Restoring Trust by Computing: Data-driven Fact-checking and Exceptional Fact Finding

Chengkai Li
Professor and Associate Chair
Director, Innovative Database and Information Systems (IDIR) Lab
CSE Department, UT-Arlington

Texas Health Resources
August 29, 2019
Outline

- A Brief History of Our Computational Journalism Research

- Computational Journalism
  - Data-driven fact-checking (ClaimBuster)
  - Other ongoing fact-checking projects
  - Exceptional fact finding (FactWatcher, Maverick)

- Graph Data Usability (Orion, GQBE, TableView, Maverick)
How it Started

Chris Paul had 16 points, 10 rebounds, 13 assists and five steals...... The only other active player to have such a game is Jason Kidd...

2009

How did they come up with that?
How it Started

“Developing the Field of Computational Journalism” Workshop, July 27-31, 2009

Let’s work on this fact-finding problem.

This is computational journalism.

Summer 2010

#@%&!#!%

Computational Journalism

By Sarah Cohen, James T. Hamilton, Fred Turner
Communications of the ACM, October 2011, Vol. 54 No. 10, Pages 66-71
CIDR 2011

The CCC Outrageous Ideas and Vision Award papers (chosen by audience voting) were:
- **3rd Place:** Computational Journalism: A Call to Arms to Database Researchers by Sarah Cohen, Chengkai Li, Jun Yang, and Cong Yu

Computational Journalism:
A Call to Arms to Database Researchers
Sarah Cohen  
School of Public Policy  
Duke Univ.  
sarah.cohen@duke.edu  
Chengkai Li  
Dept. of Comp. Sci. & Eng.  
Univ. of Texas at Arlington  
dli@uta.edu  
Jun Yang  
Dept. of Comp. Sci.  
Duke Univ.  
junyang@cs.duke.edu  
Cong Yu  
Google Inc.  
congy@umich.edu

**fact-finding**
Interestingly, there is one area of news where a specialized reporter’s black box has been immensely successful—sports. It is amazing how commentaries of the form “player X is the second since year Y to record, as a reserve, at least α points, β rebounds, γ assists, and δ blocks in a game” can be generated seemingly instantaneously. Replicating this success for investigative

**fact-checking**
We will be able to *fact-check* stories quickly. Fact-checking exposes misinformation by politicians, corporations, and special-interest groups, and guards against errors and shady practices in reporting. Consider the following example from FactCheckED.org
The US stock market is off to its best start of the year since the early 1990s

By Jason Karaian • February 17, 2019
Stocks book their worst year since the financial crisis and worst December since the Great Depression

Jonathan Garber
© Dec. 31, 2018, 08:30 AM
The “Fake News” Problem

It was always there, since “Yellow Journalism” started in 1890s.

- Exaggerated headlines, clickbait, computational propaganda, misinformation, disinformation

But it became a daily talking point since 2016.

- Pizzagate
- “filter bubble” and “echo chamber” exacerbated by social media
- The many false claims in political discourses
- Russian meddling with U.S. election
- Twitter bots, Facebook ads, trendy topic algorithms
Outline

- A Brief History of our Computation + Journalism Research
- Computational Journalism
  - Data-driven fact-checking (ClaimBuster)
  - Other ongoing fact-checking projects
  - Exceptional fact finding (FactWatcher, Maverick)
- Graph Data Usability (Orion, GQBE, TableView, Maverick)
Toward the Holy Grail of Automated Fact-checking

idir.uta.edu/claimbuster
The Holy Grail: Automated, Live Fact-Checking

Source: Bill Adair
"... our Navy is smaller than it's been since 1917", said Republican candidate Mitt Romney in third presidential debate in 2012.
Fact-checking is Hard

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

http://en.wikipedia.org/wiki/Mitt_Romney
http://en.wikipedia.org/wiki/United_States_Navy
PolitiFact “Buffet” of Factual Claims

