

Научная статья
Статья в открытом доступе
УДК 331.101.1: 004.9
doi: 10.30987/2658-4026-2025-1-93-99

Анализ культурных контекстов образовательных данных для рекомендации изучения иностранных языков

Игорь Юрьевич Коцюба^{1✉}, София Емельянова²

^{1,2} Национальный исследовательский университет ИТМО, Санкт-Петербург, Россия

¹ ikotciuba@itmo.ru

² oktyasof@gmail.com

Аннотация.

В статье рассмотрен подход к формированию рекомендаций по изучению иностранных языков с помощью мобильных приложений на основе анализа культурных контекстов образовательных данных пользователей. Проведен анализ специфики изучения иностранных языков в различных существующих образовательных сервисах. Особое внимание уделено анализу культурных данных на примере профессиональных культур для определения специфики дизайна профессионально направленных программных обучающих сервисов. По результатам проведенного в работе анализа были сформированы рекомендации для изучения иностранных языков для различных категорий специалистов, а также сформулированы требования к дизайну таких обучающих сервисов.

Перспективой дальнейших исследований является необходимость учёта языковой картины мира, которая согласно исследованиям А.А. Леонтьева, имеет трехуровневый характер и включает, в том числе, психологический и эргономический планы. Эргономическая и психологическая составляющие позволят обеспечить требуемый уровень знания иностранного языка будущим специалистом с учётом профессиональных и культурных контекстов образовательных данных.

Ключевые слова: культурный контекст, образовательные данные, психолингвистика, мобильные приложения, уровень знаний обучающая программа

Для цитирования: Коцюба И.Ю., Емельянова С. Анализ культурных контекстов образовательных данных для рекомендации изучения иностранных языков // Эргодизайн. 2025. №1 (27). С. 93-99. <http://dx.doi.org/10.30987/2658-4026-2025-1-93-99>.

Original article
Open access article

Analysing Cultural Contexts in Educational Data for Foreign Language Learning Recommendations

Igor Yu. Kotsyuba^{1✉}, Sofia Emelyanova²

^{1,2} National Research University ITMO, Saint Petersburg, Russia

¹ ikotciuba@itmo.ru

² oktyasof@gmail.com

Abstract.

The article explores an approach to generating foreign language learning recommendations via mobile applications, based on analysing cultural contexts within users' educational data. The work presents a comparative analysis of the specific features of foreign language learning in various existing educational services. The work pays particular attention to analysing cultural data, using professional cultures as an example, to determine the design specifications of professionally oriented software training services. Based on the analysis results, the authors formulate recommendations for foreign language learning for different categories of specialists, along with formulating design requirements for such training services.

The prospect for further research lies in the necessity of considering the linguistic world picture, which, according to the research of A.A. Leontiev, has a three-level structure, encompassing, among other aspects, psychological and ergonomic planes. The ergonomic and psychological components will enable achieving the required level of foreign language proficiency for future specialists, taking into account both professional and cultural contexts of educational data.

Keywords: cultural context, educational data, psycholinguistics, mobile applications, proficiency level, training program

For citation: Kotsyuba I.Yu., Emelyanova S. Analysing Cultural Contexts in Educational Data for Foreign Language Learning Recommendations. Ergodizayn [Ergodesign]. 2025;1(27):93-99. Doi: 10.30987/2658-4026-2025-1-93-99.

Введение.

Современное образование нацелено на максимальную индивидуализацию, стремление сделать процесс обучения комфортным и учитывающим различные факторы, влияющие на успешность процесса обучения [1], [2], [3], [4]. Стадия дизайна (проектирования) информационных систем должна охватывать учет большого количества требований, особенно в сфере информатизации образования, в том числе требования по контенту, структуре обучающего приложения. Анализ данных позволяет выявить нетривиальные зависимости особенностей аудитории, контента, локализации и других факторов, которые позволят сделать дизайн информационной системы менее трудоемким и более эффективным.

