Title

Thomas Maillart
School of Information
University of California, Berkeley, 102 South Hall
Berkeley, CA 94720
thomas.maillart@ischool.berkeley.edu

ABSTRACT
Abstract

Categories and Subject Descriptors
H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms
Theory

Keywords
ACM proceedings, LaTeX, text tagging

1. INTRODUCTION
2. METHOD
1 .The current implementation involved sampling analyzing XXXX questions from 83 StackExchange Websites (`http://stackexchange.com/sites#`) to characterize the statistical properties of answer timelines, as a measure of collective problem solving efficiency.

2. StackExchange, and StackOverflow in particular, is a widely recognized platform on which people ask questions and get answers by the community. Community members are incentivized to help and thoroughly answer questions through a system of points and badges [**?**].

3. The questions were selected randomly and sufficiently quantity to ensure a minimum 20% sampling for each StackExchange website (at the exception of Stack Overflow, less than 10%). The main statistics are reported in Table **??**. For each question, the following measures were extracted :

- when question was posted
- number of answers, their number of votes (+ and -)
- number of comments
- when best answer has been posted
- number of users involved in the response process

Layout Table 1 (General Statistics) columns : StackExchange Site, #questions, #questions sampled, percentage answered, median number of users involved, median number of answers, median time to accepted answer.

4. The data were collected through the StackExchange API, stored, processed and analyzed on Amazon Web Services. 5. The main metric of interest for each website is the probability distribution of waiting times before a question is answered. The other metrics described above shall be considered as control variables.

6. The distribution of waiting times is characterized by a heavy-tail, which is quantified using standard tools [**?**].

7. This method is particularly robust for .... of interest here.

8. Questions posted less than 6 months before the beginning of data collection (October 2013) were ignored.

9. The StackExchange API does not mention when the "best answer" has been accepted by the person asking the question. This has however a limited impact on our results since we care to know when the best answer was first provided. The best answer as chosen by the person who initially asked the question, might not be the one preferred by the number of votes by the community. We measure the effects of this distinction. We also care that some questions are not "forgotten" by the community, as the time we want to measure is the minimum time required to find a solution given a problem. Actually, unanswered questions are presented by StackExchange to the users to maximize the probability that they will answer them if they find a solution (find link documenting the process).

3. LIMITATIONS
4. CONCLUSIONS
APPENDIX
A. HEADINGS IN APPENDICES

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
\$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage