

## Impact Factor:

ISRA (India) = 3.117  
ISI (Dubai, UAE) = 0.829  
GIF (Australia) = 0.564  
JIF = 1.500

SIS (USA) = 0.912  
ПИИИ (Russia) = 0.156  
ESJI (KZ) = 8.716  
SJIF (Morocco) = 5.667

ICV (Poland) = 6.630  
PIF (India) = 1.940  
IBI (India) = 4.260  
OAJI (USA) = 0.350

SOI: [1.1/TAS](#) DOI: [10.15863/TAS](#)

### International Scientific Journal Theoretical & Applied Science

p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online)

Year: 2019 Issue: 06 Volume: 74

Published: 30.06.2019 <http://T-Science.org>

QR – Issue



QR – Article



**Oleg Yurievich Sabinin**

Peter the Great St. Petersburg Polytechnic University  
PhD in Computer Science

**Julie Vladimirovna Shabalina**

Peter the Great St. Petersburg Polytechnic University  
Master of Computer Science student

**Irina Vladimirovna Tarasova**

Peter the Great St. Petersburg Polytechnic University  
Lecturer

#### SECTION 4. Computer science, computer engineering and automation.

## EXTRACTING METHOD OF AUTOMATIC TEXT SUMMARIZATION ON THE BASIS OF NEURAL NETWORKS

**Abstract:** The purpose of the article is to describe and to analyze the new extract method of automatic text summarization on the basis of neural networks. The neural network is trained to determine key text sentences on the basis of property vector and to decide, whether the sentence should be included in summary or not.

**Key words:** text summarization, neural networks, machine learning.

**Language:** Russian

**Citation:** Sabinin, O. Y., Shabalina, J. V., & Tarasova, I. V. (2019). Extracting method of automatic text summarization on the basis of neural networks. *ISJ Theoretical & Applied Science*, 06 (74), 601-607.

**Soi:** <http://s-o-i.org/1.1/TAS-06-74-73> **Doi:**  <https://dx.doi.org/10.15863/TAS.2019.06.74.73>

### ИЗВЛЕКАЮЩИЙ МЕТОД АВТОМАТИЧЕСКОГО АННОТИРОВАНИЯ ТЕКСТОВ НА ОСНОВЕ АППАРАТА НЕЙРОННЫХ СЕТЕЙ

**Аннотация:** Целью данной статьи является разработка и анализ извлекающего метода автоматического аннотирования текстов на основе аппарата нейронных сетей. Задачей нейронной сети является определение ключевых предложений текста на основе свойств предложения для принятия решения о включении предложения в аннотацию.

**Ключевые слова:** аннотирование текста, нейронные сети, машинное обучение.

#### Введение.

#### Постановка проблемы

Задачи аннотирования текстов приобретают в последнее время всё большую актуальность, как для сети Интернет, но и для других хранилищ информации, например, библиотек или баз знаний различных организаций.

Аннотирование текстов помогает выделить ключевые части текста и сократить объёмы просматриваемой информации.

Огромное количество и большой объём материалов затрудняет быстрое получение аннотаций по текстам, так как формирование кратких ёмких сводок вручную требует

значительных затрат времени и человеческих ресурсов.

В связи со сказанным, задача реализации эффективных методов автоматического аннотирования текстов приобретает всё большую важность.

#### Анализ последних исследований и публикаций

С самого начала активного использования электронно-вычислительных машин первого поколения, то есть с середины пятидесятых годов прошлого века, стали предприниматься попытки решения задач обработки текста на естественном языке. Одной из первых задач по обработке

## Impact Factor:

ISRA (India) = 3.117  
ISI (Dubai, UAE) = 0.829  
GIF (Australia) = 0.564  
JIF = 1.500

SIS (USA) = 0.912  
РИИЦ (Russia) = 0.156  
ESJI (KZ) = 8.716  
SJIF (Morocco) = 5.667

ICV (Poland) = 6.630  
PIF (India) = 1.940  
IBI (India) = 4.260  
OAJI (USA) = 0.350

естественно-языковых текстов при помощи ЭВМ стало автоматическое аннотирование.

