

### ОТЗЫВ

**официального рецензента на диссертационную работу Ахметова Искандера  
на тему «Разработка модели информативного реферирования научных текстов на английском языке, предоставленную на соискание  
степени доктора философии (PhD) по специальности 6D075100 – «Информатика, Вычислительная Техника и Управление».**

№п/п	Критерии	Соответствие критериям (необходимо отметить один из вариантов ответа)	Обоснование позиции официального рецензента
1.	Тема диссертации (на дату ее утверждения) соответствует направлениям развития науки и/или государственным программам	1.1 Соответствие приоритетным направлениям развития науки или государственным программам:  3) Диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан (указать направление)	Диссертационная работа соответствует приоритетному направлению развития науки «информационные, коммуникационные и космические технологии»
2.	Важность для науки	Работа вносит существенный вклад в науку, а ее важность хорошо раскрыта	В работе выполнен всесторонний сравнительный анализ существующих систем автоматического реферирования текста. Наглядность и убедительность в оценке новизны полученных моделей придают численные эксперименты над разрабатываемым алгоритмом, а также оценка его качества и сложности. Модель автоматического экстрактивного реферирования научных текстов на основе жадного алгоритма вносит существенный вклад в науку и имеет высокое практическое значение, повышая эффективность автоматического создания краткого изложения научных текстов.
3.	Принцип самостоятельности	Уровень самостоятельности: 1) Высокий;	Уровень самостоятельности автора диссертационной работы заключается в анализе литературных данных, выполнении экспериментальной части работы, обобщении и интерпретации полученных экспериментальных данных и выводов.
4.	Принцип внутреннего единства	4.1 Обоснование актуальности диссертации: 1) Обоснована;	Исследование и разработка подходов к реферированию документов имеют высокую степень актуальности и неопределимое значение для облегчения жизни людей, поскольку избыточные данные могут усложнять и замедлять поиск необходимой информации.
		4.2 Содержание диссертации отражает тему диссертации: 1) Отражает;	В диссертационной работе предложена модель автоматического экстрактивного реферирования научных текстов на основе жадного алгоритма, что полностью отражает тему диссертации.

*Ахметов*

		<p>4.3. Цель и задачи соответствуют теме диссертации: 1) соответствуют;</p>	<p>Основной целью и идеей диссертационной работы является разработка модели для информативного реферирования научных текстов на английском языке, которая поможет экономить время профессионалам, работающим с информацией, сокращая объем информации для переработки, и одновременно сохраняя информационную ценность текста, что соответствует теме диссертации.</p>
		<p>4.4 Все разделы и положения диссертации логически взаимосвязаны: 1) полностью взаимосвязаны;</p>	<p>Диссертационная работа является завершенным научным исследованием, состоит из введения, обзора литературы, основной части, заключения, списка используемой литературы и двух приложений. Структура изложения исследовательского материала в диссертации логически правильно выстроена.</p>
		<p>4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями: 1) критический анализ есть;</p>	<p>В диссертационной работе выполнен всесторонний сравнительный анализ существующих систем автоматического реферирования текста.</p>
5.	Принцип научной новизны	<p>5.1 Научные результаты и положения являются новыми? 2) частично новые (новыми являются 25-75%);</p>	<p>Степень новизны каждого научного результата, сформулированного в диссертации в том, что диссертация содержит новые результаты по исследуемой теме, и в целом, в решении основных прикладных задач.</p>
		<p>5.2 Выводы диссертации являются новыми? 2) частично новые (новыми являются 25-75%);</p>	<p>Результаты и выводы подтверждаются результатами вычислительных экспериментов по разработанным методам, алгоритмам и моделям.</p>
		<p>5.3 Технические, технологические, экономические или управленческие решения являются новыми и обоснованными: 1) полностью новые;</p>	<p>Наглядность и убедительность в оценке новизны полученных моделей придают численные эксперименты над разрабатываемым алгоритмом, а также оценка его качества и сложности.</p>
6.	Обоснованность основных выводов	<p>Все основные выводы основаны на весомых с научной точки зрения доказательствах либо достаточно хорошо обоснованы (для qualitative research и направлений подготовки по искусству и гуманитарным наукам)</p>	<p>Достоверность полученных результатов подтверждается верификацией проведенных экспериментов с помощью общепринятых метрик, публикацией в рейтинговых журналах, имеющих ненулевой импакт-фактор по данным базы Scopus и/или Web of Science, а также выступлением на международных конференциях (ICVNS 2021, LKE 2021, MICA1 2021, ICECO 2021).</p>
7.	Основные положения, выносимые на защиту	<p>Необходимо ответить на следующие вопросы по каждому положению в отдельности: Положение 1. Метод эвристической оценки уровня качества реферата метрикой ROUGE-1.  7.1 Доказано ли положение? 1) доказано; 7.2 Является ли тривиальным?</p>	<p>Положение 1. Представлено, что метод эвристической оценки уровня качества реферата метрикой ROUGE-1, достижимого при помощи экстрактивных методов автоматического реферирования текстов дает результат в 0.59, что значительно выше текущего уровня в 0.46 у самых современных методов использующих нейронные сети. На данный период полученный практический результат оценки является наилучшим.</p>

- 2) нет  
7.3 Является ли новым?  
1) да;  
7.4 Уровень для применения:  
3) широкий  
7.5 Доказано ли в статье?  
1) да;

Положение 2.  
Разработан метод Автоматического Экстрактивного Реферирования (АЭР) GreedSum.

