

Майре АХО

Хельсинки, Финляндия

Библиотекарь-специалист Славянской библиотеки
Национальной библиотеки Финляндии

Член правления Финно-угорского общества и
Общества М. А. Кастрена

Многоязычие в библиотечной сфере Финляндии

Национальная библиотека Финляндии (бывшая Библиотека Хельсинкского университета) основана в 1640 г. в городе Турку (Або) одновременно с университетом, в 1707 г. получила право обязательного экземпляра всех выходящих на территории Шведского королевства изданий. В 1828 г. после пожара в Турку университет и библиотека были переведены в Хельсинки. В 1828 г. по указу Николая I библиотека получила право обязательного экземпляра всех выходящих на территории Российской империи изданий:

- Славянская библиотека (обязательные экземпляры за 1828–1917 гг.),
- Коллекции национальных меньшинств Российской эры: языки и культуры национальных меньшинств (1828–1917 гг.).

Оцифровка, микрофильмирование и консервация осуществляются в городе Миккели, в отделе, который был создан в 1990 г. В библиотеке собраны цифровая коллекция финских газет (газеты, опубликованные за 1771–1890 гг. на финском, шведском и немецком языках), микрофильмы местных газет, оцифрованные промышленные каталоги и ценники (на разных языках, несколько на русском), музыкальные записи (база данных Raita), в разделе Doria lentolehtiset – военные листовки на русском, украинском, финском, шведском, немецком и эстонском языках. В рамках кампании «Сохрани книгу» любой участник может предложить редкую книгу для оцифровки и стать спонсором этого процесса. Результаты можно увидеть в Интернете,

Славянская библиотека

Ситуация с русскими материалами в Финляндии значительно отличается от ситуации в других скандинавских странах благодаря продолжительной, 110-летней общей истории с Россией.

В сфере библиотечного обслуживания наиболее важным для нас был указ, в котором говорилось, что типографии обязаны предоставлять по одной копии

каждого издания Хельсинкскому университету. Благодаря профессору Якову Гроту этот указ неукоснительно выполнялся, в результате чего была собрана исключительно ценная коллекция книг, интересных для исследователей российской истории и культуры. Коллекция систематически пополнялась с 1827 по 1917 г.

Недавно нам представилась возможность закрыть лакуны в собрании: Россия согласилась заплатить долги прежнего СССР Финляндии. Славянская библиотека получила много журналов, детские литературные и провинциальные материалы в тысячах микрофильмов, две с половиной тысячи книг периода перестройки и в дополнение ко всему этому серьезное онлайн-обслуживание, которое используется в двадцати городах Финляндии.

В собрании Славянской библиотеки приблизительно полмиллиона томов, 600 редких книг, книги художников-авангардистов, старые и новые карты, эмиграционная литература, например все выпуски газеты «Руль» и большое собрание литературы «Шанхай – Харбин».

В каталогизации мы используем кириллицу. Материалы Славянской библиотеки можно найти в Интернете, в базе данных HELKA, используя либо латиницу, либо кириллицу.

Фонды комплектуются в первую очередь гуманитарной литературой.

Коллекции национальных меньшинств Российской империи

Это материалы дореволюционного периода на разных языках: еврейских (5000), балтийских (например, коллекция Lettonica содержит 10 000 экземпляров, Lithuanica – 500), тюркских (2800) и кавказских (Georgica – 2000, Armenica – 3000, Caucasia – 40), а также на арабском (600). Общее число – 50 000 наименований.

Уникальное тюркское собрание включает, например, Коран и религиозную литературу, беллетристику, биографии, книги путешественников, учебники, периодические издания и календари.

С помощью Российской национальной библиотеки оцифрован один из татарских журналов, «Шура», а с помощью фирмы «ИнфоКомм» (Санкт-Петербург) была создана поисковая система для этого журнала. Подобную поисковую систему финская фирма «Via Media» разработала для коллекции изданий на персидском языке – Persica. Оцифрованы также еврейские газеты (на идиш) «Der Fraind», «Leben un Visenshaft» и «Unser Leben».

На базе этих коллекций на языках национальных меньшинств мы разработали еще более крупный проект, направленный на изучение языков и культур

рассеяния и технологии культуры. Одновременно с языками обнаруживаем широчайший спектр религиозно-культурных комплексов: мусульмане и шииты в коллекциях Arabica, Caucasica, Persica и Turcica; евреи и караимы в коллекции Hebraica; инославные в Armeniaca, Georgica, Russica, Estonica, Lettonica и Lithuanica. Представлены многочисленные системы письма. Взять хотя бы латиницу с использованием сложных систем диакритики в коллекциях Estonica, Lettonica и Lithuanica, армянскую, грузинскую письменность.

