Mapping Factoid Adjective Constraints to Existential Restrictions over Knowledge Bases

Jiwei Ding, Wei Hu*, Qixin Xu and Yuzhong Qu*

Nanjing University, China

2019.10.30

In ISWC 19

Introduction - Background

Question answering (QA) systems over knowledge bases (KBs) transform natural language questions to structural queries (e.g., SPARQL, description logics).

Entity linking: map proper nouns to entities in given KBs;

Relation mapping: map relational phrases to relations (or relation chains) in given KBs;

Introduction - Background

Question answering (QA) systems over knowledge bases (KBs) transform natural language questions to structural queries (e.g., SPARQL, description logics).

Entity linking: map proper nouns to entities in given KBs;

Relation mapping: map relational phrases to relations (or relation chains) in given KBs;

How to find appropriate representations for adjectives?

Introduction - Example

Natural language question:

List all American actors who are alive.

Description logics:

$$\text{Actor} \sqcap \exists nationality.\{United\_States\} \sqcap \neg\exists deathDate.\top$$

Introduction - Example

Natural language question:

List all American actors who are alive.

Description logics:

Actor ⊓ $\exists nationality.\{United\_States\}$ ⊓ $\neg\exists deathDate.\top$

Logic form	Examples	
$\exists r.\top$	(Married people):	$dbo{:}Person \sqcap \exists dbo{:}spouse.\top$
$\exists r.\{a\}$	(Chinese cities):	$dbo{:}City \sqcap \exists dbo{:}country.\{dbr{:}China\}$
$\neg\exists r.\top$	(People who are alive):	$dbo{:}Person \sqcap \neg\exists dbo{:}deathDate.\top$
$\neg\exists r.\{a\}$	(Hot food):	$dbo{:}Food \sqcap \neg\exists dbo{:}servingTemperature.\{\text{“}Cold\text{”}\}$

In most cases, the meaning of the input adjective can be captured by the existence (or nonexistence) of a certain property or fact.

Introduction - Problem Definition

Goal: Find appropriate representations (in form of existential restrictions) for adjectives over knowledge bases.

Introduction - Problem Definition

Goal: Find appropriate representations (in form of existential restrictions) for adjectives over knowledge bases.

Each adjective may have different representations when modifying nouns from different classes, e.g.,

(American, actor) $\longrightarrow$ $\exists nationality.\{United\_States\}$

(American, city) $\longrightarrow$ $\exists country.\{United\_States\}$

and may have several appropriate representations for each class.

(alive, actor) $\longrightarrow$ $\neg\exists deathDate.\top$

(alive, actor) $\longrightarrow$ $\neg\exists deathPlace.\top$

Introduction - Current Approaches

Some QA systems handle adjectives just like proper nouns.

List all American actors who are alive.

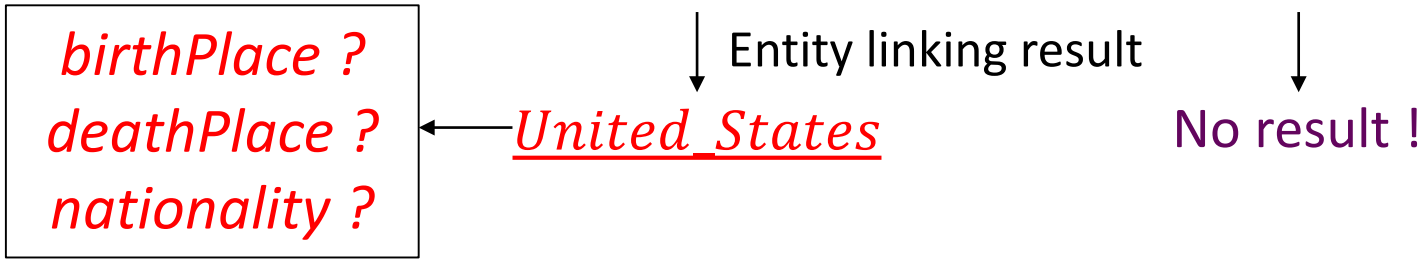
↓ Entity linking result

United_States

Introduction - Current Approaches

Some QA systems handle adjectives just like proper nouns.

