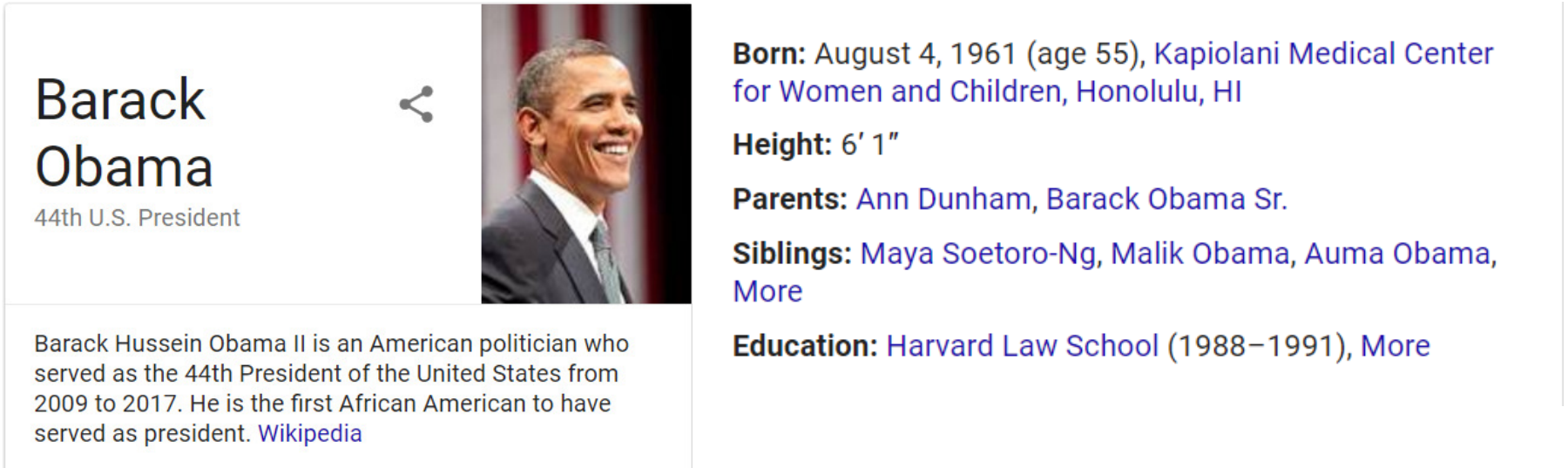
“Tell Me More” Using Ladders in Wikipedia

Siarhei Bykau (Bloomberg L.P.)
Jihwan Lee (Purdue University)
Divesh Srivastava (AT&T Labs-Research)
Yannis Velegrakis (University of Trento)

What is "Tell Me More"?

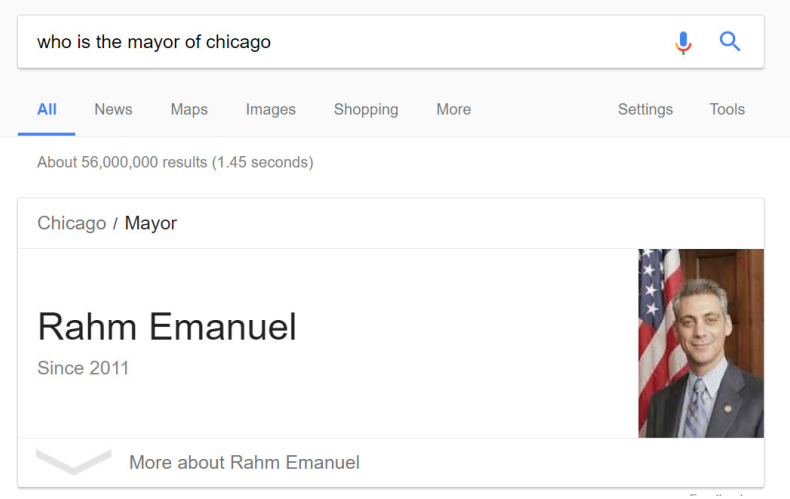
- We often ask entity queries such as "Who is Barack Obama?"
- Given the query answer, we would typically like to **know more**.



What is "Tell Me More"?

- Consider a **role query** "Who is the mayor of Chicago?"

- In this case, what should "tell me more" return?
  - More about the answer entity?



Born: November 29, 1959 (age 57), Chicago, IL

Height: 5′ 7″

Nationality: American

Spouse: Amy Rule (m. 1994)

What is "Tell Me More"?

