

Data In, Fact Out: Automated Monitoring of Facts by FactWatcher

Naeemul Hassan¹, Afroza Sultana¹, You Wu², Gensheng Zhang¹, Chengkai Li¹, Jun Yang², Cong Yu³
¹University of Texas at Arlington, ²Duke University, ³Google Research



Motivation

Factual statements backed by data are leads to news stories:

- **Social Media:** The social world's most viral photo ever generated **3.5 million likes, 170,000 comments and 460,000 shares** by Wednesday.
- **Weather:** 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.
- **Politics:** Dick Durbin is one of the only three candidates who have raised more than **25% from PAC contributions and 25% from self-financing**.



Concepts and Objectives

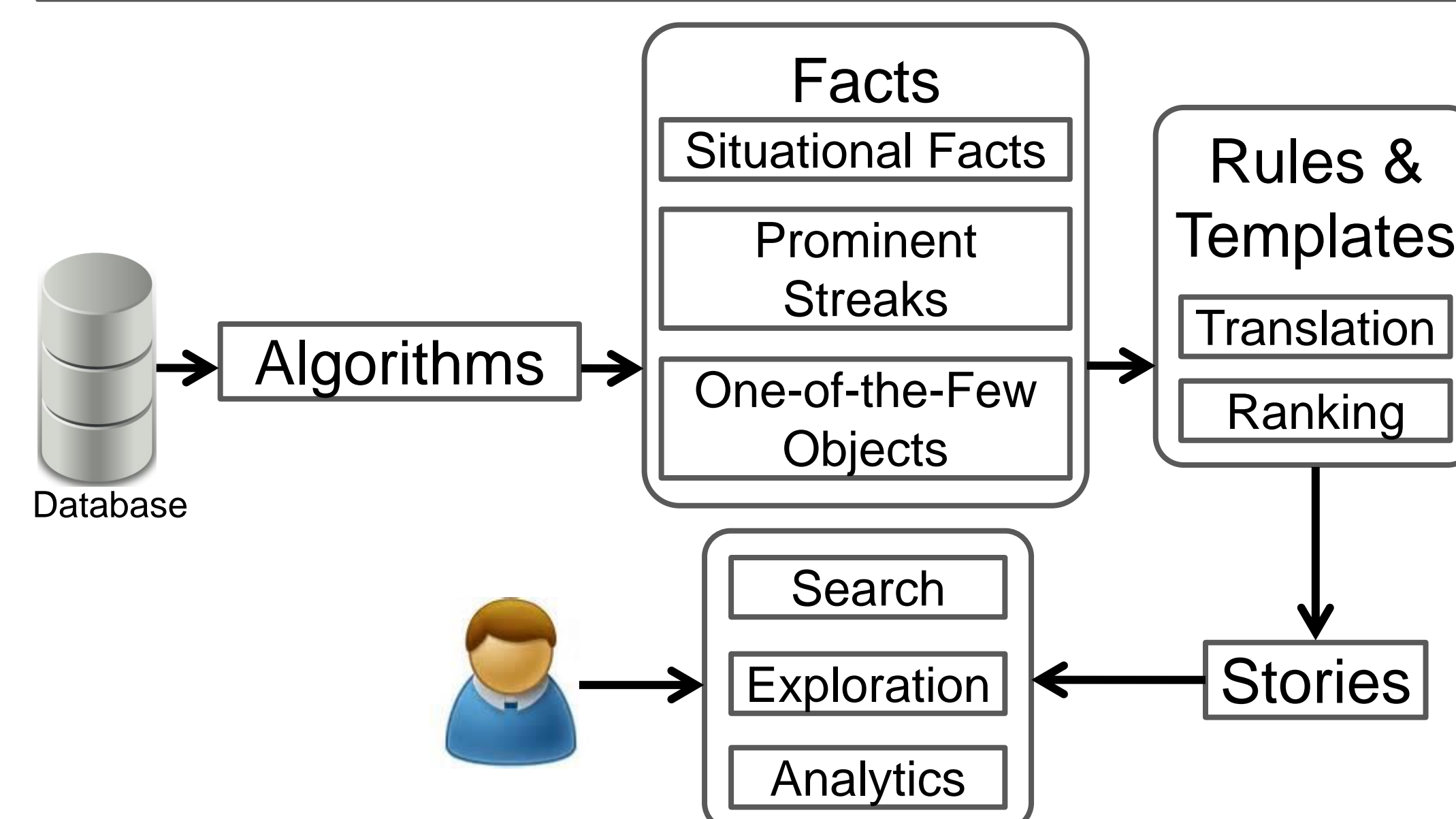
- Append-only database $R(\mathcal{D}, \mathcal{M})$, latest tuple t
 - Dimension attributes $\mathcal{D} = \{\text{Player, Team, Season}\}$
 - Measure attributes $\mathcal{M} = \{\text{Points, Assists, Blocks}\}$
 - $t = t_g: \{\text{Lamar Odom, Miami Heat, 2004, 30, 11, 11}\}$
- C : a conjunctive expression on \mathcal{D} , M : a subspace of \mathcal{M}
