Spatial web search with Semantic Web Technology

Huda Abdalslam Banour [1], Prof. Iraky Khalifa[1] , Dr. Mohamed Haggag[2]

Prof of computer science Department [1], Asoc. prof of computer science Department[2]
[1, 2]Faculty of Computers and Information, Helwan University , Cairo, Egypt
mmmh13282@yahoo.com, Dr_Iraky@hotmail.com, m-h-haggag@yahoo.com

Abstract: Growth in web technology has seen the emergence of the Semantic Web which has been envisioned to provide solutions to users of web resources, especially in the area of data and service discovery. Many GIS (Geographic Information Systems) exist which facilitate sharing of data over the web in distributed environments. However search and retrieval of data and services is difficult due to the vocabulary used in different SDI which lead to semantic heterogeneity problems when only simple keyword-based search is employed [1].
[Huda Abdalslam, Banour, Iraky Hussein Khalifaand, Mohamed Hassan Haggag. **Spatial web search with Semantic Web Technology.** *Life Sci J* 2013;10(1):2926-2933]. (ISSN: 1097-8135). http://www.lifesciencesite.com. 356

Key words: GIS, Semantic Web, Ontology, Semantic Annotation, Ontology Visualization.

1. Introduction

In early warning information systems, GI (Geographic Information) play a pivotal role in effective planning and decision making. The information is made available by data, maps and satellite images allowing users to discover them, however not much meaning is attached to them. It is left up to the user to make meaning of what is made available on the maps such that they may really never discover the functionality and usage of such information [2]. The search and retrieval of useful spatial data and services remains a challenge to the user community due to the diversity in meaning of data and web services [3]. It is not easy for a user to find the right information with the right functionality for his/her specific purpose [4].

Catalogues exist that provide searchable repositories of information and services but mechanisms to support discovery and retrieval are insufficient [5]. Interoperability of platforms for discovery and accessing data/services have been upheld and supported by the OGC through established protocols and interface specifications thereby offering support for the important task of discovery and retrieval of information that meets the user's needs. Retrieval methods that are currently in use are typically limited to keyword search or sub-string matching only. The search methods only account for the syntax of the search terms without taking into account the underlying conceptualizations. Using these methods, the information required is poorly defined in the search and results often do not satisfy the user's needs. As a result users may often miss critical information when searching for spatial data.

The search mechanism so far is achieved through static key word matching without full exploration of underlying semantics, such as hierarchical relationships among spatial data entities [6]. When using key word searching, where retrieval is based on syntax matching. More has to be done to allow machines to be able to interpret meaning of terms used thus achieving semantic interoperability. This has resulted in the emergence of the semantic web.

The Semantic Web is an extension of the current World Wide Web (www), but it is based on the idea of exchanging information with explicit, formal and machine-accessible descriptions of meaning [7] using such languages as eXitensible Markup Language (XML). In this case, documents and services are provided with well-defined meaning (semantics) and laid down in formal descriptions. Full implementation of the Semantic Web requires ontologies in which widespread availability of semantic annotations for existing and new documents on the Web have been made.

Ontologies play a pivotal role in making web content understandable and available for machine processing through the encoding of the meaning of concepts in a particular domain by detailing the relationship between the concepts [8]. Before publishing a resource in the web, it has to be annotated with a descriptive data to make it accessible. Users are able to find the resource using search engines and evaluate if the discovered resource satisfies their current information needs. There is need to move from the conventional methods that leave users questing for better search mechanisms that return results meeting the user's needs. The user will not need to scan through the results to weed out those that are not relevant. In order to curb the users' frustrations, semantic web technologies can be explored and their potential augmented in web GIS system in order to enhance the discovery of data and services.

The ontologies will be used to reveal the implicit and hidden knowledge. This research proposes strategies for integrating ontologies and semantic annotations into a GIS system so as to improve search mechanism allowing users to retrieve the most relevant information when required.

The main objective of the research is to investigate how semantic web technology can support geographic services through semantic annotation. Byreviewing the formalization techniques for better representationand semantic annotation of geographical data. And developing a method for integrating Semantic Web Services technology in GIS for the purpose of enhancing the search functionality.

