

Международный библиотечный опыт в решении проблем многоязычия

Вопрос о том, какая библиотечная проблема является самой сложной, наверное, некорректен. И время, и место, и многие другие обстоятельства существенно меняют степень сложности проблемы и, соответственно, способы ее разрешения. Однако в отношении многоязычия мнения сходятся — это действительно одна из наиболее трудно решаемых проблем. Причем многие библиотеки просто закрывают на нее глаза, ссылаясь на отсутствие необходимых ресурсов — людских и финансовых — и, к сожалению, в какой-то степени они правы. Потому что попытки решить вопросы многоязычия иногда лежат вне сферы возможностей библиотекарей, вспомним те недавние времена, когда не было Unicode, а соответственно и возможности использовать символы отдельных языков. И лишь технический прогресс дал библиотекам реальный шанс представить свои данные, в частности, библиографические, в мировой сети. Но и до сих пор, символы не всех языков и типов письма (например, глаголицы) представлены в Unicode, равно как и не все библиотечные системы (АБИС) его поддерживают. Но если с представлением символов разных языков и их отображением человечество худо-бедно справилось, то проблема поиска в многоязычном Интернете будет стоять перед ним еще долго.

Поиск релевантной информации — особая проблема в контексте многоязычия. Для начала определимся с тем, что такое многоязычный поиск. Многоязычный поиск — это процесс, при котором запрос, сделанный на одном языке, дает возможность получить релевантный ответ на других языках. Т.е. пользователь, набравший поисковый запрос, скажем, на татарском, получает результат на русском, башкирском, английском и т.д. Можно с точностью сказать, что такие возможности практически отсутствуют. Пользователь сможет получить сведения лишь о тех документах (по их библиографическим записям), ко-

торые содержат предметные рубрики, или аннотации, или нотации классификационной системы на нескольких языках. В российских библиотеках эта практика отсутствует. Я знаю, правда, о немногочисленных случаях перевода названия документа с одного языка на другой. Но название не всегда точно отражает содержание документа, поэтому и результаты поиска едва ли совпадут с потребностями пользователя.

Что же делают библиотеки мира, чтобы хотя бы в каком-то приближении решить проблему многоязычия? Начнем с самого близкого Российской государственной библиотеке (РГБ) проекта Европейской библиотеки.

24 августа 2006 г. РГБ и Европейская библиотека подписали рассчитанный на три года контракт о совместной деятельности и, следовательно, о полном членстве в составе Европейской библиотеки. 15 января 2008 года электронный каталог РГБ (ОРАС) стал доступен с портала Европейской библиотеки и открыт для поиска всем гражданам Европы и мира в целом.

Создание Европейской библиотеки было инициировано Европейской Комиссией в 2003 г. В марте 2005 г. Европейская библиотека начала свою реальную работу. Сегодня Европейская библиотека — это совместный некоммерческий проект 47 национальных библиотек Европы (из которых 32 национальных библиотеки являются полными участниками Европейской библиотеки и 15 позиционируются в качестве будущих участников), являющихся членами CENL (Конференции директоров национальных библиотек Европы). Целью CENL является углубление и укрепление роли национальных библиотек в Европе. Философией проекта «Европейская библиотека» является предоставление распределенного доступа к цифровым документам, физически хранящимся на серверах отдельных национальных библиотек Европы за счет механизмов единого доступа. К настоящему моменту Европейская библиотека предоставляет доступ к 150 миллионам записей по всей Европе на 20 языках. На конец 2007 г. зарегистрировано более 500 000 обращений к сайту Европейской библиотеки. Офис Европейской библиотеки находится в Нидерландах.

Европейская Комиссия в своих рекомендациях по вопросам оцифровки и онлайн-доступу к документам, имеющим историко-культурную значимость, а также по вопросам сохранности цифровой информации предложила использовать Европейскую библиотеку в качестве отправной точки для создания Общевропейской цифровой библиотеки. В Общевропейскую цифровую библиотеку войдут не только национальные библиотеки, но и другие библиотеки, а также музеи и архивы. CENL, в работе которой Российская государственная библиотека принимает активное участие, сделала официальное заявление в поддержку стратегии Европейской комиссии i2010. Основное направление этой стратегии заключается в продвижении открытой и конкурентноспособной цифровой экономики, а также использовании информационных технологий как способа улучшения жизни граждан.

