



Build a **Question Answering** system overnight

Jens Lehmann, Gaurav Maheshwari, Priyansh Trivedi, Mohnish Dubey, Denis
Lukovnikov,

Outline

Introduction to QA

QA Approaches Overview

Neural Ranking Approach

Hands on



**Who is the father of
Luke Skywalker?**



*What is the best DESSERT
in the world?*

**Name
one ex
president
of United States?**



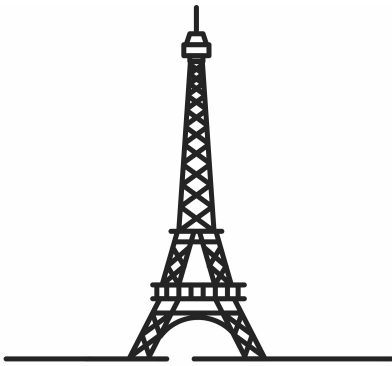
**Where is
ESWC 2018 held?**



Introduction

All NLP/AI task can be reduced to Question Answering

Richard Socher,
Deep Learning Summit 2016, SF

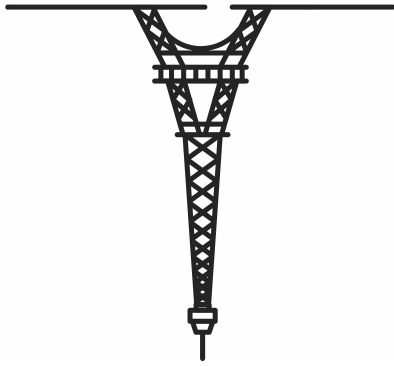


Factoid Questions

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 soooooo 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 Venice to Portoroz?

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

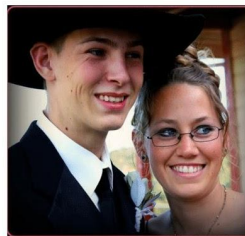
What is the polish word for wreaths?

Visual Question Answering

Who is wearing glasses?

man

woman



Where is the child sitting?

fridge

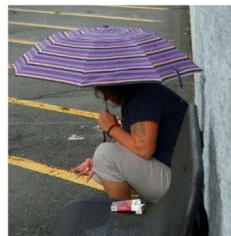
arms



Is the umbrella upside down?

yes

no



How many children are in the bed?

2

1



QA over Knowledge Bases

Given:

- a natural language question

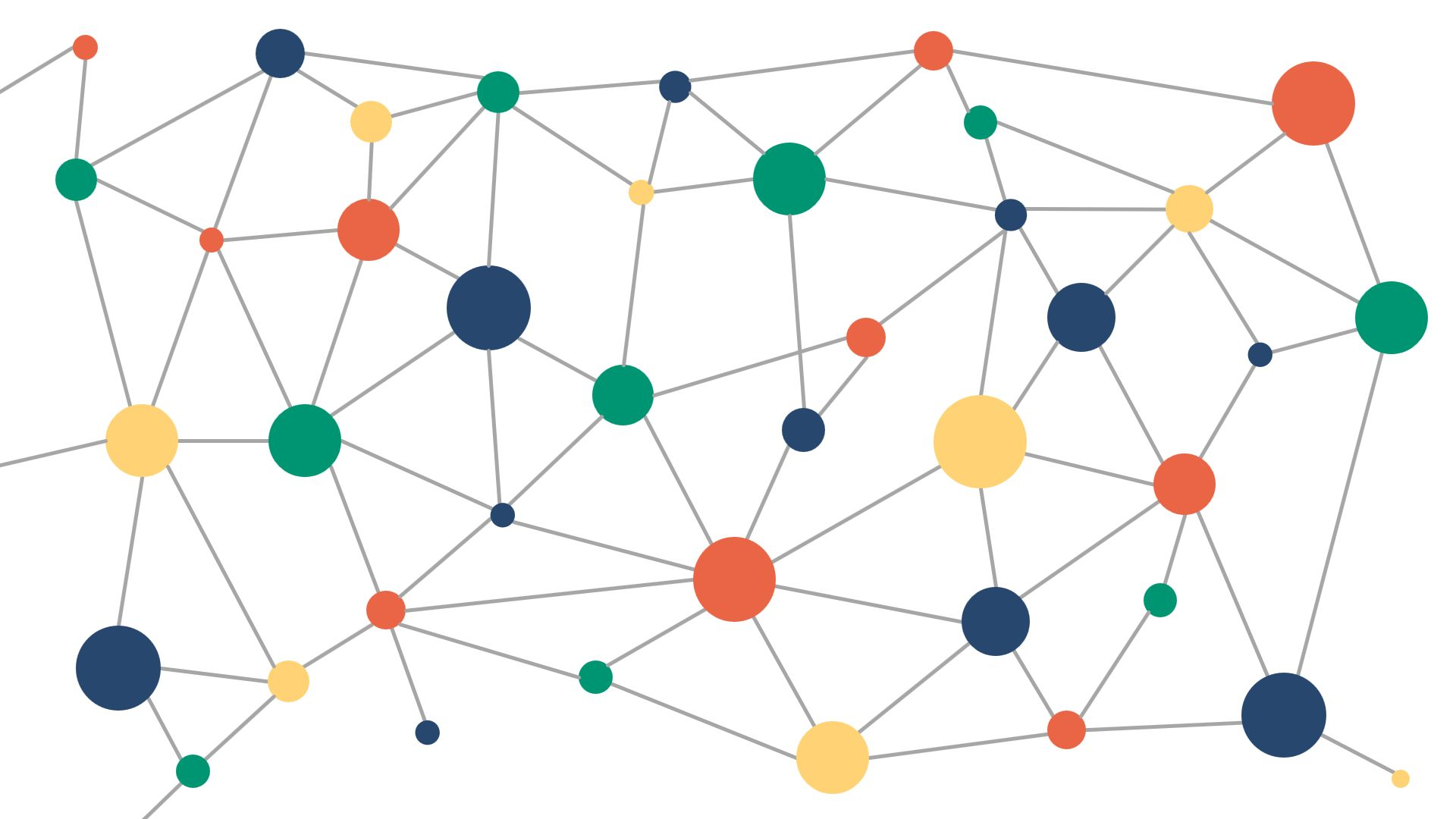
- a source knowledge base

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

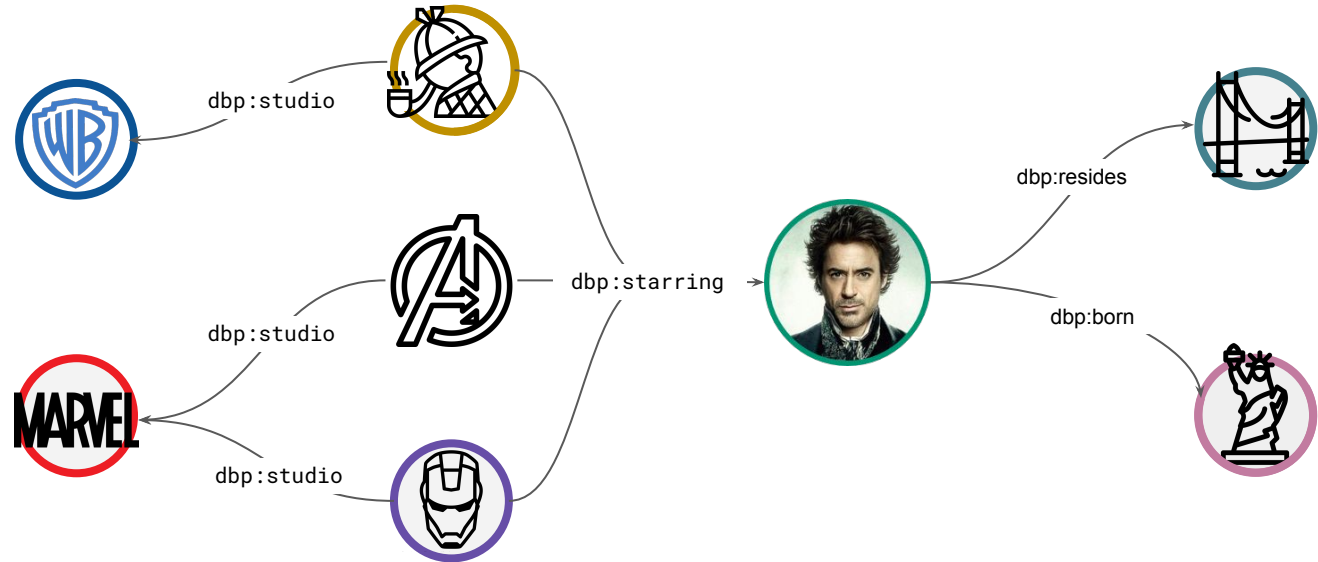
A background network graph with various sized nodes (circles) in shades of gray and white, connected by thin gray lines. The nodes are distributed across the slide, with some clusters and some isolated nodes.