2. Related Work

OGC has established and developed standards to allow interoperability of data formats shared by SDI and accessed through CSW. The OpenGIS Catalogue Services Interface Standard (CAT) supports the ability to publish and search collections of descriptive information (metadata) about geospatial data, services and related resources [9]. Providers of resources register service descriptions using their choice model. The hope has been to help users access the resources through client applications in an efficient way.

The WMS provides a simple Hypertext Transfer Protocol (HTTP) interface through which geo-registered map images from distributed geospatial databases can be requested. In the request, the user defines the layers of data required and area of interest. The response comes as a map image which can be displayed in a browser. There is room to combine data layers from several SDIs in producing a single map. The interpretation of the map images lies with the user in every case; no semantic descriptions are made when registering the services in catalogue services. The OpenGIS Web Feature Service Interface Standard (WFS) defines an interface where requests for retrieving geographic features are made and the operations involved. In catalogues this information is registered as metadata elements. To complement OGC work are International Standards Organization (ISO) standards which have established relationships within metadata elements and format for their registration.

ISO 19115:2003 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Though mostly used for digital data ISO 19115 is applicable even for maps. ISO 19139 was established to close the gap in ISO 19115, that of the formatting of the data through an XML schema. While all these standards have been developed and implemented, syntax matching in query and retrieval has been achieved, however semantic interoperability remains unaddressed in these standards.

Realizing the standards established by OGC and how they can be augmented with semantic web technologies for better service description and discovery, W3C has come up with recommendations for Semantic Annotations for WSDL and XML Schema (SAWSDL). The specification defines how to annotate WSDL interfaces and operations with categorization information that can be used to publish a Web service in a registry. The annotations on schema types can be used during Web service discovery and composition [10]. Some work has been done on including semantic metadata descriptions in catalogues such as GeoBrain.

GeoBrain, is a multidisciplinary system aimed at mobilizing NASA data and information through Web Service and knowledge management technologies allow users to dynamically and collaboratively develop interoperable and Web-executable geospatial service chains. To efficiently classify, register, describe, discover and access geospatial information, a semantically enabled OGC Catalog Service for Web (CSW) was used.

Klien proposed an architecture for ontology based discovery and retrieval of geographic information [5]. The user is allowed to formulate a query for metadata and geodata. The architecture can be extended further to allow spatial and temporal reasoning and to allow nested queries to be executed.

[11] developed a new algorithm, Service Aggregation Matchmaking (SAM) for composition-oriented discovery of Web services. Its deployment has been stalled by the lack of descriptions of services.

The purpose of the Semantic Geo-Catalogue (SGC) project [16] - promoted by the Autonomous Province of Trento (PAT) in Italy with the collaboration of Informatica Trentina, Trient Consulting Group and the University of Trento - was to develop a semantic geo-catalogue as an extension of the existing geo-portal of the PAT. It was conceived to support everyday activities of the employees of the PAT. The main requirement was to allow users to submit queries such as Bodies of water in Trento, run them on top of the available geographical resources metadata and get results also for more specific features such as rivers and lakes. This is clearly not possible without semantic support.

In this research, we will study how the existing search functionality works, and its limitation and how it can be improved using web technologies. We will build a GIS web application with the existing search functionalities (Metadata search, Spatial Search). Then integrating ontologies and semantic annotations into the system. Ontologies play an important role in spatial search by overcoming the limitations of agreements of textual specifications [4].

3. The Proposed system:

The proposed system is an integration of semantic web technologies in a GIS system to provide powerful search capabilities. The user interface should provide functionalities to search and retrieve data. Queries are done through free text search. Visualization of ontologies should be possible and flexible on the amount of detail viewed depending on user needs.

We have built a GIS application for geographic dataset visualization and spatial search. Then Integrate of OWL semantics into the GIS application.

There are 2 main methods for search in any GIS system:

- Metadata Search: search in the metadata xml files.
- Spatial Search: search using spatial criteria

We have implemented these 2 methods in our system. Then add support for semantic web GIS.Web GIS do not avoid traditional issues attached to non-semantic applications; indeed (i) requires considerable manual work, (ii) the lack of semantics limits the precision of queries, and (iii) limited expressiveness usually drastically limits functionality.

Current search mechanisms have been based on key words only without using the semantics and hierarchical nature of concepts. The research shows a way of improving search mechanism in a distributed environment through the use of ontologies and semantic annotations. They are the basis on which datasets and services are described in this research.

