

Развитие информационной системы организации с использованием семантических технологий

Черний А.В., Тузовский А.Ф.¹

¹Институт Кибернетический Центр, ул. Советская 84, г. Томск, 634050, Россия

cherny@tpu.ru, tomo@osu.cctpu.edu.ru

Аннотация. Статья посвящена применению методологии *Semantic Web* в организациях для решения задачи интеграции разнородных ресурсов информации и данных. Описан подход к разработке такой системы, предложенная архитектура, используемые технологии и программные продукты. Система основана на наборе онтологий и нацелена на извлечение, интеграцию, категоризацию, поиск разнородных объектов знаний организации на основе имеющихся в организации информационных систем и документов. Система позволяет интегрировать знания нескольких организаций-партнеров. Описанная методология и система представляют собой универсальную платформу для создания Систем Управления Знаниями и могут быть применены в любой предметной области.

Ключевые слова: онтология, система управления знаниями, аннотирование, семантический поиск, семантическая близость.

1 Введение

В настоящее время ведутся активные исследования и разработки в области реализации концепции *Semantic Web*. Данная концепция предложена консорциумом W3C в качестве новой модели развития Web сети – *Semantic Web*. Основным содержанием данного подхода является использование семантических моделей (онтологий), для описания метаданных ресурсов всемирной сети. Появились стандарты языков описания онтологий (например, RDF, OWL, SPARQL) и метаданных (например, RDF, OWL, SPARQL), а также множество инструментов для работы с онтологиями, таких как редакторы онтологий (например, Protégé) и системы логического вывода (например, Pellet, Racer, Joseki) [1]. Однако следует отметить трудность реализации концепции *Semantic Web* в рамках глобальной сети (Интернет) в ближайшее время, ввиду проблем с созданием и поддержкой описаний, а также в связи со сложностью интеграции ресурсов. Но идеи и существующие технологии *Semantic Web* могут быть использованы для развития информационных систем в рамках локальных сетей (интранет). В данной статье описан подход к развитию существующих информационных систем организации с помощью семантических технологий, использующихся в глобальной сети, таких как смысловой поиск, категоризация, навигация по ресурсам, интеграция и т.д.

2 Описание основных понятий

Все информация и данные организации содержатся во множестве разнообразных источников R (документы, файлы, базы данных, внешние ресурсы, на которые есть ссылки) $R = \{R_1, \dots, R_n\}$. Кроме этого, в организации имеется множество потребителей информации (специалисты, каталоги, бизнес-процессы и входящие в них задачи), которые заинтересованы в своевременном получении некоторого вида информации.

Для организации работы с разнотипной и распределенной информацией и данными предлагается использовать онтологическое моделирование [2]. Под онтологической моделью (онтологией) O в данной статье понимается знаковая система $\langle C, T, P, F, L, A \rangle$, где C – множество элементов, которые называются понятиями; T – частичный порядок на множестве C , задающий отношения «подкласс» и «суперкласс»; P – множество элементов, которые называются свойствами (двуместными предикатами); F – функция, которая назначает каждому элементу множества P множество элементов из множества C (с учетом их иерархии в T), к которым оно приме-

нимо (область действия, область, domain) 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, range); $L = \{L^C, L^P, \alpha^C, \alpha^P\}$ – множество текстовых меток, которые определяют профессиональные термины организации и их соответствие, соответствие α^C – элементам множества C , α^P – элементам множества P ; A – набор аксиом онтологии – утверждения об элементах предметной области, которые считаются верными, выраженных с использованием соответствующего логического языка.

Создание единой онтологии для детального описания модели знаний организации является весьма трудоемкой задачей. Для решения задачи построения онтологии организации предлагается использовать онтологическую модель следующей структуры [3]: $O = \{O_o, O_p, O_s\}$, где O_o – онтология организации, O_p – онтология информационных ресурсов, а $O_s = \{O_1, \dots, O_m\}$ – иерархически организованная, последовательно расширяемая система онтологий основных областей знаний O_i , значимых для работы организации. Выделение иерархии областей знаний организации дает возможность создавать отдельно онтологий разных подобластей знаний, которые могут иметь разную детальность, в зависимости от потребностей их моделирования. Онтология организации O_o включает основные понятия, которые описывают структуру, состав элементов и работу организации (подразделения, специалисты, заказчики, проекты, бизнес-процессы и пр.). Онтология информационных ресурсов O_p включает описание всех видов ресурсов данных и информации организации (документы, файлы, базы данных, программы и пр.).

