Geographic Knowledge Representation Using Conceptual Graphs

Athanasios Karalopoulos, Margarita Kokla, Marinos Kavouras
National Technical University of Athens
157 80 Zografos Campus, Athens, Greece
http://ontogeo.ntua.gr
{akaralop, mkokla, mkav}@survey.ntua.gr

SUMMARY

Geographic concept definitions are a rich source of knowledge with special structure and content. Conceptual Graphs are a visual, advanced knowledge based representation formalism grounded on philosophical, linguistic and object-oriented principles (Sowa, 1984 & 2000). We introduce a straightforward and easy-to-implement methodology for the representation of geographic knowledge, expressed by definitions, using conceptual graphs. The proposed methodology consists of two steps: tagging and parsing. In the first step, we follow appropriate rules to tag every word in the concept definition expressing specific geographic knowledge. In the second step, we apply a deterministic algorithm in order to parse the tagged definition and create the corresponding conceptual graph. Hence, we take advantage of conceptual graphs' potentials to represent semantic information extracted from definitions of geographic concepts and explicate their semantics.

KEYWORDS: *Geographic knowledge, concept definition, conceptual graphs.*

INTRODUCTION

Geographic concept definitions are a rich source of knowledge with special structure and content. Concept definition analysis constitutes an essential step for deriving systematic knowledge about concepts. Various approaches in the field of Natural Language Processing are based on definition analysis (Jensen & Binot, 1987, Vanderwende, 1995). In the geospatial domain, definitions have been used in order to identify similarities and heterogeneities between geographic categories (Kavouras et al., 2003). Semantic interoperability is greatly facilitated by an effective computational representation structure of knowledge. Consequently, since concept definitions contain a great deal of a concept's knowledge, it is necessary to implement efficient representations of concept definitions.

Conceptual Graphs are a visual, advanced knowledge based representation formalism grounded on philosophical, linguistic and object-oriented principles (Sowa, 1984 & 2000). They are a diagrammatic and expressive way of knowledge representation that was firstly introduced for the representation of contents of natural language texts.

During the last years, research has been done in order to represent and extract information about geographic concepts. Approaches on geographic knowledge representation include methodologies that are based on analysing geographic concept definitions and finding effective representations. These can be found among others in (Kuhn, 2002, Kokla & Kavouras, 2002).

The purpose of the present research is to develop a methodology for the representation of geographic knowledge using conceptual graphs. Our goal is to introduce an algorithm that takes a geographic concept definition as input and produces the corresponding conceptual graph representation. By representing geographic knowledge using conceptual graphs, we will be able to take advantage of their potentials in making explicit the semantics of geographic concepts.

CONCEPTUAL GRAPHS

A conceptual graph is a network of concept nodes and relation nodes (Conceptual Graph Standard, 2002). The concept nodes represent entities, attributes, or events (actions) while the relation nodes identify the kind of relationship between two concept nodes. The main characteristics of conceptual graphs are:

- Concepts and relations replace predicates and arguments from predicate logic.
- A relation's signature defines what concept types it can relate.
- Concepts allow referents to specify an individual or a set of individuals of a certain type.
- A type hierarchy can be defined on concepts.
- Different graphs can be related through a coreference link on an identical concept.
- Manipulation algorithms are defined (maximal join, graph generalization) which make use of the type hierarchy to determine concept similarity for finding common information between different graphs.
- Quantifiers are dealt with: plural, for all, there exist, some, 9 exactly.

Figure 1 shows the conceptual graph representation of the following sentence: "*Pineios River which discharges into the Aegean Sea crosses the Tembi valley which is between the mountains Olympus and Ossa*".

