

МЕТА2: ОНТОЛОГИИ КАК ЯДРО ИНТЕРНЕТ-СЕРВИСОВ

А.В. Малых, А.А. Манцивода

В докладе рассматривается концепция построения и ключевые принципы реализации системы Мета2. Кратко систему Мета2 можно охарактеризовать как платформу, которая ориентирована на организацию работы в тех предметных областях, которые обладают значительными объемами разнообразных по структуре знаний и данных. Примерами таких предметных областей являются научные исследования, образовательный процесс, а также создание продвинутых веб-сервисов и проектов.

Система Мета2 базируется на использовании методов математической логики. Ключевой проблемой, которая препятствует широкому распространению выразительных средств логики, являются ее непривычность, сложность и абстрактность. Слишком далека логика от среднестатистического пользователя, слишком высок барьер, который следует преодолеть. Это обеспечивает логике совсем ненужную ей (в данном контексте) элитарность. Поэтому ключевой задачей при создании системы Мета2 была реализация "принципа айсберга", когда мощные и выразительные механизмы логики скрыты, а над поверхностью видна лишь небольшая верхушка. С принципом айсберга связан также механизм "постепенного таяния". Это означает, что пользователь должен с легкостью начать работать в системе, а процесс изучения уточненных инструментов системы должен быть мягким и постепенным. Понятность и естественность базовых механизмов взаимодействия человека и системы были ключевыми вызовами для разработчиков Меты2.

В решении этих проблем ключевую роль играет интерфейс в широком смысле - как система сценариев взаимодействия человека и компьютера. Следует отметить, что разработка интерфейсов системы Мета2 заняла около пяти лет, и эта работа продолжается. Базовые механизмы интерфейса (деревья и записи), в которые конвертируется внутренняя логическая кухня системы, просты и понятны для пользователей. Люди с самого начала могут решать на Мете2 свои задачи. Постепенно проходит процесс "втягивания" пользователя в возможности системы, понимания им более продвинутых возможностей, для использования которых уже требуется определенная квалификация. Разные "слои" системы обеспечивают работу пользователей разной квалификации, которые вольны выбрать тот уровень абстракции, который им наиболее комфортен и эффективен.

В Мету2 встроены механизмы построения специализированных интерфейсов, которые основаны на технологии JavaFX. JavaFX - это бурно развивающийся под эгидой компании Sun Microsystems подход к созданию интерфейсов. Мета2 устроена таким образом, что она адаптирует созданные пользователем интерфейсы не хуже, чем свои собственные. Поэтому пошаговое добавление интерфейсов обычно ведет Мету2 к миграции в сторону специализированных систем со своими интерфейсами, ориентированными на решение задач в определенной предметной области.

Образно говоря, универсальность системы Меты2 основана на том, что она дает пользователю не рыбу, а удочку, причем такую, с которой легко могут справиться не только профессионалы-рыбаки, но и любители. Разработчики системы в качестве целевой аудитории Меты2 рассматривали не только профессиональных разработчиков информационных систем, но и, скажем, специалистов в области естественных наук, гуманитариев, учителей, реализаторов совместных проектов, и вообще, пользователей, которые хотели бы иметь инструмент для системной организации своей работы в самых разных областях деятельности. Методологически была сделана попытка формирования такой среды работы, при которой не требуется опускать предметную область до уровня работающего в системе формализма (как это получается, например, в базах данных, когда предметную область необходимо разложить в набор плоских записей). Наоборот, сценарий состоит в постепенном "подтягивании" Меты2 к предметной области до тех пор, пока она не станет адекватно отображать и моделировать ситуацию в последней. Процесс работы в Мете2 заключается в том, что пользователь постепенно строит онтологию, моделирующую предметную область (хотя в некоторых вариантах интерфейсов он может и не подозревать этого). На эту логическую модель, как плоть на скелет, наращиваются сервисы, интерфейсы и ресурсы. Мета2 предоставляет джентльменский набор стандартных инструментов для работы с логической средой (в частности, набор стандартных интерфейсов), которых во многих случаях достаточно для организации эффективной работы. При необходимости пользователь строит свои модули и интерфейсы.

Концептуально Мета2 находится в сфере влияния Semantic Web [1]. Логическим мотором системы являются диалекты дескриптивных логик [2]. Онтологии в формате языка OWL [3] легко импортируются в систему и экспортируются из нее. Логическим ядром концепции Меты2 является автоматический решатель, представляющий собой систему автоматического доказательства, основанную на логическом исчислении резолюционного типа, который работает на дескриптивных терминах. Ядро системы, написанное на Java, безразлично к способу применения - будь то локальный компьютер, сервер, распределенная среда онтологий и т.д.

