

# TITANIS API

TITANIS предоставляет функции вычисления текстовых параметров и обработки естественного языка для оценки психоэмоционального состояния автора текста или группы авторов. Помимо широко используемых подходов обработки естественного языка, таких как tf-idf и анализ тональности, TITANIS включает в себя психолингвистический, семантический, дискурсивный и другие виды анализа, которые позволяют обнаруживать различия в текстах пользователей с разными психоэмоциональными состояниями.

## 1 Внешний интерфейс

### Входные данные

Параметр	Тип	Default	Требования к другим параметрам	Описание
text	string			Непустой текст на русском языке в кодировке UTF-8. Для каждой функции установлен свой минимальный размер текста в символах.
<b>Оptionальные параметры, определяющие список вызываемых обработчиков</b>				
psy_cues	bool	False		Рассчитывать психолингвистические/морфологические маркеры
psy_cues_normalization	string	words	psy_cues=True	Условия нормализации для психолингвистических/морфологических маркеров
psy_dict	bool	False		Расчет словарных маркеров
psy_dict_normalization	string	words	psy_dict=True	Условия нормализации для словарных маркеров
syntax	bool	False		Расчет синтаксических признаков
discourse	bool	False		Расчет дискурсивных признаков
emotive_srl	bool	False		Выявление семантических ролей при эмотивных предикатах.

<code>frustration_clf</code>	bool	False		Предсказать состояние фрустрации у автора текста
<code>depression_clf</code> *	bool	False		Предсказать депрессивность автора текста по формату эссе
<code>rosenzweig</code> *	bool	False		Выявление типа реакции на фрустрацию по Розенцвейгу

## Выходные данные

Результат имеет формат словаря с ключами для обработчиков:

- 'PsyCues' – [Психолингвистические/морфологические маркеры](#)
- 'PsyDict' – [Словарные маркеры](#)
- 'Syntax' – [Синтаксические признаки](#)
- 'Discourse' – [Дискурсивные признаки](#)
- 'EmotiveSRL' – [Семантические роли при эмотивных предикатах](#)
- 'Frustration' – [Оценка состояния фрустрации у автора текста](#)
- 'Depression' – [Оценка депрессивности автора эссе](#) \*
- 'Rosenzweig' – [Тип реакции на фрустрацию по Розенцвейгу](#) \*

\* обработчик доступен только в расширенной версии библиотеки.

## 2 Внутренний интерфейс обработчиков

### 2.1 Расчет психолингвистических/морфологических маркеров

#### Входные данные

Параметр	Тип	Описание
text	String	Непустой текст на русском языке
normalization	string, {'words', 'sentences', 'abs'}, default: 'words'	Условия нормализации для маркеров по частям речи: 'words' - по кол. слов; 'sentences' - по кол. предложений; 'abs' - абсолютные значения.

#### Основа

Леммы слов, морфологическая аннотация mystem, словарь тональности linis-crowd.

## Используемые анализаторы

Название	Ключи
MyStem	lemma, postag (не конвертированный)

## Выходные данные

Параметр	Тип	Описание
char_count	Int	Число символов
word_count	Int	Число слов
sentence_count	Int	Число предложений
unique_words_count	Int	Число уникальных слов
punctuation_count	Int	Число знаков препинания
punctuation_per_word	float	Число знаков пунктуации / Число слов
uppercase_rate	float	Доля прописных символов
mean_word_len	float	Средняя длина слов (в количестве символов)
mean_sentence_len	float	Средняя длина предложения (в количестве слов)
unique_words_rate	float	Число уникальных слов / Число слов
verbs_1p_rate	float	Доля глаголов 1 лица
verbs_2p_rate	float	Доля глаголов 2 лица
verbs_3p_rate	float	Доля глаголов 3 лица
verbs_past_tense_rate	float	Доля глаголов прошедшего времени
infinitives_rate	float	Доля инфинитивов
pro_1p_rate	float	Доля местоимений 1 лица
pro_1p_sing_rate	float	Доля местоимений 1 лица ед. числа
pro_1p_plural_rate	float	Доля местоимений 1 лица мн. числа
pro_2p_rate	float	Доля местоимений 2 лица
pro_3p_rate	float	Доля местоимений 3 лица
trager_coef	float	Коэффициент Трейгера
logical_coh_coef	float	Коэффициент логической связности
verbs_per_nouns_coef	float	Коэффициент опредмеченности действия (кол.

