Using Semantic Similarity to Improve Information Discovery in Spatial Data Infrastructures

Fabio Gomes de Andrade[1], Cláudio de Souza Baptista[2]

[1]Departamento de Informática – Instituto Federal de Educação Ciência e Tecnologia da Paraíba (IFPB)
58900-000 – Cajazeiras – PB – Brazil

[2]Departamento de Sistemas e Computação
Universidade Federal de Campina Grande (UFCG) – Campina Grande, PB – Brazil

fabio@ifpb.edu.br, baptista@dsc.ufcg.edu.br

Abstract. *In the recent years, several works have been proposed with an approach to the use of semantics to improve the process of discovering geographic resources offered by spatial data infrastructures. However, semantic queries may return a large number of results, what causes the necessity for efficient ways to evaluate the relevance of each result retrieved. This paper proposes a framework that uses ontologies and thematic relevance to suggest a measurement that allows evaluating how relevant is each resource offered by the infrastructure to the user's query. This feature allows the results retrieved in a query to be organized through a ranking, in such a way that the most relevant resources are presented to the user first.*

1. Introduction

In recent years, spatial data infrastructures (SDIs) [Williamson et. al 2003] have been developed in order to ease the discovery of spatial data and improve the interoperability of spatial data supplied by different information sources. The development of open standards to the geospatial domain has reduced the problems concerning data interoperability. However, discovering information which is already available at spatial data providers is still a hard task.

A limitation of current SDIs is that their catalog services perform their queries based uniquely on keywords. This characteristic leads to the execution of queries with low recall, as the resources described with terms related to the keywords used to generate the query are not retrieved. Also, low precision is obtained, since many irrelevant resources which have the terms of the query in their description end up being retrieved. The difficulty in locating existing information makes many companies to spend much time and money with the production of data already made available by other providers and which could be made at lower costs or, in some cases, with no cost.

As a way to overcome the limitations of the present catalog services, it is increasingly common the application of semantic web concepts to improve the discovery of spatial data. The objective of the semantic web [Berners-Lee et. al 2001] is to use formal means to describe the semantics of resources published in the web, improving the data sharing among applications. The semantic web principles have been implemented through ontologies [Guarino 1995]. Ontologies are formal conceptualizations of an application domain, which makes its semantic understandable to both human and machines.

Usually, applications that use ontologies to discover information use an approach based on a semantic relationship known as *subsumption*. This kind of solution consists in locating all resources whose description is *subsumed* by a certain search concept. The great advantage of this kind of solution is that it improves the recall of queries, since inference rules may be used for runtime inference of new knowledge. Nevertheless, this kind of solution considers that all of the retrieved results have the same relevance to the user. As a semantic query may return a large number of results, information which is more relevant to the user may be shown at the end of the result and, eventually, may not even be analyzed. For example, if a user requests feature types about a concept *WaterCourse*, feature types associated to subclasses of this concept (which offer only a part of the information requested), can be presented earlier than types linked exactly to the concept search, which are probably more relevant to user. This feature produces the necessity for the development of mechanisms that permit the evaluation of the relevance that each retrieved result has to the user.

To tackle this limitation, this paper proposes an approach that uses notions of similarity as a way to improve the discovery of information in spatial data infrastructures. The main contribution consists of the development of a similarity measurement that enables to evaluate the relevance of each resource featured by the SDI to the user's query. Still, the paper develops a new measurement to evaluate similarity among concepts defined in ontologies, and shows how some ideas of classic information retrieval can be reused and adapted to the spatial domain.

The remaining of the paper is organized as follows. Section 2 discusses related works. Section 3 addresses the process of semantic annotation used to describe semantics of the feature types. Section 4 describes the approach used to evaluate the relevance of each resource offered to user query. Section 5 shows the implementation and the results obtained. Finally, section 6 concludes the paper and highlights further work to be undertaken.

2. Related work

Over the years, several works have been proposed to solve the problem of the discovery of information in SDIs. Though these works are related to the same research area, they differ from each other with respect to the type of resource discovered and the approach used to discover these resources.

