Mathematical Formulation for the Nile River Geographical Ontology

Enas M.F. El Houby
Systems& Information Dept., Engineering Division
National Research Centre,
Dokki, Cairo, Egypt

Ahmed M. D. E. Hassanein
Systems& Information Dept., Engineering Division
National Research Centre,
Dokki, Cairo, Egypt

ABSTRACT
Nowadays, the proliferation of geographic information systems has caused great interest in geographical ontologies. Geographical ontologies have been introduced to facilitate knowledge sharing and to assist in recognizing spatial terms employed in a query. In this research the full model of the Nile River geographical ontology has been developed to meet the needs of the recognition of the terminologies and the semantic relationships between geographical terms related to Nile River. It will enable search engine to perform a spatially aware search with providing support for query expansion and relevance ranking which will result in the retrieval of relevant web resources.

The proposed design of Nile River ontology model has been developed based on the analyzing and finding relations between different parts of Nile River. A mathematical form has been used to describe the Nile River geographical Ontology. Then the model has been implemented using protégé tool, the related data has been collected and finally the system has been tested.

General Terms
Geographic Information Retrieval (GIR), geographical ontologies

Keywords
Nile River ontology, OWL, Semantic Web, protégé

1. INTRODUCTION
The World Wide Web holds vast amounts of information. However users do not always get the information which they expect when searching the Web. One of the reasons for this problem is that a query is not expressed in terms that match the ones contained in some of the relevant documents. For example, synonymy between the terms of the original query and those of a relevant document do not match in letters. The documents are increasing in tremendous growth and so the need for more intelligent query tools is required. The Semantic Web is one of essential proposed solutions to resolve this problem so that the existing web searching systems can be extended to be efficient to find the resources in query.

Ontology, which plays a key role in the semantic web, offers an advanced approach for managing, retrieving and processing information. In modern computer science parlance, ontology is defined as a formal specification of a conceptualization of a particular domain as laid out by Gruber [1]. Ontologies are necessary for knowledge representation and knowledge exchange. This means that ontology describes concepts and relations that exist in a domain. To enable knowledge exchange, it is necessary to describe these concepts and the relationships between the concepts in a better way than just ordering them in taxonomy. However; taxonomies are important, since they form the "backbone" of ontology, but they are not enough for knowledge sharing [2].

With the increasing application of geographic information system (GIS), GIS is faced with the problem of inefficient management and incomprehensive application of the spatial information which comes from different resources and in different forms. In order to solve these problems, ontology is introduced into GIS field as a concept model which represents objects on semantic and knowledge level. In this context, Geographical Ontology (Geo-ontology) is defined as the formalization of concepts shared among the GIS field. It represents the concepts and the relationships between the concepts which are abstracted from real geographic space. Geo-ontology does not only describe spatial data using the semantic encoding method to be more easily understood by computers, but also integrates geographical data coming from different sources and in different forms for reasoning. After formalization, geo-ontology is readable to both humans and computers [3], [4].

A number of researchers tackle the semantic web and geo-ontology problems. Cardoso et al. [5] propose a method for the geographical expansion of queries exploiting spatial relationships. Another work [6] proposes the use of ontology to decide where and what should be searched from different data sources. Rui Wang & Günter Neumann [7] build geographic information retrieval systems (GIR) to retrieve documents for topics with a geographic specification. The authors focus more on the query expansion. They parse each topic into the event part and the geographic part and use different ontologies to expand both parts respectively. In [8] Gaihua Fu et al. study how to build a geo-ontology which plays a central role in intelligent spatial search on the Web. It serves as a shared vocabulary for spatial mark-up of Web resources. Their research is part of SPIRIT project. In [2], they present a method for designing ontologies that are based on formal concept analysis [9] which is a theory of data analysis identifying conceptual structures among data sets. This method allows the users to know how to discover the necessity for new concepts and relations in ontology. Farhan H. et al. [10] address policy interoperability issues across multiple organizations using geo-spatial domain dependent ontologies. Agustina Buccella et al. [11] propose a novel system (called GeoMergeP) to integrate geographic sources by formalizing their information as normalized ontologies. This process includes structural, syntactic and semantic aspects. Their integral merging process assists users in finding the more suitable correspondences. Shawn Bowers et al. [12] describe a novel approach to construct large ontologies of ecological terms (classes and properties) which allows ecologists to use common spreadsheet tools to describe different aspects of ontology. Many other GIR researchers, who are recognized in literature such as [13], [14], [15] and [16] use ontologies to assist spatial search.

