

A CONTEXT ANALYTICAL INDEX



PREFACE

On the convention of the notations adopted

The syntax and commands have `COURIER NEW` font, optional parts are enclosed in [square brackets] and variables are in *italics*.

Examples are also in `COURIER NEW` font.

Warning

All scripts (SOFTWARE PRODUCT) are provided by El Condor "as is" and "with all faults." El Condor makes no representations or warranties of any kind concerning the safety, suitability, lack of viruses, inaccuracies, typographical errors, or other harmful components of this SOFTWARE PRODUCT. There are inherent dangers in the use of any software, and you are solely responsible for determining whether this SOFTWARE PRODUCT is compatible with your equipment and other software installed on your equipment. You are also solely responsible for the protection of your equipment and backup of your data, and El Condor will not be liable for any damages you may suffer in connection with using, modifying, or distributing this SOFTWARE PRODUCT.

You can use this SOFTWARE PRODUCT freely, if you would you can credit me in program comments:

El Condor – CONDOR INFORMATIQUE – Turin

Comments, suggestions and criticisms are welcomed: mail to rossati@libero.it

Arguments table

1 CONTEXT ANALYTICAL INDEX.....	1
1.1 Introduction.....	1
1.2 How use the package.....	1
1.2.1 Document.....	2
1.2.2 Paragraph style.....	2
1.2.3 Words file.....	2
1.2.4 Context Index Font properties.....	2
1.2.5 The page numeration.....	2
1.2.6 Links creation.....	2
1.2.7 The context index.....	3
1.2.8 Call directly the index generator.....	3
2 TECHNICAL REFERENCE.....	4
2.1 MS Basic.....	4
2.1.1 Regular Expression.....	4
2.1.2 Enable Macro developing.....	4
2.1.3 Enable link on PDF.....	4
2.2 OoBasic.....	4
2.2.1 File dialog.....	4
2.2.2 Inspect objects.....	4
2.2.3 Dialog methods.....	4
2.2.4 ThisComponent.....	5
2.2.5 Dispatch commands.....	5
2.2.5.1 Table commands.....	5
2.2.5.2 Text commands.....	5
3 INDEXES.....	6

1 Context Analytical Index

1.1 Introduction

This package create an analytical index where the words of the index are shown with the part of the phrase which contains. It has developed for a OpenOffice/LibreOffice writer and for MS Word documents.

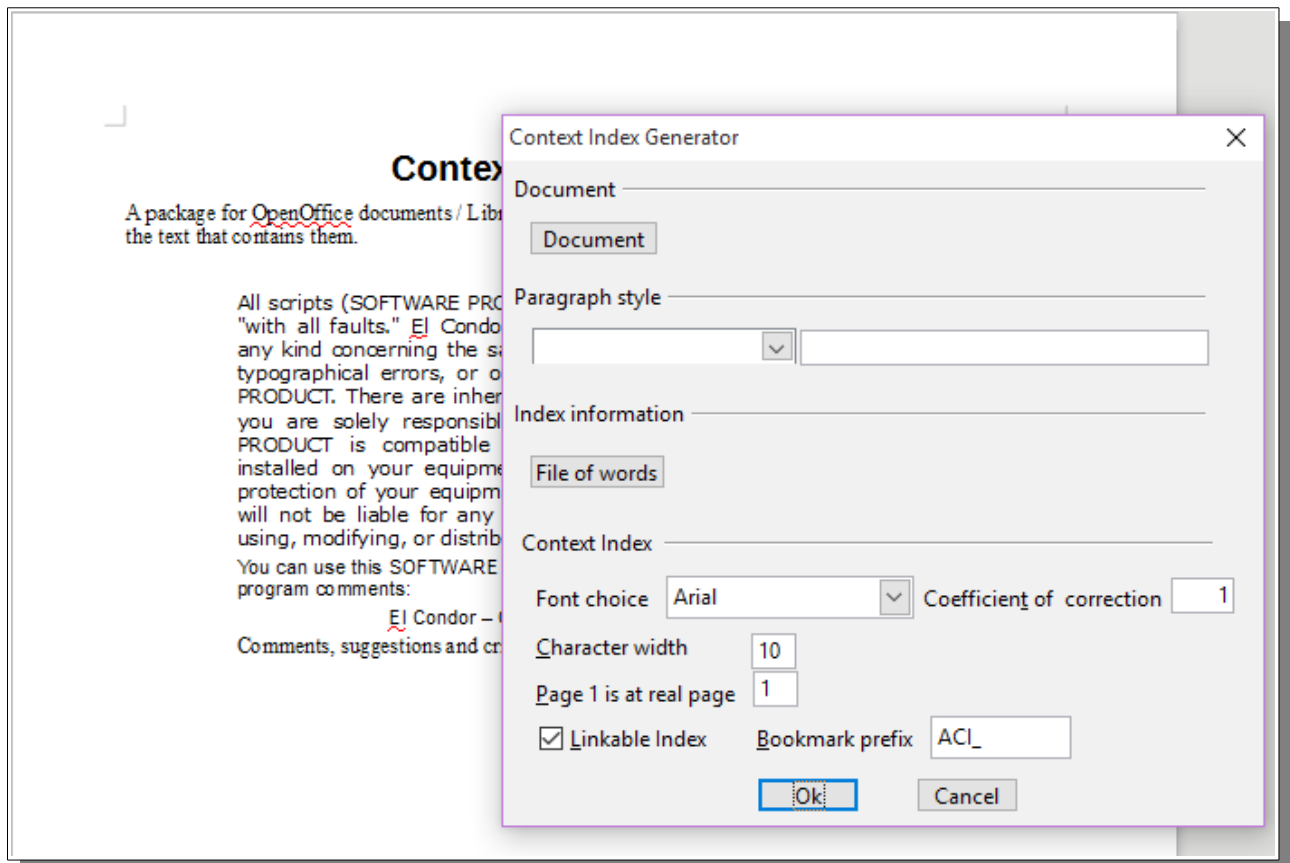
Context Analytical Index is based on a set of macro Basic for scanning an odt or doc or docx file in order to generates a context index based on a list of words.