[Smits and Friis-Christensen 2007] developed a work for discovery of data in SDIs. In that work, the semantic annotation of the resources is done by associating them to concepts defined in a thesaurus. After that, the user can browse the terms of the thesaurus to discover resources associated to it. In another work [Stock et. al 2010], the feature types of the SDI, as well as the operations that can be performed with them are described through a feature type catalog (FTC), which maintains links to the services that implement each operation defined in the FTC. The retrieval of information is done by browsing this FTC. [Lutz and Kolas 2007] developed a work that uses rules to perform the semantic annotation and the retrieval of spatial data spread on different data sources. In another work, [Lutz et al. 2008], WFS data types are semantically annotated through mapping records. After that, a reasoner using Description Logic is used to retrieve feature types which are subsumed by the user's search concept. In another work [Klien et al 2006], the authors use Comprehensive Source Descriptions to describe the semantic features of geoservices used in the discovery of data to solve the management of disasters. All of the works cited above improve the discovery of resources in SDIs by using semantics to describe their resources. However, they no not offer the means to evaluate the relevance of each retrieved resource. Other important works present the same limitation, such as [Athanasis et. al 2009], [Lutz and Klien 2006] and [Wiegand and Garcia 2007].

Janowicz et al [Janowicz et al, 2008] developed a similarity-based solution to discover spatial data supplied by SDIs. The proposed work uses a framework to evaluate the similarity between a concept defined in the user's query and the concepts associated to spatial data types offered by the infrastructure. However, this work is directed towards a very specific ontology language. Besides, it does not take into account the relevance of the theme during the retrieval process. This way, resources annotated with the same concept are judged with the same relevance. This is a drawback, especially in queries that return a large number of results.

The analysis of related work shows that the discovery of information in SDIs is still an open problem. The use of ontologies enables to explore semantics to enhance the quality of queries. However, queries can produce a large number of results which need to be evaluated by the user before being completely retrieved. When this happens, data that are potentially more relevant to the user can be shown at the end of the result, and cannot be judged by the user. This problem leads to the necessity to develop efficient mechanisms to evaluate the relevance of each result retrieved. Such solution is based on a ranking approach, in which more relevant resources are presented first. This ranking reduces the time spent during the result evaluation process and facilitates both data discovery and reuse.

3. The semantic annotation proccess

Before describing the proposed approach for evaluating the relevance of spatial resources supplied by SDIs, it is necessary to understand the kind of information that can be discovered. Aiming to standardize the access to geographic data offered by several spatial data sources, SDIs offer a set of web services that enables the access to spatial data in several different formats. Examples of such services are: Web Map Service (WMS), for the access to vector maps layers; Web Feature Service (WFS), for the access to spatial data in GML format; and Web Coverage Service (WCS), to provide access to raster data. The approach proposed by this paper focuses on discovery of feature types offered by geospatial services, in which each feature type can be a vector map layer, a GML feature type or a raster image, depending of the kind of the offered service.

The framework has a relational database which contains information about all spatial data services offered by information sources registered in the infrastructure. This information is registered at the time a data source registers its resources in the infrastructure. When the service is registered, the framework stores on its database the information about each data type it offers. For each data type offered, the framework stores information such as name, title, textual description, the type (vector map layer, feature type or coverage) and the bounding-box of the geographic region it covers. All this information is retrieved automatically at the time the service is registered, through the execution of its *getCapabilities* operation.

After this information is retrieved, the framework shows to the user, who is registering the service, a page containing the information about all data types available. Then, the framework asks the user to perform the semantic annotation of all types offered. For that, the user must choose, among the existing concepts in the domain ontologies used by the infrastructure, the one that better represents the information supplied by that data type. The URI of the concept chosen for annotation of each data type is stored together with its information and used during the process of information discovery. For example, a feature type that offers information about water reservoirs may be annotated with the concept *WaterCourse*, while one that describes only rivers may be annotated with the concept *River*. After all data types are annotated the registering is finished, and the information concerning the service and the data types it offers becomes available to the discovery process.

4. An approach to evaluate relevance

The main objective of the work described in this paper is the description of a measurement that permits to evaluate how relevant each data type offered by the infrastructure is for the user's query. The verification is done in three steps:

(i). to verify how much the concept used to annotate the data type which is under evaluation is similar to the search concept defined by the user;

(ii). to verify how relevant the theme requested by the query is to the spatial service that offers the data type under evaluation; and

(iii). to combine the values of both measurements to evaluate the relevance of the data type under evaluation for the query.

