Semantics of Place: Ontology Enrichment

Bruno Antunes[1], Ana Alves[1,2], and Francisco C. Pereira[1]

[1] CISUC, Department of Informatics Engineering
University of Coimbra, Portugal
{bema,ana,camara}@dei.uc.pt
[2] ISEC, Coimbra Institute of Engineering, Portugal
aalves@isec.pt

Abstract. In this paper, we present an approach to a challenge well known from the area of Ubiquitous Computing: extracting meaning out of geo-referenced information. The importance of this "semantics of place" problem is proportional to the number of available services and data that are common nowadays. Having rich knowledge about a place, we open up a new realm of "Location Based Services" that can behave more intelligently. Our approach builds on Ontology Engineering techniques in order to build a network of semantic associations between a place and related concepts. We briefly describe the KUSCO system and present some preliminary results.

Keywords: Point of Interest, Place, Semantic of Place, Ontologies, Ontology Enrichment, Ontology Evaluation.

1 Introduction

The current ubiquitous availability of localization technologies (particularly GPS) has driven to the emergence of many new applications (the "Location Based Services" or LBS) and enormous amounts of geo-referenced data. However, although we have already available rich and sophisticated knowledge representation and techniques (e.g. Semantic Web and Ontology Engineering) that allows for elaborate uses, information on location or *place* tends to be poor, with little or no directly associated semantics (e.g. the typical Point Of Interest or POI simply has a description and a generic type; in other cases we only have the latitude/longitude pair; and sometimes an LBS has its own purpose driven semantics, unusable to others). The association of a set of semantic concepts to a place should allow the application of those sophisticated techniques and foster the quality of current and future LBS (e.g. with better indexing). For example, knowing simply that "Goldener Adler" is placed at "47.268430 N 11.392230 E" and that it is a restaurant misses much of its essential meaning: what kind of food do they make? how expensive is it? where exactly is it? is it open now?

KUSCO is a system that intends to find and associate a set of semantic tags to Points Of Interest (or POI's). A POI is a geo-referenced tag that contains a latitude/longitude pair, a name and a type (e.g. restaurant, museum, gas station).

H. Geffner et al. (Eds.): IBERAMIA 2008, LNAI 5290, pp. 342–351, 2008.

Such information is used by KUSCO, which applies a number of techniques to automatically extract information from the Web about that POI. The system starts by doing a web search (e.g. using Yahoo), then it extracts and calculates statistics about the main words used. Afterwards, it associates these words to concepts (of WordNet [1]), and finally it determines the relationships of these concepts with the POI by using Place Ontologies.

In this paper, we describe the Ontology Enrichment process of the KUSCO system. This process is necessary to establish the association between the Generic Place Ontologies and specific POI instances, which is also normally called Ontology Instantiation. Given a POI and an Ontology, KUSCO seeks for the associations between the Ontology *Terms* and the more relevant words found in the Web pages most related to the POI. In the next section, we present "Semantics of Place" problem, well known in the area of Ubiquitous Computing. Then, we present some Ontology Engineering concepts that are essential to this work: Learning and Evaluation. In section 4, we present KUSCO and show some preliminary results. We end the paper by discussing next steps (section 5) and final remarks (section 6).

2 Semantics of Place

First introduced in [2], Jeffrey Hightower argues that location must have more associated information than simply the absolute position in a global coordinate system. Location representation needs more human-readable information including geographic, demographic, environmental, historical and, perhaps, commercial attributes. The meaning of place derives from social conventions, their private or public nature, possibilities for communication, etc. [3,4]. As argued by [5] on distinguishing the concept of place from space, a place is generally a space with something added - social meaning, conventions, cultural understandings about role, function and nature - having also temporal properties, once the same space can be different places at different times. Thus, a place only exists if it has some meaning for someone and the construction of this meaning is the main objective of our research.

