

Editorial

Semantic Informational Technologies

The term *semantic informational technologies* (or, shorter, *semantic technologies*) emerged in the 2000s as a generic concept for the qualification of a group of quickly progressing technologies including, in particular, semantics-oriented natural language processing technologies, the use of ontologies in the Semantic Web project and in many other projects of applied intelligent systems, cross-language conceptual information retrieval, ontology-based images recognition and retrieval, the generation of natural language (NL) texts, proceeding from the inner representations of their meanings, the elaboration of content representation languages as a part of agent communication languages in multi-agent systems, and the development of formal means for representing the records of e-negotiations and forming the contracts in the subfield of electronic commerce called e-contracting.

The common features of the technologies from this group is either processing of NL-texts with respect to the fact that lexical items have the meanings (i.e., are associated with one or several semantic items) or/and processing information with respect to an ontology (i.e., with respect to a set of interrelated formal records corresponding to the concepts and the connections of concepts underpinning natural language processing by people).

Overview of the issue

This special issue of *Informatica – an International Journal of Computing and Informatics* contains 7 papers submitted by the researchers from Bulgaria, Czechia, France, Russia, Serbia, Singapore, and Slovenia. The papers were carefully selected on the basis of peer reviews.

Two distinguished features of this issue as a whole are as follows. Firstly, the papers from this special issue describe the studies pertaining to the main branches of semantic informational technologies and, as a consequence, give a rather good initial look at the current state of this field. Secondly, the spectrum of described and discussed subjects is very large: from the industrial applications of the methods and models developed under the framework of the Semantic Web project to the strategy and formal tools of transforming the existing World Wide Web into a Semantic Web of a new generation.

Biology and medicine (biomedicine) are the fields where the methods of semantics-oriented natural language processing (NLP) are being very intensively developed and applied to solving practical tasks. The paper "Obtaining Status Descriptions via Automatic Analysis of Hospital Patient Records" by S. Boytcheva, I. Nikolova, E. Paskaleva, G. Angelova, D. Tcharaktchiev, and N. Dimitrova from Sofia, Bulgaria pertains just to these fields. The paper describes the progress of the

study aimed at automatic extraction of patient status data from medical texts in Bulgarian language. It is shown that certain patient-related facts can be relatively easily extracted from the texts.

The paper "Corpus and Web: Two Allies in Building and Automatically Expanding Conceptual Classes" by N. Béchet, J. Chauché, V. Prince, M. Roche (Montpellier, France) describes an original application of the methods of NLP to building and expanding conceptual classes. To find the effective solutions to this problem is important not only for biomedicine but also for many other fields. The main method of the study was to investigate a semantic-syntactic dependency in a sentence between a verb *Vb1* and its object *Ob1*, proceeding from the semantic dependency between a semantically close verb *Vb2* and its object *Ob2*. As a whole, the paper contributes to bridging a gap between Web-based and corpus-based approaches to forming and expanding conceptual classes.

The paper "Theory of K-representations as a Comprehensive Formal Framework for Developing a Multilingual Semantic Web" by V.A. Fomichov (Moscow, Russia) formulates an original strategy of transforming the existing Web into a Semantic Web of a new generation with the well-developed mechanisms of understanding NL-texts (or a Meanings Understanding Web, or a Multilinguistic Semantic Web). Besides, the paper indicates the basic formal tools being necessary for the realization of this strategy. Firstly, the paper grounds the possibility of using a mathematical model being the kernel of the theory of K-representations and describing a system of 10 partial operations on conceptual structures for building semantic representations (or text meaning representations) of, likely, arbitrary sentences and discourses in English, Russian, French, German, and other languages. The possibilities of using SK-languages (standard knowledge languages), defined by the theory of K-representations, for building semantic annotations of informational sources and for constructing semantic representations of discourses pertaining to biology and medicine are illustrated.

Secondly, the paper describes the correspondence between the inputs and outputs of an original algorithm of semantic-syntactic analysis and indicates its advantages; the semantic representations of the input texts are the expressions of SK-languages. The input texts can be the statements, questions, and commands from the sublanguages of English, Russian, and German.

The next paper "Wikipedia2Onto – Building Concept Ontology Automatically, Experimenting with Web Image Retrieval" by H. Wang, X. Jiang, L.-T. Chia, and A.-H. Tan from Nanyang Technological University, Singapore describes an original approach of the authors to using ontology for better understanding the visual images stored on the Web. This approach includes the construction of a large-scale multi-modality ontology

from Wikipedia for Web images classification. The generated ontology allows for extracting additional information from the Web pages and for increasing the accuracy of concept detection.

The paper "A Service Oriented Framework for Natural Language Text Enrichment" by T. Štajner, D. Rusu, L. Dali, B. Fortuna, D. Mladenić, and M. Grobelnik (Ljubljana, Slovenia) sets forth an original method of complementing the free NL-texts with an enrichment being a set of the triplets of the form *subject, predicate, object*. Due to the use of the triplets, the enrichment can be presented with the help of RDF – one of the basic languages of the Semantic Web project. On the basis of this set of triplets, a semantic graph of a text is constructed. As an example, an enrichment of a short article from Wikipedia is considered. The document's semantic graph is a starting point for automatically generating a document summary. The proposed method is implemented in the applied computer system Enrycher. Several directions of experimenting with this system are outlined.

Two final papers of this special issue will be of particular interest to many readers, because these papers describe the industrial applications of the methods, models, and language means elaborated under the framework of the Semantic Web project. The paper "Applications of Semantics in Agent-Based Manufacturing Systems" by M. Obitko, P. Vrba, V. Mařík, M. Radaković, and P. Kadera (Prague, Czech Republic) shows the advantages of using semantic models of application domains in the design of distributed intelligent control systems in comparison with traditional centralized manufacturing architectures. One of the precious features of the paper is that it contains a substantial discussion of the role of semantics, RDF-based and OWL-based ontologies, and architectures of Semantic Web Services in the design of distributed intelligent industrial systems. A new ontology for manufacturing domain is described; this ontology provides a semantic model of production planning and scheduling, material handling, and customer order specification. The integration of this model with an agent-based simulation and control system MAST is set forth.

The subject of the paper "The Role of the Semantic Web for Knowledge Management in the Construction Industry" by Igor Svetel and Milica Pejanović (Belgrade, Serbia) is the applications of RDF and OWL-based ontologies in the architecture, engineering, and construction industry (AEC industry). It is shown that the principal advantage of this approach is the contribution to preventing construction time delays, unforeseen work and, as a consequence, the exaggerated cost of buildings. The paper gives an overview of the standards developed for providing interoperability and flexibility in the AEC industry and of the standards elaborated under the framework of the Semantic Web project.

The guest editor would like to thank Professor Matjaz Gams for providing the opportunity to prepare this special issue on Semantic Informational Technologies. Finally, many thanks to the authors of the

papers for their contributions and to all of the referees for their precious comments ensuring the high quality of the accepted papers and making the reading as well the editing of this special issue a rewarding activity.

Vladimir A. Fomichov
 Professor of Computer Science
 Department of Innovations and Business
 in the Sphere of Informational Technologies
 Faculty of Business Informatics
 State University – Higher School of Economics
 Kirpichnaya street 33, 105679 Moscow, Russia
 Email: vfomichov@hse.ru