

DOI [10.24069/konf-23-26-04-2019.08](https://doi.org/10.24069/konf-23-26-04-2019.08)



Программная оболочка системы терминологических словарей

Н. Е. Каленов^{1,2}, А. М. Сенько¹

¹ Библиотека по естественным наукам Российской академии наук, г. Москва, Россия

² Межведомственный суперкомпьютерный центр Российской академии наук, г. Москва, Россия

Резюме: В статье рассматривается интерактивная система, представленная в Интернет, содержащая более 12000 терминов, относящихся к 69 тематическим разделам («словарям») науки и техники, соответствующим верхнему уровню Государственного рубрикатора научно-технической информации (ГРНТИ). Каждый термин представлен своими определениями (допускается несколько определений одного термина), ссылками на источники определений. В систему введены рубрикаторы УДК и ББК (частично), и определено соответствие индексов этих классификационных систем терминам, включенным в словари; представлены связи между терминами как внутри одного словаря, так и с терминами других словарей. Система предусматривает для авторизованных пользователей возможности ввода данных, их редактирования, просмотра истории изменения информации; обеспечивает развитую навигацию между элементами системы. Работа поддерживается грантами РФФИ и в перспективе может служить основой для формирования онтологии научных направлений, входящей в ядро единого цифрового пространства научных знаний.

Ключевые слова: терминологические словари; классификационные системы; онтология научных знаний; рубрикаторы; УДК, ГРНТИ; научные базы данных; интерактивная система; интернет

Для цитирования: Каленов Н. Е., Сенько А. М. Программная оболочка системы терминологических словарей // Научное издание международного уровня – 2019: стратегия и тактика управления и развития: материалы 8-й Международ. науч.-практ. конф., Москва, 23–26 апреля 2019 г. Екатеринбург: Изд-во Урал. ун-та, 2019. С. 58–62. DOI: 10.24069/konf-23-26-04-2019.08.

The program shell of the terminological dictionaries system

N. E. Kalenov^{1,2}, A. M. Senko¹

¹ Library for Natural Sciences of the Russian Academy of Sciences, Moscow, Russia

² Joint Supercomputer Centre of the Russian Academy of Sciences, Moscow, Russia

Abstract: An interactive system presented on the Internet is being considered. It contains more than 12,000 terms relating to 69 thematic sections (“dictionaries”) of science and technology corresponding to the upper level of the State Rubricator of Scientific and Technical Information (SRSTI). Each term is represented by its own definitions (several definitions of one term are allowed), with references to sources of definitions. The system includes the rubrics of the UDC and the Library Bibliographic Classification (partially) and determines the correspondence of the indices of these classification systems to the terms included in the dictionaries; connections between terms both within one dictionary and the different dictionaries are presented. The system provides for authorized users the ability to enter data, editing, view the history of information changes; provides advanced navigation between elements. The work is supported by RFBR grants and in the future can serve as a basis for the formation of the scientific fields ontology as the part of core of the unified digital scientific knowledge space.

Keywords: terminological dictionaries; glossaries of terms; classification systems; subject heading; ontology of scientific knowledge; UDC; SRSTI; scientific space; scientific data bases; interactive system; Internet

For citation: Kalenov N. E., Senko A. M. On the development of a Single Russian electronic space of knowledge. In: *World-Class Scientific Publication – 2019: Strategy and Tactics of Management and Development: Proc. 8th Int. Sci. & Pract. Conf., Moscow, April 23–26, 2019*. Ekaterinburg: Ural University Press, 2019, pp. 58–62. DOI: 10.24069/konf-23-26-04-2019.08.

Введение

Текущая версия интерактивной системы терминологических словарей (СТС) разрабатывается как элемент для создания онтологий, описывающих определенные предметные области науки. Достаточно подробно цель создания СТС и основные методологические подходы к ее реализации отражены в работах [1–6]. Пилотная версия СТС была создана в 2017–2018 гг. в рамках проекта, поддержанного РГНФ. Проект предусматривал загрузку в реляционную базу данных (БД) терминологических словарей, подготовленных специалистами ВИНТИ в виде совокупности файлов в формате MS Word, и разработку интерактивной системы, обеспечивающей возможность редактирования, ввода новых данных и навигации по ее ресурсам.

В результате реализации проекта была создана интерактивная СТС, содержащая информацию, относящуюся к 69-ти терминологическим словарям, каждый из которых соответствует одному из разделов верхнего уровня Государственного рубрикатора научно-технической информации (ГРНТИ) [7]. БД СТС содержат термины, относящиеся к данному разделу науки, связанные с соответствующим словарем, определения терминов и связанные с ними источники, из которых взяты определения.

