12. *Makarova V., Petrushin V.* Ruslana: a database of Russian emotional utterances, *7th International Conference on Spoken Language Processing*, 2002, pp. 2041-2044.
13. Russian emotional speech dialogs (RESD). Available at: https://www.kaggle.com/datasets/ar4ikov/resd-dataset.
14. *Sahidullah M., Saha G.* Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition, *Speech Communication*, 2012, Vol. 54, Issue 4, pp. 543-565.
15. *Jagtap S., Desai K., Patil J.* A Survey on Speech Emotion Recognition Using MFCC and Different Classifier, 2022.
16. *Badr Y., Mukherjee P., Thumati S.* Speech Emotion Recognition using MFCC and Hybrif Neural Networks, *13th International Conference on Neural Computation Theory and Applications*, 2021.
17. Librosa – librosa 0.10.2 documentation. – Режим доступа: https://librosa.org/doc/latest/index.html.
18. *Hochreiter S., Schmidhuber J.* Long Short-Term Memory, *Neural Computation*, 1997, No. 9 (8), pp. 1735-1780.
19. *Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N., Minkin F*. Large raw emotional dataset with aggregation mechanism, *ArXiv (Cornell University)*, 2022.
20. *Lemaev V.I., Lukashevich N.V.* Avtomaticheskaya klassifikatsiya emotsiy v rechi: metody i dannye [Automatic classification of speech emotions: methods and data], *Litera. Nota bene*, 2024, No. 4, pp. 159-173.

Букина Полина Германовна – Томский государственный университет систем управления и радиоэлектроники; e-mail: bukina.polina2014@gmail.com; г. Томск, Россия; кафедра безопасности информационных систем; студент.

Меринов Арсен Арзуманович – Томский государственный университет систем управления и радиоэлектроники; e-mail: merinovarsen@mail.ru; г. Томск, Россия; кафедра безопасности информационных систем; студент.

Харченко Сергей Сергеевич – Томский государственный университет систем управления и радиоэлектроники; e-mail: kss@fb.tusur.ru; г. Томск, Россия; кафедра безопасности информационных систем; к.т.н.; доцент.

Костюченко Евгений Юрьевич – Томский государственный университет систем управления и радиоэлектроники; e-mail: key@keva.tusur.ru; г. Томск, Россия; кафедра безопасности информационных систем; к.т.н.; и.о. зав. кафедрой.

Bukina Polina Germanovna – Tomsk State University of Control Systems and Radioelectronics; e-mail: bukina.polina2014@gmail.com; Tomsk, Russia; the Department of Information Security Systems; student.

Merinov Arsen Arzumanovich – Tomsk State University of Control Systems and Radioelectronics; e-mail: merinovarsen@mail.ru; Tomsk, Russia; the Department of Information Security Systems; student.

Kharchenko Sergey Sergeevich – Tomsk State University of Control Systems and Radioelectronics; e-mail: kss@fb.tusur.ru; Tomsk, Russia; the Department of Information Security Systems; cand. of eng. sc.; associate professor.

Kostyuchenko Evgeny Yurievich – Tomsk State University of Control Systems and Radioelectronics; e-mail: key@keva.tusur.ru; Tomsk, Russia; the Department of Information Security Systems; cand. of eng. sc.; acting head of department.

УДК 004.89 DOI 10.18522/2311-3103-2025-6-248-262

Ж.Х. Мохаммад

МЕТОДИКА ПОСТРОЕНИЯ И ОЦЕНКИ ОНТОЛОГИЧЕСКОГО ПРОФИЛЯ ДЛЯ СИСТЕМ ПЕРСОНАЛИЗАЦИИ КОНТЕНТА

Данная статья посвящена разработке и апробации методики построения онтологического профиля, предназначенного для использования в системах персонализации контента. В работе детально описана модульная архитектура веб-системы персонализации, иллюстрирующая методы и алгоритмы обработки и анализа текста на каждом этапе, а также представлен пошаговый алгоритм создания онтологии. Методика включает первичную обработку данных: извлечение ключевых слов и словосочетаний, их иерархическую кластеризацию для выявления семантической

структуры предметной области. Далее следует этап определения пороговых значений для отсева малозначимых связей, извлечения и формализации взаимосвязей между концептами с использованием методов обработки естественного языка, таких, как разрешение лексической неоднозначности и извлечение связей на основе семантического сходства. Для реализации этого процесса был разработан интегрированный конвейер (pipeline), объединяющий усовершенствованные алгоритмы, предложенные автором в предыдущих исследованиях, а именно: алгоритм извлечения ключевых фраз из отдельного текста на основе семантического сходства и модифицированный алгоритм разрешения лексической многозначности слов. В рамках данного конвейера также были оптимально интегрированы все необходимые инструменты обработки естественного языка, обеспечивающие эффективную работу указанных методов в процессе автоматического построения онтологии из текста. Особое внимание в исследовании уделяется комплексной оценке полученной онтологии с использованием специализированного набора критериев, позволяющих объективно оценить качество, полноту и непротиворечивость построенного профиля. Важной частью работы является проведение вычислительного эксперимента, который наглядно демонстрирует влияние каждого из этапов обработки данных на итоговое качество и эффективность онтологии. Показано, что предложенная методика позволяет построить практичную, масштабируемую и релевантную онтологию, готовую к промышленному внедрению и интеграции в системы персонализации для повышения их точности и адаптивности.

Онтологический профиль; персонализация контента; извлечение ключевых слов; иерархическая кластеризация; оценка онтологий; семантическая модель.

J.H. Mohammad

METHODOLOGY FOR CONSTRUCTING AND EVALUATING AN ONTOLOGICAL PROFILE FOR CONTENT PERSONALIZATION SYSTEMS: STAGES AND EVALUATION CRITERIA

This article presents the development and testing of a methodology for building an ontological profile designed for content personalization systems. It details the modular architecture of a web-based personalization system, illustrating the text processing and analysis methods and algorithms employed at each stage, and provides a step-by-step procedure for ontology creation. The methodology encompasses primary data processing, including the extraction of keywords and phrases, followed by their hierarchical clustering to reveal the semantic structure of the domain. Subsequent stages involve defining thresholds to filter out insignificant connections, and extracting and formalizing relationships between concepts using natural language processing techniques such as word-sense disambiguation and semantic similarity-based relationship extraction. An integrated pipeline was developed to implement this process, combining improved algorithms proposed by the author in previous studies, namely, an algorithm for extracting key phrases from individual text based on semantic similarity and a modified algorithm for word sense disambiguation. This pipeline also optimally integrated all necessary natural language processing tools, ensuring the efficient operation of these methods in the process of automatically constructing an ontology from text. The study places particular emphasis on a comprehensive evaluation of the resulting ontology using a specialized set of criteria designed to objectively assess the profile's quality, completeness, and consistency. A important component of the work is a computational experiment that clearly demonstrates the impact of each data processing stage on the final quality and efficacy of the ontology. The results show that the proposed method enables the construction of a practical, scalable, and relevant ontology, suitable for industrial deployment and integration into personalization systems to enhance their accuracy and adaptability.

Ontology profile; content personalization; keyword extraction; hierarchical clustering; ontology evaluation; semantic model.

