DESIGN LESSONS FROM THE
FASTEST Q&A SITE IN THE WEST

Lena Mamykina
Bella Manoim
Manas Mittal
George Hripcsak
Björn Hartmann

With a special guest appearance by Michael Terry
University of Waterloo

CHI 2011
Is there a regular expression to detect a valid regular expression?

Is it possible to detect a valid regular expression with another regular expression? If so, please give example code below.
8 Minutes Later:

Is there a regular expression to detect a valid regular expression?

Is is possible to detect a valid regular expression with another regular expression? If so please give example code below.

This is a recursive regex, and is not supported by many regex engines. PCRE based ones should support it.
digital_boy  Digital Boy  by codinghorror
Thanks to StackOverflow, I just solved an issue that I've been trying to resolve for 6 months. Why didn't I post sooner? cc @codinghorror
10 Nov

brheal  Bryan Rhea  by codinghorror
once again #stackoverflow comes through. It's like having a giant hallway full of experts at work. @codinghorror
14 Jan

halr9000  Hal Rottenberg  by codinghorror
RT @cmille19 I hate using forums other than StackOverFlow or ServerFault. MSDN should just outsource their forums (cc @codinghorror)
2 Dec
What Factors Drive This Success?
Method

- **Statistical data analysis** of the first two years (July’08 – July’10)
- **Interviews** with founders (2), designers (4), and users (6, Summer 2010)
Outline

- How Stack Overflow differs from other Q&A sites
- How Stack Overflow works
- Statistical analysis
- Three themes from interviews
- Discussion and implications
Trends in Large Q&A Sites

- Q&A sites often turn **conversational** and perform poorly on technical questions  
  [Nam; Rodrigues; Harper – all CHI’09]

- Social network Q&A is best suited for opinions and subjective answers  
  [Morris CHI’10; Horowitz WWW’10]
Stack Overflow: Overview

- Users ask questions in freeform text and answer questions of others
- Users vote for questions and answers
- Active participation earns points that translate into reputation and badges
When to prefer JSON over XML?

My requirement is just to display a set of values retrieved from database on a spread. I am using jQuery.

Favor XML over JSON when any of these is true:

- You need message validation
- You’re using XSLT
- Your messages include a lot of marked-up text
- You need to interoperate with environments that don’t support JSON

Favor JSON over XML when all of these are true:

- Messages don’t need to be validated, or validating their deserialization is simple
- You’re not transforming messages, or transforming their deserialization is simple
- Your messages are mostly data, not marked-up text
- The messaging endpoints have good JSON tools
When to prefer JSON over XML?

My requirement is just to display a set of values retrieved from database on a spread. I am using jQuery.

13 Answers

Favor XML over JSON when any of these is true:

- You need message validation
- You're using XSLT
- Your messages include a lot of marked-up text
- You need to interoperate with environments that don't support JSON

Favor JSON over XML when all of these are true:

- Messages don't need to be validated, or validating their deserialization is simple
- You're not transforming messages, or transforming their deserialization is simple
- Your messages are mostly data, not marked-up text
- The messaging endpoints have good JSON tools

Why don't you want marked-up text in json data? – bobobobo Jan 9 '10 at 21:32

4 JSON doesn't offer any advantage over XML in handling marked-up text. But I see your point; that's maybe overstated. – Robert Rossney Jan 10 '10 at 4:32
Amass enough reputation points and Stack Overflow will grant you additional privileges:

15  Vote up
15  Flag for moderator attention
50  Leave comments†
100 Edit community wiki posts
125 Vote down (costs 1 rep)
200 Reduced advertising
250 Vote to close, reopen, or migrate your questions
500 Retag questions
1000 Show total up and down vote counts
1500 Create new tags
2000 Edit other people’s posts, vote to approve or reject suggested edits
3000 Vote to close, reopen, or migrate any questions
5000 Vote to approve or reject suggested tag wiki edits
10000 Vote to delete closed questions, access to moderation tools
15000 Protect questions to prevent answers by new users
20000 Vote to delete negatively voted answers and stronger question deletion votes
Quantitative analysis
A Very Active Community:

July 2008 - July 2010:
300k Registered Users
833k Questions
2.2M Answers
2.9M Comments
7.8M Monthly Visitors
92.6% of Questions are Answered
Answers are Fast

<table>
<thead>
<tr>
<th>First Answer</th>
<th>Median:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11:00 min</td>
</tr>
</tbody>
</table>
“If you complained about a lack of response on a newsgroup after 24 hours you were labeled impatient; now you can realistically expect an answer within 20 minutes.”
(User 4)
Answer Dynamics: First 48 Hours

Most Answers Are Given in the First Four Hours

Cumulative histogram: time elapsed before questions receive answers (first 48 hrs)
It’s Been This Fast for Many Months
Infrequent Users
Frequent Users
Who Provides the Answers?
Who Provides the Answers?

