Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 4 (2012) 375 – 383

Procedia Technology

C3IT-2012

Semantic Annotations for Customizing Geographical Maps Based on WordNet

Anthoniraj Amalanathan[a], Shanmugam. S[b], Senthilnathan. M[b], Harikumar. S[b]

[a]*Assistant Professor of School of Computing Science and Engineering, VIT University, Vellore-632014, India.*
[b]*M.Sc. Student of School of Computing Science and Engineering, VIT University, Vellore-632014, India.*

Abstract

Geographical maps are an ideal technique to present and visualizethe real world information which is retrieved from various resources such as satellite and sensors. These data are effectively used in complex applications like Google andOpenStreet Map. Every location or place in a map is represented by latitude and longitude coordinates. Thislocation based search methodology can be utilized to find distance or path between two places, weather and traffic information, hotel, café and restaurant. This paper proposes an approach to describe and retrieve information by annotating location and places using semantic technologies. It is expected to retrieve more precise results by adding semantic annotation to the location. Users can also customize themap using this method to look up cities, addresses and other map related information.

Keywords: WordNet; Semantic web; Geographic Maps; LESK Algorithm; Ontology.

1. Introduction

1.1. Semantic Annotation

Semantic Annotation is a mechanism to add a label to content or a document or to an image. The labels are name of the field, comments and descriptions. In case of Map, the labels are tagged about the location, street, geo-position, shortest route to reach the destination and so on. In addition to these data further users can add up more live data known as “Metadata”.

Semantic Annotation assists to connect the ambiguity of the natural language by insisting the system, how the tagged labels are associated and these links could be assessed automatically which get into process the composite filter and search procedures. This tagging is used to retrieve the exact information or metadata in a fast access method from repository. Semantic annotation is linked to the integrated knowledge of a domain [5, 9].

2212-0173 © 2012 Published by Elsevier Ltd.
doi:10.1016/j.protcy.2012.05.058

1.2. Geographical Maps

Geographical Map in the internet is called as "Web mapping". There are various geographical maps available as a web-service such as Google Maps, Yahoo Maps, OpenStreet maps, Wikimapia and so on. For instance, Google maps provide various customizations to users. It also provides different services where developer can do geo codingto get the direction and traffic flow. The main advantage of these maps gives Interoperability in the form of web-service. Google maps API is easy tool to access the map. Marker is the method to mark a custom place on the map given by the user (Example, OpenStreet Map Marker). In the semantic web, the metadata could be added to the map in the form of map markers. This geographical map can be customized effectively by adding semantic annotation markers.

1.3. WordNet

WordNet is a database consists of word sense id, synonyms, nouns, verbs, adverbs and adjectives. Synsetis thegroup words into set of synonyms. Synsets are interconnected by means of semantic abstract and the relationship between the lexical. It is widely used in semantic web for lexical analysis.

1.4. LESK Algorithm

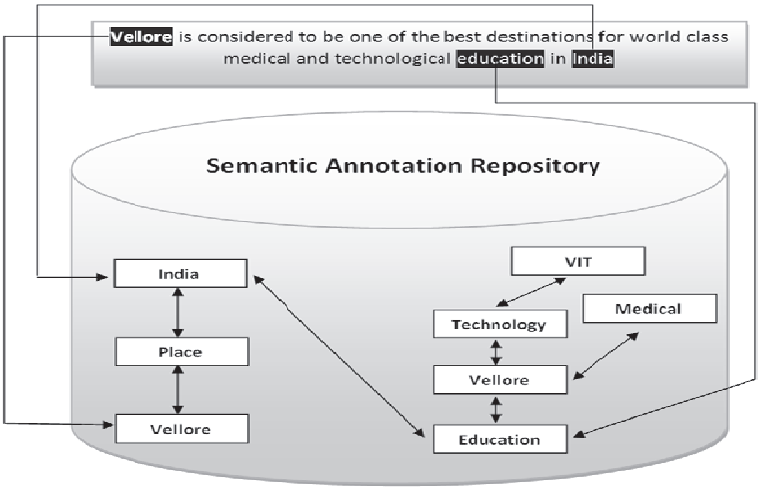
To find the best sense of the word in the given sentences there exits various algorithms such as Smoothing [6] and LESK. LESK algorithm is used to evaluate the senses of the sentences. Wordnet is the best lexical application gives the sense of a word. For Instance, In Wordnet the word "man" is a noun which has 11 senses, whereas the word "good" as a noun has 4 senses but the word "good" as an adjective has 21 senses.

