

В. А. Гостюнина, аспирант, e-mail: lerusya93@list.ru,
Н. А. Давидюк, канд. техн. наук, доц., e-mail: davidyuknv@bk.ru,
Астраханский государственный технический университет

Разработка и исследование комбинированного метода классификации текстовой информации по деструктивным индикаторам

Рассматривается задача идентификации деструктивной информации, которая сводится к задаче классификации анализируемой текстовой информации по признаку наличия или отсутствия деструктивных индикаторов. Описан универсальный комбинированный метод классификации текстовой информации по деструктивным индикаторам. Предложенный метод состоит из трех этапов: нормализация данных, модифицированный поиск по словарю и байесовский классификатор. Проведена оценка эффективности работы описанного комбинированного метода.

Ключевые слова: Интернет, деструктивный контент, деструктивный индикатор, классификация текстовой информации, комбинированный метод, нормализация данных, автоматизированный поиск, байесовский классификатор, машинное обучение

Введение

Прогресс науки и техники в настоящее время сопровождается интенсивным внедрением во многие сферы человеческой деятельности новых информационных технологий. Развитие сети Интернет приводит к неконтролируемому росту количества различной информации, в том числе

представленной в текстовом виде. Среди полезных данных в сети Интернет существует значительный объем деструктивного контента [1].

Результат анализа действующих нормативно-правовых актов РФ показал отсутствие четкого определения деструктивной информации и эталонного списка деструктивного контента. Соответственно, имеются вопросы, связанные

с определением перечня деструктивной информации, в отношении которой устанавливается соответствующий правовой режим ее оборота [2, 3].

В рамках работы под деструктивным контентом (от лат. *destructio* — разрушение, нарушение нормальной структуры) понимаем информацию, которая оказывает разрушающее воздействие на психику отдельного индивидуума, группы лиц или общественное сознание, а также представляет угрозу конституционным правам и свободам человека и гражданина в области духовной жизни и нравственного здоровья.

Для создания такого списка необходимо ввести еще одно понятие — "деструктивный индикатор", т.е. критерий наличия в информации деструктивной направленности, с помощью которого можно отнести информацию к конкретному классу, а также проверить необходимость цензуры информации. В статье [4] приведен агрегированный перечень основных выделенных деструктивных индикаторов на основе анализа законодательства Российской Федерации.

Автоматизированный поиск и идентификация деструктивного контента являются крайне не актуальными направлениями обеспечения информационной безопасности из-за внедрения "суверенного Рунета" и активизации распространения информационных вбросов различными нелегальными сообществами [1, 5].

По сути, поиск и идентификация деструктивного контента сводятся к задаче классификации информации по признакам наличия либо отсутствия в ней деструктивных индикаторов, таких как насилие и жестокость, антиобщественные, противоправные действия или преступления, ненормативная лексика и др. [2, 3]. Задача усложняется тем, что не все множество индикаторов может быть идентифицировано без участия экспертов, однако потребность в минимизации их участия очевидна [6].

Таким образом, разнородный характер контента и использование различных оснований классификации приводят к необходимости разработки унифицированного метода, позволяющего эффективно классифицировать входящую текстовую информацию на основе анализа наличия/отсутствия деструктивных индикаторов.

Комбинированный метод классификации текстовой информации

Для решения задачи предлагается комбинированный метод классификации текстовой информации для обнаружения деструктивных индикаторов, разработанный с учетом резуль-



Рис. 1. Этапы комбинированного трехэтапного метода классификации текстовой информации

татов анализа применимости существующих поисковых методов к решению задачи [7, 8].

Сущность разработанного метода состоит в обработке входных данных тремя модулями (рис. 1).

Предложенный метод позволяет провести поэтапный анализ текста, а затем и классификацию по одному или нескольким деструктивным индикаторам, причем 2-й и 3-й этапы метода могут применяться в комбинации или отдельно, если это требуется.

1-й этап осуществляет предварительную подготовку входных данных. При загрузке исследуемого контента проверяемый текст часто представлен в неструктурированном виде, а также не все слова могут "нести" искомую семантику. Поэтому проводится предобработка текста.

2-й этап предназначен для первоначального анализа текста на содержание деструктивных индикаторов по словарю. Часто в этих целях применяется посимвольный перебор по словарному набору для нахождения точного совпадения. Для решения задачи классификации текстов для обнаружения деструктивных индикаторов применение тематического поиска в "чистом" виде нецелесообразно [11]. В связи с этим предлагается модифицированный поиск по словарю.

