

Научная статья

УДК 004.421

АРХИТЕКТУРА ИНФОРМАЦИОННОЙ СИСТЕМЫ АНАЛИЗА ЗАКОНОМЕРНОСТЕЙ В ИНФОРМАЦИОННЫХ ПОТОКАХ

✉Самарин Максим Александрович;

Максимов Александр Викторович.

Санкт-Петербургский университет ГПС МЧС России, Санкт-Петербург, Россия

✉samarin2024@list.ru

Аннотация. В последнее время значительно увеличилась роль различного рода социальных сетей как источника массива разнородных данных, в том числе при возникновении происшествий и чрезвычайных ситуаций. Анализ всего массива таких гетерогенных данных, аккумулирующихся в социальных сетях, позволяет на его основе принимать управленческие решения и вырабатывать сценарии превентивных действий спасательных служб при возникновении происшествий. Это требует разработки специализированной информационной системы, направленной на решение данной задачи. В настоящей статье предложена оригинальная архитектура данной информационной системы, на основе которой в дальнейшем планируется реализация программного продукта.

Ключевые слова: большие данные, чрезвычайная ситуация, программное средство, коэффициенты и индикаторы, структуризация

Для цитирования: Самарин М.А., Максимов А.В. Архитектура информационной системы анализа закономерностей в информационных потоках // Науч.-аналит. журн. «Вестник С.-Петерб. ун-та ГПС МЧС России». 2023. № 2. С. 177–185.

Scientific article

ARCHITECTURE OF THE INFORMATION SYSTEM FOR THE ANALYSIS OF PATTERNS IN INFORMATION FLOWS

✉Samarin Maxim A.;

Maksimov Alexander V.

Saint-Petersburg university of State fire service of EMERCOM of Russia, Saint-Petersburg, Russia

✉samarin2024@list.ru

Abstract. Recently, the role of various kinds of social networks has significantly increased as a source of an array of heterogeneous data, including in the event of incidents and emergencies. Analysis of the entire array of such heterogeneous data accumulated in social networks makes it possible to make managerial decisions on its basis and develop scenarios for preventive actions by rescue services in the event of an accident. This requires the development of a specialized information system aimed at solving this problem. This paper proposes the original architecture of this information system, on the basis of which the implementation of the software product is planned in the future.

Keywords: big data, emergency, software tool, coefficients and indicators, structuring

For citation: Samarina M.A., Maksimov A.V. Architecture of the information system for the analysis of patterns in information flows // Scientific and analytical journal «Vestnik Saint-Petersburg university of State fire service of EMERCOM of Russia». 2023. № 2. P. 177–185.

© Санкт-Петербургский университет ГПС МЧС России, 2023

Введение

В современном мире количество разноформатных (разнотипных) данных постоянно растет экстремально высокими темпами. Во многом это связано с увеличением числа инструментов как технического типа (смартфоны, смарт-часы, IoT и т.д.), так и технологического (социальные сети, мобильные приложения, компьютерные сети, облачные технологии и т.д.), которые позволяют создавать, собирать, хранить и передавать данные [1–3]. Всё это многообразие средств ежесекундно синтезирует огромный объем неструктурированной информации, обобщение и анализ которой позволил бы решать многие актуальные проблемы человечества, в том числе задачи защиты и обеспечения безопасности людей при возникновении происшествий различного характера [4, 5].

Своевременно осуществленный анализ данных с технической, экономической и исследовательской точек зрения поможет осуществить поддержку принятия решений на новом уровне. Сделать это, призваны новые технологии «Big Data», основными преимуществами которых являются [6, 7]:

- обнаружение скрытых тенденций в большом объеме разрозненных данных;
- прогнозирование на основе выявленных тенденций.

Сегодня технологии «Big Data» используются в разных сегментах государственной деятельности, влияние которых постоянно расширяется. В настоящей статье представлена архитектура разрабатываемой информационной системы анализа и выявления закономерностей в информационных потоках, которая после программной реализации позволит повысить эффективность работы МЧС России [8, 10], а именно: значительно облегчить сбор большого количества данных; оптимизировать их анализ; представлять сценарии превентивного характера для возможных происшествий природного и техногенного типа; сокращать расходы ресурсов (временных, финансовых и др.); оптимизировать управленческие предложения; обосновывать принятые решения, основываясь на конъюнктуре большого числа параметров и математического аппарата.

