

студентом и преподавателем. Преподаватель сможет выкладывать свои материалы для ознакомления студентами. В свою очередь, студенты смогут задавать вопросы по материалу, просто выделив непонятный участок мышкой. При этом выделенный участок помечается определённым цветом и открывается ветка обсуждения вопроса, в которой могут участвовать как студенты, так и преподаватели.

На основе этой системы будут построены и другие элементы системы:

- Преподаватель сможет оценивать хорошие вопросы для дальнейшего поощрения студента во время экзамена или зачёта, при чём преподаватель сможет либо отображать свою оценку студенту, либо скрыть её для всех, кроме себя.

- Преподаватель практических и лабораторных работ будет иметь возможность ведения ведомости оценок, которые играют роль при итоговых экзаменах. Такую ведомость смогут просматривать как студенты, так и преподаватель лекционного материала.

- Система фильтрации повторяющихся вопросов. Если при создании вопроса студентом такой же вопрос уже имеется в базе, то вопрос создан не будет, а студент перенаправляется на страницу обсуждения данного вопроса.

- Система разбиения лекций по датам. На основе уже имеющегося расписания занятий преподаватель сможет определять конкретную тему лекции к конкретному дню в расписании. При чём при выборе определённой даты будет открыт материал, который был прикреплен преподавателем.

С помощью вышеописанного веб-приложения планируется повысить эффективность подачи материала, наладить диалог между преподавателем и студентом. В конце концов, нужные вопросы всегда всплывают уже после лекции. Никогда не поздно задать их, сидя дома, перед этим правильно их сформулировать. В свою очередь, преподаватель имеет возможность анализировать непонятные моменты в своём материале и подстраивать его для более лёгкого усвоения.

Список использованных источников:

1. Данилова, Г. В. Программное средство управления формированием IT-компетенций / Г. В. Данилова // Дистанционное обучение – образовательная среда XXI века: материалы IX международной научно-методической конференции (Минск, 3-4 декабря 2015 года). – Минск: БГУИР, 2015. – С. 300 - 301.

2. Живицкая, Е. Н. Информационная система подготовки IT-специалистов / Е. Н. Живицкая, Г. В. Данилова // Информатизация образования. – 2017. – № 1 (79). – С. 54–72.

## МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ БАНКОВСКОЙ СФЕРЫ И ОБЛАСТИ БИЗНЕС-АНАЛИЗА

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Стасюк А.В.*

*Медведев С.А. – к.т.н., доцент*

Информация всегда играла чрезвычайно важную роль в жизни человека.

Общеизвестно высказывание о том, что тот, кто владеет информацией, тот владеет и миром. Иное сообщение стоит дороже жизни. По преданию, 13 сентября 490 года до н.э. греческий воин-гонец, прибежавший из Марафона в Афины, не останавливаясь в пути, упал замертво, но донес весть о победе над персами.

С давних времен сбор и систематизация сведений об окружающем мире помогали человеку выживать в нелегких условиях – из поколения в поколение передавался опыт и навыки изготовления орудий охоты и труда, создания одежды и лекарств. Информация постоянно обновлялась и дополнялась – каждое изученное явление позволяло перейти к чему-то новому, более сложному.

С течением времени роль информации в жизни человека становилась все существеннее. Нужно было изучать и понимать уже не только законы природы, но и понятия и ценности человеческого общества – литературу, искусство, архитектуру и т.д. Сейчас, в первой половине 21-ого века роль информации в жизни человека является определяющей – чем больше навыков и знаний он имеет, тем выше ценится как специалист и сотрудник, тем больше имеет уважения в обществе.

Важность принятия правильного решения финансовым аналитиком или бизнесменом, готовящимся вложить крупную часть своего капитала в дело, сложно переоценить. Иногда на кон ставятся экономики не только предприятий и холдингов, а стран и регионов.

### **Анализ текста.**

Анализ текста, определение многих его параметров – семантический анализ, извлечение сущностей, анализ тональности текста и так далее – есть одна из основных задач данного приложения. Определяющий функционал.

Получение этих параметров приведет к возможности частичного ранжирования соответствующим интересам клиента.