В нашем мире присутствует бесчисленное количество различных культур и народов. Многообразие наций порождает большой спектр языков, на которых разговаривают люди. Как следствие процесса всемирной интеграции, когда аспекты различных сфер жизни отдельных стран приобретают глобальный масштаб, возникает все больше взаимодействий среди людей. Многие начинают изучать иностранные языки [5]. Вследствие этого расширяются границы их мышления, так как в обиход приходят понятия, которых нет в реальности человека и которые, соответственно, трудно перевести. Происходит оказание положительного влияния на умственные способности человека. Также существует множество профессий, для которых знание иностранного языка – это необходимость. Ведь множество материалов могут быть не переведены на родной язык человека.

С появлением информационных технологий одним из популярнейших способов для изучения иностранных языков стали мобильные приложения. В настоящий момент данные программные средства предоставляют большой выбор упражнений для обучающихся. Множество людей не имеют достаточного количества времени и материальных ресурсов для того, чтобы посещать полноценные языковые курсы. Так что мобильные приложения являются единственной возможностью обучения для данных потребителей [6]. Также программные средства, позволяющие изучать иностранные

языки, могут быть дополнением к основному учебному процессу [7], ведь они позволяют больше прорабатывать различные аспекты языка. Мобильные приложения дают возможность сделать обучение более увлекательным и гибким, а пользователи могут развивать свои навыки в любое удобное для них время.

Работа с данными

Для сбора данных с Amazon были использованы два инструмента: Amazon Reviews Scraper [8] (извлечение данных об отзывах на специализированную литературу) и Amazon Product Scraper [9] (извлечение описания специализированной литературы). При парсинге отзывов на приложения из Google Play был использован пакет google_play_scraper [10].

Далее были осуществлены следующие шаги по предобработке и очистке данных:

- выбор необходимых столбцов или удаление неиспользуемых столбцов,
- перевод отзывов на английский язык,
- приведение к нижнему регистру,
- удаление пунктуации,
- токенизация,
- фильтрация по частоте/длине/соответствию регулярному выражению,
- лемматизация,
- векторизация,
- удаление стоп-слов [11].

Анализ данных

1. Анализ тональности

Анализ тональности отзывов является одним из ключевых способов оценки мнения об определенном продукте или услуге [12]. Чаще всего присутствует три категории: положительные, нейтральные и негативные тексты. Результаты анализа тональности позволяют обращать внимание на сильные и слабые аспекты товара или услуги, а также подготовить рекомендации для пользователей [13].

На основе языка Python была сформирована функция [14], позволяющая оценить эмоциональную окраску отзывов на приложения и на специализированную литературу. Выделяется три категории тональности: положительная, нейтральная, негативная. Результаты анализа были добавлены в новый столбец result. Таким образом, была получена таблица (рис.1).

	reviewTitle	reviewDescription	ratingScore	date	country	result
0	Gut	The content itself is good. For those who, lik...	5	2014-11-01	Germany	Positive
1	Un libro, sin mas	A book, without more	5	2019-02-09	Spain	Positive
2	Spezialwörterbuch	This is an expensive special dictionary. Bough...	4	2011-07-24	Germany	Positive
3	gute Inhalte aber für den Preis zu teuer	I wouldn't buy the book because it's a waste o...	1	2017-05-16	Germany	Negative
0	A really good ESL book	A really good ESL book, and the only one I cou...	5	2014-12-17	United Kingdom	Positive
...
9	Excelente	Good book recommended	5	2020-12-13	Spain	Positive
0	Five Stars	Everything OK! Times respected and the book is...	5	2014-12-08	Italy	Positive
1	Purchased for a relative in Poland	The student we sent it to greatly appreciated it.	5	2023-10-18	United States	Positive
2	Five Stars	Excellent price and delivery.	5	2015-04-24	United Kingdom	Positive
3	In need of an update!!!	This book is passable. However it is in seriou...	3	2017-02-07	United States	Negative

Рис. 1. Пример датасета со столбцом тональности
Fig. 1. An example of a dataset with a key column

Было принято решение отделения положительных отзывов на литературу для представителей различных специальностей, с целью выявления тем, которые нравятся специалистам в подходах, предоставляемых авторами литературы. При анализе отзывов на приложения по изучению иностранных языков также были использованы выделенные как положительные отзывы пользователей. Данный подход необходим для более точного соотнесения аспектов специализированной литературы и ПО.

