JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(3): 21-30, 2019
https://doi.org/10.1633/JISTaP.2019.7.3.2

Operational Experience in DB "TERMIN"

Natalya N. Shaburova*

Rzhanov Institute of Semiconductor Physics, Siberian Branch
of the Russian Academy of Sciences, Novosibirsk, Russia
E-mail: shaburova@isp.nsc.ru

ABSTRACT

Information about the formation and filling (in 2014 to 2016) of a terminological dictionary on electronics and radioengineering and collective work (in 2017 to 2018) with a data bank "TERMIN" is presented in this article. In purpose of creating an instrument of navigating the modern scientific-technical space a net of terms with set semantic links is described. This set is based on the analysis of terms' definitions (each term is checked for inclusion in the definitions of all other terms; the definitions were borrowed from reputable reference editions: encyclopedias, dictionaries, reference books). The created model of a system that consists of different information sources, in which it (information) is indexed by the terminology of Russian State Rubricator of Scientific and Technical Information rubrics and/or keywords, is described. There is an access for the search in all these sources in the system. Searching inquiries are referred to in the language of these rubrics or formulated by arbitrary terms. The system is to refer to information sources and give out relevant information. In accordance with this model, semantic links of various types, which allow expanding a search at different modalities of query, should be set among data bank terms. Obtained links will have to increase semantic matching, i.e., they can provide actual understanding of the meaning of the information that is being sought.

Keywords: Universal Decimal Classification, State Rubricator of Scientific and Technical Information, Library-Bibliographic Classification, information retrieval, semantic links, electronics

Open Access

Received date: March 18, 2019
Accepted date: June 06, 2019

***Corresponding Author:** Natalya Nikolaevna Shaburova
Head of the Scientific Library
Rzhanov Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, Pr. Acad. Lavrentjeva, 13, Novosibirsk 630090, Russia
E-mail: shaburova@isp.nsc.ru

1. INTRODUCTION

The topicality of the task for a relevant search of scientific information is proved by the fact that the existing different kinds of search cannot be called sufficiently resultative (quality search). Some foreign approaches to solutions for this problem are mentioned in the introduction, but many of the arguments in this paper refer to the case of Russia.

The reasons for unsatisfactory searches are due to a mismatch between queries and documents. For instance, the “bag-of-words approach” can often be useless, as a high matching degree at this level does not necessarily mean high relevance. According to Furnas, Landauer, Gomez, and Dumais (1987), if one person assigns the name of an item, other untutored people will fail to access it in 80 to 90 percent of their attempts, and two people will often name the same concept with different representations. Contributions to the creation of semantic web, where the automated search is based on “free lexicon” as well, were actualized. Also, common searching machines (Yandex, Google) are based on the lexical search, and word matching still functions as their main mechanism.

As a way out, some tools to help us in forming a query are proposed, e.g., such a dictionary as one created by Bol’shakov and Gel’bukh (2011). It includes technical terms and basic concepts of science, humanities, business, and economics. Among its applications there is the possibility to form queries for Internet search engines. Besides this, there are some well-studied approaches: matching by query reformulation (Huang & Efthimiadis, 2009), matching with a term dependency model (Lee, Chen, Kao, & Cheng, 2009), matching with a translation model (Gao, He, & Nie, 2010), matching with a topic model (Barathi & Valli, 2014), and matching with a latent space model (Van Gysel, de Rijke, & Kanoulas, 2016). Unfortunately, they also cannot provide complete satisfaction for a search query (Li & Xu, 2013). The problem is that when searching with “free lexicon” or term-conjoined models, semantic links of concepts are not considered.

Semantic matching has been studied in the long history of information retrieval. Problems related to the so-called semantic search are discussed in a variety of different communities (Guha, McCool, & Miller, 2003; Makela, 2008; Bast, Buchhold, & Haussmann, 2016). So there are many approaches to the way of searching. Among them, according to the search algorithm of Bast, Baurle, Buchhold, and Haussmann (2013), the user interface guides the user in incrementally constructing queries by instant suggestions of definite objects and relations that lead to good hits in a semantic search. A nearly similar system is suggested by Zenz, Zhou, Minack, Siberski, & Nejdl (2009). Users start with a keyword query and then are guided through a process of incremental refinement steps to specify the query intention.

Another approach is when users do not need to do anything different from a conventional search, and semantic matching is carried out inside the search engine. In this case the possibility of semantic links should be put in advance, including representation knowledge in a way suitable for meaningful retrieval. One of the proposals is called a concept search (Giunchiglia, Kharkevich, & Zaihrayeu, 2009), which extends syntactic search based on the computation of string similarity between words, with semantic search, i.e., search based on the computation of semantic relations between concepts. The key idea of concept search is to operate on complex concepts and to maximally exploit the semantic information available, reducing to syntactic search only when necessary, i.e., when no semantic information is available.

Semantic links, set in classifications and thesauri, allow selecting documents containing the required information during a search, but to the extent that it is reflected in a document search image. It should be taken into account that the vast majority of bibliographical descriptions of documents, made in previous years in library cataloguing and constituting most retrospective documentary scientific information arrays, have only library classification indexes as document search image elements revealing their contents. Thereat, obtaining relevant information is complicated, because information is distributed in the space of dissociated scientific sources, not systematized uniformly. At present, in Russia there is no unity of means to describe a subject (thematics). Part of the sources is systematized on the Universal Decimal Classification (UDC) schemes, and part of them on those of the Library-Bibliographic Classification (LBC). There is a patent classification for patent searches. Searches for materials on physics published abroad should be assisted by the PACS (Physics and Astronomy Classification Scheme) scheme, etc. It is not at random that different variants of integrating information resources on their base, e.g. the creation of queries dispatching service based on the system of semantic correspondences between the elements of used thematic (subject) classifications or working out information retrieval thesauri based on the lexicon of classification systems, are proposed by Beloozerov and Shaburova (2015a).

