J Data Semantics manuscript No.
(will be inserted by the editor)

Capturing Place Semantics on the GeoSocial Web

Ehab ElGindy · Alia Abdelmoty

Received: date / Accepted: 2 January 2014

Abstract Massive interest in geo-referencing of personal resources is evident on the web. People are collaboratively digitising maps and building place knowledge resources that document personal use and experiences in geographic places. Understanding and discovering these place semantics can potentially lead to the development of a different type of place gazetteer that holds not only standard information of place names and geographic location, but also activities practiced by people in a place and vernacular views of place characteristics. In this paper a novel framework is proposed for the analysis of geo-folksonomies and the automatic discovery of place-related semantics. The framework is based on a model of geographic place that extends the definition of place as defined in traditional gazetteers and geospatial ontologies to include the notion of place affordance. The derived place-related concepts are compared against an expert formal ontology of place types and activities and evaluated using both a user-based evaluation experiment and by measuring the degree of semantic relatedness of the derived concepts. To demonstrate the utility of the proposed framework, an application is developed to illustrate the possible enrichment of search experience by exposing the derived semantics to users of web mapping applications.

Keywords Place semantics · Geographic information retrieval · Geo-social web

Ehab ElGindy
School of Computer Science and Informatics
Cardiff University
E-mail: ehab.elgindy@cs.cardiff.ac.uk

Alia Abdelmoty
School of Computer Science and Informatics
Cardiff University
E-mail: a.i.abdelmoty@cs.cardiff.ac.uk

1 Introduction

Geo-tagging of resources on the web has become prevalent. Geographic referencing has evolved to become a natural method of organising and linking information with the aim of facilitating its discovery and use. GPS-enabled devices allow people to store their mobility tracks, tag photos, and events. In response, many applications on the web are enabling geo-tagging of resources, e.g. geo-locating photos on Flickr[1] and tweets on Twitter[2], and people are collaboratively building their own map resources and gazetteers (e.g. GeoNames[3] and OpenStreetMap[4]). Whereas typical place name resources provided by mapping agencies, referred to as geographic thesauri, record the name and map coordinates of a place, collaborative mapping on the Social Web provides an opportunity for people to create maps that document their social and personal experiences in a place. Thus university buildings may be a place of work and study for a group of people, a conference venue for another group, and a sports facility for a different group. Understanding and encoding the information provided by users for place name resources can eventually result in a different type of place gazetteer that documents not only where a place is located, but also what happens at a place, and hence providing an opportunity for a much richer, and possibly personalised, search experience. In this paper we focus on geo-folksonomies created on web mapping applications. A *geo-folksonomy* records the tags used by users to annotate place resources on geographic maps. Some examples of applications that generate such folksonomies are Tagzania[5] and Wikimapia[6].

[1] http://www.flickr.com
[2] http://www.twitter.com
[3] http://www.geonames.org
[4] http://www.openstreetmap.org
[5] http://www.tagzania.com
[6] http://www.wikimapia.org/

Research on folksonomies produced methods for tag analysis that mainly reflect the frequency of utilisation and association of tags to resources [40]. Further analysis of these tags can be done to discover semantic relationships and thus build taxonomies to reflect vocabulary of annotation for different contexts [9, 31]. These methods can be very helpful in guiding the search and querying of these resources and the visualisation of their content [13]. Two categories of semantics associated with geographic places can be identified; spatial semantics and non-spatial semantics. Spatial semantics are those related to the definition of spatial location, boundaries and shape of the geographic place. Non-spatial semantics are used here to refer to other properties of a geographic place that are not spatial, e.g. a place name or type. Recently, some efforts have targeted the identification and discovery of the spatial aspects of place definition from web resources, [19, 37],as well as some non-spatial aspects, such as vernacular place names. The notion of place affordance, defined as the purpose the place serves for its users or the activities that can be carried out in a place is recognised as an important dimension of place definition in the geo-community. Whereas some general or standard notion of affordance can be associated with a class of geographic places, for example, associating a school with learning and teaching and a bank with money lending, etc., different individual experiences of users for the same place can be recognised and identified by analysing their tagging behaviour of place resources on the Social Web. In this paper, a framework is proposed for discovering non-spatial place semantics in geo-folksonomies. The framework is based on a model of place that encodes the notion of activities and services afforded by a place, as well as users' sentiment reflection of experience in a place. The work builds on and extends existing work on folksonomy analysis and suggests a geographically-oriented and semantics-guided approach to tag resolution and ontology building from geo-folksonomies. Due to the nature of data collection and the inaccuracy of resource allocation by users, an important initial step was needed to cluster place resources and reconstruct the geo-folksonomy. Existing ontological resources are used for matching and identification of place type and activity concepts, and statistical methods of folksonomy analysis guide the discovery of relationships between activity and place type concepts. A realistic geo-folksonomy resource is used for evaluation. The resulting place activity and place type ontologies are compared against standard ones developed by experts and used by national mapping agencies. A user-based evaluation is carried out to establish the validity of derived ontological relationships. The utility of the proposed framework is demonstrated with a prototypical application that projects the discovered place semantics alongside the traditional tag clouds associated with place resources in geo-mapping applications. Results illustrate the potential of the approach for the dynamic discovery of user-induced place semantics that essentially offers a different and complementary view to that provided by traditional formal place information resources.

The rest of the paper is structured as follows. Related work on folksonomy analysis and semantics of geographic place is reviewed in Section 2. A model of place that encapsulates the notion of place type and affordance is presented in Section 3. A framework for inferring a place ontology from a geo-folksonomy is proposed in Section 4. Data collected and evaluation experiments are described in Section 5. An example application to demonstrate the utility of the approach is presented in section 6 and conclusions and an overview of future work are given in Section 7.

2 Related Work

Folksonomy Analysis

Vast amounts of data are generated by users' collaboration and interaction on Web 2.0 applications. For example, Flickr has thousands of photos uploaded every minute (about 4.5 million daily)[7]. The folksonomy structure generated by these applications is made up of three entities; users, resources and tags, as well as relationships between them [17]. Recognising the value of the implicit semantics in these data, research work has recently been targeted at extracting and structuring these semantics [9, 22, 38]. Semantics extracted from folksonomies capture users' perception of a specific domain, which can be different from the formal information models representing that domain. Such semantics can be utilised to enhance the user experience on the web, e.g. semantic tag recommendation systems [3]. Different statistical methods are used to build taxonomies or thesauri of concepts from these folksonomies. For example, Mika [26] introduced a method based on Social Network Analysis (SNA) which makes use of different relationships between all the entities in a folksonomy. Other research works focussed mainly on analysing relationships between resources and tags and ignored the user dimension. For example, Schmitz [34] introduced a probabilistic model of subsumption, based originally on a subsumption model by Sanderson and Croft [33], to extract the parent-child relationships between tags and resources. The work in [24] considered the user dimension by introducing a pre-processing (aggregation) step, where the folksonomy is transformed from the tripartite structure of users, tags and resources to a bipartite structure of tags and resources while the users' relationships are represented as weighted edges between tags and resources. In general, semantics captured from folksonomies are represented in the form of a thesaurus, where relationships between concepts are defined by the monolingual the-

[7] http://www.flickr.com/photos/franckmichel/6855169886/

sauri standard (ISO 2788), such as "broader than", "narrower than" and "related to".

Semantics of Geographic Places

Basic geospatial models of geographic space capture the notion of geographic features and their identity. This is achieved through reference to properties defining locations of features in space and their geographic classification or type. For example, the OGC Reference Model (ORM)[8] provides a general feature model designed to characterise features, feature types and the relations between features. Over the past decade, there have been many different attempts to create a geospatial RDF standard to support the representation and sharing of geo-referenced information on the web. Several different organizations, including the W3C, research groups, and triplestore vendors have created their own ontologies and strategies for representing and querying geospatial data. For example, the Basic Geo Vocabulary was proposed by the W3C Geospatial Incubator Group [9]. It follows the GeoRSS feature model[10] to allow for the description of points, lines and polygon geometries and their associated features. The group also produced the GeoOWL ontology[11] which provides a detailed and flexible model for representing geospatial concepts [6].

The above approaches focused primarily on modelling the spatial aspects of geographic features, particularly capturing the location and spatial extension of features in space. Recently, collaborative web mapping 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 10 million[12] geographic names. Also, several research works have considered the problem of building gazetteers from user-generated data on Web 2.0 [31]. On the Semantic Web, place name (or toponym) ontologies are employed to facilitate the utilisation of gazetteers to support geographic information retrieval (GIR) tasks, such as disambiguation and expansion of terms in search engine queries [1]. Ballatore and Bertolotto [5] considered the combined use of the dbpedia ontology and volunteered geographic information resources to inform spatial exploratory queries by providing a view of the semantic content of the spatial data of interest to the user.

An ontology of place names is defined as a model of terminology and structure of geographic space and named place entities [1]. 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 geographic place [29]. In addition to place qualification using place type categorisation, qualitative spatial relationships commonly used in search queries, such as, *inside* and *near*, are also modelled to relate place instances.

