

Алгоритм онлайн мониторинга репутации по поисковым запросам

Аннотация: В докладе представлен краткий анализ методов оценки репутации. На основе исследованных источников был составлен общий алгоритм анализа репутации. Предложен собственный алгоритм анализа репутации различных объектов (компании, продукты, персоны) на базе индекса онлайн-репутации (ex-индекса), используя лишь данные сайтов поисковой выдачи. Реализация алгоритма показана на конкретном примере. В конце обозначены перспективы и ограничения предложенного алгоритма.

Ключевые слова: репутация, мониторинг, индекс онлайн репутации, поисковая выдача, тональность

Введение

По анализу многочисленных работ можно сформулировать следующее рабочее определение репутации – это сложившееся общее мнение о достоинствах и недостатках кого-либо или чего-либо. При этом в качестве субъекта репутации могут выступать компания, бренд, персона и т.д. Репутация базируется на выявлении реальных достоинств и недостатков субъекта, рациональном представлении о субъекте, а также на опыте взаимодействия с ним. [1]. Управление репутацией направлено на то, чтобы влиять на репутацию компании или отдельного человека и контролировать ее. Поэтому отслеживание упоминаний компании или личности в сети является одной из основных задач в этой области. Например, знание того, окажет ли данная новостная статья негативное или позитивное влияние на репутацию организации, имеет решающее значение для предотвращения ущерба [2].

Среди информационных технологий выделяются системы мониторинга он-лайн репутации (online reputation monitoring ORM), реализующие поиск и отслеживание информации, касающейся какой-либо сущности (компании, бренда, персоны). Его цель – отслеживать все, что сказано в сети о каком-либо объекте для

определения состояния репутации и путей дальнейшего воздействия на нее [2].

Можно выделить следующие источники формирования репутации: сотрудники компании, отзывы, форумы, блогеры, СМИ [2].

1. Обзор методов и алгоритмов мониторинга репутации

Методы репутационного анализа опираются на экспертизу и на оценку по некоторым количественным показателям. Для измерения репутации используются индексы, аккумулирующие интегральную оценку репутации по нескольким факторам. Приведем также основные государственные стандарты, связанные с оценкой репутации. В России можно выделить несколько ГОСТов, связанных с оценкой опыта и деловой репутации [4-6]. Существуют и международные стандарты, в которых рассматриваются факторы связанные с оценкой репутации, например стандарт по управлению рисками (ISO 31000:2018), который включает анализ репутационных рисков.

Ниже приведены различные методы анализа репутации.

Таблица 1 – Методы анализа репутации

Методика Global Reputation Index	Методика оценки эффективности коммуникационной активности компании. При расчете индекса оцениваются следующие факторы: конкурентная позиция на рынке, имидж, известность брендов, частота упоминаний в прессе, контекст упоминаний. Методика предполагает определение 2-х индексов: индекс популярности и индекс доверия [3]
Индекс деловой репутации	Обобщенная характеристика надежности компании, опирающаяся на стандарт по оценке репутации [4]. Индекс выражается числом от 0 до 100 и складывается из пяти факторов, которые, в свою очередь, состоят из субфакторов. К факторам относятся: история, средства, кадры, имидж, достоверность [4]

Мониторинг СМИ	Включает количественный (фиксируется количество и динамика упоминаний, охват аудитории) и качественный (оценивается тональность публикаций) анализ. Позволяет оценить медиарепутацию компании [3]
Индекс информационного благоприятствования (ИИБ)	Представляет собой показатель качественного состояния информационного поля, формируемого СМИ вокруг компании, бренда или персоны. Значение ИИБ определяется для каждого объекта в каждом сообщении СМИ. Имеет диапазон от -1000 до 1000 для каждого сообщения в зависимости от тона упоминания. При анализе ИИБ за период все индексы по всем сообщениям суммируются. Включает 3 основных составляющих: индекс цитируемости СМИ, позитив / негатив, заметность сообщения [3]

