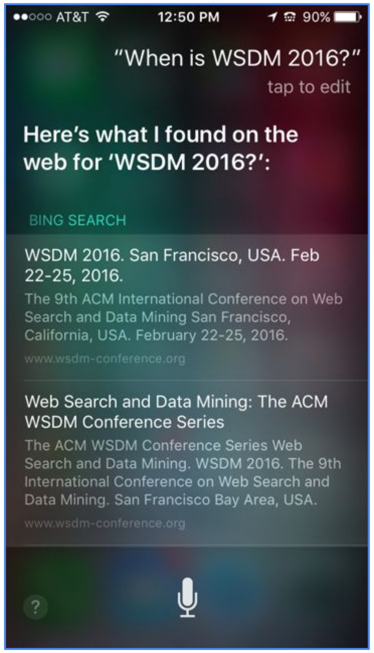
# CRQA: Crowd-powered Real-time Automated Question Answering System

Denis Savenkov
Emory University
dsavenk@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

HCOMP, Austin, TX
October 31, 2016

# Volume of question search queries is growing[1]



"When is WSDM 2016?"

Here's what I found on the web for 'WSDM 2016?':

WSDM 2016. San Francisco, USA. Feb 22-25, 2016.

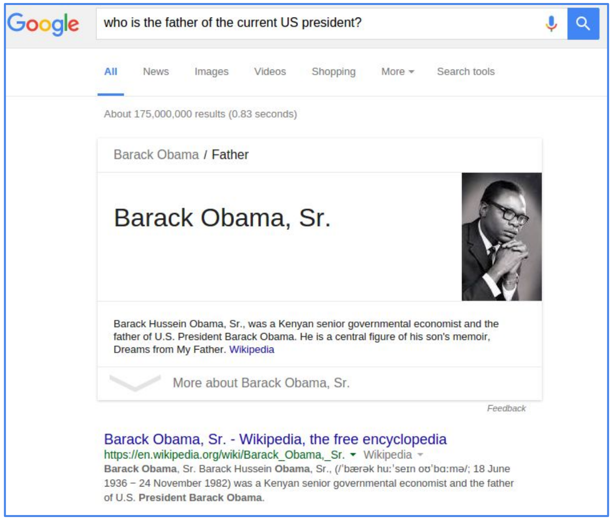
The 9th ACM International Conference on Web Search and Data Mining San Francisco, California, USA. February 22-25, 2016.

www.wsdm-conference.org

Web Search and Data Mining: The ACM WSDM Conference Series

The ACM WSDM Conference Series Web Search and Data Mining. WSDM 2016. The 9th International Conference on Web Search and Data Mining. San Francisco Bay Area, USA.

www.wsdm-conference.org



who is the father of the current US president?

About 175,000,000 results (0.83 seconds)

Barack Obama / Father

Barack Obama, Sr.

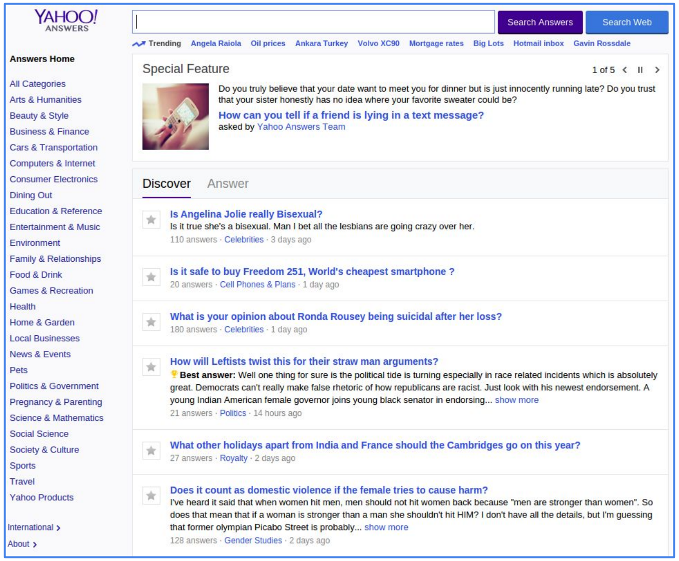
Barack Hussein Obama, Sr., was a Kenyan senior governmental economist and the father of U.S. President Barack Obama. He is a central figure of his son's memoir, Dreams from My Father. Wikipedia

More about Barack Obama, Sr.

Barack Obama, Sr. - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Barack_Obama,_Sr. ▾ Wikipedia ▾

**Barack Obama**, Sr. Barack Hussein **Obama**, Sr., (/ˈbærək huːˈseɪn oʊˈbɑːmə/; 18 June 1936 – 24 November 1982) was a Kenyan senior governmental economist and the father of U.S. **President Barack Obama**.



Trending Angela Raiola Oil prices Ankara Turkey Volvo XC90 Mortgage rates Big Lots Hotmail inbox Gavin Rossdale

**Answers Home**

All Categories
Arts & Humanities
Beauty & Style
Business & Finance
Cars & Transportation
Computers & Internet
Consumer Electronics
Dining Out
Education & Reference
Entertainment & Music
Environment
Family & Relationships
Food & Drink
Games & Recreation
Health
Home & Garden
Local Businesses
News & Events
Pets
Politics & Government
Pregnancy & Parenting
Science & Mathematics
Social Science
Society & Culture
Sports
Travel
Yahoo Products

International ›
About ›

Special Feature

Do you truly believe that your date want to meet you for dinner but is just innocently running late? Do you trust that your sister honestly has no idea where your favorite sweater could be?

**How can you tell if a friend is lying in a text message?**
asked by Yahoo Answers Team

**Is Angelina Jolie really Bisexual?**
Is it true she's a bisexual. Man I bet all the lesbians are going crazy over her.
110 answers · Celebrities · 3 days ago

**Is it safe to buy Freedom 251, World's cheapest smartphone ?**
20 answers · Cell Phones & Plans · 1 day ago

**What is your opinion about Ronda Rousey being suicidal after her loss?**
180 answers · Celebrities · 1 day ago

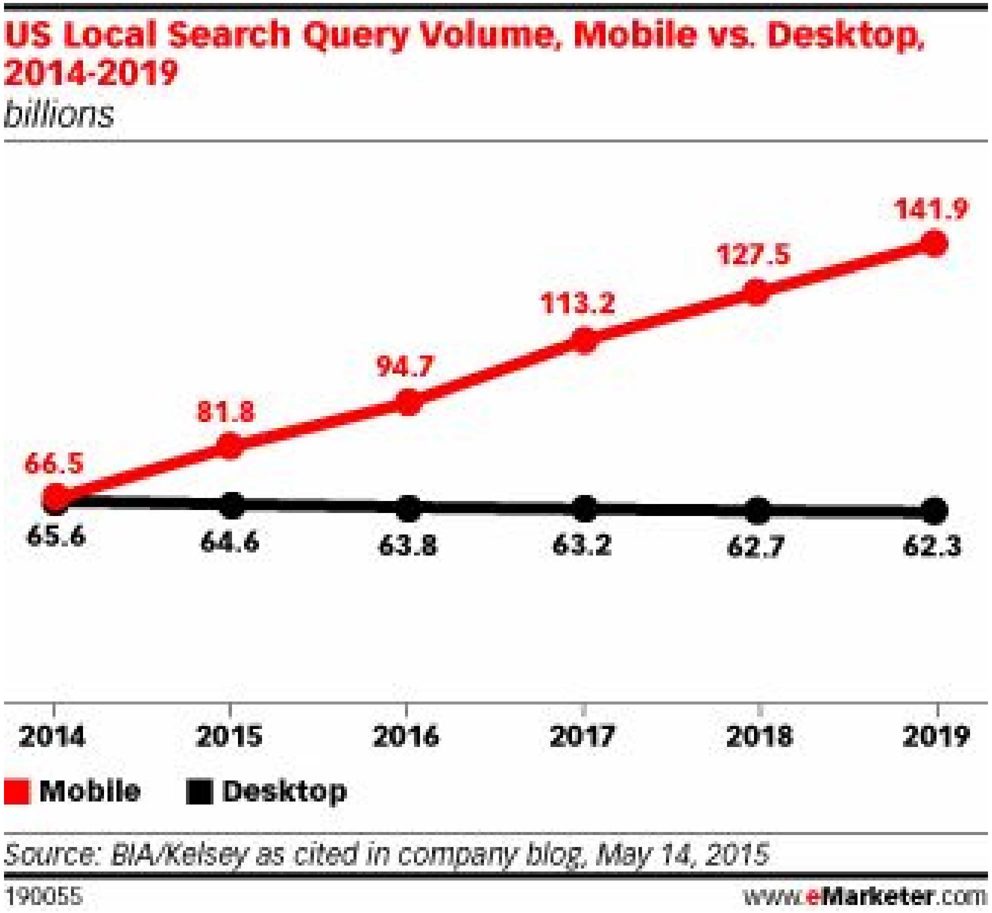
**How will Leftists twist this for their straw man arguments?**
**Best answer:** Well one thing for sure is the political tide is turning especially in race related incidents which is absolutely great. Democrats can't really make false rhetoric of how republicans are racist. Just look with his newest endorsement. A young Indian American female governor joins young black senator in endorsing... show more
21 answers · Politics · 14 hours ago

**What other holidays apart from India and France should the Cambridges go on this year?**
27 answers · Royalty · 2 days ago

**Does it count as domestic violence if the female tries to cause harm?**
I've heard it said that when women hit men, men should not hit women back because "men are stronger than women". So does that mean that if a woman is stronger than a man she shouldn't hit HIM? I don't have all the details, but I'm guessing that former olympian Picabo Street is probably... show more
128 answers · Gender Studies · 2 days ago

[1] "Questions vs. Queries in Informational Search Tasks", Ryen W. White et al, WWW 2015

# And more and more of this searches are happening on mobile

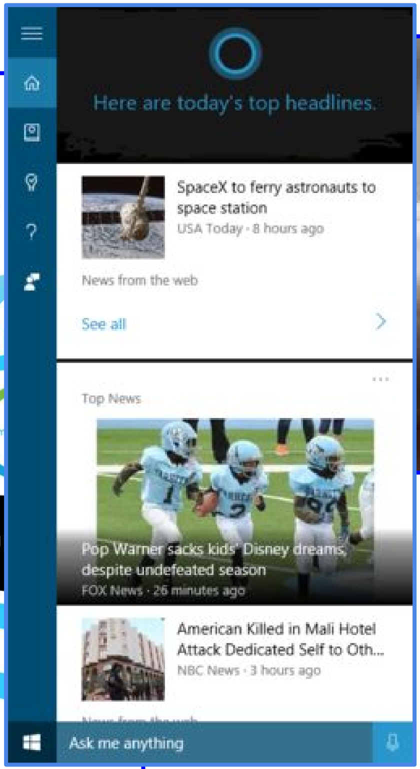


**US Local Search Query Volume, Mobile vs. Desktop, 2014-2019**
*billions*

*Source: BIA/Kelsey as cited in company blog, May 14, 2015*

190055 www.eMarketer.com

# Mobile Personal Assistants are popular


Here are today's top headlines.

SpaceX to ferry astronauts to space station
USA Today · 8 hours ago

News from the web

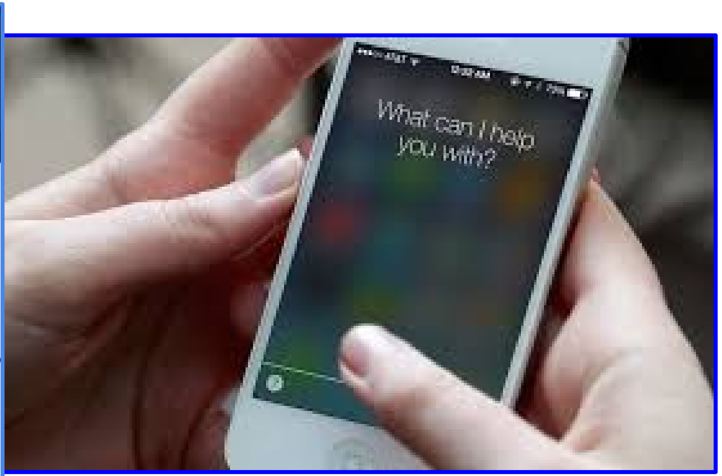
See all

Top News

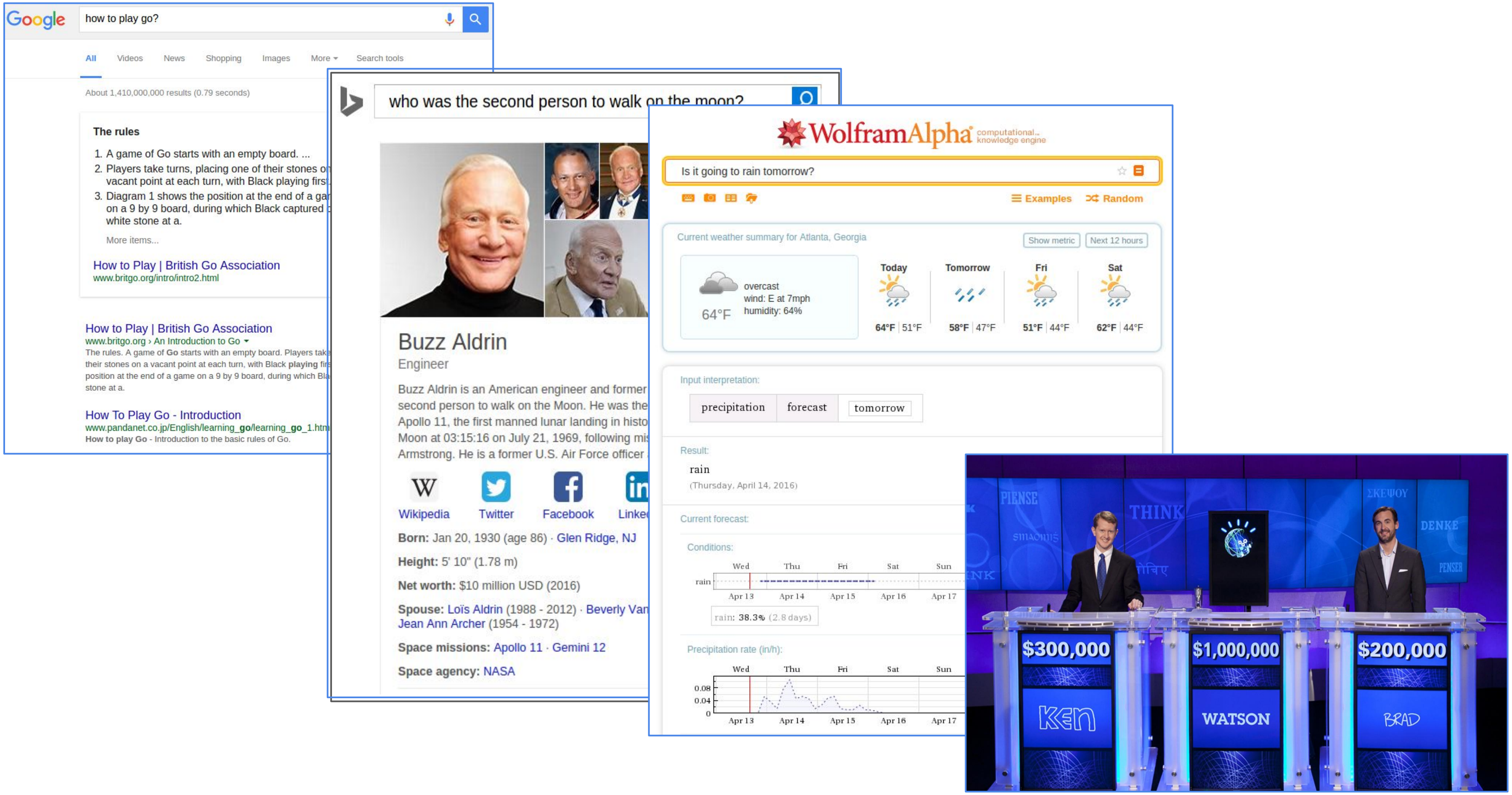
Pop Warner sacks kids' Disney dreams, despite undefeated season
FOX News · 26 minutes ago

American Killed in Mali Hotel Attack Dedicated Self to Oth...
NBC News · 3 hours ago

Ask me anything



# Automatic Question Answering works relatively well for some questions



Google

how to play go?

All Videos News Shopping Images More Search tools

About 1,410,000,000 results (0.79 seconds)

**The rules**

1. A game of Go starts with an empty board. ...
2. Players take turns, placing one of their stones o vacant point at each turn, with Black playing firs
3. Diagram 1 shows the position at the end of a ga on a 9 by 9 board, during which Black captured white stone at a.

More items...

How to Play | British Go Association
www.britgo.org/intro/intro2.html

How to Play | British Go Association
www.britgo.org › An Introduction to Go
The rules. A game of **Go** starts with an empty board. Players tak their stones on a vacant point at each turn, with Black **playing** fir position at the end of a game on a 9 by 9 board, during which Bla stone at a.

How To Play Go - Introduction
www.pandanet.co.jp/English/learning_**go**/learning_**go**_1.htm
**How to play Go** - Introduction to the basic rules of Go.

who was the second person to walk on the moon?

Buzz Aldrin
Engineer

Buzz Aldrin is an American engineer and former second person to walk on the Moon. He was the Apollo 11, the first manned lunar landing in histo Moon at 03:15:16 on July 21, 1969, following mis Armstrong. He is a former U.S. Air Force officer

Wikipedia Twitter Facebook Linke

**Born:** Jan 20, 1930 (age 86) · Glen Ridge, NJ

**Height:** 5' 10" (1.78 m)

**Net worth:** $10 million USD (2016)

**Spouse:** Lois Aldrin (1988 - 2012) · Beverly Van Jean Ann Archer (1954 - 1972)

**Space missions:** Apollo 11 · Gemini 12

**Space agency:** NASA

WolframAlpha computational knowledge engine

Is it going to rain tomorrow?

Examples Random

Current weather summary for Atlanta, Georgia

Show metric Next 12 hours

64°F overcast
wind: E at 7mph
humidity: 64%

| Today | Tomorrow | Fri | Sat |
|---|---|---|---|
| 64°F \| 51°F | 58°F \| 47°F | 51°F \| 44°F | 62°F \| 44°F |

Input interpretation:
precipitation | forecast | tomorrow

Result:
rain
(Thursday, April 14, 2016)

Current forecast:

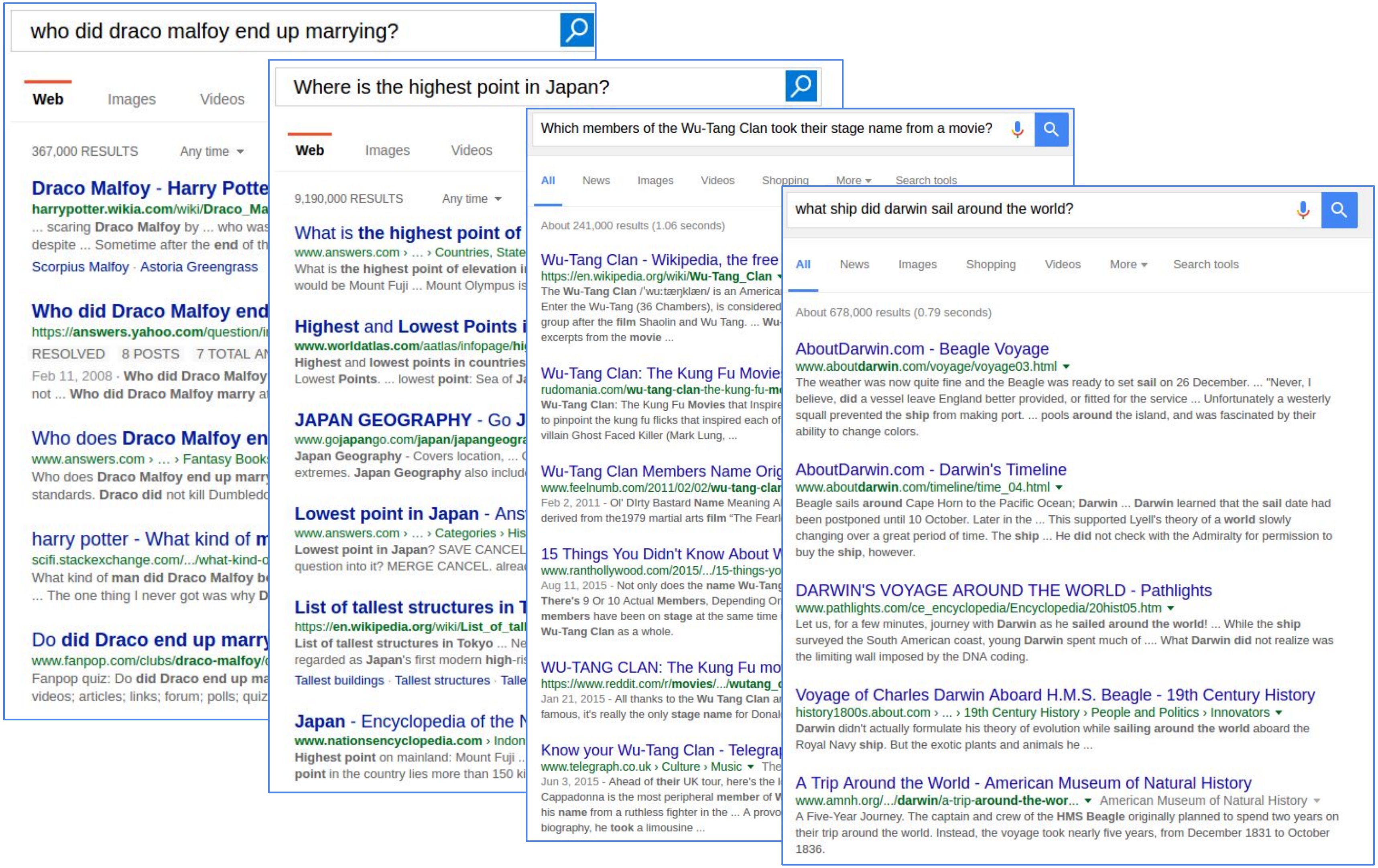
Conditions:

rain: 38.3% (2.8 days)

Precipitation rate (in/h):

(AP Photo/Jeopardy Productions, Inc.)

# … but not sufficiently well for many other questions



who did draco malfoy end up marrying?

367,000 RESULTS Any time

Draco Malfoy - Harry Potte
harrypotter.wikia.com/wiki/Draco_Ma
... scaring Draco Malfoy by ... who was
despite ... Sometime after the end of th
Scorpius Malfoy · Astoria Greengrass

Who did Draco Malfoy end
https://answers.yahoo.com/question/i
RESOLVED 8 POSTS 7 TOTAL AN
Feb 11, 2008 - Who did Draco Malfoy
not ... Who did Draco Malfoy marry a

Who does Draco Malfoy en
www.answers.com › ... › Fantasy Book
Who does Draco Malfoy end up marr
standards. Draco did not kill Dumbledo

harry potter - What kind of m
scifi.stackexchange.com/.../what-kind-o
What kind of man did Draco Malfoy be
... The one thing I never got was why D

Do did Draco end up marry
www.fanpop.com/clubs/draco-malfoy/
Fanpop quiz: Do did Draco end up ma
videos; articles; links; forum; polls; quiz

Where is the highest point in Japan?

9,190,000 RESULTS Any time

What is the highest point of
www.answers.com › ... › Countries, State
What is the highest point of elevation i
would be Mount Fuji ... Mount Olympus is

Highest and Lowest Points i
www.worldatlas.com/aatlas/infopage/hi
Highest and lowest points in countries
Lowest Points. ... lowest point: Sea of Ja

JAPAN GEOGRAPHY - Go J
www.gojapango.com/japan/japangeogra
Japan Geography - Covers location, ... C
extremes. Japan Geography also includ

Lowest point in Japan - Ans
www.answers.com › ... › Categories › His
Lowest point in Japan? SAVE CANCEL
question into it? MERGE CANCEL. alread

List of tallest structures in T
https://en.wikipedia.org/wiki/List_of_tall
List of tallest structures in Tokyo ... Ne
regarded as Japan's first modern high-ri
Tallest buildings · Tallest structures · Talle

Japan - Encyclopedia of the N
www.nationsencyclopedia.com › Indon
Highest point on mainland: Mount Fuji ...
point in the country lies more than 150 ki

Which members of the Wu-Tang Clan took their stage name from a movie?

About 241,000 results (1.06 seconds)

Wu-Tang Clan - Wikipedia, the free
https://en.wikipedia.org/wiki/Wu-Tang_Clan
The Wu-Tang Clan /ˈwuːtæŋklæn/ is an America
Enter the Wu-Tang (36 Chambers), is considered
group after the film Shaolin and Wu Tang. ... Wu-
excerpts from the movie ...

Wu-Tang Clan: The Kung Fu Movie
rudomania.com/wu-tang-clan-the-kung-fu-m
Wu-Tang Clan: The Kung Fu Movies that Inspire
to pinpoint the kung fu flicks that inspired each of
villain Ghost Faced Killer (Mark Lung, ...

Wu-Tang Clan Members Name Orig
www.feelnumb.com/2011/02/02/wu-tang-clan
Feb 2, 2011 - Ol' Dirty Bastard Name Meaning A
derived from the1979 martial arts film "The Fearl

15 Things You Didn't Know About W
www.rantholllywood.com/2015/.../15-things-yo
Aug 11, 2015 - Not only does the name Wu-Tang
There's 9 Or 10 Actual Members, Depending On
members have been on stage at the same time
Wu-Tang Clan as a whole.

WU-TANG CLAN: The Kung Fu mo
https://www.reddit.com/r/movies/.../wutang_c
Jan 21, 2015 - All thanks to the Wu Tang Clan ar
famous, it's really the only stage name for Donal

Know your Wu-Tang Clan - Telegra
www.telegraph.co.uk › Culture › Music ▾ The
Jun 3, 2015 - Ahead of their UK tour, here's the l
Cappadonna is the most peripheral member of W
his name from a ruthless fighter in the ... A provo
biography, he took a limousine ...

what ship did darwin sail around the world?

About 678,000 results (0.79 seconds)

AboutDarwin.com - Beagle Voyage
www.aboutdarwin.com/voyage/voyage03.html ▾
The weather was now quite fine and the Beagle was ready to set sail on 26 December. ... "Never, I believe, did a vessel leave England better provided, or fitted for the service ... Unfortunately a westerly squall prevented the ship from making port. ... pools around the island, and was fascinated by their ability to change colors.

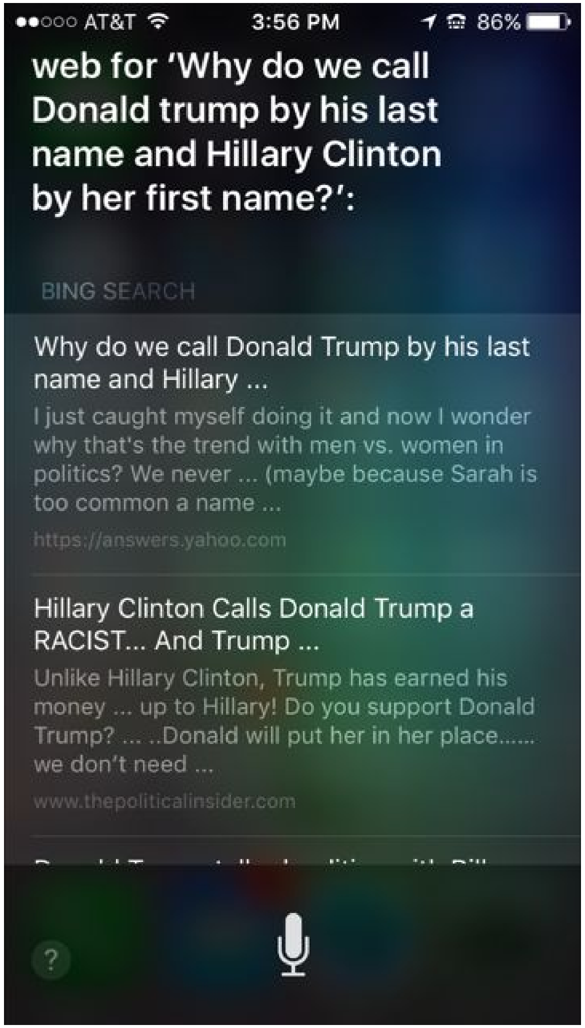
AboutDarwin.com - Darwin's Timeline
www.aboutdarwin.com/timeline/time_04.html ▾
Beagle sails around Cape Horn to the Pacific Ocean; Darwin ... Darwin learned that the sail date had been postponed until 10 October. Later in the ... This supported Lyell's theory of a world slowly changing over a great period of time. The ship ... He did not check with the Admiralty for permission to buy the ship, however.

DARWIN'S VOYAGE AROUND THE WORLD - Pathlights
www.pathlights.com/ce_encyclopedia/Encyclopedia/20hist05.htm ▾
Let us, for a few minutes, journey with Darwin as he sailed around the world! ... While the ship surveyed the South American coast, young Darwin spent much of .... What Darwin did not realize was the limiting wall imposed by the DNA coding.

Voyage of Charles Darwin Aboard H.M.S. Beagle - 19th Century History
history1800s.about.com › ... › 19th Century History › People and Politics › Innovators ▾
Darwin didn't actually formulate his theory of evolution while sailing around the world aboard the Royal Navy ship. But the exotic plants and animals he ...

A Trip Around the World - American Museum of Natural History
www.amnh.org/.../darwin/a-trip-around-the-wor... ▾ American Museum of Natural History ▾
A Five-Year Journey. The captain and crew of the HMS Beagle originally planned to spend two years on their trip around the world. Instead, the voyage took nearly five years, from December 1831 to October 1836.

# … when there is no answer, digging into “10 blue links” is even harder on mobile devices



●●○○○ AT&T 3:56 PM 86%

web for ‘Why do we call Donald trump by his last name and Hillary Clinton by her first name?’:

BING SEARCH

Why do we call Donald Trump by his last name and Hillary ...

I just caught myself doing it and now I wonder why that's the trend with men vs. women in politics? We never ... (maybe because Sarah is too common a name ...

https://answers.yahoo.com

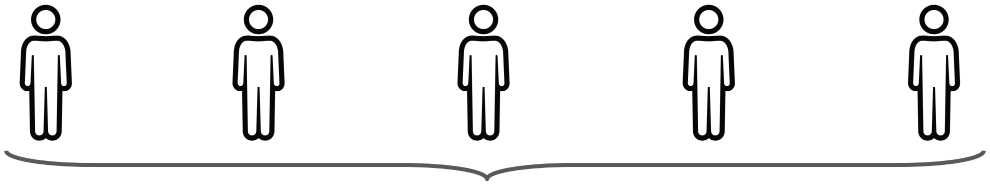
Hillary Clinton Calls Donald Trump a RACIST... And Trump ...

Unlike Hillary Clinton, Trump has earned his money ... up to Hillary! Do you support Donald Trump? ... ..Donald will put her in her place...... we don’t need ...

www.thepoliticalinsider.com

It is important to improve question answering for complex user information needs

# Goal of **TREC LiveQA** shared task is to advance research into **answering real user questions in real time**

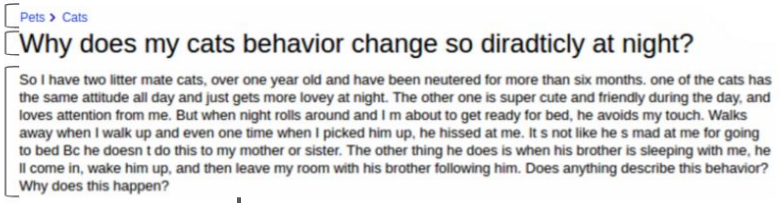


YAHOO! ANSWERS

category

title

body

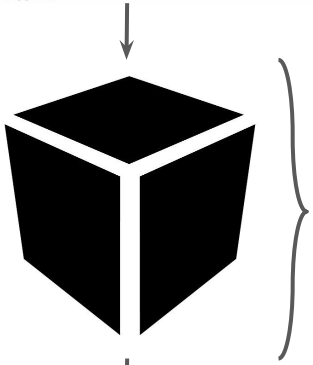


Pets › Cats

Why does my cats behavior change so diradticly at night?

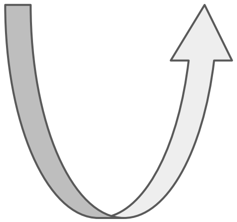
So I have two litter mate cats, over one year old and have been neutered for more than six months. one of the cats has the same attitude all day and just gets more lovey at night. The other one is super cute and friendly during the day, and loves attention from me. But when night rolls around and I m about to get ready for bed, he avoids my touch. Walks away when I walk up and even one time when I picked him up, he hissed at me. It s not like he s mad at me for going to bed Bc he doesn t do this to my mother or sister. The other thing he does is when his brother is sleeping with me, he ll come in, wake him up, and then leave my room with his brother following him. Does anything describe this behavior? Why does this happen?

Question Answering System



1 minute

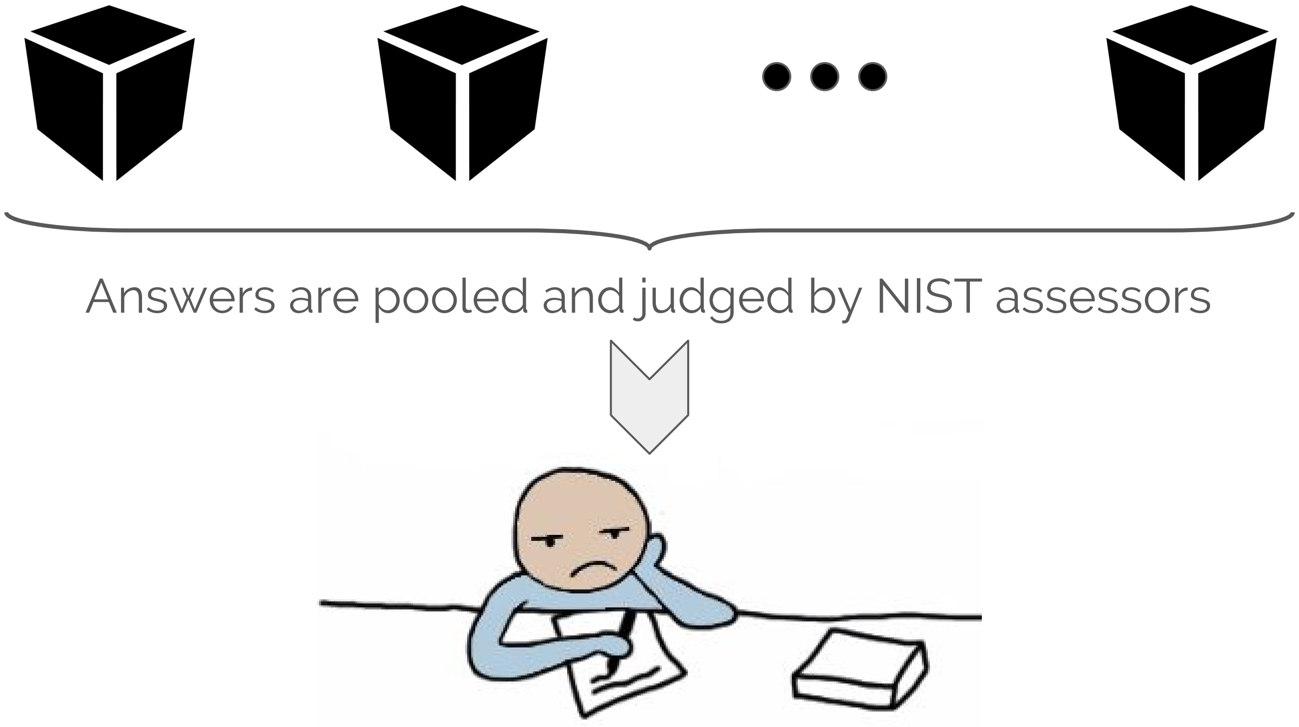
24 hours



≤ 1000 chars

https://sites.google.com/site/trecliveqa2016/

# LiveQA Evaluation Setup



Answers are pooled and judged by NIST assessors

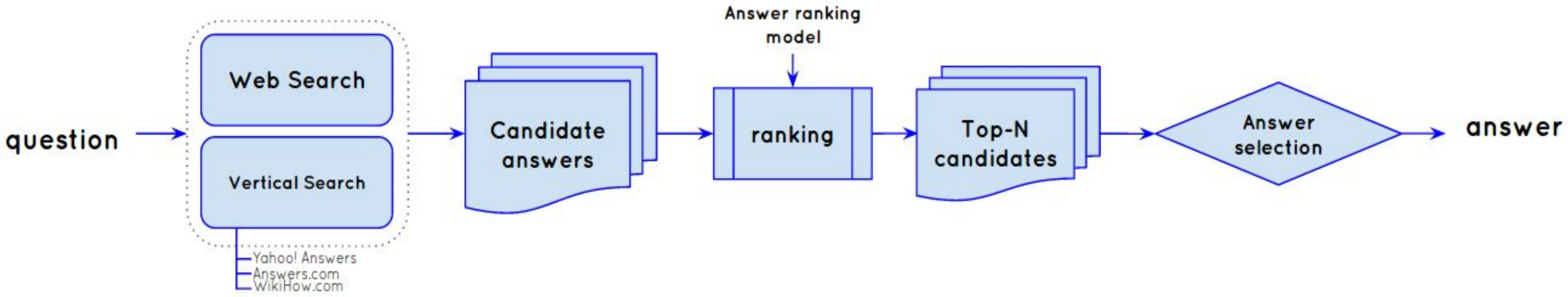
- **1: Bad** - contains no useful information
- **2: Fair** - marginally useful information
- **3: Good** - partially answers the question
- **4: Excellent** - fully answers the question

# LiveQA 2015:

## Even the best system returns a fair or better answer only for ~50% of the questions!

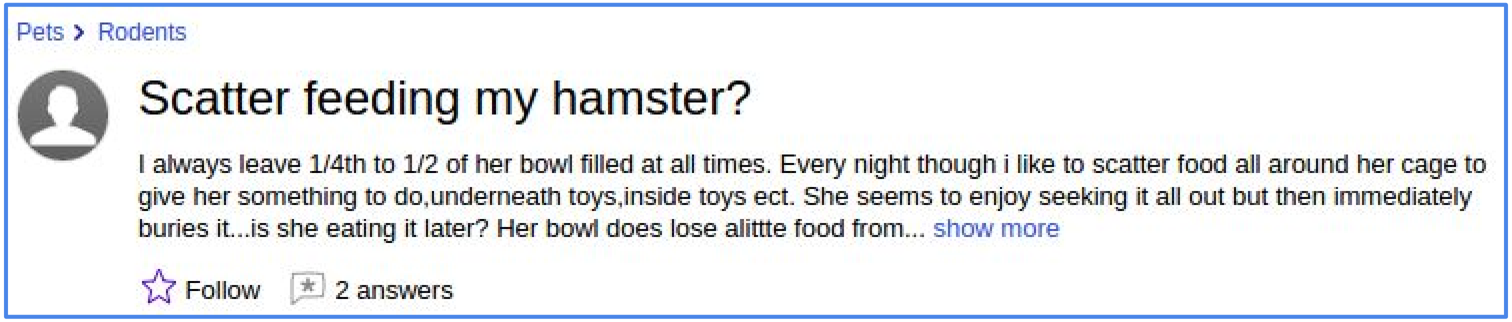
| | Avg score (0-3) | % questions with fair or better answer | % questions with excellent answer |
|---|---|---|---|
| Best system | 1.08 | 53.2 | 19.0 |

# The architecture of baseline automatic QA system



1. Search data sources
    a. CQA archives
        i. Yahoo! Answers
        ii. Answers.com
        iii. WikiHow
    b. Web search API
2. Extract candidates and their context
    a. Answers to retrieved questions
    b. Content blocks from regular web pages
3. Represent candidate answers with a set of features
4. Rank them using LambdaMART model
5. Return the top candidate as the answer

# Common Problem: Automatic systems often return an answer about the same topic, but irrelevant to the question



Pets > Rodents

Scatter feeding my hamster?

I always leave 1/4th to 1/2 of her bowl filled at all times. Every night though i like to scatter food all around her cage to give her something to do,underneath toys,inside toys ect. She seems to enjoy seeking it all out but then immediately buries it...is she eating it later? Her bowl does lose alittte food from... show more

Throwback to when my friends hamster ate my hamster and then my friends hamster died because she forgot to feed it karma

# Incorporate crowdsourcing to assist an automatic real-time question answering system

Or: combine human insight and automatic QA with machine learning

# Existing research

- ✓ **“Direct answers for search queries in the long tail”** by M.Bernstein et al, 2012
  - ○ Offline crowdsourcing of answers for long-tail search queries
- ✓ **“CrowdDB: answering queries with crowdsourcing”** by M.Franklin et al, 2011
  - ○ Using crowd to perform complex operations in SQL queries
- ✓ **“Answering search queries with crowdsearcher”** by A.Bozzon et al, 2012
  - ○ Answering queries using social media
- ✓ **“Dialog system using real-time crowdsourcing and twitter large-scale corpus”** by F. Bessho et al, 2012
  - ○ Real-time crowdsourcing as a backup plan for dialog
- ✓ **“Chorus: A crowd-powered conversational assistant”** by W.Lasecki, 2013
  - ○ Real-time chatbot powered by crowdsourcing

... and many other works

# Research Questions

- **RQ1.** Can crowdsourcing be used to improve the performance of a near real-time automatic question answering system?

# Research Questions

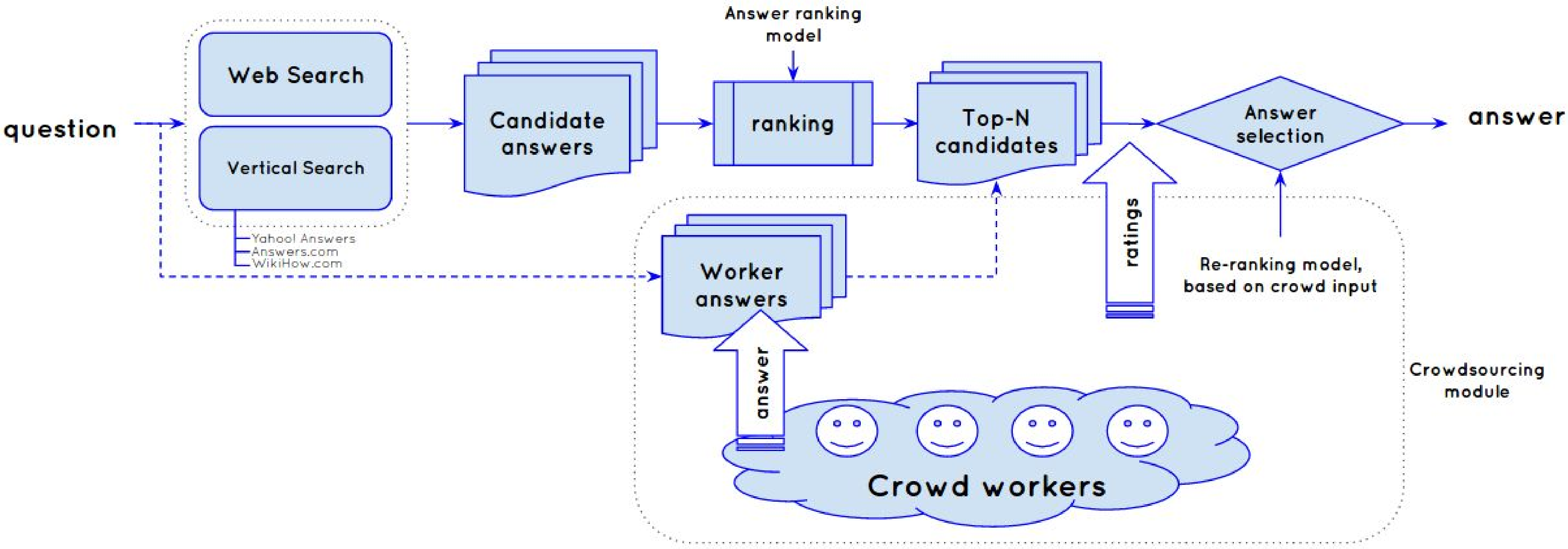
- **RQ1.** Can crowdsourcing be used to improve the performance of a near real-time automatic question answering system?

- **RQ2**. What kind of contributions from crowd workers can help improve automatic question answering and what is the relative impact of different types of feedback to the overall question answering performance?

# Research Questions

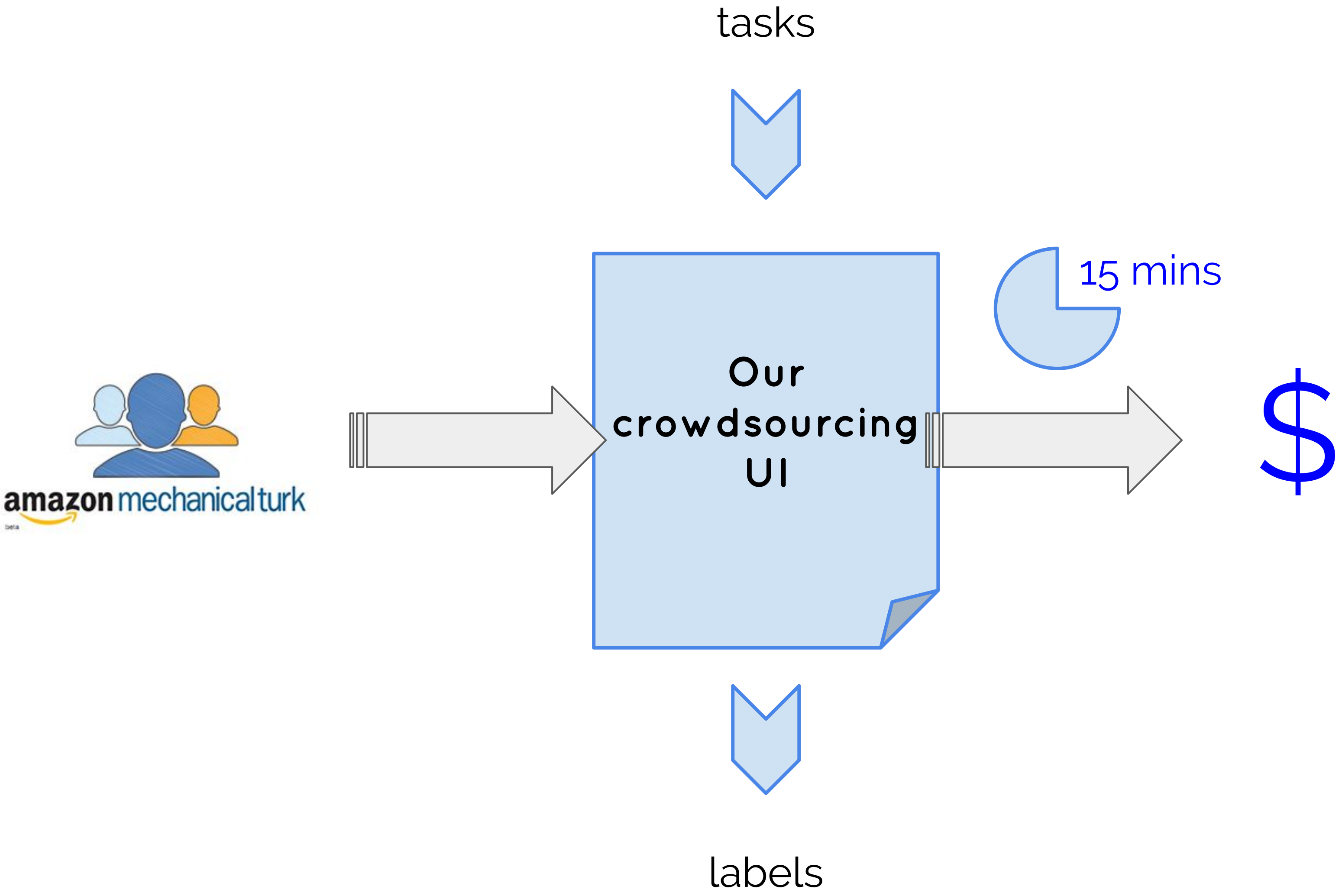
- **RQ1.** Can crowdsourcing be used to improve the performance of a near real-time automatic question answering system?
- **RQ2**. What kind of contributions from crowd workers can help improve automatic question answering and what is the relative impact of different types of feedback to the overall question answering performance?
- **RQ3**. What are the trade-offs in performance, cost, and scalability of using crowdsourcing for real-time question answering?

# CRQA: Integrating crowdsourcing with automatic QA system



1. After receiving a question, it is forwarded to the crowd
2. Can start working on the answer, if possible
3. When system ranks candidates, top-7 are pushed to workers for rating
4. Rated human and automatically generated answers are returned
5. System re-rank them based on all available information
6. Top candidate is returned as the answer

# We used the retainer model for real-time crowdsourcing



tasks

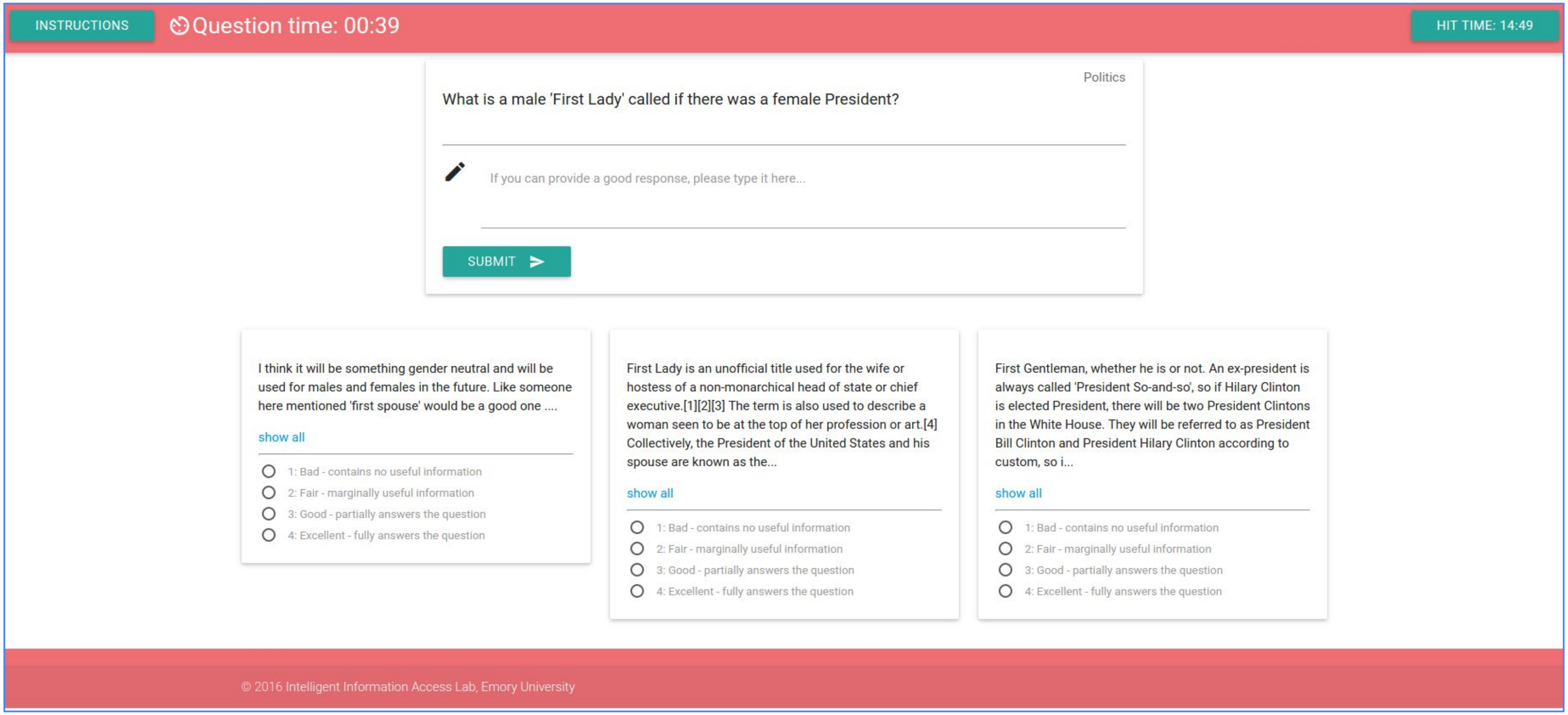
Our crowdsourcing UI

15 mins

$

labels

# UI for crowdsourcing answers and ratings



Politics

What is a male 'First Lady' called if there was a female President?

If you can provide a good response, please type it here...

I think it will be something gender neutral and will be used for males and females in the future. Like someone here mentioned 'first spouse' would be a good one ....

show all

○ 1: Bad - contains no useful information
○ 2: Fair - marginally useful information
○ 3: Good - partially answers the question
○ 4: Excellent - fully answers the question

First Lady is an unofficial title used for the wife or hostess of a non-monarchical head of state or chief executive.[1][2][3] The term is also used to describe a woman seen to be at the top of her profession or art.[4] Collectively, the President of the United States and his spouse are known as the...

show all

○ 1: Bad - contains no useful information
○ 2: Fair - marginally useful information
○ 3: Good - partially answers the question
○ 4: Excellent - fully answers the question

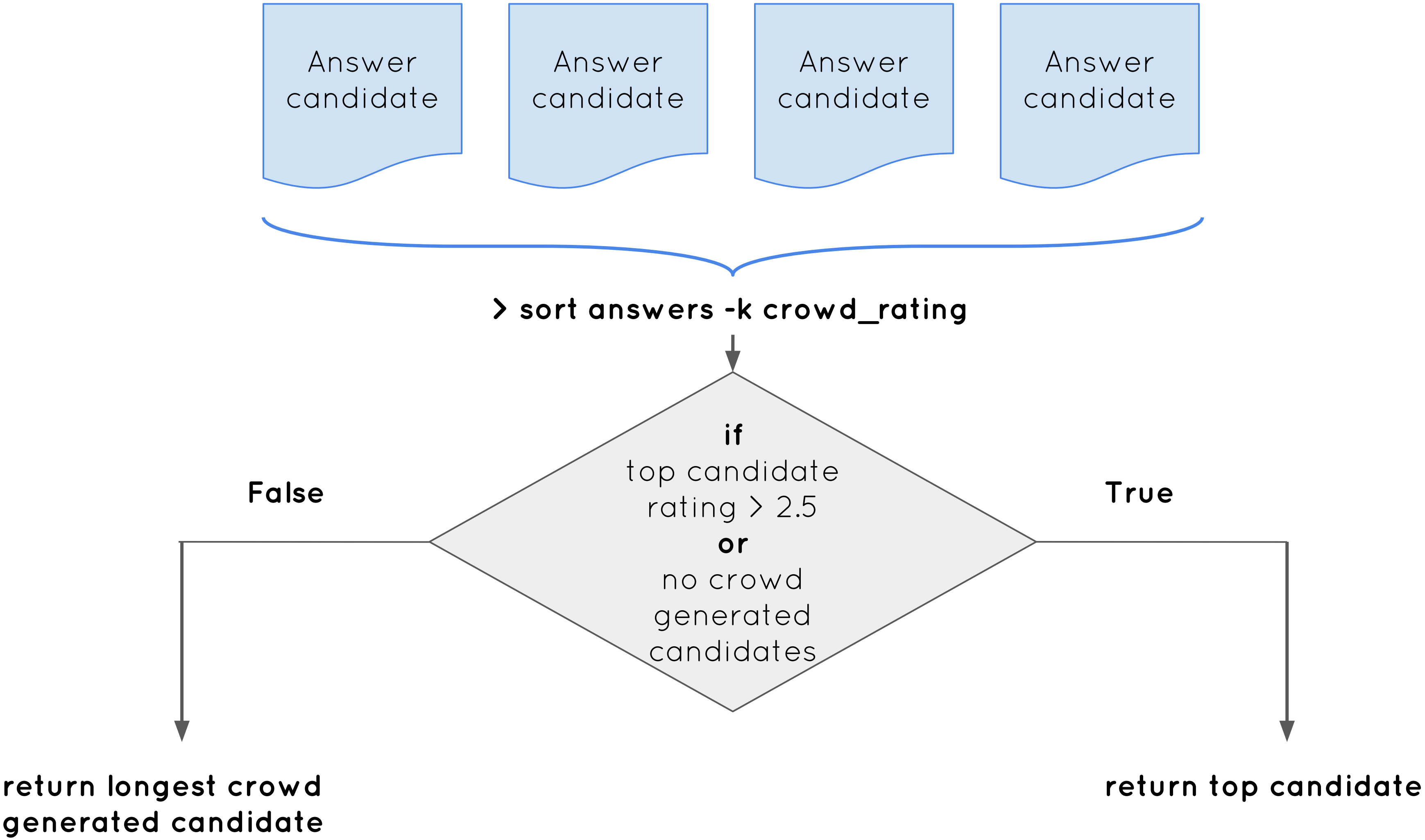
First Gentleman, whether he is or not. An ex-president is always called 'President So-and-so', so if Hilary Clinton is elected President, there will be two President Clintons in the White House. They will be referred to as President Bill Clinton and President Hilary Clinton according to custom, so i...

show all

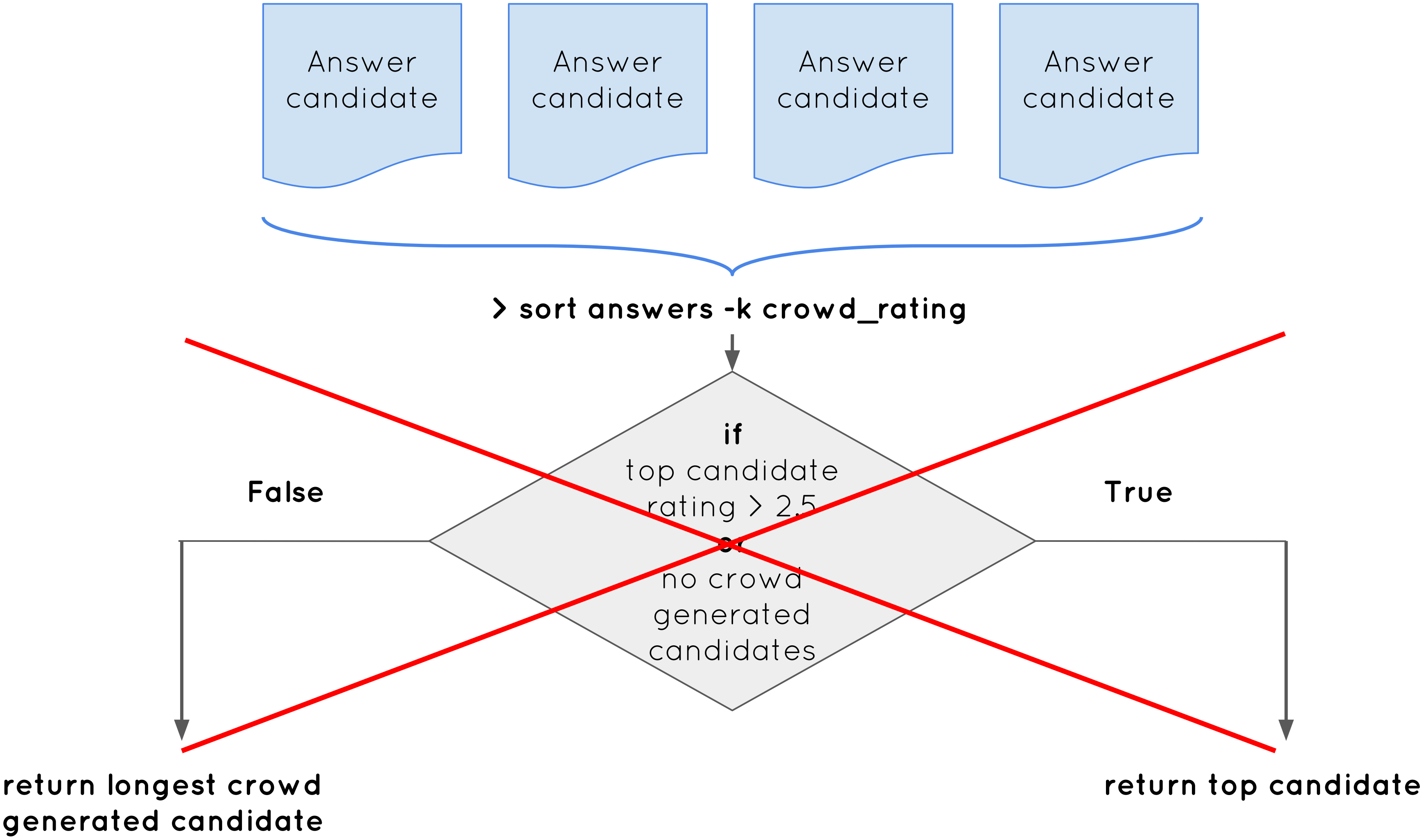
○ 1: Bad - contains no useful information
○ 2: Fair - marginally useful information
○ 3: Good - partially answers the question
○ 4: Excellent - fully answers the question

© 2016 Intelligent Information Access Lab, Emory University

# Heuristic answer re-ranking (during TREC LiveQA)



# CRQA uses a learning-to-rank model to re-rank



Answer candidate

Answer candidate

Answer candidate

Answer candidate

> sort answers -k crowd_rating

**if**
top candidate rating > 2.5
**or**
no crowd generated candidates

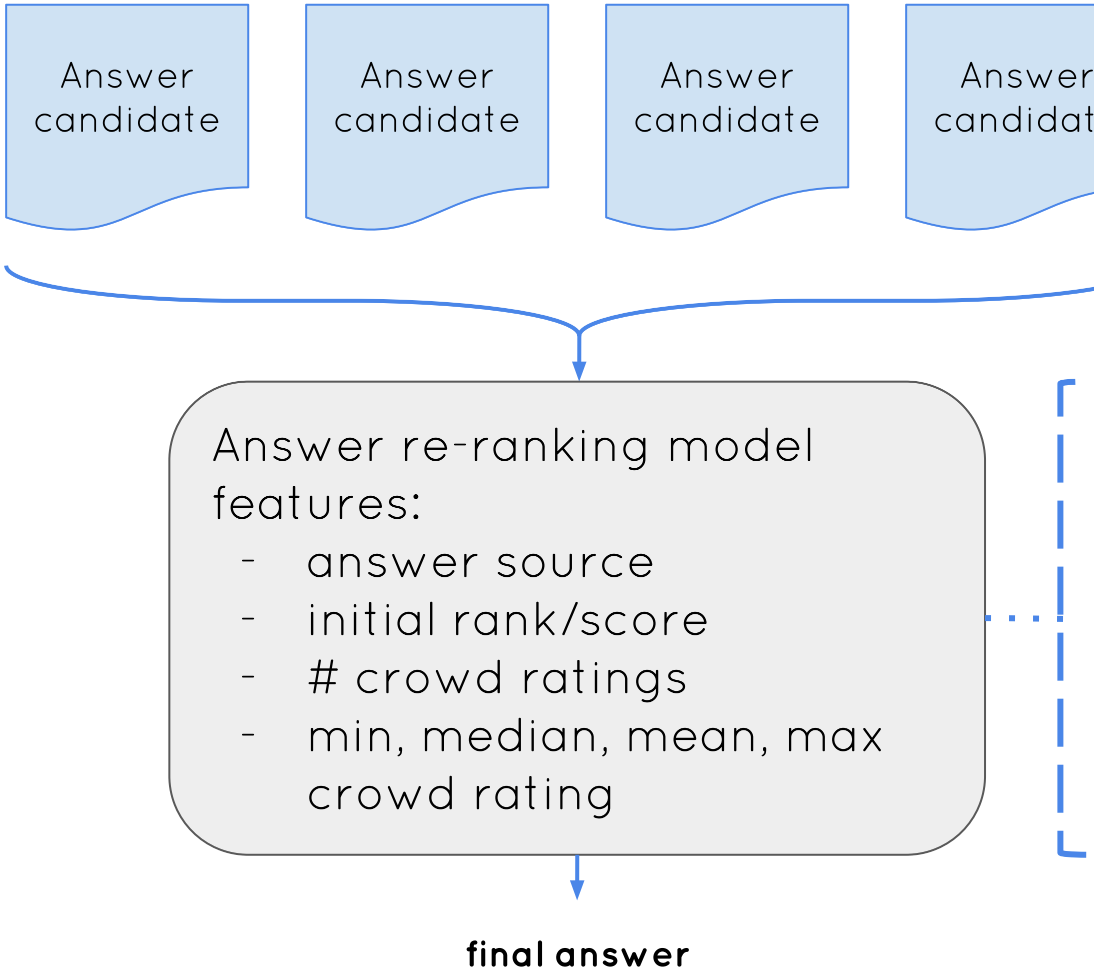
**False**

**True**

**return longest crowd generated candidate**

**return top candidate**

# CRQA uses a learning-to-rank model to re-rank



Answer candidate

Answer candidate

Answer candidate

Answer candidate

Answer re-ranking model features:

- answer source
- initial rank/score
- # crowd ratings
- min, median, mean, max crowd rating

**final answer**

- Offline crowdsourcing to get ground-truth labels
- Included Yahoo!Answers community response, crawled 2 days after challenge
- Trained GBRT model, 10-fold cross validation

# Evaluation

# Evaluation setup

**Methods compared**:

- Automatic QA
- CRQA (heuristic): re-ranking by crowdsourced score
- CRQA (LTR): re-ranking using a learning-to-rank model
- Yahoo! Answers (crawled 2 days later)

**Metrics**:

- avg-score: average answer score over all questions
- avg-prec: average answer score
- success@i$^+$: fraction of questions with answer score ≥ i
- precision@i$^+$: fraction of answers with score ≥ i

# Dataset

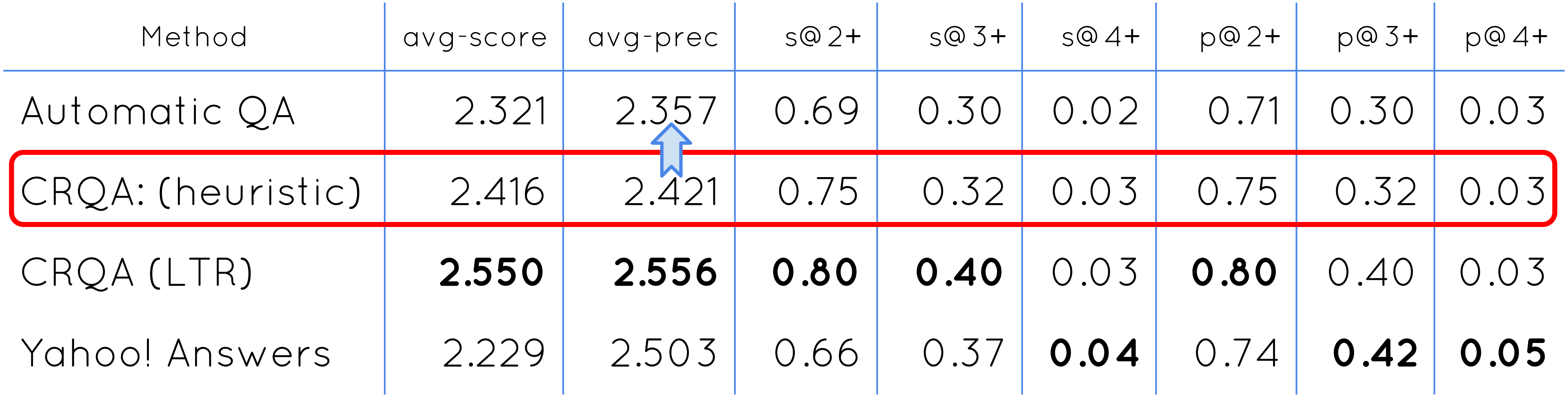
- ➢ 1,088 questions from LiveQA 2016 run
- ➢ Top 7 system and crowd-generated answers
- ➢ Answer quality labelling on a scale from 1 to 4
  - offline
  - also using crowdsourcing (different workers)

| | |
|---|---|
| Number of questions received | 1,088 |
| Number of MTurk 15 minutes assignments completed | 889 |
| Average number of questions per assignment | 11.44 |
| Total cost per question | $0.81 |
| Avg number of answers provided by workers per question | 1.25 |
| Average number of ratings per answer | 6.25 |

# Main Results

| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA: (heuristic) | 2.416 | 2.421 | 0.75 | 0.32 | 0.03 | 0.75 | 0.32 | 0.03 |
| CRQA (LTR) | **2.550** | **2.556** | **0.80** | **0.40** | 0.03 | **0.80** | 0.40 | 0.03 |
| Yahoo! Answers | 2.229 | 2.503 | 0.66 | 0.37 | **0.04** | 0.74 | **0.42** | **0.05** |

# Crowdsourcing improves performance of automatic QA system



| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA: (heuristic) | 2.416 | 2.421 | 0.75 | 0.32 | 0.03 | 0.75 | 0.32 | 0.03 |
| CRQA (LTR) | **2.550** | **2.556** | **0.80** | **0.40** | 0.03 | **0.80** | 0.40 | 0.03 |
| Yahoo! Answers | 2.229 | 2.503 | 0.66 | 0.37 | **0.04** | 0.74 | **0.42** | **0.05** |

# Learning-to-rank model allows to more effectively combine all available signals and return a better answer

| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA: (heuristic) | 2.416 | 2.421 | 0.75 | 0.32 | 0.03 | 0.75 | 0.32 | 0.03 |
| CRQA (LTR) | **2.550** | **2.556** | **0.80** | **0.40** | 0.03 | **0.80** | 0.40 | 0.03 |
| Yahoo! Answers | 2.229 | 2.503 | 0.66 | 0.37 | **0.04** | 0.74 | **0.42** | **0.05** |

# CRQA reaches the quality of community responses on Yahoo! Answers

| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA: (heuristic) | 2.416 | 2.421 | 0.75 | 0.32 | 0.03 | 0.75 | 0.32 | 0.03 |
| CRQA (LTR) | **2.550** | **2.556** | **0.80** | **0.40** | 0.03 | **0.80** | 0.40 | 0.03 |
| Yahoo! Answers | 2.229 | 2.503 | 0.66 | 0.37 | **0.04** | 0.74 | **0.42** | **0.05** |

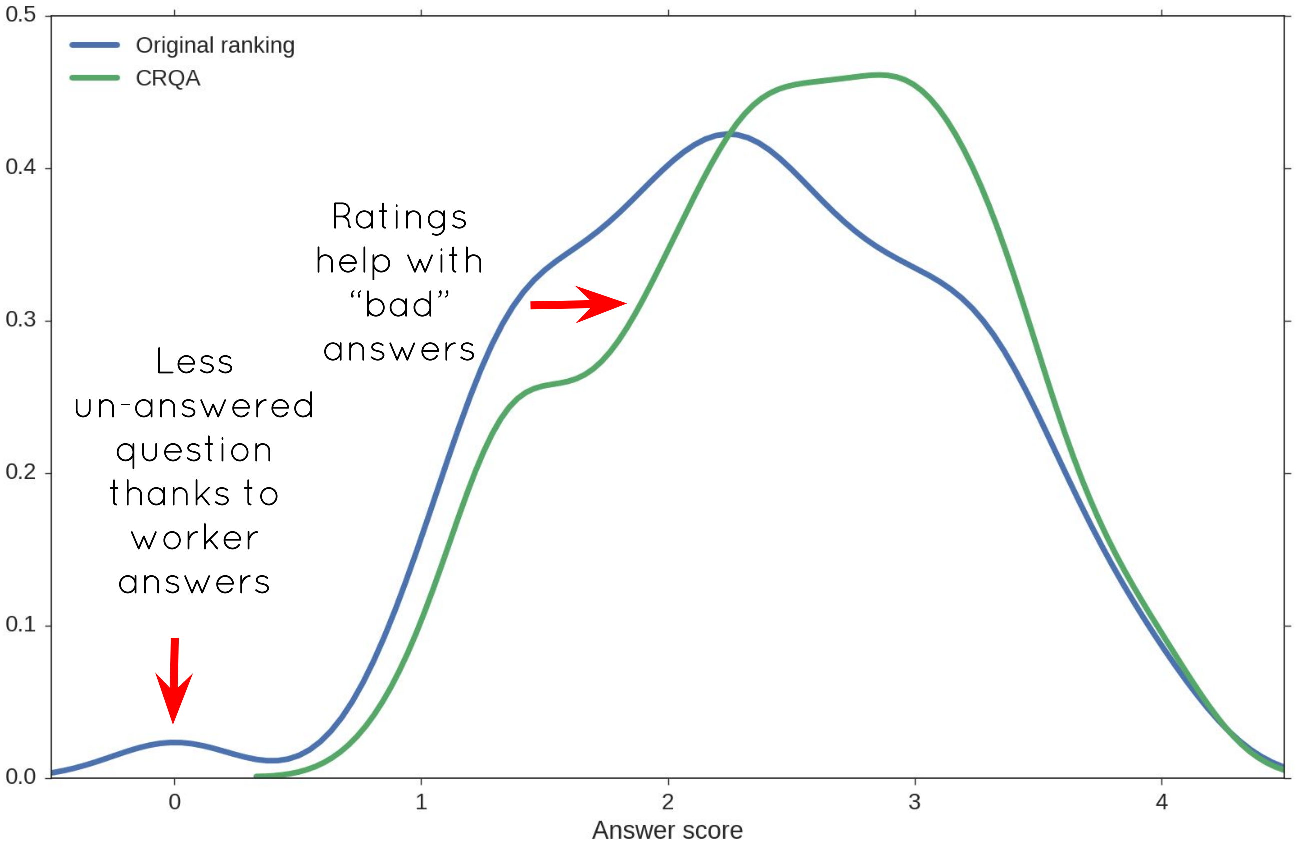
# … and it has much better coverage

| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA: (heuristic) | 2.416 | 2.421 | 0.75 | 0.32 | 0.03 | 0.75 | 0.32 | 0.03 |
| CRQA (LTR) | **2.550** | **2.556** | **0.80** | **0.40** | 0.03 | **0.80** | 0.40 | 0.03 |
| Yahoo! Answers | 2.229 | 2.503 | 0.66 | 0.37 | **0.04** | 0.74 | **0.42** | **0.05** |

# Both worker answers and ratings make an equal contribution to the answer quality improvements

| Method | avg-score | avg-prec | s@2+ | s@3+ | s@4+ | p@2+ | p@3+ | p@4+ |
|---|---|---|---|---|---|---|---|---|
| Automatic QA | 2.321 | 2.357 | 0.69 | 0.30 | 0.02 | 0.71 | 0.30 | 0.03 |
| CRQA (LTR) | 2.550 | 2.556 | 0.80 | 0.40 | 0.03 | 0.80 | 0.40 | 0.03 |
| no worker answers | 2.432 | 2.470 | 0.75 | 0.35 | 0.03 | 0.76 | 0.35 | 0.03 |
| no worker ratings | 2.459 | 2.463 | 0.76 | 0.35 | 0.03 | 0.76 | 0.36 | 0.03 |

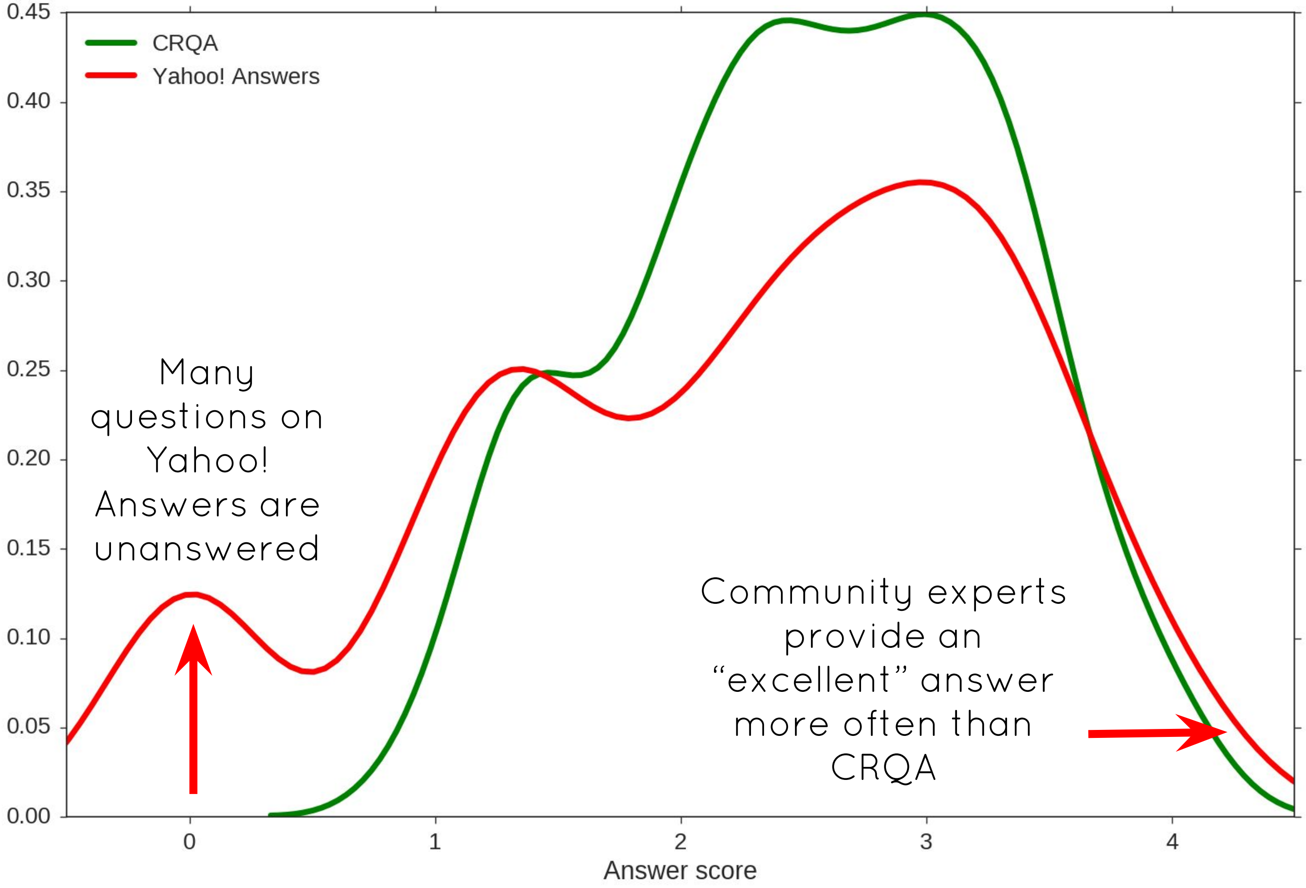
# Crowdsourcing helps to improve empty and low quality answers



Ratings help with "bad" answers

Less un-answered question thanks to worker answers

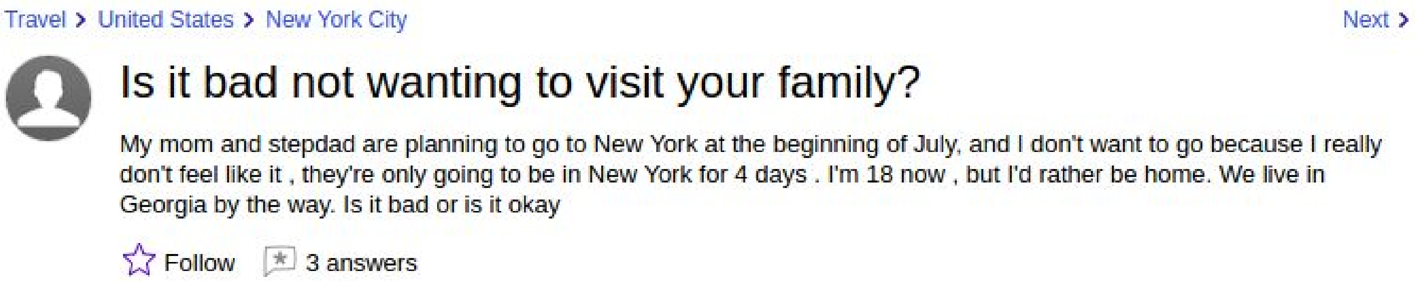
# Yahoo! Answers have both higher percentage of excellent and missing and low quality answers



Many questions on Yahoo! Answers are unanswered

Community experts provide an "excellent" answer more often than CRQA

# Crowdsourced answers are especially good for general knowledge questions



Travel > United States > New York City

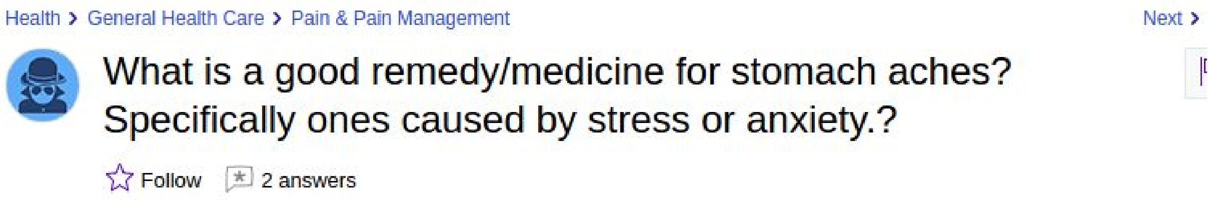
Next >

Is it bad not wanting to visit your family?

My mom and stepdad are planning to go to New York at the beginning of July, and I don't want to go because I really don't feel like it , they're only going to be in New York for 4 days . I'm 18 now , but I'd rather be home. We live in Georgia by the way. Is it bad or is it okay

☆ Follow 3 answers

Is it bad not wanting to visit your family? It's nt bad. Just be honest with them. They may be upset but they should understand



Health > General Health Care > Pain & Pain Management

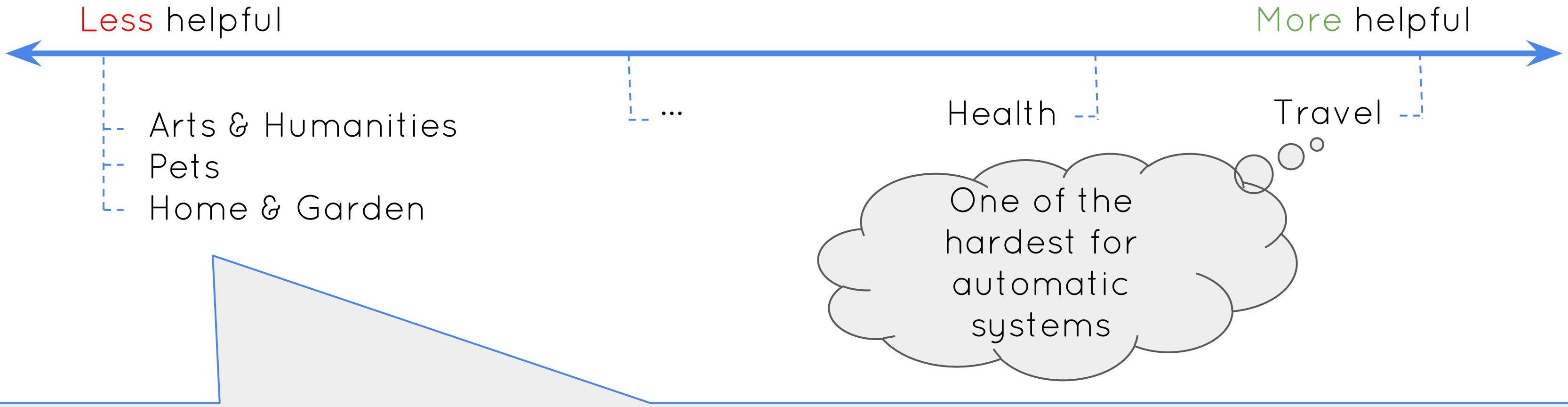
Next >

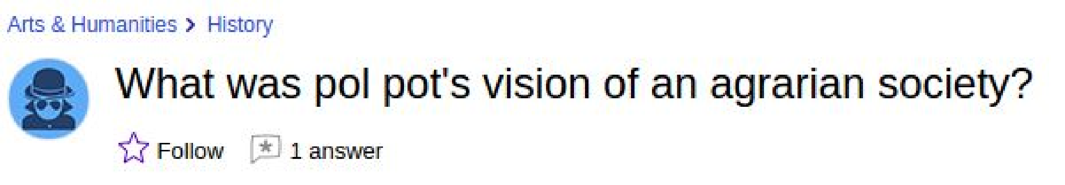
What is a good remedy/medicine for stomach aches? Specifically ones caused by stress or anxiety.?

☆ Follow 2 answers

Chamomile tea should help

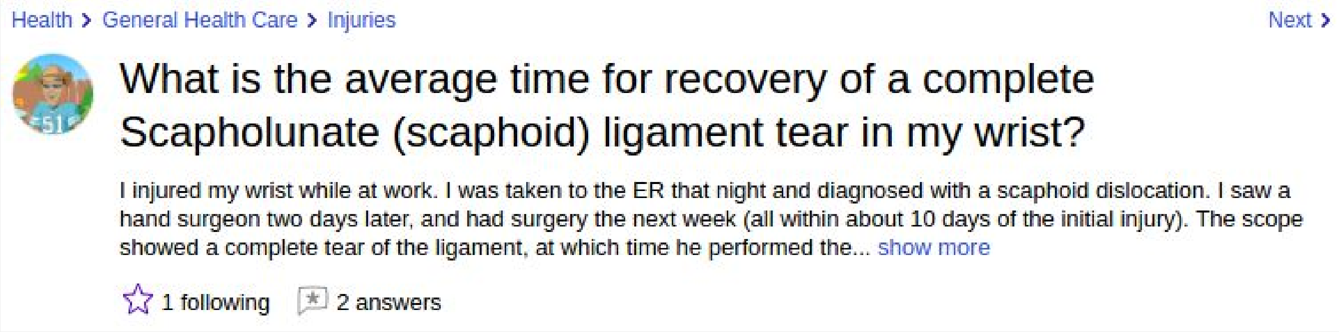
# But less effective for questions which require domain expertise




Arts & Humanities > History

What was pol pot's vision of an agrarian society?

☆ Follow 1 answer


Health > General Health Care > Injuries Next >

What is the average time for recovery of a complete Scapholunate (scaphoid) ligament tear in my wrist?

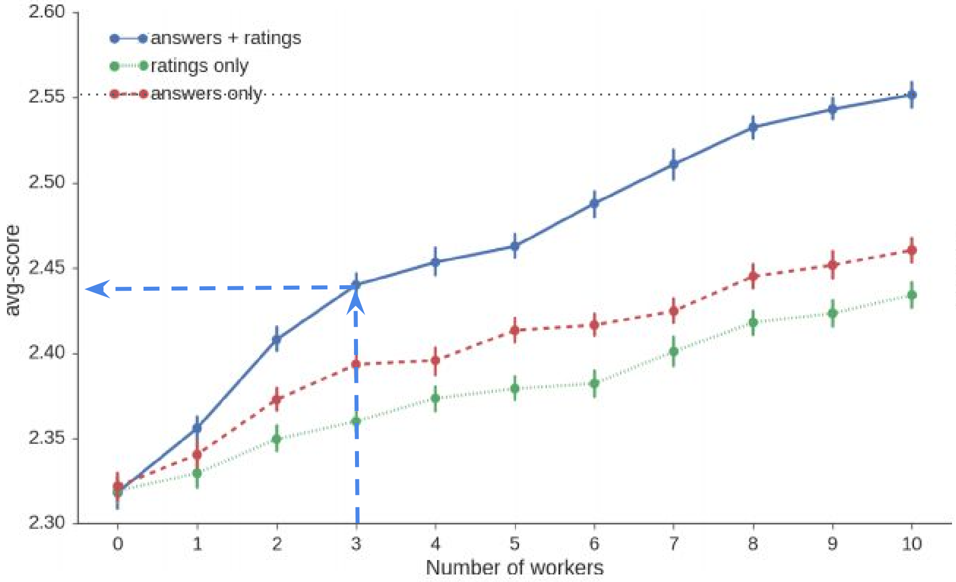
I injured my wrist while at work. I was taken to the ER that night and diagnosed with a scaphoid dislocation. I saw a hand surgeon two days later, and had surgery the next week (all within about 10 days of the initial injury). The scope showed a complete tear of the ligament, at which time he performed the... show more

☆ 1 following 2 answers

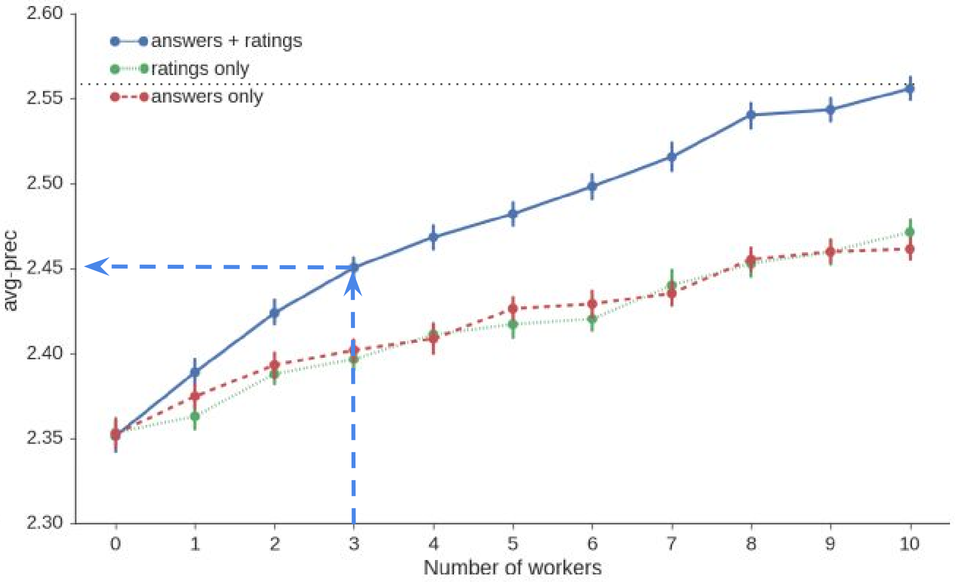
Ok, but what about the costs?

$0.81 per question is a lot of money

# Half of the overall improvements can be achieved with only 3 workers per question (30% of cost)



(a) avg-score: Average score per question



(b) avg-prec: Average score per answer (ignoring non-answered questions)

# Limitations and Future work

- Limitations
  - Fixed and uniform load for the system over 24 hours
    - Need variable size pool of workers based on the current load
- Ideas
  - Allocate crowdsourcing resources based on expected performance of the automated system
  - Use other types of feedback:
    - Search query generation
    - Key phrases to look for the the answer
    - ...
  - Online learning from crowd feedback
  - Cost optimization
    - Decide which feedback, in what amount and when to request

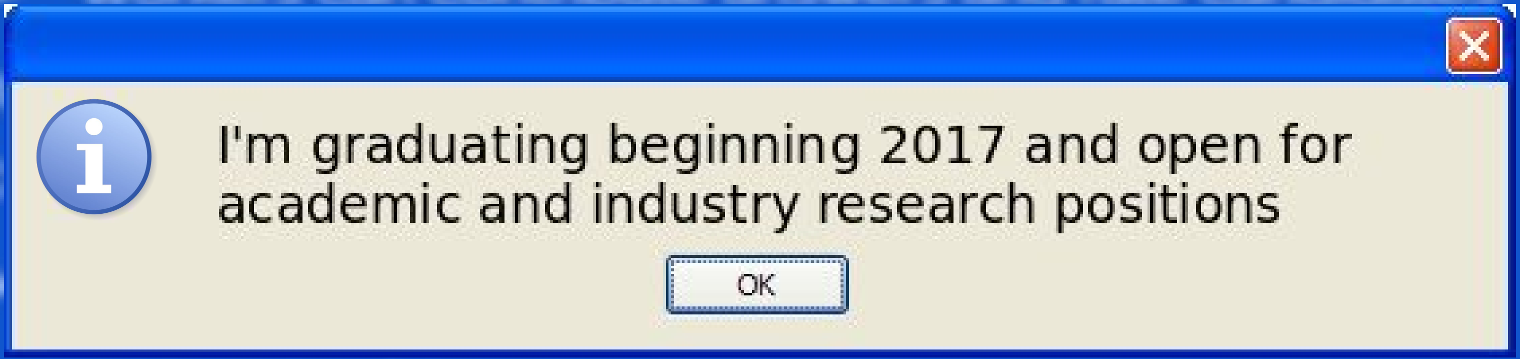
We conducted large scale experiments on real user questions, which showed:

- ○ Crowdsourcing helps for real-time QA
    - ➢ Workers can contribute answers and rate candidates
    - ➢ Humans can immediately reject off-topic candidates
- ○ Answers from our system are often even preferred to community answers
    - ➢ Which are collected 2 days after
    - ➢ With 20% of the questions were still unanswered by the community

Thank you!

We conducted large scale experiments on real user questions, which showed:

- Crowdsourcing helps for real-time QA
  - Workers can contribute answers and rate candidates



I'm graduating beginning 2017 and open for academic and industry research positions

  - With 20% of the questions still unanswered

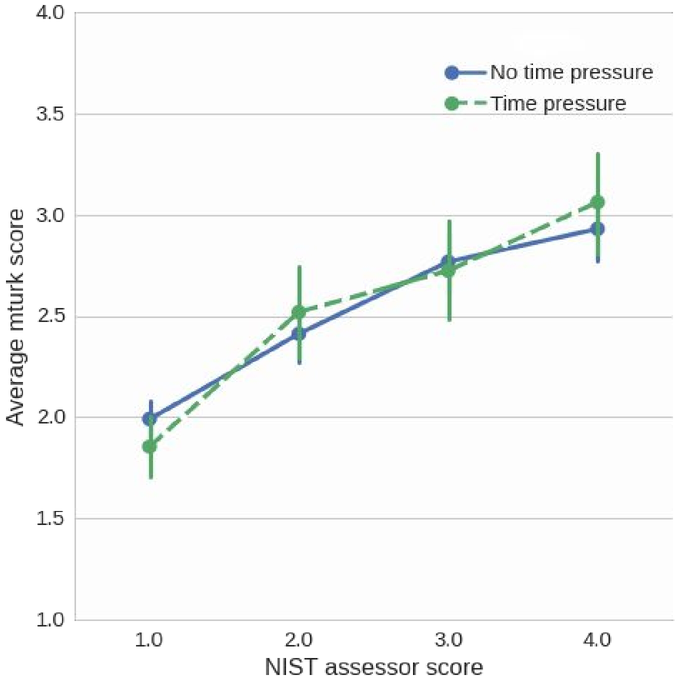
## Thank you!

# It’s better to present candidates ordered by their predicted quality

Average answer score if presented in different order:

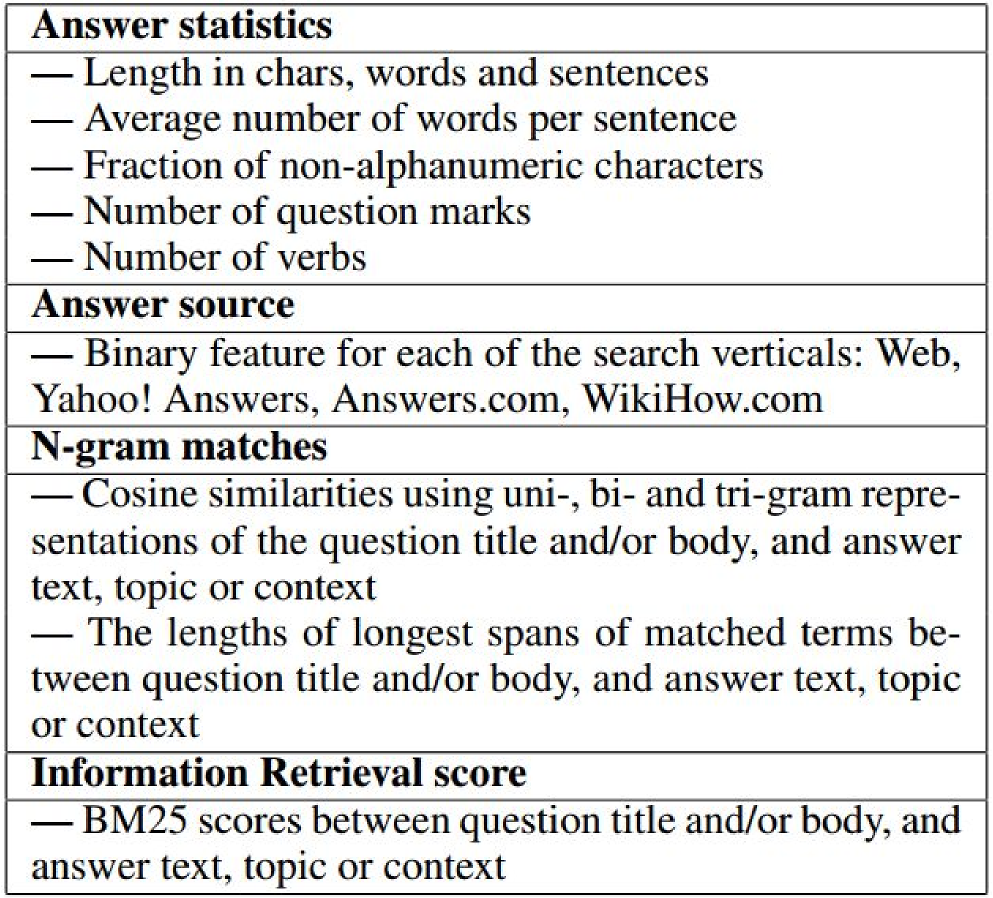
| | |
|---|---|
| Sorted by rank | 2.539 |
| Shuffled | 2.508 |

# [Backup] Crowdsourced labels correlate well with NIST assessor scores ($\rho$=0.52)



- ✓ Workers prefer to give intermediate scores (2, 3), while NIST assessors gave more extreme scores (1 and 4)
- ✓ There is no significant difference in quality between groups with and without time pressure

# Features [backup]



| |
|---|
| **Answer statistics** |
| — Length in chars, words and sentences<br>— Average number of words per sentence<br>— Fraction of non-alphanumeric characters<br>— Number of question marks<br>— Number of verbs |
| **Answer source** |
| — Binary feature for each of the search verticals: Web, Yahoo! Answers, Answers.com, WikiHow.com |
| **N-gram matches** |
| — Cosine similarities using uni-, bi- and tri-gram representations of the question title and/or body, and answer text, topic or context<br>— The lengths of longest spans of matched terms between question title and/or body, and answer text, topic or context |
| **Information Retrieval score** |
| — BM25 scores between question title and/or body, and answer text, topic or context |