In this paper a full description using mathematical formalization is presented for the "geographical ontology for the Nile River (NR geo-ontology)" model whose prototype has been previously introduced in [17]. In section 2, the design of the NR geographical ontology is described and mathematical formalization is presented. The implementation of the NR ontology using the Protégé is shown. In section 3, the usage of the ontology is discussed using the query tool of the protégé. Finally in section 4, a conclusion and the possible future work is drawn.

2. DEVELOPING THE NILE RIVER GEOGRAPHICAL ONTOLOGY

A list of the geographical places which faces the path of water in the NR as it goes from south to north is first collected. By examining the map of the Nile River and the available information about it over the internet, the places were categorized into body of water, body of land, city, country and human made structures. The geographic features of each place were systematically examined (i.e. how its existence affect the Nile river path) to group them into entities.

2.1 Analysis of the Nile River Geography

The reasoning which leads us to introduce the entities of NR model was discussed in [17]. By analyzing the meaning of each entity and its relation with the NR, the relations between the different entities of the places on the NR were introduced.

2.2 The Design of the Nile River Ontology

Each geographic entity described in the analysis is represented as a class as shown in figure1 in the hierarchical ontology model. The classification is based on the following two principles:

Geographic Motivation: A central class is created and called 'Rivers'. Any other class is created to serve the central class. For example, a river takes fresh water from a lake and throws it in the sea. It would be irrelevant here to classify the lake and sea under two different classes namely fresh water and salty water. However, classifying them under a class called 'Influx_Outflux_Point' would give the reader an idea about the relation of the lake and sea with respect to a river.

Practicability: The degree of granularity for breaking down classes into further subclasses is based upon the principle of practicability. The classification should be self-evident to a high degree. That is to say, the person reading the classification should not think long about it to understand the differences between the different classes. For example, classes should not split into further subclasses as a person can always argue about the difference between a village and a town or a river and a creek.

At this point, the necessity for a clear and formal definition of the NR ontology arises. The Nile River Ontology is defined in a mathematical form. The used mathematical definitions are borrowed from Stumme et al. [18]. However, the definitions are freely modified to serve the aims of this paper.

The NR ontology is a structure which can be mathematically described as follows:

NR: = {C, <c, R, σr, <R, A, σa, τ}

Where

C, R, A and τ are sets whose elements are class identifiers, relation identifiers, attribute identifiers and data types, respectively.

<c is called class hierarchy or taxonomy, and is used with the elements of C to describe which class is the super-class and /or subclass of the other classes.

A function σr is called relation signature, and it is used to describe the elements of R.

A partial order <R is called relation hierarchy, and is used to describe the dependence of relations of R on each other.

A function σa is called attribute signature, and is used to describe the elements of A.

τ is a set of data types such as strings, integers, etc.

If C1 <c C2, for C1, C2 ϵ C, then C1 is a sub-class of C2, and C2 is a super-class of C1.

If C1 <c C2 and there is no C3 ϵ C with C1 <c C3 <c C2, then C1 is a direct sub-class of C2, and C2 is a direct super-class of C1 which can be denoted as C1 <D C2.

If C1 <c C2 and there is C3 ϵ C with C1 <c C3 <c C2, then C1 is an indirect sub-class of C2, and C2 is an indirect super-class of C1 which can be denoted as C1 <I C2.

2.2.1 Mathematical Description of the Classes of NR Ontology

The Nile River Geographical Ontology includes the following set C of classes:

C := {Geographical_Area, Country, City_Province, Ecosystem, Swamp, Forest, Wetland, Influx-Outflux_Point, Delta, Falls, Highland, Lake, Sea, Rivers, Structures, Dams, Bridges, Internal, External, Geographical_Data }.

As depicted in figure1, these classes are arranged in a hierarchy or taxonomy with the top element of concept is NR.

Accordingly, The sets of direct relations <D can be described mathematically as follows:

{Swamp, Forest, Wetland}<D Ecosystem
{Country, Ecosystem, City_Province} <D Geographical_Area
Geographical_Area <D NR
Geographical_Data <D NR
{Delta, Falls, Highland, Lake, Sea} <D Influx-Outflux_Point
Influx-Outflux_Point <D NR
Rivers <D NR
{Internal, External} <D Bridges
{Bridges, Dams} <D Structures
Structures <D NR

Where the direct relation

{Swamp, Forest, Wetland}<D Ecosystem

means that the three classes {Swamp, Forest, Wetland} are direct subclasses of the class Ecosystem. In other words, the Ecosystem class is a direct super-class of the three classes {Swamp, Forest, Wetland}.

