



# 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 @TODO



Where is  
ESWC 2018 held?



Who is the father of  
Luke Skywalker?



What is the best dessert  
in the world?

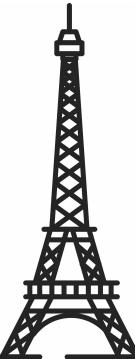
Name  
one ex  
president  
of United States?



# Introduction

# All NLP/AI task can be reduced to Question Answering

Richard Socher,  
Deep Learning Summit 2016, SF

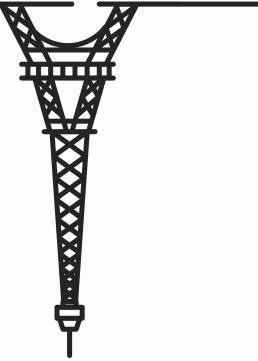


# Factoid Questions

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

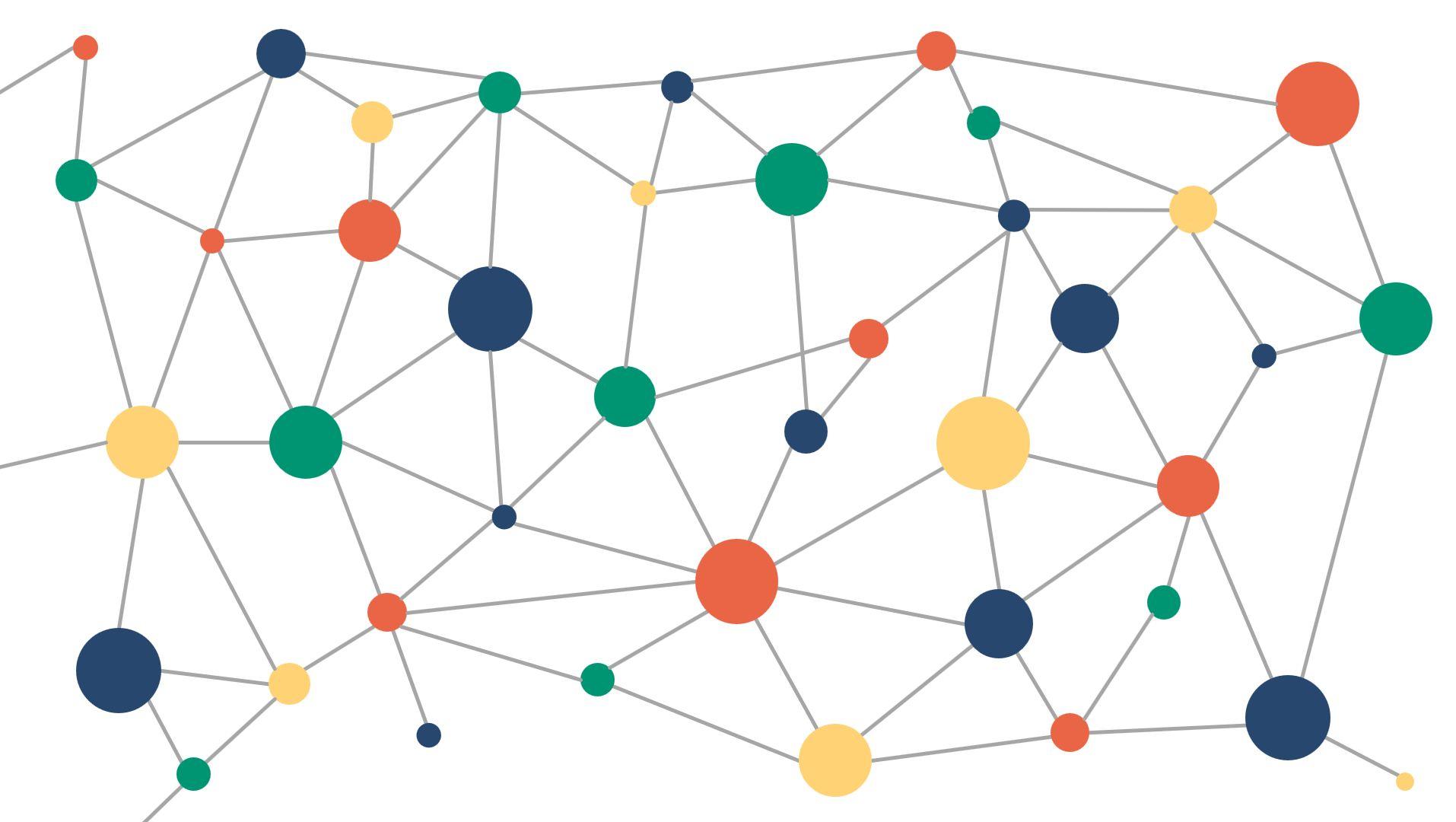
Given:

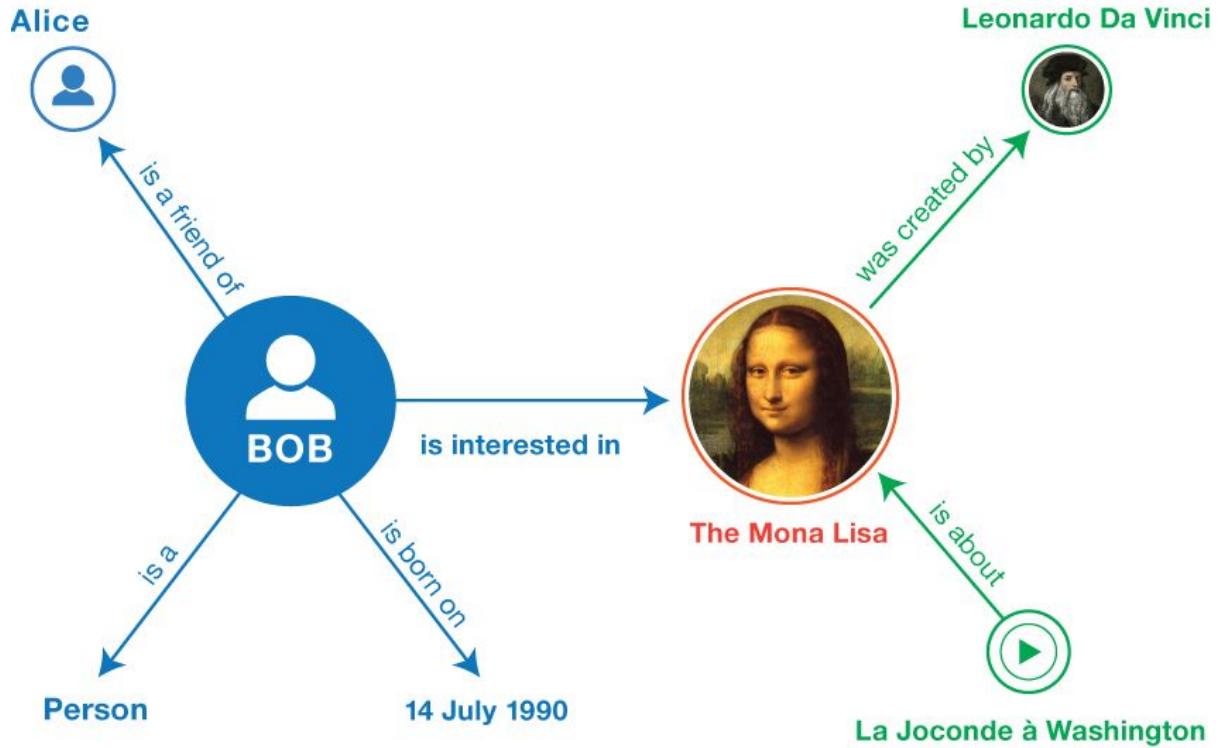
- a natural language question
- a source knowledge base

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

# Knowledge Graphs

*"Things instead of strings"*







# QA over Knowledge Bases

## 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



[People also search for](#)



**"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. Multilingual.



Freebase: 1.9B triples. Deprecated. Extremely popular.



Wikidata

# 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.

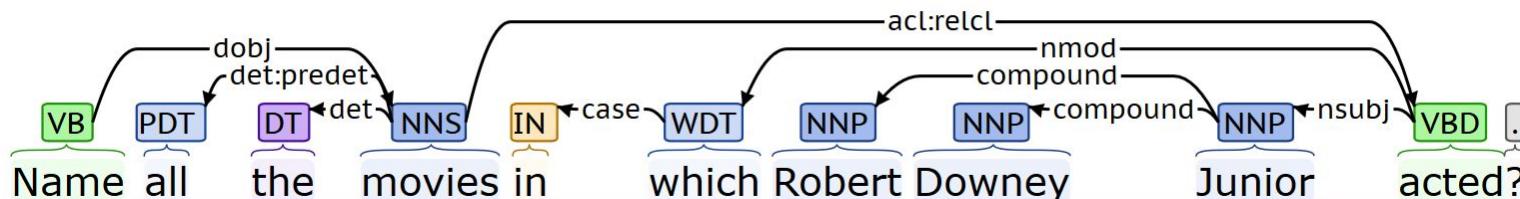
Semantic parsing?

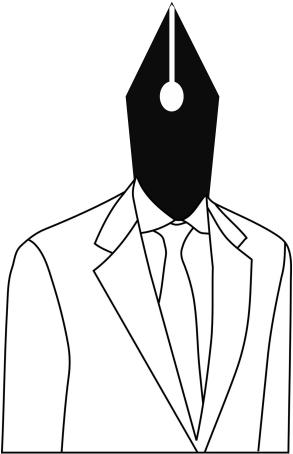
# Question Answering

# 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.

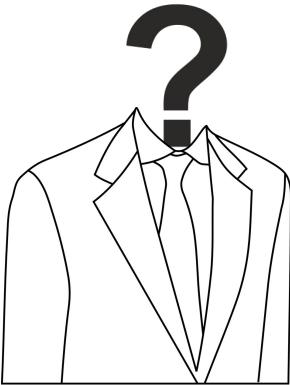




# Formal Languages

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

Eg. SPARQL; Python (most programming languages);  
 $\lambda$  - DCS.



# Formal Languages

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

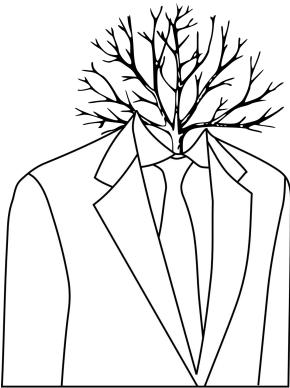
Eg. SPARQL; Python (most programming languages);  
 $\lambda$  - DCS; **user-defined**.



