

УДК [81'33+81'37]:004.065:621.373.8/38

Святошиц Марина Игоревна

научный сотрудник
Межотраслевой научно-практический центр
систем идентификации
и электронных деловых операций
Национальной академии наук
Республики Беларусь
младший научный сотрудник
Объединенный институт проблем информатики
Национальной академии наук
Республики Беларусь
исследователь, преподаватель кафедры
межкультурной экономической коммуникации
Белорусский государственный
экономический университет
г. Минск, Беларусь

Marina Svyatoshchik

Research Scientist
the Intersectoral Scientific
and Practical Center for Identification
Systems and Electronic Business Operations
of the National Academy
of Sciences of Belarus
Junior Research Scientist
the United Institute of Informatics Problems
of the National Academy of Sciences of Belarus
Researcher, Lecturer
at the Intercultural Economic
Communication Department
Belarus State Economic University
Minsk, Belarus
svyatoshchikm@mail.ru

Гибкий Павел Валерьевич

младший научный сотрудник
Межотраслевой научно-практический центр
систем идентификации
и электронных деловых операций
Национальной академии наук
Республики Беларусь
исследователь, преподаватель кафедры
теории и практики китайского языка
Белорусский государственный университет
иностраннных языков
г. Минск, Беларусь

Pavel Gibkij

Junior Research Scientist
the Intersectoral Scientific and Practical Center
for Identification Systems
and Electronic Business Operations
of the National Academy
of Sciences of Belarus
Researcher, Lecturer
at the Chinese Language
Theory and Practice Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
pavel.gibkiy@bk.ru

АВТОМАТИЗАЦИЯ ФОРМИРОВАНИЯ И СТРУКТУРИРОВАНИЯ КОЛЛЕКЦИИ ТЕКСТОВ КАК ОСНОВА СЕМАНТИЧЕСКОГО АНАЛИЗА

В статье представлены результаты первого этапа научно-исследовательской работы по автоматизации семантической разметки текстов научно-технической направленности. Цель настоящего исследования – формирование репрезентативной коллекции текстовых материалов в области фотоники и ее систематизация в специализированной компьютерной базе данных. Описана методология сбора и предобработки данных, включающая парсинг веб-источников, очистку текстов от нефункциональных элементов, OCR-распознавание, лемматизацию и определение языка. Детализирована архитектура базы данных на основе СУБД PostgreSQL, оптимизированной для хранения метаданных документов и полнотекстового поиска посредством GIN-индексов. Сформированная коллекция включает 50 научных статей на русском языке за период 2015–2025 гг., преобразованных в структурированный формат JSON. Результаты работы создают фундаментальную основу для последующего применения алгоритмов слабой семантической разметки и теоретических разработок Минской школы вычислительной семантики

(Универсального семантического кода В. В. Мартынова; Теории автоматического порождения архитектуры знаний А. Н. Гордея). Прделанная работа позволяет решать прикладные задачи библиометрического анализа, выявления исследовательских трендов и поддержки принятия решений в области фотоники. Обнаружены такие проблемы, как неоднородность форматов исходных данных и необходимость ручной верификации, намечены пути их решения через внедрение методов активного обучения.

К л ю ч е в ы е с л о в а: фотоника; обработка естественного языка; сбор данных; база данных; PostgreSQL; семантическая разметка; Универсальный семантический код; ТАПАЗ; научно-технические тексты; полнотекстовый поиск.

AUTOMATION OF TEXT COLLECTION FORMATION AND STRUCTURING AS A BASIS FOR SEMANTIC ANALYSIS

This article presents the results of the research project first stage to automate the semantic tagging of scientific and technical texts. The goal of this study is to create a representative collection of text materials in the field of photonics and systematize it in a specialized computer database. A methodology for collecting and preprocessing data is described, including web source parsing, text cleansing from non-functional elements, OCR recognition, lemmatization, and language detection. The architecture of the database, based on the PostgreSQL DBMS, optimized for storing document metadata and full-text search using GIN indexes, is detailed. The resulting collection includes 50 scientific articles in Russian from 2015–2025, converted to the structured JSON format. The results of this work provide a fundamental basis for the subsequent application of weak semantic markup algorithms and theoretical developments of the Minsk School of Computational Semantics (V. V. Martynov's Universal Semantic Code; A. N. Hardzei's Theory for Automatic Generation of Knowledge Architecture). This work enables the solution of applied problems in bibliometric analysis, identifying research trends, and supporting decision-making in the field of photonics. Issues such as heterogeneity of source data formats and the need for manual verification were identified, and solutions through the implementation of active learning methods were outlined.