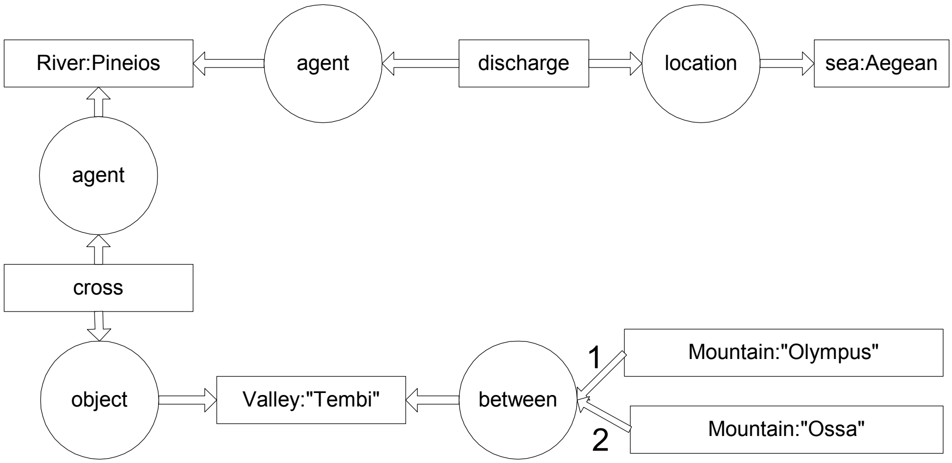

Figure 1: Conceptual Graph representation.

A lot of methods have been suggested for measuring the similarity between two conceptual graphs. Montes-y-Gomez et al. (2001) proposed a method for comparison of any pair of conceptual graphs. This method works in two main stages: matching conceptual graphs and measuring their similarity. Matching is mainly based on the generalization rules of conceptual graph theory while the second stage is based on the Dice coefficient.

EXPLORING CONCEPT DEFINITIONS

Every concept definition (not only a geographic one) is usually given by a few sentences that contain up to three general types of information: the genus, the differentia and a specific example. The genus, or hypernym, specifies the class in which the concept is subsumed and contains information that is frequently used for concept taxonomy construction.

The differentia specifies how different that concept is from the other concepts in the same class. It is a set of attributive adjectives and prepositional phrases that differentiates words with the same genus. It can also provide the purpose, the location, the look and many other aspects of general knowledge through the existence of one or more sentences (sub-clauses), each one giving a different kind of general information.

Finally, the example presents a typical situation using the concept defined and it involves specific instances of other concepts. Table 1 shows the genus, the differentia and the example of the definition: *A River is a large natural stream of water used for transportation and irrigation ('Pineios river is navigable for 30 miles")*. This is the definition of the concept *'River'* as it appears in lexical databases.

Genus	*Stream*
Differentia	*large, natural* (attributive adjectives) *of water* (prepositional phrase) *used for transportation and irrigation* (sub-clause)
Example	*Pineios river is navigable for 30 miles*

Table 1: Definition's genus, differentia and example.

In the next sections we consider that every definition of a geographic concept consists of three parts: the *main*, the *secondary* and the *example* part. The *main* part of the definition is the sentence that contains the genus, its attributive adjectives and the prepositional phrases describing the genus. The *secondary* part contains the given sub-clauses, which describe further the geographic concept and the *example* part of the definition is the example itself.

Table2 shows the *main*, the *secondary* and the *example* part of the previous definition:

Main part	A River is a large natural stream of water
Secondary part	used for transportation and irrigation (sub-clause)
Example part	Pineios river is navigable for 30 miles

Table 2: Definition's main, secondary and example parts.

The *main* part consists of the *determinant* section, which follows the general form [{article}+{concept name}+{is}], and the *attributes* section. The *attributes* section is the descriptive sentence of the *main* part that contains the genus, the attributive adjectives and the prepositional phrases. The *attributes* section of the *main* part has the general form: [{attributive adjective}*+{genus}+{prepositional phrase}*], where the asterisk declares one-or-many.

For the given definition of the concept '*River'*, the *determinant* section is the phrase '*A River is'* and the *attributes* section is the phrase *'a large natural stream of water'*.

Furthermore, the *secondary* part of a definition contains one or more sentences. Every sentence provides a particular kind of information (purpose, location, etc.). Each sentence in the *secondary* part contains a reserved phrase (for example: used for, located at, etc.) that indicates the semantic relation of the provided information (Kavouras et al. 2003). In the above example, the *secondary* part contains only one sentence (*'used for transportation and irrigation'*) in which the deserved phrase *'used for'* declares that the sentence describes the purpose of the described concept.