		глаголов / кол. существительных)
participles_gerunds_coef	float	Причастия + деепричастия / Число предложений
negation_rate	float	Число отрицательных приставок и форм (не,ни,бес,без) / Число слов
postag_A	float	Часть речи: прилагательное
postag_ADV	float	Часть речи: наречие
postag_ADVPRO	float	Часть речи: местоименное наречие
postag_ANUM	float	Часть речи: числительное-прилагательное
postag_APRO	float	Часть речи: местоимение-прилагательное
postag_COM	float	Часть речи: часть композита - сложного слова
postag_CONJ	float	Часть речи: союз
postag_INTJ	float	Часть речи: междометие
postag_NUM	float	Часть речи: числительное
postag_PART	float	Часть речи: частица
postag_PR	float	Часть речи: предлог
postag_S	float	Часть речи: существительное
postag_SPRO	float	Часть речи: местоимение-существительное
postag_V	float	Часть речи: глагол

## 2.2 Расчет словарных маркеров

### Входные данные

Параметр	Тип	Описание
text	String	Непустой текст на русском языке
normalization	string, {'words', 'sentences','abs'}, default: 'words'	Условия нормализации встречаемости слов из словарей: 'words' - по кол. слов; 'sentences' - по кол. предложений; 'abs' - абсолютные значения.

### Основа

Леммы слов и словари.



## Используемые анализаторы

Название	Ключи
MyStem	Lemma

## Выходные данные

Параметр	Тип	Описание
<p>Наименования словарей:</p> <ol style="list-style-type: none"> <li>1. tgw_positive_assessment</li> <li>2. tgw_positive_social</li> <li>3. tgw_positive_emotions</li> <li>4. tgw_negative_assessment</li> <li>5. tgw_negative_social</li> <li>6. tgw_negative_emotions</li> <li>7. tgw_motivation_activity</li> <li>8. tgw_cognitive_communication</li> <li>9. tgw_destructive_activity</li> <li>10. tgw_affect_lex</li> <li>11. tgw_bodily_states_emotions</li> <li>12. tgw_invectives</li> <li>13. tgw_soft_invectives</li> <li>14. tgw_obscene_lex</li> <li>15. tgw_youth_jargon</li> <li>16. tgw_hcs</li> <li>17. tgw_economics</li> <li>18. tgw_catastrophes</li> <li>19. tgw_security_structures</li> <li>20. tgw_healthcare_demography_ecology</li> <li>21. tgw_authority</li> </ol>	float	<p>Частотность встречаемости слов из словарей:</p> <ol style="list-style-type: none"> <li>1. Рациональная оценка положительная</li> <li>2. Социальность положительная</li> <li>3. Эмоции позитивные</li> <li>4. Рациональная оценка отрицательная</li> <li>5. Социальность отрицательная</li> <li>6. Эмоции негативные, стеничные</li> <li>7. Мотивация, деятельность и напряжение</li> <li>8. Когнитивная деятельность и коммуникация</li> <li>9. Деструктивная активность</li> <li>10. Аффектогенная лексика</li> <li>11. Эмоции и телесные состояния негативные пассивные</li> <li>12. Инвективы</li> <li>13. Мягкие инвективы</li> <li>14. Обценная лексика</li> <li>15. Молодежный жаргон</li> <li>16. Тематическая ЖКХ</li> <li>17. Тематическая Экономика</li> <li>18. Тематическая Катастрофы</li> <li>19. Тематическая Силовые структуры</li> <li>20. Социальность предметная Здравоохранение, демография и экология</li> <li>21. Тематическая Власти</li> </ol>
positive	float	Частотность позитивных эмотивов в тексте
negative	float	Частотность негативных эмотивов в тексте
ambivalent	float	Частотность амбивалентов в тексте

de_emotives	float	Частотность деэмотивов в тексте
sentiment_rate	float	Тональность текста на основе словаря linis-crowd

## 2.3 Расчет синтаксических признаков

### Входные данные

Параметр	Тип	Описание
text	String	Непустой текст на русском языке
normalization	string, {'words', 'sentences', 'abs'}, default: 'words'	Условия нормализации для маркеров по частям речи: 'words' - по кол. слов; 'sentences' - по кол. предложений; 'abs' - абсолютные значения.

