

Trillions PDFs to Make Accessible - PDFix Breakout Session Transcript

Presentation

Diana

How could we ensure accessibility for trillions of PDFs? This was the challenge we tackled when we launched PDFix a few years back. Our solution streamlining automation wherever possible. We developed a suite of tools designed to address diverse workflows and scenarios for creating, distributing and publishing accessible PDF documents, minimizing the need for manual intervention. Hi, I'm Diana from PDFix and I'm here to introduce you to our PDF Remediation Software, which significantly reduces workload and enhances productivity in making PDF documents accessible.

Whether you're a company offering remediation services or a large organization seeking an in-house solution, we believe we have the right solution for you. We also aim to simplify working with PDFs for all users. Our software is capable of addressing every stage in a document's lifecycle and enhancing its accessibility. Furthermore, it is compatible with nearly every platform. The usual scenario often begins with classification, which entails outlining the required steps for document processing.

These steps might include OCR, pre-processing, or identifying templates for large volumes of documents. I'll touch on templates later in my discussion. For untagged PDF files, the initial stage involves text recognition and automatic tagging for structural elements. The utilization of a template, content or layout matrix can greatly streamline the remainder of the process. The remaining highlighted tasks such as manual remediation, validation or fixing delineate the actions necessary for existing tagged PDFs that need adjustments due to structural inaccuracies or lack of compliance with PDF UA or WCAG ensuring that the document achieves the desired level of accessibility.

While these steps typically demand manual involvement, our software facilitates batch processing, instant validation, and automates rectification procedures. Let's start by exploring several helpful tools and features incorporated into our desktop application aimed at aiding you in manual remediation tasks. We've simplified the tagging process for you by introducing a tag as feature that leverages cutting edge content recognition technology. Simply select the content by drawing a rectangle, highlighting text or choosing objects, and then select tag as.

You can choose from a variety of common structures such as paragraphs, headings, figures, lists, tables, and more. The software automatically identifies the selected content and generates the appropriate tag structure. This structure is then placed within the tag tree at your specified location. A short video on the following slide will showcase this procedure.

Computer Voice

Select the area and choose tag as list from the context menu or accessibility pane. The created list will appear at a specified position in a tag tree.

Diana

I'm certain you've encountered the challenge of tagging a complex table. We can demonstrate that the process of tagging table can be significantly simplified with our table editor. Whether you're working from plain content or an existing table tag, our table editor ensures a seamless user experience. You can easily organize table rows, columns, split and merge cells, designate row and column headers, or establish associated headers in more intricate tables.

Once the table design is finalized, the table tag is generated within the tag tree at your specified position. A short video on the following slide will showcase this process.

Computer Voice

To generate a new table. Use the table tool and enclose the desired content by drawing a rectangle around it for tagging. During this phase you can organize, insert or delete rows and columns. Furthermore, you have the option to split or merge cells to suit the layout of the table. By selecting multiple cells, you can conveniently designate the column or row header via the context menu.

After finalizing the design to your satisfaction, proceed to confirm the creation of the table. The created table tag will then appear in the specified position within the tag tree. Accessing the table editor for an existing table tag ensures a seamless user experience granting you the ability to finally tune and adjust the table according to your design preferences.

Diana

Validation is a crucial aspect of the remediation process. Having immediate access to the compliant state is a significant advantage. We've seamlessly integrated validation directly into the program, enabling you to assess the status without saving the file and evaluating issues in a separate application. From the validation results pane. You can pinpoint problematic sections of the document or content. Furthermore, you can directly address some of these issues directly from the reported notifications.

PDFfix desktop enables the use of validation profiles like ISO 32000, PDF UA, or WCAG. Additionally, you have the flexibility to create a custom validation profile tailored to your specific accessibility needs. You have the option to produce a validation report in HTML or PDF format for each document. The validation is powered by veraPDF, an industry supported validation, is a purpose built, open source, file format validator covering all PDF A and PDF UA parts conformance levels.

veraPDF is designed to meet the needs of digital preservationists and is supported by the PDF software developer community. You can find out more at verapdf.org. Here's another brief video explaining the concept.

Computer Voice

To examine the PDF document compliance, open the validation pane and click validate. The validation results enable you to showcase and highlight the problematic sections of the document. Certain issues can be addressed directly from this point by selecting fix error from the context menu.

