

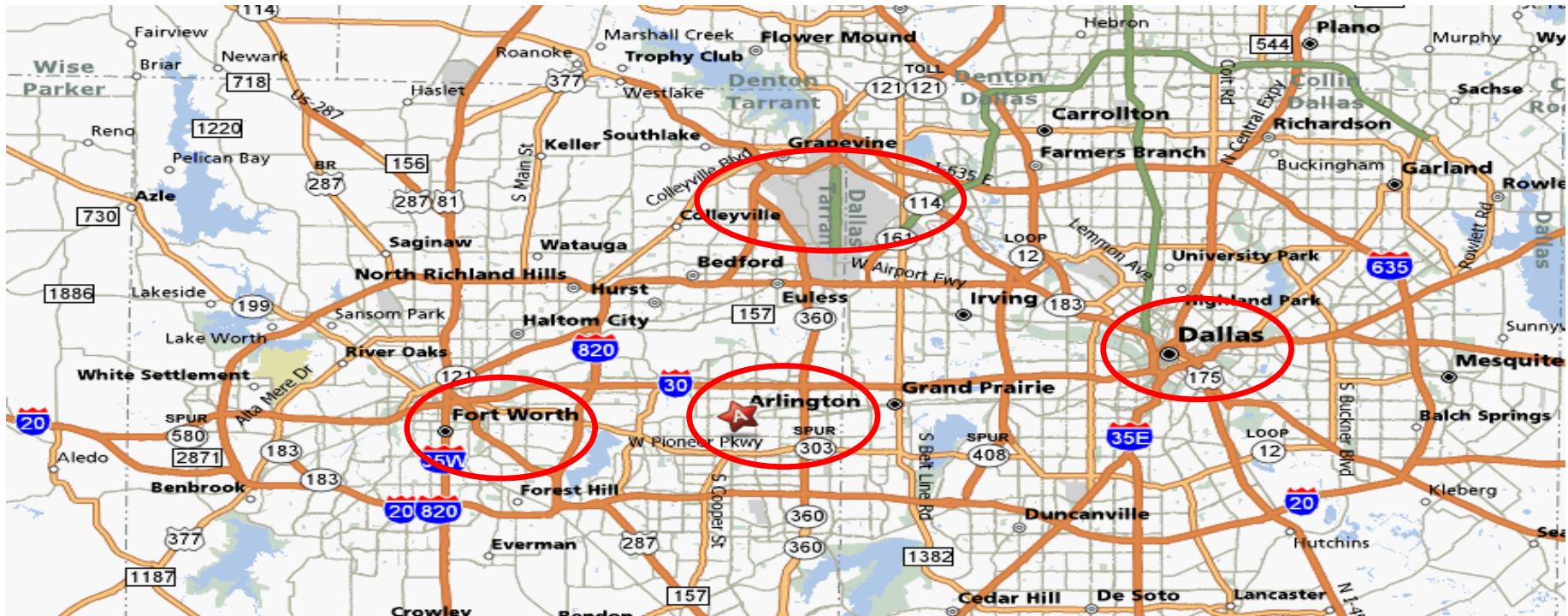
# Enabling Computational Journalism: Automated Fact-Checking and Story-Finding

Chengkai Li

Associate Professor, Department of Computer Science and Engineering  
Director, Innovative Database and Information Systems Research (IDIR) Laboratory  
University of Texas at Arlington

Shandong University, Oct. 9<sup>th</sup>, 2015  
Nanjing University, Oct. 13<sup>th</sup>, 2015

# Dallas-Fort Worth Metroplex, Texas



©2015 The University of Texas at Arlington. All Rights Reserved.



# The University of Texas at Arlington

“The University of Texas at Arlington is a growing research powerhouse committed to life-enhancing discovery, innovative instruction, and caring community engagement. An educational leader in the heart of the thriving North Texas region, UT Arlington nurtures minds within an environment that values excellence, ingenuity, and diversity.”

- Founded in 1895
- One of the 9 academic institutions in The University of Texas System
- 51,000 students (37,008 on-campus), second-largest institution within the UT System
- Classified by the Carnegie Foundation as a “Research University/High Research Activity” institution
- One of the fastest growing public research universities in the nation



# Department of Computer Science and Engineering

- 29 tenured and tenure-track faculty, 11 non-tenure track faculty, 17 adjunct faculty
- ~110 Ph.D. students, ~800 M.S. students, ~1000 B.S. students

## Ranking

- #19 Data Mining, #19 Machine Learning and Pattern Recognition, #35 Databases (source: Microsoft Academic Search, 2014)
- Computer Engineering (Graduate) #83 (source: U.S. News, 2016)
- Computer Science (Graduate) #90 (source: U.S. News, 2014)

**Major areas of research:** Big Data and Large-Scale Computing, Biocomputing and Health Informatics, Computer Networks, Computer Vision and Multimedia, Database and Information Systems, Embedded Systems and Mobile Computing, Machine Learning and Data Mining, Robotics and AI, Security and Privacy, Software Engineering, Sustainable Computing



# The Innovative Database and Information Systems Research (IDIR) Laboratory

## Research areas

- o Big Data and Data Science (Database, Data Mining, Web Data Management, Information Retrieval)

## Theme of current research

- o building large-scale human-assisting and human-assisted data and information systems with high usability, low cost and applications for social good

## Research directions

- o computational journalism
- o crowdsourcing and human computation
- o data exploration by ranking/skyline/preference queries
- o database testing
- o entity search and entity query
- o graph database usability



# Our Computational Journalism Project

Started in 2010. Collaborative project with Duke, Google Research, HP Labs, Stanford

- **Story finding:** finding and monitoring number-based facts pertinent to real-world events. The facts are leads to news stories.
- **Fact checking:** discovering and checking factual claims in political discourses, social media, and news.



# Publications

- Detecting Check-worthy Factual Claims in Presidential Debates. Naeemul Hassan, Chengkai Li, Mark Tremayne. CIKM 2015, pages 1835-1838.
- The Quest to Automate Fact-Checking Naeemul Hassan, Bill Adair, James Hamilton, Chengkai Li, Mark Tremayne, Jun Yang and Cong Yu. 2015 Computation+Journalism Symposium.
- Online Frequent Episode Mining Xiang Ao, Ping Luo, Chengkai Li, Fuzhen Zhuang, and Qing He. ICDE 2015, pages 891-902.
- Data In, Fact Out: Automated Monitoring of Facts by FactWatcher. Naeemul Hassan, Afroza Sultana, You Wu, Gensheng Zhang, Chengkai Li, Jun Yang, and Cong Yu. VLDB 2014, pages 1557-1560. Demonstration description. (**excellent demonstration award**)
- Finding, Monitoring, and Checking Claims Computationally Based on Structured Data. Brett Walenz, You (Will) Wu, Seokhyun (Alex) Song, Emre Sonmez, Eric Wu, Kevin Wu, Pankaj K. Agarwal, Jun Yang, Naeemul Hassan, Afroza Sultana, Gensheng Zhang, Chengkai Li, Cong Yu. 2014 Computation+Journalism Symposium.



# Publications (cont'd)

- Toward Computational Fact-Checking. You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, Cong Yu. VLDB 2014, pages 589-600.
- iCheck: computationally combating "lies, d-ned lies, and statistics". You Wu, Brett Walenz, Peggy Li, Andrew Shim, Emre Sonmez, Pankaj K. Agarwal, Chengkai Li, Jun Yang, Cong Yu. SIGMOD 2014, pages 1063-1066.
- Incremental Discovery of Prominent Situational Facts. Afroza Sultana, Naeemul Hassan, Chengkai Li, Jun Yang, Cong Yu. ICDE 2014, pages 112-123.
- Discovering General Prominent Streaks in Sequence Data. Gensheng Zhang, Xiao Jiang, Ping Luo, Min Wang, Chengkai Li. ACM TKDD, 8(2):article 9, June 2014.
- Discovering and Learning Sensational Episodes of News Events. Xiang Ao, Ping Luo, Chengkai Li, Fuzhen Zhuang, Qing He, and Zhongzhi Shi. WWW 2014, pages 217-218.

# Publications (cont'd)

- On "One of the Few" Objects. You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, Cong Yu. KDD 2012, pages 1487-1495.
- Prominent Streak Discovery in Sequence Data. Xiao Jiang, Chengkai Li, Ping Luo, Min Wang, Yong Yu. KDD 2011, pages 1280-1288.
- Computational Journalism: A Call to Arms to Database Researchers. Sarah Cohen, Chengkai Li, Jun Yang, Cong Yu. CIDR 2011, pages 148-151. **(3rd place in best Outrageous Ideas and Vision (OIV) Track paper competition)**





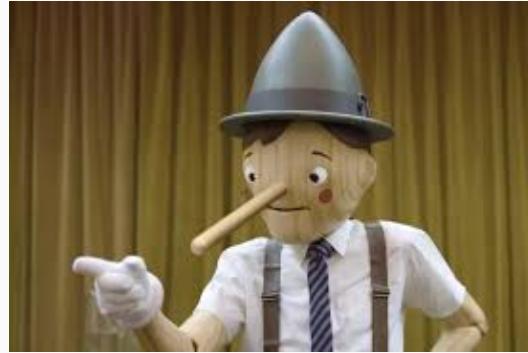
# CLAIMBUSTER

## The Quest to Automate Fact-Checking



# People Make Claims All The Time

“... our Navy is smaller than it's been since 1917”, said Republican candidate Mitt Romney in third presidential debate in 2012.



[http://en.wikipedia.org/wiki/Mitt\\_Romney](http://en.wikipedia.org/wiki/Mitt_Romney)  
<http://www.thebrainchildgroup.com/>

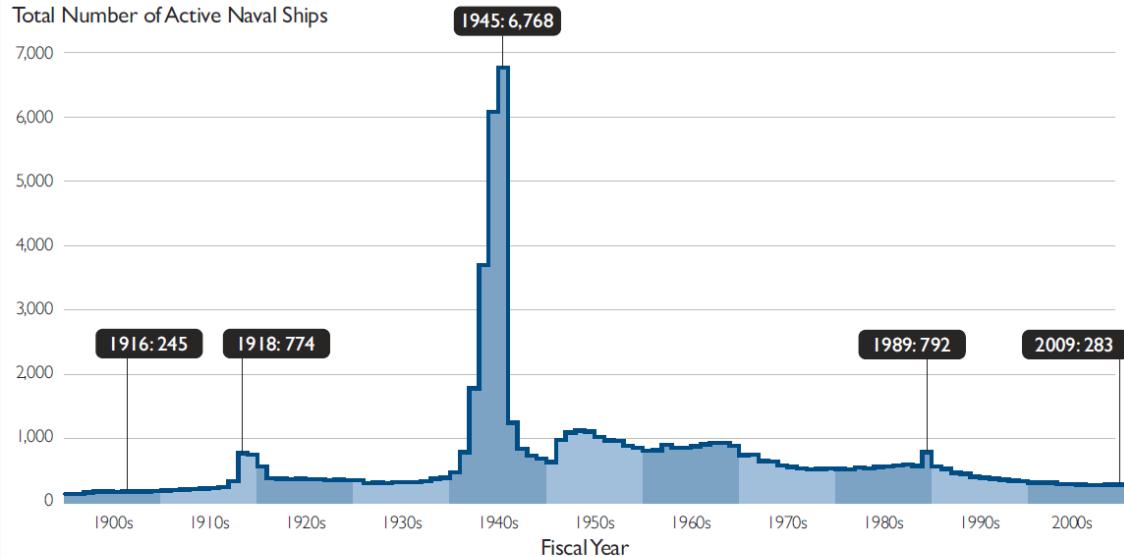
©2015 The University of Texas at Arlington. All Rights Reserved.

