Masters Program
in Geospatial Technologies

Master Thesis

Enhancing Information Retrieval in Folksonomies Using Ontology of Place Constructed from Gazetteer Information

Rania Sabrah

March 2009

Institute for Geoinformatics
University of Muenster

First Advisor: Prof. Dr. Werner Kuhn
University of Muenster, Germany

Second Advisor: Mohamed Bishr
University of Muenster, Germany

External Advisor: Prof. Dr. Fernando Bação
University Nova de Lisboa, Portugal

Table of contents

1. Introduction ... 1
1.1. Background ... 1
1.2. Motivation ... 2
1.3. Context of the Research problem ... 2
1.4. Problem statement ... 3
1.5. Research Hypothesis ... 3
1.5.1. Research objectives ... 3
1.5.2. Research questions ... 3
1.6. Methodology ... 4
1.7. Expected Results ... 4
1.8. Conclusion ... 4
2. Related work ... 5
2.1. Folksonomy ... 5
2.2. GeoSpatial Information extracted from Folksonomy: ... 7
2.3. Folksonomy characteristics ... 9
2.4. Gazetteer ... 10
2.4.1. Gazetteer components ... 10
2.4.2. GeoNames ... 11
2.4.3. Semantics of spatial relationships in gazetteer ... 12
2.5. Ontology of place: ... 12
2.6. Conclusion ... 13
3. Folksonomy, Problems and Approaches ... 15
3.1. Folksonomy Definition ... 15
3.2. Tagging Behavior ... 16
3.3. Folksonomy aspects ... 17
3.3.1. Tagging Relationship ... 18
3.3.2. Tag Frequencies ... 19
3.4. Folksonomy Problems ... 19
3.4.1. Synonyms and homonyms ... 20
3.4.2. Basic Level Variation ... 20
3.4.3. Lack of Hierarchy ... 20
3.4.4. Absence of controlled vocabulary: ... 20
3.5. Approaches ... 21
3.6. Conclusion ... 21
4. Ontology of Place ... 23
4.1. Ontology definition ... 23
4.2. Ontology of Place ... 24
4.3. Ontology Design ... 24
4.4. Implementation Issues ... 29
4.5. Conclusion ... 29
5. Formalizing Ontology of Place ... 31
5.1. Selecting Country for case study ... 31
5.2. Tagging system (Flickr) ... 32
5.2.1. Search for photos ... 32
5.2.2. Retrieve tags of the photos ... 33
5.2.3. Edit Tags ... 34
5.2.4. Check Tags ... 34
5.3. Querying Gazetteer (GeoNames) ... 35
5.3.1. Study of Feature Types ... 35
5.3.2. Gazetteer reports styles ... 35
5.3.3. Query Gazetteer ... 36
5.3.4. Check report ... 37
5.4. Ontology of place (OWL) ... 37

5.4.1. Constructing the ontology 37
5.4.2. Checking the ontology 39
5.4.3. Edit ontology in protégé 39
5.5. Conclusion 40
6. Prototype of a Search Tool 41
6.1. Developing a search tool for ontology enhanced folksonomy 41
6.1.1. Search for a place 41
6.1.2. Suggestions of related places 43
6.2. Evaluation and Results 45
6.2.1. Evaluation method 45
6.2.2. Ontology of place testing for precision and recall 45
6.2.3. Test Results 46
6.2.4. Results discussion 47
6.3. Conclusion 48
7. Conclusions and Future Work 49
7.1. Achieved results against research outline 49
7.1.1. Can the dynamic knowledge provided by folksonomies be used as a resource for acquiring bottom-up knowledge? 49
7.1.2. What is the ontology of place? And how to formalize such an ontology? 50
7.1.3. How can we enhance information retrieval in folksonomies by establishing explicit spatial "Part-Of" relationships? 50
7.2. Future work 50
References 52
Appendix A: CD-ROM 55

List of Figures

Figure 1.1: Image retrieval 2
Figure 2.1: London city boundary…………… 8
Figure 2.2: Boundaries of top 100 places in London center …………….. 8
Figure 2.3: Search in Google maps 9
Figure 2.4: GeoNames Feature classes.. 11
Figure 3.1: Users in web 1.0 and web 2.0 15
Figure 3.2: Photo from Flickr.. 16
Figure 3.3: Flickr Interface.. 17
Figure 3.4: Schematic depiction of the collaborative tagging process 18
Figure 4.1: Gruber Ontology Definition.. 23
Figure 4.2: Ontology of Place conceptual design... 25
Figure 4.3: Taxonomy of gazetteer feature types... 26
Figure 4.4: Geographic Feature Type Ontology conceptual design...................................... 28
Figure 5.1: Approach workflow 31
Figure 5.2: Geographic Feature Type ontology in protégé 38
Figure 5.3: Ontology of Place in protégé 40
Figure 6.1: Search tool interface 42
Figure 6.2: Result list 42
Figure 6.3: Result of searching for photos 43
Figure 6.4: Suggested places that are related to Sacramento, capital of California state………… 43
Figure 6.5: Suggested places that are related to Sacramento, McLean county, Kentucky state.............44
Figure 6.6: Result set of search using, a: place name only, b: using place name and parent name...…45
Figure 6.2: Result from search for “Disneyland”... …47

List of Tables

Table 4-1: Feature types used to construct the Geographic Feature Type Ontology 26
Table 6-1: suggested places types for each feature type 44
Table 6.2: search result from Flickr 47

List of Code Snippets

Code snippet 5.1: Request to Search for photos... 32
Code snippet 5.2: Request to get tags list for a photo 33
Code snippet 5.3: gazetteer search request to retrieve geographic information of place names 36
Code snippet 5.4: constructing the geographic feature ontology 38

List of Results

Result 5.1: one photo record returned from the search... 33
Result 5.2: Tags list of one of the photos 33
Result 5.3: Sample from resulting XML file with unique tags 34
Result 5.4: search result retrieved from gazetteer for a place name... 37
Result 5.5: OWL class constructed using the developed tool 39

Abstract

Folksonomy (from folk and taxonomy) is an approach to user metadata creation where users describe information objects with a free-form list of keywords (‘tags’). Folksonomy has have proved to be a useful information retrieval tool that support the emergence of “collective intelligence” or “bottom-up” light weight semantics. Since there are no guiding rules or restrictions on the users, folksonomy has some drawbacks and problems as lack of hierarchy, synonym control, and semantic precision. This research aims at enhancing information retrieval in folksonomy, particularly that of location information, by establishing explicit relationships between place name tags. To accomplish this, an automated approach is developed. The approach starts by retrieving tags from Flickr. The tags are then filtered to identify those that represent place names. Next, the gazetteer service that is a knowledge organization system for spatial information is used to query for the place names. The result of the search from the gazetteer and the feature types are used to construct an ontology of place. The ontology of place is formalized from place name concepts, where each place has a “Part-Of” relationship with its direct parent. The ontology is then formalized in OWL (Web Ontology Language). A search tool prototype is developed that extracts a place name and its parent name from the ontology and use them for searching in Flickr. The semantic richness added to Flickr search engine using our approach is tested and the results are evaluated.

Acknowledgment

First and foremost, I would like to express my thankfulness and humility to Allah most high for giving me the power and will to finalize this research.

I would like to express my gratitude to my three advisors. My deepest thanks to my first advisor Prof. Dr. Werner Kuhn for his continuous support and valuable comments. Also my appreciation for the highly scientific classes I attended with him. Special thanks are due to my second advisor Mohamed Bishr for his close follow up with all the work done in this research, his bright constructive ideas to improve it, and for teaching me how to make a scientific research in 2005 and everything else he taught to me since then. I would also like to express my thankfulness to my external advisor Prof. Dr. Fernando Bação for his remarkable suggestions and comments to enhance the work.

Special thanks to Dr. Krzysztof Janowicz for his scientific discussions and recommendations that were always fruitful. Also special thanks to Sven Schade for his interest and valuable hints.

I am deeply indebted to all my professours and lecturers in Universitat Jaume I (UJI), Castellón, Spain and in Westfälische Wilhelms-Universität Münster (WWU), Institute for Geoinformatics (ifgi), Germany. Through each and every one of them I have expanded my knowledge and learned a lot.

I would also like to salute all my friends that I met in the geospatial technologies program. I would like to thank them for introducing their experiences and cultures, also for being a second family, brothers and sisters.

Finally yet importantly, I would like to thank my family for their continuous support, encouragement, and for putting up with me in all times. Their prayers were the main source of inspiration, motivation and encouragement to continue my study.

Rania Sabrah
February 2009

List of Acronyms

ADL	Alexandria Digital Library
AI	Artificial Intelligence
API	Application Programming Interface
CA	Canada
CASE	Computer-aided Software Engineering
DB	Database
DL	Description Logic
FTT	Feature Type Thesaurus
GNIS	Geological Survey Geographic Names Information System
NGA	The National Geospatial-Intelligence Agency
NLP	Natural Language Processing
OWL	Web Ontology Language
RDF	Resource Description Framework
UML	Unified Modeling Language
USA, US	United States of America
XML	Extensible Markup Language

1. Introduction

The goal of this research is to build a search mechanism that enhances information retrieval in folksonomy. In the following we describe the research background, state the motivation, and explain the methodology that will be used.

1.1. Background

The term “Web 2.0” became notable and widespread after the first “Web 2.0 conference” in 2004. Among the notions that emerged with web 2.0 are folksonomies. Folksonomy (from folk and taxonomy) or collaborative tagging is an approach to user metadata creation where users describe information objects with a free-form list of keywords (‘tags’) (Speller, 2007). Tags serve as way of organizing content for future use, including search and navigation. Tags are called "social tags" when they are available to others, to view or search, after being created by their owner. Social tagging systems such as "Flickr"[1], for photo sharing, is becoming more and more popular with a huge number of participants sharing and tagging a large number of photos. In June 2005 when Yahoo acquired Flickr (since then the users numbers haven’t been announced) it was announced that Flickr has 775,000 registered users and 19,5 million photos and a 30 percent monthly growth rate. More recent statistics shows that there are almost 3 billion photos. In fact, more than 160,000 images were tagged in one morning in October 2008, and more than 2.8 million images Geotagged in that month. Yet, folksonomy is not a formal taxonomy as there is no hierarchy, and no directly specified parent-child or sibling relationships between tags (Mathes, 2004), in other words folksonomies are devoid of formal semantics.

On the other hand, few years earlier Tim Berners-Lee articulated the semantic web vision (Berners-Lee, Hendler, & Lassila, 2001). "Ontology" is one of the main themes of the semantic web, and was defined by (Guarino & Giaretta, 1995) as “A logical theory which gives an explicit, partial account of a conceptualization” where conceptualization is an intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality. An important observation regarding the Semantic Web is that it still suffers from knowledge sparseness (i.e., it presents good coverage for certain topics, but very low coverage for others) (Angeletou, Sabou, Specia, & Motta, 2007). Nevertheless, we should not think that Web 2.0 and the Semantic Web, tags and rdf, folksonomies and ontologies are competing for the same space. The important question with respect to semantic web technology and Web 2.0 is not how to manage a trade-off, but rather, how to use them together for the best advantage (Szomszor et al., 2007). It has been proposed by researchers in different fields that integrating these technologies will preserve the simplicity and easiness of folksonomies and at the same time enhances its structure and value by accomplishing explicit relations between tags. A considerable number of investigations are motivated by the vision of “bridging the gap” between the Semantic Web and Web 2.0 by means of ontology-learning based on folksonomy annotations (Cattuto, Benz, Hotho, & Stumme, 2008)

Giving names to places is the simplest form of georeferencing, and was most likely the one first developed by early hunter-gatherer societies (Longley, Goodchild, Maguire, & Rhind, 2005). Most people use place names to refer to geographical locations, and will usually be entirely

[1] http://www.flickr.com

ignorant of the corresponding coordinates. The 150 most popular tags on Flickr are tabulated and listed on the site. As of October 15, 2008, this list included Over 25% (39 out of 150) proper place names like cities or countries, and this percentage was checked in other dates and found to be almost always the same. Users who search for photos also use place names and often they extend their search to other related places, but it is done iteratively as there are no explicit relations between places. Gazetteers solve this problem as they are knowledge organization systems for spatial information. They deliver feature types and geographic footprints for searched place names (Hill, 2006), as well as spatial relations indicating that a place is "part of" another (e.g. Los Angeles is part of California and California is part of United States)[2] which is an explicit relationship between place names that can be used in related search.

1.2. Motivation

Searching large databases of images is a well-known problem commonly referred to as image retrieval (Aurnhammer, Hanappe, & Steels, 2006). In a search for tagged photos, a user needs to find photos of places that are related to the same place name she/he typed e.g. the user will search to find photos taken in "Muenster" (tagged with "Muenster") and would also like to see photos taken in places around the city or in Germany in general. Current folksonomies do not provide relations between place names, making it difficult to efficiently search by place.

1.3. Context of the Research problem

Websites using folksonomies allow users to search their database for resources tagged by the searched keyword. Figure 1.1 shows the sequence of steps to add and retrieve photos from the database:

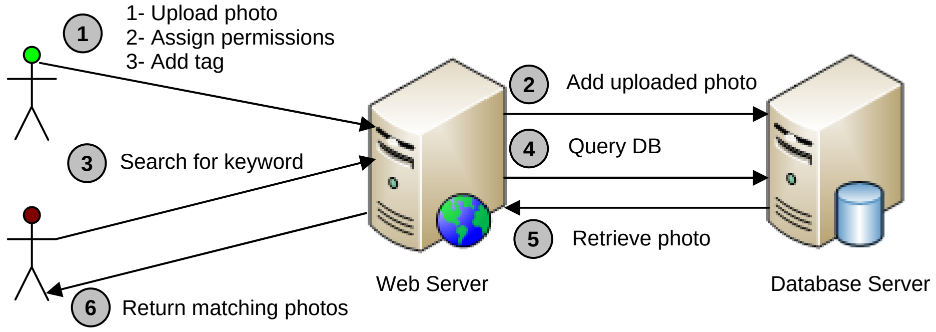


Figure 1.1: Image retrieval

1- Users in Flicker take the following steps to add their photos
    a. Upload photos through the web site
    b. Assign permissions to be public (visible to anyone), or private (visible to family only, friends only, both of them, or only for the user).
    c. Add tags with any keyword, description, or change any of the properties of the photo.
2- The uploaded photo and its attributes are added to the database

[2] http://www.alexandria.ucsb.edu/

3- Later, other users can search for photos by submitting a search keyword
4- The database is queried to find photos tagged with this keyword
5- Photos that match the search is retrieved from the database
6- The resulting matching photos are returned to the user depending on the permissions assigned to the photos, photos assigned as private are not returned to public users.

This scenario raises two challenges

1- Finding the appropriate result for the searched keyword. If it is a place name all photos tagged on this place should be returned to the user.
2- Suggesting to the user other related tags, of related places (related in the sense of location not as words). This permits users to navigate through a hierarchy of places when conducting place based search.

1.4. Problem statement

Folksonomy lacks hierarchy, synonym control, and semantic precision (Bishr & Kuhn, 2007). Searching for objects tagged with a word in folksonomies, is based on keyword search. Therefore, a related search (suggesting related tags) can not be given to the user. Currently there are no available mechanisms to manage, organize, and relate the place name tags. The available clustering of tags is done based on statistical methods and doesn't take location into consideration. If a user needs to extract all objects for related places, she/he has to repeat the search as many times as the number of places required to cover all the possible tags that could have been used by taggers. Another problem may face the user if she/he doesn’t know the exact name of all the places, only some of them, or if these objects are tagged by a local name or synonym (e.g. using Firenze for Florence).

1.5. Research Hypothesis

Lightweight ontologies of place extracted from folksonomies enhance information discovery, by improving their structure, and add semantic richness by defining explicit relationships between place names.

1.5.1. Research objectives

1- Extract place name tags from other tags.
2- Improve the structure of the folksonomy by establishing explicit “Part-Of” relations between place name tags and formalizing them in ontology of place.
3- Use the ontology of place to enhance information discovery and retrieval in folksonomies.

1.5.2. Research questions

1- Can the dynamic knowledge provided by folksonomies be used as a resource for acquiring bottom-up knowledge?
2- What is the ontology of place? And how to formalize such an ontology?
3- How can we enhance information retrieval in folksonomies by establishing explicit spatial “Part-Of” relationships?

1.6. Methodology

Our methodology is to automate an approach for acquiring tags from folksonomy, disambiguate place name tags using a gazetteer service, and finally formalize an ontology of place using these place names and their spatial relationships retrieved from the gazetteer service. The ontology of place will then be used to enhance folksonomy and improve information retrieval. To achieve this we follow these steps:

1. Tags from a folksonomy website (Flickr) will be collected using the API (Application Programming Interface) of the website, and then stored in an XML (Extensible Markup Language) file
2. Place names have to be mined from this file. A gazetteer API will be used to search for place name tags; the names which correspond to a place will return spatial information.
3. Spatial relationships have to be discovered between related places. From the returned locations, relations between places "Part-Of" can be identified.
4. The places and their relations have to be organized in a concept-relationship manner. This is accomplished by formalizing the place concepts and their spatial relationships in a lightweight ontology.
5. The ontology will be tested and evaluated by implementing a search engine for finding photos tagged with place names.

1.7. Expected Results

This mechanism is expected to enhance information retrieval in folksonomies as it will allow for semantically interpreting place names. Also using ontology of place will increase precision and recall, i.e. the search results will be improved. Besides, the ontology will allow for suggesting hierarchy of related places to the user that she/he can navigate through them.

1.8. Conclusion

Folksonomy or social tagging is one of the technologies offered by web 2.0 with a potential to enhance information indexing and retrieval. Folksonomy lacks semantic precision, hierarchy, and explicit relations between tags, which hinder its use. Formalizing ontology from tags and defining explicit relations between them will add semantic richness, and improve the search results.

Place name tags implicit relations can be made explicit by the aid of gazetteer. Querying a gazetteer with a place name returns its location, feature type, and relationships between places. An ontology of place can then be formalized from the place name tags and their spatial relations acquired from the gazetteer. This procedure will preserve the simplicity of tagging systems and at the same time add the missing semantic precision and hierarchy.

This research will extract lightweight ontology of place from folksonomy, by consulting a gazetteer service. The lightweight ontology is feedback to the folksonomy search environment, to enhance the semantic search for places.

2. Related work

In this chapter we review the research done in the areas our research is concerned with. This includes folksonomy, gazetteer, and ontology of place. More importantly, we focus on the work done to integrate these technologies. Many researchers have studied integrating folksonomy with ontology and gazetteer with ontology, but to the best of our knowledge no research used data from folksonomy with gazetteer service.

2.1. Folksonomy

This research is motivated by the vision introduced in (Bishr & Kuhn, 2007) explaining the changing role of users from information consumers to both consumers and producers. Folksonomy defined to be community based metadata, and a means of establishing semantics. The authors pointed to other work done in this field, such as (Mika, 2005) and (Gruber, 2005), that combines folksonomies with ontologies. They stated that the resulting lightweight ontologies (from folksonomy) can be used in a variety of applications, including enhancing the structure and value of a folksonomy, by allowing users to search for more abstract or specialized concepts within it.

In (Mika, 2005) a tripartite Actor-Concept-Instance model of ontologies was developed, where actor is the user, concept is the tag and instance is object annotated (e.g. photo, website, ...). Networks of folksonomies were represented in this tripartite graph with hyperedges, two lightweight ontologies are then extracted from the graph, one is based on overlapping communities (ontology of actors and concepts) and the other is overlapping sets of instances (ontology of concepts and instances). The results of the two case studies show impressive results on the emergence of semantics. The author concluded by noting the potential application of the results to improve tagging systems e.g. by offering search and navigation based on broader/narrower terms. Also the ontologies emerging from folksonomies have a large potential for enriching established, but slowly evolving linguistic ontologies such as Wordnet (Fellbaum & NetLibrary, 1998). This set of defined triples of user-tag-resource was widely used in most of the research in folksonomy done after that, as it will be shown in the following review.

The second pointed work is (Gruber, 2005) where the model of (Mika, 2005) is extended to include the source of the tag (the system where the tag originated) offering a Tagging (object, tag, tagger, source) construct in an ontology for tags dubbed TagOntology. The “source” was introduced to allow the idea of integrating environments of social tagging i.e. enable analyzing and reasoning over tag data across applications.

