Open Domain Question Answering via Semantic Enrichment

Huan Sun*, Hao Ma#

Scott Wen-tau Yih#, Chen-Tse Tsai+, Jingjing Liu#, Ming-Wei Chang#

*University of California, Santa Barbara

#Microsoft Research, Redmond

+University of Illinois, Urbana-Champaign

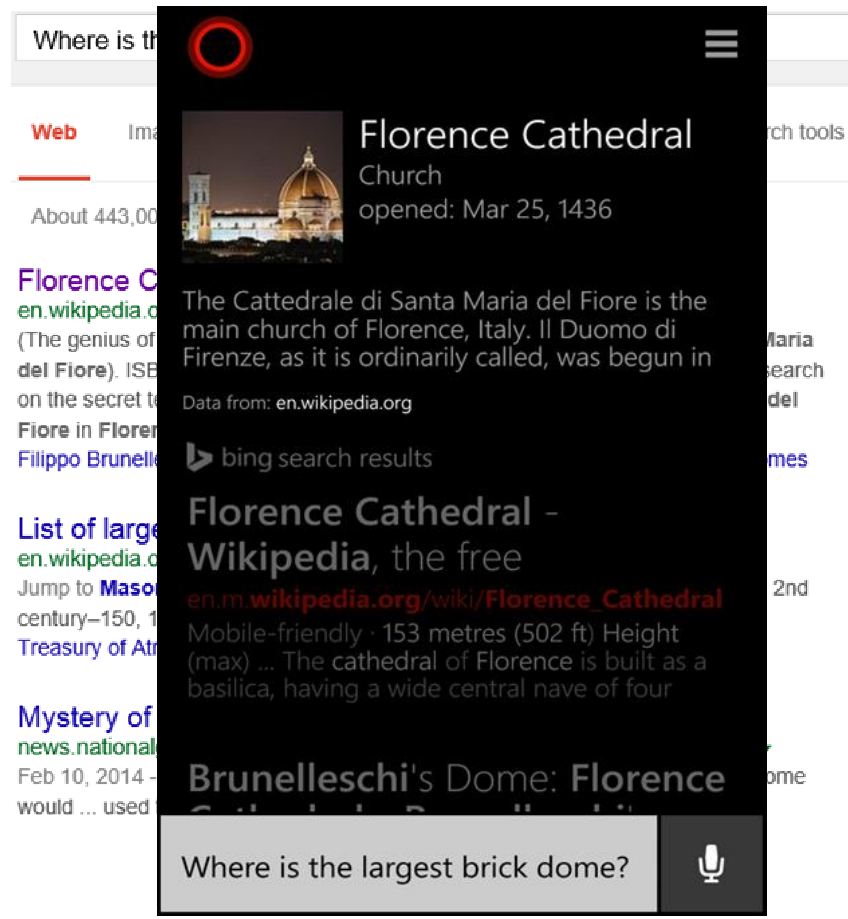
Open-domain Question Answering

Q: Where is the largest brick dome?

