E-Commerce Website Personalisation Based on Ontological Profiling

Natalia N. Edwards

School of Computer Science & Informatics
Cardiff University

This thesis is submitted for the degree of Doctor of Philosophy

2015

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed .... (candidate) Date ....

STATEMENT 1

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

Signed .... (candidate) Date ....

STATEMENT 2

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

Signed .... (candidate) Date ....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate) Date ....

Abstract

Electronic commerce has become an important part of our consumer lives, and we increasingly choose to do more and more of our shopping online. Along with the growth of online sales, the number of e-commerce retailers has also increased. This has inevitably put additional demands on existing companies as well as new market entrants to ensure that their growth (if not just survival) as well as competitiveness are sustainable and evolving. Web personalisation has been adopted as a means to support business sustainability and competitiveness. It is now increasingly common and has been recognised by e-commerce businesses and consumers as a feature and functionality, expected to be offered as 'standard'.

Recent World Wide Web technology advances have greatly improved the way e-commerce websites are designed and deployed. However, the analysis of academic literature and professional practices shows that these advances are not used to their full potential. This research gap is an opportunity for this community to consider how techniques such as ontologies could be used to enhance personalisation of e-commerce websites.

This thesis presents a novel approach to e-commerce website personalisation (*PERSONTO*), and in particular, personalisation of content presentation. Personalisation is achieved by means of an ontology-based e-shopper profiling. For this purpose, a reusable, extendible and Semantic Web compatible customer profiling ontology *OntoProfi* is designed and implemented.

A 'proof-of-concept' prototype of *PERSONTO* confirmed the feasibility of the proposed approach. The analysis of achievements of the research objectives and outcomes showed that the approach is flexible, extendible and reusable, and that it was achieved by using systematic methods in the system design and implementation of the prototype. The evaluation of the acceptance of the proposed approach suggests there is a high level of acceptance of the approach by the prospective end users and e-commerce developers.

Acknowledgements

This work would not be possible without the help, support and encouragement of so many people.

I would firstly like to express my deepest gratitude to my first supervisor, Professor Alex Gray. He made this journey enjoyable and fruitful. I remain indebted to him for his expert advice and guidance, generosity, unlimited support and patience. Thank you ever so much for making possible what at the start seemed to be the impossible!

My thanks are also due to my second supervisor, Dr Wendy Ivins for giving me her valuable time and support, for inspiring and encouraging me.

I am grateful to many of my colleagues. To Professor Chris Jones and Professor Paul Rosin for their help at the start of my PhD journey. To the late Professor Nick Fiddian, who, although did not participate in this project directly, always inspired me. To my progress reviewers, Dr Andrew Jones and Professor David Marshall for their valuable advice. To Mrs. Helen Williams for always being very helpful and kind.

I would also like to thank all those people, who participated in my case study and project evaluation for giving me their valuable time.

Last but definitely not least, my deepest love, affection and thanks go to my family. I owe a lot to my parents, my husband, my daughter and my brother for their unconditional love, understanding, continuous help and support. I thank my parents for making me the person I am and for believing in me. I thank my brother for being there for me. My very special thanks go to my husband and my daughter for their endless love, help, support and encouragement, and also for bearing my endless hours spent on the computer programming and writing, for sacrificing the well-deserved holidays and outings, and for having to learn how to cook. I am very fortunate to have you in my life.

To my family

Table of Contents

Declaration **i**

Abstract **iii**

Acknowledgements **v**

List of Figures **xix**

List of Tables **xxii**

List of Acronyms **xxiii**

1 Introduction **1**

1.1 Research Background and Motivations 1

1.2 Research Aim and Objectives 5

1.3 Research Approach 6

1.4 Research Achievements 11

1.5 Research Focus and Scope 14

1.6 Thesis Outline 18

2 Web Personalisation in E-Commerce Research and Practice **21**

2.1 Introduction 21
2.2 Web Personalisation in the Literature 22
2.2.1 Literature Review Methodology 22
2.2.1.1 Literature Review Sources 22
2.2.1.2 Performing Searches 26
2.2.1.3 Criteria for Selection 27
2.2.1.4 Literature Analysis 27
2.2.2 Literature Review Findings 28
2.2.2.1 Top-Level Classification and Characteristics of Web Personalisation Research 28
2.2.2.2 More Focused Classification of Research on Web Personalisation 31
2.2.2.3 Semantic Web and Ontology-Based Web Personalisation 38
2.2.2.3.1 Why Ontology? 39
2.2.2.3.2 Ontologies on the Semantic Web 41
2.2.2.3.3 Ontological Personalisation in E-Commerce 42
2.3 Web Personalisation in Practice 45
2.3.1 Methodology for Surveying Current Professional Practice . . . 45
2.3.2 Findings from Surveying Current Professional Practice 47
2.4 *"Hows"* of Web Personalisation: Approaches, Methods and Techniques 54
2.4.1 Process of Web Personalisation 54
2.4.1.1 Data collection 55
2.4.1.2 Data Analysis 59

2.4.1.3 User Profile Generation, Classification and Maintenance 62

2.4.1.4 System Implementation 64

2.5 Conclusions 66

2.6 Chapter Summary 69

3 *PERSONTO* Requirements Specification and System Design 71

3.1 Introduction 71

3.2 Purpose and Scope of *PERSONTO* 72

3.3 Requirements Elicitation and System Design Methodology 72

3.4 Requirements Specification for *PERSONTO* 73

3.4.1 *PERSONTO* Stakeholders and Use Cases 73

3.4.2 System Requirements 75

3.5 System Design of *PERSONTO* 79

3.5.1 *PERSONTO*'s Top-Level Architecture 80

3.5.2 The Process of E-Commerce Website Personalisation in *PERSONTO* 83

3.5.3 Specification of System Functionality 88

3.5.3.1 Pre-Processing Phase: Batch Creation of Initial Profiles 88

3.5.3.2 Run-Time Phase 89

3.5.3.3 Run-Time Architecture 90

3.5.3.4 Profile Access and Control 91

3.5.3.4.1 Adding Additional Individual Profiles 91

3.5.3.4.2 Profile Updating 91

3.5.3.4.3 Profile Deletion 97
3.5.3.4.4 Profile Viewing and Importing 98
3.5.3.5 Personalisation of Web Pages 98
3.6 Chapter Summary 99

4 *OntoProfi*: Ontology for Profiling Online Shoppers 101

4.1 Introduction 101
4.2 *OntoProfi*'s Engineering Methodology 102
4.3 *OntoProfi* Design Specification 105
4.3.1 Ontology Representation 105
4.3.1.1 OWL Overview 106
4.3.1.2 OWL's Building Blocks 108
4.3.1.3 *OntoProfi*, Reasoning and OWL DL 110
4.3.2 *OntoProfi* Class Hierarchy 113
4.3.2.1 “Customer” Sub-domain 114
4.3.2.2 “Product” Sub-domain 118
4.3.2.3 “Business” Sub-domain 120
4.3.2.4 “System” Sub-domain 121
4.3.3 *OntoProfi* Object and Datatype Properties 123
4.4 Chapter Summary 125

5 Elicitation of Customer Profiles 127

5.1 Introduction 127
5.2 Case Study Research Methodology 128

5.2.1 Research Design Phase 129

5.2.2 Data Collection Phase 134

5.2.3 Data Analysis Phase 138

5.2.3.1 Data Analysis Process 138

5.2.3.2 Data Coding and Conceptualisation 139

5.2.4 Case Study Outcomes Phase 141

5.3 Characteristics of Online Shopping: Case Study Findings 141

5.3.1 Common (Typical) Factors and Aspects for Online Shopping . 142

5.3.1.1 Motivating Factors for Online Shopping 143

5.3.1.2 De-Motivating Factors for Online Shopping 149

5.3.1.3 High Focus 154

5.3.1.4 Habitual Shopping 155

5.3.1.5 Non-Spontaneity and Non-Impulse Buying 157

5.3.2 Differentiating Aspects and Factors 158

5.3.3 Personalisation and E-Shoppers' Profiles 169

5.3.3.1 What Can We Profile and Personalise? 170

5.3.3.2 E-Shoppers Profiles for E-Commerce Personalisation 172

5.3.3.2.1 *Budget* vs *Quality* Profile Entity 173

5.3.3.2.2 *Textual* vs *Visual* Profile Entity 174

5.3.3.2.3 *Satisficer* vs *Maximiser* Profile Entity 174

5.4 Extending *OntoProfi* Ontology 174

5.4.1 Extending the `Customer_Profile` Class 175

5.4.2 Customers as *OWL Individuals* 178

5.5 Chapter Summary 179

6 Testing Feasibility of Approach: Prototype of *PERSONTO* 181

6.1 Introduction 181

6.2 Prototype Implementation 182

6.2.1 Development Setup: Languages, Technologies, Data and Environment 182

6.2.2 Website Information Architecture 184

6.2.3 Database Model 186

6.2.4 Website Personalisation: Experiments and Results 187

6.2.4.1 Customer Profiles Test Data: Creation, Assertion and Inference 188

6.2.4.2 Personalisation of Web Pages 190

6.3 Assumptions and Limitations of Prototype 196

6.4 Chapter Summary 197

7 Evaluation 199

7.1 Introduction 199

7.2 Evaluation Strategy 199

7.3 Feasibility of Ontology-Based E-Commerce Website Personalisation . 202

7.4 Flexibility and Extendibility 203

7.5 Reusability of the Approach 204

7.6 Stakeholders' Acceptance 205

7.6.1 Evaluation of *PERSONTO*'s Acceptance by End User (Customer) 205

7.6.1.1 Survey Instrument 205
7.6.1.2 Survey Data Collection: Research Setting and Procedures 209
7.6.1.3 Analysis and Results 210
7.6.2 Evaluation of *PERSONTO*'s Acceptance by E-Commerce Developers 218
7.6.2.1 Participants, Methodology and Procedure 218
7.6.2.2 Findings and Discussion 219
7.7 Research Limitations 222
7.7.1 Research Approach 223
7.7.2 Project Focus 226
7.7.3 System Design and Implementation 226
7.7.4 Evaluation 227
7.8 Chapter Summary 230

8 Conclusions and Future Work 231
8.1 Introduction 231
8.2 Synopsis 231
8.3 Fulfilment of Research Objectives 233
8.4 Contribution to Research and Practice 237
8.5 Future Work 238
8.6 And Finally... 241

References 243

Personal Communications **260**

Appendix A Case Study Protocol **261**

Appendix B “Screening” and “Products” Questionnaires **265**

Appendix C Initial Product List for Participant Observation **269**

Appendix D Origins of *OntoProfi* Ontology Classes **273**

Appendix E Alternative Languages, Technologies and Tools for Development of a Web Personalisation System **277**

Appendix F User Evaluation Survey Questionnaire **279**

Appendix G Research Quality and Rigour **283**

List of Figures

2.1 Topics in Literature on Web Personalisation 29

2.2 Classification of Web Personalisation 34

2.3 Web Personalisation: Process and Approaches 55

3.1 Use Case for *PERSONTO* 74

3.2 Architecture of *PERSONTO* 80

3.3 E-Commerce Website Personalisation in *PERSONTO*: Process and Approaches 84

3.4 Pre-Processing Phase: Creating Initial Customer Profiles 89

3.5 Top-Level Profile Processing (*Run-Time*) 90

3.6 Customer Profile: Access and Control Functions 91

3.7 “Profiling Survey” page. 93

3.8 Survey-Based Profile Processing (*Run-Time*; update of Fig. 3.5) 94

3.9 Updating of Customer Profile (*Survey-Based*) 95

3.10 Updating Customer Profile (*“Non-survey", Manual Update*) 96

3.11 Deleting Customer Profile 97

3.12 Customer Profiling Activity Diagram 99

4.1 Ontology Engineering Process *(adapted from (Noy & McGuinness 2001))* . . . 102

4.2 XML-based Ontology Languages Hierarchy *(adapted from (Jones, et al. 2002))* . . . 106

4.3 Venn Diagram of OWL 2 Syntactic Subsets *(Source: W3C OWL Working Group (2009))* . . . 112

4.4 Top-Level Sub-Domains of *OntoProfi* . . . 114

4.5 Class Hierarchy of **Customer-Related** Sub-Domain . . . 115

4.6 Class Hierarchy of **Product-Related** Sub-Domain . . . 118

4.7 Class Hierarchy of **Business-Related** Sub-Domain . . . 120

4.8 Class Hierarchy of **System-Related** Sub-Domain . . . 122

4.9 Object and Datatype Properties . . . 123

4.10 Key Relationships in *OntoProfi* (high-level) . . . 124

5.1 Overview of Online Shoppers Case Study Research Process . . . 129

5.2 Characteristics of Online Shopping . . . 142

5.3 Common (Typical) Factors and Aspects for Online Shopping . . . 143

5.4 Motivating Factors and Aspects for Online Shopping . . . 144

5.5 De-Motivating Factors and Aspects for Online Shopping . . . 149

5.6 Differentiating Aspects and Factors for Online Shopping . . . 158

5.7 Extended `Customer_Profile` class . . . 175

5.8 Key Relationships in Extended *OntoProfi* (high-level) . . . 176

6.1 Information Architecture of *PERSONTO* website . . . 185

6.2 EER Model for *PERSONTO* Database . . . 186

6.3 The process of Customer Profiles Test Data Creation 188

6.4 Example of customers inferred by the reasoner as beloning to a particular customer profile - here as `Maximiser_bv` profile *(visualised in Protégé v. 4.1)* 189

6.5 Default "Home" page *(customer is not logged in)*. 191

6.6 "Home" page for *"Maximiser-budget-textual"* profile. 192

6.7 "Category" page for *"Maximiser-quality-textual"* profile. 193

6.8 "Category" page for *"Satisficer-quality-textual"* profile. 194

6.9 "Single Product" page for *"Maximiser-budget-visual"* profile. 195

6.10 "Single Product" page for *"Satisficer-budget-visual"* profile. 195

7.1 Design-Science Research Guidelines (Hevner, et al. 2004) 200

List of Tables

1.1 Research Strategy and Research Objectives 10

2.1 List of Academic Journals for Literature Review 24

2.2 Aspects Coverage in Extant Reviews of Web Personalisation Research 32

2.3 Personalisation Features Offered by Top 10 E-Commerce Clothing E-Tailors 48

2.4 Advantages and Disadvantages of Personalisation Approaches *(adapted from (Gao, et al. 2010))* 62

4.1 OWL Class Descriptions and Axioms, Required in *OntoProfi* 111

5.1 Number of Case Study Respondents *(per age group)* 137

5.2 Evaluation of Possibility of Personalising Particular Differentiating Factors and Aspects 170

5.3 E-commerce Shoppers' Profiles 173

7.1 Application of Hevner et al.'s (2004) Guidelines to This Research . . . 201

7.2 Survey Instrument Questions 207

7.3 Cronbach's Alpha Reliability Coefficients 208

7.4 PERSONTO Survey Respondents 209

7.5 *PERSONTO* Customer Acceptance Survey: Central Tendency and Variability . . . 211

7.6 PERSONTO Customer Acceptance Survey: Frequency Distribution of Survey Items and Corresponding Bar Charts . . . 216

7.7 PERSONTO Customer Acceptance Survey: Frequency Distribution of All Survey Items. . . . 217

7.8 MSc Students Workshop ONE Responses *(N=17)* . . . 228

7.9 MSc Students Workshop TWO Responses *(N=15)* . . . 229

8.1 Fulfilment of Research Objectives . . . 234

G.1 Ensuring Quality in Interpretivist Research . . . 285

List of Acronyms

API	Application Program Interface
B2B	Business-to-Business
B2C	Business-to-Consumer
C2C	Consumer-to-Consumer
CBF	Content-Based Filtering
CF	Collaborative Filtering
CMS	Content Management System
CRM	Customer Relationship Management
CS	Computer Science
CSS	Cascading Style Sheets
CWA	Closed World Assumption
DB	Database
DL	Description Logic
EC	Electronic Commerce
EU	European Union
GTM	Grounded Theory Methodology
IDE	Integrated Development Environment
IS	Information Systems
IT	Information Technology
ITU	Intention to Use
JSP	Java Server Pages
OWA	Open World Assumption
OWL	Web Ontology Language
PC	Personal Computer
PEOU	Perceived Ease of Use
PU	Perceived Usefulness
RBF	Rule-Based Filtering
RDBMS	Relational Database Management System
RDF	Resource Description Framework

REF	Research Excellence Framework
SOAP	Simple Object Access Protocol
UI	User Interface
UML	Unified Modeling Language
URL	Uniform Resource Locator
XML	Extensible Markup Language

Chapter 1

Introduction

"#1: Focus on the user and all else will follow"

Google. *"Ten Things We Know to Be True"* (1)

1.1 Research Background and Motivations

Electronic commerce has become an important part of our consumer lives. We increasingly choose to do more and more of our shopping online; and our internet spending grows year on year. UK online retail sales in 2013 amounted to £91 billion (a rise of 16% compared with 2012), and have been predicted to grow further - the forecast for 2014 online sales is £107 billion (a market growth of 17%) (IMRG 2014).

Along with the growth of online sales, we have witnessed an increase in the number of e-commerce retailers. The number of such businesses more than doubled in five years from 2008 to 2012 (Internet Retailing 2013). This has inevitably led to increased competition, putting additional demands on existing companies as well as on new market entrants to find ways to ensure that their growth (if not just survival) and competitiveness are sustainable and can evolve.

(1) `https://www.google.co.uk/about/company/philosophy/`

Businesses employ a variety of Information Systems/ Information Technology strategies to deal with competitive forces to ensure a company's competitive advantages over their rivals. Web personalisation has become an important strategy in these efforts.

In fact, web personalisation is now considered to be *"a critical element of contemporary electronic businesses"* (Koutsabasis, et al. 2008) and *"an important approach element in customer relationships and Web strategies"* (Fan & Poole 2006). It has become a practice that is increasingly common and something that e-shoppers expect to have, and without which *"sites can't hope to keep buyers around long"* (Eisingerich & Kretschmer 2008). So, the number of websites, which offer personalised and adaptive content to their users, is increasing (Lee 2000).

The benefits of web personalisation can be assessed and discussed from two perspectives: *"business"* and *"customer"*. (1)

From the *business perspective*:

- Personalisation has been proven to be a successful marketing tool that is capable of **maximising immediate and future business opportunities** and **offering competitive advantages** to online businesses (Pramataris, et al. 2001), (Tam & Ho 2006), (Blasco-Arcas, et al. 2013).

- Personalisation can help online businesses with **improving their promotion** and **sales efforts** (Schonberg, et al. 2000), (Tam & Ho 2005); and is seen as a way to **generate an additional revenue** (Ihlström & Palmer 2002).

- It can also ensure the **customer involvement** and their **increased participation** with an e-commerce company (Blasco-Arcas et al. 2013), (Blasco-Arcas, et al. 2014).

(1) The words in bold font in this section are the author's synthesis of the aspects identified in the literature. The bold formatting is used to enhance readability, and not to denote direct quotes. Direct quotes are shown using the quotation marks.

- Personalisation can help to build customer **loyalty** (Kwon & Kim 2012), (Martínez-López, et al. 2010) as well as meaningful and strong **one-to-one working, ongoing relationships** (Riecken 2000), (Shen & Ball 2009).

From the *customer perspective*, e-commerce personalisation can be justified by its impact on the e-commerce users' attitudes and behaviour:

- We now have evidence that e-commerce customers have **diverse motivational needs** for using e-commerce websites, have **a variety of requirements** for website design (MacAulay, et al. 2007) and **evaluate website design in different ways** (Shergill & Chen 2005). This suggests that personalising e-commerce websites can bring benefits and advantages to the users.

- There is strong evidence that e-shoppers **perceive personalisation as useful**, and are **more willing to further explore content** that has been personalised for them (Tam & Ho 2006).

- Moreover, tailoring information and content to a particular user is known to **alleviate *"the burden of information overload"*** (Gauch, et al. 2007) (Liang, et al. 2006), (Sieg, et al. 2007), and **improve consumers' decision making** (Tam & Ho 2006) - consequently leading to **increased satisfaction** with the online merchant (Kwon & Kim 2012), (Liang et al. 2006), (Rust & Miu 2006).

Overall, web personalisation can make the Web more intelligent (Malik & Fyfe 2012), which can help both perspectives to achieve their goals and reap the benefits of it.

This project is motivated by the fact that despite the proven benefits and the presence of a substantial number of academic and industry solutions to web personalisation (see Chapter 2), there are still gaps that need to be addressed. The analysis of literature and practice shows that there is a noticeable bias towards researching and implementing personalisation of website content, while the research and practice on personalisation of website content presentation is limited.

Looking at using the Semantic Web technologies for web personalisation, such as ontologies, is of particular interest, as these technologies can guarantee flexibility, interoperability and reusability of the web-based systems, which is of significant importance due to the fact that the Web is constantly evolving. The same can be said about businesses - the business world is highly dynamic, making the system's qualities of flexibility, interoperability and reusability important (if not vital). Nevertheless, as literature and practice both show, the Semantic Web technologies, which could assist this aim, are still under-utilised.

Ontologies are considered to be one of the core Semantic Web technologies, capable of enhancing knowledge discovery, exchange and management. Ontologies and ontological user profiling have been successfully used in e-commerce personalisation. However, their use focused predominantly on product or service recommendation (i.e. personalisation of content), e.g. in (Adda, et al. 2006), (Cantador, et al. 2008a), (Cena, et al. 2012); and there is a limited use of ontology-based profiling for personalisation of content presentation in e-commerce.

This project set out to address these gaps, and in particular, personalisation of content presentation that utilises the Semantic Web technologies. It investigates a novel, flexible, extendible and reusable approach to the design and development of e-commerce personalisation that is based on ontological user profiling. We design and implement an e-commerce personalisation system, which we named *PERSONTO* (1); and a customer profiling ontology *OntoProfi* (2), which enables personalisation in our system. The acceptability of the proposed approach by the prospective stakeholders is also evaluated. The results of the evaluation suggest a high level of acceptability of the proposed approach and system.

(1) **PERS**onalisation based on **ONTO**logy
(2) **Onto**logy for **Profi**ling customers

1.2 Research Aim and Objectives

The **aim** of the research is:

> ***to propose, test and evaluate the feasibility and acceptability of a flexible, extendible and reusable ontology-based e-commerce personalisation approach.***

To achieve this aim, the following **objectives** have been identified:

OBJ.1: Develop rich insights into online shopping, customer profiling and e-commerce websites personalisation:

(a) Evaluate current academic and professional developments in the area of web personalisation through a survey of academic literature and commercial practices.

(b) Acquire a holistic, comprehensive and up-to-date understanding of key factors that influence e-commerce users when shopping online, which is achieved by conducting a multi-case study of e-commerce consumers:

i. Identify, describe and explain common factors that influence B2C e-commerce customers.

ii. Determine differentiating factors that affect e-commerce website users, based on different users' needs and need recognition level, goals, experiences, shopping strategies, etc. - in order to determine what online shopping factors/ aspects can be personalised.

iii. Identify different customer profiles to be used in a customer profiling ontology.

OBJ.2: Design, implement and test a system that personalises an e-commerce website for its customers by means of ontology-based customer profiling:

(a) Propose conceptual design of such a system, including system architecture and functionality.

(b) Design and implement a flexible and extendible customer profiling ontology, which will enable personalisation of various types of interactions between an EC system and its users.

(c) Implement and test a *'proof-of-concept'* prototype to demonstrate the most important and representative functionality and usage of the proposed system.

OBJ.3: Evaluate the research outcomes against the research aim:

(a) Evaluate the achievement of feasibility, flexibility and extendibility of the system; as well as 'systematicity' and reusability of the proposed e-commerce web personalisation approach.

(b) Evaluate the relevant stakeholders' acceptance (*'buy-in'*) of the proposed approach and system.

1.3 Research Approach

The nature of this research and its aim and objectives, position our research as multidisciplinary, integrating the traditions and elements of the Information Systems (IS) and Computer Science (CS) disciplines. More specifically, our research follows the *"Design Science"* paradigm (Hevner et al. 2004), where the main emphasis is on the creation of an innovative artefact.

A number of research approaches and strategies are traditionally used in the two disciplines and in design-science projects. At the top of the research approach hierarchy, research can be classified as quantitative, qualitative and mixed (i.e. a combination of the two). Research can be also classified as positivist and interpretivist.

To achieve their aims and objectives or to answer research questions, researchers employ various strategies, the most common (standard) of which include: survey, "design and creation", experiment, case study, action research, grounded theory and ethnography (Oates 2006), (Creswell 2003), (Yin 2003).

Upon examination and analysis of the research aims and objectives, a mixed research approach that combines qualitative and quantitative research methods was chosen for the project. This was dictated by the fact that our project involved understanding and evaluating complex behavioural and attitudinal phenomena, and examination of these using a mixed research approach is common in IS research. Moreover, using purely quantitative or qualitative approaches is typically seen as too limiting (Venkatesh, et al. 2013), therefore, using purely quantitative methods might not provide us with rich insights into the phenomenon of online shopping, while using only qualitative methods would not allow us to evaluate user acceptance of the proposed personalisation.

This project employs a research strategy, which uses a combination of: ***"design and creation"***, ***case study***, and ***survey*** approaches:

- **"Design and creation"** *(see Chapters 3, 4 and 6)*:

  Choosing *Design Science* as the main research paradigm determines the use of a *"design and creation"* approach. This approach is suitable if an experimental implementation of a system (such as prototype) is set as a research objective, which is the case in this project.

The focus of the *"design and creation"* part of the project is on the design and implementation of an ontology-based e-commerce personalisation system. More specifically, this part involves: proposal of a system architecture and determination of appropriate approaches and processes for deployment of e-commerce personalisation in *PERSONTO (see Chapter 3)*, conceptual design and implementation of the *OntoProfi* ontology *(see Chapter 4)*, and implementation of a proof-of-concept prototype of *PERSONTO*, which demonstrates the most representative functionality of the proposed system *(see Chapter 6)*.

- **Case Study** *(see Chapter 5)*:

  In order to build a customer profiling ontology, we need to acquire a comprehensive and up-to-date understanding of common and differentiating factors that influence shoppers to buy online. Identification of common factors is needed to provide a better understanding of the context of online shopping, while the differentiating factors allow determination of what can be personalised and what e-shopper profiles can be employed in the ontology *(see Chapter 4)*.

  To do this a multi-case case study of Business-to-Consumer (B2C) e-commerce shoppers is chosen as the most appropriate strategy for this part of the research. Our case studies use semi-structured interviews and observations of a sample of e-shoppers to collect qualitative and quantitative data necessary for the data analysis *(see § 5.2)*. The data analysis is accomplished by following the interpretivist paradigm.

  This approach is considered to be appropriate because case studies allow researchers and developers to focus on and research in-depth a particular issue, situation or phenomenon so that contemporary, up-to-date, rich insights into complex relationships and processes are uncovered, described or explained (Oates 2006), (Yin 2003). In particular, we chose the use of multiple case studies, where each interview and observation of users is viewed as a single case study. This way, *"more sophisticated descriptions and more powerful explanations"* can be developed (Miles & Huberman 1994).

- **Survey** *(see Chapter 7)*:

  Surveys are commonly used to evaluate or validate the research outcomes in a systematic and standardised way (Oates 2006). We use surveys to evaluate the acceptance of the proposed system, an aspect specified in the project's aim and objectives. These evaluations use quantitative and qualitative methods.

  The PERSONTO stakeholders' acceptance is evaluated quantitatively and qualitatively: evaluation of end-users' (potential e-shoppers) acceptance of the proposed system is based on analysis of quantitative data gathered through a questionnaire (*see* § 7.6.1); whilst the industry's (web developers') acceptance is evaluated by analysing qualitative data gathered by means of structured questionnaires and follow up interviews (§ 7.6.2).

In addition, the current state of academic developments in the area of web personalisation is evaluated by conducting a literature review, while the use of web personalisation techniques by the industry practitioners is analysed by evaluation of top 10 e-commerce companies (see §§ 2.2.2 and 2.3).

The fulfilment of the particular system characteristics (qualities) specified in the research aim (i.e. flexibility, extendibility and reusability) is evaluated using the author's analysis of the outcomes of the project (see §§ 7.3, 7.4 and 7.5).

Table 1.1 summarises where each research strategy and associated components addresses the project's research objectives (*see* § 1.2):

Research Strategy/ Component	Objectives Addressed
Literature review; analysis of professional practice (industry)	*OBJ. 1 (a)*
"Design and creation" (*PERSONTO* system):	
- System design and functionality	*OBJ. 2 (a)*
- Customer profiling ontology (*OntoProfi*)	*OBJ. 2 (b)*
- Prototype	*OBJ. 2 (c)*
Case Study	*OBJ. 1 (b) i - iii*
Survey	*OBJ. 3 (b)*
Author's analysis/ evaluation	*OBJ. 3 (a)*

Table 1.1: Research Strategy and Research Objectives

Each stage of the project follows a particular **research methodology** or **strategy**, appropriate to the project's aims and objectives; and these are covered in the relevant sections of this thesis, as follows:

- methodology for the **literature review** - *see* § 2.2.1;
- **requirements elicitation and system design** methodology - see § 3.3;
- ***OntoProfi* ontology** engineering methodology - *see* § 4.2;
- methodology used for the **case study** - *see* § 5.2;
- **evaluation strategy** - *see* § 7.2.
- a standard **'proof-of-concept' prototyping** is used for testing of the feasibility of the proposed approach (Chapter 6)

1.4 Research Achievements

- **The research contributes to the area of ontology-based personalisation of the content presentation in e-commerce websites by offering a novel, feasible and flexible approach.**

As discussed in §§ 2.2.2 and 2.5, the vast majority of existing research focuses on the personalisation of content (i.e. product recommendation).

Within the research on the content presentation personalisation (i.e. personalisation of the layout/ user interfaces, and the amount and type of product information), some research has been produced, however, it has significant limitations with regard to the domain coverage and methods used for user (customer) profiling. More specifically, research in this area either covers the domain of e-commerce (explicitly), but does not use ontological profiling; or it uses ontological profiling, but does not focus on e-commerce. The research that combines the two aspects is scarce: in our sample of literature (discussed in Chapter 2) there were only two papers that offered the coverage of ontological content presentation personalisation in e-commerce: (Gaulke & Ziegler 2015) and (Lei, et al. 2003). Nevertheless, the main focus of the first paper was on the mobile context environment, and the proposed approach was mainly "conceptual", with no "experimental" demonstration or testing (e.g. by a prototype implementation). The second paper presented a conceptual framework of ontology-based personalisation of web pages. The capabilities of the framework were demonstrated in a case study, however, the case study's focus was not an e-commerce system.

Our research contribution is in addressing the two limitations, i.e. the bias towards the personalisation of content and limited amount of research on the content presentation personalisation in e-commerce, - by proposing and testing an e-commerce personalisation system, which uses ontological customer profiling to personalise web pages layout and the amount and type of product information.

We demonstrate that the proposed approach is *feasible*, by showing that is it possible to implement it. This is tested in the *'proof-of-concept'* prototype (Chapter 6). The approach is *flexible* - as there is no reliance on a particular platform or programming language; and some parts (in particular, the reasoner) can be substituted with an alternative (see § 6.2.1 and Appendix E).

- PERSONTO is designed and implemented using **a systematic, reusable and extendible** procedures, processes tools and techniques, **acceptable to the relevant stakeholders**.

  PERSONTO's prototype is designed and implemented by using *systematic*, *reusable* and *extendible* procedures, processes, tools and techniques. This is another contribution of our research that is envisaged to be of interest to research and professional practice.

  All stages of *PERSONTO*'s design and implementation are specified and discussed throughout this thesis, and can be replicated and adapted. For example, the proposed case study protocol for elicitation of customer profile types, which specifies the process and activities to be carried out in a case study (see § 5.2.1, Appendix A), can be adapted to the needs of future projects and in different research and practice settings. Similarly, the system acceptance survey can be adapted and reused in the future in other projects. It is believed that the protocol and survey can be of use to researchers and practitioners to save time and effort on their development for another project.

The evaluation of the project outcomes suggest an overall high level of acceptance of the proposed approach by the evaluation group (see § 7.6.1). The evaluation also indicates that all parts of the proposed approach can be treated as a *'template'* in future work.

- The research develops a **reusable, extendible and Semantic Web compatible customer profiling ontology**.

  The implemented ontology offers a domain representation of e-commerce, and enables customer profiling by making use of reasoning capabilities. The ontology contains a number of classes (§ 4.3.2) that are typical for the e-commerce domain to enable the ontology's reuse and extensions in other future e-commerce projects. The ontology also specifies a number of restrictions (rules), which facilitate customer profiling in *PERSONTO* (see § 5.4.1). As demonstrated in § 5.4.1, the class hierarchy and restrictions can be reused, adapted and extended to suit the business model and rules of an e-commerce company, or objectives of future research projects, removing the need to develop a customer profiling ontology from *'scratch'*.

  The design and implementation of the ontology relies on the technologies compatible with the Semantic Web. In particular, *OntoProfi*'s representation uses OWL (Web Ontology Language), the *de facto* ontology standard for the Semantic Web (see § 4.3.1). This ensures that *OntoProfi* can be used to merge or align with external ontologies, which e-commerce companies use in their systems.

Our experiments demonstrate that the proposed and implemented *OntoProfi* ontology can be successfully used for customer profiling and subsequent website personalisation (see § 6.2.4).

- The research presents a convenient **platform for investigation of effects of personalisation**.

  In a wider research context, the presented system to personalise websites can be adopted and used in Behavioural Science research for theoretical and empirical investigations of the role and effects of personalisation, for example, to examine whether the use of personalisation increases the overall website usage, improves customers' perception of a website, affects the amount of purchases. Due to its flexibility and reusability, the prototype can be adapted to new research aims and objectives, without the need for a *'green-field'* implementation, saving research effort and resources.

1.5 Research Focus and Scope

Web personalisation can be investigated from three distinct perspectives - *business*, *behavioural* and *technology*:

- **Business Perspective:**

  From a *business perspective*, personalisation is seen as a *'what'*, and is typically defined as a business strategy, business process or business feature or practice (see (Mobasher, et al. 2000), (Eisingerich & Kretschmer 2008)). In this case, web personalisation is seen predominantly as a *marketing strategy* to generate new or enhanced business opportunities (Ho 2006) - by designing and implementing business processes for tailoring *products*, *services* or *content* so that a particular customer's requirements or wishes are met (Turban, et al. 2006), (Cingil, et al. 2000). In fact, often, personalisation is considered to be part of a company's Customer Relationship Management (CRM) programme (Ho 2006), (Martínez-López et al. 2010), (Mulvenna, et al. 2000). In this situation, personalisation is about employing various personalised marketing strategies and CRM activities such as, for example: tailored communication between the company and customers; highly targeted promotional campaigns or other marketing communication; personalised information presentation; adaptive interfaces.

- **Behavioural Perspective:**

  The *behavioural perspective* deals with developing and testing theories that explain and predict the user's intentions to use the system, perceived usefulness and benefits. This perspective is associated with Behavioural Science research, and, often, the objects of their study are the artefacts (i.e. systems, applications, tools), which were developed as the *'technology perspective'* entities.

- **Technology Perspective:**

  From a *technology perspective*, web personalisation can be defined as a *'how'*, i.e. a tool or application feature (Kramer, et al. 2000). In this case, web personalisation covers aspects of system design and implementation, e.g. development of particular functionality to enable the aforementioned marketing strategies of offering personalised content, products, and services.

  Achievement of website/ web page personalisation relies on two elements: 1) *"parameters"* of what needs to be personalised; 2) a *"trigger"* (request) for personalisation to start. The two elements can be implemented as either *system-driven* or *user-driven* [1].

The *system-driven* implementation does not require the user's participation. Instead, the personalisation *"parameters"* and *"trigger"* are executed automatically. For example, the system determines the most suitable user's parameters, based on the user's data, already collected by the system; and starts and completes website personalisation at a pre-determined moment, e.g. when the user is logged in.

The *user-driven* personalisation depends on the explicit user's input, i.e. the system requests the user to specify the explicit personalisation "parameters" (e.g. the desired product characteristics or web page layout preferences), and waits for the user input to start personalisation.

A combination of the two approaches is also possible. For instance, the users specify their preferences explicitly (*user-driven* approach), and then the system personalises the web pages without any further user input (*system-driven*). Alternatively, the system determines the personalisation *"parameters"*, but lets the users trigger when the personalisation should be done.

The first perspective is typically of interest to Business and Marketing research. Research belonging to the second and third perspectives is found predominantly in the Information System (IS) and Computer Science (CS) disciplines. Hevner et al. (2004) classified the research within these perspectives as *"Behavioural Science"* and *"Design Science"*.

The focus of our research is on the **"technology perspective"**. In particular, web personalisation in our system is implemented using *user-driven* "parameters", but

(1) The "system-driven" approach is also referred to as *"implicit"* and the user-driven as *"explicit"* (Fan & Poole 2006). Sometimes, the user-driven approach is also differentiated as *"customisation"* (Mobasher 2007).

a *system-driven* "trigger" (see § 3.5.2). We also focus on personalisation of e-commerce website's **content presentation**, due to the lack of research and commercial development in the area (see §§ 2.2.2 and 2.3). Extending e-commerce personalisation to other aspects, such as personalisation of content (e.g. product recommendations) and supporting business-related activities (e.g. business rules, CRM and marketing) is taken into consideration in the design of the proposed system and development of the ontology, but not experimentally tested in this research. Implementation and execution of the system-driven personalisation "parameters" and user-driven "triggers" needs future investigation.

In addition, some aspects of e-commerce system development, such as non-functional system quality characteristics (security and data quality assurance) are not explicitly covered in this research, as these are outside of the project's aim and scope. These are not believed to affect the research findings, and present an opportunity for interesting future work.

1.6 Thesis Outline

This thesis consists of eight chapters, and is structured as follows:

Chapter 1 "Introduction": introduces the research by providing research background and motivations, specifying the project's aims and objectives, outlining the research methodology, stating the research scope and focus, and highlighting research contributions.

Chapter 2 "Web Personalisation in E-Commerce Research and Practice": reviews the current *state-of-the-art* of web personalisation in the literature and professional practice, with a particular focus on e-commerce, Semantic Web and ontology-based user profiling. It discusses limitations of current research and practice in this domain, and outlines how these are addressed in our research.

Chapter 3 "*PERSONTO* Requirements Specification and System Design": describes and discusses the system requirements and system design of the proposed web personalisation system. It also outlines the proposed approach; and details the proposed architecture, processes, methods and techniques; and system functionality to achieve e-commerce website personalisation in this project.

Chapter 4 "*OntoProfi*: Ontology for Profiling Online Shoppers": describes and discusses the design and implementation of an ontology to profile e-commerce users, *OntoProfi*. The chapter outlines an engineering methodology used to create *OntoProfi*; explains the reasons for choosing Web Ontology Language (OWL) to represent *OntoProfi*; and describes the conceptual design and implementation of *OntoProfi*.

Chapter 5 "Elicitation of Customer Profiles": discusses the design and deployment of a case study as a means for eliciting online customer profiles. In particular, the chapter specifies the research design, data collection, analysis and data conceptualisation strategies and processes. The second part of the chapter covers the findings from the case study, such as common and differentiating factors that influence e-shoppers. It concludes with an analysis of which factors can be used for user profiling and e-commerce personalisation, and a proposal of a set of e-shoppers' profiles and how these are used to modify the customer profiling ontology, *OntoProfi*.

Chapter 6 "Testing Feasibility of Approach: Prototype of *PERSONTO*" discusses the implementation of PERSONTO's prototype, including a description of the prototype architecture and implementation details; and an examination of the assumptions and limitations.

Chapter 7 "Evaluation": evaluates the achievements of the research against the research aim, including evaluation of the system characteristics stated in the hypothesis (such as feasibility, flexibility, extendibility and reuse); examines the acceptance of the proposed system by two stakeholders, i.e. potential e-customers and e-commerce developers. It also discusses the research limitations.

Chapter 8 "Conclusions and Future Work": summarises the work presented in this thesis, including the achievement of the set research objectives and contributions to research. It also highlights areas for future work.

Chapter 2

Web Personalisation in E-Commerce Research and Practice

2.1 Introduction

Web personalisation is typically defined as the delivery of tailored website content that matches an individual user's requirements, needs, interests or wishes (Chaffey 2011). In the e-commerce context, this includes an ability to offer personalised products, services, marketing communications, order fulfilment and other Customer Relationship Management (CRM) activities, individualised to each customer (Adomavicius & Tuzhilin 2005), (Ho 2006), (Turban et al. 2006).

Over a relatively short period of time, web personalisation has become a fast growing research and practice area, which can be attributed to the increase in demand for customer-centric services (Goy, et al. 2007). The area of web personalisation comprises a sizeable body of academic research and professional practice, covering a great number of variables and aspects, offering a number of diverse perspectives, and employing a considerable variety of methods and approaches.

This chapter presents a critical review of the research in the area of web personalisation and compares it with the professional practices of e-commerce companies.

More specifically, the chapter:

- identifies and explains the methodologies used for reviewing research and professional practice;
- presents a classification of research on web personalisation, which is used to structure our literature review, focusing on the areas that are of particular relevance to this project;
- reviews the current research on web personalisation, with a particular focus on e-commerce, Semantic Web and deployment of ontology-based user profiling;
- surveys the use of personalisation by e-commerce companies (professional practice);
- examines diverse approaches, methods and techniques (the *"hows"*), employed to achieve web personalisation;
- identifies and discusses gaps and limitations in current research and practice, and the areas that provide the foundations for this project's aim and objectives.

2.2 Web Personalisation in the Literature

2.2.1 Literature Review Methodology

2.2.1.1 Literature Review Sources

Our literature review used a systematic and rigorous review process for the analysis and synthesis of extant academic works. To locate these articles, relevant literature was identified using four sources: journals publishing in e-commerce research, electronic databases, the *Zetoc* alert service (1), and relevant articles cited in the papers retrieved using the first three sources.

- **Journals publishing in e-commerce research:**
  - 20 academic, peer-reviewed journals were selected from the list of publications, specified by Bharati & Tarasewich (2002) as the most relevant, high quality and recognised as impactive outlets for e-commerce research. Five of these specialise in the e-commerce domain. (See Table 2.1; the journals are arranged in alphabetical order.)

(1) http://zetoc.mimas.ac.uk

	Journal Acronym	Journal Title	Published from	Specialise in Electronic Commerce?
1	BIT	Behavior & Information Technology	1982	No
2	CAIS	Communications of the Association for Information Systems	1999	No
3	DS	Decision Sciences	1970	No
4	ECR	Electronic Commerce Research	2001	Yes
5	EJIS	European Journal of Information Systems	1991	No
6	HBR	Harvard Business Review	1994	No
7	IM	Information and Management	1977	No
8	ISJ	Information Systems Journal	1991	No
9	ISM	Information Systems Management	1984	No
10	ISR	Information Systems Research	1990	No
11	IJEC	International Journal of Electronic Commerce	1996	Yes
12	IR	Internet Research	1991	No
13	JAIS	Journal of the Association for Information Systems	2000	No
14	JECR	Journal of Electronic Commerce Research	2000	Yes
15	JMIS	Journal of Management Information Systems	1984	No
16	JOCEC	Journal of Organizational Computing and Electronic Commerce	1991	Yes
17	MS	Management Science	1955	No
18	MISQ	MIS Quarterly	1977	No
19	QJEC	Quarterly Journal of Electronic Commerce	2000 (ceased publication in 2002)	Yes
20	SMR	Sloan Management Review	1960	No

Table 2.1: List of Academic Journals for Literature Review

- **Electronic databases**:
  - To ensure the rigour and completeness of our literature review, searches for relevant literature were also made using electronic databases that are well known and widely used in Computer Science and Information Sys-

tems research, such as: *ABI/INFORM Global (Proquest)* (1), *ACM Digital Library*(2), *EBSCO Business Source Premier* (3), *Swetwise* (4) and Scopus (5). These databases were the source of additional literature to those published in the 20 journals.

- **Mimas Zetoc Alert Service:**
  - To ensure keeping abreast of recent developments in the area throughout the project, *Zetoc* email alert service was set up at the beginning of the project. *Zetoc* is a service that provides access to British Library's Electronic Table of Contents and gives access to *"over 29,000 journals and more than 52 million article citations and conference papers"* (Zetoc n.d). It allows researchers to specify a set of search criteria (e.g. search keywords or journal titles) and alerts them via email about new publications that match the specified criteria.

    For this project, the search criteria used for setting up the alert service were the same as the ones used for performing searches on the selected journals and electronic databases (see § 2.2.1.2). Additionally, *Zetoc Alert* was configured to send an email notification whenever a new issue of the chosen 20 journals was published (see Table 2.1).

- **Papers, cited in the retrieved articles:**
  - The papers retrieved using the previous three sources were themselves a source of finding additional papers. Using the technique of *"snowball"* (or *"chain"*) sampling (Miles & Huberman 1994) that is common in research, citations used in these papers were evaluated for their relevance

(1) http://www.proquest.com/products-services/abi_inform_global.html
(2) http://dl.acm.org
(3) http://www.ebscohost.com/academic/business-source-premier
(4) http://www.swetswise.com
(5) http://www.scopus.com/

and suitability to this research, applying the same criteria for the literature selection as used for those papers that were discovered using other sources (see § 2.2.1.3). The papers found this way were added to the pool of papers to be analysed and synthesised.

2.2.1.2 Performing Searches

To select relevant literature, the following query terms were used to locate papers containing these terms in their titles and abstracts: *"personalisation/ personalization"*, *"user profile/ user profiling"*, *"ontology/ ontologies/ ontological/ ontology-based"* and *"consumer/ customer behaviour/ behavior"* (1) .

To limit the number of irrelevant papers being retrieved, the *"electronic commerce"*, *"e-commerce"* and *"online"* query terms were also added when performing searches on *'non e-commerce'* journals, i.e. the journals that did not specialise in electronic commerce research (see Table 2.1).

As well as the *'automated'* search using search engines, a manual check of all issues of each chosen journal was also performed to ensure that there were no omissions of relevant papers, which were missed by the search engines. This check did not locate any major papers, missed by the search engine.

The *Zetoc Alert* service was also set up using the same search keywords.

Locating related literature was an ongoing process - from the beginning of the project right up to the submission of this thesis.

(1) The North American spelling of certain words was searched for explicitly, since not all search engines are capable of treating the different spelling variations as synonyms.

2.2.1.3 Criteria for Selection

The selection of papers was based on the criteria that they covered research on web personalisation and use of ontologies in electronic commerce. The choice of this criteria was determined by the project's research aim and objectives (see § 1.2). Thus, the primary focus of the literature review was on the papers that dealt with e-commerce. *'Non-ecommerce'* papers were considered, but were only selected for our literature review, if it was evident that a particular non-ecommerce paper had a seminal contribution to the area of personalisation in general, and had influence (direct or potential) on developments in e-commerce personalisation. For example, a paper was selected if it had a substantial research impact, such as a high citation count, or if it discussed a methodology or technique that was also relevant to the e-commerce domain. Nevertheless, in some research areas, and in particular, with regard to literature on ontology-based personalisation, inclusion of non-ecommerce works was essential due to a lack of research in this area, and the necessity to identify and discuss related approaches and techniques and how this area was related to wider research.

2.2.1.4 Literature Analysis

To accomplish a comprehensive and systematic analysis of the retrieved literature, each chosen paper was coded for the following features: *common theme/ topic*; *independent* and *dependent variables*; *research strategy*, *methodology* and *implementation*; *respondents' sample size* and *sample source*; *type of e-commerce business* and *product type/ category* used in the empirical investigation or testing; and a brief description of each *study's findings*. Such coding is widely used and found to be fruitful for producing a comprehensive and systematic literature review (Cheung, et al. 2003), (Darley, et al. 2010). This coding enabled us to synthesise and classify academic research on web personalisation, and the outcomes are presented in § 2.2.2.

2.2.2 Literature Review Findings

2.2.2.1 Top-Level Classification and Characteristics of Web Personalisation Research

A total of 255 papers were selected as relevant for inclusion in the literature review. Initially, these papers were classified into three broad level topics:

- ***theoretical and empirical research***: research that belongs to this category examines general, strategic and tactical aspects of personalisation, consumer attitudes and behaviour, and justification for and effect of web personalisation on business and their customers;
- ***experimental research***, which designs and develops applications (systems);
- ***ontologies***: research that presents design and development of various ontologies [(1)].

Fig. 2.1 shows subcategories associated with each of the topics and the number of papers [(2)].

(1) Strictly speaking, this research category is a subcategory of the *"Experimental research"*, but due to the project's aim and objectives and the need to discuss research on ontologies separately, the category is presented as a separate entity.
(2) Some papers were classified into more than one category; literature reviews were excluded from the count.

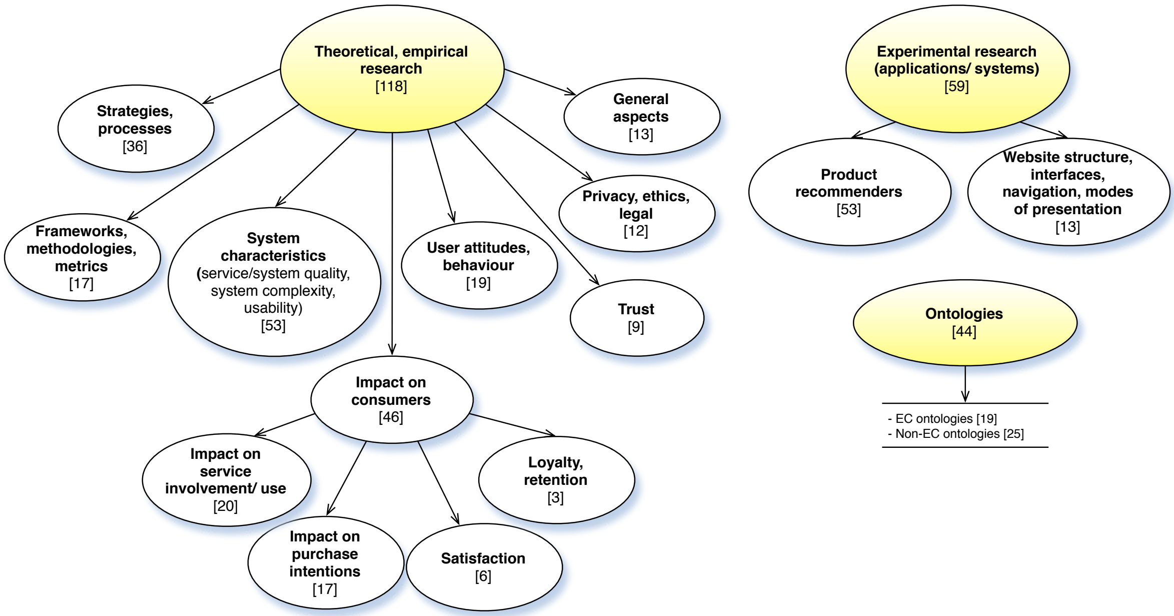


Figure 2.1: Topics in Literature on Web Personalisation

"Theoretical, empirical research" is the category with the largest number of papers. They focus on building and testing theories that cover a variety of web personalisation aspects and topics. For example, the papers that fall within this category propose and investigate the reasons why web personalisation should (or should not) be used; or examine the effect of variables such as user characteristics and behaviour, web design, web quality and usability principles on e-commerce adoption, use and reuse. A substantial amount of research in this area deals with evaluation of the impact of separate, distinct variables, e.g. the effect of usability and web design on the user satisfaction (Zviran, et al. 2006) or the effect of the product presentation format on online consumers' intention to return to a website (Jiang & Benbasat 2007). There is, however, a lack of research that investigates the impact of web personalisation on e-shoppers' behavioural intentions and traits in a more holistic manner. And yet, richer insights and understanding of this relationship is a key to developing effective personalised websites.

