

Application of Thesaurus for the Identification of the Specific Situations

Natalia P. Tuchkova¹, Alexander A. Muromskij²

*Dorodnicyn Computing Centre, FRC CSC RAS
Vavilova 40
Moscow, Russia*

¹ *E-mail: natalia_tuchkova@mail.ru*

² *E-mail: murom@ccas.ru*

Abstract

The research is devoted to the use of symbolic expressions for informational queries. The problem of information search, is discussed with use of designations and special symbols in various subject domains. It is proposed to use specialized thesauri, where along with natural language definitions are present symbolic expressions. This approach allowed refining a retrieval, regardless of the language of the original source and can also be applied to the identification of critical situations on the basis of an analysis of symbolic indicators underlying the structured data. Proposed technology comparison texts by analogy and mechanism for replenishment of the thesaurus, as well as the possibility of establishing additional semantic relationships that may speed up the search for information in a critical situation. It is offered to use these data for finding of 'special points' in subject domains and to operate them.

Keywords: Informational retrieval, thesaurus, subject area, semantic search, ambiguous situation.

1. Introduction

Even in the 50's of the XX century, Calvin Mooers [1] formulated not only the notion of 'information retrieval' (as 'it search is unknown where what is unknown exists or not', but also the condition when this search makes sense. Namely, the availability, ease and convenience of obtaining information should form the basis of the information system, and most of the failed projects, according to Mooers, did not take into account these factors.

An empirical 'Mooers' Law' was proposed [2]. It says that some information retrieval [IR] systems are used, while others are not available and this is decided by users. Mooers also has the idea of hashcoding and searching by pointers, and not by the data itself. Over the past period, in more than 60 years, there has been a long way in improving the accuracy and convenience of IR- systems.

The qualitative barrier was overcome by moving from indexation on the basis of traditional thematic headings to the use of subject domain (SD) thesauri. To implement an effective search for subject domain data (SD data) representation, pointers, dictionaries and thesauri are used.

This approach allows to provide access to information through secondary documents and to search not for the document itself, but for their

'information images'. In this case, intellectual search, expert systems and artificial intelligence systems are often mentioned. Logical-linguistic models in management that would allow to formalize information (knowledge) on the object and the subject of management were proposed by D. A. Pospelov [3].

One of the continuations of these works was the study of the analysis of large data flows in order to identify trends [4]. The applications of these studies are very broad from economic and natural sciences to social ones, connected with processes in management.

In this paper, it is proposed to formulate a description of the context of the situation and then get to formulate options for changing the situation based on known or identified trends. On the example of use of information technologies in management identification of critical, 'deadlock' situations in development of scientific disciplines is analyzed.

2. Tracking of a Situation on the Basis of the Thesaurus of Incidents

Description of the situation in the form of a list of terms with a taxonomy and symbolic indicators is created at the stage of information gathering and lexical-semantic analysis of data. To identify the situation, it is suggested to use mnemonic codes and symbols as a

convolution of information. Earlier authors discussed convolution of information and it is supposed to use the methods offered in work [5].

2.1. *The purpose of controlled vocabulary*

The main function of thesaurus of situations is accumulation of data on the situations close to critical, incidents and the conflicts. It is possible to learn about features of dictionaries of controlled lexicon in works [6, 7]

As a result of such model it would be desirable to receive the following:

- understanding the nature of conflicts of interests and clearer wording of rules of an exit from them
- identification of trends, fixation of a situation on the basis of accounting of attributes
- formulation of criteria and parameters for assessment of situations
- coding of descriptions of situations and their estimates for fast information transfer about a situation and search of similar situations

Using links in the thesaurus, it is possible to model regular or critical situations, to define whether there is a conflict of interests, to choose option of a way out. For this purpose situations and options of an entrance (exit) need to be set in the description, with attributes, numbers (codes, identifiers) that search and information transfer were made more effectively.

2.1.1. *Application to the situation description*

The application is for the development of the information support systems and making decision systems. The questions that discuss

- description of a situation with use of controlled lexicon and/or symbolic notation
- special controlled lexicon for subject domain

This approach allows to specify a query as independent of original language

2.1.2. *Methods*

- *Logic - linguistic models* in management which would allow to formalize information (knowledge) of an *object* and a *subject* of management (prof. D.A. Pospelov [3])
- *Semantic technology and analysis of big data flows* for the purpose of identification of trends (prof. V. F. Horoshevskij [4])
- Description of the situation in the form of a *list of terms with a taxonomy, symbolic indicators and linked open data* (LOD technology, Tim Berners-Lee [8])
- To use mnemonic codes and symbols as a *convolution of information* (coding theory) (prof. A. I. Cherny [9])

Let see on critical and conflict situations in scientific research because they can be allocated from regular, routine, ordinary circumstances and to present at the information level to the IR system.

