

ПРИМЕНЕНИЕ ОНТОЛОГИЙ ДЛЯ ПОСТРОЕНИЯ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА К СИСТЕМАМ WEB-ПОИСКА

Приводится обзор основных подходов организации пользовательского интерфейса при Web-поиске. Анализируется применение лексических синонимов для выявления семантических отношений между концепциями, выражаемыми на естественном языке. Предложен метод повышения эффективности использования существующих текстовых поисковых Web-систем путем представления пользовательского запроса в виде сети термов и интерактивного компонента для модификации запросов, использующего онтологию WordNet для установления семантических связей терминов пользовательского запроса.

Введение

Основная проблема, мешающая эффективному поиску текстовой информации, заключается в принципиальном отличии представления пользовательского поискового запроса от способа его интерпретации [9]. Человек, составляя запрос, оперирует семантическими единицами, в то время как для поисковой машины термины запроса являются синтаксическими единицами. С одной стороны, качество результатов Web-поиска зависит от алгоритмов, реализуемых поисковыми машинами (при этом конкретное содержание этих алгоритмов зачастую недоступно для внешнего анализа). С другой стороны, очевидно, что на качество получаемых результатов влияние оказывает и качество самих поисковых запросов. С этой точки зрения можно отметить сосуществование двух принципиально разных подходов к повышению эффективности поиска:

1. Разработка поисковой системы, ориентированной на семантический поиск.
2. Разработка интерактивных средств, помогающих пользователю составить эффективный запрос.

Эти два подхода могут быть совмещены посредством использования такого формального языка запросов, в котором смысл каждой синтаксической конструкции однозначно определен. При этом такой язык запросов должен быть максимально приближен к естественному языку. Самой поисковой системе в этом случае по-прежнему не требуется производить семантический анализ, а пользователь обходится без вспомогательных интерактивных средств. На таком гибридном подходе, в сущности, и основываются технологии Semantic Web.

И в том, и в другом случае разработчики большое внимание уделяют реализации интерфейса пользователя. Обзор поисковых систем общего назначения, а также специализированных поисковых систем, показывает, что представление результатов поиска и наличие удобных средств взаимодействия с пользователем (а не только текстовых полей для ввода запроса) также влияют и на популярность поисковой системы, и на то, насколько выдаваемые результаты соответствуют ожиданиям пользователей [25]. В дополнение к стандартным средствам текстовой строки, необходимо упомянуть следующие возможности, поддерживаемые современными системами Web-поиска:

- разрешение омонимии и основанные на этом средства выбора подходящего смысла термина;
- использование весовых коэффициентов для терминов пользовательского запроса;
- лингвистические текстовые запросы (например, уравнения химических реакций, математические выражения, фрагменты исходного кода компьютерной программы);
- интерактивные сети термов, составляющих запрос, и термов, извлекаемых из языковой онтологии, используемые для визуализации процесса расширения пользовательского запроса;

- облака тегов, другие подходы, основанные на теговой идентификации;
- средства ввода нетекстуальных запросов (например, музыкальных фрагментов, изображений и видео).

В данной работе содержится обзор подходов к конструированию сайтов, предоставляющих поисковые сервисы и совмещающих реализацию семантического поиска и интерфейсных средств, нацеленных на выявление семантики пользовательских запросов.

Поисковые системы и семантизация процесса поиска

Классические поисковые системы (такие как Google, Yandex, MSN, Infoseek и др.) ориентированы на использование текстового запроса и представление результатов в виде упорядоченного множества ссылок. Решение задач интерпретации запроса и оценки результатов поиска образует обширное пространство для исследований, в следующих областях представления и обработки знаний:

- анализ получаемых документов с точки зрения их релевантности запросу и взаимной семантической релевантности [12, 19, 27];
- автоматизации процесса создания и модификации запросов [25];
- повышение качества генерируемого краткого содержания извлекаемых документов [24];
- оценка релевантности документов на основе информации, извлекаемой из онтологий [7, 26].

Используя стандартный текстовый интерфейс, система поиска может обеспечивать интерпретацию запроса и представление извлекаемых данных с учетом их специфики. Примером может служить поисковый сервис на *nigma.ru*, который обеспечивает обработку запросов, представляющих собой формулы химических реакций и математические выражения и уравнения (для них в визуальной форме выводится объяснение решения). Другой пример – Wolfram Alpha, подключаемый компонент к браузеру, использующий подход, который разработчики называют вычислимостью знаний (knowledge computability). В ходе анализа запроса данный компонент пытается определить область знания, а затем извлечь информацию, имеющую отношение именно к этой области. Например, обнаружив в строке запроса Google название города, Wolfram Alpha выводит карту соответствующей местности, информацию о площади и населении, другие метаданные.

Наряду с подобными системами активно развиваются подходы, основанные на визуализации процесса поиска как для обеспечения интерактивного ввода запроса, так и для представления результатов поиска.

В отдельных поисковых системах используется модель облаков тегов для улучшения интерактивности процесса поиска. Появившись первоначально на сайтах, предоставляющих сервисы размещения и просмотра изображений (таких как Flickr на *www.flickr.com* или «Yandex.Фотки» на *foto.yandex.ru*), подход с облаками тегов стал применяться и в поисковых системах общего назначения. Примерами таких реализаций могут служить компоненты дополнений CloudLet (*www.getcloudlet.com*) и DeeperWeb (*www.deeperweb.com*) для браузера Firefox. Технологии использования подобной категориальной идентификации в настоящее время являются одним из основных методов улучшения пользовательского интерфейса в поисковых системах и системах, связанных с извлечением знаний [10, 11].

