**BC/NW 2023 № 1 (40):8.2**

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ НА ПРИМЕРЕ РУССКОЯЗЫЧНЫХ ТЕКСТОВ**

Челышев Э. А., Раскатова М.В

Накопленный человечеством объем данных неуклонно возрастает. Так, например, к 2025 году прогнозируемый объем данных составит уже 175 зеттабайтов [1]. По этой причине все более ощущается потребность в средствах автоматической обработки информации, в частности программных средствах и методах автоматического реферирования.

Реферированием текстов называется процесс подготовки реферата, то есть текста меньшего объема, чем исходный текст, позволяющего при этом понять суть излагаемого материала [2].

Целью данной работы является проведение сравнительного анализа алгоритмов автоматического реферирования на примере русскоязычных текстов.

В работе рассматриваются следующие алгоритмы автоматического реферирования текстов: алгоритм Луна, TextRank, LexRank и LSA (англ. Latent Semantic Analysis, рус. латентный семантический анализ). С использованием каждого из методов были подготовлены рефераты ряда русскоязычных текстов. Качество полученных рефератов было оценено при помощи таких метрик, как расстояние Дженсена-Шеннона и косинусное расстояние [3]. В работе обосновывается выбор данных метрик, а также рассматриваются альтернативные способы оценки качества получаемых рефератов.

В данном докладе представлены результаты сравнительного анализа ряда методов автоматического реферирования применительно к текстам на русском языке с использованием таких метрик, как расстояние Дженсена-Шеннона и косинусное сходство.

Литература

1. Reinsel, D. The Digitalization of the World / D. Reinsel, J. Gantz, J. Rydning. – 2018. – 28 с.: [Электронный ресурс]. – URL: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. (Дата обращения: 11.10.2022).

2. Батура, Т. В. Методы и системы автоматического реферирования текстов / Т. В. Батура, А. М. Бакиева. – Новосибирск : Новосибирский национальный исследовательский государственный университет, 2019. – 110 с. – ISBN 978-5-4437-0974-1.

3. Брюховецкий, А. А. Модель обнаружения аномальных данных на основе информационного критерия / А. А. Брюховецкий // Дневник науки. – 2021. – № 4(52). – EDN UOQDOY.