- 7.1 Доказано ли положение?  
1) доказано;  
7.2 Является ли тривиальным?  
2) нет  
7.3 Является ли новым?  
1) да;  
7.4 Уровень для применения:  
3) широкий  
7.5 Доказано ли в статье?  
1) да;

Положение 3.  
Техника тонкой настройки управляющего параметра `minimum document frequency (min_df)` работы GreedSum.

- 7.1 Доказано ли положение?  
2) скорее доказано;  
7.2 Является ли тривиальным?  
2) нет  
7.3 Является ли новым?  
1) да;  
7.4 Уровень для применения:  
3) широкий  
7.5 Доказано ли в статье?  
1) да;

Положение 2.

Доказан разработанный метод Автоматического Экстрактивного Реферирования (АЭР) GreedSum, который показывает результат 0.43 по метрике ROUGE-1 на дата сете arXiv. Приведены примеры экспериментов в приложении.

Замечания и рекомендации по данному приложению:

- 1) Нужно было бы показать при входе в систему выполняется кластеризация текстов или разработанный метод не зависит от структуры и типа данных. И может ли разработанный метод применен для слабоструктурированных датасетах ?
- 2) Нужно было показать разработанный метод Автоматического Экстрактивного Реферирования (АЭР) GreedSum является ли языкозависим и может ли адаптирован для других типов языков.

Указанные замечания не затрагивают основных положений, выводов и научных результатов работы, и несут рекомендательный характер для улучшения диссертационной работы.

Положение 3.

Представлена техника тонкой настройки управляющего параметра `minimum document frequency (min_df)` работы GreedSum, или минимальная частота вхождения слов в предложения реферируемого текста для их учета в создании словаря для построения TFIDF матрицы. На выборке в 376 текстов из датасета arXiv путем простого перебора оптимальное значение `min_df` было определено как 0.042 (т.е. слово должно появляться как минимум в 4.2% предложениях).

8.	Принцип достоверности Достоверность источников и предоставляемой информации	8.1 Выбор методологии - обоснована ли методология, достаточно ли подробно описана 1) да;	В работе подробно описана методология, а именно гетерогенные наборы данных, используемые в разработке моделей автоматического реферирования; метрики оценки качества; методы автоматического реферирования, а также численные эксперименты над разрабатываемым алгоритмом.
		8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий: 1) да;	Эксперименты выполнены при помощи языка программирования Python 3.6 (Pandas, Numpy).
		8.3 Теоретические выводы, модели, выявленные взаимосвязи и закономерности доказаны и подтверждены экспериментальным исследованием (для направлений подготовки по педагогическим наукам результаты доказаны на основе педагогического эксперимента): 1) да;	Глава 3 диссертации посвящена численным экспериментам над разрабатываемым алгоритмом, оценке его качества и сложности, а также обсуждению вопроса его практического применения.
		8.4 Важные утверждения подтверждены/частично подтверждены/не подтверждены ссылками на актуальную и достоверную научную литературу	Диссертационная работа содержит библиометрический анализ выбранной темы, сравнительный анализ моделей автоматического реферирования текстов, обзор существующих информационных систем автоматического реферирования текстов, а также оценку верхнего предела качества экстрактивных алгоритмов суммаризации в литературе.
		8.5 Используемые источники литературы достаточны/не достаточны для литературного обзора	Используемые источники литературы достаточны для литературного обзора
9	Принцип практической ценности	9.1 Диссертация имеет теоретическое значение: 1) да;	Выполнен анализ современных подходов к автоматическому реферированию текстов. В ходе исследования определено, что для текстов на английском языке практически реализуемыми являются методы, основанные на подходе экстракции.
		9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике: 1) да;	Предложенная технология и модель экстрактивного реферирования научных текстов на основе жадного алгоритма весьма конкурентоспособна по своим основным техническим характеристикам и может быть реализовано в широком диапазоне информационно-технических инфраструктур и платформ.
		9.3 Предложения для практики являются новыми? 1) полностью новые;	Предложенная модель может быть использована в науке (автоматизация обзора литературы, генерация аннотации статьи, обновление научной информации и др.), образовании (автоматическое создание заметок, презентаций, генерация тестовых вопросов), бизнесе (формирование протокола собрания, резюме больших объемов текста и др.)

*Виктор*

