Networks and data as catalysts for interactive systems

Michael Bernstein
18 April 2013
[{
  "id": "0001",
  "type": "donut",
  "name": "Cake",
  "ppu": 0.55,
  "batters": [
    {
      "id": "1001",
      "type": "Regular"
    },
    {
      "id": "1002",
      "type": "Chocolate"
    },
    {
      "id": "1003",
      "type": "Blueberry"
    },
    {
      "id": "1004",
      "type": "Devil's Food"
    }
  ],
  "topping": [
    {
      "id": "5001",
      "type": "None"
    },
    {
      "id": "5002",
      "type": "Glazed"
    },
    {
      "id": "5005",
      "type": "Sugar"
    },
    {
      "id": "5007",
      "type": "Powdered Sugar"
    },
    {
      "id": "5006",
      "type": "Chocolate with Sprinkles"
    },
    {
      "id": "5003",
      "type": "Chocolate"
    },
    {
      "id": "5004",
      "type": "Maple"
    }
  ]
},
{
  "id": "0002",
  "type": "donut",
  "name": "Raised",
  "ppu": 0.55,
  "batters": [
    {
      "id": "1001",
      "type": "Regular"
    }
  ],
  "topping": [
    {
      "id": "5001",
      "type": "None"
    },
    {
      "id": "5002",
      "type": "Glazed"
    },
    {
      "id": "5005",
      "type": "Sugar"
    },
    {
      "id": "5003",
      "type": "Chocolate"
    },
    {
      "id": "5004",
      "type": "Maple"
    }
  ]
},
{
  "id": "0003",
  "type": "donut",
  "name": "Old Fashioned",
  "ppu": 0.55,
  "batters": [
    {
      "id": "1001",
      "type": "Regular"
    },
    {
      "id": "1002",
      "type": "Chocolate"
    }
  ],
  "topping": [
    {
      "id": "5001",
      "type": "None"
    },
    {
      "id": "5002",
      "type": "Glazed"
    },
    {
      "id": "5003",
      "type": "Chocolate"
    },
    {
      "id": "5004",
      "type": "Maple"
    }
  ]}]}
Crowd-scale interaction
1) Design
2) Data
Personalization via friendsourcing

Bernstein, Tan, Smith, Czerwinski, and Horvitz. TOCHI 2010.
Bernstein, Marcus, Karger, and Miller. CHI 2010.
Bernstein, Tan, Smith, Czerwinski, and Horvitz. UIST 2009.
crowdsourcing can solve problems at massive scale.
Crowdsourcing can solve problems at massive scale.
crowdsourcing can solve problems at massive scale.
crowdsourcing can solve problems at massive scale.
crowdsourcing

struggles when only a small group is qualified.
crowdsourcing struggles when only a small group is qualified.
crowdsourcing struggles when only a small group is qualified.
crowdsourcing struggles with personalized applications.

Ask a question and I'll find someone to answer

When's the next MIT Idol singing competition?

Example questions

Ask someone

Christopher Jobst
This sounds fantastic! Half off in a city like SF is a great deal!

Max 'Electronic' Van Kleek
Why are american CEOs not like this? [ Japan Air Line's CEO gets paid less than pilots, interacts directly with staff at all levels ] - http://bit.ly/fHfEW1

Facebook News Feed

My6Sense

Android Geek Gadgets...
First iPad commercial at the 82nd Annual Academy Awards http://b...

eMarketer Articles
Mobile Users Want Personalized Services

Relevance Time

The Daily Beast - Blogs...
10 Technologies That Empower Women

Mashable!
Foursquare iPhone App Vanishes from App Store Due to “Hiccup”

Enon Landenberg
@ the Facebook party http://mypict.m
Your friends, family and colleagues know your interests, activities, and personality.
Your friends, family and colleagues know your interests, activities, and personality – and use it to help you.
Our Goal: weighted descriptors

Greg Smith
Microsoft
Stanford Alum ‘92

input

output
poker
MSR
hacker
band
dogs
vegas
groupbar
c#
Collabio
Tag Kurt to reveal each hidden item. One point for each tag, another point for each other friend who used the same tag to describe Kurt!

