

# Some Approaches to the Development of Information Influence and Hidden Communications Detection Systems in Wiki-Environment

Irada Alekperova

Institute of Information Technology of ANAS, Baku, Azerbaijan  
airada.09@gmail.com

**Abstract**— The overall structure of hidden connections and information influences detection system in a wiki-environment is presented in this article. Several concepts are available for detection of information influences, as well as hidden connections between users: OLAP, Data Mining, also advantages of given mechanisms are listed.

**Keywords**— *Wikipedia; information war; information influence; data warehouse; analytical systems; OLAP; Data Mining*

## I. INTRODUCTION

Nowadays information wars conducted using computer networks, as well as internet are gaining a daunting character. Information wars, which are the natural results of development of information technologies, create preconditions for development of new internet-technologies: social networks, blogs, virtual encyclopedias etc.

These technologies do not only unite people, they also give them ability for information exchange, its free discussion and self-expression. They also serve as means of misinformation, propaganda, opinion control, in a word, for conduction of information war in online mode [1, 2].

## II. CAPACITIES AND PROBLEMS OF WIKIPEDIA

Wikipedia virtual encyclopedia presents a doubtless interest for research of information wars and influences in internet-environment, in particular for searching task of hidden communication among users. There are many reasons allowing to successfully working with this project, research communications among users:

- Wikipedia is aimed to create an online encyclopedia, written users themselves. Each visitor of the site can bring in their strong contribution: correct an article, add or delete information, participate in discussions [3];
- Page history is set, where it is possible to examine all changes in the article and editing pages for each article;
- Each article belongs to one or several categories, determining thematic directions. Categories for articles are selected by article authors from the set of existing categories. New categories can be created, connecting it with existing ones.

Main principal of Wikipedia – is complete democracy, that's why a beginning author can participate in the project along with experienced experts [4]. Overall observance demonstrates that articles on conflict subjects, that have hundreds and thousands of corrections become battle fields [5]. The sizes of discussion pages often exceed the size of the article itself.

There are more than 280 Wikipedia in different languages, with different numbers of crosses and connections, which continuously grow by adding new crosses and creation of new connections [6].

Information influence in Wikipedia can be classified by different signs: on influence character on users and articles, on influence goal, conditions of influence implementation situation, position of impacted subject, influence scale etc.

Many edits on wiki-articles are of political character [7]. Analyzing number and character of corrections in one article, as well as studying editors we can create a clear image of relations among nations and countries amid an information war [8].

During investigation of information relations in Wikipedia, it is important to study functional connections among events and analyze hidden influences. For this reason, it is proposed to use statistical data that can be obtained from Wikipedia pages.

## III. OVERALL STRUCTURE OF HIDDEN CONNECTIONS DETECTION SYSTEM

Detection of informational impacts on users of Wikipedia and hidden influences is based on forecasting determination of possible changes in articles, as well as on user pages and subsequent observation of all changes.

Herewith, different active audit means view the same incident from different points of view. Similar impacts and hidden connections among events in wiki-environment can be detected using hidden connections and information influences detection system.

Information influence detection system collects information from Wikimedia Foundation data bases, as well as from site that have an access to Wikimedia Foundation

data bases, which collect information on edits of articles and authors of these edits [9]. Analyzing this information we can detect different knowledge, such as misinformation implementation attempt, as well as real violation of neutral point of view which is considered as the main rule of Wikipedia.

System is intended for development of intellectual analysis methods based on messages collected in wiki-environment. In such intellectual system, following main elements are distinguished logically: data warehouse, analytical data analysis system and data presentation module.

Data warehouse is used to collect primary information on activity of Wikipedia users. Analytical data analysis system carries out the search for hidden connections.

Structural data system consists of one or more analysis modules – analyzers. Several analyzers can also be used. Presence of several analyzers is required to increase the effectiveness of detection. Each analyzer carries out the search for hidden connections or influences of a certain type. For analyzer, input data is information from information collection subsystem or from other analyzer.

