

УДК 004.738

*Т.А. Гаврилова, Т.Е. Гелеверя, В.А. Горовой*

Санкт-Петербургский государственный политехнический университет, Россия

## Онтологии как средство концептуализации web-порталов

В работе рассматриваются возможности использования онтологий в качестве средств концептуализации web-порталов. Приводятся различные аргументы в пользу данного подхода. Предлагается программный инструментарий, призванный помочь системным аналитикам использовать онтологии для проектирования web-порталов.

### Введение

В последние годы все большую часть пространства Интернет завоевывают крупные и средние web-порталы – интернет-сайты, снабженные механизмами избирательного поиска и объединения информационного содержания интернет/интранет-сетей. Однако по мере быстрого накопления информации пользователи сети столкнулись с проблемой практически полного отсутствия структуризации, что затрудняет как поиск, так и логическое осмысление содержания. К сожалению, довольно часто приходится сталкиваться с тем, что очень сложно, а иногда и почти невозможно найти нужную информацию даже в пределах одного портала. Часто этот поиск превращается в многочасовой марафон, который к тому же не гарантирует получение ответов на поставленные вопросы. Это, как правило, является следствием слабой структурированности интернет-сайтов. А ведь одна из основных характеристик любого портала – это возможность легко и быстро добраться до любой информации, содержащейся в нем. Конечно, можно воспользоваться картой сайта, которая, как правило, имеется на любом крупном портале, но, к сожалению, это не всегда помогает в поиске.

В качестве примера неудачно структурированного сайта можно привести <http://www.nwgs.com>. Мало кому приходит в голову, что служба отправки sms может находиться в подразделе раздела «О компании» (рис. 1.).

К сожалению, это далеко не единичный случай плохо организованного web-портала, он лишь подтверждает общую тенденцию – существующая на сегодня интернет-технология, включая порталы, заставляет пользователя работать в слабоструктурированной среде, перекладывая на него основной груз перехода от примитивных к более сложным формам и интеллектуальным методам работы с сетевой информацией.

Следовательно, самым острым образом встает вопрос о концептуализации и структуризации web-порталов.

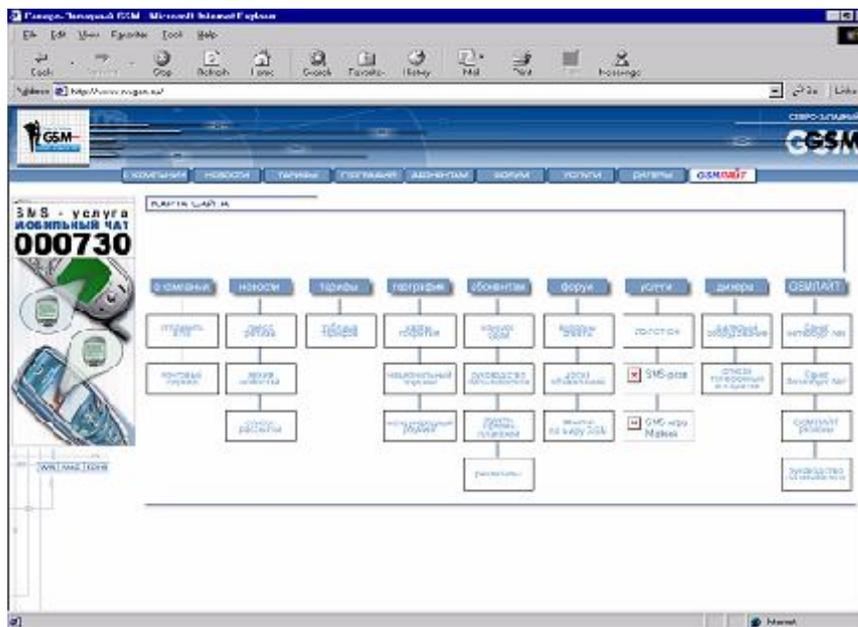


Рис. 1. Пример неудачно структурированного сайта

## 1. ОНТОЛОГИИ

Искусственный интеллект всегда служил плацдармом новых идей и понятий для других ветвей информатики. Так и сейчас – достаточно активно развивающееся новое направление – онтологии используются в смежных с ИИ областях. Их применение оказалось очень полезным и актуальным в knowledge management [1], в поддержке процессов обработки естественного языка: английского, немецкого и итальянского (система Generalized Upper Model [2] и система Plinius [2], предназначенная для полуавтоматического извлечения знаний из текстов в области химии). Также онтологии можно использовать как средство приобретения и оценки полученных знаний.