Kurt's friends have tagged him with:

- Camera
- Laptop
- Tent
- MacBook Air

My Score: 0 points
Start tagging Kurt to get points!
# Design Space

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag visibility</td>
<td>Never visible</td>
<td>Visible when guessed</td>
<td>Always visible</td>
</tr>
<tr>
<td>Tag anonymity</td>
<td>Anonymous</td>
<td>Public to tagged user only</td>
<td>Public to all users</td>
</tr>
<tr>
<td>Point mechanism</td>
<td>None</td>
<td>Reward uncommon information</td>
<td>Reward common information</td>
</tr>
<tr>
<td>Bootstrapping untagged users</td>
<td>None</td>
<td>Game mechanism (e.g., extra points)</td>
<td>Social mechanism (e.g., pretending another user has tagged first)</td>
</tr>
<tr>
<td>Tag Deletion</td>
<td>Send notification</td>
<td></td>
<td>Silent deletion</td>
</tr>
<tr>
<td>Synchronicity</td>
<td>Synchronous (e.g., ESP Game)</td>
<td></td>
<td>Asynchronous</td>
</tr>
<tr>
<td>Tagging Yourself</td>
<td>Allowed</td>
<td></td>
<td>Not allowed</td>
</tr>
</tbody>
</table>
# Design Space

<table>
<thead>
<tr>
<th>Tag visibility</th>
<th>Never visible</th>
<th>Visible when guessed</th>
<th>Always visible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag anonymity</td>
<td>Anonymous</td>
<td>Public to tagged user only</td>
<td>Public to all users</td>
</tr>
<tr>
<td>Point mechanism</td>
<td>None</td>
<td>Reward uncommon information</td>
<td>Reward common information</td>
</tr>
<tr>
<td>Bootstrapping untagged users</td>
<td>None</td>
<td>Game mechanism (e.g., extra points)</td>
<td>Social mechanism (e.g., pretending another user has tagged first)</td>
</tr>
<tr>
<td>Tag Deletion</td>
<td>Send notification</td>
<td>Silent deletion</td>
<td></td>
</tr>
<tr>
<td>Synchronicity</td>
<td>Synchronous (e.g., ESP Game)</td>
<td>Asynchronous</td>
<td></td>
</tr>
<tr>
<td>Tagging Yourself</td>
<td>Allowed</td>
<td>Not allowed</td>
<td></td>
</tr>
</tbody>
</table>

## Design Space

<table>
<thead>
<tr>
<th></th>
<th>Never visible</th>
<th>Visible when guessed</th>
<th>Always visible</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tag visibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tag anonymity</strong></td>
<td>Anonymous</td>
<td>Public to tagged user only</td>
<td>Public to all users</td>
</tr>
<tr>
<td><strong>Point mechanism</strong></td>
<td>None</td>
<td>Reward uncommon information</td>
<td>Reward common information</td>
</tr>
<tr>
<td><strong>Bootstrapping untagged users</strong></td>
<td>None</td>
<td>Game mechanism (e.g., extra points)</td>
<td>Social mechanism (e.g., pretending another user has tagged first)</td>
</tr>
<tr>
<td><strong>Tag Deletion</strong></td>
<td>Send notification</td>
<td>Silent deletion</td>
<td></td>
</tr>
<tr>
<td><strong>Synchronicity</strong></td>
<td>Synchronous (e.g., ESP Game)</td>
<td>Asynchronous</td>
<td></td>
</tr>
<tr>
<td><strong>Tagging Yourself</strong></td>
<td>Allowed</td>
<td>Not allowed</td>
<td></td>
</tr>
</tbody>
</table>
Viral spread on Facebook
July 2008 – March 2009

• 29,307 tags on 3,831 individuals
• 825 taggers, 1,350 installs
• Median tagger received 11 tags
Viral spread on Facebook
July 2008 – March 2009

• 29,307 tags on 3,831 individuals
• 825 taggers, 1,350 installs
• Median tagger received 11 tags

Collabio’s most valuable tags were the ones in the long tail: accurate and not available on Facebook or the web
Can we take advantage of existing social interactions instead of designing new ones?
Friendsourced content sharing

Related to your research

Michael Bernstein to Katrina
Who could be sharing more?

