

И. С. Демин, докт. экон. наук, профессор департамента анализа данных и принятия решений, Финансовый университет при Правительстве РФ, г. Москва, idemin@fa.ru

Е. А. Федорова, докт. экон. наук, профессор департамента корпоративных финансов и корпоративного управления, Финансовый университет при Правительстве РФ, департамент финансов, НИУ ВШЭ, г. Москва, ecolena@mail.ru,

О. Ю. Rogov, младш. науч. сотрудник, Научно-исследовательский центр УНПК МФТИ, г. Москва, olegrgv@yandex.ru

Применение словарей тональности для текстового анализа¹

В данной статье оценивается применимость словарей тональности AFINN, NRC и Loughran and McDonald Word List для определения влияния тональности новостей на состояние фондового и валютного рынков России. Эмпирическую базу исследования составляют 2,5 млн новостей о России из различных источников, аккредитованных агентством *Thomson Reuters*. На основе метода текстового анализа, известного как «мешок слов», проведена оценка тональности каждой из статей по всем исследуемым словарям, определено, насколько колебания каждого из показателей коррелируют с основными индикаторами рынков. Доказано, что тональность новостей о России имеет заметное влияние на состояние рынков. Выявлено, что наилучшие результаты при оценке такого влияния показывает словарь NRC.

Ключевые слова: текстовый анализ, тональность новостей, фондовый индекс, коэффициент корреляции, мешок слов, текст, тематические словари.

Введение

Текстовый анализ — общее название множества методологий для обработки массивов текстовых данных. Интерес к таким методологиям существенно возрос в последнее десятилетие, и текстовый анализ находит все больше применений в различных сферах человеческой деятельности. Многие из направлений использования текстового анализа связаны с финансовыми рынками, где объем информации увеличивается с каждым днем, и традиционные методы анализа способны обработать лишь малую их часть. Вместе с тем анализ корпоративных отчетов,

новостей о компаниях, интервью с топ-менеджментом и многих других текстовых данных позволяет экономистам лучше понимать и объяснять происходящие изменения на финансовых рынках. Одним из перспективных методов текстового анализа является сентимент-анализ, или анализ тональности текстов, в основе которого лежит выявление эмоционально окрашенной лексики в текстовых массивах и оценка степени эмоциональности текстов. В западной литературе первые исследования с использованием сентимент-анализа были опубликованы в конце 1990-х гг. С начала 2010-х гг. количество публикуемых по этой теме работ начало стремительно расти, что объясняется лавинообразным ростом количества регистрируемых данных и существенным повышением вычислительных мощностей.

¹ Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финансового университета 2018 г.