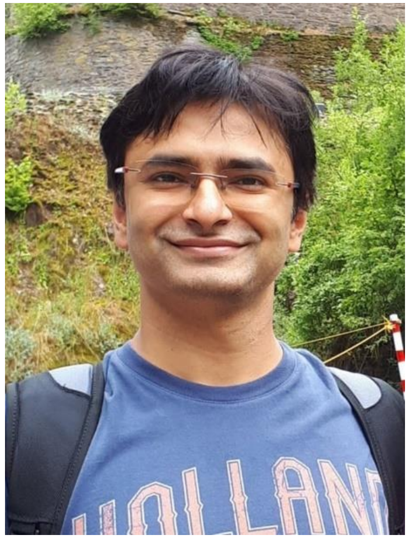
Question Answering over Knowledge Graphs

Rishiraj Saha Roy
Max Planck Institute for Informatics, Germany

Introduction to Seminar on Selected Topics in Question Answering

Saarland University, Winter Semester 2020/21

About us



Rishiraj Saha Roy

Senior Researcher

Max Planck Institute for Informatics

Saarbrücken, Germany

http://people.mpi-inf.mpg.de/~rsaharo/

Magdalena Kaiser

Graduate Student

Max Planck Institute for Informatics

Saarbrücken, Germany

http://people.mpi-inf.mpg.de/~mkaiser/

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

Representative methods from each task

Families of algorithms to build up repertoire for approaching KG-QA

Focus on methods (and not evaluation)

Understand how to go from question to answer

Prerequisites:

Basic IR, NLP, ML, DB

Understanding of core neural techniques

Interactivity ☺

Methodology

Focus on a few instantiations for each method

- Templates
- Graph embeddings
- Subgraph computations
- Belief propagation
- Graph traversal
- Sequence-to-sequence models

QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

Representative methods from each task

Families of algorithms to build up repertoire for approaching KG-QA

Focus on methods (and not evaluation)

Understand how to go from question to answer

What is question answering over knowledge graphs all about?

Question Answering: Vital for Search

What are some films directed by Nolan?

Google Assistant

Hey Siri

Hi. I'm Cortana.
Ask me a question!

amazon alexa

Question Answering: Vital for Search

What are some films directed by Nolan?

Google Assistant

Hey Siri

Hi. I'm Cortana.
Ask me a question!

amazon alexa

Christopher Nolan / Films directed

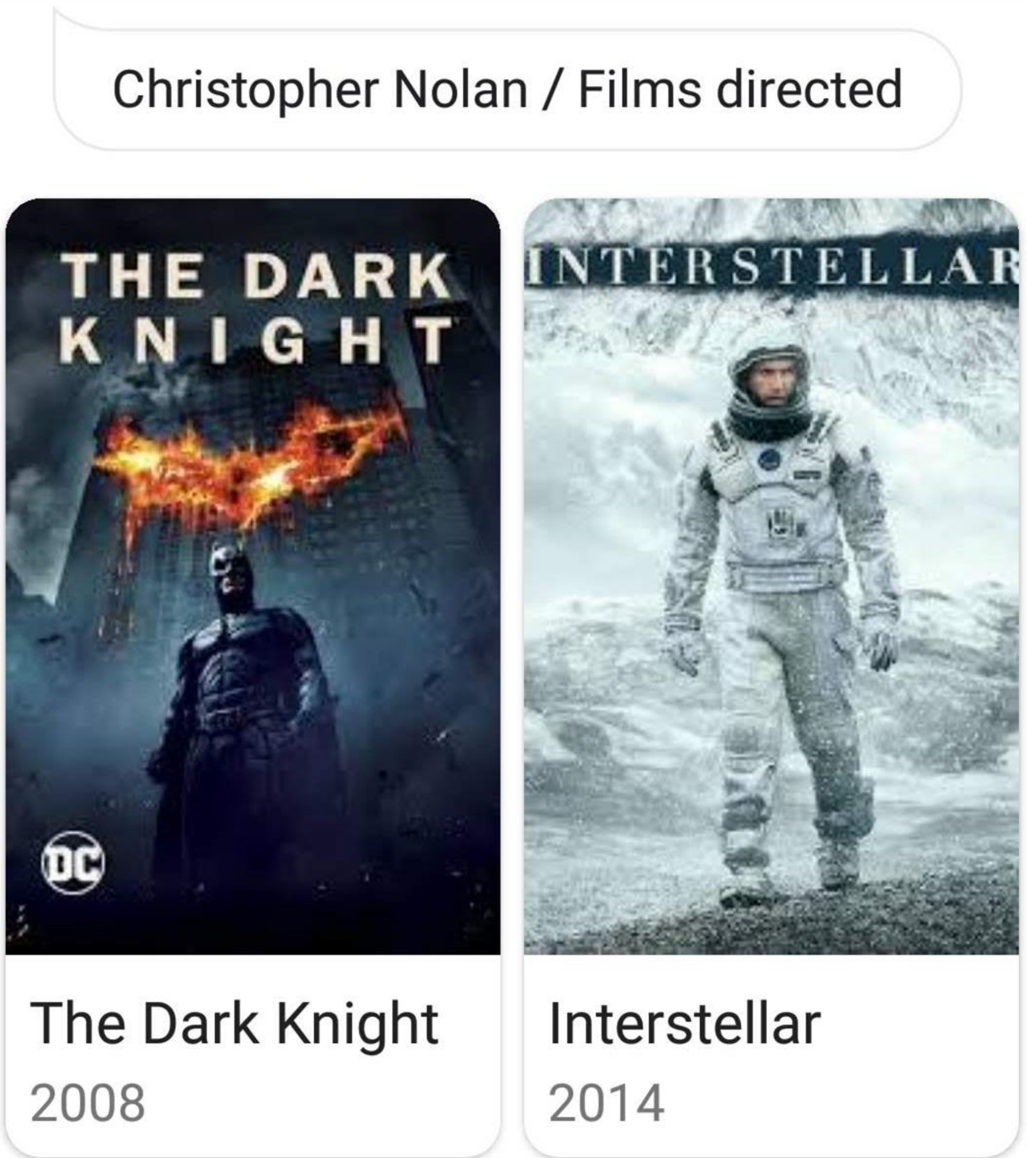
The Dark Knight
2008

Interstellar
2014

Question Answering: Vital for Search

What are some films directed by Nolan?

- Direct answers to questions
- Enabled by knowledge graphs
- Saves time and effort
- Natural in voice UI



QA over KGs

what is the capital of Germany

which club does Messi play for

LIONEL MESSI / CURRENT TEAM

FC Barcelona

what is the currency of India

what is the population of USA

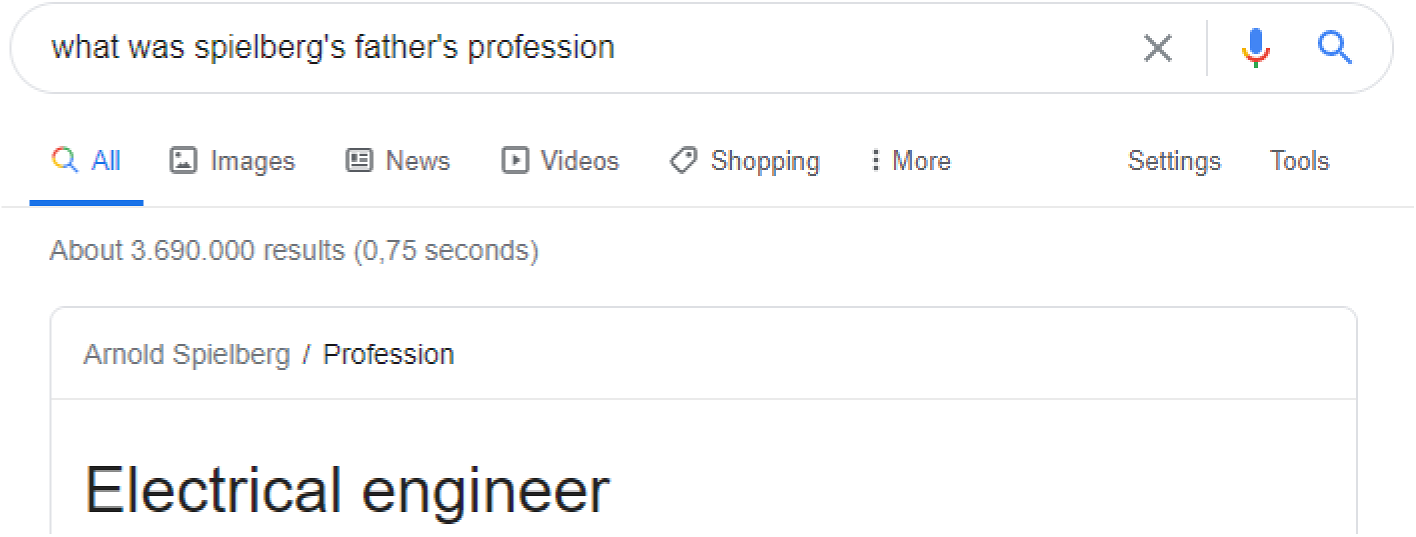
United States / Population

328.2 million (2019)

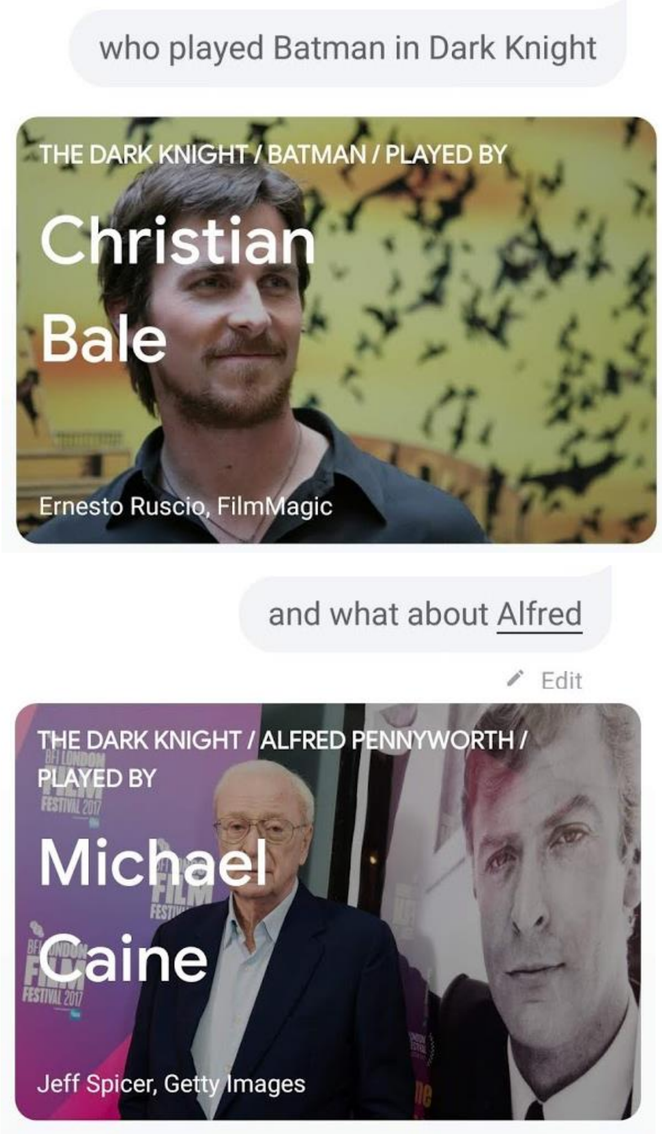
where was Albert Einstein born

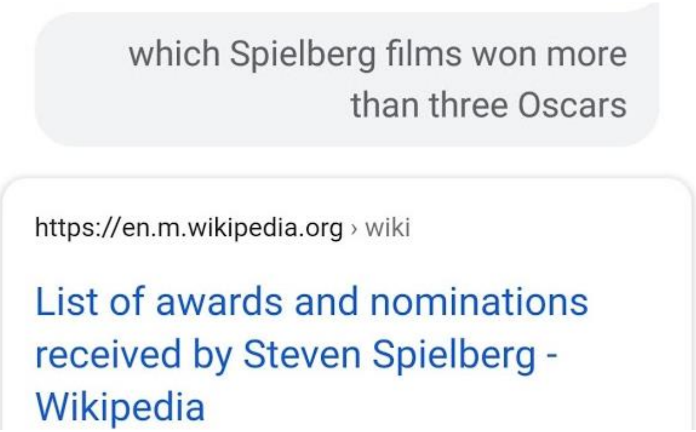
Simple questions involving one entity and relation

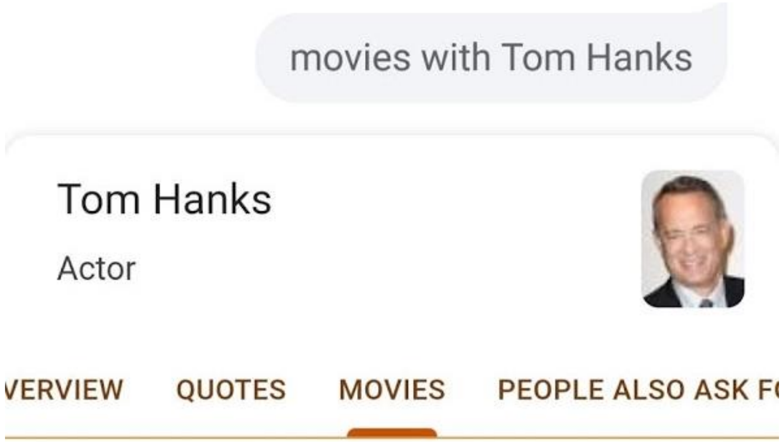
Complex questions involving multiple entities and relations

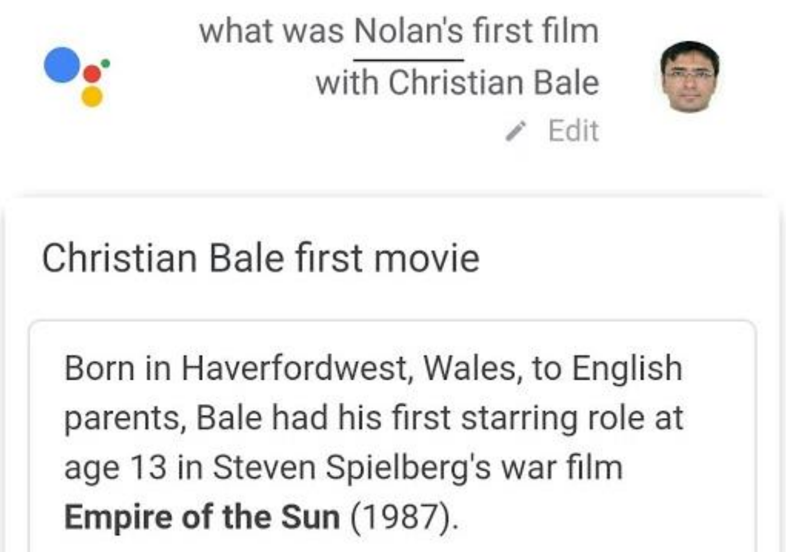


Conversational questions with implicit context

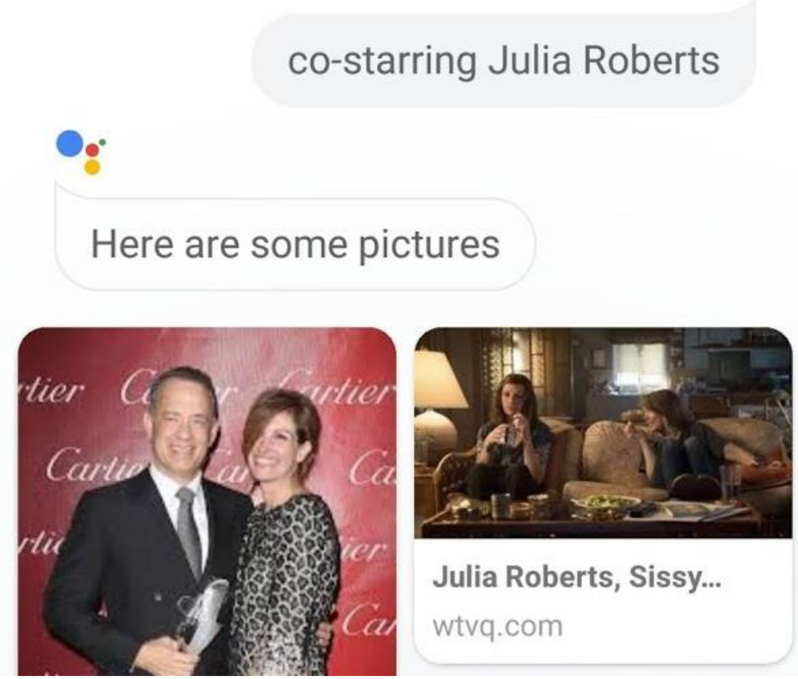








Play with QA: Try out different formulations, entities, domains, complexities, assistants, sources, languages…. to expose brittleness of SoTA and take community forward!



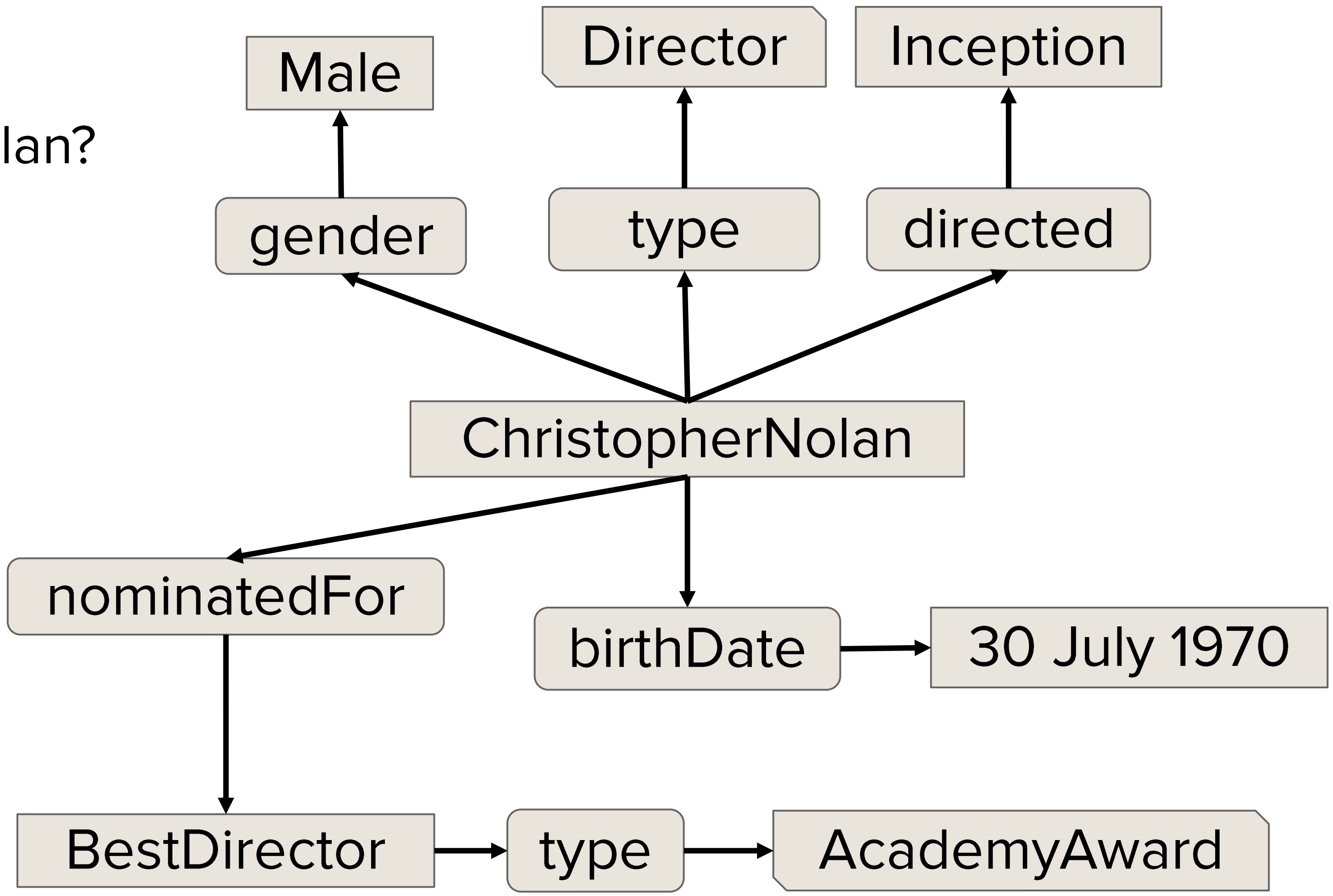
Significant progress has been made on **knowledge base contruction** over the last fifteen years or so; but for **question answering**, which is one of the most valuable applications of KBs, we are still at the **tip of iceberg**!

QA over knowledge graphs (KG-QA)

What are the Oscar nominations of Nolan?

QA over knowledge graphs (KG-QA)

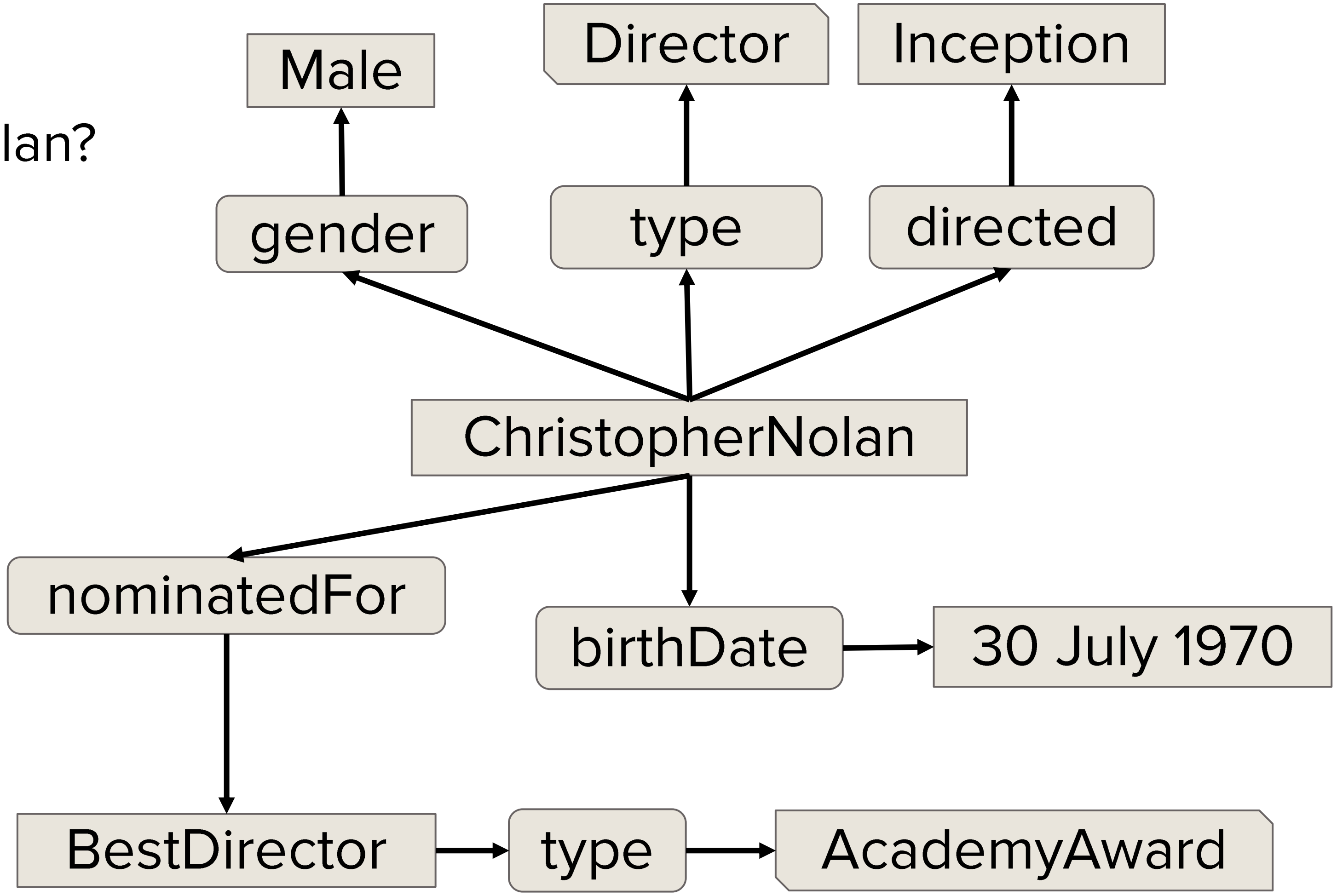
What are the Oscar nominations of Nolan?



QA over knowledge graphs (KG-QA)

What are the Oscar nominations of Nolan?

- YAGO [Suchanek et al. 2007]
- DBpedia [Auer et al. 2007]
- Freebase [Bollacker et al. 2008]
- Wikidata [Vrandečić and Krötzsch 2014]



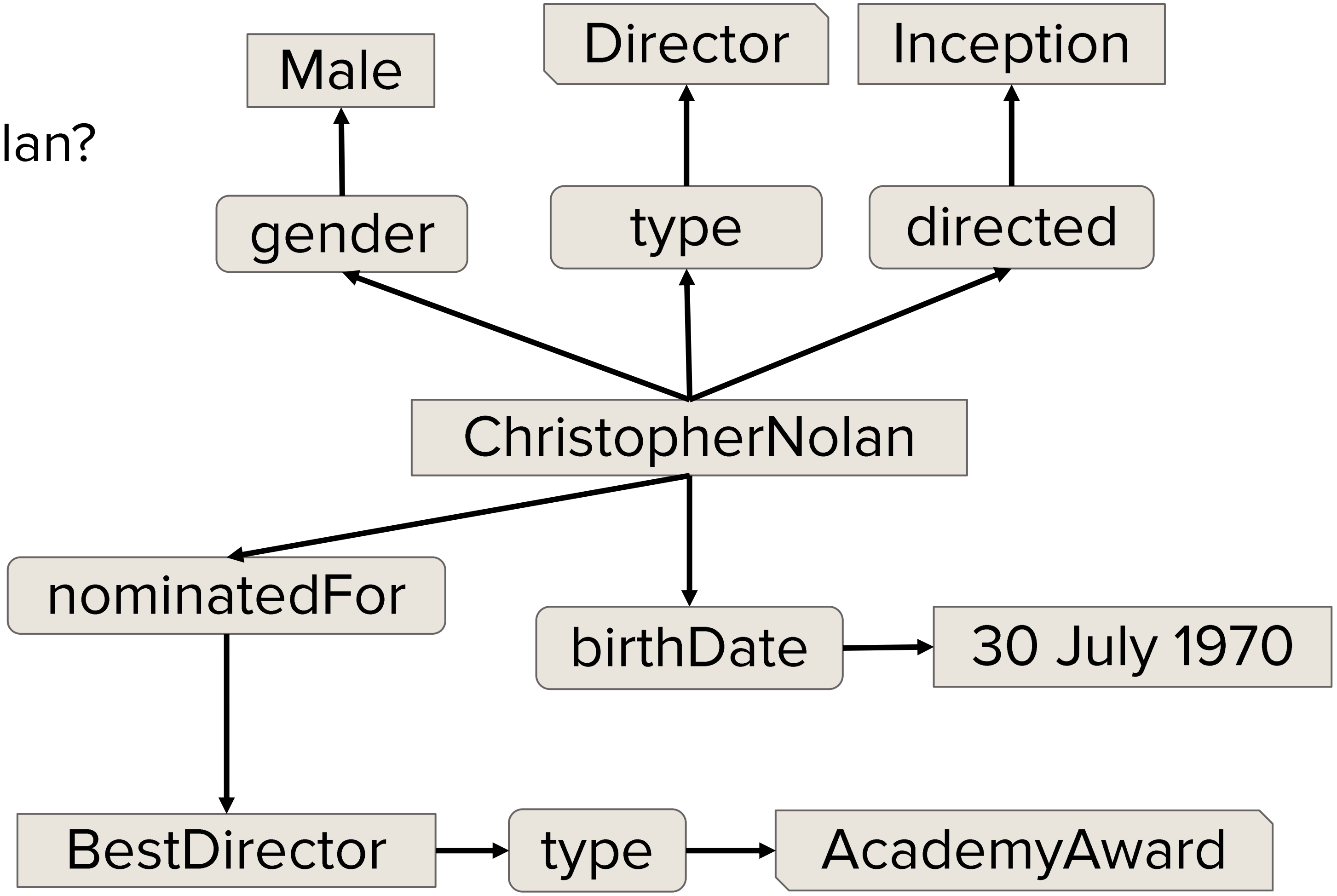
QA over knowledge graphs (KG-QA)

What are the Oscar nominations of Nolan?

- YAGO [Suchanek et al. 2007]
- DBpedia [Auer et al. 2007]
- Freebase [Bollacker et al. 2008]
- Wikidata [Vrandečić and Krötzsch 2014]

Terminology varies across KGs

Here: Entities, predicates, types, literals

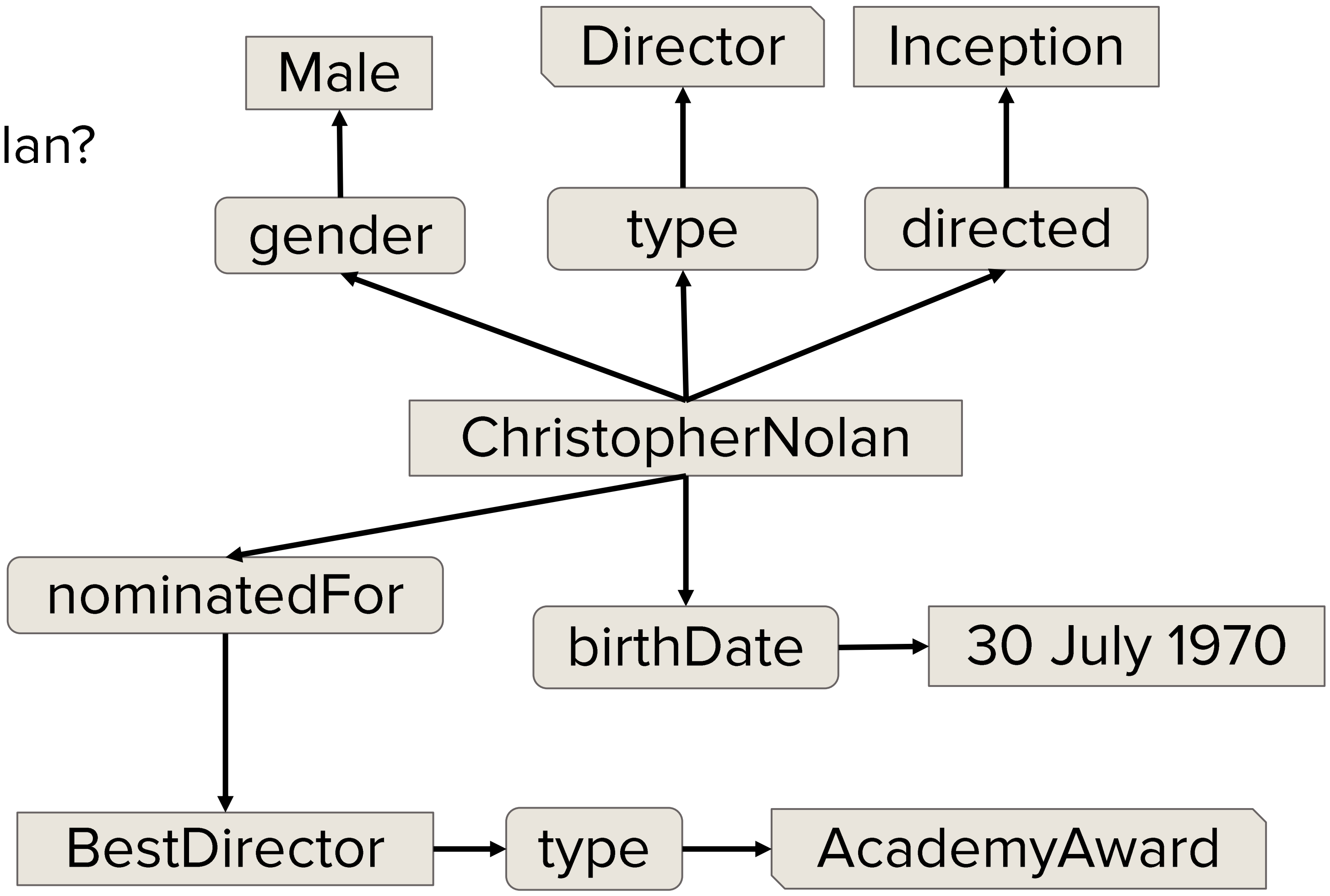


QA over knowledge graphs (KG-QA)

What are the Oscar nominations of Nolan?

- YAGO [Suchanek et al. 2007]
- DBpedia [Auer et al. 2007]
- Freebase [Bollacker et al. 2008]
- Wikidata [Vrandečić and Krötzsch 2014]

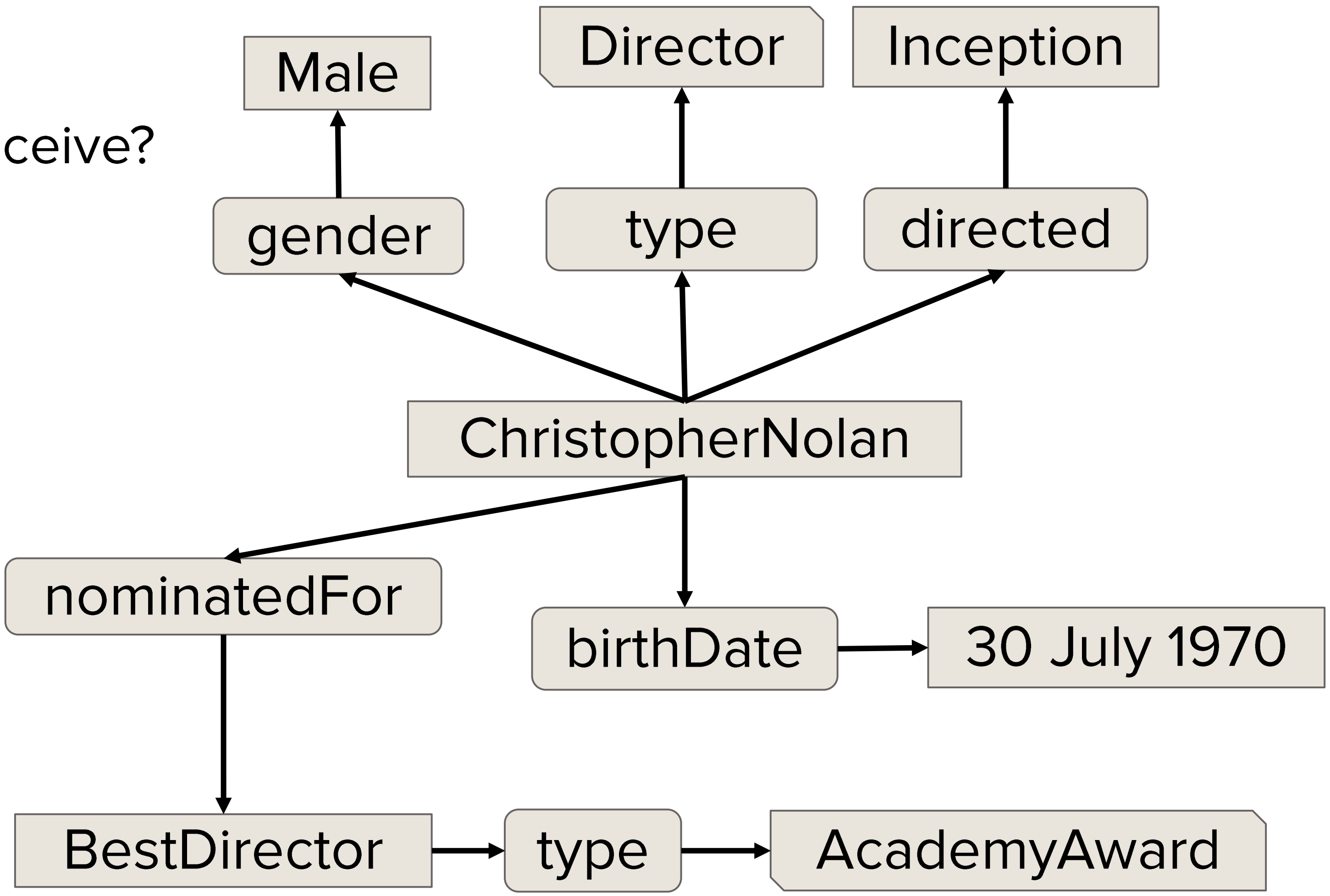
Wikidata: 12B facts, 84M entities, 7k predicates, 69k types



KGs and KBs are equivalent

Which Oscar nominations did Nolan receive?

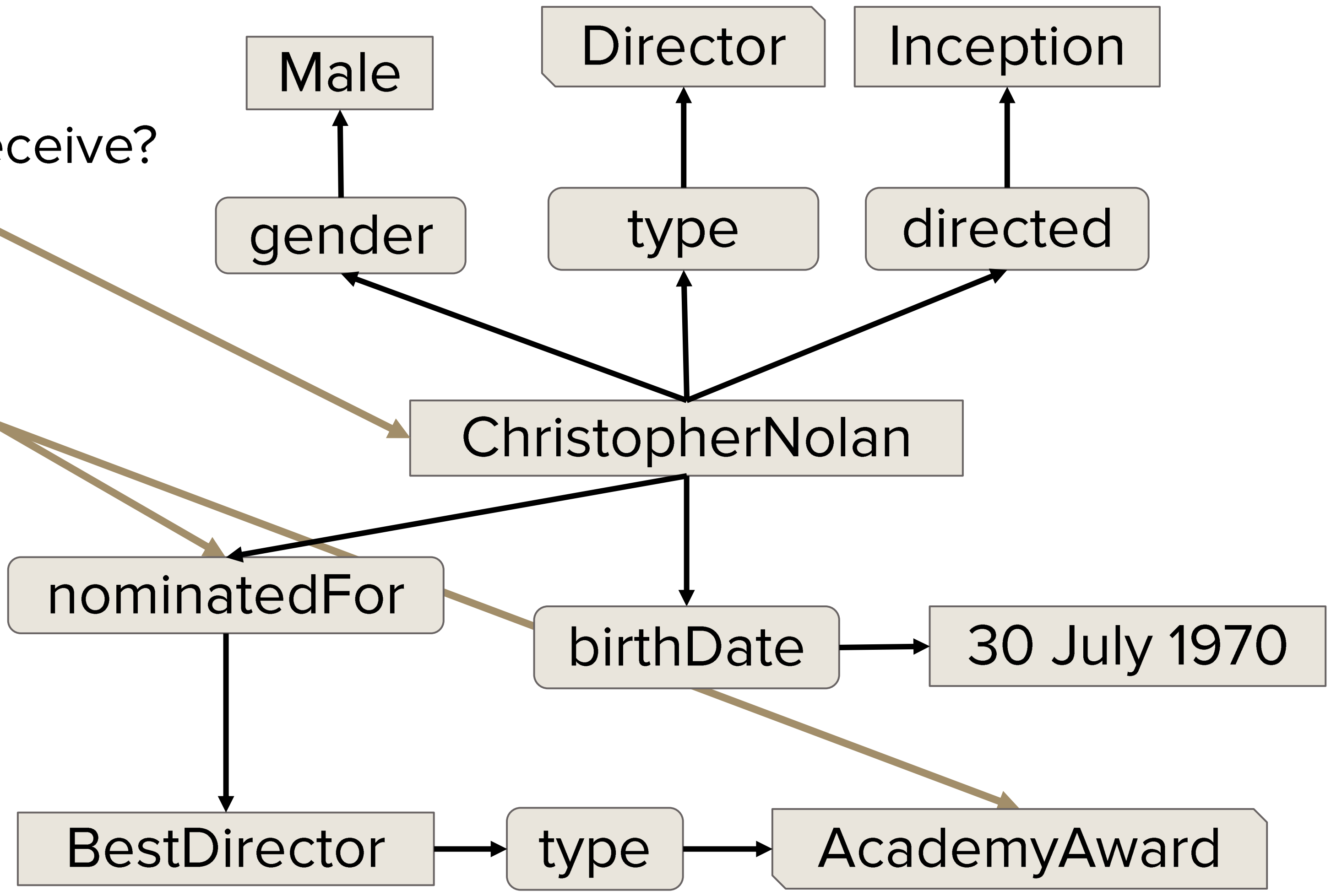
<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>



KG-QA Challenge 1: Bridge vocabulary gap

Which Oscar nominations did Nolan receive?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>



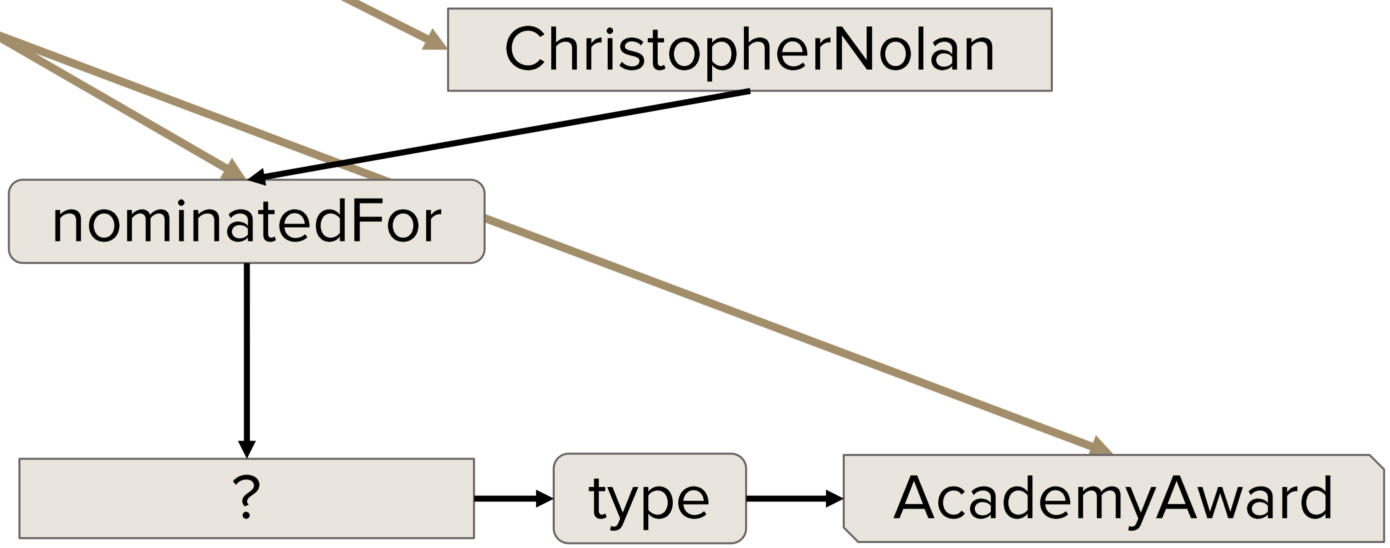
KG-QA Challenge 2: Query formulation

Which Oscar nominations did Nolan receive?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

SPARQL

```
SELECT ?ANS
WHERE {
ChristopherNolan nominatedFor ?ANS .
?ANS type AcademyAward    }
```



KG-QA Challenge 2: Query formulation

Which Oscar nominations did Nolan receive?

