



ВОЗМОЖНОСТИ АВТОМАТИЧЕСКОГО АНАЛИЗА ТЕКСТА В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ ПСИХОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ АВТОРА

КОВАЛЁВ А.К.

*Федеральный исследовательский центр «Информатика и управление»
Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация,
ORCID: <https://orcid.org/0000-0001-7309-7382>,
e-mail: alexeykkov@gmail.com*

КУЗНЕЦОВА Ю.М.

*Федеральный исследовательский центр «Информатика и управление»
Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация,
ORCID: <https://orcid.org/0000-0001-9380-4478>,
e-mail: kuzjum@yandex.ru*

ПЕНКИНА М.Ю.

*Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ),
г. Москва, Российская Федерация,
ORCID: <https://orcid.org/0000-0001-7046-6963>,
e-mail: mpenkina@mail.ru*

СТАНКЕВИЧ М.А.

*Федеральный исследовательский центр «Информатика и управление»
Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация,
ORCID: <https://orcid.org/0000-0003-0705-5832>,
e-mail: maxastan95@gmail.com*

ЧУДОВА Н.В.

*Федеральный исследовательский центр «Информатика и управление» Российской академии наук
(ФИЦ ИУ РАН), г. Москва, Российская Федерация,
ORCID: <https://orcid.org/0000-0002-3188-0886>,
e-mail: nchudova@gmail.com*

С помощью разработанного в ФИЦ ИУ РАН инструмента автоматического анализа текста и методов машинного обучения получены первые результаты в задаче выявления текстовых параметров, специфичных для людей с определенными психологическими особенностями. Инструмент корпусных лингвостатистических исследований, опирающийся на использование реляционно-ситуационного анализа, психолингвистических показателей и словарей, охватывающих лексику эмоциональной и рациональной оценки, позволили получить значения для 177 текстовых признаков эссе, написанных 486 испытуемыми. Для получения данных об уровне выраженности характерологических и личностных особенностей испытуемых применялся ряд психологических опросников. При обработке данных использовались алгоритмы бинарной классификации — методы опорных векторов (SVM) и «Случайный лес» (Random Forest). Полученные результаты позволяют сделать выводы о перспективности использования некоторых текстовых параметров в задачах популяционной психодиагностики и об адекватности примененных алгоритмов классификации.

Ключевые слова: автоматический анализ текста, личностные особенности, методы бинарной классификации.



Финансирование: Работа выполнена при частичной финансовой поддержке РФФИ (проект № 17-29-02247 «Создание методов диагностики распространения фрустрации в сетевых дискуссиях» и проект № 18-00-00233 «Методы комплексного интеллектуального анализа информации различных типов для социогуманитарных исследований в социальных медиа»).

Для цитаты: Ковалев А.К., Кузнецова Ю.М., Пенкина М.Ю., Станкевич М.А., Чудова Н.В. Возможности автоматического анализа текста в задаче определения психологических особенностей автора // Экспериментальная психология. 2020. Т. 13. №. 1. С. 149–158. DOI: <https://doi.org/10.17759/exppsy.2020130111>

POSSIBILITIES OF AUTOMATIC TEXT ANALYSIS IN THE TASK OF DETERMINING THE PSYCHOLOGICAL CHARACTERISTICS OF THE AUTHOR

ALEXEY K. KOVALEV

*Federal Research Center 'Computer Science and Control' of the Russian Academy of Sciences,
Moscow, Russia,*

ORCID: <https://orcid.org/0000-0001-7309-7382>,

e-mail: alexeykkov@gmail.com

YULIYA M. KUZNETSOVA

*Federal Research Center 'Computer Science and Control' of the Russian Academy of Sciences,
Moscow, Russia,*

ORCID: <https://orcid.org/0000-0001-9380-4478>,

e-mail: kuzjum@yandex.ru

MARINA Y. PENKINA

*Moscow State University of Psychology and Education (MSUPE),
Moscow, Russia,*

ORCID: <https://orcid.org/0000-0001-7046-6963>,

e-mail: mpenkina@mail.ru

MAKSIM A. STANKEVICH

*Federal Research Center 'Computer Science and Control' of the Russian Academy of Sciences,
Moscow, Russia,*