Введение. В условиях стремительного развития технологий искусственного интеллекта (ИИ), экспоненциального роста объёмов данных и распространения цифровых сервисов (социальные сети, Интернет вещей, большие данные, облачные вычисления) адаптация информации под индивидуальные потребности пользователей приобретает ключевое значение. Пользователи сталкиваются с серьёзными трудностями при поиске релевантной информации, что обусловлено беспрецедентным ростом объёмов доступного контента и его крайним разнообразием по форматам и источникам. Эта ситуация приводит к феномену информационной перегрузки, под которой понимается состояние, когда пользователь подвергается воздействию чрезмерного количества данных, значительная часть которых не соответствует его актуальным потребностям и интересам [1, 2], а также разнообразия форматов и источников информации. Эти факторы приводят к существенному снижению эф-

фективности информационного поиска и росту когнитивной нагрузки, которая необходима для фильтрации нерелевантного контента, навигации в условиях информационного шума и принятия решений в условиях семантической неоднозначности.

Указанные обстоятельства актуализируют необходимость разработки более совершенных технологий фильтрации и персонализации контента. Это подчеркивает важность Веб-Инженерии, которая направлена на создание веб-систем, соответствующих потребностям пользователей [3, 4]. Персонализация является одним из приложений Веб-Инженерии, включающим набор информационных процессов, направленных на обеспечение персонализированного поиска для каждого пользователя с учётом его предпочтений, поведения и других соответствующих данных [5–8]. Она включает динамическую адаптацию веб-контента, рекомендаций и пользовательских интерфейсов для соответствия конкретным потребностям и интересам каждого пользователя.

Основным ключом для персонализации веб-контента является создание профиля пользователя**,** который включает сбор и анализ данных об отдельных пользователях для понимания их предпочтений, интересов и поведения [8–11]. Современные подходы к построению пользовательских профилей опираются на две парадигмы: методы на основе концептов и методы на основе онтологических моделей. Методы на основе концептов, в основном, заключаются в извлечении ключевых слов или фраз из пользовательских данных, анализе их частоты и определении значимых терминов [12]. Несмотря на простоту и вычислительную эффективность, эти методы ограничены в своей способности учитывать базовую семантику текстового контента. Они рассматривают слова как изолированные единицы и игнорируют контекстные связи между ними. Это ограничение становится особенно проблематичным в случаях *неоднозначности* на уровне предметной области, когда один и тот же термин может иметь разные интерпретации в разных областях. Например, термин «cell» (клетка) может относиться к биологической клетке в медицинском контексте или к мобильному телефону в контексте телекоммуникаций. Это приводит к созданию профилей, перегруженных шумом и двусмысленностью, что повышает частоту ошибок алгоритмов функционирования систем персонализации.

Для устранения ограничений, связанных с использованием только ключевых слов, в некоторых работах применяются внешние источники данных, такие, как проект открытого каталога (англ. Open Directory Project, ODP). ODP предоставляет собой обширную иерархическую структуру категорий веб-сайтов, которая служит основой для создания онтологий и классификации контента. ODP используется для сопоставления терминов со значимыми концептами. Например, в [13–16], пользовательские интересы извлекаются из веб-документов и представляются в виде взвешенных категорий из ODP. Эти методы являются шагом к подходам, основанным на онтологиях, но им не хватает семантических связей между концептами.

Методы онтологического подхода, напротив, используют формальные онтологии для построения пользовательских профилей и структурируют интересы пользователей в виде онтологических концептов, используя формальные семантические отношения, например, связывая «глубокое обучение» с «искусственным интеллектом». Однако такие методы часто опираются на статические, предопределённые онтологии, которые не способны адаптироваться к динамичному характеру предпочтений пользователей или новым предметным областям. В [14] пользовательские профили создаются на основе URL-адресов и сопоставляются с категориями из онтологий *OpenDNS* и *DBpedia* [17, 18]. Это обогащает пользовательские профили семантическими связями, но может не обеспечивать точности в захвате конкретных интересов.

Соответственно, методы, основанные на концептах, делают акцент на простоту, но им не хватает семантической глубины, в то время как методы, использующие онтологический подход, предлагают более богатую семантику за счет вычислительной эффективности и адаптивности к изменяющемуся поведению пользователей или новым предметным областям. В результате, существующие методы по-прежнему недостаточно эффективны для обработки текстовых данных и извлечения знаний, что не позволяет достичь необходимого уровня эффективности систем веб-персонализации и делает их неспособными снизить когнитивную нагрузку на пользователя.

Для выявления указанных ограничений в данной статье представлен аналитический обзор методов построения профилей пользователей в веб-системах персонализации. Особое внимание уделяется технологиям обработки текстовых данных для задач адаптации контента, с подробным описанием применяемых методов на каждом этапе. В качестве практического вклада исследования была разработана программная реализация методики автоматического построения онтологии из текста с использованием технологий обработки естественного языка. Это включает извлечение ключевых фраз, разрешение лексической многозначности и выявление семантических связей на основе оценки контекстуального сходства. Для реализации данной методики был создан комплексный конвейер обработки данных (pipeline), интегрирующий усовершенствованные алгоритмы, предложенные авторами в предыдущих работах, с признанными инструментами текстовой аналитики.

1. Этапы персонализации контента веб-сайтов. Процесс персонализации веб-контента включает в себя ряд шагов, которые позволяют предоставлять пользователям индивидуализированные впечатления на основе их предпочтений и поведения [19]. На рис. 1 представлена последовательность этапов процесса в системе персонализации веб-контента на основе методов искусственного интеллекта и машинного обучения, которые подробно описаны ниже.

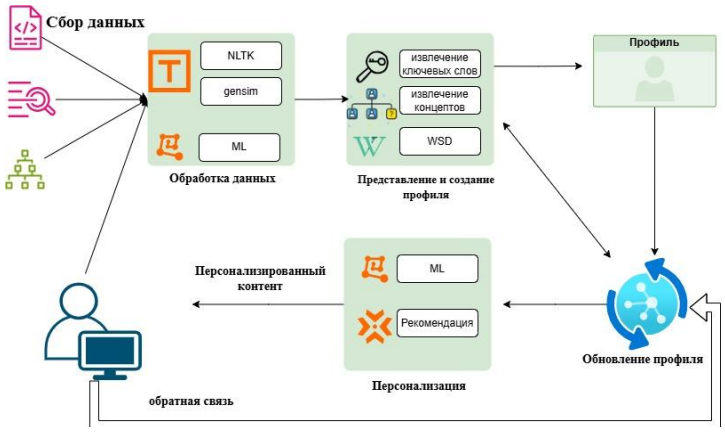

Рис. 1. Архитектура системы персонализации веб-контента

1.1. ***Сбор данных****.* Первый этап включает в себя сбор необходимых данных о пользователе. Это может включать как явные данные, предоставленные пользователем при регистрации или в ходе опросов, так и неявные данные, собранные в результате взаимодействия пользователя с веб-сайтом или приложением. Неявные данные могут включать историю просмотров, поисковые запросы, клики, историю покупок и активность в социальных сетях [8, 20].

