

Метод количественной оценки влияния новостей разных стран на цены на нефть на основе упоминаний в Интернете*

А.В. Болдырева¹

¹ Московский физико-технический институт (Национальный исследовательский университет), г.Москва, Россия

Аннотация. Важность уверенного предсказания динамики цен на нефть определяется важностью нефтяной отрасли для экономики России. В работе предлагается метод обработки новостной ленты разных стран, который позволяет: оценить их позитивное или негативное влияние на цены; дать количественную оценку важности новостей, связанных с нефтяной отраслью. Рассматриваются новости из 23 стран мира на интервале 2016-2018 гг с шагом рассмотрения один месяц. Для каждой страны отбираются 200 дескрипторов (терминов), циркулирующих в Интернете и связанных с нефтью марок WTI и Brent. Вводится понятие рейтинга и антирейтинга новостей, которые рассчитываются на основе числа значимых терминов из указанных наборов, и степени их положительной или отрицательной корреляционной связи с динамикой цен. Значимость введенных понятий доказывается совпадением стран-лидеров по рейтингу и антирейтингу их новостей со странами лидерами и анти-лидерами нефтяного рынка. Предложенная методика может уточнить количественные прогнозы, полученные на формальных моделях, с одной стороны, и дать дополнительную информацию экспертам нефтяного рынка, работающим со СМИ, с другой.

Ключевые слова: упоминания в Интернете, топ-выборка, рейтинг новостей, WTI, Brent.

DOI: 10.14357/20790279210207

Введение

Объектом рассмотрения статьи являются новостные ленты из нефтедобывающих стран. Влияние этих новостей на цены на нефть рассматриваются с точки зрения их предсказательной способности, а именно, предшествуют ли они положительному или отрицательному тренду цен. Поскольку динамика рассматривается с месячным шагом по времени, то, предсказательные свойства полученных результатов могут рассматриваться на месячном горизонте прогноза.

Уверенный прогноз цен на нефть определяется важностью нефтяной отрасли для экономики России и нефтедобывающих стран, где экспорт нефти представляет собой значительный или ведущий источник валютных поступлений.

Известно, цены на нефть складываются под влиянием многих факторов, среди которых можно выделить экономические, климатические, военно-политические и социальные. Эконометрика учитывает эти факторы в прогнозных моделях, которые отражены в многочисленных публикациях, число которых постоянно растет. Здесь мы отметим только одну известную монографию [1], где дается комплексный анализ исторического, текущего и перспективного развития ценообразования на мировом рынке нефти. Такие модели показывают хорошие прогнозные свойства в краткосрочной перспективе (день и неделя), но оказываются менее успешными в среднесрочной перспективе (месяц, квартал), где для прогноза необходимо привлекать дополнительную субъективную информацию – мнение экспертов о динамике цен на нефть.

В своих оценках эксперты используют материалы государственных органов, рейтинговых агентств и социальных сетей, которые представ-

* Работа выполнена при поддержке гранта компании British Petroleum, Московское отделение (2018 г, приказ РАНХиГС 02-900 от 19.12.2017) и гранта Благотворительного Фонда Вл.Потанина (2019-2020 гг., грант ГК190001777).

лены в новостях на страницах Интернета. Опыт работы с такими новостями пока не отражен в каких-либо публикациях. Единственное правило, которое можно сформулировать, слушая комментарии экспертов нефтяного рынка по ТВ и читая эти комментарии в Интернет публикациях, состоит в том, что новости из одних нефтедобывающих стран ведут скорее к росту цен на нефть, а из других – к падению. При этом, есть группа стран на нефтяном рынке, новости которых мало сказываются на динамике цен. Очевидно, что такой качественный анализ новостей весьма ограничен, и было бы полезно предложить модели, которые могли бы дать количественную оценку влиянию новостей разных стран на цены на нефть. Это позволило бы уточнить количественные прогнозы, полученные на формальных моделях, с одной стороны, и дать дополнительную информацию экспертам нефтяного рынка, работающим с новостями Интернета, с другой.