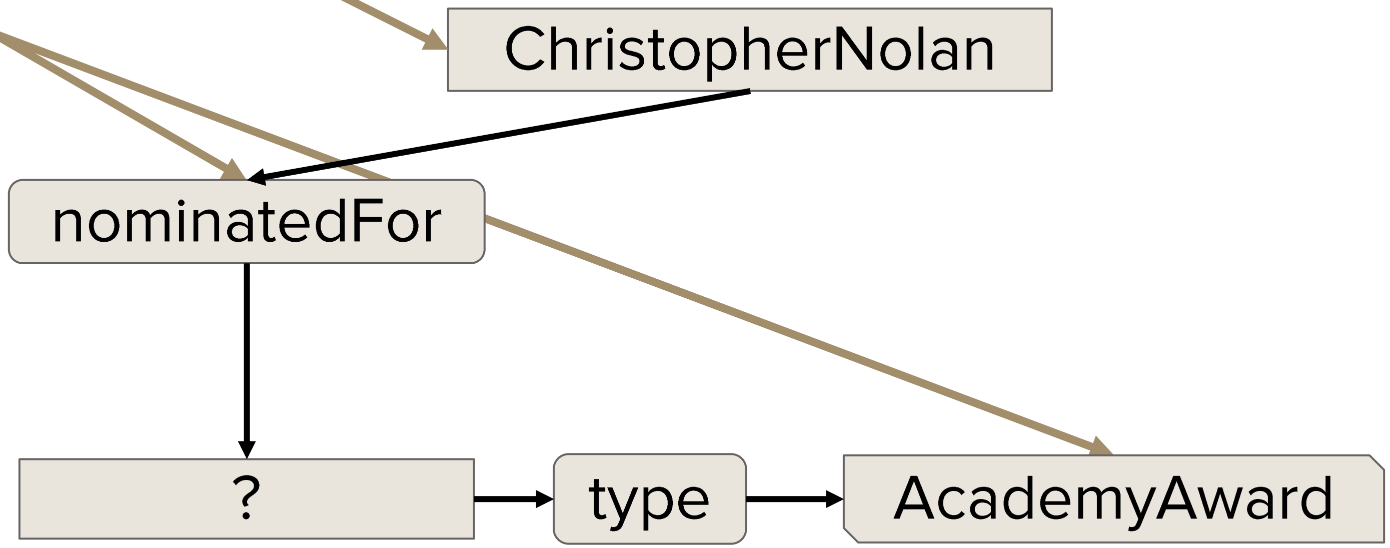
Named Entity Recognition and Disambiguation (NERD) systems (aka Entity Detection and Linking): TagME, AIDA, Dandelion, Google NL API, MS Text Analytics, IBM NLU

Named Entity Recognition (NER): Stanford NER, spaCy

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

SPARQL

```
SELECT ?ANS
WHERE {
ChristopherNolan nominatedFor ?ANS .
?ANS type AcademyAward    }
```



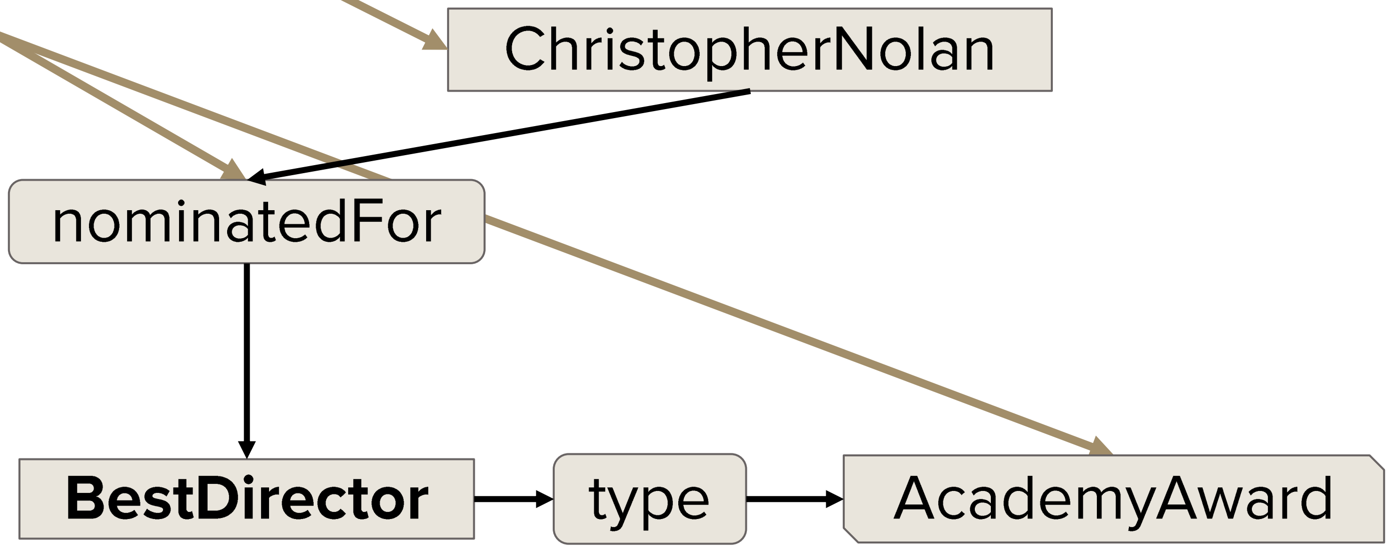
Answering with query

Which Oscar nominations did Nolan receive?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

SELECT ?ANS
WHERE {
ChristopherNolan nominatedFor ?ANS .
?ANS type AcademyAward }

SPARQL



BestDirector

Structured queries and logical forms

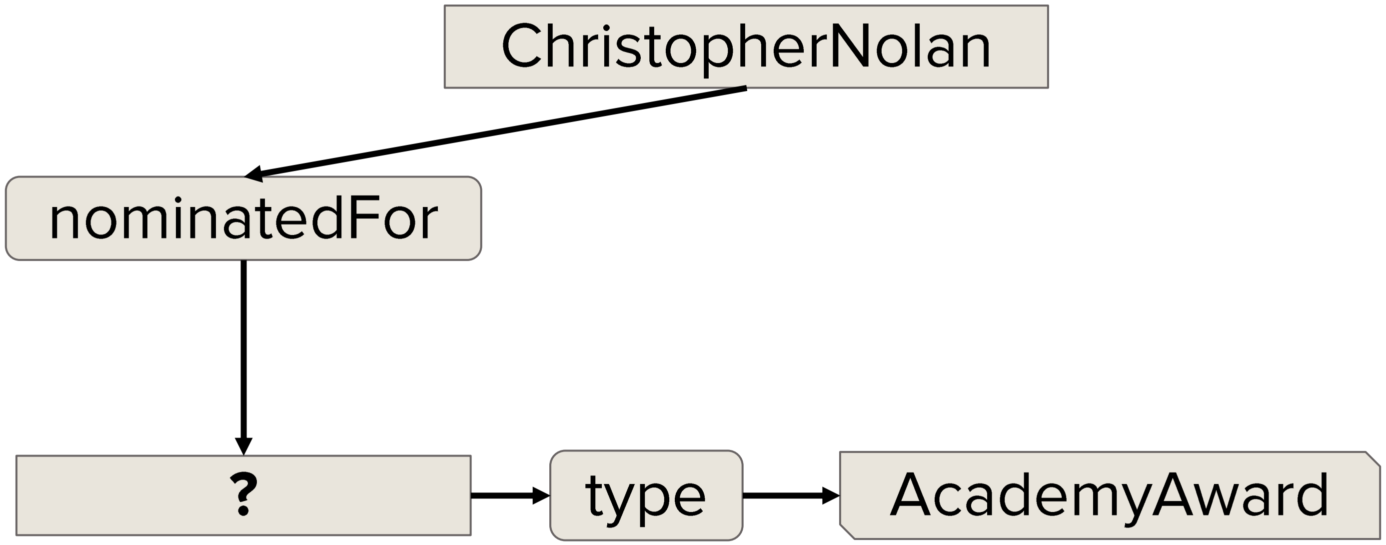
Which Oscar nominations did Nolan receive?

Lambda-calculus $\lambda x.\, nominatedFor(ChristopherNolan, x) \wedge Type(x, AcademyAward)$

Neo4j CYPHER Graph QL

Lambda-DCS $nominatedFor.ChristopherNolan \sqcap type.AcademyAward$

```
MATCH
(subj {name: 'ChristopherNolan' })-[:nominatedFor]-
>(obj:AcademyAward) RETURN obj.name
```



SPARQL BGP

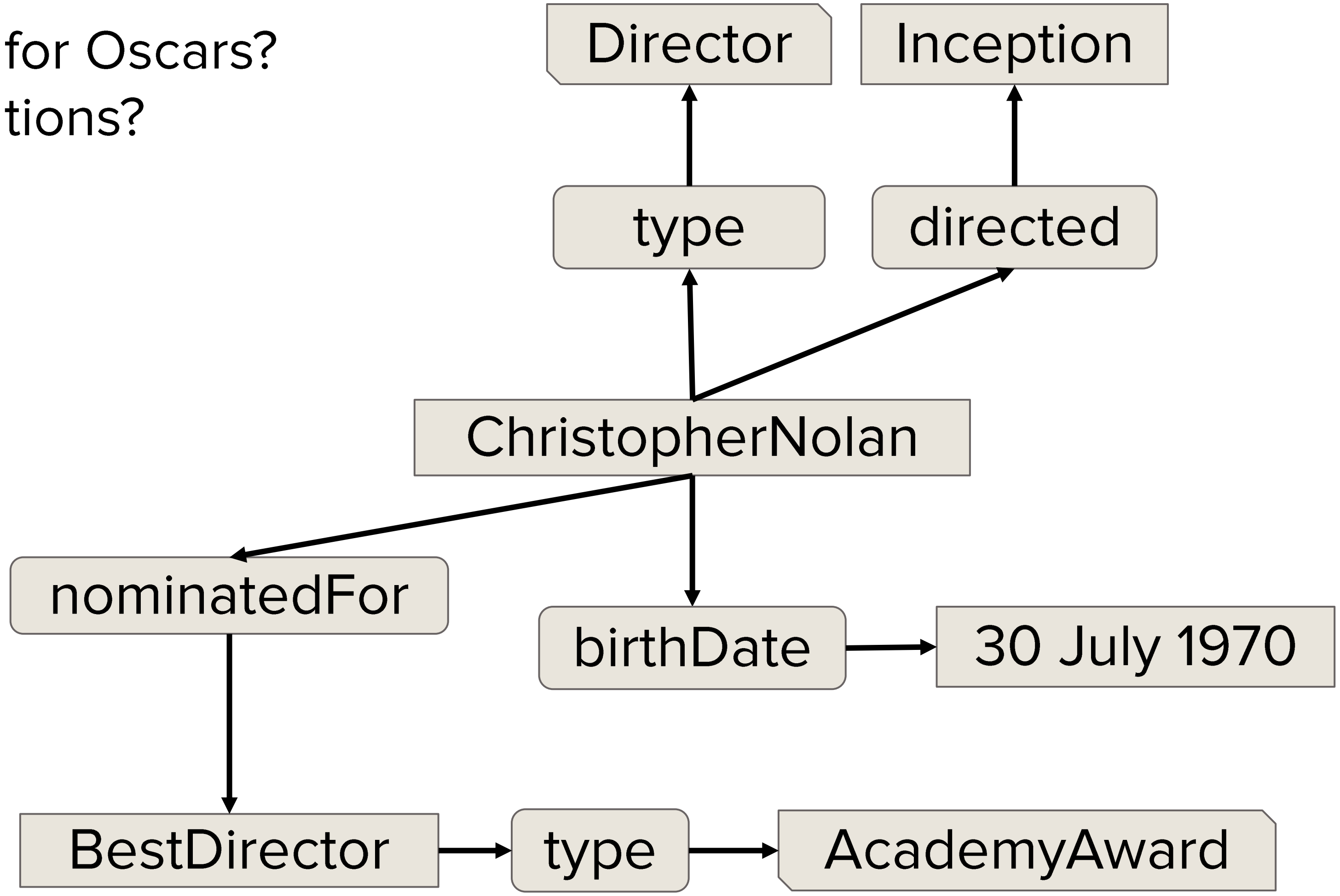
```
SELECT ?ANS
WHERE {
ChristopherNolan nominatedFor ?ANS .
?ANS type AcademyAward     }
```

BestDirector

Reification: n-ary information in KGs

For which films was Nolan nominated for Oscars?
When did Nolan get his Oscar nominations?

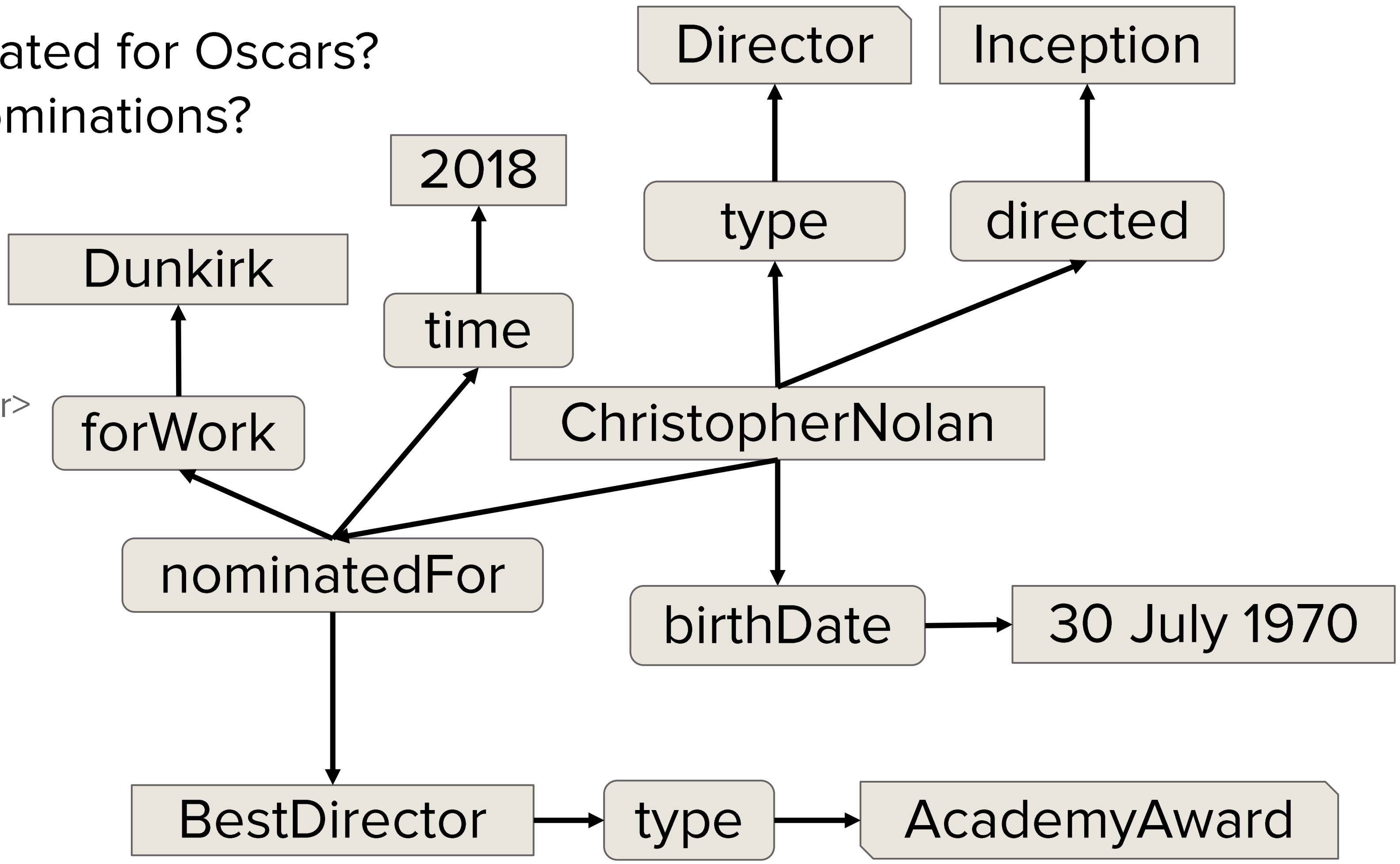
<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>



Reification: n-ary information in KGs

For which films was Nolan nominated for Oscars?
When did Nolan get his Oscar nominations?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

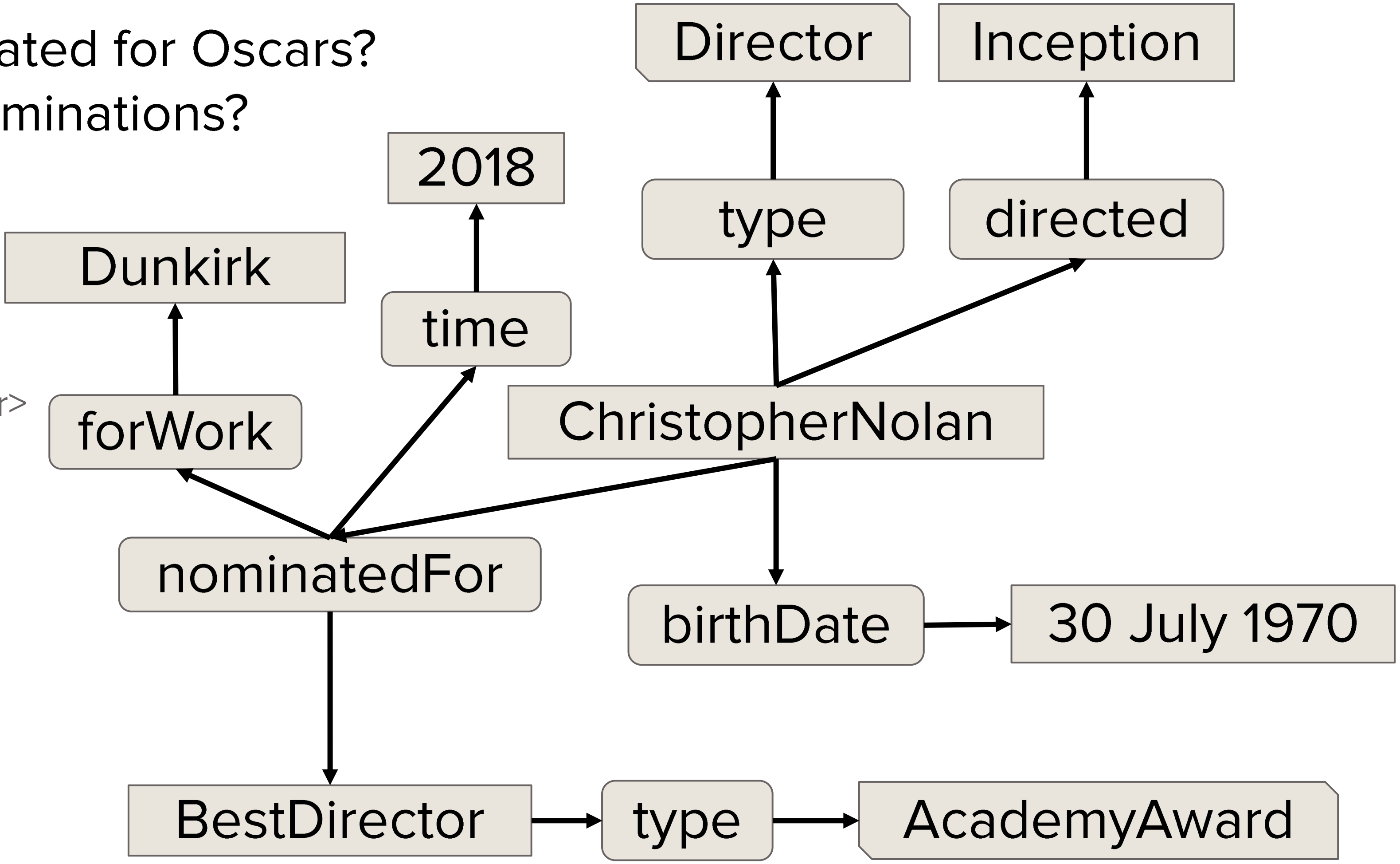


Reification: n-ary information in KGs

For which films was Nolan nominated for Oscars?
When did Nolan get his Oscar nominations?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

<ChristopherNolan, nominatedFor, 123>
<123, nominatedFor, BestDirector>
<123, forWork, Dunkirk>
<123, year, 2018>

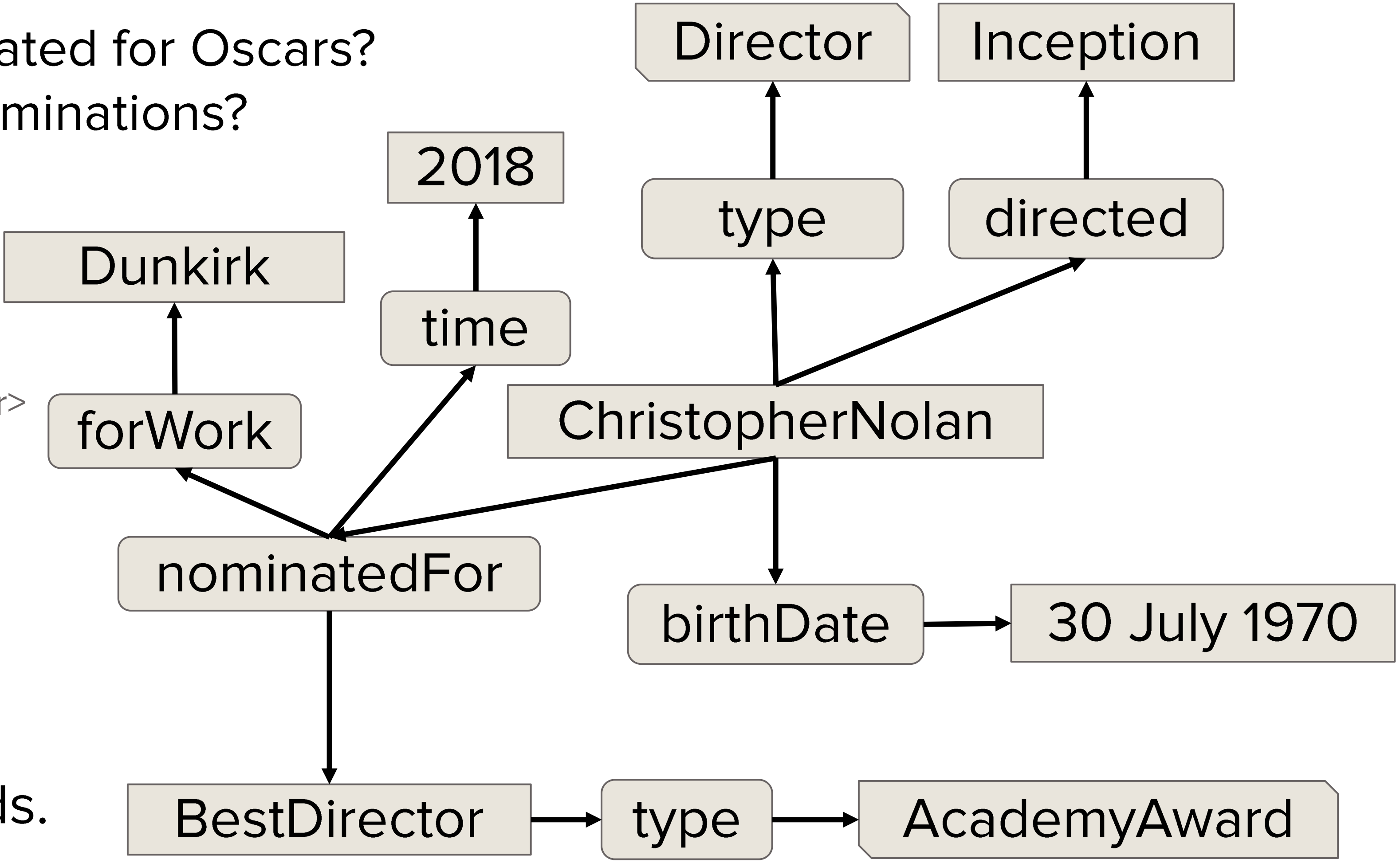


Qualifiers are a huge part of Wikidata

For which films was Nolan nominated for Oscars?
When did Nolan get his Oscar nominations?

<ChristopherNolan, gender, Male>
<ChristopherNolan, type, Director>
<ChristopherNolan, directed, Inception>
<ChristopherNolan, nominatedFor, BestDirector>
<BestDirector, type, AcademyAward>
<ChristopherNolan, birthDate, 30 July 1970>

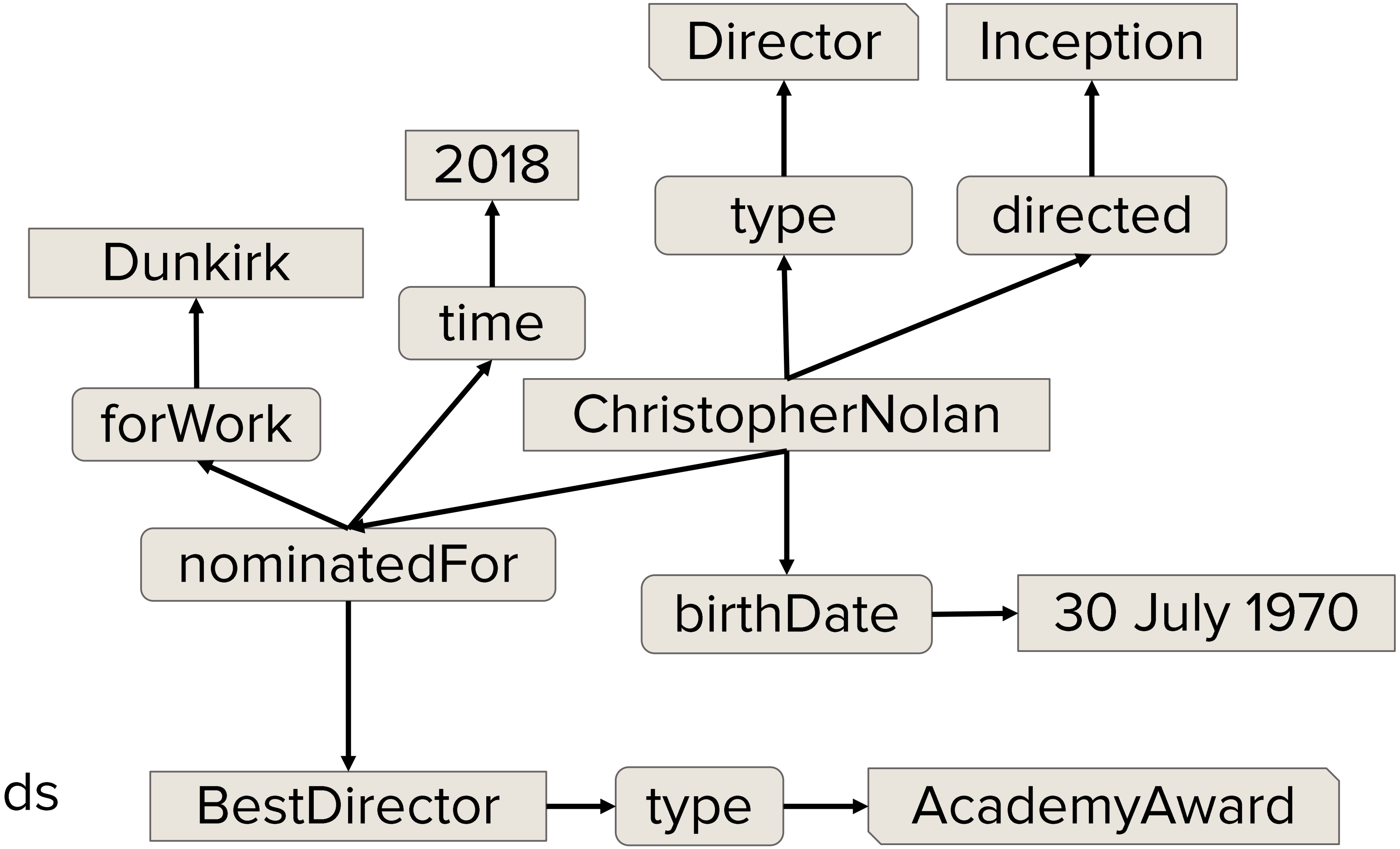
<ChristopherNolan, nominatedFor, 123>
<123, nominatedFor, BestDirector>
<123, forWork, Dunkirk>
<123, time, 2018>



Wikidata: Qualifiers, Statement-Ids.
6B triples part of reified facts!!

Questions that need reified triples

Who played Cobb in Inception?
Who did Leo play in Inception?
When did Neymar join PSG?
Who was Trump's first wife?
US president in 2016?
…



Wikidata: Qualifiers, Statement-Ids
6B triples part of reified facts!!

Explore Wikidata

Entity name / Subject

Entity id

The Dark Knight (Q163872)

Entity desc

2008 British-American superhero film directed by Christopher Nolan

TDK | Dark Knight

Entity aliases

Type

Type predicate

instance of	film	
genre	action film	
director	Christopher Nolan	

Predicate

Object

nominated for	Academy Award for Best Supporting Actor	
	statement is subject of	81st Academy Awards
	nominee	Heath Ledger
	point in time	22 February 2009

part of the series	The Dark Knight Trilogy	
	follows	Batman Begins
	followed by	The Dark Knight Rises
	series ordinal	2

cast member	Christian Bale	
	character role	Bruce Wayne
	11 references	
	Michael Caine	
	character role	Alfred Pennyworth
	4 references	
	Heath Ledger	
	character role	Joker
	9 references	

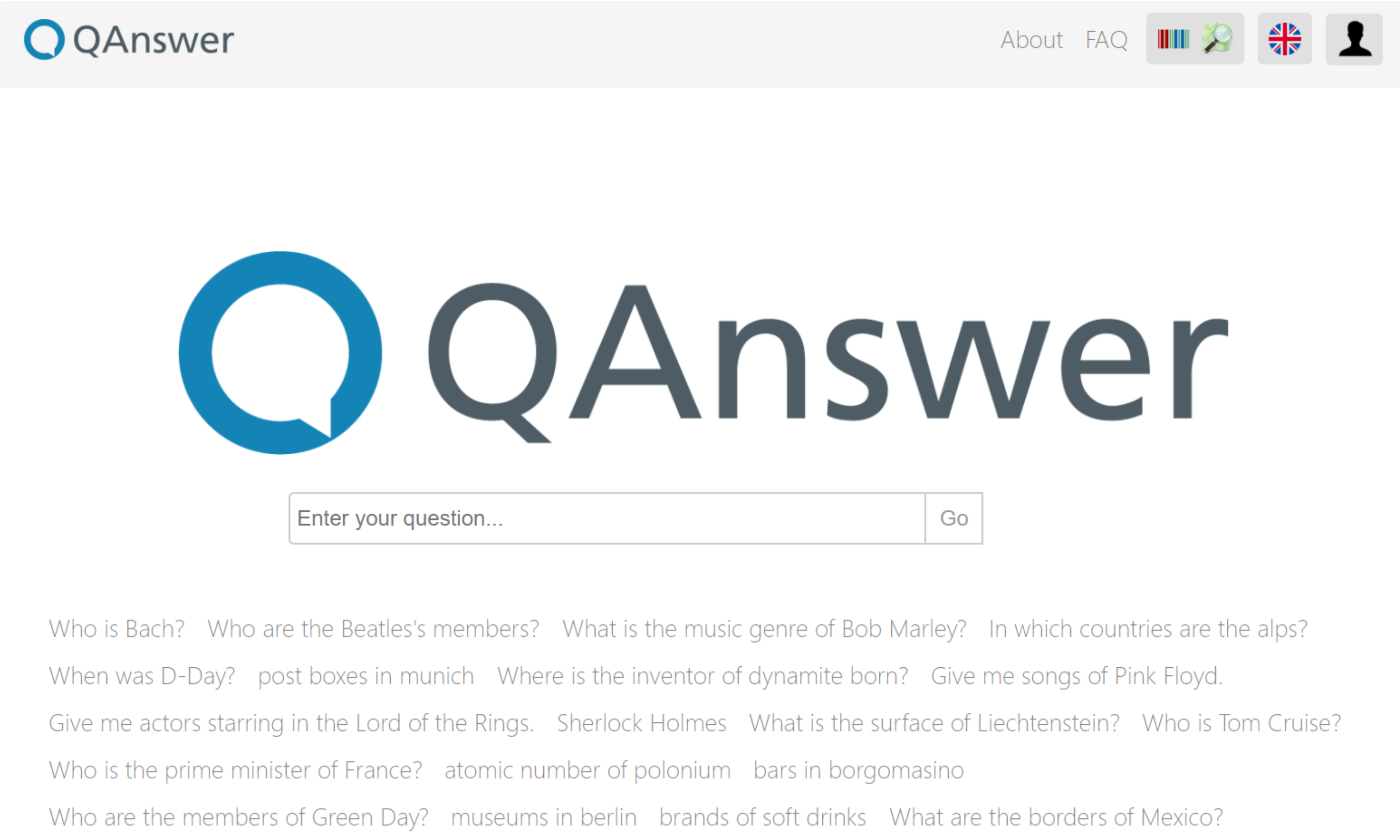
Qualifier predicate

Qualifier object

Explore Wikidata like a pro

- Wikidata: https://www.wikidata.org/wiki/Wikidata:Main_Page
- Wikidata data model: https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer
- Wikidata dumps: https://www.wikidata.org/wiki/Wikidata:Database_download
- Download latest **n-triples** dump: https://dumps.wikimedia.org/wikidatawiki/entities/
- Wikidata SPARQL Endpoint: https://query.wikidata.org/
- Wikidata statistics: https://stats.wikimedia.org/#/wikidata.org
- More stats: https://www.wikidata.org/wiki/Wikidata:Statistics

Play with QA (over Wikidata)



Diefenbach et al., QAnswer: A Question answering prototype bridging the gap between a considerable part of the LOD cloud and end-users, WWW 2019. Available at https://qanswer-frontend.univ-st-etienne.fr/

Benchmarks

- **Simple questions**
  - WebQuestions (Berant et al. 2013) over Freebase
  - SimpleQuestions (Bordes et al. 2015) over Freebase
- **Complex questions**
  - LC-QuAD 2.0 (Dubey et al 2018) over Wikidata + DBpedia
  - MetaQA (Zhang et al. 2018) over Freebase
- **Conversational questions**
  - ConvQuestions (Christmann et al. 2019) over Wikidata
  - CSQA (Saha et al. 2018) over Wikidata

Benchmarks

- **Simple questions**
  - WebQuestions* (Berant et al. 2013) over Freebase
  - SimpleQuestions (Bordes et al. 2015) over Freebase
- **Complex questions**
  - LC-QuAD 2.0* (Dubey et al 2018) over Wikidata + DBpedia
  - MetaQA (Zhang et al. 2018) over Freebase
- **Conversational questions**
  - ConvQuestions* (Christmann et al. 2019) over Wikidata
  - CSQA (Saha et al. 2018) over Wikidata

Recent benchmarks over Wikidata

More realistic benchmarks are smaller but harder

Much higher numbers on semi-synthetic benchmarks

"Vulnerable" to neural methods

* Need reified triples for answering

Benchmarks

- **Simple questions**
  - WebQuestions* (Berant et al. 2013) over Freebase
  - SimpleQuestions (Bordes et al. 2015) over Freebase
- **Complex questions**
  - LC-QuAD 2.0* (Dubey et al 2018) over Wikidata + DBpedia
  - MetaQA (Zhang et al. 2018) over Freebase
- **Conversational questions**
  - ConvQuestions* (Christmann et al. 2019) over Wikidata
  - CSQA (Saha et al. 2018) over Wikidata

Many, many more:

LC-QuAD
(Trivedi et al. 2017)

ComQA
(Abujabal et al. 2019)

GraphQuestions
(Su et al. 2016)

QALD
(Usbeck et al. 2018)

TempQuestions
(Jia et al. 2018)

ComplexWebQuestions
(Talmor and Berant 2018)

WikiMovies
(Miller et al. 2016)

ComplexQuestions
(Bao et al. 2016)

Benchmarks: WebQuestions

- Real questions: Collected using the Google Suggest API
- Mostly simple questions using one fact or reified triple
- 3778 train, 2032 test questions
- Available at: https://nlp.stanford.edu/software/sempre/

who was richard nixon married to?

what high school did harper lee go to?

what was the capital city of the east roman empire?

who plays ken barlow in coronation street?

where is the fukushima daiichi nuclear plant located?

Benchmarks: LC-QuAD 2.0

- Sampled SPARQL queries via templates, verbalized by crowdworkers
- Complex (and simple) questions involving multiple entities and relations
- 23954 train, 6046 test questions
- Available at: http://lc-quad.sda.tech/

What city is the twin city of Oslo and also the setting for “A Tree Grows in Brooklyn”?

What Empire used to have Istanbul as its capital?

How long was Shirley Temple the United States Ambassador to Ghana?

Were Dutch and Hungarian the official languages of the Holy Roman Empire?

Who replaced Albus Dumbledore as headmaster of Hogwarts?

Benchmarks: ConvQuestions

- Natural conversations by crowdworkers after choosing topic
- Both simple and complex
- Five domains
- 6720 train, 2240 dev, 2240 test conversations
- Available at:

https://convex.mpi-inf.mpg.de/

Books	Movies	Soccer	Music	TV series
When was the first book of the book series The Dwarves published ?	Who played the joker in The Dark Knight?	Which European team did Diego Costa represent in the year 2018?	Led Zeppelin had how many band members?	Who is the actor of James Gordon in Gotham?
2003	Heath Ledger	Atlético Madrid	4	Ben McKenzie
What is the **name of the second book?**	When did he die?	Did they win the Super Cup the previous year?	Which was released first: Houses of the Holy or Physical Graffiti?	**What about Bullock?**
The War of the Dwarves	22 January 2008	No	Houses of the Holy	Donal Logue
Who is the author ?	**Batman actor?**	**Which club was the winner?**	**Is the rain song and immigrant song there?**	Creator?
Markus Heitz	Christian Bale	Real Madrid C.F.	No	Bruno Heller
In which city was he born ?	Director?	Which English club did Costa play for before returning to Atlético Madrid?	Who wrote those songs?	**Married to in 2017?**
Homburg	Christopher Nolan	Chelsea F.C.	Jimmy Page	Miranda Phillips Cowley
When was he born ?	**Sequel name?**	**Which stadium is this club's home ground?**	**Name of his previous band?**	**Wedding date first wife?**
10 October 1971	The Dark Knight Rises	Stamford Bridge Stadium	The Yardbirds	19 June 1993

Metrics

- Answers as sets (for systems using explicit structured queries)
  - Precision, Recall, F1-Score
- Answers as ranked lists (systems w/o explicit queries: approx. graph search)
  - Precision@1, MRR, MAP
  - Hit@5
- Single answer
  - Accuracy

break duration ?x .
?x measuredIn minutes .

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

Getting started: Templates and embeddings

Foundational work in KG-QA

- Templates over RDF (Unger et al. 2012)
- DEANNA (Yahya et al. 2012, 2013)
- SEMPRE (Berant et al. 2013)
- PARALEX + OQA (Fader et al. 2013, 2014)
- Subgraph embeddings (Bordes et al. 2014)
- STAGG (Yih et al. 2015)
- AQQU (Bast and Haussman 2015)

Templates for KG-QA

- Interpretable

Question	**Question template**
Who is **Inception**’s **director**?	Who is **<NOUN1>**’s **<NOUN2>**?
Query	**Query template**
?ANS **director Inception**	?ANS **<PRED1> <ENT1>**

1 SPARQL triple pattern

Templates for KG-QA

- Generalizes to new domains

Who is **Libya**’s **president**?

Who is **Messi**’s **manager**?

Question Who is **Inception**’s **director**?	**Question template** Who is **<NOUN1>**’s **<NOUN2>**?
Query ?ANS **director Inception**	**Query template** ?ANS **<PRED1> <ENT1>**

1 SPARQL triple pattern

Templates for KG-QA

Question Who **plays** the **role of Cobb in Inception**?	**Question template** Who <**VERB**> <DT> **<NOUN1> <PREP1> <NOUN2>** <**PREP2**> **<NOUN3>**?
Query **Inception castMember** ?VAR ?VAR **castMember** ?ANS ?VAR **characterRole Cobb**	**Query template** **<ENT1>** <**PRED1**> ?VAR ?VAR <**PRED1**> ?ANS ?VAR **<PRED2> <ENT2>**

Multiple SPARQL triple patterns

Limitations of templates

- Hand-crafted by experts (Fader et al. 2013, 2014; Unger et al. 2012)
- Restricted coverage
- **Solution:** Learn templates
  - Question templates
  - Query templates
  - Slot alignments
- Proposed in the QUINT+NEQA framework (Abujabal et al. 2017, 2018)

Distant supervision from QA pairs

Question: Which Oscar award nomination did Nolan get for the film Dunkirk?
Answer: Best Director

NERD system

Dunkirk

NERD system

ChristopherNolan

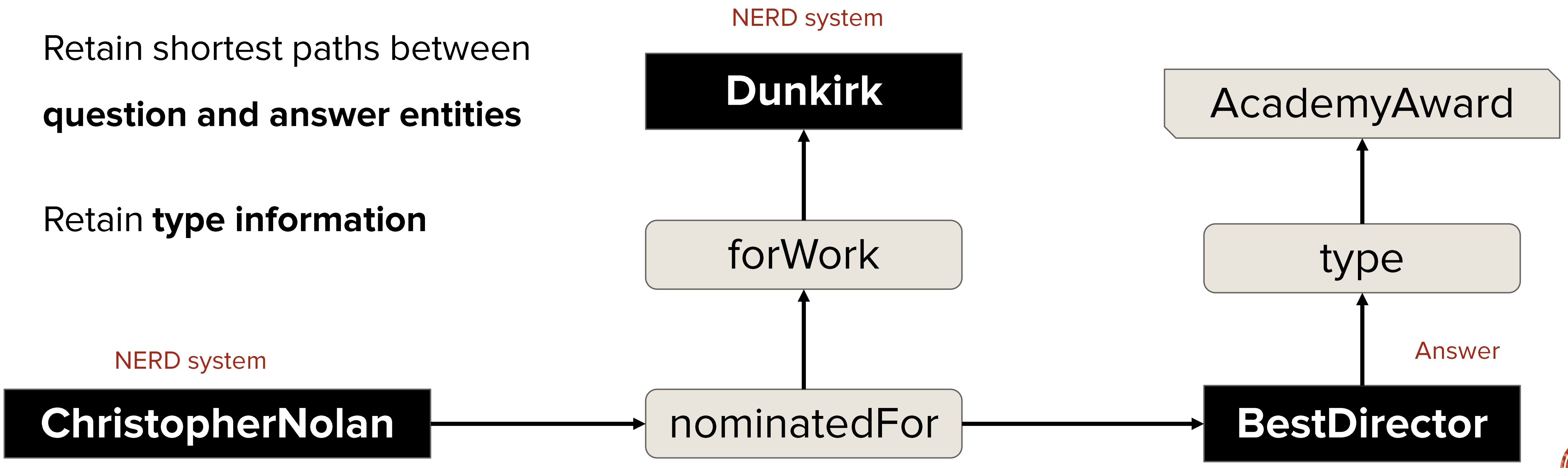
Answer

BestDirector

Distant supervision from QA pairs

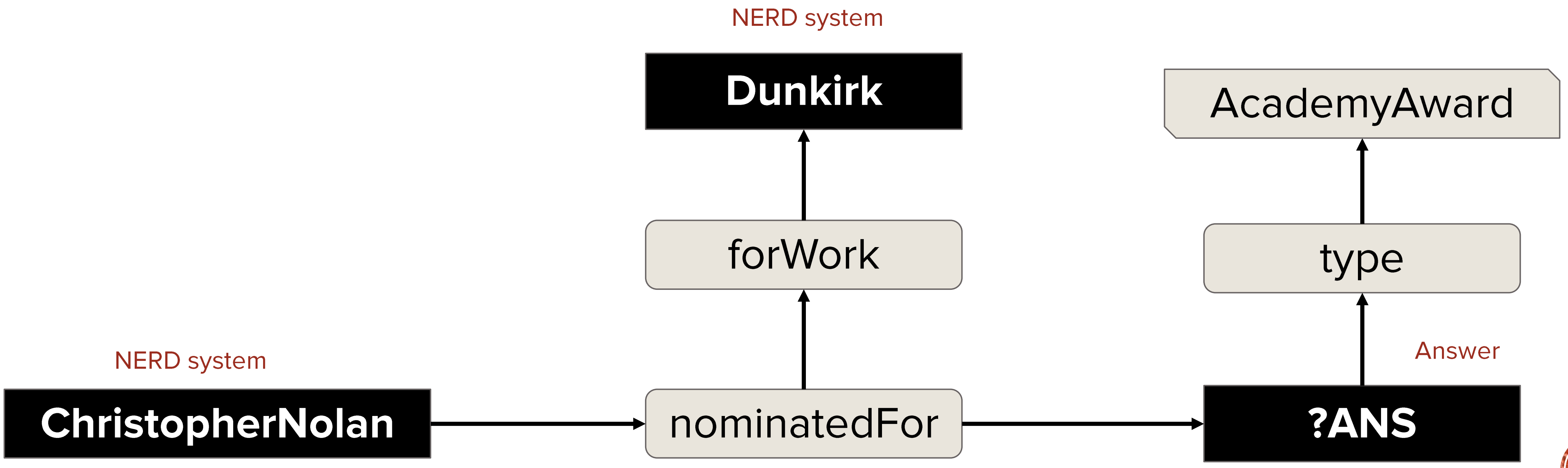
Question: Which Oscar award nomination did Nolan get for the film Dunkirk?
Answer: Best Director

- Retain shortest paths between **question and answer entities**
- Retain **type information**

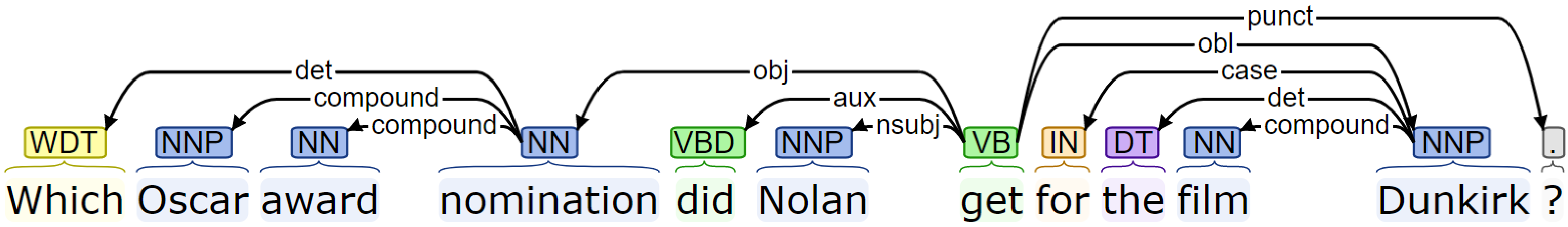


Distant supervision from QA pairs

Question: Which Oscar award nomination did Nolan get for the film Dunkirk?
Answer: Best Director
Query: SELECT ?x WHERE { ChristopherNolan nominatedFor ?VAR .
?VAR nominatedFor ?ANS .
?VAR forWork Dunkirk .
?VAR type AcademyAward . }



