Valid formation of scientific schools

Iryna Zheliznyak

Information Systems and Networks Department,
Lviv Polytechnic National University, UKRAINE, Lviv,
S. Bandery street 32a, E-mail: iryna.zhelizniak@gmail.com

Abstract – In this paper are described the methods of clustering papers by scientific schools. The algorithm of clusterizing abstracts based on the developed models and methods of frequency analysis of terms in sentences and determining the weight of sentences was reviewed. Also, in this paper i analyze the algorithm of summarization of natural language documents using the concept of importance coefficients.

Key words – clustering, text information, scientific schools, ontology, scientific direction, electronic libraries.

I. Introduction

The task of processing text data requires the use of different algorithms of division. You also need to use different mechanisms for additional data analysis and partitioning into different subcategories for further work.

Using the selection of keywords from the text makes it possible to search for relevant information long enough period of time.

Keyword - a word or phrase sustainable natural language used to express some aspect of the content of the document (or query); a word that has significant meaning. It can serve as a key when searching for information on the Internet or on the page.

Art keyword selection, extracting the most important or characteristic fragments of one or many information sources, has become an integral part of our lives. News offered - a summary of world events of the day. Quotes securities - "dry residue" of information about the sale, which continually generates market. Although some manufacturers already offer tools for keyword selection, volume information in the quickly growing and getting it correct construction is becoming more difficult. Tools such as Autosummarize feature in Microsoft Office, System IBM Intelligent Text Miner, Oracle Context and Inxight Summarizer (a component of search engine Altavista), certainly useful, but their options are limited allocation and selection of original fragments of the original document and connecting them short text [1]. Training is a summary aims to describe the main content of the text.

II. Data integration in electronic systems

The main difference between the means of the selection of keywords is that they essentially form a summary or collection of quotes from a particular material. Both types of presentation have two main objectives: to identify the most important point of full text and highlight key word.

Until recently, the general disciplines enjoyed increasing popularity, however, distribution mechanisms, and full-text search filtering information adapted to the requirements of specific users, lead to the fact that the essays that are set, are becoming increasingly important.

Using abstract methods to optimize the size of the text, but does not provide for automatic search documents. You can define the 3 most important points are not taken into account in the allocation of keywords from the text:

a) the division into parts on keywords (formatting);

b) allocation of keywords (by weight);

c) forming a synthesis document (in the form of statistics)

During data integration in electronic systems to provide a common understanding of the unified interface for user access to a set of independent sources that tend to have heterogeneity on some of their properties. The original class systems integration represent a system in which the basis for the Open Archives Initiative technology (Open Archive Initiative - OAI). Most famous of this category of information resources are a collection of text documents, especially scientific publications independently formed at the nodes of the global network, supported and managed by their owners. [2]

According to the technology OAI, provides for integration embodied in a single repository of information resources not of interest to users of the system integration, and presented some standard way of metadata describing the collection of information resources and sources of the archive collections of the individual elements. Collecting these metadata repository is made according to a specially developed protocol Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), which provides global access services and search.

There are several approaches to the problem of creation of electronic libraries with integrated information sources. Among them are the following two classes of systems:

- integrated Management of resources;
- driving with distributed resources.

Today there are two conceptual solution to this approach. First shows existence of a mechanism for cross-searching multiple archives, when all resources, bibliographic descriptions and search engine located in the organization. This search is made by direct appeal to all or selected users E in order to be able to provide centralized in one place search-based libraries, followed by the construction of the results in a single list. Second - offers to collect metadata describing information resources collected metadata. In fact it is some analog integrated electronic directory.

III. Concept and importance of scientific schools

From development and research work focused largely depends scientific image and quality of university students. Dictionary defines school as a scientific direction in science-related unity of common beliefs, principles and techniques of continuity [3]. By definition K.Lanhe, scientific school - is an informal research group, formed around the renowned scientist at the research institution that combines for collective development of certain scientific ideas, problems and direction of a number of separate research teams.

