О НОВОМ ПОДХОДЕ К СЕМАНТИЧЕСКОМУ ПРЕОБРАЗОВАНИЮ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ЗАПРОСОВ ПОИСКОВЫХ СИСТЕМ

А.В. Кириллов,
аспирант кафедры Инноваций и бизнеса в сфере информационных технологий факультета бизнес-информатики Национального исследовательского университета «Высшая школа экономики»,
e-mail: antonv.kirillov@gmail.com.

В.А. Фомичев,
доктор технических наук, профессор кафедры Инноваций и бизнеса в сфере информационных технологий факультета бизнес-информатики Национального исследовательского университета «Высшая школа экономики»,
e-mail: vfomichov@hse.ru.
Адрес: г. Москва, ул. Кирпичная, д. 33/5.

В статье рассматриваются существующие подходы к поиску информации, анализируются их основные недостатки. Предлагается новый подход к поиску информации, основанный на семантическом преобразовании поисковых запросов. В рамках предлагаемого подхода выделяются классы естественно-языковых запросов, предлагаются формальная модель проблемно-ориентированной системы первичных единиц концептуального уровня и метод построения семантического расширения поискового запроса, а также иллюстрируется применение метода.

Ключевые слова: семантический поиск, семантический анализ, теория К-представлений, естественно-языковые запросы, поисковые системы.

Введение

Несмотря на большое разнообразие подходов к поиску информации [1, 2, 3], фундаментальной проблемой при разработке поисковых систем является определение релевантности – ситуации, когда документ соответствует запросу. Для определения релевантности документов используется большое разнообразие методов, таких как VSM (векторно-пространственная модель), функции BM25 и BM25F (учитывающие различные весовые факторы слов в документе), функции Okapi, Ponte, алгоритм LCA и другие.

Однако системы, реализующие поиск по ключевым словам, сталкиваются с серьезными проблемами, связанными с качеством результатов поиска. Часто возникает ситуация, когда результаты поиска различны для двух запросов с одинаковым смыслом, но отличающихся грамматически.

Семантическая поисковая система в ее классическом виде оперирует с мета-данными, описывающими документы. Данное описание хранится в таких форматах, как RDF, RDFS, OWL и другие. Мета-информация позволяет получить семантическое описание содержимого документов. В настоящее время семантические поисковые системы используют большое разнообразие методов и технологий, большинство которых основано на анализе естественного языка. Подходы, применяемые в последние годы в семантическом поиске, весьма разнообразны: увеличение семантической релевантности посредством дополнительного синтаксического анализа и использования обнаруженных данных RDF [3], анализ поисковых запросов и документов на основе триплетов с использованием онтологии предметной области [4], использование автоматически сгенерированных онтологий при поиске [5], вопросно-ответные системы на основе семантических графов и анализа триплетов [6], извлечение семантических отношений естественного языка с помощью шаблонов грамматических зависимостей [7] и многие другие.

Хотя сегодня имеется много различных подходов к семантическому поиску в Интернете, остаются не полностью решенными следующие задачи: разработка естественно-языковых анализаторов поисковых запросов, определение типа вопроса, определение объекта интереса поискового запроса, определение предметной области либо областей, к которым может относиться поисковый запрос, определение принадлежности найденных документов той же предметной области, что и объект интереса поискового запроса, поиск синонимов объекта интереса поискового запроса и некоторые другие. Решение данных задач достигается предлагаемым подходом к семантически-ориентированному поиску.

Описание предлагаемого подхода

Предлагаемый подход к реализации семантически-ориентированного поиска базируется на следующих положениях: в силу того, что семантическое описание документов в большинстве случаев недоступно, алгоритмизация создания данного описания является весьма трудоемкой задачей, технологии семантического веба и гипертекстового поиска развиваются параллельно и независимо друг от друга, предлагается реализовать семантически-ориентированную поисковую систему, выполняющую преобразование поискового запроса, в зависимости от его типа, в форму, позволяющую синтаксической поисковой системе найти адекватное подмножество документов, семантически соответствующих ожиданиям пользователя. Данный подход оперирует с разными типами вопросов на естественном языке и позволяет создавать семантическое описание для каждого из них. В случае, если какой-то из введенных пользователем вопросов не может быть проанализирован, пользователь получит результаты работы системы поиска по ключевым словам.