На основе такой онтологической модели знаний O , описание ресурса (объекта) R_i может быть представлено в виде набора семантических метаданных [3] следующей структуры $M_i = (M_{ki}(O), M_{ci}(O))$, где $M_{ki}(O)$ – это *контекстные метаданные* объекта знаний, описывающие взаимосвязи объекта с другими объектами и понятиями организации или литералами, а $M_{ci}(O)$ – *контентные метаданные* ресурса, описывающие информацию, которая содержится в объекте. *Контекстные метаданные* соответствуют набору значений свойств понятия $c_j \in C$, экземпляром которого является описываемый объект в онтологии организации, т.е. $M_{ki}(O) = (p_1(O_i, v_1), \vee p_2(O_i, v_2) \vee \dots \vee p_r(O_i, v_r))$, где утверждение $p_i(O_i, v_i)$ состоит из предиката (отношения) $p_i \in P$, описанного в онтологии организации, URI (универсальный идентификатор ресурса) описываемого объекта O_i , для которого определяются метаданные и значения v_i , которое может быть либо URI некоторого экземпляра понятий онтологии организации, либо некоторый литерал (текст, число, дата) в соответствии с областью действия и областями возможных значений свойства p_i . *Контентные метаданные* $M_{ci}(O)$ описываются наборами утверждений из O , т.е. семантические метаданные объектов знаний описываются отношениями и понятиями из онтологий основных предметных областей знаний организации в виде набора кортежей: $M_{ci}(O) = (\{p_1(s_1, v_1), k_1\} \vee \{p_2(s_2, v_2), k_2\} \vee \dots \vee \{p_k(s_k, v_k), k_k\})$, где $\{p_i(s_i, v_i), k_i\}$ кортеж, включающий утверждение $p_i(s_i, v_i)$ (соответствующее RDF триплету (s_i, p_i, v_i)) и k_i – важность данного утверждения для описания контента объекта знаний i . Утверждение $p_i(s_i, v_i)$ состоит из предиката (отношения) $p_i \in P$, описанного в онтологии областей знаний, объекта s_i , которым может быть понятие онтологии $c_j \in C$ или экземпляр онтологии i (ссылка на контекстные метаданные некоторого экземпляра онтологии) и значения v_i , которое может быть либо URI некоторого экземпляра, либо некоторым литералом (текстом, числом, датой).

При этом следует отметить, что не все экземпляры понятий онтологии будут соответствовать ресурсам информации или данных организации, например, в онтологию также такие понятия, как Специалист, Заказчик, Проект и т.п.

Совокупность описаний онтологической модели и метаданных всех ресурсов организации составляют *онтологическую базу знаний* системы.

2.1 Измерение близости метаописаний ресурсов.

Учитывая, что модели описаний объектов знаний связаны за счет использования единой модели знаний организации, то имеется возможность оценки их подобия (сходства) между собой на основе некоторой метрики подобия $\Phi(M_i, M_j)$. В статье предложены методы оценки подобия контекстных и контентных метаданных объектов знаний. Данный функционал описывает близость между отдельными элементами знаний, содержащимися в разных объектах знаний.

Формализованное представление онтологической модели знаний, а также метаописаний ресурсов создает возможность для измерения близости (подобия) ресурсов в интеллектуальном пространстве организации. Подобие между метаданными $\Phi(Mz_b, Mz_j)$ может быть определено через подобие входящих в них утверждений:

$$\Phi(Mz_i, Mz_j) = \sum_{I_i \in MD_i} \sum_{I_j \in MD_j} sim(T_i, T_j), \quad (1)$$

где $\Phi(Mz_i, Mz_j)$ – величина близости метаописаний объекта i и объекта j ; $sim(T_i, T_j)$ – величина близости утверждений (триплетов) T_i и T_j , входящих в сравниваемые метаописания. Величины $sim(T_i, T_j)$ могут быть определены с использованием следующего выражения:

$$sim(T_i, T_j) = sim((c_i, r_j, i_k k_i), (c_x, r_y, i_z k_w)) = \left(\frac{simc(c_i, c_x) + simr(r_j, r_y) + simi(i_k, i_z)}{3} \right) \cdot f(k_l, k_w), \quad (2)$$

где $simc(c_i, c_k)$ – семантическая близость понятий, используемых в утверждениях; $simc(r_i, r_j)$ – семантическая близость отношений онтологий; $simi(i_k, i_z)$ – семантическая близость контекстных метаданных экземпляров понятий онтологий; $f(k_l, k_w)$ – функция учета коэффициентов важности утверждений (используются разные варианты).