Within this category, one particular area that is under-researched is investigation of the design and deployment of appropriate and effective metrics to measure the impact of personalisation through better integration of users' feedback into the personalisation process (Adomavicius & Tuzhilin 2005). Further research in this area would help improve the quality of personalisation.

There is also a considerable amount of research on general and specific issues and principles that deal with development and impact of strategy, processes, frameworks/ methodologies, privacy, ethics, trust, security and legal aspects. These aspects are important as e-commerce companies need to be aware of and understand the effect they can have on business and consumers. For instance, perception of privacy, security and trust has an impact on the consumer's willingness to adopt and use personalised services, e.g. a favourable perception of information privacy on a website has a positive effect on the online consumer's willingness to disclose their personal information and their willingness to buy from the website (Faja & Trimi 2006). Also, online shoppers are affected by privacy concerns, but are willing to supply their personal details if they explicitly see the benefits from personalisation (Awad & Krishnan 2006) and believe that their web experiences are improved. However, most of the research in this area describes aspects of and requirements for privacy, ethics, security, trust and law in general terms, and there is a lack of detail on more particular issues, e.g. on standardisation of these aspects across different systems (Kobsa 2001).

Overall, there is a wealth of material that covers various distinct aspects and factors of e-commerce adoption and success. However, all these aspects and factors are distinct and, more often than not, are covered in isolation from each other. The end result is that we do not have a *'big'* picture, and, more importantly, we do not know what weights the customers put on these factors. Additionally, we do not know the impact of a particular research project setting or its participants. For example, Rianto, et al. (2015) examined the criteria for personalised e-commerce, but this research took place in Indonesia, and it is not known whether the same criteria apply to the UK. Overall, the material found in the “Theoretical, empirical” research category prompted us to look at the issue of the adoption of e-commerce from a more holistic and *non-biased* perspective, leading to broader research questions (§ 5.2.1).

“Applications/ Systems” category studies the design, development, tools and techniques for deploying web personalisation, particularly product recommenders and website structure, interfaces, navigation and modes of presentation.

“Ontologies” category presents research on design and development of ontologies for e-commerce and other domains (e.g. E-learning, Health, Information Retrieval).

While, the first category (*“Theoretical, empirical research”*) is useful for informing this project, it is the last two categories (*“Applications/ Systems”* and *“Ontologies”*) that are of particular interest in our literature review, and are the focus for the remainder of this section.

2.2.2.2 More Focused Classification of Research on Web Personalisation

Fig. 2.1 presented a general overview of the retrieved literature. However, a more particular classification was needed in order to provide a more focused structure of our literature review - to fit our aim and to explain how this project relates to existing work. Several classifications/ taxonomies of web personalisation research have been proposed in previous literature reviews, but these focused on different aspects of web personalisation. Table 2.2 summarises the aspects covered by the previous literature reviews.

Covered Web Personalisation Aspects	**Sources** (in alphabetic order)
Methods and techniques to achieve personalisation (e.g. recommendation systems/ agents, AI techniques, data mining, user modelling, system adaptation, etc.)	(Gao et al. 2010), (Jeevan & Padhi 2006), (Kobsa, et al. 2001), (Mobasher 2007), (Montaner, et al. 2003), (Prasad 2003), (Rojbi & Soui 2011), (Wei, et al. 2007)
Processes of web personalisation (steps, workflow needed to deploy personalisation)	(Malik & Fyfe 2012), (Pierrakos, et al. 2003), (Schubert & Koch 2002), (Weinmann, et al. 2013)
Combined view of personalisation - as a strategic component and technological entity	(Fan & Poole 2006), (Goy et al. 2007), (Koutsabasis et al. 2008)
Business/ marketing strategy	(Montgomery & Smith 2009), (Murthi & Sarkar 2003)

Table 2.2: Aspects Coverage in Extant Reviews of Web Personalisation Research

Table 2.2 shows the diversity of aspects that are used for web personalisation research classification. The diversity is explained by the fact that research in the area spans various domains and disciplines, such as Business, Information Systems and Computer Science, and therefore the classification depends on the domain and aims of the researchers. The diversity is also reflected in the number of aspects covered in the literature reviews and in the lack of a single, *standard* classification of web personalisation research. The majority of the proposed classifications focus on the methods and techniques used to personalise web-based systems (such as the technological/ computational aspects - the *"hows"* of personalisation), while others take into consideration the *"whats"* of personalisation such as its positioning and role in business strategic and marketing activities. In addition, some covered much broader areas than others. For example, Gao et al. (2010) offered a comprehensive review of approaches to personalisation, while Wei et al. (2007) concentrated on surveying the developments of e-commerce recommender systems.

None of these reviews offered a classification that provided a perfect fit with our project's research objectives and our particular focus on e-commerce and its *"technological"* perspective (§§ 1.2 and 1.5), so an alternative classification of web personalisation research was developed, reflecting our interests.

Our classification was created by combining the content analysis and synthesis of the retrieved literature with the findings of existing literature reviews, namely, in (Brusilovsky 2001) (1), (Kobsa et al. 2001) and (Pierrakos et al. 2003). Our main focus was on the areas of the project's domain of interest, i.e. creation of an application/ system that personalises e-commerce websites. The resulting classification is shown in Fig. 2.2 (2) (3).

---

(1) This paper does not cover web personalisation explicitly. Instead, it covers the related "Adaptive Hypermedia" research area. Therefore, it was excluded from Table 2.2, but was used for our classification due to its relevance.

(2) The number of papers belonging to each category is shown in brackets. Some papers in the *"Content Presentation Personalisation"* category were classified into multiple categories.

(3) The count of the *"Content Presentation Personalisation"* papers in Fig. 2.2 is based on the *"Experimental research (applications/ systems)"* ('non-ontology') papers, shown in Fig. 2.1, plus six additional *"Ontologies"* papers, which were determined as relevant for inclusion. Hence, the count of the papers in *"Content Personalisation"* rose to 54 (53 of the 'non-ontology' "Content Personalisation" papers + 1 "Ontologies" paper), and in "Personalised Content Presentation" to 18 (13 of the 'non-ontology' + 5 "Ontologies").

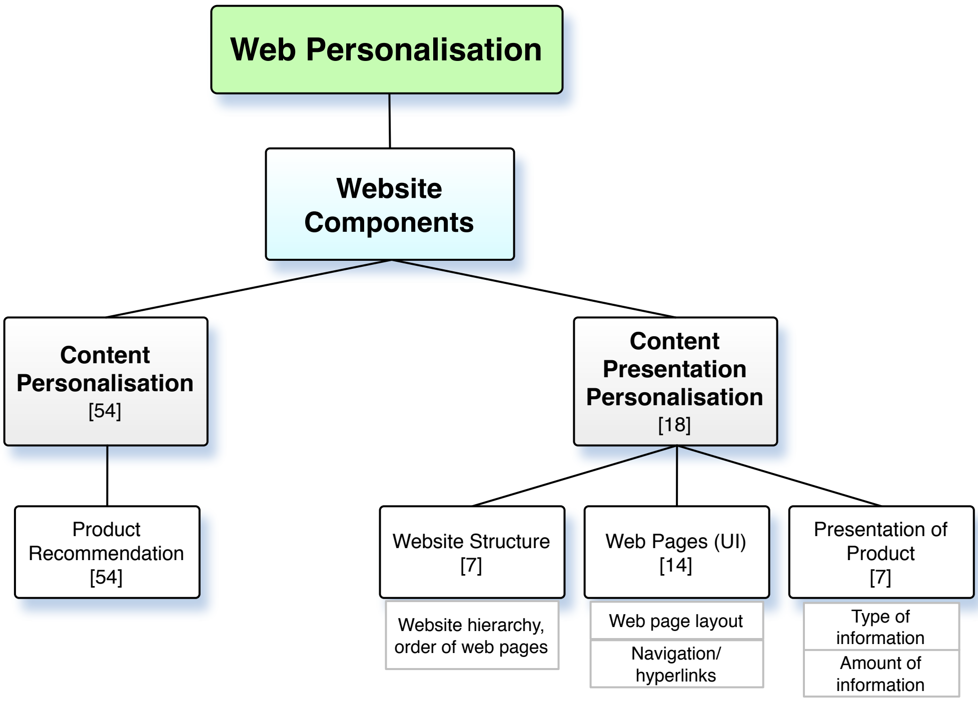


Figure 2.2: Classification of Web Personalisation

Our classification identifies two types of website components that can be personalised: ***content*** and ***content presentation***

Personalisation of content refers to the process of recommending to users a particular mix of products or services. This process is typically based on some predefined or computed criteria, e.g. by discovering patterns in the user's preferences, past shopping history or other characteristics specified in the user's profile. Research in this area examines the issues related to the design and deployment of recommendation systems (agents). Examples of content personalisation include: (Felden & Linden 2007), (Kim, et al. 2002), (Niu, et al. 2002) and (Srikumar & Bhasker 2004).

Personalisation of content presentation refers to the ability of a website to dynamically adapt the structure, web pages and the way the content is presented to the user; and is related to the *Adaptive Web/ Adaptive Hypermedia* research area (Brusilovsky 2001). There are three subcategories in this area:

- **Personalisation of website structure**. This is the adaptation of the whole or parts of a website, such as a hierarchy and order of the web pages in a system. Applications and systems that implement adaptive website structures are presented in (Baraglia & Silvestri 2007), (Robal & Kalja 2007), (Yen & Kong 2002).

- **Personalisation of web pages (UI)**. This deals with adaptation of the web page layout and navigation.

  - Typically, personalisation of the web page layout is the dynamic restructuring of how and what content is displayed on a web page, based on the user's preferences that are specified explicitly by the user or inferred by the system (e.g. computed by a software agent). It is possible to personalise the layout to display the products in a particular location of a web page (Koutsabasis & Darzentas 2008); or to personalise the type of information that is shown to the user (Ardissono, et al. 1999). Other options include filtering of content, e.g. blocking particular web page segments (Kuppusamy & Aghila 2012). A combination of various personalisation strategies is also possible, e.g. personalisation of the web page *'skins'* (themes), layout, font sizes and position of the content boxes (Kwon & Kim 2012).

  - It is also possible to adapt how the navigation is presented to a user based on the user's browsing behaviour, explicit preferences, needs or requirements. This involves predicting web pages and dynamic generation of the hyperlinks to these pages. Baraglia & Silvestri (2007), Flesca, et al. (2005), Gupta & Mathur (2002), Perkowitz & Etzioni (2000) demonstrate the use of personalised navigation.

- **Personalisation of product presentation**. This is the process of how to present the product information to a particular user or group of users. Typically, this type of content presentation offers personalisation of:

- Types of information, such as textual, graphical, multimedia;
- Amounts of information, e.g. varied amounts of text and product image types or sizes (bigger/ smaller product photos, more/ less text for product description). Examples of this personalisation are described in (Ardissono et al. 1999) and (Kuo & Chen 2001). The former displays the amount of information on a product, based on the classification of the users into *"technology experts"* and *"non-experts"*; the latter offers two types of personalisation of the amount of information - *"large"* and *"small"*.

This discussion and the examples present the personalisation of content and its presentation as two separate entities. However, it is important to note that while it works for a classification model, in practice, it is hard to separate them due to a strong relationship and interdependence between them. Often (if not always), e-commerce systems are designed in a holistic manner where the type of product, offered for sale, dictates how it should be displayed on a website. Also, intuitively, personalisation of how a product is presented warrants the change in layout; or the change in a website structure needs to be reflected in the navigation. This phenomenon is reflected in research practice, where projects used several types of content presentation personalisation, e.g. Baraglia & Silvestri (2007), Perkowitz & Etzioni (2000) and Yen & Kong (2002) combine adaptive web structure with personalisation of navigation; while Ardissono et al. (1999) and Kuo & Chen (2001) offer personalisation of web pages and product presentation. Other works describe systems that use a combination of content and its presentation, e.g. in (Koutsabasis & Darzentas 2008), (Schubert 2000) and (Yen & Kong 2002).

Overall, the amount of research that examines the design and development of web personalisation is sizeable (59 papers), and presents strong evidence that personalisation of both, content and content presentation, is useful and effective, and in particular for *different levels of a customer's familiarity with products* (Chau, et al. 2000), and *different stages of product acquisition (purchasing)* (Konradt, et al. 2003), (Detlor, et al. 2003). Additionally, content and content presentation personalisation offers other benefits, such as: *reflecting differences in users' web skills* (Hampton-Sosa & Koufaris 2005); *matching their level of "Technology Readiness"* (1) (Massey et al. 2007); *reducing system usage complexit*y (Guo & Poole 2009); and *enhancing and strengthening the higher level of website users' "arousal and pleasure"* (Cheng, et al. 2009). All in all, such personalisation can lead to an *improved usability of the website system*, which some researchers and practitioners see as an important pre-requisite for a commercial website's success (Goy et al. 2007), (Zviran et al. 2006).

Nevertheless, the comparison of the amount of research on content and content presentation reveals that there is significantly more research on content personalisation and less research on personalised content presentation, and in particular, on personalisation of product presentation. As shown in Fig. 2.1, out of the 59 *"Experimental research (applications/ systems)"* papers, 53 dealt with *content* personalisation and 13 with *content presentation* personalisation, while 7 papers covered both categories. In the *"Ontologies"* category, there were additional 6 papers: 1 paper (2) covered product recommendations, and 5 content presentation.

These figures suggest that there is a bias towards the research on the product recommenders. Within the *"Content Presentation Personalisation"* category, there is also a bias towards personalisation of web pages: the *"Website Pages (UI)"* category is covered by 14 papers, compared with 7 papers in each, *"Website Structure"* and *"Presentation of Product"* categories. In addition, these figures indicate that there is also a lack of research on integration of the two types of web personalisation (Yen & Kong 2002). Therefore, there is an opportunity for more research to be done on personalisation of content presentation and its integration with personalisation of

(1) Technology Readiness is defined as *"the propensity of customers to embrace and use new technologies for accomplishing goals"* (Massey, et al. 2007).

(2) (Hella & Krogstie 2011)

content.

2.2.2.3 Semantic Web and Ontology-Based Web Personalisation

In Computer Science, Information Science and Informatics, an ontology is typically defined as an explicit formal representation of concepts (or *classes* [(1)]) within a particular domain of interest, as well as relationships between these concepts (Gruber 1993). Ontologies are used to formalise and systematise the representation of a domain of interest (Mizoguchi 2003), (Noy & McGuinness 2001) by conceptualising knowledge about this domain, and expressing such knowledge in a machine-readable (Noy & McGuinness 2001) as well human-readable format.

Ontologies are widely used for user profiling to achieve web personalisation. Their use for this purpose was of particular interest and relevance to our project.

The development of Semantic Web technologies is a growing area of research and practice, and there is a clear link between the use of ontologies and ontological profiling and further enhancement of the Semantic Web.

The philosophy behind the *'original'* Web 1.0 was primarily about *'human reading'*, and the philosophy behind Web 2.0 was about *'human reading, writing, communicating and collaborating'*; while, it can be said that, in addition to the *'human'* communication, the philosophy of the Semantic Web (or Web 3.0) is about *'machine reading, writing, communicating and collaborating'*. The main vision (and promise) of the Semantic Web is to turn the current Web into *"the Web of linked data"* (W3C n.d.), so that data would become available for access, sharing and analysis, and computers could search, combine and process the data intelligently (Hitzler, et al. 2010, p. 11).

(1) Often, the term *"class"* is used instead of *"concept"* when describing an ontology, and in particular, while discussing OWL-based ontologies. From now on, we will follow this tradition, and predominantly use the term *"class"*. This, however, should be treated as a synonym to *"concept"*.

The ontologies have become of increasing importance for supporting systems, compliant with Semantic Web, have become an integral part of Semantic Web (W3C n.d.), and are now considered to be one of the core elements of the Semantic Web stack and the Semantic Web's *"key enabling technology"* (Davies, et al. 2002). Ontologies are used to provide *"the backbone of the Semantic Web, defining the semantics of the data and Web resources"* (Dameron, et al. 2004) and to enable reasoning capability and knowledge discovery (Koutsomitropoulos, et al. 2006). They are considered to be *"the best answer to the demand for intelligent systems that operate closer to the human conceptual level"* (Obrst, et al. 2003).

2.2.2.3.1 Why Ontology?

Along with the advent of the Semantic Web, we are witnessing the ever increasing abundance of data resources, which need to be stored, processed and accessed. However, there is the lack of a single, unified standard for their modelling, structure and retrieval support, which has led to problems of interoperability and flexibility of different standards in encoding digital data objects (Fensel, et al. 2001), (Obrst et al. 2003). To address these issues, an ontology-based approach has been widely investigated as one of the possible solutions to improve access to and management of such data (Chen, et al. 2010), (Gocić 2008), (Rosaci & Sarnè 2014).

Ontologies are considered to be an effective way to model, manage and make better use of Internet-based digital content. In comparison with other methods, such as machine learning and data mining techniques, ontology-based systems offer a deeper domain-based knowledge (Kearney, et al. 2005), more precise, correct and explicit domain conceptualisation and representation (Mizoguchi 2003), (Noy & McGuinness 2001); and a richer, less ambiguous semantic representation of user models (Cantador, et al. 2008b). Rich ontologies can capture wider types of relationships (Gauch et al. 2007). Ontology-based knowledge and domain representation provides *“a solid foundation on which we can build sharable knowledge bases for wider usability than that of a conventional knowledge base”* (Mizoguchi 2003). Similarly, ontological user models can also be easily reused and shared between different parts of the system or different systems (Cantador et al. 2008b). This facilitates the flexibility, reusability and interoperability of the system design and implementation.

Another advantage of using on ontology is use of the Open World Assumption (OWA) assumption, which treats the missing information as unknown, compared with the Closed World Assumption (CWA), which treats the missing information as false (non-existent), and which is more common in machine learning and data mining. The OWA paradigm is much better suited to web-based systems, where new data is added every day and is highly distributed, so we cannot guaranteed the data completeness or prove its absence - if we haven’t found it, it doesn’t mean the data does not exist (Allemang & Hendler 2011).

Ontology-based systems can also take advantages of the powerful inference mechanisms to facilitate data analysis. Ontologies use axioms (rules), enabling the extraction of not only explicit but also implicit knowledge, i.e. acquisition of tacit information from the existing data. Moreover, ontologies have been found to perform better in knowledge representation and user profiling when they are compared with other methods used. For example, Adda et al. (2006) reported enhanced performance of their video personalisation system when an ontology was used to describe the semantic content of video data (as opposed to using a suite of standards to describe the multimedia content, MPEG-7 [(1)]). While Middleton, et al. (2004) found that using ontological inference to profile users, yields a superior performance when compared with employing a flat list of unstructured topics.

(1) `http://mpeg.chiariglione.org/standards/mpeg-7`

2.2.2.3.2 Ontologies on the Semantic Web

The success of the Semantic Web is argued to depend on the proliferation of ontologies (Antoniou, et al. 2012). Over the last few years, an increased number and variety of ontologies, compatible with the Semantic Web, have been developed. These include both, common (generalised) ontologies, and more specific ones that target a particular domain. For example, a considerable number of ontology-based *Information Retrieval (IR)* systems have been proposed, e.g. (Cantador et al. 2008b), (Daoud, et al. 2008), (Jiang & Tan 2009), (Sieg et al. 2007) and (Stamou, et al. 2009). In Cardiff, the SPIRIT (*"Spatially-Aware Information Retrieval on the Internet"*) project developed a geographical ontology to enable geographically-aware user query interpretation and disambiguation, metadata extraction and spatial indexing (Jones et al. 2002). *E-learning* is another research area, where the use of an ontology-based system has been investigated (Fok 2005), (Gomes, et al. 2008), (Gaeta, et al. 2009). Other domains include: *Health* (e.g. (Abidi & Abidi 2013)); *Adaptive Web* (Robal & Kalja 2007); user profiling in the context of *Big Data* (Hoppe, et al. 2013); and *Web Services*, such as for instance, *OWL-S* (1) and *Web Services Modeling Ontology (WSMO)* (2); as well as systems to bring ontologies and Web Services together (Dameron et al. 2004).

The availability of these ontologies progresses the development of the Semantic Web by making it more intelligent, accessible and 'personalisable' - by providing an opportunity for creating new, novel systems that reuse the existing ontologies. For example, ontologies, which model Web Services, e-learning and geographical entities, can be combined to create a geographically-aware system for discovery and exchange of e-learning products or services offered by different vendors. The reuse of existing ontologies means that the development time, effort and costs can be substantially reduced.

(1) http://www.w3.org/Submission/OWL-S/
(2) http://www.w3.org/Submission/WSMO/

2.2.2.3.3 Ontological Personalisation in E-Commerce

E-commerce is identified as an area that could particularly benefit from the Semantic Web (Fensel, et al. 2002), and there is now a growing body of research that has proposed the use of ontologies specifically in e-commerce.

The use of ontologies emerged as *"a natural choice to represent knowledge about users in E-commerce context"* due to its *"powerful knowledge representation and inference mechanism"* (Liu, et al. 2008). Ontology-based user profiling has been found useful and beneficial due to its potential to improve profile accuracy, its ability to allow the users to provide their feedback, and its ability to merge with other types of ontologies (Middleton et al. 2004). It has been suggested that *"the usage of ontologies enhances the usability of user profiles and product profiles and improves recommendation results and customer satisfaction"* (Felden & Linden 2007). It can also reduce the work of web developers *"dramatically"* (Goy et al. 2007).

In line with research in other domains, research that focuses on ontological e-commerce personalisation can be classified as belonging to the categories of *"content personalisation"* and *"content presentation personalisation"* in our web personalisation classification (see § 2.2.2). Thus, ontologies are used for recommending various products, for example, diving equipment in (Felden & Linden 2007), food in (Hella & Krogstie 2011), and computer hardware in (Liu et al. 2008). Ontology-driven personalisation is also used to adapt (modify) web page layout and navigation in (Flesca et al. 2005), (Gaulke & Ziegler 2015), (Lei et al. 2003), (Robal & Kalja 2007), (Valencia, et al. 2013); and for product presentation personalisation (type and amount of information) in (Gaulke & Ziegler 2015) and (Lei et al. 2003).

Another area of interest to this project is research on the design and implementation of ontology-based electronic catalogues of products. Intuitively, product e-catalogues are an essential e-commerce system component, and ontology-based product catalogues can be used to enable product recommendations and other e-commerce personalisation.

Some of the work in this area created generalised ontologies such as the *eClassOWL* ontology, which models different types and services (Hepp 2005); and the *GoodRelations* ontology, which is used for describing product and service offerings (e.g. during a promotional campaign) (Hepp 2008). Other work describes more specific ontologies, targeting a particular system, situation, issue or domain. For instance, in (Nowakowski & Stuckenschmidt 2010) the *eClassOWL* ontology is used as the base for a system that matches a customer's product requirements with corresponding products found in the ontologised electronic product catalogue. A similar system to match a customer's preferences with products/ services offerings is proposed by Chen et al. (2010), where the matching is done using two ontologies, albeit both are bespoke creations: a user personalised catalogue ontology (containing user preferences) and a domain e-catalogue ontology (containing product categories and descriptions).

Other research on e-catalogues presents a way to ontologise a B2B company's product e-catalogue from its XML-based representation (Gocić 2008). Similarly, Rosaci & Sarnè (2014) propose a B2C personalisation system, which uses an XML-Schema based ontology to represent products, their categories and associated relationships. Whereas, Beneventano & Montanari (2008), in addition to a proposal on ontologising a product and services catalogue through creation of *"reference ontologies"* (i.e. ontology-based annotation of an e-catalogue), also offer a solution to alignment of the newly created ontologies and subsequent mapping of these to the existing product catalogues. Lee, et al. (2006) describe ontologising a product catalogue for a government procurement agency, whereas Liu et al. (2008) present a manually created ontology that represents a computer hardware components catalogue. Electronic products catalogues then can be used in combination with user profiling ontologies to personalise e-catalogues to an individual or a group of users (Chen et al. 2010).

Ontologies and ontological user profiling have also been used for personalisation of B2B seller-buyer matchmaking, negotiation and contract formation and fulfilment (Gonzalez Castillo, et al. 2001), (Trastour, et al. 2003); and for better understanding of a user's navigational behaviour so that the accuracy and relevance of product recommendations and personalisation of web navigation can be enhanced (Kearney et al. 2005).

Overall, the analysis of the literature shows that despite some effort, the area of ontology-based e-commerce web personalisation still needs more investigation, particularly with regard to the use of reusable and extendible ontological user profiling. Most current work involves development of ontologies, which are created for the purposes of a particular project, only applicable to the context of that project, and are not suitable or reusable in other contexts or settings. This is somewhat surprising considering that ontologies and ontological reasoning services are capable of providing *interoperability* and *flexibility* between different web technologies and components. In particular, *semantic interoperability* in the context of the Semantic Web is regarded as an important reason for choosing ontology-based development solutions (Chen et al. 2010), (Mizoguchi 2003). Such interoperability can facilitate standardisation, discovery, management, sharing, integration and reuse of Semantic Web data (Horrocks & Patel-Schneider 2011). *Flexibility* is a closely related property, which gives an opportunity to design and implement solutions that are easily reusable, extendible and scalable. For example, an existing system can be easily scaled up or extended by merging several ontologies. Overall, use of Semantic Web technologies is seen as a promising solution to the limitations of the current Web technologies, and in particular in e-commerce (Fensel et al. 2002). For these reasons, ontology-based profiling might prove to be a more efficient and effective solution for e-commerce web developers.

2.3 Web Personalisation in Practice

The previous section discussed the development of web personalisation in academic literature. The author's personal communications with the industry representatives ([Pers.Comm.1], [Pers.Comm.2], [Pers.Comm.3]) and personal experience of using e-commerce indicated that there is a limited use of research in the commercial world. To check if this is the case, a survey of whether and how practitioners (i.e. e-commerce businesses) use personalisation was undertaken. This section summarises our analysis of the usage of web personalisation in the industry.

2.3.1 Methodology for Surveying Current Professional Practice

We evaluated ten e-commerce websites, which were ranked by `alexa.com` as *"Top 10"* websites in the "Shopping/Clothes" category (1). This ranking was based on a combined measure of an estimated number of page views over a period of three months and an estimated number of daily unique visitors (Alexa Internet n.d). All these companies are well-known, large e-tailers, which belong to the Business-to-Consumer (B2C) e-commerce segment. (The list of chosen companies is in Table 2.3).

The choice of B2C e-tailer was influenced by the fact that this is one of the most prominent type of online business model. IMRG, the UK's Industry Association for Online Retail reported that in November 2014 in the UK out of 50 top online retailers 39 websites (2) (i.e. 78%) belonged to the B2C e-tailer category (IMRG 2015). The

(1) `http://www.alexa.com/topsites/category/Top/Shopping/Clothing`; ranking of July 2014.
(2) This figure includes e-commerce companies that sell "physical" goods (e.g. clothes, books, etc.),

"Shopping/Clothes" category was also chosen due to its prominence: it is consistently one of the top B2C categories - if not *'the'* top one, as stated in (UK Office for National Statistics (ONS) 2010) for the UK and in (Nielsen 2014) for some countries and continents.

Our analysis of the current use of web personalisation functionality and features by the chosen e-commerce companies was based on our web personalisation research classification (see § 2.2.2), and examined the presence of:

- **Content Personalisation**:
  - System-driven content personalisation, i.e. presence of recommendations, such as links to related products based on various filtering techniques or algorithms (e.g. *"We also recommend these items..."* or *"Other customers also bought these items..."*), or based on some business rules (e.g. cross-selling other products to form a complete outfit).
  - User-driven content personalisation, i.e. a facility for a customer to create their own, bespoke "wishlists" - a list of products that the customer is not prepared to buy there and then, but would like to buy in the future or just interested in seeing.
- **Content Presentation Personalisation**:
  - A facility to allow customers to choose how many products they want to see on a web page, e.g. a certain number of products or all products at once.
  - A facility to allow customers to choose different types of product information and different amounts of this information.

but excludes companies that offer services (e.g. rail, cinema and air tickets, travel services or software).

- A facility for sorting products on a web page, e.g. sorting by price, size, colour, material/ fabric.
- A facility for filtering products by types and categories.
- A facility to personalise the website's interface or layout, e.g. specification of the positioning of website elements and dynamic navigation.

Each website was accessed in July 2014, in the order specified in Table 2.3. A customer account was created for each company to enable evaluation of the presence of personalisation functionality and features, offered by these companies to their registered customers. The evaluation started after logging in as an existing customer, and involved visual inspection of the product categories, single product and the customer account web pages. The presence of the personalisation functionality and features, specified above, was then noted in a table (see Table 2.3).

In addition, all websites were re-checked on 7 April 2015 for the presence of any major changes in personalisation features and functionality. This was not the case: the examined websites did not have such changes.

2.3.2 Findings from Surveying Current Professional Practice

Table 2.3 summarises the findings from our survey. (The green tick means that a feature or functionality is present, the red cross means it is absent.)

Table 2.3: Personalisation Features Offered by Top 10 E-Commerce Clothing E-Tailors

	Website (country)	**Content Personalisation**		**Content Presentation Personalisation**					
		System-driven recommenda-tions	**User-driven personalisa-tion** (e.g. wishlist)	**Products display per webpage**	**Product information choice**	**Products sorting filters**	**Product type or category filtering**	**Personalisa-tion of layout or navigation**	**Other/ comments**
1	zappos.com (USA)	✓	✓	✗ products are displayed over multiple pages (determined by the system) - no choice of selecting how many items to display per page	✗ initial amount of information on an individual product is limited to 4 lines, to display the complete product information the user needs to click on "Show more information" button	✓ relevance, customer rating, bestsellers, price, brand name	✓ size, colour, occasion, brand, materials, theme, etc.	✗	
2	gap.com (USA)	✓	✓	✗ all products displayed on one page - no choice of choosing to display products over multiple pages	✗ no option to choose from	✗	✓ *Index page:* "Deals", "Sale", "New and Now", "Categories", "GapBody", "GapFit" *Product category pages*: style, size	✗	*"Quick look"* functionality
3	hm.com (UK)	✓	✗	✗ all products displayed on one page - no choice of choosing to display products over multiple pages	✗ no option to choose from	✓ colour, size, concept	✓ size, colour, concepts	✗	*"Shop Now"* functionality

(Table 2.3 continued)

	Website (country)	Content Personalisation		Content Presentation Personalisation					
		System-driven recommenda-tions	**User-driven personalisa-tion** (e.g. wishlist)	**Products display per webpage**	**Product information choice**	**Products sorting filters**	**Product type or category filtering**	**Personalisa-tion of layout or navigation**	**Other/ comments**
4	6pm.com (USA)	✓	✗	✗ products are displayed over multiple pages (determined by the system) - no choice of selecting how many items to display per page	✗ no option to choose from	✓ newest, most popular, name, lowest price, highest price, % off	✓ subcategory, size, occasion, brand, colour, top styles, brand, material, pattern, accents	✗	
5	victoriasecret.com (USA)	✓	✓	✗ all products displayed on one page - no choice of choosing to display products over multiple pages	✗ no option to choose from	✓ sort by recommend-ed, price, top rated, new arrivals, sales and specials	✓ size, style, colour	✗	Items can be added to the customer's wishlist from the product category page.
6	forever21.com (UK)	✗	✓	✓ choice of 30, 60 or 100 products per page	✗ no option to choose from	✓ *"what's new"*, high price, low price	✓ size	✗	Items can be added to the customer's wishlist from the product category page.

(Table 2.3 continued)

	Website (country)	**Content Personalisation**		**Content Presentation Personalisation**					
		System-driven recommenda-tions	**User-driven personalisa-tion** (e.g. wishlist)	**Products display per webpage**	**Product information choice**	**Products sorting filters**	**Product type or category filtering**	**Personalisa-tion of layout or navigation**	**Other/ comments**
7	shopbop.com (USA)	✓	✓	✓ choice of viewing 40 or 100 products per page	✓✗ Several sections ("Description", "Size and Fit", "Designer", "Shipping & Returns"), which are collapsible. Product description opens by default. To access other sections the user needs to expand these manually. However, the amount of information cannot be personalised.	✓ view all, price low to high, favourites, editors' picks, designer exclusives, ratings	✓ designers, sizes, colours	✗	Cross selling products (*"Wear it with"* functionality) No option of sorting by *"high to low price"*
8	yoox.com (UK)	✓	✓	✗ products are displayed over multiple pages (determined by the system) - no choice of selecting how many items to display per page	✗ no option to choose from	✓ latest arrivals, high price, low price	✓ Latest news, price range, designers, categories, sizes, colours, materials	✗	Functionality to create a personalised profile by specifying clothes size - each search will then only display products with those characteristics

(Table 2.3 continued)

	Website (country)	Content Personalisation		Content Presentation Personalisation					
		System-driven recommendations	User-driven personalisation (e.g. wishlist)	Products display per webpage	Product information choice	Products sorting filters	Product type or category filtering	Personalisation of layout or navigation	Other/ comments
9	net-a-porter.com (USA)	✓	✓	✗ products are displayed over multiple pages (determined by the system) - no choice of selecting how many items to display per page	✓✗ "Editor's Notes & details" (non-personalisable). "Size & Fit" (additional information on the product), which can be expanded by the user.	✓ new in, price high to low, price low to high	✓ product sub-categories, designer, colour, size	✗	Cross selling products (*"How to wear it"* functionality)
10	modcloth.com (USA)	✓	✓	✓ choice of viewing 50 products per page	✓✗ Minimum product description is displayed as default. The user needs to expand "Details & Measurements" section - otherwise, it is collapsed and hidden on opening a product page.	✓ newest, lowest price, highest price, most loved, top rated, must see	✓ subcategory, colour, size, designer, price, "Fresh Faves"	✗	Items can be added to the customer's wishlist from the product category page.

Overall, the evaluated e-commerce companies offer some personalisation features and functionality - with some companies offering more than others.

With regard to providing their users with website **content personalisation**, the surveyed companies are unanimous. The vast majority offer both, the system-driven functionality, i.e. product recommendations (9 out of 10 companies) and user-driven functionality, i.e. “wishlist” (8 out of 10).

However, when it comes to personalisation of the website **content presentation**, there is more variation between the companies.

All companies offer a facility to select products by types or categories, but the number of filters vary. `zappos.com`, `gap.com` and `6pm.com` provide a comprehensive list of filters, while `victoriasecret.com` and `shopbop.com` offer only three choices, and `forever21.com` only one (*“sort by size”*).

All companies but one (`gap.com`) offer a product filtering facility, such as product sorting by price, colour and size. As with selecting products by types and categories, the number and types of choices that a customer can select vary from company to company.

Three companies (`forever21.com`, `shopbop.com` and `modcloth.com`) allow their customers to choose how many products should be displayed on one web page, but the number of choices and the number of products per page vary.

When it comes to the choice of the amount and type of information, the majority of the surveyed companies do not provide this functionality. Three companies (`shopbop.com`, `net-a-porter.com` and `modcloth.com`) display an incomplete product description with a choice for the user to expand it, e.g. they display a minimum of information by default and offer additional information through a “collapsible” menu, which the user needs to click on to display this information. However, the amount of information cannot be personalised.

`yoox.com` is the only company that provides their customers with a functionality to create a personalised product search profile. A user can create their profile based on sizes, and when the user browses the website, the search results are personalised by displaying only those products that fulfil the user's profile criteria. The user can also save their searches and turn these into an email notification (although the number of saved searches is limited to 10). The other companies do not offer similar personalisation features. However, this functionality only controls the choice of products to be displayed on a web page. No personalisation of the website interface or navigation is offered.

None of the companies offer any functionality or tools to personalise the layout or navigation (adaptive interface). For instance, nobody offers the users an ability to specify and save their preferences for how products are displayed and arranged, e.g. to display products always arranged by the lowest (or highest) price or by other product characteristics the user specified as preferred - so that when the user revisits the website, the web pages always take the user's preferences into consideration. Yet, a functionality like this is intuitive to offer as the users will most likely welcome it, and, arguably, is not very difficult to implement.

Compared with the substantial progress and contribution of research on web personalisation (Gao et al. 2010), the personalisation features and functionality that are used in the commercial e-commerce systems appear to be minimal. This confirms our conjecture that there is a lack of research impact and a mismatch between the research offerings and commercial development.

2.4 *"Hows"* of Web Personalisation: Approaches, Methods and Techniques

In § 2.2.2, we discussed what components and aspects of a website can be personalised (the *"whats"* of web personalisation). Researchers and commercial practitioners have also developed diverse approaches, methods and techniques (i.e. the *"hows"*) to achieve this. The choice of such methods and approaches depends on various factors, such as the purpose of the research, the type of website and content, target demographics, and available technological advances. In professional practice, web developers' experience and preferences also play an important role in deciding how a website personalisation should be accomplished.

This section synthesises and discusses the approaches that are developed by current research to personalise websites. The findings from our analysis will form the basis for the processes, techniques and approaches to be used in our proposed system, *PERSONTO* (see § 3.5.2).

2.4.1 Process of Web Personalisation

The literature, reviewed in § 2.2, covers various stages of web personalisation systems' development. These can be aggregated as a four-phase website personalisation process, consisting of the following distinct phases:

- data collection;
- data analysis;
- user profile generation and maintenance;
- system implementation.

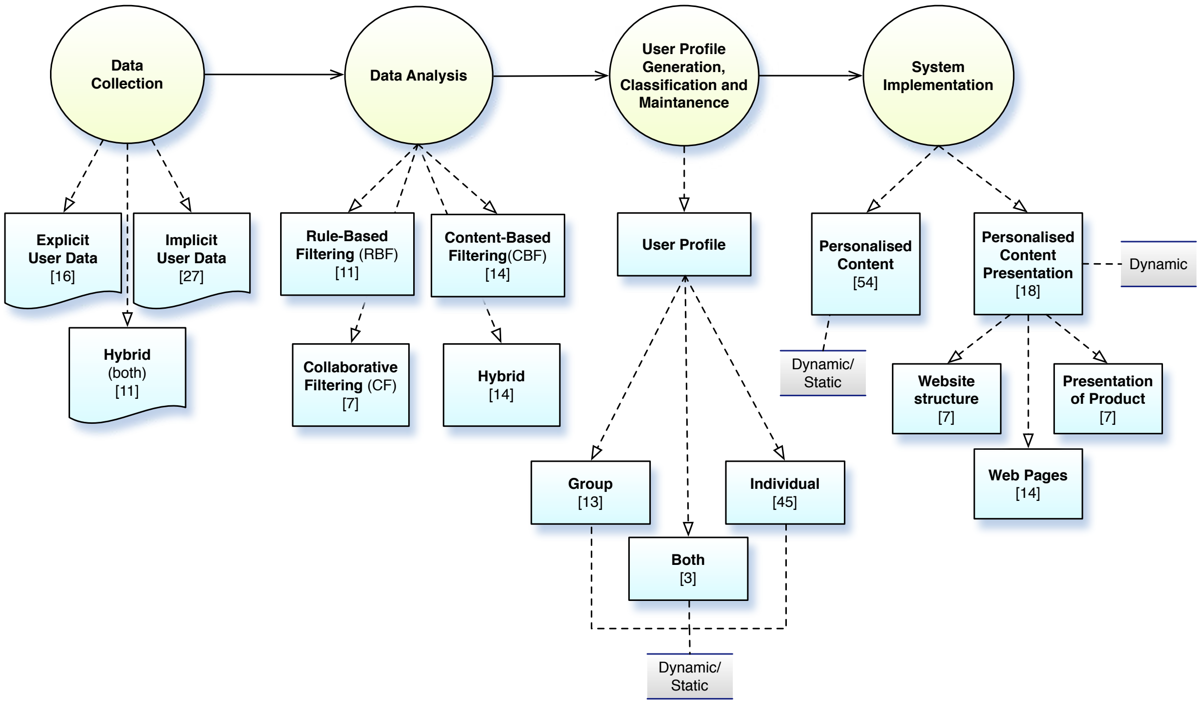


Figure 2.3: Web Personalisation: Process and Approaches

Each of these phases uses a number of strategies, approaches and techniques, as shown in Fig. 2.3. (The count of papers that use each strategy, approach or technique, shown in brackets, is based on the same pool of 59 *"Experimental research"* papers and additional 6 papers (1) from the "Ontologies" category, which were analysed, synthesised and discussed in § 2.2.2.1. Some papers were classified as using more than one strategy, approach or technique. Some papers did not specify which type of data collection, data analysis, user profile generation or system implementation strategies they used, and were excluded from the count.)

2.4.1.1 Data collection

To ensure web personalisation effectiveness and success, a certain type and amount of appropriate data must be collected.

(1) As explained in § 2.2.2.2.

Data collection strategy involves decisions on which user data to collect, and it is important to ensure that the collected data is appropriate and accurate, and in particular, that it provides relevant support for the phases that follow, i.e. the data analysis, user profile generation, classification and maintenance, and ultimate system implementation. Typically, such user data is collected either ***explicitly*** or ***implicitly***:

- ***Explicit*** user data collection is carried out through soliciting for and collecting direct user input manually, e.g. by carrying out user interviews or surveys, and requires the active participation of the user (Mobasher 2007). Other explicit data could have already been provided by customers (e.g. customer demographic information), and this can be harvested and analysed - albeit this might be sufficient only for a certain type of personalisation functionality. Thus, in the majority of cases, data collected in this way should be viewed as *primary data*. User profiling based on explicit user data is also considered to be *"knowledge-based"* (Middleton et al. 2004).

  The advantage of this method is that there is no need to have a sufficient amount of *'upfront'* data, i.e. the system's *"cold-start"* problem (1) can be avoided (Koutsabasis et al. 2008), (Middleton et al. 2004). For instance, recommendation engines and agents need a sizeable amount of data to provide users with appropriate and useful recommendations. This might be problematic for new companies that only have a short trading history or for companies with a small customer base. Another advantage of employing explicit data collection, which has been argued by some researchers (e.g. by Tao, et al. (2011)), is the fact that involving users directly can increase the quality and validity of generated user profiles due to the high level of user involvement. There is also evidence that the users not only do not mind being

(1) *"Cold-start"* problem refers to the situation, where the system cannot generate relevant or useful recommendations or personalisation because it does not have enough data for a user.

asked by the website, but are also *“pleased to cooperate with the website”* and regard their dialogue with the website as something that indicates the website’s *“competent attention to their needs”* (Mussi 2006). Explicit elicitation of the user preferences is indicated as an important factor that influences (increases) a perceived value of and positive attitude towards product recommendations (Gretzel & Fesenmaier 2006).

However, this method also suffers from a number of disadvantages, which need to be taken into consideration when developing web personalisation systems. For instance, gathering explicit data might involve additional effort, financial costs and resources that the system’s developers or owners need to invest and may be reluctant to commit to. Some researchers also argue that explicit data cannot be guaranteed or assumed to be of high quality, for example Kobsa et al. (2001) warned that it is hard to ensure that the user supplies error-free data, while Montaner et al. (2003) expressed their data quality concern due to the users withholding information or providing false data. Additionally, it might be difficult to get users involved at the data gathering stage for a number of reasons, such as: the effort required to provide this data; their privacy and data protection concerns; users simply may not know what they want, or may be unwilling to give away their personal information or reluctant to spend any time on providing such data (Gauch et al. 2007). Nevertheless, the positive stakeholders’ attitudes and high level of acceptance of our proposed system, discovered during our evaluation (see § 7.6), suggests that the user reluctance might not be an issue if the users can see the benefits of the system.

- ***Implicit*** user data is collected without any direct user input. For instance, the gathered data can be the result of unobtrusive observations of users as they interact with the website, e.g. by recording click-stream data (users’ browsing history, shopping history, website browsing patterns) in website and user logs. The logs are subsequently mined and analysed. Such data can be described as *"behaviour-based"* (Middleton et al. 2004), and is typically *secondary data*.

One of the advantages of this method is that it removes the need for user input, thus alleviating the user's burden when supplying information explicitly (Gauch et al. 2007). It might also reduce the unnecessary additional effort, financial costs and resources, which have been identified as disadvantages of the explicit data collection. However, the downside of collecting data implicitly is a possibility of being affected by the *"cold-start"* problem), since the system might not have enough user data that is useful for a recommendation agent. Additionally, comparing it with explicit data, the quality of implicit data might be inferior (e.g. due to users' errors (Felden & Linden 2007), or unusual and atypical *one-off* usage pattern), and therefore less relevant to the system users (customers).

Some systems use only one type of data. For example, the user profiling systems described in (Katifori, et al. 2008), (Mussi 2006) and (Smyth & Cotter 2000) use explicit collection of user data, while Abbattista, et al. (2002), Aghasaryan, et al. (2010), Jiang & Tuzhilin (2009) and Kuo & Chen (2001) collect user data implicitly. However, other systems employ a ***"hybrid" user data*** acquisition approach, combining explicit and implicit data collection. For example, Gupta & Mathur (2002), Liu et al. (2008), Middleton et al. (2004) and Srikumar & Bhasker (2004) use a combination of both types of data. In these cases, the use of the hybrid approach is seen as an opportunity to maximise the benefits and limit or overcome the disadvantages of explicit and implicit data collection methods. Nevertheless, there might be even bigger effort and financial costs implications due to the need to collect and analyse two types of data.

Overall, finding other, more effective and efficient methods and techniques for capturing user data is one of the challenges in web personalisation. One particular area that is currently an active research area is the collection and use of data available on social networks. Such data is readily available, and there is a substantial amount of it. Accessing the *'social media'* data might overcome the issues associated with collection of sufficient and relevant data, and getting the users to supply it. In our project, this presents an opportunity for future work (see § 8.5).

2.4.1.2 Data Analysis

Data analysis is the step that follows the data collection, and is needed for user profile classification or re-classification. Typically, data analysis is accomplished using three approaches, *Rule-Based*, *Content-Based Filtering* and *Collaborative Filtering* (Mobasher 2007):

- **Rule-Based Filtering (RBF)** approach:

  This approach relies on deploying a set of rules that are created by the system architects. The rules can be based on particular characteristics of customers (e.g. demographics) or on specific business objectives (e.g. a particular marketing campaign). Typically, RBF-based systems are static in the sense that the rules require explicit specification and subsequent computation, and any changes in the user profiling or personalisation strategy need to be specified in a new set of rules.

  Examples of rule-based systems include: a semantic personalized e-Catalog service system (Chen et al. 2010); a personalized recommendation procedure (Kim et al. 2002); and a shopping comparison engine (Yuan 2003).

- **Content-Based Filtering (CBF)** approach:

  This approach works by matching content description of items found within the system (e.g. product characteristics that are specified in the company's product catalogue) with the content description of items specified by an individual user. Often, these systems deploy document modelling techniques, which have been widely used in Information Retrieval (IR) and Information Filtering (IF) (Manning, et al. 2008).

  Examples of systems that have utilised a content-based filtering method include: delivery of a personalised search experience (Hawalah & Fasli 2015); personalised product hierarchies (Niu et al. 2002); and a book recommending system (Mooney & Roy 2000).

- **Collaborative Filtering (CF)** approach:

  This approach involves collection and analysis of ratings or preferences from a community of users, and finding similarities in ratings or preferences of individual users. One of the notable characteristics of this approach is its reliance on the presence of a large amount of data.

  Examples that use the collaborative filtering approach are: a movie recommendation system (Bae, et al. 2012); multimedia news retrieval (Cantador et al. 2008b); dynamic generation of personalised links to the recommended web pages (Gupta & Mathur 2002); and ranked product recommendations (Srikumar & Bhasker 2004).

None of the above methods on their own have provided a perfect solution, and for that reason researchers have developed algorithms and systems that rely on hybrid approaches so that researchers can take advantage of benefits these approaches offer and minimise or overcome their limitations, e.g. to compensate for limitations of a single approach with regard to the system performance and scalability (Fink, et al. 2002), quantity and quality of data (Anand & Mampilli 2014), or the type of data (Park & Chang 2009). In particular, dense data is an issue for the performance of the CF-based approaches, while CBF approaches are highly dependent on the quality of content description (Anand & Mampilli 2014); neither is CBF approach suitable for multimedia data, such as video, audio or photos (Park & Chang 2009).

Thus, Smyth & Cotter (2000) reported improved success in personalising TV listing content after combining Content-Based and Collaborative Filtering. A hybrid of CBF and CF was also used for generating recommendations for TV news items recommendations (Fink et al. 2002), news and ads stream on a website (Zhong, et al. 2015), and for recommending online academic research papers (Middleton et al. 2004); whereas a combination of Collaborative Filtering and Case-Based Reasoning [(1)] was used to generate personalised information content in (Chedrawy & Abidi 2006). Burke (2007) examined and compared the performance of 41 different systems that use a hybrid approach to implementing recommender systems and found that the hybrid approach works well, provided that developers evaluate the accuracy and consistency of each component in order to *"determine its best role in a hybrid system"*.

Overall, each of the above approaches has particular advantages and disadvantages, and these are summarised in Table 2.4:

(1) Case-Based reasoning is an approach that relies on a specific knowledge, gained from *"previously experienced, concrete problem situations (cases)"*, which are then used to solve similar new problems (Aamodt & Plaza 1994). In our sample, this approach is represented by one paper only, hence was not covered in detail.

Approach	Advantages	Disadvantages
Rule - Based	• it is easy for systems to use "`if-then`" rules to select personalised information • captures common reasons for making recommendations	• it is not easy for administrators to pre-specify rules • it is not flexible • it is for groups or specific objectives (e.g. demographic groups, particular marketing campaign, etc) - i.e. not "real", truly "personal" personalisation
Content - Based	• most effective in text-intensive domains	• not suitable for multimedia information • cannot find potential interests of users
Collaborative Filtering	• popular in the e-commerce domain • can find potential interests for users • not limited to textual information	• “cold start” problem • new user problem • new item problem
Hybrid approach	• combines the advantages of different methods (e.g. content-based and collaborative filtering)	• it may be too complicated to achieve (implement) this approach

Table 2.4: Advantages and Disadvantages of Personalisation Approaches
(adapted from (Gao et al. 2010))

2.4.1.3 User Profile Generation, Classification and Maintenance

User profiling is a process that seeks to segregate users into particular groups or types which is based on the analysis of data represented in a user model.(1) User profiling lies at the heart of personalisation (Schubert & Ginsburg 2000), and is a key aspect in developing successful web personalisation (Trajkova & Gauch 2004), (Mobasher 2007).

The process of profiling users involves three major phases (*adapted from (Montaner et al. 2003)*):

1. **User profile generation** phase, which includes decision on how the user profiles should be represented (e.g. user data collection strategy and user model building) as well as generation of initial user profiles.

2. **Profile classification** (or *learning*) phase, during which generated user profiles

(1) In some literature, the two terms, *“user modelling”* and *“user profiling”*, are used interchangeably.

undergo an analysis and subsequent classification into a particular type.

3. **Profile maintenance** phase, which entails ongoing collection and analysis of up-to-date user data and, if necessary, user profile re-classification.

