

---

---

**ЭЛЕКТРОННАЯ КОМПОНЕНТНАЯ БАЗА, ФОТОНИКА  
И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ УПРАВЛЕНИЯ**

---

---

УДК 004.032.26

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ СУЩЕСТВУЮЩИХ  
НЕЙРОСЕТЕВЫХ МЕТОДОВ РЕЗЮМИРОВАНИЯ ТЕКСТОВ**

© 2021 г. А.Г. Сбоев<sup>1,2</sup>, А.А. Селиванов<sup>2</sup>, Р.Б. Рыбка<sup>1,\*</sup>,  
И.А. Молошников<sup>1</sup>, А.В. Грязнов<sup>1</sup>

<sup>1</sup> *Национальный исследовательский центр «Курчатовский институт», Москва, Россия*

<sup>2</sup> *Национальный исследовательский ядерный университет «МИФИ», Москва, Россия*

\*E-mail: Rybka\_RB@nrcki.ru

В исследовании рассмотрены достижения последних лет в области автоматического резюмирования текстов на основе глубоких нейронных сетей, в том числе с применением обучения с подкреплением. Показано, что модели на основе экстрактивного подхода дают более высокую точность по метрике ROUGE, однако отмечено, что абстрактный подход концептуально более перспективен. Отмечено, что текущие точности методов на основе обучения с подкреплением ниже, чем у методов на основе классического обучения, что может быть объяснено использованием менее совершенных моделей в качестве агентов и значительно возрастающей вычислительной сложностью. Указаны недостатки, свойственные классическим метрикам оценки на основе совпадения подстроки (rouge, blue), представлено описание альтернативных метрик, которые могут в перспективе лечь в основу более совершенных моделей резюмирования текстов, но на текущем этапе также обладают рядом недостатков.

Рассмотрены наборы данных, которые используются в литературе для оценки методов автоматического резюмирования текстов. В результате сравнения заявленных точностей на новостных статьях CNN/Dailymail, можно сделать вывод, что на текущем этапе наиболее совершенным методом резюмирования текстов (по оценке на основе метрики rouge-1) является метод с использованием языковой модели типа BART и механизма редуцирования коллапса представления (BART+R3F), который достигает значения rouge-1, равного 41.17. В работе выделены технические решения в разных методах, совместное использование которых может обеспечить прирост в точности решения задачи резюмирования текстов.

## ВВЕДЕНИЕ

В настоящее время доступный объём текстовой информации в открытом доступе растёт (научные и новостные статьи, тексты в социальных сетях, отзывы), что делает её своевременный анализ трудоёмким или вовсе невозможным. Одним из вариантов решения данной проблемы является использование методов автоматического резюмирования текстов.

Под автоматическим резюмированием текста (в англ. литературе: text summarization) понимают генерацию текста, длина которого меньше исходного, а информативность сопоставима с рефератом, составленным людьми.

Перспективными для решения данной задачи являются методы на основе нейронных сетей, которые зарекомендовали себя в широком спектре задач обработки естественного языка (Natural Language Processing, NLP).

В данной работе будут рассмотрены подходы к автоматическому резюмированию текстов на основе искусственных нейронных сетей и метрики оценки этих методов (раздел 2); современные методы автоматического резюмирования текстов на основе нейронных сетей (раздел 3), которые реализуют различные подходы; наборы данных, которые используются в современных работах для оценки методов резюмирования текстов (раздел 4), результаты на наиболее популярном наборе данных методов, вошедших в обзор (раздел 5).

## ПОДХОДЫ К АВТОМАТИЧЕСКОМУ РЕЗЮМИРОВАНИЮ ТЕКСТОВ

Методы автоматического резюмирования текстов реализуют следующие основные подходы: экстрактивный, абстрактный или гибридный. Экстрактивный подход предполагает создание ре-

зюме на основе извлечения слов и фраз исходного текста. Абстрактный подход предполагает генерацию резюме с использованием стороннего словаря, что позволяет включать в резюме слова и фразы, которые отсутствуют в исходном тексте. Гибридный подход является объединением двух предыдущих подходов.

В литературе модели резюмирования текстов оцениваются на основе совпадения подстроки сгенерированного и эталонного резюме: метрики Rouge 1, Rouge 2, Rouge L [1], BLEU [2]. Rouge 1 и Rouge 2 сравнивают униграммы и биграммы соответственно, BLEU использует и униграммы, и биграммы, а Rouge L использует наиболее длинную общую подстроку для оценки сгенерированного резюме.