Многие разработки ориентированы на визуальный интерфейс взаимодействия с пользователем. Java-апплет TouchGraph Google Browser отображает зависимости Web-сайта в форме редактируемого графа с интерактивными гиперссылками и возможностями просмотра краткой информации о Web-сайте (рис. 1). Продукт ThinkMap Visual Thesaurus (*www.visualthesaurus.com/howitworks*) использует модель графа для визуализации отношений между терминами: поддерживаются обобщение, специализация, синонимы, антонимы, производные формы и т. д. Непосредственно не являясь элементом поисковой системы, данный компонент может использоваться в процессе модификации запроса для улучшения его качества и семантического содержания (рис. 2). Визуальные интерфейсы, реализуемые в данных инструментах, основаны соответственно на графе и семантической сети отношений и представляют интерес с точки зрения использования для разработки элементов интерфейса поисковой системы.

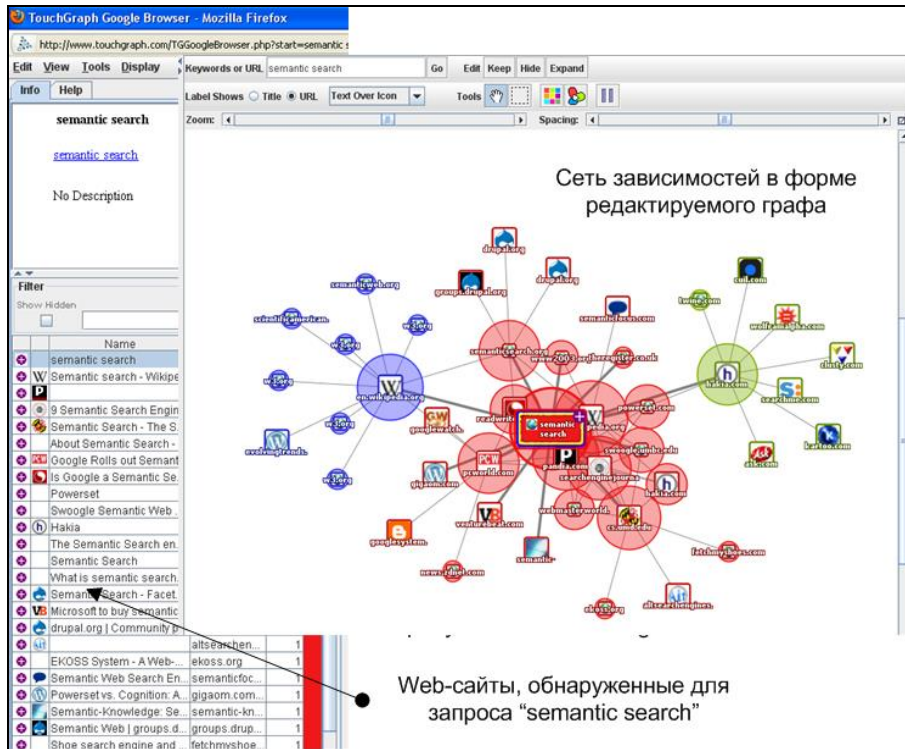


Рис. 1. Визуальный интерфейс TouchGraph

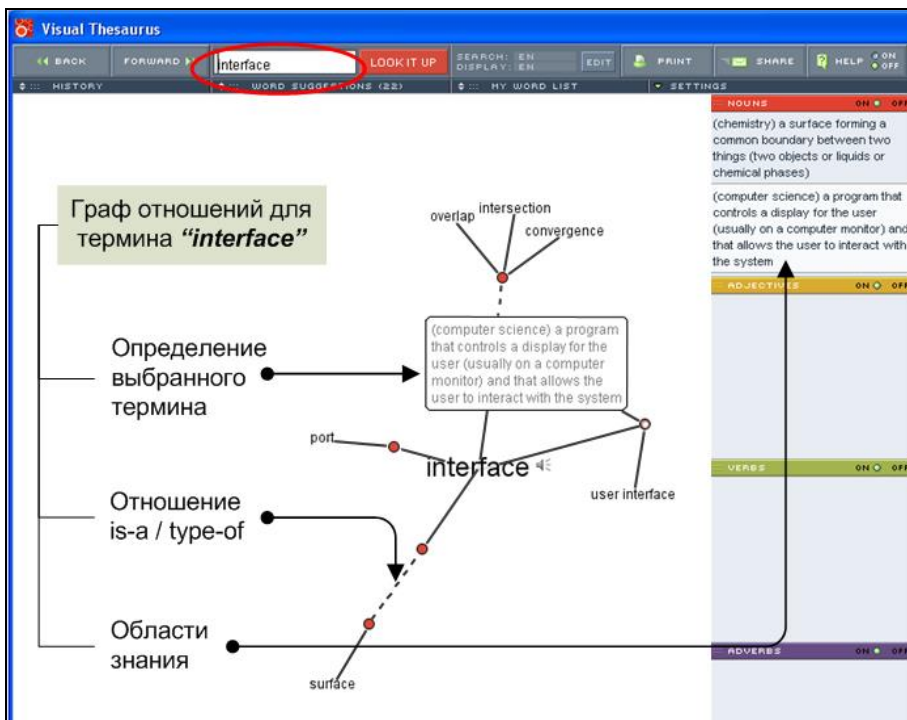


Рис. 2. Интерфейс системы ThinkMap Visual Thesaurus

Интерактивность интерфейса может достигаться не только за счет использования визуальных представлений, но с помощью семантической модификации запроса. Например, поисковый сервис *digger.com* интерпретирует пользовательский запрос и предоставляет возможность уточнить смысловую область поиска. При этом обеспечивается упрощенная для пользователя модель весовых коэффициентов предпочтения выбранных смыслов. Проведенные эксперименты показывают, что смыслы, вычисляемые *digger.com*, имеют сильное сходство с результатами, получаемыми с использованием онтологии WordNet.

Процесс семантического доопределения задачи поиска может быть реализован с использованием технологий семантической паутины (Semantic Web), основная идея которой и заключается в обеспечении ориентированных на машинное восприятие данных для существующих документов с целью улучшения процессов извлечения и формализации знаний [5, 6]. В настоящее время применение технологий семантической паутины предполагает не только автоматизацию процессов доступа к Web-ресурсам, но ориентированность на обеспечение интеграции и взаимодействия бизнес-процессов и информационных систем. Достижение такой интеграции возможно за счет создания глобальной инфраструктуры обмена документами и данными на основе управления метаданными, определяемыми для имеющихся Web-ресурсов. При этом онтологии являются формальной основой и ключевой технологией для развития семантической паутины [1, 8].

Поисковый агент *www.semanticwebsearch.com* использует принцип «раз мы не умеем понимать то, что от нас хочет пользователь, давайте заставим пользователя писать то, что мы понимаем». Пользователь имеет возможность определить тип ресурса, значение определенных свойств, связанных с типами метаданных, поддерживаемых семантической паутиной.

