to its main diagnosis Read Code term.
    - **RDFS.label:** this is used to represent a has_Lay_Term_Synonym relationship. It links a Read Code diagnosis term to all its lay term synonyms.
    - **RDFS.seeAlso:** this is used to represent has_Medical_Term_Syonym and links a Read Code diagnosis term to all its medical term synonyms.

  These relationships are implemented using the Jena addProperty Resource method (e.g. diag.addProperty(RDFS.subClassOf, m.createRsource("Urological Cancer")). PDO is saved to a file as a Jena OntModel object in an RDF/XML format and manipulated by PHB operations by reading it into a Jena OntModel object and querying it using different PDOS methods.

- **Refresh PDO:** this is used to apply changes in CT data to PDO instances. This is currently implemented by recreating PDO based on the new ISCO and CT diagnosis data.

- **Query PDO:** retrieves a lay, medical or generic PDO diagnosis term's synonyms.

- **Add To PDO:** handles operations concerning adding lay, medical or generic PDO diagnosis term synonyms.

- **Delete From PDO:** manages deleting lay, medical or generic term synonyms of a given PDO diagnosis concept.

PDO data is represented as Jena Statements. Adding to PDO is implemented using Jena addProprty method whereas query operations are implemented using getProperty method. Deleting from PDO is implemented by removing the relevant RDF statement.

7.3.4 Patient Health Base Manager (PHBM)

The Patient Health Base Manager (PHBM) defines operations that manage interactions with the PHB database tables. It covers four database operation types; get_items, add_items, update_items and delete_items. These are implemented using corresponding SQL constructs through a JDBC interface. Figure 7.2 describes major PHBM operation types:

PHBM Operation type	**Description**	**Affected tables**
get_items	Selects given record(s)	Almost all PHB tables.
add_items	Inserts new record(s)	Trusted, Favorites, URAC, Macmillan, Truste, Refinements, Charity, Concept, Scientific, and English.
delete_items	deletes given record(s)	Same tables described in the “add_items” operation.
update_items	updates values in given record (s)	Gateway, for updating a gateway search link (see Section 7.3.6).

Figure 7.2: Major PHBM Operation Types

7.3.5 Customised Google Search (CGS)

CGS is an internal PHB search engine. It is geared to execute PerIS internal search tools, introduced in this study. It operates by restricting Google Search to a selected websites list defined by individual search tools. This is achieved by connecting to the Google search engine using Google API [42] methods and running the search query with the Google Site Query Modifier[19] [41] set for each website item in the list. Partial search results are combined into a single distinct CGS search results list.

[19] Restricts the Google search to specific website (s) (e.g. cancer site:http://www.healthline.com).

CGS implements the Hospital Websites Search (HWS), the Favorite Websites Search (FWS), the Charity Websites Search (CWS) and the Specific Website Search (SWS), discussed in Section 7.4.3.4.4 by restricting the search only to the websites contained in the respective lists. CGS search results size is adjusted according to the CGS website list size, as specified in Figure 7.3.

```
if(URLs.size()<5){
    ResultSize=10;
   }else if(URLs.size()<10){
    ResultSize=5;
   }else if(URLs.size()<20){
    ResultSize=4;
   }else{
    ResultSize=3;
   }
```

Figure 7.3: Adjusting CGS Search Results Size Algorithm

In addition, CGS executes an unrestricted Google search for the PHB Google search tool with two options:

- **UK only search:** based on Google API setRestrict("countryUK") method.
- **Search language:** based on Google API setLanguageRestrict(langcode) method.

As CGS is part of the PHB system implementation, its search results are manipulated by the PHB system to allow patients to add search results to their Favorites List.

7.3.6 Gateway Wrapper (GW)

The Gateway Wrapper (GW) component searches key health gateways from within the PerIS interface. As we lack direct access to the underlying gateway databases and search engines, our search approach is based on the gateway's search result's URL. GW retrieves gateway URLs from the PHB Gateway table as previously saved by an information staff member (see Appendix D.4). For a given patient query, a search result URL, for the selected gateway with the current patient search query, is constructed and opened in a new browser window. For instance, searching for "brain cancer" on the NHS Direct Online involves the following steps:

1. Retrieving NHSDirectOnline search URL value from the PHB database. This consists of two parts:

   Search URL Part1:

   http://search.nhsdirect.nhs.uk/kbroker/nhsdirect/nhsdirect/search.lsim?qt=

   Search URL Part2:

   &hs=0&sm=0&ha=1054&sc=nhsdirect&mt=0&sb=0&nh=3

2. Constructing a Search Results URL by concatenating the patient search phrase (e.g. brain cancer" with the retrieved search URL values according to the format used by NHSDirectOnline. This step yields the following URL value:

   http://search.nhsdirect.nhs.uk/kbroker/nhsdirect/nhsdirect/search.lsim?qt=brain+cancer&hs=0&sm=0&ha=1054&sc=nhsdirect&mt=0&sb=0&nh=3

3. Opening a new browser window for the constructed URL value. This displays the NHS Direct Online search results for the query "brain cancer".

7.3.7 Login Authentication

The PHB system offers a secure user access using hospital personal identity numbers. Patients' ISCO Ids are used as usernames and currently for our prototype experimentation all patients have one password "test". However, the ISCO version utilised in this study does not contain hospital staff data. Hence, we created a sample staff data list, stored in the PHB database.

7.4 User Components

User components service requests from user GUI webpages. It defines operations that execute a user request by invoking appropriate system component methods. As shown in Figure 7.1, user components are defined for three types of user interfaces; the Patient interface, the Staff interface and the Information staff interface.

7.4.1 The Information Staff Interface

The Information staff interface is a crucial user interface. It establishes and updates data required by essential PHB components such as CT, STR, GL and several accredited websites lists. It executes the following operations:

- **update_CT:** manages CT update operations for adding, deleting and editing new CT concepts data. CT data is stored in three relational PHB tables Concept, Scientific and English denoting CT unique concepts, a concept's medical term synonyms and a concept's lay term synonyms respectively. It defines three operations:
  - **add_CT_concept:** inserts a new distinct concept term into the PHB Concept table.
  - **Edit_CT_concept:** retrieves a given CT concept's medical and lay terms from the PHB Scientific and English tables respectively. Changes to these data are saved to the respective tables.
  - **delete_CT_concept:** deletes a CT concept from the PHB Concept table and its medical and lay terms from the Scientific and English tables.
- **verify_PDO:** manages PDO data verification operations. It reads the PDO file into a Jena OntModel object and invokes PDOS methods. It covers the following operations:
  - **get_PDO:** retrieves the entire PDO ontology model object.
  - **delete_synonym:** deletes a medical, lay or generic diagnosis term synonym of a given PDO diagnosis concept using relevant PDOS method. This operation is used to delete improperly constructed diagnosis term synonyms from CT data.
  - **add_synonym:** adds a new medical, lay or generic term synonym to a PDO diagnosis concept using relevant PDOS operations.
- **update_SR, update_ML, update_TL, update_UL:** manage update operations to the Search Refinements (SR), Macmillan List (ML), Truste List (TL) and URAC List (UL). Each update operation retrieves respective lists from the PHB database and handles add and delete operations.

7.4.2 The Staff Interface

The staff interface includes one user component, i.e. STW, that manages the Staff Trusted Websites (STW) list. STW covers four operations:

- **get_STW:** retrieves STW list from the PHB database so it can be displayed and modifed in the STW Webpage.
- **delete_from_STW:** removes one or more items from STW list.
- **add_to_STW:** adds one or more items to the STW list. PHB facilitates STW construction by incorporating lists of authenticated health websites, managed by an information staff member. Figure 7.4 describes six methods by which a staff member can add an STW item.
- **label_STW_Item:** ascribes a health condition category label to a staff Trusted website item. The trusted website category label is eventually used for customising Hospital-trusted Website (HTW) list to individual patients according to the patient's health condition.

STW Add item Method	**Reason**
Add_from_HTW	It allows for a health website item to be trusted and, hence, recommended by more than one staff member which increases the significance of a website.
Add_from_PFW	It indicates patients' selected websites to staff members so they can explore, label and recommend them.
Add_from_Mamillan	It indicates Macmillan selected key health websites so they can be recommended by Staff members.
Add_from_Truste	It indicates Truste accredited health websites to staff members.
Add_from_URAC	It indicates URAC accredited health websites to staff members.
Add_own_items	Allows adding health websites identified by individual Staff members.

Figure 7.4: STW Add Items Methods

7.4.3 The Patient Interface

The Patient interface is the target user interface. It defines four patient user components that handle patient online services including Summary Medical Record (SMR), Hospital-trusted Websites (HTW), Patient Favorite Websites (PFW) and Personal Internet Search (PerIS). Many Patient interface operations are based on data generated by the Staff and Information staff interfaces.

7.4.3.1 Summary Medical Record (SMR)

SMR user component retrieves patient diagnoses, treatment and cancer management plan data from ISCO EPRs by invoking PDE methods get_diagnosis, get_treatment and get_cancer_management_plan respectively. SMR data is presented to a patient in separate GUI webpages including Diagnosis, Treatment and Cancer Management Plan Webpages (see Section 7.5.3.1.1).

7.4.3.2 Hospital Trusted Websites (HTW)

The HTW user component builds a patient-customised list of Hospital-trusted websites created by hospital staff members. It extracts all HTW values from the PHB database table Trusted. It then uses the HTW category value to customise the list according to the patient diagnosis information. A patient can access HTW list from "Your Velindre Trusted Websites" (see Section 7.5.3.2).

7.4.3.3 Patient Favorite Websites (PFW)

PFW manages PFW list's access and update operations through the PHB Favorites table. Three operations are defined on PFW items:

- **add_to_favorites:** Four methods facilitate this operation:
    1. *Using HTW websites list:* offers a patient-customised list of hospital-trusted websites.
    2. *Using Macmillan websites list:* offers Macmillan key health websites.
    3. *Selecting websites from PerIS search results:* adds from PerIS's search results.
    4. *Entering new websites in a textarea:* allows a patient to enter new websites.
- **delete_from_favorites:** deletes one or more Favorite items from the PHB Favorites table.
- **retrieve-favorites:** retrieves all PFW items from the PHB Favorites table for display in the patient GUI.

7.4.3.4 Personal Internet Search (PerIS)

PerIS is the central user component addressed in this study. It is a major user component and manipulates many PHB operations. PerIS internal architecture is shown in Figure 7.5 and consists of five subcomponents:

1. Personalised Search Topic Constructor (PSTC): establishes patient personalised search ideas.

2. Search Topic Refiner (STR): focuses the main search by enriching it with a patient selected health search refinement information.

3. Diagnosis Term Enricher (DTE): retrieves a given diagnosis's related terms from PDO

4. Search Tool Manager (STM): coordinates the execution of a wide range of Web search tools.

5. Search Mode Controller (SMC): controls the activation and execution of PerIS search modes.

7.4.3.4.1 Personalised Search Topic Constructor (PSTC)

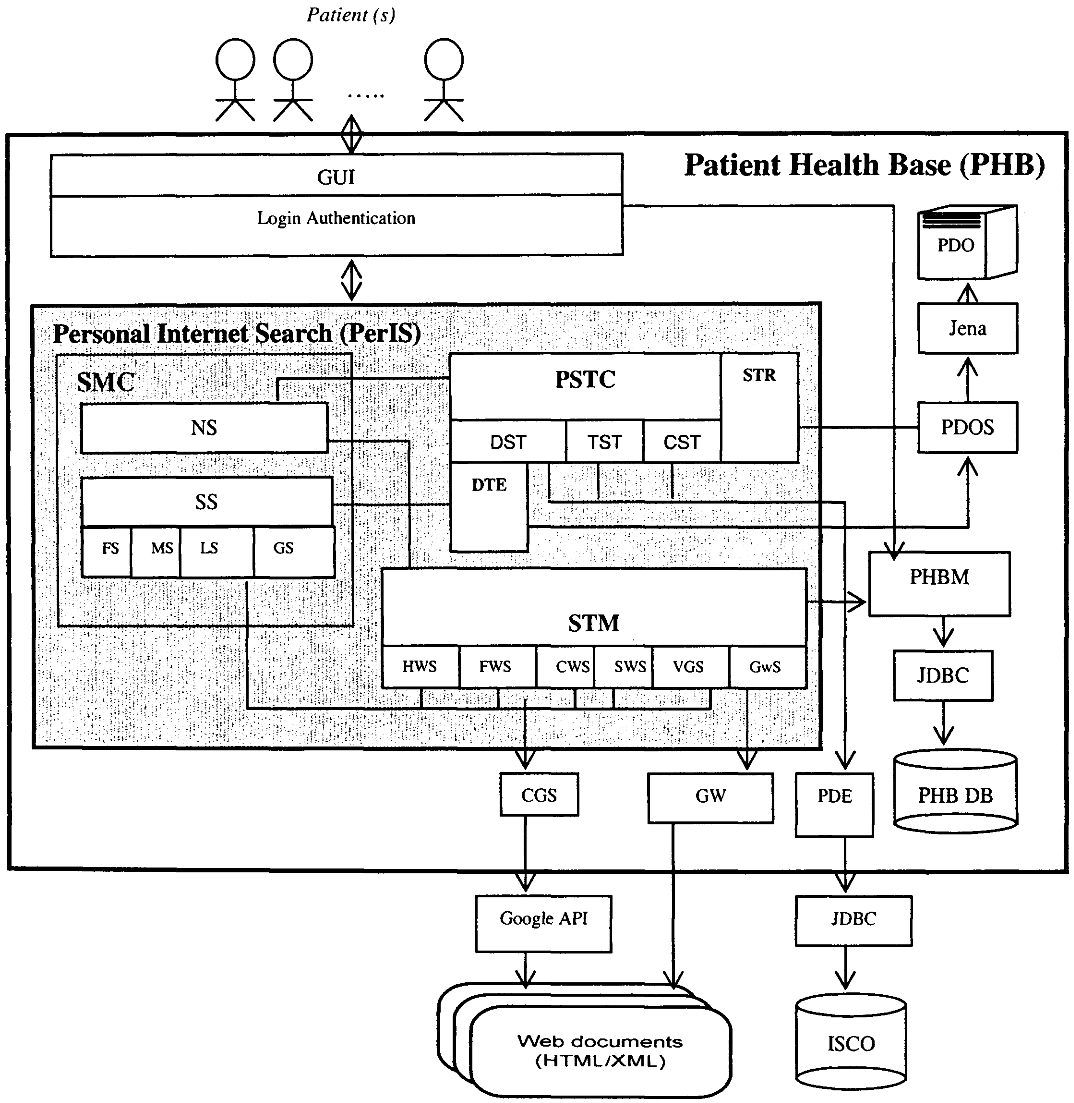
PSTC constructs patient-personalised search topics from ISCO patient diagnosis, treatment and cancer management plan data, extracted using PDE system component. Figure 7.6 describes techniques used in establishing potential combinations of different personalised search topics.

7.4.3.4.2 Search Topic Refiner (STR)

STR retrieves the STR list from the PHB database Refinements table and connects it to the PerIS tool so it can be browsed by a patient.

7.4.3.4.3 Diagnosis Term Enricher (DTE)

DTR expands a given diagnosis search information with related terms from PDO. It extracts PDO medical, lay and generic terms for the selected diagnosis using get_Lay_Terms, get_Medical_Terms and get_Generic_Terms PDOS operations and makes them available for a patient to browse and/or to employ in a PerIS semantic search.



CGS:	Customised Google Search
CST:	Cancer Management Plan Search Topics
CWS:	Charity Websites Search
DST:	Diagnosis Search Topics
DTE:	Diagnosis Topic Enricher
FS:	Full Semantic Search
FWS:	Favorite Websites Search
GS:	Generic Semantic Search
GUI:	Graphical User Interface
GwS:	Gateway Search
HWS:	Hospital Websites Search
LS:	Lay Semantic Search
MS:	Medical Semantic Search
NS:	Normal Search
PDE:	Patient Data Extractor
PDO:	Patient Diagnosis Ontology
PDOS:	Patient Diagnosis Ontology Server
PerIS:	Personal Internet Search
PHB:	Patient Health Base
PHBM:	Patient Health Base Manager
PSTC:	Personalised Search Topic Constructor
SMC:	Search Mode Controller
SS:	Semantic Search
STR:	Search Topic Refiner
SWS:	Specific Websites Search
TST:	Treatment Search Topics
VGS:	Velindre Google Search

Figure 7.5: PerIS Internal Architecture and Components

Personalised Search Topics	Search Topics Techniques
Diagnosis Search Topics (DST)	Diagnosis lay term Related diagnosis terms obtained by DTE
Treatment Search Topics (TST): ▪ Chemotherapy TST	 treatment type drug name treatment type + drug name
▪ Radiotherapy TST	treatment intent machine name treatment intent + "radiotherapy" treatment intent + "radiotherapy" + treatment site treatment intent + "radiotherapy" + machine name treatment site + machine name "radiotherapy" + treatment site + machine name
▪ Surgery TST	treatment intent treatment intent + "surgery"
▪ Palliative Care TST	Care aim
Cancer Management Plan Search Topics (CST)	Plan intent +modality

Figure 7.6: Potential Patient Search Topic Combinations

7.4.3.4.4 Search Tool Manager (STM)

STM establishes several Web search mechanisms within PerIS that can focus a patient's search information requirement. PerIS search tools offer a rich guided and patient-centred approach to Web search that allows a patient access to key health gateways and additional patient-customised search engines. Thus, a patient can utilise various search tools from a single access point. STM manipulates three search mechanism types:

i Gateway Search (GwS)

ii Customised Google Search (CGS)

iii Open Google Search

- **Gateway Search (GwS):** executes external search tools covering a wide range of medical and key health gateways, incorporated within the *Key Health Gateways* PerIS search tool, and are executed using the Gateway Wrapper system component (see Section 7.3.6) Currently, PerIS covers four types of key health gateways; namely:

  ▪ ***Accredited-Information Search Engines***: search for health information accredited by a third party (e.g. HON). PerIS links to the HONCode search engine [51] enabling the search of HON accredited health websites.

- *Medical Search Engines*: search a collection of medical databases often evaluated by the providing medical community. PerIS links to the MedHunt search engine [88].

- *National Health Gateways:* health networks that enable the search of a wide collection of medical and health information. Currently, PerIS integrates with:

    - *NHS Direct Online [108]:* a UK NHS online information service offering high quality advice and information, and details of NHS services.
    - *MedlinePlus [91]:* A US online service providing easy access to medical journal articles and extensive information on drugs, a medical encyclopaedia, and latest health news.

- *Evidence-based Search Engines:* offer reliable evidence-based information. PerIS links to the Cochrane library [19] for evidence-based information on healthcare.

- **Customised Google Search (CGS):** implements and executes an internal search engine based on Google API. CGS executes by passing a given websites list to the CGS system component (see Section 7.3.5). PerIS incorporates four internal CGS-based search tools:

    - *Hospital Trusted Websites Search (HWS):* operates CGS on the patient-customised HTW list as extracted from the PHB Trusted table.
    - *Favorite Websites Search (FWS):* operates CGS on the PFW list as retrieved from the PHB Favorites table.
    - *Charity Websites Search (CWS):* operates CGS on the CW list retrieved from the PHB Charity table.
    - *Specific Website Search (SWS):* operates CGS on one website item selected from a list of key health websites. This enables the search to be restricted to one specific website. This is useful if a patient wishes only to search for information from a single website (e.g. Cancerbackup.org.uk). This tool combines the Velindre List with Macmillan Cancer Support [85].

- **Open Google Search:** executes an open Google search. The aim is to offer unrestricted Web search mechanism for a patient. Two open Google search tools are incorporated in PerIS:
  - *Web Google Search (WGS):* links to the Web Google search using its search URL via the Gateway Wrapper component which directs a patient to the Google website in a new browser window. This benefits a patient if a patient wants to use Google Website features.
  - *Velindre Google Search (VGS):* executes open Google search using Google API via CGS. This option allows a patient to incorporate open Google search results in the Favorite Websites list defined in PHB.

7.4.3.4.5 Search Mode Controller (SMC)

SMC controls the activation and execution of PerIS search modes. PerIS executes two search modes:

- **Normal Search (NS):** performs a normal keyword search. NS is permitted for all search topic categories and search tools incorporated in PerIS.
- **Semantic Search (SS):** extends normal search results by running the search for various related search terms. SS currently affects only diagnosis search topics offered by PerIS as it manipulates PDO data that is only defined for diagnosis concepts. In addition, it executes using the PerIS internal search tools operated by our CGS component. This is due to the possibility of manipulating the underlying search mechanism provided by CGS. The SS mode is not currently executing on external search tools such as health gateways as this requires running a gateway search for different related search terms and combining search results, which is not feasible as we lack access to the gateway search mechanism. Partial SS search results are combined in a single distinct search results list. In addition, SS handles fine-grained semantic search options including:
  - **Full Semantic Search (FS):** executes the semantic search on all related terms.
  - **Medical Term Search (MS):** executes the semantic search only using the diagnosis medical term synonyms.

- **Lay Term Search (LS):** executes the semantic search only with diagnosis lay term synonyms.
- **Generic Term Search (LS):** executes the semantic search only with diagnosis generic/broad term synonyms.

7.5 Graphical User Interface (GUI)

The Graphical User Interface (GUI) component represents the PHB front-end that facilitates users' access to the PHB operations through a network connection. It consists of several webpages describing different PHB user components and operations. The PHB system defines three user GUIs the patient GUI, the staff GUI and the information staff GUI, corresponding to the three PHB user interfaces; discussed in Section 7.4.

Figure 7.7 illustrates the main GUI Webpage with the login operation. PHB users hold individual accounts and are authenticated based on their given usernames and passwords as discussed in Section 7.3.7. Only users with valid login information can access PHB services. Upon successful login, each user group is directed to its respective GUI.

Sections 7.5.1 and 7.5.2 describe the Staff and information Staff GUIs respectively. These establish important features required for the patient GUI, discussed in Section 7.5.3.

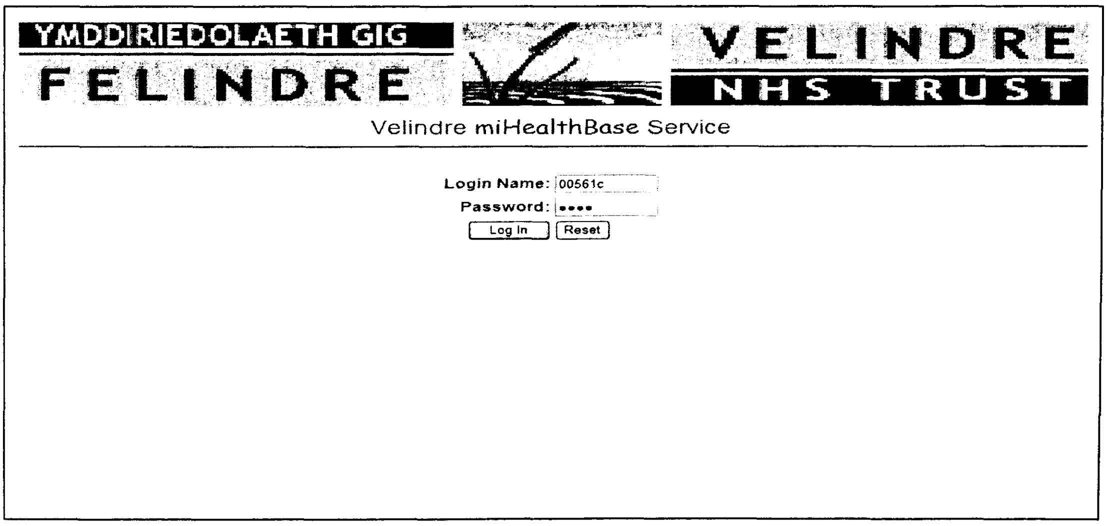


Figure 7.7: PHB Login Webpage

7.5.1 The Staff GUI

The staff GUI is key to the PHB system functionality. It is geared to support and update the patient GUI operations. Two staff GUIs are defined in our PHB system:

1) A common staff GUI which is used for building a trusted websites list for a patient by interested hospital staff members.

2) An information staff GUI that supports additional tasks relating to the management of health information vocabulary tools and various system updates (see Section 7.5.2).

The common staff GUI, shown in Figure 7.8, offers the means for establishing and maintaining individual Staff Trusted Website (STW) lists. Individual STW lists are combined in a single HTW list being customised to individual patients according to a patient's health problems (see Section 7.4.3.2).

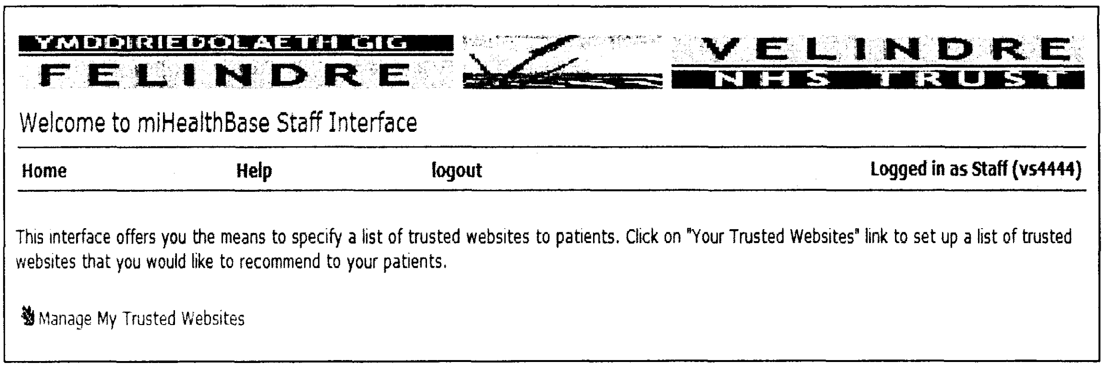


Figure 7.8: The Common Staff GUI Main Webpage

An example STW list is shown in Figure 7.9. STW data is retrieved from the PHB relational table Trusted. The Staff GUI covers three basic STW operations:

i Adding STW Items using the *Add to Trusted Websites* link. The PHB system aids this process by offering a staff member lists of accredited health websites to select from (see Appendix D.1). Newly added STW items are assigned "general" category which can be changed to a more appropriate category by editing the category value in the main STW webpage (see Figure 7.9).

ii Deleting STW Items using the *Delete from Trusted Websites* link (see Appendix D.2).

iii Categorising an STW item by specifying a website category in the *Modify Category* section. Website category is used in customising the HTW list to individual patients by matching a website category with patient diagnoses information.

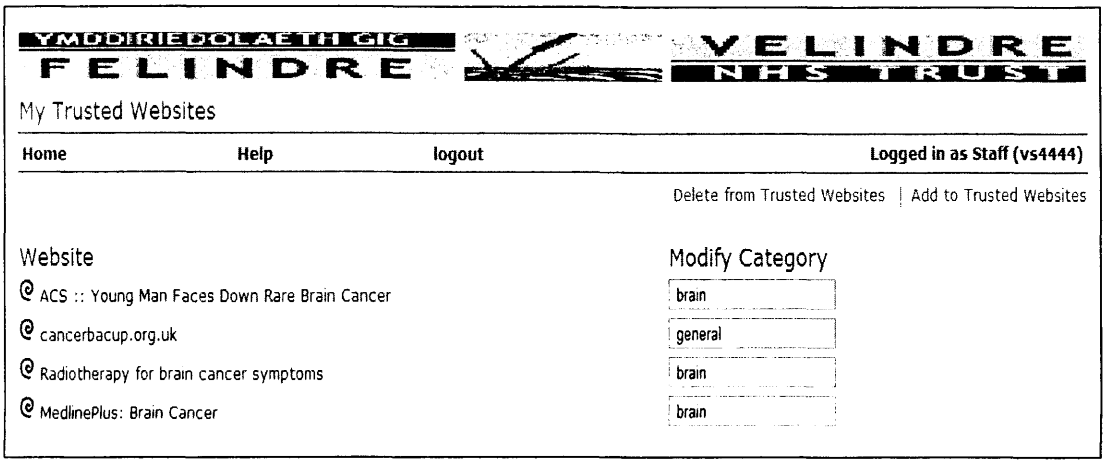


Figure 7.9: Staff "VS4444" Trusted Websites Webpage

7.5.2 The Information Staff GUI

The Information Staff GUI, shown in Figure 7.10, is a special staff interface. It is used for managing PHB updates and vocabulary-related functions. The information staff role can be assigned to a patient information specialist, information librarian or informationist as discussed in Section 2.3.1.

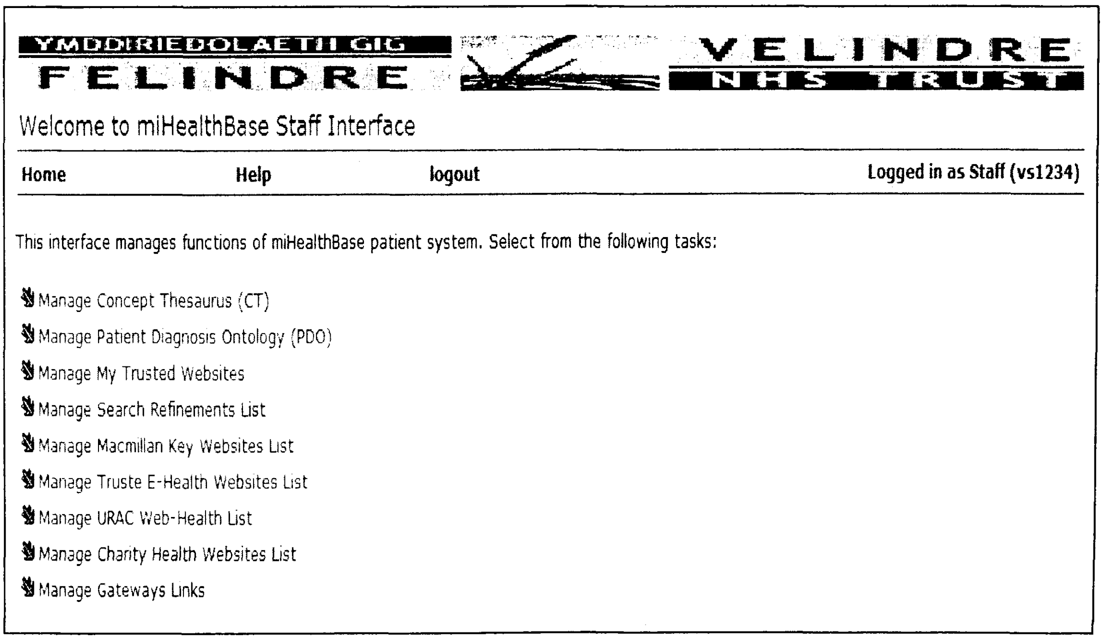


Figure 7.10: Information Staff GUI Main Webpage

The information staff GUI extends the common staff GUI with the following operations:

- Managing the Concept Thesaurus (CT) (see Section 7.5.2.1).
- Managing the Patient Diagnosis Ontology (PDO) (see Section 7.5.2.2).
- Managing Search Refinements List (see Appendix D.3).
- Managing Gateways Links (see Appendix D.4).
- Managing the Macmillan, Truste E-Health, URAC Web-Health and Charity Websites lists (see Appendix D.5).

7.5.2.1 Managing Concept Thesaurus (CT)

The Concept Thesaurus (CT) represents a mechanism through which an information staff member, a patient information specialist, can define medical-to-lay term mappings (e.g. "malignant neoplasm" versus "cancer). This study defines three types of associations among terms in the Concept Thesaurus; namely:

- Medical-to-lay term mapping (e.g. malignant neoplasm versus cancer).
- Medical term synonyms (e.g. malignant neoplasm, carcinoma).
- Lay term synonyms (e.g. tummy, belly).

The Concept Thesaurus (CT) is used in generating lay and additional medical diagnosis terms for the Patient Diagnosis Ontology (PDO). Figure 7.11 illustrates the Concept Thesaurus (CT) webpage in which information staff manage thesaurus concepts and their medical and lay labels. CT is managed by three main operations namely:

a. Create a new thesaurus concept (see Appendix D.6).
b. Edit an existing thesaurus concept (see Appendix D.6).
c. Delete a thesaurus concept by ticking the concepts to be deleted and clicking the *Delete Thesaurus Concept* button (see Figure 7.11).

Changes in CT data can be applied to the Patient Diagnosis Ontology (PDO) using the *Refresh* Diagnosis Ontology button shown in Figure 7.12.

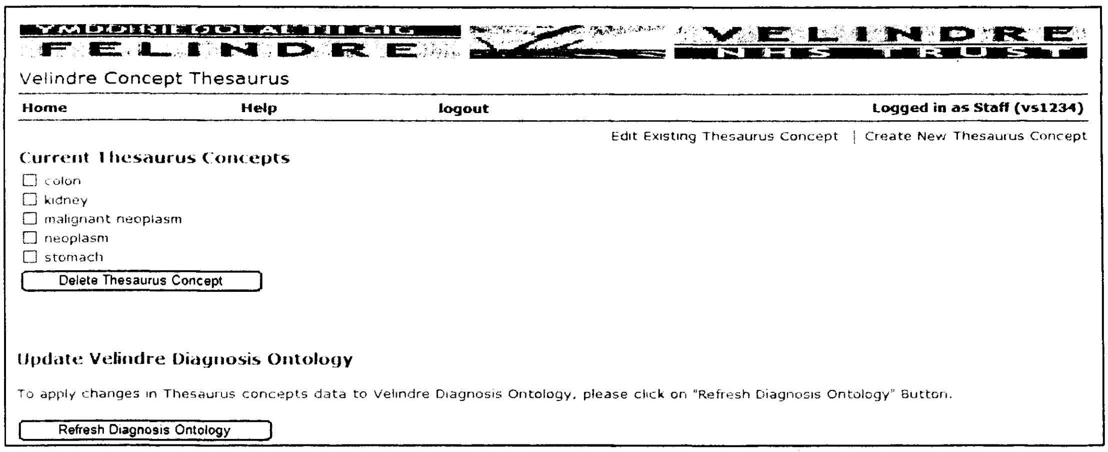


Figure 7.11: Managing Concept Thesaurus (CT) Data in Information Staff GUI

7.5.2.2 Managing Patient Diagnosis Ontology (PDO)

The Patient Diagnosis Ontology (PDO) encodes diagnosis concepts' related terms. It covers lay and medical diagnosis terminology, synonyms and specific/generic classes. PDO data is constructed automatically from the ISCO database diagnosis classification data and the Concept Thesaurus (CT) term mappings using the PDOS's "create_PDO" method. However, as PDO is eventually manipulated by the patient GUI services, there's a need to ensure that it contains valid and meaningful diagnosis term descriptions. Two main operations are managed within the information staff GUI:

- Verifying PDO diagnosis terms' meaningfulness whereby a malformed diagnosis-related term can be deleted and its proper form can be added (see Appendix D.7).
- Uploading PDO so that its data can be accessed by the information staff GUI and the patient GUI (see Appendix D.8)

7.5.3 The Patient GUI

The patient GUI is the central user GUI, accessible by ISCO-registered patient users. Figure 7.12 shows the main patient GUI webpage, listing four patient services:

- **Summary Medical Record (SMR)** – displays a patient's personal medical information stored in ISCO EPRs. This is to allow a patient to revisit essential personal medical information, as required.

- **Personal Internet Search (PerIS)** – provides a patient-personalised Internet search mechanism that aids a patient in focusing the search information requirement.
- **Favorite Websites (FW)** – manages a patient's preferred websites list.
- **Hospital Trusted Websites (HTW)** – accesses a patient-customised hospital-trusted health websites list.

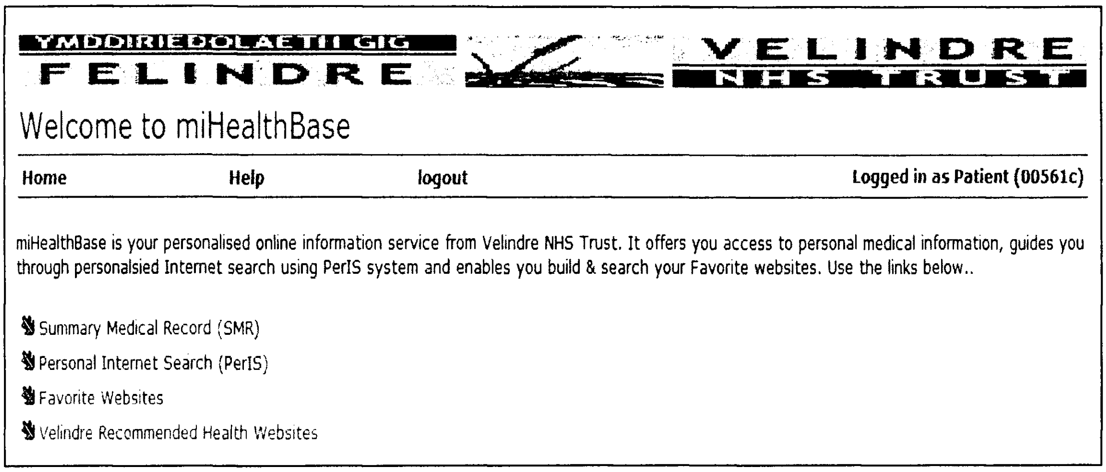


Figure 7.12: The Patient GUI Interface Homepage

7.5.3.1 Summary Medical Record (SMR) Webpage

The Summary Medical Record (SMR) webpage (Figure 7.13) offers access to essential patient personal medical information. As a tentative study, we focused on extracting medical information that is mostly sought by patients. Surveys [145, 229] report that patients usually seek information on their health problems and treatments. Hence, SMR offers access to three types of patient personal medical information, provided in separate Webpages (see Appendix D.9 for examples) covering:

- **Diagnoses:** presents a patient diagnosis in either medical or lay terms.
- **Treatment Episodes:** presents essential patient treatment details.
- **Cancer Management Plan:** presents information about a proposed patient treatment.

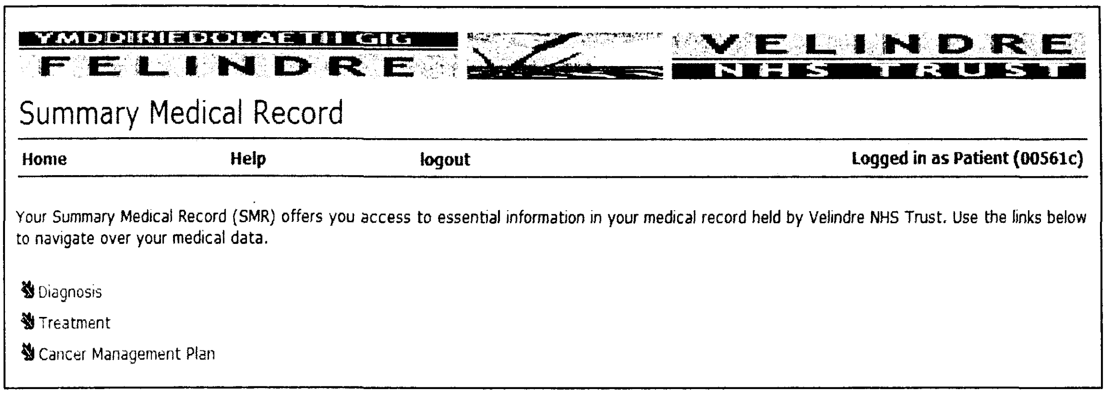


Figure 7.13: Patient "00561c" SMR Webpage

7.5.3.2 Hospital Trusted Websites (HTW) Webpage

The Hospital Trusted Websites Webpage presents a health websites list that is trusted by hospital staff members and customised to a patient's health condition. Figure 7.14 shows patient "00561c" HTW webpage. HTW items can be accessed by clicking on a website item link which opens the website in a new browser window.

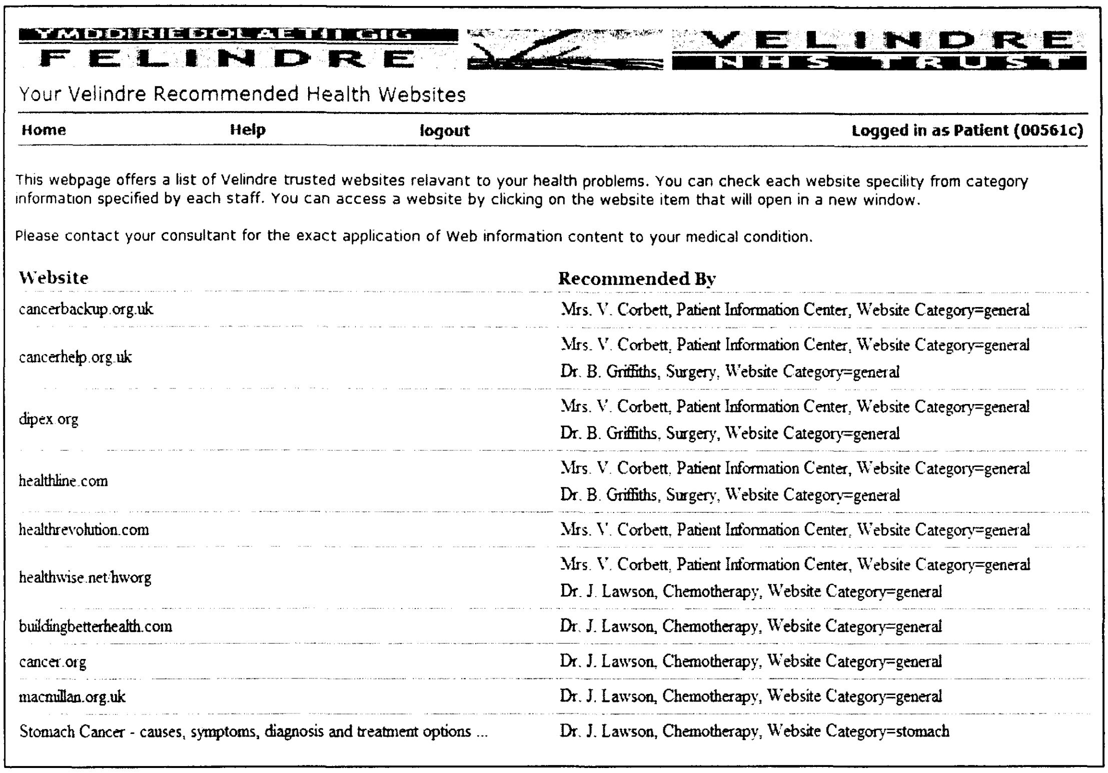
7.5.3.3 Patient Favorite Websites (PFW) Webpage

The Patient Favorite Websites (PFW) webpage (see Figure 7.15) allows a patient to maintain a Favorite websites list relating to his/her health condition. The PFW list is utilised by the PHB system in three ways:

- Accessing PFW items in the PFW webpage by clicking on the website item.
- Searching all PFW list using PerIS "Your Favorites" search tool.
- Adding from a combined PFW list when constructing a STW list (see Appendix D.1).

Two main features are supported by the PFW webpage:

- Viewing and accessing the PFW items by clicking on a website link.
- Adding/deleting PFW items (see Appendix D.10).



YMDDIRIEDOLAETH GIG FELINDRE · VELINDRE NHS TRUST

Your Velindre Recommended Health Websites

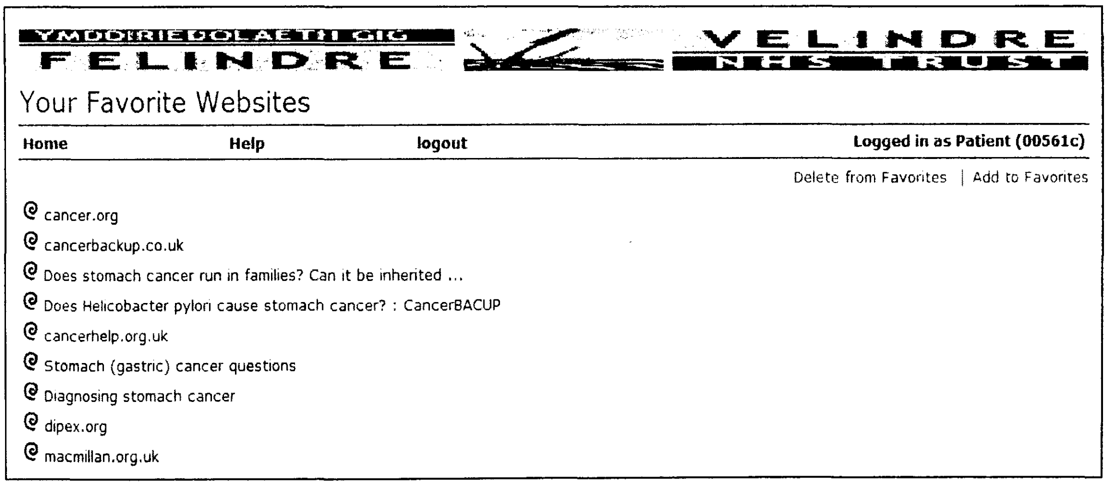
Home | Help | logout | Logged in as Patient (00561c)

This webpage offers a list of Velindre trusted websites relavant to your health problems. You can check each website specility from category information specified by each staff. You can access a website by clicking on the website item that will open in a new window.

Please contact your consultant for the exact application of Web information content to your medical condition.

Website	Recommended By
cancerbackup.org.uk	Mrs. V. Corbett, Patient Information Center, Website Category=general
cancerhelp.org.uk	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
dipex.org	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
healthline.com	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
healthrevolution.com	Mrs. V. Corbett, Patient Information Center, Website Category=general
healthwise.net/hworg	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. J. Lawson, Chemotherapy, Website Category=general
buildingbetterhealth.com	Dr. J. Lawson, Chemotherapy, Website Category=general
cancer.org	Dr. J. Lawson, Chemotherapy, Website Category=general
macmillan.org.uk	Dr. J. Lawson, Chemotherapy, Website Category=general
Stomach Cancer - causes, symptoms, diagnosis and treatment options ...	Dr. J. Lawson, Chemotherapy, Website Category=stomach

Figure 7.14: Patient "00561c" Hospital Trusted Websites Webpage



YMDDIRIEDOLAETH GIG FELINDRE · VELINDRE NHS TRUST

Your Favorite Websites

Home | Help | logout | Logged in as Patient (00561c)

Delete from Favorites | Add to Favorites

- cancer.org
- cancerbackup.co.uk
- Does stomach cancer run in families? Can it be inherited ...
- Does Helicobacter pylori cause stomach cancer? : CancerBACUP
- cancerhelp.org.uk
- Stomach (gastric) cancer questions
- Diagnosing stomach cancer
- dipex.org
- macmillan.org.uk

Figure 7.15: Patient "00561c" Favorite Websites Webpage

7.5.3.4 The Personal Internet Search (PerIS) Webpage

PerIS is the main patient user service provided by the PHB system. It is geared to address patient Internet search challenges and facilitate a patient's access to relevant Internet information. PerIS main webpage, shown in Figure 7.16, exhibits four search features:

- **Search Topic Category:** allows the focusing of a search query using three features:
  - Patient-personalised search ideas that can be selected from three search information categories, based on a patients' own EPR data: Diagnoses, Treatment or Cancer Management Plan depending on available ISCO EPR data. For instance, Figure 7.16 shows only Diagnosis and Treatment search information categories as the ISCO database lacks records on patient "00561c" cancer management plan.
  - Diagnosis search term enrichment, as extracted from PDO data.
  - Search Topic Refinements list as retrieved from the PHB database.
- **Search Tool:** incorporates a wide range of health information search tools. This is to allow a patient to focus the Web search domain to certain Web information features. PerIS includes external search tools covering key health gateways and internal patient-customised search tools (see Section 7.4.3.4.4).
- **Search Buttons:** execute either normal or semantic search modes.

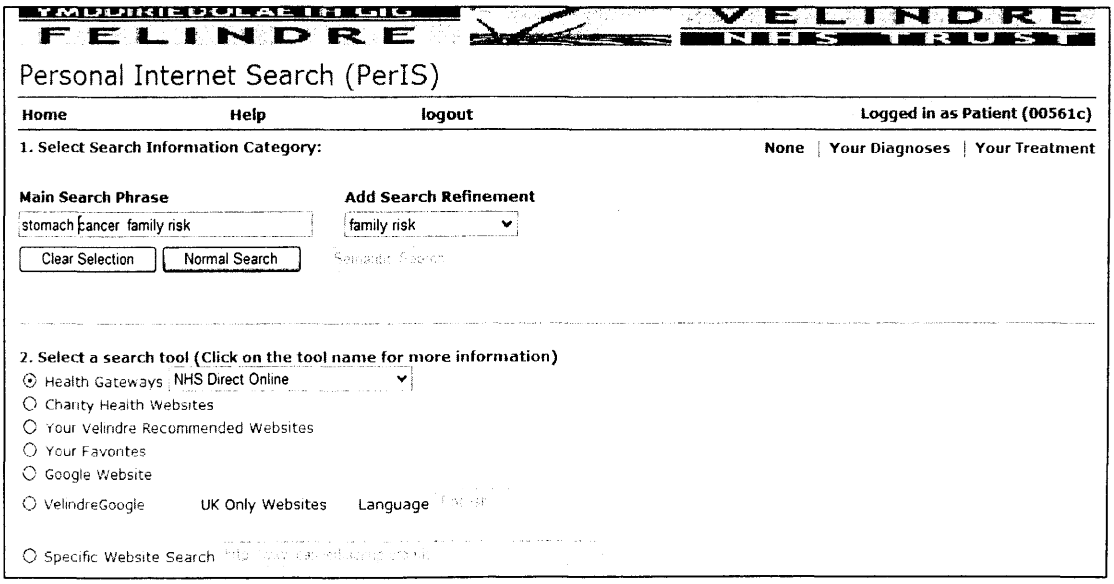


Figure 7.16: Patient "00561c" PerIS Webpage

Generally-speaking, the PerIS Internet search mechanism is distinguished by introducing the following Internet medical search features:

a. **EPR-based patient-personalised search ideas:** these are constructed directly from patient EPR data and cover three health information categories; diagnosis, treatment and cancer management plan.

b. **Diagnosis search term enrichment:** enriches diagnosis search information with various related lay, medical and generic terms. These can be used individually in a normal search or together in a semantic search to enhance normal search results.

c. **Search term refinement:** this further focuses the main search topic with patient-oriented health information types often sought by patients (e.g. side effects, risk factor).

d. **Key health gateways search:** links to key health gateways to enable a patient focus the search on prominent authenticated online health information.

e. **Hospital trusted websites search:** focuses a patient online search on authenticated online health information that is trusted by hospital officials.

f. **Charity websites search:** focuses the search on recognised non-official but authenticated online health information aimed at patients and laypeople.

g. **Patient-Favorite Websites search:** focuses the search on websites determined by a patient.

h. **Specific website search:** focuses the search on a single selected health website (e.g. cancerbackup.org.uk).

i. **Fine-grained semantic search options:** distinguish between semantic search types by allowing medical, lay, generic or full semantic search.

In addition, PerIS incorporates Google Search to offer a patient unrestricted Web search mechanism. PerIS Internet search exercise consists of three steps namely:

1. Specifying Search Information: This can be selected from PerIS patient-personalised search ideas offered by Diagnosis, Treatment and Cancer Management Plan search information categories. Alternatively, a patient can enter his/her own search information in the Main Search Phrase textbox, using the None Search information category. More search information ideas can be selected from diagnosis related terms (see Figure 7.16).

