

Вопросы работы алгоритма добавления записей в проекте «Сводный каталог периодики и аналитики по медицине»

Представлены основные вопросы и технологии построения алгоритмов – создания и проверки алгоритмических кодов, слияния библиографических записей аналитических описаний, преобразования предметных рубрик – при работе проекта «Сводный каталог периодики и аналитики по медицине» в Сибирском государственном медицинском университете.

Ключевые слова: система автоматизации библиотек ИРБИС, Сибирский государственный медицинский университет, Сводный каталог периодики и аналитики по медицине, корпоративные ресурсы, слияние аналитических описаний.

Сегодня никого не удивить работой в проекте по созданию корпоративных ресурсов. Однако немногие специалисты могут в деталях представить, как работает тот или иной проект. В этой публикации рассмотрим: технологию работы алгоритма автоматического слияния записей от библиотек – участник проекта «Сводный каталог периодики и аналитики по медицине», некоторые идеи, разработки, настройки и вопросы использования алгоритмов и программ, на основе которых созданы технологии слияния записей проекта в научно-медицинской библиотеке Сибирского государственного медицинского университета (СибГМУ, Томск).

Идеология слияния записей и построения сводного электронного каталога

Для изучения истории вопроса стоит обратиться к идее распределённой аналитической росписи периодики. В первых проектах превалировала задача полного представления журнала в виде отдельных записей статей. Такие проекты строились на базе взаимного распределения наименований периодических изданий для каталогизации между участниками и добавления записей аналитических описаний статей в некий сводный каталог. Такую модель в то время реализовать было проще. Однако оставался вопрос: «А что будет с такими аналитическими описаниями при работе системы книговыдачи?» Ведь не секрет, что для её нормальной работы необходимо описать журнал с помощью многоуровневой технологии каталогизации, т.е. должна существовать связь между библиографическими записями «Описание журнала» – «Подшивка» – «Номер журнала» – «Аналитическое описание статьи» [1].

В результате проведенных в 2002 г. исследований в СибГМУ, принято решение для проекта «Сводный каталог периодики и аналитики по медицине» не использовать существовавшую на тот момент модель взаимодействия, а постараться реализовать технологию взаимной аналитической росписи периодики таким образом, чтобы в сводном ЭК восстанавливать взаимные связи между записями. В таком случае получаемый в результате проекта ЭК начинал обладать новыми свойствами.

Технология привязки аналитических описаний статей к конкретному номеру журнала

В 2002 г. получен алгоритм восстановления связи между записями аналитических описаний и описанием номера журнала. Технология разработана на основе механизма и возможностей языка глобальной коррективы записей, реализованной на тот момент в системе ИРБИС32 [2].

Алгоритм восстановления связи между записями был построен на взаимной замене кодов в полях записи описания журнала (903) и в полях аналитических описаний статей (462^w). На коррективу было необходимо отобразить записи аналитических описаний, заранее импортированных в каталог, и определить код номера журнала, к которому нужно «привязать» эти аналитические описания.

Немного о версиях MeSH и самом рубрикаторе

Если рассматривать классификационные системы, то одной из их технологических вершин будет «Тезаурус медицинских рубрик – MeSH». С точки зрения построения классификаций целой отрасли знаний, у медицинских специалистов есть солидное преимущество: не каждая отрасль может похвастаться столь стройной классификационной таблицей терминов, рубрик и их взаимосвязей. Поэтому не секрет, что для медицинских библиотек аналогов систематизации документов такого уровня, как MeSH, сегодня нет [3].

Однако чем сложнее система, тем сложнее в ней разобраться, а тем более сложно её сопровождать. Постоянные изменения связей между рубриками от версии к версии, развитие отрасли в плане открытия новых препаратов и методик требуют постоянного изменения в MeSH. Кроме того, сам тезаурус MeSH в рамках проекта NLM (Национальной медицинской библиотеки США) ежегодно предоставляет таблицы изменений для возможности актуализировать текущее состояние предметного описания документа в библиотеках.

Реалии библиотечной отрасли, к сожалению, не позволяют библиотеками ежегодно обновлять версии тезауруса MeSH. Однако кроме обновления самого MeSH, необходимо ежегодно корректировать записи ЭК, с тем чтобы в предметных рубриках записей присутствовали ссылки и коды на последнюю актуальную версию MeSH.