2. Тематическое моделирование

Тематическое моделирование является одним из важнейших методов при анализе отзывов. Данное средство используется для выявления основных тем и аспектов, которые встречаются в тексте [15]. Тематическое моделирование позволяет находить скрытые паттерны и тенденции, на основе которых

можно делать выводы о предпочтениях и потребностях пользователей.

В ходе данной работы тематическое моделирование было применено относительно положительно оцененных отзывов на литературу по изучению иностранного языка для представителей разных сфера, а также отзывов на приложения по изучению иностранных языков.

Создается словарь с полученными во время векторизации триграммами. Затем определяется частота документа в корпусе, и строится тематическая модель с указанием количества тем. Далее темы визуализируются.

Сначала было проведено тематическое моделирование для положительных отзывов на специализированную литературу. Это позволило выявить по три темы для каждого из пяти культурных контекстов. На графике представлена LDA модель для медицинских специалистов (рис.2).

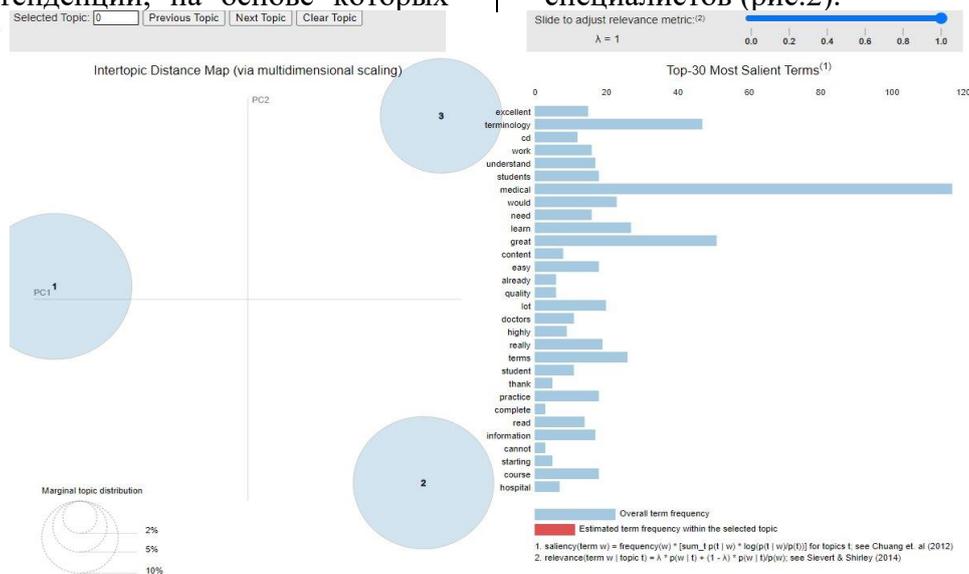


Рис. 2. Тематическое моделирование на основе позитивных отзывов на литературу для изучения иностранных языков для медицинских специалистов
Fig. 2. Thematic modeling based on positive reviews of literature for learning foreign languages for medical professionals

По выделенным словам, был составлен список тем, встречающихся в отзывах на специализированную литературу по изучению английского языка для медицинских специалистов:

- 1) специальная терминология,
- 2) базовая, полезная информация,
- 3) практика словарного запаса.

Аналогично были выявлены основные мотивы отзывов на литературу для других специалистов. Далее для поддержания

