

# ВИКОНТ: ВИзуальный Конструктор ОНТологий для структурирования семантической информации

Гаврилова Т.А., Лещева И.А.

Институт Высокопроизводительных Вычислений и Баз Данных,  
Санкт-Петербург

E-mail: gavr@fn.csa.ru, irina@fn.csa.ru

## 1 Введение

Internet, как всеобъемлющая сеть электронных ресурсов, предоставляет возможность наиболее широкого доступа к информации, хранящейся в электронных библиотеках. Основным инструментом для нахождения необходимых документов в Internet является работа с поисковыми серверами. Поисковые сервера в данное время осуществляют поиск информации по ключевым словам. Такой способ поиска данных часто оказывается крайне неэффективным, поскольку очень редко удается указать смысловую нагрузку набором слов. Составление баз данных, уже обладающих смысловой структурной информацией о других серверах и электронных библиотеках требует больших машинных и человеческих (машина не умеет сама структурировать знания и осуществлять поиск по смыслу) ресурсов.

Организация более эффективного поиска опирается в проблему плохой структурированности данных. До настоящего времени большая часть электронной информации, находящейся в недрах Internet, семантически плохо структурирована. Построение каталогов электронных библиотек осуществляется по формальным признакам, и осмысление информации, содержащейся в электронных документах, авторами не производится. В результате, значительная часть документов, относящаяся к слабо структурированным предметным областям, оказывается в семантическом отношении поверхностной, и читатель таких документов вынужден блуждать в лабиринте перекрестных ссылок. Использование механизма онтологий позволило бы авторам электронных публикаций избежать хаоса в создаваемых документах.

Проблема структурирования информации до сих пор не имеет общего решения, удовлетворительного во всех случаях. Каждый исследователь, сталкивающийся с ней, реализует метод действия, подходящий в своем случае. По-видимому, решить проблему создания единой методики структурирования информации так и не удастся ввиду всевозрастающей сложности самой информации и ее представлений. Тем не менее, существует потребность в создании единых структур для представления однотипных знаний, подходящих для различных задач. Доступ-

ность таких структур позволит совместно использовать знания, в них заложенные, а также заимствовать их для многократного использования в более сложных структурах. Несмотря на активное развитие направлений исследований, посвященных сбору и обработке информации, процесс структурирования данных и знаний так и не получил должной поддержки в созданных программных продуктах.

Проблема структурирования данных, в частности, широко обсуждалась в инженерии знаний, основным объектом исследования которой являются знания, их извлечение, представление и т.п. В приложениях, использующих знания, существует разделение знаний на собственно знания и процедуры вывода на знаниях. Эти процедуры зависимы от представления знаний, а не от самих знаний, и могут разрабатываться и развиваться отдельно. Любая современная оболочка для создания экспертных систем предлагает (или навязывает) структуру представления знаний и обеспечивает соответствующими программами вывода (inference engine).

В последнее время из собственно знаний стали выделять уровень метазнаний (метаданных). Под метазнаниями понимается структура знаний в данном представлении, что представляет собой онтологию. Термин **онтология** был введен в искусственный интеллект Томом Грубером [3] как спецификация концептуализации. Подобно формальным спецификациям программных продуктов, онтология — это описание концептов и связей, которые могут существовать. Предполагалось, что онтологии будут использованы исключительно для облегчения взаимодействия интеллектуальных агентов. В начальном значении онтология — это формальный словарь агентов и соглашения по его использованию. Затем понятие онтологии было расширено. В частности, иерархическая структура классов в объектно-ориентированном программировании представляет собой онтологию.

Авторами [4] был предложен язык для представления онтологий KIF. KIF — это формальный язык с декларативной семантикой для обмена знаниями и взаимодействия между программами. Также широко проводятся работы по использованию имеющихся онтологий в различных системах автоматизированного извлечения знаний из экспертов. В работе [5] описан инструмент PROTEGE для автоматизированного построения предметно-ориентированных систем извлечения знаний. Он использует имеющуюся онтологию для построения диалога с пользователем, который будет вводить информацию в базу знаний, а также предоставляет среду для создания онтологий. Несмотря на то, что онтологии уже

Первая Всероссийская научная конференция  
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:  
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,  
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ  
19 - 21 октября 1999 г., Санкт-Петербург

нашли себе применение, работы по их созданию находят-ся на самой ранней стадии.

