What can I do there? Towards the automatic discovery of place-related services and activities

Ahmed N. Alazzawi*, Alia I. Abdelmoty and Christopher B. Jones

Cardiff School of Computer Science & Informatics, Cardiff University, Cardiff, Wales, UK

(*Received 4 December 2010; final version received 5 June 2011*)

The current web is rich in geographically referenced data. Mining, retrieving and sharing these data raises the need for rich geographical place name resources that record spatial and thematic elements of geographical places. Here, possible services offered at a place and human activities that can be practised there are considered useful concepts to discover and encode in place name resources. Recognising this dimension of place description can enhance information retrieval tasks by extending the range of possible queries and search criteria that relate to different place instances. This work proposes an automatic approach for the identification and extraction of service and activity-related concepts from multiple resources of textual descriptions of geographical place types. Frequent affordance patterns are identified and then applied to a corpus of resources to extract service and activity types associated with specific geographical place types. The evaluation experiments undertaken demonstrate the potential value of the approach.

Keywords: geographical place types; place semantics; human activities and services

1. Introduction

Knowledge resources that maintain information about named places have become increasingly important to support systems for geographical information retrieval (GIR). They are used to recognise place names in user queries and text documents, to disambiguate between places with the same name and to generate coordinates for, that is, geocode, place names. They can also be used for reverse geocoding to find place names associated with coordinates and to assist in tasks of geodata integration, particularly when matching data items that refer to the same place but with different terminology (Hill *et al.* 1999, Jones *et al.* 2001). Typical place name resources are referred to as gazetteers, geographical thesauri or place name ontologies and record at least the name and map coordinates of a place, usually in combination with the place type and its parent in a geographical or an administrative hierarchy. Some of these resources are richer with regard to their conceptual, terminological and geospatial modelling capacity and hence of potentially greater value in assisting with processes of GIR.

In this article, we focus on supplementing the normal content of place name resources with information about what happens at a place, in particular the sorts of services and opportunities for activity that are associated with or afforded by a place. Thus, for example, a school affords learning and teaching, while a bank affords services for storing, saving, accessing or borrowing money. The potential importance of maintaining knowledge of affordance has long been recognised (Jordan *et al.* 1998); but with the increasing scope and opportunity for geographically related information services, there is a strong motivation to encode place affordance knowledge as an integral component of place name knowledge resources. Such knowledge can be expected to be of particular value in enhancing GIR services to help users find places that service their personal interests or requirements.

This work considers methods for mining knowledge of activities and services, commonly associated with place categories or types, from a variety of geographically focused resources. Starting with an existing typology of place types, a single relatively rich source of place service descriptions, the learning corpus, is used to identify frequent affordance language patterns. These patterns have been applied to several other resources, the construction corpus, to extract place-type specific services and activities. The results have been integrated within a data model that associates place types with their typical affordances of services and activities. An evaluation of the resulting associations has been performed using a web-based survey of typical activities that people regard as taking place at, or being provided by, particular place types. The approach has proven to produce good results for the extracted service types.

The remainder of the article is organised as follows: A review of place name resources and related work on the extraction of place semantics is described in Section 2. A model of service and activity concepts used in this work is given in Section 3, followed by a description of the data-mining methodology used for the extraction of these concepts. The resulting set of concepts is presented in Section 4, where the output of the evaluation experiments undertaken to gauge the quality of the derived concepts is also discussed. The article concludes in Section 5 with an outlook on future research work.

2. Related work

There are many facets to the current use and management of geographical place information on the web. Geographic referencing and linking of resources is becoming a commonplace activity, facilitated by simple mapping applications such as *Google Maps*. A lot of information in web documents refers to place names and geographical locations. The need to disambiguate geographical names in web search queries prompted interest in employing gazetteers, such as the Getty Thesaurus of Geographic Names (TGN).[1] In addition to the association of a place name with a representative point location, TGN also links a place to a set of related place types and provides an administrative/political hierarchy of place names.

Recently, collaborative mapping web applications have emerged where users are contributing to the development of web gazetteers as well as providing detailed descriptions of places and related information. A prominent example of a web gazetteer is *GeoNames*, currently containing around 8 million geographical names. A common feature of place models underlying these web gazetteers is the explicit modelling of spatial relationships between place instances. For example, GeoNames encodes spatial containment hierarchies, relating a place instance to its parent places, as well as proximity relationships in the form of neighbouring places.

On the semantic web, place name (or toponym) ontologies are employed to facilitate the utilisation of such gazetteers to support GIR tasks, such as disambiguation and expansion of terms in search engine queries (Gregory *et al.* 2002, Jones *et al.* 2004, Smart *et al.* 2010). An ontology of place names is defined as a model of terminology and structure of geographical space and named place entities (Egenhofer 2002). It extends the traditional notion of a gazetteer to encode semantically rich spatial and non-spatial entities, such as the historical and vernacular place names and events associated with a geographical place (Perry *et al.* 2006). In addition to place qualification using place-type categorisation, qualitative spatial relationships, commonly used in search queries, are also modelled to relate place instances. An example of a place name ontology is given in Abdelmoty *et al.* (2009). In this example, a geographical place is distinguished by belonging to a category of place types, for example, roads and streets, rivers and streams, as well as by the identification of its name, its shape and its relative or absolute location representation.

Functional differentiation of geographical places, in terms of the possible human activities that may be performed in a place or place affordance, has been identified as a fundamental dimension for the characterisation of geographical places (Relph 1976). For Relph, the unique quality of a geographical place is its ability to order and to focus human intentions, experiences and actions spatially. It has been argued that place affordance is a core constituent of a geographical place definition, and thus ontologies for the geographical domain should be designed with a focus on the human activities that take place in the geographical space (Kuhn 2001, Frank 2003). The term 'action-driven ontologies' was first coined by Câmara *et al.* (2000) in categorising objects in geospatial ontologies. Affordance of geospatial entities refers to those properties of an entity that determine certain human activities. In the context of spatial information theory, several works have attempted to study and formalise the notion of affordance (Rugg *et al.* 1997, Raubal and Kuhn 2004, Kuhn 2007, Sen 2007, 2008, Scheider and Kuhn 2010). The assumption is that affordance-oriented place ontologies are needed to support the increasingly more complex applications requiring semantically richer conceptualisation of the environment. Realising the value of the notion of affordance for building richer models of geographical information, the Ordnance Survey (the national mapping agency for the United Kingdom) proposed its utilisation as one of the ontological relations for representing their geographic information (Hart *et al.* 2004) and made an explicit use of a 'has-purpose' relationship in building their ontology of buildings and places. [2]

2.1. Extracting place-related information

There exist several approaches in the literature which address the issue of extracting place semantics, ranging from understanding natural language descriptions, experimentation with user-based population of place semantics and domain experts evaluations, to a hybrid of those techniques. Some of those methods address ontological issues, for example, the identification and classification of geographical place categories (Tversky and Hemenway 1983, Smith and Mark 2001, Edwards and Templeton 2005), while others focus on populating the ontologies, such as the web-mining techniques employed in the domain of GIR (Popescu *et al.* 2008).

Hierarchies of human actions in a geographical domain were extracted by analysing formal textual resources, such as the German traffic code documents in Kuhn (2000, 2001) and Sen (2007). Kuhn (2001) suggested a stepwise methodology to develop ontologies in support of human activities. The aim is to extract human action hierarchies, such as shown in Figure 1. Sen (2008) built on this work and employed a word frequency co-occurrence model on two traffic code texts to extract nested and sequential affordances.