Предварительно определим следующую терминологию, которую мы далее используем в статье:

- **Дескрипторы** – это части слов, слова или фразы, отражающие предметную область;
- **Упоминания в Интернете** – дескрипторы, появляющиеся на страницах Интернета, их динамика сравнивается с динамикой индикаторов;
- **Индикаторы** – показатели, предоставляемые правительственными или частными организациями, их динамика сравнивается с динамикой упоминаний;
- **Топовая выборка дескрипторов** – выборка упоминаний на страницах Интернета, наиболее высоко коррелирующих (позитивно или негативно) с заданными индикаторами;
- **Статистические сервисы поисковых машин** – сервисы, предоставляющие статистические данные о пользовательской активности по определенному дескриптору в абсолютном числе или в нормализованном формате.

В соответствии с целями исследования в упоминания мы включаем только те дескрипторы, которые связаны с экономикой и нефтяной отраслью и значимо проявляются в Интернете.

Статистические сервисы подсчитывают количество страниц в Интернете, где упоминание встретилось хотя бы один раз. Отметим, что никакой анализ текстов, как это принято в компьютерной лингвистике, не проводится.

В работе рассматривается два индикатора – это цены на нефть марок WTI и Brent. Именно эти марки рассматриваются как основные при оценке поведения нефтяного рынка. Динамика цен на WTI

и Brent используется при отборе дескрипторов, коррелирующих с ними.

Для каждого индикатора мы формируем несколько топовых выборок. Каждая из них связана с одним фиксированным смещением динамики упоминаний по сравнению с динамикой индикатора.

Рассматривались поисковые системы Яндекс и Гугла [2,3]. Поисковая система Яндекса слабее системы Гугла. Однако Яндекс отслеживает самые актуальные новостные порталы Интернета, поэтому в полученных результатах меньше информационного шума.

Вводится понятие рейтинга новостей – количественная оценка влияния новостей из разных стран на цены нефти марки WTI и Brent (в относительных единицах) и предлагается оценка рейтинга на основе:

- количества дескрипторов, которые отбираются в топовые выборки по корреляции с индикатором;
- среднего уровня корреляции в этих дескрипторах.

Для учета положительного и отрицательного влияния новостей из каждой страны. Кроме понятия рейтинга, мы рассматриваем понятие антирейтинг. Здесь позитивное влияние состоит в том, что новости из страны говорят в пользу роста цен на нефть и при оценке рейтинга мы используем положительную корреляцию. С негативным влиянием мы связываем противоположную тенденцию, когда новости говорят о снижении цены на нефть. Здесь мы используем отрицательную корреляцию.

Выдвигается гипотеза, что рейтинги и антирейтинги правильно отражают влияние новостей из разных стран на динамику цен. Задача исследования состоит в *проверке этой гипотезы*.

1. Данные

Период рассмотрения был выбран с января 2016 г. по май 2018 г., с шагом один месяц. Такой период рассмотрения определялся временными рамками выполнения гранта Московского отделения компании British Petroleum, которая поддерживала настоящее исследование. Таким образом, временные ряды дескрипторов и индикаторов сохранили по 29 временных отсчетов.

Для анализа влияния новостей разных стран на динамику цен мы отобрали 23 страны, связанные с нефтедобычей и нефтепереработкой

Временные ряды индикаторов, то есть цен на нефть марок WTI и Brent, были получены из официальных источников, доступных в Интернете.

Временные ряды упоминаний выбирались по материалам всего Интернета, при этом фиксировалось не количество упоминаний, а количество страниц, где появлялись упоминания. До начала обработки эти ряды прошли процедуру препроцессинга.

