

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

Lena Mamyk  
Bella M  
Man  
Ge  
Björn

With a special guest  
appearance by **Michael Terry**  
University of Waterloo

CHI 2011

 COLUMBIA UNIVERSITY  
IN THE CITY OF NEW YORK

**Bard**  
A PLACE TO THINK



Berkeley

# Question:

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

▲  
45

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



regex

☆  
21

[link](#) | [edit](#) | [flag](#)

asked Oct 5 '08 at 17:07



psytek

258 ● 3 ● 6

# 8 Minutes Later:

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

▲  
45  
▼

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

regex



link | edit | flag

asked Oct 5 '08 at 17:07



psytek

258 ● 3 ● 6

▼ 1

answered Oct 5 '08 at 17:15



MizardX

9,249 ● 17 ● 31

```
/(?:[^\?+*{}()[\]\|]+|\|.|\[(?:[^\]\|]+|\|.)*\]|\\((?:\\?[:=!]|\\?<[=!]|\\?>)?(?:R\\)))(?:[:(
```

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]



**kevindente** Kevin Dente  by codinghorror

From what I've seen so far, Stack Exchange is where people go for facts. Quora is where people go for opinions.

24 Jan




# 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

# How Stack Overflow Works

 **stackoverflow** [Questions](#) [Tags](#) [Users](#) [Badges](#) [Unanswered](#)

**Top Questions** [active](#) [162 featured](#) [hot](#) [week](#) [month](#)

0 votes	2 answers	26 views	<a href="#">Help with Javascript to C# converter</a>	<a href="#">c#</a> <a href="#">javascript</a> <a href="#">converter</a>	13m ago <a href="#">griegs</a> <b>6,893</b>
6 votes	6 answers	76 views	<a href="#">How to comment/uncomment in HTML code</a>	<a href="#">html</a> <a href="#">coding-style</a> <a href="#">coding</a> <a href="#">comment</a>	19m ago <a href="#">Jeff Knecht</a> <b>1</b>
2 votes	1 answer	45 views	<a href="#">Contents of a static library</a>	<a href="#">c++</a> <a href="#">c</a> <a href="#">library</a> <a href="#">gcc</a>	40m ago <a href="#">Jonathan Leffler</a> <b>61.3k</b>
1 vote	2 answers	24 views	<a href="#">Most Efficient way to 'look up' Keywords</a>	<a href="#">c++</a> <a href="#">search</a> <a href="#">efficiency</a> <a href="#">lexer</a>	2m ago <a href="#">ergosys</a> <b>3,567</b>
2 votes	3 answers	11 views	<a href="#">unknown css effects in moz and webkit</a>	<a href="#">css</a> <a href="#">webkit</a> <a href="#">effects</a> <a href="#">tips-and-tricks</a> <a href="#">mozilla</a>	2m ago <a href="#">Pandiya Chendur</a> <b>6,317</b>

# Answers


## 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.

40 [javascript](#) [jquery](#) [xml](#) [json](#)

▼ link | flag

☆ 35

edited Nov 28 '08 at 5:23  
 [Dustin](#)  
9,766 ● 11 ● 43

asked Nov 28 '08 at 5:14  
 [sarego](#)  
604 ● 8 ● 21  
100% accept rate

### 13 Answers

oldest newest **votes**

▲ Favor XML over JSON when any of these is true:

54

▼


✓

- 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

link | flag

answered Dec 1 '08 at 0:34  
 [Robert Rossney](#)  
15.9k ● 2 ● 14 ● 54

Why don't you want marked-up text in json data? – [bobobobo](#) Jan 9 at 21:32

# When to prefer JSON over XML?

**liquid** Complete XML Toolkit  
XSD | XQuery | XSLT | WSDL | DTD | Docs

DOWNLOAD FREE TRIAL

NEW FOR 2011

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

47 [javascript](#) [jquery](#) [xml](#) [json](#)

link | edit | flag

46

edited Nov 28 '08 at 5:23

[Dustin](#)  
12.5k ● 16 ● 53

asked Nov 28 '08 at 5:14

[sarego](#)  
678 ● 10 ● 22  
100% accept rate

## 13 Answers

active oldest votes

63 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

link | edit | flag

answered Dec 1 '08 at 0:34

[Robert Rossney](#)  
22k ● 3 ● 21 ● 67

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 4:52

### Hello World!

This is a collaboratively edited question and answer site for professional and enthusiast programmers. It's 100% free, no registration required.

[about](#) » [faq](#) »

tagged

[javascript](#) × 95357

[jquery](#) × 82093

[xml](#) × 22279

[json](#) × 9571

asked 2 years ago

viewed 5,593 times

active 8 months ago

How can I draw the outline of a path in tikz?

