Tutorial: Using Crowdsourcing Effectively for Social Media Research
17th May 2016
ICWSM 2016, Cologne,Germany

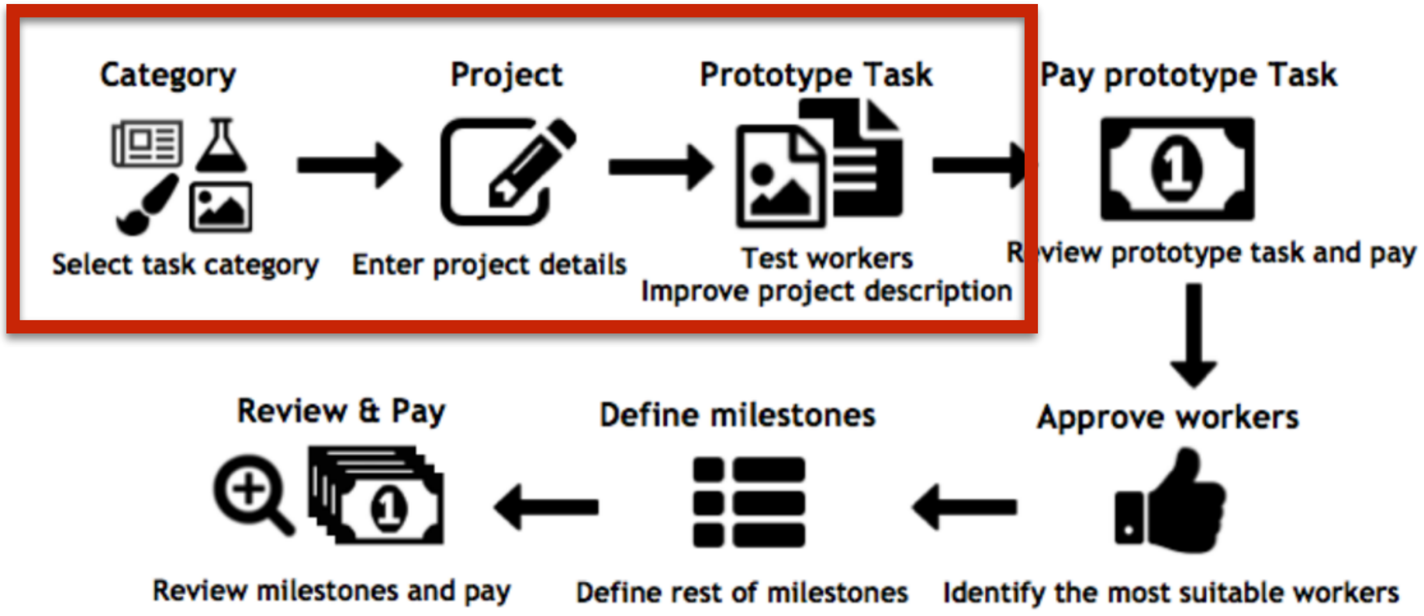
Part III Crowdsourcing and Social Media

Michele @pirroh Catasta
EPFL, Switzerland

Crowdsourcing 1.0

- no social component in MTurk/Crowdflower/etc.
- no notifications / no recommendations
- lack of economical incentives?

What is CS 2.0?



Category → Project → Prototype Task → Pay prototype Task

Select task category

Enter project details

Test workers
Improve project description

Review prototype task and pay

Approve workers

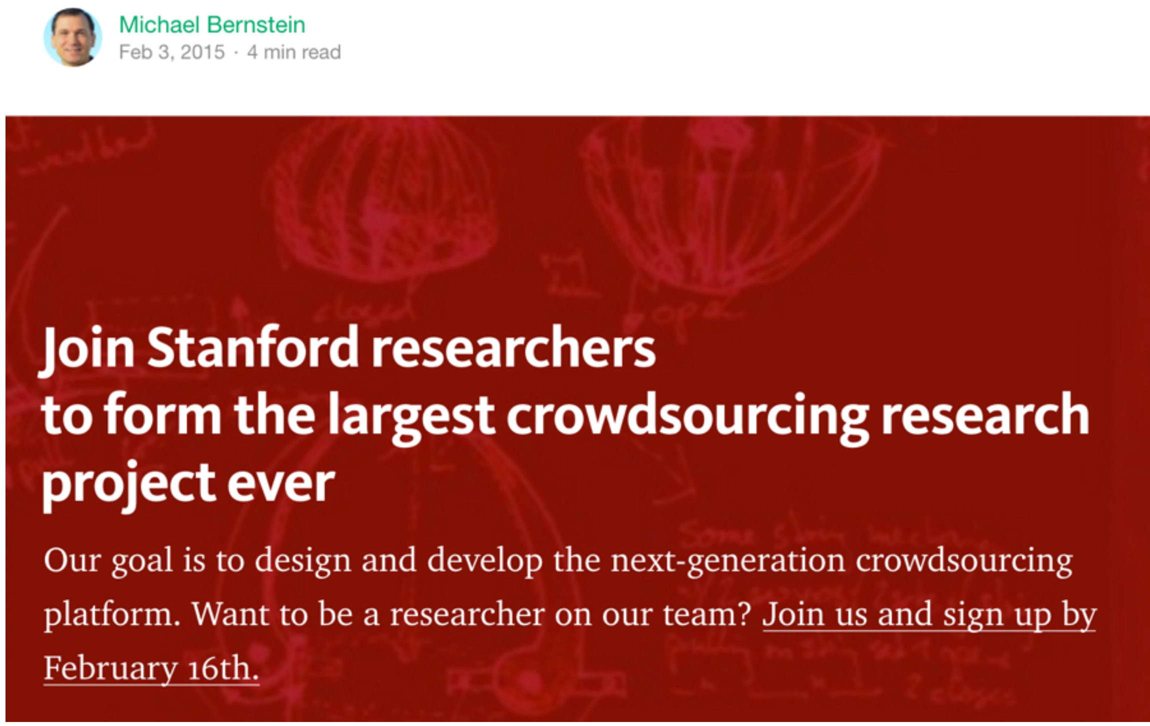
Identify the most suitable workers

Define milestones

Define rest of milestones

Review & Pay

Review milestones and pay



Michael Bernstein
Feb 3, 2015 · 4 min read

Join Stanford researchers to form the largest crowdsourcing research project ever

Our goal is to design and develop the next-generation crowdsourcing platform. Want to be a researcher on our team? Join us and sign up by February 16th.

- task prototyping vs custom apps
- Crowdflower partially covers this space, but we need an open source framework for tasks

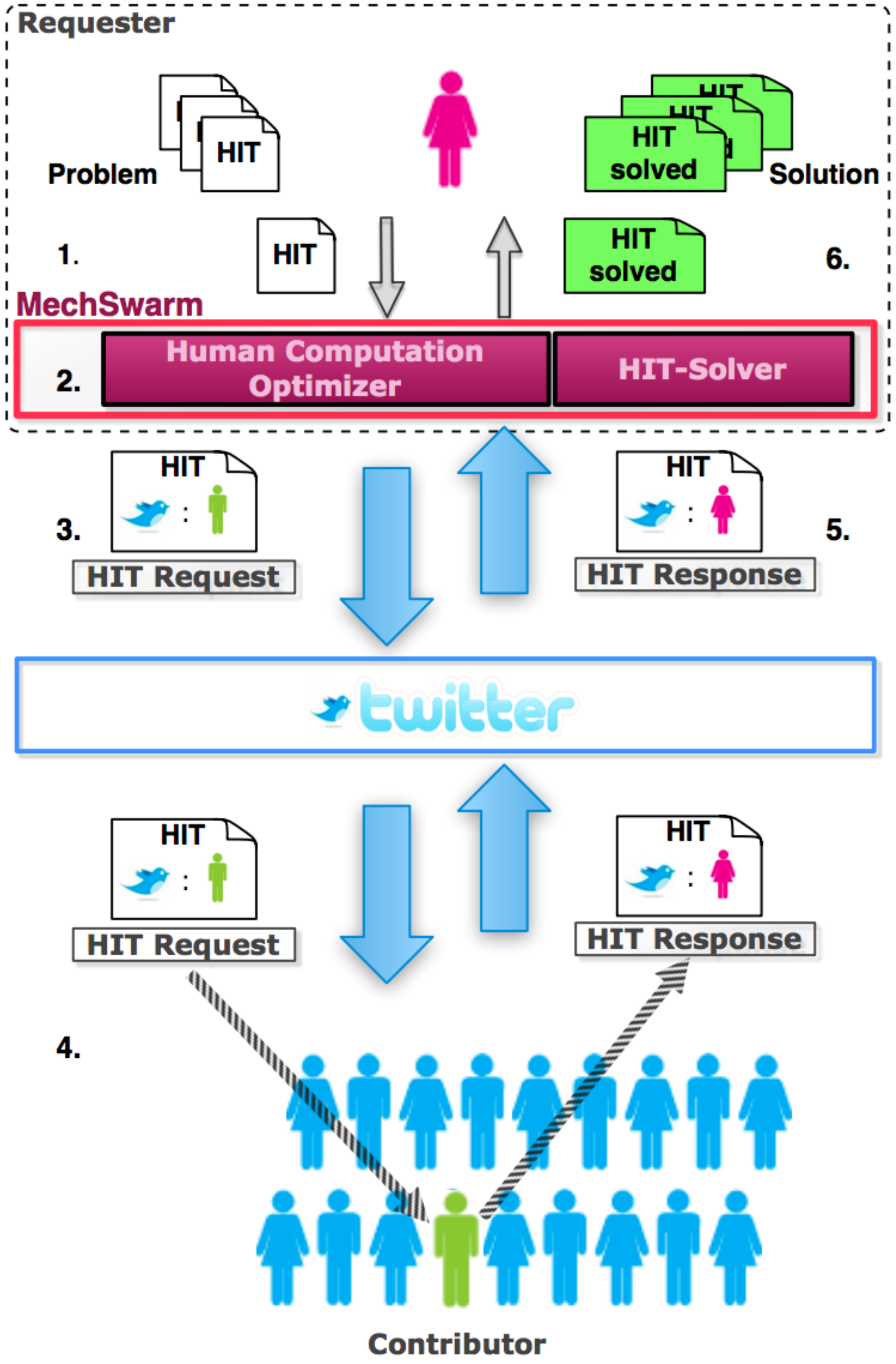
Stanford Crowd Research Collective
Daemo: a Self-Governed Crowdsourcing Marketplace. UIST 2015.

Social Media for Crowdsourcing

- Novel Decentralized Architecture
- The *Push-Crowdsourcing* paradigm
- How do we ensure quality?