Методы исследования

На данный момент в МЧС России существует множество различных баз данных. Данные сервисы, а также имеющиеся иные специализированные хранилища данных могли бы составить единую централизованную сеть, способную обрабатывать и анализировать разнообразные данные о происшествиях, в том числе об их концентрации и зависимости, например, от территории поступления сигнала. Это стало бы респектабельным информационным фундаментом для разрабатываемого программного продукта, где он получал весь требуемый массив данных.

В настоящей статье основное внимание направлено на структуризацию и анализ текстовых данных в виде сообщений, тегов и текстов из различных источников информации, как упомянутая выше единая централизованная сеть.

Процесс структуризации и анализа информации о происшествиях в разрабатываемой системе будет осуществляться посредством синтеза нескольких существующих подходов и методик оценки текстовой информации [3, 9]: индикатор дивергенции Кульбака-Лейблера; коэффициент агрессивности; индикатор информационной энтропии; индикатор выделения общеупотребительных слов; коэффициент Трейгера; индикаторы, основанные на распределении Бернулли; формула Флеша-Кинсайда; индикатор семантического алгоритма Гинзбурга; коэффициент определенности действия; индикатор связанности по Гинзбургу и т.д.

Архитектура планируемой к разработке информационной системы будет состоять из трех основных модулей:

- модуль фильтрации;
- модуль смыслового анализа и отбора данных;
- модуль анализа уровня эмоциональной напряженности.

Задачей первого модуля является фильтрация всей имеющейся в сети информации по определенным параметрам, таким как временной интервал публикации сообщения, привязка к определенным географическим координатам, фильтрация по тегам и т.д. Далее, отобранные сообщения и тексты передаются во второй модуль.

Второй модуль анализирует текстовую информацию по смысловой нагрузке, производится поиск сообщений по ключевым словам и словосочетаниям с целью составления выборки именно о происшествиях. Далее, ранжированные тексты передаются оператору и параллельно в третий программный модуль.

Третий модуль анализирует информацию по уровню эмоциональной нагрузки, выделяя сообщения с повышенной эмоциональной насыщенностью, и передает эту информацию оператору.

Рассмотрим более подробно процесс анализа и структуризации данных в каждом программном модуле.

После получения оператором МЧС России сообщений с ключевыми словами, такими как пожар, происшествие, чрезвычайная ситуация и т.д., он отправляет в систему запрос на фильтрацию поступающих сообщений по времени их публикации, геопозициям, тегам и т.д. (работает первый программный модуль).

Далее, в работу включается второй программный модуль, который с помощью ключевых слов и словосочетаний отфильтровывает список сообщений и текстов по требуемому запросу. Модуль базируется на вычислении следующих индикаторов, которые позволят ранжировать выделенные слова и словосочетания:

Индикатор дивергенции Кульбака-Лейблера [3] позволяет провести сравнение распределения терминов (реальное и теоретическое). Рассчитывается с помощью следующей формулы:

$$Kl(z) = \sum_{x \in D} p_t(z, d) * \ln \left(\frac{p_t(z, d)}{p_n(d)} \right),$$

где z – реальное распределение термина; p_n – вероятность найти термин z во множестве исследуемых сообщений относительно длины определенного сообщения d , вычисляется с помощью формулы:

$$p_n(d) = \frac{N(d)}{\sum_{x \in D} N(x)},$$

где $N(d)$ – сумма терминов в сообщении d ; $\sum_{x \in D} N(x)$ – сумма терминов x во всем массиве сообщений D ; $p_t(z, d)$ – вероятность появления термина z в документе d , вычисляется с помощью следующей формулы:

$$p_t(z, d) = \frac{tf(z, d)}{F(z)},$$

где $tf(z, d)$ – вероятность употребления термина z в сообщении d ; $F(z)$ – вероятность употребления термина z в массиве сообщений D .

