

Matching-concepts: An ontology Approach to Retrieve Geographical Information

Felix Mata¹ and Serguei Levachkine²

Geoprocessing Laboratory (GEOLAB)
Centre for Computing Research (CIC), National Polytechnic Institute (IPN)
Av. Juan de Dios Bátiz s/n Unidad Profesional “Adolfo López Mateos”
07738, México, D.F., MEXICO
migfel@sagitario.cic.ipn.mx¹, sergei@cic.ipn.mx²

Abstract. Nowadays one of the research directions in Geographical Information Retrieval (GIR) is focused on finding information related to particular geographical locations. Some approaches have treated this problem using geo-referenced information available from several sources, e.g. digital maps and spatial databases. However, reported in literature results are not very relevant to the user’s expectations, because the free-text queries have computed without considering properties and relations between geographic objects, e.g. topology of spatial data. If users express queries by names of places, prepositions, e.g. “near”, “along”, etc., then we will require additional mechanisms, e.g. spatial semantics treatment, to catch the nature of geographic objects appeared in these queries. In this work, we propose a method to match concepts using ontologies. In other words, our approach to the information retrieval is based on the spatial semantic properties and relations between objects and not only on text analysis. This approach improves the relevance of results for free-text queries, which contain geographic or spatial objects. We describe a method to retrieve spatial information based on the semantics of geographic components presented in a free-text query. Our approach represents an alternative to the keyword-matching. The idea consists of extracting the concepts presented in queries, and then matching them to the corresponding contexts. These contexts are embedded in ontologies. The best matching concept is obtained when the related to it contexts converge within an ontology. We semiautomatically built ontology from a universe of documents and propose a method to scan the ontology until finding the suitable concept-matching in accordance to the context of free-text query on a specific domain. In particular, we use queries of location and position. We present some results and show that the relevance of these results is closer to the user’s expectations. Finally, we compare our approach to the approaches that are not considered the semantics of spatial data.

1 Introduction

Presently, the retrieval of geographic information related to particular geographical locations is based on geo-referenced information sources (i.e., linked information to geographic coordinates) available on several sources (maps, geographically indexed

books, etc.) But, people do not make queries using geographic coordinates, very often a search starts submitting free-text queries to database systems, search engines or Web information systems. A free-text query contains several terms related to geographical or spatial objects, but not in the way that a GIS defines operations with spatial data, for example: within or inside can be interpreted as synonyms, but these can represent a buffer operation or other topological operation, the decision of which chosen can be solved by semantics on the query. With this sample we can see the importance of considering the semantics in free-text queries that involve geographic objects.

The goal of a search is to find information about any subject or activity related with our daily life, although commonly the results obtained do not satisfy the expectations of users. This problem occurs mainly because the techniques to solve queries are based on exact-match or keyword-match, these techniques are used in database systems and search engines respectively. So, these approaches present some problems such as the ambiguity of words.

Here is where others research lines have built techniques to avoid or solve these problems, one of them is the information retrieval (IR) where the answer to queries are based on the approximations to expected results. These expected results are denominated as *relevant*, additionally; the results are not ordered but ranked, by means of several methods (most of them using syntax techniques).

The main goal of IR consists of retrieving text documents that are relevant to a given query, where a document is considered relevant when it contains one or several words that also appear in queries. Nevertheless, that approach does not ensure finding the suitable answer, because the lack of search terms in some documents does not necessarily mean that the documents are not relevant. Besides, another disadvantage of current IR systems is that they are based on exploiting the nature of text, therefore the semantics of a query cannot be considered. Therefore alternative methods to improve that retrieval process are required, for example considering context, semantics, etc. The case of spatial data is not the exception to the problem described below, besides the methods used in traditional information retrieval systems (IRS) are not appropriate to geographic information, because if we consider that the classical information retrieval models (e.g., Vector Space, Probabilistic, Boolean) [14] are based on lexicographic term matching, then there is no way to consider relationships or properties of geographic objects (required to extract the semantic of geographic query).

Although, there are approaches in which the searches are performed with the aid of ontologies [1], the used approach cannot be applied in Geographical information systems, because the nature of spatial data requires a special treatment [2]. For example, two terms can be semantically different although they are lexicographically similar (near is semantically different when the topography is considered and when it is not considered). Then, retrieving documents by classical retrieval methods will fail, if the semantics of terms is not considered. Besides, in GIS the semantic processing approach has been widely used and proves significant results [18, 19] in that way semantic processing promises to be an interesting alternative to Retrieval information focused in geographical aspects, this field of research is known as Geographic Information Retrieval.