Успешность применения концепций и технологий семантической паутины в существенной степени зависит от качества метаданных, определяемых для того или иного ресурса. До настоящего времени число сайтов, поддерживающих такие технологии как RDF, OWL, SPARQL, FOAF и др. все еще ограничено, однако активные исследования в этой области, а также стандартизация синтаксической и семантической разметки документов консорциумом W3C, обеспечивают рост популярности технологий семантической паутины среди разработчиков. Следует отметить, что внедрение семантической информации в HTML-страницы имеет ряд ограничений. Очевидно, что поддержка метаданных для представления биометрической, географической, календарной, контактной и др. информации, относящейся к так называемой легковесной семантике (light-weight semantics), весьма полезна, хотя сами метаданные (во многих случаях создаваемые людьми) не всегда совершенны. В общественном сознании семантическая паутина еще не заняла ведущее место: статистика поисковых запросов Google показывает, что запросы, относящиеся к технологиям Semantic Web, встречаются реже чем, например, запросы о языках Prolog и Fortran [29]. Таким образом, вопросы о точности и достоверности метаданных семантической паутины, равно как и проблемы, связанные с распространением персональных данных, остаются открытыми, тем не менее, по меткому замечанию Щербака, «низкоуровневый семантический Веб безусловно лучше, чем никакого».

Из описания большинства современных поисковых систем, упоминаемых в данной работе, следует что все они относятся к классу семантически-ориентированных систем. Среди возможностей, реализация которых позволяет авторам системы поиска классифицировать ее как семантическую систему, упоминаются, в частности, следующие:

- учет морфологических вариаций терминов пользовательского запроса;
- учет синонимов, извлекаемых из языковых и предметных словарей и энциклопедий;
- расширение пользовательских запросов с добавлением терминов, извлекаемых из онтологий (типа онтологии WordNet [15]);
- поиск концепций, связанных с терминами пользовательского запроса, как на основе использования онтологической информации, так и на основе информации,

извлекаемой из баз знаний, к которым в настоящее время вполне можно отнести обширные ресурсы Wikipedia [14, 16, 17];

- учет аспектов абсолютного и относительного времени («в прошлом году», «на следующей неделе» т.п.) при интерпретации пользовательского запроса.

Обеспечение семантической ориентированности процесса поиска не обязательно предполагает использование полноценного семантического анализа. Косвенным образом в выявлении смысловой направленности пользовательского запроса может помочь сбор и применение некоторой статистической информации. Например, интерфейсы большинства современных поисковых машин обычно предлагают пользователю список подсказок, ускоряющих набор текстовой строки запроса. Чтобы сэкономить время, пользователь иногда выбирают готовую подсказку, даже если она и не совсем точно соответствует его первоначальным намерениям. Сбор информации о страницах, посещенных пользователем, выбравшим конкретную подсказку (равно, как и проигнорированных им), может в дальнейшем стать одним из факторов принятия решений при определении порядка ссылок, выдаваемых другим пользователям, выбравшим ту же подсказку. Авторам неизвестно, руководствуются ли разработчики Google или Yandex подобными рассуждениями, но такой подход представляется вполне логичным способом частичной семантизации обработки пользовательского запроса на основе статистической информации.