As a formal definition, location models can be classified into four main types [6]: Geometric, Symbolic, Hybrid or Semantic. While the first three models (the third considers both geometric and symbolic) are mainly devoted to spatial relationship between locations, the last one, the Semantic Location Model, is orthogonal to symbolic and geometric representations. The semantic representation provides other information around its place, such as a bus route or a snapshot of interest. For instance, a hybrid location model was proposed by Jiang and Steenkiste [7] where they decompose the physical environment in different levels of precision and feature a self-descriptive location representation to each level. At a lower level of decomposition, they use a local and 3D coordinate system (Latitude, Longitude, Altitude) to define points or areas for which there is no name in the hierarchical tree. As an example, they can identify a specific printer in the Carnegie Mellon University campus by the identifier *cmu/wean-hall/floor3/3100-corridor#(10,10,0)*. In a different perspective, the HP Cooltown [8] introduces a semantic representation of locations. Its main goal is to support web presence for people, places and things. They use Universal Resource Indentifies (URIs) for addressing, physical URI beaconing and sensing of URIs for discovery, and localized web servers for directories in order to create a location-aware ubiquitous system to support nomadic users. In the same line, Ubiquitous Web [9] was envisioned as a pervasive web infrastructure in which all physical objects are socially tagged and accessible by URIs, providing information and services that enrich users experiences in their physical context as the web does in the cyberspace.

While the focus of our work is the Semantic aspect of Location Representation, we also take advantage of information available on the Web about public places. With the growth of the World Wide Web, we think that almost every commercial and non-commercial entities of public interest are or tend to become present on-line by proper web sites or referred by other related institutions. This should become even more relevant for places considered interesting for a group of people (i.e. they are in a sense *Points Of Interest*). But differently from the two previous semantic models, we don't assume that the Semantic Web is already a reality, with all information semantically structured and tagged. Actually, it is widely accepted that the majority of on-line information is composed of unrestricted user-written texts, so we get mainly dependent on the Information Extraction (IE) capabilities. IE is a research subtopic within Information Retrieval (IR) devoted to extract useful information from a body of text, including techniques like Term Extraction and Named Entity Recognition. In the Natural Language Processing field, there are other techniques which will be further used in our work, including part-of-speech tagging and word sense disambiguation to discover meaningful key concepts from the Web and contextualize it within a Common Sense Ontology (see the following sections).

3 Ontology Engineering

The growing amount of information available on the web demands for the development of efficient and practical information extraction approaches, in order to avoid the actual user's overloading of information. This need for new ways of extracting information from the web stimulated a new vision, the Semantic Web [10], where resources available have associated machine-readable semantic information. For this to come true, a knowledge representation structure for representing the semantics associated to resources would be necessary, and that was where ontologies [11] assumed a central role in the movement of the Semantic Web. Because it is nearly impossible to design an ontology of the world, research focused on the development of domain-specific ontologies, in which construction and maintenance are time-consuming and error-prone when manually done. In order to automate this process, research on ontology learning has emerged, combining information extraction and learning methods to automatically, or semi-automatically, build ontologies.

3.1 Ontology Learning

According to [12], ontology learning can be described as *"the process of automatic or semi-automatic construction, enrichment and adaptation of ontologies"*. It relies on a set of algorithms, methods, techniques and tools to automatically, or semi-automatically, extract information about a specific domain to construct or adapt ontologies. The process of ontology learning comprises four different tasks: ontology population, ontology enrichment, inconsistency resolution and ontology evaluation. Ontology population is the task that deals with the instantiation of concepts and relations in an ontology, without changing its structure. On the other hand, ontology enrichment is the task of extending an ontology by adding new concepts, relations and rules, which results in changes on its structure. Because errors and inconsistencies can be introduced during ontology population and enrichment, inconsistency resolution aims to detect these inconsistencies and generate appropriate resolutions. Finally, the ontology evaluation task assesses the ontology by measuring its quality with respect to some particular criteria (see section 3.2).