1. Sharers are those who seek out large volumes of web content
2. Sharers are especially social individuals
What explains sharing interest?

Method

- Survey (N=100) on Amazon Mechanical Turk
- Vetted for cheaters
- Paid $0.05
What explains sharing interest?
Method: 4 scales of 10 questions each
What explains sharing interest?
Method: 4 scales of 10 questions each

Sharing:

“I often tell people I know about my favorite web sites to follow.”
What explains sharing interest?
Method: 4 scales of 10 questions each

Sharing:
“I often tell people I know about my favorite web sites to follow.”

Seeking:
“I often seek out entertaining posts, jokes, comics and videos using the Internet.”
What explains sharing interest?

Method: 4 scales of 10 questions each

Sharing:
“...I often tell people I know about my favorite web sites to follow...”

Seeking:
“...I often seek out entertaining posts, jokes, comics and videos using the Internet...”

Bridging social capital:
“...I come in contact with new people all the time...”

[Ellison et al. 2007]
What explains sharing interest?
Method: 4 scales of 10 questions each

Sharing:
“I often tell people I know about my favorite web sites to follow.”

Seeking:
“I often seek out entertaining posts, jokes, comics and videos using the Internet.”

Bridging social capital:
“I come in contact with new people all the time.”

Bonding social capital:
“There is someone I can turn to for advice about making very important decisions.”

[Ellison et al. 2007]
What explains sharing interest?

Results

Predicting outcome on sharing scale:

<table>
<thead>
<tr>
<th>factor</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seeking</td>
<td>0.74</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Bridging social capital</td>
<td>0.22</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Bonding social capital</td>
<td>0.01</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Adj. $R^2 = 0.56$
What explains sharing interest?

Results

Predicting outcome on sharing scale:

<table>
<thead>
<tr>
<th>factor</th>
<th>$\beta$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seeking</td>
<td>0.74</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Bridging social capital</td>
<td>0.22</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Bonding social capital</td>
<td>0.01</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Adj. $R^2 = 0.56$

+ Sharers are those who seek out large volumes of web content
+/- Sharers are especially social individuals
Support and empower heavy content seekers in directed content sharing
• Aid the few people who read and share content constantly to learn about the others
• Lower the activation barrier to sharing through recommendations
• Address inhibitions in sharing old news
Recommendations
Annotate each post with friends who might be interested in the content
Recommendations

Lifehacker: Share with friends using MIT’s FeedMe

msbernst@mit.edu
1 FeedMe today

rcm@mit.edu
0 FeedMes today

karger@mit.edu
5 FeedMes today

Type a name...
Recommendations

Lifehacker: Share with friends using MIT’s FeedMe

- msbernst@mit.edu 1 FeedMe today
- rcm@mit.edu 0 FeedMes today
- karger@mit.edu 5 FeedMes today

Type a name...

Add an optional comment...

Now | Later
Recommendations

Lifehacker: Share with friends using MIT’s FeedMe

msbernst@mit.edu  1 FeedMe today
rcm@mit.edu      0 FeedMes today
karger@mit.edu   5 FeedMes today

Type a name...
Add an optional comment...

Now  Later

[FeedMe] Lifehacker: Share with friends using MIT's FeedMe

via Lifehacker:
Share with friends using MIT's FeedMe
Send Michael a One-Click Thanks if you liked this post!

A group of researchers presented today at the prestigious CHI 2010 conference a new sharing system embedded in Google Reader. Rather than spamming all your friends, FeedMe focuses on helping you share with your inner circle.

