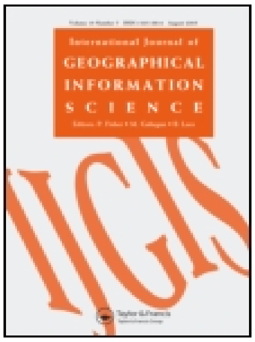


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Enriching user profiles using geo-social place semantics in geo-folksonomies

Ehab ElGindy & Alia Abdelmoty

International Journal of Geographical Information Science, 2014
Vol. 28, No. 7, 1439–1458, http://dx.doi.org/10.1080/13658816.2014.894194

Enriching user profiles using geo-social place semantics in geo-folksonomies

Ehab ElGindy* and Alia Abdelmoty

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

(*Received 30 January 2013; accepted 26 January 2014*)

Geo-folksonomies link social web users to geographic places through the tags users choose to label the places with. These tags can be a valuable source of information about the user's perception of place and can reflect their experiences and activities in the places they label. By analysing the associations between users, places and tags, an understanding of a place and its relationships with other places can be drawn. This place characterisation is unique, dynamic and reflects the perception of a particular user community that generated the geo-folksonomy. In this work, an approach is proposed to analysing geo-folksonomies that builds on and extends existing statistical methods by considering specific concepts of relevance to geographic place resources, namely, place types and place-related activities, and by building a place ontology to encode those concepts and relationships. The folksonomy analysis and evaluation are demonstrated using a realistic geo-folksonomy data set. The resulting ontology is used to build user profiles from the folksonomy. The derived profiles reflect the association between users and the specific places they tag as well as other places with relevant associated place type and activities. The methods proposed here provide the potential for many interesting and useful applications, including the harvesting of useful insight on geographic space and employing the derived user profiles to enhance the search experience and to identify similarities between users based on their association to geographic places.

Keywords: place ontology; geo-social web; user profile

1. Introduction

Collaborative tagging and social-bookmarking applications on the web allow users to tag objects with keywords to facilitate retrieval by users. Examples of some popular applications include Delicious, Flickr and Amazon. A simple form of shared vocabularies emerges in these applications and categories of tags used to characterise some resource by users are commonly referred to as 'folksonomies' (Golder and Huberman 2006). Recognising the value in this data, research work have recently been targeted at extracting and structuring embedded semantics in folksonomies (Heymann and Garcia-Molina 2006, Specia and Motta 2007, Chen *et al.* 2010) and utilising these in application of semantic tag recommendation systems (Adrian *et al.* 2007).

Geo-folksonomies are a special kind of folksonomies where people can create and tag geographic places on maps. Examples of such map creation and sharing applications include, Tagzania, Wikimapia, GeoNames and OpenStreetMap (OSM). The type of information people associate with geographic places will differ according to the purpose

*Corresponding author. Email: ehab.elgindy@cs.cardiff.ac.uk

of the application. While in applications such as OSM, users are driven by the purpose of creating maps, and thus mainly provide information on place names and place types; in other socially driven applications, such as Tagzania, there are no restrictions on the sort of information people associate with geographic places. Tags can thus reflect users' perception of the place, actual experiences and activities carried out in a place. Recently, works have addressed the problem of disambiguating concepts related to geographic classification in OSM (Ballatore and Bertolotto 2011, Mülligann *et al.* 2011, Ballatore *et al.* 2013), while interest in other socially driven location-sharing application has mostly revolved around place identification and their use as a possible resource for building web gazetteers (Keßler *et al.* 2009).

Geo-folksonomies encode relationships between users and the geographic places they label. Studying those tags can potentially provide an understanding of the characteristics of individual geographic places as perceived by users over time. In contrast to the information held in traditional gazetteers where geographic places are normally assigned generic categories of place types, geo-folksonomies can be used to build a different sort of gazetteer where categories of place types, services and activities are determined collaboratively by users. In this work, a new approach is proposed to the analysis of geo-folksonomies. The approach extends conventional statistical methods to analyse a folksonomy with the purpose of inferring place-related concepts, in particular place types and place-related human activities, and encoding these as an ontology that reflect the folksonomy structure and relationships.

The paper proposes a novel framework for analysing geo-folksonomies that is guided by place-related semantics. The approach involves several stages of folksonomy cleaning and preparation to address specific problems associated with noise and redundancy of place resources in the folksonomy. The identification and resolution of tags in the folksonomy is done by matching against a prepared reference data set of place types and activity information collected from existing ontological resources. The resolved tags are used to populate a place ontology and relationships between tags are recorded that map the structure of the underlying folksonomy. The induced 'folkontology' (Van Damme *et al.* 2007) is used to build user profiles and discover relationships between users of the folksonomy. The derived user profiles will suggest which place type and activity concepts users are associated with and the strength of their associations with these concepts.

The paper is structured as follows: a review of related work on extracting semantics from folksonomies, constructing user profiles in social-tagging applications as well as an overview of the notion of place semantics are described in Section 2. A model of place to be used as a basis for extracting information from geo-folksonomies is outlined in Section 3. The proposed approach for geo-folksonomy analysis and for building user profiles is described in Section 4. In Section 5, the data set used for evaluating the approach is described. Analysis of the derived place information is discussed and the possible utility of the generated user profiles is demonstrated.

2. Related work

2.1. Discovering semantics in folksonomies

Several approaches have been proposed in the literature for building taxonomies or thesauri of concepts from folksonomies. Mika (2007) used social-network analysis to extract relationships between the different entity types in a folksonomy. Other works focussed on analysing relationships between resources and tags only and ignored the user dimension (Heymann and Garcia-Molina 2006, Schmitz 2006, Specia and Motta 2007). Schmitz (2006) introduced a probabilistic model of subsumption, based originally on a subsumption model by Sanderson and Croft (1999), to model possible parent–child relationships between tags and resources inherent in the folksonomy structure. Markines *et al.* (2009) considered the user dimension by introducing a pre-processing (aggregation) step, where the folksonomy is transformed from a tripartite structure of users, tags and resources to a bipartite graph of tags and resources, and the users' relationships are modelled as weights on the edges of this graph. This was shown to enhance the accuracy of the induced relationships.

In some web 2.0 photo-sharing applications such as Flickr and Panoramio, users annotate their uploaded images with tags representing the place where the photos were taken. In such applications, a fair proportion of the tags refer to place names and hence they are a good source for the automatic building of gazetteers. In (Popescu *et al.* 2008) simple text, analysis approaches were used to identify place names and types, for example, nouns in the title of the photo. In other works, approaches were proposed for analysing geo-folksonomies to extract place-related events. For example, in Rattenbury *et al.* (2007) the feasibility of automatically extracting events and place semantics from Flickr tags was tested. Burst-analysis and scale-structure identification techniques were used to recognise the spatial and temporal tagging patterns of event and successful identification of place names from tags was demonstrated. Intagorn *et al.* (2010) proposed an approach for learning geospatial concepts and relationships from Flickr, where place names are first identified and then tags associated with the relevant photos are analysed using conventional folksonomy-analysis methods.

