EXT: General Office Displayer

Extension Key: **rlmp_officeimport**

Copyright 2000-2002, Robert Lemke, <rl@robertlemke.de>

This document is published under the Open Content License
available from http://www.opencontent.org/opl.shtml

The content of this document is related to TYPO3
- a GNU/GPL CMS/Framework available from www.typo3.com

Proudly made in France! *)

*) Fait en France avec fierté

Table of Contents

EXT: General Office Displayer.....1
Introduction.....1
What does it do?.....1
Where was it created?.....1
More Features.....2
Screenshots.....2
Administration.....4
About Microsoft Office 2003 file formats.....4
Users manual.....7
FAQ.....9
Installation.....9
Configuration.....10
Reference.....10
Cascading Stylesheets.....11
Testing file formats.....11
FAQ.....11
Known problems.....11
To-Do list.....11
Changelog.....11

Introduction

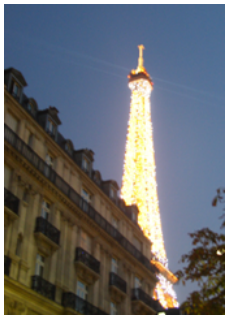
What does it do?

This extension provides import functionality for Microsoft Office 2003 Word and Excel documents (XML file type) as well as for Open Office Writer and Spreadsheet files. The contents of these files will be displayed in the layout of your website. They are even editable through the Rich Text Editor.

Where was it created?

This extension was kickstarted by Kasper Skårhøj and Robert Lemke on a 38°C hot summer evening on floor 1 of the Eiffel Tower in Paris[1], France, refreshed by the gentle breeze of the wings of History. The further development took place in boring, climatized offices though.

1 Paris is a city... :-)

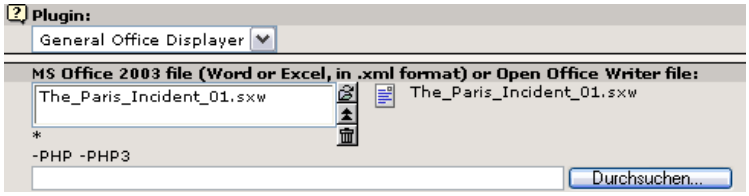


More Features

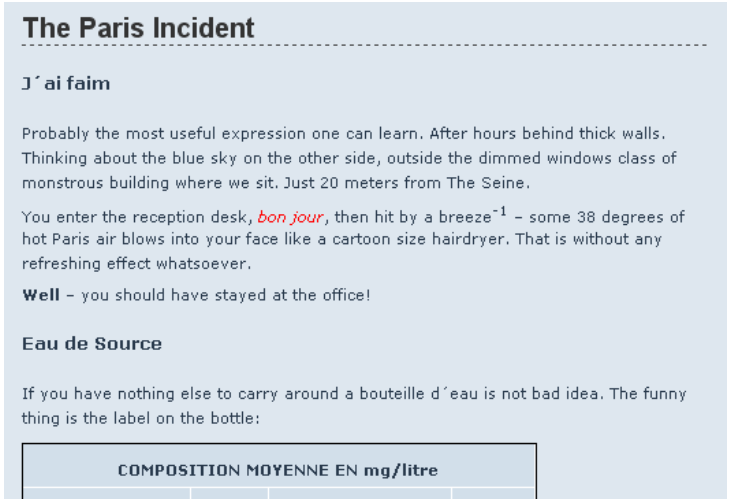
A powerful set of basic functions is compiled into this release. However there are many ideas for the future of this extension, and you probably have some custom needs for implementing it into your website.
The time spent so far on developing this plugin has not been sponsored nor payed by any client, it's just a gift to the TYPO3 community. If you need more than the basic functionality you might consider sponsoring new features by donating time or money.

Screenshots

There are two ways of uploading office files. The plugin-way displays the file directly on your webpage:



Actual output on your website:



The Paris Incident

J´ai faim

Probably the most useful expression one can learn. After hours behind thick walls. Thinking about the blue sky on the other side, outside the dimmed windows class of monstrous building where we sit. Just 20 meters from The Seine.

You enter the reception desk, *bon jour*, then hit by a breeze[-1] – some 38 degrees of hot Paris air blows into your face like a cartoon size hairdryer. That is without any refreshing effect whatsoever.