As any information system can gain advantage from the use of semantics. In GIS, the use of semantic concepts enhances the search. The Proposed Semantic Web GIS (SWGIS) is a system which answers geographically oriented queries in a smart way while integrating multiple information sources.

The proposed application is integration between GIS catalog and semantic web to improve the search mechanism in a distributed environment through the use of ontologies and semantic annotations. They are the basis on which datasets and services are described in this research.

PostgreSQL/PostGIS is used as a centralized database management system to store both spatial and non-spatial data. Postgis is used as a backend spatial database for our GIS system.

The geospatial data used in our research is a sample dataset of Egypt, it is a set of layers in shape file format (countries, Egypt cities, Egypt districts, Heliopolis roads, Heliopolis landmarks). These layers are transformed from shape file format to postgis format using the "shape file to postgis importer" plugin.

3.1 System Architecture:

The proposed system architecture is shown in Figure 1 where annotation service and ontology visualization forms components in our system and PostgreSQL/PostGIS is the data repository.

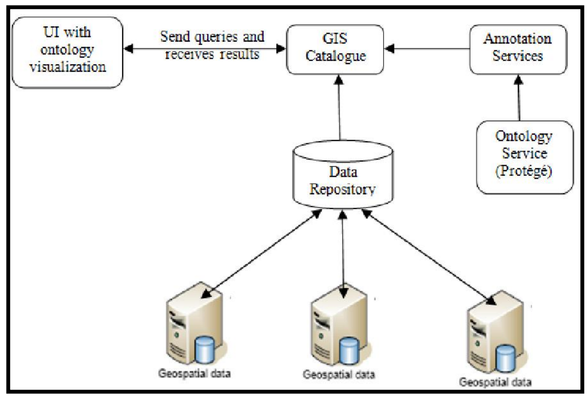

Figure 1: System Architecture

3.2 Use Case Identification

The user will have the ability to get the nearest landmark to another landmark. Also get list of landmarks from certain type that are spatially located within certain street or district.

3.3 Ontology:

The ontology service form an extra components in our system (the spatial relations ontology is developed using protégé software) then the ontology concepts are used to develop the semantic annotations.

Ontologies are used to capture knowledge about some domain of interest. Ontology describes the concepts in the domain and also the relationships that hold between those concepts.

Spatial Relation specifies how some object is located in space in relation to some reference object.

In our research, The Ontology file represent the spatial relationships between different spatial objects. The general purpose of the Places ontology is to provide a library of high level concepts and

Topographic ontology. The ontology also describes the spatial and non-spatial relationships and instances [12]. Table 1shows some spatial relations which are used in our research.

Table 1. SpatialRelationship Terms

Relationship Term	Natural language text definition
Is Located At	The object is positioned at the location of another object [13]. (e.g. landmark is located at street)
Is Located In	The object is completely within a human created legal entity or similar such as a county [13]. (e.g. city is located in country, district is located in city). This spatial relation is equivalent to is contained in relation
Is Located Near	The object has a position close to another object or physically touches another object [13]. (e.g. landmark is located near another landmark)

We have created a set of classes using C# language to read the ontology file structure format then import it to the postgres database tables.

3.4 Semantic Annotation:

A semantic annotation of a geospatial resource must provide semantic descriptions about geographic characteristics of this resource. In the development of semantic annotations, the following requirements were considered [15].

- Semantic annotations should be defined by domain experts during modeling of ontologies.
- Users should be able to understand the semantic information they deal with.
- The semantic annotations should be derived from appropriate domain ontologies In developing semantic annotation questions regarding to what objects, and actions on the domain ontologies they refer to were addressed.

The annotation procedure consists of the following steps [15]:

- All concept definitions that contain spatial relations are identified in the ontology. From each of these concept definitions, the characteristic spatial relations are extracted. This extraction (and the subsequent analysis) is “controlled”, in the sense that we will analyze the dataset by looking explicitly for the concepts defined in the ontology (rather than performing an “uncontrolled”
- For each spatial relation, the corresponding spatial analysis method will be extracted.For example, Located at is implemented as a sequence of GIS operations
- The GIS operations are applied to the geodataset to be annotated.
- The spatial entities that meet the characteristic spatial relations are stored as the result of this analysis step.
- The final result set is created by intersecting the result sets of all analysis steps.
- The ontological descriptions are then created.

