

Университет ИТМО, факультет программной инженерии и компьютерной техники
Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата прошлой лекции	Номер прошлой лекции	Название статьи/главы книги/видеолекции	Дата публикации	Размер статьи	Дата сдачи
10.09.25	1	Использование изменения системы отсчета для улучшения результатов анализа с помощью закона бенфорда	31.03.2025	1208	24.09.25
24.09.25	2	Метод регенерационного блочного кодирования	13.10.2023	3356	08.10.25
08.10.25	3	Многоцелевая фильтрация текста с использованием регулярных выражений	10.06.2024	2124	05.11.25
22.10.25	4	Формат PARQUET: Обзор структуры и методы оптимизации для высокой производительности в распределённых системах	11.04.2025	3651	05.11.25
05.11.25	5	Классификация определений в математических LaTeX статьях	17.05.2024	3471	22.11.25
19.11.25	6				
03.12.25	7				

Выполнил(а) Бых Даниил Максимович, № группы P3109, оценка _____

Прямая полная ссылка на источник или сокращённая ссылка

<https://cyberleninka.ru/article/n/klassifikatsiya-opredeleniy-v-matematicheskikh-latex-statyah/viewer>

Теги, ключевые слова или словосочетания

Обработка естественного языка, LaTeX, математические тексты.

Перечень фактов, упомянутых в статье

- 1) Математический текст сильно структурирован, символы и слова меняют смысл в зависимости от контекста.
- 2) Большинство математических статей создаются в формате LaTeX.
- 3) Для распознавания определений в LaTeX текстах построен классификатор на основе машинного обучения с векторными представлениями.
- 4) Лучший результат показала модель MathBERT с градиентным бустингом, обученная на корпусе математических текстов.

Позитивные следствия и/или достоинства описанной в статье технологии

- 1) MathBERT повышает качество распознавания определений в математических текстах.
- 2) Автоматизация классификации облегчает поиск и извлечение знаний из научных баз.
- 3) Машинное обучение позволяет визуализировать структуру и семантику математических текстов.

Негативные следствия и/или достоинства описанной в статье технологии

- 1) Классы определений и не определений слабо разделены в векторном представлении, что усложняет обучение.
- 2) Малое количество размеченных данных снижает эффективность нейросетевых моделей.
- 3) Универсальные текстовые модели плохо работают с математическими текстами, требуя специализированных решений.

Ваши замечания, пожелания преподавателю или анекдот о программах

:)