Key words: photonics; Natural Language Processing; data collection; database; PostgreSQL; semantic markup; Universal Semantic Code; TAPAZ; scientific and technical texts; full-text search.

Современные тенденции развития искусственного интеллекта (ИИ) концентрируются вокруг трёх магистральных направлений: интеллектуализация Интернета (Semantic Web), обработка больших данных (Big Data) и машинное обучение, включая языковые модели (LLM). Отставание в разработке и внедрении технологий семантического анализа текстов создает существенные барьеры для научно-технического прогресса [1, с. 400].

В период стремительного развития информационных технологий всё более очевидным становится тот факт, что именно задачи автоматической обработки текстовой информации играют ключевую роль в формировании нового поколения интеллектуальных систем. Такое внимание к текстовым данным обусловлено тем, что естественный язык служит средством представления и преобразования знания, а речь (его реализация) – основным механизмом коммуникации между людьми [2, с. 18]. С помощью нее человек передаёт знания, эмоции и идеи, язык же позволяет структурировать восприятие мира, накапливать опыт, фиксировать результаты исследований и формулировать выводы.

В условиях цифровизации всех сфер человеческой жизни естественный язык приобретает еще одну критически важную роль – он становится универсальным средством взаимодействия человека с вычислительными системами. Если всего несколько десятилетий назад использование ЭВМ требовало специальных навыков, то сегодня информационные технологии проникли в жизнь практически каждого человека, независимо от его уровня образования или профессиональной подготовки.

По этой причине возможность общения с компьютером на естественном языке перестаёт быть экспериментальной задачей и превращается в одну из самых насущных потребностей. Пользователь ожидает, что система поймёт запрос, сформулированный им без специальной подготовки и использования сложных технических команд.

Область автоматической обработки естественного языка (Natural Language Processing) развивается особенно быстро. Интенсивный рост исследований в этой сфере обусловлен не только интересом к науке, но и растущими коммерческими потребностями современного общества. Каждый интернет-пользователь ежедневно обрабатывает огромные объемы текстов: новости, документы, сообщения в переписках, комментарии, технические отчеты, публикации в социальных сетях. Для освоения таких потоков информации вручную было бы необходимо привлекать гигантские человеческие ресурсы, поэтому создаются промышленные системы обработки естественного языка, способные в автоматическом режиме анализировать тексты, извлекать элементы смысла, выявлять ключевые факты, определять специфические семантические характеристики и классифицировать содержание [3, с. 255].

Важно отметить, что автоматическая обработка текста в современных условиях требует от систем не только высокой скорости работы, но и достижения глубокого уровня понимания его содержательной стороны. Для проведения качественного анализа недостаточно разделить текст на лексемы или определить их грамматические формы. Нужно понимать контекст, учитывать многообразие языковых конструкций, верно интерпретировать намерения автора, обрабатывать двусмысленные выражения, выявлять скрытые смысловые связи. Кроме того, необходимо грамотно сочетать методы машинного обучения, семантического анализа и моделирования знаний.

Отдельное внимание обратим на то, что естественный язык используется как форма описания реальности для интеллектуальных систем. Он позволяет задавать параметры задач, формировать запросы к базам данных, описывать процессы, взаимодействовать с автоматизированными устройствами, проводить аналитическую работу. Именно благодаря этому появились виртуальные ассистенты, интеллектуальные поисковые системы, автоматические переводчики, сервисы анализа документов и многие другие технологии, которые помогают человеку работать с растущими массивами данных.

Важность всестороннего изучения обработки естественного языка возрастает в том числе в связи с тем, что информационные потоки постоянно увеличиваются не только по объему, но и по сложности. Пользователи ожидают получения мгновенного доступа к нужным данным, а компании – к оперативной обработке документов и сообщений. Соответственно, современные системы должны обеспечивать работу в режиме, близком к реальному времени, адаптироваться к новым темам, стилям и типам текстов, работать в режиме многозадачности, сохраняя при этом высокую точность анализа.