1.2. ***Обработка данных****:* после сбора данные должны пройти обработку для удаления шума, заполнения отсутствующих значений и приведения в необходимый формат для анализа. Этот этап включает такие процессы, как очистка данных, нормализация и создание признаков. Для этого используются инструменты NLP:

- NLTK [21]: Библиотека Python для токенизации, стемминга, лемматизации и синтаксического анализа;
- WordNet [22]: Лексическая база для поиска синонимов, лемматизации и разрешения неоднозначностей;
- - SpaCy: это эффективная библиотека для обработки естественного языка (NLP) на Python [23].

1.3. ***Представление и создание профиля.*** Персонализация контента в современных системах основывается на построении профиля пользователя, который отражает его информационные потребности и предпочтения [24]. Эффективность рекомендаций напрямую зависит от способа представления этого профиля. Наиболее распространенным методом является

векторная модель, где интересы пользователя описываются взвешенными векторами ключевых слов или концептов, извлеченных из просмотренных документов. Этот подход отличается простотой реализации и вычислительной эффективностью, но сталкивается с проблемами полисемии и синонимии, что может снижать точность рекомендаций [25–27].

Для преодоления ограничений векторной модели применяются семантические сети, отображающие концепты и связи между ними. Этот подход, требующий сопоставления слов концептам через онтологии или машинное обучение, эффективно решает проблемы многозначности и синонимии, несмотря на сложность реализации. Альтернативой выступает иерархическая модель, организующая интересы пользователя в виде древовидной структуры, однако её гибкость ограничена зависимостью от чёткой иерархии.

Более совершенным решением является граф знаний, интегрирующий преимущества семантических сетей и иерархических моделей. Простые графы (без атрибутов) подходят для анализа топологии данных, например, в социальных сетях. Расширенные графы (с атрибутами) включают метаданные для точного описания сущностей и широко используются в рекомендательных системах и медицинских базах знаний, обеспечивая глубинный контекстный анализ. Недостатком является высокая ресурсоёмкость построения таких графов и зависимость от качества алгоритмов обработки данных.

В данной работе профиль пользователя строится на основе онтологической модели (онтологии) с использованием структуры графа знаний для хранения концептов, ключевых фраз и атрибутов данных. Онтология представляет собой формальную, явную спецификацию разделяемой концептуализации [28], определяющую классы, свойства, отношения и ограничения в рамках конкретной предметной области. В отличие от этого, граф знаний является практической реализацией, формирующей сеть реальных сущностей (узлов), связанных между собой отношениями (рёбрами). Известные примеры включают Google Knowledge Graph и Wikidata [17].

Основные операции, выполняемые на этом этапе, заключаются в анализе содержания описаний объектов, с которыми взаимодействовал пользователь, для выявления основных характеристик его предпочтений. Этот аналитический процесс использует ряд алгоритмов обработки естественного языка (NLP), включая извлечение ключевых фраз, идентификацию концептов и связывание ключевых фраз и концептов на более высоком уровне абстракции. Кроме того, процесс включает в себя выявление взаимосвязей между ключевыми словами и применение методов устранения неоднозначности слов для точного определения контекста интересов пользователя.

Предложено решение, предлагающее автоматически извлекать концепты и отношения из текста с помощью алгоритмов обработки и анализа текстов на естественном языке.

Для реализации конвейера разработки онтологии из текста в данной работе применяются алгоритмы, созданные автором ранее. В их число входят:

1. **Метод FBKE** [29, 31, 33] (Frequency and Bert-based Keyword Extraction) для извлечения ключевых фраз, использующий векторы встраивания на основе контекста для представления ключевых фраз, что позволяет вычислить значения сходства фразы с контекстом документа. FBKE включает два этапа: на первом выбираются кандидаты n-грамм на основе их частоты встречаемости в тексте документа; на втором этапе кандидаты, после присвоения им веса, ранжируются в соответствии с их близостью к контексту документа.

2. **Алгоритм Lesk-S-BERT** [34, 35] для устранения неоднозначности смыслов извлеченных слов (униграмм). ***Lesk-S-BERT*** использует векторы S-BERT вместо встраивания слов и дополняет аннотацию синсетов из WordNet примерами. Использование расширенной аннотации синсетов повышает точность определения правильного значения слова, поскольку эти примеры отражают контекстуальное использование слова, что является ключевым для понимания его значения в конкретном контексте. Использование S-BERT как при извлечении ключевых фраз, так и при устранении неоднозначности слов снижает вычислительные затраты алгоритма. Также, это гарантирует, что процесс устранения неоднозначности слов основан на том же контекстуальном понимании, которое использовалось для извлечения ключевых фраз, что приводит к более точной интерпретации интересов пользователя. Разработанные модифицированные метод и алгоритм интегрированы в создание и расширение пользовательского профиля с целью улучшения процесса персонализации веб-контента.

3. **Алгоритм извлечения семантических отношений** (его описание представлено в разделе, посвященном разработке и вычислительному эксперименту).

1.4. Обновление профиля. Профили пользователей не являются статическими и должны постоянно обновляться, чтобы отражать изменения в предпочтениях и поведении пользователя. Этот этап включает мониторинг взаимодействия пользователя, сбор новых данных и соответствующее обновление профиля. Для этого, используются те же методы и техники, которые применялись на этапе сбора данных и построения профиля пользователя. Это включает в себя анализ и обработку описаний объектов, извлечение ключевых фраз, которые характеризуют эти объекты, и устранения их неоднозначности.

1.5. Персонализация. Последний этап – использование профиля пользователя для персонализации веб-контента или рекомендаций. На этом этапе используется комбинация техник для генерации рекомендаций, адаптированных к потребностям пользователя на основе полученной из анализа профиля информации. Эти техники включают методы оценки близости или семантической схожести, которые оценивают актуальность новых объектов в отношении пользователя. Кроме того, применяются алгоритмы машинного обучения для разработки автоматизированных моделей персонализации, способных выявлять скрытые паттерны как из данных пользователя, так и из данных других пользователей. Анализируя эти паттерны, система может предоставлять рекомендации, которые соответствуют предпочтениям и требованиям пользователя.

В следующем разделе описывается методика построения онтологии из текста (онтологического профиля) с использованием набора инструментов обработки и анализа текстов на естественном языке.

2. Реализация методики построения онтологии из текста. Для построения онтологии применяются разработанные в предыдущих работах автора алгоритмы обработки и анализа текстов на естественном языке, в т.ч. модифицированный метод извлечения ключевых фраз, алгоритм устранения неоднозначности смысла слов и алгоритм построения профиля пользователя. Полный процесс реализации алгоритма построения онтологии описан ниже. На рис. 2 и 3 представлен пользовательский интерфейс для построения и оценки онтологий, который позволяет выбрать подходящие методы и функции для выполнения каждого из описанных этапов.