## Example: SPARQL

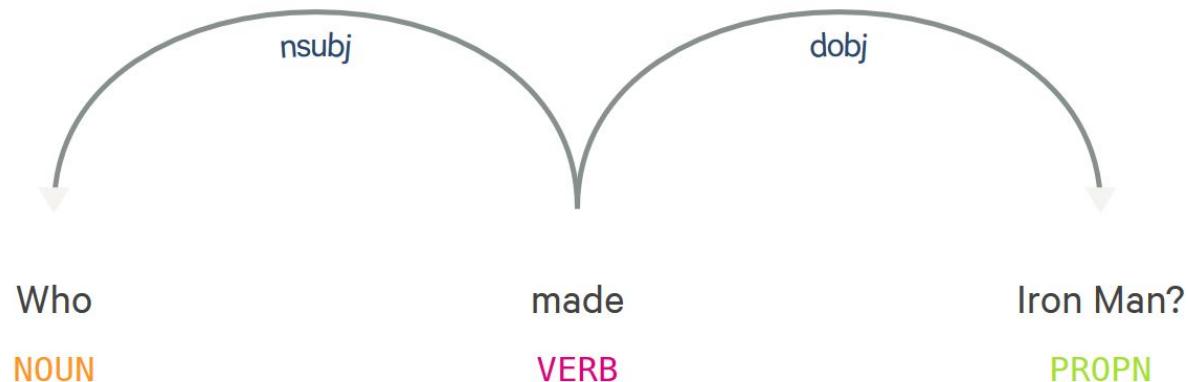
Who **made** **Iron man**?

```
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*.



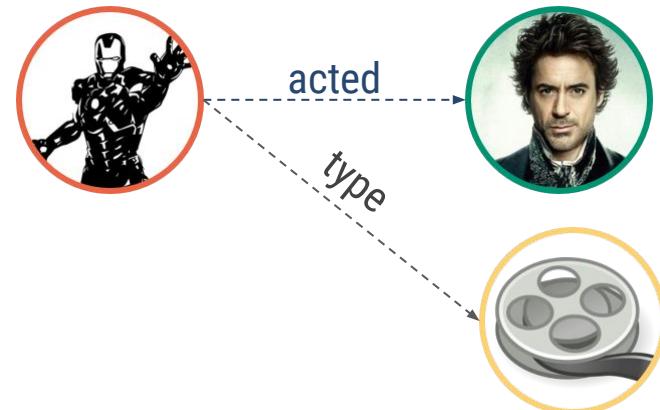
# Semantic Parsing

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



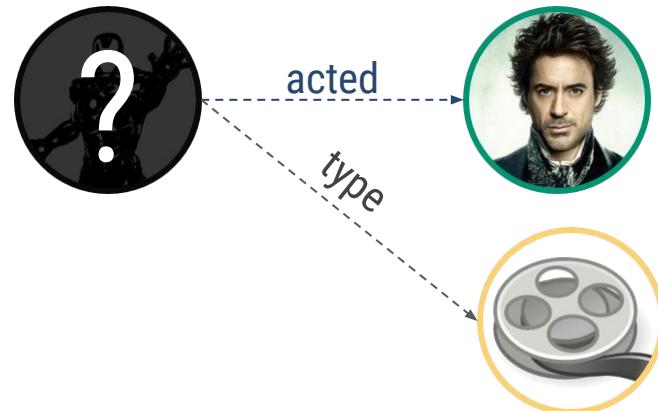
# Semantic Parsing

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



# Semantic Parsing

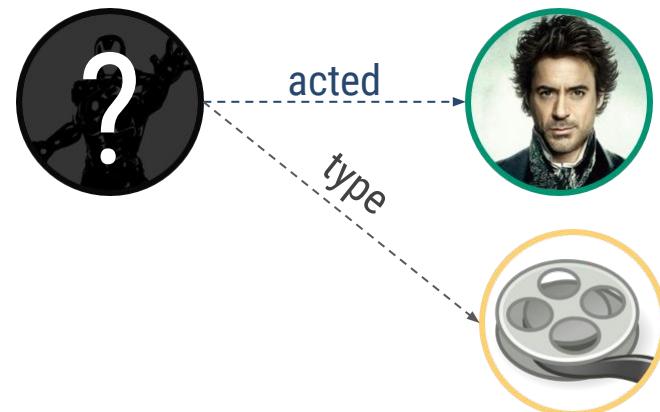
Name all the movies in which Robert Downey Jr Acted?



# 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?

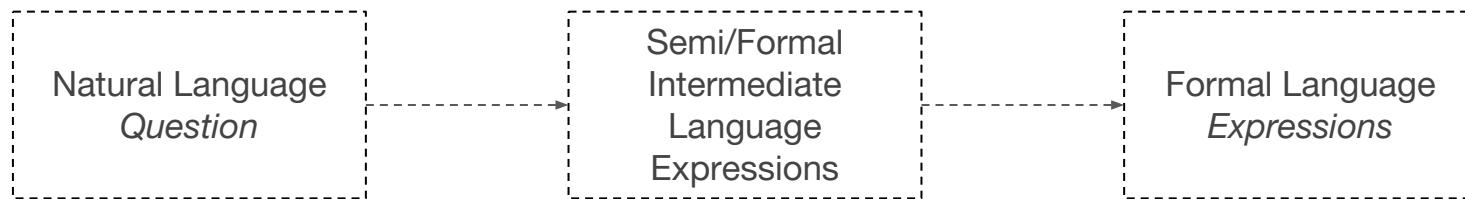
RDJ has acted in which movies?

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

\* rdf:type constraint omitted for brevity's sake



# Two step process



Natural  
Language  
*Question*

Semi/Formal  
Intermediate  
Language  
Expressions

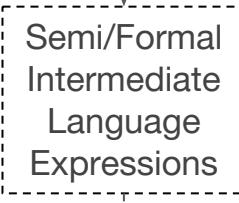
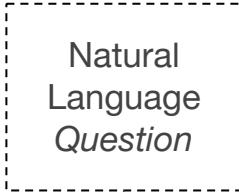
Formal  
Language  
*Expressions*

# 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.

Natural  
Language  
*Question*

Semi/Formal  
Intermediate  
Language  
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**

Formal  
Language  
*Expressions*

\* 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



Natural  
Language  
*Question*



Semi/Formal  
Intermediate  
Language  
*Expressions*

Formal  
Language  
*Expressions*

# 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.



Natural  
Language  
*Question*

Semi/Formal  
Intermediate  
Language  
Expressions

Formal  
Language  
*Expressions*