Применение онтологий при проектировании электронных библиотек и поиске данных в них позволит удобно организовать структурирование документов в электронной библиотеке и осуществлять поиск "по смыслу".

## 2 Применение онтологий для организации поиска информации в электронных библиотеках

Мы будем подразумевать под онтологией иерархическую структуру понятий (концептов). В отличие от классического определения, это понятие сужено до онтологии с определенным наполнением. В результате, онтология состоит из набора концептов и набора связей (ссылок) между ними и может быть представлена в виде набора мультидеревьев. Каждому концепту сопоставляется набор параметров (уровень, резюме, список документов). Уровень характеризует расстояние от головы дерева до данного концепта. В резюме можно кратко описать документы, на которые содержатся ссылки в под-онтологиях.

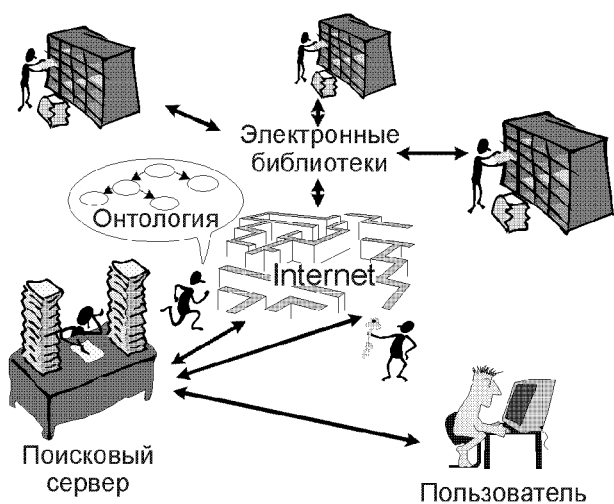


Рис. 1: Процесс поиска с использованием онтологий

Поиск необходимой информации, используя построенные унифицированные онтологии (задание подробных унифицированных онтологий позволит избежать хаоса при их развитии), схематически показан на Fig.1 и осуществляется следующим образом. Пользователь обращается к поисковому серверу, и в окошке браузера осуществляет поиск необходимой ему тематики (концепта) спускаясь и поднимаясь по мультидереву онтологии (движение может осуществляться по любому типу ссылок). Различные деревья, входящие в мультидереву, обозначаются разными цветами. В каждый момент времени пользователь видит на экране определенный концепт и все концепты, связанные с ним ссылками. В любой момент можно запустить процесс поиска, результатом которого являются все документы, соответствующие текущему концепту. Параллельно возможно осуществлять и поиск по ключевым словам в найденной совокупности документов или в документах под-дерева. Приписанное каждому концепту резюме поможет при беглом просмотре.

Мультидереву строится и дополняется в соответствии с онтологиями, построенными на web-сайтах. Поисковые сервера могут запускать в сеть агентов. Агент, сверяясь с имеющейся у него онтологией, сообщает на поисковый сервер об изменениях, произошедших с онтологиями исследуемых сайтов. Работа таких агентов избавляет от необходимости сообщать о всех изменениях онтологии каждого сайта. Достаточно, чтобы сайт поддерживал принятую онтологию. Если дополнения онтологии на нескольких web-северах совпадают, поисковый сервер может дополнить соответствующими концептами свою онтологию.

Каждый web-сервер может также строить свои вложенные онтологии, которые не отражаются в онтологии поискового сервера, и подключаются только при обращении к данному серверу. Работа с такими под-онтологиями осуществляется по тому же принципу. В результате на каждом сайте информация оказывается частично семантически структурирована, что облегчает работу с ней как web-мастеров, так и пользователей. Благодаря уменьшению просматриваемой информации, каждый поисковый сервер способен обрабатывать большее число web-сайтов, и поиск информации оказывается не только интеллектуальным, но и более быстрым.

Такой поиск, осуществляемый по всем библиотекам, приводит к выигрышу как времени (смотреть надо небольшое количество структурированной информации) так и качества (найденные данные структурированы и осмыслены).

## 3 ВИКОНТ

Для создания онтологий необходим удобный инструмент, облегчающий структурирование информации. Онтологии, создающиеся для того, чтобы сделать возможным совместное и многократное использование знаний, как ничто более нуждаются в продуманной и удобной структуре, а, следовательно, в инструментарии, облегчающем их построение и структуризацию.

