В. М. Василевская

ПРОБЛЕМА АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ РУССКОЯЗЫЧНЫХ НАУЧНЫХ ТЕКСТОВ В СФЕРЕ ИКТ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ ОНЛАЙН-СЕРВИСОВ НА БАЗЕ НЕЙРОСЕТЕЙ

В современном мире, где объем информации в области информационно-коммуникационных технологий (ИКТ) постоянно растет, задача реферирования текстов становится особенно актуальной. Эффективное извлечение ключевых идей и результатов из массива публикуемых материалов необходимо для всех, кто стремится оставаться в курсе последних тенденций и разработок.

Целью настоящего исследования является проведение сравнительного анализа различных онлайн-сервисов, функционирующих на базе нейросетей и поддерживающих функцию реферирования русскоязычных текстов, с акцентом на их эффективность в обработке материалов из области ИКТ, регулярно содержащих математические формулы, алгоритмы и большое количество аббревиатур. **Эмпирическую базу** исследования составили 15 абстрактивных автоматических рефератов, полученных нами на основе текстов ИКТ-тематики, отобранных из научной электронной библиотеки «КиберЛенинка». Исследование проводилось на базе пяти различных сервисов, отвечающих следующим условиям: доступность для пользователей белорусского интернет-сегмента, бесплатный доступ к функционалу, поддержка обработки русскоязычных текстовых материалов.

Критерии технической оценки включали: тип реферирования (абстрактивное или экстрактивное), максимально допустимое к обработке количество символов с пробелами, наличие функции регулировки степени сжатия и средний объем реферата относительно исходного текста при настройке либо запросе сжатия на 50 % (таблица).

Технические характеристики сервисов, использованных для сравнительного анализа эффективности обработки ключевых элементов русскоязычных научных текстов тематики ИКТ

Сервис/Критерий	Тип реферирования	Кол-во символов с пробелами	Выбор степени сжатия	Средний объем реферата от оригинала, %
1. YaGPT	абстрактивное	4000	да	19,2
2. Neural Writer	абстрактивное	10000	да	20,4
3. Word Counter	абстрактивное	>15000	нет	24,0
4. Myneuralnetworks	абстрактивное/ экстрактивное	10000	да	29,1
5. GPT-4	абстрактивное	>15000	да	46,0

Содержательная оценка работы анализируемых сервисов предполагала анализ их способности корректно воспроизводить в реферате ключевые элементы ИКТ-текстов, включая математические формулы, алгоритмы и специализированные аббревиатуры, характерные для данной предметной области.

1. **Передача алгоритма.** Алгоритмы часто имеют сложную структуру, включающую множество шагов, условий и циклов. Ограниченная способность нейросетевых моделей к адекватному распознаванию формальной логики алгоритмов негативно влияет на качество их представления в ре-

фератах. В исходном тексте фрагмент с алгоритмом включал: вербальное описание принципа работы алгоритма, инициальный шаг с перечислением условий его выполнения, два следующих алгоритмических шага.

Результаты исследования демонстрируют, что ни один из проанализированных сервисов реферирования не обеспечил полной и точной передачи алгоритмических структур. При сохранении общей структуры шагов и условий в большинстве случаев наблюдались существенные семантические искажения исходного содержания и частичная утрата важных деталей. Наибольшие отклонения от исходного текста зафиксированы в сервисе myneuralnetworks, реферат которого полностью исключил вербальное описание алгоритма, последовательность шагов выполнения и условия перехода. Относительно лучшие показатели продемонстрировал сервис YaGPT, сохранивший полную структуру, базовую логику выполнения и ключевые условия, хотя и с частичными смысловыми искажениями, связанными с утратой деталей второстепенной значимости.

2. **Передача формулы.** Формулы могут иметь сложную структуру, включающую различные операторы, функции и переменные. Нейросети могут испытывать трудности при распознавании и интерпретации этих элементов, что обусловлено ограниченной способностью моделей к корректному распознаванию формальных выражений и адекватной интерпретации сложных взаимосвязей между переменными. В реферируемом материале было два фрагмента текста, содержащих формулы. Каждый фрагмент текста с формулой включал: вербальное описание формулы, собственно формулу, описание переменных формулы.