**Реальные версии рубрикатора в библиотеках и особенности предметизации на их основе.
Вопрос идентификации рубрик MeSH: теория и факты.
Технология автоматического обновления кода рубрики до последней версии рубрикатора**

При разработке алгоритма восстановления связи между записями аналитических описаний и описанием номера журнала было допущено следующее предположение: в библиотеках – участницах сводного каталога могут быть установлены разные версии MeSH. Поэтому одним из условий нормальной работы «Сводного каталога периодики и аналитики по медицине» стало автоматическое преобразование в записи рубрики предыдущей версии MeSH в состоянии, когда имеется рубрика последней версии MeSH.

Были изучены модели работы сотрудников и механизмы предметизации с точки зрения технического специалиста. Подмечен один факт. Предметизация (присвоение конкретной рубрики MeSH) идет от термина, который рождается в голове у предметизатора на основе анализа документа, личного опыта и знаний. Основываясь на этом, ищется рубрика в рубрикаторе MeSH и затем в автоматизированную библиотечную систему добавляется ссылка на конкретную рубрику с её кодом, описанием, условиями применения (квалификаторами) и прочими атрибутами.

Таким образом, решение проблемы обновления кодов рубрик строится на предположении, что запись в БД останется навсегда, имея конкретную рубрику с её буквенным описанием в полях предметных рубрик. И на основании буквенного значения рубрики, неважно на каком языке, в последней версии тезауруса MeSH находится последний современный вариант этого названия рубрики, и в соответствии с ней ставится новый код или коды рубрики из последней актуальной версии MeSH.

Дополнительные рубрики «на все случаи жизни». При импорте записей от библиотеки – участницы сводного каталога, кроме контроля на ошибки (о нем сказано ниже), необходимо добавлять рубрики такие, как «Человек», «Взрослые», «Дети» и пр. Поскольку невозможно автоматически определять, какую из подобных рубрик добавлять при импорте записи, принято решение идти от обратного, т.е. в каждую запись добавлять все рубрики, а специалист при импорте удалит ненужные, по его мнению.

Автоматическое исправление типичных ошибок при вводе данных

Первый опыт работы «Сводного каталога периодики и аналитики по медицине» ещё раз подтвердил: люди – не идеальны. Встречались ошибки и неточности в описаниях и при вводе данных. Частые ошибки, например пробел в конце поля, в дальнейшем могли повлиять на работоспособность всей технологии – через ошибки в полях связи. Такие ошибки приводили к невозможности проверить на равенство значения рубрики для

поиска её аналога в последней версии тезауруса, что в свою очередь приводило к ошибкам отображения рубрик. Было принято решение разработать алгоритм удаления ошибочного символа «последний пробел» во всех полях. Кроме того, в записи аналитических описаний добавлялись данные о языке и правилах каталогизации, типе документа – заполнять эти поля во многих библиотеках забывали.

Разница в библиографических записях с рубриками MeSH системы ИРБИС32 и ИРБИС64. Методы устранения противоречий

Стабильная работа отстроенного механизма слияния записей для «Сводного каталога периодики и аналитики по медицине», построенная на системе ИРБИС32, была нарушена с появлением ИРБИС64, где применена новая структура отдельных меток для полей предметных рубрик MeSH, а также логика сохранения в записях БД только кода рубрики MeSH с ошибочным предположением, что этот код постоянен во времени. Кроме того, в первых версиях MeSH для ИРБИС64 не было возможности добавлять в рубрику не только её русское название, но и английский аналог.

Многие годы я модифицировал тезаурус MeSH, поставляемый в системе ИРБИС64, до состояния, когда он соответствовал MeSH «как в ИРБИС32», поскольку в том формате с ним было удобнее работать. Модификация проводилась на основе данных, получаемых напрямую из NLM. Это позволяло не только получать новый вариант MeSH для библиотек, но и восстанавливать оригинальные названия (английские термины), примечания, связи между английскими синонимами и др.