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. For Relph [32], 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 geographic space [20]. The term "action-driven ontologies" was first coined by Camara et al. [7] 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 [35]. 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 geographic information, the Ordnance Survey (OS) (the national mapping agency for the UK) proposed its utilisation as one of the ontological relations for representing their geographic information [15] and made an explicit use of a "*has-purpose*" relationship in building their ontology of buildings and places [13].

The work in this paper combines and extends research works in the general area of folksonomy analysis and the area of discovering place semantics from web resources. A model of place is utilised that captures, in addition to basic spatial representation of location, the notion of place affordance. The model then serves as a base for a framework that follows a geographically-oriented approach to discovering semantics from folksonomies. The results of this work also complements efforts in building gazetteers of geographic features from user-generated data.

3 Modelling Place Semantics

Geographic places are normally associated with specific functions, services, economic activities or other human activities that they provide to individuals. This dimension of a geographical place definition is typically evident in catalogues of place type specifications produced by national mapping

[8] `http://www.opengeospatial.org/standards/orm`

[9] `http://www.w3.org/2005/Incubator/geo/XGR-geo-ont/`

[10] `http://www.opengeospatial.org/pt/06-050r3`

[11] `http://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/W3C\_XGR\_Geo\_files/geo\_2007.owl`

[12] `http://www.geonames.org/`

[13] `http://www.ordnancesurvey.co.uk/oswebsite/ontology`

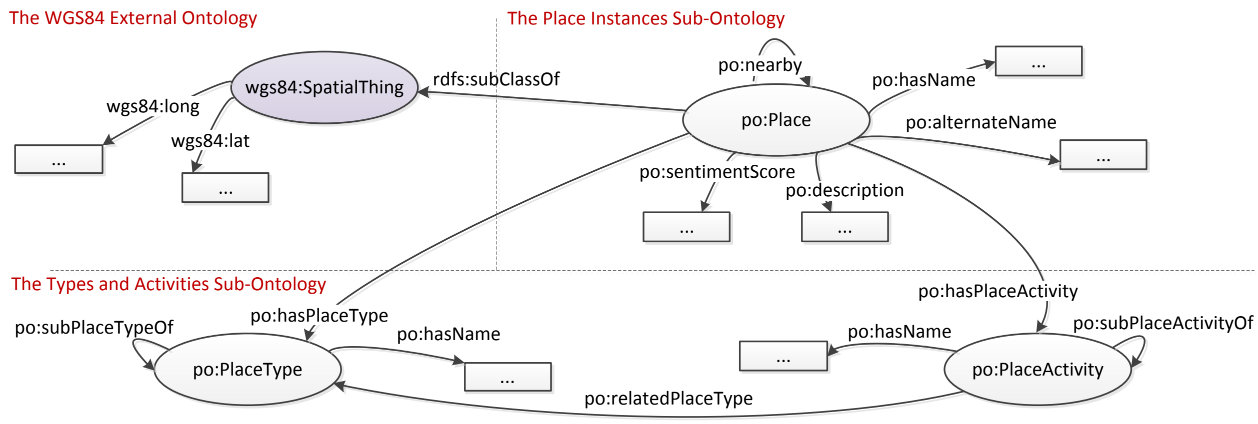


Fig. 1 Place ontology represents the place semantics captured from folksonomies

and other geographic data collections 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[14]: *Amusement park; a permanent site providing entertainment for the public in the form of amusement arcades, water rides and other facilities*, and a *Comprehensive school; a state school for teenagers, which provides free education*.

Whereas these formal classification of place types and services are useful and needed for many contexts, they are general and are not intended to capture any specific experiences of users in a place. There is an emergent need for recognising and sharing the experiences of people in geographic places, evident from the ever-growing volumes of data and applications that allow users to check-in and tag places. Such experiences are associated with particular instances of geographic place and may not be generalised.

Proposed Place Ontology

In this work, we adopt a model where a geographic place can be associated with possibly multiple place types and place activities. Place types and place activities may themselves form individual subsumption hierarchies. A place type may be associated with more than one type of activity and vice versa. A distinguishing characteristic in this model is that it allows for a specific place instance to be associated with an activity that may not be derived from its association with a specific place type. Hence, for example, a specific instance of a school may be associated with several place types, such as, primary school, public school, nursery, from which it can derive activities, such as learning and teaching, but also be associated with activities, such as, dancing, weight training, and adult education, where it offers external services to the community after school hours, etc. The former list is derived from the association with a particular place type, but the later list may come from direct annotation by users of the place. The model is encoded as a place ontology as shown in Figure 1.

The ontology contains three concepts: *Place*, *Place Type* and *Place Activity* as well as properties and inter-relationships between them. The spatial location of a place is modelled by extending the WGS84 *SpatialThing* concept to inherit the spatial properties *lat, long*. A *Place* has a *name* and possibly 0 or more *alternate names* and may be involved with different types of spatial relationships with other place instances. Spatial relationships are adopted in various proposals of place ontologies such in SPIRIT [18], TRIPOD [2] and GeoNames. It is noted that the ontology extends previous proposals, for example, that of the Ordnance Survey Building and Place ontology (OSBP)[15], where a similar notion of place activity is explicitly modelled and associated with a place type through a relationship “has-purpose”. The difference in this paper is that a place concept is introduced which also exhibit separate relationships between types and activities. In addition, inter-relationships between place types and place activities were not modelled in the OS ontology.

The design of the place ontology is implemented using OWL and all classes and properties are qualified with the prefix **po**[16]. Note that in general, the associations in this model are dynamic as accumulation of users' experiences and annotation accumulate. Hence, the relationships $po : hasPlaceType$, $po : hasPlaceActivity$ and $po : relatedPlaceType$ would be time-stamped. However, the time aspect is not con-

[14] `http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap`

[15] `http://www.ordnancesurvey.co.uk/oswebsite/ontology`

[16] Ontology can be downloaded at `http://cs.cardiff.ac.uk/2010/place-ontology#`

sidered in this current work and is the subject of future research.

4 A Framework for Discovering Place Semantics From Geo-Folksonomies

The goal of the approach proposed here is to derive an understanding of implicit place semantics from geo-folksonomies. Starting with “raw” folksonomy resources, the framework proposed involves three main stages: a) folksonomy pre-processing, b) tag resolution, and c) semantics association and ontology building. A particular characteristic of geo-folksonomies is the possible redundancy in place resource creation and the resulting fragmentation of folksonomy relationships that can affect the quality of the analysis. The first stage in the proposed approach thus involves two main tasks; a) cleaning the tags to filter out noise such as stop words, and b) clustering of place resources and the reconstruction of the folksonomy structure. The tag resolution stage involves domain-dependent analysis tasks for resolving and isolating tags that refer to domain concepts. The approach proposed here is to utilise existing domain ontologies for matching domain concepts. In our case, the process involves identification and building place type and human activity ontology bases and using those as reference sources for matching against the tag collection. The final stage is the semantics association and ontology building stage, where the individual identified domain-dependent tag collections are first analysed to derive relationships and create ontologies using the folksonomy structure. In our case, a place type sub-ontology and a place activity sub-ontology are created to represent a folksonomy-specific view of these concepts. A tag integration process is then applied to link the tags from both sub-ontologies using the inherent folksonomy relationships. The resulting structures are associated with the clustered place resources from the first stage and used to populate the place ontology. Further semantic analysis can be applied to the tag collection. Here, a sentiment analyser is developed to estimate a sentiment score for each place resource. An outline of the framework is shown in figure 2 and the different stages are described in more detail below.

4.1 Folksonomy Pre-processing Stage

A data collection process is first used to build a local geo-folksonomy repository. A crawler software is developed to process pages from Tagzania[17]. Tagzania is a geo-social tagging application where users are able to collaboratively create and annotate geographic places on a background map.

[17] http://www.tagzania.com

4.1.1 Tag Cleaning

Social tagging applications do not normally support input validation on the tags provided by users. This model of interaction is intentional and is expected by users to increase flexibility of use. Table 1 lists some identified problems in the tags and examples thereof.

Problem	Example Tags
Stop words such as articles and pronouns	a, an, the, we
Dialect	center, centre
Morphological forms of same word	shop, shops, shopping
Numbers	20, 505, 2007
Synonyms	chair, seat
Homonyms	mean
Abbreviations	UK, EU
Concatenated terms	CardiffUniversity
Non-alpha-numeric letters	"ball
URLs	www.google.co.uk

Table 1 Sample of possible problems in the tag collection.

Other mis-conceptions of tag usage include wrapping a whole sentence in quotes. For example, a tag such as "this is my house", will result in 4 separate tags for each word (including the quotes) in the sentence. The cleaning process used here involves the following sequence of steps:

1. Removal of special characters. All non alphanumeric characters are removed from tags. For instance, the tag *Cardiff&* is changed to *Cardiff*.
2. Filtering of all tags that are just one character in length.
3. Filtering of tags that represent URLs.
4. Filtering of stop-words. A list of 116 stop words, published by Microsoft [18] is used.
5. Removal of duplicate tags. Duplicate are removed in such a way as to preserve the relations between place resources and users.

