2013 IEEE International Conference on Software Maintenance

On the Personality Traits of StackOverflow Users

Blerina Bazelli, Abram Hindle, Eleni Stroulia
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
{bazelli,hindle1,stroulia}@ualberta.ca

***Abstract*—In the last decade, developers have been increasingly sharing their questions with each other through Question and Answer (Q&A) websites. As a result, these websites have become valuable knowledge repositories, covering a wealth of topics related to particular programming languages. This knowledge is even more useful as the developer community evaluates both questions and answers through a voting mechanism. As votes accumulate, the developer community recognizes reputed members and further trusts their answers. In this paper, we analyze the community's questions and answers to determine the developers' personality traits, using the Linguistic Inquiry and Word Count (LIWC). We explore the personality traits of StackOverflow authors by categorizing them into different categories based on their reputation. Through textual analysis of StackOverflow posts, we found that the top reputed authors are more extroverted compared to medium and low reputed users. Moreover, authors of up-voted posts express significantly less negative emotions than authors of down-voted posts.**

I. Introduction

Question and Answer (Q&A) websites have gained significant ground as the preferred forums for developer interactions, in part due to support they offer for users to up-vote/down-vote questions and answers, accept answers as correct and edit the posts of others [1], [2]. StackOverflow.com[1] (StackOverflow) is one of the most popular Q&A websites focused mainly on questions related to programming languages.

Since StackOverflow posts are written in natural language, comprehensibility metrics and textual context analysis tools can potentially provide us with valuable information on what makes a posting perceived as trustworthy by the community. Thus, to take into account the semantics of the text, we have also applied a text analysis tool, the *Linguistic Inquiry and Word Count* (LIWC), which uses an embedded contextual dictionary [3], [4]. The LIWC tool has been used to identify the personality traits of Facebook users [5] as well as the anti-social personality types [6].

This study is a partial replication of Rigby and Hassan's work [7] (Questions 1 and 2). They analyzed the Apache web-server developers' personalities using their development mailing-list. They compared the personalities of the top 4 developers with two categories of developers having sent more than 30 messages and less than 30 messages respectively. The extra features of Q&A websites (compared to mailing-lists) allow us to also examine the personality types of the authors from different perspectives (Questions 3 and 4). The main goal of our study is to analyze and compare the authors' personality types, based on the five most common personality traits [3]. Furthermore, we compare our results with Rigby and Hassan's work to determine differences and commonalities of personality traits among mailing-list developers and StackOverflow authors.

[1] http://stackoverflow.com/

Our research questions are:

1) Do the top reputed authors' personality types differ?
2) Do author personalities vary by reputation?
3) Do authors of posts belonging to the same topics share similar personality types?
4) Do personality traits have an impact on the process of up-voting/down-voting questions and answers?

II. Related Work

There is much research based on StackOverflow's open data-set [8]. Mamykina et al. [1] analyzed StackOverflow data in order to find the reason why StackOverflow has grown so rapidly since its inception. They found that most of the questions asked were quickly answered. Via interviews they found that factors such as reputation scores and badges motivated participation. Treude et al. [2] categorized the questions being asked on StackOverflow. They found that questions belonging into the categories of "review", "conceptual" and "how-to" are most likely to be answered. Barua et al. [9] analyzed the LDA topics of StackOverflow users' posts. They discovered topic trends of increasing popularity (e.g., Android, iPhone development) and decreasing popularity (e.g., Perl, Blackberry development).

Pennebaker et al. [3], [4] tried to associate words with one's personality traits based on a study performed on 1,203 introductory psychology students. They developed the LIWC tool, which includes a dictionary and 72 language dimensions. Each dimension has a value which is calculated based on the frequency of words related to this particular dimension. However, only some of these are related to the "Big Five Personality Traits" (Neuroticism, Extroversion, Openness, Agreeableness and Conscientiousness). LIWC was used in a study by Summer et al. [6] to analyze and predict anti-social traits ("Dark Triad Personality Traits") of Twitter users. Kramer et al. [5] used LIWC to study emotional expressions on Facebook and found evidence of emotional contagion (spreading of similar sentiment) between users.

This study aims to replicate the work done by Rigby and Hassan [7] on the Open Source Software (OSS) developers, by applying it to StackOverflow developers. Rigby and Hassan gathered data from the Apache server mailing list and used LIWC to determine the personality traits of OSS developers and whether their personality changes as they become more active. Moreover, they aimed to discover what personality traits are associated with the developers' emotional state before and after an Apache version is being released. According to their results, 2 out of the top 4 developers had similar personality traits and differ from the general population. Moreover, a decrease in the developers' positive emotions correlated with their imminent departure from the project. Finally, the textual analysis of the developers' e-mails revealed that before an Apache version was released, their e-mails were composed of words that expressed mostly optimistic feelings.

