

СИСТЕМЫ КАТАЛОГИЗАЦИИ И ИНФОРМАЦИОННОГО ПОИСКА

УДК 025.4

<https://doi.org/10.33186/1027-3689-2025-5-81-103>

Принципы, методы, алгоритмы интеллектуальной системы библиографического поиска

А. А. Боряев

*Балтийский государственный технический университет «ВОЕНМЕХ»
им. Д. Ф. Устинова, Санкт-Петербург, Российская Федерация,
sasa1953@yandex.ru, <https://orcid.org/0000-0001-6403-710X>*

Аннотация. В статье представлены научно обоснованные принципы построения и алгоритмы интеллектуальной системы библиографического поиска на основе современных интеллектуальных технологий (интеллектуальный лингвистический анализ текстов, интеллектуальное управление базами данных) и методов поиска в интернете. В рамках теории библиографии впервые с учётом особенностей современного интеллектуального программного обеспечения разработан новый алгоритм комплексной автоматизированной системы библиографического поиска и поиска специализированных изданий для публикации научных работ. Разработанные методы включают создание нормализованной базы ключевых слов на основе рубрик Универсальной десятичной классификации (УДК) и присвоение автоматически ключевых слов научным работам из этой нормализованной базы данных интеллектуальной системой лингвистического анализа текстов. Предлагаемое использование ключевых слов в качестве информационных меток в виде названий рубрик УДК, охватывающих все отрасли знаний, позволяет наиболее полно и конкретно маркировать и характеризовать научные работы (статьи) и научные издания (журналы, сборники статей научных конференций). Поиск научных работ, осуществляемый предлагаемой системой библиографического поиска, и поиск специализированных изданий для их публикации осуществляются системой поиска автоматически по нормализованным ключевым словам, присвоенным научным статьям и научным журналам за счёт интеллектуального лингвистического анализа тестов научных работ. Исследование представляет собой социальный вклад, демонстрируя важность разработки и использования лучших практик, процессов и стратегий библиографического поиска для развития науки в современном обществе. При этом библиографическая информация и в социальном плане явля-

ется одним из методов решения проблемы поиска связи между личностным и общественным знанием. Реализация результатов исследования, в соответствии с представленными алгоритмами, заключается в создании структуры специального общедоступного интернет-поискового ресурса.

Ключевые слова: библиографический поиск, ключевые слова, профильные издания, публикация, формализация библиографического поиска, автоматизированная интеллектуальная система

Для цитирования: Боряев А. А. Принципы, методы, алгоритмы интеллектуальной системы библиографического поиска // Научные и технические библиотеки. 2025. № 5. С. 81–103. <https://doi.org/10.33186/1027-3689-2025-5-81-103>

CATALOGING AND SEARCH SYSTEMS

UDC 025.4

<https://doi.org/10.33186/1027-3689-2025-5-81-103>

The principles, methods and algorithms for bibliographic search intelligence system design

Alexander A. Boryaev

*“VOENMEH” D. Ustinov Baltic State Technical University,
St. Petersburg, Russian Federation,
sasa1953@yandex.ru, <https://orcid.org/0000-0001-6403-710X>*

Abstract. The author discusses the algorithms and scientifically substantiated principles for designing bibliographic search intelligent system based on modern intelligent technologies, e. g. intelligent linguistic text analysis, smart database management, and Internet search methods. For the first time, in the context of bibliographic theory and capabilities of modern intelligent software, the new algorithm for the integrated system of bibliographic and specialized periodicals search for publishing research papers is developed. The proposed methods comprise automatic generation of normalized keywords database on the basis of the UDC headings and assignment of the keywords to the papers in this

normalized database by the intelligent linguistic text analysis system. Using the keywords as the tags in the form of UDC headings comprising every field of knowledge enables to tag and characterize scientific papers (articles) and scientific publications (journals, conference proceedings) in the most complete and specific way. The proposed bibliographic search system searches for scientific articles and specialized serials automatically by the normalized keywords, assigned to the scientific articles and journals through the intelligent linguistic text analysis. The study is to make social contribution by demonstrating the significance of design and use of the best practices, processes and strategies of bibliographic search that make the key element of modern science advancement. Moreover, in the social aspect, the bibliographic information is a way to find connections between the personal and social knowledge. The study findings can be implemented, in accordance with proposed algorithms, in designing the structure of specialized public Internet-based search resource.

Keywords: bibliographic search, keywords, core publication, publication, bibliographic search formalization, automated intelligence system

Cite: Boryaev A. A. The principles, methods and algorithms for bibliographic search intelligence system design // Scientific and technical libraries. 2025. No. 5, pp. 81–103. <https://doi.org/10.33186/1027-3689-2025-5-81-103>

1. Введение

В современных условиях развития информационного общества и лавинообразного роста потоков информации задача повышения эффективности информационного поиска является актуальной и весьма значимой. Представленная в статье система библиографического поиска имеет значение с точки зрения разработки информационно-аналитического инструментария для информационной поддержки исследований в научно-технической сфере.

В настоящее время основной тенденцией при построении информационно-аналитического инструментария для поиска информации является сочетание библиографического поиска с поиском по ключевым словам и кодам классификации, а также с использованием систем автоматической тематической классификации полнотекстовых ресур-

сов на основе искусственного интеллекта. Такими исследованиями достаточно успешно занимаются не только за рубежом, но и в России, например, в ГПНТБ России, ВИНИТИ РАН и др., о чём свидетельствуют многочисленные публикации.

