Телекомунікаційні системи

УДК 004.822

В.В. Артамонов, В.А. Тертышный

Кременчугский национальный университет имени Михаила Остроградского, Кременчуг

РАЗРАБОТКА МОДЕЛИ ИНФОРМАЦИОННОГО ПОИСКА С ИСПОЛЬЗОВАНИЕМ СВЯЗАННЫХ ДАННЫХ

В статье предлагается модель поиска, основанная на использовании связанных данных для построения архитектуры системы и выбора подхода к поиску, в основе которого лежит поиск сущностей и формирование фактов. Данный подход позволяет формировать стратегию поиска для определенных специализированных предметных областей. Представлена архитектура поисковой системы и практические аспекты реализации. Разрабатываемая модель может быть реализована в корпоративных системах документооборота.

Ключевые слова: *поиск, сущность, семантическое пространство, связанные данные, RDF.*

Введение

Постановка проблемы. Информационно-поисковая система (ИПС) – это комплекс программных средств, обеспечивающих избирательный отбор по заданным признакам документов, хранимых в электронном представлении.

Модель поиска – это комплекс трех составляющих (рис. 1), а именно: образ представления документа, представление поисковых запросов и набор критериев определения релевантности.

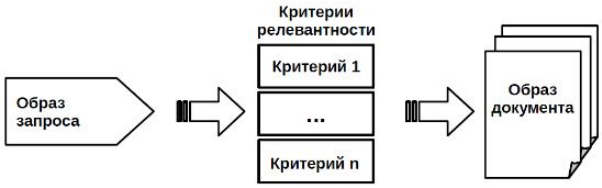

Рис. 1. Схематическое изображение модели поиска

Анализ последних исследований и публикаций. На сегодняшний день существует ряд моделей поиска информации, которые используют обширный список методов, алгоритмов, и подходов как современных, так и классических. Среди наиболее известных – модели на основе ссылочного ранжирования, модели на основе метаописаний, онтологий, использования тезаурусов и другие. В теории поиска информации уже существует набор алгоритмов и методов поиска: лексикографические, древовидные, реляционные и другие [1 – 3].

Существенным подспорьем для построения поисковых систем является использование подхода на основе связанных данных (Linked Data). Данный подход предоставляет разработчику набор инструментов для моделирования и разработки систем информационного поиска на основе графовых структур. В последнее время в связи с возрастанием требований к поисковым системам, а именно возможности сбора структурированной информации, а также полноты поисковой выдачи возникла необходимость перехода от ссылочно-накопительной к семантической модели поискового аппарата.

Первостепенной необходимостью для информационного поиска является возможность сбора информации пользователем из различных по семантическим характеристикам источников. Основой такой технологии является семантическая сеть, содержащая объекты и отношения между ними, которые могут интерпретироваться чтобы обработать смысл запроса пользователя [1, 4, 9, 12].

Поэтому, одним из важных параметров высокоэффективных поисковых систем является сбор фактографической информации об определенном объекте, событии или же набора комплексной информации.

В данной статье предложена формальная модель специализированного поиска, предоставляющая возможность сбора информации в контекстном ключе.

Наиболее популярные модели базируются на теории множеств [3]. Некоторые подходы используют векторную алгебру. Оба подхода достаточно эффективны на практике, однако у них есть общий недостаток, который следует из основного упрощающего предположения, заключающегося в том, что смысл документа, его основное содержание определяется множеством ключевых слов – терминов и понятий, входящих в него. Безусловно главным достоинством такого подхода является быстрый поиск и возможность группировки по формальным

признакам, но при этом происходит потеря контекстного концепта.

Самой распространенной является булева модель поиска (БМП), которая основывается на теории множеств и булевой алгебре [1, 5, 8]. Популярность этой модели обусловлена в первую очередь простотой реализации. С помощью БМП можно проводить индексацию и выполнять поиск в массивах документов большого объема.

