RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS
No. 85 — 2009
Advanced Information Technologies for Management – AITM 2009

Kinga Kajfosz

Karol Adamiecki University of Economics, Katowice, Poland
e-mail: kinga.kajfosz@ae.katowice.pl

TOWARDS USING SEMANTIC METHODS IN EVALUATION OF INTERNET-BASED INFORMATION TRUSTWORTHINESS

Abstract: One of the serious difficulties we meet using the Internet is the trustworthiness of found information. With regard to this there is a request for a method of Internet-based information evaluation – especially for software agents. In the paper the author described an ontology-oriented approach of information trustworthiness evaluation. The author showed the ontology as a semantic method of knowledge representation. Afterwards the usefulness of the ontology-oriented approach in creation of Internet-based information evaluation systems was proven.

1. Introduction

Nowadays the Internet is one of the most important information sources. It ensures a wide and unlimited access to its resources. In addition, every user of the international network can create some information and publish them on web pages. Due to this fact the Internet is perceived as a collection of unreliable and false information too. Web 1.0 and Web 2.0 technologies do not provide appropriate tools for supporting evaluation of the Internet-based information trustworthiness. Therefore a solution to this problem could be using of semantic methods (Web 3.0).

Tim Berners-Lee – the creator of the Semantic Web (Web 3.0) – defined it as a network which has some intelligence features [Stalmach 2005, p. 151]. It means that Web 3.0 delivers only compatibility with user-defined semantic context information and services. It is possible by usage of advanced filtration mechanisms.

The semantic web is also named "a knowledge network". The data meaning – the most important role in Web 3.0 – is possible to evaluate by dint of a complicate metadata system with knowledge and high trustworthiness. Having this kind of metadata collections causes formation of an environment in which the network works like a semantic pattern transfer channel [Nahotko 2005, p. 146].

Web 3.0 tries to solve the main problem of the Web by creation of programmes called intelligent software agents, which are able to automatic information acquisition from their area. The Internet is the part of this area too. Evaluation of

the acquired information, from its trustworthiness point, constitutes challenge for research in the knowledge engineering. We proposed one of challenging approaches to the information trustworthiness evaluation – ontology-oriented one. First of all we present same fundamental issues connected with the knowledge representation.

2. Ontologies as a method of conceptual knowledge semantic representation

An ontology is fundamental to provide a way for managing, organizing and comparing metadata in the semantic network. In the knowledge engineering the ontology is known as a formal and open representation of general conceptualization [Gołuchowski 2007]. This conceptualization means an abstract model of some occurrences which are identified with regard to their important terms. "Open" means that terms and rules of exploitation are defined [Nahotko 2005, p. 147]. Briefly the ontology defines terms and rules of processing which are used to describe some knowledge area [Koprowska, Sawzdargo 2004, p. 167]. The ontology also should be formal (understandable for computer) and common (accepted by most of scientists) [Nahotko 2005, p. 147].

Ontologies are used to solve unfamiliarity to computers and programming problems. They also help computers to understand relations between data which are part of the knowledge. In the semantic web ontologies represent conceptual models passed by the network [Nahotko 2005, p. 147]. The ontology can be used as the useful mean of the knowledge representation for the information trustworthiness evaluation. We need to create a suitable ontology.

There are known a lot of ontologies classifications. Guarino has proposed following ontologies division (with regard to the level of generalization): a top-level, a domain, a task and an application [Kusztina et al. 2007]. The top-level ontology describes general knowledge about the world and provides general concepts. The domain ontology describes knowledge which is characteristic of defined domain or related to selected reality sector. Used terms are solution of specialization defined in the top-level ontology terms. The task ontology characterizes dictionary connected with task or activity. This type of ontology differs from the domain, because task ontology may use terms from different domains to solve problem. The application ontology contains terms which are required for description of the knowledge used in particular applications. This type of ontology is often a generalization of other, related ontologies (see [Sobczak 2009; Kusztina et al. 2007]).

Ontologies also can be classified according to the following main dimensions [Sobczak 2009]:

– level of formality – by which vocabulary is created and a meaning is defined,
– aim – intended use of ontology,
– character of object which is described by ontology.

Four types of ontology in formal dimension are distinguished [Sobczak 2009]:
- informal ontology which is expressed in natural language,
- ontology with informal structure which is received after addition of some level of structure to natural language,
- semi-formal ontology which is expressed in artificial language,
- formal ontology which is expressed by precisely defined terms with formal semantic.

Sobczak has also distinguished lightweight and heavyweight ontologies [Sobczak 2009]. The first one has defined terms taxonomy and relations between them. The second one is expressed by axioms, it has inference mechanism and characterizes a high level of formalization.

Ontologies can be used in many ways: knowledge representation, natural language processing by information retrieval, digital libraries, databases and multiagents systems [Nahotko 2005, p. 148]. Ontologies include metadata which describe semantic data. By dint of conceptualization software agents can represent and exchange the knowledge – this is cooperation on the knowledge level.

Ontologies are also used in knowledge bases. Software agents can communicate with knowledge bases in order to enrich, repeatedly exploit and manage them [Nahotko 2005, p. 148].

