



# SEARCHING: FAST AND SLOW

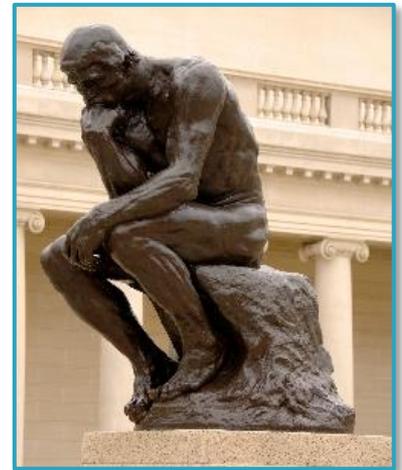
Susan Dumais

<http://research.microsoft.com/~sdumais>

#TAIA2014 Jul 11, 2014

# Searching: Fast and Slow

- Tremendous engineering effort aimed at making search fast
  - ▣ ... and for good reason
  - ▣ But, many compromises made to achieve speed
- Not all searches need to be fast
- How can we use additional time to improve search quality?



# Speed Focus in Search Important

- Schurman & Brutlag, Velocity 2009  
(Arapakis, Bai & Cambazoglu, SIGIR 2014)
- A/B tests increasing page load time (at server)
- Increasing page load time by as little 100 msecs influences search experience substantially
  - ▣ Decreased searches per user, clicks, and revenue
  - ▣ Increased abandonment, and time to click
- Effects are larger with longer latency and persist after delays are removed