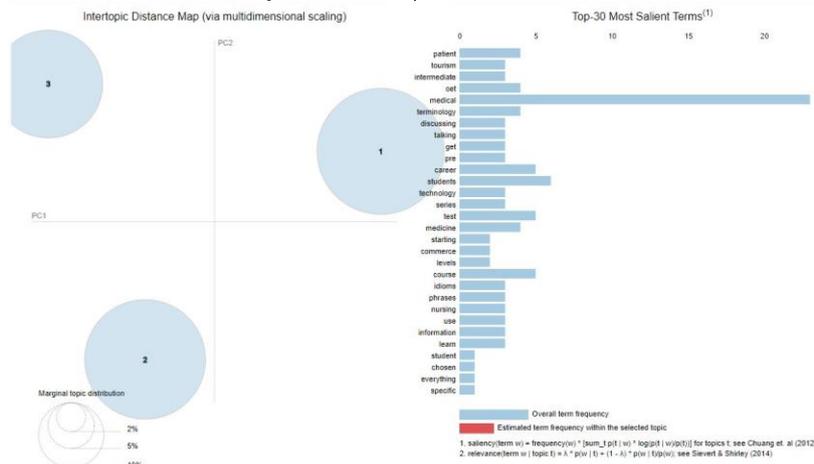


Рис. 3. Тематическое моделирование на описания литературы по изучению иностранных языков для медицинских специалистов

Fig. 3. Thematic modeling based on literature descriptions on learning foreign languages for medical professionals

Был выделен следующий список тем:

- 1) изучение идиом и фраз,
- 2) изучение определенных тем и слов,
- 3) разговорный язык для общения с пациентами.

Для каждой из выбранных профессий было дважды составлено тематическое моделирование, это позволило выявить больше аспектов, отличающих представителей разных специальностей, при изучении иностранных языков. Для IT специалист ключевыми являются факторы:

- умение коммуницировать посредством письменной и разговорной речи,
- присутствие грамотности в речи,
- хорошее владение грамматикой и наличие большого словарный запаса.

Для медицинского специалиста важными являются аспекты:

- разговорный язык для общения с пациентами,
- знание специальной терминологии и идиом,
- развитие словарного запаса.

По выделенным словам был составлен список тем, встречающихся в отзывах на HiNative:

- 1) постановка вопроса, получение ответа,

сделанных выводов относительно ключевых аспектов для различных профессий было также создано тематическое моделирование на основе описания специализированной литературы. Для каждого культурного контекста была построена LDA модель. На рис.3 представлен график, показывающий результаты анализа описания учебников по изучению английского языка для медицинских специалистов.

- 2) носители языка, сообщество и люди в нем.

Для маркетолога ключевыми являются факторы:

- хорошее владение грамматикой,
- письменная и разговорная коммуникация,
- знание различных слов и выражений.

Для бизнесмена важными являются аспекты:

- знание терминологии, идиом и грамматики,
- письменная коммуникация,
- следование тенденциям и совершенствование навыков.

Для юриста ключевыми являются факторы:

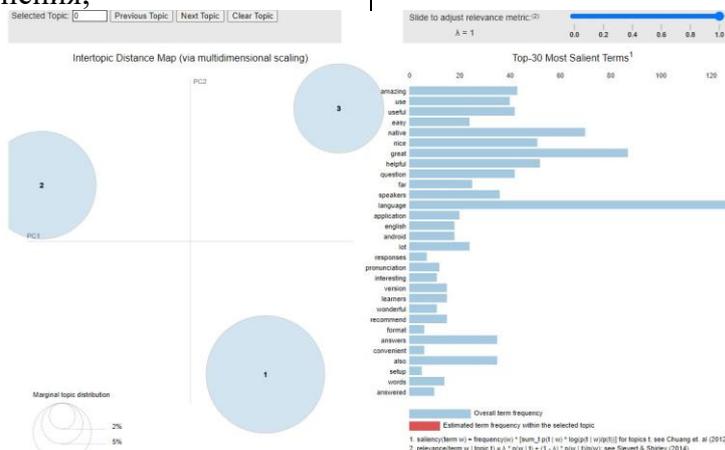
- навыки чтения,
- развитие словарного запаса,
- письменная коммуникация.

Далее было составлено тематическое моделирование для положительных отзывов на приложения по изучению иностранных языков. Было выделено по три темы для каждого ПО. На рис. 4. представлена LDA модель для приложения HiNative.

