



Build a **Question Answering** system overnight

Jens Lehmann, Andreas Both, Ioanna Lytra, Mohnish Dubey, Denis Lukovnikov,
Kuldeep Singh, Gaurav Maheshwari, Priyansh Trivedi



ESWC 2018, Heraklion

Outline

Introduction to Question Answering

Hands on with QANARY

Introduction to DL for QA

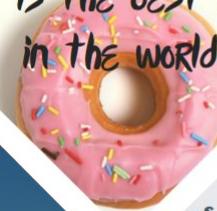
Hands on



Where is
ESWC 2018 held?

Who is the father of
Luke Skywalker?

What is the best dessert
in the world?



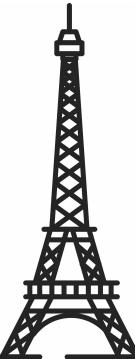
Name
one ex
president
of United States?



Introduction

All NLP/AI task can be reduced to Question Answering

Richard Socher,
Deep Learning Summit 2016, SF

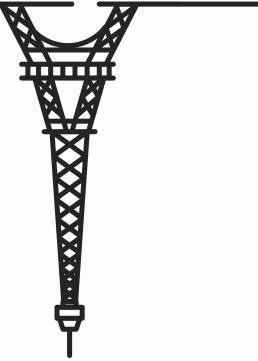


Factoid Questions

How high is the Eiffel Tower?

When was LOTR released?

Where is ESWC being held in 2018?



Non Factoid Questions

Why is the Eiffel tower in Paris?

Why is LOTR sooooo good?

Where should ESWC 2019 be held?

Factoid?

Hello, how may I help you today?

Hi. I just wanted to order a large pizza

Sure, pepperoni, a plain margherita or something else?

Pepperoni please. Also how many toppings can I get while keeping it under 10 bucks?

Domain Specific QA

Q: How do I get from Heraklion to Athens?

Q: Why is my car making this weird noise?

Q: How much did the consumer price index differ b/w greece and EU average in 2008?

... a festival called **Wianki** (Polish for **Wreaths**) have become a tradition and a yearly event in the programme of cultural events in Warsaw. The festival traces its roots to a peaceful pagan ritual where maidens would float their **wreaths** of herbs on the water to predict when they would be married, and to whom ...

Reading Comprehension QA

What is the polish word for wreaths?

Visual Question Answering

Who is wearing glasses?

man

woman



Is the umbrella upside down?

yes

no



Where is the child sitting?

fridge

arms



How many children are in the bed?

2

1



QA over Knowledge Bases

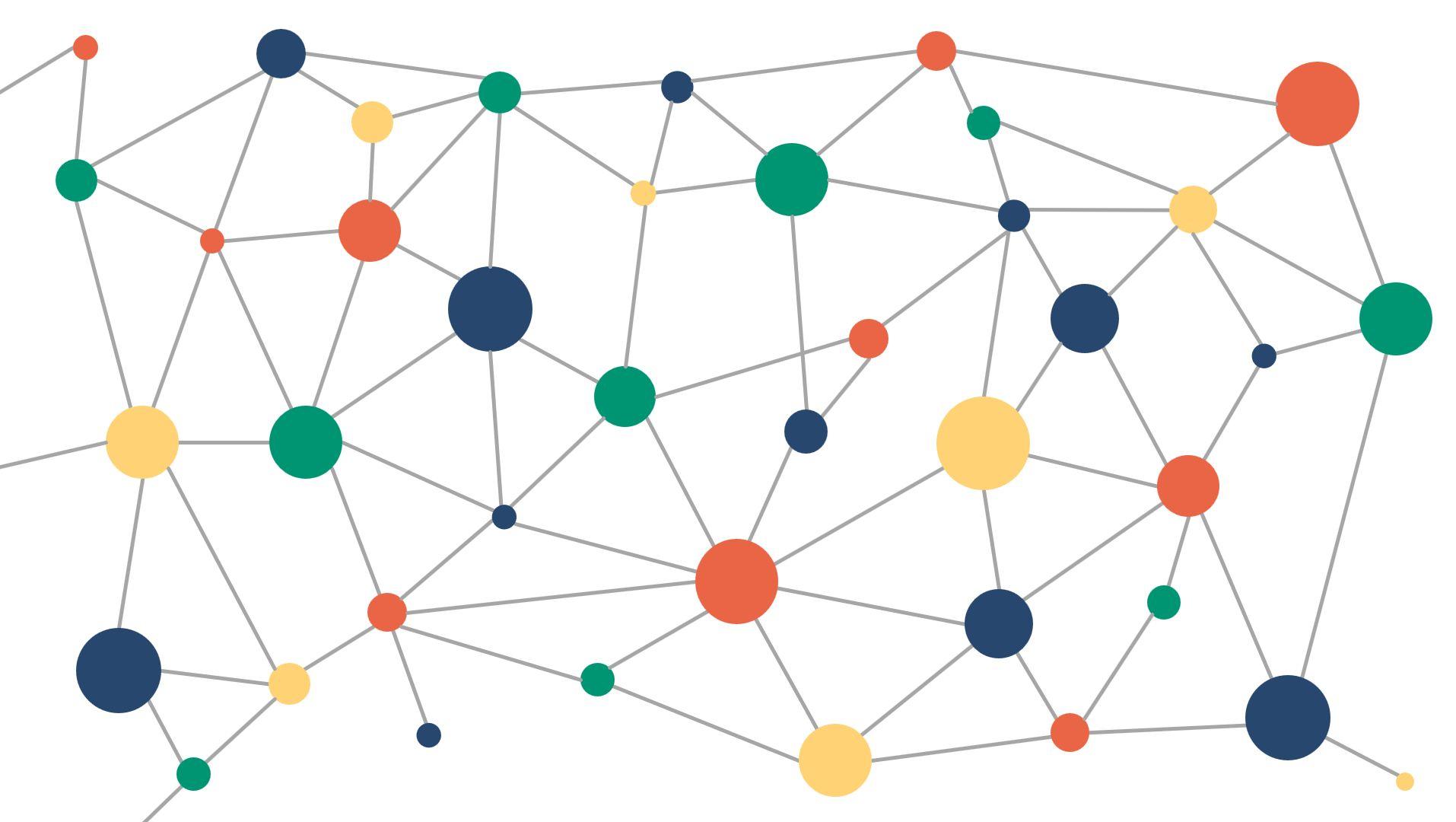
Given:

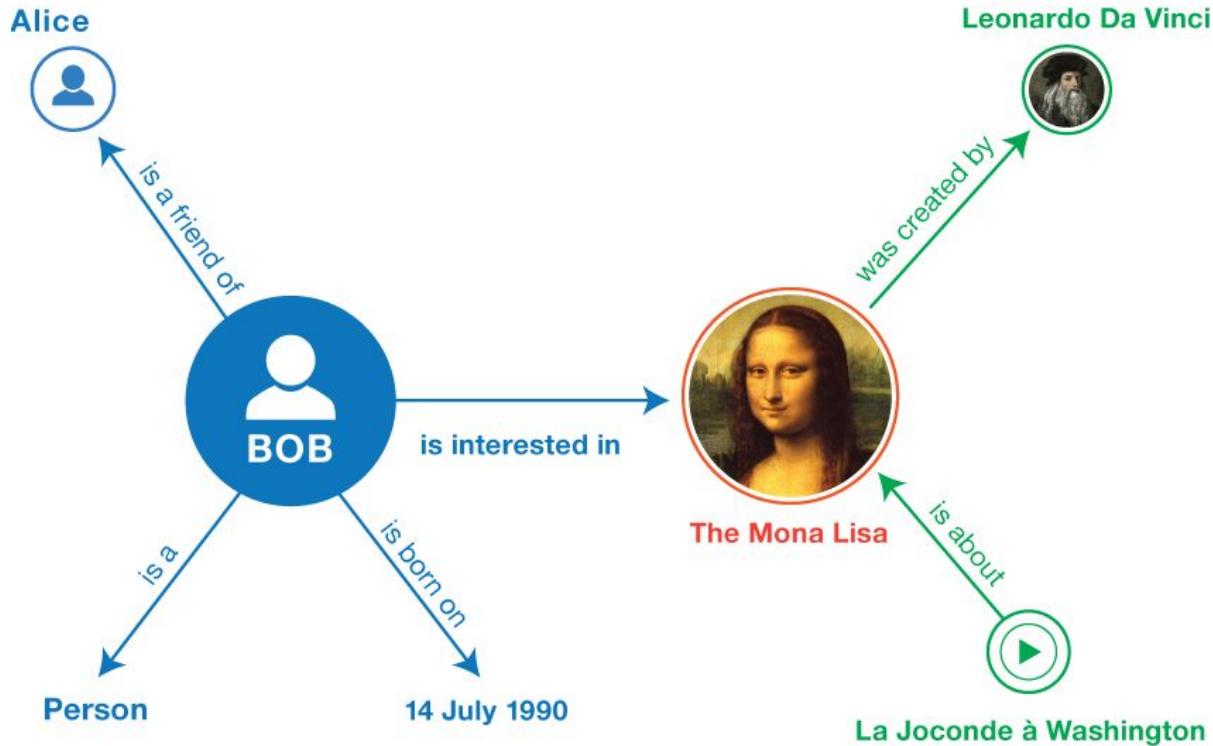
- a natural language question
- a source knowledge base

Find the subset of the *knowledge base* intended to be the answer of the question.