- Consider a **role query** "Who is the mayor of Chicago?"

- In this case, what should "tell me more" return?
  - More about the answer entity? More about the role?

Born: November 29, 1959 (age 57), Chicago, IL

Height: 5′ 7″

Nationality: American

Spouse: Amy Rule (m. 1994)

On February 23, 2011, former Illinois Congressman and White House Chief of Staff, Rahm Emanuel, won the mayoral election, beating five rivals with 55 percent of the vote,[82] and was sworn in as Mayor on May 16, 2011.

Focus of this Paper

- "Tell me more" **about the role** of role queries.
  - Role query answerable using Infobox attribute in Wikipedia page.
  - E.g., "Mayor" in "Chicago" page.

- Look at text on Wikipedia page.
  - Find **attribute** in text.
  - E.g., find mayor Rahm Emmanuel.
  - Return surrounding text.

- Raises interesting challenges.

Location in the United States	
Coordinates: 41°50′13″N 87°41′05″W[1]	
Country	United States
State	Illinois
Counties	Cook and DuPage
Settled	1780s
Incorporated (town)	August 12, 1833
Incorporated (city)	March 4, 1837
Founded by	Jean Baptiste Point du Sable
Named for	Miami-Illinois: *shikaakwa (wild onion or wild garlic)*
Government	
• **Type**	Mayor–council
• **Body**	Chicago City Council
• **Mayor**	Rahm Emanuel (D)
• **City Clerk**	Anna Valencia (D)
• **City Treasurer**	Kurt Summers, Jr. (D)

Role Linkage in Wikipedia Page

- Approach 0: find Infobox attribute value in Wikipedia page text.
  - Works for "**Rahm Emanuel** is the mayor of Chicago".
  - Value is typically a Wikilink.

> On February 23, 2011, former Illinois Congressman and White House Chief of Staff, Rahm Emanuel, won the mayoral election, beating five rivals with 55 percent of the vote,[82] and was sworn in as Mayor on May 16, 2011.

Location in the United States	
Coordinates: 41°50′13″N 87°41′05″W[1]	
Country	United States
State	Illinois
Counties	Cook and DuPage
Settled	1780s
Incorporated (town)	August 12, 1833
Incorporated (city)	March 4, 1837
Founded by	Jean Baptiste Point du Sable
Named for	Miami-Illinois: *shikaakwa (wild onion or wild garlic)*
Government	
• **Type**	Mayor–council
• **Body**	Chicago City Council
• **Mayor**	Rahm Emanuel (D)
• **City Clerk**	Anna Valencia (D)
• **City Treasurer**	Kurt Summers, Jr. (D)

Role Linkage in Wikipedia Page

?

- Approach 0: find Infobox attribute value in Wikipedia page text.
  - Does it work for "**New York City** is the largest city of US"?

largest immigrant population.[26] The capital is Washington, D.C., and the largest city is New York City; nine other major metropolitan areas—each with at least 4.5 million inhabitants and the largest having more than 13 million people—are Los Angeles, Chicago, Dallas, Houston, Philadelphia, Miami, Atlanta, Boston, and San Francisco.

The United States and its territories

Capital	Washington, D.C. 38°53′N 77°01′W
Largest city	New York City 40°43′N 74°00′W
Official languages	None at federal level[fn 1]
National language	English[fn 2]

Role Linkage in Wikipedia Page

X

- Approach 0: find Infobox attribute value in Wikipedia page text.
  - Does it work for "**New York City** is the largest city of US"?

On September 11, 2001, Al-Qaeda terrorists struck the World Trade Center in New York City and the Pentagon near Washington, D.C., killing nearly 3,000 people.[165] In response,

Ellis Island in New York City was a major gateway for European immigration.

The United States and its territories

Capital	Washington, D.C. 38°53′N 77°01′W
Largest city	New York City 40°43′N 74°00′W
Official languages	None at federal level[fn 1]
National language	English[fn 2]

Role Linkage in Wikipedia Page

?

- Approach 1: Approach 0 + synonym of attribute name in vicinity.
  - Does it work for "**New York City** is the largest city of US"?

On September 11, 2001, A Qa da terrorists struck the World Trade Center in New York C and the Pentagon near Washington, D.C., killing nearly 3,000 people.[165] In response,

Ellis Island in New rk City was a major gateway f Eu pean immigration.