PROPOSED METHODOLOGY

We introduce a methodology for transforming the definition of a geographic concept into the corresponding conceptual graph without losing any of the information contained in the definition. The methodology consists of two steps: *tagging* and *parsing*. In the first step, we follow appropriate rules to tag every word of the concept definition. In the second step, we apply a deterministic algorithm in order to parse the tagged definition and create the corresponding conceptual graph.

Alshawi (1989) was the first who developed the idea of using a hierarchy of phrasal patterns to identify formulas in concept definitions. Later on, other researchers (Klavans et al., 1990, Montemagni & Vanderwende, 1992) proposed the method of parsing the definition first, and then doing a search to locate defining formulas and use some heuristics to find the words involved in the relations. This paper is based on the last approach. We parse a geographic definition sentence before we transform it into a conceptual graph and then perform further steps at the graph level.

We will separately tag and parse the *main* and the *secondary* part of a geographic concept definition. In that way, we will produce two conceptual graphs, one corresponding to the *main* part of the definition and the other to the *secondary* one. By joining them, we will construct the conceptual graph representation of both the *main* and the *secondary* part of the definition, in which we will have to add appropriately the example's conceptual graph representation in order to result in the complete conceptual graph representation of the geographic concept.

STEP ONE: DEFINITION TAGGING

The first step in analysing a geographic concept definition before the parse step is the tagging step. Every definition is made of tokens. The exactly chosen parts of speech (tags) that we associate with the words of the *main* and the *secondary* part of the geographic concept definition are included in Table 3. The difference between 'vb' and 'v' tags is that 'vb' always belongs to the *determinant* section of the *main* part and represents the special verb that introduces the definition of the geographic concept.

article	noun	verb "be"	verb	Adjective	preposition	conjunction
art	n	vb	v	adj	prep	conj

Table 3: Tags used in step one.

As regards the *determinant* section of the *main* part of a concept definition, which always consists of an {article}, the {concept name} and the verb {is} (for example: *'A River is'*), we must notice that it is tagged using the abbreviations 'n' and 'vb'. The [{article}+{concept name}] section is tagged with the 'n' abbreviation and the {is} with the 'vb' abbreviation. Therefore, the tagging step for the *determinant* section of the given example produces the output: '{A(art) river (n)} {is (vb)}'.

Concerning the *attributes* section of the *main* part of the concept definition, which contains the genus, the attributive adjectives and the prepositional phrases and is classified into the general form [{attributive adjective}*+{genus}+{prepositional phrase}*], it is tagged using the abbreviations 'adj' for all attributive adjectives, 'n' for the genus and 'prep', 'n' for the prepositional phrase. Consequently, the tagging process on the *attributes* section of 'River' produces: '*{a(art) large (adj)} {natural (adj)} {stream (n)} {of (prep)} {water (n)}*'.

The *secondary* part of a concept description contains one or more sentences. We must apply the tagging process in each one of them. The abbreviations for the reserved phrase are usually 'v' and 'prep' (used for, located at, etc.) and for the rest words of the *secondary* part are 'n', 'adj' and 'conj'.

Hence, the tagging process on the *main* and *secondary* parts of the given definition of the concept '*River*' results in:

'{A river (n)} {is (vb)} {a large (adj)} {natural (adj)} {stream (n)} {of (prep)} {water (n)} {used (v)} {for (prep)} {transportation (n)} {and (conj)} {irrigation (n)}.

STEP TWO: DEFINITION PARSING

The *parsing* step in the creation of a conceptual graph that represents a geographic concept definition is implemented through an algorithmic procedure consisting of:

1. The *parsing* of the tagged *determinant* and *attributes* sections of the *main* part of the description and creation of the corresponding conceptual graph.
2. The *parsing* of all the clauses that belong to the tagged *secondary* part of the definition and creation of the corresponding conceptual graph for each clause.
3. The combination of the previously created conceptual graphs in a single conceptual graph which represents both the *main* and the *secondary* part of the concept.
4. Integration of the example's conceptual graph representation in order to create the complete conceptual graph for the representation of the entire concept definition