Language-related issues such as synonyms, homonyms and dialects are not considered here, but can be considered in a more detailed tag cleaning process in the future.

4.1.2 Clustering Place Resources

Implications of uncontrolled data input in geo-tagging applications can affect the accuracy of the place resources defined and used. In particular, imprecision is evident in two aspects of place definition as follows:

1. Imprecise place locations, where users do not have the knowledge (or keenness) to define and digitize a precise location for a place using the map interface provided. Hence, multiple approximate points could refer to the

[18] http://msdn.microsoft.com/en-us/library/bb164590(v=vs.80).aspx

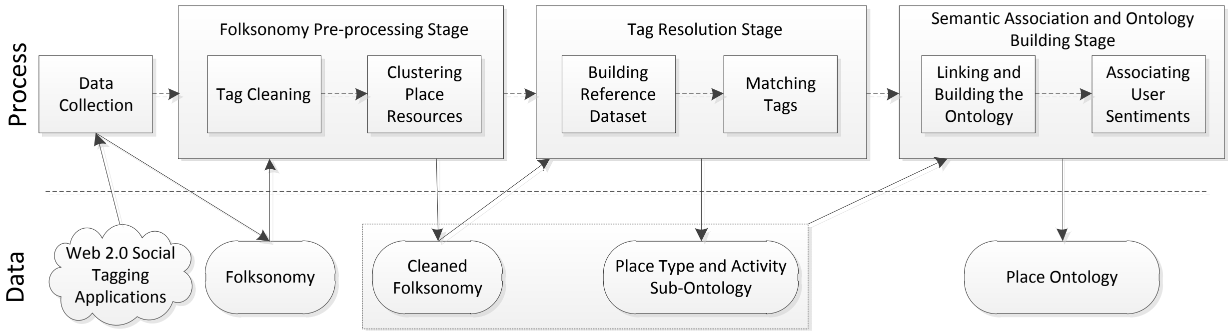


Fig. 2 The process of building a place ontology from a geo-folksonomy

location of the same place instance. The problem is related to the size of the geographic places considered. For example, it is harder for a user to identify a point representing a city than a point representing an individual building. The problem is also related to the scales of the maps offered to users and the complexity of matching precise locations across different map scales.

2. Imprecise, vernacular and multilingual place names, where users commonly use non-standard names and abbreviations for geographic places. Hence, multiple names are used to refer to the same place instance e.g. "Cardiff" and "Caerdydd".

The above problems lead to misclassification and duplication of place resources in the folksonomy which would affect its quality. Hence, a process of clustering similar place resources is needed to enhance the certainty of the contained information in the folksonomy. A two-step clustering process based on the analysis of assigned spatial location and place names is used as follows:

1. First a spatial clustering process is applied using a spatial similarity measure to group place resources based on their relative proximity.
2. This is followed by a textual clustering process to isolate resources from the identified groups above based on similarity of given place names.

Spatial Clustering: The main objective of using a spatial similarity measure is to find place instances that are in close proximity to each other. This can be achieved by using cluster analysis algorithm or by consulting external reverse geocoders to assign a unique area code for each place resource, and then area codes can be used as clusters identifiers.
The Quality Threshold (QT) clustering algorithm [16] is used here. It has the advantage of not requiring the number of clusters to be defined apriori, compared for example to other classical clustering approaches, such as the K-means clustering [10]. In general, the QT algorithm assigns a set of objects into groups (or clusters), where objects in the same cluster satisfy a pre-defined threshold function. In our case, place resources are added to a cluster if they are located within 500 meters from the centre of that cluster. Two methods are considered for reverse geo-coding the point locations of place resources (i.e. to identify a place given its spatial location); the Yahoo Where on Earth ID (*WOEID*) service and a postcode reverse geo-coding service. The *WOEID* web service provides a unique identifier for any geographic location based on the closest street to that location. Hence, place resources with the same *WOEID* can be considered close, as they all have a common closest street. The postcode reverse geo-coding service, published by GeoNames[19], provides a method that returns the postcode of any given spatial location. The service is used to resolve the postcodes of the places resources used in this experiment.

ID	WOEID	Unit Level PC	District Level PC
31758	44417	SW1A 0AA	SW1A
31759	44417	SW1A 0AA	SW1A
31760	44417	SW1A 2JR	SW1A
31761	44417	SW1A 2JR	SW1A
31762	44417	SW1A 0AA	SW1A
49775	44417	SW1A 2JR	SW1A
49776	44417	SW1A 0AA	SW1A
49777	44417	SW1A 0AA	SW1A

Table 2 Place resources referring to *Big Ben* in London, with their corresponding derived WOEIDs, postcodes (PC).

In table 2 place resources are shown representing the clock tower of "Big Ben", located in the Palace of Westminster in London. Each resource is shown with its derived WOEID and postcode. As shown in the table, all instances are grouped into one WOEID, while the postcode divides the resources into two groups, with a common district-level code (*SW1A*), but separate unit-level codes. The unit-level postcode divisions are too restrictive in this context. Also,

[19] `http://www.geonames.org/export/web-services.html`

the district-level postcodes are much too broad and are also likely to produce wrong clusters. In addition, postcode systems vary from one country to another, whereas the WOEID system of identification is more universal. All the resources in the table were identified as belonging to a single cluster using the QT clustering.

Further experimentation with the data set confirmed that both the qualitative clustering using the WOEID and the QT clustering method are both highly successful in producing correct clusters. The QT method is however, computationally expensive with a time complexity of $O(knt_{dist})$ where k is the number of clusters, n is the number of place resources and t_{dist} is the time needed to calculate the distance between the place resources, which limits its application for large data sets.

Textual Clustering: After an initial clustering of place resources using their spatial location, a second step of filtering out the clusters is applied based on place name similarity. The Levenshtein distance [21] is a method used for measuring text similarity. The Levenshtein distance or "edit distance" between two strings is the minimum number of edits needed to transform one string into another, where the allowed edit operations are insertion, deletion or substitution of a single character. Unlike folksonomy tags, a place name can be made up of multiple words, e.g. "Cardiff University" and in some cases the words are used in different order, e.g. "University of Cardiff". The traditional Levenshtein distance between these two names will be high and they will not be detected as similar. An improved version of the Levenshtein distance [12], that is based on the word level matching as opposed to character level matching, is used here and is defined as follows.

$$\sigma_t(n(r_1), n(r_2)) = 1 - \frac{LD(n(r_1), n(r_2))}{Max((n(r_1), n(r_2)))} \quad (1)$$

where σ_t is the text similarity to be calculated, n is the place name of the resource r_i, LD is the Levenshtein Distance function and Max is the maximum length of place names of the instances compared.

4.2 Tag Resolution Stage

The tag resolution stage involves a process of tag classification and filtering of tag collections. In particular, the process is guided by pre-defined assumptions of possible semantics associated with the resources. Hence, the tag resolution stage involves first, identifying and collecting place type and place activity reference data sets and using those as bases for matching and classification of the tag collection.

4.2.1 Building Reference Data Sets

A place type is a basic concept used for classification purposes in any place gazetteer. Here, two different sources are used for collecting place type information, 1) an official data source, produced by the Ordnance Survey (OS), the national mapping agency of Great Britain, and b) the GeoNames web gazetteer. The OS Buildings and Places ontology (OSBP) that is used to describe building features and place types surveyed with the intention of improving use and enabling semi-automatic processing of this data. OSBP provides over 200 place types such as: (University, Hotel, Market and Stadium). Geonames also have a place ontology that associates places with a hierarchy of place type represented as feature codes. Geonames provides over 600 unique feature codes such as: (Store, School and University). Identifying possible human activities associated with a place is a not a simple task. Some research work has addressed this issue previously [4], where an approach was shown to automatically extract possible types of services and activities from definitions of place types. Here, two resources are also used for identifying possible human activities that can be associated with geographic places: a) the OSBP ontology includes a property *os:purpose* that are defined by experts to represent the possible service(s) associated with the place types, and b) the OpenCyc ontology[20], an open source version of the Cyc project that assembles a comprehensive ontology of everyday common sense knowledge. Each place type in the OSBP ontology is attached with one or more *purpose*. Table 3 shows example records of the place type and purpose associations. The OpenCyc ontology contains human activity concepts and offers a classification of different possible activities as follows:
(*cyc:HumanActivity, cyc:CommercialActivity, cyc:OutdoorActivity, cyc:RecreationalActivity, cyc:CulturalActivity*). Figure 3 shows a sample of the SPARQL queries used to retrieve the activity types from both ontologies. Approximately 400 distinct activities are retrieved from both ontologies. Examples of the extracted place activities are: Boating, Eating, Fishing, Traveling, Working and Walking.