Do you haz teh codez?  
(If you need to find programmers, we have a site for that) → [Get Programmers](#)

### CAREERS 2.0

by stackoverflow

[Gameloft Programmer](#)  
Gameloft  
New Orleans, LA

[Frontend Developer](#)  
Kongregate  
San Francisco, CA

I use JSON unless I'm required to use XML. It's simpler to understand, and (because it requires less

Amass enough reputation points and Stack Overflow will grant you [additional privileges](#):

- 15** Vote up
- 15** Flag for moderator attention
- 50** Leave comments<sup>†</sup>
- 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



- Questions
- Tags
- Users
- Badges
- Unanswered

Ask Question

### Top Questions

bugs reqs **active** 8 featured hot week month

- 7 votes **16** answers 247 views **What's your up/down vote ratio?** discussion voting 36m ago Jarrod Roberson 136
- 4 votes **2** answers 65 views **Down voted after reputation cap was reached** discussion reputation down-votes up-votes 56m ago kiamlaluno 1,242
- 4 votes **1** answer 29 views **Should active bounty questions be protectable?** discussion bounty protected-questions 58m ago Rosinante 7,452
- 2 votes **1** answer 42 views **Linked account +100 rep bonus also applied to account with 200** bug status-bydesign reputation bonus-reputation 1h ago kiamlaluno 1,242
- 132 votes **23** answers 3 kviews **Encouraging people to explain down-votes** feature-request status-completed voting down-votes 1h ago Danny 122
- 1 vote **0** answers 34 views **'You are on the Featured tab' - but there's no Featured tab [closed]** bug tabs 1h ago GrzegorzOledzki 211
- 0 votes **2** answers 42 views **Displaying only favorite tags on homepage** support tags favorites filtering favorite-tags 1h ago Jeff Atwood 132k
- 12 votes **2** answers 429 views **Which Area 51 proposals are migration proposals?** discussion area51 1h ago moberley 384

### Welcome!

This site is intended for **bugs, features, and discussion** of [Stack Overflow](#) and the Stack Exchange Q&A engine. You must have an account on a [Stack Exchange network website](#) to participate.

[about](#) » [faq](#) »

### Frequently Asked

- How to get attention for your old, unanswered questions
- How does "Reputation" work?
- FAQ for Stack Exchange sites
- Where can I ask questions that aren't Stack Overflow questions?
- How to deal with Google questions
- Which tools and technologies build the Stack Exchange Network?
- How to ask and answer homework questions?
- What are "Community Wiki" posts?
- How does the bounty system work?
- The Many Memes of Meta

more [faq](#) »

