

*Абукова Г.А.*

*студент 4 курса факультета иностранных языков*

*Дагестанский государственный университет*

*Алиев З.Г.*

*старший преподаватель*

*Россия, г. Махачкала*

## **СОВРЕМЕННЫЕ МЕТОДЫ И ПОДХОДЫ АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ**

***Аннотация:** В данной статье дается обзор и классификация основных методов автоматического реферирования. Анализируются преимущества и недостатки каждого из подходов. Среди современных методов реферирования выделяются методы по типу получаемого реферата (экстракция, абстракция); по уровню анализа исходного текста: поверхностный уровень, уровень сущностей текста, уровень дискурсной структуры текста и т.д. Дается характеристика каждому методу.*

***Ключевые слова:** автоматическое реферирование, аннотирование, автоматическое извлечение, ключевые слова, способы и методы автоматического реферирования.*

*Abukova G.A.*

*student 6<sup>th</sup> year, faculty of foreign languages , Dagestan State University*

*Aliev Z.G.*

*senior lecturer*

*Russia, Makhachkala*

## **MODERN METHODS AND APPROACHES OF AUTOMATIC REFERENCE**

***Abstract:** This article provides an overview and classification of the main methods of automatic summarization. The advantages and disadvantages of each of the approaches are analyzed. Among modern methods of abstracting, methods are*

*distinguished according to the type of abstract received (extraction, abstraction); by the level of analysis of the source text: surface level, level of text entities, level of discourse structure of the text and so on.*

**Keywords:** *automatic summarization, annotation, automatic extraction, keywords, ways and methods of automatic summarization.*

На сегодняшний день из-за информационной перегрузке и нехватки времени, автоматическое реферирование и аннотирование пользуется популярностью. С помощью различных методов мы можем за короткое время усвоить огромный объем информации.

В зависимости от способа методы могут различаться. Так, по способу построения текста методы автоматического реферирования и аннотирования делятся на две группы: **извлекающие** (методы на основе машинного обучения и методы на основе теории графов), где из текста выделяются наиболее важные исходные фрагменты, и **генерирующие**, способные создавать новый текст, не представленный явно в тексте исходного документа.

Основными типами методов являются **статистические** и **гибридные** (статистические методы обработки документов дополняются одной или несколькими лингвистическими процедурами и лингвистическими базами знаний различной глубины). Особенности статистического подхода являются универсальность алгоритмов извлечения ключевых слов и отсутствие необходимости в трудоемких процедурах построения лингвистических баз знаний. Но качество результата использования такого метода в за частую неудовлетворительный. Данный подход применим к языкам с бедной морфологией.

К числу гибридных методов извлечения ключевых слов можно отнести методы на основе машинного обучения, где задача извлечения ключевых слов рассматривается как задача классификации. Среди методов на основе машинного обучения можно отметить: байесовские методы; метод опорных векторов; деревья решений; использование нейронных сетей.

Среди современных методов реферирования выделяют:

1) по типу получаемого реферата:

— **экстракцию** (извлечение информационных блоков, Sentence Extraction, квазиреферирование). Все существующие системы автоматического реферирования промышленного масштаба реализованы в рамках данного подхода. Недостаток: выбранные информационные блоки никак не связаны между собой.

— **абстракцию** (извлечение содержания, Content Extraction) - генерация реферата с порождением нового текста, содержательно обобщающего первичный документ или документы. Данный подход характеризуется тремя неотъемлемыми этапами: анализ исходного текста с генерацией внутреннего представления, семантическое сжатие внутреннего представления и синтез нового текста (реферата). В рамках данного подхода можно выделить два основных направления: *абстракция на основе лингвистического сжатия и с опорой на знания*.

2) по уровню анализа исходного текста:

— **поверхностный уровень** (Surface-level approach). Подходы данного уровня характеризуются поверхностным анализом текста. Предложения рассматриваются как линейные последовательности слов (в некоторых случаях, словосочетаний), которые несвязно собираются в текст.

— **уровень сущностей текста** (Entity-level approach). Здесь детально рассматривают модель текста, его элементы - «сущности», их

взаимосвязи, совместную встречаемость, кореферентность, анафорические отношения и т. д., опираясь на модель структурной связности текста (когезию).

— **уровень дискурсной структуры текста** (Discourse-level approach) предполагает качественно новый уровень анализа исходного текста, его глобальной структуры, его связи с коммуникативными целями. В основе данного подхода лежит анализ содержательной модели связности текста (когеренции). Когеренция представляет собой семантико-прагматические аспекты смысловой и деятельностной (интерактивной) связности дискурса как локальной, так и глобальной.

3) по критерию использования опоры на знания:

— **методы без опоры на знания** (Knowledge-poor approach) не предполагают создания специальных баз знаний какой-либо предметной области с целью понимания смысла текста на естественном языке. К данному направлению относится большинство существующих на текущий момент методов извлечения предложений и ряд методов абстракции на основе лингвистического сжатия.

— **методы с опорой на знания** (Knowledge-rich approach). Предполагают использование специальных баз знаний, содержащих наборы правил и эвристик какой-либо специфичной предметной области для формирования реферата.

4) по технологии построения реферата:

— **подходы «сверху-вниз»** (Top-down approaches) на основе извлечения информации предполагают трансформацию исходного текста в некоторое более содержательное внутреннее представление, а затем манипуляцию с этим представлением и формирование итогового реферата.

— **подходы «снизу-вверх»** (Top-down approaches) используют методику информационного поиска и пытаются выделить релевантные фрагменты из исходного текста и сформировать из них итоговый реферат.

5) по ориентации на предметную область:

— **подходы без ориентации на предметную область** (Domain-independent approaches).

— **подходы с ориентацией на конкретную предметную область** (Domain-dependent approaches).

На данном этапе развития автоматического реферирования и аннотирования классификация современных методов выглядит так, довольно условная. Каждый метод сочетает в себе комбинацию нескольких подходов разных направлений с некоторыми нововведениями. Применимы они к различного типа текстам различных языков.

#### **Использованные источники:**

1. Шереметьева С.О., Осминин П.Г. Методы и модели автоматического извлечения ключевых слов // Вестник ЮУрГУ. Серия: Лингвистика. 2015. №1. [Электронный ресурс] – URL: <https://cyberleninka.ru/article/n/metody-i-modeli-avtomaticheskogo-izvlecheniya-klyuchevykh-slov>.
2. Тарасов С.Д. Современные методы автоматического реферирования // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика, телекоммуникации и управление. 2010. №6 (113). [Электронный ресурс] – URL: <https://cyberleninka.ru/article/n/sovremennye-metody-avtomaticheskogo-refirovaniya>.