



68th IFLA Council and General Conference

August 18-24, 2002

Code Number:	006-122-R
Division Number:	IV
Professional Group:	Classification and Indexing
Joint Meeting with:	-
Meeting Number:	122
Simultaneous Interpretation:	-

Предметно-ориентированное взаимодействие: проблемы, связанные с проектом Тезауруса высокого уровня (HILT)

Деннис Николсон
Strathclyde University
Glasgow, UK

Введение: от HILT-фаза I к HILT-фаза II

Проблемы, связанные с предметно-ориентированным взаимодействием, затронутые в данном докладе, проистекают из двух проектов, ныне называемых HILT-фаза I и HILT-фаза II. Проект HILT-фаза I (ранее известный просто как проект HILT) был представлен в ноябре 2001. Он финансировался совместно Объединенным комитетом по информационным системам (Joint Information Systems Committee – JISC) и Исследовательской программой поддержки библиотек (Research Support Libraries Programme – RSLP) и длился чуть больше 1 года. Вопрос о финансировании проекта HILT-фаза II комитетом JISC должен быть решен в апреле 2002. Фаза II продлится 12 месяцев и будет использовать работу HILT-фаза I, а также навыки и опыт команды, которая эту фазу выполняла, для создания на основе группового консенсуса, достигнутого в ходе разработки HILT-фаза I, пилотной службы установления терминологических соответствий или маршрутной карты с уделением особого внимания развитию Распределенного национального электронного ресурса (Distributed National Electronic Resource – DNER), включая – но не обязательно этим ограничиваясь – предметную терминологию, ориентированную на высшее образование (HE) и совершенствование образования (FE) и предназначенную для описания коллекций в информационной среде (IE), планируемой JISC. Оценка пользователей и анализ затрат и прибылей различных уровней службы будут характерными особенностями этого проекта.

Проблема

Гарантия, что FE- и HE-пользователи информационной среды (IE) смогут найти соответствующие учебные, исследовательские и информационные ресурсы по предмету, является одной из основных задач, стоящих перед JISC, DNER, Сетью обнаружения ресурсов (Resource Discovery Network – RDN), а также различных провайдеров ключевых информационных и учебных услуг через архивы, библиотеки, музеи и домены электронных служб. Как показала NILT-фаза I, различные провайдеры услуг используют ряд предметных классификаций (от общих классификаций типа LCSH, UNESCO, DDC и AAT до специальных классификаций типа MeSH) для того, чтобы отвечать требованию адекватного и последовательного описания своих ресурсов для точного поиска. Если известно, что перекрестный поиск и просмотр осуществляется последовательно для пользователей информационной среды (IE), то эти схемы должны быть связаны с друг с другом, возможно, с использованием общего 'хребта', такого как DDC, имеющей международное использование, переведенной на многие языки и обладающей потенциалом в плане упрощения сетевого обмена (M2M). Возможно, еще более важным моментом является то, что терминология в представлении различных FE- и HE-пользователей должна быть 'непротиворечивой'¹, а затем переведенной в термины, используемые службой, которые необходимы пользователям для перекрестного поиска или просмотра группы услуг, релевантных их запросу. Цель проекта NILT-фаза II – построение и оценка пилотной службы, которая свяжет этот процесс в качестве 'Совместной службы' DNER в информационной среде (IE).

NILT-фаза I

NILT-фаза I показала, что:

- Много различных предметных классификаций и методик используются в службах Соединенного Королевства, которые считают, что предметный поиск через их посредство представляет ценность как для их пользователей, так и для их сотрудников.
- Была достигнута договоренность между сообществами архивов, электронных служб, библиотек и музеев о том, что целесообразно создать имеющий более практичную направленность пилотный проект, который определит затраты и прибыли и будет развивать сетевую, отвечающую требованиям машины и пользователя, интерактивную терминологическую маршрутную карту, используемую этими сообществами, и связи между терминологиями (ранее известную в проекте NILT как 'пилотная служба соответствий' – см. Схему терминологической маршрутной карты (TeRM) в Приложении А, где в общих чертах описано, что это такое и как будет работать).
- Необходимо провести дальнейшее исследование в области эффективности, уровня и характера потребностей пользователя, практичности требований к дизайну и затрат и прибылей при данном подходе, прежде чем взять на себя долгосрочные обязательства по созданию возможно дорогостоящей службы. Было определено, что

¹ Процесс определения того, какую информацию ищет пользователь, напечатавший 'лотос', – информацию об автомобиле, пакете прикладных программ, цветке, или он имеет в виду другое значение этого термина.

