Procedia Computer Science

Volume 80, 2016, Pages 2392–2396

ICCS 2016. The International Conference on Computational Science

A Bilingual Semantic Network of Computing Concepts

Eugene Khenner[1] and Olfa Nasraoui[2]

[1]*Perm State University, Russia*[*]
[2]*University of Louisville, USA*
ehenner@psu.ru, olfa.nasraoui@gmail.com

Abstract
We describe the construction of a bilingual (English-Russian /Russian-English) semantic network covering basic concepts of computing. To construct the semantic network, we used the Computing Curricular series created during 2000-2015 under the aegis of ACM and IEEE and the current standards of IT specialists training in Russia, as well as some other English language and Russian language sources. The resulting network can be used as a basic component in an intelligent information system that allows processing bilingual search queries while considering their semantics and to help support and guide automated translation efforts of academic texts from one language to the other. The network can also be useful to support comparative analysis and integration of the programs and teaching materials for Computing and IT education in Russia and English speaking countries. This network can support cross-lingual information retrieval, knowledge management and machine translation, which play an important role in e-learning personalization and retrieval in the computing domain, thus allowing to benefit from online educational resources that are available in both languages.

Keywords: Computing, Semantic Networks, IT Education

1 Introduction

One of the factors believed to be contributing to widen the talent gap in computational sciences and high performance computing in non-English-speaking countries is the difficulty of context-dependent interpretation of specific terms and their combinations, which frequently appear in the domain of computational science in English-speaking countries, especially in the USA. This problem is particularly urgent in the educational domain, for example when comparing studies of Informatics in

[*] This research was funded by the C-26.004.08 project awarded by the government of Perm Krai, Russia

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016

doi:10.1016/j.procs.2016.05.460

Russian schools with similar studies in the U.S. (Khenner E. & Semakin I., 2014) or IT-specialist training in Russian universities with similar training in the U.S. (Khenner E. & Stallmann M., 2013). Solving this problem is expected to support the development of high performance computing environments as well as university courses and professional training programs.

One significant contribution to the solution of this problem would be the creation of an intelligent information system that can support the processing of bilingual search queries while considering their semantics, and that can also support translating academic texts from one language to the other. A bilingual (in our case English/Russian and Russian/English) knowledge base covering basic concepts of computing should serve as a basis for such a system. The developed network can also be useful for various bilingual semantic web and semantic web mining applications.

Bilingual semantic networks have been successfully used to solve the problem of matching concepts in different languages. Examples include the system of semantic networks for European languages EuroVordNet (Vossen P., 1998), MultiWordNet (Pianta E. et al, 2002).

Creation of the computing bilingual semantic network is a preliminary stage toward the development of a bilingual knowledge base in the area of computing and for the development of methods that support the automated conversion of traditional information systems into intelligent ones using ontological engineering. The subject area of "Computing" has been chosen in this paper as a to illustrate the methodology due to the following reasons:

1) An ever increasing number of Internet users has led to an increased demand for multilingual services. Unlike thesauri and electronic bilingual dictionaries, providing translation and explanation of concepts that is dependent upon a given known context, to be pre-specificed by the author, the proposed approach makes it possible to dynamically obtain and automatically account for the necessary context directly during the process of information search and analysis. To reach these goals, one can use text mining techniques and exploit semantic similarity measures to take into account meaning. Our ultimate goal is building a bilingual (Russian and English) ontology of the basic terms in the area of computing, in the *owl* format standard. This will make it possible not only to embed the ontology into existing and future IS to improve information search quality. but also to improve the translation quality, as there are currently no means available to obtain adequate sense matches between English and Russian terms in the subject area of computing.

2) The computing field is of great importance at the present stage of science and technology development. In fact, computing concepts have been applied in many other subject areas of natural, social, economic, human and engineering sciences and their applications; hence the ontology under development will likely have a broad impact in many peripheral areas.

3) Because the computing field has been developing rapidly, there are many terms that still do have firmly established definitions. Therefore, these terms are often interpreted differently by different researchers and IT users, which can add ambiguity and hinder successful cooperation.

4) "Computing", as a field, can be considered to be bilingual for a qualified Russian-speaking IT specialist, as most of the terms in scientific, technical and academic literature in Russian, are calques of English-language terms, but the meaning of these calques does not always conform to the original connotations. Other foreign IT specialists might face the same problem.

5) The subject area of "Computing" gave rise to a broad educational field, covering all levels and trends in education. For this reason, it is urgent to develop and use intelligent information systems when developing academic standards and programs, as well as when implementing them into those already existing in educational establishments, without altering their initial code.

For computing and IT education, the correspondence of basic terms and use of semantic structures, having a uniform interpretation, is very important. The majority of terms in the area are of English origin. Hence IT education, provided in some other language, must have clear and accurate correlations between the terms in both languages. To solve this task, a bilingual semantic network of the subject area could be beneficial.