Based on these models, more research was done, all with the aim of integrating folksonomy with semantic web, but with different approaches, one approach was to combine folksonomy with lexical resources to formalize ontologies as in (Van Damme, Hepp, & Siorpaes, 2007), (Schmitz, 2006), (Specia & Motta, 2007), and (Specia, Angeletou, Sabou, & Motta, 2007), or formalize rdf annotation as in (Maala, Delteil, & Azough, 2007), or even by developing interfaces for users to edit tags and use an existing ontology as in (Peters & Weller, 2008). The rest of this section reviews these approaches.

A mash up of different resources was introduced in (Van Damme et al., 2007) the authors followed the model introduced by (Gruber, 2005) and combined it with lexical resources like dictionaries, Wordnet, Google and Wikipedia to disambiguate the tag sets obtained from different systems. The tag sets are enriched by trying to establish mappings to elements in existing ontologies. Also, the explicit relationships in existing ontologies were reused. As a final step a semi-automated approach was used, in which the aforementioned techniques are combined with collective human intelligence i.e. community confirms the resulted ontology and contribute with the missing information (e.g. missing relations between tags).

In (Schmitz, 2006) from Yahoo![3] Research team, he introduced ontology from Flickr[4] tags by using the statistical model for subsumption derived from the co-occurrence of tags, where a condition is used to define one term subsuming another. In his approach he considered a tag x subsumes another tag y if the probability of x occurring given y (the probability of finding tag x, in documents tagged with y) is above a certain threshold and the probability of y occurring given x is below that same threshold, he expressed this relation using the following equations

$$P(x|y >= t) \text{ and } P(y|x < t),$$
$$Dx >= Dmin, Dy >= Dmin,$$
$$Ux >= Umin, Uy >= Umin$$

Where:

t is the co-occurrence threshold, Dx is the number of documents in which term x occurs, and must be greater than a minimum value $Dmin$, and Ux is the number of users that use x in at least one image annotation, and must be greater than a minimum value $Umin$. He used this approach to develop what they called revised, probabilistic model. The subsumption model is applied on sets of tags acquired from Flickr to build a graph of possible parent-child relationships.

In (Specia & Motta, 2007) a more sophisticated pre-processing of the tags is applied. They also identified groups of related tags, and investigated the nature of these relationships by exploiting information available on the semantic web in order to give semantics both to the tags themselves and to the relationships between tags. Their methodology consisted of three steps. First step is pre-processing, accomplished by setting some rules for text verification (e.g tags must start with a letter), using similarity metric (implemented in the package SimMetrics[5]), and Filter out tags occurring less than a certain number of times. Second step is clustering of tags statistically by organizing tags in co-occurrence matrix ($n \times n$ symmetric matrix) and applying space statistics method. The last one is Concept and Relation Identification by mapping pairs of tags to concepts in existing ontologies and then using wikipedia[6] and google[7] to establish relations. (Specia et al., 2007) continued the same work by focusing on the third step, they used each tag (not in pairs) to extract all Semantic Web Terms (SWT) whose label or local name matches the tag, then identify their relations.

(Maala et al., 2007) presented a new method to convert Flickr tags describing a picture into RDF annotations. Tags are classified into six clusters (location, time, event, people, camera, activity) using Wordnet. For the set of places in order to understand the meaning of the tags and correctly build an RDF annotation, the authors used a database built from crawling several websites (like for instance Yahoo! Meteo) to obtain lists of cities, with the countries and continents in which

[3] http://www.yahoo.com
[4] http://www.flickr.com
[5] http://sourceforge.net/projects/simmetrics/
[6] http://www. Wikipedia.com
[7] http://www. google.com

they are located, and also used an ontology of things where people can be (e.g. people can be in a car, that can be on a road, that can be in a state, ...). Then all tags grouped in the location category are ordered from the smallest to the largest, say (l1 ≤ l2 ≤ . .. ≤ ln). The generated triples are: (r, in, l1), (l1, in, l2). . . (ln−1, in, ln), where r is denoting the photo.

Activities to edit and organize tags have been described as "tag gardening” in (Peters & Weller, 2008), they argued that the structured folksonomies are able to enhance recall but fail in enhancing the precision of search results due to the lack of linguistic processing of the tags which has to be performed in advance of the semantic disambiguation of tags, and because automatic development or extraction of tag relations is the differentiation of the various associative relations or the allocation of somehow related tags. They introduced manual activity, performed by the users to manage folksonomies and gain better retrieval results. This approach propose using thesauri or lexical databases to detect synonyms and the user is asked to confirm that, and chose one of the tags for indexing, then thesauri or ontologies may be used for query expansion and query disambiguation. A tool is developed to allow the user to edit his tags from different platforms in the same environment.

2.2. GeoSpatial Information extracted from Folksonomy:

Research in geoinformatics has also studied folksonomies, particularly how to integrate it with current technologies offered by semantic web. Analyzing place semantics using folksonomy was introduced in (Schlieder, 2007) and continued in (Schlieder & Matyas, 2008). Three initiatives have used geotags to generate boundaries (Grothe & Schaab, 2008), (Wilske, 2008), and (Cope, 2008a).

An approach for modeling the collaborative semantics of geographic folksonomies was introduced in (Schlieder, 2007), it is based on multi-object tagging (one tag is used to describe more than one object) to analyze consequences for the semantics of place concepts. The author used the ternary tagging relation (object, tag, user) developed in (Gruber, 2005) but with a collection as first argument: ageing({object1, … , objectN}, tag, user) a user-to-user similarity, in place conceptualization, is then measured based on the objects (photos in this use case) selected by each user using Tanimoto measure for similarity, it resulted in communities of geospatial information. With the same aim of looking at spatial choices to uncover spatial conceptualizations, but more statistical approach (Schlieder & Matyas, 2008) automatically analyzed web-based collections of images of geographic objects located in cities to gain insight into the spatial choices of the photographers. This approach was based on spatial clustering and has been implemented in a software tool, the Heatmapper. They discussed the concept of (photographic) popularity of a place and introduced a measure of popularity for the points of view used by the photographers.

An approach towards an automated generation of Spatial footprints of vague places with imprecise boundaries was introduced in (Grothe & Schaab, 2008), based on the statistical evaluation of a set of points acquired from geotagged photographs (has latitude/longitude coordinates) from Flickr that are assumed to lie in the region. Two classes of statistical methods are applied, kernel density estimation (KDE) and support vector machines (SVM), the two methods were used to estimate footprint of known regions to test their performance. Then the methods are used to estimate footprints of Alps, Black Forest, and Rocky Mountains, the methods were evaluated with quantitative measures, for which the authors used recall, precision, and F-scores.

The results showed that the SVM approach outperformed the KDE in the majority of cases.

With the same aim of approximating the spatial footprint of neighborhoods but using different method, (Wilske, 2008) used geotags to define vague boundary. First the geographic center of the region was determined using the spatial median of the set of points, and then the "egg yolk" representation approach was used to approximate the boundary. A vague region is represented as a pair of concentric regions with determinate boundaries: A core region (the "yolk") that contains all locations that definitely belong to this region and a surrounding hull region (the "egg") that contains all locations whose membership to the region is indeterminate.

Flickr team of developers also published some articles on Flickr developers blog[8] describing their work in defining boundaries for continents, countries, cities, and neighborhoods. (Cope, 2008a) describes this approach, geotags are used with Alpha shape method; it is a geometric concept which is mathematically well defined, using this methods most of the boundaries shown in this blog looked good compared to the exact boundaries. But it has to be taken into consideration that boundaries has to be defined for each area separately, i.e. defining boundaries of all neighborhoods inside a city will not define the city boundaries as shown in another blog called "GEOBLOGGERS" (Catt, 2008). Figure 2.1 shows the city of London defined using Alpha shape for London geotags, Figure 2.2 shows city of London defined by top hundred places (neighborhoods) in London center.

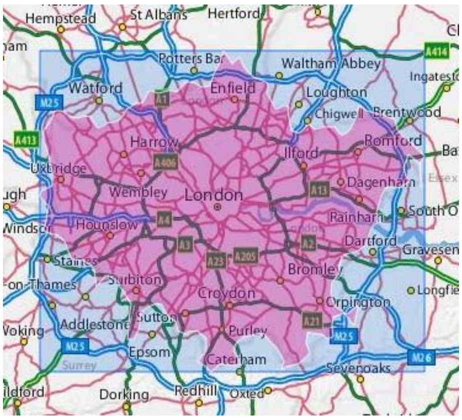


Figure 2.1: London city boundary

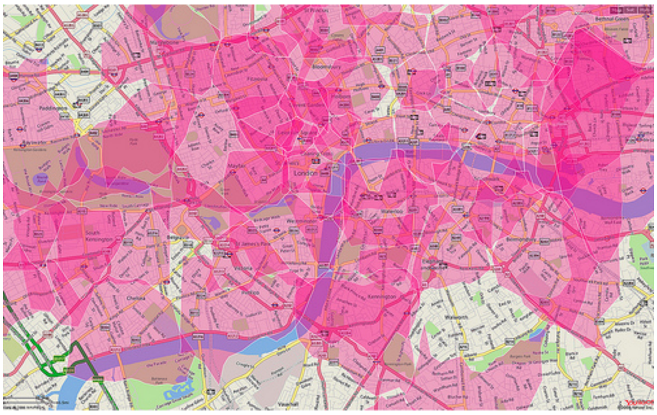


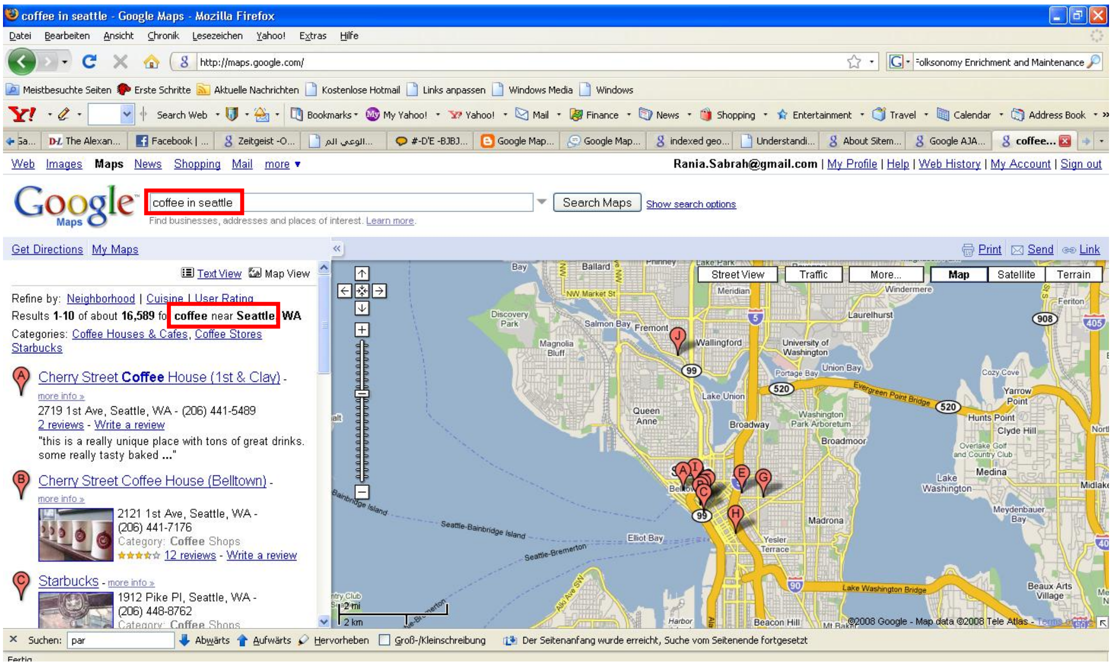
Figure 2.2: Boundaries of top 100 places in London center

Figure 2.2 is the zoomed center of London. The overlapping boundaries shows that some photos are either tagged with inaccurate tags or located in wrong coordinates, and the gaps are because these are only 100 places, and because some locations don't have photos. This made Flickr in the same blog (Cope, 2008a) invite users to organize what they called "PhotoWalk" around the edges of their neighborhood and add them to the map. These created boundaries are saved as shape files and available to download using Flickr API.

Flickr also offer a method in their API, which is given a place ID as argument and returns names and IDs of children places. The hierarchy defined for this parent child relation is neighborhoods, localities (cities or towns), regions (states) and countries. This hierarchy is clustered by place type, as defined in (Cope, 2008b) they call this process inverse geocoding, the article doesn't explain more details about how this clustering is done.

[8] http://code.flickr.com/blog/

On the other hand Google maps[9] allows users to search for business in or near a location (e.g. coffee in Seattle, or restaurants near Münster). In both cases as it is shown in Figure 2.3 the search on the map displays result for business near the location, i.e. searching for “coffee in Seattle” or “coffee near Seattle” the results are displayed for “coffee near Seattle”. This method is used to search for businesses only, it can’t be used to search for other geographic features as lakes, rivers, ...etc. And if it is used to search for “Lake near Münster” for example the returned result set are business places that has the word “Lake” as part of its name. Google allows the users to access this information and imbed Google maps functionalities in websites through Google maps API, but the documentation about how these functionalities are implemented is not available. The only source is the “Google Maps API discussion group” where it is explained that finding a business near a city is done by searching a circle range around the centroid of the city. (Williams, 14 November 2007)



coffee in seattle - Google Maps - Mozilla Firefox

http://maps.google.com/

Rania.Sabrah@gmail.com | My Profile | Help | Web History | My Account | Sign out

coffee in seattle

Find businesses, addresses and places of interest. Learn more

Refine by: Neighborhood | Cuisine | User Rating

Results 1-10 of about 16,589 fo coffee near Seattle WA

Categories: Coffee Houses & Cafes, Coffee Stores Starbucks

A Cherry Street Coffee House (1st & Clay) - more info »
2719 1st Ave, Seattle, WA - (206) 441-5489
2 reviews - Write a review
"this is a really unique place with tons of great drinks. some really tasty baked ..."

B Cherry Street Coffee House (Belltown) - more info »
2121 1st Ave, Seattle, WA - (206) 441-7176
Category: Coffee Shops
12 reviews - Write a review

C Starbucks - more info »
1912 Pike Pl, Seattle, WA - (206) 448-8762
Category: Coffee Shops

Suchen: par

Der Seitenanfang wurde erreicht, Suche vom Seitenende fortgesetzt

Datei Bearbeiten Ansicht Chronik Lesezeichen Yahoo! Extras Hilfe
Folksonomy Enrichment and Maintenance
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten Kostenlose Hotmail Links anpassen Windows Media Windows
Search Web Bookmarks My Yahoo! Yahoo! Mail Finance News Shopping Entertainment Travel Calendar Address Book
Web Images Maps News Shopping Mail more
Search Maps
Show search options
Get Directions My Maps
Print Send Link
Text View Map View
Street View Traffic More... Map Satellite Terrain
Seattle-Bainbridge Island
Elliott Bay
Lake Washington
Madrona
Medina
Lake Union
Fremont
Wallingford
Queen Anne
Broadway
Laurelhurst
2 mi
2 km
©2008 Google - Map data ©2008 Tele Atlas
Abwärts Aufwärts Hervorheben Groß-/Kleinschreibung
Fertig

Figure 2.3: Search in Google maps

2.3. Folksonomy characteristics

A literature review was introduced in (Speller, 2007), he defines folksonomy as an approach to user metadata creation where users describe information objects with a free-form list of keywords (‘tags’). This is done often to allow users to organize and retrieve the objects at a later date. The paper lists some of the implementations of folksonomies such as “Del.icio.us” for bookmarking web links, “Flickr” for photo-sharing, “CiteULike” for bookmarking scholarly writing and journal articles in particular, “Last.fm” for music and “YouTube” for video. The author argues that folksonomy can make an important contribution to digital information organization, but that it may need to be integrated with more traditional organization tools to overcome its current weaknesses.

In other research (Cattuto et al., 2007) used statistical tools to gain insights into the underlying tagging dynamics and introduced a stochastic model of user behavior embodying two main aspects of collaborative tagging: (i) a frequency-bias mechanism related to the idea that users are exposed

[9] http://maps.google.com/

to each other's tagging activity; (ii) a notion of memory, or aging of resources, in the form of a heavy-tailed access to the past state of the system. Remarkably, their simple modeling is able to account quantitatively for the observed experimental features with a surprisingly high accuracy. This points in the direction of a universal behavior of users who, despite the complexity of their own cognitive processes and the uncoordinated and selfish nature of their tagging activity, appear to follow simple activity patterns. They considered this approach a starting point upon which more cognitively informed studies can be based, with the final goal of understanding and engineering the semiotic dynamics of online social systems. Devising methods to measure the semantic relatedness between tags and characterizing it semantically is studied later in (Cattuto et al., 2008). In this paper, they consider the three following measures for the relatedness of tags: the co-occurrence count, the cosine similarity of co-occurrence distributions, and FolkRank. They map the tags of del.icio.us to synsets (sets of synonyms that represent one concept) of WordNet and use the semantic relations of WordNet to infer corresponding semantic relations in the folksonomy. The paper concluded that the three relatedness measures are best for studying three semantic characteristics:

- Cosine similarity to discover synonyms;
- FolkRank and co-occurrence relatedness for algorithms to extract taxonomic relationships between tags (concept hierarchy);
- FolkRank to discover multi-word lexemes.

2.4. Gazetteer

The core elements of a digital gazetteer are the place name, the type of place it labels, and a geographic footprint representing its location and possibly its extent. Such gazetteer data is an important component of indirect geographic referencing (or inverse geo-coding as called by Flickr) through place names. (Hill, 2000) describes the main components of gazetteers and explains the development of Alexandria Digital Library (ADL) Gazetteer. (Janee, 2006) discusses the limitations of the ADL gazetteer and the semantics of the spatial relationships defined between places. Developing an ontology as a modification of an existing feature type thesaurus is introduced in (Janowicz & Keßler, 2008)

A gazetteer is defined by (Hill, 2000) as geospatial dictionaries of geographic names with the core components of

- A name (could have variant names also)
- A location (coordinates representing a point, line, or areal location)
- A type (selected from a type scheme of categories for places/features).

2.4.1. Gazetteer components

Place Name: A place name is used to reference a particular geographic object.

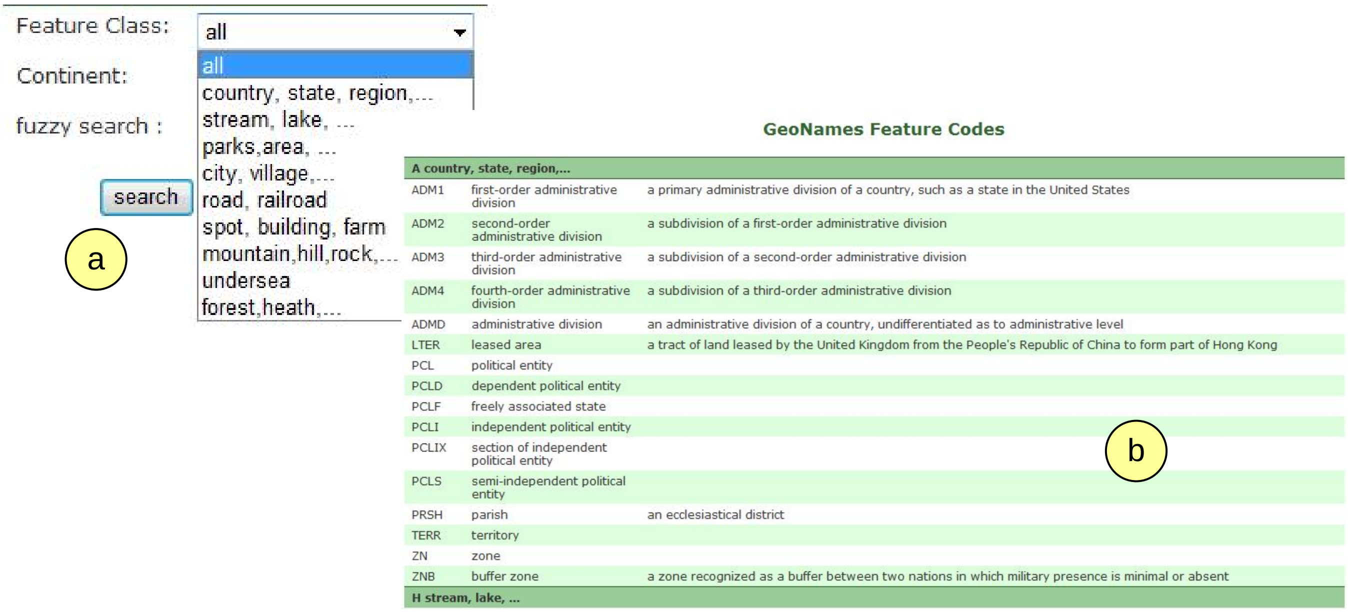
Place Location: A geographic footprint representing the location of a named place, is the other component of a digital gazetteer. This footprint, in latitude and longitude coordinates, can be point, bounding box, line, polygon, or grid representation.