List all American actors who are alive.



- Property remains unknown!
- Cannot generate representations with large lexical gap.

Introduction - Current Approaches

Some QA systems handle adjectives just like proper nouns.

List all American actors who are alive.

Entity linking result

United_States

birthPlace ?
deathPlace ?
nationality ?

No result !

- Property remains unknown!
- Cannot generate representations with large lexical gap.

Some QA systems use neural networks to learn the similarity between the adjective and the target representation.

- Lack of training data!
- A very large search space!

Introduction – Main Idea

Observation: Entities that embody the meaning of the input adjective may have some difference in fact distributions compared with entities that do not.

- Both “dead actors” and “living actors” have facts about birthDate.
- “dead actors” have facts about deathDate, while others do not.

Introduction – Main Idea

Observation: Entities that embody the meaning of the input adjective may have some difference in fact distributions compared with entities that do not.

- Both “dead actors” and “living actors” have facts about birthDate.
- “dead actors” have facts about deathDate, while others do not.

Our approach:

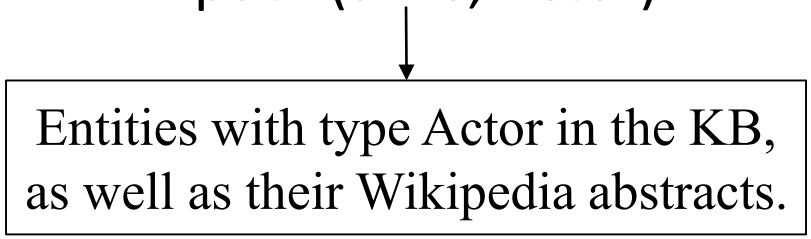
1. Collect two entity sets (one for living actors, one for dead actors);

2. Calculate fact distributions over these two entity sets, and compare the difference. Find some discriminative facts and generate existential restrictions from these facts.

Our Approach (Adj2ER)

Step 1: Collect two sets of adjective-related entities.

Input : (alive, Actor)

