

## Лабораторная работа №3

Разработать приложение под Apache Spark в режиме standalone (работа на одной машине без полноценного Hadoop окружения). Провести разбор и анализ литературного произведения (по своему варианту):

1. Очистка текста;
2. WordCount по тексту;
3. Топ50 наиболее и наименее часто встречающихся слов;
4. Стемминг для слов;
5. Топ50 наиболее и наименее часто встречающихся слов по стемминг.

## Лабораторная работа №4

Провести более детальный анализ с применением техник анализа естественного языка (natural language processing - NLP):

1. Выделение сущностей из текста произведения;
2. Сущностями могут выступать действующие лица (места, временные метки и т.д.);
3. Можно использовать самописные правила, библиотеки NLP, библиотеки ML и т.д.;
4. Частотный анализ по сущностям (Топ50, АнтиТоп50).

## Варианты

**Афанасьев Алексей**

4: Михаил Александрович Шолохов – Тихий Дон

**Бармёнков Павел**

5: Александр Иванович Герцен – Былое и думы

**Бычков Андрей**

11: Алексей Николаевич Толстой – Гиперболоид инженера Гарина

**Загородняя Ирина**

3: Алексей Николаевич Толстой – Эмигранты

**Залалов Наиль**

8: Михаил Афанасьевич Булгаков – Мастер и Маргарита

**Зверяко Владимир**

19: Иван Сергеевич Тургенев – Записки охотника

**Земниэкс Глеб**

10: Николай Васильевич Гоголь – Мертвые души

**Казорин Сергей**

13: Александр Иванович Куприн – Поединок

**Кириленко Полина**

16: Михаил Евграфович Салтыков-Щедрин – Господа Головлевы

**Колбина Анастасия**

12: Максим Горький – Фома Гордеев

**Кочергин Сергей**

6: Максим Горький – Мать

**Макеенков Александр**

15: Лев Николаевич Толстой – Война и мир

**Радаев Александр**

17: Федор Михайлович Достоевский – Преступление и наказание

**Рябов Петр**

9: Лев Николаевич Толстой – Воскресение

**Рытов Руслан**

14: Иван Александрович Гончаров – Обломов

**Телегин Егор**

1: Лев Николаевич Толстой – Анна Каренина

**Тимошина Марина**

20: Николай Васильевич Гоголь – Вечера на хуторе близ Диканьки

**Файзрахманов Айдар**

2: Федор Михайлович Достоевский – Бесы

**Шершнёв Роман**

7: Иван Сергеевич Тургенев – Отцы и дети

**Шитов Дмитрий**

18: Алексей Николаевич Толстой – Хожение по мукам