The development of linguistic means for searching scientific information is directly connected with the development of information technologies, full-text, bibliographical, and factographic databases. In recent studies, ways of expansion of the search queries by means of developing classifications-based systems are considered, for instance: the Dewey Decimal

Classification (DDC) used in different libraries, online tools as SciGator (Lardera, Gnoli, Rolandi, & Trzmielewski, 2017), aggregating metadata and automatic classifying mapping (Lin et al., 2017), or generating tags on the base of DDC-extracted terms to support interoperability diverse digital repositories into a unified information infrastructure (Khoo et al., 2012). These tools could give the possibility to expand searches to national and international catalogues or be useful for digital libraries. In any way they concern browsing and search of related concepts.

A positive aspect is that modern scientific publications have not only their obligatory indexes of these or those classifications, but they are additionally accompanied by the keywords assigned to them by authors or editors.

Information retrieval thesauri, actively developed for different fields of science and engineering, became one of the information retrieval languages at the end of the twentieth and beginning of the twenty-first centuries. A number of modern contributions, both in Russia and abroad (Antopol'skii & Markarova, 2017; BARTOC.org, n.d.; Dobrov, Ivanov, Lukashevich, & Solov'ev, 2009; Gendina, 2015; YARN, n.d.), are devoted to information retrieval thesauri-related investigations. In our country they are developed according to a special Russian state standard (GOST 7.25, 2002), but in practice, thesauri are currently little used, although they are a developed semantic tool for information retrieval.

In China for a long period the Chinese Classified Thesaurus has been widely used. It is a knowledge organization tool in libraries as well as being an integrated structure of the Chinese Library Classification and the Chinese Thesaurus. However, its complicated knowledge structure and relation maps within are implicit to end-users (Wei, Shuqing, & Qing, 2013), and now attempts at adapting the existing system to the semantic web environment are being made.

The ontological representation can be viewed as a perspective for the semantic integration of information sources and adequate interpretation of a content of textual documents. In modern understanding, ontology is a way of representing the conceptual system of a subject field that generalizes thesaurus, classification, and linguistic approaches. The ontological approach consists of building a terminological system of concepts connected by heterogeneous links realizing a terms classification on different categories and bases for comparison. The theory, practice, and methodology of building up ontologies and the ontological approach to the development of linguistic provision of information retrieval are described in modern literature in detail, particularly in the following resources (About the linguistic ontology "Thesaurus RuThes", n.d.; Gomez-Pere, Fernando-Lopez, & Corcho, 2004; W3C, 2004; Rubashkin, 2013; Denny, 2004).

The task of a modern information retrieval developmental stage consists in the necessity for integrating several types of linguistic provision means (library and information classifications being predominant in bibliographical databases, and thesauri that have a most developed system of semantic relations) along with the "free lexicon" mainly used in queries of information consumers.

In worldwide practice, such an approach was realized, particularly in the U.S. National Medical library. The Unified Medical Language System (UMLS) includes the Metathesaurus, the Semantic Network, and the Specialist Lexicon and Lexical Tools. These tools have been developed to investigate the contributions that natural language processing techniques can make to the task of mediating between the language of users and the language of online biomedical information resources. The Metathesaurus is the biggest component of the UMLS. It is a large biomedical thesaurus that is organized by concept or meaning, and it links similar names for the same concept from nearly 200 different vocabularies. The Metathesaurus also identifies useful relationships between concepts and preserves the meanings, concept names, and relationships from each vocabulary (U. S. National Library of Medicine, n.d.).

The COLI-CONC project of creating net program modules, integrating different knowledge-organizing systems (ontologies, thesauri, and classifiers) for their effective control and practical use, was started in Germany (COLI-CONC, n.d.).

Thus, a unified electronic knowledge space should be built on a logical and linguistic basis, and the ontology should be based on the vocabulary and paradigm of information languages practically used within this space in Russia (Antoshkova, Beloozerov, Dmitrieva, & Shapkin, 2017).

One of the best solutions for developing a universal instrument of thematic retrieval and navigating the modern scientific-technical information space was supported by the Russian Foundation of Basic Researches (RFBR) grant 17-03-12013-SHS "Interactive system of creating and supporting the ontology of scientific knowledge based on a dynamic complex of terminological dictionaries." The participants of the project are the Library of Natural Sciences, Russian Academy of Sciences (LNS RAS); All-Russian Institute of Scientific and Technical Information, Russian Academy of Sciences (ARISTI RAS); Institute of Scientific Information on Social Sciences, Russian Academy of Sciences (ISISS RAS); and Rzhanov Institute of Semiconductor Physics, Siberian Branch, Russian Academy of Sciences (ISP SB RAS). The head of the grant is professor Nickolai Yevgenyevich Kalenov.

The project objective is working out a model version of an interactive ontological system, including various subject areas, which is aimed to provide an increase of information retrieval efficiency (completeness and accuracy) in integrated scientific resources. The originality of the approach, implemented during the project, is that an adequate representation of ontology is a set of information retrieval thesauri built with the help of terminological dictionaries. The project novelty is conditioned by the fact that the terminology from metadata of information resources, including keywords, rubrics, and class names of library-bibliographic and information classifications, is used as descriptors. Regular thesaurus links are established between descriptors based on the semantic intersection of the volume and/or content of the concepts expressed by the descriptors. Working with metadata vocabulary (and not with full texts) largely solves the problem of an excess of information to be analyzed.

The experimental approbation of the model is carried out in 6 thematic fields on real heterogeneous information for the future realization of the aim. To prepare a full-fledged platform for testing, laborious work was carried out for 3 years to collect the material, and then 2 years to process it. This applied activity is described in the present paper.