Вероятно, наиболее близка Мете2 система Protege, разработанная в Стэнфордском университете [4]. Ключевым отличием между системами (если не сравнивать объемы накопленного опыта и ресурсов, которых у Protege несравнимо больше) является целевая аудитория. Интерфейс Protege более прямолинейно привязан к структуре онтологий, поэтому работа в системе требует большей подготовки. В Мете2 логический механизм спрятан за системой интерфейсных сценариев. С другой стороны, с Метой2 можно работать и как с системой управления объектными базами данных, если интерпретировать сами онтологии как ОБД. Но, конечно, полный функционал Меты2 простирается намного дальше функционала СУОБД.

Архитектура системы. Мета2 строится как модульная система. К основным модулям могут легко привязываться дополнительные. Базовыми модулями являются модули хранения, модуль объектной модели и модуль пользовательского интерфейса.

Одной из ключевых частей системы Мета2 является т.н. объектная модель, которая предлагает низкоуровневый доступ к данным онтологий. Фактически, объектная модель представляет собой связующую часть, через которую происходит взаимодействие модулей. Например, модули интерфейса (на Java или JavaFX) взаимодействуют с модулем хранения через объектную модель. Объектная модель используется как параметр при вызове методов различных модулей, так и в роли возвращаемого значения.

В качестве базового модуля хранения используется модуль хранения онтологий в БД. Его развитие происходило в несколько этапов. На первом этапе хранение данных онтологии почти полностью было переложено на БД, но затем для улучшения производительности были добавлены различные алгоритмы кэширования данных в оперативной памяти. Как альтернативный вариант сейчас развивается модуль хранения, который использует оперативную память как основное рабочее место хранения онтологий, тогда как состояние данных сохраняется в специальные файлы, которые представляют собой журнал изменения. Этот подход более требователен к объему оперативной памяти, но выполняет операции по изменению или получению данных из онтологии намного эффективнее. Кроме того, появляется возможность реализовать транзакции не на уровне БД, а на уровне объектной модели. Заметим, что данное представление абсолютно необходимо и в тех случаях, когда активно используется логический редактор, манипулирующий с содержимым онтологий.

Объектная модель хорошо подходит для низкоуровневых операций, но при этом порождает чересчур громоздкий код при реализации даже типовых действий. Для решения этой проблемы разрабатывается язык запросов MetaPath, играющий роль, аналогичную SQL в базах данных. Язык запросов обеспечивает компактный, но достаточно выразительный способ доступа к данным. Он базируется на идеях XPath и дескриптивной логики. Семантика языка строго определена, а синтаксис покажется знакомым людям, знающим XPath.

Проекты. Как показывает опыт, накопленный при реализации разнообразных приложений, само понятие онтологии редко живет в своем замкнутом мире. В сложных предметных областях, как правило, работают несколько онтологий, которые взаимодействуют друг с другом. Например, типичным можно было бы считать подход, когда структура некоторых данных (классы, свойства) определяется в одной (базовой) онтологии, а дальнейшее расширение происходит уже в других -- "объектных" -- онтологиях (создаются объекты, расширяются и наследуются классы и т.п.).

С другой стороны, работа с большой базой знаний, которая содержит значительное количество разрозненных онтологий, не комфортна. Люди намного эффективнее работают с небольшим количеством различных сущностей. Кроме того, невозможность локализовать пакет онтологий, относящийся к конкретной предметной области, не позволяет формировать на базе Меты2 "отчуждаемые" модели предметных областей, которые можно было бы передавать и распространять в информационных средах. Для решения этих проблем в Мете2 была реализована концепция "проекта".

Проект - это набор онтологий, интерфейсов и настроек, который технически реализуется как отдельная, независимая от других база знаний. Проект является одной из базовых сущностей с точки зрения пользователя. Проекты можно переносить из одной системы Мета2 в другую - при помощи механизма импорта и экспорта. Понятие проекта позволяет легко отчуждать сформированную среду, передавать ее другим пользователям и распространять в информационных сетях, что обеспечивает механизм построения распределенных сред знаний и данных (включая семантические гриды).