Therefore, at this point our work is addressing: to retrieve information by extracting the semantics implicit on free-text queries, and matching concepts on ontology, where we test our approach using location queries (e.g “the bank near the conference”, “hotel 1km of walking distance”)

1.1 BACKGROUND AND RELATED WORK

Nowadays, the search process is a fundamental component of retrieval information systems, the case of geographic information retrieval (GIR) is not the exception where several works have been oriented to searches, in this direction one of the main trends consists of improving the techniques and algorithms to extract relevant information (*best* answers to certain queries) where we find enough efforts to achieve that goal, for example in [3] the GIR is performed by means of spatial Bayesian algorithms, focused on workspaces of a commercial GIS, where a workspace is integrated by several layers of spatial data, the idea is to extract the relations present between these layers , in order to be exploited by the retrieval process .

Others’ works are addressed to the web, for example in [4] a geographic search was proposed using query-expansion by means of an API provided by a keyword-web search engine, one of the disadvantages of this approach is that expansion of queries (number of terms) is constrained by a search engine, then the retrieval process is affected in a negative way when many terms are required. The expansion approach has offered good results only in particular sceneries, where the number of terms was small, but when the number of terms is bigger, additional inconvenient and problems are presented, then the solution become another problem.

Other proposals are focused on solving the problem of ambiguity of words; the proposed solutions are based on a knowledge representation, such as: hierarchies of terms, taxonomies, and ontologies, but most of them are solutions based on text or syntax properties, while others describe treatment semantics without considering spatial relations, as in [5].

Into of this group several semantic approaches have been proposed, where one of the main contributions consists of including ontologies and semantic annotation, an example is described in [6].

The ontologies [7, 14] have been widely used in several semantic approaches; they are applied in, practically, any domain and of course in GIS field [8, 15]. Nevertheless, these approaches not consider processes and algorithms to explore ontologies, because an ontology describes domain theories for the explicit representation of the semantics of data [9], then we can use the ontology to know the semantics of query. Then, we need algorithms to explore these ontologies and getting the semantics required. In [10] the authors propose a way to match a location expression with certain places. It process is done using a database of places, the database is divided in two groups, the first one is formed by terms driven popularly and the second one is formed by terms very specialized. They show how geocoding can be implemented over incomplete and possibility inaccurate addressing data. Additionally, the paper shows a way to treat qualitative and quantitative data, but as a disadvantage a geocoding process is required previously. Other approaches are focused in ranking algorithms where very often the parameters considered are:

number of visits to sites and pages, or links and references, etc. An example is [11] where geo-data are used. Then, it is the point where our approach is addressed, no use geocoding, because not all users can express in explicitly way the geographical coordinates for a site or place (location queries). Moreover, a significant growth of geographically oriented search, local search, is undeniable as is described in [17], and then we need to develop better mechanisms to solve the location queries without expressing the places' coordinates. Furthermore, Many works have dealt on unstructured information where data contains some geographic reference for example in [23, 24].

Additionally, although the GIS research community have suggested and made emphasis on the use and treatment of spatial relations, few studies have been addressed in that direction, some examples are published on [12,13] while that recent work has focused on qualitative spatial reasoning, a sample can be found in the often-cited model of topological relations among point sets [18].

In the field of retrieval information many models have been used, where the term-based Vector Space Model (VSM) is the state-of-the-art document retrieval method [16] and it is based on lexicographic term matching. While that in our approach the matching is performance by concept. In particular Information Extraction (IE) and Information Retrieval (IR) are used in conjunction to built new tools that offer better results in the search process [13]. The first one has the task of organize indexes to be exploited by IR. And the second one, IE can be described as the process of populate a structured information repository (index) from an un-structured information source [21]. That task is performance in automatic way in two modalities: the first one extracting either whole information of a document, where every term of a document is treated and a weight is associated to each term [7]. While that in second one some fragments of document are extracted using predefined rules to find out specific information [21]. IR works with models, techniques, mechanisms to extract information that has already been processed, stored and (e.g. plain text files, databases, XML files). In IR the fast processing of queries is possible because the index structure was previously built [22].

