

Python programming — course introduction

Finn Årup Nielsen

Department of Informatics and Mathematical Modelling

Technical University of Denmark

Lundbeck Foundation Center for Integrated Molecular Brain Imaging

Neurobiology Research Unit,

Copenhagen University Hospital Rigshospitalet

August 31, 2009

Python programming

DTU course 02820 Python programming

Project course with a few introductory lectures, but mostly self-taught.

Deliverables: A report, a poster and an oral presentation at the poster about a Python program you write in a group.

Autumn 2009 first run for the course: The course is somewhat experimental.

Teachers: Bart Wilkowski, Marcin Szewczyk, Finn Årup Nielsen

Schedule for autumn 2009

31. August. Introduction to Python. (Finn)

7. September. Installation party. (Marcin)

14. September. “Python as Matlab”: NumPy, SciPy, MatPlot (Finn, Marcin)

21. September. Web and text processing, “natural language processing” (Finn, Marcin, Bart)

5. October. Mobile Python. Use on Nokia smartphones (Bart)

(19. October Machine learning with Python)

Other courses

Introductory programming and mathematical modelling (linear algebra, statistics, machine learning)

DTU 02827: Mobile Application Prototype Development. Also with Python.

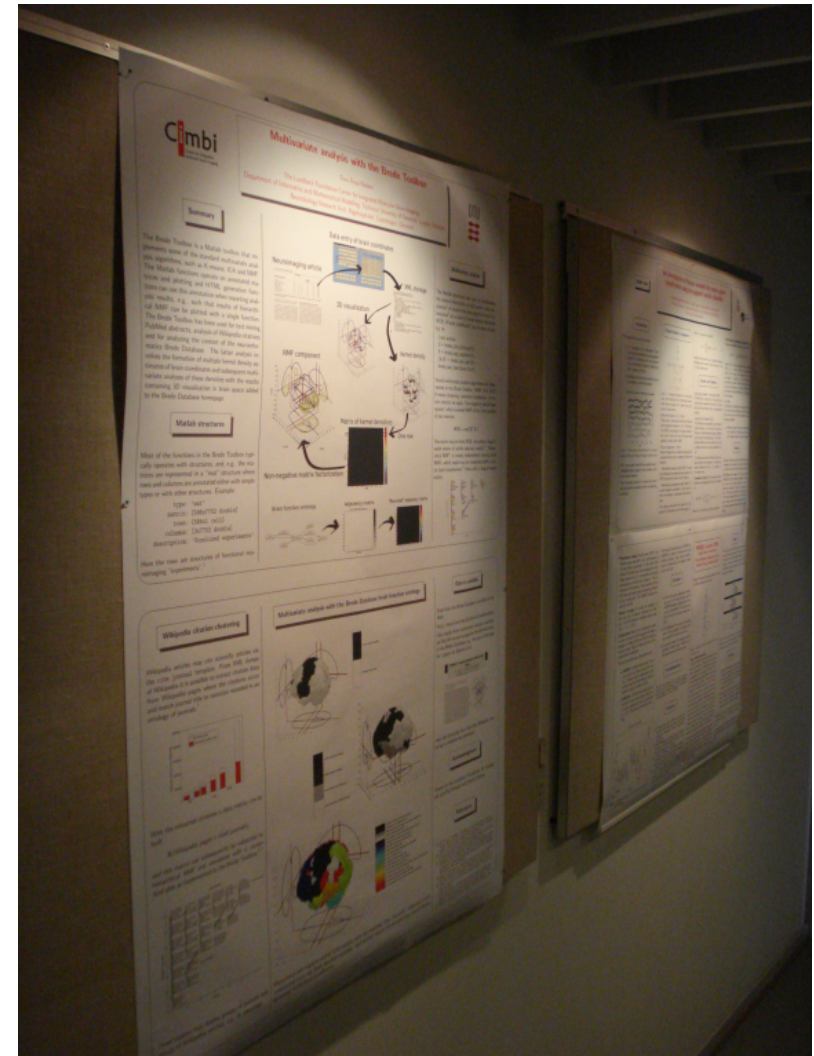
DTU 02457: Non-Linear Signal Processing. Machine learning

Poster

Construct a poster.

“Defend” the poster, i.e., give a relatively short oral presentation of the poster and answer questions.

Inspired from DTU course 02459 Machine Learning for Signal Processing



Why Python?