Place Type: A Place type represents classes of geographic objects described by Place names, such as city or river. Different classifications of Place types may be used to serve different contexts of use, e.g. topographic and administrative. Hierarchical classifications of Place types are often used with sub-class and super-class relationships, e.g. a motorway is-a road. Place types may also be spatially related, most commonly by containment relationships, modeled through "Part-Of" hierarchies. However, other types of relationships may also be possible, e.g. intersection between road objects.

Note the difference in the semantics of "Part-Of" relationships, where spatial "Part-Of" denotes physical containment, whereas semantic "Part-Of" does not, e.g. a faculty (including a set of people) may be "Part-Of" a university but with no specific spatial relationship, while a departmental building may have a "Part-Of" relationship of physical containment within a university campus (G. Fu, Abdelmoty, & Jones, 2003).

2.4.2. GeoNames

In this section we review GeoNames[10] which is an online gazetteer; as it is going to be used in our research to retrieve spatial information about place name tags. It is a geographic database with over eight million geographic names.



GeoNames Feature Codes

A country, state, region,...		
ADM1	first-order administrative division	a primary administrative division of a country, such as a state in the United States
ADM2	second-order administrative division	a subdivision of a first-order administrative division
ADM3	third-order administrative division	a subdivision of a second-order administrative division
ADM4	fourth-order administrative division	a subdivision of a third-order administrative division
ADMD	administrative division	an administrative division of a country, undifferentiated as to administrative level
LTER	leased area	a tract of land leased by the United Kingdom from the People's Republic of China to form part of Hong Kong
PCL	political entity	
PCLD	dependent political entity	
PCLF	freely associated state	
PCLI	independent political entity	
PCLIX	section of independent political entity	
PCLS	semi-independent political entity	
PRSH	parish	an ecclesiastical district
TERR	territory	
ZN	zone	
ZNB	buffer zone	a zone recognized as a buffer between two nations in which military presence is minimal or absent
H stream, lake, ...		

Figure 2.4: (a) GeoNames Feature classes drop-down list, (b) extract from the feature codes classification

In the context of gazetteers, a feature is a real world entity. The feature type which is selected from a typing scheme or ontology is used for feature categorization. A named geographic place is an abstract entity defined to refer to a physical region (extent) in space and categorized (typed) according to commonly agreed upon characteristics (Janowicz & Keßler, 2008). As shown in (Figure 2-4), in GeoNames all features are categorized into one out of nine groups, called Feature classes (it is not structured as a thesaurus; it is a simple list), and further subcategorized into one of 645 feature codes. These feature classes are adapted from The National Geospatial-Intelligence

Agency (NGA) Feature Designation Codes. But the names of the classes were not adopted. The most important sources of information in GeoNames are:

- NGA : National Geospatial-Intelligence Agency's (NGA) and the U.S. Board on Geographic Names (most names except US and CA)
- GNIS : U.S. Geological Survey Geographic Names Information System (names in US).
- www.geobase.ca (names in CA)

2.4.3. Semantics of spatial relationships in gazetteer

Semantics of spatial relationships is discussed in (Janee, 2006), the research shows that the ADL protocol provides two ways of expressing containment constraints: searching spatially (e.g., find “Santa Barbara” spatially contained within California’s footprint) and searching relationally (find “Santa Barbara” that has a Part-Of relationship to California). Conceptually, these two types of queries were shown to be equivalent (or, at any rate, any difference between them is surely splitting semantic hairs).

Development of feature type ontology based on feature typing was proposed in (Janowicz & Keßler, 2008) to improve both gazetteer interoperability and reasoning capability. Their approach is to take advantage of an existing feature typing scheme – the Alexandria Digital Library’s (ADL) Feature Type Thesaurus (FTT) – to create a portion of such ontology. Difficulties in mapping from feature type thesauri to ontologies are pointed out, a feature type ontology starting with the ADL FTT is created, and finally a description of how the proposed feature type ontology can be integrated into the gazetteer communication paradigm is shown. The same approach will be followed in this research to construct a geographic feature type ontology, that will be populated by place name tags to formalize ontology of place.

2.5. Ontology of place:

Ontology of place or geo-ontology is an ontology that holds geographic information, place name, type, category and relations between places. Geo-ontology plays an important role in the development of “geospatial” semantic web, as it facilitates search for geographic information and resources. In the following we review the challenges of constructing this ontology in (Fu, Jones, & Abdelmoty, 2005), and (Abdelmoty, Smart, & Jones, 2007). And study constructed place ontologies in (Jones, Alani, & Tudhope, 2001) and (Henriksson, Kauppinen, & Hyvönen, 2008)

(Fu et al., 2005) argues that a geo-ontology plays a key role in the development of spatially-aware search engines, with regards to providing support for query disambiguation, query expansion, relevance ranking and web resource annotation. In this research the authors focused on the problem of integrating multiple datasets for constructing the ontology, and propose similarity measures for the integration. The geo-ontology constructed in this research supports multiple names, maintains more than one geometric footprint for each place, maintains classification categories of places, and encodes containment relationships between places. Four similarity measures are proposed, two related to the thematic properties of a place (place name and place type) and two are related to the spatial properties of a place (footprint and geographical hierarchy which is derived from containment relationships between places). The authors of this paper

[10] http://www.geonames.org/

published another research (Abdelmoty et al., 2007) that discussed the limitations of the OWL ontology language for the representation of Place. In this paper, they discussed the particular requirements of place geo-ontologies and the need for combining OWL and spatial reasoning rules to support their development and maintenance. Two frameworks for the development of Place ontology management systems are proposed. The first approach assumes a centralized view of ontology development, where the instance store (or ABox - assertion box which records observations of the world) is populated from available data sources. In the second approach, no (or limited) instance store is assumed and the place information is derived from the integration of multiple data resources.

In a different approach (Jones et al., 2001) present an ontology of place that combines limited coordinate data with qualitative spatial relationships between places. The aim is to match a specified place name (place is any geographic phenomena, i.e. city, lake, river, …etc) with place names that refer to equivalent or nearby locations but not finding places that are conceptually similar but possibly entirely separate in location. This similarity is determined based on a spatial closeness measure which is the combination of two distance measures, regional hierarchy distance and Euclidean Distance between centroids. In order to measure the semantic similarity between non-spatial concepts, they introduce a thematic distance. It is based on the principle of measuring the weighted distance between a pair of classification terms by the shortest number of links that separate them in the semantic net of classification terms. The results showed that an automatic ranking of places can be achieved by this approach.

In (Henriksson et al., 2008) they examine 1-the scope of geo-ontologies used for the purposes of information retrieval on the Web, 2-the core geographical concepts and their mutual relations, and 3-the concepts properties. They also discuss the Finnish geo-ontology and its development. They pointed out that geo-ontologies should contain classes that describe the spatial aspects of places (e.g. location), regional geography (e.g. administrative regions), patterns based on human interaction with nature (e.g. land use), and aspects related solely to the physical environment (e.g. landforms).

2.6. Conclusion

To sum up the review, this research is motivated by the vision of harvesting semantics offered by folksonomy as proposed in (Bishr & Kuhn, 2007) and presented by several researchers that integrated folksonomy with semantic.

We reviewed the researches done in folksonomy, indicating the problems encountered while using the system and the approaches taken to solve these problems and benefit from the emergent semantics that is offered by it. We focused in approaches taken to formalize semantic structures from folksonomy, and also the approaches for acquiring spatial information from geotags. Gazetteer service is studied, and the semantics of its spatial relationships is noted. Finally the different approaches to formalize ontology of places are reviewed, pointing out how they were used to enhance information retrieval from the web.

The review of past researches will influence our research approach as we will follow the same approach as offered by (Janowicz & Keßler, 2008) to construct ontology of place from administrative hierarchy levels defined in Feature Type Thesaurus, and linked by “Part-Of” relationship in the sense of spatial containment as explained by (Janee, 2006). The constructed ontology of place will follow the outlines proposed in (Fu et al., 2005) and (Jones et al., 2001). Place name tags acquired from Flickr will be disambiguated using a gazetteer service and will be used to populate the constructed ontology of place. Furthermore, the ontology will be used for the development of spatially-aware search engine that can be used to disambiguate search for place names in Flickr.

3. Folksonomy, Problems and Approaches

Tagging, folksonomy, distributed classification, ethnoclassification, are all labels for the same technology. The concept of users creating and aggregating their own metadata is gaining ground on the internet (Speller, 2007). In this chapter folksonomy definition and aspects are going to be discussed in more details. Also, the problems that arouse with tagging systems and hinder its use are pointed out to be taken into consideration when developing our research approach. Finally the approaches that are taken by different researchers to minimize these problems and maximize the benefit of folksonomy are explained.

3.1. Folksonomy Definition

Folksonomy is a method of indexing information over the web. Folksonomies are described in (Gruber, 2005) as an emergent phenomenon of the Social Web. They arise from data about how people associate terms with content that they generate, share, or consume. He also pointed out that such tagging systems attracted many users because they are easy to use, has no limit (tags can be as many or as few as a user wants), and there is no wrong answer.

The popularity of folksonomy was also due to the change in role of the user in the emerging technologies of the social web, as articulated in (Bishr & Kuhn, 2007). Users' roles switched from being data consumers to become data producers, or become "prosumers" - producers and consumers (Peters & Stock, 2007), prosumers collaborate not only for the purpose of creating content, but to index these pieces of information as well. Besides, tags are a form of metadata which allow searchers to easily find images concerning a certain topic such as place name or subject matter (Maala et al., 2007).

As it shown in figure 3.1, this change in user role, led the way to "collective intelligence". In this context it is stated that "With content derived primarily by community contribution, popular and influential services like Flickr and Wikipedia represent the emergence of "collective intelligence" as the new driving force behind the evolution of the Internet." (Weiss, 2005)

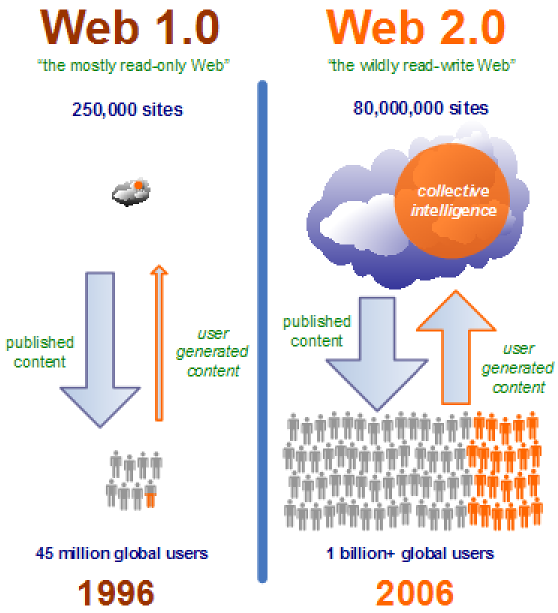


Figure: 3.1 Users in web 1.0 and web 2.0, Diagram source: http://web2.wsj2.com/

"Collective intelligence" arises from joint efforts of a group of authors or users in a so-called "collaborative services". Such services can be summarized under the tag "Web 2.0"(O'Reilly, 2005). They offer possibilities for keeping or searching diaries (Weblogs, Technorati), for the construction of encyclopedias (Wikipedia) and the management of bookmarks (Del.icio.us), photos (Flickr) or videos (YouTube). The collaboration does not stop with providing content but includes indexing of provided knowledge in some Web 2.0 services as well.

Flickr is a digital image storage and management website. It is a place to organize photos into albums, tag them with descriptive keywords, and view other users images. For example Figure 3.2 shows a photo in Flickr and its tags. Searching in Flickr for any of these tags (e.g. Las Vegas) will result in a set of photos that have "Las Vegas" as one of their tags, including the photo in figure 3.2. Flickr allows navigation by tag or user, as well as by group. Groups are places for users who share similar interests to post their images such as "Las Vegas Vacations" and "Movies" groups (Kroski, 2006). However, groups are more flexible than the traditional folder-based method of organizing files, as one photo can belong to many groups, or one group, or none at all (the concept is directly analogous to the better known "labels" in (Google's Gmail). Flickr's "Groups" then represent a form of categorical metadata rather than a physical hierarchy (Maala et al., 2007).

Figure 3.2: Photo from Flickr, tagged by (Vegas, Las Vegas, Nevada, america, uprightkangaroo, New York, statue, liberty)

3.2. Tagging Behavior

Studying the tagging behavior, (Maala et al., 2007) defined the following users tagging habits:

- Very few tags: some photos contain no tag at all or very few tags (one or two).
- Sentence tagging: users can use quotes to enter a full sentence as a tag, such as "You can not miss it" (in case no quotes are used, space is understood by Flickr as a separator between tags). Figure 3.3 shows Flickr interface.
- Vertical sentence tagging: it is the same case as the previous one, but users forget to (or intentionally did not) put the sentence between quotes. Thus the sentence can be read vertically, because Flickr has understood each space-separated word to be a different tag.
- Too many tags: contrary to the previous case, the information attached to the photo is very rich and describes many different aspects (content, location, etc.). In Flickr users can add up to seventy five (75) tags for each photo.
- Nonsense tags: these tags correspond to something not understandable for a human being not knowing the annotator universe of thinking such as "cs-tkl"
- Space free tagging: the users write a sentence by concatenating words in order to put the whole sentence on the same line; for example a user has written the tag "Ilovenature". These users may not be aware of the possibility of using quotes.
- Collective tagging: due to the interface Flickr provides, it is possible to tag several photos concurrently. Therefore it sometimes happens that a photo is described with a tag that does not apply directly to it but to a photo that has been uploaded at the same time.

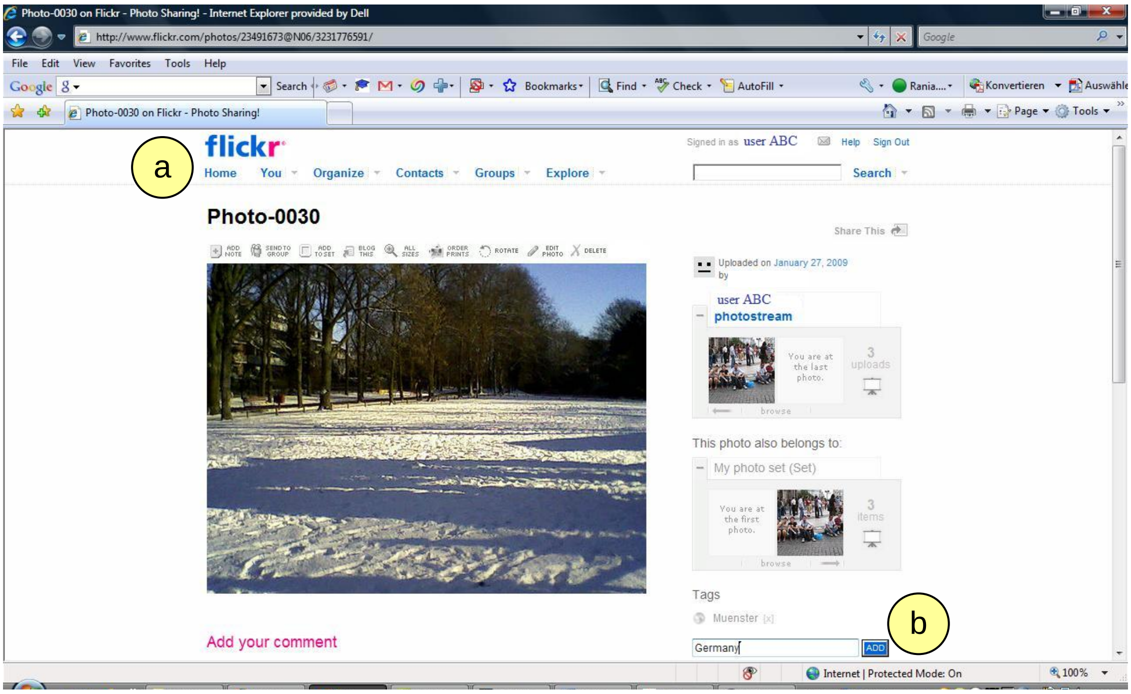


Photo-0030 on Flickr - Photo Sharing! - Internet Explorer provided by Dell
http://www.flickr.com/photos/23491673@N06/3231776591/
File Edit View Favorites Tools Help
flickr
Home You Organize Contacts Groups Explore
Signed in as user ABC Help Sign Out
Search
Photo-0030
Share This
Uploaded on January 27, 2009 by
user ABC photostream
You are at the last photo.
3 uploads
This photo also belongs to:
My photo set (Set)
You are at the first photo.
3 items
Tags
Muenster
Germany
ADD
Add your comment
Internet | Protected Mode: On
100%
a
b

Fig: 3.3 Flickr interface, a: tools to upload and organize user photos. (b): adding tags to photo

Although tags are user-defined and freeform, Flickr tags tend to naturally fall into five categories as defined by (Maala et al., 2007) and (Winget, 2006):

- Place: the location can be described at varying levels of granularity. At the largest level of granularity, the continent, the country, the region, the city, a mountain range are found frequently. At a smaller level of granularity, description of the building or the immediate natural site the photo was taken can be found: a building, a university, a house, a beach. Finally at the smallest level of granularity, there can be a description of a room or a piece of furniture: bed, chair , etc.
- Time: the time can also be described at different levels of granularity. The year, the season and the month are the most frequently found. The exact day is much less frequent. Some times of the day are (sunrise, sunset, etc.).
- Event: the holydays (Christmas), the birthdays, the weddings . . .
- Name: people names (Emma, Jean) or nicknames.
- Camera: many tags indicate the make or the model of the camera (Nokia, Canon), the colors (black & white), artistic judgments on the photo.

3.3. Folksonomy aspects

The idea of folksonomy is to allow users to add keywords to a system to describe an object or a resource. These objects differ from one application to the other depending on the goal of the site. Despite these differences in applications' goals, objects to be tagged, users of each application and the tags they use to describe the objects, folksonomy possesses some general aspects that will be discussed in the following.

3.3.1. Tagging Relationship

The uncoordinated tagging activity of the users creates a dynamic correspondent relationship between a resource and a set of tags, i.e. it creates an emergent categorization of the resources. This means that the tagging process develops social features and complex interactions(Cattuto et al., 2007). This relationship is shown by figure 3.4, users add tags to a resource in the system, and these tags define categorization of the resource.

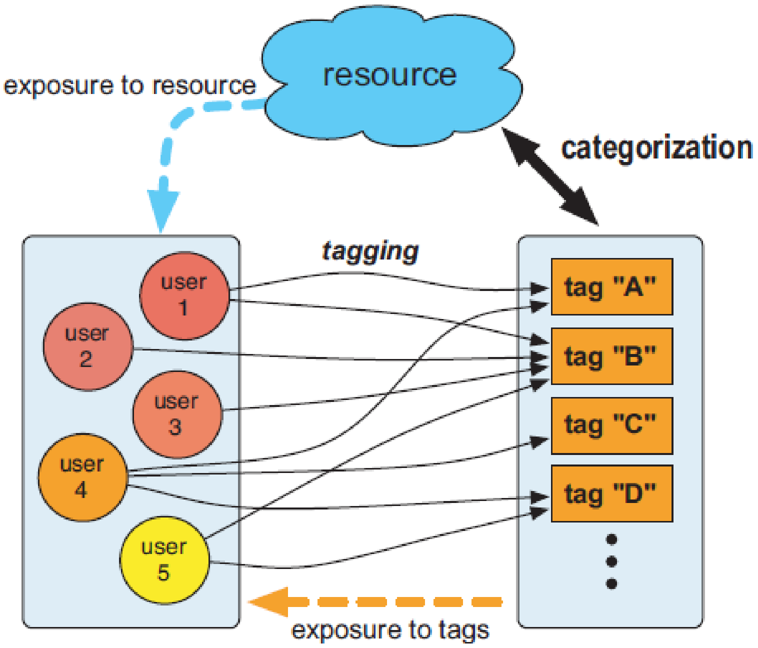


Fig: 3.4 Schematic depiction of the collaborative tagging process, Diagram source: (Cattuto, Loreto, & Pietronero, 2007)

Figure 3.4 represents a tagging system, one resource is tagged by different users (user 1, user 2, etc.) each of the users' added one or more tags, e.g. (User 1) tagged the resource with (tag "A"), and (tag "B"), (user 2) tagged it with (tag "B") and so on. Such relation was formalized by (Gruber, 2005) as

Tagging (resource, tag "A", User 1)

Tagging (resource, tag "B", User 1)

Tagging (resource, tag "B", User 2)

Generally it is expressed as: Tagging (object, tag, tagger). This relation is valid for one application, in case of exchanging this data between more than one application the relation is expressed as a four-place relation, with source (application) as part:

Tagging (Object, tag, tagger, source)

This allows us to say something about a collection of tag data, independent of the specific applications they come from. This triple of (object, tag, tagger) is called "Post" by (Cattuto et al., 2007) who argued that usually a post also contains a temporal marker indicating the (physical) time of the tagging event, so that temporal ordering can be preserved in storing and retrieving posts. Adding the time to the relation

Tagging (Object, tag, tagger, source, time)

Most of the objects tagged are associated with a location. For example in photo-sharing applications such as Flickr, the photos are taken in a certain place (location). Adding location to the objects will allow for clustering them according to their location. Hence we get a six-place relation of the form

Tagging (Object, tag, tagger, source, time, location)

This relation states that, the tagging relationship between a tag and an object creates categorization of objects in the system. The relationship is identified uniquely in an application by considering the tagger (user who added the tag). In order to be able to use data from different applications, the source is taken into consideration. The forth component in this relation is the timestamp indicating the time the relation was created. Finally the location associated with the relationship adds the spatial dimension to the relationship where this object is located. Our research falls in the context of studying this spatial dimension in tagging systems, which was not widely addressed before by other researches.

