

*На правах рукописи*



Мохаммад Жуман Хуссайд

**МОДЕЛЬ, МЕТОД И АЛГОРИТМЫ ВЕБ-ИНЖЕНЕРИИ ДЛЯ  
ИНТЕЛЛЕКТУАЛЬНОЙ ПЕРСОНАЛИЗАЦИИ КОНТЕНТА ВЕБ-  
САЙТОВ**

Специальность – 1.2.1. Искусственный интеллект и машинное обучение

**Автореферат**  
диссертации на соискание ученой степени  
кандидата технических наук

Таганрог – 2026

Работа выполнена в ФГАОУ ВО «Южный федеральный университет» на кафедре систем автоматизированного проектирования Института компьютерных технологий и информационной безопасности

Научный руководитель:

доктор технических наук, доцент

**Кравченко Юрий Алексеевич,**

ФГАОУ ВО «Южный федеральный университет»,  
профессор кафедры систем автоматизированного  
проектирования (г. Таганрог).

Официальные оппоненты:

доктор технических наук, доцент

**Гусев Павел Юрьевич**

ФГБОУ ВО "Воронежский государственный  
технический университет", заведующий кафедрой  
искусственного интеллекта и цифровых  
технологий (г. Воронеж).

кандидат технических наук

**Нагоев Залимхан Вячеславович**

ФГБНУ «Федерального научного центра  
«Кабардино-Балкарского научного центра  
Российской академии наук» генеральный  
директор (г. Нальчик).

Ведущая организация:

ФГБОУ ВО «Волгоградский государственный  
технический университет» (ВолГТУ,  
г. Волгоград).

Защита диссертации состоится «17» декабря 2026 г. В 10<sup>00</sup> на заседании объединенного совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук 99.2.107.02 на базе ФГАОУ ВО «Южный федеральный университет», ФГБОУ ВО «Южно-Российский государственный политехнический университет (НПИ) имени М. И. Платова» по адресу: 347922, г. Таганрог, пер. Некрасовский, 44, ауд. Г-439.

С диссертацией можно ознакомиться в зональной научной библиотеке ФГАОУ ВО «Южный федеральный университет» по адресу: 344015, г. Ростов-на-Дону, ул. Зорге, 21Ж, а также на портале электронных ресурсов ЮФУ: <https://hub.sfedu.ru/diss/show/1357268/>.

Отзыв на автореферат, заверенный гербовой печатью организации, просим направлять ученому секретарю объединенного диссертационного совета 99.2.107.02 по адресу: 347922, г. Таганрог, пер. Некрасовский, 44.

Автореферат разослан «\_\_» \_\_\_\_\_ 2026 г.

Ученый секретарь объединенного  
диссертационного совета 99.2.107.02,  
доктор технических наук, доцент

Ю. А. Кравченко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования и степень её разработанности.** В условиях стремительного развития технологий искусственного интеллекта (ИИ), экспоненциального роста объёмов данных и распространения цифровых сервисов (социальные сети, Интернет вещей, большие данные, облачные вычисления) персонализация представления веб-контента 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. Эти факторы приводят к существенному снижению эффективности 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. Персонализация является одним из ключевых направлений веб-инженерии и представляет собой совокупность информационных процессов и методов, направленных на формирование индивидуализированного пользовательского взаимодействия на основе анализа предпочтений, поведения и других релевантных данных 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, поведения и демографических характеристик. 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. 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. 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 исследованиях применяются внешние классификационные источники. Они позволяют соотносить термины с более общими концептами и представлять пользовательские интересы в виде взвешенных тематических категорий. 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». При этом ключевым структурным элементом подобных онтологий является синсет (synset) – набор когнитивных синонимов, объединённых общим смысловым значением и представляющих единый концепт в семантической сети (например, в WordNet). Синсеты обеспечивают точное разграничение значений многозначных слов и служат основой для устранения семантической неоднозначности. Однако такие методы часто опираются на статические, предопределённые онтологии, которые не способны адаптироваться к динамичному характеру предпочтений 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. Это связано с неспособностью этих методов совместить семантическую точность, адаптивность и масштабируемость при построении профилей пользователей.

В данном исследовании предложено комплексное решение задачи персонализации контента веб-сайтов на основе применения методов искусственного интеллекта, в том числе, алгоритмов автоматизации построения онтологий. Эти алгоритмы выполняют три ключевые функции: во-первых, осуществляют извлечение ключевых фраз и понятий; во-вторых, автоматически устраняют возникающую семантическую неоднозначность; и, в-третьих, интегрируются в единый конвейер обработки текстовых данных. Основу решения составляют методы анализа контекстуальной семантической близости, которые принципиально отличаются от статических или синтаксических тем, что учитывают конкретный контекст употребления языковых единиц, а не только их абстрактное значение или формальные связи.

Следовательно, развитие моделей, методов и алгоритмов для обработки и анализа текстов на естественном языке и извлечения знаний из них, позволяющих повысить эффективность систем персонализации веб-контента в условиях семантической неоднозначности, является актуальной научной задачей.

Ряд работ российских и зарубежных учёных посвящен разработке методов персонализации контента веб-сайтов с использованием методов

интеллектуального анализа текста, поиска информации и обработки естественных языков. Источниками исследования диссертации послужили следующие труды известных ученых по основам построения профиля пользователя и персонализации контента веб-сайтов: М. Гонтек, Б. Мобашер, А. Хавала, С. Гауч, Д. Уэстон и Д. Юрафски. Работы Х. Шютце и К. Д. Мэннинга, К. С. Джонса, М. А. Хёрста, В.М. Курейчика, А. Панченко, И. Д. Игнатов и А. Кутузова вносят значительный вклад в разработку методов ИИ, таких как построение онтологий и извлечение знаний из текста.

**Целью диссертационной работы** является повышение эффективности систем персонализации контента веб-сайтов. Под эффективностью понимается минимизация процента ошибок алгоритмов функционирования систем персонализации.

Для достижения поставленной цели были решены следующие основные **задачи**:

1. Проведён аналитический обзор современных подходов к персонализации контента веб-сайтов на основе методов обработки и анализа текстов.

2. Разработан модифицированный метод извлечения ключевых фраз из отдельных документов, основанный на новой весовой функции, которая учитывает частоту встречаемости фраз и их сходство с документом.

3. Разработан модифицированный алгоритм устранения неоднозначности смысла слов, основанный на методах контекстуального представления и расширенных аннотациях синсетов.

4. Построена онтологическая модель, предназначенная для автоматизации процесса выявления связей между понятиями, с целью отражения индивидуальных особенностей интересов пользователя в его профиле.

5. Разработан алгоритм персонализации интересов пользователя на основе онтологической модели.

6. Разработано программное приложение для проведения вычислительного эксперимента и подтверждение достоверности и эффективности полученных основных результатов.

**Объект исследования** – тексты на естественном языке.

**Предмет исследования** – методы и алгоритмы обработки и анализа текстов для решения задачи персонализации контента веб-сайтов.

**Методология и методы диссертационного исследования.** При выполнении диссертационной работы использовались методы интеллектуального анализа данных, методы обработки и анализа текстов на естественном языке, теория информационных систем, методы искусственного интеллекта и машинного обучения, а также методы объектно-ориентированного программирования.

**Степень достоверности результатов.** Достоверность научных результатов работы подтверждается непротиворечивостью и согласованностью с известными фактами и исследованиями в рассматриваемой области, высокой степенью сходимости теоретических

результатов с данными экспериментов, и определяется применением теоретических и методологических основ разработок ведущих ученых в области создания интеллектуальных систем, корректным и обоснованным использованием математического аппарата, экспериментальными исследованиями разработанных модели, метода и алгоритмов.