Interpreted, readable (usually clearer than Perl), interactive, many libraries, runs on many platforms, e.g., Nokia smartphones and Apache Web servers.

With Python one can construct numerical programs, though with a bit more boilerplate than Matlab.

Google and Yahoo! is (has been?) using it. 2.73% of Open Source code written in Python (Black Duck Software, 2009).

“Without [Python] a project the size of Star Wars: Episode II would have been very difficult to pull off.” — <http://python.org/about/quotes/>

XKCD 353: “I wrote 20 short programs in Python yesterday. It was wonderful. Perl I’m leaving you.”

Why Python? Interactive language!

Interactive session

```
$ python
```

```
Python 2.4.4 (#2, Oct 22 2008, 19:52:44)
```

```
[GCC 4.1.2 20061115 (prerelease) (Debian 4.1.1-21)] on linux2
```

```
Type "help", "copyright", "credits" or "license" for more information.
```

```
>>> 1+1
```

```
2
```

However, Matlab-like computation is not straightforward, e.g., what is the result of

```
>>> 1/2
```

Why Python? Interactive language!

Interactive session

```
$ python
```

```
Python 2.4.4 (#2, Oct 22 2008, 19:52:44)
```

```
[GCC 4.1.2 20061115 (prerelease) (Debian 4.1.1-21)] on linux2
```

```
Type "help", "copyright", "credits" or "license" for more information.
```

```
>>> 1+1
```

```
2
```

However, Matlab-like computation is not straightforward, e.g., what is the result of

```
>>> 1/2
```

```
0 # Integer division!
```


Example projects for inspiration

1. Characterize external links from DTU's Web-site.
2. Characterize internal link structure on DTU.
3. A search engine for DTU Web-pages.
4. Sentiment analysis of Tweets.
5. Sentiment analysis of blogs.
6. A wiki-based database for brain activations

7. A Web-service for visualization of brain activations.

8. Sound segmentation

More information

Google, Bing or Yahoo. <http://www.python.org/>

Google for error messages, “Python tutorial”

Books

Python Essential References (Beazley, 2000): Introduction and list of Python functions with small examples. Somewhat old

Python cookbook (Martelli et al., 2005): Short program examples for somewhat specific problems.

Practical Programming. An introduction to computer science using Python, (Campbell et al., 2009) Introductory programming.

Specialized books relevant for the course

Python scripting for computational science (Langtangen, 2005): Python book with many examples especially for numerical processing. Not fully up to date on numerical Python.