Для успешного выполнения преобразования поискового запроса к расширенному виду необходимы следующие обязательные элементы:

1. Математическая модель проблемно-ориентированной системы первичных единиц концептуального уровня, используемых поисковой системой.
2. Математическая модель многообразия смысловых структур как ориентир для построения семантических представлений естественно-языковых поисковых запросов.
3. Математическая модель лингвистической базы знаний, позволяющая связать семантическое представление поискового запроса с его грамматическим и семантическим окружением.
4. Алгоритм преобразования поискового запроса в первичное семантическое представление с последующим анализом и расширением.
5. База знаний, отражающая взаимосвязи семантических единиц, а также включающая механизм определения отношений между ними.

Анализ научной литературы показал, что теория К-представлений, предложенная В.А. Фомичевым [1, 8], по ряду характеристик наиболее удобна для решения поставленной задачи создания семантического представления поискового запроса.

Типизация запросов на естественном языке

По результатам проведенных исследований были выявлены, проанализированы и сгруппированы вопросы на естественном языке, а также вы-

делены следующие группы вопросов и методы их обработки:

✦ общие (традиционные вопросы, не касающиеся специфики того или иного объекта интереса). В процессе преобразования запросов будут использованы синонимы, гипонимы и гиперонимы объекта интереса запроса, на этапе анализа возвращаемых документов будут использоваться антонимы объекта интереса

✦ аспектно-ориентированные (вопросы, касающиеся характеристик объекта интереса, либо его особенностей). При анализе определяется принадлежность объекта интереса к предметным областям, используются определения, синонимы, антонимы и т.д.

✦ вопросы, касающиеся сохранения или изменения состава того или иного множества; при анализе используется база знаний, содержащая некоторый набор сведений о тех или иных объектах и множествах объектов.

✦ вопросы достижения целей (вопросы, связанные с успехами и неудачами тех или иных интеллектуальных систем).

Формализация обработки перечисленных вопросов, помимо общих, ранее в доступной научной литературе не рассматривалась.

В рамках проводимого исследования была выделена группа наиболее актуальных и практических значимых вопросов для их детального анализа и разработки алгоритма семантического преобразования, а также алгоритма обработки данного вида вопросов. Таковыми являются вопросы аспектно-ориентированного типа. Введем формальное определение аспектно-ориентированных вопросов.

Аспектно-ориентированными вопросами будем называть вопросительные предложения, в которых запрашивается информация, касающаяся различных аспектов того или иного объекта или системы. Такими аспектами могут являться характеристики, условия существования или функционирования, назначение, структурная организация, функции, области применения, принадлежность к какому-либо классу, принципиальные отличия, особенности и возможности различных объектов и систем.

Для представления различных аспектов необходимо ввести реляционные символы, строго соответствующие тому или иному аспекту объекта или системы. Были выделены 11 основных аспектов, и для каждого из них введен специальный реляционный символ. Рассмотрим данные символы, их смысл и пример вопроса, в соответствие которому может быть поставлен каждый реляционный символ (под X и Y будем понимать объекты интереса поискового запроса, если не указано другого):

(1) *Описание_структуры* – данный символ предназначен для представления вопросов вида «Как устроен X?». Например: «Как устроен двигатель внутреннего сгорания?»;

(2) *Описание_характеристик* – данный символ предназначен для представления вопросов вида «Каковы характеристики X?». Например: «Каковы характеристики автомобиля Mercedes ML 350?»;

(3) *Описание_работы* – данный символ предназначен для представления вопросов вида «Как работает X?». Например: «Как работает аппарат магнитно-резонансной томографии?»;

