

## ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКОЙ МОДЕЛИ ПРИ СЕМАНТИЧЕСКОМ ПОИСКЕ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ

Ю.В. Рогушина

*Институт программных систем НАН Украины, Киев, Украина  
ladamandraka2010@gmail.com*

### Аннотация

В статье рассматривается моделирование системы интеллектуального взаимодействия между информационными ресурсами и потребителями информации с использованием внешних и внутренних баз знаний. Анализируются существующие подходы к оценке уровня интеллектуальности информационных систем и целесообразность их применения к информационно-поисковым системам. На основе этого анализа предлагается критерий сравнения уровня интеллектуальности различных приложений, который базируется на таких параметрах, как тип и количество обрабатываемых в системе атомарных элементов и связей между ними, а также сложность алгоритма обработки. Разработана онтологическая модель, которая описывает взаимодействие пользователей и информационных ресурсов Web при семантическом поиске, формально описаны её элементы и связи между ними. Предложены источники и методы пополнения этой модели. Рассматривается, каким образом использование этой модели позволяет интеллектуализировать систему семантического поиска. Описана программная реализация системы персонализированного и коллаборативного семантического поиска, которая базируется на этой онтологической модели. Предложены пути использования этой модели для интеграции подсистемы семантического поиска в прикладные информационные системы (на примере задачи сопоставления компетенций, которая является составной частью таких проблем, как поиск работодателем подходящих исполнителей работ, сравнение квалификации специалистов, оценка возможности перехода студента из одного учебного заведения в другое и т.д.).

**Ключевые слова:** *семантический поиск, информационный объект, онтологическая модель, тезаурус, коллаборативный поиск.*

### Введение. Системы семантического поиска

В наиболее обобщенном понимании *информационный поиск* – это проблема, состоящая из двух составляющих: 1) сопоставления представления пользователя о нужных ему знаниях с контентом доступных *информационных ресурсов* (ИР); 2) построения на основе этого сопоставления необходимого пользователю *информационного объекта* (ИО) с конечным набором свойств, значения которых извлекаются из этих ИР.

Пользователь имеет часть информации об ИР и пытается дополнить её сведениями, извлеченными из различных источников.

В общем случае ИР – это информация, сконцентрированная и собранная в определенном формализованном виде, которая представляет определенную ценность и может быть оценена подобно материальным ресурсам.

В данной работе под ИР предполагаются различные документы, представленные в электронной форме и доступные *информационно-поисковой системе* (ИПС), а под ИО – явно или неявно содержащиеся в ИР сведения о различных физических и виртуальных объектах, которые являются результатом работы ИПС.

Частными случаями проблемы информационного поиска можно считать такие задачи, как распознавание образов, понимание речи или изображений, семантический анализ и перевод естественно-языковых текстов, исследование и композицию сервисов.

Все эти задачи в общем представлении являются очень сложными, и поэтому на практике целесообразно рассматривать относительно суженные подзадачи, в которых заранее накладывается ряд ограничений на потребности пользователя и на те информационные объекты, которые должны быть сгенерированы в результате поиска.

На основе анализа существующих ИПС можно выделить следующие основные тенденции совершенствования поиска [1]:

- от формального – к семантическому;
- от унифицированного – к персонифицированному;
- от индивидуального – к коллаборативному;
- от закрытого – к открытому (с привлечением внешних баз данных и знаний).

Сегодня термин «семантический поиск» широко используется, однако часто под ним понимают любые поисковые системы, каким-либо образом использующие знания [2, 3]. Кроме того, очевидно, что результатом поиска должна быть не абстрактная информация, а информация, несущая в себе прагматическую составляющую, т.е. полезная при решении задачи, стоящей перед пользователем.

Кроме того, для современных ИПС характерно осуществление не только синтаксической, но и семантической и прагматической фильтрации контента [4]. При семантической фильтрации рассматривается содержание информации, а при прагматической учитывается оценка информации с учетом ее субъективной ценности для конкретного индивида.

В данной работе под термином *семантический поиск* будет рассматриваться информационный поиск, в котором сопоставление и построение информационных объектов выполняются на семантическом уровне, т.е. с использованием знаний.

Основное отличие семантического поиска от традиционного – это явное использование *знаний* об объекте поиска, пользователях, ИР и *предметной области* (ПрО) поиска, а также возможность обнаружения не данных, а знаний.

Семантический поиск базируется на достижениях в области искусственного интеллекта (ИИ), в частности на общей теории представления и обработки знаний, распознавания образов и логического вывода, методах математической статистики и социопсихологии.

Предмет исследования данной работы – процесс семантического поиска в распределенной среде Web, его объекты, субъекты и отношения между ними, а также методы сопоставления информационных ресурсов с информационными потребностями пользователей.

Сегодня обнаружение знаний в Web является составной частью многих интеллектуальных приложений, которые ориентированы на работу в открытом информационном пространстве и нуждаются в постоянном обновлении информации. Это обусловлено тем, что большинство современных приложений ориентированы на использование и преобразование знаний об интересующей пользователя ПрО, которые не закладываются в *информационную систему* (ИС) при её разработке, а извлекаются динамически из доступных ИР.

В данной работе под интеллектуальными ИС будем понимать различные ИС, способные решать задачи, которые традиционно считаются творческими и требуют от человека знаний из соответствующей ПрО. Именно с изучением таких систем связана группа наук, которую объединяет название ИИ, которая изучает методы, способы и приемы моделирования и воспроизведения с помощью компьютерных средств умственной деятельности человека.

Как правило, доступ к различным ИР обеспечивается при помощи Web, где для их представления могут использоваться самые разные модели формализации, способы и форматы хранения, условия доступа и методы обработки. Это означает, что ИС могут обрабатывать

сведения, которые им предоставляют различные информационные ресурсы Web – текстовые, мультимедийные, структурированные.

Информацию, представленную в структурированном виде (онтологии, метаописания, семантически размеченные файлы и т.п.), обрабатывать намного удобнее, но, к сожалению, относительно мало документов представлено в таком виде. Поэтому поиск среди таких ИР часто не позволяет находить нужные сведения.

Значительная часть знаний, накопленных в результате развития человеческого общества и описывающих различные предметные области, содержится в документах в виде текста на естественных языках. В Web присутствует также большое количество мультимедийных документов, и с распространением различных подключенных к Интернет устройств их объем растет опережающими темпами. Вместе с тем в таких документах содержится значительно меньше полезной информации (например, подавляющее большинство различных фотографий и видеороликов, как правило, интересны лишь тем, кто их снимал).

Следует отметить, что предметом поиска в большинстве случаев является не какой-то конкретный документ, а некие сведения о каком-либо объекте – реальном либо виртуальном. При этом у пользователя присутствует часть информации об этом объекте, о его свойствах и структуре, а недостающие сведения он рассчитывает получить в результате поиска.

ИО – это информационная модель какого-либо реального или виртуального объекта ПрО (предмета, существа, события, процесса и т.д.) в информационном пространстве, которая определяет структуру, атрибуты, ограничения целостности и, возможно, поведение этого объекта. Например, объектом может быть человек, публикация, Web-сайт, организация, город, а ИО – их описание.

При семантическом поиске пользователь может указать, экземпляром какого класса является тот ИО (или группа ИО), информация о котором ему нужна, например, при помощи ссылки на класс какой-либо онтологии ПрО.

В концепции Semantic Web [5, 6] и Web-пространстве информационными объектами являются и такие объекты нематериального мира, как онтологии, программные агенты, Web-сервисы, информационные ресурсы, метаданные, базы данных и т.д.

Как основу для представления структуры ИО можно использовать классы соответствующей онтологии, а из ИР извлекать сведения для создания экземпляров ИО.

При семантическом поиске ИО может иметь заранее заданную сложную структуру, формализованную в виде класса соответствующей онтологии. Примеры ИО – организация, учебное заведение, человек-эксперт, Web-сервис, аналитическая модель бизнеса-процесса. Наличие таксономии ИО даёт возможность пользователю ИПС формально указать, какую именно информацию ему нужно найти и какими сведениями о ней он обладает.

Например, при поиске человека можно указать, что «Рогущина» – это фамилия, «Киев» – место жительства, а «Келли Лада де Мандрака» – имя принадлежавшей когда-то этому человеку собаки. При этом структура ИО обеспечивает семантическую разметку самого запроса.

При этом возникает проблема поиска онтологии, которая отображает структуру ИО, знания о котором необходимы пользователю.

Если же сравнивать усилия, необходимые для извлечения знаний (семантики), то естественно-языковые ресурсы обрабатывать значительно легче, чем мультимедийные, а сама форма представления обеспечивает первичный поиск релевантных проблеме ресурсов (например, поиск по ключевым словам в тексте).