In the above approaches, hierarchies of action entities have been considered independently of the geographical entities they are associated with, and a link from the derived action hierarchies needs to be made to the relevant entity classes that afford those actions. Specific text documents relevant to a particular domain were used in Kuhn (2001), where

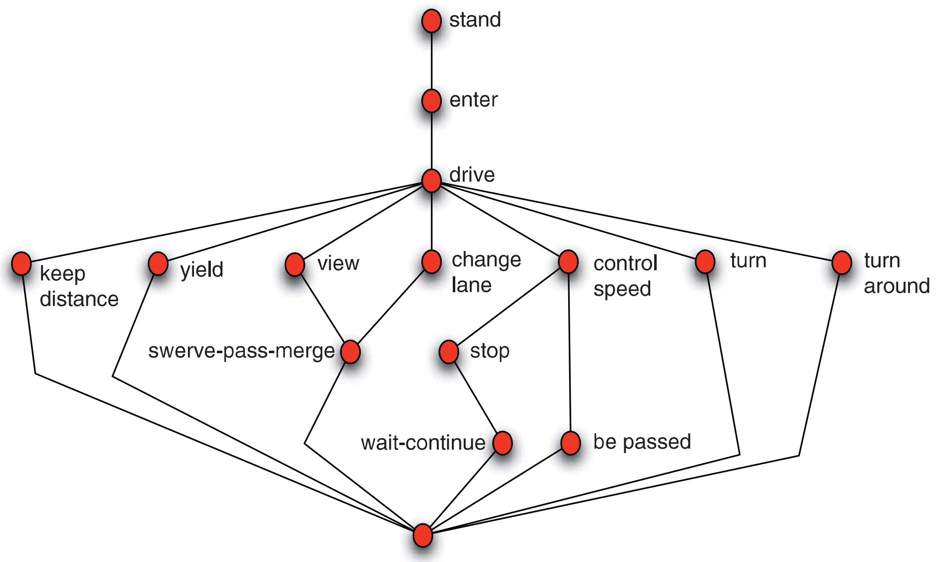

Figure 1. Action hierarchy derived from traffic code texts in Kuhn (2001) (reproduced from kuhn (2001)).

the automation of text analysis is noted to be problematic and manual intervention is needed. Sen (2007) investigated a simple case study of linguistic analysis of noun and verb phrases in formal texts of traffic codes and proposed a probabilistic approach to linking two parallel taxonomic hierarchies of entities and actions.

The work in this article is complementary to the above approaches in seeking to extract further classifying attributes of geographical place. However, the motivation is for associating place-type hierarchies with concepts of human activities and services particularly relevant to certain place types, as opposed to creating separate abstract action hierarchies for classifying place concepts. A linguistic approach is also adopted in analysing text documents, considering multiple resources of formal textual documents as well as web documents containing place-type definitions to automatically identify relevant service and activity concepts, as described later in the text.

3. Approach and methodology

Geographical places are normally associated with specific functions, economic activities or services they provide to individuals. These dimensions of a geographical place definition are typically evident in catalogues of place-type specifications, produced by national mapping and other geographical data collection agencies, and are used for the purpose of classification of place entities. For example, the following descriptions are parts of the definitions associated with place types in the Ordnance Survey Mastermap specification. [3]

> **Amusementpark:** A permanent site providing entertainment for the public in the form of amusement arcades, water rides and other facilities.
> **Comprehensive school:** A state school for teenagers, which provides free education.

Annotations assigned by people to a particular geographical place or place type are another valuable resource from which service and activity concepts can be harvested. For example, the *DBpedia* data set, created by extracting structured data from *Wikipedia*, provides detailed geographical information for over 400,000 places.[4] On the other hand, Wordnet and OpenCyc knowledge bases contain several descriptions of place types and human activities. In a mobile context, people annotations of geographical places are seen as a source of place semantics (Espinoza *et al.* 2001, Wang and Canny 2006). Classification of economic activities of business establishments is often used for place-type categorisation. For example, national bodies such as the Office of National Statistics of the United Kingdom (ONSUK)[5] and Eurostat (the statistical office of the European commission) produce classifications and definitions of economic activities, for classifying business establishments by the type of economic activity in which they are engaged.[6] Notably, a business place can be associated with a number of services, where some of these are principal activities that determine their primary classification while others are ancillary activities (such as accounting, transportation, purchasing, repair and maintenance) that exist solely to support the principal ones.

Hence, in this work, service concepts are modelled explicitly and linked to place types to which they are associated, as shown in Figure 2. The class of service types in this model is used to encompass both the notions of economic and other human activities commonly associated with a place type. Hierarchical relationships between instances of the place-type class (e.g. a school and a university are both types of an educational establishment) and those of the service-type class (e.g. swimming and tennis are both types of sport activities) are captured in the model using a self-association relationship.

Relevance of particular services to a place instance may depend on the application or context in which the model is used and which can therefore determine the strength of the association between the place type and its associated services and activities. One way of modelling this relation is by qualifying the association in the model (e.g. to indicate whether the service is primary or secondary and the time period in which it is offered in the place). In this work the focus is mainly on the identification of possible service and activity types associated with place types. The degree of relevance of the relationships between these concepts is, however, outside the scope of the current work.

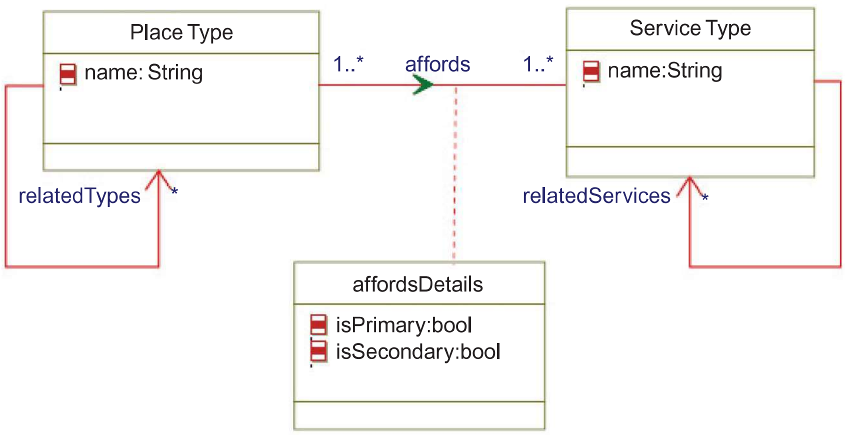

Figure 2. Modeling service and activity concepts with place types.

3.1. Methodology

The methodology of the approach is based on frequency-based lexical analysis of a corpus of place-type descriptions, denoted the *learning corpus*, to discover candidate patterns for service-and activity-type identification. The set of patterns identified is then applied on a corpus of multiple resources of place-type definitions, denoted the *construction corpus*, to extract possible service and activity types, which are then evaluated. The methodology is illustrated in Figure 3 and summarised in the following steps:

- A catalogue of place-type descriptions from the OS (referred to as the Real World Object Catalogue (OSRWOC) [7]) is used to populate place-type concepts. A total of 550 place types were identified, linked together in a hierarchy and stored (step 1 in Figure 3). The associated place-type descriptions from the catalogue form the *learning corpus*. The resulting place model will henceforth be referred to as the Real World Place Model (RWPM).
- Linguistic analysis of the *learning corpus* is applied to determine patterns of place descriptions. This involves two steps: First, lexical analysis of the place-type descriptions is undertaken using a Part of Speech Tagging (POST) system (step 2). In this work TreeTagger is used.[8] Second, analysis of the annotated corpus is undertaken to determine frequent keywords and expression patterns for place descriptions (step 3 in Figure 3).