Индикатор информационной энтропии [3] характеризует равномерность распределения термина по сообщению, вычисляется с помощью следующей формулы:

$$Ie(z) = \sum_{d \in D} p_t(z, d) * \ln \left(\frac{1}{p_t(z, d)} \right).$$

Если данный показатель $Ie(z) > 0$, то термин равномерно представлен в коллекции документов, если $Ie(z) = 0$, то термин z встречается только в одном тексте или сообщении.

Индикатор выделения общеупотребительных слов [3] демонстрирует отличие распределения терминов z в эталонном массиве и в Национальном корпусе русского языка. Индикатор вычисляется по следующей формуле:

$$R(z) = \frac{p_e(z)}{p_k(z)},$$

где p_e – относительная частота встречаемости термина в эталонном массиве; p_k – относительная частота встречаемости термина в Национальном корпусе русского языка.

Данный индикатор позволяет выделить большую часть общих слов, если они представлены на портале. Для общеупотребительных терминов индикатор имеет значение около 1, для малоупотребительных $\gg 1$.

Индикаторы основанные на распределении Бернулли [3] вычисляются с помощью сравнения реального распределения терминов в массиве с теоретическим распределением Бернулли по следующим формулам:

$$\left\{ \begin{array}{l} Z_1(z) = \sum_{x \in D} Z_{risk,1}(z, x) \\ Z_2(z) = \sum_{x \in D} Z_{risk,2}(z, x) \\ Z_{risk,1}(z, d) = \frac{-\log_2 Prob_{norm}(z, d)}{tf(z, d) + 1} \\ Z_{risk,2}(z, d) = \frac{F(z)(-\log_2 Prob_{norm}(z, d))}{df(z)(tf(z, d) + 1)} \end{array} \right. ,$$

где $df(z)$ – количество документов в массиве, содержащих термин z .

$Prob_{norm}$ рассчитывается по следующим формулам:

$$\left\{ \begin{array}{l} Prob_{norm}(w, d) = \frac{Prob(w, d)}{\sum_{x \in D} Prob(w, d)} \\ Prob(w, d) = 2^{-\log_2 Prob_1(w, d)} \\ Prob_1(w, d) = B(N, F, X) = \left(\frac{F(w)}{tf(w, d)} \right) p^{tf(w, d)} q^{F(w) - tf(w, d)} \end{array} \right. .$$

Индикатор Флеша-Кинсайда оценивает читабельность текста с помощью формулы:

$$FRE = 206,835 - 1,3 \left(\frac{a}{b} \right) - 60,1 * \left(\frac{c}{a} \right),$$

где a – количество слов в документе; b – количество предложений в документе; c – количество слогов в документе.

Чем выше полученное значение, тем легче восприятие информации.

Индикатор семантического алгоритма Гинзбурга [3] определяет семантическую близость двух терминов относительно их окружения в пределах сообщения. Индикатор вычисляется с помощью следующей формулы:

$$ind(z|c) = \frac{N_{zc}N_t}{N_{tc}N_z},$$

где N_{zc} – встречаемость термина z со словом c ; N_t – общее число терминов в массиве сообщений; N_{tc} – сумма терминов в окружении термина c ; N_z – встречаемость термина z в массиве сообщений.

Индекс значимости рассчитывается для всех терминов Z , встречающихся в определенном предложении C . Если $ind(z|c) > 1$ – то данный показатель значим при расчёте.

Индикатор связанности по Гинзбургу [3] определяет силу семантической связи между двумя словами. Рассчитывается на основе индексов значимости по следующей формуле:

$$ginz(Z, C) = 1 - \frac{sum(Z) + sum_{razn} + sum(C)}{sum_{all}},$$

где $sum(Z)$ – сумма индексов значимости с термином $Z > 1 \mid Z \notin C$; sum_{razn} – сумма абсолютных значений разностей индексов значимости в общей части; $sum(C)$ – сумма

индексов значимости с термином $C > 1 \mid C \notin Z$; sum_{all} – сумма всех индексов значимости больших 1.

Значения индикатора связанности по Гинзбургу лежат в интервале $0 < ginz(Z, C) < 1$ (0 – слова не связаны, 1 – слова связаны максимально).