Кроме того, при распознавании мультимедийных ИР вначале значительная часть информации преобразуется в естественно-языковой текст. Это обуславливает важность развития автоматизированных методов анализа естественно-языковых документов, которые позволили бы корректно соотносить документы с определённой ПрО, распознавать их смысл и из-

влекать из них полезные для пользователя сведения. Такие методы должны обеспечивать интеллектуальный анализ структуры и контента ИР.

*Системы семантического поиска (ССП)* – это ИПС, в которых поиск информации осуществляется с учетом семантики. В ряде случаев СПП представляют собой не самостоятельные ИПС, а надстройку над существующими ИПС различных типов, которая обеспечивает интеллектуализацию этих систем.

## 1 Критерии сравнения уровня интеллектуальности поисковых систем

Одной из наиболее важных составляющих информационной революции является развитие *интеллектуальных ИС*, которые обладают способностями рассуждать, учиться на своем или чужом опыте и принимать разумные решения без человеческого вмешательства. Как правило, в этих системах применяются логический вывод и нечёткая логика, характерная для человеческих рассуждений.

В.М. Глушков определил такие важные свойства систем ИИ, как адаптация, самоорганизация и самосовершенствование. Интеллектуальность ИС оценивается – по аналогии с поведением человека – через меру сложности обрабатываемых объектов ПрО, пространственно-временных и причинно-следственных отношений между ними и алгоритмов, которые используются для этого. Интеллектуальность ИС непосредственно связана со степенью структурированности информации и методов её переработки [8], со способностью системы к выбору целей, формированию планов и целенаправленному их осуществлению.

Актуальность ИС, управляемых знаниями, обусловлена тем, что сейчас всё большую важность приобретают задачи, для решения которых априори трудно локализовать информацию, достаточную для достижения нужного результата. В ходе решения таких задач может понадобиться заранее непредвиденная информация, которая удовлетворяет тем ли другим требованиям: выходными данными для таких задач может считаться вся информация, к которой такая ИС способная получить доступ (например, через среду Web).

Примером задач, которые решают управляемые знаниями ИС, являются информационно-поисковые задачи, в которых по описанию некоторых свойств и элементов структуры ИО нужно извлечь из ИР другие характеристики этого объекта, необходимые пользователю.

Не следует отождествлять понятия системы, управляемой знаниями, и понятие системы, которая базируется на знаниях. Понятие системы, которая базируется на знаниях, является *более обобщенным*, так как база знаний здесь не обязательно должна быть активной и семантически структурированной. Это означает, что управление в системе, которая базируется на знаниях, может осуществляться программой, а не ситуациями и событиями в обрабатываемой базе знаний. Можно рассматривать системы, управляемые знаниями (и системы семантического поиска в том числе), как частный случай интеллектуальных систем.

Повышение эффективности современных ИС непосредственно связано с их уровнем интеллектуализации. В частности, в литературе часто используются термины «интеллектуальная ИПС», «интеллектуализация поиска» и т.п. Однако при этом термины «интеллект», «интеллектуальность» рассматриваются в интуитивном понимании и обычно не определены формально. Это приводит к проблеме оценки уровня интеллектуальности разных ИС и затрудняет их сравнение и анализ.

Основные типы действий, которые выполняются при решении интеллектуальных задач: преобразование связей ИО; создание новых ИО; поиск ИО, который удовлетворяет заданным требованиям; вычисление значений параметров ИО; поиск необходимой информации; логический вывод; поиск заданных элементов в модели ИО; формирование понятия.

Считается, что интеллектуальные ИС могут функционировать автономно, а *интеллектуализированные* – с участием пользователя. Но провести чёткую грань между этими типами ИС сложно. Поэтому более корректно говорить не об интеллектуальных и интеллектуализированных ИС, а об уровне интеллектуальности систем: любая ИС имеет уровень интеллектуальности, отличный от нуля (в неё закладываются определённые знания разработчика относительно ПрО и методов обработки информации). Верхней же границы уровня интеллектуальности ИС не существует.

На уровень интеллектуальности системы влияют способность к самообучению и широта её ПрО [9]. Чем выше уровень интеллектуальности ИС, тем эффективнее она способна достигать поставленную цель (при этом более высокая эффективность может и не быть достигнутой из-за неудачной реализации или неудачной модели).

Оценка уровня интеллектуальности ИС является нетривиальной задачей и в значительной мере определяется областью использования этой ИС.

Например, Л. Заде вводит термин «коэффициент машинного интеллекта» MIQ (*Machine Intelligence Quotient*), по аналогии с IQ человека по Айзенку для описания ИС, в которых главными компонентами являются нечёткая логика, нейровычисления, генетические вычисления и вероятностные вычисления. Интеллектуальной при этом считается ИС с высоким MIQ [10]. Определение такого коэффициента – это прямое развитие психометрического подхода на случай искусственных систем [11].

Основное отличие между традиционным коэффициентом интеллектуальности IQ и MIQ состоит в том, что IQ является более или менее стабильным, а MIQ ИС может изменяться со временем и машинно-зависим.

В.К. Финн [12] рассматривает свойства интеллектуальных ИС, среди которых можно выделить такие важные для семантического поиска способности, как:

- упорядочивание знаний;
- целеполагание и планирование поведения;
- извлечение следствий из имеющихся знаний, т.е. способность к рассуждению;
- аргументация принятых решений;
- рефлексия, оценка знаний и действий;
- синтез процедур решения задач;
- обучение;
- рационализация идей;
- создание целостной картины ПрО;
- адаптация в условиях изменения ситуаций и знаний.

При этом интеллектуальная ИС рассматривается как совокупность собственно решателя задач, информационной среды (включающей базу данных и базу знаний) и интеллектуального интерфейса.

*Интеллектуализация* ИС – это процесс повышения уровня интеллекта ИС. Предлагаем следующее определение [13]. Пусть есть две ИС –  $A$  и  $B$ , способные находить решения  $R_A(d)$  и  $R_B(d)$  соответственно в ситуациях  $d$  из определенной ПрО  $D$ . Система  $B$  получена вследствие интеллектуализации  $A$ , или  $B = \text{Int}(A)$ , если:

- система  $B$  всегда получает решение в ситуациях, в которых его находит и система  $A$ , т.е.  $\exists R_A(d) \Rightarrow \exists R_B(d)$ ;
- существуют ситуации, в которых  $A$  не находит решения, а  $B$  – находит.

Существует большое количество публикаций, посвященных количественным оценкам и сравнению уровня интеллектуальности разных ИС. Используемые подходы отличаются набором параметров, по которым строятся оценки.

Например, в [14] уровень интеллектуализации ИС  $I(A)$  определяется наличием у нее свойств  $S_A \subseteq S = \{s_1, \dots, s_9\}$ :

- $s_1$  – автономность работы;
- $s_2$  – умение взаимодействовать с другими объектами;
- $s_3$  – способность воспринимать информацию об окружающем мире;
- $s_4$  – способность применять абстракции;
- $s_5$  – способность использовать знания о ПрО;
- $s_6$  – адаптивность поведения;
- $s_7$  – способность к обучению;
- $s_8$  – толерантность к ошибкам;
- $s_9$  – способность к общению на естественном языке.

Такой подход позволяет в некоторых случаях сравнивать уровень интеллектуализации ИС из различных ПрО (если у системы  $A$  есть все свойства системы  $B$ , но у системы  $B$  есть свойства, которых нет у системы  $A$ , то  $I(B)$  – уровень интеллектуализации системы  $B$  – выше, чем у системы  $A$ ):

$$\exists (S_A \subseteq S, S_B \subseteq S, S_A \subseteq S_B) \Rightarrow I(B) \geq I(A).$$

Однако многие ИС оказываются при таком подходе несравнимыми:

$$\exists (S_A \subseteq S, S_B \subseteq S, S_A \not\subseteq S_B, S_B \not\subseteq S_A) \Rightarrow I(B) \text{ и } I(A) \text{ несравнимы.}$$

Например, сложно сравнить уровень интеллектуализации систем, способных к общению на естественном языке, но не способных к обучению (различные чат-имитаторы), и систем индуктивного извлечения новых знаний, не поддерживающих естественно-языковой диалог. Кроме того, все эти параметры могут быть присущи ИС в большей или меньшей степени, а оценить количественно некоторые из них (например, способность использовать знания о ПрО) крайне затруднительно.

В [15] предлагается количественная оценка уровня интеллектуализации ИС:

$$I(A) = \sum_{i=1}^6 s_i \cdot w_i,$$

где  $w_i$  – вес  $i$ -го свойства, а  $s_i$  – его значение. При этом:

- $s_1$  – использование модели окружающего мира для формирования планов собственных действий;
- $s_2$  – составление альтернативных вариантов при планировании действий;
- $s_3$  – способность реконструировать план во время его выполнения, если выполняемые действия приводит к нежелательным последствиям;
- $s_4$  – использование собственного опыта для расширения и коррекции модели ПрО;
- $s_5$  – общение с пользователем на естественном языке;
- $s_6$  – допустимость периода выполнения плана решения задачи.