Relative size and answer activity of different user groups

- Always low: 34.4% of answers, 94.4% of users
- Bursty: 21.9% of answers, 4.2% of users
- Continuously high: 27.8% of answers, 1.0% of users

User group (by activity signature)
Qualitative analysis
Method

- Semi-structured Skype interviews
  - Founders (2), design team (4), users (6)
Findings

- Productive competition
- Credibility within the community
- Evolutionary approach to design
Findings

- **Productive Competition**
  Prioritizing information over conversation

“What it [voting system] was doing on those sites wasn’t very successful because when you have comments and there’s a conversation going on, if you say, ‘Well, these are valuable comments and these are not valuable comments,’ then the only way to get a valuable thing to read is to take everything that’s highly voted. Then you’re skipping interim conversation.” (Founder 1)
Findings

- **Productive Competition**
  - External incentives

  “I am very competitive and you give me indication that a high number is good and I will try to get a high number. I don’t think it’s to do with reputation so much as, ‘This is a game.’” (User 4)

  “Stack Overflow - it’s like World of Warcraft, only more productive.” (User 5)
Findings

- **Productive Competition**
  Enabling an ecology of user activity
  - Shooting star vs. continuously active users
Findings

- **Credibility in the community**
  Achieving critical mass

  “So on the first day, the first question I could come up with had already been asked and answered, and there were three or four answers, and some voting had happened. The best answer had already been voted to the top. So on the first day when I saw everything working, I knew that we were in really good shape.” (Founder 2)
Findings

- **Credibility in the community**
  Acceptance and negotiation of founders’ vision

  “You know, when you hear that these guys came up with something you wanna go check it out.”
  *(User 6)*
Findings

- **Evolutionary approach to design**
  Tight feedback loop with users

  "We pretty much had to forget all the software engineering processes we learned." (Designer 2)
Findings

- **Evolutionary approach to design**
  
  Rapid design iterations

  “We pretty much release new versions every day. Sometimes they are really small changes; the bigger ones often get announced on Meta.” (Designer 1)
Can this success be replicated?

- Stack Exchange 1.0:
  Licensed the SO engine for profit: **FAILED!**
Stack Overflow’s success

- Prioritizes information over social networking
- Impressive track record (percentage of questions answered, time to answer)
- Prominence and popularity in the community
Summary

- Stack Overflow’s success
- Patterns of activity
  - Asking and answering (infrequent users ask, active users answer)
  - User signatures (“shooting star”)

Summary

- Stack Overflow’s success
- Patterns of activity
- Design choices and process
  - Incentives
  - Community participation
  - Evolutionary design
Challenges for Research

Design matters…
…but community involvement matters more.

What role should academic systems research play in social computing?
Thank You

Questions:
lena.mamykina@dbmi.columbia.edu
bjoern@eecs.berkeley.edu
Findings

- **Challenges**
  - Higher entry barrier for noobs
  - Stack Overflow Fatigue
  - Persistence of accepted answers
Some questions are not well supported

1. Questions about *obscure* technologies.
2. Questions that are *tedious* to answer.
3. Problems that are *hard to reproduce* with a small code fragment.
4. Questions that do not have a clear best answer and *invite discussion*. 
Obscure Topics Take Longer

Tags as indicators of fast (median < 10min) and slow (median > 1 day) answers to questions
Activity Signatures

- **Activity signatures**
  capture user’s activity (#answers) over time

- Each user month coded as High (H) or Low (L) into activity string, e.g.:
  \text{LLHHHLLL}

- Matched strings into groups using regular expressions, e.g.:
  Bursty: \text{L* (H | HH | HHH) L*}
Related Work: Q&A Site Analyses

- **Large scale algorithmic analyses:**
  Yahoo! Answers
  [Adamic WWW’08, Bouguessa KDD’08, Gyöngyi QAWeb’08]

- **Mixed Methods & Qualitative Approaches:**
  KiN [Nam CHI’09]
  Unanswered Questions [Dearman CHI’10]

- **Social Network Q&A:**
  Facebook & Twitter [Morris CHI’10]
  Aardvark [Horowitz WWW’10]