

**В. М. Василевская**

**АНАЛИЗ МАШИННЫХ РЕПРОДУКТИВНЫХ РЕФЕРАТОВ,  
ВЫПОЛНЕННЫХ НЕЙРОСЕТЬЮ «АЛИСА»  
(на материале газетных текстов разных жанров)**

В наше время Интернет является неотъемлемой частью жизни социума и выступает одним из главных источников получения информации о людях и событиях. Темп жизни неумолимо возрастает, новости публикуются и изменяются с огромной скоростью, а современному человеку жизненно необходимо получать актуальную информацию без необходимости тратить много времени на чтение и переработку текстов.

Автоматическое реферирование газетных текстов основано на использовании специальных алгоритмов и искусственного интеллекта, которые позволяют выделить ключевые и наиболее важные сведения из текста, создавая его краткую версию. Этот процесс может включать в себя выделение основных тем и идей, имен и дат, а также оценку значимости информации.

Основываясь на результате реферирования, существуют два основных подхода к обобщению текста: экстрактивный (подразумевает извлечение некоторого количества наиболее важных предложений из реферируемого текста) и абстрактивный (включает интерпретацию и анализ исходного текста, выделение основной информации и генерацию нового, более короткого текста). Современные системы автоматического реферирования стремятся использовать именно абстрактивный подход, позволяющий обрабатывать тексты подобно тому, как это делает человек. Тем не менее они создают преимущественно репродуктивные рефераты, частично сокращая текст и технически отсекая второстепенную информацию. Создание продуктивных же рефератов представляет собой гораздо более сложную задачу для машины, поскольку такое реферирование предполагает краткую оценку исходного текста и при необходимости сжатие его до нескольких строк или даже одного предложения.

Наиболее перспективными и популярными сервисами для автоматического реферирования текстов являются сервисы, основанные на нейронных сетях. Нейросеть – это интеллектуальная система, принцип работы которой имитирует работу человеческого мозга. Она состоит из огромного количества искусственных нейронов, объединенных в слои, которые взаимодействуют между собой, обучаясь на входных данных.

В русскоязычном пространстве Интернета для получения авторефератов новостных статей на базе нейросети широко используется YaGPT 2 (сервис «Алиса, давай придумаем!»), который встроен в голосовой помощник «Алиса», созданный российской компанией «Яндекс».

В этой связи интересно проанализировать, насколько «Алиса» справляется с автоматическим реферированием газетных текстов разных жанров

(информационные и аналитические как наиболее широко представленные в СМИ) в «полевых условиях», то есть без проведения предобработки исходного текста и постобработки полученного реферата.

Материалом для исследования послужили три текста с официального сайта БЕЛТА. Отбор текстового материала был обусловлен в том числе относительно небольшим количеством символов, доступных «Алисе» для обработки. В связи с технической невозможностью включить тексты примеров в полном объеме здесь даны их описания, взятые с официального сайта, что позволяет получить представление о содержании текстов примеров и сопоставить их с полученными авторефератами.

Рассмотрим примеры двух текстов, относящихся к информационным жанрам. Первый текст является интервью. Его ценность для читателя/слушателя/зрителя заключается в экспертном мнении специалиста, которого интервьюируют.

Приведем описание, размещенное на сайте перед подробным интервью:

*Николай Шерстнев*

*Белорусский футбол: свежий взгляд в будущее – интервью*

*Председатель ассоциации “Белорусская федерация футбола”*

*Кажется, что футбольный сезон завершился только что, а вот уже и новый стучит мячами – до старта всего ничего: 2 марта пройдет матч за Суперкубок. Для белорусских любителей спорта номер один, также, как и для футбольных профессионалов, прошлый год выдался неоднозначным. Были и радости, и неудачи, и скандалы... О своем взгляде на итоги сезона-2023 и перспективах на будущее корреспонденту БЕЛТА рассказал председатель ассоциации “Белорусская федерация футбола” Николай Шерстнев, который занял пост в марте прошлого года.*

Далее рассмотрим автореферат текста интервью, выполненный нейросетью «Алиса».

*Белорусский футбол продолжает радовать болельщиков, несмотря на неоднозначный прошлый год с его радостями, неудачами и скандалами. Председатель ассоциации “Белорусская федерация футбола” Николай Шерстнев поделился своим взглядом на итоги сезона-2023 и перспективы на будущее. Шерстнев занял этот пост в марте 2022 года, отметив, что любая руководящая должность предполагает большую ответственность за результаты. Он подчеркнул, что старался следить за спортивными соревнованиями, но имеет свое мнение о том, что является хорошим и где можно улучшить.*

Как мы видим, текст автореферата практически полностью повторяет текст аннотации и не содержит сведений, позволяющих узнать мнение эксперта по вопросам, освещенным в интервью.

Попытка воспользоваться функцией «Ответь иначе» привела к получению еще менее информативного автореферата:

*Председатель ассоциации “Белорусская федерация футбола” Николай Шерстнев рассказал о своем взгляде на итоги футбольного сезона-2023 и перспективах на будущее. Он отметил, что занимает новую для себя должность, но уже успел погрузиться в тему и привыкнуть к ответственности.*

Таким образом, результат автореферирования текста интервью нельзя считать удовлетворительным.