По результатам 2-го этапа в случае нахождения хотя бы одного слова из словаря или его модификации (с определенной вероятностью) дальнейший процесс обработки завершится, и текст классифицируется как деструктивный. В противном случае анализ текста продолжается на 3-м этапе, базирующемся на методах машинного обучения.

Трехэтапный комбинированный метод позволяет классифицировать тексты в зависимости от конкретной задачи как по точному совпадению найденных слов с заданными деструктивными индикаторами, так и по определенной вероятности совпадения одного или нескольких индикаторов в совокупности. Рассмотрим подробнее 2-й и 3-й этапы метода.

Нормализация данных. Машинный анализ любой неструктурированной информации требует предварительной ее обработки. Основной трудностью анализа текстов является большое число слов в проверяемом тексте, так как далеко не все слова могут "нести" искомую семантику. Проблема значительно усугубится, если входной поток информации не "унифицирован", возникнет временная избыточность работы алгоритма и некорректные результаты. На рис. 2 представлена пошаговая подготовка входных данных к анализу на содержание деструктивной информации.

В проведении предобработки текста используются следующие виды алгоритмов стемминга [9, 10]:

- алгоритмы поиска (полный перебор основы слова по списку);
- усечение окончаний (имеется небольшой список правил, по которым происходит нахождение основы слова);
- лемматизация (приведение слова к канонической или первоначальной словарной форме — лемме);
- стохастические алгоритмы (вероятностное определение корневой формы слова);
- статические алгоритмы (анализ N -грамм, алгоритмы сопоставления).

Модифицированный поиск по словарю. После прохождения этапа нормализации текст готов к непосредственно анализу на содержание деструктивных индикаторов (рис. 3).

Предлагаемый модифицированный поисковый метод основывается на вычислении отношения числа символов анализируемого слова, соот-

ветствующих символам слова из словаря, к общему числу символов в слове и отнесении на этой основе текста к деструктивным/недеструктивным [12]. Для этого необходимо определить вероятность принадлежности рассматриваемого слова к деструктивному классу. Механизм расчета вероятности представляет следующий алгоритм.

Рассмотрим часть текста X , состоящую из определенного набора слов $\{X_1, \dots, X_i, \dots, X_n\}$, $i = 1, \dots, n$, где n — число слов в тексте; каждое слово текста X_i имеет конечное число символов: $X_i = \{X_1^i, \dots, X_q^i, \dots, X_m^i\}$, $q = 1, \dots, m$, m — число букв в слове. Векторное представление слов из словаря имеет вид $C = \{C_1, \dots, C_j, \dots, C_k\}$, $j = 1, \dots, k$, где k — число деструктивных слов в словаре; каждое деструктивное слово C_j из словаря имеет конечное число символов $C_j = \{C_1^j, \dots, C_w^j, \dots, C_t^j\}$, $w = 1, \dots, t$, t — число букв в деструктивном слове.

Шаг 1. Осуществим последовательное сравнение букв деструктивного слова C_j с буквами анализируемого слова X_i . Для повышения точности отнесения слова к деструктивному не допускается при посимвольном сравнении двух слов превышение длины анализируемого слова над длиной деструктивного.

Если буквы совпадают, то "счетчик" результата сравнения R_j^i увеличивается на 1. Значение R_j^i , равное сумме совпадений для всех букв слова X_i , принадлежит $[0, m]$. Можно ввести величину P_j^i , которая вычисляется следующим образом:

$$P_j^i = \begin{cases} \frac{R_j^i}{m}, & \text{если } m \leq t; \\ \frac{R_j^i}{t}, & \text{в иных случаях,} \end{cases}$$

$$i = 1, \dots, n; j = 1, \dots, k,$$

где R_j^i — значение, заданное алгоритмом своего вычисления, описанного выше.

Установим интервал P_{Π} , на основе принадлежности к которому устанавливается факт наличия деструктивной информации в тексте. Примем значение для этого показателя, установленное эмпирическим путем, равным $[0,5; 1]$.

