

**ИНФОРМАЦИОННО-РЕСУРСНАЯ ИНФРАСТРУКТУРА
СИСТЕМЫ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ**

**THE INFORMATION AND RESOURCE INFRASTRUCTURE
OF THE SYSTEM OF SCIENTIFIC AND TECHNICAL
INFORMATION**

УДК [025.4.02+025.4.06](476)

<https://doi.org/10.47612/978-985-880-283-7-2022-11-18>

Л. Л. Астапович, Е. Н. Гирко

*ГУ «Центральная научная библиотека имени Якуба Коласа
Национальной академии наук Беларуси», Минск, Беларусь*

**ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ
СМЫСЛОВОЙ ОБРАБОТКИ ДОКУМЕНТОВ В
ЦЕНТРАЛЬНОЙ НАУЧНОЙ БИБЛИОТЕКЕ
НАН БЕЛАРУСИ**

Аннотация. Представлен опыт формирования и использования лингвистических средств в Центральной научной библиотеке НАН Беларуси, обеспечивающих процессы смысловой обработки документов, включающих классификационное и координатное индексирование. Дана краткая характеристика рабочих таблиц классификации, составляемых на основе таблиц Универсальной десятичной классификации. Описана технология работы по созданию авторитетных записей и ключевых слов, необходимых для составления поискового образа документов, обозначены критерии целесообразности их формирования. Рассмотрены вопросы ведения локального рубрикатора библиотеки, применяемого для индексирования научных статей.

Ключевые слова: лингвистические средства, смысловая обработка документов, поисковый образ документов, рабочие таблицы классификации, Универсальная десятичная классификация, авторитетные записи, ключевые слова, локальный рубрикатор.

Для цитирования. Астапович, Л. Л. Лингвистическое обеспечение процессов смысловой обработки документов в Центральной научной библиотеке НАН Беларуси / Л. Л. Астапович, Е. Н. Гирко // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий. Тема 2022 года – «Библиотеки в системе информационных и социальных коммуникаций»: докл. V Междунар. науч. конф., Минск, 1–2 дек. 2022 г. / Белорус. с.-х. б-ка им. И. С. Лупиновича Нац. акад. наук Беларуси ; редкол.: Ю. О. Каракулько (отв. ред.) [и др.]. – Минск, 2022. – С. 11–18.

L. L. Astapovich, E. N. Girko

Central Scientific Library named after Yakub Kolas of the National Academy of Sciences of Belarus, Minsk, Belarus

LINGUISTIC SUPPORT OF THE SEMANTIC PROCESSING OF DOCUMENTS IN THE CENTRAL SCIENTIFIC LIBRARY OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

Abstract. The article presents the experience of creating and using of linguistic tools in the Yakub Kolas Central Scientific Library of the NAS of Belarus for supporting the semantic processing of documents including classification and coordinate indexing. The publication provides a brief description of the classification worksheets compiled on the basis of the Universal Decimal Classification tables. It also describes the technology used for creation of authority records and keywords needed for compilation of a document search description as well as the criteria for their appropriateness. The article considers the issues concerning maintenance of the library local subject headings list used for indexing of scientific publications.

Keywords: linguistic tools, semantic processing of documents, document search description, classification worksheets, Universal Decimal Classification, authority records, keywords, local subject headings list.

For citation. Astapovich L. L., Girko E. N. Linguistic support of the semantic processing of documents in the Central Scientific Library of the National Academy of Sciences of Belarus. Libraries in the information society: preserving traditions and developing new technologies. The theme for 2022 – «Libraries in the information and social communication system»: proceedings of the V international scientific conference, Minsk, December 1–2, 2022. Minsk, 2022, pp. 11–18 (in Russian).

Одним из видов деятельности библиотеки, направленной на раскрытие фонда и обеспечение его доступности для пользователей, является каталогизация документов. Отражать в электронном каталоге (ЭК) информацию о тематике и назначении документов, образующих фонд библиотеки, позволяет смысловая обработка, в ходе которой формируется поисковый образ документов (ПОД). Для обеспечения процессов смысловой обработки используются лингвистические средства, включающие совокупность информационно-поисковых языков (ИПЯ), используемых при классификационном и координатном индексировании, и методов перевода информации о содержании документов с естественного языка на ИПЯ.