2.2. *Subjects and objects of a situation in science*

We will consider situations in science and related areas where scientific achievements are directly connected with practice of their application (for example, medicine, psychology, information technologies, etc.).

In the settled sense the subject in science is the researcher (the certain person or the scientific organization), and an object is a research problem (a task, process, the phenomenon) is considered. It is possible to find more general approach to such definitions, for example, in work [10]. Managing is exercised by support (financing) of the subject and during this period there are conflicts of interests, in particular, depending on the status of subjects.

It is no wonder that critical situations in science result from the conflict of interests. For example, it is specified in the work [11] devoted to problems of human development in Russia that lobbying of interests in the hydrocarbonic energy sector leads to the conflict between the innovative and raw strategy of development of economy and as result, to braking of development of science and technology.

The analysis of the various relations of the groups and persons involved in the system of medical care are present in work [12]: drug manufacturers, doctors, patients, students, teachers and public organizations. There some definition of the conflict of interests is given, the index of terms is made, models of conflict situations and their permissions are offered. Definitions from work [12], it is possible to refer also to other fields of science and research activity in general, namely: 'Conflicts of interest are defined as circumstances that create a risk that professional judgments or actions regarding a primary interest will be unduly influenced by a secondary interest'.

The word is 'unduly' in this definition use as it is about medical practice, and unconfirmed medicine (a treatment method, etc.) can really be illegal. If to treat the word 'unduly' as it 'impossible', then we will receive definition of the general conflict in scientific research.

Let consider that 'the main interest', is scientific result, opening and its application in some sphere of human activity. The personal conflicts, the conflicts of 'schools of sciences', financial and legal aspects get to 'secondary'.

If the scientific idea, has passed a way from a hypothesis, through the proof to concrete result and was demanded by business, then from 'a situation in science'

we appear in 'a situation in business'. However, it is known enough results which have outstripped the time and/or unclaimed.

It means that researches in some SD can be stopped though, it can be the incorrect decision from the point of view of 'the main interest'. The scheme of such conflict, is very simple and to describe it terminologically simply

Conflict resolution is consists in two options:

- (i) to continue researches,
- (ii) not to continue a research

The first case isn't required to be explained. And in the second, there are options. For example, it is possible to change criteria for assessment of result and to make the decision that the research is important for SD, that is, to continue financing.

Thus, criteria for estimation of researches and the practical importance of results are necessary for situation resolve.

2.3. Features of management in science

Modern administrative structures in science actively use scientometric indicators for assessment. There were publications where for assessment of development of some disciplines in science in the different countries the quantity of 'Hot papers' and 'Papers with delayed recognition' is analyzed [13]. These are indicators of activity of publishers more than researchers which nevertheless influence an economic component of development of science [14].

Assessment of the situation, decision-making and managements by means of an information system except a scientometric requires the following

- SD information model;
- use of expert estimates;
- possibility of correction of the purpose of researches taking into account ensuring 'the main interest'.

There are important features of management in science, and it is connected with self-organization of research teams. Distinctive feature of research, scientific activity is the self-organization based on combination of the efforts connected with interest in some scientific problem. It is impossible without this interest and only on the basis of financing to organize researches. In the book by Fritjof Capra [15] he told about a social phenomenon of knowledge, and this aspect can't be ignored in the analysis of a situation. In science there were always 'singles' and modern individual exists separately, now with digital communication. Some scientists would be ready 'to look for the truth in silence' and that nobody operated them. But time of singles has passed in science. Isolation in science leads to degradation therefore the

aspiration to association around a scientific idea, formation of 'schools of sciences', followers and opponents of the idea is natural. In this capacity, science is conservative.

One more distinctive the fact that in science compete (clash), both subjects, and objects of management. The ideas and researchers are compete. And the scientometrics reflects this feature especially.

In all these cases the initial task appears: identification of sense of the competition, conflict, critical situation.

The scheme of interaction of objects and subjects of management of a situation in science is shown in fig. 1. In the scheme it is reflected that two "polar" phenomena, such as management and interest in researches can exist separately. Influence of management can strengthen or weaken interest in science generally. It repeatedly occurred in the history during the different historical periods of flourishing and decline of science [16].