Determinant and attributes parsing

The tagged *determinant* section (*{article (art) concept name (n)}{is (vb)}*] always corresponds to the conceptual graph general form of Figure 2:

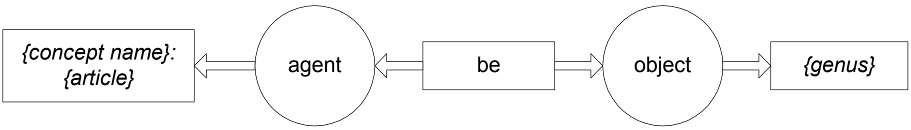

Figure 2: Conceptual graph general form for the *determinant* section.

The concept type *{genus}* refers to the genus contained in the *attributes* section of the tagged *main* part. Figure 3 shows the conceptual graph for the representation of the *determinant* in the phrase *'A River is a …stream…':*

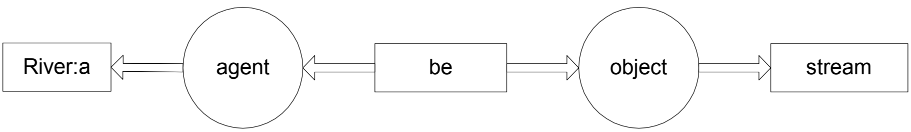

Figure 3: River's determinant conceptual graph.

As concerns the attributive adjectives (tagged with *'adj'*) in the *attributes* section, we define one concept type for each one of them, which is connected to the genus concept type via a concept relation of type *'atr'* (Figure 4).

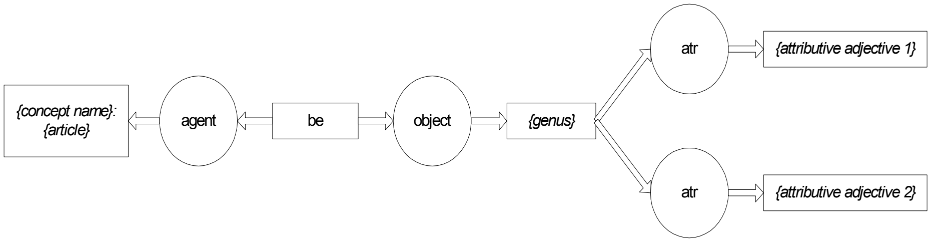

Figure 4: Conceptual graph general form for attributive adjectives.

For every tagged prepositional phrase in the *attributes* section, we introduce a conceptual relation of type *'preposition'* which is also connected to the genus of the definition and to the graph that corresponds to the remaining terms of the prepositional phrase. In general, a tagged prepositional phrase consists of one preposition (tagged with *'prep'*), one or more attributive adjectives (tagged with *'adj'*) and nouns (*'n'*): *{preposition}{attributive adjectives}*{noun}**. The attributive adjectives (if exist) characterize the noun (for example: *'a stream of water'* or *'a stream of cold water'*).

The general form of the conceptual graph corresponding to the prepositional phrase of type *{preposition}{attributive adjective}{noun}* is shown in Figure 5.

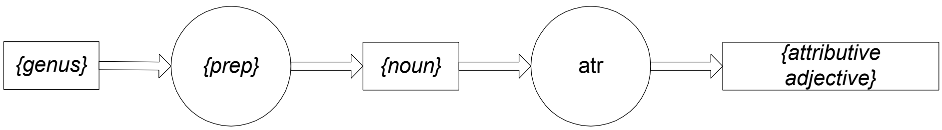

Figure 5: Conceptual graph general form for the prepositional phrase.

Hence, the general form of the conceptual graph representation of a definition's *main* part is:

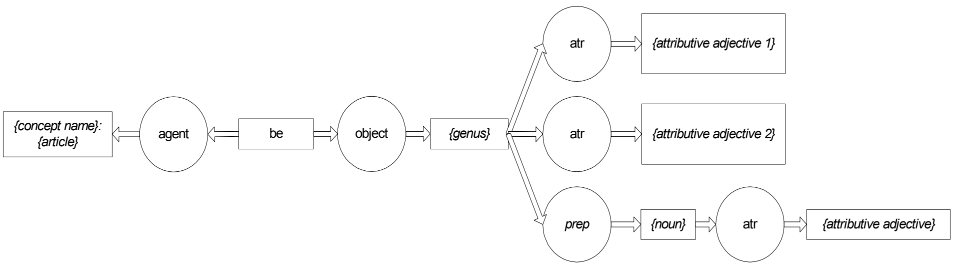

Figure 6: Conceptual graph general form for the *main* part.

