Graph-Based Formalisms for Knowledge Representation

Mile PAVLIĆ
Department of Informatics, University of Rijeka,
Rijeka, Croatia

Ana MEŠTROVIĆ
Department of Informatics, University of Rijeka,
Rijeka, Croatia

and

Alen JAKUPOVIĆ
Bussines Department, Polytechnic of Rijeka,
Rijeka, Croatia

ABSTRACT

Graph-based formalisms provide an intuitive and easily understandable vehicle for knowledge representation. In this paper several existing graph-based formalisms are described. Furthermore, a new graph-based formalism for knowledge representation is defined. Basic concepts for graphical representation (nodes and links) as well as their variation are described. Context node, context link and process node are defined in order to represent different kinds and levels of knowledge. A knowledge network diagram defined from sentences expressed in natural language is presented. The main goal of this formalism is to present knowledge stored in textual fields in the databases.

Keywords: knowledge representation, knowledge based systems, conceptual graphs, semantic networks, natural language representation

1. INTRODUCTION

Knowledge representation and reasoning is an important field in the domain of artificial intelligence. It is concerned with how knowledge can be represented symbolically and manipulated in an automated way by reasoning programs [1] and it involves machine-interpretable representation of the world [2]. In [3] the goal of knowledge representation and reasoning is described as to symbolically encode human knowledge and reasoning in such a way that this encoded knowledge can be processed by a computer via encoded reasoning to obtain intelligent behavior. It can be the knowledge of a single person, of an expert in some domain, shared knowledge of ordinary people (common sense knowledge), social knowledge accumulated by generations, e.g., in a scientific domain, etc.

Knowledge representation is important for the different scientific disciplines [4]: the theoretical foundations of artificial intelligence, linguistics (in connection with the formal description of the semantics of natural language expressions and for the formalization of lexical knowledge), cognitive psychology (to model conceptual structures and the processes of reasoning), natural language processing (e.g. natural language understanding, semantic search, question answering systems, machine translation systems) and knowledge based systems (KBS).

This paper deals with knowledge representation formalisms based on graphic notation. The main concepts used in graph-based knowledge representation are nodes (represented as vertices) and links (represented as arcs). In graph-based approach to knowledge representation graphs are considered for knowledge modeling and for computation. A benefit of this approach is that labeled graphs, schemas and drawings provide an intuitive vehicle for knowledge representation. There is a wide range of applications of graph-based methods such as such as the ER model, UML, Topic Maps in the domain of database and information systems development. The main reason is that graphs are easily understandable for users, knowledge engineers or specialists in an application domain [3].

In this paper we propose a graph-based formalism that is based on semantic network. The main motivation of this approach was to define a formalism that may capture knowledge stored in the text fields in the databases. That way this formalism can be viewed as an extension of the ER model. Introducing an additional model for knowledge representation of textual fields in the database will enable more precise reasoning with the data from the database. In our approach, we introduce different kinds of nodes for term representation. Link between nodes enable grouping terms into more complex expressions. Specifically, a process node is defined as an aggregation point for the representation of knowledge described in sentences.

The paper is organized as follows. Section 2 summarizes the related work and existing methods. Section 3 introduces the main concepts for graph-based knowledge representation formalism that we propose. Section 4 describes how to represent complex language structures using described graph-based concepts. Finally, section 5 presents the conclusions and future work.

2. RELATED WORK

Different applications need different formalisms and languages that may range from simple to complex and very expressive knowledge representation formalisms. Over the time, many different approaches, formalisms, methods and languages for knowledge representation have been developed. Traditionally, there are two main approaches to knowledge representation that are defined in the domain of artificial intelligence: declarative and procedural. The declarative approach includes logic schemas, network schemas and frames; while the procedural approach is related to production systems, also known as rule systems. Apart from these traditional approaches, there is connectionist approach [5,6] in the field of knowledge representation that introduces neural networks.

Network schemas include different formalisms based on graphical notations. The most influential are semantic networks (SN) introduced by Quillian in 1968 [7] and conceptual graphs (CG) introduced by Sowa in 1976 [8] and later developed and elaborated in [9, 10]. Both formalisms are derived from existential graphs developed by Peirce in 1896 to express logical sentences as graphical diagrams [6]. Later on, similar notations have been defined, all differing slightly in syntax and semantics. Furthermore, there is a number of graph-based methods that have their origins in these two main methods: Cognitive Semantic Networks, Structured Inheritance Networks, Multilayered Extended Semantic Networks (MultiNets), Basic Conceptual Graphs (BG), Simple Conceptual Graphs (SG), Full Conceptual Graphs (FCG), Hierarchical Semantic Form (HSF), Resource Description Framework (RDF), etc. A short overview of the main existing graph-based formalisms is given below.

