

Using terminal and Python

1 Overview

1.1 Launching the command line

- `cmd` on Windows
- Terminal app on Mac OS
- Linux users know what to do :)

1.2 Commands

- `pwd` gives current directory (also written near the cursor)
- `ls` lists files in the current directory (`dir` on Windows)
- `cd` changes current directory
 - `cd` gets you to home directory
 - `cd dir_name` gets you to a directory `dir_name` in the current dir
 - `cd /dir/name/...` (`C:/dir/name` on Windows) gets you to a directory `dir_name` relative to the root
 - `cd ..` gets you one level up
- `wget url` downloads the file from `url` to the current directory

2 WikiExtractor

- `wgethttp://medialab.di.unipi.it/Project/SemaWiki/Tools/WikiExtractor.py`
- `https://dumps.wikimedia.org/backup-index.html` — dumps of all the wikimedia projects
- WikiExtractor saves them as text
- `cd` to its directory (optional)
- `python WikiExtractor.py -cb 250K -o extracted dewiki-latest-pages-articles.xml.bz2`
— extracts German Wikipedia to the directory `extracted`

3 BRAT

3.1 Overview

- `http://brat.nlplab.org/`
- Installation: `http://brat.nlplab.org/installation.html`
- Usage: `http://brat.nlplab.org/manual.html`

3.2 Usage

- `cd` to its directory
- `python standalone.py`
- Go to the browser