Работы, реализованные в рамках создания пилотной версии СТС, в настоящее время продолжаются при поддержке грантов РФФИ 17-07-00153а и 18-00-00372 (КОМФИ). Они развиваются в двух направлениях – расширение информационной базы системы путем включения в нее индексов других классификационных систем (УДК и ББК) и установление тезаурусных связей между словарными терминами и различными классификационными индексами.

Ниже представлены характеристики функционирующей в настоящее время СТС и приведены примеры работы с системой.

Структура СТС

СТС, поддерживаемая на сервере БЕН РАН и доступная по адресу <http://class.labs.benran.ru> (рис. 1), осуществляет операции со следующими видами объектов: пользователи, словари, термины, определения терминов, источники определения терминов, индексы УДК, индексы ББК, связи между терминами. Ядром системы является реляционная база данных; в качестве СУБД используется MySQL 5.0. При создании системы использовался «каркас веб-приложений» Ruby on Rails версии 4.1 [8].

Работа с системой

Для работы с системой необходимо авторизоваться. Логин и пароль пользователю присваивает администратор, он же определяет права на выполнение тех или иных операций для данного пользователя (просмотр данных, редактирование данных, модификация структуры БД). После входа в систему пользователь попадает на основную страницу системы со списком словарей (рис. 1).

Каждое наименование словаря является активной ссылкой, при переходе по которой открывается окно с сообщением о количестве терминов, относящихся к данному словарю. В примере на рис. 2 показано количество терминов, относящихся к словарю «Информатика» (387).

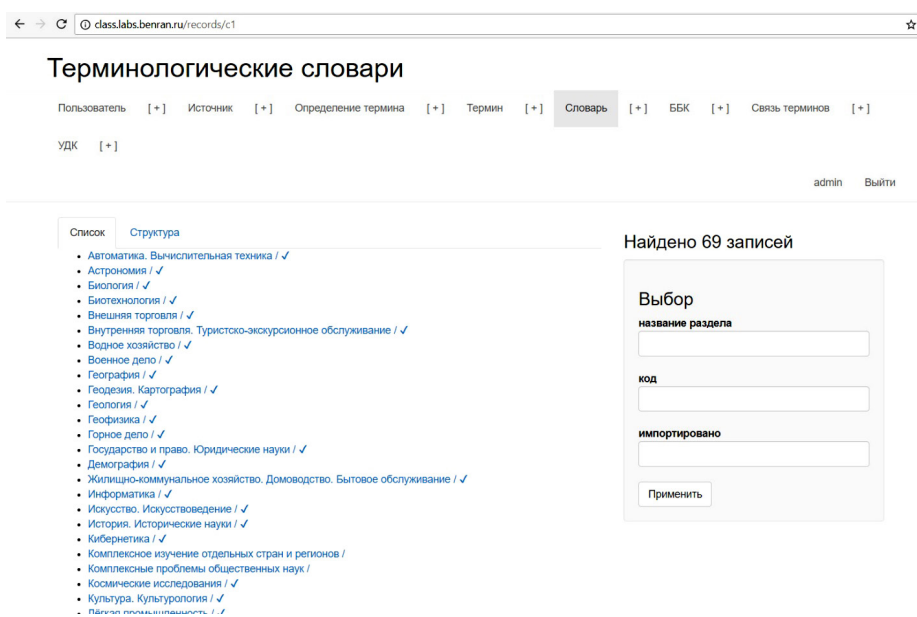


Рис. 1. Начальная страница ТСД

Информатика / ✓

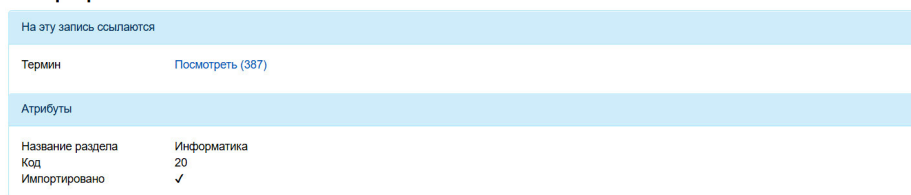


Рис. 2. Страница выбранного словаря

Перейдя по ссылке «Посмотреть», пользователь видит алфавитный список соответствующих терминов (рис. 3), каждый из которых является активной ссылкой, раскрывающей окно с подробной информацией, относящейся к данному термину.