3.3.2. Tag Frequencies

In his research (Schlieder, 2007) proved that upon examining the tags by sorting them in order of decreasing frequency of use, the tag frequency followed a power law. Also on their experiment (Cattuto et al., 2007) showed that on plotting the number of distinct tags as a function of the total number of inserted tags, a clean power-law behavior without ever reaching a steady state plateau can be observed throughout the full history of the resource.

In striking contrast to this, in the same experiment (Cattuto et al., 2007) showed that the relative proportion of tags associated with a given resource quickly approaches a quasi-stationary condition. Once the number of posts associated with a resource is sufficiently large, single tagging events have a negligible effect on the global distribution of tags, so that the existing distribution is reinforced, generally becoming more and more stable. This robustness is a very important property of collaborative tagging. On the one hand, the fact that tag fractions stabilize quickly allows the emergence of a clearly defined categorization of the resources in terms of tags, with a few top-ranked tags defining a semantic “fingerprint” of the resource they refer to. On the other hand, the long-term stability of tag proportions makes the emergent categorization robust against noise. Both aspects contribute greatly to the actual usability of collaborative tagging systems.

These observations show that there is an emergent semantics through categorization. Capturing this semantics and formalizing it, in ontology for example, will significantly enhance the system performance.

3.4. Folksonomy Problems

“Folksonomy” name was first coined by Thomas Vander Wal in a discussion on an information architecture mailing list (Smith, 2004). It is a combination of “folk” and “taxonomy”. However, it is argued that folksonomy is a misleading term as the systems in question bear very little relation to taxonomies or ontologies as clarified by (Merholz, 2004a) and explained by Vander Wal later in (Vander Wal, 2005). Folksonomy has some drawbacks as pointed out by many writers (Schlieder, 2007), (Kroski, 2006), (Peters & Stock, 2007), (Mathes, 2004), and (Merholz, 2004b). These problems hinder the optimal use of folksonomy. The next sections discuss these problems:

3.4.1. Synonyms and homonyms

There is no synonym (different words, same meaning) control in tagging systems. This leads to tags that seemingly have similar intended meanings, like “CA” and “California” both are used to refer to “California State” in USA, this means that in a search for “CA” objects that are tagged with “California” will not be in the result of the search and vice versa. Plural vs. singular is often a problem, as seen in the popular tags on Flickr, both “flower” and “flowers” were listed (Mathes, 2004). Web 2.0 services such as Technorati, Flickr and YouTube are used almost all over the world (Peters & Stock, 2007). Many users in non-English-speaking countries tag documents using their own language like “Germany“, “Deutschland”, “Alemania”. This leads to the problems of trans-language synonymy (i.e., translation).

On the other hand, homonym (same word, different meaning) represents a different but related problem to synonyms (Speller, 2007). The presence of homonyms in the system means that search precision is reduced. One example is that a search for ‘Orange’ may give results about both fruit and “Orange county” which is a place in “California state”, one of which is bound to be irrelevant to the searchers’ needs (Weinberger, 2005). Flickr tries to improve precision by automatically generating ‘clusters’ of terms to disambiguate different meanings of the search term, which can work well: for example, look at the clusters formed around the ‘orange’ tag. Of course, this increase in precision results in a loss of recall as some photos may be tagged ‘orange’ only and as such will not be found but, as (Kroski, 2006) asserts, this is a necessary trade-off.

This problem limits the precision and recall of an information retrieval system as it decreases its ability to retrieve information.

3.4.2. Basic Level Variation

A thorny semantic problem is the issue of ‘basic level variation’ described by (Kroski, 2006). This relates to how much detail an individual will go into when tagging a resource. In (Golder & Huberman, 2006) they give the example of an average person naturally tagging a photo as ‘bird’ when a birdwatcher would have naturally tagged it ‘robin’. This difference means that the two users may find each other's tags next-to-useless because they are at the wrong level of specificity for their needs. This problem naturally recalls the notions of information communities and domain ontologies that are used to resolve such issues.

3.4.3. Lack of Hierarchy

Folksonomies are flat systems. There are no parent-child relationships, no categories and subcategories. Hierarchy is a distinguishing trait of traditional taxonomies that are able to provide a deeper, more robust classification of entities. Such systems allow users a finer granularity in searching for resources.

3.4.4. Absence of controlled vocabulary:

According to (Guy & Tonkin, 2006) most users don't give much thought to the way they tag resources, and bad or "sloppy" tags exists heavily in folksonomies. Also they observed that, there exists misspellings, incorrect encodings, and compound words, testing against multilingual dictionary software, they found that 40% of Flickr tags and 28% of del.icio.us tags were either misspelt, from a language not available via the software used, encoded in a manner that was not understood by the dictionary software, or compound words consisting of more than two words or a mixture of languages.

Folksonomies are discovery systems, these mentioned problems decrease their ability to discover and retrieve information. As mentioned in chapter 1 our research objective is to enhance folksonomy by improving its information retrieval, this can be achieved by solving these problems or decrease its effect.

The result of these problems is an uncontrolled and chaotic set of tagging terms that do not support searching as effectively as more controlled vocabularies (Guy & Tonkin, 2006). Despite this fact, (Merholz, 2004b) argues that these problems should be addressed, as there are encouraging potential benefits. He compares folksonomy to foot-worn paths or “desire lines” that appear in a landscape over time, these are trails that demonstrate how a landscape’s users choose to move, which is often not on the paved paths. A smart landscape designer will let wanderers create paths through use, and then pave the emerging walkways, ensuring optimal utility. In the context of folksonomy, this indicates that for optimal systems, rules shouldn’t be imposed on users. Else users should be free to add any tag to index resources, then these tags have to be formalized in a semantic structure as ontology, and fed back to the system to enhance its performance. This point of view is also supported by the observation from (Cattuto et al., 2007) discussed in section 3.3.2

3.5. Approaches

There are different approaches aiming to solve these problems of folksonomy some of them were discussed in the related work in section 2.1. One approach is to educate users to improve “tag literacy” (Guy & Tonkin, 2006). For training of the users to do this, it might be useful to suggest some tags. Tag-suggestions can operate on a syntactical level (e.g., a user attaches “graph” and the system suggests “graphics”) or even on a relational level (e.g., a user attaches “graphics” and the system suggests “image”, because both words do often co-occur in documents’ tag clouds). This approach involves establishing user research concerning folksonomies and studying the “semiotic dynamics” underlying collaborative tagging as in (Cattuto et al., 2007).

Another approach is to consider tags as elements of natural language and treats them by means of automatic methods of natural language processing (NLP), this can be reached by using thesauri or lexical databases to detect synonyms semantics, as the approach taken by (Peters & Weller, 2008) and (Van Damme et al., 2007). It can also involve merging ontologies, e.g. a geographic classification system, with a folksonomy to allow using the tags in their hierarchical relations as well (Gruber, 2005).

3.6. Conclusion

Folksonomy is easy to use, and has no rules, this is the reason there are increasing numbers of folksonomy applications over the web that are used by millions of users. We discussed the tagging relationship with six dimensions to be: Tagging (Object, tag, tagger, source, time, location) with the focus of this research on the “location” dimension of tagging systems. Studying tags frequency showed that on plotting the number of distinct tags as a function of the total number of inserted tags, a clean power-law behavior without ever reaching a steady state was observed, and on the other hand the relative proportion of tags associated with a given resource quickly approaches a quasi-stationary condition, this implies that formalizing a semantic structure of these tags as ontology, will highly affect the system performance. Nevertheless, folksonomy has many problems that hinder their use and hide the arising “Collective intelligence”, as synonym and homonym problems, basic level variation between tags, folksonomy lack of hierarchy structure, and uncontrolled vocabulary of users. These problems limit the system ability for information retrieval and decrease their precision and recall. Many approaches have been taken to solve these problems and add semantic structure to these tagging systems without enforcing rules on the user. The results of these efforts have proven that there is emerging semantics in folksonomy. Our approach in this research is another attempt to overcome folksonomy drawbacks and add semantic structure to it. The approach is discussed in more details in chapter five, and its evaluation and results in chapter six.

4. Ontology of Place

The most common way when searching for a location is to use place names. In order to assist in recognizing place names, it is proposed to employ an ontology that encodes geographical terminology and the semantic relationships between geographic terms. The idea is that the ontology of place will enable the search engine in folksonomy to detect that the query refers to a geographic location and to perform a search which will result in the retrieval of photos that refer to the specified location (Jones et al., 2001)

4.1. Ontology definition

"An ontology is an explicit specification of a conceptualization" as defined by (Gruber, 1993) . A 'conceptualization' refers to an abstract model of some phenomenon in the world by identifying the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined, figure 4.1 plots this definition. Basically, the role of ontologies in the knowledge engineering process is to facilitate the construction of a domain model. An ontology provides a vocabulary of terms and relations with which a domain can be modeled.

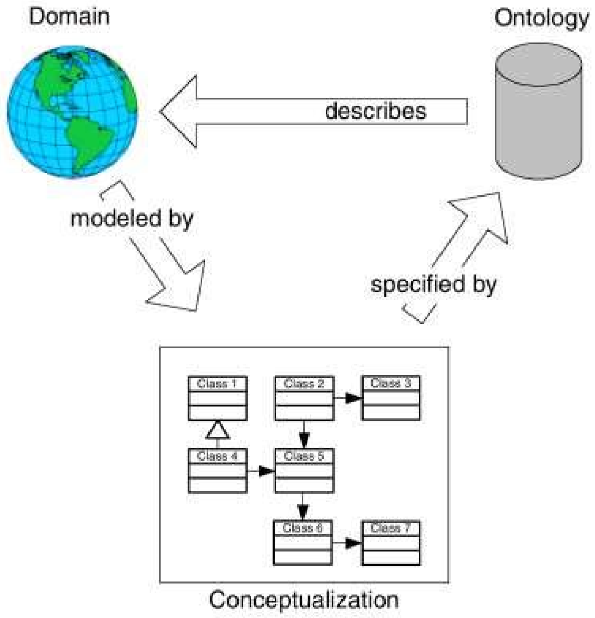


Figure 4.1: Gruber Ontology Definition (Lacy, 2005)

(Guarino & Giaretta, 1995) studied the different definitions given to "Ontology" to clarify what is meant by it. They came to a conclusion that ontology can be defined as "A logical theory which gives an explicit, partial account of a conceptualization" where conceptualization is an intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality.

4.2. Ontology of Place

Ontology of place or Geo-Ontology encodes names of places and spatial relationships. It should be noted that places occur at multiple levels of detail ranging from continents and oceans down for example to small villages, streets, buildings, monuments and streams.

(Sowa, 1996) distinguishes between formal ontologies, that define concepts with axioms and logic, and terminological ontologies that may use hierarchical structures but with limited formality. Formal ontologies provide the potential to reason automatically with the concepts and terminology of a domain. The idea of encoding terminology within hierarchical relationships is found in thesauri, which have been around for a long time in the context of information retrieval. A thesaurus may be regarded as a structured vocabulary, and as such is a terminological ontology. Terms in the vocabulary are associated with each other via broader term (BT) and narrower term (NT) hierarchical relationships. A distinction can be made between different types of BT/NT relationships, including sub-type and "Part-Of", but these distinctions are not always specified in particular thesauri (as in GeoNames) and hence may restrict their use in automatic reasoning (G. Fu et al., 2003).

4.3. Ontology Design

As indicated above, in designing ontology of place, we are concerned with a conceptualization of a place on the Earth.

4.3.1 Ontology modeling

In this research UML (Unified Modeling Language) is going to be used as a technique for modeling ontologies as discussed by (Rumbaugh, Jacobson, & Booch, 1998). Some of the reasons to use this language in ontology construction are: UML is easy to understand and use for people outside the Artificial Intelligence (AI) community, there is a standard graphical representation for UML models, and many Computer-aided Software Engineering (CASE) tools are available (Gómez-Pérez, Fernández-López, & Corcho, 2004).

(Kogut et al., 2002) showed how to use UML to represent lightweight ontologies. According to his proposal, UML class diagrams will be used for representing concepts (and their attributes), and relations between concepts (both taxonomic relationships and ad hoc ones).

In UML diagram, classes are represented with boxes divided into three parts, the name, the attributes and the operations of the class. Operations are not used for representing ontologies. Aggregation relationships of classes can be specified (which is equivalent to the "Part-Of" relationship). The aggregation relationship is denoted with an open diamond, the aggregate class (the class with the white diamond touching it) is the "whole", and the other class in the relationship is "Part-Of" that whole, this relationship is also one-to-many relationship (the "whole" is one that has many "parts") e.g. in "continent-country" relationship the continent is the "whole" and the country is "Part-Of" the continent, and one continent has many countries.

4.3.2 Conceptual design

It is proposed here that the ontology of place is composed of two ontologies as shown in Figure 4.2. The first ontology is the geographic feature type ontology which encodes the various feature types (extracted from gazetteer feature type thesaurus). The second ontology is the geographic feature ontology which encodes the concrete geographic features in a given geographic space.

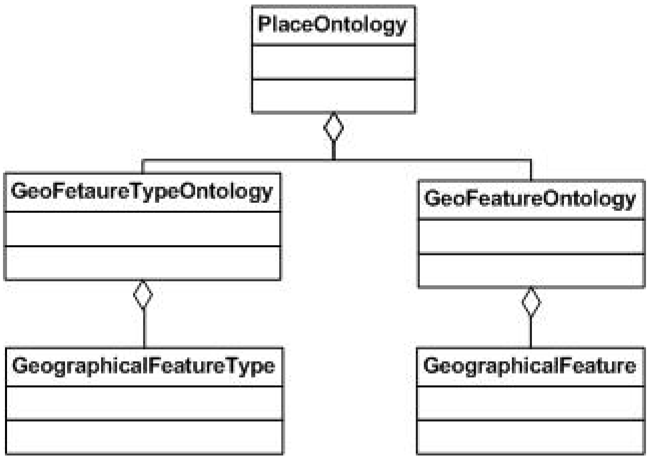


Figure 4.2: Ontology of Place conceptual design

First component: Geographic Feature Type Ontology

An extracted portion of the feature type thesaurus is used to construct the ontology. The feature types selected based on the following factors:

a. In gazetteer, feature types include abstract features such as boundaries (Country, State, County …) and physical features such as hydrological features (river, stream …) it is not a recommended practice to mix both in the same ontology. Besides, this contradicts an ontology definition as domain specification.
b. Features in gazetteer are usually represented with one point even for line features. Then the relationship “Part-Of” cannot be applied directly. For example a river will be defined by one point, which is contained by the boundary of one place, while in fact it runs in several places that could be counties or states.
c. The aim of the research is to identify a structure for place names used in tagging systems, i.e. disambiguate a place based on its hierarchy and siblings. This implies that the selected feature types have to be related to each other either as siblings or by parent-child relationship. The model showing the selected feature types and their relationships is shown in figure 4.3. The figure shows the feature types that will be used to construct the ontology. The relationships between the feature types is inferred from the gazetteer feature types definition.

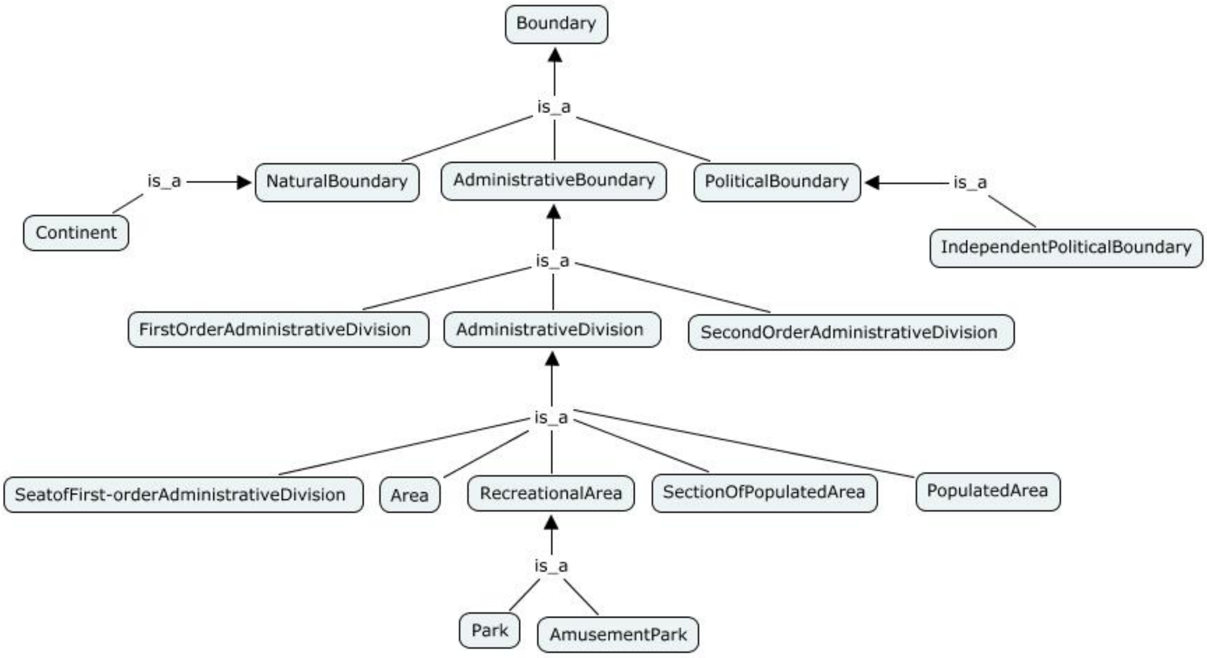


Figure 4.3: Taxonomy of gazetteer feature types

The following table defines the extracted feature types as explained by GeoNames gazetteer. Some feature types did not have a definition as “Independent political entity” and “Capital of a political entity”

	Feature type	Definition[11]
1	Continent	continent : Europe, Africa, Asia, North America, South America, Oceania,Antarctica
2	Independent political entity	No definition
3	Capital of a political entity	No definition
4	First-Order Administrative Division	a primary administrative division of a country, such as a state in the United States
5	Second-Order Administrative Division	a subdivision of a first-order administrative division
6	Seat Of a First Order Administrative Division	No definition
7	Populated Place	cities, towns, villages, or other agglomerations of buildings where people live and work
8	Section Of a Populated Place	No definition
9	Area	a tract of land without homogeneous character or boundaries
10	Park	an area, often of forested land, maintained as a place of beauty, or for recreation
11	Amusement park	Amusement Park are theme parks, adventure parks offering entertainment, similar to funfairs but with a fix location

Table 4-1: Feature types used to construct the Geographic Feature Type Ontology and their definitions

[11] Source of definition GeoNames, http://www.geonames.org/export/codes.html

Mapping from feature type thesauri to ontologies leads to some assertions as discussed by (Janowicz & Keßler, 2008). First, because the relationships in the thesauri are not explicitly defined, existing feature typing schemes can be converted to ontologies only through a process that includes validation of the relationships as required by ontologies. Second, in several cases these relations are not sufficient to disambiguate concepts, so that textual definitions and instances have to be taken into account.