2. Refining the main search information using the drop list *Add Search Refinement*. Figure 7.17 shows the addition of the word "diet" to the main

search term "gastric neoplasm". This focuses the search information requirement to documents covering *diet* and *gastric neoplasm.*

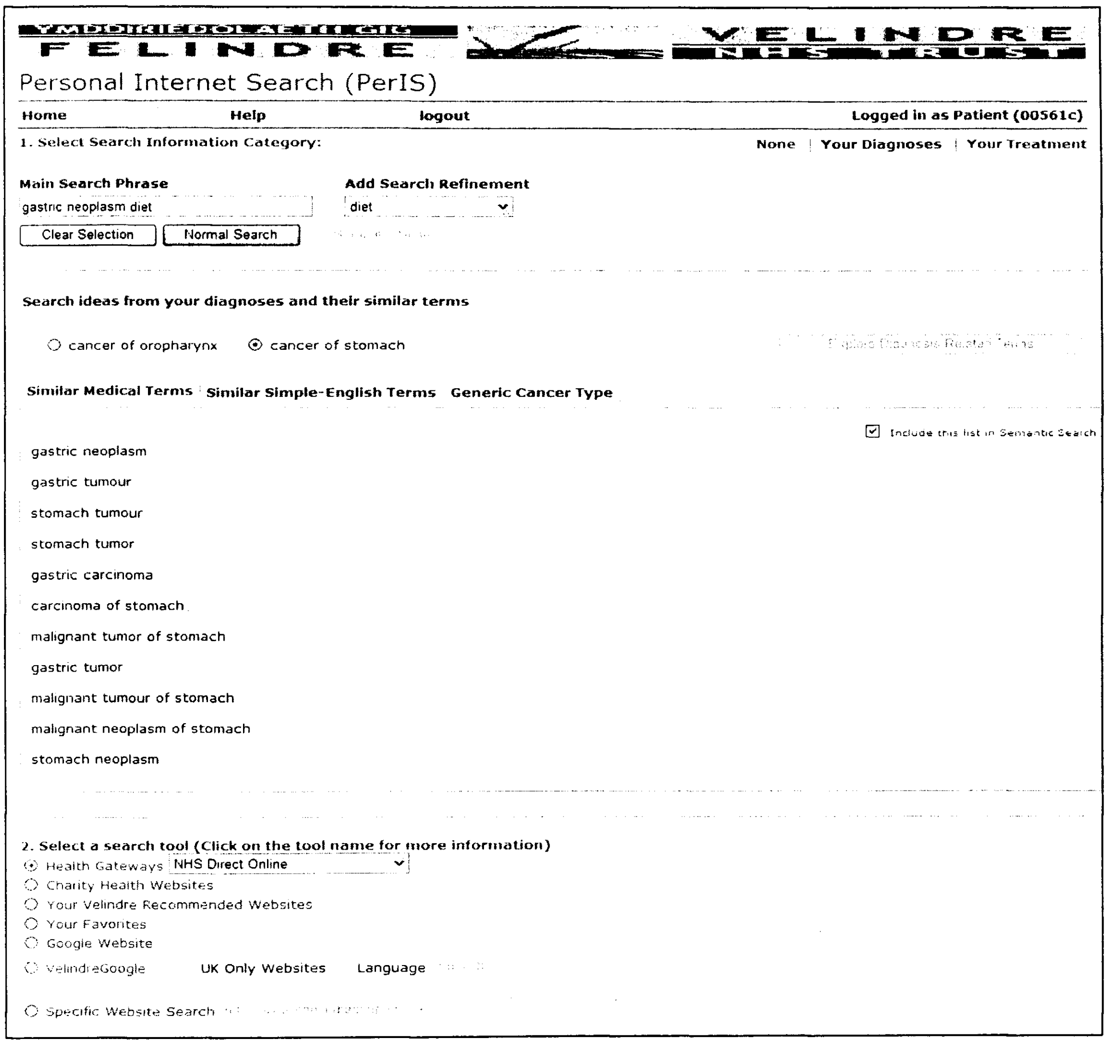


Figure 7.17: Specifying Search Information from Diagnosis Related Information

3. Selecting a preferable search tool from a wide range of key health gateways and patient-customised search tools.

4. Executing the search using either *Normal Search* or *Semantic Search* buttons.

   - ***Normal Search:*** performs a normal Web search on the specified search phrase using the selected search tool. It executes in all PerIS's search information categories and search tools.

   - ***Semantic Search:*** this is only activated for the *Your Diagnoses* information search category and PerIS's internal search tools.

A sample PerIS session, demonstrating PerIS features, is given in Appendix E.

7.6 Implementation Issues

This section describes the hardware, software and programming language implementation of the Patient Health Base (PHB) prototype system. The PHB system is developed on a PC computer.

7.6.1 Software

PHB is implemented using a three-tier client/server architecture, consisting of a client tier, a server tier and a middle tier:

- **Server Tier:** The *server tier* forms the back-end of the PHB system. It includes of the patient database server – an SQL server 2000, which is the DBMS of the Velindre ISCO database – and the Web, which is a collection of hypertext documents.
- **Client Tier:** The *client tier* denotes PHB user GUIs. It is developed using JSPs [78]. The choice of JSP technology is based on its capability to handle dynamic content and on its platform independence. The client tier interacts with the middle-tier using the Apache TOMCAT [5] Web Server that passes client requests to the middle tier components.
- **Middle Tier:** The *middle tier* performs the business logic and data processing of the PHB system, and coordinates the interactions between the client and server tiers. It is implemented as a suite of JSPs. It covers the PHB's system and user components discussed in Section 7.2.
    - **PHB User components:** execute client requests and are implemented using JSP servlets within JSP pages.
    - **PHB system components:** are implemented as JAVA classes.

The middle tier interfaces with two internal data sources:

- The relational PHB database – manages central PHB data – using the JDBC wrapper technology.
- The RDF/XML PDO file using the Jena Ontology API.

Three additional wrapper interfaces are supported with the server tier, as follows:

- JDBC to interface with the ISCO database.
- Google API to access the Google Web search engine.
- Gateway Wrapper component to execute key health gateways (see Section 7.3.6).

7.6.2 Programming Languages

The PHB implementation utilises the following programming languages:

- *Structured Query Language (SQL)*: used to interact with the relational ISCO patient database system through JDBC.
- *JAVA Language*: used to code several system components (classes) that are needed to implement the business logic and data processing functionality of the system.
- *JAVA Server Pages (JSPs)*: used to code several webpages that are responsible for the dynamic content of HTML pages.
- *Hypertext Markup Language (HTML)*: used to develop the presentation aspect of the system user interface.
- *JAVA Script (JScript)*: used to code Webpage's dynamic presentation aspects not supported by JSP.
- *Jena RDF [81]*: constructs and manipulates RDF statements constituting the Patient Diagnosis Ontology (PDO). As a JAVA RDF API, Jena is selected for its ability to construct RDF statements dynamically from JAVA structures.

7.7 Summary

This chapter presented the PHB architecture, describing main user and system components and the GUI webpages. A sample PerIS search scenario is demonstrated in Appendix E. Subsequently, PHB implementation issues are explored. A thorough evaluation of the PHB prototype system follows in Chapter 8.

CHAPTER 8

Research Evaluation

8.1 Introduction

The work presented in this thesis is based on the following hypothesis:

"Linking integrated Electronic Patient Records (EPRs) with Internet information sources enriches the patient Internet search environment and leads to an improved patient Internet search system when compared to traditional patient Internet searching."

To evaluate this hypothesis we need to demonstrate the following issues:

- Traditional patient Internet health information searching using generic search engines (e.g. Google) and health gateways is problematic.
- The feasibility of linking EPRs information with Internet information sources.
- Linking EPRs information with Internet information sources enriches the patient Internet search environment.
- Linking EPRs information with Internet information sources improves the patient Internet information search process.

This chapter evaluates our work against the hypothesis. Section 8.2 outlines the problems hindering traditional patient Internet search. Section 8.3 demonstrates the feasibility of linking EPRs to Internet health information resources. Section 8.4 illustrates how linking EPRs to Internet information sources enriches the patient Internet search environment. Section 8.5 demonstrates the improvements in our patient Personal Internet Search (PerIS) system as compared to traditional Internet search systems. Section 8.6 revisits traditional patient Internet search challenges. Section 8.7 evaluates the fulfilment of our original research goals, and finally Section 8.8 highlights the research limitations.

8.2 Traditional Patient Internet Search is Problematic

Traditional patient Internet information search is characterised as:

a. A patient may access Internet health information using Web search engines (e.g. Google), health gateways (e.g. NHS Direct Online), medical and accredited search engines or charity websites.

b. Not all patients recognize and are familiar with existing Internet health information search tools.

c. At the Velindre Hospital – Patient Information Centre, patients are guided to key Internet health websites using a printed list of key UK and International health websites. This requires a patient to type a website URL into a Web browser address bar, and browse the website for relevant information. Usually, this is done for every site individually.

d. Most patients seek online information using search engines [206]. However, some patients (especially highly educated ones) may wish to utilise medical search engines which are usually designed for professionals and qualified medical staff, while average patients are expected to use charity websites, which are aimed at patients and carers.

e. Traditional Internet search tools do not offer a patient (or user) – personalised search topics.

f. Traditional Internet search tools do not address the need for a patient health information vocabulary preference (e.g. medical, lay, generic).

g. Online patients (or e-Patients) are usually not guided to quality health websites.

Figure 8.1 summarises the challenges affecting adversely patient Internet searching as discussed in Section 4.4.2.

1. Inaccessibility of patient personal medical information.
2. Variable patient search information requirements.
3. Generic health websites list, utilised by ISCO patients.
4. Laborious, manual and generic nature of patient Internet search.
5. The wide-ranging and disparate nature of Internet health information search tools.
6. Internet information quality – difficulty in identifying trusted Internet information.
7. Health information vocabulary – difficulty in expressing medical and lay terms and identifying related terms.
8. Internet information overload – there are a large number of online data sources and/or large size of search result sets.
9. Internet information pollution – misinformation, unclear information or irrelevant details on the Internet sources.
10. Lack of Internet information coordination and sharing between patients and professionals.

Figure 8.1: Patient Internet Medical Search Challenges

8.3 The Feasibility of Linking EPRS to Internet Health Information Sources

This study proposes a new approach to improving patient Internet searching by:

1. Enabling a patient to access his/her own EPR, and
2. Linking information in EPRs so it can be used in searching Internet health information sources.

The first enables a patient to access their personal medical information so they can view it and use it when required. The second utilises a patient's personal medical information to customise patient Internet searching and focus the search to relevant Internet information.

To demonstrate the hypothesis, a Web-based patient personalised health information prototype system is developed, called the Patient Health Base (PHB). The PHB system interfaces between the patient database records and the Internet. It offers a patient access to essential personal medical information such as diagnoses and treatment episodes, extracted from the EPR. In addition, it incorporates a patient-personalised Internet search service that links the patient's medical data to relevant Internet health websites. PHB utilises the emerging PHR technology to offer a patient online access to personal medical information and a personalised Internet search facility.

Section 8.3.1 evaluates the feasibility of enabling a patient online access to EPR, whereas section 8.3.2 evaluates the feasibility of linking EPR data with Internet information sources.

8.3.1 The Feasibility of Enabling Patient Access to EPR

This study is conducted within a cancer oncology centre – the Velindre NHS Trust – that records patient information using the ISCO/CaNISC system. ISCO/CaNISC is an integrated health record system for cancer patients covering all Wales. As it is an investigational study, it utilises an anonymised version of the ISCO system. Currently, patients are not permitted electronic access to their ISCO/CaNISC personal medical records. However, a new health information infrastructure is being developed within the Information Healthcare (IHC) Program in Wales, to enable a patient to access their own integrated medical records. Eventually, all patients in Wales will be able to access a summary personal medical record through a Web portal called "MyHealthOnline" [95].

Our study is inline with the upcoming national health information changes in Wales. It can be thought of as an extension to the emerging PHR system to incorporate a personalised patient Internet search service that utilises the patient's personal medical information held by a PHR and/or EPR.

The online patient personal health information system, developed in this study, is a proof of concept system that only extracts essential medical information deemed useful for patient education and empowerment. It can be implemented as a Web-portal above a patient medical record system (e.g. ISCO). No changes are required to the underlying (ISCO) patient database. Patient medical information is extracted from the patient database using wrapper technology (e.g. JDBC) and transformed to a Web format (e.g. HTML, XML) using JSP technology.

8.3.2 The Feasibility of Linking EPR Data to Internet Information Sources

Our approach to linking EPR data to Internet information sources follows a mediator architecture that defines a mediation layer interfacing between the relational patient database (e.g. ISCO system) and Internet information sources. The mediator layer is made up of a set of tools (or components) handling the extraction of patient personal data from ISCO EPRs, the enrichment and transformation of this data into variant query methods suited to the underlying Internet search tools.

The mediator layer is implemented as a suite of JSPs running on a TOMCAT Server. The choice of JSP is based on the requirement to interface between a relational patient database and a platform independent client interface (Web browser). As JSP is based on the platform-independent JAVA Virtual Machine (JVM), it makes JSP pages executable on any machine by installing JVM.

Access to the patient database EPRs is achieved using JDBC technology. JDBC executes SQL queries against the relational patient database and stores query results in JAVA data structures. Search results are then manipulated in the JSP environment and presented to the patient in a Web format (e.g. HTML), as this is more useful to a patient. This is to alleviate a patient from the technicalities involved in querying the patient database directly. On the other hand, linkage of medical data to Internet information sources is achieved using two techniques: Google API technology, and website search URL.

8.3.2.1 Evaluating the Use of Google API

In order to automate patient Internet searching, a mechanism is required to allow the passing of patient-personalised search queries to an Internet search engine. Google offers Google API, a JAVA API to the Google search engine, to enable executing and manipulating the Google search engine from JAVA programs. In this study, Google API is used to execute patient search queries, developed using data extracted from EPRs. In addition, Google API to the Google search engine is used to create customised search engines within the PerIS system as discussed in Section 7.3.5.

Google API offers the means to customise a Google search for a patient with respect to the search language, search domain, running multiple Google searches hidden from the user and other ways. Figure 8.2 compares ten features supported by PerIS against the features of a Google website search. These features are designed to meet patient health information requirements identified in this study.

Patient Search Information Requirement	**PerIS**	**Google**
Patient-personalised search topics/ideas	•	X
Patient preferred search language (e.g. English, French)	•	•
Specific website search	•	•
Multiple website search	•	X
Hospital recommended websites	•	X
Patient preferred websites	•	X
Semantic search	•	X
Medical term only semantic search	•	X
Lay term only semantic search	•	X
Generic term only semantic search	•	X

• Supported
X Not Supported

Figure 8.2: Comparison of PerIS Search Features to Google Website Search

Google search only supports two of the identified PerIS search features: specific website search and setting a search language. In addition, setting a search language or specific website search (i.e. site restriction) on Google is not as straightforward as in PerIS. For instance, setting site restriction in Google requires familiarity with the Google site restrict construct "site:" and typing the correct form of the Google search query (e.g. stomach cancer site:http://www.cancerbackup.org.uk). Figures 8.3 and 8.4 illustrate setting a restricted website search in PerIS and Google respectively.

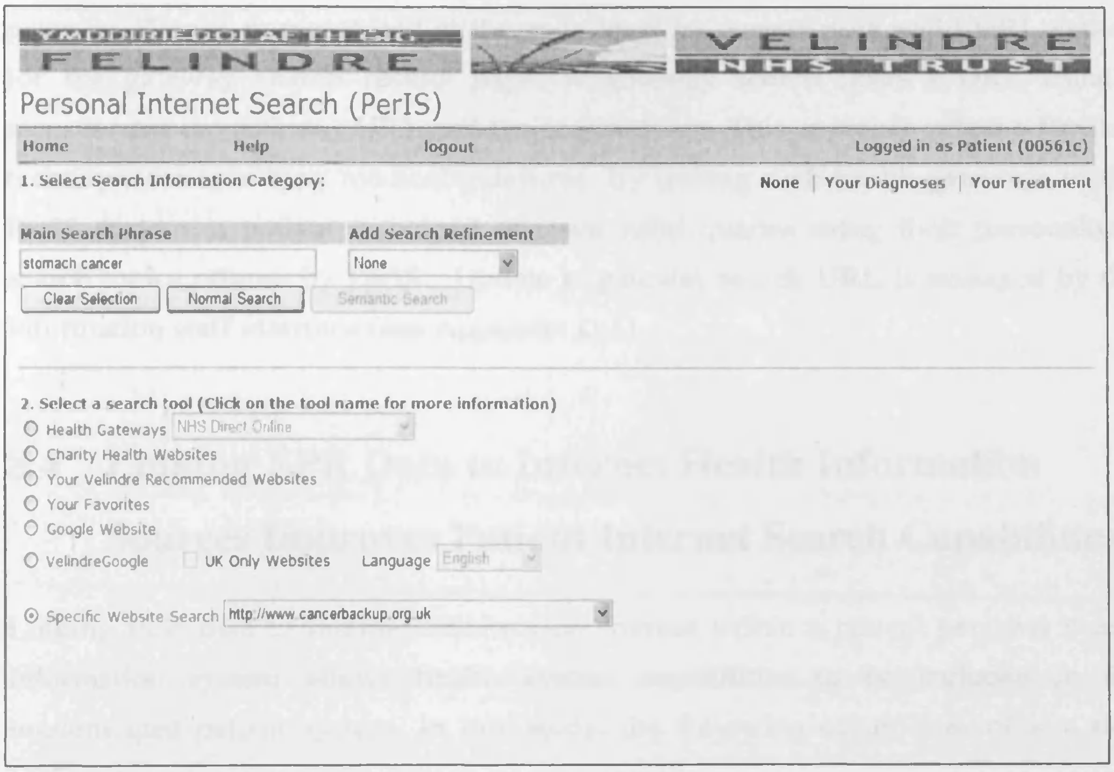


Figure 8.3: Setting Specific Website Search to cancerbackup.org.uk on PerIS

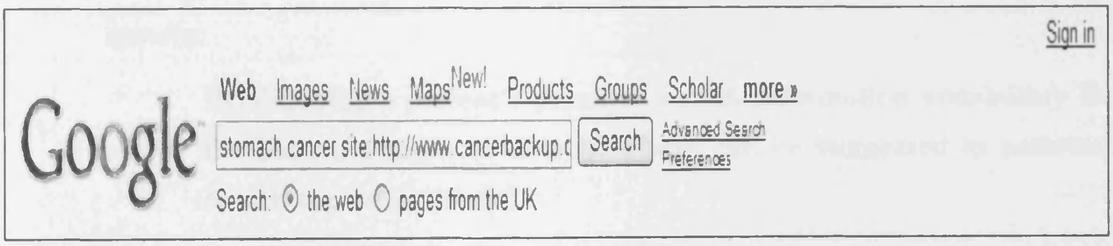


Figure 8.4: Setting Website Restrict Search on Google, Searching for" stomach cancer" only on cancerbackup.org.uk

8.3.2.2 Evaluating the Use of Website Search URL

PerIS integrates with key health gateways and medical search engines (e.g. MedlinePlus, NHSDirectOnline, HONCode) to aid patient access to key Internet health information sources more efficiently. Such Internet information sources could either be unknown to a patient and/or accessed separately in multiple Internet search sessions. Unlike Google search, health gateways and medical search engines do not provide an API to utilise their underlying databases and search capabilities.

Hence, PerIS connects directly to key health gateways and medical search engines using their search URL. PerIS passes the patient search phrase to the respective gateway. This is implemented at the code level by constructing valid URL syntax for the gateway search results page. A gateway search result's URL usually incorporates the gateway URL and the search query. This approach offers a flexible technique for searching medical gateways. By linking such health gateways to the PerIS system, a patient is helped to issue valid queries using their personalised search topics offered by PerIS. Update to gateway search URL is managed by the information staff interface (see Appendix D.4).

8.4 Linking EPR Data to Internet Health Information Sources Improves Patient Internet Search Capabilities

Linking EPR data to Internet information sources within a patient personal health information system allows health system capabilities to be included in the implemented patient system. In this study, the following capabilities offer a rich environment for improving patient Internet search:

- **EPR**: Offers three personalisation features in the patient Internet search system:
    - Establishing a patient's personal health information vocabulary from the patient's medical details. These can be suggested to patients as search topics (or ideas).
    - Customising a Hospital-trusted health websites list for a patient according to the patient's EPR medical details.
    - Customising semantic knowledge for a patient so that the patient is only presented with semantic data that are relevant to his/her medical information (diagnosis) concept. In a non-tailored semantic knowledge model, patients usually browse an entire semantic knowledge model (e.g. thesaurus, ontology) to select relevant concepts or information whereas this limits the presented terms by using the EPR knowledge.
- **EPR medical classification system:** This can be used to establish a valid medical terminology and semantic knowledge model for a patient to utilise during Internet searching. This assists formulating correct medical terms in potential patient search queries. In addition, it allows the identification and use of similar and hierarchical related medical terms stored in the database.

- **Information staff/Specialist:** A specialist information staff role is essential in ensuring proper medical-to-lay term mappings from both the medical and lay communities' perspectives. Medical-to-lay term mapping can improve patient health information understanding and enrich health information search results.
- **Healthcare professionals:** Interested healthcare providers can establish a trusted health websites list to guide their patients to trusted health information. A hospital-trusted health websites list allows a patient to focus their search to safe Internet information as opposed to the potentially unsafe generic Web search. This can reduce Internet information overload for patients and guide them to officially trusted Internet information. In fact, the hospital-trusted health websites list offers a second chance of verifying a health website which has been verified by third-party accreditation organisations. The utilisation of EPR also helps customise such a hospital-trusted websites list to patients.
- **Coordinating Internet information between patients and professionals:** Both patients and professionals are usually concerned about Internet health information quality. Allowing both parties to establish and share lists of trusted (or preferred) health websites can better guide patients to trusted and patient-relevant health websites. Healthcare professionals can benefit from patient Internet research whereas patients are guided to hospital-trusted and recommended Internet information which adds an authoritative aspect and support to the Internet health information access patients undertake.

8.5 Linking EPR Data to Internet Health Information Sources Improves Traditional Patient Internet Search

The improvement in patient Internet search using PerIS can be evaluated by comparing PerIS search capabilities, focusing techniques and search results against the traditional patient Internet search. Section 8.5.1 evaluates the PerIS Internet Search capabilities whereas Section 8.5.2 evaluates PerIS search focusing techniques. Finally, Section 8.5.3 demonstrates improvement in terms of search results by using PerIS.

8.5.1 Evaluating PerIS Search Capabilities against Traditional Patient Search Capabilities

This section compares PerIS search capabilities against those of a traditional patient Internet search outlined in Section 8.2. PerIS capabilities are designed according to patient information search requirements as investigated in this study.

PerIS offers direct and personalised access to numerous information sources from within the PerIS interface. This allows a patient to run personalised search queries using various Internet search tools more efficiently. Figure 8.5 illustrates the PerIS search capabilities and compares them against standalone search tools utilised by this study, to indicate whether the tool has the capability in a fully-supported, partially-supported or not supported state.

As shown in Figure 8.6, the majority (67%-95%) of PerIS capabilities are not supported by any of the standalone Internet search tools investigated in this study. Google offers the maximum (14%) full support of PerIS capabilities. This could be due to Google's popularity and competitive strategy to address as wide a range of user needs as possible. Among different Internet health gateways, HONCode offers the maximum (24%) partial support. This highlights the significance of HONCode for health information searching. Figure 8.7 clearly demonstrates the absence of patient search information requirements provided by PerIS capabilities in standalone non-patient-tailored Internet health information search sources. This clearly demonstrates that PerIS offers an improved patient Internet search system in terms of search capabilities.

No	Search Capabilities	PerIS	Google	HONCode	MedHunt	MedlinePlus	NHSDirect	CancerBackup	CancerHelp
1	Personalised search topics/ideas from patient diagnosis	•	X	X	X	X	X	X	X
2	Personalised search topics/ideas from patient treatment	•	X	X	X	X	X	X	X
3	Personalised search topics/ideas from patient CMP	•	X	X	X	X	X	X	X
4	Rich diagnosis vocabulary	•	X	X	X	X	X	X	X
5	Personal health information vocabulary	•	X	X	X	X	X	X	X
6	Focus search on specific information types (e.g. family risk)	•	•	X	X	O	X	X	X
7	Search language (e.g. English, French)	•	•	O	O	X	O	O	X
8	Single website restrict search	•	•	X	X	X	X	X	X
9	Multiple website restrict search	•	X	X	X	X	X	X	X
10	Hospital-trusted and recommended websites search	•	X	X	X	X	X	X	X
11	Patient preferred websites search	•	X	X	X	X	X	X	X
12	Charity websites search	•	X	X	X	X	X	X	X
13	Semantic search	•	X	X	X	O	O	X	X
14	Medical term only semantic search	•	X	X	X	X	X	X	X
15	Lay term only semantic search	•	X	X	X	X	X	X	X
16	Generic term only semantic search	•	X	X	X	O	X	X	X
17	Saving search results to patient Favorites	•	X	X	X	X	X	X	X
18	Individual search of key health gateways	•	O	O	O	O	X	X	X
19	Individual search of medical search engines (e.g. MedHunt)	•	O	O	•	O	X	X	X
20	Individual search of third-party accredited search engine (e.g. HONCode)	•	O	•	O	X	X	X	X
21	Individual search of charity websites	•	O	O	X	X	O	O	O

• Fully Supported
O Partially Supported
X Not Supported

Figure 8.5: Comparison of PerIS Search Capabilities to Stand-alone Internet Search Tools used by PerIS

Feature	Google	HONCode	MedHunt	MedlinePlus	NHSDirect	CancerBackup	CancerHelp
Fully Supported	14%	0%	0%	5%	5%	0%	0%
Partially Supported	19%	24%	14%	19%	14%	10%	5%
Not Supported	67%	76%	86%	76%	81%	90%	95%

Figure 8.6: Ratio of PerIS Capabilities Supported by Various Internet Search Tools

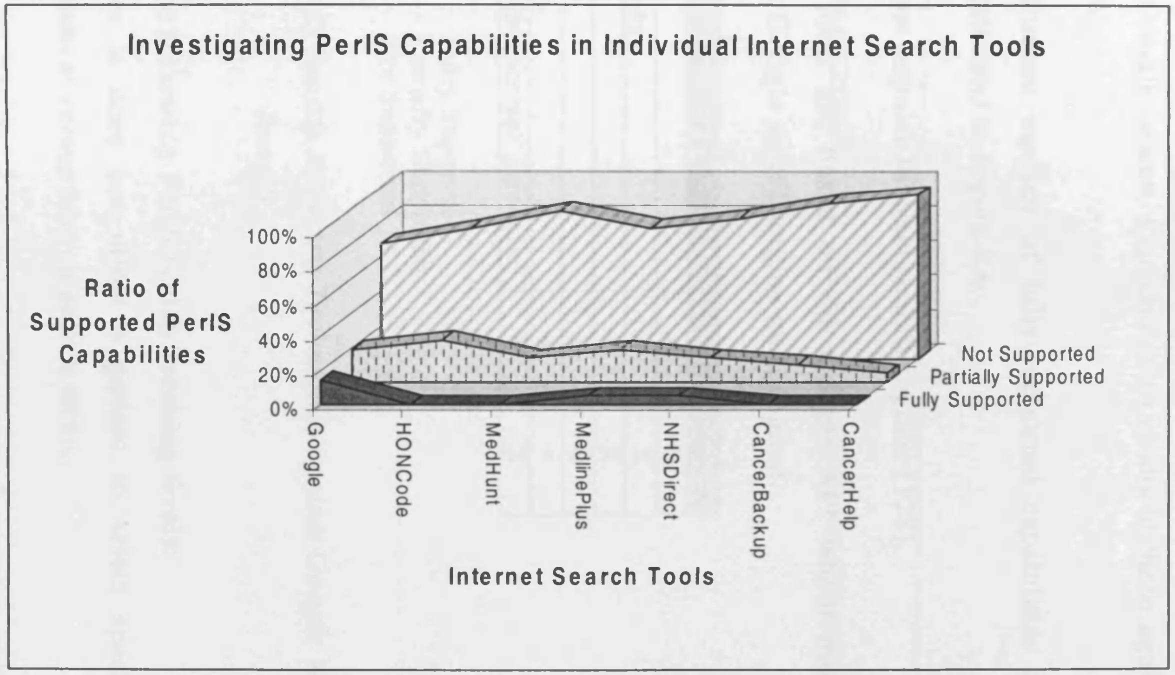


Figure 8.7: PerIS Capabilities Supported by Various Internet Search Tools

8.5.2 Evaluating PerIS Search Focusing Techniques

PerIS is geared to support patient information requirements for online medical search. It allows a patient to focus the search at several levels, by providing multiple search focusing techniques. PerIS offers six approaches to search focusing: Search topics, Search vocabulary, Search tool, Search domain, Search language, and Search mode. Figure 8.8 identifies the extent to which PerIS search focusing techniques are provided in a Google search. Our choice to evaluate these against Google is based on three factors:

- Google offers the maximum number of fully-supported capabilities that overlap with PerIS as indicated in Figure 8.6.
- Google is one of the most popular Internet search engines [128].
- PerIS internal search tools are based on the Google API which makes comparison to the main Google search engine more logical.

Search Focusing Technique	PerIS	Google
Search topics	•	O
Search vocabulary	•	X
Search tool	•	X
Search domain	•	O
Search language	•	•
Dual Search mode (normal, semantic)	•	X

•	Fully Supported
O	Partially Supported
X	Not Supported

Figure 8.8: Comparison of PerIS Search Focusing Techniques against Google Web Search

Google offers no support for the following PerIS search focusing levels:

- **Search vocabulary:** as it does not allow a patient to select specific vocabulary type (e.g. medical versus lay) on search terms.
- **Search tool:** Google does not allow a patient to focus the Internet search to certain Web information gateways (e.g. HONCode, NHSDirectOnline).
- **Search mode:** Google operates in a single keyword search mode and has no support for semantic search.

Google fully supports search language focusing and partially supports focusing search domain (see Section 8.5.2.3) and search topics (see Section 8.5.2.4). The following subsections evaluate PerIS in terms of each of the outlined search focusing levels as compared to Google.

8.5.2.1 Evaluating PerIS Search Term Vocabulary Focusing Technique

Focusing search term vocabulary allows a patient to adjust the reading level and understandability of Internet information. As a non-professional, the average patient often uses lay language when seeking online information. However, other patients seek health information using medical terminologies, usually as a result of consultation with a doctor.

Google uses a keyword-based search that searches only for specified keywords, and suffers the following limitations:

- It does not address search term semantics,
- It does not apply search term semantic enrichment, and
- It does not distinguish between medical and lay health information terminology.

In contrast, PerIS offers rich information vocabulary capabilities:

- It utilises search term semantics to extend the search information vocabulary for a patient allowing a patient to view and utilise similar and related search information terms.
- It offers a semantic search facility to augment normal search results with semantic data results.
- It incorporates two patient information requirements regarding search term vocabulary, as it offers:
  - Distinct medical and lay search term forms.
  - A generic search term form to aid a patient to explore generic health information on the search term.

Semantic search term capabilities are currently applied to diagnosis search terms. For instance, a patient seeking information on a health condition such as "cancer of stomach" can view numerous similar and related terms in medical, lay and generic term forms, and utilise them individually or collectively. Figure 8.9 illustrates different search terms and categories offered by PerIS to focus and/or enrich the diagnosis search term: "cancer of stomach".

Search Vocabulary Focusing level	PerIS	Google
Lay terms		
cancer of stomach	•	X
stomach cancer	•	X
Medical Terms		
gastric neoplasm	•	X
gastric tumour	•	X
stomach tumour	•	X
stomach tumor	•	X
gastric carcinoma	•	X
carcinoma of stomach	•	X
malignant tumor of stomach	•	X
gastric tumor	•	X
malignant tumour of stomach	•	X
malignant neoplasm of stomach	•	X
stomach neoplasm	•	X
Generic Cancer Terms		
Upper gastrointestinal cancer	•	X
cancer of digestive organs	•	X
Search focusing by Semantic Search Category		
All diagnosis vocabulary (full Semantic Search)	•	X
All medical diagnosis terms	•	X
All lay diagnosis terms	•	X
All generic diagnosis terms	•	X
All medical and lay terms	•	X
All medical and generic terms	•	X
All lay and generic terms	•	X

• Supported
X Not Supported

Figure 8.9: Comparison Search term vocabulary focusing levels in PerIS and Google for main search term "cancer of stomach"

While Google offers no search term enrichment and vocabulary focusing, PerIS enables a patient to conduct 22 focused searches related to the main search term vocabulary as illustrated in Figure 8.10.

Search Term vocabulary focusing dimension	Possible Searches
Lay terms	2
Medical terms	11
Diagnosis Generic Cancer Terms	2
Semantic Search Category	7
Total:	**22**

Figure 8.10: Breakdown of the number of potential searches

8.5.2.2 Evaluating PerIS Search Language Focusing Technique

Information language preference is vital for patients [164] especially for ethnic communities and in countries using multiple languages (e.g. Canada, Wales-UK). This section evaluates search language focusing techniques in PerIS and Google. Google fully supports search language focusing. However, search language setting in Google is only available in the advanced search webpage (Figure 8.11) whereas PerIS offers this feature directly in the main PerIS webpage (Figure 8.12).

PerIS applies language focusing in the "VelindreGoogle" search tool, an internal PerIS Web search tool. Language focusing could not be applied to both external and specific domain internal PerIS search tools. External search tools (e.g. health gateways, such as HONCode) do not enable access to their internal search engines, and manipulating their search results is unreliable as Web page presentation could change and this would involve modifying PerIS. PerIS VelindreGoogle utilises the same set of languages used by Google. However, not every patient preferred search language is supported by Google (e.g. Welsh, Somali, Swahili). This can be addressed in a future work study.

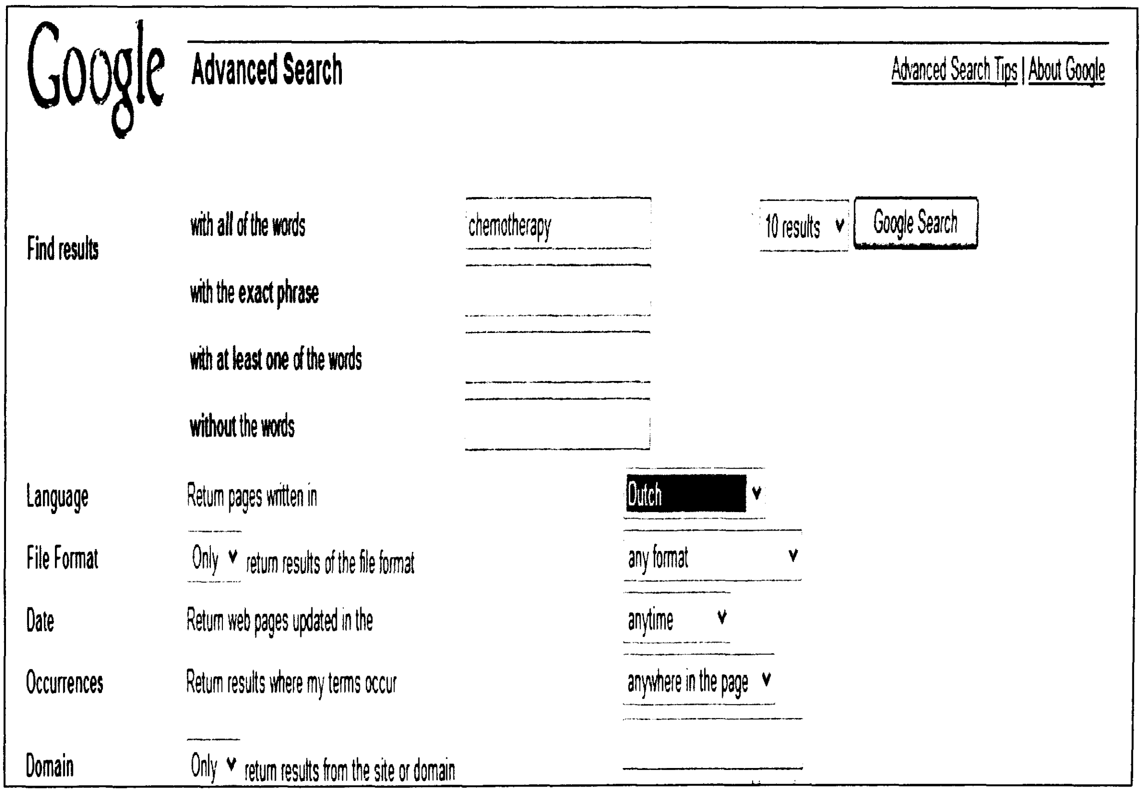


Figure 8.11: Setting search language in Google

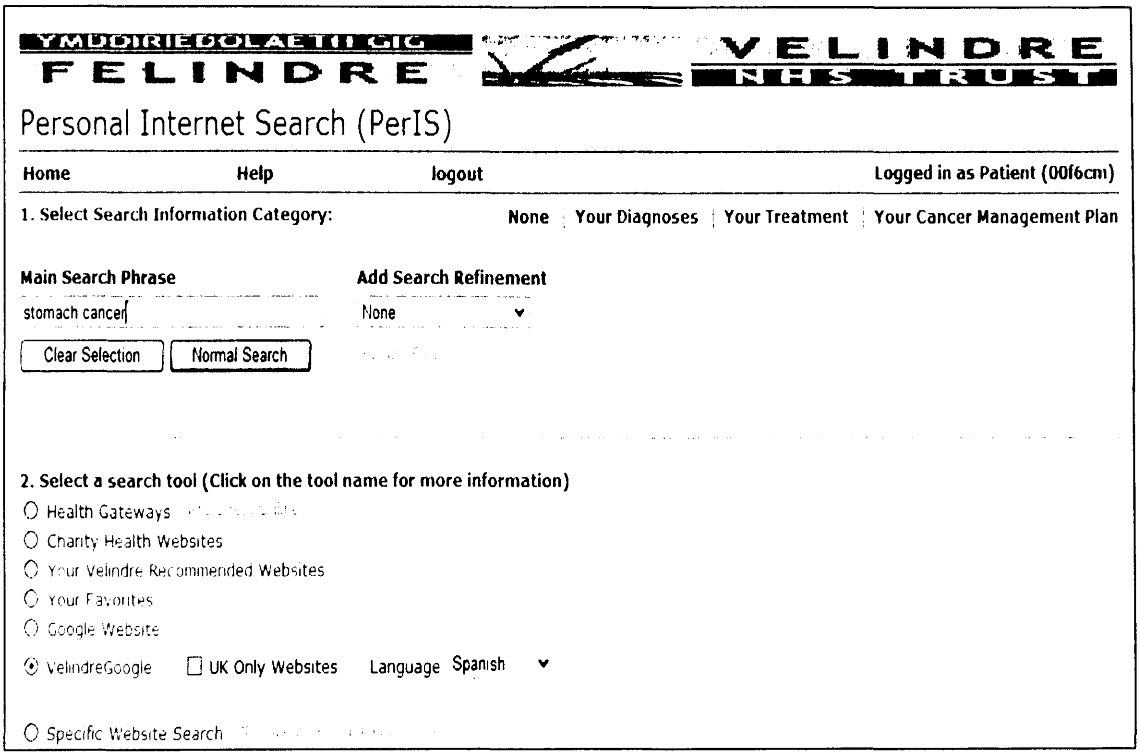


Figure 8.12: Setting search language in PerIS

8.5.2.3 Evaluating PerIS Search Domain Focusing Technique

This section evaluates search domain focusing in PerIS and Google. Search domain denotes the number of websites utilised in a given search query. Search domain can be focused to a single website, multiple websites or the entire Web. Domain search focusing is useful if users wish to retrieve information or documents only from certain information sources. This can improve information overload problems for a patient by reducing the number of results. Figure 8.13 compares search domain focusing in PerIS and Google.

Search Domain Focusing level	**PerIS**	**Google**
Single Website restriction	●	●
Multiple Websites (Examples)		
.gov (US government)	X	●
.edu (US universities)	X	●
.ac.uk (UK universities)	X	●
.nhs.uk (UK NHS)	X	●
.org	X	●
.org.uk	X	●
Hospital Trusted Websites	●	X
Patient Preferred Websites	●	X

● Supported
X Not Supported

Figure 8.13: Comparison of search domain focusing levels in PerIS and Google

Both PerIS and Google support single website search. However, this operation is straightforward and simpler in PerIS where a patient only needs to select a website from a list of key health websites (see Specific Website Search in Figure 8.14).

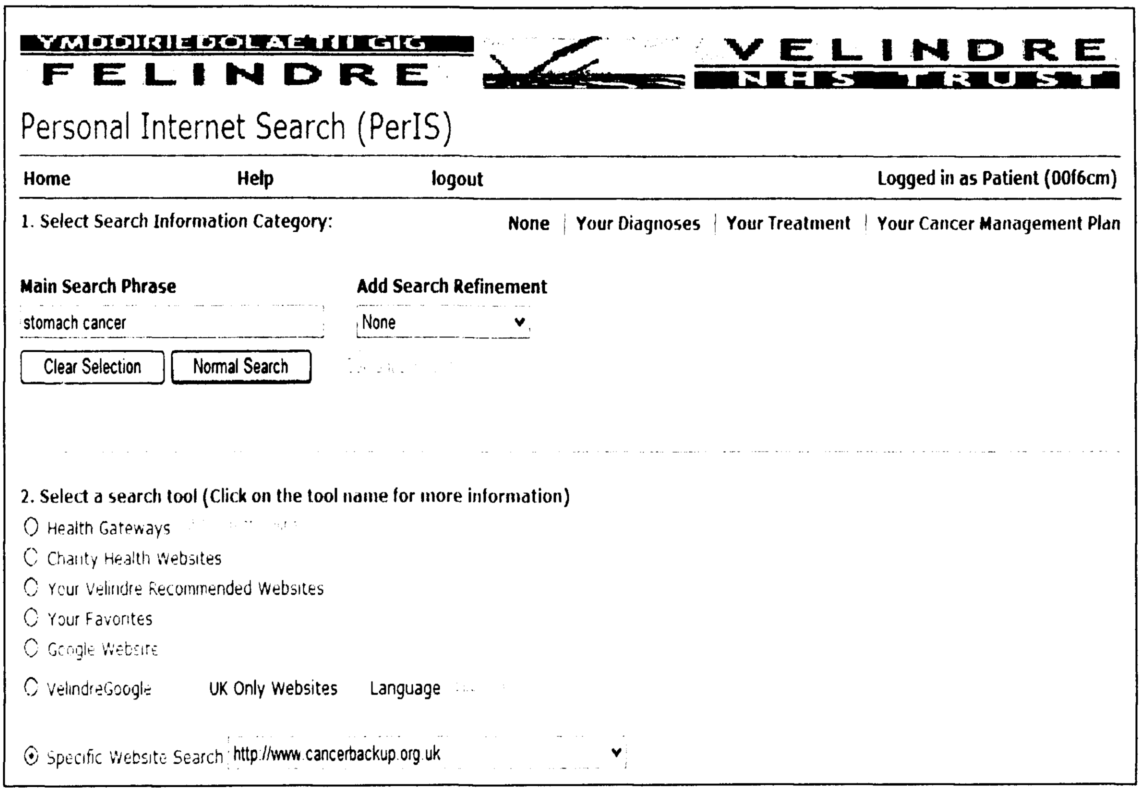


Figure 8.14: Specific Website Search in PerIS

In Google, website restriction can be specified in the search phrase or in advanced search page. Google allows a single website search using the restrict search method ("site:"), issued in the query phrase. For example, the Google search phrase:

Stomach cancer site:http://www.cancerhelp.org.uk

Would search for "stomach cancer" only on http://www.cancerhelp.org.uk. All documents retrieved by this search query are from the specified website. Thus, applying website restriction on Google puts the onus on the patient to learn and apply this command correctly. PerIS saves a patient from this burden and performs this operation implicitly. PerIS only requires a patient to specify the search phrase, select the "Specific Website Search Tool" followed by selecting a given website from those presented.

Multiple website restriction is enabled in both PerIS and Google but at different levels. Google enables multiple websites search for known domains (such as .gov (US government), .edu (US academic), .nhs.uk (UK NHS)). Such techniques may not be known to all Web users including patients. They are also not natural domains for medical sites and for patients to identify. Such techniques are not currently implemented by PerIS but they can be easily added to the list utilised by the specific website search. This will only require a few hours of coding.

PerIS enables special patient-oriented multiple website searches, a feature not currently supported by Google. A set of hospital-trusted health websites (Your Velindre Recommended Websites) and a set of a patient's preferred websites (Your Favorites) (see Figure 8.14). These techniques offer the patient and the hospital more control over the search domain than the generic domain search techniques supported by Google for all Web users. Very recently, Google offered a Customised Search Engine (CSE) [39] service that enables a user to set the websites to be searched. This can be similar to PerIS patient's Favorites search tool. However, it lacks linkage to EPRs, and thus utilising patient personalised search topics as in PerIS is not straightforward.

8.5.2.4 Evaluating PerIS Internet Search Topic Focusing Technique

This section evaluates the search topic focusing in PerIS. First we evaluate multiple search topic focusing methods in PerIS against Google (Section 8.5.2.4.1). Second, we present an illustrative evaluation of search topics offered by PerIS and Google for a patient seeking Internet information on "stomach cancer" (Section 8.5.2.4.2).

8.5.2.4.1 Evaluating Different PerIS Search Topic Focusing Methods

Google partially supports search topic focusing. As a general-purpose search engine, Google usually offers no user-personalised search topics. However, Google assists medical Internet search by presenting the user with some search topic refinements (Figure 8.15). Google supports this feature only when a user specifies a medical condition search term but not other medical information such as treatment (e.g. radiotherapy).

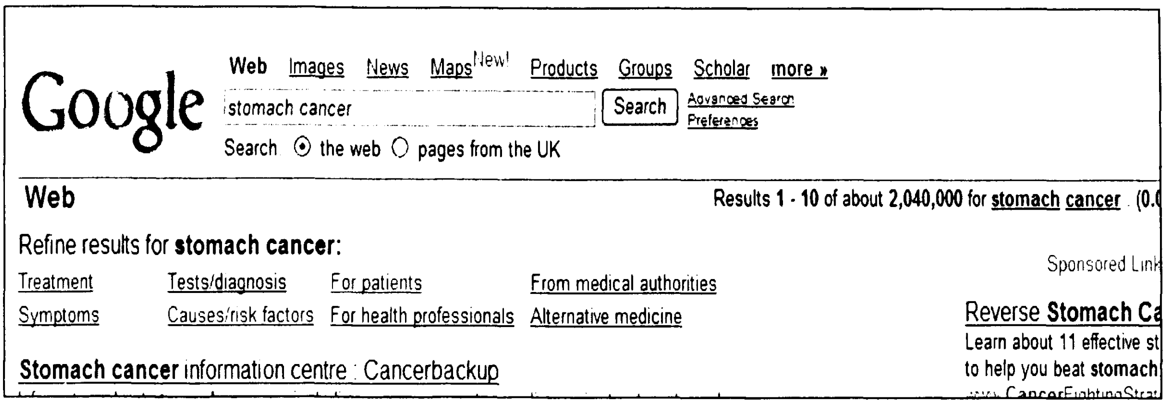


Figure 8.15: Google Offering Eight Search Topic Refinement Titles for Patients on Main Search Phrase "stomach cancer"

In contrast, PerIS is a patient-personalised Internet search facility. It offers three methods for personalising and focusing a search topic for a patient, namely:

a. **Personalised search topics (i.e. ideas) from patient medical details:** three types of personal medical information are used from the EPR, namely:
   - Patient diagnosis,
   - Patient treatment episodes, and
   - Patient cancer management plan.

b. **Search topic enrichment:** by extending diagnosis-based search information with synonyms and hierarchical terms related to a given diagnosis.

c. **Search topic refinement:** can be achieved in three ways:
   - By adding a specific health information type to the main search topic, from an extensive list of health information types often sought by patients (e.g. tumour marker, family risk).
   - By selecting a certain personalised search topic.
   - By selecting a particular diagnosis synonym, hierarchical term or lay or medical search category (Figure 8.13).

Figure 8.16 compares the search topic focusing techniques supported by PerIS and Google. Google only partially supports one of the PerIS search topic focusing techniques; search topic refinement, by asking a user to select specific health information types to refine the main search topic. Google offers less information types for refining the main search phrase than PerIS. For instance, Google offers 8 specific information types (or categories) to refine a patient search for "stomach cancer" (see Figure 8.15) whereas PerIS offers a list of 35 specific information types to focus the main search topic (see Figure 8.17).

Search Phrase Focusing Technique	PerIS	Google
Patient-Personalised Search Ideas	•	X
Search Phrase enrichment	•	X
Search Phrase refinement	•	O

• Fully Supported
O Partially Supported
X Not Supported

Figure 8.16: Comparison of Search Phrase Focusing Techniques between PerIS and Google

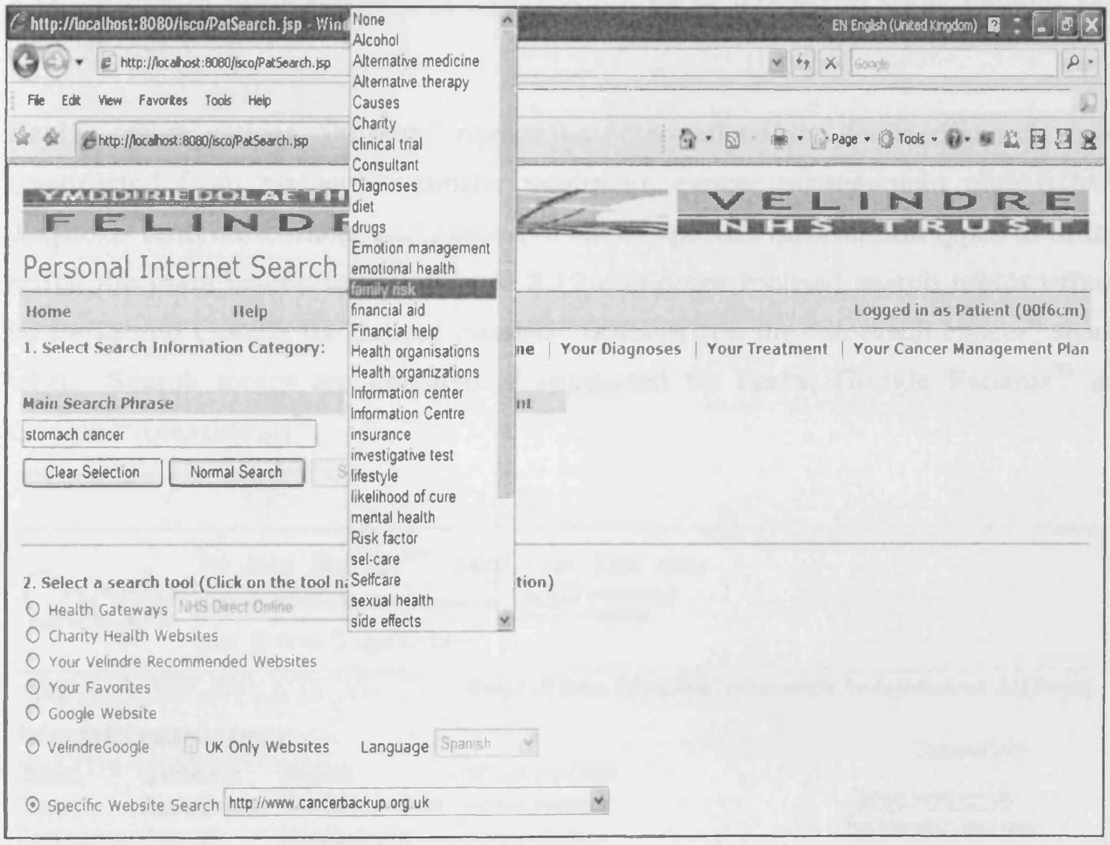


Figure 8.17: Specific Information Types Suggested by PerIS for Focusing Patient Main Search Phrase "stomach cancer"

8.5.2.4.2 Evaluating PerIS Focused (Patient-Personalised) Search Topics

This section evaluates PerIS in terms of search topics offered to a patient and compares them with the search topics offered by Google. We demonstrate this evaluation through an example of ISCO patient "00f6cm" seeking Internet information on "stomach cancer".

Google: Google offers no initial search topics. If the patient now enters the search term "stomach cancer", then Google, subsequently, offers search term refinement as discussed in Section 8.5.2.4.1. Google presents this patient with 10 refinement topics to refine the main search topic on a health condition. These topics are offered using eight topic refinement titles: Treatment, Tests/Diagnosis, For patients, From medical authorities, Symptoms, Causes/risk factors, For health professionals and Alternative medicine (see Figure 8.15). Seven of these titles are meant for patients since the title "For health professionals" is aimed at professionals. In addition, clicking on "For health professionals" will show more refinement topics: Patient handouts, Clinical trials, Continuing education, and Practice Guidelines (see Figure 8.18). Some of these refinement topics could be of interest to some patients (e.g. clinical trials).

PerIS: offers patient "00f6cm" numerous personalised and focused search topics constructed from his/her diagnosis, treatment, cancer management plan (CMP), diagnosis term enrichment and extensive list of specific information types to further refine the main search topic. Figure 8.19 compares focused search topics offered by PerIS and Google for patient number "00f6cm" on the "stomach cancer" search term. Search topics are checked if suggested by PerIS, Google Patients[20] and Google Professionals[21].