Следующий текст представляет собой заметку. Это самый распространенный информационный жанр. Задача заметки – сообщить о важном факте либо событии общественной жизни. Такие тексты отличаются сжатостью и оперативностью.

Приведем размещенное на сайте описание заметки:

*“Самое главное – это большие профессионалы”. Какие кандидатуры предлагает Минск в Совет Республики*

*14 марта, Минск /Корр. БЕЛТА/. На совместном заседании президиума Минского городского Совета депутатов и Минского городского исполнительного комитета выдвинуты кандидаты в члены Совета Республики Национального собрания восьмого созыва, передает корреспондент БЕЛТА.*

Текст заметки содержит большое количество имен и фамилий, а также несколько цитат, содержащих второстепенную информацию.

Рассмотрим автореферат заметки.

*В Совет Республики Национального собрания восьмого созыва от Минска выдвигаются кандидаты: председатель Совета Республики Наталья Кочанова, главный врач Республиканского клинического медицинского центра Ирина Абельская, член Совета Республики седьмого созыва Дмитрий Басков и другие. От Минска избирается восемь членов Совета Республики, выборы проводятся в два этапа: выдвижение кандидатов и голосование.*

В целом нейросеть успешно справилась с выделением главной информации, исключив цитаты. В то же время она не предоставила полного списка кандидатов, заменив часть имен выражением *и другие*.

Вторая попытка оказалась более успешной:

*В Совет Республики от Минска выдвигаются восемь кандидатов. Это Наталья Кочанова, Ирина Абельская, Дмитрий Басков, Дмитрий Жук, Марина Ильина, Виктор Кондратович, Олег Руммо и Сергей Сивец. Избрание членов Совета Республики проходит в два этапа – выдвижение кандидатов и голосование.*

Здесь читатель получил необходимую информацию о ходе выборов и выдвигаемых кандидатах.

Примером газетного текста аналитического жанра является комментарий. Он используется для оперативного разъяснения важных событий общественной жизни и строится, как правило, вокруг одного факта либо цепи однозначных фактов.

Приведем описание комментария, размещенное на сайте БЕЛТА.

*Дворец гражданских обрядов главного управления юстиции Мингорисполкома в следующем году отметит 60-летие. Долгое время он был единственным учреждением в столице, где регистрировали браки с иностранцами. Тысячи свадеб, сотни золотых и серебряных юбилеев, невесты и женихи из разных стран... Здесь помнят много красивых историй любви и верности. Изменились ли предпочтения молодоженов, какие дни выбирают для торжества, востребованы ли услуги выездной регистрации – эти и другие вопросы корреспондент БЕЛТА задала руководителю Дворца Юлии Кочиной.*

Помимо прочего комментарий содержит большое количество дат и чисел, написанных как цифрами, так и буквами.

Рассмотрим автореферат данного комментария.

*В следующем году Дворец гражданских обрядов главного управления юстиции Мингорисполкома отметит свое 60-летие. Он долгое время был единственным в столице учреждением, где регистрировались браки с иностранцами. Руководитель Дворца Юлия*

*Кочина рассказала о предпочтениях молодоженов и востребованных услугах. По ее словам, у каждой пары есть свои счастливые числа для похода в ЗАГС, но часто выбираются красивые даты, содержащие комбинации двоек и троек. Оказываются востребованными как будние, так и выходные дни. В 2024 году ожидается высокий спрос на даты с двойками и четверками, как это было 24 февраля 2023 года.*

Текст автореферата не содержит в себе часть вопросов, отмеченных в описании и освященных далее в комментарии. Также оказался полностью проигнорирован целый абзац с наибольшим количеством чисел, в котором автор комментария сравнивала возраст женихов и невест и разные годы при вступлении в первый либо повторный брак.

На основании проведенного исследования с точки зрения пользователя можно обозначить ряд проблем, касающихся автоматического реферирования газетных текстов разных жанров, выполняемого нейросетью «Алиса»: 1) сравнительно небольшой объем исходного текста, допустимый к автореферированию. Данная нейросеть не способна обработать большой объем материала; 2) неточная обработка числовых данных, при этом числа, написанные буквами распознаются лучше написанных цифрами; 3) если в тексте освещены несколько разных вопросов, то «Алиса» может проигнорировать некоторые из них. Новостные заметки автореферированы гораздо лучше текстов интервью и комментариев. Это связано с тем, что заметка представляет собой один связный текст, посвященный рассмотрению одного вопроса, а интервью и комментарий состоят из множества микротекстов (вопросно-ответные реплики, пояснения нескольких фактов, не всегда связанных одной тематикой); 4) в случае наличия в исходном тексте перечисления однородных членов предложения «Алиса» может сократить их перечень, заменив часть из них, например, конструкцией *и другие*. Это в некоторых случаях недопустимо, поскольку такой автореферат предоставит читателю неполную либо искаженную информацию.

Обозначенные проблемы связаны с недостаточной обученностью нейросети «Алиса» на сегодняшний день и могут быть решены в ходе дальнейших исследований и разработок.