

РЕЦЕНЗИЯ
на выпускную квалификационную работу обучающегося СПбГУ
Ашихминой Алёны Сергеевны по теме

**«Определение семантической связности юридических документов на
примере договоров»**

Выпускная квалификационная работа Ашихминой Алёны Сергеевны посвящена анализу семантической связности договоров на сайте государственных закупок с целью нахождения опечаток и несоответствий. Это безусловно актуальная практическая задача.

В первой части работы автор описывает имеющиеся данные и приводит способы их обработки, такие как, разделение документа на логические части. Приводится описание известного алгоритма TextTiling и утверждается, что этот алгоритм применим в случае отсутствия разметки документа по пунктам. Далее автор ставит задачу и готовит тестовые данные: в имеющиеся документы подмешиваются блоки текста из других документов. Задача сводится к поиску инородных блоков в тестовой выборке.

Далее приводится описание четырех разработанных автором алгоритмов поиска:

1. На основе известной тематической модели LDA.
2. На основе векторного представления слов.
3. С использованием собственной тематической модели.
4. С использованием явной маркировки блоков текста.

Для предложенных алгоритмов приводятся критерии их релевантности: точность и полнота.

К достоинствам работы можно отнести актуальность решаемой проблемы, приемлемую точность работы алгоритмов, наличие качественных схем и иллюстраций.

В то же время результате ознакомления с выпускной квалификационной работой следует сделать замечания:

1. Обзор литературы мог бы быть более полным.
2. Автор искусственно конструирует «аномальные» документы, не совсем понятно может ли подобный подход работать для реальных документов.

Несмотря на имеющиеся замечания, считаю, что работа удовлетворяет требованиям к ВКР бакалавра и заслуживает оценки **«отлично»**.

Рецензент:

к.ф-м.н., м.н.с.

ФТИ им. Иоффе РАН

Кац Владимир Наумович