### Recent Tags



# Quantitative analysis

# A Very Active Community:



**July 2008 - July 2010:**

300k Registered Users

833k Questions

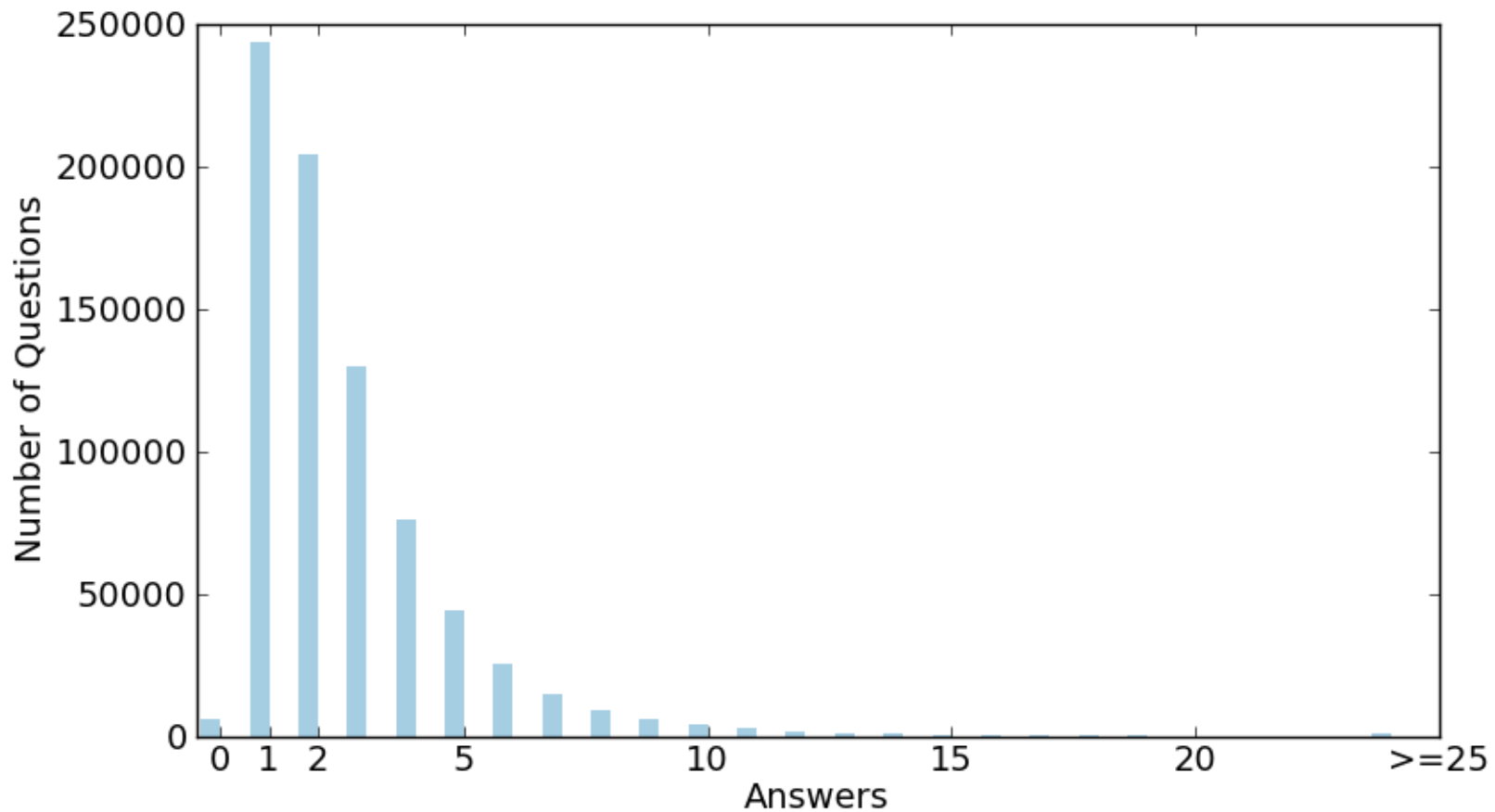
2.2M Answers

2.9M Comments

7.8M Monthly Visitors



# 92.6% of Questions are Answered



# Answers are Fast



	<b>Median:</b>