Однако пользователь ИС может самостоятельно оценить только два последних параметра, и поэтому ему приходится полностью полагаться на утверждения разработчиков системы, которые не всегда объективны и корректны.

На основе приведённого выше анализа публикаций предлагается использовать следующий критерий оценки уровня интеллектуальности ИС:

$$I(A) = f(T, X, L, C, P, K),$$

где

- $T$  – тип атомарных элементов;
- $X$  – количество атомарных элементов;
- $L$  – количество связей между элементами;
- $C$  – количество команд в нормализованном алгоритме;

- $P$  – соотношение между количеством успешных экспериментов и количеством проведенных;
- $K$  – класс задач, которые решает ИС.

Сложность структуры данных, которые способна обрабатывать ИС, характеризуют тип атомарных элементов  $T$ , их количество  $X$  и количество связей между ними  $L$ .

Сложность методов преобразования данных характеризует количество команд в нормализованном алгоритме  $C$  (нормализованный алгоритм – это любая формализация понятия алгоритма, например, нормальные алгоритмы Маркова или машина Тьюринга).

Вероятность того, что действия ИС приводят к цели, которая декларируется системой, оценивается как соотношение количества успешных экспериментов с количеством проведенных  $P$ . Этот параметр пользователь обычно может оценить довольно точно. Например, для ИПС этот параметр определяется через пертинентность поиска – какую часть из предложенных результатов выполнения запроса пользователь считает удовлетворительными ответами на свой вопрос.

В предлагаемом критерии применяются параметры, значения которых могут оценивать количественно не только разработчики ИС, но и пользователи системы – с точки зрения своих индивидуальных преимуществ и целей.

Оценивать эти параметры системы можно независимо, трансформируя критерий интеллектуальности следующим образом:

$$I(A) = f(T, X, L, C, P, K) = g(T, X, L) \cdot h(C) \cdot P \cdot \phi(K).$$

Для простоты оценки мы используем  $g(T, X, L) = g_1(T) \cdot g_2(X, L)$ , где функции  $g_1(T)$  и  $g_2(X, L)$  задаются таблично. Аналогично таблично определяются  $h(C)$  и  $\phi(K)$ . Эти оценки задаются пользователями ИС априорно на основании собственного опыта использования и анализа информационных технологий в разных областях и требуют дальнейшей детализации, расширения и уточнения.

$T$  позволяет учитывать, какие ИО используются в ИС как входные и выходные данные, насколько сложную структуру они имеют и насколько сложные методы нужны для их анализа и генерации. Для этого пользователь может обращаться к разным таксономиям ИО. Если в ИС обрабатываются ИО разных типов, то в качестве оценки уровня интеллектуальности ИС может быть принята оценка наиболее сложного из ИО:

$$g_1(T) = g_1(IO_1, \dots, IO_n) = \max_{i=1}^n g_1(IO_i).$$

Однако такой подход не будет удовлетворительным для ситуаций, когда нужно сравнить две ИС, одна из которых является интеллектуальным развитием другой. Действительно, оценка, определяемая максимальным значением, может и не измениться, хотя общая картина используемых структур данных изменится. Поэтому целесообразно использовать критерий, который учитывает суммарную оценку ИО в ИС:

$$g_1(T) = g_1(IO_1, \dots, IO_n) = \sum_{i=1}^n g_1(IO_i).$$

Такой критерий даёт более полезные результаты, но остаётся проблема разделения влияния на уровень интеллектуальности сложности входных и выходных данных системы. Действительно, создать сложную структуру данных значительно труднее, чем извлечь из уже существующей структуры определённые данные. Например, сгенерировать по естественно-языковому тексту онтологию сложнее, чем использовать уже существующую онтологию для семантической разметки этого текста, но значение критерия интеллектуальности будет согласно суммарной оценке ИО для обеих ИС одинаковым.

Поэтому целесообразно использовать два отдельных критерия – для входных данных и для выходных данных. Выходными данными ИС будем считать те ИО, которые создаются

или изменяются вследствие работы ИС, а входными – все другие данные, изменения в которых могут повлиять на формирование выходных данных:

$$\begin{aligned} g_1(T) &= a_{input} \cdot g_{input}(T) + b_{output} \cdot g_{output}(T) = \\ &= a_{input} \cdot g_{input}(IO_1, \dots, IO_{n_{input}}) + b_{output} \cdot g_{output}(IO_1, \dots, IO_{n_{output}}) = \\ &= a_{input} \cdot \sum_{i=1}^{n_{input}} g_1(IO_i) + b_{output} \cdot \sum_{i=1}^{n_{output}} g_1(IO_i), \end{aligned}$$

где  $\bigcup_{i=1}^{n_{input}} IO_i \cup \bigcup_{i=1}^{n_{output}} IO_i = T$ , а коэффициенты  $a_{input}$  и  $b_{output}$  оценивают важность входных и выходных данных и зависят от специфики ИС, как правило,  $a_{input} < b_{output}$ .

Если в более ранних интеллектуальных ИС обычно использовались достаточно простые типы атомарных элементов, которые соответствовали типам данных в языках программирования (такие, как целые и действительные числа, символы и строки символов, таблицы, имена файлов и каталогов) или связанные с моделями представления знаний (фреймы, семантические сети, графовые структуры и т.п.), то для современных семантических Web-применений характерны ИО с более сложной структурой, которые связаны с определёнными объектами реального или виртуального мира (программные агенты, Web-сервисы, семантически размеченные ресурсы, элементы Web of Things, онтологические описания и т.п.).

Рассмотрим пример интеллектуализированных метапоисковых систем, которые обрабатывают результаты поиска, получаемые от одной общей внешней ИПС.

Первая ИПС упорядочивает найденные по запросу пользователя документы, учитывая его персональные свойства, выявляемые на основе обработки информации из истории взаимодействия с пользователем. Вторая ИПС упорядочивает найденные по запросу пользователя документы, используя для формализации сферы интересов пользователя онтологию соответствующей ПрО, а третья ИПС не только предоставляет пользователю упорядоченный список ссылок на документы, но и выполняет семантическую разметку найденных документов терминами из онтологии ПрО.

Входные данные первой ИПС – это запрос (строка символов), набор ссылок на документы, полученный от внешней ИПС, и набор естественно-языковых документов (предшествующие результаты поиска), а выходные – переупорядоченный набор ссылок на документы. Для второй ИПС входные данные – это запрос, набор ссылок на документы, полученный от внешней ИПС, и онтология ПрО, а выходные – переупорядоченный набор ссылок на документы. Для третьей ИПС входные данные – это запрос, набор ссылок на документы, полученный от внешней ИПС, онтология ПрО и лексическая онтология ПрО, а выходные – переупорядоченный набор ссылок на документы и сами семантически размеченные документы. Третья ИПС способна использовать и генерировать более сложные ИО. Это позволяет отнести третью ИПС к наиболее интеллектуальным.

Следует отметить, что это касается только формальных оценок, а реальные результаты работы ИС зависят от того, насколько те модели представления информации и алгоритмы их обработки, которые в них применяются, соответствуют целям ИС, и от того, насколько удачно и корректно они реализованы. Например, если ССП использует эффективные методы поиска и обрабатывает онтологии ПрО поиска, но переадресует запросы к внешней ИПС с небольшой индексной базой и слабыми поисковыми механизмами, то получаемые результаты не удовлетворят пользователя, несмотря на её высокий уровень интеллектуальности.

Оценка  $g_2$  зависит от  $X$  и  $L$ . Например, можно использовать  $g_2(X, L) = X \cdot L$  или  $g_2(X, L) = X^L$ .

Пользователь не может точно оценить эти параметры, но может дать им приблизительную качественную оценку, введя шкалу интервалов для значений параметров: «очень мало», «мало», «средне», «много», «очень много». Их обработка базируется на теории нечётких

множеств, где мультипликация нечётких значений определяется как произведение их относительных весов, а объединение – как их сумма.

Кроме формальных параметров, пользователь при оценке уровня интеллектуальности может учитывать, насколько сложную задачу решает эта ИС.

Если ИС  $B$  является интеллектуальной надстройкой над ИС  $A$ , тогда уровень интеллектуальности ИС  $B$  с точки зрения пользователя определяется как сумма уровней интеллектуальности этих двух систем. Например, если  $A$  – информационно-поисковая система, которая предоставляет результаты по запросу пользователя, а ИС  $B$  упорядочивает эти результаты соответственно персональным вкусам этого пользователя, то пользователь не имеет возможности оценивать в отдельности свойства любой из этих систем, и может подсчитать по очерченной выше методике  $I(A \cup B)$ .

Таким образом, можно считать, что уровень интеллектуальности системы в значительной мере определяется тремя параметрами:

- 1) способностью использовать знания со сложной структурой;
- 2) наличием развитых методов обработки таких знаний (например, нечеткой логики);
- 3) способностью приобретать новые знания из открытой среды.