# Fact Checking is not Easy

“... our Navy is smaller than it's been since 1917”, said Republican candidate Mitt Romney in third presidential debate in 2012.



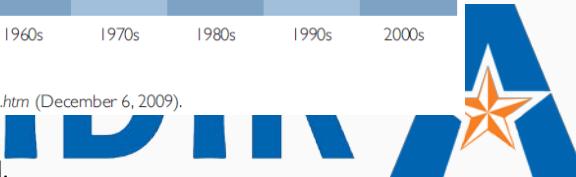
**U.S. Navy Has Smallest Number of Ships Since 1916**



[http://en.wikipedia.org/wiki/Mitt\\_Romney](http://en.wikipedia.org/wiki/Mitt_Romney)

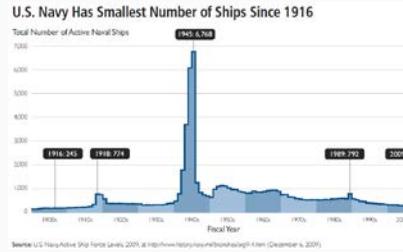
[http://s3.amazonaws.com/thf\\_media/2010/pdf/Military\\_chartbook.pdf](http://s3.amazonaws.com/thf_media/2010/pdf/Military_chartbook.pdf)

©2015 The University of Texas at Arlington. All Rights Reserved.



# Fact Checking is not Easy

“... our Navy is smaller than it's been since 1917”, said Republican candidate Mitt Romney in third presidential debate in 2012.



VS



[http://en.wikipedia.org/wiki/Mitt\\_Romney](http://en.wikipedia.org/wiki/Mitt_Romney)

[http://s3.amazonaws.com/thf\\_media/2010/pdf/Military\\_chartbook.pdf](http://s3.amazonaws.com/thf_media/2010/pdf/Military_chartbook.pdf)

[http://en.wikipedia.org/wiki/United\\_States\\_Navy](http://en.wikipedia.org/wiki/United_States_Navy)

# Existing Fact Checking Projects

Journalists and reporters spend good amount of time on fact checking



The U.S. military is at risk of losing its "military superiority" because "our Navy is smaller than it's been since 1917. Our Air Force is smaller and older than any time since 1947."



*— Mitt Romney on Monday, January 16th, 2012 in a Republican presidential debate in Myrtle Beach, S.C.*

PolitiFact <http://www.politifact.com/>

FactCheckEU <https://factcheckeu.org/>

FullFact <http://fullfact.org/>

Snopes <http://www.snopes.com/info/whatsnew.asp>

Factcheck <http://www.factcheck.org/>

# Numerous Claims to Check. Rise of Fact-Checkers

Republican candidate debate, August 6, 2015.<sup>1</sup>

9 facts checked by factcheck.org

8 facts checked by CNN

24 facts checked by PolitiFact

64 active fact-checking sites in 2015, 44 in 2014.<sup>2</sup>

1. <http://time.com/3988276/republican-debate-prime-time-transcript-full-text/>
2. <http://reporterslab.org/snapshot-of-fact-checking-around-the-world-july-2015/>



# Limitations of Current Fact-Checking Practices

- Journalists spend hours going through documents to identify claims.
- Significant time gap between speech and reporting times. Audience doesn't get correct information.
- Requires advanced writing skills to persuade readers. Such skilled writers are sparse.
- Lack of Structured Journalism and use of old publishing frameworks hinders Semantic Web applications.



# The Holy Grail: Automated, Live Fact-Checking



# The Holy Grail



Source: Bill Adair



# The Holy Grail



Source: Bill Adair

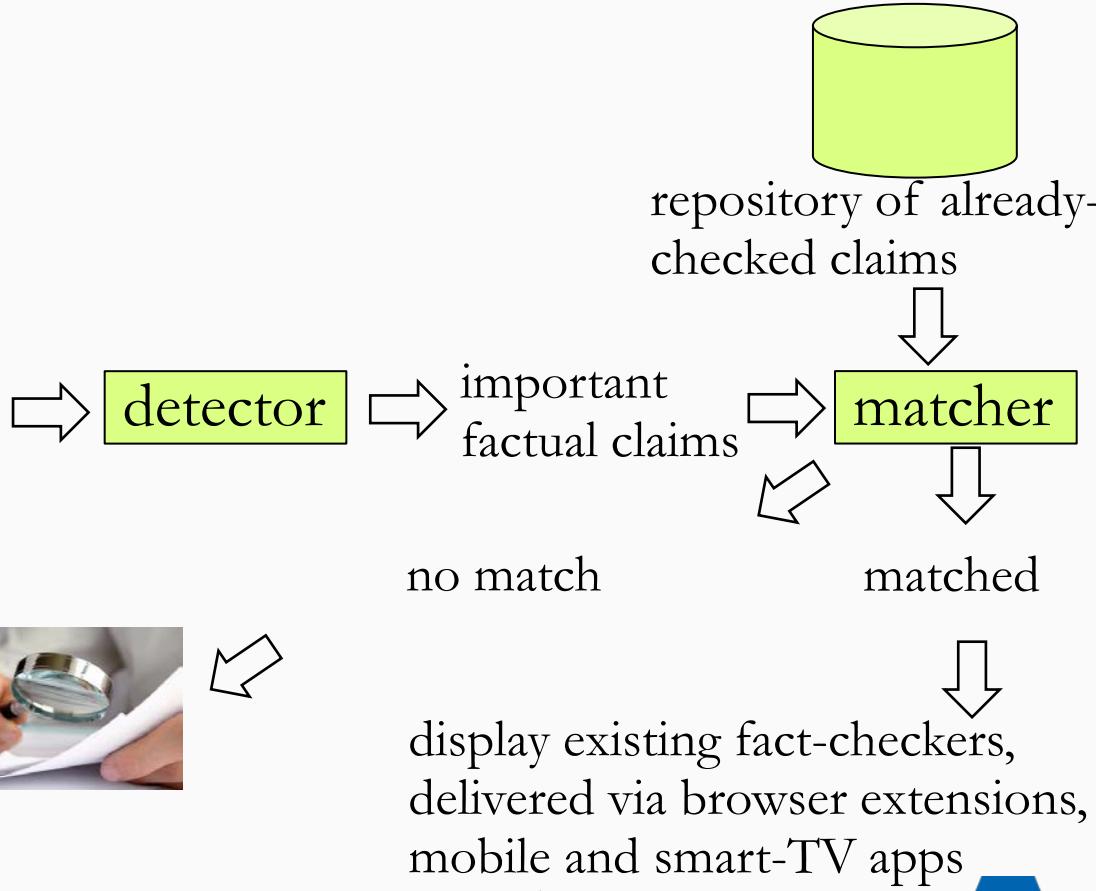
# The Holy Grail



Source: Bill Adair

# ClaimBuster

- political discourses (debates, interviews), advertisements, live events on TV and online video streams
- social media (e.g., twitter)
- web pages
- news articles

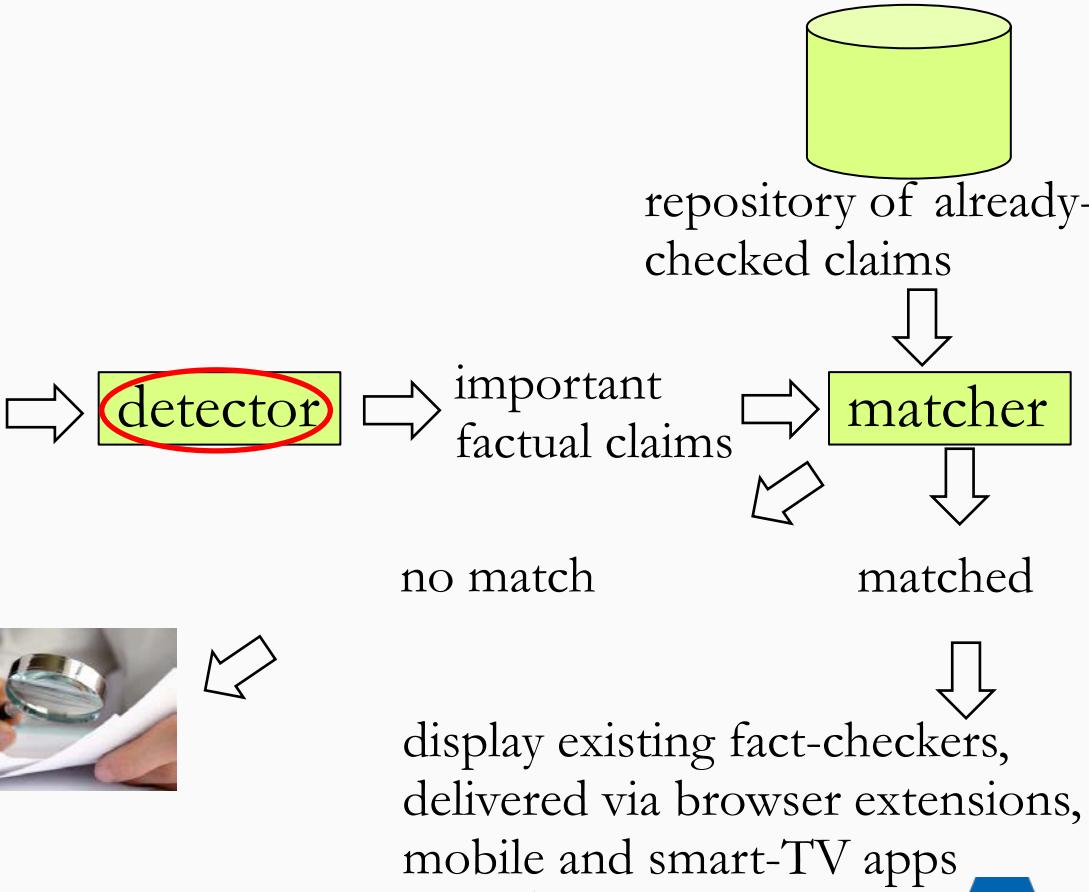


assist data analysis; solicit analyses from professionals

# ClaimBuster

- political discourses (debates, interviews), advertisements, live events on TV and online video streams
- social media (e.g., twitter)
- web pages
- news articles

assist data analysis; solicit analyses from professionals



# iCheck (Led by Duke)

## ✓ Kay Hagan is overly partisan.

Republicans suggest her achievements are thin soup. "The only thing Kay Hagan has accomplished in Washington is becoming an automatic 'yes' vote for whatever new tax or regulation President Obama wants," North Carolina GOP Chairman Claude Pope said.