<b>First Answer</b>	<b>11:00 min</b>

# Answers are Fast

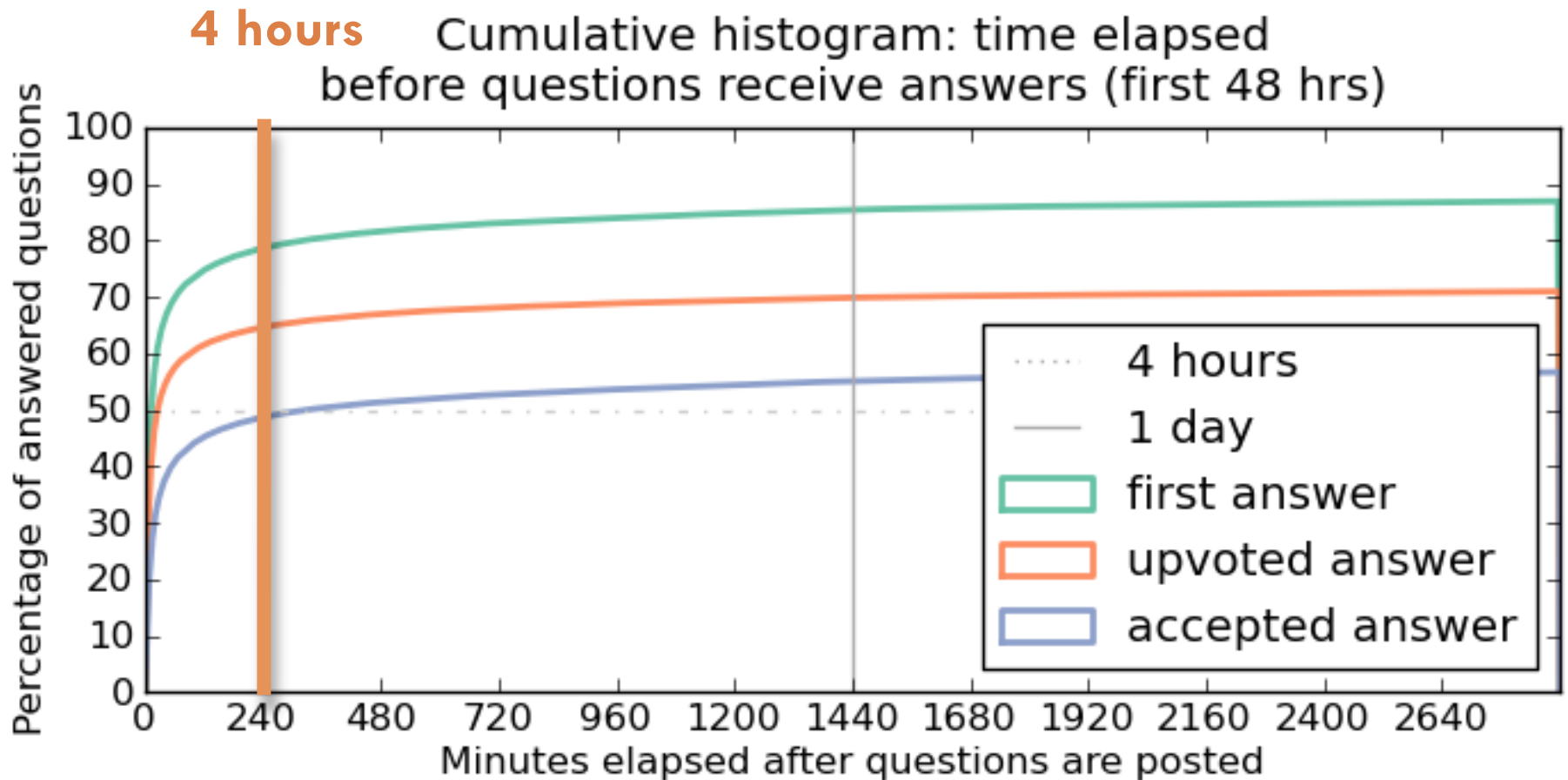


*“ If you complained about a lack of response on a newsgroup after 24 hours you were labeled impatient; now you can realistically expect an answer within 20 minutes.”*

