CSC 9010: Text Mining Applications
November 14 Lab

Dr. Paula Matuszek
Paula.Matuszek@villanova.edu
Paula.Matuszek@gmail.com
(610) 647-9789
Two Different Lab Exercises

- Privacy
- Summarization
Look at the privacy policies for at least two social media sites.

- How hard were they to find?
- How long were they?
- Who sees what you post by default?
- Who sees your location by default?
- What can you change? Is the change retroactive?
- Can you limit visibility of everything to a specified set of people? How much does that limit your use of the site?
Privacy 2

• Compare the sites you have looked at for
  • How broad the default access is
  • How hard it is to change the default access
  • How comfortable you are with this policy.

• We will review and discuss this in about 20 minutes.
Summarization

• Look at at least two online summarization sites and try them out.
  • What kind of summarization is it? Extractive or abstractive? Or something else?
  • What seemed to work well?
  • What seemed to work badly?
• We will discuss some of what you found.
Some systems to try

- http://summarizer.intellexer.com
- http://ultimate-research-assistant.com/
- http://iResearch-reporter.com (only one search without registering.)
- http://newsblaster.cs.columbia.edu/
- http://www.extractorlive.com/on_line_demo.html
- http://textcompactor.com
- http://www.splitbrain.org/services/ots