В Центральной научной библиотеки им. Я. Коласа Национальной академии наук Беларуси (ЦНБ НАН Беларуси)

поисковый образ составляется на документы, поступающие в фонд библиотеки, и при аналитической росписи статей научных журналов. В качестве ИПЯ используются таблицы классификаций, авторитетные записи (АЗ), ключевые слова (КС). Лингвистические средства для составления ПОД формируются и поддерживаются в основном в отделе лингвистического обеспечения (ОЛО) библиотеки.

Классификационное индексирование в ЦНБ НАН Беларуси осуществляется по Универсальной десятичной классификации (УДК). Однако для обеспечения унифицированного подхода к применению и формированию сложных и составных индексов, использованию отдельных индексов и определителей, при необходимости собрать в определенном разделе новые термины, еще не включенные в таблицы УДК, составляются и ведутся рабочие таблицы классификации (РТК). Отраслевые РТК, создаваемые в электронной форме для ЦНБ НАН Беларуси на основе полных таблиц УДК, отражают особенности структуры и степень детализации электронного каталога (ЭК) библиотеки. Для представления в РТК понятий, отсутствующих в УДК, принимаются решения по индексированию с применением правил общей методики УДК.

Рабочие таблицы составляются или перерабатываются после методического редактирования ЭК по разделам УДК. В процессе индексирования документов и работы с ЭК учитывается информация, представленная в выпусках бюллетеня «УДК. Изменения и дополнения», поэтому при редактировании в каталог вводятся новые индексы и исключаются разделы, наполнение которых переходит в другие отделы, подходящие по тематическому содержанию, что также находит отражение в отраслевых РТК.

Отраслевые таблицы структурно переработаны, снабжены ссылочным аппаратом, методическими рекомендациями и пояснениями, раскрывающими содержание заголовков классификационных разделов. Электронный вариант РТК позволяет оперативно вносить в таблицы все произошедшие изменения в УДК, что делает возможным поддержание актуальности рабочих таблиц [1, с. 225].

Координатное индексирование в библиотеке выполняется посредством АЗ и КС, которые позволяют детально и многоаспектно раскрыть тему документа и обозначить ее в ПОД.

Для описания смыслового содержания документов и обеспечения полноты поискового образа используются АЗ из следующих словарей сводного электронного каталога библиотек Беларуси (СЭК): «Тематический предмет», «Имя лица», «Наименование организации/коллектива», «Родовое имя», «Географическое название», «Торговая марка», «Унифицированное название», «Имя/Заглавие», «Имя/обобщающее унифицированное заглавие», «Форма, жанр и физические характеристики».

АЗ создаются библиотеками-участницами системы корпоративной каталогизации (СКК), в том числе и ЦНБ НАН Беларуси, с учетом требований формата BELMARC/Authorities. Для АЗ составляются справочные примечания, которые включают сведения, уточняющие и конкретизирующие объект. Справочная информация содержит определение термина, варианты наименования и синонимы, выше- и нижестоящие термины, эквиваленты на других языках. Обращение к справкам помогает определиться с выбором АЗ для составления ПОД и для формирования поисковых запросов пользователями.

При необходимости создания новых АЗ специалисты библиотеки формируют заявки с указанием научных понятий, которые надо отобразить в ПОД, и передают их в ОЛО, где выполняется логико-семантическая обработка заявленных терминов. В отделе записи создаются на наименования: «тематический предмет», «заглавие», «имя/заглавие», «имя/типовое заглавие», «родовое имя», «торговая марка», «форма, жанр или физические характеристики». После проверки на наличие в словарях СЭК наименований на географические объекты, персоналии и организации заявки на формирование АЗ на «географические названия» передаются в Национальную библиотеку Беларуси, а на «имя лица» и «наименование организации» – в отдел научной обработки ЦНБ НАН Беларуси.