6) In the computing field, there is a large body of structured standardized data both in the English and Russian academic literature, that can be used as a good basis for the automatic selection of key concepts and the creation of databases and ontologies. We can mention for instance English language reviews published by the Association for Computing Machinery (ACM) and the IEEE Computer Society: Computing Curricula 2001, Computer Science Volume; Information Systems 2002; Computer Engineering 2004 (CE2004); Software Engineering 2004 (SE2004); Information Technology 2008 (IT2008); Computing Curricula 2005 (CC2005); Computer Science 2008 (CS2008); Information Systems 2010 (IS2010); Graduate Software Engineering 2009 (GSwE2009); Computer Science Curricula 2013. These contain a great variety of standardized units in the "Computing" field. Among sources in Russian, there are IT encyclopedias and dictionaries as well as a number of textbooks in which some attempts to standardize computing terms have been made.

2 The problem of identification of computing basic concepts in the Russian language

In the Russian scientific and academic literature, the term "Computing" is of seldom use; the term "Informatics" is more frequent but it is polysemic, i.e. having several meanings. Along with "Informatics", one can sometimes come across the term "Computer Science", which is a complete calque from the English language.

The problem of using an adequate terminology is more essential for the IT field than for other scientific and academic disciplines, as the majority of basic terms are of foreign origin, mostly English. Of course, one can identify Russian "Informatics" with the generally accepted English term "Computer Science", however they are not completely equivalent in meaning. In view of this fact, one of the university specializations, in the field of IT, was called "Mathematics and Computer Sciences" by its developers from the methodological committee. Using the mono semantic calque "Computer Science", they avoided the term "Informatics", which could have different interpretations.

The evolution of interpretations of the term "Informatics" in Russia has been described in a number of publications. However, at the time of writing this paper, the term "Informatics" appears to be more poly-semantic than many researchers, IT teachers and IT specialists can imagine.

Polysemy, or multiple meaning, is the ability of a word to be used in various meanings, and that of the word "Informatics" is being investigated in (Yu.Yu. Chyorny, 2010). Let us quote: *"In the USSR/Russia informatics had been identified as an academic discipline by 1966 and since then it has been redefined twice: in the mid-1970s and in the early 1990s. However, informatics-3 (let us call the last version this way) did not "cancel" informatics-2, the same way as informatics-2 did not "cancel" informatics-1. As a result, we have three different fields of science with their own subject areas, leading specialists, research institutions, periodicals and training courses"*. "Informatics-3" is considered by Yu.Yu. Chyorny as fundamental science of information processes in nature, society and technical systems; "Informatics-1" is identified as the theory of scientific and information work; and "Informatics-2", in its turn, is understood as science dealing with data processing machines and their application.

When speaking about informatics, most researchers, teachers and IT specialists and users mean the above definition of "Informatics-3". However, even in its narrow sense, the term is not mono-semantic. So its ambiguity is aggravated when it is necessary to identify it with the similar, but not completely adequate, English terms.

Therefore in recent years, some leading Russian scientists put forward the idea that it would be reasonable to introduce a mono-semantic term, having the same meaning as "Informatics-3". The search for such a term resulted in the concept "computing", that is interpreted in the same manner as in the Computing Curricula series (Computing Curricula 2005) as an integral discipline, covering a wide

range of specialized applied scientific sub-disciplines such as computer sciences, artificial intelligence, computer networks, numerical mathematics, database technologies, information systems, multimedia, bioinformatics, etc.

3 General Structure of the Computing Domain Semantic Network

Having analyzed a large set of computing documents, e.g. (Computing Curricula 2005), we developed a general structure of the bilingual system of computing basic terms.

The structure is based on the following principles:

1. A complete coverage of the subject area.
2. Minimization of crossings at the top levels of the structure.
3. Integration of related concepts.

It should be noted that the network structure can be built in different ways, hence it is impossible to determine which of them is the best one a priori.

Figure 1 shows the general structure of a small part of the network under development. At the top level, the network is represented by a set of 12 basic trees, whose "leaves"(objects) could be interconnected at lower levels of the hierarchy.

"The part-whole" relationship is the dominant type ("Has Part", "A Kind Of"), for example: "Algorithms"-"Basic algorithms"-"Sorting algorithms". In the initial stage of development, the network is uniform (having a single type of relationship) and binary (relationships that link two objects). In the process of development, more relationship types, such as "object-unit" kind (e.g., "Sorting algorithm"-"Shaker sorting algorithm") appear, as well as n-ary (with $n > 2$) relations connecting objects that belong to different basic trees. However, the network is not fully tree-structured since it contains cross-sectional connections between objects located in different basic trees. The next stage in building the network will naturally add translational relations linking equivalent concepts from English to Russian between the left and right sub-networks. These relations play an important role in enabling bilingual and cross-lingual information retrieval and machine translation, which can in turn support knowledge management and e-learning search and personalization in the computing domain.

