

**МОДЕЛЬ СПЕЦИАЛИЗИРОВАННОЙ СИСТЕМЫ ПОИСКА СУЩНОСТЕЙ
НА ОСНОВЕ СВЯЗАННЫХ ДАННЫХ****В. А. Тертышный**

Кременчугский национальный университет имени Михаила Остроградского
ул. Первомайская, 20, г. Кременчуг, 39600, Украина. E-mail: mainhousepost@hotmail.com

Рассматриваются основы модели системы и методики семантического поиска с использованием связанных данных, позволяющих применять формальные подходы для построения семантического пространства. Отличием данного подхода от известных подходов (тезаурусы, онтологии, таксономии) является унификация и гибкость для проектирования и интеграции поисковых систем в общей архитектуре корпоративных систем. При этом существует возможность использования структурного распределения данных внутри общего информационного пространства. Метод поиска базируется на обработке (индексации и итеративном извлечении) набора RDF триплетов ориентированного графа, в котором отображается связь сущностей семантического пространства и их иерархическое соотношение. Также описывается метод ранжирования индексированных данных, обработка запросов и представление поискового результата.

Ключевые слова: поиск, сущность, семантическое пространство, RDF-триплеты, граф.

МОДЕЛЬ СПЕЦІАЛІЗОВАНОЇ СИСТЕМИ ПОШУКУ СУТНОСТЕЙ НА ОСНОВІ ЗВ'ЯЗАНИХ ДАНИХ**В. О. Тертышный**

Кременчуцький національний університет імені Михайла Остроградського
вул. Першотравнева, 20, м. Кременчук, 39600, Україна. E-mail: mainhousepost@hotmail.com

Розглядаються основи моделі системи і методики семантичного пошуку з використанням зв'язаних даних, що дозволяють застосовувати формальні підходи для побудови семантичного простору. Відмінністю даного підходу від відомих підходів (тезауруси, онтології, таксономії) є уніфікація і гнучкість для проектування та інтеграції пошукових систем в загальній архітектурі корпоративних систем. При цьому існує можливість використання структурного розподілу даних у середині загального інформаційного простору. Метод пошуку базується на обробці (індексції та ітеративному отриманню) набору RDF триплетів орієнтованого графа в якому відображається зв'язок сутностей семантичного простору та їх ієрархічне співвідношення. Також описується метод ранжирування індексованих даних, обробка запитів та подання пошукового результату.

Ключові слова: пошук, сутність, семантичний простір, RDF-триплет, граф.

АКТУАЛЬНОСТЬ РАБОТЫ. На сегодняшний день поиск сущностей представляет собой набор методов и алгоритмов анализа различного рода наборов данных в виде информационных объектов в контексте определенной тематики. К подобным объектам относятся как различного рода документы, так и базы и хранилища данных, которые наиболее приближены по своим свойствам и описанию к объектам и понятиям реального мира на основе запроса пользователя. Зачастую конечный пользователь системы использует неточные, неформальные запросы по ключевым словам. В последнее время возник интерес к семантическому поиску на основе структурированных данных.

Для конечного пользователя использование семантического поиска может оказаться непосильной задачей. Потому что основные поисковые системы на ориентированы на пользователя, который предпочитает простые, быстрые, неформальные запросы и при этом, он не обязан знать структуру данных и необходимую архитектуру запроса. Поисковые системы на данном уровне представляют собой гибридные системы, состоящие из поисковых алгоритмов полнотекстовых запросов и обработки объектов, имеющих описательную структуру и их связей. В свою очередь наборы данных может иметь при этом высокую степень гетерогенности [1, 2], что характерно, прежде всего, для корпоративных поисковых систем. Подобного рода системы представляют собой программно-аппаратные комплексы, имеющие высокоуровневую архитектуру.