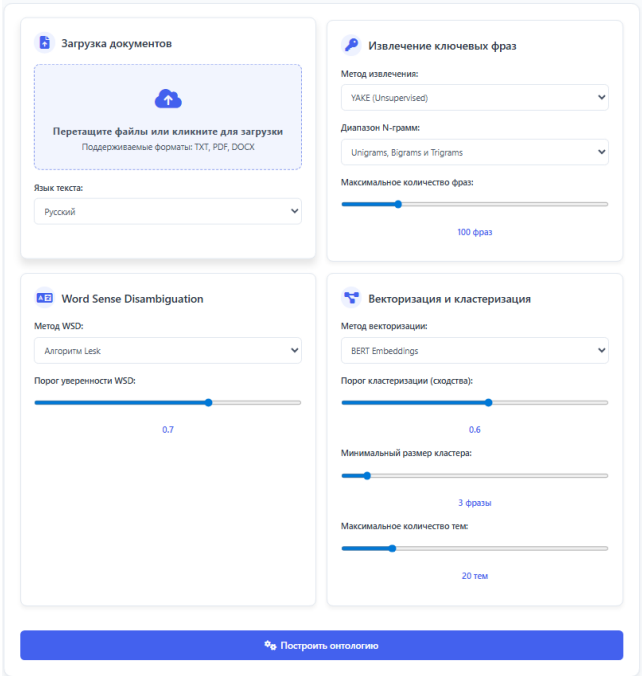

Рис. 2. Интерфейс приложения для построения и оценки онтологии

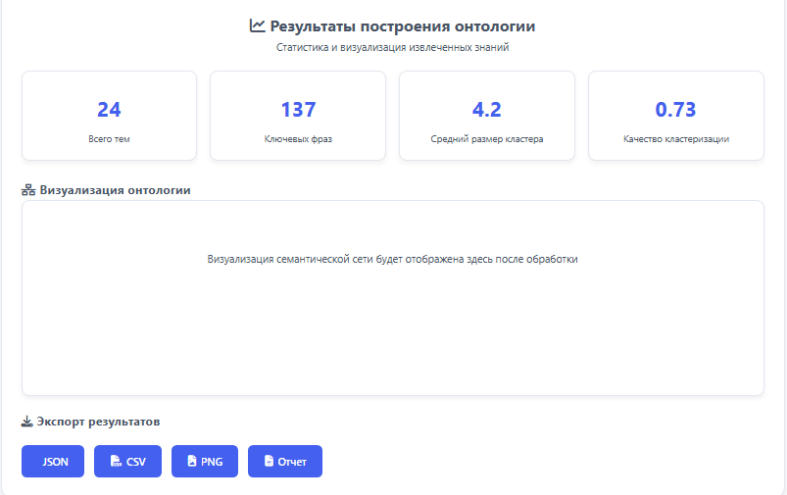

Рис. 3. Интерфейс приложения для построения и оценки онтологии

Приложение было разработано с использованием HTML, JavaScript и CSS для интерфейса, в то время как серверная часть была реализована на Python с использованием библиотеки Streamlit.

Процесс начинается с этапа предварительной обработки текстовых документов, которые представляют собой результаты поиска или любой другой текст, введенный пользователем для последующей обработки и преобразования в онтологию. Для извлечения ключевых фраз можно выбрать любой метод, например, Yake, Rake, TF-IDF, keyBERT или FBKE. В текущем эксперименте для извлечения ключевых фраз используется разработанный автором метод FBKE, который использует S-BERT для векторного представления предложений и возвращает для каждого документа список ключевых фраз вместе с их векторами. Данный этап может быть расширен за счет проведения семантического анализа с использованием библиотек типа SpaCy или других. Чем качественнее извлеченные ключевые фразы, тем лучше получается онтология, что и подтвердится при сравнении KeyBERT с нашим методом.

Интерфейс позволяет выбрать подходящий алгоритм устранения неоднозначности смысла слов с указанием порога уверенности алгоритма. В данной реализации был выбран модифицированный алгоритм LESK-BERT, предложенный автором в [30], для определения подходящих синсетов (наборов синонимов) для слов-униграмм и корректировки их векторных представлений в соответствии с их значениями для приближения к контексту, в котором они используются.

Для формирования тем и подтем применяется агломеративная кластеризация со следующими параметрами:

- Метрика расстояния: косинусное расстояние.
- Метод связи: *average linkage*, где расстояние между двумя кластерами вычисляется как среднее арифметическое расстояний между всеми парами объектов в этих кластерах.

При этом интерфейс позволяет управлять следующими параметрами:

- Максимальное количество тем.
- Пороговое расстояние (порог кластеризация): distance_threshold = 0.4.
- Минимальный размер кластера: min_cluster_size = 3.

Для создания профиля пользователя применяется иерархический алгоритм кластеризации векторов ключевых фраз для построения пользовательского профиля. Это позволяет выявить тематические интересы пользователя, где каждый кластер соответствует определенной теме. Алгоритм осуществляет многоуровневое разделение фраз на кластеры и подкластеры на основе вычисления семантической близости между векторными представлениями.

Критически важным параметром является порог сходства, определяющий объединение фраз в единую тему. На рис. 4 представлен механизм создания тем интересов с использованием алгоритма иерархической кластеризации. Экспериментальным путем установлено, что оптимальное значение данного порога составляет 60-70%.

Рис. 4. Создание тем интересов с использованием иерархической кластеризации ключевых фраз

Как показано на диаграмме, при пороге сходства 68% данные разделяются на 5 тематических кластеров интересов, каждый из которых содержит соответствующие ключевые фразы. Внутри каждого кластера наблюдаются подгруппы, которые согласно алгоритму выявления связей могут формировать узкоспециализированные тематические подкатегории (подтемы), а также группы семантически эквивалентных терминов (синонимичные ряды).

Построение отношений. Отношения устанавливаются на основе семантического сходства с пороговой фильтрацией в соответствии с алгоритмом извлечения семантических отношений, описанным ниже.

Алгоритм извлечения семантических отношений

Вход*: KWS //Список ключевых фраз*
Выход*: **Relation** // Список извлекаемых отношений*

```
for i in range(0, length(KWS)) do:
    for j in range(i + 1, length(KWS) − 1) do:
        kw1, kw2 = KWS[i], KWS[j]
        s = cos_sim(kw1, kw2)
        if s > 0.9: Relation(kw1, kw2) = synonym// синонимичность
        if s < 0.6: Relation(kw1, kw2) = ∅// отсутствие связи
        if 0.6 < s < 0.9:
            Relation(kw1, kw2) = search_wikiData(kw1, kw2)
            if Relation(kw1, kw2) == ∅:
                    Relation(kw1, kw2) = related
    end
end
Return Relation
```

Для ключевых фраз с оценкой близости в диапазоне от 0.6-0.7 до 0.9 используется Wikidata для проверки определенных отношений (например, синонимия, часть-целое, подкласс) между парами фраз. Пары, не имеющие определенных отношений в Wikidata, маркируются обобщенной меткой "has-relation", обозначающей наличие тематической связи подчиненного характера.

Таким образом, данный алгоритм извлекает следующие типы отношений:

- синонимия (сходства >90% или 60-90% при подтверждении через Wikidata);
- отношения: часть-целое (part of); более широкое понятие (Broader); более узкое понятие (Narrower); subClassOf верифицируются через Wikidata при сходстве 60-90%;
- общая (нетипизированная) связь (с доверительным уровнем 60-90%);
- отсутствие связи ∅.

