

Access to Mathematics in Open Document Format (ODF) and OpenOffice.org

@Science Workshop "Access to Mathematics and Science",
Paris, France, 27 February 2008.

Author: **Christophe Strobbe**, Katholieke Universiteit Leuven, Belgium
(Christophe.Strobbe@esat.kuleuven.be).

Introduction

This presentation explores some issues of working with formulas in OpenOffice.org. This open-source office suite supports the Open Document Format (ODF), which became an ISO standard in 2006 (ISO/IEC 26300:2006)¹ and a number of other formats (for example those from Microsoft Office and the original StarOffice suite). When working with formulas in text documents, two components of the office suite come into play: the word processor "Writer" and the equation editor "Math".

The Math editor can be used on its own for creating and editing formulas. It can save formulas as OpenDocument Formula (.odf), OpenOffice.org 1.0 Formula (.sxm), StarMath 1.0 (.smf) en MathML 1.01 (.mml). Formulas can also be exported to Portable Document Format (PDF).

Writer allows users to embed formulas in text documents by going through the menus Insert > Object > Formula. This opens a Math window in Writer (the menus and the toolbar at the top are those of the Math component, but the save as and export options are those of Writer).

OpenOffice.org Accessibility Issues

OpenOffice.org Writer

- loses alternative text for formula objects between saves (both for normal text and for LaTeX code);
- sometimes (?) loses alternative text for images between saves;
- does not output alt attribute on images for formulas when saving as HTML if no alt text is available (invalid HTML).

1 Information technology -- Open Document Format for Office Applications (OpenDocument) v1.0: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485

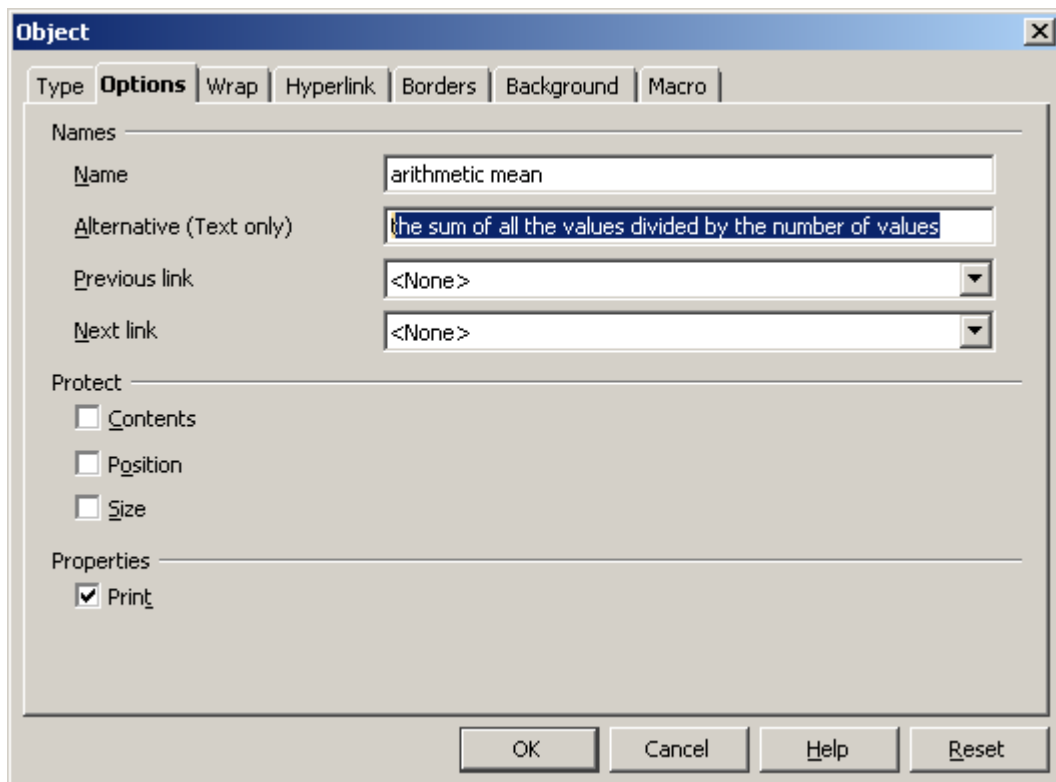


Illustration 1: Object properties of a formula: Options tab with name, alternative text and other properties

How ODF Stores Mathematics

ODF format used by Writer:

- basically an archive file (usually compressed, like a Zip file)
- that contains several XML files:
 - 'content.xml',
 - 'meta.xml',
 - 'settings.xml',
 - 'styles.xml' and
 - a few other files and directories.

OpenOffice.org stores a formula

- as an object
- referenced from the XML code in the "content.xml" file.

Example:

```
<draw:object xlink:href="./Object 1" xlink:type="simple"
xlink:show="embed" xlink:actuate="onLoad"/>
<draw:image xlink:href="./ObjectReplacements/Object 1"
xlink:type="simple" xlink:show="embed" xlink:actuate="onLoad"/>
```

- './ObjectReplacements/Object 1' = binary file;
- './Object 1' = directory with:
 - 'content.xml' = **MathML**;
 - 'settings.xml' = view settings, configuration settings (e.g. settings related to fonts).

ODF format used by Math:

- same archive format as Writer, but
- formula is stored in 'content.xml', which uses a modified MathML 1.01 format.

Exporting ODF Documents from Writer

Mathematical Markup Language (MathML)

'content.xml' files for formulas:

- one per formula,
- using MathML 1.01 (not the latest MathML version).

How to export MathML from ODF Writer document?

- Not possible through Writer's Save or Export dialogues!
- Unzip ODF file and locate correct 'content.xml' files: only OK for nerds.
- Select the formula in Writer, doubleclick the formula to open it in the editing interface of OpenOffice.org Math, copy the Math code, close the Writer document, open OpenOffice.org Math, paste the Math code, and save the file as MathML.
 - Too complicated.
 - Possible through keyboard interface??
 - Keep formulas in OpenOffice.org Formula files (created and maintained with OpenOffice.org Math) instead?
And copy them to Writer only when needed?

Extensible Hypertext Markup Language (XHTML)

Using "Export as ... XHTML":

- documents with formulas are saved as "XHTML 1.0 Strict", not with "XHTML + MathML" doctype;
- formulas are not exported in any way ...
- unlike images, which are exported with incomplete path to image file.

Hypertext Markup Language (HTML)

Using "Save as ... HTML Document":

- Writer warns about potential information loss in conversion;
- formulas are saved as GIF files ...
 - with alt attribute if alternative text was available;
 - without alt attribute if alternative text was not available (invalid HTML).

Examples

Below are some example of formulas that were used to explore the math features of "Writer" and "Math".

Arithmetic Mean

OpenOffice.org Math language:

```
overline x = {sum from i=1 to n x_i} over n
```

LaTeX code (text alternative):

```
\overline{x} = \frac{\sum_{i=1}^n x_i}{n}
```

Formula:
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Other alt text used for the object (cf. alt text bug mentioned above):
"the sum of all the values divided by the number of values".

(<http://en.wikipedia.org/wiki/Average>)

Variance

(<http://en.wikipedia.org/wiki/Variance>)

Sample Variance

OpenOffice.org code:

```
s^2 = { sum from i = 1 to n (x_i - overline x)^2} over { n - 1 }
```

LaTeX code (text alternative):

```
s^2 = \frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}{n-1}
```

Formula:
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Standard Deviation

OpenOffice.org code:

```
s = sqrt{s^2}
```

LaTeX code (text alternative):

```
s = \sqrt{s^2}
```

Formula: $s = \sqrt{s^2}$