Knowledge Graphs

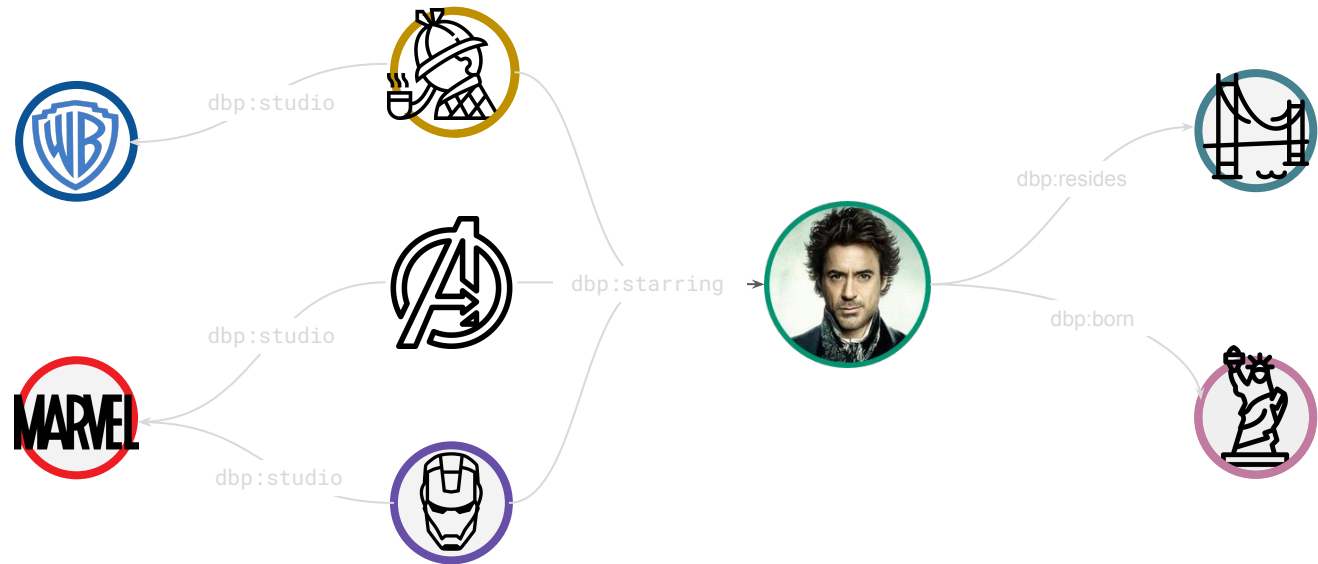
"Things instead of strings"



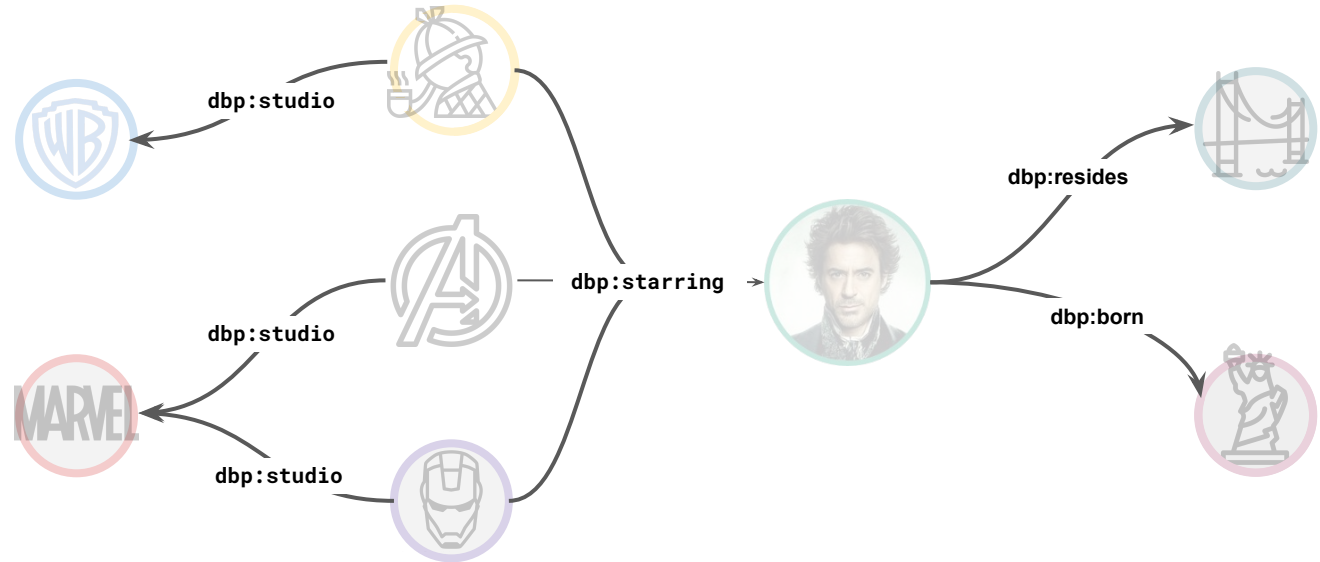
Knowledge Graph (KG)



KG Entities



KG Predicates



KGs are large



DBpedia[1]: < 6.0 million entities (2016-04 edition).



Freebase[2]: 1.9B triples. Depreciated.



Wikidata[3]: 47M entities.

KGs are large **and expressive**

Need **formal query language** to manipulate or extract information from KG.

Examples SQL, Datalog

Formal Query Languages

Formal Query Languages

Structure (Grammar)

Compositional Semantics

Example:

Natural Language
(not formal)

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

Example:

Natural Language
(not formal)

λ -DCS

```
dbp:creators(dbr:Iron_Man, ?answer)
```

Example:

Natural Language
(not formal)

λ -DCS

SPARQL

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



SPARQL 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?

```
SELECT ?uri WHERE {  
    ?uri dbp:studio dbr:Warner_Bros .  
    ?uri rdf:type dbo:Film .  
}
```

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 (formal) queries.