Совершенно очевидно, что такой огромный информационный массив на разных европейских языках требует особых механизмов поиска. Самым первым шагом в реализации возможностей многоязычного поиска была виртуальная клавиатура, инсталлированная на портале и давшая возможность набора символов языков всех полноправных участников Европейской библиотеки (расширенной латиницы и расширенной кириллицы) и, соответственно, возможность осуществлять поиск библиографических записей на языке оригинала, как это принято у нас говорить.

Одновременно с этим интерфейс портала будет переведен на 27 европейских языков, для того чтобы европейские пользователи, говорящие только на своем родном языке, смогли понять структуру портала, способ поиска документов, хранящихся в электронных каталогах европейских национальных библиотек (НБ), увидели то, чем занимаются НБ Европы, как они выглядят, какие сокровища хранятся в их фондах. На 12 языков переведено Руководство пользователя порталом. В стадии перевода находятся описания коллекций НБ Европы, представленных на портале, что тоже будет способствовать ориентации пользователя в многоязычном контенте (содержании) НБ Европы.

Но, пожалуй, самым сложным проектом, за который взялись участники Европейской библиотеки, это проект MACS, кото-

рый ставит своей целью предоставить многоязычный предметный доступ к каталогам библиотек. Через механизм маппирования (сопоставления) или связки языков предметных рубрик (SHLs) пользователи, направив запрос на одном из языков (английском, немецком или французском), могут одновременно искать релевантные документы, проиндексированные на этих языках. Многоязычный поиск становится возможным благодаря связям, установленным между предметными рубриками на разных языках. Механизм связывания был разработан 10 лет назад и соотносился с единственным на тот момент стандартом по многоязычию ИСО 5964 – Руководство по созданию и разработке многоязычного тезауруса. В настоящее время проект MACS входит в свою промышленную стадию, принимая новых партнеров. Поэтому важно, чтобы профессиональное сообщество было знакомо с основными принципами, лежащими в основе проекта. А они, в свою очередь, базируются на проекте BS 8723 и Руководстве ИФЛА для многоязычных тезаурусов.

К числу основных принципов MACS относятся:

- Равенство языков и языков предметных рубрик (SHLs) и автономность каждого SHL.
- Интероперабельность достигается путем установления эквивалентности между предметными рубриками. Поисковые термины (предметные рубрики) никогда не переводятся для достижения эквивалентности с другим языком. MACS не полагается на методику использования в случаях, когда термин, используемый в языке-источнике, не имеет своего эквивалента. Эквиваленты устанавливаются между соответствующими предпочтительными терминами каждого языка.
- Эквивалентные связи воспринимаются как понятийные кластеры. Система MACS содержит соответствия, рассматриваемые как понятийные кластеры, и они идентифицируются только с помощью цифрового идентификатора. Понятийные кластеры представляют отношения терминов на разных языках предметных рубрик с отдельным понятием.

- Последовательность результатов как основа для действенности связей. Эквиваленты MACS должны позволить пользователям осуществлять поиск последовательных и полных групп библиографических записей из разных каталогов. Вопросы установления связей не ограничиваются только семантическими и культурными аспектами. Чтобы быть действенной, связь (по крайней мере, в тех случаях, где существует некоторая двусмысленность) должна быть протестирована в библиографической базе данных партнера, содержащей предметные рубрики на соответствующем языке. Существующие данные в библиографических записях подтвердят значение предметных рубрик, снимут терминологическую двусмысленность и укажут на прагматическую релевантность связей.

Центральным звеном в стратегии создания связей в проекте MACS является база данных «Интерфейс управления связями», которая служит в качестве внешней системы по отношению к авторитетным файлам предметных рубрик на разных языках. Предполагается, что каждый партнер MACS будет устанавливать связи, используя собственный язык предметных рубрик как язык-источник. Целевой язык будет выбираться в соответствии с несколькими факторами: знанием языка предметных рубрик (семантики и структуры), знанием языка и наличием словарей. Авторитетные записи целевого языка предоставят важную информацию для процесса установления связей. Рамки предметной рубрики и семантические отношения предоставят информацию для оценки качества связей.

На этапе предварительной оценки проекта MACS была разработана методология по выбору предметных рубрик из области спорта и театра. Она состояла в использовании иерархической структуры рубрики для сбора всех специфических разделов, связанных с широкой рубрикой. В проекте также использовался алфавитный подход к выбору рубрики, близкой к тем, которые были выбраны в соответствии с иерархическим подходом. Ценность методологии состояла в том, что предметизатор имел в своем распоряжении список предметных рубрик в семантической и алфавитной среде конкретной рубрики.