Следует особо отметить проведённые в этом направлении исследования и результаты деятельности Группы развития классификационных систем и стандартизации ГПНТБ России, руководимой ведущим научным сотрудником, председателем Методического совета по классификационным системам НТИ канд. техн. наук Е. Ю. Дмитриевой, о чём также свидетельствует ряд имеющихся публикаций.

Мир вступил в информационную фазу своего развития, характеризующуюся совершенствованием информационных технологий и средств коммуникации, среди которых совершенствование способов библиографического поиска в направлении его автоматизации и использование интеллектуальных технологий занимают особое место.

Предлагаемый в статье способ библиографического поиска – это библиографический поиск, основанный на комплексном использовании следующих технологий:

- методов поиска информации в сети Интернет,
- технологий интеллектуального анализа текстов с целью выделения ключевых слов,
- интеллектуальных систем управления базами данных.

В настоящее время обострилось противоречие между накопленной информацией и малой долей информации требуемой, что привело к необходимости перехода от изучения имеющихся проблем к исследованию проблем автоматизированного поиска информации. Эволюция библиографического поиска обусловлена его спецификой и представляет собой проблему нахождения соотношения между общественным и личностным знанием [1].

В фундаментальной работе А. В. Соколова [2] обсуждаются проблемы библиографического поиска как объекта автоматизации, развития идей автоматизации семантического библиографического поиска, автоматизированных библиографических систем, а также перспективы развития автоматизированного библиографического поиска. В диссертации Д. Г. Лахути [3] рассмотрены проблемы, возникающие при интеллектуализации информационно-поисковых систем.

Библиографический поиск – это процесс интерпретации и идентификации документов по их формальным и содержательным признакам, осуществляемый на основе использования библиографических ресурсов и средств с целью удовлетворения библиографических потребностей, при этом библиографический поиск – процесс библиографической деятельности.

Структура поискового образа – запроса представляется как совокупность поисковых признаков, необходимых для интерпретации и идентификации документа в соответствии с запросом. При проведении автоматизированного библиографического поиска требуется предельная формализация поискового образа запроса, сведение его к элементам, необходимым и достаточным для осуществления процесса, то есть создание поискового предписания.

Предлагаемый в статье способ библиографического поиска относится к системам и средствам поиска информации. Реализация способа относится к системам поиска информации (библиографического поиска) о научной и исследовательской деятельности, таким как статьи, диссертации, монографии и прочие публикации, а также к системам поиска профильных изданий для публикации научного произведения.

Известно большое количество способов оптимизации процесса библиографического поиска. Однако проблема больших временных затрат на традиционный поиск по статьям, диссертациям и прочим источникам научной и исследовательской информации, обусловленная их колоссальным объёмом и отсутствием достаточной систематизации, остаётся актуальной. Существуют автоматизированные системы библиографического поиска научной информации. Например:

система библиографического поиска Российского индекса научного цитирования (РИНЦ): <https://elibrary.ru/querybox.asp?scope=newquery>;

единая библиографическая и реферативная база данных рецензируемой научной литературы Scopus издательства Elsevier: <https://journalfinder.elsevier.com/>, <https://www.sciencedirect.com/> и <https://www.scimagojr.com/>;

поисковая система реферативной базы Web of Science: <https://clarivate.com/cis/solutions/web-of-science/> и прочие.

Однако поиск по указанным системам значительно затруднён из-за большой вариативности входных данных для поиска, в результате чего системы поиска выдают большое количество информации (многие

десятки и сотни ссылок), среди которой информация, отвечающая требованиям авторов направлениям поиска, составляет лишь малую часть.

В литературе имеется значительное количество источников, где рассматриваются отдельные элементы, которые могут быть включены в автоматизированную интеллектуальную систему библиографического поиска.

Среди известных патентов, в которых рассмотрены различные системы поиска и обработки информации, следует выделить следующие способы: способ поиска информации в политематических массивах неструктурированных текстов [4]; способ двухуровневого поиска информации в предварительно преобразованном структурированном массиве данных [5]; способ поиска информации в массиве текстов, при котором производится сравнение фраз по смыслу. При перечисленных выше способах не происходят анализ содержания и систематизация текстов, по которым происходит поиск. Это приводит к необходимости ручного перебора множества результатов поиска и, следовательно, большим временным затратам при осуществлении информационного поиска в некоторых случаях, в частности, по текстам, содержащим информацию о научной и исследовательской деятельности [6]. Известен способ, где для улучшения ранжирования поиск производится с использованием информации о статье. Недостатком этих способов является то, что для сортировки и ранжирования результатов поиска используются данные, собранные на клиентской стороне, которые могут отсутствовать в принципе или быть недоступными для некоторых поисковых систем [7].

Запатентован способ [8] поиска информации в массивах необработанных текстов. Однако в описываемом в патенте способе не конкретизируются понятие «база терминов» и методика её составления и распространения, а также способ выбора терминов для присваивания конкретному научному произведению. При этом «база терминов» используется для анализа текста названия научного произведения и дальнейшего ранжирования по соответствию данных поискового запроса. Недостатком в данном случае является то, что используемый терминологический признак является единственным, что не отражает многие другие логические связи, которые могут быть между текстами в множестве искомых. Также не учитывается выбор автором научного произведения ключевых слов, которые являются дополнительными

техническими данными, позволяющими охарактеризовать научное произведение. Отсутствует также метод определения профильного издания для публикации научного произведения.