### **Научная новизна.**

Научной новизной проведенного исследования являются модель, метод и алгоритмы обработки и анализа текстов на естественном языке, направленные на решение научной задачи повышения точности персонализации веб-контента на основе онтологического представления индивидуальных особенностей интересов пользователя, что имеет важное значение для развития информатики, а именно:

1. Разработан модифицированный метод извлечения ключевых фраз из одного документа, **отличающийся** применением новых правил для определения весов ключевых фраз на основе частоты их встречаемости и контекстуальной семантической близости, **позволяющий** извлекать наиболее релевантные ключевые фразы (пункт 5 паспорта специальности 1.2.1, страницы 52–63 диссертации);

2. Разработан модифицированный алгоритм устранения неоднозначности смысла слов, **отличающийся** применением контекстных векторных представлений и расширенных аннотаций синсетов, **что позволяет** снизить процент ошибки при определении правильного значения неоднозначного слова (пункт 5 паспорта специальности 1.2.1, страницы 63–74 диссертации);

3. Построена онтологическая модель для представления индивидуальных особенностей интересов пользователя, **отличающаяся** применением новой функции расчета весовых коэффициентов, а также контекстно-зависимых отношений, **что позволяет** найти более релевантные темы (пункт 5 паспорта специальности 1.2.1, страницы 75–83 диссертации);

4. Разработан алгоритм персонализированного поиска веб-сайтов, **отличающийся** применением весовых коэффициентов в формуле ранжирования документов, учитывающей иерархические уровни онтологического профиля пользователя, **позволяющий** снизить процент ошибки при решении задачи персонализации веб-контента (пункт 4 паспорта специальности 1.2.1, страницы 83–95 диссертации).

Новые результаты, полученные в ходе выполнения диссертационного исследования, соответствуют пунктам 4, 5 паспорта научной специальности 1.2.1. Искусственный интеллект и машинное обучение.

### **Положения, выносимые на защиту:**

1. Модифицированный метод извлечения ключевых фраз из одного документа на основе частоты их встречаемости и контекстуальной семантической близости позволяет извлекать наиболее релевантные ключевые фразы;

2. Модифицированный алгоритм устранения неоднозначности смысла слов на основе контекстных векторных представлений и расширенных

аннотаций синсетов позволяет снизить процент ошибки при определении правильного значения неоднозначного слова;

3. Онтологическая модель для представления индивидуальных особенностей интересов пользователя позволяет найти более релевантные темы интересов;

4. Алгоритм персонализированного поиска веб-сайтов позволяет снизить процент ошибки при решении задачи персонализации веб-контента.

**Теоретическая значимость работы** заключается в разработке модели, метода и алгоритмов обработки естественно-языковых текстов, которые обеспечивают: эффективную обработку и анализ текстов на естественном языке; извлечение и фильтрацию ключевых фраз, а также устранение их неоднозначности; исследование и разработку средств персонализации контента веб-сайтов. Полученные научные результаты существенно развивают методологический аппарат в области искусственного интеллекта и машинного обучения, предлагая новые решения важной научной задачи увеличения объема релевантной текстовой информации в условиях семантической неоднозначности.

**Практическая значимость** заключается в разработке и реализации моделей и алгоритмов в виде программного приложения, позволяющего использовать разработанные модель, метод и алгоритмы обработки и анализа текстов на естественном языке для минимизации процента ошибки при решении задачи персонализации веб-контента.

**Реализация и внедрение результатов работы.** Теоретические и практические результаты работы внедрены в информационные процессы ИТ-компании ООО «Ит-Эффект». Полученные в работе научные результаты позволили повысить эффективность решения задачи персонализации контента веб-сайтов в условиях семантической неоднозначности, включающей разработку комплекса алгоритмов извлечения ключевых фраз, устранения их неоднозначности и извлечения семантических отношений между ними. Результаты работы также используются в учебном процессе института компьютерных технологий и информационной безопасности Южного федерального университета.

**Апробация результатов диссертации.** Основные положения и отдельные результаты исследования докладывались и обсуждались на следующих конференциях: XVIII и XX Всероссийская конференция молодых ученых аспирантов и студентов «Информационные технологии, системный анализ и управление ИТСАУ» (Россия, Таганрог, 2019–2022); VI Всероссийская научно-техническая конференция «Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности» (Россия, Таганрог, 2020); II Научно-методическая конференция НПР «Современные компьютерные технологии» (Россия, Таганрог, 2021 и 2022 гг.); XII Международная научно-техническая конференция «Технологии разработки информационных систем (ТРИС–2022)» (Россия, Таганрог, 2022); International Russian Automation Conference RusAutoCon, (Россия, Сочи, 2023); (Россия, Москва, 2022); «5<sup>th</sup> International Scientific Convention UCIENCIA»

(Куба, Гавана, 2023); «6<sup>th</sup> International Scientific Convention UCIENCIA» (Куба, Гавана, 2025).

**Публикации.** По теме диссертации опубликовано **17** научных работ, из которых: **4** статьи опубликованы в издании из перечня рекомендованных ВАК (К2), в т.ч. **2** статьи опубликованы без соавторов; **3** статьи – в изданиях из международных баз данных Scopus и/или Web of Science. Получены **2** свидетельства о государственной регистрации программ для ЭВМ. В трудах всероссийских и международных конгрессов и конференций опубликовано **10** работ.

**Личное участие соискателя в получении результатов.** Все выносимые на защиту результаты и положения, составляющие основное содержание диссертационной работы, разработаны и получены лично автором или при его непосредственном участии. В работах, опубликованных в соавторстве, соискателю принадлежит определяющая роль в развитии аппарата ИИ, а также моделей и методов обработки и анализа текстов на естественном языке.

**Структура и объем работы.** Диссертация состоит из введения, **4** разделов, заключения, списка литературы, содержащего **139** наименований, и **2** приложений. Основная часть работы содержит **142** страниц, включая **24** рисунка и **11** таблиц.

**Во введении** сформулирована цель работы, обоснована актуальность темы диссертации, описаны основные научные положения, выносимые на защиту, научная новизна, теоретическая и практическая ценность, апробация диссертационной работы, реализация и внедрение, а также структура диссертации.

**Первая глава** диссертационной работы посвящена анализу исследуемой предметной области. Проведен аналитический обзор научных исследований в области обработки и анализа текстов на основе методов искусственного интеллекта и машинного обучения при решении задачи персонализации контента веб-сайтов. Также рассмотрены существующие методы персонализации контента веб-сайтов. Основное внимание уделяется методам и алгоритмам извлечения ключевых фраз и устранения их неоднозначности.

Проведённый анализ исследований показал, что методы что методы обработки и анализа текста, применяемые в задачах персонализации контента веб-сайтов, не способны сочетать семантическую точность, неоднозначность и масштабируемость при построении профилей пользователей, что приводит к неэффективному представлению интересов пользователей в системах персонализации веб-сайтов. Конкретно, методы, основанные на ключевых словах, подчеркивают простоту, но им не хватает смысловой глубины, в то время как методы, основанные на онтологиях, предлагают более богатую семантику за счет вычислительной эффективности и адаптируемости.

Таким образом, для достижения цели диссертационного исследования, направленной на повышение эффективности моделей, методов и алгоритмов персонализации контента веб-сайтов, применено комплексное решение. Это решение используется в задаче персонализации на основе применения

методов искусственного интеллекта, в том числе, алгоритмов автоматизации построения онтологий.

