Cognitive Ontology of Information Security Priorities in Social Networks

Aktayeva A.
Almaty technological university,
Almaty, Kazakhstan
Email: aakhtaewa@gmail.com

Makulbek N.,
Shatenova G.
Kazakh Academy
of Transport and Communications
named after M.Tynyshpayev,
Almaty, Kazakhstan
Email: nmakulbek@mail.ru
Email: shatenova94@mail.ru

Galiyeva N.,
Naraliyev N.,
Baiman G.
Belarusian state university
of informatics and radioelectronics,
Minsk, Belarus
Email: nggaliyeva@gmail.com
Email: nishonali@gmail.com
Email: bgb$_z ht@mail.ru$

***Abstract*—This article considers the importance of the field of knowledge used to describe the information security ontology, and the structure and the basic principles of a technology to create a competent SPARQL-based identification profile for a user of the Social Internet Network.**

***Keywords*—Social network, Internet, linguistics, semantics, SPARQL, ontology, identification.**

I. Introduction

The information security process should be comprehensive and based on a thorough analysis of possible negative consequences. This analysis implies the obligatory identification of possible threat sources, and those factors, which contribute to their vulnerabilities, and, as a consequence, the identification of the relevant threats to the information security. Based upon this principle, it is advisable to model and classify sources of threats and their manifestations relying on the analysis of such logical chain interaction as given below (Figure 1).

Figure 1. The logical chain analysis of the interaction

In this regard, the term “threat” means a potential threat or a factual threat of committing any action against an information resource protection object, which causes damage to the owner or user, and makes itself evident in a risk of distortion and/or loss of information.

The information relations subjects determine a great deal of information resources that must be protected from attacks of any kind whatsoever. They are the result of threat implementation and made through various protection vulnerabilities and are probable (risk - attacks).

From the analysis of the protection vulnerabilities, the threat source properties and the probabilities of a possible implementation of the said threats in a particular environment, risks should be identified for a given set of information resources. This, in turn, enables the protection policy to be specified, which is set by the security policy. The security policy determines a consistent totality of security mechanisms and services adequate to the values to be protected and the environment where they are used.

II. System information security - ontology

The knowledge representation using ontologies in information retrieval systems for information security makes it possible to execute multi-aspect structured queries that can be represented as a graph. Based on the ontology technology, the information security system may offer to concretize, or vice versa, to expand a query if it is satisfactory too much (too little) to objects. Furthermore, the ontology allows us to propose the user himself to overview the field of knowledge using the concept navigation, moving from one concept to another due to the links between them.

Language classes based on different mathematical models to define the object domain ontology are shown in Figure 2.

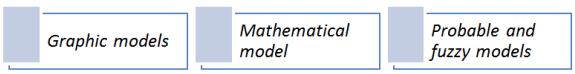

Figure 2. Language Classes of Mathematical Models

In the systems based on the ontologies, the methods used to extract information are shown in Figure 3.

The major advantages of using ontologies for knowledge representation may be noted as follows:

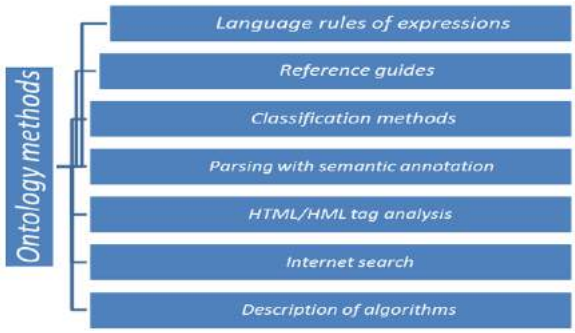

Figure 3. Available Information Extraction Methods

- Flexibility of a data model, which enables the model to be modified and extended with relative ease.
- Possibility to re-apply the existing ontologies.

The ontology is a method of representing knowledge by using a finite aggregate of concepts and relationships between them (Figure 4).

Figure 4. General Ontology Conceptualization

III. The Methodology of Using Ontology

If the ontology is a formal object domain model expressed, for example, as a graph of concepts and relations, then it generalizes a hierarchical data structure usually used for the filling in the extraction task. The information extraction is traditionally to find data, which describes some domain knowledge specified by the data structure.

To extract information, a prepared source selection-based training and / or heuristic models may be used that may be based, for example, on the usage of predetermined lexical and syntactic patterns, or ontologies.

The information extraction direction by using ontologies has stood out from the general information extraction problem relatively recently but has been already marked as a perspective trend towards the development of information extraction systems.