 - $C: \{\text{Player} = \text{Lamar Odom, Team} = *, \text{Season} = 2004\}$
 - $M: \{\text{Points, Blocks}\}$
- **Goal:** Find 3 types of facts pertinent to the arrival of t
 - **Situational Facts**^[2]: (C, M) pairs where t is in the skyline
 - $(C: \{\text{Player} = *, \text{Team} = *, \text{Season} = 2004\}, M: \{\text{Assists, Blocks}\})$
 - Lamar Odom scored 11 assists and 11 blocks. No one made a better performance in season 2004 so far.
 - **Prominent Streaks**^{[3],[5]}: (C, M) pairs where t generates prominent streaks
 - $(C: \{\text{Player} = \text{Lamar Odom, Team} = 2004, \text{Season} = *\}, M: \{\text{Length, Points, Assists}\})$
 - Lamar Odom had at least 28 points and 9 or more assists for 4 consecutive games; the longest such streak in 2004.
 - **One-of-the-Few Objects**^[4]: (C, M) pairs where t is in the top-T skyband
 - $(C: \{\text{Player} = \text{Lamar Odom, Team} = *, \text{Season} = *\}, M: \{\text{Points, Blocks}\})$
 - Lamar Odom scored 30 points & 11 blocks. Only 2 other performances have ever beaten this record. The same claim can be made for three performance records only.