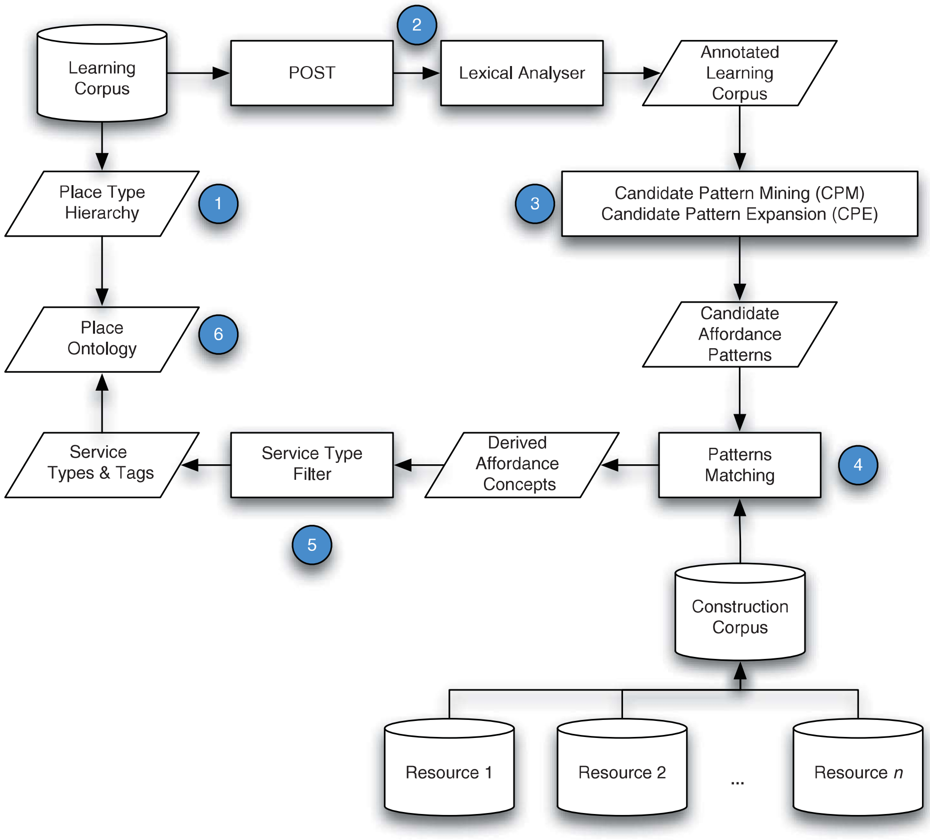

Figure 3. System architecture for service types extraction.

Table 1. Data sources for place-type definitions used as a construction corpus.

Data source	No. of place types
Ordnance Survey RWOC	555
Wikipedia	345
Oxford English Dictionary (OED)	273
OpenCyc	191
WordNet	192

- The extracted patterns are matched against a corpus of place-type descriptions, de noted the *construction corpus* (step 4). This is a store of place-type descriptions built using different resources, listed in Table 1. The result of this pattern-matching process is a set of linguistic expressions denoting affordance concepts (service/activity types and general tags) associated with place types.
- Finally, verb phrases, denoting the service/activity-type concepts, are filtered out from the derived collection above and stored with their corresponding place types (steps 5 and 6).

An overview of the processes involved in linguistic analysis is given in the following sections.

3.2. Candidate pattern mining (CPM)

The algorithm starts by iterating over each set of tokens, for each place type, constructing forward combinations of lemmas. For example, if the token set is $sT := \{l_1, l_2, l_3\}$, then the forward lemma combinations are $cP := \{l_1, l_2, l_3, l_1l_2, l_1l_3, l_1l_2l_3, l_2l_3\}$. The frequency of these lemmas in the learning corpus is calculated. Thresholds for the length and the frequency of occurrence of the resulting lemmas are set to determine a set of plausible patterns (a frequency threshold of 5 is used in this experiment). Further filtering of the patterns is done to exclude redundancy (patterns subsumed by others) (see Algorithm 1).

3.3. Candidate pattern expansion (CPE)

The output of the CPM is a set of candidate patterns, where each is a set of lemmas corresponding to a set of tokens from which the lemmas were derived, for example, *a place where* or *a place for*. Typically, a set of lemma tokens is followed by different tokens in different sentences of place-type descriptions. An expansion process is then applied to retrieve all those possible tokens in the place-type definitions. The algorithm concatenates a set of lemma tokens with the POST of its proceeding tokens. For example, a candidate pattern $\{l_1l_2\}$ from CPM may be expanded as follows: $\{l_1l_2DT\ VBN, l_1l_2VB\ NNS, l_1l_2 \ldots\}$, where *DT*, *VBN*, *VB* and *NNS* are examples of standard POST. A full list is given in Table 2.

Algorithm 1 Candidate Patterns Mining (CPM)
Input: Corpus of place types (pT), where each pT has a set of string tokens (sT), and each sT is annotated with its *POST* and *lemma*, no. of tokens threshold nT and frequency threshold fT.

Table 2. List of part of speech tags (POST) used for lexical analysis (Santorini (1990)).

POST	Definition	POST	Definition
CC	coordinating conjunction	DT	determiner
IN	preposition or subordinating conjunction	JJ	adjective
JJR	adjective, comparative	JJS	adjective, superlative
NN	noun, singular or mass	NNS	noun, plural
NP	proper noun, singular	NPS	proper noun, plural
TO	to	VB	verb, base form
VBD	verb, past tense	VBG	verb, gerund or present participle
VBN	verb, past participle	VBP	verb, non-3rd person singular present
VBZ	verb, 3rd person singular present	WDT	determiner
WP	pronoun	WRB	adverb

Output: A set of candidate patterns (cP)
1: for each sT in each pT, construct cP of forward combinations of lemmas
2: for each cP, count its frequency in corpus
3: sort cP set by frequency
4: for each cP, delete cP if its no. of lemmas $< nT$
5: for each cP, delete cP if its frequency $< fT$
6: sort cP set by no. of lemmas
7: for each cP, if cP is part of next cP and its frequency the frequency of next cP then delete cP

Algorithm 2 Candidate Patterns Expansion (CPE)
Input: A set of annotated place types pT and a set of candidate patterns (cP)

Output: Each candidate pattern is expanded with its proceeding POST in each token set in pT
1: for each cP do steps 2–3
2: for each pT do step 3
3: for each token set in pT, if cP is part of token set lemmas, then concatenate cP with proceeding POST until the end of token set

A pattern takes the form of {*place-keyword linker . . . POST**}. The list of *place-keywords* compiled from the place descriptions in the learning corpus is *place, building, establishment, facility, institution and organisation*. A *linker* is a word which may immediately follow a place-keyword, such as *where* or *of*. POST* is list of one or more part of speech tags and " . . . " indicates a nonconsecutive occurrence of tokens in the pattern. The list of POSTs identified in the patterns are then used to infer relevant place concepts. These take the form of concatenated sets of nouns and verb phrases. It is important to note that action verbs on their own are not sufficient to identify a service or activity. Instead, it is more meaningful to formulate a verb phrase, composed of a combination of a verb and one or more nouns. For example, a pattern such as *place where . . . NN* . . . VBN*, for example, *place where tennis is taught or played*, will produce the concepts *teachTennis* and *playTennis* instead of the verbs *teach* and *play*. The set of patterns identified in this experiment are listed in Section 3.4. The patterns identified are used to match against place-type descriptions in the construction corpus and to identify and automatically select relevant service concepts from the extracted nouns and verb phrases. It is possible that place descriptions in different resources in the construction corpus will yield similar verb phrases. These are filtered out manually to avoid redundancy.