(4) *Описание_функций* – данный символ предназначен для представления вопросов вида «Каковы функции X?». Например: «Каковы функции системы менеджмента качества?», «Какие функции выполняет сервер локальной сети?»;

(5) *Описание_назначения* – данный символ предназначен для представления вопросов вида «Для чего предназначен X?», «Каково назначение X?». Например: «Для чего предназначен реостат?», «Каково назначение межкомпьютерной связи?»;

(6) *Описание_применения* – данный символ предназначен для представления вопросов вида «Где используется X?», «Как применять X?». Например: «Где используется Java?», «Как применять активную XSS?»;

(7) *Описание_принадлежности* – данный символ предназначен для представления вопросов вида «К какому классу принадлежит X?», «К какой категории относится X?». Например: «К какому классу соединений относятся жиры?», «К какой категории относятся офисы?»;

(8) *Описание_различий* – данный символ предназначен для представления вопросов вида «Чем отличается X от Y?», «В чем разница между X и Y?». Например: «Чем отличается архитектура x86 от x64?», «В чем разница между процессорами Dual Core и Core 2 Duo?»;

(9) *Описание_общих_характеристик* – данный символ предназначен для представления вопросов вида «Что общего у X с Y?», «Каковы общие черты X и Y?». Например: «Что общего у резины и каучука?», «Каковы общие черты финансов и денег?»;

(10) *Описание_особенностей* – данный символ предназначен для представления вопросов вида «Как ведет себя X [в ситуации Y]?», «Каковы особенности работы X [в условиях Y]?», где X – объект

интереса поискового запроса, а Y – опциональная часть вопроса, уточняющая вопрос, служащая дополнительным условием (ограничением). Например: «Как ведет себя аргон при повышенном давлении?», «Каковы особенности работы буровой установки при высокой температуре?»;

(11) *Описание_возможностей* – данный символ предназначен для представления вопросов вида «Каковы возможности X?». Например: «Каковы возможности платформы .NET?».

Следует отметить, что перечисленные виды вопросов могут и должны дополняться новыми типами для обеспечения большей степени покрытия потребностей поисковой системы. Для поддержки построения семантического представления поисковых запросов необходима гибкая расширяемая математическая модель системы первичных единиц концептуального уровня. Построим такую модель, используя в качестве отправной точки определения сортовой системы и концептуально-объектной системы из [1, 8].

Математическая модель проблемно-ориентированной системы первичных единиц концептуального уровня

В монографиях [1, 8] вводится базовая математическая модель для описания системы первичных единиц концептуального уровня, используемых прикладной интеллектуальной системой. Эта модель определяет новый класс формальных объектов, называемых *концептуальными базисами (к.б.)*. Каждый к.б. *B* строится для формализации определенной группы предметных областей и задает формальный язык *Ls(B)*, называемый *СК-языком (стандартным концептуальным языком) в базисе B*. Язык *Ls(B)* предназначен для построения семантических представлений (СП) произвольно сложных текстов, относящихся к рассматриваемой группе областей. Произвольный к.б. *B* является упорядоченной тройкой вида *(S, Ct, Ql)*, где *S, Ct* – формальные объекты, называемые соответственно *сортовой системой и концептуально-объектной системой*, а *Ql* – формальный объект, называемый *системой кванторов и логических связок*.

Сортовой системой в [1, 8] называется произвольная упорядоченная четверка *S* вида *(St, P, Gen, Tol)*, где *St* – конечное множество символов; *P* – элемент множества *St; Gen* – непустое бинарное отношение на *St*, являющееся частичным порядком на *St* (т. е. рефлексивным, транзитивным и антисимметричным); *Tol* – бинарное отношение на *St*, являющееся антирефлексивным и симметричным, и выполняется несколько дополнительных условий.