Knowledge Graphs

"Things instead of strings"





Images from
<https://www.w3.org/TR/rdf11-primer/>

[More images](#)

Bertrand Russell

Philosopher



Bertrand Arthur William Russell, 3rd Earl Russell, OM, FRS was a British philosopher, logician, mathematician, historian, writer, social critic, political activist and Nobel laureate. [Wikipedia](#)

Born: 18 May 1872, Trellech

Died: 2 February 1970, Penrhyneddraeth

Full name: Bertrand Arthur William Russell

Spouse: Edith Finch Russell (m. 1952–1970), more

Quotes

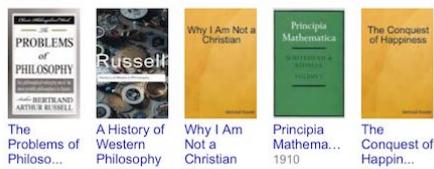
[View 7+ more](#)

To conquer fear is the beginning of wisdom.

The trouble with the world is that the stupid are cocksure and the intelligent are full of doubt.

Most people would sooner die than think; in fact, they do so.

Books

[View 45+ more](#)

People also search for

[View 15+ more](#)

QA over Knowledge Bases

"Entity" = Nodes representing things, e.g. persons, places

"Predicate" = Edges representing relations, e.g. "Works at"

Natural and efficient way to map, represent and query knowledge.

General Purpose KBs



DBpedia: <3.8B triples.



Freebase: 1.9B triples. Deprecated.



Wikidata: 47M entities.

Querying KBs with **SPARQL**

KBs are large

Billions of triples. Millions of entities.

Need formal query language to express queries.

SPARQL

W3C standard.

Formal Query Language with SQL-like grammar.

Example

How many **Marvel** **movies** was **Robert Downey Jr.**
casted in?

```
SELECT COUNT(?uri) WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbo:starring dbr:Robert_Downey_Jr  
}
```



Example

How many **Marvel** **movies** was **Robert Downey Jr.**
casted in?

```
SELECT COUNT(?uri) WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbo:starring dbr:Robert_Downey_Jr  
}
```



Example

How many **Marvel movies** was **Robert Downey Jr.**
casted in?

```
SELECT COUNT(?uri) WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbo:starring dbr:Robert_Downey_Jr  
}
```

All
marvel
movies

Every
thing
starring
RDJ

Find the
intersection

Example

How many **Marvel movies** was **Robert Downey Jr.**
casted in?

```
SELECT COUNT(?uri) WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbo:starring dbr:Robert_Downey_Jr  
}
```

All
marvel
movies

Every
thing
starring
RDJ

Find the
intersection

Count the
entities
left

Example

How many **Marvel movies** was **Robert Downey Jr.**
casted in?

```
SELECT COUNT(?uri) WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbo:starring dbr:Robert_Downey_Jr  
}
```

Example: Simple Questions

Who made Iron man?

```
SELECT ?uri WHERE {  
    dbr:Iron_Man dbp:creators ?uri.  
}
```

Example: Complex Questions

Name all **Warner Brothers** movies.

```
SELECT ?uri WHERE {  
    ?uri dbp:studio dbr:Warner_Bros .  
}
```

Example: Complex Questions

Name all Warner Brothers movies
released post 1990?

```
SELECT ?uri WHERE {  
    ?uri dbp:studio dbr:Warner_Bros .  
    ?uri dbo:releaseDate ?date .  
    FILTER (?date >= xsd:date("1990-01-01"))  
}
```

Example: Complex Questions

Name all Warner Brothers **movies**
released post 1990?

```
SELECT ?uri WHERE {  
    ?uri dbp:studio dbr:Warner_Bros .  
    ?uri rdf:type dbo:Film .  
    ?uri dbo:releaseDate ?date .  
    FILTER (?date >= xsd:date("1990-01-01"))  
}
```

Example: Boolean Queries

Did Robert Downey Junior act in Iron Man?

```
ASK WHERE {  
    dbr:Iron_Man_(2008_film)  
    dbo:starring  
    dbr:Robert_Downey_Jr  
}
```

QA as NL to SPARQL

QA can be seen as a task of converting natural language question to SPARQL queries.

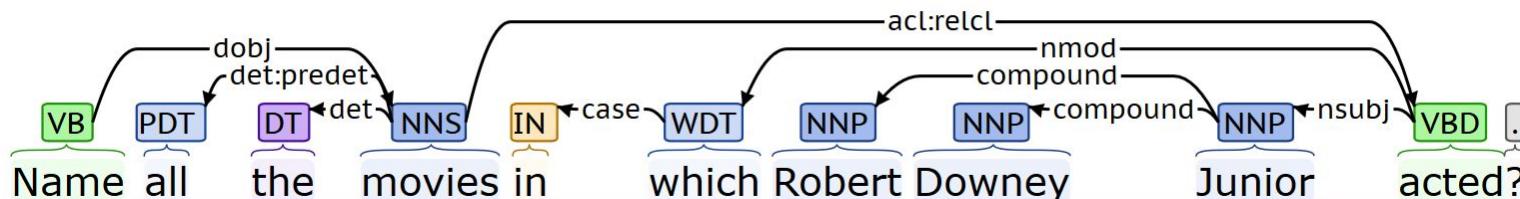
Semantic parsing?

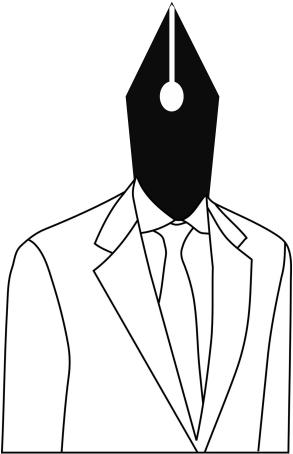
Question Answering

Semantic Parsing

Process of transforming natural language to a **formal representation** of it's meaning.

Process is generally governed by a target grammar, which can be unique for each QA system.

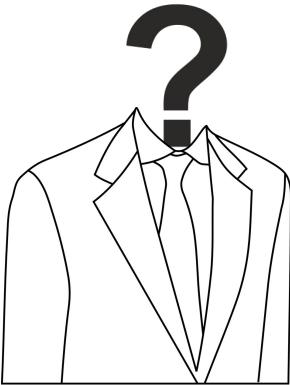




Formal Languages

Combination of well defined syntax (alphabets, symbols) and semantics (rules dictating the meaning expressed by the language)

Eg. SPARQL; Python (most programming languages);
 λ - DCS.



Formal Languages

Combination of well defined syntax (alphabets, symbols) and semantics (rules dictating the meaning expressed by the language)

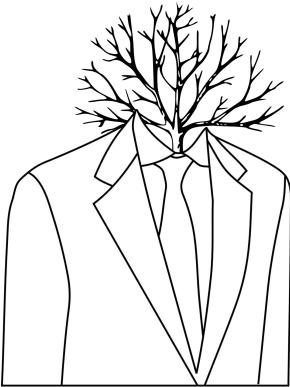
Eg. SPARQL; Python (most programming languages);
 λ - DCS; **user-defined**.



Example: SPARQL

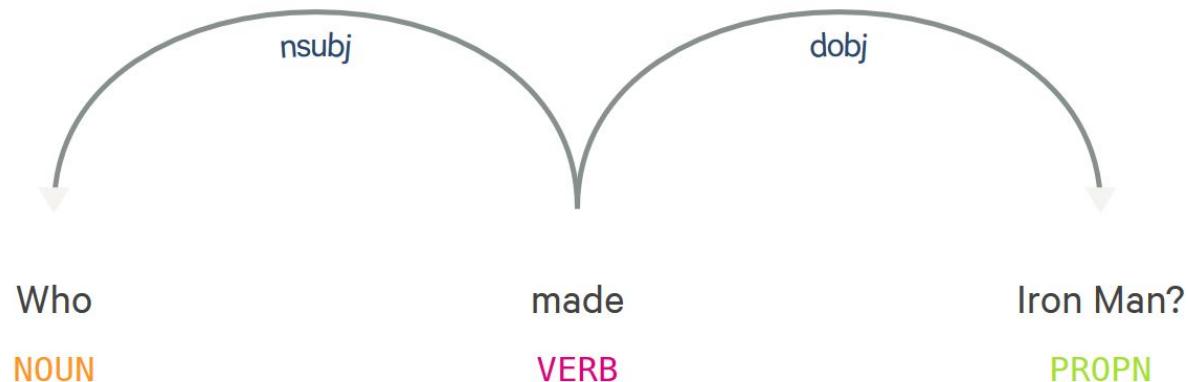
Who **made** **Iron man**?

```
SELECT ?uri WHERE {  
    dbr:Iron_Man dbp:creators ?uri.  
}
```



Example: Dependency Parse Tree

Who made Iron man?





Example: Custom

```
{  
  "x": 2,  
  "y": 2,  
  "formulae": "x+y",  
  "result": 5  
}
```

Semantic Parsing

Iron Man is a movie in which Robert Downey Jr acted.

Semantic Parsing

Iron Man is a movie in which Robert Downey Jr acted.



Semantic Parsing

Iron Man is a movie in which Robert Downey Jr acted.



Semantic Parsing

Iron Man is a movie in which **Robert Downey Jr** *acted*.



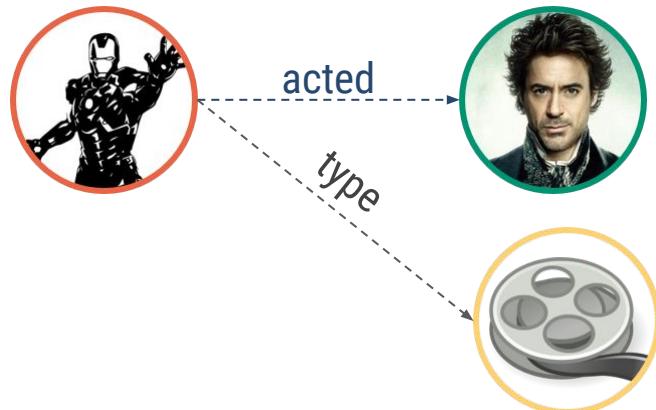
Semantic Parsing

Iron Man is a movie in which Robert Downey Jr acted.