The findings from this research project have been published in an research paper available here on the project's homepage.
Models built without recipient involvement

MIT HCI Research

Computer Science Education
Models built without recipient involvement
Models built without recipient involvement
Recommendation details

On Archiving, Curating, and Republishing Public Twitter Conversations

Once you've amassed enough of a following, one of the best uses of hot social networking app Twitter is getting instant answers to any question on your mind. When you post a question on Twitter and get a dozen replies within the next 10 minutes from live humans--some of whom you know and trust--it's waaayyy better than impersonal and sometimes out-of-date Google search results.
On Archiving, Curating, and Republishing Public Twitter Conversations

Once you've amassed enough of a following, one of the best uses of hot social networking app Twitter is getting instant answers to any question on your mind. When you post a question on Twitter and get a dozen replies within the next 10 minutes from live humans—some of whom you know and trust—it's wayy better than impersonal and sometimes out-of-date Google search results.

twitter: 38
tweet: 30
social: 27
post: 23
conversation: 19
answers: 10
blog: 3
google: 1
On Archiving, Curating, and Republishing Public Twitter Conversations

Once you've amassed enough of a following, one of the best uses of hot social networking app Twitter is getting instant answers to any question on your mind. When you post a question on Twitter and get a dozen replies within the next 10 minutes from live humans—some of whom you know and trust—it's waayyy better than impersonal and sometimes out-of-date Google search results.

joe@sixpack.com:
- sports: 200
- baseball: 150
- sox: 132
- lacrosse: 89
- workout: 41
- hiking: 23
- vitamin: 22

rcm@mit.edu:
- design: 184
- tweet: 170
- web: 79
- twitter: 48
- social: 43
- friendfeed: 32
- blog: 25
- developer: 23

twitter: 38
tweet: 30
social: 27
post: 23
conversation: 19
answers: 10
blog: 3
google: 1
Awareness indicators

rcm@mit.edu 0 FeedMes today
rcm@mit.edu 5 FeedMes today
rcm@mit.edu Seen it already

Address concerns about volume:
“How much are we sending them?”

Give an indication of whether it’s old news
“Oh, somebody already sent it to them?”
Digests: managing volume
Share without overwhelming the inbox

Now  Later

[FeedMe] Personalized Newspaper: ControlPad Turns Your Number Pad into an Application Launcher [Downloads]

FeedMe to msbernst
show details 8/18/09  Reply

Your friends using MIT's FeedMe tool thought that these posts might be interesting to you, but didn't want to bother you with a separate email at the time. Here's your weekly newspaper, as authored by your friends!

ControlPad Turns Your Number Pad into an Application Launcher [Downloads], via Lifehacker from foc@gmail.com: "cool application~"
Send feng a One-Click Thanks if you liked this post!

StarCraft II Single-Player Details Revealed, via Slashdot from karger@mit.edu

Google GMail Passes AOL, Becoming Third Most Popular E-mail - InformationWeek, via Google News - Sci/Tech from rcm@mit.edu
Send Mitchell a One-Click Thanks if you liked this post!

Designing SMS apps for mobile Africa, via Official Google Mobile Blog from kp@csail.mit.edu: "in case you guys didn't get the previous messages, here is one more."
Send Min a One-Click Thanks if you liked this post!
One-click thanks
Low-effort recipient feedback

Michael Bernstein to haystack-group show details Mar 19

Upcoming ICWSM paper:

via ReadWriteWeb:

The Million Follower Fallacy: Audience Size Doesn't Prove Influence on Twitter
Send Michael a One-Click Thanks if you liked this post!

A group of researchers have proven something we already expected to be the case: your Twitter follower count is somewhat of a meaningless metric when it
One-click thanks
Low-effort recipient feedback

Michael Bernstein to haystack-group

Upcoming ICWSM paper:
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

via ReadWriteWeb:

The Million Follower Fallacy: Audience Size Doesn't Prove Influence on Twitter

Send Michael a One-Click Thanks if you liked this post!

A group of researchers have proven something we already expected to be the case: your Twitter follower count is somewhat of a meaningless metric when it
One-click thanks
Low-effort recipient feedback

Upcoming ICWSM paper:

via ReadWriteWeb:

**The Million Follower Fallacy: Audience Size Doesn't Prove Influence on Twitter**

Send Michael a One-Click Thanks if you liked this post!