Crowdsourcing on Twitter

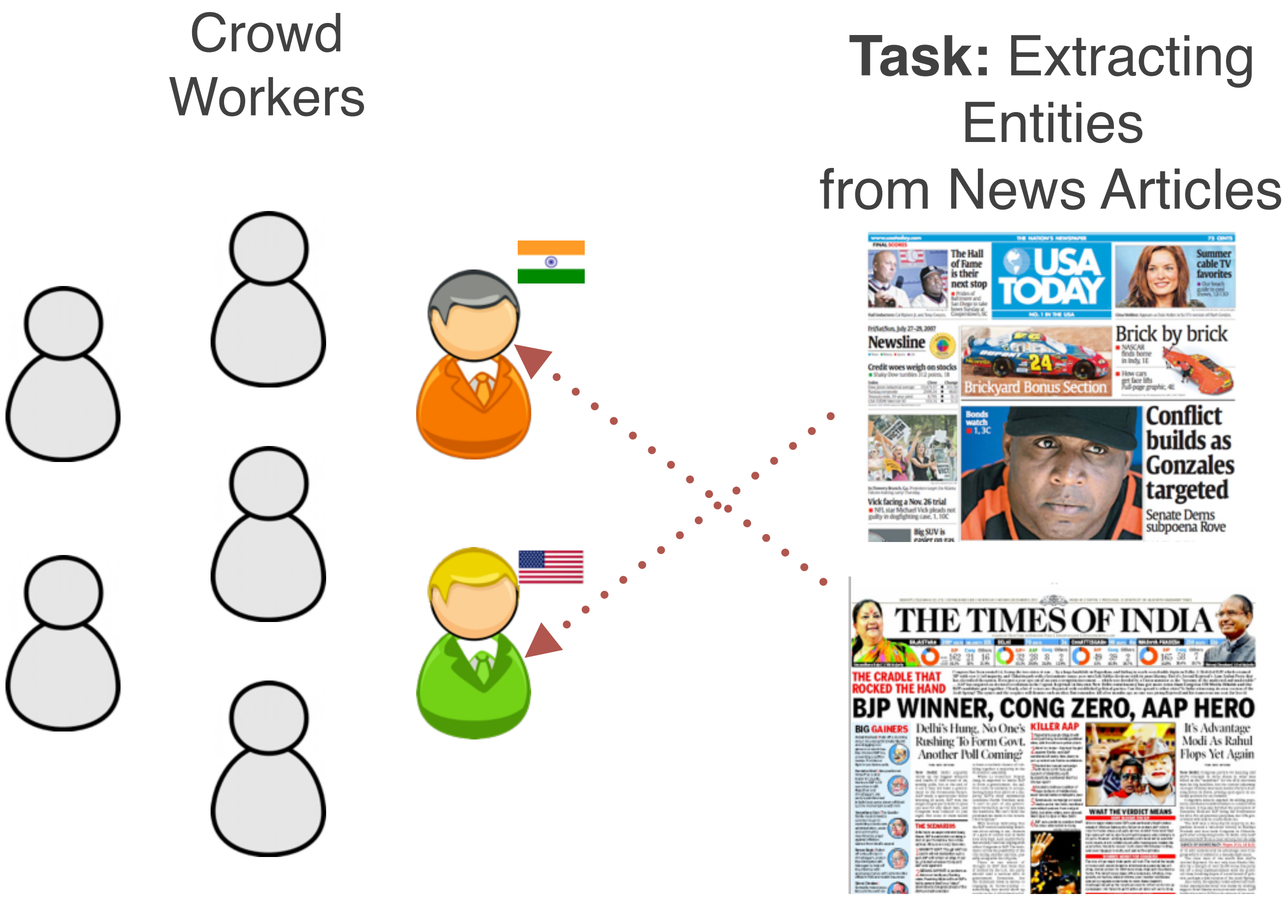
- large user base
- “notification” system
- assignment problem:
  “who is the best worker for a certain HIT?”



Diaz-Aviles et al.
Exploiting Twitter as a Social Channel for Human Computation. CrowdSearch 2012.

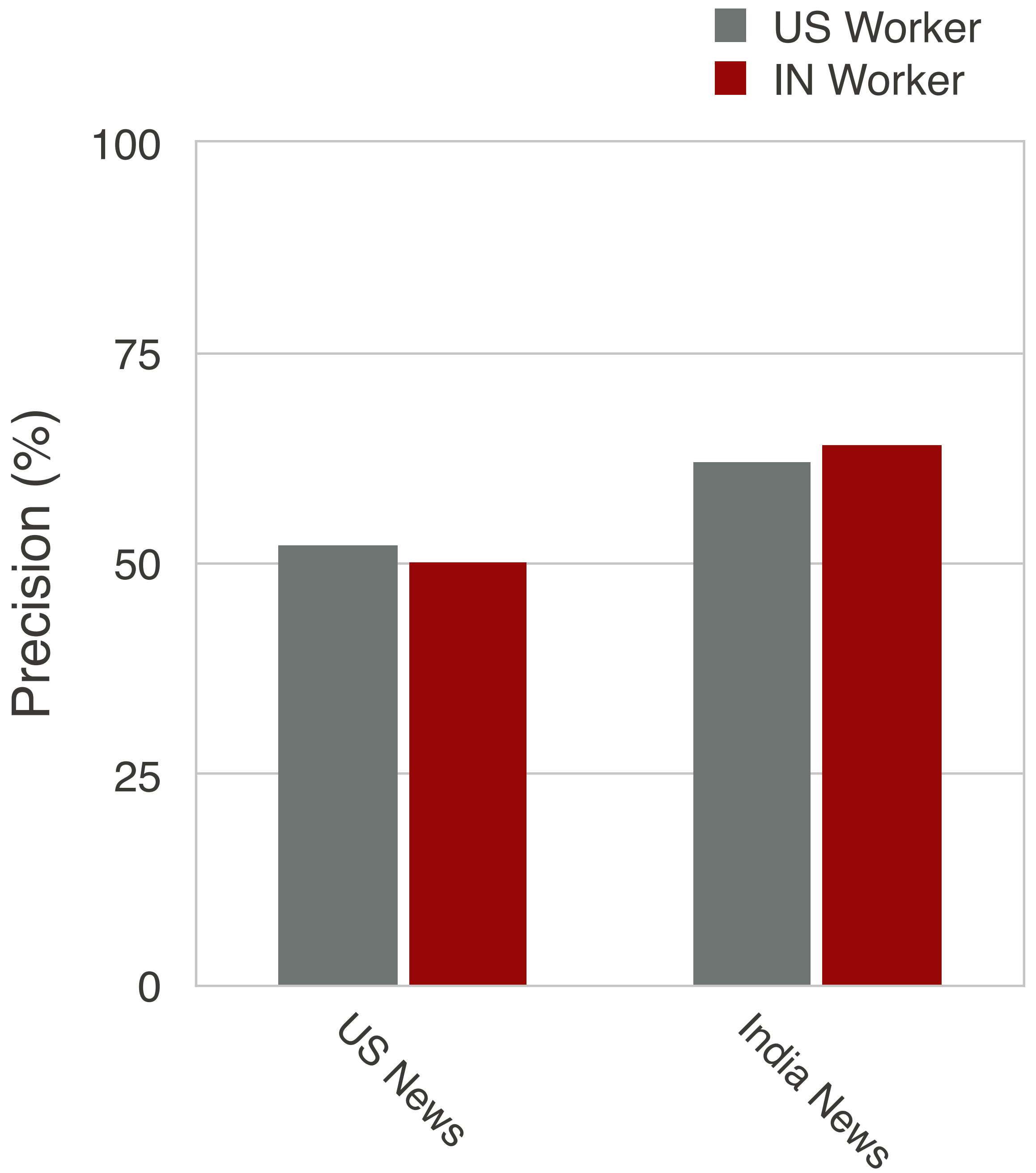
Task Routing

Selecting the Crowd We Need



Crowd Workers

Task: Extracting Entities from News Articles



Difallah, Demartini, Cudré-Mauroux
Pick-a-Crowd: Tell me what you like and I'll tell you what to do. WWW. 2013.

Task Routing

170 Registered

12K
Pages

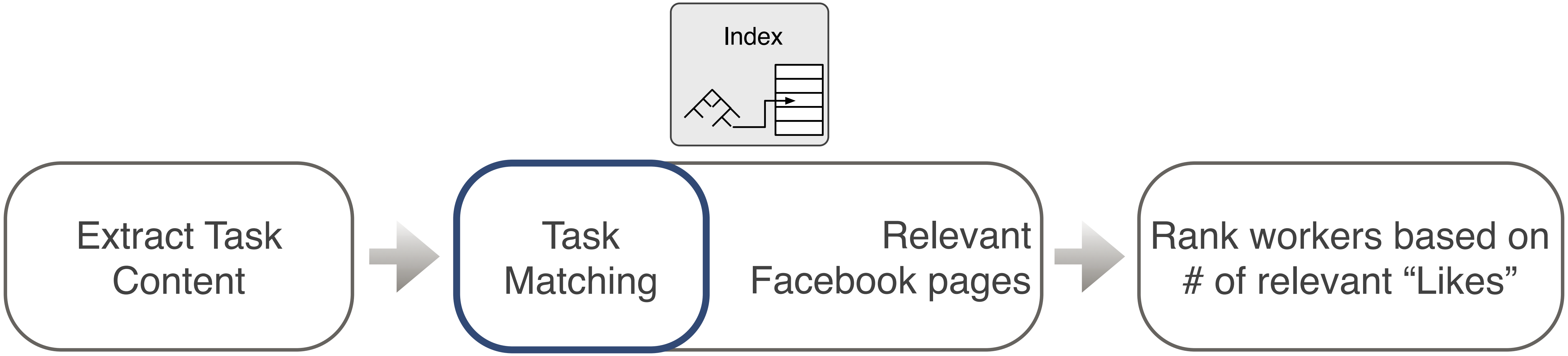
Category
Title
Description

Index

Profile Database

Difallah, Demartini, Cudré-Mauroux
Pick-a-Crowd: Tell me what you like and I'll tell you what to do. WWW. 2013.

