

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дискретной математики и информационных технологий

**РАЗРАБОТКА ОТНОЛОГИЧЕСКОЙ МОДЕЛИ ДЛЯ ОЦЕНКИ
ИКТ-КОМПАНИЙ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Абрамова Олега Александровича

Научный руководитель

доцент, к. э. н.

Г. Ю. Чернышова

Заведующий кафедрой

доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2020

ВВЕДЕНИЕ

В последние несколько лет прямое влияние интернет-технологий на все сферы применения вычислительной техники вызвало к жизни новую тенденцию, заключающуюся в переносе акцента с традиционных методов представления информации на представление информации в форме, которая удобна для ее использования в среде Интернет. Основная идея бурно развивающейся технологии Интернет – построение информационной системы на базе web-протоколов. Многие разработчики коммерческих систем управления базой данных уже включили и активно развивают поддержку импорта и экспорта данных с использованием технологии XML.

Данная тенденция затронула и сферу технологий искусственного интеллекта. Развитие технологии RDF (Resource Description Framework) на базе XML, позволило существенно упростить взаимодействие интересов специалистов и объединить специфические формы и форматы представлений объектов и процессов в таких предметных областях, как представление знаний, обработка естественно-языковой информации, принятие решений, экспертные системы и других решений на базе интернет-технологий. Как показывает практика, сближение с интернет-технологиями практически любой научной сферы существенно ускоряет темпы ее развития, что подтверждается, в частности, наблюдаемой в данный момент вспышкой интереса к проблеме представления знаний. Одной из самых последних разработок в этой области, базирующейся на направлении Semantic Web является язык web-онтологий OWL(Ontology Web Language).

Возникает необходимость представления знаний в формах, которые не только естественны для восприятия людьми, но и могут быть использованы компьютером. Один из наиболее перспективных путей решения этой проблемы – использование в информационных системах баз знаний, основанных на онтологиях. Онтология представляет собой формальную концептуальную модель приложения, разрабатываемого экспертами с использованием соответствующих формальных языков.

Целью работы является разработка web-приложения для анализа деятельности ИКТ-компании с помощью онтологической базы знаний.

Задачи представляют собой:

1. анализ существующих подходов к построению онтологических моделей;

2. выбор инструментальных средств для реализации приложения на основе онтологической модели;
3. формирование онтологической модели ИКТ-организаций;
4. разработка приложения для извлечения и анализа данных об ИКТ-организациях;
5. реализация сложных запросов с помощью разработанного приложения.

В качестве методологических основ использовались работы Кафтанниковой И.Л., Коровиной С.Е., Матюшиной М.М., Вакурина Т.Г., Гуськова А., Chantrapornchai Ch., Chidchanok Ch., Тузовского А.Ф., Ямпольского В.З., электронная документация по программным продуктам.

Теоретическая значимость работы, заключается в апробация современных подходов к проектированию и разработки онтологических моделей для анализа предметной области.

Практическая значимость работы, состоит в возможности применения разработанных онтологических моделей для решения прикладных задач, в частности для осуществления оценки деятельности ИКТ-компаний.

Структура и объём работы. Выпускная квалификационная работа состоит из введения, трех глав, заключения, списка использованных источников и шести приложений. Общий объем работы – 61 страница, из них 39 страниц – основное содержание, включая 9 рисунков и 1 таблицу, список использованных источников информации – 22 наименования.

Краткое содержание

В первом разделе выпускной квалификационной работы – «Обзор методов и средств построения онтологии» рассмотрены основные современные исследования в области онтологии, даётся определение онтологии с научной точки зрения, обзревается уже разработанные онтологии в различных сферах деятельности, рассматриваются основные методы построения онтологии, описывается абстрактная схема построения онтологии в общем случае, а также показываються различные уточнения для базовой схемы. Разработка онтологических моделей подходит при работе с предметными областями различной сложности и целевой направленности. Онтологические базы знаний могут быть применены при структуризации большого набора данных, что позволяет частично решить проблему нехватки памяти и скорости поиска в предметной области. Онтологии применимы ко многим задачам, которые связаны с бизнес-процессами, а так же обработкой естественных языков.

Изучение онтологии – это перспективное направление при построения логики и моделирования многих процессов. Многие информационные системы используют онтологии для структуризации данных и для упрощения взаимодействия с хранимой информацией. Для удобства визуализации и наглядного, понятного проектирования используются специальные конструкторы онтологических моделей. При проектировании онтологической модели при помощи конструктора используются те же подходы, что и при проектировании без конструктора. Конструктор онтологических моделей позволяет лишь наглядно показать как устроена онтологическая модель на каком либо этапе разработки этой онтологической модели. Существуют некоторые формы представления онтологической модели: в виде графа, в виде иерархической цепочки, в виде набора значений, а так же сохраняется возможность просмотра кода на языке RDF. Конструктор онтологической модели должен иметь функционал для того, что бы при проектировании онтологической модели конструктором можно было использовать те же подходы, что и проектировании онтологических моделей без конструктора.