Анализ результатов обработки формул выявил существенные ограничения: в большинстве случаев содержательные фрагменты сводились к кратким вербальным описаниям (1–2 предложения), имеющим значительные смысловые искажения. Наиболее эффективным оказался сервис GPT-4, который в одном из случаев корректно идентифицировал и передал все ключевые элементы формулы, однако при непосредственном включении ее в реферат допустил добавление посторонних символов, что привело к нарушению исходной семантики. Условно приемлемый результат продемонстрировал YaGPT, экстрактивно сохранивший в реферате один из формульных фрагментов целиком, однако полностью проигнорировал все остальные смысловые блоки исходного текста, что делает полученный реферат семантически неполноценным и не соответствующим базовым требованиям к научному реферированию.

3. **Передача аббревиатур.** Нейросетевые модели испытывают трудности при обработке текстов с множеством аббревиатур из-за недостатка контекста и их многозначности в разных областях. Это приводит к ошибкам в расшифровке и искажению смысла при реферировании специализированных материалов.

Наиболее низкую эффективность при обработке текстов с высокой концентрацией аббревиатур продемонстрировал сервис Neural Writer, допустивший грубую ошибку в интерпретации термина *NLP* (Natural Language

Processing), некорректно транслитерировав его как *НЛП* (нейролингвистическое программирование), что свидетельствует о неспособности системы учитывать тематический контекст и приводит к полному искажению смысла исходного материала. Дополнительным недостатком данного сервиса стала тенденция к исключению из рефератов фрагментов с множественными аббревиатурами, что повлекло за собой потерю значимых микротем. Наилучшие результаты показал GPT-4, обеспечивший корректную передачу большинства аббревиатур, тогда как сервисы Word Counter и Myneuralnetworks, несмотря на удовлетворительное качество интерпретации переданных аббревиатур, продемонстрировали потерю некоторой части аббревиатур вместе с содержащими их смысловыми блоками оригинального текста.

Проблемы, возникающие у нейросетей при реферировании текстов, содержащих алгоритмы, формулы и аббревиатуры, приводят к значительному сокращению объема реферата. Сложные для обработки элементы текста часто исключаются, что негативно сказывается на полноте и точности представляемой информации. При запросе сжатия текста оригинала на 50 % лучший результат показал сервис GPT-4, средний объем сгенерированных им рефератов составил 46 % от оригинала. Наименьший объем имеют рефераты, созданные при помощи YaGPT – в среднем 19,2 % от оригинала.

Основные выводы:

1. Автоматические рефераты, генерируемые сервисами на базе нейросетей, являются абстрактивными. Такой тип реферирования слабо поддается контролю со стороны пользователя и оказывается малоэффективным в тех случаях, когда особенно важны полнота и точность передачи информации оригинала. 2. Специфические элементы текстов в области ИКТ (алгоритмы, формулы, аббревиатуры) представляют значительные трудности для нейросетевых моделей при реферировании. Это приводит к сокращению объема итогового текста, а также к неполной и неточной передаче содержания оригинала. Наибольшие трудности возникают при передаче алгоритма, наименьшие – при передаче аббревиатур. 3. Использование сервиса Neural Writer для реферирования текстов ИКТ представляется нецелесообразным из-за низкой эффективности в передаче специфических элементов. Средний объем создаваемых этим сервисом рефератов (20,4 % от объема оригинала) является недостаточным для воспроизведения всех существенных данных исходного текста. 4. Сервис YaGPT не подходит для реферирования текстов ИКТ, что обусловлено ограничением по количеству символов, доступных к обработке (всего 4000), а также возникающими ошибками при работе с текстом, содержащим специфические элементы. 5. Требуется проведение дополнительных исследований с привлечением расширенного корпуса текстов для оценки эффективности сервисов реферирования на базе нейросетей (Word Counter, Myneuralnetworks, GPT-4 и аналогичных) в области ИКТ. Сгенерированные рефераты должны быть подвергнуты анализу в рамках лексико-семантического подхода (обработка ключевых элементов, контекста

их употребления, синтаксической структуры и межфразовой семантики) с целью определения степени адекватности передачи содержания исходного текста и учета выявленных ошибок при разработке модели автоматической компрессии русскоязычных научных текстов в области ИКТ.