В вышеуказанных системах письма направление текста – слева направо. Но есть письменности, в которых текст располагается справа налево: иврит и идиш, арабский язык, встречающийся в обычной форме и в форме, дополненной диакритическими знаками, которая была принята для персидского, турецкого и многих кавказских языков.

Эти материалы на языках национальных меньшинств представляют громадную ценность для ученых, живущих в разных странах. Однако для крупномасштабного проекта нет средств.

Саамская библиотека

Региональная библиотека Лапландии в Рованиеме отвечает за информационное обслуживание и работу библиотек сети по материалам о саамах и на саамском языке. Кроме того, в фондах библиотеки хранится литература по Северному Калотту и арктическим областям, например коллекция материалов по Гренландии и инуитам.

Информация о саамском народе, саамском языке и публикации, имеющие к ним отношение, собраны в коллекции «Саами», одной из самых полных в Скандинавии. Коллекция содержит художественную и справочную литературу, данные исследований и научные статьи. Особое внимание уделено публикациям о финских и инарских саамах, а также саамах-сколтах и северных саамах. Информация о коллекции зарегистрирована в базе данных Larponica.

Одной из функций специальной библиотеки «Саами» является регистрация финских публикаций на саамском языке в базе данных Fennica Национальной библиотеки Финляндии.

Интересные порталы

Баренц-портал. Баренц-регион является самым крупным регионом межрегионального сотрудничества в Европе, объединяющим Швецию, Финляндию, Норвегию и северо-запад России. Главной целью Баренц-сотрудничества является усиление инфраструктуры стран региона, установление межличност-

ных контактов и как следствие значительный вклад в экономическое, культурное и социальное развитие региона. Информационное обслуживание портала ведет Арктический центр Университета Лапландии.

«Карелия.инфо» – это интернет-гид по Республике Карелия. Техническую поддержку интернет-службы обеспечивает Северо-карельский политехнический институт в городе Йоенсуу.

Самым интересным порталом является «RUSSIAinfo». Это система поиска информации, охватывающая онлайн-ресурсы, в которых содержится информация о России. Содержание портала основано на результатах исследования, проведенного финскими специалистами-россиеведами, и представлено в форме сообщений, статей и обзоров. Эти материалы используются также на российских и зарубежных веб-сайтах, имеющих строгие качественные критерии отбора публикуемой информации.

Поисковая система RUSSIAinfo – это проект, начатый финским Министерством просвещения в 2003 г. с целью обобщения информации о России, получаемой от различных организаций. Обслуживание было начато в сентябре 2005 г. и поддерживается Александровским институтом Хельсинкского университета.

В Республике Карелия разработан портал «Финно-угорские библиотеки России». Идея создания этого портала родилась в ноябре 2003 г. в г. Хельсинки, где состоялось совещание финно-угорских библиотек России и Финляндии. Один из наиболее востребованных сервисов портала – доступ к электронной коллекции изданий на финно-угорских языках, которая создается с 2005 г. и включает оцифрованные материалы на карельском и вепсском языках. С мая 2006 г. было зафиксировано более 23 500 посещений сайта пользователями из 55 стран мира.

У проекта появился новый партнер – Финно-угорский культурный центр Российской Федерации, работающий в г. Сыктывкаре, который уже создал первый филиал финно-угорской виртуальной библиотеки, содержащей материалы на десяти языках.

Многоязычные словари, тезаурусы и онтологии

В 1987 г. мы начали проект Uralica, который длился семь лет. Цель проекта состояла в том, чтобы подготовить библиографии по различным областям финно-угорских исследований, проводившихся в Советском Союзе и Финляндии.

Для общего проекта академиями наук России и Финляндии (а на последнем этапе также независимой Эстонии) были созданы общие списки ключевых слов в таких областях, как археология, этнология и фольклор, языкознание и литературная критика. Словари разрабатывались на русском, финском и английском языках. Таким образом, в проекте использовались три рабочих языка и рассматривалось 20 народностей со своими языками и культурами.

Библиографии покрывали все исследования по вопросам финно-угорских народов, опубликованные в Финляндии и Советском Союзе с 1917 по 1987 г.

