

УДК 025.4(477)

О. О. Сербин,

зав. отделом НБУВ,

кандидат исторических наук

**НЕКОТОРЫЕ АСПЕКТЫ РАЗВИТИЯ ДЕСКРИПТОРНЫХ
ИНФОРМАЦИОННО-ПОИСКОВЫХ ЯЗЫКОВ В УКРАИНЕ
(на примере информационно-поискового тезауруса
Национальной парламентской библиотеки Украины)**

В статье рассмотрены вопросы развития дескрипторных информационно-поисковых языков в Украине на опыте «Информационно-пошукового тезауруса» Национальной парламентской библиотеки Украины. Представлены сведения о разработке тезауруса, проанализирована его структура, определено влияние «Информационно-пошукового тезауруса» Национальной парламентской библиотеки Украины на развитие дескрипторных информационно-поисковых языков в библиотеках Украины.

Ключевые слова: тезаурус, дескриптор, аскриптор, индекс, классификация, систематизация.

The questions of development of descriptor information retrieval languages in Ukraine for example of «Information retrieval thesaurus» of the National Parliamentary Library of Ukraine, are investigated. Historical data of emergence and development of «Information retrieval thesaurus» of the National Parliamentary Library of Ukraine, and analysis of its structure and components are presented. Significance and role of the «Information retrieval thesaurus» of the National Parliamentary Library of Ukraine in the development of descriptor information retrieval languages in Ukraine are defined.

Keywords: thesaurus, descriptor, ascriptor, index, classification, systematization.

В основу информационно-поисковых языков классификационного типа в Украине положены «Универсальная десятичная классификация» [14], которая была переведена и адаптирована специалистами Книжной палаты Украины имени Иван Фёдоров и «Рубрикатор Национальной библиотеки Украины имени В.И. Вернадского» [11] (первый выпуск бумажной версии [13] был презентован в 2005 г.), который является, по сути, рабочими таблицами НБУВ, разработанными на основе Библиотечно-библиографической классификации [1].

Дескрипторный ИПЯ до 2004 г. не был представлен ни в одном фундаментально-направленном отдельном издании, с перспективой широкого использования в работе библиотек Украины. Ситуация изменилась, когда в феврале 2004 г. вышел в свет «Информационно-пошуковий тезаурус»

(далее Тезаурус), подготовленный Национальной парламентской библиотекой Украины (НПБУ). Это издание способствовало решению проблемы обеспечения эквивалентности поиска в автоматизированных информационно-библиотечных системах, формализации процедуры перевода информации с естественного языка на ИПЯ и наоборот, для ее записи и поиска.

Объектом данной публикации является Тезаурус. Предметом, данной публикации является исследование подходов к созданию Тезауруса и его структурный анализ. Целью данной публикации является освещение возможностей использования данного Тезауруса в практике библиотек Украины.

В общем понимании информационно-поисковый тезаурус – это нормативный словарь дескрипторного информационно-поискового языка с зафиксированными в нем парадигматическими отношениями лексических единиц [9], или же – контролируемый словарь терминов с зафиксированными семантическими отношениями, охватывающий одну или несколько специальных областей знания [10]. Учитывая выше указанную информацию, можно сделать вывод, что разработанный специалистами НПБУ Тезаурус – это универсальный, нормативный, контролируемый словарь лексических единиц украинского языка, с зафиксированными семантическими отношениями, который сегодня, в период активного создания электронных каталогов, БД, имеет практическую ценность для библиотек Украины.

Сначала специалисты НПБУ изучили мировой и отечественный опыт по созданию словарей дескрипторных ИПЯ. Были проанализированы аспекты создания «предметного каталога» [6, с. 58] в Харьковской государственной библиотеке им. В. Г. Короленко; история создания и использования «Library of Congress Subject Headings» – «Предметные рубрики Библиотеки Конгресса» – информационного продукта Библиотеки Конгресса США; опыт издания и использования универсального «Словаря языка предметных рубрик Библиотеки Народовой» [3, 7]; опыт по разработке тезаурусов по общественным наукам Института научной информации по общественным наукам РАН. Были изучены стандарты по данной тематике. В результате проведенной исследовательской работы специалисты НПБУ определили основные подходы к созданию собственного украинского тезауруса. В конце 1997 г. группа специалистов НПБУ приступила к работе. Первый этап по созданию тезауруса заключался в «отборе терминов из разных областей знаний, которые отражают предмет и установлению связей между понятиями, представленными этими терми-

нами» [7]. Для создания Тезауруса были использованы терминологические словари, энциклопедии, справочные издания, учебные пособия, терминологические стандарты, таблицы Универсальной десятичной классификации и Библиотечно-библиографической классификации. Британский Совет в Украине предоставил НПБУ Британский стандарт BS 5723:1987 «Guide to establishment and development of monolingual thesauri» – «Создание и ведение одноязычных тезаурусов», который является идентично-соответствующим Международному стандарту ISO 2788:1986 «Documentation; Guidelines for the establishment and development of monolingual thesauri» – «Документирование – Методические советы по созданию и ведению одноязычных тезаурусов». Конкретизируя информацию относительно стандартов, следует отметить, что в работе также использовались: Государственный стандарт Украины ДСТУ 2398-94 «Информация и документация. Информационные языки. Термины и определения»; Межгосударственный стандарт ГОСТ 7.74-96 «Информационно-поисковые языки. Термины и определения» и Государственный стандарт Украины ДСТУ 3966-2000 «Принципы и правила разработки стандартов на термины и определения».

О втором этапе, по нашему мнению, можно говорить, как о работе, которая началась в конце 1999 г., уже после того, как было отобрано около 10 тыс. терминов и был проведен процесс инверсирования естественного языка для формирования устойчивого предметного ядра того или иного понятия. С этого момента началось редактирование Тезауруса в традиционной, карточной форме и ведение в электронном виде. В этот же период НПБУ выиграла конкурс на получение гранта от фонда «Возрождение». В результате в распоряжении коллектива НПБУ были деньги на материалы для издания тезауруса, на оргтехнику, на изучение (путем командирования специалистов) ИПЯ в библиотеках Польши, на оплату консультаций высококвалифицированных специалистов различных отраслей, которые активно помогали сотрудникам НПБУ в создании Тезауруса.

Содержательное наполнение Тезауруса является адекватным отражением основного фонда НПБУ. Так как пополнение фонда происходит постоянно и непрерывно, формирование основного массива Тезауруса – процесс динамичный и непрерывный, требующий фиксации нового знания. Поэтому, чтобы приступить к заключительному этапу – непосредственному изданию информационно-поискового словаря – специалисты НПБУ приняли решение к концу 2003 г. завершить формирование основного массива Тезауруса. В ноябре 2003 г. Тезаурус и методические реко-

мендации по его использованию рекомендованы к изданию. В феврале 2004 г. Тезаурус был издан как в традиционной форме, так и в электронном варианте на CD.

Таким образом, в Украине появилась информационно-поисковая система украиноязычных терминов (дескрипторов – 15,1 тыс. и аскрипторов – 5,1 тыс.), которая физически очерчена в отдельное кодексное и электронное издание. Дескрипторы в этой системе ни что иное, как ключевые слова, лексические единицы, именующие определенную рубрику, класс, в который входят близкие по смыслу или синонимично термины. Аскрипторы – это тоже лексические единицы, но они подлежат замене на соответствующие дескрипторы для осуществления информационно-поисковой работы. В отличие от дескрипторов, аскрипторы (нондескрипторы) [12, с. 397, 398] не могут быть использованы для координатного индексирования. Таким образом, можно говорить о принципе дескриптности как главном приеме построения Тезауруса.

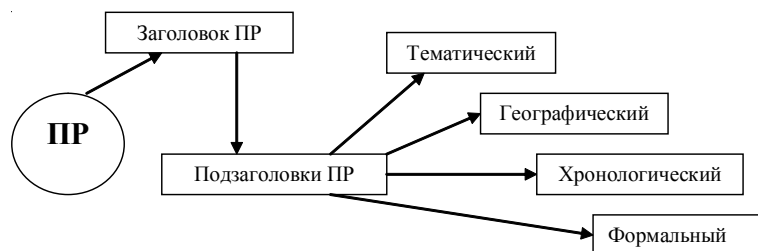
Что касается структуры Тезауруса, то она представлена двумя частями: основной – алфавитный массив (из терминов и статей к ним) и вспомогательной – ряд указателей уточняющего характера.

Алфавитный массив состоит из соответственно упорядоченных дескрипторов и аскрипторов. Имеют место и словарные статьи, отражающие предмет в лаконичной форме, что обеспечивает унификацию отображения информационного содержания документов и запросов [2]. Они могут иметь примечания (энциклопедического, уточняющего, комплексного и шаблонного характера), которые объясняют значение дескриптора на языке индекса.