Поэтому развитие прикладных ССП в направлении их интеллектуализации связано именно с разработкой подсистем управления знаниями, средств обучения и алгоритмов анализа знаний, представленных сложными онтологическими структурами.

## 2 Постановка задачи

При построении ССП необходимо разработать формальную модель, которая адекватно описывает взаимодействие между основными субъектами и объектами семантического поиска: с одной стороны – это пользователи и их информационные потребности с персональными предпочтениями; с другой – ИР, содержащие сведения об интересующих пользователей ИО. Для такого сопоставления модель должна использовать знания, извлекаемые из внутренних и внешних баз знаний. Это требует разработки методов формирования и пополнения в процессе поиска соответствующих баз знаний, а также способов использования этих знаний для повышения пертинентности поиска.

## 3 Онтологическая модель семантического поиска

Анализ публикаций показывает, что в большинстве существующих разработок, связанных с семантическим поиском, базы знаний используются неявным для пользователя способом, не оставляя ему возможности самостоятельно влиять на выбор тех знаний, которые он считает важными для поиска. Это приводит к следующим проблемам: с одной стороны, пользователям непонятны пути получения результатов, и это снижает доверие к поисковым системам; а с другой, во многих случаях такие системы ищут совсем не то, что нужно пользователю. Кроме того, возникает проблема *обнаружения* источников достаточно формализованных и структурированных знаний, а также и проблема динамического *обновления* знаний, которые используются для поиска.

Представляется, что с этой целью могут использоваться онтологии [16-18]. Но следует учитывать, что применение онтологий, несмотря на такие преимущества, как явное представление семантики и строгий математический базис (дескриптивные логики), имеет и значительные недостатки, к которым относится сложность обработки и логического вывода на них. Поэтому целесообразность использования онтологий и их частных случаев для пред-

ставления знаний относительно разных элементов семантического поиска является предметом отдельного исследования.

Нужно отметить ещё один важный аспект – разработчики средств семантического поиска, как правило, движутся путём усовершенствования традиционных поисковых средств, не учитывая опыты создания интеллектуальных приложений (начиная от экспертных систем и заканчивая созданием мультиагентных систем). Так от систем, функционирующие которых подобно интеллектуальной деятельности человека, пользователи требуют объяснений путей получения результатов (системы пояснений в экспертных системах) и понятной наглядной модели поведения и использования знаний [19]. В семантическом поиске пользователю тоже нужно явным образом и наглядно показывать, какие именно его знания и предпочтения влияют на полученный результат.

При семантическом поиске формируется информационная модель пользователя, в которой отображаются:

- персональные предпочтения пользователя, его знания и способность к восприятию информации;
- информационные потребности пользователя, особенности стоящей перед ним задачи и тех ИО, обнаружение сведений о которых позволит ему решить эту задачу;
- опыт взаимодействия пользователя с поисковой системой;
- сведения, характеризующие отношение пользователя к опыту других пользователей и определяющие целесообразность использования коллаборативного поиска.

Эта модель сопоставляется с моделями ИР, в которых могут содержаться полезные для пользователя сведения. Модель ИР может включать как семантическую разметку и метаданные об ИР, так и оценки этих ИР различными пользователями. Для семантической разметки применяется онтологическая модель ПрО, интересующей пользователя, т.е. такая разметка должна выполняться заново для каждой новой ПрО.

Использование онтологического подхода обеспечивает интероперабельность и повторное использование знаний – все эти модели в ССП могут динамично пополняться за счёт информации из внешних онтологий и анализа внешних ИР. Знания могут импортироваться из других интеллектуальных ИС (например, из семантических Wiki-ресурсов) и экспортироваться в другие системы, с которыми работают те же пользователи (например, можно переносить сведения о пользователе из ССП в персонифицированную систему дистанционного обучения или в систему оценки квалификации).

Наиболее важные классы онтологической модели ССП - это:

- пользователь – лицо, нуждающееся в информации;
- ИО – структурированный образ той информации, которую пользователь намерен получить от ССП [20].

Кроме того, в онтологической модели взаимодействия пользователей и ИР присутствуют такие классы как:

- онтология ПрО, которая описывает область, к которой относятся информационные потребности пользователя;
- лексическая онтология ПрО, позволяющая распознавать сведения о терминах онтологии в естественно-языковых текстах;
- тезаурус задачи – термины онтологии, совокупность которых характеризует ту конкретную задачу из ПрО, которую в данный момент решает пользователь, и их вес;
- информационная потребность пользователя, для удовлетворения которой и производится поиск информации;
- запрос – явным образом переданное пользователем сообщение о наличии информационной потребности (как правило, это множество ключевых слов, характеризующих одну из

информационных потребностей пользователя, связанных с конкретной задачей при помощи тезауруса);

- тема – множество запросов, связанных с одной информационной потребностью разных пользователей, позволяющее объединять семантически связанные запросы;
- результат запроса – ссылки на ИР и их оценки;
- ИР;
- рекомендация – сведения, предоставляемые пользователю проактивно, вследствие действий других пользователей, участвующих в коллаборативном поиске;
- рекомендуемый элемент – ссылка на ИР, входящая в рекомендацию;
- информационная среда – совокупность всех доступных ИР, их свойств (включая их оценки пользователями) и связей между ними;
- группа пользователей.

Таким образом, для описания информационной потребности пользователя надо построить онтологическую модель, классы которой соответствуют перечисленным выше элементам, а отношения – связям между ними (рисунок 1). Формализованное описание элементов модели позволит обеспечить их автоматизированную обработку.

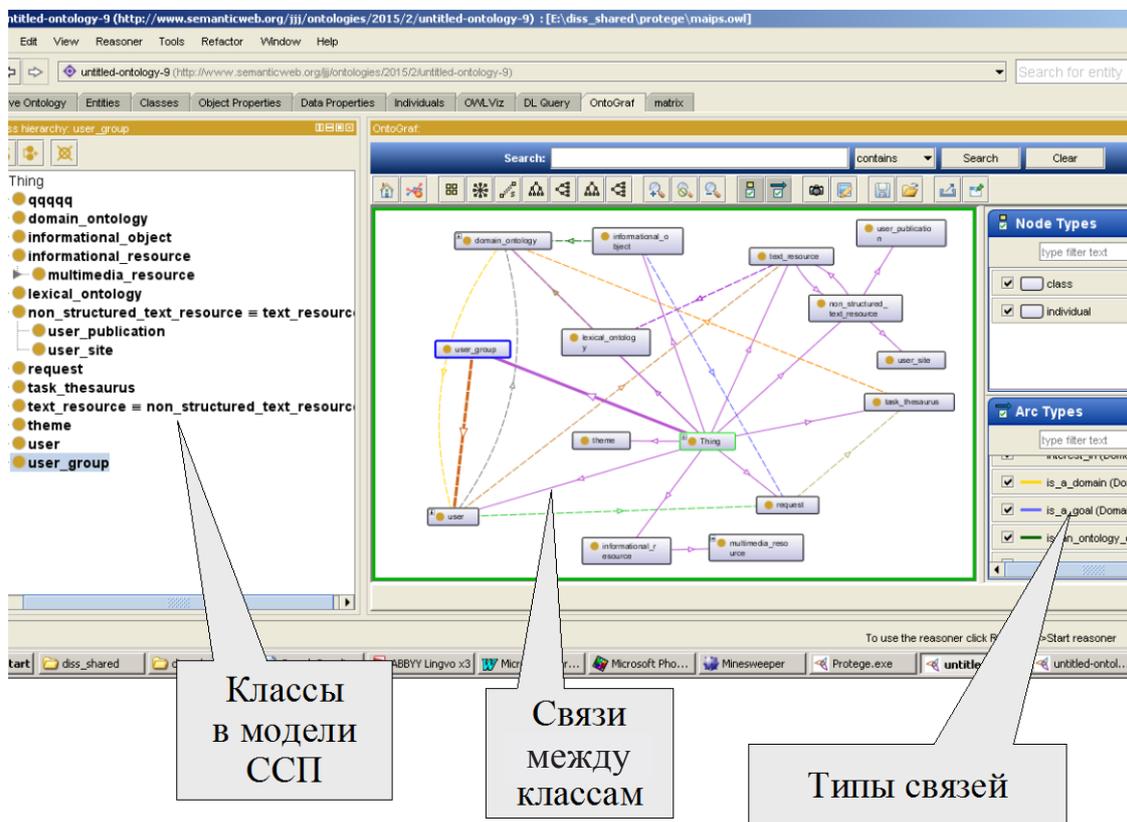


Рисунок 1 – Элементы онтологической модели взаимодействия пользователей и информационных ресурсов при семантическом поиске

Для того чтобы описать знания и интересы пользователя, в онтологическую модель вводятся два служебных класса – «компетенция» и «атомарная компетенция». Набор экземпляров атомарных компетенций может импортироваться из различных онтологий ПроО, связанных с учебными заведениями, трудоустройством и научно-исследовательскими организациями [21].