Среди большого количества литературы по автоматизированному лингвистическому анализу текстов следует выделить работы, содержащие описание методов, средств и алгоритмов лингвистического анализа текстов научных произведений интеллектуальной системой с целью выделения ключевых слов, иногда называемых «выделением смыслов» [9–12].

Особенностям современных методов библиографического поиска посвящено достаточно большое количество работ [13–16].

Анализ представленных различными авторами проблем библиографического поиска показал, что приоритетным направлением исследований станет совершенствование интеллектуальных поисковых технологий с привлечением современных интеллектуальных технологий, учитывающих два типа классификации методов библиографического поиска в электронной среде – в зависимости от типа используемой поисковой системы или избранной поисковой стратегии.

Известные к настоящему времени исследования не дают информации о том, как осуществлялся процесс формирования методологических и теоретических представлений о библиографическом поиске, на основе которых можно было бы создать автоматизированные системы.

Следует отметить, что в последнее время происходит интенсивное развитие интеллектуальных технологий во многих областях науки и техники и, в частности, технологий интеллектуального анализа текстов с целью выделения ключевых слов и интеллектуализации управления базами данных. Предлагаемый в статье способ библиографического поиска основан на комплексном использовании следующих технологий:

методов поиска информации в сети Интернет,

технологиях интеллектуального анализа текстов с целью выделения ключевых слов,

интеллектуальных системах управления базами данных.

В настоящее время интернет предоставляет широкие возможности для осуществления библиографического поиска на основе следующих преимуществ:

1. Библиографический поиск проводится на базе широкого круга источников.

2. Поиск отличается высокой оперативностью.

3. Такой поиск предполагает возможность большой мобильности (быстрая смена разных библиотек, поисковых систем, баз данных, электронных каталогов, возможность оперативной смены сформулированных первоначальных поисковых запросов и т. д.).

4. Появляются новые, ранее сложно реализуемые возможности библиографического поиска (выявление динамики публикаций по определённому научному направлению, обобщение статистики по годам и по месту публикации, а также по отдельным видам публикаций и т. д.).

В настоящее время разработаны методы интеллектуального лингвистического анализа текстов с целью выявления ключевых слов или выделения лексических единиц (выделения смыслов) для анализируемых текстов, которые также могут быть использованы для определения ключевых слов для поисковой системы.

Известно, что лексическая единица – это обозначение отдельного понятия в естественном или специально созданном искусственном языке. Лексические единицы могут иметь вид слова, устойчивого словосочетания, аббревиатуры, символьного кода и т. п.

По ключевым словам можно определить общий смысл, настрой и направление текста. Они выступают основой любого текстового материала. Как ключевые знаки в тексте могут выступать не только слова, но и словосочетания и даже предложения. Под ключевыми словами эпохи в исследовании понимаются также «особые лексические маркеры, отражающие актуальное состояние концептосферы языка и обладающие весомым прагматическим содержанием» [19].

Таким образом, отличие ключевых слов от лексических единиц заключается в том, что лексические единицы – это более общее понятие, а ключевые слова – это конкретная категория слов, которые обладают определёнными характеристиками и функциями.

Например, семейство технологий Text Mining. Методами обработки текстов на естественном языке эффективно решены задачи выделения ключевых слов для построения поисковых индексов на их основе, а также создание нормализованной базы ключевых слов в соответствии с различными отраслями и направлениями знаний; присваивания этих ключевых слов произведениям науки; их выбора и присвоения произведениям науки с помощью лингвистического анализа текстов

научных произведений; присваивание нормализованных ключевых слов изданиям в соответствии с их профилем.

Использование баз данных (БД) и информационных систем (ИС) становится неотъемлемой составляющей любой информационной технологии. Это требует применения теоретических положений и инструментальных средств разработки БД и ИС. Общую тенденцию интеллектуализации и расширения функций систем управления базами данных (СУБД) обеспечивают две важные возможности, которыми обладают практически все современные СУБД, – поддержка хранимых процедур и применение триггеров. Число виртуальных процессоров определяет администратор БД, исходя из реальных ресурсов вычислительной системы и сети клиентов. Если вычислительная система является многопроцессорной, то разные виртуальные процессоры могут обслуживаться разными реальными процессорами. В настоящее время предложен совершенно новый тип моделирования и показана возможность алгебраизации теории нейронных сетей. Это позволит поднять её на качественно более высокий уровень, а также улучшит согласование идеи нейрокомпьютера с современными тенденциями суперкомпьютеров, GRID-вычислений, агентных систем, компьютерных кластеров и т. д. Благодаря многочисленным исследованиям представляется возможным создание имитационной модели СУБД, основанной на новом подходе в области бесконфликтного использования ресурсов с применением новой технологии проектирования.

В настоящее время в лингвистике не существует чёткого определения синонимичных и не синонимичных понятий.

Проблема синонима – одна из старейших в филологической науке. Человеческая мысль интересуется ею около двадцати четырёх столетий – начиная с софиста Продика, учителя Сократа, и заканчивая лингвистами наших дней. С момента появления самого термина «синонимы» (от греч. *synonymos* – одноимённый), введённого ещё античными греческими философами, споры о характеристике функций синонимов, об их роли в языковом общении не утихают и по сей день. Даже у наиболее видных семасиологов настоящего времени нет единого мнения по этому вопросу. Можно спорить о том, повлияла ли на научную традицию составленная древними греками характеристика синонимов как слов, близких по значению, но, отдавая должное наблюдательности, а также изобретательности наших предков, заметим, что данное

ими определение осталось практически без изменения до настоящего времени.