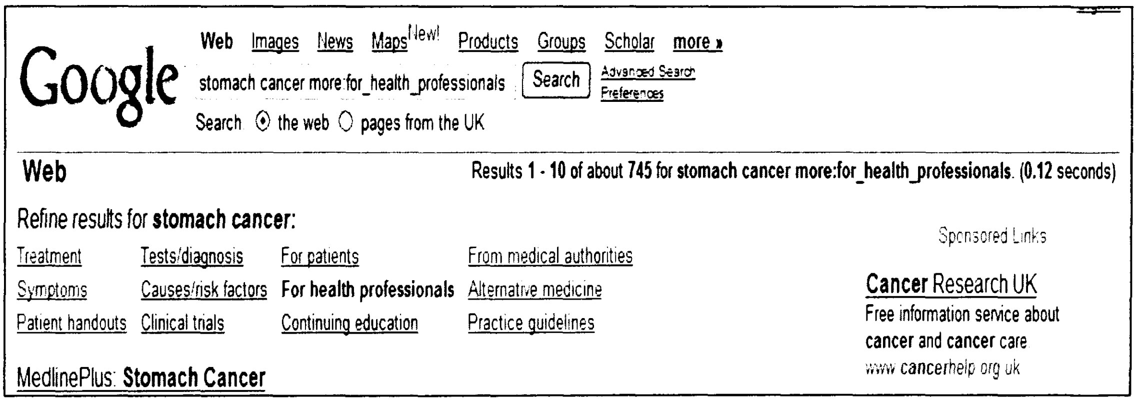


Figure 8.18: Google Offering Twelve Search Topic Refinement Titles for Professionals on Main Search Phrase "stomach cancer"

[20] Google search refinement topics displayed when clicking on "For Patients" (see Figure 8.18)
[21] Google search refinement topics displayed when clicking on "For Health Professionals" (see Figure 8.18)

No	**Suggested Search information**	**PerIS**	**Google patients**	**Google professionals**
	From Patient Personal Medical Data			
	From Diagnosis Data			
1	cancer of stomach	•	X	X
	From Treatment Data			
2	Chemotherapy	•	X	X
3	Palliative care	•	X	X
	From Chemotherapy Treatment Data			
4	chemotherapy ECF	•	X	X
5	ECF	•	X	X
	From Palliative care			
6	Palliative care	•	X	X
	From Cancer Management Plan			
7	Palliative chemotherapy	•	X	X
	From Diagnosis Term Enrichment			
	From Diagnosis Medical Term Synonyms			
8	Gastric neoplasm	•	X	X
9	Gastric tumour	•	X	X
10	stomach tumour	•	X	X
11	stomach tumor	•	X	X
12	gastric carcinoma	•	X	X
13	carcinoma of stomach	•	X	X
14	malignant tumor of stomach	•	X	X
15	gastric tumor	•	X	X
16	malignant tumour of stomach	•	X	X
17	malignant neoplasm of stomach	•	X	X
18	stomach neoplasm	•	X	X
	From Diagnosis Lay Term Synonyms			
19	cancer of stomach	•	X	X
20	stomach cancer	•	X	X
	Diagnosis Generic Cancer Type Term Synonyms			
21	Upper gastrointestinal cancer	•	X	X
22	cancer of digestive organs	•	X	X
	From Search Topic Refinements – specific information types			
23	Alcohol	•	X	X
24	Alternative medicine	•	•	•
25	Alternative therapy	•	X	X
26	Causes	•	•	•
27	Charity	•	X	X
28	Clinical trials	•	>	•
29	Consultant	•	X	X
30	Diagnoses	•	X	X
31	Diet	•	X	X
32	Drugs	•	X	X
33	Emotion management	•	X	X
34	Emotional health	•	X	X
35	Family risk	•	X	X
36	Financial aid	•	X	X
37	Financial help	•	X	X
38	Health organisation	•	X	X
39	Information centre	•	X	X
40	Health organization	•	X	X
41	Information center	•	X	X

42	Insurance	•	X	X
43	Likelihood of cure	•	X	X
44	Mental health	•	X	X
45	Risk factors	•	•	•
46	Selfcare	•	X	X
47	Self-care	•	X	X
48	Sexual health	•	X	X
49	Side effects	•	X	X
50	Smoking	•	X	X
51	Stage of disease	•	X	X
52	Support	•	X	X
53	Support groups	•	X	X
54	Symptoms	•	•	•
55	Tests	•	•	•
56	Treatment	•	•	•
57	Treatment options	•	X	X
58	For patients	X	•	>
59	For Professionals	X	>	•
60	From medical authorities	X	•	•
61	Patient handouts	X	>	•
62	Continuing Educations	X	>	•
63	Practice guidelines	X	>	•

• Offered
X Not offered
> Indirectly Offered (Not meant for but accessible)

Figure 8.19: Comparison of Suggested Search Topics and Refinements offered by PerIS and Google for ISCO Patient "00f6cm" on "stomach cancer"

Figure 8.20 aggregates the number of search topics suggested by PerIS and Google in terms of the search topic focusing techniques. Google does not support personal search topics or search topic enrichment. Thus, a patient using PerIS can utilise seven personal search topics from patient medical information and fifteen search topics from diagnosis information enrichment.

However, both Google and PerIS offer search topic refinement. Google search topic refinements are displayed when a user enters a health condition (e.g. "stomach cancer") possibly in medical language. For instance, using the search term "womb cancer" on Google does not display search refinement topics presented when using the corresponding medical term "uterus cancer". This may indicate that Google supports search refinements only for medical search terms but not lay search terms. This constitutes a limitation in Google medical search as patients are more likely to use lay terms (e.g. womb cancer rather than uterus cancer). However, PerIS's search topic refinements are offered to a patient regardless of search term topic or vocabulary.

Search Topic Focusing Technique	PerIS	Google Patients	Google Professionals
Personal Search Topics	7	0	0
Search Topic Enrichment	15	0	0
Search Topic Refinement	35	8	12
Total:	**57**	**8**	**12**

Figure 8.20: Breakdown of the Search Topics offered by PerIS and Google based on Search Topic Focusing Technique

Figure 8.21 depicts the overlap in search term refinement topics suggested by PerIS and Google. PerIS search term refinement topics are identified from our investigation into topics often sought by patients through interviews with information specialists and a literature survey of patient information needs and patient Internet access. Amid the 35 PerIS search topic refinements, Google supports six for Patient Search and seven for Professionals Search. Google lists the refinement topic "clinical trials" under "For Professionals" while it can also be useful for patients, especially highly educated patients. This topic is already included in PerIS search term refinement topics.

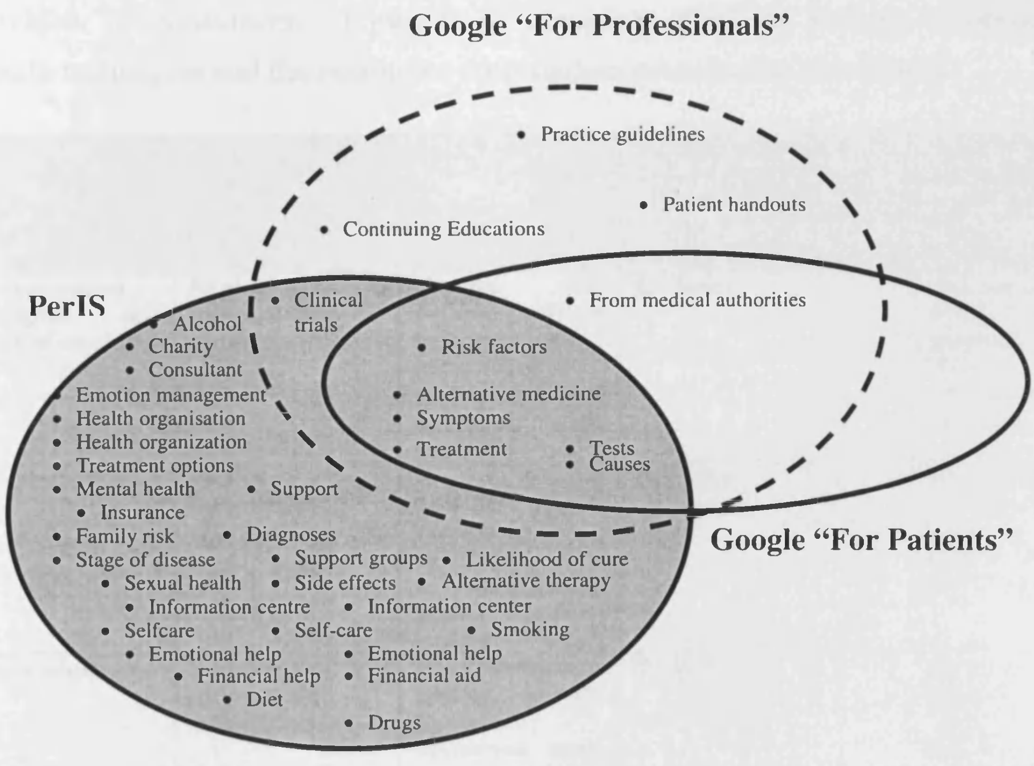


Figure 8.21: Overlap in Search Term Topic Refinements between PerIS and Google

8.5.2.5 Evaluating PerIS Search Tool Focusing Technique

PerIS Search tools are designed to serve two aims:

a) Guiding a patient to quality Internet information sources. This is evaluated in Section 8.5.2.5.1.

b) Allowing a patient to focus an Internet search to a given search tool (see Section 8.5.2.5.2).

8.5.2.5.1 Evaluating PerIS Approach to Guiding Patients to Quality Internet Information

Internet health information quality is a challenge due to the unverified and unstable nature of the Internet. As patients surf the Web independently, educating patients about verified health websites empowers patients to locate trustworthy Internet information. On the other hand, healthcare providers demand Internet health information verification to safeguard patients against unsafe Web information. Classically, Internet information quality is addressed using several techniques as discussed in Section 2.4.3. Such tools are usually aimed at Internet information providers or consumers. Figure 8.22 compares different Internet information quality techniques and the reason for their inclusion/exclusion from PerIS.

Internet Information Quality Verification Technique	Actor	Action	Included/Excluded From PerIS	Reason For Inclusion/ Exclusion From PerIS
Recommended principles and codes of conducts	Health information provider	Ensures that website content follows recommended principles and codes of conducts	Excluded	Not applicable to PerIS. PerIS is a gateway rather than a health information provider.
Self-evaluation	Health information provider	A website provider applies explicit evaluation tool to mark/certify website content quality	Excluded	Not applicable to PerIS as PerIS is a search system rather than a health website.
User-evaluation	Internet information consumer/user	The Internet user applies explicit Internet evaluation tools to mark/certify website content quality	Excluded	Stressful to patients as a patient is required to learn evaluation tools (e.g. Net Scoring) and use them for every site they access.

Third-party evaluation	Independent third-party accreditation organisation	Third-party evaluation organisation verifies website content against their accreditation program principles and grant conforming websites an accreditation seal (e.g. HONCode, URAC, Truste)	Included	Offers access to independently (non-official health) evaluated Internet health information. PerIS links to HONCode search engine for patient search and URAC and Truste accredited health websites for building hospital-trusted websites.
Metadata-based and semantic Web evaluation	Health website provider	Health website providers annotate their Web pages with a mark-up (metadata) indicating the website information quality. This aids machine automatic identification and utilisation of quality mark-up.	Excluded	Not applicable to PerIS. Outside scope of study. Also, it is difficult to annotate the entire Web as this task adds more workload to Internet information providers to annotate their pages with quality mark-up, a task not all Internet information providers welcome and it is self-evaluation.

Figure 8.22: Comparison of Internet Information Quality Techniques

Patients as Internet information consumers may have difficulty applying user evaluation tools (e.g. Checklist or Net Scoring). User evaluation requires training and considerable time and effort which can be stressful for some patient. PerIS is not designed as an Internet information evaluation tool or a provider. Instead, PerIS represents a patient-oriented gateway to existing trusted and relevant Internet information. It is designed to assist a patient access trusted and relevant Internet information in a simplified and personalised fashion.

PerIS addresses Internet information quality by educating and guiding patients to evaluated health websites. In addition, PerIS offers a patient a direct individual search to evaluated health websites. PerIS utilises two typical Internet information evaluation techniques:

a) Self-evaluated health websites (e.g. national health gateways and medical search engines).

b) Third-party accredited websites (e.g. HONCode).

Furthermore, PerIS introduces three new Internet health information quality notions/marks:

a) Hospital-Trusted websites that are customised to a patient condition.
b) Patient-Trusted (or preferred) websites.
c) Charity (non-official) health websites.

Hospital-trusted websites and Patient-trusted websites enable the incorporation of professional and patient perceptions of Internet information quality, respectively. Charity health websites, on the other hand, utilise specialist charity websites perceptions on Internet information quality to offer a patient a non-official patient- or lay-oriented perspective of Internet health information quality.

8.5.2.5.2 Evaluating PerIS Approach to Focusing Search Tool

The type of Internet search domain sought by a patient could be influenced by five factors:

- **Website quality level:** Self-evaluated websites, third-party accredited websites, hospital-trusted websites, user (e.g. Patient) trusted websites, charity websites or open Web access.
- **Website vocabulary type:** Medical (professionally-oriented) versus lay (patient-oriented)
- **Website domain capacity:** Specific single website, multiple websites, national health gateway, open Web or UK only websites for UK users.
- **Customisation:** Websites relevant to, or of interest to the patient.
- **Website language:** Websites written in a language understood and preferred by the user.

Accordingly, PerIS search tools are designed to reflect the above factors. Thus, by selecting appropriate PerIS search tools, a patient is able to meet his/her search requirement. The aim is to aid a patient to make an informed decision about the underlying search tool characteristics. PerIS offers a patient seven search tool categories:

- Health Gateways

- Charity Websites
- Your Velindre Recommended Websites
- Your Favorites
- Google
- Velindre Google
- Specific Website Search

The following sections evaluate our approach to focusing Internet search tool based on the above five outlined factors.

8.5.2.5.2.1 Focusing PerIS by Search Tool Information Quality

Typically, a patient gets no indication of search result quality when utilising common search engines (e.g. Google). Users need to check the search result credentials (e.g. website provider, update date). Google's search result rank, however, indicates the significance of the linked document in terms of its access popularity but not its quality. Thus, Google search results can not always be trusted. Health gateways, on the other hand, represent evaluated health websites and, hence, ensure health information quality of their search results.

PerIS is designed to link a patient to key and trusted health information sources. PerIS informs a patient about website quality level through search tool name or category. A brief search tool description is given when a patient points at a search tool (Figure 8.23). A more detailed search tool description is offered by the PHB system Help webpage by clicking on the tool name.

PerIS aids a patient to search verified information through numerous search tool categories: Health gateways, Charity Websites and Your Velindre Recommended Websites. "Health Gateways" offers a patient a list of accredited, national, medical and speciality health gateways. Hence, selecting any of these search tools ensures patient access to trusted and verified information.

In addition, PerIS allows a patient to define and search his/her own trusted health websites through two techniques: Your Favorites and Search Specific Websites. Furthermore, PerIS offers a patient open unverified Web search using two tools: Google and Velindre Google. Google offers a patient direct access to the Google search engine whereas Velindre Google represents a customised Google search engine manipulated by PerIS. VelindreGoogle enables a patient to set a preferred search language for the Google search from within the PerIS system. In addition, it allows a patient to save search results to the Favorite Websites in the PHB system.

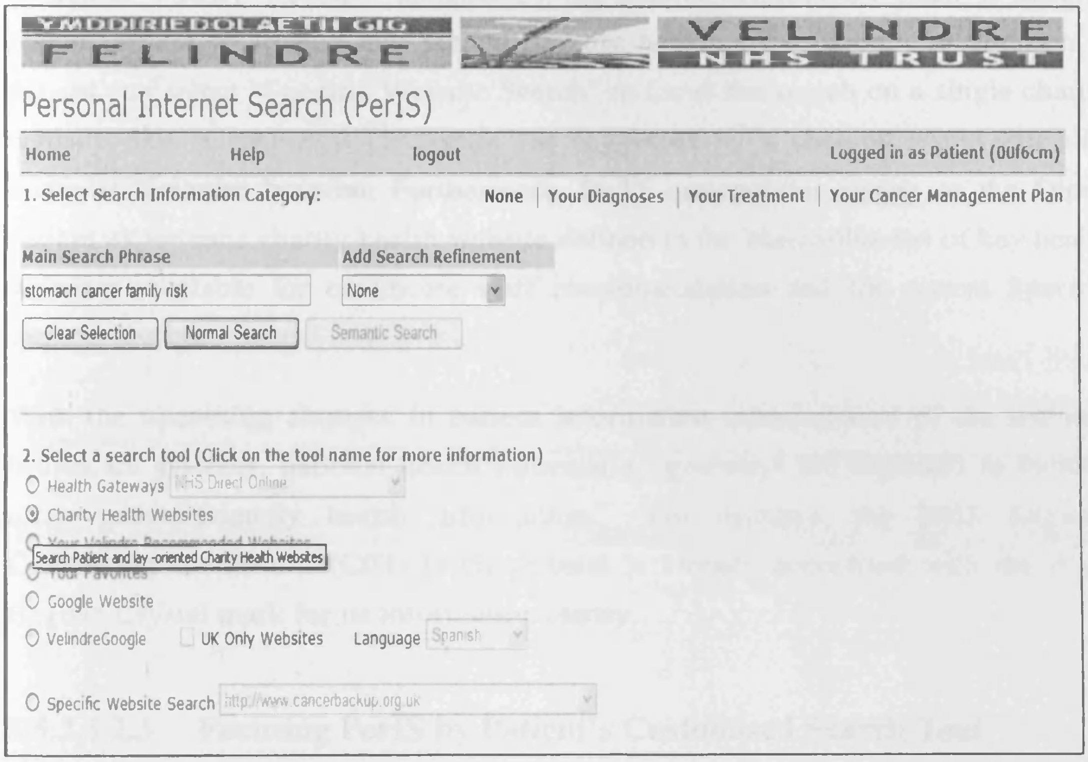


Figure 8.23: Search Tool Description Appears as Pointing at "Your Velindre Recommended Websites" Search Tool

8.5.2.5.2.2 Focusing PerIS by Search Tool Information Vocabulary Type

In classical Web searching, Internet health information is not categorised based on vocabulary level. Health information could be written in medical or lay terminologies. Medical terminology can usually be found in websites written for or by professionals such as national health gateways and medical search engines. However, lay terminology is typically used by charity websites, patient websites, lay websites and websites adhering to the Plain English Campaign (PEC) [120] regulations. Investigating the PEC website, few patient information websites are currently accredited for PEC. In addition, patient and lay health websites, though offering patient-oriented information based on experience, can be unverified, and either self or third-party verified for quality.

PerIS has a "Health Gateways" search tool to enable a patient to search health information written mostly in scientific and medical vocabulary. In addition, PerIS defines a new search tool type: "Charity Health Websites" to enable a patient to focus the search only on trusted health information websites written in a patient-friendly (or lay) language. The "Charity Health Websites" search tool groups a list of cancer charity websites updated by the patient information staff. It ensures patient access to trusted and patient-friendly health information. In addition, a patient can select "Specific Website Search" to focus the search on a single charity website like "CancerBACUP.org.uk" as suggested by a chemotherapy specialist nurse at Velindre hospital. Furthermore, PerIS incorporates access to the Dipex patient experience charity health website defined in the Macmillan list of key health websites available for healthcare staff recommendation and for patient Specific website search.

With the upcoming changes in patient information infrastructure of the national healthcare systems, national health information gateways are expected to include more patient-friendly health information. For instance, the NHS England Connecting for Health (CfH) [105] website is already accredited with the Plain English Crystal mark for its information clarity.

8.5.2.5.2.3 Focusing PerIS by Patient's Customised Search Tool

Generic Web Search tools and key health gateways usually offer one set of sources to all users. PerIS, on the other hand, offers two patient customised search tools:

- Hospital Recommended Websites (Your Velindre Recommended Websites) that is customised to a patient condition, and
- Patient Preferred Websites (i.e. Your Favorites) that are of interest to a patient.

8.5.2.5.2.4 Focusing PerIS by Search Tool Domain Capacity

Typically, generic search engines search the entire Web, whereas medical health gateways search selected evaluated medical websites. PerIS gives a patient the choice over both Internet search tools. In addition, PerIS allows a patient to focus the search on a single or multiple websites preferred by a patient, and multiple websites recommended by the hospital. Figure 8.24 illustrates the features of the PerIS search tools with respect to factors characterising patient Internet searching.

Search Engine	**Generic Web Search Engine (e.g. Google)**	**Patient-Preferred Websites**	**Charity Health Websites (e.g. Cancerbackup.org.uk, cancerhelp.org.uk)**	**National Health Gateways (e.g. NHS Direct, MedlinePlus)**	**Medical Gateways (e.g. MedHunt**	**Third-Party Accredited Search Engine/Websites (e.g. HON,**	**Hospital-Trusted Websites**
Category Type	General-purpose	Patient-preferred	Charity (i.e. Unofficial) and Patient-Oriented	National	Medical	Accredited	Hospital-Trusted and Patient-Customised
Quality Level	Not Indicated	Patient's Trusted	Verified	Verified	Verified	Verified	Verified and Recommended
Authentication System / Method	PageRank as in Google and Yahoo	Patient Judgment	Self-Accreditation	Self-Accreditation + Charity Expert Knowledge (e.g. NHS Direct)	Self-Accreditation	Third-Party Accreditation	Healthcare Professionals Evaluation and Recommendation
Vocabulary Type	Mixed	Patient-Choice	Mostly Lay and Patient Oriented	Usually mixed but recently moving towards Lay vocabulary	Medical and Professional-Oriented	Mostly medical	Should be lay and patient-oriented
Domain capacity	Web	Patient's Selected Websites	Collection of Charity Websites	Specific National Gateway	Specific Medical Gateway	Specific Medical Gateway	Multiple Websites
Customisation	Untailored	Patient Self-tailored	Untailored	Untailored	Untailored	Untailored	Patient-Tailored
Internet Information language	Patient-preference	Patient-preference	English	English	English	English	English

Figure 8.24: Comparison of PerIS Search Tools in Terms of Factors influencing a Patient Internet Search

8.5.2.6 Evaluating PerIS Search Mode Focusing

Traditional Web search engines (e.g. Google, Yahoo) usually operate in a single keyword search mode searching only for Web documents containing keywords included in the search term. With the advent of the Semantic Web (SW) [141], this traditional keyword search is extended to utilise semantic data related to the search term concept (s). Hence, semantic search applications [147, 223, 238] emerged to enrich traditional keyword search results with information based on the semantic knowledge encoded by the SW conceptual structure (e.g. ontology, thesaurus)

Typically, semantic search utilises a generic non user-tailored SW knowledge model. In such a case, the search application uses all SW knowledge related to the concept (s) indicated by the search term. This approach might extend the traditional search with information that is irrelevant or not of interest to a patient, as it may lack focus.

In contrast, in this study, the notion of the SW is used to build a patient-tailored diagnosis ontology to enrich the medical knowledge and search results using the patient diagnosis. Our Patient Diagnosis Ontology (PDO) is designed to include only information (objects and relationships) that are relevant to the diagnosis concept and patient health information requirements.

Furthermore, we have drawn our diagnosis concept objects from the EPR system and its underlying medical classification system to ensure the maximum diagnosis object term relevance to the patient diagnosis terminology. The PDO design incorporates seven patient health information requirements:

1. Patients mostly seek information on health conditions.
2. Patients need to express the correct form of a medical term when searching for health information.
3. Patients need to use the correct lay term corresponding to a medical term.
4. Patients need to distinguish between similar terms and specific/generic terms.
5. Patients' lay terminology may produce misleading or partial search results.
6. Patients demand healthcare professionals support on accessing relevant information.
7. Patients have variable information vocabulary requirements (medical versus lay vocabulary).

PerIS operates in two search modes:

- **Normal search:** similar to traditional keyword search but improved by personalised search topics and numerous generic and specialised search tools.
- **Semantic Search:** works in the Diagnosis Category search. It augments traditional search results on the patient diagnosis term with search results based on relevant terms from the PDO. A patient is offered additional medical, lay and generic terms based on the given diagnosis.

In addition, our semantic search approach ensures the following additional requirements:

8. Semantic search should utilise only relevant SW knowledge in order to retrieve relevant semantic search results.

Figure 8.25 illustrates how the above patient health information requirements are supported by PerIS and compares them to traditional keyword Web search and generic semantic search.

Patient Health Information Vocabulary Requirement	Traditional Search	Generic Semantic Search	PerIS Semantic Search
Patients mostly seek information on health conditions (i.e. diagnosis)	Not Supported	Patients either enter his/her diagnosis information or choose his/her diagnosis information from an entire diagnosis ontology or classification system (e.g. MeSH) which could be long, confusing and result in erroneous selection.	Patients can search using Diagnosis category that presents a patient with patient's own diagnosis information as recorded in EPR.
Patients need to express the correct form of a medical term when searching for health information.	Not Supported	A patient may enter diagnosis term incorrectly or have difficulty choosing the correct medical term denoting their diagnosis.	A patient diagnosis is presented to a patient in its medical form as recorded in the patient database. In addition, patient gets similar medical terms (synonyms) as recorded in the ISCO keyv2 table or generated by PDO.
Patients need to use the correct lay term corresponding to a medical term.	Not Supported	Not Supported. Generic medical semantic knowledge is usually aimed at clinicians and represented in a classification or ontology system that uses medical terms. It does not address user lay terms information requirements.	The Patient Diagnosis Ontology (PDO) utilised by PerIS Semantic Search associates lay descriptions to medical term diagnosis that are computed from a Concept Thesaurus defining medical to lay term mappings. Thus, patient can utilise their diagnosis in online search in either medical or lay terms. An information specialist verifies diagnosis lay terms.
Patients need to distinguish between similar terms and specific/generic terms.	Not Supported	Can offer patient similar and hierarchical terms defined by the underlying SW system.	Offers patient similar and generic terms constructed from the patient database terminology system and by an information specialist. This ensures a patient has a more correct, focused and controlled terminology set.
Patients' lay terminology may produce misleading or partial search	Not Supported	Not Supported	PDO utilises patient information staff specialist knowledge to define

results.			correct mapping between a medical and lay term concepts. In addition, information staff verifies that PDO contains correct terms computed from patient DB and the auxiliary CT. There are several lay term synonyms and a patient can utilise preferable lay terms or search by all lay terms.
Patients seek healthcare professionals support on accessing relevant information	Not Supported	Not Supported	By utilising the patient database in producing patient-personalised search topics (ideas) and Involving information staff to define and verify the correct medical, lay, and generic term mappings.
Patients have variant information vocabulary requirements (medical versus lay vocabulary)	Not Supported	Not Supported	By further customising semantic search to execute only a single or combinations of semantic data categories (e.g. medical, lay). This allows patient to select their preferred vocabulary categories.
Semantic search utilises only relevant semantic data	Not Applicable	Utilises all related generic semantic knowledge even if not relevant or of interest to the patient condition or current information requirement.	PDO is designed according to patient information requirements and EPR medical knowledge. As PerIS links PDO to EPRs, PerIS employs only semantic data that are relevant to patient personal medical information (i.e. diagnosis) and current search information requirement.

Figure 8.25: Comparison of PerIS Semantic Search Support for Patient Information Requirements to Traditional and Generic Semantic Search

8.5.3 Evaluating PerIS Search Results against Traditional Patient Search Results

This section evaluates PerIS by comparing PerIS search results to traditional generic Internet search results. PerIS can improve the traditional patient Internet search by using its underlying capabilities, namely:

a. Patient personalised search topics,
b. Search term enrichment,
c. Search tool focusing, and
d. Combined effect of the above techniques.

PerIS's patient-personalised search topics offer search terms from the patient personal medical details. This leads to more relevant search results that are focused to the patient's health condition.

PerIS uses a semantic enrichment technique to generate new search terms relating to a patient's diagnosis information. Due to the time constraints on this study, semantic enrichment is applied only to search terms representing patient diagnoses but can be easily extended to other search terms.

The main search term is enriched with semantic data from our semantic knowledge model – the Patient Diagnosis Ontology (PDO) built in this study. In addition, PerIS enables a patient to utilise semantic data at variable granularity (e.g. lay, medical, generic or combinations).

Furthermore, the PerIS approach to semantic enrichment utilises a patient-tailored semantic knowledge model, which is designed according to patient information requirements and allows patient choice. This ensures that search results based on semantic data are relevant and of interest to a patient. PerIS search tools can improve the traditional patient Internet search by enabling focusing of the search on certain websites or domains featuring information quality, vocabulary, language and patient-customised information sources.

Comparing PerIS search results to Google search results based on patient-personalised search topics is impractical as Google does not support such a capability. In addition, patients specify their own sought search queries that might be erroneous, ineffective in terms of its terms, or fail to utilise various patient personal medical information. Hence, PerIS patient-personalised search topics, based on EPRs, are highly likely to produce more focused results than normal search queries entered by patients.

Hence, in this section, the PerIS search results are evaluated against normal Google search results by evaluating the effects of search term semantic enrichment and PerIS search tool focusing techniques on search results. This evaluation is run for a patient diagnosis search term ("stomach cancer") to utilize its semantic data. Three assumptions are used during this evaluation:

a. Patients might use lay terms, possibly "cancer of stomach" or "stomach cancer".
b. Patients are more likely to use medical terms used by the medical community or patient database (e.g. "malignant neoplasm of stomach").
c. An average patient usually checks the first ten Google search results returned for a single search query.

PerIS is evaluated against a normal Google search by measuring overlaps in search results between potential PerIS search result sets and normal Google search result sets. PerIS can produce multiple search results according to the utilised semantic data, search modes and its variable search tools. Zero overlap indicates that PerIS produces search results not produced by Google. PerIS searches are more likely to produce useful search results as they utilise semantic information designed according to patient information needs (e.g. medical, lay and generic terms). In addition, PerIS semantic data are described using the correct term forms as defined in the patient database or by patient information staff. This evaluation is carried out over three steps:

1. Computing three sets of normal Google search results for three potential search terms that could be used by patients, namely:
    - malignant neoplasm of stomach,
    - cancer of stomach, and
    - stomach cancer.
2. Computing different PerIS Search result sets for all semantic data defined for the "stomach cancer" diagnosis concept using different PerIS search modes (e.g. normal, semantic), namely:

    Obtaining PDO semantic data on "stomach cancer".

Executing a normal search query for each semantic data term on each PerIS search tool.

Executing semantic search on semantic data categories.

3. Computing overlaps between each PerIS search result set and the main Google search result set. This operation is repeated for the three main Google searches on the medical term "malignant neoplasm of stomach" and its two lay search terms. This allows measuring and evaluation of the PerIS capabilities impact on a patient medical or a lay traditional Google search.

Google traditional search results used in this evaluation can be computed using the Google Website or the PerIS Google API-based search tool (VelindreGoogle). The two sets of Google search results are relatively close with an overlap of 70-80% between the first ten search results (see Figure 8.26).

Search Term	**Overlap between Google Website and Google API**
Malignant neoplasm of stomach	80%
Cancer of stomach	70%
Stomach cancer	70%

Figure 8.26: Overlap between Google Website and Google API (VelindreGoogle) Search Results

Hence, throughout this evaluation, VelindreGoogle is used in order to facilitate automatic comparison of Google search results with internal PerIS search results. Sections 8.5.3.1 and 8.5.3.2 discuss overlaps between PerIS and Google search results for medical and lay term searches respectively. Section 8.5.3.3 evaluates the Search Tool Focusing technique effects on the normal Google Search.

8.5.3.1 Comparing PerIS Search to Google Medical Term Search

Patients seeking health information on "stomach cancer" might use lay or medical terms depending on education level and the sought Internet information requirement. Highly educated patients may seek health information from medical literature [184]. However, a patient may have difficulty using or typing the correct medical term for a given health condition. Generally, we assume that patients might use medical terms used by the patient medical community or the patient database (e.g. "malignant neoplasm of stomach" is used by the ISCO/CaNISC database for "stomach cancer"). Figure 8.27 illustrates the first ten Google search results for "malignant neoplasm of stomach".

Rank	Search Result	Overlap With
1	Basic Summary for **Stomach** cancer - WrongDiagnosis.com http://www.wrongdiagnosis.com/s/stomach_cancer/basics.htm	HONCode
2	**MALIGNANT-NEOPLASM-OF-STOMACH** http://www.ugr.es/~oncoterm/csdata/MALIGNANT-NEOPLASM-OF-STOMACH.html	
3	Statement of Principles **MALIGNANT NEOPLASM** OF THE **STOMACH** http://www.rma.gov.au/SOP/03/007.pdf	
4	[Double primary **malignant neoplasm** of renal cell carcinoma and ... http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=4083218&dopt=Abstract	HONCode
5	**neoplasm** – Encyclopedia.com http://www.encyclopedia.com/doc/1E1-neoplasm.html	
6	Gastrointestinal Carcinoid Tumors http://www.icongrouponline.com/health/Gastrointestinal_Carcinoid_Tumors.html	
7	ICD-10: Block C15-C26 http://www.who.int/classifications/apps/icd/icd10online/gc15.htm	
8	Medical Review Guidelines Magnetic Resonance Imaging – Abdomen ... http://www.ohca.state.ok.us/provider/updates/pdflib/MRG_MRI_Abdomen.pdf	
9	Table 2 http://www.paho.org/English/DD/AIS/HSA2006_Table2.pdf	
10	International Classification of Diseases, Revision 8 (1965): List D http://www.wolfbane.com/icd/icd8d.htm	

Figure 8.27: First 10 Google Search Results for the Term "malignant neoplasm of stomach"

Google search results as appearing in Figure 8.27 are not very useful. This could be due to:

- The medical term "malignant neoplasm of stomach" is not used much in online medical and scientific literature or it is not the popular medical term for "stomach cancer".
- There is not much medical and scientific literature concerning "stomach cancer".

Hence, a patient is challenged to identify alternative medical terms for stomach cancer. PerIS, on the other hand, simplifies this situation for a patient by presenting a patient with numerous semantic data terms relating to the diagnosis search term. In addition, a patient can explore different semantic data categories and perform a partial or full semantic search using different PerIS search tools. Figure 8.28 explores the overlaps between various potential PerIS search results and the results of the traditional Google medical search for "malignant neoplasm of stomach".

Diagnosis Term Semantic Data	**Google**	**HONCode**	**MedHunt**	**MedlinePlus**	**NHSDirect**	**Cancerbackup**	**Cancerhelp**	**Velindre**	**Favorites**	**Charity**
gastric neoplasm	0/10	0/10	0/10	0/10	0/10	0	0	0 /20	0 /13	0
gastric tumour	0 /10	0/10	0/10	0/10	0/10	0 /3	0 /10	0 /36	0 /25	0 /13
stomach tumour	0 /10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /60	0 /54	0 /20
stomach tumor	0 /10	0/10	0/10	0/10	0/10	0 /2	0 /2	0 /35	0 /35	0 /4
gastric carcinoma	0 /10	0/10	0/10	0/10	0/10	0 /1	0 /3	0 /24	0 /18	0 /4
carcinoma of stomach	0/10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /50	0 /43	0 /20
malignant tumor of stomach	0 /10	0/10	0/10	0/10	0/10	0 /5	0	0/31	0 /31	0 /5
malignant tumour of stomach	0 /10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /48	0 /40	0 /20
gastric tumor	0 /10	0/10	0/10	0/10	0/10	0 /1	0	0 /22	0 /14	0 /1
malignant neoplasm of stomach	**10/10**	**2/10**	0/10	0/10	0/10	0	0 /1	0 /19	0 /12	0 /1
stomach neoplasm	0 /10	0/10	0/10	0/10	0/10	0	0 /1	0 /22	0 /14	0 /1
cancer of stomach	0 /10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /70	0 /72	0 /20
stomach cancer	0/10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /70	0 /68	0 /20
cancer upper gastrointestinal	0 /10	0/10	0/10	0/10	0/10	0 /10	0 /5	0 /40	0 /37	0 /15
cancer of digestive organs	0 /10	0/10	0/10	0/10	0/10	0 /10	0 /10	0 /46	0 /35	0 /20
Combined Scientific	**10/98**	NA	NA	NA	NA	0 /29	0 /32	0 /226	0 /193	0 /61
Combined Lay	0 /14	NA	NA	NA	NA	0 /10	0 /11	0 /85	0 /87	0 /21
Combined Generic	0 /20	NA	NA	NA	NA	0 /19	0 /14	0 /81	0 /71	0 /33

X/Y

X: Overlap between Google "malignant neoplasm of stomach" search results and this query search results

Y: Total number of this query search results explored in this comparison

NA: Not Applicable

Figure 8.28: Overlap between Google "malignant neoplasm of stomach" Search Results and PerIS Search Results

Findings:

- PerIS enriches traditional Google search with potential medical search terms.
- There is almost no overlap between the various PerIS search results and Google search results for the search term concept "malignant neoplasm of stomach". Hence, PerIS offers new search results that can be further focused using particular PerIS semantic data or search tools.
- A PerIS medical only semantic search can augment a traditional Google medical search with new search results based on all the medical semantic terms. Thus, PerIS semantic enrichment can augment a traditional Google medical search and lead to more related search results based on similar medical terms.
- Additionally, PerIS offers a patient the option to explore lay and generic information on the sought medical term. PerIS lay and generic search terms show no overlaps with the Google medical search for "malignant neoplasm of stomach". This feature informs a patient about more ways of locating information relating to the main medical search term.

8.5.3.2 Comparing PerIS Search to Google Lay Term Search

Patient-oriented health information is usually written in a clear lay language to facilitate information readability and patient understanding. On the other hand, average patients seeking online information usually use lay terminology. This could occur for two reasons:

- A patient can not express the medical term and instead resorts to lay terms (e.g. "cancer of stomach" or "stomach cancer").
- A patient particularly wishes to access patient-oriented literature described in lay (or patient-friendly) terminology.

This section, evaluates a PerIS search against a Google lay term search. We explore the overlaps between PerIS search results and Google search results for both lay terms that might be used by patients on "stomach cancer". Figures 8.29 and 8.30 present the first ten Google search results for "cancer of stomach" and "stomach cancer" search terms, respectively.

Rank	Search Result	Overlap with
1	**Stomach cancer** http://www.netdoctor.co.uk/diseases/facts/stomachcancer.htm	HONCode
2	**Stomach cancer** information centre : Cancerbackup http://www.cancerbackup.org.uk/Cancertype/Stomach	Cancerbackup
3	**Stomach** (gastric) **cancer** http://www.cancerhelp.org.uk/help/default.asp?page=3887	Cancerhelp
4	**Stomach Cancer** http://www.emedicinehealth.com/stomach_cancer/article_em.htm	HONCode
5	**Stomach cancer** - Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Stomach_cancer	
6	Gastric **Cancer** (**Stomach** Cancer), The **Cancer** Information Network http://www.cancerlinksusa.com/stomach/index.asp	
7	**Stomach cancer** – MayoClinic.com http://www.mayoclinic.com/health/stomach-cancer/DS00301	HONCode
8	ACS :: What Is **Stomach Cancer**? http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_is_stomach_cancer_40.asp	HONCode, MP
9	**Stomach Cancer** - causes, symptoms, diagnosis and treatment options ... http://www.medicinenet.com/stomach_cancer/article.htm	HONCode
10	**Stomach Cancer** on Yahoo! Health http://health.yahoo.com/topic/stomachcancer	

MP: MedlinePlus

Figure 8.29: First 10 Google Search Results for Term "cancer of stomach"

Rank	Search Result	Overlap with
1	**Stomach Cancer** - causes, symptoms, diagnosis and treatment options ... http://www.medicinenet.com/stomach_cancer/article.htm	HONCode
2	**Stomach cancer** – Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Stomach_cancer	
3	**Stomach cancer** – MayoClinic.com http://www.mayoclinic.com/health/stomach-cancer/DS00301	HONCode
4	MedlinePlus: **Stomach Cancer** http://www.nlm.nih.gov/medlineplus/stomachcancer.html	HONCode, MP
5	**Stomach Cancer** http://www.emedicinehealth.com/stomach_cancer/article_em.htm	HONCode
6	**Stomach cancer** information centre : Cancerbackup http://www.cancerbackup.org.uk/Cancertype/Stomach	Cancerbackup
7	What You Need To Know About? **Stomach Cancer** - National Cancer ... http://www.cancer.gov/cancertopics/wyntk/stomach	MH, MP
8	Gastric **Cancer** Treatment - National **Cancer** Institute http://www.cancer.gov/cancertopics/pdq/treatment/gastric/patient/	MH, MP
9	**Stomach cancer** http://www.netdoctor.co.uk/diseases/facts/stomachcancer.htm	HONCode
10	**Stomach Cancer**, Gastric Cancer - Overview – oncologychannel http://www.oncologychannel.com/gastriccancer/	HONCode

MP: MedlinePlus, MH: MedHunt

Figure 8.30: First 10 Google Search Results for term "stomach cancer"

Figures 8.31 and 8.32 illustrate overlaps between the PerIS search result sets and Google search result set for the lay terms "cancer of stomach" and "stomach cancer" respectively.

Findings:

- Generally, PerIS search results overlap with Google search results for both lay terms which indicates that these terms are common in online literature. However, PerIS can extend a single Google lay term search by utilising all alternative lay terms. Additionally, PerIS enables a patient to enrich a lay search with medical and generic information.
- No overlap is observed between the PerIS generic term search and Google lay term search. Thus, PerIS offers a patient search results not addressed by Google.
- Google search results overlap with PerIS search results for certain PerIS semantic data and search tools. This indicates that PerIS has the potential to improve the search by focusing the search based on the underlying capabilities. PerIS offers a patient more control over what semantic data to explore on what Internet information domain (or search engine).
- The particular Google lay term search results (Figures 8.29 and 8.30) come from different health gateways incorporated in PerIS. This demonstrates the genuineness of health gateways utilised by PerIS. Thus, focusing the search on a certain health gateway can give faster patient access to significant information.
- Maximum overlap of 5-6 occurs between the Google and HONCode search engine. This indicates that the HONCode search engine could be a potential patient Internet search tool. Thus, PerIS search tools can offer a patient more focused and significant search results than those generated by Google.

Section 8.5.3.3 explores the effectiveness of the PerIS search tool focusing technique in allowing a patient to retrieve significant search results.

Diagnosis Term Semantic Data	Google	HONCode	MedHunt	MedlinePlus	NHSDirect	Cancerbackup	Cancerhelp	Velindre	Favorites	Charity
gastric neoplasm	0/10	0/10	0 /10	**1/10**	0/10	0/0	0 /0	0 /20	0 /13	0 /0
gastric tumour	0/10	0/10	0/10	**1/10**	0 /10	0 /3	0 /10	**1 /36**	0 /25	0 /13
stomach tumour	**2/10**	**2/10**	0/10	**1/10**	0 /10	0 /10	0 /10	**1 /60**	0 /54	0 /20
stomach tumor	**1/10**	**1/10**	0/10	**1/10**	0 /10	0 /2	0 /2	**1 /35**	0 /35	0 /4
gastric carcinoma	**1/10**	0/10	0/10	**1/10**	0/10	0 /1	0 /3	0 /24	0 /18	0 /4
carcinoma of stomach	0/10	0/10	0/10	**1/10**	0/10	0 /10	0 /10	0 /50	0 /43	0 /20
malignant tumor of stomach	**2/10**	**1/10**	0/10	**1/10**	0/10	0 /5	0 /0	**1 /31**	0 /31	0 /5
malignant tumour of stomach	**2/10**	**1/10**	0/10	**1/10**	0/10	0 /10	0 /10	**1 /48**	0 /40	0 /20
gastric tumor	0/10	0/10	0/10	**1/10**	0/10	0 /1	0 /0	**1 /22**	0 /14	0 /1
malignant neoplasm of stomach	0/10	0/10	0/10	**1/10**	0/10	0 /0	0 /1	**0 /19**	0 /12	0 /1
stomach neoplasm	**1/10**	0/10	0/10	**1/10**	0/10	0 /0	0 /1	0 /22	0 /14	0 /1
cancer of stomach	**10/10**	**5/10**	0/10	**1/10**	0/10	**1 /10**	**1 /10**	**3 /70**	**1 /72**	**2 /20**
stomach cancer	**6 /10**	**4/10**	0/10	**1/10**	0/10	**1 /10**	**1 /10**	**3 /70**	**1 /68**	**2 /20**
cancer upper gastrointestinal	0 /10	0/10	0/10	**1/10**	0/10	0 /10	0 /5	**1 /40**	0 /37	0 /15
cancer of digestive organs	0 /10	0/10	0/10	**1/10**	0/10	0 /10	0 /10	**1 /46**	0 /35	0 /20
Combined Scientific	**3 /98**	NA	NA	NA	NA	0 /29	0 /32	**1 /226**	0 /193	0 /61
Combined Lay	**10/14**	NA	NA	NA	NA	**1 /10**	**1 /11**	**3 /85**	**1 /87**	**2 /21**
Combined Generic	0 /20	NA	NA	NA	NA	0 /19	0 /14	**1 /81**	0 /71	0 /33

X/Y

X: Overlap between Google "cancer of stomach" search results and this query search results

Y: Total number of this tool search results explored in this comparison

NA: Not Applicable

Figure 8.31: Overlap between Google "cancer of stomach" Search Results and PerIS Search Results

Diagnosis Term Semantic Data	Google	HONCode	MedHunt	MedlinePlus	NHSDirect	Cancerbackup	Cancerhelp	Velindre	Favorites	Charity
gastric neoplasm	0 /10	0/10	**1/10**	**2/10**	0/10	0 /0	0 /0	0 /20	0/13	0 /0
gastric tumour	0 /10	0/10	0/10	**2/10**	0/10	0 /3	0 /10	0 /36	0 /25	0 /13
stomach tumour	**2 /10**	**2/10**	0/10	**1/10**	0/10	0 /10	0 /10	0 /60	0 /54	0 /20
stomach tumor	**1 /10**	**1/10**	**1/10**	**2/10**	0/10	0 /2	0 /2	0 /35	0 /35	0 /4
gastric carcinoma	**1 /10**	0/10	**1/10**	**2/10**	0/10	0 /1	0 /3	0 /24	0 /18	0 /4
carcinoma of stomach	0 /10	0/10	0/10	**2/10**	0/10	0 /10	0 /10	0 /50	0 /43	0 /20
malignant tumor of stomach	**2 /10**	**1/10**	0/10	**2/10**	0/10	0 /5	0 /0	0 /31	0 /31	0 /5
malignant tumour of stomach	**2 /10**	**1/10**	0/10	**2/10**	0/10	0 /10	0 /10	0 /48	0 /40	0 /20
gastric tumor	0 /10	0/10	**1/10**	**2/10**	0/10	0 /1	0 /0	0 /22	0 /14	0 /1
malignant neoplasm of stomach	0 /10	0/10	**1/10**	**1/10**	0/10	0 /0	0 /1	0 /19	0 /12	0 /1
stomach neoplasm	**1 /10**	0/10	0/10	**1/10**	0/10	0 /0	0 /1	0 /22	0 /14	0 /1
cancer of stomach	**6 /10**	**5/10**	1/10	2/10	0/10	**1 /10**	0 /10	**1 /70**	**1 /72**	**1 /20**
stomach cancer	**10/10**	**6/10**	1/10	2/10	0/10	**1 /10**	0 /10	**1/70**	**1 /68**	**1 /20**
cancer upper gastrointestinal	0 /10	0/10	0/10	**2/10**	0/10	0 /10	0 /5	0 /40	0 /37	0 /15
cancer of digestive organs	0 /10	0/10	0/10	**2/10**	0/10	0 /10	0 /10	0 /46	0 /35	0 /20
Combined Scientific	**3 /98**	NA	NA	NA	NA	0 /29	0 /32	0 /226	0 /193	0 /61
Combined Lay	**10/14**	NA	NA	NA	NA	**1 /10**	0 /11	**1 /86**	**1 /87**	**1 /21**
Combined Generic	0 /20	NA	NA	NA	NA	0 /19	0 /14	0 /81	0 /71	0 /33

X/Y

X: Overlap between Google "stomach cancer" Search Results and This Query Search Results

Y: Total Number of This Tool Search Results Explored in This Comparison

NA: Not Applicable

Figure 8.32: Overlap between Google "stomach cancer" Search Results and PerIS Search Results

8.5.3.3 Evaluating Impact of PerIS Search Tool Focusing on Search Results Significance

This section explores the impact of the PerIS search tool focusing approach on the significance of search results obtained by a patient by investigating the ranks of the PerIS search tools' results among Google search results. Figure 8.33 summarises overlaps between Google and various PerIS search tools for the three search terms "malignant neoplasm of stomach", "cancer of stomach" and "stomach cancer".

Search Tool	malignant neoplasm of stomach	cancer of stomach	stomach cancer
Google	*10/10*	*10/10*	*10/10*
HONCode	*2/10*	*5/10*	*6/10*
MedHunt	0/10	0/10	*1/10*
MedlinePlus	0/10	*1/10*	*2/10*
NHSDirect	0/10	0/10	0/10
Cancerbackup	0	*1 /10*	*1 /10*
Cancerhelp	0 /1	*1 /10*	0 /10
Velindre	0 /19	*3 /70*	*1/70*
Favorites	0 /12	*1 /72*	*1 /68*
Charity	0 /1	*2 /20*	*1 /20*

X/Y

X: Overlap between Google Search and Given PerIS Search Tool

Y: Total Number of Given Tool Search Results Explored in this Comparison

Figure 8.33: Overlap between Google Search Results and Given PerIS Search Tool Search Results

In Figure 8.33, a high overlap of 5-6 out of ten search results occurs between Google and HONCode search results largely with lay search terms. Minimal overlap occurs between Google and MedlinePlus, MedHunt, Cancerbackup. This indicates the significance of such key health information sources, as Google usually assigns higher rank to highly referenced documents. However, not all Google search results come from valid or authenticated information sources. In contrast, PerIS offers a patient the option to focus the search on authenticated and patient-customised Internet information sources. Thus, PerIS enables a patient to have faster access to significant verified and relevant information.

This section evaluates the PerIS search effectiveness in retrieving significant information by investigating the PerIS search results' ranks among Google search results. The aim is to verify that the PerIS search tool focusing approach is more effective than Google in retrieving significant search results. Four PerIS search tools are explored in this evaluation:

- **HONCode Search Engine:** Searches authenticated health information sources accredited with HONCode Internet health information quality seal. Additionally, it highly overlaps with Google.
- **MedHunt Search Engine:** Medical information search engine offering access to medical and scientific literature.
- **CancerBackup Website:** Europe-leading patient-oriented cancer charity health website.
- **NHS Direct Gateway:** UK National health information Service.

8.5.3.3.1 Investigating HONCode Search Results in Google

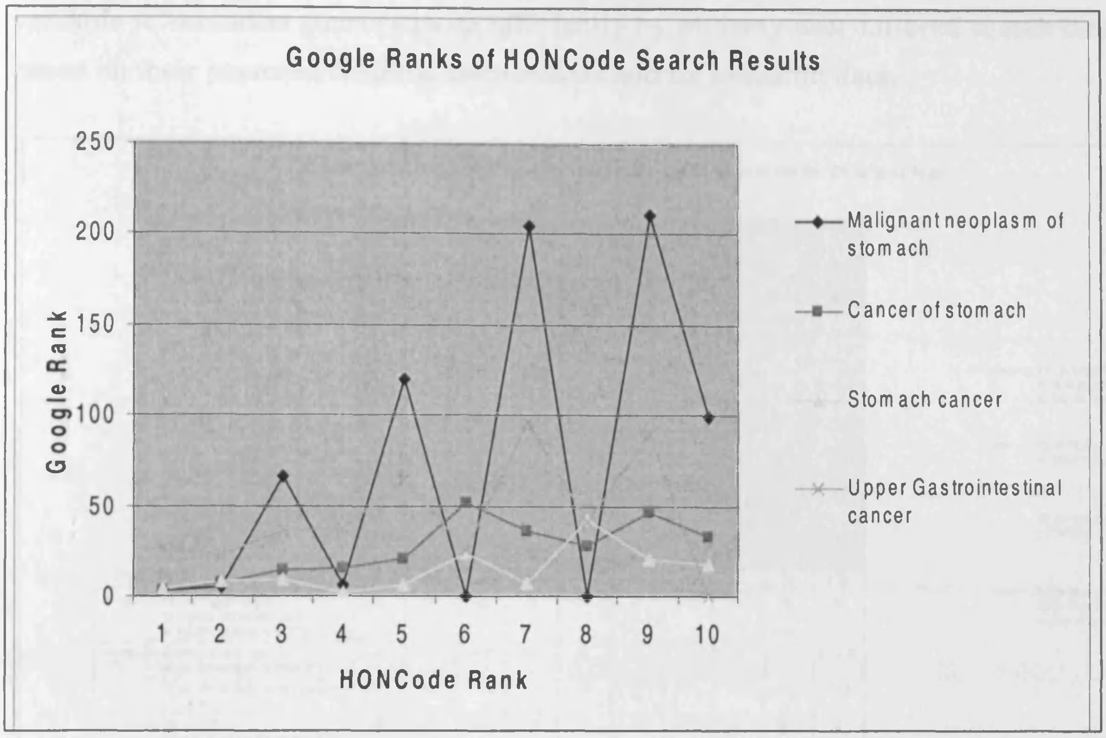
The first ten HONCode search results (Figure 8.34 and 8.35) span a wide Google search results range (e.g. 3 – 205 for "malignant neoplasm of stomach"). This can be ineffective in a medical search as patients might have less patience to inspect distant Google search results. Thus, Google medical search could be ineffective with some medical search terms.

Search results based on lay search terms span lower Google rank ranges than those based on medical or generic search terms. This certainly depends on search term popularity. Popular or highly used medical or lay terms might hit higher Google ranks. Some HONCode search results are missed by Google (see "-" value in Figure 8.34).

