

## МЕТОД ОЦЕНКИ ЭФФЕКТИВНОСТИ МЕРОПРИЯТИЙ ПО ФОРМИРОВАНИЮ КОМФОРТНОЙ ГОРОДСКОЙ СРЕДЫ

**А.В. Мельников**<sup>1</sup>, [MelnikovAV@uriit.ru](mailto:MelnikovAV@uriit.ru), <https://orcid.org/0000-0002-1073-7108>  
**К.В. Галаган**<sup>2</sup>, [galagankv@gmail.com](mailto:galagankv@gmail.com), <https://orcid.org/0000-0003-3587-8334>

<sup>1</sup> Югорский научно-исследовательский институт информационных технологий,  
Ханты-Мансийск, Россия

<sup>2</sup> Югорский государственный университет, Ханты-Мансийск, Россия

**Аннотация.** В статье рассматривается актуальная проблема несоответствия реализуемых мероприятий по формированию комфортной городской среды реальным потребностям и ожиданиям граждан. Анализируются существующие методы оценки таких мероприятий, включая индексный метод оценки качества городской среды, применяемый в рамках национального проекта «Жилье и городская среда», а также различные социологические исследования и опросы. Отмечаются недостатки этих подходов, такие как ограниченность выборки респондентов и тематики опросов. **Цель исследования** – разработка метода оценки мероприятий по формированию комфортной городской среды на основе семантического сравнения мнений граждан из социальных сетей, поисковых запросов и описаний самих мероприятий с использованием алгоритмов обработки естественного языка. **Материалы и методы.** Для реализации поставленной цели используется комплексный подход, состоящий из четырех основных этапов. Этап 1. Предобработка исходных текстовых данных – удаление шумов, приведение слов к начальной форме (лемматизация) с помощью библиотеки `ru morphology2`, определение частей речи (POS-tagging). Этап 2. Извлечение ключевых словосочетаний (N-грамм) при помощи алгоритма TF-IDF с учетом частоты употребления в пределах отдельных сообщений и во всем массиве текстов. Расчет ранга значимости N-грамм. Этап 3. Получение векторного представления (word embeddings) для каждой ключевой N-граммы с использованием предобученной нейросетевой модели SBERT. Этап 4. Вычисление меры семантического сходства векторных представлений N-грамм из разных текстовых массивов (мнений граждан и описаний мероприятий) на основе косинусного расстояния. В качестве исходных данных используются тексты мнений граждан и описания мероприятий, направленных на формирование комфортной городской среды, относящиеся к Ханты-Мансийскому автономному округу. **Результаты.** Проведенный эксперимент показал, что большинство проанализированных мероприятий, направленных на формирование комфортной городской среды, слабо коррелируют с реальными потребностями граждан. **Заключение.** Предложенный метод может использоваться в системах поддержки принятия решений для оценки и выбора наиболее эффективных мероприятий.

**Ключевые слова:** семантический анализ, TF-IDF, трансформеры, sentence transformer, комфортная городская среда, NLP

**Для цитирования:** Мельников А.В., Галаган К.В. Метод оценки эффективности мероприятий по формированию комфортной городской среды // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2024. Т. 24, № 1. С. 87–95. DOI: 10.14529/ctcr240108

Original article  
DOI: 10.14529/ctcr240108

## METHOD OF ASSESSING THE EFFECTIVENESS OF MEASURES TO FORM A COMFORTABLE URBAN ENVIRONMENT

A.V. Melnikov<sup>1</sup>, MelnikovAV@uriit.ru, <https://orcid.org/0000-0002-1073-7108>  
K.V. Galagan<sup>2</sup>, galagankv@gmail.com, <https://orcid.org/0000-0003-3587-8334>

<sup>1</sup> Ugra Research Institute of Information Technologies, Khanty-Mansiysk, Russia