The phases for dataset processing are shown in Figure 2. The first step towards the construction and population of the semantic annotation is to analyze the data provided by the Dataset, extract the main geographical classes and corresponding locations [15]. We retrieved the main Dataset classes (Country, City, District, Road, and Landmark).

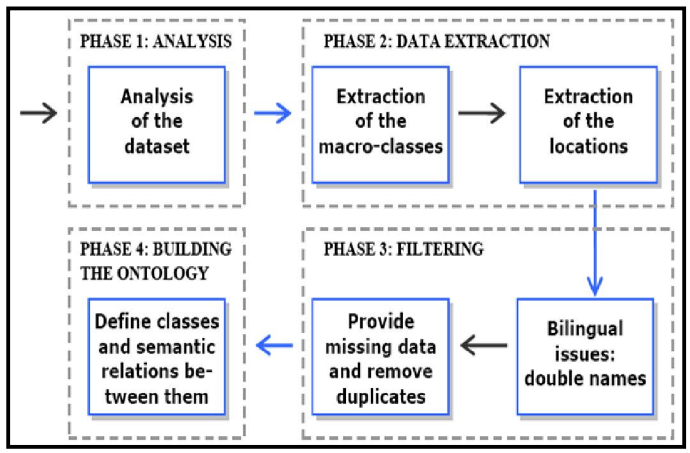

Figure 2. Phases for the dataset processing

We are focusing on spatial relations so that we extract all concept definitions from the geospatial domain ontology that contain spatial relations. Then translate spatial relations (e.g. located at) into a corresponding spatial analysis method, which is implemented as a sequence of GIS operations. Then apply these spatial analyses on the geodataset to be annotated. We have used 3 spatial analysis functions (i) Spatial Join function in QGis software to relate 2 GIS layers based on spatial locations of features in both layers, (ii) Get Nearest Line function to calculate at which road, each landmark is located at. We have built it using C#, (iii) Get Nearest Landmark function to generate “is lactated near” relation for all landmarks. We have built it also using C#.

Bilingual issues: Locations are provided with a name and possibly some alternative names [15]. The same location may be represented by 2 names (e.g. English and Arabic name). So that we must provide the names in both languages if we wants to search in both languages.While importing the entities in the database. There may be a missing or duplicate data which must be corrected before applying semantic annotation to produce correct annotation.

A semantic annotation of a geospatial resource must provide semantic descriptions about geographic characteristics of this resource [14]. Semantic annotation means that applying ontology terms to the data. In other words, create a link or association between the ontology terms and the geospatial data [14]. We have used data model and data entities levels of semantic annotation. The semi-automatic mechanism is used, algorithms of how the annotation is done are provided, and there is little human intervention in the process.

Applying ontology terms to the GIS data means create a link or association between the ontology terms and the geospatial data [14]. We have created a user interface in the web application to do that. The user of this user interface must have more experience in how to apply the correct ontology terms to the desired geographic terms. The annotation service's goal is to link the ontology concepts to vector geospatial data. The annotation method employed in the development of semantic annotations should be as uniform as possible considering all kinds of content, but flexible enough to allow exploitation of the semantics of each content. Within the annotation services, annotation procedures are defined for the different kinds of geospatial data sources and the level of semantic annotation being used. For the purpose of this research we used level 2 and level 3, semantic annotation of data model and data entities. The procedure is stored as a workflow. Each workflow consists of annotation schema to be used, the ontology concepts describing the data and the steps on execution and storage of generated annotations. Once the annotations are generated and linked to ontology concepts they are published and stored separately in the database.

3.5 Semantic Search:

Semantic search is applied on the semantically annotated data. The user can enter the search criteria in free text form, and then click the search button. We have build a semantic search functions that extracts the main classes and properties from the search text, and then access the postgresql database which store the semantically annotated GIS data to retrieve the records that match the search criteria. The query returns results meeting the criteria. The user interface allows display of linked concepts as well and map content of retrieved data. The sequence of activities which is done in semantic search is shown in Figure 3.