4.1 The information discovery process

The information discovery process occurs in the following manner as depicted in Figure 1. In a graphic interface, the user selects a theme and a geographic region of interest. This theme represents one concept defined in one of the domain ontologies used by the infrastructure. During the information discovery process, this concept is called search concept (SC). After, a spatial query is executed to filter, among the available data types, those whose bounding-box intersects the geographic region defined for the query. After that, the relevance of each filtered data type for the user's query is evaluated. Once a query can produce too many results, a threshold value describing the minimum desired relevance is defined, and the next stage in the discovery process is to filter the data types that have this degree of relevance greater than the threshold. Finally, the final results are organized in descending order of relevance and presented to the user.

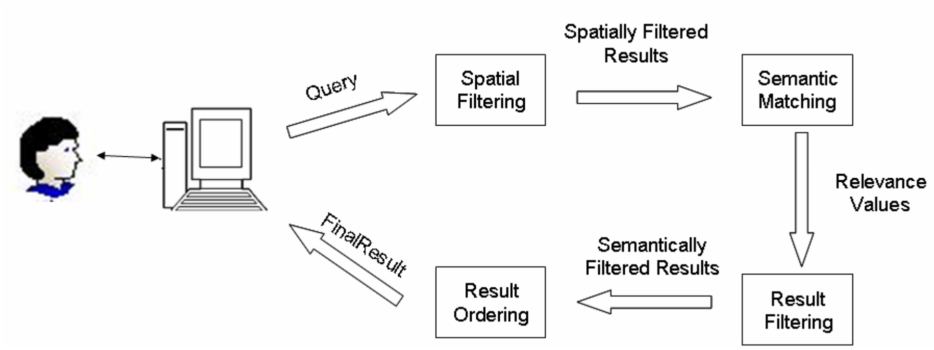

Figure 1. The information discovery process

4.2 Evaluating the similarity between the concepts

The first stage of the process used to measure the relevance consists in evaluating the similarity between the search concept of the user's query and the concept chosen for the semantic annotation of the data type under analysis. The approach used to evaluate this similarity has the following characteristics:

- **support to other types of relationships:** many works that propose to evaluate the similarity between concepts consider just the inheritance relationship. However, other relationship types, such as composition, cannot be neglected, since they denote an association of the concepts involved. As the composition is not such a strong relationship as the inheritance, weights are necessary to distinguish the relationship types. For example, two concepts associated by an inheritance relationship must have a degree of similarity greater than that existing between two concepts associated by other relationship type.
- **asymmetry:** symmetrical similarity measurements consider that the similarity between two pairs of concepts is the same, independently of the comparison order. However, for the problem studied in this paper, it was considered that the symmetry is not an interesting feature. For example let us suppose that a concept B is a sub-concept of a concept A. We may state that all data associated to B are relevant to the user, since every instance of concept B is also an instance of concept A. Nevertheless, if the user is looking for data associated to concept B, not all data associated to concept A are relevant to the user, since not all instances of concept A are instances of concept B. This characteristic requires, in the second case, that the symmetry must be smaller than in the first case, due to the existence of information of no interest to the user. The same idea is applied to the composition relationship;
- **degree of generalization:** ontologies are described through concepts that are organized in a hierarchical form, through inheritance relationships. Let us suppose that there is a hierarchy from a search concept SC in the ontology under analysis. As we go through this hierarchy, we find concepts that are more and more specialized with respect to SC and, consequently, have more difference with respect to it. Thus, the similarity of concepts must decrease gradually as the deepness of the concept increases.

The evaluation of the similarity between concepts is performed through a semantic network, generated from the parsing of the ontology at the time it is added to the SDI. The construction of this network takes into consideration two types of semantic relationship existing between the ontology concepts: inheritance and composition. The following algorithms in Table 1 present how a semantic network may be generated. The first algorithm is used to start the process of generation of the network and the second one to expand the production of the network to new concepts obtained from new concepts which are processed by the algorithm. In the first algorithm, the network is generated from each root concept (RC) in the ontology. A root concept is a concept that has no superclass in the ontology.

