

Usable Alternate Formats for Scientific Materials

Dan Zingaro
OISE/UT

November 16, 2008

Outline

Alternate Formats from Publishers

What can we do with PDF Files?

What About Math?

LaTeX

Conclusion

The Process

1. Student requests book in alternate format

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start
3. Look at budget
 - ▶ Budget? What budget?

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start
3. Look at budget
 - ▶ Budget? What budget?
4. Request book from publisher in e-format

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start
3. Look at budget
 - ▶ Budget? What budget?
4. Request book from publisher in e-format
5. Wait

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start
3. Look at budget
 - ▶ Budget? What budget?
4. Request book from publisher in e-format
5. Wait
6. Wait more

The Process

1. Student requests book in alternate format
2. Look at calendar
 - ▶ Yikes! Only two weeks until classes start
3. Look at budget
 - ▶ Budget? What budget?
4. Request book from publisher in e-format
5. Wait
6. Wait more
7. Process files so they are useful

PDF Files

- ▶ It's easiest for a publisher to get you PDF
- ▶ Example from Cengage Learning Permission Request: “It may take 10-15 business days to receive files, although PDFs may take less time.”
- ▶ Compared to Word and ASCII, PDF is probably the best option anyway
- ▶ The Word or ASCII you get will just be a conversion from PDF, so we would lose flexibility

Outline

Alternate Formats from Publishers

What can we do with PDF Files?

What About Math?

LaTeX

Conclusion

Reading PDF Files Directly

- ▶ We can load a PDF file into Adobe Acrobat for direct reading with screen readers
 - ▶ Easy navigation through tables (sometimes)
 - ▶ Limits flexibility (i.e. cannot make changes to document)
 - ▶ Hard to search for text

Converting PDF Files to Text

- ▶ Various programs can convert PDF to text
 - ▶ Adobe Acrobat Reader
 - ▶ Adobe Acrobat Professional
 - ▶ XPDF (free, best so far for me)
- ▶ Student can add notes
- ▶ Text file can be cleaned up
- ▶ Sometimes these programs cannot extract the text (i.e. PDF is an image, PDF font info is not available)
- ▶ OCR software can be used instead with varying results

Outline

Alternate Formats from Publishers

What can we do with PDF Files?

What About Math?

LaTeX

Conclusion

Sample Math PDF File

For a polynomial of the form $ax^2 + bx + c$, we can solve for x using the quadratic formula: $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

- ▶ No matter what we do, this little bit of math makes a mess
- ▶ For example, XPDF gives the following

For a polynomial of the form $ax^2 + bx + c$,
we can solve for x using the quadratic v b2
formula: $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Outline

Alternate Formats from Publishers

What can we do with PDF Files?

What About Math?

LaTeX

Conclusion

LaTeX

- ▶ Math and related material is really awkward to type in a standard word processor
- ▶ Mathematicians, computer scientists, physicists, etc. usually use LaTeX (or another markup language) instead
- ▶ LaTeX is a text-only format that lets the author focus on the content, not the layout
- ▶ It is the publishers' equivalent of HTML (the language of the web)
- ▶ The publisher takes the LaTeX code and generates the PDF files that they give us
- ▶ This means that a version of a book probably exists in LaTeX, containing a completely accessible copy of the book's contents

Author's LaTeX of Sample PDF

For a polynomial of the form ax^2+bx+c , we can solve for x using the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Outline

Alternate Formats from Publishers

What can we do with PDF Files?

What About Math?

LaTeX

Conclusion

My Experience

- ▶ CS courses at graduate level don't use one text
- ▶ They use material from many books, journals, conferences
- ▶ Students will rely on even more material for their course and thesis research
- ▶ It is impossible to have this organic mass of material produced!
- ▶ My solution: cut out all intermediate work and ask publishers and authors for LaTeX

References

- ▶ Adobe: www.adobe.com
- ▶ XPDF: www.foolabs.com/xpdf