Ontologies can function as a repository which organizes the knowledge in defined communities. In this situation they are used as knowledge acquisition tools. Ontologies are helpful in re-using of the knowledge in new systems [Nahotko 2005, p. 148].

3. An ontology for evaluation of Internet-based information trustworthiness process

With regard to ontology's features and opportunities of using them, ontologies can describe every area of life – also Internet-based information. Evaluation of information trustworthiness ontology should deliver software agents which are searching information in Internet, valuation of information. By dint of it this is less possible to add unreliable or false information to the knowledge base. This ontology should be universal – it means that every information from the Internet is able to be described by it.

Figure 1 shows schema of a system proposed by the author, which should support software agents. Every information in the Internet will be described by special ontology for evaluation of information trustworthiness. Information retrieval software agent sends queries to network. The World Wide Web returns as an answer described by ontology information to another software agent. This agent evaluates trustworthiness of Internet-based information. Received valuation has an influence on adding information to the knowledge base created by an agent. If information is trustworthy, it will be added to the knowledge base.

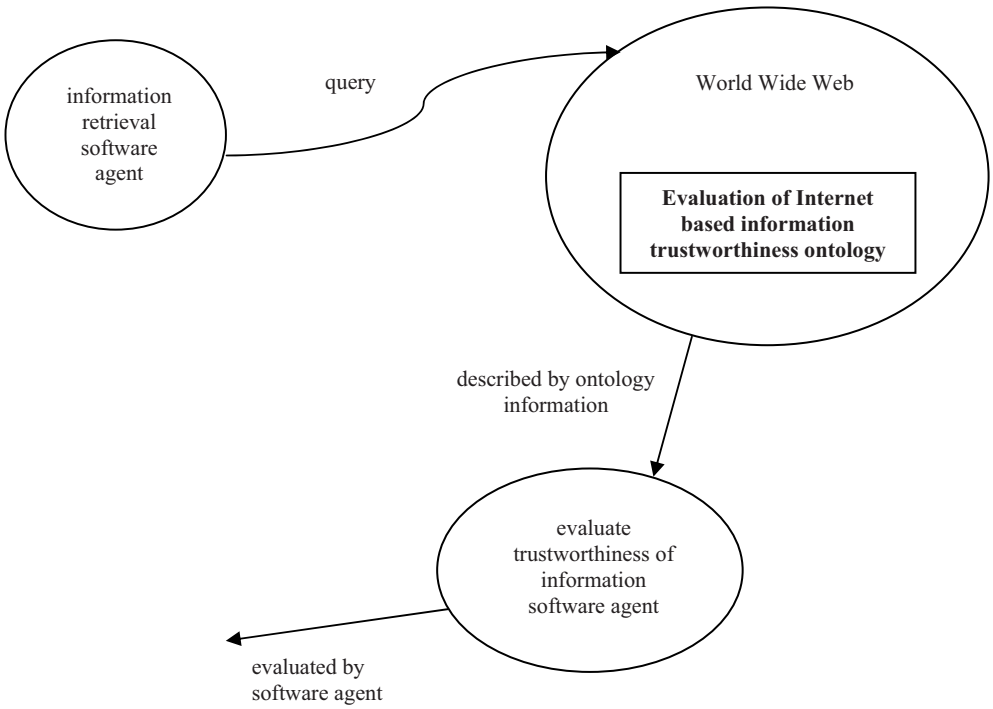

Figure 1. Schema of a system of verification information trustworthiness

Source: own elaboration.

Evaluation of Internet-based information trustworthiness ontology is a task ontology, because it concerns specified activity – evaluation of information trustworthiness. This ontology should be expressed in artificial language – it has semi-formal character. With regard to formality level this should be heavyweight ontology. In case of evaluation of information trustworthiness problem there is necessary to use an inference mechanism which will give software agents valuation of information trustworthiness.

Before we build the ontology, we have to distinguish what has an influence on information trustworthiness. Factors determining information trustworthiness are:

– date – particularly event, creation and modification date,
– creation place,
– author,
– information source,
– aim of information,
– recurrences,
– popularity.

