

ЗАНЯТИЕ 2

ОСНОВЫ SQL



Александр Джумурат, онлайн кинотеатр «ivi.ru»



adzhumurat@ivi.ru



adzhumurat

Джумурат
Александр

Разработчик рекомендательной системы,
онлайн-кинотеатр ivi.ru

Цель занятия:

Применение SQL. Простые запросы. Превращаем метрики в SQL. Сложные запросы: JOIN.

В КОНЦЕ ЗАНЯТИЯ СМОЖЕТЕ:

- Строить простые выборки из БД
- Формировать сложные выборки: JOIN
- Фильтрация выборки: WHERE, LIKE и т.д.

О ЧЁМ ПОГОВОРИМ
И ЧТО СДЕЛАЕМ

1. Простые выборки
2. Сложные выборки с условиями.
3. Агрегация данных.

Часть 1

Простые выборки

Простые выборки

Выборки из БД (реляционных) выполняются с помощью SQL, основной оператор - **SELECT**.

Для фильтрации результатов **SELECT** применяем **WHERE** (условный оператор).

Условия фильтрации бывают разные: **=**, **<>**, **>**; **IN**, **LIKE** и т. д.

Набор данных для экспериментов

Данные - рейтинги пользователей из [соревнования Kaggle](#)

Набор данных содержит несколько csv и json файлов с разнообразными данными по фильмам и информацию о рейтингах, которую ставили пользователи

Пример

Простые выборки

Объединение выборок

Простые выборки можно сделать более гибкими, если соединить в одном запросе несколько выборок

UNION - объединяем все строки

EXCEPT - выключаем из запроса-1 все строки, которые есть в запросе-2

INTERSECT - только строки, которые есть в обоих запросах

Примеры UNION, EXCEPT, INTERSECT

Часть 2

Сложные выборки: JOIN

Сложные выборки: JOIN

Для соединения двух таблиц (отношений) используется оператор JOIN

Основные виды: INNER, LEFT (RIGHT), OUTER

JOIN позволяет выбрать поля из нескольких таблиц, связав таблицы по значению поля или набора полей.

Пример

Сложные выборки: JOIN

Часть 3

Расчёт фичей:
агрегирующие функции

Агрегация данных: базовые статистики

До сих пор мы занимались простыми выборками из БД.

Для создания ML моделей требуется создавать на базе выборок фичи.

Агрегация данных: GROUP BY

GROUP BY используем, когда нужно статистику по данным в различных срезах (а не по всем данным сразу)

Агрегация данных: **HAVING**

HAVING - оператор, который помогает проводить фильтрацию по полям с агрегирующими функциями

Пример

Агрегация данных

Тема 5

Подзапросы и СТЕ

Прочие темы: CTE. Подзапросы.

CTE - способ вынести часть логики запроса в отдельное выражение, с которым можно работать как с таблицей

Подзапросы - приём в котором результаты одного **SELECT** переиспользуются в другом выражении.

Пример

Подзапросы и СТЕ

ДОМАШНЕЕ ЗАДАНИЕ

[Ссылка на github](#)

ПОЛЕЗНЫЕ МАТЕРИАЛЫ

ПОЛЕЗНЫЕ МАТЕРИАЛЫ

1. [SQL : базовые конструкции](#)
2. [SQL exercises](#)



НЕТОЛОГИЯ
групп

Спасибо за внимание!

Джумурат Александр



adzhumurat@ivi.ru

В