For the given definition of the *'River'* concept, the *main* part is represented as follows:

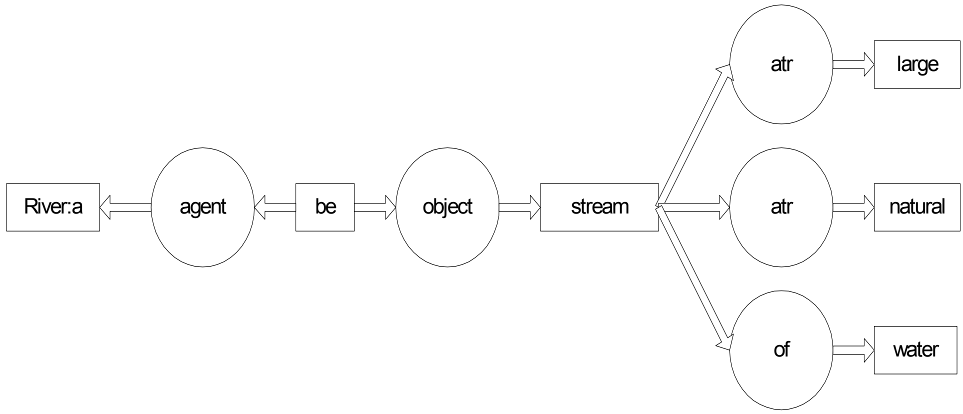

Figure 7: River's main part conceptual graph.

Secondary part parsing

Every sentence in the *secondary* part, as resulted from the tagging process, consists of a reserved phrase that reveals the sentence's semantic relation type and the remaining part providing the information itself or value of the relation (for example *'used for transportation and irrigation'*). In the parsing procedure, the tagged reserved phrase is transformed into a concept of type as the corresponding information type (for example *'purpose'*). This concept is related to the genus concept via a concept relation of type *'agent'* and to the concept types that correspond to other structural elements of the sentence via a concept relation of type *'object'*. Figure 8 shows the general conceptual graph representation form of a definition's *secondary* part that:

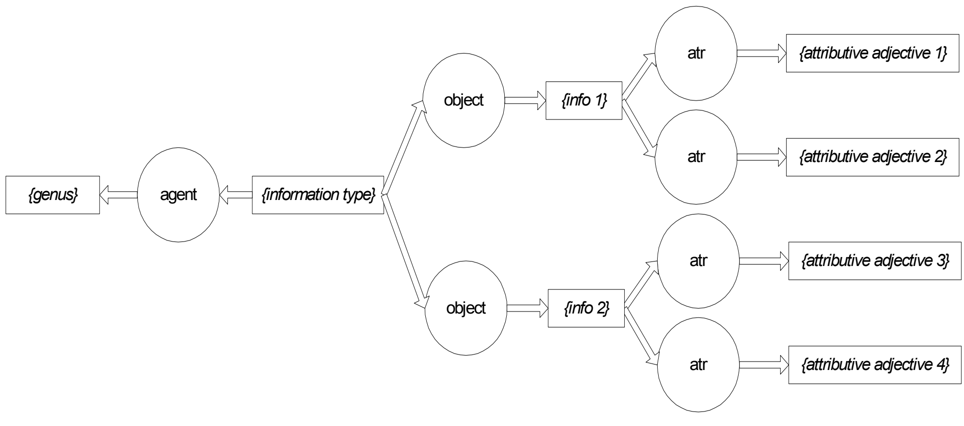

Figure 8: Conceptual graph general form for the *secondary* part.

In the above paradigm, we consider that the general type of every sentence in the *secondary* part is: *{reserved phrase}({attributive adjectives}{information})**, where the *'information'* is represented with the concepts *'info 1', 'info 2'*.

