

Siu_Wei_CSE160_Final_Project_0

February 7, 2019

0.0.1 Chianson Siu, Wei Liao

###

Prevalent Interpersonal Issues in 2018

Introduction

As each year progresses and people change, so do the prevalent social issues among society. However, this can be difficult to study since gathering data may require lots of interviews, questionnaires, or web crawlers that are surfing various social networks. As such, one method of research could be looking at data from forums focused on social topics. At Stack Exchange, Inc., the company has 173 online communities discussing things from board game rules to coding questions. For our topic of study, which focuses on prevalent social issues, their [Interpersonal Skills](#) community serves as a good source of data. The members of this community post questions daily, asking about a variety of social topics such as improving co-worker relations or helping a friend with depression. Each question also has different aspects: a score which designates a well-thought out question, answers, comments, date asked, etc. Using the data from the Interpersonal Skills community archive, our group hopes to answer the following research questions.

Research Questions

1. What types of problems do people have the most? (occurrence data, not score)

To answer this research question, we could count the number of problems in each type of category and list out the results. We would like to analyze this question because the answer can help us understand what common problems modern humans face. The research question may be of interest to international mental health organizations, where the researchers in the organizations may need to understand contemporary problems to provide meaningful and effective solutions.

2. What types of problems are unsolvable? i.e. do not have acceptable answers.

To answer this research question, we could look through the data to find questions that do not have an accepted answer and find the categories that these questions belong to, as well as prominent themes in the question itself. We would like to analyze this question because the answer may give us insight into the psychology of modern humans. The research question may be of interest to social science researchers, who may want to study contemporary human problems.

3. Can we detect users that are consistently posting spam?

To answer this research question, we could search through the content of all posts and see whether there are multiple duplicated or unmeaningful (i.e. extremely short post that consist of

one word) posts by the same user. We would like to analyze this question because users that post spam are disruptive to other users and it would be helpful to detect them and potentially block them. The research question may be of interest to IT personnels managing the platform, who may want to ban the detected spam users.

4. Can we scan and flag users that consistently post racist/sexist/harmful messages?

To answer this research question, we could search for racist/sexist/harmful keywords in posts and find the corresponding users. We would like to analyze this question because we would like to make the forum a safer and healthier place for all users. The research question may be of interest to IT personnels managing the platform, who may want to improve the environment of the platform.

5. What is the most viewed or highest scored questions?

To answer this research question, we could count the top 10 highest viewed and scored questions and list the questions out. We would like to analyze this question because we would like to know what interest modern humans. The research question may be of interest to designers like HCDE (Human Centered Design & Engineering) researchers or marketing personel, who may want to know what modern humans want and need.

6. What type of user has high reputation? (demographics, race, gender, name, age).

To answer this reserch question, we could list out the top 100 or so users that have the highest reputations, then look into their user information to see whether patterns emerge. We would like to analyze this question because we would like to see whether implicit cues of race, gender, age etc. have an impact on how a user is evaluated. The research question may be of interest to psychologist who conducts implicit association research.

Dataset

Our group will be using the Interpersonal Skills data dump from the Stack Exchange, Inc. archive. It can be downloaded here: <https://archive.org/download/stackexchange>.

- interpersonal.meta.stackexchange.com.7z – 20.9 MB
- interpersonal.stackexchange.com.7z – 138 MB

Each site on Stack Exachange comes with two data files; the site's data, and the sites metadata. Both of which are stored as XML files. Each zip contains 8 XML files: badges, comments, post history, post links, posts, tags, users, and votes.