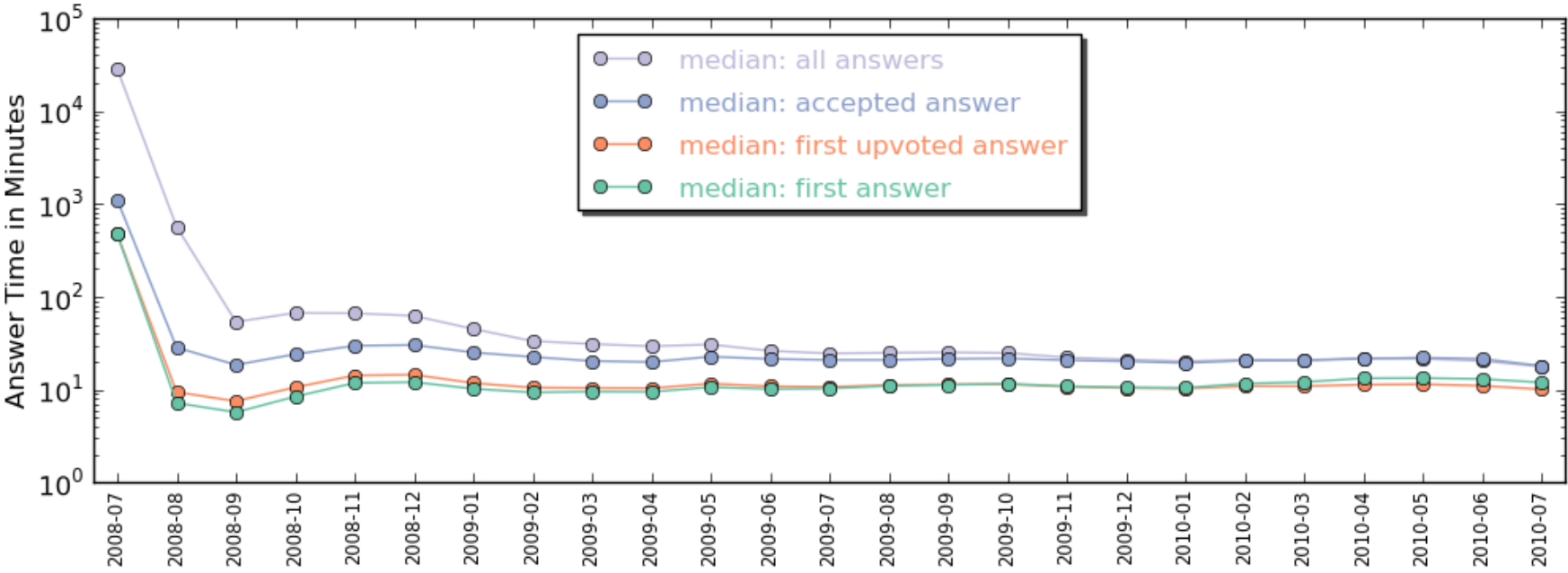
*(User 4)*

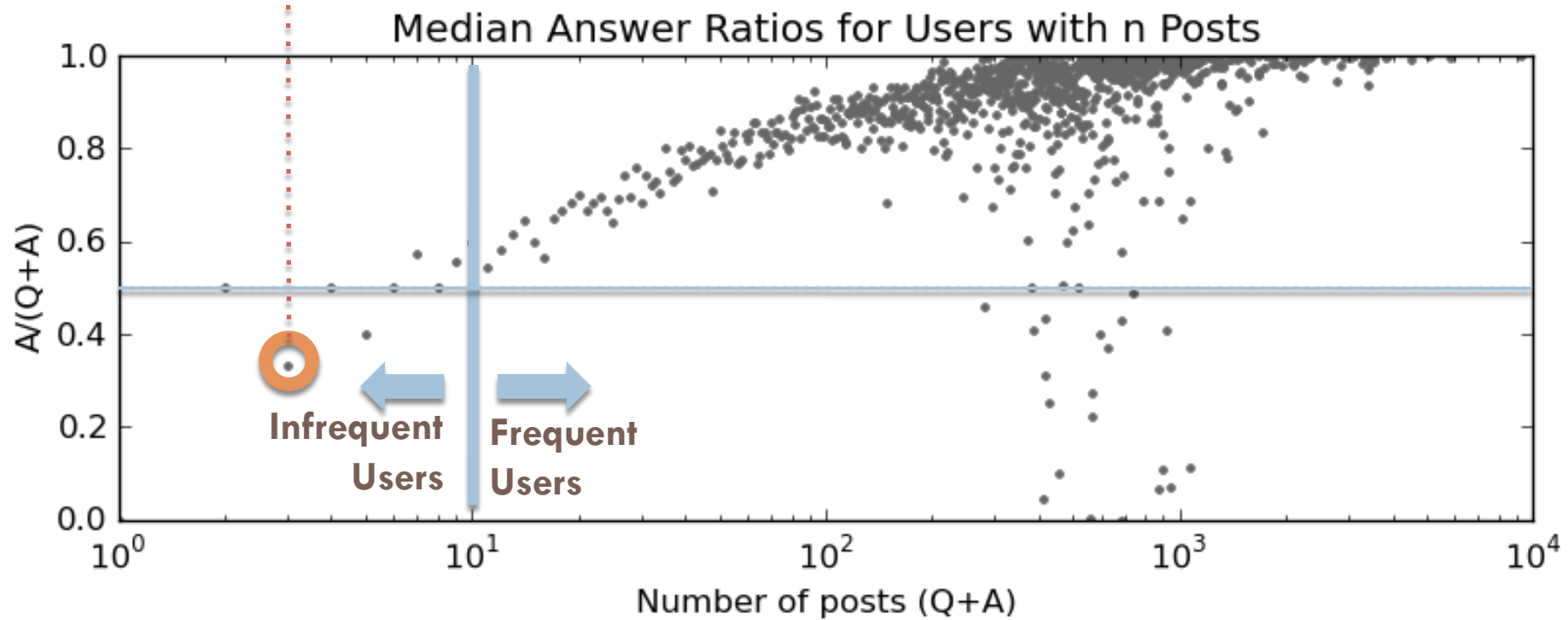
