

The ltx2word plugin

Transforming L^AT_EX to OOXML

Lukas Kohlhasse

January 16, 2014

Abstract

Since Donald Knuth created TeX in 1987, it has grown to be the major language to write mathematical papers in. However since then, the main distribution channels have changed: Nowadays many people use the web for their main information needs. With the advent HTML5 and the simultaneous standardization of MathML and Openmath, it has become possible to adequately display mathematics on the internet.

However there is currently no browser that can directly interpret L^AT_EX, making it necessary to develop tools to convert L^AT_EX to a web compatible format. Some attempts have been made at doing this with PDF or PNGs, however these attempts are essentially purely presentational and lose the rich semantic structure that can be found in L^AT_EX documents. The public domain L^AT_EXML system fixes this particular flaw by converting the L^AT_EX documents to an XML based format that conserves much of the meaning and is readable on the web.

However, while the majority of scientists from STEM fields use L^AT_EX for their typesetting needs, some conferences and almost the entire field of the social sciences still uses other formats such as OOXML (Microsoft Word). Since switching from L^AT_EX to these other formats is unacceptable for some scientists, it is necessary to create good, i.e. semantics preserving, methods to convert L^AT_EX documents to OOXML documents. The method that we use is to first convert the L^AT_EX documents to ltxml using L^AT_EXML and then use XSLT to transform the ltxml files to OOXML. Much progress has already been made on this front, with only list and reference functionalities still necessary to reproduce the core L^AT_EX macros. It is to be expected that this extension to L^AT_EXML will be functional in the coming year.

In the future, it would be viable to extend this treatment to more than just the core L^AT_EX packages and finally to ODT documents as well. This would essentially cover the breadth of conventionally used WYSIWYG file formats, allowing researchers to use L^AT_EX for all their typesetting needs, regardless of the requested fileformat.