Такой подход ориентирован на относительно узкое подмножество поисковых систем, а именно на те, которые предназначены для поддержки научной деятельности и ориентирован на пользователей со стабильными, сложно структурированными знаниями, навыками и информационными потребностями в достаточно узкой ПрО. При этом через набор компетенций формализуются не все накопленные в процессе обучения и работы знания, а лишь то их небольшое подмножество, которое необходимо и активно используется при практической деятельности.

Например, если исследователь специализируется в сфере программной инженерии, то для него будет сформирован набор компетенций именно по этой дисциплине (а также часть компетенций из других дисциплин, которые непосредственно им используются – например, из теории алгоритмов или системного анализа), а также из тех научных работ, которые он сам пишет либо изучает.

За счет использования класса «компетенция» и его свойств можно более формально описать свойства класса «пользователь» таким образом, который допускает более простое сопоставление с онтологиями ПрО и свойствами ИР. Чтобы различать компетенции реально используемые и когда-либо изученные, предполагается использовать систему весов, оценивающих уровень владения пользователем каждой из компетенций. Чтобы снизить вычислительную сложность задачи, при поиске специалистов того или иного уровня все веса ниже порогового значения заменяются нулевыми.

Чтобы упростить формирование набора компетенций каждого конкретного пользователя, целесообразно использовать иерархию компетенций, в которой на самом верхнем уровне определяются компетенции специальности или области знаний, на более низких – компетенции отдельных дисциплин или областей деятельности, а на самом нижнем – атомарные компетенции.

Значительная часть атомарных компетенций пользователя может автоматически импортироваться из множества компетенций его специальности (или специальностей), подтвержденных определёнными дипломами и сертификатами, но с использованием их весов, позволяющих учитывать их текущее состояние для пользователя. Например, если пользователь на первом курсе университета изучал матанализ, но затем никогда его не применял, то соответствующие этому предмету компетенции получают не нулевые, но достаточно низкие веса. Тем не менее, отличие этого пользователя от тех, кто вообще не знаком с данным предметом, также может быть важно и потому будет явно зафиксировано.

Некоторая часть компетенций пользователя может быть импортирована из социальных сетей, где фиксируются его области интересов.

Следует учитывать, что большинство специалистов, имеющих достаточный опыт работы, значительную часть своих компетенций получают и развивают именно в процессе своей профессиональной деятельности, а не при обучении. Поэтому иногда целесообразно импортировать компетенции из описаний занимаемых должностей (хотя для многих специальностей - особенно связанных с научно-исследовательской работой, - такие описания имеют настолько общий характер, что не пригодны для анализа), а при определении весов учитывать опыт работы в данной области.

Кроме элементов, в модели должны отображаться связи между:

- задачами пользователей и ПрО – как отношение между тезаурусом задачи и онтологией ПрО;
- ИР и ПрО – как отношение между тезаурусом ИР и онтологией ПрО;
- запросами и задачами – как отношение между запросом и тезаурусом задачи;
- различными пользователями из группы.

Поясним эти замечания на примере ССП РЕПС, предназначенной для выполнения многообразных информационных запросов в фиксированных ПрО.

РЕПС представляет собой ССП, поддерживающую персонафицированный и коллаборативный поиск. Система ориентирована на пользователей, которые имеют устоявшиеся информационные интересы и требуют постоянного поступления соответствующей информации из Web. Функционально РЕПС реализует выполнение сложных многообразных запросов в довольно узких областях, связанных с профессиональными или научными интересами пользователей. Запросы таких пользователей могут повторяться от сеанса к сеансу или изменяться, но ПрО поиска, в которых пользователи являются экспертами, практически не изменяются.

Для формализации сферы интересов пользователей используются онтологии ПрО, а для конкретизации запросов, относящихся к решению конкретной задачи, – тезаурусы задачи, построенные на основе описания задачи и онтологии соответствующей ПрО [22].

РЕПС является развитием интеллектуальной поисковой системы МАИПС [20], расширенной возможностями коллаборативного поиска (в частности, критериями для создания групп пользователей) и более специализированными моделями пользователей (включающей психофизиологические, социальные и компетентностные параметры) и ИО (в частности, пользователь может задавать желаемый уровень читабельности текста).

#### 4 Источники сведений об элементах модели

Каждый из элементов информационной модели РЕПС является классом соответствующей онтологии  $O_{РЕПС}$ , а сведения об их значениях сохраняются как описание экземпляров этих классов:

$$O_{РЕПС} = \langle T_{РЕПС}, R_{РЕПС}, F_{РЕПС} \rangle.$$

Для целей поиска информации используется довольно простой частный случай:

$$O_{РЕПС} = \langle T_{РЕПС}, R_{РЕПС}, \emptyset \rangle.$$

Сведения об экземплярах классов, которые входят во множество  $T_{РЕПС}$ , накапливаются в процессе взаимодействия РЕПС с пользователями, которые с помощью этой системы стараются удовлетворить собственные информационные потребности.

Рассмотрим подробнее, откуда именно попадает эта информация в онтологию информационной модели РЕПС и в каком порядке. Разумеется, ещё к началу взаимодействия с пользователем эта онтология содержит определённые знания, нужные для осуществления поиска на семантическом уровне (в процессе функционирования системы эти знания могут пополняться).

**Источники сведений об онтологиях ПрО.** Для представления знаний в РЕПС используются онтологии двух видов – внутренние и внешние.

Внутренние онтологии создаются непосредственно разработчиками РЕПС и могут пополняться в процессе взаимодействия РЕПС с пользователями. Основной особенностью таких онтологий является то, что разработчикам полностью известна их структура и содержимое, поэтому можно прогнозировать конечность вычислений.

Кроме того, априорные знания об отношениях между классами и экземплярами таких онтологий могут обрабатываться и анализироваться специализированными сервисами (например, для пополнения знаний относительно области интересов пользователя может использоваться сервис семантического анализа выбранных ИР, а для определения психофизиологических особенностей пользователя – специализированные сервисы экспресс-тестирования).

Внешние онтологии позволяют интегрировать в РЕПС динамические и распределённые знания, доступ к которым обеспечивает Web. Поиск таких онтологий может осуществляться

в разнообразных репозиториях, или же они могут быть сформированы в процессе работы пользователя с другими интеллектуальными приложениями. В общем случае информация о сложности структуры таких онтологий, о том, на каких дескриптивных логиках они базируются, насколько полными являются такие знания, отсутствует. Это не позволяет прогнозировать время работы алгоритмов для глубокого анализа и обработки таких онтологий. Поэтому в РЕПС для обработки внешних онтологий применяются упрощённые алгоритмы, которые используют только наиболее простые свойства онтологий (например, обрабатываются только отношения класс-подкласс).

Для подавляющего большинства задач информационного поиска этого достаточно, но для поиска совокупностей сложных ИО – например, для исследования и композиции семантических Web-сервисов, поиска Web-сущностей или формирования мультиагентных систем, – нужно использовать и более сложные структурные связи. Именно для таких ситуаций в состав РЕПС входят онтологии сложных ИО.

В элементах класса «онтология ПрО» содержатся ссылка на внешние онтологии  $O_{\text{ПрО}}$  (в отличие от внутренних онтологий РЕПС, сведения о структуре и сложности таких онтологий недоступны) и описания их тематики. Эти сведения заносятся в онтологию  $O_{\text{РЕПС}}$  разработчиками системы вручную и базируются на непосредственном анализе содержимого таких онтологий или экспортируются из внешних репозиториях, какие разработчики РЕПС признали надёжными.

Если в процессе работы РЕПС у пользователей возникает интерес к тем ПрО, которые не отображены в имеющихся онтологиях, то пользователь может подать запрос разработчикам относительно пополнения списка онтологий, нужным ему для работы.

Непосредственно пользователь не может пополнять список онтологий без согласования с разработчиками. Это вызвано тем, что обычно пользователь не может оценить вычислительную сложность онтологии ПрО (например, определить, на какой именно дескриптивной логике она базируется) и потому не может предусмотреть, допускает ли такая онтология эффективную обработку алгоритмами, встроенными в РЕПС.

**Источники сведений о лексических онтологиях ПрО.** Лексические онтологии ПрО разрабатываются автоматизированно для каждой онтологии ПрО, входящей в состав РЕПС.

Источник сведений о лексических онтологиях ПрО – специализированный блок лексического анализа, который в полуавтоматическом режиме связывает с терминами онтологии ПрО релевантные им фрагменты естественно-языковых текстов.

Каждый экземпляр класса «лексическая онтология» содержит множество экземпляров класса «лексическое соответствие».