All three phases require a strategy on data collection and data analysis methods, which were discussed in § 2.4.1.1 and § 2.4.1.2.

Typically, two **types of user profiles** are distinguished, *group* and *individual*:

- **Group**(1) profiles involve clustering users (stereotyping) into groups based on certain criteria (e.g. similar demographics, interest, experience and context), and are characterised by their ‘non-unique’, generic nature. Such profiles can be applicable to a group of people, who share some particular characteristics, needs, requirements or interests. Examples of group profiles are found in (Gupta & Mathur 2002), (Jiang & Tuzhilin 2009), (Raghu, et al. 2001) and (Robal & Kalja 2007).
- **Individual** profiles are based on each user’s personal characteristics and/or traits. Such a profile belongs to one single person, and therefore typically is unique. Example systems that use individual profiles are: (Abbattista et al. 2002), (Kuo & Chen 2001), (Liu et al. 2008), (Rosaci & Sarnè 2014), and (Trajkova & Gauch 2004).
- It is also possible to develop a system that uses a combination of individual and group profiles, see (Park & Chang 2009).

The use of group profiles is common in marketing. With regard to system development, one of the advantages of the group profiles is the fact that they can help with the *“cold-start”* problem (Cena et al. 2012). It can be also be argued that group

(1) Also called *“stereotypical”* and *“communal”* in some literature.

profiles require less overhead effort in terms of design, implementation and computational power. (1)

Deploying individual profiles might increase the complexity of the system and affect the efficiency of the system run-time behaviour. However, individual user profiles might have a much better potential in offering more targeted, more appropriate and more accurate personalisation than group profiles (2). It is likely for this reason that deploying individual profiles has been more widespread in current research than group ones: there are 42 papers that use the individual user profile strategy compared with 11 that deploy the group profiles (3).

Either type of user profile (group or individual) can be designed and implemented as *static* or *dynamic* (4). Static user profiles do not change over time and are less costly to implement and maintain (from the point of view of computational power). However, static profiles might become less relevant ('decay') over time, e.g. as users' interests, experience, context or behaviour change. Dynamic profiles, on the other hand, do take such changes into consideration, but this type of profile is more costly and harder to implement and maintain.

2.4.1.4 System Implementation

Personalising content and its presentation can be achieved ***statically*** or ***dynamically***: through pre-computing elements or entities that are to be personalised or through continuous analysis of user interactions with the website and adapting the system *'on-the-fly'*.

(1) This, however, is an observational argument based on our own experience and not backed by a comprehensive analysis or evaluation in research. None of the papers that use the group user profiling justify the reasons for this choice of the user profiling strategy.
(2) Ditto.
(3) 3 papers use both types of profiles.
(4) It is interesting to note that although, in theory, it is possible to have dynamic group and static individual profiles, in practice, group profiles are typically designed and implemented as static, while individual profiles are mostly dynamic.

Static personalisation of content involves product recommendations, which are predetermined by the website's owners as recommendations, which the owners decide to be relevant to an individual or groups of customers. An example of this is the process that is commonly known as *'up-selling'*, i.e. when a company offers a product or service that can have additional (*'add-on'*) or upgradable components or features (common when selling, for example, computers or insurance, when the offered basic 'model' of a product is configurable). Another example when static personalisation of content is common is *'cross-selling'*. This is often used by clothing e-tailers to suggest additional products (e.g. accessories) that they decide would go with the product the user is looking at on a web page. The recommendations are then precomputed (*'hard coded'*) by the website developers.

Dynamic personalisation of content, involves computing product recommendations automatically by the system according to a specified logic or algorithm, without the need to specify the product recommendations manually, i.e. the dynamic content personalisation does not rely on a prior, explicit decision of what products can be suggested, but instead involves deployment of sophisticated information filtering algorithms.

Static and dynamic methods of content personalisation have similar advantages and disadvantages: dynamic content personalisation can be more accurate or relevant to a user at a specific moment in time, but requires more computational run-time overhead. Static, on the other hand, reduces the need for the run-time power, but might not be able to scale up in a system that employs complex personalisation features (Bunt, et al. 2007). Neither would it stand a good *'test of time'* in the fast changing environment of e-commerce. It is also interesting to note that static personalisation of content is not very common in research - none of the systems we have reviewed and analysed for our literature review use this approach to personalise content. This suggests that the static content personalisation has much less (if any) research value, which can be explained by the fact that the main source of data and algorithms for implementing the static personalisation is the explicit human input, while the system only plays the role of a *'dumb'* facilitator.

*Personalisation of **content presentation*** is typically dynamic. This is a reflection of how modern e-commerce websites are designed and operate. Some web page elements can and typically are pre-programmed (i.e. static), e.g. the page header, footer, and top, bottom and side navigation columns. However, the main body of a page is generated dynamically (*'on-the-fly'*).

2.5 Conclusions

Substantial academic and commercial work has been carried out in the area of web personalisation. Nevertheless, there are particular areas and aspects that we have identified as under-researched, under-developed or not covered, and therefore needing further investigation. This project addresses the limitations of the research belonging to the *"Applications/ Systems"* web personalisation research category, particularly:

(1) Imbalance of research on web personalisation:

The analysis of literature on web personalisation shows that there is a bias towards personalisation of content, and the content presentation personalisation is the area that is under-developed. For this reason, some of these limitations, namely investigation of novel and feasible ways to personalise web pages and product presentation are worth investigating.

Addressing these limitations also reflects the author's personal research interests. This project builds on the author's previous MSc work on flexible and alternative web interfaces (Stebletsova-Edwards 2003).

Research on content presentation (product recommendation) and website structure can also benefit from more research. In our project, these are taken into consideration, but not implemented. Instead, these were selected as areas for future investigation (see § 8.5).

(2) Ontology-based web personalisation:

Ontology-based profiling is one of the ways to personalise websites. This area is promising because of the belief that profiling users this way can offer the development of web personalisation that is more flexible, extendible, reusable, easier to implement, maintain and integrate, and compatible with the Semantic Web. For example, existing ontologies can be reused and merged to extend and enrich e-commerce domain knowledge, as well as to save on the development time and effort.

Moreover, ontology-based profiling can provide an opportunity to ensure compatibility with other web technologies that are widely used in e-commerce. E-commerce systems operate in an environment, where numerous technologies and standards are used to provide various functionality for data management, retrieval and exchange, such as XML [(1)], RDF [(2)], SOAP [(3)] and various Web Services tools and standards. Interoperability between these technologies and standards is vital for successful operation of e-commerce systems.

(1) "E**X**tensible **M**arkup **L**anguage" - `http://www.w3.org/XML/`
(2) "**R**esource **D**escription **F**ramework" - `http://www.w3.org/RDF/`
(3) "**S**imple **O**bject **A**ccess **P**rotocol" - `http://www.w3.org/TR/soap/`

(3) Limited use of research in the commercial world:

Comparison of the state-of-the-art literature with the use of web personalisation in the commercial world reveals the limited use of research by e-commerce companies.

Research on web personalisation has made substantial progress and contribution to the area (Gao et al. 2010). However, with regard to the current state-of-the-art in the *'real'*, commercial world, there is still considerable scope for further development of web personalisation. There are noticeable differences between what web personalisation features and functionalities are offered by the commercial companies (see § 2.3). Overall, the personalisation features and functionality that are used in the commercial world appear to be minimal, especially if these features and functionality are compared with the number and variety of those that have been proposed in academia.

One possible reason for this is the fact that practitioners are not keen to adopt research offerings due to the problem of integrating new components with the ones that already exist within the company's enterprise system (Goy et al. 2007). Other potential reasons are: the need for additional time, effort and funds to prove feasibility of a novel solution; and unwillingness to cause a disruption to customers by introducing changes (especially, if the feasibility and effects of the changes are unknown). Further research on feasibility of integration of different components can help improve this research's impact in the commercial world.

Research impact is an issue that has always been important. This is reflected in the fact that impact measures are now included in the REF (Research Excellence Framework) [(1)] for the UK universities. Thus, we have an opportunity in this project to investigate the feasibility of designing and implementing an e-commerce personalisation system, and to evaluate its acceptance by the users and industry.

Throughout this chapter, limitations in other areas were also identified, namely: more holistic investigation of the impact of web personalisation, development of metrics to measure the personalisation effectiveness and impact (*"Theoretical, Empirical Research"* category - see Fig. 2.1); and further research on privacy, ethics, trust, security and law (*"Others"* category), also call for further investigation. These limitations, however, are outside of this project's scope and should be tackled in a separate research project.

2.6 Chapter Summary

This chapter has presented a survey of research and commercial developments in the area of web personalisation. Focussing on e-commerce, we retrieved, reviewed and synthesised a considerable amount of related literature (255 papers).Building on previous literature reviews and our own analysis of existing academic work in the area, an alternative classification of web personalisation was presented. This was used to help structure a review of existing methods, tools and techniques used for personalisation of websites and to identify gaps in current research and practice.

Our analysis of the current state-of-the-art developments in academia and practice has revealed that a considerable amount of research and commercial developments has occurred in this area, but there are still areas and aspects of web personalisation

(1) `http://www.ref.ac.uk`

that need further exploration. These limitations were discussed, and the gaps that can be investigated have been identified.

Chapter 3

PERSONTO Requirements Specification and System Design

3.1 Introduction

The aim of the research, presented in this thesis, is to *"propose, test and evaluate the feasibility and acceptability of a flexible, extendible and reusable ontology-based e-commerce personalisation approach"*. This chapter deals with the first task, i.e the proposal for the system.

The chapter explains the purpose and scope of *PERSONTO*, describes the methodology used for the system requirements and system design, and provides the details of these requirements, design, top-level architecture and the website personalisation processes and approaches used in this project.

3.2 Purpose and Scope of *PERSONTO*

PERSONTO is a system to personalise e-commerce websites. Its feasibility will be demonstrated by the design and implementation of a *'proof-of-concept'* prototype.

As discussed in § 2.2.2.2, it is possible to offer personalisation of a number of the website's features. *PERSONTO* concentrates on the development of personalisation of content presentation (one of the gaps identified in § 2.5), with a provision to extend it to personalisation of content in the future.

3.3 Requirements Elicitation and System Design Methodology

The process of designing e-commerce systems is no different to designing any other software system. The design process typically follows a particular strategy, methodology and approach, and involves: elicitation of system requirements; defining a suitable system architecture and associated components; specifying functionality, content (i.e. data), content presentation (i.e. user interface), and approaches of how the components will be implemented - all of which satisfy identified business, user and system needs and requirements.

Requirements elicitation and specification is the first step in designing a system. In this project, we started with **identification of *PERSONTO*'s stakeholders** to define their typical interactions with the system, which is accomplished by using UML Use Cases (§ 3.4.1) [(1)]. Use Cases are extensively used for elicitation and analysis of system requirements (Fowler 2003). In our project, Use Cases were used to help us understand and structure the system requirements for PERSONTO, and

(1) Unified Modelling Language, `http://www.uml.org/`

subsequently to derive particular system characteristics and necessary functionality.

We then provided a **specification of the system requirements**, based on: *a)* the analysis of the project's aim and objectives (see § 1.2), and *b)* findings from our survey of research and commercial practices (§§ 2.2.2 and 2.3). The system requirements specification included the description of the requirements and explanation of the rationale for each requirement.

Upon completion of the system requirements, *PERSONTO*'s **system functionality** was specified (see § 3.5.3). For this task, we used UML Sequence and Activity diagrams to define and explain how web personalisation was to be achieved in this project. The specification included the description of the run-time architecture and behaviour, user (customer) profile access and control functions, and personalisation of the web pages.

3.4 Requirements Specification for *PERSONTO*

3.4.1 *PERSONTO* Stakeholders and Use Cases

PERSONTO has two types of stakeholder: *"Business"* and *"Customer"*.

A "`Business`" stakeholder is the entity that covers both, an e-commerce business and a web development company. Combining these into one stakeholder was deemed appropriate because of the similarity of the aims and objectives of these two types of business entities. For example, both entities seek to develop or re-develop an e-commerce system, both have various business constraints, such as budgets, deadlines and profitability. Moreover, it is possible to have an e-commerce company that implements and maintains their website in-house.

The "`Customer`" stakeholder is the entity that represents the customers using the website.

The UML Use Cases diagram in Fig. 3.1 illustrates a typical interaction between *PERSONTO* and its users:

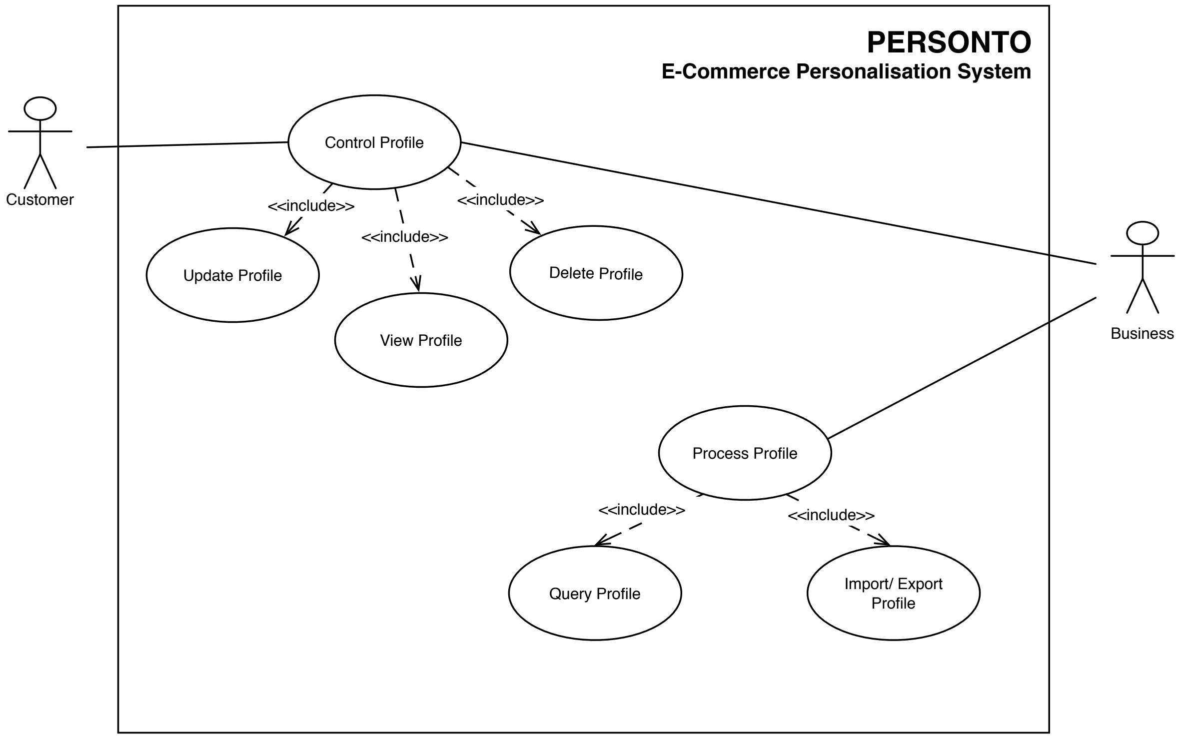


Figure 3.1: Use Case for *PERSONTO*

The actors (i.e. "`Customer`" and "`Business`" stakeholders) have direct access to user profiles, so that user profiles can be accessed and controlled. However, the "`Customer`" can only update, view and delete their own profile, while the "`Business`" stakeholder can update, view and delete any customer's profile. Additionally, the "`Business`" stakeholder has an ability to process user profiles, i.e to query and import or export all customer profiles.

3.4.2 System Requirements

Based on the analysis of the project's aim and objectives and the limitations of current research and professional practices, and the identification of the stakeholders and associated use cases, the following system requirements were identified.

[REQ.1] Customer (user) profile creation, access and control:

PERSONTO **shall provide a functionality to create, access and control customers' profiles**.

[1.1] A profile shall be created for each customer.

Initially, *PERSONTO* shall create a pool of initial customer profiles on the system's first run (one-off process). Additional profiles (e.g. for new customers) shall be created individually.

[1.2] ***PERSONTO*** **shall enable customers and system administrators to access and control customer profiles.**

A customer or system administrator shall be able to **update**, **delete**, and **view** customer profiles, stored in *PERSONTO*, which they have a permission to access and control, i.e. a customer is restricted to having access to or control over only their own profile, while a system administrator shall have access to all customer profiles.

A system administrator shall be provided with functionality to **query** and **import and export customer profiles**, stored in *PERSONTO*.

Rationale for [REQ.1]: This requirement addresses the typical interaction between the system's actors (stakeholders) and the system.

[REQ.2] Customer profile processing and storage:

***PERSONTO* shall provide a functionality to process customer profiles in order to classify each user into one of the predetermined types (profiles).** This classification shall be based on the class hierarchy and rules and restrictions, encoded in an ontology using appropriate representational and reasoning mechanisms, and automatically performed by a reasoner.

[2.1] ***PERSONTO* shall provide a facility to collect data from the customer to create a profile.**

[2.2] ***PERSONTO* shall provide a mechanism to process customers' profiles.**

[2.3] **PERSONTO shall provide a facility to store the customers' profiles in a repository.**

Rationale for [REQ.2]: Ontological user profiling has been found to be beneficial due to the advantages it offers (see § 2.2.2.3). The ontology needs to use an appropriate, Semantic Web compatible language to encode the class hierarchy, rules and restrictions. This language should provide sufficient capability for encoding class and property equivalence, class subsumption and disjointness, so that the customer's class membership (i.e. profiling into one of the specified profiles) can be inferred by the reasoner.

The ontology is required to store the customer data, which the reasoner (as the mechanism for processing the customer profiles) can access and analyse so that each customer is classified into one of the predetermined customer profiles, based on the rules and restrictions encoded in the ontology. The inferred class hierarchy is then used to personalise the web pages. The customer profiles also need to be stored to ensure the reasoner's persistent access to the data to enable re-computation of the class hierarchy (e.g. when the customers wish to update their preferences, or when the rules and restrictions in the ontology are changed).

[REQ.3] Dynamic web pages personalisation:

***PERSONTO* shall personalise web pages dynamically (at run-time) and display these to the website customers.**

Rationale for [REQ.3]: dynamic (*'on-the-fly'*) website personalisation is the most common and widely used type of personalisation in modern e-commerce systems (see § 2.4.1.4).

[REQ.4] Compatibility with existing systems:

***PERSONTO* shall be compatible with existing e-commerce systems**, so that *PERSONTO* can connect to existing back-end components without affecting the working order of these components or functionality they provide. Such connection shall be achieved and handled by a software component (programming code or package), which provides the communication between the existing system back-end and *PERSONTO*.

Rationale for [REQ.4]: This requirement addresses the *"Limited use of research in commercial world"* limitation of the current research (discussed in § 2.5); and the industry needs ([Pers.Comm.1] , [Pers.Comm.2], [Pers.Comm.3]).

As e-commerce development textbooks and professional practice show, it is common practice that the components are developed separately (or are bought as *'off-the-shelf'* products, developed by external vendors) and then connected together at the implementation/ maintenance system development stage. At times, particular components are replaced by new versions or makes, and additional components, which were not included in the initial stage of system design, are added at a later date. Therefore, there is an obvious need for compatibility in designing an e-commerce system, so that additional components and functionality can be integrated (be *"pluggable"* - (Goy et al. 2007)) into the existing system without affecting its overall working order.

[REQ.5] System Flexibility, Reusability, Extendibility:

PERSONTO **shall be designed and implemented to be flexible, reusable and extendible** to accommodate future development. For instance, future developments may include changes to different system's components (e.g. to a different reasoner); and extension of the current ontology by adding new classes, relationships and restrictions, and aligning or merging it with other existing or newly developed ontologies.

Rationale for [REQ.5]: This requirement addresses the project research aim to develop a flexible, reusable and extendible approach to e-commerce personalisation (see § 1.2).

The rationale for including this requirement is similar to **[REQ.4]**. To be of use in the commercial world of e-commerce or future research, the proposed system needs to be flexible, reusable and extendible. The system needs to offer a solution that is technology and implementation independent, so that, regardless of what technology is used for creation of web pages or what the website hosting setup is, it can link easily with them. It also needs to be able to deal with the variability of the business environment (e.g. changes in the number of customers or product ranges), and to respond to the changes (e.g. by re-evaluation of business strategy and processes, or business size scaling).

[REQ.6] Cost- and Effort-Effective Solution:

***PERSONTO* shall be designed and implemented to offer e-commerce businesses and developers a cost- and effort-effective solution, by using appropriate existing techniques and tools.**

Rationale for [REQ.6]: The rationale for this requirement is to ensure the *"Business"* stakeholder's acceptance (*'buy-in'*) of the proposed approach. The personal communications ([Pers.Comm.1] , [Pers.Comm.2], [Pers.Comm.3], [Pers.Comm.4]) suggest that minimising efforts and the costs involved in the design and re-design of their systems is an important critical success factor if it is to be adopted in the domain.

3.5 System Design of *PERSONTO*

This section provides details on the system design of *PERSONTO*. It forms the basis for the prototype implementation (see Chapter 6). The section describes the top-level architecture, explains the process of e-commerce personalisation in *PERSONTO*, and presents the specification of its system functionality.

3.5.1 *PERSONTO*'s Top-Level Architecture

The system requirements, specified in § 3.4.2, determined that our system to personalise an e-commerce website, *PERSONTO*, requires three components to perform website personalisation: an ontology, a customer profile repository and a reasoner. It also needs to connect to an existing e-commerce system. Fig. 3.2 shows the proposed top-level architecture of *PERSONTO*.

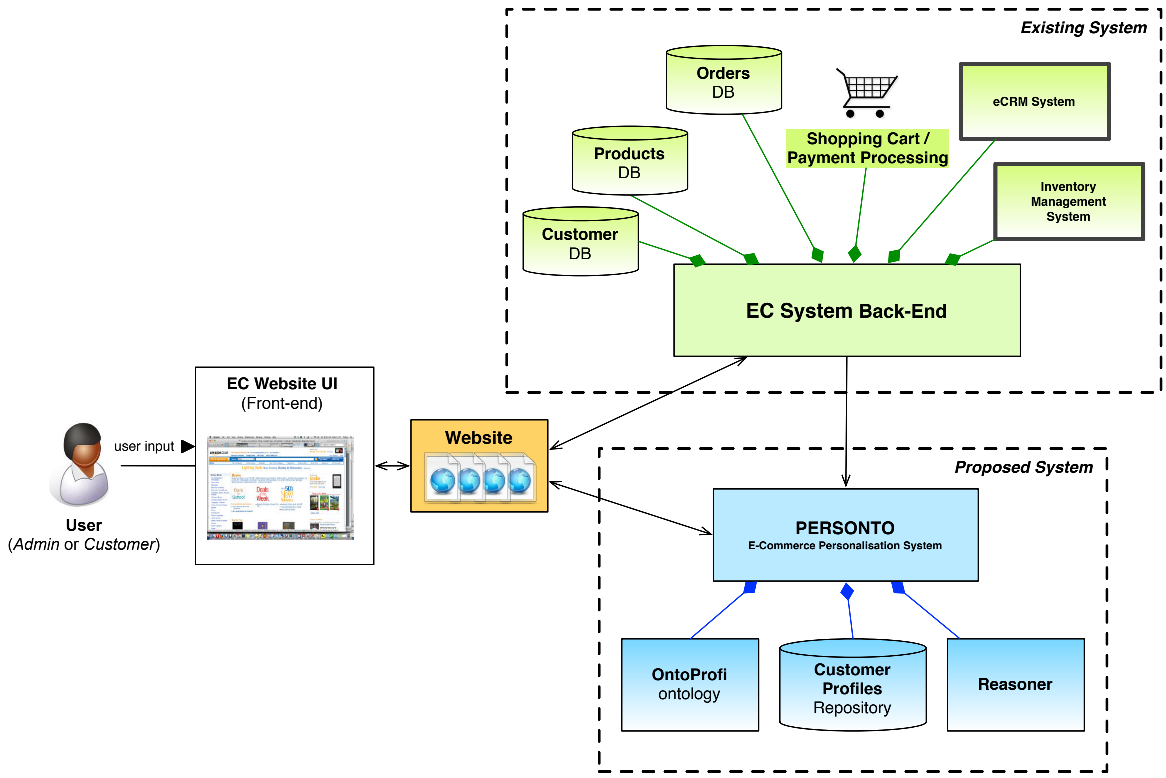


Figure 3.2: Architecture of *PERSONTO*

The *"Existing System"* box in the above figure represents typical entities, found in an e-commerce back-end system. The box shows a number of software and hardware components, such as various customer and order management facilities, e.g. product catalogue(s), inventory, CRM, payment processing and order fulfilment. These entities already exist. The typical front-end entities, such as a collection of web pages that form the website [(1)] and website user interface also exist, but were not included in the the *"Existing System"* box because they need to be modified (see § 3.5.3.5).

PERSONTO is shown in the *"Proposed System"* box, and consists of:

- **"*OntoProfi* Ontology"**:

  The personalisation in *PERSONTO* is based on profiling customers into particular types (groups), and this profiling is managed by a customer profiling ontology. As specified by the system requirements **[REQ.1]** and **[REQ.2]**, the role of *OntoProfi* is to create, store and modify each customer's individual profile. Thus, *OntoProfi* holds the class hierarchy, which represents e-commerce system entities (e.g. products, order fulfilment, communications) and relationships between these entities. It also holds a variety of rules and restrictions to enable the reasoner to classify each customer into a particular category (i.e. *"profile"*). Further detail on *OntoProfi*'s design and implementation are given in Chapter 4.

- **"Customer Profiles Repository"**:

  This component acts as a repository to store all asserted *OntoProfi*'s representation of a customer (as *OWL individuals* - see § 5.4.2) (**[REQ.2.3]**).

- **"Reasoner"**:

  This software component performs the actual process of profiling customers into a particular group (type) (**[REQ.2]**, and in particular, **[REQ.2.2]**). The reasoner evaluates (interprets) the encoded class hierarchy and rules defined in *OntoProfi* (see §§ 4.3.2 and 5.4), and computes the inferences, such as class hierarchy and class membership, based on the assertions in the ontology

(1) Websites are typically hosted on a web server, which, for the sake of clarity and simplicity of explanation of what needs to be modified, is not shown on our diagram explicitly.

(**[REQ.2.2]**). The reasoner's inferences are then used to personalise a website (**[REQ.3]**) (see § 3.5.3.3, where the role of the reasoner in *PERSONTO*'s run-time architecture is specified).

In the current version of PERSONTO's prototype, these elements connect to the back-end components of an existing e-commerce system, i.e. databases containing details of orders, customer information, products, etc., to acquire the necessary data, e.g. customer names, product specification, so that customer profiles can be created, controlled and analysed (**[REQ.1]**, **[REQ.2]**). They also interact with the system front-end, *"EC Website UI"* (via the *"Website"*) to acquire the customers' preferences for web pages layout and type of information, and, when the customer profile is known, to trigger personalisation of web pages (§ 3.5.3). Future versions of the prototype will require modification of the architecture, for instance, to enable writing back customer preferences collected by *OntoProfi*, explicitly or implicitly, and needed by the existing back-end entities. This will implement "Business perspective" personalisation (see § 1.5), such as preferences for payment or marketing communication.

Customer profiling ontology (*OntoProfi*) is the only component of *PERSONTO* that does not exist and needs to be implemented. The customer profiles repository, which is represented in the above figure as a separate entity, is, in fact, a *'conceptual'* entity, which is automatically created and updated by the *PERSONTO*'s programming code, based on the rules specified in *OntoProfi*. This provides the support for satisfying **[REQ.6]**. The reasoner, used in *PERSONTO* to analyse and classify data that is specified in the ontology, already exists. In this project, we use an open-source HermiT reasoner, which is available at `http://hermit-reasoner.com/`. However, other existing open-source reasoners can be used without affecting the outcomes of reasoning about *OntoProfi* [(1)], e.g. FaCT++ [(2)], Pellet [(3)]. These reasoners are interchangeable, which satisfies our aim and requirement to create a flexible system that is not dependent on one, very specific and bespoke reasoner (**[REQ.5]**). Connecting a reasoner to the customer profiling ontology is accomplished by using an API supplied with the reasoner (which supports **[REQ.6]**); or writing a bespoke piece of programming code.

3.5.2 The Process of E-Commerce Website Personalisation in *PERSONTO*

The process of designing and implementing e-commerce personalisation in *PERSONTO* is based on the four-stage approach to web personalisation used in research (see § 2.4.1), and is shown in Fig. 3.3, which displays the implemented components (green colour), partially implemented (yellow colour) and planned for future development (blue colour) [(4)].

(1) i.e. they will give the same results
(2) `http://owl.man.ac.uk/factplusplus/`
(3) `http://pellet.owldl.com/`
(4) Fig. 3.3 is an update of Fig. 2.3

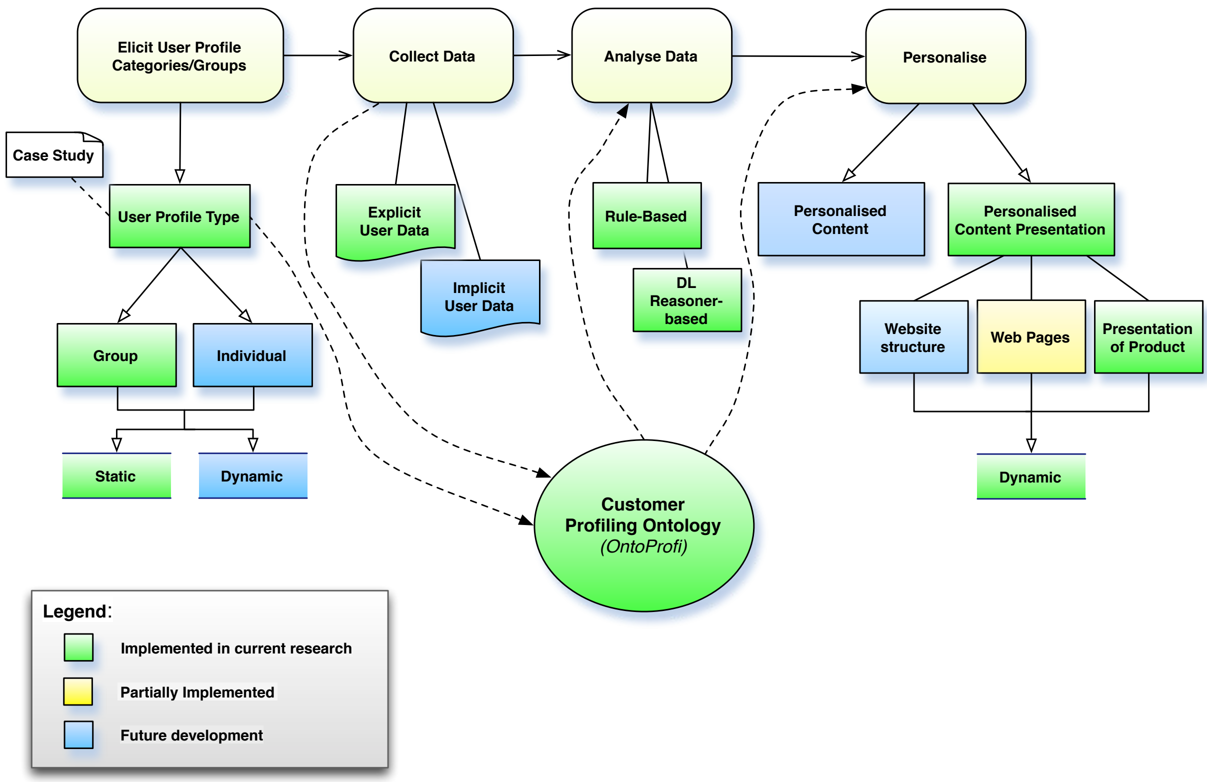


Figure 3.3: E-Commerce Website Personalisation in *PERSONTO*: Process and Approaches

- **Elicitation of user profiles categories or groups:**
  - The elicitation of user profiles categories/ groups is the first stage of implementing website personalisation in *PERSONTO*. It seeks to identify users' traits, behaviour, motivating or demotivating factors - so that these characteristics can be used to establish suitable profile groups or categories.

    Both, group and individual user profiles (static and dynamic), which were discussed in § 2.4.1.3, can be implemented in *PERSONTO*. Currently, user profiles are generated as static group profiles, due to the project's scope and limitations, and the prototype's purpose to test the feasibility of the proposed approach to website personalisation. In addition, it was believed that group profiles were more relevant and appropriate for personalisation of the content presentation (the focus of this project), whilst individual profiles may be a better choice for content personalisation (product recommendations). This belief was founded on the fact that there is a limited number of choices of: ***a)*** what content is available to be displayed on a web page (i.e. web interface components, such as page header and footer, product photos and description, side-bar menus); and ***b)*** the way the content is presented - even if the customers were offered an individual content presentation personalisation, there would be only a few interface layouts possible. In contrast, personalisation of content presentation (product recommendations) on an individual basis is more 'doable' (and appropriate) because product catalogues offer a greater variety of choices. Extending user profiling to dynamic and individual types represents interesting and promising future work (see § 8.5).

To facilitate individual profiles, the *OntoProfi* ontology will need to be extended by specifying additional object properties, datatype properties and rules (restrictions), which relate to new entities, e.g. by asserting a customer's preferences for product brands, communications types/ methods.

The elicitation of user profile categories or groups is accomplished by conducting multiple case studies of a group of online shoppers. As explained in § 1.3, this approach allows us to gain rich insights into the customer attitudes and behaviour while using e-commerce websites and shopping online, and elicit user profiles based on the rich data. Another important reason for choosing the case study approach is that the use of qualitative approaches is more common in professional practice than the quantitative approaches - due to a lack of resources or unwillingness to allocate the additional resources needed to involve a large number of respondents during website development. From the personal communications with the industry, companies prefer to use approaches that involve a limited number of potential system users, such as focus groups, individual interviews, walk-throughs and testing ([Pers.Comm.2], [Pers.Comm.3]). Thus, the case study approach is appropriate for use in industry.

- **Data Collection:**
  - *Explicit* and *implicit* data collection approaches can be used in *PERSONTO* to take advantage of both of these methods.

    The current version of the prototype uses explicit data as its main data source, and it comes from two sources,

    * Customer data that already exists in the system, such as customer name and other information on the customer (e.g. customer ID, name, address). This is used to create initial customer profiles (see § 3.5.3.1).
    * Data that is collected by the system, which is supplied by the customers, who wish to create their profiles and fill in the customer profiling survey (see § 3.5.3.4.2 and Fig. 3.7).

    The use of explicit data was considered to be relevant, appropriate and sufficient for the purpose of demonstrating functionality to personalise content presentation. *Implicit data* is not used in this version. Nevertheless, our ontology, *OntoProfi* (see Chapter 4) has been designed to allow future development, e.g. profiling customers based on analysis of a customer's shopping history and behaviour.

- **Data Analysis:**
  - Several web personalisation data analysis methods and approaches are used in research (see § 2.4.1.2). In this project, choosing an ontology-

based web personalisation approach determined the use of ontological-based reasoning, and more specifically DL-based. This supported the processing of the assertions made in our customer profiling ontology, *OntoProfi*. The DL-based reasoning was combined with rule-based data analysis to ensure that the rules specified by developers can be processed. The details of the assertions and rules, used in *OntoProfi* are given in §§ 4.3.2 and 4.3.3.

- **Website Personalisation**
  - Personalisation of content and content presentation in *PERSONTO* is designed to be dynamic to personalise a website *'on-the-fly'*, based on each customer's profile. Except for *'hard coding'* of the web pages header, footer and navigation bars, the static personalisation, such as manual selection of products (for content personalisation) or the *'hard coding'* of the main body of web pages (for content presentation personalisation) was not considered in this project since it is not common in research and practice. However, should static personalisation be of interest in future development, *PERSONTO* is flexible enough to accommodate this approach, if desired.

    The current prototype focuses on personalisation of content presentation, where adaptation of personalisation of product information was fully implemented, while an adaptive interface was partially implemented - personalised layout, but there is no personalised navigation.

    A future development will offer the personalisation of the content.

- **Customer Profiling Ontology:**
  - The *OntoProfi* ontology lies at the heart of *PERSONTO*. The ontology

needs input from the first two stages of the e-commerce website personalisation process to create and populate the ontology; and it provides data outputs required for the third and fourth stages.

3.5.3 Specification of System Functionality

To enable a personalisation functionality, *PERSONTO* requires implementation of two main phases:

- **Pre-processing phase**: this is necessary for a one-off creation of the bulk of initial profiles for existing customers, i.e. creation of the *PERSONTO*'s *"Customer Profiles Repository"* (see § 3.5.1).
- **Run-time phase**: this provides the rest of *PERSONTO*'s functionality, such as profile access and control functions. These are used to profile customers and, subsequently, to personalise the website.

3.5.3.1 Pre-Processing Phase: Batch Creation of Initial Profiles

During the first system run, *PERSONTO* creates initial profiles for all existing customers, using the details that customers provided when registering with the EC company, i.e. details such as their names and addresses, which are available in the company's database.

Creation of the initial customer profiles does not depend on any direct or explicit user input. It is run as a process, *'invisible'* to the users, and generates basic profiles containing basic information. More specifically, as is explained in § 5.4.2, a profile for each customer is created as a distinct *OWL Individual*, whose name is unique - a combination of `Customer_ID` and `Customer_Surname` stored in the `Customer` table of the company's database:

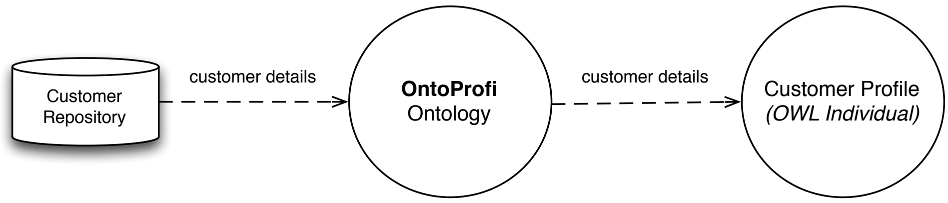


Figure 3.4: Pre-Processing Phase: Creating Initial Customer Profiles

Each customer profile is then stored in PERSONTO's *"Customer Profiles Repository"* (see § 3.5.1 and Fig. 3.2).

The initial profile does not contain any preferences specified for a customer, i.e. it is intended to be a template used for populating the customer profile with their preferences when these are specified. This is accomplished during the *"Profile Updating"* phase (see § 3.5.3.4.2).

The pre-processing phase is a one-off process when the system is created. Subsequent addition of customer profiles, e.g. for newly registered customers, will need to be run separately. This can be done either as a repeat of this phase (i.e. bulk re-creation of all customer profiles), or new profiles can be added individually (see § 3.5.3.4.1).

3.5.3.2 Run-Time Phase

The run-time phase is responsible for providing the main functionality to enable the website personalisation.

3.5.3.3 Run-Time Architecture

The overall run-time architecture connects all the prototype's functional components (see Fig. 3.5). Individual components are shown along with related aspects, described in more detail in the remainder of this section.

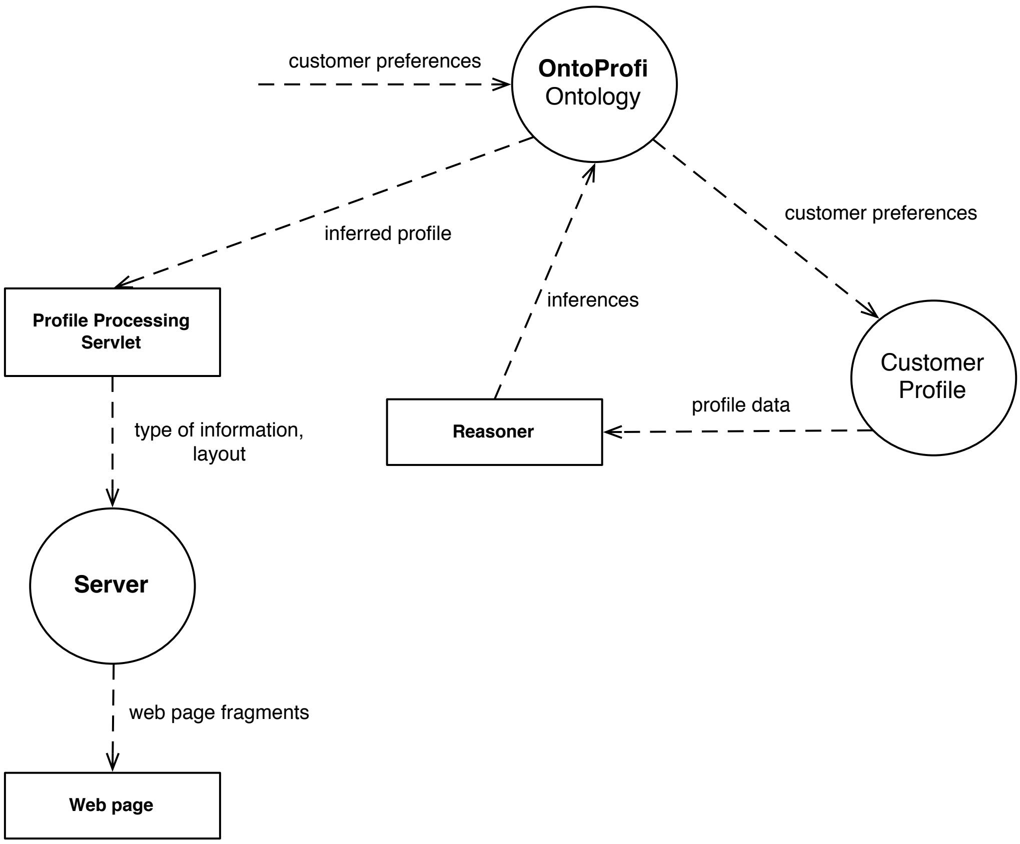


Figure 3.5: Top-Level Profile Processing (*Run-Time*)

When customer preferences are specified, they are stored in each customer's individual profile. The preferences are then passed to the reasoner, which computes (infers) the class hierarchy for the customer profile. The reasoner passes the inferences back to *OntoProfi*, which, in turn, sends it to the profile processing servlet. The servlet determines what type of information and web page layout is appropriate for the inferred customer profile and instructs the server to assemble the resulting web page from the appropriate web page fragments.

3.5.3.4 Profile Access and Control

To satisfy the specified use cases (see § 3.4.1) and **[REQ.1]**, *PERSONTO*'s run-time functionality involves a number of customer profile access and control functions, such as creation of additional profiles, profile updating, deletion, viewing and importing (see Fig. 3.6).

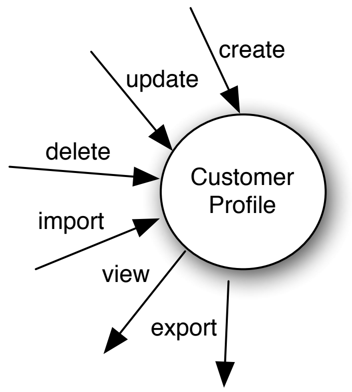


Figure 3.6: Customer Profile: Access and Control Functions

3.5.3.4.1 Adding Additional Individual Profiles

Additional individual profiles can be created by a process similar to the one described for the pre-processing phase. The only difference is that we only need to retrieve an individual customer, and add his or her profile to the existing bulk of the customer profiles already created.

3.5.3.4.2 Profile Updating

The initial customer profile, created at the pre-processing phase, is only of use if we tailor it to each customer's needs and preferences. Therefore, each initial profile needs to be updated to take into consideration each customer's unique preferences.

In the current version of *PERSONTO*, the customer's initial profile is updated by asking the customer to complete a survey. Survey-based user profile updating is a common practice in this research area, e.g. (Blasco-Arcas et al. 2014), (Golemati, et al. 2007), (Gupta & Mathur 2002), (Ntawanga, et al. 2008) and (Skillen, et al. 2014)). It is also wide-spread in e-commerce practice. E-commerce companies, such as *yoox.com*, *theoutnet.com*, *www.amazon.co.uk*, *www.shopbop.com* encourage their customers to fill in web-based forms (surveys) to specify their preferences. For example, *yoox.com* allows a customer to create an individual profile based on the customer's specified preferences for a type of product and clothing size. This profile can then be used to filter search results. *theoutnet.com* invites customers to specify sizing preferences so that the marketing emails will be personalised to show only the items in the specified sizing. *www.amazon.co.uk* and *www.shopbop.com* offer personalisation of e-mail subscriptions, which the users can choose to subscribe to, e.g. *"Daily Editorials"*, sale and special offers, or e-mail alerts on specific types of products or brands.

Fig. 3.7 shows the survey, used in this project to obtain (or update) users' preferences for type of product and information and to re-assert the customer's profile.

Figure 3.7: "Profiling Survey" page.

The overall run-time architecture, specified in Fig 3.5, is updated to include survey-based profile updating (see Fig. 3.8).

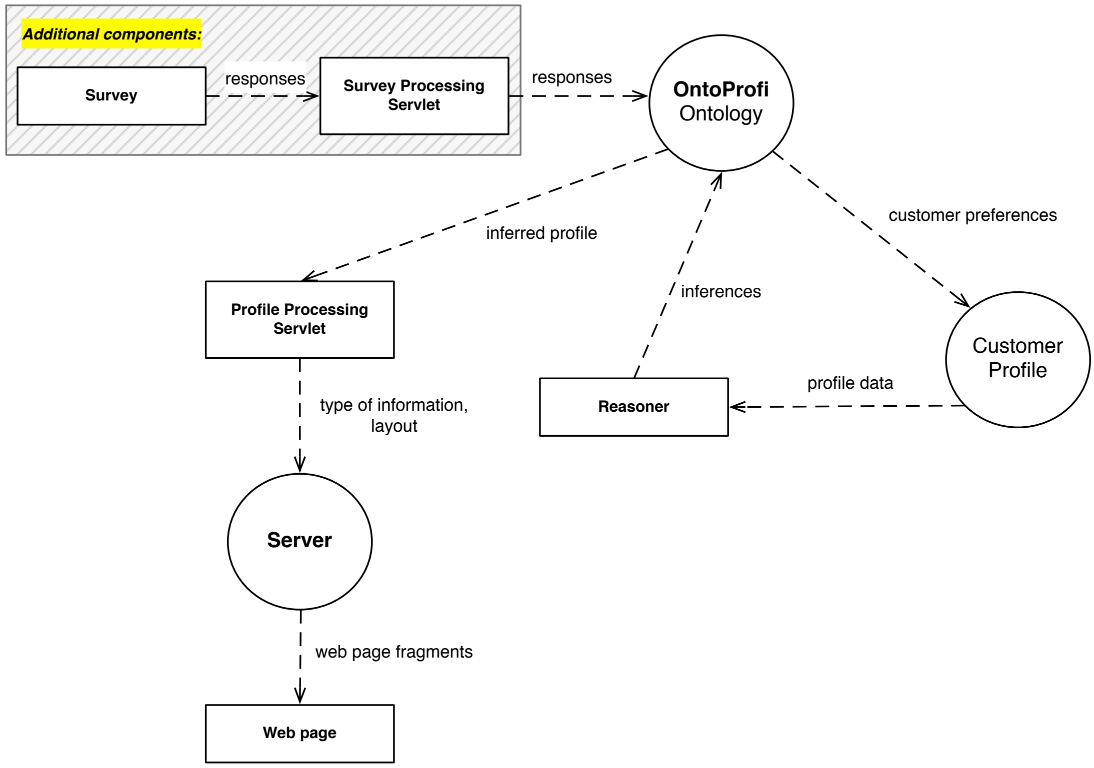


Figure 3.8: Survey-Based Profile Processing (*Run-Time*; update of Fig. 3.5)

More specifically, the UML Sequence diagram (Fig. 3.9) illustrates the process of updating customer profiles, based on their responses to the survey.

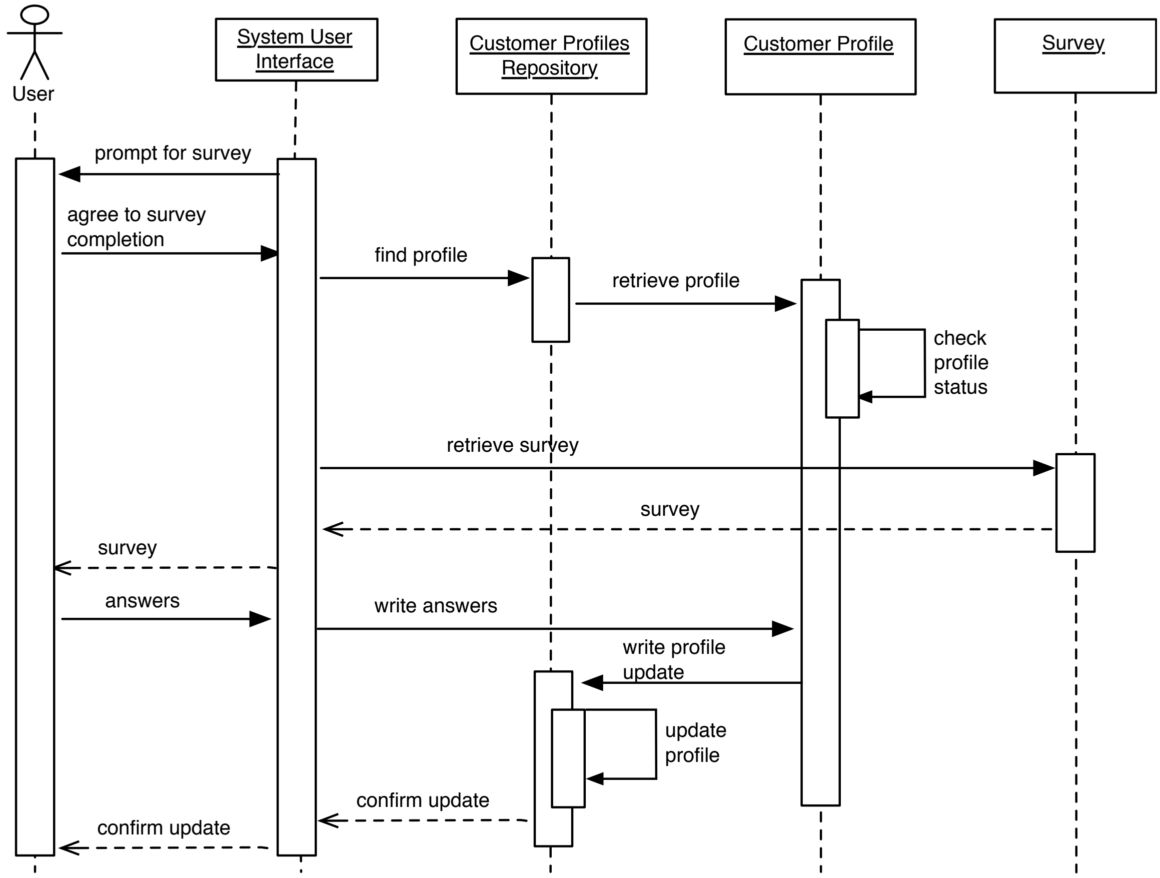


Figure 3.9: Updating of Customer Profile (*Survey-Based*)

For the purpose of this diagram, we assume that a customer is logged in (i.e. the system already knows who this customer is). The customer is invited to complete the survey, and when she or he clicks on the link to the survey page, the customer's profile details are checked and retrieved. On completion of all survey questions and clicking on the "Submit" button, the customer's answers are written to the customer's profile and the profile is updated (stored) in the customer profiles repository.