Наиболее известными представителями таких систем являются Google Enterprise Search компаниями Google, Oracle Enterprise Search Engine, The Microsoft Business Productivity Platform. Именно, на уровне корпоративных поисковых систем (КПС) можно использовать необходимый уровень формализации, где структура данных имеет предсказуемо однородный и относительно целостный характер, не смотря на возможную логическую и физическую распределенность. Подобные системы могут интегрироваться в различные комплексы ERP-, EDI-систем и обеспечивать человеко-машинный интерфейс для доступа к содержимому хранилищ данных и систем аналитической обработки. Причем интеграция возможна также на более низком локальном уровне, например, поиск по источникам, данным на основе различных разметок (сайты, хранилища). Главным аспектом использования таких систем является возможность пользователя создавать некоторый набор программных модулей, используемых ресурсами предприятия для взаимодействия с базой поисковой системы.

Вопрос создания адаптивных методов и моделей поиска к условиям распределенности и неоднородности источников данных внутри информационного пространства определенного предприятия является актуальной и востребованной задачей, особенно в условиях структурной неопределенности.

Структурная неопределенность (СН) представляет собой меру распределенности структурных единиц или рассредоточения источников данных предприятия. Рассредоточение данных относительно определенных структурных и административных единиц предоставляет возможность создания иерархически организованных сущностей. В связи с чем, целью данной работы является создание математической модели системы корпоративного поиска, обеспечивающей высокий уровень адаптации к распределенным источникам в условиях их структурной неопределенности. Одним из средств устранения подобного рода неопределенности является применения интеллектуальных методов и моделей концепции связанных данных (англ. Linked Data). Использование связанных данных в этом случае обеспечивает базис для разработки методов извлечения и поиска сущностей по их связям и иерархиям объектов на основе онтологий предметной области.

Цель работы – создание модели специализированной системы поиска сущностей на основе связанных данных.

МАТЕРИАЛЫ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ.

В данной работе рассматривается описание использования сущностного поиска на основе связанных данных. Необходимо отметить, что существенным недостатком подобных КПС является высокая стоимость и необходимость определенного уровня знаний структуры данных и понимания основных принципов архитектуры программного комплекса. Поэтому реализация альтернативных моделей систем является актуальной и необходимой [1, 3]. В этом случае связанные данные могут послужить хорошим подспорьем для решения задач структуризации и иерархического объединения ресурсов в условиях СН в едином информационном пространстве предприятия. Сам программный комплекс может быть разработан на основе определенной модели сконфигурированной под потребности клиента и включать в себя различные функциональные модули и возможности. А именно:

- использование вопрос-ответных интерфейсов, уточняющих и конкретизирующих запросы к системе;
- учет синонимичности терминов;
- автодополнение;
- полнотекстовый поиск;
- индексирование и кластеризация документов и т.п.

Разработка любой системы поиска основывается на специальной модели. Одним из самых важных вопросов в реализации поисковиков является определение релевантности для выбранных информационных объектов относительно запроса. При этом любая система поиска должна выполнять три базовых функции: формирование запроса, индексирование, сравнение. В общем случае модель состоит из следующих компонентов:

$$M = [L, Q, F, R(q_i, d_j)], \quad (1)$$

где L – множество логических представлений документов коллекции; Q – множество логических

представлений информационных запросов, определяющих потребности пользователя; F – платформа моделирования представлений документов, запросов и отношения между ними [4]; $R(q_i, d_j)$ – функция ранжирования, задающая вещественное число для запроса q_i и представления документа d_j – функция близости. Данное ранжирование определяет степень (порядок) соответствия документов данному информационному запросу q_i .

Для воплощения модели используем концептуализацию схемы данных предприятия в виде онтологий на основе связанных данных. Концептуализация схемы представляет собой формальное описание сущностных связей между объектами с помощью ключевых особенностей предметной области, в которой она используется. Пример таких особенностей является построение иерархии структурных узлов системы или предприятия, юридические и бюрократические требования, особенности хранения и доступа информации. Главной целью использования концептуализации в этом случае является объединение сущностей в системе с учетом структурной неопределенности. Данный процесс называется «юнионизация» ресурсов (англ. Union of Resources). Инструментом для реализации этого механизма является использование моделей связанных данных.