Extract question phrases

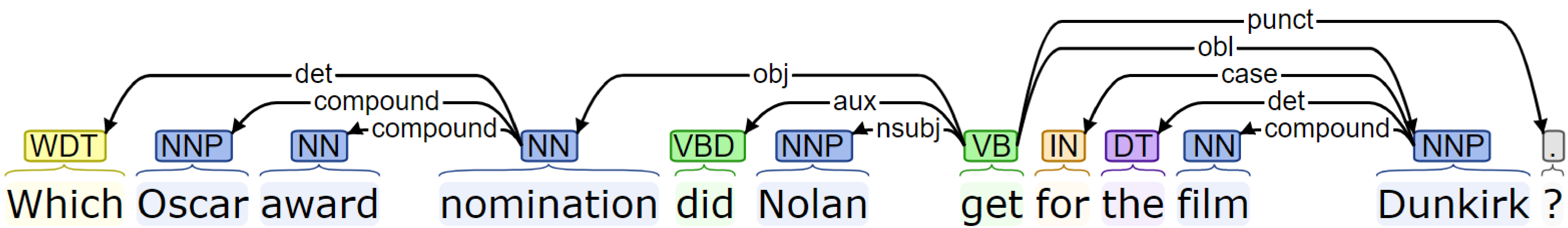


which nomination	oscar nomination	get nomination	for film	oscar	get

which	did get	did	oscar award	nomination	award	award nomination

Dependency parsing: https://web.stanford.edu/~jurafsky/slp3/15.pdf

Extract query items



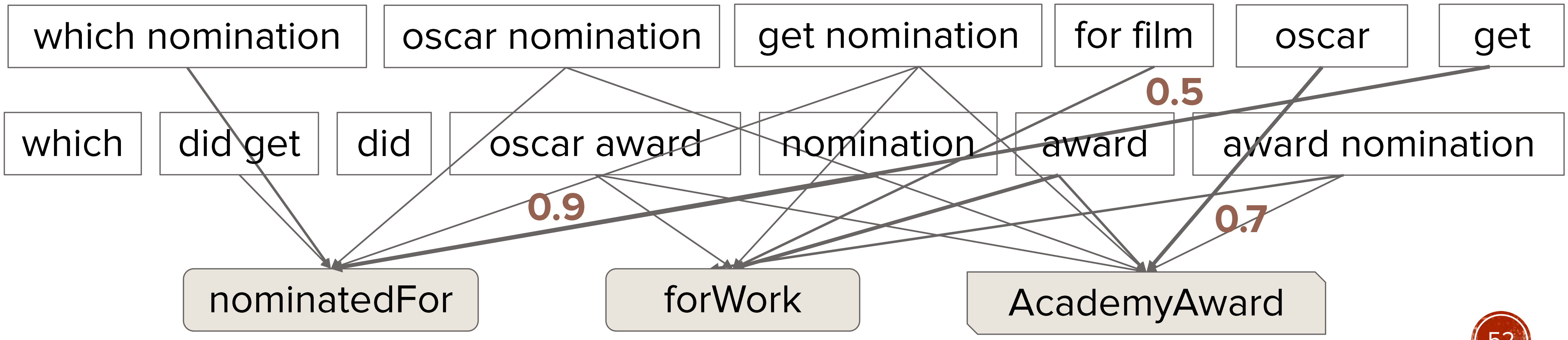
which nomination | oscar nomination | get nomination | for film | oscar | get

which | did get | did | oscar award | nomination | award | award nomination

nominatedFor | forWork | AcademyAward

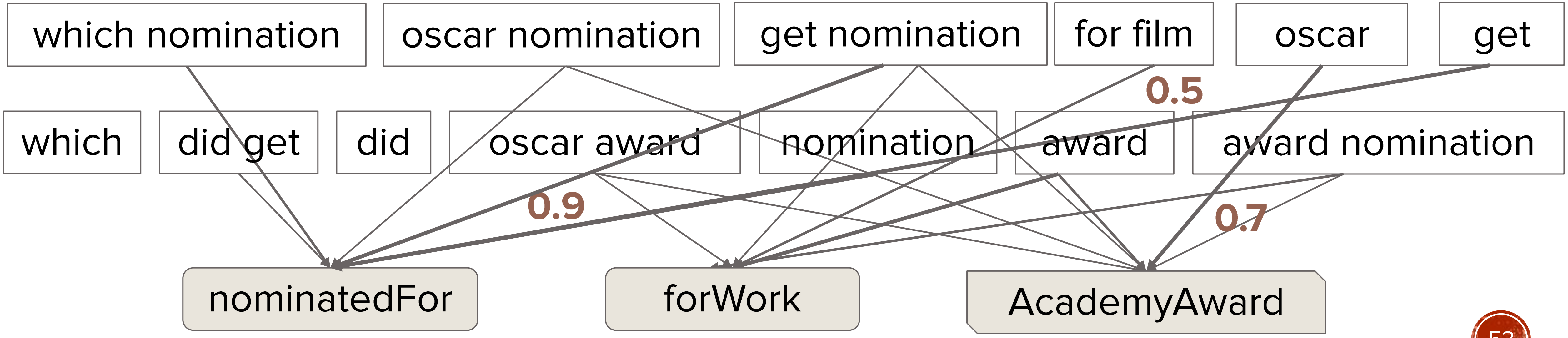
Create candidate alignments

- **Bipartite graph** with edge weights (Yahya et al. 2012)
- **Weights** from lexicons L_P and L_T (Abujabal et al. 2017, Berant and Liang 2013)



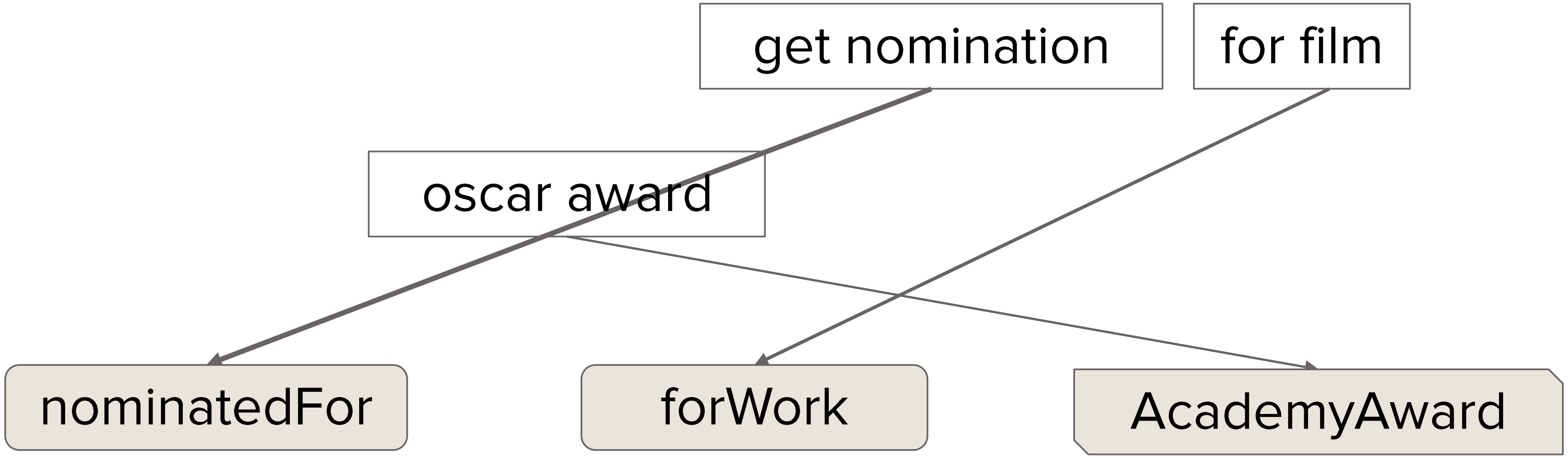
Optimal mapping via Integer Linear Program (ILP)

- Best alignment of items with **Integer Linear Program (ILP)**
- **Constraint 1:** Each KG item obtained from at most one phrase
- **Constraint 2:** Token contributing to entity cannot contribute to any other phrase
- **Constraint 3:** One phrase can map to at most one type

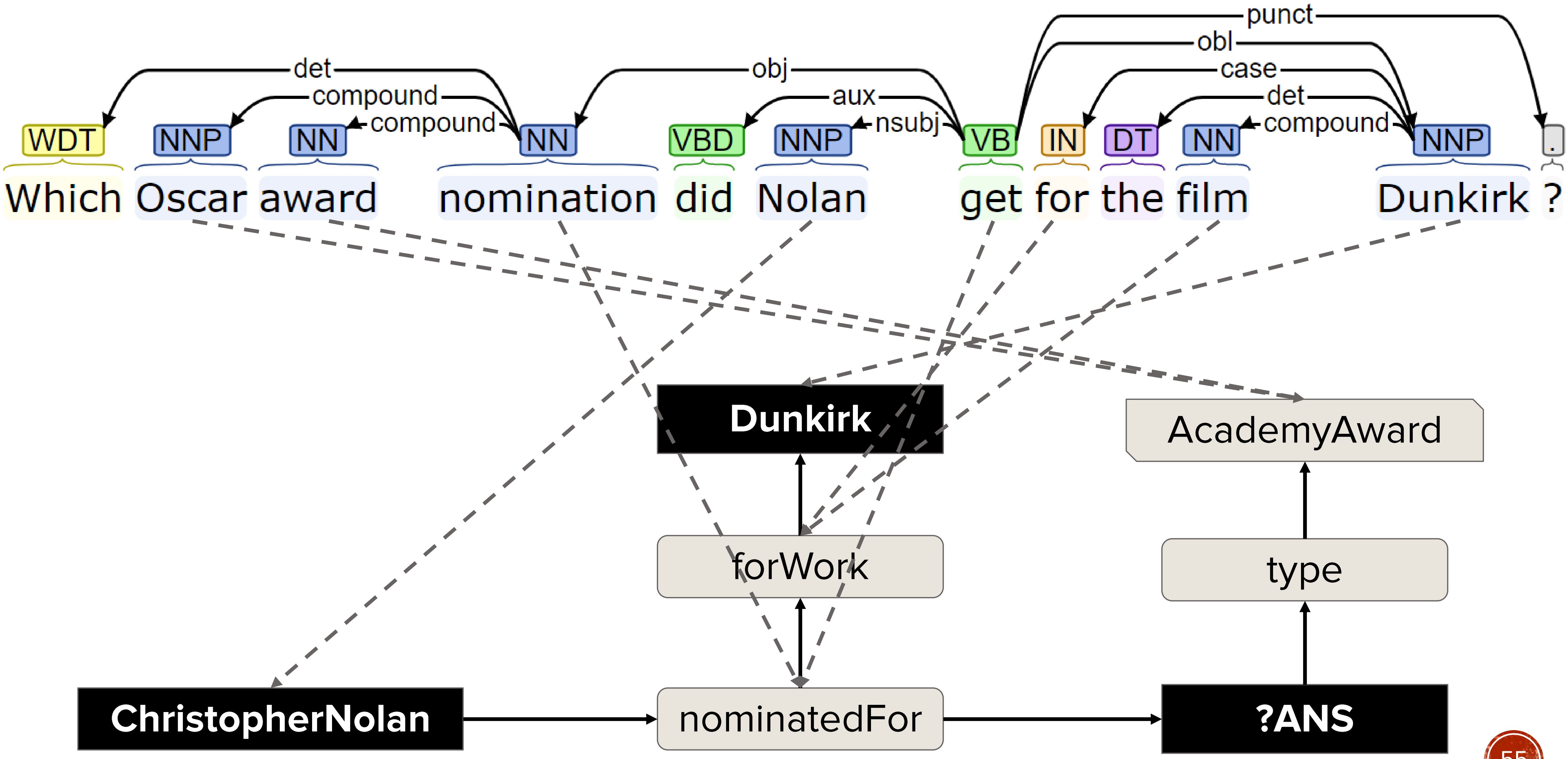


Optimal mapping via Integer Linear Program (ILP)

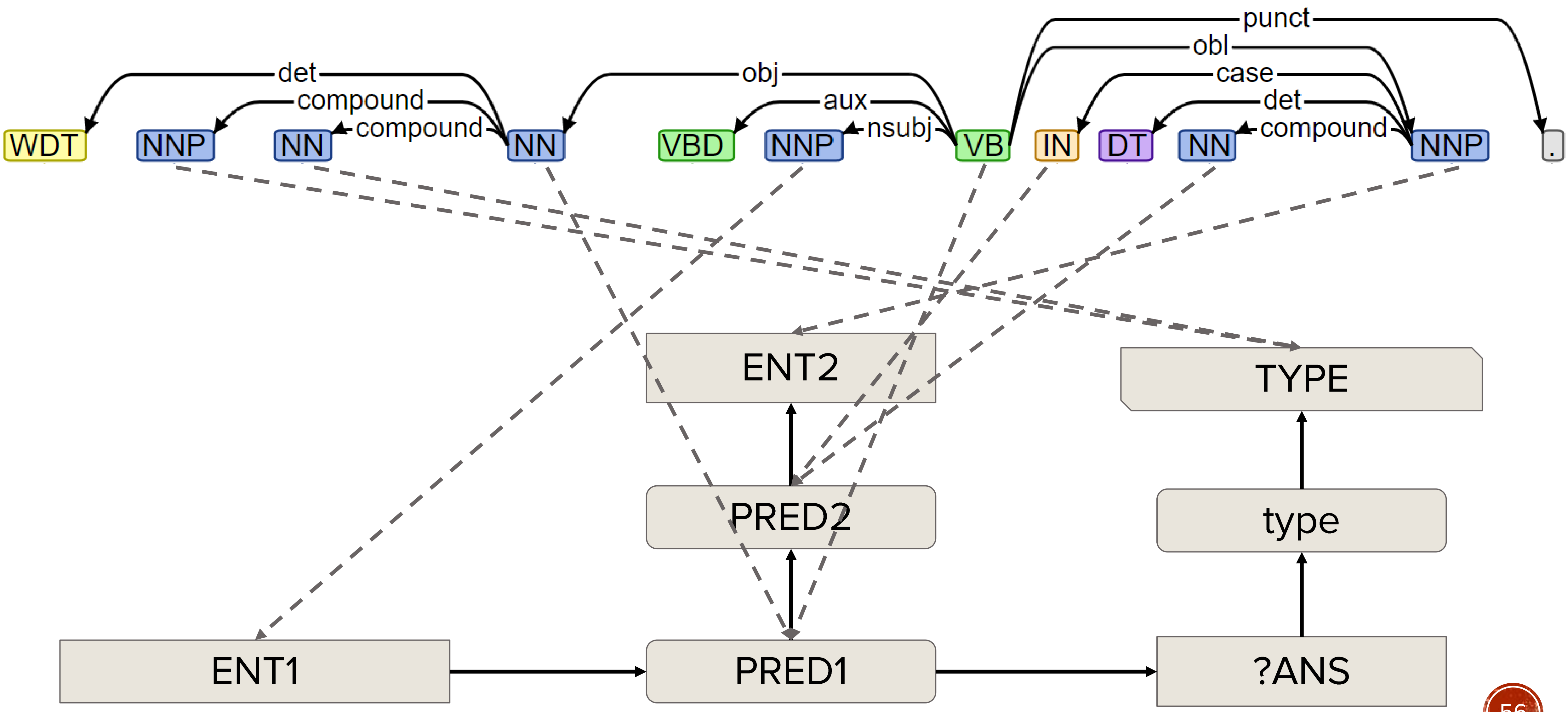
- Best alignment of items with **Integer Linear Program (ILP)**
- **Constraint 1:** Each KG item obtained from at most one phrase
- **Constraint 2:** Token contributing to entity cannot contribute to any other phrase
- **Constraint 3:** One phrase can map to at most one type



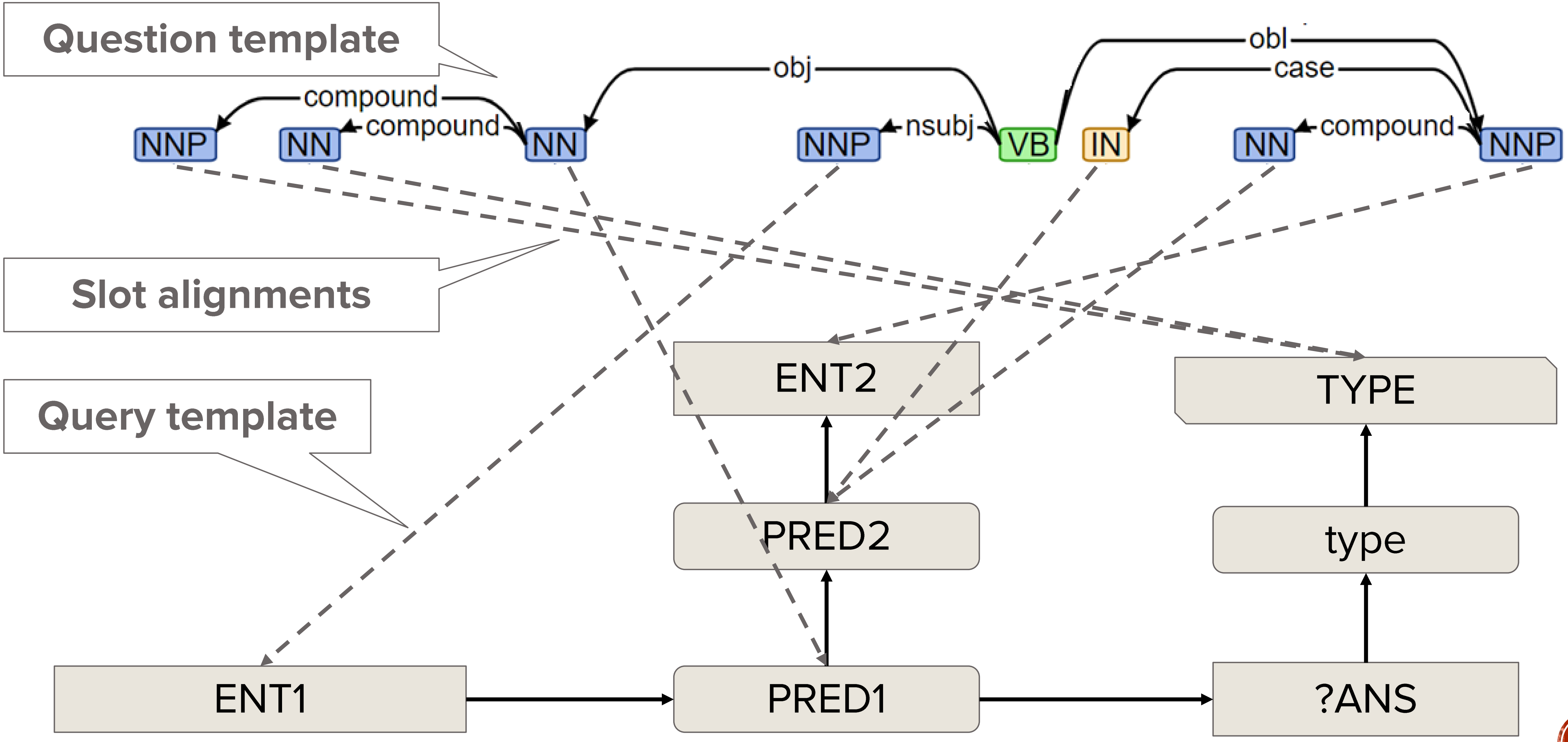
Apply alignment to question-query



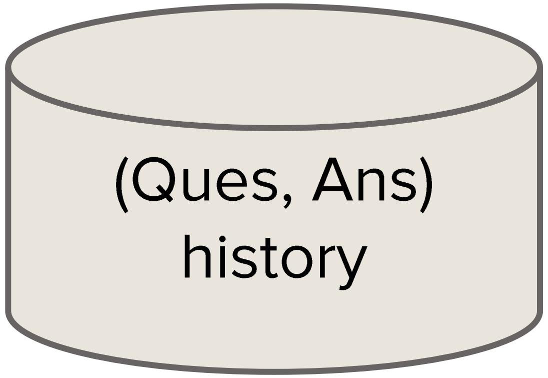
Replace concrete items by roles



Drop unnecessary question words



A continuous learning framework



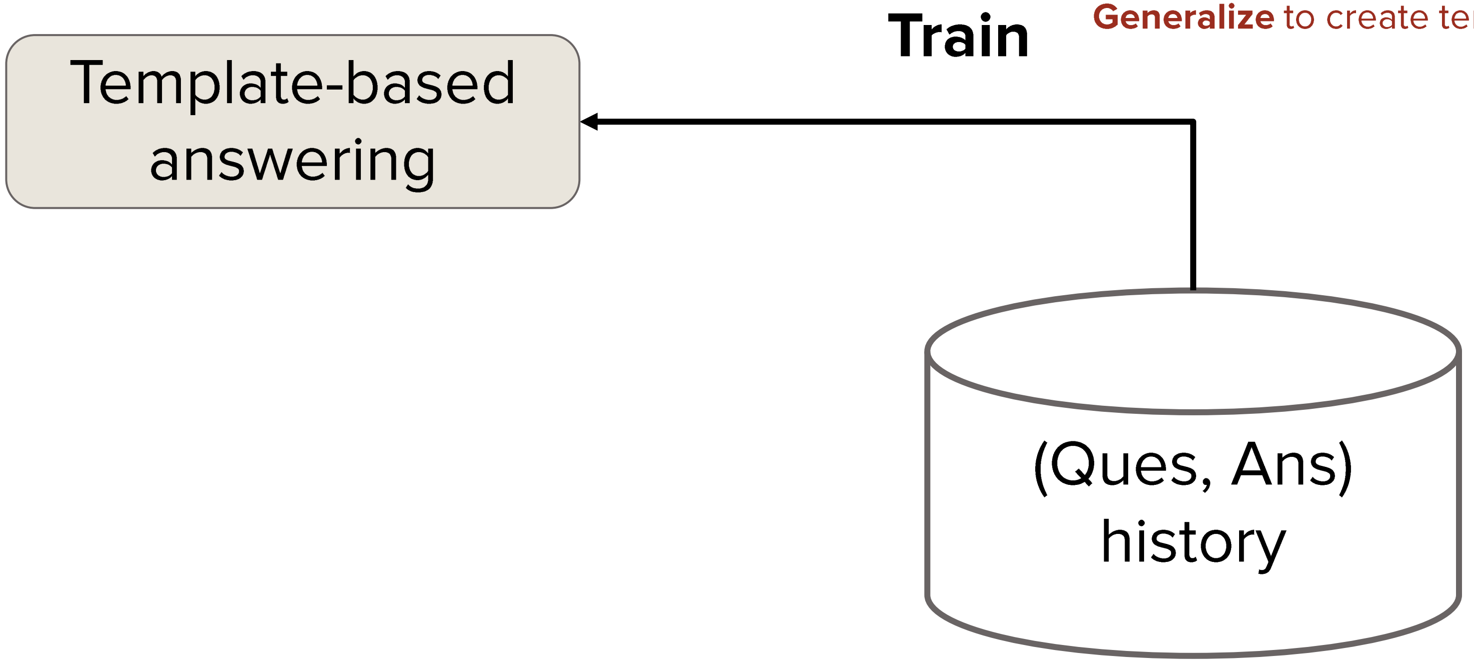
Abujabal et al., Automated template generation for question answering over knowledge graphs, WWW 2017.
Abujabal et al., Never-ending learning for open-domain question answering over knowledge bases, WWW 2018.

Train system using (Q, A) pairs

Distant supervision to go from (Ques, Ans) to (Ques, query)

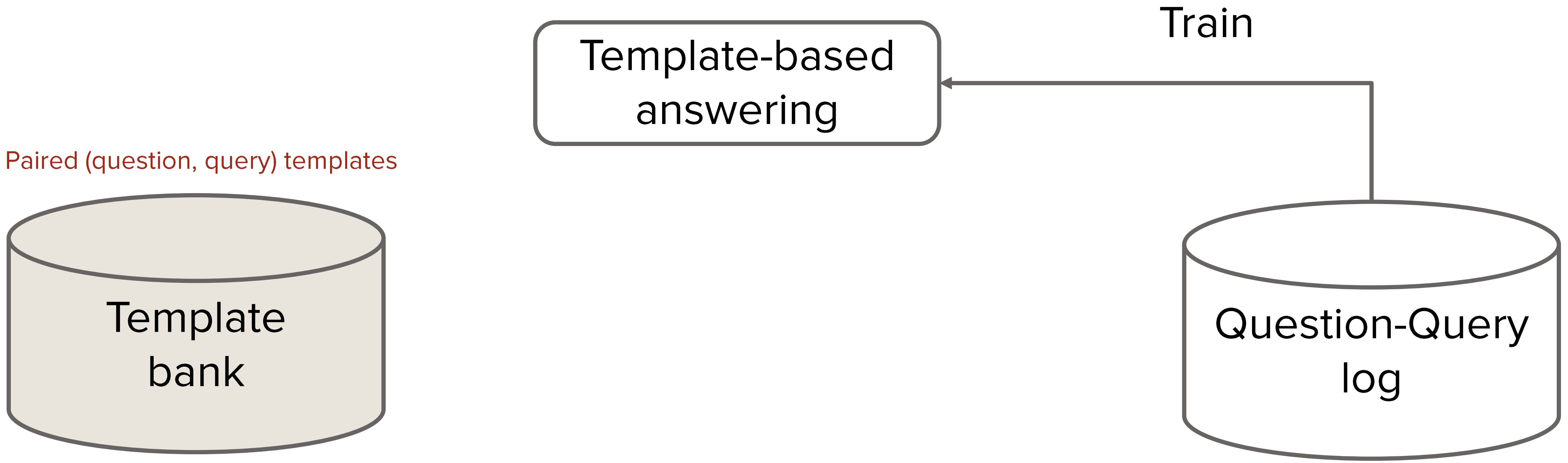
Align ques tokens to query tokens via lexicons and ILP

Generalize to create template



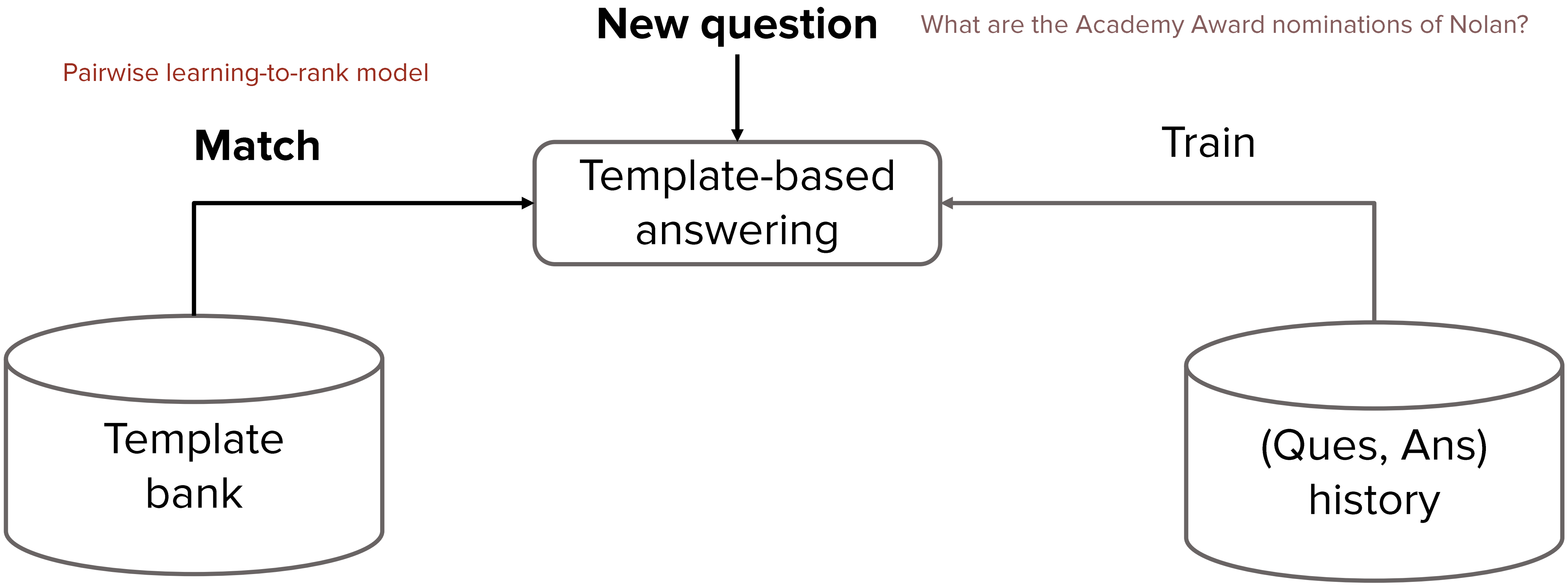
Abujabal et al., Automated template generation for question answering over knowledge graphs, WWW 2017.
Abujabal et al., Never-ending learning for open-domain question answering over knowledge bases, WWW 2018.

Learn a template repository



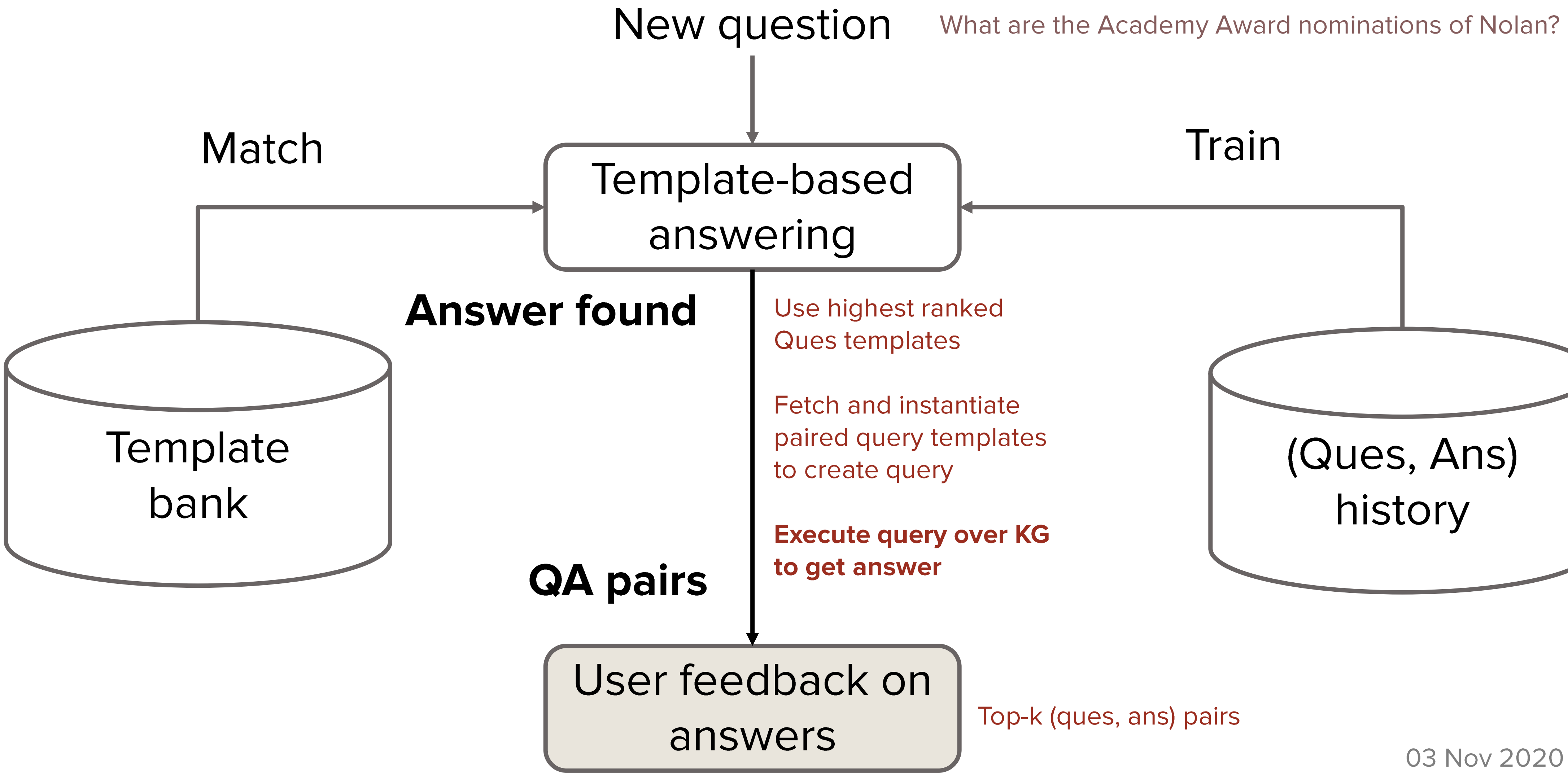
Abujabal et al., Automated template generation for question answering over knowledge graphs, WWW 2017.
Abujabal et al., Never-ending learning for open-domain question answering over knowledge bases, WWW 2018.

Answering with templates

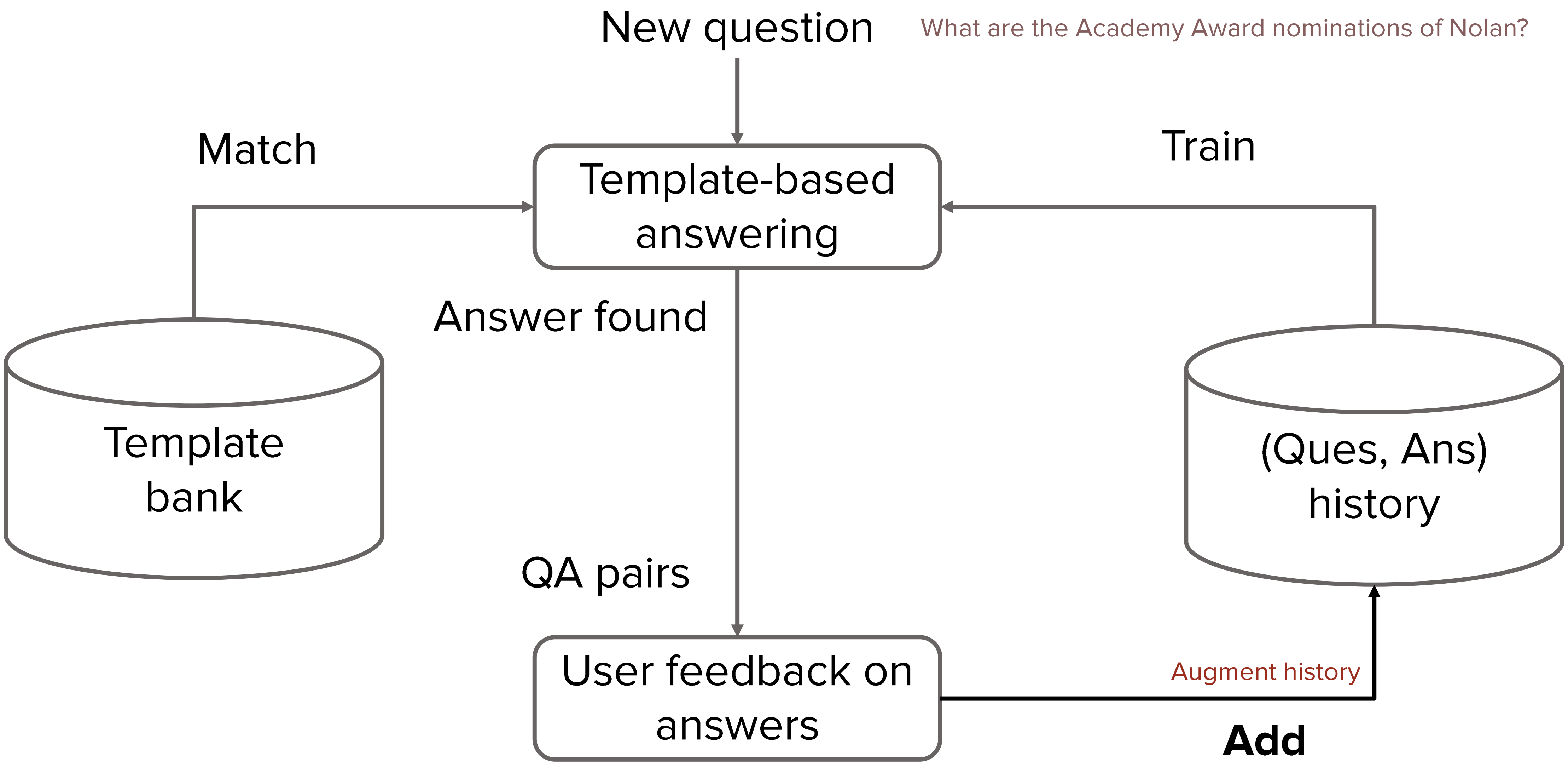


Abujabal et al., Automated template generation for question answering over knowledge graphs, WWW 2017.
Abujabal et al., Never-ending learning for open-domain question answering over knowledge bases, WWW 2018.

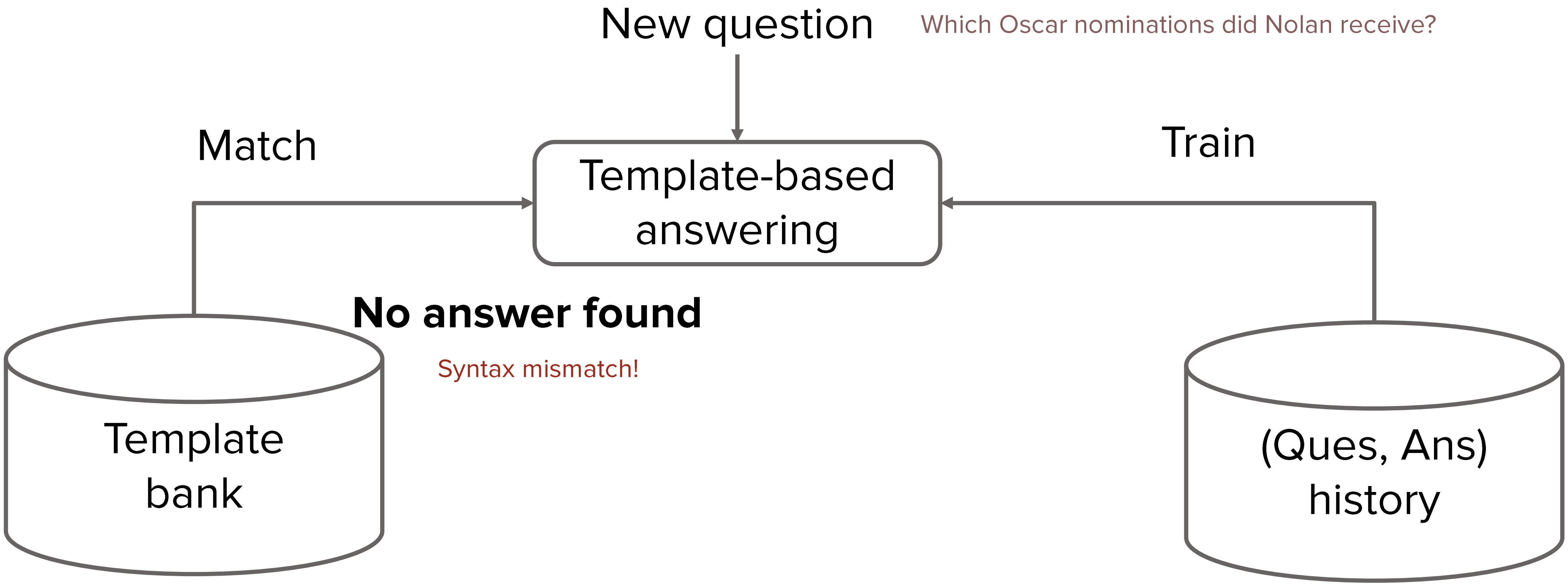
Close the loop with user feedback



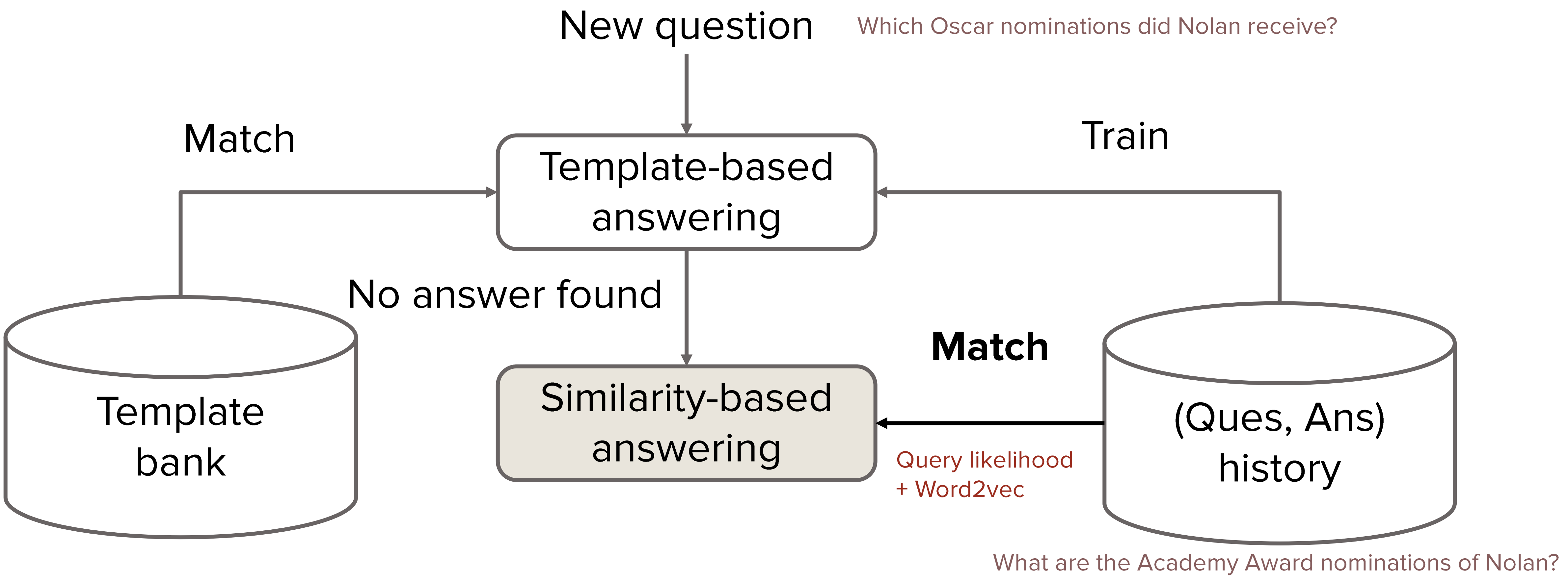
Augment history on positive feedback



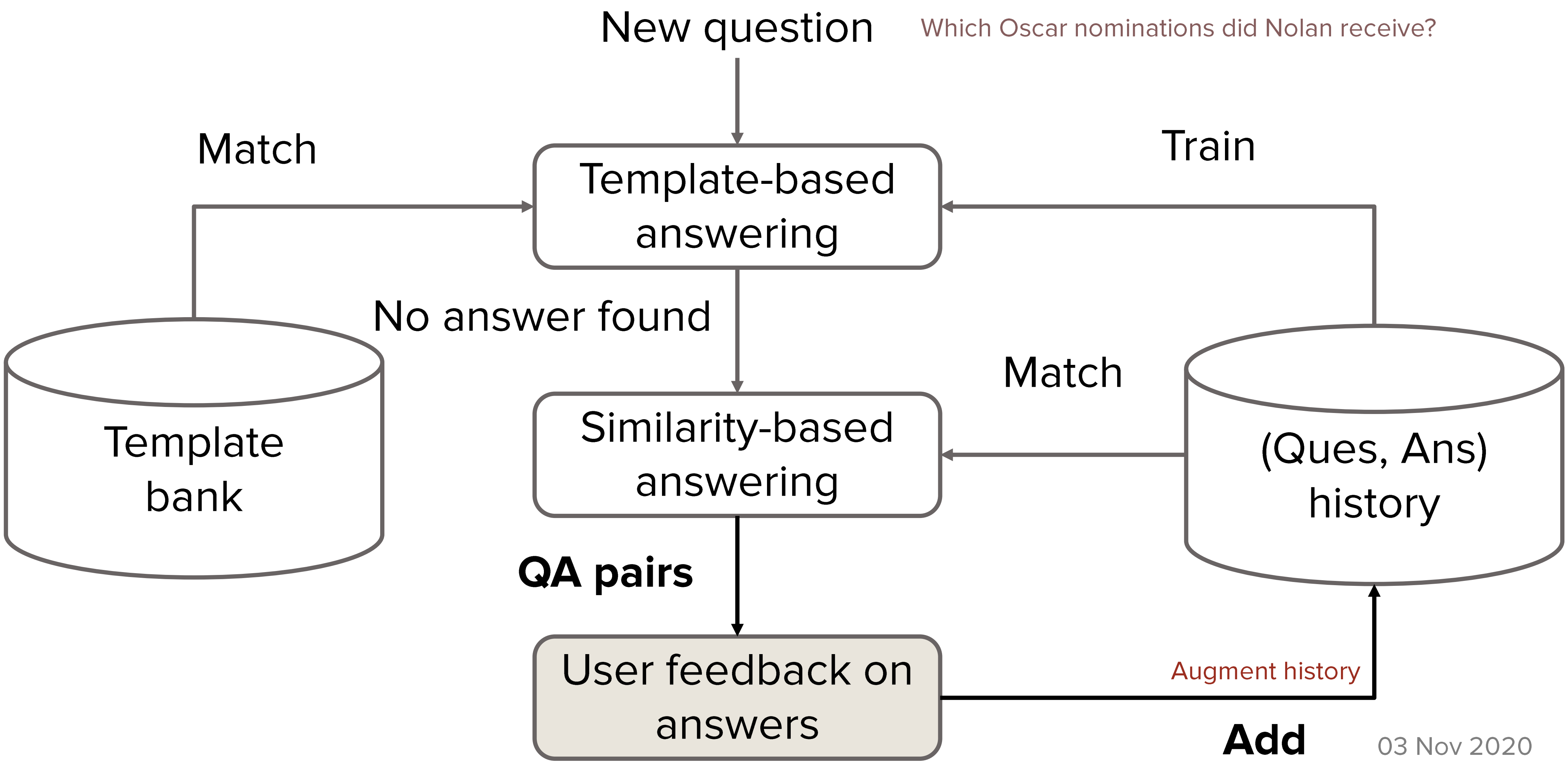
Templates can fail



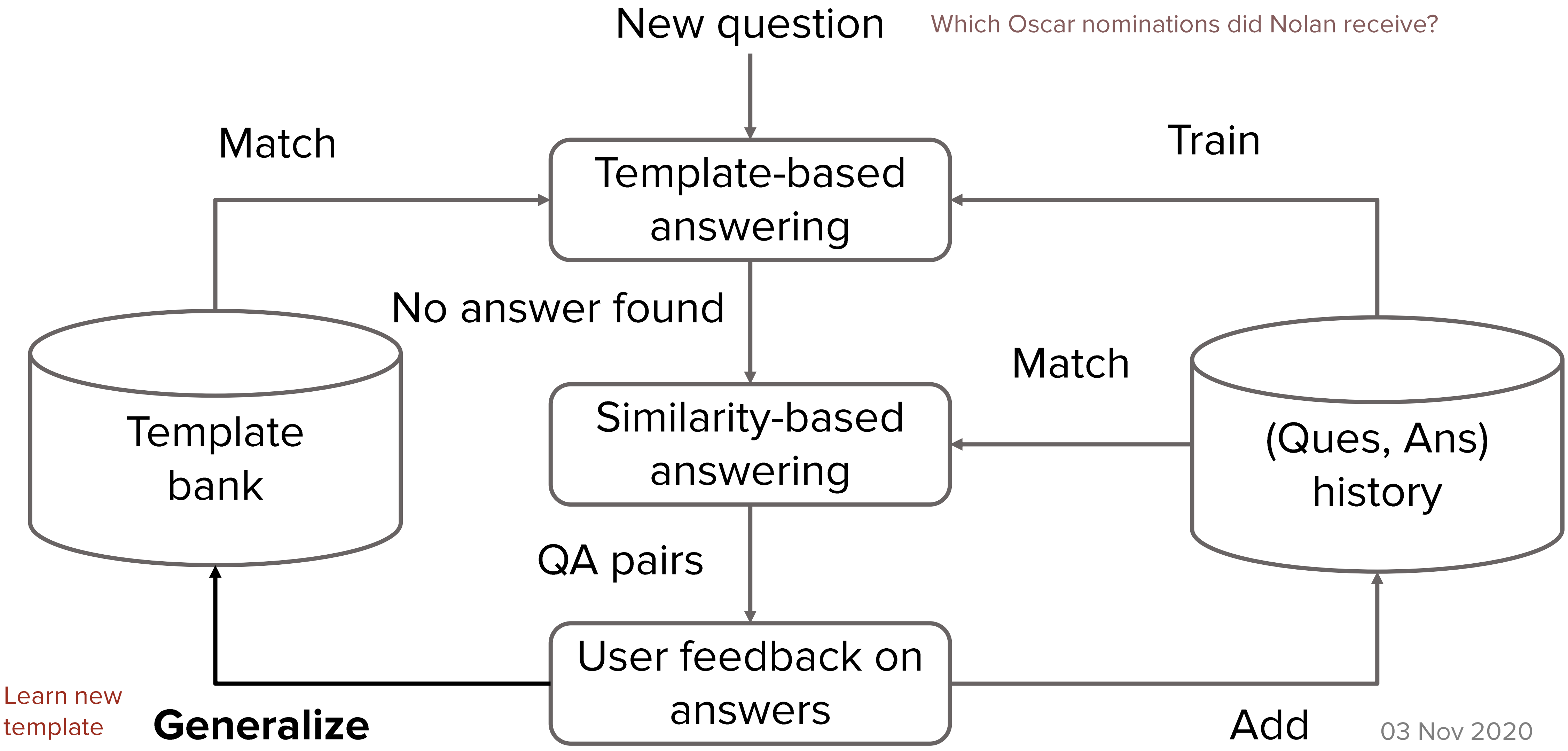
Invoke similarity-based answering



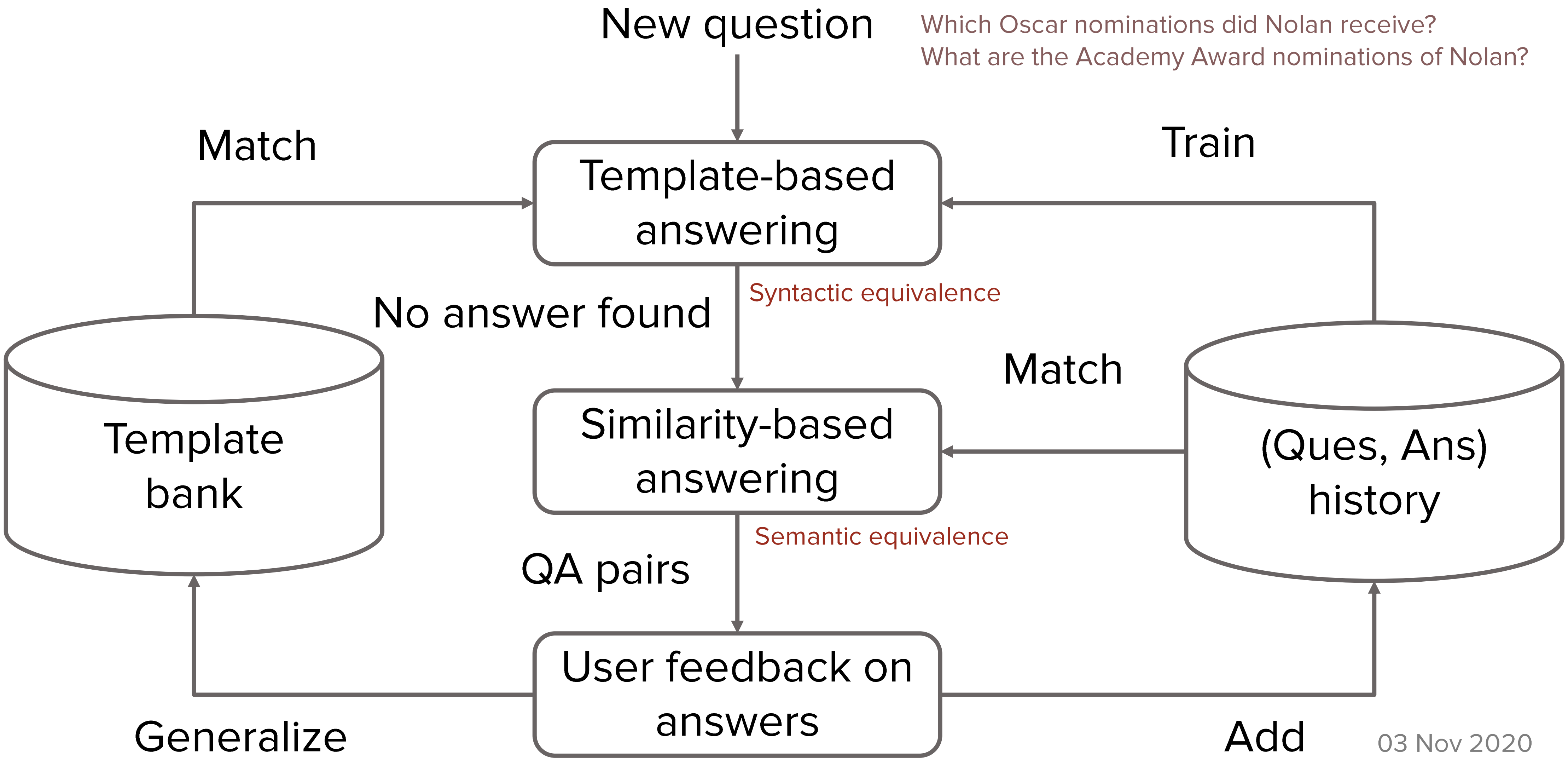
Augment history



Learn new template



Never-ending learning with NEQA



Templates: Wrap-up

- Key ideas: **Distant supervision** via shortest paths to go from (Question, Answer) to (Question, query) pair, **joint disambiguation via** Integer Linear Program
- **Template learning** also explored by Cui et al. (2017) and Hu et al. (2017)
- Works well for simple questions, but limited for **complex questions** (initial ideas in Abujabal et al. 2017, Cui et al. 2017, Hu et al. 2017)
- **Distant supervision** gets harder for complex cases
- **Similarity functions** and feedback extending scope of templates useful beyond QA?
- **Feedback in QA** subsequently investigated in QApedia (Kratzwald and Feuerriegel 2019) and IMPROVE-QA (Zhang et al. 2019)