ORCID: <https://orcid.org/0000-0003-0705-5832>,

e-mail: maxastan95@gmail.com

NATALIA V. CHUDOVA

*Federal Research Center 'Computer Science and Control' of the Russian Academy of Sciences,
Moscow, Russia,*

ORCID: <https://orcid.org/0000-0002-3188-0886>,

e-mail: nchudova@gmail.com

Using a tool for automatic text analysis and machine learning methods developed at the Federal Research Center 'Computer Science and Control' of the Russian Academy of Sciences, the first results are obtained in the task of identifying text parameters specific to people with certain psychological characteris-



tics. The tool of corpus linguistic and statistical research, based on the use of relational-situational analysis, psycholinguistic indicators and dictionaries covering the vocabulary of emotional and rational assessment, allowed us to obtain values for 177 textual attributes of the essay written by 486 subjects. To obtain data on the severity of characterological and personality characteristics of the subjects, a number of psychological questionnaires were used. When processing the data, binary classification algorithms were used – the support vector method (SVM) and the Random Forest method. The results allow us to draw conclusions about the prospects of using some textual parameters in problems of population psychodiagnostics and the adequacy of the applied classification algorithms.

Keywords: automatic text analysis, personality characteristics, binary classification methods.

Funding: This work was partly supported by the Russian Foundation for Basic Research (project No. 17-29-02247 “Development of methods for diagnosing the spread of frustration in network discussions” and project No. 18-00-00233 “Methods for the integrated intellectual analysis of various types of information for social and humanitarian research in social media”).

For citation: Kovalev A.K., Kuznetsova Y.M., Penkina M. Y., Stankevich M.A., Chudova N.V. Possibilities of automatic text analysis in the task of determining the psychological characteristics of the author. *Экспериментальная психология = Experimental psychology (Russia)*, 2020. Vol. 13, no. 1, p. 149–158. DOI: <https://doi.org/10.17759/exppsy.2020130111>. (In Russ.)

Введение

Проблема определения личностных особенностей автора по созданному им тексту не нова для гуманитарных наук [1]; во второй половине XX в. на стыке психологии и лингвистики возникла новая отрасль знания — психолингвистика; в последние годы исследования в этой области получили новый импульс в связи с появлением инструментов автоматического анализа текстов (см., например: [13; 8]; а также обзор в нашей работе: [2]). В настоящей работе представлены результаты оценки личностных особенностей автора по его текстам, полученные с помощью разрабатываемого в ФИЦ ИУ РАН инструмента корпусных лингвостатистических исследований, основанного на использовании реляционно-ситуационного анализа, психолингвистических показателей и словарей, охватывающих лексику эмоциональной и рациональной оценки [4]. Результаты пилотажных исследований, выполненных с помощью данного инструмента, представлены в работах [5; 2], где для обработки данных были использованы методы математической статистики, позволившие получить первые представления о наличии устойчивой связи между текстовыми параметрами и свойствами личности. Однако корреляционный анализ и анализ различий не могут быть признаны релевантными задаче выделения текстовых признаков, на основании которых можно выделить и описать психологические особенности написавшего текст субъекта. Сложившаяся к настоящему времени традиция анализа текстов в психодиагностических целях подразумевает привлечение на этапе установления связей методов обработки данных, более соответствующих уровню сложности анализируемого материала, в том числе, действующих возможности искусственного интеллекта: машинное обучение с последующим использованием классификаторов на основе деревьев решений, наивного байесовского классификатора, SVM, нейронных (рекуррентных и сверточных) сетей и т. д. [12]. Данные методы позволяют количественно оценить достоверность установления того или иного психологического свойства по присутствию в текстах анализировавшихся признаков.



По сравнению с указанным пилотажным исследованием в настоящей работе коллекция текстов была увеличена более чем в 3 раза и к обработке данных лингвистического анализа были привлечены методы машинного обучения.

