**Оценка сложности текста: подходы, методы, эксперименты**

Обершт София Дмитриевна

Научный руководитель – д.т.н., к. филол.н. Гавенко Ольга Юрьевна

Оценка сложности текста имеет важное прикладное значение, поскольку позволяет адаптировать его под целевую аудиторию для более успешного понимания и, соответственно, передачи знаний, что особенно актуально при составлении учебной литературы, документов, технической документации. Само понятие сложности текста является по сути своей комплексным: в качестве характеристик сложности используются такие определения, как трудность, сложность (в другом значении), читабельность и понятность. В качестве детерминантов употребления терминов предложены параметры текста, использующиеся в исследовании: качественные и количественные.

Оценка сложности текста в задачах обработки естественного языка (автоматический анализ текста, машинный перевод, изучение языка как иностранного) может трактоваться как задача классификации; традиционно дескрипторами выступают количество лет академического образования, общеевропейские компетенции владения иностранным языком, разделение на "простой/сложный" классы. Помимо статистических методов оценки сложности (формул читабельности) применяются классические методы машинного обучения, нейронные сети (сверточные, с прямой связью и др.) и гибридные модели, включающие в себя оба подхода. В качестве параметров модели используют отобранные вручную лингвистические признаки текста (лексические, семантические, синтаксические и морфологические), n-граммы, TF-IDF и нейронные кодировки представления текста. Однако в большинстве исследований остается открытым вопрос обоснования выбора модели и отбора признаков. В докладе представлен обзор подходов и методов оценки сложности текста, описан ряд задач компьютерной лингвистики, в которых используется параметр сложности, показаны результаты экспериментов по оценке сложности текста методами машинного обучения с использованием количественных параметров.