В рамках БМП документы и запросы предоставляются в виде множества морфемных основ ключевых слов – термов t_i. Имеющийся набор документов индексируется относительно термов. Набор термов называется словарем.

Словарь по своей сути – это проиндексированная база всех термов:

$$T = \{t_1, t_2, ..., t_i\}. \quad (1)$$

Тогда документ в этом случае является подмножеством словаря: $D \subseteq T$. Запрос – это булево выражение и может иметь вид:

$$t_5 \text{OR} t_7 \text{AND NOT} t_{12}. \quad (2)$$

Выражение (2) означает, что необходимо найти документы, которые включают пятый или седьмой термы словаря, но не включают двенадцатый. Если выражение срабатывает для определенного документа, то можно говорить о релевантности запроса и документа.

Недостатком БМП является отсутствие гибкости и непригодность для ранжирования документов.

В векторной модели поиска (ВМП) словарь термов (1) представлен в виде n-мерного евклидового пространства термов документов, обработанных поисковой машиной, вектор документов d, вектор запроса q (рис. 2) [7].

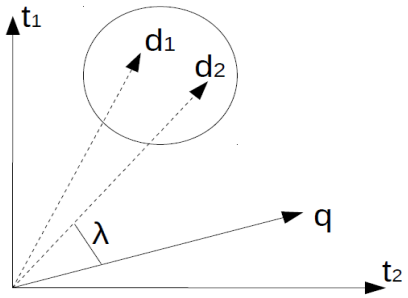

Рис. 2. Схематическое изображение векторной модели

Мера λ – мера релевантности, вычисляется как косинус угла между векторами запроса и документа. Чем ближе векторы, тем больше релевантность.

Достоинства векторной модели:

– четкая оценка соответствия документа запросу;

– косинусная метрика, удобная для ранжирования;

– учет весов повышает эффективность поиска.

Недостатки векторной модели:

– термы не могут иметь ортогональной меры в связи зависимостью друг друга;

– нет достаточного обоснования и методологических возможностей построения пространства термов.

Еще одной из распространенных моделей есть вероятностная модель, использующая “Принцип ранжирования вероятностей” по которому вычисляется оценка вероятности того, что документ является релевантным по отношению к запросу на основе списка документов, ранжированного по убыванию вероятности полезности для пользователя [3, 5].

Вероятность того, что документ d – релевантный, вычисляется на основе теоремы Байеса:

$$P(R \mid d) = \frac{(P(d \mid R) \cdot P(R))}{P(d)}, \quad (3)$$

где P(R) – вероятность того, что документ d – релевантный из коллекции D, P(d|R) – вероятность случайного выбора именно документа d из множества релевантных документов, P(d) – вероятность случайного выбора документа d из коллекции документов D.

Документы ранжируются по величине P(R|d).

Вероятностной модели присущи недостатки:

– структура документа описывается только термами;

– оптимальные результаты можно получить только в процессе обучения системы на основе имеющейся информации о релевантности;

– необходима информация о релевантности или её приближенные оценки.

На практике чаще всего используются гибридные решения, которых объединены возможности различных подходов. Также в современных информационно-поисковых системах все чаще используются методы семантической обработки информации.

На основе вышеописанного можно выделить основные понятия и цели для построения эффективной модели поиска.

Во-первых, необходимым фактором для построения модели является формальное описание документа. Ведь документ – это определенный объем информации, которой оперируют информационно-поисковые системы. Сам по себе документ может представлять содержательно законченную единицу информации, представленную на каком-либо естественном языке, которая идентифицируется уникальным образом.

Как указывалось выше, задача поиска состоит в выборе ранжированного набора документов A по определенному признаку из общего множества имеющихся объектов B базы данных (рис. 3), то есть $A \subset B$.