In the above works, place is used to reference the resource (e.g. photos) and thus analysis focussed on first identifying the place reference and then using it to classify and analyse the folksonomy structure associated with the resources. The difference in the work proposed in this paper is that the places themselves are the resources to be analysed and hence are prime components of the folksonomy structure. Here, an alternative place-focussed folksonomy analysis approach is proposed that tailors the conventional statistical methods to suit geo-folksonomy structures.

2.2. Constructing user profiles in social-tagging applications

Social tags can be used to build user profiles. Sen *et al.* (2009) argue that social-tagging activities can be considered as an implicit rating behaviour. In other words, social tags can represent the interests and express the preferences of individual users. A user profile built from folksonomies is denoted by the set of tags representing the user interests with corresponding weights. The weight of a tag in the user profile represents the strength of the relationship between the user and that tag. Weights can be simplified by using a binary weighting approach, such as in Bogers and Van den Bosch (2008), or they can be calculated using methods such as Term Frequency–Inverse Document Frequency (TF–IDF).

There are different approaches to building user profiles from social tags. For example, Tso-Sutter *et al.* (2008) proposed a user-profiling approach that relates users to tags by converting the three-dimensional folksonomy relations into an expanded user-tag rating matrix. Niwa *et al.* (2006) extended this approach and proposed a method of building clusters of tags that are highly related based on tag similarity, then the clusters are used to expand user profiles. In Au Yeung *et al.* (2009) a method, called 'personomy', is proposed in which a cluster of all popular tags of the resources annotated by a user is used to profile topics of interest to that user.

In this work, a new approach to building users profiles is proposed that utilises the derived place semantics embedded in the folksonomy. The methods proposed aim to infer, in addition to place instances directly annotated by users, place type concepts and human activity type concepts that the user may be associated with based on their tagging behaviour, as well as the behaviour of the user community in annotating the place resources.

2.3. Semantics of geographic places

One viewpoint of place is as a concept that relates geography to human existence, experiences and interaction (Relph 1976, Agnew 2011). 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 Open Geospatial Consortium (OGC) Reference Model (ORM)[1] provides a general-feature model designed to characterise geographic features, types and the relations between features. Recently, some efforts have targeted the identification and discovery of the spatial aspects of place definition from web resources, for example, possible vernacular place location and extension in space (Smart *et al.* 2010).

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 (1976), 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 (Kuhn 2001). The term 'action-driven ontologies' was 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 (Sen 2008). The assumption is that affordance-oriented place ontologies are needed to support the increasingly more complex applications requiring semantically richer conceptualisation of the environment.

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.

3. Place-related semantics in geo-folksonomies

Tags that people associate with geographic places on the geo-social web can be a valuable resource for discovering people's perception of a place, their experiences, activities and sentiments about the place. These perceptions are associated with particular place instances and may vary over time.

To encode the place-related concepts represented in a folksonomy, a model of place is adopted where a geographic place can be associated with possibly multiple place types and place activities. Two different types of semantic relationships are used in this model:

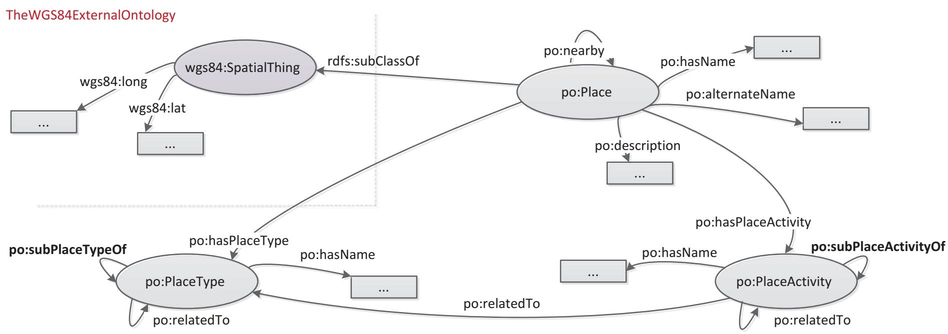


Figure 1. A model of place to represent place-related semantics in geo-folksonomies.

First, place types and place activities may themselves form individual subsumption hierarchies and second, association relationships, where a place type may be related to more than one other place type or activity concept (e.g. a place type 'school' may be related to activities such as 'learning' and 'teaching', etc.).

A distinguishing characteristic of 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 and nursery, from which it can derive activities such as learning and teaching, but it can also be associated with activities such as dancing, weight training and adult education, where it offers external services to the community after school hours. The former list is derived from the association with a particular place type, but the latter list may come from direct annotation by users in a geo-folksonomy.

The model is shown in Figure 1. Three types of entities are represented: *Place, Place Type* and *Place Activity* as well as properties and interrelationships between them. One possible representation of the spatial location is by extending the WGS84 *SpatialThing*[2] concept to inherit the spatial properties *lat, long*. This is sufficient to capture the representative point location of places in the data sets of interest to this work, but the model can be extended to allow for multiple spatial representations of geographic place. 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.

The model extends previous proposals, for example, that of the Ordnance Survey Building and Place (OSBP) ontology,[3] where a similar notion to place activity is explicitly modelled and associated with a place type through a defined relationship 'has-purpose'. The difference in the above model is the association of separate relationships between a place and place types and activities. Hence, a place may be associated with activities that are not derived from its relationship with a place type. In addition, interrelationships between place types and between place activities were not modelled in the OSBP ontology.

An ontology of place that captures the concepts and relationships in the model is implemented using the W3C Web Ontology Language (OWL). All classes and properties are qualified with the prefix **po.**[4] Note that, in general, the associations in this model are dynamic as a result of the accumulation of users' annotations. Hence, the relationships *po : hasPlaceType, po : hasPlaceActivity* and *po : relatedTo* would be time stamped. However, the time dimension is out of the scope of the current study and is the subject of future research.

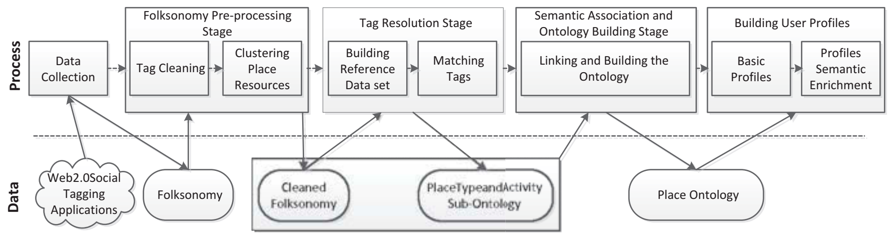


Figure 2. The process of building user profiles from geo-folksonomies.

4. Extracting user profiles from geo-folksonomies

The process of building user profiles is shown in Figure 2. Starting with a raw collected geo-folksonomy data set, the aim is to discover place-related semantics in the folksonomy based on the place model suggested above. An example of a geo-folksonomy data set collected for the purpose of this work is given in Section 5. The instantiated place model will then be used to create individual profiles for users who contributed the data in the folksonomy. User profiles will reflect the possible association of the user with individual place instances based on the inferred properties of those instances as identified from the folksonomy. The approach involves four main stages: a folksonomy pre-processing stage to filter out noise and handle specific problems associated with data input in geo-folksonomies, a tag-resolution stage where tags in the folksonomy are mapped to concepts of interest in the place model proposed, a semantics association and ontology building stage, where relationships between tags are identified and encoded in the model and finally a user profile creation stage using the model created. The different processes are described in more detail in the following sections.

4.1. Folksonomy pre-processing

A pre-processing stage of tag cleaning is needed before analysing folksonomy data. The flexibility of data input offered by folksonomy-generating applications, where no input-validation methods are used, leads to quality issues in the tags collected which need to be addressed. Basic issues may include tags with spelling mistakes, stop words and numbers. Hence, a first step in the cleaning process involves the identification, correction and filtering of noise data from the folksonomy (Van Damme *et al.* 2007, Plangprasopchok and Lerman 2009, Intagorn *et al.* 2010). A further step of linguistic analysis (lemmatisation) is also used to identify similar (as well as duplicate) tags expressed in different morphological forms, for example, the three tags: shop, shops and shopping will be identified as being similar. The folksonomy structure is updated to reflect the identified relationships between tags.

In the case of geo-folksonomies, a further complexity arises due to the possible redundancy in the creation of the place resources themselves. In particular, users are able to create duplicate place instances that essentially refer to the same geographic place on the ground. Figure 3 shows an example of this problem, where several instance of the same place, clustered in the highlighted box in the figure, were created separately by different users.

This problem again stems from the flexibility of data input offered in these applications, combined with the inability of users to recognise or digitise precise locations of place instances. This redundancy leads to fragmentation of the folksonomy structure and degradation of analysis results. An important pre-processing step with geo-folksonomies is

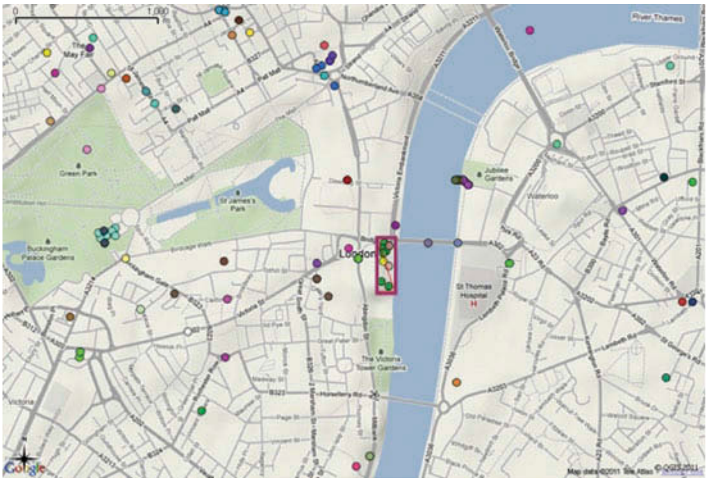


Figure 3. Cluster of place instances representing the clock tower of Big Ben, located in the Palace of Westminster in London, UK, created by users on Tagzania.

therefore the identification and clustering of duplicate place resources and the restructuring of the folksonomy accordingly. A two-step clustering process is used here as follows:

(1) First, a spatial clustering process is applied using a spatial similarity measure to group place resources based on their relative proximity. One possible approach used in this work is to initially group place instances with the same Yahoo Where on Earth ID (WOEID),[5] where a similar WOEID is given to places located within close proximity to the same street. Other spatial proximity approaches could also be used.
(2) Spatial clustering is then followed by a textual clustering process to isolate place instances from the identified spatial clusters based on the similarity of given place names. An improved version of the Levenshtein distance (French *et al.* 1997) that is based on word-level matching, as opposed to character-level matching is used here as follows.

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

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

In the tag-resolution stage, tags which correspond to concepts of place type and activity, as defined in the place model, are identified. This stage involves first identifying and collecting existing place type and place-activity reference data sets and using those as a basis for matching and classification of the tag collection. Two different sources are used for collecting place type information: (1) an official data set from the Ordnance Survey (OS), the national mapping agency of the United Kingdom, and (2) the GeoNames web

Table 1. Example place types and corresponding purposes from OSBP.

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

gazetteer, built collaboratively by users and containing over 10 million place names. The OSBP ontology contains over 200 place types that are used to describe building features and place types surveyed with the intention of improving use and enabling semi-automatic processing of this data. GeoNames also supports a place ontology that associates places with a hierarchy of place types represented as feature codes. It contains over 600 unique feature codes corresponding to place types such as: store, school and university, etc.

Two resources are also used for identifying possible human activities that can be associated with geographic places: (1) 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 (2) the OpenCyc ontology,[6] 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 1 shows example records of the place type and purpose associations.

Approximately, 400 distinct activity concepts are retrieved from both ontologies. An online implementation of the SPARQL Protocol and RDF Query Language (SPARQL) endpoint used to access both data sources can be found at.[7] Examples of the extracted place activities are: boating, eating, fishing, travelling, walking, etc.

Tags in the folksonomy are matched against the lists of extracted place type and activity concepts and matched tags are used to populate the place model. Matching is carried out on stemmed tags, using Porter stemming algorithm. Tags that may correspond to either a place type (e.g. shop) or an activity (e.g. shopping) are added as instances of both classes.

4.3. Semantic association and ontology building

Here, analysis is carried out to extract the relationships between the identified tag collection of place types and activities from the previous stage. A place-type sub-ontology and a place-activity sub-ontology are created to represent a folksonomy-specific view of these concepts (denoted as folkontology). 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.

Subclass hierarchical relationships between place-type ontology instances and between place-activity ontology instances are defined using a probabilistic model of subsumption, originally introduced by Sanderson and Croft (1999), where for any given concepts/tags x and y, $x$subsumes y if

$$P(x|y) = 1 \text{ and } p(y|x) < 1$$

In other words, $\underline{x}$ subsumes y if the place resources with which tag y is used are a subset of the resource with which x is used. Because x subsumes y and because it is more frequent,

x is represented as the parent of y in the hierarchy. Through informal analysis of the possible term pairs satisfying the subsumption conditions in the data set used, the condition $P(x|y) = 1$ was relaxed to $P(x|y) = 0.8$, as was also adopted in Sanderson and Croft (1999). The value was found to be sufficiently high to allow for the co-occurrence relationships between tags to be captured in this case.