A group of researchers have proven something we already expected to be the case: your Twitter follower count is somewhat of a meaningless metric when it

Thanks: [FeedMe] The Million Follower Fallacy

marcua@mit.edu says thanks for sharing this post!

via ReadWriteWeb

**The Million Follower Fallacy: Audience Size Doesn't Prove Influence on Twitter**

Send Michael a One-Click Thanks if you liked this post!
1) Design
2) Data
Crowd and social data for long-tail design

weather boston
weather boston

**Boston, MA weather | Boston.com**
www.boston.com/weather/
Complete weather for Boston, Massachusetts, and the world.
Extended forecast for Boston - Late-season storm has a wallop - Blizzard of '78? - 10

**Boston Weather Forecast and Conditions**
www.weather.com/weather/today/Boston+MA+USMA0046
Boston weather forecast and weather conditions. Today's Boston weather plus a 36 hour forecast and Doppler radar from weather.com
weather boston

Weather for Boston, MA

28°F | °C
Overcast
Wind: S at 14 mph
Humidity: 41%

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
</tr>
</thead>
<tbody>
<tr>
<td>32°</td>
<td>29°</td>
<td>47°</td>
<td>34°</td>
</tr>
<tr>
<td>36°</td>
<td>14°</td>
<td>31°</td>
<td>25°</td>
</tr>
</tbody>
</table>

Detailed forecast: The Weather Channel - Weather Underground - AccuWeather

Boston, MA weather | Boston.com
www.boston.com/weather/
Complete weather for Boston, Massachusetts, and the world.

Extended forecast for Boston - Late-season storm has a wallop - Blizzard of '78? - 10

Boston Weather Forecast and Conditions
www.weather.com/weather/today/Boston+MA+USMA0046

Boston weather forecast and weather conditions. Today's Boston weather plus a 36 hour forecast and Doppler radar from weather.com.
the girl with the dragon tattoo

The Girl With the Dragon Tattoo showtimes for Cambridge, MA

The Girl with the Dragon Tattoo is the first film in Columbia Pictures' three-picture adaptation of Stieg Larsson's literary ...

AMC Loews Harvard Square 5 - 10 Church Street, Cambridge, MA - Map
1:55 - 5:10 - 8:30pm

Somerville Theatre - 55 Davis Square, Somerville, MA - Map
4:00 - 7:10pm

memorial day 2012

Memorial Day is on Monday, May 27.

AAPL

AAPL - Apple Inc. (NASDAQ)

419.81  -1.58  (-0.37%)

Jan 13 4:00pm ET - Disclaimer

Open: 419.70  Volume: 8,077,107
High: 420.45  Avg Vol: 10,438,000
Low: 418.66  Mkt Cap: 390.18B
The Girl with the Dragon Tattoo

The Girl With the Dragon Tattoo showtimes for Cambridge, MA

★★★★☆ 60 reviews - 2hr 40min - Rated R - Drama
The Girl with the Dragon Tattoo is the first film in Columbia Pictures' three-picture adaptation of Stieg Larsson's literary ...

AMC Loews Harvard Square 5 - 10 Church Street, Cambridge, MA - Map
1:55 - 5:10 - 8:30pm

Somerville Theatre - 55 Davis Square, Somerville, MA - Map
4:00 - 7:10pm

Memorial Day is on Monday, May 27.

AAPL

AAPL - Apple Inc. (NASDAQ)

419.81 -1.58 (-0.37%)
Jan 13 4:00pm ET - Disclaimer
Open: 419.70 Volume: 8,077,107
High: 420.45 Avg Vol: 10,438,000
Low: 418.66 Mkt Cap: 390.18B

Chilton & Teevan 2009
the girl with the dragon tattoo

Memorial Day is on Monday, May 27.

AAPL - Apple Inc. (NASDAQ)

419.81 -1.58 (-0.37%)
Jan 13 4:00pm ET - Disclaimer
Open: 419.70  Volume:     8,077,107
High: 420.45  Avg Vol: 10,438,000
Low: 418.66   Mkt Cap: 390.18B

only popular query types
answers are high cost and high maintenance
Prevalence of Uncommon Searches

No answers for many information needs

- molasses substitutes
- increase volume windows xp
- dissolvable stitches speed
- dog body temperature
- CHI 2014 deadline

...
molasses substitutes
Molasses Substitute Recipe
frugalliving.about.com/od/.../qt/Molasses_Sub.htm
Note: These substitutions may alter the taste of your recipe a bit. If the molasses flavor is vital to the success of your recipe, try the brown sugar substitute.