Лексические синонимы в лингвистике и системах поиска

В когнитивной лингвистике и ее прикладных областях изучение синонимических рядов является одним из важнейших аспектов семантически ориентированной обработки текстов, в том числе, в следующих направлениях:

- Выявление лингвистических концепций, распознавание особенностей речевого и письменного стиля, изучение этимологических зависимостей [21].
- Исследование отношений между терминами в процессе изучения языка и его структуры [15].
- Исследование синонимических рядов как метод познания языка и объективизации мира [4].
- Исследование жанрово-стилистических функций и их проявления в языковом образе человека как область антропоцентрического подхода в когнитивной лингвистике [3].
- Семантический анализ применительно к задачам анализа Web-документов, включая проблему автоматической генерации релевантного краткого содержания обнаруженных документов [19, 24].

Синсет в онтологии определяется как синонимический ряд слов, обозначающих один и тот же концепт в заданном контексте. Например, для описания таксономических связей между существительными используются следующие основные типы отношений:

- Термин А является гиперонимом (hypernym) термина В, если А выражает более общую концепцию по отношению к В (в объектно-ориентированном анализе в этом случае речь идет об отношении обобщения), соответственно В является гипонимом (hyponym) термина А, то есть В выражает более специфическую концепцию (отношение специализации).
- Концепции, выражаемые двумя терминами А и В, находятся в отношении «часть-целое» (в объектно-ориентированном анализе и проектировании это отношение выражается агрегированием или композицией), соответственно А является голономом (holonym) В, а В – меронимом (meronym) А.
- Термины А и В являются родственными, если имеют связи с другими терминами, выступающими в свою очередь в качестве общих гиперонимов, гипонимов и т. д. для А и В.
- Термин А является свойством или атрибутом для концепции, выражаемой термином В.

Отметим, что синсет не всегда позволяет выразить основные свойства, присущие конкретному типу отношения, наиболее полно. Так, отношение меронимии может подразумевать семантически существенно отличающиеся разновидности [18, 28]:

1. Функциональная меронимия, когда отношение части и целого имеет функциональный характер: колесо является частью велосипеда, при этом отделение этой части позволяет колесу оставаться колесом, в то время как велосипед перестает выполнять свою основную функцию (и в определенной степени перестает быть велосипедом).
2. Квалификационная меронимия, отражающая определенную роль, исполняемую частью по отношению к целому: каждый супруг является неотъемлемой частью семейной пары, при отделении одной из частей смысл целого утрачивается.
3. Пространственно-временная меронимия, например, дерево является частью леса, при этом при отделении части от целого, и часть, и целое продолжают существовать (дерево остается деревом, а лес — лесом).
4. Меронимия, выражающая результат деления на части, при этом часть может наследовать функции целого: кусок пирога является частью пирога, при этом кусок пирога — это тоже пирог.
5. Меронимия, выражающая материальную структуру объекта. При этом отделение части от целого может быть невозможным без разрушения целого и, наоборот, в процессе формирования целого смысл отдельных частей утрачивается: когда части мозаики собираются в единую картину, эти части, фактически, перестают существовать [20].

Каждый тип отношений порождает определенный тип онтологии. Несмотря на то, что в общем случае смешивать различные типы онтологий не рекомендуется [2], в конкретных реализациях подобное смешение встречается. Например, онтология WordNet фактически является объединением нескольких типов онтологий (в том числе, таксономии, партономии и др.).

Идея использования семантических связей, основанных на синсетах, обусловлена двумя обстоятельствами. Во-первых, определение соответствующего фрагмента таксономии позволяет более точно установить контекст употребления термина в поисковом запросе. Во-вторых, обнаруженные связи позволяют сформировать подсказки, позволяющие пользователю улучшить свой запрос, используя более общий, или, наоборот, более специальный термин. Таким образом, смысл пользовательского запроса становится более понятным, в том числе, и самому пользователю, при этом детальный семантический анализ отношений, влияющих на выбор термина, пользователю, как правило, не требуется.

Интерактивный компонент модификации пользовательского запроса

Вспомогательные компоненты Web-поиска обычно реализуются или в форме расширения (browser plug-in), или в форме независимой Web-страницы, выполняющей роль интерфейса к поисковой машине. Первый вариант обеспечивает прозрачную интеграцию с возможностями браузера, второй предпочтительнее с точки зрения обеспечения совместимости с разными программами просмотра Web-страниц.

Многие пользователи рассматривают строку запроса как предложение на естественном языке, поэтому они не используют возможности языка запросов наиболее эффективно и правильно. Для повышения качества и более эффективного использования возможностей языка поисковых запросов предлагается использовать специализированный пользовательский интерфейс к существующим поисковым системам, позволяющий выполнять два типа преобразования исходного запроса: 1) преобразование формы (дать пользователю более гибкий способ ввода запроса, чем текстовая строка) и 2) преобразование семантики.

На рис. 3 представлен разработанный интерактивный интерфейс к поисковой системе Google. Компонент реализован в виде HTML страницы, разделенной на два фрейма. Верхний фрейм содержит JavaFX-апплет, выполняющий две основные функции: помощь в составлении запроса и формирование поисковой строки для поисковой машины (в нашем примере – Google). Результаты поиска отражаются в нижнем фрейме. Такой подход позволяет обеспечить независимость от используемого браузера (благодаря спецификации HTML), от операционной системы (поскольку код выполняется на виртуальной Java-машине), а также от физического размещения самого компонента (так

как апплет может располагаться локально или на удаленном сервере благодаря протоколу загрузки JNLP).