сделать это лучше с помощью пилотного проекта, который проработает эти и родственные вопросы.

Более подробно с проектом HILT-фаза I можно ознакомиться на web-сайте HILT², в частности, по Заключительному Отчету HILT³.

HILT-фаза II: Цели

HILT-фаза II продвигает рассматриваемый процесс на стадию пилотного проекта, в центре внимания которого – как рекомендовано на стадии HILT-фаза I – находятся требования к терминологии и тезаурусу на уровне коллекций, но с учетом потребностей поиска отдельных документов. На этой стадии будет учтена работа, проведенная в рамках HILT-фаза I, а также будет учтен опыт и навыки команды разработчиков с целью создания службы пилотной терминологической маршрутной карты или TeRM, подобной той, которая была предложена на стадии HILT-фаза I и имеющей своей целью:

- а) Обеспечить практическую экспериментальную базу, на основе которой можно исследовать и определять требования к службе по выработке предметной терминологии для информационной среды (IE) JISC с учетом DNER, RDN, пользователя, уровня коллекций, международной совместимости и аспектов доступа на местном, региональном, национальном уровне и на уровне Соединенного Королевства.
- б) Выработать рекомендации в форме пожеланий будущей службе, принимая во внимание ряд факторов, включая уровень и характер потребностей пользователя, практическую сторону дела, конструктивные требования, эффективность, выгодность покупки пакетов программ по сравнению с разработкой собственного программного обеспечения и, самое главное, все за и против для пользователей терминологической службы FE and HE, прежде всего сконцентрированной на потребностях на уровне коллекций.

HILT-фаза II: Участники

HILT-фаза II будет продолжаться в течение 12 месяцев и будет осуществляться приблизительно теми же участниками, что и HILT-фаза I, но с привлечением большего количества представителей из DNER, RDN и FE, а именно:

- Центр исследования электронных библиотек (Centre for Digital Library Research – CDLR) Стрэтклайдского университета – ведущая организация;
- Представитель DNER
- mda (бывшая Ассоциация по музейной документации (Museums Documentation Association));
- Национальный совет по архивам (National Council on Archives – NCA);
- Национальная сеть образования Шотландии (National Grid for Learning (NGfL) Scotland);
- Онлайн-компьютерный библиотечный центр – Online Computer Library Center – OCLC);

² <http://hilt.cdli.strath.ac.uk/>

³ <http://hilt.cdli.strath.ac.uk/Reports/FinalReport.html>

- Представитель RDN
- Представитель FE
- Шотландский библиотечно-информационный совет (Scottish Library and Information Council – SLIC);
- Шотландский университет промышленности (Scottish University for Industry – SufI);
- Управление Соединенного Королевства по библиотечным и информационным сетевым технологиям (UK Office for Library and Information Networking – UKOLN);
- Эксперты по терминологии Alan Gilchrist и Leonard Will (внешний эксперт).

В результате вовлечения в групповой проект CAIRNS⁴ (в котором поддержка коллекций использовалась для мини-групп), проекты SCONE и SEED⁵, объединенные для построения межсекторной базы данных коллекций⁶, и проект HILT⁷ фаза I, лидирующий сайт – Центр исследования электронных библиотек Стрэтклайдского университета⁸ – имеет опыт использования описаний на уровне коллекций в динамически распределенной среде и решения связанных терминологических проблем. Сюда входит распределенный каталог CAIRNS с участием университетов, Национальной библиотеки Шотландии (NLS), NGfL, SLAINTE, базы данных электронной библиотеки Глазго (GDL), предметно-ориентированный механизм отражения коллекций, база данных коллекций SCONE, e-print сервер Инициативы открытых архивов (Open Archives Initiative – OAI), базы данных Фонда новых возможностей (New Opportunities Fund – NOF) и другие проектные цифровые базы данных, а также потенциально другие базы данных с использованием Z39.50. Другие участники – особенно UKOLN, mda, NCA, RDN DNER и эксперты по терминологии HILT – придали дополнительную глубину и широту охвата данной команде. Вдобавок OCLC дал согласие помогать разработке, обеспечивая доступ к машиночитаемым маршрутным картам от предметных рубрик библиотеки Конгресса (LCHS) до DDC и соответствующий доступ для проведения экспертизы. CDLR тесно работает также с десятью FE-колледжами Глазго в проекте RSLP GDL.