The macro extracts the text paragraphs for a selected style(s), together the relative page and search in the text the words; for every word found it is created a row in a spreadsheet containing the word, the context fragment and the page. At the end of the scanning the sheet is ordered by word and it is used for create a context index at the end of the document itself.

This software has been developed on LibreOffice (portable) version 5.0.2.2, OpenOffice 4.1.1 and Ms Word 2010.

1.2 How use the package

The package is contained in the document contextIndexBuilder.odt, (contextIndexBuilder.doc for MS Word), starting his Main subroutine appears a dialog form where input the requested data:



- Document: it is the document of which we want to get the index.
- Index information: this file contains the words for create the index.
- Paragraph style: the combo box is populate by the style of the Document; when a style is selected it is inserted in the text field at right.
- Context Index: information for the style of the index.

1.2.1 Document

Document must be chosen before the choice of the paragraphs.

1.2.2 Paragraph style

You must choose the style or styles to index, the choices appears in the text field at right.

OpenOffice/LibreOffice The style names that appear in the document are in English, except of course those created by the user, for example an Italian writer see this paragraph style as “Stile predefinito” where in the combo it appears as “Standard”.

The package consider only the paragraph text, not the possibly text on tables.

MS Word The style names are on the user language.

1.2.3 Words file

This file contains a list of words, one for line, for which the index is to be created.

The words can have a wild-card (*, +, ?) both at start both at end; they allow you to enter in the index not only words but also words with prefixes or suffixes, for example *link* can intercept link, delink, links etc.. Obviously the order in the index is given by the words without wild-cards.

Wild-cards are translated into a regular expression for intercept only letters:

- * becomes [a-zA-Z]* i.e. 0 or multiple letters,
- + becomes [a-zA-Z]+ i.e. 1 or multiple letters,
- ? becomes [a-zA-Z]? i.e. 0 or 1 letter.

The search ignore the case and it is bounded at word, for example the item fontname? generates the regular expression \bfontname[a-zA-Z]? \b¹,

If we need a complex search, the word must be a regular expression, in this case the word for sort must be indicated after the regular expression, separated by a tabulation character (hexadecimal 09); for example for search wolves and wolf the regular expression can be: \b(wolf|wolves) \b.

At right a sample of the file of words.

```
style
\bfont ?name\w? fontname
\bfont ?size\w? fontsize
bookmark?
regular expression
*link*
OpenOffice/LibreOffice
MS Word
\b(wolf|wolves) \b    wolf
PDF
```

1.2.4 Context Index Font properties

It indicate the Font Name and Fontsize of the index generated. It is possible input a Font name not present in the list of Font names.

The context is about 45 characters before and after the word and it is based on Arial Font of dimension 10; the length of the context is rearranged if the Fontsize changes; the number of characters can be further adjusted, this correction is proposed for the following fonts:

- Courier new 0.94
- Times New Roman 1.06
- Verdana 0.93

1.2.5 The page numeration

OpenOffice/LibreOffice The page number obtained is the absolute page number, if it is different from the page number printed, it can be adjusted by the dialog field page 1 is at real page.