QA with graph embeddings

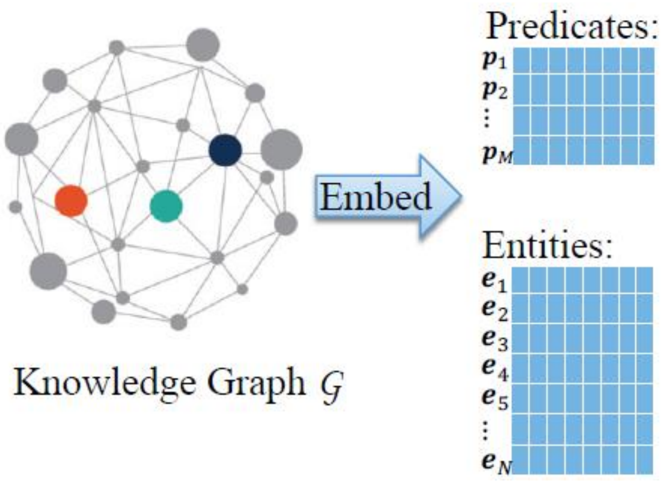
- The **KEQA** model (Huang et al. 2019)
- Leverages knowledge graph embeddings (+ word embeddings)
- Uses the TransE Model (or TransE-like …)
- Simple questions, no qualifiers
- Seminal work on neural QA in Bordes et al. (2014), Yih et al. (2015)

Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Outline

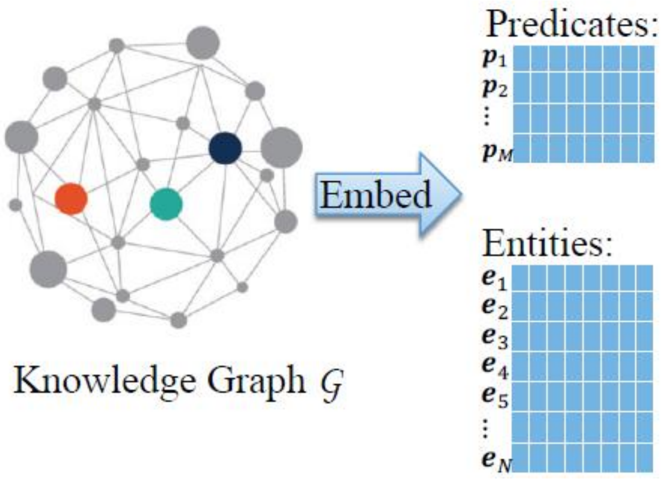
Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Learn KG embeddings



Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Using TransE



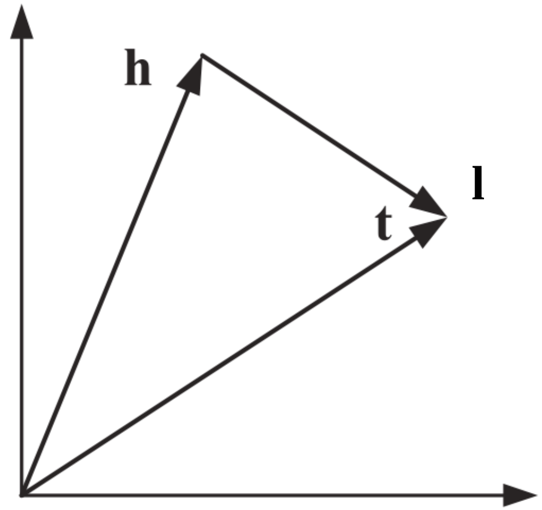
TransE (or TransE-like model) (Bordes et al. 2013)

- Head entity, predicate, tail entity
- Loss function using correct and corrupted triples

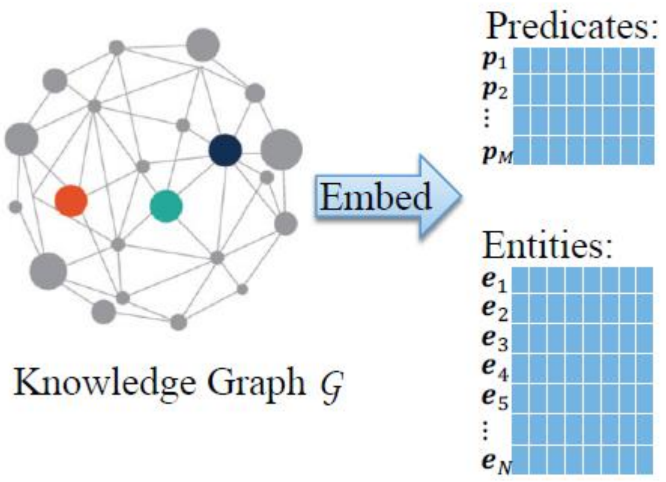
$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$

$$S'_{(h,\ell,t)} = \left\{(h',\ell,t)|h' \in E\right\} \cup \left\{(h,\ell,t')|t' \in E\right\}$$

- L2-norm of entity embeddings 1, predicates unconstrained



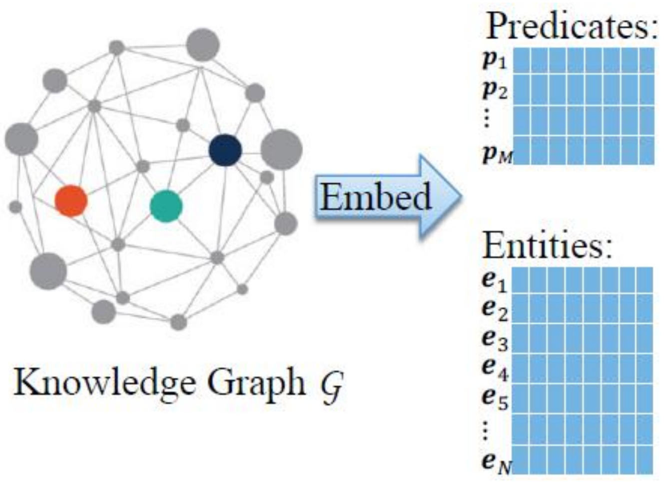
KEQA: Input question

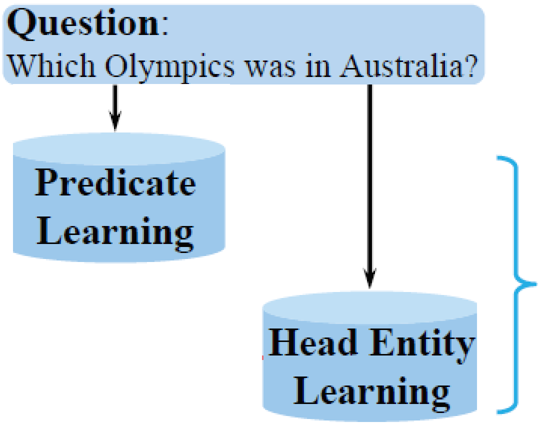


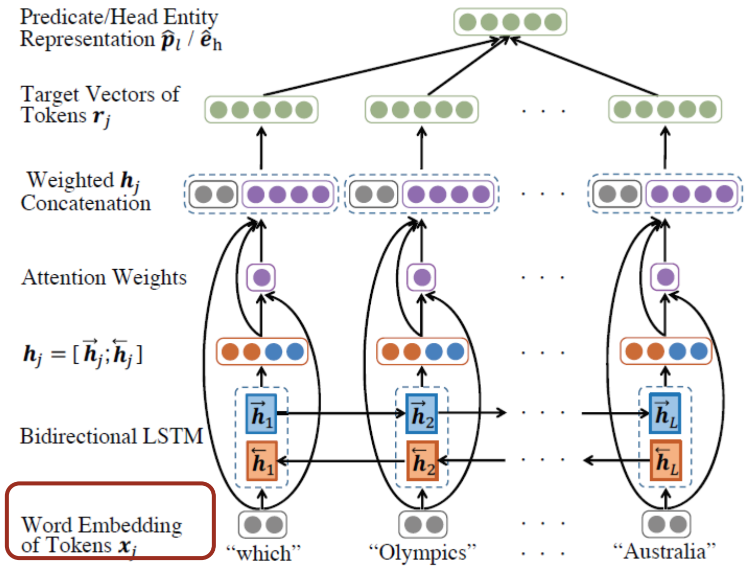
Question:
Which Olympics was in Australia?

Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Learn to predict head and body





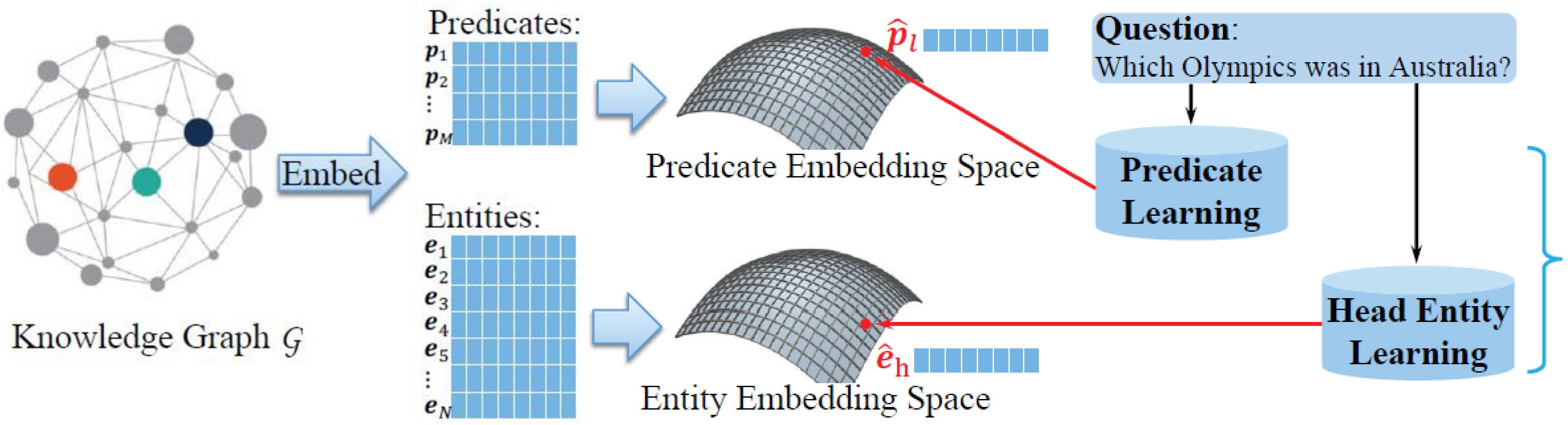


Bi-LSTM for word order

Attention for word importance

Learning representations generalization to unseen predicates at test time

KEQA: Use learnt models for prediction

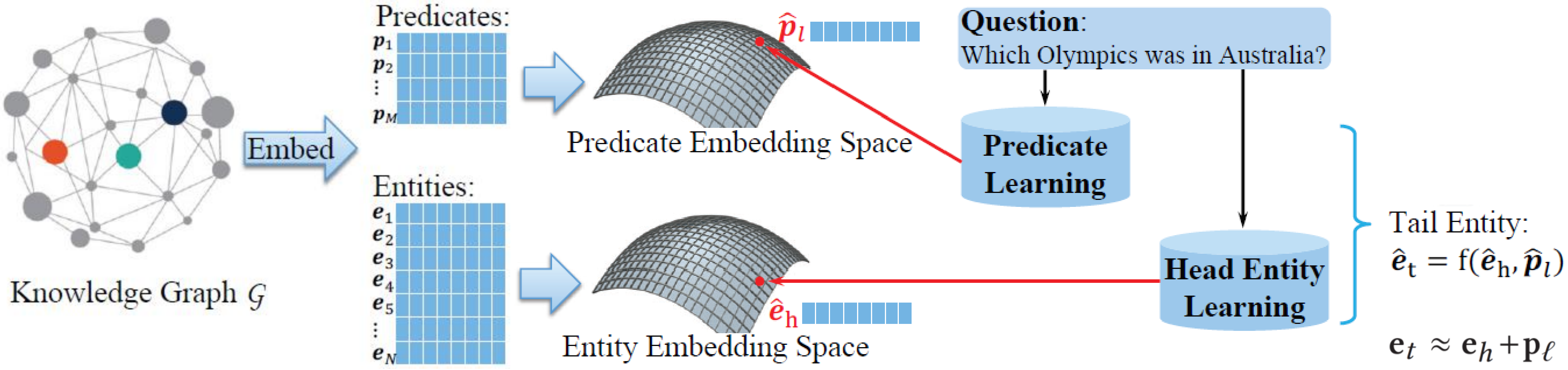


TransE (or TransE-like model) (Bordes et al. 2013)

- Head entity, predicate, tail entity
- L2-norm of entity embeddings 1, predicates unconstrained

Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Obtain tail from head and body



$\mathbf{e}_t \approx \mathbf{e}_h + \mathbf{p}_\ell$

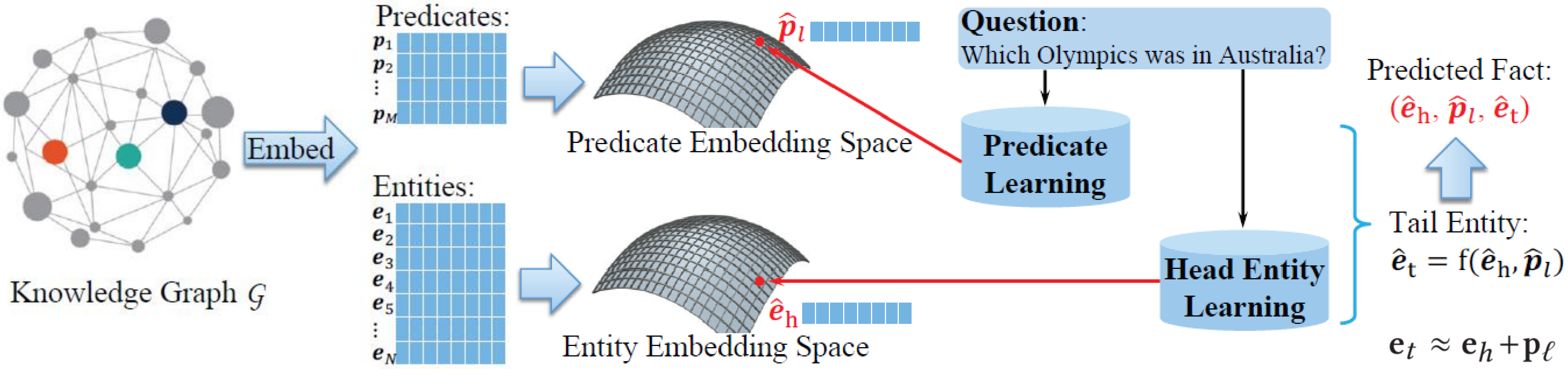
For TransE

TransE (or TransE-like model) (Bordes et al. 2013)

- Head entity, predicate, tail entity
- L2-norm of entity embeddings 1, predicates unconstrained

Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Put (head, body, tail) together

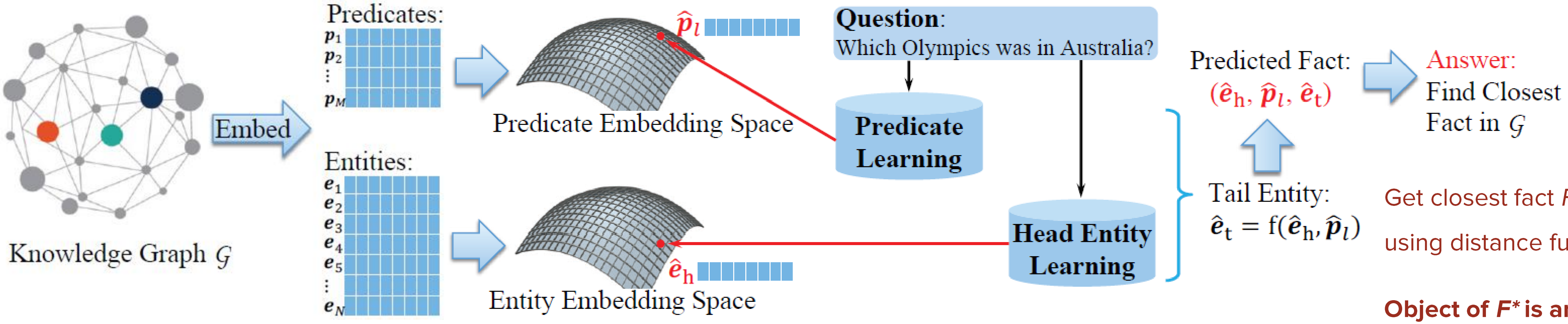


TransE (or TransE-like model) (Bordes et al. 2013)

- Head entity, predicate, tail entity
- L2-norm of entity embeddings 1, predicates unconstrained

Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Search for closest fact in KG



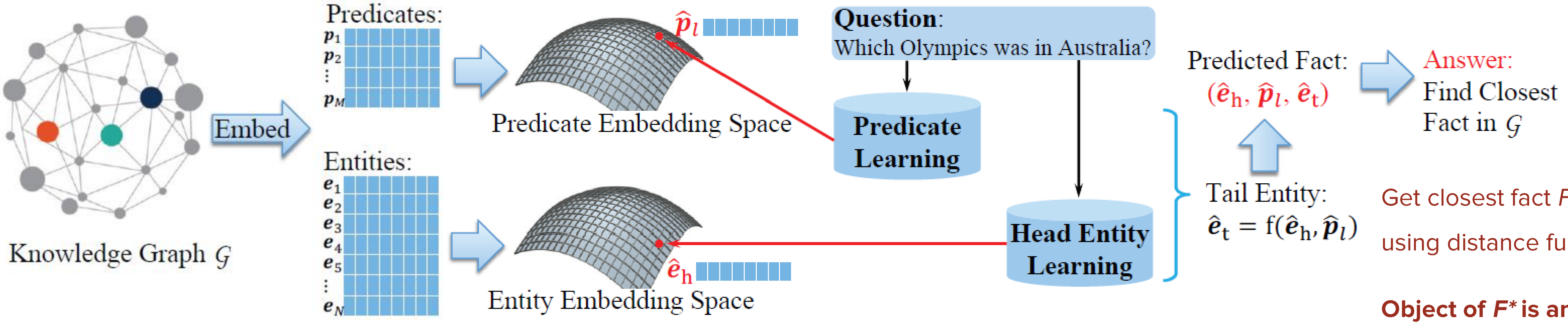
Get closest fact F^* in KG using distance function

Object of F^* is answer

$$\underset{(h,\ell,t)\in C}{\text{minimize}} \quad \|\mathbf{p}_\ell - \hat{\mathbf{p}}_\ell\|_2 + \beta_1 \|\mathbf{e}_h - \hat{\mathbf{e}}_h\|_2 + \beta_2 \|f(\mathbf{e}_h, \mathbf{p}_\ell) - \hat{\mathbf{e}}_t\|_2$$

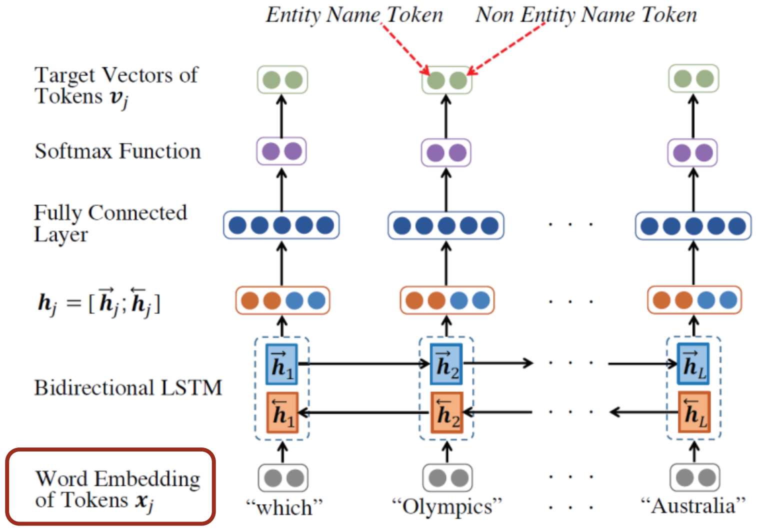
Huang et al., Knowledge graph embedding based question answering, WSDM 2019.

KEQA: Closest fact to answer



Get closest fact F^* in KG using distance function

Object of F^* is answer



$$\underset{(h,\ell,t)\in C}{\text{minimize}} \quad \|\mathbf{p}_\ell - \hat{\mathbf{p}}_\ell\|_2 + \beta_1 \|\mathbf{e}_h - \hat{\mathbf{e}}_h\|_2 + \beta_2 \| f(\mathbf{e}_h, \mathbf{p}_\ell) - \hat{\mathbf{e}}_t \|_2$$
$$- \beta_3 sim[n(h), \text{HED}_{\text{entity}}] - \beta_4 sim[n(\ell), \text{HED}_{\text{non}}],$$

Incorporate string similarity

Addl. neural model for head entity detection (HED)

Embeddings: Wrap-up

- Graph embeddings useful for simple questions, not clear for complex cases
- Embeddings and neural methods are ubiquitous now
- Much more than using pre-trained embeddings
- Leveraging sequence models (Bi-LSTMs, transformers) with attention

break duration ?x .
?x measuredIn minutes .

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

How can we answer more complex questions with multiple entities and predicates?

Complex questions

- Two basic types
  - Star joins
    - Who played for Barcelona and Real Madrid?
  - Chain joins
    - What is the profession of Messi’s father?

```
SELECT ?x WHERE
?x playedFor Barcelona .
?x playedFor RealMadrid .
```

Single variable

```
SELECT ?y WHERE
?x fatherOf Messi .
?x profession ?y .
```

Two or more variables

Complex questions

- **Much more:** Aggregations, comparatives, superlatives, reasoning, existential, temporal, ….
- Focus on **substructures** in questions and queries (Bhutani et al. 2019, Ding et al. 2019, Sun et al. 2020)
- Often rely on question **decomposition** (Bao et al. 2016, Talmor and Berant 2018, Sun et al. 2020)
- **Joint disambiguation** of question concepts (Yahya et al. 2012, Lu et al. 2019)

Which female **actor played in Casablanca and is married to** a writer who was born in Rome?

Where is the **founder of Tesla born**?

Who was the **second wife** of Tom Cruise?

Which **Portuguese speaking countries** import **fish from Brazil**?

Who wrote **more books:** Enid Blyton or Agatha Christie?

Which is the **third highest** mountain in Asia?

How many **movies have the same director** as The Shawshank Redemption?

How many movies were directed by the graduate of Burbank High School?

Did any cosmonauts die in the same place they were born in?

Complex questions

- Early efforts in Yahya et al. (2012)
- Further explorations in Bao et al. (2016), Abujabal et al. (2017) and Cui et al. (2017)
- Dedicated methods for complex questions in Ding et al. (2019), Hu et al. (2018), Luo et al. (2018), **Bhutani et al. (2019)**, **Lu et al. (2019)**, **Vakulenko et al. (2019)**, …

Which female **actor played in Casablanca and is married to** a writer who was born in Rome?

Where is the **founder of Tesla born**?

Who was the **second wife** of Tom Cruise?

Which **Portuguese speaking countries** import **fish from Brazil**?

Who wrote **more books:** Enid Blyton or Agatha Christie?

Which is the **third highest** mountain in Asia?

How many **movies have the same director** as The Shawshank Redemption?

How many movies were directed by the graduate of Burbank High School?

Did any cosmonauts die in the same place they were born in?

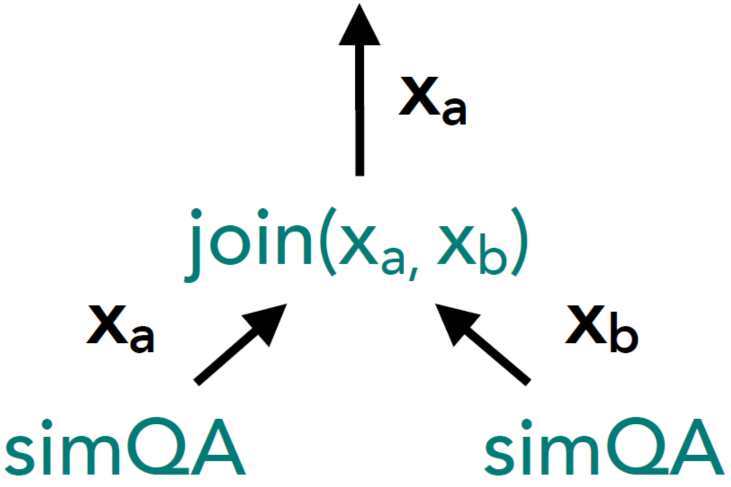
Complex QA: Structured query generation

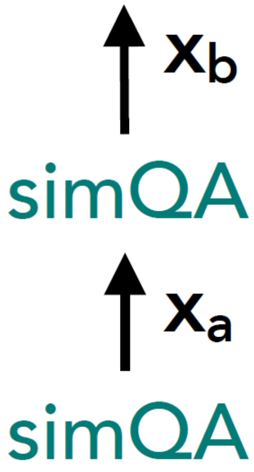
- The **TextRay** system (Bhutani et al. CIKM 2019)
- Learning complex query patterns difficult for **data sparsity**
- **Decompose-execute-join** approach to complex questions
- Constructs complex query patterns using **simple queries**
- **Semantic matching** model learns simple queries

  using **distant supervision** from QA pairs

Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Computation plan





Predict computation plan upfront with supervised method

Or

with linguistic cues

Single variable

Star join

SELECT ?x WHERE
?x playedFor Barcelona .
?x playedFor RealMadrid .

Two or more variables

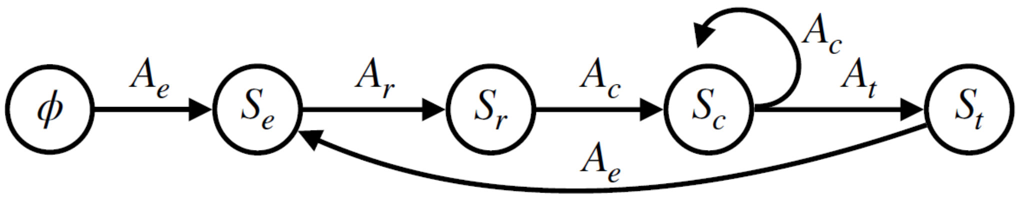
Chain join

SELECT ?y WHERE
?x fatherOf Messi .
?x profession ?y .

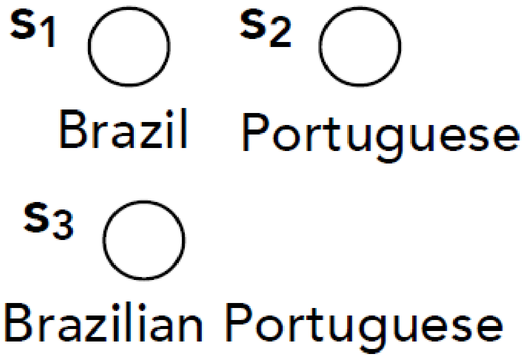
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Walkthrough

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)



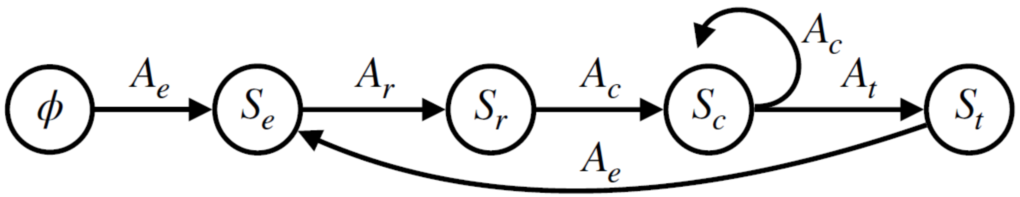
a) Identify seed

Top-k entities

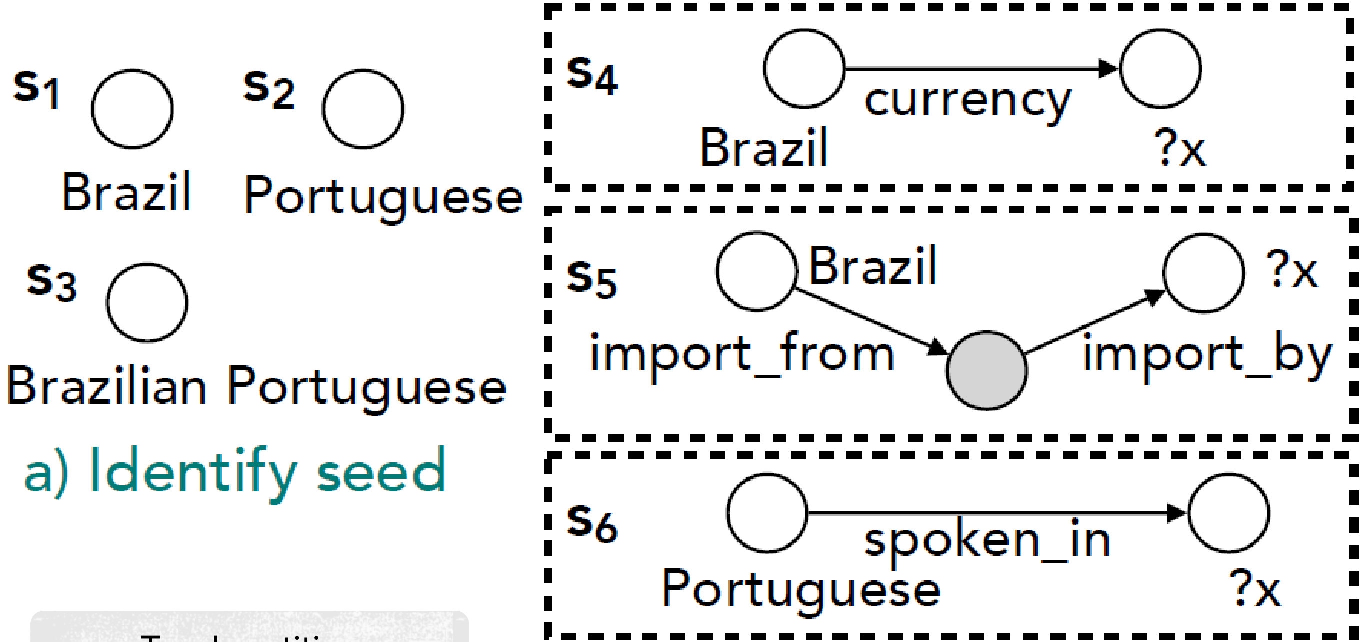
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Partial query graph

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)



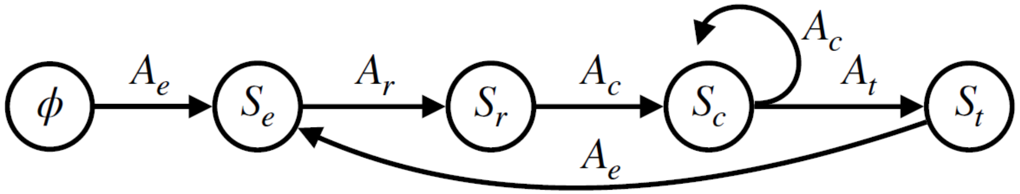
Top-k entities

b) Identify main relation path

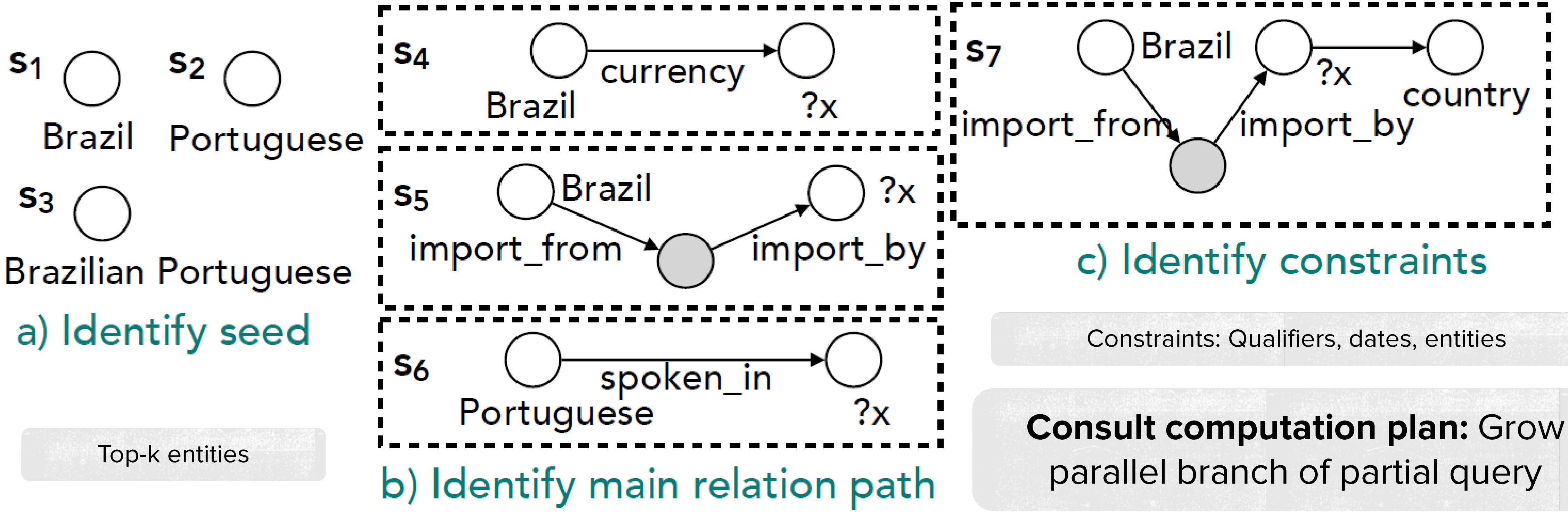
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Partial query graph

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)



Top-k entities

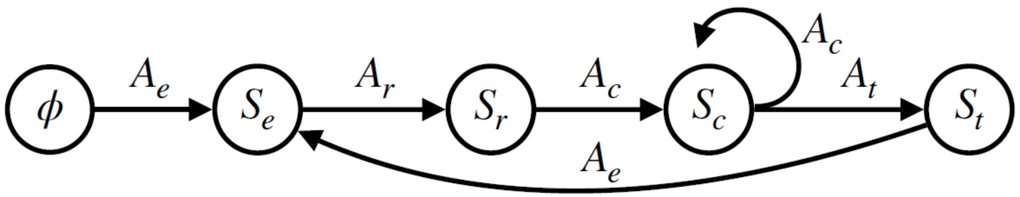
Constraints: Qualifiers, dates, entities

Consult computation plan: Grow parallel branch of partial query

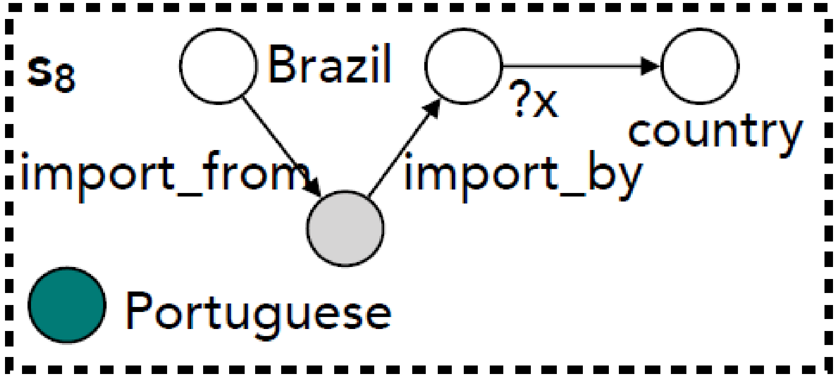
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Partial query graph

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)

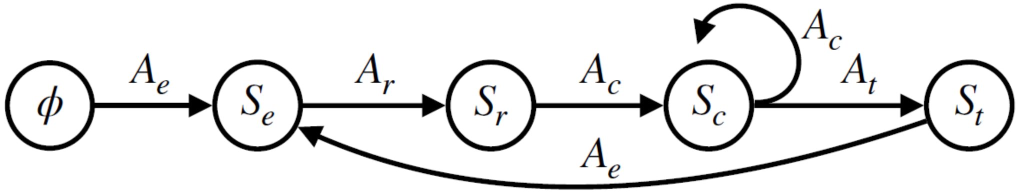


a) Identify seed

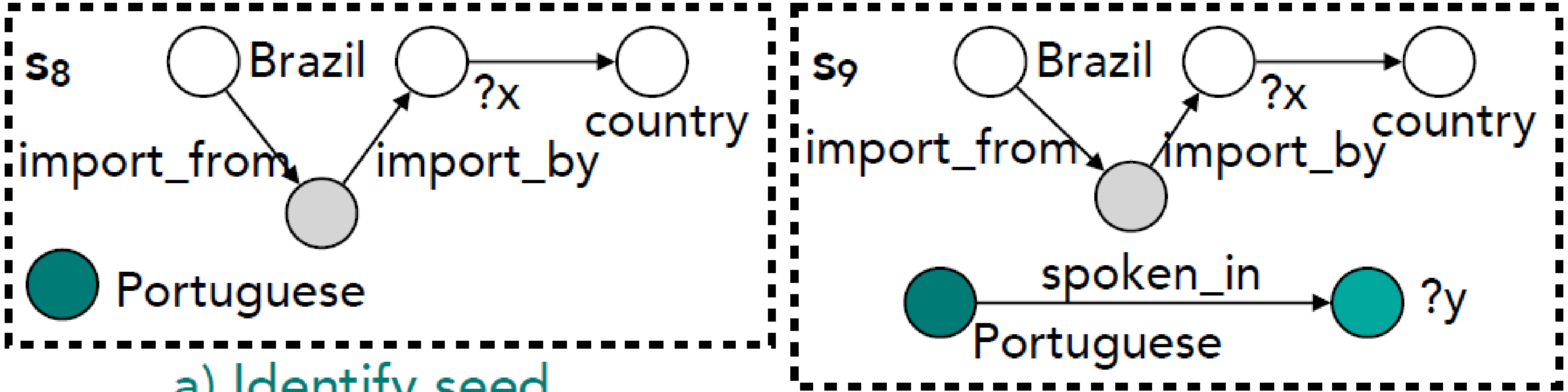
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Partial query graph

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)



a) Identify seed

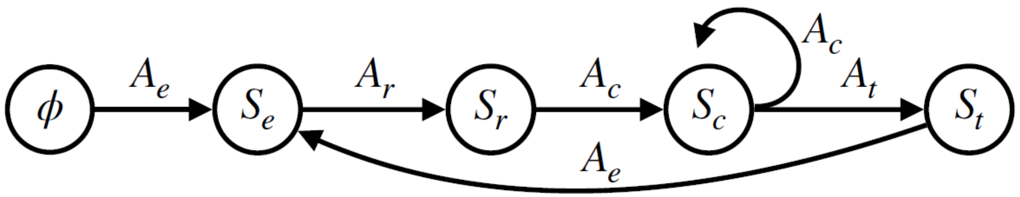
b) Identify main relation path

c) Identify constraints

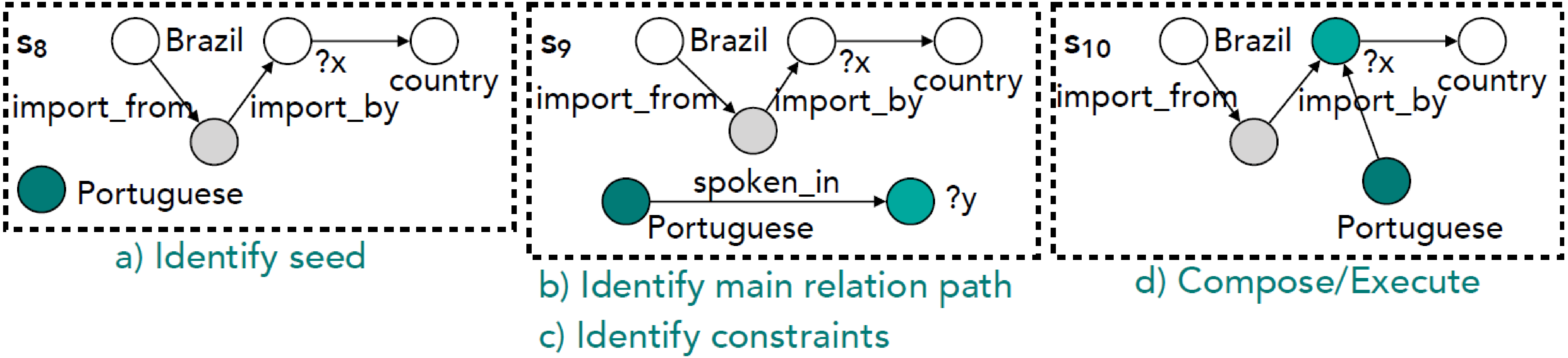
Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

TextRay: Partial query graph

Which Portuguese speaking countries import fish from Brazil?