Далее, в работу подключается алгоритм поиска ключей [3, 10], который состоит из следующих шагов:

Шаг 1 – Выборка кандидатов N наиболее часто встречающихся терминов в эталонном массиве (ключевые термины).

Шаг 2 – Вычисление ключевых терминов с помощью дивергенции Кульбака-Лейблера.

Шаг 3 – Вычисление ключевых терминов с помощью информационной энтропии.

Шаг 4 – Вычисление ключевых терминов с помощью выделения общеупотребительных слов.

Шаг 5 – Вычисление ключевых терминов с помощью распределения Бернулли.

Шаг 6 – Вычисление ключевых терминов с помощью индикатор Флеша-Кинсайда.

Шаг 7 – Нормализация полученных значений индикаторов для вычисления единого значения по каждому кандидату.

Шаг 8 – Ранжирование полученных значений по каждому кандидату.

Шаг 9 – Инициализация терминов наивысшего ранга как ключевые термины.

Шаг 10 – Формирование на основе ключевых терминов биграмм и триграмм.

Шаг 11 – Вычисление биграмм и триграмм методами шагов 2–6.

Шаг 12 – Вычисление биграмм и триграмм с помощью индикатора связанности по Гинзбургу.

Шаг 13 – Ранжирование полученных значений по каждому словосочетанию.

Шаг 14 – Выбор биграмм и триграмм с наивысшим значением ранга, присвоение им значения – ключевые.

Шаг 15 – Фильтрация полученных результатов применяя «стоп слова».

Шаг 16 – Конец алгоритма, вывод ключевых слов и словосочетаний.

Далее, с помощью полученных значений производится отбор сообщений и текстов. Оператору представляется отфильтрованный список по требуемой теме. Результаты тематического поиска могут содержать большое число сообщений и текстов с целью их сокращения и конкретизации под установленные задачи (например, срочность, неотложность) запускается третий модуль, основанный на оценке эмоциональности сообщений.

Модуль оценки уровня эмоционального напряжения основан на подходе, характеризующем степень возбуждения автора в момент написания текста. Для анализа текстов имеется список маркеров ($\{M\} = \{m_1, \dots, m_n\}$, $m_i, i = \overline{1, n}$ – отдельный маркер, n – мощность множества M), способных отражать степень эмоционального напряжения автора [10–12]:

- соотношение количества глаголов к количеству существительных в единице текста;
- наличие ненормативных слов;
- количество слов в тексте;
- средний размер предложений;
- количество знаков восклицания в документе;
- наличие иконок с эмоциями.

Также имеется список коэффициентов, которые наиболее сильно отражают эмоциональную возбужденность [10, 13, 14]:

– коэффициент агрессивности вычисляется с помощью следующей формулы (нормальное значение не превышает 0,6):

$$Ka = \frac{V + Vf}{\sum w},$$

где V – количество глаголов; Vf – количество глагольных форм (причастий и деепричастий); $\sum w$ – общее количество всех слов;

– коэффициент Трейгера [15], вычисляется с помощью следующей формулы (оптимальное значение близкое к единице):

$$Kt = \frac{V}{A},$$

где V – количество глаголов; A – количество прилагательных;

– коэффициент определенности действия [16, 17] вычисляется с помощью следующей формулы (оптимальное значение близкое к единице):

$$Koa = \frac{V}{Nn},$$

где V – количество глаголов; Nn – количество существительных.

Также с целью повышения детализации анализа могут использоваться следующие коэффициенты [18]: отношение глаголов к прилагательным; отношение суммы существительных и глаголов к сумме прилагательных и наречий; соотношение предлогов к общему количеству слов; соотношение существительных и прилагательных к количеству глаголов и причастий; соотношение предлогов к общему количеству предложений.

Данные маркеры и коэффициенты имеют общие характеристики при диагностировании, а именно если их значения завышены, то у автора имеется эмоциональное беспокойство или волнение.

Оценка эмоциональной нагрузки сообщения или документа состоит из следующих шагов:

Шаг 1 – Предобработка (текст разбивается на предложения).

Шаг 2 – Очистка текста от бессмысловых частей.

Шаг 3 – Выделение признаков (определяются части речи слов).

Шаг 4 – Расчет значения маркеров.

Шаг 5 – Расчет значения коэффициентов.