В интегрированной сетевой среде этот метод может использоваться для просмотра семантической и алфавитной среды конкретного целевого языка или языков. Можно просмотреть авторитетную запись на термин, являющийся кандидатом в потенциальный термин-связку, на одном или двух целевых языках, чтобы увидеть узкие рубрики (а также и широкие) и выявить термин, более соответствующий термину языка-источника. Алфавитный подход (там, где он возможен) может использоваться для сканирования связанных между собой рубрик. Он может быть полезен при выборе некоторых предметных рубрик, выраженных фразой, или рубрика + подрубрика, которые не связаны должным образом с некоторыми рубриками, просматриваемыми с помощью иерархического подхода, а также для локализации «рубрик-сирот» (не соотнесенных с первоначально просматриваемой рубрикой).

Мы не будем дальше вдаваться в тонкости этого проекта. Важно одно – 70 000 предметных рубрик на немецком, английском и французском языках связаны между собой силами четырех национальных библиотек (НБ Швейцарии, НБ Франции, Британской библиотеки, НБ Германии) и загружаются на портал Европейской библиотеки для осуществления многоязычного поиска. Россия могла бы и должна присоединиться к этому проекту, хотя бы потому, что на портал загружен весь электронный каталог РГБ, который делает присутствие русского языка и языков народов Российской Федерации вполне зримым.

Мне кажется важным затронуть еще один глобальный проект, который способствует продвижению языков народов мира в Интернет. Это всемирный проект Википедия. Пользователи очень хорошо знают эту свободную энциклопедию, значительно меньше с ней работают библиотекари. И, скорее всего, работают как пассивные пользователи, а не как редакторы или поставщики информации. Русский сегмент Википедии значительно меньше английского, французского, немецкого, польского, шведского, китайского. О языковых сегментах народов, населяющих Россию, и говорить не приходится. Какие-то зачатки этого присутствия есть (аварский, башкирский, чувашский, коми, удмуртский, татарский, бурятский языки). Многие библиотекари отмахиваются

от этого ресурса, как ненадежного, неполного, упрощенного, но ничего не делают, чтобы его изменять к лучшему, как это делают наши зарубежные коллеги. Например, НБ Германии и Чехии загрузили авторитетные файлы в свои национальные сегменты Википедии. Таким образом, помимо собственных сайтов, библиотеки имеют замечательную возможность размещать материалы на языках народов России и в этом раскрученном интернетовском ресурсе. По аналогии с Википедией можно использовать проект ВикиЗнание, который, начиная с 2002 г., создает свою базу знаний, загружая в нее статьи из энциклопедических словарей.

Интернет, Веб 2.0 и Библиотека 2.0 – реалии сегодняшнего дня. И к этой виртуальной реальности следует относиться со всей серьезностью, как по отношению к нашим традиционным профессиональным ценностям – каталогам, фондам, зданиям.

Коротко об авторах

Агаев Таймыр Джалилович, заместитель директора Национальной библиотеки Республики Дагестан им. Р. Гамзатова по научно-методической работе

Алушкина Марина Мереевна, заведующая отделом краеведения и национальной библиографии Национальной библиотеки Республики Алтай

Бальхаева Ирина Хубисхаловна, директор Национальной библиотеки Республики Бурятия

Бибоева Ирина Гамболовна, заведующая отделом краеведения Национальной научной библиотеки РСО – Алания

Богданова Елена Викторовна, заместитель министра Республики Карелия по вопросам национальной политики и связям с религиозными объединениями

Бойнич Валерий Дмитриевич, заместитель премьер-министра Правительства Республики Карелия

Бушкова Надежда Ивановна, заведующая отделом национальной, краеведческой литературы и библиографии Национальной библиотеки имени С.Г. Чавайна Республики Марий Эл

Валеева Зия Рахимьяновна, заместитель премьер-министра, министр культуры Республики Татарстан

Володина Наталья Ивановна, заместитель Председателя Кабинета Министров, министр культуры, по делам национальностей, информационной политики и архивного дела Чувашской Республики

Волженина Светлана Юрьевна, заместитель директора Государственной библиотеки Югры по научной и инновационной работе

Жуковская Людмила Николаевна, заместитель директора по библиотечной работе Государственной универсальной научной библиотеки Красноярского края

Кисловская Галина Александровна, директор по международной деятельности Российской государственной библиотеки