3.4. Extracted patterns

The result of mining the learning corpus is a set of candidate place definition patterns. The most frequent patterns that were identified are listed below:

Pattern 1 is defined as

$$place - keyword\ where \ldots NN^{*} \ldots NNS^{*} \ldots VBN \ldots NN^{*} \ldots.$$

Concepts extracted are in the form of $VB^{*}+NN^{*}$ and $VB^{*}+NNS^{*}$. An example is *Tennis centre is an establishment where tennis is taught and played*. The lemma form is *an establishment where tennis be teach and play*; the POST form is *DT NN WRB NN VBZ VBN CC VBN SENT*; and the service types extracted are *teachTennis* and *playTennis*.

Pattern 2 is defined as

$$place - keyword\ of \ldots NN^{*} \ldots NNS^{*} \ldots.$$

Concepts extracted are in the form of NN^{*} and NNS^{*}. An example is *Winter Garden is an indoor garden containing exotic plants, or a place of entertainment, usually associated with coastal resorts*. The lemma form is *an indoor garden contain exotic plant, or a place of entertainment, usually associate with coastal resort*; the POST form is *DT JJ NN VBG JJ NNS, CC DT NN IN NN, RB VBN IN JJ NNS SENT*; and the concepts extracted are *entertainment* and *coastalResorts*. Note the adjective concatenation in *coastalResorts*.

Pattern 3 is defined as

$$place - keyword\ of \ldots VBG \ldots NN^{*} \ldots NNS^{*} \ldots.$$

Concepts extracted are in the form of $VBG^{*}+NN^{*}$ and $VBG^{*}+NNS^{*}$. An example is *Polytechnic: an institution of higher education offering courses at degree level or below.* The lemma form is *an institution of high education offer course at degree level or below*; the POST is *DT NN IN JJR NN VBG NNS IN NN NN CC RB SENT*; and the concepts extracted are OfferingHigherEducation, OfferingDegree, OfferingDegreeLevel and OfferingCourses.

Pattern 4 is defined as

$$place - keyword\ for \ldots NN^{*} \ldots NNS^{*} \ldots.$$

Concepts extracted are in the form of NN^{*} and NNS^{*}. An example *Hall is a large building for meetings or entertainment*. The lemma form is *a large building for meeting or entertainment*; the POST form is *DT JJ NN IN NNS CC NN SENT*; and the service types extracted are *entertainment* and *meetings*.

Pattern 5 is defined as

$$place - keyword\ in\ which \ldots NN^* \ldots VBN \ldots NNS^* \ldots NN^* \ldots.$$

Concepts extracted are in the form of *VB+NN** and *VB+NNS**. An example is *Corn mill. A gristmill or grist mill is a building in which grain is ground into flour, or the grinding mechanism itself*. The lemma form is *a gristmill or grist mill be a building in which grain be grind into flour, or the grind mechanism itself*; the POST form is *DT NN CC NN NN VBZ DT NN IN WDT NN VBZ VBN IN NN, CC DT VBG NN PP SENT*; and the concepts extracted are *grindGrain*, *grindFlour* and *grindMechanism*.

Pattern 6 is defined as

$$place - keyword\ used\ to \ldots NN^* \ldots VB \ldots NN^* \ldots.$$

Concepts extracted are in the form of *VB+NN**. An example *a smelter is an establishment used to extract metal from ore by melting. see building*. The lemma form is *an establishment use to extract metal from ore by melting. see building*; the POST form is *DT NN VBN TO VB NN IN NN IN NN SENT VB NN SENT VB NN SENT*; and the concepts extracted are *extractMetal*, *extractOre*, and *extractMelting*.

Pattern 7 is defined as

$$place - keyword\ used\ for \ldots NN^* \ldots VBG \ldots NN^* \ldots.$$

Concepts extracted are in the form of *VBG+NN*. An example is *A meteorological Centre is a facility used for studying and recording facts about the weather*. The lemma form is *a facility use for study and record fact about the weather*; the POST form is *DT NN VBN IN VBG CC VBG NNS IN DT NN SENT*; and the concepts extracted are *studyingWeather* and *recordingWeather*.

Pattern 8 is defined as

$$place - keyword\ VBG \ldots NN^* \ldots NNS^* \ldots.$$

Concepts extracted are in the form of *VBG+NN** and *VBG+NNS**. An example is *a civic centre is a building containing municipal offices*. The lemma form is *a building contain municipal office*; the POST form is *DT NN VBG JJ NNS SENT*; and the concept extracted is *ContainingMunicipalOffices*. Note that there is no *linker* in this pattern.

Pattern 9 is defined as

$$place - keyword\ that/which \ldots VB/VBG\ to/for \ldots NN^* \ldots NNS^*.$$

Concepts extracted are in the form of *VB/VBG+NN** and *VB/VBG+NNS**. An example is *a Leisure Pool is a facility that may include swimming pools, fun pools, flumes and other recreational activities*. The lemma form is *a facility that may include swimming pool, fun pool, flume and other recreational activity*; the POST form is DT NN WDT MD VB NN NNS, NN NNS, NNS CC JJ JJ NNS SENT VB NN SENT; and the concept extracted is IncludeSwimming, IncludeFun, IncludeSwimmingPools, IncludeFunPools, IncludeFlumes and IncludeRecreationalActivities.

4. Results and evaluation

The patterns identified were applied to place descriptions compiled in the construction corpus. A set of just over 1550 concepts was extracted containing a mixture of verb phrases and nouns. Verb phrases were filtered out and used as service and activity concepts, and nouns were stored as general tags of their corresponding place types. In many cases, the latter set of tags included references to the typical consumers and providers of the services for a particular place type. No attempt to refine this set was made in this work, but it should be possible to identify pattern expressions to decode some of these concepts in the future.

The system was developed using the Jena Semantic Web Toolkit[9] and the place model was represented in OWL. Figure 4 shows a snapshot of the model, showing a sample of place types and associated service concepts.

Employing multiple data sources in the construction corpus was valuable, as in many cases the sources complemented one another with regard to missing place types and service/activity concepts, as shown in Table 3. For example, for the place type *Arsenal*, the tag extracted from Wikipedia was *Weapons*, while the OSRWOC provided the concepts *StoreWeapons* and *MakeWeapons*.

Two evaluation tests were undertaken to measure the quality of the developed services and concepts. The first is a qualitative user-based evaluation and the second is an evaluation against an ontology created by experts. A set of 20 place types was used in both experiments for comparison.