Постановка задачи сформулирована следующим образом:

Пусть:  $U = \{u_1, u_2, \dots, u_{|U|}\}$  – множество пользователей,  $I = \{i_1, i_2, \dots, i_{|I|}\}$  – множество текстовых элементов, с которыми взаимодействовал пользователь  $J = \{j_1, j_2, \dots, j_{|J|}\}$  – множество новых текстовых элементов, предлагаемых для рекомендации.

Требуется построить модель, позволяющую сформировать для каждого пользователя упорядоченный список элементов, максимально соответствующих его интересам. Модель должна обеспечить повышение эффективности поиска элементов знаний (концептов и ключевых фраз) и упрощении процесса оценки семантической близости между ними. В итоге, применение онтологической модели должно привести к снижению процента ошибки функционирования систем персонализации ( $E \rightarrow \min$ ).

**Во второй главе** описана разработка модифицированного метода извлечения ключевых фраз и модифицированного алгоритма устранения неоднозначности слов. Метод извлечения ключевых отличается применением новых правил для определения весов ключевых фраз на основе частоты их встречаемости и контекстуальной семантической близости.

Как показано на рисунке 1 метод выполняет предварительную обработку текста в несколько этапов: сначала проводится очистка текста, затем токенизация, перевод всех слов в нижний регистр и удаление стоп-слов. Текст сегментируется на предложения и подблоки по знакам препинания, чтобы сохранить структурную целостность. Затем формируются  $n$ -граммы, с использованием окна переменной длины.  $N$ -граммы, не принадлежащие одному блоку, удаляются, затем ключевые фразы фильтруются по шаблонам частей речи с использованием синтаксического анализатора, удаляя лишние слова, такие как глаголы и артикли. Фразы-кандидаты выбираются по частоте встречаемости, рассчитываемой отдельно для каждого типа  $n$ -грамм. Частота вычисляется по формуле:

$$TF_{n-gram} = \frac{n_t^i}{\sum_k n_k^i}, \quad (1)$$

где  $n_k$  относится к общему количеству  $n$ -грамм определенного типа (с элементами  $i$ ), а  $n_t$  представляет количество вхождений термина  $t$  в документ.

Важность фразы в документе измеряется с помощью эвристической функции, которая вычисляет сходство между документом и всеми его фразами-кандидатами. Для этого и кандидат, и документ представлены в виде векторов S-BERT, которые позволяют измерить расстояние между данными векторами с помощью косинусной меры сходства:

$$s = \text{similarity}(U, V) = \cos(\theta) = \frac{U \cdot V}{\|U\| \cdot \|V\|} = \frac{\sum_{i=1}^n U_i V_i}{\sqrt{\sum_{i=1}^n U_i^2} \sqrt{\sum_{i=1}^n V_i^2}}, \quad (2)$$

где  $U_i$  и  $V_i$  – компоненты вектора  $X$  (вектор запроса) и  $Y$  (вектор документа), соответственно. Чтобы учесть влияние длины фразы, сходство нормализуется по формуле:

$$S_{norm}^j = S^j \cdot e^{S^j/|c^j|}, \quad (3)$$

где  $S_j$  – сходство.  $|c^j|$  – длина фразы.

Окончательная релевантность рассчитывается как среднее гармоническое сходство  $S_{norm}^{j,d}$  и частоты  $tf_{c_i}$ :

$$rel(d, c^j) = (1 + \beta^2) \cdot \frac{tf_{c^j} \cdot S_{norm}^{j,d}}{\beta^2 \cdot tf_{c^j} + S_{norm}^{j,d}}, \quad (4)$$

Параметр  $\beta$  регулирует приоритет частоты (при  $\beta < 1$ ) или сходства (при  $\beta > 1$ ). При  $\beta = 1$  оба фактора равнозначны. Кандидаты ранжируются по релевантности, и выбираются лучшие.

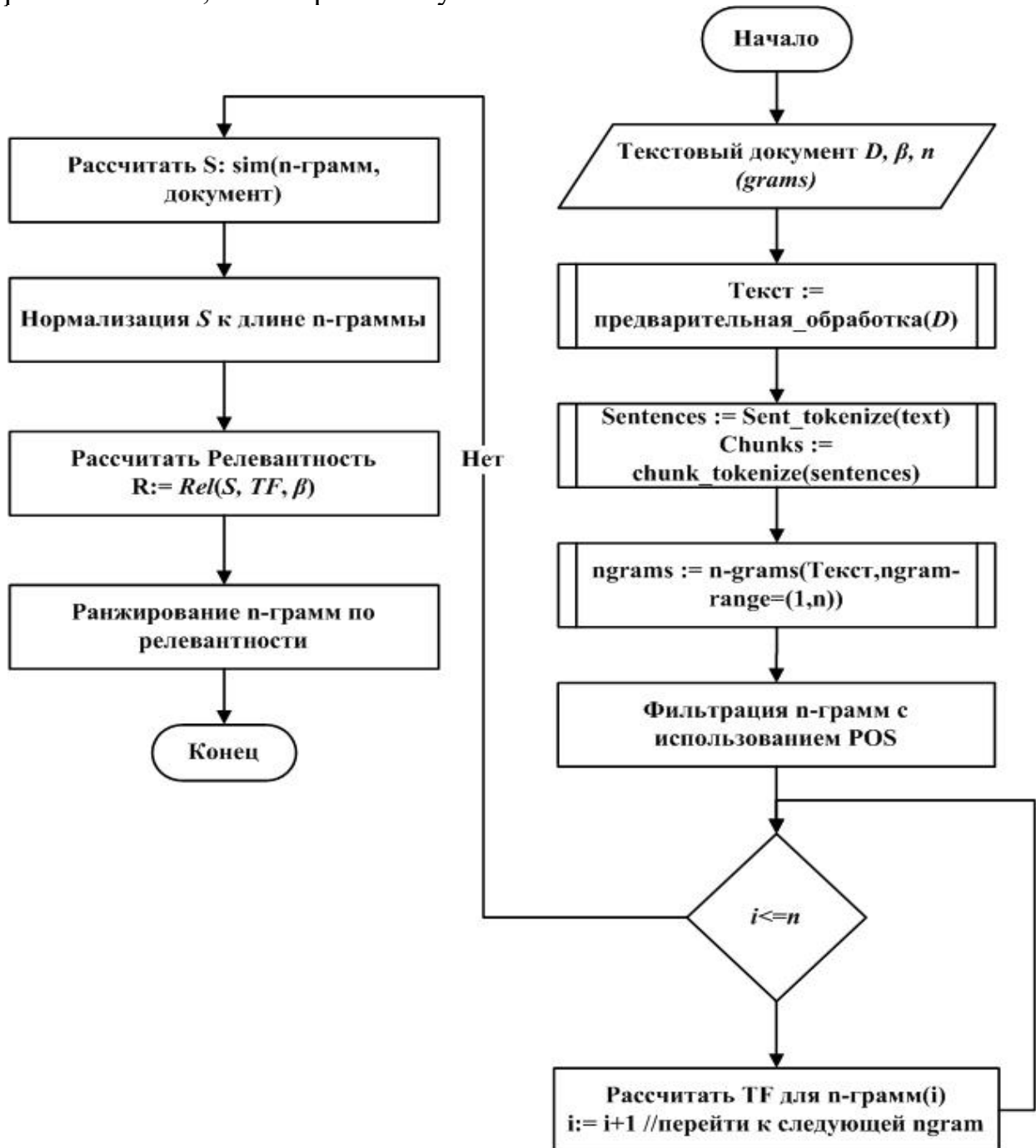


Рисунок 1 – Модифицированный метод извлечения ключевых фраз

Модифицированный алгоритм устранения неоднозначности смысла слов модифицирует алгоритм Леска (с встраиванием). Он расширяет аннотацию синсетов в WordNet, включив в нее не только определения, но и примеры использования. Это обусловлено тем, что примеры лучше отражают контекстуальное употребление слова. Алгоритм также отличается применением контекстуальной семантической близости для определения соответствующего значения целевого неоднозначного слова.