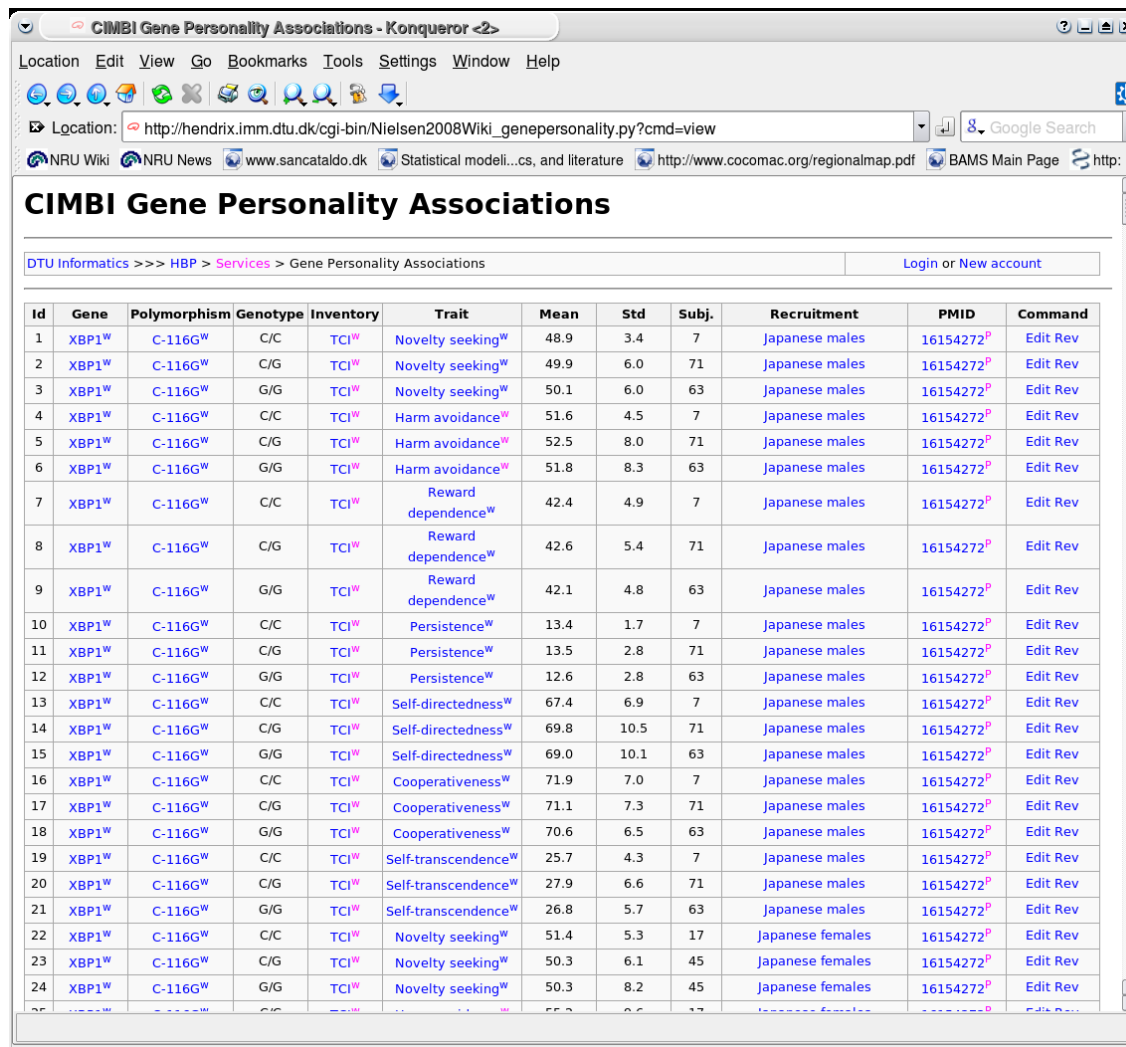
Programming collective intelligence (Segaran, 2007): Python and machine learning with data from the Web.

Natural language processing with Python (Bird et al., 2009): Text mining with Python.

Mobile Python (Scheible and Tuulos, 2007): Mobile Python.

Other O'Reilly titles: Python in a Nutshell, Python Pocket Reference, Learning Python, Programming Python?

Example: A fielded wiki . . .



| Id | Gene | Polymorphism | Genotype | Inventory | Trait | Mean | Std | Subj. | Recruitment | PMID | Command |
|----|-------------------|---------------------|----------|------------------|---------------------------------|------|------|-------|------------------|-----------------------|----------|
| 1 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Novelty seeking ^W | 48.9 | 3.4 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 2 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Novelty seeking ^W | 49.9 | 6.0 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 3 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Novelty seeking ^W | 50.1 | 6.0 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 4 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Harm avoidance ^W | 51.6 | 4.5 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 5 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Harm avoidance ^W | 52.5 | 8.0 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 6 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Harm avoidance ^W | 51.8 | 8.3 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 7 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Reward dependence ^W | 42.4 | 4.9 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 8 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Reward dependence ^W | 42.6 | 5.4 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 9 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Reward dependence ^W | 42.1 | 4.8 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 10 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Persistence ^W | 13.4 | 1.7 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 11 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Persistence ^W | 13.5 | 2.8 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 12 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Persistence ^W | 12.6 | 2.8 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 13 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Self-directedness ^W | 67.4 | 6.9 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 14 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Self-directedness ^W | 69.8 | 10.5 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 15 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Self-directedness ^W | 69.0 | 10.1 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 16 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Cooperativeness ^W | 71.9 | 7.0 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 17 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Cooperativeness ^W | 71.1 | 7.3 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 18 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Cooperativeness ^W | 70.6 | 6.5 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 19 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Self-transcendence ^W | 25.7 | 4.3 | 7 | Japanese males | 16154272 ^P | Edit Rev |
| 20 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Self-transcendence ^W | 27.9 | 6.6 | 71 | Japanese males | 16154272 ^P | Edit Rev |
| 21 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Self-transcendence ^W | 26.8 | 5.7 | 63 | Japanese males | 16154272 ^P | Edit Rev |
| 22 | XBP1 ^W | C-116G ^W | C/C | TCI ^W | Novelty seeking ^W | 51.4 | 5.3 | 17 | Japanese females | 16154272 ^P | Edit Rev |
| 23 | XBP1 ^W | C-116G ^W | C/G | TCI ^W | Novelty seeking ^W | 50.3 | 6.1 | 45 | Japanese females | 16154272 ^P | Edit Rev |
| 24 | XBP1 ^W | C-116G ^W | G/G | TCI ^W | Novelty seeking ^W | 50.3 | 8.2 | 45 | Japanese females | 16154272 ^P | Edit Rev |