The United States and its territories

Capital	Washington, D.C. 38°53′N 77°01′W
Largest city	New York City 40°43′N 74°00′W
Official languages	None at federal level[fn 1]
National language	English[fn 2]

Role Linkage in Wikipedia Page

- Approach 1: Approach 0 + synonym of attribute name in vicinity.
  - Does it work for "**New York City** is the largest city of US"?

largest immigrant population.[26] The capital is Washington, D.C., and the largest city is New York City; nine other major metropolitan areas—each with at least 4.5 million inhabitants and the largest having more than 13 million people—are Los Angeles, Chicago, Dallas, Houston, Philadelphia, Miami, Atlanta, Boston, and San Francisco.

The United States and its territories

Capital	Washington, D.C. 38°53′N 77°01′W
Largest city	New York City 40°43′N 74°00′W
Official languages	None at federal level[fn 1]
National language	English[fn 2]

Role Linkage in Wikipedia Page

X

- Approach 1: Approach 0 + synonym of attribute name in vicinity.
  - Does it work for "Chicago was incorporated on **August 12, 1833**"?

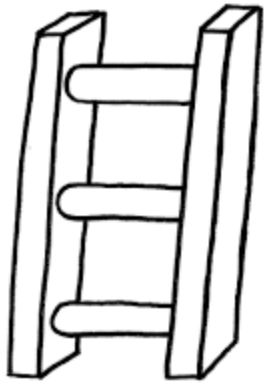
On August 12, 1833, the Town of Chicago was organized with a population of about 200.[28] Within seven years it grew to more than 4,000 people. On June 15, 1835, the first public land sales began with Edmund Dick Taylor as U.S. Receiver of Public Monies. The City of Chicago was incorporated on Saturday, March 4, 1837 a

Country	United States
State	Illinois
Counties	Cook and DuPage
Settled	1780s
Incorporated (town)	August 12, 1833
Incorporated (city)	March 4, 1837
Founded by	Jean Baptiste Point du Sable
Named for	Miami-Illinois: *shikaakwa (wild onion or wild garlic)*

Role Linkage in Wikipedia Page

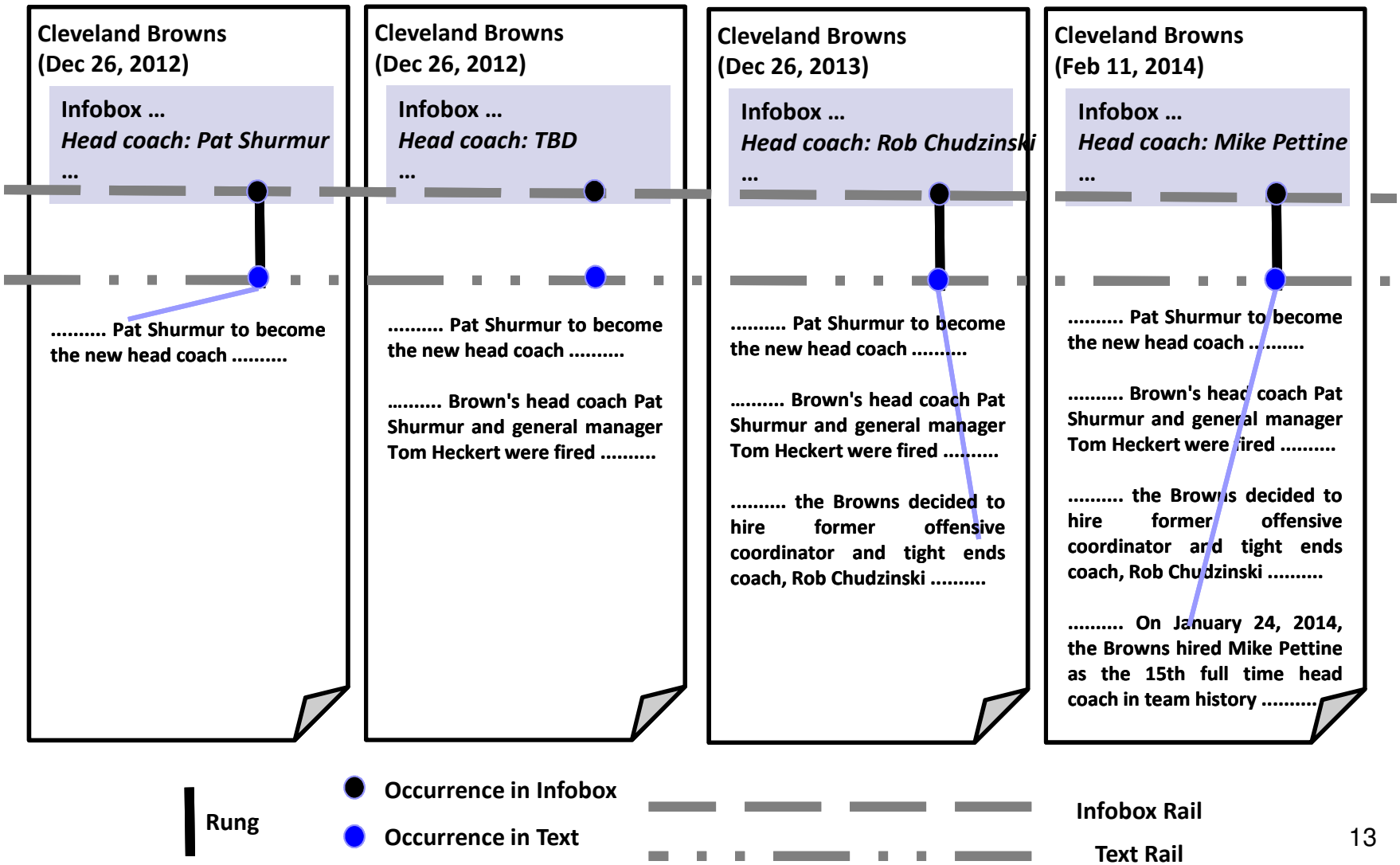
- Approach 2: Use ladders to capture **coordinated attribute value changes between IB and text** across revisions of Wikipedia page.