Id	\mathcal{D}			\mathcal{M}		
	Player	Team	Season	Points	Assists	Blocks
t_1	Lamar Odom	LA Clippers	2003	30	11	12
t_2	Eddie House	Miami Heat	2003	29	7	8
t_3	Eddie House	Miami Heat	2003	28	6	9
t_4	Lamar Odom	Miami Heat	2004	32	9	13
t_5	Lamar Odom	Miami Heat	2004	28	11	6
t_6	Lamar Odom	Miami Heat	2004	29	9	7
t_7	Eddie House	LA Clippers	2004	30	11	10
t_8	Lamar Odom	Miami Heat	2004	30	11	11

Challenges

- Huge space of possible facts
 - 216 (C, M) pairs in the above tiny table.
- Fast, monitoring algorithm for reporting facts.
- Unified fact data model.
- Unified fact ranking.

System Architecture

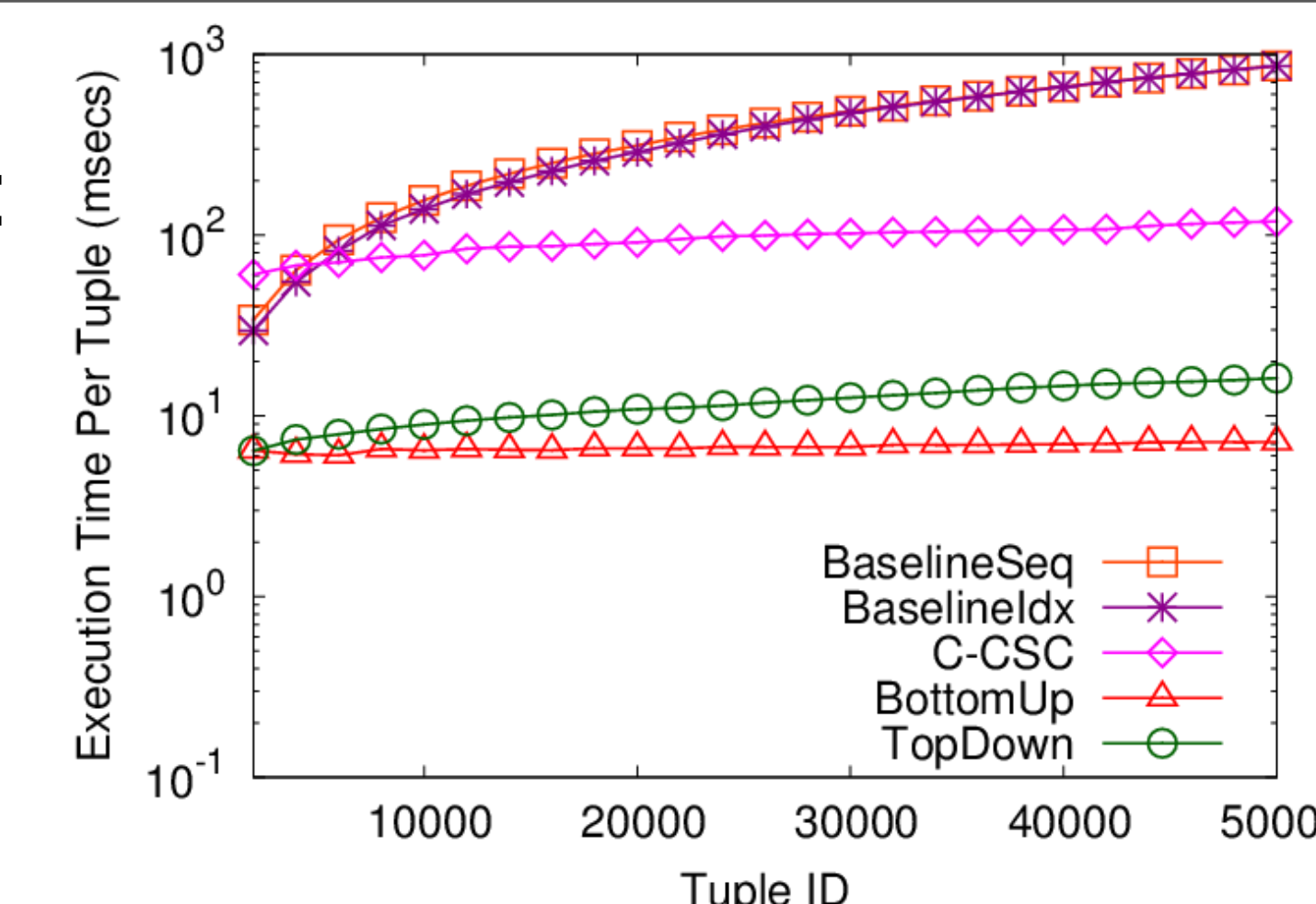