In case when analyzer informs about detection of hidden influences, some additional information may occur upon exit. Usually this information contains conclusions confirming fact of influence presence, entry of misinformation in Wikipedia pages or direct attacks – deletion of important texts from articles, insulting of users.

Information influence detection system can be used for solution of following tasks: for exposure of vandalism and agitation, as well as collection of information about abnormalities on Wikipedia. Herewith, vandalism cases on Wikipedia can be divided in two categories: accidental and intentional.

Accidental, incorrect edits of users or abnormal operation of the system is often developed as deviation from normal operation. Deliberate article distortion: entry of misinformation, agitation, advertising, spam can be described by means of features and signatures.

OLAP and Data mining are proposed to be used in detection system of hidden connections among users and their influence on the article. Data mining is one of the most important data warehouse applications.

OLAP (On-Line Analytical Processing) allows conducting fast, complete and effective analysis of large volumes of data. OLAP, as well as Data mining, is an integral part of any process of decision making support

system. Use of OLAP is designed to provide “access” to multidimensional data [10].

Data mining systems are used for exposure analysis of hidden data. Both concepts can be used in proposed system, considering that both OLAP and Data Mining are decision making support systems. If necessary, it is possible to use proposed system to drill down to contents on multidimensional data in order to obtain detailed information.

#### IV. CONCLUSION

Proposed system is designed to provide reliability to Wikipedia and increase the quality of Wiki-articles. Detection of hidden connection among users, as well as articles which create and edit in real time scale requires a significant amount of data.

Having an information storage and data mart, system is able to expose different types of information influences and detect abnormalities in Wikipedia using clustering. Main advantage of given mechanism is that after determining comparable behaviors of participants, it is possible to minimize a large number of false alarms related to information input and article editing in Wikipedia.

#### REFERENCES

- [1] Changwang Zhang, Jianping Yin, Zhiping Cai, Weifeng Chen, RRED: Robust RED Algorithm to Counter Low-rate Denial-of-Service Attacks // IEEE Communications Letters, vol. 14, 2010, pp. 489-49
- [2] Howard J.D., An Analysis of Security Incidents on the Internet: 1989-1995, doctoral dissertation, Dept. Eng. and Public Policy, Carnegie Mellon Univ., Pittsburgh, [www.cert.org/research/JHThesis/Start.html](http://www.cert.org/research/JHThesis/Start.html), 2000
- [3] Buriol L.S., Castillo C., Donato D., Leonardi S., Millozzi S., Temporal Analysis of the Wikigraph // Web Intelligence (2006), Publisher: IEEE CS Press., pp. 45-51
- [4] Alguliyev R.M., Alekperova I.Y., Kasumova R.T. Collective knowledge in Internet/Sixth international conference “Internet – Education – Science – 2008”. Volume 1, Vinnitsa, 2008, p. 5-7
- [5] Pnina Shachaf, Noriko Hara, Beyond vandalism: Wikipedia trolls // Journal of Information Science, 36(3), 2010, pp. 357-370
- [6] [http://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](http://meta.wikimedia.org/wiki/List_of_Wikipedias)
- [7] Alexander Halavais, Derek Lackaff, An Analysis of Topical Coverage of Wikipedia // Journal of Computer-Mediated Communication, Volume 13, Issue 2, 2008, pp. 429-440
- [8] Steffen Blaschke, Klaus Stein, Methods and Measures for the Analysis of Corporate Wikis: A Case Study / In: Proceedings of the 58th Annual Conference of the International Communication Association ICA, May 22-26, Montréal, Canada, 2008, pp 1-24.
- [9] [http://en.wikipedia.org/wiki/Wikimedia\\_Foundation/](http://en.wikipedia.org/wiki/Wikimedia_Foundation/)
- [10] Alguliyev R.M., Alekperova I.Y., Kasumova R.T. Multidimensional model of data in MS SQL SERVER 2000 // IX International scientific conference “Theory and technology of transmittal, reception and processing of information”, Ukraine, KNURE, 2004, p. 28-29.