<sup>2</sup> Yugra State University, Khanty-Mansiysk, Russia

**Abstract.** The article deals with the urgent problem of inconsistency of the implemented measures to form a comfortable urban environment with the real needs and expectations of citizens. It analyzes the existing methods of assessing such measures, including the index method of assessing the quality of urban environment, used within the framework of the national project “Housing and Urban Environment”, as well as various sociological studies and surveys. The shortcomings of these approaches, such as the limited sample of respondents and survey topics, are noted. **The aim of the study** is to develop a method for evaluating measures to form a comfortable urban environment based on semantic comparison of citizens' opinions from social networks, search queries and descriptions of the measures themselves using natural language processing algorithms. **Materials and methods.** To realize the set goal, an integrated approach consisting of four main stages is used. Stage 1: Pre-processing of initial text data – noise removal, reduction of words to their initial form (lemmatization) using the pymorphy2 library, identification of parts of speech (POS-tagging). Stage 2: Extraction of key word combinations (N-grams) using the TF-IDF algorithm, taking into account the frequency of usage within individual messages and in the whole text array. Calculation of N-grams significance rank. Stage 3: Obtaining a vector representation (word embeddings) for each key N-gram using the pre-trained SBERT neural network model. Step 4: Computing a measure of semantic similarity of vector representations of N-grams from different text arrays (citizens' opinions and event descriptions) based on cosine distance. The texts of citizens' opinions and descriptions of measures aimed at the formation of a comfortable urban environment related to the Khanty-Mansi Autonomous Okrug are used as input data. **Results.** The experiment has shown that most of the analyzed measures aimed at the formation of a comfortable urban environment are poorly correlated with the real needs of citizens. **Conclusion.** The proposed method can be used in decision support systems to evaluate and select the most effective measures.

**Keywords:** semantic analysis, TF-IDF, transformers, sentence transformer, comfortable urban environment, NLP

**For citation:** Melnikov A.V., Galagan K.V. Method of assessing the effectiveness of measures to form a comfortable urban environment. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*. 2024;24(1):87–95. (In Russ.) DOI: 10.14529/ctcr240108

### Введение

В настоящее время формирование качественной городской среды невозможно представить без участия граждан, жителей города. Более того, вовлечение граждан в мероприятия по формированию качественной городской среды (МФКГС) является одной из целей национального проекта «Жилье и городская среда» (Нацпроект) [1]. Значимость данной цели для Нацпроекта подчеркивается утверждением Минстроем России методических рекомендаций по вовлечению граждан и иных лиц, их объединений в решении вопросов развития городской среды (Методика) [2]. Для граждан вовлеченность в МФКГС означает возможность преобразовать городское пространство согласно своим предпочтениям и потребностям.

Методика предписывает организацию обратной связи от граждан на всех этапах реализации МФКГС. Для этого предполагается, в частности, организовать электронные площадки для сбора обратной связи и выбора МФКГС для реализации. Исследованию вопроса оценки и эффективности подобных платформ посвящена работа [3]. Как показывает исследование, описанное в работе [4], такой способ взаимодействия с гражданами низкоэффективен.

На получение более широкой и качественной обратной связи направлено проведение различных социологических исследований, выявляющих потребности жителей, связанные с комфортной городской средой (КГС). Например, в работах [5–8] представлены результаты таких исследований. При исследованиях в области КГС такой способ получения обратной связи может иметь ряд ограничений. Прежде всего это ограниченность географии и выборки респондентов. Также в социологических опросах жестко ограничены тематика вопросов и варианты ответов.

Снять эти ограничения позволяет получение обратной связи от неограниченного количества респондентов с неограниченным количеством тематик и вопросов. В работах [9, 10] авторы, используя нейросетевые методы обработки естественного языка (NLP), извлекают полезную информацию из сообщений пользователей социальной сети Twitter, что позволяет говорить о существенном расширении географии, аудитории и тематик. Также в работе [11] для получения обратной связи предлагается анализировать сведения об использовании горожанами определенных мобильных приложений.

Такие методы получения обратной связи более оперативны, чем классические социологические опросы.

В настоящей статье представляется описание метода оценки эффективности реализации МФКГС (МОЭР). Метод позволяет на основании мнений граждан, связанных с комфортной городской средой (МГКГС), в социальных медиа, запросов в поисковой системе «Яндекс», а также описаний МФКГС оценить, насколько предлагаемые для реализации МФКГС отвечают ожиданиям граждан. Метод основан на использовании алгоритмов и нейросетевых моделей NLP.

Также произведена экспериментальная проверка метода с использованием прототипа информационной системы (Система).

## 1. Описание МОЭР

Предлагаемый метод заключается в использовании семантического анализа текстовых сообщений с помощью алгоритмов и нейросетевых моделей NLP. Существенная особенность МОЭР состоит в синтезе алгоритма TF-IDF [12] для нахождения и извлечения из текста семантически значимых фраз и словосочетаний (N-грамм, где N – число слов во фразе или словосочетании) и применения нейросетевых моделей-трансформеров, основанных на архитектуре BERT [13]. Такой подход позволяет сохранить семантическое ядро сообщений, ранжировать полученные N-граммы, получить и проанализировать семантически значимые векторные представления N-грамм на семантическое сходство.