Элементы множества *St* называются *сортами; P – сортом «смысл сообщения»*; $\mathrm{Gen} \subset St \times St$ *– отношением общности;* $Tol \subset St \times St$ *– отношением совместимости* (толерантности). Если пара *(u, t)* входит в *Gen*, то можно использовать эквивалентную запись $u \to t$ и говорить, что *t* – конкретизация сорта *u*, а *u* – обобщение сорта *t*. Если $(s, u) \in Tol$, то используется запись $s \perp u$ и говорится, что сорт *s* совместим с сортом *u*.

Отношение общности *Gen* отражает существование иерархии (по степени общности) наиболее общих понятий – сортов. Например, для некоторого к.б. *B* отношение *Gen* может включать пары *(физич. объект, динамич. физич. объект), (физич. объект, физич. объект)*. Отношение совместимости *Tol* отражает существование различных, несопоставимых («ортогональных») семантических характеристик некоторых сущностей из рассматриваемой группы предметных областей. Например, человек одновременно является интеллектуальной системой и динамическим физическим объектом. Поэтому для некоторого к.б. *B* отношение *Tol* может включать пару *(интел. система, динамич. физич .объект)* и, в силу рефлексивности отношения, пару *(динамич. физич. объект, интел. система)*.

Пусть *S* – сортовая система вида *(St, P, Gen, Tol)*. Тогда произвольная упорядоченная четверка *Ct* вида *(X, V, tp, F)* в [1, 8] называется *концептуально-объектной системой*, согласованной с сортовой системой $S \Leftrightarrow$ когда выполняются следующие условия:

(1) *X, V* – счетные непересекающиеся множества символов; *tp* – отображение $X \cup V \to Tp(S)$;

(2) *F* – непустое подмножество множества *X*, для каждого $r \in F$ цепочка *tp(r)* начинается с подцепочки «{(« и заканчивается подцепочкой «)}»;

(3) *St* – непустое конечное подмножество множества *X*, и для любого $s \in St$ выполняется соотношение $tp(s) = \uparrow s$;

(4) $\{v \in V \mid tp(v) = [\text{сущ}]\}$ - счетное множество.

Множество *X* называется *первичным информационным универсумом*, элементы множеств *V* и *F* называются соответственно *переменными и функциональными символами*. Если элемент $d \in X \cup V$, $tp(d) = t$, то будем говорить, что *t – тип* элемента *d*.

Элементы множеств *X* и *V* интерпретируются как элементарные блоки, из которых (и нескольких

служебных символов) будут строиться семантические представления предложений и дискурсов. Например, *X* может включать элементы *город, отгрузка1, 125, 3/тонна, зелен, контейнер1, Столица, Вес, Цена, Часть, Элемент-множества, V* может содержать символы *x1, y3, z12, F* может включать элементы *Столица, Вес, Цена*.

Определение 1. Пусть *S* – произвольная сортовая система вида *(St, P, Gen, Tol)*, где *St* – множество сортов, *P* – выделенный сорт «смысл сообщения», *Gen* – отношение общности на *St, Tol* – отношение совместимости на *St* (см. [1, 8]). Тогда сортовую систему *S* будем называть **аспектно-ориентированной** ⇔ когда

(1) *St* включает выделенные, попарно различные сорта *техн.устр, физ.об;*

(2) (*физ.об, техн.устр)* ∈ *Gen*;

(3) $\{u \in St \mid (P, u) \in Gen\} \cap \{физ.об, техн.устр\} = \varnothing$.

Сорта *техн.устр* и *физ.об* интерпретируются как обозначения понятий «техническое устройство» и «физический объект».

Определение 2. Пусть *S* – произвольная аспектно-ориентированная сортовая система, *Ct* – концептуально-объектная система вида *(X, V, tp, F)*, согласованная с сортовой системой *S*, где *X* – множество символов (первичный информационный универсум), *V* и *F* – множества *переменных* и *функциональных символов* соответственно. Если элемент $d \in X \cup V$, $tp(d) = t$, то будем говорить, что *t* — тип элемента *d*. Тогда упорядоченная пятерка вида *Ctmw = (X, V, tp, F, Qf)* называется **слабо размеченной концептуально-объектной системой**, согласованной с сортовой системой *S* ⇔ когда выполняются следующие условия:

(1) X\F включает подмножество $Qf = \{r_1, ..., r_{11}\}$, $n = 11$, где

r_1 = описание_структуры,
r_2 =описание_характеристик,
r_3 =описание_работы,
r_4 =описание_функций,
r_5 =описание_назначения,
r_6=описание_применения,
r_7 =описание_принадлежности,
r_8 =описание_различий,
r_9 =описание_общих_характеристик,
r_{10} =описание_особенностей,
r_{11} =описание_возможностей;

(2) $tp(r_1)=tp(r_2)=...tp(r_{11}) = \{(физ.об, P)\}$.

Определение 3. Пусть *S* – произвольная аспектно-ориентированная сортовая система, *Ctmw* – слабо размеченная концептуально-объектная система вида *(X, V, tp, F, Qf)*, согласованная с *S*. Тогда упорядоченный набор *Cobs* вида *(X, V, tp, F, Qf, Chr, Qnf, Fgn)* будем называть **размеченной концептуально-объектной системой**, согласованной с сортовой системой *S* ⇔ когда выполняются следующие условия:

1. Набор *(X, V, tp, F, Qf)* является слабо размеченной концептуально-объектной системой, согласованной с сортовой системой *S*;

2. *Chr* – выделенное конечное подмножество множества унарных функциональных символов *F[1]* (интерпретируется как множество характеристик объектов заданной предметной области);

3. *Qnf* – конечное подмножество множества *F[1] \ Chr* (смысл элементов этого множества заключается в представлении характеристик, не принадлежащих объекту интереса поискового запроса);

4. Пусть *Concepts* – множество всех таких *d* из *X,* что тип *tp(d)* начинается с символа ↑(т.е. *d* – обозначение понятия).Тогда *Fgn* – это функция, ставящая в соответствие произвольному *s* из *Concepts* упорядоченную четверку *Gn = (Ge, Concr, Syn, An),* где *Ge* – множество обобщающих понятий объекта интереса поискового запроса, *Concr* – множество конкретизирующих понятий объекта интереса поискового запроса, *Syn* – множество синонимов (на концептуальном уровне) объекта интереса поискового запроса, *An* – множество антонимов объекта интереса поискового запроса, причем *Ge, Concr, Syn, An* – это подмножества множества *Concepts*. Функцию *Fgn* будем называть *детерминантом концептуального окружения*; если *s* – элемент множества *Concepts,* то упорядоченная четверка *Gn(s) = (Ge, Concr, Syn, An)* будет называться *концептуальным окружением термина s*.

Метод построения семантического расширения поискового запроса

Рассмотрим шаги, необходимые для построения семантического расширения поступившего поискового запроса на основе предлагаемой математической модели.

На первом шаге необходимо проанализировать поступивший на вход поисковый запрос ω с целью определения его типа. Для этого необходимо использовать определение слабо размеченной

концептуально-объектной системы вида *Ctmw=(X, V, tp, F, Qf)*, а именно, установить соответствие поискового запроса ω одному из элементов множества $Qf = \{r_1, ..., r_{11}\}$. Например, поисковому запросу ω = «Каковы характеристики платформы J2EE?» будет соответствовать реляционный символ $r_2 \in Qf \mid tp(r_2) = \{(физ.об, P)\}$, что означает, что данный запрос имеет тип «*Описание_характеристик*». После того, как определен тип вопроса, необходимо выделить первостепенный и второстепенный объекты интереса поискового запроса ω. Первостепенным объектом интереса будет являться ω_1 = «характеристика», а второстепенным ω_2= «платформа J2EE».

После определения основных характеристик запроса можно переходить к созданию множества вторичных поисковых запросов, порождаемых запросом ω, т.е. к построению семантического расширения входного запроса. Построение данного множества происходит на основании размеченной концептуально-объектной системы *Cobs* вида *(X, V, tp, F, Qf, Chr, Qnf, Fgn)*. Расширение поискового запроса происходит при помощи детерминанта концептуального окружения *Fgn*. Таким образом, необходимо построить набор

$Gn_{\omega_1} = (Ge_{\omega_1}, Concr_{\omega_1}, Syn_{\omega_1}, An_{\omega_1})$.