Semantic networks

A semantic network was developed as a model of human (associative) memory [2] in order to provide a structural representation of statements about a domain of interest.

It is represented using nodes, which represent concepts, and labeled arcs, which represents semantic relations. Therefore, a semantic network represents semantic relations between concepts. There are two special relations distinguishing semantic network applications: *is-a* and *a-kind-of*. The *is-a* relation is a particular relation which links individuals to their classes. A *a-kind-of* relation denotes a subsumtion. A subsumption link connects two general concepts and denotes specialisation or generalisation. Furthermore, other specific relations may be used, such as *has-a-property*, or any other relation that specifies relation between two concepts.

There are different kinds of semantic networks that vary from informal to highly formal. The most common kinds of semantic networks are: definitional network, assertional network, implicational network, executable network, learning network and hybrid network.

The main drawbacks of semantic networks include lack of formal semantics and limited reasoning tools [3].

Conceptual graphs

Conceptual graphs were initially defined to represent conceptual schemas used in database systems but after that they had a wide range of applications in artificial intelligence, computer science, and cognitive science.

CG are formally defined in an abstract syntax that is independent of any notation, but the formalism can be represented in three different notations: graphical display form (DF), the formally defined conceptual graph interchange form (CGIF), and the compact, but readable linear form (LF).

A basic conceptual graph is composed of two kinds of nodes, i.e., *concept* nodes representing entities and *relation* nodes representing relationships between these entities.

In [3], some specific graphical methods based on conceptual graphs are defined and described, such as the basic conceptual graphs and the simple conceptual graphs methods.

Multilayered extended semantic networks

Multilayered Extended Semantic Networks is defined as a formalism for the semantic representation of natural language expressions which can be used as a universal knowledge representation paradigm in human sciences [4].

In semantic network formalisms every node represents a concept and vice versa. In MultiNet formalism representatives of concepts that can be designated by a single word are called lexicalized concepts (e.g. *Peter*, *Discussion*) and all other concepts are called nonlexicalized concepts (e.g. *in front of my house*).

A concept can be generally characterized by three components: a word or a word group designating the concept and representing it externally, a collection of relations to other concepts or a complex pattern. Furthermore, each node belongs to a certain class from a predefined classification of nodes, which is called conceptual ontology (consists of 29 classes). Each node possesses 7 predefined attributes, whose values place it in a semantic space (motivation for this approach is the analogy with the point in Euclidean space). Links are established between nodes by using one of 89 predefined types of links.

Hierarchical semantic form

In [6], a knowledge representation technique named Hierarchical Semantic Form is introduced as a method for representing patterns in natural language sequences. The HSF method consists of two basic concepts: group and link. They are graphically represented with circles (empty for groups, full for links) interconnected with arrows (to show the direction of reading). The concept of groups is used in denoting a particular sign, group of signs, words, semantic categories and complex samples. Basically, it is used in order to show the sequence of denoted terms on different levels of abstraction (the group shows the link to the first element of the sequence). The same group can appear at different levels of abstraction. The concept of links is used in the creation of a sequence of signs, group of signs, words, semantic categories etc. on different levels of abstraction. The primary task of a link is to show a sample (group) in different contexts. For example, for a sample "Student", the link has to show the correct sequence of a particular sign in the sample. Similar would apply to words, sentences etc. The HSF method provides semantics for a particular sequence of terms by linking them to a concept of groups which contains a semantic description (e.g. "is a part of" or "part of the day" etc.).

Resource description framework

The RDF method has several display types, one of which is an RDF graph [11]. It is another graph-based method for knowledge representation which focuses on knowledge from web resources. It shows the relationship between web resources by using a named property and its value. Various data, documents, pictures etc. can appear as web resources. The named property refers to the observed resource. The name of the property is defined by a Uniform Resource Identifier (URI) and therefore the property can be observed as a new resource. The value of the property can be another resource or some data (e.g. string, number, date etc.).

The RDF graph method consists of three basic concepts: a node which represents the resource, a node which represents the data and an arrow which connects the nodes. An ellipse is a graphic symbol for a resource, whereas a rectangle is a graphic symbol for data. An arrow represents a property and the name of the property is written on it. The arrow points to the value of the property. Since resources have properties, at the start of an arrow there is always an ellipse (resource node), whereas at the end of the arrow there is an ellipse or a rectangle (therefore, resource or data node). The RDF method can be used to show statements with the subject-object-predicate structure.