---

Алгоритм устранения неоднозначности слов (*Леска-S-BERT*)

---

**Вход**     - Целевое слово  $w$   
               - Текст документа  $D$   
               - Предварительно извлеченные ключевые фразы  $KP$   
               - Лексическая база WordNet

**Выход**  $S_{best}$  Конкретное значение слова  $w$  в данном контексте

---

*//определить контекст  $K$  слова.*

```

1:   if  $w$  является частью  $n$ -граммы в  $KP$ :
2:   |   вектора контекста  $K^B = \text{существующий вектор из } KP$ 
   else:
3:   |   контекст  $K = \text{извлечь\_окружающие\_слова}(w, D, win = 5)$ 
4:   |   вектора контекста  $K^B = S\text{-Bert}(K)$ 

```

---

*// обработка синсетов*

```

5:    $S = \text{WordNet.get\_synsets}(w)$  //Список возможных синсетов слова
    $S = (S_0, S_1, \dots, S_N)$ 
6:   for each synset in  $S$  do:
7:   |    $O = \text{synset.definitions}$  //извлекается определение каждого
   |    $O = (o_1, o_2, \dots, o_N)$  синсета
8:   |    $\Pi = \text{synset.examples}$  // $E = (e_1, e_2, \dots, e_N)$  //извлекается набор
   |   примеров каждого синсета
9:   |   if not  $E$ :  $E = \text{get\_related\_examples}()$ 
   |    $T_{\text{аннотации}} = \text{concatenate}(E, O)$  //объединять тексты

```

*End*

```

10:   $N = \text{len}(S)$  //количество синсетов
11:   $\text{Similarity\_scores} = []$ 
12:  for  $i = 0$  to  $N$  do:
13:  |    $t_i^B = S\text{-Bert}(t_i)$  // Представление синсета  $S[i]$  в виде вектора
   |    $\text{Sim}_{S\text{-Bert}} = \cos(\vec{t_i^B}, \vec{K^B})$ 
14:  |    $\text{Similarity\_scores}[i] \leftarrow \text{sim}$ 
15:  |   end
16:   $S_{best} = \text{max}(\text{Sort}(S, \text{Similarity\_scores}))$  //Ранжирование синсетов

```

Сначала определяется контекст ключевого слова – окружающие его слова. Для контекста  $C$  и целевого неоднозначного слова  $w$  извлекается набор синсетов  $S = (S_0, S_1, \dots, S_N)$  с помощью WordNet. Затем для каждого синсета извлекается определение  $O = (o_1, o_2, \dots, o_N)$  и набор примеров  $E = (e_1, e_2, \dots, e_N)$ . Если в WordNet нет примеров использования текущего синсета, функция *get\_related\_examples* «поиск релевантных примеров» ищет ближайшие примеры в связанных синсетах (синонимах, гипонимах или гиперонимах). На рисунке 2 показан процесс поиска примеров для расширения аннотации синсета в WordNet, у которого отсутствуют примеры употребления.

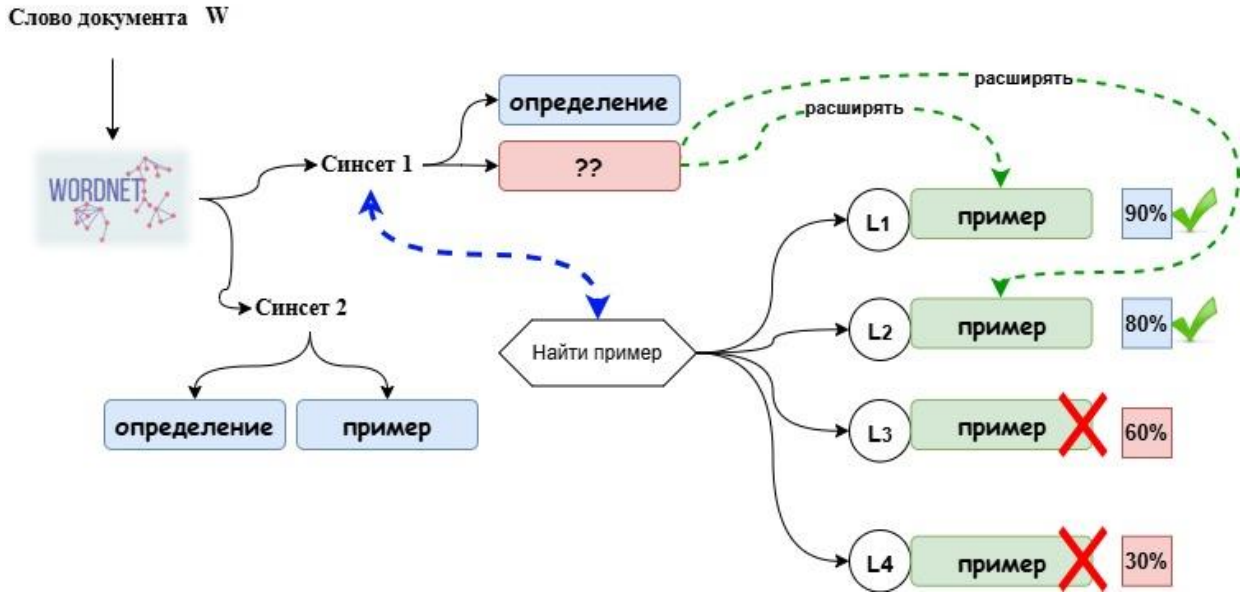


Рисунок 2 – Схема расширения аннотации синсета

Определение и примеры объединяются в единый текст, называемый аннотацией синсета:

$$T_{\text{аннотации}} = \text{concatenate}(E, O) = (t_1, t_2, \dots, t_N). \quad (5)$$

Текст контекста и аннотации синсетов преобразуются в числовые векторы S-BERT. Сходство между контекстом  $K$  и аннотацией  $t_i^{SBERT}$  оценивается через косинусное сходство, а итоговый показатель сходства рассчитывается как среднее значение.

$$\text{sim} = f(\text{Sim}_{\text{GloVe}}, \text{Sim}_{\text{Bert}}). \quad (6)$$

Синсеты ранжируются по показателям сходства, и выбирается синсет с наивысшим значением.

**Третья глава** посвящена разработке онтологической модели для представления индивидуальных особенностей интересов пользователя и алгоритма персонализированного поиска веб-сайтов. Онтологическая модель отличается применением новой функции расчета весовых коэффициентов, а также контекстно-зависимых отношений.

Рассмотрим формальное описание предложенной онтологической модели  $O$ , которое определяется как кортеж:

$$O = \langle I, C, R, W, X, D, A \rangle, \quad (7)$$

где  $I$  – множество экземпляров, представляющих конкретные сущности или ключевые фразы;  $C$  – множество концептов (классов), представляющих общие категории или темы интересов;  $R$  – множество семантических отношений (является «is-a», часть-целое «part-of» и др.), импортированных из эталонных онтологий, определяющих связи между концептами и/или экземплярами;  $R$  включает как иерархические, так и неиерархические отношения. Предлагаемая онтология характеризуется дополнительными элементами, включающими:

- $D$  – множество динамически контекстно-зависимых извлекаемых отношения, не зафиксированные в  $R$ , полученных из текста;
- Весовые функции:

$W: C \cup I \rightarrow [0,1]$ , присваивает весовые коэффициенты концептам ( $C$ ) и экземплярам ( $I$ ), количественно отражающие уровень интереса пользователя к соответствующим темам. Данные веса рассчитываются на основе анализа поведения пользователя и его взаимодействия с текстовыми данными;

$\chi: D \rightarrow [0,1]$  присваивает весовые коэффициенты, определяющие степень достоверности извлекаемых отношений.