In many systems, the information extraction includes the following steps:

- Information search;
- Extraction of terms;
- Extraction of name groups such as names of people, organizations, and geographic positions;
- Extraction of words and word-combinations, which designate one and the same object (coreference resolution);
- Normalization of terms, which makes it possible to connect them with a formal description of the object domain;
- Extraction of semantic relationship between the terms;
- Duplicate record search;
- Normalization of records, i.e., reduction thereof to the standard form (see Figure 5). Possibility to re-apply the existing ontologies.

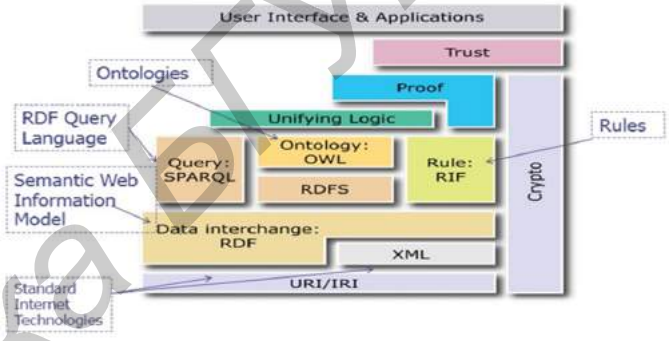

Figure 5. W3C Semantic Web Stack

The building and population of ontologies are closely related to the information extraction by using ontologies. The building of ontologies includes extraction of sets of concepts and names of relations as well as instances of the said concepts and relations between them, and the population of ontologies implies concepts, relations, and is aimed at the search of instances of concepts and relations between them.

It should be noted that the problems of building and populating ontologies are currently very pressing for the following reasons:

1. The construction and population of ontologies require the development of algorithms for an automatic highlighting of information from texts in a natural language. The most part of information in the Internet is contained just in this form. The manual processing of such data requires a good deal of permanently increasing human resources due to the huge volumes of stored information. This is precisely why the intelligent algorithms that automate the said process, acquire great importance;

2. The populated ontology is a stock information resource for a semantic web. To realize vision of the said semantic web, automatic metadata generation means are required. The semantic annotation makes it possible in future to process the said information by machines embodying the concept of the Semantic Web;

3. The population of ontology may be used to improve its quality. The basic idea is that if the ontology helps to

effectively extract the necessary information from texts, one can conclude that the ontology adequately describes the field of knowledge.

The major issue related to the information extraction is that the existing systems either need the learning under such documents as marked manually by experts, or only enable the data to be extracted from structured texts. The notion or the essence is a class of individual objects or instances, and types of relationship between the concepts are shown in Figure 6 [4].

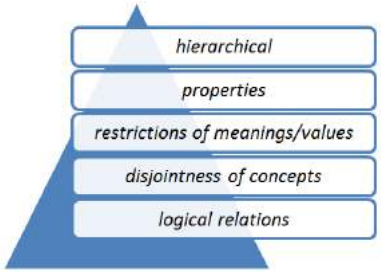

Figure 6. Types of relationship between the concepts in the ontology

IV. Modern problems of information security systems: ontology

Recently, the researchers have been paying more and more attention to the problem of extracting information from the World Wide Web. The Web HTML-pages have the advantage over straight texts that they contain markup elements such as lists, headings and tables.

Another feature of Web-documents is the fact that information therein is usually generated automatically from some databases that changes the process of extracting information from such texts to the process of "decoding". Finally, we should mention a great deal of data available in the Internet and their heterogeneity.

The above-listed features determine a number of methods usually applied when extracting data from the World Wide Web. If in the traditional systems of extracting information from unstructured texts the natural language processing methods are usually applied such as dictionaries and grammars, algorithms for the machine learning and extraction of templates are used more often in the systems of extracting information from the Internet, which are based on the syntactic properties and visual structure of Web-pages (Figure 7).

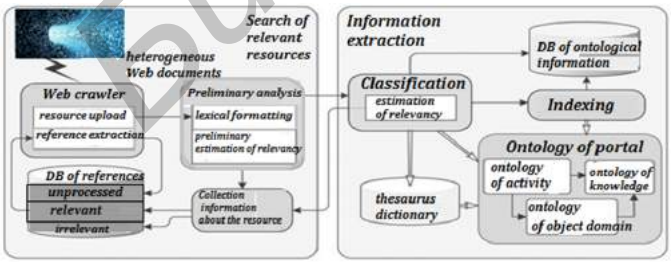

Figure 7. Operational diagram for information object when working with users

Information on the Internet Web-pages is divided into two classes:

1. Information generated automatically from the structured databases;

2. Information published manually.

The information search is to select the relevant documents from a large collection. In this task, the document is nothing more than mere words, whose meanings and relations are not considered. The search engine does not generally enable the complex analytical queries to be executed that require the analysis of the content of documents. The data extraction direction, which is less developed, just assigns a task to single out a structure, i.e. a value of information from unstructured texts.

