



A11yPDF: Bridging the Gap to Inclusive PDFs

Wajdi Aljedaani
University of North Texas

Sandeep Kumar Rudhravaram
University of North Texas

Akhila Chintham
University of North Texas

Abdulrahman Habib
Saudi Data and Artificial Intelligence
Authority

Marcelo M. Eler
University of São Paulo

ABSTRACT

Portable Document Format (PDF) files are platform-independent; thus, they are commonly used for sharing documents that need to be viewed or printed exactly as intended by the author. As with any other content format, they can present accessibility issues that hinder people with disabilities from fully perceiving or understanding the information displayed. There are many accessibility evaluation tools for PDF files that can help content creators to make them more accessible. Still, they have some limitations: lack of evaluation of certain elements of the document, cost-prohibitive paid licenses, and complex usage and reports. This paper presents A11yPDF, a free tool that can effectively evaluate the accessibility of PDFs and provide users with practical reports to help authors make their content more accessible. Compared to the currently available tools, A11yPDF is free, covers the evaluation of all elements considered important in the literature, and provides easy-to-read reports.

CCS CONCEPTS

• **Human-centered computing** → **Web Accessibility in PDF.**

KEYWORDS

Accessibility, PDF accessibility, Tools, WCAG.

ACM Reference Format:

Wajdi Aljedaani, Sandeep Kumar Rudhravaram, Akhila Chintham, Abdulrahman Habib, and Marcelo M. Eler. 2024. A11yPDF: Bridging the Gap to Inclusive PDFs. In *The 21st International Web for All Conference (W4A '24)*, May 13–14, 2024, Singapore, Singapore. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3677846.3677859>

1 INTRODUCTION

Portable Document Format (PDF) files are known for their versatility and widespread adoption as a file format for documents, ensuring consistent presentation and formatting across various devices and platforms. Nevertheless, they may present accessibility issues for people with diverse abilities using assistive technologies [4]. Navigation and understanding of information on PDF files are hindered by the frequent lack of essential accessibility features, including structured tags [8, 9], labeled headers, and alternate text

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

W4A '24, May 13–14, 2024, Singapore, Singapore

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1030-8/24/05

<https://doi.org/10.1145/3677846.3677859>

for the images. Additionally, complex PDF layouts and non-linear structures can further disrupt screen readers and magnification tools. Despite established guidelines like PDF/UA-1 [5] and the Matterhorn Protocol [7], numerous studies reveal a low compliance rate for accessibility standards in PDF documents [1–3]. While numerous tools exist to evaluate and improve PDF accessibility (e.g., PDFa Inspector¹, PAVE², ABBYY FineReader³, Common Look⁴, Adobe Acrobat Pro Accessibility Checker⁵), their widespread adoption is hindered by many limitations [6], such as cost-prohibitive paid licenses, complex usage and reports, and the lack of evaluation of certain elements of the document.

To address these concerns, this research proposes the development of a user-centered PDF Accessibility Analyzer tool, namely, A11yPDF. This tool aims to improve the accuracy, efficiency, and intuitiveness of accessibility evaluation, empowering even novice users to identify and address accessibility barriers within their PDFs. By effectively tackling these challenges, the proposed tool has the potential to significantly improve the accessibility of PDF documents, empowering individuals with diverse abilities to access information and participate fully in education and society.

2 A11YPDF

We developed A11yPDF, a web-based application that receives a PDF file and conducts a comprehensive evaluation based on the WCAG 2.2 guidelines presented in Table 1. A11yPDF assesses ten key features against accessibility criteria, including (1) Page Header, (2) Page Footer, (3) Page Number, (4) Link & Hyperlinks, (5) Page Contrast, (6) Image Description (Caption), (7) Table Description (Caption), (8) Dyslexia Friendly, (9) Font Size, and (10) Color Blindness Adaptations. Accessibility ratings are assigned for each feature, which are aggregated in a single score to reflect the overall accessibility level of the PDF document. The strategy to evaluate the accessibility of each element is presented as follows.

Contrast. The evaluation process begins by extracting images from each page of the PDF, forming a collection denoted as I_{page} . For each image within this collection, we extract RGB values of the top-left and bottom-right pixels, represented as $P_{\text{top-left}}$ and $P_{\text{bottom-right}}$, respectively. Using these RGB values, we calculate the contrast ratio CR_i for each image. This calculation involves transitioning the RGB values to a different color space to ascertain luminance values, which is crucial for determining the contrast ratio.

¹<https://github.com/pdfa/PDFaInspector>

²<https://pdf.abbyy.com/>

³<https://pdf.abbyy.com/>

⁴<https://monsido.com/monsido-commonlook-partnership>

⁵<https://www.adobe.com/accessibility/products/acrobat/using-acrobat-pro-accessibility-checker.html>

Table 1: WCAG 2.2 standard guideline with PDF elements.

No.	Principle	Success	Level	Guideline	PDF Element
1	Operable	2.4.9	AAA	Link Purpose (Link Only)	link
2	Perceivable	1.4.11	AA	Non-text Contrast	Contrast
3	Operable	2.4.6	AA	Headings and Labels	Header
4	Operable	2.4.6	AA	Headings and Labels	Footer
5	Operable	2.4.6	AA	Headings and Labels	Page Number
6	Perceivable	1.4.4	AA	Resize Text	Font Size
7	Perceivable	1.3.1	AA	Info and Relationships	Image Caption
8	Perceivable	1.3.1	AA	Info and Relationships	Table Caption
9	Understandable	3.1.5	AAA	Reading Level	Dyslexia
10	Perceivable	1.4.1	A	Use of Color	Color Blindness

Font Size. The evaluation process finds occurrences of each unique font size within the document and stores them on a matrix. Next, we categorize each occurrence based on their compliance with the WCAG criteria (e.g., 14px font size).