The set of indirect relations in NR ontology are:

{Swamp, Forest, Wetland} <I Geographical_Area
{Swamp, Forest, Wetland, Country, Ecosystem, City_Province, Delta, Falls, Highland, Lake, Sea, Internal, External, Bridges, Dams} <I NR
{Internal, External} <I Structures

Where the indirect relation

{Swamp, Forest, Wetland}<I Geographical_Area

means that the three classes {Swamp, Forest, Wetland} are indirect subclasses of the class Geographical_Area. In other words, the Geographical_Area class is an indirect super-class of the three classes {Swamp, Forest, Wetland}.

Accordingly, a set of class hierarchy <c relations which include direct <D and indirect <I relations is defined as:

<c := <D U <I

Where U means union.

Each of the following sets of classes is disjoint:

Set 1 := (Forest, Wetland, Swamp),
Set 2 := (Delta, Falls, Highland, Lake, Sea)
Set 3 := (Bridges, Dams)
Set 4 := (Internal, External)

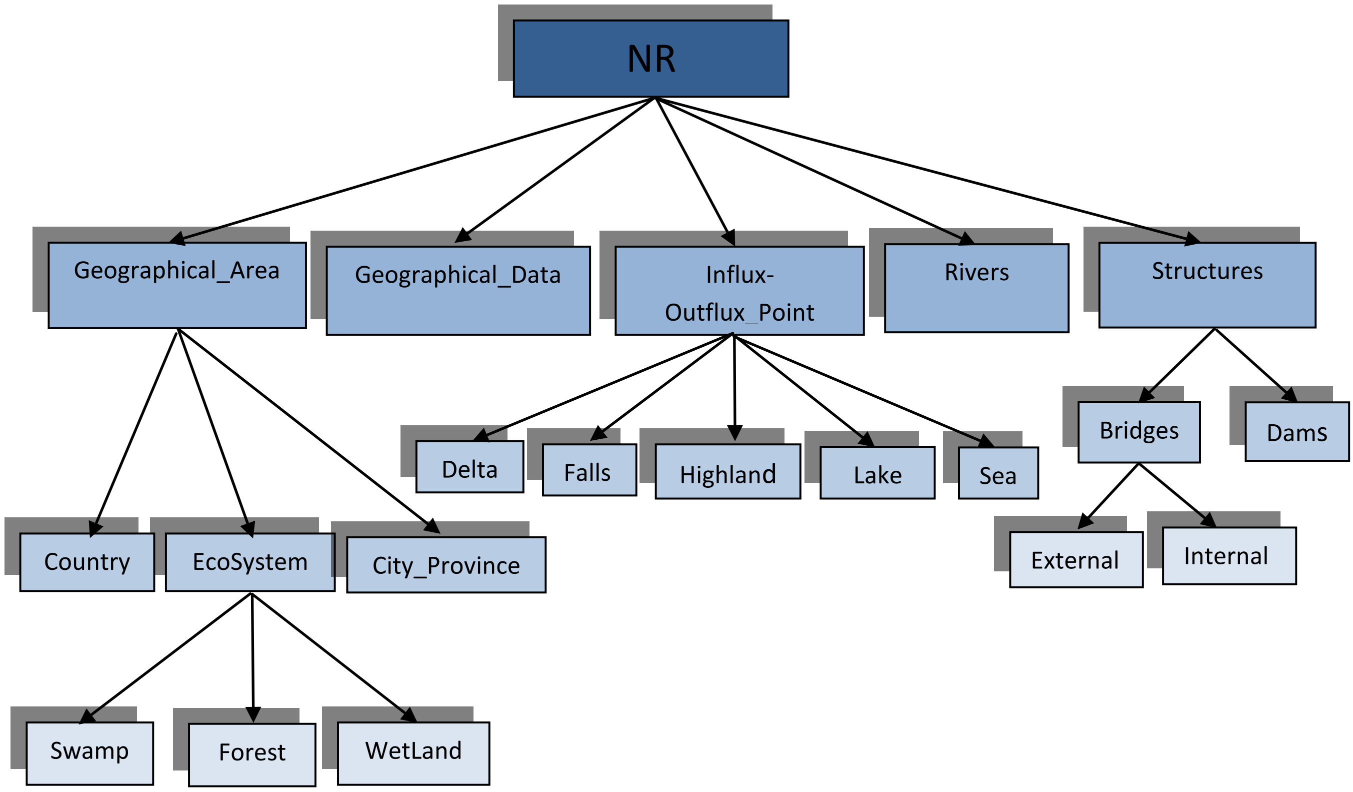

Figure 1: Class Hierarchy of the Nile River Ontology

Disjoint is denoted by “¬”, accordingly the first of the above disjoint sets must fulfill the following:

$\forall$ x {Forest(x) ¬ (Wetland(x), Swamp(x))}

and $\forall$ x {Wetland(x) ¬ (Forest(x), Swamp(x))}

and $\forall$ x {Swamp(x) ¬ (Forest(x), Wetland(x))}

where the disjoint relation:

$\forall$ x {Forest(x) ¬ (Wetland(x), Swamp(x))}

means that every element x which belongs to the class Forest can't belong to the classes Wetland or Swamp and so on for the other disjoint sets.

2.2.2 Mathematical Description of the Object/Data type properties of NR

The set of NR relations between classes (object properties) is:
R: = { Hindered_by, Built_across, contains, Passes_through, Distination_is, Is_fed_by, Feeds, Source_is, Followed_by, Preceded_by, Has_bridge, inCity, include, Is_part_of, Passes_over, Passes_under, Has_geographical_Data}

The description of signature σr of object properties is as follow:
σr (Hindered_by)= (Rivers, Dams) ⟺ σr(Built_across)= (Dams, Rivers)

The object property hindered_by has inverse property in the model named built_across.

Where σr(Hindered_by)= (Rivers, Dams)

means that Rivers are "Hindered_by" Dams. The property Hindered_by has domain Rivers and Range Dams. The inverse property of Hindered_by is Built_across. The Domain of the property Built_across is Dams and the range is Rivers. The property and its inverse is written as follows: σr (Hindered_by) = (Rivers, Dams) ⟺ σr (Built_across) = (Dams,Rivers)

where the bidirectional arrow '⟺' means 'inverse of'. The domain of property hindered_by is the range of property built_across and vice versa. The full set of object properties is shown in table1.

The set of attributes or data type properties are:
A:={Built_underConstruction_proposed(), Is_Basin_River_Country(), Latitude, Longitude, Synonyms}

The attributes have the following attribute signatures σa:
σa (Built_underConstruction_proposed) = (Structures, string)

The Built_underConstruction_proposed attribute describes the status of different structures across or above the Nile River. It is a string which can take one of the following possible values {Built, underConstruction, proposed, Removed}. The full set of data type properties is shown in table2. Figure 2 depicts the conceptual design of Nile River ontology which shows the different classes, properties and relations.

2.3 Implementation of the NR Geographical Ontology

Various ontology representation languages and tools have been used in the literature for specifying and modeling ontologies. One of these tools is a visual ontology modeling software namely Protégé 3.4.6 developed by Stanford university [19] which is built on the top of OWL language. Protégé 3.4.6 is an open source ontology development platform supporting features such as constraint checking. It has been used to implement the Nile River geo-ontology according to the proposed design. Different classes and properties have been implemented according to conceptual design shown in figure 2. According to Protégé Tool, the root class of the NR ontology which is called 'NR' is mapped to the root class 'Thing'. Both data type properties which describe classes and object properties which describe the relations between classes are implemented. Different conditions and disjoints classes have been specified. The NR ontology contains the 11 countries through which the Nile River passes or the Nile basin exists. In addition, 36 cities or Provinces inside the countries in which the Nile River exists are included.

Table 1: The full set of object properties

Object Properties	*Range* / **Domain**	*Domain* / **Range**	*Inverse*
Built_across	Dams	Rivers	Hindered_by
contains	Geographical_Area	Rivers	Passes_through
Distination_is	Rivers	Delta, Lake, sea, falls, Rivers	Is_fed_by
Feeds	Falls, Highland, Lake, Delta, Rivers	Rivers	Source_is
Followed_by	Country, City-province	Country, City-province	Preceded_by
Has_bridge	City-province	bridges	inCity
include	Country	City-province, EcoSystem, influx_outflux_point	Is_part_of
Passes_over	bridges	Rivers	Passes_under
Has_geographical_Data	Geographical_Area	Geographical_Data	