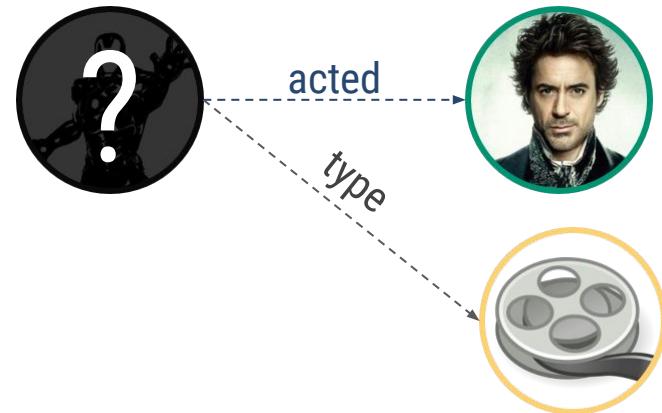
Semantic Parsing

Iron Man is a movie in which Robert Downey Jr acted.



Semantic Parsing

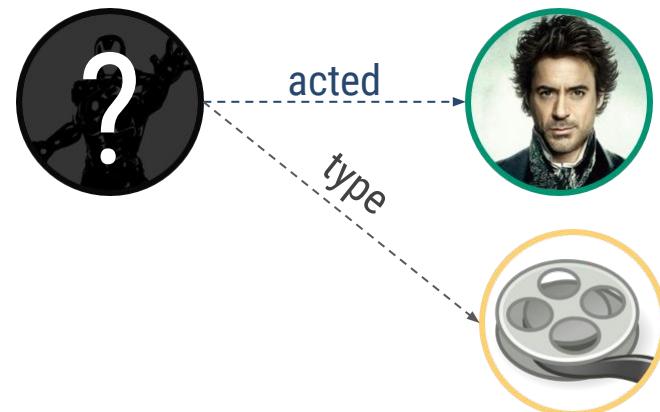
Name all the movies in which Robert Downey Jr Acted?



Semantic Parsing

Name all the movies in which Robert Downey Jr Acted?

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
    ?uri rdf:type dbo:Film .  
}
```



Natural Language is difficult to parse.

Name all the movies in which Robert Downey Jr Acted?

Which movies have RDJ?

Flicks where I can see Robert DJ?

Find me all the films casting Rober Downey Jr ?

List all the movies starring Robert Downey Junior?

RDJ has acted in which movies?

Natural Language is difficult to parse.

Name all the movies in which Robert Downey Jr acted?

Which movies have RDJ?

Flicks where I can see Robert DJ?

Find me all the films casting Rober Downey Jr ?

List all the movies starring Robert Downey Junior?

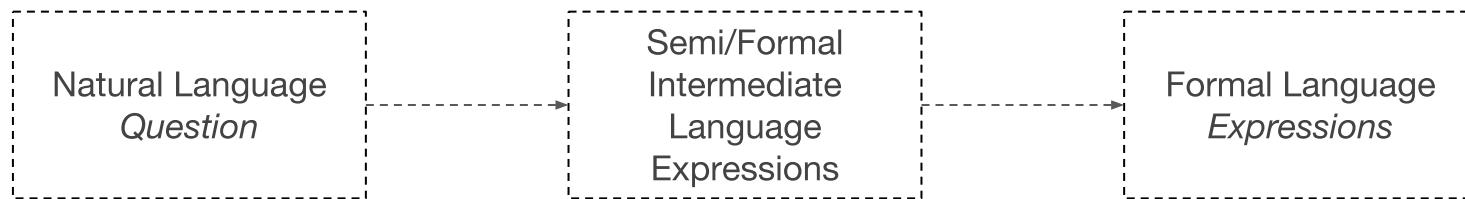
RDJ has acted in which movies?

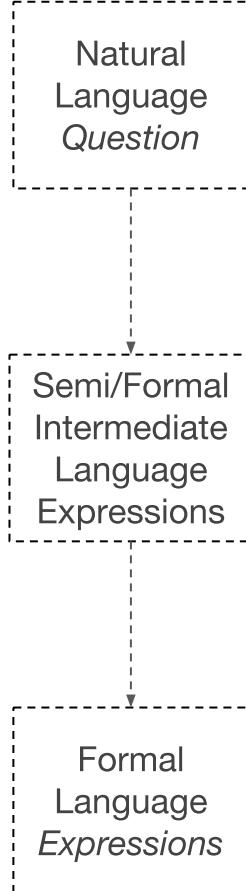
```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
}
```

* rdf:type constraint omitted for brevity's sake



Two step process



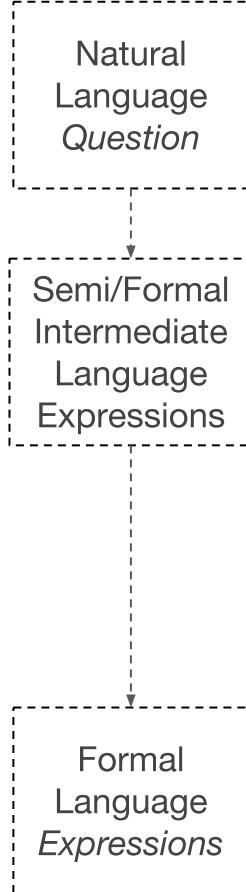


Intermediate Language Expressions

Representation of the question which eases conversion to queries.

Broadly categorized (for QA purposes) into:

- KB independent expression
- KB dependent expression



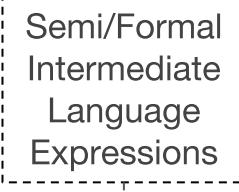
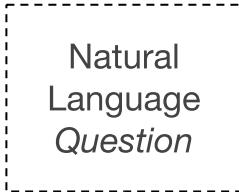
KB Independent Expressions

Represent different syntactic, lexical variations of the question as one expression

Focus: Structure of the question

Ambiguous: Structure of the KB

Eg. AskNow's [?cite] NQS, XSER [?cite], DCS expressions.



Ex: AskNow NQS

Name all the movies in which Robert Downey Jr. acted?

Query_input = Robert Downey Jr.

Query_desire = acted

Query_type = List

* rdf:type constraint omitted for brevity's sake

Natural
Language
Question

Semi/Formal
Intermediate
Language
Expressions

Formal
Language
Expressions

Parses and KB structure

The parse *ignores* the underlying KB structure



Natural
Language
Question

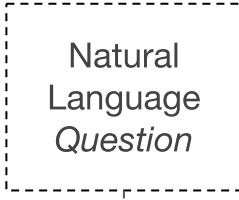
Semi/Formal
Intermediate
Language
Expressions

Formal
Language
Expressions

Parses and KB structure

The parse *ignores* the underlying KB structure





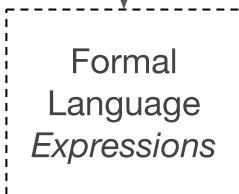
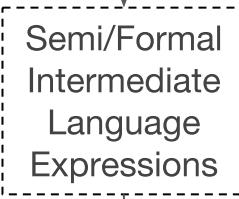
KB Dependent Expressions

Keeps the entities, predicates and structure of KB in mind while creating parses.

Focus: Structure of the KB

Ambiguous: Structure of the question

Examples - Query Graphs, SPARQL,[?cite]



Natural
Language
Question



Semi/Formal
Intermediate
Language
Expressions

Formal
Language
Expressions

Eg. Query Graphs

Name all the movies in which **Robert Downey Jr.** **acted?**

```
dbr:Robert_Downey_Jr <- dbo:starring  
dbr:Robert_Downey_Jr <- dbo:artist
```

* rdf:type constraint omitted for
brevity's sake

Natural
Language
Question

Semi/Formal
Intermediate
Language
Expressions

Formal
Language
Expressions

Parses and KB structure

The parse *ignores* the underlying KB structure



Natural
Language
Question

Semi/Formal
Intermediate
Language
Expressions

Formal
Language
Expressions

Parses and KB structure

The parse is always consistent with the underlying KB structure.



Components and Challenges

Difficult Task

Name all the movies in which Robert Downey Jr acted?

Which movies have RDJ?

Flicks where I can see Robert DJ?

Find me all the films casting Rober Downey Jr ?

List all the movies starring Robert Downey Junior?

RDJ has acted in which movies?

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
}
```

* rdf:type constraint omitted for brevity's sake

Difficult Task

Correctly understand the intent of the question.

Juxtapose this understanding with the KB schema.

Entity Linking

Identify all the spans which refer to an entity in a given KB.

Name all the movies in which **Robert Downey Jr** acted?

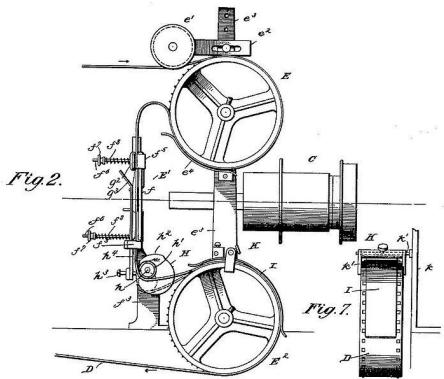


dbr:Robert_Downey_Jr

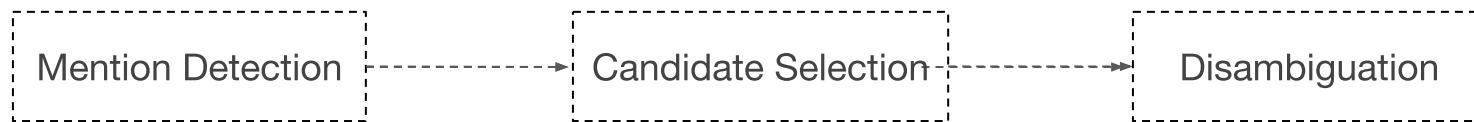
Entity Linking



- Name all the movies in which **Robert Downey Jr** acted?
- Which movies have **RDJ**?
- Flicks where I can see **Robert DJ**?
- Find me all the films casting **Rober Downey Jr** ?
- List all the movies starring **Robert Downey Junior**?
- RDJ** has acted in which movies?