Поисковый запрос предлагается составлять в форме сети термов, которая состоит из AND-узлов (рис. 3, а) и OR-узлов (рис. 3, б). Вставка новых элементов поддерживается как с помощью мыши, так и с помощью клавиатуры. Для добавления узлов с помощью мыши используются элементы интерфейса, представленные на рис. 3, в, г (для AND- и OR- узлов соответственно), которые автоматически появляются для активного узла графа (того узла, который находится в фокусе). При вводе с клавиатуры пробел интерпретируется как вставка нового AND-узла справа, что эквивалентно как *методу* ввода запроса в обычную поисковую строку, так и *семантике* введенного запроса. Чтобы удалить ненужный узел достаточно удалить из него весь текст или просто щелкнуть по нему правой кнопкой мыши. Все эти элементы интерфейса имеют отношение только к преобразованию формы запроса.

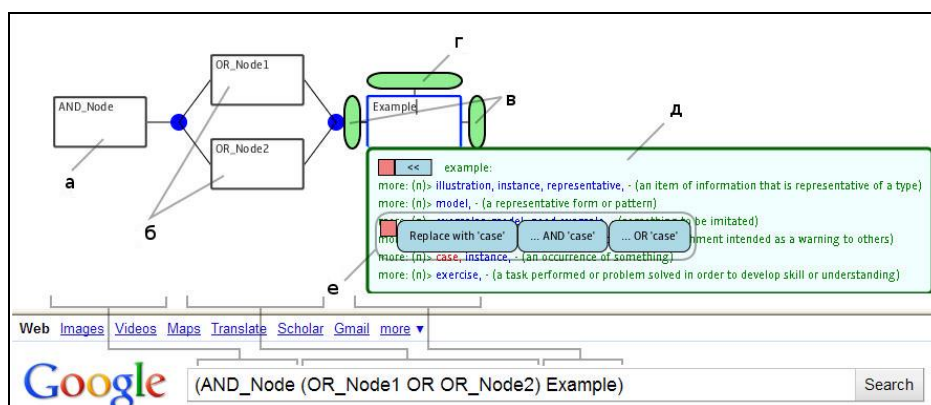


Рис. 3. Интерфейс вспомогательного компонента Web-поиска

Семантическое преобразование запроса выполняет сам пользователь с использованием интерактивного помощника (рис. 3, д), который предоставляет доступ к онтологии WordNet (sense navigator). Логически вся область помощника состоит из двух частей: заголовка и рабочей области. Заголовок содержит управляющие элементы и само исследуемое слово. Рабочая область (sense selector) содержит описание исследуемого слова на основе онтологии WordNet. Каждая строка соответствует одному смыслу и содержит следующие элементы:

- «more (<часть речи>)» – интерактивный элемент, загружает в рабочую область понятия, связанные с конкретным смыслом исследуемого слова (напр., гиперонимы, гипонимы, голонимы и т. д. для существительных).
- Синонимический ряд исследуемого слова – интерактивный элемент. Щелчок левой кнопкой мыши по синониму загружает в рабочую область описание выбранного синонима. Щелчок правой кнопкой мыши вызывает меню (рис. 3, е), позволяющее заменить слово в активном узле, добавить узел OR с выбранным синонимом или добавить узел AND с выбранным синонимом после активного узла.
- Описание значения – элемент, содержащий пояснения к смыслу слова.

Когда пользователь закончил ввод поискового запроса, сеть термов преобразуется в запрос к поисковой машине, составленный с использованием логических операторов. В зависимости от результатов, процесс семантического изменения запроса может быть повторен.

Основная задача помощника заключается не столько в том, чтобы быстро менять слова в запросе, сколько в том, чтобы дать возможность пользователю объективно оценить свой запрос. Результат поиска может быть неудовлетворительным не потому, что в Интернете мало страниц на интересующую тему, а потому, что первыми в списке результатов оказались страницы, которые содержат все слова, но, возможно, в других смыслах. Это должно подтолкнуть пользователя заменить многозначное слово синонимом,

либо вообще исключить его из запроса, чтобы оно не мешало при оценке релевантности результатов поисковой машиной (как это и демонстрируется в следующем разделе).

Пример использования

Рассмотрим процесс взаимодействия пользователя и разработанного компонента на примере. Ввиду того, что текущая реализация использует английскую версию WordNet, мы приводим пример поиска информации на английском языке.

В ходе работы над системой поиска музыкальной информации [13], нам потребовалось уточнить корректность использования латинской буквенной нотации для записи нот. Несмотря на определенное музыкальное образование (см., например, [22, 23]), нам не было известно, что музыканты используют для этого термин «pitch notation», причем буквенная нотация существует в нескольких формах, например нотация Гельмгольца, научная нотация, MIDI-нотация. Мы начали поиск с предположения, что разумными запросами могут быть «letter music system» или «letter music notation». Представление результирующего запроса иллюстрирует рис. 4.