- What is a ladder?
  - Rung: spans between an Infobox attribute and text.
  - Rail: spans between adjacent temporal revisions.

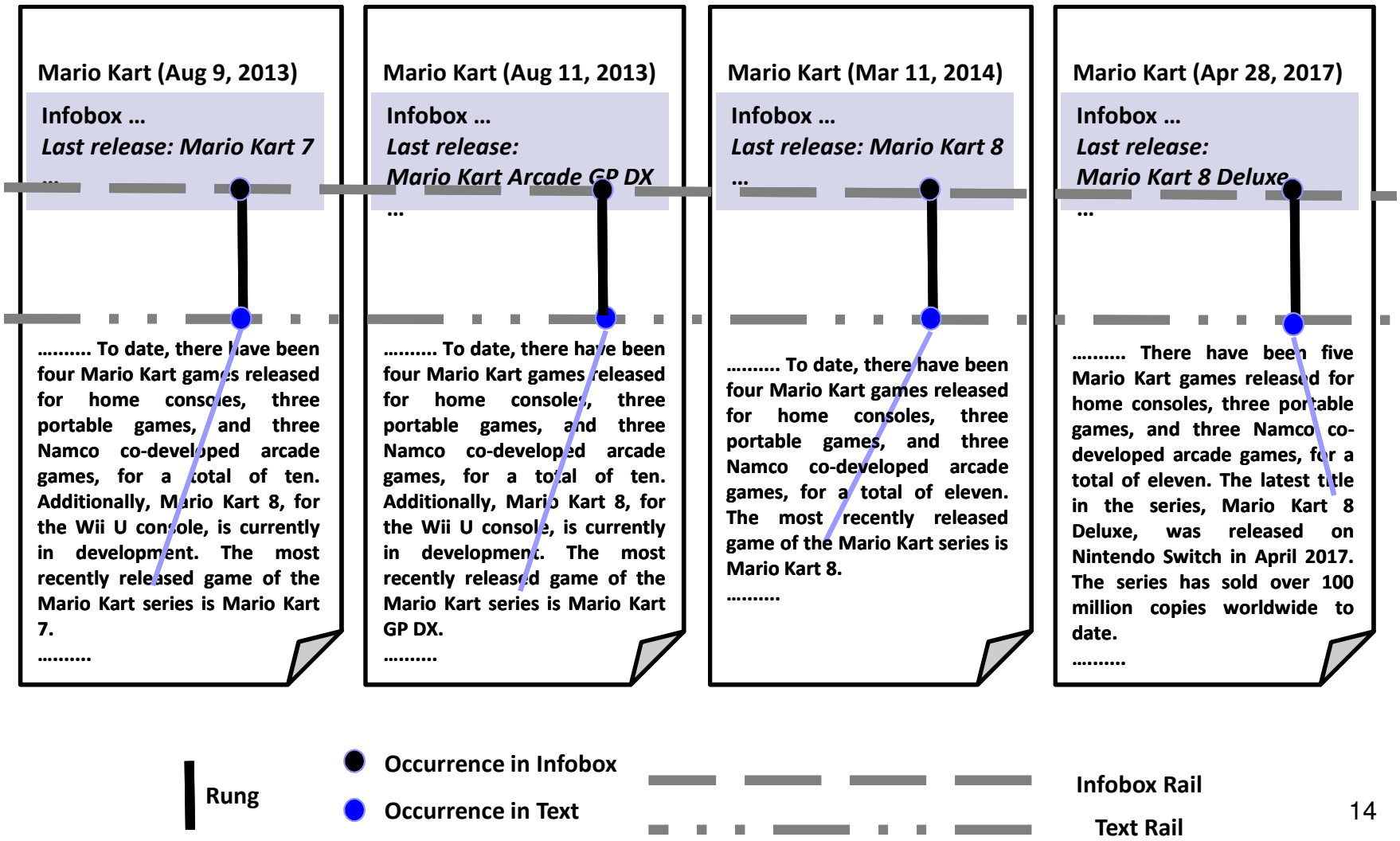


- Role linkage has three components:
  - Spatial linkage: use Approach 0 or 1 to find candidate rungs.