Предлагаемый инструментарий ВИКОНТ основан на технологии, названной CAKE (Computer Aided Knowledge Engineering). Эта технология была с успехом применена при создании инструмента KEW (Knowledge Engineering Workbench) [1, 2]. Этот инструмент предлагает пользователю-инженеру по знаниям среду для быстрого создания прототипа экспертной системы и ее тестирования. Использование KEW не предполагает знание какого-либо формального языка. Пользователь "рисует" концептуальный граф предметной области, непосредственно взаимодействуя с экспертом, и строит правила-продукции, по которым машина вывода делает заключения, что позволяет непосредственно тестировать прототип экспертной системы.

ВИКОНТ также позволяет создавать онтологию, просто "рисую" ее структуру на экране. Структура онтологии представляется в виде полуйерархического семантического графа (кроме ссылок с верхнего уровня на нижележащие, возможны ссылки различных типов и на тот же и верхние уровни). Пользователь может задать набор типов связей в графе, что позволит создавать альтернативные онтологии, использующие одни и те же концепты (мультидеревья). На Fig.2 показаны этапы создания онтологии. Построенная онтология может быть переведена на формальный язык и использоваться другими приложениями. Например, онтология может быть формально записана в виде списка типа: [(концепт) (уровень) (ре-

зюме) (список документов) (дерево, (тип ссылки, список концептов)). В конце следует описание (резюме) каждого типа ссылок и мультидеревьев. Предлагаемый инструмент ВИКОНТ является простым и удобным средством для создания и развития онтологий.

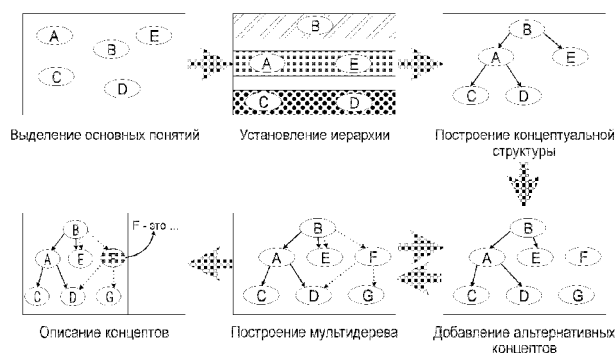


Рис. 2: Этапы создания онтологии

При работе с ВИКОНТом возможно два режима работы: графический и текстовый. Текстовый режим облегчает работу с «широкими» онтологиями (когда число ссылок велико). В графическом режиме на левой панели пользователем строится граф в виде концептов и ссылок (стрелочек). Может изображаться как вся онтология, так и определенное дерево, и любые под-деревья. В частности, может быть изображен выделенный концепт, и все концепты с исходящими и входящими ссылками. Также может включаться и отключаться режим изображения ссылок определенного типа. На правой панели изображается резюме к выделенному концепту. В текстовом режиме все связанные с выделенным концепты приводятся в виде списков по определенному типу ссылок.

Также ВИКОНТ может быть применен для подготовки электронных публикаций. Заполняя поля концептов, пользователь создает публикацию уже обладающую структурой и семантически осмысленную.

#### 4 Заключение

Построение электронных библиотек на основе онтологий позволит объединить множество сетевых ресурсов в единую систему, что значительно облегчит поиск необходимой информации. Благодаря заданию контекста поиска, найденная информация будет соответствовать той предметной области, которая требуется пользователю, и с развитием электронных библиотек на основе онтологий поиск будет становиться все более интеллектуальным и осмысленным.

#### Библиография

- [1] Т.А. Гаврилова, А.Ю. Золотарев, С.И. Еременко. *АРМ инженера по знаниям* // Сборник трудов III конф. по искусственному интеллекту. Тверь, 1992.
- [2] Т.А. Гаврилова, А.М. Яшин. *Некоторые аспекты методологии автоматизированного проектирования баз знаний* // Сб. "Научные труды СПбГТУ". СПбГТУ, 1995.

- [3] T. R. Gruber. *A translation approach to portable ontologies*. Knowledge Acquisition, 5(2): 199–220, 1993.
- [4] Anonymous. *Система Представления Знаний Ontolingua* [<http://ksl.stanford.edu>]
- [5] H. Eriksson, A. R. Puerta, M. A. Musen. *Generation of Knowledge-Acquisition Tools from Domain Ontologies*, 1994. [[http://ksl-web.stanford.edu/KSL\\_Abstracts/KSL-93-56.html](http://ksl-web.stanford.edu/KSL_Abstracts/KSL-93-56.html)]