Our work is addressed in *geographic queries* in particular those that include terms of proximity, i.e. *near*, *distance*, *behind*, *at the side of*, *in*, where these terms can be interpreted based on their meaning and in that form to offer a relevant answer. Moreover, we consider frame reference (those that helps to denote a location or position), i.e. "*within easy walking distance to Azteca Stadium*", "*five minutes from Art museum*".

The remainder of this paper is structured as follows: Section 2 describes the process to extract the semantics of queries. Section 3 shows the semi-automatic process to built ontology. In section 4 an overview of retrieval and search strategy (matching concepts) is described, where some examples are presented. In Section 5 preliminary results are presented. Finally, some conclusions are drawn and we discuss possibilities for future work.

2 SEMANTICS EXTRACTION OF QUERIES

Very often people talk and write about places using references or approximations to describe locations, ways, routes, etc. But, these descriptions lacks of precision, the following expression is a clear example of that: “*The restaurant is **in front of** Mayan Hotel*” in that case, if the direction is not expressed, the meaning of **in front of** can be different depending on the user’s position. Other sample occurs when place-approximations are used such as: “*The school is **near** main Avenue*” the meaning of **near** can change if the topography is considered or if is not considered. In these expressions a common characteristic is the use of a spatial relation (e.g *in front of*) and reference frames, or places widely known.

They are used for offering a major detail or precision about the location. These places widely known and spatial relations are rich on semantics, they can be extracted to help in the retrieval process, but a previous knowledge is required.

This previous knowledge is about the spatial relations and the places widely known (that knowledge can be got from ontology, dictionaries and gazetteers)

We explain that with the following scenario: suppose that a speaker wants to know if there are hotels near the conference venue, if he use a traditional system (e.g. Google) the results will be only documents that contain one or more words of the location query, but if we consider the possible relations associated to “*near*”, we will have, additionally, the relations and properties needed to contextualize the search.

These relations and properties are extracted from dictionaries, gazetteers and ontology. In the case of ontology, they were extracted by exploration of it, while that in the case of dictionaries and gazetteers, will be extracted by sentence analysis of definitions by each term.

The ontology exploration consists of finding (matching) the query’s terms in the ontology nodes. If the terms are found in the ontology, the relations and properties from them are extracted.

In the table 1 we can find the properties and relations extracted, for each term, from dictionaries and gazetteers to the location query: “*hotels **near** conference venue*”. In section 3.1 the extraction process is described. While that in section 2 the ontology construction process is explained.

Table 1. Relations and properties extracted to query : “*hotels near conference venue*”.

QUERY: “HOTELS NEAR CONFERENCE VENUE”.	
CONCEPT	PROPERTIES AND RELATIONS
NEAR	Within reach
	Related to center
	Related to periphery
	Related to time
	Related to place
	Relies in some sort of distance
	Synonyms: close, approximately, vicinity
HOTEL	Related to Tourism
	Related to services

CONFERENCE VENUE	Is a Place
	Related to lodging
	Has rooms
	Is a Resort
	Synonyms: guesthouse, boarding house
	Related to meeting
	Related to Topic
	Related to Event
	Has schedule
	Has location
	Is a Place
	Synonyms: place of trial, setting of event

The relations and properties shown in the table 1 were used to construct the ontology, the bold font indicates that they are *concepts* (have relations and properties). Previously, the *concepts* were manually processed, therefore, the ontology contains them (by definition ontologies are built based on concepts and relations) where that ontology will be enriched with additional relations (obtained from documents) during the rest of process. Moreover, the ontology will be used in conjunction with traditional retrieval approach to offer better results compared with those obtained by traditional systems.

3 THE ONTOLOGY CONSTRUCTION

In this section, we will explain the process to build the ontology. Basically, it is an approach based on two methods to achieve the goal. The first one is addressed to document collection and the second one is addressed to location query.

We processed the query and the documents employing these methods. In the case of documents we use an extraction method based on surrounding terms (EMST) and to the case of the query we apply an extraction method based on concept (EMC). The process was performance in semi-automatic way, using a program to extract the terms based on the two methods, afterward manually (according to the experience of specialist GIS) some terms are selected to be included in the ontology. The methods CEM and EMST are described in sections 3.1 and 3.2 respectively.