Шаг 2. Определим показатель P отнесения текста X к деструктивным равным максимальному значению из множества полученных оценок для всех слов, входящих в него:

$$P = \max\{P_j^i\}, i = 1, \dots, n; j = 1, \dots, k.$$

Шаг 3. Сравним найденный показатель P и пороговое значение P_{Π} и классифицируем текст как деструктивный или недеструктивный. Такой способ классификации применим



Рис. 2. Подготовка входных данных к анализу на содержание деструктивной информации



Рис. 3. Контекстная диаграмма работы модифицированного поиска по словарю

при "строгом" отборе, когда обнаружение хотя бы одного деструктивного слова в тексте X является достаточным условием отнесения текста к деструктивному классу.

Шаг 4. После установления порогового значения выделим деструктивные слова в множество D : $X_i \in D$, если $P_j^i \in P_{\text{п}}$. Данное множество выявленных деструктивных слов в анализируемом тексте X может представлять самостоятельный интерес для исследования. Например, когда обнаружение хотя бы одного деструктивного слова в тексте X является недостаточным условием деструктивности, то показатель отнесения текста X к деструктивному классу предлагается вычислить на основе отношения числа выявленных деструктивных слов $X_i \in D$ к числу уникальных слов исходного текста X .

Необходимо обратить внимание на возможность пополнения словаря новыми деструктивными словами. Найденное слово в тексте, относящееся к классу деструктивных, является новым в том случае, если значение его показателя отнесения к заданному классу принадлежит интервалу P_c , который найден эмпирическим способом и необходим для отсеивания слов, отличающихся от существующих словарных слов только падежным окончанием или множественным числом. Данный интервал принимается равным $[0,5; 0,75)$.

Шаг 5. Определим принадлежность найденного деструктивного слова словарю деструктивных индикаторов:

$$\begin{cases} X_i \in C, \text{ если } P_j^i \in P_c; \\ X_i \notin C, \text{ в иных случаях.} \end{cases}$$

Шаг 6. При $P_j^i \in P_c$ добавим новое слово в словарь деструктивных индикаторов на основе расчетов вероятностей отнесения слов к деструктивным.

По результатам работы модифицированного поиска по словарю анализируемому тексту присваивается класс "деструктивный" или "недеструктивный", а также, в случае необходимости, обновляется словарь деструктивных индикаторов. Если текст классифицирован как деструктивный, то итоговый результат в виде присвоенного класса к анализируемой текстовой информации считается достигнутым. В противном случае текстовая информация передается для дальнейшего анализа на следующем этапе "Байесовский классификатор".

Байесовский классификатор. Суть работы байесовского классификатора заключается в автоматическом определении наличия деструктивных индикаторов в тексте на основе обучающей выборки (рис. 4) [13, 14]. Поэтому



Рис. 4. Контекстная диаграмма работы байесовского классификатора

метод сводится к определению наиболее вероятного класса (к которому принадлежит анализируемый текст) при использовании оценки апостериорного максимума. Для этого необходимо вычислить вероятности классов и осуществить выбор класса, который обладает максимальным значением вероятности.

Обучающая выборка. Перед началом работы байесовского классификатора необходимо провести его обучение на выборке, состоящей из деструктивных и недеструктивных текстов, которые являются эталонными по заданной задаче.

Во время обучения байесовского классификатора происходит пополнение двух словарей деструктивными и недеструктивными индикаторами. Каждому слову из словаря присваивается переменное значение, которое является частотой встречаемости его в анализируемых ранее текстах соответствующего класса. Эти переменные при пополнении определенного словаря увеличиваются на значение, которое являлось частотой встречаемости в текущем тексте. Благодаря этой функции Байесовский классификатор является самообучающимся [13, 14].

Процесс обучения классификатора позволяет также вычислить показатель отнесения анализируемого текста к деструктивному/недеструктивному классам. Для этого необходимо ввести учет общего числа проанализированных текстов, что является суммой числа текстов из обучающей выборки, и числа текстов, которые отнесены к определенному классу. Значения также корректируются после окончания анализа каждого нового текста.

Таким образом, обучающая выборка постоянно пополняется и делает байесовский классификатор более точным в анализе поступающих текстов. Во время пополнения обучающей выборки эксперт может проверять правильность работы байесовского классификатора. Корректировать классификатор во время обучения эксперт может на основе анализа новых эталонных текстов.

Алгоритм обработки информации при решении задачи байесовским классификатором следующий.

