

# Программные средства психоэмоционального анализа текстов\*

И. В. Смирнов

Федеральный исследовательский центр «Информатика и управление» Российской академии наук,  
г. Москва, Россия

**Аннотация.** В работе рассматривается проблема психоэмоционального анализа текстов, направленного на выявление психологических особенностей автора текста и выявление эмоциональных характеристик текста на основе методов психолингвистики и искусственного интеллекта. Описан программный инструмент психоэмоционального анализа текстов на русском языке, представлен пример применения инструмента для анализа реакции пользователей ВКонтакте на фейковые сообщения.

**Ключевые слова:** психолингвистический анализ текстов, выявление эмоций, социальные сети, реакция на фейки.

DOI 10.14357/20718632230103

## Введение

В связи с растущим распространением электронных средств коммуникации (социальные сети, мессенджеры и т.д.) становится актуальной проблема анализа содержимого публикуемых и пересылаемых текстовых сообщений. Одним из направлений, представляющих научный и практический интерес, является определение эмоциональной направленности дискуссий, психологического и эмоционального состояния коммуникантов, выявление признаков наличия психологического неблагополучия или личностных особенностей у авторов текстовых сообщений. Решение таких задач необходимо, например, для профилактики психических заболеваний, выявления реакций

пользователей соцсетей и мессенджеров на значимые события, для оценки степени социальной напряженности и других задач обеспечения информационно-психологической безопасности.

В статье рассматриваются существующие постановки задач, методы и подходы в области психоэмоционального анализа текстовых сообщений в сети Интернет с помощью обработки естественного языка и искусственного интеллекта. Представлен программный инструмент психоэмоционального анализа текста TITANIS, описаны его функции и архитектура, представлены результаты применения инструмента для определения эмоциональной реакции пользователей социальной сети ВКонтакте на фейковые сообщения.

\* Работа выполнена в рамках научной программы Национального центра физики и математики (направление №9 «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах»).

## 1. Современные методы психоэмоционального анализа текстов

Под психоэмоциональным анализом текста будем понимать выявление по тексту психологических особенностей автора текста (личностные черты, склонность к различным видам поведения и т.д.), признаков психологического неблагополучия у него (болезненные или предболезненные состояния – депрессивность, шизофрения и т.п.), определение эмоциональных состояний автора или эмоциональной направленности текста в целом (эмоциональная напряженность, агрессивность, фрустрированность и т.п.). Взаимосвязь текстовых характеристик и психологических особенностей довольно давно исследуется клиническими психологами [1] на примере устной и письменной речи. В наше время задачи автоматического психоэмоционального анализа текстов решаются с помощью методов компьютерной лингвистики и искусственного интеллекта. Как правило, постановка задачи сводится к классификации текстов, где классами являются психологические и эмоциональные состояния, а наиболее эффективное решение достигается применением методов машинного обучения и нейросетевых подходов к размеченным корпусам текстов. К особенностям зарубежных психолингвистических исследований можно отнести использование ресурса LIWC [2]. Принцип использования LIWC заключается в определении частотности слов, принадлежащим к категориям «части речи» и «тематические слова» (категории «негативный аффект», «позитивный аффект», «человек», «время» и т.д.). Последние 5 лет преобладают подходы на основе глубокого обучения, но лингвистические ресурсы (словари, тезаурусы) также применяются. Рассмотрим далее некоторые задачи подробнее.

Для определения личностных черт по тексту используются общепринятые подходы [3, 4]: словарные; на основе машинного обучения с текстовыми признаками; с использованием нейросетевых подходов. Среди исследований на примере русского языка отметим работу [5], где использовалось машинное обучение с текстовыми признаками в виде последовательностей слов и весов TF-IDF.

Выявление эмоций в тексте сводится в простейшем случае к определению тональности (положительные, отрицательные, нейтральные эмоции), в более продвинутой постановке определяется эмоциональность на основе одной или нескольких психологических моделей, например, модели эмоций Плутчека, включающей радость, грусть, страх, доверие, ожидание, удивление, злость, неудовольствие. Отметим, что выявляемые эмоции могут относиться не только к автору текста, но и приписываться автором другим субъектам. Для выявления эмоций используется весь спектр подходов - от словарей до нейронных сетей [6, 7]. Отдельно решается задача выявления по тексту агрессии и ее разновидностей (троллинг, буллинг) [8, 9]. Среди российских работ можно отметить [10], где исследуется классификатор русскоязычных интернет-текстов на 8 классов в соответствии с базовыми эмоциями, выделяемыми шведским биологом Гуго Левхеймом. Используются следующие вербальные маркеры, характерные для отдельных классов эмоций: сочетания слов, частотность пунктуационных знаков, наличие лексем из лексико-семантических полей «болезнь», «смерть» и т.п. В работе [11] решается задача выявления эмоций по недискретной шкале на специально созданном наборе текстов. Использовались различные классификаторы, в том числе на основе нейронных сетей была выполнена векторизация текстов с помощью трех наиболее популярных моделей векторного представления. Интерес представляет задача выявления в тексте субъекта переживания и причины упоминаемых эмоций. Эта задача часто решается с помощью методов установления семантических ролей [12, 13].

Выявлению психопатологии или психологического неблагополучия по тексту также посвящено много работ. В основном представляет интерес выявление признаков депрессивности и риска суицида [14], при этом последнее время акцент смещается в сторону анализа текстов социальных сетей [15], что подтверждается в том числе тематическими семинарами и соревнованиями на крупных конференциях по анализу текстов CLEF eRisk, CLPsych, ACL. Как и для других рассматриваемых задач, для определения психологических отклонений у авто-

ров текстов по-прежнему используется LIWC и другие лингвистические ресурсы, машинное обучение [16] и нейросетевые подходы.