1063-6773/13 $26.00 © 2013 IEEE
DOI 10.1109/ICSM.2013.72

III. Data Collection and Text Analysis

We have analyzed the data of the six XML files provided by the 2013 MSR challenge [8], [10]. They include all the questions and answers posted on StackOverflow from August 2008 to August 2012, along with their authors' details.

The way a person writes and the different kind of words they select can reveal, to some extent, their personality types [3]. We use a text-analysis tool, LIWC [3], to analyze the posts and consequently to define the personality traits of the authors. LIWC includes a dictionary of 2700 words and word stems. The dictionary is divided into several categories (such as social processes, affective processes, cognitive mechanisms etc.) and sub-categories (such as insight, causation, discrepancy and tentativeness). At this point, it should be noted that when we refer to an author as neurotic, extroverted, open, agreeable or conscientious, we refer to their LIWC scores corresponding to their text corporae and not the authors themselves.

According to Pennebaker et al. [3], [4], some of the LIWC measures are correlated with the "Big Five Personality Traits":
a) "Neuroticism": is associated with negative emotions such as anxiety, anger or envy and therefore is correlated to the presence of negative emotional words.
b) "Extroversion": expresses an emotional state where the person feels the need to be more sociable and interactive with others. Therefore, "Extroversion" depends on the presence of social and positive LIWC measures as well as the absence of tentative and negative emotional measures.
c) "Openness": characterizes people who are open to new ideas and is positively correlated with tentativeness and negatively with the causation LIWC measure.
d) "Agreeableness": describes people who tend to agree with others and it was found that the dimension of articles is the most significant factor that determines this personality trait.
e) "Conscientiousness": is negatively correlated with negations and negative emotional LIWC measures.

Pennebaker et al. [4] found that some LIWC factors are correlated with each one of the personality traits. The correlation between a language dimension and a personality trait may be positive or negative. For example, the personality trait of extroversion is linearly modeled as: $Extroversion = -Tentativity - Negations + Social + PositiveEmotion$. The full models can be found in Pennebaker's et al. work [4].

IV. Methodology

In order to answer our research questions, we have divided the authors into several groups according to their reputation (top 10, high, medium and low). We also have categorized the authors based on their posts' content (defined by the keywords with wich they have been tagged) and also by the posts' votes.

The default reputation when a user registers on StackOverflow is 1. The data revealed that nearly 35% (457 627) of the users have neither asked any question nor written any response to existing questions. Since we lack information about these users, we do not include them in our analysis. For each StackOverflow question we applied the LIWC on each post separately. More specifically, our analysis method involves three steps:

First, as the posts within the XML file are in HTML format, we discard the HTML tags from the posts before applying LIWC. We also discard code snippets (text between `<code>` and `</code>`).

Second, we apply the LIWC tool on the text from the previous step. Based on the LIWC language dimensions, we compute the values of the "Big Five Personality Traits".

Third, we compare the distributions of the personality traits' values by using the ANOVA and Tukey's HSD test.

V. Results

A. Do the top reputed authors' personality types differ?

We selected the top 10 reputed users in order to see if they share similar personality traits. The reputation of the top reputed users ranges from 214 774 to 465 166. Next, we select all the posts belonging to these users (both questions and answers). By applying the LIWC tool to each post separately, we are able to calculate the values of the "Big Five Personality Traits". In order to compare the means of the distributions, we apply a one-way ANOVA test. The two hypotheses are:

- H_0: The means among top authors are equal
- H_1: The means among top authors are not equal

According to the ANOVA test, we reject the Null Hypothesis ($Pr(> F) < 2e^{-16}$) for all the personality traits tested. To further analyze these results we use the Tukey's test (a post-hoc test that compares all the possible combinations of the means). See Figure 1a.

Based on the TukeyHSD test with respect to neuroticism, the range of the 95% confidence intervals of neuroticism for all authors was from -0.67 to 1.02 (0 being none). This indicates there is a mild difference among authors in terms of neuroticism. The results of Tukey's test for the rest of the personality traits reveal that there are several combinations of authors who indeed share similar personality traits in terms of extroversion, openness, agreeableness and conscientiousness.