С тех пор было проведено множество исследований по разработке автоматизированных методов и моделей аннотирования [1,2]. Решением проблемы занимались такие отечественные исследователи как Н.В. Лукашевич, Р.Г. Пиотровский, П.Г. Осминин, С.А. Тревгода, В.А. Яцко и др. Из зарубежных исследователей стоит обратить внимание на Н.Р. Luhn, Н.Р. Edmundson, R. Mihalcea, J. Kupiec, E. Lloret, G. Salton и др.

В настоящее время можно выделить два основных подхода к автоматическому аннотированию:

- экстракция – извлекающие методы, основанные на извлечении из первичных документов наиболее информативных фрагментов [3];
- абстракция – генерирующие методы, предусматривающие создание нового текста, обобщающего первичные документы [4].

**Извлекающие методы** работают путём определения наиболее важных фрагментов текста (предложения, абзацы). При этом данные фрагменты не обрабатывают, а извлекают в таком порядке и виде в каком они приведены в тексте. Основные сложности, связанные с данным подходом, заключаются в определении ключевых предложений текста, и затем связи этих предложений в единый, удобочитаемый текст.

Извлекающие методы можно разделить на две большие группы:

- поверхностные методы, которые не прибегают к сложному лингвистическому анализу,
- глубокие методы.

К **поверхностным методам** относятся, например:

- методы, которые для выбора предложений используют статистические характеристики [5, 6];
- методы, основанные на представлении документа в виде графа, вершинами которого являются предложения или слова из текста [7];
- методы, использующие деревья решений, опорные векторы и нейронные сети [8];
- методы на основе скрытых марковских моделей, в которых при анализе предложения учитывается, входит ли предыдущее предложение в аннотацию [9].

К **глубоким методам** относятся, например, методы с использованием латентно-семантического анализа, которые анализируют взаимосвязь между предложениями текста и терминами, содержащимися в них, выявляют тематики, присутствующие в тексте, и в аннотацию выбирается определенное количество предложений от каждой тематики [10].

**Генерирующие методы** в отличие от извлекающих методов, направлены на создание нового материала, явно непредставленного в тексте исходного документа. Другими словами, они интерпретируют и исследуют текст с помощью методов обработки естественного языка, чтобы создавать новые структурные единицы текста, которые передают самую важную информацию из исходного документа. При использовании генерирующих методов текст аннотации строится на правилах, предполагающих наличие лингвистической базы знаний.

Для генерирующих методов можно выделить несколько направлений:

- использование шаблонов,
- сжатие предложений,
- полноценная абстракция.

*Подходы на основе шаблонов* используют заранее подготовленные шаблоны для представления документа. Лингвистические паттерны или правила извлечения используются для заполнения пропусков в этом шаблоне.

*Сжимающие методы* выполняют извлечение наиболее важных предложений из текста, но либо удаляют из них лишнюю информацию, либо объединяют несколько предложений, пытаясь при этом сохранить связность и смысл текста.

Существующие работы на данную тему предлагают различные способы решения этой задачи, например, в работе [11] исходный документ представляется в виде вложенного дерева, которое состоит из двух видов структур: дерева документа и дерева предложения. Это дерево строится на основе теории риторической структуры, разработанной в 1980-е годы американскими лингвистами Уильямом Манном и Сандрой Томпсон. Эта теория предлагает описание структуры дискурса (текста) в виде сетей дискурсивных единиц, соединенных семантическими отношениями [12]. Теория риторической структуры используется для построения алгоритма аннотирования текста также в работе [13].

*Для полноценной абстракции* наиболее перспективно выглядит модель кодировщик-декодировщик, которая основывается на использовании рекуррентных нейронных сетей.

