

# Темы для ИДЗ

Из чего складывается сложность:

1. Доступность датасета
2. Необходимость использовать несколько датасетов
3. Необходимость использовать СУБД не совсем по назначению
4. Алгоритмы

Темы в порядке возрастания субъективной сложности. Поле "Комментарий" содержит отправную точку для формулировки окончательного задания.

№	Название	Предлагаемая СУБД	Набор данных/API	Комментарий
24	Каталог объявлений о продаже автомобилей	Mongo	Синтетические данные	Задача - сервис, позволяющий размещать объявления о покупке или продаже автомобилей, связывать продавцов и покупателей. Необходимые (но не достаточные фичи) - аккаунты продавцов и покупателей, рейтинги и отзывы, страница подробных данных об автомобиле (пробег, номера двигателя ... , марка машины, год, фото ....). Пользователи - администраторы, клиенты.
24	Сервис редактирования и автоматического составления расписания в ВУзе	Mongo	Расписание ЛЭТИ	Задача - сервис, позволяющий принимать пожелания от преподавателей, принимать данные об аудиториях и их вместимости, данные учебных планов (предметы / курсы / группы), данные о студентах (сколько в какой группе человек, какие группы к каким направлениям относятся). Необходимые (но не достаточные фичи) - автоматическое составление расписания (исходя из предоставленных ограничений), визуализация и редактирование расписания, логика согласования расписания (препод высказал пожелание, админ поправил, препод одобрил), анализ составленного расписания (насколько равномерно используются аудитории, насколько студентам / преподам нужно между парами бегать в новую аудиторию). Пользователи - преподаватели, администраторы, студенты.
24	Сервис оценки земельных участков	Neo4j	Avito + OSM	Задача - подготовить веб-сервис, который реализует каталог участков, а также механику их оценки. Смысл сервиса - позволить добавлять в систему полные данные об участках координаты всех узловых точек и проводить на базе формы, окружающего ландшафта и соседей оценку, например: сколько соседских домов, как далеко водорем / остановка / магазин, насколько участок ниже / выше соседей. Необходимые (но не достаточные) фичи - личные страницы пользователей, рейтинги и отзывы пользователей и участков, визуализация и поиск на карте, страницы отдельных участков с подробными данными.
24	Сервис "Crowd-редакция"	Neo4j	Синтетические данные + wikibooks	Задача - подготовить веб-сервис, который реализует вычитку и редактирование книг силами сообщества (crowdsourcing). Одни и те же пользователи могут быть как авторами, так и редакторами. Необходимые (но не достаточные) фичи - личные страницы пользователей, рейтинги и отзывы пользователей, статистика рецензий, списки рецензий, страницы отдельных рецензий, страницы книг.
24	Сервис поиска врачей частной практики	MongDB	Синтетические данные	Задача - подготовить веб-сервис, которые сводят пациентов и врачей, ведущих частную практику. Категории пользователей - врачи, пациенты, админы. Необходимые (но не достаточные) фичи - личные страницы врачей и пациентов, рейтинги и отзывы, статистика, страницы для финансовой стороны вопроса (реальные платежи подключать не нужно) - оплаты и статистика, карточки и анализы пациентов.
24	Телеграм-бот для вывода данных из таблиц	MongDB	Синтетические данные	Задача - сделать простого бота, который сможет решать задачу "покажи мне строку по запросу А из таблицы Б". Для определенности можно считать, что такой бот нужен для курсов МОЭВМ, чтобы скрывать от студентов всю таблицу успеваемости и показывать студенту только его собственную строку. Таблиц / студентов / преподов может быть много.
24	Телеграм-бот для напоминаний преподавателям о новых записях в таблице	MongoDB	Синтетические данные	Задача - сделать простого бота, который сможет решать задачу "пришли пользователю уведомление, если его _неким_ образом упомянули в таблице". Для определенности можно считать, что такой бот нужен для преподавателей МОЭВМ, которые ждут появления студенческих отправок в определенной таблице. Преподов и таблиц может быть много.
24	Сервис хранения данных биржевой торговли forex	MongoDB	Синтетические данные	Задача - веб приложение, которое будет хранить / обеспечивать поиск / анализ и визуализацию данных о рынке по указанным периодам / запросам / валютным парам.
24	Сервис сбора и визуализации логов Apache2	InfluxDB	Синтетические данные	Задача - создать приложение, которое агрегирует логи Apache2 в influx. <a href="https://github.com/influxdata/telegraf/blob/master/plugins/inputs/tail/README.md">https://github.com/influxdata/telegraf/blob/master/plugins/inputs/tail/README.md</a> , <a href="https://github.com/influxdata/telegraf/blob/master/plugins/parsers/grok/README.md">https://github.com/influxdata/telegraf/blob/master/plugins/parsers/grok/README.md</a> . Необходимо поддержать одновременно все файлы логов apache - access.log, error.log, other_vhosts_access.log, а также время загрузки страниц как один из элементов данных.
24	Сервис локального (на базе нескольких гитхаб-репо) антиплагиата для текстов на естественном языке	MongDB	Синтетические данные	Задача - веб приложение, которое будет выгружать, версионировать (Используя данные гита) данные из набора гитхаб-репо (кастомизируются пользователями), хранит эти данные для удобства синтаксического анализа и формирует отчеты по возможному плагиату между пользователями репо.
24	Сервис хранения экспериментов инструмента Sumo	Neo4j	<a href="https://sumo.dlr.de/docs/index.html">https://sumo.dlr.de/docs/index.html</a>	Задача - сделать простое веб-приложение для импорта / хранения / поиска / визуализации результатов экспериментов в эмуляторе.
24	Сервис хранения экспериментов инструмента pybulletdrones	MongoDB	<a href="https://github.com/utiasDSL/gym-pybullet-drones">https://github.com/utiasDSL/gym-pybullet-drones</a>	Задача - сделать простое веб-приложение для импорта / хранения / поиска / визуализации результатов экспериментов в эмуляторе.