Semantic parsing?

Question Answering

Natural Language Variations

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?

Entity Linking (EL)



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 **Robert Downey Jr** ?

List all the movies starring **Robert Downey Junior**?

RDJ has acted in which movies?

EL - Disambiguating Entities



Who is the CEO of **Apple**?



Apple belongs to which genus?

movie character



Downey played **Iron Man** in which year?

Who is the alter ego of **Iron man**?

comic character

Relation Linking (RL)

Name all the movies in which **Robert Downey Jr** **Acted**? ^{dbo:starring}

Which movies have **RDJ**?

Flicks where I can see **Robert DJ**?

Find me all the films **casting** **Robert Downey Jr** ?

List all the movies **starring** **Robert Downey Junior**?

RDJ **has acted** in which movies?

RL - Implicit Predicates

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

Which movies have **RDJ**?

Flicks where I can see **Robert DJ**?

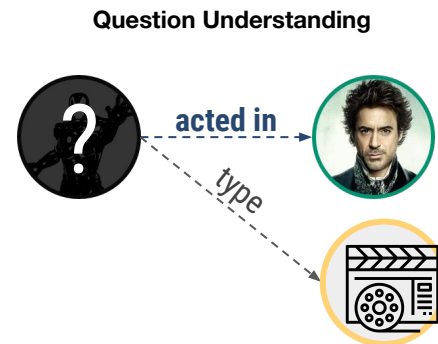
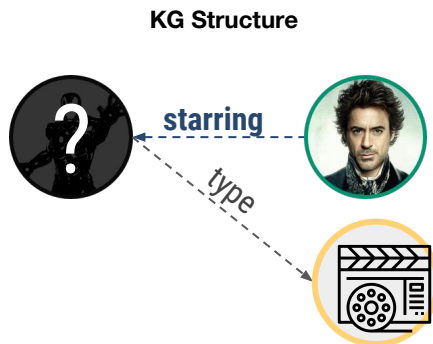
Find me all the films **casting** **Robert Downey Jr** ?

List all the movies **starring** **Robert Downey Junior**?

RDJ **has acted** in which movies?