2.1.1 Оценка семантической близости двух понятий

В онтологии O на множестве понятий C задано отношение нестрогого – частичного порядка T_C . Утверждение $T_C(c_k, c_l)$ означает, что c_k предшествует c_l или что c_l следует за c_k . Причем T_C задано так, что среди элементов множества C существует *единственный минимальный* (базовый) элемент $c_{top} \in C$. Иерархия понятий с единственной вершиной (таксономия понятий), заданная отношением T_C , используется для определения семантической близости понятий.

Для каждого понятия $c_i \in C$ существует множество $C_{ANC}(c_i)$, являющееся подмножеством C и содержащее понятия, предшествующие понятию c_i , а также само понятие c_i :

$$C_{ANC}(c_i) = \{ c_j \in C \mid T_C(c_j, c_i) \vee c_j = c_i \}. \quad (3)$$

Для оценки семантической близости двух понятий $sim_C(c_k, c_l)$ вводятся два показателя, основанные на сравнении множеств $C_{ANC}(c_i)$:

$$sim_C(c_k, c_l) = k_{st} * \frac{|C_{ANC}(c_k) \cap C_{ANC}(c_l)|}{|C_{ANC}(c_k) \cup C_{ANC}(c_l)|}, \quad (4)$$

$c_{top} \in C_{ANC}(c_k) \cap C_{ANC}(c_l)$ и $c_{top} \in C_{ANC}(c_k) \cup C_{ANC}(c_l)$ при любых $c_k \in C$ и $c_l \in C$.

$$k_{st} = \begin{cases} 1, & \text{если } C_{ANC}(c_k) \cup C_{ANC}(c_l) = C_{ANC}(c_l), \\ 0, & \text{иначе;} \end{cases} \quad (5)$$

$$sim_C(c_k, c_l) \in [0; 1]. \quad (6)$$

2.1.2 Оценка семантической близости двух экземпляров

В онтологии O для каждого экземпляра $i_x \in I$ существует:

- непустое множество $C_{INST}(i_x)$, включающее понятия, к которым относится экземпляр i_x :

$$C_{INST}(i_x) = \{ c_j \in C \mid Is(i_x, c_j) \}, C_{INST}(i_x) \neq \emptyset; \quad (7)$$

- множество $R_{INST}(i_x)$, включающие все конкретизированные отношения экземпляра i_x :

$$R_{INST}(i_x) = \{ ri_j \in R_I \mid ri_j(i_x, i_y) \} \quad (8)$$

Оценка семантической близости $sim_I(i_x, i_y)$ двух экземпляров i_x и i_y складывается из их реляционной близости и близости их типов. Реляционная близость $sim_{IL}(i_x, i_y)$ позволяет оценить схожесть двух экземпляров исходя из их отношений с другими экземплярами онтологии:

$$sim_{IL}(i_x, i_y) = \begin{cases} \frac{|R_{EQU}(i_x, i_y)|}{|R_{INST}(i_x) \cap R_{INST}(i_y)|}, & \text{если } R_{INST}(i_x) \cap R_{INST}(i_y) \neq \emptyset \\ 0, & \text{иначе} \end{cases}, \quad (9)$$

$$R_{EQU}(i_x, i_y) = \{ri_i \in R_I \mid ri_i(i_x, i_z) \wedge ri_i(i_y, i_z)\}. \quad (10)$$

Близость типов $sim_{IC}(i_x, i_y)$ основана на семантической близости понятий, к которым относятся экземпляры, в онтологической модели.

$$sim_{IC}(i_x, i_y) = \frac{\sum_{c_i \in C_{INST}(i_x)} \max_{c_j \in C_{INST}(i_y)}^{m} (sim_C(c_i, c_j))}{n}; \quad (11)$$

$$sim_{IC}(i_x, i_y) \in [0; 1]. \quad (12)$$

И, наконец, семантическая близость двух экземпляров определяется следующим образом:

$$sim_I(i_x, i_y) = \frac{k_{IC} * sim_{IC}(i_x, i_y) + k_{IL} * sim_{IL}(i_x, i_y)}{k_{IC} + k_{IL}}. \quad (13)$$

С учетом того, что $k_{IC} \in (0; 1]$ и $k_{IL} = 1 - k_{IC}$, получаем $sim_I(i_x, i_y) \in [0; 1]$.