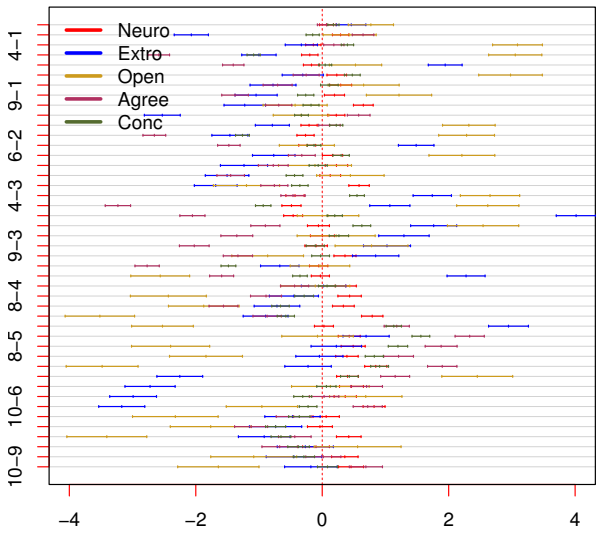
B. How do the personalities vary by reputation?

Are there differences in the personality traits of top, medium and low reputed users? We decided to investigate the relationship between personalities and reputation. We found out that the distribution of the authors' reputation was following a skewed distribution, similar to many countries distribution of wealth. There are few authors (the top reputed authors) who have very high reputation. On the other hand, the vast majority of StackOverflow users have a low reputation score. Therefore, we distinguish the authors into three main categories as follows: 1% top reputed authors, 10% medium reputed authors and the remaining 89% low reputed authors. Our hypotheses for the ANOVA test are:

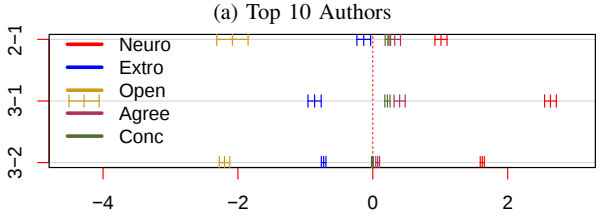
- H_0: The means of top-medium-low reputed authors are equal
- H_1: The means of top-medium-low reputed authors are not equal

The ANOVA test exposed that there are statistically significant differences among top, medium and low reputed users ($Pr(> F) < 2e^{-16}$). See Figure 1b. A further analysis with Tukey's HSD test exposed that there are differences among all groups as we expected. More specifically, the highest difference is between low and top reputed users. The score of neuroticism of low reputed users is much higher than top reputed users. This may suggest that authors who express more neuroticism through their text corpus are not being "awarded" by other users in terms of reputation points. Furthermore, we observe that there is a difference between medium and low reputed users. Again, the more reputed an author, the lower their measured neuroticism. Finally, despite the fact that there is a statistically significant difference between top and medium reputed users (top users less neurotic than medium reputed users), the mean is not much higher.

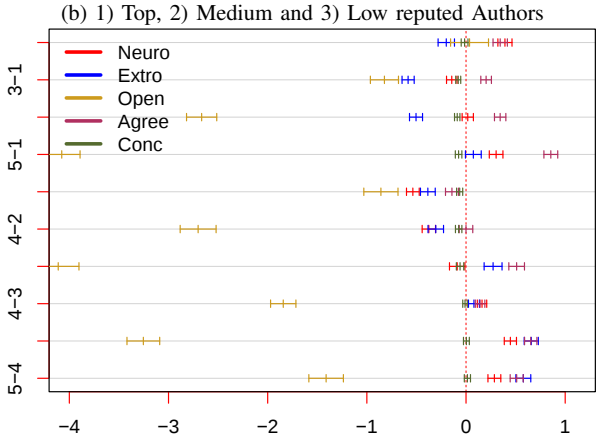
In terms of extroversion the highest discrepancy occurs between top and low reputed users. Furthermore, top reputed users exhibit more extroversion compared to medium reputed users who exhibit more extroversion compared to the less reputed ones. These results and the neuroticism results, hint that highly reputed authors exhibit more extroversion and lower neuroticism.



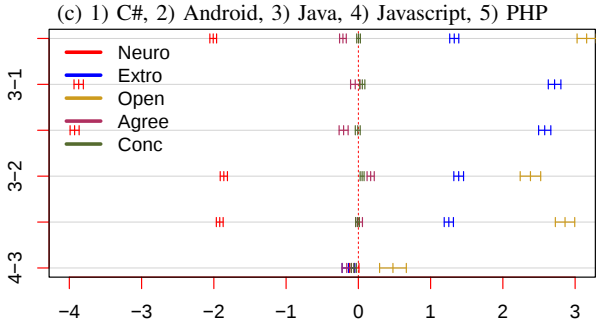
(a) Top 10 Authors



(b) 1) Top, 2) Medium and 3) Low reputed Authors



(c) 1) C#, 2) Android, 3) Java, 4) Javascript, 5) PHP



(d) 1) Down-voted questions, 2) Up-voted questions, 3) Down-voted answers and 4) Up-voted answers

Fig. 1: 95% Confidence Intervals of difference of means between Top 10 Authors, Reputed Authors, Tags, Down/Up Questions/Answers. 2-1 means the difference of means between 2) and 1) (e.g. C# and Java or Author 1 and 2).