<table>
<thead>
<tr>
<th>Statement</th>
<th>Speaker</th>
<th>Date</th>
<th>Location/Source</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>232 I haven’t heard one (republican) say that they’re in concentrated</td>
<td>o’malley</td>
<td>4/20/2015 morning edition</td>
<td><a href="http://www.npr.org/blogs/its">link</a></td>
<td></td>
</tr>
<tr>
<td>233 concentrated wealth and capital have been put into</td>
<td>o’malley</td>
<td>4/20/2015 morning edition</td>
<td><a href="http://www.npr.org/blogs/its">link</a></td>
<td></td>
</tr>
<tr>
<td>234 it is not true that regulation holds poor people down or</td>
<td>o’malley</td>
<td>4/20/2015 morning edition</td>
<td><a href="http://www.npr.org/blogs/its">link</a></td>
<td></td>
</tr>
<tr>
<td>235 Every three weeks we bring online as much as solar</td>
<td>Obama</td>
<td>4/18/2015 Weekly address</td>
<td><a href="https://www.whitehouse.gov">link</a></td>
<td></td>
</tr>
<tr>
<td>236 Our carbon pollution has fallen by 10 percent since</td>
<td>Obama</td>
<td>4/19/2015 Weekly address</td>
<td><a href="https://www.whitehouse.gov">link</a></td>
<td></td>
</tr>
<tr>
<td>237 “Benghazi has had more hearings, more documents”</td>
<td>Claire McCaskill</td>
<td>4/19/2015 ABC this week</td>
<td><a href="http://thehill.com/policy/inte">link</a></td>
<td></td>
</tr>
<tr>
<td>238 With fast track trade “I would be receiving the same</td>
<td>Obama</td>
<td>4/17/2015 remarks joint press</td>
<td><a href="https://www.whitehouse.gov">link</a></td>
<td></td>
</tr>
<tr>
<td>239 Loretta Lynch “has been now sitting there longer than</td>
<td>Obama</td>
<td>4/17/2015 remarks joint press</td>
<td><a href="https://www.whitehouse.gov">link</a></td>
<td></td>
</tr>
<tr>
<td>240 I just counted 19 Republicans who are likely to run for</td>
<td>Chris Cillizza</td>
<td>4/20/2015 the twitters</td>
<td><a href="https://twitter.com/TheFix/st">link</a></td>
<td></td>
</tr>
<tr>
<td>241 He [Rand Paul] wanted to cut all our defense</td>
<td>John McCain</td>
<td>4/20/2015 Fox News</td>
<td><a href="https://www.youtube.com/w">link</a></td>
<td></td>
</tr>
</tbody>
</table>

Bill Adair is the founder of the Pulitzer Prize-winning website PolitiFact and Knight Professor of the Practice of Journalism and Public Policy at Duke University, where
Finding Important Factual Claims: A Supervised Learning Task

Presidential Debate Transcripts (1960-2012) → Human Annotation → Ground Truth → Feature Extraction → Feature Vectors → Learning Algorithm → CLAIMBUSTER → 2016 Presidential Debates → Important factual claims
Classification and Ranking by Check-worthiness

CFS: 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.”

UFS: Unimportant factual claims
“I was in Iowa yesterday.” “My mother enjoys cooking.” “I ran for President once before.”

NFS: No factual claims (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%.”
Your estimated total payment is $37.39. You are paid approximately 6.70¢ per sentence on average. Your total payment is ranked 8 out of 54 participants. Click 'Leaderboard' to see details. This message and the leaderboard are updated every 15-30 sentences. See tips for improving your pay rate at the bottom of this page.

RO: Now, I voted for the Patriot Act.

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.**
Ground Truth Collection

- 20788 sentences
- 374 coders
- 76552 labels
- 86 top-quality coders
- 52333 labels

- 20 months, 374 coders, ~$4,000 paid
- 30 training sentences
- 1032 screening sentences (731 NFS, 63 UFS, 238 CFS) to detect spammers & low-quality coders

<table>
<thead>
<tr>
<th>Class</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFS</td>
<td>4849</td>
<td>23.52%</td>
</tr>
<tr>
<td>UFS</td>
<td>2097</td>
<td>10.17%</td>
</tr>
<tr>
<td>NFS</td>
<td>13671</td>
<td>66.31%</td>
</tr>
</tbody>
</table>

http://www.engineersdaily.com/2014/03/basics-of-soil-compaction.html
2016 Third Presidential Debate Live. Oct. 19, 2016, 9 p.m. EST
Participants: Donald Trump, Hillary Clinton
Moderators: Chris Wallace