Методы и процедура исследования

В Институте проблем искусственного интеллекта ФИЦ ИУ РАН разрабатывается система когнитивного анализа текста, основанная на методах реляционно-ситуационного анализа, лексико-частотного анализа и психолингвистического анализа русскоязычных текстов научного, публицистического и бытового жанров – «Машина РСА». Метод РСА опирается на синтаксемный анализ Г.А.Золотовой [6] и на аппарат неоднородных семантических сетей Г.С. Осипова [9]. РСА работает на основе Словаря предикатов (2,7 тыс. глаголов, причастий, деепричастий и девербативов) и Определителя семантических ролей (75 семантических отношений) и семантических связей (32 семантические связи) [10]. Кроме того, в Машине РСА используется анализатор PLATIn [3], разработанный для лексико-частотного и психолингвистического анализа текстов, в том числе для автоматического анализа веб-страниц. Лексико-частотный анализ опирается на специально созданные в ФИЦ ИУ РАН словари, содержащие лексику экспрессии, лексику отрицательной эмоциональной оценки, лексику отрицательной рациональной оценки, лексику социального стресса; всего в состав созданных 29 словарей входит более 51 тыс. лексических единиц. Психолингвистический анализ осуществляется по 27 психолингвистическим показателям, находящимся, как показали результаты проведенного нами пилотажного исследования, в статически значимой взаимосвязи с более чем 50 различными психологическими свойствами личности. В настоящий момент в Машине РСА используется еще 14 частеречных показателя, которые, возможно, пополнят список психолингвистических показателей.

В ходе проведенного исследования (с участием 487 испытуемых – студентов и взрослых г. Москвы и г. Кургана) испытуемым было предложено написать эссе на тему «Я, другие, мир» объемом в одну страницу и заполнить десять опросников. Результаты для такой черты, как агрессивность, отражены в нашей публикации [7], а в настоящей работе представлены данные, полученные на основе следующих семи опросников: опросник нарциссических черт личности (НЧЛ) Н.М. Клепиковой, О.А. Шамшиковой; опросник «Стиль саморегуляции поведения» (ССПМ) В.И. Моросановой; тест жизнестойкости (ТЖ) С. Мадди в адаптации Д.А. Леонтьева и Е.И. Рассказовой; опросник черт характера (ОЧХ) В.М. Русалова и О.Н. Маноловой; а также русскоязычные версии методик: Personal Need for Structure Thompson (Naccarato, Parker, & Moskowitz (шкала «Потребность в структуре», ШПС)), Multidimensional scale of anomie (Heydari, Davoudi, & Teymoorei (Шкала аномии, ША)), New Personal Fable Scale Lapsley (Fitzgerald, Rice, & Jackson (опросник «Личный миф», ЛМ)) в адаптации Ю.М. Кузнецовой.

Тексты эссе были обработаны с помощью инструмента «Машина РСА», так что текст каждого автора оказался численно охарактеризован по 177 параметрам.

Данные представляют собой набор из 177 признаков для 487 испытуемых. Признаки разбиты на шесть групп:

- 1) психолингвистические показатели (PI) – 27 признаков;
- 2) семантические роли (SR) – 75 признаков;
- 3) семантические связи (SL) – 32 признака;
- 4) словари оценки и состояния (AS) – 20 признаков;



5) тематические словари (SD) — 9 признаков;

6) части речи (PS) — 14 признаков.

В качестве целевых признаков выступали показатели 69 шкал использованных в исследовании опросников. Для каждого целевого признака высчитывались значения первого и третьего квартиля. Испытуемые, у которых значение целевого признака больше первого квартиля, но меньше третьего квартиля, исключались из выборки. Значение больше третьего квартиля считалось высоким, и таким испытуемым присваивалась метка класса 1; значения меньше третьего квартиля считались низкими, и таким испытуемым присваивалась метка класса 0. Таким образом, задача предсказания значения целевого признака сводилась к задаче классификации.

Также признаки, которые принимали значение «ноль» в более чем 70% случаев, бинаризовались — значения, не равные нулю, приравнивались к единице и играли роль индикатора.