Place Type	**Purpose(s)**
University	Education
Hotel	Accommodation
Market	Trading
Stadium	Racing, Playing

Table 3 Example place types and corresponding purposes from OSBP

[20] http://www.opencyc.org/

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX os: <http://www.ordnancesurvey.co.uk/ontology/
BuildingsAndPlaces/v1.1/BuildingsAndPlaces.owl#>
PREFIX cyc: <http://sw.opencyc.org/2010/08/15/concept/en/>

SELECT ?placeActivity WHERE {
{ ?placeActivity rdfs:subClassOf os:Purpose. }
UNION
{ ?placeActivity rdfs:subClassOf cyc:HumanActivity.}
UNION
{ ?placeActivity rdfs:subClassOf cyc:CommercialActivity. }
UNION
{ ?placeActivity rdfs:subClassOf cyc:OutdoorActivity. }
UNION
{ ?placeActivity rdfs:subClassOf cyc:RecreationalActivity. }
UNION
{ ?placeActivity rdfs:subClassOf cyc:CulturalActivity. }
}
```

Fig. 3 SPARQL query for retrieving place activities from OSBP ontology.

4.2.2 Matching Tags

To match the tags in the folksonomy to the extracted lists of place types and place activities, the lists are first prepared as follows. Types of activities composed of multiple words are concatenated and added to the list. For example, the place type “Coffee Shop” is transformed to “CoffeeShop”. Matching is carried out on stemmed tags against the list of stemmed types and activities, using the Porters stemming algorithm[30]. The corresponding type or activity or both are then added to the ontology. For example, a tag “shop” can match a place type “shop” and a place activity “shopping” and hence both instances are created in the corresponding type and activity ontologies. The matching process resulted in 325 place type instances and 161 place activity instances.

4.3 Semantics Association and Ontology Building Stage

In this stage, the identified tag collections are structured in two steps. Firstly, subsumption relationships within individual tag collections of place types and activities are extracted and used to populate their respective sub-ontologies, and secondly, inter-relationships between types and activities are derived using the folksonomy structure. The place ontology is then populated with the resources and their associated tags from both the type and activity ontologies. Thus, the resulting place ontology reflects the associations between tags, resources and users in the folksonomy. The final step in this stage is enriching the place instances with the user sentiments.

4.3.1 Inferring Subsumption Relationships

This process infers the subclass hierarchical relationships in place type ontology instances and in place activity ontology instances represented by the properties *po:subPlaceTypeOf* and *po:SubPlaceActivityOf*. A probabilistic model of subsumption, originally introduced by Sanderson and Croft [33], can be used to derive concept hierarchies from text documents where for any given concepts/tags x and y: x subsumes y if

$$P(x|y) \geq 0.8 \; and \; P(y|x) < 1 \tag{2}$$

In other words x subsumes y if all the documents which contain y is a subset of the documents that contain x.

This model was extended for folksonomies [34] by including users and resources in the subsumption equation as follows. x subsumes y if

$$\begin{aligned} P(x|y) >= t \; and \; P(y|x) < t, \\ R_x \geq R_{min} \; , \; R_y \geq R_{min} \\ U_x \geq U_{min} \; , \; U_y \geq U_{min} \end{aligned} \tag{3}$$

Where t is the co-occurrence threshold, R_x is the number of resources tagged using x, and U_x is the number of users that use the tag x. In [34], it was proposed to set R_{min} to a value between 5 and 40, U_{min} to a value between 5 and 20, and the threshold t to 0.8, similar to the value determined empirically in [33]. The model was applied on the identified type and activity collections, resulting in the creation of 162 subsumption relationships, of which 143 are between place types and 19 are between place activities.

4.3.2 Inferring Inter-Ontology Relationships

Relating two tags in a folksonomy can be achieved by measuring the similarity between them in the sense that the higher the similarity value between two tags, the more related they are. Cosine similarity is used o measure the similarity between tags based on their co-occurrence with users and resources in the folksonomy [24] as follows.

$$\sigma(t_1, t_2) = \frac{|R_1 \cap R_2|}{\sqrt{|R_1| \cdot |R_2|}} \tag{4}$$

Where t_i represents a tag and R_i represents the resources associated with the tag t_i in the folksonomy.
A *po:relatedPlaceType* relation is created in the place ontology between a place activity instance and a place type instance if the Cosine similarity between their corresponding tags was found to be equal or above 0.8, a threshold found empirically to be sufficient in this work. A total of 393 relationships are created, linking instances between the place type and the place activity sub-ontologies.

The process of building the place ontology involves linking the results from all the previous sub-processes and populating a place ontology with the identified semantics. A place instance of type (*po:Place*) is created for every place cluster in the restructured folksonomy and its properties are populated as follows.

– **po:hasName**: is the most commonly used place name among the folksonomy place resources in the cluster.
– **po:alternateName**: each distinct name of the folksonomy place resources in the cluster other than the most commonly used name is represented by this property.
– **po:description**: is a concatenation of the comments attached to folksonomy place resources in the cluster.
– **wgs84:long** and **wgs84:lat**: is calculated by finding the centre location of the folksonomy place resources represented by the cluster.

4.3.3 Associating User Sentiments

Folksonomy tags can reflect the opinions of users about places. The aim of sentiment analysis in this step is to calculate the sentiment score for each place resource in the folksonomy. The sentiment score for a place resource measures the positive, negative or neutral users' opinions about this place. Sentiment analysis has been used in similar research works to capture users' opinions from the interaction and collaboration activities on Web 2.0. Research works on microblogs [28], more specifically Twitter, target the problem of capturing users' opinions from posts of similar structure. In contrast to previous work, the sentiment analysis method developed here considers the influence of users and their tagging behaviour in the equations as described below.

A semantic classifier based on the Naïve Bayes classifier is used here. It assumes conditional independence among features (tags in this context), which is fitting with the nature of folksonomies. Unlike other classifiers (such as Support Vector Machines), it requires a small amount of training data. The classifier is based on Bayes' theorem as follows:

$$P(S|T_1, ...T_n) = P(S)\prod_{i=1}^{n} P(T_i|S) \quad (5)$$

where S is a sentiment, T_i is a tag and n is number of tags associated with the place resource. Assuming an equal probability of positive, negative and neutral opinions, the equation can be simplified as follows:

$$P(S|T_1, ...T_n) = \prod_{i=1}^{n} P(T_i|S) \quad (6)$$

The output of the classifier depends on the way the features are selected. Here, a simple class feature model is used. However, considering different feature models such as N-Grams can be tested in the future. The data used to train the classifier is the AFINN wordlist [14, 27] which contains 2477 words and phrases with valence between -5 and +5. Sentiment classes are defined as follows; a positive class includes words with valence between +5 and +1, a neutral class with valence of 0 and a negative class with valence between -1 and -5.

After training the classifier, the following algorithm is applied to calculate the sentiment score for place clusters using the tags assigned to each place cluster.