Для ω_1 = «характеристика» Gn_{ω_1} будет включать следующие элементы:

Ge_{ω_1} = {отзыв, рекомендация},

$Concr_{\omega_1}$ = {описание},

Syn_{ω_1} = {портрет, описание}, $An_{\omega_1} = \varnothing$.

Как только расширенное множество запросов сформировано, оно передается в традиционную поисковую систему, возвращающую множество релевантных документов. При анализе документов будут использоваться такие компоненты размеченной концептуально-объектной системы, как множества *Chr* для определения степени соответствия информации в документе характеристикам, присущим заданному объекту интереса, и *Qnf* для фильтрации документов, содержащих нерелевантную информацию. Кроме того, для фильтрации нерелевантных документов используется множество антонимов An_{ω_1} из концептуального окружения термина ω_1. Предполагается, что данные множества формируются в зависимости от предпочтений пользователя, т.е. в зависимости от его поведения и выбора тех или иных результатов поиска.

Пример построения семантически преобразованного множества поисковых запросов

Проиллюстрируем на примере построение сначала концептуального окружения термина *s* из отдельно взятого поискового запроса на естественном языке и затем – семантического расширения запроса. Пусть задан запрос ω = «Каковы особенности компьютера MacBook Pro?». Данный поисковый запрос относится к типу (10), описанному ранее, и соответствует реляционному символу *Описание_особенностей*. Первостепенным объектом интереса данного запроса являются особенности определенного объекта, в данном случае – компьютера MacBook Pro, а не сам компьютер. Пусть *s* = *«особенности»*, тогда для определения лексикографического окружения нам необходимо привести данное слово к нормальной форме (именительный падеж, единственное число, т.к. это существительное). Информация для преобразования такого типа традиционно содержится в лингвистической базе знаний, описание которой не затрагивается в данной работе, однако наличие таковой необходимо для реализации предлагаемого подхода. Нормализованный термин *s'*= «*особенность*», в таком случае *Gn(s')* = ({*свойство, черта*},{}, {*непохожесть, отличие*}, {*сходство, похожесть*}).

Как видно из построенного концептуального окружения, данный термин не имеет конкретизирующих понятий. Построение семантического расширения поискового запроса осуществляется следующим образом: сначала добавляется первоначальный поисковый запрос, затем поисковый запрос, использующий синонимы объекта интереса, затем поисковый запрос с конкретизирующими понятиями и, наконец, поисковый запрос, построенный с использованием обобщающих понятий термина *s*. Следует отметить, что построение концептуального окружения проводилось для нормализованного термина *s'*. Для построения грамматически равносильных поисковых запросов необходимо выполнить денормализацию лексикографического окружения и привести его в ту форму, в которой находился термин *s* в поисковом запросе, т.е. *Gn(s)* = ({*свойства, черты*},{},{*непохожести, отличия*},{*сходства, похожести*}). Результирующее множество поисковых запросов будет выглядеть так: *Wse* = {«Каковы особенности компьютера MacBook Pro?», «Каковы непохожести компьютера MacBook Pro?», «Каковы отличия компьютера MacBook Pro?», «Каковы свойства компью-

тера MacBook Pro?», «Каковы черты компьютера MacBook Pro?»}.

Однако, несмотря на проведенные преобразования, результат работы поисковой системы не будет удовлетворительным из-за вопросительной структуры поискового запроса, которая редко встречается в документах, содержащих описания подобных характеристик заданного объекта. Поэтому необходимо выполнить стемминг поисковых запросов.

