

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ НА ПРИМЕРЕ ПАТЕНТНОЙ ДОКУМЕНТАЦИИ

Д.П. Зрелова¹

¹Объединенный институт ядерных исследований, г. Дубна

тел.: +7 (496) 216 73 52, e-mail: zrelova@jinr.ru

Ключевые слова: патентный анализ, обработка естественного языка (NLP), нейросетевые модели, машинное обучение, семантический анализ, прогнозирование, технологические тренды.

Патентная документация представляет собой ключевой источник информации о технологических инновациях, поскольку, в отличие от научных публикаций, непосредственно отражает промышленно применимые технические решения, прошедшие экспертизу на новизну и изобретательский уровень. Будучи юридически оформленными результатами НИОКР, патенты объективно свидетельствуют о практической реализации технологических разработок. Однако содержащийся в них массив технических знаний изложен в минимально достаточном для правовой защиты объеме и соответствующим образом структурирован, что создаёт существенные сложности для системного анализа. Применение методов интеллектуального анализа данных позволяет декомпозировать эту информацию, выявляя содержательные технологические взаимосвязи и закономерности направлений инновационного развития.

Целью данного исследования является подбор и апробация оптимальных методов интеллектуального анализа патентной информации для выявления устойчивых технологических трендов

[1]. В ходе работы особое внимание уделяется статистическому анализу патентных метаданных, включающему изучение распределения патентной активности по компаниям-заявителям, индивидуальным авторам, географическим регионам и технологическим областям [2]. Параллельно разрабатывается специализированный алгоритм сбора и обработки патентных данных, предусматривающий их унифицированное представление в формате JSON с четкой структурой полей, содержащих идентификационную, техническую и правовую информацию.

Для семантического анализа патентных текстов используется комплекс NLP-методов, включающий создание тематического корпуса документов, построение терминологического словаря и обучение векторных моделей Word2Vec [3] и Doc2Vec [4] как для представления слов, так и целых документов. Проводимая работа направлена на оптимальный подбор и адаптацию существующих подходов к особенностям патентной документации.

Апробация методики проводится на репрезентативных выборках патентов в областях квантовых технологий и интеллектуальной робототехники. Выбор этих направлений обусловлен их высокой инновационной активностью и значительным объемом патентной документации, что обеспечивает достоверность получаемых результатов.

Проведенное исследование позволяет систематизировать существующие подходы к обработке патентной информации, оценить их эффективность для решения аналитических задач и сформулировать практические рекомендации по их применению. Разработанный методический инструментарий может быть использован для выявления технологических трендов, анализа конкурентной среды и определения перспективных направлений для патентования, включая недостаточно развитые в настоящий момент области («белые пятна»), имеющие, однако, высокий инновационный потенциал.

Литература

1. Попов А. В. Основы патентной аналитики [Электронный ресурс] / Федеральный институт промышленной собственности (ФИПС). — URL:

Международные научно-технологические отношения

https://www1.fips.ru/about/vptb-otdelenie-vserossiyskaya-patentno-tehnicheskaya-biblioteka/tematicheskie-vstrechi/popov_basics_patent_analytics.pdf (дата обращения: 10.03.2025).

2. Chae S., Gim J. A Study on Trend Analysis of Applicants Based on Patent Classification Systems [Electronic resource] // Information. — 2019. — Vol. 10, No. 12. — P. 364. — DOI: 10.3390/info10120364 (дата обращения: 10.03.2025).

3. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space [Electronic resource]. — arXiv, 2013. — arXiv:1301.3781v3 [cs.CL]. — URL: <https://arxiv.org/abs/1301.3781> (дата обращения: 10.03.2025).

4. Le Q., Mikolov T. Distributed Representations of Sentences and Documents [Electronic resource]. — arXiv, 2014. — arXiv:1405.4053v2 [cs.CL]. — URL: <https://arxiv.org/abs/1405.4053> (дата обращения: 10.03.2025).