**Выделение нерешенных ранее частей общей проблемы**

Несмотря на множество проведенных исследований, проблема разработки формальных методов и моделей для автоматического аннотирования еще не решена, ввиду того, что задача формализации естественного языка достаточно трудоемка, а сам язык является неоднозначным, неограниченным и эволютивным.

## Impact Factor:

ISRA (India) = 3.117  
ISI (Dubai, UAE) = 0.829  
GIF (Australia) = 0.564  
JIF = 1.500

SIS (USA) = 0.912  
РИИЦ (Russia) = 0.156  
ESJI (KZ) = 8.716  
SJIF (Morocco) = 5.667

ICV (Poland) = 6.630  
PIF (India) = 1.940  
IBI (India) = 4.260  
OAJI (USA) = 0.350

Вышеупомянутые характеристики естественного языка играют особенно большую роль при исследовании материалов на русском языке, так как именно для русского языка характерен свободный порядок слов, морфологическая сложность, подвижное разноместное ударение и высокая степень сегментной редукции [14].

На текущий момент из методов автоматического аннотирования текстов на русском языке наиболее распространены различные статистические и графовые методы, являющиеся представителями экстрагирующего подхода. Аннотации, полученные с помощью экстрагирующих подходов, зачастую характеризуются недостаточно высоким качеством текста, бессвязностью. Абстрагирующие же подходы потенциально способны обеспечить лучшее качество текста аннотации, но они чрезвычайно трудны для практической реализации и находятся на уровне исследовательских разработок.

Поскольку большинство текстов обладают достаточно выраженной структурой, ключевые части документа можно представить путем выбора предложений на основе их свойств и характеристик. В работе [15] был предложен схожий подход и рассмотрены такие методы машинного обучения с учителем для решения задач автоматического аннотирования, как наивный байесовский классификатор, метод опорных векторов. Исследователем были получены обнадеживающие результаты, поэтому в данной статье было решено использовать вышеупомянутый экстрактивный подход для автоматического аннотирования текстов на русском языке, только в качестве классификатора, в отличие от [15], были выбраны искусственные нейронные сети.

### Цель статьи

Разработка метода автоматического аннотирования текстов на русском языке на основе экстракции из текста наиболее значимых элементов при помощи аппарата нейронных сетей.

### Изложение основного материала

Предлагаемый метод предполагает, что исходный документ представляет собой набор предложений, а сами предложения рассматриваются как набор свойств и характеристик. Среди этого набора выбираются те предложения, которые нейронная сеть сочтёт более релевантными. Результатом является некоторое подмножество предложений исходного текста.

Изначально необходимо выполнить следующее:

- определить рассматриваемые свойства и характеристики предложений, значения которых

будут являться входными данными для нейронной сети;

- создать размеченный тестовый корпус текстов для последующего обучения нейронной сети;
- произвести непосредственно само обучение сети.

### Определение рассматриваемых свойств и характеристик предложений.

Каждое предложение аннотируемого текста представляется в виде вектора, состоящего из  $h$  характеристик  $[f_1, f_2, \dots, f_h]$ :

- отношение порядкового номера абзаца, к которому принадлежит предложение, к общему числу абзацев исходного документа ( $f_1$ );
- отношение порядкового номера предложения в абзаце к общему числу предложений в абзаце ( $f_2$ );
- отношение количества символов рассматриваемого предложения к количеству символов самого длинного предложения текста ( $f_3$ );
- отношение количества ключевых слов в предложении к общему количеству тематических слов предложения ( $f_4$ );
- отношение количества совпадающих тематических слов данного предложения и предыдущего к общему количеству тематических слов рассматриваемых предложений ( $f_5$ );
- отношение количества совпадающих тематических слов данного предложения и предыдущего к общему количеству тематических слов двух рассматриваемых предложений ( $f_5$ );
- отношение количества совпадающих тематических слов данного предложения и последующего к общему количеству тематических слов двух рассматриваемых предложений ( $f_6$ ).