Table 2: The full set of data type properties

Data type Properties	**Domain**	**Range**	**Allowed values**
Built_underConstruction_proposed	Structures	String	Built, underConstruction, proposed, Removed
Is_Basin_River_Country	Country	String	Basin, River, Neither
Latitude	Geographical_Data	String	
Longitude	Geographical_Data	String	
Synonyms	Geographical_Data	String	

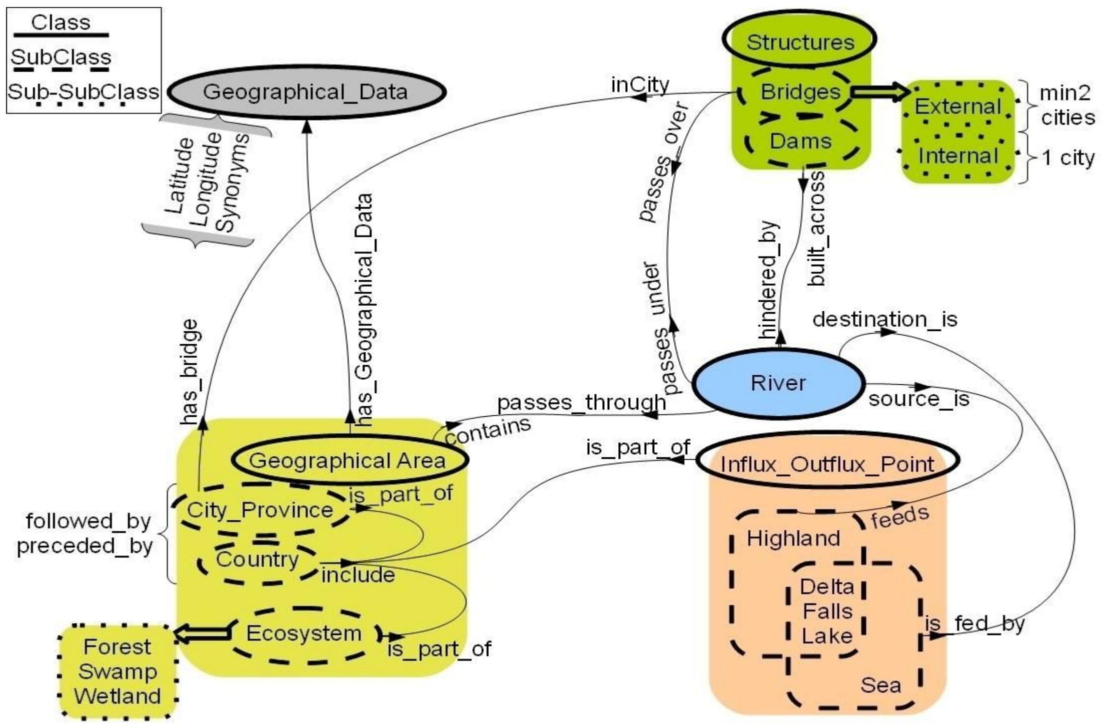

Figure 2: the class model of the NR Ontology

The OWL description for one of the countries on the NR is shown in figure 3. Burundi country is described as an instance under the class 'Country'. It includes a 'city_province' instance called Bururi_Province. Also, a lake called Lake_Tanganyika is shown to exist in the country. Also the data type property shows that it is both a Basin and river country. The geographical data can be seen for the country. Latitude and longitude at which Burundi exists can be seen.

A snapshot of NR Ontology developed using the Protégé Software is shown in figure 4.

```
<Country rdf:ID="Burundi">
  <has_Geographical_Data rdf:resource="#Burundi_"/>
  <include rdf:resource="#Bururi_Province"/>
  <include rdf:resource="#Lake_Tanganyika"/>
  <is_Basin_River_Country rdf:datatype="&xsd;string">Basin</is_Basin_River_Country>
  <is_Basin_River_Country rdf:datatype="&xsd;string">River</is_Basin_River_Country>
</Country>
<Geographical_Data rdf:ID="Burundi_">
  <Latitude rdf:datatype="&xsd;string">3.30 S</Latitude>
  <Longitude rdf:datatype="&xsd;string">30.00 E</Longitude>
</Geographical_Data>
```