Во введениях печатных библиографий главные редакторы, ученые-исследователи отмечают, что словари составлялись по возможности параллельно, но из-за различных традиций исследования или количества описываемых культур количество слов в словарях различается. Например, русско-английский словарь по археологии объемнее, чем финско-английский, потому что российские исследования охватывают гораздо большее количество культур. С другой стороны, финско-английский словарь в области фольклористики и этнологии объемнее русско-английского.

Создание таких словарей потребовало большого количества времени и многочисленных обсуждений: некоторые слова рассматривались очень тщательно в ходе общих встреч, которые длились часами. Это была интересная, захватывающая работа. Проект научил нас критическому подходу к привычным методам работы и парадигмам.

Каждый термин объяснялся на другом языке, даже если данное понятие в нем отсутствовало. Было совершенно очевидно, что все партнеры единогласно решили использовать контекстуальное индексирование.

Результаты работы по использованию терминов для индексации высоко оценены, так как это в значительной степени облегчает проведение исследований. Например, поиск определенного времени и жанра в области литературной критики становится очень простым.

Эта работа имела также огромное значение для небольшого количества дисциплин в Финляндии. Мы не имели до этого финского словаря ключевых терминов языкознания. А теперь список тысяч таких терминов был создан.

Результаты проекта находятся в базе данных ARTO в Финляндии и в базах ИНИОН в России.

Некоторые элементы, которые требуют особого внимания

Особую остроту проблеме придает фонетическая транскрипция, используемая в некоторых заглавиях статей по лингвистике, например, диакритические знаки,

которые помещаются над или под буквой. Финская база данных по уралистике весьма ограничена в этом отношении. Вместо наиболее сложных заглавий там просто стоит «xxx», а содержание статьи объясняется в примечаниях. Такое решение обусловлено тем, что было замечено: транскрипция в большинстве случаев слишком сложна, чтобы ее можно было использовать при поиске. Алфавитный порядок не создает трудностей в базе данных, однако если кто-то захочет получить часть этого многоязычного материала в печатной форме, придется учитывать алфавитный порядок. Даже такие языки, как финский, венгерский и эстонский, которые пользуются латинским алфавитом, имеют небольшие различия.

Авторитетные списки имен: культурные различия

Некоторые исторические или литературные имена в Финляндии и Венгрии адаптированы к национальной системе имен и фонетической системе. Например, Карл пишется Kaarle в Финляндии и Károly в Венгрии, а венгерский писатель Jokai Мор известен в Финляндии под именем Mauri Jokai.

В русском языке – официальном языке регионов, населенных преимущественно представителями урало-алтайских народов, – адаптация строится на фонетическом принципе.

Самый распространенный способ представления имени в публикациях на финно-угорских языках – имя плюс фамилия: Lasse Koskela, Anu-Reet Hausenberg. Венгерские имена строятся в обратном порядке – Lavotha Ödön, Nagy Marika, – причем имя и фамилия не разделяются запятой.

Существует и более архаичная структура имени, в которой первым компонентом выступает патроним, матроним, топоним или этноним, за которым следует личное имя: патроним: Tumaen' Virjas (Тумайсов Вирьяс – по-мордовски), Ez'es Mikol Mikol Volod' (Эзев Миколов Миколов Володя – коми патроним с четырьмя поколениями). Матроним: Elle-Ovlla (Эллина Овлла – саамский). Топоним: Nebdinsa Vittor (Виттор с Севера – коми). Этноним: Olyk Iraj (Ипай из луговых мари – марийский). Некоторые из этих примеров – реальные имена писателей. С развитием национальных культур этот тип имени получает все большее распространение. Немногочисленные ненецкие писатели в основном пользуются русскими именами, потому что их первоначальные имена табуированы. Единственное исключение – Тыко Вылка. Его фамилия означает «оленок».

В результате этого финско-эстонско-русского проекта получен ряд данных, с которыми мы продолжаем работать. ЮНЕСКО организовала научную программу по исчезающим языкам и культурам мира. В нее включены и самые северные урало-алтайские языки – за их положением ведутся наблюдения, результаты которых отражаются в отчетах.

Финляндия оказывает поддержку культурам народов урало-алтайской группы, живущим в России, и сотрудничество библиотек играет значительную роль в этой деятельности.

Финляндия – два языка, одна культура

Двуязычный, но монокультурный пример мы можем найти в Финляндии.

Мы используем в нашей индексации общий финский тезаурус YSA (где вышеупомянутые специальные словари либо объединены, либо используются по отдельности). Тезаурусы – это структурированные словари, где связи между словами отмечены разными кодами.