Свойства  $f_1$ - $f_2$  основываются на местоположении предложения в документе или в его абзаце. Ожидается, что эти параметры поспособствуют выбору ключевых предложений, так как аннотации, состоящие из первых предложений абзацев, превосходят аннотации, составленные с помощью других методов статьи [16], а предложения, расположенные в начале и конце абзацев, имеют высокие шансы попасть в итоговый текст [17].

Свойство  $f_3$  поможет избавиться от слишком коротких вводных предложений, которые вряд ли попадут в аннотацию [18].

Свойство  $f_4$  зависит от количества ключевых и тематических слов в предложении. Тематические слова получают следующим образом: из документа производится выборка всех существительных, прилагательных и глаголов, которые в последствии сводятся к их начальной форме. Для получившегося набора слов высчитывается их встречаемость в тексте.

## Impact Factor:

ISRA (India) = 3.117	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	РИИЦ (Russia) = 0.156	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 8.716	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 5.667	OAJI (USA) = 0.350

Ключевыми словами считаются 25% тематических слов, но не больше 10, что соответствует объему оперативной памяти человека [19]. Ожидается, что с помощью этого свойства увеличится вероятность выбора ключевых предложений, поскольку термины, которые часто встречаются в документе, вероятно, связаны с его темой [5]. Для выделения тематических слов был использован Томита-парсер компании Яндекс, имеющий встроенную поддержку русского языка, доступную документацию и используемый в таких популярных сервисах как Яндекс.Новости и Яндекс.Работа.

Свойства  $f_5-f_6$  основываются на симметричном реферировании, то есть на определении количества связей между предложениями [20].

Рассмотренные свойства могут быть изменены или дополнены. Выбор рассматриваемых свойств предложений определяет, какие предложения попадут в итоговую аннотацию и влияет на работу нейронной сети.

### Обучение нейронной сети

Обучение нейронной сети проводится для изучения типов предложений, которые должны быть включены в аннотацию. Обучение проводится на тестовом корпусе текстов, где каждое предложение отмечено как входящее в аннотацию или же нет.

Нейронная сеть ищет закономерности, присущие предложениям, которые должны быть включены в аннотацию. Используется нейронная сеть прямого распространения с тремя слоями, которая, как было доказано, является универсальным функциональным аппроксиматором [21]. Сеть может обнаруживать паттерны и аппроксимировать функцию любых данных с точностью до 100%, если в наборе данных нет противоречий.

Создание нейронной сети проводилось в NeurophStudio. Входной слой разрабатываемой нейронной сети состоит из шести нейронов, где каждый нейрон соответствует одному из свойств предложения, пяти нейронов скрытого слоя и одного нейрона выходного слоя. В качестве активационной функции используется сигмоида, обучение сети проводится методом обратного распространения ошибки.

Для создания тестового корпуса было использовано 62 статьи на различные тематики, найденные в сети Интернет. Каждый текст состоял от 27 до 102 предложений, в среднем – из 49. Всего было проанализировано 3076 предложений. 565 предложений было помечено, как ключевые, в среднем 9 на текст. Нейронная сеть была успешно обучена за 23 итерации.

Итоговая среднеквадратичная ошибка для тестового корпуса составила 0.16733. Точность нейронной сети составила 88,76% по сравнению с ручной выборкой для тестового корпуса. Для корпуса из 10 новых текстов точность составила 82,31%.

### Оценка эффективности системы

Задача оценки эффективности автоматического аннотирования текстов также является чрезвычайно важной и сложной. Не существует общего алгоритма оценки аннотации, исходя из конечного набора признаков и правил, поэтому современные подходы к оцениванию результатов автоматического аннотирования основываются на сравнении полученных автоматически аннотаций с модельными, вручную созданными аннотациями.

Для сравнения автоматически полученных аннотаций с полученными вручную аннотациями обычно используют набор метрик ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [22].

Все метрики этого набора основываются на идее максимального покрытия автоматическими аннотациями ручных и наоборот. Для вычисления покрытия используются N-граммы. N-грамма - это последовательность из N элементов, в данном случае слов.

