

**Информационно-поисковый тезаурус Парламентской библиотеки:  
этапы разработки, ведение, применение и дальнейшие перспективы**

**Information Retrieval Thesaurus of the Parliamentary Library:  
Development Stages, Maintenance, Application, and Prospects**

**Інформаційно-пошуковий тезаурус Парламентської бібліотеки:  
етапи розробки, провадження, застосування та подальшої перспективи**

*Т. А. Москаленко, Н. А. Мякова*

*Парламентская библиотека, Государственная Дума, Москва, Россия*

*Tatyana Moskalenko and Nataliya Myakova*

*Parliamentary Library, State Duma, Moscow, Russia*

*Т. А. Москаленко, Н. А. Мякова*

*Парламентська бібліотека, Державна Дума, Москва, Росія*

Информационно-поисковый тезаурус Парламентской библиотеки (ИПТ) является расширенной русской версией тезауруса Eurovoc. С 2000 г. ИПТ используется для индексирования и поиска информации в электронном каталоге Парламентской библиотеки, а английские и французские эквиваленты дескрипторов позволяют осуществлять поиск в зарубежных ресурсах. В докладе и сопровождающей его презентации показаны этапы разработки, структура, применение, технология ведения тезауруса, а также перспективы использования его Web-представления.

The information retrieval thesaurus of the Parliamentary Library (IRT) is an expanded Russian version of the Eurovoc thesaurus. Since 2000 the IRT has been used for indexing and information search in the Parliamentary Library electronic catalog. The English and French equivalents of descriptors enable search in foreign information resources. The paper and the corresponding presentation describe the steps of the IRT development, its structure, application, management and maintenance as well as the prospects of its web version.

Інформаційно-пошуковий тезаурус Парламентської бібліотеки (ІПТ) є розширеною російською версією тезауруса Eurovoc. З 2000 р. ІПТ використовується для індексування та пошуку інформації в електронному каталозі Парламентської бібліотеки, а англійські та французькі дескриптори дозволяють здійснювати пошук у зарубіжних ресурсах. У доповіді та супровідній презентації представлено етапи розробки, структуру, застосування, технологію провадження тезауруса, а також перспективи використання його Web-подання.

Многоязычный тезаурус Eurovoc был специально разработан для индексирования и поиска информации в документальных и библиографических базах данных ЕС. Тематически он охватывает все сферы деятельности европейских институтов. В настоящее время пользователями этого информационно-лингвистического средства являются Европейский парламент, Бюро официальных публикаций ЕС, парламентские библиотеки, информационно-аналитические и документационные службы ряда европейских стран, в том числе в последнее время и стран Центральной и Восточной Европы. В 1995 г., когда Парламентская библиотека начала заниматься проектом перевода тезауруса Eurovoc на русский язык, его официальная версия была представлена на 11 языках, в настоящее время – уже на 21 официальном языке Европейского Союза, а также на албанском, русском, хорватском и украинском.

В 1998 г. в Парламентской библиотеке была подготовлена русская версия тезауруса Eurovoc, был осуществлен его перевод и адаптация, вышло издание многоязычного представления тезауруса. В 2000 г. было осуществлено поэтапное объединение русской версии с собственным тезаурусом библиотеки. Объем полученного тезауруса оказался более 11 тыс. лексических единиц и около 16, 5 тыс. семантических связей. В 2001 г. был осуществлен выпуск нового трехтомного издания информационно-поискового тезауруса, включающего алфавитно-пермутационное, тематическое и многоязычное (на русском, французском, английском языках) представления.

Эта новая редакция тезауруса была одобрена на заседании Научно-консультативного совета по систематизации, кодификации законодательства и правовой информации при Председателе Государственной Думы и было рекомендовано продолжить работу как в направлении практического применения тезауруса в парламентских информационных ресурсах, так и дальнейшего его развития для обеспечения информационной, лингвистической и терминологической поддержки законотворческой деятельности депутатов.

В 2004 г. в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам была осуществлена официальная регистрация БД Тезаурус Парламентской библиотеки «ЛингваПарл» (свидетельство № 2004620218; правообладатель Аппарат Государственной Думы Федерального Собрания Российской Федерации).

По состоянию на конец апреля 2008 г. в составе ИПТ Парламентской библиотеки около 13 тыс. лексических единиц и около 24 тыс. семантических связей.

Структура ИПТ Парламентской библиотеки аналогична структуре тезауруса Eurovoc, за исключением двух типов новых связей в тезаурусных статьях аскрипторов.

Отличительной чертой этой тезаурусной структуры является ее интегральный характер, а именно: соединение двух типов информационно-поисковых языков: предметно-классификационного и дескрипторного. Структура тезауруса на верхнем уровне представлена иерархической классификацией – 21 тематическим разделом и 127 микротезаурусами, входящими в их состав, на нижнем уровне – лексическими единицами и их тезаурусными статьями. В тезаурусных статьях помимо примечания, уточняющего значение и употребление лексической единицы, указываются: тематическая принадлежность дескриптора, отношения условной эквивалентности, иерархические и ассоциативные отношения.