Существуют различные инструменты взаимодействия с базой знаний. В зависимости от целей могут быть выбраны разные наборы инструментальных средств. Инструменты могут быть скомбинированы. При правильном проектировании должны использоваться соответствующие инструменты разработ-

ки. Неотъемлемой частью проектирования онтологической модели является построения сравнительного анализа инструментов. Во втором разделе выпускной квалификационной работы – «Разработка онтологической модели» описывается процесс формирования предметной области для построения онтологической базы знаний, выделяются основные критерии по которым производится анализ онтологии после её построения, рассматривается процесс разработки онтологии с использованием программы Protege.

При проектировании онтологической модели ИКТ-организаций важно понимать, что как и любые другие предметные области, предметная область набора ИКТ-организаций должна содержать ряд базовых значений для анализа предметной области. Набор таких базовых значений может содержать основные характеристики такие как: название организации, уникальный код организации, доход организации за интересующий промежуток времени, физический адрес главного офиса организации, набор основных конкурентов на рынке, статус компании на рынке, владелец организации.

Составленная онтологическая модель может быть использована при построении онтологической базы знаний путём добавления в предметную область экземпляров классов, которые соответствуют каждому классу исходя из поставленной задачи онтологической базы знаний.

В третьем разделе выпускной квалификационной работы – «Разработка системы управления онтологической моделью» описывается процесс разработки модуля для взаимодействия с онтологической базой знаний при помощи библиотеки Jena API, процесс разработки модуля для извлечения данных о ИКТ-организациях из открытых источников с использованием Dadata API, создание пользовательского интерфейса программы с использованием HTML и шаблонизатора Mustache, применяются методы анализа онтологической базы знаний с использованием SPARQL запросов, а так же применяется способ внедрения SPARQL запросов в приложение при помощи Jena API.

Разработка серверной части приложения начинается с основного модуля, который позволит взаимодействовать с файлом онтологической базы знаний на жёстком диске. Для этого можно использовать основной функционал Jena API, который позволяет создавать экземпляры классов и связывать их уже заготовленными отношениями.

Разработанная система позволит взаимодействовать с онтологической

базой знаний для добавления, чтения различных частей базы знаний. Так же разработанная система позволяет выполнять SPARQL как для запросы для анализа данных, так и для простого извлечения информации.

Для построения онтологической базы знаний необходимо брать информацию из отдельных источников. Эта информация нужна для заполнения базы объектами. Для извлечения данных из открытых источников необходим так же отдельный API. Для заполнения онтологической базы знаний информацией из открытых баз данных об организациях подходит Dadata API. Dadata API – это набор инструментов для поиска данных об организациях по ключевым данным. Такими данными могут быть как ИНН, так и само название компании.

Разработка пользовательского интерфейса необходима для реализации удобной работы пользователя приложения. При разработке HTML документа используются стандартные подходы к вёрстке страницы. Используются CSS стили и стандартные теги HTML. При обработке HTML документа Mustache ищет специальные метки для внедрения данных в HTML документ и внедряет данные. При использовании Spring Boot существует возможность настраивать добавления, которые внедряются в HTML документ. Эти настройки производятся в контроллерах. Контроллер – это Java-класс, помеченный специальной аннотацией Controller, который образует соотношение адресом ресурса и методом класса-контроллера.

Для анализа построенной онтологической базы знаний удобно использовать SPARQL запросы. SPARQL позволяет построить запрос по поиску RDF-троек таким образом, что бы можно было выявить какие либо аналитические сведения исходя из полученных результатов. При использовании SPARQL запросов знать понимать набор значений, которые интересны для разрабатываемого приложения.

SPARQL-запрос является средством анализа построенной базы знаний. Запрос выступает в качестве формулы, по которой вычисляется запрошенная информация о ИКТ-компании. Перечисленные характеристики можно вычислить в приложении, используя модуль, для взаимодействия с базой знаний, который содержит инструменты для обработки SPARQL запросов.

ЗАКЛЮЧЕНИЕ

Для разработки приложения анализа деятельности ИКТ-компании осуществлён анализ существующих подходов к построению онтологических моделей, в результате которого были выявлены основные этапы построения онтологических моделей для каждого подхода. При построении приложения для анализа деятельности ИКТ-компаний может быть использован один из подходов:

1. предоставление максимума информации при минимуме когнитивных усилий;
2. формирование классов на основе прототипов;
3. использование межязыковых соответствий.

Кроме анализа существующих подходов проведён обзор инструментальных средств, позволяющих разрабатывать онтологические модели и базы знаний любого уровня сложности. Рассмотрены недостатки и преимущества для каждого инструмента разработки онтологических моделей. В качестве инструментов для разработки приложения для анализа деятельности ИКТ-компаний использованы следующие технологии:

1. Protege – программа для взаимодействия с онтологической базой знаний с помощью графического интерфейса;
2. Jena API – библиотека, разработанная на языке Java для построения приложений для автоматической генерации данных в базе знаний;
3. SPARQL – язык запросов к базе знаний.