Terminological dictionaries on all branches of the Russian national economy were preliminarily compiled during 2014 to 2016 (Beloozerov, 2015; Kalenov & Beloozerov, 2015). Their archive is saved in the zip format, accessible at http://systemling.narod.ru/terminolog/slovari_Minobrazin-2016-01-20.zip. Each of the compiled dictionaries corresponds to one of the State Rubricator of Scientific and Technical Information (SRSTI) sections in its thematics (code SRSTI for "Electronics. Radioengineering" is 47) and includes the terminology of its subordinate rubrics (headings), and the main keywords used for a thematic indexation of scientific documents, definitions of terms, and direct hyperlinks to the reference publications that correspond to them. The ISP library staffers took part in the filling of the dictionary "47 Electronics. Radioengineering." The thematics covers the ISP direction of investigations and developments.

Later, in 2017, the materials of these dictionaries served as a base for the data bank (DB) "TERMIN" formation, the aim of which was to make the dictionaries a ground for building up a net of semantically linked terms as a model of information space ontology for scientific and technical information. The structure of such a net was to be based on the semantic correspondences between descriptions (definitions), and later this network was enriched with vocabulary and semantic relations of previously developed information retrieval thesauri. A detailed description of it by the team of authors is in Antopol'skii, Beloozerov, Kalenov, Shaburova, and Yakshin (2017).

At the stage of the current preliminary experiment, a limited number of relations—"Enters," "Contains," "Equivalent," "Crosses," and "Weak link"—are used between semantic links. Semantic relations of previously developed information retrieval thesauri are included in the model. Thus the terminology of the dictionary allowed establishing 5 associative kinds of definitive links for expanding the possibilities of searching scientific information, and the query document mismatch challenge can be conquered.

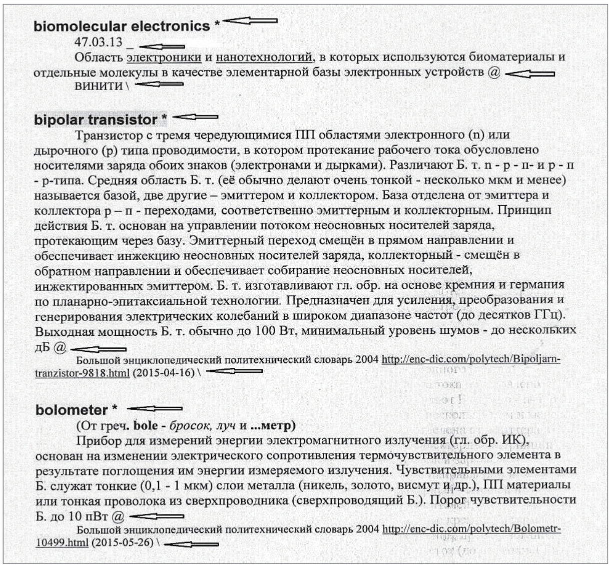
2. DESCRIPTION OF THE WORK

2.1. Building Up the Dictionary

The dictionary 47 Electronics. Radioengineering, which primarily consisted of 450 terms, was filled by the ISP library staffers. The procedure appeared as follows. First, a review of SRSTI codes was made. This is the National Hierarchical Classification System. It embraces the whole 'universum' of knowledge and it agrees to the traditional structure of the National Economy of Russia. The SRSTI is obligatory for indicating the thematic content of all scientific materials in automated systems. The Rubricator is being developed and modernized, and it is freely accessible at http://grnti.ru. Recently, it has been proposed, for example, as a base for comparing rubrics of other bibliographical classifications and determining their correspondences (Beloozerov & Shaburova, 2016).

The semantic terms of the second and third hierarchies, with which further work was carried out, were extracted for the composition of the vocabulary from the SRSTI lexicon. The rest of the dictionary was filled up with keywords that reflect the thematic content, being not the lexical units of the Rubricator, but important for information retrieval. The algorithm for work with them had to be realized in the other sequence: A word or a phrase, to be ciphered using SRSTI and UDC codes, was entered. It is not unimportant that most part of the terms of dictionary 47 Electronics. Radioengineering is accompanied by UDC codes. The hierarchical structure of classifications served as an additional instrument in further work with the DB on establishing links between terms.

Further work with the terms and keywords consisted of searching a significant and valuable interpretation of their essence (semantics). Both print encyclopedias, dictionaries, reference books, and electronic sources were used. The latter were considered to be more preferable due to their ergonomics and time savings. Specialized reference sources, such as *Encyclopedia of physics and engineering* (http://www.femto.com.



biomolecular electronics *

47.03.13 _

Область электроники и нанотехнологий, в которых используются биоматериалы и отдельные молекулы в качестве элементарной базы электронных устройств @

ВИНИТИ \

bipolar transistor *

Транзистор с тремя чередующимися ПП областями электронного (n) или дырочного (р) типа проводимости, в котором протекание рабочего тока обусловлено носителями заряда обоих знаков (электронами и дырками). Различают Б. т. n - р - п- и р - п - р-типа. Средняя область Б. т. (её обычно делают очень тонкой - несколько мкм и менее) называется базой, две другие – эмиттером и коллектором. База отделена от эмиттера и коллектора р – п - переходами, соответственно эмиттерным и коллекторным. Принцип действия Б. т. основан на управлении потоком неосновных носителей заряда, протекающим через базу. Эмиттерный переход смещён в прямом направлении и обеспечивает инжекцию неосновных носителей заряда, коллекторный - смещён в обратном направлении и обеспечивает собирание неосновных носителей, инжектированных эмиттером. Б. т. изготавливают гл. обр. на основе кремния и германия по планарно-эпитаксиальной технологии. Предназначен для усиления, преобразования и генерирования электрических колебаний в широком диапазоне частот (до десятков ГГц). Выходная мощность Б. т. обычно до 100 Вт, минимальный уровень шумов - до нескольких дБ @