Answer

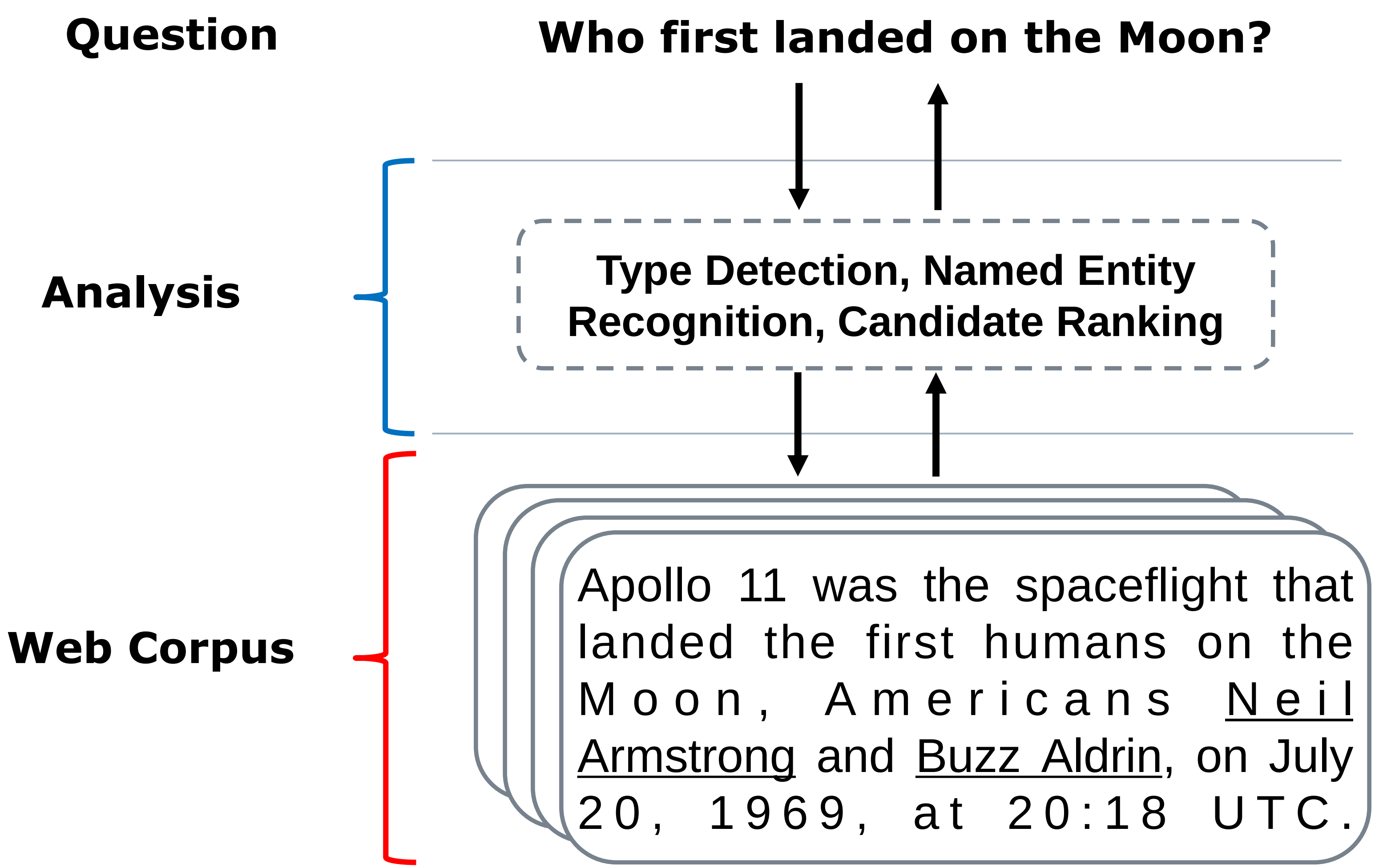
Florence Cathedral

The Cattedrale di Santa Maria del Fiore is the main church of Florence, Italy. Il Duomo di Firenze, as it is ordinarily called, was begun in 1296 in the Gothic style to the design of Arnolfo di Cambio and completed structurally in 1436 with the …
en.wikipedia.org



QA Systems via Querying the Web

[Kwok+ 2001; Brill+ 2002]



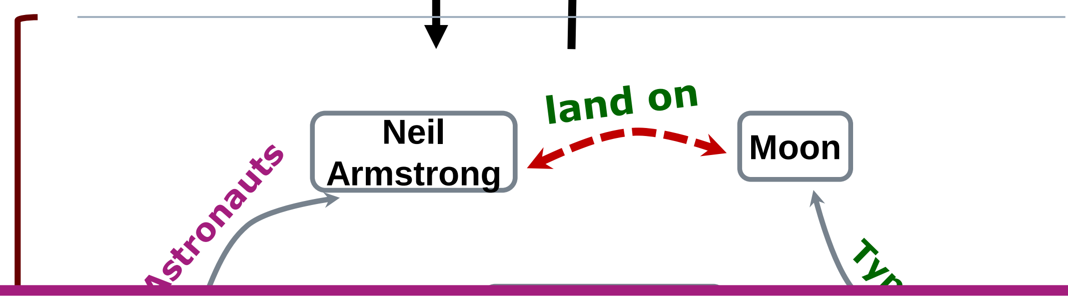
QA Systems via Querying Knowledge Bases

[Berant et al., ACL'14 & EMNLP'13]