Apart from survey-based profile updating, it is also possible to update a customer profile manually - by direct profile editing. In this case, instead of providing a customer with a survey, the customer updates individual parts of his or her profile. The sequence of interactions between the customer and relevant entities of *PERSONTO* is shown in Fig. 3.10.

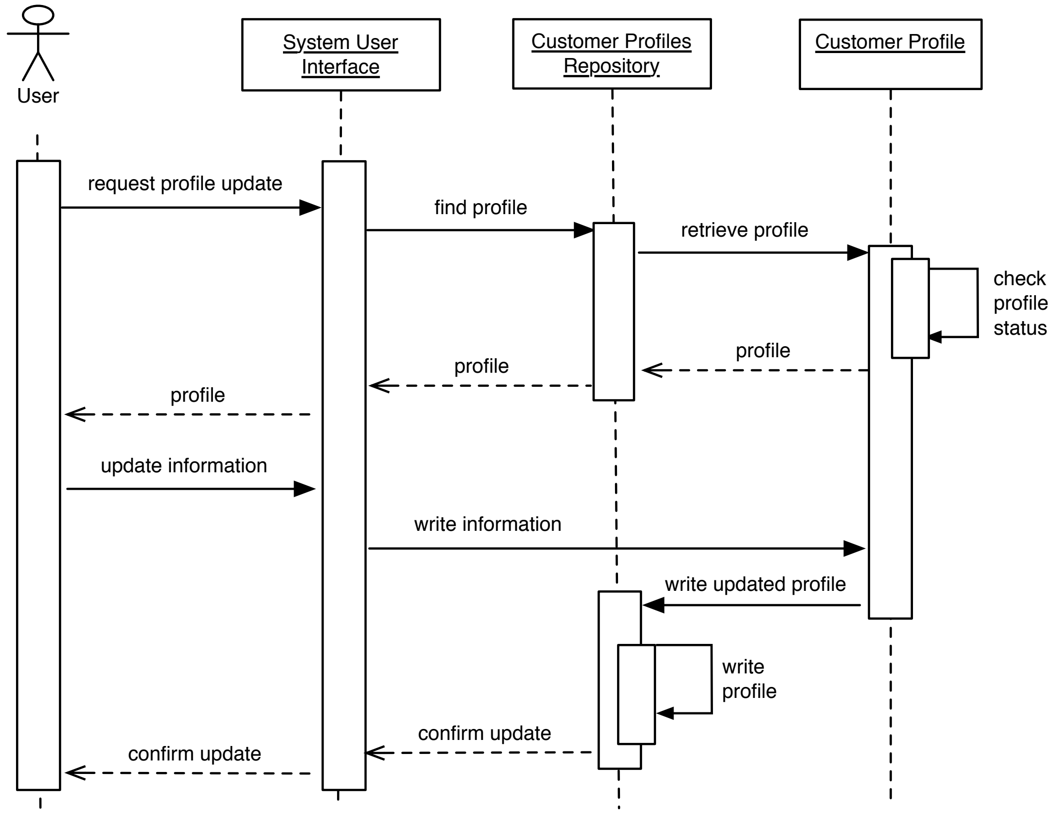


Figure 3.10: Updating Customer Profile (*"Non-survey", Manual Update*)

The two ways of updating customer profiles, just described, can both be categorised as *manual* due to the need for explicit user involvement. A third way to update a customer profile is to do it *automatically*, for example by analysing a customer's interaction with the website, such as examining browsing patterns or inspecting past purchasing trends; the system will then store and analyse these without explicit user involvement. This, however, is not covered in the current version of the prototype, and is a matter for future work (see § 8.5).

3.5.3.4.3 Profile Deletion

"Profile Deletion" functionality is provided for customers, who express a wish to delete their profiles. Such functionality is desirable and in some cases may also be necessary for ethical and legal reasons, for example, to comply with legal requirements, such as the UK's Data Protection Act 1998 (1) and the EU's Electronic Commerce Directive 2000/31/EC (2)

A typical sequence of interactions for deleting a customer profile is shown in Fig. 3.11.

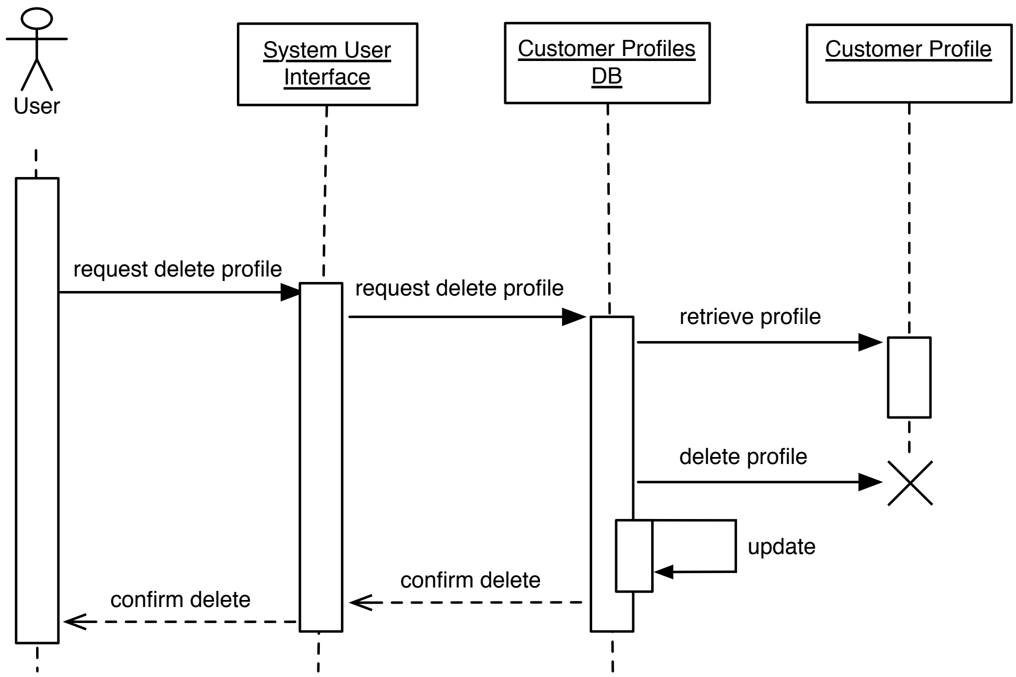


Figure 3.11: Deleting Customer Profile

It is important to note that:

– Profile deletion does NOT remove customer data held somewhere else in the e-commerce system - only information that is contained in this user profile.

(1) http://www.legislation.gov.uk/ukpga/1998/29/contents
(2) http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0031&from=EN

Customer data is held in a separate location (e.g. in the company's *"Customers DB"*), and only copied (duplicated) to a customer's profile, when this is created (pre-processed). To delete customer data completely, the whole customer account needs to be deleted. However, this is beyond *PERSONTO*'s current responsibility and implementation.

– On the other hand, it is assumed that, should a customer account be deleted, the customer profile, associated with this customer, will be removed from the customer profiles repository - by searching for this particular customer's profile in the *OntoProfi* ontology and removing this profile as shown in Fig. 3.11.

3.5.3.4.4 Profile Viewing and Importing

"Profile Viewing" functionality is designed to allow customers, developers or e-commerce company owners to access data that is held about each individual customer. *"Profile Importing"* functionality is only useful for developers or e-commerce company, for example, during system backup or migration. It is not envisaged to be useful to customers.

Viewing and importing individual, selection of, or all profiles is a *'trivial'* process that can be accomplished by implementing simple `get` functions, for example, by coding functions to search for all or individual records, stored in the *PERSONTO*'S *"Customer Profiles Repository"*, displaying these, and copying them into a new location.

3.5.3.5 Personalisation of Web Pages

Personalisation of web pages is accomplished as a server-side functionality. Fig. 3.12 illustrates the process of personalising web pages based on an inferred customer profile.

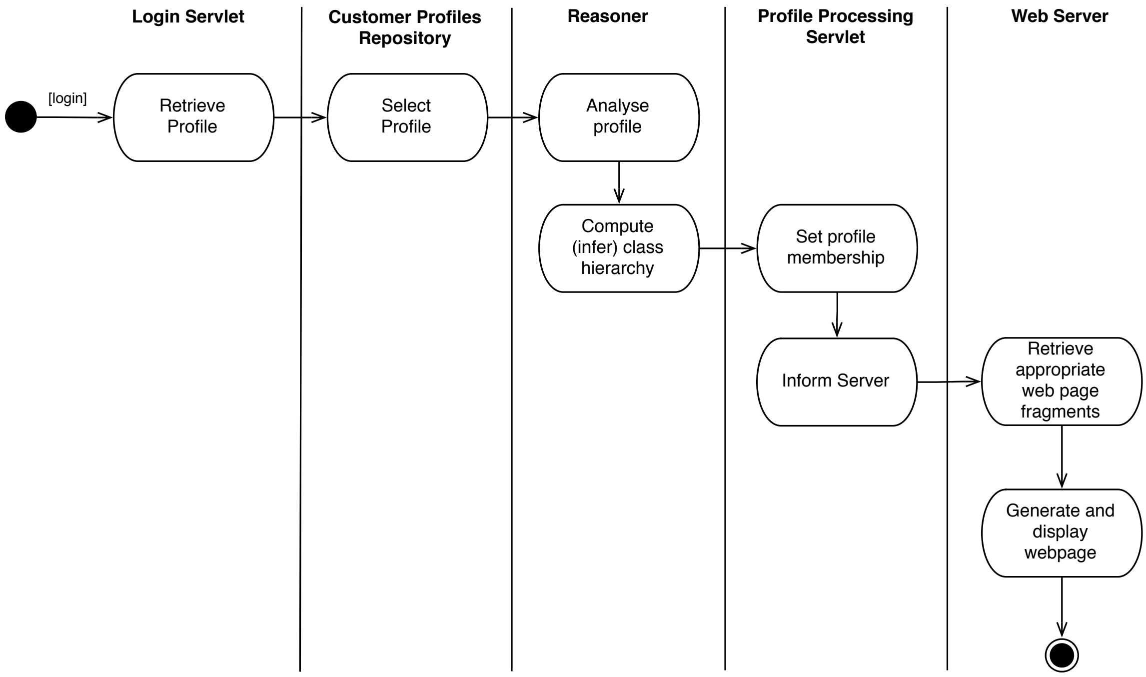


Figure 3.12: Customer Profiling Activity Diagram

After a customer's successful login, the customer profile is retrieved from the *"Customer Profiles Repository"* and sent to a Description Logic-based reasoner. The reasoner analyses the profile, computes the class hierarchy, based on the asserted profile properties (i.e. customer preferences), and sends the customer's *inferred* profile to the profile processing servlet. The servlet then associates (sets) the inferred class membership with the retrieved profile and informs the web server, which generates a web page by assembling appropriate page fragments.

3.6 Chapter Summary

This chapter described and discussed the system requirements and system design for our proposed web personalisation system, *PERSONTO*. It provided the explanation of the purpose and scope of the system, the methodology used to elicit the system requirements, and the specification of the system requirements and system design. This specification forms the basis for the implementation of the identified system's components, in particular, of the customer profiling ontology, *OntoProfi* (covered in Chapter 4), and of the system's prototype (see Chapter 6).

Chapter 4

OntoProfi: Ontology for Profiling Online Shoppers

4.1 Introduction

As specified in the previos chapter, website personalisation in *PERSONTO* relies on the use of a customer profiling ontology as a key component that enables personalisation of content and content presentation.

This chapter discusses the methodology and design of the *OntoProfi* ontology. Particularly, it specifies the methodology used to engineer *OntoProfi*; explains the reasons behind choosing *OWL 2* as the language to represent *OntoProfi*; provides a detailed description of the conceptual design of the ontology; and gives examples of how *OntoProfi* can be used in future projects.

4.2 *OntoProfi*'s Engineering Methodology

Ontology engineering typically follows a process that consists of six stages (Noy & McGuinness 2001). These stages are used to determine the ontology domain and scope, select the ontology representation language, enumerate domain terms, define appropriate classes and the class hierarchy, encode the properties of classes, their relationships and restrictions, and create instances of classes (see Fig. 4.1).

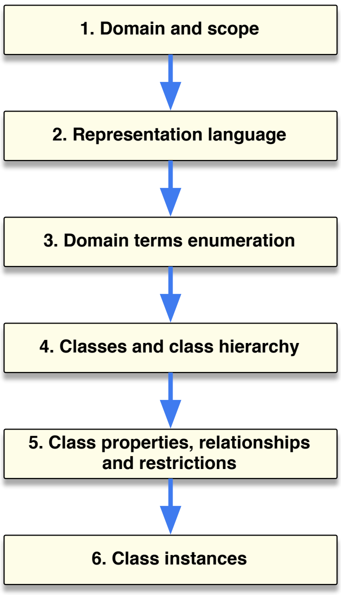


Figure 4.1: Ontology Engineering Process *(adapted from (Noy & McGuinness 2001))*

This methodology was followed in the project, and the outcomes of each stage were:

1. **Determination of *OntoProfi*'s domain and scope:**

    - The determination of *OntoProfi*'s domain and scope was based on the analysis of: the project's aim and objectives (§ 1.2); project's focus (§ 1.5); literature review findings (§ 2.2.2); and the identified system requirements (see § 3.4.2, in particular, **[REQ.1]** and **[REQ.2]**. Thus, it was determined that *OntoProfi*'s domain coverage is the domain of online shopping (e-commerce). In this project, we particularly focus on Business-to-Customer (B2C) type of e-commerce due to the characteristics of the case study participants, involved in the determination of the profiles e-commerce customers can be profiled into (discussed in Chapter 5). However, most (if not all) of the encoded entities, their relationships and other assertions are also considered to be relevant to other types of e-commerce, e.g. Business-to-Business (B2B), and can be extended to account for the specifics of a particular company or industry.

The scope of *OntoProfi* was established as: to enable profiling of e-commerce customers into a particular profile. More specifically, firstly, our ontology was required to support our project's focus on *"technology perspective"* of personalisation (i.e. content and content presentation personalisation), and, secondly, it needed to enable the future support of the *"Business perspective"'* (i.e. business-related activities, such as marketing, CRM and business rules strategies; § 1.5).

2. **Selection of ontology representation language:**

   - The choice of a language to represent an ontology is typically dictated by the outcomes of the first stage of the ontology engineering process. Therefore, the ontology language for *OntoProfi* needed to be appropriate to the ontology's domain and scope specified above. It also needed to satisfy the system requirements for: compatibility with an existing e-commerce system (**[REQ.4]**); *PERSONTO*'s flexibility, extendibility and reusability (**[REQ.5]**); profile creation, access, control, processing and storage (**[REQ.1]** and **[REQ.2]**) (see § 3.4.2). In addition, the chosen representation language needed to be compatible with Semantic Web technologies due to the benefits they offer (as discussed in § 2.2.2.3). *Web*

Ontology Language (OWL), specifically the DL variant, was chosen as the most appropriate ontology representation language for *OntoProfi*. A detailed justification for this choice is in § 4.3.1.

3. **Enumeration of terms that belong to the chosen domain:**

    - The enumeration of terms involves the identification of concepts that belong to the chosen domain to enable the creation of an appropriate class hierarchy.

      In this project, the initial terms were derived from the analysis of relevant literature and professional e-commerce practices (see Chapter 2), as well as from the author's own experience in researching and teaching e-commerce. Appendix D shows the origin of the top-level classes, which were chosen to be included in *OntoProfi*.

      On completion of the online shopping case studies, the collected case study data (see §§ 5.2.3 and 5.3) was used to extend *OntoProfi* by identifying additional classes and restrictions to enable customer profiling and website personalisation (see § 5.4).

4. **Definition of appropriate classes and the class hierarchy:**

    - This stage deals with conceptualisation of ontological classes and the class hierarchy. The terms, identified in the previous stage, were analysed to arrive at appropriate class definitions and a hierarchy in *OntoProfi*. The resulting subclasses were checked for relevance against our case study data, related literature and professional practice. The overall process of creating the class hierarchy underwent several iterations, and resulted in a class hierarchy containing four top-level sub-domains (see § 4.3.2).

5. **Encoding of properties of classes, their relationships with each other, and any required restrictions:**

    - This stage dealt with extending *OntoProfi* by encoding properties, relationships and restrictions on the classes, specified in the previous stage. In particular, the focus of this stage was to assert the restrictions needed for classifying the customers into different types. This stage was completed after the findings of the online shopping case studies were finalised, i.e. after the customer profiles were elicited (see § 5.3.3), and is described in § 5.4.

6. **Creation of instances of classes:**

    - In *OntoProfi* instances of classes represent individual customers and their associated profiles. This stage was also completed after the completion of our case studies. The creation of individuals and the encoding of underlying logic that enables the classification of individuals into the proposed profiles is discussed in § 5.4.2.

4.3 *OntoProfi* Design Specification

4.3.1 Ontology Representation

As already stated, the OWL 2 *Web Ontology Language* (W3C OWL Working Group 2012), and more specifically, its *OWL 2 DL* variant, was chosen as the most appropriate representation and implementation language for *OntoProfi*. This section provides an overview of OWL and its building blocks, and explains the reasons for choosing this ontology language.

4.3.1.1 OWL Overview

OWL is a family of ontology languages developed to represent knowledge. OWL is a semantic markup language, which is used for encoding, using and sharing ontologies, particularly targeting Semantic Web-based systems and applications.

OWL was developed to extend the vocabulary of its predecessors, such as RDF [(1)], and its design was influenced by and incorporated many features from other ontology languages, such as DAML+OIL [(2)], which OWL superseded. Fig. 4.2 shows the positioning of OWL within the XML-based ontology languages hierarchy (the arrows in the figure show antecedents).

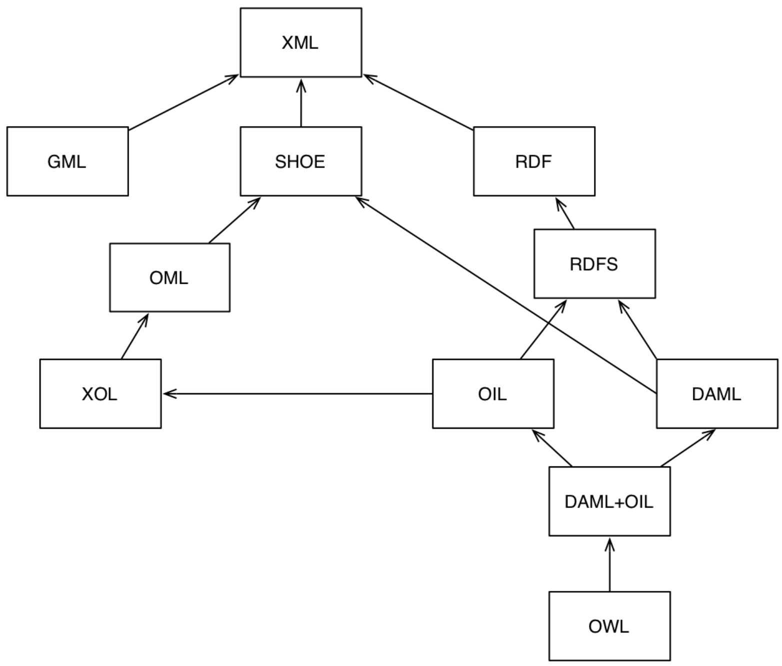


Figure 4.2: XML-based Ontology Languages Hierarchy
(adapted from (Jones et al. 2002))

Besides being classified as an XML-based language, (and like its predecessor OIL) OWL also belongs to the Description Logic (DL) family of languages (Horrocks &

(1) Resource Description Framework - `http://www.w3.org/RDF/`
(2) `http://www.w3.org/TR/daml+oil-reference`

Patel-Schneider 2011).

The latest version of the language is *OWL 2* (W3C OWL Working Group 2012). This version is the result of substantial revision, extension and improvement of the *'original'* OWL, which was first proposed by The World Wide Web Consortium (W3C) in 2004 (Bechhofer, et al. 2004). *OWL 2* was developed to overcome expressivity limitations and syntax issues, which *OWL 1* had (Grau, et al. 2008). For example, *OWL 2* (1) offers additional features, which allow metamodelling, multipart properties and multiple inverse functional properties (Allemang & Hendler 2011).

OWL was chosen as the most appropriate language for this project because:

- OWL was developed as a Semantic Web language, and is now considered to be the ***de facto standard*** for the Semantic Web. OWL is a replacement for its predecessor, DAML+OIL (2), which, at present, is not often (if at all) used any more in research and practice.

  In addition, as discussed in § 2.2.2.3.3, e-commerce is a particular area that can benefit from the use of ontologies, and since the language has been recognised as the *de facto* standard for ontology representation, it was regarded as the most appropriate for this project.

- OWL has a **well-defined *syntax***, **formal *semantics***, and **substantial *expressive powers*** (Hitzler, et al. 2012). The language was developed to fully exploit the results of extensive research on Description Logic (DL), and to support ontological reasoning performed by existing DL reasoners (Baader, et al. 2008).

- The use of OWL is **widespread** and it is **supported by a wealth of readily available tools and resources**, such as reasoners, APIs, extensions, devel-

(1) For the sake of readability, from now on we will refer to OWL 2 as just *"OWL"* and OWL 2 DL as *"OWL DL"* (i.e. dropping the "2" from the term). However, if and when it becomes necessary to distinguish between different OWL versions, we will do this explicitly.

(2) `http://www.w3.org/TR/2001/NOTE-daml+oil-reference-20011218`

opment frameworks and environments. In our project, this meant that we had an extensively tested, powerful research tool that could be used without the need for *'reinventing the wheel'*, i.e. we could implement an ontology by using the existing constructors, and could reason about our ontology by using existing reasoners. This meant that the overhead of development efforts was significantly reduced (supporting **[REQ.6]**, see § 3.4.2).

- OWL ensures interoperability (compatibility) between different Semantic Web technologies and tools, the importance of which for this project was discussed in § 2.5; and which is also one of the system requirements for this project, **[REQ.4]**.

- OWL ontology is **flexible** and **extendible**, which, again, satisfies our project requirement of flexibility, specified in the project's aim and system requirements (§ 1.2 and **[REQ.5]**). Moreover, flexibility was also needed to provide opportunities for future development, so that *OntoProfi* could be extended by aligning or merging with other existing or future ontologies (see § 8.5), for example, with the *GoodRelation* ontology that targets the e-commerce domain (covered in § 2.2.2.3). Interoperability and flexibility is helped by the fact that reliable and efficient mechanisms/ tools for ontology aligning, merging and intercommunication already exist and are readily available.

4.3.1.2 OWL's Building Blocks

A typical OWL ontology consists of: *classes*, *object properties* and *datatype properties*.

Classes are the main building blocks of an OWL ontology, and are defined as *sets* containing *individuals* (instances of a class).

Arguably, classes can be viewed as the only compulsory element in an ontology. Other elements, such as class properties and constraints, are optional, and may or may not be encoded. However, encoding properties and constraints are necessary for creation of meaningful class definitions and relationships to enable useful and accurate modelling of the chosen domain.

Classes are typically organised into a *taxonomy*, i.e. "superclass - subclass" (or "`is-a`" and/ or "`part-of`") hierarchy. OWL language allows both assertion and inference of the *"superclass - subclass"* relationship. In other words, a class taxonomy can either be encoded manually by an ontology engineer, or automatically computed (inferred) by a reasoner. (*OntoProfi*'s class hierarchy is discussed in § 4.3.2.)

Object and ***Datatype properties*** are used for asserting relationships between classes and individuals. *Object properties* specify the relationship between two classes (or two individuals). *Datatype properties* bind an individual to particular datatypes, such as RDF literals of `xsd:string`, `xsd:integer`, `xsd:date`, etc.

Object properties can be encoded to exhibit particular characteristics, such as having *functional, inverse functional, transitive, symmetric, asymmetric, reflexive,* and *irreflexive* properties. An *Object property* may also have a corresponding *inverse* property, which is encoded as `owl:inverseOf`. *Datatype property* can only have one optional property - *functional*. (Object and datatype properties used in *OntoProfi* are discussed in § 4.3.3.)

To support and enable reasoning about an ontology, classes can be described using various ***restrictions***. These include Universal and Existential quantification (i.e. $\forall$, or the OWL equivalent of `owl:allValuesFrom`; and $\exists$, or `owl:someValuesFrom`); as well as a special case of the `owl:someValuesFrom` restriction - `owl:hasValue`.

Set operations (*union, intersection, complement*), cardinality, disjointness, class equivalence, class partitioning, and domain and range can be used to make powerful and accurate representation of the domain, modelled in an ontology.

The restrictions used in *OntoProfi* are discussed in the remainder of the chapter, and in § 5.4, where we explain how *OntoProfi* is extended by including specific customer profiling classes.

4.3.1.3 *OntoProfi*, Reasoning and OWL DL

OWL provides powerful mechanisms for asserting class and property equivalence, subsumption, disjointness, and a variety of property characteristics as well as restrictions. These assertions are used by a reasoner for automatic computation (*"inference"*) of subsumption, equivalence or non-equivalence of classes or properties; class disjointness; as well as equivalence/ non-equivalence of individuals. In this project, the support for reasoning was important due to requirement to provide a cost- and effort-effective solution to achieve the industry's buy-in (**[REQ.6]**). By using an existing reasoner to reason about *OntoProfi*, customer profiling is accomplished without the need for *'hand-coding'* of how each customer is profiled, which ensures the costs and effort effectiveness of the system implementation and, therefore, improved attractiveness of the proposed approach to e-commerce businesses and developers.

More specifically, *OntoProfi* must represent the complexity of relationships between the customer and various e-commerce entities, e.g. *"customer - e-commerce business"*, *"customer - website"*, *"customer - products"*, *"customer - order fulfilment"*, by asserting (encoding):

- an appropriate class hierarchy to specify the relevant class descriptions and statements (OWL axioms), such as "superclass - subclass" relationships, sets, property and datatype descriptions;

- appropriate property restrictions to define the classes in the class hierarchy, for example, by asserting existential and universal quantification, object and data properties' hierarchy and their characteristics;
- a set of individuals to represent each customer and associated class descriptions and property restrictions.

Table 4.1 provides the "translation" of these requirements into OWL-specific constructs, which are needed to implement these requirements.

	OWL Constructs (Functional-Style Syntax)
OWL Constructors	`intersectionOf`
	`unionOf`
	`complementOf`
	`oneOf`
	`allValuesFrom`
	`someValuesFrom`
	`hasValue`
	`maxCardinality`
	`minCardinality`
	`inverseOf`
OWL Axioms	`subClassOf`
	`equivalentClass`
	`disjointWith`
	`subPropertyOf`
	`domain`
	`range`
	`FunctionalProperty`
	`SameIndividual`
	`DifferentIndividuals`

Table 4.1: OWL Class Descriptions and Axioms, Required in *OntoProfi*

OWL 2 has two variants: *OWL Full* and *OWL DL*. The latter, *OWL DL* is subdivided into three further *"profiles"*: *OWL EL* [(1)], *OWL QL* [(2)] and *OWL RL* [(3)]. These are 'trimmed down' versions of OWL, and were developed as a trade-off between expressive power and efficiency of reasoning (Motik, et al. 2012). Fig. 4.3 depicts OWL 2 variants.

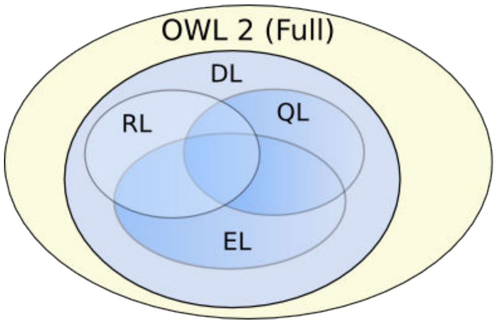


Figure 4.3: Venn Diagram of OWL 2 Syntactic Subsets
(Source: W3C OWL Working Group (2009))

In this project, the ***DL*** *'flavour'* of OWL was chosen over *OWL Full* because it is the *"DL"* subset of the "full" OWL that provides an ability to reason - reasoning over OWL Full is considered to be undecidable (Hitzler et al. 2012), and there are no reasoners currently in existence, which are able to reason over an ontology encoded in OWL Full.

The other aforementioned profiles were too restrictive for our project due to the limited semantics they support. These profiles were developed to offer different degrees of representation language expressivity (Hitzler et al. 2010) and to cover a variety of situations where the structure of an ontology makes reasoning inefficient or even impossible (i.e. to reduce the computational costs). *OWL EL* was designed as a language to represent ontologies with a very large class and properties hierarchies. *OWL QL* is aimed at data-driven applications that use ontologies with *"a very large volumes of instance data"* (Motik et al. 2012). OWL QL-based ontologies are used in

(1) `http://www.w3.org/TR/owl2-profiles/#OWL_2_EL`
(2) `http://www.w3.org/TR/owl2-profiles/#OWL_2_QL`
(3) `http://www.w3.org/TR/owl2-profiles/#OWL_2_RL`

the projects, which implement conjunctive query answering using conventional relational databases, and where answers to these queries is *"the most important reasoning task"* (Motik et al. 2012). *RL* targets the applications, which require connection to or migration from rule-based reasoning engines.

The three profiles have different restrictions imposed onto them. For example, all these profiles disallow the use of the universal quantification (except for OWL RL, which allows its use, but only on the object side). The use of negation is not available on the subject side for either of the profiles. It is available on the object side, but only for OWL RL and OWL QL. Neither is `owl:unionOf` construct permissible for these profiles, with exception of the subject side for OWL RL. (1) As specified earlier in this section (also cf. Table 4.1), all of these OWL constructs were required in our ontology, therefore, the three profiles were not suitable, and the "full" OWL DL was required in *OntoProfi*.

4.3.2 *OntoProfi* Class Hierarchy

Ontology engineers use three approaches to development of a class hierarchy: *top-down*, *bottom-up* and the combination of the two (i.e. *mixed*) (Noy & McGuinness 2001). In this project, we adopted the latter, i.e. the *mixed* approach to derive OntoProfi's classes and organise these into a hierarchy. Initial classes were identified (enumerated) inductively from our analysis of literature, commercial practices, and own experience of teaching and using e-commerce ("bottom-up"). Each possible candidate was checked for their suitability, using the project's aim, system requirements, the ontology's domain and scope as selection criteria. The enumerated classes were then grouped into a sub-domain, resulting in four sub-domains: ***customer, product, business*** (see Fig. 4.4). However, some classes, namely, the cus-

(1) It is outside of scope of this project to discuss the restrictions imposed on these profiles in depth. The interested reader is kindly referred to Hitzler et al. (2010, pp. 141-153) and (Krötzsch 2012) for more detailed explanation of restrictions and limitations of these subsets.

tomer profile, needed to be extended using the "top-down" approach, but this could only be undertaken after the analysis of our case study's data. This will be discussed in § 5.4.

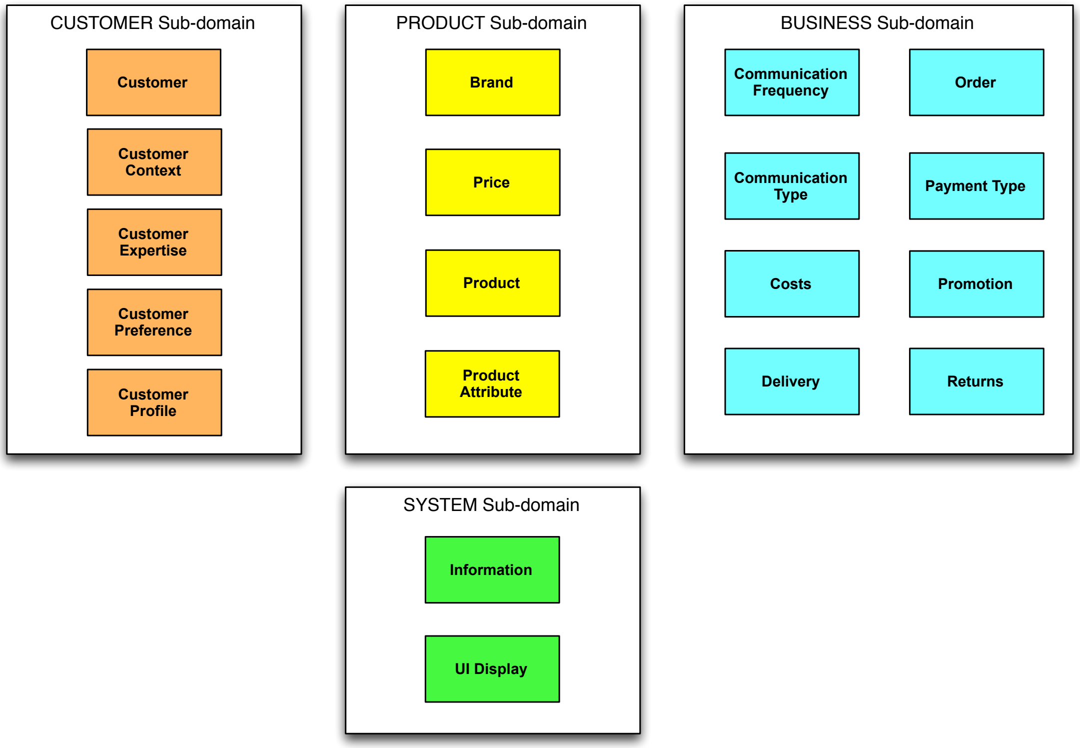


Figure 4.4: Top-Level Sub-Domains of *OntoProfi*

4.3.2.1 "Customer" Sub-domain

This sub-domain contains the classes necessary to model instances, related attributes and relationships of a customer. The classes belonging to this sub-domain play a pivotal role in facilitating user profiling, and can be used for personalisation of both, content and content presentation. The sub-domain's class hierarchy that includes associated subclasses is shown in Fig. 4.5.

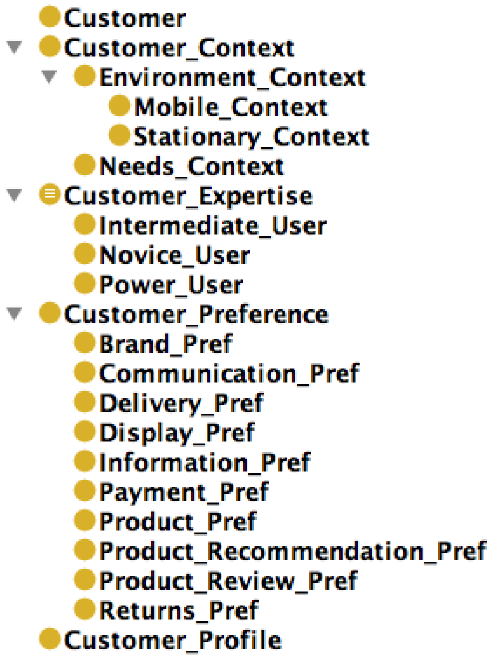


Figure 4.5: Class Hierarchy of **Customer-Related** Sub-Domain

The "Customer" sub-domain consists of five classes: `Customer`, `Customer_Context`, `Customer_Expertise`, `Customer_Preference` and `Customer_Profile`:

1. `Customer` subclass is needed to instantiate the company's customers and to allow specification of various relationship between a customer and other entities in the ontology in all four sub-domains. For example, we can assert the relationships within the "Customer-Related" sub-domain (e.g. between `Customer` and `Customer_Expertise`), or with the classes in other sub-domains (e.g. `Customer` and `Communication_Type` of the "Business-Related" sub-domain). This class can also be used to assert and infer various other customer's characteristics. For instance, we can define a relationship between a customer and a business rule. Suppose that we want to find out who the most profitable customers are so that they can be rewarded by an incentive such as a discount off their next purchase or some other promotion, and suppose that the company's rule states that a profitable customer is the customer, whose total spend to date is £1000.

The following rule can then be asserted in *OntoProfi* [1]:

$$\texttt{Most\_Profitable\_Customer} \equiv \geq \texttt{1000.00 orderTotal} \tag{4.1}$$

OntoProfi can then infer the individual customers (i.e. instantiations of `Customer`), who satisfy this rule, and offer them a reward. Other rules can be introduced to classify customers into various groups, e.g. based on the customer's shopping habits, interest, past shopping behaviour. The business rules can be specified in *OntoProfi*, or extracted from an external business rule system that the e-commerce company uses.

2. `Customer_Context` subclass allows the assertion and inference of various environmental and customer needs contexts in order to personalise content presentation. For instance, ubiquitous devices, such as tablets or mobile phones have screen sizes and resolutions different to those found on the *'stationary'* (desktop) devices, so the *'device context'* can be taken into consideration and the way the web page is displayed can be adapted to that particular context. The `Needs_Context` subclass allows expression of a customer's different needs, such as shopping for seasonal products or to accommodate customer's disabilities, so that content and content presentation is adapted accordingly.

3. `Customer_Expertise` class is used to classify customers into particular expertise categories. Here, *'expertise'* is considered to be technological proficiency in using computers (or other devices) and level of experience of shopping online.

   Inclusion of this class is due to the fact that research found that website usage and satisfaction depend on computer expertise. For instance, Yi Maggie & Klein (2009) observed that a user's skills level were very important for on-

(1) Due to its compactness and readability, DL syntax is used throughout to explain assertion of axioms, rules and restrictions.

line shopping tasks. Similarly, the importance of an e-shopper's self-efficacy beliefs was found to increase significantly as they acquire more experience in shopping online (Hernández-Ortega, et al. 2008). Web skills were also found to affect shopping enjoyment and concentration (Koufaris 2002).

Our ontology specifies 3 levels of expertise: `novice`, `intermediate` and `expert`. A customer's level of expertise can be determined in two ways: either through a user-driven or system-driven approach. The user-driven approach requires explicit input from the user to self-assess their level of expertise (e.g. by getting them to fill in a questionnaire). The system-driven approach works implicitly by monitoring and analysing a customer's online behaviour.

The user-driven method is easier to implement, but arguably, is more prone to issues such as users over-exaggerating their ability. It can also be seen as more intrusive since user input is required. Also, it is predominantly static and requires further input from the user if re-assessment of the expertise level is needed.

The system-driven approach is more costly in terms of extra effort required of system developers to implement it and will be more computationally intense, but it could prove to be more dynamic and yield more accurate and objective results.

4. `Customer_Preference` class provides the capability to assert and infer a variety of customer preferences, such as preferences for specific products, order fulfilment options, ways of communication, and what and how information is delivered.

5. `Customer_Profile` is, perhaps, the most important class in our ontology as it is used to classify each customer into a particular profile. This classification

is performed automatically by the reasoner, and is based on the asserted and inferred customer's characteristics and preferences.

Initially, `Customer_Profile` does not have any sub-classes. To define these, the types of profiles, which the customers can be segregated into, need to be discovered. In our project, this is done by analysing and synthesising the data collected during the case studies of online shoppers. The discovered customer profiles are then used to extend the `Customer_Profile` class. The details of how this is achieved is in § 5.4.

4.3.2.2 "Product" Sub-domain

The ***product-related*** sub-domain contains four classes `Brand`, `Price`, `Product` and `Product_Attributes`, and is shown in Fig. 4.6.

Figure 4.6: Class Hierarchy of **Product-Related** Sub-Domain

These classes are needed and primarily used for personalisation of content (i.e. product recommendations), based on the preferences specified by a customer, or on business rules defined by a company. For example, a customer may specify that she prefers a particular product brand and has a particular budget. The system will then find products that match these criteria.

In addition, the products can be described and classified using various rules. For example, we can distinguish between cheap and quality (e.g. expensive) products by defining them as low- or high-priced:

$$\texttt{CheapProduct} \equiv \texttt{Product\_Type} \sqcap (\exists \texttt{hasPrice.Low\_Price}) \tag{4.2}$$

$$\texttt{QualityProduct} \equiv \texttt{Product\_Type} \sqcap (\exists \texttt{hasPrice.High\_Price}) \tag{4.3}$$

Suppose that the company knows that their customers regard a low price as less than or equal to £20 and a high price as more than or equal to £100. Using the `price` datatype property restriction, the `Low_Price` and `High_Price` classes can be defined as:

$$\texttt{Low\_Price} \equiv \leq \texttt{20.00 price} \tag{4.4}$$

$$\texttt{High\_Price} \equiv \geq \texttt{100.00 price} \tag{4.5}$$

A `Medium_Price` can be asserted with the datatype property `price`, whose value lies between the cheap and high price.

The `Reasonable_Price` *OntoProfi*'s class is an example of the use of the equivalent class restriction that can help with ensuring semantic interoperability, e.g. when aligning or merging *OntoProfi* with other ontologies.

All of the above restrictions will allow the reasoner to infer the products that satisfy these rules.

The `Product_Attribute` and its subclasses (`Bulky`, `HighCost`, `LowCost`, `Perishable`) are also examples of how the product-related domain can be modelled and used for e-commerce personalisation. These can be extended and modified to reflect a particular company's type, strategy, business rules and types of products the company sells. Additionally, the classes of this sub-domain can be used for federation of data from other sources, for example, by using external data from other companies' ontologies, such as populated *eClassOWL* (1) or *GoodRelations* (2) ontologies.

4.3.2.3 "Business" Sub-domain

The ***business-related*** sub-domain is used for personalisation of marketing and order fulfilment activities. The sub-domain is represented by the classes and sub-classes shown in Fig. 4.7.

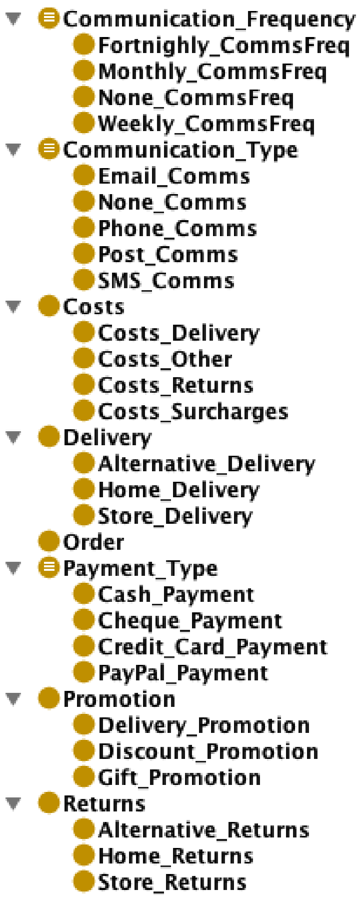


Figure 4.7: Class Hierarchy of **Business-Related** Sub-Domain

(1) `http://www.heppnetz.de/projects/eclassowl/`
(2) `http://www.heppnetz.de/projects/goodrelations/`

The classes within this sub-domain are used to model various activities and features to represent an e-commerce business marketing strategy and tactics, such as order fulfilment, payment, communication and promotional activities. For example, e-commerce customers may want to specify preferences for how they receive or return their products (e.g. *"home"* or *"store"* - an option, which some *'click-and-mortar'* EC businesses with high street shop presence currently offer); how they would like to pay for their orders (e.g. credit card, PayPal, cheque); and how they receive marketing communication (if any). The other classes in this sub-domain are also used to specify various business rules, for instance, personalised promotional offerings (such as discounts, gift and order promotions). These classes can then be linked with the `Customer` and `Customer_Profile` classes.

4.3.2.4 "System" Sub-domain

The ***system-related*** sub-domain is used for enabling personalisation of content presentation, and comprises the `Information` and `UI_Display` top level classes (see Fig. 4.8).

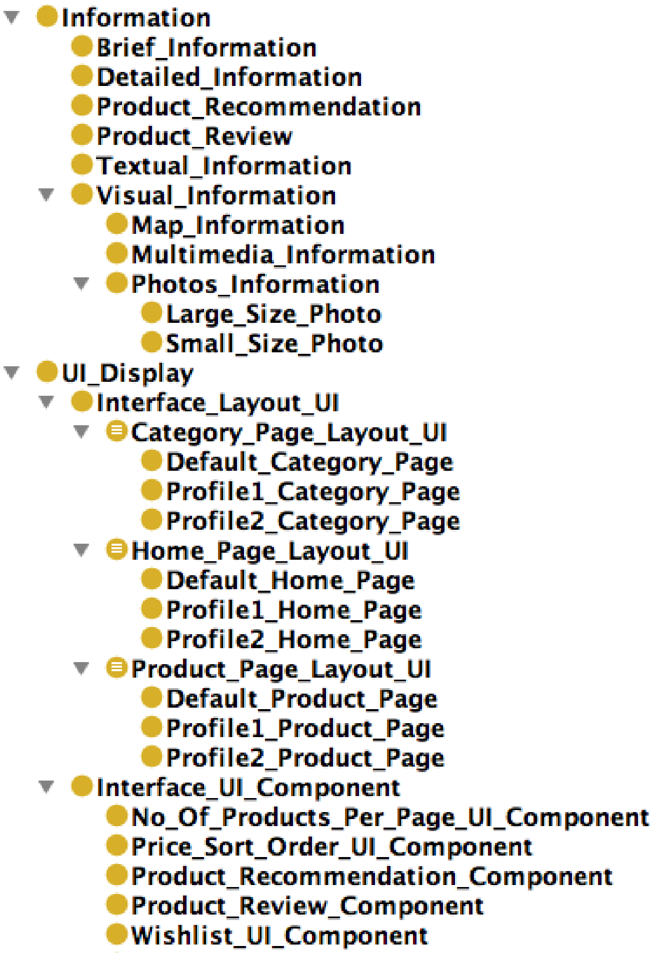


Figure 4.8: Class Hierarchy of **System-Related** Sub-Domain

`Information` is used for modelling different types and amounts of information displayed on a website, such as *"textual"* vs. *"visual"*, *"brief"* vs *"detailed"*, and product recommendations and reviews. `UI_Display` class encodes various elements of the user interface found on a commercial website, such as different types of web pages (including examples of personalised pages - prefixed with "`Profile1_`" and "`Profile2_`"), position and choice of menus, and the product sorting order.

4.3.3 *OntoProfi* Object and Datatype Properties

To express relationships between the classes and individuals, *OntoProfi* uses a number of object and datatype properties (see Fig. 4.9).

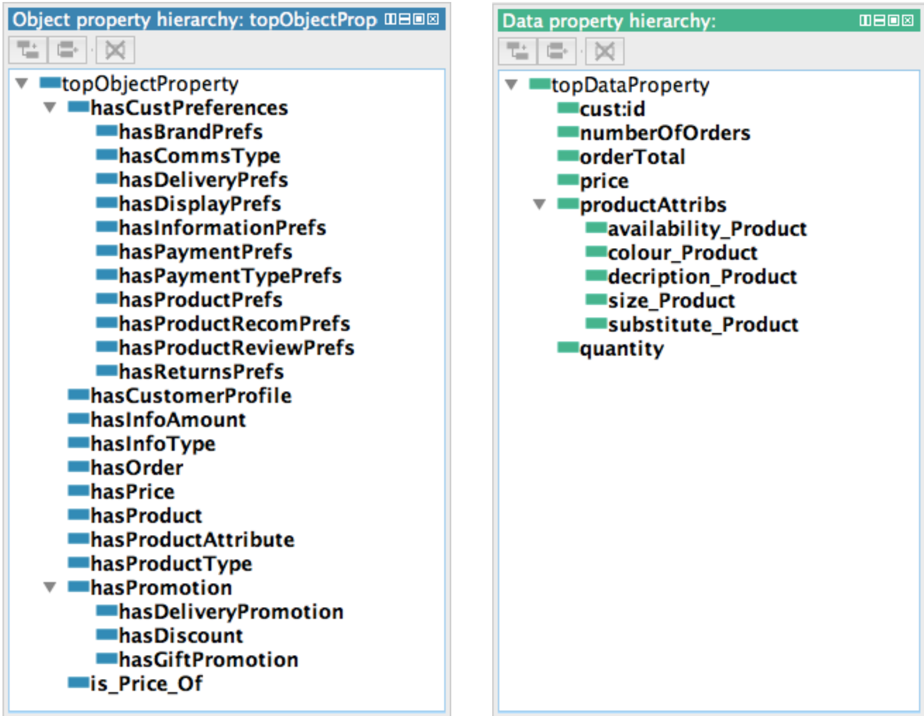


Figure 4.9: Object and Datatype Properties

Fig. 4.10 presents a diagram showing the key high-level relationships in *OntoProfi*. (The rectangles marked with an orange dot represent the *OntoProfi*'s classes; the blue rectangle symbol before a property name represents object properties, and the green rectangle symbol - datatype properties.)

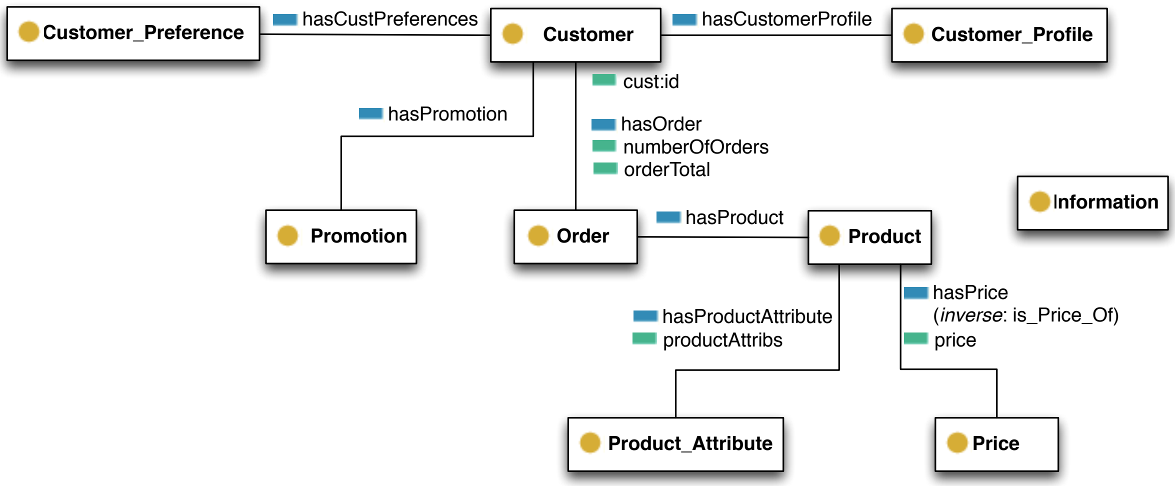


Figure 4.10: Key Relationships in *OntoProfi* (high-level)

Every `Customer` is identified by their unique `cust:id` functional datatype property, and has their unique `Customer_Profile` and `Customer_Preference`(s), which are linked by the `hasCustomerProfile` and `hasCustPreferences` object properties. The `Customer` can express their preferences for particular brands, communication type, delivery, information, etc. These preferences are asserted by using the related sub-properties of the `hasCustPreferences` top object property (i.e. `hasBrandPrefs`, `hasCommsPrefs`, `hasDeliveryPrefs`, `hasInformationPrefs`, etc. - see Fig. 4.9).

The object and datatype properties of `hasPromotion`, `hasOrder`, `hasProduct`, `numberOfOrders`, `orderTotal` can be used for assertion and inference of various business rules, e.g. identification of the customers, who can be given incentives (promotional discounts, gifts and other personalised offers). `hasProductAttribute`, `hasPrice`, `productAttribs` and `price` are used for enabling product recommendations.