Original claim made by: 1

### Supporting Arguments

7  

[Frank R. Lautenberg\(D\)](#) and [Kay Hagan\(D\)](#) agreed on 90.02% of the votes they cast between 2011-01-05 (start of session 2011) and 2013-01-01 (end of session 2012)

3  

[Thomas Harkin\(D\)](#) and [Kay Hagan\(D\)](#) agreed on 92.6% of the votes they cast between 2013-01-03 (start of session 2013) and 2014-05-09 (end of session 2014)

### Counter Arguments

#### Generated Counter Arguments



Bernard Sanders(Independent) and [Kay Hagan\(D\)](#) agreed on 85.77% of the votes they cast between 2011-01-05 (start of session 2011) and 2013-01-01 (end of session 2012)



[Kay Hagan\(D\)](#) voted 89.83% of the time with the Democrat party majority vote between 2008-01-02 and 2012-01-02.

# iCheck (Led by Duke)

In 2011, [Miguel Cabrera](#) had 197 in hits, 30 in homeruns, 0.34 in batting average; only 6 other players have ever beaten this record;

 Bump!

 Tweet!

Views: 1 | Bumps: 0 | Tweets: 0

Vladimir Guerrero: 197 in hits, 44 in homeruns, 0.35 in batting average in 2000

Todd Helton: 216 in hits, 42 in homeruns, 0.37 in batting average in 2000; 209 in hits, 33 in homeruns, 0.36 in batting average in 2003

Mike Piazza: 201 in hits, 40 in homeruns, 0.36 in batting average in 1997

Albert Pujols: 212 in hits, 43 in homeruns, 0.36 in batting average in 2003

Alex Rodriguez: 215 in hits, 36 in homeruns, 0.36 in batting average in 1996

Larry Walker: 208 in hits, 49 in homeruns, 0.37 in batting average in 1997

## Responses



The same claim (i.e. "no more than 6 other players have ever beaten this player's record in some year in 'hits', 'homeruns', 'batting average') can be made for 41 other players.

The other player are: [Albert Belle](#) in 1995 (4), in 1998 (1); [Adrian Beltre](#) in 2004 (1); [Dante Bichette](#) in 1995 (6), in 1998 (3); [Barry Bonds](#) in 2001 (0), in 2002 (0), in 2003 (1), in 2004 (1); [Bret Boone](#) in 2001 (6); [Ellis Burks](#) in 1996 (2); [Vinny Castilla](#) in 1998 (1); [Carlos Delgado](#) in 2000 (4); [Jacoby Ellsbury](#) in 2011 (4); [Darin Erstad](#) in 2000 (0); [Nomar Garciaparra](#) in 2000 (1); [Adrian Gonzalez](#) in 2011 (4); [Luis Gonzalez](#) in 2001 (0); [Ken Griffey](#) in 1997 (3), in 1998 (6); [Vladimir Guerrero](#) in 2000 (1), in 2002 (4), in 2004 (4); [Tony Gwynn](#) in 1995 (2), in 1997 (0); [Josh Hamilton](#) in 2010 (3); [Todd Helton](#) in 2000 (0), in 2001 (1), in 2003 (1), in 2004 (4); [Matt Holliday](#) in 2007 (1); [Ryan Howard](#) in 2006 (1); [Derek Jeter](#) in 1999

# ClaimBuster to be 2016-Ready



2016  
Presidential  
Debates  
(Speeches, debates,  
interviews, social  
media, news)



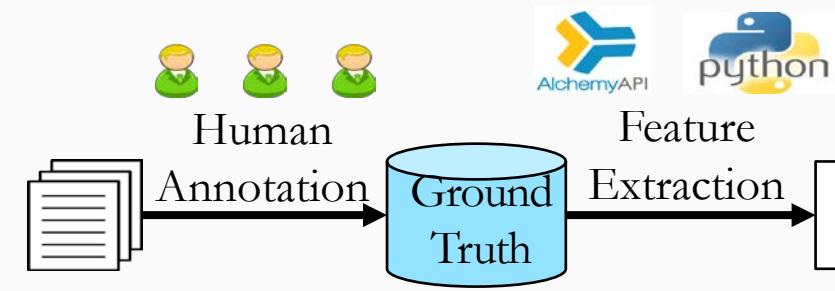
→ Factual claims  
recommended  
to be checked →



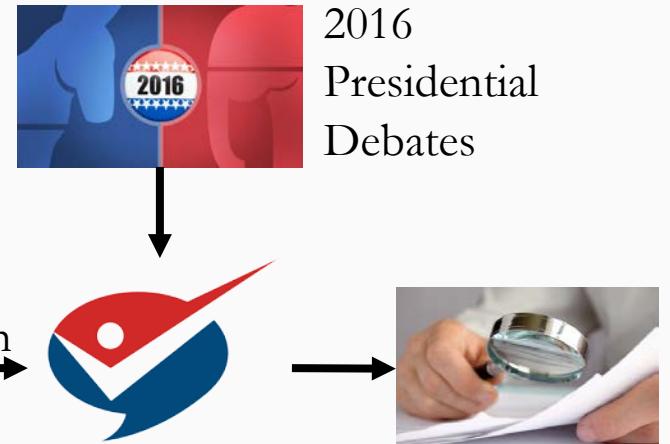
Journalists  
investigate  
the claims  
(or checked by  
algorithms,  
citizens, crowd)



# Finding Important Factual Claims: A Classification Problem



Presidential  
Debate  
Transcripts



Important  
factual claims



# Dataset: Presidential Debate Transcripts

- Source: <http://www.debates.org/index.php?page=debate-transcripts>
- All 30 debates (11 elections) in history: 1960, 1976—2012
- 20k sentences by presidential candidates: removed very short (< 5 words) sentences

# 3 Classes of Sentences

## Important factual claims

“We spend less on the military today than at any time in our history.” “The President’s position on gay marriage has changed.” “More people are unemployed today than four years ago.”

## Unimportant factual claims

“I was in Iowa yesterday.” “My mother enjoys cooking.” “I ran for President once before.”

## Sentences with no factual claims (just opinions, questions & declarations)

“Iran must not get nuclear weapons.” “7% unemployment is too high.” “My opponent is wishy-washy.” “I will be tough on crime.” “Why should we do that?” “Hello, New Hampshire!” “Our plan is to reduce tax rate by 10%.”

Goal: Given a future sentence, find  
the class it belongs to.

# Ground Truth Collection

- Developed a data collection platform [bit.ly/claimbusters](http://bit.ly/claimbusters).
- In 3 months, we accumulated 226 participants.
- Used 600 screening sentences to detect spammers & low-quality participants.
- Admitted sentences which are agreed by at least 2 top-quality participants.
- 8015 such sentences.

Class	Count
CFS	1673
UFS	482
NFS	5860



# Ground Truth Collection Website

OI: Wages are going up for the first time in a decade.

[More Context](#)

*Will the general public be interested in knowing whether (part of) this sentence is true or false?*

- There is **no** factual claim in this sentence.
- There is a factual claim but it is **unimportant**
- There is an **important** factual claim.

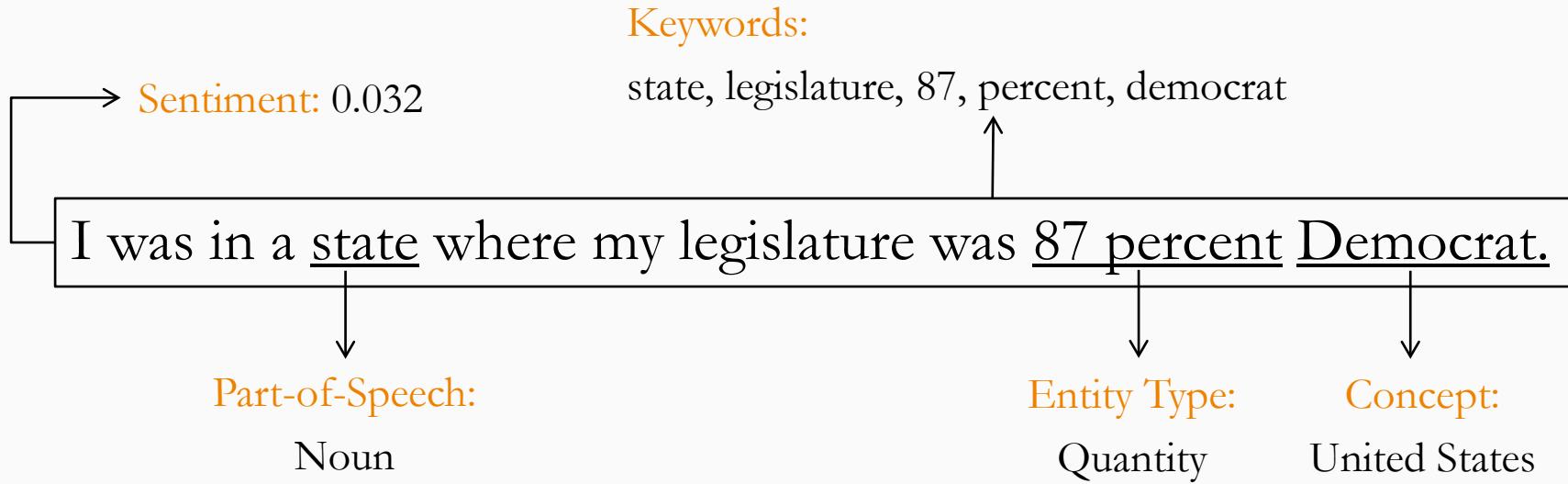
[Submit](#)

[Skip this sentence](#)

[Modify My Previous Responses](#)



# Feature Extraction



Sentiment: [-1.0 to 1.0]

Words: tf-idf scores of 6130 words (excluding rare words)