# Final Query Construction

**Should this be here? What content?**

# 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?

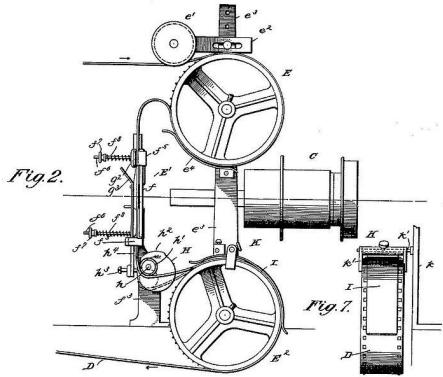


dbr:Robert\_Downey\_Jr

# Entity Linking

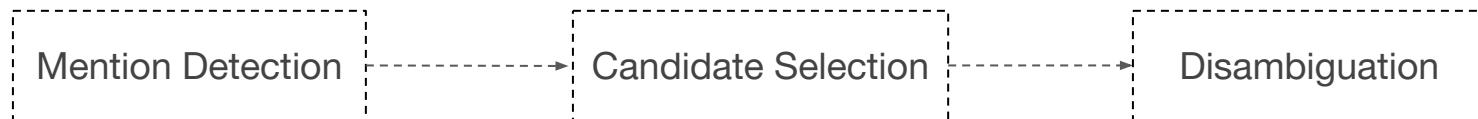


- Name all the movies in which **Robert Downey Jr** acted?
- Which movies have **RDJ**?
- Flicks where I can see **Robert DJ**?
- Find me all the films casting **Rober Downey Jr** ?
- List all the movies starring **Robert Downey Junior**?
- RDJ** has acted in which movies?



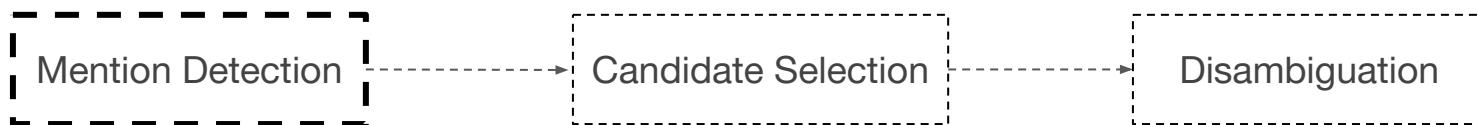
# Anatomy

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

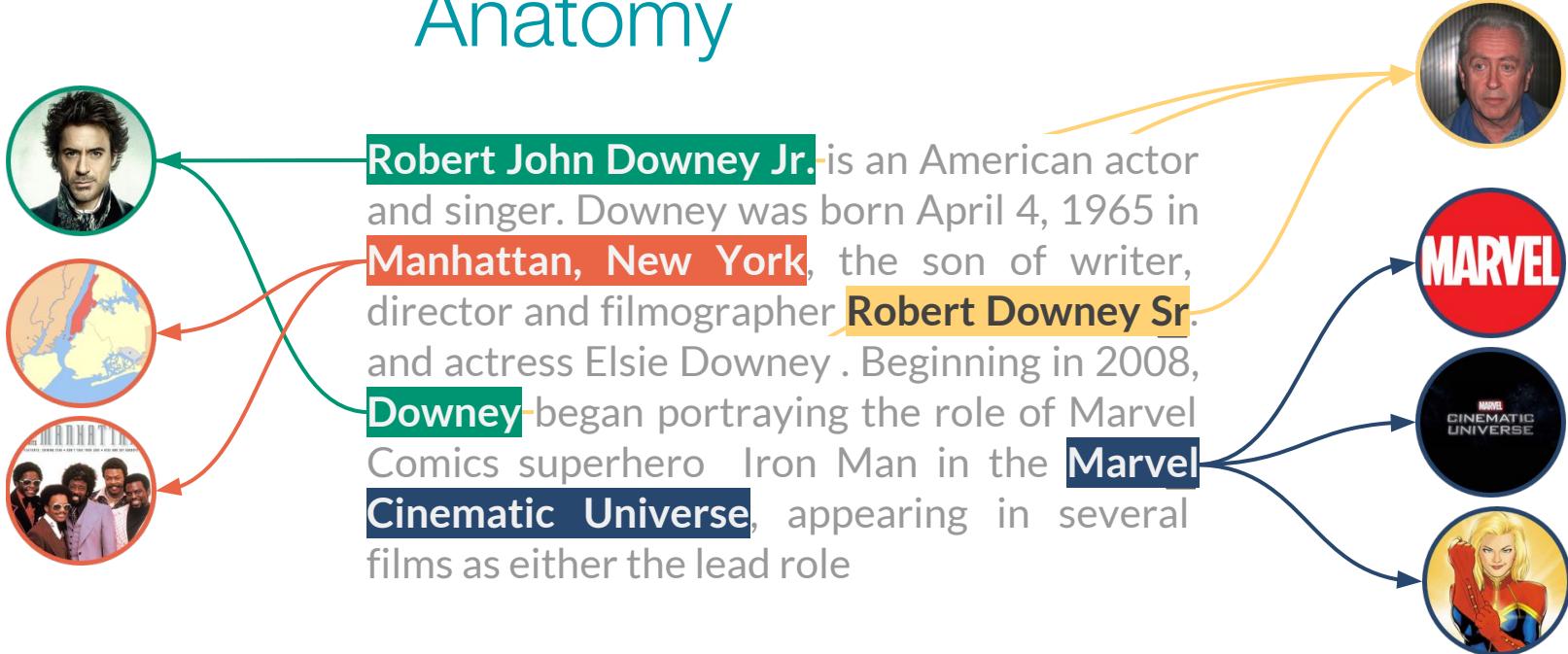


# Anatomy

**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

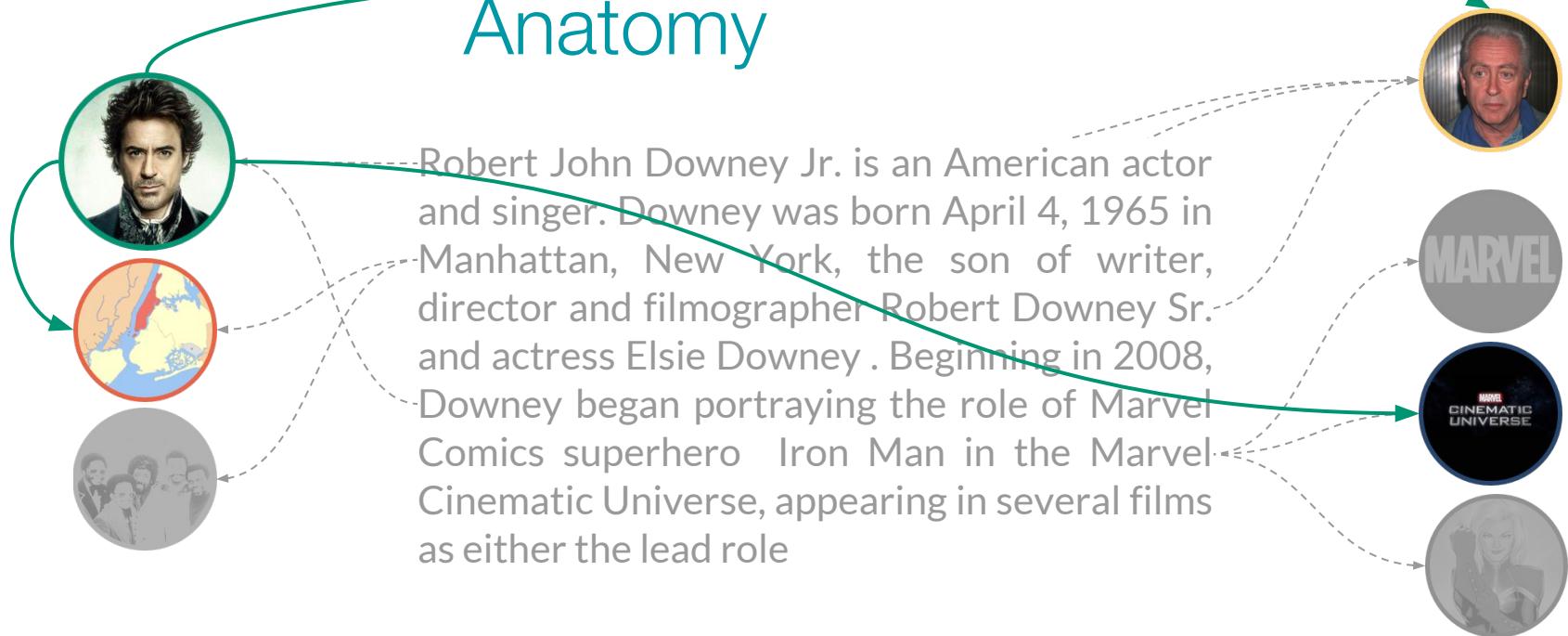


Mention Detection

Candidate Selection

Disambiguation

# Anatomy



Mention Detection

Candidate Selection

Disambiguation



Confidence:

 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?

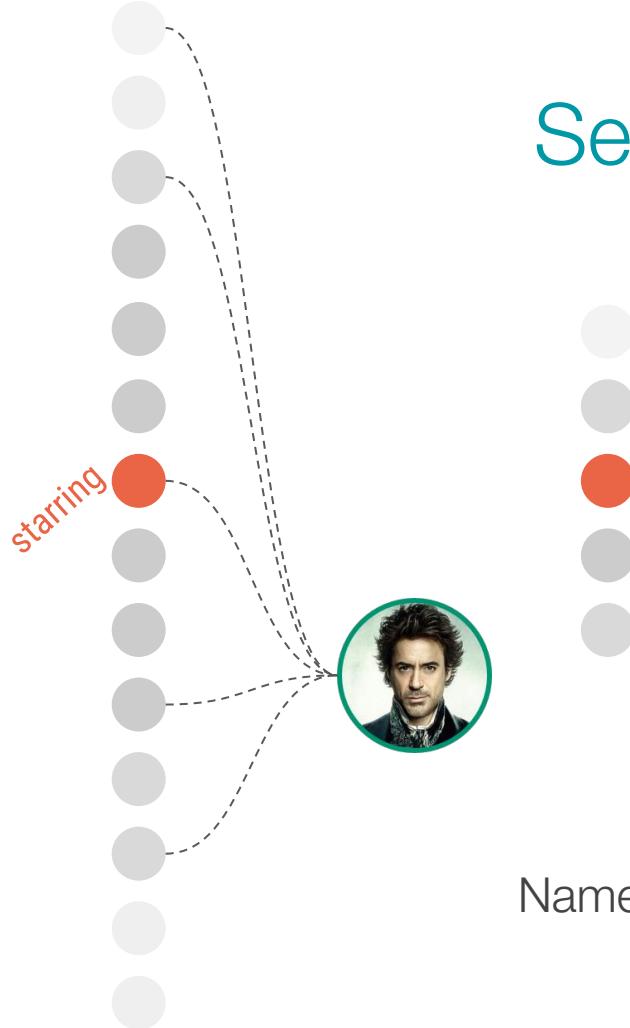
# Sequential RL

starring



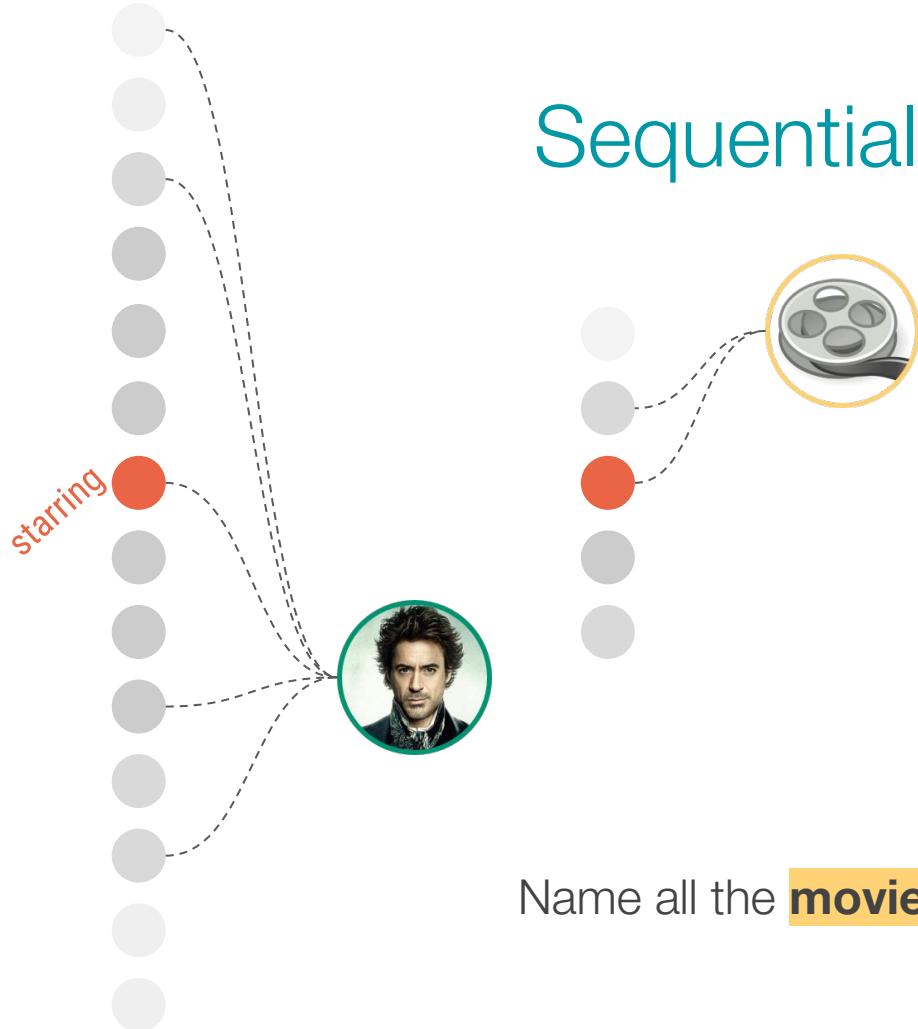
Name all the movies in which **Robert Downey Jr.** acted?

# Sequential RL

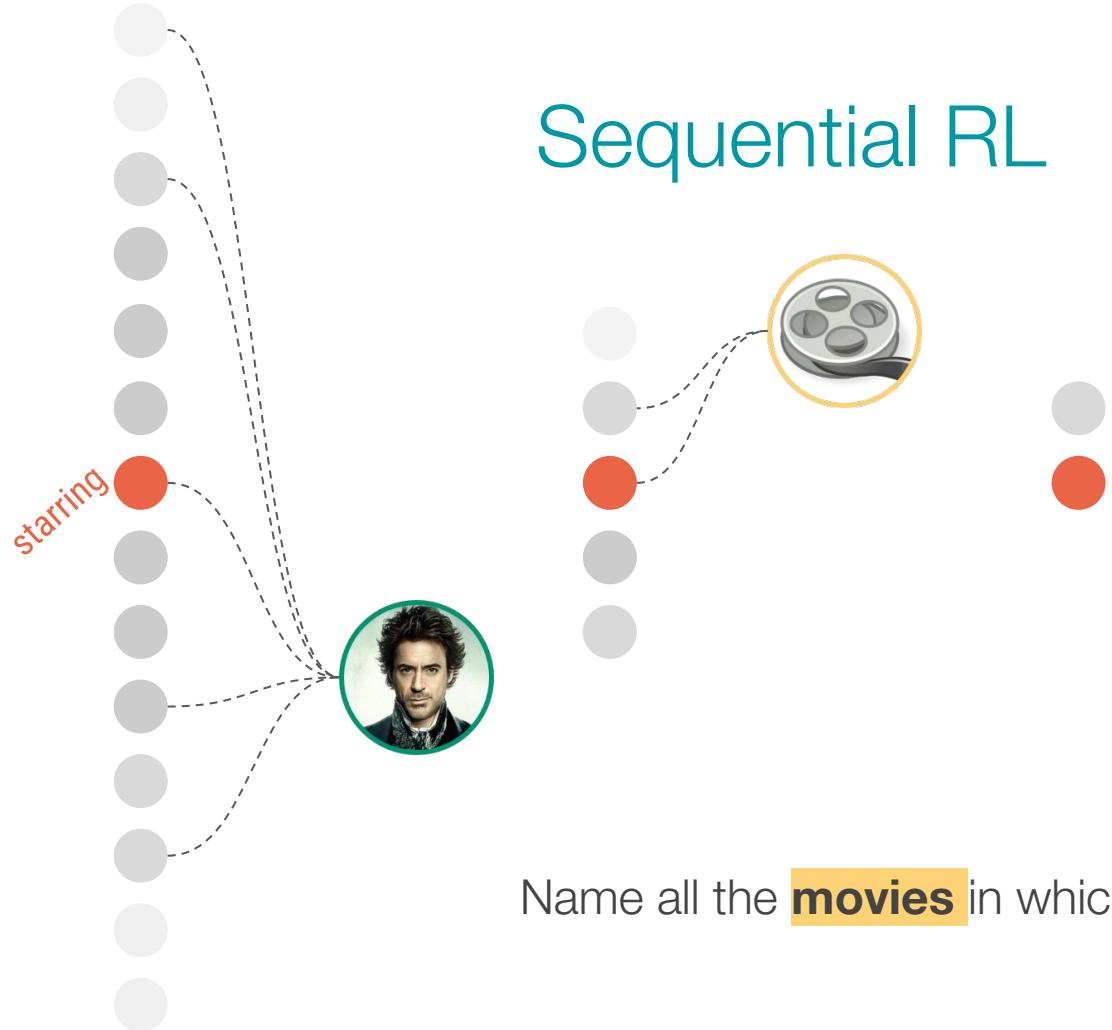


Name all the movies in which **Robert Downey Jr.** acted?

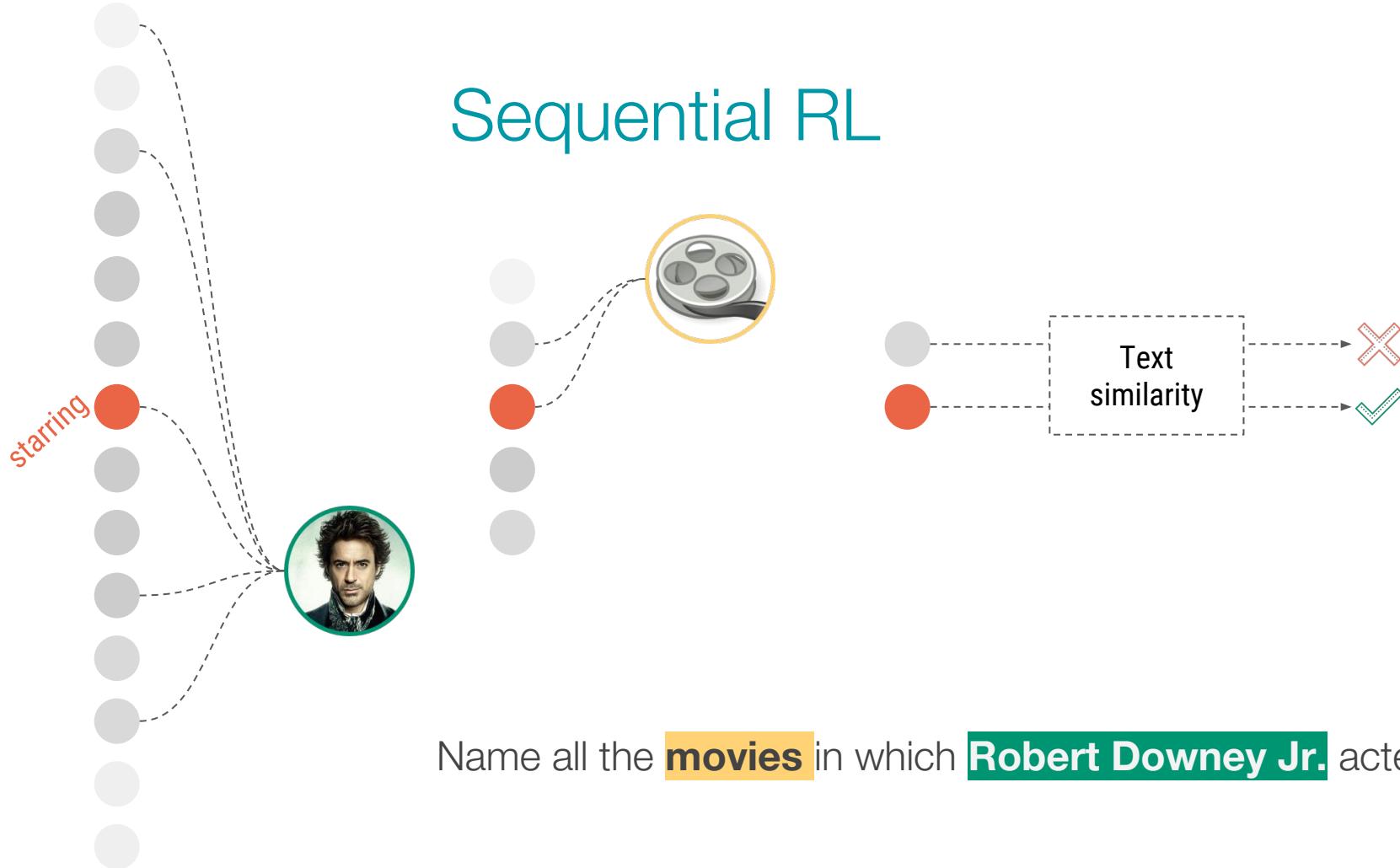
# Sequential RL



# Sequential RL

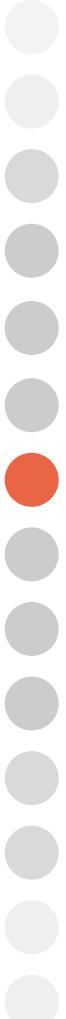


# Sequential RL



# Sequential RL

- + Allows schema verification
- + Reduced candidate search space
  
- Cannot use relation's info in entity linking process