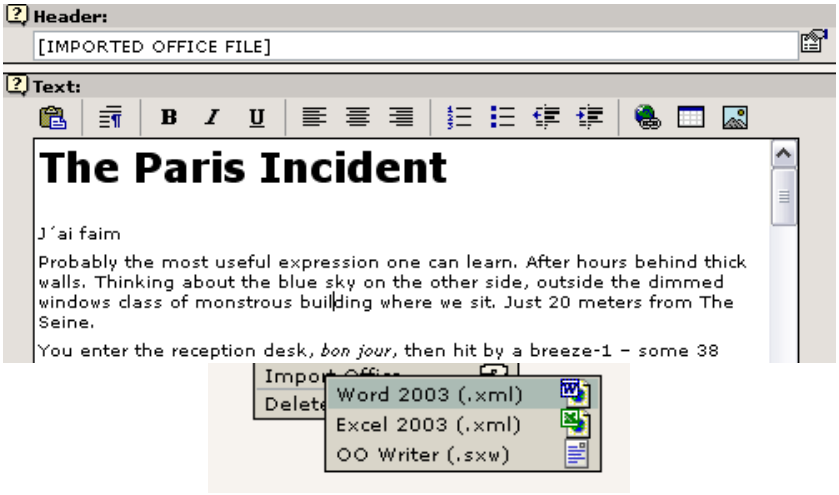
Well – you should have stayed at the office!

Eau de Source

If you have nothing else to carry around a bouteille d´eau is not bad idea. The funny thing is the label on the bottle:

COMPOSITION MOYENNE EN mg/litre

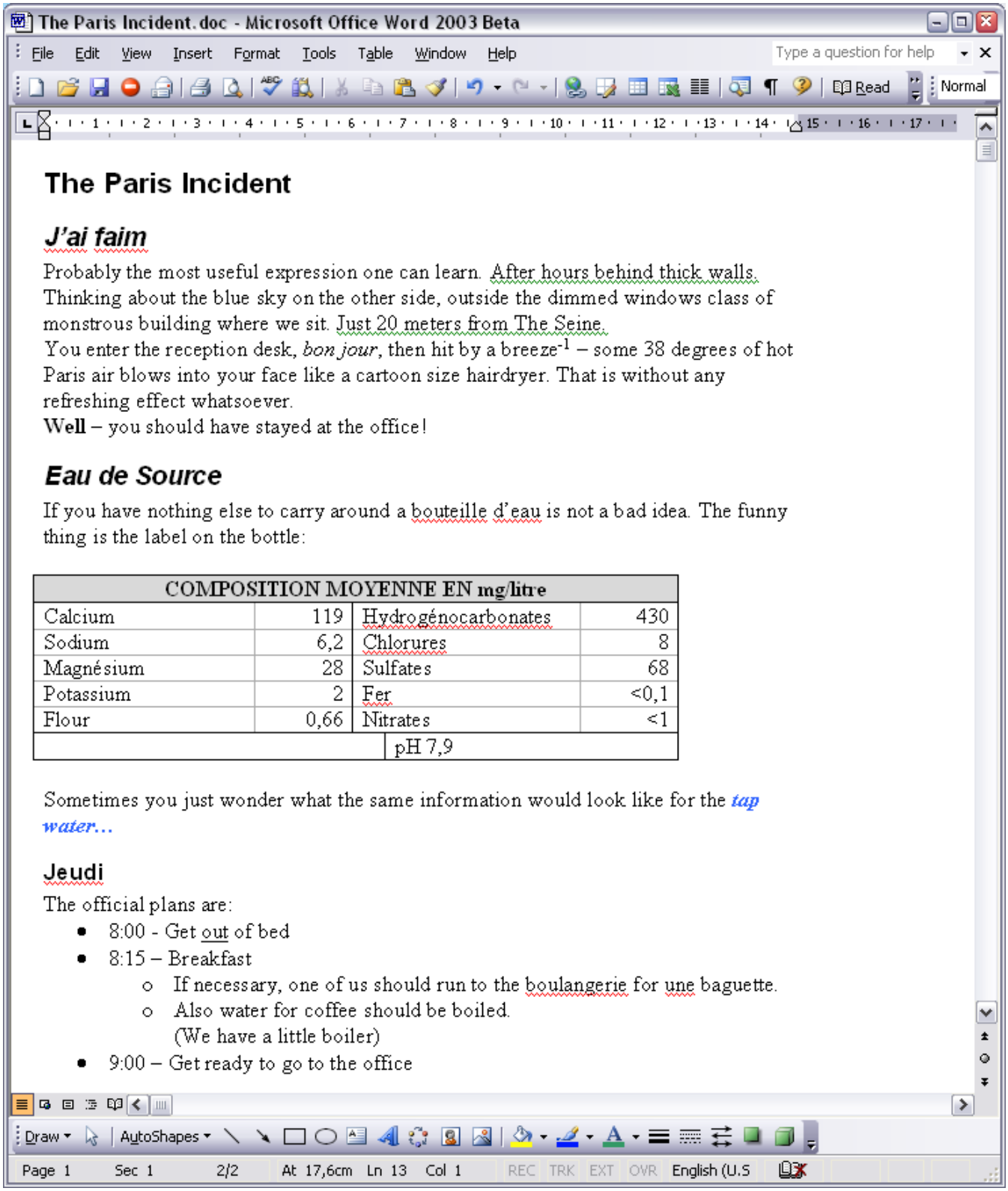
Doing it the import-way lets you edit the file's content in the Rich Text Editor:


Header:
[IMPORTED OFFICE FILE]
Text:
The Paris Incident
J´ai faim
Probably the most useful expression one can learn. After hours behind thick walls. Thinking about the blue sky on the other side, outside the dimmed windows class of monstrous building where we sit. Just 20 meters from The Seine.
You enter the reception desk, bon jour, then hit by a breeze-1 - some 38
Delete
Word 2003 (.xml)
Excel 2003 (.xml)
OO Writer (.sxw)

Administration

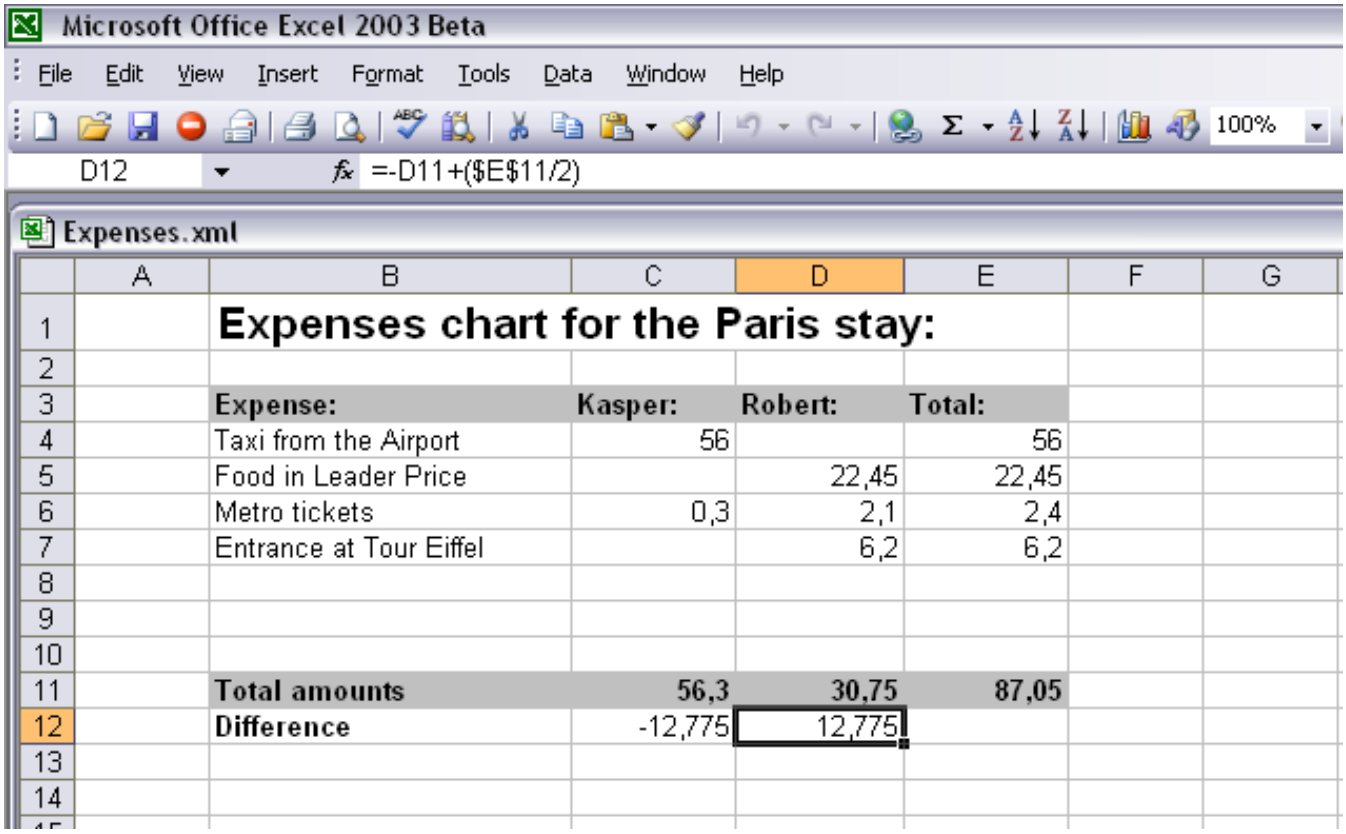
About Microsoft Office 2003 file formats

First of all, for those of you who are curious to see what MS Word 2003 looks like, here is a screenshot:



This screenshot shows the first part of the sample document found in the "Samples" folder of this extension ("The Paris Incident.xml" - there are also versions in .doc format and Open Office .sxw).