2016 Third Presidential Debate. Oct. 19, 2016, 9 p.m. EST
Participants: Hillary Clinton, Donald Trump
Moderators: Chris Wallace

2016 Second Presidential Debate. Oct. 9, 2016, 9 p.m. EST
Participants: Hillary Clinton, Donald Trump
Moderators: Martha Raddatz, Anderson Cooper, Chris Wallace

2016 Vice Presidential Debate. Oct. 4, 2016, 9 p.m. EST
Participants: Tim Kaine, Mike Pence, Hillary Clinton, Donald Trump
Moderators: Elaine Quijano, Martha Raddatz, Anderson Cooper

2016 First Presidential Debate. Sep. 26, 2016, 9 p.m. EST
Participants: Hillary Clinton, Donald Trump
Moderators: Lester Holt

2016 Democratic Party Presidential Debate. April 14, 2016, 9 p.m. EST
Participants: Bernard Sanders, Hillary Clinton, Donald Trump
LSTM-based Model for Claim Spotting

- Model based on an LSTM (long short-term memory) network
- Tested several different word embeddings but found no significant difference in overall performance
  - Different embeddings did have different affinities (i.e., some produce models which score text with numbers higher on average).
- We used nonsensical sentences to make the model more resistant to attacks. The neural models outperformed the SVM model by 29% in precision and 12% in recall.
Adversarially Trained LSTM

- Training the LSTM model on adversarial (modified) examples.

- Adversarial noise (perturbations), designed to maximize the chance of misclassification by the network, was calculated using backpropagation.

- Perturbation was applied to word-embeddings rather than directly to the input.
Adversarially Trained LSTM

Recall

Precision

F1-score
| GET | /all/{claim} | All |
| GET | /se/{claim} | Claim Checker - Search Engine |
| GET | /fm/{claim} | Claim Matcher |
| GET | /kb/{claim} | Claim Checker - Knowledge Bases |
| GET | /score_text/{text} | Claim Spotter - Get Score |
| GET | /score_url/{url} | Claim Spotter - Get Score URL |
From the Congressional Record

Statements by members of the House and Senate as officially transcribed from floor speeches and debates.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Claim</th>
<th>Chamber</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. REED</td>
<td>On June 4, 1919, citizens of Tioga County joined together and signed a charter to create the Owego Chamber of Commerce.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Ms. ESCOBAR</td>
<td>According to Mr. Bradford’s son, for more than 30 days, Mr. Bradford avoided capture and eventually was able to find his way back to the American forces.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Ms. ESCOBAR</td>
<td>Madam Speaker, I rise today to honor the life of an American hero, Bloyce Bradford, who passed away at 90 years old on Tuesday, January 29, 2019.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Ms. ESCOBAR</td>
<td>While in Korea, Mr. Bradford was wounded multiple times and, at one point, was considered missing in action.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>The PRESIDING OFFICER</td>
<td>The senior assistant legislative clerk read the following letter: U.S. Senate, President pro tempore, Washington, DC, February 15, 2019.</td>
<td>Senate</td>
<td>Link</td>
</tr>
<tr>
<td>Mr. REED</td>
<td>Given the above, I ask that this Legislative Body pause in its deliberations and join me to celebrate the Tioga County Chamber of Commerce's one hundred years of service.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>The PRESIDING OFFICER</td>
<td>To the Senate: Under the provisions of rule I, paragraph 3, of the Standing Rules of the Senate, I hereby appoint the Honorable Ben Sasse, a Senator from the State of Nebraska, to perform the duties of the Chair.</td>
<td>Senate</td>
<td>Link</td>
</tr>
<tr>
<td>Mr. REED</td>
<td>The business show brings new business to existing businesses together, putting them in contact with each other and the general public.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Mr. REED</td>
<td>By purchasing properties and then improving those properties, the Chamber helped to create environments that welcomed new businesses.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Ms. ESCOBAR</td>
<td>His legacy lives on through his sister Gwenn, whom he raised five children and the many nieces and nephews from his long marriage to his companion of 43 years and her family.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Mr. REED</td>
<td>During World War II, the Chamber assisted with the setting of war bonds to aid in the war effort.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Ms. ESCOBAR</td>
<td>He went on to enroll in the Army, where he served during the Korean War.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>The Owego Chamber of Commerce</td>
<td>The management of real estate.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>Mr. REED</td>
<td>I congratulate the Tioga County Chamber for supporting businesses in our area.</td>
<td>House</td>
<td>Link</td>
</tr>
<tr>
<td>The PRESIDING OFFICER</td>
<td>Mr. Sasse thereupon assumed leadership of the House.</td>
<td>Senate</td>
<td>Link</td>
</tr>
</tbody>
</table>