Molasses Substitutions, Measures, Tips and Cooking Hints
homecooking.about.com/od/specificfood/a/molassesstips.htm
Tail Answers
Direct results for queries in the long tail

molasses substitutes

Substitute for molasses
Replace one cup of molasses with one of the following: 1 cup dark corn syrup, honey or maple syrup; 3/4 cup firmly packed brown sugar or 3/4 cup granulated sugar, plus 1/4 cup water.
Source: http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm

Molasses Substitute Recipe
frugalliving.about.com/od/.../qt/Molasses_Sub.htm
Note: These substitutions may alter the taste of your recipe a bit. If the molasses flavor is vital to the success of your recipe, try the brown sugar substitute.

Molasses Substitutions, Measures, Tips and Cooking Hints
homecooking.about.com/od/specificfood/a/molassestips.htm
The Long Tail of Answers

Information needs

# occurrences

weather movies

chi 2017 location
The Long Tail of Answers

# occurrences

- weather
- movies

Tail Answers

Information needs

chi 2017 location
The Long Tail of Answers

Scenarios where structured information is hard to find
Not enough query volume for dedicated teams
Green Apple Calories
There are approximately 35 calories in a green apple.
Source: http://www.livestrong.com/thedailyplate/nutrition-

Inventor of First Light Bulb
The first electric light was made in 1800 by Humphry Davy, an English scientist. He experimented with electricity and invented an electric battery. When he connected wires to his battery and a piece of carbon, the carbon glowed, producing light. This is called an electric arc.
Source: http://www.enchantedlearning.com/inventors/edison/lightbulb.shtml

Substitute for molasses
Replace one cup of molasses with one of the following: 1 cup dark corn syrup, honey or maple syrup; 3/4 cup firmly packed brown sugar or 3/4 cup granulated sugar, plus 1/4 cup water.
Source: http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm

Disolvable Stitches
It typically takes at minimum one week for the suture to dissolve, i.e. be absorbed by the body.
Source: http://answers.yahoo.com/question/index?

How to Mute Audio on Windows Movie Maker
On the Audio or Audio/Music track of the timeline, click the audio clip that you want to mute. To select multiple clips, press and hold down the CTRL key as you click clips. Click Clip, point to Audio, and then click Mute.
Source: http://windows.microsoft.com/en-US/windows-vista/Adjusting-audio-

IRS Milage
The IRS allows reimbursement for business miles driven at a rate of for 51 cents per mile.
Source: http://www.irs.gov/newsroom/article/0,,id=232017,00.html

How to Turn Up Volume on Your Computer
Start> All Programs> Accessories> Entertainment> Volume Control> Wave Setting. Increase it and the Volume should go higher.
Source: http://answers.yahoo.com/question/index?

Fish Frying Temperature
350 degrees for 3 minutes is the ticket! Also, make sure to put just enough fillets in the basket to cover the bottom of it.

Area Code 407
Area code 407 is the area code for the Orlando metro area including all of Orange, Osceola, and Seminole counties, as well as small portions of Volusia and Lake counties.

Ireland Currency
Euro (EUR)
Source: http://www.greenwichmeantime.com/time-zone/europe/european-

New York City Sales Tax 2010
New York City sales tax rate is 8.875%
Source: http://ny.rand.org/stats/govtfin/salestax.html
Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
Tail Answers Pipeline

1 Identify answer candidates

2 Filter candidates

3 Extract answer content
Identify Answer Candidates

Crowd data: 75 million search sessions

All information needs are answer candidates: queries leading to a clickthrough on a single URL

- force quit mac
- force quit on macs
- how to force quit mac
Example Answer Candidates

- force quit mac
- force quit on macs
- how to force quit mac
- 410 area code
- area code 410 location

URL1

URL2
Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
Filtering Answer Candidates