МОЭР включает в себя следующие этапы:

Этап 0. Предобработка и очистка текстовых сообщений.

Этап 1. Извлечение N-грамм и расчет ранга их значимости  $\mathbb{R}$ .

Этап 2. Получение вектора контекстного представления (ВКП) для каждой N-граммы.

Этап 3. Расчет семантического сходства ВКП.

На рис. 1 представлена схема этапов и задач метода МОЭР.

### 1.1. Этап 0. Предобработка и очистка датасета текстовых сообщений

На данном этапе выполняются следующие задачи:

- формирование набора датасетов путем разделения исходного датасета по тематикам сообщений;
- очистка исходных текстовых сообщений от спецсимволов и латиницы с помощью регулярных выражений;
- преобразование слов в сообщениях к их начальной форме (лемматизация);
- частеречная разметка, определение частей речи (POS-tagging).

Лемматизация и POS-tagging производятся с применением методов Python-библиотеки `ru morphology2` [14]. Приведение текстов сообщений к начальной форме необходимо для унификации работы алгоритма TF-IDF и последующего ранжирования N-грамм. POS-tagging используется для управления частеречным составом N-грамм.

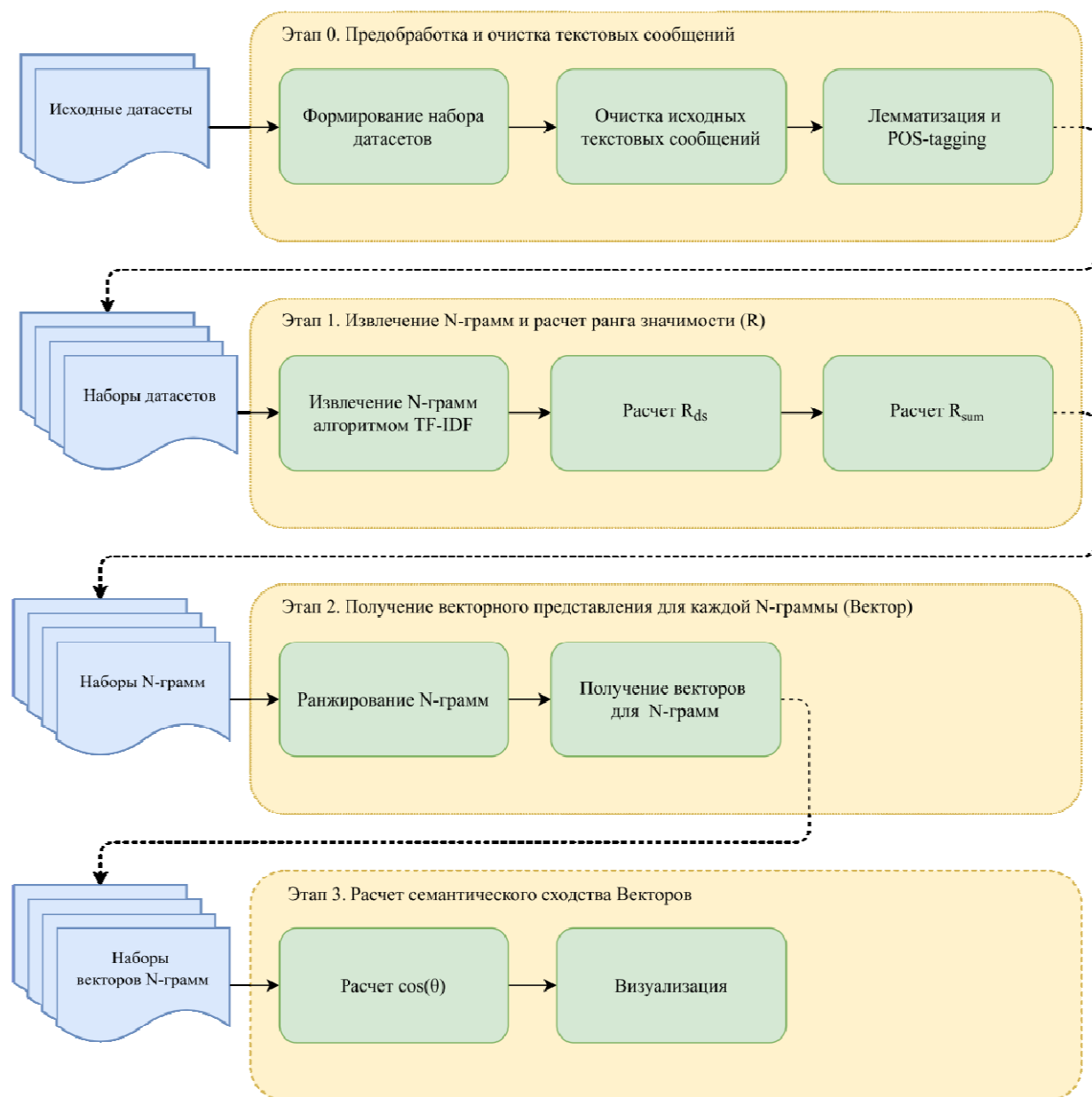


Рис. 1. Схема этапов и задач метода МОЭР  
Fig. 1. Scheme of stages and tasks of the method MEI

### 1.2. Этап 1. Извлечение N-грамм и расчет их ранга значимости

На данном этапе выполняются следующие задачи:

- извлечение N-грамм из датасетов;
- расчет показателя  $R$  N-граммы для датасета;
- расчет показателя  $R$  N-граммы для набора датасетов.

Для извлечения N-грамм из датасетов используется алгоритм TF-IDF. Расчет значения  $R$  N-граммы для датасета ( $R_{ds}$ ) рассчитывается по формуле

$$R_{ds} = TF_{n-gramm} \times IDF_{n-gramm}, \quad (1)$$

где  $R_{ds}$  – ранг N-граммы в датасете;  $TF_{n-gramm}$  – частота N-граммы в пределах отдельного сообщения;  $IDF_{n-gramm}$  – инверсия частоты, с которой N-грамма встречается во всех сообщениях датасета. Расчет значения  $R$  для набора датасетов ( $R_{sum}$ ) производится по формуле

$$R_{sum} = \sum_{k=1}^m R_{ds_k}, \quad (2)$$

где  $R_{sum}$  – суммарный ранг N-граммы в наборе;  $R_{ds}$  – ранг N-граммы в датасете.

**1.3. Этап 2. Получение ВКП для каждой N-граммы**

На данном этапе выполняются следующие задачи:

- ранжирование N-грамм;
- получение ВКП для каждой N-граммы.

Получение ВКП для N-грамм является ключевой особенностью МОЭР. Для получения ВКП N-грамм используется нейросетевая модель, основанная на архитектуре BERT. Это позволяет сформировать векторное пространство, пригодное для различного рода математических преобразований, в частности, расчета меры сходства векторов.

**1.4. Этап 3. Расчет семантического сходства ВКП**

На данном этапе выполняются следующие задачи:

- расчет косинусного расстояния;
- визуализация.

Для расчета меры сходства ВКП в Методе предлагается использовать метрику косинусного расстояния. Метрика принимает значение от 0, что соответствует максимально далекому расстоянию, до 1, что соответствует максимально близкому расстоянию между векторами. Косинусное расстояние вычисляется по формуле

$$\text{similarity} = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}, \tag{3}$$

где A и B – сравниваемые векторы N-грамм;  $\cos \theta$  – косинусное расстояние.

**2. Данные и эксперимент**

Цель эксперимента – определить семантическую близость между мнениями граждан о проблемах городской среды и описаниями мероприятий, направленных на формирование комфортной городской среды. Данные представляют собой две коллекции датасетов с текстами МГКГС, относящихся к территории Ханты-Мансийского автономного округа – Югры, и описания МФКГС, проводимые властями муниципальных образований. Для каждого текстового сообщения в датасете устанавливается категория сообщения, наименование муниципального образования, месяц и год публикации сообщения. В табл. 1 представлены краткие характеристики коллекций датасетов.

Таблица 1  
Table 1

**Краткие характеристики коллекций датасетов**  
Brief characteristics of dataset collections

Коллекция датасетов	Количество датасетов	Общее количество сообщений в наборе
МГКГС	6224	73 897
МФКГС	108	290

Первичная обработка датасета проводится с помощью регулярных выражений. Удаляются все символы, кроме кириллических и знаков препинания. Также удаляются «стоп-слова»:

- фразы общепринятых обращений;
- фразы приветствий и прощаний;
- названия населенных пунктов, улиц;
- имена собственные.