PolitiFact “Buffet” of Factual Claims

<table>
<thead>
<tr>
<th>Statement</th>
<th>Speaker</th>
<th>Date</th>
<th>Location/Source</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>I haven’t heard one (Republican) say that they’re in concentrated wealth and capital have been put into</td>
<td>o’malley</td>
<td>4/20/2015</td>
<td>morning edition</td>
<td><a href="http://www.npr.org/blogs/gts">http://www.npr.org/blogs/gts</a></td>
</tr>
<tr>
<td>it is not true that regulation holds poor people down or</td>
<td>o’malley</td>
<td>4/20/2015</td>
<td>morning edition</td>
<td><a href="http://www.npr.org/blogs/gts">http://www.npr.org/blogs/gts</a></td>
</tr>
<tr>
<td>every three weeks we bring online as much as solar</td>
<td>Obama</td>
<td>4/18/2015</td>
<td>Weekly address</td>
<td><a href="https://www.whitehouse.gov">https://www.whitehouse.gov</a></td>
</tr>
<tr>
<td>our carbon pollution has fallen by 10 percent since</td>
<td>Obama</td>
<td>4/19/2015</td>
<td>Weekly address</td>
<td><a href="https://www.whitehouse.gov">https://www.whitehouse.gov</a></td>
</tr>
<tr>
<td>Bengtson has had more hearings, more documents</td>
<td>Claire McCaskill</td>
<td>4/19/2015</td>
<td>ABC this week</td>
<td><a href="http://thetot.com/politivizere">http://thetot.com/politivizere</a></td>
</tr>
<tr>
<td>with fast track trade “I would be receiving the same</td>
<td>Obama</td>
<td>4/17/2015</td>
<td>remarks joint press</td>
<td><a href="https://www.whitehouse.gov">https://www.whitehouse.gov</a></td>
</tr>
<tr>
<td>I love a lunch “it has been so good,” longer than</td>
<td>Obama</td>
<td>4/17/2015</td>
<td>remarks joint press</td>
<td><a href="https://www.whitehouse.gov">https://www.whitehouse.gov</a></td>
</tr>
<tr>
<td>I just counted 19 Republicans who are likely to run for congress</td>
<td>frye</td>
<td>4/20/2015</td>
<td>the twitter</td>
<td><a href="https://twitter.com/TheFix/st">https://twitter.com/TheFix/st</a></td>
</tr>
<tr>
<td>We [Rand Paul] wanted to cut all our defense</td>
<td>John McCain</td>
<td>4/20/2015</td>
<td>Fox News</td>
<td><a href="https://www.youtube.com/w">https://www.youtube.com/w</a></td>
</tr>
</tbody>
</table>

25
Duke Reporters’ Lab uses ClaimBuster API in creating daily newsletters that recommend to The Washington Post, New York Times, PolitiFact, and other fact-checkers the most check-worthy claims in CNN programs, Tweets, Facebook posts, and the Congressional Record.

“There were 10 million people that would have been on some path to staying in this country for indefinitely in the proposal that [Sens.] Dick Durbin and Lindsey Graham put forward.”
— Former senator Rick Santorum (R-Pa.), in remarks on CNN’s “State of the Union,” Jan. 21, 2018

“I had actually looked at that transcript but I missed that claim,” said Glenn Kessler, who runs The Fact Checker. “So it would have been lost to history if it had not been for ClaimBuster.”
Claim Spotter
The First-ever End-to-end Fact-checking System

[CIKM15a, KDD17, VLDB17demo, IJCNN18]
* First Runner-Up, SIGMOD17 Student Research Competition
The biggest complaint they have -- it's with all of the problems going on in the world, many of the problems caused by Hillary Clinton and by Barack Obama.