The ontology learning process can be performed through three different major approaches [12]: the integration of ontologies by capturing the features that are shared between them; the construction of a new ontology from scratch, based on the information extracted from data about a specific domain; and the specialization of a generic ontology by adapting it to a specific domain.

Following the work of Buitlaar et al. [13], the ontology learning process deals with six different aspects related with the structure of an ontology: *Term Identification*, which is in the basis of every ontology learning process; *Synonym Identification*, for identifying sets of terms that refer to the same concept or relation; *Concept Identification*, which is done through the realizations of the concept (i.e. terms); *Taxonomy Construction*, using inclusion relations (usually known as "is-a" relations); *Semantic Relations Extraction*, for identifying non-taxonomic relations that connect semantically related concepts; and *Rule Acquisition*, the least explored aspect of ontology learning.

3.2 Ontology Evaluation

The need for well defined techniques of ontology evaluation arise from the fact that different ontology conceptualizations can be constructed from the same body of knowledge. Much of the work developed in this field came from the context of ontology learning and enrichment, where different evaluation approaches were explored to evaluate the resulting ontologies. Also, the increasing development of semantic-aware applications, that make use of ontologies, uncovered the need to evaluate the available ontologies and choose the one that best fits the specific needs of the application.

According to [14], there are four different categories of techniques used for ontology evaluation: those based on a "golden standard", those based on the results of an application that makes use of the ontology, those based on the use of a corpus about the domain to be covered by the ontology, and those where evaluation is done by humans.

When the semantic characterization of place involves the construction and enrichment of place ontologies, it becomes necessary to apply some of the techniques developed for ontology evaluation, in order to assess the quality of the ontology produced and validate the proposed ontology enrichment approach.

4 KUSCO

The problem of "position to place" is a well known challenge within the area of Ubiquitous Computing and relates deeply with the connection humans have with places, their functionality and meaning. Attached to a tag name, even when a category is included, a place needs a richer semantic representation in our perspective in order to be understood. This knowledge can be used for whatever processes that demand semantics of place (e.g. understanding POI's while in navigation; searching for a place that has specific characteristics; route planning using locations with specific functionalities; inferring user's activity, etc.). We formally name this process as Semantic Enrichment of Place and it consists of using available Common Sense Ontologies and Web information to build a collection of generic and instance facts about these places. Figure 1 shows the internal architecture of KUSCO System. The following sections will explain in detail each component of this system.

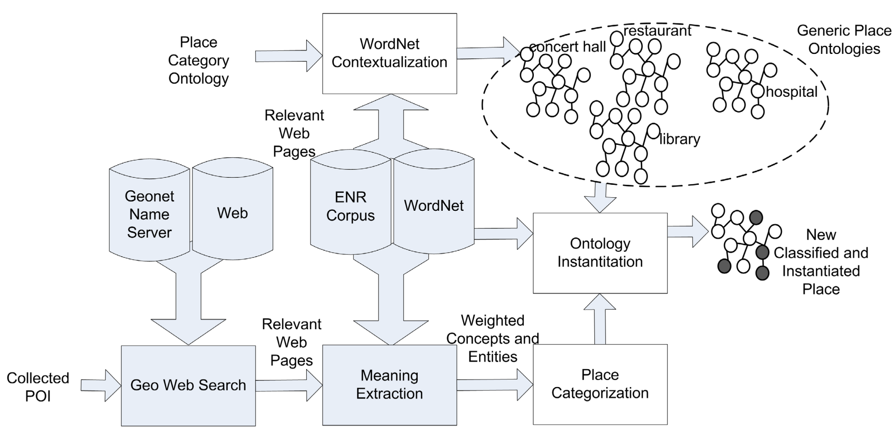

Fig. 1. The architecture of KUSCO

4.1 Generic Place Ontologies

The module of Generic Place Ontologies represents a collection of commonsense and generic information about well-known place categories, like restaurants, cinemas, museums, hotels, hospitals, etc. At a first stage, this information is manually collected from well-known and shared Ontologies (retrieving and selecting the most popular using ontology search engines like [15]). But as the system is used, it is dynamically fed by new examples, and thus instantiated and populated by specific facts about these instances that represent real-world places (see fig. 1). At a first stage, only concepts are extracted from Web describing places. Furthermore we also want to instantiate relations between these concepts using the original context where they appear. In order to infer place meaning, ontologies are contextualized on WordNet. For each term in an ontology, a WordNet's definition will be looked for (see section 4.4 for an example of this process).