```
places ← GetPlaceResources()
for p_i in places do
    users ← GetUsersOfPlace(p_i)
    usersCnt ← 0
    SntScore ← 0
    for u_i in users do
        usersCnt ← usersCnt + 1
        tagSet ← GetTagSet(p_i, u_i)
        SntScore += GetSntScore(tagSet)
    end for
    SntScore ← SntScore/usersCnt
    SaveSntScore(p_i, SntScore)
end for
```

The algorithm starts by retrieving all the place resources in the dataset and finding the associated users for each place resource. For each place-user pair the associated tags are retrieved and stored in the *tagSet*. The *tagSet* is used to calculate the sentiment scores for each place-user pair using the trained classifier, and then the average score is assigned to the place resource to neutralise the influence of individual user's scores. The sentiment score is a real value representing the overall users' sentiment about a place. The value ranges from -1 to +1. Where -1 indicates that all the tags attached to a place are classified as negative sentiments, while +1 indicates that all the tags attached to a place are classified as positive sentiments. The sentiment score is the sum of the classifier output averaged by the number of users who annotated a given place. For example, a sentiment score with value 0.8 indicates a strong positive sentiment value while the value -0.2 indicates a weak negative sentiment value. An evaluation of the sentiment analysis process is presented in the following section.

5 Results and Evaluation

The folksonomy dataset collected using the developed crawler contains 22,126 place instances in the UK and USA, 2,930 users and 12,808 distinct tags. The total number of folksonomy records is 68,437. A total of 10,119 unique WOEIDs were derived for place resources in the folksonomies. The text similarity is calculated between all place resources in each spatial cluster, all place resources having text similarity less than 80%, empirically found to be sufficient for the purpose of the present study, are filtered out from the cluster. The data cleaning stage resulted in identifying 19,614 clusters and corresponding unique places resources. Approximately, 11% (2,512) of the total number of place resources were merged.

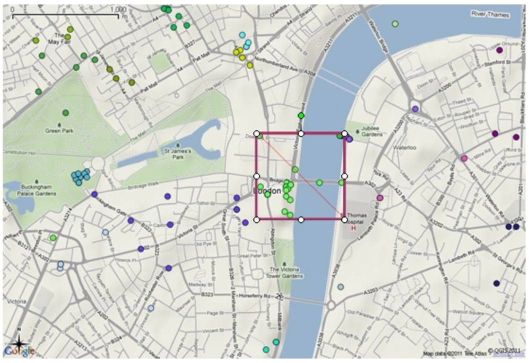


Fig. 4 Spatial clustering of place resources using WOEID.

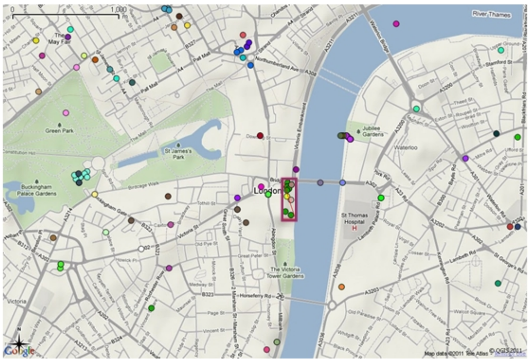


Fig. 5 Modified place clusters after applying the textual similarity method.

Figures 4 and 5 show a map of the area around the *Big Ben* tower in London with the place resources in Table 2. In Figure 4, the place resources are colour-coded according to the identified WOEID spatial clusters. In Figure 5, the same resources are shown in different clusters after applying the textual clustering method. The bounding box in Figure 4 has a diagonal of about 750 meters and includes all resources with same WOEID. The box in Figure 5 contains only those resources that refer to the place *Big Ben* and spans an area of approximately a 1/3 of the first box.

Figure 6 shows the results of classifying the tags using the proposed framework. 32% of the tags are place names. 18% of the tags were classified as user's opinions and are processed by the sentiment analysis process. 2% of the tags correspond to place types and 3% correspond to place activities. The rest of the tags (45%) do not fit in any of the above categories.

The distribution of the tags in the geo-folksonomy dataset follow a power law distribution. The frequency of tag usage is shown in Figure 7, where it is can be seen that more than 85% of the tags are used less than 5 times. This is similar to

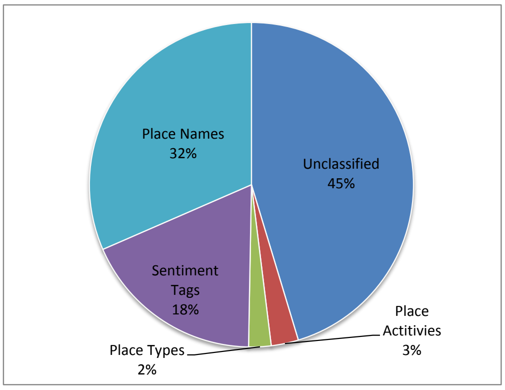


Fig. 6 Tag classification chart.

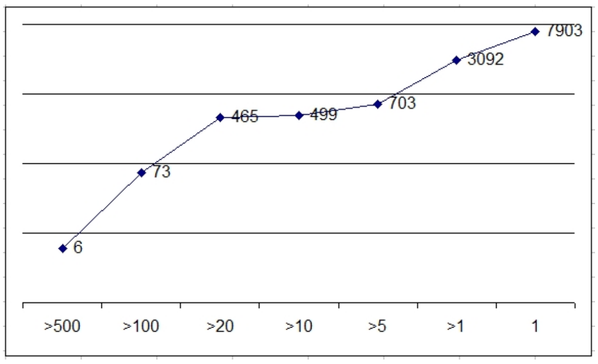


Fig. 7 Frequency of tag usage grouped on a log scale over the entire geo-folksonomy data set

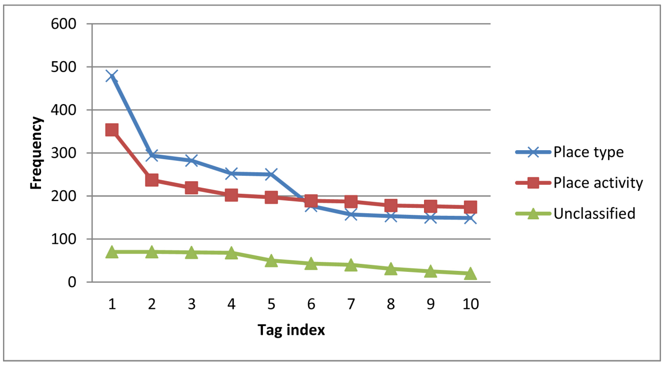


Fig. 8 Detailed tag usage frequency of the 10 most used tags.

the results reported by other empirical studies [8]. It is noted that although the percentages of place type and activity tags are low, these tags are used more frequently than unclassified tags as shown in Figure 8, which plots the frequency distribution of the 10 most used tags in each category. Table 4 lists the top 10 frequently used tags in each category. 79% of the unclassified tags contribute to the long tail of the Zipf frequency graph as they were found to be used only once or twice. The unclassified tags include possible reference to temporal concepts, such as 2008 and *summer*,

Rank	Place Type	Place Activity	Unclassified
1	food	housing	north
2	restaurant	travelling	clock
3	school	marketing	new
4	store	sale	1
5	hotel	visiting	family
6	university	servicing	TimeForPublicSpace
7	park	camping	apple_store
8	airport	socializing	high
9	museum	buying	2008
10	shop	business	recitation

Table 4 A sample of frequently used tags.

possible abbreviations (e.g. st. for street), or noise (e.g. two letter words: nv, vc, xy). The tag resolution stage resulted in identifying 346 activity types in the folksonomy, using a set of approximately 400 activity types in the reference data sets. It is interesting to observe that although 927 tags are identified as verbs using WordNet, only 107 of those corresponded to possible activity types from the compiled list using the external ontology resources. Some examples of the unclassified verb tags include, *arm*, *arrest*, *assign*, *back* and *coin*.

Figure 9 shows a subset of the derived place semantics, in which 24 place types and 16 place activities are presented and their corresponding association and subsumption relationships.

5.1 Evaluating The Data Preparation Process

To evaluate the effectiveness of the proposed tag cleaning and place clustering stage, the information gain is calculated for the geo-folksonomy before and after using the proposed methods. Shannon's information gain [36] is used to measure the uncertainty in the folksonomy structure as follows:

$$I(t) = -\sum_{i=1}^{m} \log_2 p(x_i) \tag{7}$$

Where t is any given tag. m is the number of places annotated by the tag t and $p(x_i)$ is defined as:

$$p(x_i) = \frac{w_{t,x}}{\sum_{j=1}^{m} w_{t,x_j}} \tag{8}$$

where $w_{t,x}$ is the weight of the link between t and place x. The value of $p(x_i)$ increases if the number of user assigning tag t to place x increases and vice versa. High values of $p(x_i)$ indicate a high degree of certainty (lower information gain) of using tag t with place x_i.
$I(t)$ was calculated to be 4011.54 before the clustering stage and 3442.716 after the clustering stage; a reduction of approximately 14%.

