"Enhancing ePub Accessibility: Automating Compliance with WCAG 2.1 Standards"

Submitted on: **23-2-2024**

Merijn van der Leek
merijn.van.der.leek@student.uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Dr. M.J. (Maarten) Marx
M.J.Marx@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

ABSTRACT

This project aims to enhance the accessibility of electronic publications (ePubs), particularly in compliance with the Web Content Accessibility Guidelines (WCAG) 2.1 standards. By leveraging automated software solutions, we seek to address the accessibility challenges faced by individuals with disabilities when accessing digital reading materials. Through a mixed-methods approach integrating qualitative and quantitative analyses, we will develop and evaluate an automated tool tailored to improve the accessibility of ePubs. The project focuses on technical aspects, emphasizing the importance of digital inclusivity and contributing to advancements in the field of digital accessibility.

1 INTRODUCTION

This project is dedicated to enhancing the accessibility of electronic books, particularly ePub formats, to comply with the Web Content Accessibility Guidelines (WCAG) 2.1. Ensuring universal accessibility for reading materials, especially for individuals with disabilities, is paramount. However, the consistent application of these standards across digital resources is lacking, highlighting a significant area for both research and technological innovation.

The potential of automated tools in assessing and improving the accessibility of ePubs represents a promising avenue for reducing the manual effort required to meet WCAG standards. This project lies at the nexus of computational linguistics and accessibility technology, aiming to narrow the gap between the current state of ePub accessibility and the WCAG 2.1 standards.

Despite the existence of tools like the Accessibility Checker for EPUBs (ACE) for verifying compliance, many ePubs still do not meet the necessary benchmarks. This shortfall underlines the need for further development in automated and semi-automated tools to encourage broader adherence to accessibility standards in digital publications.

Research Questions

This study is driven by a central research question and additional sub-questions to evaluate the effectiveness of automated tools in enhancing the accessibility compliance of ePubs with WCAG 2.1 standards:

Main Research Question:

- What level of effectiveness do automated tools have in ensuring ePub files meet WCAG 2.1 accessibility criteria?

Sub-Questions:

(1) In what ways do automated accessibility enhancements compare with manual processes in terms of time, resources, and effectiveness?
(2) How do various automated enhancement strategies compare in terms of effectiveness, efficiency, and user satisfaction?

This research aims to contribute to the advancement of ePub accessibility technology and to establish an evaluative framework for the effectiveness of these tools in practical applications, thereby significantly enhancing the field of digital accessibility.

2 RELATED WORKS

Recent advancements in web accessibility have underscored the importance of developing and implementing standards and tools that ensure digital content is universally accessible. This effort is crucial for integrating physically challenged individuals into the digital realm, particularly in the context of electronic publications. Studies by Othman et al. [2], Shah [4], and the work by Park et al. [3] highlight significant strides in this domain.

The innovative application of Large Language Models (LLMs) by Othman et al. [2] for automating web accessibility remediation showcases the potential of artificial intelligence in enhancing digital inclusivity. Their research demonstrates the effectiveness of LLMs, such as ChatGPT, in identifying and rectifying web content accessibility flaws, aligning with WCAG 2.1 guidelines and offering scalable solutions to pervasive accessibility challenges.

Shah [4] delves into the automated tools' efficacy in WCAG 2.1 compliance migration, presenting a comprehensive framework for assessing, identifying gaps, and implementing accessibility criteria. This work underscores the transformative potential of automation in maintaining high accessibility standards across digital platforms, emphasizing the critical role of technological advancements in facilitating digital accessibility.

Central to this discourse is the study by Park et al. [3], which addresses the specific challenges faced by physically challenged individuals in accessing electronic books. Their development of an electronic book accessibility standard, based on the EPUB 3.0 accessibility guideline, and the subsequent creation of an automatic accessibility inspection library represent a significant contribution to the field. This study not only establishes a standard document for the Telecommunication Technology Association (TTA) of Korea but also assesses the accessibility of electronic books in the Korean market, offering guidelines for enhancing e-book production with accessibility in mind.

Adding to this discussion, Garrad-Cole [1] explores digital accessibility challenges and inequalities, particularly in the context of pre-arrival platforms for higher education. Highlighting the importance of equitable access to digital technologies, this study provides a comprehensive review of literature on the development and implementation challenges of digital accessibility for physically challenged individuals, emphasizing the critical need for inclusive digital environments.

3 METHODOLOGY

This research adopts a mixed-methods approach, integrating qualitative and quantitative analyses to examine the effectiveness of automated software solutions in enhancing the accessibility of electronic publications (ePubs) in compliance with the Web Content Accessibility Guidelines (WCAG) 2.1 standards. The methodology unfolds through a series of interconnected stages, including data collection, tool development, tool implementation, and a comprehensive evaluation process.