Для построения модели предметной области, связанной с анализом деятельности компании, выделены основные компоненты, которые и использованы для построения предметной области. Предложены основные компоненты для онтологической базы знаний.

На основе анализа предметной области разработана схема онтологической базы знаний в RDF-формате. Предложена структура классов для заполнения базы знаний объектами на основе сведений из открытых источников. Для реализации этого процесса использовался Dadata API.

Для извлечения данных по организациям разработан модуль заполнения базы знаний при помощи сервиса Dadata, который позволяет получить доступ к открытой базе данных об организациях. Данные о компаниях записываются в отдельные объекты для анализа организаций на рынке ИКТ-

технологий.

Для взаимодействия пользователя с серверной частью приложения разрабатывается пользовательский интерфейс на стороне клиента. Это необходимо при построении открытых приложений для анализа деятельности компаний. Это позволит разделить функционал пользователя и разработчика.

Для добавления и анализа данных разрабатывается система взаимодействия с базой знаний на базе инструмента Jena API. Для обращения к базе знаний создан модуль, который выступает в качестве промежуточного звена между инструментами Jena API и разрабатываемым серверным приложением. Это необходимо для удобства и обработки исключительных ситуаций, и удобства взаимодействия с базой знаний.

Из открытых источников данных об организациях извлекаются данные для помещения в онтологическую базу знаний. Необходимо набрать первоначальный запас данных для более точного анализа, поскольку изначальная база знаний пуста. При извлечении данных для первоначального анализа задействуются разработанные сервисы серверного приложения, которые базируются на инструменте для доступа к открытым данным об организациях – Dadata API. Данные, полученные при помощи Dadata API, не структурированы для дальнейшего анализа. Для этого данные помещаются в разработанную онтологическую базу знаний.

Для анализа деятельности компаний разработаны соответствующие SPARQL-запросы, которые позволяют проанализировать организации, например, выявить основных конкурентов компании. Также строятся запросы для извлечения общих сведений по компаниям из базы знаний.

Заполнение онтологической базы знаний позволяет структурировать данные, полученные из различных источников. При структурировании базы знаний производится подготовка данных к дальнейшему анализу. SPARQL-запрос позволяет составить аналитический запрос для получения необходимых знаний. Чем больше заполнена база знаний, тем меньше запросов необходимо делать к стороннему API, следовательно ускоряется работа приложения. Заполнение базы знаний происходит по мере поступления новых запросов от пользователей автоматически.

Разработанное приложение можно использовать для анализа рынка ИКТ-технологий. Выполнение запросов приводит к расширению базы знаний. Дан-

ную систему можно модифицировать путём добавления SPARQL-запросов. Запросы могут быть добавлены вместе с классами-процессорами, которые поочерёдно формируют выходные данные программы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Хоай, Л., Тузовский, А.Ф. Использование онтологий в электронных библиотеках / Л. Хоай. Томск: Томский политехнический университет, 2012.
- 2 Кафтанников, И.Л., Коровин, С.Е. Перспективы использования web-онтологий в учебном процессе / И.Л. Кафтанников. Челябинск: Южно-уральский государственный университет, 2014.
- 3 Матюшин, М.М., Вакурина, Т.Г. Методы и средства построения онтологий для визуализации связанных и информационных объектов произвольной природы в сложных информационно – аналитических системах / М.М. Матюшин. Энергия М.: 2015.
- 4 Гусарова, Н. Ф., Полянская, А. Д., Сысоева В. В. Технология построения низкоуровневой онтологии предметной области с применением концептов реинженеринга бизнес – процесса / Н. Ф. Гусарова. Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2015.
- 5 Chantrapornchai, Ch., Chidchanok, Ch., Ontology construction and application in practice case study of health tourism in Thailand. URL: <https://springerplus.springeropen.com/articles/10.1186/s40064-016-3747-3>. (Дата обращения: 21.03.2020).
- 6 Бёрнс, Т., Статья Linked Data Issues [Электронный ресурс] URL: <https://www.w3.org/DesignIssues/LinkedData.html>. (Дата обращения: 24.03.2020).
- 7 Анцибор, Д.В., Терновой, М.Ю., Шункевич, Д.В., Онтологический подход к проектированию биллинговых систем / Д.В. Анцибор. Киев: Киевский политехнический институт, 2013.
- 8 Юрин, А.Ю. WEB – сервис для автоматизированного формирования продукционных баз знаний на основе концептуальных моделей / А.Ю. Юрин. Иркутск: Институт динамики систем и теории управления, 2014.
- 9 Богуславский, И., Диконов, В., Тимошенко, С. Онтология для поддержки задач извлечения смысла из текста на естественном языке / И. Богуславский: ИППИ РАН, 2014.

- 10 Грибова, В.В., Петряева М.В. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решения. В.В. Грибова. СПб.: Институт автоматизации и процессов управления Дальневосточного отделения РАН, 2016.