Большой энциклопедический политехнический словарь 2004 http://enc-dic.com/polytech/Bipoljarn-tranzistor-9818.html (2015-04-16) \

bolometer *

(От греч**. bole -** *бросок, луч* и **...метр)**

Прибор для измерений энергии электромагнитного излучения (гл. обр. ИК), основан на изменении электрического сопротивления термочувствительного элемента в результате поглощения им энергии измеряемого излучения. Чувствительными элементами Б. служат тонкие (0,1 - 1 мкм) слои металла (никель, золото, висмут и др.), ПП материалы или тонкая проволока из сверхпроводника (сверхпроводящий Б.). Порог чувствительности Б. до 10 пВт @

Большой энциклопедический политехнический словарь 2004 http://enc-dic.com/polytech/Bolometr-10499.html (2015-05-26) \

Fig. 1. Fragment of dictionary "Electronics. Radioengineering."

ua), *Engineering encyclopedia* (http://enciklopediya-tehniki.ru/tehnika), and *Big encyclopedic polytechnical dictionary* (http://www.endic.ru/polytech/default.htm), turned out to be more useful for the filling of our dictionary. A part of the terms was proposed by ARISTI RAS. This part was promptly accompanied by definitions and classification codes (cipherings), but its quantity was relatively innumerous.

So, the dictionary 47 Electronics. Radioengineering contains entries that consist of a term of the SRSTI lexicon or a keyword, its definition, and the name of the information source. The terms are headings of subordinate rubrics of SRSTI Section 47. Each keyword can be viewed as a specific heading that contains all the documents that are indexed by this keyword.

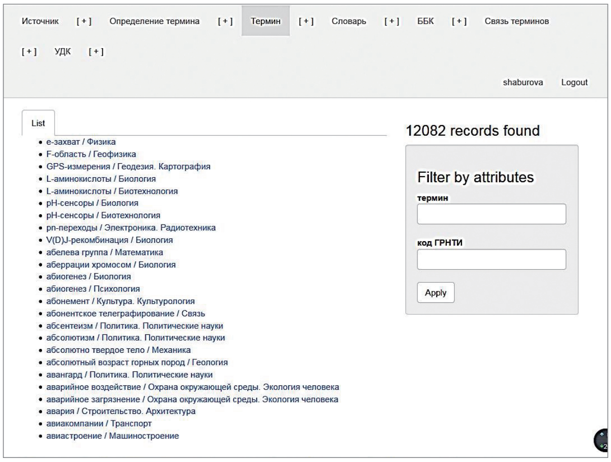
The compiled dictionaries had to be made into an instrument capable of giving out a valuable relevant search result. To transform this structure into a net of semantically linked terms, an automated analysis was, certainly, planned. However, a preliminary and only manually possible preparation was necessary for machine processing. To help it after each element of entries, marking signs *, _, @, \ were placed (arrow-marked in Fig. 1). This part of the work in the dictionary 47 Electronics. Radioengineering was carried out completely in ISP SB RAS.

2.2. Building the DB

DB TERMIN was realized on the Scirus platform developed in LNS RAS (Yakshin, 2015) and is accessible at http://class.labs.benran.ru. Further work with the BD was in two stages. In Fig. 2, one can see the DB interface (further on, in figures there will be screenshots of the TERMIN system pages). In the upper panel, there are sections "Source," "Definition," "Term," and "Dictionary." Each of them provides information as a list of the corresponding names—each of which can be entered and one can continue a necessary work direction—and indications of their quantities. These are 69 dictionaries, more than 11,000 terms, over 12,000 definitions, and more than 7,500 of their sources. Attention should be paid (Fig. 2) so that each term is accompanied by an SRSTI thematic section name of the dictionary.

At the first stage, which was being realized during 2017, a manual modification, regarding a check of whether all terms of the initial dictionaries are included in the DB, and a deciphering of the abbreviations of all notional words taken from reference sources in the definitions for their accurate machine readout, were required. In addition it was necessary to mark some indications with "conventional signs," taking into account that only a person can understand differences in such cases as the following, for example: disposition "x0" may denote temperature, geographical co-ordinates, or degree, when the main numbers should have been in some special way separated from the powers they were raised to.

At the beginning of the second stage in the work with the DB TERMIN (end of 2017 to 2018), an automated analysis was carried out and establishing proper links between the terms and keywords through their definitions was achieved. There appeared a section "Terms link" in the DB, and it was a list of about 320,000 units. As the machine processing algorithm consisted of identifying similar words in texts without determining their semantic role, there appeared a big number of false semantic links, and the following intellectual modification was required. It consisted of viewing each formally connected pair of words and deleting false links. The primary number



12082 records found

термин

код ГРНТИ

- е-захват / Физика
- F-область / Геофизика
- GPS-измерения / Геодезия. Картография
- L-аминокислоты / Биология
- L-аминокислоты / Биотехнология
- pH-сенсоры / Биология
- pH-сенсоры / Биотехнология
- pn-переходы / Электроника. Радиотехника
- V(D)J-рекомбинация / Биология
- абелева группа / Математика
- аберрации хромосом / Биология
- абиогенез / Биология
- абиогенез / Психология
- абонемент / Культура. Культурология
- абонентское телеграфирование / Связь
- абсентеизм / Политика. Политические науки
- абсолютизм / Политика. Политические науки
- абсолютно твердое тело / Механика
- абсолютный возраст горных пород / Геология
- авангард / Политика. Политические науки
- аварийное воздействие / Охрана окружающей среды. Экология человека
- аварийное загрязнение / Охрана окружающей среды. Экология человека
- авария / Строительство. Архитектура
- авиакомпании / Транспорт
- авиастроение / Машиностроение

Fig. 2. List of terms with the indication of thematic section names.