Синонимичные понятия – это понятия, которые имеют одинаковый смысл. Синонимия – это тип семантических отношений в языке, который подразумевает полное или частичное совпадение значений языковых выражений.

Известно, что различают синонимию по сигнификату и синонимию по денотату.

Синонимия по сигнификату состоит в совпадении (частичном или полном) тех понятий, которые обозначаются языковыми выражениями. Например, слова «лингвистика» и «языковедение» обозначают одно и то же понятие (гуманитарная наука, изучающая естественный язык), а потому являются полными синонимами.

Синонимия по денотату представляет собой совпадение предметных значений языковых выражений. В этом смысле синонимичны выражения «Наполеон Бонапарт», «Император Франции с 1804 по 1814 год», «Полководец, потерпевший поражение под Ватерлоо», обозначающие один и тот же единичный объект; или выражения «страус» и «самая крупная птица на Земле», обозначающие один и тот же класс объектов.

Синонимы отражают реальную действительность не непосредственно, а через тождество, которое выступает в качестве его содержательной, понятийной, семантической стороны. По указанной причине синонимия действительно не может не оказывать известного обратного влияния на тождество и познавательную деятельность человека. Поэтому всё в языке прямо существует только для того, чтобы передавать какое-то содержание, какую-то информацию, какой-то смысл. В языкознании нет формальных правил для определения сходств и различий по смыслу между словами в тексте, отождествление их остаётся интуитивным.

В связи с изложенным выше для описания системы библиографического поиска в статье нами выбраны общеупотребительные термины: метод, технология, система, которые понятны не только лингвистам и философам и позволяют с точки зрения лексического анализа говорить о правильности и однозначности данного определения.

Литературных источников, посвящённых использованию технологий интеллектуального анализа текстов с целью выделения ключевых слов совместно с интеллектуальными системами управления базами

данных для создания способа библиографического поиска, нами не обнаружено.

Наиболее полно предлагаемая в статье система библиографического поиска представлена в статье автора: Boryaev A. A. (2024). Development of intelligent system of global bibliographic search // Journal of Librarianship and Information Science, 0 (0). <https://doi.org/10.1177/09610006231223509> [17].

2. Описание принципов, методов и алгоритмов способа библиографического поиска

В настоящее время в ходе научной и исследовательской деятельности большое количество времени тратится авторами на поиск и анализ информации, описывающей результаты проведённых исследований или других видов научной деятельности. Известные способы библиографического поиска в электронных библиографических и реферативных БД не решают в достаточной мере проблему фильтрации большого количества результатов поиска, поскольку фильтрация осуществляется, как правило, по формальным признакам и не охватывает смысловое содержание работ. Большое количество времени может быть затрачено на поиск профильного издания для публикации научного произведения.

Результатом данного исследования является сокращение времени поиска информации о научной и исследовательской деятельности за счёт её формализации и использования автоматизированных интеллектуальных систем. Использование предлагаемого способа библиографического поиска позволяет также значительно сократить время на поиск профильного научного издания для публикации научного произведения.

Указанный результат достигается за счёт предварительного создания нормализованной базы ключевых слов в соответствии с различными отраслями и направлениями знаний и присваивания этих ключевых слов произведениям науки авторами или выбора их и присваивания произведениям науки с помощью лингвистического анализа текстов интеллектуальной системой, а также присваивания нормализованных ключевых слов изданиям в соответствии с их профилем. При проведении библиографического поиска из нормализованной базы ключевых слов автором или интеллектуальной лингвистической системой на ос-

нове всего научного произведения или его части выбираются ключевые слова, соответствующие требуемой тематике поиска, и загружаются в поисковую систему для осуществления библиографического поиска. При поиске профильных изданий для публикации научного произведения производятся предварительный анализ ключевых слов, используемых авторами опубликованных произведений в конкретных изданиях и сравнение их с имеющейся нормализованной базой ключевых слов с целью выбора соответствующих профилю изданий нормализованных ключевых слов и размещения их на сайтах изданий и в специальной поисковой системе научных изданий для осуществления поиска профильных научных изданий.

В качестве нормализованной базы ключевых слов предлагается использовать названия рубрик Универсальной десятичной классификации (УДК) – международной системы классификации информации, которая доступна по адресу: <https://udc-hub.com/index.php>. UDC Online – это полная стандартная обновлённая версия Универсальной схемы десятичной классификации. Она предлагает расширенный поиск и просмотр веб-страниц и поддерживает управление номерами UDC. UDC Online всегда соответствует последней официальной версии Основного справочного файла UDC. Сервис предоставляет доступ к UDC на нескольких языках.

В качестве альтернативы (в связи с тем, что обновления в виде изменений и дополнений к таблицам УДК для актуализации эталона УДК на русском языке начиная с 2019 г. Международным консорциумом УДК (UDCC) не передаются) для создания базы ключевых слов может быть использована УДК, опубликованная в России (Универсальная десятичная классификация. Москва : Издательство стандартов, 1972 – без обновлений).

В качестве другой альтернативы для создания нормализованной базы ключевых слов могут быть использованы стандарты СИБИД, регламентирующие термины и определения в конкретных областях научного и технического знания.

Базовой классификацией Государственной системы научно-технической информации (ГСНТИ) является Государственный рубрикатор научно-технической информации (ГРНТИ), эталонная версия которого актуализируется каждые три года (ГОСТ Р 7.0.49). ГРНТИ также

может быть использован для создания нормализованной базы ключевых слов.