Отдельно следует отметить и условные обозначения, используемые для оптимизации и конкретизации информационно-поисковой работы. Основные из них это «У» – означает «употребляется», т. е. символизирует о прямой дескрипторности лексической единицы; «НУ» – «не употребляется», т. е. символизирует антагонистичность дескриптора и ставится перед аскриптором; «Ш» – расшифровывается как «более широкий термин», т. е. речь идет о вышестоящем дескрипторе, обозначающем либо родовое понятие, либо целое в отношении данного дескриптора, является обозначением части этого целого; «Н» – расшифровывается как «низкий термин», т. е. нижестоящий (узкий) дескриптор, обозначающий либо видовое понятие, либо часть в отношении к понятию, что представлено вышестоящим дескриптором [9]; «А» – понимается как «ассоциативный термин», т. е. ассоциативный дескриптор, связанный с другим семантической связью, характер которой не указывается.

В Тезауусе содержатся следующие функционально-вспомогательные указатели: указатель формальных подзаголовков; указатель персоналий; указатель российско-украинских соответствий; указатель латинско-украинских соответствий; указатель учреждений и организаций; указатель географических названий.

В процессе индексирования формируется предметная рубрика (ПР) с использованием соответствующих компонентов Тезаууса для обеспечения содержательной полноты данной конкретной рубрики. Происходит это по принятой мировым сообществом схеме:



Что касается словарной статьи, содержание которой и является источником формирования ПР, то словарная статья или тезаурусная статья – это элемент лексико-семантического указателя ИПЯ, состоящий из дескриптора или нондескриптора и всей касающейся его информации [12]. То есть можно говорить о целом ряде содержательно-дополняющих данных, характеризующих тот или иной термин, под семантическим выражением которого и собраны эти данные. Чтобы проанализировать структуру и содержание тезаурусной статьи, рассмотрим следующий пример последней:

абразивные материалы¹

Вещества повышенной твердости, которые применяются в цельном или измельченном состоянии для механической обработки (шлифование, резание, полирование и тому подобное) других материалов

рос: абразивные материалы

НВ абразивы

Ш материалы

металлорежущие инструменты

¹ В целях обеспечения полной информативности дескрипторная статья и толкование её составляющих подаётся на русском языке, а не на украинском, как в оригинале (Прим. автора).

сверхтвердые вещества

Н природные абразивные материалы

синтетические абразивные материалы

А абразивные изделия

абразивные инструменты

абразивные материалы; **природные**

абразивные материалы; **синтетические**

абразивные материалы; **искусственные**

В синтетические абразивные материалы

Теперь, рассмотрев вышеприведенную тезаурусную статью детально, можно дать характеристику каждому ее структурному элементу. Так: «абразивные материалы» – не что иное, как заглавие тезаурусной статьи – дескриптор-словосочетание, под которым собраны содержательно-вскрывающие и вспомогательные элементы;

далее после дескриптора-словосочетания, указано краткое толкование данного дескриптора, где рассказывается о физических свойствах, методе, способе и цели использования;

«рос: абразивные материалы» – перевод на русский лексической единицы, которая рассматривается;

«НУ абразивы» – после соответствующего обозначения указан аскриптор (нондескриптор);

«Ш материалы; металлорежущие инструменты; сверхтвердые вещества» – после соответствующего обозначения, приведены термины более широкого значения относительно предмета, дескриптора, который рассматривается;

«Н природные абразивные материалы синтетические абразивные материалы» – после соответствующего обозначения приведены термины «узкого характера», т. е. те, которые характеризуют аспектное понимание лексического выражения рассматриваемого предмета;

«А абразивные изделия; абразивные инструменты» – после соответствующего обозначения, приведены термины ассоциативного характера, т. е. те, которые семантически сходны с рассматриваемым дескриптором;

«абразивные материалы; природные абразивные материалы; синтетические» – данные термины представлены как пермутационная ссылка²;

² Ссылки, в которых в алфавитном порядке перечислены все отдельные слова – компоненты словосочетаний, обозначающих дескрипторы и для каждого из них указаны все дескрипторы, в состав которых входят эти слова [9, с. 5].

«абразивные материалы; искусственные» – данные термины представлены как ссылка на дескриптор;

«В синтетические абразивные материалы» – после соответствующего обозначения, приведены термины, которые являются дескрипторной лексической единицей, которая рассматривается.

Таким образом, совокупность данных тезаурусных (словарных) статей составляет лексико-семантическую, общую часть Тезауруса.