And for Excel here is another screenshot:



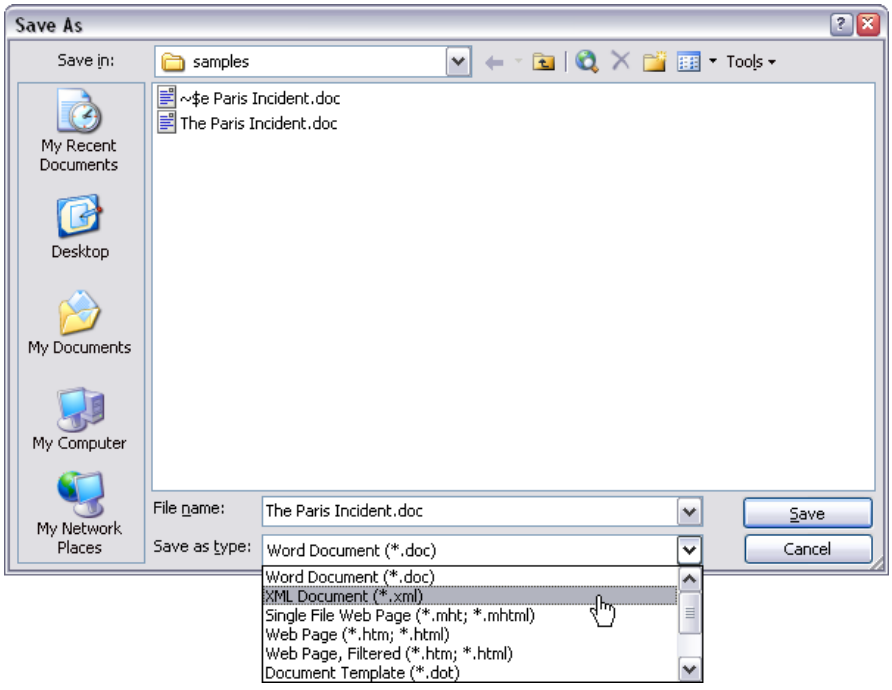
	A	B	C	D	E	F	G
1		**Expenses chart for the Paris stay:**					
2							
3		**Expense:**	**Kasper:**	**Robert:**	**Total:**		
4		Taxi from the Airport	56		56		
5		Food in Leader Price		22,45	22,45		
6		Metro tickets	0,3	2,1	2,4		
7		Entrance at Tour Eiffel		6,2	6,2		
8							
9							
10							
11		**Total amounts**	**56,3**	**30,75**	**87,05**		
12		**Difference**	-12,775	12,775			
13							
14							

This is how the Excel sample document looks (“Expenses.xml” - there are also versions in “.xls” and Open Office .sxc formats)

Saving in XML format

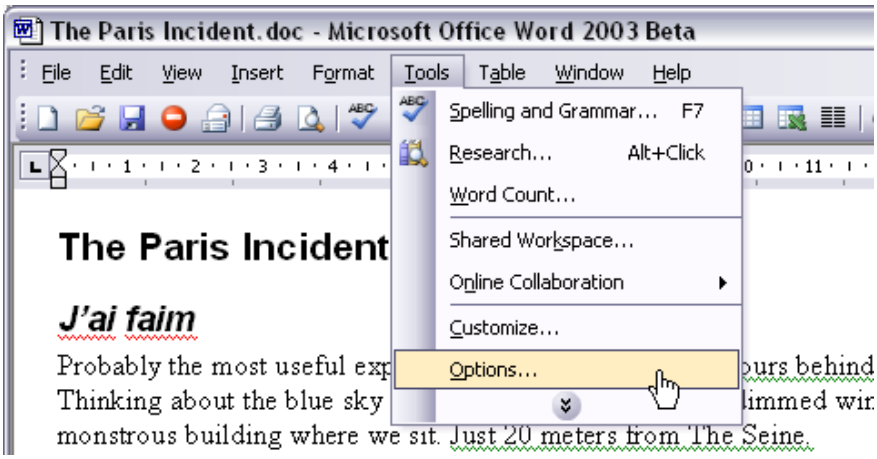
So does MSoffice 2003 really support XML natively? Well, yes and no on more levels.

On the user level you could say that it does *not* unless you configure it to do so! By default MS Office 2003 still writes the “.doc” or “.xls” formats and that is not XML at all. So you will have to ask you users to save in the new “.xml” format:

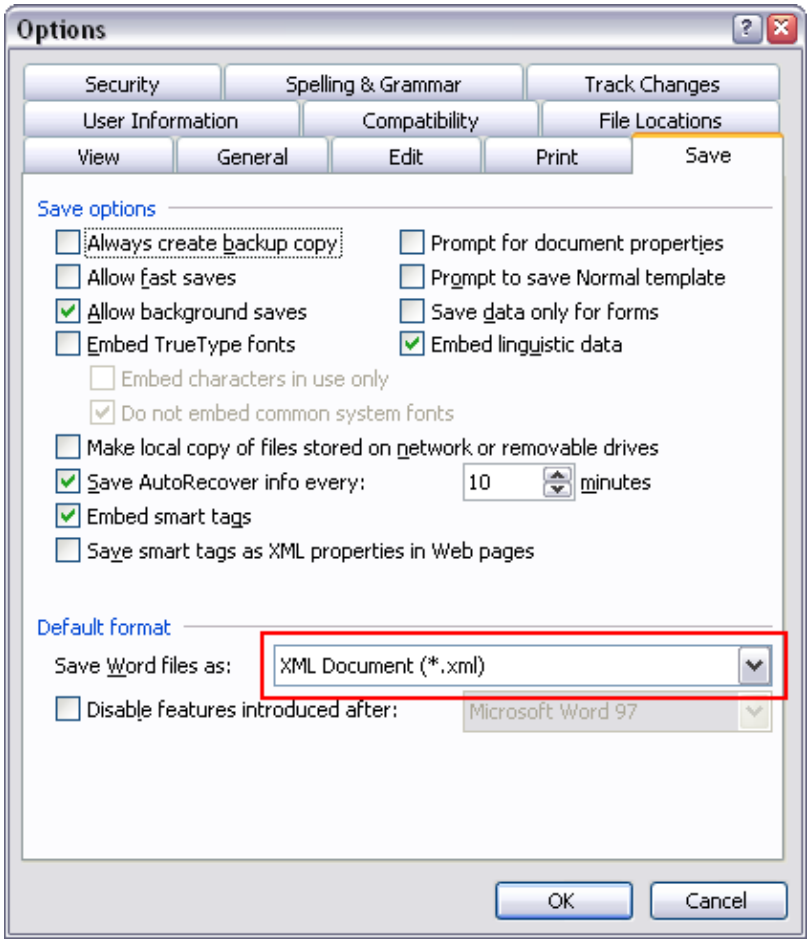


However that is easy enough as you can see.

Luckily there is an option to set the ".xml" format as the default! This was to great surprise for us as it just totally doesn't fit the usual subtle attempts from Microsoft to pseudo-support standards. But this time they did it right. To set the default saving format go to "Tools > Options"...



... select the "Save" pane and select ".xml":



Technically about the MSOffice 2003 XML format

Another part of Microsofts XML support regards whether you can actually get a schema for their XML files. Without such a thing you cannot be sure to support their formats fully. As far as I know these schemas are not publicly available. And so "XML support" looses a bit of its credibility.

However that has no serious effect on our support for this format since with a little intelligence applied and a lot of testing you can figure it out by yourself. In fact that is fun.

The new XML format of MS Office 2003 is different from the format of Open Office. Not only by the internal structure but also in the way binary files are stored. This is the different approaches they have chosen:

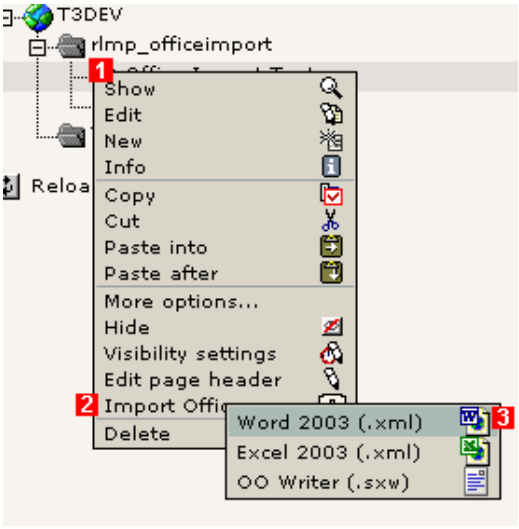
Product	Implementation	Advantages	Disadvantages
MS Office 2003	The ".xml" format of MS Office 2003 is a plain XML document in UTF-8 encoding. All binary data is also contained in the XML document by simply base64 encoding the data (and thus it fits nicely into a markup document).	One file, no need to unzip it, very easy to handle (at least for small documents).	It is large! *)
Open Office	The ".sxw" and ".sxc" formats from Open Office are in fact zip-archives. You can unzip them with WinZip if you like and inside you will find a set of XML files of which the one called "content.xml" contains the main content. The additional xml files carries meta information and stylesheet information. In a subfolder inside the zip-archive you can find all the binary files. The content.xml file only contain references to those files.	Low file size, since it's a zip archive.	Harder to handle since it must be unzipped first.