Вернемся к рассмотрению части текста X , состоящей из определенного набора слов $\{X_1, \dots, X_i, \dots, X_n\}$, $i = 1, \dots, n$, где n — число слов в тексте. Запишем векторное представление слов из словарей с деструктивными и недеструктивными индикаторами:

$$Q = \{Q_1, \dots, Q_p, \dots, Q_q\}, p = 1, \dots, q, \quad (1)$$

где q — число слов в словаре с деструктивными индикаторами;

$$W = \{W_1, \dots, W_t, \dots, W_w\}, t = 1, \dots, w, \quad (2)$$

где w — число слов в словаре с недеструктивными индикаторами.

Шаг 1. При $X_i \in Q$ или $X_i \in W$ определим показатель отнесения слова X_i к деструктивному/недеструктивному классам K_j по формуле

$$P(X_i|K_j) = \frac{X_{iK_j}}{\sum X_{K_j}}, j = 1, 2,$$

где X_{iK_j} — переменная, характеризующая частоту встречаемости X_i слова в текстах класса K_j ; $\sum X_{K_j}$ — переменная, характеризующая общее число слов, находящихся в текстах класса K_j ;

j — идентификатор класса, в частности, 1 — деструктивный класс, 2 — недеструктивный класс.

Шаг 2. При $X_i \notin Q$ применим сглаживание Лапласа, предполагающее искусственное добавление единицы к частоте появления каждого слова:

$$P(X_i|K_j) = \frac{X_{iK_j} + 1}{\sum X_{K_j} + V}, j = 1, 2,$$

где V — переменная, характеризующая число уникальных слов в обучающей выборке.

Шаг 3. Вычислим показатель отнесения анализируемого текста X к деструктивному/недеструктивному классам с помощью переменных, которые получены в результате работы метода Байеса на обучающей выборке:

$$P(K_j) = \frac{T_{K_j}}{T}, j = 1, 2,$$

где T_{K_j} — переменная, характеризующая число текстов, относящиеся в классу K_j ; T — переменная, характеризующая общее число текстов в обучающей выборке.

Шаг 4. Вычислим конечные показатели по следующей формуле:

$$K^* = \arg \max_j \left[P(K_j) \prod_i P(X_i|K_j) \right],$$

$$i = 1, \dots, m, j = 1, 2,$$

где m — число слов в текстах класса K_j ; j — идентификатор класса, в частности, 1 — деструктивный класс, 2 — недеструктивный класс.

Таким образом, анализируемый текст X отнесется к тому классу K_j , показатель K^* которого максимален.

При получении результата работы байесовского классификатора переменные обучающей выборки обновляются, а также пополняются словари деструктивного/недеструктивного классов.

В итоге анализируемый текст относится к деструктивному/недеструктивному классу без обязательного постоянного присутствия и участия эксперта, что позволяет увеличить точность и скорость анализа объемных текстов.

Оценка эффективности комбинированного метода

Необходимо оценить эффективность приведенного комбинированного метода классификации текстовой информации по деструктивным индикаторам. В ходе эксперимента требуется для каждого текста из тестовой выборки доказать или опровергнуть основную гипотезу H_0 о том, что текст является деструктивным. Кроме основной гипотезы для оценки эффективности необходимо ввести альтернативную гипотезу H_1 о том, что текст является не деструктивным.

В результате проверки текста могут быть следующие ситуации [15, 16]:

- TruePositive — ситуация, когда нулевая гипотеза принимается верно, т. е. в тексте обнаруживаются деструктивные индикаторы, которые присутствуют.
- TrueNegative — ситуация, когда нулевая гипотеза верно отклоняется, т. е. в тексте не обнаруживаются деструктивные индикаторы, которых нет.

Однако при проверке нулевой гипотезы можно допустить ошибки I или II рода.

Ошибкой I рода (FalsePositive) называют ситуацию, когда нулевая гипотеза принимается ошибочно, т. е. в тексте обнаруживаются деструктивные индикаторы, которых нет ("ложная тревога").

Ошибкой II рода (FalseNegative) называют обратную ситуацию, когда нулевая гипотеза ошибочно отвергается, т. е. в тексте не обнаруживаются деструктивные индикаторы, которые присутствуют ("пропуск цели").

Для проверки эффективности комбинированного метода отобрано 100 текстов:

Результаты проверки текстов

Методы	TruePositive (TP)	FalsePositive (FP)	FalseNegative (FN)	TrueNegative (TN)	TP + FP	FN + TN
Поиск по словарю	1	11	0	88	12	88
Модифицированный поиск по словарю	2	10	0	88	12	88
Нейронная сеть	8	4	5	83	12	88
Байесовский классификатор	9	3	4	84	12	88
Комбинированный метод	10	2	4	84	12	88

- 12 текстов деструктивной направленности (класс 1);
- 88 текстов недеструктивной направленности (класс 0).