Question **Who first landed on the Moon?**

Analysis

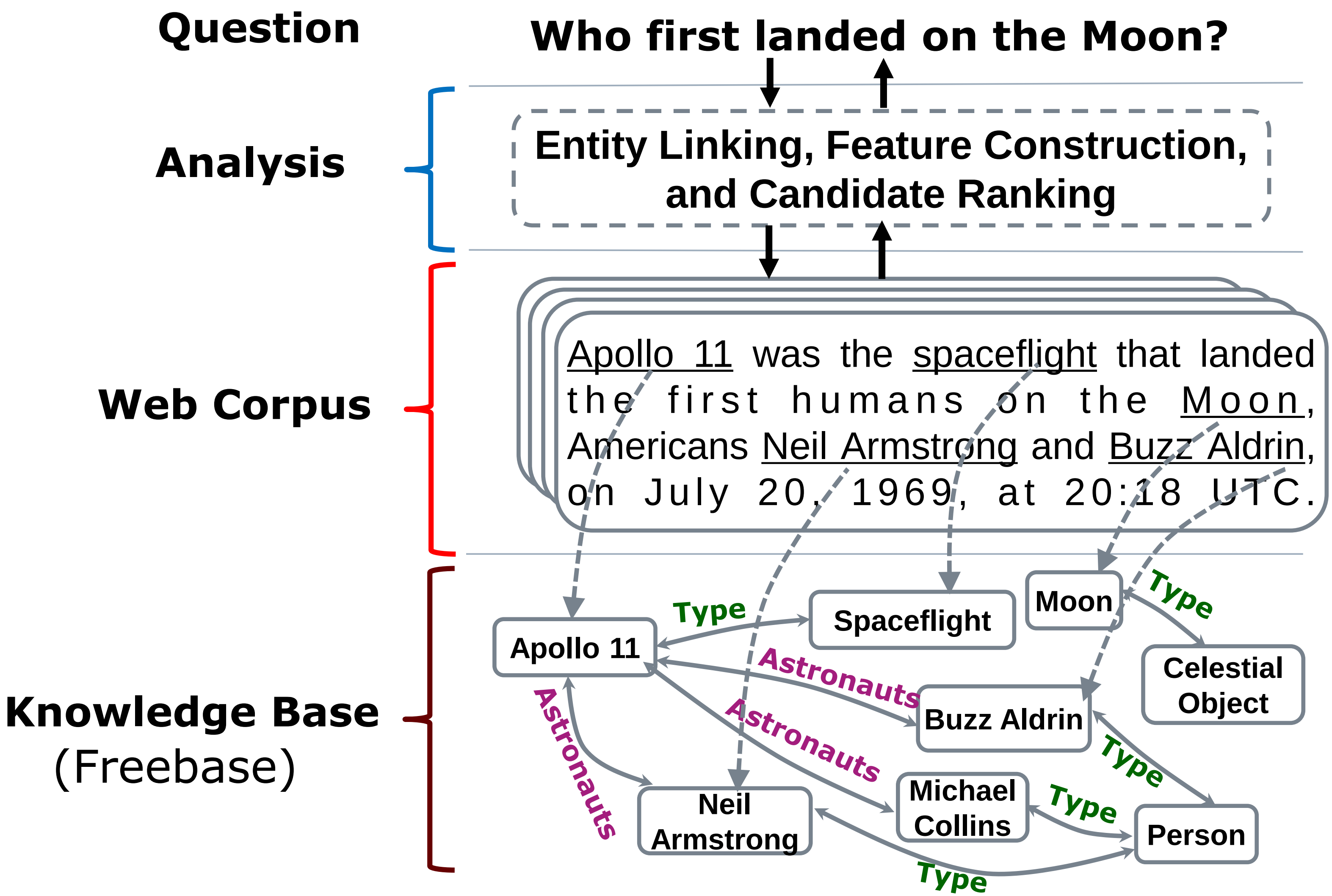
SELECT ?p
WHERE {?p *land-on* ?m . ?m *target* Moon . ?m *date* ?t .}
ORDER BY ?t LIMIT 1



Issues:

- Semantic parsing is difficult due to ontology mismatch
- Knowledge base is incomplete (missing entities/relations)

Question Answering via Semantic Enrichment



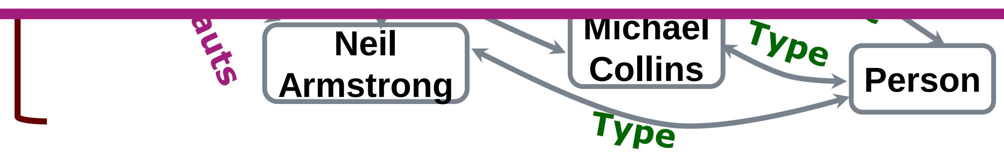
Question Answering via Semantic Enrichment

Question **Who first landed on the Moon?**

Advantages:

- Generate better answer candidates
  - Entities in Freebase
  - Mentions of the same entity merged to one candidate
- Able to leverage entity information in Freebase
  - Semantic text relevance features for ranking
  - More fine-grained answer type checking