The process of subsequent structuring can fairly quickly lead to the transformation of the tree-like structure of computing terms into a network structure.

4 Conclusion

We illustrated one way to help narrow the language and terminology gap between computational sciences and high performance computing in Russia and English-speaking countries, in particular the USA. The developed bilingual semantic network of computing concepts can become the foundation for building a bilingual knowledge base in the area of computing and can support the automated conversion of traditional information systems into intelligent ones using ontological engineering.

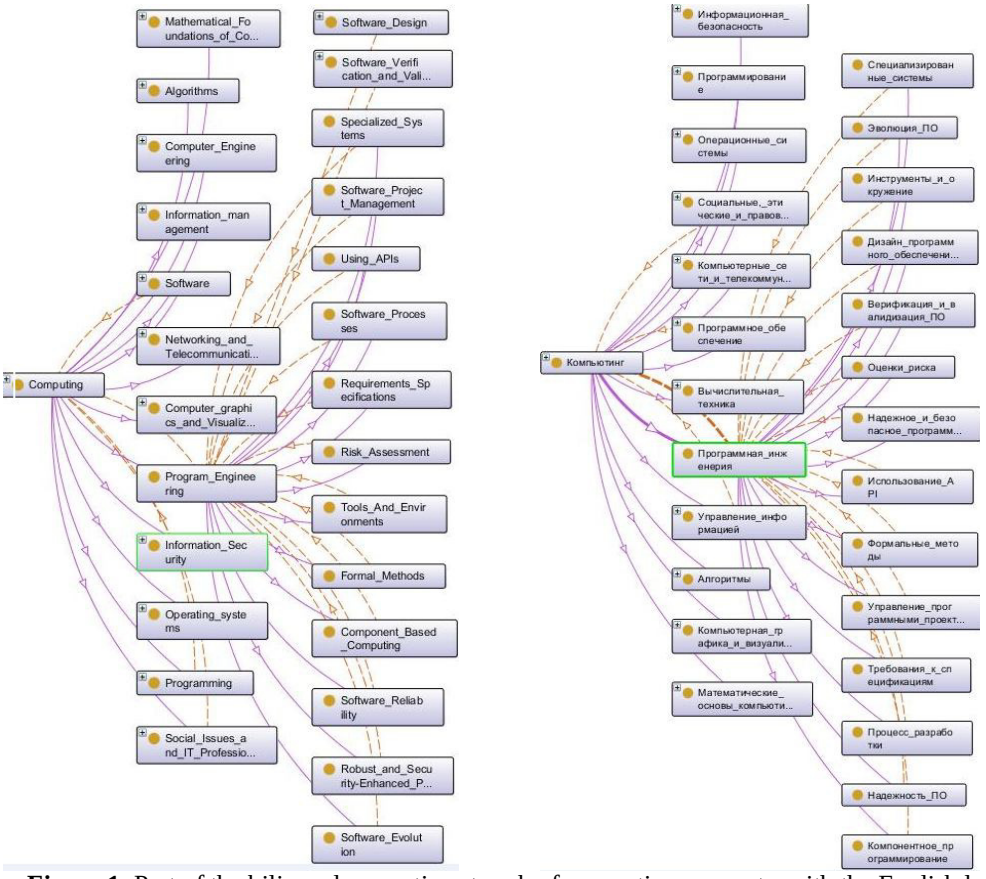

Figure 1: Part of the bilingual semantic network of computing concepts, with the English language sub-network on the left side and the Russian language sub-network on the right side.

REFERENCES

Chyorny Yu.Yu. Polysemy in Science: When Is It Harmful? (a Case Study of Informatics). Open Education. 2010, № 6, pages 1-9. In Russian.

Computing Curricula 2005. The Overview Report covering undergraduate degree programs in Computer Engineering, Computer Science, Information Systems, Information Technology, Software Engineering. ACM, AIS, IEEE-CS.

Khenner E., Semakin I. School subject Informatics (Computer Science) in Russia: Educational Relevant Areas. ACM Journal “Transactions on Computing Education” (TOCE). Special issue on computing education in K-12 schools from a cross-national perspective. June 2014, Volume 14 Issue 2. P.14:1-14:10.

Khenner E. & Stallmann M. Training of IT-specialists: Russia and USA (in Russian). Open Systems. 2013, N3, pages 58-62. In Russian

Pianta E., Bentivogli L. & Girardi C. MultiWordNet: Developing an aligned multilingual database. 2002. In Proc. of GWC-02, pages 21–25.

Vossen P. EuroWordNet: building a multilingual database with wordnets for European languages. The ELRA Newsletter, February 1998, Vol. 3 n.1, ISSN: 1026-8300. Paris. p. 7-10.