№	Название	Предлагаемая СУБД	Набор данных/API	Комментарий
24	Сервис хранения экспериментов инструмента ripes	MongoDB	<a href="https://github.com/mortbopet/Ripes">https://github.com/mortbopet/Ripes</a>	Задача - сделать простое веб-приложение для импорта / хранения / поиска / визуализации результатов экспериментов в эмуляторе.
24	Построение маршрутов по рекам и каналам	Neo4j	OpenStreetMaps	Задача - максимально смоделировать данные о реках / каналах СПб, а также объектах инфраструктуры (причалы, пирсы, знаки для водного транспорта, мосты), обеспечить построение маршрутов и вычисление статистики
25	Юридические документы	Neo4j	Синтетические данные	Смысль проекта - предоставить интерфейс для хранения, версионирования и анализа (грубого) юридических документов на русском языке. Варианты анализа: разбиение на предложения, абзацы, идентификация пунктов по ключевым словам, поиск формулировок-обязательств / требований ....
27	ИС Справочник Музеев	Memcached	<a href="https://opendata.mkrf.ru/opendata/7705851331-museums">https://opendata.mkrf.ru/opendata/7705851331-museums</a>	Тема аналогична "(Сложно) ИС Справочник образовательных организаций Санкт-Петербурга"
28	ИС Справочник образовательных организаций Санкт-Петербурга	Memcached	<a href="https://petersburgedu.ru/institution/">https://petersburgedu.ru/institution/</a>	Цель - создать каталог, хранящий максимум информации обо всех организациях. Задачи - хранение всех доступных полей, организация поиска, агрегации, изменения данных каталога. Модель данных нужно построить так, чтобы на клиенте требовалась минимальная обработка (скорее всего, вам предстоит <u>сделать</u> очень много дублирования данных)
30	Построение маршрутов с учетом плохой парковки	Neo4j	Открытые данные OpenStreetMap	Необходимо взять в качестве демо данных квадрат карты, где находится Петроградский район и для него создать приложение, которое позволит строить маршруты из точки А в точку Б так, чтобы 1) при подъезде к точке Б нужно вести водителя не наикратчайшим путем, а кругами вдоль перспективных для парковки мест 2) при всем этом поездка по маршруту не была супердолгой

## Своя тема

- Интересный, содержательный и нужный** пользователям сценарий использования
- Должен быть понятный, достаточный и доступный **источник данных** - публичный датасет, публичное API доступа к хранилищу данных
- Должно присутствовать решение задач **хранения/анализа/представления/агрегации/импорта и экспорта** данных
- Дополнительный плюс, если есть **связь с существующим проектом** и/или бакалаврской работой
- Информация об **аналогах**

## Источники данных

Помимо ранее упомянутых датасетов, можно также использовать для своей темы:

- [Список SPARQL-endpoint](#)
- <https://github.com/blog/2298-github-data-ready-for-you-to-explore>
- <http://www.foodpedia.tk/about?lang=ru>
- <http://www.imdb.com/interfaces>
- [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)
- Выгрузка данных OSM
  - <https://www.openstreetmap.org/search?query=saint-petersburg#map=10/59.9471/30.2495>
- [Kaggle - каталог наборов данных](#)
- [Datahub - каталог наборов данных](#)

From:  
<http://se.moevm.info/> - **se.moevm.info**

Permanent link:  
[http://se.moevm.info/doku.php/staff:courses:no\\_sql\\_introduction:course\\_work:topics?rev=1692869928](http://se.moevm.info/doku.php/staff:courses:no_sql_introduction:course_work:topics?rev=1692869928)

Last update: **2023/08/24 10:38**