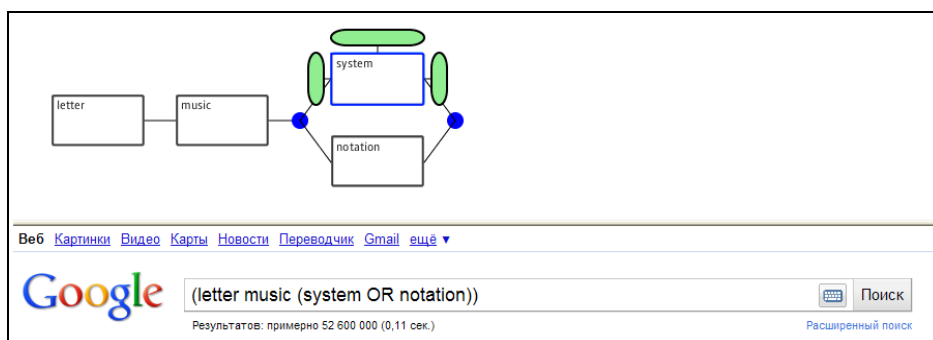


Рис. 4. Начало поиска

Проанализировав краткое содержание полученных страниц, мы пришли к выводу, что они не содержат требуемой информации. Используя панель интерактивного помощника, мы получили набор смыслов термина «music», извлеченных из WordNet (см. рис. 5).

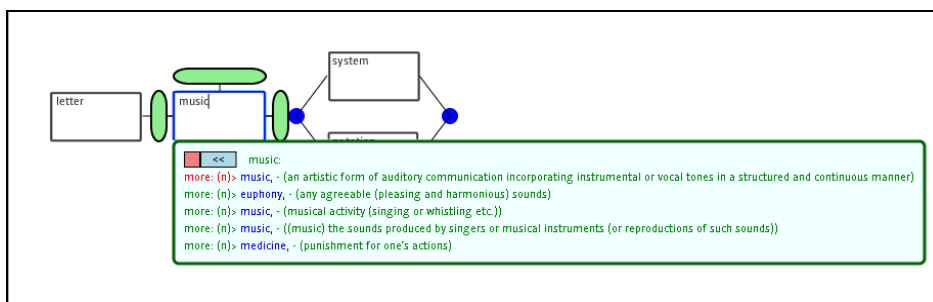


Рис. 5. Извлечение смыслов термина из онтологии

После выбора первого смысла мы получаем новое состояние рабочей области со списком ассоциаций, связанных с выбранным смыслом (рис. 6). В данном синонимическом ряду наиболее перспективным представляется термин «tune», поскольку его определение (succession of notes forming a distinctive sequence) явно соответствует нашей области поиска. Для замены узла «music» на узел «tune» используется всплывающая панель инструментов.

Анализируя результаты, выданные Google (рис. 7), мы обнаруживаем, что направление поиска правильное, но полученные сайты содержат только тривиальные примеры. Изучая термин «tune», выбираем тот смысл, который представляется наиболее

PREPRINT

подходящим для достижения нашей цели (the property of producing accurately a note of a given pitch). Открытие детализированного описания (подсказка «more:») позволяет обнаружить что «tune» в указанном смысле имеет синоним «pitch» (рис. 7). Используя всплывающую панель инструментов снова, заменяем «tune» на «pitch».

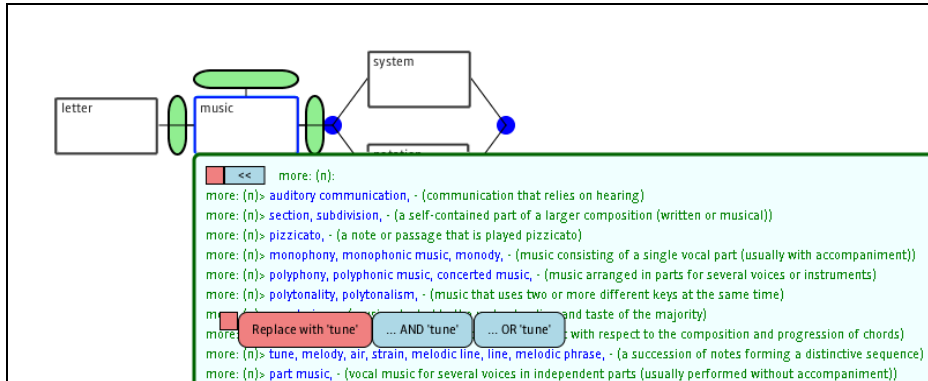


Рис. 6. Список ассоциаций для выбранного смысла

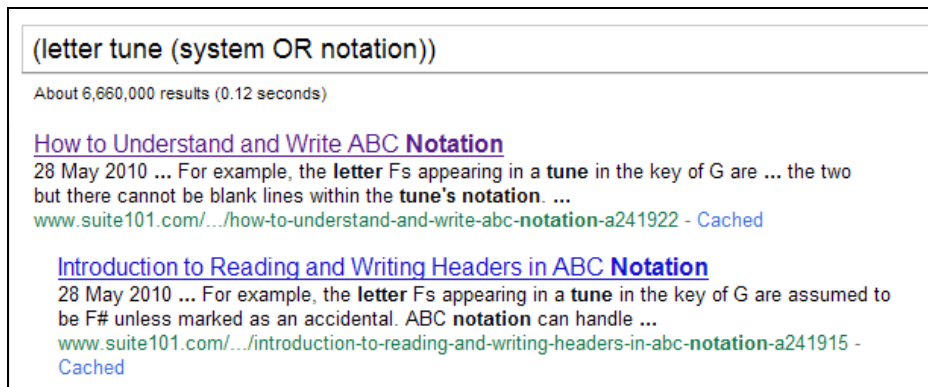


Рис. 7. Результаты, выданные поисковой системой после модификации запроса

Полученная сеть термов изображена на рис. 8. Фрагмент списка результатов, выданного поисковой машиной, представлен на рис. 9. Отметим, что первые две ссылки ведут на сайты, содержащие исчерпывающее описание упомянутых в начале раздела музыкальных нотаций.