Количество ключевых слов, присваиваемых одному произведению науки, в предлагаемом способе библиографического поиска не более десяти. С целью большей конкретизации результатов присвоение ключевых слов производится не только конечными терминами УДК (вниз по иерархии), но и терминами из названий рубрик и глав УДК (вверх по иерархии), так как, например, слово «нефть» встречается в нескольких разделах: в разделе 547 УДК «Органическая химия» находится понятие «Химия нефти», в разделе 553 УДК «Месторождения полезных ископаемых» – «Месторождения нефти», в разделе 622 «Горное дело» – «Добыча нефти», в разделе 662 УДК «Взрывчатые вещества. Топлива» – «Нефть в качестве топлива» и т. д.

Примеры присваиваемых ключевых слов:

«нефть; органическая химия»;

«нефть; месторождение полезных ископаемых; месторождения нефти»;

«нефть; горное дело; добыча нефти».

Таким образом, будут учтены синонимичные термины, определяемые одними и теми же дескрипторами.

Общая компьютерная среда для реализации предлагаемого библиографического поиска, алгоритм работы и функционирования всех алгоритмических процедур, которые подробно описаны в разделе 3 и которая решает в том числе задачи поиска и формирования перечня присваиваемых ключевых слов, как важный информационно-аналитический инструмент для функционирования способа в целом, в настоящее время находится в стадии создания коллективом специалистов, способных решить данную задачу в том числе из-за рубежа.

3. Алгоритмы осуществления библиографического поиска

3.1. Алгоритм создания нормализованной базы ключевых слов

1. Осуществляется создание нормализованной базы ключевых слов в соответствии с различными отраслями и направлениями знаний.

Создание нормализованной базы ключевых слов осуществляется путём использования в создаваемой базе в качестве ключевых слов соответствующих номеров и названий рубрик УДК [18]. В данном случае создаваемая нормализованная база ключевых слов – это совокуп-

ность названий рубрик УДК. Создаваемая база данных должна поддерживать области применения в соответствии с обращениями, предусмотренными алгоритмом поиска.

2. Создаваемая нормализованная база ключевых слов в соответствии с различными отраслями и направлениями знаний и присваивание этих ключевых слов произведениям науки авторами или выбор их и присваивание произведениям науки с помощью лингвистического анализа текстов интеллектуальной системой, а также присваивание нормализованных ключевых слов изданиям в соответствии с их профилем производятся по следующему алгоритму:

2.1. Создание нормализованной базы ключевых слов осуществляется путём использования в создаваемой базе в качестве ключевых слов соответствующих номеров и названий рубрик УДК [18]. В данном случае создаваемая нормализованная база ключевых слов – это совокупность названий рубрик УДК. Создаваемая база данных должна поддерживать следующие области применения в соответствии со следующими алгоритмами:

2.2. При обращении в нормализованную базу ключевых слов автора научного произведения база данных ключевых слов должна допускать поиск, просмотр и самостоятельный выбор ключевых слов (названия рубрик УДК) автором в соответствии с отраслью и направлению знаний созданного автором научного произведения. В дальнейшем автор самостоятельно вводит выбранные ключевые слова в своё научное произведение.

2.3. Выбор ключевых слов из базы данных ключевых слов и присваивание их произведениям науки с помощью лингвистического анализа текстов интеллектуальной системой осуществляется по следующему алгоритму.

3.2. Алгоритм лингвистического анализа текста научного произведения интеллектуальной системой

Выбор ключевых слов из базы данных ключевых слов и присваивание их произведениям науки с помощью лингвистического анализа текстов интеллектуальной системой осуществляется по следующему алгоритму:

1. С помощью компьютерной программы производится лингвистический анализ текста научного произведения интеллектуальной систе-

мой (интеллектуальный поиск) с целью интеллектуального анализа слов и словосочетаний текста научного произведения и сравнения их с ключевыми словами из нормализованной базы ключевых слов и сохранение полученной информации в виде перечня совпавших ключевых слов.

2. Компьютер формирует перечень ключевых слов, найденных по п. 1 из базы данных ключевых слов в соответствии со словами из текста научного произведения, фиксируя количество совпадений по каждому ключевому слову, и осуществляет их сортировку по убыванию количеств совпадений.

3. Производится вывод на экран монитора результатов поиска нормализованных ключевых слов с количеством совпадений.

4. Производится компьютерный анализ результатов поиска ключевых слов на принадлежность к одному из десяти основных классов УДК [18] по нормализованной базе ключевых слов с учётом количества совпадений.

5. Принадлежащие к одному классу УДК ключевые слова присваиваются анализируемому произведению науки.

3.3. Алгоритм присваивания ключевых слов изданиям в соответствии с их профилем

Присваивание ключевых слов изданиям в соответствии с их профилем осуществляется по следующему алгоритму:

1. С помощью компьютерной программы производится лингвистический анализ текста или аннотации научного издания (как правило, аннотация размещается на интернет-сайте научного издания) интеллектуальной системой (интеллектуальный поиск) с целью интеллектуального анализа слов и словосочетаний текста аннотации научного издания и сравнения их с ключевыми словами из нормализованной базы ключевых слов. Программа сохраняет полученную информацию в виде перечня совпавших ключевых слов.

2. Компьютер формирует перечень ключевых слов по п. 1 из базы данных ключевых слов, соответствующих словам из текста научного произведения, фиксируя количество совпадений по каждому ключевому слову, и осуществляет их сортировку по убыванию количеств совпадений.