As mentioned above the personality trait of *openness* characterizes one who is open to new ideas and/or shares their ideas with others. Prior to this study we theorized that top reputed authors could exhibit more openness as they were the most experienced users on StackOverflow. Our results show that top reputed authors are more open compared to both, medium and low reputed users with the highest difference among top and low reputed authors.

All the three categories also differ significantly in terms of agreeableness. Top reputed authors are more agreeable compared to the other two categories with the less agreeable authors to be the low reputed ones.

Finally, medium and low reputed users tend to have similar degree of conscientiousness according to Tukey's HSD test. Nonetheless, there is a difference among top and medium reputed authors with the top reputed authors appearing more conscientious whereas medium reputed authors appeared more conscientious than low reputed users.

C. Do authors of posts belonging to similar topics share similar personality types?

A tag is a label that describes the content of the question being asked and thus, helps categorize questions by topics such as "Web Development", "JavaScript", "Authentication" etc. When a user asks a question, StackOverflow forces users to add at least 1 tag. We examined the case of tags to be misleading; as StackOverflow allows most of the users to edit others' posts, in such cases "moderators" edit the vast majority of posts by replacing misleading tags with appropriate tags. Moreover, new tags are automatically removed if they are not used by at least one other question in a 6-month period. We found that the first 5 most common tags in descending order are: C#, Java, PHP, JavaScript and Android.

We found all the posts tagged as each one of the above tags for each author. Then, we computed the values of the "Big Five Personality Traits" for each post. As long as an author has multiple posts that correspond to one of the five most popular tags, we compute the mean of these values per author; therefore, we end up with a distribution composed of the mean values for the personality traits for each user. For instance, for an author who has written 5 posts tagged as "Java", we compute the mean of the personality traits' values and associate the author with the "Java" category.

Before applying the ANOVA test, we define the following hypotheses:

- H_0: The means of the authors' personality traits tagged for each of the most popular 5 tags are equal
- H_1: The means of the authors' personality traits tagged for each of the most popular 5 tags are not equal

The ANOVA test applied on these distributions showed that we should reject the null hypothesis ($Pr(> F) < 2e^{-16}$). See Figure 1c. With that being said, there are statistically significant differences among the authors belonging in different tag categories. The Tukey's HSD test exposed that authors with posts tagged as "Android" tend to be slightly more neurotic compared to authors who had posted posts related to "Java", "JavaScript" and "PHP". Also, according to Tukey's HSD test, authors with posts tagged as "C#" are less neurotic than authors having posted posts related to "Android" and "PHP". Finally, there is no statistically significant difference among owners of posts associated with "C#" and "JavaScript".

Furthermore, we explore the rest of the personality traits in order to see if the tags follow a pattern similar to the authors belonging into the three categories according to their reputation (top, medium and low). As we can see in Figure

3b, the personality traits of extroversion varies among different tags. More specifically, authors of posts tagged as "C#" exhibit more extroversion than those with posts related to "Android", "Java" and "JavaScript". Authors related to "C#" programming language, follow the same pattern with authors belonging to the top reputed users (Less Neurotic-More extroverted). On the other hand, authors of "PHP" related posts are more extroverted when compared with "Java", "JavaScript" and "Android" related posts.

Although there is no statistically significant difference among authors who have been written posts related to "C#" and "Android", the former are more open to new ideas than authors of posts related to "Java", "JavaScript" and "PHP" while "PHP" related authors are less open compared to "JavaScript", "Java" and "Android" related ones.

As for the personality trait of Agreeableness, the less agreeable authors are those related to "C#" and "Java" posts. The most interesting personality trait is "Conscientiousness", as authors belonging in several tag categories appear to be equally conscientious ("C#"-"Android" and "Java"-"JavaScript"-"PHP"). Although, Tukey's test showed that there is statistically significant difference among the other combinations, the difference between the means is very small.

D. How much do personality traits differ between up-voted and down-voted authors?

Since the personality types among authors of questions and answers may not be similar, we divided the posts into questions and answers based on their votes. We came up with 4 distinct categories: *Down-voted questions*, *Up-voted questions*, *Down-voted answers* and *Up-voted answers*.