Например, косинусное сходство между векторами S-BERT для фразы «машинного обучения[1]» и «искусственного интеллекта» составляет 0.72. Это указывает на то, что требуется дополнительный шаг для определения конкретного типа отношения. Для уточнения взаимосвязи используется WikiData, которая подтверждает, что «машинное обучение» является подклассом «искусственного интеллекта». В данном эксперименты рассматриваются только два типа отношений: синонимы и общее отношение (related), указывающее на наличие семантической связи без использования WikiData.

relation_type = {
синонимы (synonymous) если $sim(v_i, v_j) \geq \theta_{syn}$,
общая связь (related) если $\theta_{rel} \leq sim(v_i, v_j) < \theta_{syn}$}, (1)

где порог синонимии: $\theta_{syn} = 0.9$, порог связанности: $\theta_{rel} = 0.6$, и минимальная сила отношения: min_strength = 0.4.

После настройки параметров и выбора подходящих методов для построения онтологии из текста можно нажать кнопку «Построить онтологию», показанную на рис. 3 интерфейса. Затем приложение выполняет построение онтологии и визуализирует её. Одновременно с этим вычисляется набор метрик, демонстрирующих степень эффективности полученной онтологии. Далее приводится подробное описание этих метрик и результатов.

Метрики оценки онтологии. Для оценки эффективности онтологии используется набор критериев, которые отражают различные аспекты качества онтологии, построенной из текста. Часть критериев относится к алгоритму кластеризации и формированию тематических групп, а другая часть – к структуре полученной сети взаимосвязей.

Иерархическая кластеризация (Agglomerative Clustering) применяется для группировки семантически близких фраз. Алгоритм работает по принципу "снизу вверх", объединяя наиболее близкие кластеры на каждом шаге. Мера расстояния между кластерами вычисляется по формуле:

$$D(C_i, C_j) = 1 - sim(\mu_i, \mu_j), \quad (2)$$

где μ_i и μ_j – центроиды кластеров.

Для оценки качества разделения кластеров используются две метрики: Силуэтный коэффициент и Индекс Калински-Харабаза [32].

Силуэтный коэффициент (Silhouette coefficient) – метрика для оценки качества кластеризации, которая показывает, насколько хорошо объекты внутри кластеров отделены от объектов других кластеров. Она измеряет, насколько похож объект на свой собственный кластер (сплоченность) по сравнению с другими кластерами (разделимость).

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}, \quad (3)$$

где a(i) – среднее расстояние внутри кластера, b(i) – расстояние до ближайшего соседнего кластера.

Индекс Калински-Харабаза (Calinski-Harabasz Index) – метрика оценки качества кластеризации, основанная на соотношении межкластерной дисперсии (разделимость) и внутрикластерной дисперсии (компактность).

[1] machine learning - Wikidata.

$$\text{CH} = \frac{\text{Межкластерная дисперсия}/(\text{k}-1)}{\text{Внутрикластерная дисперсия}/(n-k)}, \quad (4)$$

где k – число кластеров, n – число объектов.

Кроме того, используются следующие структурные метрики:

- Средний размер кластера, который показывает степень детализации онтологии:

$$ACS = \frac{1}{n}\sum_{i=1}^{n}|C_i|; \quad (5)$$

- Плотность отношений измеряет связанность онтологии:

$$RD = \frac{2\cdot|E|}{|V|\cdot(|V|-1)}, \quad (6)$$

Семантическая когерентность:

Внутрикластерное сходство оценивает семантическую однородность:

$$\text{IntraSim}(C_i) = \frac{2}{|C_i|\cdot(|C_i|-1)}\sum_{x,y\in C_i} sim(x,y); \quad (7)$$

Межкластерное сходство измеряет дискриминативность кластеров:

$$\text{InterSim} = \frac{1}{k(k-1)}\sum_{i\neq j} sim(\mu_i,\mu_j), \quad (8)$$

Коэффициент когерентности показывает отношение внутрикластерной к межкластерной близости:

$$\text{Coherence\_Ratio} = \text{IntraSim}/\text{InterSim}, \quad (9)$$

Онтология представляется как взвешенный граф G(V, E, w), где: V = {v_1, v_2, ..., v_n} – множество концептов (вершин); E ⊆ V × V – семантические отношения (ребра); w: E → [0,1] – функция весов, определяющая силу отношений

Коэффициент кластеризации измеряет локальную плотность:

$$C_i = \frac{2\cdot e_i}{k_i\cdot(k_i-1)}, \quad (10)$$

где e_i – число связей между соседями вершины i, k_i – степень вершины.

Связные компоненты показывают структурную целостность онтологии. Компонента связности – это максимальное подмножество вершин, где между любой парой вершин существует путь.

Временная сложность алгоритма построения онтологии:

- Извлечение ключевых фраз: $O(n\cdot m)$ где n – количество документов, m – средняя длина документа.
- Кластеризация: $O(n^2\log(n))$ для иерархической кластеризации.
- Построение отношений: $O(k^2)$ где k – количество кластеров.

3. Результаты вычислительного эксперимента по оценке онтологии. На вход алгоритма подается коллекция текстов на естественном языке. Выходом приложения является построенная онтология и расчетные метрики её оценки. Система демонстрирует стабильные результаты при различных параметрах:

- Силуэтный коэффициент: 0.3-0.6.
- Средний размер кластера: 4-8 фраз.
- Плотность отношений: 3-8 связей на концепт.
- Внутрикластерное сходство: ≥ 0.7.

В результате проведенного эксперимента была сформирована оптимальная кластерная структура, состоящая из 18 семантических кластеров, которые объединили 70 ключевых фраз. Средний размер кластера, составивший 3.89 элемента, свидетельствует о достижении баланса между излишней дробностью и чрезмерным обобщением, так называемой «зоны Златовласки». Данная структура демонстрирует репрезентативную группировку концептов, полноценное покрытие предметной области и смысловую целостность каждого кластера.

Таблица 1

Показатели качества построения отношений

Количество документов	02
Среднее количество слов в документе	20.5
Количество извлеченных фраз	082
Число кластеров	18
Число фраз	70
Средний размер кластера	3,89
Среднее внутрикластерное сходство	84,1
Оценка силуэта	0,444
Количество связей	404
Средняя сила связи	0,803
Плотность связей	5,77

Важным результатом является высокая семантическая когерентность модели. Показатель внутрикластерного сходства, достигший значения 0.841, указывает на исключительное качество кластеризации. Анализ выявил наличие сильных семантических связей внутри кластеров, их четкую тематическую согласованность и формирование осмысленных семантических концептов в каждой группе.

Качество семантических отношений в построенной сети также оценивается как высокое. Средняя сила установленных связей находится на уровне 0.803, что подчеркивает их значимость и информационную ценность. Сеть характеризуется низким уровнем шума и подтверждает эффективность примененного принципа «качество над количеством».

Оценка метрических показателей подтверждает устойчивость модели. Значение силуэтного коэффициента (0.444) указывает на удовлетворительное разделение кластеров с минимальным перекрытием и четкими границами, что оставляет потенциал для дальнейшей оптимизации до уровня 0.5+ (рис. 5).