Staged query graph generation (Yih et al. ACL 2015)



Beam search to maintain top-k best derivations + **Semantic similarity** learned via LSTMs with attention

Bhutani et al., Learning to Answer Complex Questions over Knowledge Bases with Query Composition, CIKM 2019.

Complex QA: Computing compact subgraphs

- The **QUEST** system (Lu et al. 2019)
- Works over open vocabulary quasi KGs (Bhutani et al. 2019b, Yin et al. 2015, Fader et al. 2013, 2014)
- Augment quasi KGs with alignments and types
- Spot question cornerstones in quasi KG
- **Unsupervised compact subgraph computation:** Compute Group Steiner Tree (GST) with cornerstones as terminals for joint disambiguation of question concepts

Lu et al., Answering Complex Questions by Joining Multi-Document Evidence with Quasi Knowledge Graphs, SIGIR 2019.

Creating a quasi KG

<Nolan, directed, Inception>
<Inception, won, Best Sound>
<2011 Oscars, announced, Best Sound>
<Inception, nominated, Best Actor>
<The movie Inception, missed out, Golden Globe Awards>
<Chris Nolan, director of, The movie Inception>
<Inception's script, edited by, Chris Nolan>
<Inception, lost to, The Social Network>
<Best Actor, declared at, 83rd Academy Awards>
<The Social Network, winner of, Best Screenplay>
<Golden Globes, announced, Best Screenplay>

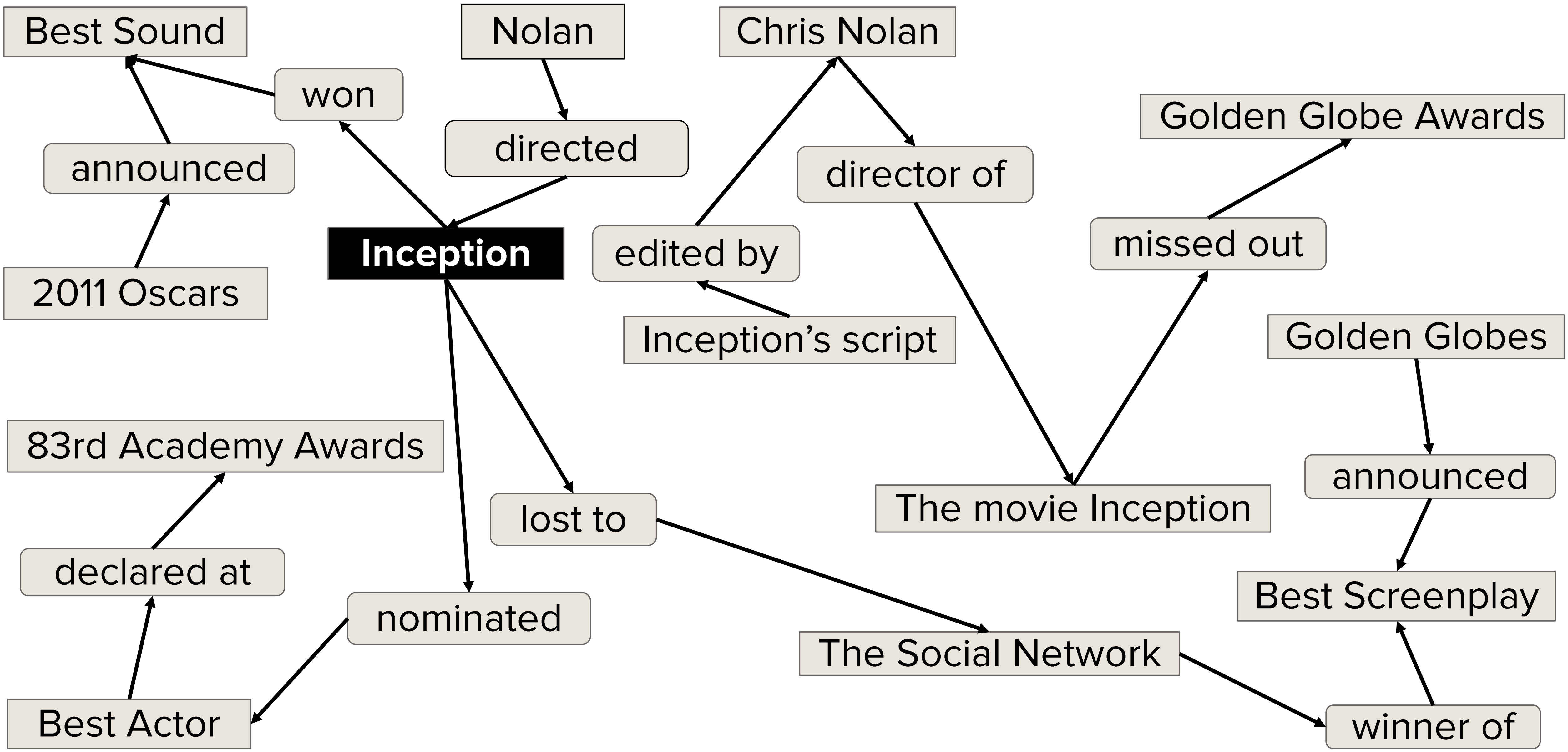
Compile an open-vocabulary triple store

Triples can ideally come from text (via Open IE), KG, or both

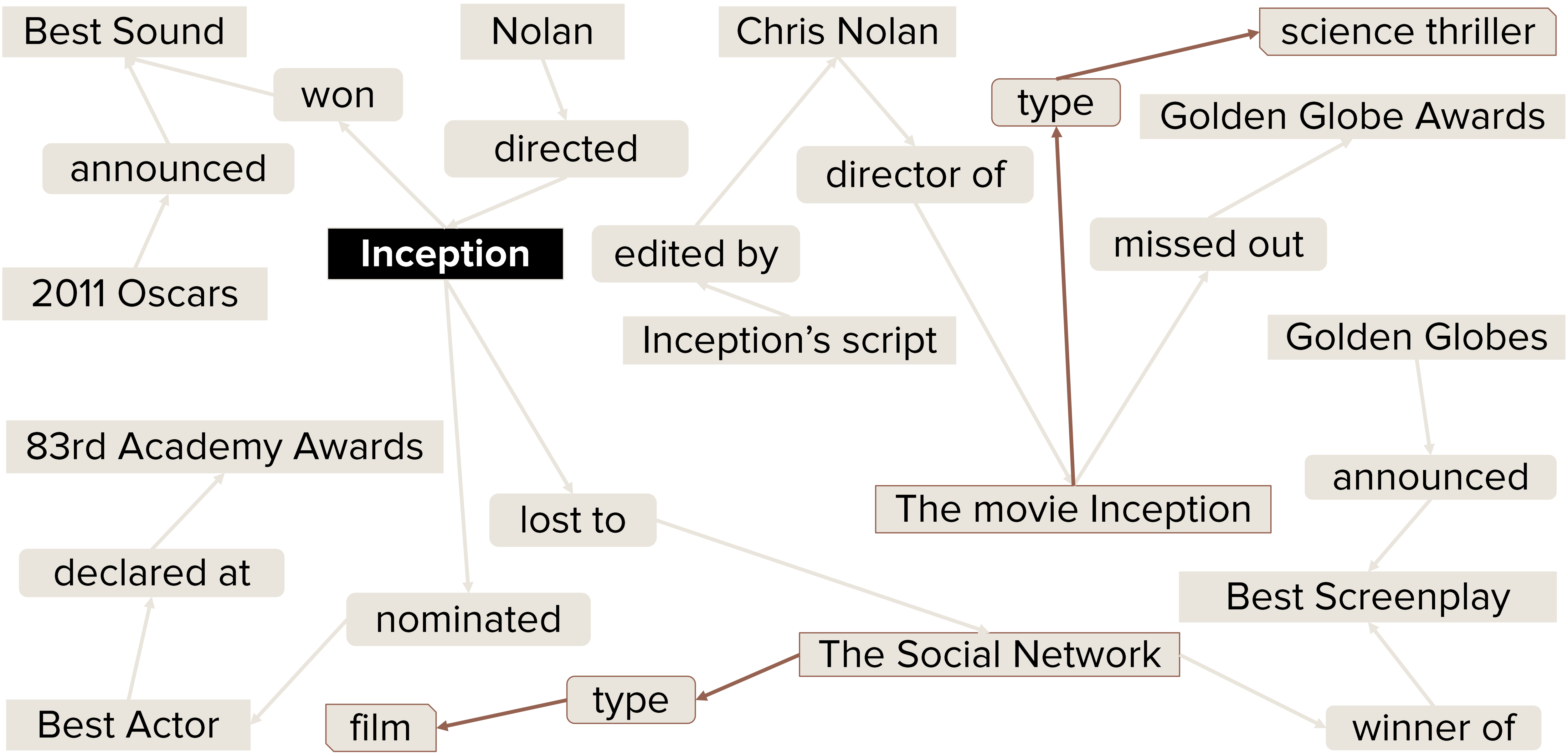
Open IE extracts KG-style triples by running pattern extraction over raw text: Stanford Open IE, ClausIE, OpenIE 5.0, ...

Lu et al., Answering Complex Questions by Joining Multi-Document Evidence with Quasi Knowledge Graphs, SIGIR 2019.

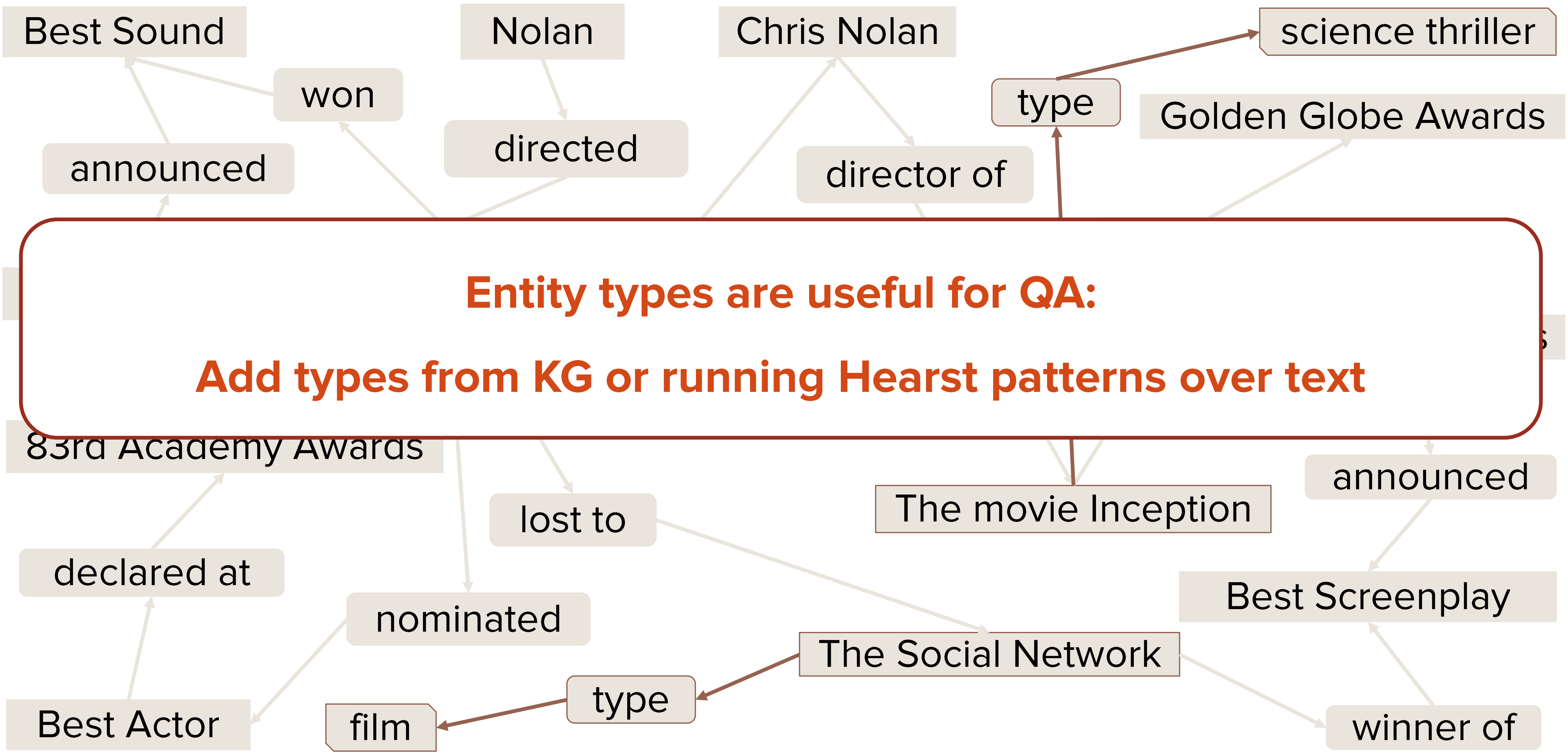
Question: Which Nolan films won an Oscar but missed a Golden Globe?



Question: Which Nolan films won an Oscar but missed a Golden Globe?



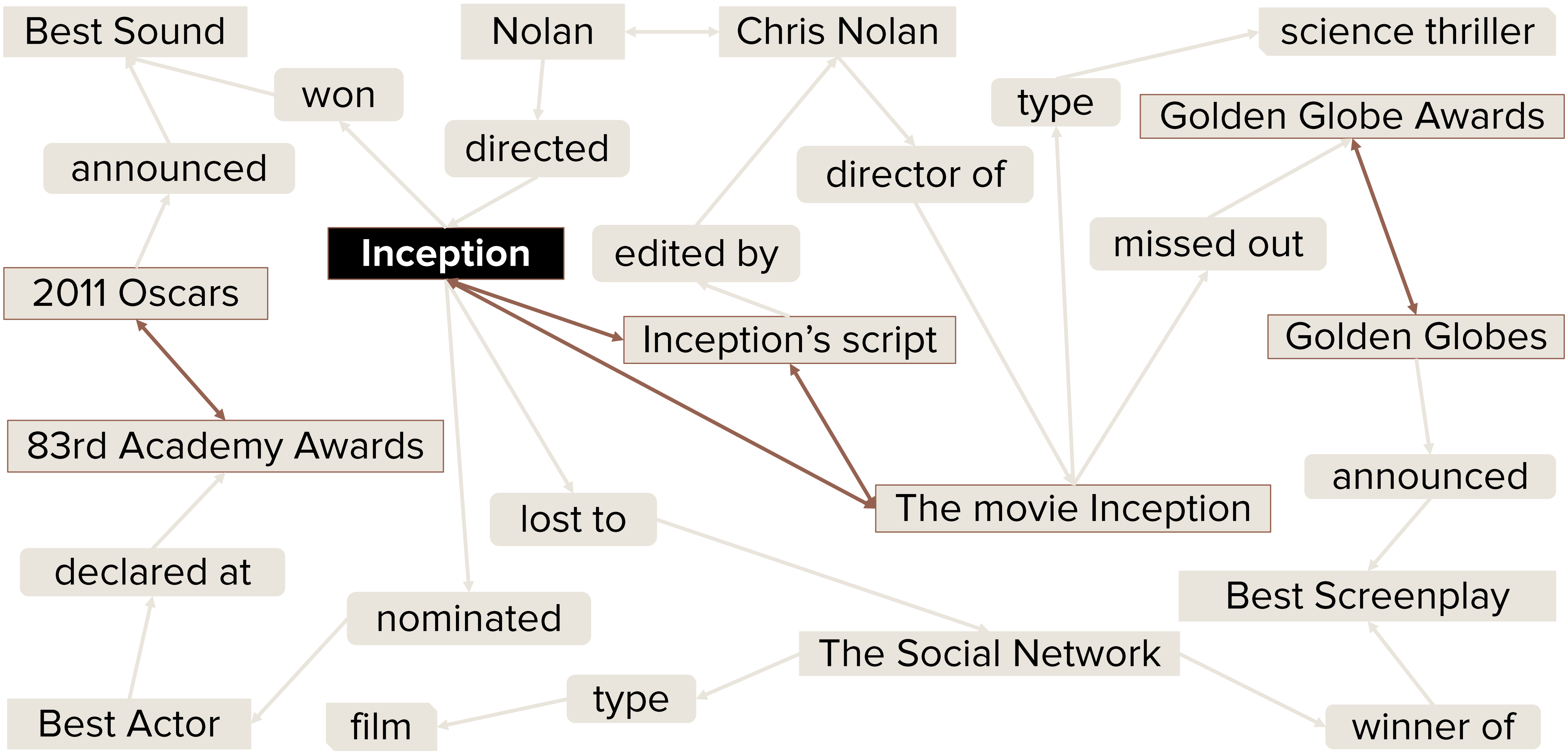
Question: Which Nolan films won an Oscar but missed a Golden Globe?



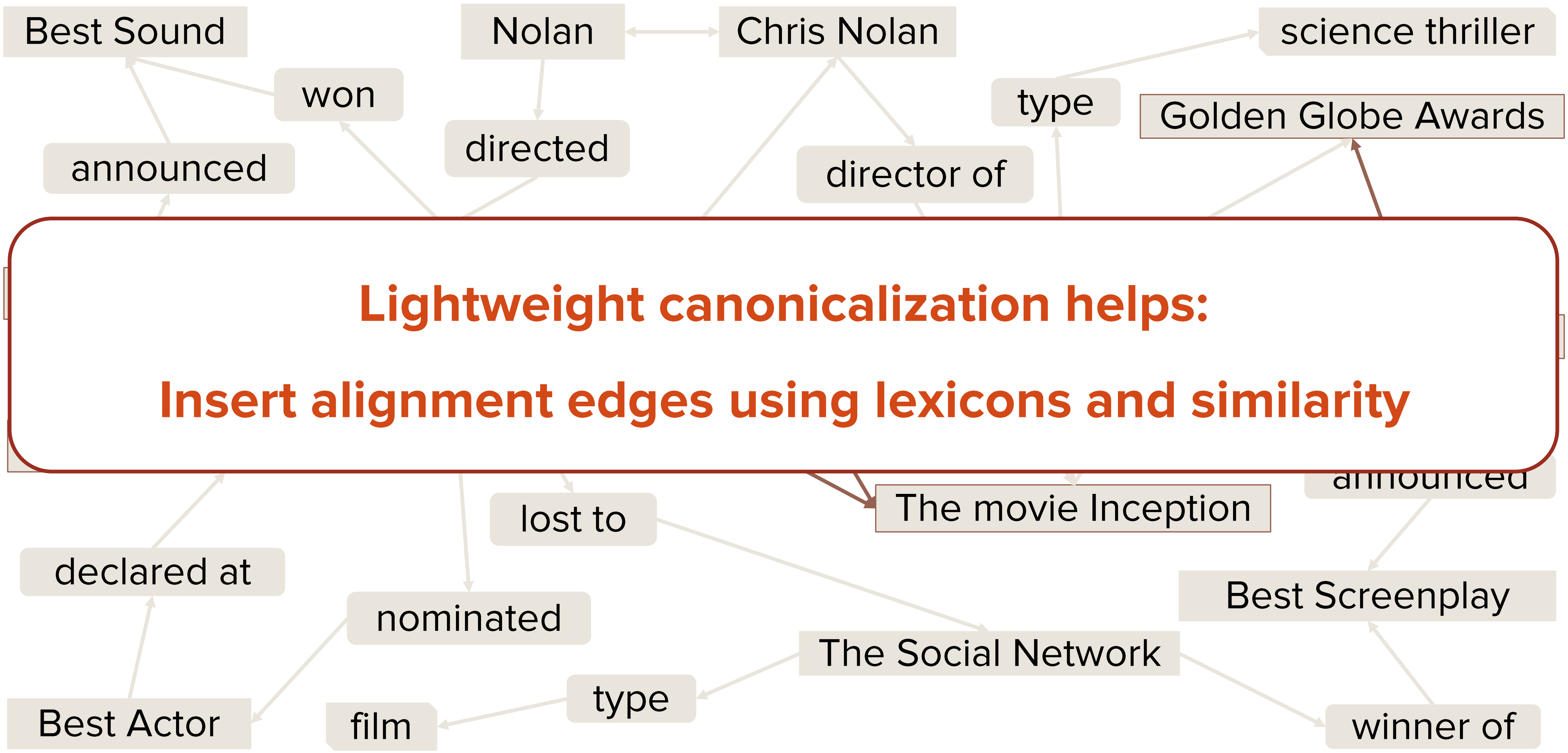
Entity types are useful for QA:

Add types from KG or running Hearst patterns over text

Question: Which Nolan films won an Oscar but missed a Golden Globe?



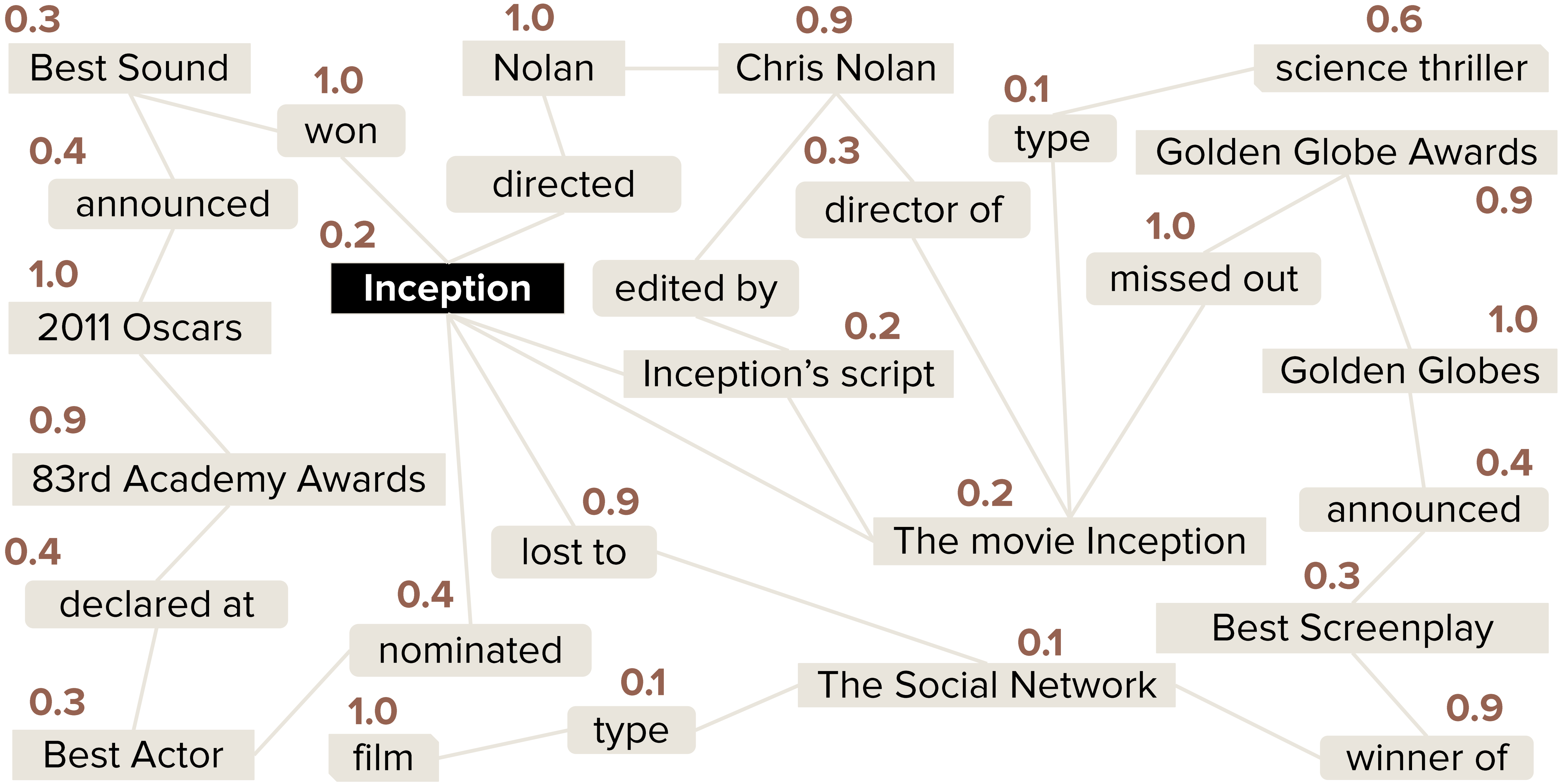
Question: Which Nolan films won an Oscar but missed a Golden Globe?



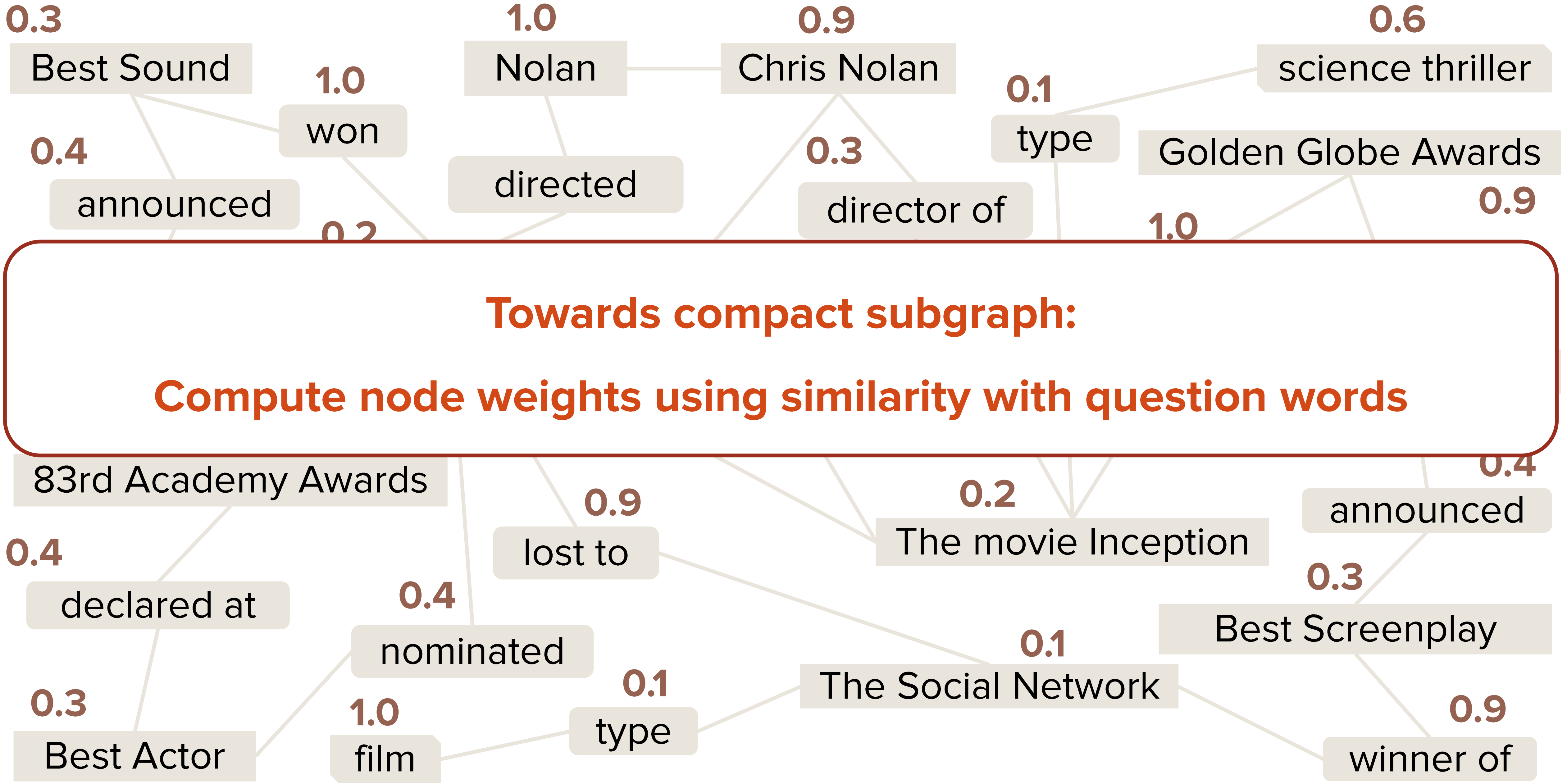
Lightweight canonicalization helps:

Insert alignment edges using lexicons and similarity

Question: Which Nolan films won an Oscar but missed a Golden Globe?

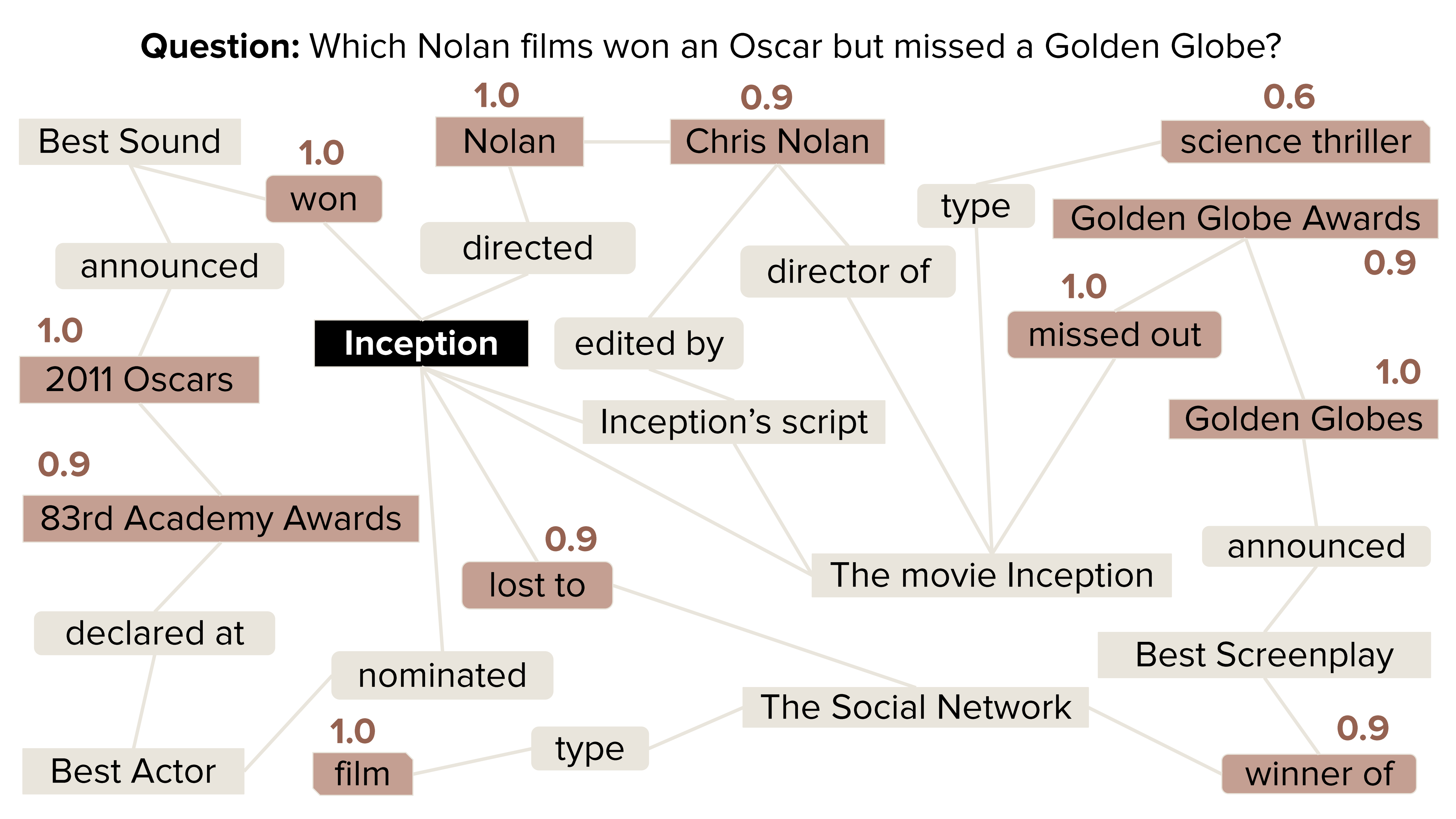


Question: Which Nolan films won an Oscar but missed a Golden Globe?

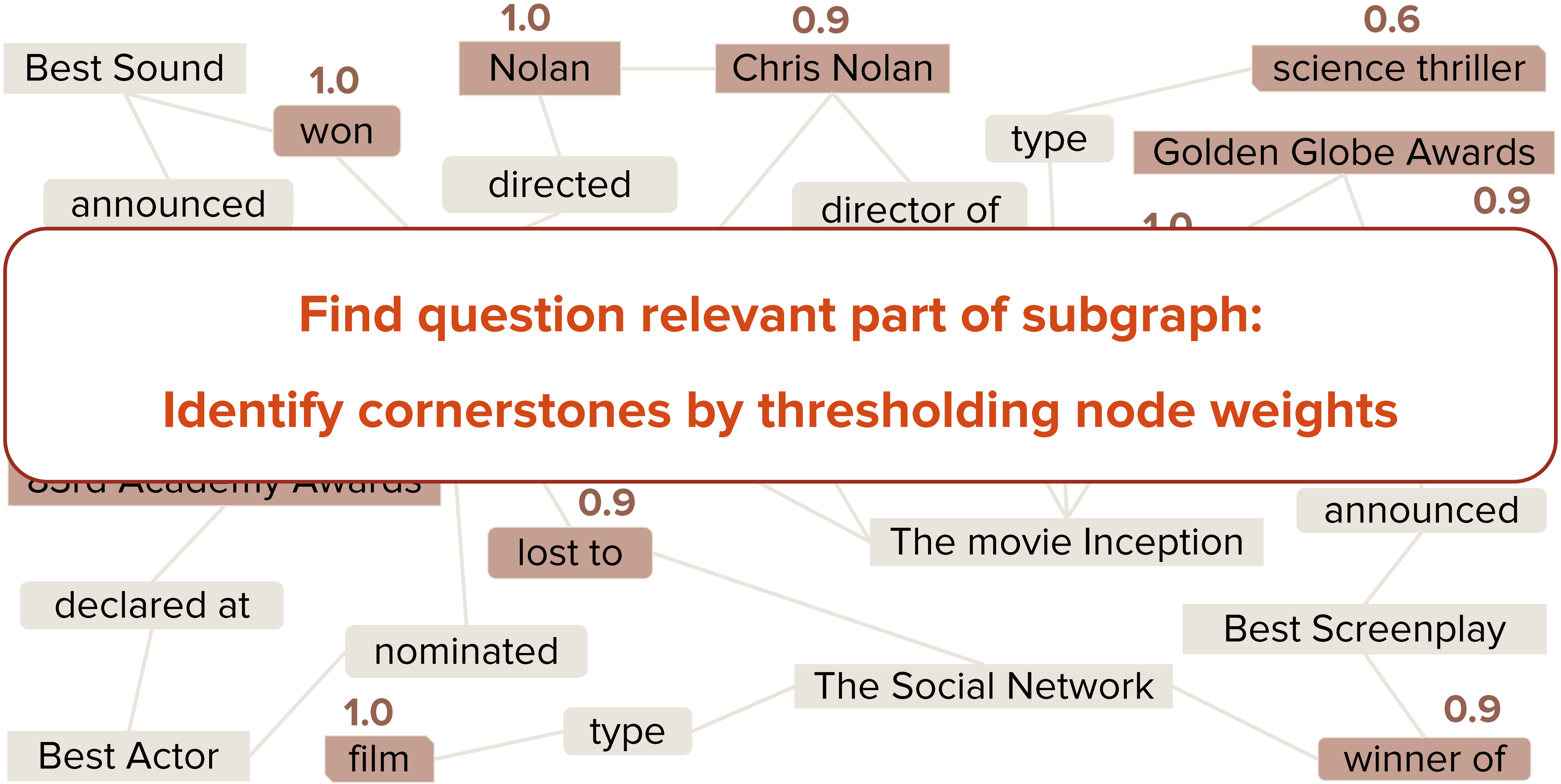


Towards compact subgraph:

Compute node weights using similarity with question words


Question: Which Nolan films won an Oscar but missed a Golden Globe?
1.0
Nolan
0.9
Chris Nolan
0.6
science thriller
Best Sound
1.0
won
directed
type
Golden Globe Awards
announced
director of
0.9
1.0
Inception
edited by
missed out
1.0
2011 Oscars
1.0
Inception’s script
Golden Globes
0.9
83rd Academy Awards
0.9
lost to
announced
The movie Inception
declared at
Best Screenplay
nominated
The Social Network
1.0
type
0.9
Best Actor
film
winner of

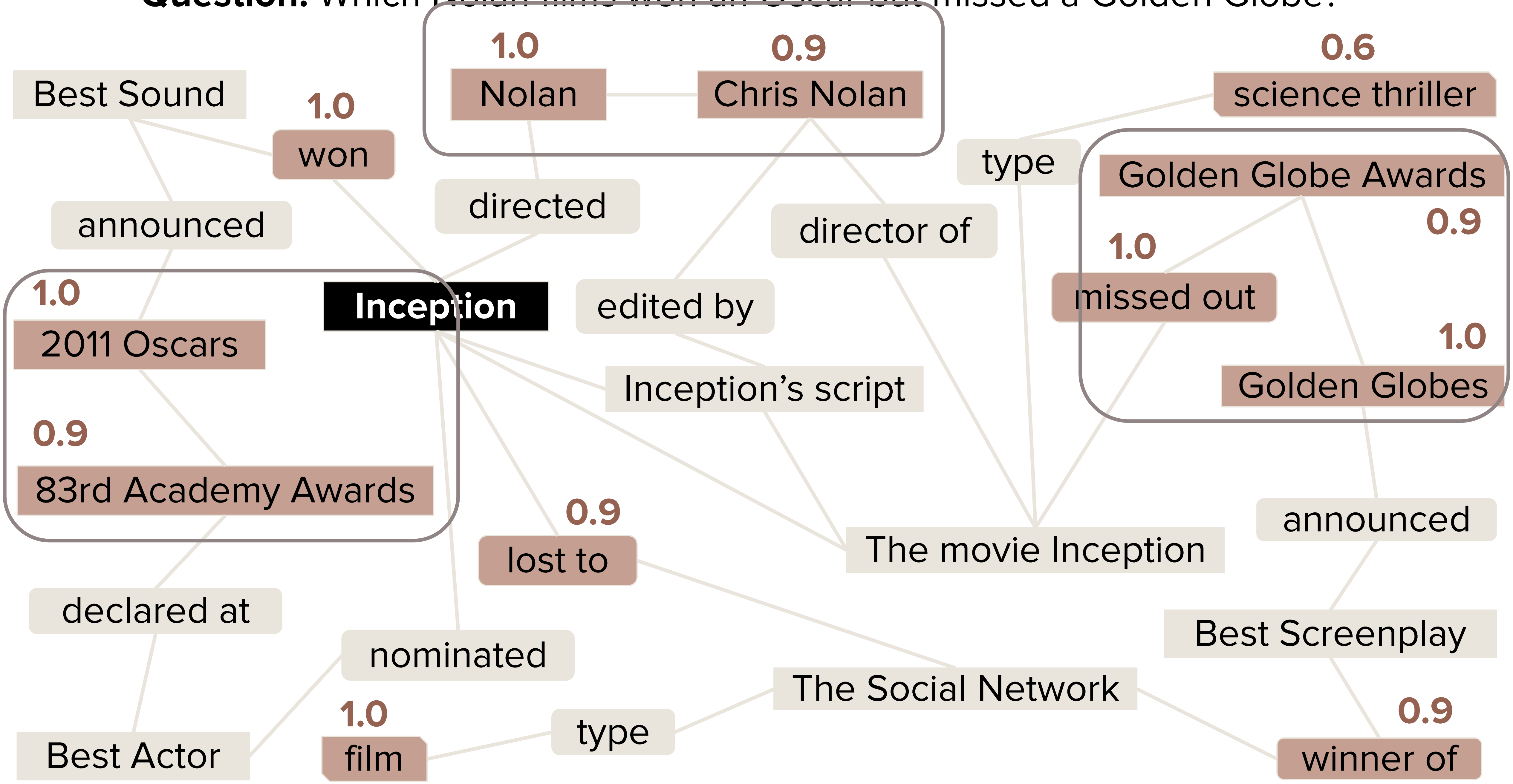
Question: Which Nolan films won an Oscar but missed a Golden Globe?



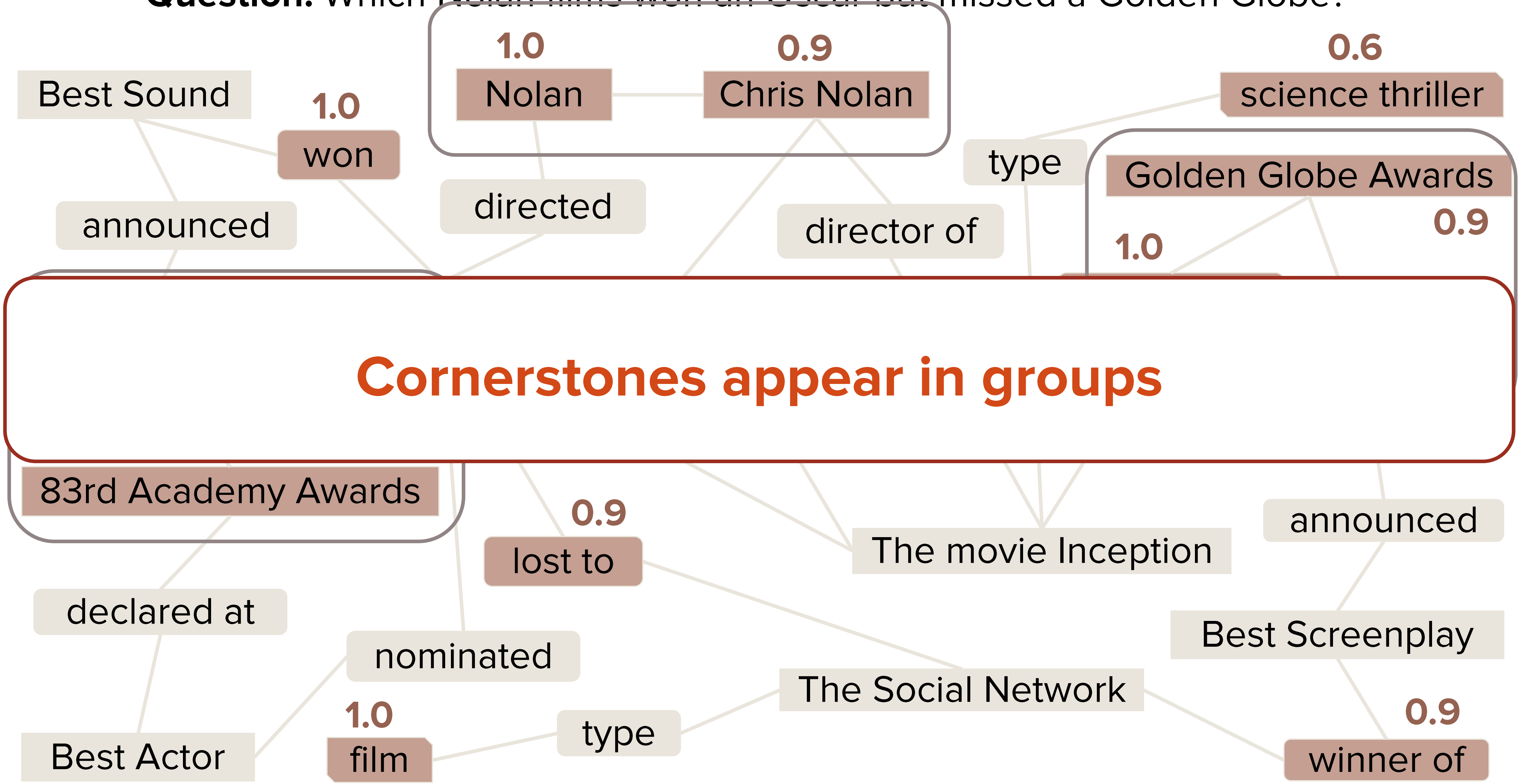
Find question relevant part of subgraph:

Identify cornerstones by thresholding node weights

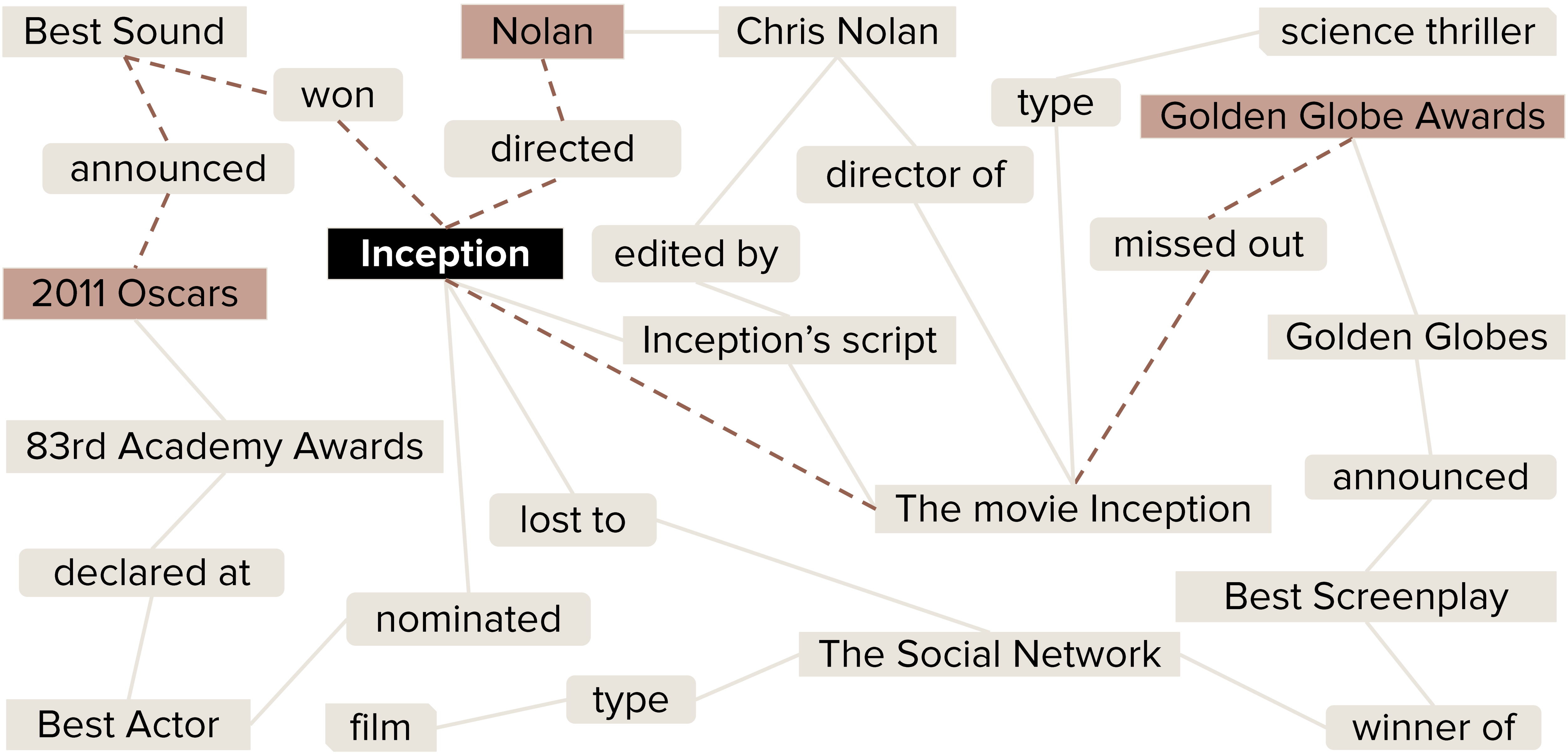
Question: Which Nolan films won an Oscar but missed a Golden Globe?