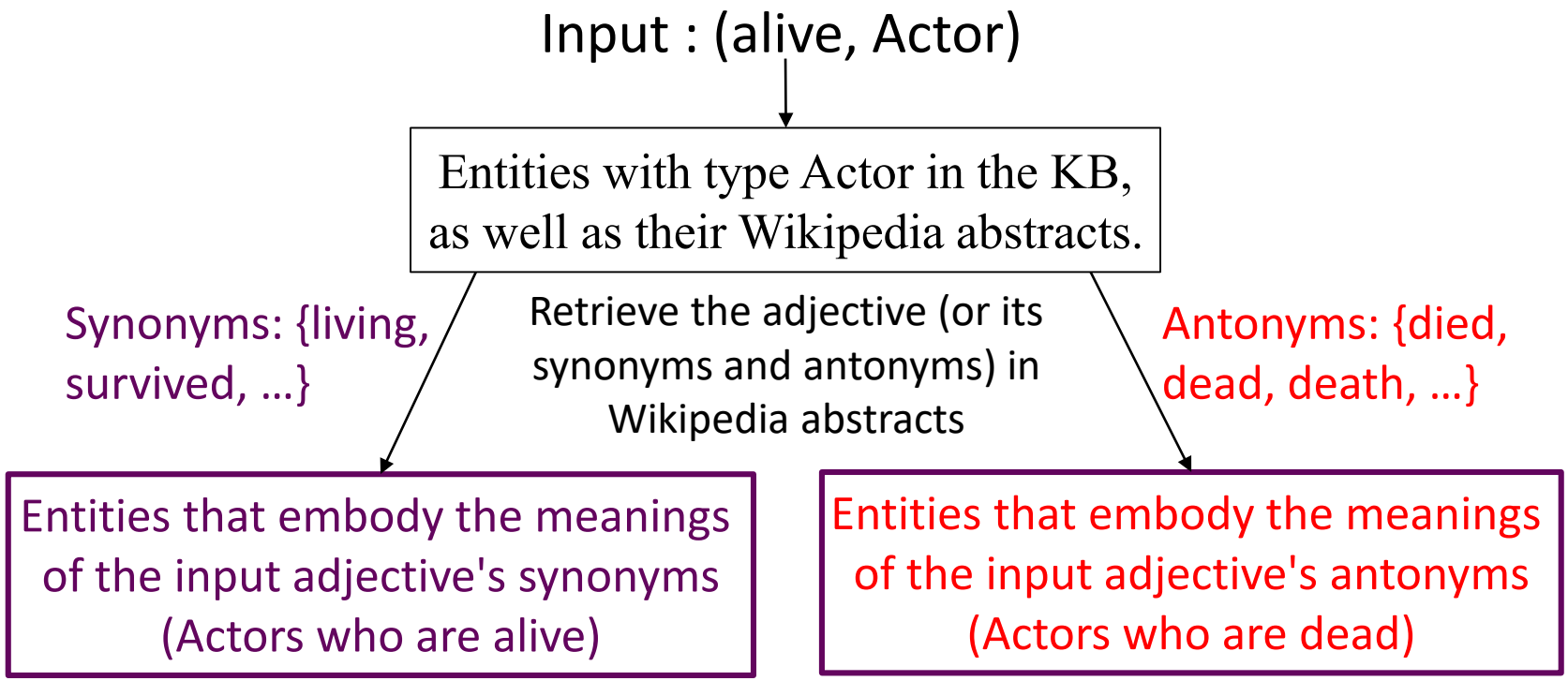


Wikipedia abstracts:

- Anthony Aston (died 1731) was an English actor and dramatist.
- Mary Porter (died 24 February 1765) was an English actress.
- Arnaud Collery (born January 1976) is a French comedian, actor and filmmaker living in Los Angeles.
- …

Our Approach (Adj2ER)

Step 1: Collect two sets of adjective-related entities.



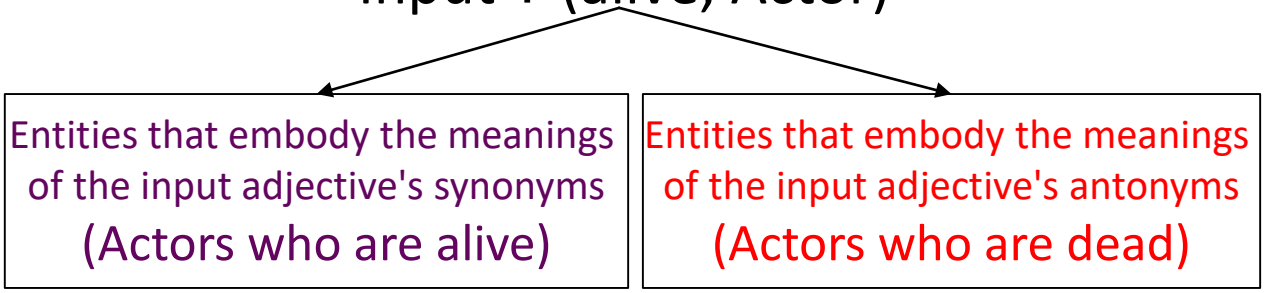
Wikipedia abstracts:

- Anthony Aston (died 1731) was an English actor and dramatist.
- Mary Porter (died 24 February 1765) was an English actress.
- Arnaud Collery (born January 1976) is a French comedian, actor and filmmaker living in Los Angeles.
- …

Our Approach (Adj2ER)

Step 2: Generate candidate existential restrictions.

Input：(alive, Actor)



- Calculate fact distributions over two entity sets.

Our Approach (Adj2ER)

Step 2: Generate candidate existential restrictions.