3.1 Data Collection

Before developing the automated tool, a diverse dataset of ePub files will be compiled from various sources. This dataset is crucial for training and testing the tool, ensuring it addresses a wide range of accessibility issues. Potential sources for ePub libraries include open-access platforms such as Gutenberg.org and StandardEbooks.org, which offer a rich repository of files known for their varying levels of accessibility. This selection aims to cover a broad spectrum of subjects, formats, and publication dates, enhancing the generalizability of the study's outcomes.

3.2 Tool Development

Following the assembly of the dataset, the focus shifts to the development of an automated tool designed to assess and amend ePub files to meet WCAG 2.1 standards. Utilizing existing open-source libraries and APIs, the project will tailor these resources to target common accessibility challenges within ePubs, including but not limited to semantic structure analysis, text-to-speech readability enhancements, and dynamic image captioning. This stage is foundational, setting the stage for the subsequent implementation and evaluation of the tool.

3.3 Tool Implementation

With the automated tool developed, its application across the curated dataset marks the next phase. This stage not only involves the automated correction of identified accessibility issues but also incorporates manual reviews to validate the accuracy and effectiveness of the tool's modifications. A subset of the ePub files will undergo in-depth manual analysis to establish a benchmark for the tool's performance assessment.

3.4 Evaluation

The evaluation of the automated tool encompasses several dimensions:

- **Compliance Improvement:** Assessing the enhancements in WCAG 2.1 compliance levels before and after the tool's application.
- **Efficiency:** Evaluating the time and resource economy offered by the automated tool in contrast to manual adjustments.
- **Fidelity to Original:** Determining if the ePubs retain their original integrity and content post-correction, ensuring that the essence of the eBooks remains unaltered while enhancing accessibility.
- **User Satisfaction:** Gathering insights from stakeholders, including publishers, authors, and readers with disabilities, through surveys and interviews to gauge feedback on the tool's usability and the quality of improvements.
- **Comparative Analysis:** Benchmarking the tool against existing solutions like the Accessibility Checker for EPUBs (ACE) to identify areas for enhancement and ascertain the tool's advantages.

4 RISK ASSESSMENT

The proposed project involves the development and application of automated software solutions to enhance the accessibility of electronic publications (ePubs) in compliance with Web Content Accessibility Guidelines (WCAG) 2.1 standards. While the project aims to address significant accessibility challenges, several risks are associated with the use of ePubs, particularly concerning legal considerations.

4.1 Legal Risk: Usage of ePubs

One of the primary risks of the project revolves around the legal implications of using ePubs sourced from various sources. Although efforts will be made to collect ePubs from open-access platforms such as Gutenberg.org and StandardEbooks.org, there may still be uncertainties regarding the legality of accessing and modifying certain ePubs. These uncertainties stem from the complex copyright landscape surrounding digital publications and the potential presence of proprietary content within ePubs.

4.1.1 Risk Mitigation. To mitigate this risk, strict adherence to copyright laws and licensing agreements will be paramount. Efforts will be made to prioritize ePubs that are explicitly labeled as open-access or licensed under permissive terms, ensuring compliance with legal requirements. Additionally, thorough documentation will be maintained regarding the sources and permissions obtained for each ePub used in the project.

4.2 Ethical Considerations

In addition to legal risks, ethical considerations surrounding data privacy and confidentiality must be addressed throughout the project. The automated tool developed as part of the project will process sensitive information contained within ePubs, including personal data and copyrighted content. Ensuring the ethical handling of this data is essential to uphold user privacy and maintain trust in the project's outcomes.

4.2.1 Risk Mitigation. To mitigate ethical risks, informed consent will be obtained from stakeholders involved in the project, including publishers, authors, and users with disabilities, regarding the use

of their data for research purposes. Transparent communication regarding data handling practices and the intended use of automated tools will foster trust and mitigate potential ethical concerns.

4.3 Project Dependencies

Furthermore, the project may be subject to dependencies on external factors, such as the availability of open-access ePubs, the stability of third-party libraries and APIs used for tool development, and the responsiveness of stakeholders for user feedback and validation. Any disruptions or delays in these dependencies could impact the project timeline and deliverables.

4.3.1 Risk Mitigation. To mitigate dependencies, proactive communication and collaboration with stakeholders will be prioritized to address potential challenges and ensure timely resolution of issues. Additionally, contingency plans will be developed to account for potential disruptions, such as identifying alternative data sources or software solutions to mitigate project delays.