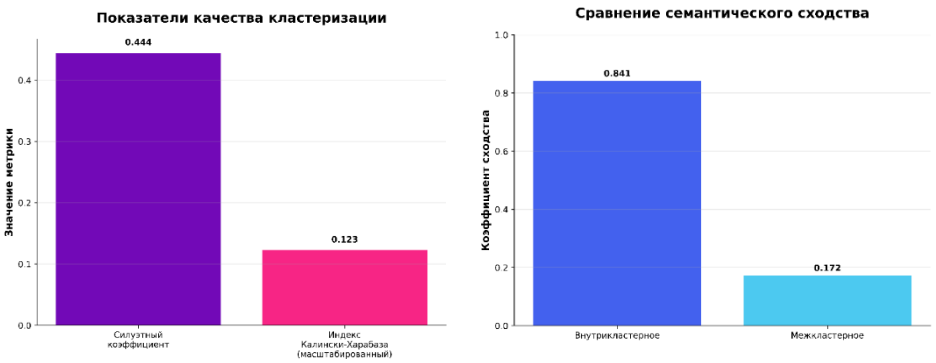

Рис. 5. Показатели качества кластеризации

Показатель плотности отношений (5.77) отражает оптимальную связность сети и сбалансированность ее структуры, обеспечивающую эффективные паттерны информационных потоков. Само собой разумеется, что количество общих связей (related) превышает количество связей-синонимов (synonyms), поскольку данная категория включает различные типы отношений, которые могут быть точно определены с использованием внешних источников знаний, таких как Wikidata. Однако в рамках данного эксперимента было решено ограничиться общей классификацией, поскольку основной целью исследования является оценка качества онтологии с точки зрения её структуры и согласованности, а также для снижения вычислительной сложности эксперимента (рис. 6).

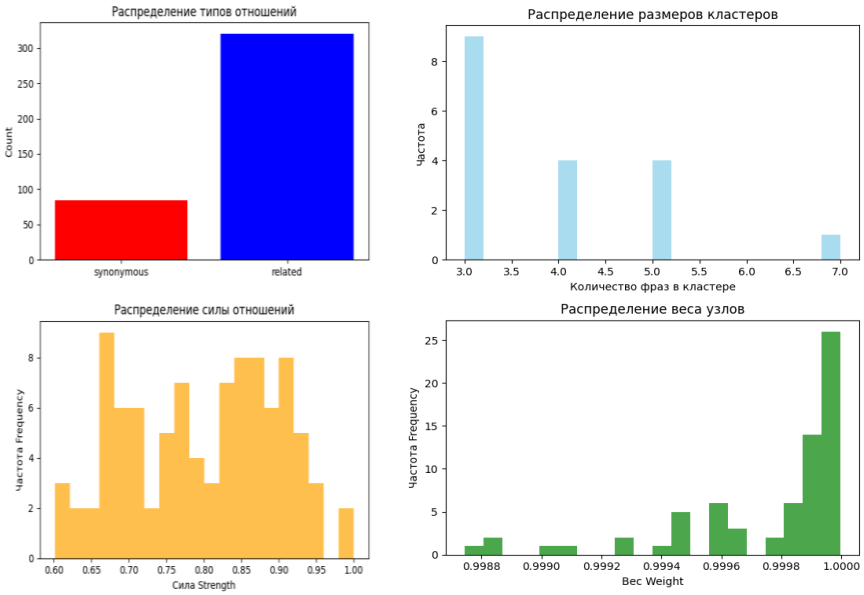

Рис. 6. Показатели качества построения отношений

Коэффициент ретенции, составивший 37% (сохранение 70 из 188 исходных фраз), свидетельствует о высокой точности фильтрации при сохранении полноты покрытия предметной области.

Практическая значимость работы заключается в высоком потенциале применения полученной онтологической модели в таких областях, как рекомендательные системы, семантический поиск, управление знаниями и категоризация контента. Модель обеспечивает пользователям четкую тематическую структуру, осмысленные семантические связи и надежную организацию знаний.

Выявленные направления для оптимизации, такие, как повышение силуэтного коэффициента и уточнение границ кластеров, незначительны и не оказывают отрицательного влияния на результаты. В целом, работа подтверждает практическую применимость разработанной методики в задачах обработки естественного языка и управления знаниями.

Система эффективно балансирует между точностью и полнотой, обеспечивая создание качественных семантических структур для практических приложений. Многоуровневый подход позволяет выявлять как структурные проблемы, так и семантические, что делает систему ценным инструментом для разработки и валидации онтологий в промышленных приложениях.

