

ОСОБЕННОСТИ ПОИСКА НАУЧНОЙ ИНФОРМАЦИИ В ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМАХ СЕТИ ИНТЕРНЕТ И КАТАЛОГАХ БИБЛИОТЕК. ИНФОРМАЦИОННО-ПОИСКОВЫЕ ТЕЗАУРУСЫ КАК ЭФФЕКТИВНОЕ СРЕДСТВО ПОИСКА ИНФОРМАЦИИ

¹Козлова Л. С., ²Бондаренко С. Н.

¹ГУ «Белорусская сельскохозяйственная библиотека НАН Беларуси»,
Минск, Беларусь, *lylenok@tut.by*,

²Белорусский национальный технический университет,
Минск, Беларусь, *bosn@)mail.ru*

Аннотация. Описаны проблемы поиска научной информации с использованием интернета и каталогов научных библиотек. Анализируются нвучные каталоги, особенности представления и организации поиска информации в них.

Одним из первостепенных требований для проведения научных исследований является умение находить необходимую информацию, адекватную научному запросу, с целью ее дальнейшей обработки, интерпретации и успешного использования в работе. В настоящее время существует большое количество возможностей поиска информации в различных информационно-поисковых системах: в сети интернет, в различных справочно-информационных ресурсах (как печатных, так и онлайн-ресурсах), а также в электронных каталогах библиотек. Необходимо отметить, что, несмотря на развитие глобальной сети интернет и наличия информационных ресурсов в свободном доступе в сети, библиотеки остаются и в настоящее время основными центрами получения необходимой информации. Связано это в первую очередь с тем, что библиотеки, особенно отраслевые, тщательно отбирают информацию в соответствии с ее значимостью в той или иной научной сфере, а также владеют инструментарием для ее представления и поиска.

Для получения информации и успешного ее использования необходимо владение определенными навыками информационного поиска, что предполагает знание терминологии предметной сферы, в которой ведется запрос информации. То есть важным представляется владение определенным лингвистическим инструментарием – информационно-поисковыми языками, необходимыми для формирования корректного информационного запроса в поисковой системе.

Информационно-поисковый язык (ИПЯ) – искусственный язык, предназначенный для выражения содержания документов или запросов или описания фактов с целью последующего поиска [1]. В большинстве случаев поиск информации пользователями происходит путем задания в поисковой системе темы из какой-либо предметной области. Темы в запросе формируются чаще всего с помощью лексических единиц (терминов), то есть используются вербальные искусственные языки на основе лексических единиц естественного языка.

Важным представляется вопрос, насколько корректны термины, применяемые для поиска, и как они соотносятся с нормированными терминами отдельных вербальных лингвистических инструментов, применяемых для индексирования (содержательной обработки) информационных ресурсов и представления их в информационно-поисковой системе.

Для того, чтобы необходимая информация была найдена и использована, информационные ресурсы должны пройти содержательную обработку с целью систематизации и представления информации о ресурсе в поисковой системе (в большинстве библиотек – в электронном каталоге). Процесс содержательной обработки называется индексированием. Индексирование – это процесс перевода содержания документов, запросов, фактов с естественного на информационно-поисковый язык [1]. Процесс происходит с использованием терминов индексирования (дескрипторов, ключевых слов, индексов и т. д.), в результате чего создается поисковый образ документа. Поисковый образ документа, в свою очередь, представляет собой совокупность ключевых элементов документа, по которым возможно осуществить его поиск в системе. Ключевые элементы включают в себя заглавие документа, автора, выходные данные, а также данные смысловой обработки: ключевые слова, индексы классификационных систем, тематические рубрики.

Несмотря на то, что одним из инструментов обработки и систематизации информации являются индексы классификационных систем (сюда относятся, например, индексы Универсальной десятичной классификации, индексы ББК – библиотечно-библиографической классификационной системы и т. д.) поиск по индексам относится скорее к более продвинутому, профессиональному поиску, когда специалист в полной мере владеет данными системами. Также такой вид поиска может оказать существенную помощь при нахождении информации в международных информационно-поисковых системах, когда поиск с помощью вербальных средств недоступен из-за незнания иностранных языков. В большинстве же случаев поиск информации осуществляется в электронных каталогах и других поисковых системах с помощью вербальных информационно-поисковых языков на родном языке пользователя. Например, при проведении анкетирования среди пользователей РНТБ было выявлено, что большинство из них при поиске информации предпочитает тематический поиск по УДК [3]. Здесь необходимо отметить, что значимое количество необходимой для исследований информации остается неиспользованным из-за неправильной формулировки поискового запроса. Поисковый запрос – это обычно одно слово или словосочетание, термин или несколько терминов, которые вводятся в строку поиска информационно-поисковой системы. Далее система автоматически обрабатывает заданную тему и предоставляет результаты поиска, содержащие термины запроса. В поисковых системах сети интернет для оказания помощи пользователю в формировании более точного запроса используются автоматические подсказки, которые выпадают в поисковом окне, например, слово «холодильники»: холодильники устройство, холодильники устойчивость, холодильники поставки и т. д. Данные подсказки основаны на частоте формулируемых запросов большинства пользователей и облегчают поиск необходимой

