

Free and/or open source OCR-related software tools

#####

[x] indicates that linux packages (rpm/deb) are available at the distribution level

[] indicates that no packages are available (most of the time zipped binaries are)

The following technologies have not been considered in this overview:

- + language related tools (NLP, lexicon building, etc.)
- + handwriting recognition systems
- + ocr tools for asian/arabic scripts

Needless to say that a quick search for “ocr” on github, sf, or googlecode will bring up countless more clones, forks, phds and other results.

Now here's the list - comments and additions welcome:

Image analysis & manipulation

[x] gimp, <http://www.gimp.org/>

[x] imagemagick, <http://www.imagemagick.org/>

[x] graphicsmagick, <http://www.graphicsmagick.org/>

[x] unpaper, <http://unpaper.berlios.de/>

[x] leptonica, <http://www.leptonica.com/>

[] tifftool, <http://sourceforge.net/projects/tifftool/>

[] agora, <http://www.rfai.li.univ-tours.fr/PagesPerso/jyramel/gb/work1.html>

Optical character recognition

[x] tesseract, <http://code.google.com/p/tesseract-ocr/>

[x] cuneiform, <http://en.openocr.org/>

[x] gocr, <http://jocr.sourceforge.net/>

[x] ocrad, <http://www.gnu.org/software/ocrad/>

[x] ocre, <http://lem.eui.upm.es/ocre.html>

[x] hocr, <http://hocr.berlios.de/>

[] simpleocr, <http://www.simpleocr.com/>

[] longan, <https://github.com/Zarkonnen/Longan>

[] asprise, <http://asprise.com/product/ocr/index.php?lang=java>

[] javaocr, <http://www.roncemer.com/software-development/java-ocr>

[] ocrchie, <http://www.cs.berkeley.edu/~fateman/kathey/ocrchie.html>

[] claraocr, <http://freecode.com/projects/claraocr>

[] kognition, <http://sourceforge.net/projects/kognition/>

[] xplab, <http://www.pattern-lab.de/>

[] eyeocr, <http://eyeocr.sourceforge.net/>

- [] plasmaocr, <http://developer.berlios.de/projects/plasmaocr/>
- [] ocr gem, <http://rubygems.org/gems/ocr>
- [] ocr for android, <http://code.google.com/p/wordsnap-ocr/>
- [] carletonocr, <http://code.google.com/p/carletonocr/>
- [] neuroocr, <http://www.codeproject.com/Articles/11285/Neural-Network-OCR>
- [] cs499ocr, <http://code.google.com/p/cs499ocr/>
- [] schnell-ocr, <https://github.com/jagd/schnell-ocr>

Frameworks

- [x] gamera, <http://gamera.informatik.hsnr.de/>
- [x] ocropus, <http://code.google.com/p/ocropus/>
- [x] ocrfeeder, <http://live.gnome.org/OCRFeeder>
- [] conjecture, <http://conjecture.sourceforge.net/>
- [] O2, <http://www.imglab.org/p/O2/>
- [] paradiit, <http://code.google.com/p/paradiit/>

Post-correction & transcription platforms

- [] fromthepage, <https://github.com/benwbrum/fromthepage/wiki>
- [] scripto, <http://scripto.org/>
- [] tb-transcription-desk, <http://code.google.com/p/tb-transcription-desk/>
- [] scribe, <https://github.com/zooniverse/Scribe>
- [] pybossa, <https://github.com/PyBossa/pybossa>
- [] textlab, <http://mel.hofstra.edu/textlab.html>
- [] t-pen, <http://digital-editor.blogspot.com/>
- [] islandora, https://github.com/Islandora/islandora_tei_editor
- [] alto-edit, <https://github.com/impactcentre/alto-editor>
- [] typewright, <http://www.18thconnect.org/typewright/documents>

Utilities

- [] ocropodium, <http://code.google.com/p/ocropodium/>
- [] isri ocr evaluation, <http://code.google.com/p/isri-ocr-evaluation-tools/>
- [] mile ocr evaluation, <http://code.google.com/p/ocr-performance-evaluator/>
- [] gedi ground truthing environment, <http://lampsrv02.umiacs.umd.edu/projdb/project.php?id=53>
- [] ground truth maker, <http://navidomass.univ-lr.fr/gtm.html>
- [] gttext, <http://code.google.com/p/gttext/>
- [] weocr, <http://weocr.ocrgid.org/>
- [] expervision, <http://www.expervision.com/ocr-software/webocr-onlineocr>
- [] hocr tools, <http://code.google.com/p/hocr-tools/>