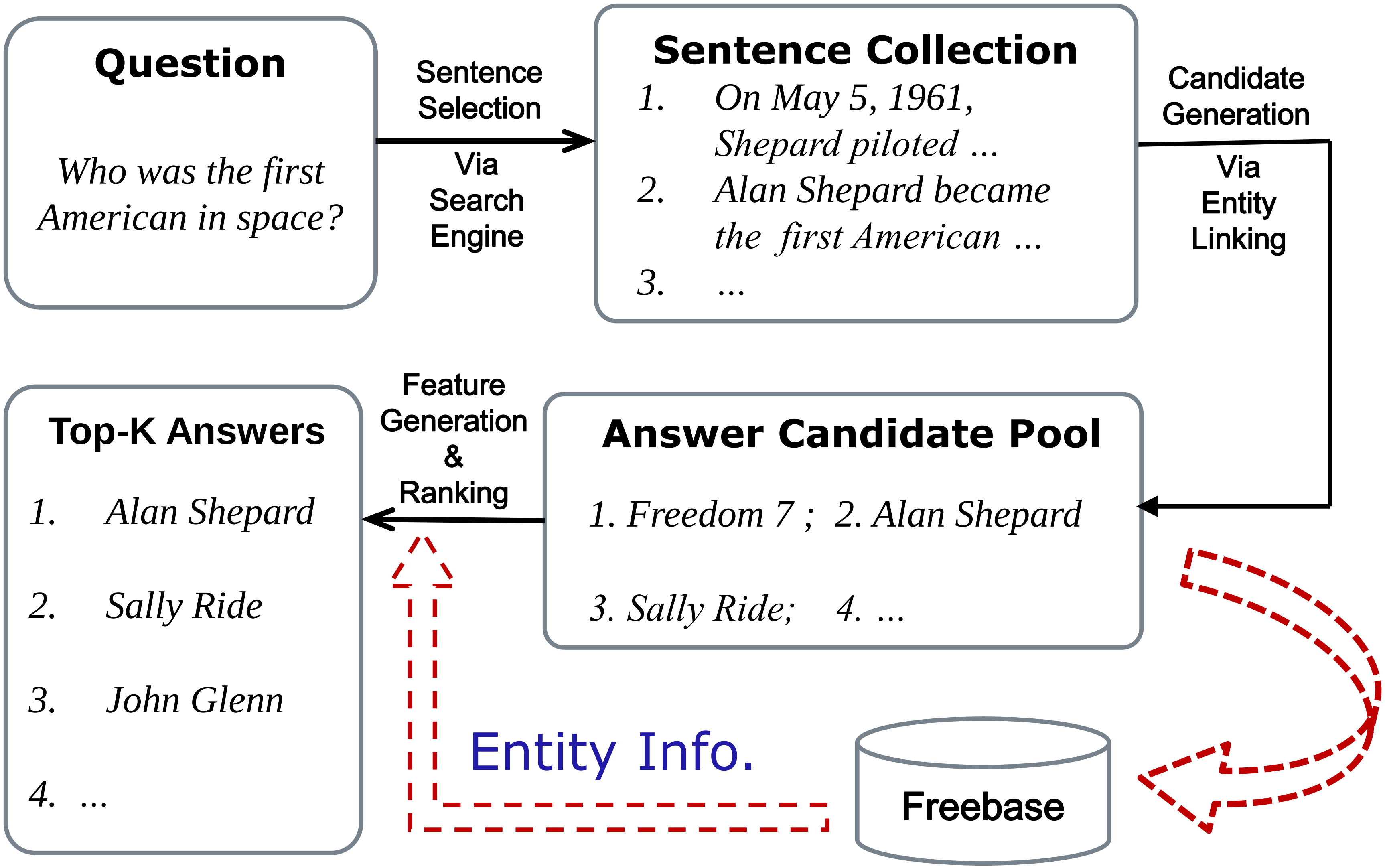
5% ~ 20% improvement in MRR



Outline

- Introduction
- System Framework
  - Identify entities as answer candidates through entity linking [Cucerzan et al., TAC'13]
  - Train an answer ranker to select the top answers
- Features enabled by KB
- Experiments
- Conclusions

System Framework



Outline

- Introduction
- System Framework
- Features enabled by KB
  - Textual Relevance (entity description)
  - Answer Type Checking (entity type)
- Experiments
- Conclusions

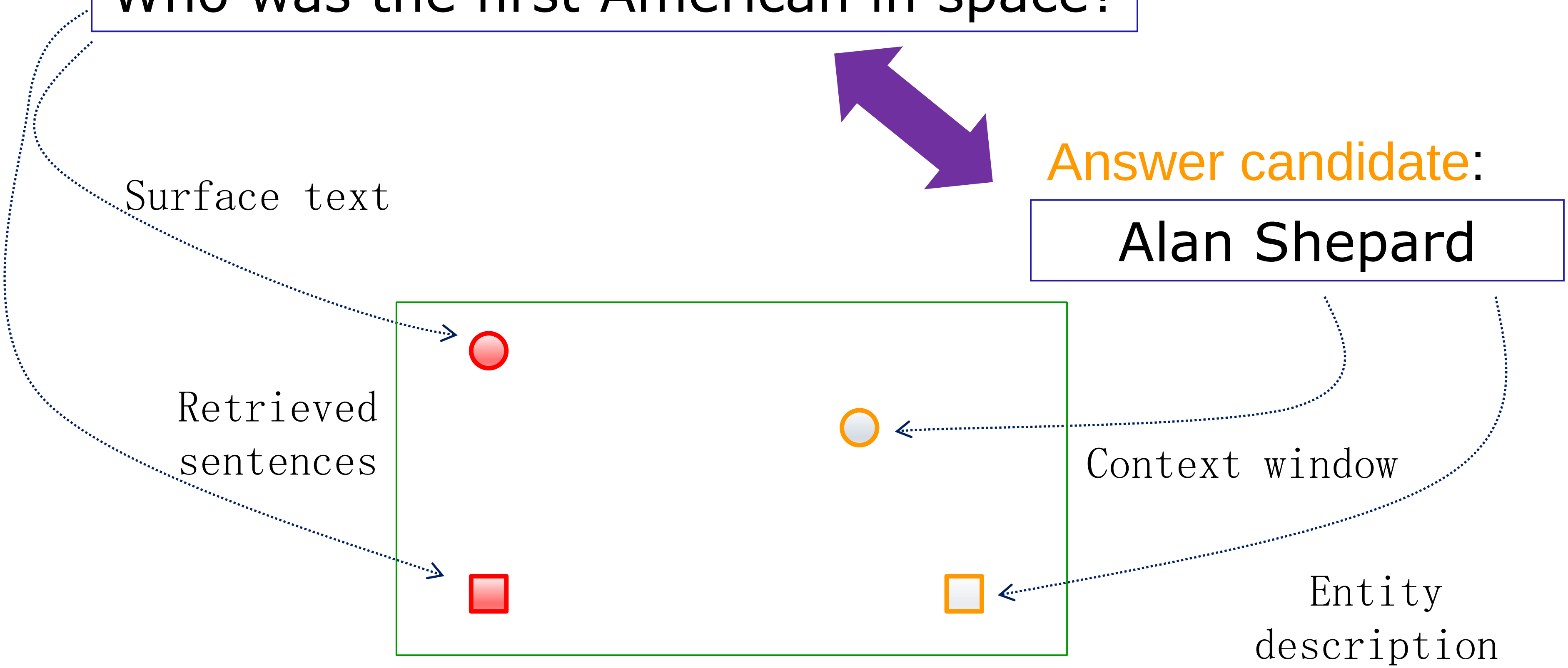
Textual Relevance between Q & A

Question:

Who was the first American in space?