Из каждого датасета алгоритмом TF-IDF извлекаются N-граммы с параметром N от 2 до 4 и количеством извлеченных N-грамм = 100. Значение  $R_{ds}$  и  $R_{sum}$  рассчитываются по формулам (1) и (2) соответственно. Значения  $R_{sum}$  для ТОП-20 извлеченных N-грамм приведены в табл. 2.

Таблица 2  
Table 2

**Значения  $R_{sum}$  для ТОП-20 извлеченных N-грамм**  
 $R_{sum}$  values for TOP-20 extracted N-grams

Значение N	Набор датасетов МГКГС		Датасет МФКГС	
	Max $R_{sum}$	Min $R_{sum}$	Max $R_{sum}$	Min $R_{sum}$
N = 2	<b>59,2314</b>	9,7534	<b>3,8184</b>	1,0173
N = 3	3,7017	1,1024	1,0693	0,5232
N = 4	2,5768	0,8919	0,9183	0,3788

Для дальнейшего анализа из каждого набора датасетов отбираются два набора ТОП-20 N-грамм для  $N = 2$ . ТОП-20 значений  $R_{sum}$  N-грамм при  $N = 2$  для каждого набора приведены на рис. 2 и 3.

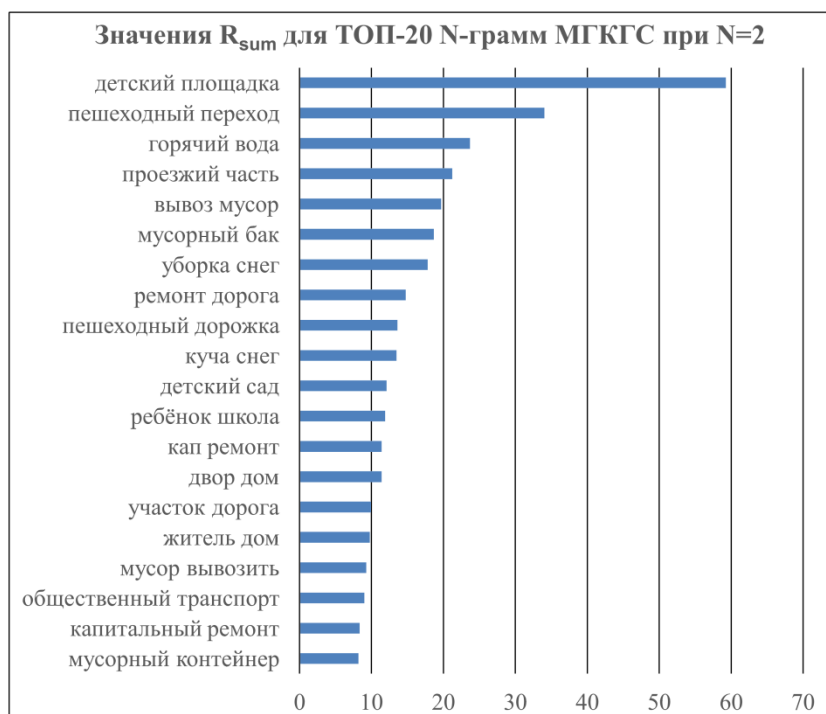


Рис. 2. Значения  $R_{sum}$  для ТОП-20 N-грамм датасета МГКГС, при  $N = 2$  (N-граммы лемматизированы)  
Fig. 2.  $R_{sum}$  values for the TOP-20 N-grams of the OCCUE dataset, with  $N = 2$  (N-grams are lemmatized)



Рис. 3. Значения  $R_{sum}$  для ТОП-20 N-грамм датасета МФКГС, при  $N = 2$  (N-граммы лемматизированы)  
Fig. 3.  $R_{sum}$  values for the TOP-20 N-grams of the MCQUE dataset, with  $N = 2$  (N-grams are lemmatized)

Для преобразования N-грамм в ВКП использована предобученная нейросетевая модель sentence-BERT [15]. Никакой тонкой настройки и дообучения для sentence-BERT не проводилось.

Для каждой N-граммы получен одномерный вектор контекстного представления размерностью 512.

Расчет косинусного расстояния проводится между векторами двух наборов ТОП-20 по формуле (3). Визуализация результата расчетов приведена на рис. 4.