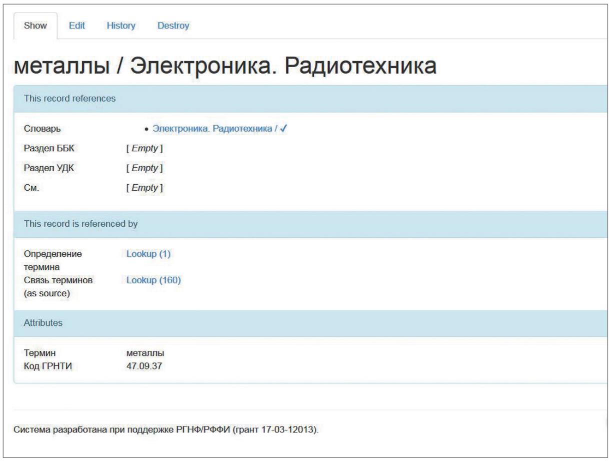
of links showed that it would be unreal to embrace all BD dictionaries with manual modification in due time. Therefore, to verify the experimental model the intellectual analysis of thematic directions was limited to several. Six BD dictionaries were chosen to check techniques based on materials of social, natural, and applied sciences. These are economics, linguistics, informatics, physics, electronics and radioengineering, and standardization, an executor being responsible for each scientific field. The SRSTI section 47 "Electronics. Radioengineering" is the initial dictionary in the compilation of which we participated in 2014 and are still modifying it now in the DB.

In the process of filling the BD dictionary "47 Electronics. Radioengineering" with the vocabulary of the UDC and LBC, the number of terms is constantly increasing. At present, there are 488 terms in the DB; this informative stage is described by the project participants (Antopol'skii, Beloozerov, Kalenov, & Markarova, 2018). Thereat, there was a very big number of links: e.g., just one term "metal" had 342 links. It is worth noting that definitive connections include two groups: 1) "Terms link" (as a source) consists of the terms in the definitions of which a term under consideration is come across (Fig. 3). If we enter the term "Definition," then one can see the second group "Terms link (as a target)." This button shows the terms found out in defining the given term. The corresponding definition appears on the same page, where the corresponding link is. As for the links of the term "Metals/Electronics. Radioengineering," there were 332 of the first group and 10 of the second one. Now, after verification of the links, there are almost twice as few of them left, i.e., 170 (160 links "as a source" and 10—"as a target"), as the links with terms of such thematic sections as "Art," "Light industry," "Housing-communal services" and, partially, "Metallurgy" were deleted.

Another example of a link to be deleted is the following. Among the automatically established links of the terms, there was such a pair as "character/Psychology" and "collective acceleration methods/Electronics. Radioengineering." The program established the link based on the coincidence of the term with a word from the text of the second term definition: "depending on the interaction character, the way and modifications of particles collective acceleration are differentiated." Accordingly, this pair was to be deleted manually. Function "Destroy" is meant for such actions. Besides this, rubrics "Show," "Edit," and "History" are placed in section "Terms link" (Fig. 3, at the top). The latter orders the responsibility for the work done and shows which of the project participants made corrections and when. It is worth paying attention to the fact that, on the page, there is an indication to the supported grant number as it is required in Russia (Fig. 3, at the bottom). After that, automatically established links were supplemented by vocabulary and relationships borrowed from previously developed information retrieval thesauri, in particular on electronics (semiconductors and nanotechnologies) (Beloozerov & Shaburova, 2015b).

The last thing being presently fulfilled within RFBR grant 17-03-12013-SHS is determining the kinds of relations among established links, i.e. hierarchy, equivalence, or association. Function "Edit" is meant to realize the procedure in section "Terms link." When entering it, an executor gets a possibility of fixing the terms' semantic relations. As it was formulated above, the intellectual activities in this field were partly based on the structure of classifications *a priori* establishing the hierarchical relations between terms.

The system offers to choose the kind of links from "Enters," "Contains," "Equivalent," or "Crosses." In the first two cases, the terms correlate as species/genus. "Enters" links the term of



металлы / Электроника. Радиотехника

This record references

Словарь • Электроника. Радиотехника / ✓
Раздел ББК [*Empty*]
Раздел УДК [*Empty*]
См. [*Empty*]

This record is referenced by

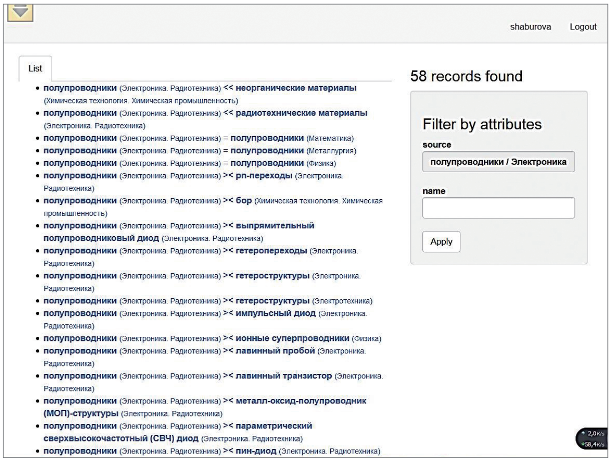
Определение термина Lookup (1)
Связь терминов (as source) Lookup (160)

Attributes

Термин металлы
Код ГРНТИ 47.09.37

Система разработана при поддержке РГНФ/РФФИ (грант 17-03-12013).

Fig. 3. Number of links for the term "metals."