Шаг 6 – Ранжирование текстов и сообщений.

Данный модуль позволяет выделить эмоционально насыщенные тексты и сообщения, которые могут свидетельствовать о критической ситуации, в которой находится автор. Полученные данные после ранжирования передаются оператору.

Архитектура планируемой к разработке информационной системы представлена на рисунке.

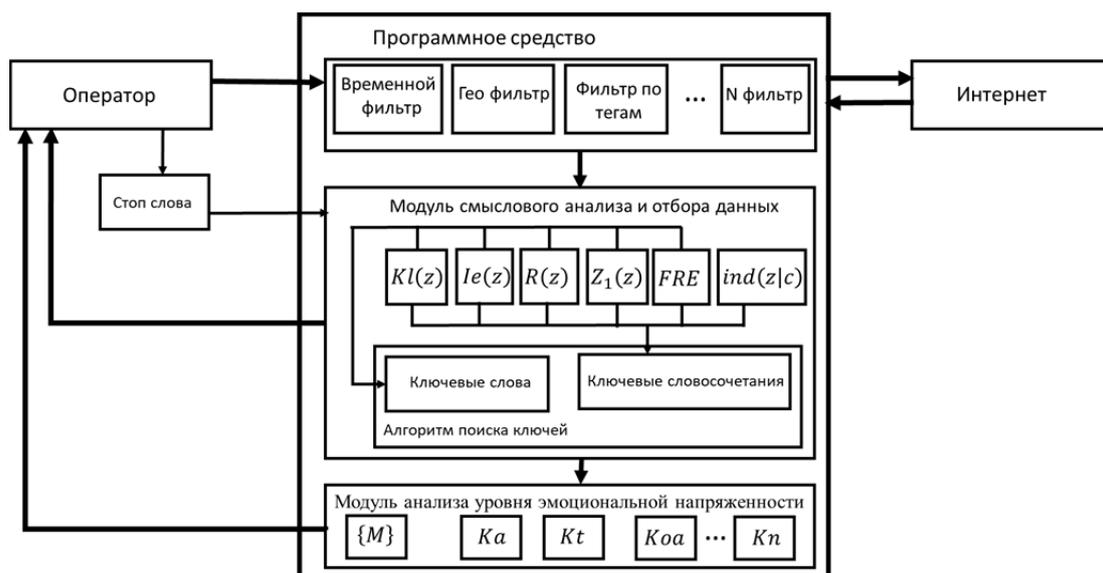


Рис. Архитектура информационной системы

Результаты исследования и их обсуждение

В дальнейшем на основе полученных результатов и представленной архитектуры планируется разработка информационной системы анализа и выявления закономерностей в информационных потоках, позволяющая автоматически производить вычисления. Источником информации для программы будут являться: новости, регулярные блоги, микроблоги, социальные сети, в которых пользователи генерируют достаточно большие объемы информации, в том числе о происшествиях, которые они увидели.

В настоящее время аналогов данной системы на рынке программного обеспечения не существует.

В основу программных алгоритмов системы легли существующие показатели оценки смысловой и эмоциональной составляющих текстов. Объединение и модернизация данных показателей позволило создать многокритериальный подход, анализирующий текстовую информацию. В ближайшем будущем также планируется модернизировать разработанные алгоритмы под анализ голосовых сообщений.

Заключение

Таким образом, в настоящей статье представлена архитектура информационной системы анализа и выявления закономерностей в информационных потоках, программные модули, входящие в её состав, и алгоритмы вычислений, функционирующие в данных модулях.

Разработка на основе представленной архитектуры информационной системы позволит значительно облегчить сбор большого количества данных; оптимизировать их анализ; разрабатывать сценарии превентивного характера для возможных происшествий природного и техногенного типа, а также сократить расходы целого спектра ресурсов (временных, финансовых, человеческих и др.).