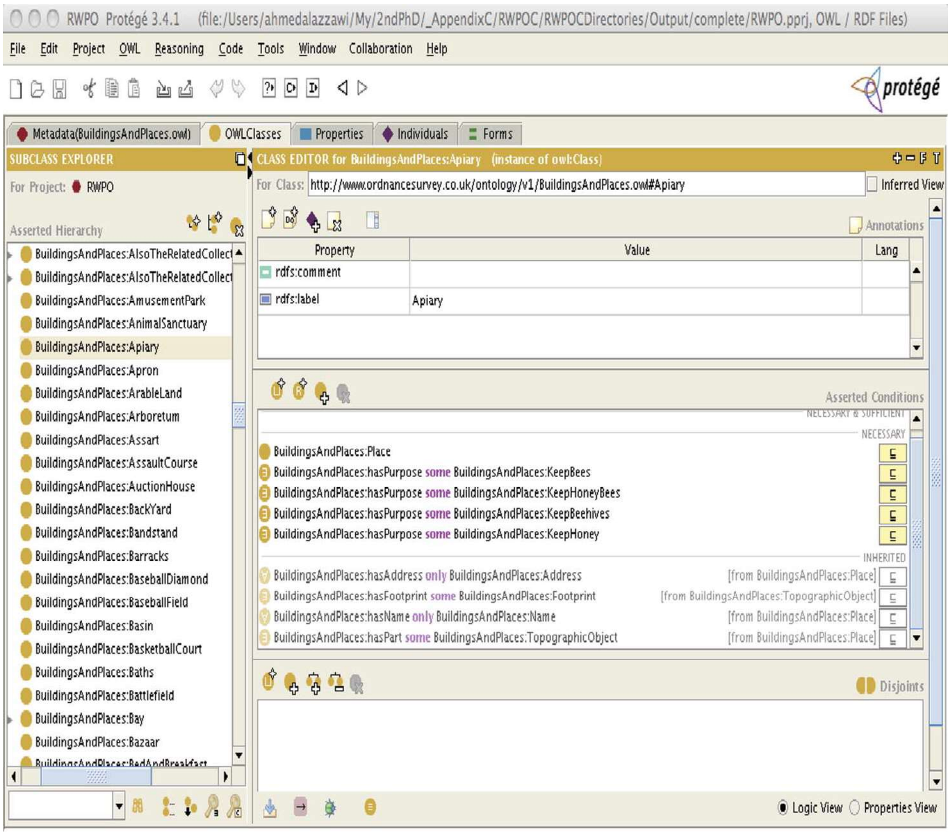

Figure 4. A snapshot of RWPO in Protégé.

Table 3. Sample of concepts extracted from individual data sources in the construction corpus.

	Service Concepts and General Tags				
Place type	OSRWOC	Wikipedia	OED	OpenCyc	WordNet
Fire station	Fire Equipment FireFighters Vehicles			KeepFireFighting Equipment KeepFirefighters	
Power station	Generate-Electricity GenerateLargeScale	Generation ElectricPower		Transformers Generators	
Stable		KeepLivestock KeepHorses	Housing HousingHorses	Farm Ranch Animals Horses	
Church	Worship ChristianWorship		ChristianWorship	HoldChristian Services HoldActivities	Public Worship
College	Students Learning		ProvidingHigherEducation ProvidingSpecializedTraining		
Weather station		ObservingWeather ObservingClimate Observing-Instruments ObservingAtmospheric Conditions ObservingWeatherForecasts			
Foundry	MeltMetal MeltGlass MeltShapes FormMetal FormGlass FormShapes			DoMetalCasting	
Glasshouse		CultivatePlants			Glass Cultivation Exhibition Plants Controlled-Conditions
Shipyard Arsenal	BuildShips RepairShips MakeAmmunition MakeMilitaryEquipment MakeWeapons StoreAmmunition StoreMilitatryEquipment StoreWeapons	BuildShips RepairShips Construction Repair Storage Ammunition Weapons	ConstructShips ConstructBoats		 ManufactureArms

4.1. User-based evaluation

A qualitative user-based evaluation in the form of an online questionnaire was carried out to assess the quality of the resulting concepts. Users were presented with a list of place types and asked to identify for each place type one or more typical service concepts that can be provided or activities that can be undertaken by people in a place of such a type. The semantic similarity between the service/activity concepts derived in this work and the concepts identified by the users was then measured. The questionnaire was hosted for a period of 2 weeks, during which students and staff from the university were invited to take part. A total of 872 people attempted the questionnaire, of which 388 (44.5%) completed the questionnaire. Around half of the participants were under 25 years old, 90.6% were native English speakers and 69.1% were females. The participants were asked to identify their familiarity with each of the given place types as either *familiar: you visit or have visited places of this type on a daily basis or a few times a week*; *somewhat familiar: you visited places of this type a number of times before*; or *not familiar: you have not visited places of this type before*. The distribution of user familiarity with place types is plotted in Figure 5a.

Semantic similarity was measured between the concepts identified by the participants and those derived and stored. The Measure of Semantic Relatedness (MSR) web service (Veksler *et al.* 2007) was employed for this purpose.[10] Two similarity measures were employed: Normalised Search Similarity (NSS) (Matveeva 2008) and Point-wise Mutual Information (PMI) (Turney 2001). Both are well-established measures that can handle large corpora and both support the calculation of the semantic distance between terms of multiple

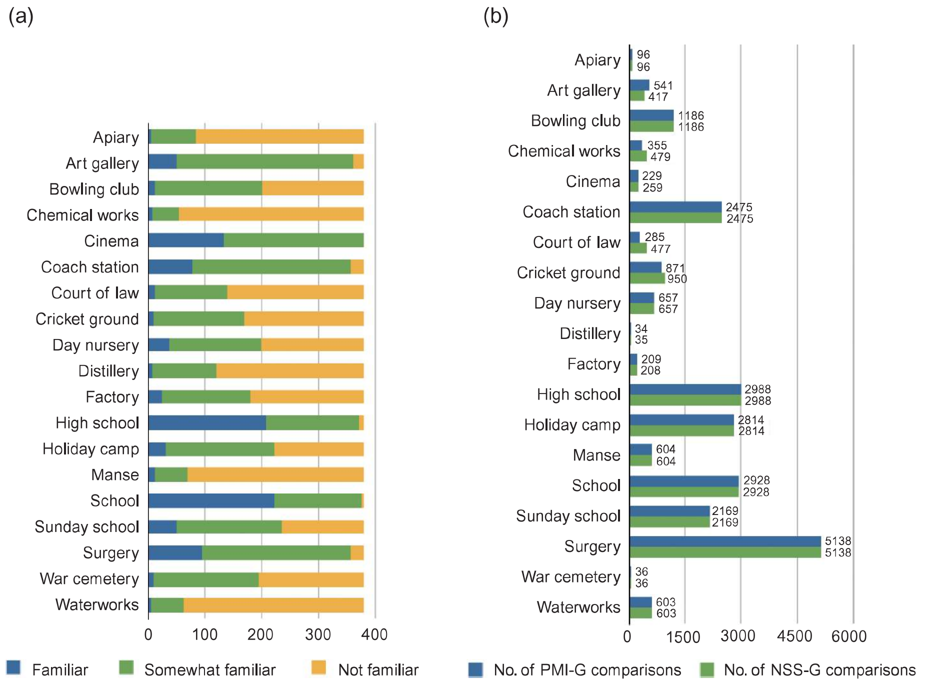

Figure 5. (a) The participants' familiarity with the place types used in the survey. (b) The number of terms compared for each place type using two measures of semantic relatedness. Note: NSS, normalised search similarity; PMI, point-wise mutual information.

words. MSRs define words in terms of their connection strengths to other words as determined by word co-occurrence. Hence, two terms are related if they often occur in the same contexts. Two terms are synonymous if their contexts are identical. However, the quality and the accuracy of the different MSRs was shown to be dependent of the size and the type of corpus used (Lindsey *et al.* 2007). A comparison of different MSRs can be found in Emadzadeh *et al.* (2010). In this work, both methods were used with the Google search engine and are thus denoted as PMI-G and NSS-G.

