

РАСПРЕДЕЛЕННАЯ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННАЯ СИСТЕМА ДЛЯ РАБОТЫ С ОНТОЛОГИЧЕСКИМИ МОДЕЛЯМИ ЗНАНИЙ

А. Ф. ТУЗОВСКИЙ

Институт “Кибернетический центр” ТПУ,

Отдел проблем информатизации ТНЦ СО РАН, Томск, Россия

e-mail: TuzovskyAF@kms.cctpu.edu.ru

A special structure of the knowledge ontology model is suggested for modeling of the diverse knowledge resources with application to an enterprise. It is based on the top level enterprise ontology and hierarchy of the domain ontologies. An architecture of the distributed information and communication software system for support of such ontology models is developed. The system allows updating ontology models and supports authorship, version and access control opportunities.

Введение

Управление знаниями становится важным видом деятельности современных организаций. Для этого разрабатываются специальные системы управления знаниями, являющиеся дальнейшим развитием информационных систем организаций [1]. Наиболее перспективен в этом направлении онтолого-семантический подход [2, 3], основанный на онтологическом моделировании знаний организации и семантических методах работы с ней. В настоящее время основное внимание в области онтологического моделирования уделяется методам разработки и описания онтологий [4]. Однако поскольку онтологические модели знаний организации включают большое количество понятий (терминов) и разрабатываются специалистами по разным направлениям деятельности, требуется создание специальных информационно-телекоммуникационных систем, поддерживающих распределенную и коллективную работу с такими моделями.

1. Онтологическая модель организации

Все знания организации содержатся во множестве разнообразных ресурсов *КО* (документы, файлы, базы данных, опыт специалистов, внешние источники, на которые есть

ссылки): $KO = \{KO_1, \dots, KO_n\}$. Знания, связанные с объектом KO_i организации, описываются в виде модели M_i , которая может иметь разный уровень формализации, начиная от записи в виде текста на естественном языке (например, в виде краткой текстовой аннотации) и заканчивая набором утверждений логики первого порядка. Каждый из ресурсов содержит только некоторую часть (фрагмент) всех знаний, важных для работы организации. Описание моделей знаний ресурсов включает термины, используемые специалистами организации (и соответствующие им понятия), и отношения между ними. Так как знания, содержащиеся в отдельных ресурсах, относятся к одной и той же организации, они должны быть согласованными. Для согласования моделей описания ресурсов их необходимо поместить в более общий контекст, т. е. следует построить некоторую общую модель знаний организации. Общая модель знаний должна включать понятия и отношения между ними, используемые в отдельных моделях ресурсов, что позволит согласовать и систематизировать знания, с которыми выполняется работа в системе управления знаниями организации.

Построение единого описания знаний организации выполняется с помощью онтологического моделирования [3]. Под онтологией O в данной работе понимается знаковая система $\langle C, T, R, F, L, A \rangle$, где:

- C — множество элементов, которые называются понятиями;
- T — частичный порядок на множестве C , задающий отношения “подкласс” и “суперкласс”;
- R — множество элементов, которые называются свойствами (двуместными предикатами);
- F — функция, которая назначает каждому элементу множества R множество элементов из множества C (с учетом их иерархии в T), к которым оно применимо (область действия, область, домен), и множество элементов из множества C или *литералов* (экземпляров примитивных типов, таких как строки и числа), которые могут быть их значениями (область возможных значений, интервал, ранг);
- $L = \{L^C, L^R, \alpha^C, \alpha^R\}$ — множество текстовых меток, которые определяют профессиональные термины организации и их соответствие, соответствие α^C — элементам множества C , α^R — элементам множества R ;
- A — набор аксиом онтологии — утверждения об элементах предметной области, которые считаются верными, выраженных с использованием соответствующего логического языка.

Создание единой онтологии для достаточно детального описания модели знаний организации — весьма трудоемкая задача. Речь идет о длительном проекте с привлечением коллектива экспертов, знания которых должны охватывать все направления деятельности исследуемой компании. Решением этой проблемы может быть отказ от глубокой декомпозиции системы и включение в онтологию только наиболее значимых понятий из рассматриваемых предметных областей или выделение одного из направлений деятельности организации и создание для этого направления детальной, но узкоспециализированной онтологии. Очевидно, что оба подхода имеют свои недостатки. Первый подход дает слишком грубую и обобщенную модель, а второй не позволяет использовать модель для взаимодействия всех подразделений компании.