In (Janowicz & Keßler, 2008) they followed the methodology described by (van Assem, Menken, Schreiber, Wielemaker, & Wielinga, 2004) to convert thesauri into RDF and OWL ontologies. Despite the fact that in GeoNames the feature types are not organized as a thesaurus, the types, hierarchy, and relationships extracted from the gazetteer can still be handled by the same methodology. This is because these types represent administrative level hierarchies with implicit defined relationships. The transformation methodology is structured into four steps:

- Preparation
- Syntactic conversion
- Semantic conversion
- Standardization

According to the syntactic conversion process, the proposed feature type ontology should preserve the structure and naming of the original thesaurus. In order to achieve this, the following steps are taken:

a. Feature types names of the original thesaurus are reused in the ontology for the concepts names.
b. No new concepts are introduced.
c. All feature types are named in singular form.

Regarding the semantic conversion process, the implicit semantics of a thesaurus has to be made explicit and interpreted in terms of the new representation format. The GeoNames gazetteer defines hierarchy for each type based on "Part-Of" relationship. The "Part-Of" relationship is equivalent to spatial containment as explained by (Janee, 2006). The ontology will interpret this relationship in the same way. To reduce the storage expense for ontology, we restrict "Part-Of" relationship only encoding geographic feature types which are directly related to the concerned feature type, rather than the whole hierarchies. For example, for "Park", the "Part-Of" relationship just encodes "County", and for "County" the "Part-Of" relationship encodes "State" and so on, rather than all the parents. The conceptual design of "Geographic Feature Type Ontology" is shown in Figure: 4.4

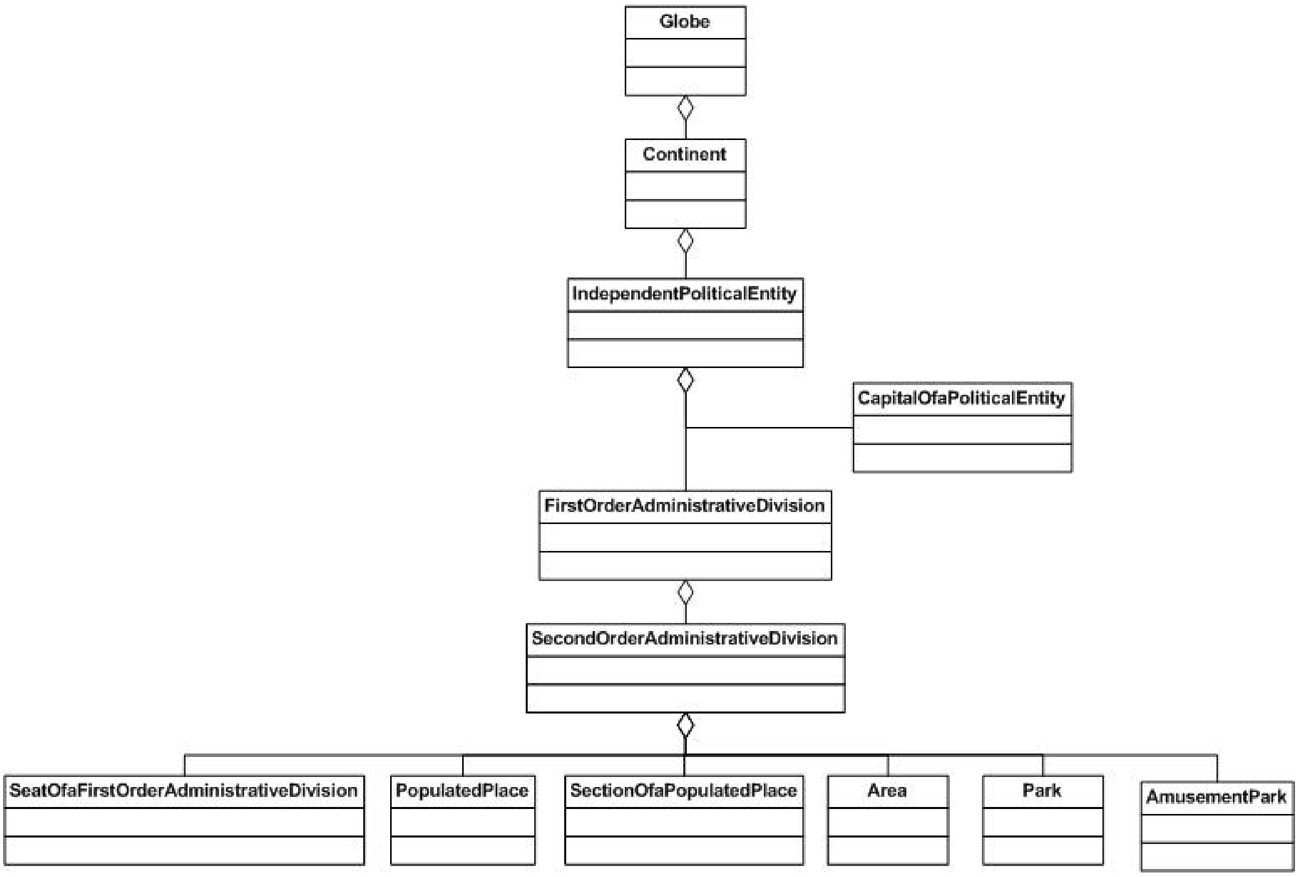

Globe
Continent
IndependentPoliticalEntity
CapitalOfaPoliticalEntity
FirstOrderAdministrativeDivision
SecondOrderAdministrativeDivision
SeatOfaFirstOrderAdministrativeDivision
PopulatedPlace
SectionOfaPopulatedPlace
Area
Park
AmusementPark

Figure 4.4: Geographic Feature Type Ontology conceptual design

The higher level of the ontology is “Globe” which is the super type. It doesn’t have a feature type or feature code defined in the gazetteer. The second level is the “Continent”, followed by “Independent Political Entity” which is the “Country” in the research Case, “First Order Administrative Division” which is the “State” and “Second Order Administrative Division” which is the “County”. The lowest level contains smaller places inside “Counties”. As explained above the aggregation relationship represents the “Part-Of” relationship in the ontology with a one-to-many relationship, and each level is related directly to the higher level.

Second component: Geographic Feature Ontology

The geographic feature ontology contains the concrete feature instances, and will encode:

- One and only one Standard-Name, which specifies a name by which a geographical feature is known. Place (or feature) names are often semantically ambiguous (Hyvonen, Lindroos, Kauppinen, & Henriksson, 2007). For example there are 18 places in “USA” having the same name “Texas”, but they have different feature types (e.g. first-order administrative division, populated place, etc.), and different hierarchy (i.e. in different counties, states). In order to disambiguate the places a suffix of the county code and state code will be added to the place names in the ontology.
- One feature type, as defined in geographic feature type ontology.
- One Spatial Relationship “Part-Of”, representing how a geographic feature is related to other geographic features (child-parent) relationship.

4.4. Implementation Issues

A step in implementing ontology of place is to select a language to express the ontology. We will focus on the Web Ontology Language (OWL-DL). OWL is a well established standard defined by the W3 Consortium, it is built on top of both Extensible Markup Language (XML) and description logic (DL), and thus it is compatible with existing Web standards and at the same time retain the formal semantics and reasoning services provided by DLs. Moreover, OWL is used by most popular ontology editors (e.g. Protégé), and most DL-reasoners (e.g. Fact++) that support subsumption reasoning for OWL-DL (Janowicz & Keßler, 2008). For these reasons OWL-DL is preferred to be used in this research.

On the other hand, OWL ontology language has limitations for representing Place ontologies as discussed by (Abdelmoty et al., 2007)

a. OWL's flat file XML representation can be insufficient when dealing with large geometric data sets.
b. In general, OWL and RDF(S) do not support all the necessary semantics for processing geo-spatial data.
c. OWL's first order, open world semantics in combination with the non unique name assumption is not suitable for constraint checking. Extensions to OWL have been proposed to overcome this limitation, for example by translating subsets of OWL to a logic program that assumes both unique name and closed world assumptions.
d. OWL can't be used to represent inference patterns of the form $\forall x, y, c : rel_1(x, y) \wedge rel_2(y, c) \rightarrow rel_3(x, c)$, so called triangular knowledge. This is a typical form of a spatial reasoning rule for composition of spatial relationships.
e. OWL does not support spatial data types. This leads to a poor representation of geometric objects using generic class and property constructs with potentially high storage overheads.
f. OWL does not support geometric processing, computation or spatial indexing and hence it is difficult to perform simple computations over geometries, such as, area or distance.

Nevertheless, it is proposed here that for many practical purposes, detailed geometric data are not necessary, as well as not being desirable, since the approach is built on place names and their “Part-Of” spatial relationship. If the ontology has to include more relationships and/or geometric data, these limitations have to be taken into consideration to overcome their consequences.

Another problem that is faced during implementation is the limitation of the ontology editor tool used (protégé), when the number of instances exceeds a certain threshold, the tool encounters unexpected errors, and doesn’t perform as expected.

The prototype included only one country and sample of place names that are retrieved from tagging system. To include all place names in a country, it wouldn’t be possible to process the ontology as one file.

4.5. Conclusion

In this chapter we used the definition of ontology to be “A logical theory which gives an explicit, partial account of a conceptualization” where conceptualization is an intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality. Ontology of place is encoding of geographic places which includes encoding place names, spatial relationships, and geographic feature types. We argued that UML can be used in conceptually designing ontology. Ontology of place is designed as aggregation of two ontologies. The first is “Geographic feature type ontology” which encodes feature types of the gazetteer service used. We showed the relationship between the feature types that are going to be used in constructing the ontology and explained their definition. We also discussed the methodology to convert thesauri into RDF and OWL ontologies. The second ontology encodes the geographic features. Finally we discussed the limitation of OWL in implementing ontology of place. The implementation of the “Geographic Feature Ontology” will use data from GeoNames in “United States of America”, resulting from querying the gazetteer with tags extracted from Flickr tagging system. The procedure of implementation and resulting ontology of place will be discussed in the next chapter.

5. Formalizing Ontology of Place

The aim of this research is to develop a bottom-up automated approach to add semantic structure to folksonomy. The approach is based on acquiring the dynamic knowledge provided by folksonomy to formalize ontology of place. This ontology will be used to discover and retrieve information about objects tagged with place names in folksonomy. To achieve this aim, we acquire a dataset of geotags from Flickr. Next, a gazetteer service is used to identify place names from these tags and extract their spatial information. Finally, Ontology of place is formalized using the place names and their "Part-Of" (parent-child) relationship. Each of these steps is automated by developing code. This code is generic and available for reuse in further research, which is part of our contribution to the field. Figure 5.1 shows the approach workflow. In the rest of this chapter we will explain each step and define its function.

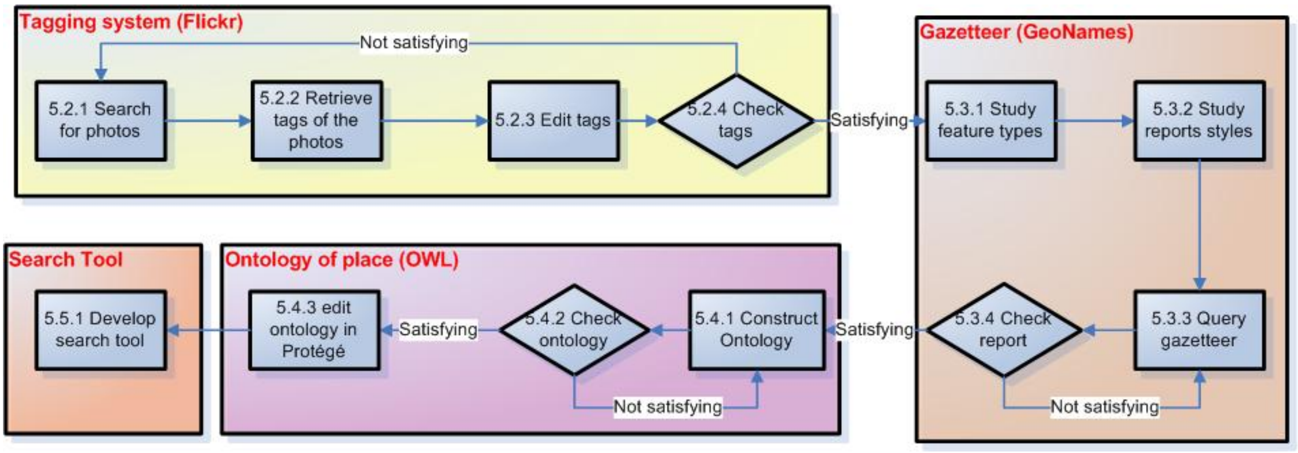


Figure 5.1: Approach workflow

The prototype runs in three stages. First is the tagging system where Flickr service is used, second is digital online gazetteer service where GeoNames service is used, and finally ontology where OWL is used within protégé editor tool. In order to request services and run queries from these three environments "Java" programming is used (for more information, see Java classes on attached CD). In the following, we discuss the steps taken in each stage, and define its procedures and results.

5.1. Selecting Country for case study

To restrict the prototype, a case study country had to be selected. The selection criteria are:

a- To retrieve data from tagging system it requires a bounding box, so a country with rectangular shape, or can be bounded by a rectangle is preferred.

b- The administrative hierarchy for each place will be extracted from the gazetteer, which implies that the selected country has to have a clear administrative hierarchy.

c- The gazetteer data is not complete for all countries of the world.

Due to these factors the "United States of America" (USA) was chosen for the prototype:

a- Its upper and lower boundaries are almost straight lines, and right and left boundaries are oceans, i.e. its boundaries can be bounded by a rectangle.

b- USA administrative hierarchy, as described in the CIA fact book[12], is fifty (50) states and one (1) district (District of Columbia), each state is divided into number of counties. This hierarchy matches the GeoNames gazetteer results of search, as it returns two administrative levels for each searched place.

c- GeoNames has complete geographic names for USA acquired from U.S. Geological Survey Geographic Names Information System (GNIS).

5.2. Tagging system (Flickr)

Flickr has an open Application Programming Interface (API) available for non-commercial use by all developers. This means that anyone can write their own program to present or use public Flickr data (like photos, video, tags, profiles or groups) in different ways. To use the Flickr API the user needs to have an application key. This is used to track API usage. We use Flickr to acquire tags added by users to their photos. To acquire tags for a certain country, two APIs have to be used in two successive steps. First, search for photos of the country. Second, retrieve tags of each photo.

5.2.1. Search for photos

To search for photos in Flickr "flickr.photos.search" method is used. This method returns a list of photos matching some criteria. If no limiting factor is passed it returns only photos added in the last twelve hours. Only photos visible to the calling user will be returned. To return private or semi-private photos, the caller must be authenticated with 'read' permissions, and have permission to view the photos. Unauthenticated calls will only return public photos. The API has thirty-two optional parameters; search for photos can be restricted by any of them. The request used is:

```
"http://api.flickr.com/services/rest/?method=flickr.photos.search
&api_key=a2a40d5fd4847950779a4a8e2fa1a183&min_upload_date=1229234400
&max_upload_date=1229320800&bbox=-122%2C30%2C-55%2C47&per_page=500&page="+i
```

Code snippet 5.1: Request to Search for photos

The parameters used

min_upload_date: Minimum upload date. Photos with an upload date greater than or equal to this value is returned. The date should be in the form of a unix timestamp. (`1229234400 is equivalent to 14 December 2008`)

max_upload_date: Maximum upload date. Photos with an upload date less than or equal to this value is returned. The date should be in the form of a unix timestamp. (`1229493600 is equivalent to 17 December 2008`)

bbox: A comma-delimited list of 4 values defining the Bounding Box of the area that will be searched. The 4 values represent the bottom-left corner of the box and the top-right corner, minimum_longitude, minimum_latitude, maximum_longitude, maximum_latitude. Longitude has a range of -180 to 180 , latitude of -90 to 90. The bounding box used is (-122, 30, -55, 47)

[12] https://www.cia.gov/library/publications/the-world-factbook/geos/us.html

per_page: Number of photos to return per page. If this argument is omitted, it defaults to 100. The maximum allowed value is 500

page: The page of results to return. If this argument is omitted, it defaults to 1. The code uses a loop to save all result pages.

Result sample

```
<photo id="2969265613" owner="51757078@N00" secret="d6824e6fe6" server="3146"
farm="4" title="Denver Union Station" ispublic="1" isfriend="0" isfamily="0" />
```

Result 5.1: one photo record returned from the search

The returned results of the search are records of photos that are geotagged inside the bounding box and match the criteria described above in the parameters. 12,000 photos were retrieved, (Result 5.1) is a sample of the photo data returned by the search. The information that is going to be reused from this result is the "id" to retrieve the tags of this photo in the following step.

5.2.2. Retrieve tags of the photos

To retrieve tags "flickr.tags.getListPhoto" method is used, this gets the tag list for a given photo. The only parameter required is the photo id.

```
"http://api.flickr.com/services/rest/?method=flickr.tags.getListPhoto
&api_key=a2a40d5fd4847950779a4a8e2fa1a183&photo_id="+value
```

Code snippet 5.2: Request to get tags list for a photo

Value: is the photo id from the result of the previous method. The retrieved tag list is as follows

Result sample

```
<photo id="2969265613">
   <tags>
      <tag id="8858729-2969271347-17017" author="8890868@N04" authorname="lbealsjr"
           raw="Photowalk" machine_tag="0">photowalk</tag>
      <tag id="8858729-2969271347-5890" author="8890868@N04" authorname="lbealsjr"
           raw="Military" machine_tag="0">military</tag>
      <tag id="8858729-2969271347-18296" author="8890868@N04" authorname="lbealsjr"
           raw="Colonial Williamsburg" machine_tag="0">colonialwilliamsburg</tag>
      <tag id="8858729-2969271347-2159" author="8890868@N04" authorname="lbealsjr"
           raw="Virginia" machine_tag="0">virginia</tag>
      <tag id="8858729-2969271347-4074" author="8890868@N04" authorname="lbealsjr"
           raw="United States" machine_tag="0">unitedstates</tag>
   </tags>
</photo>
```

Result 5.2: Tags list of one of the photos

The returned result contains information about the user (id and name). Tags information returned is "raw" which is the raw version of the tag as entered by the user, it might contain spaces and punctuation (e.g. Colonial Williamsburg). The tag body is the processed version of the tag, where spaces are removed and all the letters are in lower case (e.g. colonialwilliamsburg).

5.2.3. Edit Tags

Preprocessing for tags is required to solve some of the problems pointed out in section (3.4) these are:

- Redundant tags: due to the collective tagging behavior indicated in section (3.2) many tags are repeated a number of times in the result. These tags had to be removed and only one copy is kept, so the file contains unique tags. The original number of tags acquired was 120,000 after this process it was reduced to 11,000, which means that the number of tags was reduced to about (9 %) of its original number.
- People names: to remove people names (as Adam, Eva…etc), the tags were checked for the most popular[13] (1000) boys' names and (1000) girls' names that are used in the "United States" in 2007 and 2008. Any tag that had these names was removed. Despite that some places are named after people's names, adding these names to the ontology would cause distortion and we opted for removing them. This is while acknowledging the loss of knowledge that might result. Our estimate however, is that the gain in precision and recall outweighs this loss as discussed in the last chapter.
- Nouns: some tags were nouns as (me, music, sun…etc). The tags were checked for the most used (1000) English words[14]. If a tag is equal to (not contain) any of these words, the tag was removed.

5.2.4. Check Tags

A code is developed to count the number of tags in the XML file, after editing the tags; the number of retrieved tags might be too small. As explained in the previous step after removing the redundant tags, the total number of tags dropped to 11,000. And after removing the popular names and popular English words the number of tags dropped again to 8,000. As mentioned in the introduction section 1.1, the statistics of tags showed that only 25 % of the tags are expected to be proper place names, which mean that at this step the expected number of place name tags is about 2,000 this would result in poor coverage of the US. In this case another iteration of acquiring tags is taken.

Result sample

```
<photo id="3153737873">
   <tags>
       <tag author="7133314@N03" authorname="txcraig75" id="7110260-3153737873-
        15306" machine_tag="0" raw="National Park">nationalpark</tag>
       <tag author="7133314@N03" authorname="txcraig75" id="7110260-3153737873-
        2407" machine_tag="0" raw="Washington">washington</tag>
       <tag author="7133314@N03" authorname="txcraig75" id="7110260-3153737873-
        3605365" machine_tag="0" raw="Canon 17-40mm">canon1740mm</tag>
   </tags>
</photo>
```

Result 5.3: Sample from resulting XML file with unique tags

[13] http://www.behindthename.com/top/lists/1000us2007.php

At the end of this stage we acquire and XML file that contains about 15,000 unique tags as shown in Result 5.3. The information in this file is user information and tag information, same as the information explained in Result 5.2. In the next stage 15,000 queries to the gazetteer service will be done using a developed java code. Each query will use one of the tags' "raw" value as a place name to query the gazetteer. The queries will return the place names spatial information that will be used in constructing the ontology as explained in next section.