Answer candidate:

Alan Shepard



Bag-of-Words Vector Space

Question Vectors

- Surface text

Who was the first American in space

1	…	1	…	1	…	1	…
who		first		american		space	

- Retrieved sentences

His 15-minute sub-orbital flight made him the first **American** in space

⋮

Alan Shepard became the first **American** in space when the **Freedom 7** spacecraft blasted off from **Florida** on May 5, 1961.

Word vector

Answer Candidate Vectors

- Context window in a retrieved sentence

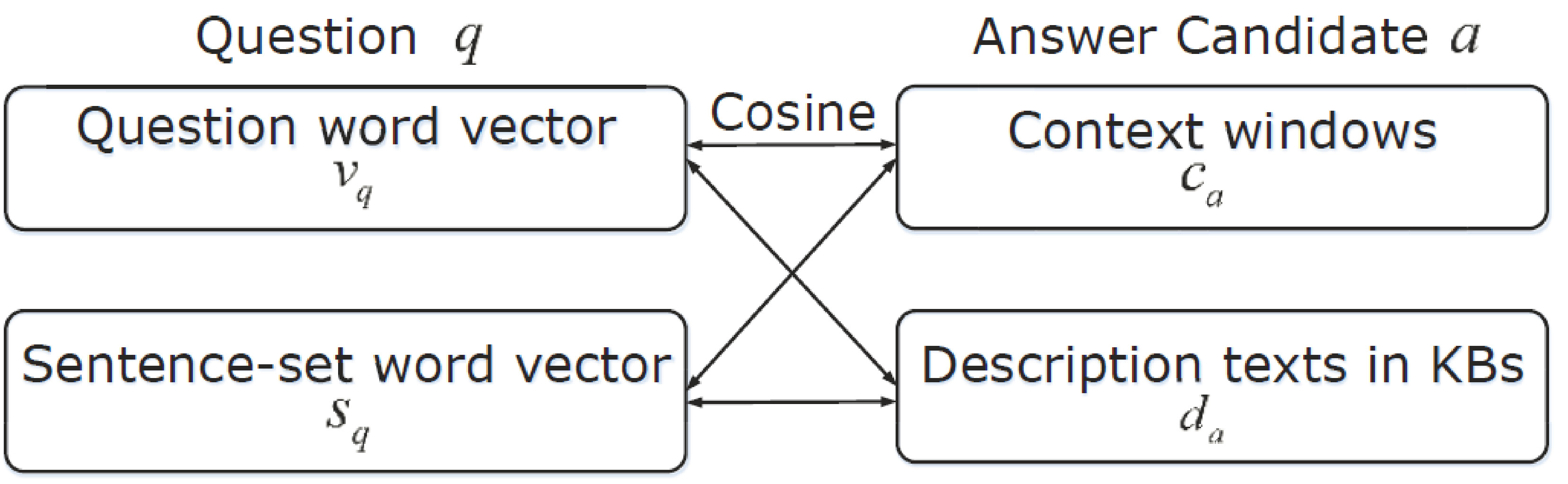
www.history.com/this-day-in-history/the-**first-american-in-space** History
From Cape Canaveral, Florida, Navy Commander **Alan Bartlett Shepard** Jr. is launched into space aboard the Freedom 7 space capsule, becoming the first American astronaut

- Description text in Freebase

Alan Shepard - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Alan_Shepard** Wikipedia
Alan Bartlett "Al" **Shepard**, Jr. (November 18, 1923 – July 21, 1998), (RADM, USN), was an American naval officer and aviator, test pilot, flag officer, one of the ...

Textual Relevance Features

- Similarity between the bag-of-words vectors of question q and answer candidate a



Outline

- Introduction
- System Framework
- Features enabled by KB
  - Textual Relevance (entity description)
  - Answer Type Checking (entity type)
- Experiments
- Conclusions

Answer Type Checking

$\boldsymbol{Q}$: Who is the first man to walk on the moon?

$$Relevance(Q, A)$$

$\boldsymbol{A}$: Apollo 11

Answer Type Checking

$\boldsymbol{Q}$: Who is the first man to walk on the moon?