Task Routing



Index

Extract Task Content

Task Matching

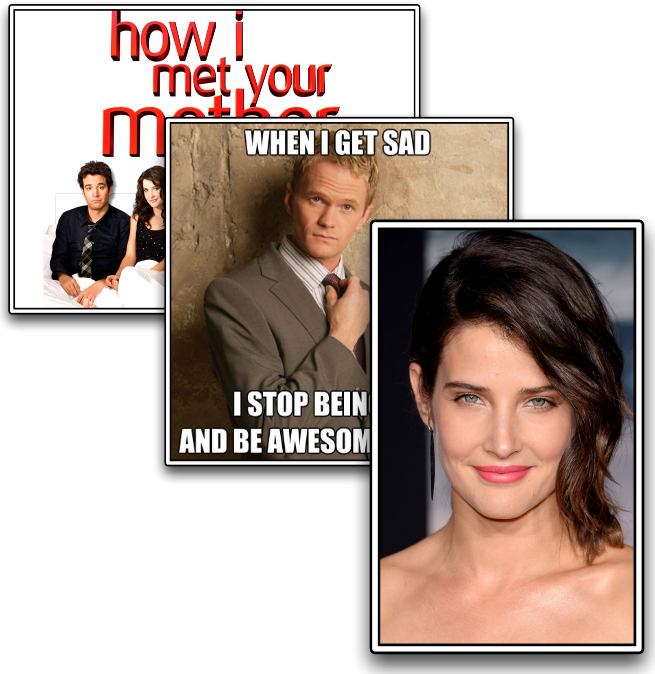
Relevant Facebook pages

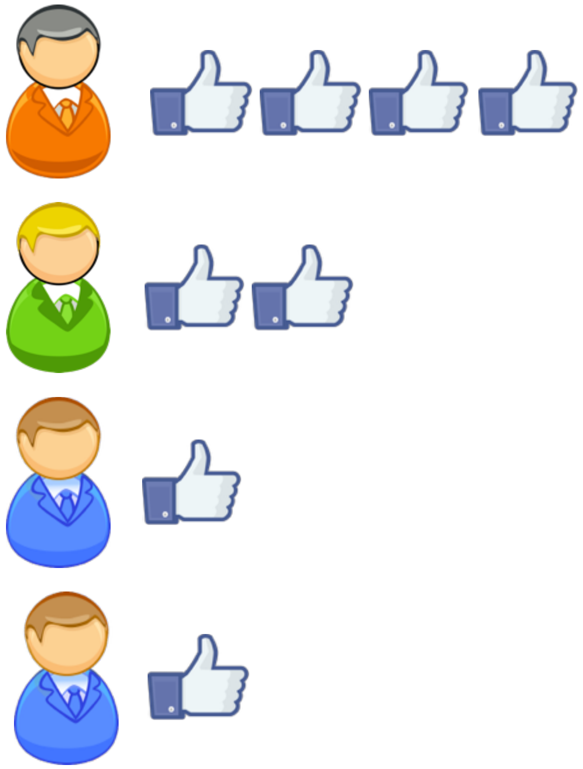
Rank workers based on # of relevant "Likes"

Input Example

Title: Actor Identification
Description: Identify Actor from the TV show "How I Met Your Mother"

Contextual:
- Neil Patrick Harris
- Cobie Smulders

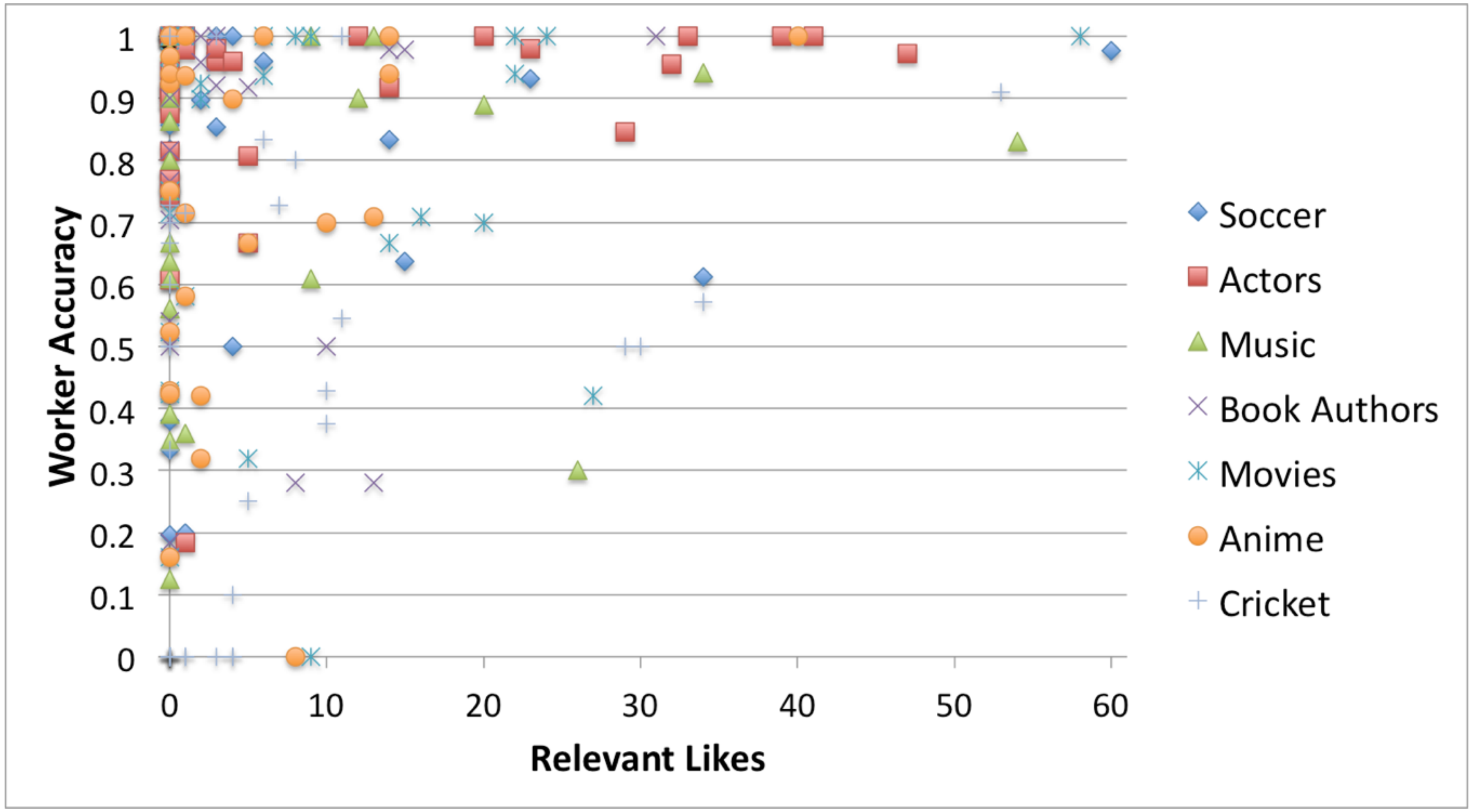




Difallah, Demartini, Cudré-Mauroux
Pick-a-Crowd: Tell me what you like and I'll tell you what to do. WWW. 2013.

Task Routing

- Workers who like more than 40 pages related to the task category have high accuracy



Difallah, Demartini, Cudré-Mauroux
Pick-a-Crowd: Tell me what you like and I'll tell you what to do. WWW. 2013.

ARE ALL TASKS THE SAME?
NO.

Answering Memory Queries
using **Transactive** Search

MEM0R1ES

“A **transactive memory** system is a mechanism through which groups collectively encode, store, and retrieve knowledge.”

Wikipedia

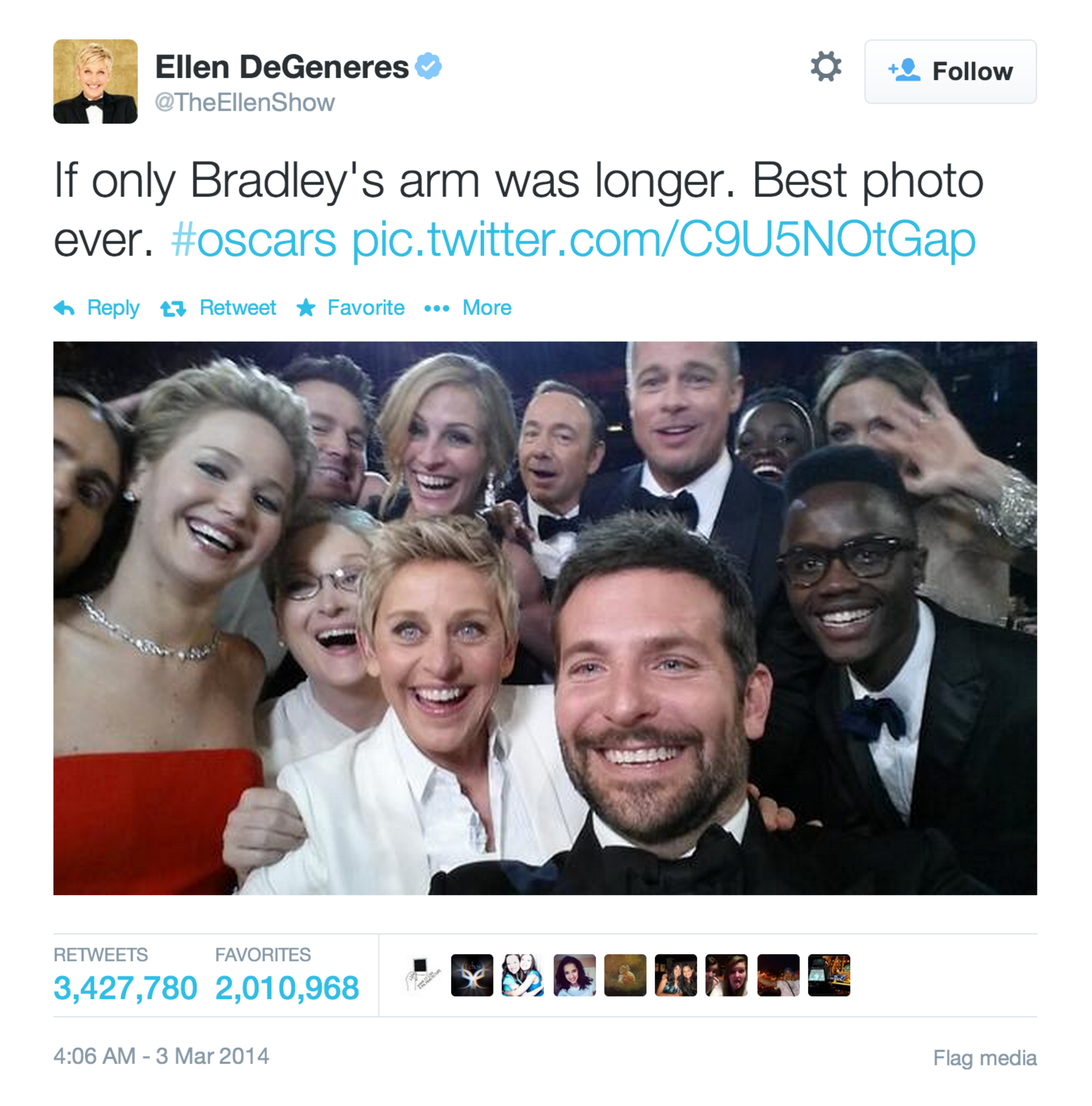
“[…] a memory system that is more complex and potentially more effective than that of any of its individual constituents.”