Коэффициенты k_{IC} и k_{IL} позволяют настраивать процедуру вычисления семантической близости двух экземпляров в зависимости от содержания *ABox* представления онтологии *O* в виде модели дескриптивной логики. Если экземпляры описаны в онтологии в основном с помощью связей с другими экземплярами или конкретными значениями, то необходимо установить соотношение $k_{IC} < k_{IL}$. В противном случае необходимо установить соотношение $k_{IC} \geq k_{IL}$.

На основе использования данного функционала могут быть решены разные задачи интеграции разнородной информации и данных: смысловой поиск документов (любой информационный ресурс, описанный с помощью метаданных), классификация объектов знаний по системе рубрик, рекомендация (предоставление) новых объектов знаний специалистам.

2.2 Семантический поиск

Семантический (смысловой) поиск состоит в поиске объектов базы знаний, у которых семантические метаданные близки (превышают некоторое пороговое значение) семантическому описанию поискового запроса. Объектом-эталоном при семантическом поиске является поисковый запрос, представленный в виде контентных метаданных. Процедура формирования множества объектов-кандидатов для выполнения среди них семантического поиска заключается в выборе пользователем тех понятий из онтологии информационных ресурсов, которым соответствуют требуемые типы объектов. В результате семантического поиска пользователю предлагается список найденных по запросу объектов, упорядоченных по величине семантической близости и дополнительно сгруппированных по типам объектов.

2.3 Категоризация

Под задачей категоризации понимается распределение ресурсов информации между иерархически организованным набором рубрик (категорий). Каждая рубрика описана семантически метаданными, которые описывают информацию, которая должна в них содержаться. Родительские и дочерние рубрики в иерархии классификации связаны между собой отношением включения (*subsumtion*) между их семантическими метаданными. При решении задачи классификации в качестве объекта-эталоны выступают метаданные рубрики классификатора. Из базы знаний извлекаются семантические метаданные всех объектов (за исключением рубрик), для которых нужно проверить соответствие рубрике. В процессе обработки результатов из множества кандидатов удаляются те объекты, контентные метаданные которых не содержат хотя бы одного понятия или экземпляра, присутствующего в семантических метаданных рубрики. Во множестве объектов-кандидатов остаются лишь те объекты, которые имеют показатель релевантности больше нуля, то есть относятся к рубрике.

3 Архитектура системы

Информационная система организации состоит из множества компьютеров объединенных в локальную сеть, которые содержат разнообразные информационные ресурсы разных форматов.

Рассматриваемая в статье система позволяет интегрировать и структурировать все общие ресурсы информации и данных. Она представляет набор серверов и клиентских приложений, которые должны устанавливаться в локальной сети организации. Структура системы показана на рис. 1.

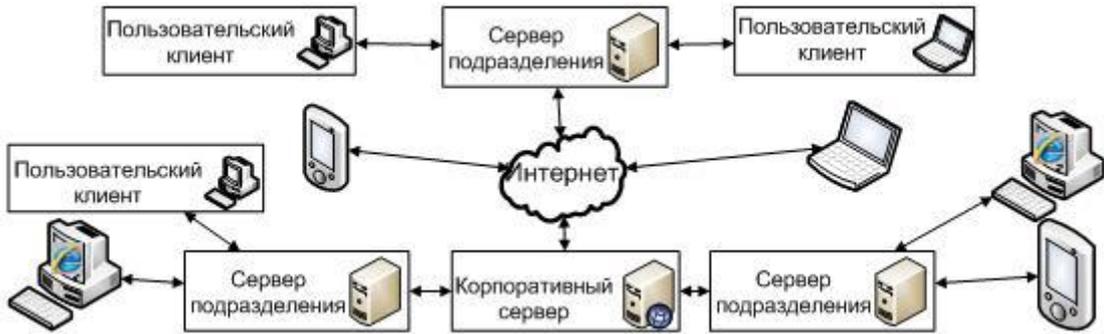


Рис. 1. Структура системы

Сервера отвечают за совместное использование онтологической базы знаний и решения базовых задач по работе с онтологической моделью (редактирование и пополнение онтологии), метаданными (формирование семантических метаданных – аннотирование, хранение базы знаний) и информационными ресурсами (поиск, категоризацию, навигация по метаданным).