$$Compatibility(Q, T)$$

$\boldsymbol{A}$: Apollo 11 $\boldsymbol{T}$: `spaceflight.space_mission`

Traditional Approach: Question Classification

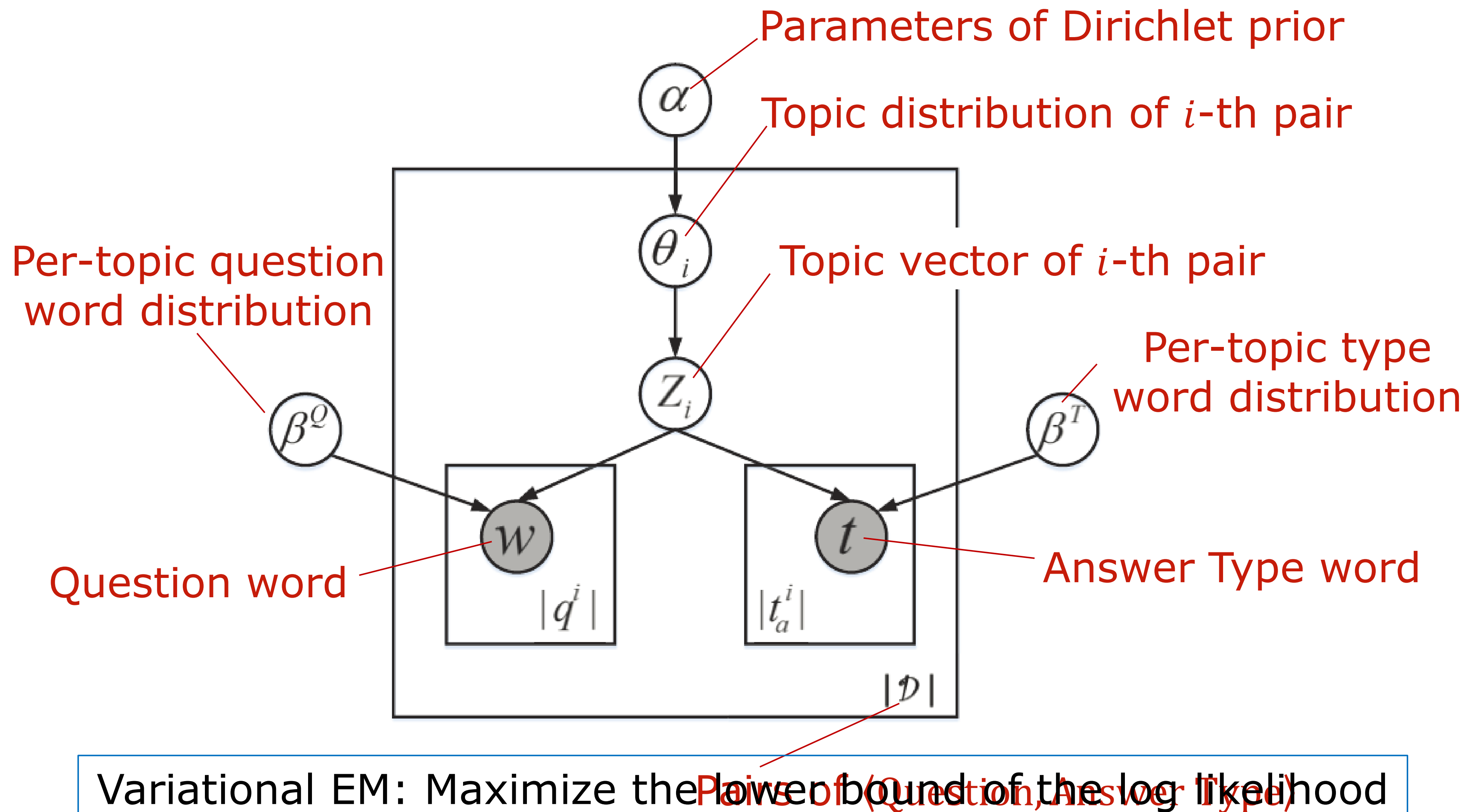
- A question is classified to a target answer type according to a predefined taxonomy.
  - e.g., `animal`, `currency`, `city`, `country`, etc.
  - Classifier trained on several thousands of labeled questions

- The number of classes is typically very small (e.g., 50 classes in [Li&Roth '02])
  - Difficult to scale to thousands of entity types in Freebase
  - Difficult to build a mapping from the coarse classes to fine-grained Freebase entity types

Joint $\langle$Question, Answer Type$\rangle$ Association

- Given pairs of question and correct answer entity
  - q = "Who is the first man to walk on the moon?"
  - e = "Neil Armstrong"
- Estimate the joint probability of observing a pair of question and entity type
  - q = "Who is the first man to walk on the moon?"
  - t = "`spaceflight.astronaut`"
- Surrogate data: click-through query logs
  - Queries that link to entity pages (e.g., Wikipedia)
  - 1.3 million pairs of question and entity type (q, t)

Joint ⟨Question, Answer Type⟩ Topic Model



Variational EM: Maximize the lower bound of the log likelihood

Outline

- Introduction
- System Framework
- Features enabled by KB
- Experiments
  - Data, Systems, Evaluation Metrics
  - Main Results & Feature Ablation Study
- Conclusions

Experiments – Data

- TREC Datasets (well-formed questions)
  - Training: 1,700 (entity) questions (TREC 8-11)
  - Testing: 202 (entity) questions (TREC 12)

Example questions:
1. What are pennies made of?
2. What is the tallest building in Japan?
3. Who sang "Tennessee Waltz"?