Для решения задачи построения онтологии знаний организации структурированную онтологическую модель предлагается разрабатывать следующим образом [5]: $O = \{O_{oo}, O_{oz}\}$, где O_{oo} — онтология организации, а $O_{oz} = \{O_1, \dots, O_m\}$ — иерархически организованная, последовательно расширяемая система онтологий основных областей

знаний O_i , значимых для работы организации. Выделение иерархии областей знаний организации дает возможность создавать отдельно онтологии разных подобластей знаний, которые могут иметь разную детальность в зависимости от потребностей их моделирования.

Онтология организации O_{oo} включает основные понятия, которые описывают структуру, состав элементов и работу организации (ПОДРАЗДЕЛЕНИЯ, СПЕЦИАЛИСТЫ, ЗАКАЗЧИКИ, ПРОЕКТЫ, БИЗНЕС-ПРОЦЕССЫ и пр.). Онтологии предметных областей включают описание понятий, связанных с ними терминов и отношений между ними. Структура онтологической модели знаний организации показана на рис. 1.

Можно выделить следующие основные проблемы, возникающие при использовании подобной онтологической модели организации:

- централизованное определение и сопровождение онтологической модели практически невозможны, так как ни одна группа специалистов не может выполнять описание всех знаний организации;
- онтологическая модель знаний организации никогда не будет полной и стабильной, в связи с этим необходимо обеспечение обратной совместимости развивающейся модели;
- существующие структуры данных и табличные схемы реляционных баз данных (БД) активно используются в организациях, и поэтому маловероятно, что онтологическая модель, которая не сможет использовать работающие структуры данных и существующие схемы реляционных БД, будет привлекательной для практического использования.

В связи с этим возникает необходимость разработки системы коллективной работы с онтологической моделью знаний организации. На основе анализа имеющихся публикаций [2, 5, 6] и проведенных исследований разработана архитектура такой системы.