5.3. Querying Gazetteer (GeoNames)

A Gazetteer service is used to extract tags that are place names and their relations. To accomplish this we developed java code to use the GeoNames search API. The tags that are place names will return a search result and those that are not will have zero result. For the place names, spatial information is retrieved from GeoNames. This information includes location (lat, lng) coordinates, administrative hierarchy (parents) of the place, and the feature type of the place.

5.3.1. Study of Feature Types

GeoNames organize geographic feature types (645 in total) in nine groups (Feature classes), as indicated in section (2.4). In the following we address remarks about these feature types:

a- Some feature types are mentioned more than once, one time as singular and another time as plural (e.g. Island & Islands, pyramid & pyramids) and in both cases the code has the same definition. While, other feature codes are mentioned once containing its plural (e.g. Quarry(-ies), wharf(-ves))

b- Some special cases are mentioned as feature types e.g. "Israeli settlement" which is a specific location. Also "postgrad&MBA" is mentioned as type. It is not clear how a location will be classified under this type and a search through the gazetteer documentation did not provide an answer to this question.

c- Some similar feature codes are mentioned under different feature classes, e.g. "Port" is a feature type in "Parks, area, …" feature class, while airport is a feature type in "spot, building, farm" feature class

The feature types used in the prototype are chosen according to the criteria mentioned in section 4.4.2.

5.3.2. Gazetteer reports styles

The GeoNames gazetteer service offers four styles of reports "Short", "Medium", "Long", "Full" each containing more information than the previous one:

a- "Short" report contains:

- name
- lat: latitude
- lng: longitude
- geonameId
- Country Code
- Fcl: Feature class
- Fcode: Feature code

[14] http://www.duboislc.org/EducationWatch/First100Words.html

b- “Medium” report contains all what exists in “Short” report and exceeds it by:

- Country name

c- “Long” report contains all what exists in “Medium” report and exceeds it by:

- fclName: Feature class name
- fcodeName: Feature code name
- Population: population of the place

d- “Full” report contains all what exists in “Long” report and exceeds it by:

- alternateNames: alternative names for the place in different languages.
- Elevation
- continentCode: Continent code
- adminCode1: administrative level one, code (e.g. state)
- adminName1: administrative level one, name
- adminCode2: administrative level two, code (e.g. county)
- adminName2: administrative level two, name
- alternateName lang: alternative names for the place in different languages, and the language used.

In the prototype the “Full” report style is used as it contains the correct level of hierarchy of the searched places.

5.3.3. Query Gazetteer

For each tag the gazetteer service is searched to return the geographic information and administrative hierarchy of the place. To accomplish this a code is developed that uses the gazetteer API to automatically run the queries for the 15,000 tags resulted from acquiring tags from Flickr as explained in section 5.2.4

```
"http://ws.geonames.org/search?country=US&style=FULL&featureClass=P&featureClass=A
&featureClass=L&name_equals="+value
```

Code snippet 5.3: gazetteer search request to retrieve geographic information of place names

The code in 5.3 includes: the country name to search for places inside (here we used “US”), the report style (full), the feature classes (P: city, village, etc. it includes 11 feature types, A: country, state, region, etc. it includes 16 feature types, L: parks,area, ... it includes 49 feature types). Finally “name_equals” to search for exact place name, and value is the place name to search for as read from the tags XML file.

Result sample

```
<totalResultsCount>11</totalResultsCount>
     <geoname>
        <name>Manhattan</name>
        <lat>40.7834345</lat>
        <lng>-73.9662495</lng>
        <geonameId>5125771</geonameId>
        <countryCode>US</countryCode>
        <countryName>United States</countryName>
        <fcl>P</fcl>
        <fcode>PPLX</fcode>
        <fclName>city, village,...</fclName>
        <fcodeName>populated place</fcodeName>
        <population/>
        <alternateNames/>
        <elevation>35</elevation>
        <continentCode>NA</continentCode>
        <adminCode1>NY</adminCode1>
        <adminName1>New York</adminName1>
        <adminCode2>061</adminCode2>
        <adminName2>New York County</adminName2>
        <timezone dstOffset="-4.0" gmtOffset="-5.0">America/New_York</timezone>
</geoname>
```

Result 5.4: search result retrieved from gazetteer for a place name

For example, the gazetteer search for “Manhattan” returned 11 places, one of them is shown in (Result 5-2). The information of the place is displayed in “Full” report style as discussed in section (5.3.2). “Manhattan” has the “populated place” feature type. And its administrative hierarchy is “New York County”, “New York” state, USA, North America, and Globe.

5.3.4. Check report

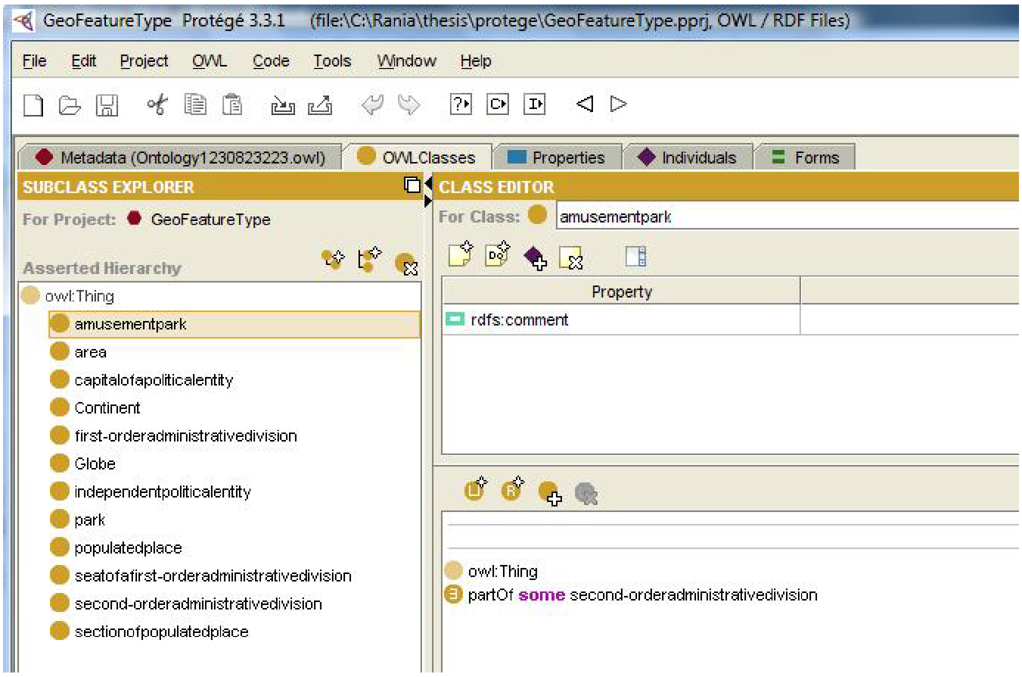
The report resulting from querying the gazetteer for all place name tags is saved in XML file. Internet browser is used to check the file in semi-automated process to ensure that there was no failure in the service while processing the search. If there is a problem with the report the gazetteer is queried again.

5.4. Ontology of place (OWL)

The result of searching the gazetteer contains the place names, their administrative hierarchy, and implicit “Part-Of” relationships. This information has to be formalized in ontology of place that explicitly defines the place names and their relationships. Accomplishing this step would allow us to use the resulting ontology in enhancing information retrieval in folksonomy. The ontology will contain the place name tags in structured form with explicit relationships. The following steps were taken to formalize the required ontology.

5.4.1. Constructing the ontology

Two ontologies have to be constructed according to the conceptual design discussed in section 4.5.2. The first ontology “geographic feature type” ontology is constructed using protégé tool according to the conceptual design in figure 4.4 and the “Part-Of” relationship is defined as restrictions between classes.


GeoFeatureType Protégé 3.3.1 (file:\C:\Rania\thesis\protege\GeoFeatureType.pprj, OWL / RDF Files)
File
Edit
Project
OWL
Code
Tools
Window
Help
Metadata (Ontology1230823223.owl)
OWLClasses
Properties
Individuals
Forms
SUBCLASS EXPLORER
CLASS EDITOR
Asserted Hierarchy
Property

For Project: GeoFeatureType

- owl:Thing
  - amusementpark
  - area
  - capitalofapoliticalentity
  - Continent
  - first-orderadministrativedivision
  - Globe
  - independentpoliticalentity
  - park
  - populatedplace
  - seatofafirst-orderadministrativedivision
  - second-orderadministrativedivision
  - sectionofpopulatedplace

For Class: amusementpark

rdfs:comment

owl:Thing

partOf **some** second-orderadministrativedivision

Figure 5.2: Geographic Feature Type ontology in protégé

The second ontology “geographic feature” ontology is constructed from the gazetteer result. This is achieved by developing a tool coded in java that changes the format and structure of the XML file into an OWL file with ontology structure. The tool is available through the work developed in this research and can be used to construct an ontology of place from any XML file that is structured as the XML file resulting from GeoNames gazetteer.

```
if (fCodeName.equalsIgnoreCase("first-orderadministrativedivision")){
             conName= adminName1+"_"+adminCode1;
             conParent= "US";
      }
else if (fCodeName.equalsIgnoreCase("second-orderadministrativedivision")) {
           conName = adminName2+"_"+adminCode2+"_"+adminCode1;
            conParent = adminName1+"_"+adminCode1;
      }
else {      conName = fName+"_"+adminCode2+"_"+adminCode1;
            conParent = adminName2+"_"+adminCode2+"_"+adminCode1;
                                 }
bw.write("<owl:Class rdf:ID="+'"'+conName+'"'+">");
bw.newLine();
bw.write("<rdfs:subClassOf>");
bw.newLine();
bw.write("<owl:Restriction>");
bw.newLine();
bw.write("<owl:someValuesFrom rdf:resource="+'"'+"#"+conParent+'"'+"/>");
bw.newLine();
bw.write("<owl:onProperty rdf:resource="+'"'+"http://www.owl-
ontologies.com/Ontology1230823223.owl#partOf"+'"'+"/>");
bw.newLine();
bw.write("</owl:Restriction>");
bw.newLine();
bw.write("</rdfs:subClassOf>");
bw.newLine();
bw.write("<rdfs:subClassOf rdf:resource="+'"'+"http://www.owl-
ontologies.com/Ontology1230823223.owl#"+fCodeName+'"'+"/>");
bw.newLine();
bw.write("</owl:Class>");
```

Code snippet 5.4: constructing the geographic feature ontology

The “fCodeName” is the feature type name, “adminName1” and “adminCode1” are the name and code of the first order administrative level (state in this case study) successively. “adminName2” and “adminCode2” are the name and code of the second order administrative level (county in this case study). “fName” is the feature name. All the information is extracted from the gazetteer search result file.

The tool checks for the feature type of the feature to be added if it is of first-order (i.e. state), the name of the feature is composed of two parts only, the feature name and its code (e.g. NewYork_NY) and the parent is set to be “US”. If the feature is of second-order (i.e. county) the feature name is three parts: name, county code, and state code (e.g. NewYorkCounty_061_NY) and the parent is set to “state name_state code” (i.e. NewYork_NY). For the rest of the feature types the feature name is three parts: name, county code, and state code (e.g. Manhattan_061_NY) and the parent is set to “countyName_countyCode_stateCode” (e.g NewYorkCounty_061_NY).

Result sample

```
<owl:Class rdf:ID=" Manhattan_061_NY">
     <rdfs:subClassOf>
         <owl:Restriction>
            <owl:someValuesFrom rdf:resource="#NewYorkCounty_061_NY"/>
            <owl:onProperty rdf:resource="http://www.owl-
              ontologies.com/Ontology1230823223.owl#partOf"/>
         </owl:Restriction>
     </rdfs:subClassOf>
     <rdfs:subClassOf rdf:resource="http://www.owl-
       ontologies.com/Ontology1230823223.owl#populatedplace"/>
</owl:Class>
```

Result 5.5: OWL class constructed using the developed tool

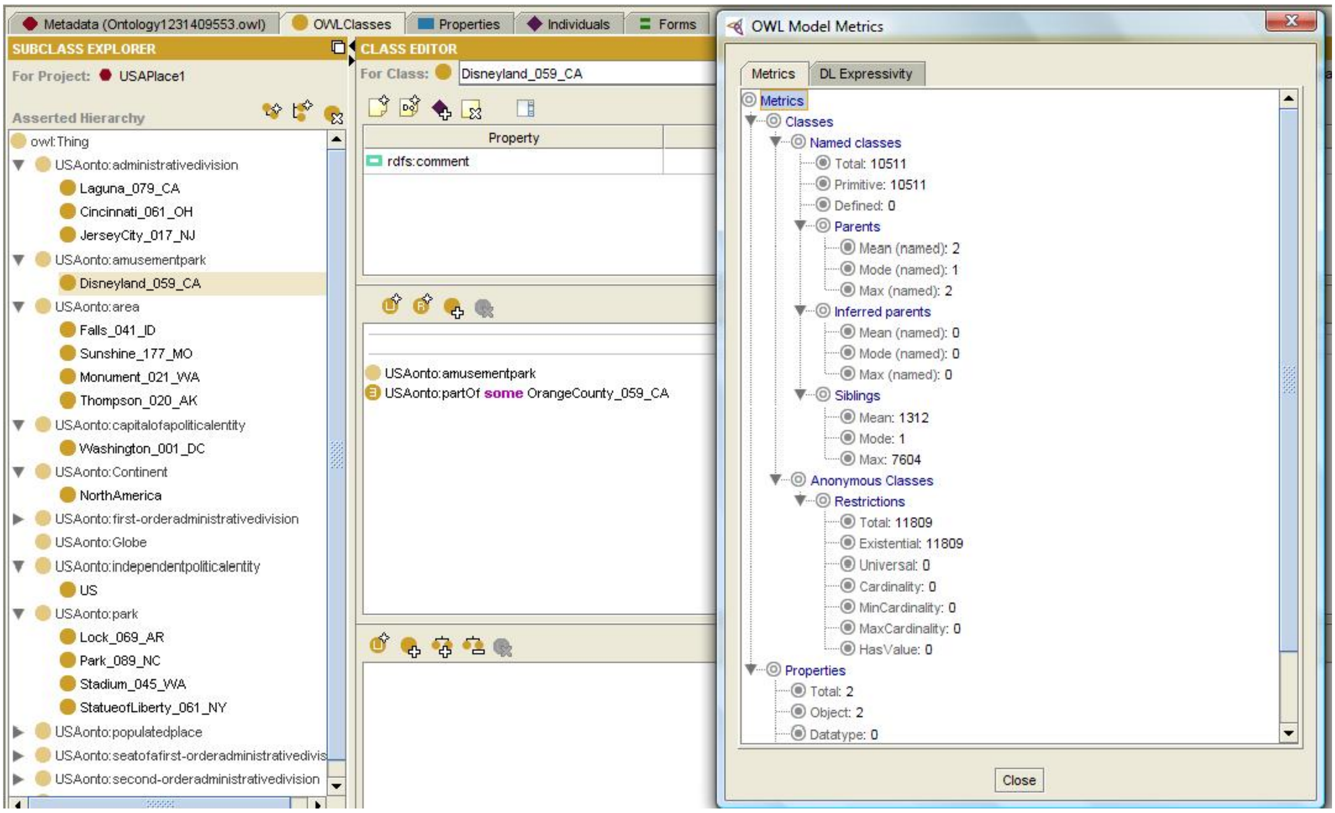
The result of all places is stored in OWL file that imports the “Geographic feature type” ontology and restricts each place as “Part-Of” its feature type, and “Part-Of” it’s direct parent administrative level.

5.4.2. Checking the ontology

The resulting ontology was checked in a text editor to add the OWL and rdf header. It was also checked in protégé to assure the classes, subclasses and their restriction relationships are in the proper syntax and the needed ontology consistency checks were also done. In the first iteration of constructing the ontology only the state code was added to the places assuming that place names are unique in each state which is not the case, this resulted in one place having different parents. So we had to add the county code as part of the place name. Moreover, the county code is unique for each state, but it might exist again in other state.

5.4.3. Edit ontology in protégé

As a final step the “Place ontology” is checked in protégé and was tested by searching for some random places and examines the result. As shown in (figure 5.3) the “Place ontology” contains more than ten thousand (10,000) places.



For Project: USAPlace1

Asserted Hierarchy

- owl:Thing
  - USAonto:administrativedivision
    - Laguna_079_CA
    - Cincinnati_061_OH
    - JerseyCity_017_NJ
  - USAonto:amusementpark
    - Disneyland_059_CA
  - USAonto:area
    - Falls_041_ID
    - Sunshine_177_MO
    - Monument_021_WA
    - Thompson_020_AK
  - USAonto:capitalofapoliticalentity
    - Washington_001_DC
  - USAonto:Continent
    - NorthAmerica
  - USAonto:first-orderadministrativedivision
  - USAonto:Globe
  - USAonto:independentpoliticalentity
    - US
  - USAonto:park
    - Lock_069_AR
    - Park_089_NC
    - Stadium_045_WA
    - StatueofLiberty_061_NY
  - USAonto:populatedplace
  - USAonto:seatofafirst-orderadministrativedivis
  - USAonto:second-orderadministrativedivision

For Class: Disneyland_059_CA

Property
rdfs:comment

USAonto:amusementpark
USAonto:partOf **some** OrangeCounty_059_CA

- Metrics
  - Classes
    - Named classes
      - Total: 10511
      - Primitive: 10511
      - Defined: 0
      - Parents
        - Mean (named): 2
        - Mode (named): 1
        - Max (named): 2
      - Inferred parents
        - Mean (named): 0
        - Mode (named): 0
        - Max (named): 0
      - Siblings
        - Mean: 1312
        - Mode: 1
        - Max: 7604
    - Anonymous Classes
      - Restrictions
        - Total: 11809
        - Existential: 11809
        - Universal: 0
        - Cardinality: 0
        - MinCardinality: 0
        - MaxCardinality: 0
        - HasValue: 0
  - Properties
    - Total: 2
    - Object: 2
    - Datatype: 0

Figure 5.3: Ontology of place in protégé

5.5. Conclusion

In this chapter we designed and implemented our research approach to enrich folksonomy. It is a bottom-up semi-automated approach to formalize ontology of place that will be used in folksonomy to add semantic richness. In the different steps of the approach we developed tools coded in java, to automate each step. This code is reusable and available on the CD attached with this research.

The ontology of place is constructed from place name tags acquired from Flickr, these tags go through a process of selection to remove the tags that might cause distortion to the ontology (as redundant tags, people's names, and English nouns). A gazetteer service, GeoNames is used to select, disambiguate, and retrieve spatial relations between places.

Finally, two ontologies are constructed. The first is a "Geographic feature type ontology" that is constructed using OWL within protégé tool, and the second is a "Geographic feature ontology". A tool is developed to automatically construct the second ontology from the result of the gazetteer. We consider this tool part of the research contribution to the field as it is reusable and automates the process. Protégé is used to construct the final result by integrating both ontologies. The resulting ontology of place contains more than 10,000 classes with the "Part-Of" relationship defined between them based on their administrative hierarchy. The ontology is checked by querying it for some place names and the results are verified by searching the web for those places. In the different stages of the approach, the different environments used are illustrated e.g. Flickr APIs and interface were described. Moreover, GeoNames APIs, reports styles, and feature types were studied and their usage illustrated.