Стоит сказать, что сегодня в технологической структуре, построенной для функционирования «Сводного каталога периодики и аналитики по медицине», используются восемь вариантов тезауруса MeSH, имеющих сложную систему взаимосвязей и позволяющих проводить восстановление рубрики в актуальном варианте, даже в том случае, если в записи присутствует только название рубрики на русском или английском языке либо запись была предметизирована с использованием бумажной версии MeSH.

Интегрирование сводного каталога с авторитетной БД авторов и вопросы идентификации записей ЭК и авторитетной БД АТНРА

Развитие «Сводного каталога периодики и аналитики по медицине» [4, 5] в 2004 г. можно рассматривать как процесс построения комплекса баз данных, связанных ссылками между записями разных БД. Задача обеспечения взаимной неизменности кодов связанных записей из разных БД намного сложнее, чем для одной БД, поскольку механизмы формирования таких кодов и отслеживания изменений в системе ИРБИС не поддерживаются. В данном случае были изменены алгоритмы формирования алгоритмического кода записи в БД АТНРА, с тем чтобы на этапе получения авторитетной записи можно было сформировать ссылку, неизменяемую в будущем во всех записях системы ИРБИС.

Алгоритмический код записи строится на основе обязательных полей, без использования генераторов случайных чисел, что позволяет добиться равенства кода записи на одну и ту же персону в авторитетном файле вне зависимости от библиотеки, где создана запись. Кроме того, такой подход к формированию кодов авторитетных записей позволяет восстанавливать ссылки в записях документов, созданных без использования авторитетной БД АТНРА.

Влияние данных технологий и результатов работы с пользователями на политику в области каталогизации и описания ресурсов в библиотеках

Использование корпоративных технологий получения сводных каталогов позволяет не только существенно экономить средства при создании ресурсов, но и сделать колоссальный рывок в области гармонизации стандартов и развития документации в совместной деятельности, стандартизации, управлении механизмами системы качества. Сегодня ландшафт медицинских библиотек России и сопредельных независимых государств невозможно представить без применения корпоративных технологий создания и использования ресурсов.

Вывод

Автором сделана попытка описания работы алгоритма по слиянию записей в базу данных «Сводного каталога периодики и аналитики по медицине», а также раскрыть моменты, повлиявшие на те или иные

решения, описать текущее состояние работы. В дальнейшем улучшение видится в применении полученных разработок для использования в дистрибутивной поставке систем ИРБИС32 и ИРБИС64, поскольку уже сегодня сложилось постоянное сообщество медицинских библиотек, которые, используя унифицированные настройки, могут получить дополнительные преимущества в работе с пользователями.

Надеюсь, что после прочтения этой публикации у пытливого читателя возникнет желание не только глубже разобраться в проблеме, но и предложить новые идеи, которые позволят библиотекам стать еще более востребованными и современными.

СПИСОК ИСТОЧНИКОВ

1. **Мешечак Н. А.** Опыт распределенной многоуровневой каталогизации периодических издания в системе ИРБИС / Н. А. Мешечак, А. С. Карауш // «Информационные технологии, компьютерные системы и издательская продукция для библиотек» : докл. и тез. докл. – Москва : ГПНТБ России, 2003. – С. 169–172.
2. **Карауш А. С.** Автоматизированная технология создания полей связи в Системе ИРБИС // «Информационные технологии, компьютерные системы и издательская продукция для библиотек» : материалы Междунар. конф. «LIBCOM-2002». – Москва : ГПНТБ России, 2002. – С. 120-121.
3. **Мешечак Н. А.** Особенности каталогизации и поиска с использованием внешнего тезауруса MeSH / Н. А. Мешечак, О. С. Колобов, А. С. Карауш // «Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса» : докл. и тез. докл. – Москва : ГПНТБ России, 2004. – 1 CD-ROM.
4. **Мешечак Н. А.** Медицинский авторитетный файл «Медики России» – новый проект для медицинских библиотек : [Электронный ресурс] // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : докл. и тез. докл. – Москва : ГПНТБ России, 2007. – 1 CD-ROM.
5. **Мешечак Н. А.** Интеграция авторитетной базы данных «Медики России» в электронный каталог НМБ СибГМУ / Н. А. Мешечак, О. С. Колобов, Л. А. Шамардина, А. С. Карауш // Науч. и техн. б-ки. – 2006. – № 7. – С. 52–56.