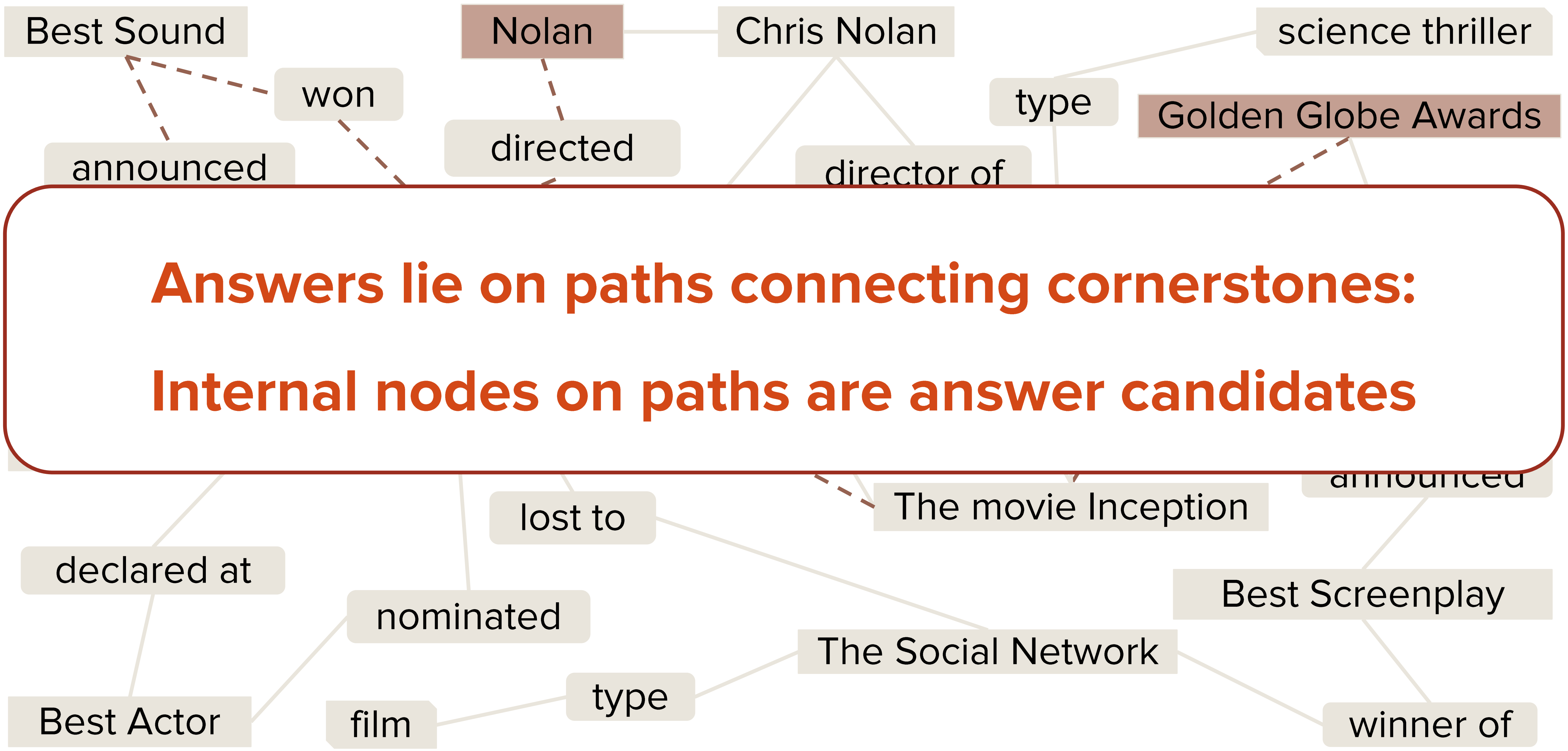
Question: Which Nolan films won an Oscar but missed a Golden Globe?



Question: Which Nolan films won an Oscar but missed a Golden Globe?



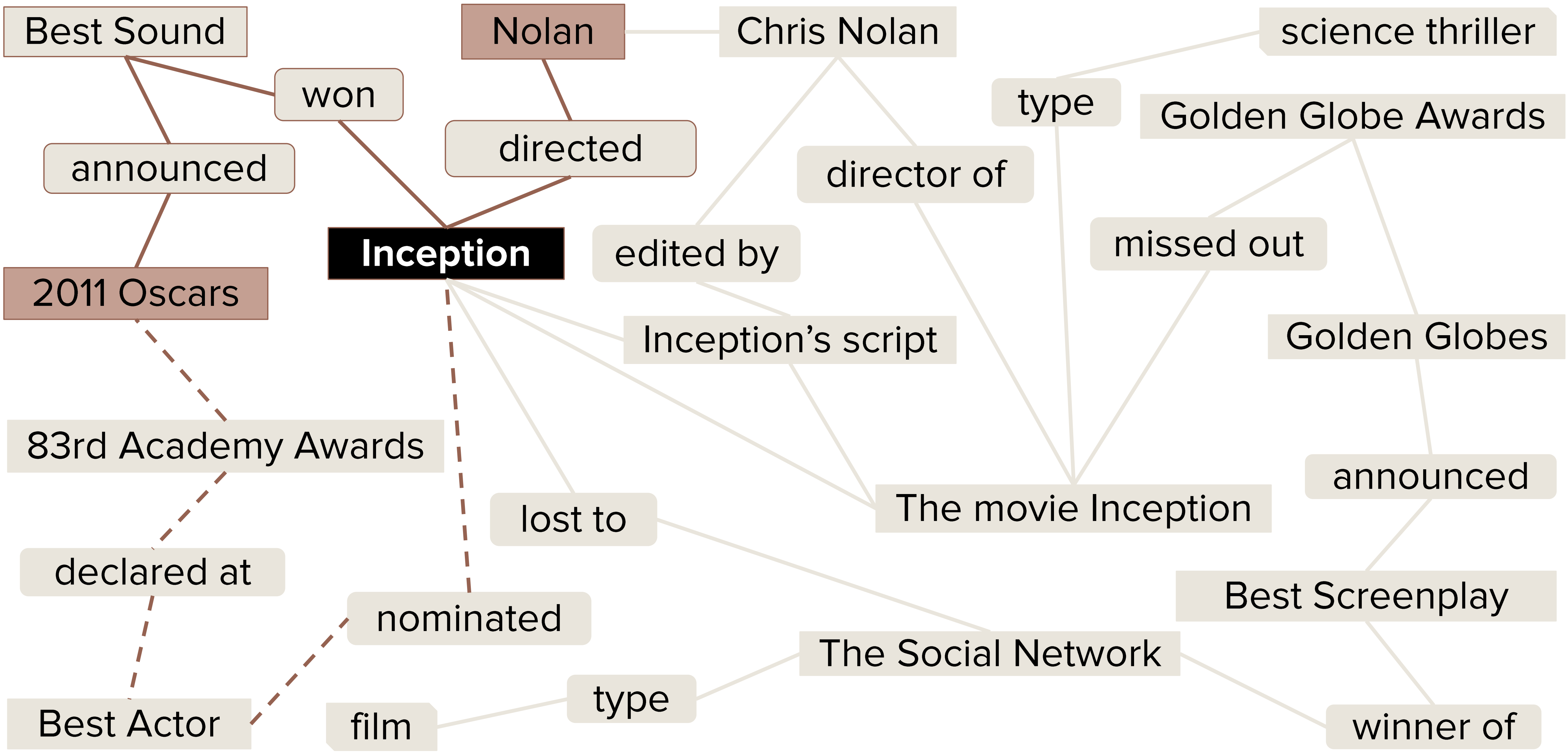
Question: Which Nolan films won an Oscar but missed a Golden Globe?



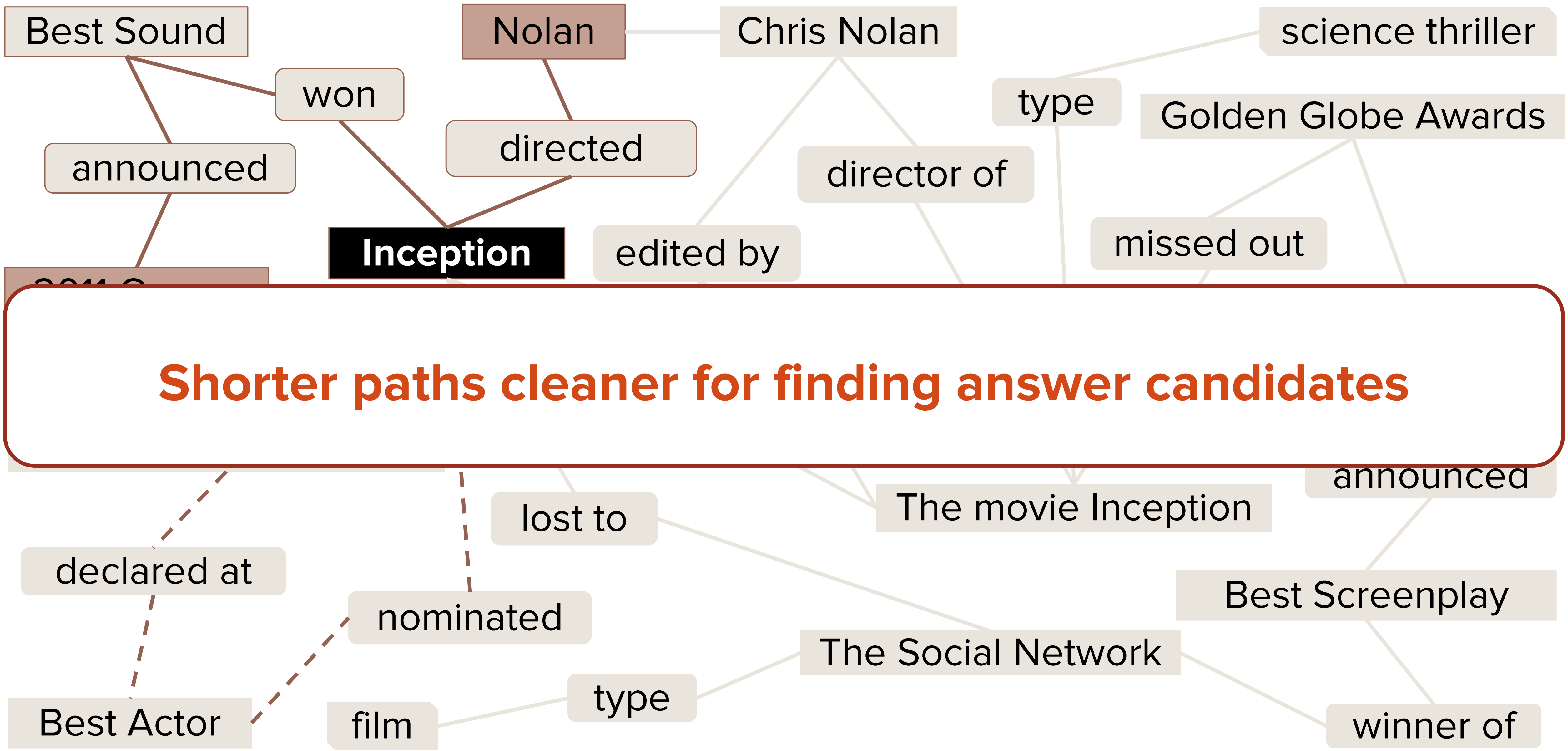
Answers lie on paths connecting cornerstones:

Internal nodes on paths are answer candidates

Question: Which Nolan films won an Oscar but missed a Golden Globe?

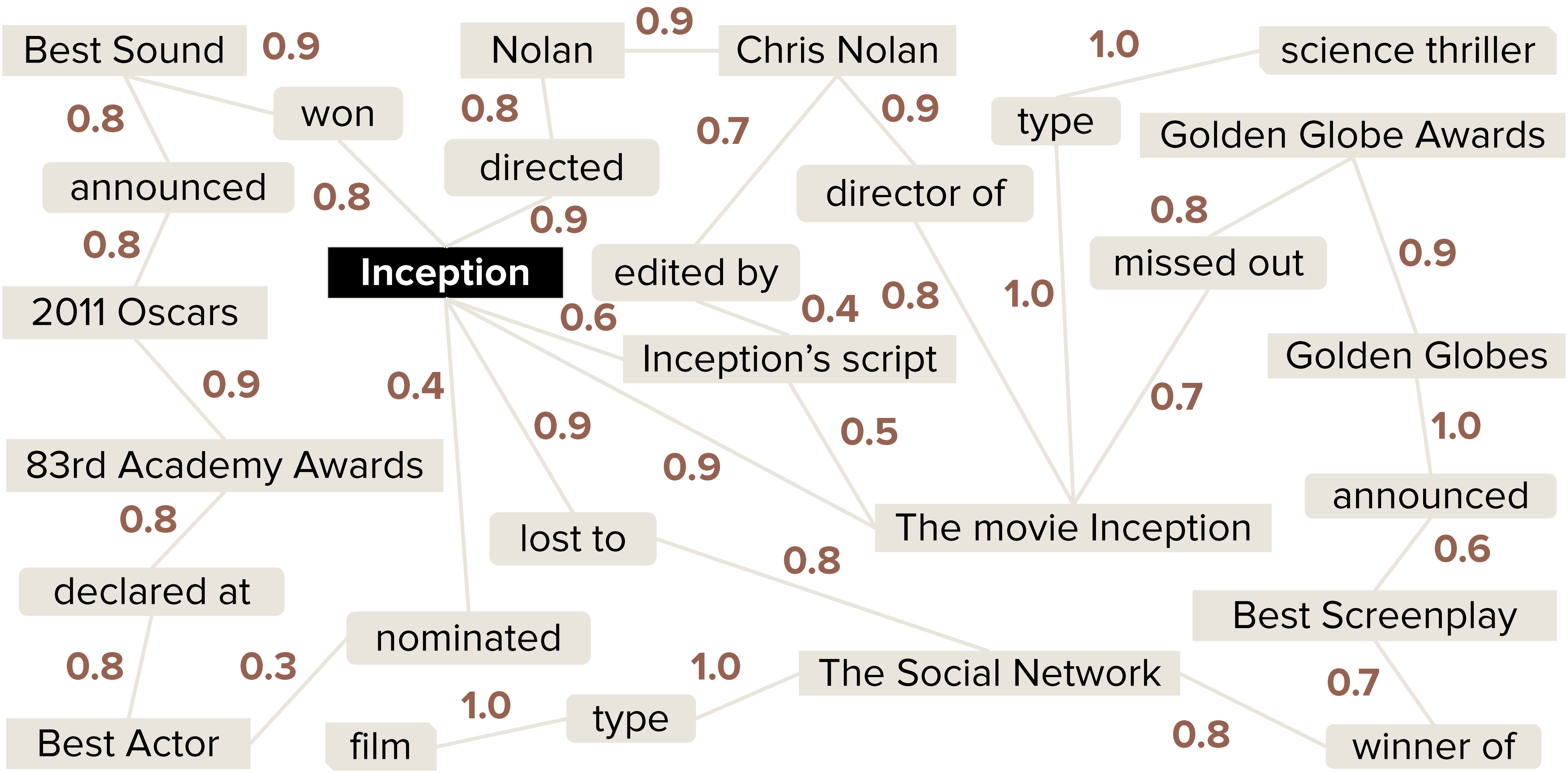


Question: Which Nolan films won an Oscar but missed a Golden Globe?

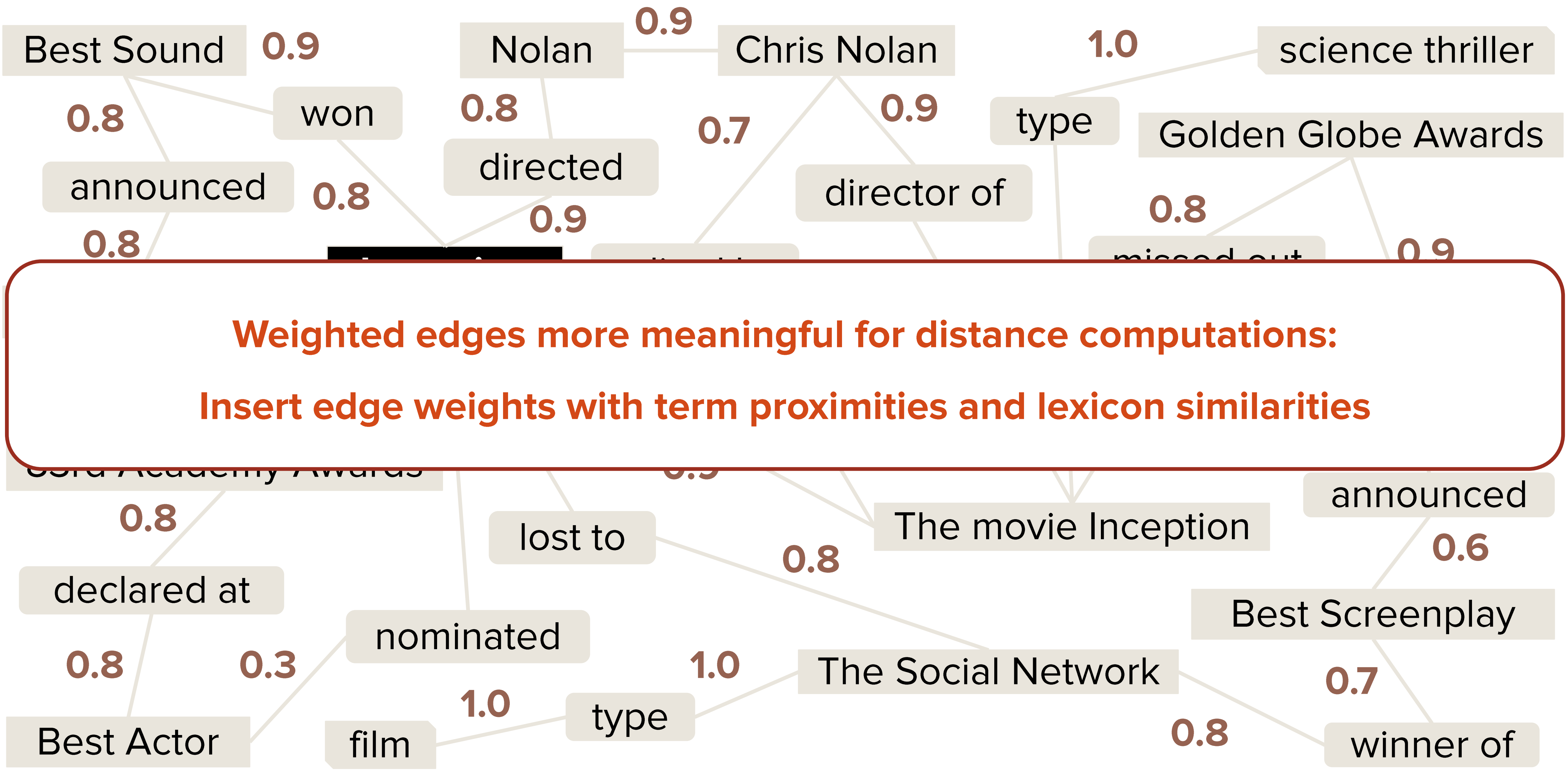


Shorter paths cleaner for finding answer candidates

Question: Which Nolan films won an Oscar but missed a Golden Globe?



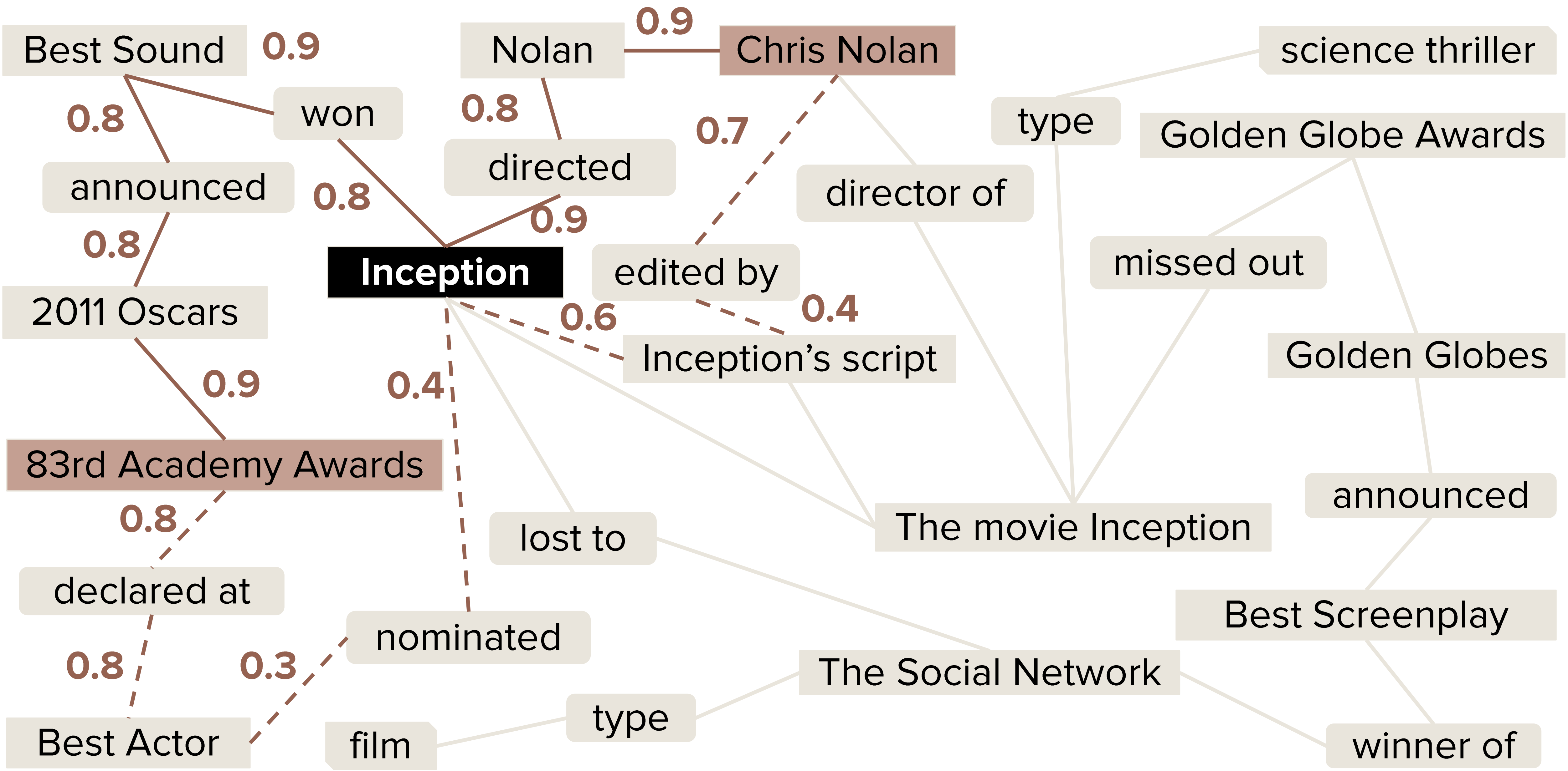
Question: Which Nolan films won an Oscar but missed a Golden Globe?



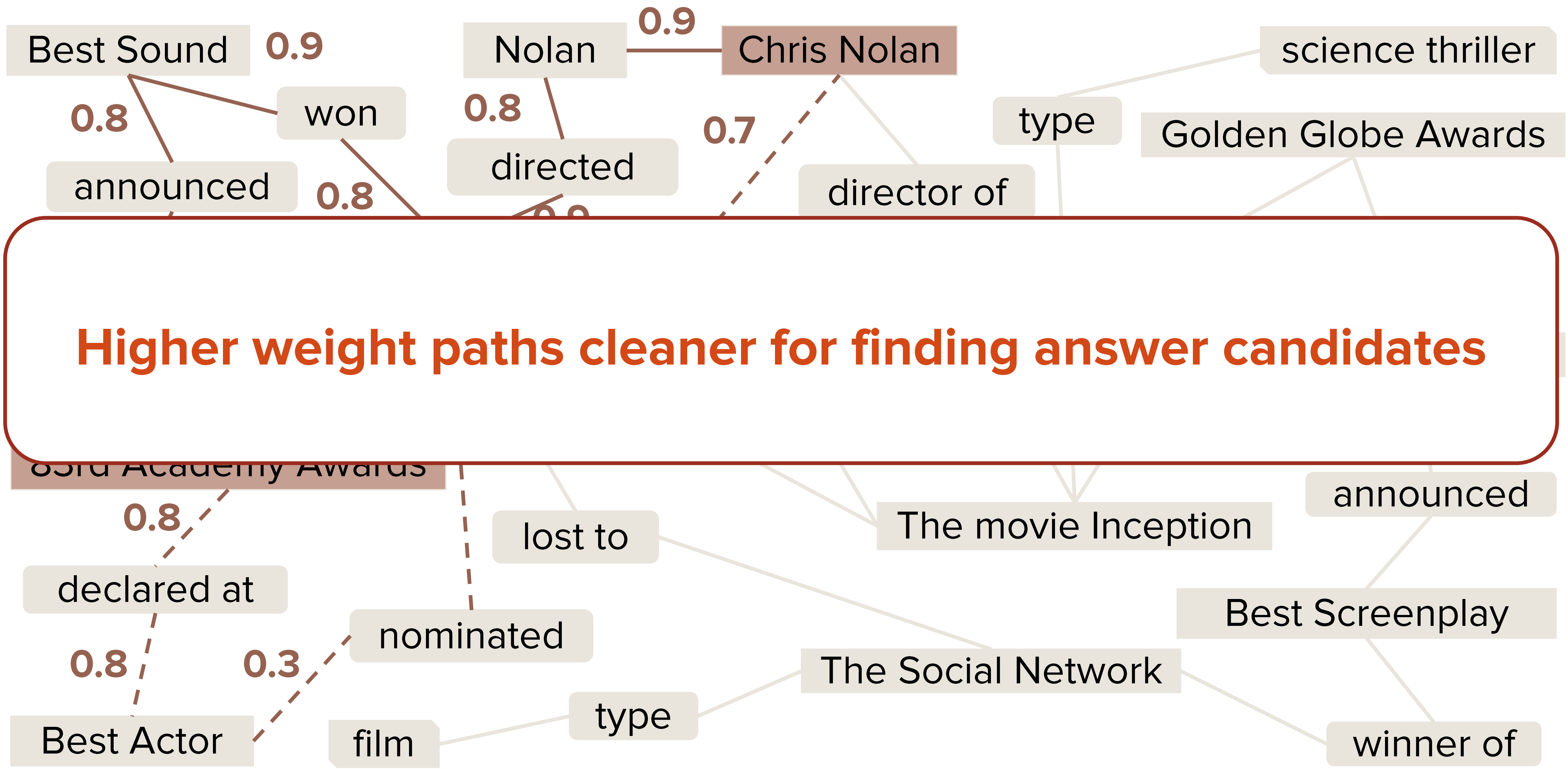
Weighted edges more meaningful for distance computations:

Insert edge weights with term proximities and lexicon similarities

Question: Which Nolan films won an Oscar but missed a Golden Globe?

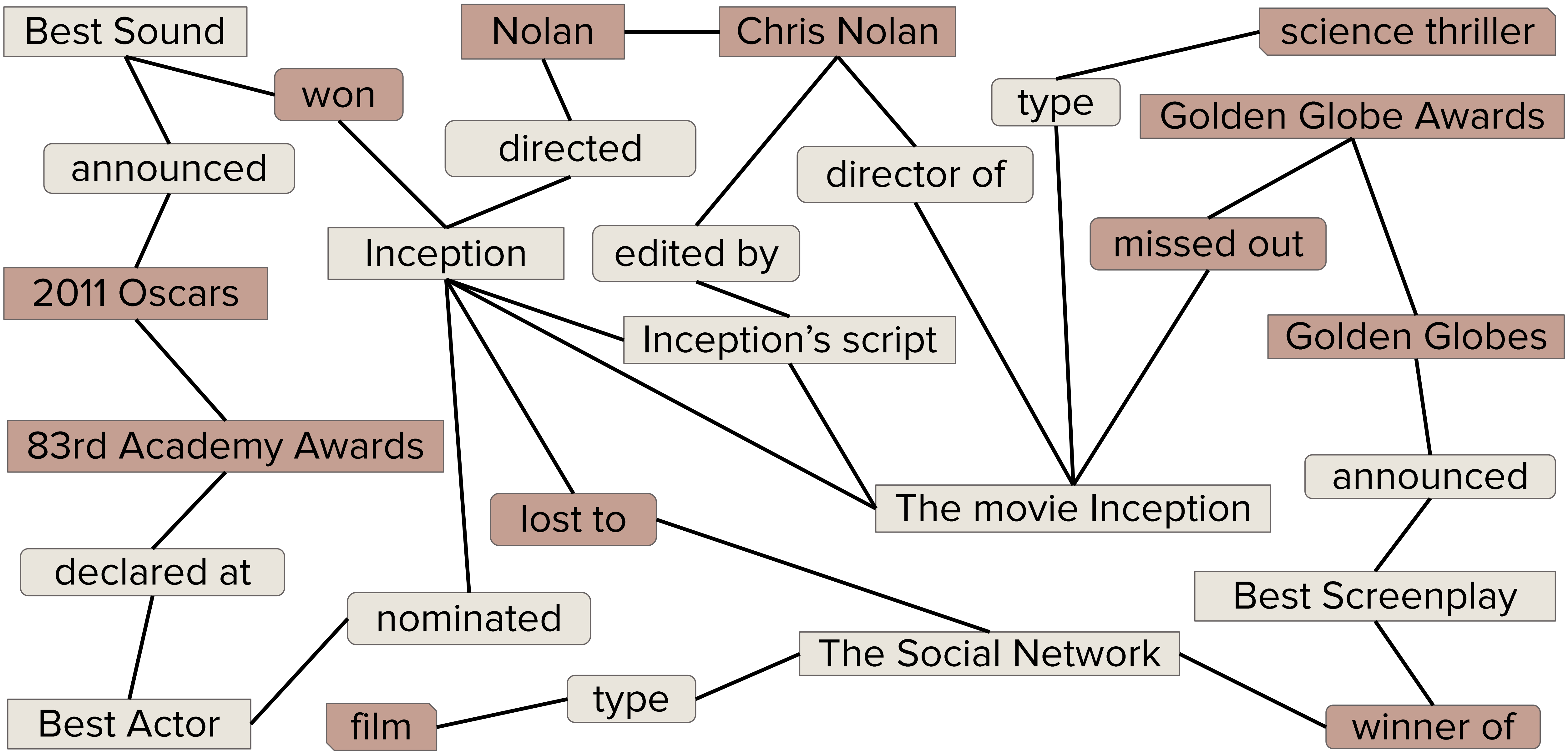


Question: Which Nolan films won an Oscar but missed a Golden Globe?

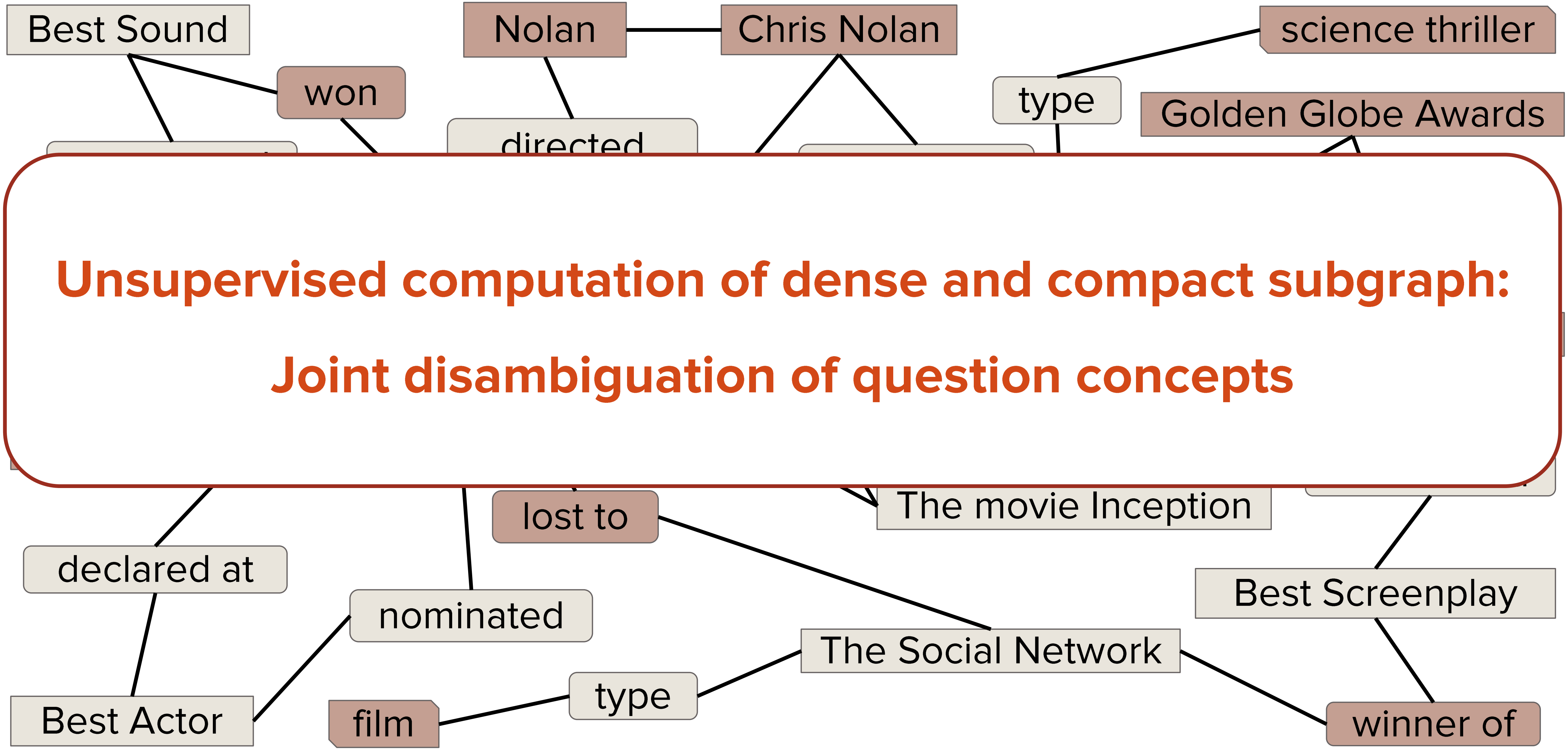


Higher weight paths cleaner for finding answer candidates

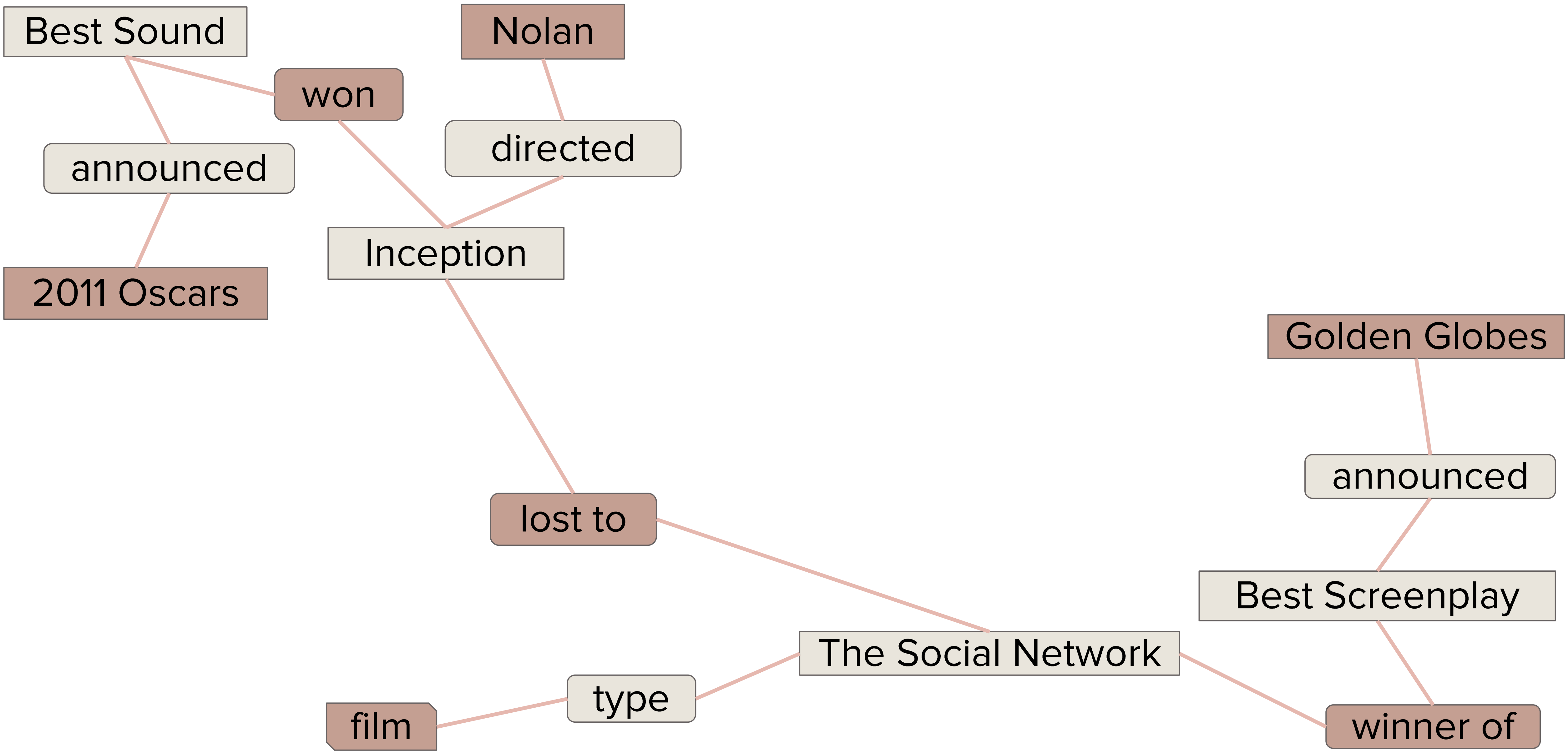
Question: Which Nolan films won an Oscar but missed a Golden Globe?



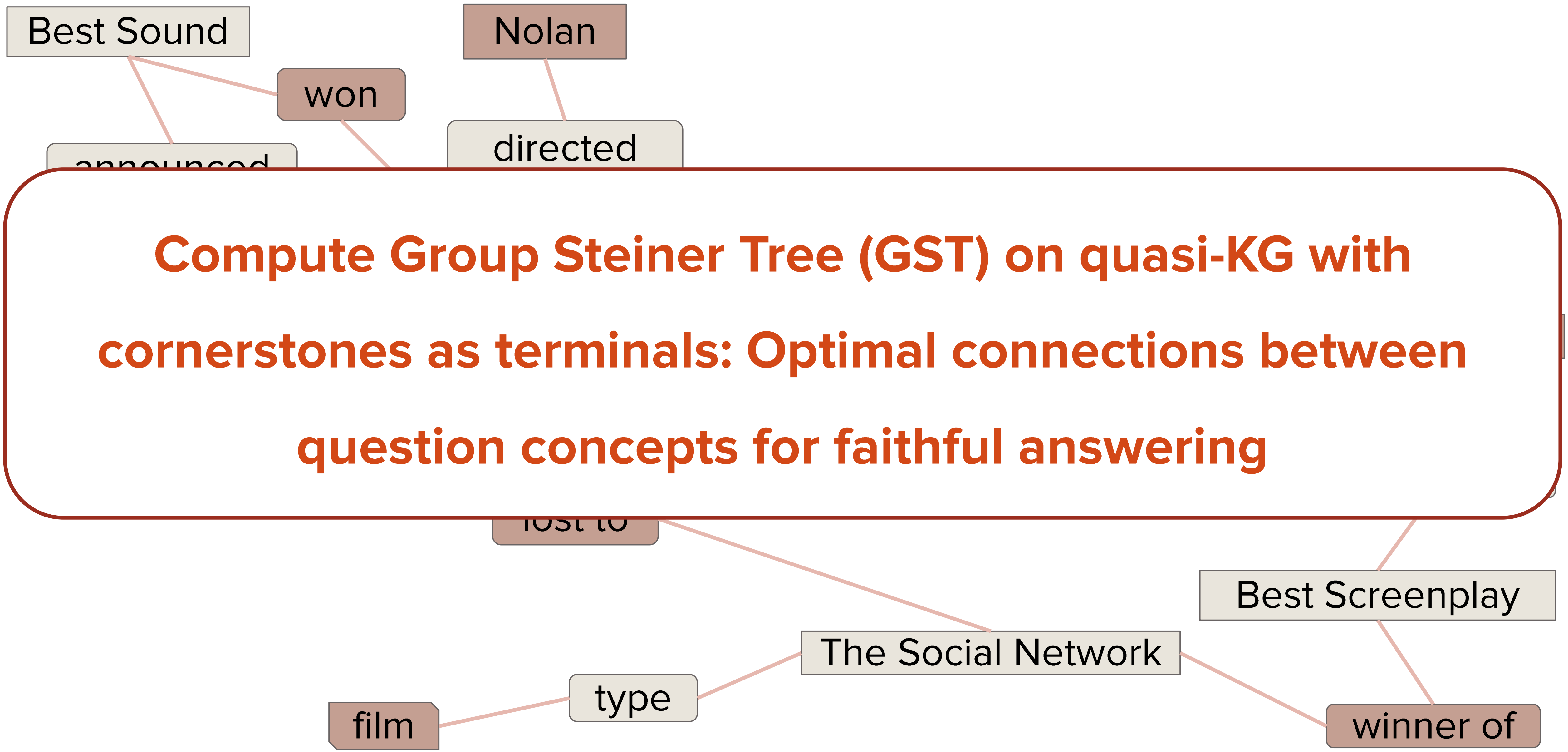
Question: Which Nolan films won an Oscar but missed a Golden Globe?



Question: Which Nolan films won an Oscar but missed a Golden Globe?



Question: Which Nolan films won an Oscar but missed a Golden Globe?



Compute Group Steiner Tree (GST) on quasi-KG with cornerstones as terminals: Optimal connections between question concepts for faithful answering

Question: Which Nolan films won an Oscar but missed a Golden Globe?

- **Answers are nodes**
- Cornerstones are not answers
- Only entities
- Must respect type constraints

Best Sound

Nolan

won

directed

announced

Inception

2011 Oscars

Golden Globes

announced

lost to

Best Screenplay

The Social Network

type

film

winner of

Question: Which Nolan films won an Oscar but missed a Golden Globe?

- Answers are nodes
- **Cornerstones are not answers**
- Only entities
- Must respect type constraints

Best Sound

won

announced

directed

Inception

lost to

type

The Social Network

announced

Best Screenplay

winner of

Question: Which Nolan films won an Oscar but missed a Golden Globe?

Best Sound

Inception

- Answers are nodes
- Cornerstones are not answers
- **Only entities**
- Must respect type constraints

Best Screenplay

The Social Network

Question: Which Nolan films won an Oscar but missed a Golden Globe?

- Answers are nodes
- Cornerstones are not answers
- Only entities
- **Must respect type constraints**

Inception

The Social Network

Question: Which Nolan films won an Oscar but missed a Golden Globe?

- **Answers ranked by aggregation**
- Best answer chosen

Inception

Number of GSTs	**5**

Number of GSTs	2

The Social Network

Question: Which Nolan films won an Oscar but missed a Golden Globe?

- Answers ranked by multiple criteria
- **Best answer chosen**

Inception

Lu et al., Answering Complex Questions by Joining Multi-Document Evidence with Quasi Knowledge Graphs, SIGIR 2019.

Complex QA: Graph-based belief propagation

- The **QAmp** system (Vakulenko et al. 2019)
- Interpretation
  - Parsing
  - Matching
- Reasoning
  - **Message passing**
  - Score aggregation

API access possible by appending the text of a question to *https://kbqa-api.ai.wu.ac.at/ask?question=*

For example, for "Name the municipality of Roberto Clemente Bridge?", use:

https://kbqa-api.ai.wu.ac.at/ask?question=Name%20the%20municipality%20of%20Roberto%20Clemente%20Bridge%20?

Vakulenko et al., Message Passing for Complex Question Answering over Knowledge Graphs, CIKM 2019.

Interpretation: Parsing by sequence labeling

Where is the **founder** of **Tesla** **born**?

P1 **E1** **P2**

Use CRFs trained on labeled data

Predict question type: **SELECT**, COUNT, ASK

Interpretation: Matching

Where is the **founder** of **Tesla** **born**?

P1 E1 P2

Use CRFs trained on labeled data

Predict question type: **SELECT**, COUNT, ASK

P1	
founder	1
founded	0.8

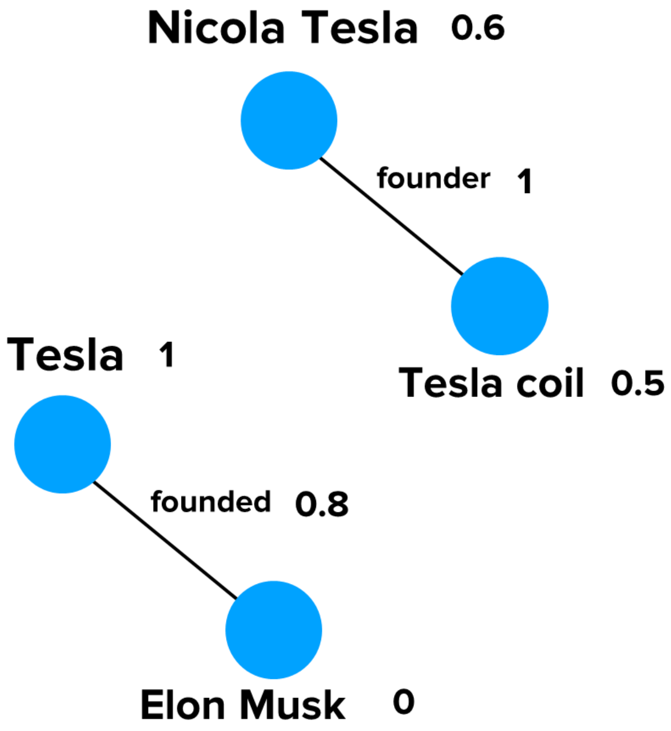
P2	
bornIn	0.8

E1	
Tesla	1
Nicola Tesla	0.6
Tesla coil	0.5

Reasoning: Message passing Hop 1

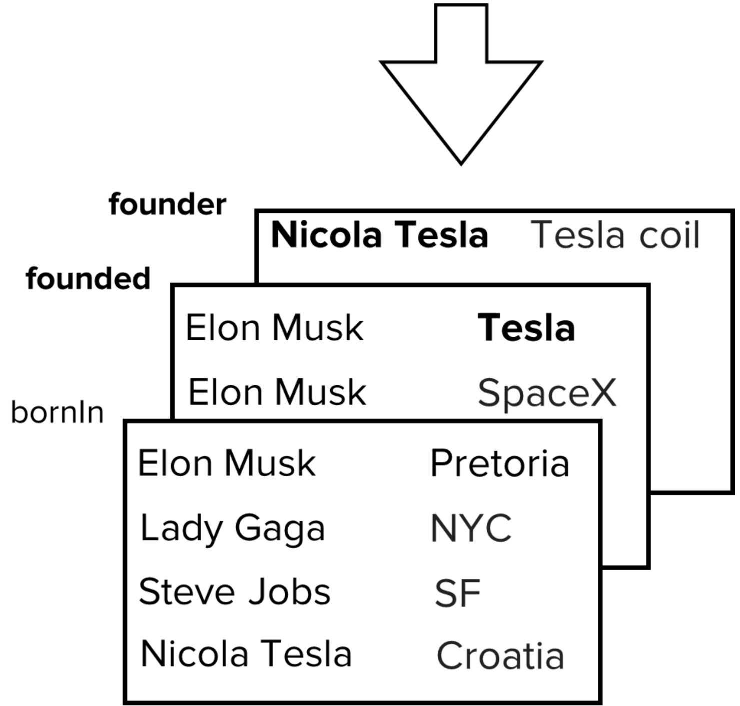
Where is the **founder** of **Tesla** **born**?

P1 — founder; E1 — Tesla; P2 — born



P1	
founder	1
founded	0.8

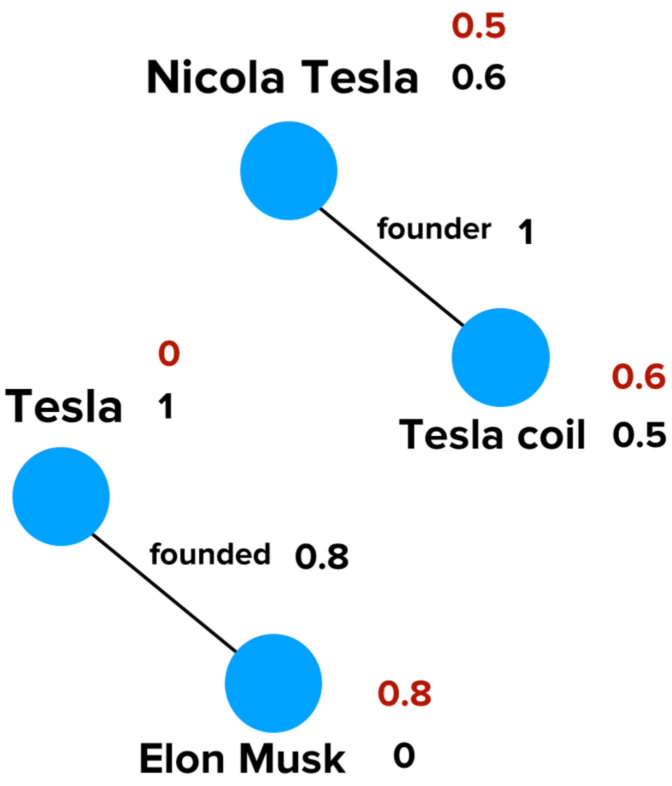
E1	
Tesla	1
Nicola Tesla	0.6
Tesla coil	0.5



Reasoning: Message passing Hop 1

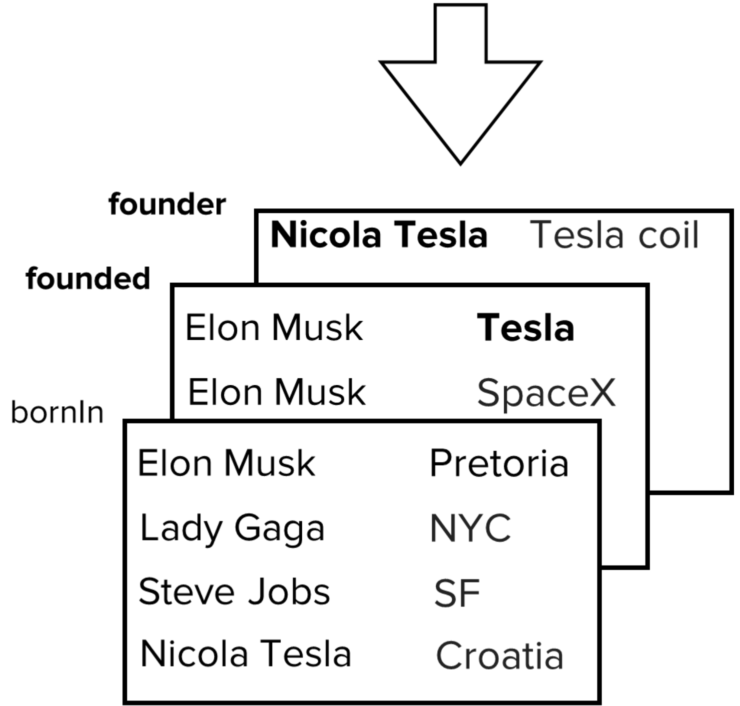
Where is the **founder** of **Tesla** **born**?

P1 E1 P2

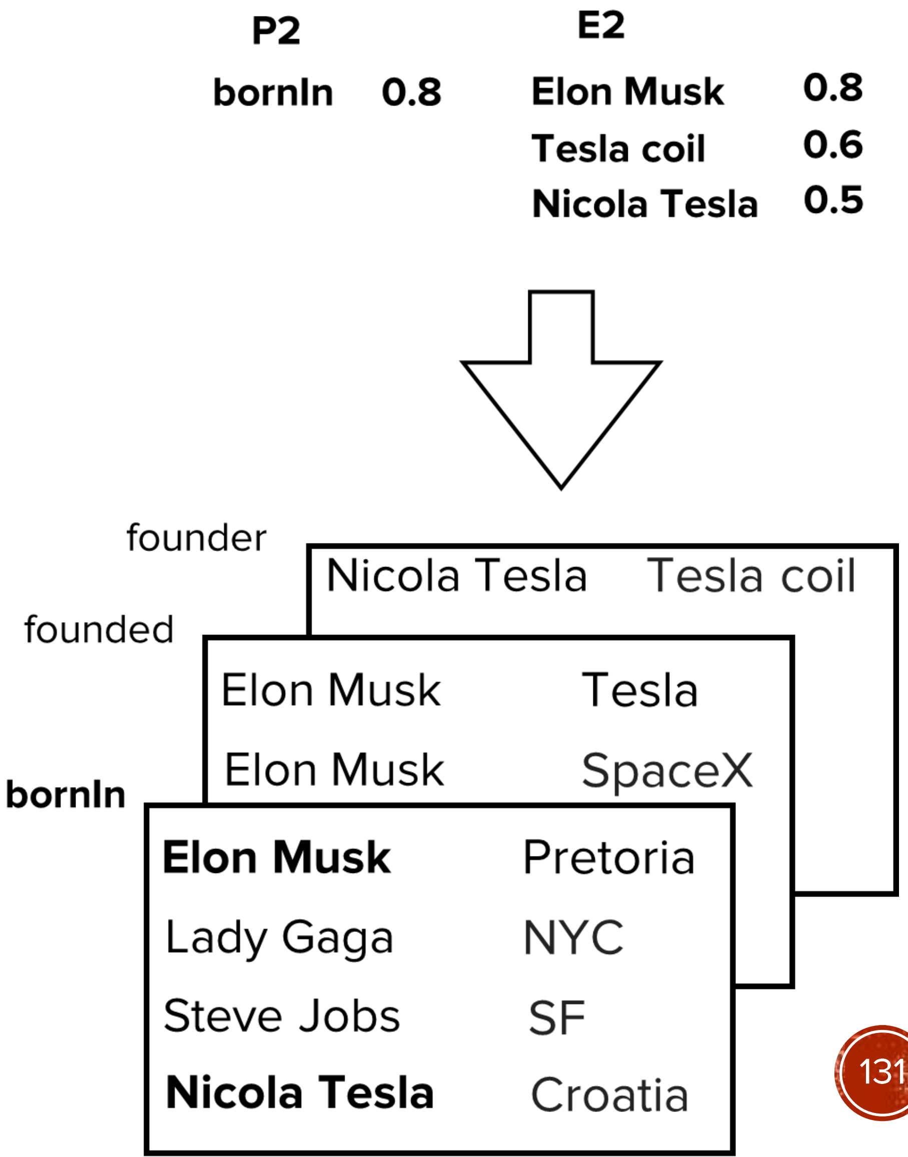


P1	
founder	1
founded	0.8

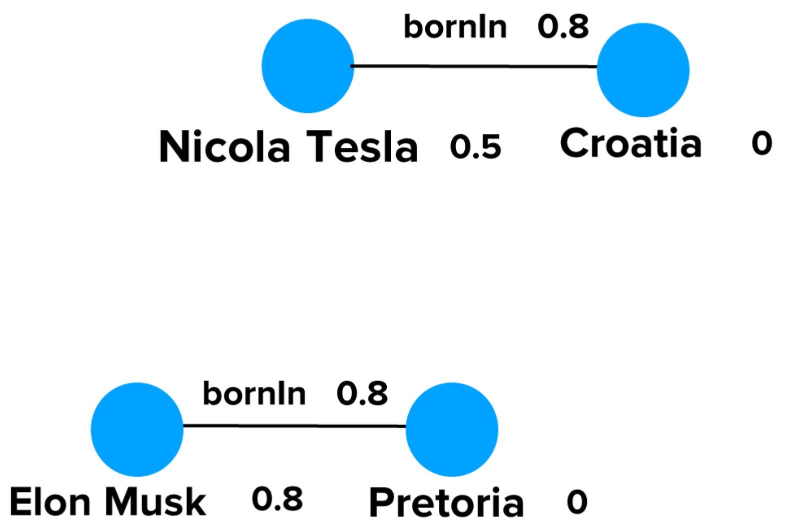
E1	
Tesla	1
Nicola Tesla	0.6
Tesla coil	0.5



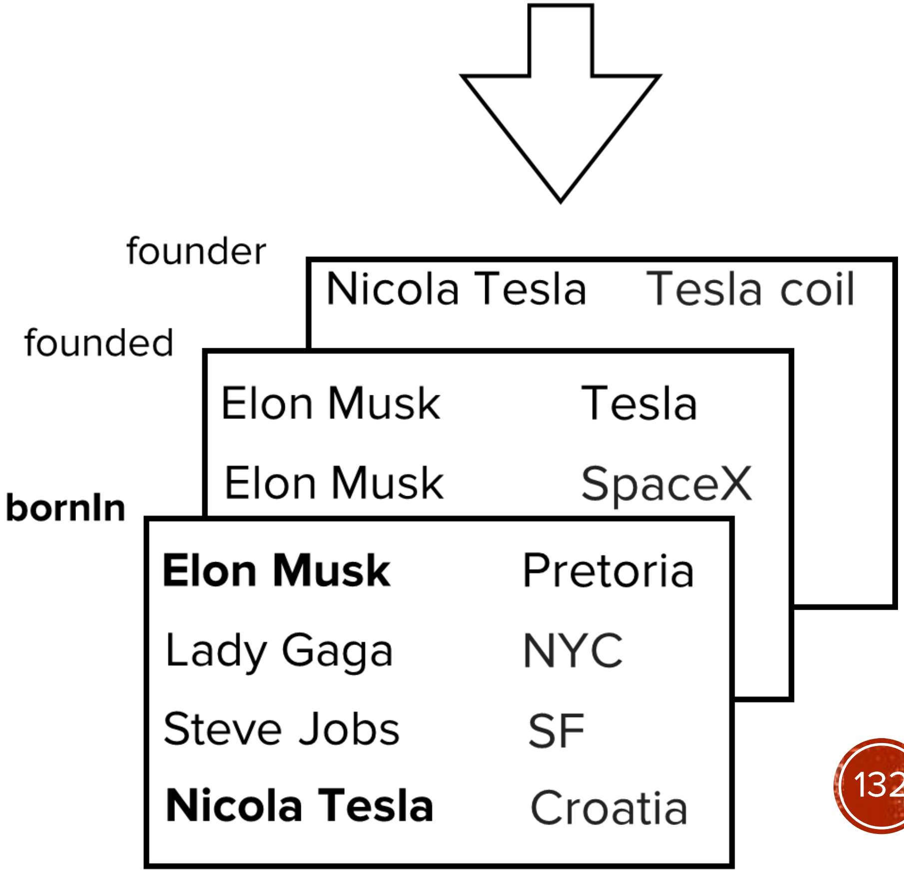
Reasoning: Message passing Hop 2



Reasoning: Message passing Hop 2



P2		E2	
bornIn	**0.8**	**Elon Musk**	**0.8**
		Tesla coil	**0.6**
		Nicola Tesla	**0.5**

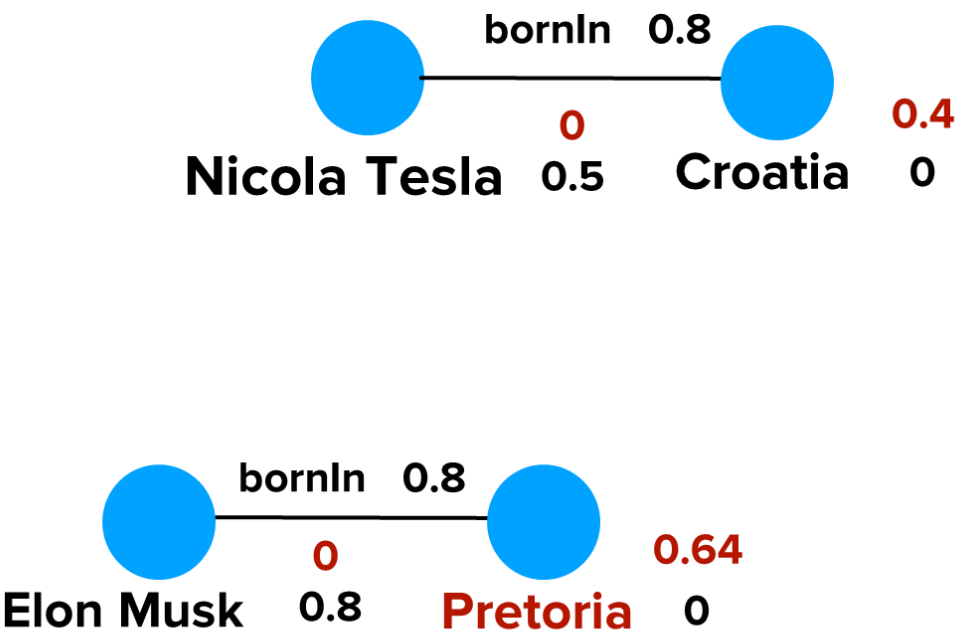


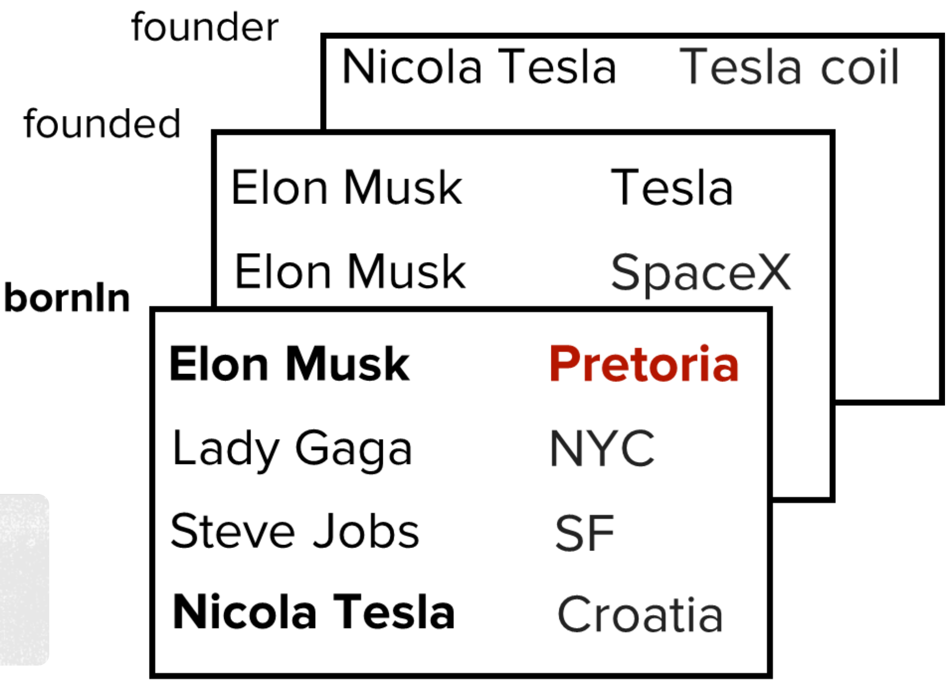
Reasoning: Message passing Hop 2

Where is the **founder** of **Tesla** **born**?

P2	
bornIn	**0.8**

E2	
Elon Musk	**0.8**
Tesla coil	**0.6**
Nicola Tesla	**0.5**





Constructs explanatory evidence

Efficient (2 seconds/question) for use of HDT + matrix operations

Complex questions: Wrap-up

- Complex KG-QA the sub-topic with the **highest attention**
- **Efficiency** generally an open issue: several partial queries executed in TextRay, a lot of similarity computations in QUEST, …
- Bias in SoTA towards **certain classes**: QAmp (**chains**), QUEST (**stars**), …
- How to reduce **large neighborhood sizes**? KGs are dense: considering full 2-hop neighborhoods often intractable due to popular entities or general entity types

break duration ?x .
?x measuredIn minutes .

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

How can we answer questions over heterogeneous sources?

QA over heterogeneous sources

- **Heterogeneous source:** System should tap into multiple KGs, or KG + Text
- **Why fuse?** Each source has its advantages and disadvantages
- **Early fusion:** GRAFT-Net (Sun et al. 2018), **PullNet (Sun et al. 2019)**
- **Late fusion:** Ferrucci et al. (2010), Baudis (2015), Sun et al. (2015), Xu et al. (2016a, 2016b), Savenkov and Agichtein (2016)
- **Unified representations:** OQA (Fader et al. 2014), TriniT (Yahya et al. 2016), UniSchema (Das et al. 2017), Nestique (Bhutani et al. 2019b), **QAnswer (Diefenbach et al. 2019)**

Heterogeneous QA: Early fusion

- The **PullNet** system (Sun et al. 2019)
- Fusion via KG **facts** and KG-**entity linked sentences**
- Built for **multi-hop** questions
- Uses **question-focused subgraph**
- **Judiciously expands context** subgraph
- Uses classifiers for **expansion points** and answers

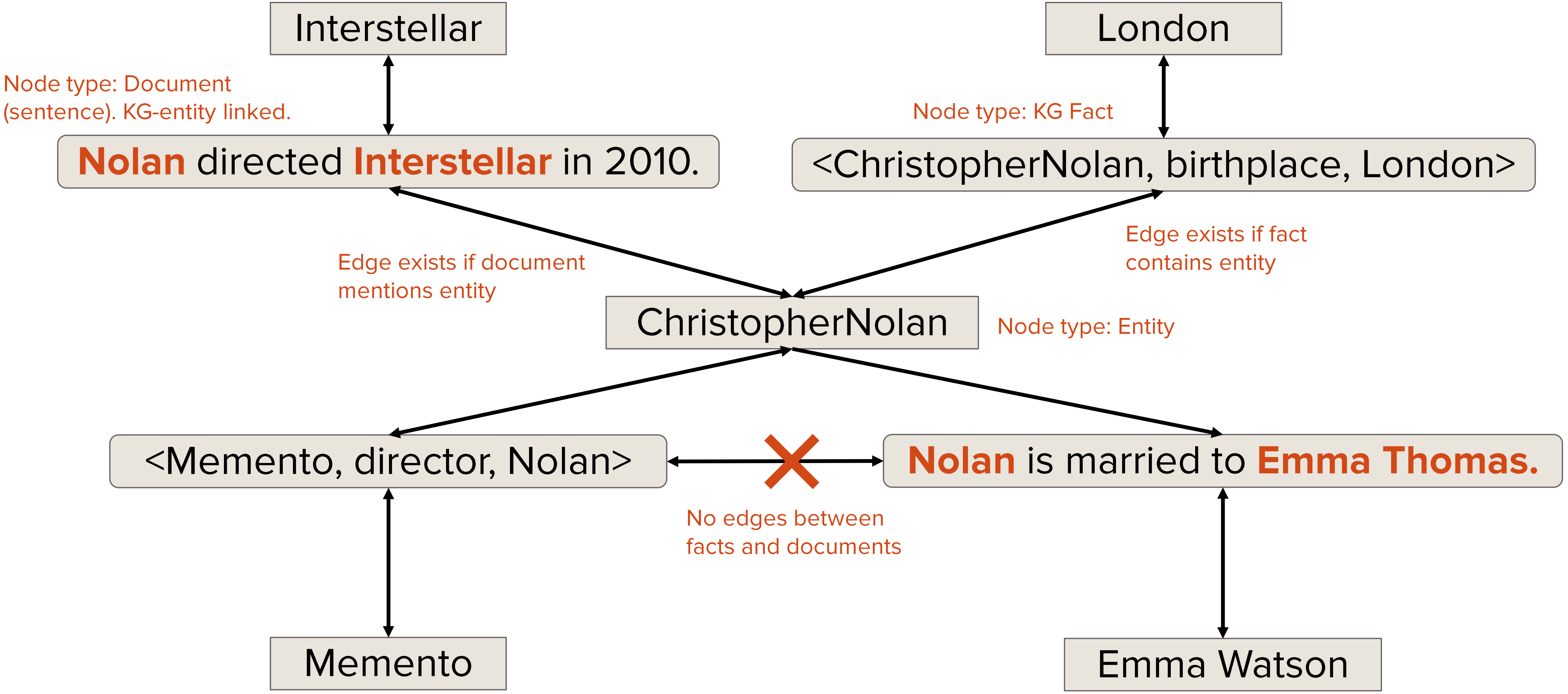
Sun et al., PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text, EMNLP 2019.

PullNet: Handling heterogeneity

KG facts	Entity-linked sentences
<ChristopherNolan, birthplace, London>	**Nolan** is married to **Emma Thomas.**
<Memento, director, Nolan>	**Nolan** directed **Interstellar** in 2010.
<Interstellar, castMember, AnneHathaway>	**Guy Pearce** was in **Memento** and **Flynn.**

Sun et al., PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text, EMNLP 2019.

PullNet: Graph model



Question: Who are the actors in movies directed by Nolan?

Sun et al., PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text, EMNLP 2019.

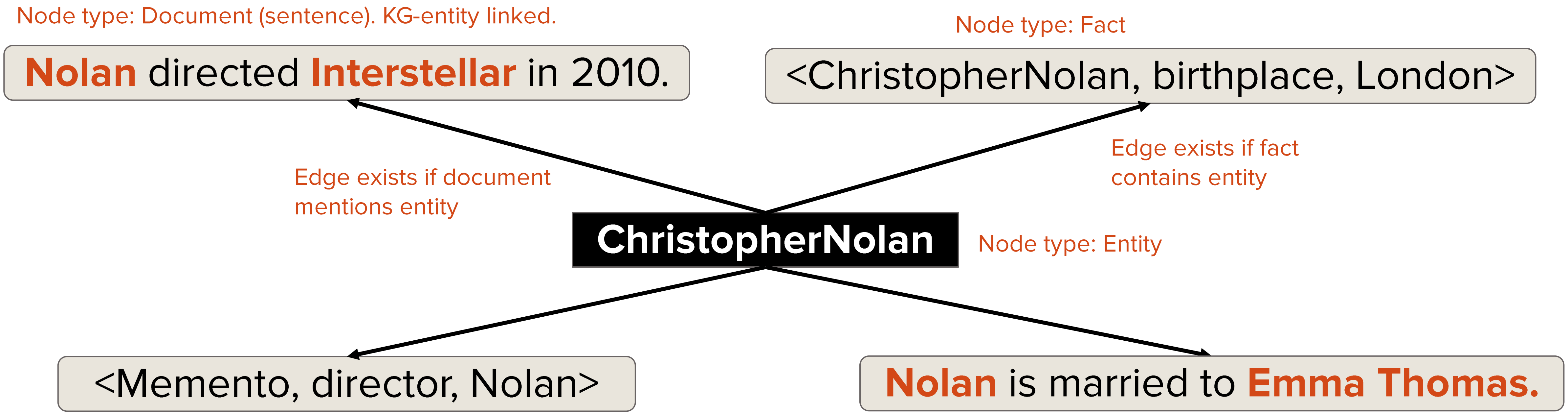
Question: Who are the actors in movies directed by Nolan?

NERD system

ChristopherNolan

Sun et al., PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text, EMNLP 2019.

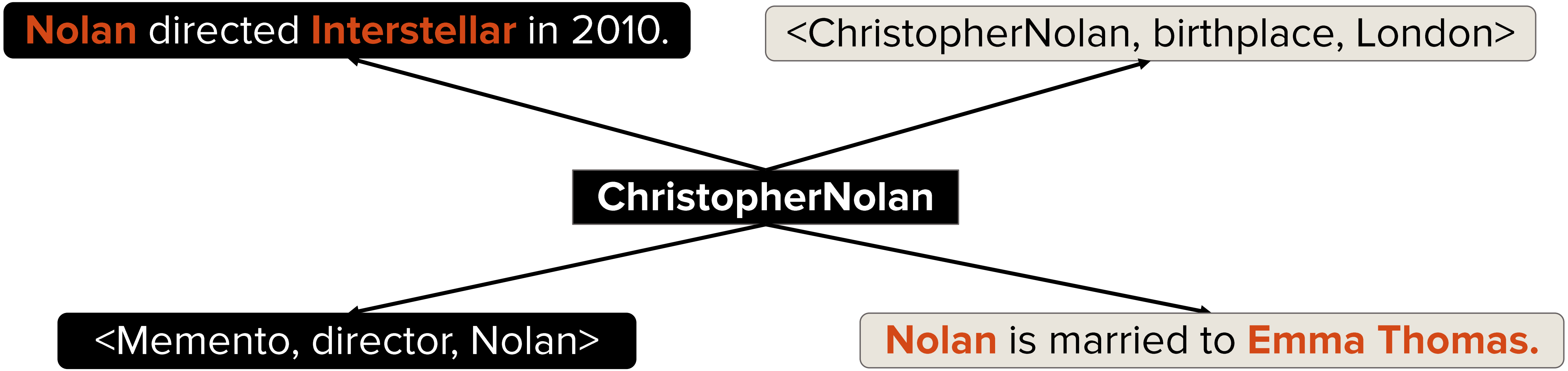
Question: Who are the actors in movies directed by Nolan?



Early fusion

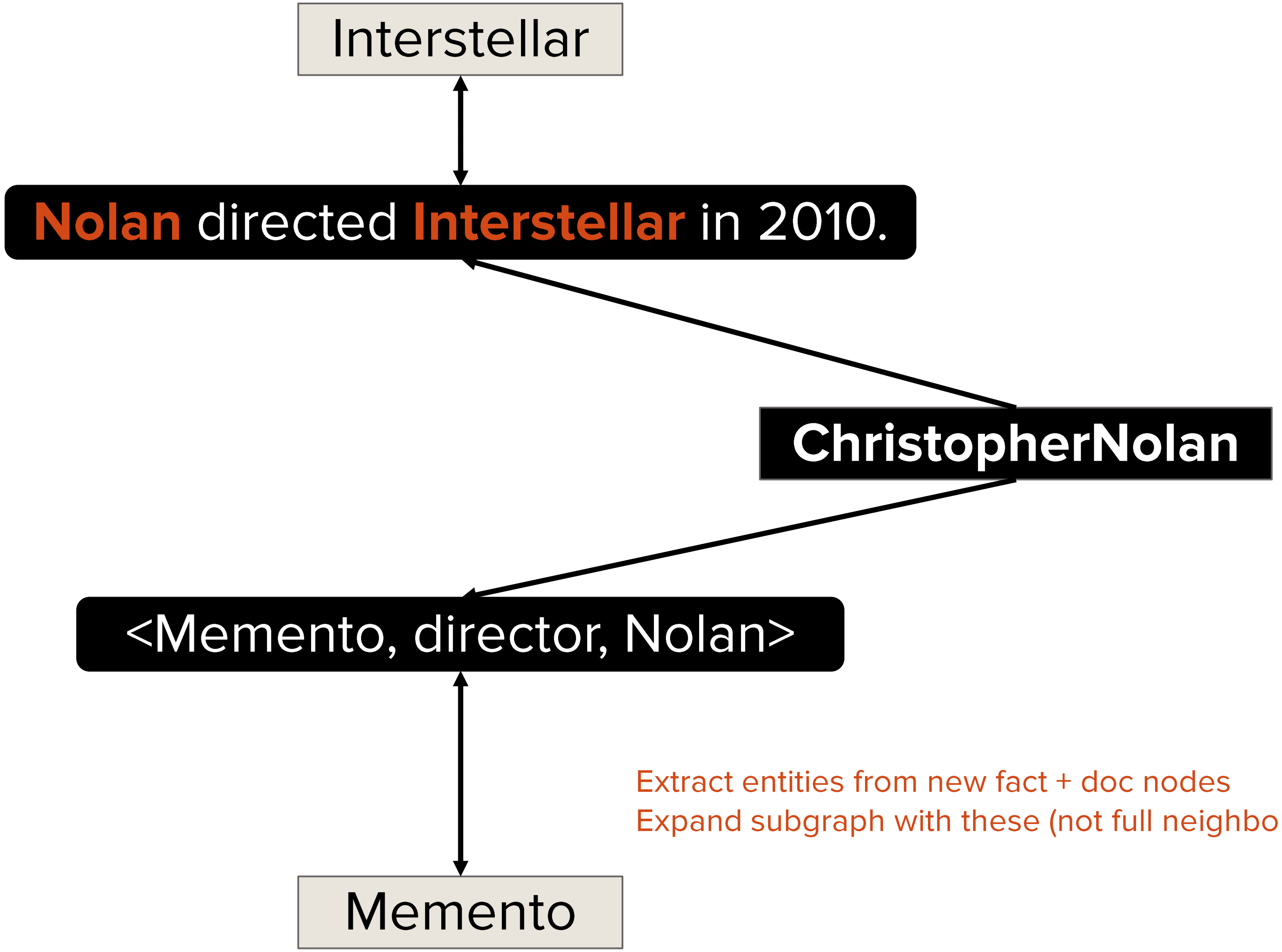
1. Pull sentences with linked entity from corpus (Lucene)
2. Pull facts of entity from the KG (using predicate similarity learned via LSTMs)

Question: Who are the actors in movies directed by Nolan?



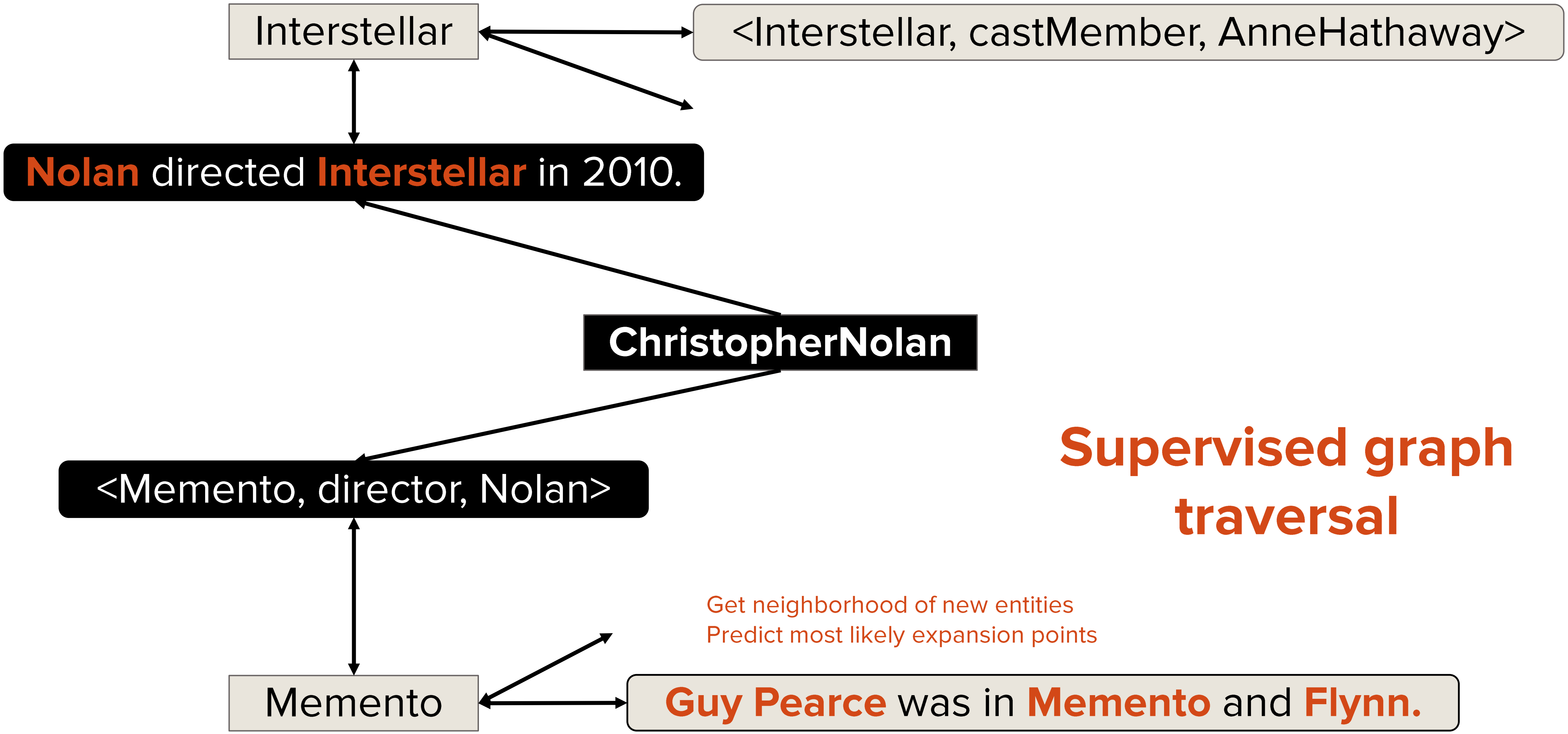
Predict most likely expansion points for next hop

Question: Who are the actors in movies directed by Nolan?



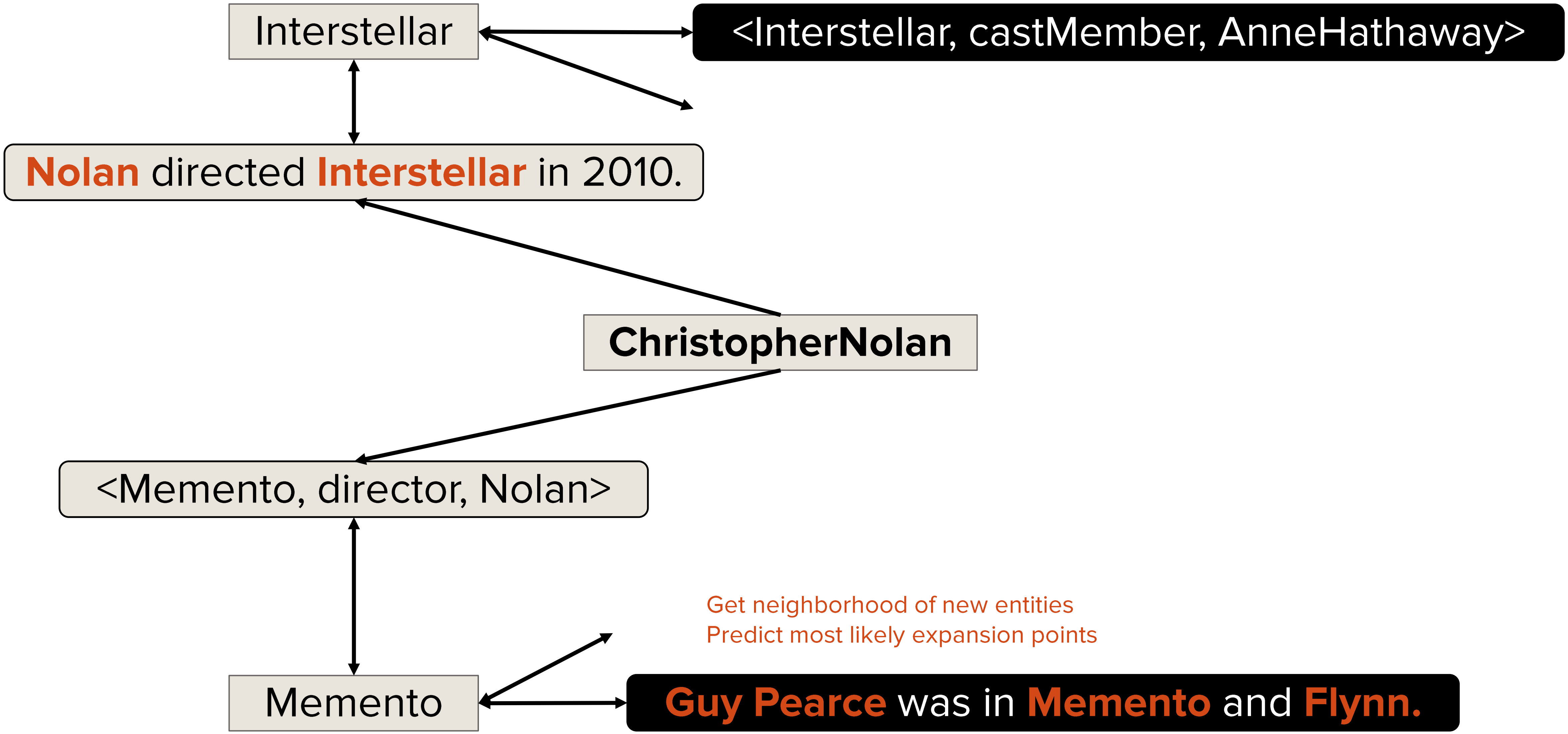
Extract entities from new fact + doc nodes
Expand subgraph with these (not full neighborhood)

Question: Who are the actors in movies directed by Nolan?

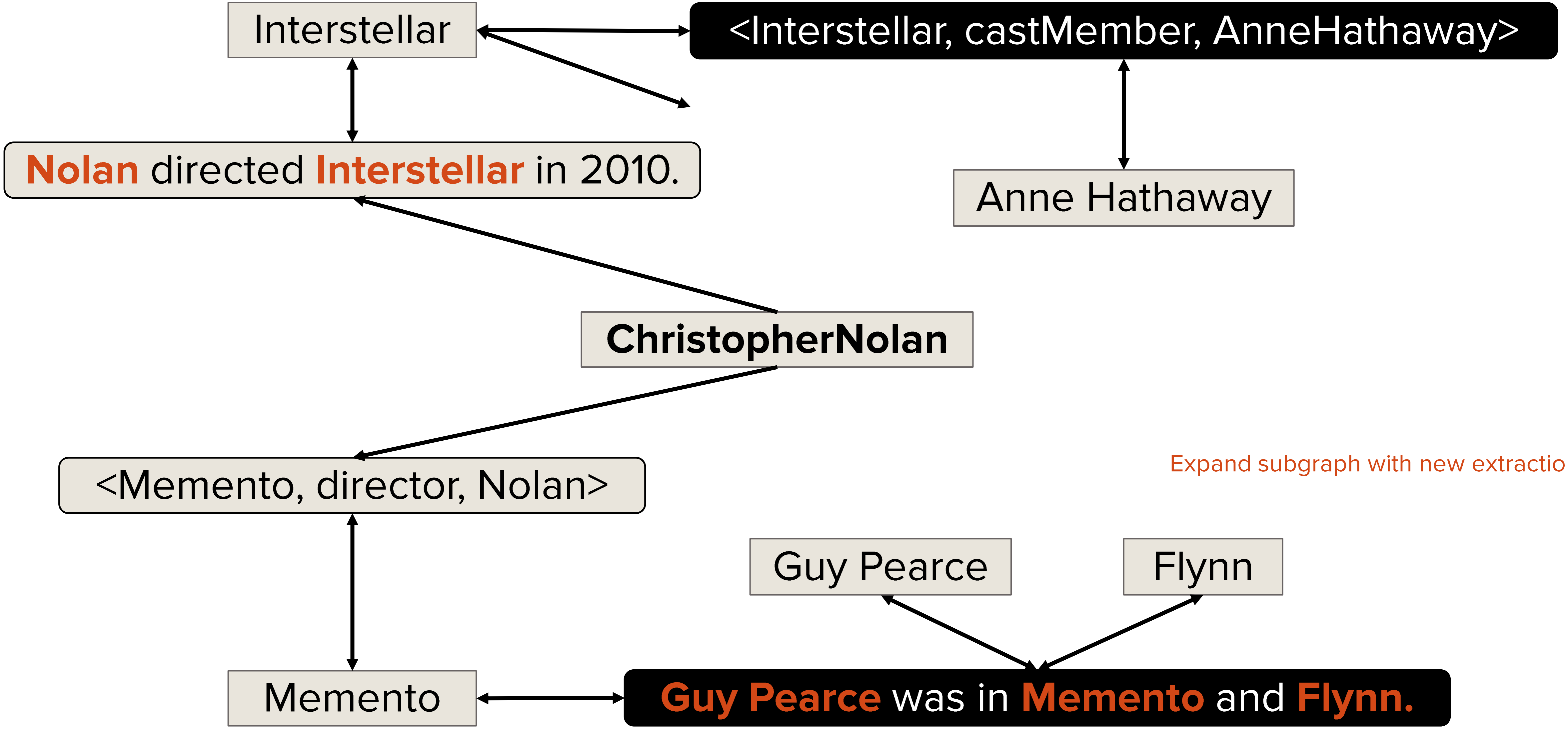


Supervised graph traversal

Question: Who are the actors in movies directed by Nolan?