Anatomy of a Entity Linking System



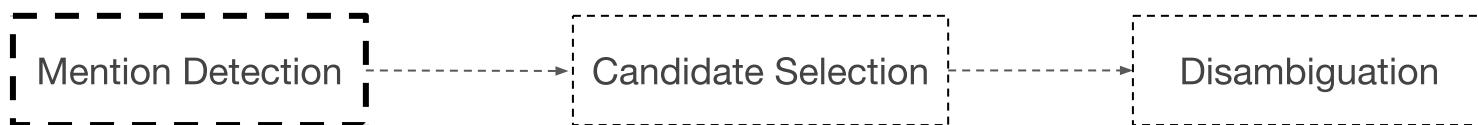
Anatomy

Robert John Downey Jr. is an American actor and singer. Downey was born April 4, 1965 in Manhattan, New York, the son of writer, director and filmographer Robert Downey Sr. and actress Elsie Downey . Beginning in 2008, Downey began portraying the role of Marvel Comics superhero Iron Man in the Marvel Cinematic Universe, appearing in several films as either the lead role

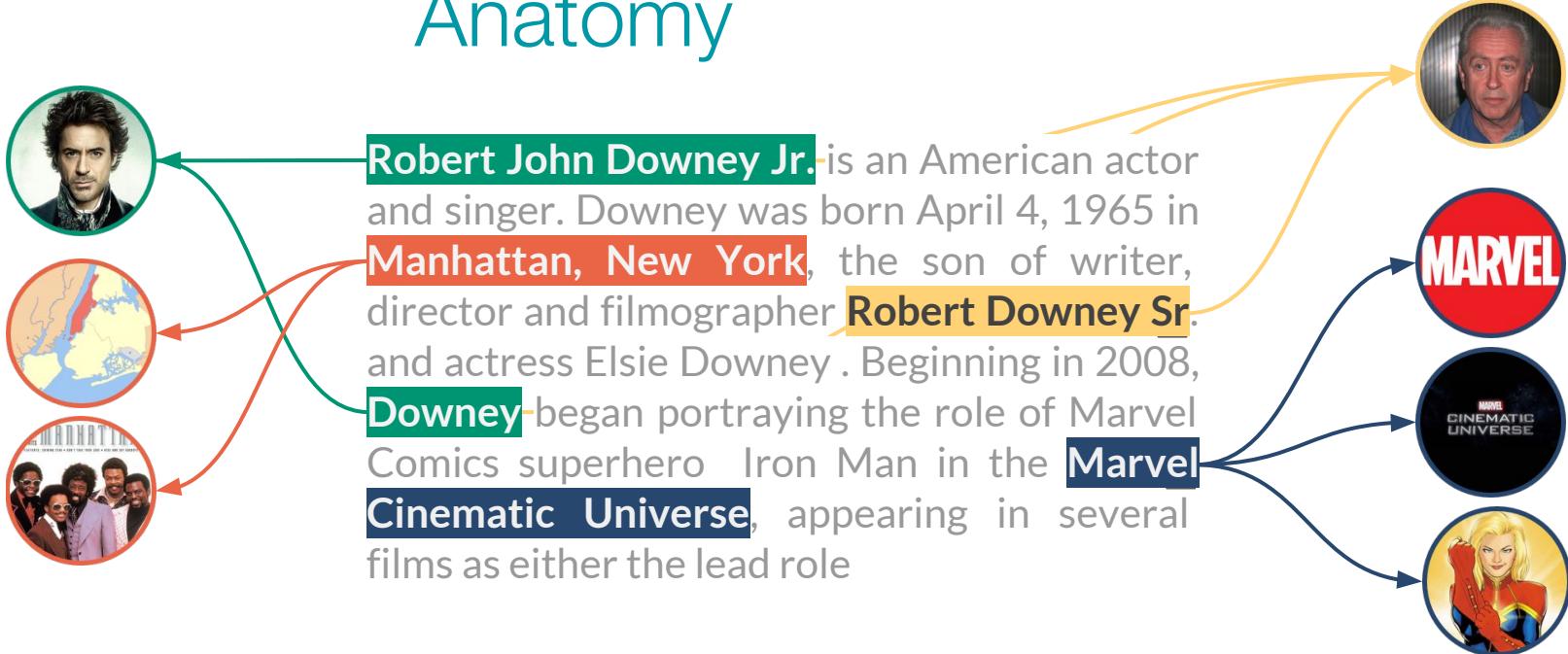


Anatomy

Robert John Downey Jr. is an American actor and singer. Downey was born April 4, 1965 in **Manhattan, New York**, the son of writer, director and filmographer **Robert Downey Sr.** and actress Elsie Downey . Beginning in 2008, **Downey** began portraying the role of Marvel Comics superhero Iron Man in the **Marvel Cinematic Universe**, appearing in several films as either the lead role



Anatomy

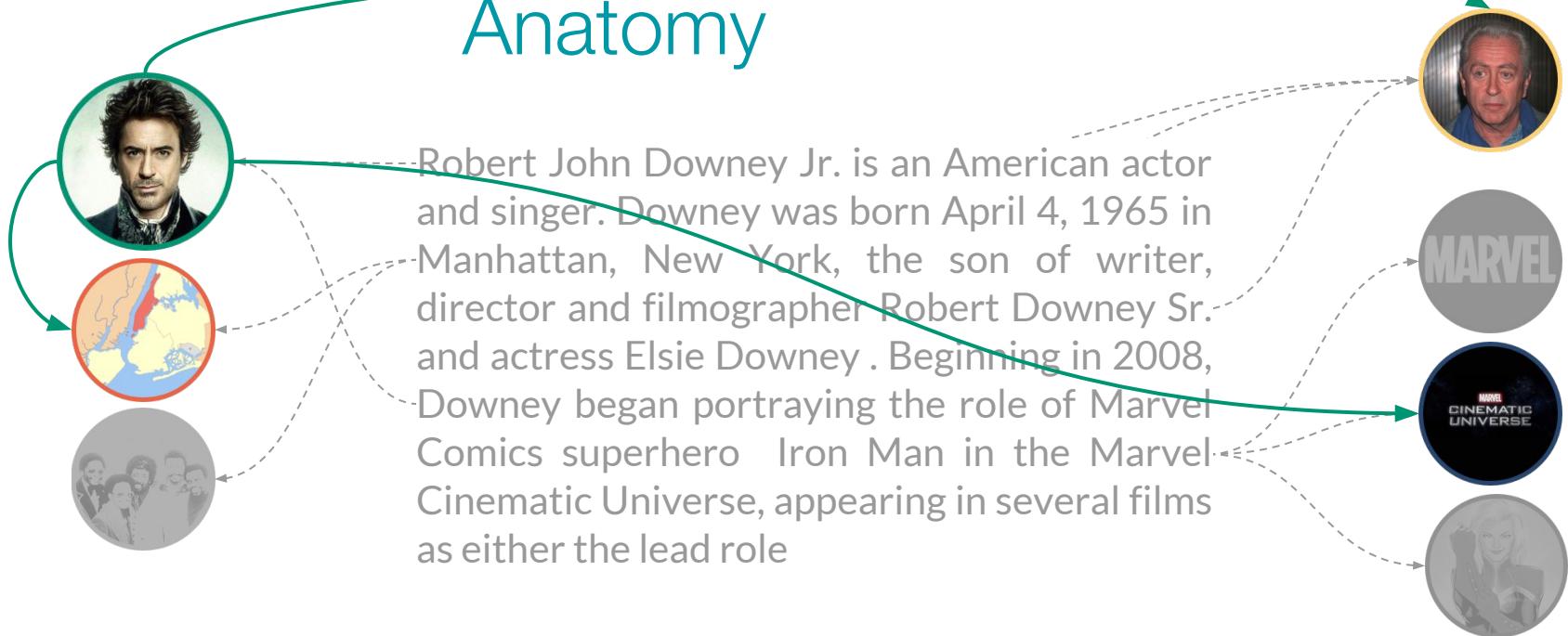


Mention Detection

Candidate Selection

Disambiguation

Anatomy



Mention Detection

Candidate Selection

Disambiguation



Confidence:

 A horizontal slider bar with a small grey square handle positioned near the left end. To its right is the value "0.5".

Language: English

n-best candidates

SELECT TYPES...

ANNOTATE

Robert [John Downey](#) Jr. is an [American](#) actor and singer. [Downey](#) was born April 4, 1965 in [Manhattan](#), [New York](#), the son of writer, director and filmographer [Robert Downey Sr.](#) and actress [Elsie Downey](#). Beginning in 2008, [Downey](#) began portraying the role of [Marvel Comics superhero Iron Man](#) in the [Marvel Cinematic Universe](#), appearing in several films as either the lead role

[BACK TO TEXT](#)

[Edit Text](#)

Language: eng Processed in: 0.1492 seconds

Robert John Downey Jr.**Normalized Entity Id:** Robert John

Downey Jr.

No wikipedia link.

No Freebase Id.

No Wikidata Id.

