Schedule for Current Topics in Digital Philology, SS 2014

http://www.informatik.uni-leipzig.de/~graebe/stdplan/block.html#DH.CTDP

Month	Semester Week	Class Dates	Topics	Readings/Resources	Exercises
April	Week 1	April 7	Grand Overview Install Python	http://www.pyzo.org/	
		April 8	Who are you? What are your interests? Github Let's Learn Python: IPython	http://chronicle.com/blogs/profhacker/tag/github101 http://try.github.io/levels/1/challenges/1 http://gitimmersion.com/ http://marklodato.github.io/visual-git-guide/index-en.html http://www.scholarslab.org/research-and-development/forking-fetching-pushing-pulling/ http://gitref.org/ http://excess.org/article/2008/07/ogre-git-tutorial/ http://git-scm.com/book https://education.github.com/	IPython Notebook chapters 1-5 here: https://github.com/DHLeipzig-CurrentTopics-SS2014/MK-Python-Course
	Week 2	April 14	Regex	regular-expressions.info docs.python.org/3/library/re.html docs.python.org/3/howto/regex.html	IPython Notebook chapter 6 here: https://github.com/DHLeipzig-CurrentTopics-SS2014/MK-Python-Course
		April 15	Web Scraping	lxml	IPython Notebook "A Short Intro to 'lxml'.ipynb here: https://github.com/DHLeipzig-CurrentTopics-SS2014/Course_Materials_SS2014/tree/master/Code
	Week 3	April 21	Ostermontag: No class		

		April 22	Web Scraping		
	Week 4	April 28	Scraping and modelling parallel Bibles	http://biblehub.com/interlinear/study/genesis/1.htm http://studybible.info/interlinear/ http://biblehub.com/interlinear/study/matthew/1.htm	Extract information as xml from these Bibles.
		April 29	Parallel Bibles discussion Group work		Align Greek-Hebrew in Old Testament using the above data.
May	Week 5	May 5	Matrices: constructing vectors and matrices	http://www.numpy.org/ http://pandas.pydata.org/	Create a large term-document matrix from several books and compare them using pairwise distance in sklearn
		May 6	Project work		
	Week 6	May 12	Semantic Information extraction - counting co-occurrents		Create a term-term matrix of how often each word co-occurs with each other word for several English novels
		May 13	Project work		
	Week 7	May 19	Presentations from Holy Cross Students		
		May 20	Project work		
	Week 8	May 26	Project Presentations		
		May 27	Project Presentations		
June	Week 9	June 2	Vectorizing Operations with Pandas		Perform moderately complex vectorized operations using Pandas (see IPython Notebook in Week 9 presentations folder)
		June 3	Project Work		
	Week 10	June 9	Pfingstmontag: No class		
		June 10	Semantic Drift – statistical significance		Create a log-likelihood ratio function and test it against a set of data trying to produce the given answers (see IPython Notebook in Week 10 Homework folder)
	Week 11	June 16	Semantic Drift – Producing semantic profiles with log		Produce a table to log-likelihood values from the co-occurrence tables that were produced in week 6.

			likelihood		
		June 17	Project Work		
	Week 12	June 23	Semantic Drift – Comparing semantic profiles with Cosine Similarity		Construct term-term matrices for each document showing the cosine similarity score of each word with each other word. Compare words in different corpora on the basis of log-likelihood and cosine similarity.
		June 24	Project Work		
July	Week 13	June 30	Bringing it all together: What questions can you ask and answer with co-occurrence profiles?		
		July 1	Visualizing your results - 2D and 3D plots	http://matplotlib.org/	
	Week 14	July 7	Presenting your results - Network visualization	http://networkx.github.io/	
		July 8	in Lausanne (DH 2014)		
	Week 15	July 14	Project presentations		
		July 15	Project presentations		