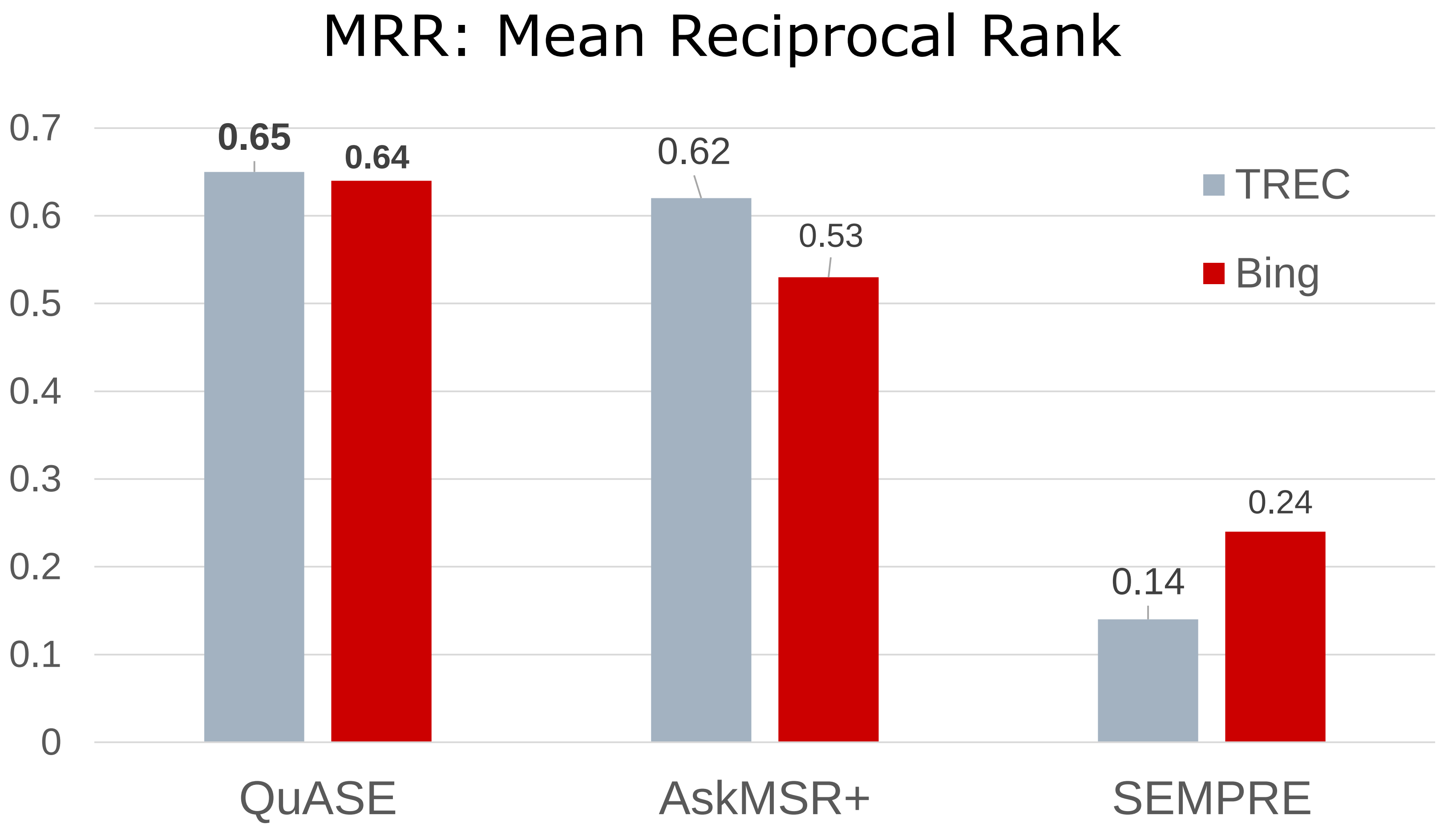
- Bing Queries (queries with question intent)
  - Training: 4,725 queries; Testing: 1,164 queries

Example queries:
1. the highest flying bird
2. indiana jones named after
3. designer of the golden gate bridge