Существует множество различных подходов к определению понятия «онтология»:

- онтологией называют эксплицитную спецификацию концептуализации [3].
- онтология – эксплицитная спецификация определенной темы [4].
- онтология – это базы знаний специального типа, которые могут «читаться и пониматься, отчуждаться от разработчика и/или физически разделяться их пользователями» [5].

Следует отметить, что понимание термина «онтология» зависит от контекста и целей его использования. Можно выделить несколько аспектов его применения, в частности, в работе [Guarino, et al., 1995a] выделяются следующие интерпретации:

1. Онтология как философская дисциплина.
2. Онтология как неформальная концептуальная система.
3. Онтология как формальный взгляд на семантику.
4. Онтология как спецификация концептуализации.
5. Онтология как представление концептуальной системы через логическую теорию.

6. Онтология как словарь, используемый логической теорией.
7. Онтология как (метауровневая) спецификация логической теории.

Резюмируя все вышесказанное, можно сказать, что онтология – это точная спецификация некоторого предмета, которая включает в себя словарь терминов предметной области и множество логических связей, которые описывают, как эти термины соотносятся между собой.

К сожалению, при использовании существующих онтологий возникает множество трудностей; в работе [7] выделяются, в частности, следующие:

- отсутствие стандартизованных идентифицирующих особенностей, которые характеризовали бы онтологии с точки зрения пользователя;
- разный уровень детализации онтологий, находящихся на одном сервере;
- отсутствие web-сайтов, использующих одинаковую логическую структуру и предоставляющих релевантную информацию об онтологиях;
- поиск подходящих онтологий сложен, занимает много времени, а также часто безрезультатен.

Эти трудности часто приводят к тому, что под конкретные задачи подчас невозможно найти подходящую онтологию из числа существующих, поэтому приходится браться за создание новой, что, естественно, является сложным и достаточно дорогостоящим процессом. Кроме того, использование готовых онтологий непосредственно людьми (в отличие от использования программными агентами) обладает еще рядом недостатков. В частности, знания разных людей могут укладываться в разные онтологии, при этом нельзя утверждать, что одна из них лучше другой. Также во многих случаях для одной предметной области (в особенности для плохо формализуемых предметных областей или в случае наличия множества спорных вопросов) можно построить несколько различных онтологий. Например, историю можно воспринимать как хронику последовательных событий, а можно как жизнь различных людей, в этих событиях участвующих. Формально это две различные онтологии: в первом случае в качестве концептов выступают исторические даты, а во втором – личности. Но по сути, оба пути полно описывают одну и ту же предметную область.

## 2. Использование онтологий в качестве инструмента концептуализации web-порталов

Подводя некоторые итоги, можно утверждать, что web-порталы нуждаются в инструментах структуризации и концептуализации их информационного наполнения. Использование онтологий в качестве одного из таких инструментов имеет ряд достаточно серьезных преимуществ.

Визуализация онтологий позволяет сделать видимыми семантические сети памяти человека. Эти сети состоят из узлов – различных концептов, а также из различных связей или соотношений, которые описывают взаимодействия между ними. В связи с этим разработка семантических сетей подразумевает глубокий анализ структурных взаимодействий между этими концептами. В свою очередь этот анализ позволит разработчику более оптимально структуризировать web-сайт, так как он будет ясно представлять себе, как связаны между собой отдельные узлы семантической сети (в частности, как связаны отдельные разделы

портала друг с другом). Еще одно преимущество использования онтологий в качестве инструмента концептуализации web-порталов является системный подход к вопросам, связанным с их проектированием.

В простейшем случае построение онтологии сводится к:

- выделению концептов – базовых понятий данной предметной области;
- построению связей между концептами – определению соотношений и взаимодействий базовых понятий;
- сравнению построенной онтологии с имеющимися – проведению параллелей с другими областями знаний.

К сожалению, в настоящее время не так много разработчиков используют онтологии в процессе создания web-порталов, поэтому провести сравнение построенной онтологии с имеющимися будет достаточно сложно в связи с нехваткой последних.

Имеется множество различных подходов и методологий по созданию онтологий. Большинство из них базируется на следующих принципах, предложенных Грубером [3]:

1. Ясность (Clarity) – онтология должна ясно передавать смысл введенных терминов (они не должны вызывать чувство когнитивного диссонанса у пользователя).
2. Согласованность (Coherence) – онтология не должна быть логически противоречивой.
3. Расширяемость (Extendibility) – онтология должна допускать возможность расширения.
4. Минимум влияния кодирования (Minimal encoding bias) – концептуализация, лежащая в основе создаваемой онтологии, должна быть специфицирована на уровне представления, а не символического кодирования.
5. Минимум онтологических обязательств (Minimal ontological commitment) – онтология должна содержать только наиболее существенные предположения о моделируемом мире.