**) And it will be interesting to see if the much larger filesize will generally demotivate people to adopt the ".xml" format as their default format of future MS Office files - if so, we can again conclude that Microsoft has gotten the best of both worlds; recognition for supporting standards while their product demotivates people to use the standards....*

Users manual

Importing into a new content element (with the Rich Text Editor)

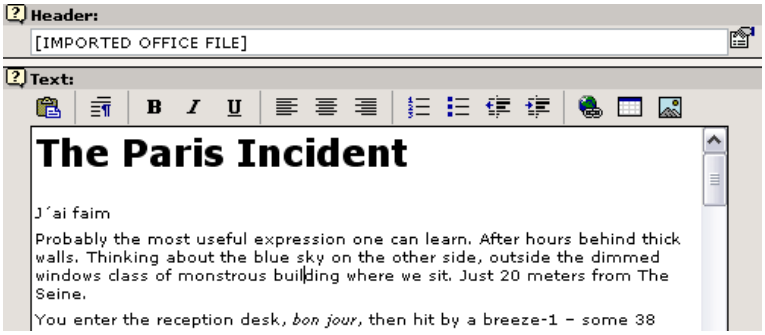
You may import an office document directly into the Rich Text Editor for further editing.



In the page tree just click on the page icon of the page which will hold the content of your document (#1). In the click menu choose "Import Office" (#2) and select the document type you are going to upload (#3).

A new dialog appears asking you for the office file. Select a file of the correct type from your harddrive:

Click upload. Transferring your document from the local harddisc to the server might take a while. When it's uploaded you'll see the content displayed in the Rich Text Editor:

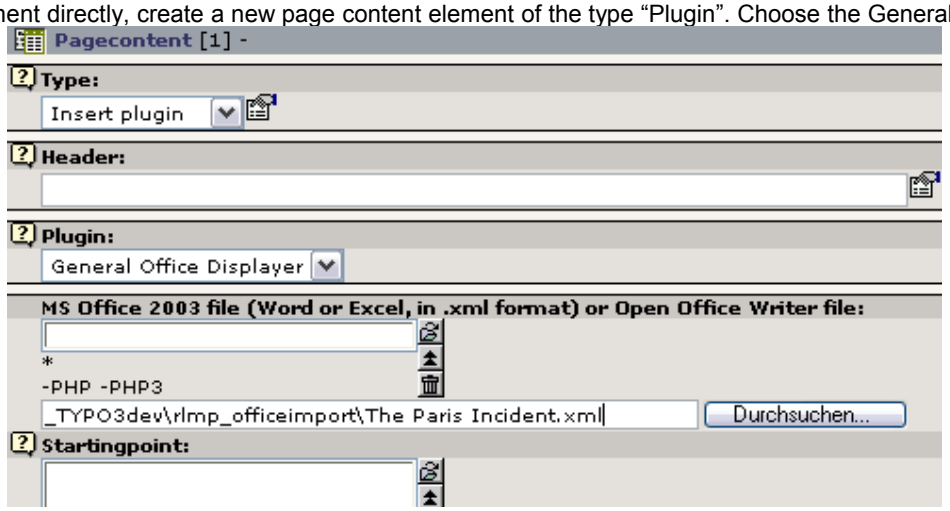


The content of the office document has now been stored in a regular content element which has been automatically created on the page.

Displaying the document by a plugin content element

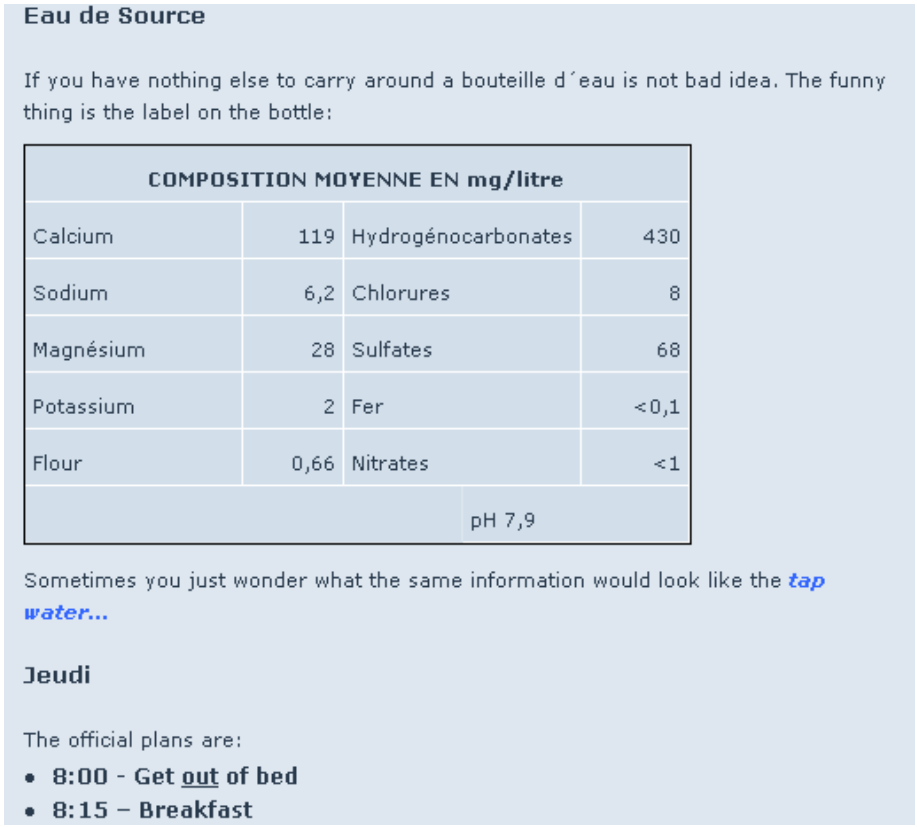
If you don't want to import your office document so that you can edit the content, you can choose to upload and display it directly. The layout will be displayed by using the design of your own website, depending on the settings your webmaster made.