6. Prototype of a Search Tool

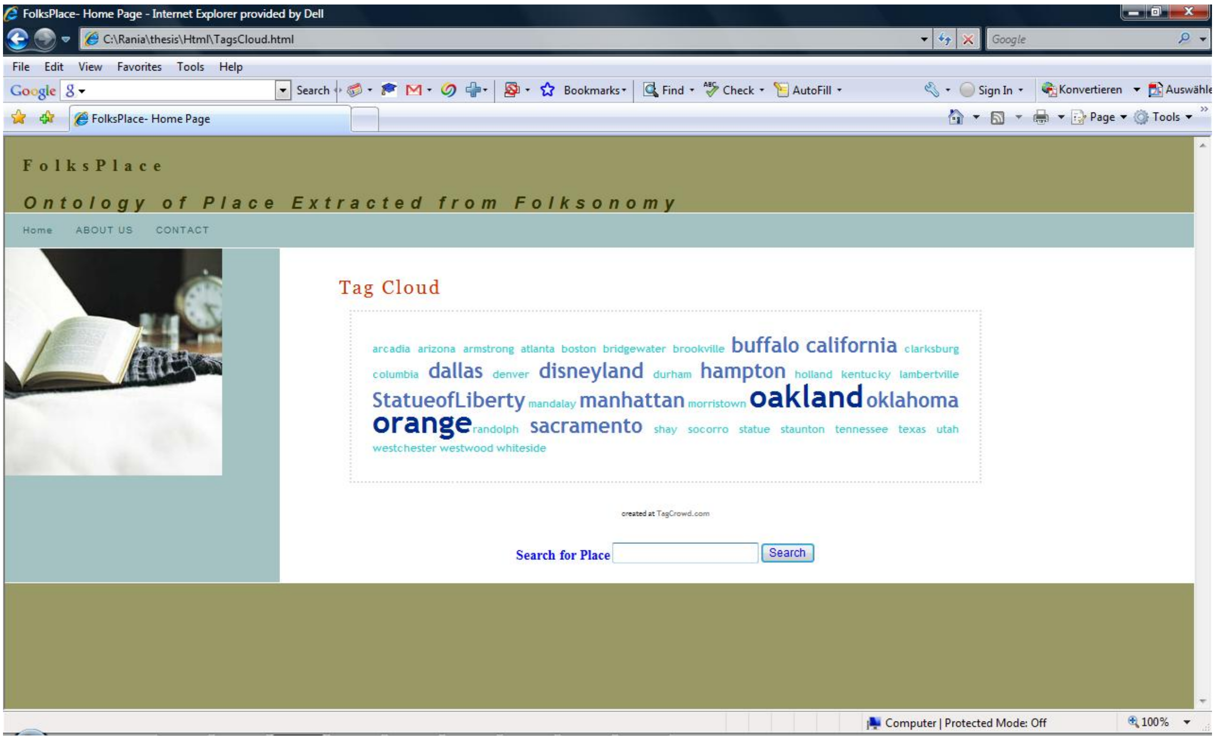
In the previous chapter we formalized ontology of place to enhance information retrieval in folksonomy. The resulting ontology is the integration of geographic feature type ontology and geographic features ontology. It contains more than 10,000 places structured as subclasses of their feature types, with explicit "Part-Of" (parent-child) spatial relationship defined between each place and its parent place (parent in the administrative hierarchy). The aim of this ontology is to enhance information discovery and retrieval by improving the system's precision and recall. In this chapter we develop a prototype that will integrate the ontology with folksonomy application. The prototype interface and functionalities are discussed, and the added benefits are described. Finally we design and run an evaluation methodology that examines the ontology of place effectiveness in retrieving the information.

6.1. Developing a search tool for ontology enhanced folksonomy

The search tool developed is associated with a web browser, to retrieve information about places from the folksonomy enriched with the ontology of place. Searching for a place using the tool, would display to the user a result set of places matching the searched place name. For each place, its name, feature type, and parents are also displayed. This information is enough to disambiguate the required place. Upon choosing one of the places from the result set Flickr is searched using the place name and its parent name and the result is displayed to the user in a new window with suggestions of related places. In the following sections we describe the developed prototype and show the screenshots of each step.

6.1.1. Search for a place

In a real world scenario the user, uses folksonomy (Flickr) interface for search. A recommended step is that the user chooses if she/he is looking for a keyword or a place. Choosing to search for a place would display the developed search tool that searches the ontology of place. The search tool interface displays a tag cloud of the most popular places extracted from the ontology, and a text area to specify a place name for search. Figure 6.1 shows the search tool interface to enter a place name.


FolksPlace- Home Page - Internet Explorer provided by Dell
C:\Rania\thesis\Html\TagsCloud.html
Google
File Edit View Favorites Tools Help
Search
Bookmarks
Find
Check
AutoFill
Sign In
Konvertieren
Auswähle
FolksPlace- Home Page
Page
Tools
Computer | Protected Mode: Off
100%

F o l k s P l a c e

Ontology of Place Extracted from Folksonomy

Tag Cloud

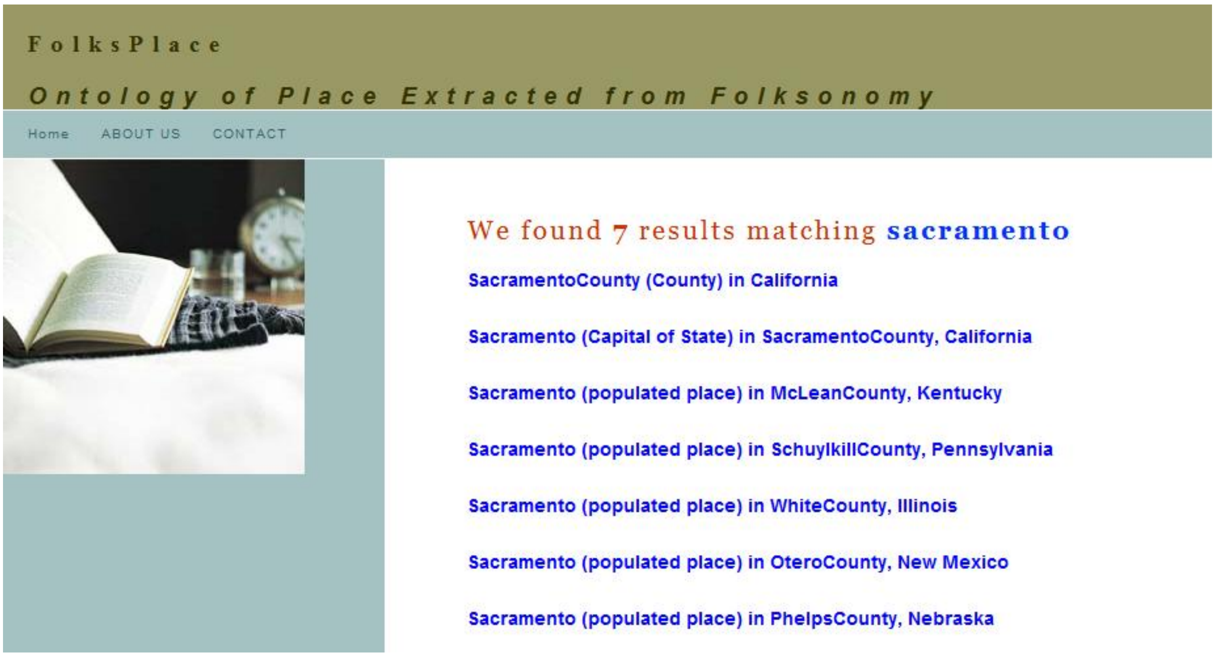
arcadia arizona armstrong atlanta boston bridgewater brookville buffalo california clarksburg columbia dallas denver disneyland durham hampton holland kentucky lambertville StatueofLiberty mandalay manhattan morristown oakland oklahoma orange randolph sacramento shay socorro statue staunton tennessee texas utah westchester westwood whiteside

created at TagCrowd.com

Search for Place

Figure 6.1: Search tool interface to enter a place to search for

Upon entering a place name the tool searches the ontology of place, and retrieves all places that contain the searched name, the results are displayed as a list. The results interface is simple and easy to understand, where each result displays the place name, feature type, county and the state it is part of (see figure 6.2). Each result from the ontology is a hyperlink that searches Flickr for the place name and state name (e.g. SacramentoCounty and California).



F o l k s P l a c e

Ontology of Place Extracted from Folksonomy

We found 7 results matching sacramento

SacramentoCounty (County) in California

Sacramento (Capital of State) in SacramentoCounty, California

Sacramento (populated place) in McLeanCounty, Kentucky

Sacramento (populated place) in SchuylkillCounty, Pennsylvania

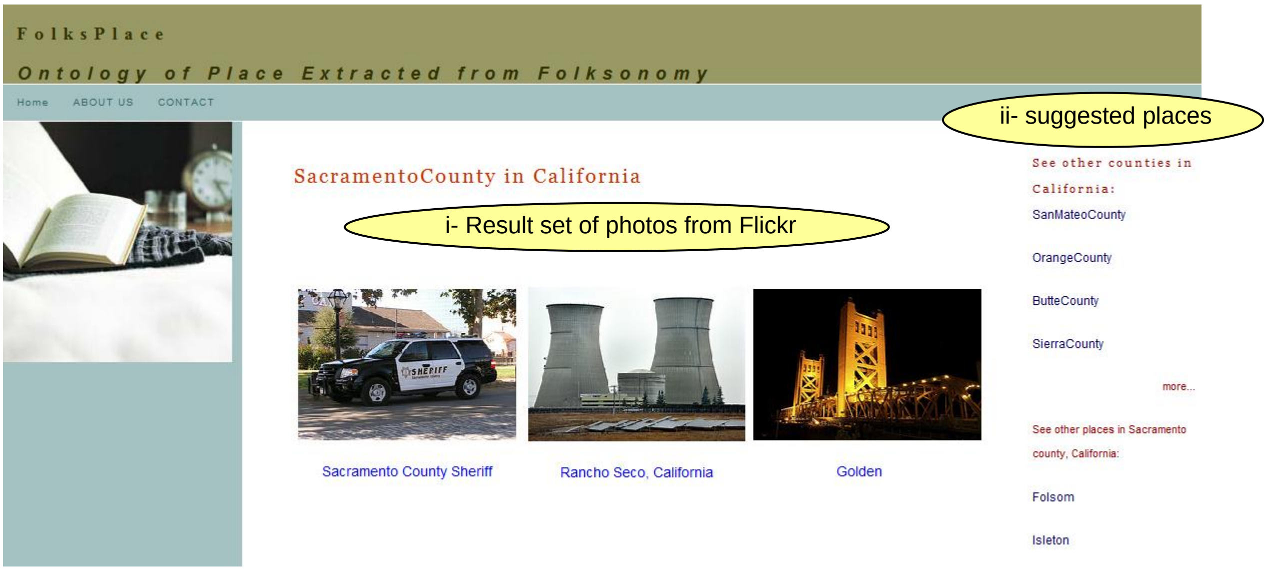
Sacramento (populated place) in WhiteCounty, Illinois

Sacramento (populated place) in OteroCounty, New Mexico

Sacramento (populated place) in PhelpsCounty, Nebraska

Figure 6.2: Result list

The search in Flickr, using its search engine, retrieves the photos tagged or described by the chosen place name and its parent name. The result is displayed in a new window as shown in figure 6.3, with the middle part of the page displaying the result set of photos, and a list of suggested places, that are related to the searched place name, on the right sidebar of the page.


FolksPlace

Ontology of Place Extracted from Folksonomy

ii- suggested places

SacramentoCounty in California

i- Result set of photos from Flickr

Sacramento County Sheriff

Rancho Seco, California

Golden

See other counties in California:

SanMateoCounty

OrangeCounty

ButteCounty

SierraCounty

more...

See other places in Sacramento county, California:

Folsom

Isleton

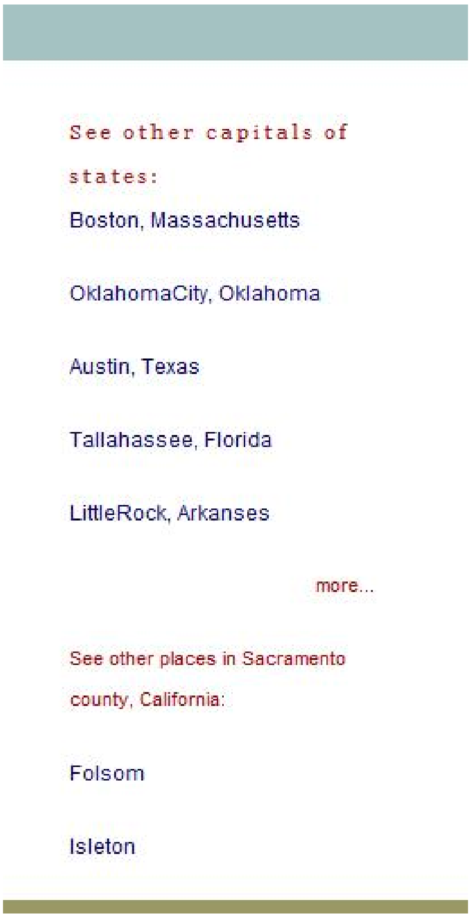
Figure 6.3: i: Result of searching for photos in SacramentoCounty and California, ii: suggestions of other places

6.1.2. Suggestions of related places

The places suggested to the user are chosen on basis that they are either semantically, or spatially, or semantically and spatially related to the searched place. The relatedness is decided according to the administrative hierarchy of the feature type of the searched place. For example, if the result is for a place of feature type “Seat of a First Order Administrative Division” (capital of state), then siblings of two types are displayed as suggestions to the user:

- Semantically related places, these are capitals of other states.
- Spatially related places, these are places of any feature type that are in the same county or state.

In figure 6.4 we show the suggested places that are displayed to the user in the right sidebar, when searching for “Sacramento” that is capital of “California” state, the first set of suggested places are capitals of other states, choosing to display “more” a list of all capitals of states is displayed (50 capitals in total). The other set is places that are part of “Sacramento” county in “California” state.


See other capitals of states:

Boston, Massachusetts

OklahomaCity, Oklahoma

Austin, Texas

Tallahassee, Florida

LittleRock, Arkanses

more...

See other places in Sacramento county, California:

Folsom

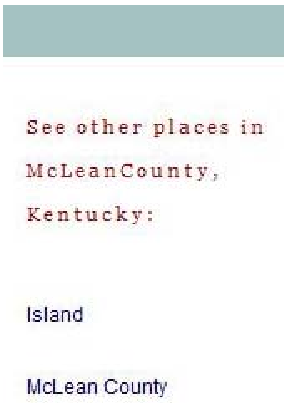
Isleton

Figure 6.4: Suggested places that are related to Sacramento, capital of California state

For another feature type that is in the lower administrative hierarchy as "populated place" the suggested places will be limited with being spatially and semantically related, as in this case the semantically related are other places of type "populated places" and these are thousands of places, that can not all be suggested to the user.

- Semantically and spatially related places: these are places with same feature type "populated place" and are part of the same county and state.

Figure 6.5 shows the places suggested to the user, that are related to a "Sacramento" populated place, in "McLeanCounty" in "Kentucky" state.



See other places in McLeanCounty, Kentucky:

Island

McLean County

Figure 6.5: suggested places that are related to Sacramento, McLean county, Kentucky state

In the following table we summarize the suggested types of places that will be presented to the user depending on the feature type of the places she/he searched for.

	Feature type	Suggested places	
		Siblings	Children
1	Continent	-	•Countries
2	Independent political entity	•Countries in the same continent	•States
3	Capital of a political entity	•States in the same country	•Places in the city
4	First-Order Administrative Division	•States in the same country	•Counties in the state
5	Second-Order Administrative Division	•Counties in the same state	•Places in the county
6	Seat Of a First Order Administrative Division	•Capitals of other states •Other places in the same county	-
7	Park	•Other parks in same state •Other places in the same county	-
8	Amusement park[15]	•Other parks in same state •Other places in the same county	-
9	Populated Place	•Populated places in same county	-
10	Section Of a Populated Place	•Section Of a Populated Place in same county	-
11	Administrative Division	•Other places in the same county	-
12	Area	•Other places in the same county	-

Table 6-1: suggested places types for each feature type

The suggested places are also hyperlinks that the user can select any of them and a result page (as in figure 6.3) will be displayed for her/him, again with suggestions of related places.

The suggestion of related places to the user is an added value to folksonomy, as the current systems do not offer such a service to the user.

[15] Parks are defined as suggestions for amusement park as this type is defined only for Disneyland in the case study.

6.2. Evaluation and Results

The idea behind this work is that the user needs to search and retrieve photos for related places that are tagged with place names, using the developed prototype she/he searches for one place name. The tool concatenates the place name with its parent name, as such the system improves precision. If the user does not know the names of all the related places, she/he will not be able to retrieve all the results. But with the suggestions list of the related places the user can retrieve results for any of these places improving the system recall. In this section we evaluate the "ontology of place" to examine the value it added to folksonomy, and observe the changes in the information retrieved when using the ontology. First we define an evaluation method that is based on two criteria, namely precision and recall. Then test the prototype according to these criteria. Finally we discuss the results of the test, pointing out the strong and weak points observed from the evaluation.

6.2.1. Evaluation method

The constructed ontology is intended to be used by the folksonomy search engine as shown in section 6.1, which is an information retrieval system. These systems are usually evaluated by employing two criterions precision and recall. (Salton & McGill, 1986)

Precision and recall: A system's precision is its ability to retrieve only relevant information units and decline irrelevant ones. The recall value on the other hand expresses the number of retrieved relevant items out of all existing relevant information (Salton & McGill, 1986).

6.2.2. Ontology of place testing for precision and recall

Since the set of objects (photos) that will be searched are on Flickr server and so are unknown to us, then the evaluation will be based completely on comparing results of search for places. This will be accomplished by searching Flickr two times for the same place using different keywords, as follows:

- Once by using the place name only, as a keyword (e.g. "Disneyland"), we will denote the retrieved result set in this case by set "a".
- And a second time by using the place name and its parent name extracted from the "ontology of place" as keywords (e.g. "Disneyland and California"). We will denote the retrieved result set in this case by set "b".

Theoretically, This method of search by using keywords, imply that the result set of the second search using place name and parent name (set "b"), is equal or subset of the result set of first search using place name only (set "a"), as shown in Figure 6.6. This is due to the fact that using more than one keyword will narrow the search result.

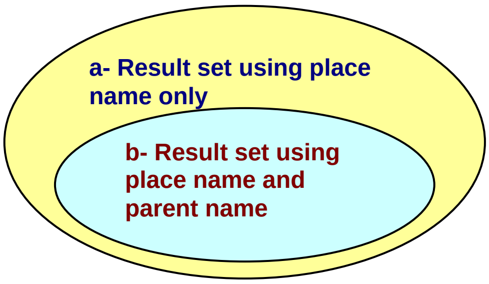

a- Result set using place name only
b- Result set using place name and parent name

Figure 6.6: Result set of search using,
a: place name only, b: using place name and parent name

In case the two sets are equal, precision and recall are neither improved nor worsened.

The other probability is that "b" is subset of "a".

$$b \subset a : \forall e[e \in b \rightarrow e \in a], \qquad (1)$$

Where “a” is the result set of searching using place name only, “b” is the result set of searching using place name and parent name, “e” is an element in “b”. The equation states that, since “b” is a proper subset of “a”, then any element that belongs to “b”, will belong also to “a”. But the inverse is not true.

Based on this, lets assume that in “b” there are some irrelevant results, these results will belong to “a” too, this means that the irrelevant results in “b” can not be more than those in “a”, this implies that the precision is either increased or not changed. On the other hand, any relevant result that belong to “b” will belong to “a” too, but the inverse is not true, this means that the recall is either decreased (i.e. some relevant results will belong to “a” and not to “b”) or not changed.

In the following we will run the test (two times for each place as explained above) using some examples from the ontology of place. For each place, each time we will manually observe and record the following:

- The total number of photos retrieved in the result set.
- The number of relevant photos from the first fifty results. Ordered by Flickr search engine in descending order of most relevant. (e.g. photos for Disneyland in California will be considered relevant)
- The number of irrelevant photos from the first fifty results. Ordered by Flickr search engine in descending order of most relevant. (e.g. photos for Disneyland in Paris will be considered irrelevant)

6.2.3. Test Results

Five diverse examples of places are chosen from the ontology of place. The diversity between the places is in their popularity and feature type; this is to study how the results change upon changing the searched place characteristics. The five places are:

- Sacramento in California state: it is a popular place, and its feature type is “capital of state”
- Sacramento in Kentucky state: it is not a popular place, and its feature type is “populated place”
- Disneyland in California state: it is a popular place, and its feature type is “Amusement Park”
- Monument in Washington State: it is a popular place as it has the Washington monument (a memorial to George Washington, first President of the United States), its feature type is “area”, and the place name “Monument” is vague and has homonym problem, as it can indicate any monument all over the world.
- Sunrise in Kentucky State: it is not a popular place, its feature type is “populated place”, and the place name has a homonym problem as it is always interpreted as the action of the sun rising from the east, not as a place.