Психоэмоциональный анализ текстов применяется сегодня прежде всего для анализа социальных сетей. Довольно часто решается задача определения эмоциональной реакции пользователей социальной сети на заданные события или кризисные ситуации, например на эпидемию COVID-19 [17], террористические атаки [18], массовые насильственные действия [19] и проч. Целью таких исследований является, в том числе, создание информационно-аналитических систем, позволяющих оценить психологическое и эмоциональное состояние населения заданного региона, представленного пользователями социальных сетей.

На сегодняшний день большинство исследований в области психоэмоционального анализа текстов ориентированы на английский язык и лишь немногие из них внедряются в промышленные системы анализа и мониторинга социальных сетей. Многочисленные коммерческие системы эмоционального анализа соцсетей, в том числе российские (YouScan, Brand24, IQBuzz и т.п.), ограничиваются как правило определением окраски текста по шкале позитивный\нейтральный\негативный и ее разновидностям. В данной работе мы представляем технологию и программный инструмент, позволяющий комплексно решать многие задачи психоэмоционального анализа текстов на русском языке.

## 2. TITANIS – инструмент психоэмоционального анализа текстов

Программный инструмент TITANIS, разработанный коллективом исследователей ФИЦ ИУ РАН в научной школе профессора Г.С. Осипова, предлагает набор текстовых параметров и методов обработки естественного языка для психоэмоционального анализа текстов. TITANIS отличается от аналогов тем, что выполняет более «тонкий» и более эффективный психоэмоциональный анализ текстов за счет использования интерпретируемых психолингвистических текстовых показателей, взятых из клинической психиатрии, и учета глубоких

уровней языка, таких как синтаксис, семантика, дискурс. Некоторые из функций TITANIS являются новыми для русского языка, включая учет дискурса, определение фрустрированности, депрессивности, причины эмоций. Методологические основы TITANIS представлены в работах [20, 21]. Опишем основные функциональные и архитектурные особенности инструмента.

### 2.1 Функциональные возможности

*Расчет психолингвистических маркеров.* TITANIS позволяет рассчитывать морфологические и психолингвистические параметры текста, характеризующие эмоциональное состояние автора текста. В состав маркеров входят различные психолингвистические коэффициенты (коэффициент Трейгера, коэффициент логической связности и др.), морфологические параметры текста (употребление различных форм глаголов, употребление местоимений в разном лице и числе, соотношение частей речи и др.), синтаксические параметры связности предложений (глубина синтаксического дерева и т.д.), дискурсивные параметры связности текста (глубина дерева риторического разбора и т.д.). Было показано, что изменение значений маркеров относительно нормы как правило связано с изменением эмоционального состояния автора текста [22, 23].

*Выявление эмоциональной направленности текста.* TITANIS позволяет оценить эмоциональную напряженность автора на момент написания текста, его склонность к эмоциональному или рациональному отношению к обсуждаемому в тексте предмету, готовность к действию (в том числе, агрессивному). Кроме этого TITANIS позволяет оценить общую эмоциональную направленность текста. В состав инструмента входят специализированные словари лексики определенных тематических категорий и лексики, характеризующей отношение автора текста к упоминаемым предметам. На основе данных словарей рассчитывается встречаемость подобной лексики в заданном тексте. Всего в TITANIS используется 21 словарь объемом около 50 тыс. слов. Примеры словарей: аффектогенная лексика (*абсолютно, зверски, неслыханно, топовый и т.п.*); эмоции негативные, стеничные (*возмутиться, зависть, ненавидеть, мерзкий, разозлить и т.п.*); социаль-

ность отрицательная (*автозак, бандит, иск, кланчить, коррумпированный и т.п.*) и т.п.

*Выявление типов эмоционального состояния.* В TITANIS одним из важнейших показателей для определения эмоционального состояния автора является использование эмотивных предикатов. Эмотивные предикаты разделены на четыре семантических подкласса: позитивы, негативы, амбиваленты (выражающие волнение), деэмотивы (выражающие выход из эмоционального состояния). На основе специализированных словарей TITANIS позволяет также выявить в тексте следующие эмоции: страх, злость, печаль, радость, удивление, презрение, стыд, смущение.

*Выявление субъекта и причины эмоций.* TITANIS позволяет определить по тексту *кто* [субъект] и от *чего или почему* [причина] испытывает эмоции, упоминаемые в тексте. Эта задача решается на основе семантического анализа текста, устанавливающего семантические роли в предикатно-аргументных конструкциях при предикатах эмоций [24]. Причину эмоции выражает семантическая роль *каузатор*, субъекта эмоции - семантическая роль *экспериментер*.

*Предсказание наличия признаков депрессии у автора текста по его небольшому сочинению.* TITANIS позволяет получить оценку наличия признаков депрессии у автора текста по небольшому сочинению о себе. Предсказание основано на модели машинного обучения, которая была обучена на сочинениях людей с клинической депрессией и людей без признаков ментальных расстройств. В качестве текстовых признаков для классификации используются в том числе психолингвистические маркеры, описанные ранее [16].

*Предсказание наличия признаков депрессивности у автора текста по текстовым сообщениям социальных сетей.* TITANIS позволяет получить предсказание наличия признаков депрессивности (как черты характера) у автора текста коротких сообщений (социальные сети, мессенджеры). Предсказание основано на машинном обучении на коллекциях текстовых сообщений тех пользователей социальной сети Вконтакте, которые прошли опросник депрессии Бека [25].

*Предсказание наличия состояния фрустрации у автора текста.* Фрустрация (от лат. frustratio — обман, тщетное ожидание) — пси-

хическое состояние, вызванное неуспехом в удовлетворении потребности, желания; наличием барьера на пути достижения цели. TITANIS позволяет получить предсказание, был ли текст написан автором в состоянии фрустрации или нет. Предсказание основывается на модели машинного обучения, которая была обучена на текстовых сообщениях 100 пользователей социальных сетей, написанных пользователями в спокойном или фрустрированном состоянии [21].