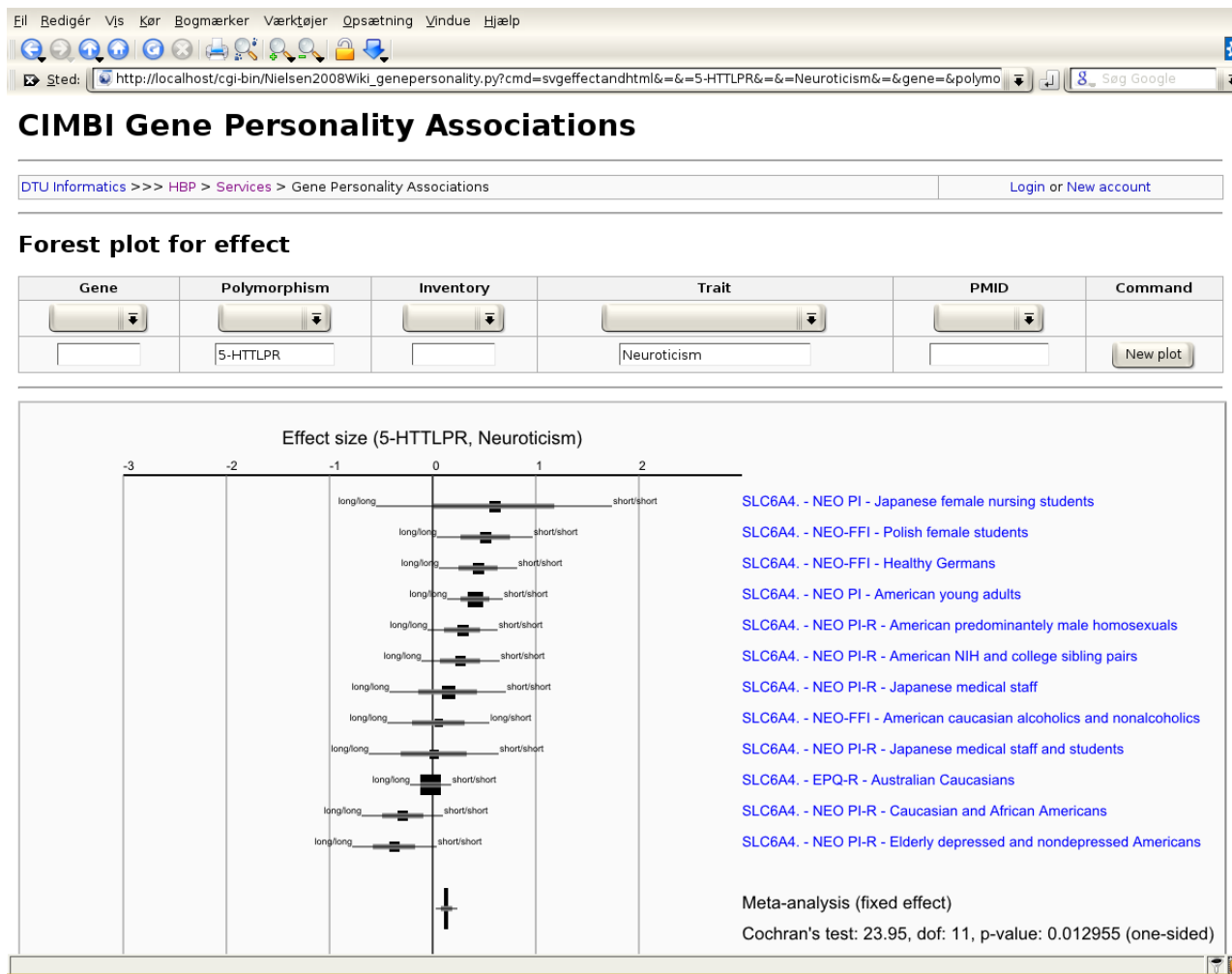
Web script in Python implementing a fielded wiki for personality genetics.