Figure 3: The OWL description of Burundi country

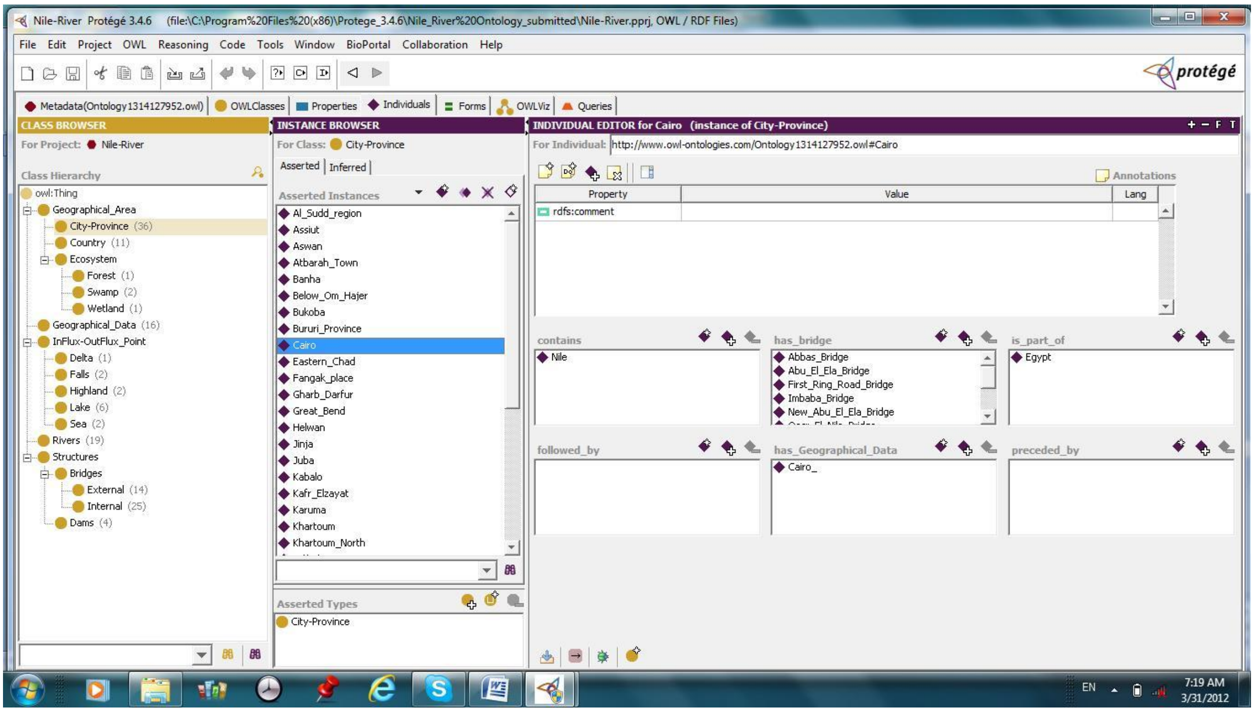

Figure 4: A snapshot of the Protégé window for the NR ontology

3. QUERY AND RESULT

In this section an application for the NR Geo-ontology model is presented. The Query tool provided by Protege_3.4.6 is used to query about the different geographical parts of the Nile River. The tool can be used to expand the terms of a searched word so that a refined meaning of the searched word can be reached. For example, the term Victoria_Nile alone is a vague term which when used in a search engine may return a lot of topics related to the river such as a social, economy or political activity related to the river. When the result of the following query (Q1) about Victoria_Nile is used to expand the search query it will retrieve meaningful results.

Q1 : built_across := {Nalubaale_Dam, Victoria_Nile}

where built_across is the relation property, Victoria_Nile is the name of the river under investigation and Nalubaale_Dam is the result of the query. That is to say, Query 1 (Q1) is used to find the dam which is built across the Victoria Nile River. If the result of Q1 namely 'Nalubaale_Dam' is used in addition to Victoria_Nile in a search engine, the results of the search engine would be more focused on the topic of interest of the user than if the term Victoria_Nile is used alone. Figure 5 shows the result of a query (Q1) run on Protégé.

The following queries were applied as well on the NR ontology:

Q2 : hindered_by := {Victoria_Nile, Nalubaale_Dam}
where in this case, hindered_by is the relation property, Nalubaale_Dam is the name of the dam under investigation and Victoria_Nile is the result of the query. Q2 is the inverse of Q1. The other queries follows:
Q3 : followed_by := {Ehtiopia, South_Sudan}
Q4 : preceded_by := {South_Sudan, Ethiopia}