- Errors in entity linking propagate

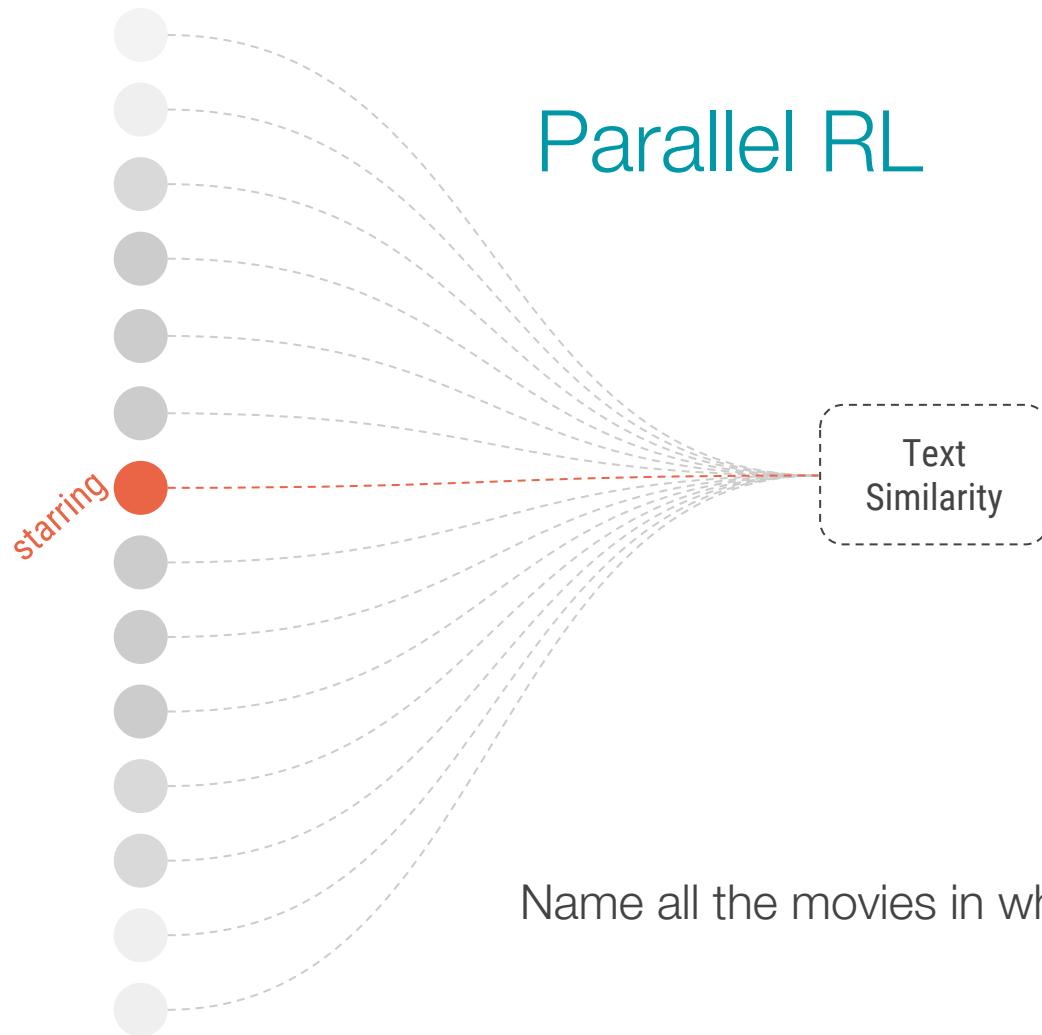


# Parallel RL

starring

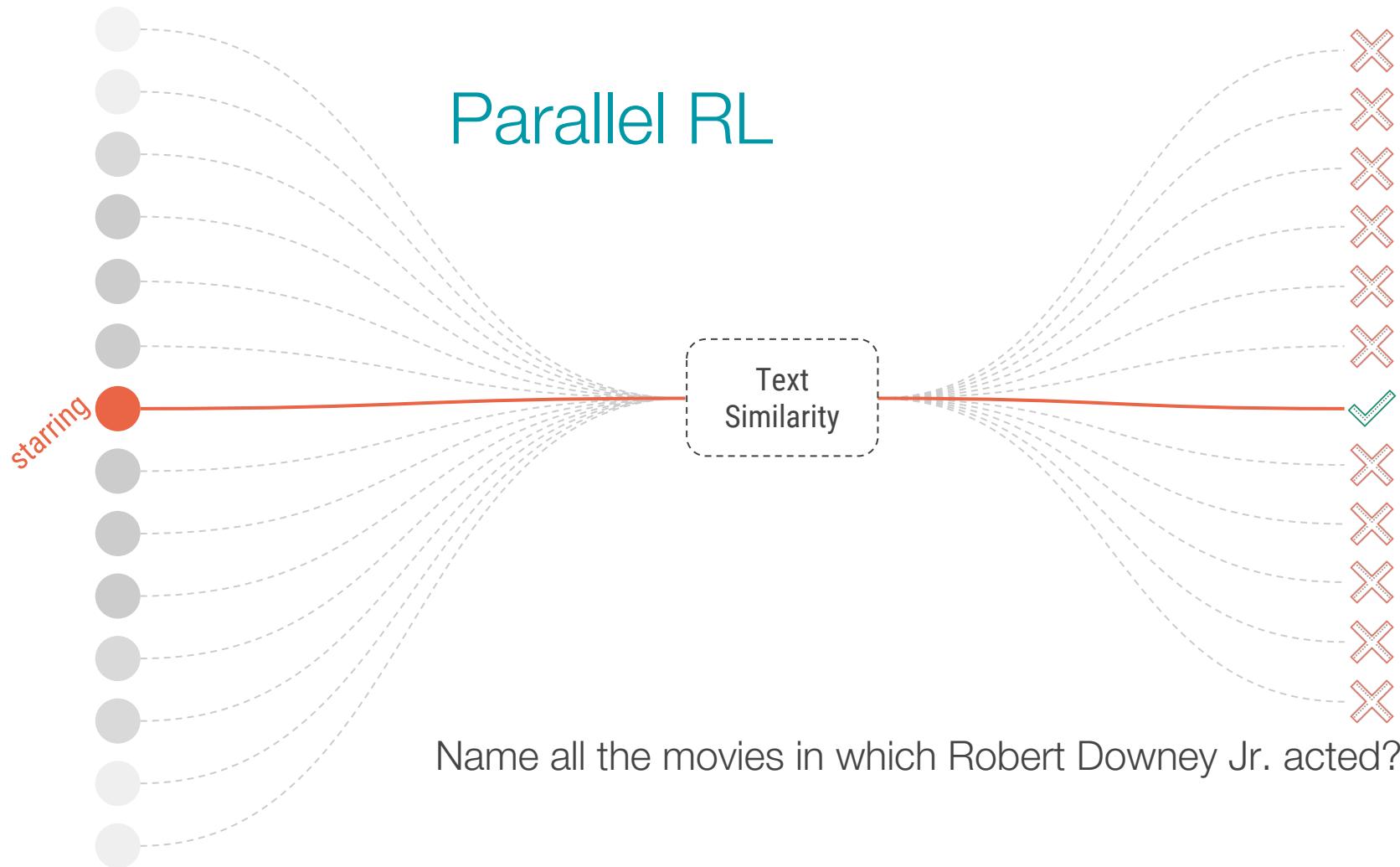
Name all the movies in which Robert Downey Jr. acted?

# Parallel RL



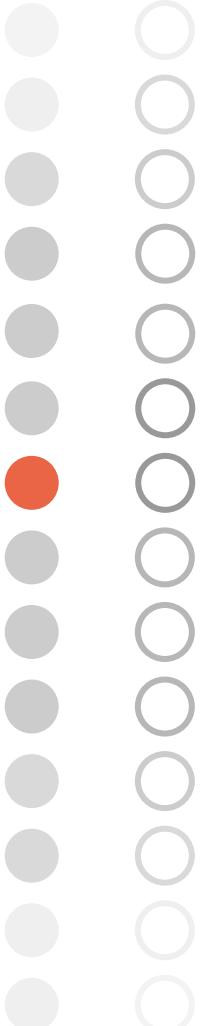
Name all the movies in which Robert Downey Jr. acted?

# Parallel RL



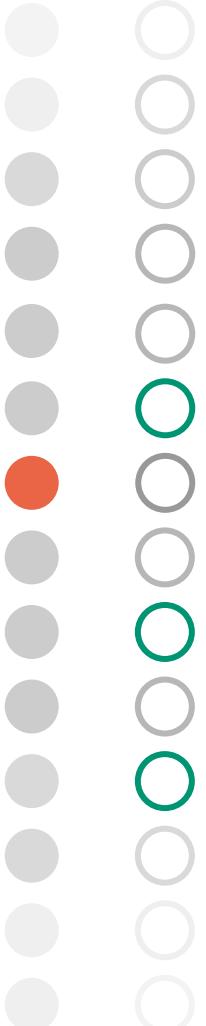
# Parallel RL

- + Computationally efficient
- + Can use results for entity linking later
  
- Result might be inconsistent with KB structure
- Generally lower precision



## Joint RL

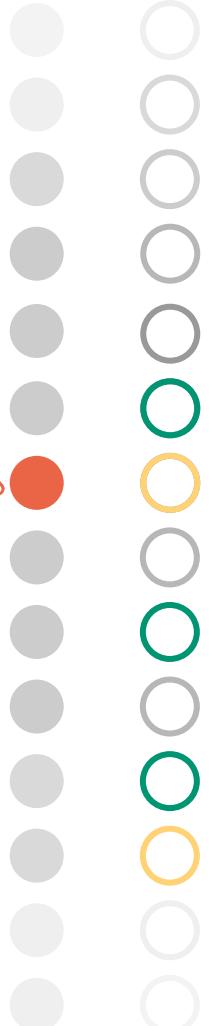
Name all the movies in which Robert Downey Jr. acted?



## Joint RL

Text  
Similarity

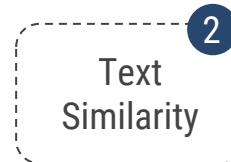
Name all the movies in which **Robert Downey Jr.** acted?



## Joint RL

starring

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





## Joint RL



2

Text  
Similarity

KB  
Structure

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



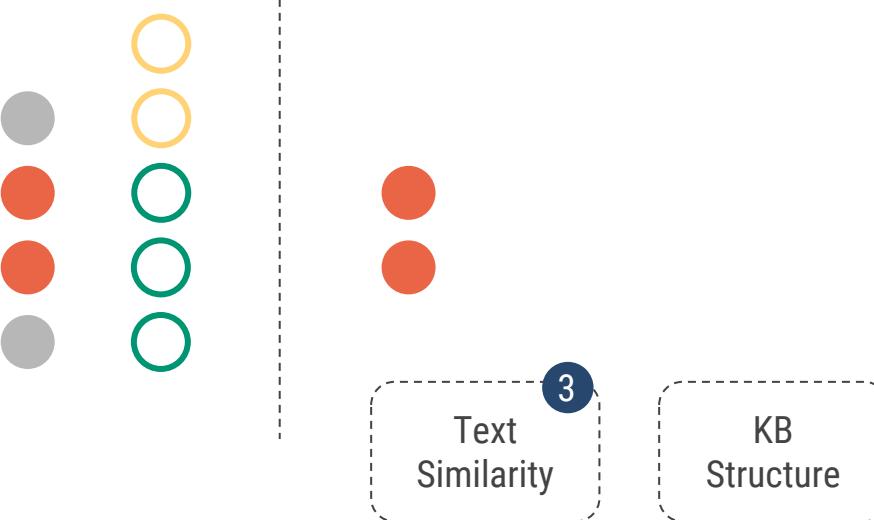
## Joint RL



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

# Joint RL

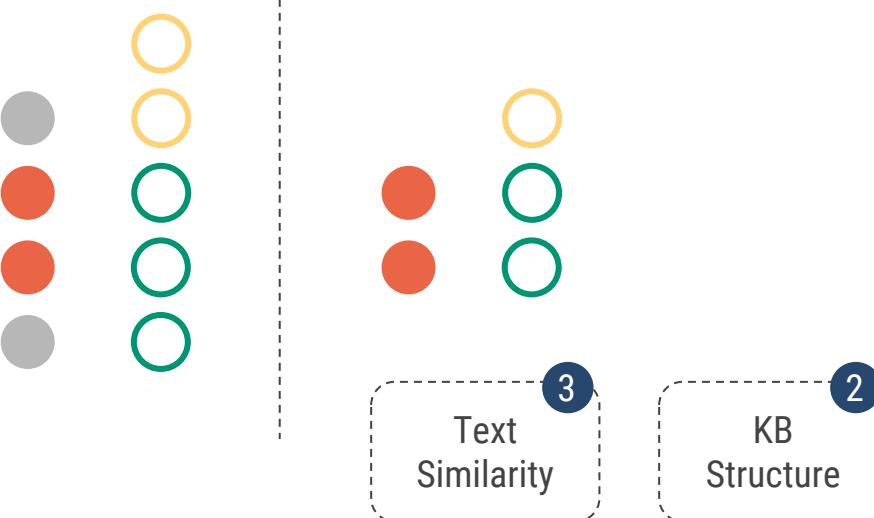
starring



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

# Joint RL

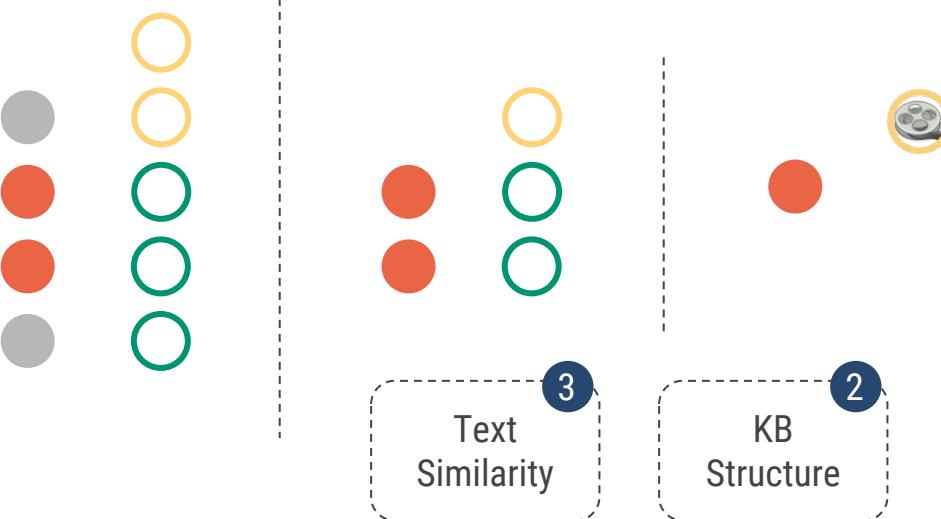
starring



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

# Joint RL

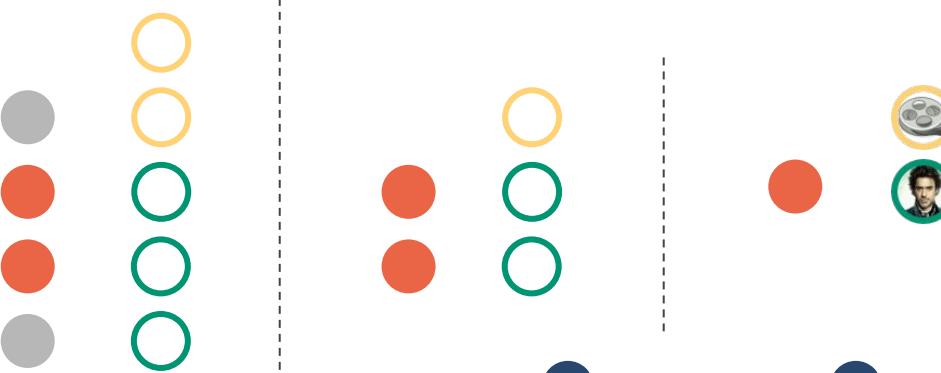
starring



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.

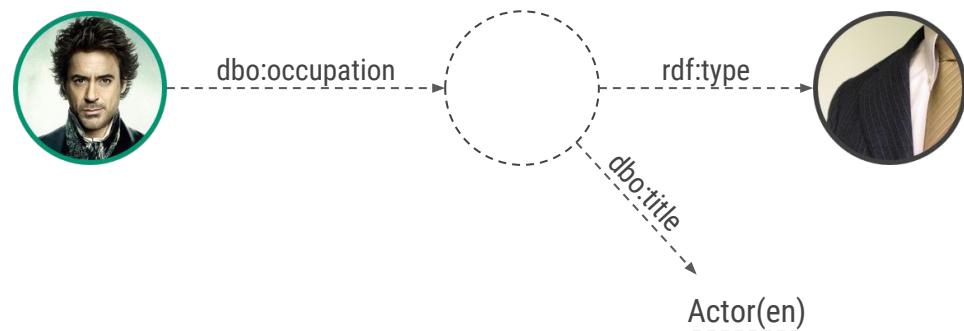
Who is the brother of Robert Downey Jr ?