Confidence Score: 0.5**Relevance Score:** 0**DBpedia types:**

Person

Freebase types:

/people/person

or and singer.[Meaning](#) [Dependency Parse](#)n, New York, the son of writer, director and filmographer Robert Downey Sr. and[Meaning](#) [Dependency Parse](#)

Beginning in 2008, Downey began portraying the role of Marvel Comics superhero Iron Man in the Marvel Cinematic Universe, appearing in several films as either the lead role

[Words](#) [Phrases](#) [Relations](#) [Entities](#) [Meaning](#) [Dependency Parse](#)**CATEGORIES**

- 0.38 arts, culture and
entertainment>arts and
entertainment>cinema
- 0.37 arts, culture and
entertainment>arts and
entertainment>literature>fiction
- 0.36 economy, business and
finance>economic
sector>media>book industry
- 0.36 arts, culture and
entertainment>arts and
entertainment>cartoon
- 0.34 economy, business and
finance>economic sector>media

TOPICS

- 1.00 **Marvel Comics**
- 1.00 **Comics**
- 1.00 **Marvel Entertainment**

Relation Linking

Identify all the spans which refer to a **predicate** in the given KB.

Name all the movies in which Robert Downey Jr **acted**?



Relation Linking



- Name all the movies in which Robert Downey Jr **acted**?
- Which movies have RDJ?
- Flicks where I can see Robert DJ?
- Find me all the films **casting** Rober Downey Jr ?
- List all the movies **starring** Robert Downey Junior?
- RDJ has **acted** in which movies?

Relation Linking Approaches

Interaction between relations and entities.

Similarity between text and the predicate surface form



Sequential RL

Name all the movies in which Robert Downey Jr. acted?

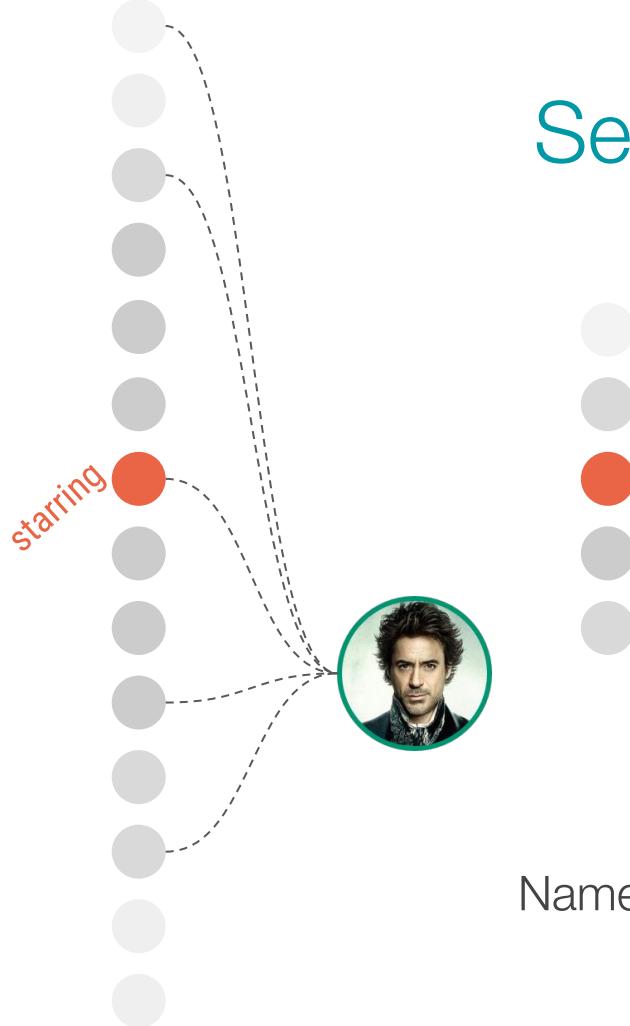
Sequential RL

starring



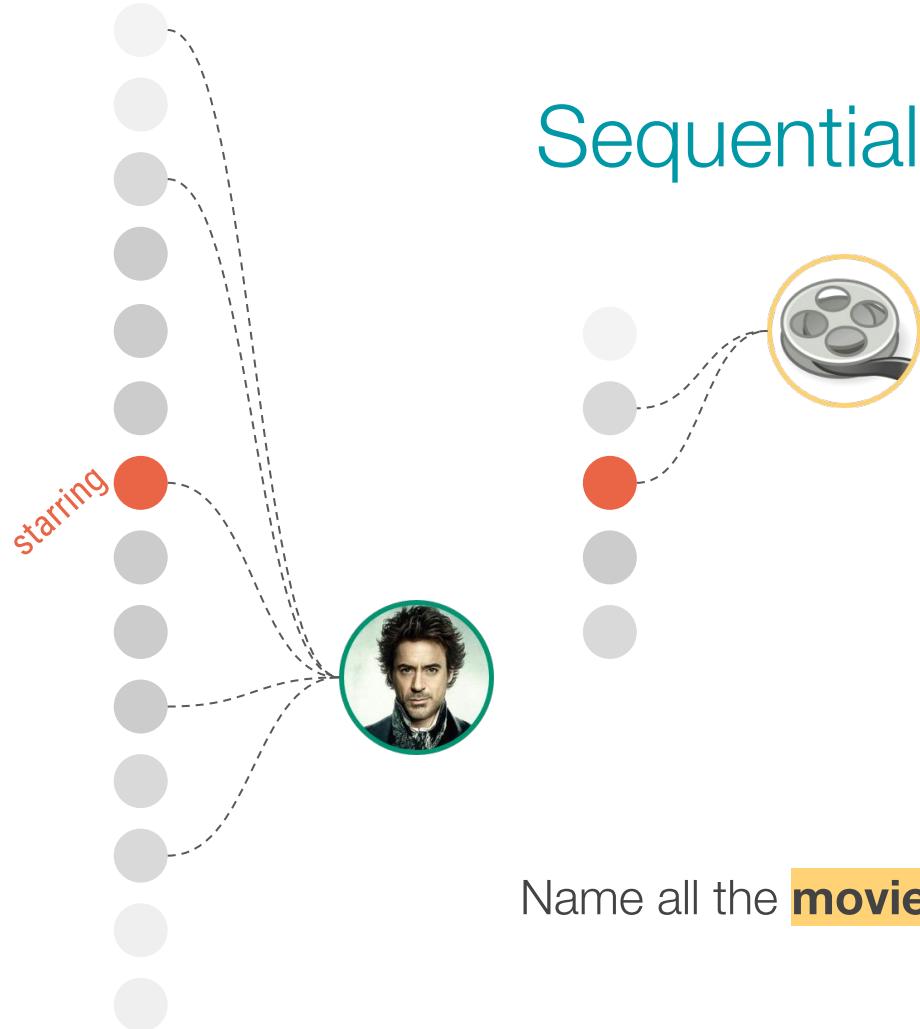
Name all the movies in which **Robert Downey Jr.** acted?

Sequential RL

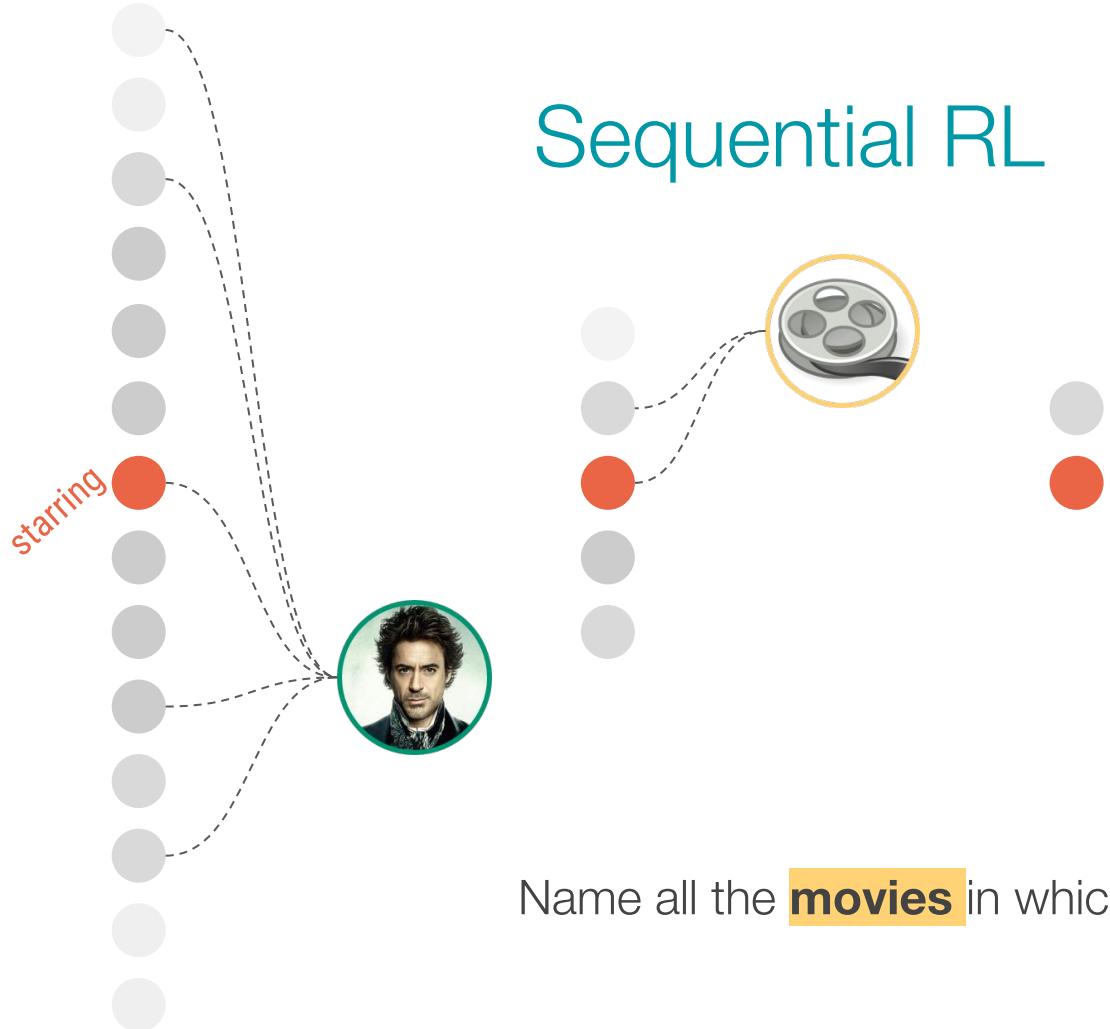


Name all the movies in which **Robert Downey Jr.** acted?