*Выявление типа реакции на фрустрацию по Розенцвейгу.* Реакция на фрустрацию – это ответ на ситуацию препятствия, когда планируемое или привычное поведение не может быть реализовано. TITANIS позволяет выявлять: экстрапунитивные реакции – направленные на внешнее окружение (человек осуждает внешнюю причину фрустрации); интропунитивные реакции – направленные на себя (человек испытывает чувство вины и ответственности за исправление сложившейся ситуации); импунитивные реакции – сложившаяся ситуация фрустрации рассматривается человеком как что-то неизбежное, незначительное, он никого не обвиняет [26].

## 2.2. Архитектура

Логическая модульная архитектура TITANIS представлена на Рис. 1. Стрелками обозначены взаимодействия между компонентами. TITANIS основан на открытой библиотеке обработки текстов ISANLP<sup>1</sup>, разработанной в ФИЦ ИУ РАН, которая предоставляет возможность создания лингвистических процессоров и конвейеров (пайплайнов) путем добавления различных обработчиков текстов в унифицированный фреймворк. ISANLP включает несколько базовых разнородных обработчиков языка:

– MORPHOPARSER – компонент морфологического анализа текстов. Позволяет использовать различные сторонние анализаторы, такие как MyStem<sup>2</sup> или UdPipe<sup>3</sup>.

– SYNTAXPARSER – компонент синтаксического анализа. Представляет собой обертку

<sup>1</sup> <https://github.com/IINemo/isanlp>

<sup>2</sup> <https://yandex.ru/dev/mystem/>

<sup>3</sup> <https://github.com/ufal/udpipe>

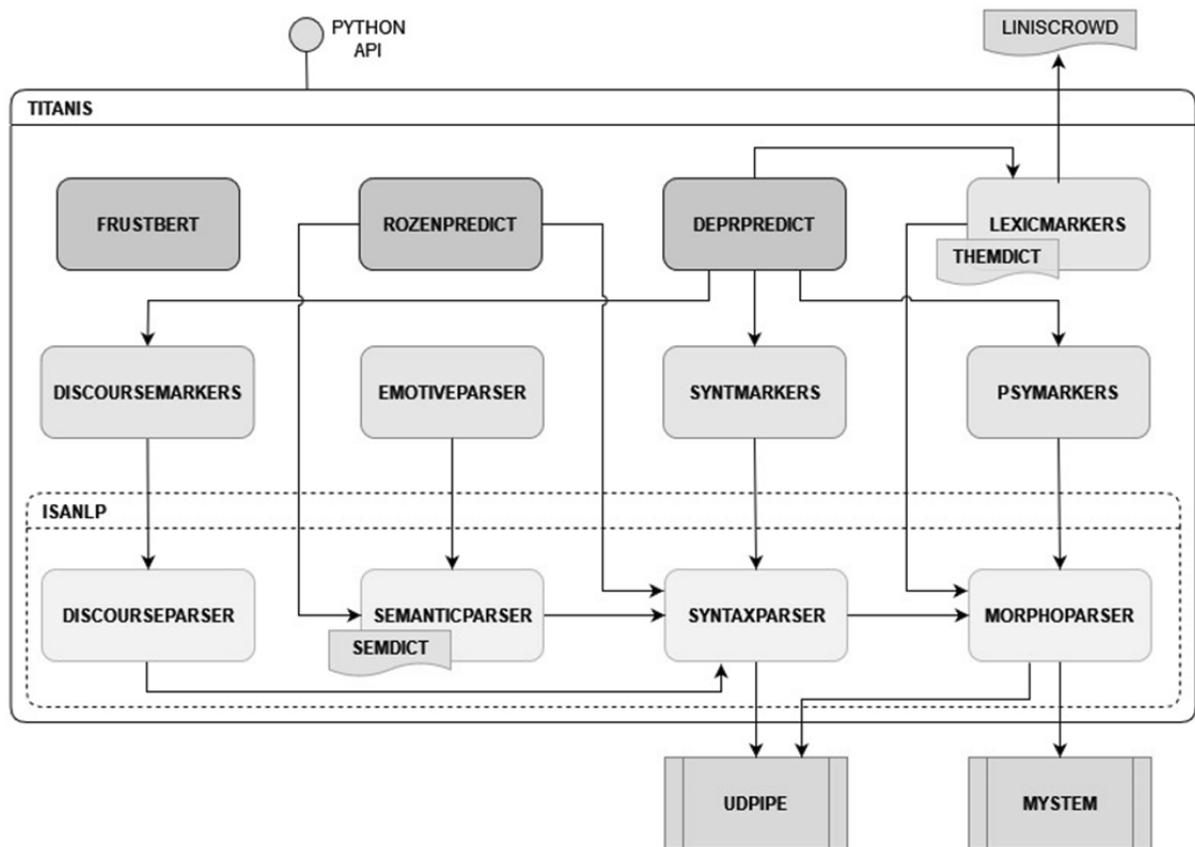


Рис. 1. Модульная архитектура TITANIS

для анализатора UdPipe, но позволяет подключать и другие синтаксические анализаторы.

- SEMANTICPARSER – компонент семантического анализа текстов, выполняющий реляционно-ситуационный анализ и установление семантических ролей [24].

- DISCOURSEPARSER – компонент дискурсивного анализа текстов. Строит дискурсивное дерево разбора текста на основе теории риторических структур [27].

Перечислим основные модули TITANIS:

- PSYMARKERS – отвечает за вычисление психолингвистических маркеров – различных текстовых показателей, информативных для определения психоэмоциональных состояний авторов текстов.

- SYNTMARKERS – вычисляет синтаксические показатели текста, потенциально влияющие на определение психоэмоциональных состояний авторов текстов.

- DISCOURSEMARKERS – компонент, отвечающий за вычисление дискурсивных марке-

ров, характеризующих риторическую структуру текста.

- LEXICMARKERS – отвечает за вычисление лексических (словарных) маркеров, характеризующих тематическую и эмоциональную направленность текста.