Links. The evaluation process entails the extraction of web links that undergo accessibility verification to ascertain their availability over the Internet, thereby assessing the relevance and utility of the referenced web resources. The culmination of this procedure is a detailed summarization that not only enumerates the total URLs identified but also delineates between those accessible and those that are not, providing a clear overview of the document's connectivity and the current accessibility status of its web references.

Header/Footer/Page Number. Navigation features such as headers, footers, and page numbering are indispensable for fostering an accessible reading environment within PDF documents. Headers succinctly convey the core idea or title of the document or its specific sections, immediately offering context and assisting readers in grasping the document's structure. Footers, on the other hand, typically provide additional information or references, enriching the document's content. Page numbers enable readers to navigate to particular sections of the document effortlessly, streamlining the process of finding specific information. To guarantee these navigation elements effectively contribute to making documents more accessible, advanced analytical methods have been introduced. These methods assess whether PDF documents correctly incorporate headers, footers, and page numbers, ensuring these crucial navigation aids are present and appropriately formatted. Such automated assessments are key in confirming adherence to the Web Content Accessibility Guidelines (WCAG), aimed at making information accessible to a broader audience, including individuals with disabilities.

Image and Table Caption. The process maps the linkages between content elements and their captions, seeking to render the dissemination of information more intuitive and approachable. It accentuates the importance of merging visual data depiction with textual narratives, thereby aiding in the generation of scientific literature that is not only more comprehensible but also user-friendly.

Dyslexia. Enhanced font readability, as recommended by the Web Content Accessibility Guidelines (WCAG), emphasizes the use of sans serif fonts such as Arial and Comic Sans to improve text clarity for a wider audience, including those with reading difficulties like dyslexia. Sans serif fonts are preferred for their straightforward style, which reduces letter crowding and enhances legibility. Recommended fonts include Arial, Comic Sans, Verdana, Tahoma, Century Gothic, Trebuchet, Calibri, and Open Sans. Following these recommendations supports WCAG's goal of making

digital content accessible and understandable to all users. Additionally, choosing the appropriate font size and color contrast is vital for further enhancing readability and ensuring compliance with WCAG accessibility standards.

Color Blindness. The evaluation process involves a comprehensive methodology for color quantization and luminance calculation using k-means clustering and matrix operations. Our approach efficiently simplifies the color palette of an image while retaining essential visual details. By calculating luminance and contrast ratios, we can assess the accessibility of visual content.

For the accessibility and content quality of PDF, we define a comprehensive set of criteria encompassing both structural and content aspects. Each criterion is quantitatively assessed through a percentage score, p_i , reflecting the extent to which the document meets specific accessibility or content standards. The overall performance, P_{overall} , rounded, is then calculated as the arithmetic mean of these scores, providing a holistic measure of the document's quality and accessibility.

3 CONCLUSION

A11yPDF analyses all important features for accessibility according to the literature on the accessibility of PDF files, which is an advance compared to the currently available tools. In addition, our tool is free and provides reports that can be easily understood even by non-experts. In its current implementation, A11yPDF has the following limitations: The A11yPDF tool does not evaluate title, and subsections elements. In future work, we intend to conduct an extensive experiment to evaluate the effectiveness of our tool compared to the related work.

REFERENCES

- [1] Wajdi Aljedaani, Mona Aljedaani, Eman Abdullah AlOmar, Mohamed Wiem Mkaouer, Stephanie Ludi, and Yousef Bani Khalaf. 2021. I cannot see you—the perspectives of deaf students to online learning during covid-19 pandemic: Saudi arabia case study. *Education Sciences* 11, 11 (2021), 712.
- [2] Wajdi Aljedaani, Mohammed Alkahtani, Stephanie Ludi, Mohamed Wiem Mkaouer, Marcelo M Eler, Marouane Kessentini, and Ali Ouni. 2023. The state of accessibility in blackboard: Survey and user reviews case study. In *Proceedings of the 20th International Web for All Conference*. 84–95.
- [3] Wajdi Aljedaani, Rrezarta Krasniqi, Sanaa Aljedaani, Mohamed Wiem Mkaouer, Stephanie Ludi, and Khaled Al-Raddah. 2023. If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: a literature review. *Universal access in the information society* 22, 3 (2023), 1027–1046.
- [4] Jeffrey P Bigham, Erin L Brady, Cole Gleason, Anhong Guo, and David A Shamma. 2016. An uninteresting tour through why our research papers aren't accessible. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 621–631.
- [5] Olaf Drümmer. 2012. PDF/UA (ISO 14289-1)—applying WCAG 2.0 principles to the world of PDF documents. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I* 13. Springer, 587–594.
- [6] Kerry A Falloon. 2020. Effectively Evaluating the Accessibility of Electronic Monographs Using VPATs and Other Resources at the College of Staten Island Library-CUNY. *Serials Review* 46, 2 (2020), 98–113.
- [7] Alicia Mason and Sakshi Bhati. 2023. Ready. gov: Who's ready, really? Examining principles of inclusivity and universal design in emergency management and disaster preparedness public information websites. *Journal of Emergency Management and Disaster Communications* (2023).
- [8] Obianuju Okafor, Wajdi Aljedaani, and Stephanie Ludi. 2022. Comparative Analysis of Accessibility Testing Tools and Their Limitations in RIAs. In *International Conference on Human-Computer Interaction*. Springer, 479–500.
- [9] Alberto Dumont Alves Oliveira, Paulo Sérgio Henrique Dos Santos, Wilson Estécio Marcílio Júnior, Wajdi M Aljedaani, Danilo Medeiros Eler, and Marcelo Medeiros Eler. 2023. Analyzing Accessibility Reviews Associated with Visual Disabilities or Eye Conditions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.