Для большинства жителей Финляндии родным языком является финский, но у нас есть второй национальный язык – шведский. Общий финский тезаурус был создан на финском языке и потом переведен на шведский.

Финляндия, как говорят, является монокультурной страной. Но когда мы смотрим на проблемы, возникшие при переводе тезауруса, ситуация не так проста. Многие финские слова не имеют шведского эквивалента, например, такие религиозные слова, как *herättäjähuhlat*. Для некоторых шведских слов в финском языке эквивалента также нет. Различие было бы еще больше, если бы мы говорили о шведском языке в Швеции.

Финские онтологии семантического веба

Это большой финский проект, направленный на изучение и развитие инфраструктуры финского семантического веба, представление и поиск информации на базе онтологии. Цель проекта FinnONTO состоит в том, чтобы при индексации перейти от нынешней тезаурусной техники к семантически более богатой технике онтологии.

В консорциум, поддерживающий проект, входят более 30 общественных организаций и компании, финансирующие исследование. Этот консорциум представляет широкий срез культурных и общественных организаций, включая музеи, библиотеки, организации здравоохранения, правительство, СМИ, образование, компании и университеты.

Национальная библиотека Финляндии была активно вовлечена в развитие финской общей онтологии YSO.

Почему тезаурусы недостаточны?

Веб-службы должны стать намного более интеллектуальными, чем сейчас. Это значит, что они должны приобрести семантику. Когда ключевые слова до-

бавлены к библиографической записи или метаданным с помощью человеческого интеллекта, реальный контекст в любом случае сохранится. Структура машиночитаемых онтологий создана при помощи человеческого интеллекта.

Моделирование онтологий для семантического веба – сложная задача, но все-таки успехи в создании семантических веб-служб становятся все более заметны.

Проблемы моделирования онтологий

Конкретные существительные достаточно просто представлять в виде онтологий. Сложнее обстоит дело с существительными абстрактными.

Для них приходится находить фактически индивидуальные решения. Для контекстуального определения приходится проверять все, потому что контекст зависит от языка. Компьютер не способен определить семантические связи слова. Возьмем для примера понятие «ребенок» («дети»). Хотя все мы разного возраста, мы являемся чьими-то детьми. В наших семьях тоже есть дети, которые относятся, скажем, к группе, называемой «школьники», и т.д.

Язык тоже постоянно меняется. Значения слов могут расширяться или сужаться, а то и существенно изменяться. Пример таких изменений – слово «вирус», ставшее столь знакомым всем нам как термин информатики.

Возникает вопрос: кто должен отслеживать семантические изменения или изменения вообще в языке или языках? Задачей онтологической работы, ведущейся в Финляндии, было показать отличительные культурные особенности. Мы отнюдь не стремились достичь идеала.

FinnONTO состоит из 20 000 понятий. Каждому понятию присвоен URI (универсальный код ресурса). Используемые языки – финский, шведский и английский, но шведские и английские эквиваленты не проверялись на соответствие онтологической структуре. Финскую версию подготовила Городская библиотека Хельсинки.

Межкультуральное объединение онтологий

Онтологии строят и развивают уже во многих частях мира. Но межкультуральное объединение онтологий не было предметом частых обсуждений. Однако чтобы не быть бесполезными, разнообразные онтологии должны быть так или иначе совместимы.

В онтологии понятия представлены терминами естественного языка. Межкультурный информационный поиск начинает работать, когда документы в

собрании содержатся на одном языке, а пользовательские запросы поступают на другом. Ведь для пользователя легче сформулировать запрос, используя слова родного языка. В результате поиск может выдать большое количество документов на языках, которых пользователь не знает. В подобных ситуациях помогают системы машинного перевода.

Есть такие слова, значения которых в разных языках совпадают, например, слово «солнце». Есть слова, значения которых зависят от культуры. Хороший пример – названия цветов. Есть понятия, которых нет в культуре языка, но слово существует, например, спаренные глаголы (это явление характерно для финно-угорских и многих восточных языков).

Разработчики компьютерных программ признают, что труднее всего научить компьютер «думать» так, как мыслит человек. Попытки построить теорию здравого смысла предпринимались в психологии, антропологии и когнитологии.

Программисты продолжают свою работу. В настоящее время есть такое перспективное направление исследований, как создание многоязычных онтологий, в которых перевод осуществляется не на уровне слов, а на уровне понятий.

Пользователи ждут более быстрого и точного обслуживания.