

Рецензия на рукопись
«Автоматическое выделение терминов для тематического моделирования»
Никитина М. А.

1. Новизна, актуальность и обоснованность работы

Четко сформулирована цель работы. Понятна и ясна ее актуальность. Указано, в чем заключается новизна. Тем не менее, стоит привести дополнительные ссылки на предшествующие исследования данной проблемы.

2. Список ошибок, недочетов и замечаний

Аннотация

Раздел написан качественно и понятно. Опечатка: узукоспециализированных \rightarrow узкоспециализированных.

Введение

Абзац 4, предложение «Нейросеть использована готовая» стоит пояснить подробнее или заменить. Абзац 6, предложение «Предшествующие исследования предлагаемого подхода ...» стоит дополнить ссылкой на соответствующие работы. Здесь и далее вместо короткого тире -- нужно использовать неразрывное длинное \sim ---. Опечатка: эмбединг \rightarrow эмбеддинг.

Постановка задачи

Задача поставлена четко, ясна проблема и мотивация.

Вычислительный эксперимент

Абзац 2, предложение «Для проведения первого эксперимента из статей удаляются числа, заголовки и ссылки на литературу ...», стоит пояснить, вручную это делается, или же написан скрипт. Рассматривается поиск терминов из одного слова. Далее проводится анализ всех терминов, но непонятно, есть ли ограничение по количеству слов в них. Абзац 3, предложение «Так как датасет состоит из статей, написанных на английском языке ...», не до конца понятно, как влияет язык текста на использование стемминга. Для ссылок на таблицы и графики стоит использовать неразрывный пробел: График $\sim \backslash \text{ref}\{\text{label}\}$.

Теоретическая часть

Теоретическое описание метода должно идти до вычислительного эксперимента. Опечатки: под формулой (8) исправить $\alpha td \rightarrow \alpha_{td}$, в формуле (9) также исправить $\varphi \omega s \rightarrow \varphi_{\omega s}$.

Анализ ошибки

В Таблице 2 во всех столбцах кроме «Порог» используются точки в качестве разделителей для десятичных дробей, поэтому и в данном столбце следует использовать их. В Таблицах 2 и 3 в пределах одного столбца приведены числа с разным количеством цифр после разделяющей точки, лучше добавить нули в конце, например: 0.1 → 0.100. Абзац 2, стоит заменить «от худшего до лучшего» → «от худшего к лучшему». Опечатка: в Абзаце 1 исправить Resqll → Recall.

Заключение

Работа находится в стадии завершения, поэтому можно ожидать скорого написания данного раздела.

3. Комментарии к коду

Код написан аккуратно, четко и понятно. Приведены подробные комментарии к функциям и конкретным фрагментам кода. Легко интерпретируется и расширяется.

4. Общее мнение о работе

Работа выполнена качественно и удовлетворяет всем требованиям, предъявляемым к статье. Читается легко, все необходимые термины введены по ходу повествования.

Рецензент:

Киселев Н. С.