Для отбора данных и препроцессинга рядов упоминаний был использован программный комплекс Quantum. Комплекс использовался ранее для прогноза динамики макроэкономических параметров России [4] и сейчас был настроен на решение задач настоящего исследования. Функции программы:

- скачивание текстовых данных с финансово-аналитических сайтов для дальнейшего анализа и пополнения баз дескрипторов;
- обработка скачиваемых текстовых наборов и поиск устойчивых словосочетаний; скачивание данных о количестве запросов статистических сайтов Google Trends и Yandex Wordstat;
- скачивание данных о количестве упоминаний в Интернете с центральных страниц поисковых систем Яндекс и Google; скачивание данных с портала Мирового Банка Данных, и с сайта Федерального резервного банка Сент-Луиса, для дальнейшего перекрестного анализа и поиска возможных взаимосвязей между отдельными индикаторами;
- парсинг скачанных данных и сведение их в таблицы Excel;
- обработка данных полученных таблиц и создание топовых выборок для исследуемых индикаторов, в соответствии с заданными алгоритмами и параметрами

Для формирования стартового набора дескрипторов и выполнения процедур препроцессинга мы следовали рекомендациям работы [5]. Предварительно из имеющихся словарей терминов были отобраны дескрипторы по следующим группам: нефтедобыча, запасы нефти, нефтяная инфраструктура, экономика, биржевые термины, ВИЭ (возобновляемые источники энергии), народные волнения. Здесь указаны группы слов и словосочетаний, которые отражают причины формирования цены на нефть

Препроцессинг данных состоит в выполнении двух шагов:

1. Отбор заданного числа дескрипторов для каждой из 23 стран.
2. Формирование топовых выборок каждого индикатора из числа дескрипторов.

Для каждой страны были отобраны 200 дескрипторов, имеющих наибольшие относительно других дескрипторов частоты упоминаний. Наши

эксперименты с другим числом дескрипторов (в частности, с 100 и 300 дескрипторами) показали, что этого числа вполне достаточно, чтобы не потерять значимых и не приобрести лишних. Таким образом, для 23 стран мы имели 4600 дескрипторов.

2. Метод и эксперименты

2.1 Формирование топовых выборок и расчет рейтинга и антирейтинга

В соответствии с определением рейтинга и антирейтинга мы должны сформировать наборы дескрипторов, которые значимо положительно и отрицательно связаны корреляционным отношением с выбранными индикаторами – это нефть марок WTI и Brent. Корреляционные отношения рассчитываются между временными рядами на указанном выше интервале наблюдения: январь 2016 – май 2018.

При расчете корреляционных отношений мы использовали сдвиги (лаги) временных рядов дескрипторов по отношению к временным рядам индикаторов. Мы полагаем, что влияние новости на динамику индикаторов не превышает одного квартала. Именно поэтому выбраны лаги от нуля до трех месяцев.

Направление влияния определяется знаком корреляции между динамикой временного ряда дескриптора и временного ряда индикатора. При положительной корреляции мы рассчитываем рейтинг, при отрицательной антирейтинг.

В работе использовались две формы расчета корреляционных отношений. В одном случае мы использовали корреляцию значений временных рядов, в другом – корреляцию тренда временных рядов. В последнем случае использовались значения +1 и -1 при возрастании или убывании. В первом случае корреляция определялась по критерию Пирсона, во втором – Спирмена.

Топовые выборки дескрипторов формировались отдельно для каждого лага, каждого знака корреляции и каждого критерия. Эти опции отражены в табл. 1. Такие топовые выборки были построены для всех 23 стран мира.

Табл. 1

Опции отбора дескрипторов

| | Опции | Кол-во | Реализация |
|---|-----------------------------|--------|-------------------------|
| 1 | Лаги при расчете корреляции | 4 | {0,1,2,3} |
| 2 | Направление влияния | 2 | полож, отриц. |
| 3 | Критерии расчета | 2 | по Пирсону, по Спирмену |

В табл. 2 показан пример отобранных дескрипторов для Катара для нефти марки Brent по алгоритму Спирмена при расчете рейтинга. В табл. 3 показан пример получения итоговых данных для расчета рейтингов для Норвегии и России. Обозначения: Н – Норвегия, Р – Россия, PS – критерий Пирсона, SP – критерий Спирмена.