Следует пояснить, что в традиционном поиске понятие стемминга термина поискового запроса означает сохранение лишь основы слова, чтобы избежать зависимости от различных словоформ, встречающихся в разных документах. Для поискового запроса в целом стеммингом будет являться сокращение данного запроса до семантически значимых составляющих, т.е. до первостепенных и второстепенных объектов интереса. В представленном примере множество *Wse* примет следующий вид: *Wse'* = {«MacBook Pro особенности», «MacBook Pro непохожести», «MacBook Pro отличия», «MacBook Pro свойства», «MacBook Pro черты»}. Изменение порядка и количества слов позволяет с наибольшей точностью найти документы, содержащие необходимую информацию. Элементы представленного множества поисковых запросов с большей вероятностью встречаются в различных документах, что позволяет избежать ограничения лишь по документам, содержащим описание непосредственно «особенностей» данного объекта.

Приведем пример результатов работы поисковой системы (в данном случае — Яndex) для первоначального поискового запроса и для множества сгенерированных запросов:

Таблица 1.

Сравнение результатов выдачи поисковой системы

w = «Каковы особенности компьютера MacBook Pro?»	Wse' = {«MacBook Pro особенности», «MacBook Pro непохожести», «MacBook Pro отличия», «MacBook Pro свойства», «MacBook Pro черты»}
re: Store — Apple MacBook Pro 13″ Ноутбук **MacBook Pro 13″**, блок питания MagSafe, шнур питания, адаптер, диски с ПО, салфетка для очистки экрана, документация. Отличительные **особенности**…	**Apple - MacBook Pro - Спецификации 15-дюймовой модели** **MacBook Pro** имеет следующие **особенности** для снижения воздействия на окружающую среду: корпус из алюминия и стекла, пригодный для вторичной переработки; экран со светодиодной подсветкой, не содержащий ртути
Ремонт ноутбуков MacBook от Apple, срочный ремонт MacBook Pro… Здравствуйте, у меня на **macbook pro** сначала перестал записивать дисковод, а потом там остался диск и он не выдает его обратно, какова ориентировочная стоимость ремонта? Ответ ...	**Apple - MacBook Pro - Производительность - Скоростные процессоры…** В **отличие** от систем, где модули памяти подключены к процессору через отдельный контроллер, в новом **MacBook Pro** используется интегрированный контроллер памяти, с которым можно подключать память напрямую к процессору.
	Apple MacBook Pro -- Обзоры -- mobi.ru Главная **черта** внешности **MacBook Pro** — минимализм. Ничего лишнего, ни единого декоративного элемента.
Мидис Запорожье:: Ноутбук Apple MacBook Pro Компьютеры и ноутбук… Всю необходимую информацию о **MacBook Pro** вы найдёте на сайте Apple.com: видеоруководства по iLife, **Mac** OS X, Aperture и другие материалы. Узнайте об **особенностях** нового **MacBook Pro**. Прочтите советы по работе с ним.	**Top List: Ноутбуки - обзор Apple MacBook Pro. Характеристика. Описание** Дорожные **свойства**. Первое, что ощущается, когда берёшь **MacBook Pro** в руки – это его совсем небольшая толщина.
	Замена жёсткого диска в MacBook Pro - Все о продукции Apple Замена жёсткого диска в моём **MacBook Pro** прошла просто замечательно… Каждая из моделей обладает своими особенностями, придающими ноутбукам определенные **свойства**, полезные в тех или иных...

Из результатов, представленных в *таблице 1*, видно, что пользователь получает более широкий набор результатов, удовлетворяющих его запросу. Более того, в результатах не будут присутствовать вопросительные слова. В дальнейшем планируется использовать специализированную базу знаний, содержащую определения объектов интереса поискового запроса и некоторую дополнительную информацию, которая может быть использована при анализе результатов работы поисковой системы.