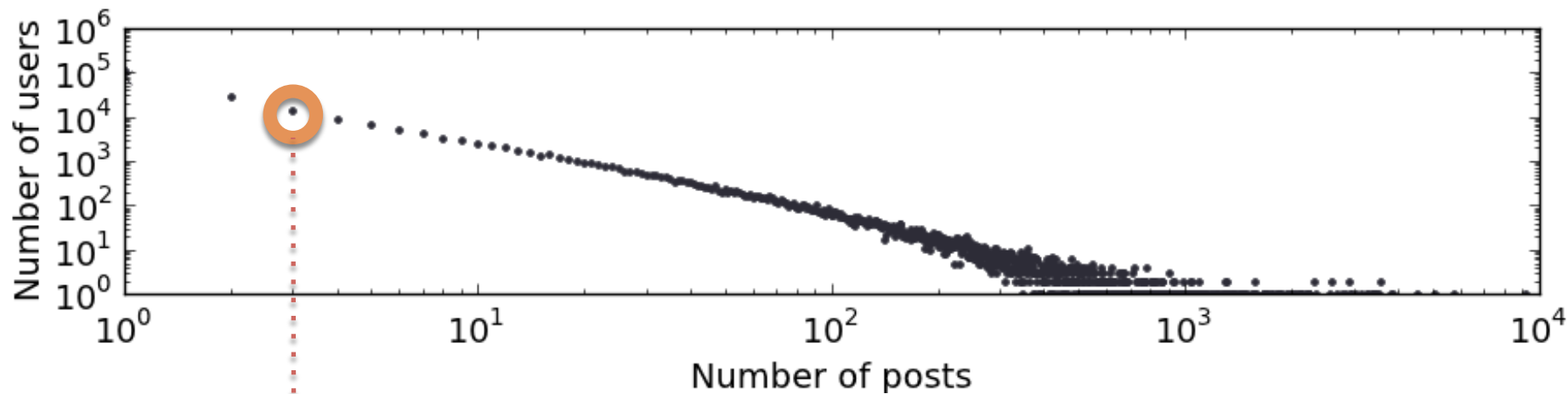
# Answer Dynamics: First 48 Hours