В качестве методов для сравнения с предлагаемым комбинированным выбраны следующие:

- поиск по словарю. В основе этого алгоритма содержится заранее подготовленный словарь деструктивных терминов. Словарь является статичным на протяжении всего эксперимента;
- модифицированный поиск по словарю. В основе этого алгоритма содержится словарь деструктивных терминов, который строится по специальному алгоритму [8]. Словарь пополняется динамически после каждого нового прогона текста;
- алгоритм на базе нейронной сети. Алгоритм, который построен на базе обучающей выборки из 90 текстов (10 деструктивных, 80 недеструктивных) с использованием NeuralNetworkWizard 1.7;
- алгоритм на базе байесовского классификатора. Алгоритм работает по принципу из работы [14].

В результате проведенного эксперимента получены результаты, которые приведены в табл. 1.

Для проверки полученных значений должны выполняться равенства:

$$TP + FP = D \text{ — число деструктивных текстов,}$$

$$FN + TN = N \text{ — число недеструктивных текстов.}$$

Как видно из табл. 1, равенства для всех методов выполняются.

Определение уровня значимости и мощности. Уровнем значимости α называется вероятность совершения ошибки I рода. Значение уровня значимости α обычно задается близким к нулю, поскольку чем меньше значение уровня значимости, тем меньше вероятность совершения ошибки I рода [15].

Вероятность ошибки II-го рода обозначается β . Однако с этой величиной тесно связана другая, имеющая большое статистическое значение — мощность критерия. Она вычисляется

по формуле $(1 - \beta)$. Таким образом, чем выше мощность, тем меньше вероятность совершить ошибку II рода.

В табл. 2 приведены отношения уровня значимости и мощность к гипотезе H_0 .

Различные значения α -уровня имеют достоинства и недостатки. Меньшие α -уровни дают большую уверенность в том, что установленная альтернативная гипотеза значима, но при этом существует больший риск не отвергнуть ложную нулевую (или отвергнуть истинную альтернативную) гипотезу (ошибка II рода, или "ложноотрицательное решение"), и таким образом получить меньшую статистическую мощность. Выбор α -уровня неизбежно требует компромисса между значимостью и мощностью, следовательно, между вероятностями ошибок I и II рода [15].

Значение α устанавливается исходя из "научных конвенций" — соглашений, принятых в научном сообществе на основе практического опыта в различных областях исследования. Например, вероятность пропустить спам у современных антиспам-систем колеблется в пределах от 1 до 30 %. Вероятность ошибочно отвергнуть валидное сообщение — от 0,001 до 3 %. Выбор системы и ее настроек зависит от условий конкретного получателя: для одних получателей риск потерять 1 % хорошей почты оценивается как незначительный, для других потеря даже 0,1 % является недопустимой.

Таблица 2

Уровень значимости и мощность [16]

Событие	Наличие различий в действительности	
	Нулевая гипотеза неверна	Нулевая гипотеза верна
Нулевая гипотеза отклоняется	Истинно положительный результат $(1 - \beta)$	Ложно положительный результат α (ошибка I рода)
Нулевая гипотеза принимается	Ложно отрицательный результат β (ошибка II рода)	Истинно отрицательный результат $(1 - \alpha)$

Таблица 3

Уровни значимости и мощности

Методы	Уровень значимости	Мощность
Поиск по словарю	0,11	1
Модифицированный поиск по словарю	0,1	1
Нейронная сеть	0,04	0,95
Байесовский классификатор	0,03	0,96
Комбинированный метод	0,02	0,96

По результатам проведенного эксперимента уровни значимости и показатели мощности сведены в табл. 3.

Результаты вычислений показывают самый низкий уровень значимости для комбинированного из всех проверенных методов, что указывает на малую вероятность совершения ошибки I рода. Кроме этого, необходимо отметить, что при низком показателе уровня значимости мощность осталась на высоком уровне, что в совокупности доказывает эффективность разработанного комбинированного метода классификации информации по деструктивным индикаторам.