The reduction in uncertainty is caused primarily by the regions that have increased place annotation activities where it is likely for multiple users to annotate the same place using similar names. Table 5 shows a sample WOEID regions, the number of places in each region and the information content before and after using the proposed method.

WOEID	Instances	(I) Before	(I) After	% Reduction
2441564	106	126	115	8.7%
2491521	86	11.7	6.9	41%
2352127	83	129	119	7.8%
2377112	80	23.6	18.8	20.3%
2480201	68	24.6	21.6	12.2%

Table 5 Information content (Uncertainty) sample.

5.2 User-Based Ontology Evaluation

A possible approach to ontology evaluation is to compare it to a "golden standard" which itself can be an ontology. The OS Building and Place ontology is used here for demonstration. Figure 10 compares the semantics related to the place type "Tourism Attraction" as defined in the OSBP ontology to those related to the place Type "Tourism" in the derived place ontology. As can be seen in the Figure, only one "purpose" (Entertainment) is associated with the "Tourism Attraction" place type in the OSBP ontology, whereas a much richer set of relationships is identified in the place ontology, reflecting the usage of the concept in the specific folksonomy dataset ("Tourism" is related to 6 other place types and 4 place activities). However, it should be noted that an absolute comparison is not realistic as both ontologies serve different purposes and, as suggested previously, the ontology derived from the folksonomy is dynamic and its structure is likely to change with time.

To further evaluate the derived ontology, a questionnaire was designed to assess the quality of the derived concepts and their relationships. Five different places in London, UK, corresponding to different possible place types, were chosen, namely, Hyde Park, Marriot Hotel, Tesco, Wagamama and the Imperial War Museum. The geographic region was chosen primarily, because of popularity and as such more users were likely to be aware of the place names and secondly because of the density of the associated tags in the folksonomy. The questionnaire was issued to university students over a period of 4 weeks. 53 students participated in the survey, of which 76% were male users, approximately 90% were under 29 years old, 96% of users have a degree above high school, 65.9% were familiar with London and 80.4% were native English speakers.

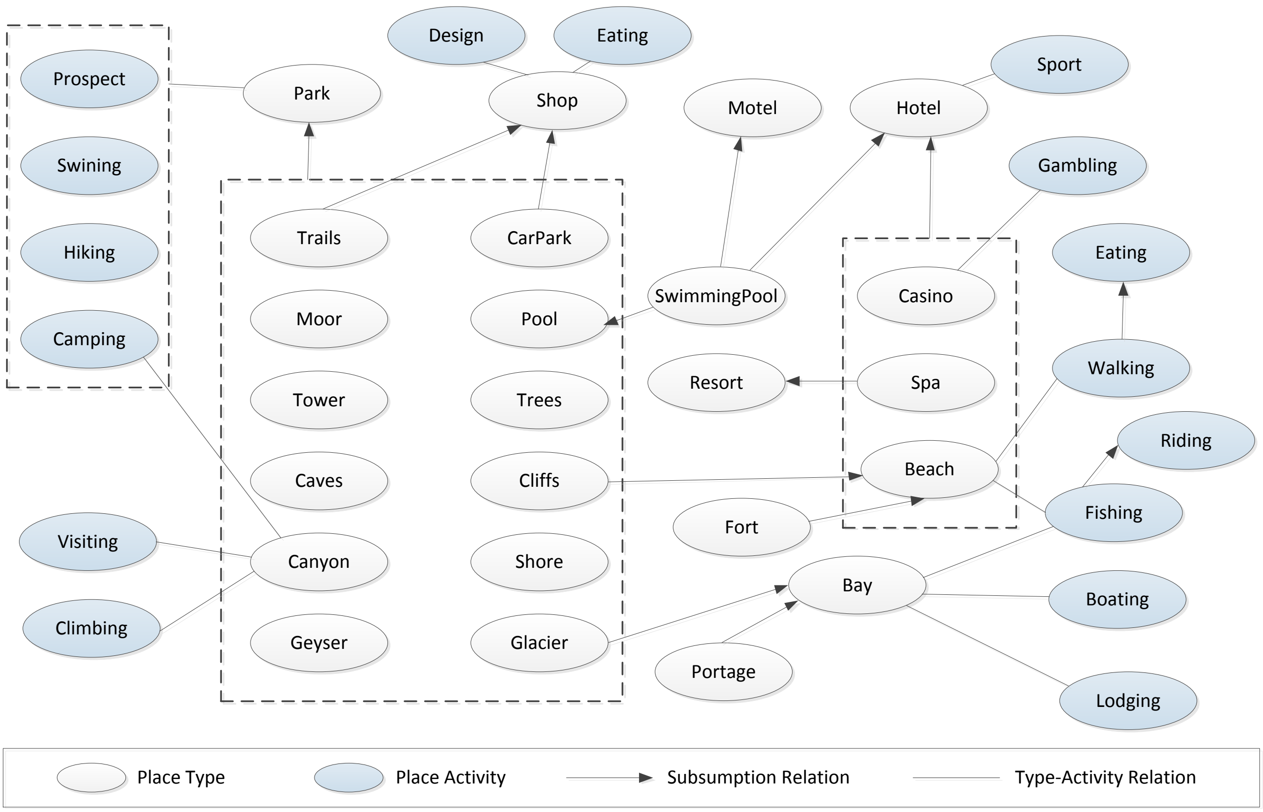


Fig. 9 A snapshot of the derived ontology showing a number of place types, their related place activities and subsumption relationships.

Two types of questions were asked for each place. The first type of questions aimed at evaluating the quality of the relationships between concepts. Figure 11 shows the responses of participants on questions about place-type relationships. The second type of questions aimed at evaluating misclassified tags by asking the user to suggest a classification for tags co-occurring with the place resource, as either a place type, a place activity, a related concept or a non-related concept. Figure 12 shows the results of the second type of questions for the place "Hyde Park". Users' responses are used to calculate the recall, precision and F1 measure for evaluation. Table 6 lists the number of true positives, false positives, true negatives and false negatives used to calculate the precision (0.8), recall (0.5) and F1 (0.615). The experiment suggests a correlation between the derived ontology and users' perception of places and related semantics. Finally, the survey also questioned the users' experiences, or impressions (if they did not visit the places), with the five places. The responses again correlated with the output of the sentiment classifier. Though the experiment is limited, the results are promising and indicative of the validity of the methods. However, a larger experiment can be pursued in the future.

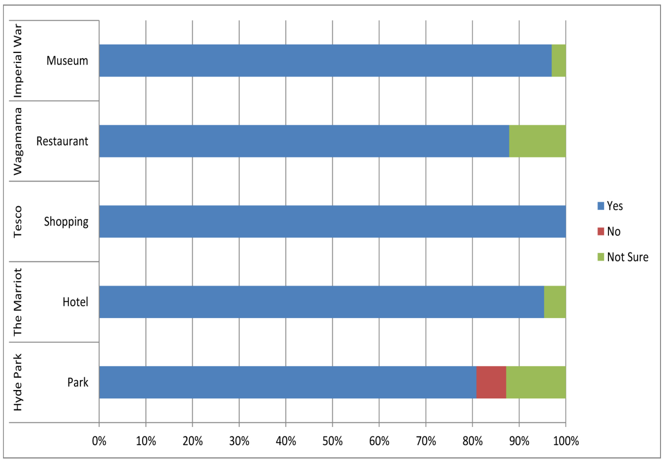


Fig. 11 Level of agreement in the questionnaire with the derived relationships between concepts for the chosen place resources.

5.3 Quantitative Ontology Evaluation Using Semantic Similarity

A quantitative evaluation experiment was designed here to measure the level of agreement between the semantics represented by the place type and place activity sub-ontologies on one side and the general semantics on the web on the other

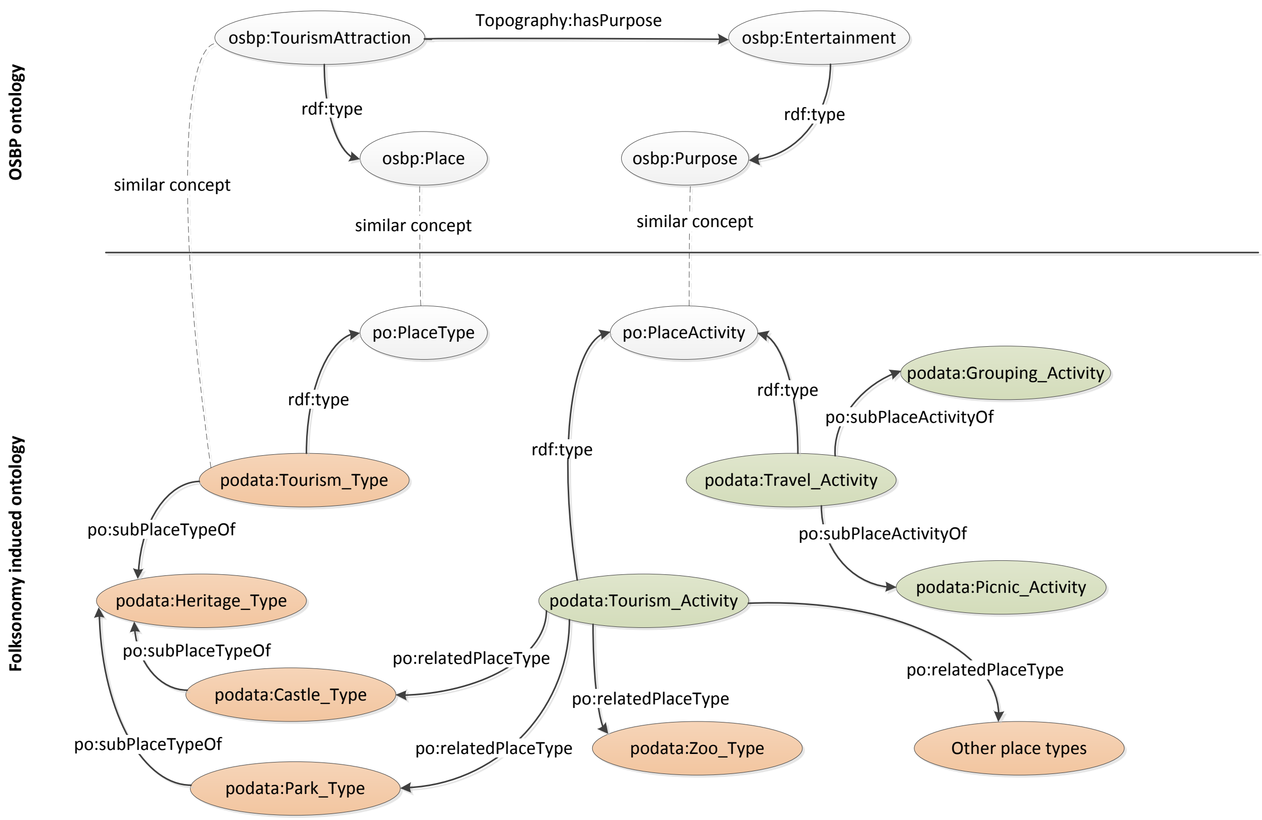


Fig. 10 An example of a place type concept "Tourism" as defined in the Ordnance Survey ontology and its computed definition in the derived place ontology.

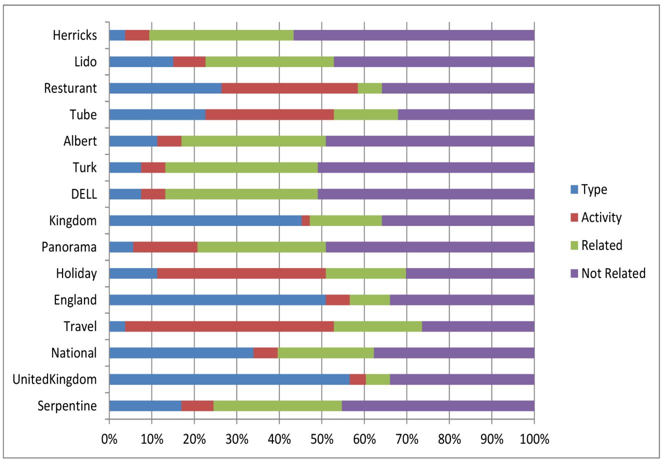


Fig. 12 A sample of the users' responses classifying tags co-occurring with the place "Hyde Park".

Place	TP	FP	TN	FN
Hyde Park	4	2	3	12
Marriot	4	0	10	5
Tesco	4	1	12	3
Wagamama	4	2	12	0
Imperial War	4	0	15	0
Total	20	5	52	20

Table 6 Evaluating the tag classification results with the questionnaire responses.

side. The Measure of Semantic Relatedness (MSR) web service [41] provides a set of methods through web-based API interface to calculate the semantic relatedness between two terms[21]. Although the MSR provides different methods of calculating the semantic relatedness, all of them are based on the same theory. The MSR assumes that the strength of the relation between two terms is proportional to the number of times the two terms co-occurred together in the same documents on the web. MSR does not employ any semantic analysis approaches and is based only on co-occurrence of the terms. It assumes that the existence of two terms in the same document implies they are in the same context. Hence, the more frequently they appear together, the more semantically related they are. The performance of the different MSR methods in terms of quality and accuracy is found to be dependent on the size and type of the input data [23]. More details and comparisons about the different MSR methods can

[21] `http://cwl-projects.cogsci.rpi.edu/msr/`

be found in [11]. In this experiment, the Point-wise Mutual Information (PMI) [39] and the Normalised Search Similarity (NSS) [25] methods are chosen to evaluate the quality of the derived tag relationships. Both methods can measure the semantic relatedness among terms in large datasets.

Place type relationships as well as place activity relationships are evaluated using both the PMI and the NSS methods. First, a set of SPARQL queries are used to retrieve the relations along with the concepts they connect. The appropriate MSR API functions are passed the two concepts of each relation to calculate the semantic similarity between them using the Google's search engine. The PMI and NSS values are computed for 500 relationships. Figure 13 shows a graph of the output of both measures along with their corresponding trend lines. Both measures show a high degree of relatedness between the identified tag relationships with average values of 0.86 for PMI (and standard deviation of 0.16) and 0.77 for NSS (and standard deviation of 0.1). The figure also shows the corresponding trend lines of both measures.