Рейтинг рассчитывается как взвешенная сумма произведений числа дескрипторов в каждой топовой выборке и среднего значения корреляции в этой выборке. Полученная сумма усреднялась по числу топовых выборок, использованных в расчете. Антирейтинг рассчитывается по той же формуле, но в нем учитывается только отрицательная корреляция.

Табл. 2

Дескрипторы и их корреляция для Катара при расчете рейтинга

| | Дескриптор | Лag 0 | Лag 1 | Лag 2 | Лag 3 |
|----|--------------------------------------|-------|-------|-------|-------|
| 1 | average price WTI Qatar | 0.66 | 0.84 | 0.91 | 0.79 |
| 2 | bullish oil prices Qatar | 0.80 | 0.81 | 0.81 | 0.77 |
| 3 | diseal fuel Qatar | 0.65 | 0.77 | 0.80 | 0.70 |
| 4 | bearish oil pipelines Qatar | 0.72 | 0.75 | 0.77 | 0.76 |
| 5 | oil gaz prices Qatar | 0.62 | 0.75 | 0.80 | 0.69 |
| 6 | Qatar nudear program | 0.58 | 0.71 | 0.76 | 0.78 |
| 7 | oil gaz service Qatar | 0.56 | 0.70 | 0.77 | 0.73 |
| 8 | decreas share crude oil Qatar | 0.53 | 0.70 | 0.82 | 0.71 |
| 9 | Qatar Stock Growth | 0.71 | 0.67 | 0.58 | 0.52 |
| 10 | ship gas oip Qatar | 0.55 | 0.66 | 0.71 | 0.60 |
| 11 | rise oil prices Qatar | 0.66 | 0.66 | 0.72 | 0.73 |
| 12 | copetotion Qatar oip market | 0.45 | 0.65 | 0.72 | 0.58 |
| 13 | Qatar export oil | 0.48 | 0.64 | 0.77 | 0.74 |
| 14 | investment Qatar oil infrastructure | 0.46 | 0.62 | 0.72 | 0.64 |
| 15 | construction Qatar oil gas pipelines | 0.41 | 0.60 | 0.69 | 0.64 |

2.2 Рейтинги новостей 23 стран мира

По правилам, описанным в предыдущем параграфе, были рассчитаны рейтинги и антирейтинги 23 стран мира. На рис.1-3 рейтинги и антирейтинги суммируются для обеих марок нефти и обоих критериев. На рис. 4 рейтинги и антирейтинги, построенные отдельно для марок WTI и Brent, затем суммировались по всем странам.

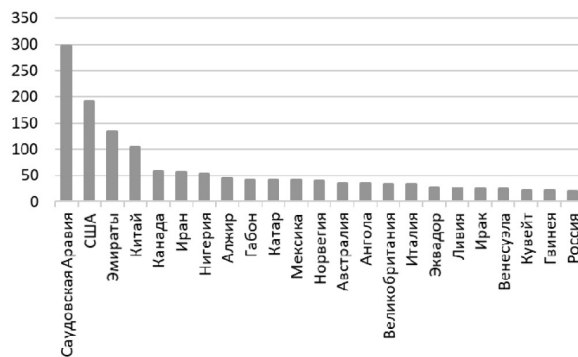


Рис. 1. Значения рейтингов 23 стран мира, указаны абсолютные значения

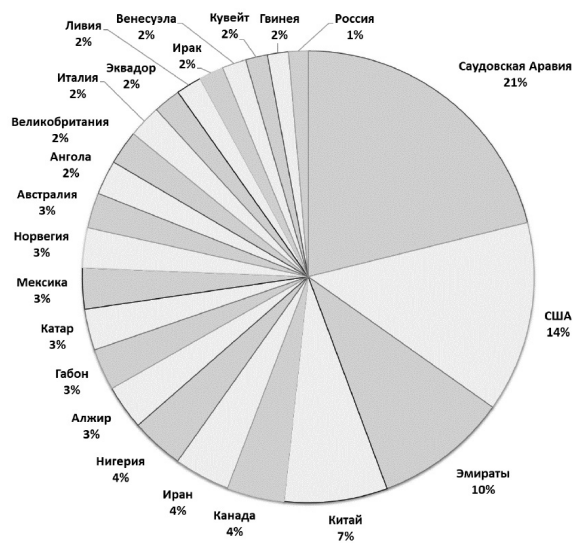


Рис. 2. Значения относительных рейтингов 23 стран мира, указаны относительные значения