Семантический анализ - этап в последовательности действий алгоритма автоматического понимания текстов, заключающийся в выделении семантических отношений, формировании семантического представления текстов. Один из возможных вариантов представления семантического представления -

структура, состоящая из "текстовых фактов" [1]. Семантический анализ в рамках одного предложения называется локальным семантическим анализом. В общем случае семантическое представление является графом, семантической сетью, отражающим бинарные отношения между двумя узлами - смысловыми единицами текста. Глубина семантического анализа может быть разной, а в реальных системах чаще всего строится только лишь синтаксико-семантическое представление текста или отдельных предложений.

Тональность — это эмоциональное отношение автора высказывания к некоторому объекту (объекту реального мира, событию, процессу или их свойствам/атрибутам), выраженное в тексте. Эмоциональная составляющая, выраженная на уровне лексемы или коммуникативного фрагмента, называется лексической тональностью (или лексическим сентиментом). Тональность всего текста в целом можно определить как функцию (в простейшем случае сумму) лексических тональностей составляющих его единиц (предложений) и правил их сочетания[2].

#### **Извлечение сущностей из текста (Entity Extraction).**

Типичная задача извлечения информации: просканировать набор документов, написанных на естественном языке, и наполнить базу данных выделенной полезной информацией.

С данной задачей на данный момент довольно неплохо справляются коммерческое решение Rosette и целая система библиотек для NLP – Spacy. Правда оба решения имеют неплохой процент качества работы именно с текстами на английском языке, хотя и поддерживают другие языки. Еще одна проблема – большой объем ресурсов, необходимый для работы Spacy. Один «поднятый» instance Spacy потребляет 2 Гб оперативной памяти. Так как в рамках приложения решается задача возможности заведения для каждого клиента своего словаря – появляется проблема с количеством доступной оперативной памяти.

#### **Обучение классификатора. SVM.**

В общем, задача машинного обучения сводится к получению набора выборки данных и, в последствии, к попыткам предсказать свойства неизвестных данных. Если каждый набор данных — это не одиночное число, а например, многомерная сущность (multi-dimensional entry или multivariate data), то он должен иметь несколько признаков.

Машинное обучение представляет собой обучение выделению некоторых свойств выборки данных и применение их к новым данным. Вот почему общепринятая практика оценки алгоритма в Машинном обучении — это разбиение данных вручную на два набора данных. Первый из них — это обучающая выборка, на ней изучаются свойства данных. Второй — контрольная выборка, на ней тестируются эти свойства.

SVM. Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

Итогом проделанной работы стала рекомендательная система, ранжирующая новости в соответствии с требованиями пользователя-клиента данного сервиса. Разработаны бизнес-правила, разработана и реализована «гибкая» архитектура приложения, позволяющая без исправления кода настраиваться под требования нового клиента.

Список использованных источников:

- 1) Building Machine Learning Systems with Python. Luis Pedro Coelho, Willi Richert. Построение систем машинного обучения на языке Python. Луис Педро Коэльо, Вилли Ричарт.
- 2) Machine Learning with Spark: Create scalable machine learning applications to power a modern data-driven business using Spark. Nick Pentreath.

## **ЗАЩИТА ДАННЫХ В ОБЛАЧНОМ ХРАНИЛИЩЕ**

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Сыч Ю.В.*

*Прохорчик Р.В. – ассистент каф. ПОИТ, м.т.н.*

Облачные хранилища данных – сервисы, предоставляющие возможность хранить свои файлы на удаленных серверах, а также получать к ним доступ из любой точки мира, где есть доступ в Интернет. В условиях роста необходимости хранения информации, данные сервисы стали популярны среди пользователей. В настоящее время множество компаний также хранит большую массу информации, необходимая для принятия решений. Она накапливается в различных источниках и хранилищах, превращаясь в опыт предприятия. Однако, отправляя свои файлы в облачное хранилище, пользователи не задумываются о защите информации в облаках от доступа третьих лиц. Стоит ли всецело доверять держателям облачных сервисов, полагаясь на предпринимаемые ими меры защиты информации – личное дело каждого.

Опасность хранения файлов в облачном хранилище заключается в том, что пользователь передает файлы на хранение третьему лицу с неизвестными, по отношению к пользователю, намерениями. А опасность именно файлов, как объекта информации, в том и заключается, что с него можно сделать копию и