Input：(alive, Actor)

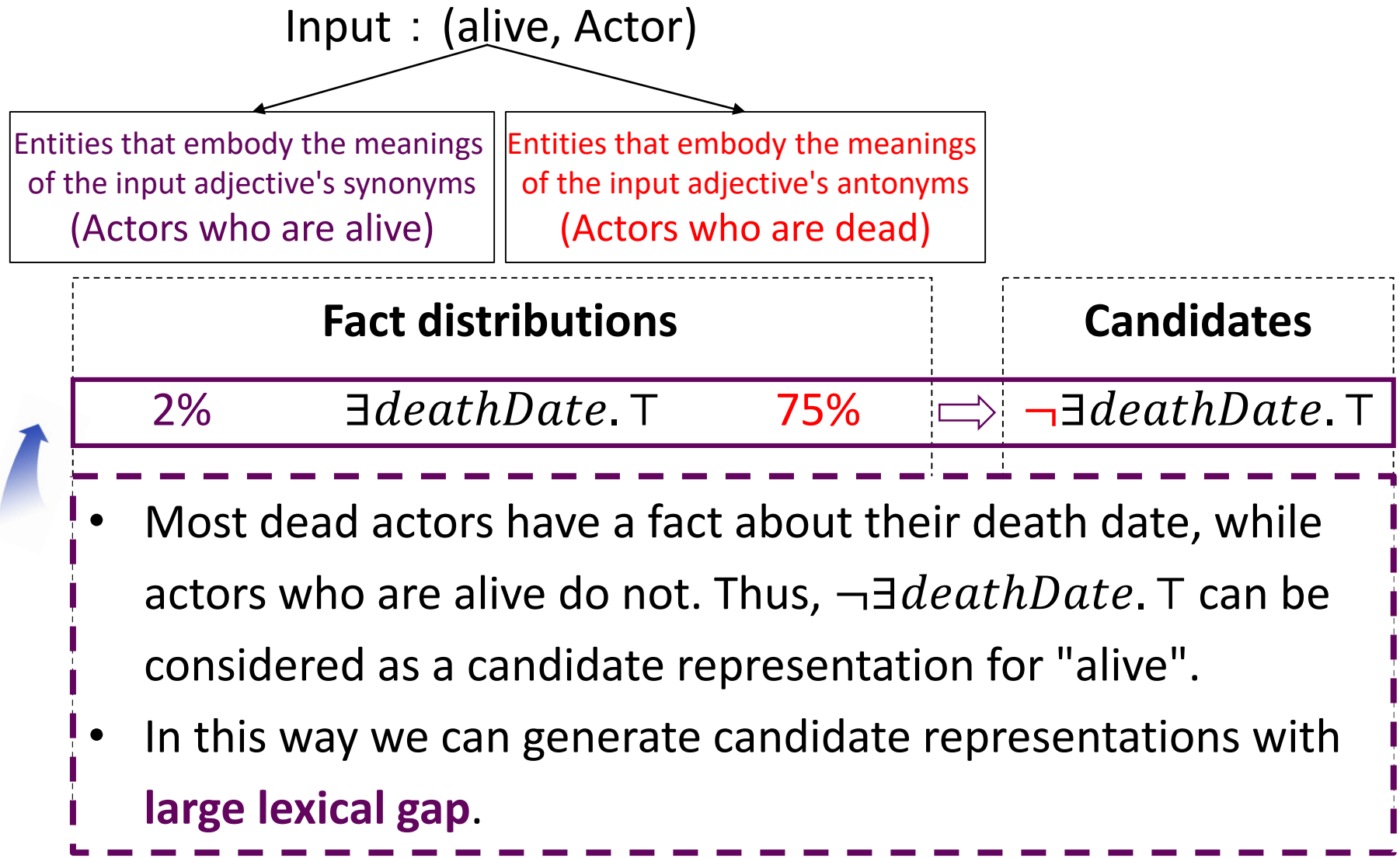
Entities that embody the meanings of the input adjective's synonyms (Actors who are alive)	Entities that embody the meanings of the input adjective's antonyms (Actors who are dead)

Fact distributions

2%	$\exists deathDate.\top$	75%
62%	$\exists birthDate.\top$	67%
5%	$\exists deathPlace.\top$	73%
62%	$\exists homepage.\top$	3%
0%	$\exists deathCause.\top$	5%

Our Approach (Adj2ER)

Step 2: Generate candidate existential restrictions.



- Most dead actors have a fact about their death date, while actors who are alive do not. Thus, $\neg\exists deathDate.\top$ can be considered as a candidate representation for "alive".
- In this way we can generate candidate representations with **large lexical gap**.