3. Производится вывод на экран монитора результатов поиска нормализованных ключевых слов с количеством совпадений.

4. Производится компьютерный анализ результатов поиска нормализованных ключевых слов на принадлежность к одному из десяти основных классов УДК [18] по нормализованной базе ключевых слов с учётом количества совпадений.

5. Принадлежащие к одному или нескольким классам УДК нормализованные ключевые слова (не менее двух ключевых слов) присваиваются анализируемому научному изданию, так как научное издание может иметь несколько тематических направлений.

3.4. Алгоритм проведения библиографического поиска

При проведении библиографического поиска из нормализованной базы ключевых слов автором или интеллектуальной лингвистической системой на основе всего научного произведения или его части выбираются ключевые слова, соответствующие требуемой тематике поиска, и загружаются в поисковую систему для осуществления библиографического поиска. При этом проведение библиографического поиска осуществляется по следующему алгоритму:

1. С помощью компьютерной программы производится лингвистический анализ текста научного произведения или его части интеллектуальной системой (интеллектуальный поиск) с целью интеллектуального анализа слов и словосочетаний текста научного произведения и сравнения их с ключевыми словами из нормализованной базы ключевых слов, и программа сохраняет полученную информацию в виде перечня совпавших ключевых слов.

2. Компьютер формирует перечень ключевых слов, найденных по п. 1 из базы данных ключевых слов в соответствии со словами из текста научного произведения или его части, фиксируя количество совпадений по каждому ключевому слову, и осуществляет их сортировку по убыванию количеств совпадений.

3. Производится вывод на экран монитора результатов поиска нормализованных ключевых слов с количеством совпадений.

4. Производится компьютерный анализ результатов поиска нормализованных ключевых слов на принадлежность к одному из десяти основных классов УДК [16] по нормализованной базе ключевых слов с учётом количества совпадений.

5. Принадлежащие к одному классу УДК нормализованные ключевые слова присваиваются анализируемому произведению науки.

3.5. Алгоритм присваивания

нормализованных ключевых слов изданиям

Присваивание нормализованных ключевых слов изданиям в соответствии с их профилем осуществляется по следующему алгоритму:

1. С помощью компьютерной программы производится лингвистический анализ текста аннотации научного издания (как правило, размещается на интернет-сайте научного издания) интеллектуальной системой (интеллектуальный поиск) с целью интеллектуального анализа слов и словосочетаний текста аннотации научного издания и сравнения их с ключевыми словами из нормализованной базы ключевых слов, и программа сохраняет полученную информацию в виде перечня совпавших ключевых слов.

2. Компьютер формирует перечень ключевых слов по п. 1 из базы данных ключевых слов, соответствующих словам из текста аннотации научного издания, фиксируя количество совпадений по каждому ключевому слову, и осуществляет их сортировку по убыванию количества совпадений.

3. Производится вывод на экран монитора результатов поиска нормализованных ключевых слов с количеством совпадений.

4. Производится компьютерный анализ результатов поиска нормализованных ключевых слов на принадлежность к одному из десяти основных классов УДК [18] по нормализованной базе ключевых слов с учётом совпадений.

5. Принадлежащие к одному или нескольким классам УДК нормализованные ключевые слова (не менее двух ключевых слов) присваиваются анализируемому научному изданию.

6. Компьютерная поисковая программа сравнивает ключевые слова, присвоенные произведениям науки в соответствии с п. 1 с ключевыми словами, присвоенными научным изданиям в соответствии с п. 2, и выводит на экран монитора перечень профильных изданий, имеющих совпадающие ключевые слова.

3.6. Алгоритм поиска профильных изданий

При проведении поиска профильных изданий для публикации научного произведения производятся предварительный анализ ключевых слов, используемых авторами опубликованных произведений в конкретных изданиях, и сравнение их с имеющейся нормализованной базой ключевых слов с целью выбора соответствующих профилю изданий нормализованных ключевых слов и размещения их на сайтах изданий и в специальной поисковой системе. Осуществляется следующий алгоритм действий:

1. Производится сбор данных по ключевым словам, используемым авторами опубликованных научных произведений в научных изданиях, и создаётся база данных ключевых слов по каждому изданию на основе совпадений ключевых слов с ключевыми словами базы данных, созданной на основе рубрик УДК (База 1).

2. Создание нормализованной базы ключевых слов осуществляется путём использования в создаваемой базе в качестве ключевых слов соответствующих номеров и названий рубрик УДК [18]. В данном случае создаваемая нормализованная база ключевых слов – это совокупность названий рубрик УДК (База 2).

3. С помощью компьютерной программы производится лингвистический анализ текста аннотации научного издания (как правило, размещается на интернет-сайте научного издания) интеллектуальной системой (интеллектуальный поиск) с целью интеллектуального анализа слов и словосочетаний текста аннотации научного издания и сравнения их с ключевыми словами из нормализованной базы ключевых слов, и программа сохраняет полученную информацию в виде перечня совпавших ключевых слов, на основе которых создаётся база данных (База 3).

4. Компьютерная программа путём сравнения ключевых слов Базы 1, Базы 3 с Базой 2 определяет совпадающие ключевые слова, которые присваиваются научному изданию и размещаются в специальной поисковой системе научных изданий для осуществления поиска профильных научных изданий для публикации научных произведений.