С помощью объединенных объектов на основе концептуальной схемы и аннотации предоставляется возможность получать и аккумулировать факты в поисковые сущности и делать логические выводы для построения поисковых результатов. Факт, в данном случае, представляет собой атрибут, который описывает свойство самого объекта или указывает на связь со свойствами других объектов.

Реализовать подобную модель и построить информационную систему возможно с помощью семантической разметки, удобной для машинной обработки. В настоящее время семантическая разметка представлена целым рядом словарей (Schema.org, OpenGraph, FOAF) и спецификаций синтаксисов (RDFa, Microdata). В основе создания системы используем построение концептуальной схемы на основе RDF и используем язык запросов SPARQL. При этом необходимо учитывать ряд особенностей данной технологии. В SPARQL нет встроенного языка запросов для полнотекстового поиска литералов, ее функциональность обычно передается с помощью «магических предикатов», которые являются особенностью хранилищ триплетов. SPARQL 1.0 позволяет сопоставление с помощью регулярных выражений, но оно, как правило, не поддерживается ИР поисковиками. В целом, хранилища триплетов не реализуют процесс ранжирования [5].

Известна модель, где набор обратных индексов используется для сопоставления ограниченных форм конъюнктивных запросов, в частности, древовидных запросов с единственной рабочей переменной в корне дерева. Такие запросы могут включать «базовые концепты», т.е. набор ресурсов,

которые имеют предикат-значение (англ. Predicate-Value) [5], содержащий заданный набор ключевых слов. Отмечено, что преобразование таких запросов может быть более эффективным, чем комбинация хранилища триплета и полнотекстового индекса на стороне сервера. Такие запросы предлагают простой алгоритм распространения для передачи оценки сопоставления ключевых слов назад к корневому узлу дерева значений, тем самым получая ранжирование результатов. Все вышеперечисленные модели предназначены для продвинутого пользователя, который знаком со структурой данных и способен обозначить информацию, в которой он нуждается с помощью структурированного запроса, т.е. предоставив графовые шаблоны для сопоставления [6].

Рассмотрим модель с четырех основных сторон, а именно со стороны способа представления запроса к поисковой системе; метода поиска по распределенным источникам; ранжирование результатов; формы поисковой выдачи запроса.

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 (характеристика 2 <5)) OR (характеристика1 = false AND (характеристика3 > 7)).

Одним из наиболее развитых механизмов уточнения запросов является использования служебных слов AND и OR.

Сбор данных в распределенных источниках представим в работе графом отображающим модель целостного образа информационного пространства предприятия, в котором узлы графа соответствуют объектам (источникам данных), а дуги – отношениям между этими объектами. Формальное такую сеть можно описать в виде:

$$H = \langle I, C, G \rangle, \quad (2)$$

где I – множество информационных единиц, C – множество типов связей, G – отображение, задающее конкретное отношения из имеющихся типов C между элементами I .

3. Для выдачи ранжированного результата используем метод [7], в основе которого лежит использование двух функции ранжирования данных по запросам. В системе они представлены объектами – агентами обработки (англ. Processing Agents, PA).

Первый объект представим функцией вычисления входных параметров на основе итерационной обработки и установление связей объекта согласно аннотационных правил. Второй объект представим, как функцию вычисления множеств I и G и направлен на обработку связанных источников данных (документов, объектов), ранжирование которых устанавливает связями I и G .

Ранжирование документа в I производится с помощью ID -функции обратной инверсной частоты

документа, которая в классической интерпретации имеет вид:

$$IDF(q_i) = \log \frac{N}{(q_i)}, \quad (3)$$

где N – общее количество документов, q_i – количество документов, которые содержат q_i .