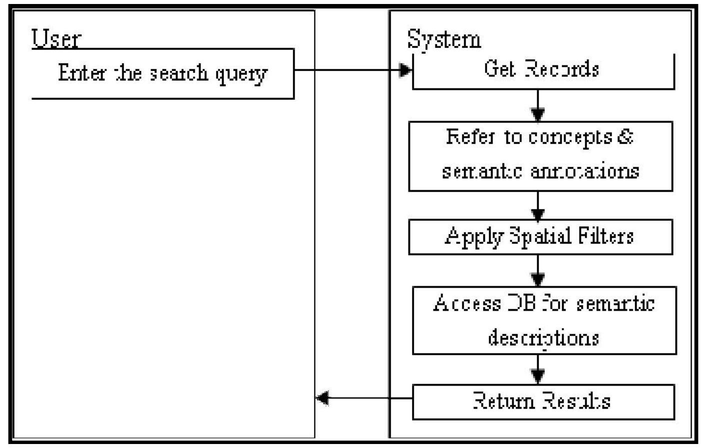

Figure 3. Activity Diagram for Semantic Search

3.6 Search Results Visualization

Certain landmark from the search result can be displayed over the map with a marker to indicate its location. Also the semantic description can be displayed in natural language as formulated in semantic annotation step. Information about the selected landmark can be displayed as an indented list hierarchy using the Tree View control in Asp.net. Also the landmark can be visualized using the different visualization options for the ontology graphs such as space tree, hyper tree. The ontology visualization is implemented using the JIT java script library.

4. Experiments and Results

The ontology service and annotation service form extra components in the semantically enabled GIS environment. The ontology service is an environment for developing ontologies (protégé). The semantic annotation service links different kinds of geospatial data to their ontologies. The ontology terms to be used in the annotation are specified and the content to be annotated is known. The semantically annotated GIS data are in natural language and the use of ontology terms guarantee uniqueness when each field is associated with a concept that semantically represent their content. Subject of semantic annotation are spatial information objects that are stored in spatial databases. They can be analyzed with spatial processing methods. Applying spatial analysis methods for information extraction avoids problems of semantic heterogeneities encountered with textual (symbol) analysis, which is obviously restricted by the ambiguities of natural language.

The web application user interface provides an abstract interface which supports the discovery and access of geospatial information. The user interface is implemented in such a way that the user enters query using free text search. The proposed ontology visualizer enables the hierarchy of the related concepts to be displayed. This allows the user to navigate about the concepts and find a suitable search term.

4.1 Search and Visualization of Results

Current search mechanisms have been based on key words only without using the semantics and hierarchical nature of concepts. The results display the frustration user's face when they have to try different keywords to get the desire results. In the ontology environment though search is by keywords, the presence of semantic annotations assists the user in query formulation and results are therefore obtained much easier than in an environment where information is not semantically defined.

4.2 Proposed Semantic Search

The user gives a query to the system in the user interface as shown in Figure 4, then the system translate the given query into triples (subject, predicate, object) Then looks for the answer using the terms applied in the semantic annotation process. The user query is converted to sql query which is executed in the postgres database at which the semantically annotated data are stored to retrieve the desired results.

The ontology visualizer displays the hierarchy of concepts related to the search terms as shown in Figure 5. The user is able to navigate about the hierarchy. He is able to get more information from the semantic annotation done and the hierarchy helps him distinguish between different terms. In the semantic annotation environment, the result returns all related concepts displayed in the visualizer. Navigating about the concepts helps the user formulate a better query. The method of ontology visualization depends on the amount of detail required by the user, level of annotation and storage capability of the geo-catalogue. Any visualization method chosen should be flexible enough to allow a general overview or detailed view of concepts.

The following are the sequence for the semantic search and results visualization:

- The user enters the free text search"أقرب مستشفى إلى بنك باريس" as shown in Figure 4.
- The semantic search function retrieves the main classes from the search criteria
- Check if the search text contains a class name.
- Check if the search text contains equivalent names for the classes' names by retrieving the classes' attributes (which are defined in the ontology file). The "Hospital" and "Bank" classes are retrieved. Then retrieve all records from semantic annotation table that have the same class's ids in the subject or object values.
- Check if the search text contains a property (relation or predicate).
- Check if the search text contains equivalent name for the property or relation name. The "Is Located Near" property is retrieved.
- Get the remaining words (not matched words). We will get "بنك باريس".
- Check for matching the remaining words with the values in the semantic annotation table. " بنك مصر باريس" will be returned because there is no bank with the name "بنك باريس".
- Also the nearest hospital to that bank will be returned.
- The search result can be visualized on the ontology viewer as shown in Figure 5.

