

Использование метрики ROUGE для автоматической оценки качества машинного перевода

Глинская Ольга Сергеевна

Студент (бакалавр)

Саратовский государственный университет имени Н.Г. Чернышевского, Институт филологии и журналистики, Саратов, Россия

E-mail: glinskayolya@gmail.com

В 2024 году объем мирового рынка машинного перевода (МП) достиг 218,8 млн долларов США, а к 2033 году, по прогнозам IMARC Group (International Market Analysis Research and Consulting Group), его объем составит 508,7 млн долларов [4]. Несмотря на то, что МП способствует научно-техническому и культурному обмену, его широкое распространение отчасти привело к некоторой деградации и девальвации языковых норм, снижению требований к грамотности текста [3]. Это обуславливает необходимость постоянной оценки динамики качества систем МП.

Цель работы – выявить эффективность использования ROUGE в качестве самостоятельного инструмента оценки качества МП.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) представляет набор показателей для оценки качества текста и основывается на сравнении n -грамм сгенерированных текстов (predictions) с эталонными (references) [2]. Метрики семейства ROUGE различаются длиной n -грамм. ROUGE-1 основана на поиске соответствий между единицами (отдельными словами) сгенерированного и эталонного текстов, ROUGE-2 – между последовательностями из двух слов. Оценка ROUGE-L основана на измерении самой длинной последовательности слов (LCS), общей для эталонного и сгенерированного текстов [2].

Для исследования были использованы 40 аннотаций статей на русском языке, опубликованных в журнале «Компьютерная лингвистика и вычислительные онтологии» за период с 2017 по 2024 годы и размещённых в российской научной библиотеке eLibrary.ru.

С помощью 5 онлайн-сервисов МП (Яндекс Переводчик, Google Переводчик, DeepL, Microsoft Translator (стандартный и официальный тоны), Reverso) были получены переводы с русского на английский язык (predictions).

Эталонными текстами (references) послужили аннотации на английском языке, представленные авторами данных статей в публикациях.

Полученные оценки ROUGE показали, что наиболее приближенные к эталону переводы для рассматриваемого жанра научного дискурса генерируются Яндекс Переводчиком, Google Переводчиком и Microsoft Translate (стандартный тон) (рис.).

В ходе анализа результатов было выявлено, что оценки ROUGE зависят как от эталонного текста, так и от исходного текста на русском языке. Некоторые семантически и грамматически верные переводы получали низкую оценку из-за использования системами МП вариативных соответствий, лексических и грамматических опущений и добавлений, синтаксических перестановок [1], несовпадающих с эталонным вариантом. Наиболее точными и объективными оказались показатели ROUGE для текстов, состоящих преимущественно из терминов. Таким образом, рассмотренная метрика не может использоваться как самостоятельный инструмент для оценки динамики качества МП, но может применяться для оценки перевода текстов требующих однозначных эквивалентов [1], что является актуальным для научного дискурса.

Источники и литература

- 1) Алексеева И. С. Введение в переводоведение: Учеб. пособие для студ. филол. и лингв. фак. высш. учеб. заведений. - СПб.: «Академия», 2004.
- 2) Мельничук Д. В., Носкина А. В. Сравнение NLP-моделей на задаче суммаризации академических текстов на русском языке // Компьютерная лингвистика и вычислительные онтологии. Выпуск 7 (Труды XXVI Международной объединённой научной конференции «Интернет и современное общество», IMS-2023, Санкт-Петербург, 26–28 июня 2023г. Сборник научных статей).—СПб.: Университет ИТМО, 2024. С. 54–59.
- 3) Берендяев М. В., Сант С. А., Светова С. Ю. Меморандум Ассоциации переводческих компаний по вопросам институционализации, стандартизации и создания нормативно-правовой базы для применения машинного перевода и прочих технологий автоматической генерации текста в качестве профессиональных инструментов в области перевода и локализации [Электронный ресурс] // Документы Ассоциации переводческих компаний. 2023. URL: <https://atcru.org/upload/iblock/ff5/fp58d29psjlsr3minttu0nuqntlay50.pdf> (дата обращения: 01.02.2025).
- 4) Machine Translation Market Report 2024–2032 [Electronic resource] // URL: <https://www.imarcgroup.com/machine-translation-market> (дата обращения: 25.12.2024).

Иллюстрации

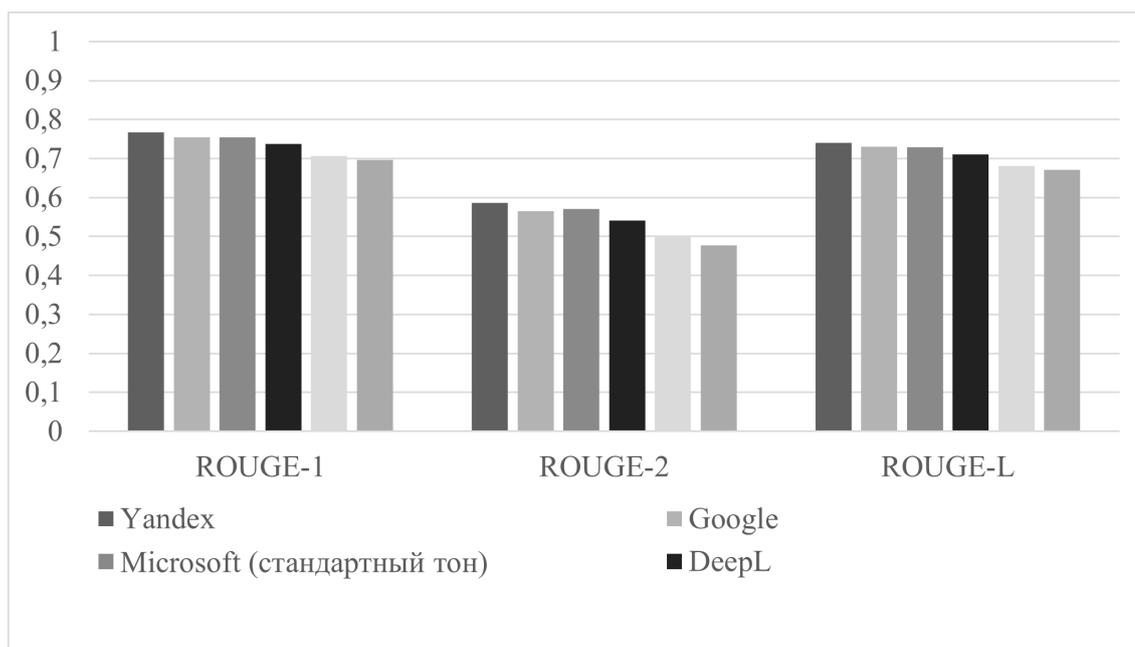


Рис. : Среднее значение показателей ROUGE