Конечным результатом работы метода является документ или объект, который максимально соответствует запросу пользователя и дополнительные связи с другими сущностями. Во время каждой итеративной обработки коллекции документов по атрибутам происходит ранжирование документа по набору фактов документа. В данном случае документ представляет собой часть объединенного узла концептуальной схемы (графа). В физическом виде представлен гипертекстовым документом с RDF разметкой. Сам документ также содержит сущности: документы (файлы), изображения, людей и т.д. Каждая из сущностей имеет аннотации сущности (атрибуты), которые составляют основу сериализации матрицы фактов (имя, связи, описание, ссылки на ресур т.п.).

Отдельно от PA используется аккумулятивный граф (АГ). Задача АГ – сбор накопленных на основе запросов данных для анализа формальных схем, связей и особенностей предметной области для дальнейшей концептуализации.

Для построения схемы следующим был создан граф отношений (ГО), главным назначением которого является описание связей графов системы со структурной неопределенностью, отображения аннотаций сущностей графов.

Поэтому, сущности системы можно классифицировать по виду назначения и структуре.

Следует отметить, что АГ выполняет назначения обработки, вычисления и анализа и также имеет функции метаданных концептуальных схем системы и документирования.

4. Форма поисковой выдачи представляет собой набор связанных данных модели RDF. Следует отметить, что существует отдельная возможность ранжирования данных по модели RDF и адаптация уже известных функций ранжирования.

Для практической реализации в работе была использована графовая база данных Neo4j, которая в полной мере позволяет реализовать необходимую модель. С помощью данной БД мы можем получить набор данных по запросу на основе модели RDF с использованием языка запросов SPARQL и подобных. Сама БД представляет собой ориентированный граф состоящий из узлов и ребер. Также в neo4j имеет следующее структуры и функции:

Node (граф, узел, нода) – объект в базе данных, узел графа. Количество узлов ограничено двумя в степени 35 ~ 34 биллиона

Nodelabel (метка графа) – используется как условный «тип узла». Например, узлы определенного типа могут быть связаны с узлами иного типа (например фильм-актер).

Relation (связь) – связь между двумя узлами, ребро графа. Количество связей ограничено 2 в степени 35 ~ 34 биллиона.

Relationidentifier (тип связи). Максимальное количество типов связей 32767.

Properties (свойства узла) – набор данных, которые можно назначить ноде. Например, если узел — это товар, то в свойствах узла можно хранить id документа из базы MySQL или любого другого источника.

Node ID (ID узла) – уникальный идентификатор узла (графа).

В таком случае работая с данной БД «поисковая фраза» может иметь следующий комплексный логический вид:

Сущности, в которых есть ссылки на документы, которые относятся к статистике за 2012 год либо сущности в которых есть ссылки на документы, которые не относятся к статистике за май 2012 года.

На рис. 1 представлен пример графовой базы данных, где вершинами графа являются сущности.

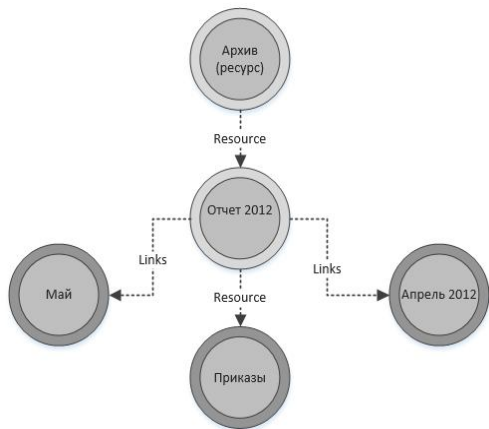


Рисунок 1 – Пример построение концептуальной схемы соотношения сущностей на основе графов

Для реализации семантического поиска предлагается использовать модель семантического пространства, в которой осуществляется поиск объектов с учетом их структурных связей и общей меры энтропии внутри получившийся развернутой в результате семантической сети [7].

Само семантическое пространство представляет набор и классификацию множества объектов и категорий между которыми в системе координатных осей производится вычисление расстояний между объектами.