MS Word The page number printed is obtained by asking the wdActiveEndAdjustedPageNumber of the paragraph; normally it does not require any correction.

1.2.6 Links creation

If the check box Linkable index is selected a link pointing to the paragraph is created for every item of the index via a bookmark associated to a paragraph whose name prefix can be specified in its text field.

1 \b tells the word boundary

1.2.7 The context index

The context index is generated in a table of three columns; at the start this table is deleted if exists and then recreated at the end of the document.

The first column contains the phrase before the word, the program try to extract the start of the phrase; the second column contains the word and the rest of the phrase, possibly until the point; the third column contains the page were is located the word, the page number can be linkable.

The table has the name `ContextIndex` in LibreOffice/OpenOffice; in MS Word `ContextIndex` is the name of a Bookmark.

1.2.8 Call directly the index generator

We can directly invoke the index generator provided to set all the required information, as in the example below (OpenOffice/LibreOffice) by the `callNoDialog` subroutine:

```
Sub callNoDialog
    fontName = "Arial"
    fontSize = 10
    pageOffset = 1-5
    Correction = 1
    bookMarkFlag = true
    bookMarkPrefix = "ACI_"
    FileDoc = ConvertToUrl("C:\D\ContextIndex\contextIndex.odt")
    fileIndx = ConvertToUrl("C:\D\ContextIndex\ContextindexNames.txt")
    ParagraphStyle = "Standard"
    extractParagraphs(linkDocument(FileDoc), fileIndx, ParagraphStyle)
End sub
```

2 Technical reference

This section contains some notes relating to product development, it is, obviously, not exhaustive.

2.1 MS Basic

2.1.1 Regular Expression

Regular expression functions must be enabled i.e. we need to add VBA reference to "Microsoft VBScript Regular Expressions 5.5":

- Select **Developer** tab
- Select **Visual Basic** icon from **Code** ribbon section
- In **Microsoft Visual Basic for Applications** window select **Tools** from the top menu.
- Select **References**
- Check the box next to **Microsoft VBScript Regular Expressions 5.5** to include in your workbook.
- Click **OK**

2.1.2 Enable Macro developing

- Select **File** tab, select **Options**
- Select **Customize ribbon**
- In the list of **Main tabs** check **Developer**
- Click **OK**

2.1.3 Enable link on PDF

In some 64-bit Microsoft Office installations, users run into trouble when trying to use the embedded links in converted PDFs. If the links don't work, you need to change some options and then reconvert the document. Go through the steps once more and check the "Minimize size (Publishing Online)" check box on the Save As page. Click on the "Options" button and then in the resulting dialog box, uncheck the "ISO 19005-1 compliant (PDF/A)" check box. Click "OK" and "Save." The embedded links in the converted PDF file will now work.

2.2 OoBasic

2.2.1 File dialog

In order to select a default directory of the FilePicker. we must use a OfficeFilePicker like reported By Andrew Pitonyak in OpenOffice.org Base Macro Programming:

In OOo version 2.0, the Win32FolderPicker has a bug, which prevents you from setting the initial directory. Use **Tools > Options > OpenOffice.org > General** to open Options dialog, and then check the *Use OpenOffice.org Dialogs* checkbox. Baring that, you can explicitly use the OfficeFolderPicker service rather than the FolderPicker service.

2.2.2 Inspect objects

For a list of methods and properties of OpenOffice objects the object generally has the properties

Dbg_Properties and Dbg_Methods.

```
oDialog = CreateUnoDialog(DialogLibraries.Standard.ContextIndexDialog)
MsgBox oDialog.Dbg_Properties
MsgBox oDialog.Dbg_Methods
```

2.2.3 Dialog methods

A dialog is created by: `oDialog = CreateUnoDialog(DialogLibraries.Standard.dialogName)`

```
oDialog.Execute()
```

```
oDialog.GetControl("controlName")
```

```
oDialog.endExecute()
```

```
oDialog.setPosSize(X,Y,width,height,wathSet)
```

1=x, 2=y,

```
oDialog.setPosSize(0,0,420,370,12) '12 sets width and height 4=width, 8=height
```

2.2.4 ThisComponent


`thisComponent.URL` the document name (and location) in form of URL

2.2.5 Dispatch commands

2.2.5.1 Table commands

Command	Parameters	Note
	Name	Value type
DeleteTable		
EntireColumn		
GoToEnd		
GoToStartOfTable		
	TableName	string
	Columns	integer
InsertTable	Rows	integer
	Flags	integer
JumpToNextCell		
SelectTable		