Figure 9 shows the representation of the *secondary* part of the *'River'* definition:

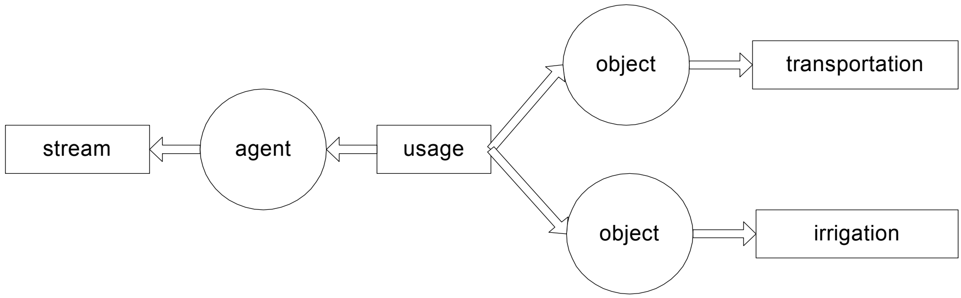

Figure 9: River's secondary part conceptual graph.

The above step draws from the methodology for analyzing definitions and extracting information in the form of semantic relations which was introduced by Jensen & Binot (1987) and further pursued by Ravin (1993) and Vanderwende (1995). This approach consists in the syntactic analysis of definitions and in the application of rules, which examine the existence of certain syntactic and lexical patterns. Patterns take advantage of specific elements of definitions, in order to identify a set of semantic relations and their values based on the syntactic analysis.

Combination

The combination of the conceptual graphs corresponding to the *main* and *secondary* parts of a geographic concept definition produces the integrated representation of the definition. It is the simplest step in the overall procedure since both of the two graphs contain the common concept *'genus'*. Figure 10 shows the general form of the conceptual graph representing a geographic concept definition:

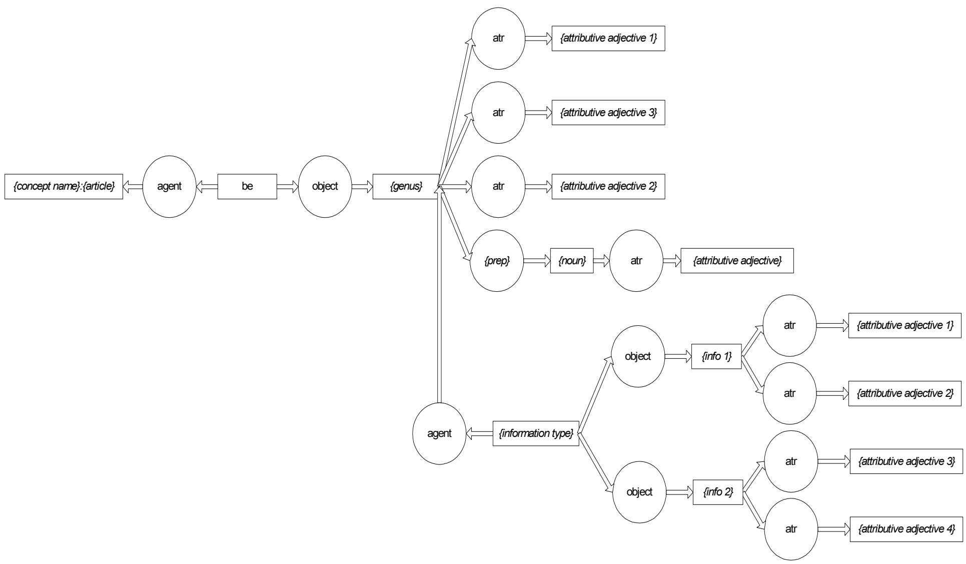

Figure 10: Conceptual graph general form after combination.