Построение терминологической маршрутной карты (TeRM)

Для целей этого проекта создан пилотный проект TeRM с использованием коммерческого программного обеспечения Wordmap⁹. Он должен (на основе опыта HILT-фаза I) продемонстрировать средства, необходимые для создания пилотной модели. Это не означает предпочтение данного программного обеспечения или поставщика, ни его рекламу в противопоставление доморощенным или открытым источникам. Целью данного проекта является разработка спецификации требований на основе оценки пользователей, фокусных групп обслуживания и внешних экспертов. Затем он должен сравнить все имеющиеся релевантные пакеты, сделав всесторонний обзор всех коммерческих и других решений. WordMap будет среди предложенного программного обеспечения, в основном отвечающего требованиям. Будет также

⁴ См. <http://cairns.lib.strath.ac.uk> – каталог университетов, NLS, NGfL и др. с использованием Z39.50

⁵ См. <http://scone.strath.ac.uk> и <http://seed.cdlr.strath.ac.uk>

⁶ См. <http://scone.strath.ac.uk/service/index.cfm>

⁷ См. <http://hilt.cdlr.strath.ac.uk>

⁸ См. <http://cdlr.strath.ac.uk>

⁹ См. www.wordmap.com

исследован вопрос о том, предпочтителен ли выбор общественного открытого источника выбору коммерческого решения.

Есть много вещей, говорящих в пользу использования коммерческого программного обеспечения на этой стадии развития. Опыт HILT-фаза I показывает, что участникам проекта легче обсуждать требования такого вида сервиса, дающего подлинный иллюстративный пример тому, на чем стоит сосредоточиться. Следовательно, он дает существенное подтверждение тому, как полезно иметь пилотный проект на ранней стадии разработки, чтобы помочь привлечь внимание и пробудить интерес пользователей и других заинтересованных сторон и дать им возможность попрактиковаться и обозначить рамки решаемой проблемы. Wordmap используется, т.к. мы хотим иметь настоящую демонстрационную модель на ранней стадии разработки для пользователей и провайдеров услуг, в рамках которой можно взаимодействовать. Попытка выработать полный набор требований без предварительной пилотной модели привела бы, думается, к худшим результатам, т.к. деятельность пользователей и провайдеров услуг не была бы достаточно простимулирована работой в конкретном контексте, и они не смогли бы представить всех проблем. Это прогматичный подход, который приведет нас к правильной оценке проблем, а также убережет нас от потенциальных потерь и риска на пути к полному решению задачи.

Терминологии и терминологические проблемы

Первоначальная иллюстративная TeRM будет основана на терминологии RDN¹⁰, терминологии, известной как часть таксономических наборов Wordmap, которые включают в себя, в частности, набор терминов, используемых интернет-пользователями, а также на отобранных наборах LCSH, DDC, UNESCO и AAT. OCLC обеспечит разработку соответствий LCSH – DDC и, может быть, сможет обеспечить разработку соответствий между DDC и предметными рубриками Conspectus¹¹. Тезаурус UNESCO доступен в онлайн-режиме¹², а выборки AAT можно получить из руководств. Целью должно стать создание выборочных соответствий, достаточных для пилотной модели на первом этапе, т.е. это не будет многоаспектной терминологической картой. Необходимо также проанализировать различные проблемы, поднятые экспертом HILT-фаза I Леонардом Уиллом (Заключительный отчет HILT, раздел 10) и два дополнительных вопроса:

1. Вопрос о том, будет или нет TeRM нуждаться в центральном 'хребте'.

