

*В. В. Скворцов, Начальник управления
Автоматизированных технологий Российской
национальной библиотеки. Член постоянного
Комитета IFLA по формату UNIMARC.
Руководитель Национальной службы развития
системы форматов RUSMARC.*

Единое лингвистическое обеспечение – национальная задача?

К уважаемому сообществу лингвистов я могу отнести себя только с очень большой натяжкой. Формат RUSMARC, которым я занимаюсь, как и другие форматы, иногда относят к лингвистическому обеспечению, а иногда нет. Таким образом, я могу отнести себя к лингвистам не только с большой натяжкой, но ещё и только иногда.

Так что же меня заставило взяться за заявленную тему?

Дело в том, что, с появлением автоматизированных библиотечных систем, вопрос о применении в них таких, например, лингвистических средств как ИПЯ, встал самым естественным образом.

Как должны работать ИПЯ в автоматизированных библиотечных системах?

С появлением корпоративных автоматизированных библиотечных и информационных систем также самым естественным образом встал вопрос о согласовании лингвистического обеспечения, таких систем. Причем здесь, пользуясь своей лингвистической некомпетентностью, я отнесу к лингвистическому обеспечению, также и форматы (имея в виду MARC-форматы, форматы метаданных, bib1 и т.д.) а также язык библиографического описания объекта.

Таким образом, мы выделили здесь два вопроса, касающихся лингвистического обеспечения, которые возникли **а.** с появлением автоматизированных библиотечных систем – вопрос о том, как должны работать ИПЯ в автоматизированных системах и **б.** с появлением корпоративных автоматизированных систем – таких как ЛИБНЕТ, например, – вопрос о согласовании лингвистического обеспечения, по крайней мере, в рамках корпорации

На протяжении нескольких лет я являюсь сопредседателем секции по лингвистическому обеспечению на Крымской конференции, и членом постоянного комитета соответствующей секции РБА и, соответственно, конференции РБА. Должен сказать, что два, поставленных выше вопроса или вообще не находят отражения в докладах на крупнейших в России библиотечных конференциях, или затрагиваются лишь частично.

Доклады о чём представляются на секциях? О состоянии и развитии тех или иных классификационных таблиц или предметных рубрик, о методике предметизации или систематизации с использованием тех или иных таблиц, о методиках обучения систематизаторов и предметизаторов. Это, так сказать, традиционное лингвистическое поле. Ветер автоматизации принёс новые темы – например, об автоматизированных рабочих местах систематизаторов или предметизаторов в автоматизированных библиотечных системах.

Что же касается озвученных выше двух вопросов, то, как сказал Высоцкий – не отражается она как-то в песнях, эта кампания.

Между тем, создание любой корпорации – это выработка и принятие правил взаимодействия в определённых целях, а также материальное обеспечение этих правил. Без этого – нет корпорации.

Правила работы с лингвистическим обеспечением - это часть общих правил взаимодействия.

Давайте попробуем оценить перспективы согласования таких важнейших компонентов лингвистического обеспечения, как язык библиографической записи (правила и традиции каталогизации), информационно-поисковые языки (ИПЯ) и машиночитаемые библиографические форматы.

Правила каталогизации.

Как известно, развитием и совершенствованием правил каталогизации или, скорее, правил описания объектов или ресурсов - на международном уровне занимается секция каталогизации IFLA – международной федерации библиотечных ассоциаций и организаций. С расширением спектра описываемых объектов – сейчас это может быть практически любой ресурс, на любом носителе – меняется сама философия описания. Разрабатываются новые принципы описания – например, т.н. франкфуртские принципы, строится новая модель описаний – FRBR (Functional Requirement for Bibliographic Record), создаются международные правила каталогизации ICC (International Cataloguing Codes).

Проще говоря, существует очевидная и активная тенденция, с одной стороны, обобщения подхода к описанию ресурса, с другой – на основании этого общего подхода – согласования национальных правил каталогизации разных стран.

Компьютеризация библиотек и развитие компьютерных сетей – создали необходимый базис для корпоративной работы, в том числе и для появления международных корпораций. Библиотеки и другие информационные центры в реальной перспективе становятся тружениками единого информационного поля. Таким образом, если ранее вопросы согласования общих правил каталогизации были просто темой для бесед, или для докладов, или, например, просто возможностью организовать поездку за рубеж и согласовать там с кем-нибудь что-нибудь: всё равно никто не проверит. Теперь - мы либо действительно выработываем общие правила, либо не работаем вместе.

Это понимается в IFLA, и этим вызвана активность секции каталогизации в последние, скажем, десять лет.

Теперь давайте посмотрим, как активность на самом верху библиотечного сообщества отражается в его глубинах. Сделаем это на примере уже упоминавшихся FRBR.

Модель считается вполне разработанной, в ряде стран проводились специальные исследования по поводу того, как её внедрять и на основе этих исследований создавались новые правила каталогизации. У нас такие исследования проводились видимо в ходе самой разработки новых правил. Исследования завершены, правила разработаны – и что?