# Schurman (Bing)

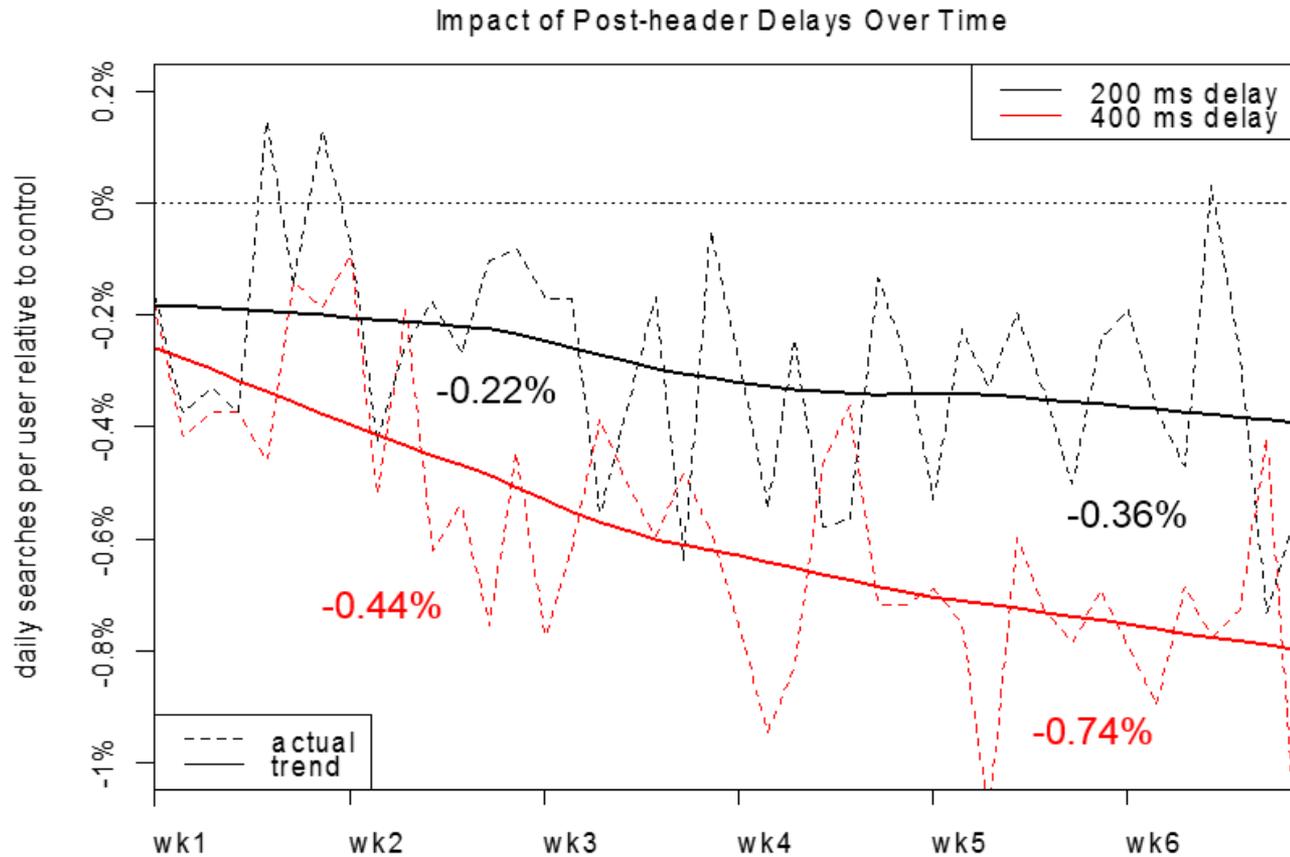
## Server Delays Experiment: Results

	Distinct Queries/User	Query Refinement	Revenue/User	Any Clicks	Satisfaction	Time to Click (increase in ms)
50ms	-	-	-	-	-	-
200ms	-	-	-	-0.3%	-0.4%	500
500ms	-	-0.6%	-1.2%	-1.0%	-0.9%	1200
1000ms	-0.7%	-0.9%	-2.8%	-1.9%	-1.6%	1900
2000ms	-1.8%	-2.1%	-4.3%	-4.4%	-3.8%	3100

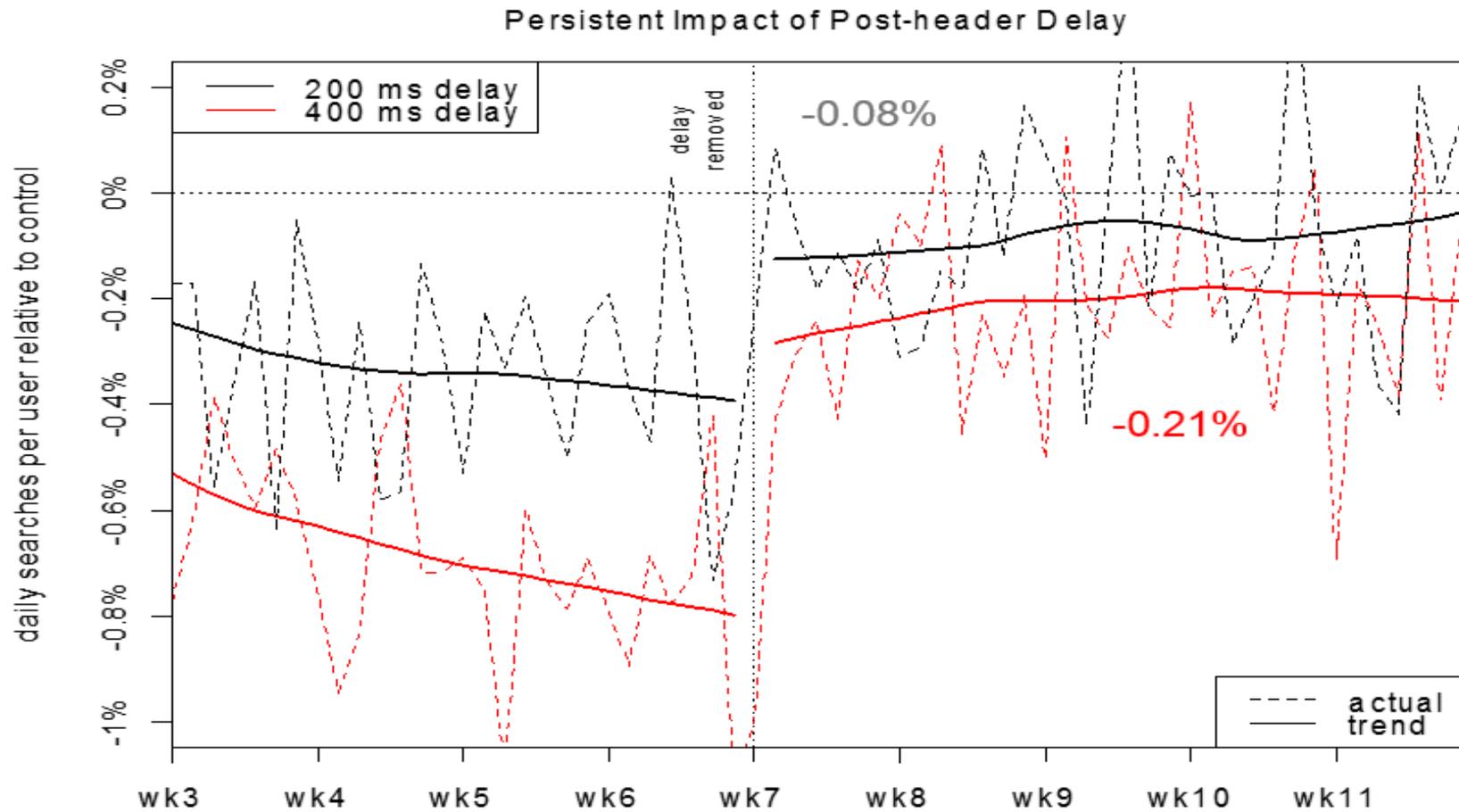
- Means no statistically significant change

- Strong negative impacts
- Roughly linear changes with increasing delay
- Time to Click changed by roughly double the delay

# Brutlag (Google)

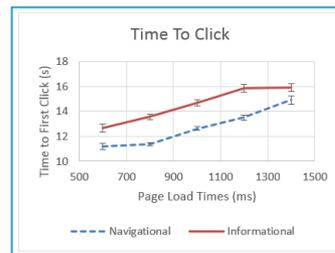


# Brutlag (Google)



# Speed Focus in Search Important

- Teevan et al., HCIR 2013
- Examined naturally occurring variation in page load time (for same query), from 500-1500 msec
  - ▣ Longer load time associated with increases in
    - Abandonment rate increased (from 20% to 25%)
    - Time first to click increased (from 1.2 to 1.6 secs)
  - ▣ Larger effects on navigational (vs. informational) queries



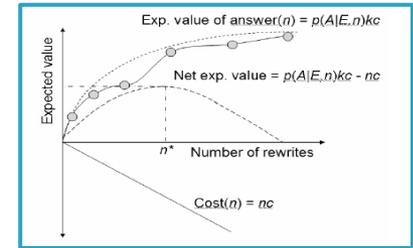
# Not All Searches Need to Be Fast

- Complex information needs
  - ▣ Long search sessions
  - ▣ Cross-session tasks
- Social search
  - ▣ Question asking
- Technology limits
  - ▣ Mobile devices
  - ▣ Limited connectivity
  - ▣ Search from Mars



# Improving Search with More Time

- By the second
  - ▣ Use richer query and document analysis
  - ▣ Issue additional queries



- By the minute
  - ▣ Include humans in the loop, e.g., to generate “answers”



- By the hour
  - ▣ Create new search artifacts
  - ▣ Enable new search experiences



- Relaxing time constraints creates interesting new opportunities for “search”

# By the Second



- Use richer query and document analysis
- Issue additional queries
- Find additional answers on “quick back”
- ...
  
- Especially helpful for
  - ▣ Difficult queries
  - ▣ Long sessions, whether struggling or exploring

# Question Answering

## □ AskMSR question answering system

### □ Re-write query in declarative form

#### ■ E.g., “Who is Bill Gates married to?”

■ “Bill Gates +is married +to” <>

■ <> “+is married +to Bill Gates”

■ “Bill Gates” AND “married to”

■ “Bill” AND “Gates” AND “married”

### □ Mine n-grams from snippets, exploiting redundancy

### □ Are multiple queries worth the cost?



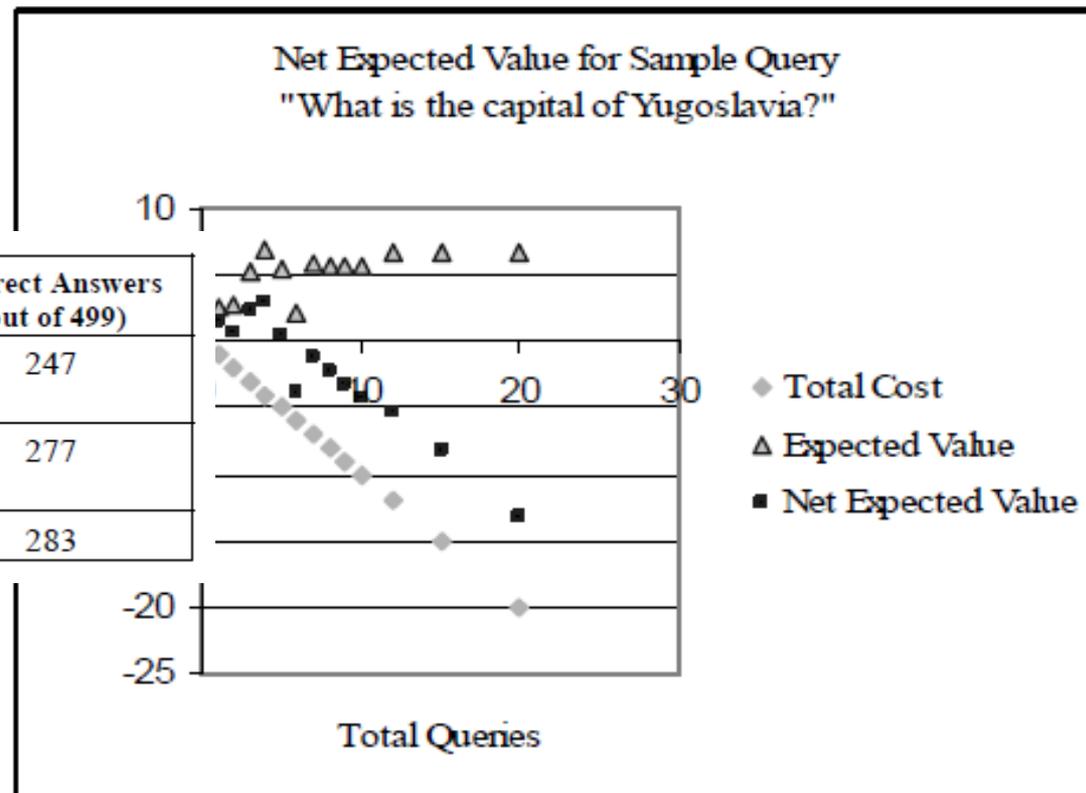
1. Melinda French	53%
2. Microsoft Corp	16%
3. Mimi Gardner	8%

# Decision-Theoretic QA

- Order query rewrites by their importance
- Assess cost and benefit of additional queries

- Aggregate results

Rewrite Policy	Cost	Correct Answers (out of 499)
Conjunctural rewrites only	499	247
Cost-benefit $k=10, c=1$	1179	277
All rewrites	3426	283



# By the Minute



- Use slower resources (like people)
- Can be used to augment many components of the search process
  - ▣ Understanding the query
  - ▣ Finding (or generating) better results
  - ▣ Understanding (or organizing) results

# People Can Provide Rich Input



- Study: Complex restaurant queries to Yelp
- People used to
  - ▣ Support deeper understand of the query
  - ▣ Organize results in a new way

# Understand Query: Identify Entities

- Search engines do poorly with long, complex queries
- Query: *Italian* restaurant in *Squirrel Hill or Greenfield* with a *gluten-free* menu and a *fairly sophisticated atmosphere*
- Crowd workers identify important attributes
  - ▣ Given list of potential attributes
  - ▣ Option add new attributes
  - ▣ Example: *cuisine*, *location*, *special diet*, *atmosphere*
- Crowd workers match attributes to query
- Attributes used to issue a structured search (to Yelp)

# Understand Results: Tabulate

- Crowd workers tabulate search results
  - ▣ Given a query, result, attribute, and value
  - ▣ Does the result meet the attribute?

**centric restaurant in manhattan, close to the empire state building.  
Argentine or Mexican food. Quiet place. With reservation.**

	Takes Reservations	Type of Cuisine	Location
<a href="#">Azul Bistro - New York, NY</a>	✓	✓	✓
<a href="#">El Gauchito - Elmhurst, NY</a>	✓	✓	✓
<a href="#">El Almacen - Brooklyn, NY</a>	✓	✓	✓
<a href="#">Pampas Argentinas - Forest Hills, NY</a>	✓	✓	✓
<a href="#">Sosa Borella - New York, NY</a>	✓	✓	✓
<a href="#">Libertador - New York, NY</a>	✓	✓	✓
<a href="#">Mexico Lindo Restaurant - New York, NY</a>	✓	✓	✓
<a href="#">El Mariachi Restaurant - Astoria, NY</a>	?	✓	✓
<a href="#">Nuchas - New York, NY</a>	✗	✓	✓
<a href="#">Empanadas Bar NYC - New York, NY</a>	✗	✓	✓

# People Can Generate New Content

## □ Bing Answers

**Weather in Bellevue, Washington**  
bing.com/weather - Data from weather.com®  
[Hourly](#) · [Monthly Averages](#) · [Radar & Maps](#)

Now	Wed	Thu	Fri	Sat	Sun	Next 5 days
46° Cloudy						
°F   °C 55° / 37°	59° / 41°	64° / 45°	73° / 48°	75° / 50°		

Change providers: [AccuWeather](#) · [Weather Underground](#) · [Compare all](#)

**News about Dog food recall**  
bing.com/news

**Pet food recall 2012 now includes 39 states**  
The Food and Drug Administration has expanded a national recall of Diamond Pet Foods dog food manufactured between Dec. 9, 2011, and April 7, 2012, from the...  
The Christian Science Monitor · 16 hours ago  
[UPDATE: Pet-food recall expanded](#) · [Seattle Times](#)  
[Pet food store adjusts to Diamond recalls](#) · [Examiner](#)  
See also: [More stories](#) · [Top stories](#) · [Related blogs](#)

## □ “Tail” Answers

**Currency Converter**  
investing.money.msn.com/investments/currency-converter-calculator

11/7/2011 5/9/2012 1 US Dollar To Chinese Yuan Convert

6.3  
6.2

1 US Dollar = 6.305965 Chinese Yuan  
1 CNY = 0.15858001 USD

**Restaurants near Bellevue, Washington**  
bing.com/local

1. Seastar Restaurant & Raw Bar · Website · (425) 456-0010 · 205 108th Ave NE Ste 100 · Bellevue · Directions · Menu
2. Gilbert's Main Street Bagel · (425) 455-5650 · 10024 Main St · Bellevue · Directions · Menu
3. Daniel's Broiler Steakhouse R... · Website · (425) 462-4662 · 10500 NE 8th St Ste 2100 · Bellevue · Directions · Menu

bing molasses substitute

bing dissolvable stitches speed

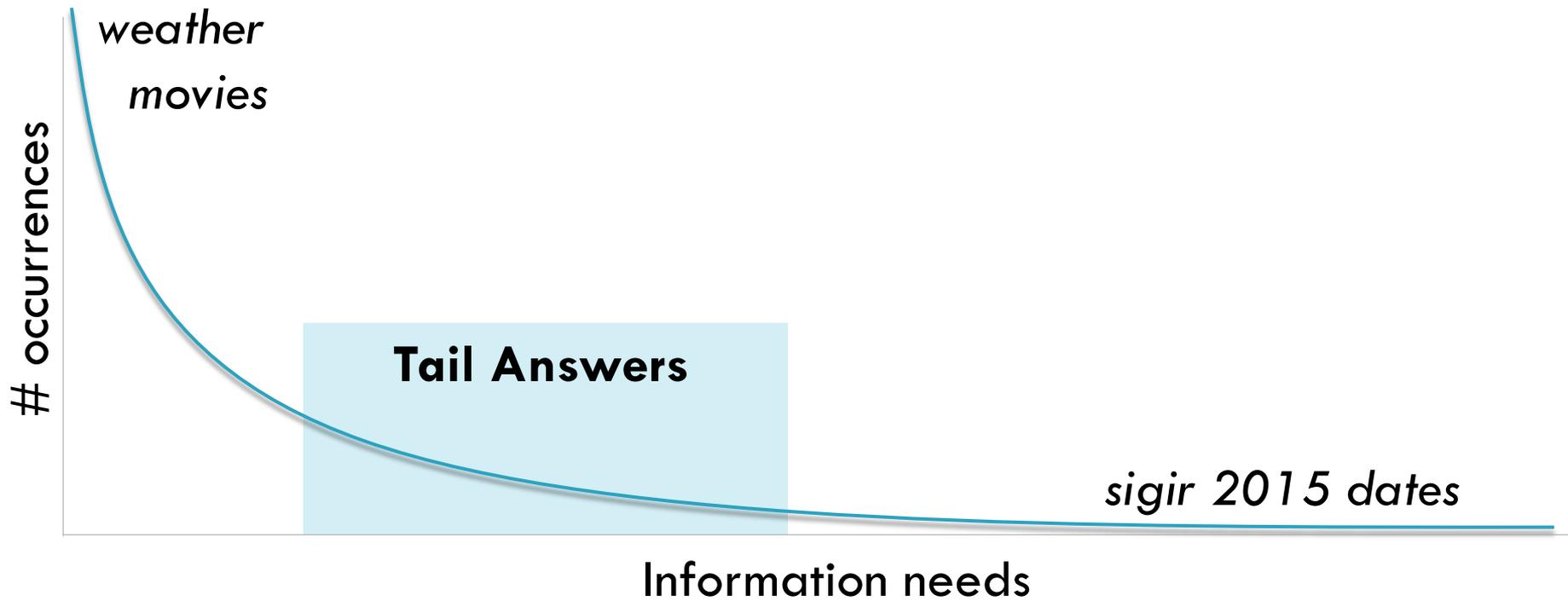
bing increase sound volume windows

bing temperature for frying fish

bing calories green apple

bing dog normal body temperature

# The Long Tail of Answers



Hard to find structured information

Not enough query volume for dedicated teams

# Tail Answers Pipeline

## 1. Identify Answer Candidates (logs)

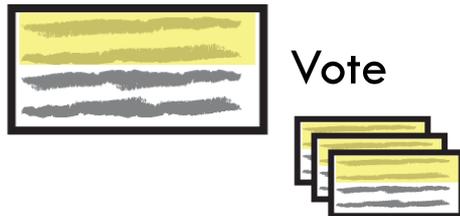
Search trails that lead to same URL     $\text{query}_1 \dots \text{query}_n \rightarrow \boxed{\text{URL}}$

## 2. Filter Candidates (crowd-powered)

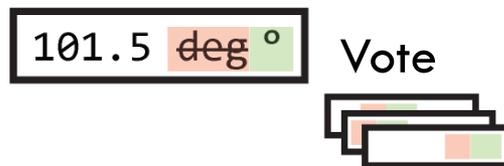
Navigational behavior    Unambiguous needs    Succinct answers

## 3. Generate Answers (crowd-powered)

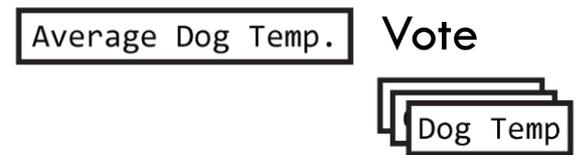
Extract



Proofread



Title



# Tail Answers Results

- Quality: 87% had no errors
- Time: minutes
- Cost: 44¢ to create answer
- Expt: result quality x presence of “tail answer”
- Tail Answers
  - ▣ Change subjective ratings half as much as good ranking
  - ▣ Fully compensate for poor rankings

## 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.

Source: <http://www.walleyecentral.com/forums/showthread.php?t=146552>

## 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.

Source: [http://en.wikipedia.org/wiki/Area\\_code\\_407](http://en.wikipedia.org/wiki/Area_code_407)

## Ireland Currency

Euro (EUR)

Source: <http://wwp.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>

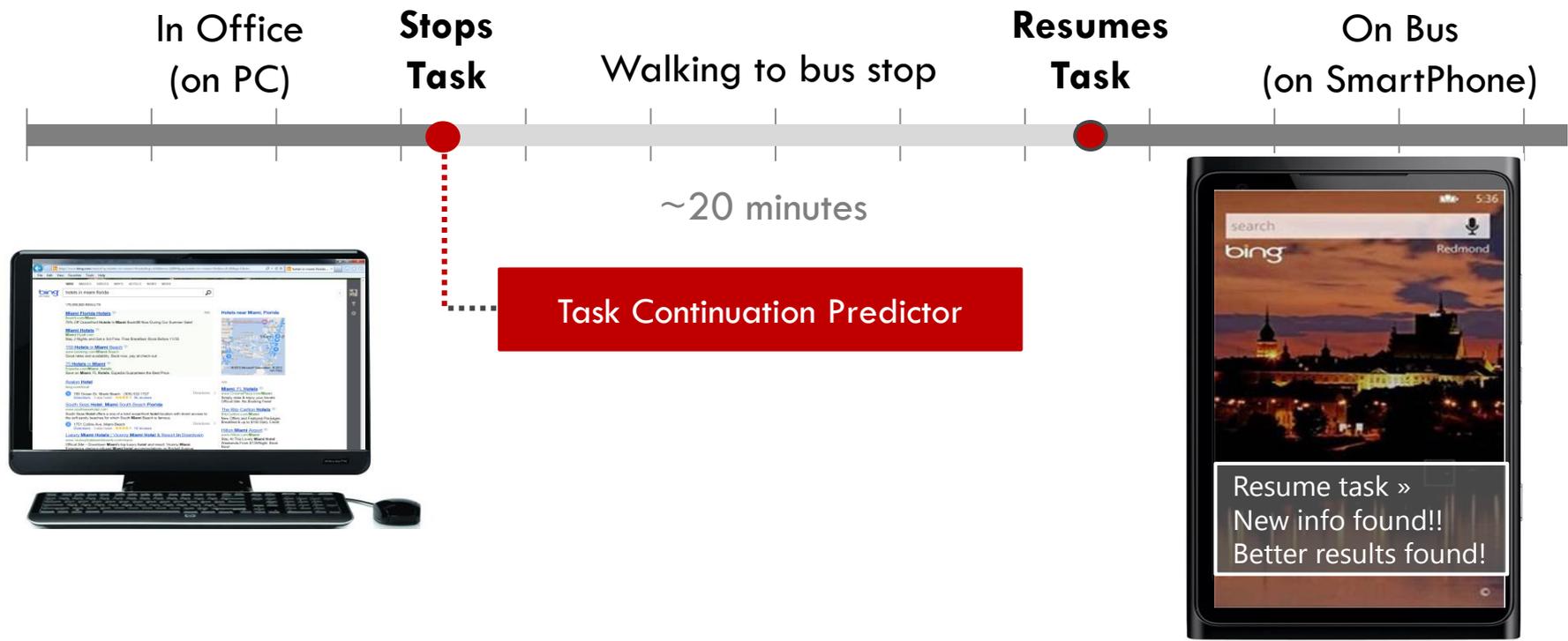
# By the Hour



- We can create new “search” experiences
- Support ongoing tasks
  - ▣ Task resumption, across sessions or devices
  - ▣ Reinststate context, generate summaries, highlight change
- Proactively retrieve information of interest
- Asynchronously answer search requests
  - ▣ Dinner reservations for tonight
  - ▣ Background material by morning

# Support Task Resumption

- 10-15% of tasks continue across sessions
- Predict which tasks will be resumed at a later time
- Reinststate and enrich context



# Searching: Fast and Slow

- Relaxing time constraints creates interesting opportunities to change “search” as we know it
- Especially useful for
  - ▣ complex information needs that extend over time
  - ▣ richer understanding and presentation of information
- Allows us to think about solutions that
  - ▣ support differential computation (e.g., CiteSight)
  - ▣ combine human and algorithmic components (e.g., TailAnswers, VizWiz)
- Requires that we break out of the search box

# Thank You !



- Questions/Comments ???
- More info, <http://research.microsoft.com/~sdumais>

# Further Reading

## □ The need for speed

- Schurman, E. and Brutlag, J. *Performance related changes and their user impact*. Velocity 2009 Conference.
- Arapakis, I., Shi, X. and Cambazoglu, B. *Impact of response latency on user behavior in web search*. SIGIR 2014.

## □ Slow search

- Teevan, J., Collins-Thompson, K., White, R., Dumais, S.T. and Kim, Y. *Slow search: Information retrieval without time constraints*. HCIR 2013.
- Azari, D., Horvitz, E., Dumais, S.T. and Brill, E. *Actions, answers and uncertainty: A decision-making perspective on web question answering*. IPM 2004.
- Lee, C-J., Teevan, J. and de la Chica, S. *Characterizing multi-click search behavior and the risks and opportunities of changing results during use*. SIGIR 2014.
- Bernstein, M., Teevan, J., Dumais, S.T., Libeling, D. and Horvitz, E. *Direct answers for search queries in the long tail*. CHI 2012.
- Wang, Y., Huang, X. and White, R. *Characterizing and supporting cross-device search tasks*. WSDM 2013.