POS Tag: 43 tags

Entity Type: 26 types

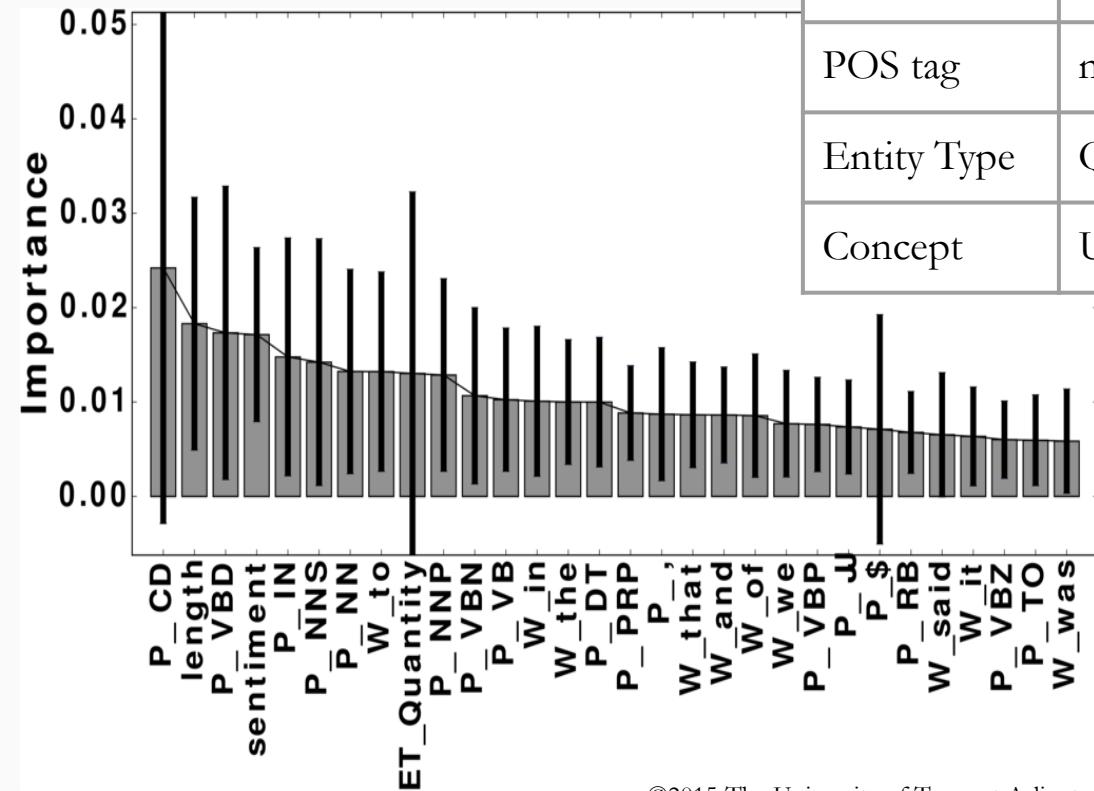


# Feature Selection

- 6201 features in total
- Used a Random Forest Classifier to calculate importance of each feature.
- Most Important Feature: POS tag ‘Cardinal Number’



# Important Features



Word	percent; people; jobs
POS tag	noun; cardinal number; past tense; preposition
Entity Type	Quantity; Country; FieldTerminology; Person
Concept	United States Senate; Barack Obama

# Implementation: Python NLP/ML Tools

## Data wrangling

- Use [NLTK](#) (Natural Language Toolkit) to transform debate files into structured data format
- Use [mysql-python-connector](#) to store extracted features into an MySQL database
- Use [matplotlib](#) to plot classifiers' performance.

## Feature extraction

- Use [AlchemyAPI](#) (Python wrapper) to extract rich features of sentences

## Classification

- Use [scikit-learn](#) to build classification models



# Evaluation: Classification

- 4-fold cross validation
- Algorithms: Naive Bayes, Random Forest & Support Vector Machine
- Support Vector Machine performed better than others in general.

	Precision	Recall	F-measure
NFS	0.90	0.96	0.93
UFS	0.65	0.26	0.37
CFS	0.79	0.74	0.77



# Evaluation: Ranking

- Measured accuracy of top-K sentences.
- ClaimBuster has a strong agreement with high-quality human coders on the check-worthiness of sentences

K	P@K	NDCG@K
25	1	1
50	1	1
100	0.960	0.970
200	0.940	0.951
300	0.853	0.881
500	0.690	0.840



# Case Study: #GOPDebate2015

- Near real-time experiment with 2015 first Republican primary debate
- Transcript grabbed from closed captions of the Fox News channel using TextGrabber
- 1393 sentences
- 71% of the fact-checks from CNN, factcheck.org & PolitiFact were ranked by ClaimBuster within top 18%.

# Case Study: #GOPDebate2015

CNN Claim	Associated sentence(s)[From TextGrabber]	Score
1	Part of this iranian deal was lifting the international sanctions on general sulemani.	0.415
2	I would go on to add – >> you don't favor – >> i have never said that.	0.511
3	A majority of the candidates on this stage supported amnesty.	0.295
4	Timely the medicaid is growing at one of the lowest rates in the country.	0.534
4	We went from \$8 billion in the hole to \$5 million in the black.	0.773
5	And the mexican government is much smarter, much sharper, much more cunning and they send the bad ones over because they don't want to pay for them.	0.215
6	[Not found in the transcript]	N/A



# Case Study: #GOPDebate2015

- Real-time experiment with 2015 second Republican primary debate
- Closed Captions from CNN channel
- Tweeted important factual claims to  
<https://twitter.com/ClaimBusterTM>, live!



©2015 The University of Texas at Arlington. All Rights Reserved.



## Tweets

 Follow



**ModerateEdge**

3m

@ModerateEdge

@realDonaldTrump If you make \$25,001,  
should you pay \$2,500 when \$25,000 you  
pay nothing? Pay only on amt over \$25K.  
#Trump2016 #DonaldTrump

 Retweeted by ClaimBuster

Expand



**BicycleBrandsDirect**

9m

@bicyclebrands

1.3 million bicycles recalled for crash  
hazard: (KMSP) - Nearly 1.3 million  
bicycles in the United States are ...  
[bit.ly/1MGT2HO](http://bit.ly/1MGT2HO)

 Retweeted by ClaimBuster

Expand



**Rich Luchette**

19m

@richluchette

\$272 million project, more than 3,000 jobs

 Tweet to @ClaimBusterTM

# Demo

<http://idir.uta.edu/claimbuster>

 ClaimBuster

Press Acknowledgement

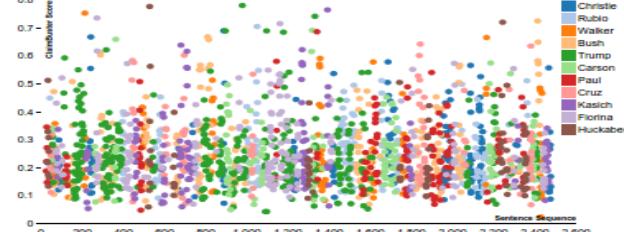
2016 Republican Party Presidential Debate. Sept. 16, 2015, 7 p.m.  
Venue: Ronald Reagan Presidential Library, Simi Valley, California. Broadcasted by: CNN.  
Speakers: Dana Bash, Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Hugh Hewitt, Mike Huckabee, John Kasich, Rand Paul, Marco Rubio, Jake Tapper, Donald Trump, Scott Walker  
Transcript Source: <http://time.com/4037239/second-republican-debate-transcript-cnn/>

Chronological Order Order by Score Most Check-worthy <=0.1 <=0.2 <=0.3 <=0.4 <=0.5 <=0.6 <=0.7 <=0.8 <=0.9 <=0.95 Least Check-worthy

0.04 The eleven leading Republican candidates for president are at their podiums.  
0.17 They are ready to face off, and if you've been watching this race, you know anything could happen over the next few hours.  
0.03 To viewers who are just joining us, welcome to the Air Force One Pavilion of the Ronald Reagan Presidential Library.  
0.29 Our thanks to the staff here and especially to former first lady Nancy Reagan for this impressive setting with Ronald Reagan's presidential plane as our backdrop.  
0.29 This debate is airing on CNN networks in the United States and around the world.  
0.30 It's also being broadcast on the Salem Radio Network.  
0.17 I know everyone is very eager to get started.  
0.23 But first, I want to explain the ground rules tonight.  
0.19 My name is Jake Tapper.  
0.19 I'll be the moderator.  
0.24 I will be joined in the questioning by Salem Radio Network talk show host Hugh Hewitt.  
0.04 He worked in the Reagan administration for six years.  
0.22 And by CNN's chief political correspondent Dana Bash.  
0.15 I will ask follow-up questions, I will attempt to guide the discussion.

  
Second 2016 GOP President...  
QUESTION  
PRESIDENTS SPEECHES ON CNN  
CNN  
SALINAS PRESIDENTIAL DEBATE  
37:23 / 2:54:35

0.24 Do they bear responsibility for this refugee crisis, and what would you have done when Bashar Assad crossed the line?





## Automated live fact-checking



### 2016 Republican Party Presidential Debate. Sept. 16, 2015, 7 p.m.

Speakers: Dana Bash, Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Hugh Hewitt, Mike Huckabee, John Kasich, Rand Paul, Marco Rubio, Jake Tapper, Donald Trump, Scott Walker



### 2016 Republican Party Presidential Debate. Aug. 6, 2015, 8 p.m.

Speakers: Bret Baier, Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Mike Huckabee, John Kasich, Megyn Kelly, Rand Paul, Rick Perry, Marco Rubio, Donald Trump, Scott Walker, Chris Wallace

Fact-check your own text

#### Tweets

[Follow](#)



**ModerateEdge**  
@ModerateEdge

@realDonaldTrump If you make \$25,001, should you pay \$2,500 when \$25,000 you pay nothing? Pay only on amt over \$25K.  
#Trump2016 #DonaldTrump

[Retweeted by ClaimBuster](#)  
[Expand](#)



**BicycleBrandsDirect**  
@bicyclebrands

1.3 million bicycles recalled for crash hazard: (KMSP) - Nearly 1.3 million bicycles in the United States are ...  
[bit.ly/1MGT2HO](http://bit.ly/1MGT2HO)

[Retweeted by ClaimBuster](#)  
[Expand](#)



**Rich Luchette**  
@richluchette

\$272 million project, more than 3,000 jobs

[Tweet to @ClaimBusterTM](#)



2016 Republican Party Presidential Debate. Sept. 16, 2015, 7 p.m.

Venue: Ronald Reagan Presidential Library, Simi Valley, California. Broadcasted by: CNN.

Speakers: Dana Bash, Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Hugh Hewitt, Mike Huckabee, John Kasich, Rand Paul, Marco Rubio, Jake Tapper, Donald Trump, Scott Walker

Transcript Source: <http://time.com/4037239/second-republican-debate-transcript-cnn/>

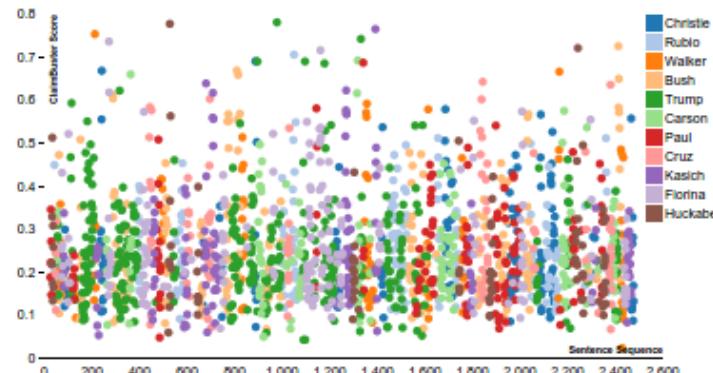
Chronological Order Order by Score

Most Check-worthy <=0.8 <=0.6 <=0.4 <=0.2 Least Check-worthy

0.48 The eleven leading Republican candidates for president are at their podiums.  
0.17 They are ready to face off, and if you've been watching this race, you know anything could happen over the next few hours.  
0.41 To viewers who are just joining us, welcome to the Air Force One Pavilion of the Ronald Reagan Presidential Library.  
0.29 Our thanks to the staff here and especially to former first lady Nancy Reagan for this impressive setting with Ronald Reagan's presidential plane as our backdrop.  
0.28 This debate is airing on CNN networks in the United States and around the world.  
0.30 It's also being broadcast on the Salem Radio Network.  
0.17 I know everyone is very eager to get started.  
0.23 But first, I want to explain the ground rules tonight.  
0.19 My name is Jake Tapper.  
0.19 I'll be the moderator.  
0.24 I will be joined in the questioning by Salem Radio Network talk show host Hugh Hewitt.  
0.44 He worked in the Reagan administration for six years.  
0.22 And by CNN's chief political correspondent Dana Bash.  
0.18 I will ask follow-up questions, I will attempt to guide the discussion.