KG Structure Mismatch

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



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"))  
}
```

Challenges

Entity Linking

Predicate Linking

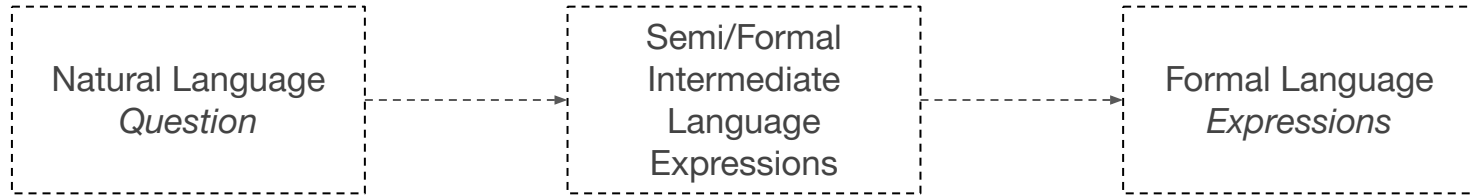
KG structure mis-match

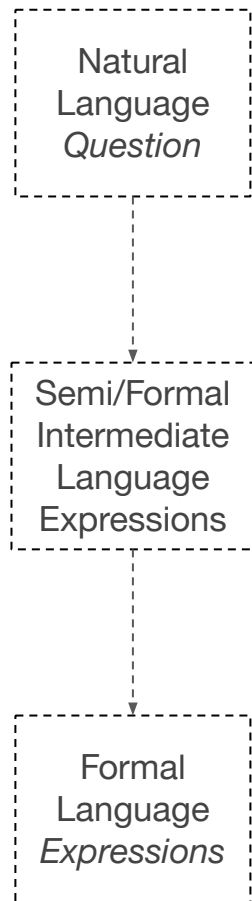
Auxiliary Constraints

Solution



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 [1] NQS, XSER [2],DCS expressions.

[1] Dubey, Mohnish, et al. "Asknow: A framework for natural language query formalization in sparql." International Semantic Web Conference. Springer, Cham, 2016.

[2] K. Xu, et al, "Answering natural language questions via phrasal semantic parsing," in Natural Language Processing and Chinese Computing, pp. 333-344, Springer, 2014.



Ex: AskNow NQS

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

Query_input = Robert Downey Jr.

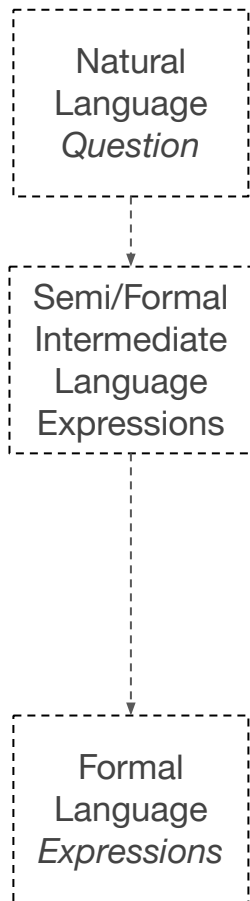
Query_desire = acted

Query_type = List

* rdf:type constraint omitted for brevity's sake

Parses and KB structure

The parse *ignores* the underlying KB structure

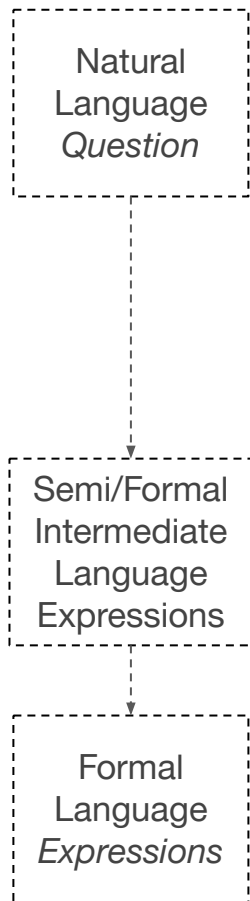


Parses and KB structure



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

KB Dependent Expressions

Specifically

- Creates a set of formal query (like SPARQL) candidates
- Rank them with respect to Question

Problems

But there could be

- **Millions** of SPARQL
- SPARQL are difficult to **represent**.

Millions of SPARQL

Only choose SPARQL which have **all the entities** mentioned in the question a.k.a **topic entities**.

Achieved by entity linking system

Restrict the number of predicates in SPARQL to just few.


Millions of SPARQL

Give me a list of everything where Robert Downey Jr Acted?

Millions of SPARQL

Give me a list of everything where **Robert Downey Jr** Acted?

Topic Entity



Millions of SPARQL

Give me a list of everything where **Robert Downey Jr** Acted?

- ✓ SELECT ?uri WHERE {**dbr:Robert_Downey_Jr** dbp:lives ?uri.}
- ✓ SELECT ?uri WHERE {**dbr:Robert_Downey_Jr** dbp:resides ?uri.}
- ✓ SELECT ?uri WHERE {?uri dbp:studio ?x.
 ?uri dbp:starring **dbr:Robert_Downey_Jr**.}
- ✗ SELECT ?uri WHERE {dbr:Iron_Man dbp:studio ?uri.}
- ✗ SELECT ?uri WHERE {?uri dbp:studio dbr:Warner_Bros.}

SPARQL is difficult to represent