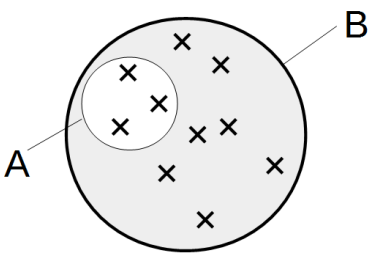

Рис. 3. Пространственное отображение ранжированного набора документов

Особый интерес возникает к поиску информации в гетерогенной, нечеткой среде.

Нечеткая среда – это среда, атрибуты которой, могут иметь нечеткие значения. Существует четыре пары групп систем поиска по взаимоотношению структурирования (табл. 1).

Таблица 1

Пары групп взаимодействия запрос-среда

Запрос	Среда
Структурированный	Структурированная
Структурированный	Неструктурированная
Неструктурированный	Структурированная
Неструктурированный	Неструктурированная

Разработанные модели поиска информации в неструктурированной среде включают в себя набор методов: лексикографический, семантико-лексический, атрибутивный опираясь в основном на полнотекстовые базы данных. Полнотекстовая БД – это, по сути, множество документов, которые связанны друг с другом перекрестными ссылками, поэтому наиболее адекватно она представляется в форме семантической сети.

Тогда $A = \{x : P(x)\}$, (4)

где P – метрическая функция отбора.

Однако для хранения детализированной информации о предметной области её необходимо дополнить множеством сущностей – понятий, которые содержатся в документах.

Здесь главной проблемой является отсутствие широкого контекстного пространства в процессе поиска. Связи имеют однонаправленный характер, в документе отсутствует информация о связи и о том, что на него ссылается другой документ. Также связи являются малоинформативными. Само по себе наличие связи свидетельствует лишь об указании ссылки на документ, но не о её характере.

Классические модели устанавливают лишь связи между текстами документов. Неинформативность таких связей приводит к отсутствию иерархически и категориально структурированных наборов массивов документов различных предметных областей. Поэтому вопрос о создании семантических технологий для повышения качества поиска возник давно. Главным фундаментом для построения модели, которая бы смогла предоставлять структурированную и подробную информацию для определенной предметной области, являются связанные данные. В чем особенность и достоинства данной технологии?

Данная технология использует семантический подход для формализации многих аспектов. В первую очередь это использование Entity-Relationship model (ER), что позволяет гибко использовать формализацию объектов предметной области. Также в отличие от стандартных подходов связи устанавливаются между объектами и являются информативными, то есть не только констатирую связь, но и указывают на её характер.

Для установки ссылок также в изложенных выше подходах используются универсальные идентификаторы (URI). Но, URI могут существовать не только у страниц, но также и у объектов или абстрактных понятий. В целом, URI глобально уникальны, поэтому могут быть использованы для установления ссылок в различных местах.

Исходя из этого связи устанавливаются не между текстами, а между данными, а точнее - между группами сущностей, объединенных по формальным признакам. Для примера это могут быть не только веб-документы, а любые объекты реального мира (люди, города, научные статьи, продукция и т. д.). Вследствие образуется так называемая сеть данных.

Описательные формы сущностей представляются в форме стандартных языков, таких как RDF. RDF является хорошей средой описания связанных данных. И в первую очередь RDF является графовой моделью, которая позволяет производить интерпретацию информации в удобные, формальные формы (сущности). RDF определяет связи между сущностями. Подобные связи могут использоваться для навигации, а также интеграции информации из различных источников предприятия и Web-узлов.

Связанные данные описываются в виде отдельных суждений, представленных триплетами (субъект-предикат-объект). Триплет может быть представлен в виде графа. В самом графе субъект и объект являются узлами, а предикат – соединительная связь направленная от субъекта к объекту.

Чтобы обеспечить однозначную компьютерную обработку таких суждений и строить графы связанных данных из отдельных троек, все сущности в RDF представляются не наименованиями, а уникальными идентификаторами (URI-универсальными идентификаторами ресурса). Внешне URI может выглядеть как привычный электронный адрес. Но, в отличие от URL, URI не ведет к ресурсу – это лишь идентифицирующая метка. Тройки с совпадающими субъектами или объектами объединяются в граф. Используя именованные связи, программа может проследить всю цепочку.