В качестве алгоритмов бинарной классификации использовались методы опорных векторов (SVM) и «Случайный лес» (Random Forest), реализованные в библиотеке «Scikit-learn». Метод опорных векторов — линейный классификатор, переводящий объекты в пространство более высокой размерности, чем исходное, и строящий в этом пространстве гиперплоскость, разделяющую объекты разных классов с максимальным зазором. Случайный лес представляет собой ансамбль моделей, в котором базовой моделью является дерево решений (decision tree), и сочетает в себе метод бэггинга (bagging), базовые алгоритмы строятся по бутстреппированным выборкам, и метод случайных подпространств — при построении базового алгоритма используется только часть признаков. Классификация происходит путем голосования — каждое дерево относит объект к одному из классов; побеждает тот класс, за который проголосовало наибольшее число деревьев. Метрика качества — доля верных ответов (accuracy). Настройка гиперпараметров (параметр регуляризации и тип ядра для метода опорных векторов, количество базовых моделей в ансамбле, максимальная глубина дерева для случайного леса) производилась подбором по сетке с пятикратным скользящим контролем. Из-за небольшого количества данных качество классификации оценивалось не на отложенной выборке, а путем усреднения значения метрики качества с каждой итерации пятикратного скользящего контроля.

Эксперименты проводились отдельно для каждой группы признаков. С помощью оценки важности признаков в алгоритме случайного леса в каждой группе отбирались 5 признаков с наибольшим значением важности. Совокупность этих признаков по всем группам составила группу наиболее важных признаков (TOP 5). Также были проведены эксперименты на всех признаках (ALL).

Результаты и их обсуждение

Проведенная обработка позволила выявить текстовые признаки, наличие которых дает возможность отличить тексты, написанные людьми с высоким уровнем измеренных в нашем исследовании личностных особенностей.

Рассмотрим вначале проблему достоверности полученных результатов. Наиболее высокую степень уверенности при оценке личностных особенностей по текстовым признакам можно иметь, выявляя такие особенности, как общая жизнестойкость (интегральный показатель теста жизнестойкости, достоверность связи — $0,707 \pm 0,086$), способность к осознанному программированию собственных действий (шкала «Программирование» опросника «Стиль саморегуляции поведения», достоверность — $0,696 \pm 0,061$), зависть к окружающим



(шкала «Сверхзанятость чувством зависти» опросника «Нарциссические черты личности», достоверность — $0,733 \pm 0,068$), надменность в отношении к окружающим (шкала «Дерзкое и заносчивое поведение» опросника «Нарциссические черты личности», достоверность — $0,695 \pm 0,081$). Эти черты распознаются по текстовым признакам при автоматическом анализе текста с уверенностью около 0,7.

Ниже приведен список черт, распознаваемых с уверенностью, лежащей в интервале 0,65–0,69:

1) стиль саморегуляции поведения (на основании показателей шкалы «Оценивание результата» — $0,664 \pm 0,128$);

2) стиль саморегуляции поведения (на основании показателей шкалы «Самостоятельность» — $0,671 \pm 0,075$);

3) нарциссические черты личности (на основании показателей шкалы «Вера в собственную уникальность» — $0,687 \pm 0,076$ и показателей шкалы Потребность во внимании $0,652 \pm 0,111$);

4) уровень жизнестойкости (на основании показателей шкалы «Вовлеченность» — $0,684 \pm 0,046$ и показателей шкалы «Принятие риска» — $0,662 \pm 0,055$);

5) потребность в структурировании деятельности и времени (на основании показателей шкалы «Независимость» — $0,671 \pm 0,086$ и показателей шкалы «Манипулятивность» — $0,68 \pm 0,084$);

6) показатели методики «Личный миф» (на основании показателей шкалы «Уникальность» — $0,669 \pm 0,03$, шкалы «Неуязвимость» — $0,66 \pm 0,092$, шкалы «Всесилие в общении» — $0,668 \pm 0,104$ и шкалы «Всесилие общее» — $0,662 \pm 0,088$).

Другие черты распознаются с уверенностью менее 0,65; в связи с этим они не представляют интереса для процедур классификации авторов на основе данных автоматического анализа их текстов.

