

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

Дата прошедшей лекции	Номер прошедшей лекции	Название статьи/главы книги/видеолекции	Дата публикации (не старше 2022 года)	Размер статьи (от 400 слов)	Дата сдачи
10.09.2025	1	Non-Standard Binary Representations and the Stern Sequence	14.08.2023	~3,800	24.09.2025
24.09.2025	2	LLM-Guided Search for Deletion-Correcting Codes	01.04.2025	~10000	08.10.2025
08.10.2025	3	Intelligent Interface: Enhancing Lecture Engagement with Didactic Activity Summaries	20.06.2024	~7500	22.10.2025
22.10.2025	4	Advancing vision-language models in front-end development via data synthesis	03.03.2025	~12000	05.11.2025
05.11.2025	5	Spreadsheet Quality Assurance: A Literature Review	22.01.2024	~7,500	19.11.2025
19.11.2025	6	NeuRaLaTeX: A machine learning library written in pure LaTeX	31.05.2025	~8500	03.12.2025
04.12.2025	7	Inconsistencies in TeX-Produced Documents	03.09.2024	~9,500	18.12.2025

Выполнил(а) Эллити Мохамед Эмад Ахмед Авад, № группы P3131, оценка                       
Фамилия И.О. студента не заполнять

**Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.)**

**Прямая ссылка на источник:** <https://arxiv.org/abs/2407.15511>

**Теги, ключевые слова или словосочетания (минимум три слова):**

TeX programming, LaTeX engines, Document reproducibility, PDF consistency, Cross-version analysis

**Перечень фактов, упомянутых в статье (минимум четыре пункта)**

- Авторы провели обширное исследование различий в PDF-документах, сгенерированных разными TeX-движками (pdfTeX, XeTeX, LuaTeX) и версиями TeX Live.
- Были выявлены типы несоответствий: пробелы между словами, разрывы строк, различия в шрифтах, изображения, ссылки и количество страниц.
- Исследование охватило 342 документа, включая разные классы документов: article, amsart, revtex, IEEEtran, elsarticle и др.
- Проведен анализ корневых причин ошибок и несоответствий, включая синтаксические ошибки автора, поведение пакетов LaTeX и обновления движков TeX Live.

**Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)**

- Обеспечивается глубокое понимание различий между движками TeX и версиями TeX Live, что важно для воспроизводимых исследований.
- Позволяет выявлять потенциальные ошибки компиляции и неконсистентность форматов документов заранее.
- Подход демонстрирует возможности автоматизированного тестирования PDF-документов с помощью инструментов diff-pdf, PyMuPDF и OpenCV.

**Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)**

- • Компиляция и анализ большого числа документов требует значительных вычислительных ресурсов.
- Различия между движками и версиями TeX могут создавать неожиданные проблемы при подготовке публикаций.
- Метод выявляет проблемы, но не всегда предлагает способы их автоматического исправления.
- Сложность воспроизведения экспериментов при изменении пакетов или обновлении TeX Live.

**Ваши замечания, пожелания преподавателю или анекдот о программистах**

“Same document, different compile.”

(Особенно когда pdfTeX и LuaTeX создают разные PDFs 😊)