The task to extract information from an unstructured text is important in the data analysis and processing direction. The relevance of the task is caused by the rapidly growing volume of unstructured information, for example, in the Internet. In general, the information extraction implies a certain data structure or template to be filled with such information as contained in the text data in the natural language, or, in other words, instances of certain classes of objects or events, and relations between them to be identified. According to the study of Russell and Norvig, the information extraction is "in the middle" between the information retrieval, which consists in the selection of those documents that meet the user's query, and the understanding of the meaning of a text implying a deep analysis of the said text with a view to identify its semantics [1, 3, 4, 10, 11, 13].

The information security ontology based on a case approach is shown in Figure 8.

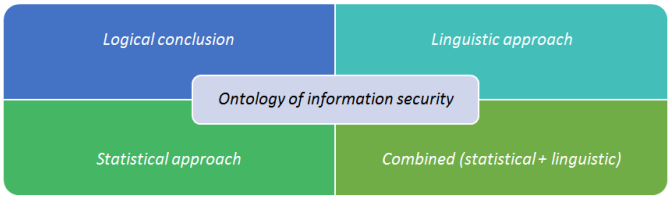

Figure 8. Ontology-based information security methods

On the one hand, it allows the use of a powerful mathematical tool and the obtainment of a theoretical proof of the efficiency of the algorithms. On the other hand, the most of these approaches have one significant drawback. Its essence is that they are based on the assumption of independence of the occurrence of words in a sentence. For the natural language, such a hypothesis is an excessively strong [7].

Unlike the linguistic approach, the major issue in the implementation of the statistical approach is a "silence" - a situation where the terms consisting of one word remain unnoticed by the system. In the studies of the recent years, statistical and linguistic methods are often combined. Here the ACABIT System and TRUCKS can be noted [7].

Thus, the use of ontologies ensures the compliance with the requirements for the developed system, which consists in the need for the inference of a new knowledge.

V. The classification of knowledge manipulation

The technological elements of the information modification system are a manipulative impact on the users of social networks. For this purpose, a complex criterion is used based on the accounting of combination of the following parameters:

1. Frequency of the use of technologies;

2. Range of application thereof;

3. Level of impact.

Subject to the complex criterion, those groups of manipulative methods may be singled out, which are the most multipurpose, with a high frequency of occurrence in different information war technologies having an area of application some information and communication situations such as public debates and group discussions, speeches at public rallies and demonstrations, in mass media, in intergroup and interpersonal conflicts characterized by a sufficiently high efficacy and impact on the human mind. These methods are characterized by a high degree of expressiveness in all three parameters of the manipulative impact on the users of social networks.

The application of text message processing methods to protect against information modification is critically important in cases where the unambiguous identification uses characteristics of information and communication facilities by calculating network data, such identification methods.

A lot of online resources and services such as forums, portals, online stores are faced with various aspects of the problem of manipulation and an artificial formation of public opinion by "organizing" purposeful thematic dialogues where a number of users have several accounting records.

The methods of the information and psychological impact on the mass consciousness are divided into the following seven basic groups of the information and psychological impact:

1. "Name calling";

2. "Glittering generality";

3. "Transfer";

4. "Reference to influencers", "on recommendation", "testimonial";

5. "Plainfolks"; "Card stacking"; "Dealing a card from the bottom of the deck";

6. "Common platform" or "bandwagon" [2,15].

The possibility to use portals and sites for dissemination of information and insufficient functionality of user identification and authentication mechanisms, those users who leave messages, determines a number of ways to improve protection systems and monitoring systems for information security and information and telecommunication facilities.

VI. Model developed Ontology System

The ontological approach to knowledge representation allows the use of the existing and approved algorithms to execute analytical queries. Execution of analytical queries to the data is ensured in the course of interaction of the system end user with the software implementation of a model describing the field of knowledge.

The information extraction problem is different from the information retrieval and query problem. The information extraction systems may be be divided into the following four types according to the degree of participation of an expert in the system development and adaptation:

1. Customizable manually where the user defines, in a certain language, rules for information extraction from specific websites.

2. Training/learning. The user manually marks a training set of documents, which is used to build an information extraction module.

3. Partial training/learning. The user does not mark the whole training set but only provides some additional information, for example, he selects a template from the options provided by the system and marks data to be extracted.

