

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

ЗАДАЧА 01

Сервис анализа патентной
активности компаний Москвы



Техническое задание «Сервис анализа патентной активности компаний Москвы»

1. Актуальность задачи

Современное развитие экономики характеризуется непрерывным увеличением роли инноваций как ключевого фактора роста. Перед городом стоит задача повышения инновационной активности компаний и исследователей.

Патентная статистика – важнейший инструмент для оценки существующего уровня научно-исследовательской активности и выявления тенденций развития технологического предпринимательства.

Однако система регистрации российских патентов не учитывает уникальный идентификатор организаций – идентификационный номер налогоплательщика (ИНН). Для повышения эффективности внедрения городских программ поддержки инноваторов требуется обогащение базы данных патентной информации нормализованными сведениями об авторах и правообладателях. Целевой потребитель сервиса - органы власти для мониторинга патентной активности и компании для мониторинга правового статуса портфелей интеллектуальных прав.

2. Описание задачи

Разработайте сервис, который позволит загружать и парсить первичные данные о патентах, а также формировать на основе этих данных перечни патентов, размеченных по авторам и правообладателям.

Ключевой частью сервиса станет алгоритм, который позволит разметить патенты по ИНН и ОГРН юридических лиц и индивидуальных предпринимателей, а также ID физических лиц.

Дополнительными возможностями сервиса и алгоритма могут являться:

2.1. Категоризация патентов (Приложение).

2.2. Построение диаграмм на основе данных статистики патентования и др.

2.3 Построение патентных ландшафтов¹ по указанным кластерам: виды компаний, регион и т.д.

3. Требования к решению

3.1. Бизнес-требования:

1. Возможность загрузки и парсинга excel-документов с данными о патентах по заданному шаблону. Шаблон будет представлен в приложении к ТЗ.

2. Возможность обогащения excel-документов данными об уникальных идентификаторах (обязательно), а также иными данными (опционально - выявление пересекающихся приоритетов, технологий, категоризация патентов по кластерам компаний др.)

3. Возможность формирования, хранения и выгрузки общей базы данных о патентах или правообладателях на основе объединения уникальных данных из разных выборок (пример: в систему загружено 2 документа, содержащих данные о 200 патентах. Общее количество уникальных компаний, собранных в общей базе из двух выборок: 50. Общее количество уникальных патентов в общей базе из двух выборок: 150.)

4. Возможность формирования и вывода ключевых статистических данных по общей базе данных, а также о результатах разметки отдельного файла в web-интерфейсе (общее количество патентов, количество размеченных патентов, общее количество правообладателей, количество по типам правообладателей: ЮЛ, ФЛ, ИП, по категориям правообладателей: вуз, НИИ, ИТ-компания и т.д.).

5. Возможность формирования и вывода дашборда с ключевыми статистическими данными по общей базе данных, а также

¹ Патентный ландшафт - это отчет о состоянии патентной активности на определенной территории(стране, регионе..) по определенной технологии(категории). Зачастую аналитические показатели накладываются на карту.

Вариантом исполнения может быть создание интерактивного дашборда - несколько ключевых диаграмм из п. 2.2. + страницы, состоящей из карты с "температурным показателем" по распределению патентной активности по выбранной категории патентов из п. 2.1.

о результатах разметки отдельного файла в web-интерфейсе (Приложение). **Опционально**

6. По произвольной выборке документов построить карту на основе анализа

- терминов, наиболее часто встречающихся в документах выборки, их «близости» в документе, «жесткости» связей между ними. Пример

(рис. 7, стр. 14) :
https://ised-isde.canada.ca/site/canadian-intellectual-property-office/sites/default/files/attachments/2022/CIPO-Patents-in-Space-Report_e.pdf

- игроков на рынке, патентующих свои разработки. Пример (рис. 15 стр. 46):
<https://www.wipo.int/edocs/pubdocs/en/wipo-pub-1075-23-en-covid-19-vaccines-and-therapeutics.pdf>

Опционально

3.2. Техническая реализация

Модуль должен состоять из отдельных компонентов:

- Компонент с моделью ML
- Компонент с API REST
- Компонент хранения баз данных и результатов алгоритма
- Компонент с web-интерфейсом для работы пользователей

Обязательно использование Docker для развертывания

3.3. Требования к стеку

- СУБД PostgreSQL

4. Пользовательский путь

- Сотрудник заходит в сервис (компонент web-интерфейса)
- Загружает файл, содержащий информацию по патентам, в строгом соответствии с шаблонным файлом.
- В результате работы сервиса получает скачиваемый файл .xlsx, содержащий размеченные патенты с ИНН и ОГРН

компаний и индивидуальных предпринимателей, ID физических лиц.