Persistence with a small SQLite database.

Some of the Python libraries used: cgi, Cookie, math, pysqlite2, scipy, sha.

One Python script with 2269 lines of code.

Example: . . . A fielded wiki



Computation of *effect sizes* (a statistical value) and comparison to statistical distributions.

Generation of interactive and hyperlinked plots in SVG (an XML format)

Structured information from Wikipedia

DTU Informatics WP Sniper - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://hendrix.imm.dtu.dk/services/wikipedia/snipper.html>

DTU Informatics WP Sniper

DTU Informatics >>> HBP > Services > Wikipedia > WP Sniper

| Name | Wikipedia | | | Region | Other resources | | | | | | |
|-----------------|-----------|--------|---------|----------|-----------------|---------|---------|---------|---------|--------|--------|
| | rsid | Gene | Chromo. | | Ensembl | dbSNP | HapMap | SNPedia | AlzGene | SzGene | PDGene |
| Factor V Leiden | rs6025 | F5 | 1 | | 6025 | 6025 | 6025 | 6025 | | | |
| Val158Met | rs4680 | COMT | 22 | Exon 3 | 4680 | 4680 | 4680 | 4680 | M O | M O | |
| T102C | rs6313 | HTR2A | 13 | Exon 1 | 6313 | 6313 | 6313 | 6313 | M O | M O | |
| -1438G/A | rs6311 | HTR2A | 13 | | 6311 | 6311 | 6311 | 6311 | | M O | |
| C-1019G | rs6295 | HTR1A | 5 | Promotor | 6295 | 6295 | 6295 | 6295 | | | |
| | rs7997012 | HTR2A | 13 | Intron 2 | 7997012 | 7997012 | 7997012 | 7997012 | | | |
| His452Tyr | rs6314 | HTR2A | 13 | Exon 3 | 6314 | 6314 | 6314 | 6314 | | M O | |
| | rs1954787 | GRIK4 | 11 | | 1954787 | 1954787 | 1954787 | 1954787 | | | |
| Val66Met | rs6265 | BDNF | 11 | | 6265 | 6265 | 6265 | 6265 | M O | M O | M O |
| G294A | rs6294 | HTR1A | 5 | | 6294 | 6294 | 6294 | 6294 | | | |
| C267T | rs1805054 | HTR6 | 1 | | 1805054 | 1805054 | 1805054 | 1805054 | M O | | |
| A779C | rs1799913 | TPH1 | 11 | Intron 7 | 1799913 | 1799913 | 1799913 | 1799913 | | M O | |
| Asp294His | rs1805009 | MC1R | 16 | | 1805009 | 1805009 | 1805009 | 1805009 | | | |
| C677T | rs1801133 | MTHFR | 1 | | 1801133 | 1801133 | 1801133 | 1801133 | M O | M O | |
| C-521T | rs1800955 | DRD4 | | Promotor | 1800955 | 1800955 | 1800955 | 1800955 | | M O | |
| | rs7341475 | RELN | 7 | Intron 4 | 7341475 | 7341475 | 7341475 | 7341475 | | M O | |
| A1287G | rs5569 | SLC6A2 | | Exon 9 | 5569 | 5569 | 5569 | 5569 | | | |

Background: An overview of single nucleotide polymorphism (genetic variations) in Wikipedia. The table is constructed from information in infoboxes in Wikipedia.