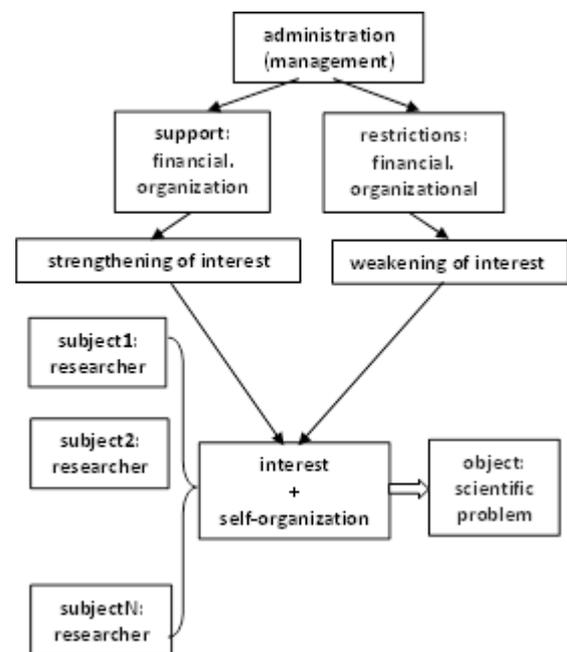


Fig. 1. The scheme of interaction of objects and subjects in science.

2.4. Information model of a situation

The aim of creation of models of situations to receive the decision on management at the result. Let use controlled vocabularies as a way to organize knowledge for the subsequent extraction of the instruction for management [17].

Information model generally includes the following blocks:

- (i) description of subject domain as the thesaurus, ontology

- (a) terms of subject domain
- (b) designations, synonyms, reductions
- (c) links to qualifiers
- (d) mnemonic identifiers
- (e) communications of terms
- (f) references to publications
- (g) references to authors
- (ii) options of situations
 - (a) the term is isolated (few communications and publications)
 - (b) the term is demanded (many communications and publications)
- (iii) behavior options
 - (a) development of the section of subject domain (grants, budgetary financing, etc.)
 - (b) giving of sign 'successful'
 - (c) termination of researches (reduction of financing and shots), giving of sign 'problem'
- (iv) block of the automatic analysis
 - (a) archiving of codes from options of behavior (iii)
 - (b) monitoring "problem" and "successful"
 - (c) elimination of "unsuccessful" project
 - (d) recommendations about support of "successful" projects
- (v) block expert analysis
 - (a) acceptance or denial of "decision" from the block of the analysis (iv)
 - (b) correction of the decision (iv)
- (vi) modification of the description of subject domain
 - (a) addition of new terms and links
 - (b) addition of publications, links, qualifiers.

In Fig. 2 the is a scheme of communications in model is shown. Data can put to the expert block after the preliminary 'automatic' decision or at once from the block of options of behavior. The expert can request it. From model we can receive at the exit of assessment of activity of researches and option of management. These estimates without expert opinion are formal.

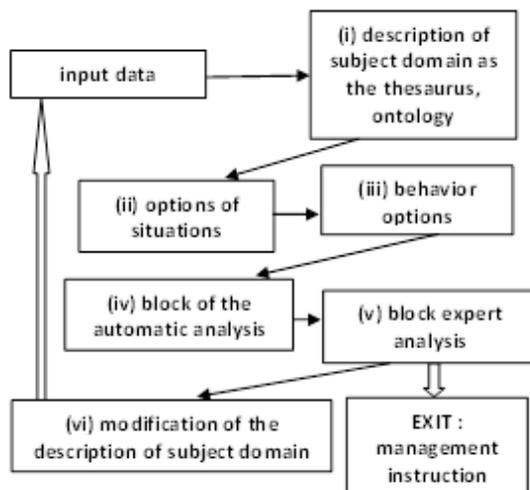


Fig. 2. The scheme of interaction of objects and subjects in science.

In total with scientometric data it is possible to receive a picture of "deadlock", crisis points in researches:

- - it is a lot of publications and research activity, but there are no 'breaks' and opening
- - 'the isolated point' in subject domain, but nobody farther and has promoted in this area

Conveniently in this sense to use tools graphical representation of knowledge where the offered metaphor of the 'isolated' points can be illustrated on the example of graphic ontologies of subject domain [18]. These situations demand decision-making unlike when 'all are successful' and 'all move ahead on the way of opening'. Analyzing these 'points', taking into account 'ensuring the main interest', it is required to understand, deadlock it is branches in development or not. It is obvious that only experts, but not information systems can make such decisions.

It is possible to mark out the conflicts of interests in groups with 'horizontal communications' and the conflicts of interests in 'hierarchical structures'.

The conflicts which aren't requiring the expert solution can be 'horizontal' if there is a collision of interests in group, it is the deadlock [19].

