Representation, processing, and sharing of the results of scientific research in information systems based on ontology

(Reprezentacja, przetwarzanie i udostępnianie wyników badań naukowych w systemach informatycznych opartych na ontologii)

Rafał Trójczak

Summary

In recent years, the number of scientific publications has significantly increased. Among scientists there is a widespread belief that so many scientific papers are being published today that nobody alone is able to get to know them all, even when limiting their own interests to a narrow specialization. The situation is probably even more difficult for scientists working at the cross-section of specializations or for practitioners who want to use the results of scientific research in their daily work, which entails a number of negative consequences. Not being acquainted with all the current research in a given field, scientists are risking duplication of the research already carried out. What is more, without knowing all research problems currently being dealt with by a given scientific community, its members are not able to design studies in such a way that they fit into the work of the community.

In order to prevent such a state of affairs the scientific community has taken a number of actions. Some scientific journals and articles provide structured abstracts which are meant to facilitate the absorption of the content of the publication. Another way to deal with the problem is publishing a collection of several "highlights" of a given publication, which succinctly render the claims made by its authors. Thus the reader can get an idea of the content of the publication without having to read the whole.

There are also IT attempts to solve this problem: representation of research results in a form that can be processed by a computer. The result of their application is the creation of a knowledge base of a chosen field or discipline. Despite numerous attempts, none of the existing ways of representing the results of scientific research brings about a satisfactory search results. The existing proposals do not provide an automatic way of deriving new knowledge from what already exists, which is indispensable when the amount of scientific information is constantly increasing. Another disadvantage of the existing solutions is the fact that they do not provide the possibility to search for contradictions between already acknowledged statements.

The main purpose of the work is to solve the aforementioned problem by creating an ontology of the laws of science, which will consist of two parts: 1) a taxonomy of laws with a description of their internal structure and 2) sets of the represented laws. The ontology is called OntoBeef Science, in brief: Science. The theoretical basis for creating a taxonomy of the laws of science are the works of Kazimierz Ajdukiewicz and Władysław Krajewski. The developed ontology have the form of a rich formal theory, which is translated into a knowledge representation language OWL, a standard of the Semantic Web, having its model in description logic. Representing the taxonomy of scientific laws in OWL language enables sharing and processing of scientific laws in information systems and allows for developing algorithms for acquiring new knowledge and searching for contradictions in statements that have been represented.

The basic method used to create the Science ontology was the analysis of the relevant literature on the subject and the analysis of the laws of science occurring in the literature in the field of agri-food science. After analysing both of these sources of knowledge, a knowledge base has been created with the use of tools of applied ontology. The purpose of the work is its application,

because its results are to facilitate the work of scientists and practitioners from the field. The pragmatic approach f the dissertation is also characteristic for applied ontology, which is described in chapter 1, and it can better justify the choice of this particular method for the outlined solution. Another feature of the chosen approach is paying a lot of attention to the possibility of to re-using the research results. This feature can be seen on the example of the Dolce ontology being used as a primary structure of the entities in the field, instead of creating a new solution of this type.

The interdisciplinary nature of the paper is also worth noticing. It contains philosophical considerations regarding the nature of the laws of science, but it also discusses various attempts to define different classifications of these laws. On the other hand, engineering ontology --- usually included in computer science --- is a key element of the findings of the paper, and its product, the Science ontology, is its key achievement. The Science ontology is a logical theory, therefore, the study of its properties should be included in formal sciences, or more strictly - logic. Finally, when it comes to particular scientific laws, they were taken from the field of agri-food sciences and that completes the picture of multidisciplinarity.

The paper consists of an introduction, five chapters and a conclusion. The first chapter contains information on ontology in philosophy and information technology, their history, characteristics and mutual connections. In the second chapter some previous attempts to solve the problem outlined above have been presented. The third chapter discusses the issue of scientific laws, their definitions, selected attempts to systematize it and presents an ontology of scientific laws, first expressed in a natural language, and then - in first order logic and description logic. The fourth chapter presents algorithms created for the processing of scientific laws. The presented algorithms are divided into two groups: 1) algorithms for acquiring new scientific laws; and 2) algorithms searching for contradictions among the laws already represented. The fifth, final chapter is devoted to providing the access to the represented scientific rights. The OntoBeef application created for this purpose is addressed primarily at scientists and practitioners. The application consists of a module presenting concepts, a module with publications, a module with scientific laws and a user module.

The ontology of scientific laws presented in the paper was created as part of a scientific project ProOptiBeef. Therefore, most examples of scientific laws presented in the paper come from the field of agri-food sciences. The the taxonomy structure itself has also been designed with the scientific laws found in papers in this field in mind. The presented taxonomy may be expanded in the future if it turns out that scientific laws in other fields of science cannot be represented in the current taxonomy.

The Science ontology structure is flexible enough for the representation of almost twelve thousand scientific laws from scientific publications on agri-food sciences. The solution presented in the work was used, on the one hand, to generate new research problems, and on the other hand - as the basis of the module of scientific laws in the OntoBeef application. I am not familiar with any other case of developing an ontology of scientific laws, I do not know of any other way of representing scientific research results that would allow to represent the internal structure of the laws using semantic technologies and on this basis to seek new laws or search for epistemic conflicts in the ones already represented. A particularly interesting part of the paper seems to be the presentation of the different ways of acquiring new knowledge from the ontology of scientific laws and a description of the possibility of finding contradictions among the represented laws. With the twelve thousands of registered laws we encounter interesting, original results of aforementioned mechanisms. Along with the increase in the size of the knowledge base, these mechanisms should prove even more effective because, as the size of the knowledge base increases the chances of obtaining new results also grow.

The research started in the dissertation can be developed in several different directions. Above all, complementing the knowledge base with new laws in the agri-food science will positively influence its usefulness and raise, as mentioned above, the effectiveness of deduction techniques. In turn, the application of the developed ontology of the laws of science to another area of interest may reveal deficiencies of the created ontology and indicate the places in which it should be extended. The application schemes presented in Chapter 4 are just a sample of what a developed ontology can be used for. Creating new schemes or extending the existing ones should not pose a major problem. Finally, as presented in Chapter 5, the OntoBeef application can be expanded to take full advantage of the presented ontology and incorporate future amendments.