

Отчёт о проделанной работе

16 октября - 22 октября

«Исследование и разработка методов машинного обучения для построения персонифицированных моделей классификации в системе защиты от массовых несанкционированных рассылок электронной почты»

Конов Михаил Алексеевич

3 курс, группа 320

1. Написан и согласован отчет о работе за неделю
2. Рассмотрены современные обзоры методов машинного обучения, используемых для фильтрации спама: [“Machine learning for email spam filtering: review, approaches and open research problems”](#) (23 стр.), [“A Survey of Existing E-Mail Spam Filtering Methods Considering Machine Learning Techniques”](#) (10 стр.)
3. Проведена работа над поиском открытых датасетов спама. В результате были найдены датасеты ENRON, PU, SpamAssassin Public Corpus. Из которых наиболее подходящим является для обучения персонализированного классификатора спама является ENRON.
4. С точки зрения алгоритмической части разобраны:
 - Алгоритм классификации спама, основанный на инкрементальном SVM с использованием метода выбора признаков TFDRC: [“Incremental personalized E-mail spam filter using novel TFDRC feature selection with dynamic feature update”](#)(13 стр.)
 - Классификация спама при помощи алгоритма взвешенного голосования трех моделей(KNN, J48, Random Forest): [“Voting-based Classification for E-mail Spam Detection”](#) (14 стр.)
 - Классификаторы ID3 и J48(C4.5): [“A comparative study of decision tree ID3 and C4.5”](#)(7 стр.)
 - Гибридный генетический алгоритм классификации спама и выбора признаков, основанный на RWN: [“An intelligent system for spam detection and identification of the most relevant features based on evolutionary Random Weight Networks”](#) (17 стр.)

Итого: 84 страницы