Что касается функционально-вспомогательных указателей, используемых для облегчения и оптимизации процессов поиска и индексирования, то их, как отмечалось выше, шесть:

- указатель формальных подзаголовков, представляет упорядоченный согласно алфавиту перечень терминов, при индексировании документов, используемых в полном соответствии с названием указателя;
- указатель персоналий содержит имена выдающихся лиц, богов, мифических и общеизвестных литературных персонажей, которые формируют перечень соответственно алфавиту и при индексировании документов используются в полном соответствии с названием указателя;
- указатель российско-украинских соответствий – перечень переведённых с украинского на русский язык лексических единиц, представленных в общем разделе Тезауруса. В данном указателе все переведенные термины упорядочены в соответствии с русским алфавитом;
- указатель латинско-украинских соответствий – перечень биологических и медицинских лексических единиц, переведённых с украинского на латинский язык и представленных в общем разделе Тезауруса. В данном указателе все переведенные термины упорядочены в соответствии с латинским алфавитом;
- указатель учреждений и организаций представлен в соответствии с украинским алфавитом перечнем названий организаций и учреждений международного и национального значения;
- указатель географических названий представлен в виде алфавитного перечня названий различных географических единиц. Попутно следует отметить, что наряду с терминами узкого значения приводятся термины (географические названия) уточняющего характера, более широкого понимания.

Завершающая часть Тезауруса – указатель терминологических источников – библиографический перечень изданий научного и справочного характера, насчитывающий 133 пункта.

Подытоживая все выше сказанное, можно резюмировать: «Информационно-пошуковий тезаурус НІБУ» – это не просто издание, в котором

физически оформлен дескрипторный ИПЯ, но и издание уникально-структурированное, что дает возможность говорить о более широких его возможностях. И это закономерно, ведь в наше время постоянной терминологизации существующих научных понятий и общепотребительного языка возникают закономерные вопросы относительно типологизации информационно-поисковых словарей [5] и дискуссии относительно предметности или классификационности поиска, осуществляемого с помощью подобных словарей. По нашему мнению, терминологически вполне возможно говорить не только о предметном, но и о классификационном поиске в Тезаурусе, исходя из следующей логической формулировки, приведённой далее.

Во-первых, говоря об информационно-поисковом словаре, т. е. тезаурусе, следует подразумевать не просто сокровище или сокровищницу (с греческого *thesaurus*) и множество раскрывающих содержание единиц некоторого языка с заданной на нём системой семантических соотношений. В данном случае стоит говорить о нормативном словаре ключевых слов и дескрипторов (словарных единиц в виде слов, словосочетаний или кодов, выражающих класс условной эквивалентности, в который включены эквивалентные и близкие по смыслу слова). Предназначен тезаурус для того, чтобы выполнить информационные запросы и координатное индексирование, т. е. *индексировать* по единичному освещению ключевых слов или дескрипторов.

Во-вторых, дескриптор представляет собой лексическую единицу, выраженную информативным словом или кодом и является именем (названием) класса синонимичных или близких по смыслу ключевых слов [4]. Многозначному слову естественного языка соответствует несколько дескрипторов, а нескольким синонимичным словам и высказываниям – один дескриптор. Дескрипторы используются для координатного индексирования документов и информационных запросов с целью последующего поиска. В свою очередь, координатное индексирование документа – это процесс выражения основного содержания текста перечнем содержательных слов, выбираемых из тезауруса, т. е. будто указанием его координат. Иными словами, дескриптор выражает определенный класс, определенную рубрику, в которой освещается содержательное толкование значения данного ключевого слова. Таким образом, под дескриптором понимается не только семантическое выражение понятия, но и содержательное его выражение в рамках данного класса – информация и метаинформация.

В-третьих, общеизвестно, что библиотечно-библиографическая клас-

сификация – это система упорядоченного расположения печатных произведений (и некоторых документальных материалов) по классам, подклассам и более мелким подразделениям в логической последовательности и подчинении на основе признаков содержания, формы издания, читательского назначения и т. п. [1, с. 5–7]. То есть наличие класса и обуславливает процесс классифицирования той или иной смыслообразующей информации, содержащейся в документе. Ведь библиотечно-библиографическая классификационная система, равно как тезаурус, не что иное, как информационно поисковая система, информация в которой подразделяется на отдельные точки индексации и поиска. Учитывая то, что индекс (лат. *index* – указатель, список) – система условных обозначений (буквенных, цифровых или комбинированных) [8], вполне оправданно можно понимать индекс как дескриптор, ведь название дескриптора и является условным буквенным обозначением любого понятия.