Для дескрипторов ИПТ Парламентской библиотеки, которые являются и дескрипторами Eurovoc, предусмотрены также эквиваленты на английском и французском языках, что позволяет осуществлять поиск в зарубежных поисковых системах (прежде всего ЕС), где используются тезаурус Eurovoc.

Организация содержательной обработки документов в Парламентской библиотеке построена таким образом, чтобы использовать возможности обоих информационно-поисковых языков, содержащихся в тезаурусе. Тезаурус применяется для аналитической обработки всех видов документов, поступающих в фонд библиотеки, и информационного поиска в базах данных:

- отечественных и зарубежных книг, авторефератов диссертаций, неопубликованных материалов;

- статей из журналов и сборников;
- публикаций официальных документов.

По состоянию на конец апреля 2008 г. в электронном каталоге Парламентской библиотеки проиндексировано с помощью тезауруса:

- около 172 тыс. книг, авторефератов диссертаций и неопубликованных материалов,
- около 201 тыс. статей,
- около 286 тыс. публикаций официальных документов.

С использованием названий разделов и микротезаурусов был подготовлен предметно-тематический ряд для представления библиографической информации в Бюллетене новых поступлений Парламентской библиотеки.

При регистрации поступающих запросов в системе контроля за выполнением запросов осуществляется их тематическая классификация в соответствии с перечнем микротезаурусов и ежемесячно осуществляется подготовка тематических рейтингов запросов (в соответствии с разделами тезауруса).

Одним из разделов информационного продукта Парламентской библиотеки, который называется «Библиодосье к законопроекту», является перечень основных понятий по теме законопроекта: в терминологии тезауруса «ЛингваПарл» на русском языке и в терминологии тезауруса Eurovoc – на английском и французском языках.

Постоянно ведется работа по анализу новой лексики, используемой при обработке поступающей в фонд Парламентской библиотеки литературы, тезаурус пополняется новыми лексическими

единицами и семантическими отношениями, при необходимости изменяется статус ранее используемых лексических единиц (с соответствующим редактированием базы данных).

Источниками пополнения и изменения лексического состава и структуры тезауруса являются:

- актуализированная официальная версия Eurovoc;
- предложения сотрудников Парламентской библиотеки, занимающихся индексированием документов;
- предложения сотрудников Парламентской библиотеки, занимающихся поиском информации в электронном каталоге библиотеки и во внешних информационных ресурсах;
- анализ запросов пользователей Парламентской библиотеки;
- статистический анализ дескрипторов, использованных при индексировании документов;
- статистический анализ неконтролируемой лексики, использованной при индексировании документов.

Технологически ведение информационно-поискового тезауруса Парламентской библиотеки обеспечивается такими процедурами, как:

- поддержание в контрольном состоянии лексического состава и структуры тезауруса;
- актуализация тезауруса для индексирования документов и поиска информации;
- подготовка новых русских версий Eurovoc в соответствии с актуализированными официальными версиями Eurovoc (при необходимости);
- подготовка и актуализация методической документации по использованию тезауруса для индексирования документов и запросов.

Для ведения тезауруса в электронном виде используется специально разработанное программное обеспечение, предполагающее возможность ввода терминов на русском, английском, французском языках с возможностью установления семантических связей, построения тезаурусных статей, необходимых сортировок и печати. По мере накопления новой лексики производится встраивание в структуру тезауруса новых терминов с соответствующими операциями в базе данных ведения тезауруса. Обновленный тезаурус выгружается из специализированной базы данных ведения тезауруса в формате, пригодном для импорта в базу данных, встроенную в базовую автоматизированную систему, используемую при обработке потока изданий и документов. В автоматизированной библиотечно-информационной системе Парламентской библиотеки ИПТ используется при индексировании документов и при поиске информации с возможностью поиска терминов тезауруса по древовидной структуре и списку терминов с навигацией по дереву тезауруса, связям и микротезаурусам.

Одним из направлений оптимизации использования ИПТ Парламентской библиотеки является разработка его Web-представления, обеспечивающего возможность навигации по тезаурусу, в том числе, по «дереву» – его иерархической структуре и поиск лексических единиц тезауруса по значимому слову (основе значимого слова), входящего в их состав. Установка Web-представления тезауруса в электронный каталог с удаленным доступом позволит существенно облегчить отбор дескрипторов, составление поискового запроса и, при необходимости, уточнение информационной потребности пользователя.

Web-представление тезауруса также позволит использовать его как терминологический словарь в Автоматизированной системе обеспечения законодательной деятельности Государственной Думы и других электронных ресурсах Государственной Думы.

В 2007 году Парламентская библиотека завершила «пилотный» проект по созданию Парламентской электронной библиотеки (ПЭБ). Для организации навигации и поиска информации в ПЭБ, в том числе и при полнотекстовом поиске, используется Тезаурус Парламентской библиотеки «ЛингваПарл». Предусмотрен поиск по тезаурусу и по дереву тезауруса с целью отбора терминов и переноса их в форму для осуществления поиска. При отборе терминов из тезауруса пользователю предоставляется возможность просматривать тезаурусную статью, осуществлять переходы по структуре тезауруса в соответствии с иерархическими и ассоциативными связями. Поиск с использованием терминов тезауруса возможен как по распознанным копиям, так и по атрибутам электронных документов, хранящимся в ПЭБ.