Недостатками данного класса метрик является, во-первых, привязка к конкретным словам эталона – если в сгенерированном резюме та же мысль выражена иными словами, оценка окажется низкой, во-вторых, слабый учёт порядка слов в резюме, особенно в случае с метриками на униграммах. То есть некорректное предложение, которое содержит те же слова, что и эталон, но в случайном порядке, получит высокую оценку.

В качестве возможных перспективных решений описанных проблем рассматриваются метрики на основе языковых моделей: Rouge-w, METEOR [3], BLEURT [4]. Данные метрики используют дополнительные модели, которые оценивают смысловую близость эталонного и сгенерированного резюме на основе их векторного представления. Rouge w использует модель word2vec [5]; METEOR – лексическую базу данных WordNet; BertScore и BLEURT – контекстные языковые модели на основе топологии Transformer [6], предварительно обученные на больших корпусах текстов (BERT [7]).

Недостатками данных метрик является, во-первых, большая вычислительная сложность по сравнению с методами на основе совпадения подстроки, во-вторых, необходимость подбора языковой модели и её параметров, чтобы оценка работала корректно, в-третьих, отсутствие гарантии эффективности работы при переносе метрики с одного набора данных на другой, в-четвёртых, часто такие модели являются привязанными к языку, что не позволяет, к примеру, использовать оригинальные модели BertScore и BLEURT для русского языка.

Все описанные метрики не являются дифференцируемыми, а значит не могут быть оптимизированы напрямую в канве классического обучения нейронных сетей. Прямая оптимизация этих метрик при обучении нейронной сети возможна при

использовании обучения с подкреплением, где в качестве награды рассматривается метрика оценки, а в качестве действий агента – генерация слов или предложений для итогового резюме.

## МЕТОДЫ АВТОМАТИЧЕСКОГО РЕЗЮМИРОВАНИЯ ТЕКСТОВ

### 1) Экстрактивные методы

#### HSASRL [8]

Данный метод реализует экстрактивный подход на основе обучения с подкреплением.

Основная идея метода – ранжирование предложений по значимости на основе прямой оптимизации метрики оценки метода (rouge) и модели, которая создаёт векторное представление предложений. Векторное представление формируется на основе biLSTM с механизмом внимания. Слова изначально представлены с помощью дистрибутивной модели word2vec, полученные вектора используются, чтобы создать внутреннее векторное представление слов с учётом контекста. Данные вектора используются для получения векторного представления предложений.

На основе векторов предложений, информации об исходном тексте и уже сгенерированной части резюме вычисляются производные характеристики предложений с использованием полносвязных слоёв: содержательность предложения ( $C_j$ ) (см. формулу 1), значимость предложения во входном тексте ( $M_j$ ) (см. формулу 2), новизна предложения относительно уже сформированного резюме ( $N_j$ ) (см. формулу 3), векторного представления позиции предложения в тексте ( $P_j$ ) (см. формулу 4).

$$C_j = W_c s_j; \quad (1)$$

$$M_j = s_j^T W_s d; \quad (2)$$

$$N_j = s_j^T W_r \tanh(o_j); \quad (3.1)$$

$$o_j = \sum_{i=1}^{j-1} s_i P(y_i = 1 \vee s_i, o_i, d); \quad (3.2)$$

$$P_j = W_p p_j; \quad (4)$$

где  $s_j$  – векторное представление предложения;

$d$  – векторное представление документа;

$o$  – векторное представление уже сформированного резюме;

$p$  – позиционное кодирование предложения в виде конкатенации закодированных векторов индексов позиции предложения от начала и от конца текста;

$W$  – матрицы весов полносвязных слоёв, которые отвечают за  $C_j$ ,  $M_j$ ,  $N_j$  и  $P_j$  соответственно.

Результаты конкатенируются и подаются в полносвязный слой с функцией активации Softmax, выходные активности которого определяют вероятность включения предложения в резюме.

В процессе обучения для каждого документа генерируется множество вариантов реферата, чтобы ускорить процесс обучения и сходимость алгоритма, это пространство сокращается до наиболее вероятных вариантов.