Table 1: Semantic Network Generation Algorithm

```
generateSemanticNetwork(O: Ontology): SemanticNetwork;
begin
      SN = new SemanticNetwork();
      rootNode = createNode("Thing");
      SN.addNode(rootNode);
      for each RCi in O do
      begin
           newNode = createNode(RCi)
           SN.addNode(newNode);
           SN.addSubclassEdge(rootNode, newNode);
           expandSemanticNetwork(SN, newNode, O);
      end;
      return SN;
end;

expandSemanticNetwork(sn:SemanticNetwork,currentNode:Node,
O:ontology): void;
begin
      SC = O.getSubClasses (currentNode.getConcept());
      for each SCi in SC do
      begin
          newNode = createNode(SCi);
          sn.addNode (newNode);
          sn.addSubclassEdge(currentNode, newNode);
      end;
      OP = O.getObjectProperties(currentNode.getConcept());
      for each OPi in OP do
      begin
          newNode = createNode(OPi.getRange());
          sn.addNode (newNode);
          sn.addAssociationEdge(currentNode, newNode);
      end;
end;
```

The result of the execution of the algorithms above is a semantic network which contains all of the concepts defined in the ontology and the semantic relationships existing between these concepts. In such network, nodes represent concepts, and arrows represent semantic relationships. The network produced has two kinds of arrow: one to define inheritance relationships and other to define composition relationships. Figure 2 shows a semantic network produced for the hydrographic domain, extracted from the GEMET ontology.

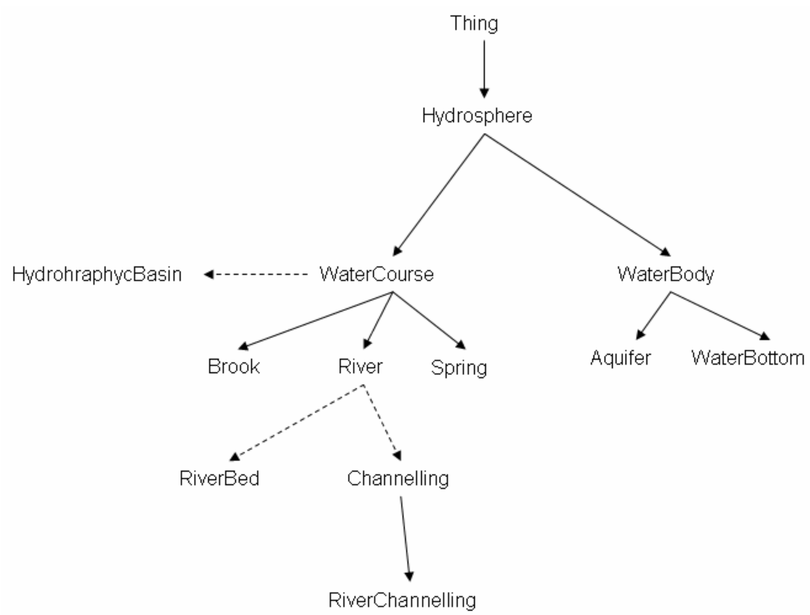

Figure 2. Semantic network produced for a hydrographic ontology

After the semantic network is produced, the framework evaluates the degree of similarity for all combinations of pairs of concepts defined in the ontology. This similarity is calculated through the analysis of the path that connects the two concepts under evaluation in the semantic network. The calculation of this similarity is performed taking into consideration two kinds of variables: the semantic relationship between the concepts and the distance between them in the network.

The objective of the semantic relationship between the concepts is to assign a greater degree of similarity to pairs of concepts that have stronger semantic relationships. As inheritance is a semantic relationship stronger than composition, concepts associated by an inheritance relationship must have a degree of similarity greater than that concepts associated by a composition relationship. To implement this constraint, a weight is assigned to each arrow of the network. For each node, two weights are possible: a normal weight and an inverse weight. The weight used to evaluate the similarity depends on the order of the concepts involved. This constraint is used to keep the asymmetry requirement. The use of two kinds of weight enables to ensure asymmetry, keeping the simplicity to discovery paths in graphs. Currently, the weight used for normal and inverse weights are, respectively, 0.8 and 0.6, for inheritance relationship, and 0.6 and 0.4, for composition relationship.

