

**INFORMATION MONITORING SYSTEM: A PROBLEM OF LINGUISTIC RESOURCES CONSISTENCY AND VERIFICATION**

**Natalya Dulina<sup>1</sup>, Olga Kozhunova<sup>2</sup>**

<sup>1</sup>Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS (CCAS RAS), Moscow, Russia

<sup>2</sup>The Institute for Informatics Problems of the RAS (IPI RAS), Moscow, Russia  
<sup>1</sup>ngdulina@mail.ru, <sup>2</sup>okozhunova@ipiran.ru

At present Information Monitoring has gained high dissemination in various spheres of human activities. It provides for a great range of opportunities to monitor, analyze, and assess subject data from relevant perspectives and tasks. Systems designed and developed according to the goals of Information Monitoring are termed as Information Monitoring Systems [1, 2].

Thus having a wide spread, Information Monitoring Systems were suggested to be applied for scientific results and performance monitoring. One of the first attempts made in Russia is the Information and Technology System of Monitoring of the Russian Academy of Sciences (ITSM RAS). Its prototype was designed and implemented in the Institute for Informatics Problems of the Russian Academy of Sciences (IPI RAS).

One of the incentives for ITSM RAS creation was a set of regulations approved by the Russian Government and Presidium of the RAS. Firstly, the system was assumed to provide Heads of the RAS and senior staff of Institutes with information on the results of formation, implementation, and monitoring of the Basic Research Program of the RAS.

This involved demand in formulating a set of terms and concepts of the field which was quite a challenge since procedure of Information Monitoring lacked any linguistic tools. Therefore, at the stage of design of the ITSM RAS necessity for relevant linguistic resources was taken into account. The scheme of the system comprising several types of resources is given at Figure 1 [3, 4].

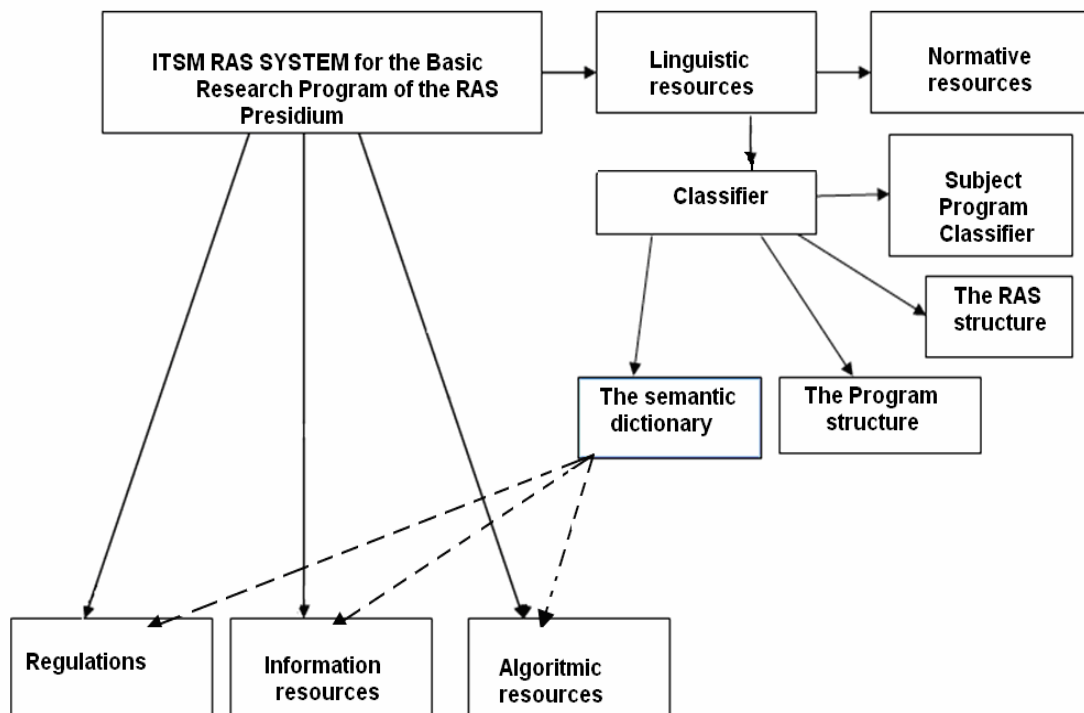


Figure 1. ITSM RAS structure and its components

In the paper we consider only one component of the Linguistic resources, i.e. the Semantic Dictionary (Fig. 1). The reason lies in its specific functions: clarification of terms meanings, calculation of Information Monitoring indicators and coordination of their meanings [3, 4]. The dictionary architecture is based on the classification scheme with subsumption and functional thesaurus relations. Such a dictionary structure allows integrating existing indicators, their variants and to support completion of the dictionary by new indicators.