The degree of relatedness between concepts is derived using co-occurrence similarity measures. A commonly used method to measure tag similarity is the cosine similarity method (Markines *et al.* 2009), where similarity between two tags is defined as:

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

where t_i represents a tag and R_i represents the set of instances of resources associated with the tag t_i in the folksonomy. An association relationship is defined between two concepts if the cosine similarity between their corresponding tags was found to be above a certain threshold.

4.4. Building user profiles

A folksonomy is defined as a quadruple $\mathbb{F}:=(U, T, R, Y)$, where U,T,R are finite sets of instances of users, tags and resources, respectively, and Y defines a relation, the tag assignment, between these sets, that is, $Y \subseteq U \times T \times R$ (Hotho *et al*. 2006, Abel 2011). A folksonomy can be interpreted as a hypergraph where each edge corresponds to a tag assignment so that $\mathbb{G} = (V, E)$, where $V = U \cup T \cup R$ is the set of vertices and $E = \{\{u, t, r\}|(u, t, r) \in Y\}$ is the set of hyperedges. Further, a folksonomy can be transformed into a tripartite undirected graph, which is denoted as folksonomy graph $\mathbb{G}_{\mathbb{F}}$.

A Folksonomy graph $\mathbb{G}_{\mathbb{F}} = (V_{\mathbb{F}}, E_{\mathbb{F}})$ is an undirected weighted tripartite graph that models a given folksonomy $\mathbb{F}$, where: $V_{\mathbb{F}} = U \cup T \cup R$ is the set of nodes, $E_{\mathbb{F}} = \{\{u, t\}, \{t, r\}, \{u, r\}|(u, t, r) \in Y\}\}$ is the set of edges, and a weight w is associated with each edge $e \in E_{\mathbb{F}}$.

One approach to model users in folksonomies is to model them by means of their *personomy* (Hotho *et al.* 2006), which represents the tagging activities that a particular user performed. The *personomy* $\mathbb{P}_u = (T_u, R_u, I_u)$ of a given user $u \in U$ is the restriction of $\mathbb{F}$ to u, where:

- T_u and R_u are finite sets of tags and resources, respectively, that are referenced from tag assignments performed by the user u and
- I_u defines a relation between these sets: $I_u := \{(t, r) \in T_u \times R_u|(u, t, r) \in Y\}$.

Personomies can be exploited to create tag-based profiles that are essentially weighted tags. The weights associated with tags are the count of how often a user u applied a given tag $t : w_u(t) = |\{r \in R_u : (t, r) \in I_u\}|$.

Here, three different approaches for creating user profiles from the folksonomy are compared. The first approach is the basic association of direct tags (DT) with users as derived from the personomy definition in this article. The second approach extends the profiles by computing tag similarity in the folksonomy structure, and the third approach extends the profiles by analysing tag similarity using the derived folkontology structural relationships.

4.4.1. User profiles with direct tags (DT)

The DT approach defines user profiles as collections of tags together with corresponding weights representing users' interest in each of these tags. Hence, a user profile PF for user u is defined as follows.

$$PF_u = \{< t_i, w_i > | t_i \in T_u, w_i = w_u(t_i)\}$$

where T_u and w_u are defined as in the personomy definition above.

4.4.2. Folksonomy-extended user profiles (FE)

A basic tag-based user profile is first constructed as above. Cosine similarity between tags in the profile and the rest of the tags in T is computed. The set of tags in the basic profile is extended with the set of all other tags with a similarity value >0. The weight assigned for each new tag is the maximum similarity value computed for that tag. The strength of the association between the user and tags can be controlled by user-defined parameters in the equation and the enriched user profile $PF_u^{'}$ is represented as follows.

$$PF_u^{'} = \left\{ < t_i, w_i > | w_i = \begin{cases} \alpha w_i & , if t_i \in T_u \\ \beta Max\left(\mathrm{sim}\left(t_i, t_j\right)\right) & , \forall\left(t_i \in \{T - T_u\} \wedge t_j \in T_u\right) \end{cases} \right. .$$

In the equation, parameters $PF_u^{'}$ and $\beta \in (0, 1]$ can be used to indicate the level of association of the tag to the user, depending on whether the tag is directly annotated by the user or it is similar to a tag annotated by the user.

4.4.3. Folkontology-extended user profiles (FOE)

A basic user profile is also constructed first using DT. However, in this case the profile is enriched only with isolated tags used to populate the place ontology from the folksonomy. Each tag in the set of DT T_u is used to query the place ontology; if a tag is identified as a place type or place activity, all related concepts to this tag, within a specified semantic distance, are retrieved and added to the profile. The weight assigned to the new tags is a function of the minimum semantic distance of that tag in the folkontology. The enriched user profile $\tilde{PF}_u$ is represented as follows.

$$\tilde{PF}_u = \left\{ < t_i, w_i > | w_i = \begin{cases} \alpha w_i & , if t_i \in T_u \\ \beta / Min\left(SemDist\left(t_i, t_j\right)\right) & \forall\left(t_i \in \{T - T_u\} \wedge t_j \in T_u\right) \end{cases} \right.$$

Where parameters α and $\beta \in (0, 1]$ and $SemDist$ is the semantic distance between the two tags t_i,t_j, defined here as the minimum number of edges connecting the two tags in the ontology relationship graph (Budanitsky and Hirst 2006).

4.4.4. User profile example

Figure 4 is an example folksonomy consisting of four users, five tags and six place resources. The tagging activity of each user is represented by a hyperedge connecting user, tag and place. A basic user profile is shown in the table below. Each row in the table

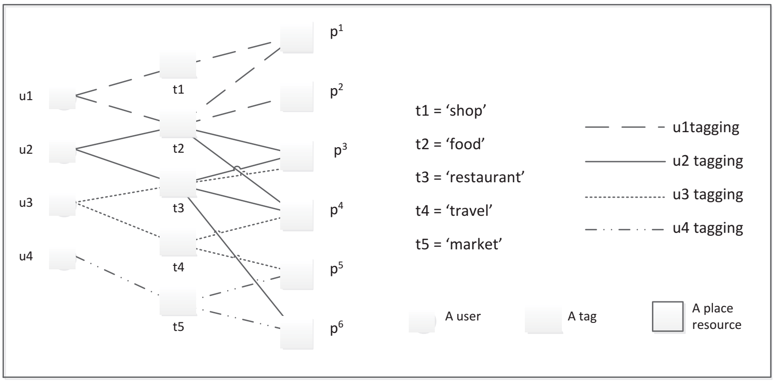


Figure 4. An example folksonomy.