### Основа

Синтаксическая аннотация.

### Используемые анализаторы

Название	Ключи
UDPipe	syntax_dep_tree

### Выходные данные

Параметр	Тип	Описание
max_synt_tree	int	Максимальная глубина синтаксического дерева
min_synt_tree	int	Минимальная глубина синтаксического дерева
mean_synt_tree	float	Средняя глубина синтаксического дерева
Наименование синтаксических связей: 1. synt_rel_flat:foreign: 2. synt_rel_csubj 3. synt_rel_acl 4. synt_rel_acl:relcl	float	Частота встречаемости синтаксических связей в тексте: 1. Синтакс. связь: flat:foreign: 2. Синтакс. связь: csubj 3. Синтакс. связь: acl 4. Синтакс. связь: acl:relcl

5. synt_rel_advcl 6. synt_rel_advmod 7. synt_rel_amod 8. synt_rel_appos 9. synt_rel_aux 10. synt_rel_aux:pass 11. synt_rel_case 12. synt_rel_cc 13. synt_rel_cc:preconj 14. synt_rel_ccomp 15. synt_rel_conj 16. synt_rel_cop 17. synt_rel_det 18. synt_rel_discourse 19. synt_rel_fixed 20. synt_rel_flat 21. synt_rel_goeswith 22. synt_rel_iobj 23. synt_rel_list 24. synt_rel_mark 25. synt_rel_nmod 26. synt_rel_nsubj 27. synt_rel_nsubj:pass 28. synt_rel_nummod 29. synt_rel_nummod:gov 30. synt_rel_obj 31. synt_rel_obl 32. synt_rel_orphan 33. synt_rel_parataxis 34. synt_rel_punct 35. synt_rel_root 36. synt_rel_xcomp 37. synt_rel_compound 38. synt_rel_flat:foreign		5. Синтакс. связь: advcl 6. Синтакс. связь: advmod 7. Синтакс. связь: amod 8. Синтакс. связь: appos 9. Синтакс. связь: aux 10. Синтакс. связь: aux:pass 11. Синтакс. связь: case 12. Синтакс. связь: cc 13. Синтакс. связь: cc:preconj 14. Синтакс. связь: ccomp 15. Синтакс. связь: conj 16. Синтакс. связь: cop 17. Синтакс. связь: det 18. Синтакс. связь: discourse 19. Синтакс. связь: fixed 20. Синтакс. связь: flat 21. Синтакс. связь: goeswith 22. Синтакс. связь: iobj 23. Синтакс. связь: list 24. Синтакс. связь: mark 25. Синтакс. связь: nmod 26. Синтакс. связь: nsubj 27. Синтакс. связь: nsubj:pass 28. Синтакс. связь: nummod 29. Синтакс. связь: nummod:gov 30. Синтакс. связь: obj 31. Синтакс. связь: obl 32. Синтакс. связь: orphan 33. Синтакс. связь: parataxis 34. Синтакс. связь: punct 35. Синтакс. связь: root 36. Синтакс. связь: xcomp 37. Синтакс. связь: compound 38. Синтакс. связь: flat:foreign
---	--	---

## 2.4 Расчет дискурсивных признаков

### Входные данные

Параметр	Тип	Описание
text	string	Непустой текст на русском языке

### Основа

Иерархическая дискурсивная аннотация в рамках Теории риторических структур.



## Используемые анализаторы

Название	Ключи
MyStem	postag, morph
UDPipe	tokens, sentences, lemma, syntax_dep_tree
IsaNLP RST	rst