4.2 Geo Web Search

This module is responsible for finding Web pages using only POI data as keywords: place name and geographical address. This last element is composed of the City name (where POI is located) and is obtained from Gazetteers [1] available on Web). This search is presently made by the freely available Yahoo API. For the moment, we are applying a simple heuristic that use the geographical reference as another keyword in the search. Thus, assuming a POI is a quadruple (Latitude, Longitude, [Category,] Name)[2], the final query to search will be: "City Name" + ["Category" +] "Name". In the near future, however, we think that works like Geotumba[17] and Wikipedia[18], with their geo-reference annotated information, will contribute to get better precision of this process, once the search on the Web can be localized only in a given geographical region. At this moment our system is very sensitive to geographical location of Place Name. For example, after looking for specific Web information for a given POI named "Carnegie Hall" in New York, we find many relevant results all referring to the same place: a concert venue. In another example, given a POI in the same city about "Mount Sinai" (a hospital), a geographical search gives us definitions such as a hospital and a metropolitan neighborhood. This shows us that this approach can become very dependent of search algorithms and of the Web's representativeness of places. At the end of this process, the N more relevant pages are selected (as suggested by the search engine).

4.3 Meaning Extraction

Having the set of Web pages found earlier, keyword extraction and contextualization on WordNet is made at this point. This processing includes POS tagging, Named Entity Recognition and Word Sense Disambiguation using available NLP tools [**?**,19,20]. On completion of these sub tasks for each web page, we are able to extract the most relevant terms (only common or proper nouns) that will be used in the categorization task (next module). The common nouns are contextualized on WordNet and thus can be thought not only as a word but more

[1] A geographical dictionary (as at the back of an atlas) generally including position and geographical names like Geonet Names Server and Geographic Names Information System [16].

[2] This category refers to the type of POI in question, a museum, a restaurant, a pub, etc. This information is optional, once sometimes it may not be present in the POI.

cognitively as a concept (specifically a synset - family of words having the same meaning, i.e., synonyms [?]). Each concept importance is computed by TF/IDF weighting [21] considering the most relevant WebPages for all POI's on that category.

4.4 Place Categorization

In order to evaluate the capacity of categorizing POI's (i.e. if they represent restaurants, museums, bars, etc.) we selected a set of ontologies using a popularity based criteria [22]. The result of this ontology selection process was a set of four ontologies about different domains: restaurants[3], museums[4], travel[5] and shows[6].

In an initial phase, already described, POI's were associated to a set of WordNet concepts. To facilitate the categorization of the POI's against this set of ontologies, we also map the concepts of the selected ontologies in WordNet. The mapping comprises three phases:

1. *Term Extraction.* The terms are extracted from the name of the concepts contained in the ontology. Because these names are usually comprised of one or more terms, they are split by upper case letters and special characters such as '-' and '_'.
2. *Term Composition.* In a preliminary analysis of the results obtained in the previous phase, we found that some of the terms, extracted of the split concept names, represented composed entities such as 'fast food' or 'self-service'. To avoid loosing these composed entities, the different terms extracted from each concept are combined and the resulting combinations are included as terms associated to the concept.
3. *Concept Identification.* The terms and combinations of terms, extracted in the previous phases, are then searched in WordNet. When more that one sense is found for each term, these are disambiguated by selecting the sense with the greatest tag count value. The tag count is a value given by WordNet for each word sense and represents the frequency of that word sense in a textual corpus.