Модель информационной системы на основе семантического пространства можно описать таким образом:

$$M_{sm} = \langle O, P(q_i, d_i), L \rangle, \quad (4)$$

где O – множество объектов семантического пространства, включающего в себя подмножества сущностей различной типизации; $P(q_i, d_i)$ – процедура определения релевантности, где q_i – представление поискового запроса, d_i – процедура агрегирования и компарирования сущностей; L –

множество связей включающих в себя подмножества связей различной типизации.

Подобную поисковую систему можно представить в виде трех направляющих на осях координат.

Координата нативности (N_i) – множество терминов описывающих структуру сущностей (объектов).

Координата связей (L_i) – множество терминов описывающих иерархическую и структурную взаимосвязь объектов имеющих весовые коэффициенты.

Статическая модель семантического пространства может иметь вид графа с помощью которого можно отобразить семантическую близость.

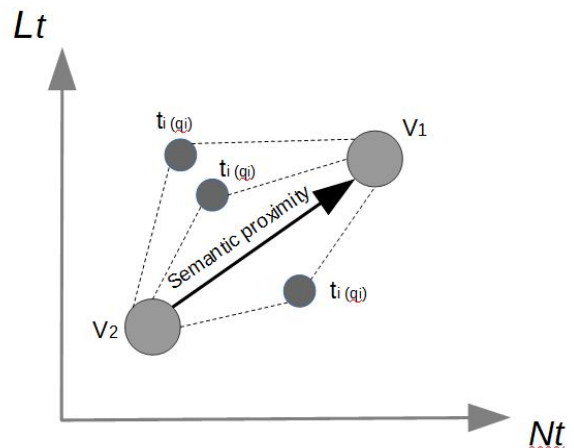


Рисунок 2 – Схема отображения связей семантической близости сущностей

Терминология задачи по поиску сущностей:

– характеристика – определенное свойство сущности. Например, объем документа.

– шаблон сущности – набор всех возможных характеристик однотипных сущностей, например – перечень возможных характеристик документа. При добавлении нового документа администратор может выбирать характеристики в рамках шаблона.

Добавить новую характеристику для одного документа не является рациональным способом [8] осуществления сущностного соотношения. Для этого нужно добавить характеристику в шаблон для этого документа. Одновременно эта характеристика будет доступна для всех товаров, использующих этот шаблон

– группа сущностей – сущности на основе одного шаблона. Например, файлы xml. Поиск и ранжирование может производиться только для документов и объектов из одной группы

– критерий – логическое правило, которое состоит из набора формальных требований к характеристикам документа.

Представление запроса к поисковой системе можно представить в виде набора терминов объединенных логическими операторами. По аналогии с моделью ключевых слов (тегов), когда запрос имеет вид $Q = a \vee b$, запрос в

семантичному пошуку можна представити в формі набору триплетів [8]:

$$Q = (p_1(s_1, o_1)k_1) + (p_i(s_i, o_i)k_i), \quad (5)$$

де $(p_1(s_1, o_1)k_1)$ – кортеж атрибута сутності, який відповідає триплету RDF, k_i – ваговий коефіцієнт атрибута сутності.

В нашому випадку експлікація кожного об'єкта семантичного простору буде представляти набір множини триплетів разом з ваговими коефіцієнтами атрибутів вибраних з загального проіндексованого масиву системою.

$$J(O) = (\sum_{i=1}^n p_i(s_i, o_i)k_i). \quad (6)$$

Саму модель семантичного простору можна представити в формі орієнтованого графа на основі моделі RDF (рис.3).