Wikipedia

A **transactive search** system discovers and aggregates the information stored in a transactive memory.

INFORMATION NEED

reconstruct the attendees' list of the 86th Academy Awards (2014)

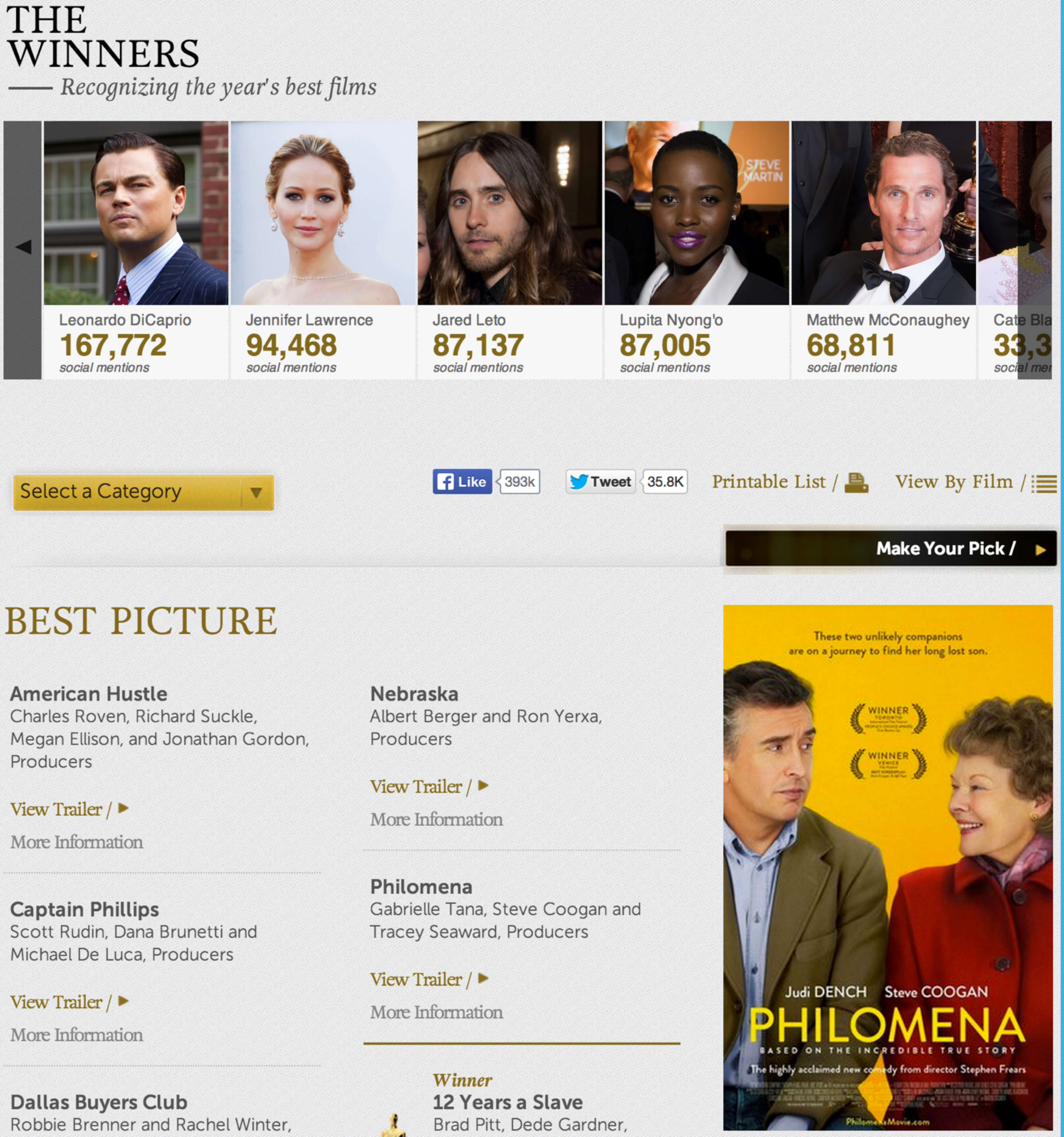


Ellen DeGeneres
@TheEllenShow

If only Bradley's arm was longer. Best photo ever. #oscars pic.twitter.com/C9U5NOtGap

RETWEETS 3,427,780
FAVORITES 2,010,968

4:06 AM - 3 Mar 2014



THE WINNERS

— *Recognizing the year's best films*

Leonardo DiCaprio	167,772	social mentions
Jennifer Lawrence	94,468	social mentions
Jared Leto	87,137	social mentions
Lupita Nyong'o	87,005	social mentions
Matthew McConaughey	68,811	social mentions
Cate Bla	33,3	social mer

BEST PICTURE

American Hustle
Charles Roven, Richard Suckle, Megan Ellison, and Jonathan Gordon, Producers

View Trailer / ▶

More Information

Captain Phillips
Scott Rudin, Dana Brunetti and Michael De Luca, Producers

View Trailer / ▶

More Information

Dallas Buyers Club
Robbie Brenner and Rachel Winter,

Nebraska
Albert Berger and Ron Yerxa, Producers

View Trailer / ▶

More Information

Philomena
Gabrielle Tana, Steve Coogan and Tracey Seaward, Producers

View Trailer / ▶

More Information

Winner
12 Years a Slave
Brad Pitt, Dede Gardner,

These two unlikely companions are on a journey to find her long lost son.
WINNER
WINNER
Judi DENCH Steve COOGAN
PHILOMENA
BASED ON THE INCREDIBLE TRUE STORY
The highly acclaimed new comedy from director Stephen Frears
PhilomenaMovie.com

MISTAKES: not all the nominees participate to the ceremony

PRECISION :-(

MISSING ENTRIES: what about all the people working “behind the scenes”?

RECALL :-(

FROM THE IDEA...

- for data that is stored in the memories of a group of people, the current query strategies are **suboptimal**
- we need a new form of human computation, different from standard crowdsourcing (i.e., no anonymous crowd)

“A taxonomy of Web Search” — A. Broder (2002)

Navigational: The immediate intent is to reach a particular Web site.

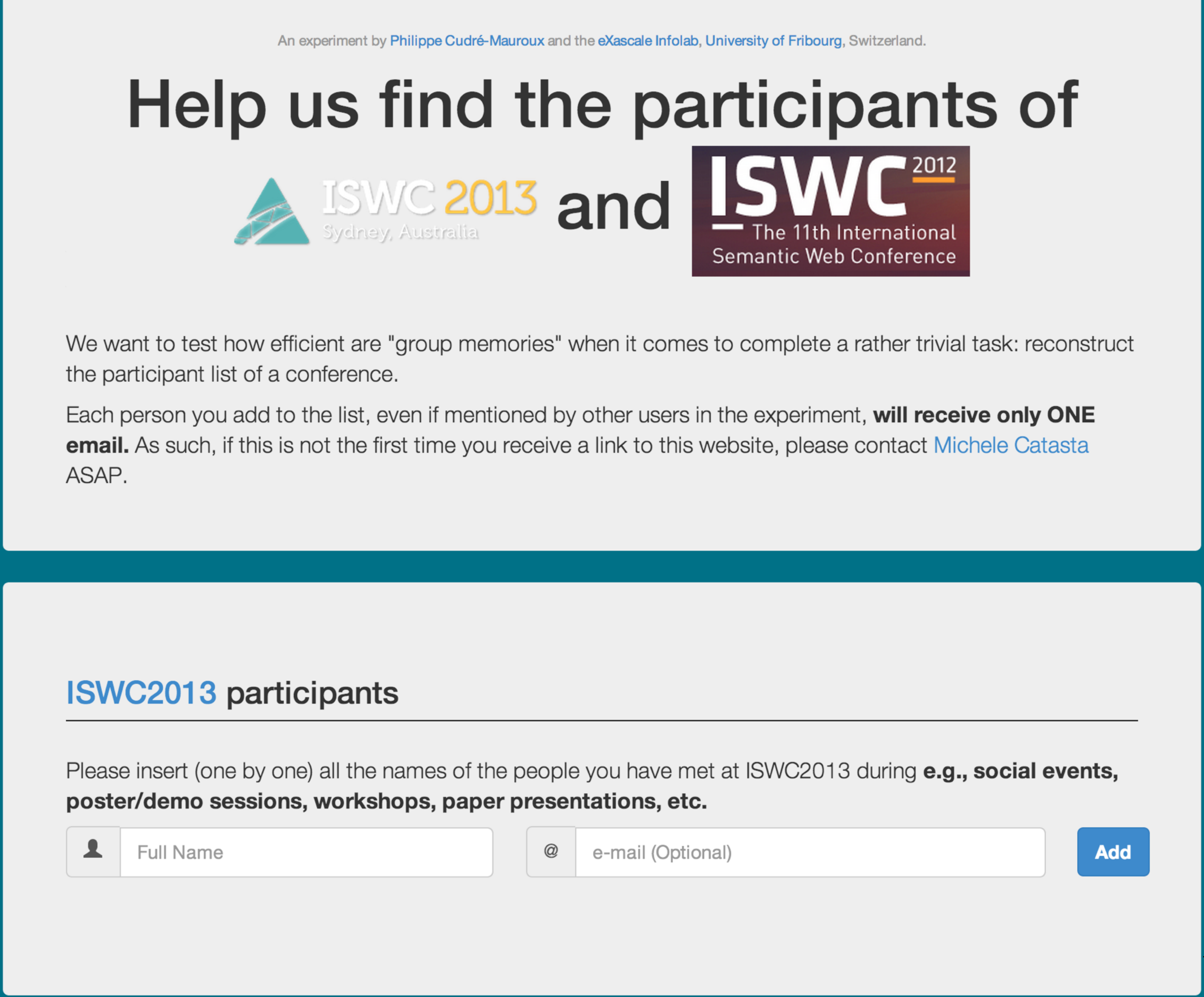
Informational: The intent is to acquire some information assumed to be present on one or more Web pages.

Transactional: The intent is to perform some Webmediated activity.

Transactive: The intent is to acquire some information that can be reconstructed **only** by an [ephemeral] social network.