В этой связи можем также отметить актуальность применения Telegram-каналов и чат-ботов в современном образовании, что обусловлено их способностью формировать легко обновляемые и хорошо структурированные коллекции текстов [4, с. 168], превращая их в ценные источники данных для последующего семантического анализа. Более того, интеграция чат-ботов в учебный процесс как интерактивных носителей информации усиливает персонализацию обучения и обеспечивает доступ к актуальным данным на постоянной основе.

Таким образом, исследования автоматической обработки текстов не только крайне актуальны для современной лингвистики и науки в целом, но и становятся мощным двигателем развития информационного общества. От качества и эффективности работы таких систем зависит, насколько полноценно человек сможет взаимодействовать с цифровой средой, как быстро будет происходить анализ языковых данных и до какой степени информационные технологии смогут учитывать потребности пользователей. Развитие методов обработки естественного языка открывает путь к созданию новых интеллектуальных систем, способных не только распознавать слова, но и понимать смысл текста, вести диалог, анализировать большие массивы данных и принимать аргументированные решения на основе обработанной информации.

Литературный обзор

В этом контексте уникальным потенциалом обладают достижения Минской школы вычислительной семантики, в частности Универсальный семантический код (УСК) В. В. Мартынова [5] и его развитие в рамках Теории автоматического порождения архитектуры знаний (ТАПАЗ) А. Н. Гордея [6]. Данный теоретический задел открывает перспективы для создания эффективных систем автоматической семантической разметки, что критически важно для управления массивами научной информации.

Универсальный семантический код, разработанный профессором Виктором Владимировичем Мартыновым, представляет собой одну из наиболее выдающихся попыток формального описания семантической структуры естественного языка. Эта система разрабатывалась в конце 60-х – 90-х гг. XX века и была направлена на создание универсального метаязыка, позволяющего точно и однозначно кодировать значения любых высказываний. УСК изна-

чально задумывался как средство преодоления ограничений традиционных лингвистических формализмов, которые лишь описывали структуру предложения, но не обеспечивали точного и воспроизводимого представления значения языковых знаков.

В основе УСК лежит идея исчисления смыслов – представления значения любого естественно-языкового выражения в виде формализованной процедуры, поддающейся машинной обработке.

Грамматика УСК задаёт правила построения смысловых цепочек, которые представляют собой последовательности формальных операторов, отражающих структуру ситуации, описываемой средствами языка.

В отличие от логических исчислений, основанных на предикатах и кванторах, грамматика УСК строится на принципах вычислительной семантики и кибернетики, являясь гибкой системой, способной моделировать разнообразные языковые конструкции.

Формальный аппарат УСК включает набор универсальных операторов, каждый из которых имеет строго определённую функцию. Эти операторы выполняют роль строительных блоков, из которых составляется описание любой языковой ситуации, независимо от её сложности.

В трудах В. В. Мартынова описываются примеры того, как естественные предложения преобразуются в формальные процедуры в УСК. Такие преобразования показывают универсальность системы и её независимость от грамматики или лексики конкретного национального языка. В. В. Мартынов стремился создать механизм, позволяющий машине самостоятельно интерпретировать всю полноту содержания, заложенного в тексте его автором, тем самым сформировав фундамент для будущих исследований в области искусственного интеллекта. В последующих версиях УСК (УСК-5 и УСК-6) система была расширена, уточнена и дополнена, что сделало её более применимой к выполнению задач автоматизации интеллектуальной деятельности и семантического анализа.

Следует отметить, что концептуальные решения, предложенные В. В. Мартыновым, предсказали необходимость универсальных, интерпретируемых и операциональных семантических моделей в вычислительных системах.

Финализацией УСК стала ТАПАЗ – теория автоматического порождения архитектуры знаний А. Н. Гордея – автора монографии «Принципы исчисления семантики предметных областей» (1998) [2], выполненной под научным консультированием В. В. Мартынова.

Теория автоматического порождения архитектуры знаний (ТАПАЗ) представляет собой развитую формализованную систему, предназначенную для моделирования, описания и порождения знания в интеллектуальных информационных системах. ТАПАЗ-2 – расширенная и существенно усовершенствованная версия первоначальной модели (ТАПАЗ), ориентированная на выполнение широкого спектра задач – от обработки и пони-

мания естественного языка до построения моделей знаний, применимых в современных интеллектуальных технологиях. Отметим также, что основанием обеих версий ТАПАЗ является формализованная теория [6, р. 18], представленная как «множество некоторых конечных последовательностей символов, называемых формулами и термами, а также множество некоторых простых операций, производимых над ними» [7, с. 9].

