

Анна Голуб, 11А.

Научный руководитель: Дмитрий Владимирович Грановский.

Место практики: проект «Открытый корпус» (OpenCorpora).

Опыт автоматизированной классификации контекстов при морфологической разметке корпуса

#### Аннотация

В рамках проекта OpenCorpora производится лингвистическая разметка корпуса русского языка с помощью краудсорсинга: пользователи выполняют задания на сайте OpenCorpora. Единичное задание имеет вид: для данного слова в данном контексте выбрать тот грамматический признак, которым оно обладает. Единообразие разметки обеспечивается, в частности, при помощи проверки ответов модераторами. Одной из важных задач является фильтрация поступающих ответов для того, чтобы на модерацию не попадали те из них, которые с очень низкой вероятностью окажутся неправильными. Одним из способов фильтрации является проверка, стоит ли данное слово в некотором заданном ошибкоопасном контексте. Соответственно, задачей является выявление критериев такого контекста.