В табл. 1 представлены данные о 10 наиболее значимых текстовых признаках, характерных для людей с высокими баллами по шкалам «Общая жизнестойкость» — ТЖ, «Программирование» — ССПМ, «Сверхзанятость чувством зависти» и «Дерзкое, заносчивое поведение» — НЧЛ.

Как видно из представленной выше таблицы, между признаками, характерными для людей с высоким уровнем жизнестойкости, и признаками, характерными для людей с высоким уровнем нарциссизма, обнаруживается пересечение по половине списка. Такие характеристики текста, как присутствие молодежного жаргона, значительное преобладание глаголов над существительными и в целом повышенная частота использования глаголов, частая встречаемость местоимения «мы», «наше» и т. п., частая встречаемость слов в позиции объекта описания, — все эти особенности характерны как для людей с проявлениями нарциссизма, так и для людей, обладающих высокой жизнестойкостью.

Далее, можно видеть, что другие признаки, характерные для текстов людей с высокими баллами по двум шкалам НЧЛ, оказываются существенны и для текстов людей, склонных к программированию своей деятельности (согласно данным по опроснику ССПМ). Таким образом, 6 из 10 текстовых признаков, позволяющих говорить о наличии у автора текста проявлений нарциссизма, оказываются плоходифференцирующими, и требуется дополнительная проверка — уже на отделение от текстов «жизнестойких» и «программирующих».

Полученные данные могут быть интерпретированы с точки зрения диагностической ценности выделенных связей между личностными свойствами индивида и признаками тек-



Тест жизнестойкости, общая шкала	Опросник стиля саморегуляции поведения, шкала «Программирование»	Опросник нарциссических черт личности, шкала «Занятость чувством зависти»	Опросник нарциссических черт личности, шкала «Дерзкое, заносчивое поведение»
Молодежный жаргон	Часть речи: местоимение-прилагательное	Доля местоимений 1-го лица мн. числа	Коэффициент определенности действия
Коэффициент определенности действия (глаголы/существительные)	Тональность слов	Семантическая связь: LOC	Семантическая роль: объект
Безысклительная и усилительная лексика	Число знаков препинания / число слов	Тональность слов	Семантическая связь: OBJ
Доля местоимений 1-го лица мн. числа	Средняя глубина синтаксического дерева	Число знаков препинания / число слов	Тональность слов
Коэффициент Трейгера	Семантическая роль: делибератив	Уголовный жаргон	Часть речи: прилагательное
Семантическая роль: предикат	Семантическая роль: предикат	Часть речи: наречие	Часть речи: глагол
Семантическая роль: объект	Часть речи: частица	Семантическая связь: COM	Семантическая роль: директив
Часть речи: глагол	Лексика социального разобщения	Семантическая роль: локатив	Доля глаголов прошедшего времени
Часть речи: предлог	Часть речи: наречие	Молодежный жаргон	Уголовный жаргон
Доля глаголов 3-го лица	Семантическая связь: CAUS	Прецедентные тексты, архаизмы, междометия и др.	Лексика мотивации, деятельности и напряжения

ста. В этом отношении определяемые с помощью нашего инструмента текстовые признаки различаются по своей способности дифференцировать носителей разных психологических особенностей. Наименее ценным в диагностическом плане является признак «Тональность слов», поскольку он связан одновременно с тремя из четырех упоминаемых в таблице психологических свойств. Данный факт заслуживает внимания, поскольку оценка тональности текста (сантимент-анализ) до настоящего времени остается базовой техникой текстового анализа в диагностических целях. Исследовательский интерес к данному параметру определяется интуитивным представлением о том, что в речи и текстах людей, у которых определенные психологические особенности выражены выше нормы, скапливается относительно большое количество аффективно заряженных слов и выражений; однако именно универсальность данного признака снижает его ценность при решении задач текстовой диагностики личности. Напротив, полезными в диагностическом плане являются «уникальные» связи, такие как, например, повышение частоты встречаемости в текстах жизнестойких людей безысклительной и усилительной лексики или в текстах людей с выраженной способностью к программированию — встречаемости частиц. Формирование текстовых «профилей», включающих в себя показатели Машины РСА, наиболее достоверно идентифицирующие психологические особенности, позволит решать задачи диагностики таких особенностей по порождаемым человеком текстам, а при анализе больших массивов текстов (в том числе сетевых) — задачи их классификации на принадлежащие авторам с выраженной и невыраженной диагностируемой особенностью.