4. Without training/learning. The system automatically marks the teaching selection and creates an information extraction module completely without the participation of the user [10, 11, 13].

A query in using ontologies may be executed automatically by the inference engines. As a query language for ontologies, SPARQL may be used.

In the paper, examples [7] of using SPARQL are given. A relation between the queries, the formal model of a system under development and the query code in the SPARQL language allows the monitoring of the impact of:

1.Modifications of the plurality of received requests in the system and the ontologies used upon the in the software system code;

2. Modifications of the software system code on the ontologies used and the requests under consideration.

The latter circumstance creates additional opportunities for an effective verification of the software at all stages of its life cycle [7].

In such cases, to identify the user, mathematical linguistics methods may be applied. Thus there occurs a need to develop:

- An identifier model for an Internet portal user based on the tuple of linguistic features of a short message;
- A method for creating a an Internet portal user based on the identifier model containing a tuple of linguistic features;
- A method for identifying an Internet portal user based on the component profile;
- A method for creating an Internet portal user's component profile, which implies a number of steps to be performed;
- Processing user communications within the Internet portal;
- Analysis of messages by the parts of speech followed by the use of templates (syntactic patterns) to select the most common phrases;

- Lexicographical analysis of a message and selection of phrases structures in accordance with the described patterns, and collection of statistics about the use of punctuation marks and special characters;
- Selection of lexical phrases the basis of words and word forms of the language, and also identification of thematic special words and phrases typical for audience of a specific form.

Once the Internet portal user's profile has been studied, as well as the essence of the methods and means applied, and the reasons, which induced to commit violations, these reasons may be either affected, or the requirements for protecting against such violations may be defined more precisely.

And the model of a violator should reflect his practical and theoretical capabilities, a priori knowledge, time and place of action etc.

In developing a user profile model, the following should be determined:

- A category of persons, to whom the violator may belong;
- Causes of the violator's actions (goals pursued by the violator);
- Qualification of the violator and his equipment (methods and tools used to commit a violation);
- Restrictions and assumptions about the nature of possible actions of the violator.

According to the results the researches revealed that, one of the most promising areas of scientific research as related to the analysis, forecasting and modeling of semi-structured and badly formalized phenomena and processes are a fuzzy logic.

The mathematical theory of fuzzy sets (fuzzysets) and fuzzy logic (fuzzylogic) are generalizations of the classical theory of sets and the classical formal logic. The fuzzy inference algorithms differ mainly in the form of rules used, logical operations and a kind of the defuzzification method. Some fuzzy inference models are developed by Mamdani, Sugeno, Larsen, and Tsukamoto [84]. Such approach allows us to solve problems of improving the functioning of different systems under the conditions of insufficiency and unreliability of information on the running processes if assessment is subjective.

The requirements for the other stages of implementing the algorithm for building ontologies and the entire algorithm overall are as follows:

1. The algorithm should extract hierarchical and associative relations between the terms;

2. The ontology should reflect the relevant state of the specified field of scientific knowledge;

3. The accuracy and completeness of extracted terms and relations should not yield to the existing and approved algorithms for building ontologies;

4. The algorithm should require no training, or there should be a possibility to receive the necessary teaching selections from the public sources without spending much effort for their processing;

5. The algorithm should not require a large scope of manual labor of experts in the predetermined object domain tuning;

6. The sources of data for the algorithm should be publicly available and regularly updated;

7. The algorithm should have a modular architecture;

8. The possibility of configuring the algorithm automatically for a specific field of knowledge.

The information security ontology must contain as much information as possible about the field of knowledge, in particular, not only the hierarchy of concepts and trends but also non-hierarchical (associative) relations.

VII. Conclusion

The formal model of the ontological object domain is a graph of concepts and relations. The use of ontologies is a natural development of the direction, in addition, mathematic models and algorithms, and architectural and technological solutions may be developed relying on the ontologies to establish a system of replenishment and storage, analysis and issue of that information on request, which characterizes the performance of the User - information on the Internet Web-pages. The hierarchical structure of data is used for the filling in/population in the information extraction task. To identify users on the Web-pages of social networks, the following methods should be applied:

- Identification, the use of which is difficult due to the possibility of changing technical characteristics of the device;
- Determination of the authorship of a text after a linguistic correction, which will require a substantial adaptation for the processing;
- Development of a tuple of linguistic features of a short message, which enables the identifier construction features to be taken into account.

By using the ontologies and SPARQL technology, queries to the system may be formally described, thus, creating guarantees of their calculation and additional features for an effective verification of the system code at all stages of its life cycle. It is necessary to create a software prototype to keep record of and to analyze data when modeling and identifying the user's profile in social networks under the conditions of the information war.