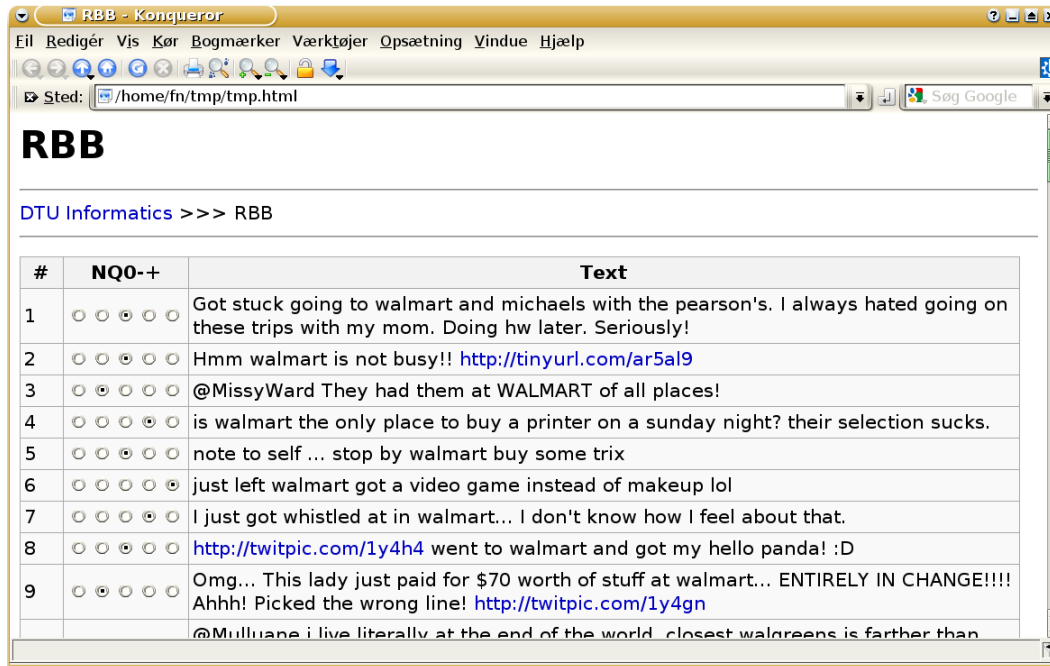
Finn Årup Nielsen, CIMBI, DTU Informatics.

Get Wikipedia pages that contain a specific template, download the page, extract information from the templates and render the result on an HTML page.

Python libraries: json, re, urllib2

Around 25 Python lines to get the data, and around 120 to render the result.

Web script for Twitter annotation



The screenshot shows a web browser window titled "RBB - Kongueror". The address bar shows the URL "/home/fn/tmp/tmp.html". The page content includes the text "DTU Informatics >>> RBB" and a table with 9 rows. Each row contains a number, a progress indicator (circles), and a text snippet from a tweet.

| # | NQ0+ | Text |
|---|-----------|--|
| 1 | ○ ○ ○ ○ ○ | Got stuck going to walmart and michaels with the pearson's. I always hated going on these trips with my mom. Doing hw later. Seriously! |
| 2 | ○ ○ ○ ○ ○ | Hmm walmart is not busy!! http://tinyurl.com/ar5a19 |
| 3 | ○ ○ ○ ○ ○ | @MissyWard They had them at WALMART of all places! |
| 4 | ○ ○ ○ ○ ○ | is walmart the only place to buy a printer on a sunday night? their selection sucks. |
| 5 | ○ ○ ○ ○ ○ | note to self ... stop by walmart buy some trix |
| 6 | ○ ○ ○ ○ ○ | just left walmart got a video game instead of makeup lol |
| 7 | ○ ○ ○ ○ ○ | I just got whistled at in walmart... I don't know how I feel about that. |
| 8 | ○ ○ ○ ○ ○ | http://twitpic.com/1y4h4 went to walmart and got my hello panda! :D |
| 9 | ○ ○ ○ ○ ○ | Omg... This lady just paid for \$70 worth of stuff at walmart... ENTIRELY IN CHANGE!!!! Ahhh! Picked the wrong line! http://twitpic.com/1y4gn |

CGI program that searches Twitter with a user-defined query, obtain tweets and present them in a Web form for manual annotation and stores the result in a SQL database.

Python libraries: codecs, json, re, cgi, urllib2, pysqlite2, xml.

500 Python lines.

Searching brain coordinates

Brede Database - Talairach coordinate search

Location search (one coordinate) e.g., 14 -9 -15

Experiment search (several coordinates)

