

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.55

Якубенко  
Александр Игоревич

Модель сложноструктурированного сайта на основе  
документоориентированной базы данных

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники  
по специальности 1-40 81-03 Искусственный интеллект

Научный руководитель  
Колб Дмитрий Григорьевич  
кандидат технических наук

Минск 2015

## ВВЕДЕНИЕ

В настоящее время Интернет занимает огромное место в жизни человека. В отличие от Интернета «раннего» периода, в настоящий момент он не представлен просто сайтами, на которых размещена та или иная информация, а также ссылки на прочие ресурсы либо каталоги файлов. Это уже и не просто совокупность сервисов наподобие онлайн-переводчиков, электронных словарей или web-интерфейсов для электронных почтовых ящиков.

В современном Интернете web-сервисы начинают интегрироваться с «настольными» приложениями, перенимая типичные для них интерфейс, поведение и возможности. Среда Интернет позволяет также организовать многопользовательскую работу с документами, проектами, обмениваться различной информацией в едином сервисе с людьми, зарегистрированными в нем и разделенными большими расстояниями.

Одним из видов сервисов коллективной работы пользователей являются социальные сети. Социальная сеть – это совокупность социальных элементов (людей, групп, организаций, сообществ), которые обмениваются между собой различными данными. Такими данными могут быть как простые текстовые сообщения, так и изображения, видео- и аудиофайлы. Кроме того чаще всего социальная сеть хранит множество информации о действиях и предпочтениях пользователей, о личных характеристиках и биографических сведениях.

Таким образом, социальные сети аккумулируют огромные объемы информации.

Социальные сети стали новой ступенью в развитии не только коммуникаций между людьми, но и в сферах маркетинга, рекламы, торговли, экономики, государственного управления. Самое главное - они все больше влияют на решение острых социальных проблем человечества.

Социальная сеть облегчает создание персонального профиля и виртуальных взаимоотношений. Социальные сети используются для поиска людей со схожими интересами и объектов этих интересов. В социальных сетях обычно используется народная классификация, практика совместной категоризации информации.

На данный момент существует огромное количество социальных сетей. В связи с этим на первый план выходит не то насколько много данных может хранить социальная сеть и какие функции предоставляет, а то каким образом интерпретирует имеющиеся данные. В этой ситуации востребованными оказываются различные интеллектуальные алгоритмы семантического анализа информации. В частности алгоритмы искусственного интеллекта. Они позволяют генерировать множество более полезной информации по заранее определенным правилам.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Актуальность темы диссертации**

В настоящее время объемы данных в сети интернет стремительно растут. В данных условиях особенно остро становится проблема структуризации и унификации имеющейся информации, а так же эффективного поиска и навигации по ней. Документоориентированные базы данных позволяют удобно хранить имеющуюся информацию произвольной структуры. Для ее обработки, структуризации и классификации становится необходимым использование новых инновационных алгоритмов, основанных на теориях искусственного интеллекта.

### **Цель и задачи исследования**

Цель магистерской диссертации – разработать сложноструктурированный сайт на основе документоориентированной базы данных. Сайтом является социальная сеть. Разработать эффективный алгоритм поиска рекомендаций.

Работа над диссертацией включает в себя следующий список задач:

- Построение онтологии социальной сети;
- Проектирование модели сайта;
- Исследование технологий разработки веб-сайтов;
- Разработка сайта;
- Разработка эффективного алгоритма поиска рекомендаций;
- Исследование методов тестирования качества рекомендательных систем;
- Разработка системы для тестирования спроектированного алгоритма;
- Анализ качества рекомендательного алгоритма, а так же скорости его работы.

### **Объект и предмет исследования**

Объект исследования – алгоритмы интеллектуального анализа данных.

Предмет исследования – методы улучшения качества алгоритмов рекомендаций для использования в крупных социальных сетях.

### **Гипотеза**

В ходе исследования было выдвинуто предположение, что данные, характеризующие активность пользователя внутри социальной сети являются

основополагающими при поиске и ранжировании рекомендаций внутри социальной сети. Данное предположение было подтверждено результатами исследования.

### **Методология и методы проведенного исследования**

Большое количество простых рекомендательных алгоритмов различных сложно-структурированных сайтов работают, основываясь лишь на данных, явно введенных пользователем. В нашем случае, когда речь идет о социальных сетях, целесообразно использовать информацию, накопившуюся в ходе различного рода активностей пользователя внутри социальной сети.

Для тестирования качества и скорости работы рекомендательного алгоритма использованы наборы данных разной величины и плотности. Для анализа практической применимости и точности разработанной рекомендательной системы были использованы метрики Half-life Utility и Precision-N.