Заключение

В работе описан трехэтапный комбинированный метод классификации текстовой информации по деструктивным индикаторам. Предложенный метод включает нормализацию текста, модифицированный поиск по словарю и байесовский классификатор. Модифицированный поиск по словарю имеет отличительную способность к самонаполнению словаря новыми словоформами, что позволяет иметь актуальный набор деструктивных индикаторов. Байесовский классификатор имеет способность к самообучению, что повышает точность классификации входной текстовой информации. Оценка эффективности комбинированного метода показала, что он имеет низкую вероятность совершения ошибок, обеспечивая высокую точность результатов в сравнении с другими методами. Разработанный комбинированный метод является универсальным методом в решении задач в сфере классификации, идентификации и поиска требуемой информации.

Список литературы

1. **Смирнов А. А.** Негативный контент: проблемы идентификации в контексте правового регулирования // Информационное право. 2015. № 2. С. 18–25.

2. **Федеральный закон** от 29 декабря 2010 г. № 436-ФЗ "О защите детей от информации, причиняющей вред их здоровью и развитию" // Российская газета, 2010 г. URL: <https://www.rg.ru/2010/12/31/deti-inform-dok.html> (дата обращения 18.01.2020).

3. **Роскомнадзор.** Концепция информационной безопасности детей. URL: <http://rkn.gov.ru/mass-communications/p700/p701/> (дата обращения: 13.04.2019).

4. **Байдулова Д. Р., Байтуменов А. З., Гостюнина В. А., Давидюк Н. В.** Анализ содержательных индикаторов для решения задачи по возрастной классификации текстовой информации // Студенческая наука для развития информационного общества. Сб. Матер. IX Всеросс. науч.-техн. конф. 2019. С. 60–68.

5. **Гостюнина В. А.** Анализ требований Федерального закона № 436 для определения возрастной оценки Web-контента // I Междунар. науч.-техн. конференция "Актуальные вопросы использования технологий анализа данных и искусственного интеллекта" (Астрахань, Астраханский государственный университет, 6–8 ноября 2018 г.) — молодежная секция. 2018. С. 39–43.

6. **Гостюнина В. А., Давидюк Н. В., Давидюк В. В., Байтуменов А. З.** Процедура оценки степени доверия эксперту на примере задачи возрастного категорирования веб-контента // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2019. № 2. С. 86–97.

7. **Gostyunina V. A., Davidyuk N. V.** The combined method of textual information analysis for the content of destructive indicators // Journal of Physics: Conference Series. 2019. Vol. 1399(1). P. 033109(1-7). doi:10.1088/1742-6596/1399/3/033109.

8. **Гостюнина В. А., Давидюк Н. В., Байдулова Д. Р.** Модуль для идентификации деструктивной информации в web-контенте с применением интеллектуальной обработки данных / Свидетельство о регистрации программы для ЭВМ RU 2019660776, 13.08.2019. Заявка № 2019619095 от 23.07.2019.

9. **Флах П.** Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2015. 400 с.

10. **Sipser M.** Introduction to the Theory of Computation. Boston: Cengage Learning, 2013. 445 p.

11. **Давидюк Н. В., Гостюнина В. А., Байдулова Д. Р.** Интеллектуальный алгоритм идентификации деструктивной информации в тексте // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2019. № 2. С. 29–39.

12. **Давидюк Н. В., Гостюнина В. А., Байдулова Д. Р.** Идентификация деструктивной информации в web-контенте с применением интеллектуальной обработки данных // Математические методы в технике и технологиях — ММТТ. 2019. Т. 12. Ч. 1. С. 206–210.

13. **Tang V.** Toward optimal feature selection in naive Bayes for text categorization // IEEE Transactions on Knowledge and Data Engineering, 28(9), 2016. С. 2508–2521.

14. **Байдулова Д. Р., Гостюнина В. А., Давидюк Н. В.** Идентификация деструктивной информации с использованием интеллектуального метода Байеса // Каспий в эпоху цифровой экономики материалы Междунар. науч.-практ. форума. 2019. С. 151–154.

15. **Савинов А. Н., Иванов В. И.** Анализ решения проблем возникновения ошибок первого и второго рода в системах распознавания клавиатурного почерка / Вестник Волжского университета им. В. Н. Татищева. 2011. № 18. С. 120–125.

16. **Выборнова О. Н.** Управление рисками обработки информации на основе экспертных оценок / Автореферат дис. кандидата технических наук. Кубан. гос. технол. ун-т. Астрахань, 2017.