The competition in scientific community (horizontal communications, to a competition of authorities)

- priority copyright of new result
- competitions of grants
- rating competitions

The 'hierarchical structures' are part of the administrative conflicts with the managing director and the financing structure - 'are vertical', it too - the deadlock [20].

The deadlock conflict can carry destroying, leading to an impasse or transferring a situation to higher stage of development of researches [21].

The conflict of the ideas in subject domain is a development. These points can be reflected in formal assessment of the item (iii). As a result we will receive model of management of a situation in researches and options of development (exit) from a situation. For compact data presentation it is offered to use the mechanism of convolution of information [5]

2.5. Example of Subject domain model

The example includes both search in formulas [22] and text processing of article for formation of a search image on subject domain of 'the equation of mathematical physics' [23].

Conflicts of management of researches in the field of classical mathematics have a long story, beginning from Leibniz and Newton. Newton and Leibniz dispute on a priority of opening of differential and integral

calculus is known to all (Leibniz–Newton calculus controversy) [24]. It is an example of one of the most known conflicts of interests in the field of researches in classical mathematics. This dispute has been taken out to the public sphere and served promoting of mathematical knowledge. The polar points of view concerning Newton or Leibniz's priority expressed historians of mathematics up to the beginning of the 20th century. Such conflicts in one subject domain contribute to the development of science and make one of science engines in general.

Importance of researches in the field of classical mathematics doesn't weaken. Now there is a set of results of their researches and the dispute on priorities came to the level of new numerical methods and applications. It belongs to applications in the field of the mixed equations.

For formation of the index the mixed equations, tasks for these equations, namely, of the equation, Tricomi of a task and their generalizations are considered by Tricomi. Links of the terms adjoining Tricomi problem [25] is established. Among them results of such authors as M.M. Smirnov [26], S. Gellerstedt [27], F. Frankl [28], P. Germain [29] and others.

The following structure of the index is chosen: all concepts, tasks, the equations, related are located in separate lines, paragraphs of articles of the index which are numbered. For each number the line (lines) in which are located also a surname of the author of work, names of work with reduction of formular dependences and also the relevant sources published, provided in a line, a surname of authors is allocated.

Example of article of the Index on the equations of the mixed type is in *Table 1*:

Table 1. Example of structure of article of the index. Index of mixed equation

Element	Value
Term ID	EMIX002
Term En	Tricomi equation
Term SYN	equation T
Term AUT	Tricomi Francesco Giacomo
MSC	35M10 Equations of mixed type
LaTex	$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$
References	1. Smirnov M.M. Equations of mixed type. American Translation of the Mathematical monographs. Vol. 51. Mathematical Soc., 31 Dec 1978. p: 232 2. Gellerstedt S. Doctoral Thesis, 1935; Jbuch Fortschritte Math. 61, 1259.

The Index is used in a IR systems LibMeta [30] for links of authors, articles, formulas:

- entrance for search - the term
- entrance for search - a formula
- entrance for search - the author

The effect of semantic links is based on use of terms from the thesaurus mathematical SD, supplied with formulas as bases for mathematical inquiry. Leaning the thesis formulated by Mooers use of the IR-system has to bring to authors benefit, for example:

- enrichment of the thesaurus of the addressee due to application of new terms from the thesaurus SD;
- use of terms from the thesaurus as keywords of the publication for specification of the place of work within SD;
- application of terms of the thesaurus in search queries by drawing up the abstract list.

Terms from the thesaurus mathematical SD, supplied with formulas, make a basis for mathematical inquiry. The terminological description SD on the basis of dictionaries and thesauruses comes to the forefront at solutions of the problem of search of new result.

The expert receives at the exit result with assessment - new, not new and makes the decision on assessment of the research project.

3. Conclusion

The economy of XXI centuries is structured around flows of information. Use of modern information technologies has exerted impact on management in various spheres. For the scientific sphere it has affected in account scientometric indicators. On the basis of these indicators many estimates and conclusions about a situation in various fields of knowledge become. Nevertheless, without features of subject domains and the conflict of interests of these indicators becomes insufficiently. Use of thesauruses of subject domains for identification of communications of researches and researchers allows to reveal tendencies and to focus researches in a certain direction.

Acknowledgements

The work are support by Russian Foundation for Basic Research, the project #17-07-00217a.