HONCode Search Result Rank	Malignant neoplasm of stomach (370)	Cancer of stomach (751)	Stomach cancer (740)	Upper Gastrointestinal cancer (552)
1	3	3	4	26
2	5	7	8	23
3	66	14	9	36
4	6	15	2	37
5	120	21	6	64
6	-	52	24	26
7	205	36	7	96
8	-	28	42	41
9	211	47	21	90
10	99	33	18	-

Google Rank -: Search Result not available in Google
(x): Google Total Search Results

Figure 8.34: Google Ranks of First Ten HONCode Search Results



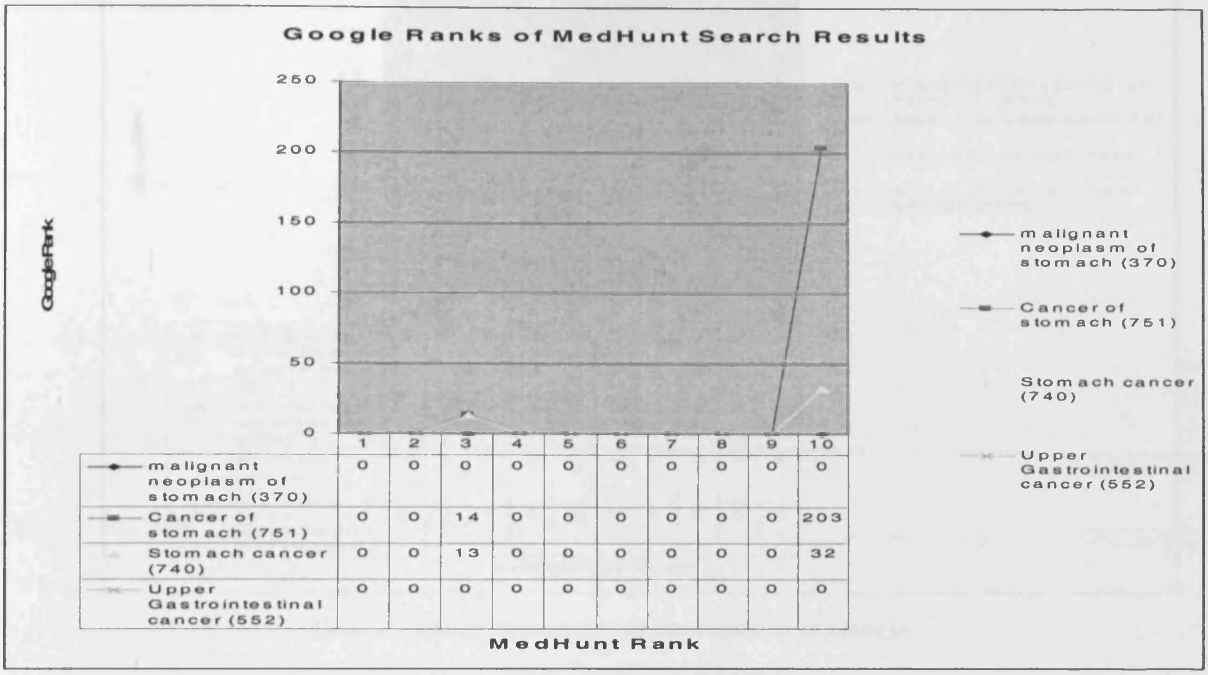
Google Rank 0: Search Result not available in Google

Figure 8.35: Google Ranks of First Ten HONCode Search Results

8.5.3.3.2 MedHunt, CancerBackup and HSDirect Search Results in Google

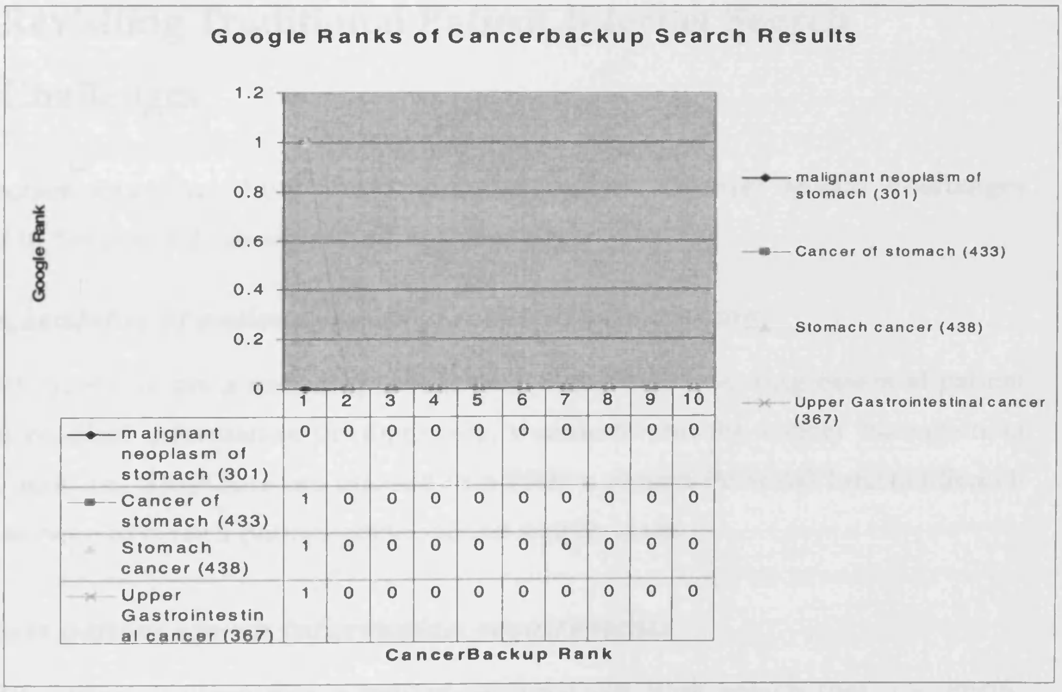
Figures 8.36, 8.37 and 8.38 show the Google Ranks of the first ten MedHunt, Cancerbackup and NHSDirect Search Results respectively. The first ten Google search results scarcely overlap with MedHunt (Figure 8.36), Cancerbackup (Figure 8.37) and NHSDirect (Figure 8.38) first ten search results. Furthermore, the two overlaps between Google and MedHunt search results occur over a very dispersed range. Hence, using a Web generic search engine, a patient has small chance of accessing useful information as identified by the investigation of health information sources.

This experiment clearly verifies the effectiveness of the PerIS search tool focusing technique which allows a patient to focus the search on a particular Web domain or search engine, thereby yielding more focused and significant search results. In addition, a patient gets a sense of information quality and authenticity unlike the case with the generic Google search. Furthermore, by integrating key health gateways in a patient-personalised Internet search system, patients can query such valuable information sources more efficiently by utilising user-tailored search terms based on their personal medical information and its semantic data.



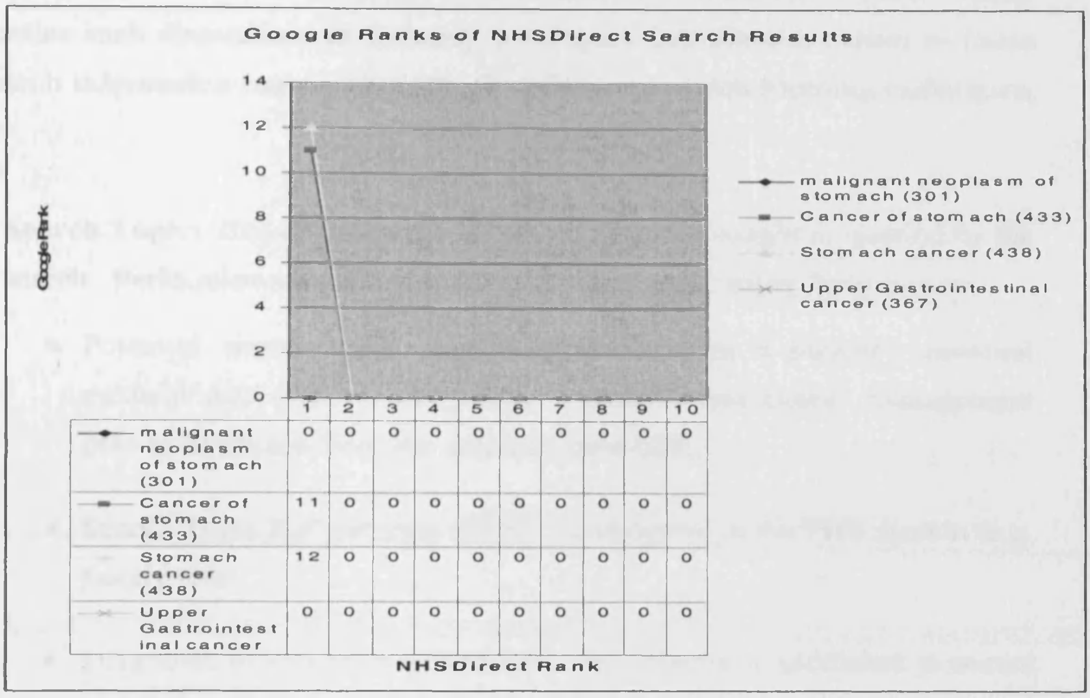
Google Rank 0: Search Result not available in Google
(x): Google Total Search Results

Figure 8.36: Google Ranks of First Ten MedHunt Search Results



Google Rank 0: Search Result Not available in Google
(x): Google Total Search Results

Figure 8.37: UK Only Google Ranks of First Ten Cancerbackup Search Results



Google Rank 0: Search Result not available in Google

(x): Google Total Search Results

Figure 8.38: UK Only Google Ranks of First Ten NHSDirect Search Results

8.6 Revisiting Traditional Patient Internet Search Challenges

This section examines how the traditional patient Internet search challenges outlined in Section 8.2 are addressed in this study:

1. Inaccessibility of patient personal medical information

The PHB system offers a patient online access to an SMR covering essential patient personal medical information on diagnosis, treatment and the cancer management plan. In addition, SMR data are utilised in a PHB's patient Personal Internet Search (PerIS) service to offer a patient personalised search ideas.

2. Variant patient search information requirements

The PHB system incorporates a patient-personalised Web search tool, i.e. PerIS, that is geared to support patient medical search requirements. Typically, a patient's search information requirements relate to several search issues (e.g. search information type, search domain, information vocabulary). Hence, PerIS incorporates such dimensions as focusing techniques that allow a patient to focus their search information requirement. PerIS defines six search focusing techniques, namely:

a. **Search Topic:** this specifies the health information sought or queried by the search. PerIS allows a patient to focus a search topic using three means:

    - Potential search topic ideas formulated from a patient's personal medical data such as diagnosis, treatment and cancer management plan as extracted from the patient's own EPR.

    - Search Topic Refinements (STR) incorporated in the PHB system (e.g. family risk).

    - Diagnosis search term enrichment that generates additional potential diagnosis-related search terms (e.g. synonyms, hierarchies).

b. **Search Topic Vocabulary:** Patient health information could be described using different but related vocabulary. In addition, in this study, we highlight variable patient information needs regarding health information vocabulary as identified by Butters [184]. Hence, a patient-oriented health information conceptual model, covering patient diagnosis concepts, is established that accommodates different patient information needs. PerIS utilises our diagnosis conceptual model, referred to as the Patient Diagnosis Ontology (PDO), for two purposes:

    - To enable a patient to focus on the desired health information vocabulary type (e.g. medical, lay).

    - To enrich search topics and/or search results for a patient.

    Focusing Internet search term vocabulary allows a patient to adjust the reading level and understandability of the Internet information. Thus, a non-specialist patient may utilise lay terms or request lay term search whereas highly-educated or professional patients may utilise the medical terms or medical term search.

c. **Search Tool:** Web health information can be queried by several means including medical search engines, national health gateways (e.g. NHS Direct

Online), accredited search engines (e.g. HONCode), charity health websites (e.g. cancerbackup.org.uk), or generic Web search engines (e.g. Google). Accordingly, PerIS incorporates numerous Web search tool categories that enable a user to focus the search according to the following factors:

- **Website Quality Level:** this covers self-evaluated websites, third-party accredited websites, hospital-trusted websites, patient preferred websites, charity websites or generic Web search.

- **Website Vocabulary Type:** this includes two types:

    - Medical or professional-oriented websites (e.g. medical and national health gateways), and

    - Lay or patient-oriented websites (e.g. charity websites).

- **Website Domain:** this restricts the search to a single website, multiple websites of the entire Web. In addition, it allows UK only websites.

- **Customisation:** restricted to websites relevant to or of interest to the patient condition.

In the current system design, PerIS incorporates six Web search tool categories:

- **Health Gateways:** this covers professional-oriented health gateways, third-party accredited search engines, national health gateways and medical search engines.

- **Charity Websites:** this covers patient-oriented health gateways (e.g. cancerhelp.org).

- **Your Velindre Recommended Websites:** A customised search engine that searches only hospital-trusted websites on a list that is customised to the patient condition.

- **Your Favorites:** A patient's Favorites list of health websites that is constructed by a patient

- **Google:** links to normal Google search.

- **Specific Website Search:** restricts the search to a single website.

d. **Search Domain:** this denotes the number of websites utilised in a given search query. Search domain focusing is useful if a patient wishes to retrieve information from certain websites. PerIS incorporates a mechanism to search the entire Web, search a group of trusted websites or restrict the search to a single website.

e. **Search Language:** this enables a patient to retrieve Web information written in a patient's preferred language.

f. **Search Mode:** this offers a patient the option to conduct a normal Web search based on the specified search query or perform a semantic search that extends the normal search with search results based on the search query's related terminology.

3. Generic Velindre Websites list, utilised by ISCO Patients

The PHB offers a patient a hospital-trusted websites list that is customised to the patient based on a patient's health problems as described in the patient EPR. This list can be accessed by a patient individually or searched using the PerIS's "Your Velindre Recommended Websites" search tool.

4. Laborious, manual and generic nature of patient Internet search

PerIS simplifies and personalises the patient Internet search process. It offers a patient-customised search ideas and patient-customised search tools. Hence, a patient does not have to memorise their medical information or formulate or type them correctly. In addition, it links to a wide range of health Internet search tools which facilitates the search of these tools for a patient.

5. The Wide-ranging and disparate nature of Internet health information search tools

PerIS incorporates and categorises key Internet health information resources and make them available for a patient to utilise in a single Internet search system. The underlying PerIS Internet information resources cater for variant patient information vocabulary and quality Interests.

6. Internet information quality – difficulty identifying trusted Internet information

PerIS incorporates a mechanism that builds a hospital-trusted websites (HTW) list that is customised for individual patients according to their health problems as extracted from the EPR. In addition, PerIS guides and enables patient choice over key Internet health information gateways and customised search engines of variable Internet information quality perception:

- Third-party accredited health information search engine.
- Self-evaluated national health gateways.
- Hospital-trusted health websites search engine.
- Non-official trusted charity health websites search engine.
- Patient-trusted websites search engine.
- Generic Internet search engine.

7. Health information vocabulary – difficulty expressing medical and lay terms and identifying related terms.

The PHB system incorporates four mechanisms to improve a patient's health information vocabulary:

a. Formulating patient-personalised search topics from EPR data to be utilised by PerIS.

b. Building and utilising a Search Topic Refinements (STR) list to enable a patient to further focus or narrow their personalised search topics.

c. Incorporating a generic Concept Thesaurus (CT) through which an information staff member defines medical-to-lay term mappings from both medical and lay perspectives.

d. Building a Patient Diagnosis Ontology (PDO) that encodes related diagnoses terms that are of interest to patients and use EPR medical classification knowledge.

8. Internet information overload – large number of online data sources and/or large size of search result sets.

The following search dimensions can reduce information overload and focus patient Internet search results:

- *Focused Internet search information topic (e.g. diagnosis, treatment) often sought by patients.*

- *Search term refinement (e.g. tumour marker – information types often sought by patients.*

- *Personalisation/Customisation:* offering patient-personalised search topic values from the patient's own EPR data and customised search tools.
- *Conceptual Framework:* can reduce information overload in two ways:
  - Enriching the search with relevant semantic data and, hence, focusing search results only to related search results.
  - Focusing the search and, hence, search results to a specific conceptual category (e.g. medical, lay).
- *Evaluated and Trusted Websites and search Engines:* guides patients to trusted and key health websites. According to Carlson [185], quality health websites can improve information overload, surely, a view shared by healthcare professionals and patient's unwilling to waste time surfing unverified search results.
- *Specific Web Search domains:* Can reduce search results for patients (e.g. searching only breastcancer.org.uk for breast cancer information, or searching Cancerbackup.org.uk for common cancer information)
- *Ease of access:* By offering a single point of access to a myriad of Internet and local information resources, a patient is offered a focused view of the Internet that facilitates access to multiple information resources from a single interface saving the patient, the time and effort required to search disparate Internet information resources separately.

9. Internet information pollution – misinformation, unclear information or irrelevant details.

The notion of Internet Information pollution is described in the literature using different perceptions:

- *Misinformation:* This is addressed by guiding patients to search tools and websites holding trustworthy and evaluated information.
- *Too much and unorganised information (i.e. information overload):* Same as information overload (see previous point).
- *Lengthy sentences and irrelevant details:* This can be tackled by building a customised search engine searching only Plain English Campaign (PEC) accredited websites. However, this technique is not addressed in this study

due to the lack of PEC accredited health information websites. This can be extended in a future work. Nonetheless, PEC accredited health websites can currently be accommodated through the Hospital Trusted websites ("Your Velindre Recommended Websites) customised search engine.

10. Lack of Internet information coordination and sharing between patients and professionals

PHB incorporates two mechanisms that allow hospital staff to build a trusted websites list and a patient to build Favorite Websites within the PHB system. In addition, PHB customises the hospital trusted websites list to patients according to the patient condition. Furthermore, PHB enables staff members to view and select from the patient preferred website when building their trusted websites list. Similarly, it enables a patient to view and select from hospital trusted websites when building a patient Favorites list.

8.7 The Fulfilment of Research Aims

Section 1.6 stated three research aims:

1. Personalising patient Internet information searching based on the patient's own medical information and health information requirements.

2. Simplifying and enriching a patient's medical search information vocabulary by use of a rich personal health information vocabulary utilising clinical data and the underlying data semantics, i.e., terminological relationships (e.g. synonyms, hierarchies).

3. Guiding a patient to quality Internet health information.

The thesis demonstrated the fulfilment of the three aims as follows:

Aim 1: Personalising patient Internet health information searching

The study offered a novel approach to personalising patient Internet health information searching by:

- extracting essential data from the patient's own medical records that are deemed useful for patient education and Internet patient search,

- enriching this with patient-tailored health diagnosis semantic data according to the patient's information needs and with specific search refinements (e.g. family risk) often sought by patients, and

- linking it to Internet information sources. This approach offers the following patient Internet search personalisation capabilities:

- Patient Personalised Internet Search (PerIS) system within a PHR system as opposed to generic Internet search.

- Patient–personalised search topics (or ideas) based on a patient's own medical details and history, data semantic and patient information needs.

- Personal health information vocabulary based on a patient's own EPR to explain and relate medical health information vocabulary relevant to a patient's own diagnosis concepts.

- Personal Web space within a Web-based personal health base (PHB) system, currently used to store their preferred Internet search results and health websites. More personal data could be incorporated in future work (e.g. diary, contacts)

Aim 2: Simplifying and enriching patient medical search information vocabulary

PerIS improves a patient's search information vocabulary by:

- Formulating and offering potential patient-personalised search topics that utilise the correct medical terminology and details as described in EPRs. Hence, a patient can not mistype or specify incorrect medical information in their search query. This simplifies, validates and, hence, improves patient medical information search query formulation.

- Developing and employing a patient-tailored diagnosis conceptual model (i.e. PDO) to explain and enrich the patient's medical diagnosis information with similar medical, lay and generic cancer terms. PDO can enrich health information for a patient at three levels by:
    a. Explaining medical diagnosis information with lay terms and relating it to similar and related medical, lay and generic terms.

b. Enriching search results for a patient using additional diagnosis semantic data.

c. Enabling a patient to focus the semantic search on a preferred semantic terminology (e.g. lay terminology only Internet Search).

Our approach to building a diagnosis conceptual model, in terms of the Patient Diagnosis Ontology, is distinct in three ways:

a. It is patient-tailored, i.e., based on a patient's information needs. This ensures that the diagnosis conceptual model incorporates only relevant and interesting semantic information.

b. It links a given diagnosis concept to both medical and lay terminology to enrich a patient's search with both medical and lay semantic data or to enable the patient to select a preferred semantic data category (e.g. lay versus medical/scientific)

c. It incorporates four semantic knowledge perceptions:
   - *Patient DB medical terminology (or classification system):* to ensure the terminology is familiar to the patient and medical communities and compatible with EPR descriptions.
   - *Patient DBA / developer(s):* to utilise additional existing diagnosis conceptual knowledge and models encoded in EPRs by database developers.
   - *Patient information specialist:* to ensure correct medical to lay diagnosis mappings that embrace a valid lay terminology perspective.
   - *Patients:* by focusing on health information types often sought by a patient, i.e., diagnosis information, and incorporating patients information vocabulary needs to distinguish between medical and lay vocabulary, and related terms.

Aim 3: Guiding patient to quality Internet health information

The patient Personal Internet Search (PerIS) system addresses Internet information quality by:

- Guiding patients to Internet information sources to focus the search on trusted Internet information.

- Incorporating wide-ranging search tools covering variable perception on Internet information quality. This offers a patient more choice over self, official and non-official trusted Internet information. PerIS incorporates three common Internet information quality perceptions:

    a. Third-Party Accredited Internet health information (e.g. HONCode)

    b. Self-evaluated health gateways (e.g. NHSDirect, MedlinePlus)

    c. Generic unverified Internet search (e.g. Google)

  Additional Internet information quality perceptions are introduced with respective search engines:

    d. Hospital-Trusted websites that are customised to a patient's condition.

    e. Patient-Own-Trusted (or preferred) websites.

    f. Charity (non-official) Trusted Health websites.

- Enabling direct search of key health gateways from a single search interface.

8.8 Research Limitations

This study has successfully fulfilled the research aims as discussed in Section 8.7. In addition, our prototype system (i.e. PHB) has demonstrated the undertaken research hypothesis. However, it has the following limitations:

1. **Effects of time constraints on the project:**

    - PerIS uses PDO to enrich the search with diagnosis term synonyms and hierarchies. However, it does not implement a solution to diagnosis term homonyms, i.e., removing search results which use the same diagnosis term but have different meaning. Usually, the medical term homonym problem is less frequent in medical Internet literature than medical term synonym or hierarchal terms. However, this issue can be addressed in future work.

- The PHB prototype was not evaluated with patients. However, this can be conducted in a future work study if it is to be taken forward.

2. Implementation limitations

- PerIS internal search tools and semantic search can be slow in some cases: internal PerIS search tools such as "Your Velindre Recommended Websites" and "Your Favorites" are implemented using Google API and the Google Search website restrict command , i.e. "site:" This requires running many background Google searches for each website defined in the underlying search tool websites list. The same applies to the semantic search options that execute Google searches for each semantic data item utilised by the semantic search. Thus, the execution of such search tools can sometimes be very slow for search queries involving a very large number (e.g. 100s) of background Google searches.

- Very occasionally, the Google-API server goes down and does not execute. However, it executed in subsequent attempts.

CHAPTER 9

Conclusions and Future Work

9.1 Conclusions

This thesis presented a novel approach to personalising patient Internet medical searching that integrates data from a patient's own EPR with relevant Internet information sources. In principle, this research was motivated by a patient's lack of information, inaccessibility of personal medical information, limitations of traditional patient information sources and problems hindering patient Internet medical searching as explored in Chapter 2.

Our approach to personalising patient Internet medical searching is determined by the following considerations:

- The current Internet popularity in public healthcare: this is based on the extensive Internet health information, high Internet health information access, and the advanced Internet technologies especially with regard to security and ease of access.

- The emerging role of the Internet as a central health information delivery platform in the newly developed national health information strategies (e.g. CfH [105], IHC [109]), is marked by information sharing and the patient-centeredness approach to healthcare. Key components of such a dramatic change in official healthcare are a staff accessible integrated EPR [160] that promotes clinical data integrity, consistency, timeliness and sharing, and a patient accessible summary Personal Health Record (PHR) that permits a patient to access essential personal medical information and promotes the delivery of patient-personalised health services using it.

- The growing role of patients as potential and equal partners in their own healthcare. This has been advocated by the Patient-Empowerment movement [195], that sees patients as equal partners in their own healthcare, and capable of handling their own personal health information. This approach has recently been adopted by official patient information strategies [69, 103, 105, 109, 114, 134].

- Both patients' and professionals' demand for simplifying and guiding a patient's access to related and trusted Internet health information.

- The feasibility of extending the PHR framework with a patient-personalised Internet search capability.

Hence, we developed an online patient health information system as a PHR prototype, called the Patient Health Base (PHB) which offers patient-personalised information services including SMR and PerIS. PerIS is the key patient-personalised service addressed in this research. PerIS functionality is supported by two staff interface types: a common staff interface that delivers a staff trusted health websites list, and an information staff interface that manages the construction of a patient-oriented diagnosis information vocabulary and lists of third-party accredited health websites, and other PHB update operations.

The following components are central to PerIS essence and functionality:

- **Personalised Search Topic Constructor (PSTC):** formulates potential search ideas from a patient's own EPR data and related PDO terminology.

- **Patient Diagnosis Ontology (PDO):** constructs a patient-oriented diagnosis vocabulary from both the medical and lay perspectives. This is to ensure a patient has access to valid medical and lay diagnosis terminology. Typically, generic medical encoding systems and medical terminologies do not cover or identify lay terminology. PDO bridges the gap between medical and lay terminology using a generic Concept Thesaurus (CT) facility managed by a patient information staff member. PDO is used in explaining and enriching medical terminologies and executing a fine-grained semantic search operation that distinguishes medical, lay and generic Internet searches.

- **Hospital Trusted Websites (HTW):** offers a hospital-trusted websites list that is customised to an individual patient based on a patient's health condition. An interested hospital staff member builds the individual Staff Trusted Websites (STW) list. As we realised that medical professionals normally have less familiarity with Internet information resources, the STW construction operation is aided by lists of third-party accredited health websites, and the process is managed by hospital information staff.

- **Customised Google Search (CGS):** CGS is central to the execution of the PerIS internal search tools such as Hospital-trusted websites, Favorite Websites, Charity websites, and specific websites. In addition, CGS functionality implements the semantic search options. CGS customises the execution of the Google search engine based on Google API search features.

- **Gateway Wrapper (GW):** establishes the linkage to key health gateways and medical search engines from within the PerIS interface.

Two integration problems addressed by the PHB functionality, were:

1. The construction of a patient-oriented diagnosis vocabulary, i.e. PDO, that integrates the medical and lay diagnosis vocabulary perspectives. We have adopted a tightly-coupled data level federated approach in developing this, whereby a patient information specialist guides the mappings between medical diagnosis terminology extracted from the patient database and lay terminology defined by an information staff member through a Concept Thesaurus (CT) Interface. This is essential in ensuring valid diagnosis medical-to-lay term mappings, as patients are usually unskilled in expressing valid medical and lay terminology.

2. The integration of a patient's own EPR data with relevant Internet information resources. A mediator loosely-coupled data-level integration approach is used to link EPR information with relevant Internet resources. Typically, Internet-based integration is complicated by the large number of online information resources, the inaccessibility of a data source's structure and conceptualisation, data source volatility and the skill level of the patient user. The PHB system is implemented as a middleware layer interfacing between the patient database and different Internet information sources. Access to, and manipulation of a patient's EPR data are undertaken by the PSTC component that formulates valid combinations of potential

patient search topics, while medical and lay terminology challenges are facilitated by the PDO component. Furthermore, an array of potential Internet search tools is incorporated in the system to offer a patient a focused view of the Internet that assists a patient in selecting search tools that match his/her current Internet search requirement.

We have demonstrated the feasibility of building PHB as a PHR prototype connecting to patient database. In addition, we have demonstrated that PerIS improves the search in the following Internet medical search dimensions:

- **Internet search capabilities:** We evaluated PerIS search capabilities (see Figure 8.4) against individual Internet search tools and health gateways utilised in this study. The majority of PerIS capabilities are not supported by any of the individual external Internet search engines incorporated in PerIS. Google offers the maximum (14%) full support of PerIS capabilities whereas HONCode offers the maximum (24%) partial support of PerIS capabilities.

- **Internet Search Focusing Techniques:** PerIS incorporates six search focusing techniques such as search topics, search vocabulary, search tool, search domain, search language and search mode. Google supports to some extent three of these capabilities (search language, search domain and search topic refinement, see Section 8.5.2).

- **Internet search results:** PerIS can improve traditional search results through semantic enrichment and search tool focusing. The thesis demonstrated the improvement in PerIS search results for medical term search, lay term search and individual gateway search:

    - *Improvement in traditional medical term search:* We demonstrated that the Google search results for the medical Read Code term "malignant neoplasm of stomach" were insignificant (see Figure 8.26), which agrees with Westberg's [290] and Abidi [149] findings, about failure of medical encoding systems in retrieving significant search results. In contrast, PerIS's medical search enrichment augments traditional medical search results with medical term synonyms, corresponding lay term synonyms and generic term synonyms which gave the user more relevant results.

- *Improvement in traditional lay term search:* PerIS demonstrated overlap with traditional Google lay term search for the term "stomach cancer". Substantial overlap occurred between PerIS's HONCode search results and Google search results, which demonstrates the potential of the HONCode search engine for patient Internet searching. PerIS extended a traditional lay term search with lay term synonyms, alternative medical terms and variant search tools which improved the output to relevant Web documents and focused websites.

- *Improvement in significant search results:* PerIS's direct linkage to medical and authenticated health gateways (e.g. HONCode, MedHunt) led to more focused and significant search results. We have investigated the ranking of potential PerIS search results among Google search results and demonstrated that the early ten PerIS's search result are more focused than the first ten Google search results (see Section 8.5.3.3). PerIS search results from external health gateways were spread over a very wide search results range in Google search. Thus, PerIS's direct linkage to key health gateways offered faster access to significant search results.

9.2 Latest Developments in PHR technology and Attendant Search Engines

There is a growing interest in personal health records among healthcare organisations and the IT industry. In the UK, PHR programmes are part of NHS healthcare programmes, as can be seen in the NHS England's Summary Care Record (SCR) initiative (see Appendix A.1) and the NHS Wales's Individual Health Record (IHR) (see Appendix A.2). In the USA, recently, the IT industry launched initiatives to facilitate the sharing of a patient's medical data among multiple organisations and patients themselves. This is because the USA, unlike the UK, lacks a national healthcare service that coordinates patient data nationally. Instead, healthcare, in the USA, is delivered by private and disparate healthcare organisations (e.g. hospitals and clinics). This makes it difficult for a patient to combine medical data recorded by different physicians working in diverse organisations. This section highlights latest developments in the area of PHR, and the accompanying search functionalities.

9.2.1 The NHS England Summary Care Record (SCR) and HealthSpace

The SCR (see Appendix A.1) is the NHS England initiative to enable patients and authorised healthcare providers to access patient medical data online. It also enables, at a national level, authorised health organisations to access a patient's record. HealthSpace is the NHS England's website that enables a patient to maintain a PHR online and connect to their GP SCR or eventually their nationally integrated record. Currently, all patients living in England can create a HealthSpace account, where they can record their health information manually. However, connection to a SCR through HealthSpace is currently only available to patients living in areas covered by the Early Adopter Programme [33]. A Health Space account (Figure 9.1) enables a patient to view and/or record demographics and health information (e.g. medications, allergies). HealthSpace incorporates two features whereby a patient can search for and access relevant health information:

- **Find** (Figure 9.2): searches for a given medication specified by the patient. This only searches a database of drugs and medications.

- **Library** (Figure 9.3): allows a patient to create a list of links to useful websites. This resembles the patient Favorite Websites list implemented in this study but our approach offers a patient, mechanisms to establish this list automatically from Hospital trusted Websites, third-party accredited websites and search results from some PerIS search tools.

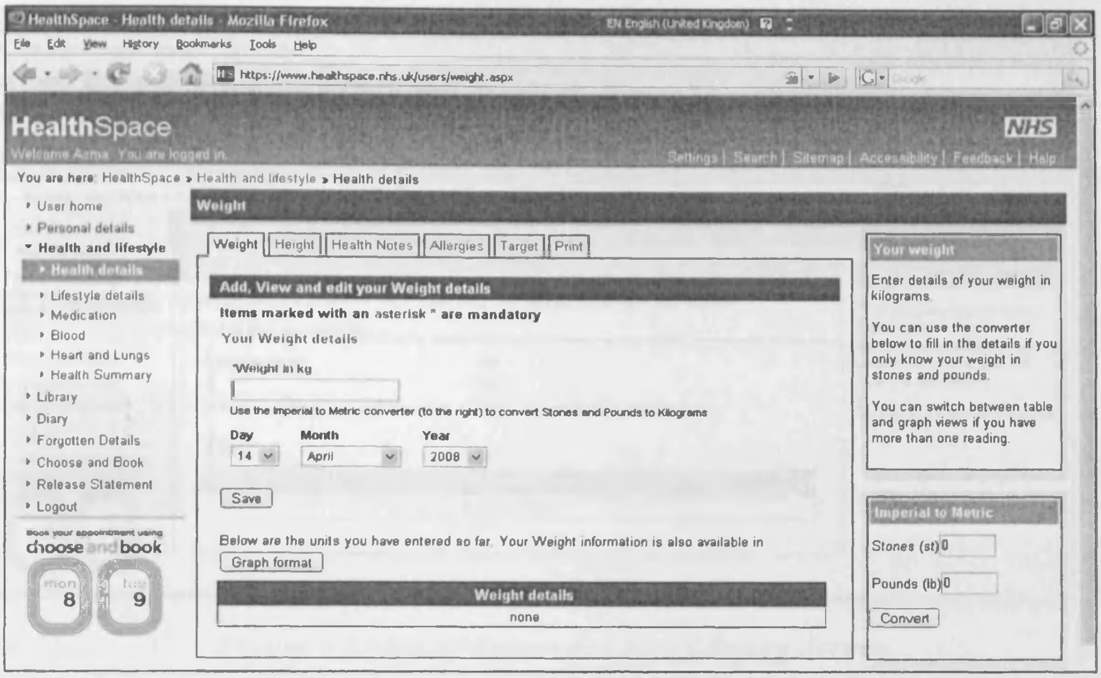


Figure 9.1: HealthSpace Account Features

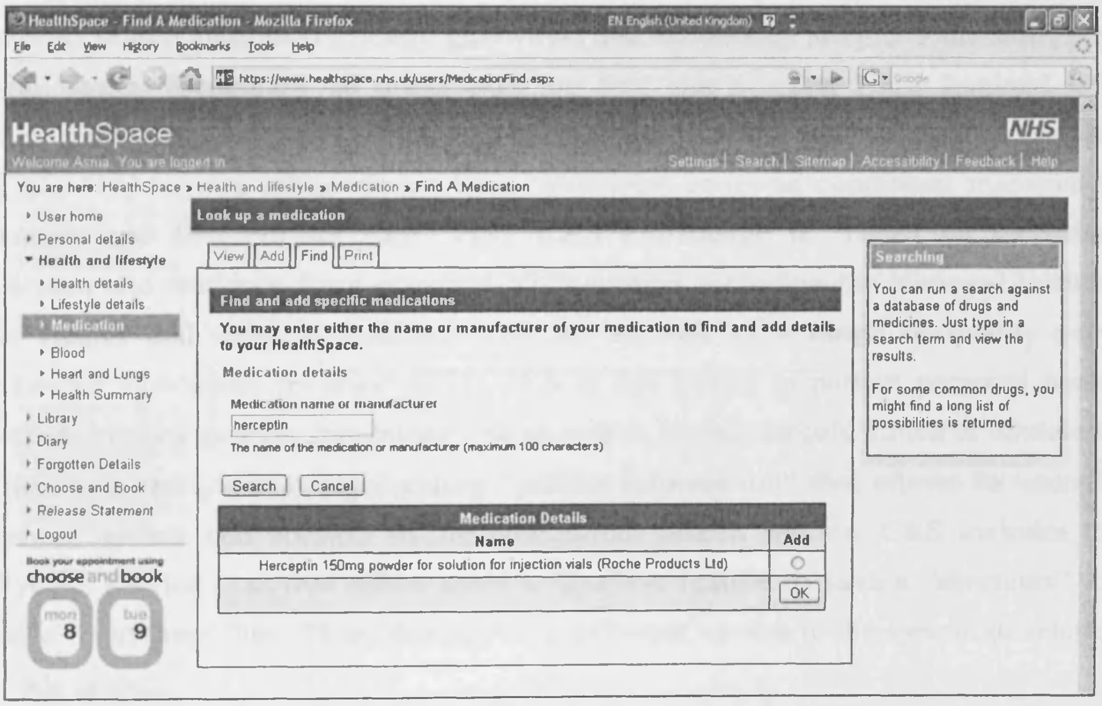


Figure 9.2: HealthSpace Medication Find Screen

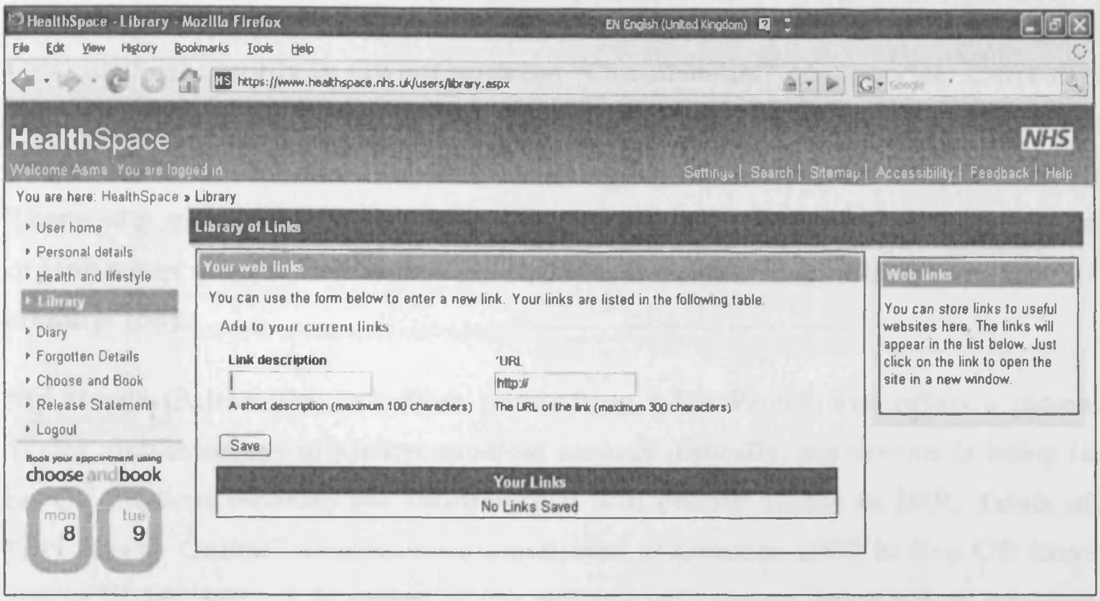


Figure 9.3: HealthSpace Account Library Screen

To the best of our knowledge, the literature shows no reports about patient-personalised Web search engine linking the HealthSpace account (or data) to Internet search engines (or health gateways) and addressing the problems associated with health vocabulary as undertaken by this study. The NHS England CfH programme has recently established a Clinical Knowledge Summaries service (CKS) [18] hosting a knowledge base "about the common conditions managed in primary and first contact care" [18]. CKS knowledge is "based on secondary research and evidence from standard NHS sources including the National Institute for Health and Clinical Evidence (NICE) as well as a range of quality peer-reviewed systematic reviews" [17]. CKS is not linked to patient personal health records but is a generic standalone online search service largely aimed at clinicians. There is a section in it highlighting "patient information" that allows its users to browse leaflets and connect to the NHSDirect Search service. CKS includes the MyCKS service [18] that offers users a Toolbox feature to save a "shortcuts" list and a "read-later" list. Thus, this is a very different system to the system developed in this project.

9.2.2 The NHS Wales Individual Health Record (IHR) and My Health Online

IHR [68] (see Appendix A.2) integrates patient data from different points of care and makes this information available for national healthcare organisations involved in patient care. A pilot was launched in Gwent-Wales, in November 2006, that links GPs' medical records in Gwent with the "Out-of-hours" care service. Currently, 76 out of 96 GP practices in Wales share their medical records with the "Out-of-hours" service, i.e. one in seven people in Wales can use the IHR service [68]. There are plans to extend this trial to share information with other health organisations (e.g. NHS Wales Ambulance Services) and/or to other areas in the country [68].

My Health Online [96] is a Web portal from NHS Wales, that offers a patient, in Wales, online access to his/her medical records. Initially, the service is being tested on GP medical records, but ultimately it will enable access to IHR. Trials of the "My Health Online" service were conducted in October 2007 in five GP surgeries across Wales [96]. A snapshot of the website [97] shows that it offers a patient the following services: update account details, book appointments, order repeated prescriptions, send messages to the GP practice and view medical records. Access to medical records has been tested on selected patients in three of the five practices [96]. We are unable to identify a search service associated with the "My Health Online" initiative from the snapshot, or the literature.

9.2.3 HealthFrame – Records for Living

HealthFrame [54] (Figure 9.4) is a PHR solution from Records for Living. This a a software program that enables a patient to record and manage personal health information including conditions, medications, visits and treatments. It is not directly linked to a patient's official medical record(s). However, a patient can obtain a copy of his/her official medical record(s) from individual health organisations and import it into the HealthFrame PHR account [64]. The problem with this approach is that not all patients update their health information regularly [285] as discussed in Section 2.5.1, Page 43.

HealthFrame incorporates the "Library Reference" search facility (Figure 9.5) that searches pricing and statistical databases and the MedlinePlus gateway. This search facility is generic, i.e. not patient-personalised (so it does not utilise a patient's personal health information stored in the HealthFrame PHR to customise the search features for a patient). The Library search requires a patient to enter a search term which is then mapped to a list of search term matches describing related topics and medical term synonyms based on UMLS and ICD-9 codes. The search only runs a single search match selected by a patient which is not necessary the original search term entered by a patient. No support is given for lay terms. For instance, entering "womb cancer" suggests no matches and zero search results (see Figure 9.6).

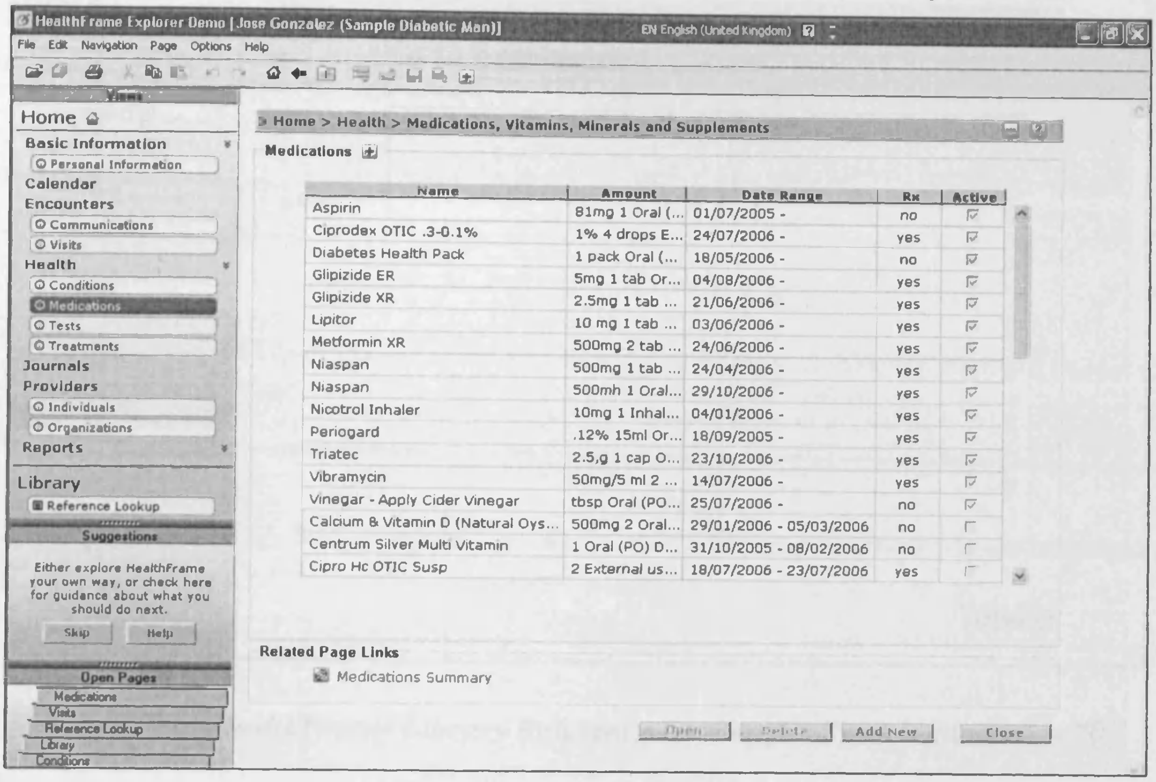


Figure 9.4: HealthFrame Screen –Medication Section

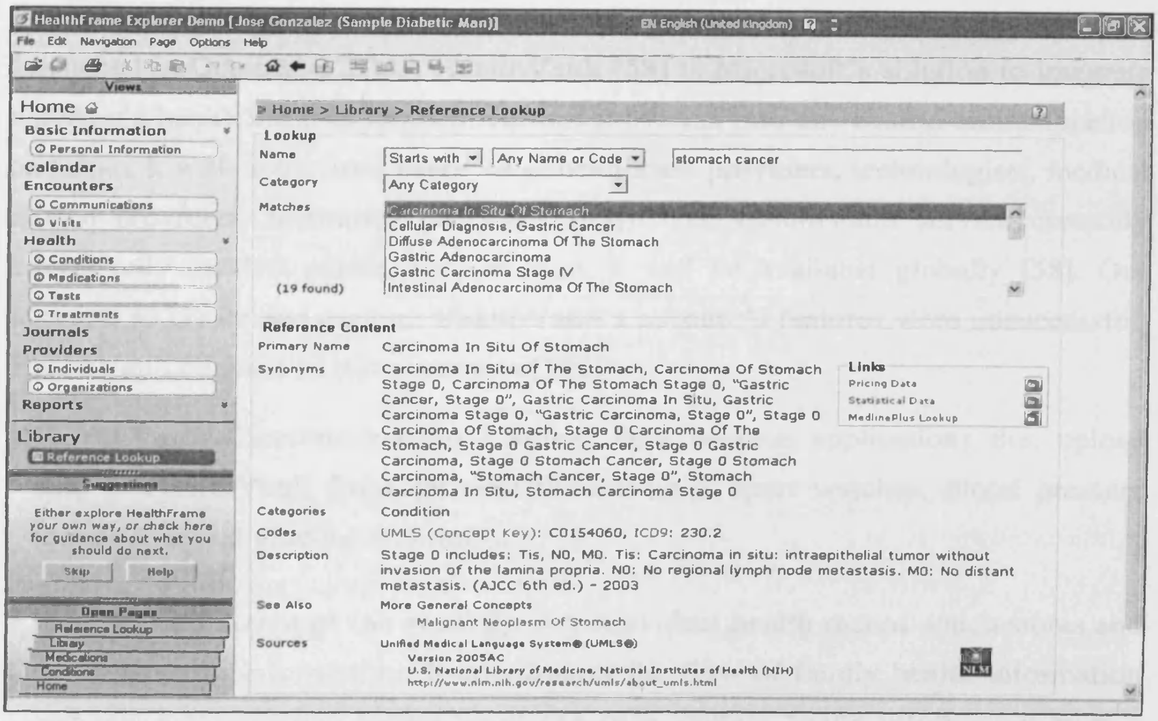


Figure 9.5: HealthFrame Library Reference Lookup

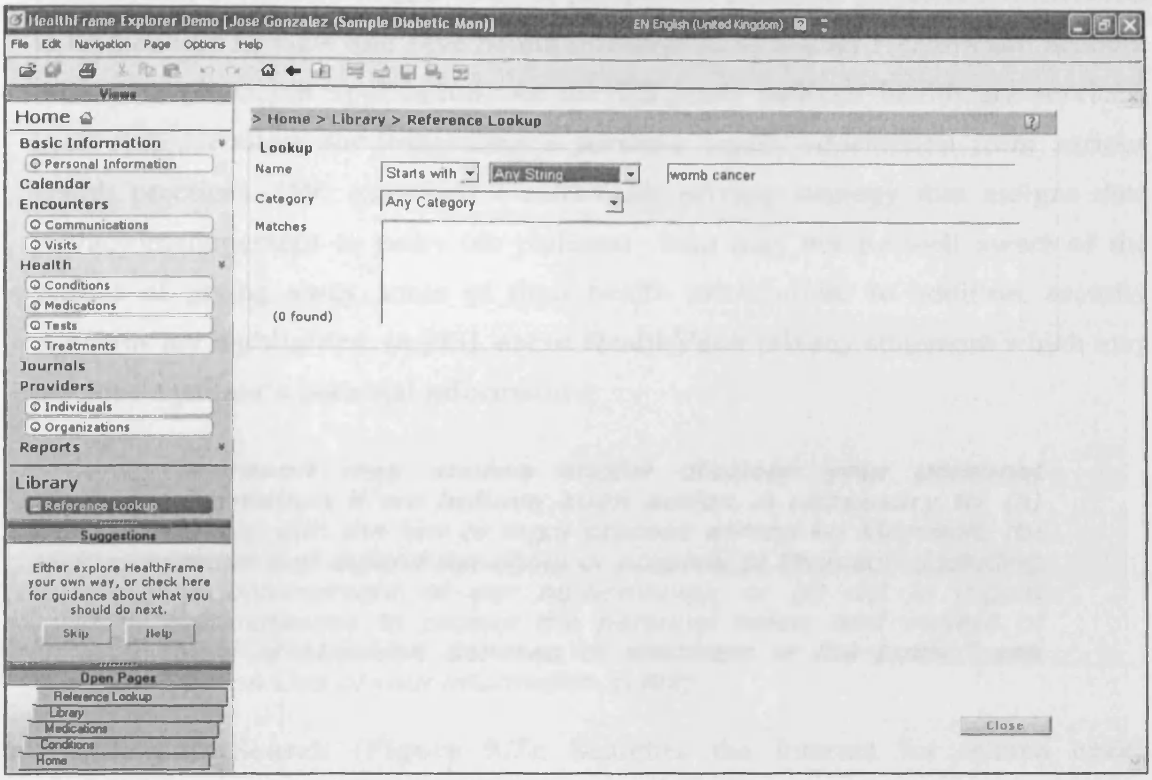


Figure 9.6: HealthFrame Library Reference Lookup for "womb cancer" – "0 Found" Search Results

9.2.4 Microsoft HealthVault

Launched in October 4, 2007, HealthVault [58] is Microsoft's solution to integrate a person's health informationfrom various providers into one central online location and share it with authorised users (e.g. healthcare providers, technologists, medical device providers, insurance providers [93]). The HealthVault service currently covers only the US public, but, in time, it will be available globally [58]. Our attempts to create and explore HealthVault's account's features were unsuccessful. HealthVault consists of three sections [136]:

- **HealthVault Communication Centre:** Free desktop applications that upload data to HealthVault from external devices (e.g. sport watches, blood pressure monitors, blood glucose monitors).
- **HealthVault Account (or record):** An individual health record which stores and updates health information. It coordinates the flow of family health information and sets authorization. HealthVault can only capture health information from a

HealthVault compliant health tool. A person can authorise physician(s) involved in his/her care to view and save health information to his/her HealthVault account through a physician application. As the US lacks national healthcare services, such systems allow for integrating a person's health information from various health practices. [59] questions HealthVault privacy strategy that assigns data privacy management to users (or patients) who may not be well aware of the pitfalls of giving away some of their health information. In addition, security concerns are highlighted, in [93], about HealthVault privacy statement which may disclose a patient's personal information:

> ***"Microsoft may access and/or disclose your personal information if we believe such action is necessary to***: *(a) comply with the law or legal process served on Microsoft; (b) protect and defend the rights or property of Microsoft (including the enforcement of our agreements); or (c) act in urgent circumstances to protect the personal safety and welfare of users of Microsoft services or members of the public." see Section Use of your information in [92]*

- **HealthVault Search (Figure 9.7)**: Searches the Internet for related health information and uses related search refinements for focusing the search.