3.1 CONCEPTS EXTRACTION METHOD (CEM)

The method consists of two steps: the first one, where the terms are extracted from query (verbs and nouns) and stored in a table. The second one: dictionaries and gazetteers are used to find-extract the relations and properties of each concept. For example, for the earlier query we have the term *hotel* where the associated definition (from dictionary) is: “*is a place which provides overnight lodging and offers other services*”. In that case, the verbs are identified as relations, while that the nouns allow

identifying properties related to the concept. Then, a preliminary output for that process is:

Relations = {provide, is a, offer} Properties related= {place, services, overnight, lodging}

Here, the sentence analysis is based on mechanisms used in traditional IR, where the stopwords are not considered, but additionally rules are added to identify relations such as: "is a" "has a" "part of" etc.

3.2 EXTRACTION METHOD BY SURROUNDED TERMS (EMST)

This method is based on the hypothesis of that the surrounded terms to location query's terms are useful to find other relations. For example: a surrounded term to **near** could be subway station, street, or suburb, in each case we have a different representation (Point, Line, or Polygon respectively) therefore the relations will be different for each representation.

Another example is when a user submits the query: "rivers in Mexico", then suppose that we have a document that contains the following fragment of text: "Colorado **River** is a **river** in the south western United States and north western Mexico".

The surrounded terms (ST) to the "River" term are: ST= {in, is a, Colorado}. These ST will be added to the table that contains the results obtained using the CEM method. We explain now the overall process of EMST: First, based on document collection from previous experiment where the documents contain at least one proximity term or names of places widely known in Mexico City.

We computed the higher term frequency (HTF) and inverted document frequency (IDF) for each concept contained in each docut. The HTF and IDF is computed only to spatial relations (related with proximity) and the surrounded words to them (verbs and nouns) for each document. The *stopwords* were not considered. The Table 2 shows the properties and relations extracted to concept "near" using EMST and EMC.

Table 2. Semantics extracted to "near"

CONCEPT	SET OF SURROUNDED WORDS (SSW) AND HTF	PROPERTIES AND RELATIONS
NEAR	Along a route, Located, located inside, nearest, quite close to, conveniently located, located right, relatively close, Surrounded, 6 miles from, ~15 minutes from, is less than 10 minutes from, is situated, Only 1 km, less 10 minutes away from,...	Within reach
		Related to center
		Related to periphery
		Relies in some sort of distance
		Synonyms

Then, using these results the ontology is enriched by adding these relations and properties. The figure 1 shows a fragment of the ontology.

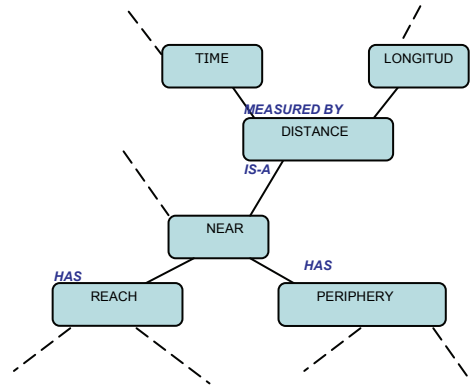


Fig. 1. Snippet of Ontology

As the Fig. 1 shows, there are properties and relations that can aid in the process of retrieval to location queries related to concept *near*, and not only by syntax properties. For example, in the table 2 the term *distance* appears in bold because *distance* is a concept in our ontology, therefore the documents that contain the concept *distance* will be processed, and also the *near* concept (because they have a relationship in the ontology).

Then, in this form, the relevant documents extraction for concept *near* will be performance by means of *distance* concept (without ontology, it would not be possible). In the following section we show an overview of the retrieval process, and a brief description of the strategy of search.

4 RETRIEVAL PROCESS (SEARCH STRATEGY)

Our mechanism of search is based on matching concepts for *geographic queries*. The strategy is divided into four tasks:

- 1.- Indexing phase.
- 2.- Construction of Inverted List of files.
- 3.- Ontology exploration.
- 4.- Correlation.

The first one is referred to the indexing phase, according to the context, in that case the context is: location. Then the sentence analysis recognizes spatial relations and location expressions by means of a simple heuristic, for example: *near* is marked as an spatial relation and “*walking distance*” as a location expression related to concept *near*. Additionally, the surrounded terms to the spatial relations, also, are extracted. The result of this step is a table of terms with three attributes: the term, spatial relation and location expression, where the possible values; for the first attribute is: the word, and for the second and third one are: a true or false value according to the word.