Therefore, combining the outputs of the parsing method for the *main* and the *secondary* parts of the definition *'A River is a large natural stream of water used for transportation and irrigation'* we construct the conceptual graph representation of the Figure 11:

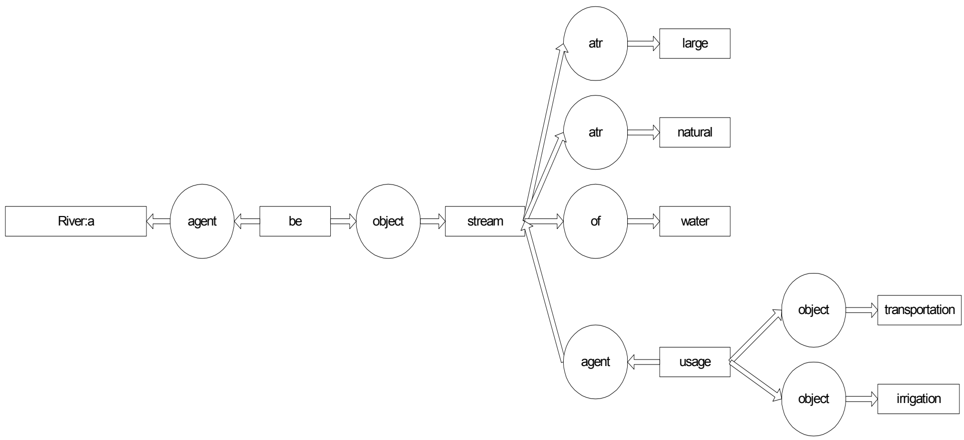

Figure 11: River's conceptual graph after combination.

Example's addition

We consider the concept type *Definition*. It is the concept whose referent can be any definition of a geographic concept. Since we represent the concept's definition using conceptual graphs, the concept *Definition* is a context (Conceptual Graph Standard, 2002) and its designator is a nonblank conceptual graph that represents the definition itself.

We also consider the concept type *Example*. It is the concept whose referent can be any example that contains a geographic concept and is used in order to explain better this concept. The *Example*'s referent will be a nonblank conceptual graph representing the example. The *Example* concept is also a context.

We assume that D is the conceptual graph representing a definition of a geographic concept. In order to integrate the definition's example, we first create the conceptual graph E that represents the example and then, we relate a concept of type *Example* (with graph E as the referent) to the concept *Definition* whose referent is the conceptual graph D. The concept relation is of type *'atr'*.

Hence, the complete conceptual graph that represents the given definition of the geographic concept *'River'* is shown in Figure 12.

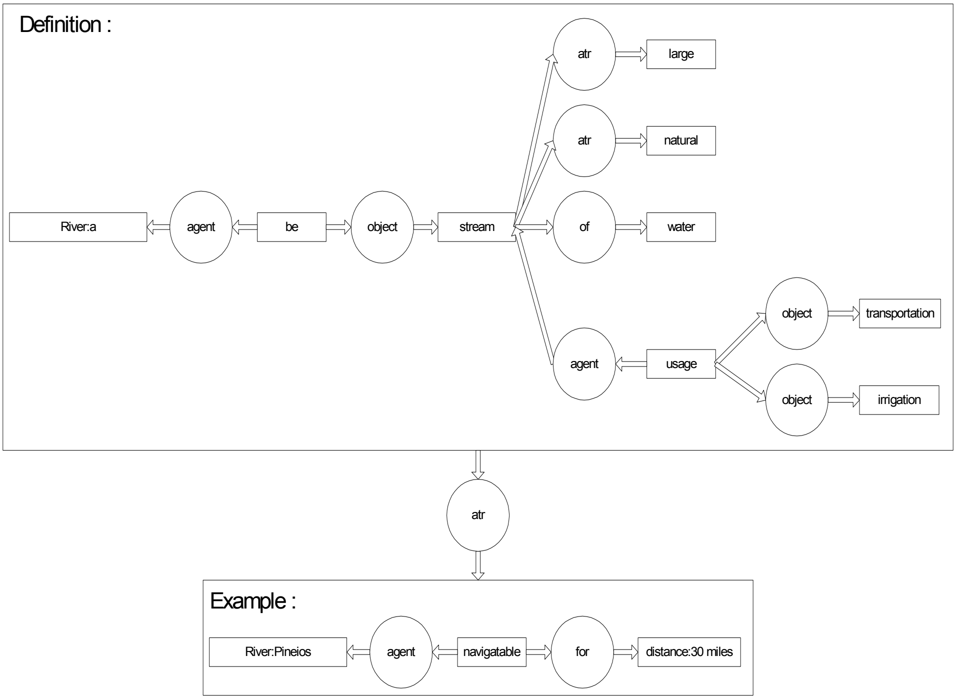

Figure 12: Conceptual graph representation of River's definition.