Nomenclature

σ	Algebraic selection operator
π	Algebraic projection operator

2. Related Works

2.1. Semantic Indexing and Retrieval



Semantic web becomes more meaningful when more metadata is available. Annotation should be specified up to certain thing rather than general knowledge. By applying the core meaning of ontology to maintain this specific annotation information provided about most relevant entity like people, industries, country, etc. Provisioning details of such things helps to correlate domain specific from inter domain senses.

Semantic Annotation repository consists of all related Meta tags that is required for given query. In some system, text repository has Knowledge Management (KM) [5,11] for the integrated annotation. Fig.1. represents semantic structure of a search phrase "Vellore is considered to be one of the best destinations for world class technology education in India". From the given phrase the repository unit will search until for corresponding domains of appropriate words in the phrases, such as "Vellore", "education", "India". Each word is correlated to its main meta-domain. i.e., "Vellore" falls under the domain "place" which falls under the main-domain "India". But In turn, "India" has another sub-domain "education". "Technology" is the subset of "Education" domain. Finally, as per the query "VIT" results as an outcome because it falls in the following domains: "Technology", "Education", "Vellore" and "India".

Fig. 1. Semantic Annotation Repository with an example phrase

When these are related with semantic annotations the most sensed word will be provided from the repository such as Vellore is a city which comes under India, where cities are sub domain of India. In India, there are many cities one among those is Vellore which arises here is Education domain, further ithas many sub domain within it; when a city Vellore is related to education,repository response with many streams of education that are available in Vellore such that Technology is relative words in this phrase which yields VIT from Technology in accordance with Vellore. [5, 11]

2.2. Why to develop ontology?

In today's World Wide Web, ontology is a usual tool to annotate programmatically. Ontology determines the essential quality of a common listing of the words used in some enterprise for researches that need to use jointly or in common information in a domain. Ontology includes Machine-Interpretable definitions of basic concepts in the domain and relations in division which is shared byeach other. Ontology shares common understanding of the constructed information of software factor. To render

capable or able for some task use again after processing of domain knowledge [7]. Ontology makes domains the act of taking possession of or power over something. Explicit it is separate into parts or portions. To analyze domain knowledge sharing common understanding of the structure of information among people or software agent in one of the more the place designated as the end in developing ontology's. Enabling reuse of domain knowledge was or of the power of driving or impelling powerful behind recent surge in ontology research making explicit domain assumptions. Ontology has the inherited methodology which is used for creating precise domain precondition where domain knowledge is distinguished from functional knowledge [7].Analyzing domain is possible once a declarative specification of the term availablean ontology. Ontology is a formal explicit act of describing a concept in a domain discourse. [12]

2.3. Web Ontology Language

OWL (Web Ontology Language) was designed by the W3C used by all kinds of domains such as Education, Medical and Social. It is built with RDF and RDF Schema languages. In case of map annotation OWL could be used to define the metadata for map dataset. For example, annotating the word "wine" and labelling it with various attributes annotation header is used. Which consist of <owlx:Annotation> and <owlx:Label> tags. A sample header annotation is depicts the annotation of a word "wine" and labeling it in different languages [1]. Ontology Indexing and Inferencingare the two web semantic search based on the frequencies [3].

```
<owlx:Classowlx:name="wine" owlx:complete="false">
<owlx:Annotation>
<owlx:Labelxml:lang="en">wine</owlx:Label>
<owlx:Labelxml:lang="fr">vin</owlx:Label>
</owlx:Annotation>
<owlx:Classowlx:name="&food;PotableLiquid" />
</owlx:Class>
```