The second step is process the table (shown in table 1) to construct an inverted list of files accord to the concepts. The result is an inverted index file shown in table 3.

The third task is exploring the ontology to find the main terms (of the extracted terms in step 1) by means of top-down exploration of ontology. The exploration starts in the root node and the node's relations represent the way to go to other concepts (related), when the term is found, the exploration finish, and then context is extracted (context = the relations and properties around the concept).

Therefore, in this step the main terms are extracted (e.g. near is the main term to "five minutes from") a term is considered as "main" when it is represented as a spatial relation in the ontology (also is a node) and has another nodes related to it. These relations and properties are searched in the table of Inverted Index files, and the documents associated to these, are returned as candidate results to the initial query.

The last task consist of correlate the set of surrounded words (*SSW*) with others *SSW*'s (shown in table 2) for example *near* has a relationship with the concept *distance*, then the *SSW* of *near* and *SSW* of *distance* can have hidden relations, therefore the extraction of these is carried out. That is the point where we say that we match by concept. The representation about this process appears in Fig. 2.

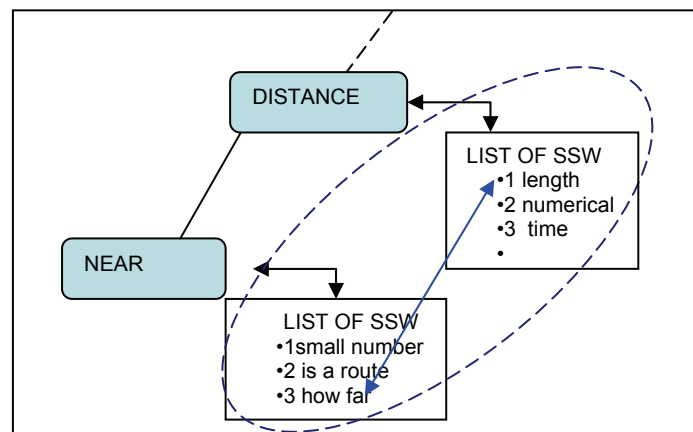


Fig. 2. Matching the Concepts NEAR and DISTANCE on ontology

The correlation process is performed using a simple relation rules, we describe two cases: the first one when documents do not contain the terms that appear in query. The first case is described in the present section and the second case is described in Section 4.1

The first case is described considering the next situation: suppose the reference frame: "Azteca stadium" in a query, now in the document collection there is a document *Y* contains the reference frame: *Azteca Stadium* and we find in the ontology exploration that "Azteca stadium" also appears in the SSW of *distance* concept. Nevertheless, other document *X* is related to the *near* concept, because document *X* contains in the SSW the name of a street (a street near of *Azteca Stadium* but the word *near* does not appear, nor *Azteca Stadium*).

Then, the traditional retrieval will not consider the document *X* because there is no matching. Although in document *X* is relevant. Then, to accord to document analysis

document X and *document Y* are not related. But, both of documents are considered as candidate to be related because in the ontology *near* and *distance* are related, and they should be appear in the final result.

The list inverted index is resumed in table 3 where the IDF (inverse document frequency), TF (term frequency) and radius search are considered to the queries related to location. After to this process will be determinate if they are correlated or not accord to the document weight. The calculation of document weight is explained in the following paragraph.

Table 3. List Inverted file used to correlate documents

Query	Document ID	Frequency proximity Word (FPW)	Reference Frame Frequency (RPF)	Resultant Frequency (RF)	Estimated Radius Search (ERS)	Final Weight Document (FWD)
1.-Near metro station Sevilla	001	12	2	6	1 km	60
2.-Close to Paseo de la Reforma Avenue	002	8	3	4	6 km	65
3.- Located inside forest Chapultepec	003	7	10	6	17 hectare	70

In table 3 we resume the process of correlation by using traditional inverted files but we added attributes to aid in the retrieval process, in particular in the processing queries. We can see in the first column three queries about places well known in Mexico City and very close in distance, but the queries using synonyms or other proximity expressions that requires a semantic processing to obtain relevant results, then in the table appears the main criteria to decide which documents are related, and which are not, the decision is taken by computing the values in each column to obtain a final weight for each document, this weight allows to decide if the correlation between document's terms is convenient to perform or not.

