

УДК 004.413.2

ПРОЕКТИРОВАНИЕ АВТОМАТИЗИРОВАННОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ КЛАСТЕРИЗАЦИИ РЕЗУЛЬТАТОВ ИНТЕЛЛЕКТУАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

Минкина В.А. (МГТУ СТАНКИН)

Научный руководитель – кандидат технических наук, доцент Бычкова Н.А. (МГТУ СТАНКИН)

Введение

В настоящее время наблюдается экспоненциальный рост объема мировой патентной документации и усиление глобальной конкуренции в сфере инноваций, в результате чего возникает потребность в регулярном патентно-информационном поиске для подтверждения патентоспособности и патентной чистоты новых изобретений. Несмотря на это, в существующих специализированных поисковых системах сохраняется проблема выдачи релевантных патентных документов. Проблема обусловлена, как правило, тем, что для поиска патентных документов используется логический поиск. Это простой и легкий в реализации метод, который имеет некоторые ограничения, такие как неспособность обрабатывать синонимы, многозначность и контекст [1]. Ключевые проблемы существующих систем кластеризации результатов интеллектуальной деятельности (РИД) заключаются в их неспособности к интегрированной обработке разнородных данных (текстов, метаданных, классификаторов), слабом использовании семантического анализа; вычислительной неэффективности систем на больших массивах информации [2], а также в отсутствии прозрачных и интерпретируемых результатов, что затрудняет их практическое использование. Решением данной проблемы может стать разработка автоматизированной аналитической системы кластеризации РИД.

Основная часть

Реализация поставленной задачи сводится к разработке системы в виде совокупности программных модулей, работа которых строится на основе средств фильтрации, поиска в базах данных патентных документов и кластеризации. В качестве входных данных используются текст реферата и метаданные заявки на объект патентного права. Текст подлежит обработке – с применением специализированных средств из него извлекаются признаки (ключевые слова). Технологии поиска по ключевым словам широко известны и применяются в большинстве поисковых систем [3]. Далее извлеченные признаки используются для определения класса международной патентной классификации (МПК). Затем для формирования поисковой выдачи патентных документов из специализированных баз данных реализуется фильтрация: с использованием ключевых слов выделяются релевантные патентные документы. В сформированный датасет патентных документов включаются входные данные. После этого для представления результатов фильтрации в однородных категориях используется кластерный анализ на основе МПК и метаданных: он применяется к датасету и сортирует его на кластеры, так что в результате один из кластеров содержит входные данные. Таким образом, на выходе пользователь получает список патентных документов, релевантных входным данным. Кластеризация позволяет автоматически выдавать однородные группы документов, радикально ускоряя процесс, за счет выявления скрытых закономерностей, которые не могут быть обнаружены при поиске с использованием ключевых слов и рубрикаторов.

Выводы

В результате реализации разработанной автоматизированной аналитической системы кластеризации результатов интеллектуальной деятельности достигается эффективная обработка, группировка и анализ результатов интеллектуальной деятельности с использованием методов кластерного анализа, которая позволяет в итоге обнаружить релевантные патентные документы.

Литература

1. Курейчик В. В., Герасименко П. С. Основные подходы к извлечению текстовой информации (обзор) //Известия Южного федерального университета. Технические науки. – 2024. – № 4 (240). – С. 6-14.
2. Алилов Р. Г., Абиева Э. Р., Зубов Д. И. Особенности взаимодействия с информационно-поисковыми сервисами для начинающих пользователей: возможности и ограничения //Вестник ФИПС. – 2025. – Т. 4. – № 1 (11). – С. 62-75.
3. Андреев С.Г. Способ кластеризации результатов поиска в зависимости от семантики: пат. РФ № 2564629. МПК G06F 17/30, G06F 17/27; патентообладатель ООО «Аби ИнфоПоиск». – Заявл. № 2014112240/08, 31.03.2014. – Оpubл. 10.10.2015. – 41 с.