В простейшем варианте (например, для тех языков, для которых в РЕПС не разработаны специальные методы лексического анализа) «лексическое соответствие» – это множество пар <термин онтологии, фрагмент текста>. Такие пары могут формироваться и вручную, и автоматизированно. Кроме того, существующие лексические онтологии могут модифицироваться пользователем с учётом специфики его задач. Затем они могут использоваться для семантической разметки текста с точки зрения интересов пользователя.

**Источники сведений о тезаурусах.** Тезаурус задачи строится непосредственно пользователем для того, чтобы отобразить специфику той задачи, в процессе решения которой у него возникла данная информационная потребность.

Пользователь может строить тезаурус по выбранной онтологии ПрО, отмечая нужные термины и определяя их вес. Пользователь может также вручную редактировать любой раньше созданный тезаурус. Кроме того, в системе реализованы теоретико-множественные операции объединения, пересечения и дополнения над тезаурусами.

Ещё один способ построения тезаурусов основан на анализе описания задачи с использованием лексической онтологии соответствующей ПрО. Можно рассматривать его как проекцию задачи на ПрО: сервис находит в тексте описания задачи фрагменты, которые соответствуют терминам онтологии ПрО, и прибавляет к тезаурусу эти термины с весом 1.

**Источники сведений о запросах.** Запрос – это набор ключевых слов, которые предоставляет пользователь, и семантический контекст, формализованный через тезаурус задачи и сведения о пользователе, его задающем. Из-за того, что система РЕПС ориентирована на удовлетворение постоянных информационных потребностей пользователей, она предусматривает многократное выполнение запросов с учётом их семантики.

**Источники сведений о темах.** Темы могут создаваться пользователями для того, чтобы объединить несколько семантически связанных запросов. Одна тема может включать запросы разных пользователей, которые базируются на разных онтологиях и тезаурусах. Тема может быть создана пользователем путём явного задания входящих в неё запросов. Другой путь формирования тем – в автоматическом режиме связываются запросы по одной онтологии ПрО или с пересекающимися тезаурусами. Значительно упростить создание и редактирование тем позволяют теоретико-множественные операции над темами: объединение, пересечение и дополнение запросов, которые входят в состав отдельных тем.

**Источники сведений о результатах запроса.** В процессе формирования результатов запроса РЕПС перенаправляет ключевые слова запроса к внешней ИПС, а потом переупорядочивает полученный список ИР с учётом онтологии ПрО, тезауруса и персональных сведений пользователей. Для этого в РЕПС разработан и реализован ряд методов и алгоритмов семантического анализа ИР.

**Источники сведений о пользователях РЕПС.** Сведения о пользователях служат основой для персонификации семантического поиска и создания рекомендаций [23]. Получение и анализ знаний о них позволяют сделать информационный поиск значительно эффективнее.

**Источники сведений о группах онтологий РЕПС.** РЕПС предназначен для персонифицированного и коллаборативного поиска информации, что требует иметь возможность группирования пользователей соответственно семантике их информационных потребностей.

С этой целью выделяют группы пользователей, которые используют одну и ту же онтологию как основу для информационных запросов. Это означает, что таких пользователей интересует один домен знаний. Однако такая классификация является очень приблизительной из-за того, что одна онтология может описывать довольно широкую ПрО, а интересы пользователей могут касаться относительно узких регионов этой ПрО. Кроме того, следует учитывать, что в одном запросе могут использоваться термины из разных онтологий, а также термины, которые пользователь прибавил лично.

С целью создания проактивных рекомендаций в системе РЕПС к онтологии вводятся дополнительные служебные классы, предназначенные для накопления коллаборативных знаний: группа онтологий; группа терминов запроса и группа пересекающихся тезаурусов. Эти классы предназначены для того, чтобы находить пользователей, которые имеют подобные информационные потребности. При этом пользователь может сам, явным образом формировать группы тех пользователей, чьи информационные потребности могут быть ему интересны по тем или иным критериям (например, пользователи, применяющие ту же онтологию для формирования тезаурусов задачи, проживающие в одной стране и использующие одинаковые ключевые слова в запросах).

К сожалению, в большинстве существующих рекомендательных систем пользователь не видит, по каким критериям формируются рекомендации и не может влиять на их получение. Из-за этого пользователю нередко приходится просматривать ненужные сообщения.

В РЕПС пользователь сам инициирует получение рекомендаций (и может в любой момент отказаться от них или изменить параметры их получения) и, кроме того, может определять, какая часть его взаимодействия с поисковой системой может быть доступна для обработки коллегами. Поэтому пользователь РЕПС может прогнозировать получаемые результаты и управлять коллаборативным поиском.

## 5 Использование онтологической модели ССП в прикладных ИС

Предложенную модель семантического поиска достаточно легко адаптировать для решения различных прикладных задач. Рассмотрим это на примере проблемы сопоставления компетенций [24], которая является составной частью таких задач, как поиск работодателем подходящих исполнителей работ; сравнение квалификации специалистов, имеющих различные специальности (в частности, соответствующие нормативам различных стран); выбор абитуриентом учебного заведения, предлагающего необходимый ему набор дисциплин; оценка возможности перехода студента из одного учебного заведения в другое (какие дисциплины из ранее изученных можно засчитывать) и т.д.

Модель ПрО поиска описывает базовые понятия и связи между ними, а также структуру ИО, который является результатом поиска (рисунок 2).

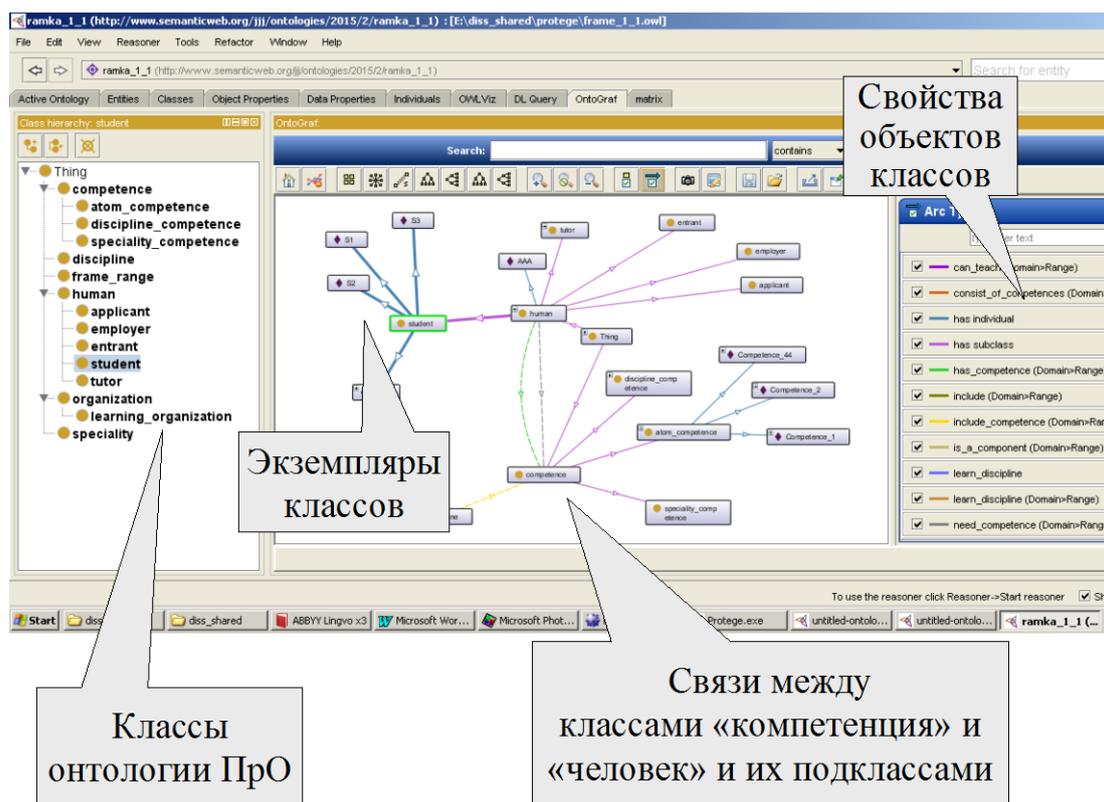


Рисунок 2 – Онтологическая модель сопоставления компетенций

Пользователь описывает свою информационную потребность, указывая класс онтологии ПрО  $A$ , к которому относится искомый ИО – человек, учебное заведение, специальность и т.д., и накладываемые на него условия – множество экземпляров класса «Атомарная компетенция», с которыми этот объект связан выбранными отношениями онтологии  $B$ , которые описывают связи между экземплярами.

Таким образом, запросы сводятся к сопоставлению значений свойств одного класса для экземпляров различных классов. Например, сопоставление свойств «включает компетенции», которое относится к классу «Атомарная компетенция», для экземпляра класса «Специалист» и для экземпляра класса «Специальность».