В качестве примера методологии создания онтологий можно привести так называемую «Enterprise Methodology». Жизненный цикл при использовании этой методологии состоит из следующих стадий:

1. Определение целей и намерений.
2. Построение онтологии:
  - *Ontology Capture* – выделение основных концептов и отношений между ними; выработка недвусмысленных определений; определение терминов для ссылки на эти концепты и отношения.
  - *Ontology Coding* – фиксирование метаонтологии; выбор языка представления; кодирование.
  - *Integrating Existing Ontologies* – как использовать и использовать ли уже готовые онтологии.
3. Оценка (evaluation).
4. Документирование.
5. Создание руководства для каждой из этих фаз (Guidelines for each phase).

В общих чертах процесс разработки онтологии показан на рис. 2.

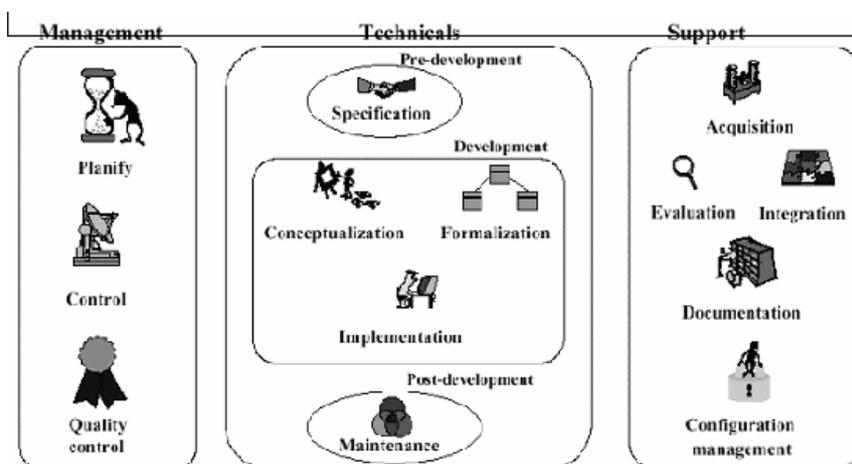


Рис. 2. Процесс разработки онтологии в рамках подхода METHONTOLOGY

В рамках этого подхода в «жизненном цикле» создания онтологий выделяются следующие процедуры: *управление проектом, собственно разработка и поддержка разработки.*

**Процедуры управления проектом** включают *планирование, контроль и гарантии качества.* Собственно **разработка** включает *спецификацию, концептуализацию, формализацию, реализацию и техническую поддержку.*

**Процедуры поддержки** включают действия, выполняемые одновременно с разработкой, без которых онтология не может быть построена: *процедуры приобретения знаний, оценки, интеграции, документирования и управления конфигурациями.*

Естественно, использование онтологий в качестве инструмента концептуализации несколько увеличит стоимость web-портала, но это как раз тот случай, когда расходы обоснованны.

В процессе использования различных ресурсов сети часто приходится сталкиваться с неравномерной иерархической вложенностью различных разделов (например, в одном из разделов содержатся ссылки на 10 подразделов, а в другом – только на 2). Это пример не очень хорошего проектирования портала и, как следствие, его плохой структурированности. Использование онтологий в качестве средства концептуализации позволит выявить эти проблемы на ранней стадии.

В проекте УЗОР нами предложен программный инструмент для проектирования порталов на основе визуализации онтологий как развитие методологии CAKE (Computer-Aided Knowledge Engineering), последовательно разрабатываемой с 1995 г. совместно с А.В. Воиновым [8], [9].

CAKE-2 представляет собой средство для визуального проектирования концептуальной модели предметной области, определения базы знаний на основании имеющихся концептов, а также работы с этой базой знаний. Основная идея системы – предоставить пользователю (инженеру по знаниям) полнофункциональное и легкодоступное средство, позволяющее решать основные задачи инженерии знаний (рис.3)

Созданный инструмент предоставляет разработчику портала (сайта и т.д.) инструмент, позволяющий визуально формировать структуру концеп-

туального базиса, а также определять связи между концептами. Инструментарий является, с одной стороны, CASE-средством, позволяющим автоматизировать некоторые этапы разработки, с другой стороны, это mindtool, облегчающий процесс структурирования и формализации информации.