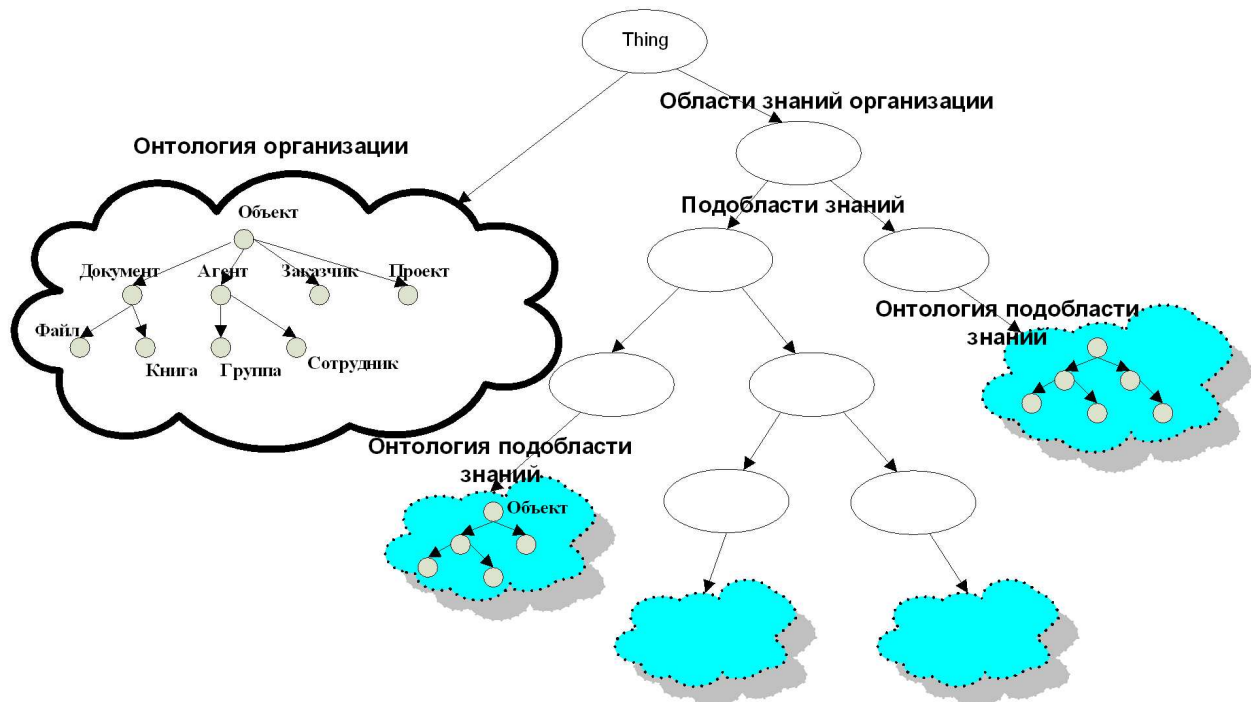


Рис. 1. Структура единой онтологической модели знаний организации

2. Распределенная онтологическая модель

Управление распределенной онтологией требует решения задач, которые не появляются с онтологией, управляемой одним владельцем. При работе с набором онтологий возникают следующие проблемы:

— *избыточность*. Более чем один участник может предоставить альтернативные модели одной и той же части знаний организации. Распределенная онтология позволяет таким альтернативам сосуществовать и предоставляет инструменты участникам для обсуждения использования онтологии и организации взаимосвязей между альтернативными моделями;

— *надежность*. Так как глобальная онтология является открытой, она может включать сеть онтологических элементов, полученных от специалистов с различными уровнями доверия. Предлагаемая распределенная онтология позволяет оценивать и фильтровать элементы онтологической модели на основе заданных уровней доверия к их создателям;

— *изменяемость*. Организация, используемые в ней термины/понятия (классы обозначаемых ими сущностей) и взаимосвязи между ними постоянно изменяются. Предлагаемая распределенная онтология включает инструменты для расширения (развития) онтологии с поддержкой обратной совместимости;

— *соответствие стандартам*. Существующие документы и данные, такие как XML Schema и схемы реляционных БД, уже содержат онтологическую информацию, хотя не соответствуют требованиям формальной онтологической модели.

Общая онтологическая модель (рис. 2) может располагаться в одной онтологической базе знаний (ОБЗ) или быть распределенной, т. е. содержаться в нескольких ОБЗ.



Рис. 2. Схема распределенной системы по работе с онтологической моделью

С каждой ОБЗ взаимодействует сервер онтологии. В составе системы также имеется центральный компьютер, который содержит дерево областей знаний (глобальный онтологический каталог — ГОК) и прикладной интерфейс (API) для поиска и решения базовых задач по работе с онтологической моделью. Центральная ОБЗ содержит онтологическую модель верхнего уровня и онтологические модели предметных областей. Отдельные онтологические модели предметных областей могут размещаться в распределенных ОБЗ (в этих БЗ есть указание на адрес центрального сервера онтологий).

Подобная система позволяет разным группам участников публиковать разные определения понятий и отношений в распределенной системе и в отдельные моменты времени их обновлять. Более того, отношения могут использовать типы различных владельцев как предметные области, тем самым создавая сеть.

Онтологические базы знаний реализуются в виде специальной базы данных на основе реляционных СУБД, в которой содержится описание онтологической модели и метаданных всех информационных ресурсов. Она также включает программное обеспечение для реализации таких наборов методов, как:

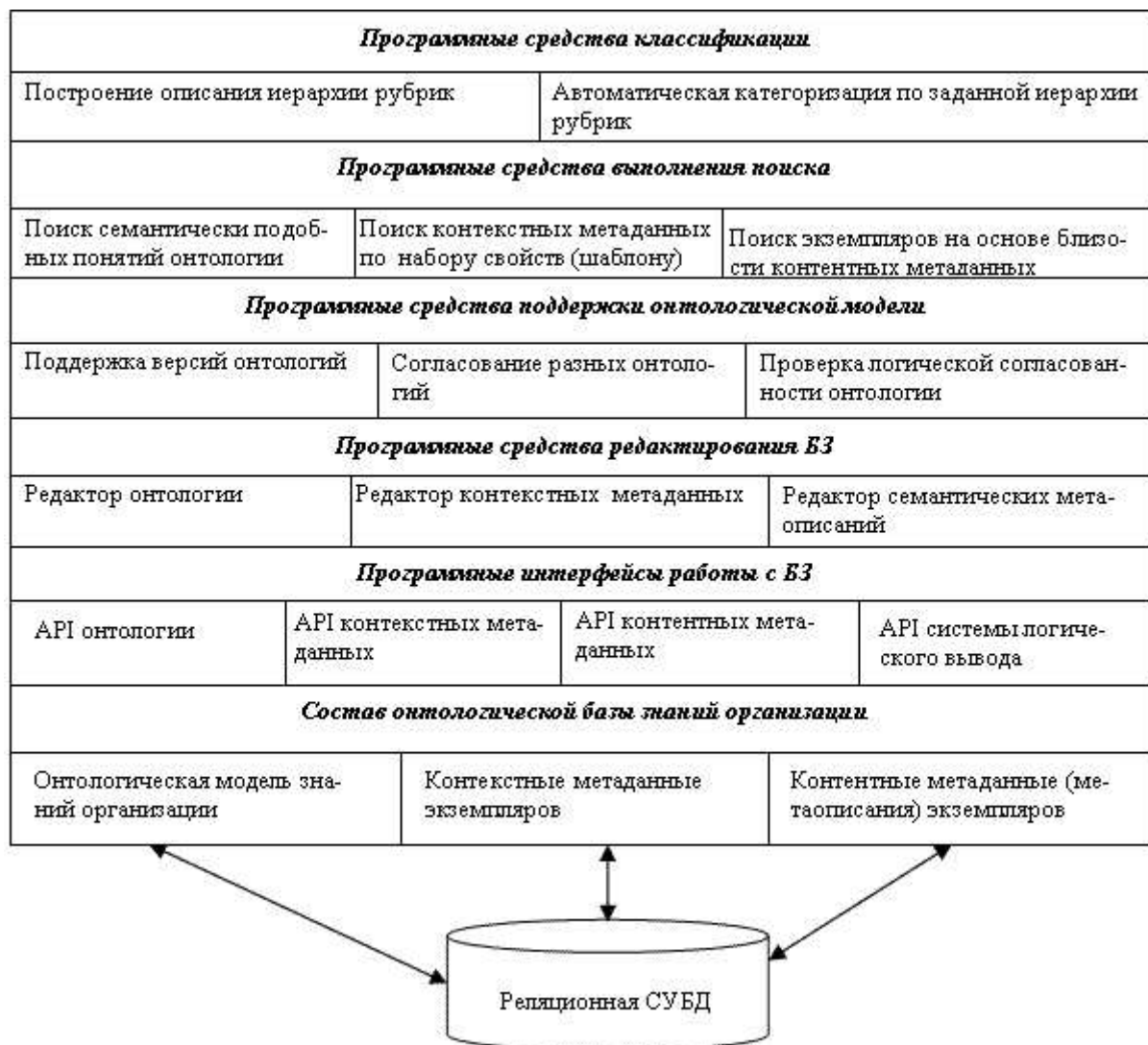


Рис. 3. Многоуровневая схема реализации онтологической базы знаний организации

- методы развития и поддержки единой онтологической модели знаний организации;
- методы описания объектов, содержащих знания (контекстных и контентных метаданных);
- методы работы с описаниями объектов, содержащих знания (контекстными и контентными метаданными).

Многоуровневая схема реализации онтологической базы знаний организации и программного обеспечения работы с ней показана на рис. 3 [7].

Содержание каждой ОБЗ делается доступным с помощью серверов онтологий, которые также содержат подсистему публикации внесенных изменений в ГОК. С помощью компонента публикации сервер может заносить определения понятий и отношений его с ОБЗ в глобальный онтологический справочник, который размещается на центральном компьютере, и тем самым предоставлять их всем пользователям для рассмотрения.