- $A$  – множество аксиом (логических ограничений), обеспечивающих правила для определения и фильтрации связей.

Далее, приводится подробное объяснение онтологии и того, как она реализована.

Динамически извлекаемые отношения между ключевыми фразами, полученные из текста с использованием *алгоритма извлечения семантических отношений*. Особенность данного алгоритма заключается в автоматическом формировании этих связей из текста с помощью методов оценки семантической близости, применяемых к векторным представлениям фраз и концептов.

---

#### Алгоритм извлечения семантических отношений

---

**Вход:**  $KWS$  // Список ключевых фраз

**Выход:**  $Relation$  // Список извлекаемых отношений

---

```

1:  for  $i$  in range(0, length( $KWS$ )) do:
2:      for  $j$  in range( $i + 1$ , length( $KWS$ ) – 1) do:
3:           $kw1, kw2 = KWS[i], KWS[j]$ 
4:           $s = \text{cos\_sim}(kw1, kw2)$ 
5:          if  $s > 0.9$ :  $Relation(kw1, kw2) = \text{synonym}$  // синонимичность
6:          if  $s < 0.6$ :  $Relation(kw1, kw2) = \emptyset$  // отсутствие связи
7:          if  $0.6 < s < 0.9$ :
8:               $Relation(kw1, kw2) = \text{search\_wikiData}(kw1, kw2)$ 
9:              if  $Relation(kw1, kw2) == \emptyset$ :
10:                   $Relation(kw1, kw2) = \text{has-relation}$ 
11:      end
12:  end
13:  Return  $Relation$ 

```

Косинусная мера оценивает семантическую близость, но для определения синонимии требуется пороговое значение  $\emptyset$ , зависящее от задачи. Для фраз при использовании S-BERT порог обычно составляет  $[0,7-0,9]$ . Сходство выше 0,9 часто указывают на почти идеальные синонимы. Для ключевых фраз с оценкой сходства в диапазоне от 0.7 до 0.9 используется Wikidata для проверки определенных отношений (например, синонимия, часть-целое, подкласс) между парами фраз. Пары, не имеющие определенных отношений Wikidata, получают общее обозначение «has-relation».

Общий процесс алгоритма персонализированного поиска веб-сайтов включает в себя создание и обновление профиля пользователя, а также сортировку документов в соответствии с этим профилем.

Для создания профиля пользователя  $U$  на основе посещенных веб-страниц  $K$  выполняется ряд шагов, которые включают в себя:

1. Извлечение текстового содержимого веб-страниц с сохранением заголовков и подзаголовков для определения важности ключевых фраз.

2. Применение модифицированного метода  $FBKE$  для извлечения семантически релевантных фраз. Вес фразы  $W(k)$  рассчитывается по формуле:

$$W(k) = f_k \cdot p_k, \quad (8)$$

где  $f_k$  – вес фразы, рассчитанный по формуле (4),  $p_k$  – позиционный вес (выше для заголовков, ниже для основного текста).

$$p_k = \begin{cases} \alpha & \text{если } k \text{ в заголовке} \\ \frac{\alpha}{i} & \text{если } k \text{ в подзаголовке } H_i \\ 1 & \text{если } k \text{ в основном тексте} \end{cases} \quad (9)$$

где  $\alpha$  – настраиваемая константа (обычно  $\alpha > 1$ ).

3. Векторы отдельных слов (униграмм) обобщены, поэтому для уточнения их значений применяется разработанный модифицированный алгоритм устранения неоднозначности смысла слов Lesk-SBERT.

4. Создание тем интереса пользователя. Ключевые фразы группируются в темы интересов методом иерархической кластеризации их векторных представлений. Алгоритм разделяет фразы на кластеры и подкластеры по семантическому сходству, используя пороговые значения для выявления взаимосвязей. После кластеризации каждой теме присваивается вес, равный среднему весу её фраз, что отражает общий интерес пользователя к данной теме. Вес *темы интереса* определяется следующей формулой:

$$w_{t_i} = \frac{1}{|P_{t_i}|} \sum_{p \in P_{t_i}} w_p, \quad (10)$$

где  $P_{t_i}$  – количество фраз в кластере  $t_i$ ; Итоговый вес ключевой фразы  $w_p$  на уровне темы интересов  $t_i$  рассчитывается как среднее значение её весов во всех соответствующих документах (веб-страницах):

$$w_p = \frac{1}{|D_p|} \sum_{d \in D_p} w_{p,d}, \quad (11)$$

где  $p$  – ключевая фраза;  $D_p$  – множество документов, в которых встречается фраза  $p$ ;  $w_{p,d}$  – вес фразы  $p$  в документе  $d$ , рассчитан по формуле (8).

Для вычислительной эффективности количество ключевых фраз в каждой теме ограничено диапазоном от 1 до 30. Фразы обновляются со временем с использованием функции временного уменьшения:

$$w_{(t)} = w_0 \cdot (1 - \lambda t) \quad (12)$$

Результатом данного процесса является семантически насыщенный онтологический профиль пользователя.

*Процедура ранжирования документов при персонализации интересов пользователя.* Для заданного набора текстовых документов  $D$ , требующих персонализированного ранжирования на основе профиля пользователя, где документы могут представлять как результаты поискового запроса, так и контент веб-сайта, данная процедура осуществляет сортировку материалов в соответствии с их релевантностью к конкретному пользователю. Далее описывается механизм сопоставления документов с онтологическим профилем пользователя для генерации персонализированных результатов, соответствующих интересам, заданным в профиле. Предполагается, что исходный профиль создан в соответствии с шагами, описанными в предыдущем подпункте и структурирован в виде трехуровневой иерархии:

- уровень 1: основные темы интересов ( $T_1, T_2, \dots, T_k$ );
- уровень 2: подтемы внутри каждой темы ( $S_1, S_2, \dots, S_m$ );
- уровень 3: индивидуальные ключевые фразы ( $K_1, K_2, \dots, K_p$ ).

1. Инициализация и предварительная обработка: все полученные документы преобразуются в векторные представления с использованием того же метода векторизации тем интересов и ключевых фраз в профиле (S-BERT).

Вводятся следующие переменные:

- весовые коэффициенты:  $\alpha = 0.5, \beta = 0.3, \gamma = 0.2$ ;
- пороговые значения:  $\theta_{topic} = 0.4$  (минимальное сходство с темой);  
 $\theta_{sub-topic} = 0.6$  (минимальное сходство с подтемой).

2. Иерархическая оценка релевантности. Для каждого документа  $d \in D$ :

а) оценка на уровне тем:

$$R_{topic}(d) = \max_{t \in T} [\cos(v_d, v_t) \cdot w_t], \quad (13)$$

если  $R_{topic}(d) < \theta_{topic} \rightarrow$  пропустить документ;

б) оценка на уровне подтем (для  $t^* = \operatorname{argmax}(R_{topic}(d))$ ):

$$R_{sub}(d) = \max_{s \in S_{t^*}} [\cos(v_d, v_s) \cdot w_s], \quad (14)$$

$$\text{если } R_{sub}(d) < \theta_{sub-topic} \rightarrow R(d) = \alpha \cdot R_{topic}(d), \quad (15)$$

иначе перейти к следующему шагу;

в) детальная оценка на уровне фраз (для  $s^* = \operatorname{argmax}(R_{sub}(d))$ ):

$$R_{phrase}(d) = \frac{\sum_{k \in K_{s^*}} \cos(v_d, v_k) \cdot w_k}{|K_{s^*}|}. \quad (16)$$

Итоговый балл:

$$R(d) = \alpha \cdot R_{topic}(d) + \beta \cdot R_{sub}(d) + \gamma \cdot R_{phrase}(d). \quad (17)$$

3. Ранжирование и выдача результатов:

- сортировка документов по убыванию  $R(d)$ ;
- отсортированный список топ-N документов предоставляется пользователю в качестве персонализированных результатов поиска.