Изначально на каждой итерации обучения из документа выбирается  $k$  предложений с наибольшим значением метрики оценки (rouge). Из данных предложений формируются все возможные последовательности длины  $L$ , которая рассчитывается как средняя длина эталона в предложениях в рассматриваемом наборе данных (например, 3 предложения для корпуса CNN/DM).

Данное множество потенциальных рефератов используется для расчёта награды и функции потерь:

$$\nabla L(\theta) \approx -r(\hat{y}) \sum_{i=1}^m \nabla \log p(\hat{y}_i \vee s_i, d, \theta),$$

где  $L$  – функция потерь;

$r$  – функция награды (rouge);

$s_i$  – предложение  $i$ ;

$\hat{Y}_i$  – предсказанная метка для предложения  $i$  (включать или нет в резюме);

$d$  – текущий рассматриваемый документ;

$\theta$  – параметры модели.

Таким образом, из метода HSASRL примечательными технологическими решениями являются:

- вычисляемые признаки предложения (содержательность, новизна относительно уже сгенерированной части резюме, значимость предложения в документе и позиция предложения в документе);
- функция потерь, которая при обучении сети учитывает не всю совокупность возможных резюме, которую генерирует модель, что позволяет ускорить работы модели и сходимость.

### НАNSum [9]

Данный метод реализует экстрактивный подход на основе графовых слоёв нейронной сети.

Входной текст представляется в виде гетерогенного графа и векторного представления слов на основе языковой модели ALBERT [10]. Граф включает три типа узлов: именованные сущности, слова, предложения.

Именованные сущности в тексте заменяются на особые токены («Person\_A», «Person\_B», «Place\_A» и т.д.). Для слов каждое вхождение слова представляется отдельным узлом. Каждый узел предложения отражает информацию об одном предложении.

В графе четыре типа связей:

1. Направленные связи от именованных сущностей к словам предложения (Next);

2. Узлы слов и сущностей имеют направленные связи с предложением (In);

3. Одинаковые именованные сущности соединяются связью (Same);

4. Два узла предложения, если они имеют общую триграмму, связью (Similar).

Полученный граф разбивается на три подграфа: граф уровня слов, граф уровня слов и предложений, граф уровня предложений, каждый из которых обрабатывается отдельным графовым свёрточным слоем с механизмом внимания (Graph Attention Layer, GAT) [11]. В результате получается векторное представление предложений с учётом контекста.

На основе векторного представления предложений определяется, какие предложения включать в резюме – для этого используется ещё два слоя GAT и полносвязный слой с сигмоидальной активацией. В резюме включаются предложения с активностью выше эмпирически подбираемого порога.

Таким образом, следующие технологические решения из метода НАNSum следует отметить:

1. Представление текста в виде гетерогенного графа, включающего узлы слов, предложений и именованных сущностей, что позволяет учитывать информацию о структуре текста;

2. Использование именованных сущностей, что даёт дополнительную информацию для модели резюмирования текстов, как правило, значимую для резюме (личности, названия мест, организаций и так далее).

### 2) Абстрактивные методы

#### Deep Communicating Agents for Abstractive Summarization [12]

Данный метод реализует абстрактивный подход на основе обучения с подкреплением.

Концепция метода заключается в анализе текста по частям несколькими агентами, которые представляют собой блоки нейронной сети. Каждый блок состоит из трёх слоёв biLSTM, которым на вход дополнительно подаётся внутреннее состояние других агентов для поддержания общего гло-

бального контекста у каждого агента (развитие топологии CommNet [13]).

Выходные вектора слов от каждого агента взвешиваются на основе каскада механизмов внимания, таким образом формируется сначала вектор контекста агента, а затем единый вектор контекста. Текущий вектор контекста конкатенируется с вектором, полученным на предыдущем шаге, и подаётся на вход «декодировщику» – одностороннему LSTM, который создаёт векторное представление для генерации нового слова реферата. Данный вектор передаётся в полносвязный слой, размерность которого соответствует словарю модели. Таким образом выход данного слоя позволяет выбрать, какое слово включить в реферат на текущем этапе.

Дополнительно используется механизм копирования [14], который позволяет включать в реферат слова из исходного текста, в том числе если они отсутствуют в словаре модели.