Табл. 3

Средние значения корреляции для топовых выборок для Норвегии и России при расчете рейтинга

| | Расчетная величина | Кол-во Лag 0 | Корр. Лag 0 | Кол-во Лag 1 | Корр. Лag 1 | Кол-во Лag 2 | Корр. Лag 2 | Кол-во Лag 3 | Корр. Лag 3 |
|---|--------------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|
| 1 | Н., WTI-PS | 4 | 0.72 | 14 | 0.71 | 14 | 0.71 | 8 | 0.76 |
| 2 | Н., WTI-SP | 4 | 0.76 | 14 | 0.73 | 22 | 0.71 | 17 | 0.73 |
| 3 | Н., Brent-PS | 6 | 0.75 | 14 | 0.72 | 15 | 0.72 | 10 | 0.75 |
| 4 | Н., Brent-SP | 8 | 0.80 | 17 | 0.75 | 27 | 0.72 | 26 | 0.72 |
| 5 | Р., WTI-PS | 3 | 0.61 | 3 | 0.57 | 9 | 0.68 | 5 | 0.69 |
| 6 | Р., WTI-SP | 0 | 0.0 | 8 | 0.66 | 13 | 0.65 | 5 | 0.64 |
| 7 | Р., Brent-PS | 2 | 0.69 | 9 | 0.71 | 11 | 0.70 | 7 | 0.72 |
| 8 | Р., Brent-SP | 1 | 0.67 | 14 | 0.67 | 15 | 0.67 | 11 | 0.68 |

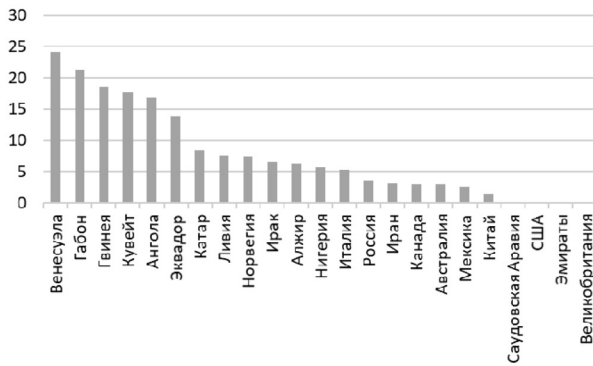


Рис. 3. Значения антирейтингов 23 стран мира, указаны абсолютные значения



Рис. 4. Суммарные рейтинг и антирейтинг влияния новостей (нефтяная отрасль и экономика в целом) по 23 странам мира для двух марок нефти; С- означает сводные рейтинг и антирейтинг

Анализ результатов показывает, что предложенная числовая форма оценки влияния новостей из разных стран на цены на нефть оказалась оправданной:

- рейтинг ожидаемо возглавили Саудовская Аравия, США, Арабские Эмираты и Китай;
- антирейтинг возглавили Венесуэла, Габон и Гвинея.

В общем случае антирейтинг не означает, что влияние этих стран мало, а означает, что новости из этих стран оказывают противоположное влияние на нефтяные цены. Однако в рассматриваемом случае оказывается, что по сравнению с положительным влиянием, отрицательное действительно мало.

Заметим, что лидеры положительного рейтинга имеют пустые отрицательные топ-выборки. То есть, рост любых новостей из этих стран, касающихся нефтяной отрасли, или экономики, все

равно соответствует росту цен на нефть в ближнем периоде с учетом временных лагов.

