

АВТОМАТИЧЕСКИЙ ПОДБОР ЗАГОЛОВКОВ ТЕМАТИЧЕСКИХ ТЕКСТОВ АНГЛИЙСКОГО ЯЗЫКА ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Ввиду увеличения объема информационных данных (в том числе массивов текстов на английском языке) возникает потребность в обеспечении их автоматической обработки в части автоматического реферирования. Построение краткого резюме может быть произведено при помощи методов в рамках одного из следующих подходов: экстрактивного (предполагает поиск и извлечение наиболее значимых частей исходного текста, из которых в дальнейшем строится реферат) и абстрактивного (заключается в генерировании нового текста). Одной из задач этого этапа является автоматическое формулирование заголовка. Для ее решения разработаны различные программные продукты, использующие технологию машинного обучения. К одной из последних разработок можно отнести языковую модель GPT-3 от OpenAI.

В ходе исследования нами были использованы такие программные продукты как t-CONSPECTUS, Open Text Summarizer, Paper Title Generator, Writefull, INK's Google Ads Headline Ideas Generator в качестве средств для автоматического анализа и обработки ряда текстов информационного характера в 5 тематических категориях: «Естественные науки», «Информационные технологии», «Лингвистика», «Культура и искусство», «Образование».

Так, автоматически сформулированные заголовки новостных статей соответствовали основным требованиям: отражали одну из идей исходного текста и характеризовались отсутствием значительных потерь в синтаксической структуре независимо от предметной области. Однако некоторые из них получились достаточно длинными, например, для статьи с заголовком «*Students who self-identity as multilingual perform better in exams*» был создан заголовок «*Young people who consider themselves 'multilingual' tend to perform better across a wide range of subjects at school*», что может быть проинтерпретировано как недостаток.

Если касаться сгенерированных заголовков информационных постов, размещаемых в блогах, то более качественный результат показали программные продукты, разработанные с опорой на принципы абстрактивного подхода к автоматическому реферированию. Они позволили создавать заголовки, характеризующиеся высокой степенью конкретизации в передаче смысла исходного текста и реализующие их основные функции (номина-

тивную, информативную и рекламную). В качестве примера можно привести оригинальный и сгенерированный заголовки: «*Understanding Generation Z on Social Media*», «*Connect with Gen Z on social media to discover their needs, interests and values*».

Таким образом, проведенный анализ сгенерированных заголовков в автоматическом режиме показывает, что технология машинного обучения является перспективным инструментом для решения поставленной задачи.