

МЕТОДИ ТА ЗАСОБИ КОНВЕРТАЦІЇ ДОКУМЕНТІВ LATEX В АДАПТИВНИЙ КОНТЕНТ HTML5 ДЛЯ ВЕБСИСТЕМ

Вінницький національний технічний університет

Анотація

В роботі розглядається сучасні підходи до перетворення документів, підготовлених у системі верстки LaTeX, у формат HTML5, що є критично важливим для створення доступного та адаптивного наукового веб-контенту. Основна увага приділяється аналізу функціональних можливостей таких інструментів, як MathJax, KaTeX, LaTeXML та Pandoc. Досліджуються методи відображення математичних формул, забезпечення доступності та підтримка стандартів, зокрема MathML.

Ключові слова: LaTeX, KaTeX, LaTeXML, Pandoc, MathML, HTML5, вебсистема.

Abstract

This work examines modern approaches to converting documents prepared in the LaTeX typesetting system into HTML5 format, which is critically important for creating accessible and responsive scientific web content. The primary focus is on analyzing the functional capabilities of tools such as MathJax, KaTeX, LaTeXML, and Pandoc. The research explores methods for displaying mathematical formulas, ensuring accessibility, and supporting standards, particularly MathML.

Keywords: LaTeX, KaTeX, LaTeXML, Pandoc, MathML, HTML5, websystem.

Вступ

Сучасна наукова документація вимагає доступності контенту на різних пристроях. Традиційний формат PDF, хоча і є стандартом для друку, проте має суттєві обмеження у вебсередовищі: він не є адаптивним, важко індексується пошуковими системами та має низький рівень доступності для людей з порушеннями зору. Використання LaTeX як джерела для генерації HTML5-контенту є актуальним вирішенням цієї проблеми.

Результати досліджень

Для ефективного відображення наукового контенту у вебсередовищі необхідно здійснювати перетворення вихідних LaTeX-файлів у формати, сумісні зі стандартами W3C, зокрема HTML5 та MathML. Основним викликом при такому переході є збереження складної ієрархічної структури документа та математичної точності [1]. Одним із найпоширеніших засобів розв'язання цієї задачі є модульний JavaScript-рушій, що забезпечує високоякісну типографіку в усіх сучасних браузерах – MathJax [2]. Завдяки використанню векторних шрифтів та SVG-графіки, MathJax дозволяє математичним виразам залишатися чіткими при будь-якому рівні масштабування сторінки, що є важливим для мобільних пристроїв [2]. Його важливою перевагою є здатність обробляти макроси користувача, що дозволяє переносити складні авторські конструкції безпосередньо у вебконтент.

Для високонавантажених вебсистем та платформ із великим обсягом даних, де важливим фактором є швидкість відгуку інтерфейсу, доцільно використовувати бібліотеку KaTeX. На відміну від інших рішень, KaTeX виконує рендеринг синхронно, що запобігає «стрибкам» контенту під час завантаження сторінки [7]. Застосування технології серверного рендерингу (SSR) у поєднанні з Node.js дозволяє передавати клієнту вже сформовану HTML-розмітку, що значно прискорює та покращує SEO-оптимізацію наукових ресурсів. Це робить KaTeX ідеальним вибором для освітніх онлайн-платформ та інтерактивних підручників.

Глибокий семантичний аналіз документа реалізовується в системі LaTeXML, яка була розроблена як засіб подолання труднощів при перетворенні документів із презентаційно-орієнтованого формату LaTeX у контентно-орієнтований XML для вебпредставлення. Цей інструмент не просто копіює зовнішній вигляд, а намагається відтворити логічну структуру документа через проміжну XML-схему, що дозволяє зберігати семантичні та презентаційні підказки [6]. Такий підхід дозволяє створювати

«розумні» документи, де математичні формули не лише візуалізуються, а й піддаються граматичному аналізу за допомогою парсера, який виводить структуру виразів. Внутрішнє представлення XMath дозволяє зберігати паралельну розмітку: презентаційний MathML для відображення та контентний MathML або OpenMath для подальших обчислень у системах комп'ютерної алгебри [3].