При этом явно указывается семантика информационной потребности, и потому можно осуществлять поиск, дифференцирующий различные отношения между искомым ИО и набором компетенций. Например, с одним и тем же экземпляром класса «человек» некоторые различные подмножества атомарных компетенций могут быть связаны отношениями «обладает», «имеет сертификат», «может преподавать», «имеет опыт использования». Это позволяет точнее удовлетворять информационную потребность пользователя, находя ИО, соответствующие его требованиям.

Для сопоставления специальностей, умений и компетенций людей и организаций разных стран целесообразно использовать набор эталонных атомарных экземпляров каждого класса. Экземпляр считается атомарным, если ни один другой экземпляр этого класса не является его подмножеством.

Например, если две компетенции  $A$  и  $B$  пересекаются, то по ним строится три потенциально атомарные компетенции  $A_1$ ,  $B_1$  и  $C$ , такие, что

$$A \cap B = C, A_1 \cup C = A, B_1 \cup C = B.$$

Поэтому целесообразно пополнить онтологическую модель ПрО дополнительным служебным классом. Класс «Атомарная компетенция» – подкласс класса «Компетенция», такой, что для  $\forall a \in$  «Атомарная компетенция» существует хотя бы один элемент класса  $\forall b \in$  «Компетенция», такой, что  $a \subseteq b$ , но для ни одного элемента из класса «Атомарная компетенция» не существует другого элемента этого класса  $c \in$  «Атомарная компетенция», такого, что  $c \subseteq a$ ,  $a \not\subseteq c$ . Класс «Атомарная компетенция» имеет свойство «входить в состав» класса «Дисциплина» и свойство «входить в» класс «Компетенция».

В таком случае пользователь описывает свою информационную потребность, указывая класс онтологии ПрО  $A$ , к которому относится искомый ИО – человек, учебное заведение, специальность и т.д., - и накладываемые на него условия: множество экземпляров класса «Атомарная компетенция», с которыми этот объект связан выбранными отношениями онтологии  $B$ , описывающими связи между экземплярами.

Важным достоинством предложенной модели поиска является то, что при таком описании явным образом указывается семантика информационной потребности. Вследствие этого можно осуществлять поиск, который дифференцирует различные отношения между искомым ИО и набором компетенций.

Например, с одним и тем же экземпляром класса «человек» некоторые различные подмножества атомарных компетенций могут быть связаны отношениями «обладает», «имеет сертификат», «может преподавать», «имеет опыт использования».

Такая дифференциация позволяет значительно более точно удовлетворять информационную потребность пользователя, обнаруживая именно те ИО, которые соответствуют его требованиям.

Формирование множества атомарных компетенций, которое требует значительных интеллектуальных усилий экспертов и может быть автоматизировано только частично, с пополнением базы знаний сведениями об экземплярах ИО, требующим постоянного обновления большого объема данных.

С учётом того, что современные интеллектуальные приложения, как правило, функционируют в открытом информационном пространстве, в качестве источников знаний об экземплярах ИО целесообразно использовать IP Web:

- различные государственные и международные стандарты и нормативы, связанные с описанием структуры и уровня;
- официальные сайты и организационные онтологии [25] учебных заведений, содержащие информацию о преподаваемых специальностях и входящих в их состав дисциплинах;
- внешние оценки различных учебных заведений и организаций, позволяющие оценить качество получаемых в них компетенций;
- персональные сведения о людях, извлекаемые из социальных сетей, Wiki-ресурсов и различных естественно-языковых документов, доступных через Web (научных публикаций, технических отчетов, методических материалов и т.д.).

## Заключение

Онтологический анализ является сегодня одной из важных составляющих развития интеллектуализированных Web-приложений [26]. Использование онтологической модели взаимодействия субъектов и объектов системы семантического поиска обеспечивает повторное использование составляющих её знаний; даёт пользователю чёткое представление как о возможностях системы, так и о её поведении; предоставляет возможность для оценки уровня интеллектуальности системы и, что наиболее важно, позволяет интегрировать ССП с различными прикладными системами, которые для своего функционирования нуждаются в семантическом поиске сложных ИО.

## Благодарности

Работа была выполнена при частичной поддержке проекта «Разработка интеллектуальной системы информационного и когнитивного сопровождения функционирования Национальной рамки квалификаций».

## Список источников

- [1] *Рогушина, Ю.В.* Знание-ориентированные средства поддержки семантического поиска в Web / Ю.В. Рогушина. – LAP LAMBERT Academic Publishing, 2014. – 214 с.
- [2] *Amerland, D.* Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence / D. Amerland. – Que Publishing, 2013. – 230 p.
- [3] Wolfram Alpha computational knowledge engine, 2009. – <http://basetechnology.blogspot.com/2009/03/wolfram-alpha-computational-knowledge.html>.
- [4] *Ясин, Е.* Теоретические проблемы развития информационных систем / Е. Ясин // Модели данных и систем баз данных: Тр. совмест. сов.-амер. семинара. – М.: Наука, 1979. – С. 5-30.
- [5] W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity/>.
- [6] Introduction to Semantic Web. – <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
- [7] *Минский, М.* Фреймы для представления знаний / М. Минский. – М.: Энергия, 1979. – 152 с.
- [8] *Глушков, В.М.* Кибернетика / В.М. Глушков // Математическая энциклопедия. Т. 2 – М.: Сов. энциклопедия, 1979. – С. 850-856.
- [9] *Васильев, С.Н.* От классических задач регулирования к интеллектуальному управлению / С.Н. Васильев // Известия Академии наук. Теория и системы управления. 2001. №2. – С. 5-21.
- [10] *Zadeh, L.A.* Roles of soft computing and fuzzy logic in the conception, design and deployment of information/intelligent systems / L.A. Zadeh // Computational intelligence: soft computing and fuzzy-neuro integration with applications. - Springer Berlin Heidelberg, 1998. – P. 1-9.
- [11] *Валькман, Ю.П.* О структуре исследований в области искусственного интеллекта / Ю.П. Валькман // Материалы X Международной научно-технической конференции «Системный анализ и информационные технологии». - К.: НТУУ «КПИ», 2008. – С. 25. – [http://lib.znate.ru/pars\\_docs/refs/81/80229/80229.pdf](http://lib.znate.ru/pars_docs/refs/81/80229/80229.pdf).

- [12] **Финн, В.К.** Искусственный интеллект: идейная база и основной продукт // Труды IX Национальной конференции с международным участием «Искусственный интеллект-2004». Т.1, 2004. – С. 11–20.
- [13] **Любич, А.А.** О выборе критериев оценки интеллектуальности информационной системы / А.А. Любич, В.Л. Плескач, Ю.В. Рогушина // УСИМ. 2005. №1. – С. 3-7.
- [14] **Гаврилова, Т.А.** Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский – СПб.: Питер, 2001. – 382 с.
- [15] **Кузьменко, Г.С.** Прагматичний підхід до оцінки рівню інтелекту інтелектуалізованих систем / Г.С. Кузьменко, В.А. Литвинов // Математичні машини і системи. 2003. №1. – С. 3-9.
- [16] **Gruber, T.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing / T. Gruber // International Journal of Human-Computer Studies. 1995. Vol. 43. Issues 5-6. – P. 907-928.
- [17] **Uschold, M.** Ontologies: Principles, Methods and Applications / M. Uschold, M. Grüninger // Knowledge Engineering Review. 1996. 11(2). – P. 93-155.
- [18] **Смирнов, С.В.** Онтологии как смысловые модели / С.В. Смирнов // Онтология проектирования. 2013. №2(8). – С. 12-19.
- [19] **Gilbert, N.**, Platforms and methods for agent-based modeling / N. Gilbert, S. Bankes // Proc. of the National Academy of Sciences, 99 (suppl 3), 2002. – P. 7197-7198.
- [20] **Рогушина, Ю.В.** Разработка онтологической модели информационной потребности пользователя при семантическом поиске / Ю.В. Рогушина // Онтология проектирования. 2014. №2(12). – С. 61-82.
- [21] **Rogushina, J.** Ontology-based competency analyses in new research domains / J. Rogushina, A. Gladun // Journal of Computing and Information Technology. 2012. Vol. 20, No. 4. – P. 277-293.
- [22] ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval / Geneva: International Organization for Standards, 2011.
- [23] **Middleton, S.** Ontology-Based Recommender Systems / S. Middleton, D.De Roure, N. Shadbolt // Handbook on Ontologies. Eds.: S. Staab, R. Studer. - Springer, 2009. – P. 779-796.
- [24] **Harzallah, B.** Knowledge management for competence management / B. Harzallah // Proc. of Int. Conf. on Knowledge Management and Knowledge Technologies I-KNOW'05, 2005. – P. 449-460.
- [25] An organization ontology, W3C Working Draft, 2012. - <http://www.w3.org/TR/2012/WD-vocab-org-20120405/>.
- [26] **Боргест, Н.М.** Онтологии: современное состояние, краткий обзор / Н.М. Боргест, М.Д. Коровин // Онтология проектирования. 2013. №2(8). – С. 49-55.