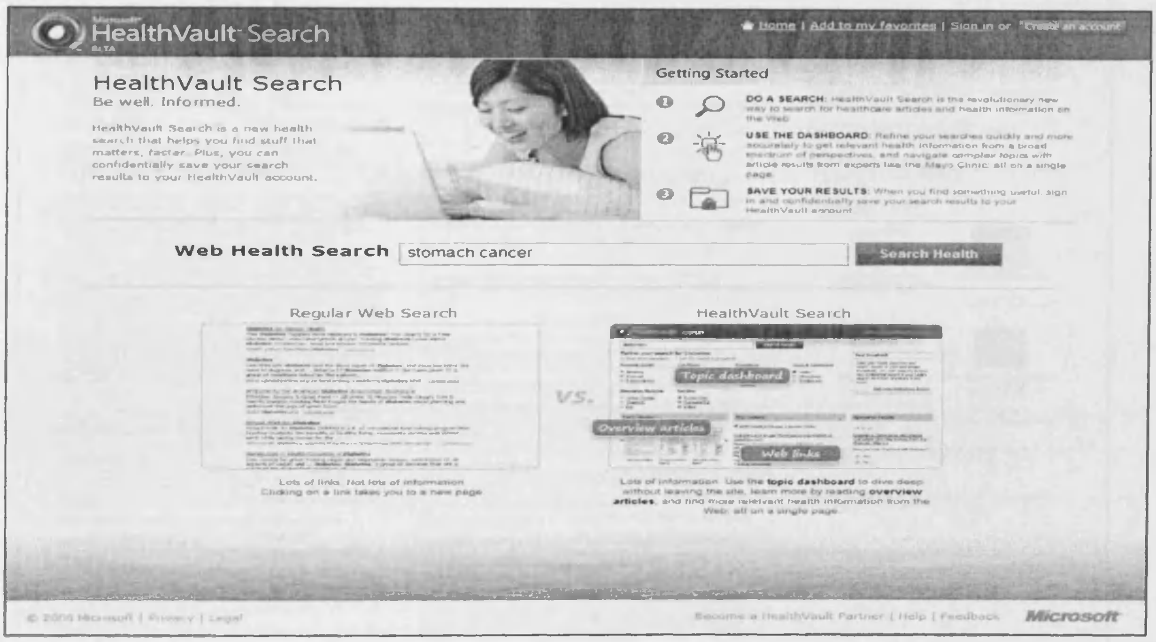


Figure 9.7: Microsoft HealthVault Search Engine

The HealthVault Search website seems generic and does not indicate personalised search features (e.g. personalised search topics, recommended health websites) based on a patient's own requirements or data as stored in the HealthVault record. However, it allows saving search results to a HealthVault record. HealthVault Search suggests some health topics related to the current search term to refine the search (Figure 9.8). This is beneficial for focusing the main search term to specific information types. However, medical terms such as “stomach neoplasm”, “gastric neoplasm” or “upper gastrointestinal cancer” are not included in the search refinements for the search term “stomach cancer” and do not appear in the search results. Hence, a patient wanting to explore medical literature on “stomach cancer” may miss some important search results using such terms. Furthermore, the HealthVault Search offers limited refinements for some search terms. For example, “womb cancer” unlike the search term “uterus cancer” which would imply that the HealthVault Search does not recognise these two terms as synonyms and/or also does not recognise lay health terminology. Thus, the overlap of this search system with the one provided by this project is minimal.

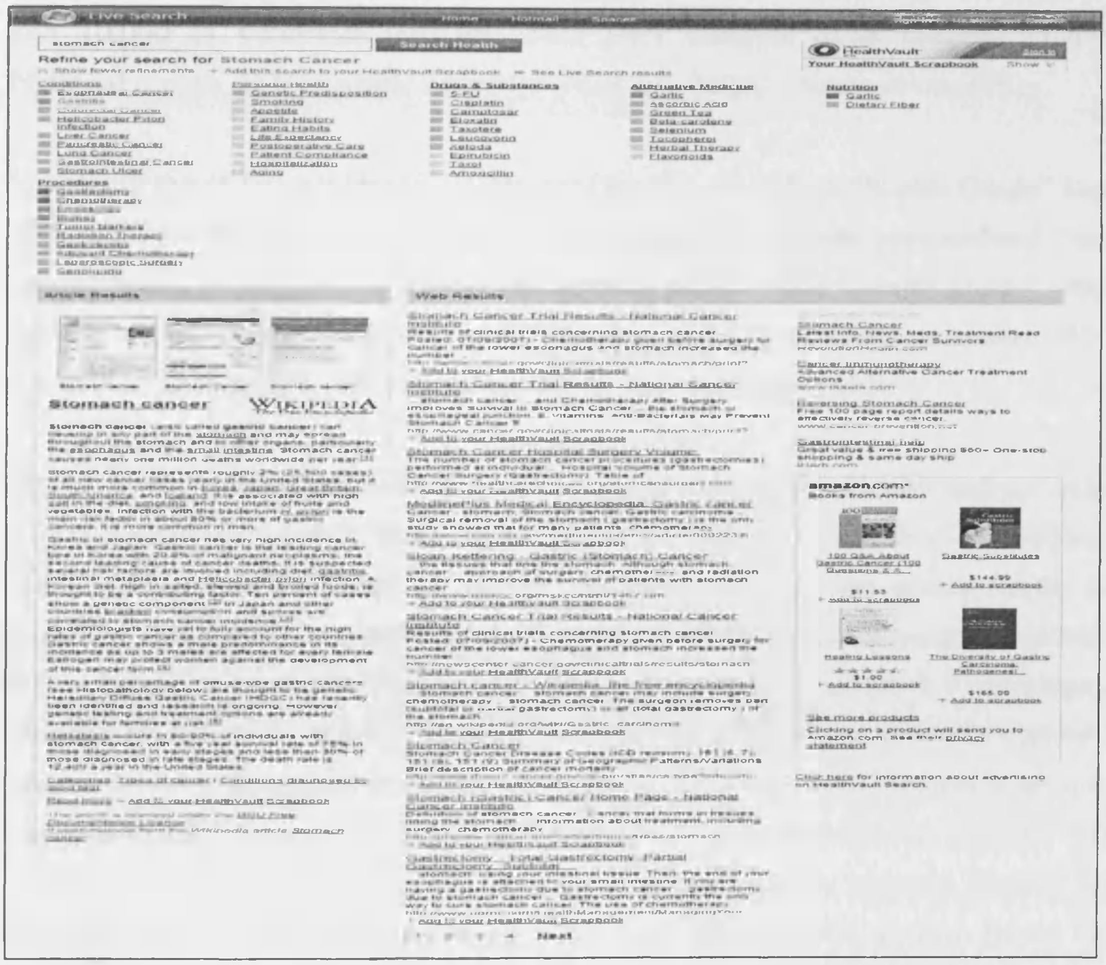


Figure 9.8: HealthVault Search Refinements and Results for the search term “stomach cancer”

9.2.5 GoogleHealth And Personal Health Records

GoogleHealth [101] integrates the PHR technology with many Google features. It adopts a consumer-centred approach that offers individuals the responsibility of managing and sharing their health information [43]. GoogleHealth was launched in February 21, 2008, as a pilot involving Google and Cleveland Clinic, and it connects Cleveland Clinic's PHR system known as eCleveland Clinic MyChart to a Google profile feature "in a live clinical delivery setting" [101]. The pilot will test the secure exchange of patient medical data between the Cleveland PHR system and Google profiles. The aim is to make the Cleveland PHRs available nationally and, hence, they can be shared "with multiple physicians, healthcare service providers and pharmacies" [101]. GoogleHealth plans to enrol a sample of between 1,500 and 10,000 of volunteer patients. This pilot outlines three benefits [101]: 1) National access, 2) Consumer empowerment and 3) 24/7 Access/Portability.

Snapshots the of GoogleHealth prototype [36, 40] describe a "health Guide" feature that searches trusted medical sources and creates a patient-personalised "health guide" based on the data stored in the patient profile. The "health guide" offers a patient information on drugs, tests, treatments and preventative measures. We are not able to analyse this search tool as it is not yet available publicly.

Furthermore, Google provides a generic Google Directory Health service (Figure 9.9) that searches trusted health information resources. In addition, it enables the search within specific health categories which could limit the search results for a patient to more relevant search results and help a patient access related information faster. However, the categories have a cascaded style (e.g. Health→ Conditions and Diseases→ cancer→ Gastrointestinal→ stomach). Thus a patient can narrow the search domain according to health categories. However, this requires a patient to understand that "stomach cancer" is a subtype of "gastrointestinal cancer" which most patients do not realise. As discussed in Chapter 2, patients usually have difficulty identifying medical terminology and relationships among terms. Such issues are addressed in this study by offering a patient similar and related terminologies of their own diagnosis and utilising this terminology in Internet searching. Furthermore, Google Directory Health lacks search refinements in terms of a specific information type related to a condition (e.g. treatment, family risk) some of which are given by a Web Google search when entering a health condition search term. Such information types can be entered by a patient in the search term after refining the search category (see Figure 9.10).

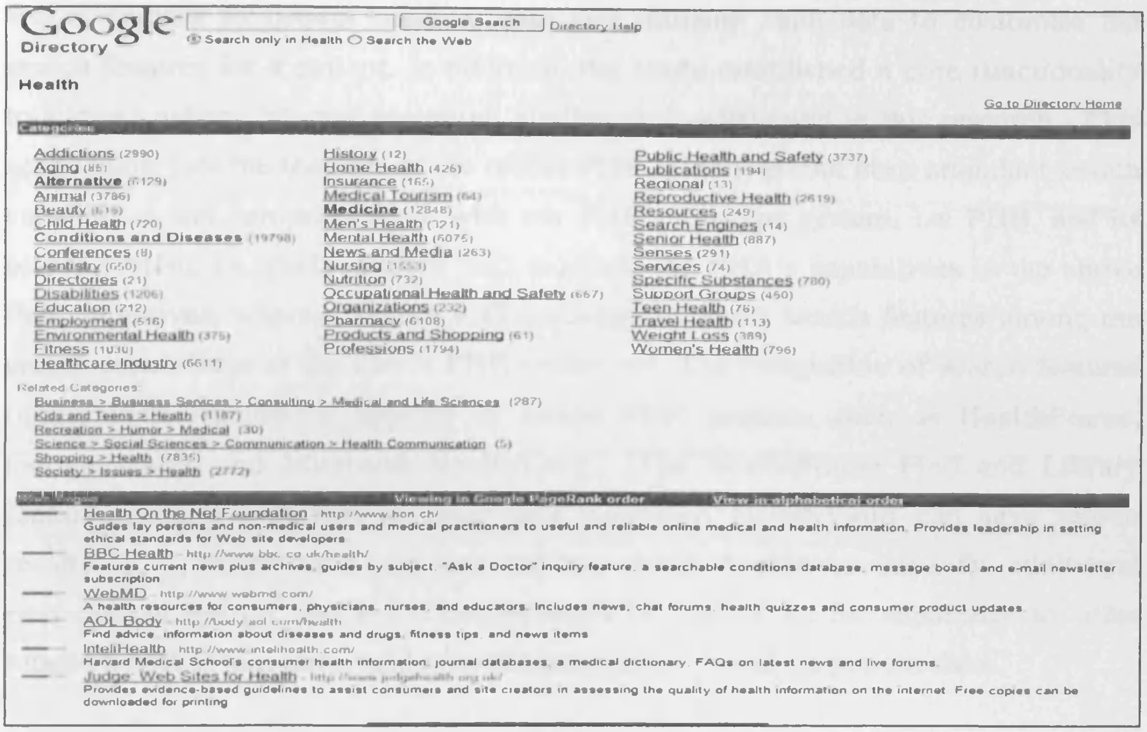


Figure 9.9: Google Directory Health Search Engine

Figure 9.10: Google Directory Health Search Results for "Treatment" in the "Stomach" category

9.2.6 Summary

This study addressed personalising patient Internet searching by linking Internet search engines to patient medical data and utilising such data to customise the search features for a patient. In addition, the study established a core functionality to address patient Internet searching challenges investigated in this research. This section analyses the features of the recent PHR initiatives and their attendant search capabilities and compares them with our PHR prototype system, i.e. PHB, and its search system, i.e. PerIS. Figure 9.11 explores the PHB's capabilities in the above PHR initiatives, whereas Figure 9.12 investigates PerIS search features among the search capabilities of the above PHR initiatives. The integration of search features in the PHR framework appears in recent PHR projects such as HealthFrame, GoogleHealth, and Microsoft HealthVault. The HealthFrame Find and Library features require a patient to enter data manually. HealthVault can save search results to a HealthVault account but we were unable to identify additional personalised search features. GoogleHealth is distinct in its capability to offer related health information using the "Health guide" based on patient data.

To the best of our knowledge, the above systems do not address the following PerIS capabilities:

1. Building a list of hospital trusted websites customised to a patient condition, and providing a search tool to search such a list.
2. Sharing interesting health websites between patients and healthcare professionals.
3. Establishing medical-to-lay term mappings on health conditions (i.e. diagnosis) to utilise in Internet searching.
4. Implementing various semantic search options that permit a patient to focus the semantic search to medical terms, lay terms or generic terms.
5. Providing separate medical health gateways and charity health websites search tools.
6. Focusing the search to a specific website or health gateway.
7. Implementing an extensive list of search refinements to focus the main search term to a specific health information type.

PHB Feature	NHS Wales My Health Online	NHS England HealthSpace	GoogleHealth	Microsoft HealthVault	Living For Health HealthFrame
EPR-Linked: links to an integrated EPR (ISCO/CaNISC).	Currently links to GP medical records but eventually to be accessible throughout authorised national health organisations.	Currently links to GP medical records but eventually to be accessible throughout authorised national health organisations. Also patients can enter/update their health details.	Can be shared by multiple physicians.	Patient permits physicians to insert information into their HealthVault record.	A patient-owned software program managed and controlled by a patient. Updates depend on patients.
Live and timely linkage to EPRs.	√	√	√	√	Patients can enter data or import a copy from their health organisation records upon request.
Personal medical information extracted: ▪ Diagnoses ▪ Treatments, ▪ Cancer management plan.	Trials indicate only access to medical records without specifying types of medical information accessed. <u>**Also allows:**</u> appointment booking, ordering repeated prescriptions, email messages to GP practice.	Portal: health details, lifestyle details, medications, blood, heart, health summary, library, diary, Choose and book. As linkage to GP records is at trial stage in selected areas, we are not able to verify if all or part of these health details are loaded from GP medical records.	GoogleHealth Profile: conditions and symptoms; medications; allergies; surgeries and procedures; test results; immunizations; age, sex and height.	CNV. Reports describe blood pressure, cholesterol levels, surgical procedures.	Access to essential personal medical information (under the Health Category): ▪ Condition ▪ Medications ▪ Visits ▪ Treatments
Patient-personalised Internet search (PerIS)	**X**	Only Medication Find feature that searches a medication database for a medication specified by a patient.	"Health Guide": offers personalised health information (see Figure 9.12).	CNV	Library Reference: largely generic but suggests related search matches using UML and ICD-9 terms.

		Also describes a Library feature that allows a patient to build a list of health websites manually.			
Building a Staff Trusted Websites (STW) list	X	X	X	CNV	X
A patient-customised Hospital Trusted Websites (HTW) list.	X	X	X	CNV	X
Building a Patient Favorite Websites (PFW) list.	X	X	X	CNV	X
Communicating interesting health websites between patients and healthcare professionals.	X	X	X	CNV	X
A Concept Thesaurus (CT) mechanism to create medical-to-lay term mappings.	X	X	X	CNV	X
A Patient Diagnosis Ontology (PDO) covering multiple medical, lay and generic term synonyms.	X	X	X	CNV	X

√: Available
X: Not Available
CNV: Can Not Verify Due to the Lack of Reports or Inaccessibility of this System to Analyse

Figure 9.11: Comparing PHB Capabilities to recent PHR projects

Search Capability	PerIS	Google Directory Health	GoogleHealth "Health Guide"	Microsoft's HealthVault Search	HealthFrame's Library Search
Search portal Type	PHR-linked and Patient-personalised. Offers: personalised search ideas, customised search tools (e.g. Hospital Trusted Websites, Favorites) and personalised semantic knowledge on diagnoses.	Generic	PHR-linked and Patient-personalised offering personalised health information covering treatment, drugs, tests and preventive measures.	PHR-Linked but generic. However, it allows saving search results to a user's HealthVault account but does not make use of HealthVault patient data to customise the search features for a patient	PHR-linked but generic, i.e. does not utilise PHR data or offer personalised search features.
Personalised search topics/ideas from a patient's own diagnosis, treatment and cancer management plan	√	X	Searches for health information based on patient data stored in the GoogleHealth profile.	X	X
Rich diagnosis vocabulary	Covers diagnosis (or health conditions) only, based on EPRs and a Concept Thesaurus (CT) integrating medical and lay vocabulary, and offers medical, lay and generic term synonyms.	Limited in terms of medical synonyms (e.g. recovers only stomach cancer, gastric neoplasm as medical terms synonyms but not other medical synonyms (e.g. stomach neoplasm and gastric neoplasm)	CNV	Limited in terms of medical synonyms (e.g. does not recognise medical term synonyms for "stomach cancer" such as "stomach neoplasm" or "gastric neoplasm"). Less search refinements are suggested for lay search terms such as "womb cancer" as compared to its medical search term "uterus cancer".	Offers mostly medical term matches based on UMLS and ICD-9 codes when selecting "look up any name or term" and the search is conducted on a specific match selected by the user. No support for lay terms (e.g. "womb cancer" returns no search results) or generic/broad terms.
Search Refinements/	Can be based on personalised search	Can refine/focus the search domain on cascaded	Possibly based on patient data and	Offers a large set of search refinements	Offers search matches based on generic categories and related

specific information types (e.g. family risk)	topics, diagnosis related terms, or an extensive list of specific information types sought by patients and investigated in this study.	subcategories of the Google Health Directory. This requires users to understand the Google Health category structure to identify category or information of interest, and to be knowledgeable in relationships between medical terms.	health categories covered by the service.	covering related terms and health topics. Less support for lay terms and limited specific information types as compared to PerIS.	terms from UMLS and ICD-9. Runs the search on a search term match. Limited support for specific health information types (e.g. family risk). Furthermore, entering the search term "stomach cancer family risk" returns no search results.
Search language	Supports focusing search result in multiple languages supported by Google API.	English only.	CNV	CNV	English only.
Linkage to trusted health websites	HONCode Search, MedHunt Search, MedlinePlus, NHSDirect Online, Hospital-trusted Websites Search.	√	√	√	Pricing and Statistics databases, MedlinePlus gateway.
Single website/gateway restrict search	√	X	CNV	X	X
Hospital-trusted and recommended websites search	√	X	CNV	X	X
Patient preferred websites search	√	X	CNV	X	X
Charity websites search	√	X	CNV	X	X
Semantic search	Covers a rich set of medical, lay and broad term synonyms.	X	CNV	Search refinements identify some related vocabulary but the search does not recover many similar term search results (e.g. stomach	Does not conduct the search on related terms. Also, there's no capability to distinguish medical and lay search terms. The search relies on the underlying gateway (e.g. MedlinePlus) semantic search

				neoplasm, gastric neoplasm versus stomach cancer).	capabilities.
Medical term only semantic search	√	X	CNV	X	X
Lay term only semantic search	√	X	CNV	X	X
Generic term only semantic search	√	X	CNV	X	X
Saving search results to PHR	√	X	CNV	√	X
Individual search of key health gateways	√	X	CNV	X	X
Individual search of third-party accredited search engine (e.g. HONCode)	√	X	CNV	X	X
Individual search of charity websites	√	X	CNV	X	X

√: Available
X: Not Available
CNV: Can Not Verify Due to the Inaccessibility of this System to Access and Analyse

Figure 9.12: Comparing PHB's PerIS Capabilities to recent PHR projects' Search Features

9.3 Research Recommendations

This research built a conceptual diagnosis knowledge model based on the ISCO diagnosis classification data to explain and enrich patient diagnosis information stored in the patient database. However, the ISCO database version utilised in this study does not record all ISCO Diag table diagnosis concepts in the ISCO Classification and Keyv2 tables that model a Read Code diagnosis concept classes and synonyms respectively. Thus, for an effective solution, ISCO Classification and Keyv2 tables need to cover every diagnosis concept recorded by the ISCO Diag table to ensure that every patient diagnosis concept can be extended with semantic knowledge using our Patient Diagnosis Ontology (PDO).

9.4 Future Work

This research demonstrated the feasibility of personalising patient Internet medical search using the contents of a patient's EPR. Its promising outcomes open several directions for future research. We discuss thirteen of them:

1. **Exploring popular online medical terminology:** this study offered a combined medical and lay diagnosis terminology to explain and enrich the patient health information vocabulary. Both medical and lay diagnosis terms proved effective in extending and enriching normal Google patient Internet searches (see Sections 8.5.3.1 and 8.5.3.2). However, our investigation into the medical diagnosis terms' search results showed that Read Code diagnosis terms did not retrieve significant Internet search results (see Figure 8.26 and 9.13). Alternative medical diagnosis synonyms stored by the ISCO DBA and those added using our system offered more significant improvement in the search results, especially from scientific and medical websites (e.g. HONCode, MedHunt). Nonetheless, medical term searches return no or insignificant search results from patient-oriented websites. For instance, the medical terms "malignant neoplasm of stomach" and "gastric neoplasm" return no search results on the Cancerbackup and Cancerhelp charity websites as shown in Figure 9.13.

Term	Term Vocabulary Type	Cancerbackup	Cancerhelp
cancer of stomach	Lay	225	4
stomach cancer	Lay	225	72
stomach tumour	Mixed	73	2
stomach tumor	Mixed	1	0
malignant neoplasm of stomach	Medical	0	0
gastric neoplasm	Medical	0	0
gastric cancer	*Mixed*	*15*	*7*
gastric tumour	Medical	2	0

Figure 9.13: Investigating Search Results using Different Diagnosis Term Vocabulary on "cancerbackup.org.uk" and "cancerhelp.org.uk" Websites

Patient-oriented websites as well as professional oriented websites seem to utilise a more readable health information vocabulary, that we call "patient-oriented" medical vocabulary (e.g. Gastric cancer, Endometrial cancer) of mixed medical and lay terms. Figure 9.13 illustrates that the term "gastric cancer" retrieves higher search results than other medical terms. This could be due to the fact that both "cancerbackup.org.uk" and "cancerhelp.org.uk" are patient-oriented websites. In fact, "gastric cancer" is a common name for "stomach cancer according to [129].

Hence, in addition to the strictly medical and lay terminology emphasised by this study, patient Internet searching needs to recognise and incorporate additional more readable medical terminology that could be popular in medical and scientific literature such as the term "gastric cancer" which is neither lay nor medical. Our current system implementation can cater for, and add such terminology to the PDO through the Information staff interface, when verifying and possibly adding new medical diagnosis synonyms. Thus, an information staff member is better able to recognise patient-friendly and popular medical terminology used in the literature, unlike a DBA who is more concerned with clinical terms.

A gynaecologist [156] explained that doctors usually use their own medical vocabulary (e.g. uterus cancer) that is different from the EPR medical classification systems (e.g. MeSH: "Endometrial neoplasm", Read Codes: "malignant neoplasm of uterus") and use lay terminology when consulting patients (e.g. Womb cancer). Thus, an investigation into the popular medical terminology used by doctors and, therefore, highly likely to occur in literature, is needed to further enrich patient Internet searching. Similarly, there is a need to investigate popular lay diagnosis terminology that is widely used in the literature.

2. **Suggesting relevant search ideas for new search terms:** PerIS is designed as a patient-personalised search tool that assumes a patient has difficulty in expressing valid medical search terminology, a situation identified in the literature. Hence, it offers search ideas in a categorised fashion based on medical information type (e.g. Your Diagnosis, Your Treatment). This ensures a patient has access to valid personalised search ideas that are focussed on a patient's health condition. Similarly, the semantic search enrichment is applied to the diagnosis search category. However, a future enhancement to PerIS's personalised search ideas could be suggestions to a patient upon entering a search term of ways of enhancing it. This requires validation of the search term for spelling, and can be used as a generic PerIS search feature that can be used by any patient or patient carer.

3. **Developing a Plain-English search tool:** implementing an internal search tool that restricts the search to the Plain English Campaign accredited health information websites. This is to ensure the retrieval of less technical or scientific information which can further reduce the information pollution problem for a patient user.

4. **Directing search tool selection based on search term terminology level:** enhancing PerIS so that it suggests potential medical and professional-oriented health gateways when a patient selects a medical term or requests a medical term semantic search. Similarly, indicating potential charity websites, PEC accredited websites and patient or lay-oriented websites for lay term searches.

5. **Extending PHB and PerIS to cover and search local hospital information sources and documents:** hospital and national patient information resources can be added to PHB and/or PerIS, such as an individualised Patient Information Pack (PIP) and/or treatment sheets.

6. **Extending PHB to cover additional personal organiser services** such as appointments, diaries, test results, important contacts.

7. **Extending PerIS to family, carers and the public:** PerIS capabilities can be incorporated into a generic online interface that can be accessed by the public or carers. Such a capability can utilise PerIS's personalised search

ideas, diagnosis ontology model, and external health gateways but care would have to be taken about access rights to EPR.

8. **Extending PHB to other health domains (e.g. diabetes):** The PHB's architecture can be easily extended to additional patient databases in other health domains. This requires exploring the database metadata to identify essential patient information, and programming the PDE component to extract this data.

9. **Enhancing PerIS search engine performance:** investigating the use of Grid technologies to improve the PerIS search engine performance.

10. **Enhancing search language options** to cover languages other than those recognised by the Google search engine API (e.g. Welsh, Somali, Swahili).

11. **Enhancing our lay diagnosis construction algorithm** to generate full lay description or partial lay description that contain mixed medical and lay vocabulary. This requires a proper definition of what constitutes medical or lay terminology.

12. **Investigating a mechanism to address medical term homonyms in a search result.**

13. **Evaluating PHB and PerIS:** An evaluation study involving patient and staff users can be conducted as a research study to investigate and evaluate user feedback on PerIS capabilities and operations. This would be needed to establish if the user community saw it as beneficial.

9.5 Final Word

The work presented in this thesis has established a new platform for delivering patient health information and patient Internet medical search capability which brings together medical, patient and lay perspectives especially with regard to information vocabulary and quality. We have addressed the diagnosis health information vocabulary by combining and integrating medical and lay terminology which proved essential when accommodating patient information needs for variable information vocabulary. Internet information quality is accommodated by using official trusted websites, charity websites, third-party accredited heath websites and a patient's Favorite websites.

Today's information world demands a modern and integrated information delivery model that links potential information stakeholders in an efficient and customised fashion. Our research approach to integrating a patient's and a professional's perspectives and operations is inline with the emerging national health information programmes for patient-empowerment and better communication between patients and professionals. The PHB architecture can be easily extended to accommodate additional patient Internet search requirements, patient databases and health gateways. The current PerIS functionality fulfils the research objectives as discussed in Section 8.7. However, it is open to further research exploration as outlined in Section 9.4.

Glossary

Electronic Medical Record (EMR)	A single patient's clinical medical record within a single health organisation (e.g. GP, hospital) that records patient's clinical data. It is only accessed by legitimate clinicians involved in patient care.
Electronic Patient Record (EPR)	A single common multi-provider integrated electronic medical record that is shared across participating health organisations and accessed only by authorised clinicians. It stores a patient's clinical data from multiple providers (e.g. ISCO/CaNISC system). This is the legal record of patient clinical data recorded by participating health organisations. EPR is also described in literature using the term Electronic Health Record (EHR) (e.g. [192])
Personal Health Record (PHR)	A patient accessible electronic health record that stores a subset of a patient's EPR data, stores clinical data that is deemed essential and useful for a patient to access (e.g. diagnoses, treatment, tests), and has patient input. Recent PHR prototype systems (e.g. miHealth [225], MyChart [216]) include additional services (e.g. prescription renewal, diaries, appointments). Similar terms (or projects) denoting this patient record type include (but are not limited to): ▪ The NHS England Summary Care Record [104, 130]. ▪ The NHS Wales Individual Health Record (IHR) [71]. ▪ The NHS Scotland National Integrated Care Record (ICR) [99]. ▪ US iHealthRecord [65]. ▪ The Patient Health Base (PHB) prototype

system developed during this research study.

Summary Medical Record (SMR)

A term used by this study to describe the subset of ISCO/CaNISC EPR (currently covering diagnoses, treatment episodes and cancer management plan) extracted by this study for a patient to view and for utilisation by the PerIS system within the Patient Health Base (PHB) prototype system developed during this study. In future work, SMR can be extended to cover tests, allergies. It is worth noting that the SMR feature is used in this study to distinguish a patient's clinical data extracted from ISCO/CaNISC EPR data, from other personal health information that can be added to our PHB (i.e. PHR) prototype system either by the patient or by the health organisation (e.g. prescription renewal, appointments, diaries).

Summary Care Record (SCR)

The NHS England initiative to offer a patient online access to personal medical information recorded by NHS England services.

HealthSpace

The NHS England website which enables a patient to store and manage their personal health information online and connect to their SCR.

Individual Health Record (IHR)

The NHS Wales initiative to make a patient's personal medical information accessible online by a patient and across authorised healthcare organisations.

My Health Online

The NHS Wales website which enables a patient to access and manage their personal health information online.

Appendix A

A Sample PHR Projects

A.1 The NHS England Summary Care Record

The NHS Summary Care Record is part of the NHS England Care Service, within the NHS Connecting for Health (CfH) programme [105]. The NHS Care Service aims to develop a secure health information system across England that is accessible by both professionals and patients. The underlying electronic health system is composed of two electronic record types: a Detailed Care Record, and a Summary Care Record. The Detailed Care Record contains detailed treatment notes made by healthcare professionals involved in a patient's care. The Summary Care Record stores selected information from the Detailed Care Record that is important to a patient such as medications and prescriptions and would be accessible to patients via a Web portal known as HealthSpace [53]. Launched in December 2003, HealthSpace provides a patient with his/her own online health organiser and by 2008 will enable access to the NHS Summary Care Record [47, 52]. Currently, HealthSpace offers a patient the following functions [52]:

- Calendar - generation of email reminders for appointments [52],
- Personal health history and health tracker [52],
- Personal library and address book [52],
- Search for local NHS service information [52].
- Prescription renewal and nominating pharmacies [47], and
- Arranging appointments and specifying referral hospitals and clinics [47].

Initially patient medical information comes from the local GP but eventually it will come form other parts in the NHS [84]. Information is added to SCR each time a patient uses NHS Services [84]. Patients are informed of those additions during routine consultations and have the option to use SCR, HealthSpace or limit access to their information.

A.2 The NHS Wales Individual Health Record (IHR)

The Individual Health Record (IHR) is the NHS Wales's patient health record project within the Informing Healthcare (IHC) programme [109]. The aim of IHR is to "integrate information at the time of care, so that patients are empowered by having the information they need to take part in the decision process about their own healthcare" [71]. Initial anticipated IHR functionalities include [71]:

- Personal details – identity and preferences,
- Care relationships – who is involved in the patient care [71],
- Information from health events (e.g. discharge summary, operation letters), and
- Current health status (e.g. current prescribed medication).

Additional services may be included when greater integration of NHS information systems is achieved, such as [71]:

- Personal health information,
- Making appointments, and
- Corresponding electronically.

IHR is designed for access by patients and healthcare professionals involved in patientcare [71]. The record will be accessible by patients through a Web-based public gateway called "My Health On-Line" [95]. Pilot Individual Health Record projects include (but are not limited to):

- Gwent Emergency Care Record, and
- A pilot maternity portal: A patient-held maternity record and a personal pregnancy record.

A.3 The NHS Scotland National Integrated Care Record (ICR)

In Scotland, the patient-held medical record notion is outlined in the "Patient Focus and Public Involvement" plan [114], it will be accessible by patients through smart Cards. The NHS Scotland National eHealth/ IM&T Strategy [99] describes an Integrated Care Record that is managed by both patients and professionals. The aim of the strategy is "to deliver an Integrated Care Record jointly managed by patients and professional NHS staff with in-built security of access governed by patient consent" [99]. Additionally, there are a number of isolated PHR projects in Scotland such as Babylink [6] and renalpatientview [124]. In terms of PHR capabilities, the "Patient-Focused NHS" document [114] outlines a number of functionalities to be included in the patient-held record:

- Access to personal health information,
- An educational material,
- A space for a patient to record information about themselves, and
- A widening range of patient information sources (e.g. Linkage to NHS24 website) and improving access to it [114].

A.4 US iHealthRecord

Launched on May 9, 2005, iHealthRecord [65] is a PHR by Medem Incorporated available to any individual in the US. iHealthRecord is available through physicians registered with the Modem network [66]. The system enables a patient to create and update their iHealthRecord online. More than 10,000 Americans built an iHealthRecord during the first weeks of its launch [257]. iHealthRecord functionality includes (but are not limited to):

- A patient can create, access and update iHealthRecord online [65],
- Ability to access medical personal information [257],
- Ability to access the iHealthRecord in an emergency [257],
- A wallet card providing emergency contact information [257],
- Access to iHealthRecord is controlled by a patient who can share his/her health information with whom he/she wants (e.g. family, physicians) [65],
- Email and online consultation with physicians [257], and
- Medication reminder via email [257].
- Education programs tailored to individuals [257]. Based on condition and medication information, a patient can receive educational information from trusted health authorities including the Food and Drug Administration (FDA), the Centres for Disease Control (CDC), the American Heart Association (AHA), and the nation's leading medical societies [65]. However, it is not clear how the educational material is delivered (online or in a printed form).

A.5 miHealth

MiHealth [225] is a personalised Web service for breast cancer patients at the Liverpool John Moores University's International Centre for Digital Content (ICDC). miHealth provides a localised information resource – a central database of accurate, up-to-date, authoritative information that is personalised to the patient's healthcare journey [234]. It is designed to reflect the individual patient needs [234]. The system offers personalised services such as [234]: miInformation; miDiary and miTreatment; miContacts and Useful Information; and miMoodStates. The miInformation service is a database of information that is structured around the breast cancer patient journey. The service enables a patient to select the information that they regard as relevant to them. In addition, it incorporates a glossary section to look up terms used in the website's main information pages.

The system provides information in rich format (e.g. text, images, audio/video-clips) and delivers to multiple communication platforms (e.g. PCs, hand-held computers, kiosks, interactive TV and mobile phones).

A.6 MyChart

MyChart at Geisinger Health System [38] is a Web portal that enables patients to view selected portions of their Electronic Health Record and exchange electronic messages with their doctor's practice [286]. MyChart offers patients the following electronic services [216]:

- Review laboratory tests, allergies, medications and healthcare problem lists,
- View their past and future office visits and review their health related histories,
- Request an appointment, prescription renewals and referrals, and
- Send messages and queries to their providers.

Appendix B

Requirement Analysis

B.1 Domain Problems

Problem number	Description
P1	EPRs data are modelled for clinicians but not patients' use.
P2	EPRs are described using medical terminology.
P3	EPRs data are of high security and should be accessed by legitimate users.
P4	Patients currently lack direct access to personal medical information.
P5	No current patient interface to the ISCO system.
P6	Patients have variant information needs.
P7	Patient Internet access at Velindre NHS Trust is laborious, manual and uncustomised.
P8	Velindre patients are guided to key Internet health information sources using generic paper list of key health websites.
P9	The Internet covers wide-ranging and disparate Internet health information search tools.
P10	Internet information quality: Internet information is unregulated and uncontrolled.
P10a	• Generic search tools do not indicate trusted websites to patients.
P10b	• No authoritative advice from healthcare providers in guiding patients to trusted Internet information sources.
P10c	• No communication between patients and healthcare providers regarding patients' Internet information research.
P10d	• Professionals require patient access to authoritative or hospital-trusted information resources.
P10e	• Patient demand unrestricted access to Internet information.
P10f	• Professionals are generally unaware of Internet health information resources.
P11	Internet health information vocabulary:
P11a	• Patients have difficulty expressing the correct medical term describing their sought information.
P11b	• Patients have difficulty formulating proper lay terms.
P11c	• Patients have difficulty identifying related vocabulary (e.g. synonyms, hierarchies).
P11d	• Conflicting patients' information needs regarding health information vocabulary; some demand medical and scientific health information while others request lay health information.
P11e	• Generic health information vocabulary is aimed at professionals and, therefore, could be ambiguous, too broad and /or technical for average patients.
P11f	• Generic Internet search tools do not locate health information described using various relevant terms.
P12	Web search tools do not offer patient personalised health topics or search ideas.
P13	Web search tools may not offer patient sufficient health information search refinements.
P14	Internet information overload: Numerous Internet information resources that deliver unfocused patient search results.
P15	Internet information pollution: Internet health information is written in medical terminology and might contain technical jargons or irrelevant details.

Figure B.1: Domain Problems

B.2 Stakeholders Needs (Functional Requirements)

Need	Description	Stakeholder	Maps to Problem (s)
N1	Need to access meaningful and useful EPR details.	Patient	P1
N2	Need to understand medical EPR terminology.	Patient	P2
N3	Need secure interface to the ISCO System.	Patient	P3
N4	Need to access essential personal medical information.	Patient	P4
N5	Need online interface to the ISCO system.	Patient	P5
N6	Need accommodating variant information needs.	Patient	P6
N7	Need improved and customised (or personalised) Internet search mechanism.	Patient	P7
N8	Need customised and electronic list of hospital-trusted or key health websites.	Patient	P8
N9	Need guidance to potential Web search tools and medical and health gateways.	Patient	P9
N10	Need guidance to quality health websites.	Patient	P10
N11	Need guidance to key health gateways.	Patient	P10a
N12	Need access to Hospital-Trusted Websites (HTW) list.	Patient	P10b
N13	Need sharing of or feedback on researched Internet information sources.	Patient	P10c
N14	Need to verify trusted Internet information sources to patients.	Staff	P10d
N15	Need open and unrestricted access to Internet information.	Patient	P10e
N16	Need guidance on key and accredited Web health websites.	Staff	P10f
N17	Need Patient-oriented Health Information Vocabulary (PHIV) that accommodates patient information needs and preferences.	Patient	P11, P11e
N18	Need to recognise correct medical terminology on health problems.	Patient	P11a
N19	Need to recognise correct lay terminology on health problems.	Patient	P11b
N20	Need to recognise health terms hierarchies, i.e., specific/generic terms.	Patient	P11c
N21	Need to recognise medical term synonyms.	Patient	P11c
N22	Need to recognise lay term synonyms.	Patient	P11c
N23	Need to recognise generic term synonyms.	Patient	P11c
N24	Need to access medical and scientific health information	Patient	P11d
N25	Need to access lay health information	Patient	P11d
N26	Need to locate or retrieve various Internet health information described in related vocabulary.	Patient	P11f
N27	Need personalised Internet search topics or ideas.	Patient	P12
N28	Need potential Health Information Search Refinements (HISR).	Patient	P13
N29	Need to access relevant Internet information.	Patient	P14
N30	Need to access less technical and clear health information that is patient-oriented.	Patient	P15
N31	Need to interpret EPR medical terminology in lay terminology.	Information Staff	P2, P11b

Figure B.2: Stakeholders Needs

B.3 Proposed System Features

Feature	Description	Maps to Need (s)
F1	The system will offer patient meaningful EPR information on diagnosis, treatment and cancer management plan.	N1
F2	The system will incorporate a mechanism to establish a PHIV conceptual model that accommodates patient information needs.	N2, N17
F3	The system will incorporate a security mechanism to allow access only to legitimate users.	N3
F4	The system will offer patient direct and electronic access to essential personal medical information	N4
F5	The system will create a patient interface as a Web-portal to the ISCO database system with additional patient-oriented functionality and features.	N5
F6	The system will incorporate variant patient information needs and execute patient preferences.	N6
F7	The system will incorporate a patient Personal Internet Search (PerIS) facility	N7
F8	The system will incorporate a mechanism that offers patient access to a customised Hospital-Trusted Websites (HTW) list.	N8, N10, N12
F9	PerIS will allow patient search key health gateways and search engines	N9, N11
F10	PerIS will allow patients search HTW list	N12
F11	The system will incorporate a mechanism to enable hospital staff view and select from patient preferred (or Favorites) health websites.	N13
F12	The system will incorporate a hospital staff interface	N14
F13	The system will incorporate a mechanism to enable interested staff build an individual trusted list of websites.	N14
F14	PerIS will enable generic unrestricted Google Web search	N15
F15	The system will incorporate a mechanism to guide professionals to key and third-party accredited health websites.	N16
F16	The system will incorporate a mechanism to allow an information staff (or a librarian) specify accredited Web health websites so it can be accessed and used by professionals when building their trusted health websites.	N16
F17	The system will incorporate a mechanism to verify PHIV before being used by patients.	N17
F18	PHIV will incorporate medical terminology describing patient health information. This feature will offer patient correct medical terminology describing their personal medical information.	N18
F19	PHIV will incorporate corresponding lay terminology describing patient health information in simple English. This feature will allow patient to use correct lay terminology describing their personal medical information.	N19
F20	PHIV will incorporate generic terminology relating to patient health problems. This will allow patient to relate generic and	N20

	specific (i.e. medical) terminology.	
F21	PHIV will incorporate medical synonyms terminology to enable patient utilise and recognise similar medical terms describing the same concepts.	N21
F22	PHIV will incorporate lay synonyms terminology to enable patient utilise and recognise similar lay terms describing the same concepts.	N22
F23	PHIV will incorporate generic synonyms terminology to enable patient utilise and recognise similar generic terms describing the same concepts.	N23
F24	PerIS will offer patient personalised search ideas from EPRs data and PHIV terminology	N27
F25	PerIS will enable patient search for health information described using medical terminology.	N24
F26	PerIS will enable patient search for health information described using lay terminology.	N25
F27	PerIS will enable patient perform a semantic search that retrieves health information described using related vocabulary	N26
F28	PerIS will offer patient a set of potential HISR. An information staff may establish such a set.	N29
F29	The system will incorporate a mechanism to establish and update the HISR set.	N28
F30	PerIS will incorporate Web search focusing mechanisms to allow patients access relevant and preferable Web information. Several focusing dimensions are investigated: personalised search ideas, potential search refinements, rich information vocabulary, variant quality websites, and search domain.	N29
F31	PerIS will incorporate a mechanism that enables patient access less technical, clear and patient-oriented health information. The system will incorporate charity health websites as key patient-oriented health websites (e.g. cancerbackup.co.uk).	N30

F32	The system will incorporate an information staff (or librarian) interface for the following tasks:
	• Establish and update HISR
	• Establish and update third-party accredited health websites.
	• Manage a Concept Thesaurus (CT) that defines:
	  ▪ Medical-to-lay term mapping
	  ▪ Medical synonyms
	  ▪ Lay synonyms
	• Verify generated PHIV
	• Establish and update gateways and charity health websites links.
		N9, N11, N16, N19 – N23, N28, N30 – N31

Figure B.3: Proposed System Features

B.4 Proposed System Constraints[22] (Non-Functional Requirements)

Constraint	Description
C1	Patients access the Web-based ISCO interface via a Web browser.
C2	The ISCO patient interface should be simple and adopts a user-friendly terminology.
C3	No technical knowledge is required for either patient or staff on using system services apart from normal Web interaction knowledge.
C4	The system should provide a user manual or help webpage on the system features.
C5	The system should offer a secure access to individual space and/or medical information.
C6	Internet information resources or search tools should be labelled to enable patient make informed decision on a particular Web search tool.
C7	The system should include a Disclaimer[23].
C8	Communication with the ISCO database through an SQL Server 2000.
C9	Communication with Google through GoogleApi.
C10	Communication with specific health gateways though website search service.
C11	Users should be able to delete from or add to Favorites (or trusted websites)
C12	User should be able to add from search results to Favorites (or trusted) websites.
C13	Patients should be able to add from HTW to Favorites.
C14	Staff should be able to add from patients Favorite Websites to individual trusted websites.
C15	System should be reliable and accessible at all times.
C16	PerIS should give patient the option to perform either normal or semantic Web search.
C17	PerIS should give patient the choice to perform medical term -or lay term-based Web search.
C18	PerIS search results should be unique.
C19	The combined list of patient Favorite Websites should be unique.
C20	The combined list of Staff trusted websites should be unique.
C21	Search results should be hyperlinked.
C22	Each search result should open in a new window to keep the actual system window current.

Figure B.4: Proposed System Constraints (Non-Functional Requirements)

[22] As an investigational study, we largely focused on the operational rather than the "look and feel" system constraints. The later needs be fully addressed in final system products.

[23] A Disclaimer notifies users that the system or (Velindre Hospital) is not responsible for the content of external websites.

Appendix C

PHB Algorithms

C.1 Building Diagnosis Lay Terms using CT

```
- get all concepts having scientific synonyms – select distinct  concept from table scientific
for all concepts
        - get concept(i)
        - get scientific synonyms of concept (i) – select synonym from scientific where concept=concept(i)
        - get lay synonyms of concept(i) – select synonym from english where concept=concept(i)

        if (concept(i) has lay synonyms)
         for all scientific synonyms of concept(i)
                - get scientific synonym(j) of concept(i)
                        for all incoming diagSyns //input parameter
                                - get diagSyns (k)
                                        if concept(i) scientific synonym(j) exists in diagSyns (k)
                                                for all lay synonyms of concept(i)
                                                        - get concept(i) lay synonym(s)
                                                        - create a new lay diagnosis synonym replacing every occurrence of
                                            concept(i) scientific synonym(j) in diagSyns(k) with concept(i) lay synonym(s)
                                                                - add new diagnosis lay synonym to new diagSyns

- return new diagSyns
```

C.2 Building Additional Diagnosis Medical Terms using CT

```
- concepts=getAllSciConcepts// select distinct concept from table scientific
if concepts not null
    for all concepts
                - get concept(i)
                - conSyns=get scientific synonyms of concept(i) – select synonyms from table KBUsers scientific where
concept=concept(i)
        for all diagSyns {from DiagClassification}{//represent medical diagnosis synonyms from ISCO corev2 & keyv2 tables
                    - get diagSyn(j)
                    while (more scientific synonyms of concept(i) && not found)
                        - get concept(i)scientific synonym (k)
                        if  concept(i)scientific synonym (k) exists in diagSyn(j) text
                            - get its pos in diagSyn(j)
                            - get its text in diagSyn(j)
                            - conceptScientificSynonymFound =true
                        if(conceptScientificSynonymFound =true)
            for all scientific synonyms of concept(i)
                - get concept(i)scientific synonym (k)
                - create new diagnosis scientific synonym replacing all occurrences of the diagsyn(j) text with the current
            concept(i)scientific synonym (k)
                    - add new diag synonym to input diagSyn vector if not already in vector.//diagSyns contains original sci syns

 - return input diagSyn //with additional synonyms
```

Appendix D

PHB GUI Operations

D.1 Adding Items to Staff Trusted Websites (STW) List

Figure D.1 shows six options for adding new STW website items. The PHB system facilitates the construction of STW lists by equipping this process with lists of key and third-party accredited health websites identified and updated by an information staff using the information staff GUI operations.

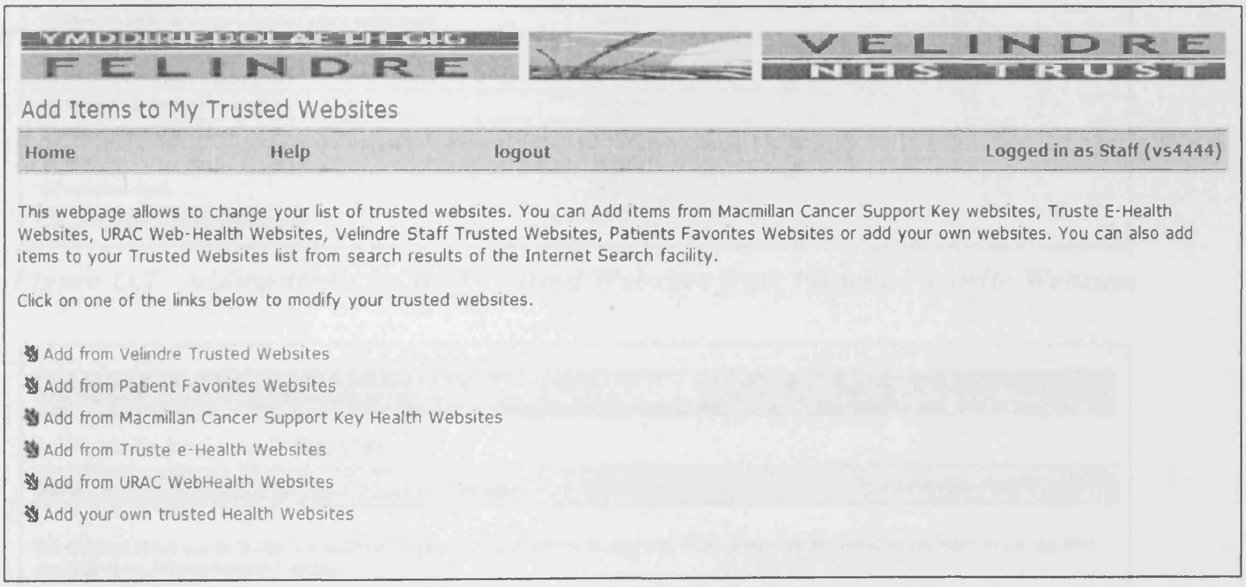


Figure D.1: STW Add Items Options

Three authenticated health websites lists are utilised in the construction of STW: Macmillan Key health websites list, URAC WebHealth Accredited Websites, and Truste e-Health Accredited Websites. In addition, a staff can add items to STW from Hospital-trusted websites, patient Favorite Websites, PerIS search results or by entering new items. Figures D.2 and D.3 demonstrate the process of adding STW items from patient Favorite Websites and Truste E-Health websites respectively.