Алгоритм анализа репутации включает процедуры сбора, обработки и интерпретации данных о том, как компанию, человека или продукт воспринимают разные группы рассматриваемой аудитории. Перечислим основные шаги:

1. Определение объекта анализа.
 2. Определение метрик для оценки репутации (индексы, экспертные оценки).
 3. Определение источников данных (поисковые системы, соцсети, новостные агрегаторы и т.д.).
 4. Сбор данных из выбранных источников.
 5. Предобработка данных (очистка, нормализация и структуризация данных).
 6. Сентимент-анализ (анализ тональности данных в шкале («позитив-негатив»)).
 7. Расчеты выбранных метрик исходя из полученных данных.
 8. Интерпретация данных и качественный анализ.
- Формулировка выводов и рекомендаций.

2. Реализация алгоритма онлайн мониторинга репутации на базе поисковых запросов на языке программирования Python

Онлайн-репутация может быть определена по результатам поиска в поисковой системе, комментариям в социальных сетях, рейтингу отзывов потребителей, статьям и новостям об организации или связанным с ними атрибутам, таким как продукты и услуги. Наиболее доступным источником информации среди перечисленных является поисковые выдачи по объекту мониторинга. Можно предположить, что оценка репутации по материалам первой страницы поисковой выдачи, является признаком для последующего анализа и обработки расширенных источников. [7]. Необходимо учесть, что поисковая выдача относится к неструктурированной информации [2].

В данной статье предложен алгоритм оценки онлайн-репутации с использованием индекса онлайн-репутации, ex [8].

Алгоритм вычисления индекса следующий:

1) Определить список поисковых запросов $x_1, x_2, x_3 \dots x_i \dots x_m$, связанных по смыслу с «А» и оценить их популярность (запрашиваемость) за отрезок времени $p(x_i)$ (месяц, год и т.д.)

2) Определить поисковые системы, по которым будет измеряться Ex -индекс (наиболее популярные: Yandex, Google, Mail)

3) Произвести подсчет положительного и отрицательного контента, осуществляя поиск по каждой поисковой фразе в каждой поисковой системе. На каждой найденной поисковой системой странице подсчитываются положительные или отрицательные упоминания объекта А и рассчитывается тональность каждого найденного результата Т. Результаты записываются в таблицу.

4) Определяется индекс репутации ex поискового запроса x для каждой поисковой системы S

$$ex(S, x) = \sum_{i=1}^n T_i(S, x) * P_i, \quad (1)$$

где x – поисковый запрос,

ex – индекс репутации поискового запроса x ,

S – поисковая система,

T – тональность страницы, P — вероятность перехода (клика) пользователя поисковой системы на результат поиска.

5) Определяется индекс репутации объекта оценки A в каждой поисковой системе по формуле

$$Ex^{S(Ya:google:mail:...)} = \frac{\sum_{j=1}^m ex(S, x_j) * p_j}{\sum_{j=1}^m p_j} \quad (2)$$

где Ex – индекс репутации объекта оценки; S – поисковая система;

ex – индекс репутации поискового запроса x; p_j – относительная частотность запроса (от 0 до 1).

Индекс **Ex** репутации в поисковых системах определяется как сумма индексов в каждой поисковой системе, умноженных на долю ее рынка, по формуле

$$Ex = \sum_S^{[...] } Ex^S * \theta(S), \quad (3)$$

где $\theta(S)$ - доля рынка поисковой системы.

Для анализа тональности было проведено сравнение библиотек, позволяющих проводить анализ тональности на языке Python: NLTK (Natural Language Toolkit), VADER (Valence Aware Dictionary and sEntiment Reasoner), TextBlob, Flair, SpaCy. Данные библиотеки различаются по скорости, точности и другим критериям. В результате была выбрана библиотека TextBlob, так как она простая и имеет высокую скорость получения тональных оценок, что важно при запуске алгоритма в режиме мониторинга.