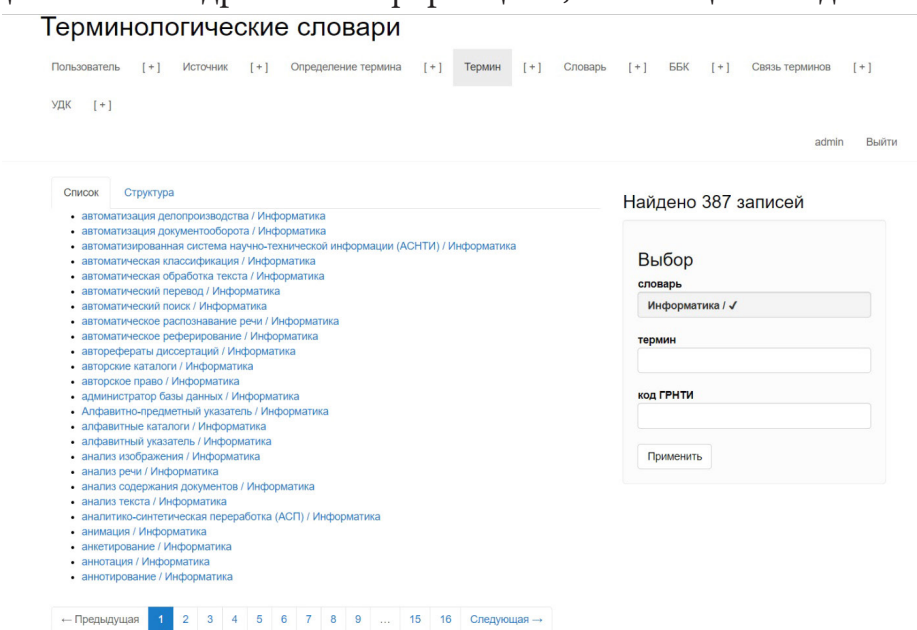


Рис. 3. Список терминов, относящихся к словарю «Информатика»

На рис. 4 показана страница, относящаяся к термину «Информационные запросы» словаря «Информатика». Страница содержит наименование термина с указанием, к какому словарю он относится (с возможностью перехода на словарь), индексы ГРНТИ, ББК и УДК, к которым относится термин (с возможностью перехода на таблицу ББК или УДК); ссылку на определение термина (а от него – на источник определения), указание на связи данного термина с другими.

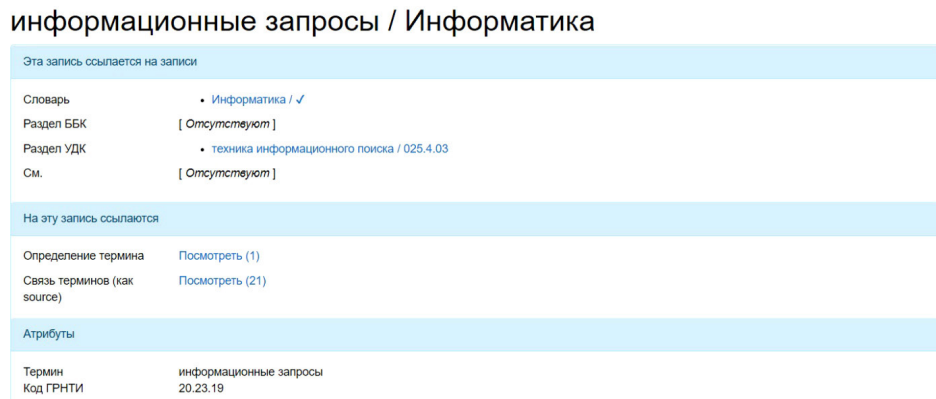


Рис. 4. Страница термина «Информационные запросы» словаря «Информатика»

В рассматриваемом примере переход по ссылке «Посмотреть» в графе «Связь терминов» открывает страницу со списком, включающим 21 термин, связанный с данным (рис. 5). Как видно из рис. 5, термин «информационные запросы» связан с терминами, относящимися к различным словарям.

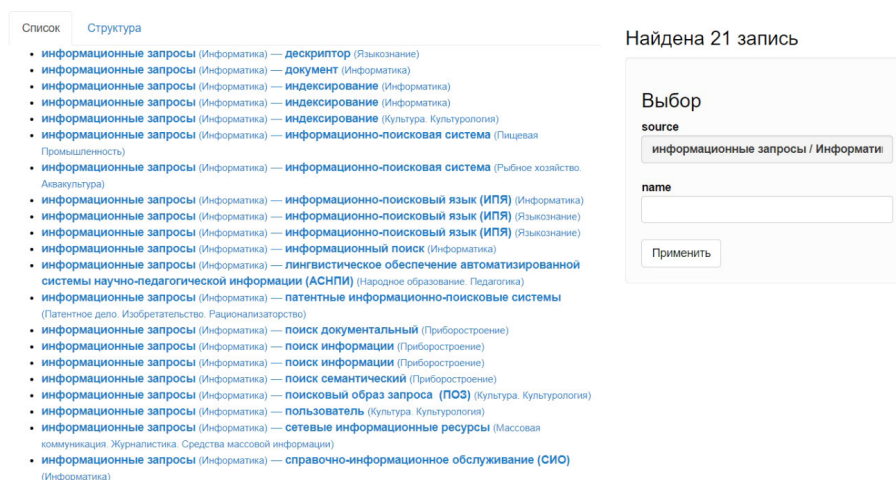


Рис. 5. Связи термина «информационные запросы»