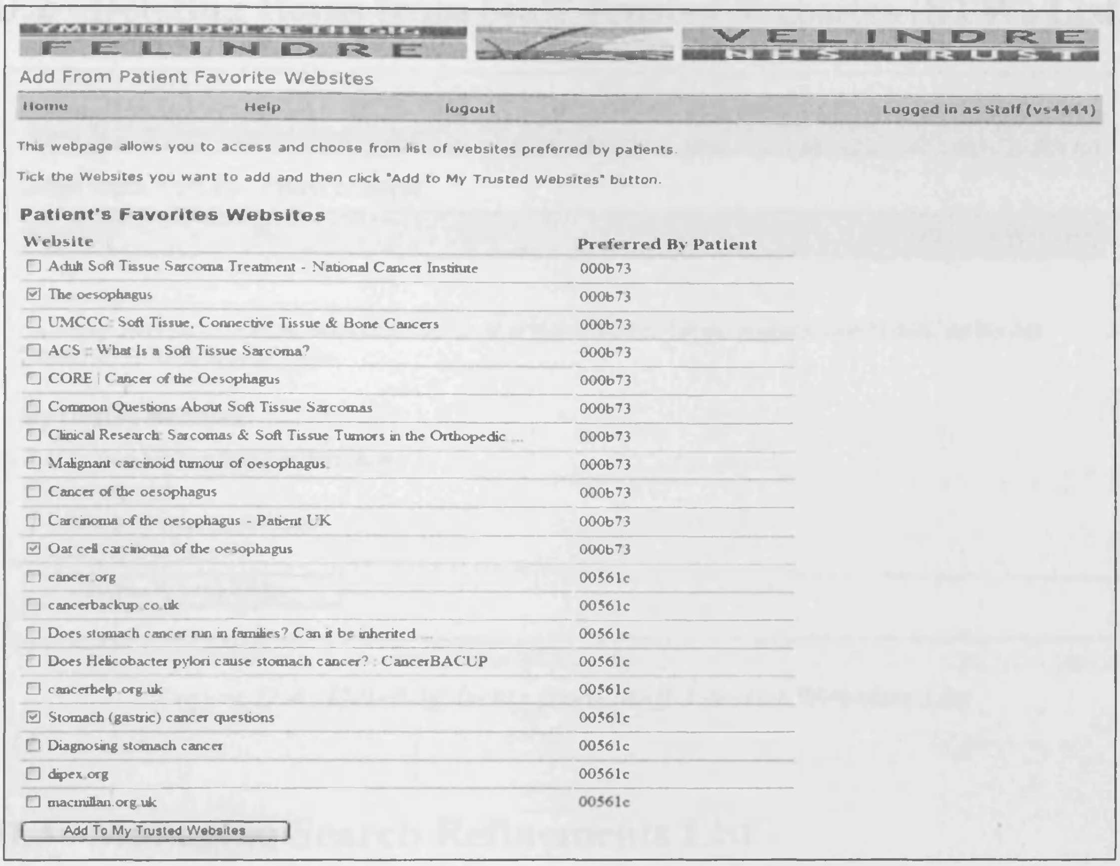


Figure D.2: Adding Items to Staff Trusted Websites from Patient Favorite Websites

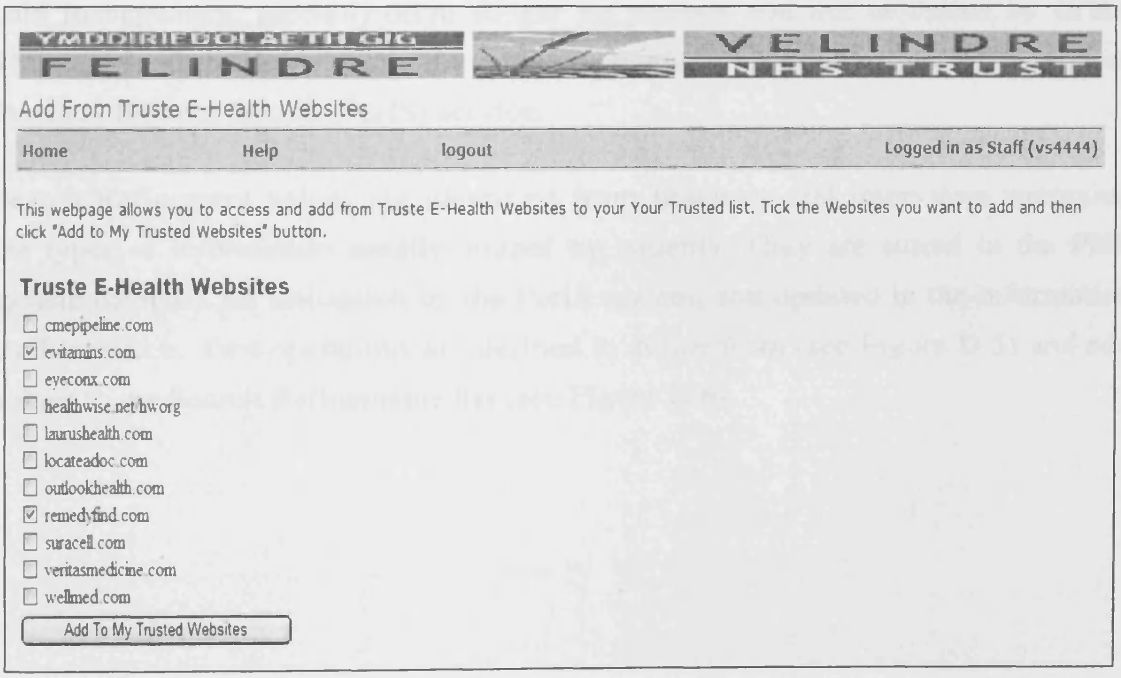


Figure D.3: Adding Items to Staff Trusted Websites from Truste E-Health Accredited Websites

D.2 Deleting Items from Staff Trusted Websites (STW) List

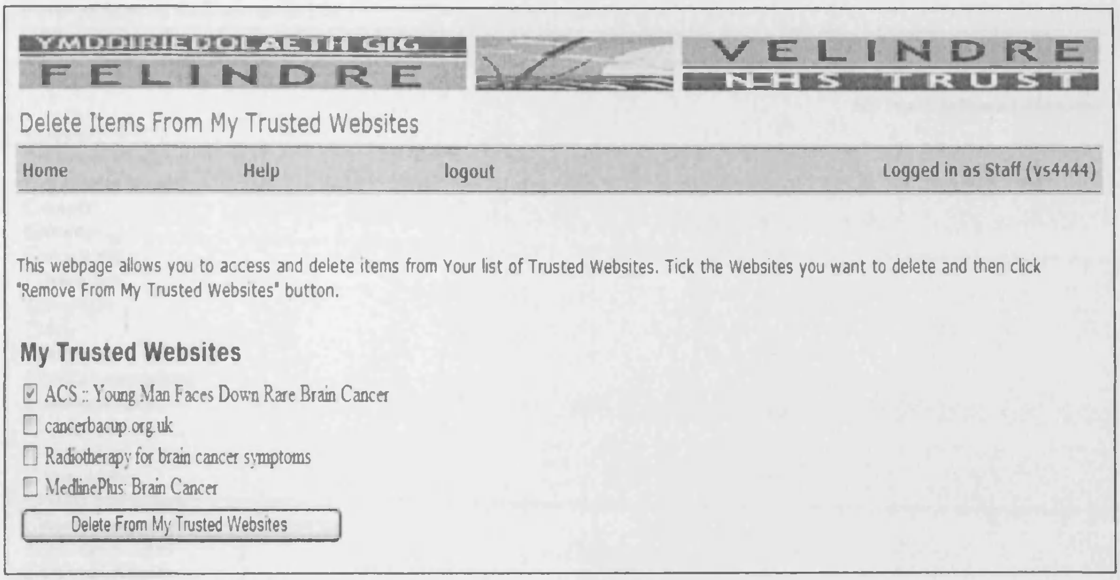


Figure D.4: Deleting Items from Staff Trusted Websites List

D.3 Managing Search Refinements List

Search Refinements denote health information types (or factors) (e.g. side effects, pain management, alcohol) often sought by patients and not modelled by EPRs. They are used to further focus the patient's search information topic in the patient Personal Internet Search (PerIS) service.

Search Refinement values are identified from literature and interviews surveying the types of information usually sought by patients. They are stored in the PHB system database for utilisation by the PerIS system, and updated in the information staff interface. Two operations are defined to delete from (see Figure D.5) and add values to the Search Refinements list (see Figure D.6).

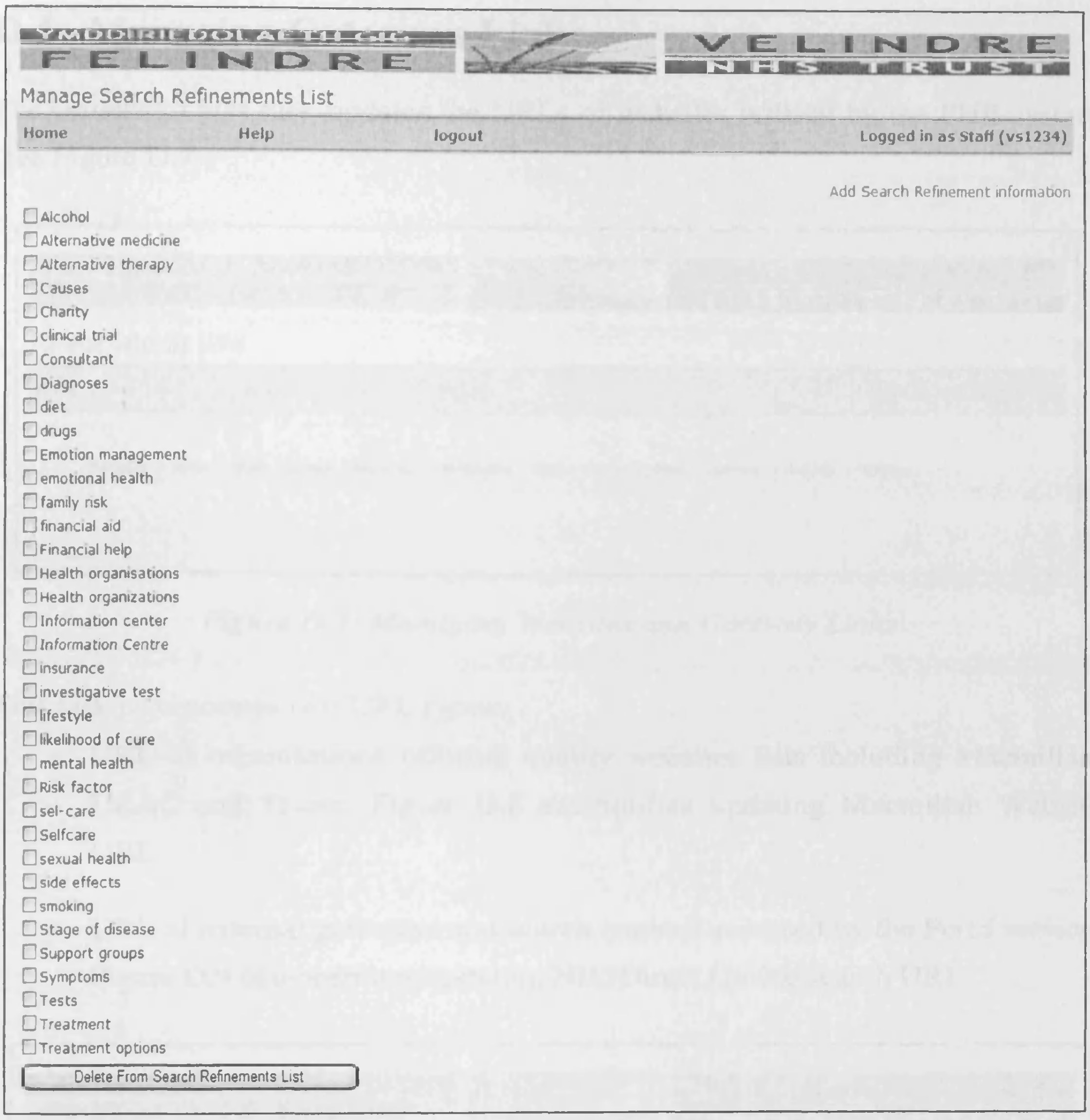


Figure D.5: Managing and Deleting from Search Refinements Webpage

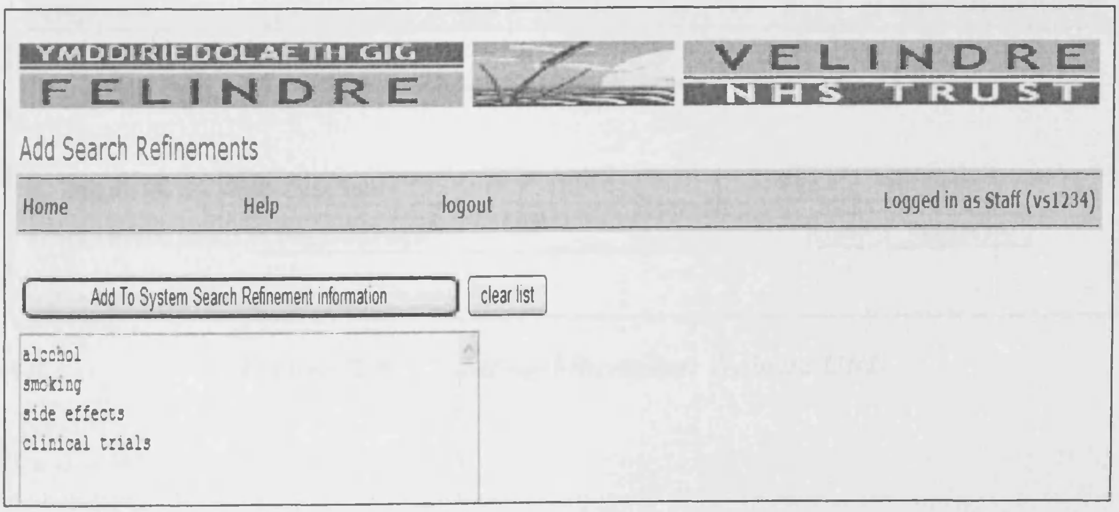


Figure D.6: Adding Search Refinements Webpage

D.4 Managing Gateways Links

The Gateway Links task updates the URLs of websites utilised by the PHB system (see Figure D.7).

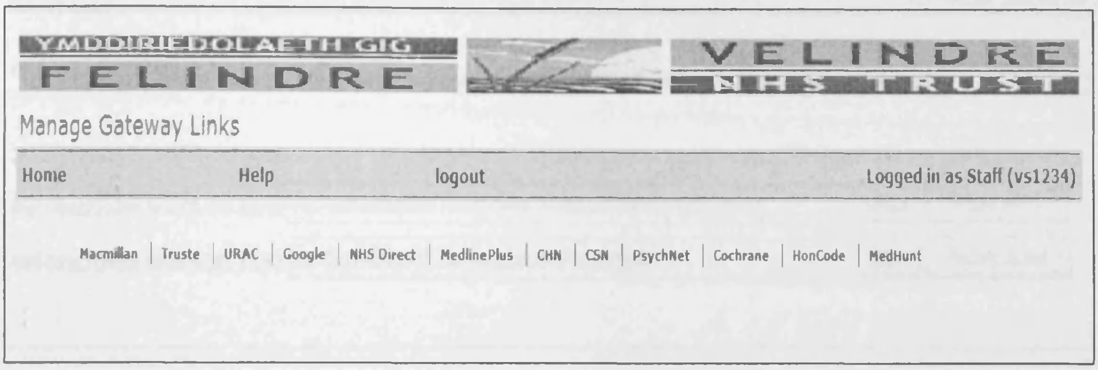


Figure D.7: Managing Websites and Gateway Links

This task incorporates two URL types:

- URL of organisations offering quality websites lists including Macmillan, URAC and Truste. Figure D.8 exemplifies updating Macmillan Website URL.
- URL of external gateways and search engines accessed by the PerIS service. Figure D.9 demonstrates updating NHSDirect Online search URL.

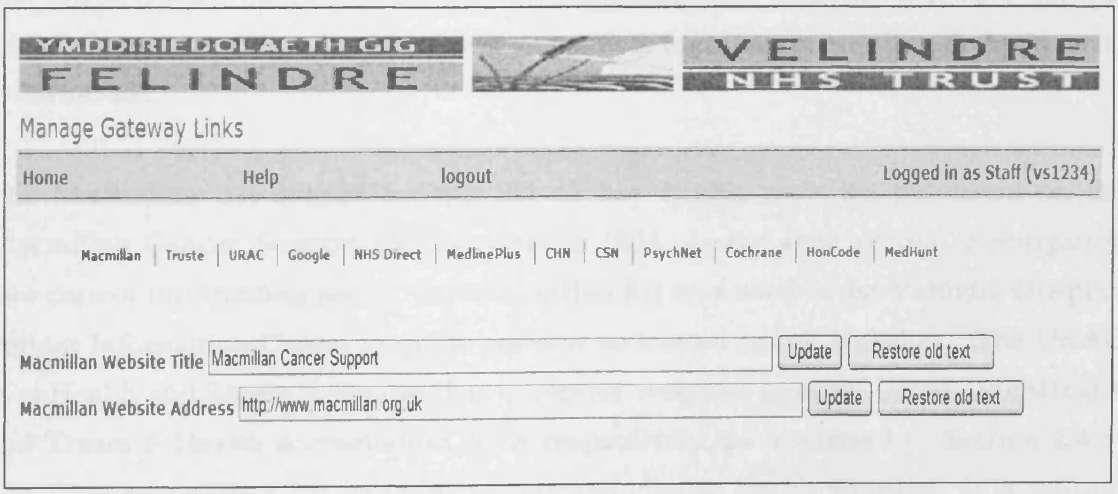


Figure D.8: Updating Macmillan Website URL

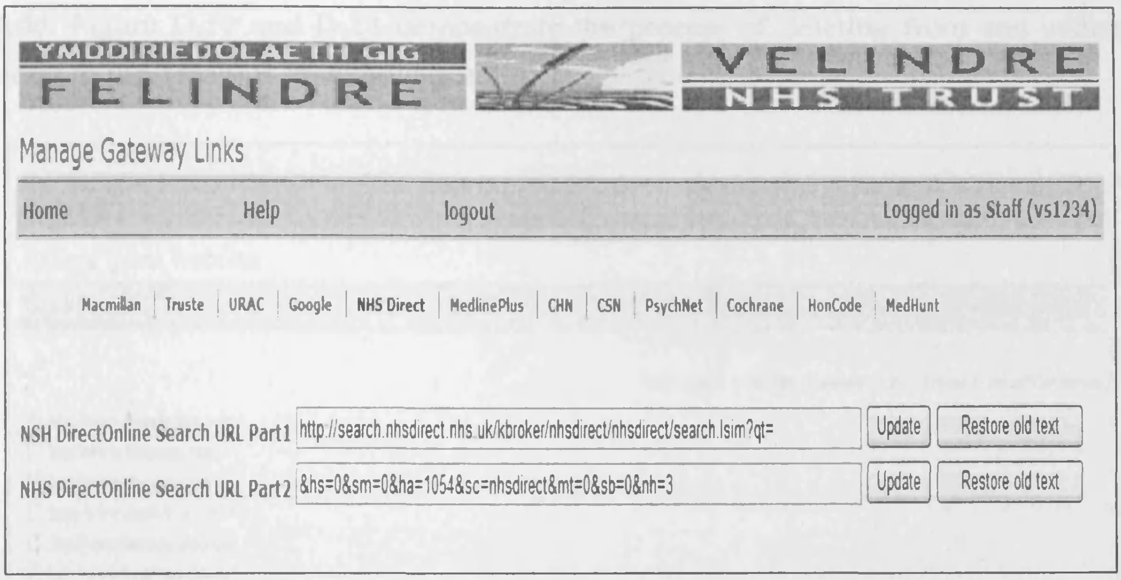


Figure D.9: Updating NHS Direct Online Search URL

D.5 Managing Macmillan, Truste, URAC and Charity Health Websites lists

The Patient Health Base (miHealthBase) utilises three third-party accredited health websites lists; Macmillan, URAC WebHealth, and Truste E-Health accredited websites. Accordingly, these lists are regarded as “good quality” health websites that hospital staff members can utilise in building their Trusted Websites list. In addition, the Macmillan list is offered to patients to utilise in building their *Favorite Websites* list.

The Macmillan list constitutes the list of key health websites published on the Macmillan Cancer Support (MCS) website [85]. At the time of our investigation into patient information needs, the Macmillan list was used at the Velindre Hospital Patient Information Centre to guide patients to trusted health websites. The URAC WebHealth and Truste E-Health lists comprise websites holding URAC WebHealth and Truste E-Health accreditation seals respectively, as discussed in Section 2.4.3. The Charity websites list contains recognised charity health websites. It is utilised in the PerIS’s Charity Websites Search.

These lists are managed and updated similarly using the relevant links in the Information Staff GUI. Each list is managed by two operations: 1) Delete and 2)

Add. Figure D.10 and D.11 demonstrate the process of deleting from and adding items to the Truste E-Health website list.

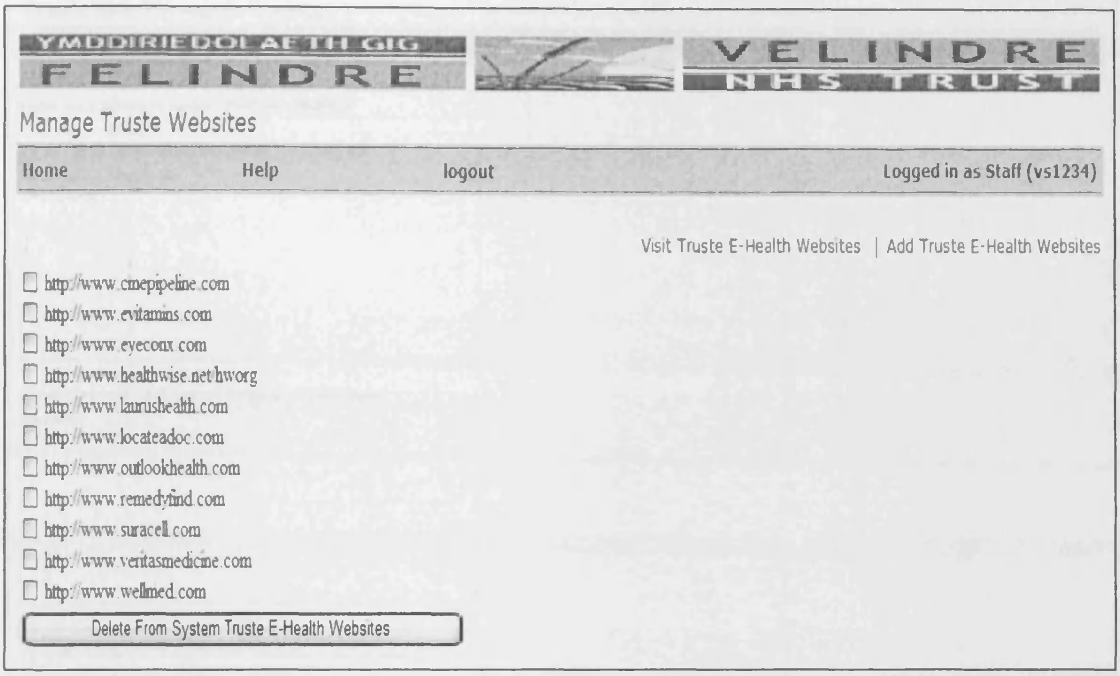


Figure D.10: Deleting from Truste E-Health list

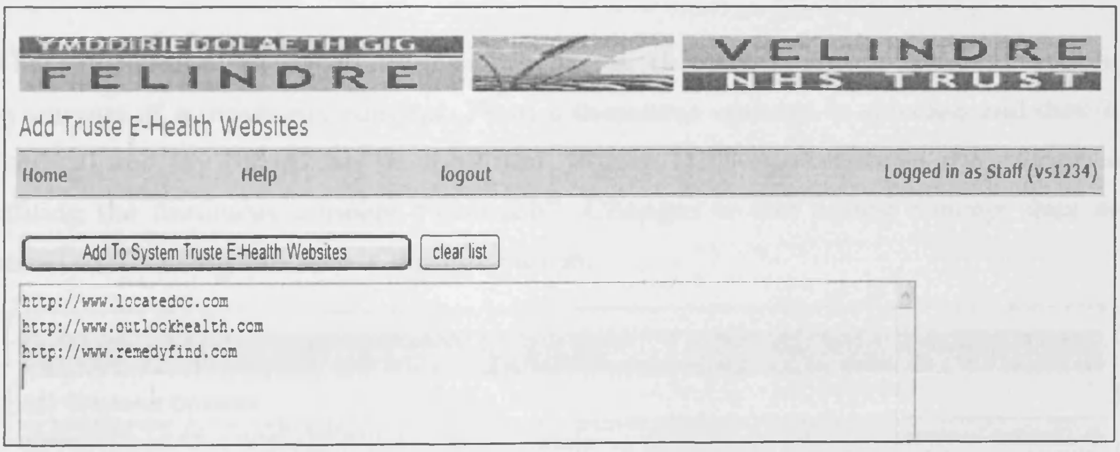


Figure D.11: Deleting from Truste E-Health list

D.6 Create and Edit Thesaurus Concepts

A new concept can be added to CT by using the *Create New Thesaurus Concept* button. The software prompts for the name of the new concept and lists of its medical and lay synonyms (Figure D.12). The name of the concept should be included in the medical labels list of that concept. New concept data are saved to the PHB DB using the *Save New Concept Data* button.

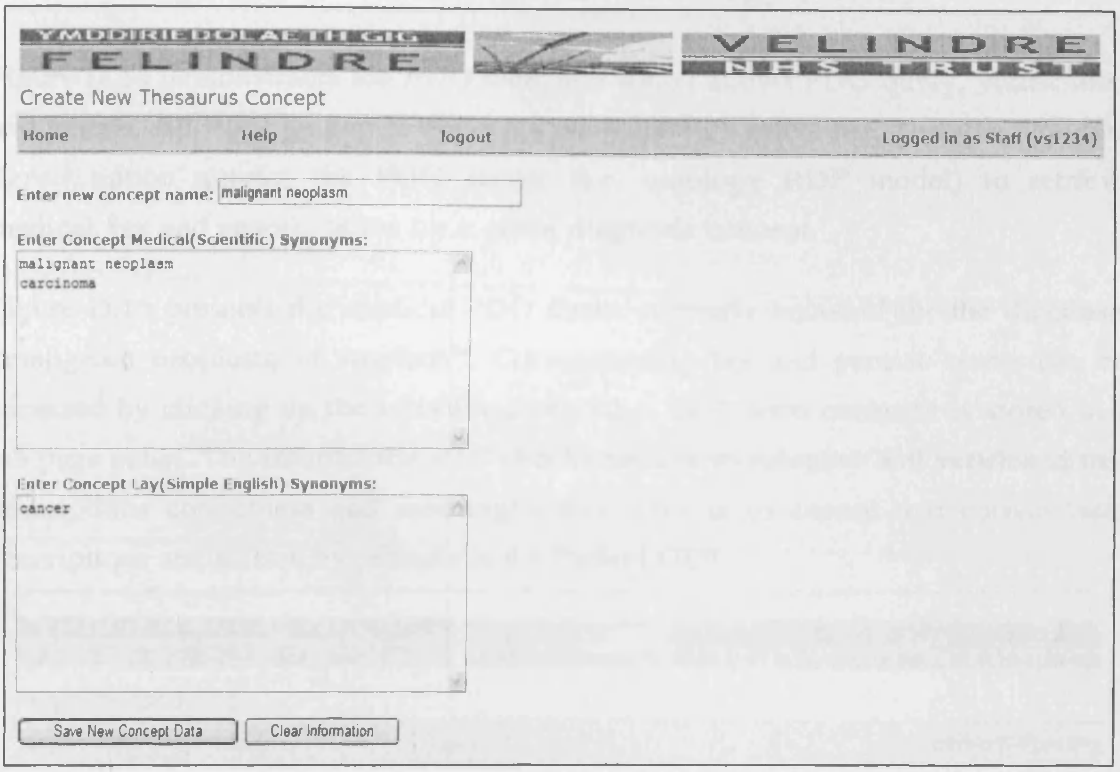


Figure D.12: Creating New Thesaurus Concept

The information staff can change the terms (labels) denoting medical and lay synonyms of a thesaurus concept. First, a thesaurus concept is selected and then its medical and lay terms can be modified. Figure D.13 demonstrates the process of editing the thesaurus concept "stomach". Changes to the edited concept data are saved to CT using the *Save Changes* button.

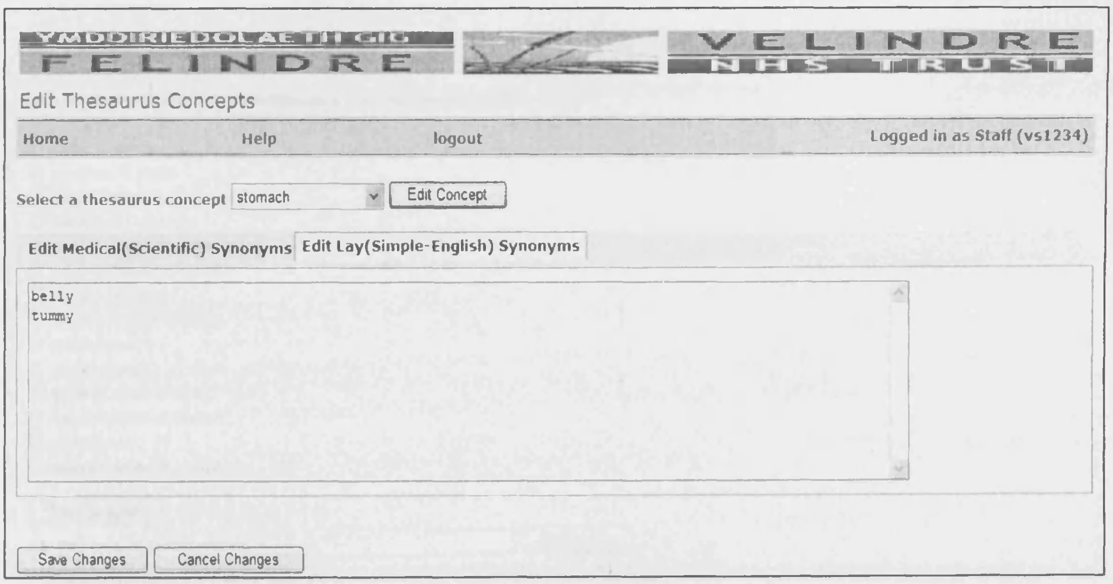


Figure D.13: Editing Thesaurus Concept

D.7 Verifying Patient Diagnosis Ontology (PDO)

Figure D.14 demonstrates the PDO webpage which allows PDO query, verification and update. All PDO diagnosis concepts are listed in a select list. The *View Related Terms* button queries the PDO server (i.e. ontology RDF model) to retrieve medical, lay and generic terms for a given diagnosis concept.

Figure D.15 presents the medical PDO terms currently encoded for the diagnosis "malignant neoplasm of stomach". Corresponding lay and generic terms can be accessed by clicking on the relevant given tabs. Each term category is stored in a tab page panel. The information staff checks each term category and verifies terms' descriptions correctness and meaningfulness. This is to ensure that correct term descriptions are utilised by patients in the Patient GUI.

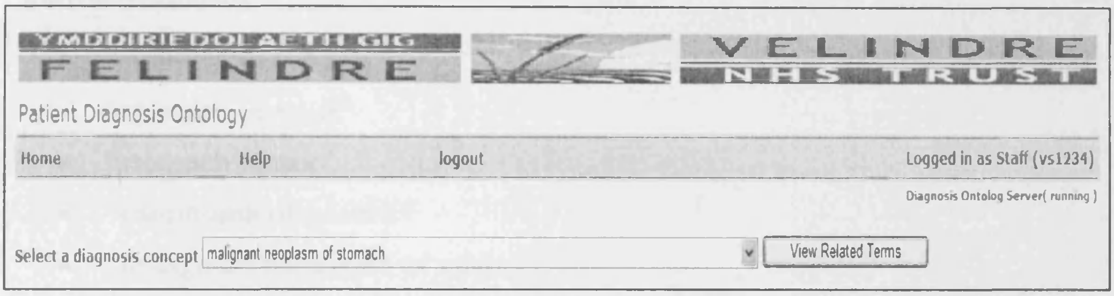


Figure D.14: Diagnosis Ontology Server in Running Status

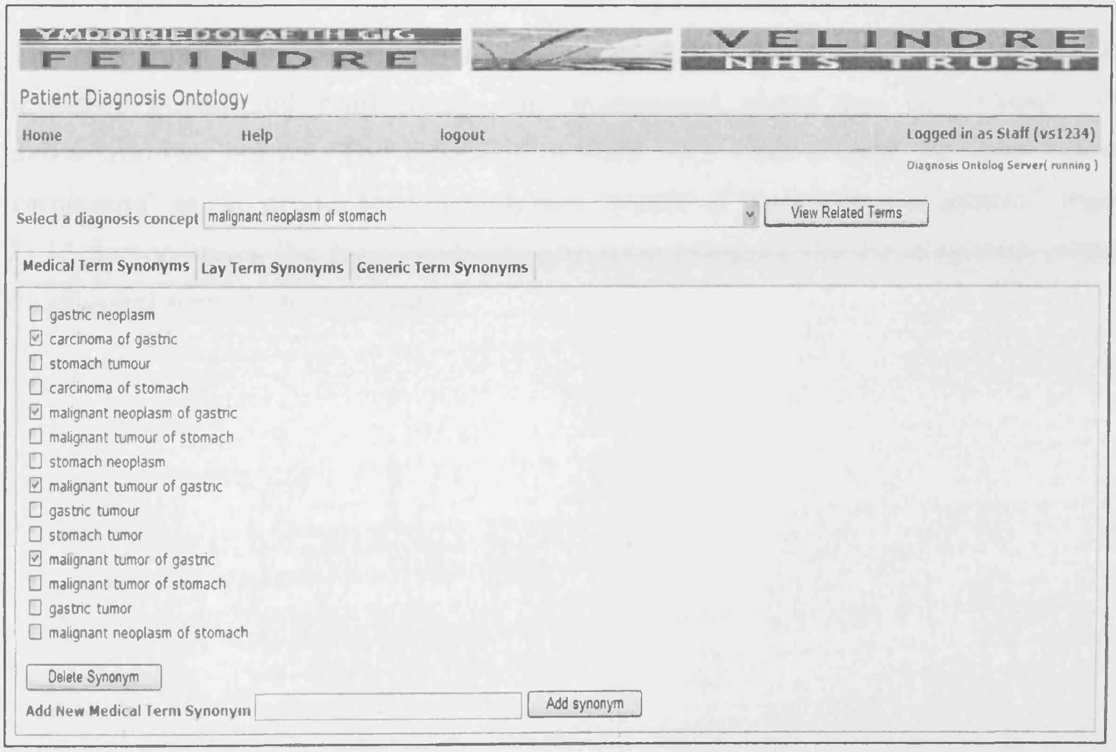


Figure D.15: Diagnosis Ontology Server in Running Status

Two operations are offered to update term synonyms in each of the medical, lay and generic term categories:

- ***deleteSynonym***: deletes (a malformed or incorrect description of) a diagnosis term synonym from a given diagnosis term category.
- ***addSynonym***: adds the correct form of a malformed diagnosis term synonym or a new diagnosis lay, medical or generic synonym. In addition, this operation enables the addition of diagnosis terms not covered by the ISCO system and/or generated from CT.

For instance, the medical synonym mapping between "gastric" and "stomach" in CT produced the following newly constructed medical term synonyms for the diagnosis "malignant neoplasm of stomach" and its ISCO medical term synonym "gastric neoplasm":

- "stomach neoplasm"
- "stomach tumour"
- "stomach tumor"
- "carcinoma of gastric"
- "malignant neoplasm of gastric"
- "malignant tumour of gastric"
- "malignant tumor of gastric"

The first three terms are linguistically acceptable while the last four are not and therefore considered malformed. The malformed terms can be deleted using *deleteSynonym* button. The information staff may want to add the term "gastric carcinoma" as the proper term description instead of "carcinoma of gastric". Figure D.16 demonstrates the final medical synonyms category for the diagnosis concept "malignant neoplasm of stomach".

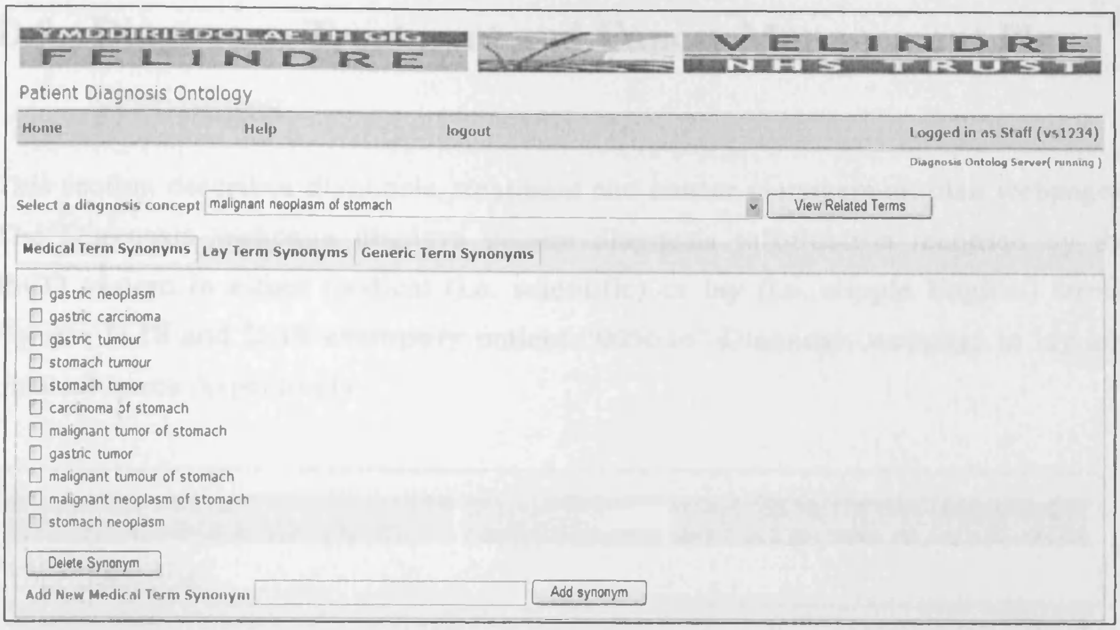


Figure D.16: Verifying PDO Medical Terms for Diagnosis "malignant neoplasm of stomach"

D.8 Uploading Patient Diagnosis Ontology (PDO)

When PDO is created, it is stored in a file system. However, if PDO is not yet created or its file can not be found, then the system indicates that the ontology server is not available and displays the *Upload* button to create and upload the diagnosis ontology (Figure D.17). This operation creates the diagnosis ontology model and makes it available to browse and update.

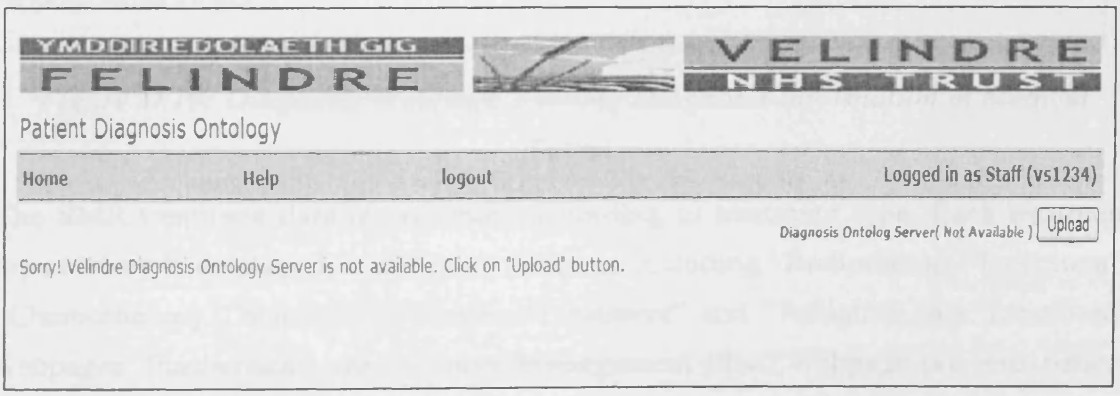


Figure D.17: Uploading Diagnosis Ontology Server

D.9 Diagnoses, Treatment and Cancer Management Plan Webpages

This section describes diagnosis, treatment and cancer management plan webpages. The Diagnosis webpage displays patient diagnosis information recorded by the ISCO system in either medical (i.e. scientific) or lay (i.e. simple English) terms. Figures D.18 and D.19 exemplify patient "00561c" Diagnosis webpage in lay and medical terms respectively.

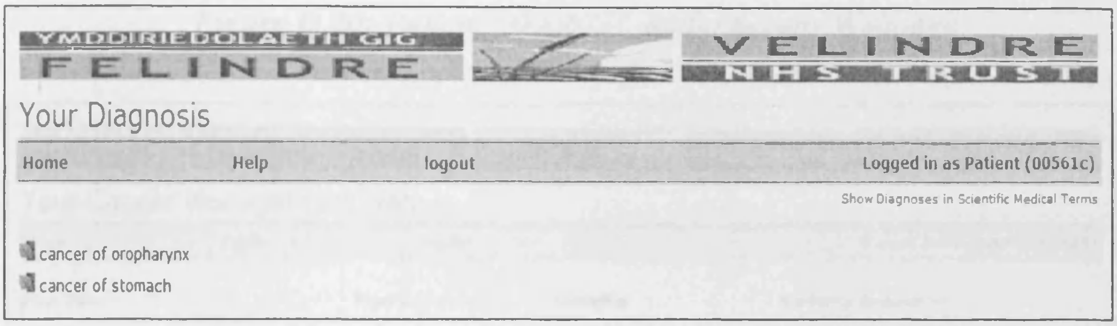


Figure D.18: Diagnosis Webpage showing Diagnosis Information in Lay Terms

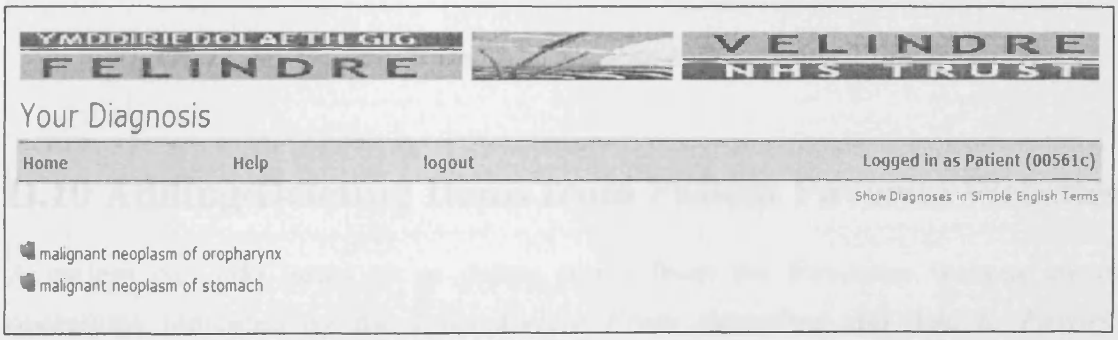
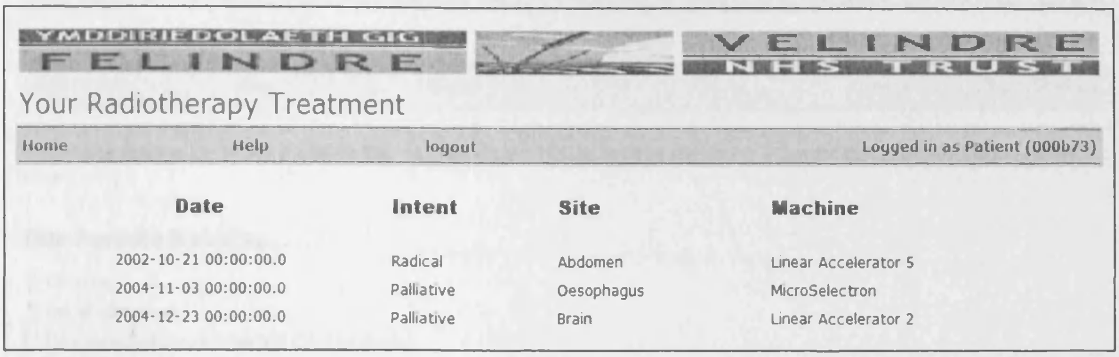


Figure D.19: Diagnosis Webpage showing Diagnosis Information in Medical Terms

The SMR treatment data is presented according to treatment type. Each treatment type history is displayed in different webpage including "Radiotherapy Treatment", "Chemotherapy Treatment", "Surgery Treatment" and "Palliative care Treatment" webpages. Furthermore, the "Cancer Management Plan" webpage presents patient proposed treatment. Figures D.20 and D.21 exemplify a patient radiotherapy treatment Webpage and cancer management plan webpage.



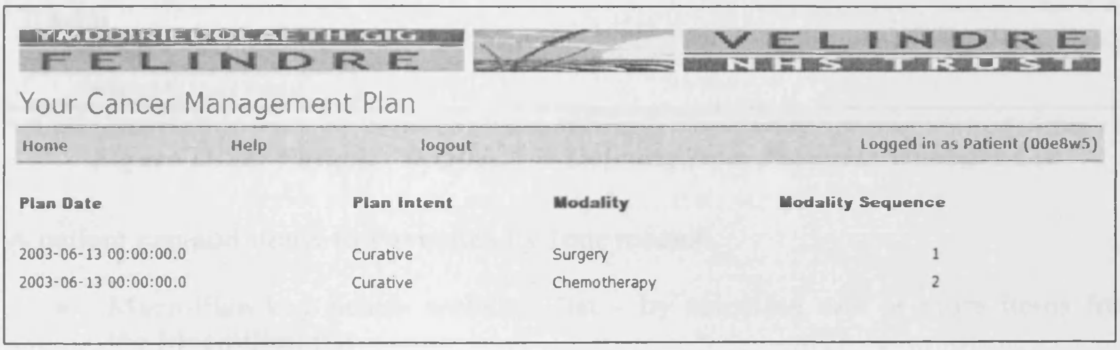
YMDDIRIEDOLAETH GIG FELINDRE VELINDRE NHS TRUST

Your Radiotherapy Treatment

Home Help logout Logged in as Patient (000b73)

Date	Intent	Site	Machine
2002-10-21 00:00:00.0	Radical	Abdomen	Linear Accelerator 5
2004-11-03 00:00:00.0	Palliative	Oesophagus	MicroSelectron
2004-12-23 00:00:00.0	Palliative	Brain	Linear Accelerator 2

Figure D.20: Patient "000b73" Radiotherapy Webpage



YMDDIRIEDOLAETH GIG FELINDRE VELINDRE NHS TRUST

Your Cancer Management Plan

Home Help logout Logged in as Patient (00e8w5)

Plan Date	Plan Intent	Modality	Modality Sequence
2003-06-13 00:00:00.0	Curative	Surgery	1
2003-06-13 00:00:00.0	Curative	Chemotherapy	2

Figure D.21: Patient "00e8w5" Cancer Management Plan Webpage

D.10 Adding/Deleting Items from Patient Favorite Websites

A patient can add items to or delete items from the Favorites website by two operations indicated by the links *Delete From Favorites* and *Add to Favorites* respectively. One or more items can be deleted from a patient's Favorites as exemplified in Figure D.22.

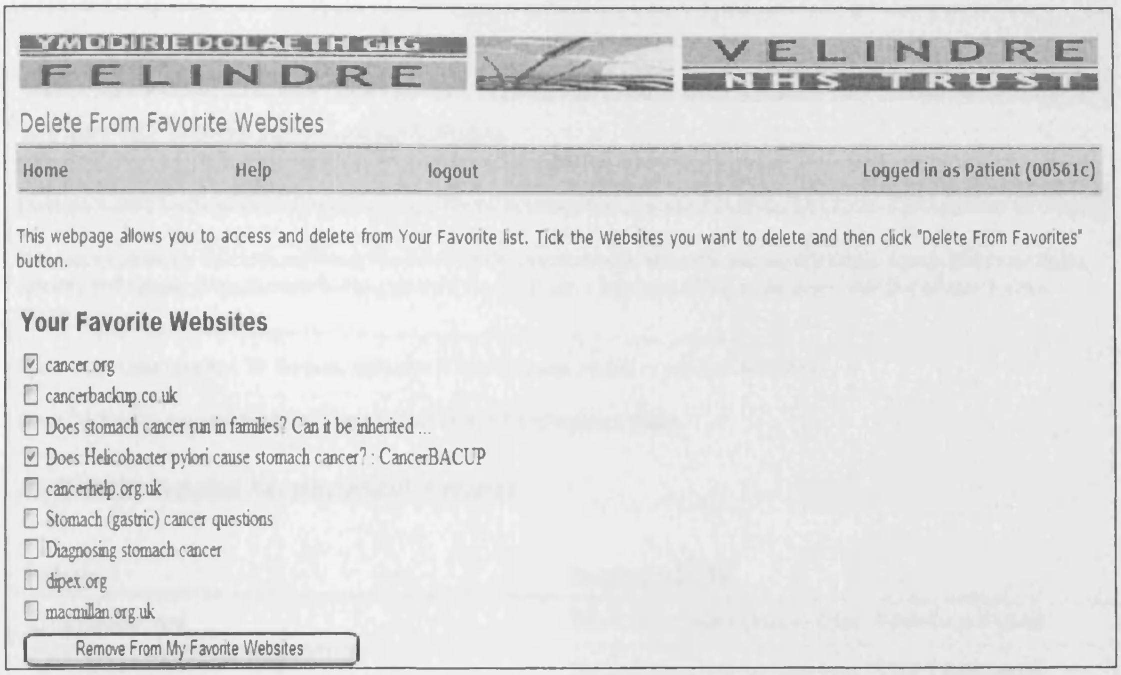
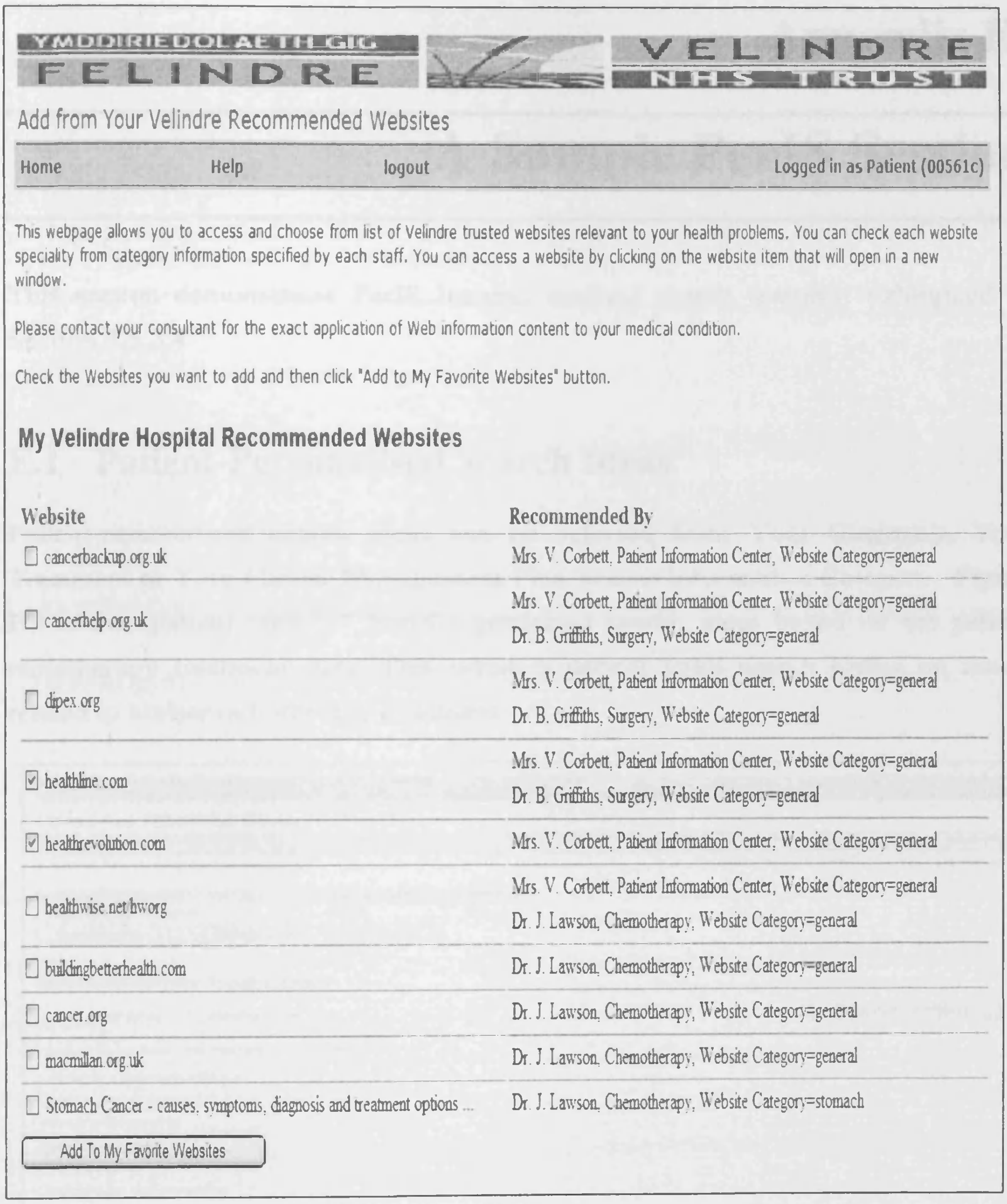


Figure D.22: Patient "00561c" - Deleting from Favorite Websites List

A patient can add items to Favorites by four means:

- Macmillan key health websites list – by selecting one or more items from the Macmillan list.
- HTW list – by selecting one or more items from the HTW list.
- Entering the patient's own websites– by entering one or more items in a textarea.
- PerIS Internet search results from PHB internal search tools – by selecting one or more items from PerIS search results.

Figures D.23 demonstrates the process of adding items to Favorites from the HTW list.



YMDDIRIEDOLAETH GIG FELINDRE — VELINDRE NHS TRUST

Add from Your Velindre Recommended Websites

Home | Help | logout | Logged in as Patient (00561c)

This webpage allows you to access and choose from list of Velindre trusted websites relevant to your health problems. You can check each website speciality from category information specified by each staff. You can access a website by clicking on the website item that will open in a new window.

Please contact your consultant for the exact application of Web information content to your medical condition.

Check the Websites you want to add and then click "Add to My Favorite Websites" button.

My Velindre Hospital Recommended Websites

Website	Recommended By
☐ cancerbackup.org.uk	Mrs. V. Corbett, Patient Information Center, Website Category=general
☐ cancerhelp.org.uk	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
☐ dipex.org	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
☑ healthline.com	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. B. Griffiths, Surgery, Website Category=general
☑ healthrevolution.com	Mrs. V. Corbett, Patient Information Center, Website Category=general
☐ healthwise.net/hworg	Mrs. V. Corbett, Patient Information Center, Website Category=general Dr. J. Lawson, Chemotherapy, Website Category=general
☐ buildingbetterhealth.com	Dr. J. Lawson, Chemotherapy, Website Category=general
☐ cancer.org	Dr. J. Lawson, Chemotherapy, Website Category=general
☐ macmillan.org.uk	Dr. J. Lawson, Chemotherapy, Website Category=general
☐ Stomach Cancer - causes, symptoms, diagnosis and treatment options ...	Dr. J. Lawson, Chemotherapy, Website Category=stomach

Add To My Favorite Websites

Figure D.23: Patient "00561c" - Adding to Favorite Websites List from the HTW List

Appendix E

A Sample PerIS Session

This section demonstrates PerIS Internet medical search features highlighted in Section 7.5.3.4

E.1 Patient-Personalised Search Ideas

Patient-personalised search ideas can be selected from Your Diagnosis, Your Treatment or Your Cancer Management Plan Search Information Category. Figure E.1 shows patient "00b73" PerIS's generated search ideas based on the patient radiotherapy treatment data. This offers a patient valid search topics on issues related to his/her radiotherapy treatment.