2.2.5.2 Text commands

Command	Parameters	Note
	Name	Value type
Bold	Bold	boolean
GoToEndOfDoc		 Works if is out of a table
InsertText	Text	string
LeftPara	LeftPara	boolean
RightPara	RightPara	boolean
TopBottomMargin	TopBottomMargin.TopMargin	integer
	TopBottomMargin.BottomMargin	integer
	TopBottomMargin.ContextMargin	integer
	TopBottomMargin.TopRelMargin	integer
	TopBottomMargin.BottomRelMargin	integer

3 Indexes

Context Analytical Index

... is created for every item of the index via a bookmark associated to a paragraph whose name prefix ...	2
... in MS Word ContextIndex is the name of a Bookmark .	3
... it is bounded at word, for example the item fontname? generates the regular expression ...	2
It indicate the Font Name and Fontsize of the index generated.	2
It is possible input a Font name not present in the list of Font names.	2
... input a Font name not present in the list of Font names .	2
It indicate the Font Name and Fontsize of the index generated.	2
... the length of the context is rearranged if the Fontsize changes; the number of characters can be ...	2
... words with prefixes or suffixes, for example *link* can intercept link, delink, links etc.	2
... or suffixes, for example *link* can intercept link , delink, links etc.	2
... for example *link* can intercept link, delink , links etc.	2
... for example *link* can intercept link, delink, links etc.	2
If the check box Linkable index is selected a link pointing to the ...	2
If the check box Linkable index is selected a link pointing to the paragraph is created for every ...	2
... is located the word, the page number can be linkable .	3
... into trouble when trying to use the embedded links in converted PDFs.	4
If the links don't work, you need to change some options and ...	4
" The embedded links in the converted PDF file will now work.	4
... for a OpenOffice/LibreOffice writer and for MS Word documents.	1
1 and Ms Word 2010.	1
doc for MS Word), starting his Main subroutine appears a ...	1
MS Word The style names are on the user language.	2
MS Word The page number printed is obtained by asking ...	2
... ContextIndex in LibreOffice/OpenOffice; in MS Word ContextIndex is the name of a Bookmark.	3
It has developed for a OpenOffice/LibreOffice writer and for MS Word ...	1
OpenOffice/LibreOffice The style names that appear in ...	2
OpenOffice/LibreOffice The page number obtained is the ...	2
... required information, as in the example below (OpenOffice/LibreOffice) by the callNoDialog ...	3
OpenOffice/LibreOffice The page number obtained is the absolute page number, if ...	2
Office The page number obtained is the absolute page number , if it is different from the page number ...	2
... page number, if it is different from the page number printed, it can be adjusted by the dialog ...	2
MS Word The page number printed is obtained by asking the ...	2
... the page were is located the word, the page number can be linkable.	3
... box, uncheck the "ISO 19005-1 compliant (PDF/A)" check box.	4
" The embedded links in the converted PDF file will now work.	4
Wild-cards are translated into a regular expression for intercept only letters: ...	2
... for example the item fontname? generates the regular expression \bfontname[a-zA-Z]?b1, ...	2
If we need a complex search, the word must be a regular expression , in this case the word for sort ...	2
... the word for sort must be indicated after the regular expression , separated by a tabulation ...	2
09); for example for search wolves and wolf the regular expression can be: \b(wolf wolves)b.	2
Regular expression functions must be enabled i.	4
... extracts the text paragraphs for a selected style(s) , together the relative page and search in the ...	1
Paragraph style : the combo box is populate by the style of the ...	1
... style: the combo box is populate by the style of the Document; when a style is selected it is ...	1
... populate by the style of the Document; when a style is selected it is inserted in the text field at ...	1
Context Index: information for the style of the index.	1
You must choose the style or styles to index, the choices appears in the ...	2
OpenOffice/LibreOffice The style names that appear in the document are in ...	2
... example an Italian writer see this paragraph style as "Stile predefinito" where in the combo it ...	2
MS Word The style names are on the user language.	2
The words can have a wild-card (*, +, ?) both at start both at end; they ...	2
... (hexadecimal 09); for example for search wolves and wolf the regular expression can be: ...	2
... 09); for example for search wolves and wolf the regular expression can be: \b(wolf wolves)b.	2
... and wolf the regular expression can be: \b(wolf wolves)b.	2
... wolf the regular expression can be: \b(wolf wolves)b.	2