

МЕТОДИКА ФОРМИРОВАНИЯ КЛЮЧЕВЫХ СЛОВ: ВЗГЛЯД НА ПРОБЛЕМУ

Последнее десятилетие прошлого века было знаменательным для библиотечного дела нашей республики. Именно в этот период началась активная работа по автоматизации библиотечно-информационных процессов. Библиотеки стали создавать электронные каталоги (ЭК) на свои фонды. С ростом объема библиографических записей ЭК перед библиотеками стала проблема усовершенствования имеющейся поисковой системы. Наибольшие трудности возникали при организации тематического поиска. Разработчики автоматизированных информационных библиотечных систем (АИБС) пошли навстречу библиотечным работникам и для этих целей предложили использование лингвистического обеспечения программы, под которым понимается «применяемая в ее технологиях совокупность информационных языков, словарных баз данных, лингвистических процессов, средств ведения и использования словарей».¹ Также были предложены дополнительные поля, в которые можно вводить аннотации, названия рубрик (тематических, хронологических и др.) и ключевые слова (КС). При формировании поискового запроса можно использовать классификаторы (УДК, ББК, ДДК, ГРНТИ и др.).

Все вышеназванные варианты осуществления тематического поиска в ЭК нашли активное применение. Я не буду подробно останавливаться на достоинствах и недостатках каждого из них. В докладе особое внимание хотелось бы уделить поиску по КС. Нужно сказать, что этот вид поиска положительно оценили индексаторы и пользователи. Последним он знаком по применению в сетях Интернет, где поиск по КС становится все более необходимым лингвистическим средством документального поиска.

В настоящее время индексирование КС стало обязательным элементом технологии библиотечно-информационных систем. В то же время этот вид индексирования является ахиллесовой пятой для библиотек, так как явно недостаточно методических материалов, если не сказать, что они практически отсутствуют. Естественно, что библиотеки пытаются найти выход из создавшегося положения, разрабатывая методики по формированию КС на основании своего опыта работы, и очень многие из них считают свои убеждения непогрешимыми.

Хотя это вопрос спорный, никто пока не провел исследования предпочтений пользователя при составлении тематического поискового предписания в библиографических базах данных. К тому же возникает вопрос о том, насколько эти методические решения отвечают общим тенденциям?

¹ Зайцева Е.М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития / Е.М. Зайцева // Науч. и техн. б-ки. – 2000. – № 3. – С. 54

Все это говорит о том, что разработка общего методического решения по формированию КС очень актуальна. Количество библиографических записей в создаваемых библиотеками ЭК растет с каждым днем. И стоит задуматься, насколько востребованы будут наши ЭК при предстоящей интеграции в мировое информационное пространство. Нужно заметить при этом: чем больше проходит времени, тем масштабнее будет редактирование, без которого не обойтись. О его необходимости говорит то, что в данный момент при индексировании документа КС большую роль играет субъективный фактор, т.е. степень раскрытия содержания документа прямо зависит от уровня квалификации индексатора, его знакомства с текстами отрасли, ее проблематикой. В основном выбор КС и форма их представления определяется по наитию и не связана ни с какими правилами. Анализ результатов индексирования документов КС в наиболее успешно занимающихся автоматизацией массовых библиотеках Москвы, Перми и др. городов России показал, что при субъективном подходе одна и та же книга, проиндексированная разными сотрудниками даже одного и того же отдела, содержит совершенно несопоставимые поисковые образы документов (ПОД). Небольшой пример из опыта работы нашей библиотеки. Библиотека ВГУ им. П.М. Машерова заимствует у библиотеки Витебского государственного технологического университета (ВГТУ) БД по экономике. Подход к составлению КС у нас разный и прежде чем данную БД влить в наш ЭК, мы вынуждены проводить редактирование поля «ключевые слова». Тем же занимаются и сотрудники библиотеки ВГТУ, заимствующие у нас БД по педагогике. Я думаю, что похожие ситуации типичны для библиотек нашей республики, да и библиотек бывшего СНГ. Это заставляет серьезно задуматься. Необходимо искать выход из создавшейся ситуации.

В большинстве случаев недостатки индексирования выявляются при использовании ЭК преимущественно при обслуживании пользователей. Чаще всего оказывается, что созданный ЭК низкого качества и мало пригоден при тематическом (особенно многоаспектном) поиске.

Учитывая, что в последнее время много говорится о корпорации библиотечно-информационных систем, хотелось, чтобы в первую очередь был разработан документ, который бы обеспечил единый подход к координатному индексированию входящего потока документов.

Ежедневно занимаясь индексированием, редактированием и введением в ЭК поступающих документов, библиотекари всякий раз обдумывают стратегию отбора и написания КС. В 2002 г. библиотека ВГУ им. П.М. Машерова приняла решение о разработке методических рекомендаций по формированию КС. Мы тщательно изучили все публикации по волнующей нас проблеме, обратились за консультационной помощью к библиотекам, которые успешно работают в данном направлении. Мы очень благодарны за оказанную помощь сотрудникам Национальной библиотеки РБ, Зональной научной библиотеки Саратовского государственного университета, библиотеке Кемеровского университета культуры. Учитывая весь изученный материал и опыт работы сотрудников нашей библиотеки, а

также возможности программно-информационного обеспечения, мы разработали «Методику представления КС для ЭК библиотеки ВГУ им. П.М. Машерова». Мы не считаем ее непогрешимой и придерживаемся мнения, что нужно как можно чаще публично обсуждать разные методики координатного индексирования и искать, в конечном счете, золотую середину.

