

WDML: Position Statement

Volker Sorge

My concern with digital math libraries is primarily from the technical point of view. I am particularly interested in the recognition, processing, exposition and reuse of mathematics encapsulated in mathematical documents.

I believe this is an important aspect of a digital mathematics library, because why should people bother with using a DML if they get no more out of it than from an ordinary digital library, or even just dedicated search engines like Google Scholar or Microsoft Academic Search? And the latter often offer a much simpler, more intuitive interface than most digital libraries.

To distinguish a digital library for mathematics it has to offer added value that is beneficial for the user and unique to mathematics. Many of these points are related to the exploitation of mathematical formulas inside documents, making them amenable to search, copy-and-paste, reuse in mathematical software, text-to-speech, etc.

Technical Challenges Today

The main technical challenges today are certainly how to find, extract, recognise and correctly interpret formulas within the context of a document, in order to make them available to further processing. These problems are particularly pronounced in retro-digitised material where content is given in image format only, but they also exist in documents that were compiled into an electronic format from some source such as \LaTeX .

Our community already has quite a number of solutions for these problems and personally I am confident that most of these challenges will be solvable in the near future.

Technical Challenges of the Future

One might think that once the above problems are solved we will have no more challenges in the future: Most publications will be available in a suitable electronic form, compiled from adequate sources, which might even be available as well and can therefore be used as a basis to further processing.

I believe the contrary will be the case: **There will be more challenges for serving mathematical documents to readers in the future, than there are today.**

Already now the ways how we use/consume/read/access/search/process mathematics are drastically changing. For example, documents will be more often read on portable devices like iPads than traditionally as paper print. Therefore, a whole range of challenges will be related to supporting these new technology. And since mathematics will always be regarded as “niche”, it will be the task of our community to solve many of the accompanying scientific and engineering problems.

Here are just a few I can think of:

Zooming and Reflowing The smaller the device, the more the need to adapt documents dynamically to the display. Zooming and reflowing is already non-trivial for regular text. However,

for artefacts like tables, figures, formulas, etc. it becomes more and more challenging.

Search using a handheld device: It is already painful to enter full text on small devices. And, of course, the concept of a “formula editor” will be totally antiquated on a 4” screen. However, it is surprisingly easy to enter complex symbols (e.g. Kanji) by touch and gestures. So we will use handwritten math to search in printed documents.

Ever changing “Standards”: Currently we are in the lucky situation to have three quasi-standards: \LaTeX for authoring mathematics, MathML for displaying mathematics in web browsers and PDF for displaying and printing documents in general. All of these are fairly robust (i.e., a PDF document looks now pretty much the same on a PDF viewer as it did five years ago) and had a relatively long lifetime.

In the age of ebook readers, tablets and smart phones, there is a myriad of formats for electronic documents and it is currently more likely that format lifetimes will decrease rather than increase. And even if there is a temporarily accepted standard, such as for example epub could become one, it is unlikely that a document that displays well today will still do so after the next revision of the standard, on the next hardware device, or simply after the next software update.