Q5 : contains := {(Sudan, Tekeze) , (Ethiopia, Tekeze)}
Q6: Destination_is := {(Ruvyironza, Rusumo_Falls), (Nyabarongo, Rusumo_Falls)}
Q7 : feeds := {Ethiopian_Highlands, Blue_Nile}
Q8 : has_bridge := {Jinja, Nalubaale_Bridge}
Q9: inCity:= {(Khartoum_Tuti_Bridge, Tuti), (Khartoum_North_Tuti_Bridge,Tuti), (Omdurman_Tuti_Suspension_Bridge, Tuti)}
Q10: include := {Uganda, Pakwach}
Q11: is_fed_by:= {(Lake_Albert, Victoria_Nile), (Lake_Kyoga, Victoria_Nile)}
Q12 : passes_through := {Kagera_River, Bukoba}
Q13 : passes_under := {Blue_Nile, Soba_Bridge}
Q14 : Source_is := {Kagera_River, Rusumo_Falls}

Using the names of places under investigation into queries such as those mentioned above would expand the terms entered in a search engine. In this case, the results of the queries mentioned above are essential to increase the depth of the search and widen the initial data used in any search engine.

4. CONCLUSION AND FUTURE WORK

The proposed design of the Nile River ontology model is developed based on analyzing and finding relations between the different parts of Nile River. The proposed model is implemented using protégé tool, the related data is collected and finally the system is tested. Different queries where shown to test the ability of the ontology to expand the information collected about many parts of the Nile River. Future work will include expanding the work to have a search engine which could make use of the ontology described in this paper.

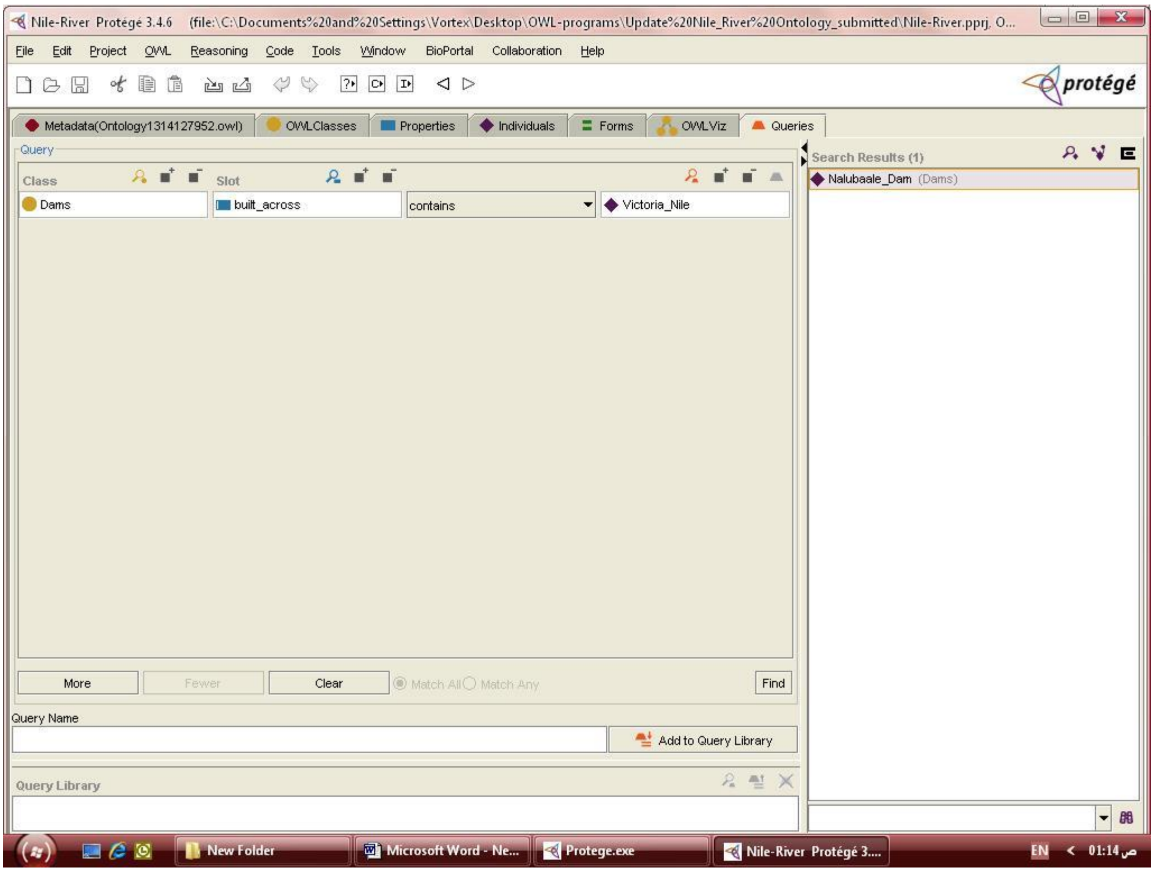

Figure5: A snapshot of Protégé showing the result of a query about what is built_across the Victoria_Nile River

5. REFERENCES

[1] T.R. Gruber. “A Translation Approach to Portable Ontology Specifications”, Knowledge Acquisition, vol. 5, pp. 199–220, 1993.