Our Approach (Adj2ER)

Step 2: Generate candidate existential restrictions.

Input ： (alive, Actor)

Entities that embody the meanings of the input adjective's synonyms (Actors who are alive)	Entities that embody the meanings of the input adjective's antonyms (Actors who are dead)

Fact distributions				Candidates
2%	$\exists deathDate.\top$	75%	⇨	$\neg\exists deathDate.\top$
62%	$\exists birthDate.\top$	67%	⇨	X

- Most dead actors and living actors have a fact about their birth date. Thus, $\exists birthDate.\top$ will not be considered as a candidate representation for "alive".
- In this way we can **reduce the search space**.

Our Approach (Adj2ER)

Step 2: Generate candidate existential restrictions.

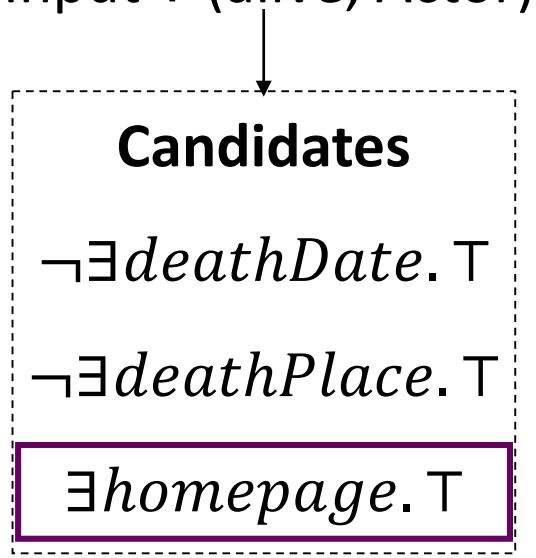
Input ： (alive, Actor)

Entities that embody the meanings of the input adjective's synonyms (Actors who are alive)	Entities that embody the meanings of the input adjective's antonyms (Actors who are dead)

Fact distributions				Candidates
2%	$\exists deathDate.\top$	75%	⇨	$\neg\exists deathDate.\top$
62%	$\exists birthDate.\top$	67%	⇨	X
5%	$\exists deathPlace.\top$	73%	⇨	$\neg\exists deathPlace.\top$
62%	$\exists homepage.\top$	3%	⇨	$\exists homepage.\top$
0%	$\exists deathCause.\top$	5%	⇨	X

Our Approach (Adj2ER)

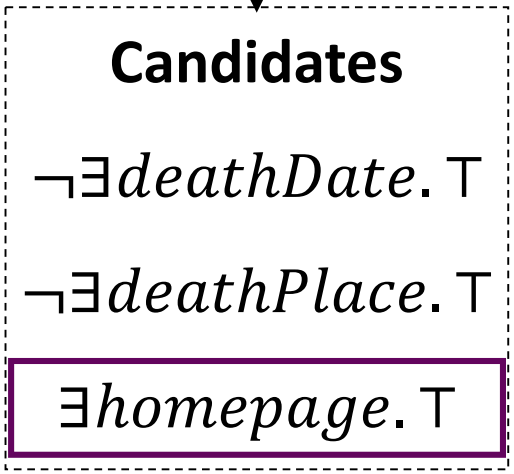
- $\exists homepage.\top$ was generated as a candidate representation for "alive" → **Each living actor should have a homepage?**



Our Approach (Adj2ER)