Онтологический справочник поддерживает мастер-каталог со ссылками на понятия и отношения в различных онтологиях, размещенных в хранилищах. Мастер-каталог в глобальном онтологическом справочнике может быть распределен иерархически между множеством уровней подобластей знаний. Описания подобластей знаний могут размещаться как на центральном компьютере, так и на разных компьютерах. После выполнения изменений отдельной онтологией глобальный онтологический справочник корректирует свой каталог для сохранения измененных понятий и отношений.

Серверы онтологий могут размещаться на разных компьютерах локальной сети или находиться на том же компьютере, где находится и центральный сервер.

Глобальный онтологический справочник может быть реализован следующими способами:

- с помощью стандартной поисковой системы, которая индексирует текстовые метки понятий и отношений по ключевым словам;
- с помощью программы (на языках C++/C#), которая централизованно воссоздает всю сеть онтологий с помощью объектов;
- с помощью реляционной БД с таблицами для понятий, отношений и наследования, которая содержит области определения для отношений, области определений и области значений для функций, а также подклассы и суперклассы для наследования.

При этом отношения индексируются по именам, областям определения и значений, а наследование индексируется по подклассам и суперклассам.

В данной системе работы с онтологической моделью также имеется набор программных средств (инструментов) для взаимодействия пользователей с онтологией с компьютеров клиентов. Данный набор инструментов позволяет выполнять поиск требуемых элементов онтологической модели в глобальном онтологическом справочнике. После определения конкретного сервера онтологий, который содержит требуемую информацию, пользователь использует набор инструментов для выполнения запросов к ОБЗ и получения подробной информации об интересующих его понятиях или отношениях. Таким образом, предлагаемая архитектура работы с онтологической моделью расширяет абстрактную онтологическую модель до модели распределенной онтологии.

Распределенная онтология позволяет построить онтологическую модель на основе сети онтологий, представленных распределенными наборами определений понятий и отношений. В целом распределенные наборы описаний понятий и отношений определяют глобальную онтологическую модель.

Базы знаний (БЗ) онтологической сети могут быть связаны друг с другом. Таким образом, понятие, описанное в одной БЗ, может быть подклассом некоторого понятия (на который делается ссылка), описанного в другой базе знаний. Аналогично отношение, определенное в одной БЗ (отношение в виде ссылки на объект), может иметь в качестве области определения понятие, определенное в другой БЗ, а в качестве области значений — понятие, описанное в третьей БЗ. В соответствии с данной архитектурой такие ссылки могут встречаться независимо от того, располагаются ли базы знаний на одном компьютере или физически распределены по разным компьютерам.

В данной работе термин “описание понятия” используется как общий способ описания понятий. Описание понятия является содержимым БЗ, включающим обязательные поля (описание на естественном языке элементов, которые относятся к данному понятию) и необязательные поля (текстовая метка понятия; набор ссылок на понятия более высокого уровня; набор ссылок на подклассы (производные понятия); ограничения (например, понятия, с которыми данное понятие не пересекается)).

Особенностью данной системы является возможность выполнять только такие корректировки описания онтологии, которые не нарушают обратную совместимость. Это гарантирует, что словарь, верный (правильный) в рамках данной онтологии, будет оставаться правильным (верным) по отношению к будущим изменениям (развитию) онтологии. Данная система использует набор правил [6], которые позволяют выполнять только “безопасные” изменения в онтологии, не приводящие к нарушению ранее сделанных метаописаний ресурсов знаний. К таким правилам относятся следующие.

1. Онтология может быть изменена, но должна сохранять обратную совместимость. В результате имеется возможность добавления новых понятий и отношений, добавления отношений наследования суперкласс/подкласс, расширения существующих отношений и функций. Это позволяет усовершенствовать онтологическую модель без нарушения доступного словаря.

2. Описание понятия $c2$ является “обновлением” понятия $c1$, если множество суперклассов и множество подклассов в $c2$, соответственно, содержит множество суперклассов и подклассов в понятии $c1$, а описание и текстовая метка остаются неизменными. Как отмечалось выше, такой способ обновления дает возможность динамически развивать и расширять онтологию, не нарушая обратной совместимости. В частности, определение понятия может обновляться таким образом, что включает новые отношения суперкласс/подкласс и новые определения понятий и отношений могут быть добавлены к онтологии без нарушения обратной совместимости.