2.3 Рейтинги новостей из России

Мы более детально исследовали содержание топовых выборок, сформированных для расчета рейтинга и антирейтинга России. В табл. 4 представлены примеры дескрипторов, попавшие в эти топ-выборки.

Табл. 4

Примеры дескрипторов из новостей по России

| Влияние | Примеры |
|---------------|--|
| Положительное | sanctions against Russia, anti Russian sanctions USA, marine fuel Russia, financial warfare Russia, shock oil market Russia, Iranian oil Russia, и др. |
| Отрицательное | drop well efficiency Russia, bankruptcy managers oil companies Russia, policy Russia, closure oil wells Russia, и др. |

Таким образом, чем больше новостей из России о санкциях, тем выше цены на нефть. Как было показано выше, это влияние все же минимально по сравнению с влиянием новостей из Саудовской Аравии или США.

Видна закономерность – чем больше новостей о снижении эффективности нефтяных скважин России, тем ниже цена на нефть. Как было показано выше, и это влияние невелико по сравнению с влиянием новостей из Саудовской Аравии или США.

По правилам, предложенным в предыдущем параграфе, были рассчитаны рейтинг и антирейтинг новостей из России, которые представлены на рис.5. Мы разделили расчеты по обоим маркам нефти и по обоим критериям.

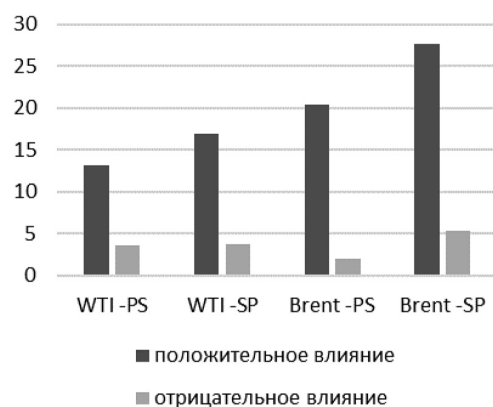


Рис. 5. Рейтинг и антирейтинг влияния новостей из России для марок нефти; PS - расчет проводился по алгоритму Пирсона, SP – расчет проводился по алгоритму Спирмена

Заключение

Прогноз цен на нефть на среднесрочном горизонте (месяц, квартал) в значительной мере определяется мнением экспертов нефтяного рынка, которые в своих предсказаниях опираются на новости Интернета. Для оценки влияния новостей стран мира на цены на нефть в настоящей работе было введено понятие рейтинга и антирейтинга новостей и предложена количественная характеристика этой оценки. Были проведены расчеты, которые показали значимость введенных понятий на основе совпадения стран-лидеров по рейтингу и антирейтингу их новостей со странами лидерами и анти-лидерами нефтяного рынка.

Можно сделать выводы по результатам выполненных расчетов:

1. Новости оказывают на нефть марки Brent более значимое влияние, чем на нефть марки WTI;
2. Применение критерия Спирмена показало более точные результаты, чем критерий Пирсона;
3. Влияние новостей из России, касающихся нефтяной отрасли, на цены обеих марок нефти оказалось минимальным среди всех исследуемых стран;
4. Аналогично общему тренду, Российские новости оказывают большее влияние на нефть марки Brent, чем на нефть марки WTI.

Как направление будущих работ можно рассмотреть создание сервиса для анализа СМИ и

использовать результаты его работы для прогнозирования нефтяных цен. При этом можно использовать не только упоминания на страницах Интернета, но и запросы в его поисковые машины. Успешность применения поисковых запросов для прогнозирования цен на нефть была показана ранее в работе [4].