To perform the comparison between the two concepts, the first step consists in locating, in the network, the path that connects the two concepts. To allow the comparison between concepts Q and D, there must be at least one path from node Q that leads to D, or vice-versa. When none of these paths exist, the concepts are considered disjoint and the degree of similarity between them is assumed to be zero. When any of these paths can be found, the framework uses weights of the nodes in this path to evaluate the semantic relationship between them. Let W={w_1, w_2, ,,,, w_n} be the set of the weights of each arrow in the shortest path that connects the concepts Q and D in the semantic network that represents the ontology in which these concepts were defined. The value of the semantic relationship can be formally defined in Equation I. In order to ensure asymmetry property, the weights values depends on the order of the concepts in the path. If the path starts with the search concept Q and ends with the concept D, the normal weights are considered. However, if the path starts with the concept D and ends with the concept Q, the inverse weight are considered.

$$sem\mathrm{Re}l = \min\{w_1, w_2, \ldots, w_n\} \quad \text{(I)}$$

The second variable used to compare the similarity between the concepts is the distance between them. Rada et al [Rada et al 1989] introduce the semantic distance as a metric to evaluate similarity among concepts in semantic networks. The goal of this variable is to guarantee that pairs of concepts which are closer in the network have a greater degree of similarity compared to more distant pairs. We use this metric to implement the constraint to the degree of generalization between the concepts (Section 4.2). This measurement is inversely proportional to the degree of similarity, that is, as increases the distance between the concepts, the similarity between them diminishes.

After evaluating the semantic relationship and the distance between the two compared concepts, the values of these variables are used to measure the similarity between the concepts. To calculate these values, a weight is assigned to each of these variables. W_1 represents the weight assigned to the semantic relationship, while w_2 represents the weight of distance. The use of these weights makes the similarity between the concepts to be evaluated through the Equation II:

$$sim(Q,D) = w1 * sem\mathrm{Re}lationship(Q,D) + w2 * \left(\frac{1}{dist(Q,D)}\right) \quad \text{(II)}$$

The solution evaluates the degree of similarity between all pairs of concepts defined in the ontology (in both directions), generating a similarity matrix. The values of these similarities are stored in a relational database. Table 2 shows the similarity matrix for an excerpt of the concepts of the semantic network depicted in Figure 2. Concepts marked with S represent the concept defined in the user query, whereas concepts marked with D represent the ones used to annotate the feature type that is being evaluated.

Table 2. Similarity matrix

	Hydrosphere(D)	**WaterCourse (D)**	**River(D)**	**RiverBed (D)**
Hydrosphere (S)	1	0.84	0.74	0.54
WaterCourse (S)	0.68	1	0.84	0.58
River (S)	0.58	0.68	1	0.68
RiverBed (S)	0.38	0.42	0.52	1

4.3 Evaluating the degree of thematic relevance

Besides the degree of semantic similarity among the concepts involved in the query, we consider the degree of thematic relevance to improve the discovery process. The objective of this measurement is to evaluate how relevant is a theme requested in a query to the service which offers the data type under analysis. Through this measurement, data types offered by services in which the theme has more relevance are shown first to the user during the presentation of results. The value of this measurement is very important, since many data types are offered by several services, especially if the user's query requests a very common theme.

The degree of relevance that a certain theme has to a service is calculated through the normalized frequency, which is a measurement used in the classical information retrieval [Baeza-Yates and Ribeiro-Neto 1999] to evaluate the relevance of a certain term in a document. To evaluate this degree, the framework registers, at the time the service is registered, the normalized frequency of each theme offered by it. This way, the degree of relevance (*relDegree*) of a theme C to a service is calculated through the proportion of the number of times the theme occurs in the service (ni) and the number of data types offered by the service (N). The value of ni for a certain concept C is calculated based on the semantic relationships defined in the ontology. Such calculation comes from Equation III. In this equation, fi(C) is the number of occurrences of the concept under evaluation, and fi(S) and fi(SC) represent, respectively, the number of occurrences of a synonym concept of C and the number of occurrences of a concept that is a sub-class of C.

$$relDegree(C,S) = \frac{fi(C) + \sum fi(S) + \sum fi(SC)}{N} \quad \text{(III)}$$

The information of relevance of each concept is stored in the database of the infrastructure. The major advantage of keeping this information stored is that this allows us to accelerate the response time of the query. Such feature also helps us to keep the scalability of the solution for large amount of data.