4. Заключение

Предлагаемый способ библиографического поиска с использованием автоматизированной системы относится к библиографическим поисковым системам и средствам, а также к системам и средствам поиска специализированных изданий для публикации научных работ. Представленный в статье способ библиографического поиска основан на комплексном использовании следующих современных технологий:

методам поиска информации в сети Интернет,

технологиях интеллектуального анализа текстов с целью выделения ключевых слов,

интеллектуальных системах управления базами данных.

Использование библиографического поиска приводит к оптимизации времени и повышению информативности результатов библиографического поиска – это достигается использованием в качестве критерия поиска ключевых слов из нормализованной базы данных, основанной на УДК. Ключевые слова из нормализованной базы данных присваиваются исследовательским работам и научным публикациям автором или интеллектуальной системой лингвистического анализа текстов научных произведений. Поиск научных статей и специализированных изданий для публикации научных статей ведётся по нормализованным ключевым словам, присвоенным научным статьям и научным изданиям (журналам, сборникам докладов конференций и т. д.). Исследование представляет собой социальный вклад, демонстрируя важность разработки и использования лучших современных практик, процессов и стратегий библиографического поиска, которые являются важными участниками развития науки в современном обществе. При этом библиографическая информация и в социальном плане является одним из методов решения проблемы нахождения связи между личностным и общественным знанием. Реализация результатов исследования в соответствии с представленными алгоритмами заключается в создании структуры специального общедоступного интернет-поискового библиографического ресурса.

Часть информации по предлагаемой в статье системе библиографического поиска представлена в работе [17].

Список источников

1. **Нещерет М. Ю.** Эволюция теоретических и методологических представлений о библиографическом поиске в отечественном библиографоведении : дис. ... канд. пед. наук. Москва, 2008.
2. **Соколов А. В.** Автоматизация библиографического поиска. Москва : Книга, 1981. 167 с.
3. **Лахути Д. Г.** Проблемы интеллектуализации информационно-поисковых систем : дис. ... д-ра техн. наук : 05.13.17 / Рос. гос. гуманитарный ун-т. Москва, 1999.
4. **Тимофеева Е. Г., Хромов А. А., Тимофеев Г. С., Силкин В. В.** Способ поиска информации в политематических массивах неструктурированных текстов. Патент RU 2409849 С2. Номер заявки: 2008130492/08. Дата регистрации: 24.07.2008. Дата публикации: 20.01.2011.
5. **Рогачёв И. П.** Способ двухуровневого поиска информации в предварительно преобразованном структурированном массиве данных. Патент RU 2571406 С1. Номер заявки: 2014126198/08. Дата регистрации: 27.06.2014. Дата публикации: 20.12.2015.
6. **Циликов И. С.** Способ поиска информации в массиве текстов. Патент RU 2392660 С2. Номер заявки: 2008114801/09. Дата регистрации: 15.04.2008. Дата публикации: 20.06.2010.
7. **Лапшин С. В., Лапшина Д. И.** Способ структурирования результатов поиска по текстам, содержащим информацию о научной и исследовательской деятельности. Патент RU 2656469. Номер заявки: 2017105652. Дата регистрации: 20.02.2017. Дата публикации: 05.06.2018.
8. **Клинцов В. П., Селедкин В. А.** Способ автоматической итеративной кластеризации электронных документов по семантической близости, способ поиска в совокупности кластеризованных по семантической близости документов и машиночитаемые носители. Патент RU 2556425 С1. Номер заявки: 2014105486/08. Дата регистрации: 14.02.2014. Дата публикации: 10.07.2015.
9. **Ефимова Т. В.** Лингвистический анализ и формальное представление содержания нарративного текста : дис. ... канд. филол. наук. Воронеж, 2004.
10. **Соколова Е. В., Митрофанова О. А.** Автоматическое извлечение ключевых слов и словосочетаний из русскоязычных текстов с помощью алгоритма KEA // Труды объединённой научной конференции «Интернет и современное общество». 2017. № 1. С. 157–165.
11. **Красавина В. Д., Мирзагитова А. Р.** Оптимизация поиска в системе Lead-Scanner с помощью автоматического выделения ключевых слов и словосочетаний // Труды международной конференции «Корпусная лингвистика – 2015». Санкт-Петербург, 2015. С. 296–306.
12. **Воронина И. Е., Кретов А. А., Попова И. В., Дудкина Л. В.** Функциональный подход к выделению ключевых слов: методика и реализация. компьютерная лингвистика и обработка естественного языка // ВЕСТНИК ВГУ. Серия: Системный анализ и информационные технологии, 2009. № 1. С. 68–72.

13. **Коровин А. М.** Интеллектуальные системы: текст лекций. Челябинск : Издательский центр ЮУрГУ, 2015.
14. **Москвитина Т. Н.** Методы выделения ключевых слов при реферировании научного текста // Вестник Томского государственного педагогического университета. 2018. № 8 (197). С. 45–50.
15. **Воронина И. Е. и др.** Функциональный подход к выделению ключевых слов: методика и реализация // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2009. № 1. С. 68–72.
16. **Бондарчук Д. В.** Алгоритмы интеллектуального поиска на основе метода категориальных векторов : дис. ... канд. физ.-мат. наук. Екатеринбург : УрГУПС, 2016.
17. **Воруаев А. А.** (2024). Development of intelligent system of global bibliographic search // Journal of Librarianship and Information Science, 0 (0). URL: <https://doi.org/10.1177/09610006231223509>.
18. **UDC: Universal Decimal Classification.** Standard edition. 2 volumes. London : BSI, 2005. Vol. 1 – Systematic tables, ISBN 0 580 45469 X. Vol. 2 – Alphabetical Index, ISBN 0 580 45470 3. UDC Online is the complete standard up-to-date version of the Universal Decimal Classification scheme. URL: <https://udc-hub.com/index.php>.
19. **Попова Л. А.** Ключевые слова современности: проблема термина // Ученые записки Петрозаводского государственного университета. 2017. № 5 (166). URL: <https://cyberleninka.ru/article/n/klyuchevye-slova-sovremennosti-problema-termina> (дата обращения: 27.04.2025).