В процессе логико-семантической обработки осуществляется и оценка перспективности применения заявленных терминов в

поисковых образах. Если предполагается их активное использование и планируется увеличение количества библиографических записей (БЗ), «привязанных» к создаваемой АЗ, то принимается решение о целесообразности формирования новой АЗ на термин. В тех случаях, когда в соответствии с анализом перспективности определяется небольшая вероятность дальнейшего использования этих терминов в БЗ, то составляются ключевые слова (КС).

Создаются КС и на понятия, которые не представлены в авторитетных источниках, но необходимы для смысловой обработки документов. Надо отметить, что КС для лингвистического обеспечения ЭК имеют существенное значение, так как составлены в основном на узкоспециальные и авторские термины. Они дополняют ПОД и позволяют наиболее полно раскрыть содержание документов [2, с. 261].

Для улучшения качества словарей СЭК и локального каталога библиотеки, обеспечения более четкого выбора дескрипторов при смысловой обработке документов каталогизаторы библиотек, участвующих в СКК и ведении СЭК, редактируют АЗ. Специалисты ОЛО при редактировании записей выполняют лингвоструктурный анализ наименований, устанавливают иерархические и ассоциативные связи АЗ. В ходе редактирования в АЗ вносятся изменения для корректировки их смыслового наполнения, составляются или дорабатываются справочные примечания, раскрывающие значение понятий. При совпадении семантического содержания нескольких записей они объединяются в одну АЗ.

Для поисковых образов, составляемых в рамках создания аналитических БЗ на статьи из научных журналов, помимо АЗ определяются коды локального рубрикатора библиотеки. Рубрикатор формировался как часть подсистемы избирательного распространения информации (ИРИ) ЦНБ НАН Беларуси и является лингвистическим средством (классификационным информационно-поисковым языком) для описания и фиксации профессиональных интересов пользователей, в первую очередь абонентов ИРИ.

Классификационная таблица рубрикатора составлена специалистами ОЛО, разработка программного обеспечения выполнена сотрудниками лаборатории информационного обеспечения научных исследований Объединенного института проблем информатики Национальной академии наук Беларуси. Рубрикатор представляет собой иерархическую политематическую классификационную систему. Каждому тематическому понятию присвоен уникальный код. С целью дальнейшего расширения состава рубрикатора при его разработке цифровые коды присваивались рубрикам не подряд, а с определенными промежутками, чтобы при необходимости можно было вводить новые разделы. Для каждой рубрики определены соответствующие индексы УДК.

Локальный рубрикатор ЦНБ НАН Беларуси создавался как прагматическая классификация, поэтому включает тематику документов, поступающих в библиотеку, и потребности пользователей библиотеки. Тематически рубрикатор разделен на четыре класса: общий отдел, гуманитарные и общественные науки; математика, естественные науки; прикладные и технические науки, медицина.

С 2016 года рубрикатор используется в ЦНБ НАН Беларуси для индексирования статей из научных периодических изданий. Работа с политематическим рубрикатором в основном ведется в онлайн-режиме. Использование онлайн-версии обеспечивает выполнение следующих операций: просмотр рубрикатора с переходом по уровням иерархии; поиск рубрик по коду или по словам из названия; поиск рубрик по УДК; копирование рубрик в 690 поля БЗ (процесс доступен каталогизаторам ЦНБ НАН Беларуси, выполняющим создание или редактирование аналитических БЗ) [3, с. 283].

Осуществляется работа и по поддержке рубрикатора, направленная на его расширение актуальными темами, необходимость которых определяется в процессе аналитической росписи журнальных статей. При этом выполняется интеллектуальный анализ новых понятий, включающий определение их лексического содержания, установление смысловой связи терминов с разделами рубрикатора. Проводится и редактирование существующих рубрик. Все изменения

вносятся в модуль ведения рубрикатора (онлайн-рубрикатор). После редакции и актуализации рубрикатора составляются таблицы изменений и дополнений для специалистов библиотеки, использующих его в работе.