В завершение отметим, что нами не были обнаружены взаимосвязи между выделяемыми с помощью Опросника черт характера взрослого человека В.М. Русалова чертами личности и параметрами текста, используемыми в Машине РСА: оба метода классификации не дают надежных результатов при работе с характерологическими особенностями испытуемого — распознавание по всем группам признаков по всем шкалам ОЧХ оказывается не выше 0,59. Другими словами, наличие ярко выраженных черт характера не проявляется в тексте. Этот вывод подкрепляется и данными проведенного нами пилотажного исследования [5], в котором между показателями шкал ОЧХ и параметрами текста не было обнаружено значимых корреляций, за исключением положительной корреляции выраженности так называемого дистимного типа характера с использованием инвектив. Связано ли это со спецификой самого понятия «характер» (в концепции В.М. Русалова и разработанном им с О.Н. Маноловой опроснике, характер понимается как интегральная структура индивидуальности, как сформировавшийся в детстве и юности инструмент адаптации к социальным требованиям) или с тем, что используемые нами текстовые признаки имеют слишком частный характер и не позволяют описывать специфику сверхфразового единства, — пока вопрос открытый.

Заключение

Результаты, полученные в данном исследовании, позволяют проводить отбор текстов, чьи авторы с некоторой долей уверенности могут быть охарактеризованы в отношении таких личностных свойств, как: стиль саморегуляции, жизнестойкость, нарциссизм, потребность в структуре, наличие личного мифа. Сами тексты при этом должны быть предварительно отнесены (или заданы испытуемым в психодиагностическом исследовании) к выполненному в эссеистическому стилю. Этот стиль определяется следующим образом: «... прозаическое сочинение небольшого объема и свободной композиции, выражающее индивидуальные впечатления и соображения по конкретному поводу или вопросу и заведомо не претендующее на определяющую или исчерпывающую трактовку предмета» [11]. Эссеистический стиль трактуется в жанроведении как стиль, в котором на первый план выступает личность автора. В связи с развитием коммуникации в социальных сетях этот стиль стал востребован многими пользователями в небольших постах Живого журнала или Фейсбука, излагающими свои взгляды на мир и значимые для них события личной и общественной жизни. Таким образом, результаты нашего исследования могут быть полезными для сетевой психодиагностики и популяционных исследований на основе данных сетевой активности.

Литература

1. Алмаев Н.А., Доронев А.Б., Малкова Г.Ю. Проявление психологической травмы в автобиографических рассказах // Экспериментальная психология. 2009. Т. 2. № 2. С. 104–115.
2. Воронцова О.Ю., Ениколопов С.Н., Кузнецова Ю.М., Чудова Н.В. и др. Лингвистические характеристики текстов психически больных и здоровых людей [Электронный ресурс] // Психологические исследования. 2018. Т. 11. № 61. URL: <http://psystudy.ru/index.php/num/2018v11n61/1622-enikolopov61.html>
3. Девяткин Д.А., Кузнецова Ю.М., Чудова, Н.В., Швец А.В. Интеллектуальный анализ проявлений вербальной агрессивности в текстах сетевых сообществ // Искусственный интеллект и принятие решений. 2014. № 2. С. 95–109.
4. Ениколопов С.Н., Кузнецова Ю.М., Смирнов И.В., Станкевич М.А., Чудова Н.В. Создание инструмента автоматического анализа текста в интересах социогуманитарных исследований. Ч. 1. Методические и методологические аспекты // Искусственный интеллект и принятие решений. 2019. № 2. С. 28–38. DOI 10.14357/20718594190203.