0.24 Do they bear responsibility for this refugee crisis, and what would you have done when Bashar Assad crossed the line?





## 2016 Republican Party Presidential Debate. Sept. 16, 2015, 7 p.m.

Venue: Ronald Reagan Presidential Library, Simi Valley, California. Broadcasted by: CNN.

Speakers: Dana Bash, Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Hugh Hewitt, Mike Huckabee, John Kasich, Rand Paul, Marco Rubio, Jake Tapper, Donald Trump, Scott Walker

Transcript Source: <http://time.com/4037239/second-republican-debate-transcript-cnn/>

Most Check-worthy <=1.0 <=0.9 <=0.8 <=0.7 <=0.6 <=0.5 <=0.4 <=0.3 <=0.2 <=0.1 Least Check-worthy

Chronological Order

Order by Score

0.24 That's a fact.

0.33 And when the people of Iowa found that out, I went to No.

0.46 1 and you went down the tubes.

0.29 Governor Walker?

0.13 Jake, yeah, absolutely, I'll take this on, because this is an issue that's important in this race.

0.31 Just because he says it doesn't make it true.

0.19 The facts are the facts.

0.75 We balanced a \$3.6 billion budget deficit, we did it by cutting taxes - \$4.7 billion to help working families, family farmers, small business owners and senior citizens.

0.23 And it's about time people in America stand up and take note of this.

0.30 If you want someone that can actually take on the special interest of Washington, which you yourself said you were part of, using the system, we need somebody that will stand up and fight for average Americans to put them back in charge of their government.

0.16 I'm the one who is taking that on.

0.23 I'll do that as your next president.

0.17 Let's move on.

0.23 Jake, Jake.

0.13 A phenomenon going on in the race right now is the political...

0.25 OK, Governor Kasich, go ahead.

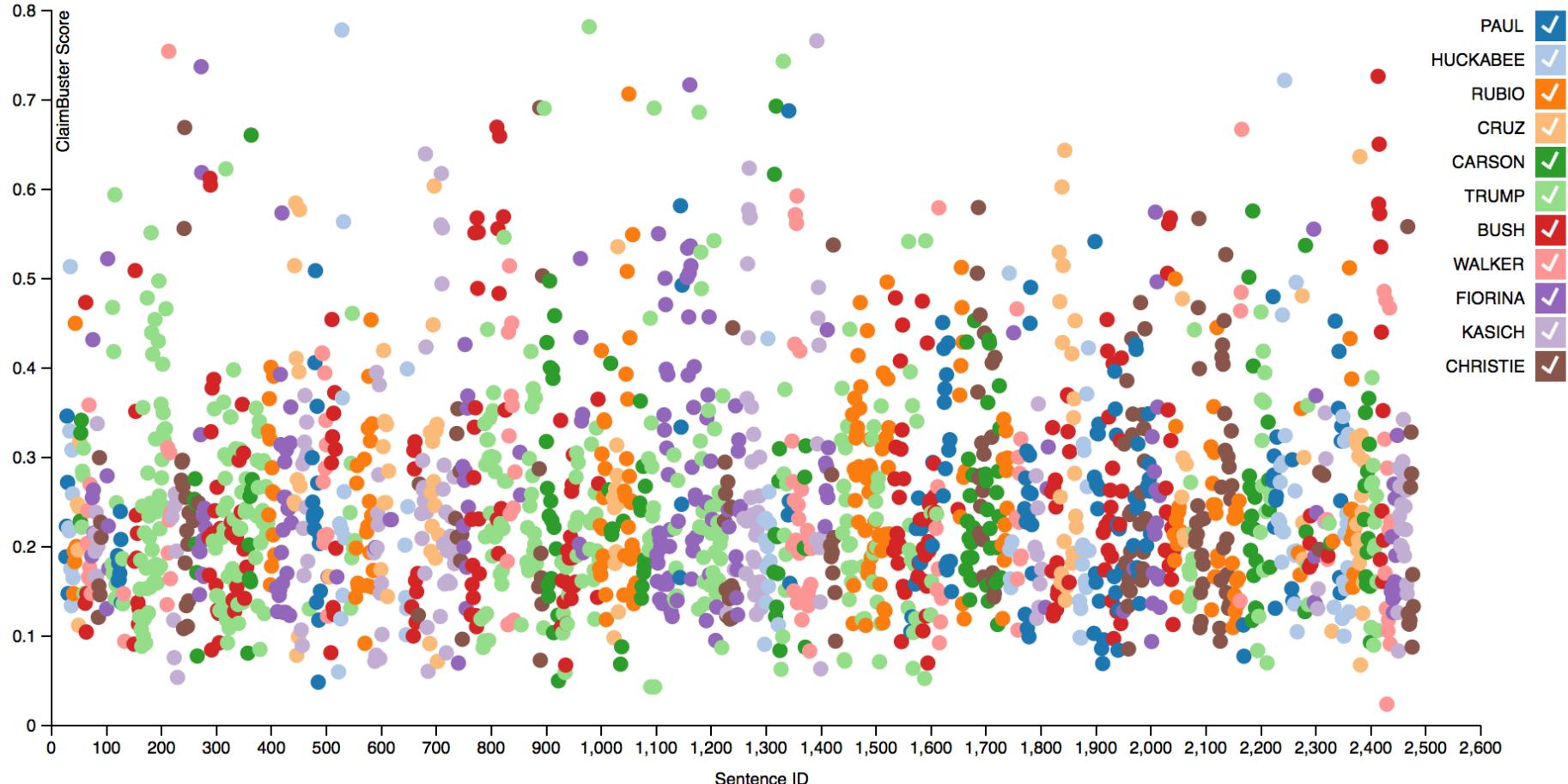


0.53

In fact, today, on the front page of the Wall Street Journal, they fired another 25 or 30,000 people saying we still haven't recovered from the catastrophe.



## Debate Timeline Graph





# You are Invited

<http://bit.ly/claimbusters>

Classify Check Worthy Fact x  
idir-server2.uta.edu/classifyfact\_survey/

ClassifyFact chengkai labeled 126 sentences Leaderboard Log Out

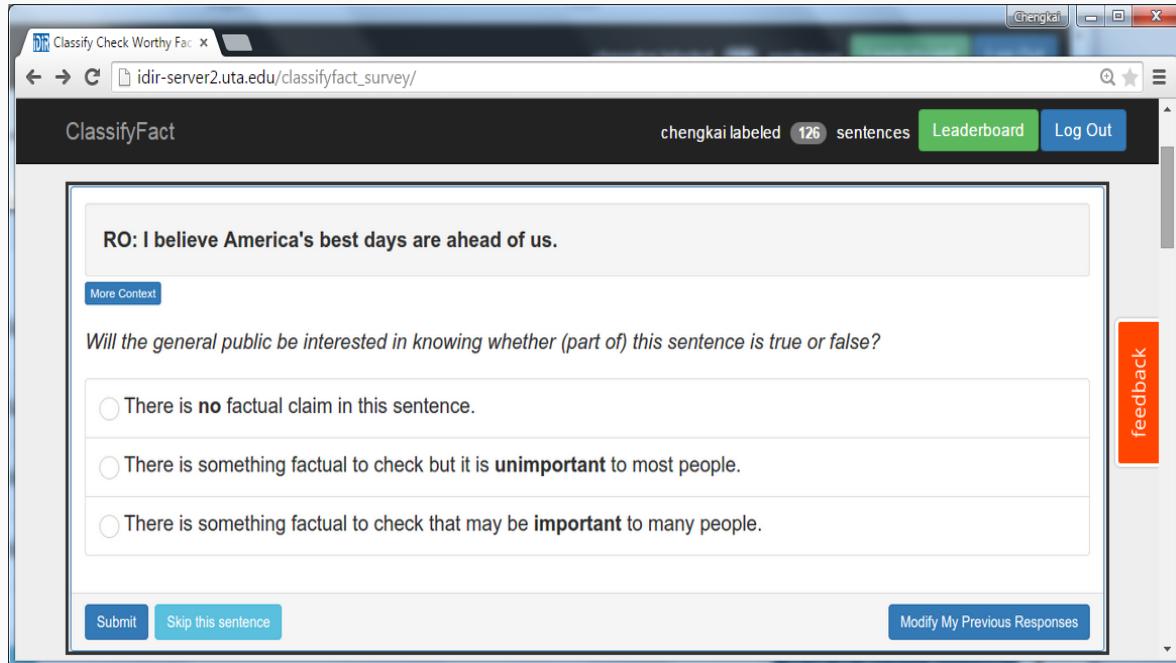
RO: I believe America's best days are ahead of us.

More Context

Will the general public be interested in knowing whether (part of) this sentence is true or false?

There is no factual claim in this sentence.  
 There is something factual to check but it is unimportant to most people.  
 There is something factual to check that may be important to many people.

Submit Skip this sentence Modify My Previous Responses



# FactWatcher

Automated Monitoring of Facts from Real-  
World Events



# FactWatcher (Fact Finding)



Tuple  $t$  for new real world event appended to database



id	player	day	month	season	team	opp_team	pts	ast	reb
$t_1$	Bogues	11	Feb.	1991-92	Hornets	Hawks	4	12	5
$t_2$	Seikaly	13	Feb.	1991-92	Heat	Hawks	24	5	15
$t_3$	Sherman	7	Dec.	1993-94	Celtics	Nets	13	13	5
$t_4$	Wesley	4	Feb.	1994-95	Celtics	Nets	2	5	2
$t_5$	Wesley	5	Feb.	1994-95	Celtics	Timberwolves	3	5	3
$t_6$	Strictland	3	Jan.	1995-96	Blazers	Celtics	27	18	8
$t_7$	Wesley	25	Feb.	1995-96	Celtics	Nets	12	13	5

Find constraint-measure pair  $(C, M)$  such that  $t$  is in the contextual skyline



Generate factual claim



Wesley had 12 points, 13 assists and 5 rebounds on February 25, 1996 to become the first player with a 12/13/5 (points/assists/rebounds) in February.

Constraint	Measure
month=Feb	pts, ast, reb
opp_team=Nets	ast, reb
team=Celtics & opp_team=Nets	ast, reb
...	...

## »LIVE UPDATE

[February 20, 1998] Todd Fuller had 1 assist, 3 steals and 1 block in the Golden State Warriors' defeat against the Denver Nuggets. It is one of the best performance made by him.

SEARCH

michael jordan

 4

Michael Adonis Jordan

**Michael Jordan**

Michael Michael Jordan

Michael Reggie Jordan

 3

Michael Thomas Jordan

Victory against the New Jersey Nets is one of the best performance made by him.

 1

[January 13, 1997] Horace Grant had 26 points and 6 assists in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.

**MORE LIKE THIS** 2

[January 13, 1997] After the Orlando Magic's win over the New Jersey Nets, for the first time in his career, Rony Seikaly had at least 20 points for 6 consecutive games, after today's game.

**MORE LIKE THIS** 1

[January 13, 1997] Horace Grant had 26 points and 2 steals in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.

**MORE LIKE THIS** 5

[January 13, 1997] Horace Grant had 26 points, 6 assists and 2 steals in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.

**MORE LIKE THIS** 3

[January 13, 1997] After the Orlando Magic's victory against the New Jersey Nets, for the first time in his career, Rony Seikaly had at least 20 points and 8 rebounds for 6 consecutive games, after today's game.

**MORE LIKE THIS** 2

[January 13, 1997] Nick Anderson had 8 assists and 2 blocks in the Orlando Magic's win over the New Jersey Nets. It is one of the best performance made by him.

**MORE LIKE THIS**

FACT TYPE &gt;

SITUATIONAL FACT PROMINENT STREAK ONE-OF-THE-FEW 

RANKING &gt;

RECENTNESS INTERESTINGNESS POPULARITY 

PLAYERS &gt;

TEAMS &gt;

SEASONS &gt;

1996-97 (9) 1994-95 (5) 1992-93 (1) 

+MORE

LESS 

Presented In



Excellent Demo Award

**COMPUTATION  
+ JOURNALISM**   
**SYMPOSIUM** 

<http://idir.uta.edu/factwatcher/>

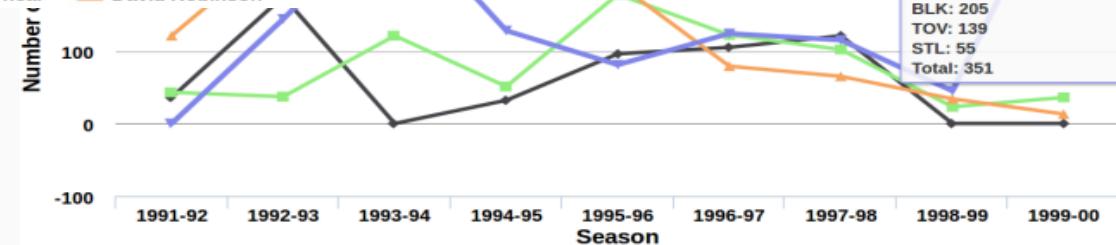
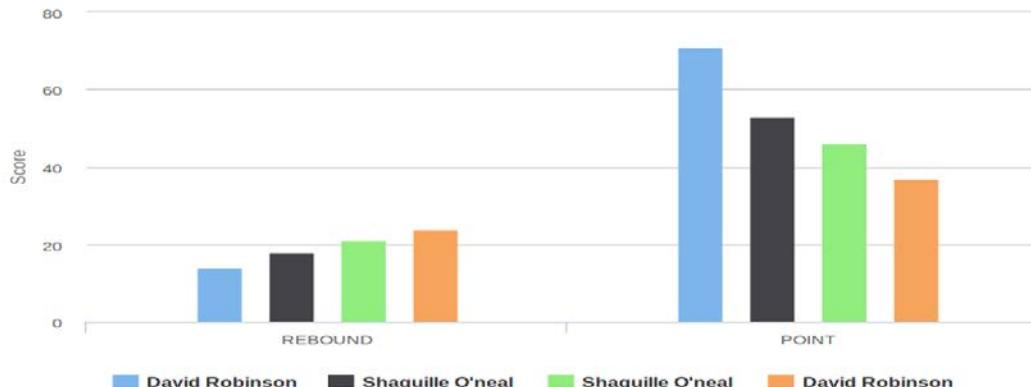
[April 24, 1994] David Robinson had 71 points and 14 rebounds in the San Antonio Spurs' victory against the Los Angeles Clippers. No one before had a better performance in NBA history.

[April 20, 1994] Shaquille O'neal had 53 points and 18 rebounds in the Orlando Magic's win over the Minnesota Timberwolves. No one before had a better performance in NBA history.

[February 16, 1993] Shaquille O'neal had 46 points and 21 rebounds in the Orlando Magic's defeat against the Detroit Pistons. No one before had a better performance in NBA history.

[February 27, 1992] David Robinson had 37 points and 24 rebounds in the San Antonio Spurs' victory against the Golden State Warriors. No one before had a better performance in NBA history.

#### Compare Similar Stories



# FactWatcher Finds Three Types of Facts (and can be Extended)

## Domains

- o sports, weather, crimes, transportation, finance, social media analytics

## Examples from Real News Media

### Prominent streaks

- o “This month the Chinese capital has experienced 10 days with a maximum temperature in around 35 degrees Celsius – the most for the month of July in a decade.”  
[http://www.chinadaily.com.cn/china/2010-07/27/content\\_11055675.htm](http://www.chinadaily.com.cn/china/2010-07/27/content_11055675.htm)
- o “The Nikkei 225 closed below 10000 for the 12th consecutive week, the longest such streak since June 2009.”  
<http://www.bloomberg.com/news/articles/2010-08-06/japanese-stocks-fall-for-second-day-this-week-on-u-s-jobless-claims-yen>



# FactWatcher Finds Three Types of Facts (and can be Extended)

## Examples from Real News Media

### Situational facts, One-of-the-few objects

- “Paul George had **21 points**, 11 rebounds and 5 assists to become the **first Pacers player** with a 20/10/5 (points/rebounds/assists) game **against the Bulls since** Detlef Schrempf in **December 1992.**”  
<http://espn.go.com/espn/elias?date=20130205>
- “The social world’s **most viral photo** ever generated **3.5 million likes**, 170,000 comments and **460,000 shares** by Wednesday afternoon.”  
<http://www.cnbc.com/id/49728455>



# Demo

<http://idir.uta.edu/factwatcher/>



# Acknowledgment

## UTA Students

- Naeemul Hassan
- Afroza Sultana
- Gensheng Zhang
- Joseph Minumol
- Fatma Dogan

## Collaborators

- Bill Adair (Duke)
- Pankaj Agarwal (Duke)
- Sarah Cohen (Columbia)
- James Hamilton (Stanford)
- Ping Luo (Chinese Academy of Sciences)
- Mark Tremayne (UTA)
- Min Wang (Google Research)
- Jun Yang (Duke)
- Cong Yu (Google Research)



# Acknowledgment

## Funding sponsors



UNIVERSITY OF  
TEXAS  
ARLINGTON

**Disclaimer:** This material is based upon work partially supported by the National Science Foundation Grants 1018865, 1117369 and 1408928, 2011 and 2012 HP Labs Innovation Research Awards, and the National Natural Science Foundation of China Grant 61370019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.



# Thank You! Questions?

<http://ranger.uta.edu/~cli>

<http://idir.uta.edu>

[cli@uta.edu](mailto:cli@uta.edu)

Please help us label the data

<http://bit.ly/claimbusters>



# Other IDIR Projects



# CrowdConsensus

Team/organization decision making and  
event organizer



# Humans' Obsession: Comparing Things

Who is the better footballer?



<http://visual.ly/messi-vs-ronaldo>

Which is the better company to work for?



# Facebook-Google Showdown

BI Facebook vs Google best < X www.businessinsider.com/facebook-vs-google-best-employer-2013-11

BUSINESS INSIDER Tech Finance Politics Strategy Life

## These Charts Prove Facebook Is A Better Place To Work Than Google

JULIE BORT | 8+ NOV. 26, 2013, 9:22 PM | 18,234 | 6

[FACEBOOK](#) [LINKEDIN](#) [TWITTER](#) [GOOGLE+](#) [PRINT](#) [EMAIL](#)

### Car Rentals From \$8 a Day

[tripbase.com/Cars-Cheap](#)  
Cheap Car Rentals Compare Deals from Top Companies

When it comes to desirable places to work in the tech industry, two companies are always at the top of the list: Facebook and Google.

But which one is *really* the better employer?

To answer that question, we compared the two



Mark Zuckerberg and Larry Page.

BI Facebook vs Google best < X www.businessinsider.com/facebook-vs-google-best-employer-2013-11

BUSINESS INSIDER Tech Finance Politics Strategy Life

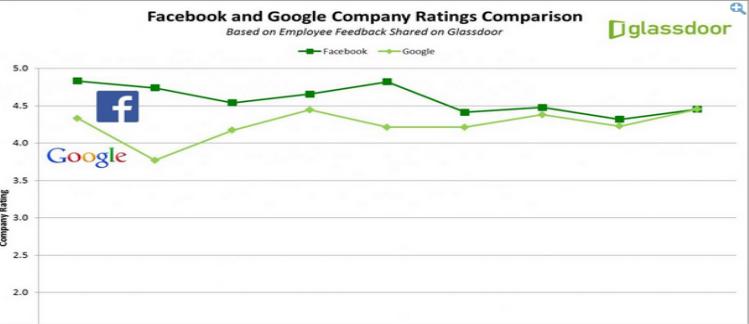
### Round 1: Overall satisfaction ratings — Facebook wins

(Score so far: Facebook - 1, Google - 0)

Facebook employees rate their employer slightly better overall (4.6) compared to how Google employees rate Google (4.1), from at least 550 company reviews per company.

Satisfaction ratings are based on a 5-point scale: 1.0=very dissatisfied, 3.0=OK, 5.0=very satisfied. So, we can see that employees are highly satisfied with both companies.

Interestingly, Google's rating has been climbing, indicating that current employers are getting happier. As of this quarter, the two companies are tied at 4.5 each, as the chart below shows:



The graph compares Facebook (green line with square markers) and Google (blue line with diamond markers) satisfaction scores over time. Both companies show relatively stable satisfaction levels around 4.5-4.6, with Google's score increasing slightly over time.

Period	Facebook Satisfaction Score	Google Satisfaction Score
Q1	4.6	4.1
Q2	4.6	4.1
Q3	4.5	4.2
Q4	4.5	4.3
Q1	4.5	4.4
Q2	4.5	4.5
Q3	4.5	4.5
Q4	4.5	4.5

# Facebook-Google Showdown

Multiple criteria

Overall satisfaction	CEO approval	Employee confidence in the future	Perks and salaries	Interview difficulty
 $\succ$ 	 $\succ$ 	 $\succ$ 	 $\succ$ 	 $\succ$ 

Which one is better?

# CrowdConsensus

Optimization techniques and algorithms that minimize questions to users for reaching consensus.

## Applications

Collecting Public Opinion

Group Decision Making

- Where for lunch, which product to use, which candidate to hire

Information Exploration

- Compare photos by color, sharpness, and landscape

# How to Get Preference Relations?

## Ask the Crowd (mobile apps)

**Between *A separation*(2011) and *The big Lebowski*(1998)  
which movie is better with regard to *story*?**

- A separation(2011).
- The big Lebowski(1998).
- no preference.