For each place type, a combined MSR is calculated as the average of relatedness of all the extracted terms in the RWPM with the terms submitted by users. To compare two terms, MSR sends search queries to the Google search engine (consisting of the original individual terms as well as of their concatenation, where a term may consist of multiple words). Based on the number of results returned, a semantic distance is calculated. Figure 5b shows the number of term comparisons for every place type using both MSRs. The number of term comparisons generally correlates with the participants' familiarity with the place types considered as shown in Figure 5. The result of the average MSR is plotted in Figure 6 for each of the place types considered using both NSS and PMI. The strength of the similarity using PMI, for most of the place types, approached or exceeded 50%. It is noted that the overall performance of the two MSRs was highly correlated (with a correlation coefficient of 0.795). A similar observation was also found in Lindsey *et al.* (2007). Figure 7 shows sample term comparisons using both NSS and PMI. For each pair of the terms compared, the first term is a service/activity type from the survey and the second is a concept from the RWPO. The experiment demonstrates how the automatically extracted service/activity types can be semantically close to the user perception of place affordance.

4.2. *Comparison with an expert ontology*

An ontology was developed by the Ordnance Survey to describe building features and the place classes surveyed, "with the intention of improving the use of the surveyed data and enabling semi-automatic processing of these data".[11] The ontology is expressed in a controlled natural language, called *Rabbit*, that supports the authoring of OWL ontologies by domain experts (Denaux *et al.* 2010).

A place concept in the ontology is defined as a kind of topographic object with the properties *purpose*, *address* and *name*. The property 'purpose' is used to hold notions of activity or processes associated with a place and is defined by experts and explicitly stored in the ontology. Table 4 shows the high correlation between the values of this property in the expert ontology with those derived and stored in the RWPM for the set of place types in this experiment. Figure 8 shows a comparison of the MSR for both the RWPO and the expert ontology with the concepts collected in the user study above.

5. Conclusion

This article describes an approach to the identification and extraction of concepts of services offered in a place and possible activities that can be practised there. The approach is based on mining a collection of services and service types from a variety of geographically focused resources, starting from an existing typology of place types. A rich resource of place service descriptions is used to identify frequent language patterns of place affordance. These patterns are then applied to several other resources – including ones from the web – to extract place type-specific services and concepts. A user-based evaluation experiment was conducted to evaluate the quality of the identified concepts. In addition,

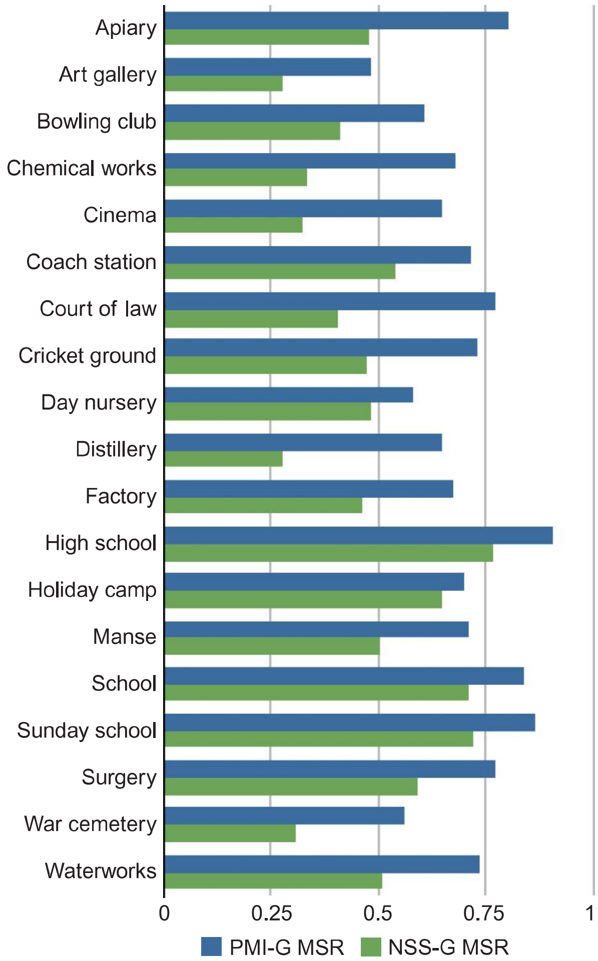

Figure 6. The average semantic relatedness between RWPO-derived concepts and those identified in the questionnaire. Note: RWPO, real world place object; PMI, point-wise mutual information; NSS, normalised search similarity; MSR, measure of semantic relatedness.

service concepts defined in an ontology created by domain experts were used for comparison. In both cases, a high degree of relevance of the automatically derived concepts was witnessed. The results of the experiments are encouraging and show the potential value of the approach to support the development of semantically rich place ontologies.

There is, however, much scope for extending and refining the approach in several directions, some of which are as follows:

- The place-keyword list can be expanded with semantically equivalent notions for *place*. For example, the use of the keyword *location* in the description of an Observatory in Wikipedia – 'a location used for observing terrestrial and celestial events' – can lead to the identification of service concepts: ObservingTerrestrialEvents and ObservingCelestialEvents, which are otherwise missed in the current system.

