Tutorial/Workshop Session 2: Susan Schreibman (@schreib100): Text Analysis with online Tools @squintar

Tools we will test

1. http://dfr.jstor.org

2. google Ngram viewer : http://books.google.com/ngrams (You cannot add your data to it, it will work with the scanned books only, mainly for comparing how particular words' usage has chnged over the period of time in the scanned books.)

3. Text Arc : http://www.textarc.org/ Brad Paley Developed "for fun" by designer, independant from academia. Based on concordance (?was she talking about this soft?) Get an overview of an entire corpus of texts

4. Wordle http://www.wordle.net/

5. IBM's ManyEyes : http://www-958.ibm.com/software/analytics/manyeyes/ micro-reading of texts, charts out the words that are led by certain words

6. Voyant : http://disc.library.emory.edu/lincoln/voyant/ developed by literary scholars. shows you also where in the text the word occurs. http://voyant-tools.org

7. Data for research : http://about.jstor.org/service/data-for-research

Other tools and Resources

8. Rapid miner http://rapid-i.com/content/view/181/

Bookworm Culturomics

HathiTrust

First step : chose a text ;) We'll each play with one single text (personal, not too short) with several different tools

-- how's that going?

.doc prepared by copy pasting the text of a PDF of a major paper in biology/medical sciences (first paper of discovery of HIV before it was called HIV) (as in Grid?)

Trying to do it with Unicode text, Wordle is not able to render it properly, just some squares and all....

My work in progress can be followed here : https://docs.google.com/document/d/1oQVt1w0OEHdBAMkvhW0lj-7opDGCnFi5xVB88-lGRaA/edit?usp=sharing

Does these things work with languages written using scripts other than Roman ?

Using ManyEyes with french .txt is kind of frustating because it doesn't understand utf-8... With Voyant it is much better. With Voyant you can add your own limits, stop words list, which is usefull.