- Focus on fact-finding information needs [Kellar 2007]

radio

pandora

pandora radio log in

pandora.com
Filtering Answer Candidates

- Three filters remove answer candidates that do not address fact-finding information needs:
  - Navigation behavior
    - Pages addressing search needs
  - Query behavior
    - Unambiguous needs
  - Answer type
    - Succinct answers
Filter by Navigation Behavior

- Destination Probability for URL:
  \[ P(\text{session length} = 2 \mid \text{URL in trail}) \]
- Probability of ending session at URL after clicking through from the search results
Filter by Navigation Behavior

- Destination Probability for URL:
  \[ P(\text{session length} = 2 \mid \text{URL in trail}) \]
- Probability of ending session at URL after clicking through from the search results
Filter by Navigation Behavior

- Destination Probability for \( \text{URL} \):
  \[ P(\text{session length} = 2 \mid \text{URL in trail}) \]
- Probability of ending session at \( \text{URL} \) after clicking through from the search results

Diagram:
- Query → URL1
- Query → URL1
- Query → URL1 → URL2 → URL4
  - URL1 destination probability = 0.5
- Query → URL1 → URL3
Filter by Navigation Behavior

• Destination Probability Filter: URLs with low probability that searchers will end their session *(Lots of back navigations, later clicks)*

• Focus on queries where searchers addressed an information need
Filter by Query Behavior

• What answers are these searchers looking for?

- dissolvable stitches
  (how long they last?)
  (what they’re made of?)

- 732 area code
  (city and state?)
  (count of active phone numbers?)
Filter by Query Behavior

- A minority of searchers use question words:
  - how long dissolvable stitches last
  - where is 732 area code

- Filter candidates with fewer than 1% of clickthroughs from question queries
Filter by Answer Type

- Can a concise answer address this need?
- Ask paid crowdsourcing workers to select:
  - **Short:** phrase or sentence
    - “The optimal fish frying temperature is 350°F.”
  - **List:** small set of directions or alternatives
    - “To change your password over Remote Desktop: 1) Click on Start > Windows Security. 2) Click the Change Password button. [...]”
  - **Summary:** synthesize large amount of content
    - Impact of Budget Cuts on Teachers
Filter by Answer Type

- Can a concise answer address this need?
- Ask paid crowdsourcing workers to select:
  - **Short: phrase or sentence**
    “The optimal fish frying temperature is 350°F.”
  - **List: small set of directions or alternatives**
    “To change your password over Remote Desktop: 1) Click on Start > Windows Security. 2) Click the Change Password button. [...]”
  - **Summary: synthesize large amount of content**
    Impact of Budget Cuts on Teachers
Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
Extracting the Tail Answer

- We now have answer candidates with:
  - Factual responses
  - Succinct responses
- However, the answer is buried:

  dissolvable stitches
dissolvable stitches how long
dissolvable stitches absorption
Crowdsourcing Workflow

- Reliably extract relevant answer from the URL via paid crowdsourcing (CrowdFlower)

[Bernstein et al. 2010, Little et al. 2010]
Quality Challenge: Overgenerating

• Typical extraction length:

Quality Challenge: Overgenerating

- Extraction length with regex-based gold:

Tail Answers Pipeline

1. Identify answer candidates
2. Filter candidates
3. Extract answer content
<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventor of First Light Bulb</td>
<td>The first electric light was made in 1800 by Humphry Davy, an English scientist. He experimented with electricity and invented an electric battery. When he connected wires to his battery and a piece of carbon, the carbon glowed, producing light. This is called an electric arc.</td>
<td><a href="http://www.enchantedlearning.com/inventors/edison/lightbulb.shtml">http://www.enchantedlearning.com/inventors/edison/lightbulb.shtml</a>&lt;br&gt;Source: <a href="http://www.enchantedlearning.com/inventors/edison/lightbulb.shtml">http://www.enchantedlearning.com/inventors/edison/lightbulb.shtml</a></td>
</tr>
<tr>
<td>Substitute for molasses</td>
<td>Replace one cup of molasses with one of the following: 1 cup dark corn syrup, honey or maple syrup; 3/4 cup firmly packed brown sugar or 3/4 cup granulated sugar, plus 1/4 cup water.</td>
<td><a href="http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm">http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm</a>&lt;br&gt;Source: <a href="http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm">http://frugalliving.about.com/od/makeyour/qt/Molasses/Sub.htm</a></td>
</tr>
<tr>
<td>Dissolvable Stitches</td>
<td>It typically takes at minimum one week for the suture to dissolve, i.e. be absorbed by the body.</td>
<td><a href="http://answers.yahoo.com/question/index?qid=20060822050444AASZ9L">http://answers.yahoo.com/question/index?qid=20060822050444AASZ9L</a>&lt;br&gt;Source: <a href="http://answers.yahoo.com/question/index?qid=20060822050444AASZ9L">http://answers.yahoo.com/question/index?qid=20060822050444AASZ9L</a></td>
</tr>
</tbody>
</table>
Evaluation: Answer Quality