Очевидно, что информативность (а следовательно и полезность) данных увеличивается по мере увеличения количества связей с другими данными.

Субъект и предикат всегда выражены только идентификатором. Объект может также быть представлен идентификатором, либо литералом–текстовой строкой. Это может быть имя лица, заглавие ресурса, наименование организации, предметная рубрика. Цепочку связанных данных всегда завершают литералы, которые используются для человекочитаемого вывода (дисплейного представления). То есть читатель видит только то, что находится на конце цепочки связанных данных (заглавие, имя автора, географическое название и т.д.); все промежуточные звенья этой цепочки от него, как правило, скрыты, и он может и не знать, что имеет дело со связанными данными. Как видим, подход RDF очень отличается от традиционной записи. В привычной нам записи различные аспекты ресурса объединяются в одну цепочку с помощью особого синтаксиса – тегов полей, индикаторов, подполей. В RDF данные разделены на отдельные суждения, которые могут обрабатываться независимо друг от друга; важно отметить, что могут использоваться суждения (триплеты), из разных источников и разных схем метаданных.

Основной материал исследования

Проблемой имеющихся моделей является отсутствие возможности адаптировать общую выработанную модель к определенной узкой предметной области. Или же наоборот, некоторые модели разработаны только под узкоспециализированные промышленные, научные и документные системы.

Следовательно, необходимо расширить формализованное представление сущности, как основной семантической единицы информационного пространства:

$$E = \{E_i, A, L_i, LP, E_a, A_E\}, \qquad (5)$$

где E_i – идентификатор сущности в семантическом пространстве, A – множество атрибутов сущности, L_i – связи сущности в семантическом пространстве с другими объектами, LP – свойства связей, E_a – действия сущности, A_E – действия над сущностью.

Сущность является частью общего семантического пространства. В данном случае семантическое пространство (СП) – это система категорий индивидуального сознания, при помощи которых происходит оценка и классификация различных объектов. Если принять определенные допущения, в частности о независимости данных категорий, то появляется возможность размещения тех или иных значений в многомерном семантическом пространстве, где можно вычислять расстояние между термами [12].

Формальная модель СП:

$$M_{sm} =< E, P(q_i, d_i), L >, \qquad (6)$$

где E – базовая сущность СП, P – функция ранжирования запроса относительно документа, q_i – запрос к СП, d_i – идентификатор документа в БД, L – имеющиеся связи в СП.

Данная модель имеет ряд преимуществ перед традиционным онтологическим подходом т. к. в ней сущности рассматриваются в контексте связанных данных.

Данный подход предоставляет возможность группировки лексем на основе имеющихся атрибутов и связей в СП. При осуществлении поиска и проведении анализа конечными результатами являются конкретные события, персоны и контекстные отношения между ними, которые можно интерпретировать как факты. В контексте работы фактом будем называть суждение, которое может быть элементарным (триплет) или сложным логическим высказыванием. Это означает, что концептуальный интерфейс семантического пространства базируется на наборе сущностей ПрО и наборе фактов. Например, имея факт Man = Human ∩ Male, можно создать схему факта:

$$\text{Father} = \text{Man} \cap \exists \text{hasChild}.$$

Модель факта в этом случае включает:

– имена сущностей E_0, E_1,… (Person, Female и др.);

– имена ролей и взаимоотношений r_0, r_1,… (has, hasChild, hasPhone и др.);

– квантор утверждения о существовании;

– логическую операцию конъюнкции.

Таким образом, предлагаемый образ документа основывается на модели СП, а также множестве сущностей и их отношений (фактов), полученных путем анализа ПрО. То есть документ представляет собой часть объединенного узла концептуальной схемы (графа).