Метрика ROUGE-N основывается на вычислении совпадающих N-грамм автоматической аннотации и модельной, где N – это длина используемых цепочек слов.

Оценка реализованной программной системы производилась с помощью метрик ROUGE-1 и ROUGE-2, основывающихся на анализе последовательностей из одного и двух слов соответственно.

Например, для предложения «мама мыла раму» можно выделить такие униграммы ( $N = 1$ ), как: мама, мыла, раму. Для этого же предложения можно извлечь следующие биграммы ( $N = 2$ ): мама мыла, мыла раму.

Само по себе число совпадающих N-грамм автоматической и модельной аннотаций не является оценкой эффективности результата автоматического аннотирования.

Для оценки эффективности результата используются характеристики:

- ROUGE Precision;
- ROUGE Recall;
- F-мера.

Характеристика Rouge Precision (точность) – оценка того, насколько хорошо модельные аннотации покрывают автоматические аннотации. Она рассчитывается по формуле:

$$\text{Precision} = \frac{C_N}{M_N}, \quad (1)$$

где  $C_N$  - количество совпадающих N –грамм;  
 $M_N$  - общее число N-грамм модельной аннотации.

## Impact Factor:

ISRA (India) = 3.117	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	РИИЦ (Russia) = 0.156	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 8.716	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 5.667	OAJI (USA) = 0.350

Характеристика Rouge Recall (полнота) – оценка того, насколько хорошо автоматические аннотации покрывают модельные. Она рассчитывается по формуле:

$$\text{Recall} = \frac{C_N}{A_N}, \quad (2)$$

где  $C_N$  - количество совпадающих N-грамм;  
 $A_N$  - общее число N-грамм автоматической аннотации.

Очевидно, что чем выше точность и полнота, тем лучше. Но на практике максимальная точность и полнота не достижимы, поэтому для объединения информации о точности и полноте вычисляется F-мера:

$$F = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (3)$$

где F – среднее гармоническое от точности и полноты.

F-мера выступает в качестве итогового значения метрики, отражающего качество полученных аннотаций.

К примеру, ручная аннотация представлена предложением «мама мыла раму», а автоматическая – «мама мыла наше окно», значения метрик ROUGE-1 и ROUGE-2, вычисленные по формулам (1) - (3), можно увидеть в табл. 1.

Таблица 1. Пример вычисления метрик ROUGE-1 и ROUGE-2

Метрика	Precision	Recall	F-мера
ROUGE-1	$2/3 \approx 0,67$	$2/4 = 0,5$	0,57
ROUGE-2	$1/2 = 0,5$	$1/3 \approx 0,33$	0,4

Для создания тестового корпуса для оценки эффективности реализованной системы было использовано 10 статей на различные тематики, найденные в сети Интернет. Выбранные тексты включали от 18 до 94 предложений, в среднем – 35. Для каждого текста вручную были написаны модельные аннотации. Всего было проанализировано 363 предложения, 102 из них было помечено, как предложения, входящие в итоговые аннотации, в среднем 10 на текст.

Эффективность реализованной системы была оценена на тестовом наборе документов путём сравнения модельных и автоматических аннотаций с помощью метрик ROUGE-1 и ROUGE-2. Максимально возможным значением характеристик является 1.

Результаты оценивания эффективности можно увидеть в табл. 2.

Таблица 2. Результаты оценивания эффективности реализованной системы

Метрика	Precision	Recall	F-мера
ROUGE-1	0,61	0,32	0,42
ROUGE-2	0,23	0,12	0,16

В соответствие с тем, что F-мера является итоговым показателем метрик, необходимо провести анализ именно этого результата.

Исходя из того, что F-мера метрики ROUGE-1 относительно близка к 1, можно сделать вывод, что автоматические и ручные аннотации были достаточно близки по набору слов. Показания F-меры метрики ROUGE-2 несколько хуже. Полученные результаты вполне оправданы в силу сложности обучения компьютера пониманию естественного языка.