To import a document directly, create a new page content element of the type “Plugin”. Choose the General Office Displayer:



Now you select an office file directly from your harddrive or use the element browser for selecting a document previously copied to your webserver.

Save and display the page and the content of your office file will be nicely displayed on your website:



Eau de Source

If you have nothing else to carry around a bouteille d´eau is not bad idea. The funny thing is the label on the bottle:

COMPOSITION MOYENNE EN mg/litre			
Calcium	119	Hydrogénocarbonates	430
Sodium	6,2	Chlorures	8
Magnésium	28	Sulfates	68
Potassium	2	Fer	<0,1
Flour	0,66	Nitrates	<1
		pH 7,9	

Sometimes you just wonder what the same information would look like the *tap water...*

Jeudi

The official plans are:

- 8:00 - Get out of bed
- 8:15 – Breakfast

Understanding the difference of the approaches

It is important to understand the difference of the two approaches of either *importing* the content (into the Rich Text Editor) or *displaying* the content (by a plugin). This is summarized here:

- *Import* a document if you just need to transfer the content and possibly edit it inside of TYPO3.

  Notice: You cannot easily update the content from the original office document later unless it is acceptable for you to do a *new* import into a *new* content element which you manually have to swap with the “old”.

- *Display* a document (with the plugin) if you need to upload revised versions of the same document many times.

  Notice: By this method it is not possible to edit the document online in TYPO3 - only to display it. Your original should be the document on your harddrive and when you have made a revised version you upload the document again as a substitute for the document currently online.

FAQ

Q: *When I try to upload an office file an error message tells me that the file is too large. How come?*

A: The maximum file size defined in this extension is 10 Megabytes. However the administrator of your Website might have chosen to restrict the file size – or you even exceeded the 10 MB allowed by the extension

Q: *The document doesn't display exactly like it does in Word / Excel / OO Writer / OO Spreadsheet.*

A: That's right. But it hopefully fits into the layout of your website. TYPO3 only uses the content and structure of document in order to display it. Only few parts of the layout will match the original.

Installation

Just install this extension like any other by using the Extension Manager. It depends on another extension called *libunzipped*. If you did not install *libunzipped* already, you'll have to do that first. This extension is necessary to work with Open Office files.

After the installation of *libunzipped* you'll be asked for an unzip application commandline. That is the command you would enter at your console to unzip a zip archive. *libunzipped* was tested on Linux and Windows servers using the tool *unzip* which is available at http://www.info-zip.org/pub/infozip/UnZip.html.

Application:	Example commandline:
unzip (Linux)	unzip -qq ###ARCHIVENAME### -d ###DIRECTORY###
unzip (Windows)	unzip.exe -qq ###ARCHIVENAME### -d ###DIRECTORY###
WinRAR (Windows)	c:/Program Files/WinRAR/winrar.exe x -afzip -ibck -inul -o+ ###ARCHIVENAME### ###DIRECTORY###

Configuration

The General Office Displayer has a bunch of default values which assure that office documents render nicely even directly after installing this extension. In fact for many well programmed websites using cascading stylesheets no changes have to be made in order to get a nice result matching your overall layout of the project.

However you might want to change some of the options being responsible for the rendering of your documents. Simply overwrite the properties using the Template Object Browser or writing them directly into your TypoScript template. The table below contains all options you currently have. Most options apply for both, Microsoft Office and Open Office documents. There are some special options though.

Working with custom styles

The default settings will handle most of the basic styles which are available in your office documents. Additionally you may use your own formats / styles and define how TYPO3 should render them. Imagine you have a custom style “MyPreformattedStyle”. Simply add the following TypoScript to your template:

```
plugin.tx_rlmpofficeimport.tagWraps.mypreformattedstyle = <pre> | </pre>
```

... and all your custom style occurences will render preformatted.

Note: You must not use spaces or special characters in your custom style names.

Reference

Property:	Data type:	Description:	Default:

imageCObject_scaledImage	IMAGE cObject	This configuration will determine how images are being displayed. Use the same configuration as for the IMAGE content object (see the TSRef).
		
		Currently this code is only used for Microsoft Office documents, except the image width which also applies to Open Office files.
		