```
SELECT ?uri WHERE {  
    ?uri dbp:studio dbr:Marvel_Studios.  
    ?uri dbp:starring dbr:Robert_Downey_Jr  
}
```

SPARQL is difficult to represent

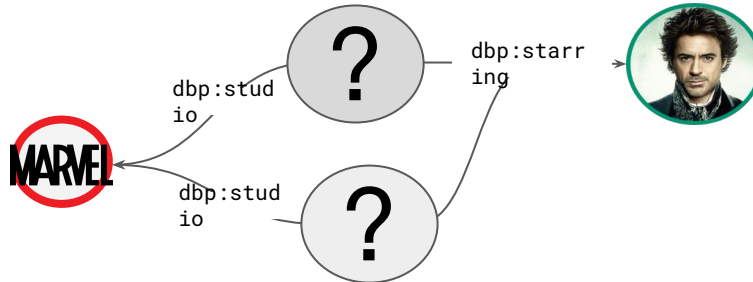
?uri dbp:studio dbr:Marvel_Studios.

?uri dbp:starring dbr:Robert_Downey_Jr

SPARQL is difficult to represent

?uri dbp:studio dbr:Marvel_Studios.

?uri dbp:starring dbr:Robert_Downey_Jr



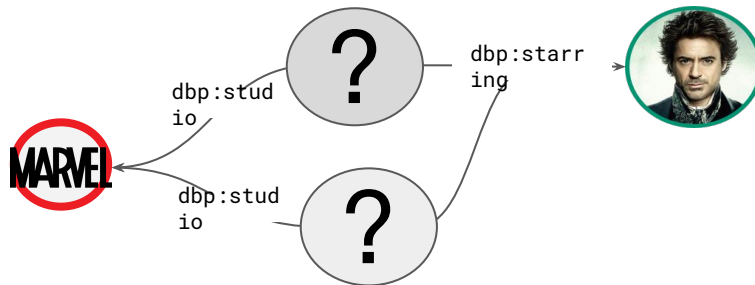
SPARQL is difficult to represent

The where clause of SPARQL represents a path(sub graph) and thus could be **linearized** a.k.a core chain

Core chain

?uri dbp:studio dbr:Marvel_Studios.

?uri dbp:starring dbr:Robert_Downey_Jr



dbr:Marvel_Studios - dbp:studio + dbp:starring - dbr:Robert_Downey_Jr

Core chain

?uri **dbp:studio** **dbr:Marvel_Studios.**

?uri **dbp:starring** **dbr:Robert_Downey_Jr**

dbr:Marvel_Studios - **dbp:studio** + **dbp:starring** - **dbr:Robert_Downey_Jr**

Core chain

`dbp:Marvel_Studios` - `dbr:studio` + `dbo:starring` - `dbr:Robert_Downey_Jr`

Core chain- dropping entities

- `dbr:studio` + `dbo:starring`

Problems

- **Millions** of SPARQL
- SPARQL are difficult to **represent**.

Solutions

- Millions of SPARQL - Use **topic entity** and **limited** number of predicate
- SPARQL is difficult to represent - linearize it as **core chain**

Approach

Overview

Give me a list of everything where Robert Downey Jr Acted?

Overview

Given: Question, Topic
Entity.

Give me a list of everything where **Robert Downey Jr** Acted?

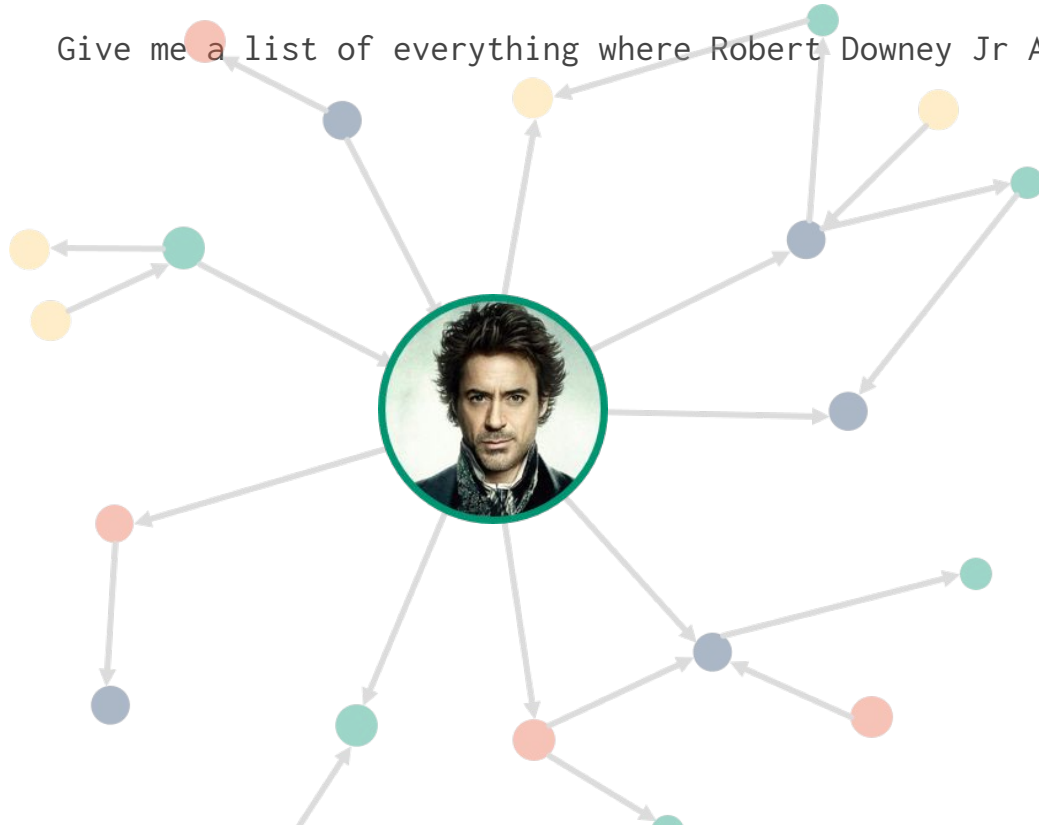


Overview

Given: Question, Topic Entity.