Сеть настраивается на основе обучения с подкреплением, в качестве награды используется метрика оценки работы модели (rouge). На каждом шаге работы сети вычисляется разность между оценкой резюме с последним сгенерированным предложением и без него, что позволяет оценить вклад в резюме последнего сгенерированного предложения:

$$r_{inc}(\hat{o}_q) = r([\hat{o}_1, \dots, \hat{o}_q]) - r([\hat{o}_1, \dots, \hat{o}_{q-1}]),$$

где  $r_{inc}$  – инкрементальная награда на текущий шаг генерации;

$\hat{o}_q$  – сгенерированное предложение под номером  $q$ ;

$r([\hat{o}_1, \dots, \hat{o}_q])$  – награда за резюме с учётом последнего сгенерированного предложения;

$r([\hat{o}_1, \dots, \hat{o}_{q-1}])$  – награда за резюме без использования последнего сгенерированного предложения.

Ещё одной частью алгоритма является механизм критики, который на основе более простой модели генерирует резюме, награда за которое вычитается из награды основной модели. Польза от данного механизма заключается в том, что основная модель «соревнуется» с критиком, стремится генерировать резюме лучше.

В случае с «механизмом самокритики» модель критика имеет общую с основной моделью часть. В рассматриваемом алгоритме механизм самокритики построен на весах внимания каждого слова в

тексте, которые генерируются на выходе агентов, на каждой итерации работы сети критик добавляет к своему резюме слово с максимальным весом.

Таким образом, в методе DCA следует отметить следующие технологические решения:

1. Множественные агенты с обменом информацией для анализа разных частей текста (что в частности позволяет обрабатывать тексты большой длины);

2. Инкрементальная награда по предложениям, которая позволяет учесть вклад каждого предложения в информативность резюме, а значит обучать сеть генерировать текст, в котором каждое предложение вносит дополнительную информацию в резюме;

3. Механизм самокритики, который позволяет обучаться модели генерировать лучше, чем выбранная модель критика, которая в данном случае имеет общую часть с основной моделью;

4. Механизм копирования, как дополнение к методу абстрактного резюмирования, что позволяет улучшить резюме в том числе дополняя его словами, которые отсутствуют в словаре модели.

## BART + R3F [15]

Данный метод реализует абстрактный подход на основе тонкой настройки (fine tuning) предобученной языковой модели.

В качестве языковой модели для генерации реферата используется BART [16]. Модель BART основана на слоях типа Transformer и состоит из кодировщика и декодировщика. Кодировщик предназначен для создания векторного представления слов входного текста с учётом контекста. Декодировщик использует как векторное представление слов, так и информацию об уже сгенерированном резюме, чтобы генерировать каждое следующее слово резюме.

В данной статье на стадии тонкой настройки языковой модели используется дивергенция Кульбака-Лейблера, рассчитанная между векторами, которые были получены до и после тонкой настройки модели. Эта мера используется как дополнительный член в функции ошибки, что позволяет ограничить изменение весов языковой модели в ходе тонкой настройки, таким образом не «испортить» её изначальное состояние. Данный механизм называют «редуцированием коллапса представления». Коллапс представления (representational collapse) – деградация обобщающего представления предобученной модели на этапе тонкой настройки. Иными словами, это явление, при котором в ре-

зультате дообучения модели снижается её обобщающая способность.

Данный метод может быть использован с любыми языковыми моделями, а также совместно с иными функциями ошибки, в том числе с учётом награды в процессе обучения с подкреплением.

### НАБОРЫ ДАННЫХ

При разработке новых моделей резюмирования текстов их оценку производят обычно на одном из популярных корпусов. Наиболее часто используемым среди них является корпус новостей ресурсов CNN и DailyMail, CNN/DM [17,18].

Корпус содержит 312084 новостных текстов, средняя длина текста – 781 слово, средняя длина эталонного резюме – 3.75 предложений или 56 слов. У набора данных существует стандартное разбиение на тренировочную (287226 текстов), валидационную (13368) и тестировочную (11490) части. Существуют версии корпуса с анонимизацией именованных сущностей (заменой личностей, организаций, географических точек на уникальные идентификаторы вида «@entity5»).