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<br>(Cai et al., 2013 )           | 917        | Yes           | Yes               | Freebase  |
| WebQuestions<br>(Berant et al., 2013)    | 5 810      | No            | Yes               | Freebase  |
| SimpleQuestions<br>(Bordes et al., 2015) | 108 442    | No            | No                | Freebase  |
| 30M Factoid<br>(Serban et al., 2016)     | 30 000 000 | No            | No*               | Freebase  |
| QALD (Unger et al., 2016)                | 450        | Yes           | Yes               | DBpedia   |
| LC-QuAD                                  | 5000       | Yes           | Yes               | DBpedia   |

# Interesting approaches

# End to End Networks for QA

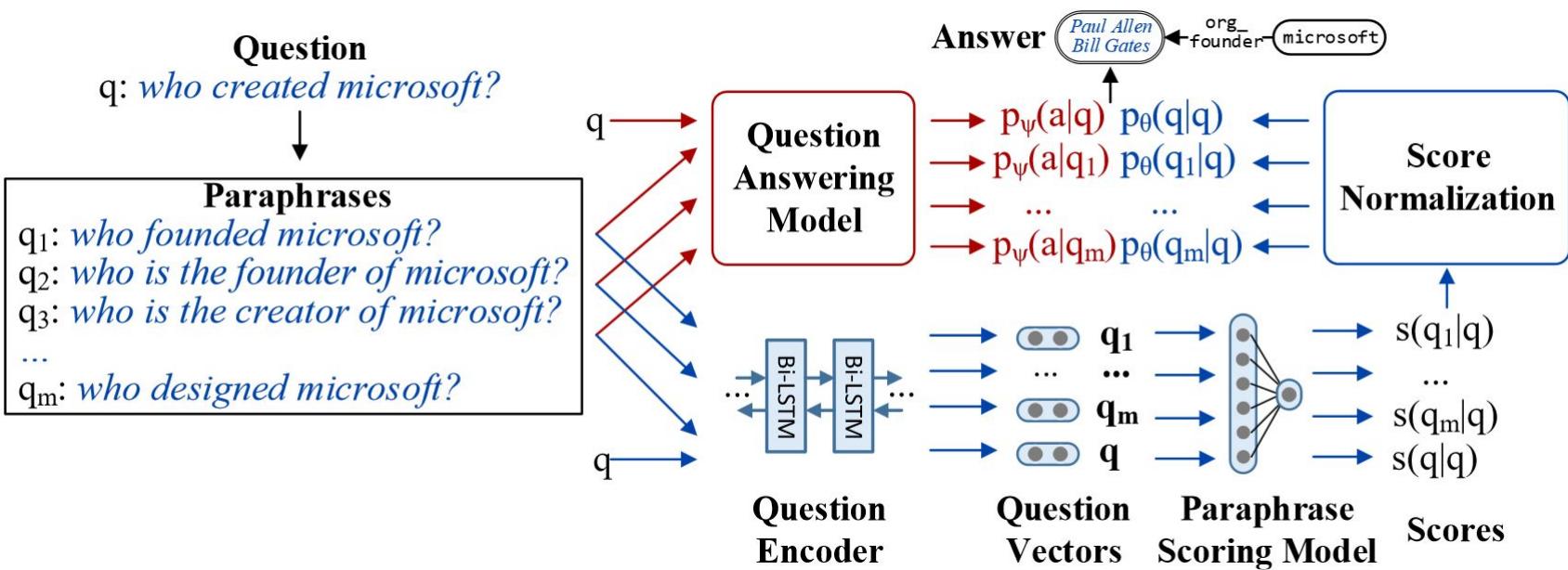
Reduce search space by simple IR techniques

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





Qanary

# Qanary

Several tasks common across QA systems.

Having these tasks as components, sharing a vocabulary can help develop better QA system.

# Qanary

A framework for developing QA system by integrating various component using a pre-defined **Question Answering vocabulary**.

# Observations

- Limited compatibility
- Predefined pipelines
- Limited semantics
- + Interoperable infrastructure
- + Exchangeable components
- + Flexible granularity
- + Isolation of components

# Goals

Easy-to-build QA systems on-top of reusable components

Establish an ecosystem of components for QA systems

# Qanary Methodology and Technical Framework

# Knowledge perspective - I

Requirements of Knowledge perspective