2.4. Semantic annotation of the gravity map

Semantics are related with artifacts like maps through which metadata are added in the form of annotation. Annotation acts as a connection between abstract of ontology's and determined artifacts. Annotations are tags mentions about the specific instances, gravity contour map extended from the gravity map scripts. It can be colligate with the contour map instances outlined in the gravity ontology, emerging tool ontology concept is being related with all kind of entities in the various domain in this case annotating concept is achieved through ontology object, Typically metadata are used to annotate with the help of ontology ideas. Metadata are the absolute terms are the legends as in Google maps, open street maps, ArcGis and XML for Image Annotations (XIMA). In Counter to Google map ArcGIS and XIMA permits users to annotate entire map, particular latitude and longitude on a map and inner regions of map by applying text or image as a captions. Inner region parts are outlined by vector points which in turns forms into polygon. By applying certain methods such as gridding, contouring datasets are linked to raw data maps semantically using the high rich ontology techniques. This result produces maps with Knowledge Provenance (KP).[10]

3. Proposed Methodology

User defined information for the particular place in the map is termed as "Metadata" by considering a geographical map of certain region. A Metadata comprises of one or more information about the location. In other words, it is called as Annotating. For our test case the lexical database of WordNet has been taken, to retrieve the best sense for annotating a location. The word sense can be determined using LESK Algorithm. The annotated metadata of a location are scanned to a related semantic word with the help of lexical database. Basically, each word has a sense; to this algorithm maximum sense metadata is correlated.

Maps are used to represent the landscape images. The proposal deals with "OpenStreet Maps" powdered by wiki world map. It is completely customizable as per user needs; since open street maps made up of vector plane points. These vector plane points are resizable to a new boundary lines. A road is extended or shrunken; in such a case corresponding vector points can be dragged to a new position, it also supports graphical legends that may be used to store user defined metadata in it.

3.1. Adding Metadata to map

Metadata could be added with the help of markers and custom locations are added to map by the user. User can append the metadata using the metadata and it could be useful to find the relationship between different locations. The following proposed algorithm is used to add a metadata to a particular location in the geographical map.

```
procedure addMetadata( )
        md← metadata
        loc(x,y)← location on the map
        loc(x,y)←addLoc(md)
end addMetadata
```

The above algorithm states how to add a metadata on a map; *md* is taken as metadata while annotating, *md* has information of a location. *addLoc* function adds the metadata (md) to a specified location in a map, with respect to the x as latitude and y as longitude.

3.2. LESK Algorithm

The disambiguation of a word using LESK algorithm and WordNet lexical database results to the best sense. Each word sense is mapped with other words in the sentence.[12]

```
procedure lesk()
        sentence←set of words
        max-synset←MAX(all words in the sentence)
        possible-synsets←high frequency word with all other words
        foreach sense in possible-synsets do
        gloss←set of words in the gloss
        best-sense←computeSense(gloss,sentence)
        end foreach
end lesk
```

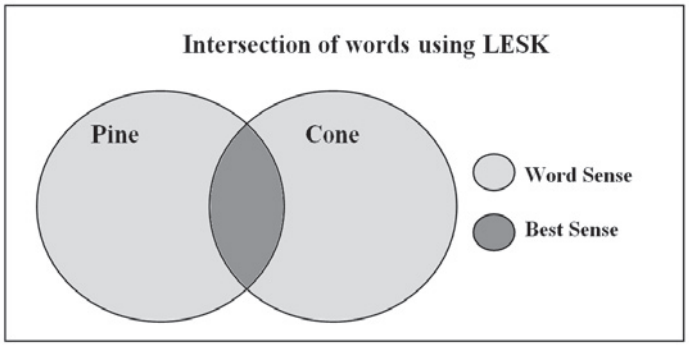


Fig.2. Intersection of words using LESK

The above algorithm describes the best-sense of a word in the given sentence. The example is also represented in the Fig. 2. For Instance, Word Pine contains x number of senses, word Cone contains the y number of senses. Among two words maximum sense is determined max(x,y). Then maximum-word is set with possible words in the sentence. In addition to the gloss of possible set of words are obtained and in which best-sense word is extracted (i.e.,$x \cap y \in \max(x, y)$)[12]