Менее распространено использование таких корпусов, как:

- Reddit TIFU short [19] – корпус текстов из социальной сети Reddit, из раздела, в котором пользователи описывают неудачные истории, которые произошли с ними. В качестве эталонного резюме рассматривается краткая выжимка истории за авторством самого пользователя. Число текстов: 79949, средняя длина текста – 342.4 слова, резюме – 9.33 слова;
- Reddit TIFU long [19] – содержит более длинные тексты и резюме, чем Reddit TIFU short. Число текстов: 42984, средняя длина текста: 432.6, средняя длина резюме – 23 слова.
- PubMed [20] – корпус медицинских научных статей, 133000 статей, средняя длина текста в словах – 3016, средняя длина резюме в словах – 203;
- Arxiv [20] – корпус научных статей на разные темы, 215000 статей, средняя длина текста в словах – 4938, средняя длина резюме в словах – 220;
- Big Patent [21] – корпус текстов заявок на патент в США, включая рефераты, написанные специалистами патентного бюро. Число текстов – 1341362, средняя длина текста в словах – 3572.8, средняя длина резюме в словах – 116.5.
- New York Times [22] – корпус новостных статей журнала New York Times, 655000 текстов, сред-

няя длина текста в словах – 530, средняя длина резюме в словах – 38.

На русском языке корпус для задачи резюмирования текстов был представлен на соревновании «Диалог» [23], корпус содержит 1003869 текстов новостного ресурса «Россия Сегодня» [24].

### РЕЗУЛЬТАТЫ

В разделе 3 описаны методы, которые реализуют различные подходы к автоматическому резюмированию текстов, при этом в рамках подхода достигают наибольшей точности (по метрикам rouge-1, rouge-2, rouge-l) на корпусе данных CNN/DM. Результаты представлены в таблице 1.

На основе результатов таблицы можно сделать

Таблица 1. Точность решения задачи резюмирования текстов на наборе данных CNN/DM

Метод	Тип метода	Rouge-1	Rouge-2	Rouge-L
BART+R3F	абстрактивный	44.38	<b>21.53</b>	<b>41.17</b>
HAHSum	экстрактивный	<b>44.68</b>	21.30	40.75
DCA	RL, абстрактивный	41.69	19.47	37.92
HSASRL	RL, экстрактивный	41.5	19.5	37.9

вывод о том, что перспективным является использование технологических решений из наиболее точных методов на основе абстрактивного и экстрактивного подходов для разработки новых методов резюмирования текстов, а также использование обучения с подкреплением совместно с данными технологическими решениями для достижения более высокого качества решения задачи.

### БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке внутреннего гранта НИЦ «Курчатовский Институт» (приказ №1055) с использованием оборудования центра коллективного пользования «Комплекс моделирования и обработки данных исследовательских установок мега-класса» НИЦ «Курчатовский институт», <http://ckp.nrcki.ru/>.

### ЗАКЛЮЧЕНИЕ

В настоящее время наибольшая точность (по метрикам ROUGE-1-f, ROUGE-2-f, ROUGE-L-f) достигается экстрактивными подходами, что частично может быть объяснено природой метрик, по которым оценивается качество моделей. С другой стороны, абстрактивные подходы развиваются как более перспективные за счет возможной генерации резюме из слов, отсутствующих в исходном тексте.

Среди экстрактивных подходов лучшим является метод HANSum – метод на основе гетерогенного графа с использованием информации об именованных сущностях, графовых слоёв внимания и языковой модели ALBERT, точность (ROUGE-1-f/ROUGE-2-f/ROUGE-L-f): 44.68 / 21.30 / 40.75.

Наиболее точным методом на основе абстрактного подхода является BART+R3F, ключевая особенность – использование механизма предотвращения коллапса представления, что обеспечивает более эффективную процедуру дообучения языковой модели. Достигнутая точность (ROUGE-1-f/ROUGE-2-f/ROUGE-L-f): 44.38 / 21.53 / 41.17. Текущий уровень может быть улучшен за счет использования RL методов в совокупности с State-of-the-art моделями резюмирования текстов, что до настоящего времени не исследовалось.

Перспективными для такого исследования являются следующие методы обучения с подкреплением: инкрементальное награждение по сгенерированным частям резюме, что позволяет учесть вклад в результат (по выбранным метрикам) каждой сгенерированной части резюме; дополнение функции потерь дистанцией между сгенерированными частями для сокращения избыточности резюме; механизм «многоорукого бандита», который обеспечивает оптимизацию нескольких функций потерь в рамках процесса обучения; добавление в награду обратной связи, при котором на каждой итерации обучения человек оценивает качество полученного резюме.