- EMOTIVEPARSER – отвечает за выявление субъекта и причины эмоций. Использует семантический анализатор для выявления экспериментера и каузатора при эмотивных предикатах.

- DEPRPREDICT – компонент, предсказывающий наличие признаков депрессии у автора текста по его небольшому сочинению или признакам депрессивности у автора текста по текстовым сообщениям социальных сетей.

- ROZENPREDICT – отвечает за выявление типа реакции на фрустрацию по Розенцвейгу на основе синтаксической и семантической структуры текста.

- FRUSTBERT – предсказывает наличие состояния фрустрации у автора текста. Обучен на текстовых сообщениях социальных сетей.

Используемые в TITANIS сторонние библиотеки обработки текстов имеют открытый исходный код и распространяются по некоммерческим лицензиям. TITANIS представляет собой набор легко развертываемых докеров, наличие Python API обеспечивает легкую интеграцию. Открытая версия с ограниченными возможностями (сокращены словари, отсутствуют некоторые функции) доступна по ссылке <https://github.com/tchewik/titanis-open>.

### 3. Применение TITANIS для анализа реакции пользователей ВКонтакте на фейковые сообщения

Инструмент TITANIS применялся для психолингвистического анализа русскоязычных комментариев к фейковым новостям в социальной сети ВКонтакте. Целью исследования является выявление текстовых показателей, характерных для обсуждения фейковых сообщений и отражающих психоэмоциональную реакцию пользователей ВКонтакте на фейковые сообщения. Схема исследования включает поиск статистически значимых различий в текстовых показателях двух выборок текстов с дальнейшей интерпретацией различий.

Для эмпирического исследования экспертами были отобраны фейковые сообщения (посты) по 15 тематикам, затем с помощью API<sup>4</sup> ВКонтакте автоматически загружались все комментарии к этим сообщениям. Всего собрано около 288 тыс. комментариев. Число уникальных пользователей, которые писали комментарии под фейками составляет 148157. В качестве фоновой (контрольной) выборки была собрана коллекция русскоязычных комментариев к роликам YouTube с развлекательным и кулинарным контентом. Предполагается, что комментарии к таким сообщениям психоэмоционально нейтральны и уровень их эмоциональности соответствуют норме. Объем фоновой выборки составил около 300 тыс. комментариев. Тексты обеих коллекций относятся к одному жанру и стилю, что позволяет их сравнивать по тематически независимым текстовым показателям.

Собранные сообщения подвергались процедуре очистки, направленной на удаление нетипичных комментариев. После процедур очистки данных итоговый объем выборки составил 549257 комментариев: 262313 комментариев к фейкам и 286944 фоновых комментариев. Очищенные комментарии обрабатывались далее с помощью описанного выше инструмента TITANIS для получения текстовых показателей.

Опишем далее наиболее значимые результаты анализа.

#### 3.1. Общая характеристика реакций на фейковые сообщения

Статистически значимые различия в текстовых показателях двух выборок текстов определялись с помощью критерия  $\chi^2$ . Было выявлено более 50 показателей, показывающих различия с уровнем значимости  $p < 0,05$ . На Рис. 2 представлены статистически значимые различия встречаемости некоторых текстовых признаков в комментариях к фейковым сообщениям по сравнению с фоновым корпусом.

Полученные статистические данные свидетельствуют о том, что текстовые особенности комментариев фейковых сообщений имеют отчетливый и системный характер, согласованно проявляясь на лексическом, синтаксическом, семантическом уровнях речевой деятельности их авторов. Прежде всего, обращает на себя внимание предпочтение авторами комментариев к фейкам лексики с определенной семантикой: выражающей негативные эмоции, негативную эмоциональную и рациональную оценку, описывающей деструктивные действия и отрицательные явления социальной жизни. При этом слова с семантикой позитивных эмоций, положительной рациональной оценки, положительной социальности и мотивации к действию встречаются в данном подкорпусе значимо реже по сравнению с фоновым. Для фейковых комментариев в большей степени характерно также употребление речевых средств, провоцирующих у читателя стеничные отрицательные аффекты: оскорбляющие («Инвективы»), обесценивающие («Мягкие инвективы»), шокирующие («Молодежный жаргон») слова и выражения.

