**Промежуточная аттестация 3**

**Проект: Вопросно-ответный поиск в промышленной отрасли**

**Исходные данные:**

Набор технологий:

* Редактор кода
* PostgreSQL сервер
* Клиент PostgreSQL

**Проект: Вопросно-ответный поиск в промышленной отрасли**

Так у автора настоящей работы очень близкая задача по чат-боту, обученному на вопросно-ответной модели, в рамках итоговой аттестации, данную задачу постарались рассмотреть также с точки зрения практического применения в работе автора.

**Постановка задачи:**

Пользователь работает с большим объемом специализированных терминов из различных нормативных правовых актов и стандартов. Пользователю требуется решение, при котором задавая вопрос системе, он будет видеть необходимые термины и их источники утверждения (нормативные правовые акты и стандарты).

**Реализация:**

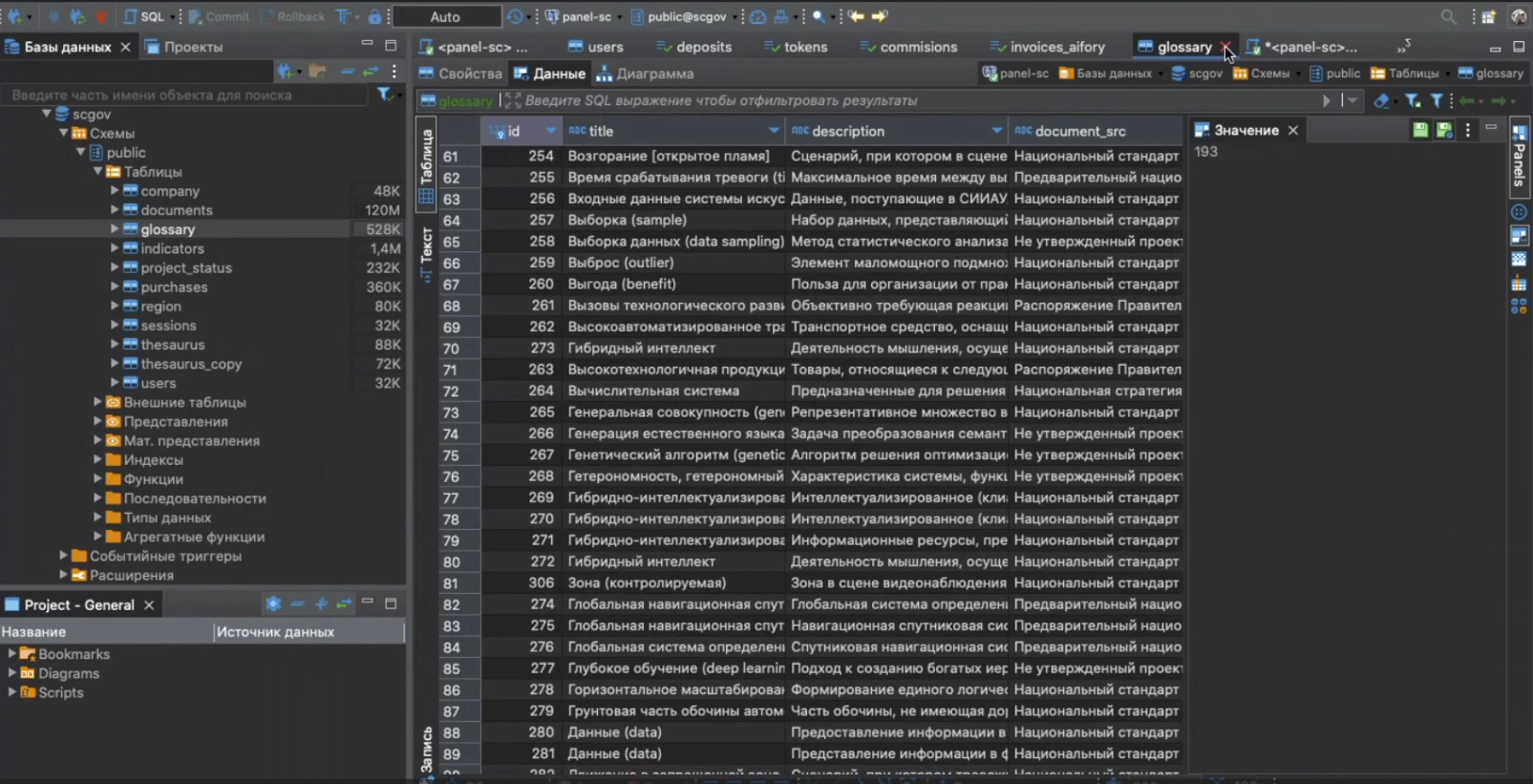
Для реализации данной задачи была выбрана программа DBeaver, так как она работает со множеством современных СУБД, например MySQL и PostgreSQL, а также является бесплатной.