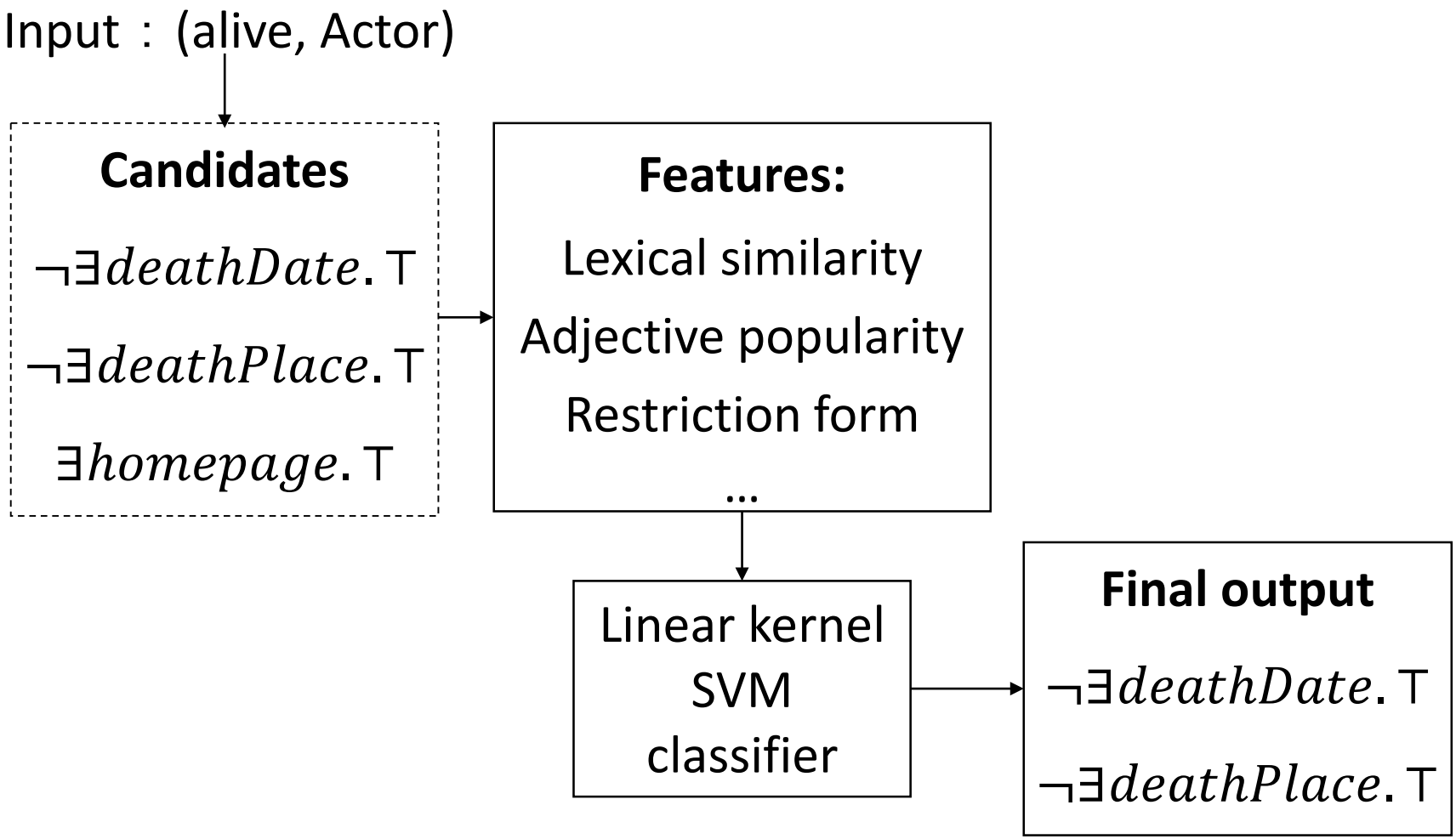
- $\exists homepage.\top$ was generated as a candidate representation for "alive" → **Each living actor should have a homepage?**
- Reason: Step 2 only considered the difference in fact distribution, but ignored the meaning of the input adjective, as well as the meaning of properties and values in the target representation.

Input：(alive, Actor)



Our Approach (Adj2ER)

Step 3: Filter out inappropriate candidate existential restrictions.



Experiments - Adjective Mapping

Goal: To verify whether Adj2ER can generate accurate mappings for most of adjectives in natural language questions.

Datasets:

QALDadj65 contains 65 (adjective, class) pairs collected from QALD 1-9;

YAadj396 contains 396 (adjective, class) pairs frequently used in Yahoo! Answers;

Reference answers are labeled by 5 Semantic Web majored students.