Scientific school - a community of people, formed under the auspices of personality - the scientist and leader who has ideas, themes for development. No leader - no school. The best schools - those where the followers of the leader involved active research in topical areas and the combined ideas, methods, research traditions, enhanced cooperation, search of new facts [3].

One of the most important results of functioning scientific schools are scientific publications of its members.

Among the factors effective functioning scientific schools are the following:

• determining the scientific direction of scientific topics relevant profile, prospects of its development;

• formation of research units (institute, department, laboratory, center), university faculties, chairs;

• formation of research teams, careful planning research;

• establish a modern logistics research base;

• publication of basic scientific works: monographs, scientific manuals, articles in professional journals, including international;

• availability of professional scientific periodical;

• annual scientific events, symposiums, conferences and seminars.

Signs scientific school is the presence of the scientific community that develops in time and space; focus on the development of new, original direction in science; common research interests, principles and methodological approaches in carrying out productive research program; multiple generations of scientists (link "teacher - student"), a recognized leader united; improving scientific skills of participants schools; publication of scientific results (publications, reports).

The criteria for recognition of academic schools are matching profile theme state priority areas of science and technology programs of the Ministry of Education, Youth and Sports, the National Association sciences industry; doctoral and master's theses in the direction of the school; availability discoveries, inventions; publication of monographs and articles in professional journals, deposit accounts; organization of scientific events: the annual international or national conferences, permanent seminars; are based on current school research and production structure of the state level.

Clustering is important in obtaining relevant information, navigation, text summarization and organization of documents available online, electronic libraries and corporate networks.

Therefore, to date, relevant is the problem of finding optimal solutions to the problem of clustering documents. Often happens when the entered user query returns a huge number of references to the text and other materials. Typically, the query results are linear lists with a simple transfer of objects, sorted by relevance to the query.

At a time when a large number of objects, natural desire is to get the user a short list of categories of documents found. Breakdown of total documents in the category subject to the problem of clustering data.

Scientific direction - is the sphere of scientific research team to address some significant fundamental problems.

Scientific school - scientific team, which aims at addressing scientific direction [4].

This study determined set of scientific school publications Sch, characterized by multiple keywords Key, multiple authors set Author and school founders Main:

$$Sch = \langle Key, Author, Main \rangle, Main \in Author$$

Input information for referring to the publication of scientific school of a file containing publication. Since the file is necessary to define the basic characteristics publication:

1. The author (s) of publication (A);
2. Scientific Institution (B);
3. Topic publication (C);
4. Keywords (D);
6. Article.

After the analysis of data and received the necessary information, you can start clustering publication.

Ontological approach - almost best to search for articles of a scientific school. However ontologies Ukrainian language is not so much. Therefore, the use of ontological search possible at the next stages of building a system analysis of scientific schools.

Conclusion

In this paper constructing of a formal model of data space and multiformat data processing techniques as well as possibility to clusterize abstract from sources of various formats are considered.

Summarizing of documents is an urgent and important task, which is very difficult for modern system of abstracting. Resources for referencing could have varied nature, so traditionally they required different pre-processing but using weights theory increases the quality of the resulting abstract. Methods of abstracting which are considered in this thesis can be successfully applied in modern information systems.

References

[1] Dunham M. Data Mining Introductory and Advanced Topics / M. Dunham ; Pearson Education. — Inc., 2003. – 64 p.

[2] Exploratory mining and pruning optimizations of constrained associations rules // Proc. of 1998 ACM–SIGMOD Conf. on Management of Data, Seattle, June 1998 y., Washington. – Washington, 1998. – 13 p.

[3] Obvintseva GI Development of information model monitoring scientific, technological and innovation activities in Ukraine with the use of electronic information resources / GI Obvintseva. - K .: STI, 2010. – 91 p.

[4] Agrawal R. Fast Discovery of Association Rules / R. Agrawal, R. Srikant // Proc. of the 20th International Conference on VLDB, September 1994, Santiago, Chile.– Santiago, 1994.– 487 p.