The information extraction direction by using ontologies has stood out from the general task of information extraction relatively recently but has already been marked as an outlook for development of information extraction systems. The information extraction traditionally aims to find data that describe a certain field of knowledge specified by the data structure.

References

[1] G. S.Jowett and V. O'Donnel *Propaganda and Persuasion*. Newbury Park, 1992

[2] D. O. Kuchumov *Problema regional'noj informacionnoj bezopasnosti (na primere osveshhenija v regional'nyh SMI terroristicheskogo akta v g. Beslan)*, No 4. Bezopasnost' Evrazii. Moskva, 2007.

[3] Stuart J. Russell *Artificial Intelligence: A Modern Approach / Stuart J Russell*, 2d ed. Peter Norvig: Pearson Education, 2003.

[4] A. Aktayeva, N. Galyieva, N. Naralyiev, and etc. *International scientific journal «Modern IT and IT - education»*. vol.12, No2, 26 pp. 2016

[5] Jim Cowie and Yorick Wilks *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text / Ed. by Robert Dale, Hermann Moisland, Harold Somers*. New York, USA: Marcel Dekker,2000

[6] K. K. Kolin *Philosophy of information: the structure of reality and the phenomenon of information*. No 4, 61 pp. Metafizika, 2013.

[7] I. M. Azhmuhamedov *Achieving information security based system analysis and fuzzy cognitive modeling*. 340 p. Monograph, Astrakhan, 2012

[8] A. D. Ursul *The nature of information: philosophical essay*.231 p. Chelyabinsk, 2010

[9] M. M. Gorbunov-Posadov *Internet-aktivnost' kak objazannost' uchenogo // Informacionnye tehnologii i vychislitel'nye sistemy*. No 3, 88 s., 2007

[10] M. E. Suhoparov *Metodika identifikacii pol'zovatelej portalov seti internet na osnove metodov matematicheskoj lingvistiki // 05.13.19 – Metody i sistemy zashhity informacii, informacionnaja bezopasnost' - Dissertacija na soiskanie uchenoj stepeni kandidata tehnicheskih nauk* Sankt-Peterburg – 2015

[11] S. Afonin *Minimal Union-Free Decompositions of Regular Languages // Language and Automata Theory and Applications / Ed. by Adrian Dediu, Armand Ionescu, Carlos Martın-Vide*. vol. 5457 of Lecture Notes in Computer Science, 8 Berlin, Heidelberg: Springer-Verlag, 2009

[12] D. D. Golomazov *Vydelenie terminov iz kollekcii tekstov s zadannym tematicheskim de-leniem*. No 2. S. 8 Informacionnye tehnologii, 2010

[13] V. A. Vasenin, i dr. *Ispol'zovanie semanticheskih tehnologij dlja obnaruzhenija grid-resource*. No 7, 2 pp. Programmnaja inzhenerija, 2011

[14] W. Yeong *Lightweight Directory Access Protocol.*Vol. 2251, No 1777, 1 pp., Search, 1995

[15] O. I. Borovikova *Ontologicheskij podhod k postroeniju sistem informacionnoj pod-derzhki nauchnoj i proizvodstvennoj dejatel'nosti // Materialy Vserossijskoj konferencii s mezhdunarodnym uchastiem «Znanija – Ontologii – Teorii» (ZONT–09)*. T.2. 93 pp. Novosibirsk: Institut matematiki im. S.L. Soboleva SO RAN, 2009

[16] A. Aktaeva, etc *Development of a mathematical model of information warfare // International Journal of Open Information Technologies* vol. 2, No11, 2014, 28-33 pp., www.injoit.org, https://doaj.org/toc/2307-8162

КОГНИТИВНАЯ ОНТОЛОГИЯ В ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ СОЦИАЛЬНЫХ СЕТЕЙ

Актаева А., Галиева Н., Наралиев Н., Байман Г., Шатенова Г., Макулбек Н.

Процесс обеспечения информационной безопасности должен быть всеобъемлющим и основан на тщательном анализе возможных негативных последствий. Этот анализ предполагает обязательную идентификацию возможных источников угроз, а также те факторы, которые способствуют их уязвимости, и, как следствие, выявление соответствующих угроз информационной безопасности. Исходя из этого принципа, то целесообразно моделировать и классифицировать источники угроз и их проявления, опираясь на анализ такого логического взаимодействия цепи. В данной статье рассматривается важность области знаний, используемых для описания онтологий информационной безопасности, а также структура и основные принципы технологии для создания компетентного SPARQL на основе профиля идентификации для пользователя сети Интернет социальной.