

УДК 004.912

ЭВОЛЮЦИЯ КОНЦЕПЦИИ СТРЕССА: АНАЛИЗ НАУЧНЫХ ПУБЛИКАЦИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Бабилов И.А (ИТМО)

Научный руководитель – кандидат технических наук, доцент Ковальчук С.В. (ИТМО)

Научный консультант – доктор биологических наук, профессор Кубряк О.В. (МЭИ)

Введение. Работа посвящена исследованию эволюции научной терминологии на примере понятия «стресс». Этот термин, изначально введенный Гансом Селье как физиологическое явление (модель общего адаптационного синдрома), претерпел значительные изменения и стал использоваться в более широком социально-психологическом контексте. Для анализа эволюции термина применяются современные методы тематического моделирования, позволяющие выявить ключевые этапы развития концепции стресса на основе анализа корпуса научных публикаций. Цель исследования — выделить основные фазы стресса по классификации Селье и проследить динамику изменения тематик, связанных с этим понятием, в научной литературе.

Основная часть. Для анализа эволюции понятия «стресс» были использованы следующие методы тематического моделирования:

- 1) LDA (Gensim) [1];
- 2) NMF (Gensim) [2];
- 3) ARTM (BigARTM) [3];
- 4) Top2Vec [4];
- 5) BERTopic [5].

Каждая модель была протестирована на выявление 3, 4, 7 и 10 тем, что позволило оценить их способность к интерпретации фаз стресса по классификации Селье [6].

Корпус данных был составлен из англоязычных статей базы PubMed, опубликованных с 1975 по 2024 год. Критерием отбора статей стало наличие ключевого слова «stress» в названии или аннотации. Общий объем корпуса составил 1 045 358 статей со средней длиной документа (abstract) 127 слов.

Наиболее интерпретируемые результаты были получены с использованием модели BERTopic (10 тем). Данная модель позволила лучше выделять фазы стресса по Селье: тревогу, сопротивление и истощение. Также достойные результаты показала классическая модель NMF с использованием эмбедингов Word2Vec [7], которая успешно идентифицировала отдельные фазы стресса тревоги и сопротивления.

Модели LDA, ARTM и Top2Vec продемонстрировали менее четкие результаты, что свидетельствует об их ограниченной применимости для анализа фаз стресса. Кроме того, было установлено, что небольшое количество тем (3 или 4) не позволяет адекватно отразить все этапы стресса по классификации Селье.

Дополнительно была проанализирована динамика полученных тем с 1975 по 2024 год. После нормализации данных по количеству статей в каждом году выявлен следующий тренд: ранее популярные направления исследований стресса утратили свою актуальность, тогда как новые тематики, связанные с социально-психологическими аспектами стресса, стали активно развиваться в последние десятилетия.

Выводы. В исследовании был проведен статический и динамический анализ корпуса англоязычных научных публикаций по теме стресса с помощью классических и современных методов тематического моделирования. Наилучшие результаты в интерпретации фаз стресса по модели Селье показала модель BERTopic. Полученные результаты могут быть применены для информационно-поисковой поддержки в научных работах по стрессологии, а также

позволяют автоматически разделить смежные направления исследования. Предложенные методы применимы не только к исследованиям стресса, но и к другим дисциплинам и текстам на различных языках.

В дальнейшем планируется расширить исследование, включив анализ эволюции содержания понятия «стресс» в контексте современных научных дискуссий.

Список использованных источников:

1. Blei D. M., Ng A. Y., Jordan M. I. “Latent Dirichlet Allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.
2. Lee, D., Seung, H. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791, 1999.
3. Vorontsov K. “Additive Regularization for Topic Models of Text Collections” Dokl. Math., vol. 89, pp. 301–304, 2013, doi: 10.1134/S1064562414020185.
4. Angelov D. “Top2Vec: Distributed Representations of Topics.” arXiv, 2020. doi: 10.48550/ARXIV.2008.09470.
5. Grootendorst M. “BERTopic: Neural topic modeling with a class-based TF-IDF procedure.” arXiv, 2022. doi: 10.48550/ARXIV.2203.05794.
6. Selye H. Stress and the general adaptation syndrome. Br Med J. 1950 Jun 17;1(4667):1383-92. doi: 10.1136/bmj.1.4667.1383. PMID: 15426759; PMCID: PMC2038162.
7. Mikolov T., Chen K., Corrado G., Dean J. “Efficient Estimation of Word Representations in Vector Space” arXiv, 2013, arXiv:1301.3781.