<sup>4</sup> <https://dev.vk.com/reference>

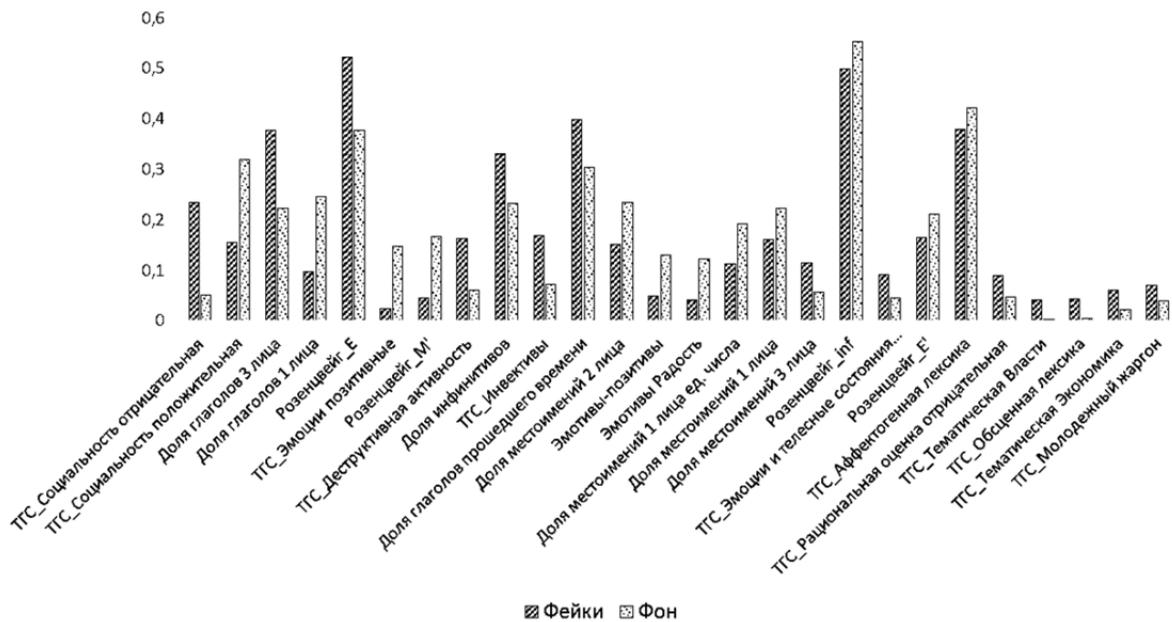


Рис. 2. Средние значения психолингвистических показателей для комментариев к фейкам и фоновой коллекции

### 3.2. Особенности реакции на фейковые сообщения различной тематики

В Табл. 1 представлены данные о количестве комментариев к фейковым сообщениям на разные темы в процентах от общего числа комментариев (тематика фейковых сообщений

была определена экспертами). Наиболее сильный отклик (почти 2/3 комментариев) вызывают фейки, посвященные трем темам: «Политическая сфера, силовики и государственная безопасность» (29.83%), «Социально значимые происшествия» (18.85%) и «Здравоохранение» (13.24%).

Табл. 1. Количество комментариев к фейковым сообщениям на разные темы

Тематика	Количество комментариев	Процент от общего числа
Политическая сфера, силовики и государственная безопасность	86059	29.83%
Социально значимые происшествия	54379	18.85%
Здравоохранение	38186	13.24%
Образование	26219	9.09%
Экология	17848	6.19%
Социальная защита и трудовая сфера	16816	5.83%
Коронавирусная инфекция (COVID-19)	16575	5.75%
Судебная система	12067	4.18%
Спорт и международные события	6609	2.29%
ЖКХ Благоустройство и стройкомплекс	6586	2.28%
Дорожно-транспортная сфера	2965	1.03%
Промышленная сфера	2145	0.74%
Информационные технологии; интернет и СМИ	1730	0.60%
Культура	289	0.10%
Международная и межконфессиональная среда	37	0.01%

Анализ текстовых психолингвистических показателей для комментариев методами статистики (различия по T-test и различия по  $\chi^2$ ) выявил следующее. Комментирование фейковых сообщений на темы, составляющие первую тройку, сопровождается повышенным эмоциональным и коммуникативным напряжением (отличаются психолингвистические коэффициенты, другие грамматические и частеречные показатели аффективного возбуждения, лексика общения). Сообщения на эти темы комментируются высказываниями, содержащими агрессию и обвинения.

На Рис. 3 показаны статистически значимые различия между темами «Здравоохранение» и «Политическая сфера, силовика и гос. безопасность» по показателям из группы «Эмотивы» (показатели нормированы по максимальным значениям среди тем). Видно, что комментариям по теме «Здравоохранение» характерны больше эмоции «страх», «печаль», «покой» и «радость», политическая тема характеризуется эмоциями «отвращение», «стыд» и «гнев».

Сравнение тем из «большой тройки» между собой показало, что фейковые сообщения на тему здравоохранения вызывают особое напряжение и переживание барьера, непреодолимого препятствия. Фейки на тему власти по сравнению с фейками на тему происшествий вызывает всплеск требований к властям разрешить проблему, в то

время как фейки, посвящённые происшествиям, актуализируют защитную реакцию отрицания («ничего страшного»).

### 3.3. Анализ фрустрационного реагирования на фейковые сообщения различной тематики

Было выявлено, что фейки оказывают вполне систематическое действие на агрессивность реакций пользователей – в комментариях к фейковым сообщениям на любые темы значимо чаще встречаются агрессия (экстрапунитивный тип фрустрационного реагирования), чем в комментариях к обычным сообщениям из фонового корпуса. При этом также систематически снижается готовность реагировать спокойно, видя в комментируемых высказываниях лишь что-то обычное, привычное, нестрашное (число импунитивных реакций с фиксацией на препятствии значимо падает в исследуемом корпусе по сравнению с фоновым). Средние значения показателей экстрапунитивных и импунитивных реакций в комментариях разных тематик и фонового корпуса представлены в Табл. 2.

Необходимо отметить, что в данном случае не исследовался фактор тематики, возможно, любые сообщения на указанные темы вызывают такую реакцию пользователей соцсетей по сравнению с фоном. Изучение этого вопроса будет предметом дальнейших исследований.

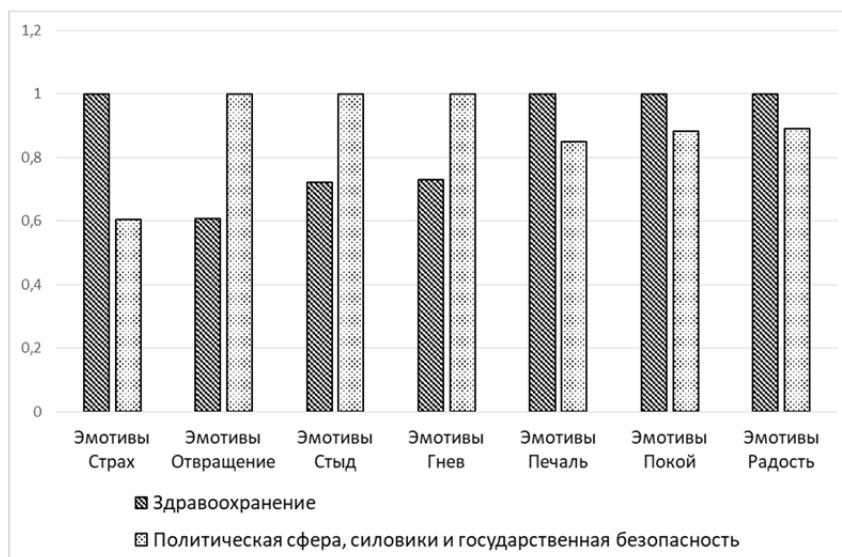


Рис. 3. Нормированные средние текстовые показатели из группы «Эмотивы» для комментариев к фейковым сообщениям на темы «Здравоохранение» и «Политическая сфера, силовика и гос. безопасность»