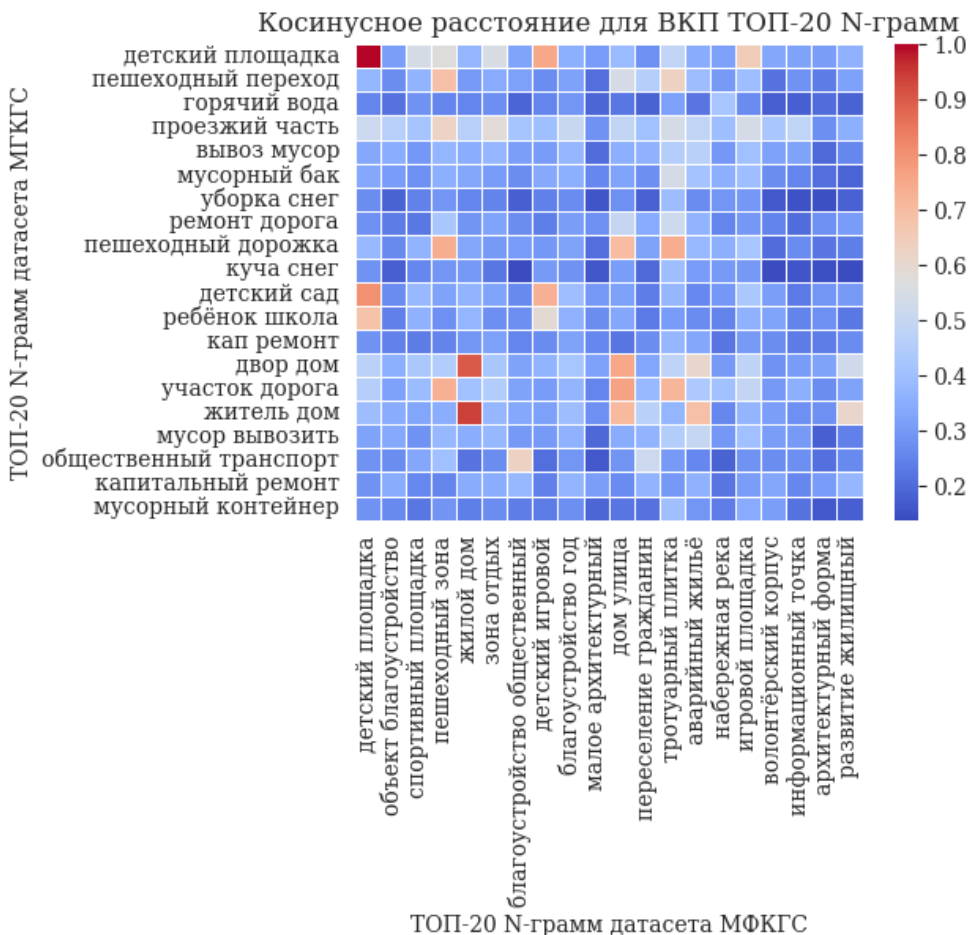


Рис. 4. Результат расчета косинусного расстояния между ТОП-20 N-грамм (N-граммы лемматизированы)  
 Fig. 4. The result of calculating the cosine distance between the TOP-20 N-grams (N-grams are lemmatized)

**Заключение**

Рассмотренный в статье метод оценки мероприятий по формированию комфортной городской среды позволяет эффективно ранжировать и сопоставлять МФКГС с актуальными потребностями граждан.

С помощью МОЭР была проведена оценка эффективности МФКГС, проведенных или планируемых к проведению в 2020–21 гг. на территории Ханты-Мансийского автономного округа Югры.

Результаты проведенного с использованием МОЭР эксперимента указывают на то, что подавляющее количество МФКГС слабо связаны с реальными потребностями граждан в вопросах КГС и, как следствие, будут низкоэффективны.

Приведенный МОЭР может быть применен в системах поддержки принятия решений при проведении оценки значимости МФКГС. Своевременная и качественная реализация МФКГС, направленных на решение реальных потребностей жителей города, положительно скажется на оценке гражданами деятельности местных властей и повысит вовлеченность граждан в вопросы управления городскими пространствами.