The formalisms described above attempt to give a graph-based knowledge representation using nodes and arcs as basic concepts, but in a different way. Figure 1 shows an example of a simple sentence in English language represented with different graph-based formalisms. In the next section we propose a new approach for knowledge representation.

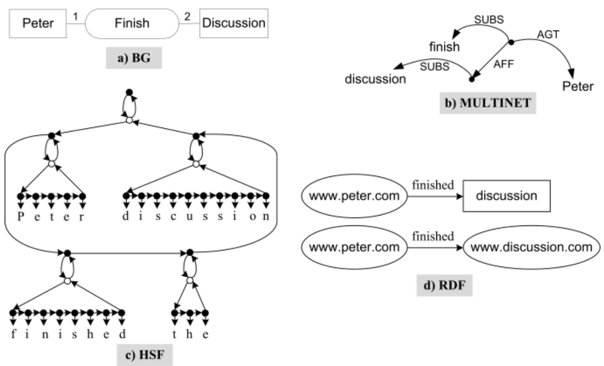

Figure 1. Graph-based methods formknowledge representation

3. BASIC CONCEPTS

In this section we describe basic concepts of the graph-based method that we propose.

Node

A Node is a drop of knowledge (term, entity) different from any other knowledge in the model. All concepts which have their own meaning are in nodes. Thus nodes can represent particular named people, things, events, actions, ideas, but also concepts on a higher level of abstraction, such as a person, a table, a sporting event, learning, a feeling, activities etc. A node is the smallest unit of knowledge which cannot be further divided. Larger units of knowledge are represented by a group of connected nodes. Also, nodes are not groups (relations, tables, classes) of similar entities. A node is not a classification of entities, but rather one individual entity. Every new term is a new node. A node may only contain one term. The meaning (name) of a node is entered in the node symbol (rectangle), as shown on Figure 2. Even the smallest semantic difference between two terms leads to the creation of a new node. A new node is also created when there is a difference in "essence" between two concepts (e.g. two different persons named Marko). The name of a node is its attribute and gives the semantic identification of the concept. The name does not have to be one word only; it can be a group of words, a compound or a derivative with a specific meaning, e.g. "I don't know", "army branch", "come to mind", company "good vibrations". A node can have several attributes. While implementing the method, it is necessary for a node to have a name, but also an identification attribute which identifies it unambiguously. Names of nodes may not be polysemantic, that is, all homonyms have to be broken down and their precise word versions have to be used in the knowledge record. In human language it is presumed that, as a result of their natural intelligence, readers will understand the knowledge which one tries to transfer to them. Our goal is to define such a formalism which relies on the reader's prior knowledge, but leaves no possibility of alternative interpretations and ambiguities. We have shown the essential requirement of the this formalism: there is a drop of knowledge (essence, concept) which has its own identity (name or identifier) and semantics (essence of the concept, the corresponding fact it replaces) different from anything else, both in reality and in the mind, and such drops are represented by nodes. There are following types of nodes: node (ordinary node, static node, entity, concept, term), context node (abstract, group, sort, class, framework, type), data node (place where data are kept) and different kinds of process nodes (binding, relative, dynamic, functional, action nodes, conditioned nodes).

Link

The second basic concept is link, which has the role of connecting a maximum of two nodes in the network. Links do not have link names, but can have role names. Role name (as described below) is the information belonging to the node, and this name questions the role of the connection between that node and another node. A link is represented by a line, with or without an arrow. A link cannot connect three or more nodes. Only binary links are allowed. If there is a need to connect three or more nodes then "process nodes", which connect several nodes by binary links, can be introduced. The cardinality of a link is always (1,1): (1,1) [12]. This means that the first node in the connection must always exist and that it is connected with a maximum of one other node and vice versa. An example of two nodes and their links, which correspond to the sentence: "City of Zagreb", but also to the sentence "Zagreb City", is shown in Figure 2. If we take another city, e.g. Sarajevo, it will be connected to City by a new link.

Context link and context node

A Context link is a special link between a context node (node on the higher level of abstraction, general, superior, class, generic, superterm) and a specialised node (described, specific, node on the lower level of abstraction, of phenomenon, of the pertinent).