Чтобы осуществить полноценный анализ эффективности системы и сформулировать

выводы относительно её применимости, было проведено сравнение показателей метрик ROUGE-1 и ROUGE-2, полученных в данной работе, с показателями данных метрик уже существующих инструментов. Для сравнения были выбраны системы автоматического аннотирования текстов, показавшие наилучшие результаты на сегодняшний день: [15], [23], [24].

Результаты работы системы, рассматриваемой в данной работе, представлены в последней строке табл. 3.

Таблица 3. Сравнение показателей реализованной системы с показателями уже существующих систем

Автор, год	ROUGE-1	ROUGE-2
Nallapati, 2017	0,39	0,16
See, 2017	0,39	0,17
Wong K., 2008	0,42	0,12

## Impact Factor:

ISRA (India) = 3.117	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	РИИЦ (Russia) = 0.156	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 8.716	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 5.667	OAJI (USA) = 0.350

Предлагаемая система, 2019	0,42	0,16
----------------------------	------	------

Как видно из табл. 3 реализованная система превзошла уже существующие, показав наивысший результат для метрики ROUGE-1 и практически наилучший для метрики ROUGE-2, что говорит о том, что наша система предоставляет более качественные автоматические аннотации, нежели другие системы. Полученные результаты позволяют утвердить применимость разработанного метода для аннотирования текстов на русском языке, а также продолжать дальнейшие исследования. Впоследствии возможно усложнение топологии нейронной сети, а также изменение или дополнение рассматриваемых характеристик предложений для улучшения качества аннотаций. Выбранные для анализа предложения характеристики, а также сформированные вручную обучающая и тестовая выборки имеют большое влияние на работу нейронной сети, а, следовательно, и всей системы в целом. Сеть обучается в соответствии со стилем читателя и в соответствии с предложениями, которые именно этот читатель считает подходящими для аннотации. Можно рассматривать данную

особенность как преимущество данного подхода, так как любой человек может обучить нейронную сеть в соответствии со своими личными предпочтениями.

### Выводы и предложения

Точность результатов предложенного метода автоматического аннотирования текстов на тестовой выборке составила 88,76%. Результаты оказались вполне удовлетворительными, что позволяет проводить дальнейшие исследования. Выбранные для анализа предложений свойства, а также выбранные независимым читателем ключевые предложения для тестового корпуса текстов имеют большое влияние на работу нейронной сети. Сеть обучается в соответствии со стилем читателя и в соответствии с предложениями, которые именно этот читатель считает ключевыми. Можно рассматривать данную особенность как преимущество данного подхода, так как любой человек может обучить нейронную сеть в соответствии со своими личными предпочтениями.

## References:

1. Shkurina, M. V., & Sabinin, O. Y. (2018). An overview and analysis of automatic text summarization methods. *ISJ Theoretical & Applied Science*, 12 (68), 282-286.
2. Shkurina, M. V., & Sabinin, O. Y. (2019). Comparative analysis of extractive text summarization methods for texts in Russian language. *ISJ Theoretical & Applied Science*, 06 (74), 164-169.
3. Radev, D. R., Hovy, E., & McKeown, K. (2002). Introduction to the Special Issue on *Summarization Computational Linguistics*. 28(4), 399-408.
4. Osminin, P. G. (2016). Postroenie modeli referirovaniya i annotirovaniya nauchno-tehnicheskikh textov, orientirovannih na avtomaticheskij perevod (Doctoral dissertation, Cheljabinsk).
5. Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*. 2(2), 159-165.
6. Aksenova, T. S., & Sabinin, O. Y. (2011). *Sistema avtomaticheskogo referirovaniya nauchno-tehnicheskikh textov na osnove Oracle Data Mining. Nauchnije issledovanija i innovacionnaja dejatel'nost'*: materialy nauchno-pract. konf. (pp.61-68). SPb.: Izd-vo Politechn. Univer.
7. Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization. *Journal of Artificial Intelligence Research*. 22, 457-479.
8. Sabinin, O. Y., & Shabalina, J. V. (2018). Method of automatic text summarization of texts in russian on the basis of the neural networks. *East European Science Journal*, 4 (32), part 1, 56-59.
9. Conroy, J. M., & O'leary, D. P. (2001). *Text summarization via hidden Markov models* Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '01. (pp.406-407). New York, New York, USA: ACM Press.
10. Ozsoy, M. G., Alpaslan, F. N., & Cicekli, I. (2011). Text summarization using Latent