3.3. WordNet

WordNet is used for lexical references for semantic annotation on a map, in practice WordNet API's or WordNet lexical database are being used rather than WordNet application, WordNet web services are much easier to use among distributed platforms and provides interoperability programming language formats such as XML (eXtensibleMarkup Language) and JSON (Java Script Object Notation). Asia WordNet Project (AWN) serves as web service for language. It offers various services for developers such as WordNet dictionary, auto complete words, semantic browser and senses. In above mentioned services JSON is pretty faster when compared to XML

PyWordNet has been developed in Python programming language for accessing lexical meaning from WordNet; it contains collection of WordNet packages which can be easily imported to a programming language. WordNet API's are available for all high level programming languages.

Below XML represents the sample code of WordNet autocomplete web-service for the string "word". Result has 50 related outcome words given accordingly to the input string.

```
<data>
<item id= "1"><lemma>word</lemma></item>
<item id= "2"><lemma>wordsworth</lemma></item>
<item id= "3"><lemma>wordsmith</lemma></item>
</data>
```

The abovemethod uses the web-service for autocomplete words using WordNet. This service could be useful while annotating the repeated the semantic information of a words in give sentence. For the same outcome the JSON result will be shorter and faster due to fewer amounts of lines.

4. Result Analysis

4.1. Select and Project Operators

Selection and project are basic algebraic operators used for representing the retrieval of contents in the form of mathematical notation. Generally, A Select operator is denoted by Sigma ($\sum$) and a project operator is denoted by Pi ($\prod$). These two standardized symbols are being used in fetching the elements as per the condition. For example, A SQL query performs the selection; mean that it fetches all values of all attributes from the relation. In case of semantic annotation, these operators could be used for representing the WordNet access method such as selection and projection[4].

4.1.1. Selection for retrieving synonym of a noun word

Table 1 depicts about the mathematical notation with the input and output. The notation is represented for the selecting the word from the database based on the condition P_K. The outcome of selection based retrieval would be the synonym of the input word *W*.[4]

Table 1 Mathematical Notation of Synonym word

Input	Output
The word W with the condition P_K	A Synonym collection S_σ of all grammars can be defined as $\{W_{ni} \in W_\sigma \mid P_K (S_{ni})\}$
Mathematical Notation $\sum P_K (W) = W_\sigma$	

4.1.2. Selection for retrieving holonyms word

Table 2 depicts about the mathematical notation with the input and output. The notation is represented for the selecting the word from the database based on the condition P_K. The outcome of projection based retrieval would be the holonym of the input word *W*.[4]

Table 2 Mathematical Notation of Holonyms word

Input	Output
The word W with the condition P_K	A Holonyms collection H_π of P_K can be defined as $\{W_{ni} \in W_\pi \mid P_K (H_{ni})\}$
Mathematical Notation $\prod P_K (W) = W_\pi$	

4.2. Precision and Recall

Synonyms and Holonyms are usedbased on WordNettechnique. This Practical method is applied to some particular task which has been developed for the Geographical information operation of accessing information from the task. It may help in determining the properties of something not directly expressed geographic information contained in texts.