### Список литературы

1. Паспорт национального проекта «Жилье и городская среда» // Минстрой России: сайт. URL: <https://minstroyrf.gov.ru/docs/221887/> (дата обращения: 21.09.2021).
2. Об утверждении методических рекомендаций по вовлечению граждан, их объединений и иных лиц в решение вопросов развития городской среды: приказ Минстроя России № 913/пр от 30.12.2020. URL: [https://minstroyrf.gov.ru/upload/iblock/315/30\\_12\\_2020\\_913\\_pr\\_Metodicheskie\\_rekomendatsii\\_po\\_vovlecheniyu.pdf](https://minstroyrf.gov.ru/upload/iblock/315/30_12_2020_913_pr_Metodicheskie_rekomendatsii_po_vovlecheniyu.pdf) (дата обращения: 31.03.2022).
3. Мухаметов Д.Р. Модели платформ вовлечения граждан для создания в России умных городов нового поколения // Вопросы инновационной экономики. 2020. Т. 10, № 3. С. 1605–1622. DOI: 10.18334/vines.10.3.110683
4. Опыт общественного участия в планировании комфортной городской среды на примере Архангельской области / А.Г. Деменев, Т.Ф. Шубина, П.В. Шубина и др. // Арктика и Север. 2018. № 33. С. 91–117. DOI: 10.17238/issn2221-2698.2018.33.91
5. Аксенова В.В. Комфортная городская среда: общественное мнение москвичей // Социальная политика и социология. 2020. Т. 19, № 4 (137). С. 76–84. DOI: 10.17922/2071-36652020-19-4-76-84
6. Богданова Л.П., Глушкова М.А. Оценка качества городской среды населением города Твери // Вестник Тверского государственного университета. Серия: География и геоэкология. 2021. № 2 (34). С. 14–24. DOI: 10.26456/2226-7719-2021-2-14-24
7. Дунаева Д.О. Дискурсивные практики горожан как коммуникативный механизм формирования образа комфортного города (опыт полевого исследования) // Вестник Томского государственного университета. Философия. Социология. Политология. 2021. № 60. С. 137–150. DOI: 10.17223/1998863X/60/13
8. Руссова О.Н., Смак Т.С., Тарасов И.А. Оценка комфортности городской среды как фактор социального самочувствия городских жителей Архангельской области // Арктика и Север. 2020. № 41. Р. 236–247. DOI: 10.37482/issn2221-2698.2020.41.236
9. Detecting citizen problems and their locations using twitter data / G. Abalı, E. Karaarslan, A. Hürriyetođlu, F. Dalkılıç // 2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG). Istanbul, Turkey, 2018. P. 30–33. DOI: 10.1109/SGCF.2018.8408936
10. Estévez-Ortiz F.-J., García-Jiménez A., Glösekötter P. An application of people’s sentiment from social media to smart cities // El profesional de la información. 2016. Vol. 25, no. 6. P. 851–858. DOI: 10.3145/epi.2016.nov.02
11. Алексеев С.А. Коммуникативно-информационные технологии в повышении качества городской среды // Управление устойчивым развитием. 2019. № 6 (25). С. 36–40.
12. Spärck Jones K. A statistical interpretation of term specificity and its application in retrieval // Journal of Documentation. 2004. Vol. 60, no. 5. P. 493–502. DOI: 10.1108/00220410410560573
13. Attention Is All You Need / Ashish Vaswani, Noam Shazeer, Niki Parmar et al. // 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA. 2017. URL: <https://arxiv.org/pdf/1706.03762.pdf> (дата обращения: 30.05.2022).
14. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages. 2015. URL: <https://arxiv.org/pdf/1503.07283.pdf> (дата обращения: 30.05.2022).
15. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 2019. URL: <https://arxiv.org/pdf/1908.10084.pdf> (дата обращения: 30.05.2022).

### References

1. *Pasport natsional'nogo proekta "Zhil'e i gorodskaya sreda"* [Passport of the national project "Housing and Urban Environment"]. *Ministry of Construction of Russia: website*. (In Russ.) Available at: <https://minstroyrf.gov.ru/docs/22188> (accessed 21.09.2021)
2. *Ob utverzhdenii metodicheskikh rekomendatsiy po vovlecheniyu grazhdan, ikh ob"edineniy i inykh lits v reshenie voprosov razvitiya gorodskoy sredy: prikaz Ministroya Rossii № 913/pr ot 30.12.2020* [On approval of guidelines for the involvement of citizens, their associations and other persons in resolving issues of the development of the urban environment: order of the Ministry of Construction of Russia No. 913/pr dated 30.12.2020]. (In Russ.) Available at: [https://minstroyrf.gov.ru/upload/iblock/315/30\\_12\\_2020\\_913\\_pr\\_Metodicheskie\\_rekomendatsii\\_po\\_vovlecheniyu.pdf](https://minstroyrf.gov.ru/upload/iblock/315/30_12_2020_913_pr_Metodicheskie_rekomendatsii_po_vovlecheniyu.pdf) (accessed 31.03.2022).
3. Mukhametov D.R. Models of citizen engagement platforms for creating new generation smart



cities in Russia. *Russian Journal of Innovation Economics*. 2020;10(3):1605–1622. (in Russ.) DOI: 10.18334/vinec.10.3.110683