5 PROJECT PLAN

This section outlines the timeline and tasks for the execution of the project, ensuring a structured approach towards achieving the objectives.

5.1 Detailed Task Breakdown

The project tasks are further detailed using a Gantt chart, providing a visual representation of task dependencies and durations.

Task Breakdown

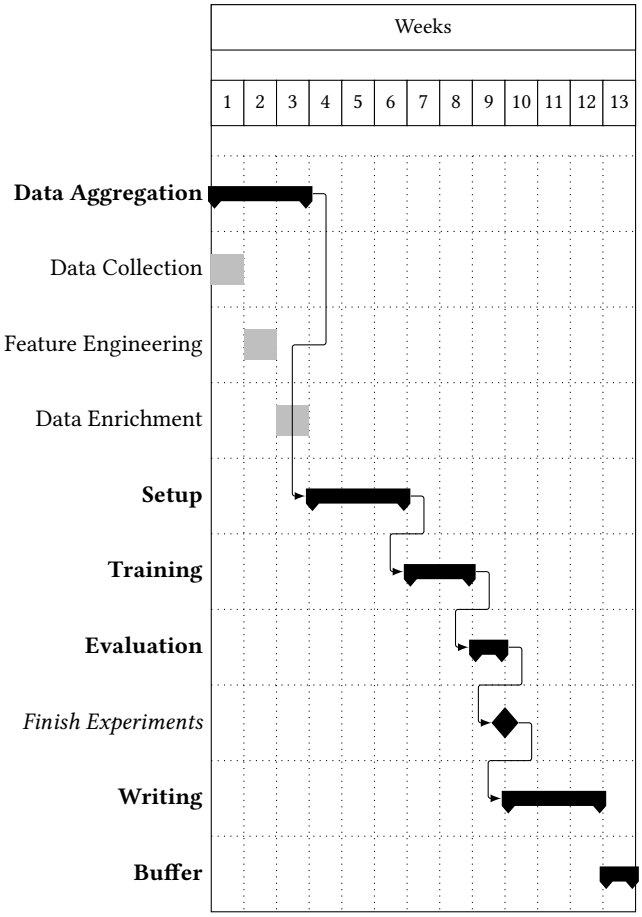
(1) **Preparatory Work (Week 1)**
- Finding Data: Gather diverse ePub files from various sources.

(2) **Tool Development (Weeks 2-6)**
- Initial Tool Development: Begin developing the automated tool for ePub accessibility enhancement.
- Experimentation: Conduct experiments and refine the tool based on feedback.
- Tool Refinement: Improve the tool's functionality and address any issues encountered during experimentation.

(3) **Evaluation (Weeks 7-10)**
- Testing and Validation: Evaluate the effectiveness and accuracy of the tool in enhancing ePub accessibility.
- Analysis and Results: Analyze the results of the evaluation and draw conclusions regarding the tool's performance.

(4) **Reporting (Weeks 11-12)**
- Writing Thesis: Compile the findings, methodology, and outcomes into a comprehensive thesis document.

(5) **Buffer (Week 13)**
- Allocate additional time as a buffer for any unexpected delays or last-minute adjustments.



REFERENCES

[1] F. Garrad-Cole. 2024. *Using a pre-arrival platform as an induction tool.* Supporting the Student Journey into Higher Education. https://books.google.com/books?hl=en&lr=&id=Cv31EAAAQBAJ&oi=fnd&pg=PA42&dq=digital+accessibility+for+physically+challenged&ots=llWUCJpaWc&sig=3qhF-U0hvl8B0F9B7IZ7fd83cpA Accessed: 18-2-2024.

[2] A. Othman, A. Dhouib, and A. Nasser Al Jabor. 2023. Fostering websites accessibility: A case study on the use of the Large Language Models ChatGPT for automatic remediation. https://dl.acm.org/doi/abs/10.1145/3594806.3596542. In *Proceedings of the 16th International Conference on Web Accessibility*. ACM. Accessed: 17-2-2024.

[3] Joo Hyun Park, Hyun-Young Kim, and Soon-Bum Lim. 2019. Development of an electronic book accessibility standard for physically challenged individuals and deduction of a production guideline. *Computer Standards & Interfaces* 64 (2019), 78–84. https://doi.org/10.1016/j.csi.2018.12.004 Accessed: 18-2-2024.

[4] H. Shah. 2023. Harnessing Web Accessibility Tools for WCAG 2.1 Migration of a Design System. https://ieeexplore.ieee.org/abstract/document/10384953. In *2023 International Conference on Computing ....* Accessed: 17-2-2024.