Оказывается, действительно нет трудностей с внедрением терминологии FRBR. Терминология внедряется и, не зная её, уже невозможно вести профессиональный разговор с зарубежными каталогизаторами. У нас – ещё можно, однако и в России терминология FRBR введена новыми правилами каталогизации.

Что же касается самой модели – тут сложнее. Во-первых, в её понимании единство каталогизаторов нельзя назвать монолитным – особенно это касается сериальных изданий.

Во-вторых, FRBR это теоретическое построение, не дающее (и вообще говоря, не предназначенное давать) конкретных инструментов, через которые модель должна внедряться в повседневную практику.

Принято считать, что таким инструментом могла бы служить некая система авторитетных файлов. Однако, вряд ли это такая система авторитетных файлов, как мы представляем её сегодня, поскольку мотивы для её возникновения – совершенно иные, чем у ныне существующей.

В-третьих, подавляющее большинство библиотек работают в стандартных автоматизированных системах, а производители таких систем – это уже другое ведомство. Что им FRBR?

Возможно, это и не всегда так и не абсолютно так, но это общий камень преткновения.

И тогда оказывается проще посмотреть, может быть в том, что уже у нас есть, что-то соответствует модели FRBR? Ага, то-то и то-то! Значит, - мы внедрили FRBR в том-то и том-то.

В результате модель FRBR активно шагает по библиотекам мира.

Я не хочу драматизировать ситуацию, в принципе идёт нормальное развитие, просто она, эта ситуация, несколько дальше от завершения, чем может показаться на первый взгляд.

Давайте посмотрим, а что же внутри страны?

У нас есть ГОСТ 7.1-2003 на библиографическую запись и единые правила каталогизации, пока не ИСС, но, тем не менее, это реальная основа для ведения каталогизации.

У нас также есть три библиографирующие организации национального уровня: РКП, РГБ и РНБ и, соответственно, три разные практики каталогизации, порождающие разные записи на одни и те же документы.

Напомню, что мы пытаемся рассуждать в контексте работы корпоративных систем. Нет сомнения, что для эффективной работы в такой системе, должна существовать единственная запись на каждый документ.

Идея скоординировать работу трёх названных организаций, сколько помню, воплощалась в разных ипостасях: Российский центр кооперативной каталогизации (РЦКК), Центр текущей национальной библиографии в ИНТЕРНЕТ (ЦТНБИ), Центр «Либнет», но беда в том, что они (РКП, РГБ, РНБ) отчасти напоминают одинаково заряженные частицы и при первой возможности – разбегаются, каждая в свою сторону.

В настоящее время в Центре «Либнет» удалось объединить и удерживать вот уже пятый год в рамках совместного проекта две из трёх названных организаций: обе национальные библиотеки РГБ и РНБ. Результат их совместной работы – единственная консолидированная запись на каждый документ, с суммарным качеством и сокращением трудозатрат примерно в 1.5 раза.

При этом, как вы понимаете, РКП работает самостоятельно и выдаёт свои варианты записей на те же документы, только быстрее.

Из трёх критериев: быстрота появления, качество записи и уникальность записи – первый был важен всегда, значение 2-го и третьего критериев с особой силой проявляется именно при корпоративной работе, и чем шире корпорация, тем сильнее они работают.

ИПЯ.

Каких только информационно-поисковых языков не существует в России! Тут и ББК, и УДК и ДДК, и предметные рубрики РНБ, и РКП, ГЦНМБ, ГПНТБ, ИНИОН, тезаурусы, ключевые слова. Мало того, что существуют 4 типа таблиц ББК, но обе наши национальные библиотеки, по-разному работают с одними и теми же таблицами.

Можно ли всё это согласовать? Науке точно не известно.

Однако, если мы хотим работать эффективно, мы должны понимать, в какой среде работаем. Значит, согласовывать нужно.

На наш взгляд в пути согласования есть три основных вопроса, которые должны решаться последовательно:

Вопрос 1: Мы должны решить, чего, собственно, ждём от поисковой системы. Как она должна работать с ИПЯ? Например, если мы допускаем одновременное существование в записи предметных рубрик универсальных библиотек и специализированных библиотек, то, скорее всего, наша поисковая система будет работать с ними просто как с ключевыми словами. А целесообразно ли в самом деле иметь в одной записи и предметные рубрики специализированных библиотек, и предметные рубрики универсальных библиотек?

Это опять-таки зависит от того, как должен проводиться поиск.

Итак, первый вопрос – это определение оптимальной стратегии поиска и, исходя из этого, определение набора ИПЯ и условий их использования.

Вопрос 2: Построение технологии индексирования и поддержания авторитетных файлов в корпоративной системе. Иначе говоря, исходя из набора используемых ИПЯ, должны быть определены организации, ответственные за соответствующие авторитетные файлы или тезаурусы и порядок взаимодействия членов корпорации в процессе индексирования текущих поступлений.