Figure 4. Semantic Search

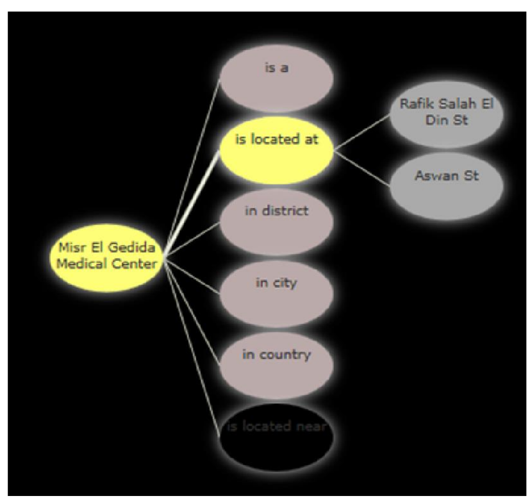

Figure 5. Ontology Visualization

- The semantic description and the indented list can be displayed for the search result as shown in Figure 6.

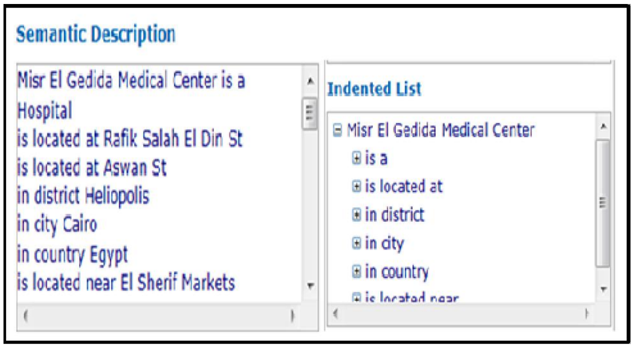

Figure 6. Semantic Description & Indented List Visualization

5. Discussions, Conclusions and Future Work

5.1 Discussions

Spatial Data infrastructures have created an open environment for sharing resources through the web, yet users have continued to face challenges in discovering the resources due to the inefficient search methods, vocabulary used and poor query formulation. The search is limited to keyword match. To overcome this problem, focus has shifted toward the semantic web where information is well defined. In this research we explored ways of bringing together semantic web technologies and geographic standards so as to improve the search functionality in GIS.

According to the research study, we were able to demonstrate the shortcomings of the search mechanisms based on keywords or substring matching. Therefore, keywords return few results, whereas substring matching returns too many results.

We demonstrated that when geographic standards and semantic web technologies (ontologies and annotations) are brought together, the search functionality within GIS improves. Inasmuch as geospatial catalogues play a pivotal role in discovery of geospatial data and services in distributed environments, they have limitations with regard to the searching way. To overcome this limitation, traditional search mechanism need to be augmented with semantic web technologies to improve data and service discovery. The semantic web technologies add meaning to resources.

Ontologies being the cornerstone of semantic web, captures relationships between concepts thus ensuring semantic interoperability in dynamic environments. In the research study we developed an ontology called Places based on spatial relations domain. The concepts used came from terms used in Ordnance Survey Spatial Relations V0.2 forecasting [13]. We explored methods of visualization of ontologies using different techniques which give users the freedom to choose the level of detail they intend to view. We proposed strategies for integrating ontologies and semantic annotation in a GIS system as a proof of concept in which semantic web technologies improve search for data complimented by visualization of ontology concepts that represent landmarks. Using the Spatial Relation ontology, we demonstrated how semantic annotation could be done at data model level and data entities level by adding pointers to the domain concepts and how the annotations assist users even non-experts in query formulation thus discovering the right information.

Semantic annotations establish a link between geospatial resources and ontologies. Annotations can be achieved at three levels and we have used level 2 (data model) and level 3 (data entities). The task of being able to formulate good queries is taken off the shoulder of the consumer through ontology visualization. User queries are referenced to the annotations. The query is executed against the meaning of supplied descriptions instead of the usual keyword based matching.

