

МАТЕМАТИЧЕСКИЕ ПУБЛИКАЦИИ В ФОРМАТЕ PDF С УСЛОВИЯМИ РАВНОЙ ДОСТУПНОСТИ

Б. М. Дубров

*Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, doubrov@bsu.by*

Публикация научных статей и учебных материалов в формате, позволяющем специальным возможностям операционных систем (Assistive Technologies) озвучивать содержание статей, включая математические формулы, для людей с ограниченным зрением остается одной из проблемных областей. В данной работе мы рассматриваем различные международные стандарты цифровых форматов, нацеленные на решение этой проблемы, а также новую версию пакета LaTeX, которая включает поддержку генерации PDF файлов с корректной семантической разметкой и математическими формулами в формате MathML.

Ключевые слова: специальные возможности; математические публикации; LaTeX; международные стандарты PDF/UA; WCAG.

Международные стандарты в области равной доступности электронных документов

Условия равной доступности электронных материалов, включая как веб сайты, так и электронные документы в формате PDF, являются одним из ключевых требований к электронным материалам. Этот вопрос регулируется на уровне законодательств большинства государств, включая Республику Беларусь [1, 2]. Мы рассматриваем технологические аспекты этого вопроса.

Для веб-сайтов ключевыми технологиями, обеспечивающими условия равной доступности, являются стандарт ARIA [3], позволяющий описывать семантику элементов на странице, а также стандарт WCAG 2.2, формализующий технические требования к разметке веб страниц [4].

Для электронных публикаций в формате PDF стандарт WCAG 2.2 не дает полного покрытия технических требований, гарантирующих, что специальные возможности (Assistive Technologies – AT), встроенные сегодня во все без исключения операционные системы, включая мобильные платформы, сможет обеспечить доступ к содержанию документа для людей с ограниченными возможностями. В связи с этим разработан международный стандарт PDF/UA [5, 6] (PDF/UA-1 для PDF документов версий 1.7 и ранее и PDF/UA-2 для PDF документов начиная с версии 2.0). Этот стандарт регламентирует, что все математические формулы в PDF документах должны быть представлены в формате MathML [7].

На данный момент стандарт PDF/UA-1 поддерживается в качестве формата экспорта как Libre Office, так и Microsoft Office. Однако ни один из этих продуктов не включает в себя поддержку математических формул. Таким образом, на сегодняшний день все математические публикации, сгенерированные с использованием этих программных продуктов, не позволяют AT озвучить математические формулы, которые, по сути, представляют собой растровые изображения или наборы символов без какой-либо дополнительной семантической разметки.

Поддержка семантической разметки PDF документов в LaTeX3

Одним из наиболее востребованных продуктов для генерации научных публикаций является LaTeX. Так, коллекция научных препринтов на портале arXiv.org включает в себя более двух миллионов публикаций в формате LaTeX, которые в автоматическом формате компили-

руются в PDF формат. Однако и эта технология до настоящего времени не предлагала возможностей генерации документов, удовлетворяющих стандарту PDF/UA, что также исключало возможность доступа к этим документам для людей с ограниченными возможностями.

Начиная с 2020 года ведется разработка новой версии LaTeX, получившая кодовое название LaTeX3 [8], которая включает в себя поддержку PDF/UA в качестве результата компиляции. Перечислим ряд возможностей этой новой версии:

- конфигурирование формата вывода: PDF 1.7, PDF 2.0, PDF/A, PDF/UA;
- задание метаданных документа, которые являются обязательными для поддержки ISO стандартов PDF/A и PDF/UA;
- задание альтернативных текстовых описаний растровой графики, что является одним из ключевых требований стандартов WCAG и PDF/UA;
- отображение семантики LaTeX документов в семантику PDF документов, описываемую стандартом PDF 2.0;
- методы конвертации формул LaTeX в MathML.

В данный момент новая версия LaTeX находится в тестовой фазе и доступна в большинстве дистрибутивов LaTeX. Финальный релиз ожидается в 2025 году. Ожидается, что новая версия будет интегрирована как в процессы архива препринтов arXiv.org, так и в большинство основанных на LaTeX редакторских систем международных журналов.

Библиографические ссылки

1. О мерах по совершенствованию использования национального сегмента сети Интернет [Электронный ресурс] / Указ Президента Респ. Беларусь, 1 февраля 2010 г., № 60 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2020.
2. Государственная программа «Социальная защита» на 2021-2025 годы [Электронный ресурс] / утв. постановлением Совета Министров Респ. Беларусь, 21 дек. 2020 г., №748 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2021.
3. ARIA in HTML [Electronic resource] / W3C recommendation. 16 February, 2024. URL: <https://www.w3.org/TR/2024/REC-html-aria-20240216/> (date of access: 05.04.2024).
4. Web Content Accessibility Guidelines (WCAG) 2.2 [Electronic resource] / W3C recommendation. 05 October, 2024. URL: <https://www.w3.org/TR/2023/REC-WCAG22-20231005/> (date of access: 05.04.2024).
5. Document management applications Electronic document file format enhancement for accessibility Part 1: Use of ISO 32000-1 (PDF/UA-1) / ISO 14289-1:2014. 2014-12. 17p.
6. Document management applications Electronic document file format enhancement for accessibility Part 2: Use of ISO 32000-2 (PDF/UA-2) / ISO 14289-2:2024. 2024-03. 42p.
7. ARIA in HTML [Electronic resource] / W3C recommendation. 16 February, 2024. URL: <https://www.w3.org/TR/2024/REC-html-aria-20240216/> (date of access: 05.04.2024).
8. Mittelbach F., Fischer U. Enhancing LaTeX to automatically produce tagged and accessible PDF / TUGboat. 2024. V.45:1. P.139-147.