4.4 Calculating the relevance

After defining the metrics used to calculate the degree of relevance of a data type to a user's query, the next step consists in defining how the values of these metrics will be used for that purpose. One possibility to solve this problem would be the representation of the user's query and the data type under evaluation as vectors in a bi-dimensional space and use the Euclidian distance to evaluate the similarity. However, this kind of metric represents similarity through a real number, corresponding to the distance between the points. Thus, in order to solve the problem, we adopted the sum of the values of the metrics, where a weight is assigned to each of the metrics. This technique, besides offering flexibility, since all weights may be altered to perform new queries, also offers similarity values between 0 and 1, which makes the evaluation of similarity more intuitive for the human being.

Thus, given a theme Q defined in the user's query and the theme D associated to the feature type under evaluation, the degree of relevance of this type for the query is calculated through Equation IV. In such equation, *sg* represents the degree of similarity of concepts Q and D, whereas *relDeg* represents the degree of relevance that the theme D has to the service by which the feature type under evaluation is offered. Finally w_1 e w_2 represent the weights that each type of measurement has to the calculation of spatial similarity. Each weight must have a value between 0 and 1, and their sum must always be equal to 1:

$$semanticSim(Q,D) = w_1 \times sg(Q,D) + w_2 \times relDeg(D,S) \quad \text{(IV)}$$

5. Implementation and results

To evaluate the proposed approach, a prototype was developed. The first step in this implementation was to define the domain ontologies that would be used for semantic annotation and discovery. In our experiments we have used ten domain ontologies, which were created from data models according to the Brazilian Spatial Data National Infrastructure. These ontologies are represented in OWL and the Jena framework is used to parse them. After that, we gathered several spatial services (WMS and WFS) offered by several providers throughout Brazil. Each service was processed and their information was stored in a database. Besides, we registered information concerning each feature type they offer. Each type was semantically annotated through the domain ontologies defined in the infrastructure. Currently, this database stores about 457 feature types, distributed among 21 geospatial web services, from 16 service providers. All this information is stored in a PostgreSQL/PostGIS database server.

To illustrate the results obtained during the evaluation process, let us suppose a simple query in which the user wants to find feature types regarding *WaterCourses* in Brazil. In the database used for evaluation, there are 70 feature types directly related either to this concept or to one of its subclasses. These types are distributed among the services offered by 8 different Brazilian sources: the National Water Agency (ANA), the Executive Agency of Water Management of the State of Paraíba (AESA), the National Agency for Electrical Energy (ANEEL), the Brazilian Institute for Environment (IBAMA), the Ministry of Fisheries and Aquaculture (MPA), the Protection System for Amazon (SIPAM), the Department of Water Resources of the State of Santa Catarina (SIRHESC) and the Federal University of Minas Gerais (UFMG), according to the Table 3. For each entry in the table, we have the provider name, the thematic relevance of the concept search (*WaterCourse*) to the service, the number of feature types annotated with the search concept, the number of feature types annotated with *WaterCourse* subclasses and the number of feature types annotated with concepts which are related to the search concept through a composition relationship.

Table 3. Data providers example concerning the hydrography concept

Provider	Thematic Relevance	Concept Search	Subclasses	Composition
ANA	1	2	4	0
AESA	0.5625	3	6	1
ANEEL	0.1621	5	5	0
IBAMA	0.1250	3	0	0
MPA	1	4	0	0
SIPAM	0.0517	2	2	0
SIRHESC	0.7045	29	2	0
UFMG	0.3076	0	3	1

After executing the query, 36 features types obtained a relevance degree greater than or equal to 90%. In this category were all feature types exactly annotated with the search concept. In this category, the types are listed in descendant order of thematic relevance. The types offered by the ANA and the MPA are listed first, with relevance of 100 %. After, the result shows the 29 types of data provided by the SIRHESC with relevance around 94% and the types offered by the AESA, with relevance around 91 % .

The second category contains feature types that have a relevance value between 80% and 90%. In this case there were two types of results. The first one contains data types that are associated with exactly the search concept, but are offered by services with low thematic relevance. Hence, there were the other two feature types offered by ANA, which had relevance of 81%. The second one is composed of data whose services have high thematic relevance, but they are annotated with concepts that represent subclasses of the search concept. In this case, we have data types offered by ANEEL, IBAMA and SIPAM. These types have gained importance around 83%, 82% and 81%, respectively.