Подчеркнем, что главное отличие семантического кода ТАПАЗ-2 от семантических сетей / фреймов заключается в том, что в них смысл выражений задаётся декларативно¹ – приписывается извне как готовый элемент онтологии или модели. В ТАПАЗ-2, напротив, семантика вычисляется процедурно – она выводится алгоритмически на основе синтаксической структуры самого выражения по специальным правилам интерпретации [9, с. 5–18]. Такой подход не имеет мировых аналогов.

В области фотоники, как и во многих других отраслях современного научного знания, наблюдается лавинообразный рост публикаций. За последние годы она превратилась в междисциплинарную сферу, охватывающую оптику, материаловедение, квантовые технологии, микро- и наноинженерию. Появление новых методов генерации, передачи и обработки света, развитие квантовых источников, оптических вычислителей и сенсоров приводит к тому, что поток данных растёт экспоненциально². В результате внедрение каждой новой разработки сразу же порождает целую серию последующих публикаций, создавая мощный каскад научной информации.

Объём материалов увеличивается такими темпами, что исследователи не успевают их обработать [11, с. 138], а тематическая фрагментация делает задачу ещё сложнее: статьи публикуются в очень разноплановых издательствах: журналах, препринт-архивах, сборниках конференций, причём качество и глубина исследований в одном и том же издании могут существенно различаться. Ручная сортировка, отбор релевантных источников и их тематическая группировка начинают занимать несоразмерно много времени, что снижает скорость поиска информации и затрудняет отслеживание передовых направлений в науке. Кроме того, современные исследования в области фотоники всё чаще требуют комплексного сравнения результатов, анализа больших массивов экспериментальных данных и поиска скрытых закономерностей. Следует также отметить, что, если выполнять перечисленные виды работ вручную, легко пропустить важные взаимосвязи между работами, не заметить новые тенденции в научных исследованиях или дублирование информации. Более того, растёт и вероятность возникновения ошибок и неточностей: человеческий фактор проявляется в неполноте обзора, субъективности выбора источников и ограниченной способности удерживать в поле внимания множество деталей. По этой причине возникает объективная необходимость перехода к автоматизированным системам обработки научной информации.

¹ См. статью «О декларативном и процедуральном представлении знаний» [8, с. 6].

² «Экспоненциальный рост» описывает процесс, при котором величина увеличивается с постоянно нарастающей скоростью [10].

Интеллектуальные инструменты, принцип работы которых построен на машинном обучении, языковые модели и методы семантического анализа способны не только ускорить сортировку, аннотирование и кластеризацию публикаций, но и выстраивать логические связи между исследованиями, выделять ключевые тренды и формировать структурированные карты знаний. Это позволяет ученым концентрироваться на содержательной стороне работы, не тратя силы на механическую обработку огромных массивов текстов. Следовательно, развитие фотоники как высокотехнологичной области требует новых подходов к управлению научными данными, чтобы обеспечить исследователям быстрый доступ к актуальным и систематизированным знаниям.

Автоматизация анализа литературы становится не просто инструментом повышения эффективности труда, но и важным фактором, определяющим темпы технологического прогресса. Без таких решений рассматриваемая предметная область рискует столкнуться с избытком «информационного шума», который замедлит внедрение инноваций, по причине чего интеграция интеллектуальных аналитических платформ является естественным этапом эволюции научных исследований в фотонике.

В таких условиях ручной анализ и структурирование накопленных данных становятся крайне затруднительными и постепенно теряют эффективность. Соответственно, создание автоматизированных инструментов для сбора, хранения и первичной семантической обработки корпусов текстов – чрезвычайно актуальная задача.

Целью данной работы, представляющей результаты первого этапа научно-исследовательской работы «Автоматизация семантической разметки текстов научно-технической направленности с использованием алгоритмов слабой семантической разметки», выполняемой в рамках гранта БРФФИ по договору № Ф25ИИ М-009 от 02 мая 2025 года, являлось формирование структурированной коллекции текстов по фотонике и разработка инфраструктуры базы данных для её хранения, что закладывает фундамент для последующего внедрения алгоритмов слабой семантической разметки.