Различие центрального сервера и серверов подразделений состоит в следующем:

- Центральный сервер хранит всю онтологическую модель организации, в то время, как сервер подразделения хранит только онтологии тех предметных областей, с которыми работает это подразделение. Это сделано для возможности их автономного редактирования и увеличения скорости аннотирования информационных ресурсов данного подразделения. Набор функциональных возможностей может быть настроен для каждого сервера в отдельности. Сервера могут находиться в одной локальной сети, или быть распределены по разным интранет сетям организации.
- Каждый сервер подразделения имеет собственный набор категорий в подсистеме категоризации, свой набор документов для поиска и т.п.

Также существует пользовательский клиент системы интеграции («толстый клиент») – набор приложений, интегрирующих возможности системы с рабочим интерфейсом конечных пользователей, делая использование системы удобным. Оно позволяет пользователю обрабатывать в контексте онтологии личные данные.

Если пользователю этот функционал не нужен, он может работать с системой с помощью тонкого клиента - браузера.

Архитектура системы приведена на рис 2.

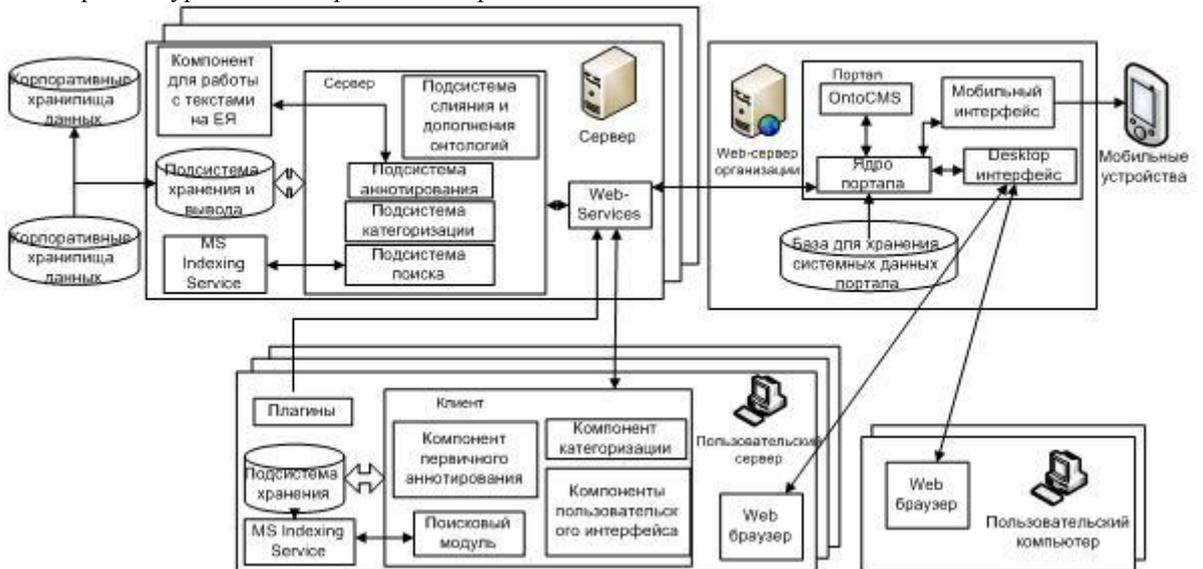


Рис.2. Архитектура системы

4 Описание компонентов системы

В состав сервера входят такие подсистемы, как подсистема хранения и логического вывода, подсистема аннотирования, подсистема поиска, подсистема пополнения онтологии, подсистема категоризации.

4.1 Подсистема хранения и логического вывода

Данная подсистема состоит из трех частей: модуля логического вывода, модуля хранения и модуля интеграции.

Модуль логического вывода отвечает за построение иерархий и создание новых триплетов на основе базовой онтологии за счет механизмов дескриптивной логики. Таким образом, в системе появляются связи, явно не заданные, но актуальные.

Модуль хранения обеспечивает хранение описания онтологической модели в реляционных базах данных. Данная подсистема устанавливается, как на серверах, так и на пользовательских клиентах потому, что на серверах хранится вся корпоративная информация и данные, а на клиентских компьютерах могут находиться только метаданные (описания документов) ресурсов пользователя, которые не предоставляются в общий доступ.