Табл. 2. Средние значения показателей экстрапунитивных и импунитивных реакций для тематик фейковых комментариев

Тематика фейков	Экстрапунитивные	Импунитивные
Комментарии YouTube (фон)	0,378	0,166
Информационные технологии; интернет и СМИ	0,428	0,028
Спорт и международные события	0,48	0,049
ЖКХ Благоустройство и стройкомплекс	0,485	0,056
Дорожно-транспортная сфера	0,485	0,056
Промышленная сфера	0,492	0,048
Социальная защита и трудовая сфера	0,493	0,0484
Коронавирусная инфекция (COVID-19)	0,498	0,047
Экология	0,508	0,037
Политическая сфера, силовики и государственная безопасность	0,526	0,042
Социально значимые происшествия	0,528	0,046
Здравоохранение	0,535	0,049
Образование	0,536	0,042
Судебная система	0,558	0,046

## Заключение

Психоэмоциональный анализ текстов является актуальным направлением научных исследований и прикладных разработок. Представленный в статье инструмент TITANIS реализует оригинальные функции психоэмоционального анализа текстов на русском языке и позволяет получать полезную информацию об эмоциональном и психологическом состоянии авторов текстов и пользователей социальных сетей, а также оценивать эмоциональную направленность текстов. Применение инструмента TITANIS показало, что фейковые сообщения вызывают рост напряжения, беспокойства и агрессии, приводят к повышению эмоционального возбуждения, спаду позитивных эмоций и положительных оценок происходящего.

На основе TITANIS планируется разработка систем, предназначенных для оперативного отслеживания реакции населения, представленного пользователями социальных сетей и мессенджеров, на законодательные, административные и хозяйственные инициативы и принимаемые решения, проекты, законы, на чрезвычайные ситуации. Использование таких систем будет способствовать повышению качества обратной связи в системе «народ - власть», обеспечит возможности для превентивной про-

работки социальных проблем и своевременное вмешательство в уже существующие или потенциальные конфликты, что необходимо для предотвращения роста социальной напряженности.

## Литература

1. Ениколопов С.Н., Медведева Т.И., Воронцова О.Ю. Оценка текстов, написанных больными эндогенными психическими заболеваниями // Психиатрия. – 2019. – Т. 81, № 1. – С. 56–64.
2. Pennebaker J. W., Francis M. E., Booth R. J. Linguistic inquiry and word count: LIWC 2001 // Mahway: Lawrence Erlbaum Associates. – 2001. – V. 71. – №. 2001. – p. 2001.
3. Štajner S., Yenikent S. A survey of automatic personality detection from texts // Proceedings of the 28th International Conference on Computational Linguistics. – 2020. –pp. 6284-6295.
4. Ahmad H., Asghar M. Z., Khan A. S., Habib A. A systematic literature review of personality trait classification from textual content // Open Computer Science. – 2020. – V. 10. – №. 1. – pp. 175-193.
5. Ярушкина Н. Г., Мошкин В. С., Андреев И. А. Алгоритм психолингвистического анализа текстовых данных социальных сетей с применением модели «Большая пятёрка» // Онтология проектирования. – 2022. – Т. 12. – №. 1 (43). – С. 82-92.
6. Acheampong F. A., Wenyu C., Nunoo-Mensah H. Text-based emotion detection: Advances, challenges, and opportunities // Engineering Reports. – 2020. – V. 2. – №. 7. – p. e12189.