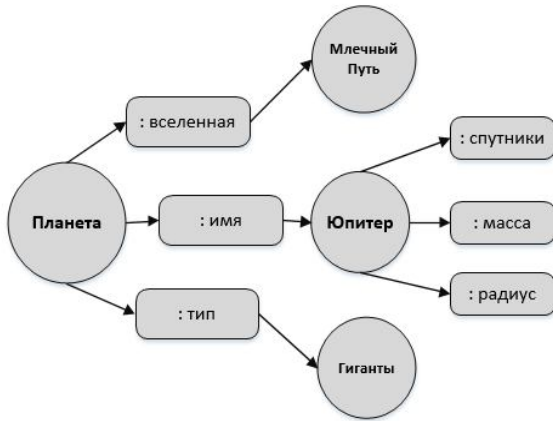


Рисунок 3 – Схематичне представлення семантичного простору

В першу чергу необхідно врахувати симетричні та інверсні відношення між вершинами графа [6]. Тоді концептуальну близькість між елементами α і ω можна виразити:

$$S(\alpha, \omega) = \max_{i=1 \rightarrow k} (S_p(\alpha, \omega)), \quad (7)$$

де k – число можливих шляхів графа від вершини α до вершини ω . Де під шляхом $P(\alpha, \omega)$ розуміється набір ребер (предикатів) ведучих від вершини α до вершини ω , з урахуванням їх напрямленості.

В цьому випадку величину шляху $P(\alpha, \omega)$ можна вважати відстанню вираження семантичної близькості між сутностями, яке має вигляд:

$$S(P(\alpha, \omega)) = \frac{P(\alpha, \omega)}{D}, \quad (8)$$

де D – глибина графа всієї топології існуючих об'єктів графа

$$D = N_1 + 2N_2, \quad (9)$$

де N_1 – кількість вершин шляху від сутності α до сутності ω , N_2 – кількість вершин до коренної сутності.

З урахуванням вищевказаних особливостей даного підходу опишемо метрику з урахуванням предметної області, в якій існує набір понять S_i зв'язаних між собою відношеннями R_i .

Вершинами V_i є сутності, а ребра L_i є семантичними відношеннями.

Маючи запит Q , проіндексованою множини RDF-триплетів T_i , представляючих собою кортеж виду $\langle S_i, R_i, O_i \rangle$, де S_i – включення в об'єднання понять і екземплярів понять, O_i – об'єднання понять і екземплярів понять з урахуванням відношень вагових коефіцієнтів.

На основі представленої метрики метод пошуку сутностей можна представити в формі набору зв'язаних між собою етапів:

- створення наборів класів на основі зв'язів з поняттями;
- вибір та вилучення з проіндексованого триплетного набору триплетів зв'язаних даних, кожен з елементів структури яких (суб'єкт, предикат, об'єкт) мають зв'язи в представленні входження в поняття і їх екземпляри);
- ранжування графів і триплетів всередині графів, елементи яких мають термінологічні мітки з певним ваговим коефіцієнтом.

Результатом програмної реалізації даного методу є набір даних про документ, виду $\langle U_i, S_i \rangle$, де U_i – ідентифікатор ресурсу (англ. URI), S_i – індекс значимості елемента відносно сформульованого запиту і має вигляд $\langle Y_i, K_i T_i, K_i G \rangle$.

Y – коефіцієнт поправки для даного запиту, $K_i T_i$ – коефіцієнт значимості зворотної інверсії, проіндексованого набору триплетів, $K_i G$ – коефіцієнт ієрархічної значимості всередині графа в відношенні до інших елементів графа.

ВИВІДИ. В даній роботі розглянуті базові аспекти використання технологій зв'язаних даних в контексті семантичного простору.

Отримано подальше розв'язання методу корпоративного пошуку сутностей на основі зв'язаних даних, що дозволяє підвищувати ефективність аналізу семантичного простору підприємства за рахунок налаштування гнучких шаблонів сутностей на основі нечітких критеріїв.

Представлено математичну модель системи корпоративного пошуку, яка дозволяє в порівнянні з існуючими, адаптуватися до розподіленої структури корпоративних джерел підприємства і забезпечувати цілеспрямований пошук зв'язаних даних в умовах структурної неопределенності.

Представлено метрики схожості об'єктів і формалізований процес ранжування і формування пошукового запиту до системи.