- После получения размеченного файла происходит редирект на веб-интерфейс с патентной аналитикой по данным из исходного файла.

5. Требования к презентации

- Презентация предоставляется в формате pptx, pdf;
- В презентации должен быть раскрыт принцип работы сервиса;
- Презентация должна раскрывать используемые технологии,
- В презентации должно быть приведено обоснование, почему выбранные технологии и принцип работы сервиса позволяют эффективно решать поставленные бизнес-задачи

6. Источники данных

- Реестр патентов на изобретения, полезные модели и промышленные образцы, включающий данные о наименовании организации, почтовом адресе, наименовании и описании патента (информация предоставляется в файле и по ссылке <https://www1.fips.ru/registers-web/>);
- Реестр юридических лиц и индивидуальных предпринимателей, включающий данные о наименовании организаций и их ИНН и ОГРН, а также реестр физических лиц-получателей поддержки (информация предоставляется в файле)
- Поисковая система ФИПС с возможностью верификации уточненных сведений по патентам (информация предоставляется в файле и по ссылке <https://www.fips.ru/iiss/>)

7. Требования к сдаче решений на платформе

- Ссылка на репозиторий с кодом;
- Ссылка на презентацию (требования указаны в шаблоне);
- Ссылка (адрес) на прототип для проверки выполненной работы;
- Ссылка на дополнительную сопроводительную документацию в случае наличия (.doc/.pdf).

8. Критерии оценки

- 1) Подход коллектива к решению задачи
 - а) Оригинальные предложения по решению задачи;
 - б) Реализация оригинального предложения в сервисе.
- 2) Техническая проработка решения
 - а) Работоспособность решения;
 - б) Поддерживаемость решения;
 - в) Качество кода;
 - д) Отчуждаемость сервиса (сервис должен иметь возможность работать самостоятельно без зависимости от конкретной инфраструктуры).
- 3) Соответствие решения поставленной задаче
 - а) Полнота реализации механизма разметки патентов по уникальным идентификаторам.
 - б) Скорость работы сервиса
 - в) Доля размеченных патентов
 - д) Доля ошибок в разметке патентов (наилучшее соответствие по коэффициенту корреляции Мэтьюса)
- 4) Эффективность решения в рамках поставленной задачи
 - а) Наличие дополнительной функциональности в сервисе (оценивается при доле размеченных патентов более 50% от исходного массива, оценивается понимание бизнес-составляющей сервиса, выбора и реализации представления графической информации)
 - і) категоризация патентов (п. 2.1)

- ii) дашборды (п. 2.2)
 - iii) кластеры (п. 2.3)
- 5) Выступление коллектива на питч-сессии (только для финальной экспертизы)

ПРИЛОЖЕНИЕ

Дополнительная категоризация исходных данных при разделении по правообладателям

Следует выделять следующие категории:

1. Виды объектов, охраняемых патентами:
 - 1) изобретение,
 - 2) полезная модель,
 - 3) промышленный образец,
2. Для изобретений и полезных моделей:
 - 1) индекс класса (из «Открытых реестров²»: поле 51(после аббревиатуры «МПК») или поле 52(после аббревиатуры «СПК»), 3 первых символа: буква и 2 цифры)
 - 2) индекс подкласса (из «Открытых реестров»: поле 51(после аббревиатуры «МПК») или поле 52(после аббревиатуры «СПК»), 4 первых символа: буква, 2 цифры, буква).
3. Для промышленных образцов:
 - 1) класс (из «Открытых реестров»: поле 51, после аббревиатуры «МКПО^{номер версии}», первые 2 цифры перед дефисом)
 - 2) подкласс (из «Открытых реестров»: поле 51, после аббревиатуры «МКПО^{номер версии}», 2 цифры, дефис, 2 цифры)
4. Применительно ко всем ОИС – разделение по количеству авторов, представленных в одной записи из реестра на категории:
 - 1) один автор,
 - 2) от 2 до 5 авторов,
 - 3) более 5 авторов.

² Раздел «Источники данных», первый пункт

5. Применительно ко всем ОИС – разделение по дате регистрации и региону для отслеживания динамики развития.