- Hand-coded for correctness (two–three redundant coders)
- 87% of Tail Answers were completely correct or had a minor error (e.g., title != content)
- Field experiment manipulating answer presence and ranking quality:
  - Tail Answers impact subjective ratings half as much as good ranking, and fully compensate for poor results.
Ongoing Challenges

- Crowd now has lots of editorial control over the search engine
Ongoing Challenges

• Crowd now has lots of editorial control over the search engine

Sea of China

Should actually be called the Sea of Japan

Ongoing Challenges

• Crowd now has lots of editorial control over the search engine

<table>
<thead>
<tr>
<th>Sea of China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should actually be called the Sea of Japan</td>
</tr>
</tbody>
</table>

Cannibalizing pageviews from the original pages
Extension: Better Result Snippets

- Improve result pages for popular queries
Extension: Better Result Snippets

- Improve result pages for popular queries

Automatically extracted

[Link to Boston Wallpaper Removal Service Reviews](http://www.angieslist.com/companylist/boston/wallpaper.htm)

Service Area: Entire Area Except Attleboro-taunton, Boxford-gloucester, Cohasset & Worcester Counties
Extension: Better Result Snippets

- Improve result pages for popular queries

Automatically extracted

**Boston Wallpaper Removal Service Reviews**
Service Area: Entire Area Except Attleboro-taunton, Boxford-gloucester, Cohasset & Worcester Counties
[www.angieslist.com/companylist/boston/wallpaper.htm](http://www.angieslist.com/companylist/boston/wallpaper.htm)

Crowd-authored

**Boston Wallpaper Removal Service Reviews**
There are members who sign up and share experiences with each other so that the user can choose the service company that's right for their job the first time around.
[www.angieslist.com/companylist/boston/wallpaper.htm](http://www.angieslist.com/companylist/boston/wallpaper.htm)
Extension: Domain-Specific Answers

- Design for specific information needs

```php
$a = "Hello ";
$b = $a . "World!";
```
Extension: Domain-Specific Answers
• Design for specific information needs

PHP: String Concatenation
```php
$a = "Hello ";
$b = $a . "World!";
```

Array Joining: PHP to Python
```php
join(" ", array('do', 're', 'mi'));
``` →
```python
's '.join(['do', 're', 'mi'])
```
Extension: Domain-Specific Answers

- Design for specific information needs

<table>
<thead>
<tr>
<th>PHP: String Concatenation</th>
<th>Python String Concatenation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>$a = &quot;Hello &quot;;</code></td>
<td><code>&quot;Hello &quot;;</code></td>
</tr>
<tr>
<td><code>$b = $a . &quot;World!&quot;;</code></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Array Joining: PHP to Python</th>
<th>Array Joining: Python to PHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHP</td>
<td>Python</td>
</tr>
<tr>
<td><code>join(&quot; &quot;, array('do', 're', 'mi'));</code></td>
<td><code>' '.join(['do', 're', 'mi'])</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Array Slicing: Python to PHP</th>
<th>Array Slicing: PHP to Python</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>PHP</td>
</tr>
<tr>
<td><code>a[2:4]</code></td>
<td><code>array_slice($a, 2, 2);</code></td>
</tr>
</tbody>
</table>
http://hci.stanford.edu/msb