4.2 Модуль интеграции

предоставляет функционал по интеграции онтологии и баз данных. Это дает нам возможность, не изменяя существующие в организации информационные системы, использовать данные этих ИС и иметь в онтологии всегда актуальные данные без приложения дополнительных усилий. Каждая таблица в БД ставится в соответствие с понятием онтологии. Каждый столбец этой таблицы является свойством понятия, а каждая строка – экземпляром (рис. 3).



Рис. 3. Формирование онтологической модели на основе БД

При интеграции в свойствах каждого понятия указывается, с каким полем таблицы связано каждое свойство этого понятия. В БД добавляются триггеры, которые позволяют отслеживать обновления экземпляров в БД и актуализировать эти значения в онтологии.

Выработанная организацией базовая онтология является доступной для партнеров организации (только онтология, без метаданных). Таким образом, становится возможным слияние онтологий из разных источников, что дает партнерам использовать все преимущества единого информационного пространства нескольких взаимодействующих организаций. И, как следствие, более полно описанные документы, семантически описанные совместные бизнес-процессы и т.п.

Слияние онтологий выполняется совместно менеджерами по работе со знаниями организаций и основано на свойствах языка OWL, который позволяет упростить этот процесс, находя, например, идентичные понятия и ставя их в однозначное соответствие и так далее.

4.3 Подсистему аннотирования

Подсистему аннотирования можно разделить на серверную часть, которая производит разбор документа и составление триплетов, отражающих его смысл, и клиентскую, которая производит первичный разбор и анализ необходимости аннотирования документа.

Серверная часть подсистемы аннотирования выделяет из документов понятия, экземпляры понятий и связи между ними. Она основывается на использовании программного обеспечения, выполняющего грамматический разбор текстов на естественном языке. На основе этого строится описание (граф) документа, который сохраняется в виде триплетов на сервере и (если позволяет формат документа) в самом документе. С каждым триплетом, полученным из документа, связывается значение, определяющее соответствие триплета смыслу документа – значимость

этого утверждения. По полученным аннотациям впоследствии производится семантический поиск.

Аннотации, полученные таким путем, считаются первичными, и пользователю предоставляется возможность после ознакомления с документом в дружелюбном пользовательском интерфейсе откорректировать и дополнить автоматическую аннотацию.

Клиентская часть подсистемы аннотирования. Ввиду того, что онтология организации постоянно пополняется новыми понятиями, связями и экземплярами, необходим механизм, который будет определять, требуется ли какому либо документу повторное аннотирование.

За это отвечает клиентский модуль аннотирования. Он просматривает документы на предмет наличия новых понятий, или, говоря точнее, их лексических меток во всех словоформах. Если они в документе найдены, то предложения, их содержащие, и их окружение отправляется на повторное аннотирование, а результаты аннотирования сохраняются на сервере и в документе (если позволяет формат).

4.4 Подсистема поиска

Как известно, семантический поиск дает намного более релевантные результаты, чем полнотекстовый. Но при этом считается [3], что лучшим вариантом является их совмещение. Причиной этому являются следующие факторы:

- семантический поиск может не описывать весь смысл документов из-за её неполноты онтологии, на основе которой они описаны;
- возможны огрехи в существующей онтологии;
- пользователь может желать найти конкретную цитату, а при использовании только семантического поиска, где запросом является набор триплетов, это не представляется возможным.

Ввиду этого приемлемой видится система, в которой пользователь может самостоятельно задать – какой из следующих видов поиска его интересует: только семантический, только полнотекстовый или совмещенный, когда пользователь задает набор триплетов и уточняющий полнотекстовый запрос, либо помечает при наборе в строку полнотекстового запроса, что его необходимо интерпретировать и как семантический. Во втором случае система разбирает его по тем же правилам, что и тексты при аннотировании и на этой основе строит триплеты для запроса. Таким образом, пользователь получает первыми результаты, которые удовлетворяют семантическому запросу и содержат искомый текст, затем прочие в порядке убывания релевантности.

4.5 Подсистема дополнения и объединения онтологий

Каждый пользователь системы имеет возможность дополнить и расширить онтологию. Но изначально все изменения, которые делает пользователь, сохраняются в его личной копии онтологии. Затем клиентская часть передает эту онтологию на сервер, где и происходит слияние. При возникновении нецелостности онтологии либо других конфликтов, ответственным за разрешение их является менеджер по работе со знаниями.