4.4 Chapter Summary

This chapter presented *OntoProfi*, a customer profiling ontology, which enables website personalisation in *PERSONTO*. The chapter explained the ontology engineering methodology used for creating *OntoProfi*, discussed the reasons behind the choice of OWL DL as a language for *OntoProfi* representation, and discussed *OntoProfi*'s class hierarchy and object and datatype properties. Building on the presented class hierarchy, the ontology needs to be modified (extended) to include additional classes and related properties to represent specific customer profiles as the basis for classifying the customers into groups. The elicitation of such profiles is based on the multi-case case study of online shoppers, described in the next chapter.

Chapter 5

Elicitation of Customer Profiles

5.1 Introduction

The *OntoProfi* ontology, presented in the previous chapter, provides the basis for enabling personalisation. However, it needs to be extended with additional models to represent specific customer profile types. To elicit these types, a multiple case study approach is used in this project.

This chapter discusses the process and outcomes of our multi-case case study of online shoppers, and in particular it:

- explains and discusses the case study research methodology that is used for data collection, data analysis and conceptualisation (including a discussion of the limitations of the chosen approaches);
- discusses the findings of common and differentiating factors that affect online shoppers;
- analyses which of these factors can be used to personalise e-commerce websites; and proposes a set of e-shoppers profile entities for our customer profiling

ontology, *OntoProfi*;

- explains how *OntoProfi* is extended to incorporate the identified profiles and to enable customer profiling in *PERSONTO*.

5.2 Case Study Research Methodology

Regardless of the type of research strategy (quantitative, qualitative or mixed), every part of a research project always follows a particular process that consists of several distinct phases. Devising and following appropriate phases of this process is typically considered to be important for acceptability and credibility of research (Oates 2006).

In this project, the research process for our case study followed a research cycle, which consisted of four phases:

- research design (covered in detail in § 5.2.1);
- data collection (§ 5.2.2);
- data analysis (§ 5.2.3);
- generation of outcomes (§ 5.3).

Fig. 5.1 represents an overview of our case study research process:

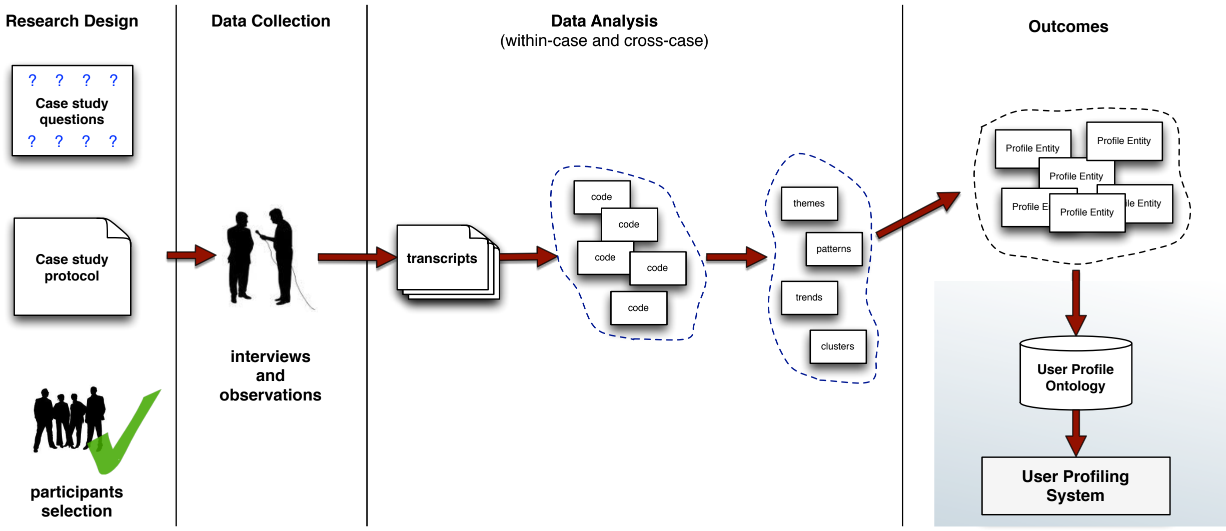


Figure 5.1: Overview of Online Shoppers Case Study Research Process

Each research phase generated particular outcomes. For example, during the *"research design"* phase, case study questions, protocol and participant selection were undertaken. The remainder of this section describes and discusses each of the research phases and sub-phases.

5.2.1 Research Design Phase

Research design is considered to be a *'logical plan'* (or *'blueprint'*) for research, which helps to ensure that collected evidence addresses the posed research objectives or questions (Yin 2003).

Our "Research Design" phase involved three activities:

- formulation of **case study questions**, which specified the main topics for our inquiry;
- devising a **case study protocol**, which specified the procedures and general rules for conducting our case study;

- **selecting participants** to provide the case study data.

Case Study Methods, Focus and Questions

Our examination of web personalisation of professional practices (§ 2.3) and later the prototype (§ 6.2.4) focused on "Clothes" e-commerce category due to its prominence. However, in our case study we needed to avoid bias and to collect as rich data as possible. Therefore, it was decided to explore other e-commerce categories, such as "Film, music", "Holiday accommodation", "Household goods", etc. (see Appendix C).

To collect data, we used face-to-face semi-structured interviews and "verbal protocol analysis" observations. Other methods, such as a "diary study" and postal surveys, were considered, but were determined to be inappropriate to our research due to worse response rates, lack of an opportunity to further probe into the participants' responses, and the fact that it is not common in professional practice ([Pers.Comm.2], [Pers.Comm.3]).

Our case study questionnaire was developed based on the analysis of the project's aim and objectives (see § 1.2). More specifically, we identified that answers to the following questions would provide the necessary insights into online shoppers' behaviour, motivating factors, and enable us to discover the types e-shoppers should be profiled into.

- *What are the motivating and demotivating factors that affect e-commerce shoppers?*
- *Are there any common and differentiating factors and aspects, which have an effect on e-commerce shoppers?*
- *Can any of these factors/ aspects be personalised?*

- *Can we segregate e-commerce shoppers into particular groups or types? And if so, what groups or types?*

Case Study Protocol

Case study protocols are generally seen as a *"major way of increasing the **reliability** of case study research"* (Yin 2003) (emphasis added), so that a project's quality and rigour can be ensured. Thus, from the research perspective, our case study protocol was developed to make sure that data collection was performed accurately, efficiently and effectively. From the perspective of application or dissemination of research in practice, we wanted to develop a procedure that can be re-used in subsequent projects - as system development and implementation using a reusable approach was a part of our research aim § 1.2). For instance, the presented protocol can be used by an e-commerce company or web-developers, who would be able to adapt our protocol to fit their particular situation or needs. A copy of the protocol is in Appendix A.

The developed case study protocol specified the processes and activities needing to be carried out to complete our case study. In particular, it has instructions on the order of activities when interviewing and observing the participants. Thus, each interview started with a *"research professionalism and ethics preamble"* to affirm we were abiding by Cardiff University's Code of Conduct for researchers and explained what it meant (this included assurance of providing participants' anonymity).

Following this, the participants were asked to fill in the *"Screening"* and *"Products"* questionnaires (see Appendix B for a copy of these questionnaires). The *"Screening"* questionnaire was used to ensure that all participants had shopped online before, i.e. they were already e-shoppers, which was our case study participants selection requirement. The questionnaire was also used to get some additional quantitative data on the participants' online shopping habits, such as: how much time each respondent typically spends online; how often they do online shopping; how many products they typically buy within a certain period of time.

The *"Products"* questionnaire was used to get information on the types of products and how often these are bought (*"often bought"*, *"seldom bought"*, *"never bought/ never would buy"*). The list of the chosen products was derived from the UK Office for National Statistics (ONS) (2010) report, and the choice of products was limited to the first 16 most commonly bought products online. These products covered a wide range of types of goods most frequently bought on the Internet, and included: frequent and infrequent purchases; tangible and intangible; durable (non-perishable) and non-durable (perishable); high outlay (high risk) and low outlay (low risk); high and low differentiation; high and low complexity; as well as search, experience and credence product types.

On completion of both questionnaires, the case study proceeded to collection of the main data, as specified in the remainder of the case study protocol (starting with the *"Initial Questions"* section of the protocol, see Appendix A). The main data came from two sources: interviews and observations. The interviews took the form of semi-structured interviews. The observations followed the *"verbal protocol analysis"* (Oates 2006), whose purpose was to get the participants to "think aloud" so that they could describe, explain and justify the decisions they make when they interact with e-commerce websites to search for and buy products. This process of data collection during the interviews and observations is explained in more detail in § 5.2.2.

Selection of Participants

Qualitative research, such as the case study, typically uses purposive sampling, i.e. non-random selection of participants (Creswell 2006), (Miles & Huberman 1994).

Purposeful sampling is further divided into a number of strategies, such as: maximum variation, homogeneous, critical case, theory based, etc.- see (Miles & Huberman 1994, p. 27-28).

Purposive sampling has several advantages and limitations:

- This type of sampling is common and has been widely accepted as appropriate for qualitative and mixed research approaches (e.g. (Lincoln & Guba 1985), (Miles & Huberman 1994), (Marshall 1996), (Onwuegbuzie & Leech 2007)). Moreover, as Eisenhardt (1989) argues, in case study research *"random selection* [of case study participants] *is neither necessary, nor even preferable"*.
- Using purposive sampling is advocated to provide *"thick descriptions"*, data richness and holism (Miles & Huberman 1994), which are needed in order to explain the complexity of a phenomenon investigated in case studies. Convenience sampling, as a sub-type of purposive sampling, is seen as a particularly appropriate method that guarantees production of rich data (Koerber & McMichael 2008).
- Purposive, non-random sampling involves less time, effort and costs compared with probabilistic random sampling due to the need in the latter to involve a significant number of respondents.
- However, collecting data using purposive sampling has a danger of introducing bias (Miles & Huberman 1994, pp. 9-10). It can also hinder generalisation and replication of research findings. These limitations and issues are common in qualitative research in general, not just in sampling strategy, and require extra care and attention on the researcher's part.

In our project, we used a *"mixed purposeful"* sampling approach (Miles & Huberman 1994, pp. 27-28), combining *"convenience sampling"* (when the first few participants were selected based on their suitability, availability and willingness to participate in the study) with a *"snowball sampling"* (when the already interviewed subjects recommended other people, who they thought would be of particular interest to our case study due to their particular experiences with online shopping, e.g. people, who buy a large number and variety of products). This data collection strategy was deemed appropriate and acceptable for our work for two reasons:

Firstly, because of the strategy's commonality and wide acceptance as appropriate (1).

Secondly, because of its common use in the industry. It appears to be a typical practice that companies involved in web system development use non-random, *"convenient"* selection of respondents for design and testing of their commercial projects ([Pers.Comm.2] and [Pers.Comm.3]).

The chosen participants selection approach has some limitations, for example, potential to introduce bias and issues with generalisability of the research outcomes. Mitigation of these limitations will be discussed in more detail in § 7.7.

5.2.2 Data Collection Phase

The main empirical data for our case study was gathered from semi-structured interviews and interactive *"think aloud"* observations of the selected participants. The interviews were combined with observations in order to gather as rich and holistic data as possible. The combination of two data collection methods was chosen so that they complemented each other: the interviews allowed us to explore the topics of interest in depth, while observations gave us an opportunity to gather additional insights and evidence of online shopper's behaviour. Data collection was carried out

(1) For instance, much research in social science uses non-random sampling techniques (Cena et al. 2012).

in the *'field'* - at participant's homes or places of work, and each interview and observation was recorded, so that they could be subsequently transcribed and used in the data analysis phase of our project (see § 5.2.3).

The process of data collection followed the case study protocol that was developed in the early stage of the research design (see § 5.2.1 and Appendix A). Interviewing and observations commenced after the completion of the Protocol's *"Before the Interview"* tasks 1) through to 5) (see Appendix A).

Each **interview** started with a *'warm up'* opening question or questions, for example, *"What makes you buy on the Internet?"* or *"What do you think about buying on the Internet?"*. The opening questions were used to establish a rapport between the interviewee and interviewer and put an interviewee at ease if that was necessary. After such *'warm up'* the participants were given freedom in what they wanted to say or show, with only minimum intrusion or interruption from the interviewer. However, as each interview progressed, interesting areas that the participants talked about started to emerge. In these instances, the interviewees were asked additional questions so that these areas could be explored in more depth. For example, if a participant mentioned a particular website that they use on a regular basis, they would be probed further about the reasons for this.

The interview was followed by an **observation** of each participant, in which they were asked to shop for certain products. During the observation, the participants needed: to decide which website was appropriate for each product; access that website; search for these products; and 'pretend' to buy them. Throughout the interaction with a website, the participants were asked to "think aloud" to explain the reasons behind their actions and to describe their decision making.

Similar to the interviews, all observations were recorded as *"screen recordings"* that captured e-shoppers keyboard and mouse activities as well as their vocal explanations of what they were doing and why.

For the first three observations, the participants were asked to imagine that they would need to buy products from each of 16 product categories listed in our *"Products Questionnaire"* (Appendix B). For this task, we compiled a list of commonly bought or best-selling products in each category(1) (see Appendix C). However, after the first three observations it became clear that some of the data, we collected during these observations, was repetitive and therefore redundant. So, a change in observation protocol was needed.

After evaluation of the outcomes from the first three observations, we decided that it would be more appropriate to adapt the protocol by reducing the number of products the participants had to search for and concentrating on asking them to show us how they buy their regular products (if they had any), their last few products they bought and how they would shop ('theoretically') for a product they would never buy. This change allowed us to focus on each particular user by exploring their shopping habits when searching for familiar products (as opposed to the ones we chose for them) as well as to reducing the fatigue from providing repetitive and redundant data. We do not believe that this change in the protocol biased the results in any way or affected later interviews, since the tasks the participants had to complete remained the same, it was only the amount and choice of products that changed. In fact, a change in case study protocol is legitimate, common, beneficial and often even necessary for the research to reflect upon the success (or otherwise) of the chosen, predetermined *"prior instrumentation"* and adapt the research protocol accordingly (Eisenhardt 1989), (Miles & Huberman 1994, pp. 34 - 38).

(1) Best-selling products as of June 2011.

Data Sources (Participants)

Data was collected over the period from June 2010 to December 2011. Interviews and observations lasted between approximately 25 minutes (*minimum duration*) and 76 minutes (*maximum duration*), with average duration of approximately 45 minutes.

The total number of respondents, who participated in our case study, was ten: seven females and three males. Table 5.1 shows the breakdown of the participants per particular age group (1):

Age Group	16 - 24	25 - 44	45 - 54	55 - 64	65+
No. of participants	0	7	1	2	0

Table 5.1: Number of Case Study Respondents *(per age group)*

All participants were *'established'* B2C online shoppers, and recently bought something on the Internet. Their general 'web surfing' durations ranged from minimum of 0.5 hours to a maximum of 7 - 8 hours a day. The time spent on online shopping was estimated by the respondents as being between 1 and 4 hours a week. The majority of respondents (9 out of 10) *'window shop'* (i.e. browse e-commerce websites without actually purchasing products) on a regular basis.

The respondents online shopping pattern varied: 60% (6 out of 10) of respondents said that they buy something online a couple of times a month, 20% (2 out of 10) - once or twice a week, and the rest (20%) less often - once every 3 - 6 months. 6 out of the 9 respondents(2) shop online at a particular time of day.

There are some research limitations due to the number and choice of participants. This will be discussed in § 7.7.

(1) Age groups used here are based on the age groups used in (UK Office for National Statistics (ONS) 2010)

(2) 1 respondent didn't answer this question.

5.2.3 Data Analysis Phase

The purpose of this phase is to code the gathered data by extraction and inductive conceptualisation (inference) of the common themes, clusters and patterns found in the data. The phase involved a particular process and a strategy for data coding and conceptualisation.

5.2.3.1 Data Analysis Process

On completion of each interview and observation, the associated recordings were transcribed verbatim, and the generated transcripts were used to analyse the gathered data. We employed a multi-pass, iterative reading and analysis of each transcript. The first pass through the data allowed us to get familiar with the data and check the accuracy of transcription. Each subsequent pass served three purposes:

1. *Data consistency and completeness*:

   This phase involves checking the collected data for consistency and completeness. This process helps with ensuring that no opportunities for exploring particularly interesting issue or phenomenon were missed if these were presented in the earlier data. If this was the case, additional data was collected in subsequent interviews and observation.

2. *Initial, within-case level data coding*:

   This pass involves getting more familiar with, and *closer* to the data. The collected data is summarised for each participant, i.e. initial, first-level codes are generated for each case study (i.e. a single participant). Thus, initial themes, trends and patterns start to emerge, but these are still at a single case level.

3. *Cross-case pattern coding*:

This pass involves a *"data with data"* comparative analysis, a technique commonly employed by qualitative researchers (e.g. in Grounded Theory Methodology (GTM), see (Charmaz 2006)) to generate inductively derived concepts, categories, attributes and patterns at a more generalised/ conceptual level. During this stage we searched for generalised themes, trends, clusters and patterns (*"meta-codes"* (Miles & Huberman 1994)) across all participants' data (i.e. cross-case analysis), and this stage resulted in our main findings - see § 5.3.

These *passes* were not executed in a linear fashion. The data was analysed as it arrived, i.e. on completion of each interview and observation. So, it was inevitable that there was an iteration between the three passes to re-evaluate the interim findings from additional data and appropriately adapt the data collection and synthesis. Additionally, memos were written throughout the *data analysis* phase to aid the data summary and synthesis, and more specifically to help us *"tie together different pieces of data into a recognisable cluster"* and *"to show that those data are instances of a general concept"* (Miles & Huberman 1994, p. 72).

5.2.3.2 Data Coding and Conceptualisation

The main data came from interviewing and observing the participants. The interviews and observations were open-ended and *"intensive"* [(1)] so that participants' experiences and attitudes would be explored - in depth and with only a minimal guidance or interference from the researcher. The participants were not asked direct, closed-ended questions that required "yes" or "no" answer (e.g. *"Do you find e-commerce easy to use?"*). Instead, they were asked indirect, open-ended questions, for instance, '*'What do you think about using e-commerce?"*. In other words, the

(1) "Intensive" interviews refer to an in-depth exploration of particular topics or experiences that researches investigate (Charmaz 2006, pp. 25-27).

participants were invited to provide their own description and interpretation of their thoughts, beliefs, feelings and attitudes. Using indirect questions and minimising any interviewer's guidance and interference was deemed necessary to minimise the interviewer's influence that might have otherwise made the interviewees' responses biased or primed.

Each interview and observation was treated and analysed as a single, separate case study (i.e. a multi-case case study approach). The resulting case study data analysis has led to the formulation of common and differentiating factors and aspects that affect online shoppers (see § 5.2.4). The actual conceptualisation of these factors and aspects was based on the assumption that respondents would mostly recall and bring the information out that they felt strongly about and/ or they thought was important to mention; and similarly, it was assumed that if something did not get a mention it was not of importance to our respondents. More specifically, during the data analysis, the following rules were used to conceptualise the common and differentiating factors and aspects of online shopping:

- *Common (typical) factors and aspects* - motivators and de-motivators (see § 5.3.1):
  - If two or more respondents mentioned a factor or aspect and there were no contradicting mentions of the same or similar factor or aspect. Thus, positive mentions were classified as *"motivating factors and aspects"*, while negative mentions were *"de-motivators"*.
- *Differentiating factors and aspects* (§ 5.3.2):
  - These were classified as *"differentiating"* similarly to the classification of *"common"* factors, i.e. if there were at least two people on either side of a characteristic, factor or aspect of online shopping.

Overall, our data coding and conceptualisation relied on the premise that interviewees talk about things that they want to tell (Atkinson 2002), and that the things they talk about are important or prominent to them.

Similar to the research limitations of the selection and number of participants, the chosen approach to data coding and conceptualisation has some limitations that needs to be addressed (see § 7.7).

5.2.4 Case Study Outcomes Phase

This was the last, concluding phase of our case study. The case study findings were based on the outcomes of the previous research phase (*"Data Analysis"*), and are discussed in the next section (§ 5.3), where we describe and discuss generalised characteristics, aspects and factors of e-shopping.

5.3 Characteristics of Online Shopping: Case Study Findings

Our within-case and cross-case analysis of the collected data revealed several notable characteristics, factors and aspects that the case study respondents reported as being influential in their online shopping decision making. We have grouped these into two top-level categories:

- common factors and aspects that affect e-shoppers' decision to adopt, use and re-use e-commerce - see § 5.3.1;
- differentiating factors and aspects that distinguish e-commerce buyers - see § 5.3.2.

Figure 5.2: Characteristics of Online Shopping

The first part of this section discusses these findings, making use of direct quotations from the transcripts of interviews with the case study participants to illustrate and support our arguments and conclusions.(1) The second part of the section explains how we arrived at the proposed e-shoppers' profiles to be used for personalisation of e-commerce systems and experience.

5.3.1 Common (Typical) Factors and Aspects for Online Shopping

The common (or typical) factors and aspects give us insights into what influences shoppers' decision to adopt, use and re-use e-commerce. Our analysis of the gathered data shows that when it comes to making online use and purchase decisions, all shoppers are affected by particular **motivating** and **de-motivating** factors. In addition, there is a strong evidence in our data that e-shoppers exhibit other common characteristics of being **highly focused**, **habitual** and **non-spontaneous**. Fig. 5.3 shows these common factors:

(1) Please note that the interviewees' grammar in quotations was preserved. However, in places where the interviewees accidentally missed some words and the meaning of the quotation was not clear, additional, explanatory words were added in square brackets to add clarity to the quote.

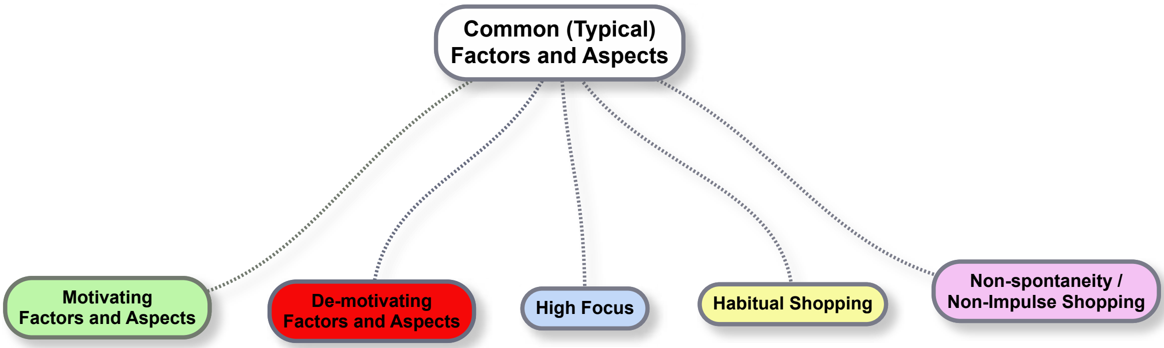


Figure 5.3: Common (Typical) Factors and Aspects for Online Shopping

5.3.1.1 Motivating Factors for Online Shopping

The motivators for shopping online are the factors, which our case study participants stated as important for their decision to adopt and use e-commerce, and subsequently make online purchases.

Fig. 5.4 depicts the identified factors that motivate people's decision to use e-commerce:

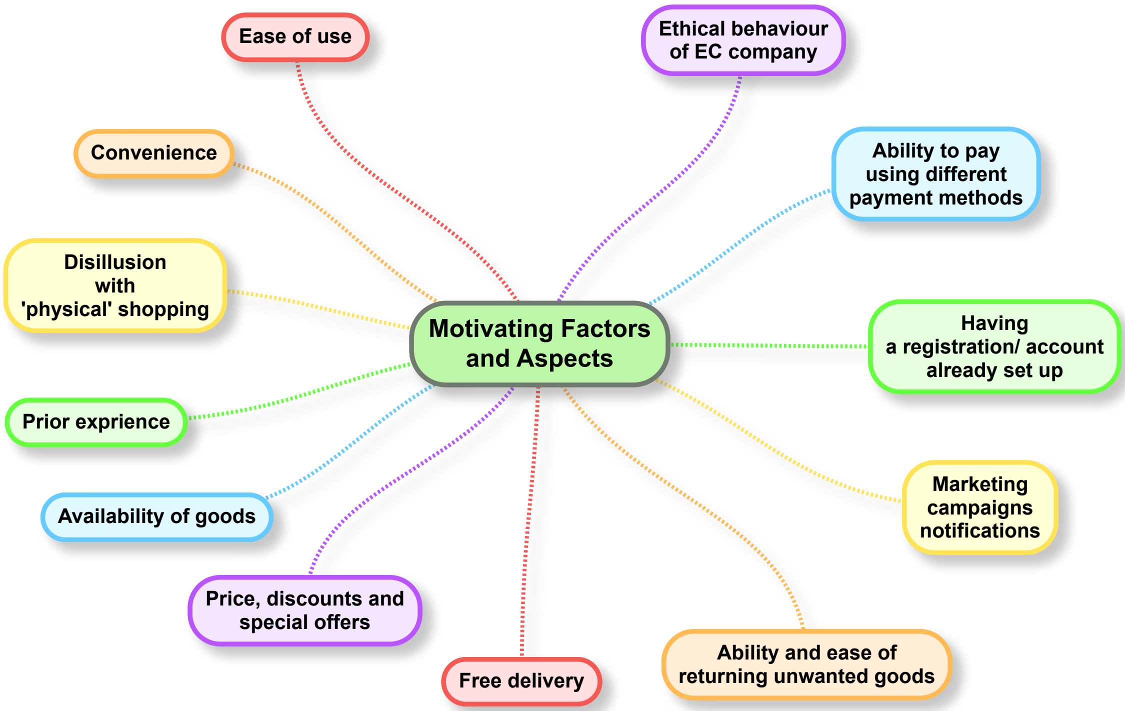


Figure 5.4: Motivating Factors and Aspects for Online Shopping

The analysis of our data showed that *"ease of use"* and *"convenience"* were the most prominent general factors and reasons why our case study participants chose to use e-commerce. For example:

> *"It's much easier. You can see everything..."*

> *"... and you can do it [shopping] without having to remember to do it later... you don't have to worry about writing it down somewhere and then thinking: 'I'm going to buy it one day!' You can actually do it there and then."*

> *"I don't have time to go to the shops ... physically, and go look for things.*

It's much easier. You can see everything ... you can find straight away, you don't need to go to five shops and look for things. You can just go [on the Internet] and straight away find it."

"... 'cause you don't want to carry them [purchases] home..."

The *"ease"* and *"convenience"* factors in our data were often linked to another prominent factor of *"disillusion with 'physical' shopping"* at brick-and-mortar stores, which some participants 'blame' as the reason for their adoption of EC:

[with e-commerce] "There is no need to go to shops physically."

"... I don't have to faff around, look around the store and wait in the queue. And it gets through the front door, and it's all very handy."

"I am getting really fed up with going shopping. Really! ... So, I'm becoming more of a shopper online now. ... and I've actually had better service online that I had from shops. ... And less stressful because, you know, I don't have to go to the shop and pay 5 quid for parking."

Additionally, some people feel that, compared with a physical store, e-commerce does not impose the same *amount of stress* on making them buy something, where they feel they have to make their decision there and then and quickly or face the situation that they would need to come back to the store.

"Prior experience" is another motivating factor that we identified as the one that affects shoppers' decisions to re-use e-commerce and make online purchases. Our case study interviews and observations showed that shoppers were more likely to

use and trust an already known website, where they shopped previously, and were satisfied with their past purchases:

> *"... and because I've had some good experience, I feel more confident doing it [buying online]."*

> *"Because I trust it [the website] ... and you know you are going to get it fairly quickly and, yeah, just from using them before, so you know what to expect and you know the cost is going to be right without having to compare."*

> *"I shopped with them previously and, you know, when the item comes, it's exactly what you've seen [on the website] and it's always beautifully packaged and its always in tip-top condition. ... It's [delivery] always been really, really quick."*

"Availability of goods" factor also featured prominently as one of the reasons the respondents chose e-commerce. This factor relates to a shopper's ability to purchase products that are available online but not available anywhere else, and in particular, in the high street shops:

> *"You go to town and they haven't got what you want or you've got to wait for it or you've got to order it..."*

> *"... I know that I need a book and I can't get it from the library. ... And I know that the bookshops are getting less in town, there are not as good, not half as good. So, I couldn't get any of my reference books there."*

"I couldn't find it [required product] in town. I looked at Cardiff, Bristol, Bath and those places, all nearby town. And this [online] was just so much more convenient and what I wanted; it was available."

"...if I've seen something in a magazine and I can't buy them in a shop - I bought it off the Internet."

The *"availability"* factor also refers to an ability to buy specialist, niche or rare products, which local brick-and-mortar stores do not stock.

"Price", *"discounts"* and *"special offers"* were identified as yet other factors that play an important role in motivating people to shop online. Shoppers choose to make their purchases online because, compared with brick-and-mortar companies, they consider e-commerce companies as more competitive and more likely to be able to offer their customers 'good' price, discounts and special offers:

"much, much cheaper than buying in shops..."

"Everything I google is often discounted."

"So, I will see it in a shop and buy it on the Internet, 'cause I know it will be cheaper."

"What I look for is for what is on offer at that particular time first ..."

"I will be more motivated to look for something what I need on the sale. ... And the motivation mostly because it was a sale and I needed it. That's all."

> *"So, the sale would come on. They send me flyers and discount cards and they do give me information, you know, email me stuff and tell me when [the sale is on]. And that's how I buy from them. Exactly the same, really."*

"Free delivery" option was the most prominent *"special offer"* factor mentioned by our respondents, e.g.:

> *"If you get free delivery, that would encourage me to purchase something."*

> *"... and I always go for the one [company] that does the free delivery."*

There were also other motivating factors that we discovered in the course of the case study; these were: *"ability and ease of being able to return unwanted goods"*; *"notifications about marketing campaigns"*, such as emails with promotional offers (*"because I get stuff* [promotional emails] *sent to me")*; *"having a registration/ account already being set up"*; and *"ability to pay using different payment methods"* (e.g. PayPal, which one of our respondents saw as safer - *"because you don't hand over any details"* and more convenient - *"you don't have to type any credit card details"* - compared with credit cards).

It is also interesting to note another factor that influences the *'to buy or not to buy'* decision, mentioned by one participant - the e-commerce company's *"ethical behaviour"* factor.(1)

(1) Despite the fact that this was a motivating factor that was only found in one interview, due to its novelty we decided to include it here along with other, more common factors. To the best of our knowledge, there has been no research on examining the role and importance of shoppers' ethical values, beliefs and opinions on their shopping behaviour and patterns, and it might be of interest to researchers to examine this factor.

5.3.1.2 De-Motivating Factors for Online Shopping

Online shoppers also have a number of factors that deter them from buying online. Fig. 5.5 shows factors we have found to be de-motivating in the online shopping environment:

Figure 5.5: De-Motivating Factors and Aspects for Online Shopping

"Not having any need" to shop for particular goods and/ or on particular websites is perhaps the most 'obvious' demotivating factor. For instance, people with children would stop buying certain goods (e.g. toys) because their children had grown up.

There are also cases when shoppers have to *"consider other people"* (e.g. in their household) while making their decision to buy or not to buy online. For example, one participant explained that she has to take her partner's habits and wishes into consideration: her partner, in her own words, is *"old fashioned" and "doesn't use a* *laptop"* and prefers to buy certain products and services using 'traditional' means, such as in person or through talking to a *'human'* company representative or by post:

> *"So, no, we have to do everything on paper."*

Shopper' beliefs that a *non-online company may offer a better deal* is another barrier to the use e-commerce. Some shoppers use e-commerce websites just to do their research and then buy products or services using 'non-online' means, e.g. through telephoning the companies or visiting them in person:

> *"I think you often get a better price verbally than on Internet."* [1]

One of the major, overarching de-motivating factors is the shoppers' general *preference for physical stores*:

> *"I'm not your normal Internet person, really, because I tend to go to shops, look around the shop, feel the product. I like to feel things and really into fabrics."*

> *"I always like going [to shops], look at different foods to see what's there. I just like going into supermarket."*

This preference for 'physical' shopping, however, often applies only to certain types of products, but the type of product, which this preference applies to, differ among shoppers. (This differentiation will be discussed in more detail in § 5.3.2.)

The preference for physical stores often stems from *not having a physical contact* with certain types of product, which, in the words of one of our participants, *"need to be seen"*. This factor is another barrier to online shopping:

(1) It is interesting to note here that often these are the same companies the shopper previously researched online.

> *"I like to look at stuff on there [on the Internet], but I wouldn't necessarily buy it. I'd probably go into town and have a look at it."*

> *"... you have to try it, you always have to. It's stuff that you don't know until you see. So, I'm a bit 'old school'. I like going out and see stuff and actually buying it rather than using the Net too much."*

The *lack of physical contact* is related to another demotivating factor - the issue of *needing to return goods*. Some people try to avoid having to send goods back by either not buying certain items or only buying when a shopper is absolutely sure about these products, e.g. by determining the size in the physical store and then placing an order online.

> *"It's really a lot of problems if you make wrong decision. Then you have to follow up and try to get your money back, and send things back, which is really big hustle. Sometimes it's actually impossible to return items."*

> *"I'm not one of those people, who go and buy loads of sizes, bring it all home, try it all on and take it back. I can't do that because I can't stand having a load of stuff in my house that I bought I don't want. I've gotta get it back as quickly as possible because I want my money back. So, I tend to be really sure about the sizes and don't buy anything unless I'm absolutely sure."*

One particular "demotivating factor" associated with the need to return goods is *having to pay for returns* - in the words of one subject, it is all about *"feeling a bit hard done because of cost of return"*.

Online shoppers also feel demotivated because they have several *"fears"*. One of the most common is the *"fear of making a mistake"*. For example, if a shopper made a mistake of ordering a product with wrong specifications in the past, they prefer not to order online in the future:

> *"I didn't know what I was doing. So I think that made me a bit more... I'd better off going into a shop for somebody to explain it to me better."*

> *"It was useful to talk to a human being! I can't remember what it was, but I'd got something seriously wrong, or would have got something seriously wrong if he [a shop assistant in a brick-and-mortar store] hadn't ordered it for me."*

> *[By dealing with a company in person] "I'd feel more sure that I hadn't made a mistake. Or if I'd made a mistake it would be somebody else's liability... I feel if I'm dealing with a human being that's offloads responsibility on to them."*

A *"fear of substitutes"* is another common demotivating "fear", although our data shows that this fear is more relevant to online grocery shopping than any other type of shopping. This *"fear"* (as well as other "fears") may lead online shoppers to do their shopping in a 'physical' store.

People are also put off online shopping due to the fear of having *"problems with availability of goods"*, and in particular, if a website does not have an item in stock but still takes orders. *"Problems with delivery"* factor also featured prominently in our respondents' statements and comments, and in particular, when shoppers do not know upfront how much delivery charges will amount to or when the delivery will arrive; or when a company breaks their promises by not delivering goods as described or on time. For example:

> *"... it took so long to arrive. I think it was practically a week after the New Year. Even though they said it was going to be there [on time]... So, I think that was the end of me doing Christmas stuff like that. I thought I am not going to be done doing that again."*

"Not providing clear information" can also play a deciding role in shoppers refusing to make purchases, for example:

> *"... they had sale and they had price in red. It's probably me making an assumption, but when I see a price in red, I assume that's the price. But it's not... It was the discount. And I thought: "I can't buy from there! They are not clear enough on what they are selling the goods for!" And I think they are trying to deliberately deceive me here. And that's how I view it. I thought they deliberately try to deceive people into in that, thinking: "Oh, I'll have that - it's only 7.99!" Instead of, actually, it's 24.99. That we have taken 7.99 off the price of something. Why would you do that? 'Cause if I see sale and then the price in red next to it, you would assume that's the sale price, but it wasn't."*

"Technical issues with websites" can also be detrimental to online shopping:

> *"I would not trust to buy it ... 'cause the site didn't work. So then you are less likely [to] trust that your details are going to be safe and secure and you actually get [the goods you ordered]."*

Some of the factors that we identified as *motivators*, also have their 'opposites' as *de-motivators*. Thus, while good, *"positive experience"* is a motivator to shop online and might lead to more online purchases, *"bad experience"* has an opposite effect.

"Ethical behaviour" is juxtaposed with *"unethical behaviour"*; and while for the majority of people having already been registered with a company was a motivator, for some other respondents *"the necessity to register"* with websites was a demotivating factor, and in particular because *"you need to remember all those passwords"*.

5.3.1.3 High Focus

Every interview and observation showed that online shopping is a **highly focused and very directed activity**. Shoppers identify their need prior to performing their searches and often stick to their plan:

> *"I already would have some idea what I need..."*

> *"I usually stick to what I want..."*

> *"... now I'm really becoming more decisive. And even the product [is] reduced, I wouldn't necessarily buy it now, because I don't need it. I think you are learning through the, like, learning curve of your shopping habits. [if] ... at that time I don't need it - that's it!"*

One possible explanation for the reasons why online shoppers' activities are so highly focused is their unwillingness to spend extra time on browsing around, as explained by the following quotes:

> *"I'm usually pretty focused on what I want to get. I haven't got time to aimlessly look around. I've usually got an idea of what I'm going to get and I'm homing quite quickly because of time."*

"I wouldn't troll through them all... 'cause it's such a waste of time."

"I'm really specific, you know. I'd know what I was looking for and I would just look for that specific thing. I wouldn't sway off it. ... I don't look for hours. ... I'd like to do the job in 5 or 10 minutes - job done!"

Another possible reason for wishing to stay focused is the shoppers' beliefs that additional effort would be required if the goods were bought on impulse, for instance because the goods would need to be returned (and this appears to have a direct link to the *"lack of physical contact"* and "need to return goods" demotivating factors - § 5.3.1.2). The following quotes illustrate this aspect of online shopping:

"It's really a lot of problems if you make wrong decision. Then you have to follow up and try to get your money back, and send things back, which is really big hustle. Sometimes it's actually impossible to return items. [(1)] ... That's why you should be careful, you should have your own sort of list, your own plan how to use it. Because otherwise you will be in trouble."

"I wouldn't want to buy something knowing that I would have to go back and take it back. I won't want to go back. Why would you not buy what you want in the first place?!"

5.3.1.4 Habitual Shopping

"Habitual shopping" is another strong characteristic of online shopping that featured prominently in our data. Based on our data analysis, the shoppers' habits

(1) This quote was also used to illustrate the *"lack of physical contact"* demotivating factor in § 5.3.1.2, and repeated here to explain the respondent's emphasis on the need to have a shopping "list" or "plan".

appear to form under the influence of the *previous experience* motivating factor that we described earlier in § 5.3.1.1:

> *"I shopped with them previously and, you know, when the item comes, it's exactly what you've seen [on the website] and it's always beautifully packaged and its always in tip-top condition. ... It's [delivery] always been really, really quick."*

> *"Because I bought from them before when I was looking for stuff and that's it .. So now, 'cause I know they sell what I need, I tend to go just for there. ... It's weird! 'Cause I have done the search, find one [website] I like - that's it. That's the only one [website] I'm buying from..."*

> *"I don't think I've looked anywhere else for books ... I've always found what I wanted."*

Based on the previous experience, all our respondents stated that they have discovered their **favourite websites**, which they shop with - sometimes predominantly or even solely. They also form beliefs about these websites, for example that a particular website offers *"best value"*, so the shoppers tend to stick with these websites:

> *"Generally, I've always found [one particular e-commerce company] just been most reliable. ... and because it's always the best value".*

> *"Because they are very competitive..."*

Moreover, online shoppers tend to use **particular websites for particular purchases**, for instance:

"I use one particular one website for the clothes..."

"I don't think I've looked anywhere else for books ... I've always found what I wanted."

"I know the sizing is right. Just suits me..."

The respondents also have their **regular items** they purchase as well as **particular items** they would never buy online. (§ 5.3.2 discusses this aspect in more detail.)

5.3.1.5 Non-Spontaneity and Non-Impulse Buying

By and large, online shoppers are **not spontaneous** in making purchases. For the majority of the respondents, there is no impulse buying and it is more about planning and taking time to decide on whether to purchase something or wait.(1) The following quote represents a typical situation:

"I would look at it [a product], think: "Hmm, that's nice!" ... think about it. (Because I don't want to spend the money.) ... And I'd look at it for a while then and then I'd wait for the sale, probably."

One of the possible explanations for lack of impulse buying is that many shoppers have learnt from past experiences that it might be worth waiting for sales and discount offers:

"... because I know they're are going to have [offers, sale] sooner or later ..."

(1) However, impulse buying appears to be more prominent when shopping in physical stores, e.g. *"...normally I just impulse buy in shops"*

Nevertheless, there are several factors that can affect spontaneity and impulse buying. For example, informing shoppers about limited availability, or running a promotional campaign can push shoppers to buy something there and then and not wait for sales.

5.3.2 Differentiating Aspects and Factors

In contrast to the online shopping common factors and aspects described in § 5.3.1, the differentiating factors and aspects are those, where the case study participants disagreed. Fig. 5.6 shows these differentiating factors [(1)].

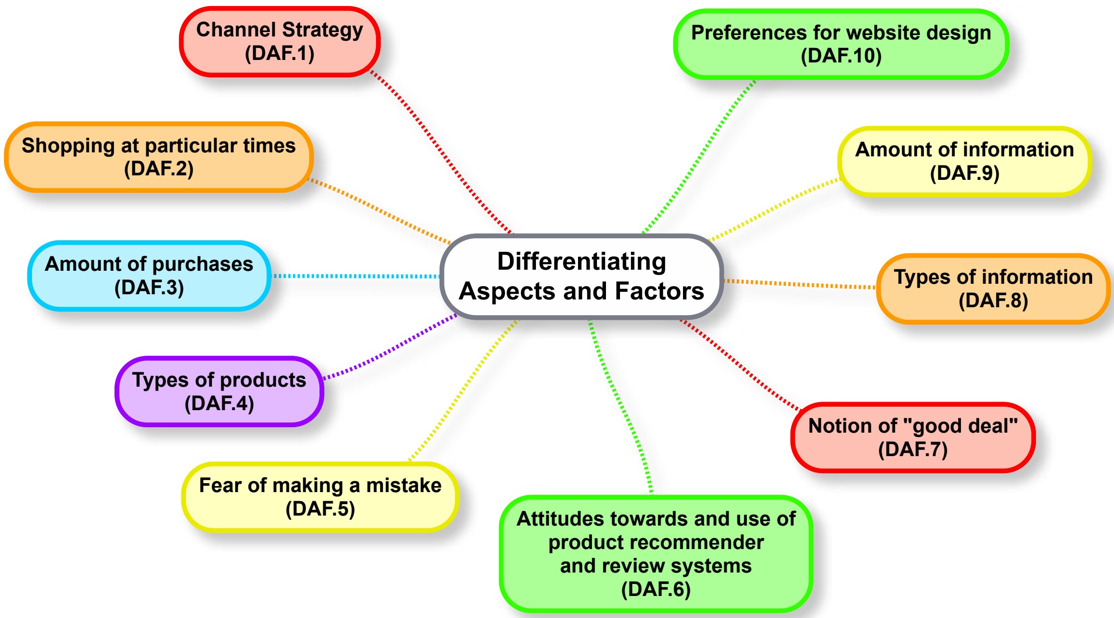


Figure 5.6: Differentiating Aspects and Factors for Online Shopping

DAF.1: Channel strategy

All participants used a multi-channel strategy of combining shopping online

(1) The labels "DAF.*X*" were added to these factors (and not to the factors, described in the preceding sections) due to the need to discuss the differentiating aspects and factors in more depth and the necessity for comprehensive cross-referencing.

and in brick-and-mortar stores, albeit to a greater or lesser extent. However, the *'source'*, where the actual purchases are made, differed. Some shoppers go into a physical store to check a product out, try it and then make their purchase online. Others do it the other way around - they research a product online and then go to the physical store to purchase it. The following examples illustrate this phenomenon:

- Research in-store, buy online:

> *"I'd look at their shop ... I'd look at their products and I just decide what [to buy] ... and then I would buy it off here [website]."*

- Research online, buy in-store:

> *"What you see online is quite often in-store as well, so I can go to the store and look..."*

> *"I like to look at stuff [online], but I wouldn't necessarily buy it. I'd probably go into town and have a look at it and get it from there [physical store]"*

Often, the choice of strategy depends on the type of products, price and risks involved. The *fear of making a mistake* factor, which we specified as one of the de-motivators in § 5.3.1.2, might have an effect here. For instance, one participant explained that when he was buying a new computer he did a lot of research online, but did not proceed with an online purchase because he felt that he *"wouldn't be able to get it right myself"* due to *"too many choices"* of different upgrades, specifications, add-on services. So, in the end, he ordered it in-store after extensive research online.

Sometimes, the channel strategy might change due to a situation change. For example, one participant used to buy goods online, but once the company opened a physical store decided that online shopping with that company wasn't for her any more:

> *"They now have opened a shop in Cardiff, so I'm probably not going to do that [buying online] any more either ... because I like actually going and have a look. So if I go into town once a week, I will actually go there [physical store]."*

DAF.2: Shopping at particular times

Online shoppers differ when it comes to whether they have particular set times for doing their online shopping. Some respondents indicated that they shop at particular times of the day, e.g. during lunch time or in the evening after having come back from work. Other respondents, do not have a particular time for their shopping. They shop at whatever time they want or need to.

DAF.3: Amount of purchases

This is another differentiator - shoppers vary considerably with regard to the amount of purchases they make. Some make purchases every week, others shop only at particular times of the year.

However (and like any other type of shopping), online shopping exhibits strong **seasonal patterns**. For example, the majority of respondents stated that they are more likely to make a greater number of purchases over the Christmas and New Year period than any other time of the year, for example: *"... at Christmas, New Year then I buy quite a lot"*. Some respondents make their purchases only during those periods.

DAF.4: Types of products

The respondents differ in the types of products and services they would or would not buy on the internet. For instance, all respondents would buy books, tickets for events or make their travel arrangements online. The majority would buy film DVDs and household goods. However, the vast majority of the respondents would never buy medicine on the Internet. There are also some types of products, which have divided the respondents. Thus, for example, clothes, shoes and furniture have come across as 'controversial' products - some people do not have any reservations about buying these online, some prefer not to buy them online, while others stated that they would be happy to buy these products online but first need to see and try these in a physical store:

> *"I just don't think about buying clothes on the internet. I like to go to a shop and browse and see what I want to try it on."*
>
> *"That's [buying in a physical store] the only way to see colours and see if it fits."*
>
> *"I would have to go feel and see it [in a store]..."*
>
> *You have to look at it to see it [furniture], touch it, you know ... unless I've seen it [previously]"*
>
> *"... I like to try them on. Shoes are the one thing I like to try on 'cause I've got 'funny' feet... and clothes I don't like... Some sportswear that I'd buy, but mostly clothes I have to go to a shop to try them..."*
>
> *"I wanna test drive it.."*

> *"I have bought one pair of boots ... But I had actually already tried them on somewhere, so I, basically, knew what size I was going to get."*

There is a similar division with regard to online grocery shopping. Some shoppers do not have a problem with this and regularly do their food shopping on the Internet. Others prefer not to buy perishable items, such as meat, fruit and vegetables online. For these shoppers the same rule of being able to see and touch in a physical store applies:

> *"I seldom buy food or groceries. I think I prefer to see what I'm buying. Especially, if it's fresh food. Like, I don't mind ordering online things like washing powder and your cleaning products and things like that. I'm quite happy to buy them. But when it comes to fresh food, I like to go and see what's there. ... I like to touch fruit."*

DAF.5: Fear of making a mistake

"Fear of making a mistake" was one of the demotivating factors described in § 5.3.1.2. While it is true that this was one of the factors that might prevent shoppers from buying online, our case study subjects differed in the intensity of this fear. For some the fear was strong enough that they decided not to buy certain products, for example, car insurance or high outlay goods. In these cases, making a mistake was considered to be too costly and outweighed the possible benefits of online lower prices or convenience. For others the fear of making a mistake was not as strong. These subjects would still make their purchase, but had to take extra precautions, such as checking the details more carefully.

DAF.6: Attitudes towards and use of product recommender and reviews systems

Product recommender and reviews systems have become a prominent (if not 'compulsory') functionality offered by e-commerce websites. Our analysis of the interviews and observations shows that shoppers' attitude to and use of recommenders and reviews when making their purchase decisions vary significantly.

Some respondents find these useful or at least do not mind them:

> *"I look at the "Recommended" [list] underneath and say: "Oh yeah, I could have that!"*

but often this only relates to a certain type of product, for instance:

> *"If you particular want one book, particular author, then you might forget that it has other ones, really good stories or... yes, this is a really good reminder" [but] "... with clothes I tend not to follow these instructions they try to match what I buy - because I already have some ideas..."*

Some people have a fairly negative attitude to recommenders and voice their categorical dislike:

> *"I don't really like when the site becomes too suggestive and it's putting pressure, a little bit of pressure on you. Even like in suggestive form. This I don't personally like."*

> *"I find it [recommendation functionality] annoying. I'll tell you why - because it's intrusive - big brother watching you. And I don't like that".*

"I don't personally like it. In particular, when some sites push you a little bit towards certain decisions [when] you don't actually have a motive, a real motive to buy."

"Sometimes they can be too helpful, and in being too helpful they can be a little off-putting... You know, I wanna be able just browse and pick what I...[want/ need]... I don't want them to make recommendations to me."

"... if I buy something, it would be one thing, and I don't like to get swayed into buying 2 or 3 more. I feel really really bad, you know, then."

"... I feel like I'm being controlled and I'm being manipulated."

Nevertheless, the majority of shoppers, even those with a negative attitude towards recommendations, appear to understand the reasons behind this functionality (or at least have tolerance with it):

"I can see what they are doing..."

"Even, you know, you don't use it, it's interesting to go so and see what. . . I don't know, being nosy probably. ... That feature is quite good even I don't use it, but I find it's just interesting how it relates to other products. If you are looking for series of books, it gives you ... that extra information."

And these shoppers even look at the recommended goods, if only out of curiosity:

"Although, I'll have a little scout [to see] what they're recommending..."

"Sometimes I looked at it [a recommendation] just to see why. Why would you buy that with that? And would you buy those things together if there are the same things, more or less? ..."

Shoppers demonstrate a similar variety of attitudes towards product reviews. Some respondents do not read or pay much attention to these, albeit often it depends on the type of a product. For instance, for books there is a widespread dislike of reviews, such as:

"I know what I want. Regardless of what somebody else thinks... I'll get it."

However, it is a different story for high outlay products. For example, one respondent was buying a new vacuum cleaner and checked every single review for a particular cleaner she was going to buy:

"I checked every single review ... If I'm spending that kind of money, it's got to be a good one."

Our analysis of the data collected through interviews and observations has revealed another interesting fact: the amount of product reviews people are willing to read varies. The observed difference stem either from personal traits and preferences of participants, or from the type of product, or sometimes both. Thus, for some participants a limited number of reviews was sufficient, e.g. only top 3 reviews or just "skimming over":

"So, I've probably skimmed it, and only registered it if it says "It's rubbish"..."

Other participants, as in the case above, prefer to read all reviews. Buying high outlay products appears to make shoppers read more reviews, compared with low outlay products.

DAF.7: Notion of a *"good deal"*

When making a purchase decision everybody wants to find a *'good deal'*, but people vary in their notions of what constitutes this *'good deal'*. For some shoppers price is more important, for others - it is superior quality, for which they are prepared to pay extra.

