

УДК 004.896

**РАЗРАБОТКА И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ВЫЯВЛЕНИЯ РЕЛЕВАНТНЫХ СООБЩЕНИЙ ДЛЯ БАНКОВСКОЙ ОРГАНИЗАЦИИ, РАЗМЕЩЕННЫХ НА ВЕБ-САЙТАХ СОЦИАЛЬНЫХ СЕТЕЙ, НА ОСНОВЕ МЕТОДОВ ВЫЯВЛЕНИЯ ИМЕНОВАННЫХ СУЩНОСТЕЙ ДЛЯ ПРИМЕНЕНИЯ В SOC БАНКОВСКОЙ ОРГАНИЗАЦИИ**

Гурьев Н.А. (Университет ИТМО)

Научный руководитель – доцент, к.т.н. Воробьева А.А.  
(Университет ИТМО)

В докладе рассмотрены способы выявления релевантных сообщений для банковской организации, размещенных на веб-сайтах социальных сетей, на основе методов выявления именованных сущностей для применения в SOC банковской организации. Произведен сравнительный анализ по отношению к решаемой задаче.

**Введение.** В современном мире многие пользователи делятся отзывами о последних обновлениях банковских приложений или о качестве работы банковских сервисов, что играет большую роль в формировании репутационных рисков.

Однако выявить информацию, связанную с рисками, нет так просто. В новостях или сообщениях важны не только упоминания событий, но и контекст. Необходимо понять, какая связь у персоны или компании с факторами, которые банк относит к источникам риска. Нужно не только проанализировать новость, но и оценить то, как она повлияет на банк и его клиентов, другими словами, понять, является ли сообщение релевантным.

Такая задача входит в круг обязанностей сотрудников Центра по управлению киберинцидентами (security operation center, SOC) в банковских организациях – ведутся работы по мониторингу Интернет для выявления в социальных сетях и на тематических порталах сообщений о технических сбоях и банковском мошенничестве. Объем обрабатываемой сотрудниками информации огромен. Человек физически не способен оперативно прочесть все сообщения, даже если речь идет только об интересующей его теме. Таким образом, необходимо разработать решение на базе текстовой аналитики для автоматизации процесса определения релевантности сообщений.

**Основная часть.** Для выявления релевантных сообщений из социальных сетей возможно применение следующих способов:

- распознавание именованных сущностей;
- векторное представление слов с учётом контекста;
- семантическая кластеризация похожих именованных объектов;
- подбор оптимальной модели, подходящей по параметрам поиска для указанной организации.

Наибольшая эффективность выявления релевантных сообщений достигается при одновременном использовании данных техник.

**Выводы.** В качестве проверки эффективности предложенных способов выявления сообщений, было проведено их испытание на тестовых данных, собранных из профилей/групп указанного банка в социальных сетях: ВКонтакте, banki.ru, Twitter. Для каждой из подзадач выявления релевантных сообщений были применены различные модели машинного обучения.

Таким образом, одновременное применение рассмотренных методов позволяет различными путями исследовать подаваемое текстовое сообщение.

Предложенный алгоритм является оптимальным и автоматизированным решением задачи выявления сообщений, несущих репутационные риски.

Гурьев Н.А. (автор)

Подпись

Воробьева А.А. (научный руководитель)

Подпись