Алгоритм оценки репутации одной поисковой выдачи по объекту мониторинга реализован в среде Python. Мониторинг реализуется на базе периодической обработки поисковых запросов по объекту мониторинга. Для вычисления индекса в программе формируются таблицы частоты запросов в месяц, ссылочный массив на отклики в поисковой выдаче, тональности контента с сайта по ссылке в выдаче и кликабельности верхнего списка поисковой выдачи. Алгоритм проверен на нескольких объектах мониторинга.

Заключение

В данной работе был предложен алгоритм мониторинга онлайн репутации для разных объектов (организации, личности, продуктов и т.д.). Преимущество данного алгоритма состоит в возможности оценки по открытым источникам (сайты в сети Интернет), а также не требует хранения больших массивов данных.

Вместе с тем такой подход не лишен ограничений. Для улучшения своей репутации организации могут продвигать свои сайты в топы поисковой выдачи, поэтому полученная оценка репутации может быть лучше действительной. Также подобные алгоритмы должны учитывать следующее: наличие информационных сайтов на первых строках поисковой выдачи (это скорее факты, чем мнения); проблема автоматического доступа к некоторым сайтам с защитой от ботов (например, капча); сложность единообразной обработки разнородного html-кода сайтов; неточности в определении тональности; отсутствие способа присвоения весов различным сайтам (исходя из объема текста и репутации самого сайта); использование одного фактора при оценке репутации; необходимость ручного составления списка ключевых слов для запроса.

Тем не менее алгоритм можно использовать для примерной оценки репутации, а указанные ограничения можно использовать как перспективные направления дальнейшей работы.

Литература:

1. *Тендит К.Н.* Основы репутационного менеджмента: учеб. пособие. – Комсомольск-на-Амуре: ФГБОУ ВПО «КнАГТУ», 2013. – 85 с.
2. *Saleiro P.* Entity Retrieval and Text Mining for Online Reputation Monitoring. – University of Chicago, 2017. – 173 p.
3. *Краснянский Д.Е.* Основы репутационного менеджмента: тексты лекций. - М.: МГТУ ГА, 2015. – 40 с.
4. ГОСТ Р 66.0.01–2017. Оценка опыта и деловой репутации субъектов предпринимательской деятельности. Национальная система стандартов. Общие положения, требования и руководящие принципы. – Введ. 2017-07-01. – М.: Стандартинформ, 2017. – 23 с.

5. ГОСТ Р 56002–2014. Оценка опыта и деловой репутации строительных организаций. Модель оценки и требования. – Введ. 2014-09-01. – М.: Стандартинформ, 2014. – 28 с.

6. ГОСТ Р 71198–2023. Индекс деловой репутации (ЭКГ-рейтинг). Порядок формирования и применения. – Введ. 2023-12-01. – М.: Стандартинформ, 2023. – 34 с.

7. *Suryavanshi P., Gangele Sh.* Sentiment Analysis View for Brand Reputation Monitoring on Social Media // International Journal of Innovative Research in Computer and Communication Engineering. – 2024. – Volume 12, Special Issue 1. – P. 110-115.

8. Будник В.Р. Понятие и методы формирования онлайн - репутации компаний // Интеллектуальный потенциал России 2025: сборник статей Всероссийской научно-практической конференции с международным участием (Иркутск, 28 мая 2025 года). – Уфа: ООО «Аэтерна», 2025. – С. 35-39.

Логинова Л.Н., Дроздов А.Д.

Подходы и инструменты сбора информации из открытых источников для мониторинга и выявления угроз информационной безопасности

Аннотация: В работе рассмотрены подходы и инструменты сбора информации из открытых источников для выявления угроз и оценки уязвимостей информационной безопасности.

Ключевые слова: информационная безопасность, открытые источники информации, анализ, фишинг, угрозы

За последние годы информационная безопасность (ИБ) стала одним из ключевых направлений в цифровой среде. Истории о кибератаках и утечках данных все чаще появляются в новостных публикациях. Злоумышленники активно используют разведку информационного пространства, сбор и анализ данных, отслеживание цифровой активности пользователей и организаций для построения векторов атак. Информационные технологии стремительно развиваются, а вместе с ними появляются новые продукты, системы и инструменты. Независимо от области