# **Инструкция к сдаче**

1. Настоятельно рекомендуем сдавать практическое задание в виде ссылки на личный репозиторий на github.
2. Рекомендуемый способ организации данных в репозитории: создать отдельные папки по темам и помещать в них отдельные файлы для каждой задачи с правильным расширением.

Ссылка на инструкцию по работе с git и сдачу практики:

<https://docs.google.com/document/d/1RAT_ukE39iOfbz1xa39QXae2hBUEZ4U6Fko_wFDdrsM/edit>

Ссылка на видеокурс по Git:

<https://geekbrains.ru/courses/66>

Если остались сложности с системой git, то обратитесь к преподавателю или наставнику.

# **Тема «POS-tagger и NER»**

**Задание 1. Написать теггер на данных с русским языком**

1. проверить UnigramTagger, BigramTagger, TrigramTagger и их комбинации
2. написать свой теггер как на занятии, попробовать разные векторайзеры, добавить знание не только букв но и слов
3. сравнить все реализованные методы, сделать выводы

**Задание 2. Проверить, насколько хорошо работает NER**

**Данные брать из** [**Index of /pub/named\_entities**](http://www.labinform.ru/pub/named_entities/)

1. проверить NER из nltk/spacy/deeppavlov.
2. написать свой NER, попробовать разные подходы.
   1. передаём в сетку токен и его соседей.
   2. передаём в сетку только токен.
   3. свой вариант.
3. сравнить свои реализованные подходы на качество — вывести precision/recall/f1\_score.