MULTI-LAYER KNOWLEDGE REPRESENTATION

Krassimira Ivanova

Abstract: *An approach for knowledge representation based on post-relation type of information bases is outlined in the paper. Explanation starts with remembering the idea of Natural Language Addressing. After that, the idea of Multi-layer Knowledge Representation by Means of Natural Language Addressing is presented.*

Keywords*: Multi-layer Knowledge Representation; Natural Language Addressing*

ACM Keywords*: I.2.4 Knowledge Representation Formalisms and Methods*

Introduction

There are a lot of approaches for knowledge representation [Sowa, 2000]. In this work we show an example of knowledge representation by means of Natural Language Addressing (NLA) [Ivanova et al, 2012a; 2012b; Ivanova et al, 2013a; 2013b; 2013c; 2013d; 2013e; Ivanova, 2013; Ivanova, 2014a]. NLA is based on the Multi-domain Information Model [Markov, 1984; Markov, 2004; Markov, 2004a].

To remember the idea of knowledge representation let start from a letter to Philip Jourdain written in 1914 by Gottlob Frege [Frege, 1980]:

> Let us suppose an explorer travelling in an unexplored country sees a high snow-capped mountain on the northern horizon.
>
> By making inquiries among the natives he learns that its name is 'Aphla'. By sighting it from different points he determines its position as exactly as possible, enters it in a map, and writes in his diary: 'Aphla is at least 5000 meters high'.
>
> Another explorer sees a snow-capped mountain on the southern horizon and learns that it is called Ateb. He enters it in his map under this name.
>
> Later comparison shows that both explorers saw the same mountain. Now the content of the proposition 'Ateb is Aphla' is far from being a mere consequence of the principle of identity, but contains a valuable piece of geographical knowledge. What is stated in the proposition 'Ateb is Aphla' is certainly not the same thing as the content of the proposition 'Ateb is Ateb'.
>
> Now if what corresponded to the name 'Aphla' as part of the thought was the reference of the name and hence the mountain itself, then this would be the same in both thoughts. The thought expressed in the proposition 'Ateb is Aphla' would have to coincide with the one in 'Ateb is Ateb', which is far from being the case. What corresponds to the name 'Ateb' as part of the thought must therefore be different from what corresponds to the name 'Aphla' as part of the thought. This cannot therefore be the reference which is the

> same for both names, but must be something which is different in the two cases, and I say accordingly that the sense of the name 'Ateb' is different from the sense of the name 'Aphla'.
>
> Accordingly, the sense of the proposition 'Ateb is at least 5000 meters high' is also different from the sense of the proposition 'Aphla is at least 5000 meters high'. Someone who takes the latter to be true need not therefore take the former to be true. An object can be determined in different ways, and every one of these ways of determining it can give rise to a special name, and these different names then have different senses; for it is not self-evident that it is the same object which is being determined in different ways.
>
> We find this in astronomy in the case of planetoids and comets. Now if the sense of a name was something subjective, then the sense of the proposition in which the name occurs, and hence the thought, would also be something subjective, and the thought one man connects with this proposition would be different from the thought another man connects with it; a common store of thoughts, a common science would be impossible.
>
> It would be impossible for something one man said to contradict what another man said, because the two would not express the same thought at all, but each his owns.
>
> For these reasons I believe that the sense of a name is not something subjective (crossed out: in one's mental life), that it does not therefore belong to psychology, and that it is indispensable [Frege, 1980].

The important knowledge in this example is [Ivanova et al, 2013c]:

- The names Ateb and Aphla refer different parts of the same natural object (mountain, let call it *Pirrin*);
- The position of the referred object (mountain) is fixed by any artificial system (geographical co-ordinates, address) which is another name of the same object;
- The names and the address correspond one to another and both to the real object but without the explorer's map, respectively – the explorer's diary, it is impossible to restore the correspondence;
- At the end, the names Ateb and Aphla are connected hierarchically to the name Pirrin and the relations are:
  - Aphla is_a_South_Side_of Pirrin;
  - Ateb is_a_North_Side_of Pirrin.

The knowledge above is unstructured. For automated processing it has to be structured following some information (data) model. Knowledge representation is closely connected to data models, i.e. the information structures used for organizing the information in the internal or external computer memory. In other words, knowledge representation is depended on the storing patterns and program tools for accessing data. Below we will outline an approach for knowledge representation based on post-relation type of information bases starting with remembering the idea of Natural Language Addressing. After that, the idea of Multi-layer Knowledge Representation by Means of Natural Language Addressing will be presented.