## Impact Factor:

<b>ISRA (India)</b>	<b>= 3.117</b>	<b>SIS (USA)</b>	<b>= 0.912</b>	<b>ICV (Poland)</b>	<b>= 6.630</b>
<b>ISI (Dubai, UAE)</b>	<b>= 0.829</b>	<b>PIHHI (Russia)</b>	<b>= 0.156</b>	<b>PIF (India)</b>	<b>= 1.940</b>
<b>GIF (Australia)</b>	<b>= 0.564</b>	<b>ESJI (KZ)</b>	<b>= 8.716</b>	<b>IBI (India)</b>	<b>= 4.260</b>
<b>JIF</b>	<b>= 1.500</b>	<b>SJIF (Morocco)</b>	<b>= 5.667</b>	<b>OAJI (USA)</b>	<b>= 0.350</b>

- Semantic Analysis. *Journal of Information Science*. 37(4), 405–417.
11. Kikuchi, Y., et al. (2014). *Single Document Summarization based on Nested Tree Structure* Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). (pp.315-320). Stroudsburg, PA, USA: Association for Computational Linguistics.
  12. Mann, W., Matthiessen, C., & Thompson, C. A. (1989). *Rhetorical structure theory and text analysis*. University of Southern California.
  13. Sabinin, O. Y., & Trevgoda, S. A. (2008). Sistemi avtomaticheskogo referirovanija teksta. *Pribori i sistemi. Upravljenje, control, diagnostika*. 1, 23-26.
  14. Bolshakova, E. I., et al. (2011). *Avtomaticheskaja obrabotka tekstov na estestvennom jazike i komputernaja lingvistika*. (p.272, p.3). Moscow: MIEM.
  15. Wong, K. F., Wu, M., & Li, W. (2008, August). *Extractive summarization using supervised and semi-supervised learning*. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1 (pp. 985-992). Association for Computational Linguistics.
  16. Brandow, R., Mitze, K., & Rau, L. F. (1995). Automatic condensation of electronic publications by sentence selection. *Information Processing & Management*, 31(5), 675-685.
  17. Baxendale, P. B. (1958). Machine-made index for technical literature—an experiment. *IBM Journal of Research and Development*, 2(4), 354-361.
  18. Kupiec, J., Pedersen, J., & Chen, F. (1995, July). *A trainable document summarizer*. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 68-73). ACM.
  19. Vanjushkin, A. S., & Grascenko, L. A. (2016). *Metodi i algoritmi izvlečenja ključevih slov*. Novie informacionne tehnologii v avtomatizirovannih sistemah, 19.
  20. Stupin, V. S. (2004). *Sistema avtomaticheskogo referirovanija metodom simmetrichnogo referirovanija*. In Kompiuternaja lingvistika i intelektualnije tehnologii, 579-591.
  21. Galizdra, V. I., & Babaev, S. B. (2011). *Nejronnije seti i aproksimacija danih*. Nauchnije i obrazovatelnije problem grazdanskoj zaschiti, 3.
  22. Lin, C. Y. (2004). ROUGE: A Package for Automatic Evaluation of summaries Conference: In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004).
  23. Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents//AAAI, 3075-3081.
  24. See, A., Liu, P. J., & Manning, C. D. (2017). *Get to the point: Summarization with pointer-generator networks* //arXiv preprint arXiv:1704.04368