Linguistic Self-Expression on Dating Websites

(Typing for Swiping)

An undergraduate research project that seeks to identify idiosyncratic features of language on the dating website OkCupid. Over 1500 user profiles have been scraped and tagged with gender, sexual orientation, location, and age as reported by the user on the website. The full text of profiles and their tags were stored in a SQLite database. Ongoing analysis includes measuring linguistic features using Natural Language Processing (NLP) and using K means clustering on the corpus of profiles. The most concise and engaging results appear when the profiles are visualized using word clouds grouped by the tags.

But the stuff for Whalen tho - by Thursday, March 10th, 5p (500 – 700 words)

1. Background : WHY?
   1. We were interested in people being sexist and using gendered language
   2. We considered a lot of options, including scraping twitter feeds and youtube comments, but things like those don’t make people clearly define their genders for us and assuming is rude
   3. So dating sites it was
      1. Also gives us a huge pool of people and we don’t have to be watching too closely on their interactions because now we’ve changed it to be how they present themselves.
2. Methods : HOW?
   1. Set up fake accounts (bi lady and bi dude)
      1. ROBOT JOKES!
      2. My dumb face is hot (both will be expressed either in the poster or our slide show, not strictly for the here and now)
   2. Scrape scrape scrape a bunch of profiles
      1. They didn’t even catch us!!!!! :D
   3. Nltk and textblob those fuckers
   4. Add to database!
      1. We’ve got gender, age, location, orientation, looking for, and 9 profile fields
         1. We took the contents of the profiles and got some base stats on them, too
   5. Research things to look for
      1. Lots of old linguists did it where you wrote down everything you thought. This feels simultaneously more natural and more planned
   6. Try looking for those things
   7. Fail
   8. Spend a week downloading numpy
   9. Word clouds!
      1. They’re cool but what do they mean?
3. Results : WHAT DID YOU END UP WITH?
   1. A database with 1558 lonely hearts
   2. Lots of words
   3. Some new skills
   4. A bunch of word clouds
4. Discussion/Conclusions : WHAT DOES IT MEAN?
   1. Things are prettier as word clouds than as print outs of numbers and stuff
   2. So far we’re all the same when we’re trying to find our one true loves
5. Bibliography/works cited
   1. Lol who needs one of these