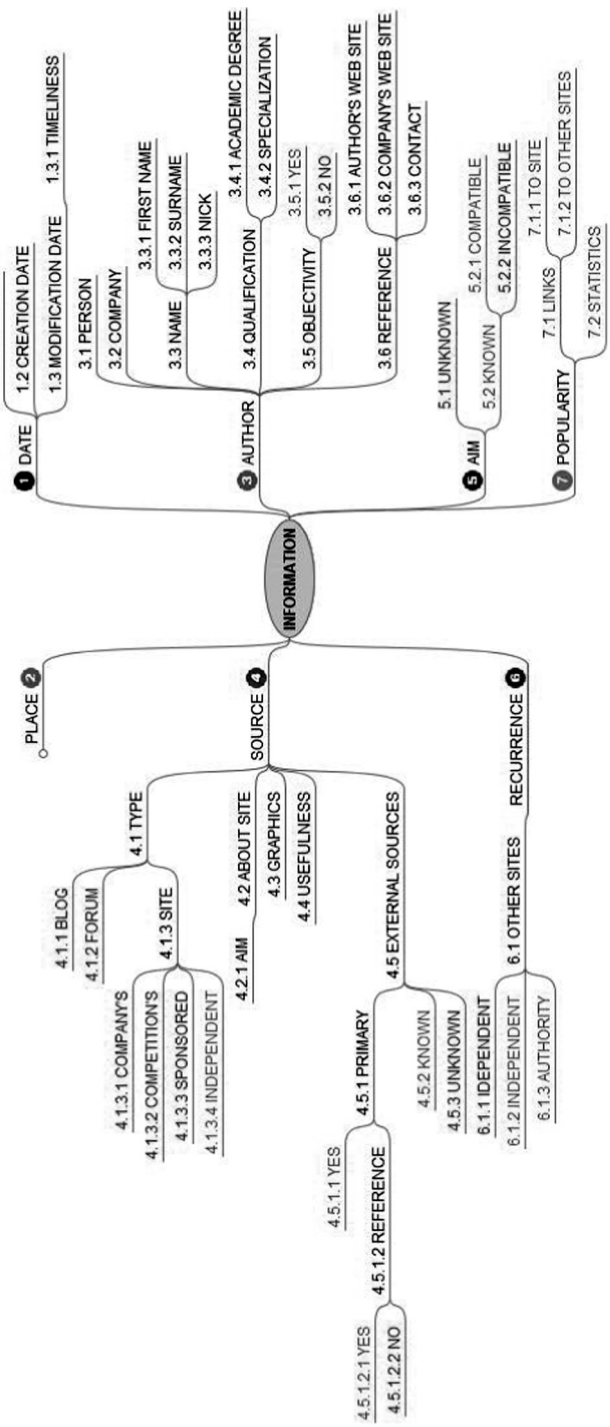

Figure 2. Mind map – factors determining information trustworthiness

Source: author's elaboration

If information is created after long time from describing event, it can be untrustworthy. After some time our mind can deform information. Modification date shows whether information is actual or not. A place, where information was created, is important in evaluation of information trustworthiness. When a region or a city is not in accordance with a country, information cannot be reliable. Also some types of organizations are more trustworthy than others, e.g. government vs. commerce. The next important factor is author. We should know his/her name, qualification and objectivity. Also a source has influence on information trustworthiness. Some types of source, e.g. a blog, forum, sponsored site, are less trustworthy, because they can be subjective. Even information about page, its graphics and usefulness is important for trustworthiness. The aim of information should be known and in accordance with the site aim. Information is trustworthy when other independence sources publish it. Also popularity of source and information have influence on reliability.

Figure 2 shows a mind map which presents factors determining information trustworthiness. Words-keys have got four colours: black, blue, green, red. Black colour informs that this factor only describes information and does not have an influence on its trustworthiness. Blue colour means that the factor can have positive or negative influence on trustworthiness. Green colour characterizes factors which have positive influence on information trustworthiness. Red represents factors with negative influence.

4. Conclusion

Evaluation of Internet-based information trustworthiness is important but very difficult. Nowadays software agents often search information in the international network, but they need some support in trustworthiness evaluation. Proposed in the paper approach to the verification of information trustworthiness can help software agents to evaluate information acquired from Internet. Evaluation of Internet-based information trustworthiness ontology plays main role in this system.

Essential advantage of proposed system is prevention of addition of unreliable or false information to knowledge base. Presented ontology is only a prototype which is under construction.

Proposed approach has some limitations: it needs universal ontology and expert system to evaluate trustworthiness of information. In future papers the author will work out evaluation of Internet-based information trustworthiness ontology and create expert system.

References

Gołuchowski J. (2007), *Technologie informatyczne w zarządzaniu wiedzą w organizacji*, AE, Katowice.

Koprowska M., Sawzdargo R. (2004), *Technologie wspierające procesy wyszukiwania informacji w przyszłej sieci WWW*, [in:] *Internet w społeczeństwie informacyjnym. Podstawowe problemy Internetu*, Ed. A. Grzywak, Wydawnictwa Naukowo-Techniczne, Warszawa.

Kusztina E., Różewski P., Ciszczyk M., Sikora K. (2007), *Struktura ontologii jako narzędzie opisu wiedzy dziedzinowej*, http://pan.wi.ps.pl/pliki/MetInfStos/200702/MetInfStos_2007_02_Art_08.pdf [21.03.2009].

Nahotko M. (2005), Metadane i ontologie w systemach zarządzania wiedzą, [in:] *Profesjonalna informacja w Internecie*, Ed. M. Kocójowa, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, pp. 144-149.

Sobczak A. (2009), *Analiza wybranych metod budowy ontologii*, http://swo.ae.katowice.pl/_pdf/271.pdf [18.03.2009].

Stalmach L. (2005), Nowe mechanizmy dystrybucji informacji w Internecie. Metadane i ich standaryzacja, [in:] *Profesjonalna informacja w Internecie*, Ed. M. Kocójowa, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, pp. 150-156.