Climbing the Matterhorn

An introduction to the definitive algorithm for PDF/UA conformance

PDF Association, PDF/UA Competence Center / **January 14, 2014**

PDF/UA is the ISO standard – published as ISO 14289 in July 2012 – defining the creation and processing of accessible PDF. This article is directed primarily at implementers, quality assurance (QA) and technical product managers interested in supporting accessibility in PDF. It describes the purpose and function of the Matterhorn Protocol, and explains how developers may use this document to address PDF/UA conformance in a systematic and reliable manner.

As a format, PDF was designed to provide reliable, high-quality visual representation of **any** two-dimensional content, regardless of peculiarity of design, source format or viewing environment. PDF does this better than any other technology; indeed, it has no serious competition.

Accessibility, however, is a broad and complex subject. In addition, the flexibility of PDF forces developers to cover an exceptionally wide range of use-cases. At the same time, the relative vagueness of Tagged PDF's definition in ISO 32000-1 has not encouraged third-party development.

The PDF/UA standard provides developers with a clear road map to understanding how to do Tagged PDF right, but nonetheless requires substantial research for those who aren't already familiar with accessibility requirements. Most major PDF software developers are willing to implement Tagged PDF, but they want to know what's really important. That's where the Matterhorn Protocol fits in.

The Matterhorn Protocol Provides Focus

A PDF Association publication, the [Matterhorn Protocol](#) specifies all possible ways to fail PDF/UA. As such, it's a set of algorithms providing the practical rules for implementing software that creates, processes or presents accessible PDF.

The Matterhorn Protocol helps set priorities in both research and execution. It enables developers not yet fully familiar with every detail in PDF/UA to get to work right away, accelerating all aspects of code and product development.

Developed by the PDF Association in cooperation with the [AIIM Committee](#) that initiated PDF/UA the Matterhorn Protocol consists of 136 distinct Failure Conditions. Each Failure Condition identifies a non-conforming condition based on PDF/UA's hard requirements ("shall" statements) as applied to documents, pages, objects or JavaScripts. 89 Failure Conditions may be assessed entirely by software; 47 require some level of human validation.

How to Use the Matterhorn Protocol

The basic approach for implementing the Matterhorn Protocol is to map the Failure Conditions to the various tasks implied by the specific word-processing, content extraction or other context.

PDF Creation

PDF creation is the ideal place to implement PDF/UA conformance for many reasons, not least because so many checkpoints requiring human validation may be inferred from the structures created by the author. The PDF generator must ensure that semantic tables in the source, for example, are properly tagged with table tags and attributes in the output PDF.

Accessibility Validation

Organizations that adopt accessibility standards want the capacity to check the accessibility status of their websites and PDF files. Implementers will need to consider how to make the human validation component as streamlined as possible while accommodating the variety of cases the software may encounter. See Access for All's PDF Accessibility Checker, PAC 2.0, for the first software implementation of a PDF/UA validator based on the Matterhorn Protocol.

Ensuring Tagged PDF Conforms to PDF/UA

It's always preferable to re-create a PDF than to edit an existing file to ensure good tagging, but it's not always possible. In many cases, existing tagged PDF files that fail human validation must be corrected rather than re-created from the source application. Depending on the precise design objectives, this sort of implementation can range from trivial to challenging. It's relatively easy to allow users to efficiently check and correct alternative text attributes in Figure tags. It's much less easy to produce a graphical user interface (GUI) allowing users to easily and reliably change the set of content enclosed within Figure tags.

Tagging Untagged PDF

Perhaps the most challenging task in the world of accessible PDF would be that of bringing untagged PDF files into conformance with PDF/UA. For such cases, human validation of logical reading order and valid structure type selection is difficult to avoid. Here, the Matterhorn Protocol provides both a means of verifying conformance and a way to document the human effort required to achieve it.

Consuming Tagged PDF

When PDF viewing implementations can rely on files validated by the Matterhorn Protocol end-users may be assured of a high-quality result in applications that use Tagged PDF. These include, besides the obvious case of Assistive Technology (AT), mobile devices, context extraction, search engines, business intelligence systems and other applications utilizing semantic information.

Download the Matterhorn Protocol Now

Designed by and for developers, the Matterhorn Protocol is a practical guide to achieving PDF/UA conformance you can start implementing today. Download the Matterhorn Protocol now.

Related Resources
- ISO 14289-1:2012 (PDF/UA), published by ISO
- PDF/UA-1 Technical Implementation Guide: Understanding ISO 14289-1, published by AIIM
- Achieving WCAG 2.0 with PDF/UA, published by AIIM
- PDF/UA in a Nutshell, published by the PDF Association
- Access for All's PDF Accessibility Checker 2.0 (PAC 2.0)
- The PDF Association's PDF/UA Competence Center