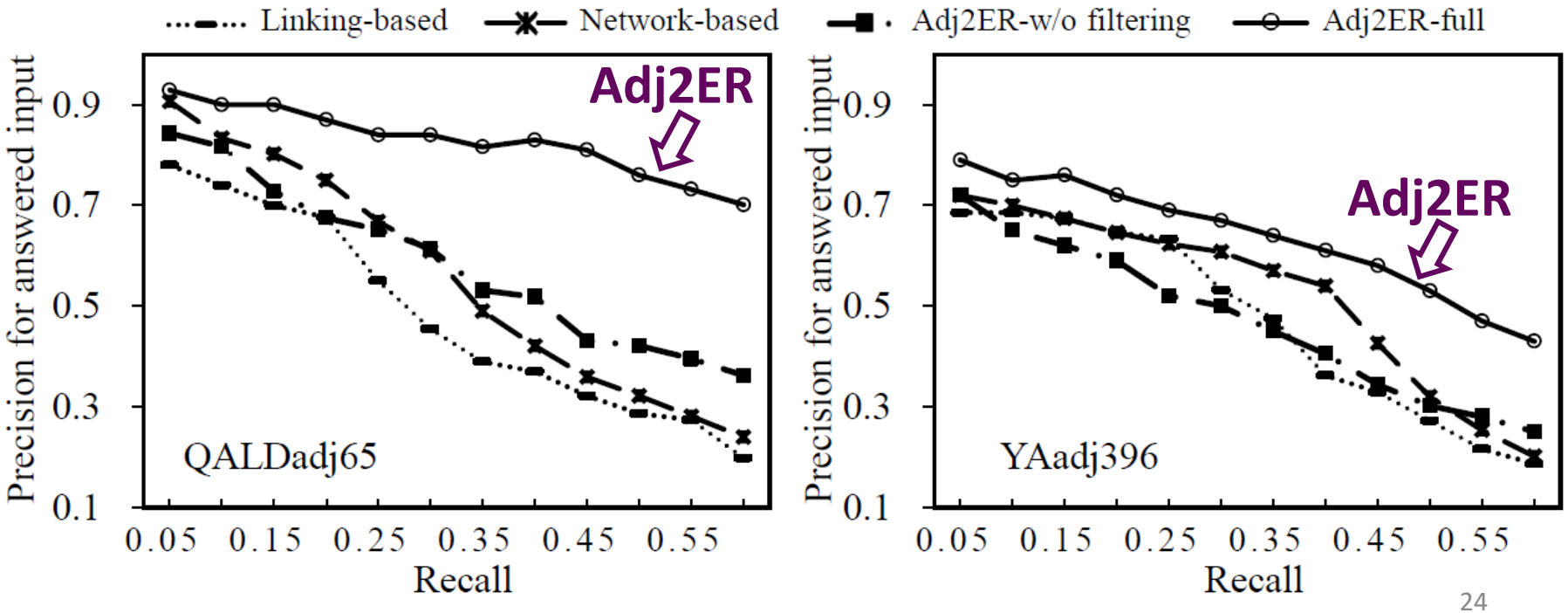
Comparative approaches:

Linking-based approach, **Network-based** approach.

Experiments - Adjective Mapping

Results for adjective mapping:

	QALDadj65			YAadj396			Time
	P	R	F1	P	R	F1	
Linking-based	31.90%	43.88%	32.66%	40.40%	34.18%	33.49%	**2.53s**
Network-based	40.26%	43.92%	36.48%	40.50%	40.54%	37.27%	89.15s
Adj2ER-w/o filtering	52.30%	36.89%	38.36%	39.98%	39.73%	36.54%	7.12s
Adj2ER-full	**71.30%**	**58.44%**	**59.65%**	**56.79%**	**46.29%**	**47.97%**	8.41s