- **полупроводники** (Электроника. Радиотехника) << **неорганические материалы** (Химическая технология. Химическая промышленность)
- **полупроводники** (Электроника. Радиотехника) << **радиотехнические материалы** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) = **полупроводники** (Математика)
- **полупроводники** (Электроника. Радиотехника) = **полупроводники** (Металлургия)
- **полупроводники** (Электроника. Радиотехника) = **полупроводники** (Физика)
- **полупроводники** (Электроника. Радиотехника) >< **pn-переходы** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **бор** (Химическая технология. Химическая промышленность)
- **полупроводники** (Электроника. Радиотехника) >< **выпрямительный полупроводниковый диод** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **гетеропереходы** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **гетероструктуры** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **гетероструктуры** (Электротехника)
- **полупроводники** (Электроника. Радиотехника) >< **импульсный диод** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **ионные суперпроводники** (Физика)
- **полупроводники** (Электроника. Радиотехника) >< **лавинный пробой** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **лавинный транзистор** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **металл-оксид-полупроводник (МОП)-структуры** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **параметрический сверхвысокочастотный (СВЧ) диод** (Электроника. Радиотехника)
- **полупроводники** (Электроника. Радиотехника) >< **пин-диод** (Электроника. Радиотехника)

58 records found

Filter by attributes

source: полупроводники / Электроника

name:

Fig. 4. Indications of the kinds of links between terms.

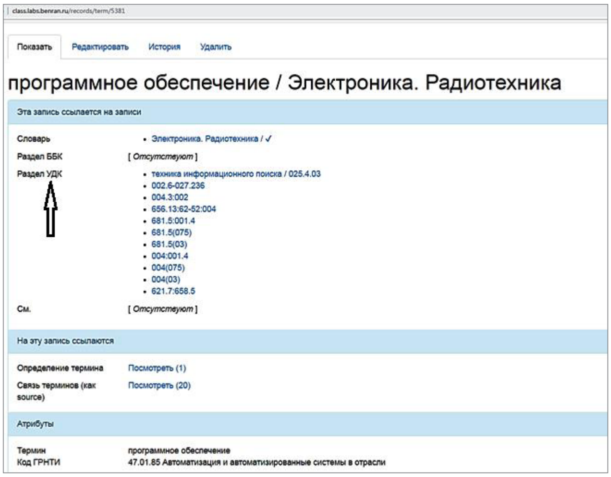
species, subclass, being on the left side, with the term of genus, superclass, on the right side in the terms pair record (Fig. 4). "Contains," *vice versa*, links the generic term on the left with the specific one on the right. In cases when links of terms connect them from different fields of knowledge, one should understand that the scopes (volumes) of the concepts for the same terms in different fields may differ, and the genus-species relations established for one field may change for the "cross" relations in the other one. "Equivalence" means we believe that the same relevant documents are meant by these keywords. "Crosses" means that considerably crossing arrays of indexed documents are meant by this pair of keywords, and, in many cases, it will be useful to widen the inquiry on this link.

Four examples:

1. "Heptode/Electronics. Radioengineering" and "mixer tube/Electronics. Radioengineering": the terms correlate as genus/species, and the first term "enters" the second one.
2. "Aerial/Electronics. Radioengineering" and "horn-reflector aerial/Electronics. Radioengineering": the first concept is wider and it "contains" the second one.
3. "Electric capacitor/Electronics. Radioengineering" and "electric capacitor/Electrical Engineering" are the equivalent concepts.
4. "Remote probing/Electronics. Radioengineering" and "measurement/Geophysics": the concepts cross each other.

In parallel, the links were rendered signs << ("enters"), >> ("contains"), = ("equivalent"), and >< ("crosses"). When entering the rubric "Terms link," it provided the visualization of the kinds of definitive links (Fig. 4).

The term links were left uncategorized and were not eliminated in cases when there was a real semantic link—which, however, is not sufficiently strong to be used for widening a query during a search (i.e., arrays of documents that underlie the terms are crossed weakly)—between the terms connected by an automatically set definitive link. They were left in the dictionary as a special kind of "weak" links as material for further processing.



программное обеспечение / Электроника. Радиотехника

Эта запись ссылается на записи

Словарь
- Электроника. Радиотехника / ✓

Раздел ББК [*Отсутствуют*]

Раздел УДК
- техника информационного поиска / 025.4.03
- 002.6-027.236
- 004.3:002
- 656.13:62-52:004
- 681.5:001.4
- 681.5(075)
- 681.5(03)
- 004:001.4
- 004(075)
- 004(03)
- 621.7:658.5

См. [*Отсутствуют*]

На эту запись ссылаются

Определение термина Посмотреть (1)

Связь терминов (как source) Посмотреть (20)

Атрибуты

Термин программное обеспечение
Код ГРНТИ 47.01.85 Автоматизация и автоматизированные системы в отрасли

Fig. 5. Universal Decimal Classification codes for term "software/Electronics. Radioengineering."

Thus, there are five kinds of links set in the dictionary "Electronics. Radioengineering" DB TERMIN LNS RAS as shown in Table 1.

The final action of the grant participants in this stage of development is rendering the codes of specialized (narrow) thematics on SRSTI to each keyword. So, as most parts of the terms of dictionary "47 Electronics. Radioengineering" are accompanied by UDC codes (in Fig. 5, marked by an arrow), the semantic network of keywords becomes combined with the system of correspondence between the headings of bibliographic classifiers, in order to construct a model of metadata connections for thematic searches and navigation in the information space of scientific knowledge.

The intermediate results of the carried-out activities and theoretical groundings, besides the report to RFBR, are formulated in Antopol'skii, Beloozerov, and Markarova (2017) and Antoshkova, Beloozerov, Dmitrieva, Smirnova, et al. (2017).