Таким образом, можно сделать вывод, что «Информційно-пошуковий тезаурус Національної парламентської бібліотеки України» – это универсальный, нормативный, контролируемый словарь лексических единиц украинского языка, с зафиксированными семантическими отношениями. В его структуре восемь разделов, семь из которых функционально-содержательно наполнены относительно процессов компетентного выполнения поиска и индексации: общая часть (лексико-семантическое собрание терминов) и шесть вспомогательных указателей (формальных подзаголовков, персоналий, русско- и латино-украинских соответствий, учреждений и организаций, географических названий). Содержательная часть сформирована из словарных статей, представляющих лексические единицы (дескрипторы и аскрипторы) со справочной и операционной информацией о них. «Информційно-пошуковий тезаурус НПБУ» построен таким образом, что позволяет оперативно осуществлять процессы поиска, индексировать отдельные документы и получать справочную информацию, что, в свою очередь, предоставляет возможность упорядочивать, систематизировать документные массивы поступлений адекватно к структуре и содержанию основного фонда библиотеки. «Информційно-пошуковий тезаурус НПБУ» отражает уровень развития дескрипторных информационно-поисковых языков в Украине, его разработка стала толчком для создания и использования в библиотеках Украины информационно-поисковых систем на основе дескрипторных ИПЯ, разработанных в соответствии с развитием систематизации научного знания.

Список использованных источников

1. Библиотечно-библиографическая классификация : табл. для науч. б-к Гос. б-ка СССР им. В. И. Ленина. – Доп. и испр. – М. : ГБЛ, 1990. – 169 с.
2. Вилегжаніна Т. І. Інформаційно-пошуковий тезаурус / Т. І. Вилегжаніна, О. М. Збанацька ; Нац. парлам. б-ка України. – Донецьк : Альфа-Прес, 2004. – 1104 с.
3. Вилегжаніна Т. І. Лінгвістичне забезпечення для створення уніфікованого пошукового образу та пошукового запиту документа/ Т. І. Вилегжаніна, О. М. Збанацька // Електронні ресурси бібліотек : за підсумками Всеукр.наук.-практ.конф. директорів держ. та обл. унів. б-к, 14–17 жовт. 2003 р. – Кіровоград, 2003. – С. 126–134.
4. Дескриптор [Электронный ресурс]. – Режим доступа: <http://dic.academic.ru/dic.nsf/enc1p/16102>. – Загл. с экрана. – Дата просмотра: 16 янв. 2012 г.
5. Збанацька О. Проблеми класифікації інформаційно-пошукових мов / О. Збанацька // Державне управління і право : зб. наук. пр. – К., 2006. – Вип. 1, ч. 2. – С. 151–153.
6. Збанацька О. М. Інформаційно-пошуковий тезаурус НПБУ : особливості терміносистеми та використання в процесі предметизації / О. М. Збанацька // Бібліотекознавство. Документознавство. Інформологія. – 2005. – № 1. – С. 57–61.
7. Збанацька О. М. Про створення «Словника уніфікованої лексики для складання предметних рубрик Національної парламентської бібліотеки України» / О. М. Збанацька // Науково-технічні бібліотеки в єдиному інформаційному просторі України : міжнар. наук.-практ. конф. – К., 2000. – С. 119–122.
8. Индекс [Электронный ресурс]. – Режим доступа: <http://oval.ru/enc/29406.html>. – Загл. с экрана. – Дата просмотра: 16 янв. 2012 г.
9. Информационно-поисковые языки. Термины и определения : ГОСТ 7.74–96. – М., 1997. – 35 с.
10. Інформація та документація. Інформаційні мови. Терміни та визначення : ДСТУ 2398–94. – К., 1994. – 73 с.
11. Сербін О. «Рубрикатор НБУВ» : історія започаткування та структурний аналіз / Олег Сербін // Вісн. Кн. палати. – 2008. – № 1. – С. 17–20.
12. Стандарты по библиотечному делу : сборник / сост. Захарчук Т. В., Петрова Л. И., Завадская Т. А., Зусьман О. М. – СПб. : Профессия, 2000. – 512 с.
13. Таблиці територіальних типових поділів / І. І. Багрій, А. Г. Бровкін (кер.), Л. В. Головіна, І. А. Зарічняк (відп. за вип.) ; НАН України, Нац. б-ка України ім. В. І. Вернадського ; голов. ред. О. С. Онищенко. – К. : НБУВ, 2004. – 769 с.
14. Універсальна десяткова класифікація : у 2 кн. : пер. з англ. Кн.1. Таблиці / голов. ред. М. І. Сенченко ; USD Consortium ; Кн. палата України. – К. : Кн. палата України, 2000. – 932 с.