Литература

1. Бушуев В.В., Конопляник А.А., Миркин Я.М. Цены на нефть: анализ, тенденции, прогноз. М.: ИД «Энергия». 2013. 344 с.
2. *Интернет* ресурс Яндекс: [https:// wordstat.yandex.ru/](https://wordstat.yandex.ru/)
3. *Интернет* ресурс Гугл: [https:// trends.google.ru/trends/](https://trends.google.ru/trends/)
4. Болдырева А. Построение прогнозных моделей экономической и социальной конъюнктуры по интенсивности запросов в поисковой сети Интернет // Современная экономика: теория, политика, инновации. Сборник студенческих научных работ. М.: РАНХиГС. 2016. С. 36–61.
5. Boldyreva A., Sobolevskiy O., Alexandrov M, Danilova V. 2017. Creating collections of descriptors based on Internet queries. Proc. of 15-th Mexican Intern. Conf. on Artificial Intelligence, Springer, LNAI, Vol. 10061, P. 303-314, Web: // https://link.springer.com/chapter/10.1007/978-3-319-62434-1_26

Болдырева Анна Витальевна. Московский физико-технический институт (Национальный исследовательский университет), г. Москва, Россия. Аспирант, ассистент в Физтех-школе аэрокосмических технологий. Количество печатных работ: 25 (в т.ч. 7 в иностранных изданиях). Область научных интересов: интеллектуальный анализ данных (data mining), интернет-социология, индуктивное моделирование. E-mail: anna.boldyreva@phystech.edu

Mentions on internet pages as a tool for assessing impact of news from different countries on oil prices

A.V. Boldyreva¹

¹ Moscow Institute of Physics and Technology (National Research University),
Moscow, Russia

Abstract. The importance of reliable predicting of oil price dynamics is determined by the importance of the oil industry for the Russian economy. This paper proposes a methodology for processing news from different countries, which allows 1) to assess their positive or negative impact on prices; (b) to quantify the importance of oil industry news from different countries. The paper considers news from 23 countries of the world for the period of 2016-2018 with the one month step. 200 descriptors (terms) for each country circulating on the Internet and associated with WTI and Brent oil are selected. The concept of rating and anti-rating of news is introduced, which are calculated on the basis of number of significant terms from the mentioned sets, and degree of their positive or negative correlation with price dynamics. The significance of the introduced concepts is proved by the coincidence of the leading countries in terms of rating and anti-rating of their news with the countries leading on the oil market. The proposed methodology can clarify quantitative forecasts of formal models and provide additional information for experts who work with media.

Keywords: mentions in Internet, top sample, news rating, WTI, Brent

DOI: 10.14357/20790279210207

References

1. *Bushuev V.V., Konoplyanik A.A., Mirkin Ya.M.* 2013. Tseny na neft: analiz, tendencii, prognoz [Oil prices: analysis, tendencies, forecast]. M.: "Energiya" Publ. 344 p.
2. *Internet resource Yandex:* [https:// wordstat.yandex.ru/](https://wordstat.yandex.ru/)
3. *Internet resource Google:* [https:// trends.google.ru/ trends/](https://trends.google.ru/trends/)
4. *Boldyreva A.* 2016. Postroenie prognoznykh modelei ekonomicheskoi i socialnoi konyunktury po intensivnosti zaprosov v poiskovoi seti Internet [Construction of predictive models for economic and social environment based on intensity of queries to the Internet search machine]. *Sovremennaya ekonomika: teoriya, politika, innovatsii. Sbornik studentcheskikh nauchnykh rabot* [Modern economics: theory, politics, innovation. Collection of student research papers]. M.: RANEPa. P. 36–61.
5. *Boldyreva A., Sobolevskiy O., Alexandrov M., Danilova V.* 2017. Creating collections of descriptors based on Internet query. *Proc. of 15-th Mexican Intern. Conf. on Artificial Intelligence, Springer, LNAI. Vol. 10061, P. 303-314, Web: // https://link.springer.com/chapter/10.1007/978-3-319-62434-1_26*

Boldyreva Anna Vitalievna. Moscow Institute of Physics and Technology (National Research University), Moscow, Russia. Ph.D. student, Assistant at the Phystech School of Aerospace Technologies. Number of publications – 25 (including 7 the foreign ones). Research interests: data mining, Internet sociology, inductive modeling. E-mail: anna.boldyreva@phystech.edu