Універсальність та потужність інструменту Pandoc у конвертації наукових документів базується на його унікальній архітектурі, в центрі якої лежить абстрактне синтаксичне дерево (AST). На відміну від простих конвертерів, які лише замінюють одні теги іншими, Pandoc спочатку повністю розбирає вихідний LaTeX-код і будує його логічну модель у пам'яті, що не залежить від кінцевого формату [5]. Це дозволяє програмі «розуміти» структуру документа: де знаходиться математична формула, де таблиця, а де посилання на літературу. Завдяки такому підходу процес трансформації в HTML5 стає інтелектуальним, оскільки розробник може втручатися в структуру документа на етапі його перебування, наприклад, автоматично змінюючи рівні заголовків або додаючи специфічні класи для адаптивного дизайну. Особливу роль відіграє вбудована система фільтрів та підтримка механізму citerproc, який автоматизує роботу з бібліографічними даними [4]. У середовищі LaTeX цитування зазвичай керуються через BibTeX, що є досить специфічним для наукового ПЗ, проте Pandoc здатний перетворити ці записи на стандартні вебелементи. Під час генерації HTML5-контенту програма автоматично створює зв'язки між текстом і списком літератури, перетворюючи статичні текстові посилання на інтерактивні внутрішні гіперпосилання. Це докорінно змінює досвід користувача, адже лінійний документ перетворюється на повноцінну гіпертекстову систему, де читач може миттєво перейти до джерела або змісту[5].

Додатковим фактором успіху адаптивної візуалізації є впровадження стандарту MathML Core. Сучасні браузері, такі як Chrome, Firefox, Safari тепер нативно підтримують відображення математики, що дозволяє значно зменшити обсяг JavaScript-коду, який завантажується на пристрій користувача [3]. Комбінація семантичного HTML5, адаптивних CSS-сіток (Grid та Flexbox) та MathML дозволяє науковим статтям автоматично перегруповувати контент. Наприклад, двоколонкові макети журналів трансформуються в одну колонку на смартфонах без втрати читабельності формул чи таблиць [3]. Таким чином, перехід від власницьких та статичних форматів до відкритого вебу на основі LaTeX стає ключовим етапом у забезпеченні глобальної доступності наукових знань.

Висновки

В результаті дослідження методів та засобів конвертації документів LaTeX в адаптивний HTML5-контент було зроблено висновок, що перехід від статичного формату PDF до гнучких вебстандартів є необхідним кроком для сучасної наукової документації. Аналіз показав, що вибір інструментів безпосередньо залежить від технічних вимог до кінцевої системи, де MathJax забезпечує найвищу якість типографіки для складних авторських макросів, тоді як KaTeX виступає незамінним рішенням для високонавантажених платформ завдяки синхронному рендерингу та відсутності візуальних стрибків контенту. Паралельно з цим система LaTeXML дозволяє здійснювати глибокий семантичний аналіз і зберігати логіку документа для подальших обчислень, а архітектура Pandoc на основі абстрактного синтаксичного дерева трансформує лінійні тексти у складні гіпертекстові середовища з інтерактивною бібліографією. Впровадження стандарту MathML Core у поєднанні з адаптивними можливостями HTML5 та сучасними методами верстки дозволяє не лише адаптувати науковий контент до будь-яких пристроїв, а й створити інклюзивне середовище, доступне для людей із порушеннями зору. Таким чином, комплексна цифрова трансформація наукових документів на основі LaTeX забезпечує інтерактивність та глобальну доступність академічних знань у сучасному інформаційному просторі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Math Tools. [Електронний ресурс] – Режим доступу: https://www.w3.org/wiki/Math_Tools (дата звернення: 07.01.2026).
2. MathJax. [Електронний ресурс] – Режим доступу: <https://www.mathjax.org/> (дата звернення: 07.01.2026).
3. MathML. [Електронний ресурс] – Режим доступу: <https://developer.mozilla.org/en-US/docs/Web/MathML> (дата звернення: 07.01.2026).
4. Pandoc filters. [Електронний ресурс] – Режим доступу: <https://pandoc.org/filters.html> (дата звернення: 07.01.2026).
5. Pandoc User's Guide. [Електронний ресурс] – Режим доступу: <https://pandoc.org/MANUAL.html> (дата звернення: 07.01.2026).
6. LaTeXML The Manual. [Електронний ресурс] – Режим доступу: <https://math.nist.gov/~BMiller/LaTeXML/manual.pdf> (дата звернення: 07.01.2026).

7. KaTeX. The fastest math typesetting library for the web. [Електронний ресурс] – Режим доступу: <https://katex.org/> (дата звернення: 07.01.2026).

Бусигіна Вероніка Павлівна – студентка групи ІПКТ-24б, кафедра автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: veronikabysygina@gmail.com

Богач Ілона Віталіївна – к.т.н., доцент, професор кафедри автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: ilona.bogach@gmail.com

Busyhina Veronika Pavlivna – student of 1PCT-24b group, Department of Automation and Intelligent Information Technologies, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: veronikabysygina@gmail.com

Bogach Ilona Vitaliivna – PhD, Associate Professor of Automation and Intelligent Information Technologies, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: ilona.bogach@gmail.com