Sequential RL

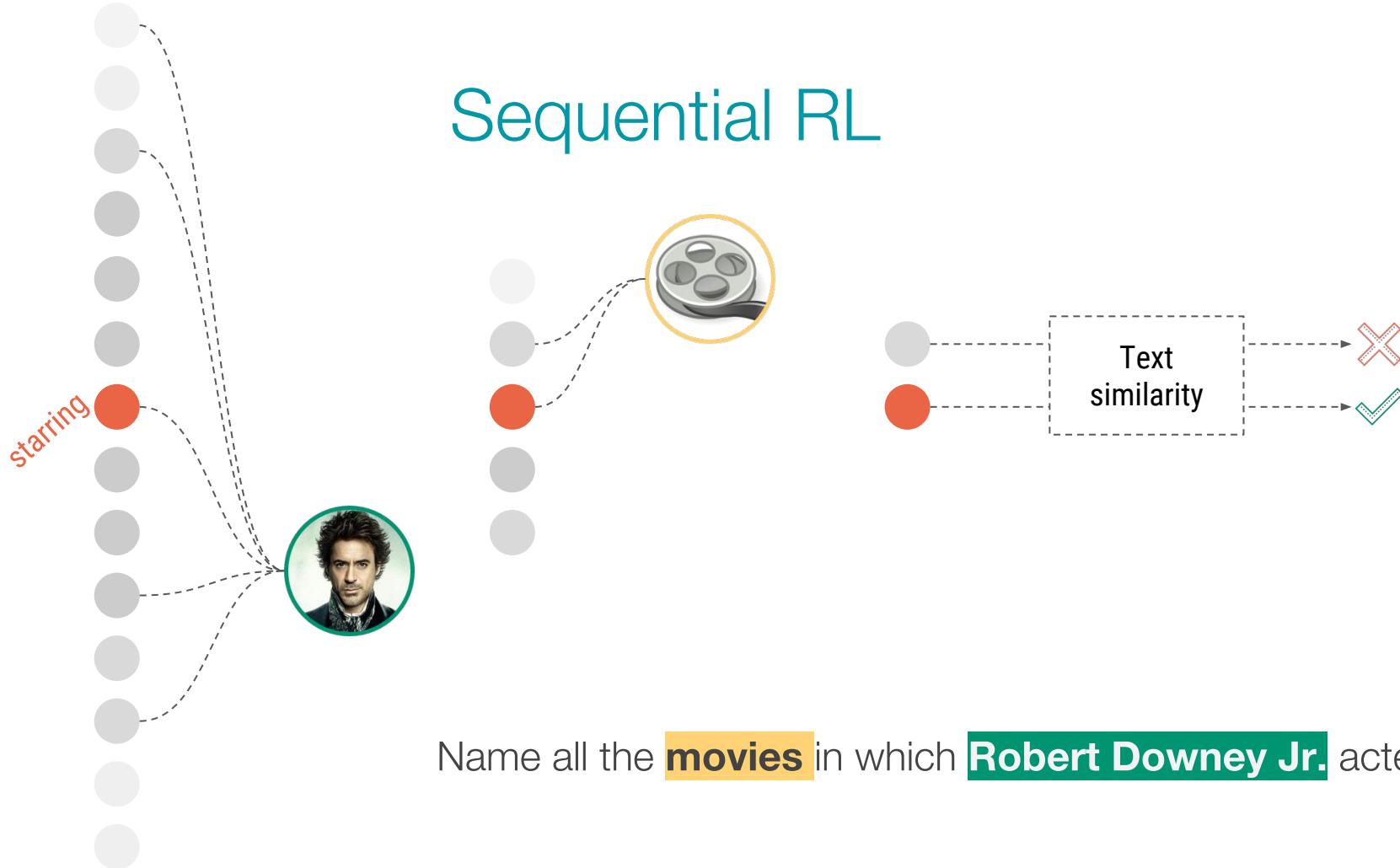


Sequential RL



Name all the **movies** in which **Robert Downey Jr.** acted?

Sequential RL



Sequential RL

- + Allows schema verification
- + Reduced candidate search space

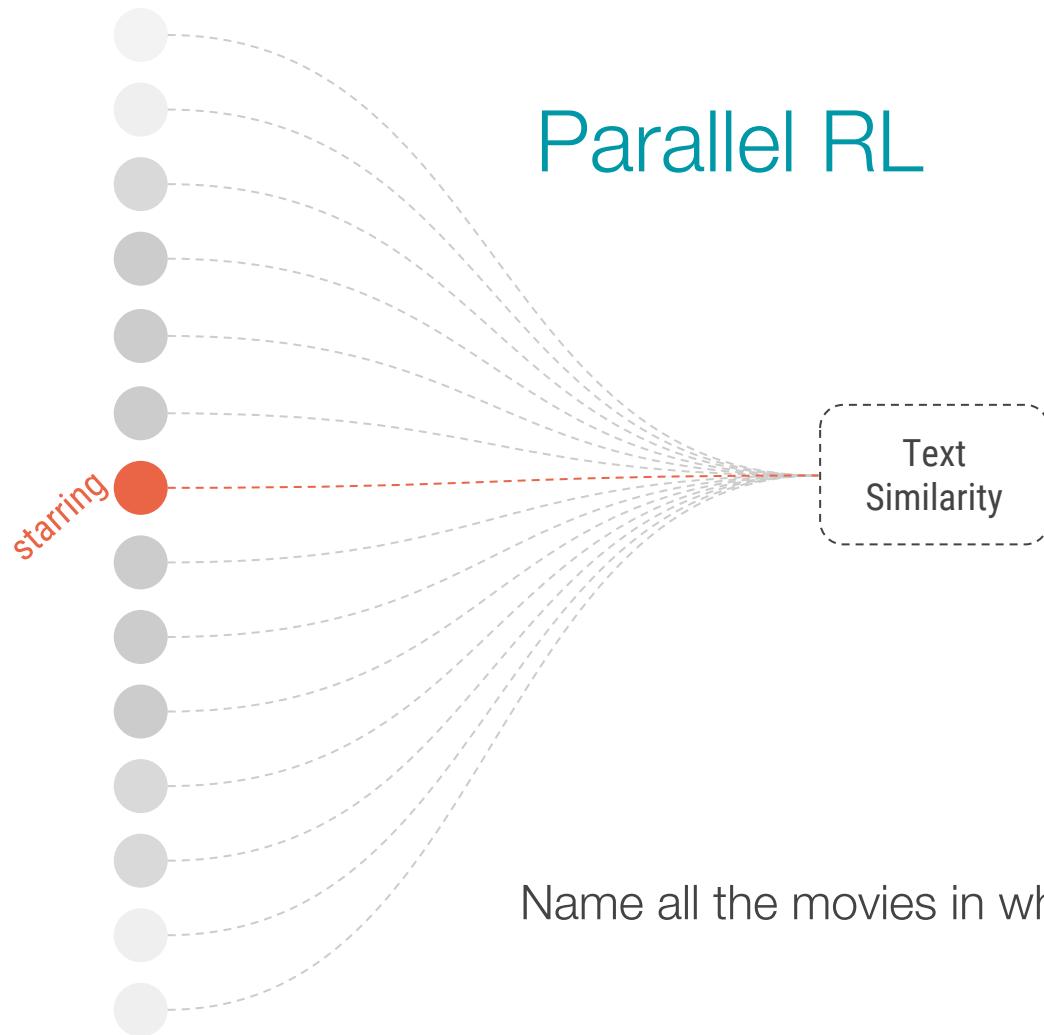
- Cannot use relation's info in entity linking process
- Errors in entity linking propagate



