

Analysis of Stack Overflow

Taylor B. Morris

January 10, 2017

Abstract

1 Introduction

2 Background

The average programmer today is familiar with Stack Overflow, a forum designed for asking and answering programming related questions. Stack Overflow is open to anyone whether they be professional or a hobbyist or just starting out. Due to its open nature, posts on Stack Overflow cover a large variety of questions over a number of languages. A quick internet search for a programming issue is likely to result in a Stack Overflow post as a top result.

Thanks to the ubiquitous nature of Stack Overflow, analysis on the site offers a way to obtain a large amount of information on programmers' practices and problems. One prior analysis of Stack Overflow utilized an alternate method of parsing and tagging of data, using an adverb-verb formulation instead of the traditional noun and verb word bagging. This method of analysis was performed on the topics of Stack Overflow posts and allowed the researchers to evaluate the types of questions seen corresponding to different programming languages and different Stack Overflow tags. This analysis shed light on what types of questions are more common between different languages; additionally, this study helped to reveal some of the most common uses of each language [1].

3 Methods

3.1 Question Data

The data to be used is a dataset of Stack Overflow questions and some of their statistics found on Kaggle.com. Each row of the data represents a single question and includes the title, body, score (a metric used by Stack Overflow to determine the quality of a question), closed date (if the question was closed), and owner ID.

Each Stack Overflow question typically contains a single sentence title. Often, a simplified version of the question appears in the title; however, the primary text and background for the coding question appears in the body. The body can contain Markdown or HTML code, including a special tag to allow for code: `<code>`

4 Results

5 Conclusions

References

- [1] Miltiadis Allamanis and Charles Sutton. Why, when, and what: analyzing stack overflow questions by topic, type, and code. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, pages 53–56. IEEE Press, 2013.