…TO THE TESTING ENVIRONMENT

- We want to **reconstruct the attendees list** of two Semantic Web conferences, ISWC2012 and ISWC2013
- We were given access to the ground truth but, in general, such lists are not publicly available
- Additional data sources: authors list (first author, last author, etc.), mentions in Online Social Networks



An experiment by Philippe Cudré-Mauroux and the eXascale Infolab, University of Fribourg, Switzerland.

Help us find the participants of

ISWC 2013 Sydney, Australia and ISWC 2012 The 11th International Semantic Web Conference

We want to test how efficient are "group memories" when it comes to complete a rather trivial task: reconstruct the participant list of a conference.

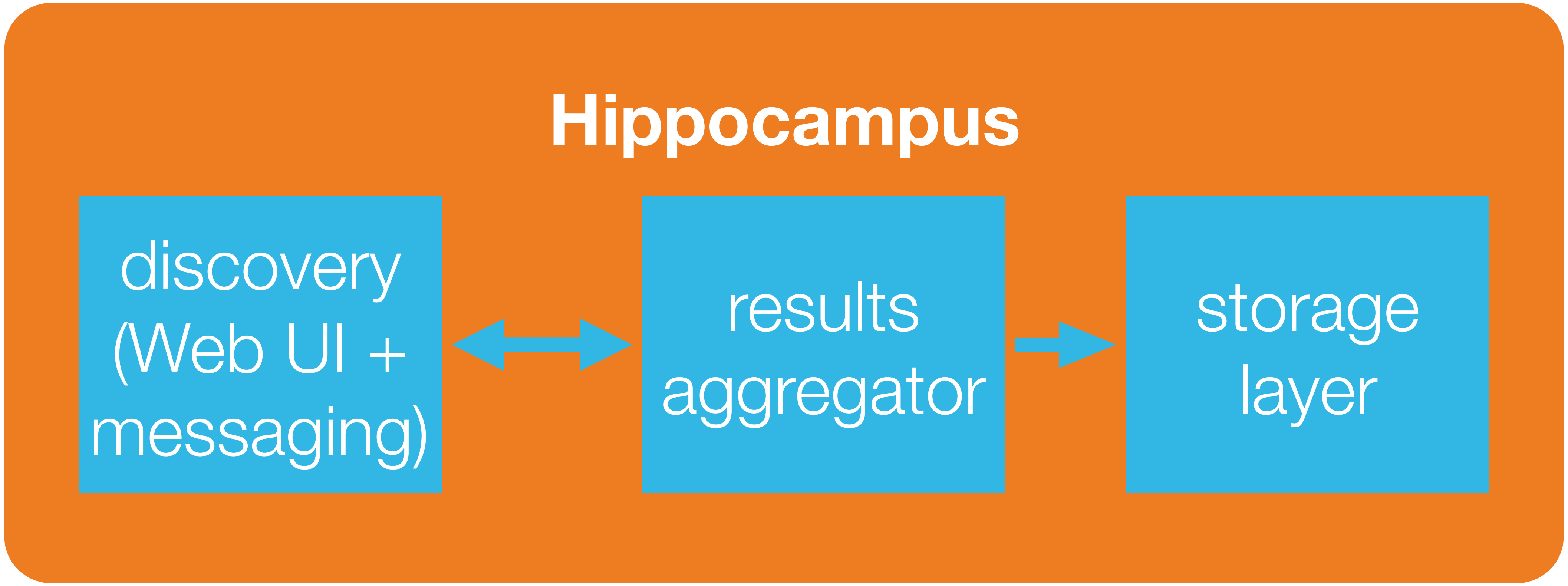
Each person you add to the list, even if mentioned by other users in the experiment, **will receive only ONE email.** As such, if this is not the first time you receive a link to this website, please contact Michele Catasta ASAP.

ISWC2013 participants

Please insert (one by one) all the names of the people you have met at ISWC2013 during **e.g., social events, poster/demo sessions, workshops, paper presentations, etc.**

EXPERIMENT ARCHITECTURE

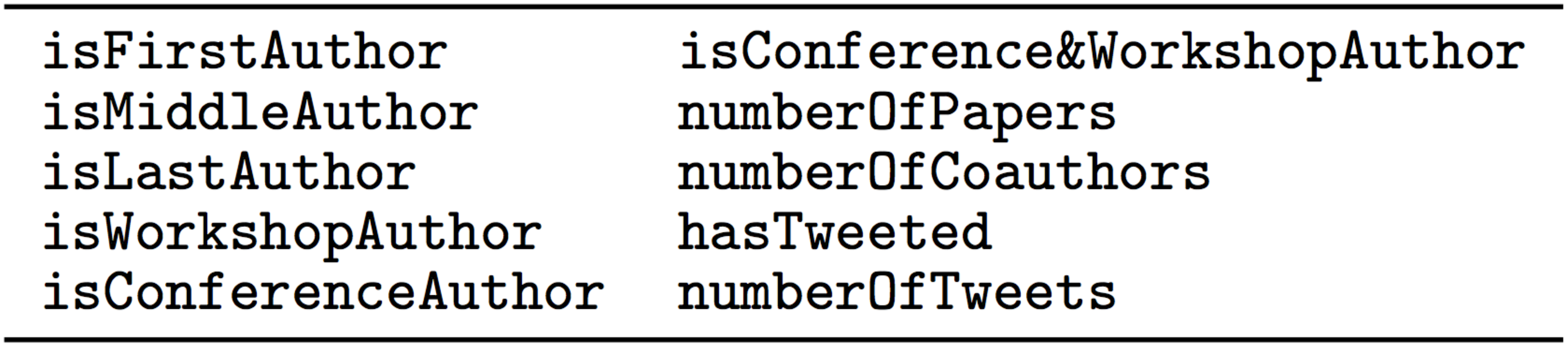
- tailored Web UI + results aggregator
- iterative reconstruction: every time a new person was mentioned, Hippocampus sent her an invitation to contribute to the attendees list



MACHINE LEARNING APPROACHES

- we collected the proceedings information and all the tweets with the conference hashtags

- we trained state-of-the-art classifiers with these features:



- isFirstAuthor
- isMiddleAuthor
- isLastAuthor
- isWorkshopAuthor
- isConferenceAuthor
- isConference&WorkshopAuthor
- numberOfPapers
- numberOfCoauthors
- hasTweeted
- numberOfTweets

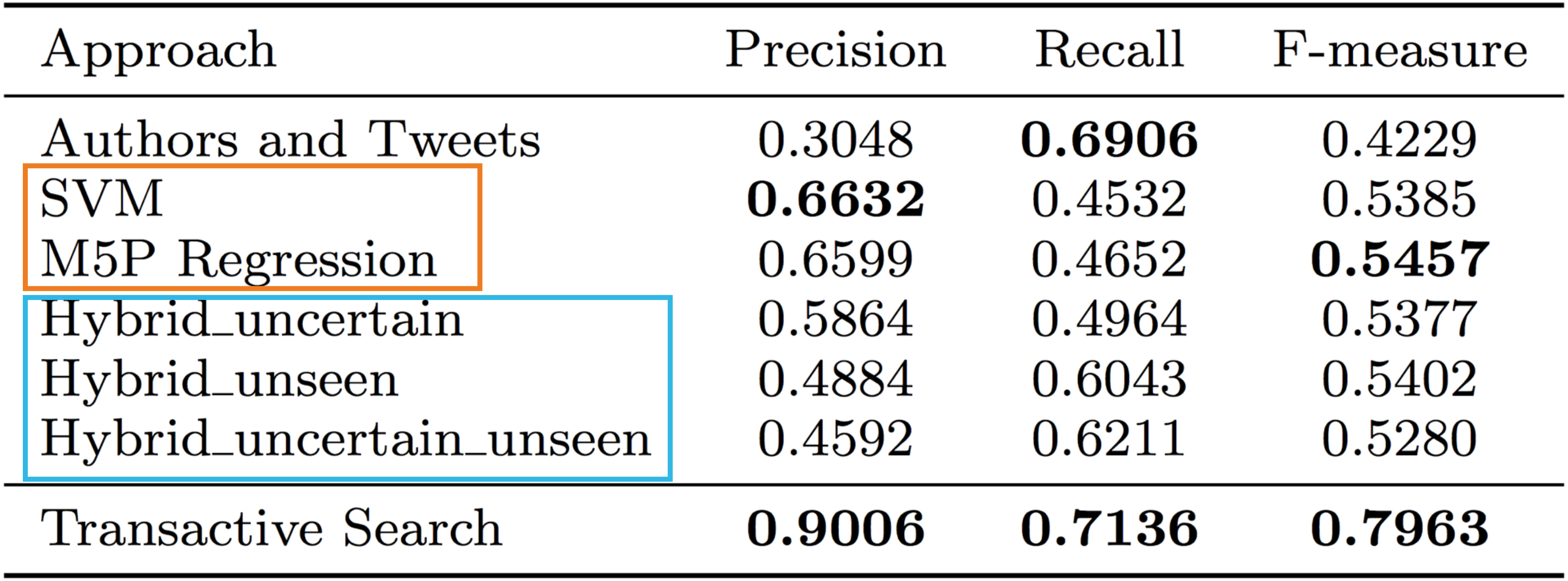
not possible without the ground truth!

ML + CROWDSOURCING APPROACHES

- **Uncertain cases** (precision): we asked the crowd to revise the low-confidence results of the ML classifier. (e.g., people that didn't attend the conference but tweeted about it)

- **Unseen cases** (recall): we asked the crowd to actively look for attendees not included in the authors list (e.g., organizers mentioned in the Web site)

the crowd has access **only** to public data on the Web!