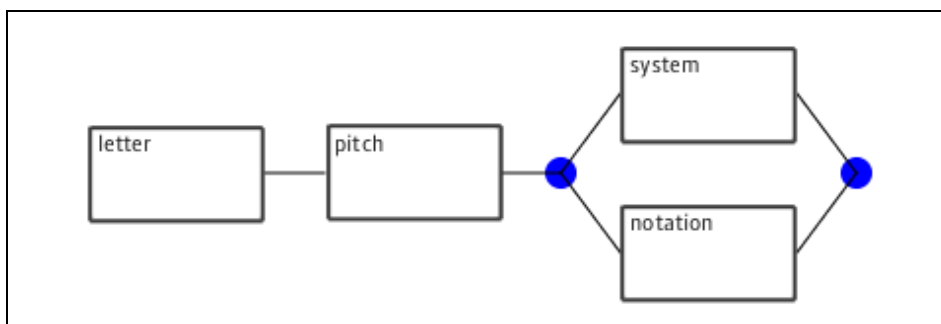


Рис. 8. Модифицированная сеть термов

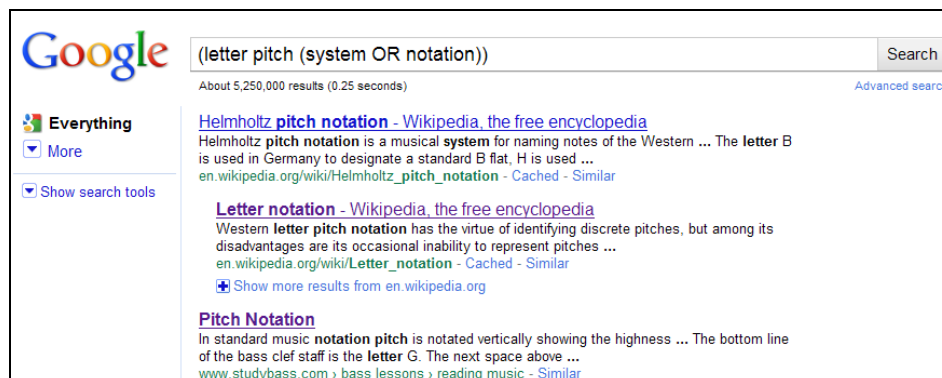


Рис. 9. Ссылки, выданные поисковой машиной

Заключение

Качество веб-поиска определяется не только качеством алгоритмов поисковой системы, но и качеством самого поискового запроса, который, в свою очередь, зависит, в том числе, и от возможностей пользовательского интерфейса. Подход, предполагающий использование информации, извлекаемой из онтологий, для модификации пользовательского запроса, известен давно. Новизна предложенного решения заключается в первую очередь в том, что мы предпринята попытка учесть субъективный характер процесса поиска. Предлагаемая сеть термов позволяет в интуитивно-понятной форме эффективно использовать возможности поисковых систем, интерпретирующих логические выражения, не требуя от пользователя знания языка запросов конкретной поисковой системы. Не менее важно, что предлагаемый помощник позволяет пользователю (с помощью онтологии WordNet) адекватно оценить и изменить сам запрос. Таким образом, ключевым аспектом здесь является не столько обеспечение роста скорости поиска, а реализация возможности включения пользователя в активный анализ предметной области поиска.

Языки поисковых запросов обычно поддерживают не только построение логических выражений с операциями И, ИЛИ, НЕ, но и другие возможности, например, метасимвольные аргументы, идентификацию категории термина, идентификацию сайта и т. п. Развитие имеющейся реализации с учетом данных возможностей представляет интерес в качестве дальнейшей разработки.

Открытым остается вопрос об оценке предложенных средств конкретными пользователями. Наряду с экспертными оценками вызывает интерес сбор и анализ объективной информации о процессе поиска. Фактически, в этом случае мы приходим к постановке задачи в области интеллектуального анализа данных (data mining). На основе выбранных характеристик, которые можно получить автоматически (например, о длине пользовательского запроса, числе открытых документов и ссылок, времени поиска и т. д.), может быть поставлена задача разработки метрик, позволяющих оценить, как изменяется процесс поиска с использованием тех или иных интерфейсных средств.

ЛИТЕРАТУРА

1. Андон Ф.И., Гришанова И.Ю., Резниченко В.А. Semantic Web как новая модель информационного пространства Интернет. <http://shcherbak.net/semantic-web-kak-novaya-model-informacionnogo-prostranstva-internet/> (2010).
2. Инновационное развитие: экономика, интеллектуальные ресурсы, управление знаниями / Под ред. Б.З. Мильзнера. - М.: ИНФРА-М, 2010.- 624 с., - (Научная мысль): Гл. 25. Инженерия знаний, с.480-500.
3. Покровская О.В. Лексические синонимы, репрезентирующие языковой образ человека, с точки зрения их жанрово-стилистических функций и участия в жанрообразовании. Вестн. Томского гос. Ун-та, 2008, № 307, с.20-23.
4. Фришберг И. Синонимическая аттракция в свете когнитивного подхода. Труды межд. конф. «Язык. Культура. Коммуникация», Челябинск, 2004, с.147-150.