We noticed that 46% and 37% of all answers and questions respectively did not have any votes. After excluding the "vote-free" posts, questions and answers with at least 1 positive vote are called Up-voted, while posts with less than 0 votes are called Down-voted. We compare the authors belonging in these categories as follows: Down-voted questions vs. Up-voted questions and Down-voted answers vs. Up-voted answers as the nature of a question being asked is different from an answer. It should be noted that some authors may be double counted as they may have both Up-voted posts and Down-voted posts.

- H_0: The means of the authors' personality traits for each category to be compared are equal
- H_1: The means of the authors' personality traits for each category to be compared are not equal

The ANOVA test results in $Pr(> F) < 5.3e^{-6}$ for all the personality traits; therefore, we reject the Null Hypothesis.

As mentioned above we focus on comparing authors of questions and answers separately. Based on Figure 1d, authors belonging to the category of down-voted questions expressed more neuroticism compared to the up-voted ones. We cannot assume the same result for the answers, as according to Tukey's HSD test there is no significant difference among up-voted and down-voted answers.

Despite the fact that we expected authors of down-voted questions to be less extroverted, these questions exhibit more extroversion than up-voted ones. Authors of up-voted answers exhibited more extroversion than the down-voted answers.

We theorized that authors of up-voted posts might be more open than authors of down-voted posts. Authors of up-voted questions and answers had higher openness means, compared to authors of down-voted questions and answers respectively. However, they exhibit less agreeableness and conscientiousness.

VI. Conclusions

In this paper we analyzed the personality properties of top, medium and low reputed authors, authors of most popular tags and most Up-voted and Down-voted posts on StackOverflow by replicating Rigby and Hassan's work [7], who analyzed the personality traits of the top contributors of the Apache web-server project using the development mailing-list. According to our results, some of the top reputed authors share similar personality traits which matches Rigby and Hassan's pattern (2 out of 4 top developers within Apache shared similar personality traits). We also found out that the top, medium and low reputed authors differ in Neuroticism, Extroversion, Openness, Agreeableness and Conscientiousness. Top reputed authors are less neurotic, more extroverted and open compared to medium and low reputed users who may just have entered the StackOverflow community. This difference may imply that posters who exhibit less neuroticism and more extroversion gain more popularity and reputation. Rigby and Hassan's [7] conclusion differ; they report similar Extroversion and Openness measures between top authors and the general population.

Furthermore, Tukey's HSD test shows that authors related to posts tagged as "Android" exhibit more neuroticism than authors with posts tagged as "Java", "JavaScript" and "PHP". Authors related to "C#" follow the same pattern as the top reputed users: less neuroticism and more extroversion. Yet authors of "PHP" related posts exhibited more extroversion than authors of "Java", "JavaScript" and "Android" posts.

These results could serve as a measure that managers can use to hire programmers who can ask and answer questions effectively. In the future, we will focus on analyzing the variation of StackOverflow authors' personalities over time. Furthermore, we would like to investigate the different kind of the personalities of question askers and question respondents and if they relate to teamwork abilities.

References

[1] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, "Design lessons from the fastest q&a site in the west," in *Proceedings of the 2011 annual conference on Human factors in computing systems*. ACM, 2011, pp. 2857–2866.

[2] C. Treude, O. Barzilay, and M.-A. Storey, "How do programmers ask and answer questions on the web?: Nier track," in *Software Engineering (ICSE), 2011 33rd International Conference on*. IEEE, 2011.

[3] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, 2001.

[4] J. W. Pennebaker, L. A. King *et al.*, "Linguistic styles: Language use as an individual difference," *Journal of personality and social psychology*, vol. 77, no. 6, pp. 1296–1312, 1999.

[5] A. D. Kramer, "The spread of emotion via facebook," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. ACM, 2012, pp. 767–770.

[6] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2. IEEE, 2012, pp. 386–393.

[7] P. C. Rigby and A. E. Hassan, "What can oss mailing lists tell us? a preliminary psychometric text analysis of the apache developer mailing list," in *Proceedings of the Fourth International Workshop on Mining Software Repositories*. IEEE Computer Society, 2007, p. 23.

[8] J. Atwood, "Stack overflow creative commons data dump," June 2009, http://blog.stackoverflow.com/2009/06/stack-overflow-creative-commons-data-dump/.

[9] A. Barua, S. W. Thomas, and A. E. Hassan, "What are developers talking about? an analysis of topics and trends in stack overflow," *Empirical Software Engineering*, pp. 1–36, 2012.

[10] A. Bacchelli, "Mining challenge 2013: Stack overflow," in *The 10th Working Conference on Mining Software Repositories*, 2013.