Список источников

1. Thaduri A. Railway Assets: A Potential Domain for Big Data Analytics // *Procedia Computer Science*. 2015. Vol. 53. P. 457–467.
2. Рабинович А.Е., Август А.В. Применение технологии Big Data в сфере железнодорожного сообщения // *Оригинальные исследования*. 2021. Т. 11. С. 155–161.
3. Сбоев А.Г. Нейросетевое моделирование и машинное обучение на основе экспериментальных и наблюдательных данных: дис. ... д-ра физ.-мат. наук. М., 2021. 389 с.
4. Вострых А.В. Когнитивная модель описания пользователей информационных систем, используемых в МЧС России // *Природные и техногенные риски (физико-математические и прикладные аспекты)*. 2022. № 2 (42). С. 47–57.
5. Вострых А.В. Метод оценки эффективности графических пользовательских интерфейсов программных продуктов // *Приборы и системы. Управление, контроль, диагностика*. 2022. № 10. С. 19–28.
6. Еременко К. Работа с данными в любой сфере. Как выйти на новый уровень, используя аналитику. М.: Альпина Паблишер, 2019. 304 с.
7. Стивенс-Давидовиц С. Все лгут. Поисковики, Big Data и Интернет знают о вас все. М.: Эксмо, 2018. 384 с.
8. Максимов А.В., Матвеев А.В. Перспективы применения искусственного интеллекта в анализе больших данных социальных сетей при возникновении чрезвычайных ситуаций // *Сервис безопасности в России: опыт, проблемы, перспективы. Современные методы и технологии предупреждения и профилактики возникновения чрезвычайных ситуаций: материалы XI Всерос. науч.-практ. конф.* 2019. С. 284–286.

9. Максимов А.В., Матвеев А.В. Перспективы использования коллективных знаний при реагировании на чрезвычайные ситуации // Науч.-аналит. журн. «Вестник С.-Петерб. ун-та ГПС МЧС России». 2019. № 4. С. 89–97. EDN QPBTLA.
10. Гудовских Д.В. Анализ эмотивности текстов на основе психолингвистических маркеров с определением морфологических свойств // Вестник ВГУ. Сер.: Лингвистика и межкультурная коммуникация. 2015. № 3. С. 92–97
11. Pang B., Lee L. Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. Vol. 2. № 1/2. P. 543–561.
12. Леонтьев А.А. Основы психолингвистики. М.: СМЫСЛ, 1997. 287 с.
13. Sadegh M. Opinion mining and sentiment analysis: A survey // International Journal of Computers & Technology 2.3. 2012. P. 171–178.
14. Beigi G. An overview of sentiment analysis in social media and its applications in disaster relief // Sentiment analysis and ontology engineering. Springer. 2016. P. 313–340.
15. Mika V.M., Graziotin D., Kuutila M. The evolution of sentiment analysis – A review of research topics, venues, and top cited papers // Computer Science Review 27. 2018. P. 16–32.
16. Mikolov T. Distributed representations of words and phrases and their compositionality // Advances in neural information processing systems. 2013. P. 3111–3119.
17. Wei X. Low-Resource cross-Domain product review sentiment classification based on a CNN with an auxiliary large-Scale corpus // Algorithms. 2017. T. 10. № 3. P. 81.
18. McAuley J. Inferring networks of substitutable and complementary products. In Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD'15). Sydney, Australia. 2015. P. 178–183.