5. Щербак С.С. Интеллектуализация обработки информации на основе технологий Semantic Web. <http://shcherbak.net/intellect/> (2010).
6. Berners-Lee T., Hendler J., Lassila O. "The Semantic Web." *Scientific American*, May 2001.
7. Budanitsky A., Hirst G. "Evaluating WordNet-based measures of lexical semantic relatedness." *Computational Linguistics*, Vol.32, Issue 1 (March, 2006), 2006, 13-47.
8. Cardoso J., Litras M., Hepp M. (Eds) "Semantic Web: Real World Applications from Industry." Springer Science, 2008.
9. Guha R., McCool R., Miller E. "Semantic search." In *Proceedings of WWW 2003*, Budapest, Hungary, May 2003.
10. Halvey M., Kean M. "An assessment of tag presentation techniques." In *Proceedings of WWW 2007*, Banff, Alberta, Canada, May 8-12, 2007.
11. Hassan-Montero Yu., Herrero-Solana V. "Improving tag clouds as visual information retrieval interfaces." In *Proceedings of the I International Conference on Multidisciplinary Information Sciences and Technologies (InSciT2006)*, Merida, Spain, October 25-28, 2006.
12. Klyuev V., Oleshchuk V. "Semantic retrieval of text documents." In *Proceedings of 7th IEEE Conference on Computer and Information Technology (CIT2007)*, 2007, 199-183.
13. Kuznetsov A., Pyshkin E. "Searching for music: from melodies in mind to the resources on the web." In *Proceedings of the 13th International Conference on Humans and Computers (HC2010)*, University of Aizu, Aizu-Wakamatsu, Japan, Dec. 8-10, 2010, 152-158.
14. Medelyan O., Milne D., Legg C., Witten I. "Mining meaning from Wikipedia." *International Journal of Human-Computer Studies*, Vol.65, Issue No.9 (September 2009), 716-764.
15. Miller, G.A. "WordNet: a lexical database for English." *Communications of the ACM*, Vol. 38, No. 11, 39-41.
16. Milne D., Witten I. "An effective low-cost measure of semantic relatedness obtained from wikipedia links." In *Wikipedia and AI workshop at the AAAI-08 conference (WikiAI08)*, Chicago, USA.
17. Milne D.N., Witten I.H., Nichols D.M. "A knowledge-based search engine powered by wikipedia." In *Proceedings of the 16th ACM Conference on Information and Knowledge Management (Lisbon, Portugal, 2007)*, 445-454.
18. Mizoguchi R. "Advanced Course on Ontological Engineering." <http://www.ci.sanken.osaka-u.ac.jp/pub/miz/part3v3.pdf> (2010).
19. Oleshchuk V., Klyuev V. "Context-Aware Summary Generation for Web-Pages." In *Proceedings of IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Rende (Cosenza), Italy. September 21-23, 2009*.
20. Perec G. "La Vie Mode d'emploi." Hachette, 1978. Имеется русский перевод: Перек Ж. Жизнь способ употребления / Пер. с франц. В. Кислова.- СПб.: Изд-во Ивана Лимбаха, 2009.- 624 с.
21. Pimenov E.A. "The analysis of the concepts by synonyms." In *Proceedings of 7th International Conference on Cognitive Modeling in Linguistics, Moscow-Varna, 2005*.
22. Pyshkin E. "Night in Ylivieska." CD Edition, Central Ostrobothnia Polytechnic, 2004. <http://aivt.ftk.spbstu.ru/info/staff/pychkin/> (2010).
23. Pychkine E. "Understanding the Object Model: Prerequisites for the Crash Course of Software Engineering." Central Ostrobothnia Polytechnic, 2003.
24. Pyshkin E., Klyuev V. "On document evaluation for better context-aware summary generation." In *Proceedings of 2nd International Symposium on Aware Computing (ISAC2010)*, Nov. 1-4, National Cheng-Kung University, Tainan, Taiwan, 2010, CD Edition.
25. Pyshkin E., Kuznetsov A. "An Approach for web search user interface based on the cognitive synonyms concept." In *Proceedings of 3rd International Conference on Human - Centric Computing*, Cebu, Philippines, 2010, CD Edition.
26. Rinaldi A.M. "An ontology-driven approach for semantic information retrieval on the web." *ACM Transactions on Internet Technology*, Vol.9, Issue 3 (July 2009), 2009, Article No.10.
27. Satake S., Ishikawa S., Vazhenin D., Klyuev V. "A semantic-based summarization tool for improving the usability of general purpose search engine." *ICPC Joint Conference on Pervasive Computing*, Taiwan, Dec. 2009, CD Edition.
28. Winston, M.E., Chaffin R., Herrmann D. "A taxonomy of part-whole relations," *Cognitive Science*, 11, 417-444, 1987.
29. Zambonini D. "The 7 (f)laws of the Semantic Web." In *O'Reilly Blogs*, 2006. http://www.oreillynet.com/xml/blog/2006/06/the_7_flaws_of_the_semantic_web.html (2010).

PREPRINT

Andrey Kuztetsov, Evgeny Pyshkin

Using Ontologies for Constructing the Web Search User Interface.

Approaches used for Web search user interface are reviewed. Recognition of language concepts relationships discovered by their lexical synonyms in its application to the web search is analyzed. The implementation of the approach to assist the query based search with using query token network and WordNet ontology is introduced.