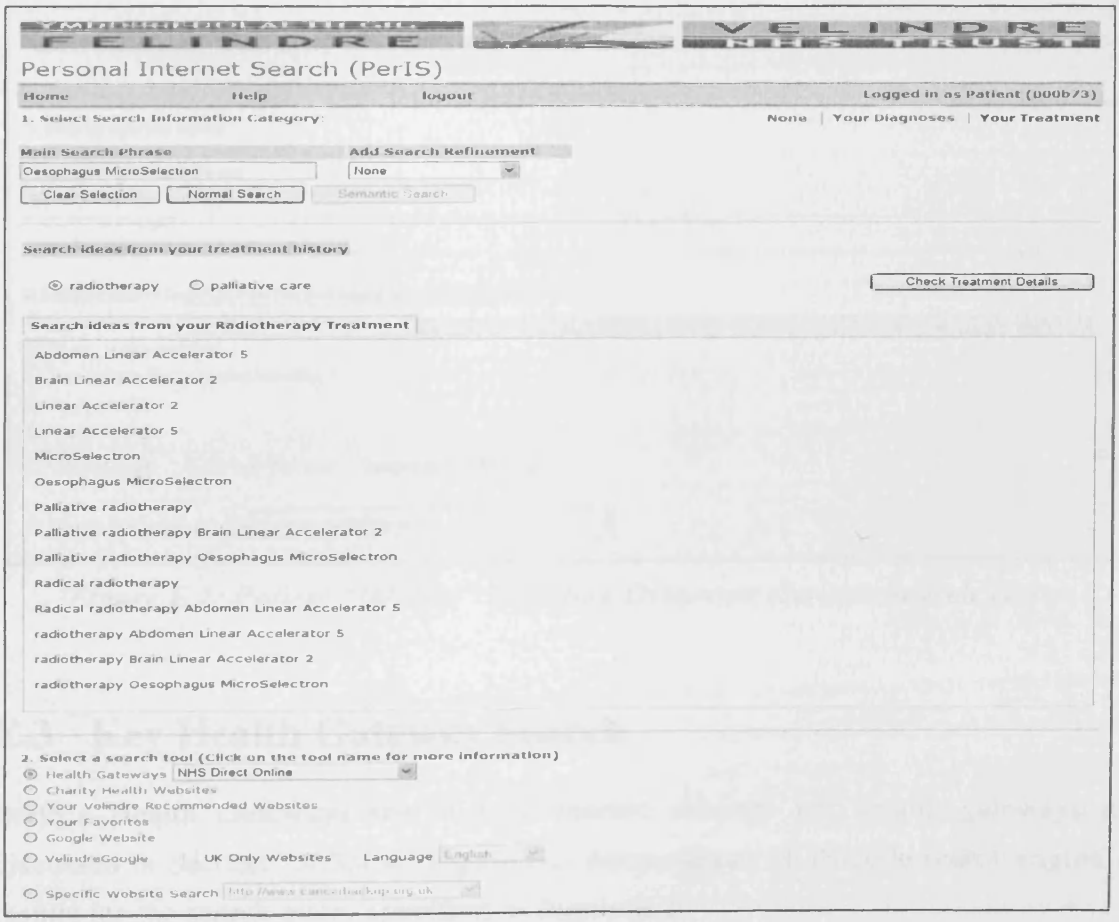


Figure E.1: Patient "00b73" Radiotherapy Treatment-based Search Ideas

E.2 Search ideas from Diagnosis Search Term Enrichment

A patient may wish to search using a given diagnosis related lay, medical or generic term. Figure E.2 illustrates a search exercise for the generic term "upper gastrointestinal cancer".

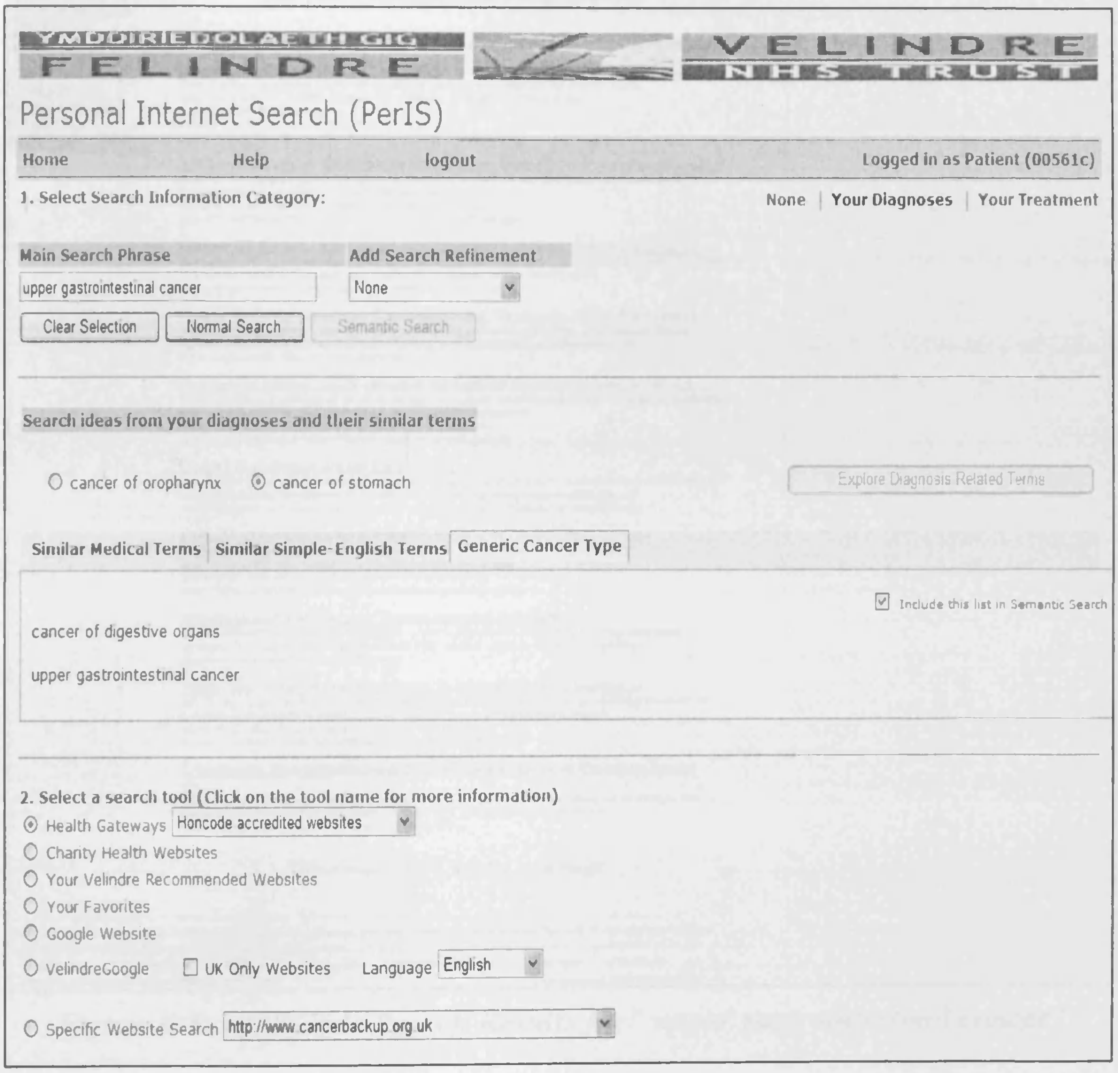


Figure E.2: Patient "00561c" Selecting Diagnosis Generic Search Term

E.3 Key Health Gateway Search

PerIS's Health Gateways search tool execute external key health gateways as discussed in Section 7.4.3.4.4. Figure E.3 demonstrates HONCode search engine's results for the search query specified in Figure E.2.

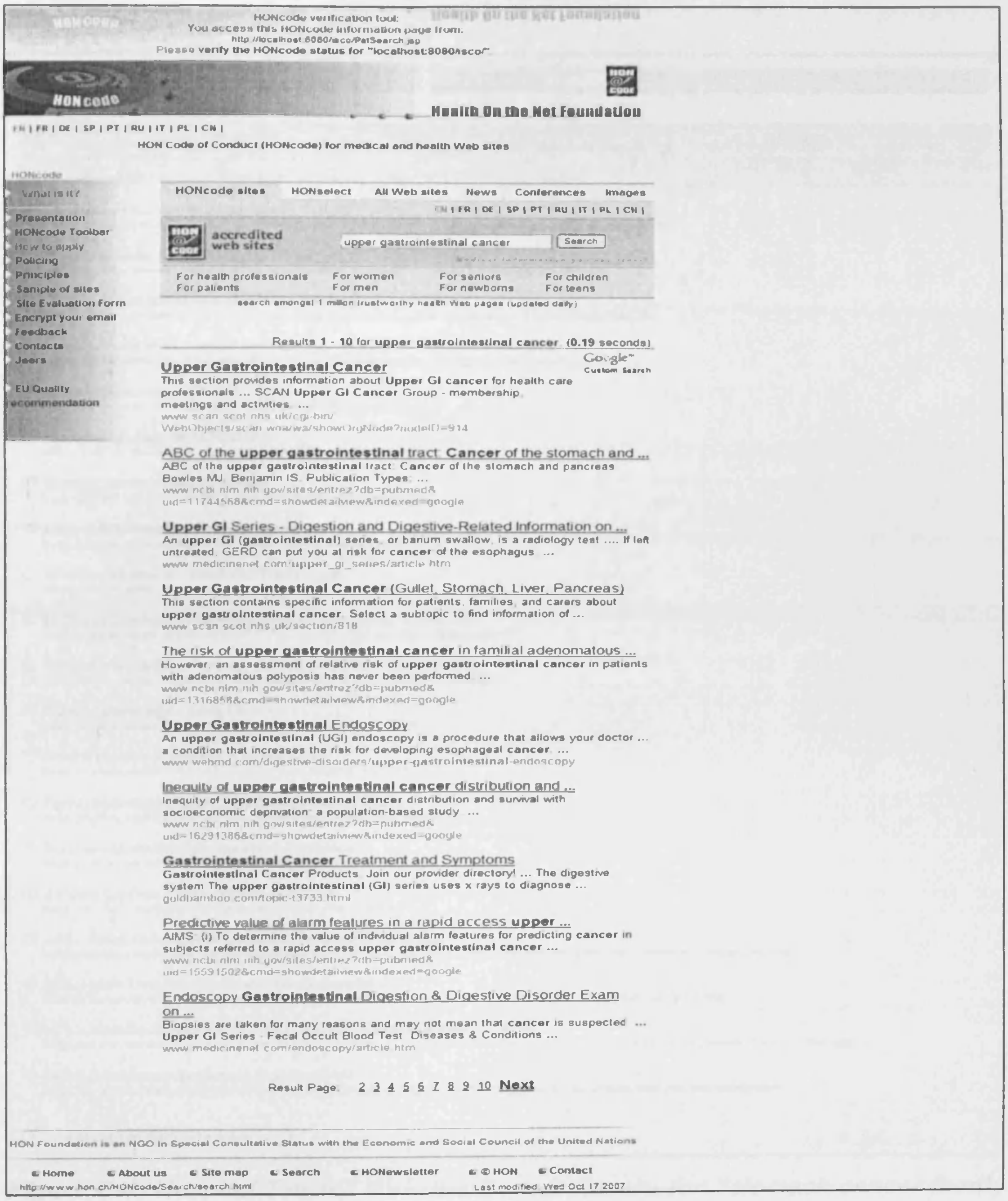


Figure E.3: HONCode Search Results for" upper gastrointestinal cancer"

E.4 Hospital Trusted Websites Search

Figure E.4 illustrates search results of "stomach cancer family risk" on patient "00561c" Hospital Trusted Websites. This query specifies a lay diagnosis search term, i.e. "stomach cancer", a search refinement, i.e. "family risk" and the search tool "Your Velindre Recommended Websites".

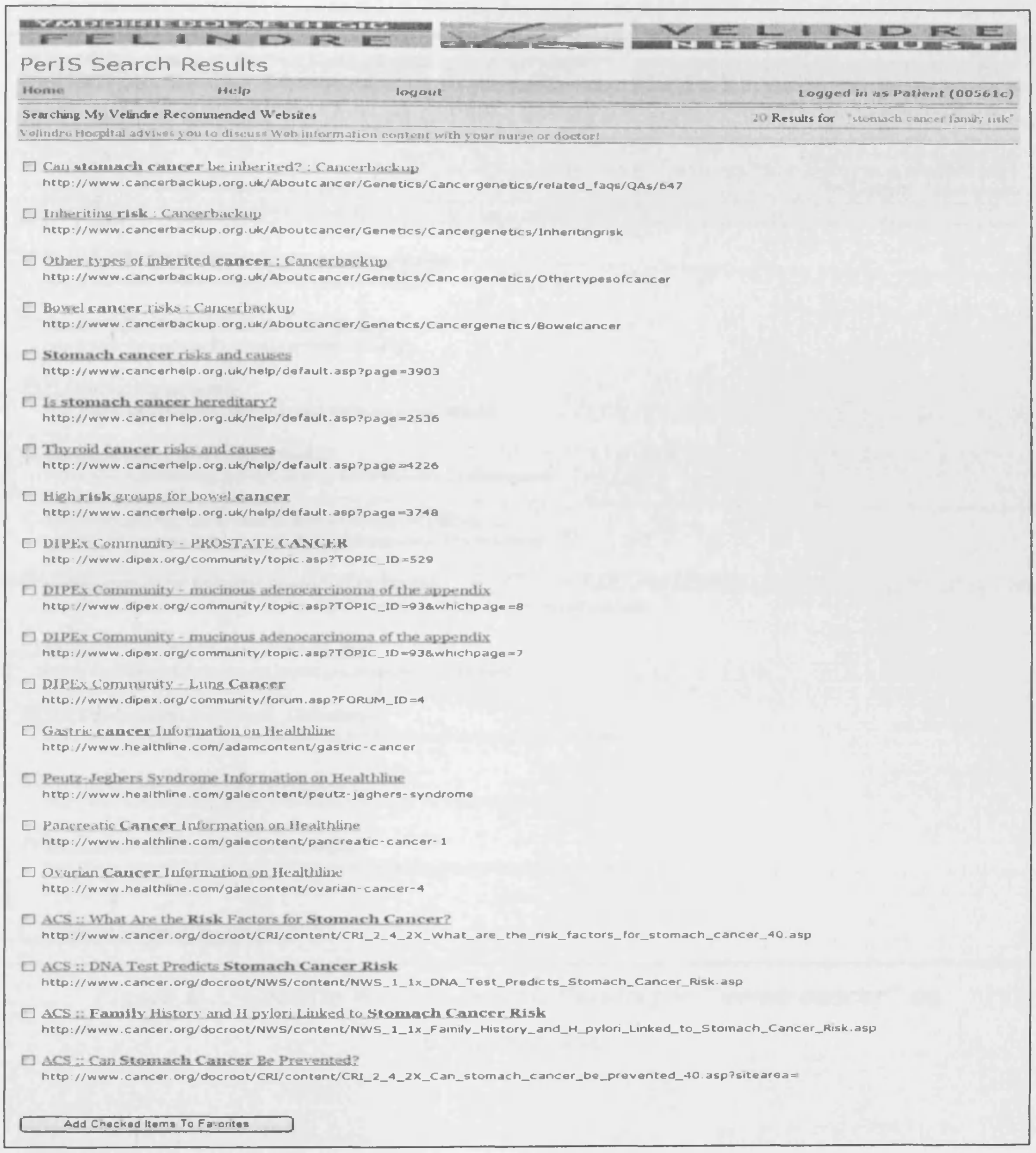


Figure E.4: Hospital Trusted Websites Search Results for "stomach cancer family risk"

E.5 Specific Website Search

Specific Website search is useful when wanting to search one single, assumingly popular, health website (or gateway) like "cancerbackup.co.uk". Figure E.5 demonstrates PerIS's Specific Website search for "womb cancer" on "cancerbackup.org.uk".

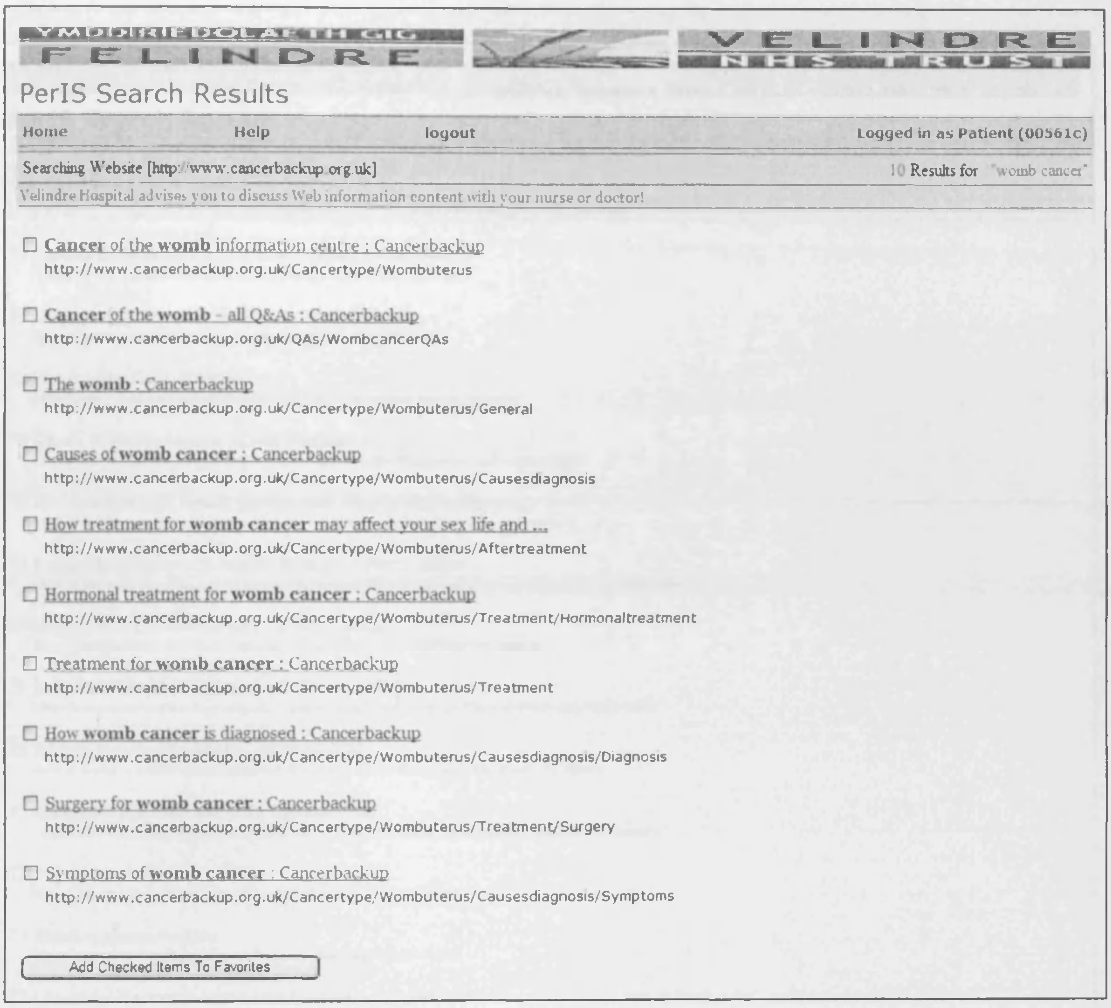


Figure E.5: Specific Website Search Results for "womb cancer" on "Cancerbackup"

E.6 Charity Websites Search

PerIS's Charity Websites search tool searches a key charity websites list. This tool is useful if a patient wishes to search trusted but non-official patient-oriented health websites. PerIS Charity websites search for "womb cancer" is given in Figure E.6. Two charity websites are currently incorporated in the list; "Cancerbackup.org.uk[24]" and "cancerhelp.org.uk[25]".

[24] Cancerbackup.org.uk is Europe's leading cancer information service, aimed at patients and their carers.

[25] Cancerhelp.org.uk is voted best health site and most popular health site in 2006. It is also a holder of Plain English Crystal mark.

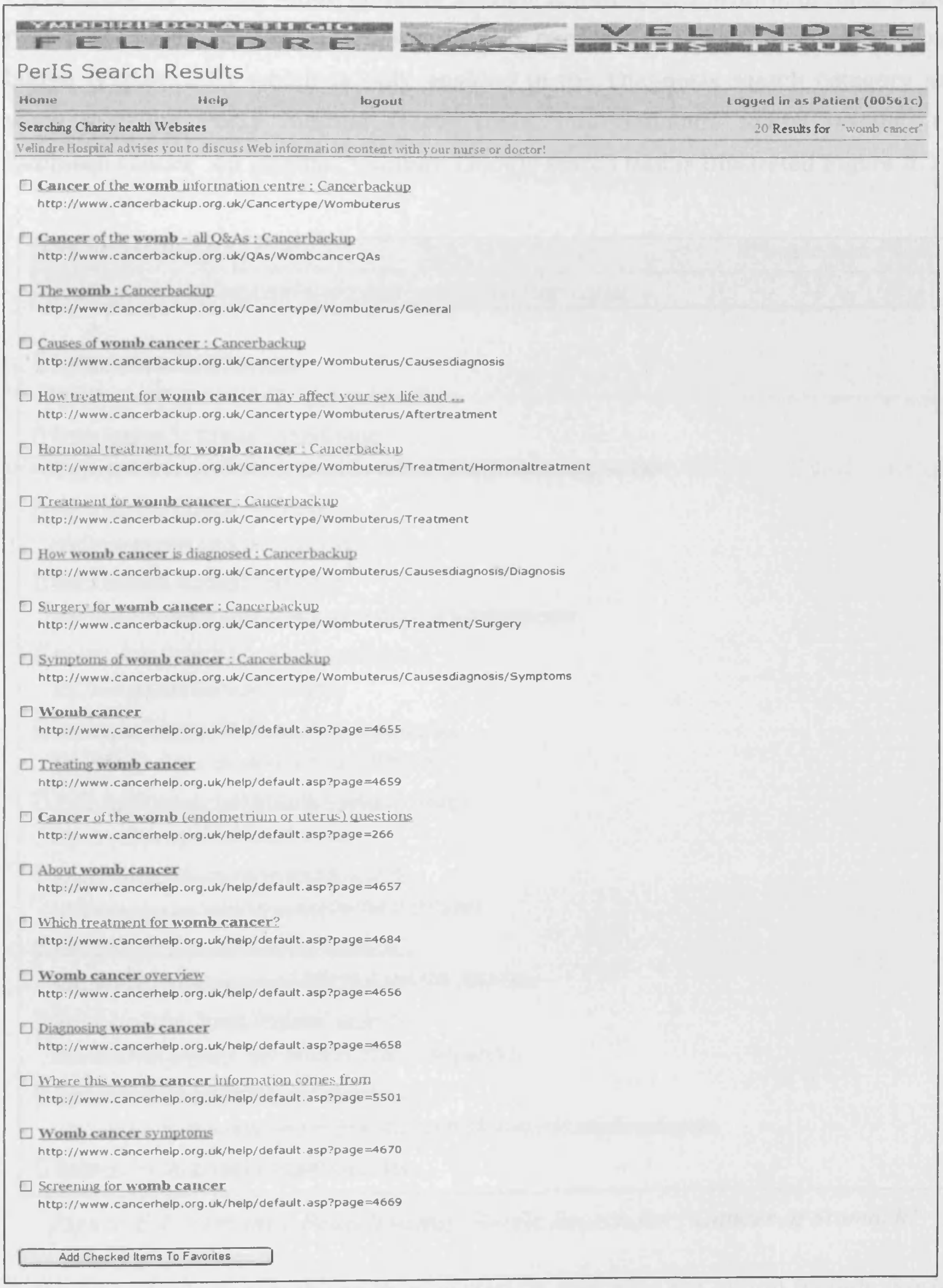


Figure E.6: Charity Websites Search Results for "womb cancer"

E.7 PerIS Semantic Search

PerIS semantic search mode enriches normal search results with additional search results based on related search terms. It is executed using the *Semantic Search* button (Figure E.2) which is only enabled in the Diagnosis search category and when selecting PerIS internal search tools. Full Semantic search results[26] for "stomach cancer" on internal Velindre Google search tool is illustrated Figure E.7.

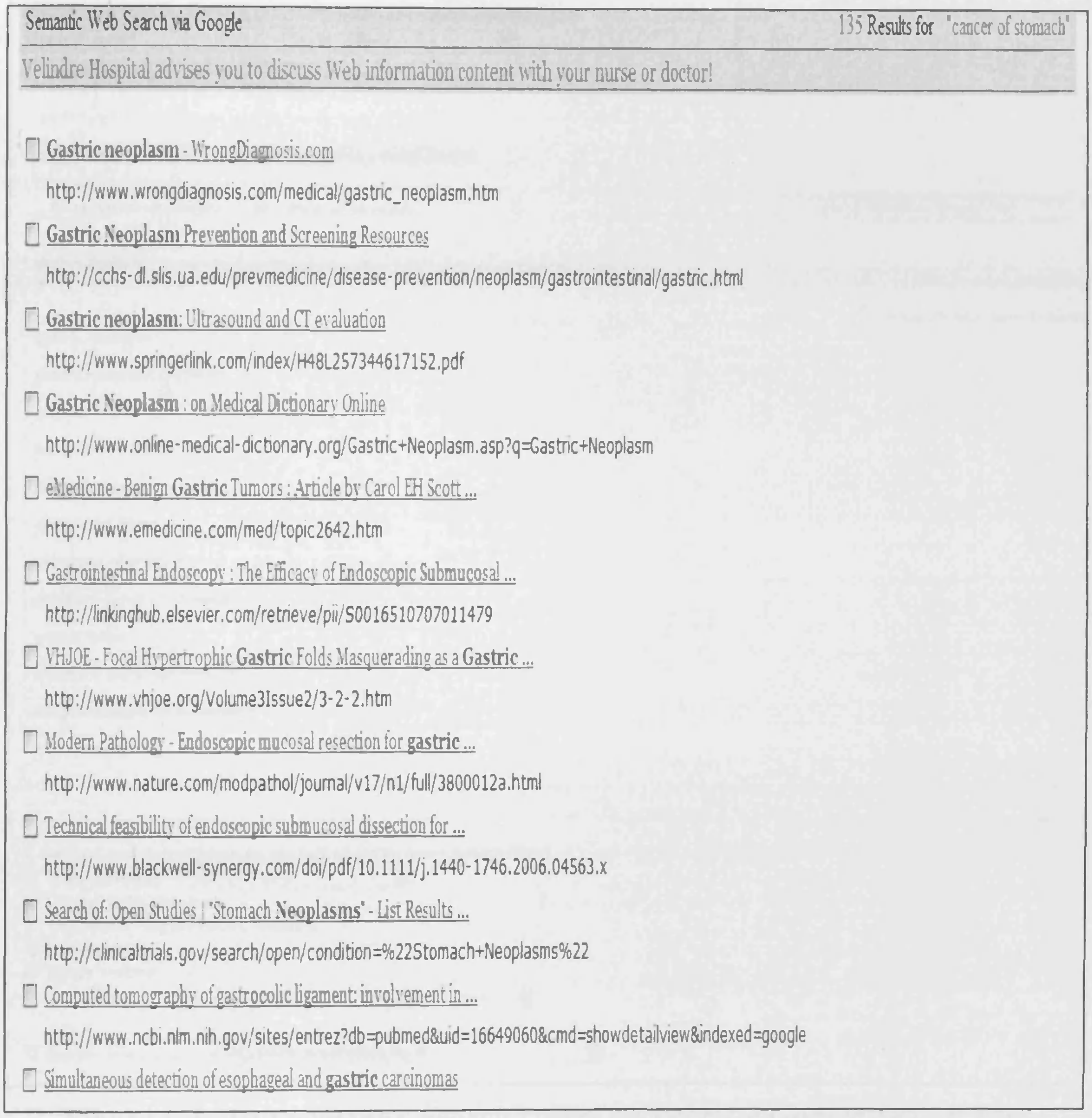


Figure E.7: Semantic Search using Google Search for "Cancer of Stomach"

A lay only semantic search can be executed by unticking the option *Include this list in Semantic Search* (see Figure E.8) from medical and generic diagnosis term tabs.

[26] For the demonstration reason, Figure E.7 only displays the first search results page as fitting all 135 search results in one figure makes them unreadable.

Lay semantic search results for "cancer of stomach" on Velindre Google search is shown in Figure E.9.

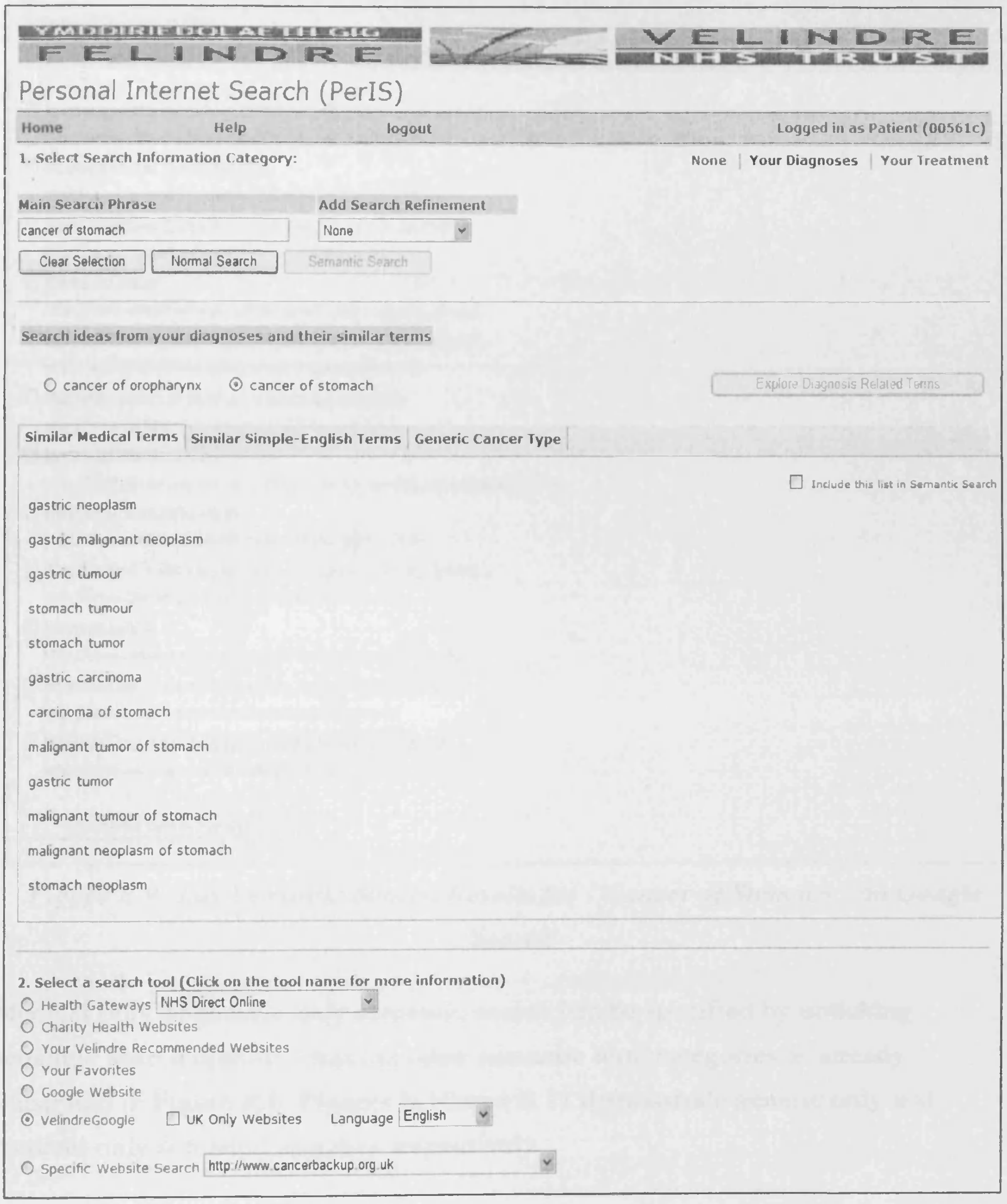


Figure E.8: Lay Semantic Search Query on Velindre Google Search Tool

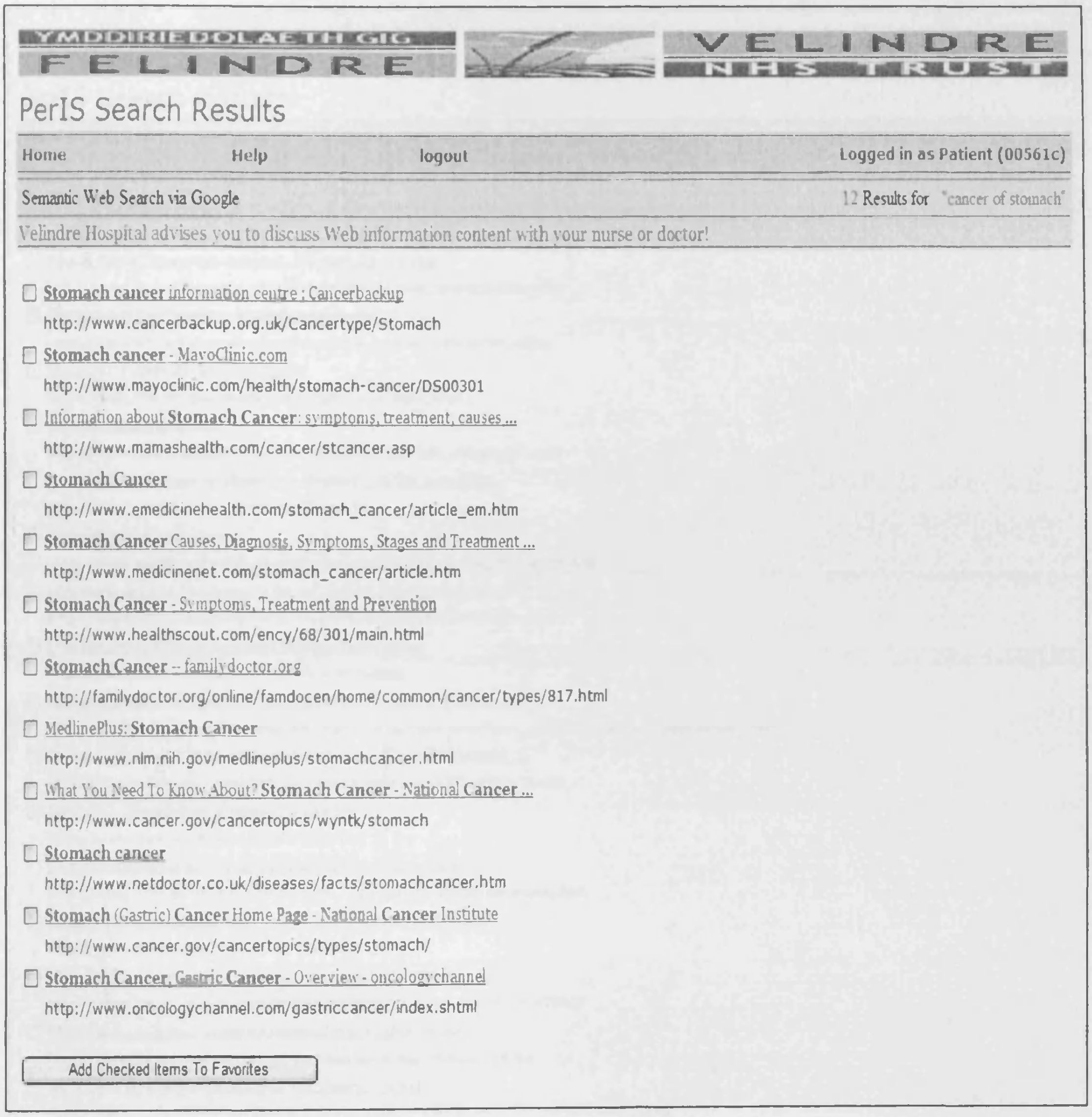
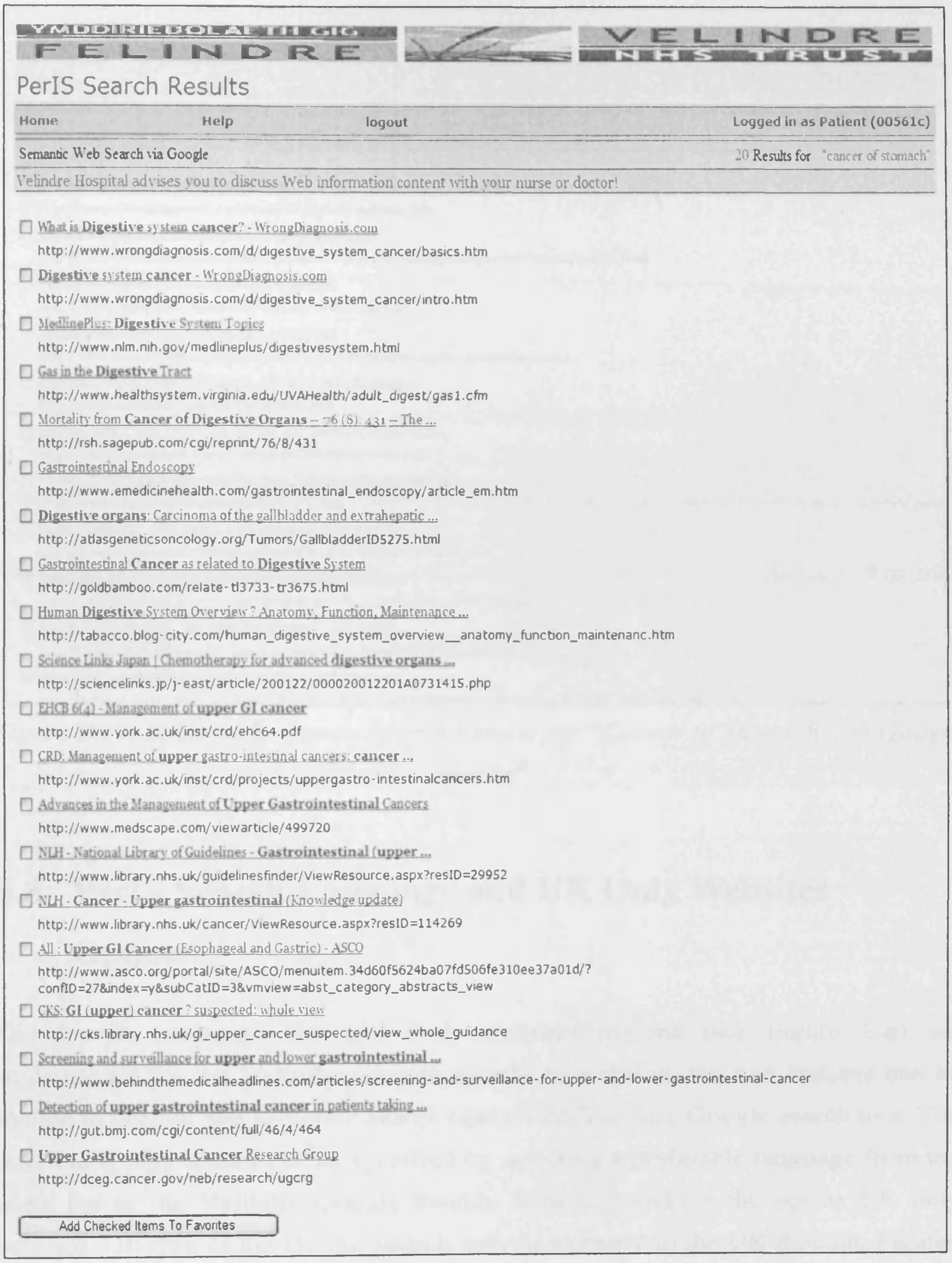


Figure E.9: Lay Semantic Search Results for "Cancer of Stomach" on Google Search

Medical only or generic only semantic search can be specified by unticking semantic search options from the other semantic term categories as already illustrated in Figure E.8. Figures E.10 and E.11 demonstrate generic only and medical only semantic searches respectively.


YMDDIRIEDOLAETH GIG FELINDRE | VELINDRE NHS TRUST

PerIS Search Results

Home | Help | logout | Logged in as Patient (00561c)

Semantic Web Search via Google | 20 Results for "cancer of stomach"

Velindre Hospital advises you to discuss Web information content with your nurse or doctor!

- What is Digestive system cancer? - WrongDiagnosis.com
  http://www.wrongdiagnosis.com/d/digestive_system_cancer/basics.htm
- Digestive system cancer - WrongDiagnosis.com
  http://www.wrongdiagnosis.com/d/digestive_system_cancer/intro.htm
- MedlinePlus: Digestive System Topics
  http://www.nlm.nih.gov/medlineplus/digestivesystem.html
- Gas in the Digestive Tract
  http://www.healthsystem.virginia.edu/UVAHealth/adult_digest/gas1.cfm
- Mortality from Cancer of Digestive Organs -- 76 (8): 431 -- The ...
  http://rsh.sagepub.com/cgi/reprint/76/8/431
- Gastrointestinal Endoscopy
  http://www.emedicinehealth.com/gastrointestinal_endoscopy/article_em.htm
- Digestive organs: Carcinoma of the gallbladder and extrahepatic ...
  http://atlasgeneticsoncology.org/Tumors/GallbladderID5275.html
- Gastrointestinal Cancer as related to Digestive System
  http://goldbamboo.com/relate-tl3733-tr3675.html
- Human Digestive System Overview ? Anatomy, Function, Maintenance ...
  http://tabacco.blog-city.com/human_digestive_system_overview__anatomy_function_maintenanc.htm
- Science Links Japan | Chemotherapy for advanced digestive organs ...
  http://sciencelinks.jp/j-east/article/200122/000020012201A0731415.php
- EHCB 6(4) - Management of upper GI cancer
  http://www.york.ac.uk/inst/crd/ehc64.pdf
- CRD: Management of upper gastro-intestinal cancers: cancer ...
  http://www.york.ac.uk/inst/crd/projects/uppergastro-intestinalcancers.htm
- Advances in the Management of Upper Gastrointestinal Cancers
  http://www.medscape.com/viewarticle/499720
- NLH - National Library of Guidelines - Gastrointestinal (upper ...
  http://www.library.nhs.uk/guidelinesfinder/ViewResource.aspx?resID=29952
- NLH - Cancer - Upper gastrointestinal (Knowledge update)
  http://www.library.nhs.uk/cancer/ViewResource.aspx?resID=114269
- All : Upper GI Cancer (Esophageal and Gastric) - ASCO
  http://www.asco.org/portal/site/ASCO/menuitem.34d60f5624ba07fd506fe310ee37a01d/?confID=27&index=y&subCatID=3&vmview=abst_category_abstracts_view
- CKS: GI (upper) cancer ? suspected: whole view
  http://cks.library.nhs.uk/gi_upper_cancer_suspected/view_whole_guidance
- Screening and surveillance for upper and lower gastrointestinal ...
  http://www.behindthemedicalheadlines.com/articles/screening-and-surveillance-for-upper-and-lower-gastrointestinal-cancer
- Detection of upper gastrointestinal cancer in patients taking ...
  http://gut.bmj.com/cgi/content/full/46/4/464
- Upper Gastrointestinal Cancer Research Group
  http://dceg.cancer.gov/neb/research/ugcrg

Add Checked Items To Favorites

Figure E.10: Generic Semantic Search Results for "Cancer of Stomach" on Google Search

Figure E.11: Medical Semantic Search Results for "Cancer of Stomach" on Google Search[27]

E.8 PerIS Search Language and UK Only Websites Options

The *Search language* and *UK Only websites* options (see Figure E.8) are implemented for the Velindre Google search. In addition, the two features can be applied to normal and semantic search options on Velindre Google search tool. The search language option can be specified by selecting a preferable language from the select list of the Velindre Google Search. Similarly, ticking the option UK only websites will restrict the Google search only to websites in the UK domain. Figures E.12 shows search results for "stomach cancer" in Spanish whereas Figure E.13 illustrates lay semantic search results for UK only websites on Google for "cancer of stomach".

[27] For the demonstration reason, Figure E.11 only displays the first search results page due to the large search result size (105).

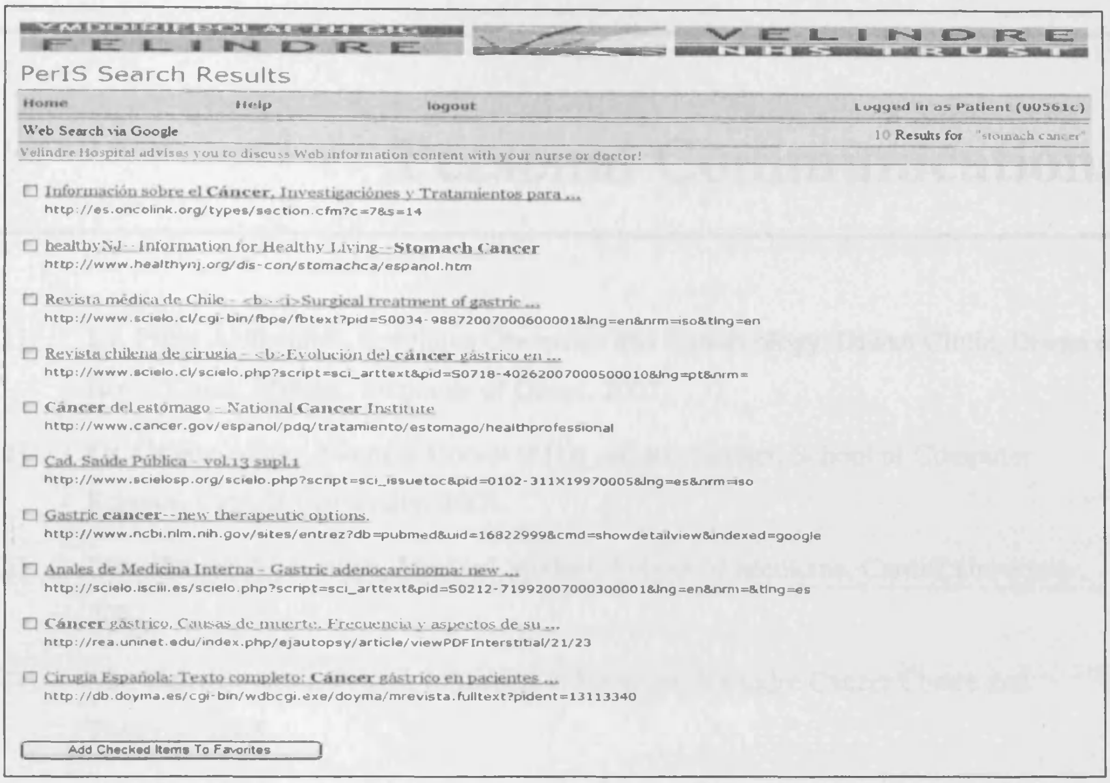


Figure E.12: Google Search Results for "Cancer of Stomach" in Spanish

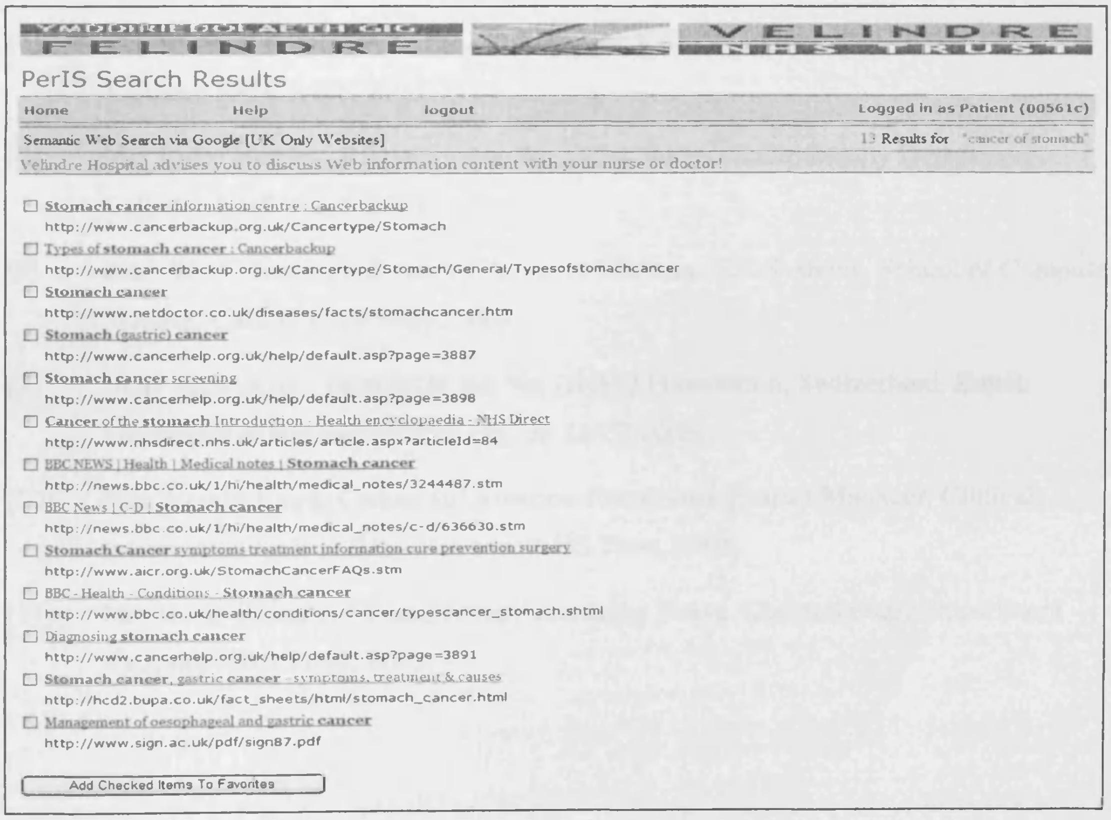


Figure E.13: UK Only Lay Semantic Search Results on Google for "Cancer of Stomach"

Personal Communications

[1] Dr. Fikra Al-Busaidi, Specialist Obstetrics and Gynaecology, Diwan Clinic, Diwan of Royal Court. Muscat, Sultanate of Oman, 2007.

[2] Dr. Omnia Allam, Medical Doctor (MD) and Researcher. School of Computer Science, Cardiff University, 2005.

[3] Miss Zainab Allawatiya, Medical Student, School of Medicine, Cardiff University, 2007.

[4] Mrs Valerie Aston, Patient Information Manager, Velindre Cancer Centre and Tenovus, 2005.

[5] Mrs Hazel Bailey, Information Analysis, Clinical Information Unit (CIU) - Velindre NHS Trust, 2005.

[6] Dr. Anthony Bater, ISCO Project Manager, Clinical Information Unit (CIU) - Velindre NHS Trust, 2005.

[7] Mrs Joyce Butters, Radiotherapy Specialist Nurse, Radiotherapy Department - Velindre NHS Trust, 2005.

[8] Prof. W. A. Gray, Professor of Advanced Information Systems, School of Computer Science, Cardiff University, 2007.

[9] HON secretariat , Health On the Net (HON) Foundation, Switzerland. Email: <honsecretariat@healthonnet.org, on 13/02/2006.

[10] Mrs Wendy Jones, Cancer Information Framework Project Manager, Clinical Information Unit (CIU) - Velindre NHS Trust, 2005.

[11] Mrs Rosie Roberts, Chemotherapy Specialist Nurse, Chemotherapy Department - Velindre NHS Trust, 2005.

Refereed Papers

[1] A. Al-Busaidi, "Supporting Multiple Integrated Views in a Web environment," presented at *The British National Conference on Databases (BNCOD 21) - Doctoral Consortium*, Edinburgh, UK, 2004.

[2] A. Al-Busaidi, A. Gray, and N. Fiddian, "Personalising Web Information for Patient: Linking Patient Medical Data with the Web via a Patient Personal Knowledge Base," presented at *The International Symposium on Health Information Management Research (ISHIMR05)*, Thessaloniki, Greece, 2005.

[3] A. Al-Busaidi, A. Gray, and N. Fiddian, "Personalizing web information for patients: linking patient medical data with the web via a patient personal knowledge base," *Health Informatics Journal*, vol. 12(1), 27-39, 2006.

[4] A. Al-Busaidi, A. Gray, and N. Fiddian, "Investigating and Utilising Patient Information Needs to Focus Internet Searching for Cancer Patients," presented at *The International Symposium for Health Information Management Research (ISHIMR06).* Halifax, Nova Scotia, Canada, 2006.

[5] M. Burgess, O. Allam, A. Gray, and A. Al-Busaidi, "Patient-Centred Approach to Focusing Online Health Information Search Results," presented at ISePHR (2006), 2006.