The result of the mapping process is that all the concepts of each ontology became associated to one or more concepts of WordNet. With the ontologies already mapped in WordNet, the categorization process proceeds with three different approaches:

– *Simple Approach.* The simple approach, as its name tells, is the most simple approach and represents the direct mapping between the concepts associated to POI's and the concepts associated to the ontologies. The mappings between concepts of the two structures are counted and the POI is categorized in the ontology with the greatest number of mappings.

[3] http://gaia.fdi.ucm.es/ontologies/restaurant.owl
[4] http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs
[5] http://protege.cim3.net/file/pub/ontologies/travel/travel.owl
[6] http://www-agentcities.doc.ic.ac.uk/ontology/shows.daml

Table 1. Percentages of correctly categorized POI's

	Simple	**Weighted**	**Expanded**
Restaurants (I)	71%	29%	59%
Restaurants (II)	70%	41%	69%
Museums (I)	0%	15%	15%
Museums (II)	14%	14%	14%

- *Weighted Approach.* The weighted approach takes advantage of the TF/IDF [21] value of each one of the concepts that are associated to POI's. The TF/IDF value represents the weight of the concept in relation to the POI it is associated to. This way, each mapping has a weight equal to the weight of the concept that originated the mapping. The POI is then categorized in the ontology with the greatest sum of mapping weights.
- *Expanded Approach.* The expanded approach is based on the idea that the expansion of the concepts to their hyponyms make the mapping more tolerant and extensive. One may argue that when searching for *restaurants*, we are implicitly searching for every kind of restaurant, such as an *italian restaurant* or a *self-service restaurant*. Following this idea, the concepts associated to POI's are expanded to their hyponyms and the concepts that result from this expansion are associated to the POI. Then, the mapping between POI's and ontolgies is performed as in the simple approach.

4.5 Preliminary Results

In order to evaluate the three categorization approaches described before, we conducted some preliminary experiments with four sets of POI's, manually categorized as restaurants and museums. We then used the three categorization approaches to categorize the POI's according to the four ontologies previously selected and mapped in WordNet. The percentages of correctly categorized POI's for each set are presented in Table 1.

Although this is a preliminary experimentation, using a total of 116 POI's, the results obtained reveal interesting hints. As expected, the quality of the ontologies is crucial to the results of the categorization process. In our experimentation scenario, the ontology representing the restaurants domain was clearly more detailed than that representing the museums domain. Furthermore, the museums ontology was very abstract, which decreases the probability of matching with the specific concepts associated to POI's. In part, this explains the bad results of the POI's representing museums. Another interesting result is that the simple approach performs better than the weighted approach in most cases. This reveals that somehow the TF/IDF value used for weighting the concepts associated to the POI's is not reflecting the real weight of the concept, which should be improved in a near future. Also, we can conclude that the expanded approach stays very close to the simple approach. In this situation, there is not an evident gain on expanding the concepts to their hyponyms. Again, the quality and detail of the ontologies used may have a strong impact in the results obtained with this approach in the way that when ontologies are not specific enough there is no point on specializing the concepts associated to the POI's.

5 Future Work

This is an ongoing work and a lot of ideas are planned to be tested in the near future, some of them were extracted from the results obtained so far. One of the things we plan to improve is the way the TF/IDF of each concept associated to the POI's is calculated. Actually, it is calculated only taking into account the set of six web pages that were selected for each POI. Clearly, this set of documents is too small for obtaining relevant TF/IDF values. We believe that calculating the TF/IDF value taking into account all the web pages selected to the whole set of POI's will improve its relevance. Another fact already referred is that the quality of the ontologies used in the process is crucial to the results obtained, which demands for a more carefully selection and evaluation of such ontologies. Some of the approaches developed in areas such as ontology evaluation (see section 3.2) may be applied, in order to guarantee the quality of the ontologies used in the system. The next step is the ontology instatiation process, which will make use of the information generated in the previous steps to create instances of concrete POI's in the Generic Place Ontologies.