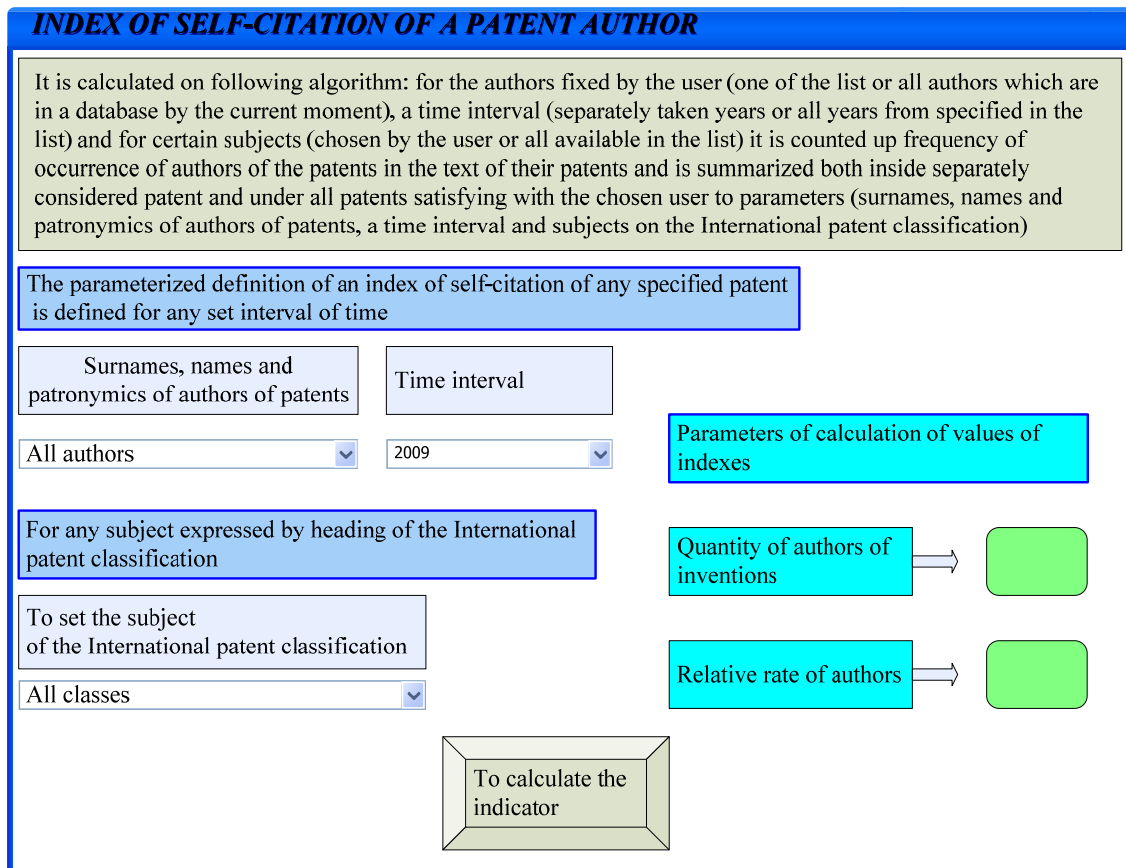
The dictionary was also designed to act as a tool of linguistic resources verification since indicators are ambiguous by their nature and hardly possess definitions. Thus the Semantic Dictionary is linked with regulations, information and algorithm components of the Evaluation system [4].

Besides, a special methodology of indicators’ meaning coordination was elaborated within the system. The process includes two stages [3].

At the first stage, the positioning of every approved (by experts) indicator within the overall classification scheme is carried out. It allows coordinating a preliminary insight of indicator’s meaning that can be extracted from the classification scheme of the dictionary.

At the second stage, the clarification of indicators’ meaning with the help of the dictionary entries linked with legislation, information and algorithm components of the Information Monitoring system is conducted.

The basic idea of the indicators meaning coordination lies in the development of the procedure of the indicators meaning clarification that considers the influence of five factors simultaneously: legislation, term, cognitive (explication of experts’ concepts and tracing for them), information, and authority [3].



**Figure 2. The Semantic Dictionary entry for the indicators “Index of self-citation of a patent author”**

To date the dictionary architecture including linkages to regulations, information, and algorithm components of the Evaluation system is being designed and developed. At the Figure 2 a sample of the Semantic Dictionary entry for the indicators "Index of self-citation of a patent author" is given. It's a test variant of the entry which comprises parameters of indicators and allows specifying their meanings and values according to clear and validated procedure. This provides controlled and consistent procedures of monitoring, analysis, and assessment within the Information Monitoring system and verified application of significant for experts indicators.

### **References**

1. Gilchrist A. Thesauri, taxonomies and ontologies an etymological note. *Journal of Documentation* 59(1), (2002), pp. 7-18.
2. Guarino N. (ed.), *Formal Ontology in Information Systems*. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, (1998), pp. 3-15.
3. Zatsman I., Kozhunova O. Evaluation system for the Russian Academy of Sciences: clarification tools. In: Abstracts of the International Conference ATLC'2007 «Atlanta Conference on Science, Technology and Innovation Policy 2007». Atlanta, USA, (2007), p. 28.
4. Zatsman I. M. & Kozhunova O. S. Evaluating for institutional academic activities: classification scheme for R&D indicators. Abstract book of the 10th International Conference on Science and Technology Indicators. Vienna: ARC GmbH, (2008), pp. 428-431.