Afterwards, you can revalidate to review the document once more and iterate through the process for any other reported issues. Following that, you can conduct another round of validation to reassess the document and continue addressing any additional reported issues. This iterative process ensures thorough scrutiny and refinement until all discrepancies are resolved, ensuring optimal compliance and quality of the document.

You also have the option to examine the document using a different validation profile.

Diana

A keyboard is an essential tool for every remediation professional with PDFix desktop. You have the ability to customize your own set of shortcuts for nearly every action from the toolbar or menu. Let's examine what sets the accessibility automation with PDFix apart. We offer a wide array of commands to be executed on one or multiple documents in a batch.

These commands are categorized into groups such as fonts, tags, content, annotations, metadata and others. A few examples: fix optional content, fix ID tree, fix parent tree, set PDF UA standard, set title, artifact content, create bookmarks, validate and many more. Each command can be customized according to the specific needs of the document set. For example, you can utilize a set annotation contents command to update links and configure it to incorporate the text under the link as its content.

Alternatively, you can set the document title based on the first heading or the file name. You have the capability to perform multiple commands simultaneously by crafting a customized command sequence tailored for a specific document set. The command sequence you've created can be saved, reused or shared with your colleagues. Lastly, all commands work with PDFix desktop batch as well as with automated workflow with the PDFix SDK.

Previously, I discussed the concept of templates. Templates are highly beneficial tools that facilitate content recognition and structure generation by utilizing known data. With templates, you can efficiently handle large document sets such as invoices, health records, or bank statements which typically adhere to machine generated layouts. Here's a comparison of the outcomes obtained from processing a file without and with utilizing the template.

The initial recognition algorithm recognizes the content as two tables delineated by a line. Using the template, you can set rules for recognizing paragraphs, headings, headers, footers, or other specific types of content. Additionally, you can specify alternate text or images, mark the content as an artifact and implement various other customizations. The template proves to be a

significant time saving tool during the remediation process of PDF documents, reducing the need for extensive manual work.

Here are some technical details regarding automated remediation and utilizing commands with the PDFix SDK. You have the flexibility to select from various integration methods and programming languages. You have the option to engage with PDFix SDK through .net, Python, C++, Java, JavaScript or via a command line interface. Feel free to reach out to us. Our developers will be happy to advise you in the most suitable technical solution for your needs.

We cannot overlook the significant assistance provided by AI in many aspects of the remediation process. We offer pre-built modules for automated tasks such as natural language detection, image captioning, or layout detection. AI represents the future of automated processing. Additionally, our plugin based interface enables you to apply your own training set to address highly specific requirements. Finally, I'd like to touch upon tagged PDFs and HTML.

The structural data preserved within a PDF as tags can precisely delineate content reflow facilitating the generation of fully responsive HTML. This proves to be optimal for distributing content and ensuring readability on smaller devices. The HTML derivation algorithm in PDFix generates compliant HTML from a tagged PDF, showcasing the effectiveness of well tagged PDF documents. This process is also beneficial for verifying the reading order of tagged PDFs.

Likewise, using PDFix, you have the ability to export the tag structure into alternative formats such as JSON or XML.

We've recently introduced a reverse conversion from HTML to PDF UA. Such conversions are frequently preferred by content creation systems and content management systems. By leveraging a combination of HTML to PDF conversion and automated accessibility fix commands, it can automatically produce documents that are fully accessible. All the information presented in my presentations is integrated into our products, which are as follows.

PDFix Desktop Lite, a free PDF accessibility checker. PDFix Desktop Professional, a remediation tool for professionals. PDFix SDK, an accessibility automation and integration tool. All of our programs are cross-platform and available on Windows, Mac OS and Linux Operating Systems.

I appreciate you taking the time to watch my presentation. All our products offer free trials. You can download them from our website at www.PDFix.net/download. Additionally, you can utilize our free services on PDFix.io, where we provide access to all the main features. Feel free to contact us if you have any questions at support@PDFix.net.

Now it's time for questions. If you have any, please don't hesitate to ask them in the chat. My colleagues are available and ready to provide answers. Moreover, you have the option to visit our booth in the conference lobby after the session, or you can speak directly with our specialists regarding any questions or requests you may have.