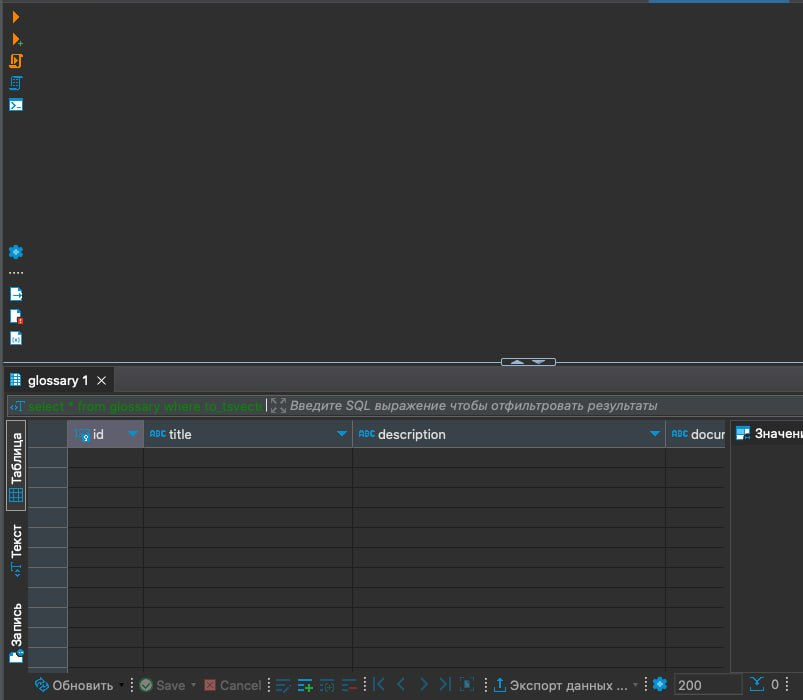
Для реализации задачи был подготовлен набор данных:



Реализация решения была основана на возможностях полнотекстового поиска PostgreSQL: <https://www.postgresql.org/docs/current/textsearch-controls.html>

Данные были загружены в систему:

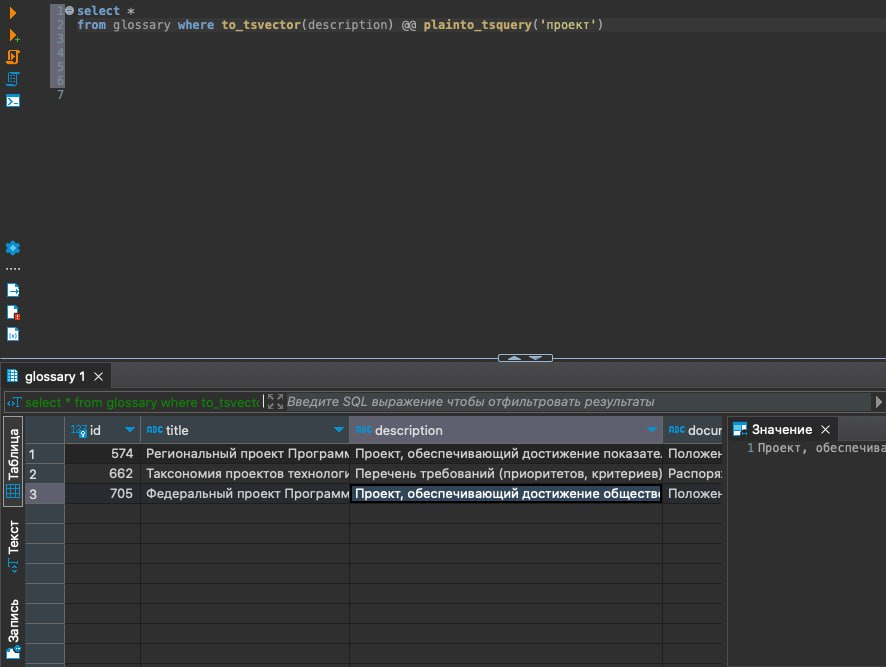




Далее был настроен запрос для пользователя:

select \*

from glossary where to\_tsvector(description) @@ plainto\_tsquery('проект')



В ответ на вопрос показателя показать контент, содержащий упоминание «проект», система выдает результаты поиска в ответе в таблице внизу экрана.

При этом система не ошибается и подсвечивает показатели скорости работы:



Преимущества использования полнотекстового поиска PostgreSQL:

1. Простота использования (для меня, как для не подготовленного пользователя, работа с PostgreSQL оказалась несравнимо проще, чем с Jupiter notebook.
2. Многие поисковые движки быстрые, но:

* не могут индексировать все документы
* не имеют доступа ко всем атрибутам
* требуют поддержки администратором
* часто должны быть сертифицированы / лицензируемы
* не обеспечивают мгновенный поиск, т. к. нужно время для переиндексации новых данных
* не обеспечивают согласованность данных, например, данные могут быть удалены из базы данных

Разработчики полнотекстового поиска PostgreSQL утверждают, что с их системой такого не бывает. И за время тестирования, я также не выявила подобную проблематику.

Так получилось, что финальную аттестацию я делала до Аттестации 3. И потратила достаточно много времени на настройку обработки набора данных.

При работе с полнотекстовым поиском PostgreSQL я быстро загрузила набор данных. И система единожды его обработала – во время поиска ответа на пользовательский запрос дополнительной обработки уже не было.

Кроме того, я обнаружила, что в полнотекстовом поиске PostgreSQL не учитываются в поиске стоп-слова. При работе с Jupiter notebook приходилось вручную подключать библиотеки и учить систему исключать стоп-слова.