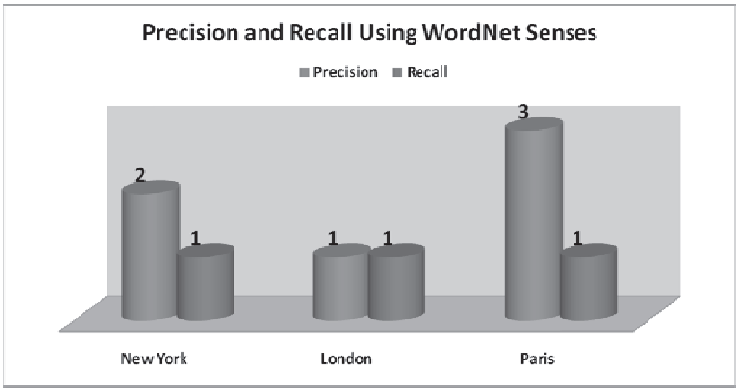


Fig.3. A sample precision and recall values in sense using WordNet

Further questionable activities, a line leading to a place or print will include the implementation of the same implies an orderly logical arrangement with a richer reserve supply that can be drawn upon when needed, the ontology product using the GNS and GNIS gazetteers in proximity with WordNet and the investigation of various ranking required that weight differently the geographical farms with a detail to the non-geographical ones [2, 9].

A sample phrases is taken as a model such as “Our Company is located in New York, London and Paris”. In this phrase each word has the senses. The related word with the precision values is determined by using WordNet. For Example, the word “Paris” has 4 senses out of which 3 sensed words are related to it. WordNet gives the following output, the precision value is 0.75 and the recall value is 0.25.

5. Conclusion

After the existence of web 3.0, semantic technologies play a vital role in all sorts of domains. Even in the medical and biological domain semantic based researches are going on. In this paper, we conclude that geographical maps such as Google Maps, Yahoo Maps and Wikimapia could be customizable using the semantic annotation technique and these maps could be enhanced more by relating to metadata of a location, based on the user preferences. The maps are annotated with the help of lexical databases such as WordNet and ontologies. These types of maps could also be implemented using ontology tools. In future, the semantic repository engine should be faster to access, correlate the metadata of all domains and also there should be a domain management system which manages the domains consisting of large entries of sub-domains.

References

1. Baader, F., Horrocks, I. and Sattler, U. "Description Logics as Ontology Languages for the Semantic Web"'Festschrift in honor of JörgSiekmann, Lecture Notes in Artificial Intelligence', Springer-Verlag, 2003, pp. 228--248.
2. Buscaldi, D., Rosso, P. and Sanchis, E. "A WordNet-based indexing technique for Geographical Information Retrieval,", 2006.
3. Jiang, X. and Tan, A.-H. "Learning and inferencing in user ontology for personalized Semantic Web search," Information Sciences (179:16), 2009, pp. 2794 - 2808.
4. Kapetanios, E., Schaal, M., Kapetanios, E. and Schaal, M. "A Model and Algebra for Collaborative Semantic Annotation in Digital Libraries".
5. Kiryakov, A., Popov, B., Terziev, I., Manov, D. and Ognyanoff, D. "Semantic annotation, indexing, and retrieval," Journal of Web Semantics (2), 2004, pp. 49--79.
6. Lapata, M. and Keller, F. "Evaluating smoothing algorithms against plausibility judgments"'In Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL, 2001), pp. 346--353.
7. Noy, N. F. and Mcguinness, D. L. "Ontology Development 101: A Guide to Creating Your First Ontology", 2001.
8. Oren, E., Delbru, R., Mцller, K., Vцlkel, M. and Handschuh, S. "Annotation and Navigation in Semantic Wikis", in Vцlkel, M. and Schaffert, S., ed.,'Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics', ESWC2006, 2006.
9. Reeve, L. "Survey of semantic annotation platforms"'Proceedings of the 2005 ACM Symposium on Applied Computing', ACM Press, 2005, pp. 1634--1638.
10. Rio, N. D., Silva, P. P. D. and Gates, A. Q. "Semantic Annotation of Maps Through Knowledge Provenance"'In Proceedings of the Second International Conference on Geospatial Semantics (GeoS 2007), pp. 29--30.
11. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E. and Ciravegna, F. "Semantic annotation for knowledge management: Requirements and a survey of the state of the art," Web Semantics: Science, Services and Agents on the World Wide Web (4:1), 2006, pp. 14 - 28.
12. Vasilescu, F., Langlais, P. and Lapalme, G. "Evaluating Variants of the Lesk Approach for Disambiguating Words"'Proceedings of Language Resources and Evaluation (LREC 2004)', Lisbonne, Portugal, 2004, pp. 633--636.