- abstract knowledge representation: qa vocabulary
- align the input/output of the each component in a QA process

# QA vocabulary

Represent all the available knowledge about a question.

- + representation of knowledge about question separated from process
- + includes trust & provenance
- + self-describing, reusable and extensible
- + enables efficient collaboration on a data-level
- + agnostic to question format (text, structured, audio)
- + agnostic to question answering processing steps and implementation

# Aligning I/O in QA processes

- + required input mapped from KB
- + computed output mapped into KB
- + mapping on a logical and sound level

# Knowledge Representation using the qa Vocabulary

# Abstract Knowledge Representation (KR)

Represent all the knowledge about a question using a RDF vocabulary

# KR requirements

- + self-describing, sound knowledge representation
- + represent provenance for (all) information
- + represent trust for (all) information

# Derived Technology stack

- + Resource Description Framework (RDF)
- + Web Annotation Data Model (WADM)
- + Question Answering vocabulary (qa)

# Resource Description Framework (RDF)

[Introduction to RDF](#) (slides by Manolis Koubarakis)

# Web Annotation Data Model (WADM)

- + oa:Annotation <myIRI> a oa:Annotation;
- + oa:hasTarget oa:hasTarget <questionIRI> ;
- + oa:hasBody oa:hasBody <TextSelector> ;
- + oa:annotatedAt oa:annotatedBy <DBpediaSpotlight> ;
- + oa:annotatedAt "...^^xsd:date ;

# QA Vocabulary

Introducing new QA-related concepts on-top of WADM  
[?cite]