**Четвертая глава** диссертации посвящена описанию разработки программного приложения и проведения вычислительного эксперимента для сравнения эффективности разработанных модели, метода и алгоритмов персонализации контента веб-сайтов (рисунок 3).

Модифицированный метод извлечения ключевых фраз *FBKE* был оценен на пяти эталонных наборах данных из различных областей, таких как сельское хозяйство и технологии, в сравнении с рядом значимых методов, включая YAKE и RAKE.

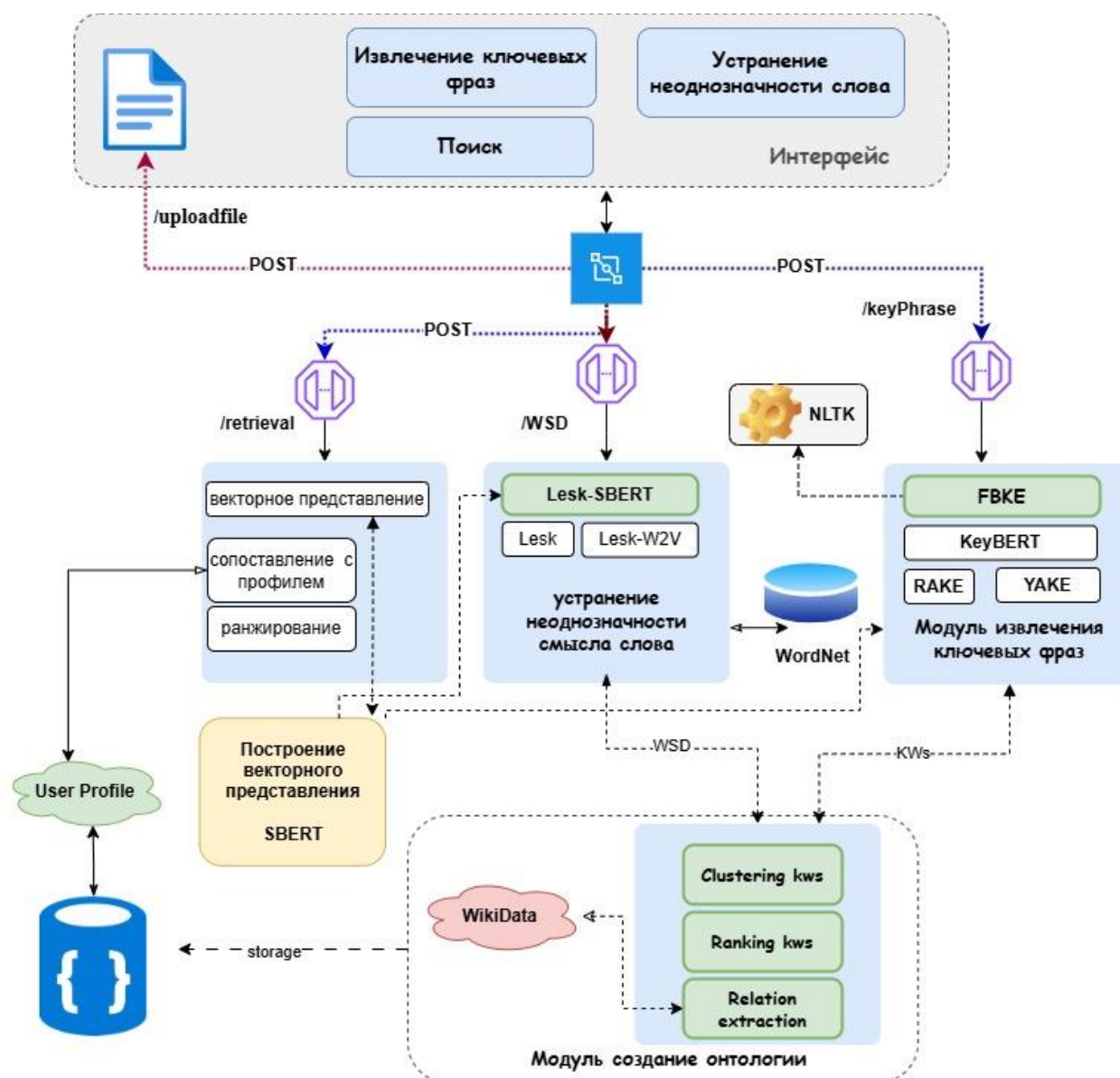


Рисунок 3 – Компонентная архитектура программного приложения системы персонализации веб-контента

Метод продемонстрировал свою эффективность и превзошел сравниваемые методы на 1–3%.

Таблица 1 – Точность ранжирования, измеренная с помощью MAP@K

|                          | @k  | wiki        | Theses100   | Fao30       | semEval2010 | Schutz2008  |
|--------------------------|-----|-------------|-------------|-------------|-------------|-------------|
| TF-Df                    | @5  | <b>20,4</b> | <b>8,4</b>  | 24,0        | 13,9        | 44,8        |
|                          | @10 | 13,8        | 7,0         | 15,4        | 9,5         | 34,0        |
| YAKE                     | @5  | 7,2         | 5,2         | 7,8         | 10,3        | 11,4        |
|                          | @10 | 7,2         | 3,9         | 4,9         | 6,9         | 7,8         |
| TEXTRANK                 | @5  | 5,6         | 4,4         | 14,9        | 3,1         | 12,4        |
|                          | @10 | 3,5         | 3,6         | 9,1         | 1,9         | 13,9        |
| KeyBERT                  | @5  | 1.5         | 6.3         | –           | –           | –           |
|                          | @10 | 0.8         | 4.9         | –           | –           | –           |
| Разработанный метод FBKE | @5  | <b>23,7</b> | <b>10,2</b> | <b>29,3</b> | <b>19,8</b> | <b>50,9</b> |
|                          | @10 | 16          | 08,7        | 20,0        | 12,4        | <b>41,6</b> |

Алгоритм устранения неоднозначности слов был протестирован на известных наборах SensEval (Se2, Se3, Se7, Se13, Se15). Он значительно превзошел оригинальный алгоритм *Леск* и показал конкурентоспособные результаты по сравнению с аналогичными алгоритмами (таб. 2).

Таблица 2 – Результаты устранения неоднозначности смысла слов, рассчитанные с использованием F-меры

| Методы                               | F1-Мера (%) |       |       |       |       |
|--------------------------------------|-------------|-------|-------|-------|-------|
|                                      | Se2         | se3   | se7   | se13  | se15  |
| WN 1st sense                         | 66.8        | 66.2  | 55.2  | 62.95 | 67.8  |
| Леск (простой)                       | 40.29       | 35.94 | 22.63 | 38.27 | 41.21 |
| Леск (расширенный)                   | 50.6        | 44.5  | 32.0  | 53.6  | 51.0  |
| Леск (с встраиванием)                | 63.0        | 63.7  | 56.7  | 66.2  | 64.6  |
| UKB                                  | 56          | 51.7  | 39.0  | 53.6  | 55.2  |
| Разработанный алгоритм (LESK-SBERT): |             |       |       |       |       |
| •С расширенными аннотациями          | 66.96       | 67.78 | 56.16 | 66.27 | 68.38 |
| •С определениями                     | 53.06       | 48.48 | 34.28 | 56.2  | 59.98 |
| •С примерами                         | 51.84       | 48.81 | 35.6  | 46.3  | 54.4  |

Онтологическая модель описывается с использованием набора показателей (рисунок 4). Полученные результаты демонстрируют высокую семантическую и структурную целостность: внутреннее сходство составляет 84%, а средняя сила связей – 80%, что отражает точность представления понятий.