Table 1. Semantic links of compared terms

Sign	Names	Meaning
A=B	coincides, equal, identical	Terms A and B indicated a possibility of searching identical documents (synonyms).
A>>B	more, wider, includes	Term A indicates a possibility of searching the documents in the set of which the terms designated by term B are included.
A<<B	less, already enters	Term A indicates a possibility of searching the documents included in the set of documents designated by term B.
A><B	crosses	Sets of documents, indexed by terms A and B, cross in a considerable part.
A–B	definitively linked	Documents designated by terms A and B are linked by pragmatic connections, but their sets cross weakly, belonging to different ontological categories.

3. CONCLUSION

In conclusion, it is possible to wind up the following: The dictionaries that contain terms, SRSTI, UDC and partly LBC codes, term definitions, and indications to definition sources have been compiled. The DB TERMIN has been created. Names of thematic sections have been added to the enumerated parameters. The automated processing of the data array, entered in the DB on the material of terminological dictionaries, provided its full coverage with a guaranteed result as a system of semantically linked scientific terms, but, alongside with this, it required large preparatory manual work and even more final intellectual verification and refinement. The semantic links have been established between terms based on analysis of their definitions, and relations between the links.

The system of five kinds of terms' semantic links has been realized in the dictionary "47 Electronics. Radioengineering." The associative kind of semantic links "cross" has the biggest number of links, and "equivalence" the smallest one. Part of the term links have not any concretized relations as their semantic links clearly exist, but they do not match the meanings of the used relations. These links correspond to a minor cross of the meanings of terms, but they can be the material for a further development of the system at introducing "pragmatic" relations in it, such as raw material-product, event-place-time, cause-consequence, etc.

As a result, the integration of several types of linguistic provision means (UDC, SRSTI, partly LBC, thesauri with semantic relations, and possibility of formulation queries by arbitrary terms) is carried out, and the framework of an exhaustive model of metadata connections (ontology) for thematic retrieval with relevant semantic matching and more coherent and easy navigation in the information space of scientific knowledge is realized. Practically, the attempt to integrate different classifications and descriptor languages for a polyfunctional information retrieval within an ontological approach has been first made. In whole, the results obtained are consistent with the global level, and the method of forming links between the terms has no analogues known to the participants of the project.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation of Basic Researches (17-03-12013-SHS).

The author would like to express her gratitude to all participants in the project for their coordinated work and provided results obtained during its joint implementation. Also, the author would like to express her gratitude to Alexander V. Zhuravlev, ISP SB RAS executive interpreter, for the English version of the present article.

REFERENCES

About the linguistic ontology "Thesaurus RuThes". (n.d.). Retrieved April 10, 2019 from http://www.labinform.ru/pub/ruthes/index.htm.

Antopol'skii, A. B., Beloozerov, V. N., Kalenov, N. Y., & Markarova, T. S. (2018). On the development of terminological database as a complex of sectorial information retrieval thesauri. *Information Resources of Russia*, 5, 22-30.

Antopol'skii, A. B., Beloozerov, V. N., Kalenov, N. E., Shaburova, N. N., & Yakshin, M. M. (2017). The development of a semantic network of keywords based on definitive relationships. *Scientific and Technical Information Processing*, 44(4), 261-265.

Antopol'skii, A. B., Beloozerov, V. N., & Markarova, T. S. (2017). On the development of the ontology based on the classifiers of scientific information and terminological dictionaries. *Information Resources of Russia*, 5, 2-7.

Antopol'skii, A. B., & Markarova, T. S. (2017). *Information languages in the twenty-first century*. Retrieved April 10, 2019 from http://www.systemling.narod.ru/informat/bibliografiya-ontologij.docx.

Antoshkova, O. A., Beloozerov, V. N., Dmitrieva, E. Y., & Shapkin, A. V. (2017). Development of ontology of STI on the basis of bibliographic classifications. In N. Y. Kalenov & V. A. Tsvetkova (Eds.), *Information provision of science, new technologies* (pp. 292-300). Moscow: LNS RAS.

Antoshkova, O. A., Beloozerov, V. N., Dmitrieva, E. Y., Smirnova, O. V., Shapkin, A. V., & Shaburova, N. N. (2017). On a method for constructing an ontology of scientific and technical information as a network of bibliographic classifications. *Scientific and Technical Information Processing*, 44(4), 266-272.

Barathi, M., & Valli, S. (2014). Topic based query suggestion using hidden topic model for effective web search. *Journal of Theoretical and Applied Information Technology*, 59(3), 632-642.

BARTOC.org. (n.d.). *About*. Retrieved April 10, 2019 from https://bartoc.org/en/content/about.

Bast, H., Baurle, F., Buchhold, B., & Haussmann, E. (2013). *Broccoli: Semantic full-text search at your fingertips*. Retrieved April 10, 2019 from https://arxiv.org/

abs/1207.2615.

Bast, H., Buchhold, B., & Haussmann, E. (2016). Semantic search on text and knowledge bases. *Foundations and Trends in Information Retrieval*, 10(2-3), 119-271.

Beloozerov, V. N. (2015). Technology of building up terminological dictionaries using lexicon of classification systems. In N. Y. Kalenov & V. A. Tsvetkova (Eds.), *Information provision of science, new technologies* (pp. 126-136). Moscow: LNS RAS.

Beloozerov, V. N., & Shaburova, N. N. (2015a). Comparison of bibliographic classifications for semiconductors and nanotechnologies in thesaurus format. *Scientific and Technical Information Processing*, 3, 47-51.

Beloozerov, V. N., & Shaburova, N. N. (2015b). Thesaurus of thematic rubrics in semiconductor physics as a model of united classificational-vocabulary system/ Comparison of LBC and UDC with the help of thematic meanings of "47 Electronics. Radioengineering" SRSTI rubrics. In N. Y. Kalenov & V. A. Tsvetkova (Eds.), *Information provision of science, new technologies* (pp. 147-152). Moscow: LNS RAS.