Методология исследования

В процессе отбора материала использовался метод сплошной выборки: коллекция формировалась на основе небольших публикаций и фрагментов монографий, доступных в открытом доступе в Интернете. Базовым критерием отбора было упоминание объектов, относящихся к предметной области «Фотоника». Предметная область упоминания при этом не ограничивалась, что дает возможность в дальнейшем исследовать междисциплинарные связи. Итоговая коллекция на данном этапе включает 50 текстовых документов на русском языке за период 2015–2025 гг. Для каждого документа в библиографическом списке фиксировалось краткое наименование и источник, позволяющий восстановить контекст оригинальной публикации.

Обработка данных, осуществлявшаяся с использованием языка программирования «Python» и специализированных библиотек, проходила 4 этапа, представленных ниже.

1. Сбор данных: парсинг веб-страниц выполнялся с помощью `BeautifulSoup` и `Scrapy`.

2. Предварительная обработка включала комплексную очистку от HTML-тегов, метаданных и служебных символов. Для сканированных PDF-документов применялось OCR-распознавание (`PyPDF2`, `Tesseract`).

3. Лингвистическая обработка: для работы с русскоязычными текстами использовались библиотеки `pymorphy2` и `natasha` для лемматизации и токенизации. Языковая принадлежность определялась с помощью `langdetect`, что позволило отфильтровать нерусскоязычные материалы.

4. Структурирование: все тексты были преобразованы в унифицированный JSON-формат с полями: `title` (заголовок), `abstract` (аннотация), `full_text` (полный текст), `source` (источник), `key words` (автоматически извлечённые ключевые слова).

Для хранения и управления коллекцией была разработана реляционная база данных на основе СУБД PostgreSQL [12], выбранная за её надёжность, производительность и мощную поддержку полнотекстового поиска.

Основной сущностью¹ является таблица `documents` (см. схему ниже).

```
```sql
CREATE TABLE documents (
 id SERIAL PRIMARY KEY,
 title TEXT,
 full_text TEXT,
 metadata JSONB,
 publication_date DATE
);
```
```

Раздельное хранение основного текста (`full_text`) и структурированных метаданных (`metadata` в формате JSONB) обеспечивает гибкость и оптимизацию выполнения запросов.

Для обеспечения высокой скорости полнотекстового поиска по содержанию документов на русском языке был создан GIN-индекс:

```
```sql
CREATE INDEX idx_full_text ON documents USING GIN (to_tsvector('russian', full_text));
```
```

¹ Согласно П. П. Чену, сущность определяется как объект, который может быть однозначно идентифицирован [13, р. 10]. В рамках моделирования предметной области сущность представляет собой объект с уникальной идентичностью, обладающий состоянием и жизненным циклом, а также отражающий значимый стереотип реального мира; в реляционной базе данных сущность, как правило, реализуется в виде таблицы, а её экземпляры – в виде строк.

Данный индекс обеспечивает выполнение сложных семантических запросов с использованием операторов PostgreSQL `@@` (для `tsquery`), что является необходимым условием для их последующего анализа.

Корректность загрузки данных проверялась скриптом, который выборочно сравнивал записи в базе данных с исходными JSON-файлами. Ошибки импорта были минимизированы за счёт этапа предварительной нормализации данных.

Результаты и обсуждение

В результате выполнения работы получены следующие результаты:

1) сформирована репрезентативная коллекция из 50 научных статей по фотонике на русском языке;

2) создана и заполнена работоспособная база данных PostgreSQL, обеспечивающая надежное хранение текстов и метаданных и эффективный полнотекстовый поиск.

3) получена основа для будущего анализа связей между документами (цитирования, тематической близости и др.).

Сформированный ресурс может быть использован для решения следующих прикладных задач:

1) проведения библиометрических исследований и анализа научных трендов;

2) выявления ключевых направлений и «горячих тем» в фотонике;

3) поддержки принятия решений при планировании новых научных проектов.

Основные трудности были связаны с неоднородностью исходных форматов данных (PDF, HTML, DOC) и ошибками автоматического извлечения терминов, потребовавшими частичной ручной верификации. В качестве пути решения на последующих этапах планируется внедрение подхода активного обучения (active learning), при котором модель будет запрашивать верификацию у эксперта только для неопределённых случаев, что значительно сократит трудозатраты.