Аналогично были выявлены основные мотивы отзывов на другие приложения по изучению иностранных языков:

- Duolingo:

- 1) продолжение обучения;
 - 2) новый элемент в уроках;
 - 3) простой в использовании язык;
- Babbel:
 - 1) разговор с использованием голоса;
 - 2) программа обучения грамматике;
 - 3) понятные объяснения;



- Busuu:
- 1) использование слов;
- 2) изучение грамматики и прохождение тестов;
- 3) практика языка;

Рис. 4. Тематическое моделирование на положительный отзыв на приложение HiNative
Fig. 4. Thematic modeling based on a positive review of the HiNative application

- Rosetta Stone:
- 1) перевод слов;
- 2) новые методы обучения;
- 3) разговорная практика;
- Mondly:
- 1) простой в использовании язык;
- 2) система обучения;
- 3) дружелюбное и понятное обучение;
- HelloTalk:
- 1) разговор голосом;
- 2) нахождение друзей;
- 3) разговорная практика;
- Drops:
- 1) словарный запас;
- 2) интересное времяпрепровождение;
- 3) использование слов в жизни;
- Pimsleur:
- 1) простой в разговоре язык;
- 2) аудирование;
- 3) практика произношения;
- Lirica:
- 1) музыка и песни;
- 2) текстовая составляющая песни;
- 3) слова и грамматика.

Заключение

По результатам проведенного в работе анализа были сформированы рекомендации для изучения иностранных языков для специалистов.

Для IT специалиста подойдут следующие ПО:

- Babbel, Rosetta Stone можно использовать для практики разговорной речи,

- с помощью HiNative появится возможность оценить грамотность благодаря общению с носителями языка,
- при использовании Babbel, Busuu, Lirica можно изучить грамматические конструкции,
- Drops позволит пополнить словарный запас.

Для медицинского специалиста будут полезны приложения:

- Duolingo, Pimsleur, Mondly позволят овладеть простым языком, который необходим при общении с пациентами,
- с помощью Babbel, HelloTalk появится возможность практики разговорного языка,
- благодаря HiNative, HelloTalk можно изучить специальную терминологию и идиомы,
- Busuu, Drops позволят пополнить словарный запас,
- также дополнительно можно обратить внимание на упражнения по чтению, представленные в Pimsleur, и упражнения по грамматике в приложениях Pimsleur, Busuu, Babbel, Drops.

Для маркетолога подойдут следующие ПО:

- при использовании Babbel, Busuu можно изучить грамматические конструкции,
- Babbel, HelloTalk, Pimsleur, Rosetta Stone можно использовать для практики разговорной речи,
- с помощью Busuu, Drops появится возможность пополнить словарный запас,
- также дополнительно можно обратить внимание на упражнения по чтению,

представленные в Pimsleur, и упражнения по грамматике в приложениях Pimsleur, Drops.

Для бизнесмена будут полезны приложения:

- благодаря HiNative можно изучить специальную терминологию и идиомы,
- при использовании Babbel, Busuu можно изучить грамматические конструкции,
- Lirica позволит обратить внимания на тенденции, так как в данном ПО присутствуют музыка и тексты,
- также дополнительно можно обратить внимание на упражнения по чтению, представленные в Pimsleur, упражнения по грамматике в приложениях Pimsleur, Drops,