5. Ениколопов С.Н., Кузнецова Ю.М., Минин А.Н., Пенкина М.Ю., Смирнов И.В., Станкевич М.А., Чудова Н.В. Особенности текста и психологические особенности: опыт эмпирического компьютерного исследования // Труды ИСА РАН. 2019. № 3. С. 91–99.
6. Золотова Г.А., Онипенко Н.К., Сидорова М.Ю. Коммуникативная грамматика русского языка. М.: Ин-т рус. яз. РАН им. В.В. Виноградова, 2004.
7. Ковалёв А.К., Кузнецова Ю.М., Минин А.Н., Пенкина М.Ю., Смирнов И.В., Станкевич М.А., Чудова Н.В. Методы выявления по тексту психологических характеристик автора (на примере агрессивности) // Вопросы кибербезопасности. 2019. № 4(32). С. 72–80.
8. Литвинова Т.А., Литвинова О.А., Рыжкова Е.С., Бирюкова Е.Д., Середин П.В., Загоровская О.В. Исследование влияния пола и психологических характеристик автора на количественные параметры его текста с использованием программы Linguistic Inquiry and Word Count // Научный диалог. 2015. № 12 (48). С. 101–109.
9. Осипов Г.С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. М.: Наука, Физматлит, 1997. 142 с.
10. Осипов Г.С., Смирнов И.В., Тихомиров И.А. Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. 2008. № 2. С. 3–10.
11. Эссе. Большая советская энциклопедия. М.: Советская энциклопедия, 1969–1978.
12. Gupta U., Chatterjee A., Srikanth R., Agrawal P. A Sentiment-and-Semantics-Based Approach for Emotion Detection in Textual Conversations [Электронный ресурс] // Neu-IR: Workshop on Neural Information Retrieval, SIGIR. 2017, ACM. URL: arXiv:1707.06996
13. Pennebaker J., Boyd R., Jordan K., Blackburn K. The development and psychometric properties of LIWC-2015 [Электронный ресурс]. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

References

1. Almaev N.A., Dorodnev A.B., Malkova G.YU. Proyavlenie psikhologicheskoy travmy v avtobiograficheskikh rasskazakh // Eksperimental'naya psikhologiya. 2009. Tom 2. № 2. S. 104–115. (In Russ.).
2. Vorontsova O.YU., Enikolopov S.N., Kuznetsova YU.M., CHudova N.V. i dr. Lingvisticheskie kharakteristiki tekstov psikhicheski bol'nykh i zdorovykh lyudej // Psikhologicheskie issledovaniya. 2018. T. 11. № 61. URL: <http://psystudy.ru/index.php/num/2018v11n61/1622-enikolopov61.html>. (In Russ.).
3. Devyatkin D.A., Kuznetsova YU.M., CHudova, N.V., SHvets A.V. Intellektual'nyj analiz proyavlenij verbal'noj agressivnosti v tekstakh setevykh soobshhestv // Iskustvennyj intellekt i prinyatie reshenij. 2014, №2, s. 95–109. (In Russ.).
4. Enikolopov S.N., Kuznetsova YU.M., Smirnov I.V., Stankevich M.A., CHudova N.V. Sozdanie instrumenta avtomaticheskogo analiza teksta v interesakh sotsio-gumanitarnykh issledovaniy. CH. 1. Metodicheskie i metodologicheskie aspekty // Iskustvennyj intellekt i prinyatie reshenij. 2019. № 2, Str. 28-38. DOI 10.14357/20718594190203. (In Russ.).
5. Enikolopov S.N., Kuznetsova YU.M., Minin A.N., Penkina M.YU., Smirnov I.V., Stankevich M.A., CHudova N.V. Osobennosti teksta i psikhologicheskie osobennosti: opyt ehmpiricheskogo komp'yuternogo issledovaniya // Trudy ISA RAN, 2019, № 3 (v pechati). (In Russ.).
6. Zolotova G.A., Onipenko N.K., Sidorova M.YU. Kommunikativnaya grammatika russkogo yazyka. M.: In-t rus. yaz. RAN im. V.V. Vinogradova, 2004. (In Russ.).
7. Kovalyov A.K., Kuznetsova YU.M., Minin A.N., Penkina M.YU., Smirnov I.V., Stankevich M.A., CHudova N.V. Metody vyyavleniya po tekstu psikhologicheskikh kharakteristik avtora (na primere agressivnosti) // Voprosy kiberbezopasnosti. 2019, № 4 (32), c. 72–80. (In Russ.).
8. Litvinova T.A., Litvinova O.A., Ryzhkova E.S., Biryukova E.D., Seredin P.V., Zagorovskaya O.V. Issledovanie vliyaniya pola i psikhologicheskikh kharakteristik avtora na kolichestvennye parametry ego teksta s ispol'zovaniem programmy Linguistic Inquiry and Word Count // Nauchnyj dialog. 2015. № 2 (48). S. 101–109. (In Russ.).
9. Osipov G.S. Priobretenie znaniy intellektual'nymi sistemami: Osnovy teorii i tekhnologii. M.: Nauka, Fizmatlit, 1997. (In Russ.).