Bibliography

[1] "130,000 Ealing patients to receive health information by mobile phone." http://www.carelink.co.uk/templates/page.aspx?id=1326, (Accessed 12/12/06)

[2] "American Health Information Management Association (AHIMA)." URL: http://www.ahima.org/, (Accessed 25/01/07)

[3] "The American Society for Testing and Materials (ASTM) International." http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3, (Accessed 28/01/07)

[4] "Analysis of 9th HON survey of health and medical Internet users," *Health on the Net Foundation (HON)*, 2005. http://www.hon.ch/Survey/Survey2005/res.html, (Accessed 13/11/05)

[5] "The Apache Jakarta Project." http://jakarta.apache.org/tomcat/, (Accessed 01/05/04)

[6] "Babylink, Scotland." http://www.babylink.info/edinburgh/NeonatalUnit/Welcome1.aspx, (Accessed 04/02/07)

[7] "Better Health Information Provision Will Lead to Improved Patient Choice," *Health Informatics Commuinty, NHS*, 2006. http://www.informatics.nhs.uk/cgi-bin/item.cgi?id=2087, (Accessed 01/12/06)

[8] "BMP2: Data Level Integration." http://www.bpm2.com/index.php?option=com_content&task=view&id=89&Itemid=0, (Accessed 11/03/07)

[9] "British Heart Foundation." http://www.bhf.org.uk/, (Accessed 07/03/07)

[10] "The British National Conference on Databases (BNCOD 21). Edinburgh, UK.," 2004.

[11] "Calman-Hine Report," *Department of Health (DOH), UK*, 1995. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4071083, (Accessed 07/02/08)

[12] "Cancer Specialist Library." http://www.library.nhs.uk/cancer/Default.aspx?pagename=HOME, (Accessed 13/04/07)

[13] "CancerBACUP online: cancer information, support & resources." http://www.cancerbacup.org.uk, (Accessed 01/10/05)

[14] "CapMed." http://www.capmed.com, (Accessed 05/02/07)

[15] "Card 3 – Information provision," *The British Medical Association (BMA)*, 2005. https://registration.bma.org.uk/ap.nsf/Content/consenttk2~card3, (Accessed 08/02/08)

[16] "The Center for Information Therapy." http://www.informationtherapy.org/, (Accessed 05/02/08)

[17] "Clinical Knowledge Summaries." http://www.connectingforhealth.nhs.uk/resources/procurement/cks, (Accessed 15/04/08)

[18] "Clinical Knowledge Summaries (CKS)." http://cks.library.nhs.uk/, (Accessed 15/04/08)

[19] "The Cochrane Library." http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME, (Accessed 23/02/07)

[20] "Connecting For Health Unites Over 100 Organizations To Bring American Healthcare System into Information Age," *Markle Foundation Press Center*, 2003. http://www.markle.org/resources/press_center/press_releases/2003/press_release_06052003.php, (Accessed 25/01/07)

[21] "Copying Letters to Patients," *Department of Health (DoH)*. http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/PatientAndPublicInvolvement/CopyingLettersToPatients/fs/en, (Accessed 10/12/06)

[22] "Copying Letters to Patients – Quotes," *Department of Health (DoH)*. http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/PatientAndPublicInvolvement/CopyingLettersToPatients/CopyingLettersToPatientsTookit/CopyingLettersTookitArticle/fs/en?CONTENT_ID=4127738&chk=S/ezSU, (Accessed 10/12/06)

[23] "Criteria for Assessing the Quality of Health Information on the Internet - Policy Paper," *Health Summit Working Group (HSWG), Mitretek Systems*, 2000. http://hitiweb.mitretek.org/docs/policy.html, (Accessed 26/02/06)

[24] "Current Procedural Terminology (CPT)." http://www.ama-assn.org/ama/pub/category/3113.html, (Accessed 08/01/08)

[25] "Data Integration With EAI Tools," *EAI Ittoolbox Popular Q&A Team*. http://eai.ittoolbox.com/documents/popular-q-and-a/data-integration-with-eai-tools-3326, (Accessed 29/10/07)

[26] "Data Protection Act, 1998," *Department of Health*, 2003. http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/RecordsManagement/DataProtectionAct1998Article/fs/en?CONTENT_ID=4000489&chk=VrXoGe, (Accessed 11/01/2007)

[27] "Diagnosis Related Groups (DRGs) and the Medicare Program: Implications for Medical Technology," *A Technical Memorandum. US Congress – Office of Technology Assessment. Washington D.C.*, 1983. http://www.princeton.edu/~ota/disk3/1983/8306/8306.PDF, (Accessed 08/01/08)

[28] "Diagnostic and Statistical Manual of Mental Disorders Fourth Edition Text Revision (DSM-IV-TR)," *American Psychiatric Association*, 2000. http://www.dsmivtr.org/, (Accessed 08/02/08)

[29] "DISCERN, Quality Criteria for Consumer Health Information ". http://www.discern.org.uk/, (Accessed 26/02/06)

[30] "Doctors back NHS 'smart' cards," *BBC*, 2000. http://news.bbc.co.uk/1/hi/health/716461.stml, (Accessed 18/12/06)

[31] "Dublin Core Metadata Element Set, Version 1.1," *Dublin Core Metadata Initaitive*. http://dublincore.org/documents/dces/, (Accessed 08/05/06)

[32] "EAI " *Free On-line Dictionary of Computing (FOLDOC)*. http://foldoc.org/?query=EAI, (Accessed 29/10/07)

[33] "Early Adopters." http://www.connectingforhealth.nhs.uk/systemsandservices/nhscrs/scr/learn/early, (Accessed 15/04/08)

[34] "eEurope 2002: Quality Criteria for Health related Websites." http://www.jmir.org/2002/3/e15/, (Accessed 26/02/06)

[35] "eHealth Code of Ethics," *Internet Healthcare Coalition (IHC)* http://www.ihealthcoalition.org/ethics/ethics.html, (Accessed 26/02/06)

[36] "First Google Health Screenshots." http://blogoscoped.com/archive/2007-08-14-n43.html, (Accessed 16/04/08)

[37] "Freedom of Information Act 2000," *Department of Health*, 2000. http://www.advisorybodies.doh.gov.uk/piag/foi.htm#foia, (Accessed 01/12/06)

[38] "Geisinger Health System." http://www.geisinger.org, (Accessed 23/10/05)

[39] "Google Custom Search Engine." http://google.com/coop/cse, (Accessed 23/06/07)

[40] "Google Health Prototype." http://googlesystem.blogspot.com/2007/08/google-health-prototype.html, (Accessed 16/04/08)

[41] "Google Site Query Modifier." http://www.google.com/help/operators.html (Accessed 12/04/07)

[42] "Google Web Search Api." http://www.google.com/apis/, (Accessed 01/09/04)

[43] "GoogleHealth ". http://healthinformatics.wikispaces.com/GoogleHealth, (Accessed 16/04/08)

[44] "Guidance for Access to Health Records Requests under the Data Protection Act 1998," *Department of Health*, 2003. http://www.advisorybodies.doh.gov.uk/ipu/ahr/dpa1998.pdf, (Accessed 11/01/2007)

[45] "Guidelines for Medical and Health Information Sites on the Internet," *The American Medical Association (AMA).* http://www.ama-assn.org/ama/pub/category/1905.html, (Accessed 26/02/06)

[46] "Health care information on the Internet – Part One," *The British Library.* http://www.bl.uk/collections/health/weblink.html, (Accessed 05/02/06)

[47] "Health care records whenever and wherever you need them," *NHS Care Records Service.* http://www.connectingforhealth.nhs.uk/publications/its_coming_leaflet.pdf, (Accessed 21/12/06)

[48] "Health IT Strategic Framework, Glossary of Selected Terms," *The United States Department of Health and Human Services. Office of the National Coordinator for Health Information Technology (ONC).* http://www.hhs.gov/healthit/glossary.html, (Accessed 02/02/07)

[49] "Health Level Seven," *The American National Standards Institute (ANSI).* http://www.hl7.org, (Accessed 28/01/07)

[50] "Health literacy: a prescription to end confusion," *Institute of Medicine*, Washington DC: National Academy Press, Vol. 2006, April 08, 2004.

[51] "Health On the Net Foundation (HON)." http://www.hon.ch, (Accessed 13/11/05)

[52] "Health Space," *Department of Health.* http://www.dh.gov.uk/PolicyAndGuidance/PatientChoice/Choice/BetterInformationChoicesHealth/BetterChoicesArticle/fs/en?CONTENT_ID=4123085&chk=e8YeBn, (Accessed 10/12/06)

[53] "Health Space," *NHS England.* http://www.healthspace.nhs.uk, (Accessed 08/02/08)

[54] "HealthFrame." http://www.recordsforliving.com/HealthFrame/, (Accessed 17/04/08)

[55] "HealthLine." http://www.healthline.com/index], (Accessed 25/02/07)

[56] "HealthMinder." http://www.healtminder.com, (Accessed 05/02/07)

[57] "HealthTracer." http://www.healthtracer.com, (Accessed 05/02/07)

[58] "HealthVault." http://www.healthvault.com/, (Accessed 14/04/08)

[59] "HealthVault (Microsoft Health Vault)." http://www.johnon.com/413/health-vault.html, (Accessed 15/04/08)

[60] "The HIQuA Tool," *Decision Systems Group, Harvard University.* http://samwise.bwh.harvard.edu/hiqua/, (Accessed 07/03/07)

[61] "HITSP Delivers Interoperability Specs for Nationwide Health Information Network," *IHS Inc.*, 2006. http://engineers.ihs.com/news/2006/hitsp-specs-nihn.htm, (Accessed 28/01/07)

[62] "HON Code of Conduct (HONcode) for medical and health Web sites." http://www.hon.ch/HONcode/, (Accessed 26/02/06)

[63] HON secretariat , Health On the Net (HON) Foundation, Switzerland, Email: <honsecretariat@healthonnet.org, on 13/02/2006, Personal Communications, 2006.

[64] "How to get a PHR." http://www.recordsforliving.com/PersonalHealthRecords/HowToGetOne.aspx, (Accessed 17/04/08)

[65] "iHealthRecord, About the iHealthRecord." http://www.ihealthrecord.org/aboutIHealthRecord.html, (Accessed 29/01/07)

[66] "iHealthRecord, Welcome to the Medem Physician Finder." http://www.medem.com/physfinder/physfinder_entrypage.cfm, (Accessed 29/01/07)

[67] "ILIAD." http://prime.jsc.nasa.gov/iliad/iliadtxt.html, (Accessed 07/03/07)

[68] "The Individual Health Record (IHR)." http://www.wales.nhs.uk/IHC/page.cfm?pid=25880, (Accessed 17/04/08)

[69] "Information for Health: An Information Strategy for the Modern NHS 1998-2005," *NHS Executive*, Leeds, 1998.

[70] "Information Overload," *Answers.com*. http://www.answers.com/information%20overload, (Accessed 25/04/07)

[71] "Informing Healthcare Achievements," *NHS Wales*, 2006. http://www.wales.nhs.uk/ihc/documents/IHCAchievements2006.pdf, (Accessed 06/01/2007)

[72] "Informing Healthcare, About Informing Healthcare," *NHS Wales*. http://www.wales.nhs.uk/IHC/page.cfm?pid=8763, (Accessed 10/01/2007)

[73] "International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)." http://www.cdc.gov/nchs/datawh/ftpserv/ftpicd9/ftpicd9.htm#guidelines, (Accessed 03/07/07)

[74] "International Classification of Primary Care, Second edition (ICPC-2)," *World Health Organisation*. http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html, (Accessed 08/01/08)

[75] "Internet Access," *National Statistics*. http://www.statistics.gov.uk/cci/nugget.asp?id=8, (Accessed 15/02/06)

[76] "INTOUCH With Health." http://www.intouchwithhealth.co.uk/delivery_kiosks.htm, (Accessed 12/12/06)

[77] "Intute: Health & Life Sciences: Medicine." http://www.intute.ac.uk/healthandlifesciences/medicine/, (Accessed 28/02/07)

[78] "Java Server Pages Technology." http://java.sun.com/products/jsp/, (Accessed 01/05/04)

[79] "JDBC Technology." http://java.sun.com/products/jdbc/, (Accessed 01/07/05)

[80] "Jena 2 Ontology API." http://jena.sourceforge.net/ontology/#creatingModels, (Accessed 28/12/07)

[81] "Jena – A Semantic Web Framework for Java." http://jena.sourceforge.net/, (Accessed 28/12/07)

[82] "Key Capabilities of an Electronic Health Record Syetsm," *The Institute of Medicine*, July 31, 2003 2003.

[83] "Kosmix." http://www.kosmix.com/index, (Accessed 25/02/07)

[84] "Learn about the Summary Care Record." http://www.connectingforhealth.nhs.uk/systemsandservices/nhscrs/scr/learn, (Accessed 14/0408)

[85] "Macmillan Cancer Support." http://www.macmillan.org.uk/Home.aspx, (Accessed 13/04/07)

[86] "Mammahealth." http://www.mamma.com/, (Accessed 08/02/08)

[87] "The Markle Foundation. Connecting for Health." http://www.connectingforhealth.org/aboutus/index.html, (Accessed 04/02/07)

[88] "MedHunt." http://www.hon.ch/MedHunt/, (Accessed 26/02/06)

[89] "MedHunt Resource Description," *Kent State University, Libraries & Media Services.* http://www.library.kent.edu/resource.php?id=505, (Accessed 26/02/06)

[90] "Medical Subject Headings (MeSH)," *The US National Library of Medicine, The National Institute for Health.* http://www.nlm.nih.gov/mesh/, (Accessed 08/01/08)

[91] "MedlinePlus." http://medlineplus.gov/, (Accessed 23/02/07)

[92] "Microsoft HealthVault Search & HealthVault.com Beta Version Privacy Statement." https://health.live.com/content.aspx?id=help/privacy.htm&rmproc=true, (Accessed 17/04/08)

[93] "Microsoft's Health Vault Allows You to Store and Share Your Electronic Medical Records - But Should You?." http://www.theinternetpatrol.com/microsofts-health-vault-allows-you-to-store-and-share-your-electronic-medical-records-but-should-you, (Accessed 14/04/08)

[94] "Mitretek Systems." http://www.mitretek.org/, (Accessed 26/02/06)

[95] "My Health On-line," *Informing Healthcare Projects, NHS Wales.* http://www.wales.nhs.uk/IHC/text/page.cfm?pid=8769, (Accessed 10/01/2007)

[96] "My Health Online." http://www.wales.nhs.uk/IHC/page.cfm?pid=25814, (Accessed 17/04/08)

[97] "My Health Online." http://www.wales.nhs.uk/sites3/Documents/266/Item_8_Appendix%201.pdf, (Accessed 16/04/08)

[98] "MyChart," *UT SouthWestern Medical Center.* https://mychart.utsouthwestern.edu/mychart/, (Accessed 05/02/07)

[99] "National eHealth/ IM&T Strategy 2004 – 2008," *NHS Scotland*, 2004. http://www.ehealth.scot.nhs.uk/, (Accessed 04/02/07)

[100] "Net Scoring: Criteria to assess the quality of Health Internet information." http://www.chu-rouen.fr/netscoring/netscoringeng.html, (Accessed 26/02/06)

[101] "New Releases. CLEVELAND CLINIC COLLABORATES WITH GOOGLE." http://www.clevelandclinic.org/media/release.asp?Press_Releases_No=930, (Accessed 16/04/08)

[102] "NHS24," *NHS Scotland.* http://www.nhs24.com, (Accessed 30/01/07)

[103] "The NHS Cancer Plan," 2000. http://www.theglobeandmail.com/special/cancerdocuments/nhscancerplan.pdf, (Accessed 01/12/06)

[104] "NHS Care Records Service." http://www.connectingforhealth.nhs.uk/systemsandservices/nhscrs, (Accessed 10/04/08)

[105] "NHS Connecting for Health (CfH)." http://www.connectingforhealth.nhs.uk/, (Accessed 10/01/07)

[106] "NHS Direct Digital TV Launched." http://www.informatics.nhs.uk/cgi-bin/item.cgi?id=1086&d=11&h=24&f=46&dateformat=%25o-%25b-%25Y, (Accessed 30/01/07)

[107] "NHS Direct Kiosk." http://www.keytechnology-solutions.com/nhs.html, (Accessed 12/12/06)

[108] "NHS Direct On-Line." http://www.nhsdirect.nhs.uk/, (Accessed 26/02/06)

[109] "NHS Wales Informing Healthcare (IHC)." http://www.wales.nhs.uk/IHC/home.cfm?CFID=12776325&CFTOKEN=18511690, (Accessed 10/01/2007)

[110] "NHS. NHSOnline." http://www.nhsonline.net/, (Accessed 30/01/07)

[111] "On-Line Medical Dictionary." http://cancerweb.ncl.ac.uk/omd/, (Accessed 12/10/05)

[112] "Our Health Our Care Our Say White Paper," *Department of Health*, 2006. http://www.dh.gov.uk/dr_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4127628.pdf, (Accessed 08/02/08)

[113] "Overview: Personal Health Records – Current Landscape and Future Visions," *Project Health Design*. http://www.projecthealthdesign.org/overview-phr, (Accessed 30/01/07)

[114] "Patient Focus and Public Involvement," *NHS-Scotland*, 2001. http://www.scotland.gov.uk/Resource/Doc/158744/0043087.pdf, (Accessed 15/12/06)

[115] "The Patient Information Advisory Group (PIAG)," *Department of Health*, 2001. http://www.advisorybodies.doh.gov.uk/piag/foi.htm#piag, (Accessed 01/12/06)

[116] "PCASSO (Patient-Centered Access to Secure Systems Online)." http://medicine.ucsd.edu/pcasso/, (Accessed 02/02/07)

[117] "Personal Health Records," *OpenClinical, Knowledge Management for Healthcare*, 2006. http://www.openclinical.org/phr.html, (Accessed 02/02/07)

[118] "Personal Health Records on the Internet," *Informatics-Review*, 2000. http://www.informatics-review.com/records.html, (Accessed 21/01/07)

[119] "PersonalMD." http://www.personalmd.com, (Accessed 05/02/07)

[120] "Plain English Campaign." http://www.plainenglish.co.uk, (Accessed 01/10/05)

[121] "Proceed with Caution: A Report on the Quality of Health Information on the Internet. Complete Study," *RAND Health, Prepared for California HealthCare Foundation.*, 2001. http://www.chcf.org/documents/consumer/ProceedWithCautionCompleteStudy.pdf, (Accessed 20/02/07)

[122] "Proceedings of the Health and Social Services Committee," *The National Assembly for Wales*, 2006. http://www.wales.gov.uk/cms/2/CommitteeMeeting/N0000000000000000000000000000053/8fbd44f947131ea0216966185d8f4fd2.html, (Accessed 28/02/07)

[123] "Quality Standards for Medical Publishing on the Web. British Healthcare Internet Association (BHIA) Quality Standards." http://www.bhia.org, (Accessed 26/02/06)

[124] "Renal Patient View," *NHS*. https://www.renalpatientview.org/, (Accessed 04/02/07)

[125] "Requirement Analysis Process: Requirement Elicitation, Analysis, Specification." http://www.outsource2india.com/software/RequirementAnalysis.asp, (Accessed 12/11/2007)

[126] "Selecting A Development Approach " *Centers for Medicare and Medicaid Services (CMS)*, 2005. http://www.cms.hhs.gov/SystemLifecycleFramework/Downloads/SelectingDevelopmentApproach.pdf, (Accessed 10/11/07)

[127] "SNOMED Clinical Terms (SNOMED CT)." http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html, (Accessed 05/11/07)

[128] "Social networks top Google search," *BBC*, 2006. http://news.bbc.co.uk/1/hi/technology/6189809.stm, (Accessed 01/05/07)

[129] "Stomach Cancer," *Wikipedia*. http://en.wikipedia.org/wiki/Stomach_cancer, (Accessed 07/08/07)

[130] "Summary Care Record (SCR)." http://www.connectingforhealth.nhs.uk/systemsandservices/nhscrs/scr, (Accessed 10/04/08)

[131] "Sustain IT Workshop. Broadband and Patient Involvement in Health Care," Wellcome Library. London, 13 February 2006.

[132] "Text Retrieval Quality: A Primer," *Oracle Technology Network*. http://www.oracle.com/technology/products/text/htdocs/imt_quality.htm, (Accessed 01/01/08)

[133] "Tomorrow's patients will have information prescribed alongside medicines," *eGov Monitor*. http://www.egovmonitor.com/node/8149, (Accessed 05/12/06)

[134] "Towards a Cancer Information Strategy," *Cancer Information Strategy Team, NHS Information Authority*, 2000. http://www.nhsia.nhs.uk/cancer/pages/documents/fulldoc.pdf, (Accessed 01/12/06)

[135] "Trust-e." http://www.truste.org, (Accessed 01/11/05)

[136] "Understanding Microsoft HealthVault." https://worlddoc.fugu.com/remote/RemoteLaunch.aspx?c=2H2G3PX54W&n=25&sh=208, (Accessed 14/04/08)

[137] "Unified Medical Language System," *The US National Library of Medicine,The National Institute for Health*. http://www.nlm.nih.gov/research/umls/, (Accessed 08/01/08)

[138] "Use of the Internet," *National Statistics*. http://www.statistics.gov.uk/CCI/nugget.asp?ID=1711&Pos=1&ColRank=1&Rank=176, (Accessed 09/04/07)

[139] "Utilization Accreditation Commission (URAC)." http://www.urac.org, (Accessed 01/11/05)

[140] "Velindre NHS Trust." http://www.wales.nhs.uk/sites3/home.cfm?orgid=34, (Accessed 15/10/07)

[141] "W3C Semantic Web Activity." http://www.w3.org/2001/sw/, (Accessed 05/11/07)

[142] "WebMD." http://www.WebMD.com, (Accessed 07/03/07)

[143] "Welcome to Hi-Ethics," *Health Internet Ethics (HI-Ethics)*. http://www.hiethics.com/, (Accessed 26/02/06)

[144] "Welcome to the National Institute for Health and Clinical Excellence (NICE)." http://www.nice.org.uk/, (Accessed 01/01/08)

[145] "Welsh Source Health Information Online," *Informing Health Care, NHS Wales*, 2006. http://www.wales.nhs.uk/ihc/Documents/Welsh%20Source%20Health%20Information%20Online%20-%20Report%20Published%20Nov%202006.pdf, (Accessed 14/02/2007)

[146] "WordNet." http://wordnet.princeton.edu/w3wn.html, (Accessed 05/11/07)

[147] J. M. Abasolo and M. Gomez, "MELISA. An ontology-based agent for information retrieval in medicine," presented at *ECDL 2000 Workshop on the Semantic Web*, Lisbon, Portugal, 2000.

[148] K. Aberer, P. Cudré-Mauroux, and M. Hauswirth, "A framework for semantic gossiping," *ACM SIGMOD*, vol. 31(4), 48 – 53, 2002.

[149] S. R. Abidi, K. Elvidge, H. Kharazi, M. Shepherd, C. Watters, and J. Duffy, "Reformulation of Consumer Health Queries with Standard Medical Vocabulary," presented at *The International Symposium for Health Information Management Research (ISHIMR06)*, Halifax, Nova Scotia, Canada, 2006.

[150] S. Abiteboul, P. Buneman, and D. Suciu, "Data on the Web: From Relations to Semistructured data and XML": *Morgan Kaufmann Publishers*, 2000.

[151] AHIMA, "The Role of the Personal Health Record in the EHR," *Journal of AHIMA*, vol. 76(7), 2005.

[152] A. Al-Busaidi, "Supporting Multiple Integrated Views in a Web environment," presented at *The British National Conference on Databases (BNCOD 21) - Doctoral Consortium*, Edinburgh, UK, 2004.

[153] A. Al-Busaidi, A. Gray, and N. Fiddian, "Investigating and Utilising Patient Information Needs to Focus Internet Searching for Cancer Patients," presented at *The International Symposium for Health Information Management Research (ISHIMR06).* Halifax, Nova Scotia, Canada, 2006.

[154] A. Al-Busaidi, A. Gray, and N. Fiddian, "Personalising Web Information for Patient: Linking Patient Medical Data with the Web via a Patient Personal Knowledge Base," presented at *The International Symposium on Health Information Management Research (ISHIMR05)*, Thessaloniki, Greece, 2005.

[155] A. Al-Busaidi, A. Gray, and N. Fiddian, "Personalizing web information for patients: linking patient medical data with the web via a patient personal knowledge base," *Health Informatics Journal*, vol. 12(1), 27-39, 2006.

[156] F. Al-Busaidi, Specialist Obstetrics and Gynaecology, Diwan Clinic, Diwan of Royal Court. Muscat, Sultanate of Oman, Personal Communication, 2007.

[157] O. Al-Ghamdi, "A Clinical Aspect of the Computer-Based Patient Record: Free Text versus Coded Data," 2007. http://www.informatics-review.com/thoughts/freetext.pdf, (Accessed 06/03/07)

[158] M. Al-Ubaydli, "Using Search Engines to Find Online Medical Information," *PloS Med* vol. 2(9), 2005.

[159] M. B. Al Mourad."Using Structure and Behavior to create Multiple Views for Interoperability in a heterogeneous Object-Oriented Multidatabase Environment."*PhD.Thesis*, Department of Computer Science, Cardiff University, Cardiff, 2003.

[160] O. Allam."A Holistic Analysis Approach to Facilitating Communication between General Practitioners and Cancer Care Teams."*PhD.Thesis*, School of Computer Science, Cardiff University, Cardiff, 2006.

[161] O. Allam, Medical Doctor (MD) and Researcher. School of Computer Science, Cardiff University. Personal Communications, 2005.

[162] O. Allam, A. Gray, A. Bater, H. Baily, and D. Morrey, "The ISCO Information System, A Model of Coexistence with the Integrated National Care Record," presented at *The International Symposium for Health Information Management Research (ISHIMR* Sheffield, UK, 2004.

[163] Z. Allawatiya, Medical Student, School of Medicine, Cardiff University, Personal Communications, 2007.

[164] V. Aston, Patient Information Manager, Velindre Cancer Centre and Tenovus. Personal Communications, 2005.

[165] D. Avison and G. Fitzgerald, "Information Systems Development, Methodologies, Techniques and Tools", 4 ed: *McGraw Hill*, 2006.

[166] T. Baeyens and P. Fricke, "Bringing BMP to the Mass Market," *Business Integration Journal*, (Mar/Apr 2006), 2006.

[167] H. Bailey, Information Analysis, Clinical Information Unit (CIU) - Velindre NHS Trust. Personal Communications, 2005.

[168] N. Bajwa, "Application Integration through Microsoft Technologies," 2007. http://www.codeproject.com/useritems/Application_Integration.asp?print=true, (Accessed 29/10/07)

[169] P. Baker, C. Goble, S. Bechhofer, A. P. Brass, N., and R. Stevens, "An Ontology For Bioinformatics Applications," *Bioinformatics*, vol. 15(6), 510-520, 1999.

[170] P. G. Baker, A. Brass, S. Bechhofer, C. Goble, N. Paton, and R. Stevens, "TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. An Overview," presented at *The Sixth International Conference on Intelligent Systems for Molecular Biology*, 1998.

[171] C. Balmer, "Cancer Patients use of the Internet," *Cancer Nursing Practice*, vol. 1(2), 28-31, 2002.

[172] J. Barnes, "Improving Health in Wales. A Plan for the NHS with its Partners," 2001. http://www.rpsgb.org.uk/wales/pdfs/imphealthinwales.pdf, (Accessed 18/02/07)

[173] A. Bater, ISCO Project Manager, Clinical Information Unit (CIU) - Velindre NHS Trust. Personal Communications, 2005.

[174] D. Bell and J. Grimson, "Distributed Database Systems": *Addison-Wesely*, 1992.

[175] D. S. Bental, A. Cawsey, and R. Jones, "Patient information systems that tailor to the individual," *Patient Education and Counselling*, vol. 36(2), 171-180, 1999.

[176] B. A. Bilodeau and L. F. Degner, "Information Needs, Sources of Information, and Decisional Roles in Women With Breast Cancer.," *Oncology Nursing Forum*,23,691-696, 1996

[177] J. S. Blair, "An Overview Of Healthcare Information Standards. IBM Healthcare Solutions.," *IBM Healthcare Solutions, Atlanta, Georgia*, 1999. http://hipaafaq.com/cpri.htm, (Accessed 08/02/08)

[178] T. Bodenheimer and K. Grumbach, "Electronic technology – a spark to revitalise primary care?," *JAMA* vol. 290, 259–64, 2003.

[179] K. E. Boulding, "The pollution of information," 1996. http://www.colorado.edu/econ/Kenneth.Boulding/pollution-of-info.html, (Accessed 04/05/07)

[180] I. K. Bray, "Introduction to Requirements Engineering": *Addison-Wesley*, 2002.

[181] F. P. J. Brooks, "No Silver Bullet: Essence and Accident in Software Engineering," *The Mythical Man-Month*,177-203: Addison Wesley, 1986

[182] M. Burgess, O. Allam, A. Gray, and A. Al-Busaidi, "Patient-Centred Approach to Focusing Online Health Information Search Results," presented at *ISePHR (2006)*, 2006.

[183] S. Busse, R.-D. Kutsche, U. Leser, and H. Weber, "Federated Information Systems: Concepts, Terminology and Architectures. ," *Technical Report 99-9*: Forschungsberichte des Fachbereichs Informatik, 1999

[184] J. Butters, Radiotherapy Specialist Nurse, Radiotherapy Department - Velindre NHS Trust. Personal Communications, 2005.

[185] C. N. Carlson, "Information overload, retrieval strategies and Internet user empowerment," *The Good, the Bad and the Irrelevant (COST 269)*,1,169-173. Helsinki, Finland: Haddon, Leslie, Eds, 2003

[186] L. Carrier, "Managing at light speed," *IEEE Computer*, vol. 32(7), 107-109, 1999.

[187] A. Cawsey, R. Jones, and P. J., "The Evaluation of a Personalised Information System for Patients with Cancer," *User Modelling and User-Adapted Interaction 2000*, vol. 10(1), 2000.

[188] S. Childs, "Internet Opportunities and Challenges for Patient Information and Support," presented at *Sustain IT workshop "Broadband and Patient Involvement in Health Care"*, Wellcome Library, London, 2006.

[189] E. Coiera, "Guide to Health Informatics". London: *Arnold*, 2003.

[190] Z. Cui, D. M. Jones, and P. O'Brien, "Semantic B2B Integration: Issues in Ontology-based Applications," presented at *SIGMOD Record*, 2002.

[191] S. Darmoni, B. Dahamna, and T. Roth-Berghofer, "Seal of transparency heritage in the CISMeF quality-controlled health gateway," *BMC Medical Informatics and Decision Making*, vol. 4, 2004.

[192] J. David, "Welcome to the RHIO Revolution! EHRs, RHIOs, and the NHIN. Putting IT all Together," *HIMSS*, 2006. http://www.idph.state.il.us/ehrtf/Jeff_David_Slides.pdf, (Accessed 02/02/07)

[193] D. Detmer, "Building the National Health Information Infrastructure for Personal Health, Health Care Services, Public Health, and Research," *BMC Medical Informatics and Decision Making*, vol. 3(1), 2003.

[194] D. Detmer and E. Steen, "Learning from Abroad: Lessons and Questions on Personal Health Records for National Policy," *Research Report, American Medical Informatics Association*, 2006. http://www.aarp.org/research/health/carequality/2006_10_phr_abroad.html, (Accessed 08/02/08)

[195] M. Dixon-Woods, "Writing wrongs? An analysis of published discourses about the use of patient information leaflets," *Soc Sci Med*, vol. 52, 1417-1432, 2001.

[196] R. M. Duwairi."Views for Interoperability in a Heterogeneous Object-Oriented Multidatabase System ".*PhD.Thesis*, Computer Science Department, University of Wales College of Cardiff, Cardiff, UK, 1997.

[197] L. Edgar, A. Greenberg, and J. Remmer, "Providing Internet lessons to oncology patients and family members: A shared project,". *Psycho-Oncology*, vol. 11, 439-446, 2002.

[198] A. Elmagarmid, M. Rusinkiewicz, and A. Sheth, "Management of heterogeneous and autonomous database systems". San Francisco, CA.: *Morgan Kaufmann Publishers Inc.*, 1998.

[199] S. Endsley, D. Kibbe, A. Linares, and K. Colorafi, "An introduction to personal health records," *Family Practice Management*, vol. 13(5), 57-62, 2006.

[200] G. Eysenbach, "Towards quality management of medical information on the Internet: evaluation, labelling, and filtering of information," *BMJ*, vol. 317, 1496–1502, 1998.

[201] G. Eysenbach, J. Powell, and O. S. Kuss, E.R., "Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. ," *JAMA 2000*, vol. 287(20), 2691-2700, 2002.

[202] G. Eysenbach, G. Yihune, K. Lampe, P. Cross, and D. Brickley, "MedCERTAIN: Quality Management, Certification and Rating of Health Information on the Net " presented at *AMIA Symp*, 2000.

[203] D. Fang, J. Hammer, and D. McLeod, "The identification and resolution of semantic heterogeneity," presented at *The First International Workshop on Interoperability in Multidatabase Systems*, Kyoto, Japan, 1991.

[204] D. Fensel, "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce": *SpringerVerlag*, 2001.

[205] D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, and M. Klein, "Oil in a nutshell," presented at *The 12th International Conference on Knowledge Engineering and Knowledge Management EKAW-2000*, 2000.

[206] T. Ferguson, "Found on the Net: What e-Patients Do Online: A Tentative Taxonomy," *The Ferguson Report*, 9, September 2002.

[207] S. Fox and L. Rainie, "The Online Health Care Revolution: How the Web Helps Americans Take Better Care of Themselves," *Pew Charitable Trusts*, 2000. http://www.pewinternet.org/pdfs/PIP_Health_Report.pdf, (Accessed 08/02/08)

[208] E. Franconi, "Introduction to Data Warehousing," 2002. http://www.cs.man.ac.uk/~franconi/teaching/2002/cs636/CS636-dw-intro.ppt, (Accessed 21/06/03)

[209] M. Garcia-Solaco, F. Saltor, and M. Castellanos, "Semantic heterogeneity in multidatabase systems," In O. Bukhres and A. Elmagarmid, *Object Oriented Multidatabase Systems: A Solution for Advanced Applications*,129-202, 1996

[210] W. A. Gray, Professor of Advanced Information Systems, School of Computer Science, Cardiff University, Personal Communications, 2007.

[211] J. Griffiths, "Cancer Services Information Framework," 2000. http://www.wales.nhs.uk/publications/cif.pdf, (Accessed 15/02/07)

[212] T. Gruber, "Toward Principles for the Design of Ontologies used for Knowledge Sharing," presented at *The Padua Workshop on Formal Ontology*, 1993.

[213] F. Hakimpour and A. Geppert, "Resolving semantic heterogeneity in schema integration," presented at *The international conference on Formal Ontology in Information Systems* Ogunquit, Maine, USA, 2001.

[214] J. Hammer, "Data Warehousing Overview, Terminology, and Research Issues," 2003. http://www.cise.ufl.edu/~jhammer/classes/wh-seminar/Overview/Overview.PPT, (Accessed 21/6/2003)

[215] J. Hammer and M. Schneider, "Genomic Algebra: A new, Integrating Data Model, Language, and Tool for processing and querying Genomic Information," *Department of Computer & Information Science & Engineering, University of Florida*, 2003. http://citeseer.nj.nec.com/cache/papers/cs/26920/http:zSzzSzwww-db.cs.wisc.eduzSzcidrzSzprogramzSzp16.pdf/genomics-algebra-a-new.pdf, (Accessed 20/06/03)

[216] A. Hassol, J. M. Walker, D. Kidder, K. Rokita, D. Young, S. Pierdon, D. Deitz, S. Kuck, and E. Ortiz, "Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging," *Journal of the American Medical Informatics Association*, vol. 11(6), 505-513, 2004.

[217] J. Heflin and J. Hendler, "Semantic Interoperability on the Web," presented at *Extreme Markup Languages 2000*, 2000.

[218] J. Heflin, J. Hendler, and S. Luke, "SHOE: A Blueprint for the Semantic Web," In D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, *Spinning the Semantic Web*. Cambridge, MA: MIT Press, 2003

[219] D. Heimbigner and D. McLeod, "A Federated Architecture for Information Management," *ACM Transaction on Information Systems (TOIS)*, vol. 3(3), 253-278, 1985.

[220] L. Q. Hieu, "Integration of Web Data Sources: A Survey of Existing Problems. Institute of Computer Science. Heinrich-Heine-University," 2005. http://dbs.informatik.uni-halle.de/GvD2005/beitraege/gvd05_LeQuang.pdf, (Accessed 20/10/07)

[221] W. Jones, Cancer Information Framework Project Manager, Clinical Information Unit (CIU) - Velindre NHS Trust. Personal Communication, 2005.

[222] M. KAMEL-BOULOS, A. V. ROUDSARI, and E. R. CARSON, "A dynamic problem-knowledge coupling semantic Web service," presented at *6th World Congress on the Internet in Medicine*, Udine, Italy, 2001.

[223] M. KAMEL-BOULOS, A. V. ROUDSARI, and E. R. CARSON, "HealthCyberMap: mapping the health cyberspace using hypermedia GIS and clinical codes," presented at *1st European Conference for Geographic Information Sciences in Public Health*, Sheffield, 2001.

[224] D. Karunaratna."Exploitation of Semantic Information in the Creation of Multiple Views of a Federated Database Systems."*PhD.Thesis*, Computer Science Department, University of Wales College of Cardiff, Cardiff, 2000.

[225] C. Kierans, " miHealth," presented at *Sustain IT workshop "Broadband and Patient Involvement in Health Care"*, Wellcome Library, London, 2006.

[226] A. Laender, B. Ribeiro-Neto, A. da Silva, and J. Teixeira, "A brief survey of Web data extraction tools," *SIGMOND Record*, vol. 31(2), 84-93, 2002.

[227] A. Langlands, "Information for Health: An Information Strategy for the Modern NHS," *NHS Executive, Health Service Circular*, 1998. http://christie.man.ac.uk/profinfo/departments/nursing/nursing_developments/usefuldocs/HSC168InformationforHealth98.pdf, (Accessed 14/02/07)

[228] O. Lassila, "Web Metadata: A Matter of Semantics," *EEE Internet Computing*, vol. 12(4), 30-37, 1998.

[229] A. Lenhart and S. Fox, "A portrait of the Internet's new storytellers " *Pew Internet & American Life Project*, 2006. http://www.pewinternet.org/pdfs/PIP%20Bloggers%20Report%20July%2019%202006.pdf, (Accessed 30/11/06)

[230] E. B. Lerner, D. V. K. Jehle, D. M. Janicke, and R. M. Moscati, "Medical communication: Do our patients understand?," *The American Journal of Emergency Medicine*, vol. 18(7), 764-766, 2000.

[231] V. Lewington and J. Farmer, "Getting better: health information provision in Guernsey," *New Library World*, vol. 96(7), 8-14(7), 1995.

[232] D. P. Lorence and A. Spink, "Semantics and the medical web: a review of barriers and breakthroughs in effective healthcare query," *Health Information and Libraries Journal*, vol. 21(2), 109-116, 2004.

[233] L. A. Maciaszek, "Requirement Analysis and System Design", 3 ed: *Addison-Wisely*, 2007.

[234] M. Mair and C. Kierans, "Evaluating 'mihealth Liverpool': Assessing the Effectiveness of New Health Information and Communication Technology in Providing Support to Patients with Breast Cancer Receiving Treatment in Liverpool," *Centre for Public Health, Liverpool John Moores University*, 2006. http://www.mihealth.info/MihealthReport_July_2006_V2.pdf, (Accessed 05/02/06)

[235] M. A. Mayer, V. Karkaletsis, P. Archer, P. Ruiz, K. Stamatakis, and A. Leis, "Quality Lableling of Medical Web Content," presented at *10'th International Symposium for Health Information Management Research (ISHIMR 2005)*. Thessaloniki, Greece, 2005.

[236] H. McCONNELL, "International efforts in implementing national health information infrastructure and electronic health records," *World Hospitals and Health Services*, vol. 40(1), 33-40, 2004.

[237] S. McEwen, "Requirements: An introduction," *Metasys Technologies, Inc.* , 2004. http://www.ibm.com/developerworks/rational/library/4166.html, (Accessed 0802/08)

[238] S. W. McRoy, A. Liu-Perez, and S. S. Ali, "Interactive computerized health care education " *Journal of the American Medical Informatics Association*, vol. 5(4), 347-356, 1998.

[239] T. Miller, G. Leroy, S. Chatterjee, J. Fan, and B. Thoms, "A Classifier to Evaluate Language Specificity of Medical Documents," presented at *Fortieth Annual Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, Big Island, Hawaii, 2007.

[240] T. Miller, G. Leroy, and E. Wood, "Dynamic Generation of a Table of Contents with Consumer-Friendly Labels," presented at *American Medical Informatics Association (AMIA) Annual Symposium*, Washington DC, 2006.

[241] M. E. Mills and R. Davidson, "Cancer patients' sources of information: use and quality issues," *Psychooncology*, vol. 11(5), 371-378, 2002.

[242] T. Mueck, "Database Integration," *Abteilung Data Engineering, Universität Magdeburg*, 1999. http://wwwiti.cs.uni-magdeburg.de/iti_bm/ibss/courses/Mue/SummerSchool.pdf, (Accessed 20/06/03)

[243] M. Murphy and R. Mann, "Secondary Uses," 2005. www.wales.nhs.uk/ihc/documents/2-2_Secondary_uses_of_care_information(MM%20RM).ppt, (Accessed 01/03/2007)

[244] D. E. Nelson, "Reducing Information Pollution in the Internet Age," *Public Health Research, Practice and Policy*, vol. 4(1), 2007.

[245] R. Neville, A. Greene, J. Mcleod, A. Tracy, and J. Surie, "Mobile phone text messaging can help young people manage asthma," *British Medical Journal (BMJ)*, vol. 325(7364), 600, 2002.

[246] J. Nielsen, "Information Pollution," *Jakob Nielsen's Alertbox*, 2003. http://www.useit.com/alertbox/20030811.html, (Accessed 03/05/07)

[247] M. North, "Read Codes Introduction." http://www.equip.ac.uk/readCodes/docs/index.html, (Accessed 25/02/06)

[248] A. Noruzi, "Information Pollution," *Iranian Journal of Information Science* vol. 15(1&2), 3-8, 2000.

[249] J. Ogden, R. Branson, A. Bryett, A. Campbell, A. Febles, I. Ferguson, H. Lavender, J. Mizan, R. Simpson, and M. Tayler, "What's in a name? An experimental study of patients' views of the impact and function of a diagnosis," *Family Practice*, vol. 20, 248–253, 2003.

[250] K. B. Oliver and N. K. Roderer, "Working towards the Informationist," *Health Informatics Journal*, vol. 12(1), 41-48, 2006.

[251] R. L. Ownby, "Influence of Vocabulary and Sentence Complexity and Passive Voice on the Readability of Consumer-Oriented Mental Health Information on the Internet," *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, vol. 2005, 585–588., 2005.

[252] M. T. Ozsu and P. Valduriez, "Principles of Distributed Database Systems": *Prentice-Hall*, 1991.

[253] Q. Parker, "Life outside NpfIT: Wales " *E-Health Insider*, 2004. http://www.e-health-insider.com/comment_and_analysis/index.cfm?ID=31, (Accessed 15/12/06)

[254] T. Philips, "Healthy respect for smart cards," *The Guardian*, 2006. http://technology.guardian.co.uk/weekly/story/0,,1802542,00.html, (Accessed 8/12/06)

[255] A. Preece, "Distributed Knowledge Technologies for Intelligent Web Services," *Departmental Seminar, Computer Science Department, University of Aberdeen*, 2001. http://users.cs.cf.ac.uk/A.D.Preece/talks/DistributedKnowledgeTechnologiesForIntelligentWebServices/DistributedKnowledgeTechnologiesForIntelligentWebServices.pdf, (Accessed 08/02/08)

[256] A. Preece, K. Hui, A. Gray, P. Marti, T. Bench-Capon, D. Jones, and Z. Cui, "The KRAFT Architecture for Knowledge Fusion and Transformation," presented at *19th SGES International Conference on Knowledgebased Systems and Applied Artificial Intelligence (Expert Systems 99)*, Berlin, 1999.

[257] D. J. Protti, "Personal Health Records And Sharing Patient Information," *NHS – Connecting For Health (CFH)*, 2006. http://www.connectingforhealth.nhs.uk/worldview/protti7, (Accessed 21/12/06)

[258] R. Rada and B. K. Martin, "Augmenting thesauri for information systems," *ACM Transactions on Office Information Systems*, vol. 5(4), 378-392, 1987.

[259] R. W. Rasmussen, "A framework for the UML meta model," *University of Bergen, Institute for informatics*, 2004. http://www.ii.uib.no/~rolfwr/thesisdoc/main1.html, (Accessed 30/04/04)

[260] A. Rector, P. Zanstra, and W. Solomon, "GALEN: Terminology Services for Clinical Information Systems," *Medical Informatics Group, Department of Computer Science, University of Manchester*, 1995. http://books.google.com/books?id=b-6ekNnTfNkC&pg=PA90&lpg=PA90&dq=GALEN:+Terminology+Services+for+Clinical+Information+Systems&source=web&ots=UMi7LOBB84&sig=kThQWqEgBs1LZhvC8Bp_u0xcozk, (Accessed 09/02/08)

[261] C. E. Rees and P. A. Bath, "The information needs and source preferences of women with breast cancer and their family members: A review of the literature published between 1988 and 1998," *Journal of Advanced Nursing*, vol. 31(4), 833-841, 2000.

[262] A. Risk and J. Dzenowagis, "Review of Internet Health Information Quality Initiatives," *Journal of Medical Internet Research*, vol. 3(4), e28, 2001.

[263] R. Roberts, Chemotherapy Specialist Nurse, Chemotherapy Department - Velindre NHS Trust. Personal Communications, 2005.

[264] S. Sastry and P. Carroll, "Doctors, patients and the Internet: time to grasp the nettle," *Clinical Medicine, Journal of the Royal College of Physicians*, vol. 2(2), 131-133, 2002.

[265] J. H. Schneider and D. Kofos, "Consumer Internet Health Records: Benefits for Paediatrics, Problems and Proposed Standards," *Baylor College of Medicine, Department of Pediatrics, Houston, TX; Methodist Children's Hospital of South Texas, Telemedicine and Critical Care, San Antonio, TX*, 2001. http://www.atmeda.org/news/2001_presentations/e-Health/Schneider%20ATA%20CIMRs.ppt, (Accessed 21/01/07)

[266] A. P. Sheth and J. A. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computing Surveys*, vol. 22(3), 183-236, 1990.

[267] E. H. Shortliffe and L. E. Perreault, "Medical Informatics: Computer Applications in Health Care and Biomedicine". New York: *Springer*, 2001.

[268] B. Sissons, "A Personalised Information System for Cancer Patients," *MSc Dissertation. School of Computer Science, Cardiff University*, 2004.

[269] L. Slaughter, C. Ruland, and A. K. Rotegard, "Mapping Cancer Patients' Symptoms to UMLS Concepts," *AMIA Annual Symposium Proceedings*, vol. 2005, 699–703, 2005.

[270] K. Smolij and K. Dun, "Patient Health Information Management: Searching for the Right Model," *American Health Information Association (AHIMA)*, 2006. http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_032723.html, (Accessed 25/01/07)

[271] D. Soergel, T. Tse, and L. Slaughter, "Helping Healthcare Consumers Understand: An "Interpretive Layer" for Finding and Making Sense of Medical Information ", 2004.

[272] I. Sommerville, "Software engineering". Boston: *Pearson/Addison-Wesley*, 2004.

[273] C. Sperberg-McQueen, C. Huitfeldt, and A. Renear, "A Practical Extraction of meaning from Markup," presented at *ACH/ALLC 2001*, New York University, 2001.

[274] P. Stavri, "Personal health information-seeking: a qualitative review of the literature," *MEDINFO 2001*, vol. 10, 1484-1488, 2001.

[275] M. L. Stead, J. M. Brown, L. Fallowfield, and P. Selby, "Lack of communication between healthcare professionals and women with ovarian cancer about sexual issues " *British Journal of Cancer (2003)*, vol. 88, 666 – 671, 2003.

[276] A. Stoop, A. Van't Riet, and M. Berg, "Using IT for Patient Education: Realizing surplus value?," *Patient Education and Counseling* vol. 54, 187-195, 2004.

[277] J. Sweetland, "Users' perceptions of the impact of information provided by a consumer health information service: an in-depth study of six users," *Health Libraries Review*, vol. 17(2), 77-82, 2000.

[278] A. R. H. Tawil."Supporting Semantic Interoperability in a Heterogeneous Multiple Information Server Environment."*PhD.Thesis*, School of Computer Science, Cardiff University, Cardiff, 2001.

[279] M. Themistocleous and G. Corbitt, "Is business process integration feasible?," *Journal of Enterprise Information Management* vol. 19(4), 434-449, 2006.

[280] L. Theodosiou and J. Green, "Emerging challenges in using health information from the internet," *Advances in Psychiatric Treatment*, vol. 9, 387-396, 2003.

[281] R. Thomas, H. Thornton, and J. Mackay, "Patient information materials in oncology: are they needed and do they work?," *Clinical Oncology*, vol. 11(4), 225-231, 1999.

[282] T. Tseab and D. Soergela, "Exploring Medical Expressions Used by Consumers and the Media: An Emerging View of Consumer Health Vocabularies,". *AMIA Annual Symposium Proceedings*, vol. 2003, 674–678, 2003.

[283] M. Uschold, "Where is the semantics on the Semantic Web? ," presented at *The Workshop on Ontologies in Agent Systems, 5th International Conference on Autonomous Agents*, 2001.

[284] A. Van't Riet, M. Berg, F. Hiddema, and K. Sol, "Meeting patients' needs with patient information systems: potential benefits of qualitative research methods," *International Journal of Medical Informatics*, vol. 64(1), 1-14, 2001.

[285] C. P. Waegemann, "Testimony National Committee on Vital and Health Statistics Regarding Personal Health Records," *Medical Records Institute*, 2002. http://www.ncvhs.hhs.gov/020724p6.htm, (Accessed 23/01/07)

[286] J. M. Walker, "Linking Patients and Physicians on the Web," *Geisinger Health System*, 2005. http://www.himss.org/Content/files/linkingpatients.pdf, (Accessed 22/10/05)

[287] D. Wanless, "Securing Our Future Health: Taking a Long-Term View. Final Report," *HM Treasury, Wanless Report. April 2002*, 2002. http://www.hm-treasury.gov.uk/consultations_and_legislation/wanless/consult_wanless_final.cfm, (Accessed 18/02/07)

[288] V. Watzlaf, X. Zeng, and P. Firouzan, "Clinical Terminologies And Structured Vocabularies: The Impact On The CPR " *Department of Health Information Management. University of Pittsburgh*, 2007. http://www.himss.org/content/files/proceedings/2002/sessions/seshouts/sesho089.pdf, (Accessed 28/01/07)

[289] C. Wentong, S. J. Turner, and Boon Ping Can, "Hierarchical Federations: An Architecture for Information Hiding," presented at *The fifteenth workshop on Parallel and Distributed simulation*, 2001.

[290] E. E. Westberg and R. A. Miller, "The basis for using the Internet to support the information needs of primary care," *Journal of the American Medical Informatics Association*, vol. 6(6), 6-25, 1999.

[291] G. Wiederhold, "Integration versus Interoperation " *Stanford University*, 2002. http://dbs.uni-leipzig.de/en/vortraege/Wiederhold.abstr.html, (Accessed 05/11/07)

[292] M. Wikgren, "Health Information on the Internet: New Trajectories? Net Paper," presented at *The 15th Nordic Conference on Media and Communication Research*, Reykjavik, 2001.

[293] P. Williams, D. Nicholas, and P. Huntington, "Home electronic health information for the consumer: user evaluation of a DiTV video-on-demand service," *Aslib Proceedings*, vol. 55, 64-74, 2003.

[294] E. Woods, "Final Paper: Delivery of Medical Information: Challenges and Potential," *Digital Libraries. Information 653*, 2006. http://www.pages.drexel.edu/~epw22/finalpaper.htm, (Accessed 28/02/07)

[295] J. C. Wootton, "The quality of information on women's health on the Internet," *Journal of Women's Health*, vol. 6, 575- 581, 1997.

[296] P. E. Zanstra, A. L. Rector, W. D. Solomon, T. Rush, W. A. Nowlan, and S. K. Bechhofer, "A terminology server for integrating clinical information systems: the GALEN approach " presented at *Current Perspectives in Healthcare Computing*, Harrogate, 1995.

[297] Q. Zeng, S. Kogan, N. Ash, and R. A. Greenes, "Patient and Clinician Vocabulary: How Different Are They?," presented at *MEDINFO 2001*, 2001.

[298] Q. T. Zeng, J. Crowell, R. M. Plovnick, E. Kim, and L. D. Ngo, E. , "Assisting Consumer Health Information Retrieval with Query Recommendations " *Journal of the American Medical Informatics Association*, vol. 13, 80-90, 2006.

[299] Q. T. Zeng and T. Tse, "Exploring and Developing Consumer Health Vocabularies," *Journal of the American Medical Informatics Association*, vol. 13, 24-29, 2005.

[300] G. H. Zhou, R. and R. King, "Generating Data Integration Mediators that Use Materialization " *Journal of Intelligent Information Systems*, vol. 6(2/3), 111-133, 1996.