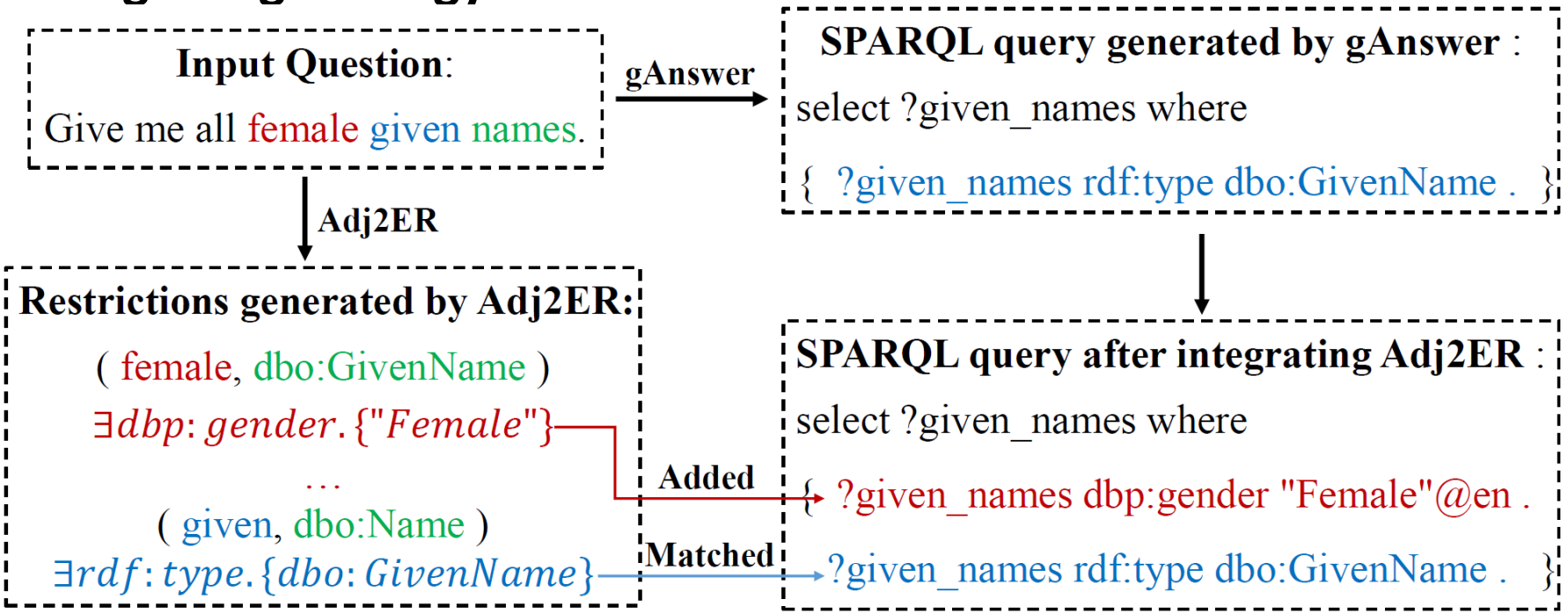
Experiments – Question Answering

Goal: To verify current QA systems (gAnswer and WDAqua) can benefit from integrating Adj2ER.

Dataset:

AdjQuestions: 70 questions from QALD 1-9, 50 questions from Yahoo! Answers;

Integrating strategy:



Experiments – Question Answering

Results for question answering:

	70 QALD questions			50 YA questions			Overall
	P	R	F1	P	R	F1	F1
gAnswer	30.49%	55.30%	29.75%	16.56%	36.26%	13.97%	23.18%
gAnswer + Adj2ER	**44.03%**	**62.25%**	**43.02%**	37.32%	**56.59%**	38.59%	**41.18%**
WDAqua	21.10%	26.64%	17.79%	23.53%	28.10%	22.04%	19.56%
WDAqua + Adj2ER	33.28%	43.86%	32.05%	**42.70%**	44.88%	**40.99%**	35.77%

- Existing QA systems gain an improvement of 16%-18% in F1 score by integrating Adj2ER.
- The accuracy for interpreting adjectives improved from 32.5% to 61.6%.

Conclusion

We proposed an approach (Adj2ER) which finds appropriate representations (in form of existential restrictions) for adjectives over knowledge bases.

List all American actors who are alive.

Actor ⊓ $\exists nationality.\{United\_States\}$ ⊓ $\neg\exists deathDate.\top$

Future work：

- Release a lexicon for adjectives representation over DBpedia.
- Apply Adj2ER to other tasks (e.g., information extraction and question generation).

Thanks for your time!

Any questions?

Contact: jiweiding@outlook.com