However, even for those looking for the cheapest price, often a *'good deal'* is not just the cheapest or best quality product, it is about a combination of the two - as seen by a shopper's formed beliefs of what combination works best for them, e.g.:

> *"the cheapest for what I want ... The one that ticks all boxes - and then ... the cheapest."*

In all cases, shoppers tend to evaluate the offerings from the point of view of past shopping experiences, for example:

> *"I will actually have rough idea what a sweater or trousers will cost approximately, and then ... and obviously, compare if I see [them] in the shop. ... Obviously, I would know how much I spent before on the same sort of items, and if it's too expensive then I'll wait for sale."*

When looking for a *'good deal'* and evaluating whether it is truly *"good"*, different shoppers employ different strategies. Most respondents use price comparison websites to aid their decision making, but some decide to stay away from them, mostly because of their unwillingness to spend extra time, and in particular, since they have already formed their opinions and beliefs about certain websites they used in the past (see § 5.3.1.4, *"Habitual Shopping"*). For instance, one participant told us that typically he doesn't bother checking other websites to compare prices and goes straight to Amazon to make certain purchases, because: *"I always found them to be very competitive"*. Similarly, another participant, having formed his opinion about companies operating in the same sector and offering similar products, doesn't want to shop around any more:

> *"... I've decided not to go for any others [companies] because they're all pretty much the same in reality."*

DAF.8: Types of information

Websites offer two types of information: textual and visual. Some respondents stated the need to read textual information, such as, for instance, product descriptions. Others emphasised their preference for visual information:

> *"I like everything with maps. I like things with good information, like the map exactly, you know..."*

> *"I'd love to be able ... [to have a] shopping basket - and to be able to go through and just drag-and-drop things into my shopping basket as I see them. ... I'd like be able to go like ... almost have it like a virtual supermarket. ... It'd be great if you could go through a supermarket as if it was a virtual store. Then you have your shelves and you just click on something and it'll pop in your basket. Instead of having an item written down, it's actually like looking at shelves of goods and then you just click on items you want. ... 'Cause you're getting an experience..."*

However, the issue of preference for a particular type of information is not clear-cut. Finding the right balance between the two appears to be key.

DAF.9: Amount of information

The amount of information is another aspect that is different for different shoppers. Some shoppers like getting a lot of information and get frustrated if they don't get that. Others get annoyed if there is too much information, for example:

> *"It used to be quite easy - used to have a choice of one book, but now there's quite a few options and it takes a bit longer than it used to. Too many choices!"*

> *"... you don't spend much time reading them [e-commerce websites]. You just ... you come across, you look... And they are very busy, really a lot of text ... "*

The shoppers, who don't like having a lot of information, are typically satisfied by a limited number of choices. Often, these are the first few choices that the shoppers are presented with:

> *"I order them [goods that the participant searched for] in price, so you only need to look at the top three of each sites. And that's it."*

> *[I would only look at the] "first three reviews. If they all saying the same thing, I'd think "OK, that's going to be typical". If there were different, then I'd read more."*

> *"I usually find what I want on the first page. I don't tend to go go beyond."*

> *[I would see] "who does provide them [products] and probably troll through five of those."*

Similar to the situation with the types of information, the preference for the amount of information might be affected by the type of product a shopper is looking for. For example, high outlay products, which involve bigger purchase risk and consequently ‘harder’ decisions, would most likely push all shoppers, even those who prefer a limited amount of information, to seek more information than for other types of products.

DAF.10: Preferences for website design

Shoppers differ in having or not having preferences for website design. Some respondents have strong opinions about what background colours, navigation menu features or page layout they would or would not want to see on a website. Others do not mind much what they see on a website, but nevertheless still have some ideas of what a *‘nice’* website should look like, albeit they were unable to express these ideas in a concrete form.

5.3.3 Personalisation and E-Shoppers’ Profiles

The previous section has detailed the characteristics of online shopping, including common and differentiating factors and aspects that e-commerce shoppers exhibit. Common factors and aspects give us a general understanding of what motivates and de-motivates our case study subjects, and thus provides us with a context of what e-commerce shopping experience and processes entail - from the point of view of e-shoppers. However, it is the analysis of the differentiating factors and aspects that allows us to identify different profiles that e-shoppers can be ‘*segregated*’ into. The profiles can then be used to extend the *OntoProfi* ontology to enable website personalisation.

5.3.3.1 What Can We Profile and Personalise?

To enable derivation of user profiles, a further analysis of which factors/ aspects could be used for user profiling was needed. Table 5.2 summarises these factors:

ID	Factor/ Aspect	Can this factor/ aspect be used for profiling?
DAF.1	Channel strategy	No
DAF.2	Shopping at particular times	No
DAF.3	Amount of purchases	No
DAF.4	Types of products	No
DAF.5	Fear of making a mistake	No
DAF.6	Attitudes towards and use of product recommender and review systems	No
DAF.7	Notion of a *"good deal"*	Yes
DAF.8	Types of information	Yes
DAF.9	Amount of information	Yes
DAF.10	Preferences for website design	Yes

Table 5.2: Evaluation of Possibility of Personalising Particular Differentiating Factors and Aspects

This table shows the differentiating factors and aspects, classified into two groups: the ones that, explicitly and directly, cannot be used for user profiling and the ones that can be. The *"Yes"* and *"No"* answers came as a result of our analysis of whether these factors were external to the shoppers (e.g. system-oriented perspective) or internal (such as a shopper's personal belief, attitude or other intrinsic characteristics). External factors can be used for user profiling and can be personalised by the system. Internal factors theoretically can be influenced by personalisation, but cannot be personalised directly.

DAF.1 through to **DAF.5** are the factors that belong to the first, *''internal factors'* group as they cover aspects of online shopping that belong to shoppers' *'strategic'* or *'in-*

ternal beliefs' categories, and as such cannot be used for any direct user profiling. The **DAF.6** factor is slightly different. This factor can be used for personalisation of content presentation, for example, by asking the users to specify their explicit preferences for product recommendations and reviews, and then show them or exclude them from the web pages. However, in our analysis, the factor came out as an *'attitudinal'*, *'internal belief'* factor, and, therefore, cannot be used for profiling.

The factors **DAF.7** to **DAF.10**, on the other hand, are classified as *''external'*, and are very suitable for user profiling.

DAF.7 "Notion of a *"good deal"* ":

Strictly speaking, it is not possible to personalise the "notion" per se - as it relates to a person's internal belief. However, it is possible to utilise the differences between shoppers' preferences for particular types of products, which our analysis of the case study data revealed (i.e. *"budget"* vs. *"quality"*). Both types of personalisation can be implemented: through user-directed (explicit data) personalisation, when users specify explicitly which products they consider to be a *'good deal'* or most relevant and appropriate; and through a system-based (implicit data) approach, when the system evaluates products' appropriateness based on a specified set of rules or criteria (the use of, and advantages and disadvantages of using explicit and implicit data in user profiling was discussed in § 2.4.1.1).

DAF.8 "Types of Information" and **DAF.9 "Amount of Information"**:

Personalisation of these two factors is straightforward. As we learnt from the case study, shoppers differ in their preferences for the type and amount of information they typically look for on e-commerce web pages. Thus, some participants preferred textual information, while others have a better response to visual presentations. Some e-shoppers are content with small amount of information, while others expect and look for more. Implementation of this functionality requires the presence of different types of information (e.g. product photos and description) and different level of detail of this information (e.g. brief and comprehensive product description). This information is most likely already present in an e-commerce system, but needs to be displayed on the web pages in new ways.

DAF.10 "Preferences for website design":

Personalisation of a website based on user preferences is a straightforward and fairly easy functionality to implement. There are a number of ways to implement website design personalisation, and all elements of a website can be personalised - although, from the website design and implementation point of view, personalisation of some website elements is more involved than others. For example, it is fairly easy to provide users with a user-directed personalisation option of changing web page background colour, font typeface and font colour. It is more difficult to implement a user-directed option to personalise navigation, in particular, for complex websites, which use dynamically generated navigation. In this case, personalisation can still be achieved by using a system-based approach, but implementation would be more costly and difficult.

5.3.3.2 E-Shoppers Profiles for E-Commerce Personalisation

Upon examination of the factors and aspects that can be used for user profiling and can be personalised, we propose the following e-shopper profile entities to be used to personalise an e-commerce website:

ID	Factor/ Aspect	E-shopper Profile Entity
DAF.7	Notion of a "good deal"	**budget** vs **quality**
DAF.8	Types of information	**textual** vs **visual**
DAF.9	Amount of information	**satisficer** vs **maximiser**

Table 5.3: E-commerce Shoppers' Profiles

In Table 5.3, only DAF.7 - DAF.9 factors are included. DAF.10 (*"Preferences for website design"*) was not included explicitly as it was determined inappropriate to do so. *"Website design"* is a concept that covers aspects belonging to different domains, for instance, it includes visual aspects, navigation, underlying functionality, etc. Moreover, a "good" design can be considered as an emergent property, and therefore does not lend itself easily to being used to segregate shoppers into several distinct profiles. So, instead of using the DAF.10 factor explicitly, particular relevant aspects and elements of website design were incorporated within DAF.8 and DAF.9. For example, our *OntoProfi* ontology contains a number of classes that cover different types and amount of information that can be personalised (§ 4.3.2, and Fig. 4.8)

5.3.3.2.1 ***Budget*** **vs** ***Quality*** **Profile Entity**

This profile entity was formed in recognition of the fact that shoppers differ in their notion of a *'good'* deal, and we profile e-shoppers into these two categories, based on their preferences for certain products - either the most competitively priced or the best quality.

5.3.3.2.2 *Textual* vs *Visual* Profile Entity

Our proposal of this profile entity was based on our findings of differences between e-shoppers' preferences of either textual or visual information. We propose that one or other type of information would be made more prominent (§ 6.2.4.2 illustrates how this was implemented in the prototype).

5.3.3.2.3 *Satisficer* vs *Maximiser* Profile Entity

"Satisficer" and *"maximiser"*(1) are the last two profile entities, whose purpose is to reflect e-shoppers' preferences for the amount of information they receive on web pages, as detailed in our discussion of DAF.9 factor in § 5.3.2. Thus, a user classified as *"satisficer"* is typically content with less amount of information compared with a *"maximiser"*; and this can be successfully modelled in a personalising system.

5.4 Extending *OntoProfi* Ontology

Having identified what profile types the customers can be segregated into, *OntoProfi* needs to be extended to model these identified profile types so that the reasoner can infer what profile each customer belongs to.

(1) The name of the first profile type, *"satisficer"*, first introduced by Simon (1956), was influenced by the term used in psychology to explain a decision making strategy, when a person makes a decision based on certain heuristics without going through all and every available alternative. The name of the second profile type, *"maximiser"*, was proposed by ourselves as a juxtaposition of *"satisficer"*. It is only on completion of the case study, we discovered through further research that the term bearing the same name already exists in psychology and decision making (e.g. (Schwartz, et al. 2002)). However, the already existing term relates to a more general behavioural aspect.

5.4.1 Extending the `Customer_Profile` Class

The identified customer profiles can now be included in *OntoProfi* as subclasses of `Customer_Profile` class. Thus two new classes of `Maximiser` and `Satisficer` and corresponding subclasses (*visual* vs *textual*, and *budget* vs *quality*) are added to the ontology (see Fig. 5.7) (1).

Figure 5.7: Extended `Customer_Profile` class

To enable customer profiling, additional object properties of `hasInfoAmount`, `hasInfoType` and `hasProductType` need to be added to *OntoProfi*. These object properties can now be used to construct `Maximiser` and `Satisficer` classes and their subclasses by linking these to the `Product` and `Information` classes (see §§ 4.3.2.2 and 4.3.2.4). Fig. 5.8 is an update of Fig. 4.10, and shows how the relationships in *OntoProfi* have been extended to incorporate these additional object properties.

(1) Postfixes `_bt`, `_bv`, `_qt` and `_qv` in the figure denote respective shorthands for *"budget-textual"*, *"budget-visual"*, *"quality-textual"* and *"quality-visual"*.

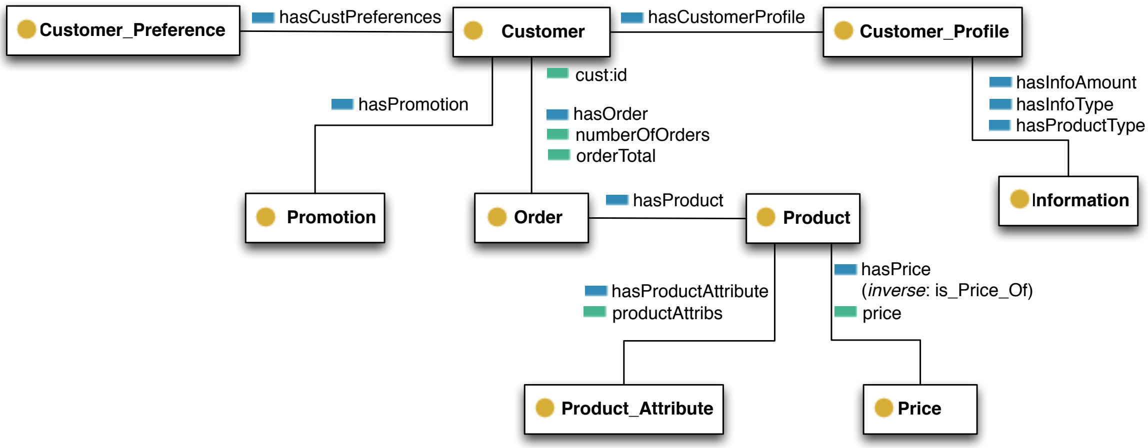


Figure 5.8: Key Relationships in Extended *OntoProfi* (high-level)

To define `Maximiser` and `Satisficer` classes and their subclasses, *OntoProfi* uses the logic-based existential `owl:someValuesFrom` restriction.

The top-level classes, `Maximiser` and `Satisficer`, are defined as:

$$\texttt{Maximiser} \equiv \texttt{Customer\_Profile} \sqcap (\exists \texttt{hasInfoAmount.Detailed\_Information}) \tag{5.1}$$

$$\texttt{Satisficer} \equiv \texttt{Customer\_Profile} \sqcap (\exists \texttt{hasInfoAmount.Brief\_Information}) \tag{5.2}$$

The subclasses of `Maximiser` and `Satisficer` are then defined as:

$$\begin{aligned}\texttt{Maximiser\_bt} \equiv \; & \texttt{Maximiser} \sqcap (\exists \texttt{hasProductType.CheapProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Textual\_Information})\end{aligned} \tag{5.3}$$

$$\begin{aligned}\texttt{Maximiser\_bv} \equiv \; & \texttt{Maximiser} \sqcap (\exists \texttt{hasProductType.CheapProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Visual\_Information})\end{aligned} \tag{5.4}$$

$$\begin{aligned}\texttt{Maximiser\_qt} \equiv\ & \texttt{Maximiser} \sqcap (\exists \texttt{hasProductType.QualityProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Textual\_Information})\end{aligned} \tag{5.5}$$

$$\begin{aligned}\texttt{Maximiser\_qv} \equiv\ & \texttt{Maximiser} \sqcap (\exists \texttt{hasProductType.QualityProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Visual\_Information})\end{aligned} \tag{5.6}$$

$$\begin{aligned}\texttt{Satisficer\_bt} \equiv\ & \texttt{Satisficer} \sqcap (\exists \texttt{hasProductType.CheapProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Textual\_Information})\end{aligned} \tag{5.7}$$

$$\begin{aligned}\texttt{Satisficer\_bv} \equiv\ & \texttt{Satisficer} \sqcap (\exists \texttt{hasProductType.CheapProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Visual\_Information})\end{aligned} \tag{5.8}$$

$$\begin{aligned}\texttt{Satisficer\_qt} \equiv\ & \texttt{Satisficer} \sqcap (\exists \texttt{hasProductType.QualityProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Textual\_Information})\end{aligned} \tag{5.9}$$

$$\begin{aligned}\texttt{Satisficer\_qv} \equiv\ & \texttt{Satisficer} \sqcap (\exists \texttt{hasProductType.QualityProduct}) \\ & \sqcap (\exists \texttt{hasInfoType.Visual\_Information})\end{aligned} \tag{5.10}$$

Additionally, these classes are made disjoint through the use of `owl:disjointWith` constructor. This disjointness is necessary to ensure that no individual is classified into more than one class by a reasoner, so the following disjointness axiom is asserted:

$$\texttt{Maximiser} \sqcap \texttt{Satisficer} \sqsubseteq \perp \tag{5.11}$$

Similarly, all corresponding subclasses are asserted as disjoint:

$$\texttt{Maximiser\_bt} \sqcap \texttt{Maximiser\_bv} \sqcap \texttt{Maximiser\_qt} \sqcap \texttt{Maximiser\_qv} \sqsubseteq \perp \tag{5.12}$$

$$\texttt{Satisficer\_bt} \sqcap \texttt{Satisficer\_bv} \sqcap \texttt{Satisficer\_qt} \sqcap \texttt{Satisficer\_qv} \sqsubseteq \perp \tag{5.13}$$

5.4.2 Customers as *OWL Individuals*

OWL individual is an instance of an *OWL class*. Like an OWL class, OWL individuals maybe be asserted as relating to each other through specified properties. OWL individuals can also be stated as the same through the use of `owl:sameAs` (this is similar to the class equivalence) or different from each other using OWL's `owl:differentFrom` property assertion.

To enable customer profiling in *PERSONTO*, we need three category of OWL individuals: to represent instances of a `Customer`, `Product` and `Information`. These individuals are created, asserted and inferred, using:

- Each customer is represented in *OntoProfi* as an *OWL individual* and encoded using the `owl:differentFrom` assertion. OWL individuals are created during the *PERSONTO*'s pre-processing stage (see § 3.5.3.1). Initially, they contain a minimum of information (a profile's *'skeleton'* that contains customer account details, such as customer ID and name), and store more information when the customer supplies it, e.g. by completing the profiling survey (see § 3.5.3.4.2).
- As each customer's usage of the system increases or when a customer specifies their preferences, a number of object properties are asserted for each corresponding OWL individual. The assertions are stored in the *"Customer*

Profiles Repository" (see § 3.5.1).

- Examples of instances of the `Product` and `Information` classes are created manually.
- To enable segregation (profiling) of customers into one of the eight customer profiles, the reasoner analyses the relationship between the newly asserted object properties and individuals and computes (infers) or recomputes the membership of each OWL individual associated with this customer, based on the analysis of the relationship between the instances of the `Customer`, `Product` and `Information` classes.

5.5 Chapter Summary

This chapter has discussed the process and outcomes of a multi-case case study of online shoppers. It has described the research methodology and research design used for elicitation of common and differentiating factors and aspects that affect e-commerce customers. The analysis of the gathered case study data revealed that there are two main groups of factors/ aspects that influence shoppers' decision to buy online, which we have classified as common (typical) and differentiating. The common factors have given us a general understanding of what drives and influences online purchases (i.e. the context of e-commerce shopping), while differentiating factors have allowed us to determine what can be personalised and what types of profiles e-shoppers can be segregated into. The identified profile entities are used to extend *OntoProfi* to include additional user profiling classes and store each customer's details.

Chapter 6

Testing Feasibility of Approach: Prototype of *PERSONTO*

6.1 Introduction

To verify and test the technical feasibility of the proposed ontology-based e-commerce website personalisation approach, a 'proof-of-concept' prototype was implemented.

This chapter discusses the prototype's implementation setup and the outcomes of the experiments to implement the most essential and representative functionality, which allows the personalisation of content presentation. The concluding part discusses limitations of the prototype and the assumptions made during the design and implementation.

6.2 Prototype Implementation

6.2.1 Development Setup: Languages, Technologies, Data and Environment

PERSONTO's prototype was implemented as a Java EE application [(1)].

The following languages, technologies, tools and data were used for the prototype implementation:

- **Development Environment**:

  NetBeans IDE v.7.3.1 (later upgraded to v.8.0.2) was used for coding, compiling and testing. NetBeans is a widely used software development platform that provides a comprehensive set of tools and support for Java, databases and web servers.

- **Website Development Languages and Technologies**:

  The website was developed as *'bespoke'* and from *'scratch'*, using Java, JSP [(2)], HTML and CSS. These technologies are common in research and practice. For example Ardissono et al. (1999), Chen et al. (2010), Flesca et al. (2005), Paik, et al. (2002),Skillen et al. (2014) and Wu, et al. (2001) use Java as the programming language to create their systems. Java is also needed for communication between various components, for instance, for interaction with an OWL ontology through an available API. Java's JDK v.1.8 was used in this project.

  JSP, HTML and CSS are de-facto standards in website development.

(1) Java Enterprise Edition - http://www.oracle.com/technetwork/java/javaee/overview/index.html
(2) http://www.oracle.com/technetwork/java/javaee/jsp/index.html

- **Database**:

  MySQL was used as the database technology, as it is freely available, widely used, provides all the necessary functionality, and is supported by NetBeans.

- **Ontology**:

  OntoProfi ontology was authored in Protégé 4.1 ontology editor [(1)]. OWL API [(2)] was used for manipulating *OntoProfi* within the *PERSONTO* prototype.

- **Reasoner**: HermiT reasoner v.1.3.6 [(3)] was used to reason about *OntoProfi*.

- **Server**: GlassFish (Open Source Edition v.3.1.2.2 and v.4.0) [(4)] was chosen as a web server to process JSP servlets and JSP pages.

- **Data**:

  The database was populated with *'simulated'* customer, product and order details, using insertion of random data.

  The images used on the *PERSONTO*'s website (in particular, product photos) were harvested from `morgueFile.com`, a website that provides free photos for use in a creative project. All images used for our website had a permission allowing copying, transmitting and adaptation.

Our choice of development languages and technologies in the *PERSONTO* prototype was determined and dictated by:

- their appropriateness, availability and maturity;

- popularity and degree of widespread usage;

- their compatibility with existing e-commerce technologies;

(1) `http://protege.stanford.edu/`
(2) `http://owlapi.sourceforge.net/`
(3) `http://hermit-reasoner.com/`
(4) `https://glassfish.java.net/`

- the author’s own expertise in the chosen languages and technologies.

It is possible to develop the same system using alternative languages, technologies or tools. In fact, all technologies and tools used can be substituted with alternatives. For example, the web pages can be written with PHP or Python, other database vendors, reasoners, web servers and development IDEs can be used without affecting the end results. Appendix E suggests alternatives (open source and commercial) that are widely used in website development. The availability of these alternatives ensures that the future system development will not be hindered should the technologies and tools, used in this project, cease to exist (although it is hard to imagine this will happen).

6.2.2 Website Information Architecture

The website for our prototype was implemented as an e-commerce website containing components that are typical for such a website. Fig. 6.1 depicts the website information architecture for *PERSONTO*:

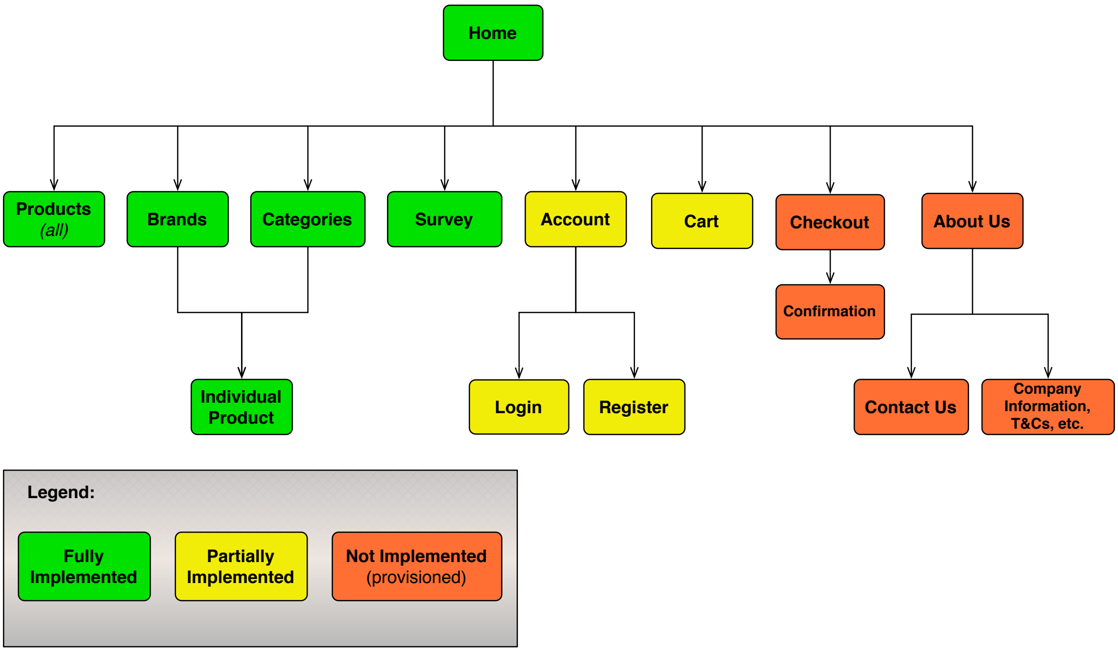


Figure 6.1: Information Architecture of *PERSONTO* website

The project's focus and scope (see § 1.5) determined that we needed to demonstrate and test the functionality that supports personalisation of the *"information search"* and *"evaluation, negotiation and selection"* stages of the online purchasing process (O'Keefe & McEachern 1998), (Turban et al. 2006) - i.e. browsing/ searching for and selecting products, but not purchasing or post-purchase evaluation. Therefore, we only implemented those components (web pages), which were necessary to fulfil this objective. In Fig. 6.1, *green* denotes fully implemented web pages; *yellow* - web pages that were created, but some functionality is limited or missing, for example, the "`Login`" page retrieves the customer's name but not her/ his password; and *orange* - web pages that were provisioned for, but not implemented. Partial or non-implementation of the selected pages or functionality was not deemed to affect the results of our experiments, as there are no dependencies between the implemented pages, which were required for our experiments to demonstrate the feasibility of content presentation personalisation, and the pages that were not implemented or partially implemented. For example, the decision not to implement the password functionality was dictated by the need to speed up testing (i.e. to retrieve the customer's details and profile without typing the customer's password). This functionality can be added at a later date. Similarly, the 'missing' pages (e.g. "`About Us`", "`Checkout`") can be coded without affecting the pages already fully implemented in the prototype.

6.2.3 Database Model

To enable the necessary and typical e-commerce functionality, such as storage and retrieval of customer, product and order data, a database was implemented, based on the Enhanced Entity - Relationship (EER) model shown in Fig. 6.2.

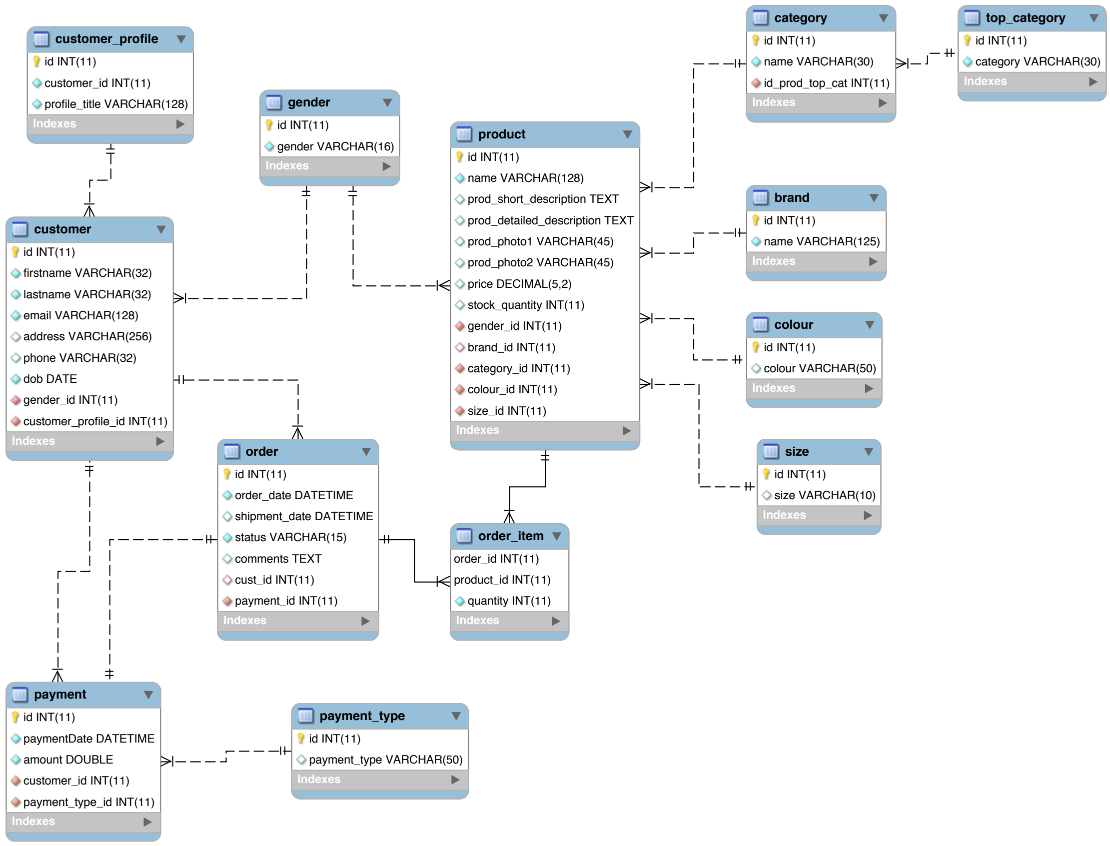


Figure 6.2: EER Model for *PERSONTO* Database

In the current version of the prototype, the data from the database is used for two purposes:

- provision of back-end services to *PERSONTO*'s website, such as dynamic generation of products web pages, and allowing product selection based on their categories and brands;
- creation of customer profiles in *OntoProfi* - by retrieving customer details, such as name and other customer-related data to populate individual profiles (see § 3.5.3.1 and § 3.5.3.4.1).

6.2.4 Website Personalisation: Experiments and Results

To test the feasibility of implementing the system and to demonstrate its representative functionality of how a website's content presentation is personalised (described in § 3.5.3), we developed and tested the following components:

- **front-end** components: a simple e-commerce website that was based on the website architecture, described in § 6.2.2;
- **back-end** components: a set of functions that provided *PERSONTO*'s run-time functionality, described in § 3.5.3.2 and § 3.5.3.5.

Both front-end and back-end components relied on the customer profiles test data, which was created at the start of our implementation.

The remainder of this section describes the creation of such test data before moving on to a description of the process and outcomes of website personalisation based on the inferred customer profiles.

6.2.4.1 Customer Profiles Test Data: Creation, Assertion and Inference

As specified in § 3.5.2, website personalisation in the current prototype of *PERSONTO* is achieved as a "group personalisation", with the customers being profiled into one of the eight predetermined profiles.

The first step that needs to be performed is to create initial customer profiles. This corresponds to the *"Pre-Processing Phase"* described earlier (§ 3.5.3.1).

To generate an initial batch of customer profiles for our experiments, we create 50 individuals, corresponding to the first 50 customers in the *"Customer"* table of the *PERSONTO* customer database - as shown in Fig. 6.3.

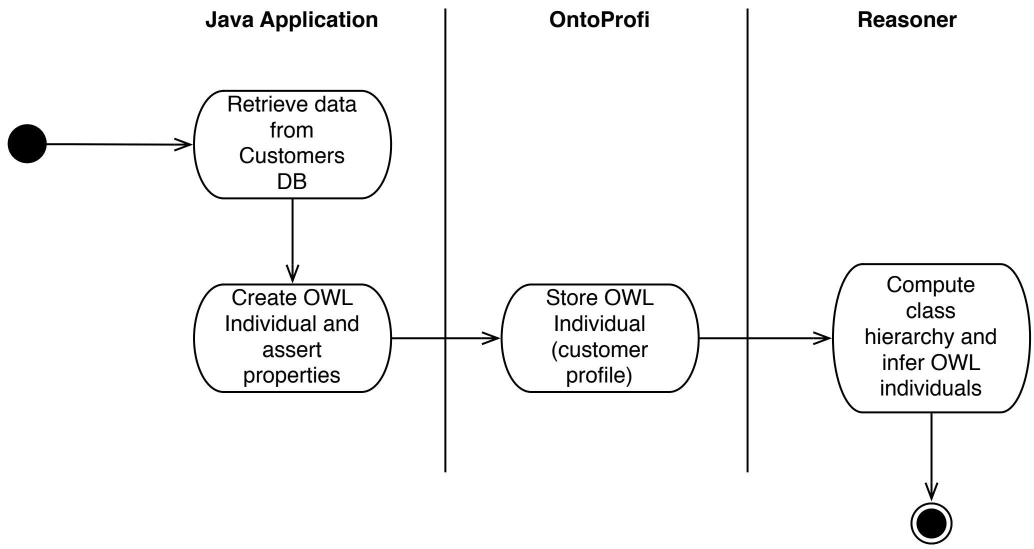


Figure 6.3: The process of Customer Profiles Test Data Creation

Firstly, a Java function retrieves the customer data, such as the customer ID number and the customer name from the *"Customer"* table of our database, and creates an OWL individual for this customer in *OntoProfi* (see § 5.4.2). This *"customer's individual"* is then stored in the *PERSONTO*'s *"Customer Profiles Repository"* (see § 3.5.1).

Secondly, as a simulation of the customer's responses to the profiling survey, we represent the customer's preferences for the type of product and information by random assertion of the relationships between each *"customer's individual"* and three individuals, which represent the instances of `Information` and `Product` classes, such as: `visual_information` or `textual_information`; `brief_information` or `detailed_information`; and `cheap_product` or `quality_product`. The relationships between the *"customer's individual"* and the `Information` and `Product` instances are asserted using the `hasInfoAmount`, `hasInfoType` and `hasProductType` object properties.

Lastly, the reasoner computes the class hierarchy, i.e. infers each customer individual as belonging to a particular profile, based on the asserted relationships. Fig. 6.4 is an example of customers that are inferred by the reasoner as `Maximiser_bv`.

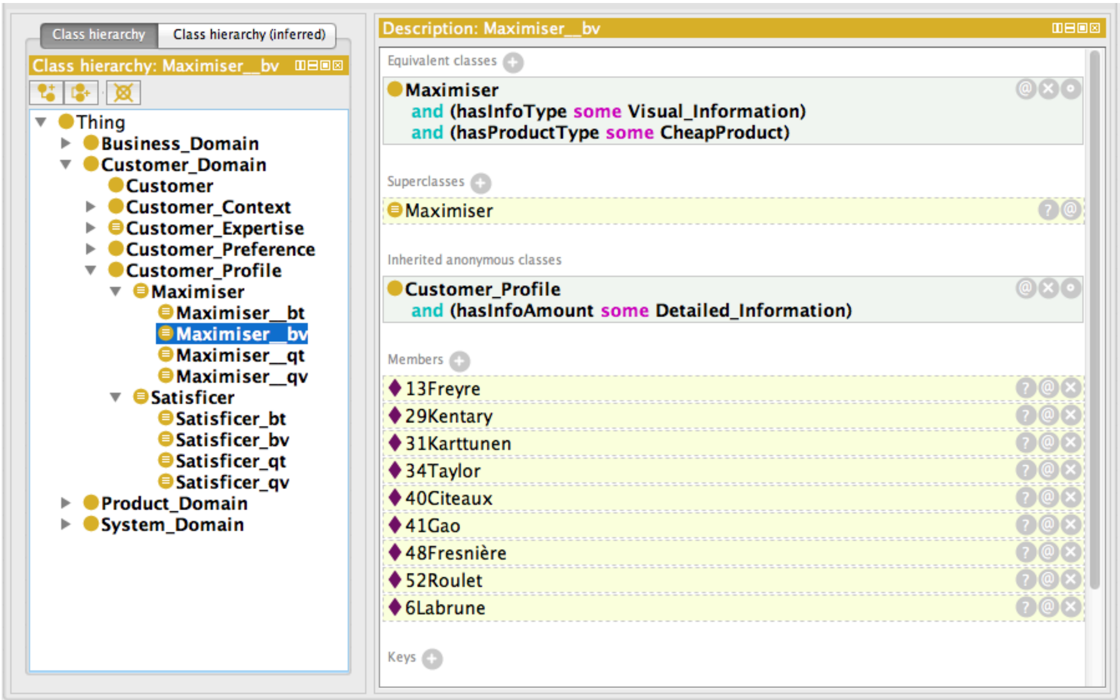


Figure 6.4: Example of customers inferred by the reasoner as beloning to a particular customer profile - here as `Maximiser_bv` profile *(visualised in Protégé v. 4.1)*

Additional *“customer’s individuals”* can be added by the completion of the customer profiling survey (see §§ 3.5.3.4.1 and 3.5.3.4.2). The reasoner will then have to be invoked again to re-compute the class hierarchy to profile the newly added *“customer’s individual”* into a profile.

6.2.4.2 Personalisation of Web Pages

Personalisation of web pages in the prototype follows the process described in § 3.5.3.5. Here, we illustrate the outcomes of such a process by demonstrating a sample of different versions of web pages, personalised according to the inferred customer profiles.

As in any modern e-commerce system (Laudon & Traver 2006), (Turban et al. 2006), the web pages in *PERSONTO* are assembled dynamically from four page fragments: the page header, page footer, left-hand side bar and main body. The page header, footer and side bar stay the same on every page of the website. The main body changes depending on the page type the client (browser) requests the server to retrieve , e.g. *home page*, *brands page*, *product category page* and *individual product page*; and on the inferred customer profile.

To receive personalisation of *PERSONTO* web pages, a customer needs to log in. On login, the customer’s profile is analysed and classified, and a JSP profile processing servlet instructs the web server to generate appropriate web pages. For example, before the customer is logged in, the *PERSONTO* home page looks like the one in Fig. 6.5.

Figure 6.5: Default “Home” page *(customer is not logged in)*.

When the logged-in customer has been classified as *“Maximiser-budget-textual"* (i.e. `Maximiser_bt`), the home page is transformed to the one shown in Fig. 6.6 to reflect the *"maximiser’s"* need for more information, and better prominence of text.(1)

(1) The “Budget” requirement of the profile can be satisfied by sorting the categories by price - this, however, is not implemented in the current version of *PERSONTO*.

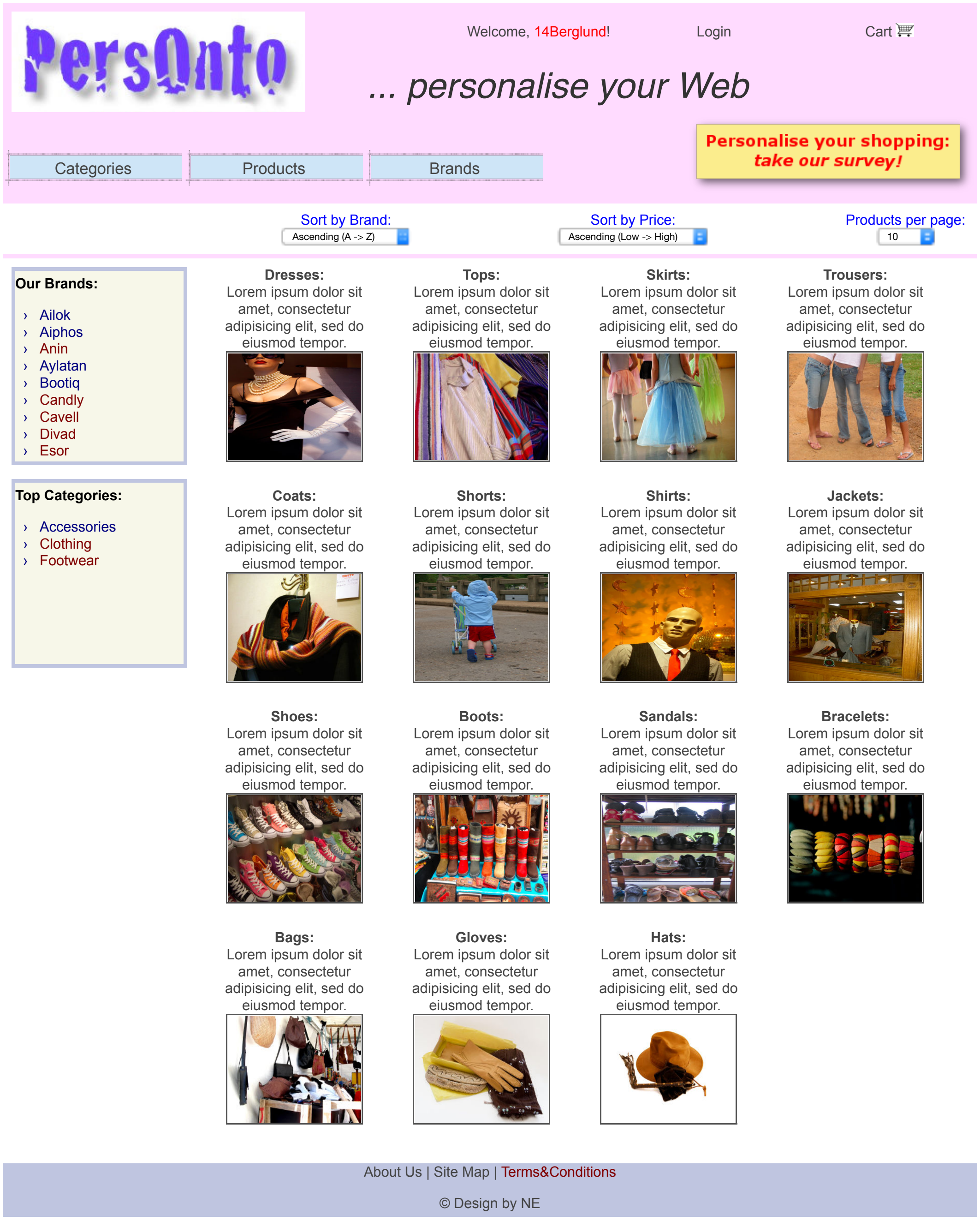


Figure 6.6: “Home” page for *“Maximiser-budget-textual”* profile.

In a similar way, we adapt other *PERSONTO* web pages according to the customer’s profile. For instance, Fig. 6.7 and Fig. 6.8 show “Categories” pages for *“Maximiser-quality-textual”* and *“Satisficer-quality-textual”*, respectively. While, Fig. 6.9 and

Fig. 6.10 illustrate what a “Single Product” page looks like for the *“Maximiser-budget-visual”* and a *“Satisficer-budget-visual”* profiles.

Figure 6.7: “Category” page for *“Maximiser-quality-textual”* profile.

Figure 6.8: “Category” page for *“Satisficer-quality-textual”* profile.

Figure 6.9: “Single Product” page for *“Maximiser-budget-visual”* profile.

Figure 6.10: “Single Product” page for *“Satisficer-budget-visual”* profile.

6.3 Assumptions and Limitations of Prototype

Our prototype was developed as a *'proof-of-concept'* prototype, and as is typical for such a system, several assumptions were made and the prototype has several limitations:

- Our focus is the *"technology"* perspective (§ 1.5), which is the main objective and focus of the project and also, arguably, the most challenging and effort- and time-consuming part of the implementation. The prototype implements the core functionality that we believed was most representative and attainable within the project's time frame and resources. In particular, we needed to establish that e-commerce website personalisation could be achieved by using an ontology-based customer profiling and reasoning approach.

  Other perspectives, and in particular, the *"business"* one (e.g. personalisation of marketing communication between the company and its customers, and product recommendations), were taken into account and implemented in the ontology, but not implemented in the current prototype.

- The e-commerce website in the prototype only contained the typical components of an e-commerce website (see § 6.2.2) since the development of a complete e-commerce website was not our aim or objective. Therefore, we coded only the web pages and functionality, required to support the prototype's purpose. (Details of this implementation are given in § 6.2.2.)

- The prototype only covered the "browsing phase" of the online purchasing process, i.e. searching for and selecting products. The "purchasing phase" (i.e. the checkout process) was not implemented, it is, however, taken into consideration in future development.

- Website personalisation is implemented as a “mass personalisation”, i.e. the customer profiling was achieved by using “group” and “non-fuzzy” profiling, when customers were categorised into specific predetermined profiles. Individual and “fuzzy” profiling was outside the scope of the project, but present interesting future research possibilities and challenges (§ 8.5).

- The customer profiling process in the prototype relied on survey-based self-assessment by a customer. In other words, the process was manual, user-driven and relied on explicit data. Automated profiling, for example, one that is based on analysis of customers activities or past history (i.e. implicit), was not implemented at this time. This is a limitation of the current implementation, which was the result of the need to avoid the *PERSONTO*'s *‘cold start’*. The potential limitations of the manual data collection approach (see § 2.4.1.1) were taken into consideration, but need further examination.

- It was assumed that customers had agreed to have their profiles created and stored. However, an option to delete a profile for customers, who did not wish their profile to be stored was provisioned in the system design (but not implemented). Similarly, an option for customers to refuse permission to have their profile created can be incorporated in the future.

6.4 Chapter Summary

The chapter described and discussed the outcomes from our experimental implementation of *PERSONTO*'s prototype. Using the specified system requirements, top-level architecture, system design and functionality (Chapter 3) and implementation of the *OntoProfi* ontology (Chapter 4), a *‘proof-of-concept’* prototype demonstrated the feasibility of the proposed approach to personalise e-commerce websites by invoking ontological user profiling. The limitations of the prototype were also discussed.

Chapter 7

Evaluation

7.1 Introduction

This chapter provides an evaluation of the work presented in the thesis. The evaluation strategy is based on "Design-Science" guidelines, proposed in (Hevner et al. 2004). The research aim and objectives are analysed and assessed against the achievements of the project. We also evaluate and discuss user acceptance of the proposed approach by analysing the data collected through a survey with potential users of *PERSONTO*. The concluding part of the chapter discusses the research limitations of our work.

7.2 Evaluation Strategy

As stated in § 1.3, this research followed the *"design science"* paradigm. Hevner et al. (2004) proposed a set of guidelines on how to conduct and evaluate this type of research (see Fig. 7.1). These guidelines formed the basis of our evaluation strategy.

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Figure 7.1: Design-Science Research Guidelines (Hevner et al. 2004)

The work presented in this thesis, has been based on the following research aim:

> ***to propose, test and evaluate the feasibility and acceptability of a flexible, extendible and re-usable ontology-based e-commerce personalisation approach.***

The above aim contains the following four distinct statements that have been investigated in this thesis and are the subject of the evaluation presented in this chapter:

STMT.1: feasibility of e-commerce website personalisation based on ontology-based customer profiling;

STMT.2: flexibility and extendibility of such a system;

STMT.3: reusability of the approach with respect to its system design and implementation.

STMT.4: acceptance of such a system and approach by the relevant stakeholders, i.e. end users (customers of an e-commerce company) and system "owners" (both: e-commerce companies and developers).

Table 7.1 summarises the application of Hevner et al.'s (2004) guidelines to this research.

Guideline (Hevner et al. 2004)	**Research Statement**	**Covered In**
Guideline 1	STMT.1, STMT.2, STMT.3	Chapter 3, § 5.3, § 4.3 and Chapter 6
Guideline 2	STMT.4	§ 7.6
Guideline 3	STMT.4	§ 7.6
Guideline 4		§§ 1.4 and 8.4
Guideline 5		§ 7.7 and Appendix G
Guideline 6		throughout this project and thesis
Guideline 7		throughout this thesis

Table 7.1: Application of Hevner et al.'s (2004) Guidelines to This Research

Using Hevner et al.'s (2004) guidelines, the assessment of the **STMT.1**, **STMT.2** and **STMT.3** statements relates to *"Guideline 1"*. We evaluate the results of our empirical and experimental (developmental) undertakings, and in particular: the proposed approach to personalisation of e-commerce websites (Chapter 3); findings from the case study (§ 5.3); design and development of the customer profiling ontology, *OntoProfi* (§ 4.3); and implementation of *PERSONTO*'s prototype (Chapter 6). The three statements are discussed in §§ 7.3 to 7.5.

The **STMT.4** statement (i.e. *"acceptance of such a system by the relevant stakeholders"*) relates to Hevner et al.'s (2004) *"Guidelines 2"* and *"Guidelines 3"*, and is evaluated by a survey of the potential end users (customers) and a sample of e-commerce business owners and website developers. The evaluation of this statement is discussed in § 7.6.

The research contributions (*"Guideline 4"*) are discussed in §§ 1.4 and 8.4, while the research quality and rigour (*"Guideline 5"*) is covered in § 7.7 and Appendix G .

The principles of *"Guideline 6"* and *"Guideline 7"* have been applied throughout the project and thesis.

7.3 Feasibility of Ontology-Based E-Commerce Website Personalisation

Feasibility of ontology-based e-commerce website personalisation has been demonstrated by proposing an approach to the design and implementation of *PERSONTO* and deploying a 'proof-of-concept' prototype (Chapters 3 and 6, respectively). The website personalisation in *PERSONTO* is achieved by incorporating an e-customer profiling ontology, *OntoProfi* (Chapter 4) within an exemplar e-commerce system.

More specifically, we have shown that it is feasible to:

- design and implement an ontology to profile e-commerce customers (Chapter 4);
- connect the *OntoProfi* ontology and the customer profiles repository to the standard e-commerce systems components, such as a website and back-end

systems, e.g. customer database (Chapter 6);

- profile customers into one of the predetermined customer profiles (Chapter 6);
- personalise various web pages (home page, product categories and single product pages) based on a particular customer profile that is computed by a reasoner (Chapter 6).

7.4 Flexibility and Extendibility

The proposed e-commerce website personalisation system, *PERSONTO*, was designed with flexibility and extendibility in mind. As specified in Chapter 3, the overall design of *PERSONTO* allows for changes (e.g. addition of new entities/ components) to be incorporated without much affect on or disruption to the overall system. This was already demonstrated by adding specific customer profiles, discovered by the online shopping case study and creating individual customer's instances (see § 5.4). In addition, it is possible to:

- Modify or extend the *OntoProfi* ontology by adding new classes, facts and axioms, for instance, to incorporate new business rules an e-commerce company may wish to implement. These are *'internal'* modifications or extensions, i.e. modifications of *OntoProfi* itself. The current version of *OntoProfi* already has some classes (see § 4.3.2) that were created to enable future development.
- Extend *PERSONTO* features and functionality to offer more personalisation features. This can be done by *'external'* modifications of *OntoProfi*, i.e. connecting to (aligning) or merging with other, external ontologies. This is achieved by using available ontology aligning and merging mechanisms, e.g. by asserting class equivalence in *OntoProfi* and other ontologies (for alignment), or by asserting the union of axioms from a set of ontologies (for merging).

- Deploy different tools and technologies. As explained in § 6.2.1, alternative programming languages, servers and ontology reasoners can be used to implement a system, analogous to the one that was demonstrated in our prototype - without affecting functionality. This shows the proposed design of *PERSONTO* is platform and technology independent.

7.5 Reusability of the Approach

The research work described in the thesis has been accomplished using a systematic and reusable approach (methodologies and techniques). More specifically, in this thesis we have specified, explained and discussed a:

- Conceptual approach to enabling website personalisation by adding several 'add-on', 'plug-and-playable' components, such as a customer profiling ontology and reasoner (see § 3.5)
- Systematic and reusable approach to conduct a case study (see § 5.2).
- Systematic and reusable approach to design and development of an e-commerce personalised system (see Chapter 3).