информации, а также освобождают от необходимости анализа огромного количества поисковых результатов с целью нахождения действительно необходимых сведений, что позволяет оптимизировать рабочие процессы, осуществляемые пользователем. При поиске информационных ресурсов в электронных каталогах библиотек пользователю приходится соотносить поисковый запрос с терминами индексирования, принятыми в библиотеке для обработки документов. Так как первостепенная задача библиотек – качественное информационное обслуживание пользователей, то основная цель обработки документа – это его грамотный поисковый образ, максимально релевантный поисковому запросу. Поэтому для научной обработки ресурсов в большинстве библиотек используют нормированные термины – то есть термины, которые признаны наиболее корректными и употребимыми в той или иной предметной области. Для качественного поиска информации пользователям необходимо освоить правила формирования правильного поискового запроса, а также навыки использования лингвистических инструментов, которые применяются для кодирования информации и соответственно, могут быть применены для поиска. Сюда относятся отраслевые тезаурусы и специализированные терминологические словари, авторитетные базы данных, содержащие специализированные термины (ключевые слова). Представленная в таких базах данных лексика является нормированной. Осуществляется анализ каждого термина, описание его с точки зрения корректности, частоте употребления в научной отрасли, а также приводятся связи с синонимичными понятиями. При поиске информации от синонимичных понятий приводятся отсылки к нормированному термину, а также к документам, содержащим это понятие. Такие термины, как правило, содержатся и в специализированных предметных словарях – тезаурусах. Несколько подробнее хотелось бы рассмотреть тезаурусы в целом как вербальное средство, которое может быть использовано и как средство обработки информации, и как справочный материал, и как инструмент для поиска необходимой информации.

Тезаурус – полный систематизированный набор данных от какой-либо области знаний, совокупность терминов, описывающих предметную область с указанием семантических связей между ними. Информационно-поисковый тезаурус – контролируемый словарь терминов предметной области, явно указывающий на отношения между ними и предназначенный для информационного поиска, а также создания поискового образа документа [2].

Основной единицей тезаурусов являются термины, предназначенные непосредственно для индексирования – авторизованные термины, или дескрипторы. Синонимичные авторизованным терминам понятия называются аскрипторами и не применяются в процессе обработки ресурсов, а, следовательно, нежелательны при поиске информации. Обычно от синонимов даются отсылки к основному понятию – дескриптору: смотри, используй. В соответствии со стандартами, предписывающими структуру и содержание информационно-поисковых тезаурусов, термин, предназначенный для индексирования, представляет собой понятие в форме существительного или именной группы (словосочетания). Дескриптором является такой термин среди синонимов, который рассматривается как основной для выражения понятия предметной области. В боль-

шинстве поисковых систем при поиске информации по синонимам автоматически «подтягиваются» поисковые результаты, содержащие основной термин (дескриптор). Обычно дескрипторы формулируются однозначно, однако из-за полисемии некоторые термины нуждаются в уточнении, так как вне контекста их значение понять невозможно. В таких случаях термин, выбранный в качестве основного для кодирования информации, снабжается краткой пометой (релятором), или кратким комментарием. Например, в Информационно-поисковом тезаурусе по сельскому хозяйству и продовольствию Центральной научной сельскохозяйственной библиотеки Российской академии наук дескриптор «бычки» имеет два значения, поэтому данный термин снабжен реляторами: бычки (рыбы) и бычки (телята). В таком же виде необходимо формулировать и поисковый запрос во избежание необходимости отбора документов, относящихся к разным областям исследований – скотоводство (крупный рогатый скот) и рыбное хозяйство.

Дескрипторы могут быть представлены отдельными лексическими единицами или же словосочетанием (два или несколько слов). В качестве дескриптора выбирается наиболее употребляемый в научной сфере термин, лексическая единица, соответствующая базовому понятию предметной области.

Информационно-поисковые тезаурусы характеризуются наличием определенных семантических связей между лексическими единицами. Иерархические связи предполагают наличие вышестоящих (широких) по смыслу и нижестоящих (более узких) терминов. Это, например, отношения род-вид, целое – часть.

Ассоциативные связи – связи между терминами близких предметных областей. Зачастую ассоциативные термины оказывают немалую помощь в создании поискового образа документа и непосредственно в поиске информации, так как дают отсылки к множеству понятий, связанных непосредственно с темой поиска, и таким образом существенно расширяют поисковый диапазон, выполняя функцию «помощника» для нахождения необходимой информации в поисковом пространстве.

Таким образом, показано, что существует достаточное количество лингвистических средств, которые позволяют в полном объеме реализовать функции не только инструментов обеспечения смысловой обработки информационных ресурсов, но и функций информационно-поисковых средств, повышающих качество информационного поиска. Одним из таких средств являются предметные информационно-поисковые тезаурусы, обеспечивающие эффективность и релевантность поиска научной информации в различных информационных системах, и в частности, в каталогах библиотек, и оказывают существенную помощь пользователям, осуществляющим научные исследования.

Литература

1. Гендина, Н. И. Лингвистические средства библиотечно-информационных технологий: учебник / Н. И. Гендина. – Санкт-Петербург: Профессия. 2015. – 440 с. – (Учебник для бакалавров).

2. Ф-т информационных технологий Новосибирского государственного университета [Электронный ресурс]. – Режим доступа: <http://www.nsc.ru/>. – Дата доступа: 18.10.2022.

3. Ванюшина, С. П. УДК в научно-технических библиотеках Беларуси / С. П. Ванюшина // Инструментарий индексатора и его применение у библиотэках Беларусі = Инструментарий индексатора и его применение в библиотеках Беларусі / Нацыянальная бібліятэка Беларусі; [склад. С. А. Пугачева; пад навук. рэд. Т. В. Кузьмініч]. – Мінск, 2016.