The headers of each column in table 3 are eloquent, as in traditional IR approaches the weight of document is computed based on the word occurrence number for each document. In our case we made some modifications to this process, for example the column RF is computed by considering the TF, IDF as in traditional IR theory, but adding the average between frequency proximity words (FPW) and reference frame frequency (RFF). The FPW is the number of occurrences in the document to words such as: "*near*" "*away*" "*5 minutes*", etc. RFF is the number of occurrences to places widely known such as: "*Central Bank*", "*Modern Art Museum*", etc. While that estimated radius search (ERS) column is computed based on the extension or length of the reference point, i.e. a station metro in Mexico City has an extension average of 1 km; in similar way to the other queries the value is assigned.

The final weight document column (FWD) is computed considering the ERS but restricted to the distance average to the main term, for example "*located inside*", has a related concept "*near*", then although the value RES for "*located inside*" is 17 hectare, the final average distance considered is 4 km because this is the average

value for “*near*” concept. A detail to realize is that FPW to *query 1* is higher to *query 2* and *query 3*, in the traditional approach the document 001 will be the more relevant, with our approach the document 003 is the most relevant.

That process is the last step in our strategy of search, although an additional case is included using logical inference (the user decides if the inference will be activated) that procedure is yet in test phase, but the preliminary results about initial tests are satisfactory, we describe the process in section 4.1

4.1 INFERENCE ON QUERIES

In this section, we show some queries where the relation’s properties are processed to extract the semantics of a spatial relation (i.e. “*in*”). That semantics will be used to enhancement the retrieval process. Here we apply axiomatic properties such as: symmetric, transitive, reflexive, inverse to spatial relations.

The goal is performing inferences by applying first logical order, which approach is well-known but the application is oriented to geographic information retrieval. The inference process allows obtaining additional and relevant results compared with traditional approach on IR.

An example is represented in the following scenario: Suppose a user who asks for “*Rivers north of Mexico*”, the semantics of *north* can be extracted with the process described below, but probably will there be concepts that do not have a direct relation with any concept.

In that case, the process can be enriched if the power of inference is used. To achieve that, we require using the results obtained from the previous queries related to the reference frames (Mexico data in that case).

These results will be stored in plain files for each query. For example, if we have the previous results for queries about *Mexico* and *Rivers*, where the documents contains expressions such as: “*USA north of Mexico*” and “*Wisconsin in USA*” then, the transitivity of “*in*” and the “*north of*” allow to infer that “*Wisconsin North of Mexico*”, in that way we can consider candidate results that match not only by the concepts “*in*” or “*north of*”, but also those whose semantics was obtained by inference. Nevertheless, additional mechanisms are required to refine and robust that phase. The table 4 shows some of the possible inferences about this particular query, according to the spatial reasoning.

Table 4. Inferences using axiomatic properties of relations

From previous results	- Wisconsin North of Mexico	- USA North of Mexico
Current query	- is Fox River in Wisconsin?	- is Wisconsin in USA ?
(Additional answers) By semantics of <i>in</i> and <i>North of</i> , we can infer that:	- Fox River North of Mexico	- Wisconsin North of Mexico

The examples shown in Table 4 describing the possible inferences to combining relations between two concepts, in particular the relations focused in direction or

location (e.g. *Northern, southern, in*) Therefore, using the transitivity property, from the previous example, is feasible infer that “*Fox River North of Mexico*”, **although these concepts are not directly associated in the ontology**. With this approach we can retrieve documents using basic logical inference.

5 PRELIMINARY RESULTS

With this approach we present the following results, the document collection, at this moment, is approximately of 100 documents where the documents are web pages and pdf files, they are from several sites of internet, the criteria to search in these sites was only that documents should contain a location term.

At this moment, the relevance of results is considered as good compared with the results obtained from web search engines, the criteria of assessing is based on test using search engines based on keywords versus our approach. Nevertheless, others tests are required to evaluate and show the statistic table from overall approach.

In the Figure 3 we summarize the overall process, with an example using a query and show the steps described earlier.

Consider the following query submitted by tourist:

q= Downtown in Mexico City is near of Benito Juarez International Airport?

Some snippets from retrieved documents are shown in Fig. 3.

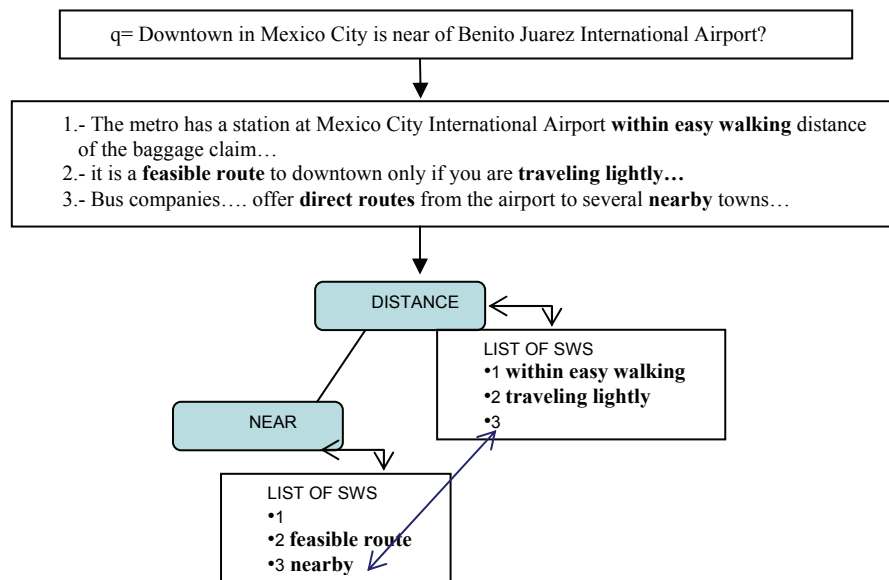


Fig. 3. Matching Concepts

In bold appears the location expression detected during the retrieval process. The first document is retrieved (using keyword match, it would not appear) in the ontology exploration phase, because the concept *near* has a relation with concept *distance*. Later, the concept *distance* has in their *SSW* list the expression: “*within easy walking*”, therefore, the correlation phase return the corresponding document as a candidate result, the process continues and document appears in the final results. The case is analogous to other documents. Nevertheless additional tests are required to formalize the process.

6 CONCLUSIONS AND FUTURE WORK

This work describes an approach to performance geographic information retrieval based on location queries, the main idea is extract and process the properties and relations from geographic objects that appear in a query and documents related to them. Our approach is based on exploring ontology, it is the method used to match the concepts by relations and properties and not only by syntactical methods.

The main goal of this work is to offer more relevant results to locations query, considering spatial relations and properties from terms such as: near, in, “to five minutes from”, etc. the justification is based on the fact that the actual retrieval process is performance without considering the nature of geographic object. For example: a query that contains the near term can be interpreted in different way if the topography is considered. Moreover, the primitives of representation for the surrounded terms (e.g. hotels, streets, counties) are rich in semantics, they can be the relations and properties involved in the meaning of a query.

It is not the same a query hotels near airport that hotels near main street (we require different spatial operations to obtain the results, although the spatial relation is the same in both queries).

The retrieval process is performance by means of exploring ontology and using mechanisms based on first order logic to make inferences. Moreover, using traditional retrieval information with support of ontologies can improve relevance of the results returned by traditional IR approaches and search engines.

The paper shows some examples about certain location and proximity queries. Nevertheless, it is necessary to perform additional testing to check the performance of our approach, in specific with a bigger document collection.

Our work is opposite to other techniques (e.g. query expansion) because the retrieval is performance by matching concept based on ontology exploration. Not by adding other extra keywords, based on the keyword-match. Moreover, we use the surrounding terms to query’s terms as a mechanism to make a better information retrieval, because they have a lot of semantics that can be used in the retrieval process. Additionally, we use the traditional inference mechanism used in spatial reasoning. It is with the purpose of solving some location queries; without using geographic coordinates in explicit way. We define a search strategy to exploit the semantic of spatial relation, in particular those that contains a location term.

The strategy and the rules used will be formalized as part of future work. This work extends the retrieval capabilities of existing methods and proposes a method to

explore ontology as a support to GIR in conjunction with techniques used in traditional IR.

ACKNOWLEDGES

The authors of this paper wish to thank the Centre for Computing Research (CIC), General Coordination of Postgraduate Study and Research (CGPI), National Polytechnic Institute (IPN) and the Mexican National Council for Science and Technology (CONACYT) for their support.