Данная модель предоставляет возможность осуществления интерпретации, семантического расширения и сохранения в виде связанных данных пользовательских запросов и документов с помощью как формальных, так и естественно-языковых форм представления. Способ представления запроса должен поддерживать сложные логические запросы, реализуемые элементами управления на стороне клиента или в виде транзакционного запроса между модулями, интегрированными в общий программный комплекс, например:

(характеристика1 = true AND (характеристика2 <5))
OR (характеристика1 = false AND
(характеристика3 > 7)) .

Механизмом уточнения и расширения запросов является использование логических связок AND и OR. Критерием релевантности документа служит мера семантической близости, описанная в работе [15]. Ранжирование документа во множестве найденных информационных единиц производится с

помощью IDF-функции – инверсной частоты терма, которая в классической интерпретации имеет вид:

$$\mathrm{IDF}(t,q) = \log\frac{N}{q_t}, \qquad (7)$$

где N – общее количество документов, q_t – количество документов, которые содержат терм t.

Таким образом, основными особенностями предлагаемой модели поиска являются:

– использование расширенных семантических данных, полученных в ходе обработки запроса и в процессе поисковой обработки и полученных в виде веб-документов;

– использование формальной модели семантического пространства, позволяющее реализовать фактографический поиск наряду с индексацией документов;

– возможность гибкой подстройки системы под узкоспециализированные предметные области путём обработки, интерпретации и систематизации запросов пользователей и найденных документов.

Весь процесс поиска представлен на рис. 4.

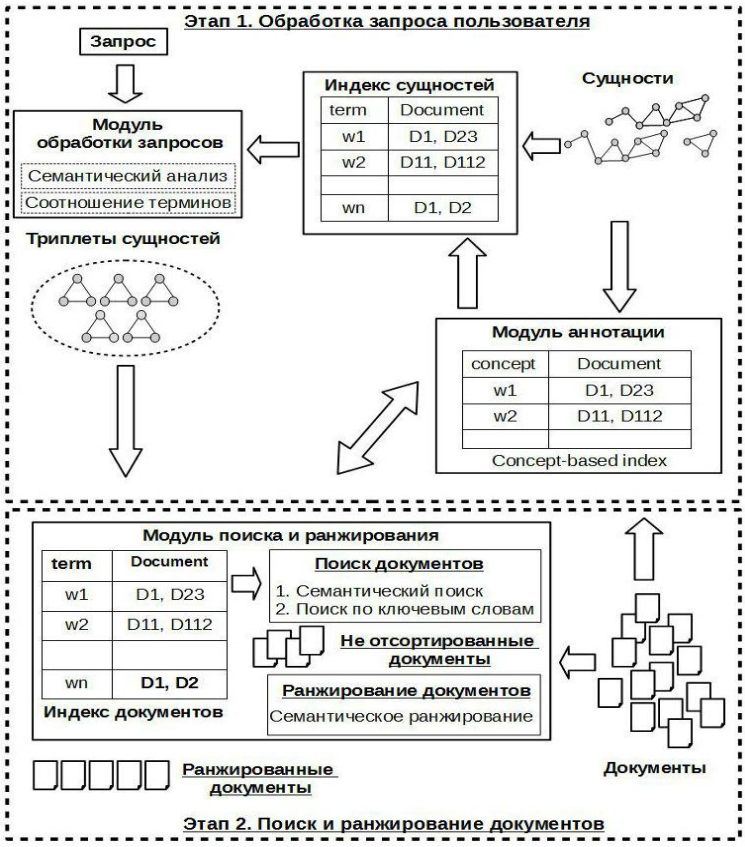

Рис. 4. Схема процесса поиска

Процесс состоит из двух этапов:

1. Обработка и интерпретация запроса пользователя.
2. Поиск и ранжирование документов и фактов.