All the above components, methodologies and techniques in developing an e-commerce personalisation system can be adapted and reused in future projects by extending and adapting the *PERSONTO*'s entities (components).

7.6 Stakeholders' Acceptance

As stated in § 7.2, our evaluation strategy to assess stakeholders' acceptance of *PERSONTO* consisted of evaluating two groups of potential users: system end users (i.e. e-commerce customer) and system "owners" (i.e. e-commerce developers). This section explains the evaluation setup, process and procedures, and discusses the results of the evaluation for each group of stakeholders.

7.6.1 Evaluation of *PERSONTO*'s Acceptance by End User (Customer)

PERSONOTO's acceptance by potential end users (i.e. online shoppers) was evaluated by means of a survey to test user acceptance of *PERSONTO*, as a website personalisation system, by measuring users' attitudes towards such a system.

7.6.1.1 Survey Instrument

Our user acceptance survey was based on a paper-based questionnaire. The basis for our survey instrument was the *Technology Acceptance Model* (TAM) instrument developed by F. D. Davis (Davis 1989), which was later updated by Venkatesh & Davis (2000) to include *"intention to use"* variables. The instrument has been extensively used to assess a system's acceptance by its users. It has been also widely validated and replicated - as of September 2015 the citations count for F.D. Davis' original article amounted to 10568 [(1)], suggesting the instrument's high quality and wide acceptance. The instrument was deemed appropriate for the purpose of our own survey, since it was designed to measure user attitudes towards a new or updated system, which fitted the purpose of our survey.

(1) Source: Scopus (`http://www.scopus.com`)

The instrument measures three factors:

- *Perceived Usefulness (PU)*, described as a user belief as to whether a particular system helps with enhancing the job performance of the user. (Davis 1989)
- *Perceived Ease of Use (PEOU)*, which refers to the *"degree to which a person believes that a particular system would be free of effort"*. (Davis 1989)
- *Intention to Use (ITU)*, which is designed to measure the likelihood of users to use the new/ updated system. (Venkatesh & Davis 2000)

To construct our instrument, we adapted the original questions, proposed in (Davis 1989) and (Venkatesh & Davis 2000), by rewording these to make the questions specific to our project (such as using the word "*PERSONTO*" in each question), but keeping the original split into three factors intact. Two items from the original Davis' (1989) scales, Item No. 2 and Item No. 3, were dropped. These questions ask the respondents to evaluate whether they perceive that the system under study improves their job performance and increases their productivity. These two aspects were irrelevant to *PERSONTO*, and therefore were not tested. Instead, we included *"PU5"* Item to test the overall participants' attitudes towards *PERSONTO* by explicitly asking them whether they thought that personalisation of web pages was a good idea.

The adapted instrument resulted in containing 13 questions, with several questions in each of the above factors, as follows:

- *Perceived Usefulness (PU)* - 5 questions
- *Perceived Ease of Use (PEOU)* - 6 questions
- *Intention to Use (ITU)* - 2 questions

Table 7.2 shows the resulting questions.

Item	Question	Source
Perceived Usefulness (PU)		
PU1	Using *PERSONTO* and personalised webpages would enable me to do my online shopping more quickly.	(Davis 1989)
PU2	Using *PERSONTO* and personalised webpages would enable me to do my online shopping easier.	(Davis 1989)
PU3	Using *PERSONTO* and personalised webpages would enhance my effectiveness in online shopping.	(Davis 1989)
PU4	Overall, using *PERSONTO* and personalised webpages would be useful for my online shopping.	(Davis 1989)
PU5	Having webpages personalised is a good idea.	this research
Perceived Ease of Use (PEOU)		
PEOU6	Learning how to use *PERSONTO* and personalised webpages would be easy for me.	(Davis 1989)
PEOU7	I would find it easy to get *PERSONTO* to do what I want to do.	(Davis 1989)
PEOU8	My interactions with *PERSONTO* and personalised webpages would be clear and understandable.	(Davis 1989)
PEOU9	I would find *PERSONTO* and personalised webpages to be flexible to interact with.	(Davis 1989)
PEOU10	It would be easy for me to become skilful at using *PERSONTO* and personalised webpages.	(Davis 1989)
PEOU11	Overall, I would find *PERSONTO* to easy to use.	(Davis 1989)
Intention To Use (ITU)		
ITU12	Assuming I have access to *PERSONTO* (or a similar type of website, which allows for webpages personalisation), I intent to use it.	(Venkatesh & Davis 2000)
ITU13	Given that I have access to *PERSONTO* (or a similar type of website, which allows for webpages personalisation), I predict that I would use it.	(Venkatesh & Davis 2000)

Table 7.2: Survey Instrument Questions

Like the F. D. Davis' (1989) original response scale, our scale was a seven-point Likert scale. To test the level of the respondent's agreement, the following standard *"level of agreement"* options were used (arranged from negative to positive): *1: Strongly disagree; 2: Disagree; 3: Slightly disagree; 4: Neutral; 5: Slightly agree; 6: Agree; 7: Strongly agree.* A copy of the complete questionnaire is in Appendix F.

To check the instrument's **internal consistency and reliability**, Crobach's Alpha coefficients for the three groups of our constructs (*PU*, *PEOU* and *ITU*) as well as a cumulative coefficient for all constructs were calculated. Table 7.3 shows the resulting statistics:

Cronbach's Alpha	N of Items
.876	5

(a)

Cronbach's Alpha for PU constructs

Cronbach's Alpha	N of Items
.838	6

(b)

Cronbach's Alpha for PEOU constructs

Cronbach's Alpha	N of Items
.902	2

(c)

Cronbach's Alpha for ITU constructs

Cronbach's Alpha	N of Items
.908	13

(d)

Cronbach's Alpha for all constructs

Table 7.3: Cronbach's Alpha Reliability Coefficients

All of the above Cronbach's Alpha coefficients are greater than 0.7 ($\alpha > 0.7$) and two of them are greater than 0.9 ($\alpha > 0.9$), therefore, as commonly accepted, the instrument's reliability was considered as having a high level of internal consistency (Nunnally & Bernstein 1994).

7.6.1.2 Survey Data Collection: Research Setting and Procedures

A total of 150 paper-based questionnaires were distributed among undergraduate (BSc) and postgraduate (MSc) students. The students were taking a range of courses, including BSc Computer Science, BSc Business Information and all MSc programmes in Cardiff's School of Computer Science & Informatics [(1)]. A total of 65 (N = 65) completed questionnaires were collected, resulting in a response rate of approx. 43% [(2)]. Two of the returned questionnaires contained incomplete answers and were discarded, yielding 63 usable responses (N = 63).

Table 7.4 shows the number and percentages of each group of respondents that returned the completed questionnaire:

Respondents Group	***N***	**%**
Undergraduate (BSc) Students	39	60%
Postgraduate (MSc) Students	26	40%
Total	**65**	**100%**

Table 7.4: PERSONTO Survey Respondents

The questionnaire was administered in a lab setting. The lab was equipped with PCs running Windows 7 operating system and an Internet access, which was required to access the *PERSONTO*'s prototype.

The respondents were provided with a brief explanation of the purpose of the survey and were given the questionnaire, which contained the instructions of what they needed to do (see Appendix F). To start the evaluation, the respondents were required to access the prototype via the given URL. After this, they were invited to

(1) `http://www.cs.cf.ac.uk/degreeprogrammes/`
(2) There is no consensus on what constitutes an average or a *'good'* response rate. For example, a comparative analysis of the survey of the survey response rate in academic studies by (Baruch 1999) reported the average of 55.6% with a standard deviation of 19.7, the median was 60% and mode 45%. However, a decline in the response rate through the years, difference between journals and countries, where research was conducted, were also reported.

complete a three-question *PERSONTO*'s *'Profiling Survey'* (see Fig. 3.7) so that the system could profile them into one of the profiles. On completion of the survey, the respondents were presented with the web pages, personalised for them based on their profile, and were asked to view these. Lastly, they were invited to give their opinions on the web pages' personalisation by answering the questions posed by the questionnaire.

Due to the fact that the server, available in the lab, was not able to parse JSP pages, which were used in the original implementation of *PERSONTO* (§ 6.2.1), the prototype was adapted to suit the experimental lab setting. All personalised web pages were *'hard coded'* in HTML, and the survey participants were shown the relevant ones based on their responses to the *"Profiling Survey"* (see Fig. 3.7), which were used to profile the participants into one of the eight customer profiles (see § 5.3.3.2).

Such prototype adaptation was regarded as appropriate and valid because the adaptation could not have affected or biased the participants' responses in any way - the resulting personalised web pages, which the respondents evaluated, were identical to the ones we discussed in § 6.2.4.2. Moreover, the purpose of the survey was to evaluate the users' attitudes, and not the system implementation - we covered the latter by the prototype development (see Chapter 6) and corresponding evaluation of feasibility of our proposed approach to e-commerce personalisation (see § 7.3).

7.6.1.3 Analysis and Results

Analysis of the collected data focused on the aim of the survey, i.e. testing the level of e-customers' acceptance of the website personalisation, which involved descriptive statistics.

Table 7.5 displays central tendency and variability; and Table 7.6 shows frequency distribution of all survey items and their corresponding bar charts:

		PU1	PU2	PU3	PU4	PU5
N	Valid	63	63	63	63	63
	Missing	0	0	0	0	0
Median		5.0000	5.0000	5.0000	5.0000	6.0000
Mode		6.00	5.00[a]	6.00	6.00	6.00
Range		5.00	5.00	5.00	5.00	4.00
Percentiles	25	4.0000	4.0000	4.0000	4.0000	5.0000
	50	5.0000	5.0000	5.0000	5.0000	6.0000
	75	6.0000	6.0000	6.0000	6.0000	7.0000

a. Multiple modes exist. The smallest value is shown

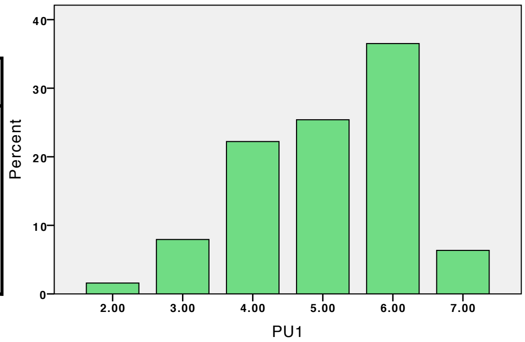
		PEOU6	PEOU7	PEOU8	PEOU9	PEOU10	PEOU11
N	Valid	63	63	63	63	63	63
	Missing	0	0	0	0	0	0
Median		6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
Mode		6.00	6.00	6.00	6.00	6.00	6.00
Range		4.00	4.00	4.00	4.00	4.00	6.00
Percentiles	25	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	50	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
	75	7.0000	6.0000	6.0000	6.0000	6.0000	7.0000

		ITU12	ITU13
N	Valid	63	63
	Missing	0	0
Median		5.0000	5.0000
Mode		6.00	6.00
Range		6.00	6.00
Percentiles	25	4.0000	4.0000
	50	5.0000	5.0000
	75	6.0000	6.0000

Table 7.5: *PERSONTO* Customer Acceptance Survey: Central Tendency and Variability

PU1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	1	1.6	1.6	1.6
	3.00	5	7.9	7.9	9.5
	4.00	14	22.2	22.2	31.7
	5.00	16	25.4	25.4	57.1
	6.00	23	36.5	36.5	93.7
	7.00	4	6.3	6.3	100.0
	Total	63	100.0	100.0	

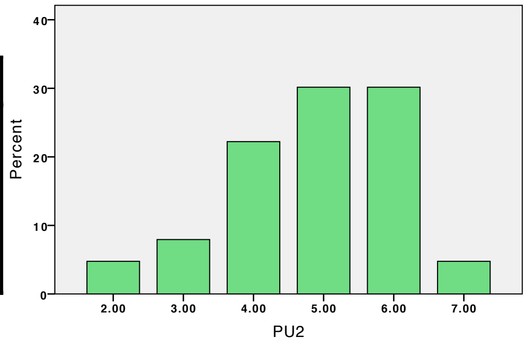


(a) **PU1** Item

"*PERSONTO* enables to do online shopping more quickly" (1)

PU2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	3	4.8	4.8	4.8
	3.00	5	7.9	7.9	12.7
	4.00	14	22.2	22.2	34.9
	5.00	19	30.2	30.2	65.1
	6.00	19	30.2	30.2	95.2
	7.00	3	4.8	4.8	100.0
	Total	63	100.0	100.0	

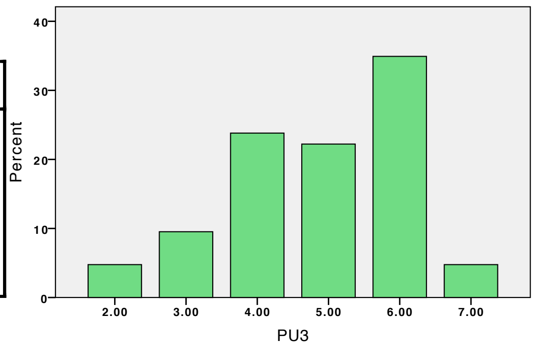


(b) **PU2** item

"*PERSONTO* enables to do online shopping easier"

PU3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	3	4.8	4.8	4.8
	3.00	6	9.5	9.5	14.3
	4.00	15	23.8	23.8	38.1
	5.00	14	22.2	22.2	60.3
	6.00	22	34.9	34.9	95.2
	7.00	3	4.8	4.8	100.0
	Total	63	100.0	100.0	



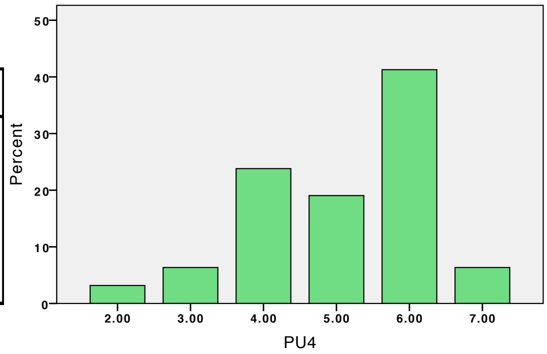
(c) **PU3** item

"*PERSONTO* enhances online shopping effectiveness"

(1) These captions represent a short summary, reflecting the essence of each question. See Table 7.2 for full wording of each question.

PU4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	2	3.2	3.2	3.2
	3.00	4	6.3	6.3	9.5
	4.00	15	23.8	23.8	33.3
	5.00	12	19.0	19.0	52.4
	6.00	26	41.3	41.3	93.7
	7.00	4	6.3	6.3	100.0
	Total	63	100.0	100.0	

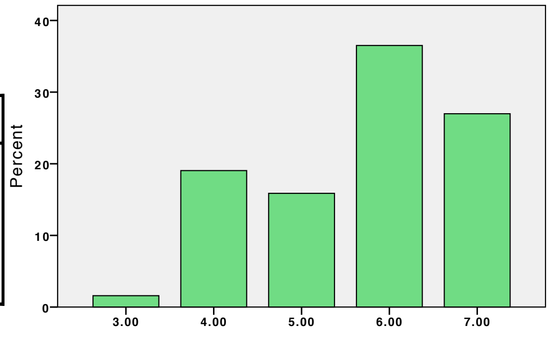


(d) **PU4** item

"*PERSONTO* and personalised web pages would be useful for online shopping"

PU5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	1.6	1.6	1.6
	4.00	12	19.0	19.0	20.6
	5.00	10	15.9	15.9	36.5
	6.00	23	36.5	36.5	73.0
	7.00	17	27.0	27.0	100.0
	Total	63	100.0	100.0	

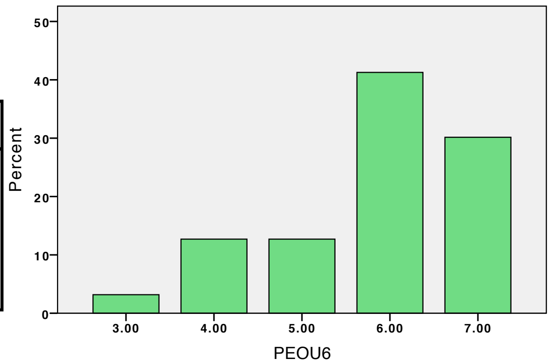


(e) **PU5** item

"Personalised web pages is a good idea"

PEOU6

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	2	3.2	3.2	3.2
	4.00	8	12.7	12.7	15.9
	5.00	8	12.7	12.7	28.6
	6.00	26	41.3	41.3	69.8
	7.00	19	30.2	30.2	100.0
	Total	63	100.0	100.0	

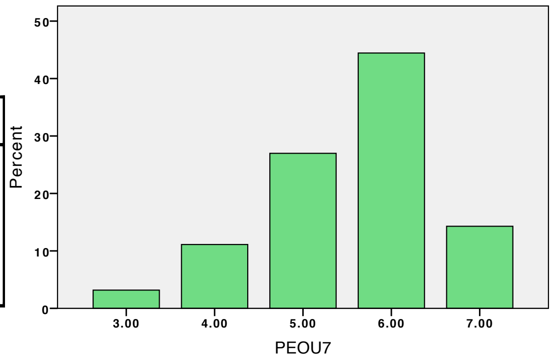


(f) **PEOU6** item

"Learning how to use *PERSONTO* and personalised web pages would be easy"

PEOU7

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	2	3.2	3.2	3.2
	4.00	7	11.1	11.1	14.3
	5.00	17	27.0	27.0	41.3
	6.00	28	44.4	44.4	85.7
	7.00	9	14.3	14.3	100.0
	Total	63	100.0	100.0	

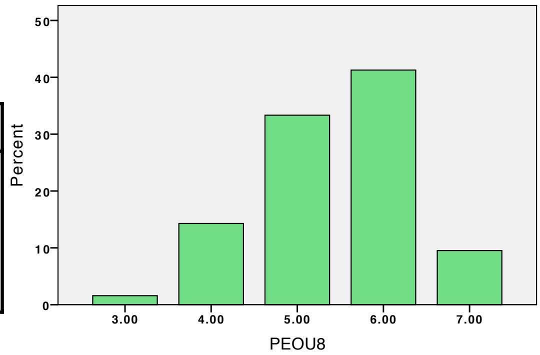


(g) **PEOU7** item

"Easy to get *PERSONTO* to do what the user wants to do"

PEOU8

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	1.6	1.6	1.6
	4.00	9	14.3	14.3	15.9
	5.00	21	33.3	33.3	49.2
	6.00	26	41.3	41.3	90.5
	7.00	6	9.5	9.5	100.0
	Total	63	100.0	100.0	

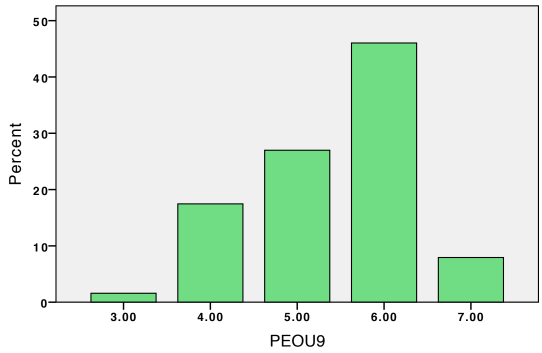


(h) **PEOU8** item

"Interactions with *PERSONTO* would be clear and understandable"

PEOU9

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	1.6	1.6	1.6
	4.00	11	17.5	17.5	19.0
	5.00	17	27.0	27.0	46.0
	6.00	29	46.0	46.0	92.1
	7.00	5	7.9	7.9	100.0
	Total	63	100.0	100.0	

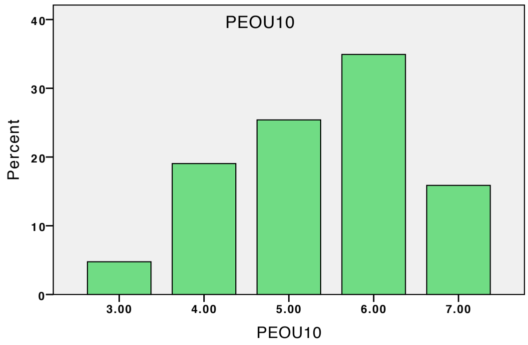


(i) **PEOU9** item

"*PERSONTO* would be flexible to interact with"

PEOU10

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	3	4.8	4.8	4.8
	4.00	12	19.0	19.0	23.8
	5.00	16	25.4	25.4	49.2
	6.00	22	34.9	34.9	84.1
	7.00	10	15.9	15.9	100.0
	Total	63	100.0	100.0	

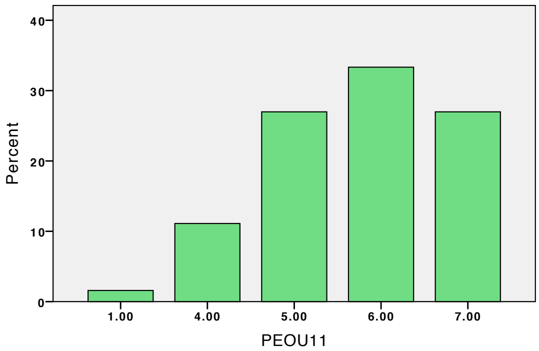


(j) **PEOU10** item

“It would be easy to become skilful at using *PERSONTO*”

PEOU11

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	1.6	1.6	1.6
	4.00	7	11.1	11.1	12.7
	5.00	17	27.0	27.0	39.7
	6.00	21	33.3	33.3	73.0
	7.00	17	27.0	27.0	100.0
	Total	63	100.0	100.0	

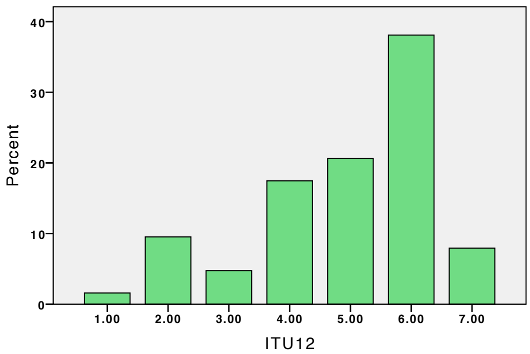


(k) **PEOU11** item

“Overall ease of use”

ITU12

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	1.6	1.6	1.6
	2.00	6	9.5	9.5	11.1
	3.00	3	4.8	4.8	15.9
	4.00	11	17.5	17.5	33.3
	5.00	13	20.6	20.6	54.0
	6.00	24	38.1	38.1	92.1
	7.00	5	7.9	7.9	100.0
	Total	63	100.0	100.0	

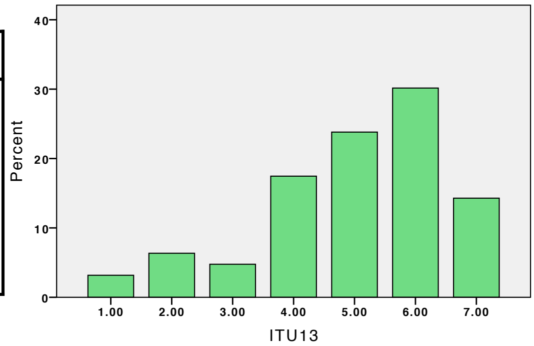


(l) **ITU12** item

“I intend to use *PERSONTO*”

ITU13

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	2	3.2	3.2	3.2
	2.00	4	6.3	6.3	9.5
	3.00	3	4.8	4.8	14.3
	4.00	11	17.5	17.5	31.7
	5.00	15	23.8	23.8	55.6
	6.00	19	30.2	30.2	85.7
	7.00	9	14.3	14.3	100.0
	Total	63	100.0	100.0	



(m) **ITU13** item
"I would use *PERSONTO*"

Table 7.6: PERSONTO Customer Acceptance Survey: Frequency Distribution of Survey Items and Corresponding Bar Charts

As seen in the Tables 7.5 and 7.6, the median of subjects' responses to all survey items ranges between **5.00** (*"Slightly Agree"*) to **6.00** (*"Agree"*). The majority of survey items' responses have a mode of **6.00** - with the exception of item "PU2", which has a multiple mode of **5.00** and **6.00**. The range varies between **4.00** (*"Neutral"*) and **6.00**. We can also note that the lower quartile (Q1) lies between **4.00** ($\approx$ *46%* of all survey items) and **5.00** ($\approx$ *54%*), and the third quartile (Q3) is between **6.00** ($\approx$ *77%*) and **7.00** (*"Strongly Agree"*) ($\approx$ *23%*).

Combining the *"Perceived Usefulness"* ("PU"), *"Perceived Ease of Use"* ("PEOU") and *"Intention to Use"* ("ITU") items into multiple response sets, shows us a similar level of acceptance, with the *"Perceived Ease of Use"* set of questions rated higher (83.1% for *"Slightly Agree"* and above, and 57.7% for *"Agree"* and above), followed by the *"Perceived Usefulness"* (68.2% for *"Slightly Agree"* and above, and 45.7% for *"Agree"* and above), and the *''Intention to Use"* (67.4% for *"Slightly Agree"* and above, and 45.2% for *"Agree"* and above).

Similarly, if we treat all survey items as one single set (see Table 7.7), the analysis of the frequency distribution shows that 74.8% of the respondents answered *"Slightly*

Agree" and above to any of the survey questions, with 51% choosing *"Agree"* and *"Strongly Agree"* answers.

All Items Frequencies

		Responses	
		N	Percent
All Items[a]	1.00	4	0.5%
	2.00	19	2.3%
	3.00	36	4.4%
	4.00	146	17.8%
	5.00	195	23.8%
	6.00	308	37.6%
	7.00	111	13.6%
Total		819	100.0%

a. Group

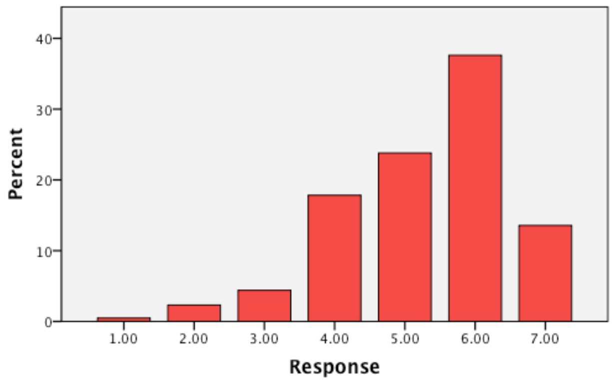


Table 7.7: PERSONTO Customer Acceptance Survey: Frequency Distribution of All Survey Items.

These figures suggest a high level of acceptance of *PERSONTO* as a website personalisation system by these potential users. Acceptance of *PERSONTO* by the survey participants also appears to be supported by the additional comments that some respondents made under the *"Other Comments"* section of the questionnaire, for example [(1)]:

> *"Many websites already attempt to personalise experiences with cookies... However, I found it interesting that this system changes the actual*

(1) Respondents' original grammar and punctuation were preserved.

elements sizes & priorities on screen, unlike the standard methods today where layout is more or less consistent.”

“I really liked the idea...”

“Sort by price is usually available on sites I use eBay & Amazon, the brief text or full description is a great idea and the picture vs. text. If I only had to personalise once and it applied to full web pages, it would be great. (1)”

7.6.2 Evaluation of *PERSONTO*’s Acceptance by E-Commerce Developers

The acceptance of *PERSONTO* by e-commerce developers was evaluated by a questionnaire. The aim of this evaluation was to gauge the buy-in of the proposed approach to e-commerce personalisation.

7.6.2.1 Participants, Methodology and Procedure

To carry out the evaluation, two industry representatives of two different design companies were asked to participate. Chris Green (2) is a creative director for *Company A*, a large multi-national interactive group of companies. Joe Brown is a digital manager for *Company B*, a small design company. Both companies offer design, implementation and maintenance of e-commerce websites. Chris Green and Joe Brown have experience in designing for and managing a variety of web development projects, from small to large in scale.

(1) NB: Personalisation in PERSONTO is designed to do just that.
(2) The names of respondents and companies have been changed to preserve anonymity.

A questionnaire approach was adopted because the respondents stated this was the only suitable approach for them to provide evaluation due to pressures of their jobs.

The respondents were sent a summary of the project, which explained the project's purpose, process and outcomes, and were asked to complete a questionnaire, which consisted of two parts:

- **Part A** asked the respondents to provide the facts about their companies, such as whether they offered any personalisation features or functionality as part of developing websites for their clients, whether they involved potential users in website design, implementation and testing as well as the methods and number of users involved.
- **Part B** asked the respondents to provide their opinions on *PERSONTO*, as an approach to design and implementation of personalisation in e-commerce systems. The questions in this part sought to gather the industry representatives' opinions on the acceptability and usefulness of the approach, proposed in the project.

On receipt and analysis of the responses, the respondents were contacted again with follow-up questions to clarify and expand on some of their answers.

7.6.2.2 Findings and Discussion

Both companies offer personalisation features and functionality, albeit in a limited way. *Company A* offers bespoke design, responsive interfaces and branding. They also offer personalisation of data and admin system management for user accounts such as bespoke access permissions based on the client's requirements and needs. *Company B* offers basic personalisation of content based on selected areas of interests (e.g. the geographical location of the client). This is achieved by using various filters in the Content Management System's (CMS) back-end.

Both companies involve users in system design and development. The users are selected as a non-random, convenience sample, and often involve the same people, participating in different projects. However, the methods and number of users involved varies. *Company A* uses focus groups, individual interviews, workthroughs and testing (such as wireframe design and evaluation). *Company B* only uses workthroughs and testing.

The number of prospective users involved in website design and development differs between the two companies. Both respondents report a minimum of 1 participant. However, the maximum varies between 15 (*Company B*) and 57 (*Company A*). On average *Company B* uses 4 *'evaluators'*, while *Company A* uses 8 -10. This difference might be explained by the difference between the companies in terms of their size - as the bigger the company and project size is, the more resources it typically has for a larger involvement of the system end-users.

With regard to the questions on whether the proposed approach to e-commerce personalisation could be considered useful and whether the approach would be considered for adoption in their website design and development, both respondents gave positive answers, commenting (1):

> *"Can see PERSONTO as being an extension to customer lifecycle marketing. Allowing the full website to be personlised to the end user in multiple dimensions, rather than simply displaying content "related" to previous purchases etc."*

> *"Personalised content and interfaces are at the very heart of capturing to-*

(1) ***Please note*** that these are direct quotes, and original English/ grammar and punctuation used by the respondents are preserved.

day's online audiences and consumers, harnessing the power of big data and granular, progressive user profiles is key to providing an any screen, anywhere experience which tailors content down to single user level."

"All aspects of the approach would be useful to inform an optimised user/-customer experience."

The most useful features of *PERSONTO* were identified as personalisation of user interface and tools (particularly, based on the level of IT competence) and the ease of integration with the back-end components and systems.

The respondents would consider adopting the following aspects and features of *PERSONTO*:

- "Auto personalisation of content display in terms of image size prioritisation - as we work with many audiences in a primarily visual way and this would enable new levels of personalisation within a feasible cost."

- "Customer profiling - creates a better understanding of the customer needs on a per site basis and removing the "assumed" knowledge. Ability to build a library of profiles which can be used across projects to deliver tried and tested personlisation."

- "Personalised payment user flow - as this is a new offering which is traditionally very linear and generic. This would be a differentiator for us as a business." (1)

(1) This was not proposed explicitly in *PERSONTO*, and appears to be an assumption of the respondent. However, the proposed future work to extend personalisation to cover the *"business perspective"* includes payment personalisation as a part of personalisation of the order fulfilment processes and activities (see § 8.5).

Other comments, made by the respondents, provided interesting insights into the industry's concerns.

> *This is an extremely competitive market and the risks associated would not be with the concept itself but more with choosing and working with technologies which are not only interoperable but which are both robust and committed to long term existence as well as platfrom/screen agnostic.*

> *Could the system be developed to automatically update / amend the customer profiles based on user history / conversions / leads etc..? Once concern would be adding another back-end element to maintain and manage as cutomer needs change / expand.*

The first comment reinforces the fact that the industry needs to ensure the technologies they use to develop systems are interoperable, available and mature. The second comment demonstrates the industry's concern for the need to minimise efforts and costs associated with developing new systems. These comments support our choice of using the Semantic Web technologies, such as ontology, and the proposed *'plug-and-playable'* architecture.

Overall, the responses by the industry representatives are encouraging and suggest the industry *'buy-in'* (acceptance) of *PERSONTO* as an approach to personalise e-commerce.

7.7 Research Limitations

The work presented in this thesis has successfully addressed its set objectives (§ 1.2) by demonstrating feasibility and acceptance of designing and implementing an ontology-based system to personalise e-commerce. However, our research has some limitations, such as those related to our research approach, project focus, system design and implementation, and evaluation.

7.7.1 Research Approach

One of the limitation of this research project arises from adopting a mixed research approach that combined qualitative and quantitative research techniques. In particular, a substantial part of our research, and specifically our case study to elicit e-shoppers profiles (Chapter 5) and subsequent modification of `Customer_profile` classes in OntoProfi ontology (§ 5.4.1), relied on a qualitative, interpretivist approach to data collection and analysis, which has been criticised as having the possibility of introducing subjectivity and bias, lacking quality and rigour, and hindering generalisation or replication of research findings (e.g. (Oates 2006), (Yin 2003)).

Therefore, utmost care was taken to mitigate these issues through the use of established principles and techniques, thus ensuring the quality and rigour of qualitative, interpretivist research. In particular, the case studies followed a methodology, which was explicitly specified prior to commencement of the research and described in detail in this thesis. The interview transcripts were rigorously and diligently checked for correctness of transcription. The analysis of data followed the established qualitative, interpretivist techniques of *"staying close to data"* and *"comparing data with data"* (Charmaz 2006, pp. 42-71). The discussion of findings provided detailed, *'rich'* (*'thick'*) descriptions using the actual data.

One particular widely-discussed issue that relates to the use of an interpretivist approach is the potential issue of generalisation of findings. Although it is true that it is highly improbable that the exact outcomes from our case study can be replicated, nonetheless, another case study can replicate the proposed methodology so that the outcomes of the two case studies can be compared. Moreover, despite the fact that it is possible that some (if not most) e-commerce customers might exhibit unique characteristics and traits, the identified profiles (classes) proposed in the *OntoProfi* ontology appeared to represent reasonably general types of online customers that are hypothesised to be applicable to most e-commerce customers. In addition, OntoProfi can also be easily extended based on outcomes of a new, different case study (§ 7.4).

Appendix G provides a detailed summary of the evidence of how the established interpretivist research principles and techniques for improving the research quality and rigour were employed in our research.

The choice of sampling technique, and also, possibly, and number and the choice of respondents (see §§ 5.2.1 and 5.2.2) were other limitations in this project. Using a *"mixed purposeful"* sampling approach and small sample size has a potential for introducing bias and a danger of not collecting enough (complete) information (Miles & Huberman 1994). Nevertheless, with regard to the sampling technique, as already discussed in § 5.2.1, this was considered appropriate and acceptable in this research for the reasons of commonality of its use in qualitative research, the guarantee of production of rich data and the assumed suitability for its use in commercial development. As the author's personal communications confirmed [Pers.Comm.1], [Pers.Comm.2], [Pers.Comm.3], it is a common practice in the industry to use non-random convenience sampling techniques.

With regard to the number of respondents, a low number of respondents in qualitative research, including case studies, is considered to be acceptable. For example, Onwuegbuzie & Leech (2007) advised to give more consideration to collecting sufficient number of words than to a sample space. Eisenhardt (1989) argued that "while there is no ideal number of cases, a number between 4 and 10 cases usually works well" and "with more than 10 cases, it quickly becomes difficult to cope with complexity and volume of data". Sarker, et al. (2013) found that among the case studies they reviewed 52% used 1 case unit, 13% - 2 and 9% - 3 units. Moreover, as the author's personal communications confirmed ([Pers.Comm.1], [Pers.Comm.2], [Pers.Comm.3]), the use of a low number of prospective system users in commercial practice is not uncommon.

The lack of representation of 65+ participants is also a potential limitation of this research. Recently, this age group has seen large growth in the overall rate of online purchases - from 16% in 2008 to 42% in 2015 (UK Office for National Statistics (ONS) 2015). However, this group is still the smallest one in all shopping categories (clothes, video games, household goods, computer hardware, etc.). The biggest age group is the 35 - 44, which reported the highest (or joint highest) rate of online purchasing in 11 out of 15 shopping categories; this age group is also the one, which makes purchases more often (UK Office for National Statistics (ONS) 2015). Nevertheless, given the growth rate of 65+ online shoppers, this group needs to be considered in future research.

The sampling techniques, number of participants and age group representation are considered acceptable for achieving this project's aim and objectives, but more detailed examination of the potential effects is needed.

The data for the case study was collected over a long period of time. However, it is not believed to have affected the case study outcomes. Internet shopping has changed the way we shop, but the actual process of purchasing online has remained 'stable' from the inception of e-commerce, and is reflected in the standard set of "tools" and functionality (product catalogue, shopping cart, payment processing), which every e-commerce website offers. For this reason, we believe that the time it took us to collect the data did not have an effect on our findings, but this requires further exploration.

The quantitative part of this work also has limitations. This part was represented by a user evaluation survey to test acceptance of the proposed e-commerce personalisation by its potential users (§ 7.6). The survey data was treated as ordinal and could only be analysed using descriptive statistics. The use of ordinal data allowed us to analyse and rank each survey category (question) to establish the trends in the respondents' attitudes to the proposed personalisation approach. However, the ordinal data does not allow the determination of the exact measurement of the difference between the options, e.g. whether the difference (distance) between *"Neutral"* and *"Slightly Agree"* is weaker (or stronger) than between *"Slightly Agree"* and *"Agree"*. The use of descriptive statistics only allows us to describe (summarise) the data, but does not permit drawing of any inferences, in particular, to investigate or determine causal relationships. So, while the use of ordinal data and descriptive statistics was appropriate owing to the set objective of the survey, further investigation into user attitudes and behaviour with regard to the proposed personalisation system (e.g. by using inferential statistics) would provide additional insights in future research (§ 8.5).

7.7.2 Project Focus

The project's focus is on B2C e-tailer type of e-commerce as the most prominent business model online (see § 1.5). Nevertheless, the research findings can also be applicable to the design and implementation of other e-commerce types, e.g. to B2B and C2C since the technologies used in e-commerce are the same regardless of the business market relationship or type.

7.7.3 System Design and Implementation

The system requirements and the resulting system design were based on the analysis of the project's aims and objectives, the limitations of extant research and practices, as well as our own expertise in e-commerce. Verification and validation of the requirements by industry representatives was not carried out due to the difficulty in gaining access to such representatives. The future work would benefit from the industry involvement.

The scope and focus of the project (§ 1.5) determined the main research emphasis being on design and implementation of the most important, representative and challenging functionality. Thus, a complete set of operations that *PERSONTO* could provide was not tested. Similarly, *PERSONTO*'s prototype was deployed as a *'proof-of-concept'* that aimed to prove feasibility of implementing the proposed system and, therefore, only had limited functionality. In addition, the prototype was implemented under *'lab'* conditions (on a local server), therefore, robustness (system recovery), reliability or efficiency (e.g. speed, response time or ease of use) in a commercial environment was not tested. Other non-functional system design and development aspects (such as: usability, security robustness and data quality assurance) were not tackled in this project as these were not believed to have an effect on our testing of the feasibility of implementing the functional components. The investigation of non-functional requirements presents an opportunity for future work (§ 8.5).

7.7.4 Evaluation

Evaluation of acceptance of the proposed approach was carried out as a preliminary (initial) investigation into a *'general'* level of acceptance. The limitations of our **"End User (Customer)" evaluation** stems from the use of descriptive statistics methods and, as discussed earlier, requires expansion to include the inferential statistics approaches.

Another potential limitation is the use of students for evaluating the acceptance of the proposed system. However, the use of students in research is widespread, and well-documented and appraised; and, along with potential limitations, might offer several, advantages, such as: the students are easily approachable; and are more used to completing various survey, which reduces problems with comprehension of questions and instructions, compared with non-student respondents (Thirumalai & Sinha 2009). The students are even seen as *"a desirable group"* to use in e-commerce research because they are *"more likely to be technologically savvy and thus more likely to make online purchases"* (Rose, et al. 2005).

The third potential issue is the concern that the chosen 'end-user' sample had not explored all eight profiles. To examine whether this was the case, we conducted two workshops with MSc students. The first workshop was with the subgroup of students, who completed evaluation of *PERSONTO* acceptance survey (§ 7.6.1.2). The total number of students in this group was 17. The second workshop involved a group of 15 MSc students; these did not participate in the survey.

In both workshops, the students were asked to fill in a questionnaire, which prompted them to select their preferences for products, information type and quantity, i.e. the same customer profiling categories used in this research. Tables 7.8 and 7.8 summarise the students' responses [(1)]

Question	**Responses**			
What is more important for you: price or quality?	price:	**8**	quality:	**11**
What type of information do you prefer?	textual:	**10**	visual:	**13**
How much information do you prefer to see on a web page?	brief:	**5**	detailed:	**11**

Table 7.8: MSc Students Workshop ONE Responses *(N=17)*

(1) The workshop participants were not explicitly asked to select only one response for each question, which resulted in several students choosing both responses.

Question	Responses			
What is more important for you: price or quality?	price:	**5**	quality:	**12**
What type of information do you prefer?	textual:	**8**	visual:	**10**
How much information do you prefer to see on a web page?	brief:	**3**	detailed:	**13**

Table 7.9: MSc Students Workshop TWO Responses *(N=15)*

The analysis of the responses from the two workshops shows that, indeed, the students from two separate groups can be personalised into eight distinct profiles, which, in turn, is a strong indication that our evaluation of acceptance of *PERSONTO* did not have an issue of having a *'too homogeneous'* group of respondents that could only be profiled into one or two profiles. (The two tables suggest that there is a potential bias towards *"quality"*, *"visual"* and *"detailed"* categories; this presents an opportunity for future research.)

Our **"E-Commerce Developers" acceptance** is limited due to the use of the qualitative approach and small number of respondents. This reflects the difficulty of gaining access to industry representatives, who, owing to business pressures, are somewhat unwilling to participate in research projects. Nevertheless, further expansion to include more respondents and the use of other evaluation methods and approaches (e.g. focus groups, in-depth interviews, qualitative surveys and walkthroughs) will provide more detailed examination of the level of acceptance of the proposed e-commerce personalisation approach, and is worthwhile given the positive responses by the respondents.

7.8 Chapter Summary

This chapter presented an evaluation of the project outcomes against the project aim and objectives. In particular, we evaluated the fulfilment of the statements made by the main aim of this research, such as feasibility, flexibility and extendibility of the proposed system to personalise e-commerce; the use of a systematic and reusable approach to such system design and development; and acceptance of the proposed system by end users (e-shoppers) and e-commerce businesses and developers. The results of our analysis of the statements proposed in the project's research aim showed that all these statements have been successfully achieved and demonstrated. Evaluation of the *'buy-in'* by end-users and industry representatives suggested a good level of acceptance of the system proposed in this research. The concluding part highlighted the limitations of our research, some of which are the subject of proposed future work in the next chapter.

Chapter 8

Conclusions and Future Work

8.1 Introduction

This thesis has presented, discussed and evaluated a novel approach to e-commerce website personalisation that is achieved by deployment of ontological customer profiling. In this concluding chapter, we provide a synopsis of the key concepts and ideas examined in the thesis, and evaluate the fulfilment of our research objectives. The chapter concludes with discussion of implications of the outcomes of this work to research and practice, and exploration of potential future development.

8.2 Synopsis

In this thesis, we have presented a novel approach to e-commerce personalisation that relies on ontology-based customer profiling.

It was argued that on the one hand, web personalisation in e-commerce was proven to offer online business and customers a wealth of benefits and advantages, but on the other hand, more approaches could be developed and tested, which offer feasible, flexible, extendible and acceptable ways to personalise e-commerce, and which potentially could improve the customer's experience. A survey of academic literature and professional practice demonstrated that there exists a wealth of tools and techniques to enable e-commerce personalisation. However, there is also a need to further develop some areas and aspects, which have limitations, such as a lack of personalisation of content presentation, limited research on ontology-based user profiling in e-commerce, and a limited uptake of research findings in the commercial web development world.

Building upon the analysis of the existing approaches and to address the current limitations in research and practice, a solution of how web personalisation can be achieved in a novel way involving ontological profiling was proposed. The proposed approach (*PERSONTO*) was designed to be compatible with existing e-commerce system architectures and extendible to allow for future development so that it could be incorporated easily into current (existing) systems.

Ontology-based customer profiling was proposed as a means to personalise e-commerce websites. To create a customer profiling ontology that contained an appropriate class hierarchy, a multi-case case study to elicit what aspects can be personalised and what types (groups) online customers can be profiled into was carried out. The case study revealed that there are a number of common and differentiating factors that affect e-customers, and the latter, differentiating factors, led to the determination of what can be personalised and what types of profiles online customers can be segregated into. This in turn, led to the creation of respective classes in *OntoProfi*, our proposed customer profiling ontology.

The thesis also demonstrated the feasibility of the proposed approach. A *'proof-of-concept'* prototype of *PERSONTO* confirmed that it is possible to personalise e-commerce websites through ontological customer profiling.

In order to evaluate the proposed approach, a structured evaluation strategy was adopted. The strategy entailed evaluation of the set research aim and objectives, and validation of the proposed approach's potential acceptability ('buy-in') by end users and industry (web developers). The outcomes of the evaluation demonstrated that the research aim and objectives have been successfully achieved. The analysis of the evaluation of the attitudes towards the proposed approach to e-commerce personalisation by prospective end-users and e-commerce developers suggested a high level of acceptability of *PERSONTO* by the study groups. Both groups indicated that the proposed approach offers a promising potential for enhancement of e-commerce website usage.

Overall, *PERSONTO* offers website developers a flexible, extendible, systematic and reusable way to personalise e-commerce websites and other activities typically carried out by an e-commerce company. The project needs and will benefit from further work and fuller examination by a wider group of users.

8.3 Fulfilment of Research Objectives

This thesis demonstrated that all the research objectives, which were stated in § 1.2, were successfully fulfilled. Table 8.1 summarises the achievements.

Table 8.1: Fulfilment of Research Objectives

Obj. No.	Objective	Where demonstrated?	Summary of achievements
OBJ.1	**Develop rich insights into online shopping, customer profiling and e-commerce websites personalisation:**	**Chs. 2, 4**	
1.a	Evaluate current academic and professional developments in the area of web personalisation through a survey of academic literature and commercial practices.	Ch. 2	A review of academic research and professional practices was presented in Ch. 2. Using systematic methodologies (§ 2.2.1 - academic research; § 2.3.1 - professional practices), we discussed a diversity of tools, methods and techniques used for web personalisation, and in particular for ontological user profiling (§ 2.2.2.3). Limitations of current research and practice were discussed (§ 2.5).
1.b	Acquire a holistic, comprehensive and up-to-date understanding of key factors that influence e-commerce users when shopping online, which is achieved by conducting a multi-case study of e-commerce consumers:	Ch. 5	The methodology and outcomes of the multi-case case study is discussed in Ch. 5.
1.b.i	Identify, describe and explain common factors that influence B2C e-commerce customers.	§ 5.3.1	The multi-case case study to elicit customer profiles identified and discussed several factors that affect online shoppers, namely, "motivating" and "de-motivating" (§ 5.3.1.1 and § 5.3.1.2, respectively), "high focus" (§ 5.3.1.3), "habitual shopping" (§ 5.3.1.4) and "non-spontaneity/ non-impulse buying" (§ 5.3.1.5)

(Table 8.1 continued)

Obj. No.	Objective	Where demonstrated?	Summary of achievements
1.b.ii	Determine differentiating factors that affect e-commerce website users, based on different users' needs and need recognition level, goals, experiences, shopping strategies, etc. - in order to determine what online shopping factors/ aspects can be personalised.	§ 5.3.2	A variety of differentiating factors were identified and discussed in § 5.3.2.
1.b.iii	Identify different customer profiles to be used in a customer profiling ontology.	§ 5.3.3	The differentiating factors that were identified to fulfil *Obj. 1.b.ii* (§ 5.3.2) were analysed to determine which of these can be personalised (see § 5.3.3.1), and subsequently three types of customer profiles were proposed (§ 5.3.3.2).
OBJ.2	***Design, implement and test a system that personalises an e-commerce website for its customers by means of ontology-based customer profiling:***	**Chs. 3, 5, 6**	
2.a	Propose conceptual design of such a system, including system architecture and functionality.	Ch. 3	Ch. 3 presented the system requirements (§ 3.4) and system design of *PERSONTO*, as a system to personalise e-commerce websites (§ 3.5). The process of e-commerce website personalisation was outlined and explained (§ 3.5.2); and system functionality was specified (§ 3.5.3).
2.b	Design and implement a flexible and extendible customer profiling ontology, which will enable personalisation of various types of interactions between an EC system and its users.	Ch. 5	Ch. 5 described the design and implementation of *OntoProfi*, an ontology to profile online shoppers.

(Table 8.1 continued)

Obj. No.	Objective	Where demonstrated?	Summary of achievements
2.c	Implement and test a *`proof-of-concept'* prototype to demonstrate the most important and representative functionality and usage of the proposed system.	Ch. 6	To test feasibility and demonstrate the most important and representation functionality of *PERSONTO*, a 'proof-of-concept' prototype was built (Ch. 6). The details of the prototype implementation (§ 6.2) and assumptions and limitations (§ 6.3) were discussed.
OBJ.3	***Evaluate the research outcomes against the research aim:***	**Ch. 7**	
3.a	Evaluate the achievement of feasibility, flexibility and extendibility of the system; as well as 'systematicity' and reusability of the proposed e-commerce web personalisation approach.	§ 7.3 - 7.5	Based on a particular evaluation strategy (§ 7.2), the evaluation of the project outcomes discussed the achievements of the proposed system's feasibility (§ 7.3), flexibility and extendibility (§ 7.4) and 'systematicity' and reusability of the presented approach (§ 7.5).
3.b	Evaluate the relevant stakeholders' acceptance ('buy-in') of the proposed system.	§ 7.6	The evaluation of acceptance by potential *PERSONTO*'s users was accomplished by means of a survey, which employed a particular survey instrument (§ 7.6.1.1) and data collection procedures (§ 7.6.1.2). The outcomes from the survey analysis were discussed in § 7.6.1.3. The evaluation of acceptance by the industry (e-commerce developers) was discussed in §7.6.2.

8.4 Contribution to Research and Practice

This research project has made a number of contributions to the domain, and these can be summarised as:

- **Novel, feasible and flexible approach to e-commerce website personalisation, in particular personalisation of e-commerce website's content presentation.**

  As discussed in §§ 2.2.2 and 2.5, there has been some research produced in this area, however, it also had limitations, such as the domain coverage or not using ontological customer profiling. This project has addressed these limitations.

  We have also demonstrated that the proposed approach is *feasible* (see chapter 6). The approach is also *flexible* - as it does not rely on a particular platform or programming language and the components and technologies (e.g. the reasoner) can be substituted with an alternative (see § 6.2.1 and Appendix E).

- **The use of systematic, reusable and extendible procedures, processes, tools and techniques for the system design and development.**

  The thesis specified and discussed particular methodologies, processes, tools and techniques, which were used for *PERSONTO*'s design and implementation, e.g. the system requirements and design (§§ 3.4 and 3.5); the case study protocol for elicitation of customer profile types (§ 5.2.1 and appendix A); the system acceptance survey (§ 7.6). These are reusable, and can be adapted to the needs of future projects and in different research and practice settings.