Natural Language Addressing

In this research we follow the proposition of Kr. Markov to use the computer encoding of name's (concept's) letters as logical address of connected to it information stored in a multi-dimensional numbered information spaces [Markov, 1984; Markov, 2004; Markov, 2004a]. This way no indexes are needed and high speed direct access to the text elements is available. It is similar to the natural order addressing in a dictionary where no explicit index is used but the concept by itself locates the definition. For this case we use the term: "Natural Language Addressing" (NL-addressing) [Ivanova et al, 2013a].

The idea of NL-addressing is to use encoding of the name both as relative address and as route in a multi-dimensional information space and this way to speed the access to stored information. For instance, let have the next definition: "**Pirrin:** A mountain in the unexplored country with co-ordinates (x, y)".

In the computer memory, for example, it may be stored in a file at relative address "50067328" and the index couple is: ("Pirrin", "50067328"). At the memory address "50067328" the main text, "A mountain ... (x,y)" will be stored. To read/write the main text, firstly we need to find name "Pirrin" in the index and after that to access memory address "50067328" to read/write the definition.

If we assume that name "Pirrin" in the computer memory is encoded by six numbers (letter codes), for instance by using ASCII encoding system Pirrin is encoded as (80, 105, 114, 114, 105, 110), than we may use these codes for direct address to memory, i.e. ("Pirrin", "80, 105, 114, 114, 105, 110").

Above we have written two times the same name as letters and codes. Because of this we may omit this couple and index, and read/write directly to the address "80, 105, 114, 114, 105, 110".

For human this address will be shown as "Pirrin", but for the computer it will be "80, 105, 114, 114, 105, 110".

Till now, NL-addressing has been presented in several publications [Ivanova et al, 2012a; 2012b; Ivanova et al, 2013a; 2013b; 2013c; 2013d; 2013e; Ivanova, 2013; Ivanova, 2014a].

Examples of models for knowledge representation

Maybe the simplest model for knowledge representation is one used for dictionaries and vocabularies.

From our example we may create a simple vocabulary (Table 1):

Table 1. A simple vocabulary

name	*definition*
Pirrin	A mountain
Aphla	The South Side of Pirrin mountain
Ateb	The North Side of Pirrin mountain
(x, y)	Co-ordinates of Pirrin mountain

The vocabulary in Table 1 is good for using by humans but it is not appropriate for processing in the computer. The same knowledge may be represented in the most popular data format – the relational one. The information from Table 1 may be represented by a relation (Table 2).

Table 2. Relational representation of the simple ontology

object	*South_Side*	*North_Side*	Co-ordinates
Pirrin	**Aphla**	**Ateb**	**(x, y)**

Knowledge from Table 1 or Table 2 may be used for creating a simple ontology with four concepts which may be represented by an ontology graph (Figure 1):

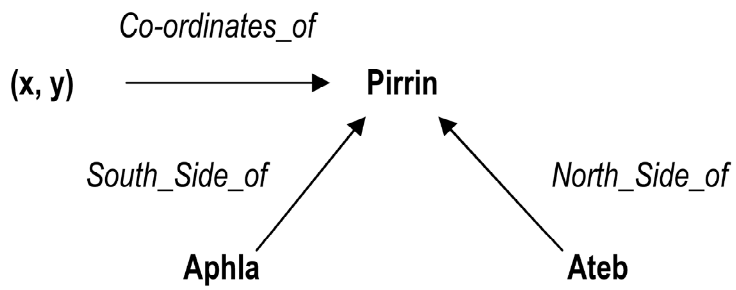

Figure 1. A simple ontology graph

Now we are ready to remember the Resource Description Framework (RDF). It is the W3C recommendation for semantic annotations in the Semantic Web. RDF is a standard syntax for Semantic Web annotations and languages [Klyne & Carroll, 2004]. The underlying structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an ***RDF graph***. This can be illustrated by a node and directed-arc diagram, in which each triple is represented as a *node-arc-node* link (hence the term "graph") (Figure 2).

Figure 2. RDF triple

Each triple represents a statement of a relationship between the things denoted by the nodes that it links. Each triple has three parts: (1) *subject*, (2) *object*, and (3) a *predicate* (also called a property) that denotes a relationship. The direction of the arc is significant: it always points toward the object. The nodes of an RDF graph are its subjects and objects. The assertion of an RDF triple says that some relationship, indicated by the predicate, holds between the things denoted by subject and object of the triple. The assertion of an RDF graph amounts to asserting all the triples in it, so the meaning of an RDF graph is the conjunction (logical AND) of the statements corresponding to all the triples it contains. A formal account of the meaning of RDF graphs is given in [Hayes, 2004].

RDF representation of our simple ontology from Figure 1 is given in Table 3.