7. Nandwani P., Verma R. A review on sentiment analysis and emotion detection from text // *Social Network Analysis and Mining*. – 2021. – V. 11. – №. 1. – pp. 1-19.
8. Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi, and Marcos Zampieri. Benchmarking Aggression Identification in Social Media // *Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018)*. – 2018. – pp. 1-11. Association for Computational Linguistics.
9. Faneva Ramiandrisoa, Josiane Mothe. Aggression Identification in Social Media: a Transfer Learning Based Approach // *Second Workshop on Trolling, Aggression and Cyberbullying*, European Language Resources Association (ELRA), May 2020, Marseille, France. pp.26-31.
10. Колмогорова А. В., Вдовина Л. А. Лексико-грамматические маркеры эмоций как параметры для sentiment анализа русскоязычных интернет-текстов // *Вестник Пермского университета. Российская и зарубежная филология*. – 2019. – Т.11. – №3. – С. 38–46.
11. Колмогорова А. В., Калинин А. А. Эмоциональный анализ постов ВКонтакте: классификатор или регрессор? // *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог»*. Москва. – 2022. – Вып. 21. – С. 311-322.
12. Ghazi D., Inkpen D., Szpakowicz S. Detecting emotion stimuli in emotion-bearing sentences // *International Conference on Intelligent Text Processing and Computational Linguistics*. – Springer, Cham. – 2015. – pp. 152-165.
13. Campagnano C., Conia S., Navigli R. SRL4E–Semantic Role Labeling for Emotions: A Unified Evaluation Framework // *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. – 2022. – pp. 4586-4601
14. Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. Natural language processing in mental health applications using non-clinical texts // *Natural Language Engineering*. – 2017. – V. 23. – №. 5. – pp. 649-685.
15. Kayalvizhi S. et al. Findings of the Shared Task on Detecting Signs of Depression from Social Media // *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. – 2022. – pp. 331-338.
16. M. Stankevich, I. Smirnov, Y. Kuznetsova, N. Kiselnikova, S. Enikolopov. Predicting Depression from Essays in Russian // *In Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019"*. – 2019. – pp. 637–647.
17. Mathur, P. Kubde and S. Vaidya, "Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19 // *5th International Conference on Communication and Electronics Systems (ICCES)*. – 2020. – pp. 845-848.
18. Becker K., Harb J. G., Ebeling R. Exploring deep learning for the analysis of emotional reactions to terrorist events on twitter // *Journal of Information and Data Management*. – 2019. – V. 10. – №. 2. – pp. 97-115.
19. Harb J. G. D., Ebeling R., Becker K. A framework to analyze the emotional reactions to mass violent events on Twitter and influential factors // *Information Processing & Management*. – 2020. – V. 57. – №. 6. – p. 102372.
20. Ениколопов С. Н., Кузнецова Ю. М., Осипов Г. С., Смирнов И. В., Чудова Н. В. Метод релеяционно-ситуационного анализа текста в психологических исследованиях // *Психология. Журнал Высшей школы экономики*. – 2021. – Т. 18. – №4. – С. 748–769.
21. Smirnov I., Stankevich M., Kuznetsova Y., Suvorova M., Larionov D., Nikitina E., Savelov M., Grigoriev O. TITANIS: A Tool for Intelligent Text Analysis in Social Media // *In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science, Springer, Cham*. – 2021. – V. 12948. – pp.232-247.
22. С.Н. Ениколопов, Ю.М. Кузнецова, А.Н. Минин, М.Ю. Пенкина, И.В. Смирнов, М.А. Станкевич, Н.В. Чудова. Особенности текста и психологические особенности: опыт эмпирического компьютерного исследования // *Труды Института Системного Анализа Российской академии наук*. – 2019. – Том 69. – №3. – С. 91-99.
23. Ениколопов С.Н., Ковалев А.К., Кузнецова Ю.М., Старостина Е.Н., Чудова Н.В. Признаки, характерные для письменных текстов, написанных в состоянии фрустрации // *Вестник МГУ. Серия 14. Психология*. – 2019. – №3. – С. 66–85.
24. Shelmanov A. O., Smirnov I. V., Methods for Semantic Role Labeling of Russian Texts // *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2014)*. Issue 13 (20). – 2014. – pp. 580-592.
25. Stankevich M., Latyshev A., Kuminskaya E., Smirnov I., Grigoriev O. Depression Detection from Social Media Texts // *Proceedings of Data Analytics and Management in Data Intensive Domains: XXI International Conference DAMDID/RCDL'2019* – 2019. – pp. 352-362.
26. Devyatkin, Dmitry A., Natalia V. Chudova, Anfisa A. Chuganskaya and Daria Sharypina. Methods for Recognition of Frustration-Derived Reactions on Social Media // *In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science, Springer, Cham*. – V. 12948. – pp. 17-30.
27. Chistova E., Shelmanov A., Pisarevskaya D., Kobozeva M., Isakov V., Panchenko A., Toldova S., Smirnov I. RST Discourse Parser for Russian: an Experimental Study of Deep Learning Models // *International Conference on Analysis of Images, Social Networks and Texts*. – *Lecture Notes in Computer Science, Springer, Cham*. – 2021. – V.12602. – pp.105-119.

**Смирнов Иван Валентинович.** Федеральный исследовательский центр "Информатика и управление" Российской академии наук, г. Москва, Россия. Заведующий отделом «Интеллектуальный анализ информации», кандидат физико-математических наук, доцент. Количество печатных работ: 120. Область научных интересов: интеллектуальный анализ текстов, искусственный интеллект. E-mail: ivs@isa.ru.

## Software for Psycho-Emotional Text Processing

I. V. Smirnov

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

**Abstract.** The paper considers the problem of psycho-emotional text processing, aimed at identifying the psychological characteristics of the author of the text and identifying the emotional characteristics of the text based on methods of psycholinguistics and artificial intelligence. A tool for psycho-emotional analysis of texts in Russian is described as well as application of the tool to analysis of the VKontakte users' reaction to fake messages is presented.

