

# *Отличие построения нарратива российских государственных и негосударственных СМИ*

Копылов Валентин Владимирович  
БПТ171

20 декабря 2018 г.

Описание проекта по курсу  
«Основы программирования в Python»

## **Что за проект?**

Данный проект ставит перед собой задачу анализа нарративов различных СМИ по определенным рубрикам, таким например как: политика, мир, экономика, спорт, культура, наука и другие. С помощью выгрузки новостных статей и их заголовков за определенный отрезок времени планируется высчитать частоту употребляемых слов в зависимости от темы и новостного ресурса. Таким образом, для пользователя будет доступен функционал взаимодействия, подразумевающий выбор новостных ресурсов (из предложенных) для сравнения и визуализации, выбор интересующей рубрики и временного отрезка за который будут собираться новости.

В итоге для пользователя будет доступен ряд визуализаций, которые он может применить для наглядного сравнения новостных ресурсов.

## **Зачем этот проект?**

Данный проект позволит взглянуть на возможно существующую разницу в формировании нарратива, которая существует в российских СМИ. В ходе данного проекта будет сделана попытка выявления паттернов использования слов и выражений в медиа среде с целью поиска видимых тенденций и отклонений.

## С помощью чего сделан?

Для осуществления проекта будут использованы:

- html-парсинг с помощью библиотек request и BeautifulSoup
- построение базы данных с помощью pandas
- визуализация полученных данных с помощью matplotlib
- работа с представленным текстом, используя Regex

## Основные информационные ресурсы:

- [Lenta.ru](https://lenta.ru)
- [РИА Новости](https://ria.ru)

*Дополнительные(возможные):*

- [NEWSru.com](https://newsru.com)
- [ВЕДОМОСТИ](https://vedomosti.ru)

# Архитектура проекта

Работа программы осуществляется в 3 этапа:

1. Выбор новостных источников (из предложенных) для анализа и дальнейшее формирование базы данных слов
2. Пользователю предоставляется на выбор возможные рубрики для анализа.
3. Выбор желаемого вывода программы  
(многоэтапный: можно получить и визуализацию и статистическую справку)