[2] Marek Obitko, V´aclav Sn´aˇsel, and Jan Smid, “Ontology Design with Formal Concept Analysis”, CLA 2004, pp. 111–119, ISBN 80-248-0597-9, VˇSB – Technical University of Ostrava, Dept. of Computer Science, 2004.

[3] Song Wei, Zhang Ming. "The compendious tutorial for semantic web", Beijing: higher education publishing company, 2004.

[4] Yandong Wang, Jingjing Daib, Jizhen Shenga, Kai Zhoua, Jianya Gong", Geo-ontology design and its logic

reasoning", Geoinformatics: Geospatial Information Science, Proc. of SPIE Vol. 6753, 675309, 2007.

[5] Cardoso, N., Silva, M.J., "Query Expansion through Geographical feature Types", In: GIR 2007, Lisboa, Portugal, November 9 (2007)

[6] Mata, F., "Geographic Information Retrieval by Topological, Geographical, and Conceptual Matching", In: Fonseca, F., Rodríguez, M.A., Levashkin, S. (eds.) GeoS 2007. LNCS, vol. 4853, pp. 98–113. Springer, Heidelberg (2007)

[7] Rui Wang and Günter Neumann," Ontology-Based Query Construction for GeoCLEF", LNCS 5706, pp. 880–884, Springer-Verlag Berlin Heidelberg, 2009

[8] Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty, "Building a Geographical Ontology for Intelligent Spatial Search on the Web", In Proceedings of IASTED International Conference on Databases and Applications, Springer Verlag, pp. 167-172, 2005.

[9] Ganter, B., Wille, R.," Formal Concept Analysis, Mathematical Foundation", Springer Verlag Berlin:. 1999.

[10] Mohammad Farhan Husain, Tahseen Al-Khateeb, Mohmmad Alam, Latifur Khan, "Ontology based policy interoperability in geo-spatial domain", Computer Standards & Interfaces 33, pp214–219, 2011.

[11] Agustina Buccella, Alejandra Cechich, Domenico Gendarmi, Filippo Lanubile, Giovanni Semeraro, Attilio Colagrossi, "Building a global normalized ontology for integrating geographic data sources", Computers & Geosciences 37, pp893–916, 2011.

[12] Shawn Bowers, Joshua S. Madin, Mark P. Schildhauer, "Owlifier: Creating OWL-DL ontologies from simple spreadsheet-based knowledge descriptions", Ecological Informatics 5, pp19–25, 2010.

[13] C. B. Jones and H. Alani and D. Tudhope , "Geographical Information Retrieval with Ontologies of Place", In Proceedings of the International Conference on Spatial Information Theory, pages 322–335. Springer-Verlag, 2001.

[14] W. Riekert. "Automated Retrieval of Information in the Internet by Using Thesauri and Gazetteers as Knowledge Sources", Journal of Universal Computer Science, 8(6):581– 590, June 2002.

[15] Vidal, V.M.P., Sacramento, E.R., de Macêdo, J.A.F., Casanova, M.A.: "An Ontology- Based Framework for Geographic Data Integration". In: Heuser, C.A., Pernul, G. (eds.) ER 2009. LNCS, vol. 5833, pp. 337–346. Springer, Heidelberg (2009)

[16] Zaki, C., Servières, M., Moreau, G.: "Combining Conceptual and Ontological Models for Representing Spatio-Temporal Data and Semantic Evolution in GIS". In: Ontologies for urban development: Future development of urban ontologies, Liège, Belgium, pp. 95–104 (2009).

[17] M.D.E. Hassanein, Enas M.F. El Houby, "Towards a comprehensive Nile River Geographical Ontology Model", Advances in Natural and Applied Sciences, 6(6): pp. 772-778, 2012.

[18] G. Stumme, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, Y. Sure, R. Volz, V. Zacharias. "The Karlsruhe View on Ontologies", Technical Report University of Karlsruhe, Institute AIFB, 2003.

[19] http://protege.stanford.edu/index.html.