**Keywords:** psycholinguistic text processing, emotion detection, social networks, reaction to fake.

**DOI** 10.14357/20718632230103

### References

- Enikolopov, S.N., Medvedeva, T.I., Vorontcova, O.Y. 2019. Ocenka tekstov, napisannyh bol'nymi endogennymi psihicheskimi zabolevaniyami [Evaluation of texts written by patients with endogenous mental illnesses]. *Psihiatriya* [Psychiatry]. 81:56–64.
- Pennebaker J. W., Francis M. E., Booth R. J. Linguistic inquiry and word count: LIWC 2001 // Mahway: Lawrence Erlbaum Associates. – 2001. – V. 71. – №. 2001. – p. 2001.
- Štajner S., Yenikent S. A survey of automatic personality detection from texts // Proceedings of the 28th International Conference on Computational Linguistics. – 2020. –pp. 6284-6295.
- Ahmad H., Asghar M. Z., Khan A. S., Habib A. A systematic literature review of personality trait classification from textual content // *Open Computer Science*. – 2020. – V. 10. – №. 1. – pp. 175-193.
- YArushkina, N. G., Moshkin, V. S., Andreev, I. A. 2022. Algoritm psiholingvisticheskogo analiza tekstovykh dannykh social'nykh setej s primeneniem modeli «Bol'shaya pyatyorka» [Algorithm of psycholinguistic analysis of text data of social networks using the "Big Five" model]. *Ontologiya proektirovaniya* [Design Ontology] 1(43):82-92.
- Acheampong F. A., Wenyu C., Nunoo-Mensah H. Text-based emotion detection: Advances, challenges, and opportunities // *Engineering Reports*. – 2020. – V. 2. – №. 7. – p. e12189.
- Nandwani P., Verma R. A review on sentiment analysis and emotion detection from text // *Social Network Analysis and Mining*. – 2021. – V. 11. – №. 1. – pp. 1-19.
- Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi, and Marcos Zampieri. Benchmarking Aggression Identification in Social Media // Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018). – 2018. – pp. 1-11. Association for Computational Linguistics.
- Faneva Ramiandrisoa, Josiane Mothe. Aggression Identification in Social Media: a Transfer Learning Based Approach // Second Workshop on Trolling, Aggression and Cyberbullying, European Language Resources Association (ELRA), May 2020, Marseille, France. pp.26-31.
- Kolmogorova, A.V., Vdovina, L.A. 2019. Leksikogrammaticheskie markery emocij kak parametry dlya sentiment analiza russkoyazychnykh internet-tekstov [Lexicogrammatical markers of emotions as parameters for sentiment analysis of Russian-language Internet texts]. *Vestnik Permskogo universiteta. Rossijskaya i zarubezhnaya filologiya* [Bulletin of Perm University. Russian and foreign philology]. 11(3):38-46.
- Kolmogorova, A.V., Kalinin ,A.A. 2022. Emocional'nyj analiz postov VKontakte: klassifikator ili regressor? [Emotional analysis of VKontakte posts: classifier or regressor?]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoj mezhdunarodnoj konferencii «Dialog»*. Moskva [Computational linguistics and intelligent technologies: Based on the materials of the annual international conference "Dialogue". Moscow] 21:311-322.
- Ghazi D., Inkpen D., Szpakowicz S. Detecting emotion stimuli in emotion-bearing sentences // International Conference on Intelligent Text Processing and Computational Linguistics. – Springer, Cham. – 2015. – pp. 152-165.
- Campagnano C., Conia S., Navigli R. SRL4E–Semantic Role Labeling for Emotions: A Unified Evaluation Framework // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2022. – pp. 4586-4601
- Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. Natural language processing in mental health applications using non-clinical texts // *Natural Language Engineering*. – 2017. – V. 23. – №. 5. – pp. 649-685.
- Kayalvizhi S. et al. Findings of the Shared Task on Detecting Signs of Depression from Social Media // Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion. – 2022. – pp. 331-338.
- M. Stankevich, I. Smirnov, Y. Kuznetsova, N. Kiselnikova, S. Enikolopov. Predicting Depression from Essays in Russian // In Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019". – 2019. – pp. 637–647.

17. Mathur, P. Kubde and S. Vaidya, "Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19 // 5th International Conference on Communication and Electronics Systems (ICCES). – 2020. – pp. 845-848.
18. Becker K., Harb J. G., Ebeling R. Exploring deep learning for the analysis of emotional reactions to terrorist events on twitter // *Journal of Information and Data Management*. – 2019. – V. 10. – №. 2. – pp. 97-115.
19. Harb J. G. D., Ebeling R., Becker K. A framework to analyze the emotional reactions to mass violent events on Twitter and influential factors // *Information Processing & Management*. – 2020. – V. 57. – №. 6. – p. 102372.
20. Enikolopov, S.N., Kuznecova, YU.M., Osipov, G.S., Smirnov, I.V., Chudova, N.V. 2021. Metod relyacionno-situacionnogo analiza teksta v psihologicheskikh issledovaniyah [The method of relational-situational text analysis in psychological research]. *Psihologiya. Zhurnal Vysshej shkoly ekonomiki* [Psychology. Journal of the Higher School of Economics] 18(4):748-769.
21. Smirnov I., Stankevich M., Kuznetsova Y., Suvorova M., Larionov D., Nikitina E., Savelov M., Grigoriev O. TITANIS: A Tool for Intelligent Text Analysis in Social Media // In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) *Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science*, Springer, Cham. – 2021. – V. 12948. – pp.232-247.
22. Enikolopov, S.N., Kuznecova, YU.M., Minin, A.N., Penkina, M.YU., Smirnov, I.V., Stankevich, M.A., Chudova, N.V. 2019. Osobennosti teksta i psihologicheskie osobennosti: opyt empiricheskogo komp'yuternogo issledovaniya [Text features and psychological features: the experience of empirical computer research]. *Trudy Instituta Sistemnogo Analiza Rossijskoj akademii nauk* [Proceedings of the Institute for System Analysis of the Russian Academy of Sciences] 69(3):91-99.
23. Enikolopov, S.N., Kovalev, A.K., Kuznecova, YU.M., Starostina, E.N., Chudova, N.V. 2019. Priznaki, harakternye dlya pis'mennyh tekstov, napisannyh v sostoyanii frustracii [Characteristic of texts written in a state of frustration]. *Vestnik MGU. Seriya 14. Psihologiya* [Bulletin of Moscow State University. Series 14. Psychology] 3:66-85.
24. Shelmanov A. O., Smirnov I. V., *Methods for Semantic Role Labeling of Russian Texts // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2014). Issue 13 (20).* – 2014. – pp. 580-592.
25. Stankevich M., Latyshev A., Kuminskaya E., Smirnov I., Grigoriev O. Depression Detection from Social Media Texts // *Proceedings of Data Analytics and Management in Data Intensive Domains: XXI International Conference DAMDID/RCDL'2019 – 2019.* – pp. 352-362.
26. Devyatkin, Dmitry A., Natalia V. Chudova, Anfisa A. Chuganskaya and Daria Sharypina. *Methods for Recognition of Frustration-Derived Reactions on Social Media // In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science*, Springer, Cham. – V. 12948. – pp. 17-30.
27. Chistova E., Shelmanov A., Pisarevskaya D., Kobozeva M., Isakov V., Panchenko A., Toldova S., Smirnov I. RST Discourse Parser for Russian: an Experimental Study of Deep Learning Models // *International Conference on Analysis of Images, Social Networks and Texts.* – *Lecture Notes in Computer Science*, Springer, Cham. – 2021. – V.12602. – pp.105-119.

**Smirnov I. V.** PhD, Associate Professor, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia, e-mail: [ivs@isa.ru](mailto:ivs@isa.ru)