6 Conclusions

It is clear that, in order to improve current and future location based services, more information must be associated to common POI's. Location representation needs more human-readable information including geographic, demographic, environmental, historical and, perhaps, commercial attributes. KUSCO, the system we are developing, implements a process that we call as Semantic Enrichment of Place, which consists of using available Common Sense Ontologies and Web information to build a collection of generic and instance facts about these places. We have described the system architecture and foccused in the process of association between the Generic Place Ontologies and specific POI instances. Interesting results were obtained in a preliminary experimentation, which revealed important hints that will be used for future improvements.

References

1. Fellbaum, C. (ed.): Wordnet, an Electronic Lexical Database. MIT Press, Cambridge (1998)
2. Hightower, J.: From position to place. In: Proc. of the Workshop on Location-Aware Computing. Part of the Ubiquitous Computing Conference, pp. 10–12 (2003)
3. Genereux, R., Ward, L., Russell, J.: The behavioral component in the meaning of places. Journal of Environmental Psychology 3, 43–55 (1983)

4. Kramer, B.: Classification of generic places: Explorations with implications for evaluation. Journal of Environmental Psychology 15, 3–22 (1995)
5. Harrison, S., Dourish, P.: Re-place-ing space: the roles of place and space in collaborative systems. In: CSCW 1996: Proceedings of the 1996 ACM conference on Computer supported cooperative work, pp. 67–76. ACM Press, New York (1996)
6. Ye, J., Coyle, L., Dobson, S., Nixon, P.: A unified semantics space model. In: Hightower, J., Schiele, B., Strang, T. (eds.) LoCA 2007. LNCS, vol. 4718, pp. 103–120. Springer, Heidelberg (2007)
7. Jiang, C., Steenkiste, P.: A hybrid location model with a computable location identifier for ubiquitous computing (2002)
8. Kindberg, T., Barton, J., Morgan, J., et al.: People, places, things: Web presence for the real world. In: Proc. of WMCSA 2000 (2000)
9. Vazquez, J.I., Abaitua, J., Ipiïa, D.L.: The ubiquitous web as a model to lead our environments to their full potential. In: W3C Workshop on the Ubiquitous Web (2006) (position paper)
10. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American 284, 34–43 (2001)
11. Zuniga, G.L.: Ontology: Its transformation from philosophy to information systems. In: Proceedings of the International Conference on Formal Ontology in Information Systems, pp. 187–197. ACM Press, New York (2001)
12. Petasis, G., Karkaletsis, V., Paliouras, G.: Ontology population and enrichment: State of the art. Public deliverable d4.3, BOEMIE Project (2007)
13. Buitelaar, P., Cimiano, P., Magnini, B.: Ontology Learning from Text: Methods, Evaluation and Applications. IOS Press, Amsterdam (2005)
14. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. Technical report, Josef Stefan Institute (2005)
15. Swoogle: Semantic web search engine (2008)
16. Geonet names server (gns). National imagery and mapping agency (2008)
17. Cardoso, N., Martins, B., Andrade, L., Silva, M.J.: The xldb group at geoclef 2005. Technical report (2005)
18. Wikipedia (2008)
19. Patwardhan, S., Banerjee, S., Pedersen, T.: Senserelate: Targetword-a generalized framework for word sense disambiguation. In: AAAI, pp. 1692–1693 (2005)
20. Krishnan, V., Manning, C.D.: An effective two-stage model for exploiting nonlocal dependencies in named entity recognition. In: ACL 2006: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, Morristown, NJ, USA, pp. 1121–1128. Association for Computational Linguistics (2006)
21. Salton, G., Wong, A., Yang, A.C.S.: A vector space model for automatic indexing. Communications of the ACM 18, 229–237 (1975)
22. Sabou, M., Lopez, V., Motta, E., Uren, V.: Ontology selection: Ontology evaluation on the real semantic web. In: Proceedings of the Evaluation of Ontologies on the Web Workshop, held in conjunction with WWW (2006)