References

1. C. Mooers, Information retrieval viewed as temporal signaling, in *Proc. of the International Congress of Mathematicians 1* (1950) pp. 572–573.
2. C. Mooers, Mooers Law, or why some retrieval systems are used and others are not, *J. Bulletin of the Association for Information Science and Technology*. 23 (1996), pp. 22–23. doi:10.1002/bult.37.
3. D. A. Pospelov, *Logic-linguistic models in control systems* (Jenergoatomizdat, Moscow, 1981).
4. V. F. Horoshevskij, Discovering new technological trends in texts collections: hybrid models and data patterns time series analysis, *Information-measuring and Control Systems*, **11**(5) (2013) 25–34.
5. A. A. Muromskij, N. P. Tuchkova, Ontological approach to the data protection for their transfer and archiving, *Ontology of Designing* **6** 2(20) (2016) 136–148.
6. What is a Controlled Vocabulary, and how is it useful? <http://www.controlledvocabulary.com>.
7. P. C. Enros, *May Dictionary and Thesaurus Projects, Scope and Status. Report Institute for the History and Philosophy of Science and Technology* (University of Toronto, Toronto, 1980).
8. <https://www.w3.org/DesignIssues/LinkedData.html>
9. A. I. Chernyj, *Introduction to the theory of information retrieval* (Nauka, Moscow, 1975).
10. Yu. V. Karyakin, The subject and subject of science *Advances in current natural sciences 1* (2008) 76–79.
11. Human potential as a critical resource of Russia, Ed. B. G. Yudin. (RAS Institute of Philosophy, Moscow, 2007).
12. Institute of Medicine (US) Committee on Conflict of Interest in Medical Research, Education, and Practice; Lo B, Field MJ. Eds. *Conflict of Interest in Medical Research, Education, and Practice*. (DC): (National Academies Press (US); Washington, 2009).
13. L. Bornmann, A. Y. Ye, F. Y Ye, Identifying 'hot papers' and papers with 'delayed recognition' in large-scale datasets by using dynamically normalized citation impact scores, *J. Scientometrics* (2018). <https://doi.org/10.1007/s11192-018-2772-0>.
14. R. R. Kinouchi. Scientometrics: the project for a science of science transformed into an industry of measurements, *J. Scientae studia* **12**(1) (2014) 147–159.
15. F. Capra. *The Hidden Connections: Integrating the Biological, Cognitive, and Social Dimensions of Life Into a Science of Sustainability* (Doubleday, NY, 2002).
16. V. I. Vernadskij *Nauchnaya mysl' kak planetnoe yavlenie. 1937-1938.* (Nauka, Moscow, 1991)
17. ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (<https://www.niso.org/publications/ansiniso-z3919-2005-r2010>)
18. WebWOVL <http://vowl.visualdataweb.org/webvowl.html>
19. V. D. Wall, L. L. Nolan, Small Group Conflict: A Look at Equity, Satisfaction, and Styles of Conflict Management, *J. Small Group Behavior* **18**(2) (1987) 188–211.
20. J. A. Wall, R. R. Callister, Conflict and its management, *Journal of management*. **21**(3) (1995) 515–558.
21. A. C. Amason, Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: resolving a paradox for top management teams, *Academy of Management Journal*. **39** (1) (1996) 123–148.
22. E. I. Moiseev, A. A. Muromskij, N. P. Tuchkova, *Internet and mathematical knowledge: representation of equations of mathematical physics in the information retrieval environment* (MAKS Press, Moscow, 2008).
23. A. N. Tihonov, A. A. Samarskij, *Equations of Mathematical Physics*, (Gostekhteorizdat, Moscow, 1953)
24. A. F. Cajori, *History of Mathematical Notations* (Paquin Printers, Chicago, 1929) **2**, pp. 211–216.
25. F. D. Tricomi. *Lectures on partial differential equations* (Foreign Languages Publishing House, Moscow, 1957),
26. M. M. Smirnov, Equations of mixed type. American Translation of the Mathematical monographs. *J. Mathematical Soc.* **51**. (1978).
27. S. Gellerstedt. Doctoral Thesis.; *Jbuch Fortschritte Math.* (1935) **61**(1259).
28. F. Frankl. On Cauchy's problem for partial differential equations of mixed elliptico-hyperbolic type with initial data on the parabolic line (Russian, English summary) *Bull. Acad. Sc. URSS, Izvestia Akad. Nauk SSSR, Sér. Math.*, **8** (1944) 195–224.
29. P. Germain, An expression for Green's function for a particular Tricomi problem, *Quart. appl. Math.* **14**. (2) (1956) 113–124.
30. V. A. Serebryakov, O. M. Ataeva, Information model of the open personal semantic library LibMeta, in *Proc. XVIII All-Russian Scientific Conference "Scientific service in the Internet" Novorossiysk, 19-24 Sep 2016.* (Moscow, Keldysh Institute of Applied Mathematics of Russian Academy of Science) 304–313.