Формирование, актуализация и применение согласно установленным методам и правилам лингвистических средств обеспечивает качественное выполнение смысловой обработки документов. Использование РТК, в которые оперативно включаются дополнения из таблиц УДК и принятые решения по индексированию, АЗ и КС дает возможность унифицировать результаты процесса индексирования, сформировать единообразные подходы к составлению поискового образа, раскрывающего смысловое содержание документов, что позволяет получить наиболее релевантную информацию при проведении тематического поиска в ЭК библиотеки.

Список использованных источников:

1. Астапович, Л. Л. Рабочие таблицы классификации ЦНБ НАН Беларуси: традиционный методический аппарат в электронной форме / Л. Л. Астапович, Ж. К. Зенькевич // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий : докл. междунар. науч. конф., 3–4 дек. 2014 г. / Белорус. гос. с.-х. б-ка им. И. С. Лупиновича Нац. акад. наук Беларуси ; редкол.: В. В. Юрченко [и др.]. – Минск, 2014. – С. 221–225.

2. Астапович, Л. Л. Ведение словаря ключевых слов электронного каталога Центральной научной библиотеки НАН Беларуси / Л. Л. Астапович // Развитие информатизации и государственной системы научно-технической информации: РИНТИ-2021 : докл. XX Междунар. конф., 18 нояб. 2021 г., Минск / Объед. ин-т проблем информатики Нац. акад. наук Беларуси ; науч. ред.: А. В. Тузиков, Р. Б. Григянец, В. Н. Венгеров. – Минск, 2021. – С. 257–261.

3. Астапович, Л. Л. Разработка политематического рубрикатора для подсистемы избирательного распространения информации в ЦНБ НАН Беларуси / Л. Л. Астапович, Г. О. Лаужель // Развитие информатизации и государственной системы научно-технической информации: РИНТИ-2017 : докл. XVI

Международ. конф., 16 нояб. 2017 г., Минск / Объед. ин-т проблем информатики Нац. акад. наук Беларуси ; науч. ред.: А. В. Тузиков, Р. Б. Григянец, В. Н. Венгеров. – Минск, 2017. – С. 280–284.

References:

1. Astapovich L. L., Zen'kevich Zh. K. Working tables of the classification of the Central Scientific Library of the National Academy of Sciences of Belarus: a traditional methodological apparatus in electronic form. *Biblioteki v informatsionnom obshchestve: sokhranenie traditsii i razvitie novykh tekhnologii: doklady mezhdunarodnoi nauchnoi konferentsii, Minsk, 3–4 dekabrya 2014 g.* [Libraries in the information society: preserving traditions and developing new technologies: proceedings of international scientific conference, Minsk, December 3–4, 2014]. Minsk, 2014, pp. 221–225 (in Russian).
2. Astapovich L. L. Maintaining a dictionary of keywords of the electronic catalogue of the Central Scientific Library of the National Academy of Sciences of Belarus. *Razvitie informatizatsii i gosudarstvennoi sistemy nauchno-tekhnicheskoi informatsii: RINTI-2021: doklady XX Mezhdunarodnoi konferentsii, 18 noyabrya 2021 g., Minsk* [Development of informatization and the state system of scientific and technical information: RINTI-2021: reports of the XVII International conference, November 20, 2021, Minsk]. Minsk, 2021, pp. 257–261 (in Russian).
3. Astapovich L. L., Lauzhel' G. O. Development of a polythematic subject headings list for the subsystem of selective dissemination of information in the Central Library of the National Academy of Sciences of Belarus. *Razvitie informatizatsii i gosudarstvennoi sistemy nauchno-tekhnicheskoi informatsii: RINTI-2017: doklady XVI Mezhdunarodnoi konferentsii, 16 noyabrya 2017 g., Minsk* [Development of informatization and the state system of scientific and technical information: RINTI-2017: reports of the XVI International conference, November 16, 2017, Minsk]. Minsk, 2017, pp. 280–284 (in Russian).

Поступила в редакцию 02.09.2022

Received 02.09.2022