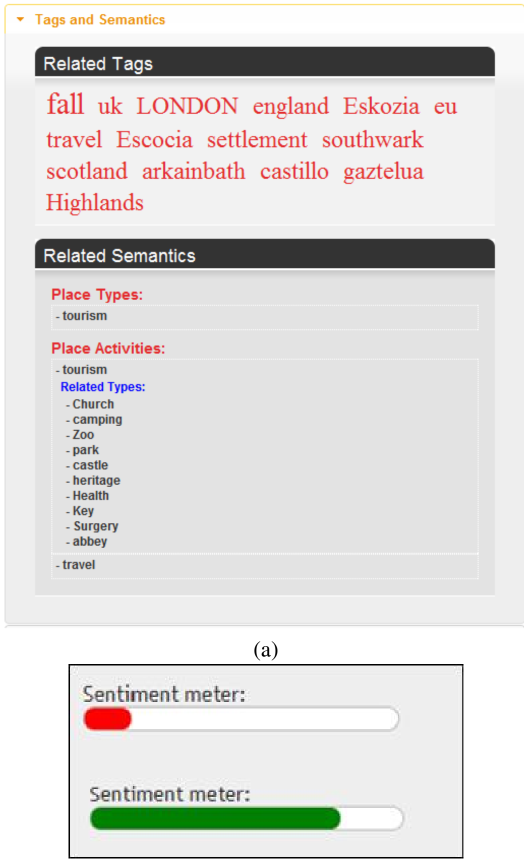
Table 7 illustrates the results of the experiment by showing a sample of the measures of PMI-G and NSS-G for 10 relationships. The experiment demonstrates the likelihood of the validity of the place semantics automatically extracted from the geo-folksonomies; i.e. that the extracted semantics are found to be similar to those expressed in general web documents.

Concept 1	Concept 2	PMI-G	NSS-G
Sale(A)	Flat(T)	69%	90%
Buy(A)	Sale(A)	100%	83%
Hotel(T)	Reservation(A)	97%	79%
University(T)	College(T)	100%	89%
Spa(T)	Hotel(T)	96%	91%
Boating(A)	Fishing(A)	100%	78%
Rock(T)	Climbing(A)	63%	65%
Casino(T)	Gambling(A)	93%	76%
Museum(T)	Park(T)	75%	80%
Rock(T)	Mountain(T)	86%	82%

Table 7 A sample of the MSR measures calculated using PMI-G and NSS-G applied on the ontology relations between places types (T) and activities (A)

6 The SemTag Application

To demonstrate the utility of the proposed framework, an application, called SemTag, was developed to display the derived place semantics. For comparison, these were displayed alongside the tag cloud for any given place resource. A tag cloud is used on social applications to display the most popular tags associated with a resource, regardless of how they are semantically related to that resource.



(a)

(b)

Fig. 14 a) Screenshot of the SemTag application showing the derived place semantics for the place "London Eye". (b) A meter gadget displaying the sentiment score for place instances.

The screen shot in Figure 14 shows part of the user interface displaying the tag cloud and the derived place types and activities for the place "London Eye". Note how the place type "tourism" and the activity "travel" are identified with this point of interest, but are not included in the tag cloud.

A sentiment meter gadget is also implemented and presented on the interface to visualise the sentiment score of a place, as shown in Figure 14. The gadget is a 'progress bar'-like component where colour is used to distinguish the score level; a red colour for a low sentiment score and a green colour for a high sentiment score.

The application demonstrates the possible utility of the proposed framework, where it can be envisaged that the derived place semantics may be used to refine search queries and combined with the sentiment score be used to rank the retrieved search results.

7 Conclusions and Future Work

Users' interactions and collaborations on Web 2.0 mapping applications generate geo-folksonomies. Geographic places

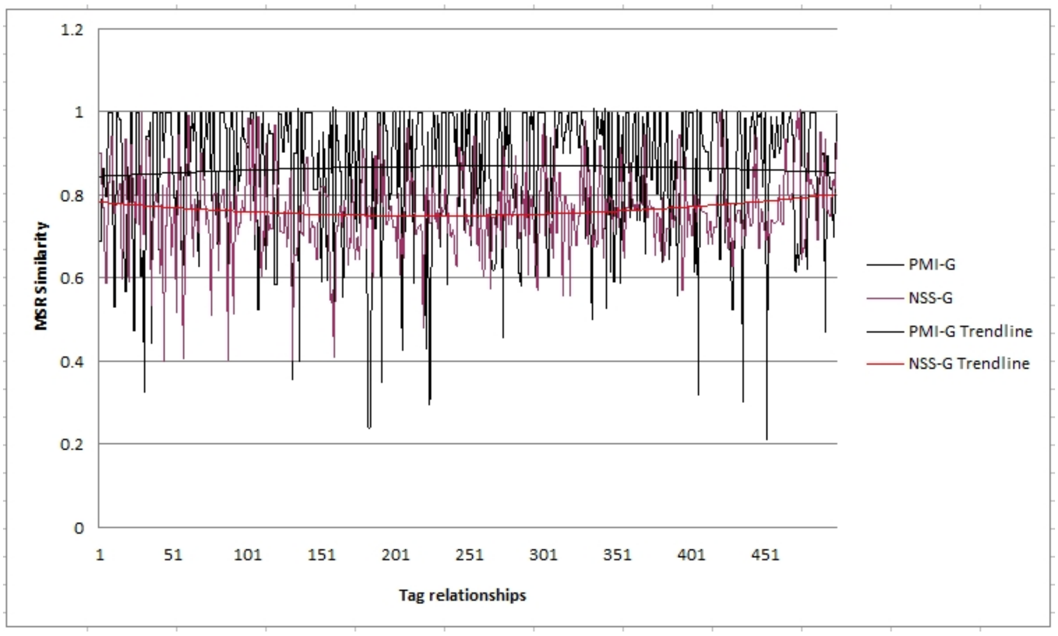


Fig. 13 Results of the PMI-G and the NSS-G semantic relatedness measures for a set of 500 derived place ontology relationships and the corresponding trend lines.

are annotated with different kinds of place semantics, including, vernacular place names, place types and activities people participate in, events, as well as personal opinions. Much interest has emerged in the geographic information retrieval community in the creation and population of place name resources to facilitate and enhance the search and retrieval of geographically-referenced information. These works focus primarily on finding place names and geographic locations of place instances. Geo-folksonomies embed rich user-oriented place semantics, which if discovered, can potentially lead to much richer place knowledge resources and more personalized search and retrieval of web information content.

The work in this paper combines and extends research works in the general area of folksonomy analysis and the area of discovering place semantics from web resources. A model of place is utilised that captures, in addition to basic spatial representation of location, the notion of place affordance and allows for the representation of possible association of a place resource and multiple place types, place activities and inter-relationships between types and activities. The model is used as a base for a framework for discovering place-related semantics from geo-folksonomies. Existing ontological resources were used in a tag resolution stage for matching and identification of place type and activity concepts. A process of semantic association with the filtered tags was then designed to extract relationships between their corresponding concepts and to build representative place ontologies. Subsumption models, folksonomy analysis and tag similarity methods were used to guide this process, resulting in the extraction of a significant number of different types of relationships between place types and place activities and their inter-relationships. The resulting place ontology thus associates specific place instances with possibly multiple place types and place activities, directly associated or inferred as a consequence of derived relationships. The resulting ontology represents the “wisdom of the crowd” of users in the folksonomy and is shown to reflect a much richer structure of concepts and relationships than those defined in a formal data source produced by experts. A limited user experiment confirms the validity of the results.

The overarching goal of this work is to build dynamic user-generated place gazetteers that can be used to resolve geographic place concepts in search engines and question-answering systems. The main contribution of this paper is the proposal of framework and demonstration of how this goal can be achieved. However, much more work still needs to be done. In particular, some possible extensions of the work include: a more detailed study of the unclassified tags in the folksonomy to identify more useful concepts, employing more ontological resources, for example, ConceptNet [22] to resolve tags, extension of the place model to include the time dimension to reflect the dynamic nature of the evolution of the folksonomy structure and further evaluation of the resulting semantics and their utilisation in useful application on the semantic and social web.

[22] http://conceptnet5.media.mit.edu/

References

1. Abdelmoty, A., Smart, P., El-Geresy, B., Jones, C.: Supporting frameworks for the geospatial semantic web. In: SSTD '09: Proceedings of the 11th International Symposium on Advances in Spatial and Temporal Databases, vol. LNCS 5644, pp. 335–372 (2009)
2. Abdelmoty, A., Smart, P., Jones, C.: Building place ontologies for the semantic web:: issues and approaches. In: Proceedings of the 4th ACM workshop on Geographical information retrieval, pp. 7–12. ACM (2007)