Запросы пользователя обрабатывается на entity-based QA аннотации. В случае если запрос вмещает несколько сущностей, то модуль аннотации переводит терминологию пользователя в терминологию доступных сущностей и выдает список сущностей в результате. Для обеспечения быстрого доступа (онлайн) к сущностям, они индексируются априори. Получение ответа происходит от потенциально неограниченного количества сущностей, которые охватывают неограниченное множество доменов. Запрос соответствующий ПрО, будет обрабатываться так:

1) выбор соответствующих сущностей, на основе контекста и доступной семантической информации;

2) формирование ответа на основе сущностей.

На втором этапе проводится поиск и ранжирование релевантных документов обработанных на основе предыдущих поисковых шаблонов сущностей. На данном этапе проводится также автоматическое индексирование сущностей. Такой подход предоставляет возможность динамического масштабирования в больших репозиториях данных.

Оба этапа осуществляются благодаря четырем модулям, входящим в архитектуру:

1. Модуль индексирования сущностей, являющийся препроцессором имеющейся семантической информации.

2. Модуль обработки запросов, формирующий набор триплетов – вариантов запроса с семантически близкими сущностями.

3. Аннотационный модуль, генерирующий концепт-индекс между документами, сущностями и фактами.

4. Модуль поиска и ранжирования, который производит поиск и ранжирование релевантных документов относительно триплетов, сформированных модулем обработки запросов.

В результате система должна выдать набор сущностей на основе запроса и дополнительный набор – ранжированный список документов.

Модуль обработки запросов (рисунок 5) интерпретирует запрос пользователя путем выявления триплетных ассоциаций, которые соотносятся с триплетами в наборе. Для этого существует карта терминов (Entity-mapping table) определенной сущности в словаре, базирующемся на множестве триплетов и их отношений. Конечным результатом является набор сгенерированных триплетов.

Триплеты являются базисными элементами для построения двусвязного, семантического пространства, являющегося структурой для хранения сущностей. Тогда в случае сложных, комплексных запросов система сможет “отвечать” атомарными фактами. Набор фактов, непосредственно отвечающих на запрос будет устанавливать соответствие не только между различными триплетами, входящими в семантическое описание, но и между сущностями, имеющими наборы искомых триплетов.

Также предусматривается уточнение запросов пользователей. В этом случае система будет адаптироваться под запросы. Исходя из набора построенных семантических связей используя язык расширенных запросов можно получить некую уже готовую часть топологии семантического пространства, что в свою очередь предоставить возможность агрегации данных и осуществления поиска по уже готовому образу (то есть паттерну). При этом можно адаптация может также включать особенности различных аспектов предметной области.

Практическое внедрение поисковой системы указанной архитектуры предполагает наличие распределённой структуры информационной системы предприятия. Схема реализации модулей приложения с использованием специализированной системы поиска изображена на рис. 6.

Система специализированного поиска интегрируется в общесистемную архитектуру корпоративной системы документооборота. Для реализации ПС в конкретной ПрО необходимо формирование предустановленного набора тезаурусов, которые далее будут пополняться автоматически.

Выводы

Проведен сравнительный анализ имеющихся моделей поиска. Выявлено, что недостатками имеющихся моделей являются:

– отсутствие возможности поиска с учетом конкретной семантической особенности слов, сущностей;

– сложность поиска по абстрактным запросам;