Заключение. В данной работе предложено комплексное решение автоматического построения онтологического профиля для решения проблем веб-персонализации. Ключевым вкладом является разработка целостного конвейера обработки данных (pipeline), который интегрирует усовершенствованные алгоритмы автора (такие как метод извлечения ключевых фраз FBKE и алгоритм разрешения лексической неоднозначности Lesk-BERT) с признанными инструментами обработки естественного языка. Проведенный вычислительный эксперимент подтвердил эффективность метода и продемонстрировал способность системы создавать непротиворечивые и релевантные онтологические структуры. Практическая значимость исследования заключается в потенциале применения онтологии в рекомендательных системах, семантическом поиске и управлении знаниями. Модель обеспечивает четкую тематическую структуру с осмысленными семантическими связями, способствуя снижению когнитивной нагрузки пользователя. Перспективы дальнейших исследований включают интеграцию онтологии в задачи персонализации веб-контента, а также проведение сравнительных экспериментов для оценки производительности алгоритмов против современных аналогов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Garrigós I., Gomez J., Houben G.-J.* Specification of personalization in web application design // Information Software Technology. – 2010. – Vol. 52, No. 9. – P. 991-1010.
2. *Мертёхин А.А.* Интернет-зависимое поведение и перегрузка информацией // Северо-Кавказский психологический вестник. – 2012. – Т. 10, № 3. – P. 24-27.
3. *Meister F., Shin D., Andrews L.* "Getting to know you": What's new in personalization technologies // E-Doc. – 2002. – Vol. 16, No. 2. – P. 8-8.
4. *Pressman R.S., Lowe D.* Web engineering // Software Engineering: A Practitioner's Perspective. – 2000. – P. 769-798.
5. *Ginige A., Murugesan S.* Web Engineering: A Holistic, Disciplined Approach to Web-Based System Development // 12 th International World Wide Web Conference. – 2003. – Vol. 3. – Web Engineering.
6. *Tao X., Li Y., Zhong N.* A personalized ontology model for web information gathering // IEEE transactions on knowledge data engineering. – 2010. – Vol. 23, No. 4. – P. 496-511.
7. *Guo Q., Chen W., Wan H.* AOL4PS: A large-scale data set for personalized search // Data Intelligence. – 2021. – Vol. 3. – AOL4PS. – No. 4. – P. 548-567.
8. *Farid M., Elgohary R., Moawad I., Roushdy M.* User profiling approaches, modeling, and personalization // Proceedings of the 11th international conference on informatics & systems (INFOS 2018). – 2018.
9. *Mobasher B.* Data mining for web personalization // The adaptive web. – Springer, 2007. – P. 90-135.
10. *Gauch S., Speretta M., Chandramouli A., Micarelli A.* User profiles for personalized information access // The adaptive web. – 2007. – P. 54-89.
11. *Cantador I., Bellogín A., Castells P.* Ontology-based personalised and context-aware recommendations of news items // 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology. – IEEE, 2008. – Vol. 1. – P. 562-565.
12. *Leung K.W.-T., Lee D.L.* Deriving concept-based user profiles from search engine logs // IEEE Transactions on knowledge and data engineering. – 2009. – Vol. 22, No. 7. – P. 969-982.
13. *Liu F., Yu C., Meng W.* Personalized web search by mapping user queries to categories // Proceedings of the eleventh international conference on Information and knowledge management CIKM02: Eleventh ACM International Conference on Information and Knowledge Management. – McLean Virginia USA: ACM, 2002. – P. 558-565.
14. *Penas P., Del Hoyo R., Vea-Murguía J., González C., Mayo S.* Collective knowledge ontology user profiling for Twitter–automatic user profiling // 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). – IEEE, 2013. – Vol. 1. – P. 439-444.
15. *Gauch S., Chaffee J., Pretschner A.* Ontology-based personalized search and browsing // Web Intelligence and Agent Systems: An international Journal. – 2003. – Vol. 1, No. 3-4. – P. 219-234.
16. *Xu Y., Wang K., Zhang B., Chen Z.* Privacy-enhancing personalized web search // Proceedings of the 16th international conference on World Wide Web WWW'07: 16th International World Wide Web Conference. – Banff Alberta Canada: ACM, 2007. – P. 591-600.
17. *Abián D., Guerra F., Martínez-Romanos J., Trillo-Lado R.* Wikidata and DBpedia: A Comparative Study // Semantic Keyword-Based Search on Structured Data Sources: Lecture Notes in Computer Science / eds. J. Szymański, Y. Velegrakis. – Cham: Springer International Publishing, 2018. – Vol. 10546. – Wikidata and DBpedia. – P. 142-154. – ISBN 978-3-319-74496-4.
18. *Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P.N., Hellmann S.,. Morsey M, Van Kleef P., Auer S.* Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia // Semantic web. – 2015. – Vol. 6, No. 2. – P. 167-195.
19. *Eke C.I., Norman A.A., Shuib L., Nweke H.F.* A survey of user profiling: State-of-the-art, challenges, and solutions // IEEE Access. – 2019. – Vol. 7. – P. 144907-144924.
20. *Purificato E., Boratto L., De Luca User E.W.* Modeling and User Profiling: A Comprehensive Survey // arXiv preprint arXiv:2402.09660. – 2024.
21. *Bird S.* NLTK: the natural language toolkit // Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. – 2006. – P. 69-72.
22. *Miller G.A.* WordNet: a lexical database for English // C. A. – 1995. – Vol. 38. – P. 39-41.
23. *Vasiliev Y.* Natural language processing with Python and spaCy: A practical introduction. – No Starch Press, 2020. – ISBN 1-71850-052-1.
24. *Lops P., De Gemmis M., Semeraro G.* Content-based recommender systems: State of the art and trends // Recommender systems handbook. – 2011. – P. 73-105.
25. *Poelmans J., Ignatov D.I., Kuznetsov S.O., Dedene G.* Formal concept analysis in knowledge processing: A survey on applications // Expert Systems with Applications. – 2013. – Vol. 40, No. 16. – P. 6538-6560.
26. *Poelmans J., Ignatov D.I., Viaene S., Dedene G., Kuznetsov S.O.* Text mining scientific papers: a survey on FCA-based information retrieval research // Advances in Data Mining. Applications and Theoretical Aspects: 12th Industrial Conference, ICDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 12. – Springer, 2012. – P. 273-287.

27. *Маннинг К.Д., Рагхаван П., Шютце Х.* Введение в информационный поиск. – 2011.
28. *Gruber T.* What is an Ontology. – 1993.
29. *Мохаммад Ж.Х. Мансур А.М., Кравченко Ю.А., Бова В.В.* Метод извлечения ключевых фраз на основе новой функции ранжирования // Информационные технологии. – 2022. – Т. 9, № 28. – P. 465-474.
30. *Кравченко Ю.А., Мансур А.М., Хуссайн М.Ж.* Модифицированный метод устранения неоднозначности смысла слов, основанный на методах распределенного представления // Известия ЮФУ. Технические науки. – 2021. – № 3.
31. *Мохаммад Ж.Х., Мансур А.М., Кравченко Ю.А., Кравченко Д.Ю.* Метод автоматического извлечения ключевых слов // Международный научно-технический конгресс «Интеллектуальные системы и информационные технологии – 2022». – 2022. – P. 91-97.
32. *Wang X., Xu Y.* An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index // IOP Conference Series: Materials Science and Engineering. – IOP Publishing, 2019. – Vol. 569. – P. 052024.
33. *Мохаммад Ж.Х., Мансур А.М., Кравченко Ю.А*. Перспективы применения метода извлечения ключевых фраз FBKE в задачах персонализации веб-контента // XX Всероссийская научная конференция молодых ученых, аспирантов и студентов «Информационные технологии, системный анализ и управление (ИТСАУ-2022). – Таганрог, 2022. – P. 206.
34. *Мохаммад Ж.Х., Мансур А.М., Кравченко Ю.А.* Модифицированный метод устранения неоднозначности смысла слов, основанный на методах распределенного представления // Известия ЮФУ. Технические науки. – 2021. – № 3 (220). – P. 92-101.
35. *Мохаммад Ж.Х., Мансур А.* Модифицированный метод устранения семантической неоднозначности слов. – Таганрог, 2022.

