

System Description: A Semantics-Aware L^AT_EX-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 L^AT_EX 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 L^AT_EX is quite trivial but not the opposite. For example if a conference requested documents in Word format, the only recourse 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 L^AT_EX to Word. The first method is to just generate a PDF file and then open this file in Word/Open Office. However this approach has several problems. First of all it does not convert semantically at all. A headline is not treated as a headline, but as a piece of centered text with larger font. Additionally math is treated as either not distinct from text in any meaningful way, or as a picture, making later editing impractical if not impossible. The same applies to references. Finally it is a two step process that would save some time if creating the pdf could be skipped. ^{3 4}

EdN:3

EdN:4

2 Implementation

^{5 6 7 8}

EdN:5

¹ EDNOTE: write something

EdN:6

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.

EdN:7

³ EDNOTE: make PDF then open it with Word. Treats Math as either formatted Text or as pictures, unsuited for searching or for editing the math afterwards. References are also just pretty printed Text

⁴ EDNOTE: Various other ways that start at the .tex file, all either are buggy/not updated anymore or can't deal with math at all.

⁵ EDNOTE: First convert to Ltxml using LaTeXML

⁶ EDNOTE: Then convert the resulting ltxml using xslt stylesheets to a word appropriate format.

⁷ EDNOTE: Finally use postprocessing to zip it all up

EdN:8

3 Conclusion

EdN:9

⁹ ¹⁰

EdN:10

⁸ EDNOTE: Don't know where to put an explanation of Word format in this place

⁹ EDNOTE: MK@MK: say something

¹⁰ EDNOTE: In Conclusion, easy to use