Ключевым элементом в обеспечении данной пилотной модели должна стать возможность 'перевода' терминов запроса пользователя, изложенного в терминах пользователя, в различные термины, используемые в распределенной среде, причем сделать это надо с умом и с пользой для дела. Потребуется взаимодействие между пользователем и TeRM, чтобы 'разрешить неоднозначность' термина или терминов, используемых пользователем, (например, имеется ли в виду 'лотос' цветок или автомобиль, или пакет прикладных программ, или что-нибудь еще. Есть два варианта ответа:

¹⁰ См. список, созданный Andy Powell на <http://www.rdn.ac.uk/cgi-bin/browse>

¹¹ Предметные рубрики Conspectus используются в базе данных CAIRNS.

¹² См. <http://www.ulcc.ac.uk/unesco/thesaurus.htm>

а) вероятный

б) лучше в терминах наиболее успешного поиска для пользователя (следует учитывать, что поиск может идти на разных языках)

на вопрос, какой характер будет носить взаимодействие между пользователем и единой центральной классификацией, с которой соотнесены все другие классификации в среде TeRM, или между пользователем и каждой конкретной классификацией. Если для достижения цели выбирается единая основная классификация, то следом встает вопрос: какая из существующих классификаций будет лучше служить поставленной цели? Наибольшие шансы у DDC (хорошо структурированной, иерархической классификации, уже переведенной на основные языки мира).

2. Вопрос, так или иначе, сводится к решению проблемы предметной терминологии в распределенной среде: возможна ли идентификация и принятие единой классификации, считающейся отвечающей всем требованиям во всех областях – это может быть как абсолютно новая классификация, так и известная классификация, по возможности приспособленная для удовлетворения требований модели.

В конечном счете это не что иное, как расширение анализа затрат и прибылей проекта службы установления терминологических соответствий – точка отсчета, из которой надо исходить при подходе к оценке модели TeRM по сравнению с другими моделями.

Построение исследовательской среды

Это будет достигаться путем присоединения DNER и других коллекций, включая коллекции RDN, архивные, музейные коллекции и местную коллекцию OAI к базе данных коллекций SCONE¹³ для создания тестовой базы данных коллекций NILT-фаза II и среды с архитектурой и перекрестным поиском на основе CLD, используя динамичный механизм создания архитектуры CAIRNS и широкий механизм поиска. Целью будет использование 'родных' предметных классификаций для коллекций в данной среде и использование пилотной TeRM для устранения неоднозначности пользовательских терминов и согласования различных классификаций. Можно воспользоваться рядом пользовательских базисных архитектур, связав их в общем виде с предметными узлами в соответствии с тематическими интересами и представляя различные виды пользовательской среды: местную, региональную, национальную, Соединенного Королевства (общую) и Соединенного Королевства (предметные узлы)¹⁴. Целью будет привязать TeRM к механизму создания архитектуры, если это возможно (CAIRNS предлагает это сделать), или хотя бы симитировать этот аспект (это будет не так красиво, но достаточно для поисковых целей проекта).