Для оценки алгоритма персонализированного поиска веб-сайтов были созданы экспериментальные конвейеры для анализа индивидуального влияния каждой разработанной алгоритмической компоненты в том числе извлечение ключевых слов, устранение неоднозначности, а также извлечения семантических отношений на основе онтологии на релевантность результатов поиска (таблицы 3 и 4).

|                                   |       |
|-----------------------------------|-------|
| Число кластеров                   | 18    |
| Число фраз                        | 70    |
| Средний размер кластера           | 3,89  |
| Среднее внутрикластерное сходство | 84,1  |
| Количество связей                 | 404   |
| Средняя сила связи                | 0,803 |
| Плотность связей                  | 5,77  |
| Силуэтный коэффициент             | 0,444 |
| Индекс Калински-Харабаза          | 0,123 |
| число отношений                   |       |
| Синонимы (Синонимы)               | 89    |
| Общие                             | 315   |

**Данные:**  
Количество документов: **20**  
Среднее количество слов в документе: **62.5**

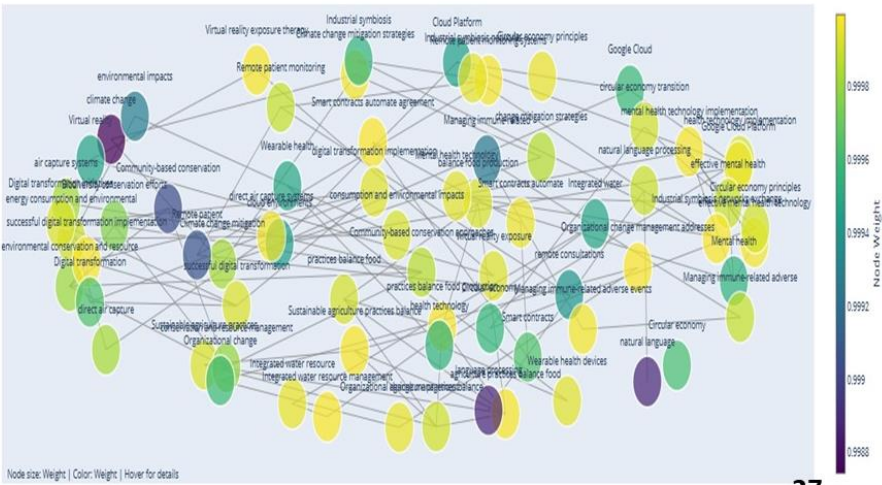


Рисунок 4 – Показатели качества построения отношений

Таблица 3 – Экспериментальные пайплайны для оценки персонализации

|            | Извлечение ключевых слов | WSD                   | Онтология | Цель                                                                                                                            |
|------------|--------------------------|-----------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------|
| Пайплайн 1 | TF-IDF                   | Леск                  | нет       | Базовый конвейер для демонстрации влияния традиционных методов извлечения ключевых слов и WSD без учета семантических отношений |
| Пайплайн 2 | FBKE                     | Леск                  | нет       | Демонстрирует улучшение при использовании FBKE вместо TF-IDF для извлечения ключевых слов при сохранении методов WSD            |
| Пайплайн 3 | TF-IDF                   | ЛЕСК <sub>SBERT</sub> | нет       | Демонстрирует улучшение при использовании ЛЕСК <sub>SBERT</sub> вместо Леск                                                     |
| Пайплайн 4 | FBKE                     | ЛЕСК <sub>SBERT</sub> | нет       | Демонстрирует комбинированный эффект FBKE и ЛЕСК <sub>SBERT</sub> на точность профиля                                           |

|            |      |                       |    |                                                                                                       |
|------------|------|-----------------------|----|-------------------------------------------------------------------------------------------------------|
| Пайплайн 5 | FBKE | ЛЕСК <sub>SBERT</sub> | да | Показывает эффект интеграции семантических отношений на основе онтологии в процесс построения профиля |
|------------|------|-----------------------|----|-------------------------------------------------------------------------------------------------------|

Таблица 4 – Экспериментальные пайплайны для оценки персонализации

|            | <i>MAP@5</i> | <i>Precision @5</i> | <i>MAP @10</i> | <i>Precision @10</i> |
|------------|--------------|---------------------|----------------|----------------------|
| Пайплайн 1 | 42.02        | 48.33               | 37.13          | 41.4                 |
| Пайплайн 2 | 52.7         | 58.6                | 47.32          | 52.8                 |
| Пайплайн 3 | 49.6         | 52.9                | 41.42          | 47.88                |
| Пайплайн 4 | 64.2         | 67.91               | 56.88          | 63.01                |
| Пайплайн 5 | <b>68.05</b> | <b>70.95</b>        | <b>59.83</b>   | <b>67.92</b>         |

В среднем предложенный автором алгоритм персонализированного поиска веб-сайтов снижает процент ошибок алгоритмов функционирования систем персонализации на **5–8%**.

**В заключении** изложены итоги выполненного исследования, даны рекомендации, а также описаны перспективы дальнейшей разработки темы.

**В приложениях** приведены свидетельства об официальной регистрации программ для ЭВМ и копии актов о внедрении результатов работы.

## ЗАКЛЮЧЕНИЕ

Диссертация посвящена решению научной задачи создания моделей, алгоритмов и методов обработки и анализа текстов на естественном языке и извлечения знаний из них, позволяющих повысить эффективность систем персонализации контента веб-сайтов. Под эффективностью понимается минимизация процента ошибок алгоритмов функционирования систем персонализации.

Разработанные в рамках данного исследования модель, алгоритмы и метод демонстрируют перспективы развития гибких механизмов обработки и анализа текстов на естественном языке в контексте разработки подходов искусственного интеллекта. Эти механизмы учитывают информационную неопределенность и семантическую неоднозначность, и открывают новые возможности для создания интеллектуальных систем, способных адаптироваться к сложным и неоднозначным данным, что особенно актуально в условиях постоянного роста объема и разнообразия информации.

Основными результатами проведенных исследований являются следующие:

1. Проведен аналитический обзор особенностей создания систем искусственного интеллекта и машинного обучения для персонализации контента веб-сайтов при обработке и анализе текстов на естественном языке. Сформулирована постановка задачи исследования;

2. Разработан модифицированный метод извлечения ключевых фраз из одного документа, отличающийся применением новых правил для определения весов ключевых фраз на основе частоты их встречаемости и контекстуальной семантической близости, позволяющий извлекать наиболее релевантные ключевые фразы;

3. Разработан модифицированный алгоритм устранения неоднозначности смысла слов, отличающийся применением контекстных векторных представлений и расширенных аннотаций синсетов, что позволяет снизить процент ошибки при определении правильного значения неоднозначного слова;

4. Построена онтологическая модель для представления индивидуальных особенностей интересов пользователя, отличающаяся применением новой функции расчета весовых коэффициентов, а также контекстно-зависимых отношений, что позволяет найти более релевантные темы;

5. Разработан алгоритм персонализированного поиска веб-сайтов, отличающийся применением весовых коэффициентов в формуле ранжирования документов, учитывающей иерархические уровни онтологического профиля пользователя, позволяющий снизить процент ошибки при решении задачи персонализации веб-контента.