## APPLICATION OF THE ONTOLOGICAL MODEL FOR SEMANTIC SEARCH OF THE INFORMATION OBJECTS

**J.V. Rogushina**

*Institute of Software Systems of the National Academy of Sciences of Ukraine, Kiev, Ukraine  
ladamandraka2010@gmail.com*

### Abstract

The article describes the modeling of an intelligent system aimed to support interaction of information resources and information consumers that use internal and external knowledge bases. Existing approaches to estimation of intelligent level of information systems and expedience of their applicability towards information retrieval systems are analyzed. On basis of this analysis, a comparison criterion of intelligent level of different applications is proposed. The criteria is based on such parameters as type and number of processed atomic elements and relations among them and processing algorithm complexity. Ontological model of interaction between Web resources and users of the semantic search system is developed. Sources and methods of development and improvement of this model are proposed. The model applications for increasing the intelligence level of the semantic search system are proposed. Software realization of personalized and collaborative semantic search system on basis of this ontological model is described. The ways of the models application for integration of search instrument with applied systems are described (on example of competence matching task that is used for employment assistance, qualification comparison, educational mobility etc.).

**Keywords:** *semantic search, information object, ontological model, thesaurus, collaborative search.*

## Acknowledgements

The work was carried out with partial support from the project «Designing of an intellectual system of informational and cognitive support of the National qualifications frame functioning».

## References

- [1] **Rogushina, J.V.** Znanie-orientirovannye sredstva podderzhki semanticheskogo poiska v Web [Knowledge-oriented means of semantic search support in the Web] / J.V. Rogushina. – LAP LAMBERT Academic Publishing, 2014. – 214 p. (In Russian).
- [2] **Amerland, D.** Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence / D. Amerland. – Que Publishing, 2013. – 230 p.
- [3] Wolfram Alpha computational knowledge engine, 2009. – <http://basetechnology.blogspot.com/2009/03/wolfram-alpha-computational-knowledge.html>.
- [4] **Yasin, E.** Teoreticheskie problemy razvitiya informacionnykh sistem [Theoretical problems of information system development] / E. Yasin // Data and data base systems models. Proc. of joint Soviet-American seminar. – Moscow: Science, 1979. – P. 5-30. (In Russian).
- [5] W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity/>.
- [6] Introduction to Semantic Web. – <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
- [7] **Minsky, M.** Freymi dlya predstavleniya znaniy [Frames for knowledge representation] / M. Minsky. – Moscow: Energy, 1979. – 152 p. (In Russian).
- [8] **Glushkov, V.M.** Kibernetika [Cybernetics] / V.M. Glushkov // Mathematical encyclopedia. T. 2. – Moscow: Sov. encyclopediya, 1979. – C. 850-856. (In Russian).
- [9] **Vasiliev, S.N.** Ot klassicheskikh zadach regulirovaniya k intellektual'nomu upravleniyu [From classic tasks of regulation to intelligent control] / S.N. Vasiliev // Proceedings of the Academy of Sciences. Control theory and systems. 2001. No. 2. – P. 5-21. (In Russian).
- [10] **Zadeh, L.A.** Roles of soft computing and fuzzy logic in the conception, design and deployment of information/intelligent systems / L.A. Zadeh // Computational intelligence: soft computing and fuzzy-neuro integration with applications. - Springer Berlin Heidelberg, 1998. – P. 1-9.
- [11] **Valkman, U.R.** O strukture issledovaniy v oblasti iskusstvennogo intellekta [About structure of research in the scope of artificial intelligence] // Proc. of X Int. scientific-technical conf. “System analysis and information technologies”, Kiev, NTUU “KPI”, 2008. – P. 25. – [http://lib.znate.ru/pars\\_docs/refs/81/80229/80229.pdf](http://lib.znate.ru/pars_docs/refs/81/80229/80229.pdf). (In Russian).
- [12] **Finn, V.K.** Iskusstvennyi intellekt: ideynaya baza i osnovnoy produkt [Artificial intelligence: ideological foundation and main product] // Proc. of IX National conf. with international participation “Artificial intelligence-2004”. Vol. 1, 2004. – P. 11–20. (In Russian).
- [13] **Lyubich, A.A.** O vybore kriteriev ocenki intellektual'nosti informacionnoy sistemy [About selection of criteria for evaluation of information system intelligence] / A.A. Lyubich, V.L. Pleskach, J.V. Rogushina // USiM. 2005. No. 1. – P. 3-7. (In Russian).
- [14] **Gavrilova, T.A.** Bazy znaniy intellektual'nykh sistem [Knowledge bases of intelligent systems] / T.A. Gavrilova, V.F. Horoshevsky. – St. Petersburg: Piter, 2001. – 382 p. (In Russian).
- [15] **Kuzmenko, G.E.** Pragmatychny pidhid do ocinki rivnyu intelektu intelektualizovanykh sistem [Pragmatic approach to estimation of intelligence level on intelligent systems] / G.E. Kuzmenko, V.A. Litvinov // Mathematical machines and systems. 2003, No. 1. – P. 3-9. (In Ukrainian).
- [16] **Gruber, T.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing / T. Gruber // International Journal of Human-Computer Studies. 1995. Vol. 43. Issues 5-6. – P. 907-928.
- [17] **Uschold, M.** Ontologies: Principles, Methods and Applications / M. Uschold, M. Grüninger // Knowledge Engineering Review. 1996. 11(2). – P. 93-155.
- [18] **Smirnov, S.V.** Ontologii kak smyslovye modeli [Ontologies as semantic models] / S.V. Smirnov // Ontologija proektirovaniya. 2013. No. 2(8). - P. 12–19. (In Russian).
- [19] **Gilbert, N.** Platforms and methods for agent-based modeling / N. Gilbert, S. Banks // Proc. of the National Academy of Sciences, 99 (suppl 3), 2002. – P. 7197-7198.
- [20] **Rogushina, J.V.** Razrabotka ontologicheskoy modeli potrebnosti pol'zovat'lya pri semanticheskoy poiske [Design of ontological model of user's demand by semantic search] / J.V. Rogushina // Ontologija proektirovaniya. 2014. №2(12). – P. 61-82. (In Russian).

- [21] **Rogushina, J.** Ontology-based competency analyses in new research domains / J. Rogushina, A. Gladun // Journal of Computing and Information Technology. 2012. Vol. 20, No. 4. – P. 277-293.
- [22] ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval / Geneva: International Organization for Standards, 2011.
- [23] **Middleton, S.** Ontology-Based Recommender Systems / S. Middleton, D. De Roure, N. Shadbolt // Handbook on Ontologies. Eds.: S. Staab, R. Studer. - Springer, 2009. – P. 779-796.
- [24] **Harzallah, B.** Knowledge management for competence management / B. Harzallah // Proc. of Int. Conf. on Knowledge Management and Knowledge Technologies I-KNOW'05, 2005. – P. 449-460.
- [25] An organization ontology, W3C Working Draft, 2012.- <http://www.w3.org/TR/2012/WD-vocab-org-20120405/>
- [26] **Borgest, N.M.** Ontologii: sovremennoe sostoyanie, kratkiy obzor [Ontologies: current state, brief review] / N.M. Borgest, M.D. Korovin // Ontologija proektirovaniya. 2013. №2(8). – P. 49-55. (In Russian).
- 

### Сведения об авторе



**Рогущина Юлия Витальевна** (1967 г.р.) окончила Киевский государственный университет им. Т.Г. Шевченко в 1989 г. Степень кандидата физико-математических наук получила в Институте кибернетики им. В.М. Глушкова, г. Киев, в 1995 г. В настоящее время работает старшим научным сотрудником в Институте программных систем Национальной академии наук Украины. Исследовательские интересы включают разработку методов семантического поиска и рекомендации; онтологический анализ; исследование поведения интеллектуальных программных агентов, разработку алгоритмов индуктивного извлечения знаний, применение технологий Semantic Web. Автор более 150 публикаций в научных журналах и на конференциях, в том числе монографии «Знание-ориентированные средства поддержки семантического поиска в Web» и нескольких учебников.

**Julia Vitalijevna Rogushina** (b. 1967). She received the M. Sc. from Kyiv Taras Shevchenko State University in 1989. Her PhD degree in Computer Science she received in Glushkov's Institute of Cybernetics, Kyiv, in 1995. She is a senior researcher at the Institute of Software Systems, National Academy of Sciences of Ukraine. Her research interests include the development and application of intelligent information systems; theory of software agents behavior, inductive knowledge acquisition, intelligent information retrieval, ontological analysis, Semantic Web technologies. She has published more than 150 publications in scientific journals and conferences, monograph “Knowledge-oriented means of the Web semantic search” and several textbooks.