**submit**

**skip this question**

**Between each pair of cities, which has the better weather overall?**

<input type="radio"/> Dallas(TX)	<input type="radio"/> about the same	<input type="radio"/> Chicago(IL)	<input type="radio"/> not sure
----------------------------------	--------------------------------------	-----------------------------------	--------------------------------

**In each following pair(each row), which picture has more interesting colors?**

Left Picture	Right Picture	Preference
		<ul style="list-style-type: none"><li><input type="radio"/> Left Picture</li><li><input type="radio"/> Right Picture</li><li><input type="radio"/> No Preference</li></ul>

# CrewScout

## Expert Team Formation

# CrewScout --- Expert Team Formation

Task Panel	Skill Panel	Parameter Panel	Display Panel																																																																																																																														
<p><b>CrewScout</b> An Expert Team Finding Tool</p> <p>database <input type="text"/> Search</p> <p>Total 9793 Available Tasks</p> <p><b>Task Detail</b></p> <p>Task ID: 177 Title: The Design and Implementation of a Sequence Database System Description: This paper discusses the design and implementation of SEQ, a database system with support for sequence data. SEQ models a sequence as an ordered collection of records, and supports a declarative sequence query language based on an algebra of query operators, thereby permitting algebraic query optimization and evaluation. SEQ has been built as a component of the PREDATOR database system Reference: 32</p> <p>Task ID: 291. Browsing and querying in object-oriented databases</p> <p>Task ID: 618. Polynomial time query processing in temporal deductive databases</p> <p>Task ID: 748. Supporting Fine-grained Data Lineage in a Database Visualization Environment</p>	<p><b>Required Skills</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> Design and Implementation</li><li><input checked="" type="checkbox"/> Database System</li><li><input checked="" type="checkbox"/> Query Optimization</li><li><input checked="" type="checkbox"/> Query Language</li><li><input checked="" type="checkbox"/> Query Evaluation</li><li><input checked="" type="checkbox"/> Complex Data</li></ul> <p><input type="button" value="Show Experts"/></p> <p><b>Aggregate Function</b> <input checked="" type="radio"/> AVG <input type="radio"/> MIN <input type="radio"/> MAX</p> <p><b>Skyline Team Size</b> <input type="text" value="3"/> <input type="button" value="Skyline Teams"/></p>	<p><b>Eligible Experts</b></p> <p>Show 20 entries</p> <table border="1"><thead><tr><th>Expert's Name</th><th>Complex Data</th><th>Database System</th><th>Query Evaluation</th><th>Query Language</th><th>Query Optimization</th><th>Total Weights</th></tr></thead><tbody><tr><td>Raghuram Krishnan</td><td>11</td><td>15</td><td>62</td><td>39</td><td>43</td><td>170</td></tr><tr><td>Leonid Libkin</td><td>11</td><td>3</td><td>54</td><td>100</td><td>0</td><td>168</td></tr><tr><td>Dirk Gucht</td><td>11</td><td>3</td><td>23</td><td>78</td><td>10</td><td>125</td></tr><tr><td>Yehoshua Sagiv</td><td>11</td><td>0</td><td>85</td><td>11</td><td>10</td><td>117</td></tr><tr><td>Caetano Junior</td><td>100</td><td>0</td><td>0</td><td>6</td><td>10</td><td>116</td></tr><tr><td>Agma Traina</td><td>100</td><td>0</td><td>0</td><td>6</td><td>10</td><td>116</td></tr><tr><td>Divesh Srivastava</td><td>11</td><td>10</td><td>46</td><td>28</td><td>19</td><td>114</td></tr><tr><td>Martin Grohe</td><td>11</td><td>0</td><td>46</td><td>44</td><td>5</td><td>106</td></tr><tr><td>Jan Bussche</td><td>11</td><td>3</td><td>8</td><td>72</td><td>0</td><td>94</td></tr><tr><td>Kian-lee Tan</td><td>11</td><td>8</td><td>31</td><td>0</td><td>43</td><td>93</td></tr><tr><td>Diego Calvanese</td><td>11</td><td>3</td><td>8</td><td>33</td><td>33</td><td>88</td></tr><tr><td>Jeffrey Yu</td><td>11</td><td>5</td><td>31</td><td>17</td><td>24</td><td>88</td></tr><tr><td>Louisa Raschid</td><td>22</td><td>0</td><td>15</td><td>22</td><td>29</td><td>88</td></tr><tr><td>Karl Aberer</td><td>22</td><td>5</td><td>31</td><td>11</td><td>19</td><td>88</td></tr><tr><td>Elisa Bertino</td><td>22</td><td>23</td><td>8</td><td>28</td><td>5</td><td>86</td></tr><tr><td>Tore Risch</td><td>22</td><td>5</td><td>0</td><td>33</td><td>24</td><td>84</td></tr><tr><td>Miron Livny</td><td>22</td><td>23</td><td>8</td><td>11</td><td>19</td><td>83</td></tr></tbody></table>	Expert's Name	Complex Data	Database System	Query Evaluation	Query Language	Query Optimization	Total Weights	Raghuram Krishnan	11	15	62	39	43	170	Leonid Libkin	11	3	54	100	0	168	Dirk Gucht	11	3	23	78	10	125	Yehoshua Sagiv	11	0	85	11	10	117	Caetano Junior	100	0	0	6	10	116	Agma Traina	100	0	0	6	10	116	Divesh Srivastava	11	10	46	28	19	114	Martin Grohe	11	0	46	44	5	106	Jan Bussche	11	3	8	72	0	94	Kian-lee Tan	11	8	31	0	43	93	Diego Calvanese	11	3	8	33	33	88	Jeffrey Yu	11	5	31	17	24	88	Louisa Raschid	22	0	15	22	29	88	Karl Aberer	22	5	31	11	19	88	Elisa Bertino	22	23	8	28	5	86	Tore Risch	22	5	0	33	24	84	Miron Livny	22	23	8	11	19	83	 
Expert's Name	Complex Data	Database System	Query Evaluation	Query Language	Query Optimization	Total Weights																																																																																																																											
Raghuram Krishnan	11	15	62	39	43	170																																																																																																																											
Leonid Libkin	11	3	54	100	0	168																																																																																																																											
Dirk Gucht	11	3	23	78	10	125																																																																																																																											
Yehoshua Sagiv	11	0	85	11	10	117																																																																																																																											
Caetano Junior	100	0	0	6	10	116																																																																																																																											
Agma Traina	100	0	0	6	10	116																																																																																																																											
Divesh Srivastava	11	10	46	28	19	114																																																																																																																											
Martin Grohe	11	0	46	44	5	106																																																																																																																											
Jan Bussche	11	3	8	72	0	94																																																																																																																											
Kian-lee Tan	11	8	31	0	43	93																																																																																																																											
Diego Calvanese	11	3	8	33	33	88																																																																																																																											
Jeffrey Yu	11	5	31	17	24	88																																																																																																																											
Louisa Raschid	22	0	15	22	29	88																																																																																																																											
Karl Aberer	22	5	31	11	19	88																																																																																																																											
Elisa Bertino	22	23	8	28	5	86																																																																																																																											
Tore Risch	22	5	0	33	24	84																																																																																																																											
Miron Livny	22	23	8	11	19	83																																																																																																																											

<http://idir.uta.edu/crewsout/>

# CrewScout <http://idir.uta.edu/crewscore/>

## Clustering Panel

Required Skills	
<input type="checkbox"/> Design and Implementation	
<input checked="" type="checkbox"/> Database System	
<input checked="" type="checkbox"/> Query Optimization	
<input checked="" type="checkbox"/> Query Language	
<input checked="" type="checkbox"/> Query Evaluation	
<input checked="" type="checkbox"/> Complex Data	

Aggregate Function	
<input checked="" type="radio"/> AVG	<input type="radio"/> MIN
<input type="radio"/> MAX	

**Skyline Team Size**

Number of Clusters:

Clustering Algorithm:

Similarity Measure:

### Skyline Teams

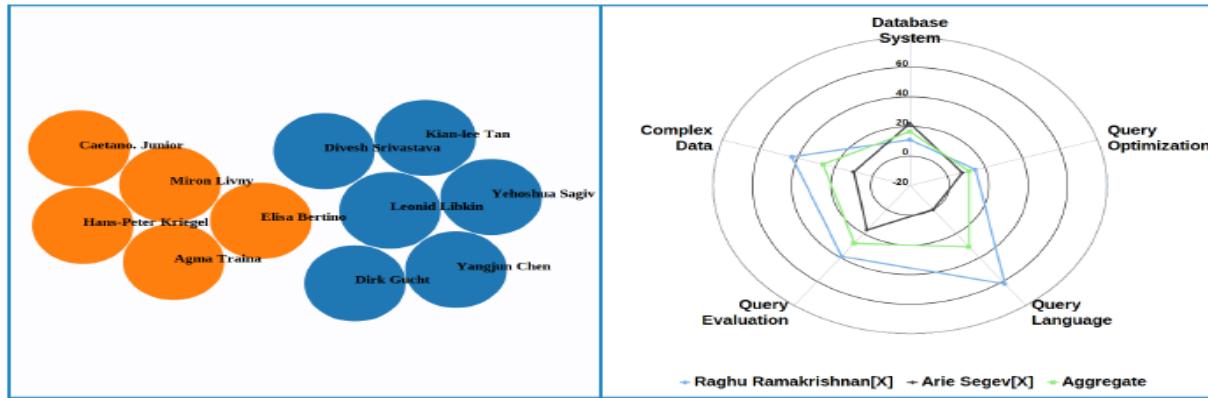
Show 20 entries

Team Members' Names	Complex Data	Database System	Query Evaluat
Dirk Gucht   Raghu Ramakrishnan   Leonid Libkin	11.00	7.00	46.33
Yehoshua Sagiv   Raghu Ramakrishnan   Leonid Libkin	11.00	6.00	67.00
Agma Traina   Raghu Ramakrishnan   Leonid Libkin	40.67	6.00	38.67
Caetano. Junior   Raghu Ramakrishnan   Leonid Libkin	40.67	6.00	38.67
Divesh Srivastava   Raghu Ramakrishnan   Leonid Libkin	11.00	9.33	54.00
Martin Grohe   Raghu Ramakrishnan   Leonid Libkin	11.00	6.00	54.00
Kian-lee Tan   Raghu Ramakrishnan   Leonid Libkin	11.00	8.67	49.00
Diego Calvanese   Raghu Ramakrishnan   Leonid Libkin	11.00	7.00	41.33
Jeffrey Yu   Raghu Ramakrishnan   Leonid Libkin	11.00	7.67	49.00
Raghu Ramakrishnan   Karl Aberer   Leonid Libkin	14.67	7.67	49.00
Raghu Ramakrishnan   Louisa Raschid   Leonid Libkin	14.67	6.00	43.67
Raghu Ramakrishnan   Elisa Bertino   Leonid Libkin	14.67	13.67	41.33
Tore Risch   Raghu Ramakrishnan   Leonid Libkin	14.67	7.67	38.67
Miron Livny   Raghu Ramakrishnan   Leonid Libkin	14.67	13.67	41.33
Yangjun Chen   Raghu Ramakrishnan   Leonid Libkin	18.33	7.67	51.33
Maurizio Lenzerini   Raghu Ramakrishnan   Leonid Libkin	11.00	7.67	41.33
Hans-Peter Kriegel   Raghu Ramakrishnan   Leonid Libkin	14.67	20.33	38.67
Dirk Gucht   Yehoshua Sagiv   Raghu Ramakrishnan	11.00	6.00	56.67
Dirk Gucht   Agma Traina   Raghu Ramakrishnan	40.67	6.00	28.33
Dirk Gucht   Caetano. Junior   Raghu Ramakrishnan	40.67	6.00	28.33