10. Osipov G.S., Smirnov I.V., Tikhomirov I.A. Relyatsionno-situatsionnyj metod poiska i analiza tekstov i ego prilozheniya // *Iskusstvennyj intellekt i prinyatie reshenij*. 2008. № 2. S. 3–10. (In Russ.).
11. *Esse. Bol'shaya sovetskaya ehntsiklopediya*. M.: Sovetskaya ehntsiklopediya 1969-1978. (In Russ.).
12. Gupta U., Chatterjee A., Srikanth R., & Agrawal P. A Sentiment-and-Semantics-Based Approach for Emotion Detection in Textual Conversations // *Neu-IR: Workshop on Neural Information Retrieval, SIGIR 2017, ACM*. URL: [arXiv:1707.06996](https://arxiv.org/abs/1707.06996).
13. Pennebaker J., Boyd R., Jordan K., Blackburn K. The development and psychometric properties of LIWC2015. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf.

Информация об авторах

Ковалёв Алексей Константинович, младший научный сотрудник, Федеральный исследовательский центр “Информатика и управление” Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-7309-7382>, e-mail: alexeykkov@gmail.com

Кузнецова Юлия Михайловна, старший научный сотрудник, Федеральный исследовательский центр “Информатика и управление” Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-9380-4478>, e-mail: kuzjum@yandex.ru,

Пенкина Марина Юрьевна, старший преподаватель кафедры общей психологии, Московский государственный психолого-педагогического университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-7046-6963>, e-mail: mpenkina@mail.ru

Станкевич Максим Алексеевич, младший научный сотрудник, Федеральный исследовательский центр “Информатика и управление” Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0003-0705-5832>, e-mail: maxastan95@gmail.com

Чудова Наталья Владимировна, старший научный сотрудник, Федеральный исследовательский центр “Информатика и управление” Российской академии наук (ФИЦ ИУ РАН), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-3188-0886>, e-mail: nchudova@gmail.com

Information about the authors

Alexey K. Kovalev, junior researcher, Federal Research Center ‘Computer Science and Control’ of the Russian Academy of Sciences, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-7309-7382>, e-mail: alexeykkov@gmail.com

Yuliya M. Kuznetsova, PhD, senior researcher, Federal Research Center ‘Computer Science and Control’ of the Russian Academy of Sciences, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-9380-4478>, e-mail: kuzjum@yandex.ru

Marina Y. Penkina, Senior lecturer, Moscow State University of Psychology and Education (MSUPE), Moscow, Russia, ORCID: <https://orcid.org/0000-0001-7046-6963>, e-mail: mpenkina@mail.ru

Maksim A. Stankevich, junior researcher, Federal Research Center ‘Computer Science and Control’ of the Russian Academy of Sciences, Moscow, Russia, ORCID: <https://orcid.org/0000-0003-0705-5832>, e-mail: maxastan95@gmail.com

Natalia V. Chudova, PhD, senior researcher, Federal Research Center ‘Computer Science and Control’ of the Russian Academy of Sciences, Moscow, Russia, ORCID: <https://orcid.org/0000-0002-3188-0886>, e-mail: nchudova@gmail.com

Получена 27.09.2019

Received 27.09.2019

Принята в печать 18.11.2019

Accepted 18.11.2019