Вычислительный центр
Дальневосточного отделения Российской академии наук

Автоматизированное формирование аналитических отчётов по данным спутникового мониторинга на основе больших языковых моделей

Шамурина Анна Игоревна, Кожевникова Татьяна Владимировна, Авершин Артем Андреевич
sh-a-ig@mail.ru

Постановка задачи

ДАННЫЕ

ИС геомониторинга, разработанная в ВЦ ДВО РАН, хранит:

- разновременные космические снимки и их метаданные
- контуры нарушенных земель после тематической обработки
- ряды площадей объектов на даты съёмки

Хранение: PostgreSQL с расширением PostGIS

ВОПРОСЫ СПЕЦИАЛИСТА

«Как менялась площадь нарушенных земель по районам?»

«Какие снимки пригодны по облачности?»

«Согласуются ли результаты разных съёмочных систем?»

Каждый вопрос требует SQL-запроса с пространственными функциями

ЦЕЛЬ

Аналитические справки по запросу на естественном языке

- корректный запрос к базе данных
- неизменность хранимых данных
- все числа отчёта — из базы данных

Метод: генерация с извлечением данных

1

Вопрос

на естественном языке

→

2 · извлечение

Контекст

схема базы и правила для данных ДЗЗ

→

3 · генерация

SQL

DeepSeek строит запрос

→

4 · извлечение

Выполнение

проверка, только чтение

→

5 · генерация

Отчёт

текст по готовым числам

ПОЧЕМУ DEEPSEEK

Генерация кода и SQL. Модели DeepSeek сильны в задачах программирования, включая построение запросов

Низкая стоимость. Цена за токен ниже, чем у сопоставимых облачных моделей

Интерфейс OpenAI и открытые веса. Замена на локальную модель без изменения программы

ОГРАНИЧЕНИЯ

Платный сервис. Каждый запрос оплачивается по числу токенов

Не локальный. Схема базы, вопрос и строки результата передаются на внешний сервер

Снимки и контуры остаются в базе; для закрытых данных — локальная модель

Шамурина А.И., Кожевникова Т.В., Авершин А.А. · ВЦ ДВО РАН · sh-a-ig@mail.ru

Пример: от вопроса к отчёту

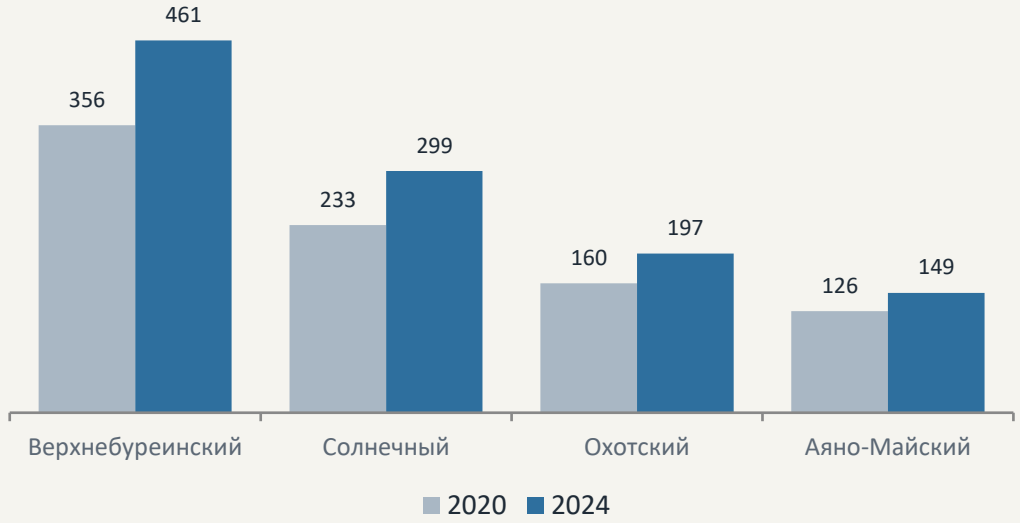
ВОПРОС

«Как изменилась площадь нарушенных земель по районам с 2020 по 2024 год?»

SQL-ЗАПРОС (ФРАГМЕНТ)

```
WITH last_obs AS (
  SELECT DISTINCT ON (object_id, EXTRACT(YEAR FROM obs_date))
         object_id, area_ha,
         EXTRACT(YEAR FROM obs_date)::int AS year
  FROM observations
  ORDER BY object_id, year, obs_date DESC)
SELECT m.district, l.year, SUM(l.area_ha)
FROM last_obs l
JOIN mining_objects m ON m.id = l.object_id
WHERE l.year IN (2020, 2024)
GROUP BY m.district, l.year
```

Площадь нарушенных земель, га (модельные данные)



ВЫВОД В ОТЧЁТЕ (ПРИМЕР)

Площадь выросла во всех четырёх районах: быстрее всего в Верхнебуреинском (на 29,5 процента), медленнее всего в Аяно-Майском (на 18,2 процента)

Результаты апробации

12 объектов

4 района

56 снимков

2 съёмочные системы

7 лет, 2019–2025

Модельная база данных со структурой ИС геомониторинга, разработанной в ВЦ ДВО РАН

Учёт специфики ДЗЗ

При нескольких снимках объекта за год суммирование наблюдений кратно завышает площадь. Явное правило — брать последнее наблюдение года — обеспечивает корректный расчёт

Облачность объяснена

23 снимка не использованы в дешифрировании — у всех облачность выше порога 30 процентов

SQL рядом с отчётом

Самоисправление устраняет ошибки СУБД, но не логические — поэтому запрос предъявляется вместе с отчётом