### **Научная новизна и значимость полученных результатов**

Научная новизна результатов заключается в абсолютно новом алгоритме поиска рекомендаций внутри социальной сети, основанном на идеях алгоритма «Дейкстры» поиска пути кратчайшего веса в графе.

Усовершенствование разработанного алгоритма рекомендаций возможно за счет более точного подбора значений приоритета для каждого из отношений в построенной онтологии. Эта возможность может являться отправной точкой для новых исследований в этой области.

### **Практическая (экономическая, социальная) значимость полученных результатов**

Разработанная социальная сеть может быть использована для безопасного общения и обмена информацией между пользователями по всему миру.

Алгоритм рекомендаций для разработанной системы может быть использован для сложноструктурированных сайтов на наборах данных произвольной величины и структуры.

### **Основные положения диссертации, выносимые на защиту**

1. Описана онтология социальной сети;
2. Предложен интеллектуальный алгоритм для поиска рекомендаций пользователю внутри социальной сети;
3. Построена и описана модель, разрабатываемого сайта;
4. Построенная модель реализована на практике;

5. Разработан программный продукт, позволяющий тестировать сайт на выборках данных различного объема;
6. Произведен анализ скорости и качества разработанной рекомендательной системы.

#### **Личный вклад соискателя**

Была разработана модель сложно структурированного сайта исходя из поставленных задач. По построенной модели был запрограммирован и развернут в сети интернет сайт, являющийся социальной сетью.

Был разработан алгоритм поиска рекомендаций для пользователя внутри социальной сети.

#### **Апробация результатов диссертации**

Доклад по теме работы был представлен на 50-й научной конференции БГУИР (2014 г.) в секции «Интеллектуальные информационные технологии».

#### **Структура и объем диссертации**

Общий объем магистерской диссертации составляет 71 страницу, включая 21 иллюстрацию, 4 таблицы, библиографический список из 54 наименований и приложения.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В главе 1 были рассмотрены основные понятия, связанные с социальными сетями, даны определения основным тезисам, проведен обзор существующих проблем. Поставлены цели и задачи исследования. Описаны основные функции разрабатываемой социальной сети.

Глава 2 посвящена практической реализации построенной модели на практике и последующего анализа полученных результатов. Широко рассмотрены все используемые технологии. Построена онтология предметной области. Описан спроектированный интеллектуальный алгоритм выбора рекомендаций для пользователя социальной сети. Проведены исследования скорости и качества разработанного алгоритма рекомендаций.

Библиотека БГУИР

## ЗАКЛЮЧЕНИЕ

В соответствии с поставленными задачами, построена модель сложно-структурированного сайта. Данный сайт представляет собой социальную сеть.

Информационную основу спроектированной социальной сети представляет онтология, с помощью которой осуществляется систематизация и структуризация информации, реализуется эффективный поиск и навигация по информационному пространству внутри социальной сети.

Рассмотрены функции портала, реализуемые на основании созданной онтологии.

По построенной модели исследованы и подобраны современные средства web-разработки необходимые для реализации спроектированной системы на практике.

Построенная модель сайта реализована на практике. Разработанный сайт развернут в сети интернет.

Спроектирован и реализован на практике алгоритм поиска рекомендаций внутри социальной сети.

Реализована платформа для конфигурирования и тестирования разработанного алгоритма рекомендации.

Таким образом, в ходе подготовки магистерской диссертации был получен полнофункциональный web-сайт, полностью готовый к применению. Данный сайт ориентирован на любого пользователя сети интернет. С его помощью пользователи могут общаться друг с другом а так же публиковать и исследовать интересующую информацию в виде статей. При размещении его в глобальной сети география распространения возрастает до масштабов всего мира.

Кроме того некоторые наработки и изменения существующих библиотек, внесенные в ходе программной реализации спроектированной системы годны для публикации в общественный доступ на сайте коллективной разработки программного кода «github» и как следствие – внедрения данных наработок в существующий код библиотек для их усовершенствования. В частности реализована поддержка работы библиотеки аутентификации «sorcery» для работы с графовой базой данных neo4j а так же работа библиотеки для быстрого построения web-форм «sorcery» с библиотекой «neo4j», являющейся ORM для neo4j.

В качестве дальнейшего совершенствования web-сайта представляется возможным разработка модулей предоставляющих возможности оставлять комментарий к статье либо на личной странице пользователя. Так же возможна доработка интерфейса сайта с целью дальнейшего повышения его информативности, привлекательности и удобства.

К минусам разработанной системы стоит отнести сложность разбиения найденных рекомендаций на страницы. Для каждой новой страницы необходимо либо пересчитывать все рекомендации предыдущих страниц, либо сериализовать текущее состояние алгоритма поиска рекомендаций и передавать это состояние между страницами сайта.

Библиотека БГУИР