ЛИТЕРАТУРА

1. Ефимова Т.В. Моделирование семантичного простору нарративного тексту в статичній і динамічній // Вестник ВГУ. – 2012. – № 1. – С. 69–74.
2. Развитие информационной системы организации с использованием семантических технологий / А.В. Черный, А.Ф. Тузовский // Матер. всерос. конф. с междунар. участием «Знания–Онтология–Теория». – Новосибирск: ЗАО «РИЦ Прайс-Курьер», 2009. –Т. 2. – С. 52–59.
3. Губин М.Ю., Разин В.В., Тузовский А.Ф. Методы создания семантических метаописаний документов с применением семантических сетей, фреймворков моделей и частотных характеристик // Доклады Томского

государственного университета систем управления и радиоэлектроники. – Томск, 2010. – Вып. 2. – С. 227–229.

4. TaoCheng, KevinChen-ChuanChang (2007) “Entity Search Engine: Towards Agile Best-Effort Information Integration over the Web”. *CIDR*. – PP. 108–113.

5. Rio Blanco, Peter Milka, SebatianoVigna (2011) “Effective and Efficient Entity Search in RDF data”. *The Semantic Web–ISWC* – Springer. – 92 p.

6. Семантико-ориентированные системы на основе баз знаний / И.П. Кузнецов, А.Г. Мацкевич. – М.: Связьиздат, 2007. – 173 с.

7. Некрашевич С.П., Божко Д.В. Представление данных в Интернет на основе семантических сетей. // Искусственный интеллект. – Киев, 2006. – Вып. 1. – С. 57–59.

8. Ле Хоай, Тузовский А.Ф. Разработка семантических электронных библиотек на основе онтологических моделей // Сборник докладов XV Всероссийской научной конференции RCDL, 15–18 октября 2013 г. Переславль-Залесский, 2013. – С. 291–299.

9. Марчук А.Г. На пути к большим RDF данным // Сборник докладов XV Всероссийской научной конференции RCDL, 15–18 октября 2013 г. Переславль-Залесский, 2013. – 2013. – С. 51–56.

10. Ландэ Д.В. Эффективный поиск знаний в интернет. Профессиональная работа. – М.: Издательский дом “Вильямс”, 2005. – 270 с.

11. Нгуен Ба Нгок, Тузовский А.Ф. Модель информационного поиска на основе семантических метаописаний // Управление большими системами. Вып. 41. – М.: ИПУ РАН, 2013. – С. 51–92.

12. Roi Blanco, Harry Halpin, Daniel M Herzig, Peter Mika, Jeffrey Pound, Henry S Thompson, T Tran Duc (2011) “Entity search evaluation over structured web data”. *Proc. of the 1st International Workshop on Entity-Oriented Search at SIGIR*. – PP. 101–103.

13. Скрэг Г. Семантические сети как модели памяти // НЗЛ. – М.: Прогресс, 1999. – Т. III. – С. 259–302.

14. Okkyung Choi, SeokHyun Yoon, Myeongeun Oh, Sangyong Han (2003). *Semantic Web Search Model for Information Retrieval of the Semantic Data*. Verlag Berlin Heidelberg. – Springer. – PP. 588–593.

15. Цыганов Н.Л., Циканин М.А. Исследование методов поиска дубликатов веб-документов с учетом запроса пользователя // Интернет-математика 2007: сбор. работ участников конкурса. – Екатеринбург: Изд-во Урал. ун-та, 2007. – С. 211–222.

SPECIALIZED MODEL OF ENTITY SEARCH BASED ON LINKED DATA APPROACH

V. Tertyshnyi

Kremenchuk Mykhailo Ostrohradskyi National University

vul . Pershotravneva, 20, Kremenchuk, 39600, Ukraine. E-mail: mainhousepost@hotmail.com

The article covers the basics of the model, system and methods of semantic search using linked data allowing applying formal approaches for semantic space creating. Differences between this approach and the other known approaches (thesauri, ontologies, taxonomies) are unification and flexibility for the design and integration of search engines in the overall architecture of enterprise systems. In this case, it is possible to use the structural distribution of data within a common information space. Search method is based on processing (indexing and iterative retrieving) of RDF-triples set in a directed graph which shows the relationship between semantic space entities and their hierarchical relationship. Also it describes a method for ranking the indexed data, query processing and search results representation.