## Выходные данные

Параметр	Тип	Описание
number_of_units	int	Количество дискурсивных единиц
number_of_trees	int	Количество дискурсивных деревьев
first_tree_edus_number	int	Количество элементарных дискурсивных единиц, входящих в первое дерево
last_tree_edus_number	int	Количество элементарных дискурсивных единиц, входящих в последнее дерево
suspended_edus_number	int	Количество элементарных дискурсивных единиц, не входящих ни в одно дерево
average_du_length	float	Средняя длина дискурсивной единицы в словах
average_edu_length	float	Средняя длина элементарной дискурсивной единицы в словах
average_tree_depth	float	Средняя глубина дискурсивного дерева
symmetric_relations	float	Доля мультаядерных (NN) риторических отношений
first_word_uniqueness	float	Доля уникальных первых слов элементарной ДЕ
du_len_imbalance	float	Среднее отношение длины правой части отношения к длине левой части
simple_sentence_ratio	float	Доля простых предложений среди ЭДЕ
Параметры - наименования риторических отношений: 1. attribution_NS 2. attribution_SN 3. background_NS 4. cause-effect_NS 5. cause-effect_SN 6. comparison_NN	int	Количество вхождений определенных риторических отношений: 1. Источник информации с ядром слева 2. Источник информации с ядром справа 3. Фон 4. Причина с ядром слева 5. Причина с ядром справа 6. Сравнение

7. concession_NS 8. condition_NS 9. condition_SN 10. contrast_NN 11. elaboration_NS 12. evidence_NS 13. interpretation-evaluation_NS 14. interpretation-evaluation_SN 15. joint_NN 16. preparation_SN 17. purpose_NS 18. purpose_SN 19. restatement_NN 20. same-unit_NN 21. sequence_NN 22. solutionhood_SN		7. Уступка 8. Условие с ядром слева 9. Условие с ядром справа 10. Контраст 11. Детализация 12. Обоснование 13. Интерпретация/оценка с ядром слева 14. Интерпретация/оценка с ядром справа 15. Конъюнкция 16. Подготовка 17. Цель с ядром слева 18. Цель с ядром справа 19. Переформулировка 20. Прерывающаяся единица 21. Последовательность 22. Решение
--	--	---

## 2.5 Семантические роли при эмотивах

### Входные данные

Параметр	Тип	Описание
text	string	Анализируемый текст, длина от 10 символов

### Используемые анализаторы

Название	Ключи
IsaNLP RST	rst

### Основа

Модель, обнаруживающая семантические структуры внутри элементарных дискурсивных единиц.

### Выходные данные

Параметр	Тип	Описание
clauses	list[dict]	Список структур, по одной для каждой клаузы в тексте
clauses['text']	string	Текст клаузы
clauses['predictions']	dict	Результат установления ролей внутри клаузы

clauses['predictions']['predicate']	list[dict]	слова-предикаты
clauses['predictions']['experiencer']	list[dict]	Аргументы с ролью экспериенцера
clauses['predictions']['causator']	list[dict]	Аргументы с ролью каузатора
clauses['predictions']['instrument']	list[dict]	Аргументы с ролью инструмента
clauses['predictions']['other']	list[dict]	Аргументы с иными ролями
clauses['predictions']['*name*']['*idx*']	dict	Структура с информацией о выявленном предикате или аргументе
clauses['predictions']['*name*']['*idx*']['idxs']	list[int]	Список индексов слов в клаузе, относящихся к выявленному предикату/аргументу
clauses['predictions']['*name*']['*idx*']['text']	string	Выявленное слово предикат/аргумент
clauses['predictions']['*name*']['*idx*']['score']	float	Показатель уверенности модели

## 2.6 Предсказание депрессивности автора текста по формату эссе

### Входные данные

Параметр	Тип	Описание
text	string	Русскоязычный текст размером от 500 до 5000 символов

### Основа

Психолингвистические и словарные признаки; классификационная модель, обученная на эссе.

### Используемые анализаторы

Название	Ключи
----------	-------

MyStem	lemma, postag
--------	---------------

### Выходные данные

Параметр	Тип	Описание
prediction	int	Предсказание модели (0 или 1)

## 2.7 Предсказание состояния фрустрации у автора текста

### Входные данные

Параметр	Тип	Описание
text	string	Русскоязычный текст размером от 2000 до 20000 символов

### Основа

Модель BERT; классификационная модель, обученная на текстовых сообщениях пользователей социальных сетей.