---

The single biggest problem in communication is the illusion that it has...
Outline

- A Brief History of our Computation + Journalism Research
- Computational Journalism
  - Data-driven fact-checking (ClaimBuster)
  - Other ongoing fact-checking projects
  - Exceptional fact finding (FactWatcher, Maverick)
- Graph Data Usability (Orion, GQBE, TableView, Maverick)
ClaimPortal: Integrated Monitoring, Searching, Checking, and Analytics of Factual Claims on Twitter

https://idir.uta.edu/claimportal/
Outline

- A Brief History of our Computation + Journalism Research
- Computational Journalism
  - Data-driven fact-checking (ClaimBuster)
  - Other ongoing fact-checking projects
  - Exceptional fact finding (FactWatcher, Maverick)
- Graph Data Usability (Orion, GQBE, TableView, Maverick)
How did they come up with that?

Chris Paul had 16 points, 10 rebounds, 13 assists and five steals...... The only other active player to have such a game is Jason Kidd...
Exceptional Fact Finding

Real-world event (sports, transportation, crime, weather, finance, social media)

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.

Number-based facts make news stories more engaging

http://en.wikipedia.org/wiki/Basketball
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.

idir.uta.edu/factwatcher

Excellent Demo Award
Fact-finding Algorithms

FactWatcher [VLDB14 demo]

- [ICDE14] Situational Facts: “No other player scored more pts and reb against DAL than Jordan.”

- [KDD12] One-of-the-Few: “Jordan scored 10 pts & 10 reb. Only 3 others had similar performance.”

- [KDD11] Prominent Streaks: “The Nikkei 225 closed below 10000 for the 12th consecutive week, the longest such streak since June 2009.”

- [TKDD14] General Prominent Streaks: “James has scored at least 20 points and handed out 10 or more assists in each of his last five games against the Hawks, the longest streak he has ever had against one team.”

Frequent Episode Mining [ICDE15]
Fact-finding Algorithms

- Many interesting facts in the real-world can be modeled as various types of skyline points.
- The gist of fact-finding is to efficiently, incrementally maintain skyline points over ever-changing data, while considering constraints such as selection conditions.

Stephan Börzsönyi, Donald Kossmann, Konrad Stocker: The Skyline Operator. ICDE 2001: 421-430
“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/eliases?date=20130205)
“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/elia?date=20130205)
## Modeling Situational Facts

### Dimension space: $\mathcal{D} = \{d_1, \ldots, d_n\}$

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

### Measure space: $\mathcal{M} = \{m_1, \ldots, m_s\}$

Skyline maintenance + Data Cube/OLAP

$$\text{Skyline maintenance} + \text{Data Cube/OLAP}$$
Outline

- A Brief History of our Computation + Journalism Research
- Computational Journalism
  - Data-driven fact-checking (ClaimBuster)
  - Other ongoing fact-checking projects
  - Exceptional fact finding (FactWatcher, Maverick)
- Graph Data Usability (Orion, GQBE, TableView, Maverick)
Tackling Graph Data Usability Challenges

Challenges
- Massive, complex graphs; millions of entities; billions of edges.
- Requires substantial understanding of schema and data and complex pre-processing, before one can fetch information or gain insights from data.

Objectives
- Make it easy to understand, query, explore, and clean graph data.