В рамках дальнейших работ предполагается развитие единого подхода на основе набора описанных выше методов для эффективного решения задачи автоматического резюмирования текстов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Lin, Chin-Yew. «Rouge: A package for automatic evaluation of summaries». In *Text summarization branches out*, pp. 74–81. 2004.
2. Papineni K., Roukos S., Ward T., Zhu W.J. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40<sup>th</sup> annual meeting of the Association for Computational Linguistics 2002 Jul.* (pp. 311–318).
3. Banerjee S., Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization 2005 Jun.* (pp. 65–72).
4. Sellam T., Das D., Parikh A. BLEURT: Learning Robust Metrics for Text Generation. In *Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics 2020 Jul.* (pp. 7881–7892).
5. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. 2013 Jan. 16.
6. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is All you Need. In *NIPS 2017 Jan.* 1.
7. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019 Jun.* (pp. 4171–4186).
8. Mohsen F., Wang J., Al-Sabahi K. A hierarchical self-attentive neural extractive summarizer via reinforcement learning (HSASRL). *Applied Intelligence*. 2020 Mar. 16:1–4.
9. Jia R., Cao Y., Tang H., Fang F., Cao C., Wang S. Neural Extractive Summarization with Hierarchical Attentive Heterogeneous Graph Network. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020 Nov.* (pp. 3622–3631).
10. Sharma S.G., Lan R.S., Chen M. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *Submitted to International Conference on Learning Representations*. <https://openreview.net/forum> 2020.
11. Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*. 2017 Oct. 30.
12. Celikyilmaz A., Bosselut A., He X., Choi Y. Deep Communicating Agents for Abstractive Summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) 2018 Jun.* (pp. 1662–1675).
13. Sukhbaatar S., Szlam A., Fergus R. Learning multiagent communication with backpropagation. *Advances in Neural Information Processing Systems*. 2016:2252–60.
14. Gu J., Lu Z., Li H., Li V.O. Incorporating Copying Mechanism in Sequence-to-Sequence Learning. In *Proceedings of the 54<sup>th</sup> Annual Meeting of the*

- Association for Computational Linguistics (Volume 1: Long Papers) 2016 Aug. (pp. 1631–1640).*
15. Aghajanyan A., Shrivastava A., Gupta A., Goyal N., Zettlemoyer L., Gupta S. Better fine-tuning by reducing representational collapse. *arXiv preprint arXiv:2008.03156*. 2020 Aug 6.
  16. Lewis M., Liu Y., Goyal N., Ghazvininejad M., Mohamed A., Levy O., Stoyanov V., Zettlemoyer L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *In Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics 2020 Jul. (pp. 7871–7880).*
  17. Hermann K.M., Kočiský T., Grefenstette E., Espeholt L., Kay W., Suleyman M., Blunsom P. Teaching machines to read and comprehend. *Advances in Neural Information Processing Systems*. 2015 Dec. 5;28.
  18. Nallapati R., Zhou B., dos Santos C., glar Gulçehre Ç., Xiang B. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. *CoNLL 2016*. 2016 Aug. 11:280.
  19. Kim B., Kim H., Kim G. Abstractive Summarization of Reddit Posts with Multi-level Memory Networks. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019 Jun. (pp. 2519–2531).*
  20. Cohan A., Deroncourt F., Kim D.S., Bui T., Kim S., Chang W., Goharian N. A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents. *In Proceedings of NAACL-HLT 2018 (pp. 615–621).*
  21. Sharma E., Li C., Wang L. BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization. *In Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics 2019 Jul. (pp. 2204–2213).*
  22. Li J.J., Thadani K., Stent A. The role of discourse units in near-extractive summarization. *In Proceedings of the 17<sup>th</sup> Annual Meeting of the Special Interest Group on Discourse and Dialogue 2016 Sep. (pp. 137–147).*
  23. Malykh V.A., Kalaidin P.S. *Headline Generation Shared Task on Dialogue'2019 report*.
  24. Gavrilov D., Kalaidin P., Malykh V. Self-attentive model for headline generation. *In European Conference on Information Retrieval 2019 Apr. 14 (pp. 87–93).* Springer, Cham.