3. Аналогично определению понятий используется “определение отношения” как обобщенный способ описания отношения. Описание отношения $r2$ является “обновлением” описания отношения $r1$, если понятия, входящие в область определения (домен) $r2$, являются суперклассами соответствующих понятий, составляющих область определения $r1$, и в области определения $r2$ и $r2$ включают $r1$. Как указывалось выше, данный способ обновления позволяет онтологии динамично развиваться и расширяться, не нарушая обратной совместимости.

В каждый конкретный момент времени глобальная онтологическая модель представляется сетью понятий и отношений, описанных в различных базах знаний и доступных с помощью серверов онтологий.

Прикладной интерфейс доступа к отдельным онтологиям реализуется с использованием технологии вызова удаленных процедур (remote procedure calls), например технологии .Net Remoting, для вызова соответствующих функций/методов в справочниках и

серверах, или с помощью запроса онтологической информации от подходящего сервера, например, используя стандартный протокол HTTP, а затем выполняя его грамматический разбор с использованием DOM или SAX интерфейса XML-анализатора (parser) для выделения нужной информации.

3. Основные сервисные подсистемы поддержки работы с онтологической моделью

Кроме средств хранения и доступа к элементам онтологий распределенная система включает:

- подсистему поддержки авторской ответственности;
- подсистему управления версиями элементов онтологической модели;
- подсистему управления правами доступа.

3.1. Подсистема поддержки авторской ответственности

Модель авторской ответственности используется для получения информации об авторах понятий или отношений, для определения степени доверия к элементам онтологии и получения ответов на такие вопросы, как: является ли автор широко признанным специалистом, чье определение, вероятно, будет признано руководством; является ли определение ясно сформулированным и будет ли определение понятия/отношения сохраняться стабильным? Модель метаданных для управления правами авторства на каждое понятие и отношение включает специальные метаданные: имя специалиста (подразделения), ответственного за содержание документа (авторство); цифровую подпись специалиста (подразделения) на документе с описанием понятия/отношения с использованием одного из общепризнанных стандартов, например, такого, как стандарт на XML-документ; сертификат или цепочку сертификатов, подписанных признанным центром сертификации авторства, который связывает цифровую подпись с конкретным автором.

Данная модель авторства позволяет пользователю указать авторов, которым он доверяет. Более того, могут поддерживаться группы пользователей на основе задания для групп “политики доверия” путем перечисления других групп, которым они доверяют. Описание групп является рекурсивным таким образом, что одна группа может включать другие группы.

Используя фильтрацию элементов онтологической модели на основе политики доверия к авторам, модель позволяет формировать подсети онтологий. Такая подсеть онтологий доступна пользователям на основе заданной ими политики доверия.

Программа редактирования онтологий включает набор программ с графическим интерфейсом для создания описаний понятий и отношений. Она позволяет просматривать сеть онтологий и добавлять к ней новые понятия и отношения. В основном набор инструментов включает прикладной интерфейс программирования (API) для поиска понятий и доступа к ним, включая поиск понятия по любой заданной комбинации параметров (ключевое слово из определения, суперкласс, подкласс, автор или подразделение в цепочке сертификации). Используя понятие, можно получить доступ к суперклассам, подклассам, отношениям, определенным для этой предметной области, определению.

Прикладной программный интерфейс также включает и аналогичные средства для поиска и доступа к отношениям.

Учитывая, что более чем один специалист может предоставить альтернативные модели одной и той же части знаний организации (предметной области), в предлагаемой системе включено отношение “эквивалентности”, аналогичное отношению наследования, для выявления понятий, которые являются эквивалентными.

Два понятия эквивалентны, если имеется взаимно-однозначная функция с одним понятием в качестве области определения и с другим понятием в качестве области значений. В частности, если два понятия эквивалентны, то экземпляры одного понятия являются эквивалентами экземпляров второго понятия.