6. Разработано программное приложение, реализующее разработанные модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта и машинного обучения, позволяющее проводить сравнительный анализ предложенных алгоритмов с существующими аналогами;

7. Выполнены вычислительные эксперименты, которые подтвердили эффективность полученных решений, превзошедших результаты известных алгоритмов. В среднем, предложенный автором персонализированного поиска веб-сайтов способствует уменьшению частоты ошибок при ранжировании поисковых результатов на 5–8%.

В целом совокупность полученных в диссертации теоретических и практических результатов позволяет сделать вывод о том, что цель исследований достигнута, сформулированная научная задача решена.

Дальнейшее развитие результатов диссертационного исследования целесообразно направить на интеграцию контент-ориентированных рекомендаций, основанных на разработанной онтологической модели, с методами коллаборативной фильтрации и обучения с подкреплением для повышения точности и разнообразия рекомендаций. Важным направлением является оптимизация алгоритмов для распределённых вычислительных сред с целью эффективной обработки больших данных и поддержки работы в реальном времени. Также расширение онтологических отношений при сохранении вычислительной эффективности позволит точнее моделировать интересы пользователей и повысить общую производительность системы.

## ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИОННОЙ РАБОТЫ

### Публикации в изданиях, рекомендованных ВАК РФ:

1. Мохаммад, Ж. Х. Методика построения и оценки онтологического профиля для систем персонализации контента / Ж.Х. Мохаммад // Известия ЮФУ. Технические науки. – 2025. – № 6 (248). – С. 248-262. (ВАК, К2)
2. Мохаммад, Ж. Х. Извлечение ключевых фраз на основе больших языковых моделей / Ж.Х. Мохаммад // Известия ЮФУ. Технические науки. – 2024. – № 5. – С. 196-187. (ВАК, К2)
3. Мохаммад, Ж. Х. Модифицированный метод устранения неоднозначности смысла слов, основанный на методах распределенного представления / А.М. Мансур, Ж.Х. Мохаммад, Ю.А. Кравченко // Известия ЮФУ. Технические науки. – 2021. – № 3 (220). – С. 92-101. (ВАК, К2)
4. Мохаммад, Ж. Х. Векторизация текста с использованием методов интеллектуального анализа данных / Ю.А. Кравченко, А.М. Мансур, Ж.Х. Мохаммад // Известия ЮФУ. Технические науки. – 2021. – № 2 (219). – С. 154-167. (ВАК, К2)

### Публикации в изданиях, индексируемых в базах

#### «Scopus» и «Web of Science»:

5. Mohammad, J. Automated Ontology Extraction from Text for Content-Based Web Personalization / J. Mohammad, A. Mansour, Y. Kravchenko // Lecture Notes in Computer Science. – LNCS. – Vol. 16328. – 2026. – P. 139-151. (Scopus, Q2)
6. Mohammad, J. Harnessing Key Phrases in Constructing a Concept-Based Semantic Representation of Text Using Clustering Techniques / J. Mohammad, A. Mansour, Y. Kravchenko, D. Kravchenko, N. Silega // Lecture Notes in Computer Science. – LNCS. – Vol. 14335. – 2023. – P. 190-201. (Scopus, Q2)
7. Mohammad, J. Algorithm for Optimization of Keyword Extraction Based on the Application of a Linguistic Parser / J. Mohammad, A. Mansour, D. Kravchenko, Y. Kravchenko, N. Pavlov // Informatics and Automation. – Vol. 23. – 2024. – no. 2. – P. 467-494. (Scopus, Q4, RSCI)

### Публикации в других изданиях, опубликовано 10 работ, основные из них следующие:

8. Мохаммад, Ж. Х. Метод извлечения ключевых фраз на основе новой функции ранжирования / Ж.Х. Мохаммад, А.М. Мансур, Ю.А. Кравченко, В.В. Бова // Информационные технологии. – 2022. – Том. 28. № 9. – С. 465-474. (ВАК)
9. Mohammad, J. Text vectorization method based on concept mining using clustering techniques / J. Mohammad, A. Mansour, Y. Kravchenko // 2022 VI International Conference on Information Technologies in Engineering Education (Inforino). IEEE. – 2022. – P. 1-10. (Scopus)
10. Mohammad, J. Generating Conceptual Semantic Vectors Based on Key Phrase Extraction Techniques / J. Mohammad, A. Mansour, Y. Kravchenko // 2023 International Russian Automation Conference (RusAutoCon). IEEE. – 2023. – P. 374-379. (Scopus)

11. Мохаммад Ж. Х. Онтологическая модель для персонализации контента веб-сайтов на основе методов обработки и анализа текстов / Ж.Х. Мохаммад // Труды международного научно-технического конгресса «Интеллектуальные системы и информационные технологии – 2025» («ИС & ИТ-2025», «IS&IT'22»). Научное издание. – Таганрог: Изд-во Ступина С.А., Т.1. – 2025. – С. 324-329.

12. Мохаммад, Ж. Х. Модифицированный метод построения семантического представления текста на основе методов кластеризации и взвешивания терминов / Ж.Х. Мохаммад, А.М. Мансур, Д.Ю. Кравченко, Ю.А. Кравченко // Труды XII международной научно-технической конференции «Технологии разработки информационных систем (ТРИС-2022)». – Таганрог: 2022. – С. 94-100.

13. Мохаммад Ж. Х. Модифицированный метод устранения семантической неоднозначности слов / Ж.Х. Мохаммад, А.М. Мансур // Труды XIX всероссийской научной конференции молодых ученых, аспирантов и студентов «Информационные технологии, системный анализ и управление» (ИТСАУ-2021). – 2021. – С. 204-207.

#### **Свидетельства о государственной регистрации программ для ЭВМ:**

1. Мохаммад, Ж. Программный модуль оптимизации работы классификатора при векторизации текста на основе биоэвристик / Ж. Мохаммад, А. Мансур, Ю.А. Кравченко, Д. Ю. Кравченко // Свидетельство регистрации программы для ЭВМ. – 12.12.2023. – № 2023687185.

2. Мохаммад, Ж. Программный модуль оптимизации извлечения ключевых слов при обработке лингвистической экспертной информации / Ж. Мохаммад, А. Мансур, Ю.А. Кравченко, К. Н. Владимировна // Свидетельство регистрации программы для ЭВМ. – 14.12.2023. – № 2023687372.

#### **Личный вклад автора.**

В опубликованных трудах:

[1, 4, 11] – автором разработаны и исследованы онтологическая модель и алгоритм построения онтологического профиля пользователя для систем персонализации результатов веб-поиска.

[2, 9, 10, 12] – автором разработаны методы и алгоритмы построения концептуального и семантического представления текстов, основанные на извлечении ключевых фраз, методах кластеризации и взвешивания терминов при обработке и анализе текстов на естественном языке.

[3, 5, 7, 8, 13] – автором разработаны метод и алгоритмы оптимизации извлечения ключевых слов, основанные на применении лингвистического парсера и методов интеллектуального анализа данных при обработке и анализе текстов на естественном языке;

[6, 14–17] – автором разработан модифицированный алгоритм, основанный на методах распределенного семантического представления для решения задачи устранения неоднозначности смысла слов и концептов при обработке и анализе текстов на естественном языке.

При создании указанных программ для ЭВМ, использованы теоретические результаты, метод, модель и алгоритмы, полученные лично автором.