

# САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математико-Механический факультет  
Кафедра информационно аналитических систем

## Суммаризация групп в социальных сетях

Дипломная работа студента 645 группы  
Чурикова Никиты Сергеевича

*Научный руководитель:*

к.ф. - м.н., доцент ГРАФЕЕВА Н. Г.

*Рецензент:*

Руководитель департамента вычислительной биологии  
ЯКОВЛЕВ П. А.

*Заведующий кафедрой:*

к.ф. - м.н., доцент МИХАЙЛОВА Е. Г.

Санкт-Петербург  
2019 г.

# Содержание

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Аннотация</b>                          | <b>1</b> |
| <b>2</b> | <b>Введение</b>                           | <b>1</b> |
| 2.1      | Постановка задачи . . . . .               | 2        |
| 2.2      | Обзор литературы . . . . .                | 2        |
| 2.3      | Полученные результаты . . . . .           | 2        |
| <b>3</b> | <b>Алгоритмы, использованные в работе</b> | <b>2</b> |
| 3.1      | Текст . . . . .                           | 2        |
| 3.2      | Изображения . . . . .                     | 2        |
| <b>4</b> | <b>Оценки качества</b>                    | <b>2</b> |
| <b>5</b> | <b>Эксперименты</b>                       | <b>2</b> |
| <b>6</b> | <b>Выводы</b>                             | <b>2</b> |
| <b>7</b> | <b>Заключение</b>                         | <b>2</b> |

## 1 Аннотация

Одной из задач обработки естественного языка является задача суммаризации текста. Ее целью является уменьшение размера исходного текста без потери ключевой информации. В данной работе мы решаем схожую проблему, но для информационных ресурсов в социальных сетях. В частности, необходимо рассмотреть задачу суммаризации текстов и картинок, поскольку это два основных источника информации. В тексте мы приводим численное обоснование выбранных методов, а также приводим оценку нашей суммаризации людьми.

## 2 Введение

В современном мире создается все больше и больше информации, которую мы можем потреблять. Новости, статьи, юмор постоянно меняются и создаются людьми. При таком потоке информации появляется потребность в инструментах, способных давать как можно больше информации с минимальными потерями.

При чтении новостей люди, как правило, не идут дальше новостных заголовков [], для популярных технических статей создают краткие описания описывающие их достижения и основные моменты [], а визуальный контент нередко подчиняется единому шаблону.

В данной работе мы показываем, как используя современные достижения в области анализа данных можно извлекать полезную информацию из новостных ресурсов в социальной сети вконтакте [1].

## **2.1 Постановка задачи**

Описать технические составляющие задачи

- Потребность в каких данных была
- Как формулировать задачу суммаризации новостного ресурса
- Для текстовых групп – это выделение ключевых слов и генерация заголовков новостей
- Для групп с изображений – это сбор похожих изображений вместе и показ некоторого одного изображения для каждой подгруппы.

## **2.2 Обзор литературы**

Рассказать про литературу, которая рассматривает задачи выше.

## **2.3 Полученные результаты**

Что является результатом работы (будет веб сервис, куда можно закинуть ссылку на группу), как оценивали качество (продолжить результаты работы алгоритмов толкерам), а также оценка качества по автоматизированным метрикам, и как они коррелируют с оценками людей.

# **3 Алгоритмы, использованные в работе**

## **3.1 Текст**

## **3.2 Изображения**

# **4 Оценки качества**

# **5 Эксперименты**

# **6 Выводы**

# **7 Заключение**

# **Список литературы**