

PDF2HTML

Enabling Wider Accessibility

T. V. Raman
Senior Computer Scientist
Advanced Technology Group
Adobe Systems
<http://access.adobe.com>



©1997 Adobe Systems Incorporated. All rights reserved.

Outline

- **PDF –access challenges.**
- **Acrobat Access.**
- **Server side accessibility solution.**



Problem Statement

Purely visual layout of PDF:

- **Difficult to use with speech.**
- **Rich visual layout breaks screenreaders.**



Visual Layout

Implicitly encodes:

- Document logical structure.
- Document reading order.

Alternative presentations a challenge.



Marked Content

A future means for:

- Capturing logical structure.
- Encode reading order.

Needed for more than *just* access.



Client-side Solution

In April 1996, we released the first Beta version of Acrobat Access for Windows.



©1997 Adobe Systems Incorporated. All rights reserved.

Acrobat Access

- A free plugin to Acrobat.
- Helps screen readers work with Acrobat.
- Attempts to present text in reading order.

Over 5,000 copies downloaded.



GUI Speech Access

Accessing the GUI with screen readers:

- **Requires significant user customization.**
- **Difficult to start using new applications.**

Needs a high threshold of indignation!



Server-side Solutions

- **Cross platform solution.**
- **Screen reader independent.**
- **Deliver incrementally better solutions.**

Has proven very popular.



Solution Overview

- **Server converts PDF to HTML.**
- **Content presented in reading order.**
- **Conversion is fully navigable.**

Integrates with current WWW access.



Delivering Solutions

- Translation proxy server.
- Forms based conversion.
- Email based conversion.

Packaging to suit diverse needs.



Translation Proxy

- Proxy: <http://access.adobe.com:8080>
- Automatically serves PDF as HTML.

Completely transparent.



Forms-based Conversion

User specifies URL to a PDF document:

- Simple form for simple conversion.
- Advanced form for more options.

Designed to work with screen readers.



Email Conversion

User emails the URL of a PDF file to:

- **pdf2html@adobe.com**
- **pdf2txt@adobe.com**

Converted document is mailed back.



Conclusion

Users retrieve information:

- **Where they want,**
- **When they want,**
- **The way they want!**