Systems
- GQBE [TKDE15, ICDE14demo]: graph query by example
- Orion [VLDB15demo, SIGMOD17tutorial]: auto-suggestion for interactive query formulation
- TableView [SIGMOD16, ICDE18demo]: generating preview tables for knowledge graphs
  (* SIGMOD Most Reproducible Paper Award)
- Maverick [SIGMOD18, VLDB18demo]: finding outliers and errors in graphs
**IDIR Projects and Demos**

**Demonstration Videos** [https://vimeo.com/channels/1024406](https://vimeo.com/channels/1024406)

- **ClaimBuster** (idir.uta.edu/claimbuster)  Automated, live fact-checking
- **ClaimPortal** (idir.uta.edu/claimportal)  Monitoring factual-claims on social media
- **CrewScout** (idir.uta.edu/crewscout)  Expert team finding by skyline groups
- **ERQ: Entity-Relationship Query** (idir.uta.edu/erq)  Structured query on Wikipedia
- **Facetedpedia** (idir.uta.edu/facetedpedia)  Faceted search interface for Wikipedia
- **FactWatcher** (idir.uta.edu/factwatcher)  Fact-finding from real-world events
- **FrameAnnotator** (idir.uta.edu/frameannotator)  Frame annotation tool
- **GQBE** (idir.uta.edu/gqbe)  Graph query by example
- **Orion** (idir.uta.edu/orion)  Auto-suggestion for interactive graph query formulation
- **TableView**  Generating preview tables for knowledge graphs
- **Maverick**  Exceptional fact finding from knowledge graphs
Current IDIR Students

Ph.D. students

Farahnaz Akrami
Fatma Arslan
Israa Jaradat
Damian Jimenez
Shadekur Rahman
Samiul Saeef
Xiao Shi
Theodora Toutountzi
Zeyu Zhang

M.S. students

Priyank Arora
Sumeet Lubal
Sarthak Majithia
Daniel Obembe
Sarbajit Roy

B.S. students

Kyrell Dixon
Jacob Devasier
Graduated Students

**Ph.D. students**

Aditya Telang
2011 (co-advised)
IBM Research India

Ning Yan
2013
Research Scientist
Huawei, Santa Clara

Naeemul Hassan
2016
Assistant Professor
University of Mississippi

Nandish Jayaram
2016
Pivotal

Gensheng Zhang
2017
Google

Afroza Sultana
2018
Teradata Lab

**M.S. Thesis students**

Tulsi Chandwani (2017, Red Hat)
Abu Ayub Ansari Syed (2017)
Ishwor Timilsina (2017, Fidelity)
Nigesh Shakya (2017)
Rohit Bhoopalam (2016, Akamai)
Fatma Dogan (2015, UTA Ph.D.)
Minumol Joseph (2015, Capital One)

Ramesh Venkataraman (2014, Amazon)
Mahesh Gupta (2012, Electronic Arts)
Jijo Philip (2012, Cerner Corporation)
Avinash Bharadwaj (2011, Copper Labs)
Quazi (Sunny) Hasan (2010, Dematic)
Jared Ashman (2010, Ambit Energy)
Muhammad Safiullah (2008, Microsoft)

**B.S. students**

Josue Caraballo (2017)
Damian Jimenez (2017, UTA Ph.D.)
Long Ly (2015)
Sidharth Goyal (2015)
Huadong Feng (2014)
Raju Karki (2012)
Angus Helm (2010)
Aakash Tuli (2010)
Collaborators on Current and Past IDIR Projects

- Bill Adair (Duke, Public Policy)
- Pankaj Agarwal (Duke)
- Xiang Ao (Chinese Academy of Sciences)
- Vassilis Athitsos (UTA)
- Sourav Bhowmick (Nanyang Technological University)
- Sharma Chakravarthy (UTA)
- Gong Cheng (Nanjing University)
- Byron Choi (Hong Kong Baptist University)
- Christoph Csallner (UTA)
- Gautam Das (UTA)
- Chris Ding (UTA)
- Ramez Elmasri (UTA)
- Leonidas Fegaras (UTA)
- Peter Fray (University of Technology Sydney, Journalism)
- James Hamilton (Stanford, Communication)
- Naemul Hassan (Mississippi)
- Wei Hu (Nanjing University)
- Arijit Khan (Nanyang Technological University)
- Angela Lee (UTD, Communication)
- Zhiqiang Lin (Ohio State)
- Ping Luo (Chinese Academy of Sciences)
- Mark Stancel (Duke, Public Policy)
- Mark Tremayne (UTA, Communication)
- Min Wang (Google)
- Xifeng Yan (UCSB)
- Jun Yang (Duke)
- Cong Yu (Google Research)
- Nan Zhang (Penn State)
Thank you!