Making Accessible PDF Documents

Heather Devine
Adobe Systems
345 Park Ave
San Jose CA 95110
hdevine@adobe.com

Andres Gonzalez
Adobe Systems
345 Park Ave
San Jose CA 95110
andgonza@adobe.com

Dr. Matthew Hardy
Adobe Systems
345 Park Ave
San Jose CA 95110
mahardy@adobe.com

ABSTRACT
Accessibility features in the Adobe Portable Document Format (PDF) help facilitate access to electronic information for people with disabilities. This workshop explores how to create accessible PDF documents, from within Adobe Acrobat and other applications; how to use the Adobe Acrobat PDF accessibility checker and repair workflow; best practices for accessibility; and how accessibility has been built into forthcoming ISO standards (PDF/UA, PDF 32000-2).

Categories and Subject Descriptors
I.7.2 Document Preparation, Adobe Acrobat.

General Terms
Documentation, Standardization.

Keywords
Accessibility, Content Reuse, Logical Structure, PDF, PDF/UA.

1. INTRODUCTION
Electronic document accessibility is crucial to allow users with disabilities access to electronic information. Making accessible PDF documents not only enables access through assistive technologies (AT), such as screen readers and magnifiers, but also provides logical structure to a document which can enable better reflow, display of documents and content reuse.

2. CREATING ACCESSIBLE PDF DOCUMENTS
There are two primary workflows for creating accessible PDF documents. The first is to add logical structure and text extraction capabilities to an existing PDF file. This process is often referred to as tagging. Tagging is, in most cases, a semi-automated process where a software inference engine creates a basic structure, and a user makes corrections and additions to that structure using manual tools.

The second is to generate an accessible PDF file from other software application source files, such as a word processor or desktop publishing application. This process is mostly automatic; the semantic information contained in the source file about the document structure is translated into a tagged PDF.

DocEng'11, September 19–22, 2011, Mountain View, California, USA.
ACM 978-1-4503-0863-2/11/09.

2.1 Existing PDF Document
In order to make an existing PDF document accessible, one must consider the characteristics of the document to determine which changes are necessary to increase accessibility. For example, scanned images of text must undergo optical character recognition (OCR) to add searchable text. Document logical structure must be identified, and its correspondence to document content and layout must be established.

Document logical structure includes an appropriate reading order, and the identification of semantic structural elements such as sections, headings, paragraphs, tables, and images. Documents should contain alternative text descriptions for images and graphics, as well as a specification of the document's language. Documents can also include navigational aids, such as bookmarks, which are particularly helpful for longer documents.

2.2 Converting Documents to PDF
The key to creating accessible PDF documents from other source files, such as files from a word processor or desktop publishing application, is to design the source document with accessibility in mind. The author of a document needs to create the proper logical structure in the original format that is being used, and the document structure will be translated into the corresponding PDF tags.

Similar guidelines apply: sections, headings, and paragraphs should be identified using the software's styles; images, graphics or animated elements should include alternative text descriptions; and navigational aids, such as bookmarks, should be included. Content that is not relevant to the text, such as background images or page numbers, should be tagged as artifacts. Document metadata can be included, such as title, author, keywords, and language. Additionally, some programs allow tab order and reading order to be defined, which can aid in accessibility.

Options for creating PDF files from programs may also impact the accessibility of the generated document. Conversion settings which effect accessibility include tag generation, standards compliance, bookmark conversion, file security, link conversion, and animation settings. If the conversion results in accessibility problems in the generated PDF file, these issues can be addressed using accessibility checker repair tools, or in some cases in the source document itself. Keep in mind that repairs made to the generated file will not be saved if the file is regenerated. Similarly, once the PDF file is generated, its accessibility should be evaluated using an accessibility checker so that any remaining accessibility issues can be repaired.

2.3 Supporting Assistive Technologies

The logical structure and content of a properly tagged PDF can be conveyed to assistive technologies such as screen readers and magnifiers via accessibility APIs. Adobe Acrobat and Reader support the Microsoft Active Accessibility (MSAA) API. In addition, the Adobe IPDDom API is used to render to assistive technologies those aspects of the structure of PDF documents that are not covered by MSAA. For details about conveying document structure and content through accessibility APIs see Gonzalez and Guarino-Reid (2005) [1].

3. ACCESSIBILITY CHECKER

An accessibility checker can help identify parts of a document which do not meet accessibility requirements. An accessibility checker can iterate through each accessibility requirement to determine whether the document meets that requirement. For example, it can examine the document for scanned text which should be made searchable, identify images and graphics which require alternative text descriptions, determine whether the document includes structure information, locate sections headings to add navigation aids, and ensure that document text is available to AT.

4. REPAIR WORKFLOW

The repair workflow can be used in conjunction with an accessibility checker to create an accessible PDF document. After using the checker to determine which parts of the document do not meet accessibility requirements, repairs can be made to improve accessibility. After evaluating the characteristics of the document and its accessibility level, there are several steps which can then be taken to improve accessibility based on the results: (1) perform OCR to convert scanned text to searchable text; (2) add navigation aids, such as bookmarks; (3) ensure that document security does not limit access to document text; (4) add tags and structure to the file; and (5) re-evaluate the document to determine if there are other areas which need improvement.

5. BEST PRACTICES

There are a number of best practices which support the creation of accessible PDF documents. These practices should be kept in mind when making an existing PDF file accessible or generating an accessible PDF file from other software programs:

- Ensure the document text is available to AT
- Identify document structure through the use of tags
- Include tooltips or alternative text for form fields, graphics, and animated elements
- Create navigation aids
- Verify the reading order for document elements
- Verify the document's accessibility

6. ACCESSIBILITY IN PDF STANDARDS

In 2008, PDF became an ISO standard (ISO 32000-1:2008). While PDF already defined mechanisms by which a PDF document could be made accessible, there was no standard for accessible PDF. However, work is now underway on a new ISO PDF subset called PDF Universal Accessibility (ISO 14289-1, PDF/UA). PDF/UA enforces the optional structures provided in the PDF standard to ensure than any PDF generated to the PDF/UA specification is potentially accessible. It also restricts certain types of content which would be difficult to interact with for someone using AT, e.g. scripts that require user input that has timing requirements.

PDF/UA further defines what information an interactive PDF processor (e.g. a PDF reading application) must make available to AT and in some cases specifies how it should interact with AT. An example of this is with multimedia playback, where the controls to a sound clip must be made available to someone who cannot see to use the controls with a mouse.

Finally, PDF/UA describes requirements for AT, to ensure that no matter which combination of interactive PDF processor and AT a user chooses, they have similar levels of access to the content and interactive elements of the document.

PDF/UA is still a work in progress and is currently an ISO draft international standard (DIS). The expectation is that it will be published as a final standard in 2012.

During the creation of PDF/UA, a number of needs were identified in PDF to facilitate accessibility. These additional needs are being addressed in PDF 2.0, ISO 32000-2. The changes primarily consist of extensions to the logical structure tags, which can now represent more types of content. Examples of these new tags include support for mathematics and document titles.

7. ACKNOWLEDGMENTS

Our thanks to Adobe Systems and Kirk Gould for supporting our participation in *DocEng'11*.

8. REFERENCES

[1] Andres Gonzalez and Loretta Guarino Reid. 2005. Platform-independent accessibility API: accessible document object model. In *Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)* (W4A '05). ACM, New York, NY, USA, 63-71. DOI=http://doi.acm.org/10.1145/1061811.1061824