3.2. Подсистема управления версиями

Подсистема управления версиями элементов онтологии позволяет отслеживать развитие онтологической модели. Управление версиями предоставляет общую модель, которая применяется к онтологии, для отслеживания изменений в модели и исключения конфликтов между изменениями, что дает возможность отслеживать постепенное развитие модели. Модель управления версиями напрямую встраивается в программные средства работы с онтологиями. В общем случае управление версиями полезно для проектов и моделей, в которых модель может быть разделена на множество компонентов (для онтологической модели компонентами являются “понятия”, отношения, наследование) и разные пользователи могут редактировать разные части модели.

Управление версиями имеет следующие цели:

- гарантировать, что не произойдет конфликта между двумя редактированиями, отслеживать, чтобы не выполнялось редактирование одного и того же компонента одновременно двумя пользователями (пессимистический контроль версий);
- разрешать, чтобы проверку временных изменений производил специалист, выполняющий редактирование, незаметно для других пользователей;
- отслеживать изменения;
- поддерживать всю историю изменений каждого компонента;
- поддерживать информацию о том, кто сделал каждое изменение, его комментарии о целях изменения и информацию об изменении, которое влияет на множество компонентов;
- задавать специальные моменты времени, когда всем компонентам содержимого (контента) задается номер версии.

Задание номеров версий основывается на следующих принципах: детальное управление версиями на уровне компонентов, которые совпадают с понятиями; пессимистический контроль версий — не поддерживается одновременная работа нескольких пользователей с одним и тем же компонентом, что необходимо для поддержки детального управления версиями; линейное задание версий — не поддерживается ветвление версий при их контроле.

3.3. Подсистема управления правами доступа

Управление правами доступа предоставляет общую модель, которая применяется к онтологии, для управления назначением прав в системе с большим количеством ресурсов, пользователей и типов прав доступа.

Элементы онтологической модели являются типами ресурсов. В принципе, управление доступом может разрешать каждому владельцу ресурсов задавать для всех своих ресурсов и пользователей списки привилегий доступа, которые разрешены данному пользователю по отношению к данному ресурсу. Однако это быстро приведет к состоянию неуправляемости. Для упрощения управления доступом вводится понятие “группа пользователей”. Это набор пользователей или других групп пользователей, который задается время от времени владельцем группы пользователей. Пользователи объединяются в группы пользователей (например, “специалисты компании”, “сотрудники конструкторского отдела”) с некоторыми общими свойствами или которым назначается некоторый общий набор привилегий. Владелец группы пользователей является некто, кому доверяют, по крайней мере, некоторые владельцы надежное управление группой пользователей, чтобы он доверял назначение привилегий всей группе.

Разрешено рекурсивное использование групп пользователей, т.е. включение группы *A* в качестве члена группы *B* означает, что все члены группы *A* обязательно удовлетворяют критерию, чтобы сделать их членами группы *B*. Например, “сотрудники компании” могут быть членами группы пользователей “сотрудники конструкторского отдела”. Тогда назначение привилегий группе *B* предполагает назначение привилегий членам группы *A*. Владелец группы *B* при назначении группы *A* в качестве члена своей группы также отслеживает, кого владелец группы *A* признает в качестве членов.

Уровень управления доступом дает возможность создать и управлять группами пользователей. Кроме этого он дает возможность любому владельцу ресурсов создавать и управлять группами ресурсов, а также позволяет любому владельцу ресурсов управлять доступом к каждой группе ресурсов.

Возможности доступа к элементам онтологической модели (в результате вызова методов API) сгруппированы следующим образом: *просмотр* того, что имеется; *чтение* (включает просмотр); *обновление* (включает чтение плюс захват (*check-out*) и некоторое редактирование) — выполнение изменений, которые сохраняют обратную совместимость; *запись* (включает обновление); *все операции* (включает запись плюс удаление).

Упрощенная схема архитектуры среды выполнения предлагаемой системы показана на рис. 4. Данная архитектура предоставляет средства для совместной работы и управления всеми участниками онтологической модели.

Для онтологической модели в оперативной памяти (RAM) также создается соответствующая ей DOM-модель. Основная копия (оригинал) онтологической модели, включающая информацию о версиях и истории и параметры прав доступа, сохраняется в

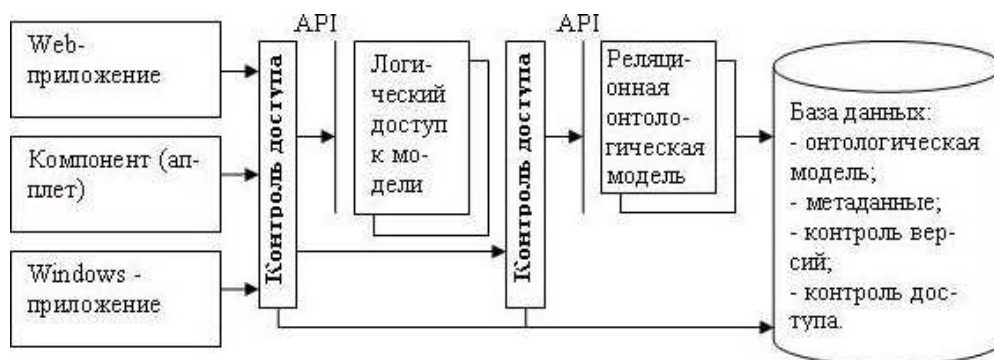


Рис. 4. Упрощенная схема работы системы с распределенной онтологической моделью

базе данных. Модуль контроля доступа читает данные по управлению доступом (ACL) из базы данных и действует в соответствии с ними. Логический уровень доступа к модели предоставляет более высокий уровень функциональности, такой как поиск и другие операции, состоящие из набора более простых операций доступа к модели. Уровень логического доступа к модели имеет свое собственное управление правами доступа.

И, наконец, пользовательский интерфейс, например, такой как web-приложения, компоненты (апплет) или локальные Windows-приложения, позволяет связываться с онтологической моделью, используя, например, Active Server Pages для web-интерфейса, или .Net Remoting (для интерфейса посредством компонентов) или интерфейс отдельного приложения.

Заключение

Использование предлагаемой распределенной информационно-телекоммуникационной системы позволяет создавать и коллективно поддерживать в актуальном состоянии единую онтологическую модель знаний организации. На основе использования данной модели могут быть описаны все ресурсы организации, содержащие знания, и реализованы базовые методы по работе со знаниями при выполнении основных бизнес-процессов организации [7].

Список литературы

- [1] Тузовский А.Ф., Чириков С.В., Ямпольский В.З. Системы управления знаниями. Методы и технологии. Томск: Изд-во НТЛ, 2005. 260 с.
- [2] TOWARDS the Semantic Web: Ontology-Driven Knowledge Management / J. Davies, D. Fensel, F. van Harmelen (Eds). John Wiley & Sons Ltd, 2003. 311 p.
- [3] Тузовский А.Ф., Козлов С.В., Ямпольский В.З. Использование онтологий в системах управления знаниями организаций // Изв. Том. политех. ун-та. 2006. Т. 309, № 3. С. 180–184.
- [4] HANDBOOK on Ontologies (International Handbooks on Information Systems) / S. Staab, R. Studer (Eds). N.Y.: Springer Verlag, 2004. 660 p.
- [5] Тузовский А.Ф., Козлов С.В. Построение модели знаний организации с использованием системы онтологий // Компьютерная лингвистика и интеллектуальные технологии: Тр. Междунар. конф. “Диалог-2006”, Бекасово, 31 мая — 4 июня 2006. М.: Изд-во РГГУ, 2006. С. 508–512.
- [6] ПАТ. 7093200 США, Instance browser for ontology / M. Z. Schreiber et. al.; заявлено 06.07.2001; опубл. 15.08.2006 // [Электронный ресурс]. 2006. <http://www.patentstorm.us/patents/7093200-fulltext.html>
- [7] Тузовский А.Ф. Разработка системы управления знаниями на основе единой онтологической базы знаний // Изв. Том. политех. ун-та. 2007. Т. 310, № 2. С. 182–185.

Поступила в редакцию 23 августа 2007 г.