Keywords: search, entity, semantic space, RDF-triples, graph.

REFERENCES

1. Efimova, T.V. (2012) “Modelling semantic space of narrative text in statics and dynamics”, *Transaction VGU*, no. 1, pp. 69–74.

2. Chorniy, A.V., Tuzovskiy, A.F. (2009) “The development of the information system of the organization, with the use of semantic technologies”, *Mater. All-Russia. Conf. with int. participation “Knowledge Ontology-Theory”*, ZAO “RIP Price Courier”, Novosibirsk, vol. 2, pp. 52–59.

3. Gubin, M.Y., Razin, V.V., Tuzovskiy, A.F. (2010) “Methods for creating semantic metadescriptions documents using semantic networks, frame model and frequency characteristics”, *Reports of Tomsk State University of Control Systems and Radio Electronics*, Tomsk, iss. 2, pp. 227–229.

4. TaoCheng, KevinChen-ChuanChang (2007) “Entity Search Engine: Towards Agile Best-Effort Information Integration over the Web”, *CIDR*, pp. 108–113.

5. Rio Blanco, Peter Milka, SebatianoVigna (2011) “Effective and Efficient Entity Search in RDF data”, *The Semantic Web–ISWC*, Springer.

6. Kuznetsov, I.P., Matskevich, A.G. (2007) *Semantiko-orientirovannye sistemy na osnove baz znaniy* [Semantic-oriented systems based on knowledge bases] Svyazizdat, Moscow, Russia.

7. Nekrashevich, S.P., Bozhko, D.V. (2006) Presentation of data in the Internet based on semantic networks, *Artificial Intelligence*, Kiev, iss. 1, pp. 57–59.

8. LeChoay, Tuzovskiy, A.F. (2013) “Development of semantic digital libraries based on ontological models”, *Proceedings of the XV Scientific Conference RCDL*, October 15–18,

Pereslavl, pp. 291–299.

9. Marchuk, A.G. (2013) “On the way to a big RDF data”, *Proceedings of the XV Scientific Conference RCDL*, October, 15–18, pp. 51–56.

10. Lande, D.V. (2005) *Effektivnyiy poisk znaniy v internet. Professionalnaya rabota* [Efficient search of knowledge on the Internet. Professional work], Izdatelskiy Dom “Williams”, Moscow, Russia.

11. Nguen Ba Ngok, Tuzovskiy A.F. (2013) “Model of information retrieval based on semantic meta descriptions”, *Managing pain-Shimi systems*, vol. 41, pp. 51–92.

12. Roi, Blanco, Harry, Halpin, Daniel, M. Herzig, Peter, Mika, Jeffrey, Pound, Henry, S. Thompson, T. Tran Duc (2011) “Entity search evaluation over structured web data”. *Proc. of the 1st International Workshop on Entity-Oriented Search at SIGIR*. – pp. 101–103.

13. Skrag, G. (1999) *Semanticheskie seti kak modeli pamyati* [Semantic networks as a model of memory], Progress, Moscow, vol. III, pp. 259–302.

14. Okkyung Choi, SeokHyun Yoon, Myeongeun Oh, Sangyong Han (2003) *Semantic Web Search Model for Information Retrieval of the Semantic Data*. Verlag Berlin Heidelberg. – Springer – pp. 588–593.

15. Tsyganov, N.L., Tsykanin, M.A. (2007) “Issledovaniye metodov poiska dublikatov web-dokumentov s uchetom zaprosa polzovatelya”, *Internet-matematika*, Izdatelstvo Ural.univ. Ekaterinburg, pp. 211–222.

Стаття надійшла 18.09.2014.