Collect 2-hop subgraph around it.

Give me a list of everything where Robert Downey Jr Acted?



Overview

Given: Question, Topic
Entity.

Collect 2-hop subgraph
around it.

Generate core-chain
candidates

Give me a list of everything where Robert Downey Jr Acted?

dbr:Robert_Downey_Jr + dbp:birthplace

dbr:Robert_Downey_Jr + dbp:parent

dbr:Robert_Downey_Jr + dbp:spouse - dbp:foundedBy

dbr:Robert_Downey_Jr - dbp:starring

dbr:Robert_Downey_Jr - dbp:starring + dbp:director

...

Overview

Given: Question, Topic
Entity.

Give me a list of everything where Robert Downey Jr Acted?

Collect 2-hop subgraph
around it.

0.10 dbr:Robert_Downey_Jr + dbp:birthplace

0.23 dbr:Robert_Downey_Jr + dbp:parent

Generate core-chain
candidates

0.04 dbr:Robert_Downey_Jr + dbp:spouse - dbp:foundedBy

0.73 dbr:Robert_Downey_Jr - dbp:starring

0.41 dbr:Robert_Downey_Jr - dbp:starring + dbp:director

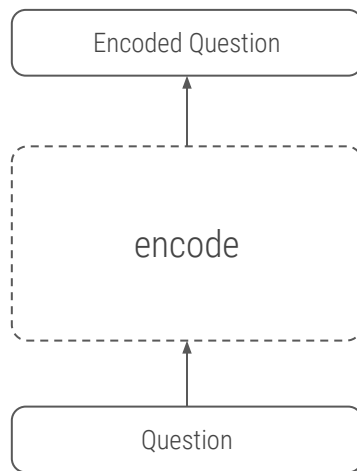
Rank Candidates based
on similarity with
questions

...

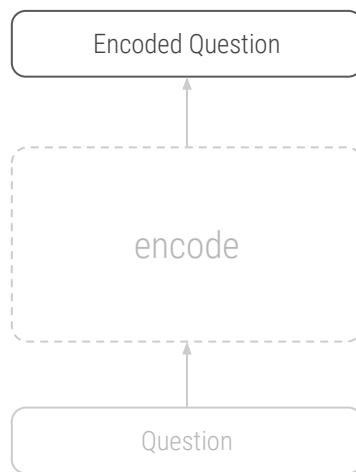
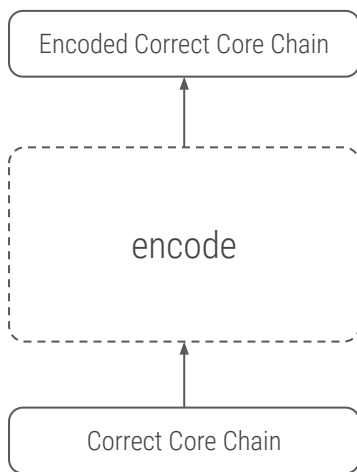
Ranking Framework

Encode core chain and question to a **vector space** such that **distance** correct core chain and the question lies **close** to one another.

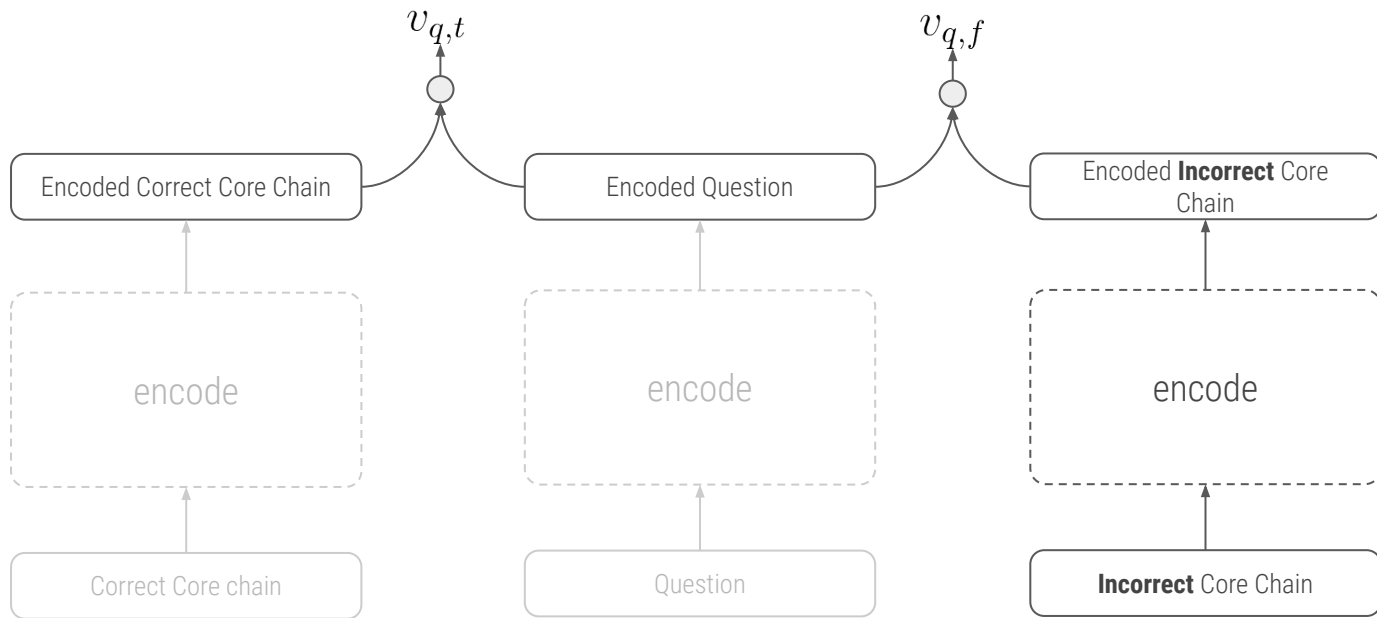
Ranking Framework



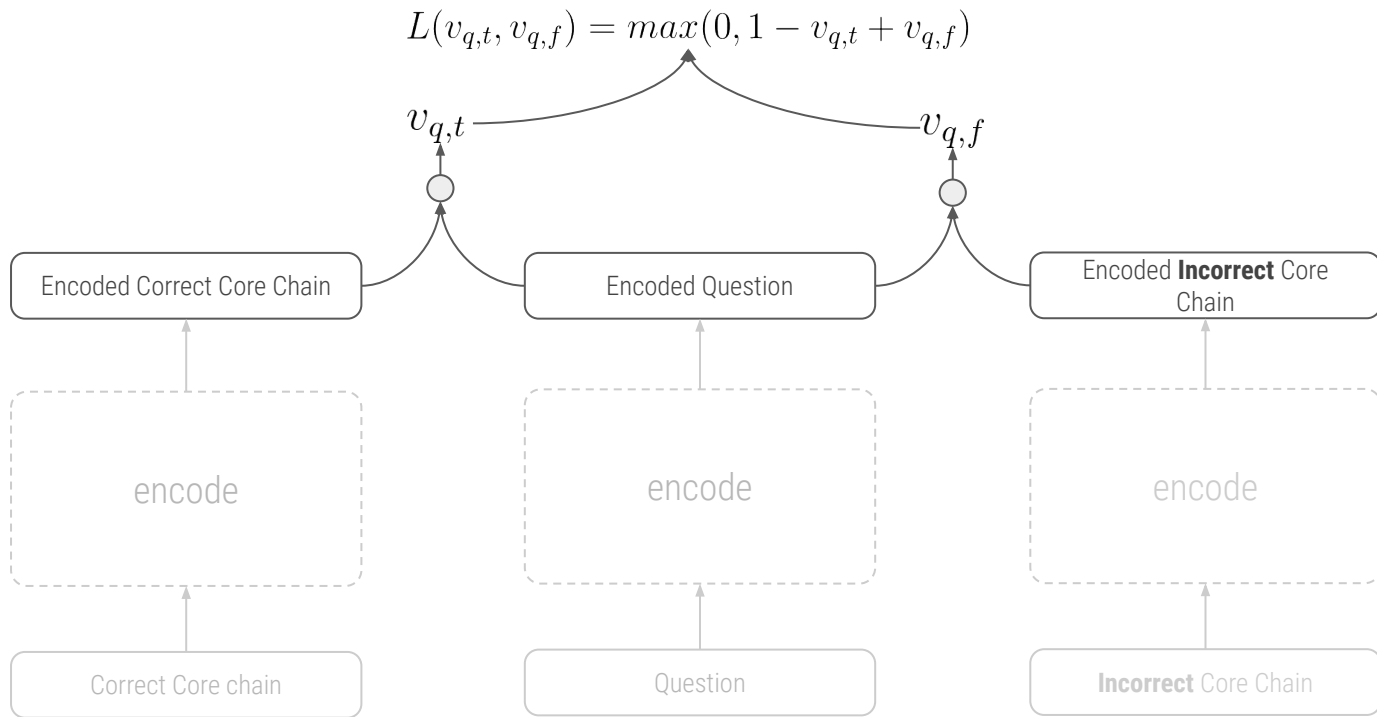
Ranking Framework



Ranking Framework



Ranking Framework



Setup and Session 2

- Word embedding based method
- Encoder based ranking framework

Setup

Setup

Recommended steps for setup