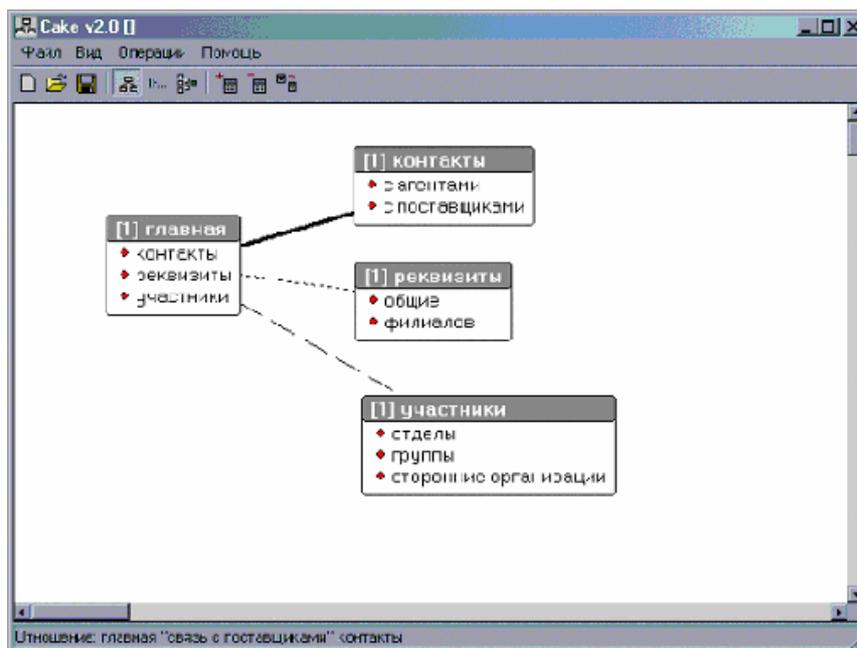


Рис. 3. Фрагмент онтологии корпоративного портала

Дальнейшее развитие инструментария CAKE, применительно к web-порталам и онтологиям, заключается в определении необходимой и достаточной функциональности, позволяющей разработчику самостоятельно решать задачи, входящие в состав цикла разработки портала.

## Заключение

Онтология может явиться удобным средством концептуализации web-порталов. Современные средства визуализации позволяют существенно упростить процесс их создания. Использование онтологий также позволяет выявить многие «узкие места» на стадии проектирования web-порталов и позволяет избежать большинства ошибок, являющихся следствием их плохой структурированности.

## Литература

1. Гаврилова Т.А., Воинов А. Визуальный концептуальный анализ знаний и данных для интеллектуальных и обучающих систем // VI Междунар. конф. «Знание – Диалог – Решение» KDS-97. – Ялта. – 1997. – С.1-8.
2. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем // СПб: Питер 2000.
3. Arpírez J., Gómez-Pérez A., Lozano A., Pinto S. (ONTO) 2Agent: An ontology-based WWW broker to select Ontologies // Workshop on Applications of ontologies and Problem Solving Methods «ECAI'98».

4. Gavrilova T., Voinov A. Visualized Conceptual Structuring for Heterogenous Knowledge Acquisition // Proc. of Int. conf. on Educational Multimedia and Hypermedia «EDMEDIA'96». – MIT. – Boston (USA). – 1996. – P. 258-264.
5. Gómez-Pérez A. Knowledge Sharing and Reuse // In the Handbook of Applied Expert Systems. – CRC Press, 1998.
6. Gruber T. A translation Approach to portable ontology specifications // Knowledge Acquisition. – Vol. 5. – 1993. – P. 199-220.
7. Guarino N., Carrara M., Giaretta P. Formalizing Ontological Commitments // 12th National Conf. on Artificial Intelligence «AAAI-94». – 1994. – P. 560-567.
8. Guarino N., Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing // IOS Press. – 1995. – P. 25-32.
9. Kühn O., Abecker A. Corporate Memories for Knowledge Management in Industrial Practice: Prospects and Challenges // IOS Press. – 1998.
10. Uschold M., Grüninger M. ONTOLOGIES: Principles, Methods and Applications // Knowledge Engineering Review. – Vol. 11, № 2. – 1996, June.

Paper presents ontologies as a tool for web-portals conceptualization. The stress is put on different aspects of this approach. Described ideas are implemented in computer programs, that may be used as special ontological engineering tools, helping experts to conceptualize web-portals. The advantages of using ontologies for construction of web-portals are discussed.

*Статья поступила в редакцию 10.07.02.*