  - Infobox temporal linkage: same attribute name across revisions.
  - Text temporal linkage: identify **corresponding text across revisions**.

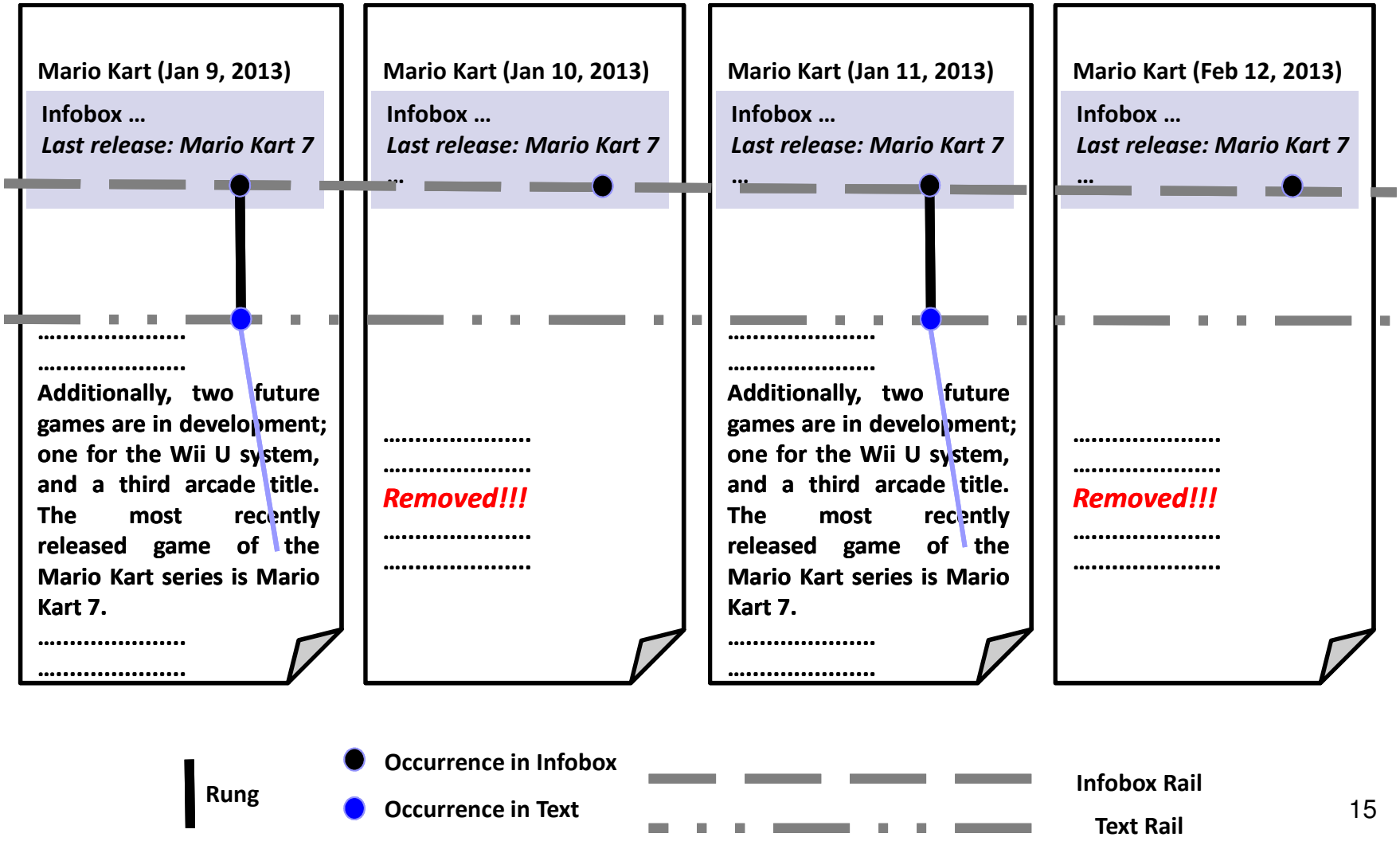
Building Ladders in Wikipedia: Append Values