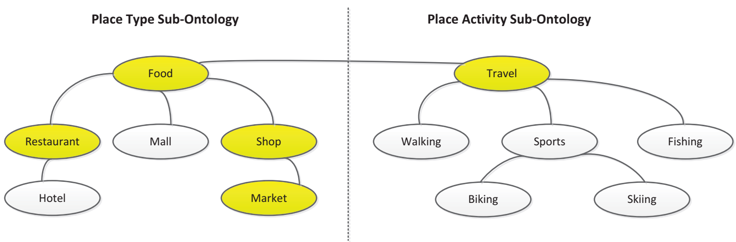


Figure 5. A snapshot of the place ontology illustrating the relations between the concepts in user profiles.

represents a user profile. The values in each cell represent the weights (defined as the frequency of use) between a user and a tag pair.

For illustration purposes, Figure 5 shows a sample of place-type and place-activity ontologies as derived from the folksonomy data set used in the experiments presented later in this work. Derived subsumption relationships are represented between tags in each ontology (e.g. between 'travel' and 'walking') and co-occurrence similarity relationships are represented between tags across the two ontologies (e.g. between 'food' and 'travel').

In Figure 5, concepts representing DT in the user profiles in Table 2 are highlighted. For demonstration, assume a semantic distance threshold of 1. The basic user profiles can be updated with related tags from the ontology as shown in the Table 3.

For example, user (U1) becomes associated with the tag 'travel' as a consequence of the association between the tags 'food' and 'travel' in the ontology, etc.

Table 2. Basic user profiles extracted for the example folksonomy in Figure 4.

User/tag	t1 (Shop)	t2 (Food)	t3 (Restaurant)	t4 (Travel)	t5 (Market)
U1	1	2	0	0	0
U2	0	2	3	0	0
U3	0	0	1	2	0
U4	0	0	0	0	2

Table 3. Modified user profiles using the relationships from the induced place ontology, where $\alpha = 1$ and $\beta = 0.5$ for demonstration.

User/tag	t1 (Shop)	t2 (Food)	t3 (Restaurant)	t4 (Travel)	t5 (Market)
U1	1	2	0.5	0.5	0.5
U2	0.5	2	3	0.5	0
U3	0	0.5	1	2	0
U4	0.5	0	0	0	2

5. Evaluation

A data-collection process is first used to build a local geo-folksonomy repository. A crawler software is developed to process pages from Tagzania.[8] The crawler is used to extract the geo-folksonomy generated by user interaction on this application. For our experiments, the collected geo-folksonomy data set included 22,126 place instances in the UK and USA, 2930 users and 12,808 distinct tags. The total number of collected geo-folksonomy tuples is 68,437. The data-cleaning stage resulted in identifying 19,614 clusters and corresponding unique places resources. Approximately, 11% (2512) of the total number of place resources were merged.

Figure 6 shows a subset of the derived place semantics, in which place types and activities are presented with their corresponding association and subsumption relationships (dashed boxes in the figure are used for simplification to indicate that a group of concepts share the relationships, thus 'beach', 'spa' and 'casino' all share the same relationship with 'hotel'). While some of the derived relationships can conceptually be recognised, the semantics of others cannot directly be associated. For example, 'beach' is associated with the activities like 'walking' and 'fishing', while it is subsumed by a

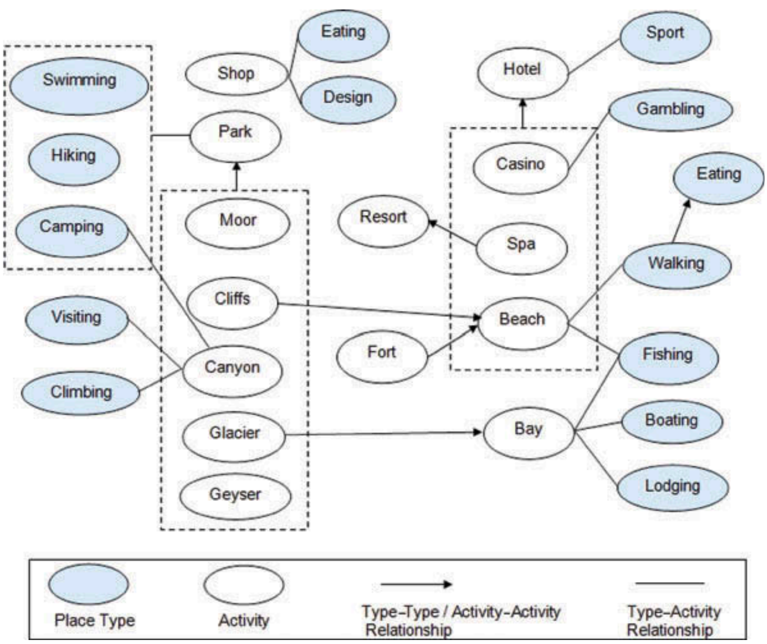


Figure 6. A snapshot of the derived ontology showing a number of place types, their related place activities and subsumption relationships.

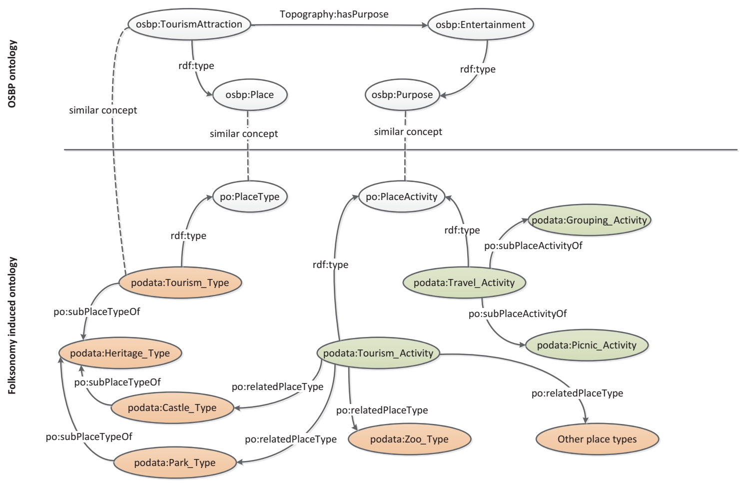


Figure 7. An example of a place type concept ‘tourism’ as defined in the Ordnance Survey ontology and its computed definition in the enriched place ontology.

‘hotel’. One can reason that ‘beach’ is not a subclass of ‘hotel’, but instances of type ‘beach’ may be located within close proximity to instances of type ‘hotel’. It is important to note that as the folksonomy data set increases, derived relationships between concepts are likely to change and refine. Thus, for the purpose of this study the subsumption conditions are used to capture possible semantic relationships between concepts in the sub-ontologies and reflect the usage patterns of the terms in the folksonomy. Further refinement of these relationships will be considered in the future.