Следует отметить, что средств, предоставляемых проектом, достаточно для того, чтобы превратить Мету2 в специализированную систему, сформированную в соответствии с архитектурой MVC. Здесь model представляется онтологией или пакетом онтологий, view - собственными интерфейсами Меты2, либо разработанными создателем проекта на основе технологии JavaFX или веб-технологий. Control может быть организован как встроенными средствами самой Меты2 (в несложных случаях), так и разработанным пользователем.

Пользовательский интерфейс и дерева. Основная часть интерфейса, с которой начинает работать пользователь - это пользовательский клиент системы Мета2. Он реализован нами как вебстарт-приложение [5]. Это означает, что для инсталляции Меты2 достаточно пройти по специальной ссылке. Если на компьютере установлена

Java, то пользовательский клиент запустится и будет работать. Кроме легкого и автоматического начала работы, использование вебстарта предоставляет возможность автоматического обновления.

Можно отметить еще одну часть клиента - это модуль работы с деревьями. Представление онтологий в древовидной форме может зависеть от задач, которые стоят перед пользователем. Мета2 отличается тем, что она предоставляет пользователю механизмы создания собственных типов ленивых деревьев, топология которых соответствует природе решаемой задачи. Для реализации настроек деревьев используются механизм правил. Правила определяют то, какие узлы открываются в дереве и в какой последовательности. Правила могут быть заданы на уровне конфигурации проекта, либо задаются в динамике при работе с деревьями.

JavaFX как инструмент создания компонент интерфейса. В качестве средства для создания модулей интерфейса можно использовать язык Java. Как альтернатива предлагается также скриптовый язык проекта JavaFX [6]. Этот язык ориентирован именно на разработку пользовательских интерфейсов, содержит в себе вспомогательные конструкции для этой цели, которые упрощают и делают более надежным код программы. Первоначально в системе Мета2 использовалась интерпретируемая версия языка, но, в связи с развитием JavaFX, был произведен переход на его компилируемую версию. Существует простой механизм подключения модулей на JavaFX в Мету2 в качестве компонент проектов.

Пилотные проекты на основе Мету2. На данный момент особый интерес у нас вызывает идея использования Мету2 для создания принципиально новых подходов к производству "интеллектуальных" Интернет-сервисов и управлению веб-контентом, базирующихся на онтологиях. В частности, сейчас активно развивается метод удаленного доступа к онтологиям через веб-интерфейс. Метод удаленного доступа должен обеспечивать:

1. Создание и редактирование терминологической структуры онтологии (работа с Т-боксом в терминологии дескриптивных логик).
2. Наполнение онтологии объектами, редактирование объектов и их удаление (работа с А-боксом).
3. Эффективный механизм удаленного получения данных и знаний из онтологии.

Механизм доступа к данным Мету2 через Интернет является ядром технологии разработки Интернет-сервисов, в которой классическая пара XML+БД заменяется на онтологии. В частности, нами рассматривается возможность построения интерфейсов через JSP и библиотеки тэгов: как стандартные JSTL (JSP Standard Tag Library), так и работающие с хранилищами Мету2. В ядре разработки лежит объектная модель онтологий, функционирующая на серверной стороне.

Приложения в области естественных наук. В настоящее время для целого ряда исследовательских направлений, связанных с изучением озера Байкал и окружающей территории, разрабатывается пакет онтологий с соответствующей информационной поддержкой. Основной идеей этих проектов является сбор и концентрация доселе разрозненных междисциплинарных данных. В настоящее время в высокой степени готовности находятся: онтология географических объектов Байкальской Сибири, трехмерная модель озера Байкал, информационная система по планктону озера Байкал, виртуальный гербарий. Начата разработка онтологий по минеральным источникам и геологическим катастрофам Байкальской Сибири. Данные работы проводятся при поддержке научно-образовательного центра Байкал [7] совместно с профильными институтами ИНЦ СО РАН, факультетами ИГУ и Бурятского государственного университета.

Мета2 и ГИС-проекты. Общая идея ГИС-систем, построенных на основе Мету2 основана на сочетании специализированных онтологий, относящихся к пространственно распределенным предметным областям и геометрических/картографических онтологий общего характера. Пока основное достижение в этом направлении - создание подробной трехмерной модели Байкала, состоящей из более 100 тыс. компонент, а также разработка общих модулей работы с элементами пространственно распределенных предметных областей как планарными геометрическими объектами.

