

**Вопросы работы алгоритма добавления записей проекта
«Сводный каталог периодики и аналитики по медицине»**

**Questions of the Activity of the Append Algorithm
of the «Union Catalog of Periodicals and Analytics on Medicine» Project**

**Питання роботи алгоритму додавання записів
у проєкті «Зведений каталог періодики та аналітики з медицини»**

A. C. Karaush

*Муниципальная информационная библиотечная система, Томск, Россия
Научно-медицинская библиотека Сибирского государственного
медицинского университета, Томск, Россия*

Alexander Karaush

*Municipal Information Library System, Tomsk, Russia
Siberian State Medical University. Scientific Medical Library, Tomsk, Russia*

O. C. Karaush

*Муниципальная информационная библиотечная система, Томск, Россия
Научно-медицинская библиотека Сибирского государственного
медицинского университета, Томск, Россия*

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

The paper presents the basic questions and technologies of elaborating algorithmic codes algorithms, algorithms of the union of bibliographic records of analytical descriptions, and subject heading transformation algorithms while applying the «Union Catalog of Periodicals and Analytics on Medicine» project in the Siberian State Medical University in Tomsk.

Представлені основні питання і технології розробки алгоритмів створення та перевірки алгоритмічних кодів, алгоритмів об'єднання бібліографічних записів аналітичних описів, алгоритмів перетворення предметних рубрик під час реалізації проєкту «Зведений каталог періодики та аналітики з медицини» в СибДМУ м. Томська.

Сегодня никого не удивит работой в проекте по созданию корпоративных ресурсов. Однако немногие специалисты могут в деталях представить то, как работает тот или иной проект. В данной публикации необходимо остановиться на технологии работы алгоритма автоматического слияния записей от библиотек-участниц «Сводного каталога периодики и аналитики по медицине».

Ниже приведены некоторые идеи, разработки, настройки и вопросы использования алгоритмов и программ, на основе которых были созданы технологии слияния записей проекта «Сводного каталога периодики и аналитики по медицине» в научно-медицинской библиотеке Сибирского государственного медицинского университета (г. Томск).

1. Идеология слияния записей и построения сводного электронного каталога

Для изучения истории вопроса стоит обратиться к самой идеи 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 добавлении записей аналитических описаний статей в некий сводный каталог. Такую модель в то время реализовать было проще. Однако, всё же, оставался вопрос: «А что будет с такими аналитическими описаниями при работе системы книговыдачи?» Ведь не секрет, что для нормальной работы системы книговыдачи, 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 году исследований в СибГМУ, было принято решение для проекта «Сводного каталога 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 электронном каталоге восстанавливать взаимные связи между записями. В таком случае получаемый в результате проекта электронный каталог начинал обладать новыми свойствами.

2. Технология привязки аналитических описаний статей к конкретному номеру журнала

В 2002 году был получен алгоритм восстановления связи между записями аналитических описаний и описанием номера журнала. Технология была разработана на основе механизма глобальной корректировки записей и возможностям языка глобальной корректировки, 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 корректировку необходимо было отобразить записи аналитических описаний, заранее импортированных в каталог и определить код номера журнала, к которому необходимо «привязать» эти аналитические описания.

3. Немного о версиях MeSH и самом рубрикаторе словами технического специалиста

Если рассматривать классификационные системы, то одной из технологических вершин таких систем будет «Тезаурус медицинских рубрик – MeSH». С точки зрения построения классификаций целой отрасли знаний у медицинских специалистов есть солидное преимущество. Ведь 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 этого, сам Тезаурус MeSH в рамках проекта NLM (Национальной медицинской библиотеки) США ежегодно предоставляет таблицы изменений для возможности библиотек актуализировать текущее состояние предметного описания документа.

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

4. 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, что 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. Автором были изучены модели работы сотрудников и механизмы предметизации с точки зрения технического специалиста. И был 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. В дальнейшем на основании этого ищется рубрика в рубрикаторе MeSH, затем в автоматизированную библиотечную

систему добавляется ссылка на конкретную рубрику с её кодом, описанием, условиями применения (квалификаторами) и прочими атрибутами.

Таким образом, решение вопроса по обновлению кодов рубрик строится на предположении, что запись в базе данных останется навсегда, имеющая 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 ставится в соответствии с ней новый код или коды рубрики из последней актуальной версии MeSH.

5. Дополнительные рубрики «на все случаи жизни»

При импорте записей от библиотеки-участницы сводного каталога кроме контроля на ошибки, о котором рассказывается дальше, необходимо добавлять рубрики, такие как «Человек», «Взрослые», «Дети» и пр. Поскольку невозможно автоматически определять какую из подобных рубрик добавлять при импорте записи, то было принято решение идти от обратного. Т.е. в каждую запись добавляются все рубрики, а специалист при импорте удаляет ненужные, по его мнению.

6. Автоматическое исправление типичных ошибок при вводе данных

Как показал первый опыт работы «Сводного каталога периодики и аналитики по медицине», люди – не идеальны. Встречались ошибки в описаниях, неточности и ошибки ввода данных. Часто встречающиеся 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 по удалению ошибочного символа «последний пробел» во всех полях.

Кроме этого в записи аналитических описаний добавлялись данные о языке и правилах каталогизации, типе документа, при заполнение этих полей во многих библиотеках забывали.

7. 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 выходом Системы «ИРБИС64», в котором была применена новая структура отдельных меток для полей предметных рубрик MeSH, а также логикой сохранения в записях базы данных только кода рубрики MeSH с ошибочным предположением, что этот код постоянен во времени. Кроме этого в первых версиях MeSH для «ИРБИС64» не было возможности добавлять в рубрику не только ее русское название, но и английский аналог.

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

Стоит сказать, что сегодня в технологической структуре, построенной для функционирования «Сводного каталога периодики и аналитики по медицине», используется 8 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, или запись была предметизирована с использованием бумажной версии MeSH.

8. Интегрирование сводного каталога с авторитетной базой данных авторов и вопросы идентификации записей электронного каталога и авторитетной базой данных авторов АТНРА

Развитие «Сводного каталога периодики и аналитики по медицине» [4, 5] в 2004 году можно рассматривать, как процесс построения комплекса баз данных, связанных ссылками между записями разных баз данных. Данная задача – обеспечение взаимной неизменности кодов связанных записей из разных баз данных, намного сложнее, чем для одной базы данных, поскольку механизмы формирования таких кодов и отслеживания изменений в Системе «ИРБИС» не поддерживаются. В данном случае были изменены алгоритмы формирования алгоритмического кода записи в базе данных 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 этого, такой подход к формированию кодов авторитетных записей позволяет восстанавливать ссылки в записях документах, созданных без использования авторитетной базы данных АТНРА.

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

Использование корпоративных технологий получения сводных каталогов позволяют получить не только существенную экономию средств при создании ресурсов, но и сделать колоссальный рывок в области гармонизации стандартов и развития документации в области 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, а также описать текущее состояние работы в этом направлении. В дальнейшем направления работы по улучшению видятся 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 этого документа у пытливого читателя возникнет желание не только подробнее разобраться, но и предложить новые идеи, которые позволят библиотекам стать еще более востребованными и современными.

Литература

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»: Доклады и тез. докладов. – М.: 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 МК «LIBCOM-2002». – М.: 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»: Доклады и тез. докладов. – М.: 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 : доклады и тез. докладов. – М. : 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. 52–26