

Санкт-Петербургский Государственный Университет
Математико-механический факультет

Кафедра информатики

Зернов Алексей Викторович

Разработка системы автоматического анализа новостных публикаций на финансовом рынке

Бакалаврская работа

Научный руководитель:
к. ф.-м. н., доцент Григорьев Д. А.

Рецензент:
позиция рецензента рецензент

Санкт-Петербург
2017

SAINT-PETERSBURG STATE UNIVERSITY
Faculty of Mathematics and Mechanics

Computer Science Department

Alexey Zernov

Development of the automatic analysis
system of a financial market's news
publications

Bachelor's Thesis

Scientific supervisor:
assistant professor Dmitry Grigoryev

Reviewer:
reviewer position reviewer

Saint-Petersburg
2017

Оглавление

Введение	4
1. Обзор существующих решений	5
1.1. IBM: Watson Developer Cloud	5
1.1.1. News Intelligence	5
1.1.2. Social Customer Care	5
1.1.3. News Explorer	5
1.1.4. Investment Advisor	6
1.2. Microsoft: Text Analytics	6
1.3. Медиалогия	6
2. Инструменты и методологии	8
2.1.	8

Введение

1. Обзор существующих решений

1.1. IBM: Watson Developer Cloud

1.1.1. News Intelligence

Одним из наиболее интересных сервисов для анализа новостных публикаций является приложение News Intelligence¹.

Приложение предлагает ввести название интересующей Вас компании, после чего предоставляются следующие результаты:

- Наиболее упоминаемые сущности (люди, темы, компании)
- Наиболее просматриваемые новостные публикации
- Анализ тональности новостных публикаций из десяти случайных источников
- Совместные упоминания и оценка их тональности

Важно отметить, что приложение News Intelligence является лишь примером использования данного инструмента, а не готовым продуктом.

1.1.2. Social Customer Care

Еще одним интересным примером использования Watson Developer Cloud является приложение Social Customer Care². Оно осуществляет мониторинг социальных медиа, определяя потребности клиента или его запросы, а также автоматически отвечает в режиме реального времени.

1.1.3. News Explorer

Следующий пример — News Explorer³. В данном приложении отображаются наиболее обсуждаемые запросы, наиболее часто встречаю-

¹<https://discovery-news-demo.mybluemix.net/>

²<https://social-customer-care.mybluemix.net/>

³<http://news-explorer.mybluemix.net/>

щиеся совместные упоминания и список свежих новостей, разбитых по категориям.

В этом приложении также можно самостоятельно задать интересный запрос, после чего будет отображена визуализированная карта взаимосвязей разных сущностей и новостных публикаций в виде графа.

1.1.4. Investment Advisor

И последним рассматриваемым примером из данной группы является приложение Investment Advisor⁴. В данном приложении есть две группы людей: инвесторы и представители компаний. На основе анализа личностных особенностей людей по их постам, строятся определенные рекомендации по вложениям и наиболее подходящим для сотрудничества представителям компаний.

1.2. Microsoft: Text Analytics

Text Analytics⁵ позволяет провести анализ тональности текста, выделив ключевые слова. Microsoft предоставляет набор методов API для работы с данным сервисом. Подробнее о работе с ними будет написано ниже.

1.3. Медиалогия

Медиалогия⁶ — разработчик автоматической системы мониторинга и анализа СМИ в режиме реального времени.

Данная платформа предоставляет такие решения, как мониторинг СМИ компании (ее брендов, конкурентов и др.) и анализ СМИ и сообщений с использованием уникальной технологии лингвистического анализа текстов.

⁴<http://investment-advisor.mybluemix.net/>

⁵<https://text-analytics-demo.azurewebsites.net/>

⁶<http://www.mlg.ru/>

Из предоставленных примеров отчетов⁷ на сайте можно увидеть, что сервис учитывает следующее:

- **Количество упоминаний.** Отслеживается динамика по кварталам и месяцам. Отслеживая динамику по дням, платформа определяет наиболее заметные информационные поводы, вызывающие более сильный всплеск упоминаний. Также учитывается цитируемость, совместные упоминания и распределение по тематическим рубрикам.
- **Качество упоминаний.** В рассмотренном отчете предоставлена информация о положительных и негативных сообщениях. Выделены пики и проанализирована связь со СМИ, которые способствуют больше благоприятному или отрицательному всплеску упоминаний.

Также в отчете были учтены распределения по уровням СМИ, по их географическому расположению, и прочее. Однако среди всего этого наиболее важным моментом является как раз анализ текста новости: является упоминание положительным или отрицательным, как это совместно упоминается с другими запросами и тому подобное.

⁷В качестве образца был взят аналитический отчет компании «Вымпелком» в СМИ за II квартал 2009 года

2. Инструменты и методологии

2.1.