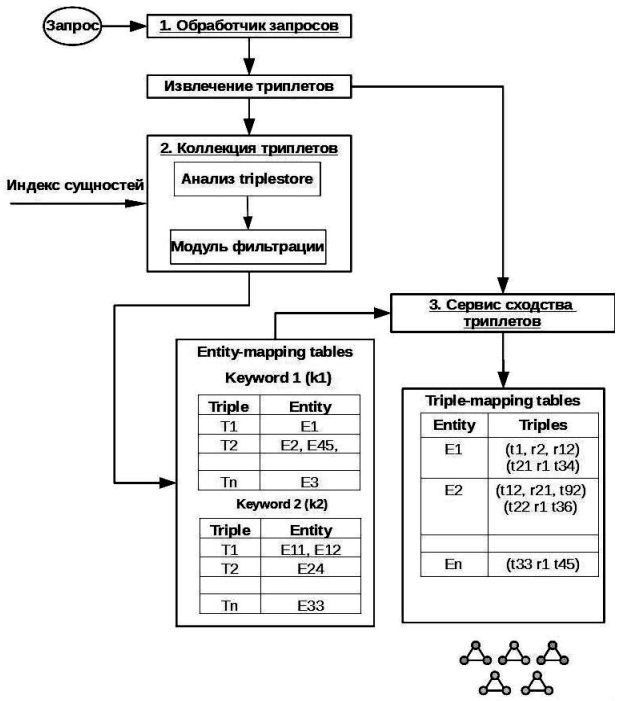

Рис. 5. Модуль обработки запросов

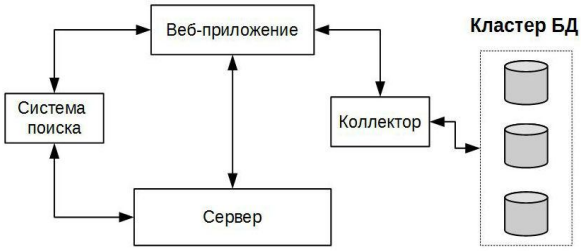

Рис. 6. Схема практической реализации специализированной системы поиска

– результаты поиска не объединяются в категории;

– в некоторых случаях, отсутствие контекстных операторов и невозможность ранжирования результатов;

В ходе исследований получена общая классификация имеющихся архитектур поисковых систем. Был выявлен ряд архитектур, использующих семантические технологии для создания поисковой системы на основе связанных данных.

Предложена модель специализированного поиска на основе связанных данных, которая, в отличие от существующих, позволяет проводить поиск информации в контекстном, узкоспециализированном пространстве предметной области, и обеспечивает возможность автоматизированной подстройки под выбранную предметную область.

Основными особенностями предлагаемой модели поиска являются:

– использование расширенных семантических данных, полученных в ходе обработки запроса и в процессе поисковой обработки и полученных в виде веб-документов;

– использование формальной модели семантического пространства, позволяющее реализовать фактографический поиск наряду с индексацией документов;

– возможность гибкой подстройки системы под узкоспециализированные предметные области путём обработки, интерпретации и систематизации запросов пользователей и найденных документов.

Основу модели представляют подсистемы обработки и поиска данных. Базовой концепцией работы модели является построение семантического пространства, использующего реальные знания о мире, построение связей между сущностями, формальное иерархическое сопоставление узлов семантического пространства.

Предложена архитектура для построения и организации смыслового поиска внутри информационного пространства предприятия. В основе архитектуры лежит использование семантического пространства с использованием иерархии сущностей и их связей между отдельно взятыми сущностями и группами сущностей. Отличительной чертой предлагаемой архитектуры является возможность использования системы под информационные нужды определенного предприятия с относительно простой и быстрой интеграцией в общесистемный стек.

Архитектура системы не требует больших финансовых затрат и обладает свойством расширения и дополнения информационной базы знаний в зависимости от информационных потребностей предприятия.

Список литературы

1. Губин М.Ю. Методы создания семантических метаописаний документов с применением семантических сетей, фреймовых моделей и частотных характеристик / М.Ю. Губин, В.В. Разин, А.Ф. Тузовский // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2010. – Т. 2, № 2. – С. 227-229.

2. Рабчевский, Е.А. Автоматическое построение онтологии на основе лексико-синтаксических шаблонов для информационного поиска / Е.А. Рабчевский // Труды 11-й всероссийской научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Петрозаводск, 2009. – С. 69-77.

3. Марчук А.Г. На пути к большим RDF данным / А.Г. Марчук // Конф. 15-18 октября 2013 г. Переславль-Залесский – 2013. – С. 51-56.

4. Ландэ Д.В. Эффективный поиск знаний в Интернет. Профессиональная работа / Д.В. Ландэ. – М.: Издательский дом "Вильямс", 2005. – 270 с.