- **Reusable and extendible customer profiling ontology:**

The thesis demonstrated that the proposed and implemented *OntoProfi* customer profiling ontology can be successfully used for website personalisation (see § 6.2.4). The ontology contains a number of classes that are typical for the e-commerce domain to enable the ontology's reuse and extensions in future projects (§ 4.3.2). *OntoProfi* was also designed to be Semantic Web-compatible (see § 4.3.1), which promotes reuse and capability to extend the ontology by merging or aligning with other existing ontologies on the Web.

- **Platform for investigation of effects of personalisation:**

  In a wider research context, the developed system to personalise websites can be adopted and used in Behavioural Science research for theoretical and empirical investigations of the role and effects of personalisation, for example, to examine whether the use of personalisation increases the overall website usage, improves customers' perception of a website, affects the amount of purchases.

8.5 Future Work

Several areas present an opportunity for future development, and these are summarised as follows:

- **Customer Profiling:**

  The current version of the prototype demonstrated how website personalisation can be achieved by using stereotypical (*'group'*), static customer profiles, which were created, based on explicit data (§§ 3.5.2 and 3.5.3). Further customer profile enrichment, such as creation of individual (*'fuzzy'*), dynamic customer profiling based on collected implicit data (e.g. logs of customer interaction with a website) or the customer data that already exists in external systems (e.g.

social networks data), represents an area for future development, and can be accomplished by extending/ modifying *OntoProfi*. In addition, customer profiles can be enriched by soliciting for and storing additional online customer's characteristics (e.g. customer expertise/ web skills), or customer's preferences (e.g. favourite colours, fabric/ material, other customers' product ratings) - some of which are already provided in *OntoProfi* (see § 4.3.2).

Further work on customer profiling and personalisation will include other phases of the online buying process, such as the *"purchasing"* phase. Customer profiling can also be extended to include personalisation of online shoppers' experience that is based on identification and recognition of external factors, such as *"context-aware"* personalisation, where the system automatically identifies the type of device used by a customer (e.g. desktop, mobile phone or tablet) and personalises the website accordingly.

- **Content Personalisation:**

  The focus of the prototype was on demonstration of the most representative functionality that personalises the content presentation personalisation, i.e. the gap that was identified by our review of research and professional practice (see §§ 2.2.2.2, 2.3 and 2.5) . Extending the prototype to personalisation of other components, such as website structure and navigation, as well as content personalisation (e.g. product recommendations) is a future step in its development.

- **User-Driven Approach:**

  Investigation of feasibility of extending the *PERSONTO*'s functionality to enable the customers to specify their preferences explicitly, e.g. by changing the website's layout, is an interesting future development challenge.

- **"Business" Perspective Coverage:**

  As stated in § 1.5, the focus of this research was on the *"Technology"* perspective. Extending *PERSONTO* to the *"Business"* perspective to cover personalisation of *'non-technology'*-related aspects such as promotional, order fulfilment and customer relationship management activities presents yet another opportunity for future development. This can include personalisation of the type and frequency of marketing communications, order progress notification, payment types, delivery and returns.

  As already stated (see §§ 3.5.1 and 7.4), *PERSONTO* and *OntoProfi* were designed so that future extension of the system components can be easily incorporated. For example, *OntoProfi* can be merged with other e-commerce ontologies, such as *GoodRelations* (Hepp 2008), which was discussed in § 2.2.2.3, to cross-sell or up-sell the company's products. Another interesting area is the integration of *PERSONTO* with other Semantic Web technologies that are commonly used in e-commerce (e.g. with Web Services). Additionally, personalisation in *PERSONTO* can be extended to incorporate other business-related aspects that are specific to a particular e-commerce company, e.g. the company's business rules, and marketing and CRM strategies. *OntoProfi* already provides support for these, particularly, through the classes of the "Business" sub domain (see § 4.3.2).

- **System Quality and Quality of Service:**

  Limited testing of system quality was identified as one of the limitations of the current prototype (§ 7.7). Further development and testing to ensure system robustness, reliability, efficiency and effectiveness in the 'real world' (as opposed to the *'lab*'), commercial environment need to be carried out. In particular, the future testing should evaluate the *quickness of adaptation* (a system's ability to personalise a website quickly), *load balancing* (distribution of the load of different personalisation components on the web servers) and failover/ failback mechanisms, which were proposed by Kobsa (2001) as significant web personalisation system's quality characteristics.

In addition, future work can be extended to ensure other non-functional system quality characteristics, such as usability, accuracy, security and data quality assurance.

8.6 And Finally...

The inspiration for this project was the discovery that, despite the advances in the related research and practice, namely, on web personalisation and ontology-based user profiling, further work was needed to enhance e-commerce personalisation for its users and developers. Building upon existing web personalisation and ontological profiling technologies, tools and techniques, developed in other domains, the project offers a potential to move this forward by presenting a novel flexible, extendible and reusable approach to e-commerce personalisation.

References

A. Aamodt & E. Plaza (1994). ‘Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches’. *AI Communications* **7**(1):39–59.

F. Abbattista, et al. (2002). ‘Improving the Usability of an E-commerce Web Site through Personalization’. In F. Ricci & B. Smith (eds.), *Recommendation and Personalization in Ecommerce, 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*, pp. 20 – 29.

S. S. R. Abidi & S. R. Abidi (2013). ‘An Ontology-Driven Personalization Framework for Designing Theory-Driven Self-Management Interventions’. In *KR4HC/ProHealth*, vol. 8268 of *Lecture Notes in Computer Science*, pp. 97 – 112. Springer.

M. Adda, et al. (2006). ‘Personalization Based on Domain Ontology’. In *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED ’06), Madrid, Spain*, pp. 199 – 204.

G. Adomavicius & A. Tuzhilin (2005). ‘Personalization Technologies: a Process-Oriented Perspective’. *Communications of the ACM* **48**(10):83 – 90.

A. Aghasaryan, et al. (2010). ‘Personalized Application Enablement by Web Session Analysis and Multisource User Profiling’. *Bell Labs Technical Journal* **15**(1):67–76.

Alexa Internet (n.d). ‘About Us. [online] Available at: <http://www.alexa.com/about> [Accessed 10 July 2014]’.

D. Allemang & J. Hendler (2011). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edn.

D. Anand & B. S. Mampilli (2014). ‘Folksonomy-Based Fuzzy User Profiling for Improved Recommendations’. *Expert Systems with Applications* **41**(5):2424 – 2436.

G. Antoniou, et al. (2012). *A Semantic Web Primer*. The MIT Press.

L. Ardissono, et al. (1999). 'A Configurable System for the Construction of Adaptive Virtual Stores'. *World Wide Web* **2**(3):143 – 159.

R. Atkinson (2002). 'The Life Story Interview'. In J. F. Gubrium & J. A. Holstein (eds.), *Handbook of Interview Research: Context and Method*, chap. 6, pp. 121 – 140. SAGE Publications.

N. F. Awad & M. S. Krishnan (2006). 'The Personalization Privacy Paradox: an Emprical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization'. *MIS Quarterly* **30**(1):13 – 28.

F. Baader, et al. (2008). 'Description Logics'. In F. van Harmelen, V. Lifschitz, & B. Porter (eds.), *Handbook of Knowledge Representation*, chap. 3, pp. 135 – 180. Elsevier.

S. M. Bae, et al. (2012). 'Movie Recommendation Based on User Profile and Reviews'. In D. Anderson, H.-J. Yang, & P. Varacha (eds.), *Latest Trends in Information Technology*. WSEAS Press.

R. Baraglia & F. Silvestri (2007). 'Dynamic Personalization of Web Sites Without User Intervention'. *Communications of the ACM* **50**(2):63 – 67.

Y. Baruch (1999). 'Response Rate in Academic Studies — A Comparative Analysis'. *Human Relations* **52**(4):421 – 438.

S. Bechhofer, et al. (2004). 'OWL Web Ontology Language Overview - W3C Recommendation'. http://www.w3.org/TR/owl-features/.

D. Beneventano & D. Montanari (2008). 'Ontological Mappings of Product Catalogues'. In *The Third International Workshop on Ontology Matching (OM), 2008*.

P. Bharati & P. Tarasewich (2002). 'Global Perceptions of Journals Publishing E-Commerce Research'. *Communications of the ACM* **45**(5):21 – 26.

L. Blasco-Arcas, et al. (2014). 'Collaborating Online: the Roles of Interactivity and Personalization'. *The Service Industries Journal* **34**(8):677 – 698.

L. Blasco-Arcas, et al. (2013). 'Adopting Television as a New Channel for E-Commerce. The Influence of Interactive Technologies on Consumer Behavior'. *Electronic Commerce Research* **13**(4):457 – 475.

P. Brusilovsky (2001). 'Adaptive Hypermedia'. *User Modeling and User-Adapted Interaction* **11**(1-2):87 – 110.

A. Bunt, et al. (2007). 'Adaptive Content Presentation for the Web'. In P. Brusilovsky, A. Kobsa, & W. Nejdl (eds.), *The Adaptive Web*, chap. 13, pp. 409 – 432. Springer-Verlag, Berlin, Heidelberg.

R. Burke (2007). 'Hybrid Web Recommender Systems'. In P. Brusilovsky, A. Kobsa, & W. Nejdl (eds.), *The Adaptive Web*, chap. 12, pp. 377 – 408. Springer-Verlag, Berlin, Heidelberg.

I. Cantador, et al. (2008a). 'Ontology-Based Personalised and Context-Aware Recommendations of News Items'. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08), Sydney, Australia*, vol. 1, pp. 562 – 565.

I. Cantador, et al. (2008b). 'A Multi-Purpose Ontology-Based Approach for Personalised Content Filtering and Retrieval'. In M. Wallace, M. Angelides, & P. Mylonas (eds.), *Advances in Semantic Media Adaptation and Personalization*, pp. 25 – 52. Springer Verlag Studies in Computational Intelligence, Vol. 93.

F. Cena, et al. (2012). 'Property-Based Interest Propagation in Ontology-Based User Model.'. In *20th Conference on User Modeling, Adaptation, and Personalization (UMAP 2012), Montreal, Canada*, vol. 7379 of *Lecture Notes in Computer Science*, pp. 38 – 50. Springer.

D. Chaffey (2011). *E-Business & E-Commerce Management: Strategy, Implementation and Practice*. Financial Times/ Prentice-Hall, 5th edn.

K. Charmaz (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications Ltd, 1st edn.

P. Y. K. Chau, et al. (2000). 'Impact of Information Presentation Modes on Online Shopping: An Empirical Evaluation of a Broadband Interactive Shopping Service'. *Journal of Organizational Computing & Electronic Commerce* **10**(1):1.

Z. Chedrawy & S. Abidi (2006). 'Case Based Reasoning for Information Personalization: Using a Context-Sensitive Compositional Case Adaptation Approach'. In *Proceedings of IEEE International Conference on Engineering of Intelligent Systems (ICEIS), Islamabad, Pakistan*.

D. Chen, et al. (2010). 'A Semantic Query Approach to Personalized E-Catalogs Service System'. *Journal of Theoretical and Applied Electronic Commerce Research* **5**(3):39 – 54.

F.-F. Cheng, et al. (2009). 'The Effect of Online Store Atmosphere on Consumer's Emotional Responses - an Experimental Study of Music and Colour'. *Behaviour & Information Technology* **28**(4):323 – 334.

C. M. K. Cheung, et al. (2003). 'Online Consumer Behavior: a Review and Agenda for Future Research'. In *Proceedings of the 16th Bled eCommerce Conference, Bled, Slovenia*, pp. 194 – 218.

I. Cingil, et al. (2000). 'A Broader Approach to Personalization'. *Communications of the ACM* **43**(8):136 – 141.

J. Creswell (2006). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications, Inc, 2nd edn.

J. W. Creswell (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications Ltd., 2 edn.

O. Dameron, et al. (2004). 'Accessing and Manipulating Ontologies Using Web Services'. In *Proceedings of Workshop on Semantic Web Services: Preparing to Meet the World of Business Applications at the Third International Conference on the Semantic Web (ISWC 2004), Hiroshima, Japan*.

M. Daoud, et al. (2008). 'Learning User Interests for a Session-based Personalized Search'. In *Proceedings of the Second International Symposium on Information Interaction in Context (IIiX'08), London, UK*, pp. 57 – 64.

W. K. Darley, et al. (2010). 'Toward an Integrated Framework for Online Consumer Behavior and Decision Making Process: a Review'. *Psychology and Marketing* **27**(2):94 – 116.

J. Davies, et al. (eds.) (2002). *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley & Sons, Inc., New York, NY, USA.

F. Davis (1989). 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology'. *MIS Quarterly* **13**(3):319 – 340.

B. Detlor, et al. (2003). 'Pre-Purchase Online Information Seeking: Search versus Browse'. *Journal of Electronic Commerce Research* **4**(2):72 – 84.

K. M. Eisenhardt (1989). ‘Building Theories from Case Study Research’. *The Academy of Management Review* **14**(4):532 – 550.

A. B. Eisingerich & T. Kretschmer (2008). ‘In E-Commerce, More Is More’. *Harvard Business Review* **86**(3):20 – 21.

S. Faja & S. Trimi (2006). ‘Influence Of The Web Vendor’s Interventions On Privacy-Related Behaviors In E-Commerce’. *Communications of AIS* **2006**(17):2 – 68.

H. Fan & M. S. Poole (2006). ‘What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems’. *Journal of Organizational Computing & Electronic Commerce* **16**(3/4):179 – 202.

C. Felden & M. Linden (2007). ‘Ontology-Based User Profiling’. In *Proceedings of the 10th international conference on Business information systems (BIS ’07), Poznań, Poland*, pp. 314–327.

D. Fensel, et al. (2001). ‘Ontologies and Electronic Commerce’. *Intelligent Systems, IEEE* **16**(1):8 – 14.

D. Fensel, et al. (eds.) (2002). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, Cambridge, MA, USA.

J. Fink, et al. (2002). ‘Putting Personalization into Practice’. *Communications of the ACM* **45**(5):41 – 42.

S. Flesca, et al. (2005). ‘Mining User Preferences, Page Content and Usage to Personalize Website Navigation’. *World Wide Web* **8**(3):317 – 345.

A. W. P. Fok (2005). ‘Ontology-Driven Content Search for Personalized Education’. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM ’05), Singapore*, pp. 1033 – 1034.

M. Fowler (2003). *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Addison-Wesley, Boston, MA, USA, 3rd edn.

M. Gaeta, et al. (2009). ‘Advanced Ontology Management System for Personalised e-Learning’. *Knowledge-Based Systems* **22**(4):292 – 301.

M. Gao, et al. (2010). ‘Personalisation in Web Computing and Informatics: Theories, Techniques, Applications, and Future Research’. *Information Systems Frontiers* **12**(5):607 – 629.

S. Gauch, et al. (2007). 'User Profiles for Personalized Information Access'. In P. Brusilovsky, A. Kobsa, & W. Nejdl (eds.), *The Adaptive Web*, vol. 4321 of *Lecture Notes in Computer Science*, chap. 2, pp. 54 – 89. Springer Berlin / Heidelberg.

W. Gaulke & J. Ziegler (2015). 'Using Profiled Ontologies to Leverage Model Driven User Interface Generation'. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '15), Duisburg, Germany*, pp. 254 – 259.

M. Gocić (2008). 'Ontologizing of Product Catalogue'. In *5th Conference on Information and Communication Technologies, Fitconf 2008*, Mostar, Bosnia and Herzegovina.

M. Golemati, et al. (2007). 'Creating an Ontology for the User Profile: Method and Applications'. In *Proceedings of the First International Conference on Research Challenges in Information Science (RCIS 2007), Ouarzazate, Morocco*.

P. Gomes, et al. (2008). 'Using Ontologies for eLearning Personalization'. *Communication & Cognition* **41**(1 & 2):127 – 142.

J. Gonzalez Castillo, et al. (2001). 'Description Logics for Matchmaking of Services'. In *Proceedings of KI2001 Workshop on Applications of Description Logics*, Vienna, Austria.

A. Goy, et al. (2007). 'Personalization in e-Commerce Applications'. In P. Brusilovsky, A. Kobsa, & W. Nejdl (eds.), *The Adaptive Web*, chap. 16, pp. 485 – 520. Springer-Verlag, Berlin, Heidelberg.

B. C. Grau, et al. (2008). 'OWL 2: The Next Step for OWL'. *Web Semantics: Science, Services and Agents on the World Wide Web* **6**(4):309 – 322.

U. Gretzel & D. R. Fesenmaier (2006). 'Persuasion in Recommender Systems'. *International Journal of Electronic Commerce* **11**(2):81 – 100.

T. R. Gruber (1993). 'A Translation Approach to Portable Ontology Specifications'. *Knowledge Acquisition* **5**(2):199 – 220.

Y. M. Guo & M. S. Poole (2009). 'Antecedents of Flow in Online Shopping: a Test of Alternative Models'. *Information Systems Journal* **19**(4):369 – 390.

A. Gupta & A. Mathur (2002). 'Adaptive Delivery of E-Commerce Web Sites'. *Intelligent Data Analysis* **6**(5):469 – 480.

W. Hampton-Sosa & M. Koufaris (2005). ‘The Effect of Web Site Perceptions on Initial Trust in the Owner Company’. *International Journal of Electronic Commerce* **10**(1):55 – 81.

A. Hawalah & M. Fasli (2015). ‘Dynamic User Profiles for Web Personalisation’. *Expert Systems with Applications* **42**(5):2547 – 2569.

L. Hella & J. Krogstie (2011). ‘Personalisation by Semantic Web Technology in Food Shopping’. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics (WIMS ’11), Sogndal, Norway*, pp. 34:1 – 34:12.

M. Hepp (2005). ‘eClassOWL: a Fully-Fledged Products and Services Ontology in OWL [poster]’. In *the 4th International Semantic Web Conference (ISWC2005)*, Galway, Ireland.

M. Hepp (2008). ‘GoodRelations: An Ontology for Describing Products and Services Offers on the Web’. In *Proceedings of the 16th international conference on Knowledge Engineering: Practice and Patterns (EKAW ’08), Acitrezza, Catania, Italy*, pp. 329 – 346.

B. Hernández-Ortega, et al. (2008). ‘Differences Between Potential, New and Experienced E-Customers’. *Internet Research* **18**(3):248 – 265.

A. R. Hevner, et al. (2004). ‘Design Science in Information Systems Research’. *MIS Quarterly* **28**(1):75 – 105.

P. Hitzler, et al. (2012). ‘OWL 2 Web Ontology Language Primer (Second Edition)’. http://www.w3.org/TR/2012/REC-owl2-primer-20121211/.

P. Hitzler, et al. (2010). *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC.

S. Y. Ho (2006). ‘The Attraction of Internet Personalization to Web Users’. *Electronic Markets* **16**(1):41 – 50.

A. Hoppe, et al. (2013). ‘Automatic Ontology-based User Profile Learning from Heterogeneous Web Resources in a Big Data Context’. *Proc. VLDB Endow.* **6**(12):1428–1433.

I. Horrocks & P. F. Patel-Schneider (2011). ‘Knowledge Representation and Reasoning on the Semantic Web: OWL’. In J. Domingue, D. Fensel, & J. A. Hendler (eds.), *Handbook of Semantic Web Technologies*, chap. 9, pp. 365–398. Springer.

C. Ihlström & J. Palmer (2002). 'Revenues for Online Newspapers: Owner and User Perceptions'. *Electronic Markets* **12**(4):228 – 236.

IMRG (2014). '£91 Billion Spent Online in 2013 (Press Release). [online] [Accessed 31 May 2014]. Avalable from: <http://www.imrg.org/index.php?catalog=539>'. `http://www.imrg.org/index.php?catalog=539`.

IMRG (2015). 'The Top 50 Online Retailers in the UK November 2014. [online] [Accessed 13 January 2015]. Avalable from: <http://www.imrg.org/index.php?catalog=1616>'.

Internet Retailing (2013). 'Online Retailer Numbers in UK Have Doubled in Five Years. [online] Available at: <http://internetretailing.net/2013/05/online-retailer-numbers-in-uk-have-doubled-in-five-years/> [Accessed 29 August 2014]'.

V. K. J. Jeevan & P. Padhi (2006). 'A Selective Review of Research in Content Personalization'. *Library Review* **55**(9):556–586.

T. Jiang & A. Tuzhilin (2009). 'Improving Personalization Solutions through Optimal Segmentation of Customer Bases'. *IEEE Transactions on Knowledge and Data Engineering* **21**(3):305–320.

X. Jiang & A.-H. Tan (2009). 'Learning and Inferencing in User Ontology for Personalized Semantic Web Search'. *Information Sciences* **179**(16):2794–2808.

Z. J. Jiang & I. Benbasat (2007). 'The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding'. *MIS Quarterly* **31**(3):475 – 500.

C. B. Jones, et al. (2002). 'Spatial Information Retrieval and Geographical Ontologies: an Overview of the SPIRIT Project'. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02), Tampere, Finland*, pp. 387 – 388.

A. Katifori, et al. (2008). 'Personal Ontology Creation and Visualization for a Personal Interaction Management System'. In *CHI Workshop on Personal Information Management (PIM)*, Florence, Italy.

P. Kearney, et al. (2005). 'Employing a Domain Ontology to Gain Insights into User Behaviour'. In *Proceedings of the 3rd Workshop on Intelligent Techniques for Web Personalization, at IJCAI 2005*.

J. K. Kim, et al. (2002). 'A Personalized Recommendation Procedure for Internet Shopping Support'. *Electronic Commerce Research and Applications* **1**(3-4):301–313.

A. Kobsa (2001). 'Generic User Modeling Systems'. *User Modeling and User-Adapted Interaction* **11**(1-2):49 – 63.

A. Kobsa, et al. (2001). 'Personalised Hypermedia Presentation Techniques for Improving Online Customer Relationships'. *Knowledge Engineering Review* **16**(2):111 – 155.

A. Koerber & L. McMichael (2008). 'Qualitative Sampling Methods: A Primer for Technical Communicators'. *Journal of Business and Technical Communication* **22**(4):454 – 473.

U. Konradt, et al. (2003). 'Usability in Online Shops: Scale Construction, Validation and the Influence on the Buyers' Intention and Decision'. *Behaviour & Information Technology* **22**(3):165 – 174.

M. Koufaris (2002). 'Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior'. *Information Systems Research* **13**(2):205 – 223.

P. Koutsabasis & J. Darzentas (2008). 'Item-Based Filtering and Semantic Networks for Personalized Web Content Adaptation in E-Commerce'. In J. Darzentas, G. A. Vouros, S. Vosinakis, & A. Arnellos (eds.), *Artificial Intelligence: Theories, Models and Applications: 5th Hellenic Conference on AI, SETN 2008, Syros, Greece, October 2-4, 2008*, pp. 148–159.

P. Koutsabasis, et al. (2008). 'A Descriptive Reference Framework for the Personalisation of E-Business Applications'. *Electronic Commerce Research* **8**(3):173 – 192.

D. A. Koutsomitropoulos, et al. (2006). 'A Methodology for Conducting Knowledge Discovery on the Semantic Web'. In *Adaptive and Personalized Semantic Web*, pp. 95–105.

J. Kramer, et al. (2000). 'A User-Centered Design Approach to Personalization'. *Communications of the ACM* **43**(8):44 – 48.

M. Krötzsch (2012). 'OWL 2 Profiles: An Introduction to Lightweight Ontology Languages'. In *Reasoning Web. Semantic Technologies for Advanced Query Answering: 8th International Summer School 2012, Vienna, Austria, September 3-8, 2012*, pp. 112 – 183.

Y.-F. Kuo & L.-S. Chen (2001). 'Personalization Technology Application to Internet Content Provider'. *Expert Systems with Applications* **21**(4):203 – 215.

K. S. Kuppusamy & G. Aghila (2012). 'A Personalized Web Page Content Filtering Model Based on Segmentation'. *International Journal of Information Sciences and Techniques* **2**(1):41 – 51.

K. Kwon & C. Kim (2012). 'How to Design Personalization in a Context of Customer Retention: Who Personalizes What and to What Extent?'. *Electronic Commerce Research and Applications* **11**(2):101 – 116.

K. Laudon & C. Traver (2006). *E-Commerce: Business, Technology, Society (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

J. Lee (2000). 'Understanding Merchandising Effectiveness of Online Stores'. *Electronic Markets* **10**(1):20 – 28.

T. Lee, et al. (2006). 'Building an Operational Product Ontology System'. *Electronic Commerce Research and Applications* **5**(1):16–28.

Y. Lei, et al. (2003). 'Design of Customized Web Applications with OntoWeaver'. In *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP '03), Sanibel Island, Florida, USA*, pp. 54 – 61.

T.-P. Liang, et al. (2006). 'Personalized Content Recommendation and User Satisfaction: Theoretical Synthesis and Empirical Findings'. *Journal of Management Information Systems* **23**(3):45 – 70.

Y. Lincoln & E. Guba (1985). *Naturalistic Inquiry*. SAGE Publications.

W. Liu, et al. (2008). 'Ontology-Based User Modeling for E-Commerce System'. In *Pervasive Computing and Applications, 2008. ICPCA 2008. Third International Conference on*, vol. 1, pp. 260 – 263.

L. A. MacAulay, et al. (2007). 'Co-Evolving E-tail and On-Line Communities: Conceptual Framework'. *International Journal of Electronic Commerce* **11**(4):53 – 77.

Z. Malik & C. Fyfe (2012). 'Review of Web Personalization'. *Journal of Emerging Technologies in Web Intelligence* **4**(3):285 – 296.

C. D. Manning, et al. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

M. N. Marshall (1996). 'Sampling for Qualitative Research'. *Family Practice* **13**(6):522–526.

F. J. Martínez-López, et al. (2010). 'Psychological Elements Explaining the Consumer's Adoption and Use of a Website Recommendation System: a Theoretical Framework Proposal'. *Internet Research* **20**(3):316–341.

A. P. Massey, et al. (2007). 'Usability of Online Services: The Role of Technology Readiness and Context'. *Decision Sciences* **38**(2):277 – 308.

S. E. Middleton, et al. (2004). 'Ontological User Profiling in Recommender Systems'. *ACM Transactions on Information Systems* **22**(1):54 – 88.

M. Miles & A. Huberman (1994). *Qualitative Data Analysis: an Expanded Sourcebook*. SAGE Publications.

R. Mizoguchi (2003). 'Tutorial on Ontological Engineering. Part 1: Introduction to Ontological Engineering'. *New Generation Computing* **21**(4):365 – 384.

B. Mobasher (2007). 'Data Mining for Web Personalization'. In P. Brusilovsky, A. Kobsa, & W. Nejdl (eds.), *The Adaptive Web*, vol. 4321 of *Lecture Notes in Computer Science*, pp. 90 – 135. Springer Berlin Heidelberg.

B. Mobasher, et al. (2000). 'Automatic Personalization Based on Web Usage Mining'. *Communications of the ACM* **43**(8):142 – 151.

M. Montaner, et al. (2003). 'A Taxonomy of Recommender Agents on the Internet'. *Artificial Intelligence Review* **19**(4):285 – 330.

A. L. Montgomery & M. D. Smith (2009). 'Prospects for Personalization on the Internet'. *Journal of Interactive Marketing* **23**(2):130 – 137.

R. J. Mooney & L. Roy (2000). 'Content-Based Book Recommending Using Learning for Text Categorization'. In *Proceedings of the Fifth ACM Conference on Digital Libraries (DL '00), San Antonio, TX, USA*, pp. 195 – 204.

B. Motik, et al. (2012). 'OWL 2 Web Ontology Language Profiles (Second Edition)'. `http://www.w3.org/TR/2012/REC-owl2-profiles-20121211/`.

M. D. Mulvenna, et al. (2000). 'Personalization on the Net using Web Mining'. *Communications of the ACM* **43**(8):122 – 125.

B. P. S. Murthi & S. Sarkar (2003). 'The Role of the Management Sciences in Research on Personalization'. *Management Science* **49**(10):1344 – 1362.

S. Mussi (2006). 'User Profiling on the Web Based on Deep Knowledge and Sequential Questioning'. *Expert Systems* **23**(1):21 – 38.

Nielsen (2014). 'E-Commerce: Evolution or Revolution in the Fast-Moving Consumer Goods World? report, August 2014'.

L. Niu, et al. (2002). 'Product Hierarchy-Based Customer Profiles for Electronic Commerce Recommendation'. In *Proceedings of 2002 International Conference on Machine Learning and Cybernetics, 2002*, vol. 2, pp. 1075 – 1080.

P. Nowakowski & H. Stuckenschmidt (2010). 'Ontology-Based Product Catalogues: an Example Implementation'. In *MKWI*, pp. 15 – 25.

N. F. Noy & D. L. McGuinness (2001). 'Ontology Development 101: A Guide to Creating Your First Ontology'. Tech. rep.

F. Ntawanga, et al. (2008). 'Maintaining Customer Profiles in an e-Commerce Environment'. In *Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology (SAICSIT '08)*, pp. 169 – 179.

J. Nunnally & I. Bernstein (1994). *Psychometric Theory*. McGraw Hill, 3rd edn.

B. J. Oates (2006). *Researching information systems and computing*. SAGE Publications, Thousand Oaks.

L. Obrst, et al. (2003). 'Ontologies for Corporate Web Applications'. *AI Magazine* **24**(3):49 – 62.

R. M. O'Keefe & T. McEachern (1998). 'Web-Cased Customer Decision Support Systems'. *Communications of the ACM* **41**(3):71 – 78.

A. Onwuegbuzie & N. Leech (2007). 'A Call for Qualitative Power Analyses'. *Quality & Quantity: International Journal of Methodology* **41**(1):105 – 121.

H.-Y. Paik, et al. (2002). 'Dynamic Restructuring of E-Catalog Communities Based on User Interaction Patterns'. *World Wide Web* **5**(4):325 – 366.

Y.-J. Park & K.-N. Chang (2009). 'Individual and Group Behavior-based Customer Profile Model for Personalized Product Recommendation'. *Expert Systems with Applications* **36**(2):1932 – 1939.

M. Perkowitz & O. Etzioni (2000). 'Adaptive Web Sites'. *Communications of the ACM* **43**(8):152 – 158.

D. Pierrakos, et al. (2003). 'Web Usage Mining As a Tool for Personalization: A Survey'. *User Modeling and User-Adapted Interaction* **13**(4):311 – 372.

K. C. Pramataris, et al. (2001). 'Personalized Interactive TV Advertising: The iMEDIA Business Model'. *Electronic Markets* **11**(1):17 – 25.

B. Prasad (2003). 'Intelligent Techniques for E-Commerce'. *Journal of Electronic Commerce Research* **4**(2):65 – 71.

T. S. Raghu, et al. (2001). 'Dynamic Profiling of Consumers for Customized Offerings over the Internet: a Model and Analysis'. *Decision Support Systems* **32**(2):117 – 134.

Rianto, et al. (2015). 'A Preliminary Study to Determine Criteria for Personalized e-Commerce'. In *Proceedings of the International HCI and UX Conference in Indonesia (CHIuXiD '15)*, pp. 73 – 77.

D. Riecken (2000). 'Personalized Views of Personalization'. *Communications of the ACM* **8**:26 – 28.

T. Robal & A. Kalja (2007). 'Applying User Profile Ontology for Mining Web Site Adaptation Recommendations'. In Y. E. Ioannidis, B. Novikov, & B. Rachev (eds.), *ADBIS Research Communications*, vol. 325 of *CEUR Workshop Proceedings*. CEUR-WS.org.

S. Rojbi & M. Soui (2011). 'User Modeling and Web-Based Customazation Techniques: an Examination of the Published Literature'. In *4th International Conference on Logistics on Logistics (LOGISTIQUA) 2011*, pp. 83 – 90.

D. Rosaci & G. Sarnè (2014). 'Multi-Agent Technology and Ontologies to Support Personalization in B2C E-Commerce'. *Electronic Commerce Research and Applications* **13**(1):13 – 23.

G. M. Rose, et al. (2005). 'On-line waiting: The role of download time and other important predictors on attitude toward e-retailers'. *Psychology and Marketing* **22**(2):127 – 151.

R. T. Rust & C. Miu (2006). 'What Academic Research Tells Us About Service'. *Communications of the ACM* **49**(7):49 – 54.

S. Sarker, et al. (2013). 'Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles'. *MIS Quarterly* **37**(4):iii – xviii.

E. Schonberg, et al. (2000). 'Measuring Success'. *Communications of the ACM* **43**(8):53 – 57.

P. Schubert (2000). 'The Participatory Electronic Product Catalog: Supporting Customer Collaboration in E-Commerce Applications'. *Electronic Markets* **10**(4):229 – 236.

P. Schubert & M. Ginsburg (2000). 'Virtual Communities of Transaction: The Role of Personalization in Electronic Commerce.'. *Electronic Markets* **10**(1):45 – 55.

P. Schubert & M. Koch (2002). 'The Power of Personalization: Customer Collaboration and Virtual Communities'. In *Proceedings of Americas Conference on Information Systems (AMCIS) 2002*, pp. 1953 – 1965, Dallas, TX.

B. Schwartz, et al. (2002). 'Maximizing versus Satisficing: Happiness is a Matter of Choice'. *Journal of Personality and Social Psychology* **83**(5):1178 – 1197.

A. Shen & A. D. Ball (2009). 'Is Personalization of Services Always a Good Thing? Exploring the Role of Technology-Mediated Personalization (TMP) in Service Relationships'. *Journal of Services Marketing* **23**(2):79 – 91.

G. S. Shergill & Z. Chen (2005). 'Web-Based Shopping: Consumers' Attitudes Towards Online Shopping In New Zealand'. *Journal of Electronic Commerce Research* **6**(2):79 – 84.

A. Sieg, et al. (2007). 'Web Search Personalization with Ontological User Profiles'. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, pp. 525 – 534, New York, NY, USA. ACM.

H. A. Simon (1956). 'Rational Choice and the Structure of the Environment'. *Psychological Review* **63**(2):129 – 138.

K.-L. Skillen, et al. (2014). 'Ontological User Modelling and Semantic Rule-based Reasoning for Personalisation of Help-On-Demand Services in Pervasive Environments'. *Future Generation Computer Systems* **34**:97 – 109.

B. Smyth & P. Cotter (2000). 'A Personalized Television Listings Service'. *Communications of the ACM* **43**(8):107 – 111.

K. Srikumar & B. Bhasker (2004). 'Personalized Product Selection in Internet Business'. *Journal of Electronic Commerce Research* **5**(4):216 – 227.

S. Stamou, et al. (2009). 'Ontology-Driven Personalized Query Refinement'. *Journal of Web Engineering* **8**(2):113 – 153.

N. Stebletsova-Edwards (2003). 'MultiSeek: Voice-Enabled Information Retrieval on the Internet'. MSc Dissertation, Department of Computer Science, Cardiff University.

K. Y. Tam & S. Y. Ho (2005). 'Web Personalization as a Persuasion Strategy: an Elaboration Likelihood Model Perspective'. *Information Systems Research* **16**(3):271 – 291.

K. Y. Tam & S. Y. Ho (2006). 'Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes'. *MIS Quarterly* **30**(4):865 – 890.

X. Tao, et al. (2011). 'A Personalized Ontology Model for Web Information Gathering'. *IEEE Transactions on Knowledge and Data Engineering* **23**(4):496 – 511.

S. Thirumalai & K. K. Sinha (2009). 'Customization Strategies in Electronic Retailing: Implications of Customer Purchase Behavior'. *Decision Sciences* **40**(1):5 – 36.

J. Trajkova & S. Gauch (2004). 'Improving Ontology-Based User Profiles'. In C. Fluhr, G. Grefenstette, & W. B. Croft (eds.), *RIAO*, pp. 380 – 390. CID.

D. Trastour, et al. (2003). 'Semantic Web Support for the Business-to-business e-Commerce Pre-contractual Lifecycle'. *Comput. Netw.* **42**(5):661 – 673.

E. Turban, et al. (2006). *Electronic Commerce 2006: a Managerial Perspective*. Pearson Prentice Hall.

UK Office for National Statistics (ONS) (2010). 'Internet Access 2010 - Households and Individuals'.

UK Office for National Statistics (ONS) (2015). 'Internet Access 2015 - Households and Individuals'.

X. Valencia, et al. (2013). 'User Individuality Management in Websites Based on WAI-ARIA Annotations and Ontologies'. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility (W4A '13), Rio de Janeiro, Brazil*, pp. 29:1 – 29:10.

V. Venkatesh, et al. (2013). 'Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems'. *MIS Quarterly* **37**(1):21 – 54.

V. Venkatesh & F. D. Davis (2000). ‘A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies’. *Management Science* **46**(2):186 – 204.

W3C (n.d.). ‘Semantic Web’. http://www.w3.org/standards/semanticweb/.

W3C OWL Working Group (2009). ‘OWL 2 Web Ontology Language - Document Overview (Working Draft 27 March 2009)’. http://www.w3.org/TR/owl2-overview/.

W3C OWL Working Group (2012). ‘OWL 2 Web Ontology Language - Document Overview (Second Edition)’. http://www.w3.org/TR/owl2-overview/.

K. Wei, et al. (2007). ‘A Survey of E-Commerce Recommender Systems’. In *Service Systems and Service Management, 2007 International Conference on*, pp. 1 – 5.

M. Weinmann, et al. (2013). ‘A Taxonomy of Web Personalization’. In *Proceedings of 19th Americas Conference on Information Systems (AMCIS) 2013, Chicago, Illinois, USA, August 15-17, 2013*.

K.-L. Wu, et al. (2001). ‘Personalization with Dynamic Profiler’. In *Proceedings of the Third International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems (WECWIS ’01), San Juan, California, USA*, p. 12.

B. P.-C. Yen & R. C. W. Kong (2002). ‘Personalization of Information Access for Electronic Catalogs on the Web’. *Electronic Commerce Research and Applications* **1**(1):20 – 40.

G. Yi Maggie & B. D. Klein (2009). ‘Beyond the Test of the Four Channel Model of Flow in the Context of Online Shopping’. *Communications of AIS* **2009**(24):837 – 856.

R. Yin (2003). *Case Study Research: Design and Methods*. Applied Social Research Methods Series. SAGE Publications.

S.-T. Yuan (2003). ‘A Personalized and Integrative Comparison-Shopping Engine and Its Applications’. *Decision Support Systems* **34**(2):139 – 156.

Zetoc (n.d). ‘About Zetoc. [online] Available at: <http://zetoc.mimas.ac.uk/about.html> [Accessed 1 August 2014]’.

E. Zhong, et al. (2015). ‘Building Discriminative User Profiles for Large-scale Content Recommendation’. In *Proceedings of the 21th ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining, KDD '15, pp. 2277 – 2286, New York, NY, USA. ACM.

M. Zviran, et al. (2006). 'User Satisfaction from Commercial Web Sites: the Effect of Design and Use'. *Information & Management* **43**(2):157 – 178.

Personal Communications [1]

[Pers.Comm.1] John Bay, Technical Director, small software development company, Cardiff

[Pers.Comm.2] Chris Green, Creative Director, a large multi-international group of companies, Carmarthenshire

[Pers.Comm.3] Joe Brown, Digital Manager, a small design company, West Sussex

[Pers.Comm.4] William Mellons, Company Director, independent department store, Wales and South West England

(1) To protect anonymity, the names of people have been changed, and the companies' names have been made anonymous.

Appendix A

Case Study Protocol

Sources of Evidence / Data:

- Questionnaire, observation and semi-structured interview

Before the Interview:

1. Explain the purpose of the study
2. Explain the methods and conduct (recording, confidentiality)
3. Present form to sign
4. Any questions before we start?
5. Ask to fill in the questionnaire (***"Screening questions"***)

Initial Questions:

1. Do you have a favourite website(s)?
2. Tell me about your experience with this (these) website(s)?
   2.1. Why do you like it (them)?
   2.2. How did you discover it (them)?
   2.3. How often do you "visit (them)"?
   2.4. Do you ever visit them just to "look at it (them)" without buying anything?
3. Have you ever come across a website that you do not like, or even hate?
   3.1. Are there any websites that stand/ stood out? Why? Have you bought anything from that/ those website(s)? Tell me about your experience with this/these website(s)?

Intermediate Questions:

4. How do you feel about buying on the Internet?
5. What makes you buy on the internet?
6. How does your experience compare with "normal" shopping?
7. Are there any specific things or times that you prefer to buy online compared with "non-online"? And the other way round?
8. What influences you while shopping online?
9. How do you make decisions when buying online?
   9.1. Do you think it is different from "non-online" shopping? (If so, why? If not, why not?)
10. What was the last item that you bought? When? Describe the process you went through...(experience, process, decisions...)
11. Do you [think you] have a typical process that you go through while shopping online Please describe this process... Where do you go first? Next? ....

Questions on Personalisation:

12. Have you ever noticed personalisation features/ functionalities on some e-commerce websites?
13. Have you ever take advantage of these/ use them?
14. What do you think about those? Do you think they influence your decision?
15. What feature do you wish e-commerce sites to have?

[Observation]

- Ask participants to complete a task on searching and buying products. Observe the buying process and ask the participants to explain their decision making process ('hows" and "whys").

Ending Questions:

16. Are there any lessons that you have learned from shopping online and would like to tell me about?
17. Where do you see yourself with shopping online in few years time - would you be shopping more? less? about the same? Are there any new products or services that you think you will be buying online?
18. Is there is anything else that you would like to add? Something that just occurred to you? or something that you think we haven't covered but is important to cover?
19. Is there anything else that I should understand or take into consideration?
20. Is there anything you would like to ask me?

Appendix B

“Screening” and “Products” Questionnaires

Buying Online

Screening Questionnaire

1. **What is your occupation?**...

2. **[Age]**

3. **Where do you typically access the Internet from?**

   home ❑ work ❑ other (*Please specify)*................................

4. **Have you ever bought anything on the Internet?** *(please circle appropriate answer)*

   YES NO

5. **Roughly how many hours do you spend using Internet?:**

   5.1.For general web surfing:

   per day .... or per week .... or per month ....

   5.2.For shopping on the Internet:

   per day .... or per week .... or per month ....

6. **Do you ever just search for a product or service on the Internet without actually buying ("window shop")?:**

   YES ☛ Go to Q. 6.1 **NO** ☛ Go to Q. 7

   6.1. If you answered "YES" to the above question, please specify how often are you likely to search for a product or service online:

   - every day ❑
   - 1 - 2 times a week ❑
   - a couple of times a month ❑
   - less or more often -> *please specify*.....................

7. **How often are you likely to buy something online?:**

   - every day ❑
   - 1 - 2 times a week ❑
   - a couple of times a month ❑
   - less or more often -> *please specify*.....................

8. **How many products or services have you bought on the Internet?:**

   - today ....
   - last week ....
   - last month ....
   - last year ....

9. **Do you shop online at a particular time or on a particular day?:**

Appendix C

Initial Product List for Participant Observation

Products List for Observations

1. Clothes, sports goods:
   - pair of shoes/ trainers
   - T-shirt
2. Films, music
   - films: *The King's Speech* (was best seller on AMZ on 07/06/2011)
   - *Tangled* (bestseller on play.com)
3. Holiday accommodation (e.g. hotels)
   - 3* hotel in London in Kensington
4. Household goods (e.g. furniture, toys)
   - a toy - birthday present for a 5 year old
5. Books, magazines, newspapers (including e-books)
   - The Dukan Diet
   - A game of Thrones
6. Other travel arrangements (e.g. transport tickets, car hire)
   - rail tickets: Cardiff to Manchester
7. Tickets for events
   - concert tickets (e.g. Adele, Alun Carr, Bryan Adams, Diversity)
8. Electronic equipment (including cameras)
   - a Cannon still photo camera
9. Food or groceries
   - a bottle of milk
10. Video games software and upgrades
    - Zumba Fitness (Wii)
    - L.A. Noire
11. Share purchases, insurance policies and other financial services
    - car insurance
    - Home content insurance
12. Other computer software and upgrades
    - Microsoft Office
    - Anti-Virus
13. Telecommunication services
    - broadband
    - TV subscription (e.g. Sky Sports)
14. Computer hardware
    - laptop
    - desktop
15. E-learning material
    - Learn French / Spanish / Italian
16. Medicine
    - a packet of paracetamol

Appendix D

Origins of *OntoProfi* Ontology Classes

Domain	Class	Origin		
		Commercial Practices (Industry)	**Literature**	**This research** (case study)
Customer	Customer	■	■	
	Customer Context	■	■	
	Customer Expertise		■	
	Customer Preferences	■	■	■
	Customer Profile	■	■	■
Product	Brand	■	■	■
	Product	■	■	■
	Product Attribute	■	■	■
	Price	■	■	■
Business	Communication Frequency	■		■
	Communication Type	■		■
	Costs	■		■
	Delivery	■		■
	Order	■		
	Payment Type	■		■
	Promotion	■		■
	Returns	■		■
System	Information	■	■	■
	UI Display	■	■	■

Appendix E

Alternative Languages, Technologies and Tools for Development of a Web Personalisation System

System component	Languages, technologies, tools used in this project	Main alternatives *(in alphabetic order)*
Website	Java, JSP	ASP, ASP.NET, Perl, PHP, Python, Ruby on Rails, XML
Database	MySQL	DB2, Microsoft SQL Server, Oracle, SQLite, PostgreSQL
Ontology	OWL API	Jena API, Protege-OWL API
Reasoner	HermiT	Fact++, JFact, Pellet, RacerPro
Server	GlassFish	Apache HTTP Server, Apache Tomcat, IBM HTTP Server, Microsoft's IIS
Development IDE	NetBeans	Eclipse

Appendix F

User Evaluation Survey Questionnaire

PersOnto: an E-Commerce Website Personalisation System
Questionnaire

Introduction:

PersOnto is a system that enables personalisation (adapting) of how webpages are displayed. The webpage personalisation is targeted at each individual user, and is based on users' self-selected preferences. Once a user completes a simple survey, each webpage is transformed, making particular webpage elements more prominent (e.g. if a user prefers textual information, text is displayed more prominently).

Instructions:

- Go to: https://project.cs.cf.ac.uk/N.Edwards/personto/start.html
- Click on "complete this short survey" link (https://project.cs.cf.ac.uk/N.Edwards/personto/survey.html)
- Complete the survey and evaluate the outcomes of webpages personalisation (which is based on your answers)
- Complete this questionnaire **-> *please select only one rating for each question. Also, please make sure to answer all questions to ensure that the questionnaire is usable for statistical analysis*!**

Please indicate your agreement with the next set of statements using the following rating scale:

1 Strongly disagree	2 Disagree	3 Slightly disagree	4 Neutral	5 Slightly agree	6 Agree	7 Strongly agree

Perceived Usefulness:

		Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
1	Using PersOnto and personalised webpages would enable me to do my online shopping more quickly	1	2	3	4	5	6	7
2	Using PersOnto and personalised webpages would enable me to my online shopping easier	1	2	3	4	5	6	7
3	Using PersOnto and personalised webpages would enhance my effectiveness in online shopping	1	2	3	4	5	6	7
4	Overall, using PersOnto and personalised webpages would be useful for my online shopping	1	2	3	4	5	6	7
5	Having webpages personalised is a **good idea**	1	2	3	4	5	6	7

Perceived Ease of Use:

		Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
6	Learning how to use PersOnto and personalised webpages would be easy for me	1	2	3	4	5	6	7
7	I would find it easy to get PersOnto to do what I want to do	1	2	3	4	5	6	7
8	My interactions with PersOnto and personalised webpages would be clear and understandable	1	2	3	4	5	6	7
9	I would find PersOnto and personalised webpages to be flexible to interact with	1	2	3	4	5	6	7
10	It would be easy for me to become skilful at using PersOnto and personalised webpages	1	2	3	4	5	6	7
11	Overall, I would find PersOnto to easy to use	1	2	3	4	5	6	7

Intention to Use:

		Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
12	Assuming I have access to PersOnto (or a similar type of website, which allows for webpages personalisation), I intent to use it	1	2	3	4	5	6	7
13	Given that I have access to PersOnto (or a similar type of website, which allows for webpages personalisation), I predict that I would use it	1	2	3	4	5	6	7

Other Comments: (use the other side of paper if necessary)

Thank you very much for completing this questionnaire!

Appendix G

Research Quality and Rigour

Table G.1 summarises the evidence of applying the interpetvist research principles of quality and rigour in this project.

Table G.1: Ensuring Quality in Interpretivist Research

Positivism *("Research Validity")*	**Interpretivism** *("Research Trustworthiness")*				**Evidence**
	in **(Lincoln & Guba 1985)**	in **(Miles & Huberman 1994)**	**Main Principles**	**Suggested Activities**	**in this Research**
Objectivity	**Confirmability**	**Objectivity/ Confirmability**	Relative neutrality; explicit acknowledgement of researcher's bias; demonstration that the findings are shaped by the respondents, and not by researcher's bias, motivation or interest.	Explicit and detailed description of research methods and procedures; "audit trail", comparison of findings with similar/ rival research studies; making data available to others.	Case study methodology (including protocol) was specified prior to commencement of data collection (§ 5.2). Tape recordings and transcripts have been kept.
Reliability	**Dependability**	**Reliability/ Dependability/ Auditability**	Findings are shown to be consistent, reasonably stable over time, and across researchers and methods; and could be repeated in other studies.	Making research questions or aims of the project clear to others; "meaningful parallelism across data sources" (Miles & Huberman 1994); ensuring completeness of data collection; researchers triangulation; data quality checking.	Research aim and objectives of the project were specified in § 1.2. Case study questions were specified in § 5.2.1. After the case study participants interviews and observations were recorded and transcribed, each transcript was checked for correctness of transcription at least twice after the transcription and (in the majority of case this was done 3 or 4 times) - see § 5.2.3.
Internal validity	**Credibility**	**Internal Validity / Credibility/ Authenticity**	Findings demonstrate 'truth' value; confidence in the 'truth' of the findings; findings of the study make sense, are credible to the people studied, researchers and wider communities/ other people.	Prolonged engagement; persistent observation; triangulation, and in particular data triangulation; peer debriefing; negative case analysis; referential adequacy; member checks (testing analytical/ conceptual outcomes with respondents).	The case study involved in-depth interviews and observations. The data analysis uses "staying close to data" and "comparing data with data" techniques (i.e. ensuring that only participants words and expressions are used; and that data is triangulated from all interviews and observations) - the techniques, which are widely used in interpretivist research, for instance, in Ethnography and Grounded Theory (e.g. (Charmaz 2006, pp. 42-71)
External validity	**Transferability**	**External Validity/ Transferability/ Fittingness**	Findings have applicability in other contexts; are transferable to other contexts; fit with what we already know; research demonstrates ability and possibility for the findings to be generalised.	Full, detailed description of the original data sample, respondents, research setting to permit comparison with other samples/ studies; "thick" description of findings; findings described as generic enough; suggestions on how findings can be replicated.	Description of participants and research setting was summarised in § 5.2.2. The analysis and discussion of findings was handled with utmost care to provide as rich ("thick") description as possible, using participants' *'voices'* (through direct quotation) throughout the discussion (§ 5.3). As discussed in § 7.7 "Research Limitations", although it is not possible to replicate the exact outcomes, the proposed methodology and approach can be reused (replicated) in a different study. Furthermore, the proposed customer profiles appear to be general enough to be applicable to most (and possibly all) e-commerce users.