СПИСОК ИСТОЧНИКОВ

1. **Lalaeng Ch., Subongkod M., Sinlapasawet M.** Individual competency causal factors affecting performance of academic personnel in higher education institution. *Procedia Computer Science*. 2024;237:502-509. DOI 10.1016/j.procs.2024.05.133. EDN JYJMAR.
2. **Jaruseviciute V., Silinskas G., Muotka J., Kiuru N.** Trajectories of adolescents' adjustment behaviors across the transition to upper secondary education: The role of individual and environmental factors. *Learning and Individual Differences*. 2024;112:102457. DOI 10.1016/j.lindif.2024.102457.
3. **Sormani E., van der Sijde P.** Configuring motivational pathways to manage academics' individual motivation for engagement: Insights from German higher education institutions. *Journal of Engineering and Technology Management*. 2023;69:101768. DOI 10.1016/j.jengtecman.2023.101768.
4. **Сафронова И.В.** Индивидуализация образовательного процесса как условие повышения качества образовательных услуг // *Фундаментальные и прикладные исследования: проблемы и результаты*. 2012. № 1. С. 102-107. EDN RDMPRZ.
5. **Никонова Е.Н.** Роль иностранного языка в эпоху глобализаций // *Grand Altai Research & Education*. 2017. № 2. С. 38-42. EDN YNNVRK.
6. **Бобылева Г.А., Бормосова Н.Е., Афанасьева Е.Г. и др.** Применение мобильных приложений в процессе обучения иностранному языку // *Ученые записки университета им. П.Ф. Лесгафта*. 2021. № 10(200). С. 33-37. DOI 10.34835/issn.2308-1961.2021.10.p33-37. EDN GITXVN.
7. **Канцур А.Г., Бердникова Н.С.** Использование мобильных приложений на уроках иностранного языка // *Проблемы романо-германской филологии, педагогики и методики преподавания иностранных языков*. 2019. № 15. С. 75-80. EDN YGETKI.
8. **Amazon Reviews Scraper. Apify.** URL: <https://console.apify.com/actors/R8WeJwLuzLZ6g4Bkk/inp-ut> (дата обращения: 30.11.2024).
9. **Amazon Product Scraper. HasData.** URL: <https://app.hasdata.com/marketplace/23> (дата обращения: 30.11.2024).
10. **Google-Play-Scraper** // PyPI URL: <https://pypi.org/project/google-play-scraper/> (дата обращения: 30.11.2024).
11. **Обработка естественного языка. Ifmo.** URL: https://neerc.ifmo.ru/wiki/index.php?title=Обработка_естественного_языка (дата обращения: 30.11.2024).

упражнения по аудированию, находящиеся в Pimsleur, Lirica, и упражнения, направленные на развитие навыков письма в Pimsleur.

Для юриста подойдут следующие ПО:

- с помощью Lirica, HiNative появится возможность развития навыков чтения,
- Busuu, Drops позволят пополнить словарный запас,
- также дополнительно можно обратить внимание на упражнения по чтению, представленные в Pimsleur, и упражнения по грамматике в приложениях Pimsleur, Busuu, Babbel, Drops, и упражнения, направленные на развитие навыков письма в Pimsleur.

REFERENCES

1. **Lalaeng Ch., Subongkod M., Sinlapasawet M.** Individual Competency Causal Factors Affecting Performance of Academic Personnel in Higher Education Institution. *Procedia Computer Science*. 2024;237:502-509. DOI 10.1016/j.procs.2024.05.133.
2. **Jaruseviciute V., Silinskas G., Muotka J., Kiuru N.** Trajectories of Adolescents' Adjustment Behaviours Across the Transition to Upper Secondary Education: The Role of Individual and Environmental Factors. *Learning and Individual Differences*. 2024;112:102457. DOI 10.1016/j.lindif.2024.102457.
3. **Sormani E., van der Sijde P.** Configuring Motivational Pathways to Manage Academics' Individual Motivation for Engagement: Insights From German Higher Education Institutions. *Journal of Engineering and Technology Management*. 2023;69:101768. DOI 10.1016/j.jengtecman.2023.101768.
4. **Safronova I.V.** Individualization of the Educational Process as a Condition for Improving the Quality of Educational Services. *Fundamental and Applied Research: Problems and Results*. 2012;1:102-107.
5. **Nikonova E.N.** The Role of a Foreign Language in the Era of Globalization. *Grand Altai Research & Education*. 2017;2:38-42.
6. **Bobyleva G.A., Bormosaeva N.E., Afanasyeva E.G., et al.** Use of Mobile Applications in the Process of Teaching a Foreign Language. *Uchenye Zapiski Universiteta Imeni P.F. Lesgafta*. 2021;10(200):33-37. DOI 10.34835/issn.2308-1961.2021.10.p33-37.
7. **Kantsur A.G., Berdnikova N.S.** Using Mobile Applications at Foreign Language Lessons. *Problems of Romano-Germanic Philology, Pedagogy and Methods of Teaching Foreign Languages*. 2019;15:75-80.
8. **Amazon Reviews Scraper. Apify** [Internet] [cited 2024 Nov 30]. Available from: <https://console.apify.com/actors/R8WeJwLuzLZ6g4Bkk/inp-ut>.
9. **Amazon Product Scraper. HasData** [Internet] [cited 2024 Nov 30]. Available from: <https://app.hasdata.com/marketplace/23>.
10. **Google-Play-Scraper** // PyPI [Internet] [cited 2024 Nov 30]. Available from: <https://pypi.org/project/google-play-scraper/>.
11. **Natural Language Processing. Ifmo.** [Internet] [cited 2024 Nov 30]. Available from: https://neerc.ifmo.ru/wiki/index.php?title=Обработка_естественного_языка.