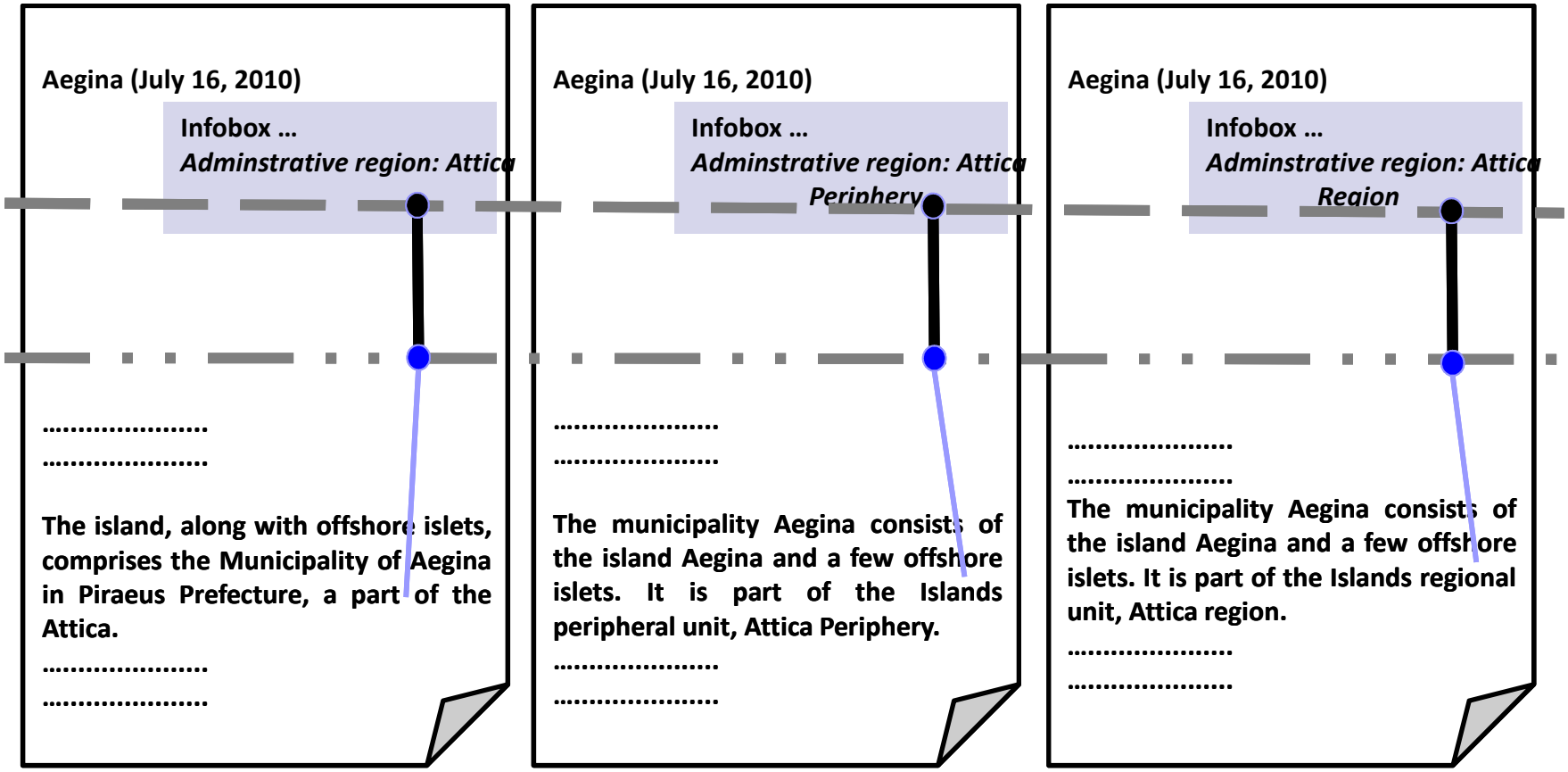
Building Ladders in Wikipedia: Replace Values