Результаты NILT-фаза II

¹³ <http://scone.strath.ac.uk/Service/Index.cfm>

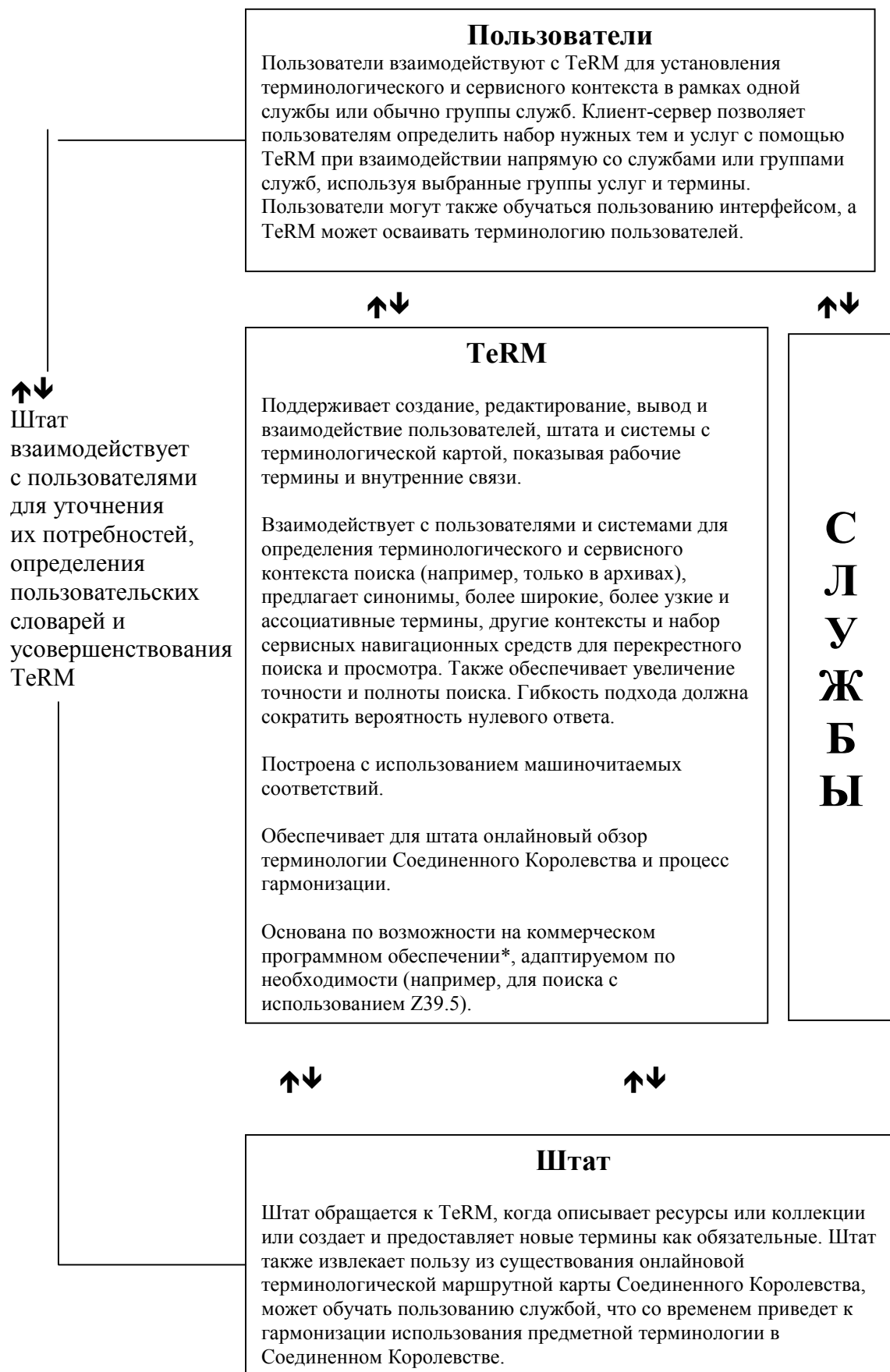
¹⁴ В частности, следующие архитектуры: университета, FE-колледжа, HE и FE в Глазго, HE и FE в Шотландии, DNER для HE и FE, RDN для HE и FE.

Конкретными результатами НИЛТ-фаза II являются:

1. Лучшее понимание проблемы и потребностей пользователей FE and HE относительно предметного поиска в проектируемой информационной среде JISC, как в рамках JISC, служб JISC, так и в сообществе в целом благодаря деятельности по распространению.
2. Более глубокое понимание требований к терминологическим соответствиям в DNER и связанных с ним службах Соединенного Королевства, охватывающее особенности местных, региональных, национальных, международных уровней, предметных узлов, FE и HE, а также архивов, библиотек, музеев и электронных служб.
3. Действующая пилотная терминологическая демонстрационная служба для JISC IE (с ограниченными функциями и с полным спектром услуг, требующим, возможно, замены программного обеспечения).
4. Требования, материальные затраты на установку и обслуживание, и затраты по отношению к прибыли для будущей службы, включая терминологические и функциональные требования пользователей и M2M.
5. Заключительный отчет по проекту вместе с необходимыми рекомендациями.

При соблюдении предполагаемого финансирования (к моменту написания доклада имеется только неформальная договоренность) НИЛТ-фаза II начнется в мае 2002 г.

Приложение А. Схема интерактивной терминологической маршрутной карты (TeRM)



***Примечание:** Примеры можно посмотреть на www.wordmap.com, www.oingo.com
и vivisimo.com