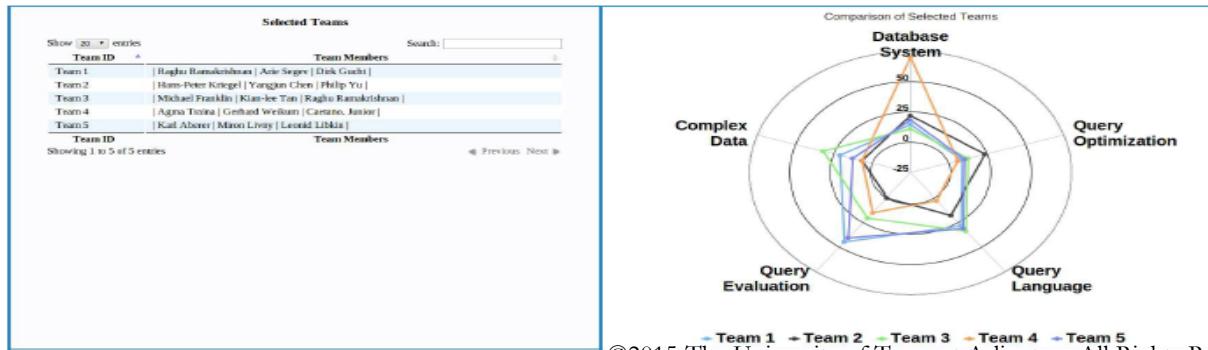
Showing 1 to 20 of 504 entries

Team Members' Names     
Previous Next

# CrewScout <http://idir.uta.edu/crewscore/>



(b) 2 experts selected



# GQBE (Graph Query-by-Example)

<http://idir.uta.edu/gqbe>

On a knowledge graph:

**Input:** one or more example n-entity tuples

**Output:** top-k most similar answer n-tuples.

The screenshot shows the GQBE interface with two input fields and a search button. The top field contains "Donald Knuth,Stanford University,Turing Award," and the bottom field contains "john mcd". A checkbox labeled "Explanatory Mode" is checked. Below the inputs is a dropdown menu titled "Select an item from the list:" containing several suggestions. The suggestion "John McCarthy" is highlighted in orange. To the right, a detailed result card for "John McCarthy" is displayed, showing a photo, his name, and a brief biography: "John McCarthy (September 4, 1927 – October 24, 2011) was an American computer scientist...".

+ Donald Knuth,Stanford University,Turing Award,

- john mcd

Select an item from the list:

John McCain	Politician
John Lennon & Paul McCartney	Theatrical Lyricist
John McClane	Fictional Character
John McCain presidential campaign	Election campaign
<b>John McCarthy</b>	Computer Scientist

[view more](#)

Explanatory Mode

**Search** **Clear**

John McCarthy

John McCarthy (September 4, 1927 – October 24, 2011) was an American computer scientist...

Computer Scientist

©2015 The University of Texas at Arlington. All Rights Reserved.



# GQBE (Graph Query-by-Example)



# GQBE



Explanatory Mode



Donald Knuth,Stanford University,Turing Award,

Search

Clear

View Maximal Query Graph

Total Answers returned: 126

Showing: 126



Stanford University



IEEE John von Neumann  
Medal



Donald Knuth

[View Answer  
Graph](#)



Stanford University



Turing Award



John McCarthy

[View Answer  
Graph](#)



Stanford University



Turing Award



Niklaus Wirth

[View Answer  
Graph](#)



Stanford University



National Medal of  
Science for Mathematics  
and Computer Science



Donald Knuth

[View Answer  
Graph](#)



Stanford University



Kyoto Prize



Donald Knuth

[View Answer  
Graph](#)



Stanford University



Turing Award



Robin Milner

[View Answer  
Graph](#)

# Orion: Auto-suggestion for Interactive Graph Query Formulation

<http://idir.uta.edu/orion>

The screenshot shows the Orion interface with a query graph on the left and a sidebar with documentation and controls on the right.

**Query Graph:**

- Nodes: "Film Actor" (grey), "?", "Film Director" (orange), and several "Film" nodes (orange).
- Edges:
  - "Film Actor" is connected to "?" via "film/produced\_by".
  - "?" is connected to "Film Director" via "writer/directed\_by".
  - "Film Director" is connected to "Film" nodes via "writer/film" and "film-starting".
  - "?" is connected to "Film" nodes via "film-starting".

**Sidebar (Edge Examples / Reverse Role):**

Click on an edge to view its source type, object type and example instances.

**Possible Actions:**

- Click on other grey nodes to be included in the query graph.
- Click on the grey edge to select it, or click on a grey edge to display the other occurrences of the grey edge.
- Click on the empty canvas to add the selected nodes and edges to the query graph while ignoring the unselected grey nodes, and display new suggestions.
- Click on selected nodes (in blue) to unselect them.

**Useful Tips:**

- To add a new node, click on empty space in the canvas.
- To add a new edge between two nodes, click on one node and drag to the other node.
- To re-position a node, hold down the shift key, then click and drag the node around.
- To remove a node, select the node by clicking on it and press the Delete button.
- To remove an edge, select the edge by clicking on it and press the Delete button.

**Buttons:**

- Survey
- Clear
- New Suggestions 3 ▾



# Entity-Relationship Queries (ERQ)

Entity-Relationship Queries - Mozilla Firefox  
File Edit View History Bookmarks Tools Help  
<http://idir.uta.edu/erq/>  
Most Visited Getting Started Latest Headlines  
Entity-Relationship Queries

## Entity-Relationship Queries

Introduction Guideline INEX Queries OWN Queries

**Entity-Relationship Query (ERQ)** is an entity-centric, structured query mechanism. You can query named entities by telling what kinds of entities you want and what are the relations among them. The current demo supports 0.75 million entities collected from Wikipedia. These entities are organized into 10 types, namely Person, Company, University, City, Novel, Player, Club, Song, Film and Award. The 2008-07-24 snapshot of Wikipedia is used as the corpus.

**Quick Examples:**

- Find who won World Chess Champion titles.
- Find Catalan universities.
- Find German cities that were in the hanseatic league.
- Find companies and their founders, where the companies are in Silicon Valley and founders are Stanford graduates.
- Find an Australian actress, an Academy Award winning film, and a Grammy Award winning song, where the actress stars the film and the song is the theme of that film.

**Four Types of Queries in ERQ**

1. SS: Single-Predicate Query (Selection Predicate)
2. SJ: Single-Predicate Query (Join Predicate)
3. MS: Multi-Predicate Query (without Join Predicate)
4. MJ: Multi-Predicate Query (with Join Predicate)

The INEX17 query set (converted from INEX topics) only contains SS and MS queries. OWN19 contains SS, MS and MJ queries. We

**Entity Type**      **Predicates**

Entity x: Person stanford graduate   
Entity y: Company "silicon valley"   
Entity z: Select

**Relationships Between Entities:**

Relationship: x and y found

**Go**

Total Answers: 25 x=18 y=14 Time 23 ms

1 [Jerry Yang](#) [Yahoo!](#)

- In January 1994, [Stanford graduate](#) students [Jerry Yang](#) and David Filo created a website named Jerry's Guide to the World Wide Web. ([see all 6](#))
- Imran is the son of Nuzhat Khan and Anil Pal, who works as a senior manager at [Yahoo](#) in [Silicon Valley](#). ([see all 7](#))
- [Jerry Yang](#) co- founded [Yahoo](#) ([see all 3](#))

2 [Scott McNealy](#) [Sun Microsystems](#)

- On February 12, 1982 Vinod Khosla, Andy Bechtolsheim, and [Scott McNealy](#), all [Stanford graduate](#) students, founded Sun Microsystems. ([see all 4](#))
- The AMBAR was founded in 2002 by a group of experienced technology entrepreneurs and business professionals from the [Silicon Valley](#) companies and venture capital firms such as [Sun Microsystems](#), Intel Capital, and Draper Fisher Jurvetson. ([see all 8](#))
- Vinod Khosla, a fellow graduate of Stanford who was an early employee at Daisy Systems Corporation convinced Bechtolsheim along with [Scott McNealy](#) to [found Sun Microsystems](#) in order to build the Sun1/ 100 workstation. ([see all 3](#))

3 [David Packard](#) [Hewlett-Packard Company](#)

<http://idir.uta.edu/erq>

# Facetedpedia <http://idir.uta.edu/facetedpedia>

In contrast, a faceted interface over the result list looks like this:



Facetedpedia  
UT Arlington

US action film  search

What kind of entities are you looking for: Film

**Facets**

<b>Living people</b>	[4]
<a href="#">(Will Smith)</a> [4]	<a href="#">(Gene Hackman)</a> [1]
<a href="#">(Bill Pullman)</a> [1]	<a href="#">(Richard Matheson)</a> [1]
<a href="#">(Jeff Goldblum)</a> [1]	<a href="#">(Tony Scott)</a> [1]
<a href="#">(Dean Devlin)</a> [1]	<a href="#">(BrIDGET Moynahan)</a> [1]
<a href="#">(Roland Emmerich)</a> [1]	

<b>Events by year</b>	[4]
<a href="#">Establishments by year</a> [4]	<a href="#">Conflicts by year</a> [3]
<a href="#">Military history by year</a> [2]	<a href="#">Disestablishments by year</a> [1]
<a href="#">Years in television</a> [1]	<a href="#">Introductions by year</a> [1]

<b>People by city in the United States by state</b>	[4]
<a href="#">People by city in Pennsylvania</a> [4]	<a href="#">People by city in New York</a> [2]
<a href="#">People by city in Illinois</a> [2]	<a href="#">People by city in California</a> [1]
<a href="#">People by city in Missouri</a> [1]	

<b>California counties</b>	[4]
----------------------------	-----

**Selected Categories:**

[\[remove\] American film actors > \(Will Smith\)](#)

**Wikipedia Articles**

• 4 Articles Selected

**Independence Day (film)**  
The *United States* military originally intended to provide personnel, vehicles, and costumes for the *film*; however, they backed out when the producers ...  
[http://en.wikipedia.org  
/wiki/Independence\\_Day\\_\(film\)](http://en.wikipedia.org/wiki/Independence_Day_(film))

**Enemy of the State (film)**  
Language, English. Budget, US\$90000000. Gross revenue, US\$250649836 .... Janet Maslin of the New York Times approved of the *film's* *action-packed* sequences, ...  
[http://en.wikipedia.org  
/wiki/Enemy\\_of\\_the\\_State\\_\(film\)](http://en.wikipedia.org<br/>/wiki/Enemy_of_the_State_(film))

**I Am Legend (film)**  
I Am Legend was released on December 14, 2007, in the *United States*, and opened to .... and that there was "a pretty heavy screenplay for an *action film* ...  
<http://en.wikipedia.org>