Каждая связь, представленная в списке, является активной ссылкой, при переходе по которой можно увидеть, какой именно тип связи («содержит», «входит в...», «пересекается с...», «эквивалентен») установлен между данными терминами.

Программная оболочка СТС позволяет искать нужный объект любого вида по фрагменту его наименования (или кода рубрики для УДК и ББК). Для этого необходимо выбрать соответствующую вкладку (вид объекта) в верхней части начальной страницы (см. рис. 1) и задать нужный фрагмент в правой части страницы.

Заключение

На текущий момент в СТС отражено 12090 терминов, относящихся к 69 словарям; 12842 определения терминов, взятых из 7661 источника; 298381 связь между терминами; 3025 индексов УДК; 559 индексов ББК.

Список литературы

1. Белоозеров В. Н. Технология разработки терминологических словарей по лексике классификационных систем. В: Каленов Н. Е., Цветкова В. А. (ред.) *Информационное обеспечение науки: новые технологии: сборник научных трудов*. М.: БЕН РАН; 2015. С. 126–136.
2. Антошкова О. А., Белоозеров В. Н., Дмитриева Е. Ю. Разработка базовых соответствий между ГРНТИ и другими классификационными системами. В: Каленов Н. Е., Цветкова В. А. (ред.) *Информационное обеспечение науки: новые технологии: сборник научных трудов*. М.: БЕН РАН; 2015. С. 105–120.
3. Антошкова О. А., Белоозеров В. Н., Дмитриева Е. Ю., Шапкин А. В. Разработка онтологии НТИ на основе библиографических классификаций. В: Каленов Н. Е., Цветкова В. А. (ред.) *Информационное обеспечение науки: новые технологии: сборник научных трудов*. М.: БЕН РАН; 2017. С. 292–300.
4. Антопольский А. Б., Белоозеров В. Н., Маркарова Т. С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей. *Информационные ресурсы России*. 2017;(5):2–7.
5. Antopolskii A. V., Beloozerov V. N., Kalenov N. E., Shaburova N. N., Yakshin M. M. The Development of a Semantic Network of Keywords Based on Definitive Relationships. *Scientific and Technical Information Processing*. 2017;44(4):261–265. DOI: [10.3103/S0147688217040062](https://doi.org/10.3103/S0147688217040062)
6. Антопольский А. Б., Белоозеров В. Н., Каленов Н. Е., Маркарова Т. С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов. *Информационные ресурсы России*. 2018;(5):22–30.
7. Государственный рубрикатор научно-технической информации. Режим доступа: <http://www.extech.ru/info/catalogs/grnti/> [Дата обращения: 11.06.2019].
8. *Ruby on Rails*. Available at: https://en.wikipedia.org/wiki/Ruby_on_Rails [Дата обращения: 18.06.2019].

Информация об авторах

Каленов Николай Евгеньевич – доктор технических наук, профессор, главный научный сотрудник Отдела системных исследований и автоматизированных технологий БЕН РАН; ведущий научный сотрудник Отдела информационных ресурсов и систем МСЦ РАН (филиал ФГУ ФНЦ НИИСИ РАН), г. Москва Россия; e-mail: nekalenov@yandex.ru.

Сенько Александр Михайлович – научный сотрудник Отдела системных исследований и автоматизированных технологий БЕН РАН, г. Москва, Россия; e-mail: alexander.senko@gmail.com.

Information about the authors

Nikolay E. Kalenov – Doctor of Sciences (Engineering), Professor, Chief Researcher, System Research & Automation Technology Department., Library for Natural Sciences of RAS; Leader Researcher, Information Resources & Systems Department of the Joint Supercomputer Centre of RAS (Branch of the Scientific Centre for System Research); Moscow, Russia; e-mail: nekalenov@yandex.ru.

Alexander M. Senko – Researcher, System Research & Automation Technology Department, Library for Natural Sciences of RAS; Moscow, Russia; e-mail: alexander.senko@gmail.com.