The resulting induced folkontology is likely to be different from a traditional place ontology designed for the purpose of map making, for example. By nature, the concepts and relationships identified from the folksonomy are user specific and dynamic, reflecting snapshots of users’ views and experiences in the geographic place. Hence, traditional ontology evaluation approaches, in particular comparing to a golden standard (Vrandečić 2009) are not directly applicable in this context.

Figure 7 compares the semantics related to the place type *Tourism Attraction* as defined in 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 data set (*Tourism* is related to six other place types and four place activities). An absolute comparison is not realistic, where the OSBP serve the specific purpose of map creation and use, whereas the folkontology concepts essentially reflect the dynamic usage of the concepts by users for specific place instances.

5.1. Evaluating the quality of the place folkontology

One way to evaluate the quality of the place concepts and relationships derived from the folksonomy is to measure the level of agreement between the derived relationships and similar ones defined by users on the general web. This can be considered a sort of data-driven

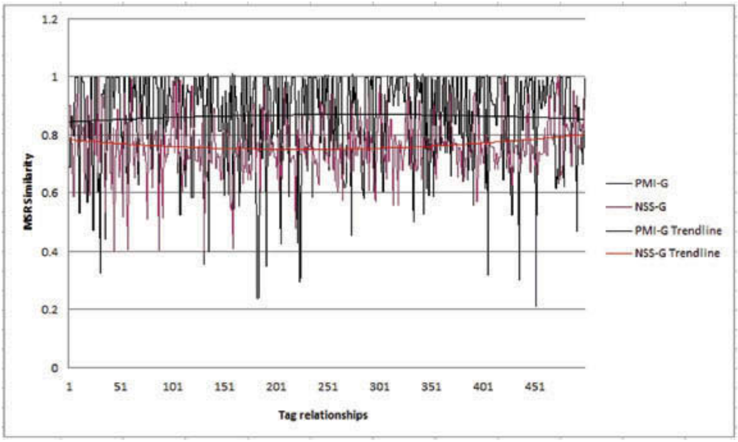


Figure 8. A graph showing the result of the semantic relatedness measures using the PMI and the NSS methods for a set of 500 relationships in the induced place ontology.

evaluation to the folkontology (Brewster *et al.* 2004), where the comparison is against data provided by users are in a context (in this case, unconstrained information provision on the web) similar to the context of usage in the social location-sharing applications.

The Measure of Semantic Relatedness (MSR) (Veksler *et al.* 2007) provides a set of methods to calculate the semantic relatedness between two terms.

MSR assumes that the strength of the relation between two terms is proportional to the number of times the two terms co-occurred in the same documents on the web. The performance of the different MSR methods in terms of quality and accuracy was found to be dependent on the size and type of the input data (Emadzadeh *et al.* 2010). Here, two of the more popular methods used to measure semantic relatedness in large data sets are employed, namely, Point-wise Mutual Information (PMI) and Normalised Search Similarity (NSS) (Matveeva 2008).

Five hundred relations in the induced ontology that link place types, place activities or both are evaluated using the PMI and the NSS methods. The average value of semantic relatedness computed for the PMI measure is 0.86 (with standard deviation of 0.16) and 0.77 for NSS (with standard deviation of 0.1). Figure 8 shows the output of both measures and their corresponding trend lines and Table 4 shows their values for a sample of 10

Table 4. A sample of the MSR measures calculated using PMI and NSS applied on the ontology relationships between places types (T) and activities (A).

Concept 1	Concept 2	PMI	NSS
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%

relationships. The high average values are indicative of a strong association between the concepts identified and used in the ontology.

The mean of the differences between both measures and the cosine similarity values between the terms (for cosine similarity ≥0.5) was also computed. This was found to be 0.267 for PMI and 0.183 for NSS, suggesting a fair degree of agreement between the methods. It is to be noted that further experiments with larger geo-folksonomy data sets need to be carried out to establish the significance of the derived relationships and their value for estimating the semantic relatedness of place concepts.

5.2. *Analysis of the user profiles*

A primary goal of this work is to identify place-related concepts and semantics in the geo-folksonomy and use these to build user profiles reflecting the relationships between users and places in the data set. Here, an analysis of the effectiveness of this approach is considered by comparing the different methods used for generating the user profiles.

Four versions of the user profiles are created and compared: basic profiles with DT, profiles enriched with similar tags using cosine similarity (FE) and profiles enriched with place-related tags from the place ontology with one-step semantic distance (FOE-SD1) and two-step semantic distance (FOE-SD2). The ratio of the place-related concepts identified against the number of distinct tags used in creating the user profiles provides a measure of the effectiveness of the methods employed. It should be noted however that the quality and relevance of the derived concepts need further evaluation, normally through a user study that tracks and builds profiles for a group of users over time. This is the subject of future work.

The data set contains 12,808 users and 2930 tags. Table 5 illustrates the output of the profiles in terms of the total number of place types and place activities against the total number of distinct tags used to build the profiles.

Enriching the basic user profiles using cosine similarity with tags that are 80% or more similar to the tags directly used by users resulted in an increase of the total number of tags used in the profiles by 3252 tags, of which 41 are place types and 34 are place activities. Although a high threshold value is used, the number of the retrieved place-related concepts is small compared to the total number of tags retrieved.

Utilising the place ontology to enrich the basic user profiles by retrieving concepts with one-step semantic distance from the tags in the profile resulted in retrieving 93 tags, of which 52 are place types and 41 are place activities. 78% of place-related concepts in the FOE-SD1 user profiles come from similarity relationships and thus overlap with the set in the FE user profiles. The rest of the tag set in FOE-SD1 user profiles come from subsumption relationships. As the semantic distance between the tags increase, more tags

Table 5. Statistics of place types and activities in user profiles constructed using DT four approaches to building the user profiles.

	Place semantics		
Method/count	Types	Activities	Distinct tags
DT	191	63	3639
FE	232	97	6891
FOE-SD1	243	104	3732
FOE-SD2	322	140	3907

Figure 9. (a) Place–user heat map with 1-step semantic distance; (b) Place–user heat map with 2-steps semantic distance.

are collected in the user profiles. With the two-step semantic distance FOE-SD2, the set of place-related concepts is almost doubled compared to FOE-SD1 (268 tags), with only 36% of overlap with the FE user profiles. The exercise demonstrates the effectiveness of the approach in detecting related concepts which are useful in the folksonomy.

Several interesting applications of the developed profiles can be envisaged. Two examples are illustrated here, namely, new place recommendation/association to users based on their extended profiles and measuring the similarity of users based on the derived association with place instances and place-related properties.

5.2.1. Place–user maps

Enriching user profiles can allow place resources in geo-folksonomies to be searchable and discoverable by more users. To illustrate this, user profiles were used to draw a heat map showing places and users related to places. Here, the association between user and place data would be based on the strength of the relation derived between the user and the place type and activity concepts of the different place instances.