Я не буду подробно излагать все пункты нашего документа, т.к. для этого нужно много времени. Остановлюсь лишь на некоторых из них. Прежде чем рассказать о нашем подходе к составлению КС, хочу оговорить, что поле КС, за отсутствием у нас некоторых тематических рубрик включает географические понятия, персоналии, хронологические данные. Но с внедрением программы АИБС «МАРК - SQL» у нас будут применены дополнительные рубрики и подрубрики и вышеназванные данные уйдут в эти поля.

В нашей Методике оговорены основные подходы к анализу содержания документа, даны типовые схемы: для гуманитарных и негуманитарных наук. Индексатору рекомендуется опираться на схему смыслового анализа документа, при этом четко определяется круг вопросов, подлежащих выявлению. Это: объект исследования или описания, цели исследования, характеристика объекта, применяемые методики исследования и т.д. Оговаривается объем ПОДа, он не ограничен, но 6-8 ключевых слов считается достаточным (Крандфилдский эксперимент показал- при увеличении количества терминов с 3 до 8 эффективность поиска увеличивается с 65% до 81%). КС могут быть однословными и в виде словосочетаний.

КС формулируются, как правило, в именительном падеже. В словосочетаниях допустимо употребление косвенных падежей, если это обусловлено типом связи, но основное значимое слово должно стоять в именительном падеже. Например: библиотеки, расследование преступлений.

При формировании КС употребляется как единственное, так и множественное число. Термины, используемые в литературном языке только в единственном или только во множественном числе, соответственно используются в той же форме. Например: морковь, финансы, риск. Если термины в общелитературном языке употребляются как в единственном, так и во множественном числе, то предпочтение отдается множественному.

КС формулируются на языке документа, исключение составляют документы на иностранных языках - КС для них формулируются на русском языке.

КС даются в обычной неинвертированной формулировке. Инверсия используется в следующих случаях:

- для словосочетаний, главное слово в которых - общенаучный термин (система, закон и т. д.). На первое место в таких словосочетаниях выносится определение. Например: Медведя культ, вероятностей теория;
- словосочетаний, в состав которых входит имя собственное. Например: Шатлина метод, Ортеги гипотеза;

- словосочетаний, в состав которых входит изменяющееся числительное.
Например: класс 1;

Имена собственные формулируются в традиционной форме, т.е. сначала идет фамилия, а за ней следуют инициалы, после инициалов точка не ставится. Каждое имя собственное включается в ПОД с соответствующим термином из списка семантических категорий, который дает социокультурную характеристику данному лицу, а также приводятся обозначение страны и века, в которые жил данный деятель. Например: Богданович М, поэты, Беларусь, 20в.

Названия научных работ, литературных произведений, государственных, партийных и т. д. документов, являющихся объектом исследования, включаются в дескрипторный ПОД в кавычках («»). При этом авторские научные работы, литературные произведения писателей и т. д. включаются вместе с фамилией и инициалами авторов на первом месте. Например: Толстой Л Н «Война и мир».

Названия газет и журналов включаются в дескрипторный ПОД в кавычках, на первое место выносится обозначение вида документа (газета, журнал). Например: газета «Труд».

В Методике даны рекомендации по представлению в дескрипторном ПОДе административно-территориальных и географических наименований, названий исторических событий, наименований организаций, учреждений и общественных объединений, сведений о музеях, хронологических данных и др.

Предлагаю всем, кто заинтересовался «Методикой представления КС для ЭК библиотеки ВГУ им. П.М. Машерова», непосредственно ознакомиться с документом. Нам будет интересно ваше мнение.

В заключение хотелось бы предложить библиотекам, а тем более координаторам корпоративных проектов обратить более пристальное внимание на данную проблему, ведь в создавшейся ситуации единственно правильным будет ее коллегиальное решение. В отношении методического объединения вузовских библиотек Витебского региона хочу сказать, что на ноябрь у нас намечен семинар по выработке единого подхода к заполнению поля «ключевые слова». Надеюсь, мы сможем принять общее решение, и это будет первым шагом к будущей единой корпоративной системе вузовских библиотек Витебского региона с возможностью обмена БД.

Список литературы:

1. Алешин Л.И. Поиск документов: анализ и новые возможности / Л.И. Алешин. // Библиография. – 2002. - №4. – С. 14 – 22
2. Визер Ж. Сам себе проверяющий, или как составлять ключевые слова / Ж. Визер // Библиотека. – 1999. - № 11. – С. 64
3. Воройский Ф. Индексирование документов в АИБС / Ф. Воройский // Библиотека. – 1996. - № 9. – С. 42 – 44

4. Демчук А.Л. Методические рекомендации по заполнению тематических полей АИБС МАРК / А.Л. Демчук. – Саратов: ЗНБ СГУ, 2002. – 4 с.
5. Жарикова Л.А. Возможности тематического поиска в электронном каталоге ГПНТБ СО РАН / Л.А. Жарикова, Г.А. Скарук, А.В. Бобко, Г.И. Сергеева // Науч. и техн. б-ки. – 2002. - № 8. – С. 30 – 38
6. Зайцева Е.М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития / Е.М. Зайцева // Науч. и техн. б-ки. – 2000. -№ 3. – С.54-57
7. Инструкция по индексированию входного потока документов для формирования массивов электронного каталога / Министерство культуры и печати РБ, Национальная библиотека Беларуси. – Мн., 1995. – 28 с.
8. Методика составления ключевых слов для электронного каталога библиотеки ВГУ им. П.М. Машерова. – Витебск, 2003. –13 с.
9. Пименов Е. О тезаурусе замолвите слово: [Лингвистическое обеспечение] / Е. Пименов// Библиотека. - 2001. - № 7. - С. 42-45
10. Щербинина Г.С. Координатное индексирование в свете интеграционных процессов / Г.С. Щербинина. – Науч. и техн. б-ки. – 2002. - № 10. – С. 72 – 76