References

1. Thaduri A. Railway Assets: A Potential Domain for Big Data Analytics // Procedia Computer Science. 2015. Vol. 53. P. 457–467.
2. Rabinovich A.E., Avgust A.V. Primenenie tekhnologii Big Data v sfere zheleznodorozhnogo soobshcheniya // Original'nye issledovaniya. 2021. T. 11. S. 155–161.
3. Sboev A.G. Nejrosetevoe modelirovanie i mashinnoe obuchenie na osnove eksperimental'nyh i nablyudatel'nyh dannyh: dis. ... d-ra fiz.-mat. nauk. M., 2021. 389 s.
4. Vostryh A.V. Kognitivnaya model' opisaniya pol'zovatelej informacionnyh sistem, ispol'zuemyh v MCHS Rossii // Prirodnye i tekhnogennye riski (fiziko-matematicheskie i prikladnye aspekty). 2022. № 2 (42). S. 47–57.
5. Vostryh A.V. Metod ocenki effektivnosti graficheskikh pol'zovatel'skikh interfejsov programmnyh produktov // Pribory i sistemy. Upravlenie, kontrol', diagnostika. 2022. № 10. S. 19–28.
6. Eremenko K. Rabota s dannyimi v lyuboj sfere. Kak vyjti na novyj uroven', ispol'zuya analitiku. M.: Al'pina Publisher, 2019. 304 s.
7. Stivens-Davidovic S. Vse lgut. Poiskoviki, Big Data i Internet znayut o vas vse. M.: Eksmo, 2018. 384 s.
8. Maksimov A.V., Matveev A.V. Perspektivy primeneniya iskusstvennogo intellekta v analize bol'shikh dannyh social'nyh setej pri vozniknovenii chrezvychajnyh situacij // Servis bezopasnosti v Rossii: opyt, problemy, perspektivy. Sovremennye metody i tekhnologii preduprezhdeniya i profilaktiki vozniknoveniya chrezvychajnyh situacij: materialy XI Vseros. nauch.-prakt. konf. 2019. S. 284–286.
9. Maksimov A.V., Matveev A.V. Perspektivy ispol'zovaniya kollektivnyh znanij pri reagirovanii na chrezvychajnye situacii // Nauch.-analit. zhurn. «Vestnik S.-Peterb. un-ta GPS MCHS Rossii». 2019. № 4. S. 89–97. EDN QPBTLA.
10. Gudovskih D.V. Analiz emotivnosti tekstov na osnove psiholingvisticheskikh markerov s opredeleniem morfolozicheskikh svojstv // Vestnik VGU. Ser.: Lingvistika i mezhkul'turnaya kommunikaciya. 2015. № 3. S. 92–97

11. Pang B., Lee L. Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. Vol. 2. № 1/2. P. 543–561.
12. Leont'ev A.A. Osnovy psiholingvistiki. M.: SMYSL, 1997. 287 s.
13. Sadegh M. Opinion mining and sentiment analysis: A survey // International Journal of Computers & Technology 2.3. 2012. P. 171–178.
14. Beigi G. An overview of sentiment analysis in social media and its applications in disaster relief // Sentiment analysis and ontology engineering. – Springer. 2016. pp. 313–340.
15. Mika V.M., Graziotin D., Kuutila M. The evolution of sentiment analysis – A review of research topics, venues, and top cited papers // Computer Science Review 27. 2018. P. 16–32.
16. Mikolov T. Distributed representations of words and phrases and their compositionality // Advances in neural information processing systems. 2013. P. 3111–3119.
17. Wei X. Low-Resource cross-Domain product review sentiment classification based on a CNN with an auxiliary large-Scale corpus // Algorithms. 2017. T. 10. № 3. P. 81.
18. McAuley J. Inferring networks of substitutable and complementary products. In Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD'15). Sydney, Australia. 2015. P. 178–183.

Информация о статье:

Статья поступила в редакцию: 12.05.2023; одобрена после рецензирования: 27.05.2023; принята к публикации: 29.05.2023

Information about the article:

The article was submitted to the editorial office: 12.05.2023; approved after review: 27.05.2023; accepted for publication: 29.05.2023

Сведения об авторах:

Самарин Максим Александрович, аспирант Санкт-Петербургского университета ГПС МЧС России (196105, Санкт-Петербург, Московский пр., д. 149), e-mail: samarin2024@list.ru

Максимов Александр Викторович, заместитель начальника кафедры прикладной математики и информационных технологий Санкт-Петербургского университета ГПС МЧС России (196105, Санкт-Петербург, Московский пр., д. 149), кандидат технических наук, доцент, e-mail: he1nze@mail.ru, <https://orcid.org/0000-0002-4607-7519>

Information about authors:

Samarin Maxim A., post-graduate student, Saint-Petersburg university of State fire service of EMERCOM of Russia (196105, Saint-Petersburg, Moskovsky ave., 149), e-mail: samarin2024@list.ru

Maksimov Alexander V., deputy head of the applied mathematics and information technologies department, Saint-Petersburg university of State fire service of EMERCOM of Russia (196105, Saint-Petersburg, Moskovsky ave., 149), candidate of technical sciences, associate professor, e-mail: he1nze@mail.ru, <https://orcid.org/0000-0002-4607-7519>