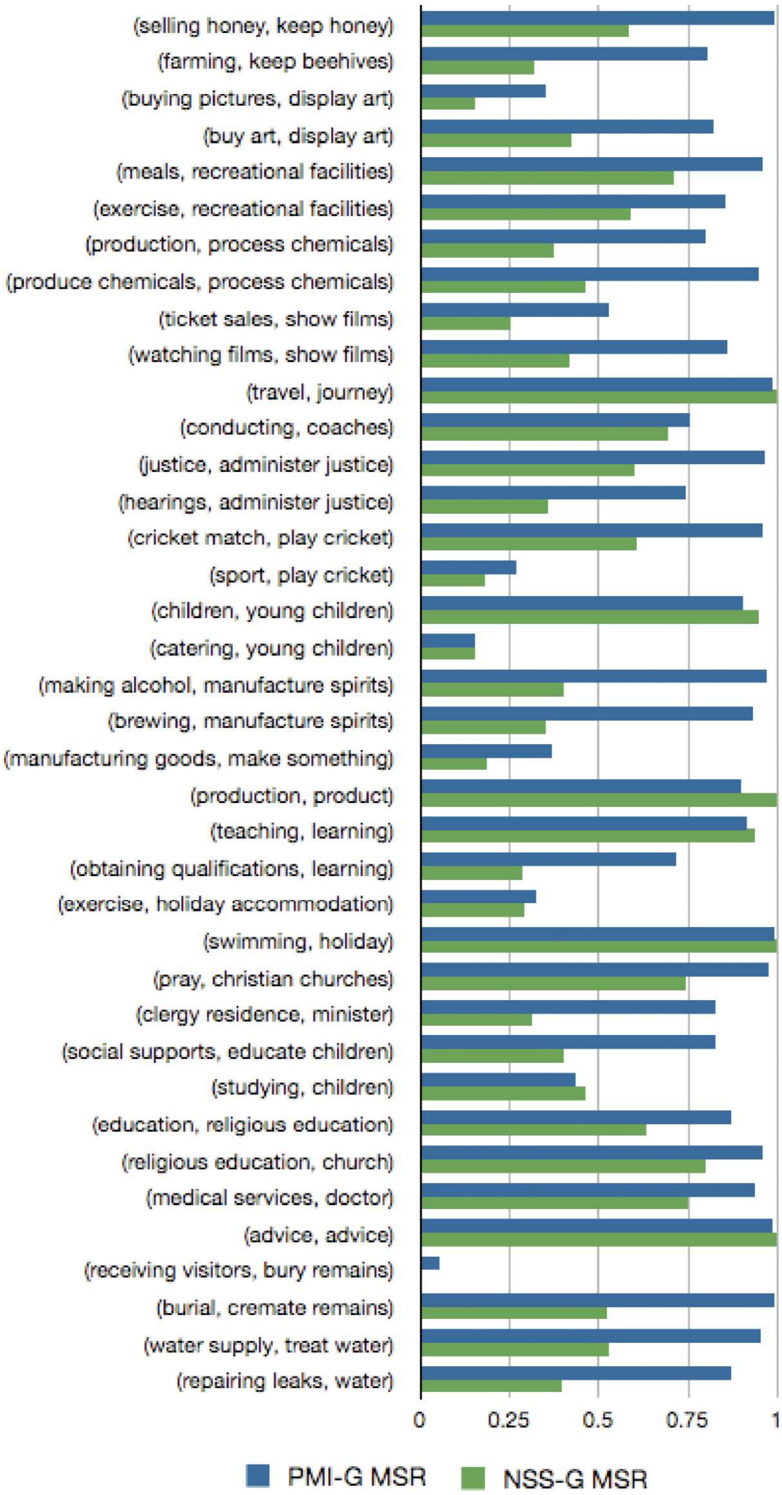

Figure 7. Sample terms comparisons using the semantic relatedness measures. The first is a term collected in the questionnaire and the second is a term in the RWPM. Note: RWPM, real world placemodel; PMI, point-wise mutual information; NSS, normalised search similarity; MSR, measure of semantic relatedness.

Table 4. Comparison of derived place concepts in the RWPO with an expert annotated ontology.

	Place Concepts	
Place type	RWPO	Expert ontology
Apiary	KeepHoney, KeepBeehives, KeepHoneyBees, and KeepBees	–
Art Gallery	DisplayArt, DisplayPublicViewing, and DisplayWorks	ArtExhibition
Bowling Club	RecreationalFacilities	PlayingBowls
Chemical Works	ProcessChemicals	ChemicalIndustrialProcessing
Cinema	ShowFilms	ShowingFilm
Coach Station	Journey, Coaches, and Passengers	CoachTransport
Court of Law	AdministerJustice	JusticeAdministration
Cricket Ground	PlayCricket	PlayingCricket
Day Nursery	YoungChildren	InfantCare
Distillery	DistilAlcoholicSpirits, and ManufactureSpirits	DistillingSpirit
Factory	MakeSomething, ConstructSomething, Product, Workers, Goods, Machines, ManufactureMachine, ManufactureGoods, AssembleMachine, and AssembleGoods	Manufacturing
High School	Learning, Teenagers, and Girls	–
Holiday Camp	Holiday, HolidayAccommodation, and Entertainment	–
Manse	Minister and ChristianChurches	HousingMinister_Religion
School	EducateChildren, Learning, and Children	Education
Sunday School	Church, ReligiousEducation, and Children	–
Surgery	Doctor, Dentist, Vet, Advice, Treatment, MedicalPractitioner, and Patients	HealthCareProvision
War Cemetery	CremateDeadBodies, CremateRemains, BuryDeadBodies, and BuryRemains	–
Waterworks	TreatWater, SupplyWater, Water, and WaterSupply	WaterTreatment and SupplyWater

- The pattern-mining process could be refined further using more resources for pattern learning. Also, the system can be trained to learn the degree of success of individual patterns in recognising relevant concepts.
- Methods for further processing of the extracted concepts are possible, for example, by ranking the relevance of the extracted service concepts in a particular application domain, or by further exploring the general tags to extract concepts such as a *service consumer* or a *service provider*.
- Finally, there is a need to model the temporal dimension of place affordance, where a place may be seen to afford different services or activities at different times.

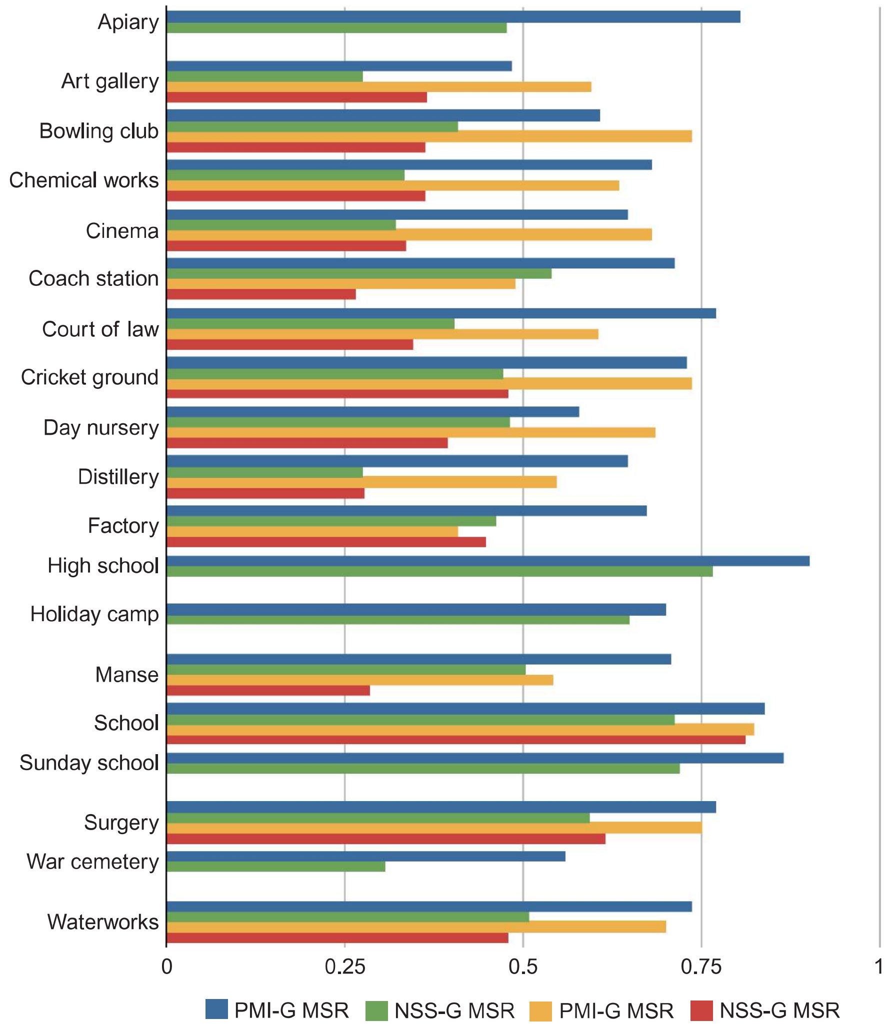

Figure 8. Average similarity measures of service concepts extracted for some place types compared with the benchmark set in the Ordnance Survey expert ontology. Note: PMI, point-wise mutual information; NSS, normalised search similarity; NSR, measure of semantic relatedness.

Similarly, the model of services and activities can be extended to support different views of affordance to serve the needs of different users, contexts and applications.

Future experiments to gauge the value of the approach are planned in the context of GIR for query expansion and retrieval tasks.

Acknowledgements

The authors thank Mark Hall from Cardiff School of Computer Science & Informatics for providing help with TreeTagger; Phil Smart for providing help with Protégé; and Vladislav Daniel Veksler, at CogWorks Laboratory, Rensselaer Polytechnic Institute, for providing help with their MSR Web service.

