

EXMARaLDA Add-In for MS Excel - Version 0.9.4.2

1 Summary

The EXMARaLDA Add-In for MS Excel is a freely available (GPL) plugin written using the Office VBA API which allows users of Excel to import data from the EXMARaLDA XML format into a spreadsheet (preserving cells and spans, as well as most types of metadata – see below for details) and back again from a spreadsheet to XML. The add-in is compatible with Office XP, 2003 and 2007 under Windows XP, Vista or Windows 7 and comes with absolutely no warranty. The latest version of the add-in can be found on <http://exmaralda.org>.

2 Installing

To install the add-in follow these steps:

1. Copy the file *exmaralda_io_0.9.4.2.xla* to the directory that holds your Excel add-ins. On an English language Windows XP installation this is usually:

```
C:\Documents and Settings\YOUR_USER_NAME\application
data\Microsoft\AddIns
```

- Other languages may have slightly different paths (e.g. “Dokumente und Einstellungen” for “Documents and Settings” on German Windows).
2. Open Excel and choose *Tools -> Add-Ins...* (again, other languages may use different names, e.g. German *Extras* for *Tools* etc).
 3. Click *Browse...* and navigate to the file *exmaralda_io_0.9.4.2.xla*. Once the file is chosen, the add-in *Exmaralda_Io_0.9.4.2* should appear on the list of add-ins.
 4. Check the box next to *Exmaralda_Io_0.9.4.2* and click OK. A new menu called *Exmaralda* should now appear in the menu bar above.

3 Uninstalling

To remove the add-in simply go to *Tools -> Add-Ins* and uncheck *Exmaralda_Io_0.9.4.2*. By pressing OK the Exmaralda menu will now be removed. You may then optionally delete the file *exmaralda_io_0.9.4.2.xla* if you wish.

4 Usage

The Exmaralda Add-In menu in Excel provides three basic functions as detailed below.

4.1 Import from Exmaralda

Choosing this function will open the Import Form. The form allows users to choose whether or not metadata applying to an entire Exmaralda document will be imported, and similarly for metadata applying to each tier within a document. Document metadata is stored in the Excel file's *Comments* property, which keeps the entire header of the original XML file.

Each tier is imported in a separate column in Excel, so that horizontal layers from EXMARaLDA are represented vertically as in the image below (this is because Excel can only accommodate 256 columns; see “Transposing” below for an EXMARaLDA style horizontal view). The user may determine where imported data will appear by selecting a cell or cells – data will then be imported starting from the top-leftmost selected cell. Each tier’s *category* attribute is used as a column header in bold. Tier metadata is stored in a comment to the header cell of each column. Note that at present **no metadata contained in <ud-tier-information> and <ud-information> tags within a <tier> element** is imported (support for this will probably be added in a future version. A further type of element which is ignored by the importer at present is the **<tierformat-table> of .exb files**, meaning that formatting information such as background colors for each tier etc. are not carried over.

	0	1	2	3	4	5	6
[word]	Dieser	Text	kommt	aus	dem	Buch	"
[lemma]	dies	Text	kommen	aus	d	Buch	"
[pos]	PDAT	NN	VVFIN	APPR	ART	NN	\$(
[matrix-satz_felder]	VF_MS		LSK_MS_1		MF_MS_1		

When importing multiple files, users can select whether the columns representing the annotation levels of each document should be imported side by side, each document in a separate Excel workbook, or continuously underneath each other (usually only suitable if all documents have precisely the same layers in the same order). When documents are imported underneath each other, a header row is only generated for the first document, and the other documents are assumed to have exactly the same columns. If multiple files are imported into the same workbook and importing of document metadata is still selected, only the first document’s header will be inserted into the *Comments* property. Finally, it is also possible to omit certain layers that should not be imported. To do so, select *do not import tiers* and input the names of the tier categories that should be ignored, separated by a semicolon, e.g. *lemma;pos*

4.2 Export to Exmaralda

A selected range of cells may be exported to EXMARaLDA XML using the *Export to Exmaralda* menu. The first row of the exported data is assumed to contain the category names for each annotation layers, and all other rows are treated as data proper. If the *Comments* field of the Excel workbook contains an EXMARaLDA header (usually generated by the importer), this may be used for the exported file, otherwise a generic header is generated automatically. Similarly, tier attributes stored by the importer in comments to the header cell of each layer may be exported, with the exception of user defined tier metadata in **<ud-tier-information> and <ud-information> tags** within tiers. If no tier attributes are found, default values are inserted, which assume that the first column is a transcription layer (type="t") and all other layers are annotation layers (type="a"), and that the display name of the layer is the same as the category name.

4.3 Transpose Selection in a New Sheet

Since an Excel spreadsheet can only contain a maximum of 256 columns but as many as 65536 rows, imported data is presented vertically by default. It is however possible to transpose a selection into a horizontal format similar to the EXMARaLDA editor using this command. The first row is assumed to contain layer headers and is rendered in bold face. If the selected data contains more than 256 rows, the rows are broken up into groups of up to 256 rows each and the headers are repeated for each block to make navigation easier. Note that some spans may cross the border of a block of 256 lines. In such cases the add-in will attempt to find a cutoff point starting at a 200 row block. Data that cannot be split up in this way (or in any way, if there is a span of more than 256 rows), will not be able to be transposed.

5 Limitations and Assumptions about the Data

For the sake of simplicity and speed in handling large files, the add-in does not actually parse the XML on import, but rather reads each file line by line, making certain assumptions about the order of elements and attributes. The add-in supports the following variations:

- Time line events' start and end attributes may stand in any order
- Time line elements may have labels and may be numbered arbitrarily
- For each tier, a category attribute is expected, which may appear anywhere

In all other cases it should be assumed that the importer expects data that looks like the data produced by the exporter. That said, the importer is somewhat robust in that it ignores anything it does not understand: this means it will happily import invalid XML if it also contains the elements it expects. If you run into problems caused by hidden assumptions made by the importer, please file a bug report, ideally with a sample of the data which caused the problem.

6 Contact

For any questions, bug reports or feature requests regarding the add-in please contact Amir Zeldes (amir {dot} zeldes {at} rz {dot} hu-berlin {dot} de).