CONCLUSIONS AND FURTHER WORK

The present research focuses on the representation of geographic concept definitions using conceptual graphs. Developing a straightforward and easy-to-implement methodology for transforming a structured geographic concept definition into the corresponding conceptual graph representation breaks many limitations and obstacles in the extraction of semantic information from definitions of geographic concepts and facilitates the implementation of a really interoperable geographic environment.

Furthermore, this work lays the foundation for the exploitation of conceptual graphs' potentials in order to identify and formalize similarities and heterogeneities between geographic concepts, design knowledge bases for geographic ontologies and fully implement methods for the semantic integration of geographic resources.

BIBLIOGRAPHY

Alshawi H., Analysing the Dictionary Definitions. In Boguraev B. and Briscoe T. (eds). Computational Lexicography for Natural Language Processing. Longman Group UK Limited: 153-170, 1989.

Conceptual Graph Standard, 2002 NCITS.T2 Committee on Information Interchange and Interpretation. http://users.bestweb.net/~sowa/cg/cgstand.htm.

Jensen K. & Binot J., Disambiguating prepositional phrase attachments by using on-line dictionary definitions. Computational Linguistics 13 (3-4): 251-260, 1987.

Jensen K., Heidorn G.E. & Richardson S.D. (Eds), 1993 Natural Language Processing: The PLNLP Approach. Kluwer Academic Publishers, Boston/Dordrecht/London. ISBN 0-7923-9279-5, pp 324.

Kavouras M., Kokla M. & Tomai E., Determination, Visualization and Interpretation of Semantic Similarity among Geographic Ontologies. In M. Gould, R. Laurini, S. Coulondre (eds). Proceedings of the 6th AGILE Conference on Geographic Information Science. Presses Polytechniques et Universitaires Romandes: 51-56, 2003.

Klavans J., Chodorow M.S. & Wacholder N, From Dictionary to Knowledge Base via Taxonomy. In Proceedings of the 6th Annual Conference of the UW Centre for the New OED: Electronic Test Research: 110-132, 1990.

Kokla M. & Kavouras M., Extracting Latent Semantic Relations from Definitions to Disambiguate Geographic Ontologies. In G. Zavala (ed.). GIScience 2002 Abstracts. University of California Regents: 87-90, 2002.

Kuhn W., Modeling the Semantics of Geographic Categories through Conceptual Integration. In M. Egenhofer, D. Mark (eds). Proceedings of the Second International Conference GIScience. Springer: 108-118, 2002.

Montemagni S. & Vanderwende L., Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries. In Proceedings of the 14th COLING: 546-552, 1992.

Montes-y-Gomez M., Gelbukh A., Lopez-Lopez A. & Baeza-Yates R. Flexible Comparison of Conceptual Graphs, 2001. In Heinrich C. Mayr, Jirí Lazanský, Gerald Quirchmayr, Pavel Vogel (eds). Proceedings of the Database and Expert Systems Applications, 12th International Conference, DEXA. Springer: 102-111, 2001.

Ravin Y, 1993 Disambiguating and Interpreting Verb Definitions. In Jensen, Heidorn & Richardson Natural Language Processing: The PLNLP Approach. Kluwer Academic Publishers, Boston/Dordrecht/London. ISBN 0-7923-9279-5, pp 324.

Sowa, J., 1984 Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley. ISBN 05214449004, pp 406.

Sowa, J., 2000 Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks Cole Publishing Co. ISBN 0-534-9496-7, pp 594.

Vanderwende L., 1995 The Analysis of Noun Sentences using Semantic Information Extracted from on-line Dictionaries. Ph.D. thesis, Faculty of the Graduate School of Arts and Sciences, Georgetown University, Washington, D.C: pp 296.