Table 3. RDF representation of the simple ontology

subject	relation	object
Pirrin	*has_South_Side*	**Aphla**
Pirrin	*has_North_Side*	**Ateb**
Pirrin	has_Co-ordinates	**(x, y)**

From examples given above we may conclude that vocabularies, taxonomies, thesauruses, relations, ontologies, and RDF-graphs, **all what they have in common** are [Pidcock & Uschold, 2012]:

- They are approaches to help structure, classify, model, and/or represent the concepts and relationships pertaining to some subject matter of interest to some community;
- They are intended to enable a community to come to agreement and to commit to use the same terms in the same way;
- There is a set of terms that some community agrees to use to refer to these concepts and relationships;
- The meaning of the terms is specified in some way and to some degree;
- They are fuzzy, ill-defined notions used in many different ways by different individuals and communities.

The **major differences** that distinguish these approaches [Pidcock & Uschold, 2012]:

- How much meaning is specified for each term?
- What notation or language is used to specify the meaning?
- What is the thing for? Taxonomies, thesauruses, and ontologies have different but overlapping uses.

At the end, some additional information may be connected to the names. For instance, it may be the type of mountain, minerals found, some photos, textual descriptions, etc. All such information is connected to names and has to be accessed by names as keywords or paths to it, i.e. its computer representation has to be organized using corresponded pointers, indexes of keyword, etc.

In this case the concept “knowledge representation” is used. As we have seen above, the ontologies are useful approach for knowledge representation, which is understandable for humans as well as for the specialized software.

Example of Multi-layer Knowledge Representation by means of Natural Language Addressing

In this point, we will illustrate an approach for storing RDF-graphs by means of the Natural Language Addressing. Taking in account the interrelations between nodes and edges on Figure 1, a special two-dimensional "multi-layer" representation of knowledge from Table 3 becomes possible (Table 4). It is usual for humans but it is not wide used in the computers.

Table 4. Multi-layer representation of the simple ontology

object / layer	Pirrin
South_Side	**Aphla**
North_Side	**Ateb**
Co-ordinates	**(x, y)**

The layers form Table 4 may be stored in different files. If we will use the possibility for NL-addressing, the names of the columns will define locations in files of layers.

To receive all knowledge for given node, we have to take node (column) name as NL-address, for instance "Pirrin", and read all information stored at location determined by its encoding ("80, 105, 114, 114, 105, 110") as NL-addresses in different layers (rows "*South_Side*", "*North_Side", and "*Co-ordinates").

In this multi-layer knowledge representation we have one very important achievement – only cells from Table 4, given in bold, will be stored in computer memory. All other information is "virtual" address information used for access to real information. This causes avoiding of supporting indexes for speeding of information search and as a result reducing of used computer resources - memory and processing time.

Note that in practical implementations the tables similar to Table 4 may have several hundred or thousand columns and rows. In addition they are sparse and many cells are empty. NL-Addressing gives possibility to work effectively with such tables not storing empty cells.

Conclusion

Concluding the examples, let point on advantages and disadvantages of the illustrated above multi-layer knowledge representation by means of the Natural Language Addressing.

The main advantages are:

- Reducing the number of tables, which represent the graph;
- Reducing the number of filled cells.

The main disadvantages are:

- The tables are sparse;
- Avoiding pointers we receive a variety of names, which have different lengths and cause difficulties for the implementations in the data bases where the fixed length is preferable;
- The number of nodes may be very great and this way needs corresponded number of columns in the table (in any cases hundreds or thousands).

The disadvantages may be avoided if we will use the Multi-Domain Information Model (MDIM) [Markov, 2004] and corresponded Multi-Domain Access Method (MDAM) [Markov, 1984]. We upgraded MDAM to NL-addressing approach to apply for storing graphs. The possibility to use coordinates is good for graph models where it is possible to replace search with addressing. Hence, the advantages of MDAM are:

- The possibility to build growing space hierarchies of information elements;
- The great power for building interconnections between information elements stored in the information base;
- The practically unlimited number of dimensions (this is the main advantage of the numbered information spaces for graphs where it is possible "*to address, not to search*");

The NL-addressing and multi-layer organization of the information are good basis for implementing this approach for real solutions.

Bibliography

[Frege, 1980] Gottlob Frege An extract from an undated letter, published in Frege's Philosophical and Mathematical Correspondence (ed.) Gottfried Gabriel, Hans Hermes. Friedrich Kanbartel. Christian Thiel, and Albert Veraart, Abridged for the English (edn.), by Brian MeGuinness, and Trans. Hans Kaal (Oxford: Blackwell. 1980), http://mind.ucsd.edu/syllabi/00-01/phil235/a_readings/frege_jourdain.html (accessed: 15.11.2012).

