


ПРОМЕЖУТОЧНЫЙ ОТЧЕТ О ВЫПОЛНЕНИИ НИР

«Совершенствование патентного поиска на основе использования методов искусственного интеллекта, искусственных нейронных сетей и машинного обучения» 1 этап

Шифр темы: **6-ИТ-2020**

Сроки проведения НИР: 1 кв. 2020 – 4 кв. 2021



Научный руководитель НИР
Советник директора
Горбунов А.В.

Ответственный исполнитель НИР
Ведущий научный сотрудник отдела
проектирования информационно-
поисковых систем, к.т.н.
Генин Б.Л.

Исполнители
Отдел 119
Отдел 31



Цель исследования

Выработка подходов к повышению качества патентного поиска, основанных на использовании методов и средств искусственного интеллекта, искусственных нейронных сетей и машинного обучения

Задачи исследования

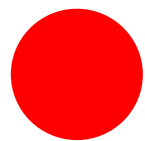
- 1) Обзор литературы по использованию методов искусственного интеллекта и по критериям оценки качества поиска
- 2) Разработка утилиты оценки качества поиска
- 3) Обновление дистрибутивного тезауруса русского языка
- 4) Создание дистрибутивного тезауруса английского языка по массивам англоязычной патентной документации



Актуальность исследования

Настоящая работа направлена на исследование актуальных проблем повышения качества патентного поиска, в частности проблем семантического поиска похожих патентных документов для целей экспертизы.





ПРОБЛЕМА: *Высокая трудоемкость информационного поиска для определения уровня техники в предметной области заявки при экспертизе заявки на изобретение по существу и необходимость высокой квалификации в области патентного поиска*

В работах ФИПС была проанализирована и доказана на примере русскоязычных массивов патентных документов перспективность и эффективность использования для повышения качества автоматического поиска дистрибутивной семантики, основанной на машинном обучении искусственных нейронных сетей и решающих деревьев. В этих работах ФИПС выявлена актуальность дифференцирования анализа качества поиска по тематическим областям техники, а также актуальность исследований по расширению применения предложенных подходов на англоязычные массивы, а в перспективе и на многоязычные массивы.





1 ЭТАП

1 кв.2020 – 4 кв.2020

Обзор литературы по использованию методов искусственного интеллекта и по критериям оценки качества поиска. Разработка утилиты оценки качества поиска. Создание и обновление дистрибутивных тезаурусов

2 ЭТАП

1 кв.2021 – 4 кв.2021

Реализация использования разработанных дистрибутивных тезаурусов в системе патентного поиска и оценка качества поиска похожих документов





Основные результаты работы:

- критерии качества поиска похожих, которые могут применяться для оценки результатов поиска с применением нейронных сетей в массивах патентных документов;
- утилита оценки качества поиска;
- обновленный дистрибутивный тезаурус русского языка;
- дистрибутивный тезаурус английского языка по массивам англоязычной патентной документации;

Результаты работы могут быть использованы в:

- При проектировании информационных систем ФИПС, разрабатываемых в рамках мероприятий Цифровой Экономики
- При разработке функциональности поиска похожих в системах патентного поиска
- При анализе качества автоматического патентного поиска



Потребители результатов исследования

Возможные пользователи: Эксперты отраслевых отделов ФИПС, Пользователи патентной информации, специалисты и разработчики новой техники и технологий в стране, Отдел проектирования информационно-поисковых систем ФИПС, Ростелеком.

★ *критерии качества поиска похожих, которые могут применяться для оценки результатов поиска с применением нейронных сетей в массивах патентных документов*

★ *Предложен критерий - Степень соответствия списка результатов автоматического поиска похожих списку документов, определяющих предшествующий уровень техники, определенных экспертизой*

★ Новизна заключается в учете специфики задачи патентного поиска для определения уровня техники и в учете особенностей патентной документации





утилита оценки качества поиска



Утилита оценки качества поиска позволяет сравнивать качество автоматического патентного поиска похожих в системах патентного поиска по предложенному критерию в режиме нормального функционирования этих систем



Новизна заключается в учете специфики задачи патентного поиска для определения уровня техники, в учете особенностей патентной документации, и в использовании предложенного критерия качества для сравнения систем



★ обновленный дистрибутивный тезаурус русского языка

★ Обновление дистрибутивного тезауруса русского языка производилось на основе обновленной (расширенной) базы данных российских патентных документов, включающей документы по 2019 год включительно, и комплекса экспериментов по оптимизации параметров.

★ Новизна обновленного дистрибутивного тезауруса русского языка заключается в том, что он построен на базе российских патентных документов с несколькими вариантами использования компонентов патентных документов с варьированием применяемых алгоритмов машинного обучения искусственной нейронной сети





По результатам исследования опубликованы:

1

Искусственный интеллект в работе патентных ведомств, Горбунов А. В., Эриванцева Т. Н., Генин Б. Л., Золкин Д. С. (Принято к публикации, журнал Форсайт)

2

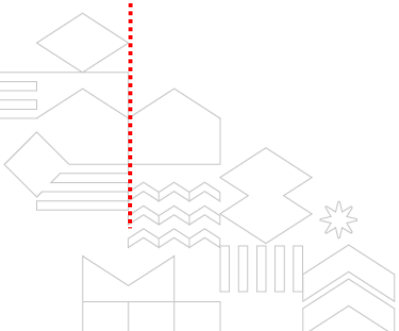
«Патентный поиск и оценки его качества», Горбунов А.В., Генин Б.Л., Золкин Д.С., Тарасов А.В., Право интеллектуальной собственности, 2020, №3 (61)



Доклады с результатами НИР на российских и международных научных мероприятиях :

1

«Искусственный интеллект в патентном поиске», XXIV Международная конференция Роспатента. «Трансформация сферы интеллектуальной собственности в современных условиях», 20-21 ОКТЯБРЯ 2020 ГОДА, МОСКВА





Важнейшим результатом НИР является научно-обоснованный по результатам более 2000 экспериментов вывод об эффективности предложенного способа использования методов и средств искусственного интеллекта для повышения качества автоматического поиска похожих патентных документов при поиске уровня техники в предметной области заявок на изобретения и полезные модели





Спасибо за внимание!