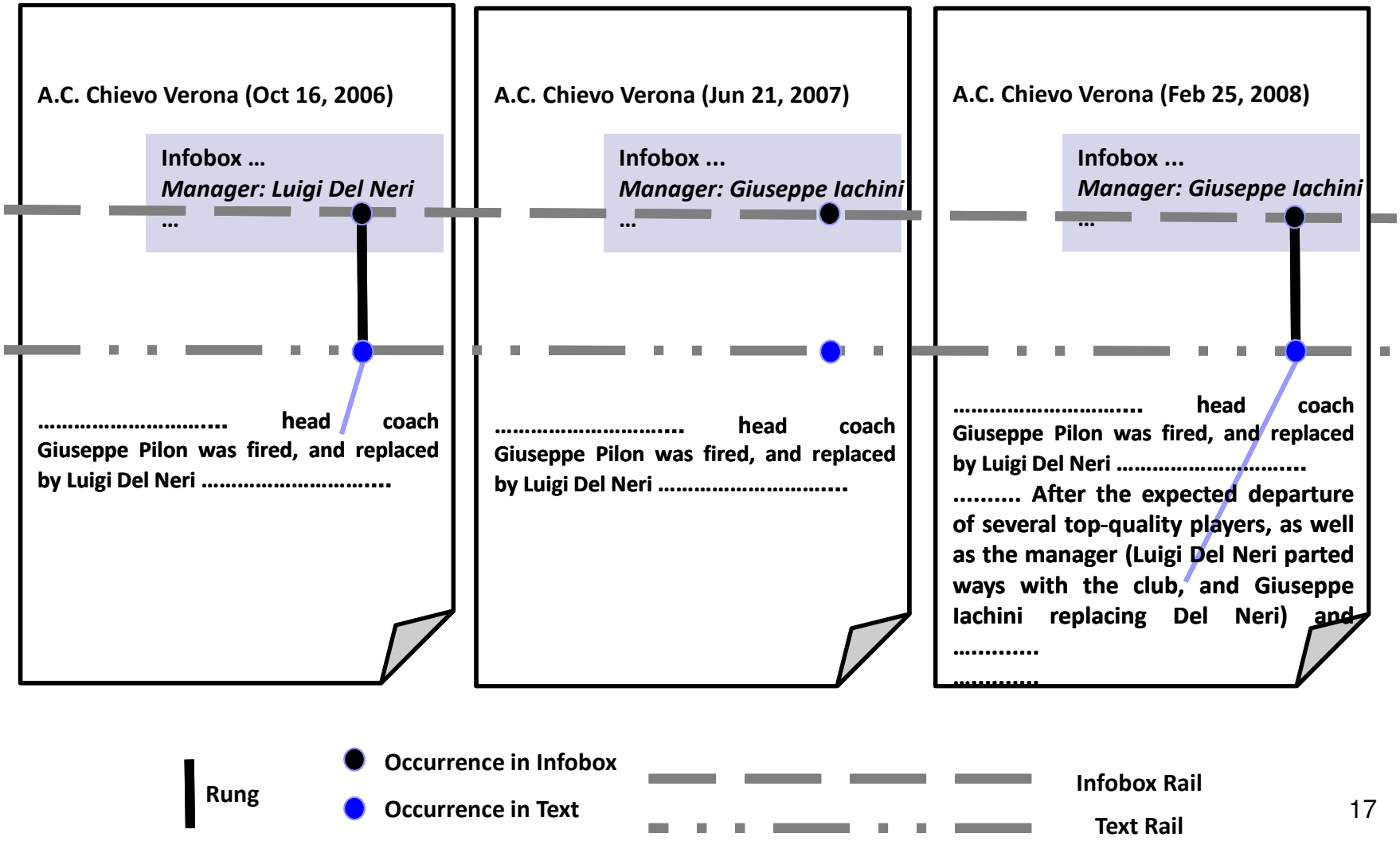
Building Ladders in Wikipedia: Vandalism



Building Ladders in Wikipedia: Controversy



Building Ladders in Wikipedia: Asynchrony



Discovering the Text Rail of a Ladder

- Goal: locate **corresponding text** across Wikipedia page revisions.