### Выходные данные

Параметр	Тип	Описание
prediction	Int	Предсказание модели (0 или 1)

## 2.8 Выявление типа реакции на фрустрацию по Розенцвейгу

### Входные данные

Параметр	Тип	Описание
text	string	Непустой текст на русском языке
postag	list[string]	Список частей речи
srl	list[isnlp.annotation.Event]	Результаты семантического анализа в виде списка объектов класса Event ()
morph	list[dict]	Морфологические признаки слов
lemma	list[string]	Нормальные формы слов анализируемого текста

syntax_dep_tree	list[isnlp.annotation.WordSynt]	Синтаксическое дерево в виде списка объектов класса WordSynt.
-----------------	---------------------------------	---

### Используемые анализаторы

Название	Ключи
MyStem	postag, morph
Udpipe	tokens, sentences, lemma, syntax_dep_tree
IsaNLP SRL FrameBank	srl

### Основа

Типология реакций на фрустрацию по-Розенцвейгу.

### Выходные данные

Параметр	Тип	Описание
type	string	Тип реакции на фрустрацию. Строка из множества: {E,E', e, M, M',m, I, I',i} E - агрессия, E'- растерянность и неприятие, e - требование к другим исправить ситуацию, M - отрицание чьей-либо вины, M' - игнорирование, отрицание значимости препятствия m - ожидание, что само всё как-то образуется, I - самообвинение I' - признание ситуации в чём-то полезной, i - "я сам исправлю ситуацию"

### 3. Сторонние компоненты

Компонент	Лицензия и ссылка
MyStem	<a href="https://yandex.ru/legal/mystem/">https://yandex.ru/legal/mystem/</a>  “3.2. Программа может использоваться в коммерческих целях для разработки/создания каких-либо сервисов или программ, включаться и использоваться по прямому функциональному назначению в составе таких сервисов или программ, а также использоваться иным образом в процессе оказания услуг/выполнения работ, за исключением: <ul style="list-style-type: none"><li>- использование Программы и/или сервисов или программ, созданных на её основе или с её использованием, для создания или распространения массовых рассылок и спама;</li><li>- использование Программы и/или сервисов или программ, созданных на её основе или с её использованием, для поисковой оптимизации сайтов в сети Интернет;</li><li>- использование Программы для создания сервисов или программ или в составе сервисов или программ, предлагающих услуги или функциональность, аналогичную программам и сервисам Правообладателя.”</li></ul>
UdPipe	Mozilla Public License Version 2.0 <a href="https://raw.githubusercontent.com/ufal/udpipe/master/LICENSE">https://raw.githubusercontent.com/ufal/udpipe/master/LICENSE</a>
isanlp	MIT License <a href="https://raw.githubusercontent.com/IINemo/isanlp/master/LICENSE">https://raw.githubusercontent.com/IINemo/isanlp/master/LICENSE</a>
isanlp_srl_framebank	MIT License <a href="https://raw.githubusercontent.com/IINemo/isanlp_srl_framebank/master/LICENSE">https://raw.githubusercontent.com/IINemo/isanlp_srl_framebank/master/LICENSE</a>
isanlp_rst	MIT License <a href="https://raw.githubusercontent.com/tchewik/isanlp_rst/master/LICENSE">https://raw.githubusercontent.com/tchewik/isanlp_rst/master/LICENSE</a>
linis-crowd sentiment	<a href="https://creativecommons.org/licenses/by-nc-sa/4.0/">https://creativecommons.org/licenses/by-nc-sa/4.0/</a>  Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

### 4. Минимальные требования

Оборудование	Системные требования
<ul style="list-style-type: none"><li>• 16 ГБ оперативной памяти</li><li>• 8 ядер ЦП</li></ul>	<ul style="list-style-type: none"><li>• python 3.9</li><li>• docker &gt;= 17.06</li><li>• docker-compose &gt;= 1.8.0</li></ul>

## 5. Рекомендуемые требования к оборудованию

- Оперативная память: 17.5 ГБ  
В том числе:
  - o 0.5 ГБ Синтаксический анализатор
  - o 4 ГБ Анализатор семантических ролей
  - o 2 ГБ Семантические роли при эмотивных предикатах
  - o 7 ГБ Дискурсивный анализатор
  - o 4 ГБ Оценка фрустрированности автора
- В расширенной версии:
  - o 0.3 ГБ Оценка депрессивности автора
  - o 4 ГБ Тип реакции на фрустрацию
- Количество ядер для стабильной работы каждого модуля: 16.

## 6. Ссылка на открытый репозиторий

<https://github.com/tchewik/titanis-open>