4. Demenev A.G., Shubina T.F., Shubina P.V. et al. Public participation in planning a comfortable urban environment on the example of the Arkhangelsk region. *Arctic and North*. 2018;33:91–117. DOI: 10.17238/issn2221-2698.2018.33.91 (In Russ.)

5. Aksenova V.V. Comfortable urban environment: public opinion of Muscovites. *Socialnaya politika i sociologiya = Social policy and sociology*, 2020;19(4(137)):76–84. (In Russ.) DOI: 10.17922/2071-36652020-19-4-76-84

6. Bogdanova L.P., Glushkova M.A. Assessment of quality of the city environment by populations of the city Tver. *Herald of Tver state university. Series: Geography and geocology*. 2021;2(34):14–24. (In Russ.) DOI: 10.26456/2226-7719-2021-2-14-24

7. Dunaeva D.O. Discursive practices of citizens as a communicative mechanism for forming the image of a comfortable city (practical research experience). *Tomsk state university journal of philosophy, sociology and political science*. 2021;60:137–150. (In Russ.) DOI: 10.17223/1998863X/60/13

8. Russova O.N., Smak T.S., Tarasov I.A. Assessment of the comfort of the urban environment as a factor in the social well-being of citizens of the Arkhangelsk oblast. *Arctic and North*. 2020;41:236–247. (In Russ.) DOI: 10.37482/issn2221-2698.2020.41.236

9. Abalı G., Karaarslan E., Hürriyetoğlu A., Dalkılıç F. Detecting citizen problems and their locations using twitter data. In: *2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*. Istanbul, Turkey; 2018. P. 30–33. DOI: 10.1109/SGCF.2018.8408936

10. Estévez-Ortiz F.-J., García-Jiménez A., Glösekötter P. An application of people's sentiment from social media to smart cities. *El profesional de la información*. 2016;25(6):851–858. DOI: 10.3145/epi.2016.nov.02

11. Alekseev S.A. Communicative information technologies to increase the quality of the city environment. *Managing Sustainable Development*. 2019;6(25):36–40. (In Russ.)

12. Spärck Jones K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*. 2004;60(5):493–502. DOI: 10.1108/00220410410560573

13. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA. 2017. Available at: <https://arxiv.org/pdf/1706.03762.pdf> (accessed 30.05.2022).

14. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages. 2015. Available at: <https://arxiv.org/pdf/1503.07283.pdf> (accessed 30.05.2022).

15. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 2019. Available at: <https://arxiv.org/pdf/1908.10084.pdf> (accessed 30.05.2022).

### **Информация об авторах**

**Мельников Андрей Витальевич**, д-р техн. наук, проф., директор, Югорский научно-исследовательский институт информационных технологий, Ханты-Мансийск, Россия; [MelnikovAV@uriit.ru](mailto:MelnikovAV@uriit.ru).

**Галаган Константин Владимирович**, аспирант кафедры цифровых технологий Института цифровой экономики, Югорский государственный университет, Ханты-Мансийск, Россия; [galagankv@gmail.com](mailto:galagankv@gmail.com).

### **Information about the authors**

**Andrey V. Melnikov**, Dr. Sci. (Eng.), Prof., Director, Ugra Research Institute of Information Technologies, Khanty-Mansiysk, Russia; [MelnikovAV@uriit.ru](mailto:MelnikovAV@uriit.ru).

**Konstantin V. Galagan**, Postgraduate Student of the Department of Digital Technologies, Institute of Digital Economy, Yugra State University, Khanty-Mansiysk, Russia; [galagankv@gmail.com](mailto:galagankv@gmail.com).

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

Авторы заявляют об отсутствии конфликта интересов.

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

The authors declare no conflicts of interests.

**Статья поступила в редакцию 11.09.2023**

**The article was submitted 11.09.2023**