5. Baziz, M. Semantic cores for representing documents in information retrieval / M. Baziz, M. Boughanem, N. Aussenac-Gilles, C. Chrisment // Proc. Of 2005 ACM symposium on applied computing. - New Mexico, 2005. – P. 1011-1017.

6. Google. The knowledge graph [Electronic resource]. – Accessed to: http://www.google.com/insidesearch/features/search/knowledge.html.

7. Guha, R.V. Semantic search / R.V. Guha, R. McCool, E. Miller // Proc. of the 12th inter. WWW conf. (WWW 2003). - Budapest, Hungary, 2003. – P. 700-709.

8. Tao Cheng, Kevin Chen-Chuan Chang (2007) "Entity Search Engine: Towards Agile Best-Effort In-formation Integration over the Web". CIDR. – P. 108-113.

9. Rio Blanco, Peter Milka, Sebatiano Vigna (2011) "Effective and Efficient Entity Search in RDF data". The Semantic Web – ISWC – Springer. – 92 p.

10. Miller A. A semantic concordance / A. Miller, C. Leacock, R. Tengi, R.T. Bunker // 93rd proc. of the workshop on Human Language Technology. - PA: USA, 1993. – P. 303-308.

11. Pedersen, T. Measures of semantic similarity and relatedness in the medical domain / T. Pedersen, S. Pakhamov, S. Patwardhan // University of Minnesota digital technology center research report DTC 2005/12.

12. Resnik, P. Semantic similarity in a taxonomy: An information-based measures and its application to problems of ambiguity in natural language / P. Resnik // Journal of artificial intelligence. – 1999. – P. 95-130.

13. Shah, U. Information Retrieval on the Semantic Web / U. Shah, T. Finin, A. Joshi, R. Cost, J. Mayfield // 10th Inter. Conf.. on Information and Knowledge Management. – N.Y., USA: ACM Press, 2003. – P. 461-468.

14. Sparck, J. Document Retrieval: Shallow Data, Deep Theories, Historical Reflections, Potential Directions / J. Sparck // 25th European Conf. on IR Research. – Pisa, Italy: Springer Verlag, 2003. – V. 2633, № 77. – P. 1-11.

15. Тертышный В.А. Модель специализированной системы поиска сущностей на основе связанных данных / В.А. Тертышный // Вісник Кременчуцького національного університету імені Михайла Остроградського. – Кременчук, 2014. – Вип. 5/2014(88). – С. 112.

Поступила в редколлегию 6.07.2015

Рецензент: д-р техн. наук, проф. А.А. Можаев, Национальный технический университет «ХПИ», Харьков.

РОЗРОБКА МОДЕЛІ ІНФОРМАЦІЙНОГО ПОШУКУ З ВИКОРИСТАННЯМ ЗВ'ЯЗАНИХ ДАНИХ

В.В. Артамонов, В.О. Тертишний

У статті пропонується модель пошуку, заснована на використанні пов'язаних даних для побудови архітектури системи і вибору підходу до пошуку, в основі якого лежить пошук сутностей і формування фактів. Даний підхід дозволяє формувати стратегію пошуку для певних спеціалізованих предметних областей. Представлена архітектура пошукової системи та практичні аспекти реалізації. Модель може бути реалізована в корпоративних системах документообігу.

***Ключові слова**: пошук, сутність, семантичний простір, пов'язані дані, RDF.*

DEVELOPMENT OF INFORMATION SEARCH MODEL WITH THE USE OF LINKED DATA

V.V. Artamonov, V.O. Tertyshnyi

The article discusses the development of a search model based on the use of related data to build the system architecture and the choice of approach to the search based on the search for essences. This approach will allow to choose a search strategy for certain specialized subject areas. The article also discusses the differences of existing models and their advantages and disadvantages. The developed model can get in the implementation in corporate document management systems.

***Keywords**: search, entity, semantic space related data, RDF.*