Visualize in INC Talairach atlas

| # | Distance | x | y | z | WOEB | Description |
|----|----------|-----|----|----|------|--|
| 1 | 0.5 | -37 | 1 | 36 | 128 | Left precentral gyrus/frontal lobe — Visuo proprioceptive conflict (WOEXP: 393) |
| 2 | 6.0 | -39 | 1 | 31 | 83 | — Silent word generation (WOEXP: 262) |
| 3 | 6.7 | -43 | 5 | 36 | 91 | Left midfrontal — Alzheimer's disease versus healthy (WOEXP: 291) |
| 4 | 7.5 | -31 | -2 | 36 | 25 | Left precentral gyrus — Mental rotation of abstract figures versus object determination or dots counting (WOEXP: 84) |
| 5 | 8.1 | -39 | 0 | 29 | 137 | — Nitroglycerin-provoked cluster headache (WOEXP: 424) |
| 6 | 8.2 | -31 | -1 | 40 | 25 | Left precentral gyrus — Mental rotation of abstract figures versus rest (WOEXP: 79) |
| 7 | 9.1 | -41 | -3 | 44 | 179 | Left precentral gyrus — Unfair game offers (WOEXP: 562) |
| 8 | 9.2 | -44 | 2 | 44 | 134 | Left dorsolateral frontal — Focused episodic memory versus semantic memory (WOEXP: 415) |
| 9 | 10.2 | -39 | 0 | 47 | 178 | Left middle frontal gyrus — Semantic versus case (WOEXP: 550) |
| 10 | 10.4 | -47 | -1 | 37 | 88 | Left precentral gyrus — Activation in sadness film viewing versus neutral film viewing (WOEXP: 282) |
| 11 | 10.5 | -33 | 11 | 36 | 110 | Left lateral frontal — Spatial intelligence (WOEXP: 339) |
| 12 | 10.7 | -40 | 8 | 28 | 176 | — Verbal fluency (WOEXP: 539) |
| 13 | 10.7 | -32 | -3 | 44 | 3 | Left frontal eye field — Saccade versus central contrast (WOEXP: 7) |
| 14 | 10.7 | -45 | 1 | 29 | 81 | Left middle frontal gyrus/precentral sulcus — Shape matching (WOEXP: 253) |
| 15 | 10.9 | -45 | -5 | 38 | 88 | Left precentral gyrus — Activation in sadness film viewing versus neutral film viewing (WOEXP: 282) |
| 16 | 11.0 | -47 | 5 | 32 | 107 | Left precentral gyrus — Drinking at predictable intervals (WOEXP: 334) |
| 17 | 11.0 | -43 | 7 | 29 | 141 | Left mid-dorsal lateral prefrontal cortex — Memory retention (WOEXP: 432) |
| 18 | 11.1 | -47 | 7 | 36 | 91 | Left midfrontal — Decline in Alzheimer's disease (WOEXP: 292) |
| 19 | 11.5 | -39 | 6 | 47 | 89 | Left middle frontal gyrus — Serotonin synthesis capacity, proportional scaling. Men versus women (WOEXP: 285) |
| 20 | 11.9 | -48 | -1 | 41 | 46 | — Visual pursuit tracking (WOEXP: 149) |
| 21 | 12.0 | -39 | 5 | 25 | 10 | Left precentral gyrus — Monotonic distance effect in number decision (WOEXP: 27) |
| 22 | 12.1 | -33 | 3 | 26 | 100 | — Warmth on left hand versus heat (WOEXP: 313) |
| 23 | 12.3 | -49 | 4 | 40 | 90 | Middle frontal — Cued recall of familiar people. Group analysis (WOEXP: 290) |
| 24 | 12.4 | -36 | -8 | 44 | 135 | Left precentral sulcus — Imagined saccades (WOEXP: 419) |

CGI script to search the “Brede Database” for coordinates.

Reading of XML files, adding data to SQLite database, search database given a user query and generate an HTML Web page.

700 lines of Perl code, but could have been written in Python.

Brede Database.

<http://muzeeker.com/>

References

- Beazley, D. M. (2000). *Python Essential Reference*. The New Riders Professional Library. New Riders, Indianapolis.
- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly, Sebastopol, California. ISBN 9780596516499.
- Campbell, J., Gries, P., Montojo, J., and Wilson, G. (2009). *Practical Programming: An Introduction to Computer Science Using Python*. The Pragmatic Bookshelf, Raleigh.
- Langtangen, H. P. (2005). *Python Scripting for Computational Science*, volume 3 of *Texts in Computational Science and Engineering*. Springer, second edition. ISBN 3540294155.
- Martelli, A., Ravenscroft, A. M., and Ascher, D., editors (2005). *Python Cookbook*. O'Reilly, Sebastopol, California, 2nd edition.
- Scheible, J. and Tuulos, V. (2007). *Mobile Python: Rapid Prototyping of Applications on the Mobile Platform*. Wiley, 1st edition. ISBN 9780470515051.
- Segaran, T. (2007). *Programming Collective Intelligence*. O'Reilly, Sebastopol, California.