Development and Research of a Combined Method for Classifying Textual Information by Destructive Indicators

As part of the work, the problem of identifying destructive information is considered, which reduces to the task of classifying the analyzed text information based on the presence or absence of destructive indicators. The article describes a universal combined method for classifying textual information by destructive indicators. The proposed method consists of three stages: data normalization, a modified dictionary search, and a Bayesian classifier. The article evaluates the performance of the described combined method.

Keywords: internet, destructive content, destructive indicator, classification of textual information, combined method, data normalization, automated search, Bayesian classifier, machine learning

DOI: 10.17587/it.26.402-409

References

1. Smirnov A. A. Negative content: identification problems in the context of legal regulation, *Information Law*, 2015, no. 2, pp. 18–25 (in Russian).
2. Federal law dated December 29, 2010 No. 436-ФЗ "On the protection of children from information harmful to their health and development", *Rossiyskaya Gazeta*, 2010, available at: www.rg.ru/2010/12/31/deti-inform-dok.html (accessed: 01/18/2020) (in Russian).
3. Roskomnadzor, The concept of children's information security, available at: <http://rkn.gov.ru/mass-communications/p700/p701/> (accessed: 01/18/2020) (in Russian).
4. Baidulova D. R., Baitumenov A. Z., Gostyunina V. A., Davidyuk N. V. Analysis of content indicators for solving the problem of age classification of textual information, *Student science for the development of information Society Collection of materials of the IX All-Russian scientific and technical conference*, 2019, pp. 60–68 (in Russian).
5. Gostyunina V. A. Analysis of the requirements of Federal Law No. 436 for determining the age rating of Web content, *I International Scientific and Technical Conference "Actual Issues of Using Data Analysis Technologies and Artificial Intelligence" Collection of materials of the International Conference* (Astrakhan, Astrakhan State University, November 6–8, 2018) youth section, Under the scientific editorship of S. V. Okladnikova, 2018, pp. 39–43 (in Russian).
6. Gostyunina V. A., Davidyuk N. V., Davidyuk V. V., Baitumenov A. Z. The procedure for assessing the degree of trust in an expert using the age-categorization problem for web content as an example, *Bulletin of the Astrakhan State Technical University. Series: Management, Computing and Informatics*, 2019, no 2, pp. 86–97 (in Russian).
7. Gostyunina V. A., Davidyuk N. V. The combined method of textual information analysis for the content of destructive indicators, *Journal of Physics: Conference Series*, 2019, vol. 1399(1), pp.033109(1–7), doi:10.1088/1742-6596/1399/3/033109.
8. Gostyunina V. A., Davidyuk N. V., Baidulova D. R. Module for identifying destructive information in web content using intellectual data processing, Certificate of registration of a computer program RU 2019660776, 08/13/2019. Application No. 2019619095 from 07/23/2019.
9. Flach P. Machine learning. Science and the art of constructing algorithms that extract knowledge from data, Per. from English A. A. Slinkin, Moscow, DMK Press, 2015.
10. Sipser M. Introduction to the Theory of Computation, Cengage Learning, 2013, 445 p.
11. Davidyuk N. V., Gostyunina V. A., Baidulova D. R. Intelligent algorithm for identifying destructive information in the text, *Bulletin of the Astrakhan State Technical University. Series: Management, Computing Engineering and Computer Science*, 2019, no. 2, pp. 29–39 (in Russian).
12. Davidyuk N. V., Gostyunina V. A., Baidulova D. R. Identification of destructive information in web content using intelligent data processing, *Mathematical Methods in Engineering and Technology – MMTT*, 2019, vol. 12, no. 1, pp. 206–210 (in Russian).
13. Tang B. Toward optimal feature selection in naive Bayes for text categorization, *IEEE Transactions on Knowledge and Data Engineering*, 2016, pp. 2508–2521.
14. Baidulova D. R., Gostyunina V. A., Davidyuk N. V. Identification of destructive information using the Bayesian intellectual method, *The Caspian in the era of the digital economy, materials of the International Scientific and Practical Forum*, 2019, pp. 151–154 (in Russian).
15. Savinov A. N., Ivanov V. I. Analysis of the solution to the problems of the occurrence of errors of the first and second kind in recognition systems for keyboard writing, *Bulletin of the Volga University. V. N. Tatishchev*, 2011, no. 18, pp. 120–125 (in Russian).
16. Vybornova O. N. Management of information processing risks on the basis of expert estimates, Abstract dis. Candidate of Technical Sciences, Kuban. state tech-nol. un-t Astrakhan, 2017 (in Russian).