Parallel RL

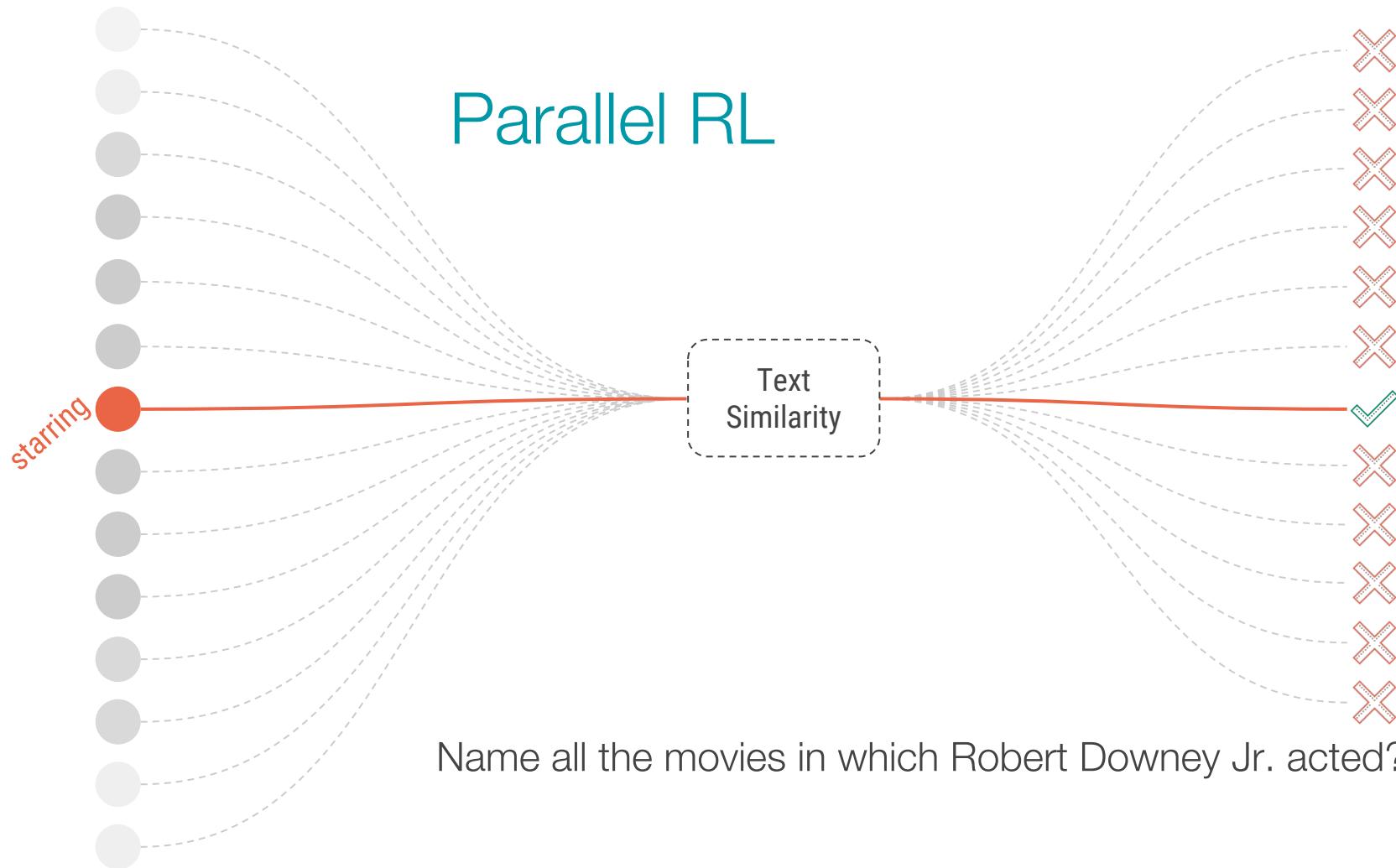
Name all the movies in which Robert Downey Jr. acted?

Parallel RL



Name all the movies in which Robert Downey Jr. acted?

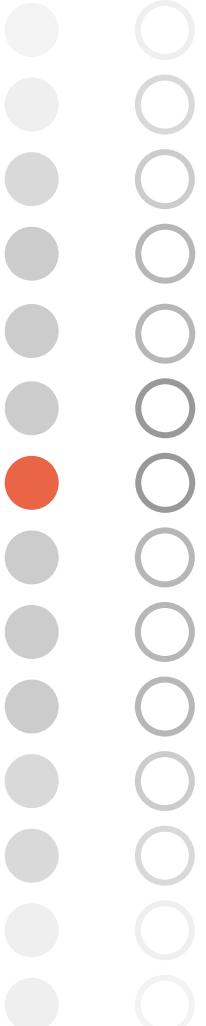
Parallel RL



Parallel RL

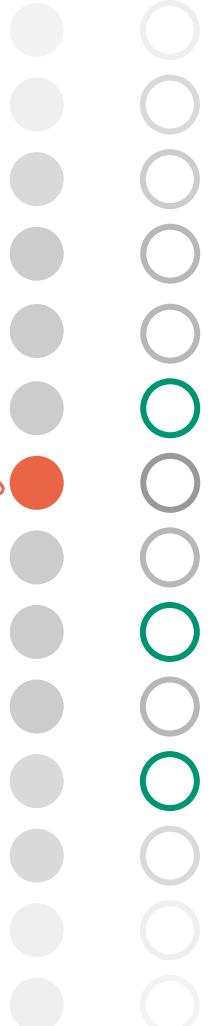
- + Computationally efficient
- + Can use results for entity linking later

- Result might be inconsistent with KB structure
- Generally lower precision



Joint RL

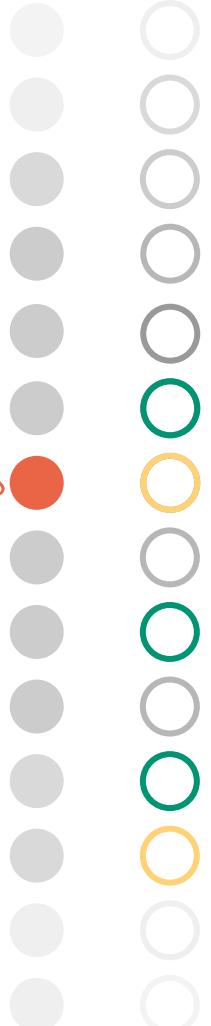
Name all the movies in which Robert Downey Jr. acted?



Joint RL

Text
Similarity

Name all the movies in which **Robert Downey Jr.** acted?



Joint RL

starring

Name all the **movies** in which **Robert Downey Jr.** acted?

2

Text
Similarity



Joint RL



2

Text
Similarity

KB
Structure

Name all the **movies** in which **Robert Downey Jr.** acted?