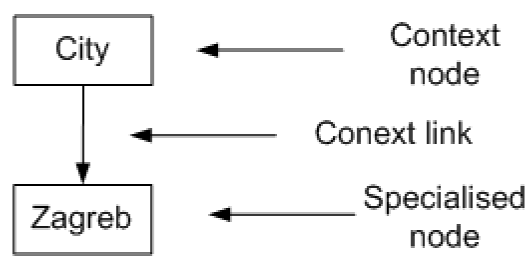

Figure 2. Relationship between a context node and a specialised node

The assumption on which this knowledge modeling concept lies upon is that knowledge can be presented and organised in nodes among which there is a certain relationship: abstract term – specific term. The complete knowledge network consists of several levels of knowledge, as it is presented in Figure 3. The first level is the level of the first (phenomenal) nodes, that is, the schemes of knowledge about relationships between particular phenomena. The second and higher levels are the levels of classes and their relationships. The higher level node is a context node with respect to the node described on the lower level. The knowledge network imposes no limitations regarding the connection of any node with any other node on any level.
It is possible to interconnect nodes from all levels; then the higher level node is called the context node with respect to the node it is connected to. This link is called the context link. This

is not a generalisation link in which supertype attributes belong to the subtype and both have several occurrences. A context link for a particular node answers to the following questions: what is the node, which sort is the node, of which type it is, which class is it, to which group it belongs, etc. If two nodes from different levels are connected, then the context link is represented by a line with an arrow. The arrow points to the lower level node. Figure 3 shows certain context links. One node can be the context node for an unlimited number of specialised nodes. One node can have an unlimited number of superior context nodes. A context link can be established between different sorts of nodes (ordinary, process nodes).

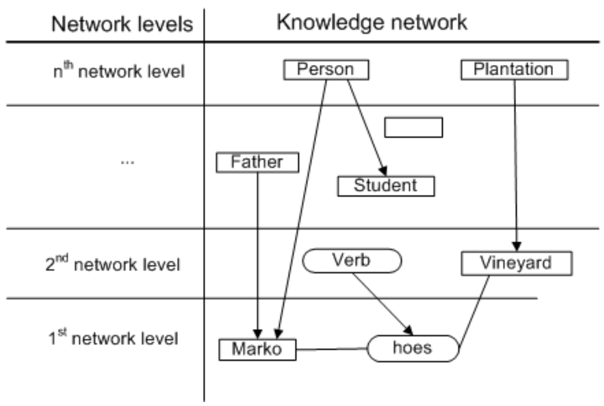

Figure 3. Knowledge network nodes at different levels

Process node

A process node is a node whose links connect the nodes and together with them create a more complex presentation of knowledge in form of aggregated knowledge. The process node is intended for representing knowledge which cannot be represented by ordinary nodes because it stands for: relationships between nodes, activities, links between several nodes, actions, occurrences, feelings etc. The process node is graphically represented by an oval (Figure 3). Names of ordinary and context nodes are usually nouns. Names of process nodes can belong to different parts of speech or word groups but are usually verbs or gerunds. Process nodes are the glue which links and connects words in a sentence into superterms. A process node can represent an action happening between the nodes. If an action is composed of several parts, then it is represented by several connected process nodes.

4. NATURAL LANGUAGE SENTENCES REPRESENTATION

An example of a simple statement representation

A simple statement can be represented using nodes and links. Simple sentences given in the *subject-predicate-object* form can be transformed into graphic notation. Subjects and objects are represented with the specialized node and predicates are represented with the process node.

Let us analyse the simple statement clause "Marko hoes the vineyard". This sentence has one verb -"hoes", and corresponds to one process node. Other two words are nouns and we introduce two nodes for them as it is shown in Figure 4 a.

This sentence can be extended with additional knowledge. One process node may have as many individual drops of knowledge as it has links as it is shown in Figure 4 b.

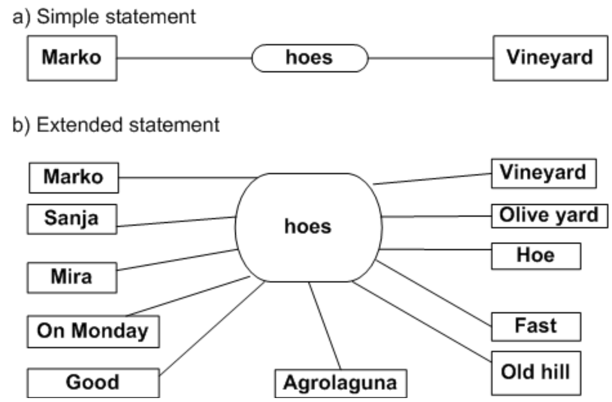

Figure 4. The process node "hoes" (Croatian: "kopa") and its links

An example of a complex statement representation

We will now show model for a sentence chosen from daily press, expand the obtained diagram with additional knowledge about those events and ask questions the answers to which can later be gathered and built into the model.

A complex sentence is: "La Peregrina, a pearl from the 16th century that Richard Burton gave to the actress as a gift, was sold for $ 11.842.500". A model of this sentence is shown in Figure 5.

