Some computational tools for the humanities

Benson Muite

benson.muite@ut.ee
http://kodu.ut.ee/~benson/

10 June 2016

Some Links

- https://github.com/Digitaalhumanitaaria/Digitaalhumanitaaria.github.io/blob/master/storage/seminarpreparation.md
- http://digitaalhumanitaaria.github.io/storage/seminarpreparation

Outline

- Git
- LaTeX
- Pandoc
- Parallel R
- Language Evolution Simulation

Git

Git

- Distributed version control
- Allow multiple people to work on a project together
- Keep a history of additions to the project
- Works best for text files where compression and tracking algorithms are effective, though can also use it to store other files

Git providers

- Github `https://github.com/`
- Bitbucket `https://bitbucket.org/`
- Gitlab `https://about.gitlab.com/`
- You?

TeX

Figure: Donald Knuth created TeX.

TeX

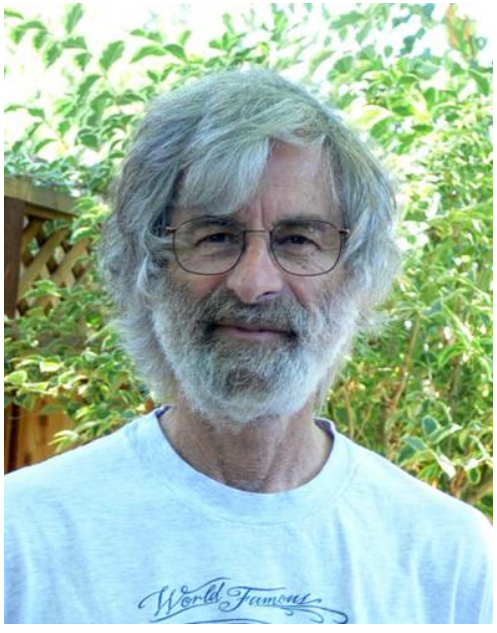

Figure: Leslie Lamport created LaTeX.

LaTeX

- Want better support in other languages
- Being open can incorporate this – but need to develop a community
- Since a markup language, can try to use extra contextual information when processing files
- One area to start with is bibliographies – biblatex package `https://github.com/Digitaalhumanitaaria/biblatex`

Pandoc

- Many different markup languages
- Want to be able to translate from one markup language to another
- In many cases, equivalent types of commands, eg. title, subtitle, bold formatting etc
- Pandoc is an attempt to do this
- Still needs some human cleanup

Parallel R

- Programming with Big Data in R (pbdR)
  `http://r-pbd.org/`
  i) open source project to allow you to do statistical analysis on large data sets
  ii) can be useful for processing large data corpora
  iii) PCA (principle component analysis) example
- `https://github.com/vijayachitrabio/biohpc`

Example from social sciences

- Abrams-Strogatz model for language death
-

$$\frac{dx}{dt} = yp_{yx} - xp_{xy}$$
$$p_{yx} = csx^a \quad p_{xy} = c(1-s)(1-x)^a$$
$$y = 1 - x$$

- x fraction of population speakers of language x, y fraction of population speakers of language x, p_{xy} probability of switching from language x to y, p_{yx} probability of switching from language y to x, c time scaling constant, a influence of population size on probability of switching language (claim $a \approx 1.3$ for many groups), s - relative status of language

Example from social sciences

- Patriarca-Heinsalu model
-

$$
\begin{aligned}
N1_t &= k(s1N1^aN2 - s2N2^aN1) + N1_{xx} + N1_{yy} \\
&\quad + \alpha N1 - \alpha N1(N1 + N2)/KK \\
N2_t &= k(s2N2^aN1 - s1N1^aN2) + N2_{xx} + N2_{yy} \\
&\quad + \alpha N2 - \alpha N2(N1 + N2)/KK
\end{aligned}
$$

- N_i population of speakers of language i, k rate coefficient, a model fitting parameter, s_i relative status, α Malthus growth rate, KK carrying capacity

Conclusions and further work

- Many of these tools work best on command line
- This takes time to learn how to use
- Advantage is automation for often repeated procedures which are slow to repeat yourself in a GUI
- Also allows customization and generation of workflows suited to you
- Get an account on an HPC cluster try some of these tools
- Provide some interesting data sets and we can help you use these tools

References

- Abrams, D.M. and Strogatz, S.H. “Modelling the dynamics of language death” Nature **424**, 900 (2013) `http://dx.doi.org/10.1088/1367-2630/13/3/033007`
- Appelbaum, J. “Donald Knuth” `https://en.wikipedia.org/wiki/Donald_Knuth#/media/File:KnuthAtOpenContentAlliance.jpg`
- Lamport, L. “Leslie Lamport” `https://en.wikipedia.org/wiki/Leslie_Lamport#/media/File:Leslie_Lamport.jpg`

References

- Linux Magazine “Linus Torvalds” https://en.wikipedia.org/wiki/Linus_Torvalds#/media/File:Linus_Torvalds.jpeg
- Patriarca, M. and Heinsalu, E. “Influence of geography on language competition” Physica A: Statistical Mechanics and its Applications **388**(2–3), 174-186, (2009) http://dx.doi.org/doi:10.1016/j.physa.2008.09.034 http://arxiv.org/abs/0807.3100v2

Acknowledgements

- Oleg Batrašev
- Els Heinsalu
- Artjom Lind
- Anders Martoja
-