Place name	Parent name	(a) Using place name only			(b) Using both names		
		# results	# relevant (from1st 50)	# Irrelevant (from1st 50)	# results	# relevant (from1st 50)	# Irrelevant (from1st 50)
Sacramento	California	262,584	45	5	62,810	50	0
Sacramento	Kentucky	262,584	0	50	1,082	49	1
Disneyland	California	930,767	30	20	136,344	48	2
Monument	Washington	858,604	8	42	133,433	50	0
Sunrise	Kentucky	1,193,275	0	50	1,010	0	50

Table 6.2: search result from Flickr

Table 6.2 shows the result of searching in Flickr for the five places. The first two columns are the place name, and its parent name that will be used in the search. The next three columns are the results of search using the place name only (e.g. Sacramento), (# results) indicates the total number of results retrieved by Flickr that matched the place name (photos that are tagged or described using the place name), (# relevant from 1st 50) is the number of photos that were found relevant in the first 50 results, (# irrelevant from 1st 50) is the number of photos that were found irrelevant in the first 50 results. The last three columns are the result of search using place name and parent name (e.g. Sacramento and California).

6.2.4. Results discussion

From the results displayed in (table 6.2) the following remarks can be pointed out:

1- In all cases using both names for search, result in a much smaller result set, i.e. in all cases "b" is subset of "a" and not equal.

2- In all cases the irrelevant results in using both names are less than those when using just the place name. This proves that the precision is increased. Its also worth noting that:

a. For popular place names, the difference in relevant and irrelevant results is not large between both search result sets "a" and "b" (e.g. Sacramento in California state)

b. For places that are popular for their parent name (e.g. Monument in Washington State), using the parent name disambiguated the place name, and the precision is significantly increased.

c. But for less popular place names there is large difference in the results (e.g. Sacramento in Kentucky State) which means that the precision is significantly increased.

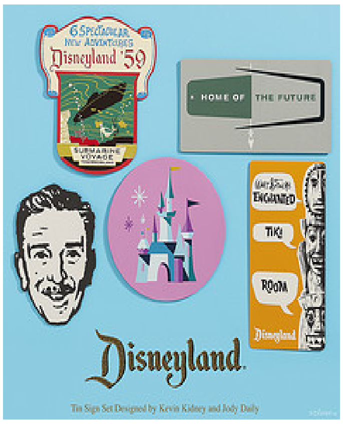


Figure 6.7: Result from search for "Disneyland"

d. In some cases, when not including the parent name, the result set included photos which are not taken in places, such as posters, designs or advertisements (e.g. Disneyland) figure 6.7 shows one of the photos in the result set for searching for Disneyland which is a poster not a photo for the place taken by user. This means that using both names restrict the search to places.

e. A contradiction to the previous point, when searching for a place which is not popular and has a name that is interpreted by the system as a different meaning, e.g. searching

3- The result set of searching for “Sacramento”, contained some photos that are tagged with “Sacramento” and “CA” (the administrative code of “California”) which is a folksonomy synonym problem, and was not retrieved in the second result set of searching for “Sacramento and California” i.e. some relevant results are missed in the second search. This proves that the recall is decreased.

6.3. Conclusion

In this chapter we designed and implemented a prototype to use ontology of place constructed from the place name tags, in retrieving information from Flickr. We showed the interface that is used to retrieve places from the ontology that matches the place name a user is searching for. Each place is described by its feature type and parents name. Each place is a hyperlink that displays a search result page containing the retrieved information from Flickr, and suggestions of related places to the place searched. The suggestions of related places are chosen based on being semantically or spatially related or both.

Next, we evaluated the effectiveness of the ontology of place in retrieving information from Flickr. The evaluation was based on two indicators precision and recall, theoretically we proved that the precision of information retrieval should be increased or not changed and the recall of information is either not changed or decreased. The results of the evaluation showed that the precision is significantly increased in most of the cases. In some rare cases due to folksonomies homonymy problem, the precision is not increased. However we expect that a full integration of the developed prototype within Flickr would improve these results as well. The results also showed that the recall is decreased in some cases, due to folksonomies synonym problem which is using different words for the same place, as tagging a photo with “CA” instead of “California” that leads to not retrieving such photo when searching for photos tagged with the place name and “California”. As discussed earlier in section 3.4.1 there is always a necessary trade-off between precision and recall, and this research makes no exception to this rule.

7. Conclusions and Future Work

The main objective of this research was to enhance information retrieval in folksonomy using ontology, to improve semantic precision. This involves integrating technologies from semantic web and social web, in an attempt to bridge the gab between them. This integration would add value to both, as from one side semantic web bottom up adaptation did not happen as the learning curve is so steep and on the other side social web lack semantic precision. The goal was to formalize ontology of place from place name tags and reuse this ontology to retrieve information from folksonomy.

To achieve this, we reviewed the related work that studied folksonomy, gazetteer and ontology, with focus on the researches that constructed ontology from folksonomy, and researches that studied folksonomy in GI context, other work that formalized ontology of place using gazetteer feature types was also reviewed. The second step was studying folksonomy in more details, discussing its characteristics, problems, and approaches taken to solve these problems. Third step was studying the ontology of place and develop its conceptual design. Next we designed and implemented an automated approach to proof the research concept and hypothesis. Finally a prototype is developed to test the approach, and evaluate the results. All the steps were accomplished successfully with satisfying results. Tags from Flickr, which is a folksonomy application, could be successfully semantically structured in ontology of place. The ontology enhanced Flickr, by adding semantic precision to the information retrieval, and partially enhanced recall by suggesting related places to the user.

7.1. Achieved results against research outline

In this section we discuss the achieved results against the research objectives and questions stated in chapter 1. The objectives included extracting place name tags from other tags. We were able to accomplish this by querying a gazetteer service that identified place names and returned a search result with their spatial information. Moreover, the prototype proved that improving the structure of the folksonomy by establishing explicit “Part-Of” relationships between place name tags and formalizing them in ontology of place is a feasible approach that enhances information discovery and retrieval in folksonomies.

7.1.1. Can the dynamic knowledge provided by folksonomies be used as a resource for acquiring bottom-up knowledge?

The dynamic knowledge provided by folksonomy develops social features and complex interactions between users, as discussed in section 3.3 and section 3.5. The collective tagging activity creates a dynamic correspondence between a resource and a set of tags, i.e. an emergent categorization in terms of tags shared by a community. Folksonomy is regarded as laboratories of semiotic dynamics, and their investigation provide valuable insights into both the analysis and the design of large communicating systems (human or artificial). Despite the problems discussed in section 3.4 that evolve with folksonomy, the attempts to solve these problems and formalize folksonomy in knowledge bases have shown encouraging potential benefits. In our approach we followed a bottom-up approach and could successfully formalize ontology of place from tags that were acquired from a tagging system (Flickr). The ontology is then used to enrich folksonomy by offering more services to the user and improve the information retrieval precision.

7.1.2. What is the ontology of place? And how to formalize such an ontology?

Ontology is defined in section 4.2 as “A logical theory which gives an explicit, partial account of a conceptualization” (Guarino & Giaretta, 1995) where conceptualization is an intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality. Ontology of place encodes names of places and the terms that describe spatial relationships. The ontology of place is formalized as aggregation of two ontologies (see section 4.5). The first is “Geographic Feature Type Ontology” that encodes administrative hierarchy levels using gazetteer feature types, and defines the explicit “Part-Of” spatial relationship between the different levels. The second ontology is “Geographic Feature Ontology” that encodes the geographic features, for each feature the ontology encodes a place name, feature type, and “Part-Of” spatial relationship that encodes geographic features which are directly related to the concerned feature, rather than the whole hierarchies.

7.1.3. How can we enhance information retrieval in folksonomies by establishing explicit spatial “Part-Of” relationships?

To establish explicit “Part-of” (containment) spatial relationship between tags, a gazetteer service was used. A gazetteer is geospatial dictionaries of geographic names with the core components of place names, location, and feature type (see section 2.4). The gazetteer service also provides the hierarchal relationship between administrative levels. Establishing such relationship between places disambiguates the place based on its name and administrative hierarchy (parent-child relationship). The prototype developed in chapter 5, and its evaluation showed that with this approach we added semantic precision, which allowed for offering better service to folksonomy user and improving information retrieval.

7.2. Future work

- As discussed earlier folksonomies problems decrease its benefits, one way to enhance this is to focus on the preprocessing of tags or as called in some studies “tag gardening”. In this way the tags used to construct the ontology of place will have fewer problems and hence result in ontology with fewer distortions, and the information recall will not be decreased.

- Adding the geographic location of each place to the ontology, will allow for more enhancements in suggesting related places to the user, as the Euclidean distance between each two places can be calculated and the suggested places will be arranged in ascending order based on the closer places (shorter Euclidean distance).

- The place location from the gazetteer can also be used to buffer the area around it and retrieve geo-tagged photos in that range. This is possible using the place coordinates as center and a predesigned radius.

- Due to the fact that many users are non-English speakers and they use their own language in tagging the objects, some objects cannot be retrieved by search engines unless the user knows all the alternative words for what she/he is looking for (e.g. firenze & Florence, köln & cologne). This implies that adding alternative place names to the ontology will improve its ability to recall information. This information is for the most part available in the Gazetteer and can be incorporated within this work.

- The place names which are not popular and has homonym problem (e.g. Sunrise) should be added to the list of nouns that are removed from the tags as discussed in section (5.2.3) as they are also considered distortion to the ontology. Such nouns can not be previously known, but removing them iteratively will result in a more distortion free ontology.

- Currently, the ontology of place encodes one spatial relationship that is "Part-Of" relationship. Including more spatial relationships will better represent how a geographic feature is related to other geographic features, this may include:
  - **adjacent-to** relationship, encoding which geographic features share a boundary with another geographic feature.
  - **overlapping** relationship, encoding which geographic features overlap with a geographic feature .

- Adding all or some of the previous suggestion must be accompanied by expanding the search tool to include these new options, to improve its service.

- One final point is to build an automated system with associated interface that automates the entire workflow developed in this thesis. This will make enriching folksonomies a standard automated task that can be easily accomplished.

References

1. Abdelmoty, Smart, & Jones. (2007). Building Place Ontologies for the Semantic Web: Issues and Approaches. Paper presented at the Proceedings of the 4th ACM workshop on Geographical information retrieval, Lisbon, Portugal.

2. Angeletou, S., Sabou, M., Specia, L., & Motta, E. (2007). Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report, Proc. of the ESWC (European Semantic Web Conference) "Bridging the Gap between Semantic Web and Web 2.0 (Vol. 2).

3. Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Integrating collaborative tagging and emergent semantics for image retrieval. Paper presented at the 15th International World Wide Web Conference, Collaborative Web Tagging Workshop, Edinburgh, Scotland.

4. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 28-37.

5. Bishr, M., & Kuhn, W. (2007). Geospatial Information Bottom-Up: A Matter of Trust and Semantics, AGILE. Aalborg, Denmark: Springer.

6. Catt, R. D. (2008). Flickr can haz (some) Shapedata. Retrieved 12 November 2008, from http://geobloggers.com/2008/10/30/flickr-can-haz-some-shapedata/

7. Cattuto, C., Benz, D., Hotho, A., & Stumme, G. (2008). Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems. Paper presented at the 3rd Workshop on Ontology Learning and Population (OLP3), July 22nd, Patras, Greece.

8. Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. Proc Natl Acad Sci U S A, 104(5), 1461-1464.

9. Cope, A. S. (2008a). The Shape of Alpha. Retrieved 10 November, 2008, from http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/

10. Cope, A. S. (2008b). Who's On First? Retrieved 14 November, 2008, from http://code.flickr.com/blog/2008/09/04/whos-on-first/

11. Fellbaum, C., & NetLibrary, I. (1998). WordNet: an electronic lexical database: MIT Press USA.

12. Fu, Jones, C., & Abdelmoty, A. (2005). Building a Geographical Ontology for Intelligent Spatial Search on the Web. Paper presented at the Proceedings of IASTED International Conference on Databases and Applications, Innsbruck, Austria.

13. Fu, G., Abdelmoty, A., & Jones, C. (2003). Spatially-Aware Information Retrieval on the Internet (SPIRIT), Design of a Geographical Ontology, http://www.geo-spirit.com/publications/SPIRIT_WP3_D5.pdf, Access date: January 2009 (No. D5 3101).

14. Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2), 198.

15. Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web: Springer.

16. Grothe, C., & Schaab, J. (2008). An Evaluation of Kernel Density Estimation and Support Vector Machines for Automated Generation of Footprints for Imprecise Regions from Geotags. Paper presented at the International Workshop on Computational Models of Place, PLACE'08, Held in conjunction with GIScience'08, Park City, Utah, USA.

17. Gruber. (1993). A translation approach to portable ontology specifications. KNOWLEDGE ACQUISITION, 5, 199-199.

18. Gruber. (2005). Folksonomy of Ontology: A Mash-up of Apples and Oranges. Paper presented at the Proceedings of First on-Line conference on Metadata and Semantics Research (MTSR).

19. Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological

20. Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up tags. D-Lib Magazine, 12(1), 1082-9873.

21. Henriksson, R., Kauppinen, T., & Hyvönen, E. (2008). Core geographical concepts: case Finnish geo-ontology.

22. Hill, L. (2000). Core Elements of Digital Gazetters: Placenames, Categories, and Footprints. LECTURE NOTES IN COMPUTER SCIENCE, 280-290.

23. Hill, L. (2006). Georeferencing: the geographic associations of information: MIT Press.

24. Hyvonen, E., Lindroos, R., Kauppinen, T., & Henriksson, R. (2007). An ontology service for geographical content.

25. Janee, G. (2006). Rethinking gazetteers and interoperability. Paper presented at the International Workshop on Digital Gazetteer Research & Practice, Santa Barbara, California.

26. Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. International Journal of Geographical Information Science, 22:10, 1129-1157.

27. Jones, C., Alani, H., & Tudhope, D. (2001). Geographical Information Retrieval with Ontologies of Place. Paper presented at the Proceedings of International Conference on Spatial Information Theory: Foundations of Geographic Information Science, COSIT, CA, USA.

28. Kogut, P., Cranefield, S., Hart, L., Dutra, M., Baclawski, K., Kokar, M., et al. (2002). UML for ontology development. The Knowledge Engineering Review, 17(01), 61-64.

29. Kroski, E. (2006). The Hive Mind: Folksonomies and User-Based Tagging. Retrieved January 2009, http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/, 14(2006), 259-284.

30. Lacy, L. (2005). Owl: Representing Information Using The Web Ontology Language: Trafford Publishing.

31. Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). Geographic Information Systems and Science (Second ed.): Wiley.

32. Maala, M., Delteil, A., & Azough, A. (2007). A conversion process from flickr tags to rdf descriptions. Paper presented at the 10th International Conference on Business Information Systems, Poznan, Poland.

33. Mathes, A. (2004). Folksonomies-Cooperative Classification and Communication Through Shared Metadata. Computer Mediated Communication, LIS590CMC (Doctoral Seminar), Graduate School of Library and Information Science, University of Illinois Urbana-Champaign.

34. Merholz, P. (2004a). Ethnoclassification and vernacular vocabularies. Retrieved January, 2009, from http://www.peterme.com/archives/000387.html

35. Merholz, P. (2004b). Metadata for the masses. Adaptive Path Retrieved January, 2009, from http://www.adaptivepath.com/ideas/essays/archives/000361.php

36. Mika, P. (2005). Ontologies Are Us: A Unified Model of Social Networks and Semantics. Paper presented at the The Semantic Web-ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, November 6-10, 2005: Proceedings, Galway, Ireland.

37. O'Reilly, T. (2005). What is Web 2.0. Design patterns and business models for the next generation of software. from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

38. Peters, & Stock. (2007). Folksonomy and Information Retrieval. Paper presented at the Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology.

39. Peters, & Weller. (2008). Tag gardening for folksonomy enrichment and maintenance. Webology,

40. Rumbaugh, J., Jacobson, I., & Booch, G. (1998). The Unified Modeling Language Reference Guide: Addison-Wesley.

41. Salton, & McGill. (1986). Introduction to Modern Information Retrieval: McGraw-Hill, Inc. New York, NY, USA.

42. Schlieder. (2007). Modeling Collaborative Semantics with a Geographic Recommender, In: Hainaut, J. & al. (eds.) International Workshop on Semantic and Conceptual Issues in Geographic Information Systems (Vol. 4802, pp. 338-347). Auckland, New Zealand: LNCS, Springer: Berlin.

43. Schlieder, & Matyas. (2008). Photographing a City: An Analysis of Place Concepts Based on Spatial Choices. Paper presented at the International Workshop on Computational Models of Place, PLACE'08, Held in conjunction with GIScience'08, Park City, Utah, USA.

44. Schmitz, P. (2006). Inducing ontology from flickr tags. Paper presented at the Collaborative Web Tagging Workshop at WWW 2006, May, Edinburgh, Scotland.

45. Smith, G. (2004). Folksonomy: social classification. Retrieved January, 2009, from http://atomiq.org/archives/2004/08/folksonomy_social_classification.html

46. Sowa, J. (1996). Ontologies for Knowledge Sharing. Manuscript of the invited talk at TKE, 96.

47. Specia, L., Angeletou, S., Sabou, M., & Motta, E. (2007). Bridging the gap between folksonomies and the semantic web: An experience report. Paper presented at the Proc. of The European Semantic Web Conference ESWC.

48. Specia, L., & Motta, E. (2007). Integrating Folksonomies with the Semantic Web. LECTURE NOTES IN COMPUTER SCIENCE, 4519, 624.

49. Speller, E. (2007). Collaborative tagging, folksonomies, distributed classification or ethnoclassification: a literature review. Library Student Journal.

50. Szomszor, Cattuto, Alani, O'Hara, Baldassarri, Loreto, et al. (2007). Folksonomies, the Semantic Web, and Movie Recommendation. Paper presented at the Proceedings of the workshop on Bridging the Gap between Semantic Web and Web 2.0 at the 4th European Semantic Web Conference (ESWC2007), Innsbruck, Austria.

51. van Assem, M., Menken, M., Schreiber, G., Wielemaker, J., & Wielinga, B. (2004). A Method for Converting Thesauri to RDF/OWL. LECTURE NOTES IN COMPUTER SCIENCE, 17-31.

52. Van Damme, C., Hepp, M., & Siorpaes, K. (2007). FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. Bridging the Gap between Semantic Web and Web, 2, 57–70.

53. Vander Wal, T. (2005). Folksonomy Definition and Wikipedia. Retrieved January, 2009, from http://www.vanderwal.net/random/entrysel.php?blog=1750

54. Weinberger, D. (2005). Taxonomies to Tags: From Trees to Piles of Leaves. January 2009, from http://www.hyperorg.com/blogger/misc/taxonomies_and_tags.html

55. Weiss, A. (2005). The power of collective intelligence. netWorker, 9(3), 16-23.

56. Williams, M. (14 November 2007). locating a business house thru API (Google Maps API discussion group). Retrieved 20 November 2008, from http://groups.google.com/group/Google-Maps-API/browse_thread/thread/788af26a509d84be/1f33109d48526f8c?lnk=gst&q=business+near#1f33109d48526f8c

57. Wilske, F. (2008). Approximation of Neighborhood Boundaries Using Collaborative Tagging Systems, GI-Days. Münster, Germany.

58. Winget, M. (2006). User-defined classification on the online photo sharing site Flickr … Or, how I learned to stop worrying and love the million typing monkeys. 17th ASIS&T SIG/CR Classification Research Workshop.

Appendix A

CD-ROM

Contents of the CD-ROM

- \Code: contains the java classes
  - getPhotos: the code for acquiring photos from Flickr within the boundaries of USA
  - getTag: the code for acquiring all tags for each photo
  - removeElement: the code to remove redundant tags.
  - nameRemoveElement: the code to remove any tag that contains any of the most famous 1000 boys names or 1000 girls names or 1000 most spoken English words
  - searchGazetteer: the code to search the gazetteer for place names and retrieve their spatial information.
  - createOntology: the code to create the ontology of place that changes the format and structure of the result from gazetteer search to format and structure of OWL.
- \Results: contains the files resulting from each stage
  - \Photos\photosInUSA.xml: the file that contains the acquired photos from Flickr
  - \Tags\editFinal.xml: the files that contains the list of tags after all the editing and the list of names that were removed
  - \Gazetteer\ searchGazetteer.xml: the result from the gazetteer search
  - \Ontology\USAplace1.owl: the OWL file of the ontology
- \FeatureTypes\featureType.doc: a complete list of all feature classes in geonames and feature codes.

Student Declaration

I declare that the submitted work has been completed by me the undersigned and that I have not used any other than permitted reference sources or materials nor engaged in any plagiarism. All references and other sources used by me have been appropriately acknowledged in the work. I further declare that the work has not been submitted for the purpose of academic examination, either in its original or similar form, anywhere else.

Declared in Münster ……………………….
(date)

……………………..…………
(Unterschrift)