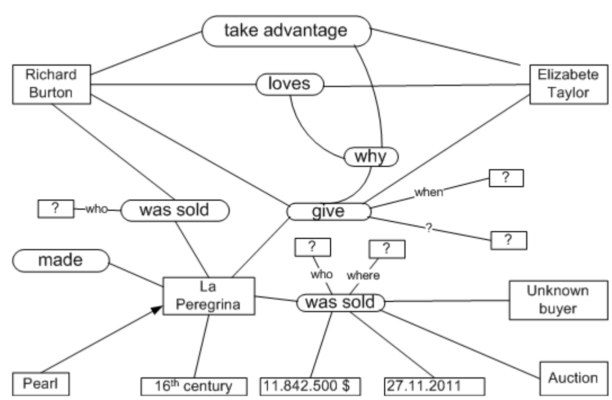

Figure 5. Knowledge network applicable to more than one sentence

The chosen sentence does not tell us which actress it refers to, but from previous sentences it can be concluded that Elizabeth Taylor's pearls were being sold. A part of knowledge can be extracted from each of the sentences in the text and added to the diagram. Several process nodes are shown. Thus we see the first sale to Richard Burton and the second sale to an unknown buyer. We see everything that was happening with La Peregrina pearl. Unlike natural language, which consists of finite logical sentences and in which the subject always has to be named, the model of a knowledge network consists of one expressed "sentence", unlimited by size and contents, in which each term appears once and only once, but with at least as many connections as contained in the corresponding text. Process nodes around an ordinary node show everything that was happening (processes, events, occurrences, emotions, activities, actions, movements) with the ordinary node in any of the roles (subject, predicate, object). Process nodes between ordinary nodes show everything that was happening between those entities.

5. CONCLUSION AND FUTURE WORK

In this paper, we present a graph-based formalism for knowledge representation. This is a new method and as such it will continue developing and upgrading. Further lines of research can go in the direction of applying the method to various sentences, groups of sentences connected by particular subjects; various types of texts, parts of an information system such as documents, databases, reports, screen forms, Web contents or business applications. It provides an alternative way of storing knowledge, different from the way the human mind stores spoken or written words, and different from other existing formalisms.

What this method introduces is a special kind of process nodes which represent a link between nodes, but at the same time a new term related to other terms. The method emphasises contextuality of all knowledge.

The further research includes realising the idea of incorporating questions in the knowledge network enables us to learn, to ask ourselves what else can be added in order to expand knowledge and retrieve it from the knowledge base. A separate issue is how to apply this method in the field of artificial intelligence for reasoning. Moreover, further research and improvements can upgrade this method to include a number of grammatical and orthographic standards (tenses, cases, pronouns, multilingualism). The knowledge network enables easier development of expert systems, particularly modules for communication between experts and knowledge bases, both for asking questions and for expanding the knowledge base.

6. REFERENCES

[1] R. Brachman, H. Levesque, **Knowledge Representation and Reasoning**, The Morgan Kaufmann Series in Artificial Intelligence, Morgan Kaufmann, 2004.

[2] S. Grimm, P. Hitzler, A. Abecker, “Knowledge Representation and Ontologies, Logic”, **Ontologies and Semantic Web Languages**, 2007.

[3] M. Chein, M.-L. Mugnier, **Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs**, Springer Publishing Company, Incorporated, 1 edition, 2008.

[4] H. Helbig, **Knowledge Representation and the Semantics of Natural Language**, Springer, Berlin, 2005.

[5] M. Stanojevic, S. Vraneš, “Knowledge representation with soul”, **Expert Systems with Applications**, Vol 33, 2007, pp. 122–134.

[6] S. Russell, P. Norvig, **Artificial Intelligence: A Modern Approach**, Pearson Education, 2003.

[7] R. Quillian, “Semantic memory”, **Semantic Information Processing**, MIT Press, 1968, pp. 216–270,

[8] J. F. Sowa, “Conceptual graphs for a data base interface”, **IBM Journal of Research and Development**, Vol 20, 1979, pp. 336–357.

[9] J. F. Sowa, “Semantics of conceptual graphs”, in **Proceedings of the 17th annual meeting on Association for Computational Linguistics**, Stroudsburg, PA, USA, 1979, pp. 39–44.

[10] J. F. Sowa, **Knowledge representation: logical, philosophical and computational foundations**, Brooks/Cole Publishing Co., Pacific Grove, CA, USA, 2000.

[11] G. Klyne, J. Carroll, **Resource description framework (RDF): Concepts and abstract syntax**, World WideWeb Consortium, 2004.

[12] M. Pavlic, **Oblikovanje baza podataka**, Department of Informatics University of Rijeka, 2011.