REFERENCES

1. *Garrigós I., Gomez J., Houben G.-J.* Specification of personalization in web application design, *Information Software Technology,* 2010, Vol. 52, No. 9, pp. 991-1010.
2. *Mertekhin A.A.* Internet-zavisimoe povedenie i peregruzka informatsiey [Internet-dependent behavior and information overload], *Severo-Kavkazskiy psikhologicheskiy vestnik* [North Caucasian Psychological Bulletin], 2012, Vol. 10, No. 3, pp. 24-27.
3. *Meister F., Shin D., Andrews L.* "Getting to know you": What's new in personalization technologies, *E-Doc,* 2002, Vol. 16, No. 2, pp. 8-8.
4. *Pressman R.S., Lowe D.* Web engineering, *Software Engineering: A Practitioner's Perspective,* 2000, pp. 769-798.
5. *Ginige A., Murugesan S.* Web Engineering: A Holistic, Disciplined Approach to Web-Based System Development, *12 th International World Wide Web Conference,* 2003, Vol. 3. Web Engineering.
6. *Tao X., Li Y., Zhong N.* A personalized ontology model for web information gathering, *IEEE transactions on knowledge data engineering,* 2010, Vol. 23, No. 4, pp. 496-511.
7. *Guo Q., Chen W., Wan H.* AOL4PS: A large-scale data set for personalized search, *Data Intelligence,* 2021, Vol. 3, AOL4PS, No. 4, pp. 548-567.
8. *Farid M., Elgohary R., Moawad I., Roushdy M.* User profiling approaches, modeling, and personalization, *Proceedings of the 11th international conference on informatics & systems (INFOS 2018),* 2018.
9. *Mobasher B.* Data mining for web personalization, *The adaptive web*. Springer, 2007, pp. 90-135.
10. *Gauch S., Speretta M., Chandramouli A., Micarelli A.* User profiles for personalized information access, *The adaptive web,* 2007, pp. 54-89.
11. *Cantador I., Bellogín A., Castells P.* Ontology-based personalised and context-aware recommendations of news items, *2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology*. IEEE, 2008, Vol. 1, pp. 562-565.
12. *Leung K.W.-T., Lee D.L.* Deriving concept-based user profiles from search engine logs, *IEEE Transactions on knowledge and data engineering,* 2009, Vol. 22, No. 7, pp. 969-982.
13. *Liu F., Yu C., Meng W*. Personalized web search by mapping user queries to categories, *Proceedings of the eleventh international conference on Information and knowledge management CIKM02: Eleventh ACM International Conference on Information and Knowledge Management*. McLean Virginia USA: ACM, 2002, pp. 558-565.
14. *Penas P., Del Hoyo R., Vea-Murguía J., González C., Mayo S.* Collective knowledge ontology user profiling for Twitter–automatic user profiling, *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE, 2013, Vol. 1, pp. 439-444.
15. *Gauch S., Chaffee J., Pretschner A.* Ontology-based personalized search and browsing, *Web Intelligence and Agent Systems: An international Journal,* 2003, Vol. 1, No. 3-4, pp. 219-234.
16. *Xu Y., Wang K., Zhang B., Chen Z.* Privacy-enhancing personalized web search, *Proceedings of the 16th international conference on World Wide Web WWW'07: 16th International World Wide Web Conference*. Banff Alberta Canada: ACM, 2007, pp. 591-600.

17. *Abián D., Guerra F., Martínez-Romanos J., Trillo-Lado R.* Wikidata and DBpedia: A Comparative Study, *Semantic Keyword-Based Search on Structured Data Sources: Lecture Notes in Computer Science,* eds. J. Szymański, Y. Velegrakis. Cham: Springer International Publishing, 2018, Vol. 10546. Wikidata and DBpedia, pp. 142-154. ISBN 978-3-319-74496-4.
18. *Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P.N., Hellmann S.,. Morsey M, Van Kleef P., Auer S.* Dbpedia–a large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic web,* 2015, Vol. 6, No. 2, pp. 167-195.
19. *Eke C.I., Norman A.A., Shuib L., Nweke H.F.* A survey of user profiling: State-of-the-art, challenges, and solutions, *IEEE Access,* 2019, Vol. 7, pp. 144907-144924.
20. *Purificato E., Boratto L., De Luca User E.W.* Modeling and User Profiling: A Comprehensive Survey, *arXiv preprint arXiv:2402.09660,* 2024.
21. *Bird S.* NLTK: the natural language toolkit, *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions,* 2006, pp. 69-72.
22. *Miller G.A.* WordNet: a lexical database for English, *C. A.*, 1995, Vol. 38, pp. 39-41.
23. *Vasiliev Y.* Natural language processing with Python and spaCy: A practical introduction. No Starch Press, 2020. ISBN 1-71850-052-1.
24. *Lops P., De Gemmis M., Semeraro G.* Content-based recommender systems: State of the art and trends, *Recommender systems handbook,* 2011, pp. 73-105.
25. *Poelmans J., Ignatov D.I., Kuznetsov S.O., Dedene G.* Formal concept analysis in knowledge processing: A survey on applications, *Expert Systems with Applications,* 2013, Vol. 40, No. 16, pp. 6538-6560.
26. *Poelmans J., Ignatov D.I., Viaene S., Dedene G., Kuznetsov S.O.* Text mining scientific papers: a survey on FCA-based information retrieval research, *Advances in Data Mining. Applications and Theoretical Aspects: 12th Industrial Conference, ICDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 12*. Springer, 2012, pp. 273-287.
27. *Manning K.D., Ragkhavan P., Shyuttse Kh.* Vvedenie v informatsionnyy poisk [Introduction to information retrieva], 2011.
28. *Gruber T.* What is an Ontology, 1993.
29. *Mokhammad Zh.Kh. Mansur A.M., Kravchenko Yu.A., Bova V.V.* Metod izvlecheniya klyuchevykh fraz na osnove novoy funktsii ranzhirovaniya [Method for extracting key phrases based on a new ranking function], *Informatsionnye tekhnologii* [Information Technologies], 2022, Vol. 9, No. 28, pp. 465-474.
30. *Kravchenko Yu.A., Mansur A.M., Khussayn M.Zh.* Modifitsirovannyy metod ustraneniya neodnoznachnosti smysla slov, osnovannyy na metodakh raspredelennogo predstavleniya [Modified method for disambiguating the meaning of words based on distributed representation methods], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3.
31. *Mokhammad Zh.Kh., Mansur A.M., Kravchenko Yu.A., Kravchenko D.Yu.* Metod avtomaticheskogo izvlecheniya klyuchevykh slov [Method of automatic extraction of keywords], *Mezhdunarodnyy nauchno-tekhnicheskiy kongress «Intellektual'nye sistemy i informatsionnye tekhnologii – 2022»* [International Scientific and Technical Congress "Intelligent Systems and Information Technologies - 2022"], 2022, pp. 91-97.
32. *Wang X., Xu Y.* An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index, *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2019, Vol. 569, pp. 052024.
33. *Mokhammad Zh.Kh., Mansur A.M., Kravchenko Yu.A.* Perspektivy primeneniya metoda izvlecheniya klyuchevykh fraz FBKE v zadachakh personalizatsii veb-kontenta [Prospects for applying the FBKE keyword extraction method in web content personalization tasks], *XX Vserossiyskaya nauchnaya konferentsiya molodykh uchenykh, aspirantov i studentov «Informatsionnye tekhnologii, sistemnyy analiz i upravlenie (ITSAU-2022)* [XX All-Russian scientific conference of young scientists, graduate students and students "Information technology, systems analysis and management (ITSAU-2022)], Taganrog, 2022, pp. 206.
34. *Mokhammad Zh.Kh., Mansur A.M., Kravchenko Yu.A.* Modifitsirovannyy metod ustraneniya neodnoznachnosti smysla slov, osnovannyy na metodakh raspredelennogo predstavleniya [A modified method for disambiguating word meanings based on distributed representation methods], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3 (220), pp. 92-101.
35. *Mokhammad Zh.Kh., Mansur A.* Modifitsirovannyy metod ustraneniya semanticheskoy neodnoznachnosti slov [A modified method for disambiguating semantic words]. Taganrog, 2022.

Мохаммад Жуман Хуссайн – Южный федеральный университет; e-mail: zmohammad@sfedu.com; mahammad.hs.juman@gmail.com; г. Таганрог, Россия; тел.: +79880158697; кафедра систем автоматизированного проектирования им. В.М. Курейчика; соискатель.

Mohammad Juman Hussain – Southern Federal University; e-mail: zmohammad@sfedu.ru, mahammad.hs.juman@gmail.com; Taganrog, Russia; phone: +79880158697; the Department of Computer Aided Design named after V.M. Kureichik; applicant.