Approach	Precision	Recall	F-measure
Authors and Tweets	0.3048	**0.6906**	0.4229
SVM	**0.6632**	0.4532	0.5385
M5P Regression	0.6599	0.4652	**0.5457**
Hybrid_uncertain	0.5864	0.4964	0.5377
Hybrid_unseen	0.4884	0.6043	0.5402
Hybrid_uncertain_unseen	0.4592	0.6211	0.5280
Transactive Search	**0.9006**	**0.7136**	**0.7963**

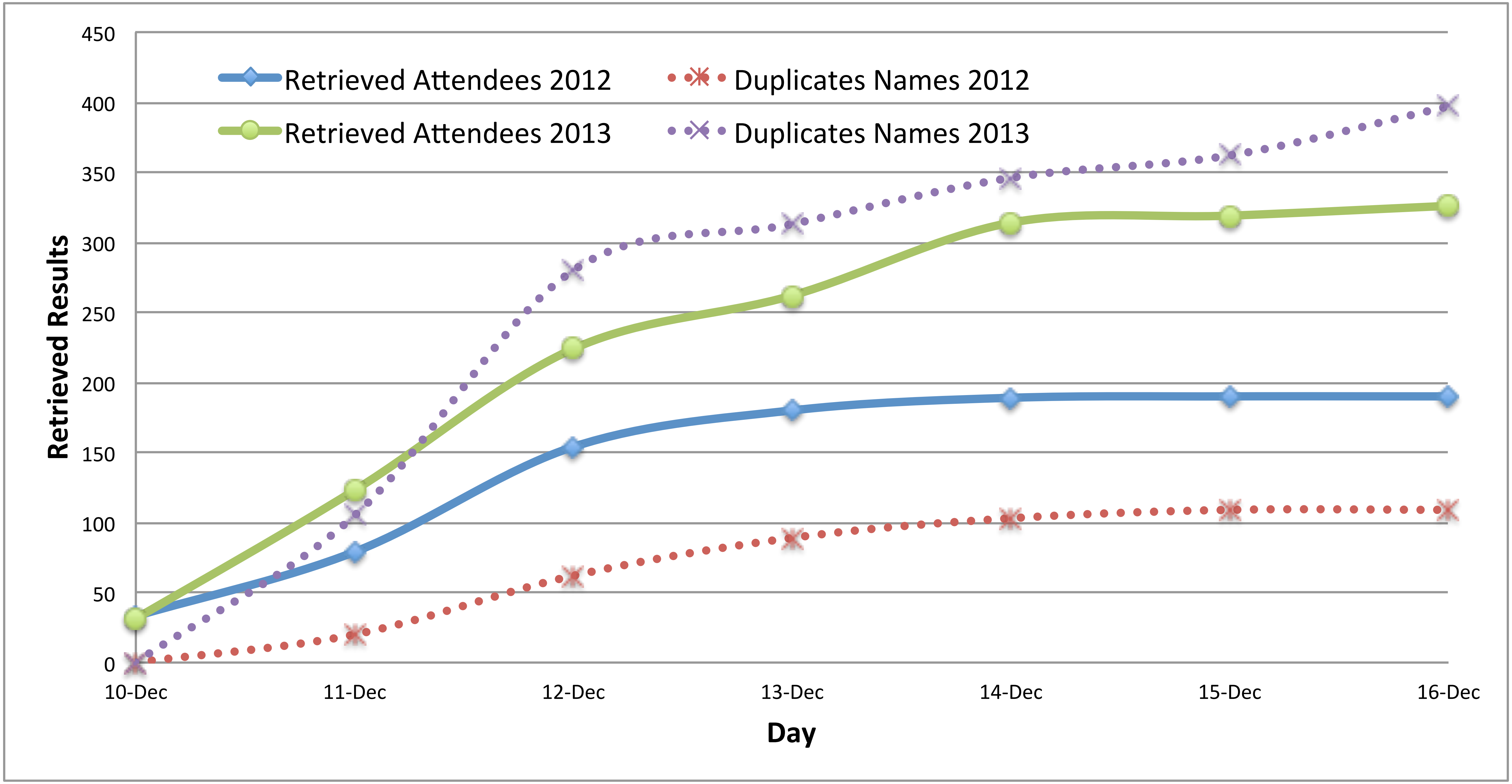
Authors and Tweets: baseline (exhaustive list of authors and twitterers)

Machine Learning: SVM, M5P Regression

Machine Learning + Crowdsourcing: Hybrid_(uncertain, unseen, uncertain_unseen)

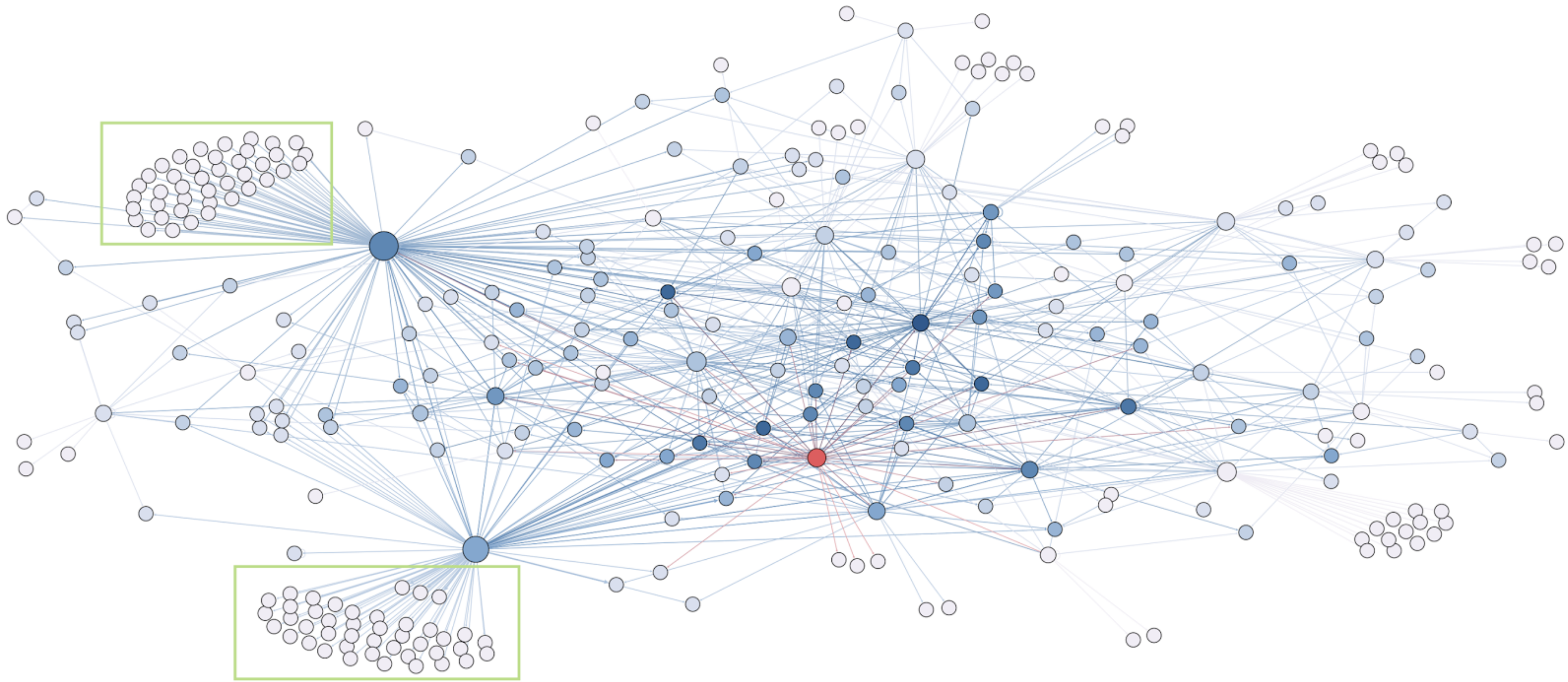
Transactive vs ML & Crowdsourcing

ISWC 2013



attendees found over time

Transactive Search



Transactive Memory Graph

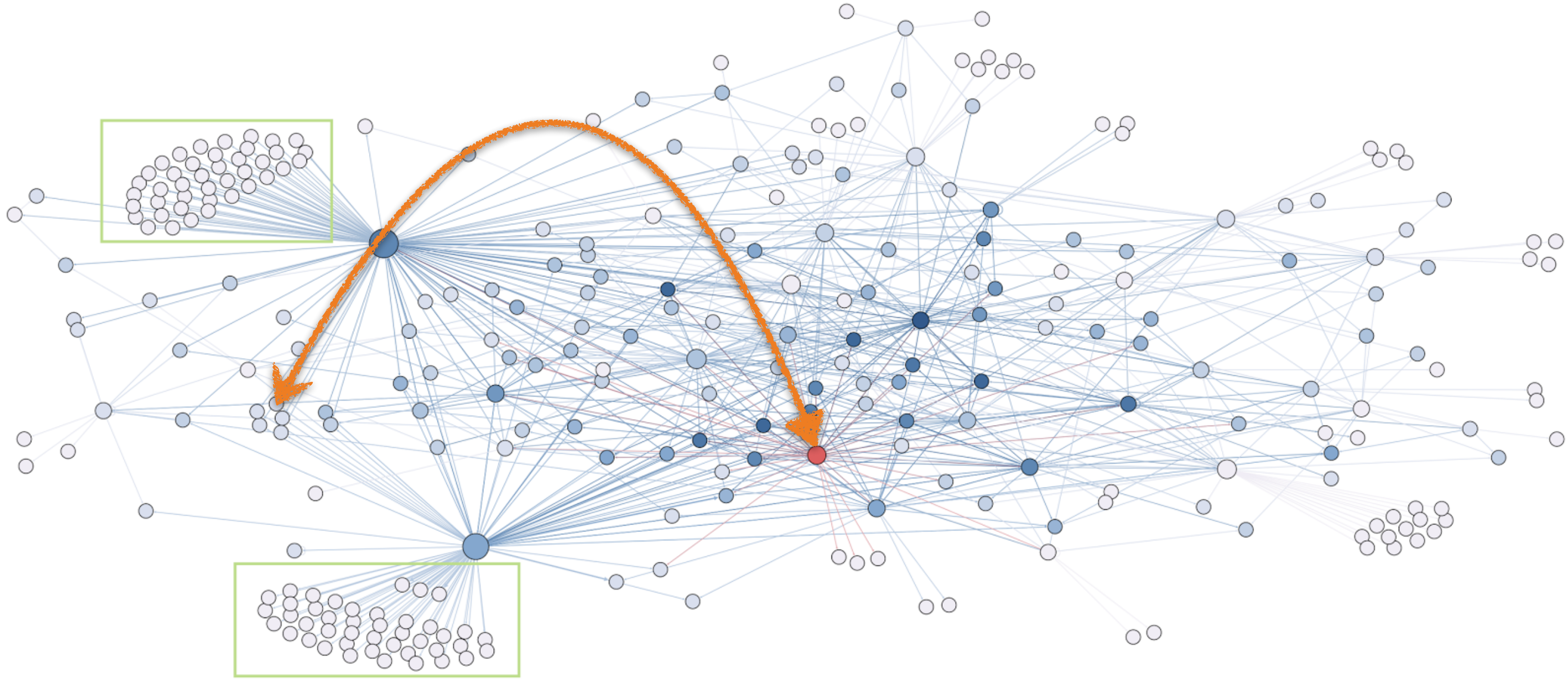
in green, two isolated “components” discovered by top-contributors

Result discussion

- for a specific class of queries, our **Transactive Search** performs up to **46%** better than the best alternative approach (i.e., Machine Learning + Crowdsourcing)

- we will explore incentives for Hippocampus, as it is currently **two orders of magnitude slower** than the alternative approaches

- we reported some initial evidences that, as human memories fade with time, our approach **works best with recent events**



Transactive Point Queries

what if the information need can be served only by one/few nodes?

