

Е.О. Филиппова, доктор технических наук

Томский политехнический университет, Томск, Российская Федерация

Е.С. Шипицына, кандидат педагогических наук

Томский политехнический университет, Томск, Российская Федерация

А.О. Аспомбитова

Томский политехнический университет, Томск, Российская Федерация

E.O. Filippova, Doctor of Technical Sciences

Tomsk Polytechnic University, Tomsk, Russian Federation

E.S. Shipitsyna, Candidate of Pedagogical Sciences

Tomsk Polytechnic University, Tomsk, Russian Federation

A.O. Aspombitowa

Tomsk Polytechnic University, Tomsk, Russian Federation

АНАЛИЗ ДАННЫХ СВОБОДНОГО АССОЦИАТИВНОГО ЭКСПЕРИМЕНТА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

ANALYSIS OF DATA FROM A FREE ASSOCIATIVE EXPERIMENT USING MACHINE LEARNING METHODS

Рассматриваются вопросы анализа данных свободного ассоциативного эксперимента, проведенного среди студентов технического университета с использованием методов машинного обучения. Проводилась лемматизация с помощью библиотеки pymystem3-Python-обертка для Yandex.Mystem. Вычисление семантической близости реализовано с помощью модели BERT и токенизатора AutoTokenizer из библиотеки transformers. Группировка слов по эмоциональному окрасу реакций выполнена с помощью библиотеки transformers (модель nlptown/bert-base-multilingual-uncased-sentiment).

Ключевые слова: ассоциативный эксперимент; машинное обучение; лемматизация; анализ данных.

The article considers the issues of data analysis of a free associative experiment conducted among students of a technical university using machine learning methods. Lemmatization was carried out using the pymystem3 library – a Python wrapper for Yandex.Mystem. Semantic proximity was calculated using the BERT model and the AutoTokenizer tokenizer from the transformers library. Words were grouped by emotional coloring of reactions using the transformers library (the nlptown/bert-base-multilingual-uncased-sentiment model).

Key words: associative experiment; machine learning; lemmatization; data analysis.

Работа посвящена анализу данных свободного ассоциативного эксперимента, проведенного среди студентов технического университета с использованием методов машинного обучения.

Цель – проанализировать данные свободного ассоциативного эксперимента с использованием методов машинного обучения.

В исследовании приняли участие 170 студентов 1 курса Томского политехнического университета. Каждому участнику эксперимента была дана анкета, содержащая слова-стимулы на свободное ассоциирование: *математика – это; высшая математика для меня; математический анализ; решать задачи я; на занятии по высшей математике; контрольные по математике я; лекции по математике вызывают у меня; функция – это; предел последовательности; предел функции; односторонние пределы; разрывы функции; бесконечно большие величины; бесконечно малые величины; замечательные пределы*. Всего было получено 1902 реакций.

Для анализа полученных результатов проводилась лемматизация с помощью библиотеки `rumystem3-Python-обертка для Yandex.Mystem`. Для объединения эквивалентных слов и нахождения синонимов был использован сервис «Яндекс.Словарь» <https://tech.yandex.ru/dictionary>. Слова группировали по эмоциональному окрасу реакций с помощью библиотеки `transformers` (модель `nlptown/bert-base-multilingual-uncased-sentiment`), разделив их на пять категорий: очень негативный, негативный, нейтральный, позитивный и очень позитивный. Вычисление семантической близости реализовано с помощью модели BERT и токенизатора `AutoTokenizer` из библиотеки `transformers`.

Согласно анализу реакций и исходя из эмоционального тона, в большей степени негативный и очень негативный окрас вызывали стимулы: *решать задачи я; разрывы функции; бесконечно малые величины*. Позитивный и очень позитивный окрас вызывали стимулы: *математика – это; высшая математика для меня; математический анализ; на занятии по высшей математике; контрольные по математике я; лекции по математике вызывают у меня; замечательные пределы*. Количество ответов с нейтральным окрасом было в меньшинстве.

Наиболее семантически близкие реакции с коэффициентом семантической близости выше 0,6 были на стимулы: *высшая математика для меня; математический анализ; решать задачи я*. Наименьший коэффициент (0,38) семантической близости реакций был на стимул *предел функции*, что свидетельствует о проявлении связей на уровне несущественных компонентов лексического значения сравниваемых слов и слабо усвоенном понимании определения предела функций.