В рамках первого этапа научно-исследовательской работы успешно создана техническая и информационная основа для автоматизации семантической разметки текстов в области фотоники. Разработанная методология сбора, предобработки и структурирования текстовых данных, а также реализованная архитектура базы данных представляют собой законченное решение, готовое к интеграции с алгоритмами слабой семантической разметки на основе ТАПАЗ. Дальнейшая работа будет направлена на разработку и тестирование алгоритмов разметки, их адаптацию к специфике научно-технических текстов и повышение степени автоматизации всего процесса обработки данных. Полученные результаты вносят вклад в развитие отечественных технологий семантического анализа и имеют практическую ценность для научного сообщества в области фотоники.

ЛИТЕРАТУРА

1. Алгоритм автоматизированной семантической разметки текстов / М. И. Святощик, А. Н. Гордей, Р. С. Панащик [и др.]. // XVIII Междунар. науч.-техн. конф. «Развитие информатизации и государственной системы научно-технической информации». – Минск : ОИПИ НАН Бела-руси, 2024. – С. 400–404.
2. *Гордей, А. Н.* Принципы исчисления семантики предметных областей / А. Н. Гордей. – Минск, 1998. – 156 с.
3. *Святощик, М. И.* Системы автоматической обработки текстов и их уровневое построение / М. И. Святощик // X Междунар. науч. конф. «Молодые ученые в инновационном поиске», 11–12 марта, Минск : МГЛУ, 2021. – С. 254–259.
4. *Гибкий, П. В.* Использование Telegram-канала и чат-бота в практике подготовки преподавателей китайского языка / П. В. Гибкий, А. А. Карпович // Вестн. Ом. гос. пед. ун-та. Гуманит. исследования. – 2025. – № 4 (49). – С. 165–169. – URL: <https://cyberleninka.ru/article/n/ispolzovanie-telegram-kanala-i-chat-bota-v-praktike-podgotovki-prepodavateley-kitayskogo-yazyuka> (дата обращения: 23.12.2025).
5. *Мартынов, В. В.* Универсальный семантический код: УСК-3 / В. В. Мартынов ; Под ред. А. Е. Михневича. – Минск : Наука и техника, 1984. – 132 с.
6. *Hardzei, A.* Theory for Automatic Generation of Knowledge Architecture: TAPAZ-2 / A. Hardzei ; transl. from Rus. I. M. Boyko // Republican Institute of Higher Education, 2017. – Rev. English edn. – Minsk : RIHE, 2017. – 50 p.
7. *Расёва, Е.* Математика метаматематики / Е. Расёва, Р. Сикорский. – М. : Наука, 1972. – 295 с.
8. *Гордей, А. Н.* О декларативном и процедуральном представлении знаний / А. Н. Гордей // Иностранные языки в высшей школе ; ФГБОУВО «Рязанский государственный университет имени С. А. Есенина». – Рязань : Редакц.-изд. центр РГУ им. С. А. Есенина, 2021. – Вып. 3 (58). – С. 5–12.
9. *Гордей, А. Н.* Вновь о семантическом подходе к обработке данных на естественном языке / А. Н. Гордей // Иностранные языки в высшей школе ; ФГБОУВО «Рязанский государственный университет имени С. А. Есенина». – Рязань : Редакц.-изд. центр РГУ им. С. А. Есенина, 2023. – Вып. 3 (66). – С. 5–18.
10. Тего-семантические основания искусственного моделирования интеллектуальной деятельности / А. Н. Гордей, Р. С. Панащик, М. И. Святощик [и др.]. // IV Междунар. форум IT-Академграда «Искусственный интеллект в Беларуси» : доклады, Минск, 13–14 октября 2025 г. – Минск : ОИПИ НАН Беларуси, 2025. – С. 138–147.
11. Exponential growth and decay // Khan Academy. – URL: <https://www.khanacademy.org/math/algebra/x2f8bb11595b61c86:exponential-growth-decay> (date of access: 08.04.2024).
12. PostgreSQL Documentation: Full Text Search. – URL: <https://www.postgresql.org/docs/current/textsearch.html> (date of access: 10.12.2025).
13. *Chen, P. P.* The Entity-Relationship Model: Toward a Unified View of Data / P. P. Chen. – Cambridge, Massachusetts : Center for Information Systems Research, MIT Sloan School of Management, 1977. – 44 p.

Поступила в редакцию 28.01.2026