Notes

1. http://www.getty.edu/research/tools/vocabularies/index.html
2. http://www.ordnancesurvey.co.uk/oswebsite/ontology
3. http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/
4. http://wiki.dbpedia.org/datasets
5. http://http://www.statistics.gov.uk
6. See The Standard Industrial Classification of all Economic Activities (SIC), http://www.statistics.gov.uk/methods_quality/sic/downloads/sic2007explanatorynotes.pdf
7. OSRWOC: http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/information/technical/rwocinfo.html
8. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html
9. http://jena.sourceforge.net
10. http://cwl-projects.cogsci.rpi.edu/msr
11. Building and Place ontology: http://www.ordnancesurvey.co.uk/oswebsite/ontology

References

Abdelmoty, A., *et al.*, 2009. Supporting frameworks for the geospatial semantic web. *In*: *SSTD '09: proceedings of the 11th international symposium on advances in spatial and temporal databases*, Vol. Lecture Notes in Computer Science 5644, Aalborg, Denmark. Berlin: Springer-Verlag, 335–372.

Câmara, G., *et al.*, 2000. Action-driven ontologies of the geographical space: beyond the field-object debate. *In*: *Proceedings 1st international conference on geographical information science, GIScience*, Savannah, GA. New York: Springer, 52–54.

Denaux, R., *et al.*, 2010. Rabbit to OWL: ontology authoring with a CNL-based tool. *In*: *Proceedings of the 2009 conference on controlled natural language*, CNL'09, Marettimo Island, Italy. Berlin, Heidelberg: Springer-Verlag, 246–264.

Edwards, J. and Templeton, A., 2005. The structure of perceived qualities of situations. *European Journal of Social Psychology*, 35 (6), 705–723.

Egenhofer, M.J., 2002. Toward the semantic geospatial web. *In*: *Proceedings of the tenth ACM international symposium on Advances in geographic information systems*, McLean, VA. McLean, VA: ACM Press, 1–4.

Emadzadeh, E., Nikfarjam, A., and Muthaiyah, S., 2010. A comparative study on Measure of Semantic Relatedness function. *In*: *The 2nd international conference on computer and automation engineering (ICCAE)*, Singapore. Singapore: IEEE, 94–97.

Espinoza, F., *et al.*, 2001. Geonotes: social and navigational aspects of location-based information systems. *In*: *Ubicomp 2001: Ubiquitous computing*, 2–17.

Frank, A., 2003. Ontology for spatio-temporal databases. *In*: M. Koubarakis *et al.*, eds. Spatiotemporal database: the chorochronos approach. Vol. Lecture Notes in Computer Science 2520, Berlin: Springer, 9–77.

Gregory, I., *et al.*, 2002. The Great Britain Historical GIS Project: from maps to changing human geography. *The Cartographic Journal*, 39 (1), 37–49.

Hart, G., Temple, S., and Mizen, H., 2004. Tales of the river bank: first thoughts in the development of a topographic ontology. *In*: F. Toppen and P. Prastacos, eds. *Proceedings of the 7th AGILE Conference*, Crete, Greece. Heraklion: Crete University Press, 165–168.

Hill, L.L., Frew, J., and Zheng, Q., 1999. Geographic names: the implementation of a gazetteer in a georeferenced digital library. *D-Lib Magazine*, p. 5.

Jones, C., *et al.*, 2004. The SPIRIT spatial search engine: architecture, ontologies and spatial indexing. *In*: *Proceedings of the third international conference on geographic information science GIScience*, Vol. Lecture Notes in Computer Science 3234, Adelphi, MD. Berlin: Springer Verlag, 125–139.

Jones, C.B., Alani, H., and Tudhope, D., 2001. Geographical information retrieval with ontologies of place. *In*: *Proceedings of the international conference on spatial information theory, COSIT*, Vol. Lecture Notes in Computer Science 2205, Morro Bay, CA. Berlin: Springer-Verlag, 322–335.

Jordan, T., *et al.*, 1998. An affordance-based model of place in GIS. *In*: T. Poiker and N. Chrisman, eds. *Eighth International Symposium on Spatial Data Handling*, Vancouver, BC. Vanconver, BC: IUG, 98–109.

Kuhn, W., 2000. Ontologies from text. *In*: M. Egenhofer and D. Mark, eds. *International Conference on Geographic Information Science (GIScience 2000)*, Savannah, GA. Boca Raton, FL: Taylor & Francis.

Kuhn, W., 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15 (7), 613–631.

Kuhn, W., 2007. An image-schematic account of spatial categories. *In*: *Spatial information theory*, Vol. Lecture Notes in Computer Science 4736, Melbourne. Berlin: Springer-Verlag, 152–168.

Lindsey, R., *et al.*, 2007. Effects of corpus selection on semantic relatedness. *In*: *8th international conference of cognitive modeling, ICCM2007*, Michigan, Boca Raton, FL: Taylor & Francis.

Matveeva, I., 2008. Generalized latent semantic analysis for document representation. Thesis (PhD). University of Chicago, Chicago.

Perry, M., Hakimpour, F., and Sheth, A., 2006. Analyzing the space, and time: an ontology-based approach. *In*: *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, Arlington, VA. Arlington, VA: GIS '06 ACM, 147–154.

Popescu, A., Grefenstette, G., and Moëllic, P.A., 2008. Gazetiki: automatic creation of a geographical gazetteer. *In*: *ACM/IEEE Joint Conference on Digital Libraries*, Pitsburgh, PN. New York: ACM, 85–93.

Raubal, M. and Kuhn, W., 2004. Ontology-based task simulation. *Spatial Cognition and Computation*, 4 (1), 15–37.

Relph, E., 1976. *Place and placelessness*. London: Pion Ltd.

Rugg, R., Egenhofer, M., and Kuhn, W., 1997. Formalizing behavior of geographic feature types. *Geographical Systems*, 4, 159–180.

Santorini, B., 1990. *Part-of-speech tagging guidelines for the penn treebank project*. Philadelphia, PA: University of Pennsylvania, School of Engineering and Applied Science, Department. of Computer and Information Science.

Scheider, S. and Kuhn, W., 2010. Affordance-based categorization of road network data using a grounded theory of channel networks. *International Journal of Geographical Information Science*, 24 (8), 1249–1267.

Sen, S., 2007. Two types of hierarchies in geospatial ontologies. *In*: *Second International Conference, GeoS 2007*, Mexico, Mexico: Springer, 1–19.

Sen, S., 2008. Use of affordances in geospatial ontologies. *In*: *Proceedings of the 2006 international conference on Towards affordance-based robot control*, Dagstuhl Castle, Berlin, Heidelberg: Springer Verlag, 122–139.

Smart, P., Jones, C., and Twaroch, F., 2010. Multi-source toponym data integration and mediation for a meta-gazetteer service. *In*: *Sixth international conference on Geographic Information Science, GIScience 2010*, Vol. Lecture Notes in Computer Science 6292, Zurich, Switzerland. New York: Springer, 234–248.

Smith, B. and Mark, D.M., 2001. Geographical categories: an ontological investigation. *International Journal of Geographical Information Science*, 15 (7), 591–612.

Turney, P.D., 2001. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *In*: *Proceedings of the 12th european conference on machine learning*, EMCL '01, Freiburg, Germany. London: Springer-Verlag, 491–502.

Tversky, B. and Hemenway, K., 1983. Categories of environmental scenes. *Cognitive Psychology*, 15 (1), 121–149.

Veksler, V., *et al.*, 2007. A proxy for all your semantic needs. *In*: *29th annual meeting of the cognitive science society, CogSci2007*, Nashwille, TN. Nashville, TN: Cognitive Science Society.

Wang, J. and Canny, J., 2006. End-user place annotation on mobile devices: a comparative study. *In*: *CHI '06 extended abstracts on Human factors in computing systems*, CHI '06, Canada. New York: ACM, 1493–1498.