The third category of results includes 18 feature types that have a relevance value between 60% and 80%. This category contains data types offered by services in which the *WaterCourse* theme is highly relevant, but have been annotated with subclasses of the search concept. The remaining data types offered by SIRHESC, AESA, UFMG, ANEEL and SIPAM are listed, in this order. Finally the last category contains the data types of services offered and AESA and UFMG that were annotated with concepts that have a composition relationship with the search concept. The relevance values were to 58% and 53% respectively.

6. Conclusion and further work

SDIs play an increasingly important role in the dissemination of geographic information offered by several organizations. However, locating geographic data offered by these infrastructures in an efficient and precise manner is still a hard task. Though ontology-based solutions have improved the discovering process, there is still the need to evaluate the relevance of the retrieved results to the user, such that the more relevant results can be exhibited first.

This paper described a framework that combines the notion of semantic similarity between concepts defined in ontologies and ideas applied to the classical information retrieval to evaluate how relevant are the spatial data offered by the infrastructure to an end-user's query. Though the present results have shown that the approach is interesting, some future works are still necessary.

On of the works necessary in the future consists in extending the notion of semantic similarity to treat more complex concepts and relationships, as, for example, concepts defined through conjunction, disjunction and negation of other concepts. Another important future work is to evaluate the user preferences once the result is presented. Besides to validate our approach, this work will enable to improve the weights used to calculate semantic similarity. Still, there is the need to evaluate the similarity between concepts defined in different ontologies, what can give an even better recall for the user's queries.

References

Athanasis, N., Kalabokidis, K., Vaitis, M. and Soulakellis, N. (2009) "Towards Semantics-based Approach in the Development of Geographic Portals". In: Computers & Geosciences, Volume 35(2), pages 301-308.

Baeza-Yates, R. and Ribeiro-Neto, B. (1999) "Modern Information Retrieval". Addison Wesley

Berners-Lee, T, Hendler, J. and Lassila, O. (2001) "The Semantic Web". http://www.scientificamerican.com/article.cfm?id=the-semantic-web, May

Guarino, N. (1995) "Formal Ontology, Conceptual Analysis and Knowledge Representation". In: International Journal of Human Computer Studies, Volume 43(5-6), pages 625 – 640.

Janowicz K., Wilkes M. and Lutz, M. (2008) "Similarity-based Information Retrieval and its Role within Spatial Data Infrastructures", International Conference on Geographic Information Science, Cova, T. V., Miller, H. J., Beard, K., Frank, A. U. and Goodchild, M. F., Germany, Springer, p. 151-167.

Klien, E., Lutz, M., and Kuhn, W. (2006). "Ontology-Based Discovery of Geographic Information Services - An Application in Disaster Management". In: Computers, Environment and Urban Systems, Volume 30(1), pages 102-123.

Lutz, M. and Klien, E. (2006) "Ontology-based Retrieval of Geographic Information". In: International Journal of Geographical Information Science, Volume 20(3), pages 233-260.

Lutz, M. and Kolas, D. (2007) "Rule-Based Discovery in Spatial Data Infrastructure". In: Transactions in GIS, Volume 11 (3), pages 317-336.

Lutz, M., Sprado, J., Klien, E., Schubert, C. and Christ, I. (2008). "Overcoming Semantic Heterogeneity in Spatial Data Infrastructures". In: Computers & Geosciences, Volume 35(4), pages 739-752.

Rada, R., Mili, H., Bicknell, E. and Blettner, M. (1989) "Development and Application of a Metric on Semantic Nets". In: IEEE Transactions on Systems, Man and Cybernetics, Volume 19 (1), pages 17-30.

Smits, P. and Friis-Christensen, A. (2007). "Resource Discovery in a European Spatial Data Infrastructure". In: IEEE Transactions on Knowledge and Data Engineering, Volume 19 (1), pages 85-95.

Stock, K., Atkinson, R., Higgins, C., Small, M., Woolf, A., Millard, K. and Arctur, D. (2010) "A Semantic Registry Using a Feature Type Catalogue instead Ontologies to Support Spatial Data Infrastructures". In: International Journal of Geographic Information Science, Volume 24 (2), pages 231-252.

Wiegand, N. and Garcia, C. (2007) "A Task-Based Ontology Approach to Automate Geospatial Data Retrieval". In: Transactions in GIS, Volume 11(3), pages 355-376.

Williamson, I., Rajabifard, A. and Feeney, M. F. (2003), Developing Spatial Data Infrastructures: From Concept to Reality, Taylor & Francis.