System Description: A Semantics-Aware LaTeX-to-WORD/ODF Converter

Lukas Kohlhase and Michael Kohlhase

Math/CS, Jacobs University Bremen

Abstract. [1] EdN:1

1 Problem & State of the Art

Many researchers in STEM fields only use LaTeX to typeset their documents. However many people still use Microsoft Word/Open office exclusively for their typesetting. When these two groups of people intersect, it can lead to friction, as transforming text to LaTeX is quite trivial but not the opposite. For example if a conference requested documents in Word format, the only recouse is often to just write the document in Word, which is a pain, especially if any Mathematics is to be included.

[2] EdN:2

There are several methods to transform papers from LaTeX to Word. The first method is to just generate a PDF file and then open this file in Word/Open Office. This achieves the goal of looking like the desired pdf document, just in Office. However when later editing the resulting document, problems turn up. The fundamental problem is that it converts the appearance of the document and loses meaning due to macro expansion. This is especially blatant when looking at the math in a document. Either it is treated as text, with no meaningful way to distinguish between math and formatted text that happens to contain some mathematical symbols, making automatic treatment of this kind of math difficult, or it is represented by an image of the relevant formulae, which makes editing extremely impractical if not impossible. The same holds true for references, they are essentially treated as parts of text with a linked number in front of them, complicating adding new references substantially.

2 Implementation

[3] [4] [5] [6] EdN:3 EdN:4

[1] EdNote: write something

[2] EdNote: Here we state the Problem, some conferences and admin want papers in word format, however LaTeX is superior for various reasons. Hence converter is needed. Two step process, wastes some time.

[3] EdNote: First convert to Ltxml using LaTeXML

[4] EdNote: Then convert the resulting ltxml using xslt stylesheets to a word apppropriate format.

EdN:5
EdN:6

3 Conclusion

EdN:7 [7] [8]
EdN:8

[5] EdNote: Finally use postprocessing to zip it all up
[6] EdNote: Don't know where to put an explanation of Word format in this place
[7] EdNote: MK@MK: say something
[8] EdNote: In Conclusion, easy to use