References

1. **Neshcheret M. Yu.** Evolution of theoretical and methodological ideas about bibliographic search in Russian bibliography : dissertation by M. Yu. Neshcheret, 2008.
2. **Sokolov A. V.** Automation of bibliographic search. Moscow : Kniga, 1981. 167 p.
3. **Lahuti D. G.** Problems of intellectualization of information search systems : dissertation of Doctor of Technical Sciences: 05.13.17 / Russian State University of Humanities. Moscow, 1999.
4. **Timofeeva E. G., Khromov A. A., Timofeev G. S., Silkin V. V.** A method for searching information in polythematic arrays of unstructured texts. Patent RU 2409849 C2. Application number: 2008130492/08. Registration date: 07/24/2008. Date of publication: 01.20.2011.

5. **Rogachev I. P.** Method of two-level information search in a pre-transformed structured data array. Patent RU 2571406 C1. Application number: 2014126198/08. Registration date: 06.27.2014. Date of publication: 20.12.2015.
6. **Tsilikov I. S.** A way to search for information in an array of texts. Patent RU 2392660 C2. Application number: 2008114801/09. Registration date: 04.15.2008. Publication date: 06.20.2010.
7. **Lapshin S. V., Lapshina D. I.** A method for structuring search results based on texts containing information about scientific and research activities. Patent RU 2656469. Application number: 2017105652. Registration date: 02.20.2017. Date of publication: 06.05.2018.
8. **Klintsov V. P., Seledkin V. A.** A method of automatic iterative clustering of electronic documents by semantic proximity, a method of searching in a set of documents clustered by semantic proximity and machine-readable media. Patent RU 2556425 C1. Application number: 2014105486/08. Registration date: 02.14.2014. Date of publication: 10.07.2015.
9. **Efimova T. V.** Linguistic analysis and formal representation of the content of a narrative text: dissertation of the Candidate of Philological Sciences. Voronezh, 2004.
10. **Sokolova E. V., Mitrofanova O. A.** Automatic extraction of keywords and phrases from Russian-language texts using the KEA algorithm // Proceedings of the joint scientific conference "The Internet and Modern Society". 2017. No. 1. Pp. 157–165.
11. **Krasavina V. D., Mirzagitova A. R.** Optimization of search in the Lead-Scanner system using automatic selection of keywords and phrases // Proceedings of the international conference "Corpus Linguistics – 2015". Saint Petersburg, 2015. Pp. 296–306.
12. **Voronina I. E., Kretov A. A., Popova I. V., Dudkina L. V.** A functional approach to keyword extraction: methodology and implementation. computational linguistics and natural language processing. VSU BULLETIN. Series: System Analysis and Information Technologies. 2009. No. 1. P. 68–72.
13. **Korovin A. M.** Intelligent systems: text of lectures. Chelyabinsk : SUSU Publishing Center. 2015.
14. **Moskvitina T. N.** Methods of keyword extraction in scientific text abstracting // Bulletin of Tomsk State Pedagogical University. 2018. № 8 (197). Pp. 45–50.
15. **Voronina I. E. and others.** A functional approach to keyword extraction: methodology and implementation // Bulletin of Voronezh State University. Series: System analysis and Information Technology. 2009. No. 1. Pp. 68–72.
16. **Bondarchuk D. V.** Algorithms of intelligent search based on the method of categorical vectors: dissertation for the degree of Ph.D. Yekaterinburg : UrGUPS, 2016.
17. **Boryaev A. A.** (2024). Development of intelligent system of global bibliographic search. Journal of Librarianship and Information Science, 0 (0). URL: <https://doi.org/10.1177/09610006231223509>.
18. **UDC: Universal Decimal Classification.** Standard edition. 2 volumes. London : BSI, 2005. Vol. 1 – Systematic tables, ISBN 0 580 45469 X. Vol. 2 – Alphabetical Index, ISBN 0 580 45470 3. UDC Online is the complete standard up-to-date version of the Universal Decimal Classification scheme. URL: <https://udc-hub.com/index.php>.

19. **Попова Л. А.** Keywords of modernity: the problem of the term // Scientific Notes of Petrozavodsk State University. 2017. No. 5 (166). URL: <https://cyberleninka.ru/article/n/klyuchevye-slova-sovremennosti-problema-termina> (date of request: 04.27.2025).

Информация об авторе / Author

Боряев Александр Александрович –
канд. техн. наук, доцент, доцент
Балтийского государственного
технического университета
«ВОЕНМЕХ» им. Д. Ф. Устинова,
Санкт-Петербург, Российская
Федерация
sasa1953@yandex.ru

Alexander A. Boryaev – Cand. Sc.
(Engineering), Assistant Professor,
“VOENMEH” D. Ustinov Baltic State
Technical University, St. Petersburg,
Russian Federation
sasa1953@yandex.ru