

Отзыв

на выпускную квалификационную работу бакалавра

Халиуллиной Лии Рауфовны

“Поиск дубликатов среди документов”

В выпускной квалификационной работе бакалавра рассматривается задача построения автоматизированной системы поиска дубликатов среди документов.

В работе представлены принципы работы существующих алгоритмов сравнения текстов: алгоритмы *sif* и *Koala*, *Long Sent*, метод «шинглов», методы на основе меры *TF-IDF*, *Heavy Sent*, *I-Match*, метод описательных слов. Для рассмотренных методов представлены результаты сравнения эффективности работы. На основе анализа специфики построения научных текстов сформулирован новый «модифицированный» метод поиска дубликатов.

Кроме того, в работе проводится анализ результатов, полученных применением различных методов к задаче поиска дубликатов курсовых работ. В результате экспериментов автором выявлено, что модифицированный метод учитывает стиль документов и их длину, но не выявляет цитирование и сноски на литературу.

К недостаткам работы можно отнести:

1. Предложенный автором модифицированный метод строится на основе специфики построения научных текстов, а именно, сводится к сопоставлению разделов документов. При этом ни в теме работы ни в постановке задачи не упоминается, что работа ведется лишь с научными текстами.
2. Описание проводимого эксперимента недостаточно. Указывается лишь информация о 15 курсовых работах и их копиях. Но не указано количество плагиата для каждого документа. В таком случае невозможно оценить эффективность методов.
3. Не приведены результаты работы существующих систем антиплагиата (например, <https://text.ru/antiplagiat>)

В целом работа заслуживает оценки “хорошо”.

Научный руководитель,
ст. преподаватель

Малинина М.А.