Построение лингвистической базы знаний

Построение семантических представлений аспектно-ориентированных поисковых запросов должно поддерживаться лингвистической базой знаний. Лингвистическая база знаний (ЛБЗ) необходима для хранения информации о грамматических свойствах слов, для определения семантического контекста того или иного слова, для построения взаимосвязей между словами, а также для добавления определений значений слов и построения иерархии понятий. Для построения такой базы требуется адекватная мате-

матическая модель. В теории К-представлений [1, 8] предлагается математическая модель лингвистической базы данных, которая частично удовлетворяет поставленным требованиям. Тем не менее, предложенная модель не учитывает взаимосвязи между словами, которые являются синонимами, антонимами, меронимами, холонимами, гиперонимами и гипонимами. Для определения данных связей между словами предложенную математическую модель ЛБД необходимо дополнить функцией, ставящей заданному слову в соответствие его расширенное лексикографическое окружение в зависимости от значения и предметной области, которой принадлежит данное слово. Помимо этого, требуется найти такой источник данных, который позволил бы наполнить построенную модель фактическими данными. Существующие в настоящее время коллективно разрабатываемые базы знаний Wiktionary и Wikipedia [9] имеют ряд преимуществ по сравнению с такими традиционными решениями, как WordNet [10] и некоторые его разновидности [9]. Тем не менее, данные источники данных не являются полностью приемлемыми в силу различных структурных и технологических ограничений. В настоящее время ведется работа над оптимизацией их использования в качестве источника данных, а также преобразования их структуры в наиболее соответствующую разрабатываемой математической модели.

Заключение

Предлагаемый подход не только решает проблемы традиционных поисковых систем, но и позволяет расширить возможности семантического поиска за счет реализации естественно-языкового интерфейса, построения семантически близких, но грамматически отличающихся запросов, а также возможности использовать базу знаний, содержащую информацию о семантических единицах.

Введенное формальное определение размеченной концептуально-объектной системы, по сравнению с введенным В.А. Фомичевым в теории К-представлений понятием концептуально-объектной системы (см. [1, 8]), позволяет:

- ✦ конструировать иерархию понятий (а не только сортов) по степени их общности;
- ✦ ставить в соответствие понятию его лексикографическое окружение;
- ✦ определять обобщающие, уточняющие и аналогичные семантически, но отличающиеся синтаксически понятия (с помощью функции «детерминант концептуального окружения понятия»);
- ✦ выделить подкласс информационных единиц с отрицательным значением.

Предложенная формальная модель полностью поддерживает процесс преобразования поискового запроса к расширенному виду, связанный с построением множества семантически близких поисковых запросов. Представляется, что продолжение работы в данном направлении позволит расширить множество анализируемых вопросов, повысить качество анализа при помощи дополнительных критериев, а также построить систему семантически-ориентированного анализа естественно-языковых вопросов, позволяющую значительно повысить семантическую релевантность результатов работы традиционных систем поиска по ключевым словам. ■

Литература

1. Фомичев, В. 2005. Формализация проектирования лингвистических процессоров. МАКС Пресс. Москва.
2. Кириллов, А. 2009. Поисковые системы: компоненты, логика и методы ранжирования. Бизнес-информатика № 4(10). С. 51—59. Москва.
3. Halpin, H and Lavrenko, V. 2009. Relevance Feedback Between Hypertext and Semantic Search. Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
4. Meij, E, Mika, P and Zaragoza, H. 2009. Investigating the Demand Side of Semantic Search through Query Log Analysis. Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
5. Fernandez, M and Lopez, V. 2009. Using TREC for cross-comparison between classic IR and ontology-based search models at a Web scale. Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
6. Dali, L, Rusu, D and Fortuna, B. 2009. Question Answering Based on Semantic Graphs. Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
7. Akbik, A and Broth, J. 2009. Wanderlust: Extracting Semantic relations from Natural Language Text Using Dependency Grammar Patterns. Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
8. Fomichov, V. 2010. Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Springer: New York, Dordrecht, Heidelberg, London.
9. Zesch, T, Müller, C and Gurevych, I. 2008. Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. Proc. Conference LREC 2008 (Marrakech, Morocco).
10. Fellbaum, C and Miller, G. 1998. WordNet. An electronic lexical database. Cambridge, MA: MIT Press; 1998. 422 p.