- Challenges: text edited by humans, correspondence is tenuous.
  - **Changes** to entity in a role in text and IB **may not be synchronized**.
  - Need to handle **vandalism**: cannot compare only adjacent revisions.

- Opportunities: text edited by humans, for human consumption.
  - Local **context changes slowly** when an entity in a role persists.
  - Local **context doesn't change much** even if entity in a role changes.

- Solution: start with a rung, gradually grow rail across revisions.
  - Use **decision tree to learn** the text rail, based on entity features.

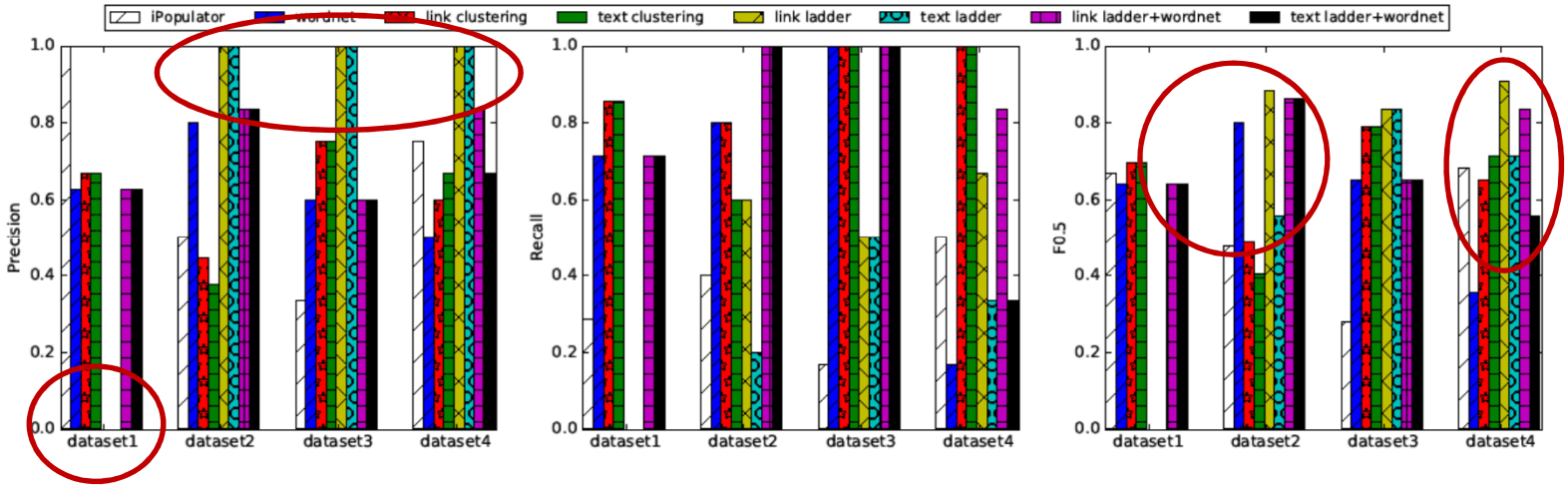
Experimental Setup

- Use real world Wikipedia datasets, crowdsourced ground truth.
  - Extracted 19,511 distinct roles from selected Wikipedia pages.
  - Partitioned based on n_rev, ambiguity, n_trans.

dataset	n_rev	ambiguity	n_trans	# of pages
dataset1	low	low	low	9,241
dataset2	high	low	low	2,708
dataset3	high	low	high	597
dataset4	high	high	high	218

- Competitive methods:
  - iPopulator: automatically populates Infoboxes using text.
  - Wordnet, link/text clustering.
  - Link/text ladder, link/text ladder + wordnet

Experimental Results: Effectiveness



- Ladder methods: high precision, inapplicable if no value changes.
  - Ladder + wordnet: successfully combines complementary ideas.

- Link-based methods are more effective than text-based methods.

Summary, Future Work

- "Tell me more" for role queries is a compelling application.

- Role linkage is a novel problem, distinct from entity linkage.
  - Proposed efficient methods to build ladders to solve this problem.
  - Use of ladder-based methods is effective in Wikipedia data.

- A lot of interesting work remains to be done.
  - Integrating related Wikipedia pages for better role linkage.
  - Identifying data quality issues: missing text in Wikipedia pages.