```
qa:Question
    rdfs:subClassOf oa:Annotation.
qa:Answer, . . .
qa:Dataset, . . .
qa:AnnotationQuestion, . . .
```

# From KR to methodology

**Conclusion:** Advantages of using an ontology

- + agnostic to question format (text, structured, audio)
- + agnostic to question answering processing steps
- + agnostic to implementation
  - + programming language
  - + component granularity

# From KR to methodology II

## Methodology:

- + abstract knowledge representation
  - + independent representation
- + align the input/output of the each component
  - + on a logical and sound level

# Qanary Methodology

# Qanary Architecture

TODO: Insert the arch diagram.

# It's about the components, stupid. - slide 33

An agile QA framework can only provide common  
features

any particular problem solving/algorithm needs to be  
separated  
from the pipeline

# Component Data Alignment- CDA

**Goal** Establish common ground for the research community

# CDA- II

Two options

<Insert diagrams>

# CDA III

Alignment of input/output of each component with qa, if component provides output using presentation as

## semantic data (RDF)

- logical representation of alignment
  - ontology alignment (OWL, DOL)
  - SPARQL query
- non-semantic data (API, JSON, XML, CSV)
  - SPARQL query

## NER/NED

- DBpedia
- Spotlight
- [?cite] (NIF)
- relation detection
  - PATTY [?cite]
- query construction
  - SINA [?cite]

# CDA: NED

## Create component input

- fetch question URI (from Qanary triplestore)

## Processing

- retrieve textual question representation from URI
- compute named entities within the text

## store component output

- for each named entity:
  - create a oa:TextSelector within the Qanary triplestore containing the positions of the particular Named Entity

# CDA: NED Benefits

Easily replace the NED component.

Measure quality against exchangeable relation detection  
and query construction components.

# CDA: Relation Detection

## Create component input

- fetch question URI (from Qanary triplestore)
- fetch Named Entities which are already available

## Processing

- retrieve textual question representation from URI
- compute relations within the text

## store component output

- for each named relation:
  - create a relation resource within the Qanary triplestore (using a oa:TextSelector to mark the positions)

# CDA: Relation Detection Benefits

Any improvement on the NED component (i.e., replace) will improve the quality here

Measure quality against exchangeable query construction components.

# CDA: Query Construction

## Create component input

- fetch Named Entities (which are already available)
- fetch Relations (which are already available)

## Processing

- compute SPARQL

## Store component output

- for each created SPARQL:
  - store a resource/SPARQL in the Qanary knowledge base

# CDA: Query Construction Benefits

Any improvement on the NED component (i.e., replace) will improve the quality here

Any improvement on the Relation detection component (i.e., replace) will improve the quality here

# Case Study

<Insert diagrams slide 41>

# CDA- II

<Insert diagrams from slide 42>

# Take Away: Qanary methodology

Qanary: knowledge-driven methodology for QA systems

- and reference implementation of methodology, too

- build on-top of the qa vocabulary

# Example of Semantic parsing

An overview/framework of the approach

Some example

Introduce Qanary that is it would be more broadly discussed in the next session

# With different paradigms comes different dataset

2 slides related to datasets

Rough

# End-2-End based approach

Something ; but I have no clue

(copied text )

- No traditional NLP ; Systems essentially know nothing about language
- No intermediate logical structures – answers are generated directly from the question („end-to-end“)
- Current approaches can answer simple questions, i.e. questions involving a particular relation and entity in the knowledge graph

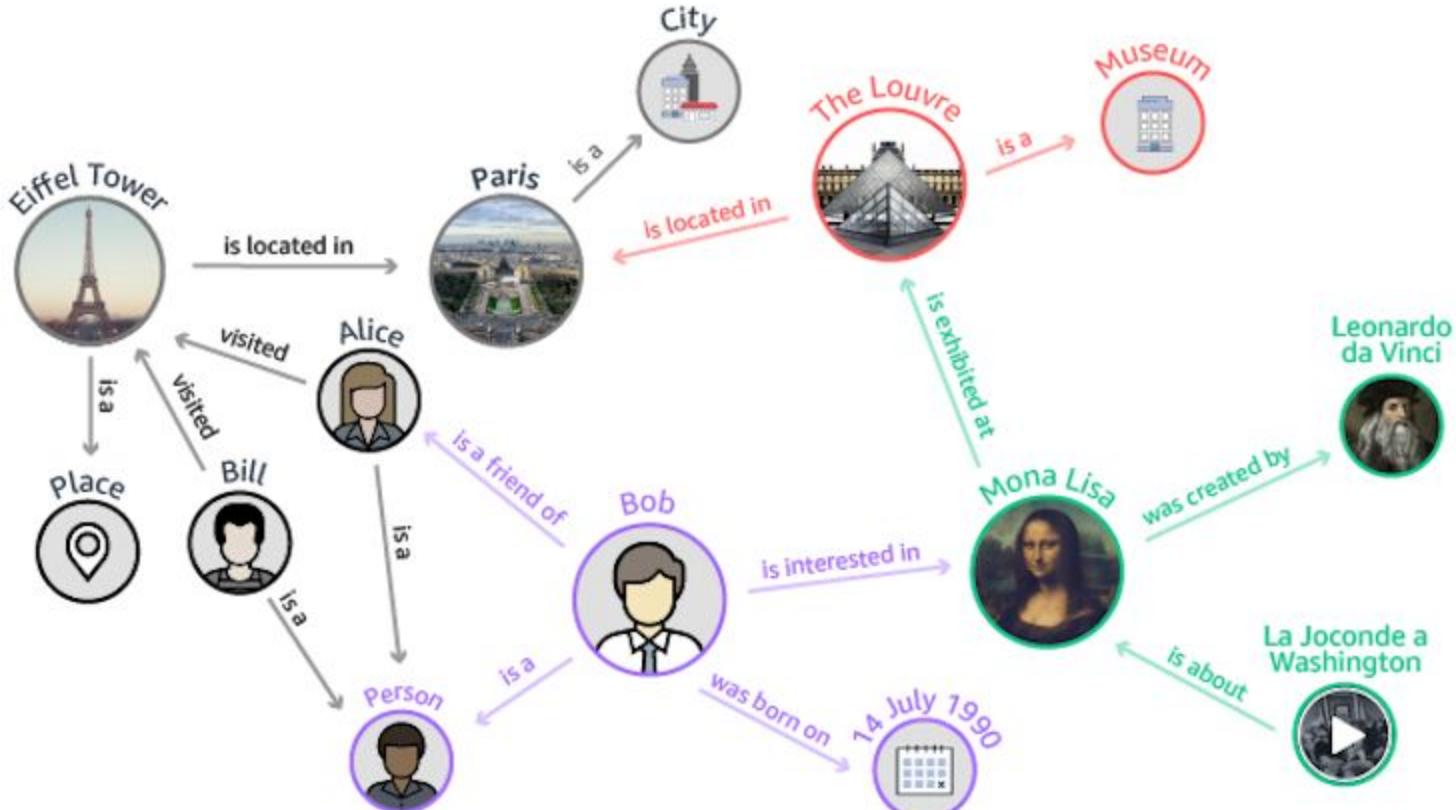
# How to build a **Question Answering** system overnight?



“All NLP/AI task *can be* reduced to  
Question Answering”

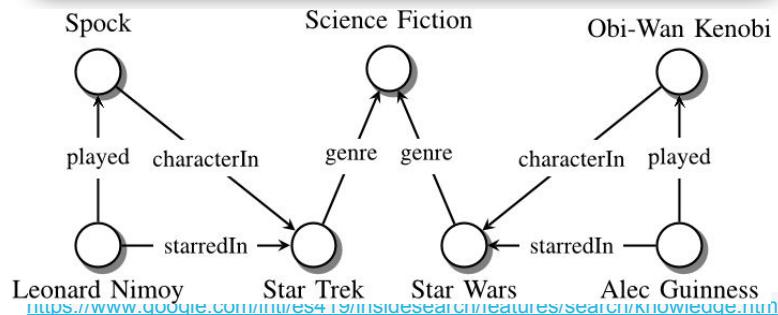
Richard Socher,  
Deep Learning Summit 2016, SF

# Kn



# Knowledge Graphs / Knowledge Graphs

Beispiel: Google Knowledge Graph / Knowledge Vault



- "Things instead of strings"
- "Entity" = Nodes, e.g. persons, places
- Edges = relations, e.g. "Works at"
- Used on Google, Facebook, LinkedIn, Twitter, DBpedia, Wikidata
- Natural and efficient way to map knowledge

# Knowledge Graphs



All Shopping Images Videos News More Settings Tools

Gandalf > Books



[Gandalf | The One Wiki to Rule Them All | Fandom powered by Wikia](#)

[lotr.wikia.com/wiki/Gandalf](https://lotr.wikia.com/wiki/Gandalf)

Gandalf portrayed by Ian McKellen in *The Lord of the Rings* trilogy and *The Hobbit* trilogy .... Ring. Gandalf was immediately suspicious of the Hobbit's story of how he acquired it. .... —*The Lord of the Rings, The Fellowship of the Ring, Book II, Chapter 5:* "The ..... Gandalf the White in Peter Jackson's *The Return of the King* ...

Race: Maia

Weapon: Glamdring, Narya, Wizard Staff

Eye color: Blue-grey

Hair color: Grey, later white

[Gandalf - Wikipedia](#)

<https://en.wikipedia.org/wiki/Gandalf>

As explained in *The Fellowship of the Ring*, Gandalf spent the years between *The Hobbit* and *The Lord of the Rings* travelling Middle-earth in search of ...

Wizard · Three Rings · Gandalf (disambiguation)

[Gandalf - Tolkien Gateway](#)

[tolkiengateway.net/wiki/Gandalf](https://tolkiengateway.net/wiki/Gandalf)

Nov 13, 2016 - Gandalf, for his part, found himself amazed by the hobbit, until then the wise .... where Gandalf read the Book of Records, which revealed the fate of ..... Gandalf in *The Lord of the Rings: The Fellowship of the Ring* (video game) ... Gandalf in *The Lord of the Rings: The Return of the King* (Game Boy Advance) ...

Height: 5'6"

Clothing: Grey robes (later white), blue hat, grey ...

Hair color: Long white, silver beard

Sailed west: 29 September T.A. 3021; Grey Ha...

[Amazon.com: Following Gandalf: Epic Battles and Moral Victory in The ...](#)

<https://www.amazon.com/Following-Gandalf-Battles-Moral-Victory/.../1587430851>



Gandalf

Lord Of The Rings character



Gandalf is a fictional character and one of the protagonists in J. R. R. Tolkien's novels *The Hobbit* and *The Lord of the Rings*. He is a wizard, member of the Istari order, as well as leader of the Fellowship of the Ring and the army of the West. [Wikipedia](#)

Portrayed by: Ian McKellen

Quotations: A wizard is never late, nor is he early. He arrives precisely when he means to., "Fool of a Took!"

Title: Servant of the Secret Fire, Elf-friend, Istar (Wizard), Wielder of the Flame of Anor, Ring-bearer

Species: Maia

Has possessed: Glamdring, Narya

Organizations: White Council, Thorin and Company, The Fellowship of

# Simple 1

List all movie RDJ has been casted in?