1. Get Anaconda: <https://conda.io/miniconda.html>
 - a. Run the miniconda .sh file
 - b. Create environment: `>> conda create -n pytorch python=2.7 pip ipython jupyter`
 - c. Activate environment: `>> source activate pytorch`
2. Get PyTorch:
 - a. Go to <https://pytorch.org> and follow instructions for your OS
3. Git clone <https://github.com/AskNowQA/QA-Tutorial>

Setup with poor internet

1. Get Anaconda: <https://conda.io/miniconda.html>
 - a. Run the miniconda .sh file
 - b. Create environment: `>> conda create -n pytorch python=2.7 pip ipython jupyter`
 - c. Activate environment: `>> source activate pytorch`
2. Get repo from stick
3. Go to /session4/pkgs and run `install_deps.sh`

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 (Icons in KG)

Sherlock holmes by Matthew Davis from the Noun Project

Empire State Building by Jake Dunham from the Noun Project

Golden gate bridge by icon 54 from the Noun Project

Statue of Liberty by Berkah Icon from the Noun Project

MARVEL is from flaticon

Sherlock Holmes is from FlatIcon

Iron Man by Tatyana Kyul from the Noun Project

Ferguson by priyanka from the Noun Project

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.

K. Xu, et al, **"Answering natural language questions via phrasal semantic parsing,"** in Natural Language Processing and Chinese Computing, pp. 333-344, Springer, 2014

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).

Dubey, Mohnish, et al. "**Asknow: A framework for natural language query formalization in sparql.**" International Semantic Web Conference. Springer, Cham, 2016.