		This is the default:
		
		`imageCObject_scaledImage {`
		`file.width = 100`
		`file.import.current = 1`
		`imageLinkWrap = 1`
		`imageLinkWrap {`
		`width = 800`
		`JSwindow = 1`
		`enable = 1`
		`}`
		`wrap = <div style="text-align:center; margin-bottom: 10px;"> | </div>`
		`}`
			see description

tagWraps	wraps	The HTML code used for rendering the office documents is being defined by this object. There are some fixed styles and you may custom styles as well. The following wraps are the built-in fixed styles:
		
		`heading1 = <h1> | </h1>`
		`heading2 = <h2> | </h2>`
		`heading3 = <h3> | </h3>`
		`heading4 = <h4> | </h4>`
		`heading5 = <h5> | </h5>`
		`paragraph = <p> | </p>`
		`bold = <strong> | </strong>`
		`italic = <em> | </em>`
		`underlined = <u> | </u>`
		`unorderedlist = <ul> | </ul>`
		`listitem = <li> | </li>`
		`superscript = <sup> | </sup>`
		`subscript = <sub> | </sub>`
		`preformatted = <pre> | </pre>`
		`indented = <blockquote> | </blockquote>`
		
		If you use a custom style in your word / writer document, you may even define an HTML code for that. Say we used some style in open office which we called “myownpreformatted”. You'll define an HTML code by adding the following wrap:
		
		`myownpreformatted = <pre class=”someclass”> | </pre>`
			See description

parseOptions	parse	See the list below for additional parse options	

[plugin.tx_rlmpofficeimport]

Property:	Data type:	Description:	Default:
renderMicrosoftSmartTags	boolean	If set, cities and places recognized by Microsoft Office 2003 will be wrapped by the following code: | *Please don't ask what this feature is for ...*	1
renderColors	boolean	If set, (font-)colors found in the document will be rendered.	1
renderBackgroundColors	boolean	If set, background colors found in the document will be rendered.	1
renderFontFaces	boolean	If set, font faces found in the document will be rendered.	1

[plugin.tx_rlmpofficeimport.parseOptions]

Cascading Stylesheets

The whole output will be wrapped into a DIV tag with class="tx-rlmpofficeimport-pi1". By that you should be able to take influence on the layout of most tags being used. Otherwise just use the tagWraps options explained above.

Testing file formats

For your convenience (and for the memory of our session at Tour Eiffel ...) we provide a document for each format you can use to test the plugin. You find them in the *samples/* folder in the extension's directory.

FAQ

Q: *Why don't you support Microsoft Word .doc files?*
A: Office 2003 provides XML output – like Open Office does just for longer time. XML is much easier to parse than a proprietary format like the .doc files. However, if there is a real need for supporting .doc files even that might be implemented some day, ideas for some basic import features already exist.

Q: *Why do you support Microsoft Office 2003 files at all? It's not Open Source, so ...?*
A: It's a wide-spread format and many customers are using it of course.

Q: *Microsoft Office 2003 is not released yet (August 2003), can you send me a copy of the beta?*
A: No, sorry.

Q: *I didn't find the Open Office Spreadsheet support you were talking about. Where is it?*
A: Well, I just didn't have the time to implement it yet. Just wait some more.

Q: *Could you please implement XXX? I really need that feature!*
A: Yes, of course. You might consider sponsoring new features or wait until maybe your feature will be implemented some day.

Q: *What do you plan for the future of this extension?*
A: You might want to look at our ToDo list. Just see the TODO.txt file in the extension's doc folder. It contains our notes about developing this extension.

Q: *I use this extension on a windows driven surver / with the quickstarter package. If I try to upload an Open Office document nothing happens, why?*
A: Open Office documents consist of many files zipped into a single file. In order to unzip these, a program called "unzip" is needed. Support for windows binaries of such an unzipping program is still untested. On Linux platforms such a binary might easily be installed (see the section above for more information about unzip).

Known problems

- Full configuration for the IMAGE object only applies to MS Office Word files
- Images are not rendered when importing through the RTE

To-Do list

- Support for Open Office Spreadsheets
- Support for writing back into the formats?
- API for export of database records into Excel / Calc spreadsheet (could be used from many places within TYPO3 where data is exported!)
- API for re-import of such a spreadsheet -> this would enable people to export data, edit in Calc/Excel and then re-import / syncronize with the database.
- Configuration of allowed elements / tags, useful for importing into the RTE
- Browse function (show content on different pages instead of one)
- Support for big documents

- Organization of many documents
- ...

Changelog

- 1.0.4 - BUGFIX: for function createContentElement (file: cm1/index.php):
- 1.0.3 - Added icons - Thanks to Netcreators for providing them!
- 1.0.2 – Cleaned up the default values for tag rendering: <b> and <i> changed to and . <tr> and <td> don't have their own classes anymore, use inherited CSS definitions instead.
- 1.0.0 – First public release