Joint RL



3

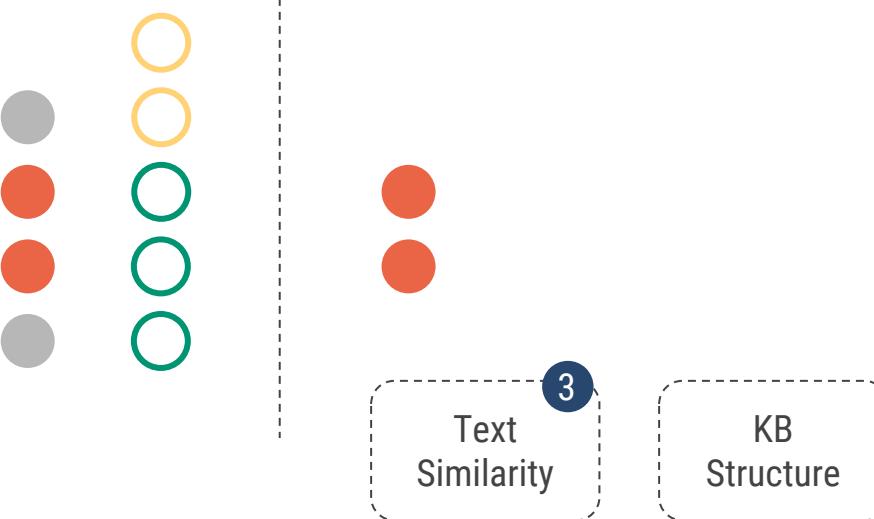
Text
Similarity

KB
Structure

Name all the **movies** in which **Robert Downey Jr. acted?**

Joint RL

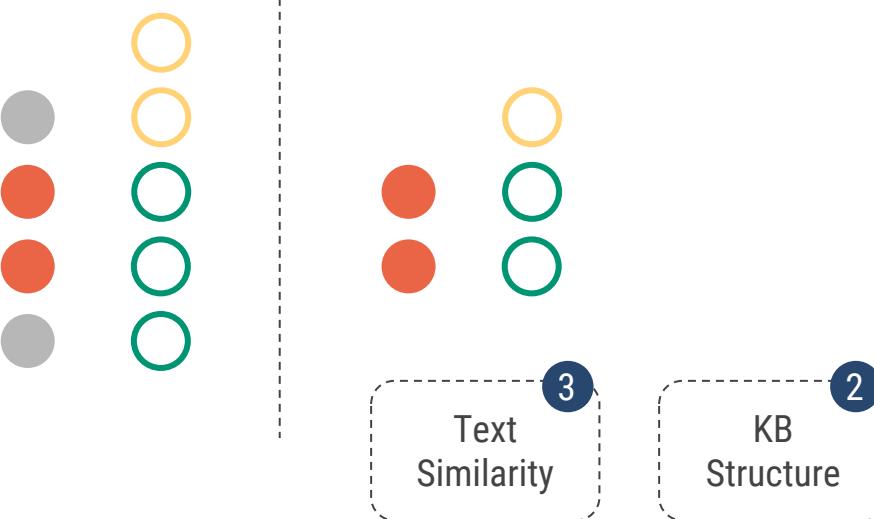
starring



Name all the **movies** in which **Robert Downey Jr. acted?**

Joint RL

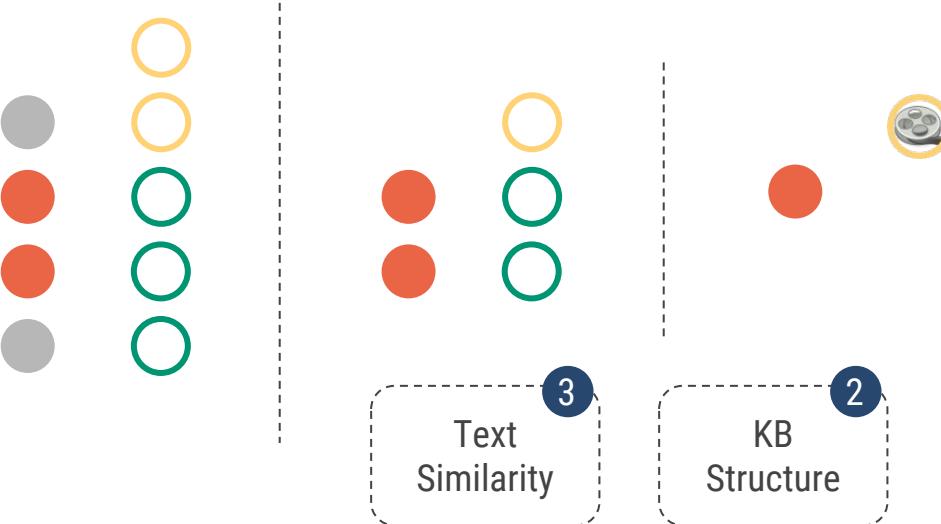
starring



Name all the **movies** in which **Robert Downey Jr. acted?**

Joint RL

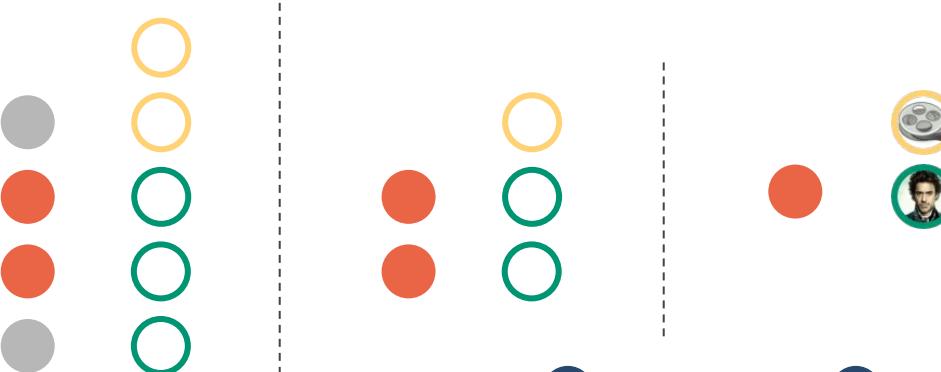
starring



Name all the **movies** in which **Robert Downey Jr. acted?**

Joint RL

starring



Name all the **movies** in which **Robert Downey Jr. acted?**

Joint RL

- + Better disambiguation
- + Allows re-ranking
- + Schema compatible results

- Computationally Inefficient
- Need probabilistic framework, with feedback mechanisms

Implicit Relations

Relations which are not explicitly present in the question but must be inferred based on KB schema and the rest of the words in the question.

Name all the movies of Robert Downey Jr. ?

dbo:starring needs to be inferred automatically.

Which movies have RDJ?

Flicks where I can see Robert DJ?

KB specific predicates

Different Kbs have different ways to handle constraints.

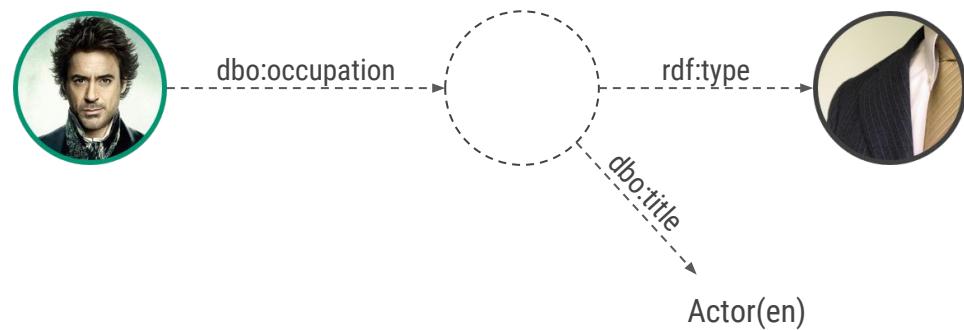
Who is the brother of Robert Downey Jr ?

Dbpedia - Brother

Freebase - Sibling + male

Handling Blank Nodes

Not a real-world entity, but is used to collect multiple fields of an event or a special relationship.



Auxiliary Constraints

Name everything where **Robert Downey Jr** **Acted?**

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
}
```

Auxiliary Constraints

Name everything where **Robert Downey Jr Acted?**

Questions might ask for:

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
}
```

Auxiliary Constraints

Name all the **movies** in which **Robert Downey Jr** **Acted?**

Questions might ask for:

Specific **type** of answers

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
    ?uri rdf:type dbo:Film  
}
```

Auxiliary Constraints

How many movies has RDJ acted in?

Questions might ask for:

Count the number of results.

```
SELECT count(?uri) WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
}
```

Auxiliary Constraints

Name the movies RDJ acted in after 2005?

Questions might ask for:

Filter the results based on some arbitrary metric.

```
SELECT ?uri WHERE {  
    ?uri dbo:starring dbr:Robert_Downey_Jr .  
    ?uri dbo:releaseDate ?date .  
    FILTER (?date >= xsd:date("2005-01-01"))  
}
```

With great datasets come great solutions

Priyansh Trivedi
ISWC 2017, Vienna



Datasets

Several datasets based on different KB having varying complexity and size addressing different aspects of Question Answering systems.

Dataset(s)

Dataset	Size	Logical Forms	Complex Questions	Target KB
Free917 (Cai et al., 2013)	917	Yes	Yes	Freebase
WebQuestions (Berant et al., 2013)	5 810	No	Yes	Freebase
SimpleQuestions (Bordes et al., 2015)	108 442	No	No	Freebase
30M Factoid (Serban et al., 2016)	30 000 000	No	No*	Freebase
QALD (Unger et al., 2016)	450	Yes	Yes	DBpedia
LC-QuAD (Trivedi et al. 2017)	5000	Yes	Yes	DBpedia

Interesting approaches

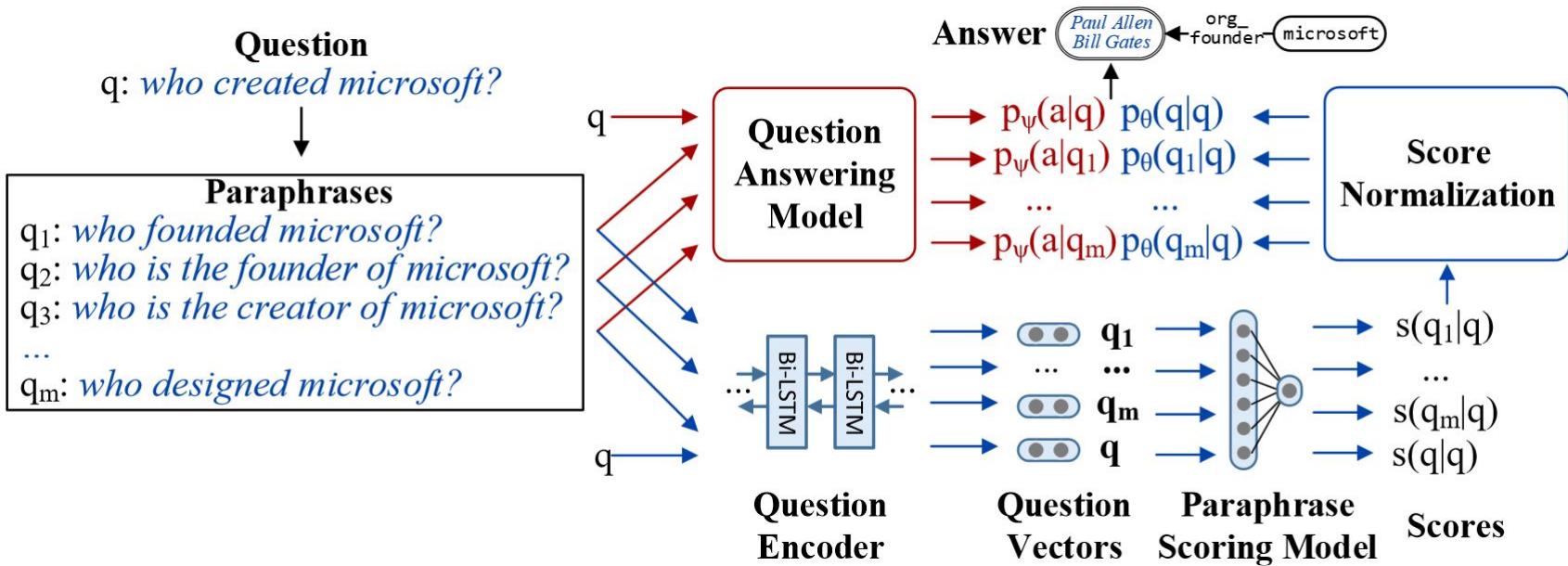
End to End Networks for QA

Golub et al., 2016

- Reduce search space by simple IR techniques
 - Substring match between question and possible entities.
- Use NN to encode the question, and KB entities and predicates
- A semantic relevance function is then used to find the most likely entity, relation pair.

Paraphrasing Questions

Dong et al., 2016



References.

Images of **manhattan** and **marvel cinematic universe** have been taken from wikipedia

https://en.wikipedia.org/wiki/Manhattan#/media/File:New_York_City_location_Manhattan.svg

https://en.wikipedia.org/wiki/Marvel_Cinematic_Universe#/media/File:Marvel_Cinematic_Universe_logo.png

Robert Downey Jr. image from slide 40 onwards has been taken from
https://commons.wikimedia.org/wiki/File:Robert_Downey,_Jr._2012.jpg

References.

Cai, Qingqing, and Alexander Yates. "**Large-scale Semantic Parsing via Schema Matching and Lexicon Extension.**" ACL (1). 2013.

Bordes, Antoine, et al. "**Large-scale simple question answering with memory networks.**" arXiv preprint arXiv:1506.02075 (2015).

Serban, Iulian Vlad, et al. "**Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus.**" arXiv preprint arXiv:1603.06807 (2016).

Unger, Christina, Axel-Cyrille Ngonga Ngomo, and Elena Cabrio. "**6th open challenge on question answering over linked data (qald-6).**" Semantic Web Evaluation Challenge. Springer International Publishing, 2016.

References.

Liang, Chen, et al. "**Neural symbolic machines: Learning semantic parsers on freebase with weak supervision.**" arXiv preprint arXiv:1611.00020 (2016).

Trivedi, Priyansh, et al. "**Lc-quad: A corpus for complex question answering over knowledge graphs.**" International Semantic Web Conference. Springer, Cham, 2017.

Golub, David, and Xiaodong He. "**Character-level question answering with attention.**" arXiv preprint arXiv:1604.00727 (2016).

Dong, Li, et al. "**Learning to paraphrase for question answering.**" arXiv preprint arXiv:1708.06022 (2017).

Qanary