[Hayes, 2004] Patrick Hayes, Editor, RDF Semantics, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/2004/REC-rdf-mt-20040210/ . Latest version available at http://www.w3.org/TR/rdf-mt/ (accessed: 21.02.2013).

[Ivanova et al, 2012a] Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. “About NL-addressing” (К вопросу о естествено-языконой адрессации) In: V. Velychko et al (ed.), Problems of Computer in Intellectualization. ITHEA® 2012, Kiev, Ukraine - Sofia, Bulgaria, ISBN: 978-954-16-0061 0 (printed), ISBN: 978-954-16-0062-7 (online), pp. 77-83 (in Russian).

[Ivanova et al, 2012b] Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. “Storing RDF Graphs using NL-addressing”, In: G. Setlak, M. Alexandrov, K. Markov (ed.), Artificial Intelligence Methods and Techniques for Business and Engineering Applications. ITHEA® 2012, Rzeszow, Poland; Sofia, Bulgaria, ISBN: 978-954-16-0057-3 (printed), ISBN: 978-954-16-0058-0 (online), pp. 84 – 98.

[Ivanova et al, 2013a] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Introduction to the Natural Language Addressing”, International Journal "Information Technologies & Knowledge" Vol.7, Number 2, 2013, ISSN 1313-0455 (printed), 1313-048X (online), pp. 139–146.

[Ivanova et al, 2013b] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Introduction to Storing Graphs by NL-Addressing”, International Journal “Information Theories and Applications”, Vol. 20, Number 3, 2013, ISSN 1310-0513 (printed), 1313-0463 (online), pp. 263 – 284.

[Ivanova et al, 2013c] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Storing Dictionaries and Thesauruses Using NL-Addressing”, International Journal "Information Models and Analyses" Vol.2, Number 3, 2013, ISSN 1314-6416 (printed), 1314-6432(online), pp. 239 - 251.

[Ivanova et al, 2013d] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “The Natural Language Addressing Approach”, International Scientific Conference “Modern Informatics: Problems, Achievements, and Prospects of Development”, devoted to the 90th anniversary of academician V. M. Glushkov. Kiev, Ukraine, 2013, ISBN 978-966-02-6928-6, pp. 214 - 215.

[Ivanova et al, 2013e] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Storing Ontologies by NL-Addressing”, IVth All–Russian Conference “Knowledge-Ontology-Theory” (KONT-13), Novosibirsk, Russia, 2013, ISSN 0568 661X, pp. 175 - 184.

[Ivanova, 2013] Krassimira Ivanova, “Informational and Information models”, In Proceedings of 3rd International conference “Knowledge Management and Competitive Intelligence” in the frame of 17th International Forum of Young Scientists “Radio Electronics and Youth in the XXI Century”, Kharkov National University of Radio Electronics (KNURE), Kharkov, Ukraine, Vol.9, 2013, pp 6-7.

[Ivanova, 2014a] Krasimira Ivanova, “Storing Data using Natural Language Addressing”, PhD Thesis, Hasselt University, Belgium, 2014

[Klyne & Carroll, 2004] Graham Klyne and Jeremy J. Carroll, Editors, Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/. Latest version available at http://www.w3.org/TR/rdf-concepts/ (accessed: 21.02.2013).

[Markov, 1984] Krassimir Markov, “A Multi-domain Access Method”, Proceedings of the International Conference on Computer Based Scientific Research, PLOVDIV, 1984, pp. 558 - 563.

[Markov, 2004] Krassimir Markov, “Multi-domain information model”, Int. J. Information Theories and Applications, 11/4, 2004, pp. 303 - 308

[Markov, 2004a] Krassimir Markov, “Co-ordinate based physical organization for computer representation of information spaces”, (Координатно базирана физическа организация за компютърно представяне на информационни пространства) Proceedings of the Second International Conference “Information Research, Applications and Education” i.TECH 2004, Varna, Bulgaria, Sofia, FOI-COMMERCE – 2004, стр. 163 - 172 (in Bulgarian).

[Pidcock & Uschold, 2012] Woody Pidcock, Michael Uschold. “What are the differences between a vocabulary, taxonomy, a thesaurus, ontology, and a meta-model?”, InfoGrid - the Web Graph Database http://infogrid.org/trac/wiki/Reference/PidcockArticle, Retrieved November 18, 2012 (accessed: 26.11.2012).

[Sowa, 2000] John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, ©2000. 594 + xiv pages; ISBN 0-534-94965-7

Authors' Information

Ivanova Krassimira *– University of National and World Economy, Sofia, Bulgaria; Institute of Mathematics and Informatics, BAS, Bulgaria; e-mail: krasy78@mail.bg*

Major Fields of Scientific Research: Software Engineering, Business Informatics, Data Mining, Multidimensional multi-layer data structures in self-structured systems