Most Answers Are Given in the First Four Hours



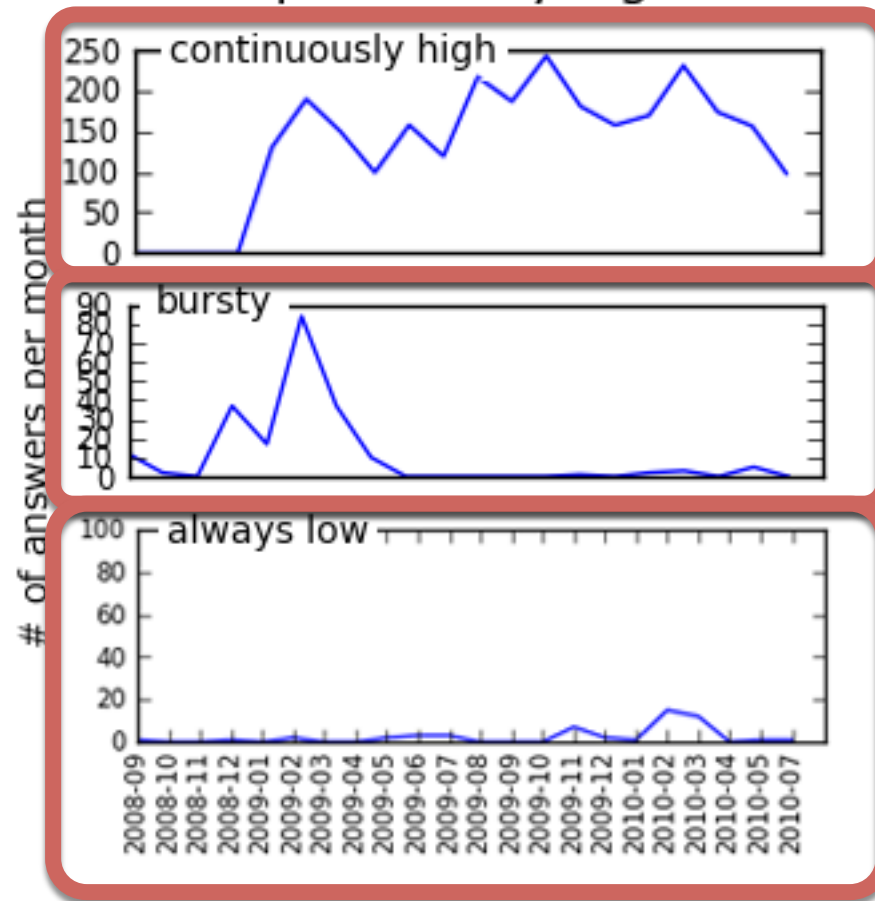
# It's Been This Fast for Many Months



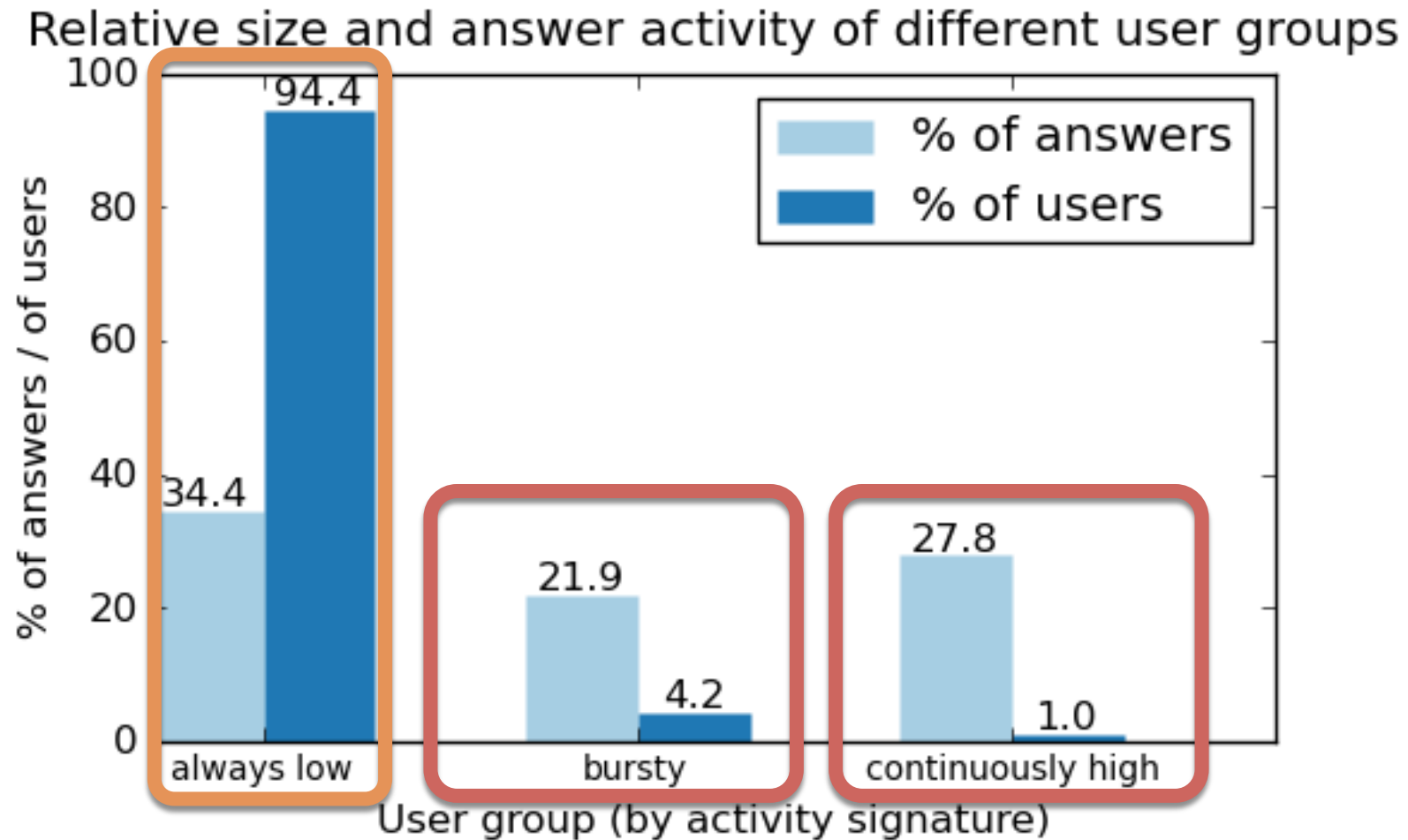


# Who Provides the Answers?

Example activity signatures



# Who Provides the Answers?







# 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)*

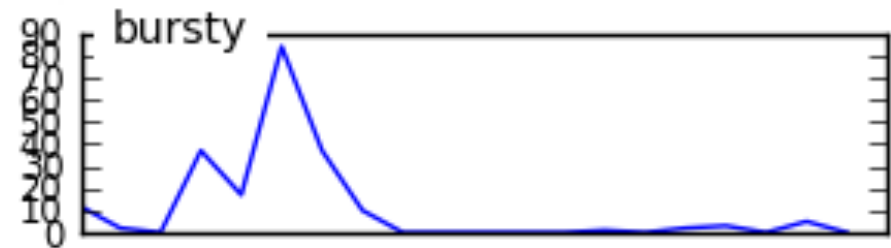
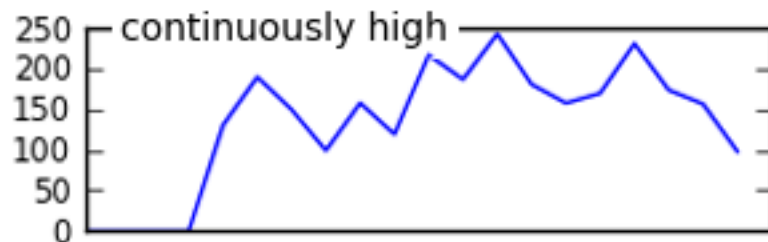
*“ Stack Overflow - it's like World of Warcraft, only more productive.” (User 5)*