4.6 Подсистема категоризации

Подсистема категоризации служит для распределения по категориям ресурсов компании (документов, сотрудников, ссылок, программ и т.п.) на основе их семантического описания.

4.7 Пользовательский клиент

Данный модуль включает возможности категоризации, поиска, первичного аннотирования (как было описано ранее), а также подсистемы взаимодействия с пользователем.

Подсистема взаимодействия с пользователем предназначена для предоставления пользователю удобного графического интерфейса по работе с системой. Основными элементами данной подсистемы являются дополнительные модули интеграции различных приложений с описываемой системой и пользовательские интерфейсы для доступа к средствам категоризации и поиска.

Модули интеграции приложений с системой обеспечивают средство интеграции различных приложений по работе с документами с серверами системы для аннотирования и поддержки связи информации документов с онтологической базой знаний. Например, в приложении Microsoft Word, модуль интеграции позволяет выделять понятия предметной области и выполнять с

ними некоторые манипуляции. Например, читая некоторый документ, в котором есть ссылка на другой документ, имеется возможность, выполнив щелчок кнопкой «мыши» на упоминание о нем, перейти к карточке документа и узнать, кем и когда он был создан, кем изменялся и пр., либо сразу открыть документ. Такого рода модули могут быть разработаны для браузеров, офисных приложений и т.п.

Пользовательские интерфейсы к средствам категоризации и поиска. Клиентское приложение обеспечивает удобные функции для категоризации и поиска. Например, вызов поискового интерфейса нажатием сочетания клавиш, представление модуля категоризации в виде простых Windows папок и т.п.

4.8 Подсистема Web-сервера

Web-сервер может присутствовать как в каждом подразделении, так и быть вынесенным на корпоративный уровень и главной составляющей в нем является семантический портал.

Семантический портал отвечает за создание и доступ к функционалу системы через web-интерфейс и состоит из базовой части и системы управления контентом на основе семантических метаданных.

Базовая часть семантического портала предоставляет набор классов и интерфейсов (API) для реализации других модулей и компонент портала. Так как система является лишь платформой для создания систем интеграции информации и данных и обеспечивает лишь базовый функционал, который есть в любой организации, то его, конечно, необходимо разрабатывать дополнительные специализированные модули, предоставляющие дополнительную (специфическую для конкретной организации) функциональность.

Система управления контентом на основе семантических метаданных (ресурсов информации и данных) представляет собой Content Management System (CMS) с расширенным функционалом для работы с онтологиями и семантическими метаданными. Например, любой элемент управления на web-странице может формировать SPARQL запрос к онтологической базе знаний и представлять полученные результаты пользователю в удобном виде. На основе онтологической CMS реализуется подсистема поддержки задач бизнес-процессов требуемыми ресурсами информации и данных. Для поддержки бизнес-процессов необходимо выполнить их семантического описания бизнес-процессов состоит в том, что бизнес-процесс в целом, и каждая его задача в частности, описывается в понятиях онтологии. Это дает множество преимуществ. При выполнении некоторой задачи бизнес-процесса сотрудник может узнать всю информацию, полезную для ее выполнения, получить описание сотрудников организации, которые выполняли данную задачу ранее (и, при необходимости, связаться с ними) и многое другое.

4.9 Реализация системы

В реализации предлагаемого подхода используются разные сторонние программные продукты. В качестве базовой подсистемы используется OWLAPI. Для аннотирования могут использоваться компоненты RML или более дорогие и более функциональные компоненты RCO. В качестве движка полнотекстового поиска мы используем входящий в состав ОС Windows Microsoft Indexing Service.

5 Заключение

В целом можно сказать, что разработанная система позволяет не только интегрировать всю информацию, которая имеется в организации, начиная от знаний о служащих, заканчивая информацией о производимых деталях или услугах, но и создавать коллективное информационное пространство, которое, впоследствии, приблизит эпоху Semantic Web.

Литература

[1] Allemang D., Hendler J., Semantic Web for the Working Ontologist Modeling in RDF, RDFS and OWL – Morgan Kaufmann Publishers, 2008. – 350 p.

[2] Тузовский А.Ф. Интеграция Информации и Данных Организации с Использованием Онтологических Моделей и Семантических Методов.// Материалы Всероссийской конференции с международным участием «Знания-Онтологии-Теория» - 2007.- Т.1 – С. 45-53

[3] <http://www.intuit.ru/department/expert/ontoth/3/>