```
SELECT ?movies where {  
?movies <http://dbpedia.org/ontology/starring> <http://dbpedia.org/resource/Robert\_Downey\_Jr.>  
}
```

# Simple 2

List all movies from Marvel Studios

```
SELECT ?movies where {  
?movies <http://dbpedia.org/resource/Marvel\_Studios>  
}
```

# Combine

List Marvel movies did RDJ has been casted in?

```
SELECT ?movies where {  
?movies <http://dbpedia.org/ontology/starring> <http://dbpedia.org/resource/Robert\_Downey\_Jr> .  
?movies <http://dbpedia.org/property/studio> <http://dbpedia.org/resource/Marvel\_Studios>  
}
```

# Have a Count

How many Marvel movies did RDJ has been casted in?

```
SELECT COUNT(?movies) where {  
?movies <http://dbpedia.org/ontology/starring> <http://dbpedia.org/resource/Robert\_Downey\_Jr> .  
?movies <http://dbpedia.org/property/studio> <http://dbpedia.org/resource/Marvel\_Studios>  
}
```

# Machine translation problem

Question → Sparql using background knowledge which is knowledge base for our case.

# Formal Language

Intermediary representation of question from which converting to query is easier.

Broadly categorized (for QA purposes) into:

- KB independent logical forms
- KB dependent logical forms

# Paradigms

**Semantic Parsing:** Convert Question to explicit logical form which can retrieve answers.

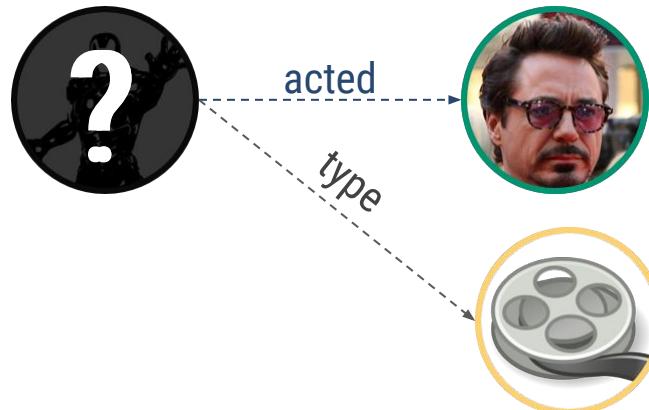
**End-to-End:** Retrieve answers directly, given a question without an explicit logical form.

# Entity Linking

Identify the entity mentioned in the question and link it to the KB.

# Entity Linking

Name all the movies in which Robert Downey Jr Acted?



# Various ways to represent same entity

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?

# Basic framework

| Approach   | Advantage  | Disadvantage  |
|------------|--|---|
| Sequential | <ul style="list-style-type: none"><li>- Reduces candidate search space for Relation Linking</li><li>- Allows schema verification</li></ul>   | <ul style="list-style-type: none"><li>- Relation Linking information cannot be exploited in Entity Linking process</li><li>- Errors in Entity Linking cannot be overcome</li></ul>        |
| Parallel   | <ul style="list-style-type: none"><li>- Lower runtime</li><li>- Re-ranking of Entities possible based on Relation Linking</li></ul>  | <ul style="list-style-type: none"><li>- Entity Linking process cannot use information from Relation Linking process and vice versa</li><li>- Does not allow schema verification</li></ul> |
| Joint      | <ul style="list-style-type: none"><li>- Potentially high accuracy</li><li>- Reduces error propagation</li><li>- Better disambiguation</li><li>- Allows schema verification</li><li>- Allows re-ranking</li></ul> | <ul style="list-style-type: none"><li>- Complexity increase</li><li>- Larger search space</li></ul>   |

# Entity linking

Identify the entity mentioned in the question and link it to knowledgebase

Problems

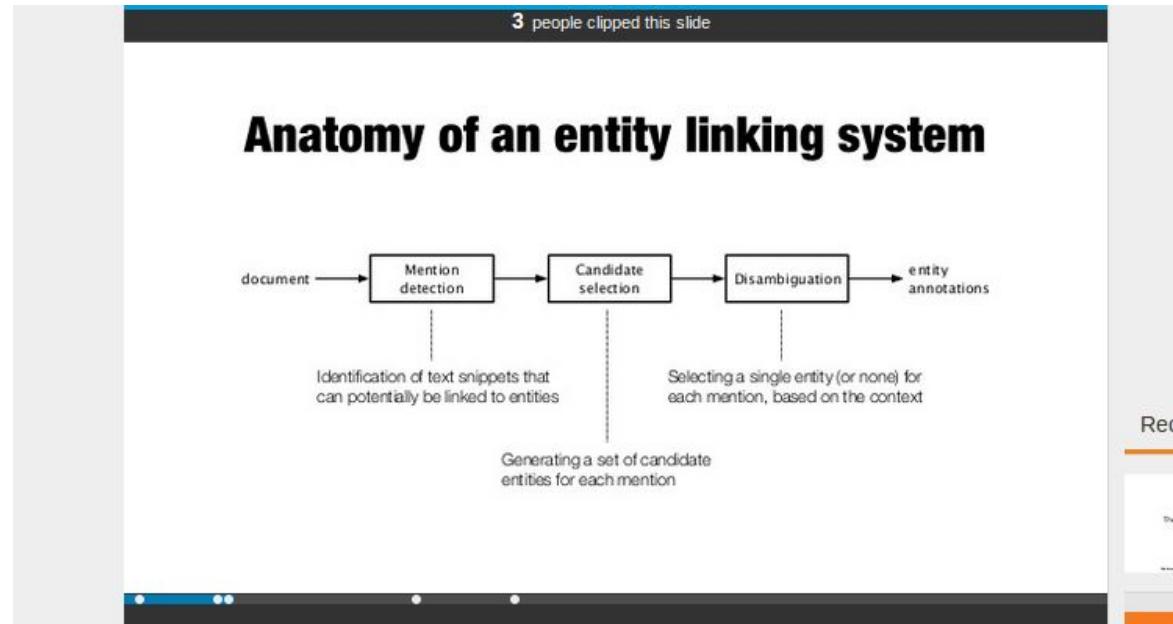
JRR Tolkien → Jrr tolkein

JRR Tolkien → tolkein

Some example of sub-set and super-set

And several other ways to represent

# Found it here



## Document

## Entity Annotation

### Mention Detection

### Candidate Selection

### Disambiguation

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 Iron Man in the Marvel Cinematic Universe, appearing in several films as the lead role.

## Document

## Entity Annotation

### Mention Detection

### Candidate Selection

### Disambiguation

JRR Tolkien, (3 January 1892 – 2 September 1973) was an English writer, poet, philologist, and university professor who is best known as the author of the classic high fantasy works The Hobbit, The Lord of the Rings, and The Silmarillion. ...

# Document

# Entity Annotation

## Mention Detection

## Candidate Selection

## Disambiguation



JRR Tolkien, (3 January 1892 – 2 September 1973) was an English writer, poet, philologist, and university professor who is best known as the author of the classic high fantasy works The Hobbit, The Lord of the Rings, and The Silmarillion. ...

# Document

# Entity Annotation

## Mention Detection

## Candidate Selection

## Disambiguation



JRR Tolkien, (3 January 1892 – 2 September 1973) was an English writer, poet, philologist, and university professor who is best known as the author of the classic high fantasy works The Hobbit, The Lord of the Rings, and The Silmarillion. ...



# DBpedia Spotlight in action



Confidence:  0.5

n-best candidates      Language: English

JRR Tolkien, (3 January 1892 – 2 September 1973) was an [English](#) writer, [poet](#), [philologist](#), and university professor who is best known as the author of the classic [high fantasy](#) works [The Hobbit](#), [The Lord of the Rings](#), and [The Silmarillion](#)

[BACK TO TEXT](#)

# Off the shelf solution

>basic algorithm of spotlight

>SOme other dumb down solution

# TextRazor in action

# Introduction

# How to build a **Question Answering** system overnight?

Jens Lehmann<sup>1,2</sup>, Andreas Both<sup>3</sup>, Ioanna Lytra<sup>1</sup>, Mohnish Dubey<sup>1</sup>, Denis Lukovnikov<sup>1</sup>, Kuldeep Singh<sup>2</sup>,  
Gaurav Maheshwari<sup>1,2</sup>, Priyansh Trivedi<sup>1,2</sup>

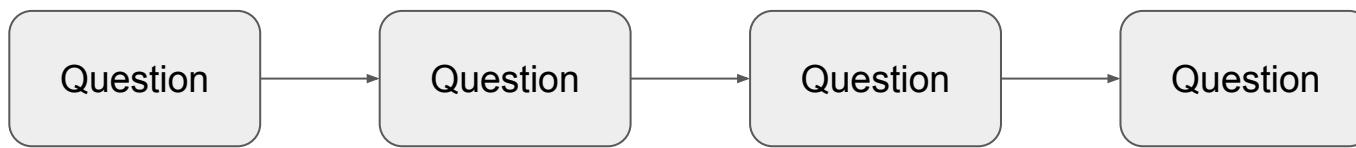
<sup>1</sup> University of Bonn, Bonn, Germany

<sup>2</sup> Fraunhofer IAIS, St. Augustin, Germany

<sup>3</sup> Poopie Pooper, Poopertown

**ESWC 2018, Heraklion**

# Anatomy of Semantic parsing based QA





how to build a

# QUESTION ANSWERING

system overnight?

eswc 2018

- Introduction
  - Team
  - Tutorial flow/index
    - Outline
    - A line about each session and the speaker
- Break down the name "QA" + "Knowledge graph"
- QA
  - Some quote(Socher said something about reducing every NLP task to that of QA)
  - Give an example
  - Describe various kind of QA
    - Visual QA
    - Table QA
    - Unstructured QA
    - QA over KG
- Knowledge graph
  - Give an example
  - RDF
  - Example of different kind of KG
- QA over KG (Formally/intuitively)
  - Concrete problem description
- Developing a framework
  - Think about a nice question
  - Entity linking
    - Introducing the problem and the basic solutions
  - Relation linking
  - Adding constraints
  - Putting everything together
- Now with concrete problems at hand. General paradigms
  - Semantic parsing based approach
    - ASKNOW (Just an intuitive explanation)- reperesents without KB
    - STAG (same as above)- with KB
  - End-to-end approach
    - ??
- With paradigms comes dataset (or is it other -way around (insert the diosurous meme))
  - LC-QaLD
  - Web-question
  - Some other Interesting Dataset (flight one, sequential QA?)
- Various ways of representing queries ??