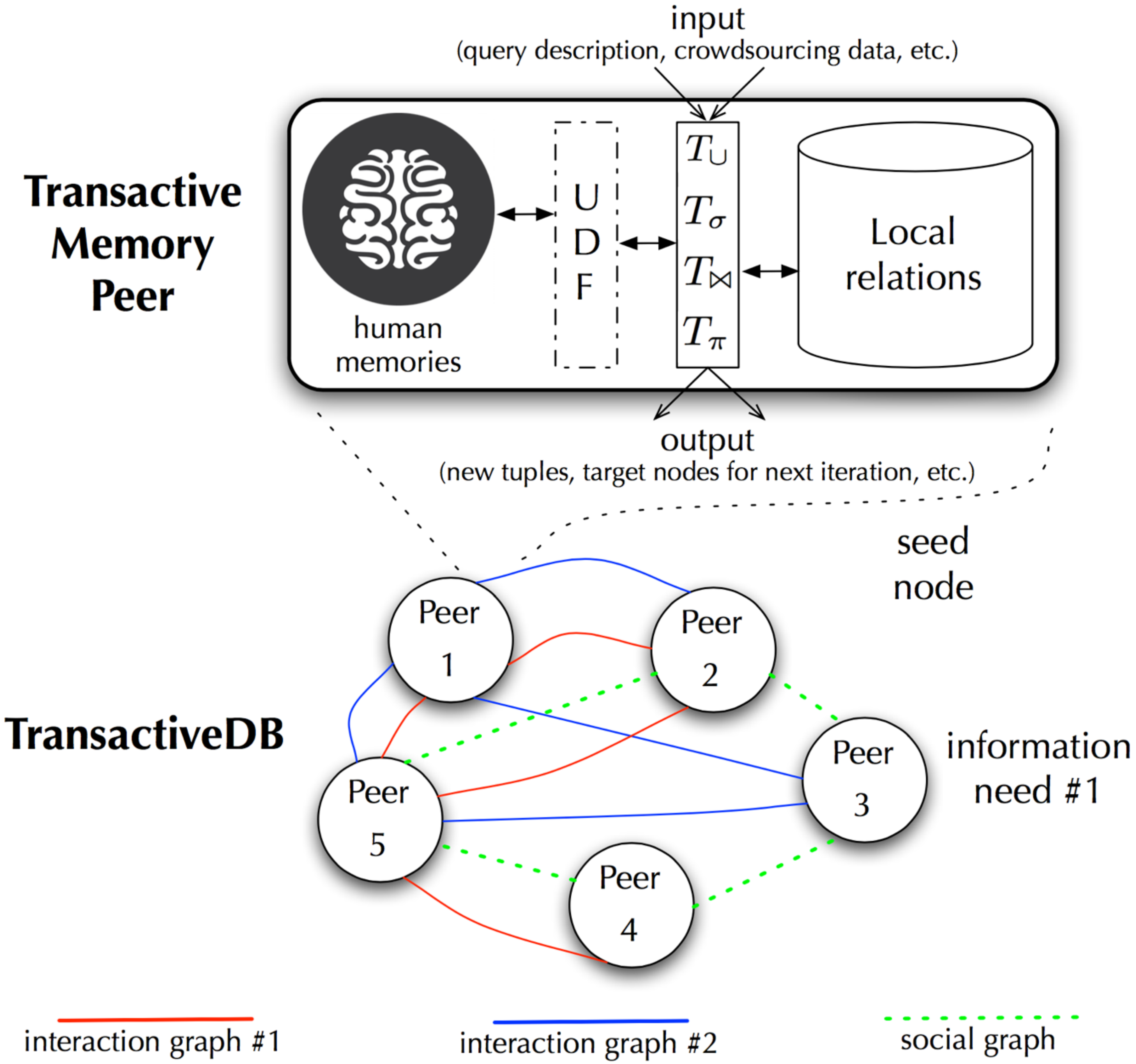
What is the name of the delicious cocktail I had during last year's gala dinner?

This information need can be unlikely satisfied by:
- a **Web search** (i.e., the conference website does not contain such information)
- a **DB query** (i.e., the transactions of the restaurant are private)
- a **crowdsourcing task** (i.e., the anonymous crowd did not participate to the conference)

But (some of) the attendees of the conference could work collectively and come up with an answer

Tapping into Collective Human Memories

- **TransactiveDB:** a **decentralized** data management system that elicits and processes memories of individuals or groups in order to answer transactive queries
- **Node:** classical DBMS + transactive operators handling the memories of a particular user (i.e., personal events, contextual data, etc.)
- **Interaction graph:** a subset of the underlying social network connecting different end-users, corresponding to a specific context (e.g., social event, family setting, etc.)



Architecture

Crowdsourcing for Social Media

- Like in many other scientific fields, crowdsourcing is playing a key role in social media research
- ICWSM2016 proceedings:
  - 37 mentions of “crowdsourcing”
  - 30 mentions of “Mechanical Turk”
  - 6 mentions of “Crowdflower”

- **DATA GATHERING:** “We first employed crowdsourcing to **collect Twitter users**’ cognitive styles using standard psychometric instruments”

- **SENTIMENT:** “Through a crowdsourcing study, we show that there are marked differences between the overall **tweet sentiment** and the sentiment expressed towards the subjects mentioned in tweets related to three crises events.”

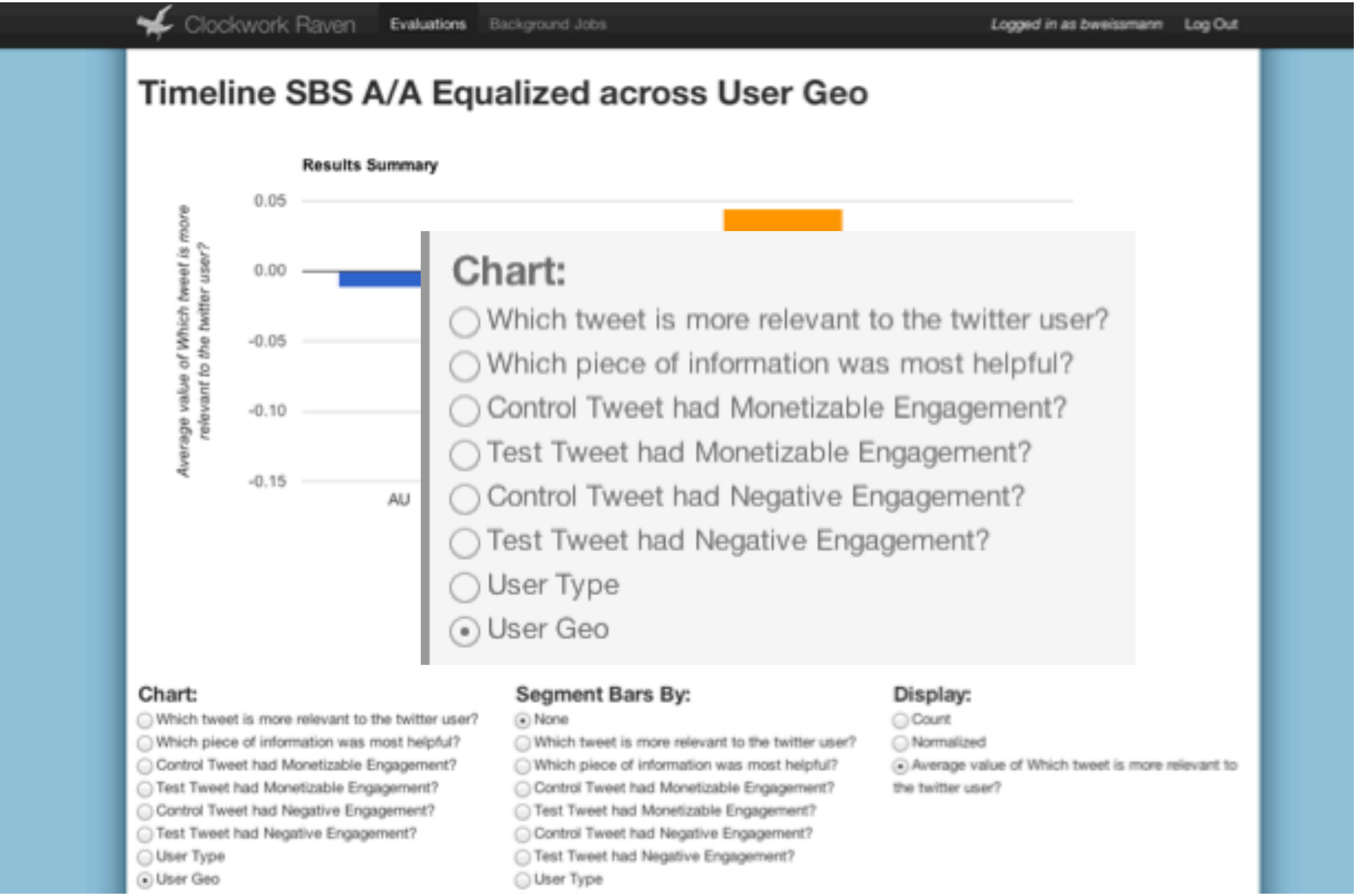
- **LEXICON:** “We built **lexical categories that capture this list of stereotypes** by mapping the 2000 most commonly occurring verbs and adjectives in our dataset onto the set of categories through a series of crowdsourcing tasks.”

- **VALIDATION:** ”**To calibrate and validate this measure**, we turn to crowdsourcing labels on Amazon Mechanical Turk. The results reveal that cosine distance is a strong predictor of similarity.”

CS excels in understanding human nuances

sentiment, sarcasm, jargon, etc.