REFERENCES

1. Wordnet: network which models the relationships between words, for example, synonyms, antonyms, hyponyms. <http://www.wordnet.com>
2. Egenhofer M., Interaction with Geographic Information Systems via Spatial Queries, *Journal of Visual Languages and Computing*, 1 (4): 389-413, 1990
3. Walker A.R., Pham B., Moody M.: Spatial bayesian learning algorithms for geographic information retrieval, *Proceedings of the 13th annual ACM international workshop on Geographic information systems, GIS'05*, November 4-5, 2005, Bremen, Germany
4. Delboni T. M., Borges K. A. V., Laender A. H. F.: Geographic web search based on positioning expressions, *Proceedings of the workshop on Geographic information*, 2005, Bremen, Germany
5. Maedche A., Stabb S., Stojanovic N., Studer R., Sure Y.: "A framewok for developing Semantic Web Portals". *Lecture notes in Computer Science* v. 2097, 2001
6. Budak A., Sheth A., and Ramakrishnan C.: Geospatial Ontology Development and Semantic Analytics, *Handbook of Geographic Information Science*, Eds: J. P. Wilson and A. S. Fotheringham, Blackwell Publishing, 2004
7. Guarino N.: Formal ontology and information systems, *In Proceedings of the 1st International Conference*, Trento , Italy, IOS press, june 1998
8. Harding, J.: Geo-ontology Concepts and Issues, *Report of a workshop on Geo-ontology*, Ilkley UK, September, 16-17, 2003
9. Koo S., Lim S., Lee S.: Building an ontology based on hub words for information retrieval, *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, (2003) 46
10. Davids Jr. C.A. ,Fonseca F.,and Borges K.: A Flexible Addressing System for approximate Geocoding. In *Proceedings of V brazilian symposium on Geoinformatics*, Campos do Jordão, Brazil, 2003
11. Stabb S. et al: "Semantic Community Web Portals", 2000, in *WWW9/ computer Networks* (Special Issue: WWW9 – Proceedings of the 9th international world wide web conference, Amsterdam, the Netherlands, Maio, 15-19, 2000)
12. Heinzle F., Kocpczynsky M., and Sester M.: Spatial Data Interpretations for the intelligent access to spatial information in the internet, in *Proceedings of the 21th international cartographic conference*, Durban, South Africa 2003

13. Schilder F., Versley Y., and Habel C.: Extracting Spatial Information: grounding, classifying and linking spatial expressions. *In proceedings of the ACM SIGIR workshop on Geographic Information Retrieval*, Sheffield, UK, 2004
14. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993
15. Hammond, A. Sheth, and K. Kochut, Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content, *in Real World Semantic Web Applications*, V. Kashyap and L. Shklar, Eds., IOS Press, 2002
16. Salton G.: Automatic Text Processing: *The Transformation Analysis and Retrieval of Information by Computer*. Addison-Wesley, 1989
17. Himmelstein M.: *Local Search: The internet is the Yellow pages*, IEEE computer, 38 (2): 26-35, 2005
18. Wildöcher, A., Faurot, E., Bilhaut, F.: Multimodal indexation of contrastive structures in geographical documents. *In RIAO 2004*, Avignon, pages p.555–570, 2004
19. Torres, M.:Semantics definition to represent spatial data. *International Workshop on Semantic Processing of Spatial Data (Geopro 2002)* 2002
20. Baeza-Yates, and Ribeiro-Neto B.. Modern Information Retrieval. New York, NY: ACM Press; 1999
21. Gaizauskas, R.: An information extraction perspective on text mining: *Tasks, technologies and prototype applications. Euromap TextMining Seminar*, Sheffield, 2002
22. Jones, C.-B., Abdelmoty, A.-I., Finch, D., Fu, G., Vaid, S. The Spirit Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. *Third International Conference - Geographic Information Science*, Adelphi, Usa, October 20-23, (2004) 125 – 139
23. Desai S., Knoblock C. A., Chiang Y., Desai K., Chen C.: Automatically Identifying and Georeferencing Street Maps on the Web, *The 2nd International Workshop on Geographic Information Retrieval (GIR'05, CIKM Workshop)*, November 2005
24. Silveira-Chaves M., Silva M. J., Martins B.: A Geographic Knowledge Base for Semantic Web Applications. 40-54, *ACM GIR '05 workshop on GIR* (2005)