

Linguistic Self-Expression on Dating Websites

Elizabeth Dellea and Kelly Sadwin

Ithaca College

4th Year Undergraduates

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.