Приложения в области образования. На базе естественнонаучных проектов, перечисленных выше, планируется создание специализированных информационно-справочных систем для студентов, которые будут находиться в открытом доступе. Кроме того, развивается ряд проектов в области математического образования. Отметим здесь два проекта. Во-первых, это проект информационно-справочной системы по математическому анализу, который создается на основе популярной онлайн-консультации [8], которая активно действует уже в течение нескольких лет. В рамках этого проекта накоплен уникальный материал в форме диалогов о задачах матанализа и их решениях. Для конвертирования этого материала в информационно-справочную систему была разработана онтология в области математического анализа, а затем диалоги онлайн-консультации были размечены элементами этой онтологии, которые сыграли роль метаданных. На основе этой разметки в Мету2 была автоматически выстроена структура, обеспечивающая быстрый доступ к диалогам, исходя из тематики, уровня сложности и т.д.

Второй пилотный проект связан с хорошо известной энциклопедией по линейной алгебре Линеал [9]. В Мету2 была построена онтология, отображающая структуру данной энциклопедии и создан соответствующий проект (в закрытом доступе, чтобы не нарушить авторских прав разработчиков Линеала).

Особый интерес для апробирования системы Мета2 представляет разработка модуля тестирования знаний на основе спецификации IMS QTI [10]. Здесь нами исследуется общий подход, основанный на формализации спецификаций и стандартов. Как уже отмечалось, онтологии являются логическим формализмом, обладающим строгой математической семантикой. С другой стороны, онтологии - это выразительный язык представления знаний, реализованный в системе Мета2. Стандарты и спецификации пишутся на естественных языках. Тем не менее, их тексты должны быть в максимальной степени формализованы - для обеспечения строгости и однозначности описаний. Идея состоит в том, чтобы использовать онтологии и Мета2 для манипуляций со стандартами и спецификациями. Во-первых, онтологии могут быть использованы при разработке самих стандартов, поскольку обеспечивают необходимое структурирование и формальную строгость. Во-вторых, "погрузив" спецификацию в онтологию, а значит, получив ее формальное описание, мы можем использовать спецификации как инструмент прямого действия, на базе которого можно строить конкретные системы, работающие на основе данных спецификаций. Аналог здесь - использование грамматик для автоматического генерирования парсеров. Но язык онтологий намного шире языка грамматик, поэтому и возможностей (и сложностей) здесь намного больше.

Нами была разработана структурная онтология и модуль по переводу XML-описаний IMS QTI в эту структурную онтологию. На основе этой онтологии на портале [7] был реализован Интернет-сервис, который обеспечивает пользовательский интерфейс для работы с тестами: их создание, редактирование, использование.

Электронная библиотека. В основе данной библиотеки лежит экспериментальная онтология библиографических данных, в которой формализованы элементы спецификаций BibTeX, LOM, а также библиографических стандартов. Над данной онтологией надстроены методы и интерфейсы, которые формируют гибкую среду электронной библиотеки. В данном проекте апробируется технология построения сложных веб-ресурсов на основе онтологий и системы Мета2. Также, как и в системе тестирования, реализация ведется через прямое использование стандартов и спецификаций, формализованных в виде онтологий. Был реализован Интернет-сервис, который позволяет пользователю производить манипуляции с библиотечными данными.

Заключение. Сентябрь 2008 года является датой, когда нами планируется открытие доступа к бета-версии системы Мета2. Также в планы включено создание платформы для обмена онтологиями и проектами, и создание репозитория стандартных онтологий. Будет осуществляться доводка естественнонаучных онтологий и доработка соответствующих проектов. На математическом уровне особое внимание будет уделено дальнейшей настройке логического исчисления и стратегиям поиска вывода в данном исчислении.

ЛИТЕРАТУРА:

1. W3C Semantic Web Activity. - <http://www.w3.org/2001/sw/>.
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. The description logic handbook: theory, implementation, and applications. - Cambridge University Press, 2003.
3. OWL Web Ontology Language Overview. - <http://www.w3.org/TR/owl-features/>.
4. Protege: A Free, Open Source Ontology Editor and Knowledge-base Framework. - <http://protege.stanford.edu/>.
5. Java Web Start Technology. - <http://java.sun.com/products/javawebstart/>.
6. JavaFX Technology – At a Glance. - <http://java.sun.com/javafx/>.
7. Научно-образовательный центр "Байкал". - <http://lake.baikal.ru>.
8. Онлайн-консультация по математическому анализу. - <http://teacode.com/forum/show-forum.jsp?forum=0>.
9. Линеал. Базовая электронная энциклопедия по линейной алгебре. - <http://lineal.guru.ru/lineal3/>.
10. IMS Global Learning Consortium. - <http://imslobal.org/>.