The heat map shown in Figure 9a illustrates the relation between users and places using the FOE-SD1 user profiles. The size of the circle representing a place increases if more users can be related to that place. A place and a user are related if there is at least one common tag between the user profile and the tags of that place. Figure 9 shows the heat map using the FOE-SD2 user profiles, allowing many more users to be related to the place resources.

5.2.2. User-similarity analysis

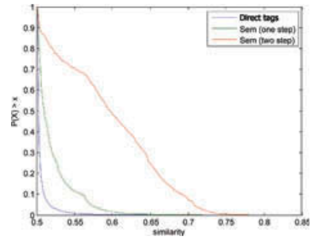
User similarity is another application of the proposed framework where user–user similarity is computed using three versions of user profiles: DT, FOE-SD1 and FOE-SD2. Table 6 shows the statistics for the user similarity based on the three profiles.

Table 6. Statistics for user similarity values using the enriched user profiles and the direct tags approach.

Profiles	Min	Max	Avg	Quartile (1st, 2nd, 3rd)
DT	0.0025	0.34	0.009	(0.0025, 0.005, 0.0075)
FOE-SD1	0.0025	0.437	0.039	(0.0075, 0.0175, 0.0525)
FOE-SD2	0.0025	0.56	0.192	(0.057, 0.185, 0.297)

Profiles	Average Similarity Value	Average Common Place Semantics
DT	0.1	0.3
SD1	0.25	1.5
SD2	0.34	2.3

(a)



(b)

Figure 10. (a) Average number of place-related tags in the top 100 similar user profiles; (b) CCDF of user similarity using the three user profile versions.

In order to understand the relationship between the similarity values and the place semantics, the top 100 user similarity relationships are further analysed in Figure 10a. Similarity calculated using DT gives more weight for a pair of user profiles if they share more tags, regardless of these tags being associated to the same places or representing place-related semantics.

Figure 10b shows the complementary cumulative distribution function (CCDF) of user similarity using the three user profile versions. The CCDF function describes the probability that a similarity value will be found at a value higher than or equal to *x*. It is noted that the enriched user profiles increase the probability of similarity matching. For instance, a probability of user similarity of value ≥0.1 is approximately 0.5 using the DT profiles and increases to approximately 0.55 with the FOE-SD1 profiles and 0.7 with the FOE-SD2 profiles. Further, increasing the semantic distance will increase the probability of user similarity, but this comes on the expense of the information content value in the profiles (value decreases as the degree of similarity between users saturates – tends to 1). Further studies need to be carried out on the relationship between the content of the user profiles of tags and places and their relevance and utility to users in different application scenarios.

6. Conclusion

Users' interactions and collaborations on the geo-social web generate geo-folksonomies that record tags used by users to label geographic places.

Interest in discovering and analysing place-related semantics implicit in this tag collection focussed on purpose-driven web-mapping applications where users collaboratively tag places to facilitate their identification and definition for the purpose of map making. In social-driven location-sharing applications, users have no restrictions on the sort of information they associate with places. Hence, tag collections in these applications can provide a rich resource of information on users' perceptions of geographic places and how it changes over time.

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 for discovering place semantics from geo-folksonomies. The approach involves several stages of folksonomy cleaning and preparation to address specific problems associated with noise and redundancy of place resources in the folksonomy. The identification and resolution of tags in the folksonomy is done by matching against a prepared reference data set of place type and activity information collected from existing ontology resources. The resolved tags are used to populate a place ontology and relationships between tags are recorded that map the structure of the underlying folksonomy. The induced folkontology is used to build user profiles and to discover relationships between users of the folksonomy. User profiles will suggest which place types and activity concepts a user is associated with and the strength of their associations with these concepts.

The framework was implemented and applied on a realistic geo-folksonomy data set. Results of the application of the different stages of the approach are presented and analysed. The value of using the framework in building enriched user profiles is demonstrated against conventional statistical methods used in folksonomy analysis. Two examples of possible applications of the enriched user profiles are also presented.

A diversity of emerging location-sharing applications is rapidly accumulating large amounts of geo-folksonomy data sets. The methods proposed in this work explore the challenges in analysing this data and demonstrate their potential value for user and place profiling and hence also for improving user experience on the web. Research still needs to be carried out to further evaluate the approach. In particular, a user study would be useful in measuring the relevance of the content of the generated user profiles. In addition, other dimensions of place-related semantics can be explored.

Notes

1. http://www.opengeospatial.org/standards/orm
2. http://www.w3.org/2003/01/geow/gs84_pos#
3. http://www.ordnancesurvey.co.uk/oswebsite/ontology/
4. http://cs.cardiff.ac.uk/2010/place-ontology\#
5. http://developer.yahoo.com/geo/geoplanet/
6. http://www.opencyc.org/
7. http://hobzy.cs.cf.ac.uk/sparql/
8. http://www.tagzania.com

References

Abel, F., 2011. *Contextualization, user modeling and personalization in the social web*. Thesis (PhD). Leibniz Universität Hannover.

Adrian, B., Sauermann, L., and Roth-Berghofer, T., 2007. ConTag: a semantic tag recommendation system. *In*: T. Pellegrini and S. Schaffert, eds. *Proceedings of the 3rd International Conference on Semantic Technologies, I-Semantics*, 5–7 September, Graz, 297–304.

Agnew, J., 2011. *Space and place*. London: Sage Publications.

Au Yeung, C., Gibbins, N., and Shadbolt, N., 2009. Contextualising tags in collaborative tagging systems. *In*: F. Menczer, ed. *Proceedings of the 20th ACM conference on hypertext and hypermedia*, 29 June–1 July, New York: ACM, 251–260.

Ballatore, A. and Bertolotto, M., 2011. Semantically enriching VGI in support of implicit feedback analysis. *In*: K. Tanaka, P. Fröhlich, and K.-S. Kim, eds. *International symposium on web and wireless GIS, W2GIS*, 3–4 March, Heidelberg: Springer, 78–93.

Ballatore, A., Bertolotto, M., and Wilson, D.C., 2013. Geographic knowledge extraction and semantic similarity in openstreetmap. *Knowledge and Information Systems*, 37 (1), 61–81. doi:10.1007/s10115-012-0571-0

Bogers, T. and Van den Bosch, A., 2008. Recommending scientific articles using citeulike. *In*: P. Pu, D.G. Bridge, B. Mobasher, and F. Ricci, eds. *Proceedings of the 2008 ACM conference on Recommender systems*, 23–25 October, New York: ACM, 287–290.

Brewster, C., *et al.*, 2004. Data driven ontology evaluation. *In*: *International conference on language resources and evaluation*, 26–28 May, Paris: Evaluations and Language Resources (ELDA), 234–248.