FactWatcher GUI

Main Interface

Experiment Results

NBA dataset:
 $|D| = 5$
 $|M| = 7$

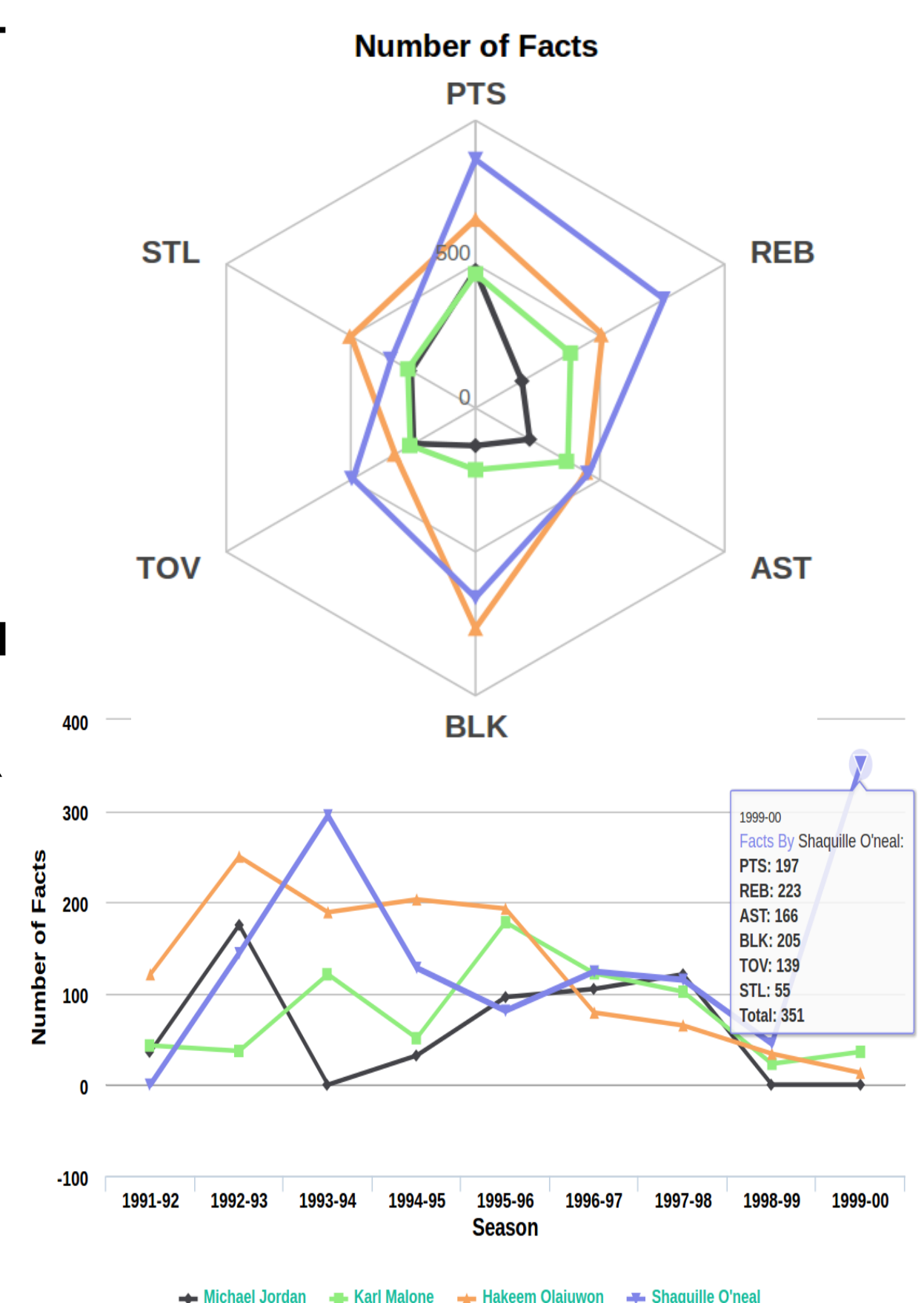


Funding



Comparison of Players

Similar Stories



Publications

1. Data In, Fact Out. Automated Monitoring of Facts by FactWatcher. Hassan et al., VLDB 2014 demo.
2. Incremental discovery of prominent situational facts. Sultana et al., ICDE 2014.
3. Discovering general prominent streaks in sequence data. Zhang et al., TKDD 2014.
4. On one of the few objects. Wu et al., KDD 2012
5. Prominent streak discovery in sequence data. Jiang et al., KDD 2011

idir.uta.edu/factwatcher