The traditional approach where users supply keywords or specify a bounding box when searching for information misses some key information that may be required for the purposes of decision making. To improve the search functionality of GIS, we proposed methods of integrating ontologies and semantic annotations in the system architecture.

5.2 Conclusion

With reference to the objectives that have driven this research project the following conclusions can be drawn:

- Integration of geographic data standards and semantic web technology results in improved search functionality.
- Standardized and rich geospatial ontologies increase the usefulness and effectiveness of geospatial data published through geo-catalogue services. The relationships existing between concepts can be made use when searching for services/data thus increasing the search space.
- Ontology visualization improves query formulation, in that a user is able to interact with the hierarchy of concepts which are linked to resources through their data descriptions from which they can choose the appropriate search concept.
- GIS system allows better search and retrieval of spatial data when integrated with semantic web technologies.

5.3 Future Work and Recommendations

Based on the results of this research, the following points are recommended for future development of geo-catalogues.

- Extend the functionality of geo-catalogue services for storage of semantic workflows in the same which automatically annotates resources to their ontology concepts.
- Apply the semantic annotation to the roads layer to be able to formulate the search query in the form "nearest landmark to certain.

Corresponding author
Huda Abdalslam Banour
Faculty of Computers and Information, Helwan University, Cairo, Egypt
mmmh13282@yahoo.com

References

[1] Klien E., M. Lutz, and W. Kuhn. Ontology-based discovery of geographic information services - an application in disaster management. Computers, Environment and Urban Systems, 30:102–23, 2006.
[2] Lemmens R.. Lost and found, the importance of modeling map content semantically. International Perspectives on Maps and the Internet, 24:377–397, 2008.
[3] Egenhofer M.J.. Toward the semantic geospatial web. In Proceedings of the 10^{th} ACM international symposium on Advances in geographic information systems, pages 1–4. ACM New York, NY, USA, 2002.
[4] Lemmens R. L. G. and M. de Vries. Semantic description of location based web services using an extensible location ontology. In Geoinformation and Mobility, in IfGI prints, pages 261–276, 2004.
[5] Klien E. and M. Lutz. An architecture for ontology-based GI discovery and retrieval of geographic information. Ontology-based Discovery and Composition of Geographic Information Services, 179(188):94–pages, 2004.
[6] Yue P., L. Di, P. Zhao, W. Yang, G. Yu, and Y. Wei. Semantic augmentations for geospatial catalogue service. In Proceedings of the 2006 IEEE International Geoscience and Remote Sensing Symposium (IGARSS06), volume 31, pages 3486–3489, 2006.
[7] Sabou C.F.M. and F. van Harmelen. Ontology-based information visualization: Toward semantic web applications. Visualizing the Semantic Web: XML-based Internet and information visualization, pages 45–pages, 2006.
[8] Dolbear C. and G. Hart. Creating Spatial Information Infrastructures: Towards the Spatial Semantic Web, chapter Opportunities and Challenges in Exploiting Semantics as an Aid to Information Integration, pages 89–101. CRC Press, 2008.
[9] OGC. Open GIS Catalogue Services Specification. http://www.opengeospatial.org/standards/cat: Accessed 12-11-2009, February 2007.
[10]W3C. Semantic Annotations for WSDL and XML Schema. http://www.w3.org/TR/sawsdl/, August 2007.
[11] Brogi A., S. Corfini, and R. Popescu. Semantics-based composition oriented discovery of web services. ACM Transactions on Internet Technology (TOIT), 8(4):1–39, 2008.
[12] Ordnance Survey (2001) Ordnance Survey Master Map Real World Objects Catalogue, v. 1.0, November 2001.
[13] OS GeoSemantics Team Definition
[14] Juliet Gwenzi., Enhancing Spatial web search with Semantic Web Technology and Metadata Visualisation, March, 2010
[15] A faceted ontology for a semantic geo-catalogue, Feroz Farazi1, Vincenzo Maltese1, Fausto Giunchiglia1 (DISI - Università di Trento, Trento, Italy), Alexander Ivanyukovich (Trient Consulting Group S.r.l., Trento, Italy)
[16] Ivanyukovich A., F. Giunchiglia, V. Rizzi, V. Maltese. SGC: Architettura del sistema. Technical report, TCG/INFOTN/2009/3/D0002R5, 2009.

1/12/2013