

This is a test for your Desktop Search software.

The aim is to get a database of extended search functionalities of various Desktop Search software's.

The last version of this database is on http://taye.fr/informatique/search/indexation_test3.xls

- Version 3.7 (with addition of docx, xlsx, pptx, xml files)
- Version 3.6 (with some changes in the "17" files : JPEG with metadata embedded)
- Version 3.5 (with an addition in OpenOffice files and Powerpoint .PPT files)
- Version 3.4 (with an addition in Supplements.doc)
- Version 3.3 (with very minor changes compared to 3.2)
- Version 3.2 (with some more files compared to the previous version which was not numbered).

Get the files used for this test on

http://taye.fr/informatique/search/indexation_test3.zip (160 kB).

Decompress this archive in a place where it is indexed by your Desktop Search software.

You will get the files:

• 01_indexation_test.doc	MS Word format
• 02_indexation_test.rtf	RTF format
• 03_indexation_test_Win.txt	text file with Windows format
• 04_indexation_test_DOS.txt	text file with DOS format
• 05_indexation_test.pdf	PDF, made with Adobe Acrobat 8 from 01_indexation_test.doc
• 06_indexation_test_ISO-8859-1.html	html with ISO-8859-1 character set
• 07_indexation_test_UTF-8.html	html with UTF-8 character set
• 08_indexation_test_ISO-8859-1.mht	mht archive (made with Internet Explorer 7 from 06_indexation_test_ISO-8859-1.html)
• 09_indexation_test_UTF-8.mht	mht archive (made with Internet Explorer 7 from 07_indexation_test_UTF-8.html)
• 10_indexation_test_text.eml	e-mail sent in text mode
• 11_indexation_test_html_OE.eml	e-mail sent in html mode with Outlook Express
• 12_indexation_test_html_TB.eml	e-mail sent in html mode with Thunderbird
• 13_indexation_test.xls	MS Excel file, text in cells. Contains also the numbers 67356976541 and 77889954321
• 14_indexation_test_comment.xls	MS Excel file, text in comments
• 15_indexation_test_graphic_zone.xls	MS Excel file, text in a graphic zone
• 16_indexation_test_text_zone.xls	MS Excel file, text in a text zone
• 17i_indexation_test_IPTC.jpg	JPEG file, text in IPTC metadata only
• 17ix_indexation_test_IPTC_XMP.jpg	JPEG file, text in IPTC and XMP metadata
• 17x_indexation_test_XMP.jpg	JPEG file, text in XMP metadata only
• 18_indexation_test_Win	text file with Windows format, without extension (like many Unix files)
• 19_indexation_test.odt	OpenOffice Write
• 20_indexation_test.ods	OpenOffice Calc
• 21_indexation_test_comment.ods	OpenOffice Calc, text in comments
• 22_indexation_test.odp	OpenOffice Impress
• 23_indexation_test.ppt	MS Powerpoint
• 24_indexation_test.xml	XML
• Supplements.doc	in order to test text with complicated (Polish) diacritics
• Z1-18_indexation_test.zip	all the files 01 to 18 prefixed by a "z" letter, in a zip archive

All these files contain the words:

unusualword
eczéma
Façade
Fenêtres
+33 4 42 24 65 56 tel. No
05 56 78 90 12 tel. No
05.11.22.34.56 tel. No

Once indexed, do the searches (the syntax depends on your search software; do not use the cotes; unusualword is used to restrict the search to the above files, since it is quite improbable that you use that word in your files):

- Search "unusualword ecz"
- Search "unusualword eczé"
- Search "unusualword ecze"
- Search "unusualword 67356976541"
- Search "unusualword 77889954321"
- Search "unusualword POZWOLIŁEM", and also "unusualword POZWOLILEM", and also "unusualword ŚRÓDŁĄCZE", and also "unusualword SRODLACZE"

And note the successful and unsuccessful searches that you get.

Then please report your results in a copy of the file

http://taye.fr/informatique/search/indexation_test3.xls

(if you don't have Excel, you can use OpenOffice)

Use comments in the cells in order to give more precise information, if necessary.

Send me the results by e-mail at gtsp01 - at - gmail.com, and I will incorporate them in the online version of

http://taye.fr/informatique/search/indexation_test3.xls

Tell me if you want that I put your name, your initials, or nothing in this report file.

Gérard Tayeb
Marseille, France

<http://taye.fr>
gtsp01 - at - gmail.com