Systems & Evaluation Metrics

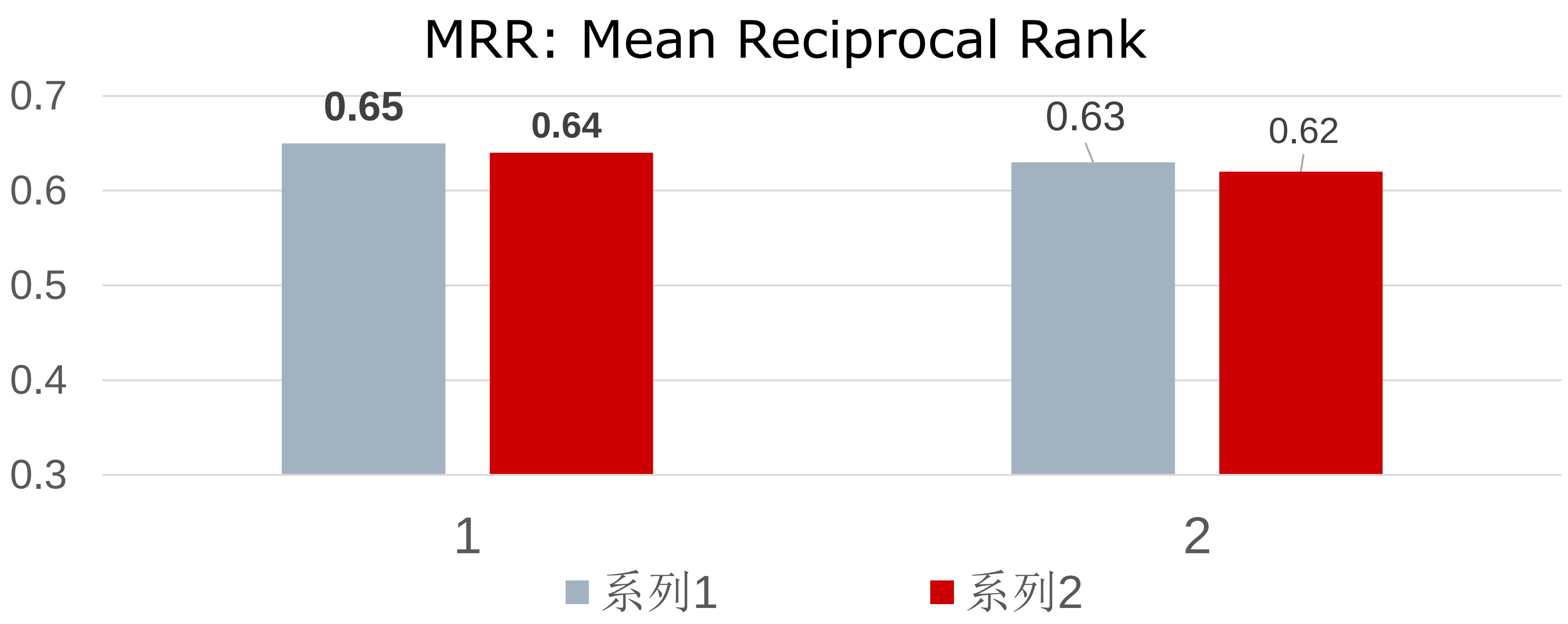
- QuASE (**Qu**estion **A**nswering via **S**emantic **E**nrichment)
  - Includes other basic features (e.g., candidate freq.)
  - Ranker learner: MART (Multiple Additive Regression Trees)
- Baselines
  - AskMSR+ [Tsai+ ‘15] – Web-based QA system
  - SEMPRE [Berant+ ‘14] – Semantic parsing QA using Freebase
- Evaluation Metrics
  - MRR: Mean Reciprocal Rank
    - Determined by the top-ranked correct answer
  - Precision/Recall/F1 (Not presented here)

Experiments – Results



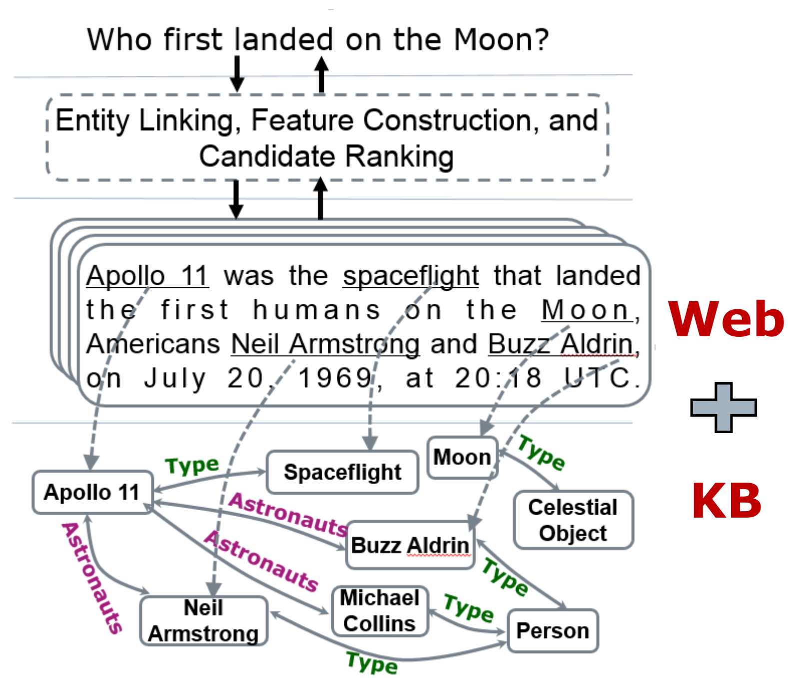
Experiments – Feature Ablation Study

- Remove KB-related features
  - Textual relevance features using entity description
  - Joint $\langle \text{Question}, \text{Answer Type} \rangle$ Association
- Answer candidate set is still from KB



Conclusions

- Question Answering via Semantic Enrichment
  - Augment Web corpus with KB information
  - Detect answer candidate via entity linking
  - Leverage KB features to improve answer ranking
  - Outperform Web-only & KB-only QA systems



- Future Work
  - Incorporate more relational information between entities (e.g., paths in KB graph)