3. Adrian, B., Sauermann, L., T., R.B.: Contag: A semantic tag recommendation system. In: Proceedings of the 3rd International Conference on Semantic Technologies, I-Semantics, pp. 297–304 (2007)
4. Alazzawi, A.N., Abdelmoty, A.I., Jones, C.B.: What can i do there? towards the automatic discovery of place-related services and activities. International Journal of Geographical Information Science **26**(2), 345–364 (2012)
5. Ballatore, A., Bertolotto, M.: Semantically enriching vgi in support of implicit feedback analysis. In: International Symposium on Web and Wireless GIS, W2GIS, pp. 78–93 (2011)
6. Becker, C., Bizer, C.: Exploring the geospatial semantic web with dbpedia mobile. Web Semantics: Science, Services and Agents on the World Wide Web **7**(4), 278–286 (2009)
7. Câmara, G., Miguel, A., Monteiro, V., Paiva, A., Cartaxo, R., Souza, M.D.: Action-driven ontologies of the geographical space: Beyond the field-object debate. In: Proceedings 1st International Conference on Geographical Information Science, GIScience, pp. 52–54 (2000)
8. Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V., Loreto, V., Hotho, A., Grahl, M., Stumme, G.: Network properties of folksonomies. AI Communications **20**(4), 245–262 (2007)
9. Chen, W., Cai, Y., Leung, H., Li, Q.: Generating ontologies with basic level concepts from folksonomies. Procedia Computer Science **1**(1), 573 – 581 (2010)
10. De Smith, M., Goodchild, M., Longley, P.: Geospatial Analysis, A Comprehensive Guide to Principles, Techniques and Software Tools. Metador (2007)
11. Emadzadeh, E., Nikfarjam, A., Muthaiyah, S.: A comparative study on measure of semantic relatedness function. In: The 2nd International Conference on Computer and Automatiom Engineering, vol. 1, pp. 94–97 (2010)
12. French, J., Powell, A., Schulman, E.: Applications of approximate word matching in information retrieval. In: CIKM, pp. 9–15 (1997)
13. Fu, G., Jones, C., Abdelmoty, A.: Ontology-based spatial query expansion in information retrieval. In: International Conference on Ontologies, Databases and Applications of Semantics, ODBASE, pp. 1466–1482. Springer (2005)
14. Hansen, L., Arvidsson, A., Nielsen, F.Å., Colleoni, E., Etter, M.: Good friends, bad news - affect and virality in twitter. In: J. Park, L. Yang, C. Lee (eds.) Future Information Technology, Communication in Computer and Information Scence, vol. 185, pp. 34–43. Springer (2011)
15. Hart, G., Temple, S., Mizen, H.: Tales of the river bank: first thoughts in the development of a topographic ontology. In: F. Toppen, P. Prastacos (eds.) Proceedings of the 7th AGILE Conference, pp. 165–168. Crete University Press, Heraklion (2004)
16. Heyer, L., Kruglyak, S., Yooseph, S.: Exploring expression data: identification and analysis of coexpressed genes. Genome research **9**(11), 1106 (1999)
17. Hotho, A., Jaschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Proceedings of the 3rd European Semantic Web Conference on The Semantic Web, ESWC'06, vol. LNCS 4011, pp. 411–426 (2006)
18. Jones, C., Abdelmoty, A., Finch, D., Fu, G., Vaid, S.: The spirit spatial search engine:architecture, ontologies and spatial indexing. In: Proceedings 3rd International Conference on Geographical Information Science, GIScience'04, vol. LNCS 3234, pp. 125–139 (2004)
19. Keßler, C., Maué, P., Heuer, J., Bartoschek, T.: Bottom-up gazetteers: Learning from the implicit semantics of geotags. GeoSpatial Semantics pp. 83–102 (2009)
20. Kuhn, W.: Ontologies in Support of Activities in Geographical Space. International Journal of Geographical Information Science **15**(7), 613–631 (2001)
21. Levenshtein, V.: Binary codes capable of correcting deletions, insertions. and reversals. Soviet Physics Doklady **10**, 707–710 (1966)
22. Lin, H., Davis, J., Zhou, Y.: An integrated apporoach to extracting ontological structures from folksonomies. In: L. Arayo (ed.) Proceedings of the 6th European Semantic Web Conference on The Semantic Web, ESWC'09, vol. LNCS 5554, pp. 654–668. Springer (2009)
23. Lindsey, R., Veksler, V., Grintsvayg, A., Gray, W.: Be wary of what your computer reads: the effects of corpus selection on measuring semantic relatedness. In: 8th International Conference of Cognitive Modeling, ICCM (2007)
24. Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G.: Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the 18th international conference on World wide web, pp. 641–650 (2009)
25. Matveeva, I.: Generalized latent semantic analysis for document representation. ProQuest (2008)

26. Mika, P.: Ontologies are us: A unified model of social networks and semantics. Web Semantics: Science, Services and Agents on the World Wide Web **5** (2007)
27. Nielsen, F.Å.: Afinn (2011). URL http://www2.imm.dtu.dk/pubdb/p.php?6010
28. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). ELRA (2010)
29. Perry, M., Hakimpour, F., Sheth, A.: Analyzing theme, space, and time: an ontology-based approach. In: Proceedings of the 14th annual ACMGIS, pp. 147–154. ACM (2006)
30. Porter, M.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)
31. Rattenbury, T., Naaman, M.: Methods for extracting place semantics from Flickr tags. ACM Transactions on the Web (TWEB) **3**(1), 1–30 (2009)
32. Relph, E.: Place and placelessness. Pion Ltd (1976)
33. Sanderson, M., Croft, B.: Deriving concept hierarchies from text. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 206–213. ACM (1999)
34. Schmitz, P.: Inducing ontology from flickr tags. In: In Proc. of the Collaborative Web Tagging Workshop (WWW Š06) (2006)
35. Sen, S.: Use of affordances in geospatial ontologies. In: Proceedings of the 2006 international conference on Towards affordance-based robot control, pp. 122–139. Springer Verlag (2008)
36. Shannon, C.: A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review **5**(1), 55 (2001)
37. Smart, P., Jones, C., Twaroch, F.: Multi-source toponym data integration and mediation for a meta-gazetteer service. In: Geographic Information Science, vol. LNCS 6292, pp. 234–248 (2010)
38. Tsui, E., Wang, W.M., Cheung, C.F., Lau, A.S.M.: A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags. Inf. Process. Manage. **46**(1), 44–57 (2010)
39. Turney, P.D.: Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In: EMCL '01: Proceedings of the 12th European Conference on Machine Learning, pp. 491–502. Springer-Verlag, London, UK (2001)
40. Van Damme, C., Hepp, M., Siorpaes, K.: Folksontology: An integrated approach for turning folksonomies into ontologies. In: SemNet, vol. 2, pp. 57–70 (2007)
41. Veksler, V., Grintsvayg, A., Lindsey, R., Gray, W.: A proxy for all your semantic needs. In: 29th Annual Meeting of the Cognitive Science Society (2007)