Вопрос 3: Что делать с ретро-массивом? Тут мы не даём никаких рекомендаций за исключением одной – рассматривать это как отдельную задачу, после ответа на первые два вопроса. В этом случае, при всей сложности задачи, она, по крайней мере, имеет определённые граничные условия.

Форматы.

О форматах вообще есть смысл говорить тогда, когда речь идет о каком-либо взаимодействии. Иначе – это внутреннее дело каждой отдельной организации. Однако, если речь идёт об обмене записями, о простом заимствовании или, например, о просмотре результатов поиска при работе по протоколу z39.50 – мы обязаны определять форматы взаимодействия. В данном случае я имею в виду библиографические машиночитаемые форматы, конечно же. Современное программное обеспечение предоставляет средства для работы в различных форматах, однако, эффективной работы без потери данных можно достичь только пользуясь единым, в рамках корпорации, форматом. Этот формат должен соответствовать международным стандартам и национальным правилам каталогизации.

Такой формат в России существует – это формат RUSMARC. Формат RUSMARC, как вы знаете, это рабочий формат национального информационного и библиографического центра «Либнет», он также – рабочий формат ассоциации АРБИКОН. В своё время обе названные корпорации заявили об этом, и я считаю, что это, наряду с работой в формате RUSMARC Российской национальной библиотеки и поддержкой тогдашнего Управления библиотек МК РФ и нынешнего ФАКК, определило ситуацию с форматом в стране. Но это, так сказать, декларативная часть. А что с программным обеспечением? Сколько не говори: «RUSMARC» - без программного обеспечения во рту сладко не станет.

Сразу после того, как РНБ перешла на работу в формате RUSMARC она выставила свои записи бесплатно для заимствования. Результат эксперимента был очевиден – эти записи в то время не взял никто. Хорошо, что не было сделано поспешных выводов: эксперимент ведь для чего-то проводился.

Вообще, ситуация с программным обеспечением для библиотек в России, я бы сказал, перспективней, чем где бы то ни было в мире. Отдельная тема – США, но это - отдельная тема. Большая часть распространённых библиотечных систем в России – именно по распространённости в России - имеют свои корни непосредственно в библиотеках: это и OPAC-Global, и Руслан, и Ирбис. Следовательно, их производители знают нужды библиотек и должны бы оперативно на них реагировать.

Первое, что было сделано, вслед за упомянутой выше декларативной частью: практически все системы объявили о поддержке формата RUSMARC.

Должен сказать, что эти RUSMARCи, хотя и носили одну фамилию, были похоже друг на друга не более, чем дети разных отцов.

Таким образом, основная задача – создание основы для корпоративной работы – была, скажем прямо, недорешена. Однако, такие системы как OPAC-Global, Руслан и система «Русская книга» РНБ изначально работали и работают в согласованном RUSMARCе.

Чтобы исправить ситуацию – Национальная Служба по формату RUSMARC уже несколько лет проводит сертификацию программных продуктов на предмет соответствия их этому формату. К настоящему моменту выдано уже 9 свидетельств. Среди получивших такие свидетельства, уже упоминавшиеся OPAC-Global и Руслан, АС Библиотека-3, АС СПб Университета Культуры. В этом году, на Крымской конференции, было выдано свидетельство системе MAPK-SQL версии 1.9. Заявку на получение свидетельства подали и создатели системы Ирбис.

Сам процесс получения свидетельства включает не просто экспертную оценку существующего в системе формата, но по сути превращается в интерактивный процесс доработки формата системы до уровня требований формата RUSMARC. В результате, те, кто работает в сертифицированных системах – не должны испытывать сложностей во взаимодействии, ни с Центром «Либнет», ни в АРБИКОНЕ, ни с РНБ, ни друг с другом.

Не всё ещё решено и не всё обстоит так гладко с согласованием форматов в стране, но ситуация явно меняется к лучшему. Немаловажным фактором этого является существование единого центра, отвечающего за развитие формата RUSMARC в стране – Национальной Службы, а также - объективной необходимости согласованного формата.

Мы обсудили с вами некоторые тенденции, проблемы и задачи, возникающие при согласовании различных компонент лингвистического обеспечения.

Два слова о рекомендуемых размерах корпорации. Как говорилось выше, корпоративные решения работают тем эффективней, чем шире корпорация. Исходя из этого и оставляя за рамками рассмотрения межнациональные корпорации, максимальная эффективность корпоративных решений достигается в общенациональной корпорации. Нам представляется, что Центр «Либнет», также как и упоминавшийся выше АРБИКОН, уже сейчас – корпорации национального уровня. Не по названию, а по существу.

Однако, если вернуться к теме нашего доклада, то лингвистическое обеспечение согласовывается в рамках той и другой корпорации - по отдельности. И это уже другая проблема. И тема другого доклада.