Budanitsky, A. and Hirst, G., 2006. Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32 (1), 13–47. doi:10.1162/coli.2006.32.1.13

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*, 28–31 October, Santa Barbara: University of California, 52–54.

Chen, W., *et al.*, 2010. Generating ontologies with basic level concepts from folk-sonomies. *Procedia Computer Science*, 1 (1), 573–581. doi:10.1016/j.procs.2010.04.061

Emadzadeh, E., Nikfarjam, A., and Muthaiyah, S., 2010. A comparative study on measure of semantic relatedness function. *In*: V. Mahadevan and Z. Jianhong, eds. *2010 The 2nd international conference on Computer and Automation Engineering (ICCAE),* Vol. 1, 26–28 February, Singapore: IEEE Press, 94–97.

French, J., Powell, A., and Schulman, E., 1997. Applications of approximate word matching in information retrieval. *In*: F. Glolshani, K. Makki, C. Nicholas, and N. Pissinou, eds. *Proceedings of the sixth international conference on Information and knowledge management*, 10–14 November, New York: ACM, 9–15.

Golder, S.A. and Huberman, B.A., 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32 (2), 198–208. doi:10.1177/0165551506062337

Heymann, P. and Garcia-Molina, H., 2006. *Collaborative creation of communal hierarchical taxonomies in social tagging systems*. Stanford: Stanford InfoLab. Technical Report 2006-10.

Hotho, A., *et al.*, 2006. Information retrieval in folksonomies: search and ranking. *In*: Y. Sure and J. Domingue, eds. *Proceedings of the 3rd European conference on The Semantic Web: research and applications ESWC'06*, Vol. 4011, 11–14 June, Heidelberg: Springer, 411–426.

Intagorn, S., Plangprasopchok, A., and Lerman, K., 2010. Harvesting geospatial knowledge from social metadata. *In*: S. French, B. Tomaszewski, and C. Zobel, eds. *7th International conference on information systems for crisis response and management*, 2–5 May, ISCRAM Association, 13–23.

Keßler, C., *et al.*, 2009. Bottom-up gazetteers: learning from the implicit semantics of geotags. *In*: K. Janowicz, M. Raubal, and S. Levashkin, eds. *GeoSpatial semantics, GeoS 2009*, Vol. 5892, 3–4 December, Heidelberg: Springer, 83–102.

Kuhn, W., 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15 (7), 613–631. doi:10.1080/13658810110061180

Markines, B., *et al.*, 2009. Evaluating similarity measures for emergent semantics of social tagging. *In*: Y. Maarek and W. Nejdl, eds. *Proceedings of the 18th international conference on World Wide Web*, 20–24 April, New York: ACM, 641–650.

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

Mika, P., 2007. Ontologies are us: a unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5 (1), 5–15. doi:10.1016/j.websem.2006.11.002

Mülligann, C., *et al.*, 2011. Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. *In*: M. Egenhofer, N. Giudice, R. Moratz and M. Worboys, eds. *Spatial information theory, 10th international conference*, COSIT 2011, Vol. LNCS 6899, 12–16 September, Heidelberg: Springer, 350–370.

Niwa, S., Doi, T., and Honiden, S., 2006. Web page recommender system based on Folksonomy mining. *In*: S. Latifi, ed. *Proceedings of the third international conference on information technology*, New Generations IEEE Computer Society, 10–12 April, IEEE Press, 388–393.

Plangprasopchok, A. and Lerman, K., 2009. Constructing folksonomies from user-specified relations on flickr. *In*: Y. Maarek and W. Nejdl, eds. *Proceedings of the 18th international conference on World Wide Web ACM*, 20–24 April, New York: ACM, 781–790.

Popescu, A., Grefenstette, G., and Mollic, P., 2008. Gazetiki: automatic creation of a geographical gazetteer. *In*: A. Paepcke, J. Borbinha, and M. Naaman, eds. *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, 16–20 June, New York: ACM, 85–93.

Rattenbury, T., Good, N., and Naaman, M., 2007. Towards automatic extraction of event and place semantics from flickr tags. *In*: C. Clarke, N. Fuhr, and N. Kando, eds. *Proceedings of the 30th annual international ACM SIGIR conference*, 23–27 July, New York: ACM, 103–110.

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

Sanderson, M. and Croft, B., 1999. Deriving concept hierarchies from text. *In*: F. Gey, M. Hearst, and R. Tong, eds. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 15–19 August, New York: ACM, 206–213.

Schmitz, P., 2006. Inducing ontology from flickr tags. *In*: *Collaborative web tagging workshop at World Wide Web*. Scotland: Edinburgh.

Sen, S., 2008. Use of affordances in geospatial ontologies. *In*: *Proceedings of the 2006 international conference on towards affordance-based robot control*, Vol. 4760. Berlin: Springer-Verlag, 122–139.

Sen, S., Vig, J., and Riedl, J., 2009. Learning to recognize valuable tags. *In*: N. Oliver and D. Weld, eds. *Proceedings of the 14th international conference on Intelligent user interfaces*, 8–11 February, New York: ACM, 87–96.

Smart, P., Jones, C., and Twaroch, F., 2010. Multi-source toponym data integration and mediation for a meta-gazetteer service. *In*: S. Fabrikan, R. Reichenbacher, M. Van Kreveld, and C. Schlieder, eds. *6th International conference geographic information science*, GIScience 2010, Vol. LNCS 6292, 14–17 September, Heidelberg: Springer, 234–248.

Specia, L. and Motta, E., 2007. Integrating folksonomies with the semantic web. *In*: E. Franconi, M. Kifer, and W. May, eds. *ESWC '07 Proceedings of the 4th European conference on The Semantic Web: Research and Applications*, 7 June, Heidelberg: Springer, 624–639.

Tso-Sutter, K., Marinho, L., and Schmidt-Thieme, L., 2008. Tag-aware recommender systems by fusion of collaborative filtering algorithms. *In*: R. Wainwright and H. Haddad, eds. *Proceedings of the 2008 ACM symposium on Applied computing*, 16–20 March, New York: ACM, 1995–1999.

Van Damme, C., Hepp, M., and Siorpaes, K., 2007. Folksontology: an integrated approach for turning folksonomies into ontologies. *Bridging the Gap between Semantic Web and Web*, 2, 57–70.

Veksler, V., *et al.*, 2007. A proxy for all your semantic needs. *In*: D.S. McNamara and G.J. Trafton, eds. *29th Annual meeting of the cognitive science society*, 1–4 August, Austin, TX: Cognitive Science Society, 1878.

Vrandečić, D., 2009. Ontology evaluation. *In*: S. Staab and R. Studer, eds. *Handbook on ontologies, international handbooks on information systems*. Berlin: Springer, 293–313.