Beloozerov, V. N., & Shaburova, N. N. (2016). Comparison of LBC and UDC with the help of thematic meanings of "47 Electronics. Radioengineering" SRSTI rubrics. In P. P. Treskova & O. A. Oganova (Eds.), *Information provision of science, new technologies* (pp. 175-187). Ekaterinburg: CSL UB RAS.

Bol'shakov, I. A., & Gel'bukh, A. F. (2011). *A large electronic dictionary as a polythematic guide and shaper of queries to the Web*. Retrieved April 10, 2019 from http://www.dialog-21.ru/media/1414/14.pdf.

COLI-CONC. (n.d.). *coli-conc: Infrastructure to facilitate management and exchange of concordances between library knowledge organization systems*. Retrieved April 10, 2019 from https://coli-conc.gbv.de/.

Denny, M. (2004). *Ontology tools survey, revisited*. Retrieved April 10, 2019 from https://www.xml.com/pub/a/2004/07/14/onto.html.

Dobrov, B. V., Ivanov, V. V., Lukashevich, N. V., & Solov'ev, V. D. (2009). *Ontologies and thesauri: Models, instruments, applications*. Moscow: BINOM publ.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11), 964-971.

Gao, J., He, X., & Nie, J.-Y. (2010). *Clickthrough-based translation models for web search: From word models to phrase models*. Retrieved April 10, 2019 from https://core.ac.uk/download/pdf/23798617.pdf.

Gendina, N. I. (2015). *Information retrieval thesauri: Structure, purpose and order of development*. Retrieved April 10, 2019 from https://nsu.ru/xmlui/bitstream/handle/nsu/8962/IPT.pdf.

Giunchiglia, F., Kharkevich, U., & Zaihrayeu, I. (2009). Concept search. In *The Semantic Web: Research and Applications. ESWC 2009* (pp. 429-444). Berlin: Springer-Verlag.

Gomez-Pere, A., Fernando-Lopez, M., & Corcho, O. (2004). *Ontological engineering*. Berlin: Springer-Verlag.

GOST 7.25 (2002). *System of standards on information, library science and publishing: Information retrieval monolingual thesaurus*. Moscow: MTC 191.

Guha, R., McCool, R., & Miller, E. (2003). *Semantic search*. Retrieved April 10, 2019 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.774&rep=rep1&type=pdf.

Huang, J., & Efthimiadis, E. N. (2009). *Analyzing and evaluation query reformulation strategies in web-search logs*. Retrieved April 10, 2019 from https://jeffhuang.com/Final_Reformulation_CIKM09.pdf.

Kalenov, N. Y., & Beloozerov, V. N. (2015). Building up terminological dictionaries based on the vocabulary of classification systems. *Scientific and Technical Information Processing*, 3, 60-70.

Khoo, M., Tudhope, D., Binding, C., Abels, E., Lin, X., & Massam, D. (2012). Towards digital repository interoperability: The Document Indexing and Semantic Tagging Interface for Libraries (DISTIL). In P. Zaphiris, G. Buchanan, E. Rasmussen, & F. Loizides (Eds.), *International conference on theory and practice of digital libraries* (pp. 439-444). Berlin: Springer-Verlag.

Lardera, M., Gnoli, G., Rolandi, C., & Trzmielewski, M. (2017). Developing SciGator, a DDC-based library browsing tool. *Knowledge Organization*, 44(8), 638-643.

Lee, C.-J., Chen, R.-C., Kao, S.-H., & Cheng, P.-J. (2009). *A term dependency-based approach for query terms ranking*. Retrieved April 10, 2019 from http://rueycheng.com/paper/term-dependency.pdf.

Li, H., & Xu, J. (2013). Semantic matching in search. *Foundations and Trends in Information Retrieval*, 7(5), 343-469.

Lin, X., Khoo, M., Ahn, J.-W., Tudhope, D., Binding, C., Massam, D., & Jones, H. (2017). Mapping metadata to DDC classification structures for searching and browsing. *International Journal on Digital Libraries*, 18(1), 25-39.

Makela, E. (2008). *Survey of semantic search research*. Retrieved April 10, 2019 from https://www.researchgate.net/

publication/225070162_Survey_of_Semantic_Search_Research.

W3C. (2004). *OWL web ontology language guide: W3C recommendation*. Retrieved April 10, 2019 from https://www.w3.org/TR/2004/REC-owl-guide-20040210/.

Rubashkin, V. S. (2013). *Ontological semantics*. Moscow: Fizmatlit.

YARN. (n.d.). *Thesaurus of Russian language: Yet another RussNet*. Retrieved April 10, 2019 from https://russianword.net/.

U. S. National Library of Medicine. (n.d.). *Unified Medical Language System* (UMLS). Retrieved April 10, 2019 from https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/.

Van Gysel, C., de Rijke, M., & Kanoulas, E. (2016). *Learning latent vector spaces for product search*. Retrieved April 10, 2019 from https://arxiv.org/pdf/1608.07253.pdf.

Wei, F., Shuqing, B., & Qing, Z. (2013). Semantic visualization for subject authority data of Chinese classified thesaurus. In A. Slavic, A. Akdag Salah, & S. Davies (Eds.), *Classification and visualization: Interfaces to knowledge* (pp. 191-206). Wurzburg: Ergon Verlag.

Yakshin, M. M. (2015). Development of the Scirus platform. In N. Y. Kalenov & V. A. Tsvetkova (Eds.), *Information provision of science, new technologies* (pp. 203-207). Moscow: LNS RAS.

Zenz, G., Zhou, X., Minack, E., Siberski, W., & Nejdl, W. (2009). From keywords to semantic queries: Incremental query construction on the semantic web. *Journal of Web Semantics*, 7(3), 166-176.