# 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!**

# Summary



- 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](mailto:lena.mamykina@dbmi.columbia.edu)

[bjoern@eecs.berkeley.edu](mailto:bjoern@eecs.berkeley.edu)



# Findings



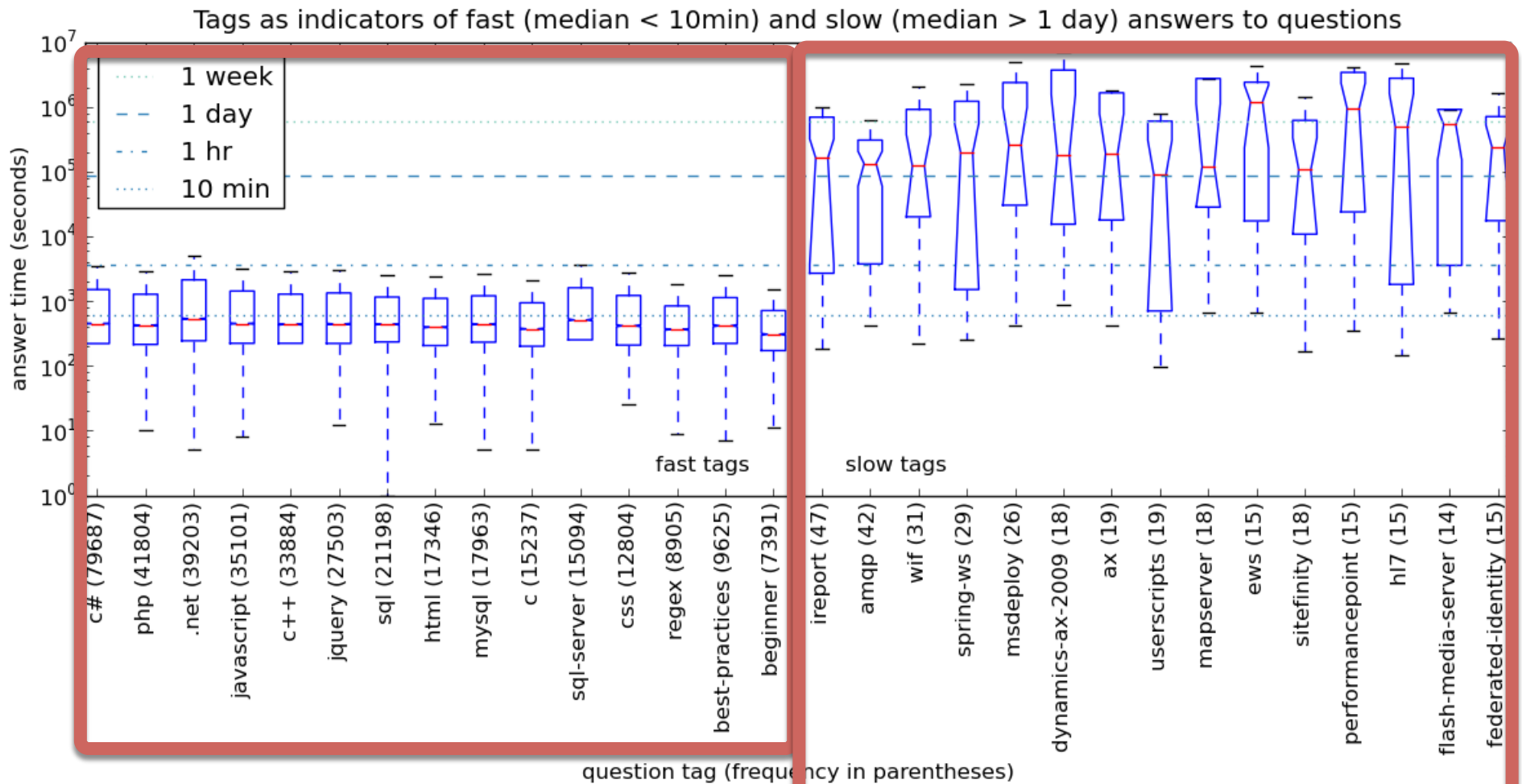
- Challenges
  - Higher entry barrier for noobs
  - Stack Overflow Fatigue
  - Persistence of accepted answers

# 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



# 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.:  
LLHHHLLL
- Matched strings into groups using regular expressions, e.g.:  
Bursty:  $L^* (H | HH | HHH) L^*$

# 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]