CS @Twitter



Timeline SBS A/A Equalized across User Geo

Results Summary
Average value of Which tweet is more relevant to the twitter user?
0.05
0.00
-0.05
-0.10
-0.15
AU

Chart:

○ Which tweet is more relevant to the twitter user?
○ Which piece of information was most helpful?
○ Control Tweet had Monetizable Engagement?
○ Test Tweet had Monetizable Engagement?
○ Control Tweet had Negative Engagement?
○ Test Tweet had Negative Engagement?
○ User Type
● User Geo

Chart:

○ Which tweet is more relevant to the twitter user?
○ Which piece of information was most helpful?
○ Control Tweet had Monetizable Engagement?
○ Test Tweet had Monetizable Engagement?
○ Control Tweet had Negative Engagement?
○ Test Tweet had Negative Engagement?
○ User Type
● User Geo

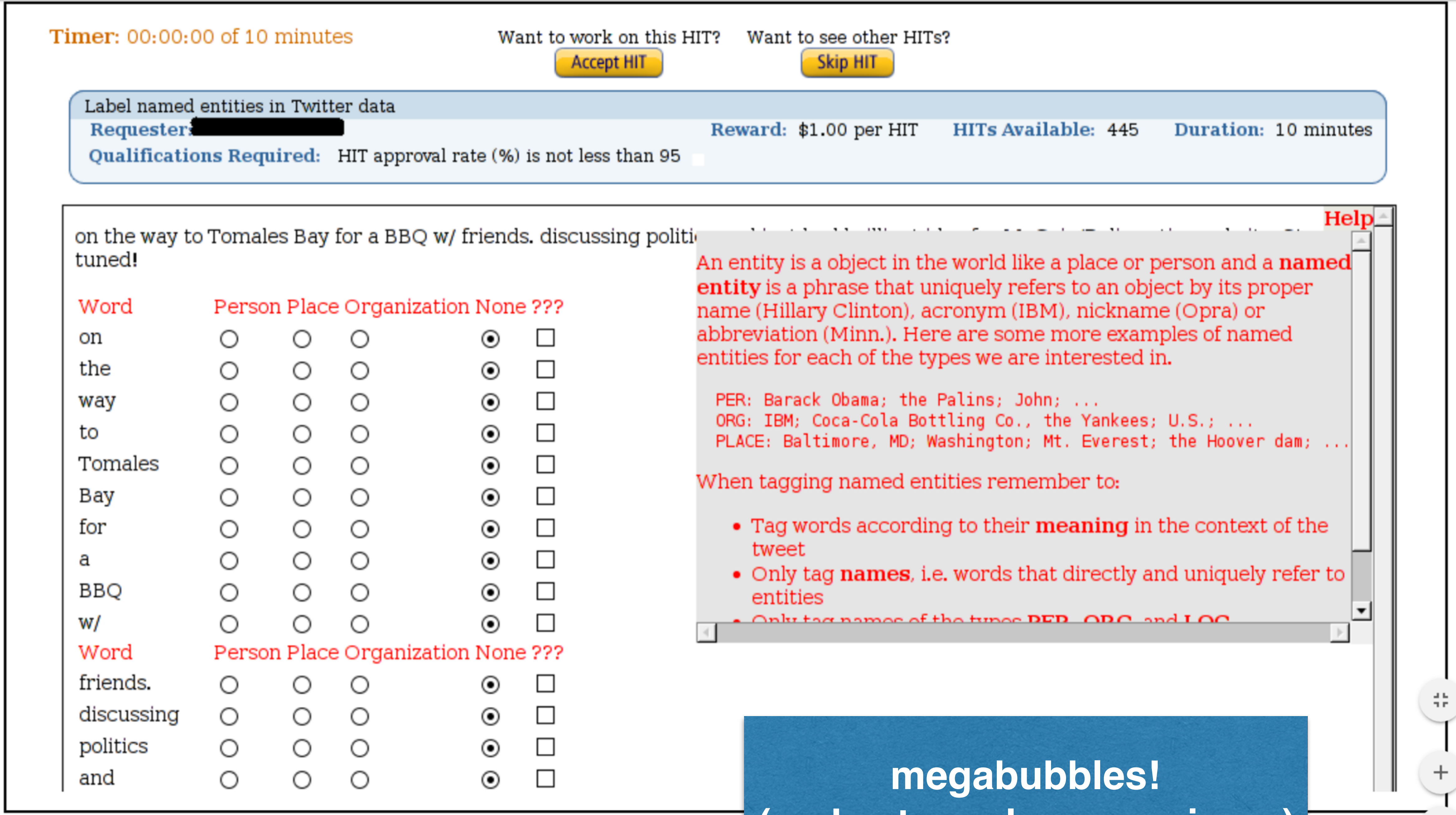
Segment Bars By:

● None
○ Which tweet is more relevant to the twitter user?
○ Which piece of information was most helpful?
○ Control Tweet had Monetizable Engagement?
○ Test Tweet had Monetizable Engagement?
○ Control Tweet had Negative Engagement?
○ Test Tweet had Negative Engagement?
○ User Type

Display:

○ Count
○ Normalized
● Average value of Which tweet is more relevant to the twitter user?

Annotate entities in Tweets (2010)



Timer: 00:00:00 of 10 minutes

Want to work on this HIT? Accept HIT

Want to see other HITs? Skip HIT

Label named entities in Twitter data

Requester: [redacted]

Reward: $1.00 per HIT

HITs Available: 445

Duration: 10 minutes

Qualifications Required: HIT approval rate (%) is not less than 95

on the way to Tomales Bay for a BBQ w/ friends. discussing politi
tuned!

Word	Person	Place	Organization	None	???
on	○	○	○	●	□
the	○	○	○	●	□
way	○	○	○	●	□
to	○	○	○	●	□
Tomales	○	○	○	●	□
Bay	○	○	○	●	□
for	○	○	○	●	□
a	○	○	○	●	□
BBQ	○	○	○	●	□
w/	○	○	○	●	□
Word	Person	Place	Organization	None	???
friends.	○	○	○	●	□
discussing	○	○	○	●	□
politics	○	○	○	●	□
and	○	○	○	●	□

Help

An entity is a object in the world like a place or person and a **named entity** is a phrase that uniquely refers to an object by its proper name (Hillary Clinton), acronym (IBM), nickname (Opra) or abbreviation (Minn.). Here are some more examples of named entities for each of the types we are interested in.

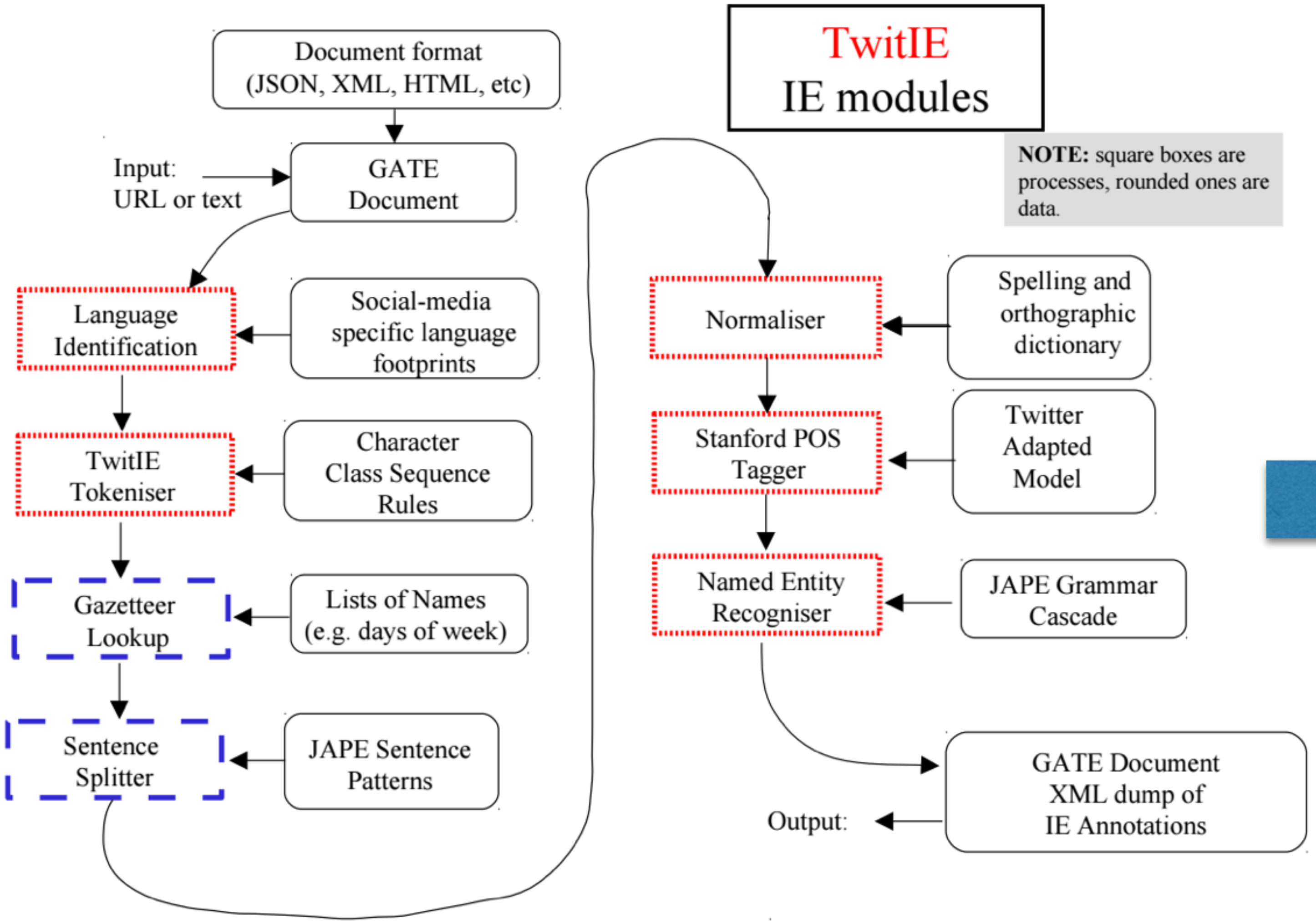
```
PER: Barack Obama; the Palins; John; ...
ORG: IBM; Coca-Cola Bottling Co., the Yankees; U.S.; ...
PLACE: Baltimore, MD; Washington; Mt. Everest; the Hoover dam; ...
```

When tagging named entities remember to:

- Tag words according to their **meaning** in the context of the tweet
- Only tag **names**, i.e. words that directly and uniquely refer to entities
- Only tag names of the types **PER**, **ORG**, and **LOC**

megabubbles!
(and extremely expensive…)

Annotate entities in Tweets (today)


TwitIE
IE modules
NOTE: square boxes are processes, rounded ones are data.
Document format (JSON, XML, HTML, etc)
Input: URL or text
GATE Document
Language Identification
Social-media specific language footprints
TwitIE Tokeniser
Character Class Sequence Rules
Gazetteer Lookup
Lists of Names (e.g. days of week)
Sentence Splitter
JAPE Sentence Patterns
Normaliser
Spelling and orthographic dictionary
Stanford POS Tagger
Twitter Adapted Model
Named Entity Recognizer
JAPE Grammar Cascade
GATE Document XML dump of IE Annotations
Output:

NLP pipeline with 80% accuracy in NER on Tweets

low confidence results go to the crowd

- faster
- slash the costs

Virtual Labs (Duncan Watts)

- “big and thin” vs “small and rich”
- SurveyMonkey/Google Forms in a crowdsourcing platform: scale up N of subjects
- what about the need for synchronicity?