12. **Акжолов Р.К., Верига А.В.** Анализ тональности текста // Вестник науки. 2019. Т. 4, № 9(18). С. 29-31. EDN CDQZRU.

13. **Трофимова Е.В., Туральчук К.А.** Разработка рекомендательной системы на основе анализа тональности текста // Актуальные проблемы гуманитарных и естественных наук. 2015. №1-1. С. 93-94. EDN TILCLD.

14. **TextBlob: Simplified Text Processing.** URL: <https://textblob.readthedocs.io/en/dev/> (дата обращения: 30.11.2024).

15. **Коршунов А., Гомзин А.** Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. 2012. №23. EDN PLUXDR.

12. **Akzholov R.K., Veriga A.V.** Sentiment Analysis of Text. Science Bulletin. 2019;4-9(18):29-31.

13. **Trofimova E.V., Turalchuk K.A.** Development of a Recommender System Based on Sentiment Analysis of Text. Actual Problems of Humanitarian and Natural Sciences. 2015;1-1:93-94.

14. **TextBlob: Simplified Text Processing.** [Internet] [cited 2024 Nov 30]. Available from: <https://textblob.readthedocs.io/en/dev/>

15. **Korshunov A., Gomzin A.** Topic Modelling for Natural Language Texts. Proceedings of the Institute for System Programming of the Russian Academy of Sciences. 2012;23.

Информация об авторах:

Коцюба Игорь Юрьевич (Санкт-Петербург, Россия)

– кандидат технических наук, доцент, факультет инфокоммуникационных технологий, Университет ИТМО (Россия, 197101, г. Санкт-Петербург, Кронверкский проспект, 49, литер А, e-mail: ikotciuba@itmo.ru) SPIN-код: 5296-3099, AuthorID: 757740.

Емельянова София (Санкт-Петербург, Россия) – студент, факультет инфокоммуникационных технологий, Университет ИТМО (Россия, 197101, г. Санкт-Петербург, Кронверкский проспект, 49, литер А, e-mail: oktyasof@gmail.com).

Information about the authors:

Kotsyuba Igor Yuryevich (Saint Petersburg, Russia) –

Candidate of Technical Sciences, Associate Professor at the Faculty of Infocommunication Technologies of ITMO University (Litera A, 49 Kronverkskiy Prospekt, Saint Petersburg 197101 Russia, e-mail: ikotciuba@itmo.ru), the author's international identification numbers: SPIN-code: 5296-3099, AuthorID: 757740.

Emelyanova Sofia (Saint Petersburg, Russia) – student at the Faculty of Infocommunication Technologies of ITMO University (Litera A, 49 Kronverkskiy Prospekt, Saint Petersburg 197101 Russia, e-mail: oktyasof@gmail.com).

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Contribution of the authors: the authors contributed equally to this article.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interests.

Статья поступила в редакцию 22.11.2024; одобрена после рецензирования 29.11.2024; принята к публикации 03.12.2024. Рецензент – Кухта М.С., доктор философских наук., профессор Томского политехнического университета, член редакционного совета журнала «Эргодизайн»

The paper was submitted for publication on the 22nd of November 2024; approved after the peer review on the 29th of November, 2024; accepted for publication on the 03rd of December, 2024. Reviewer – Kukhta M.S., Doctor of Philosophical Sciences, Professor of National Research Tomsk Polytechnic University, member of the editorial board of the journal “Ergodesign”.