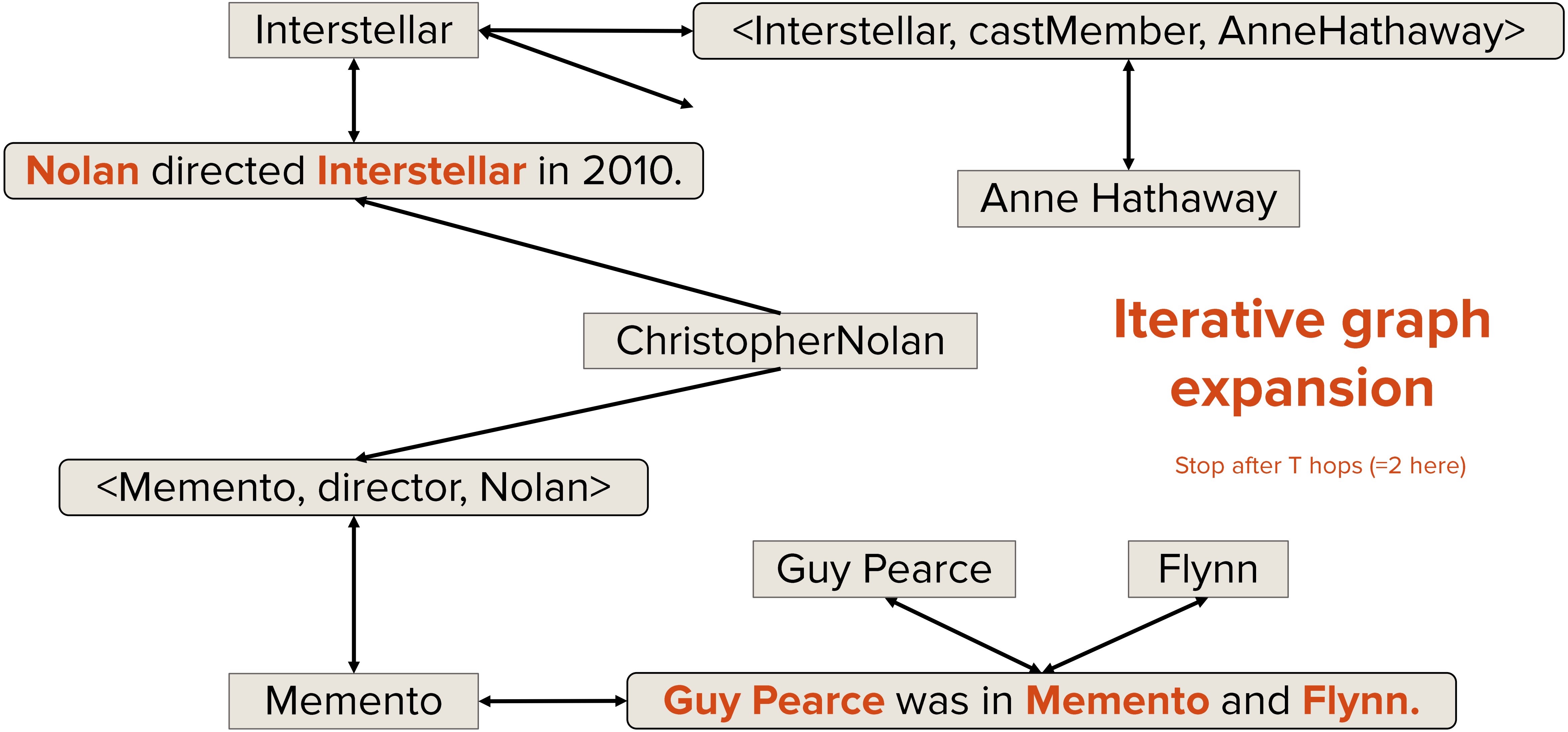


Question: Who are the actors in movies directed by Nolan?



Expand subgraph with new extractions

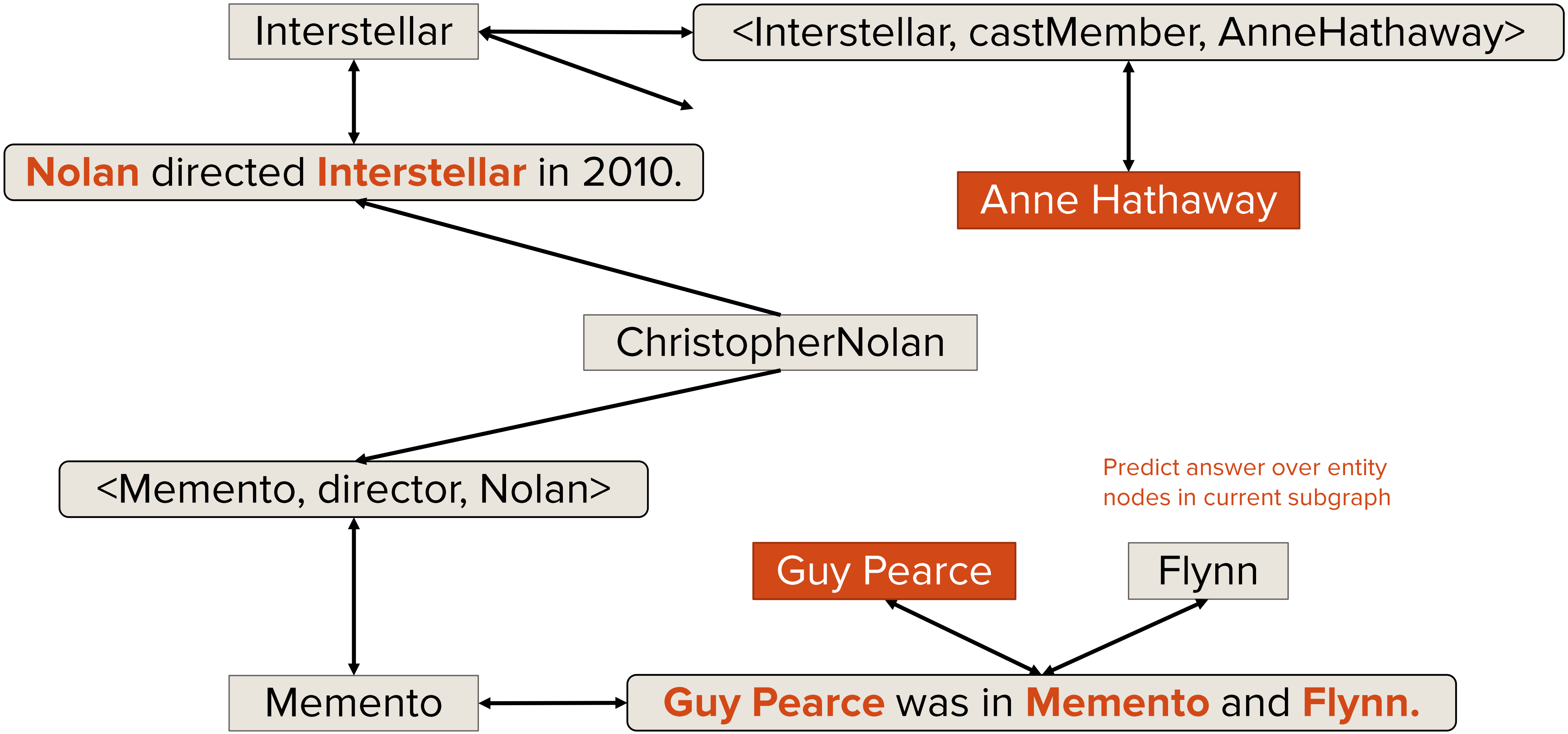
Question: Who are the actors in movies directed by Nolan?



Iterative graph expansion

Stop after T hops (=2 here)

Question: Who are the actors in movies directed by Nolan?



Question: Who are the actors in movies directed by Nolan?

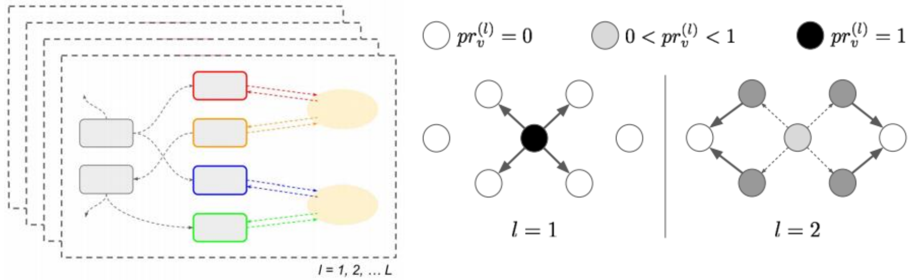
Interstellar

<Interstellar, castMember, AnneHathaway>

Nolan directed **Interstellar** in 2010.

Anne Hathaway

ChristopherNolan



Graph CNN-based classifier (GRAFT-Net) for both expansion point and answer prediction (Sun et al. 2018)

<Memento, director, Nolan>

Guy Pearce

Flynn

Memento

Guy Pearce was in **Memento** and **Flynn.**

PullNet: Training

- Distant supervision with QA pairs
- Uses shortest paths between Q and A entities **in KG**
- Gold expansion points: Intermediate nodes on shortest paths
- Uses teacher forcing
- Gold answers: From benchmark

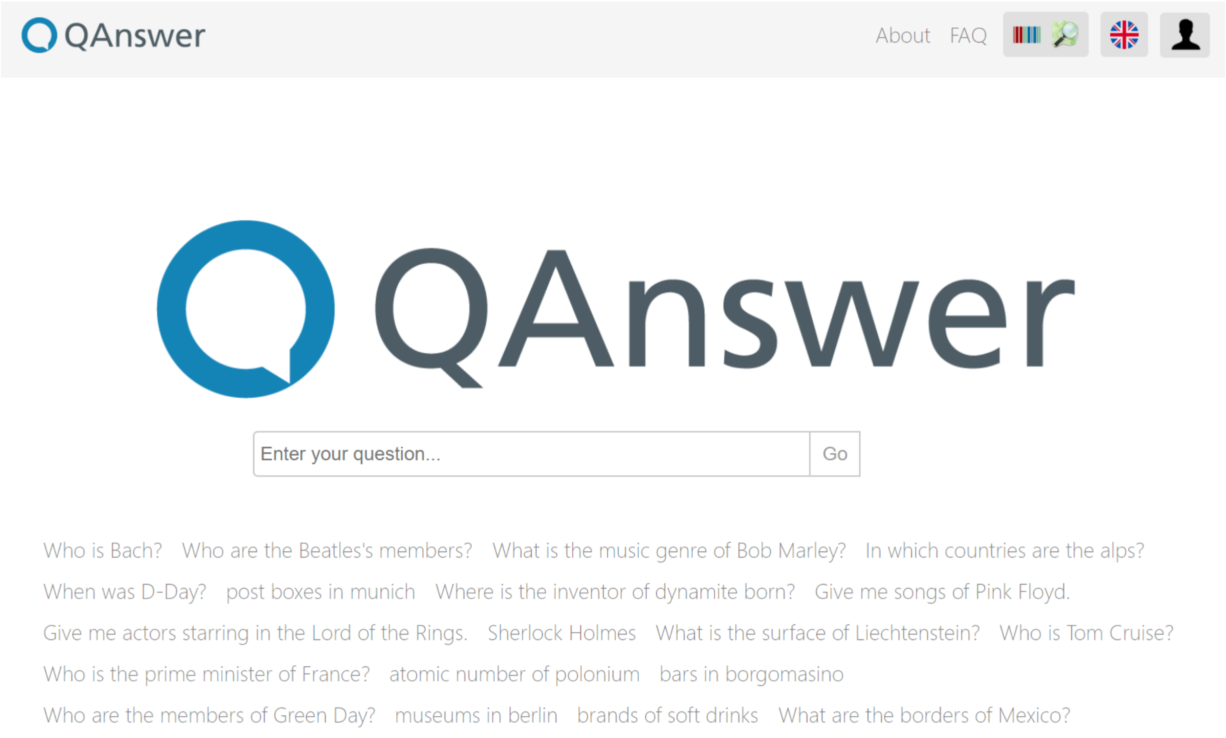
Sun et al., PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text, EMNLP 2019.

Closely related to multi-hop KGR

- Multi-hop knowledge graph reasoning (KGR) and knowledge graph completion (KGC) closely associated with multi-hop QA
- Bridge between neural and symbolic space
- MINERVA (Das et al. 2018) [Reinforcement learning]
- SRN (Qiu et al. 2020) [Reinforcement learning]
- DrKIT (Dhingra et al. 2020)
- Similar ideas explored for multi-hop MRC (Asai et al. 2020)

Heterogeneous QA: Unified resource

- **QAnswer** (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA



Diefenbach et al., QAnswer: A Question answering prototype bridging the gap between a considerable part of the LOD cloud and end-users, WWW 2019.

Structured querying over multiple KGs

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA

Give me actors born in Berlin.

Question

Diefenbach et al., QAnswer: A Question answering prototype bridging the gap between a considerable part of the LOD cloud and end-users, WWW 2019.

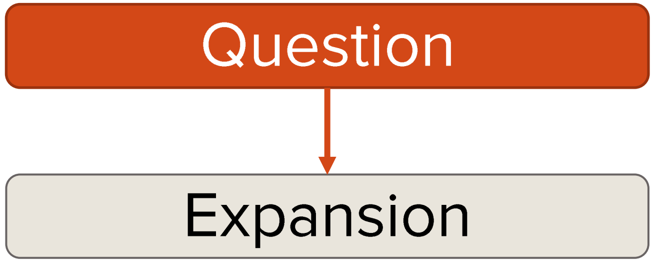
Expand question with candidate KG concepts

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA

Give me actors born in Berlin.

R = {actor, TV actor, bornIn, Born, Berlin, Berlin Univ, West Berlin}

Lucene-based lookup



Diefenbach et al., QAnswer: A Question answering prototype bridging the gap between a considerable part of the LOD cloud and end-users, WWW 2019.

Generation of SPARQL queries with candidates

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA

Efficient construction of SPARQL queries using a BFS of depth 2 on the KG **(exhaustive but valid)**

Enabled by HDT + additional indexing of KG (distances between object pairs)

Give me actors born in Berlin.

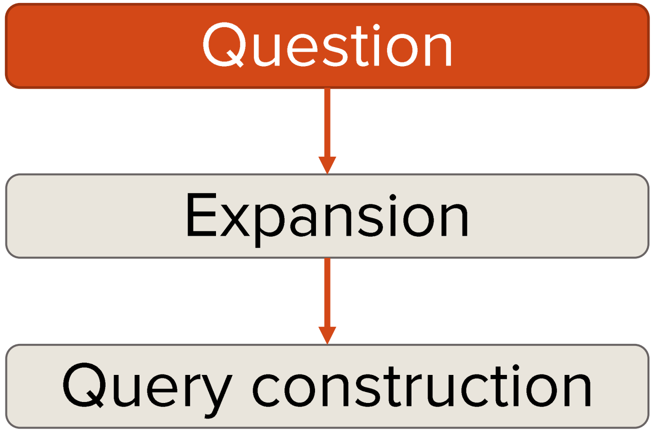
R = {actor, TVActor, bornIn, Born, Berlin, BerlinUniv, WestBerlin}

```
SELECT / ASK ?x
WHERE {s1 s2 s3}

SELECT / ASK ?x
WHERE {s1 s2 s3 . s4 s5 s6 .}

SELECT ?x
WHERE {   ?x    bornIn Berlin .
          ?x    ?y     actor}

SELECT ?x
WHERE {   ?x    ?y     BerlinUniv .
          ?x    ?y     TVActor .}
```



Structured querying over multiple KGs

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA

Give me actors born in Berlin.

Question

↓

R = {actor, TV actor, bornIn, Born, Berlin, Berlin Univ, West Berlin}

Expansion

↓

SELECT / ASK ?x
WHERE {s1 s2 s3 . s4 s5 s6 .}

Query construction

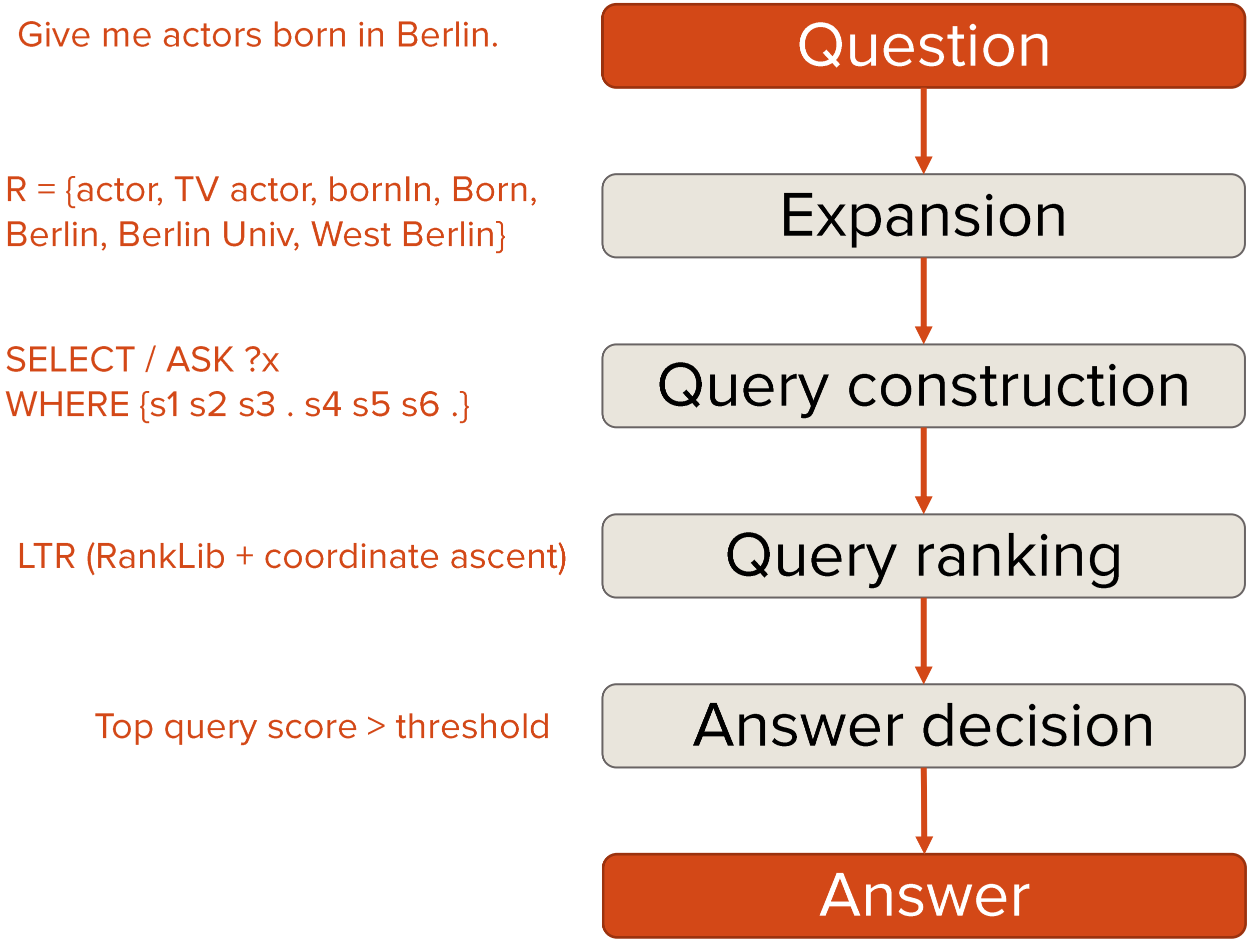
↓

LTR (RankLib + coordinate ascent)

Query ranking

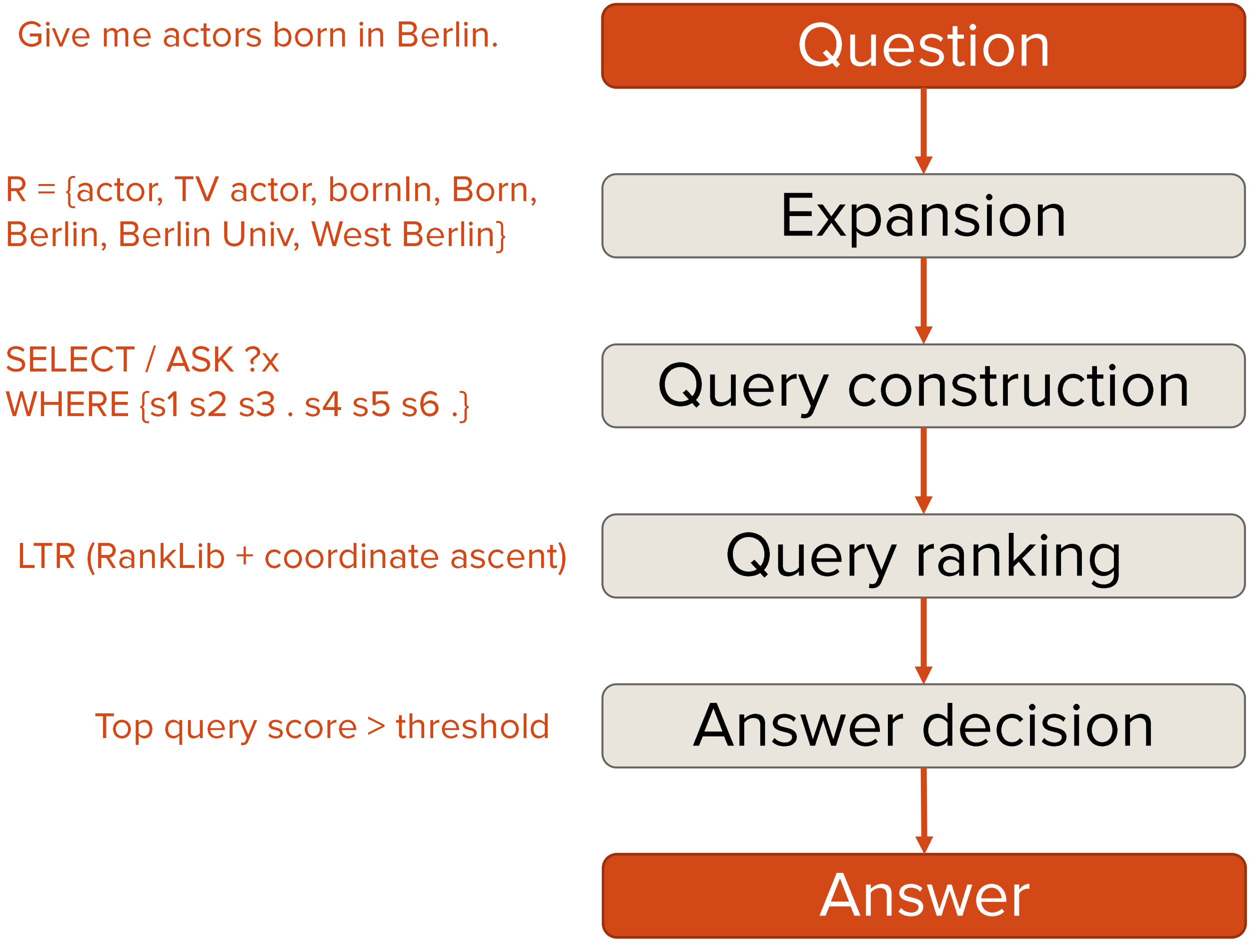
Structured querying over multiple KGs

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA



Structured querying over multiple KGs

- QAnswer (Diefenbach et al. 2019)
- Multiple KGs as unified triple store
- KG-agnostic approach for QA
- Extremely efficient due to HDT and additional KG indexing ☺
- Syntax agnostic ☹



Heterogeneous QA: Wrap-up

- Early fusion and unified representation truer to spirit of heterogeneous QA than late fusion
- PullNet uses early fusion ☺ and deals with complex questions ☺
- But in principle only for chain joins ☹
- Efficiency is an open issue (too many predictions), no. of hops assumed to be known ☹
- QAnswer is efficient ☺ and works over multiple KGs (largely unexplored) ☺
- But works mostly for relatively simple questions ☹
- Current systems still **not truly unified:** reliance on **KG entities** for linking and distant supervision, and a **triplified view** of knowledge

break duration ?x .
?x measuredIn minutes .

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KG and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

How can we deal with information needs spread over multi-turn conversations?

Conversational KG-QA

Which actor voiced the character Unicorn in The Last Unicorn?

Mia Farrow

Which role was voiced by Alan Arkin in the Last Unicorn?

Schmendrick

Who performed the songs in the movie The Last Unicorn?

America

What is the genre of the band that performed the songs in The Last Unicorn?

Folk rock

Who was the director of the movie The Last Unicorn?

Jules Bass

Conversational KG-QA

Which actor voiced the character Unicorn in The Last Unicorn?

Mia Farrow

And Alan Arkin was behind ...?

Schmendrick

The songs were by...?

America

Genre of this band?

Folk rock

By the way, who directed the movie?

Jules Bass

Conversational KG-QA

- Information needs **rarely one-off**
- Sequence of **follow-up questions** on a topic
- Analogous to **search sessions** and **interactive retrieval**
- Users want to simulate **natural experience** with assistant
- **Leave context unspecified** in follow-ups

Conversational KG-QA

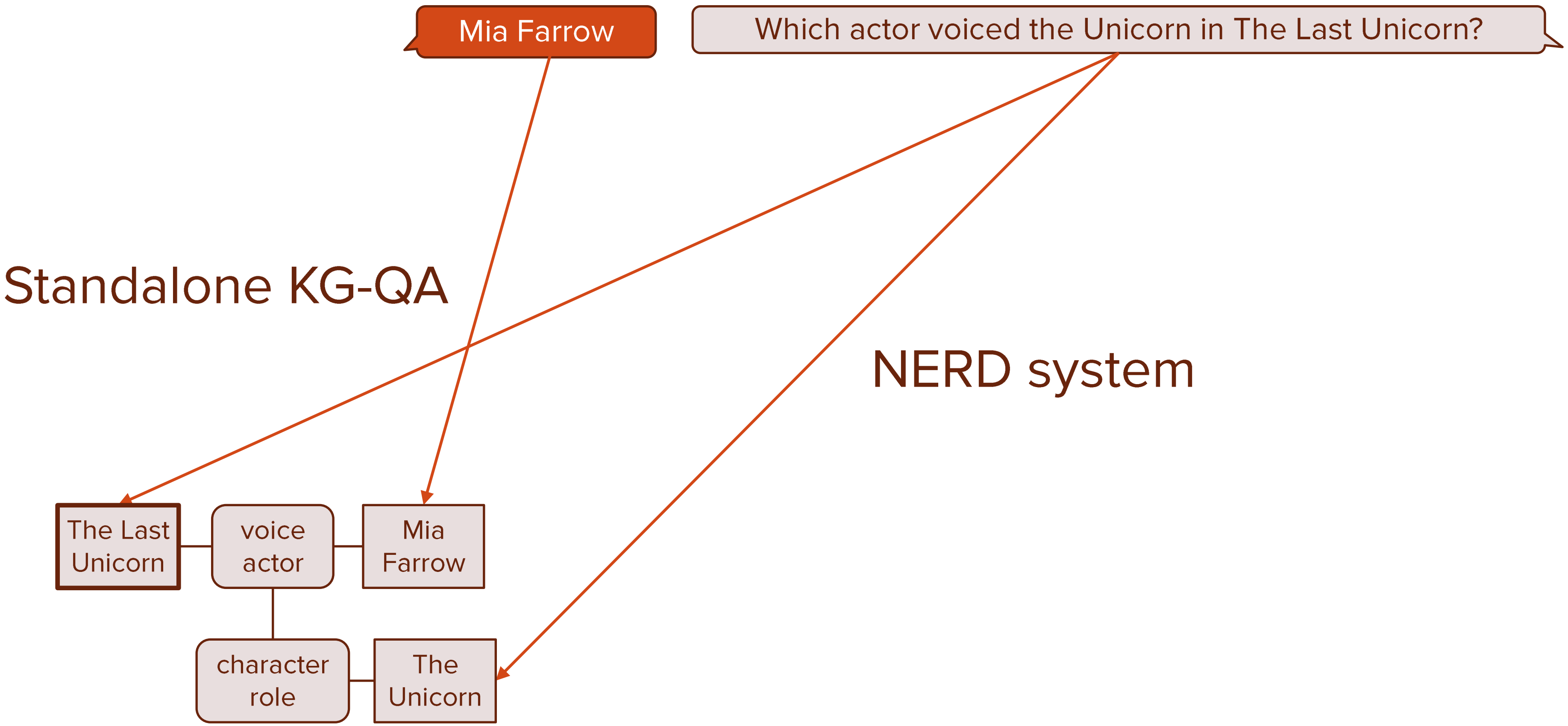
- **Key challenges** in conversational (KG-)QA
  - Infer implicit context
  - Handle ad hoc formulations
- Initially explored over **small tables** as sequential QA (Iyyer et al. 2017)
- Key direction for KG-QA now (Saha et al. 2018, Guo et al. 2018, **Christmann et al. 2019**, **Shen et al. 2019**)

Conversational QA: Graph traversal

- The **CONVEX** system (Christmann et al. 2019)
- Large topic jumps in conversations are rare: establish **localized KG context**
- Harness **KG-connectivity:** No need to complete/rewrite questions
- **Expand context judiciously** with relevant entities and predicates in neighborhood
- **Unsupervised** iterative graph traversal (c.f. supervised graph traversal in PullNet)
- CONVEX works on top of **any KG-QA system** to handle conversations

Christmann et al., Look before you Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion, CIKM 2019.

Initial context

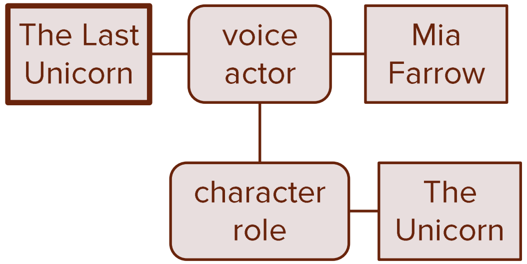


Initial context

Mia Farrow

Which actor voiced the Unicorn in The Last Unicorn?

And Alan Arkin was behind . . .?



How to expand the context?

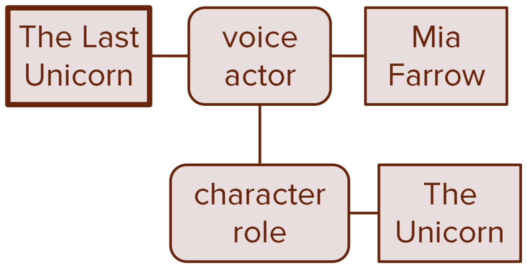

Neighborhood of
Mia Farrow
Neighborhood of
The Last Uni
borhood of
Unicorn

Judicious context expansion

Mia Farrow

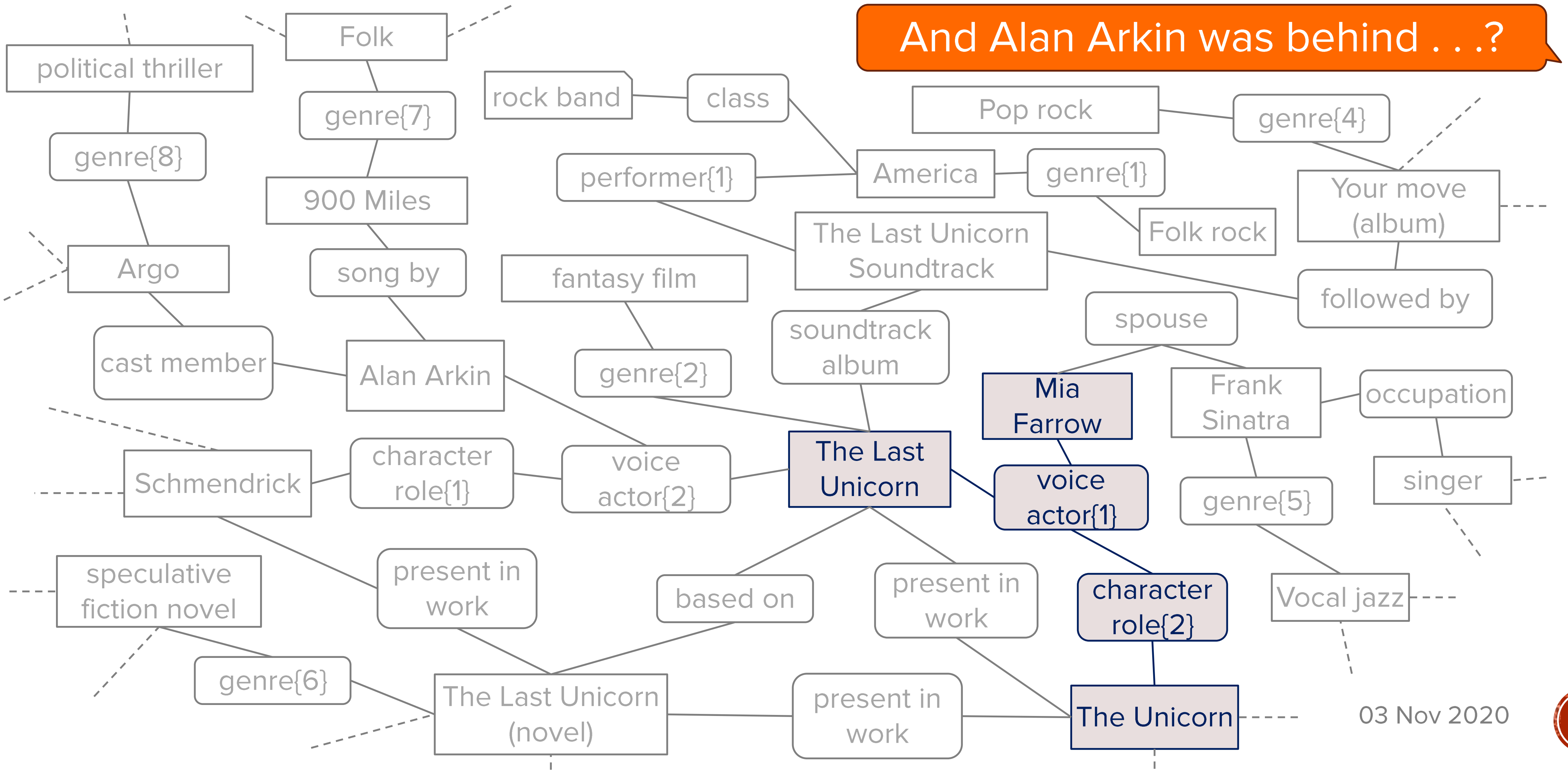
Which actor voiced the Unicorn in The Last Unicorn?

And Alan Arkin was behind . . .?

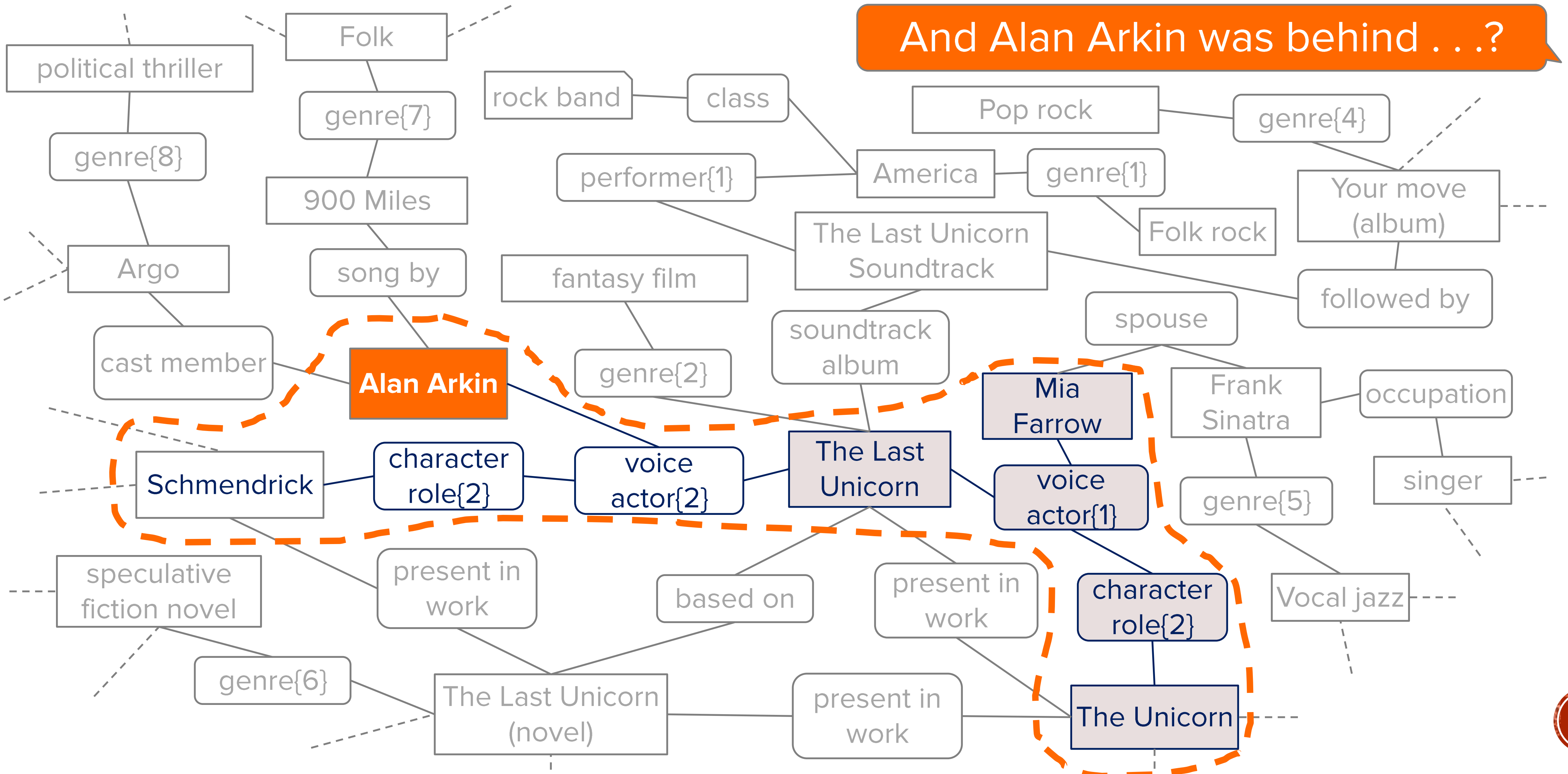


Do not expand with the complete neighborhood!

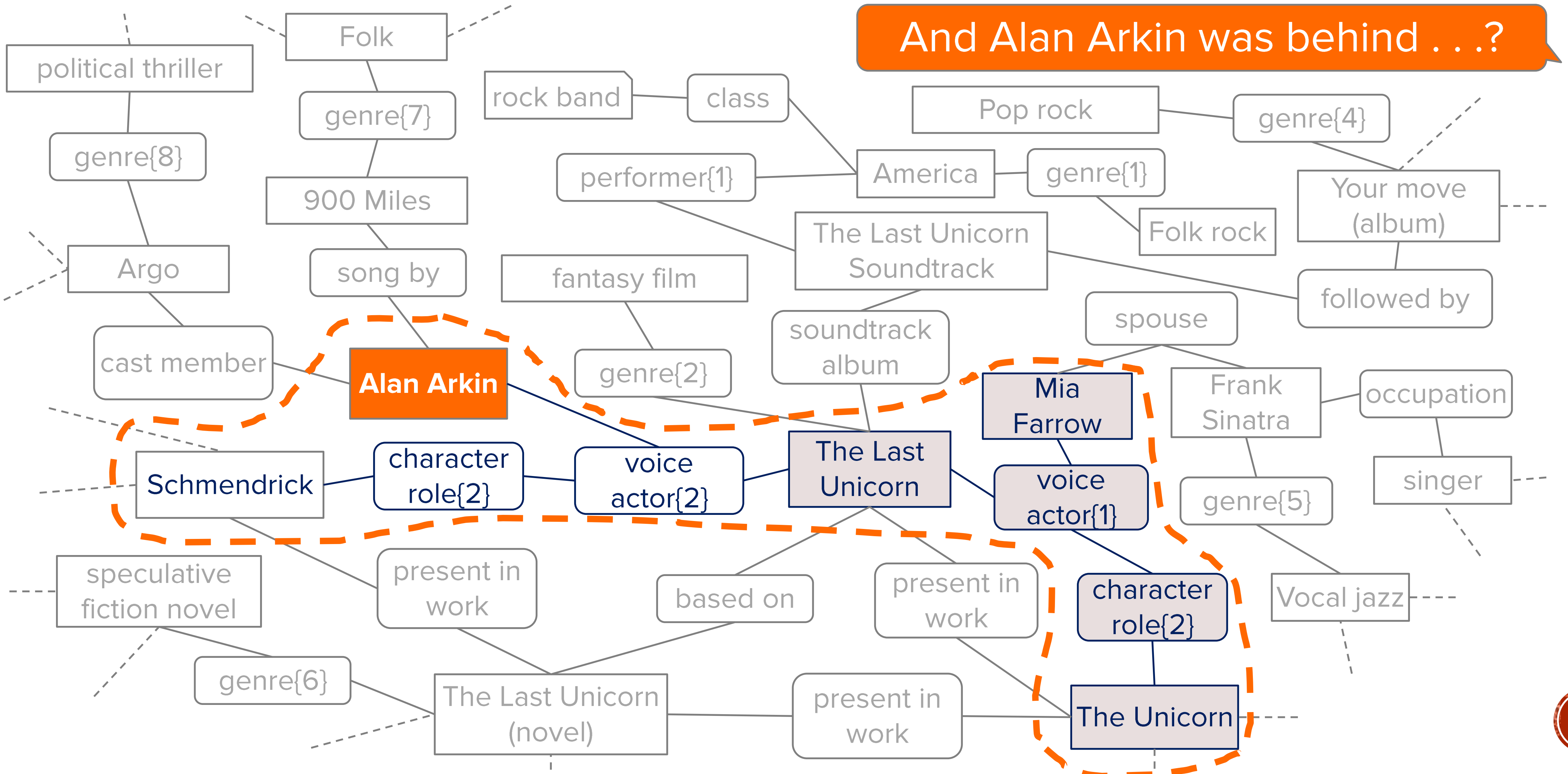
Exploring context neighborhood



Exploring context neighborhood



Find frontier nodes to define expansion border

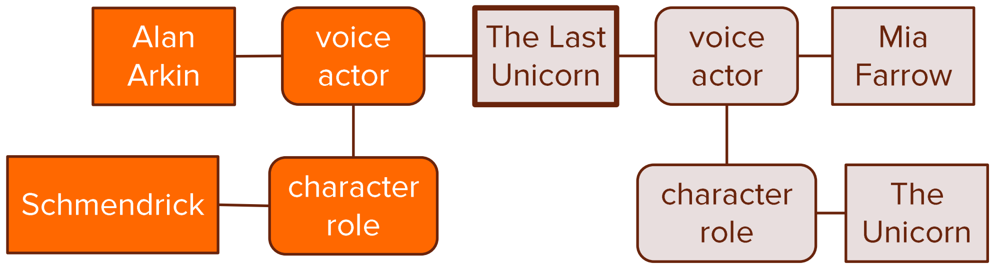


Context graph

Mia Farrow

Which actor voiced the Unicorn in The Last Unicorn?

And Alan Arkin was behind . . .?



Graph expanded with relevant facts only

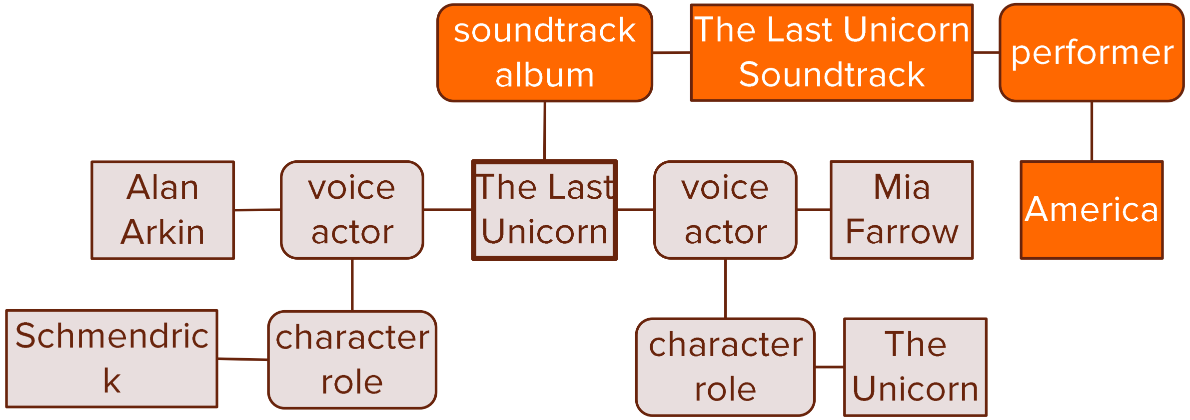
Context graph

Mia Farrow

Which actor voiced the Unicorn in The Last Unicorn?

And Alan Arkin was behind . . .?

So who performed the songs?



Graph expanded with relevant facts only

Context graph

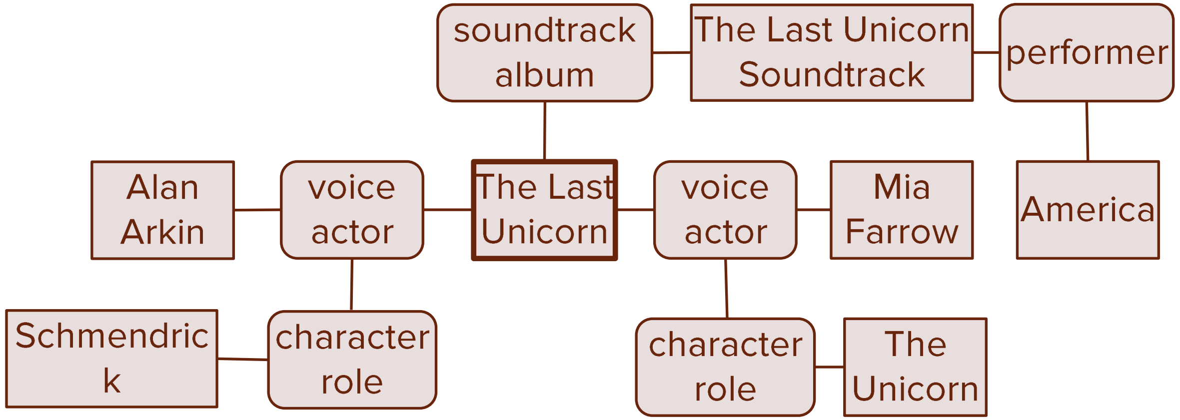
Mia Farrow

Which actor voiced the Unicorn in The Last Unicorn?

And Alan Arkin was behind . . .?

So who performed the songs?

Genre of this band?



How to determine Frontier nodes?

Frontier score

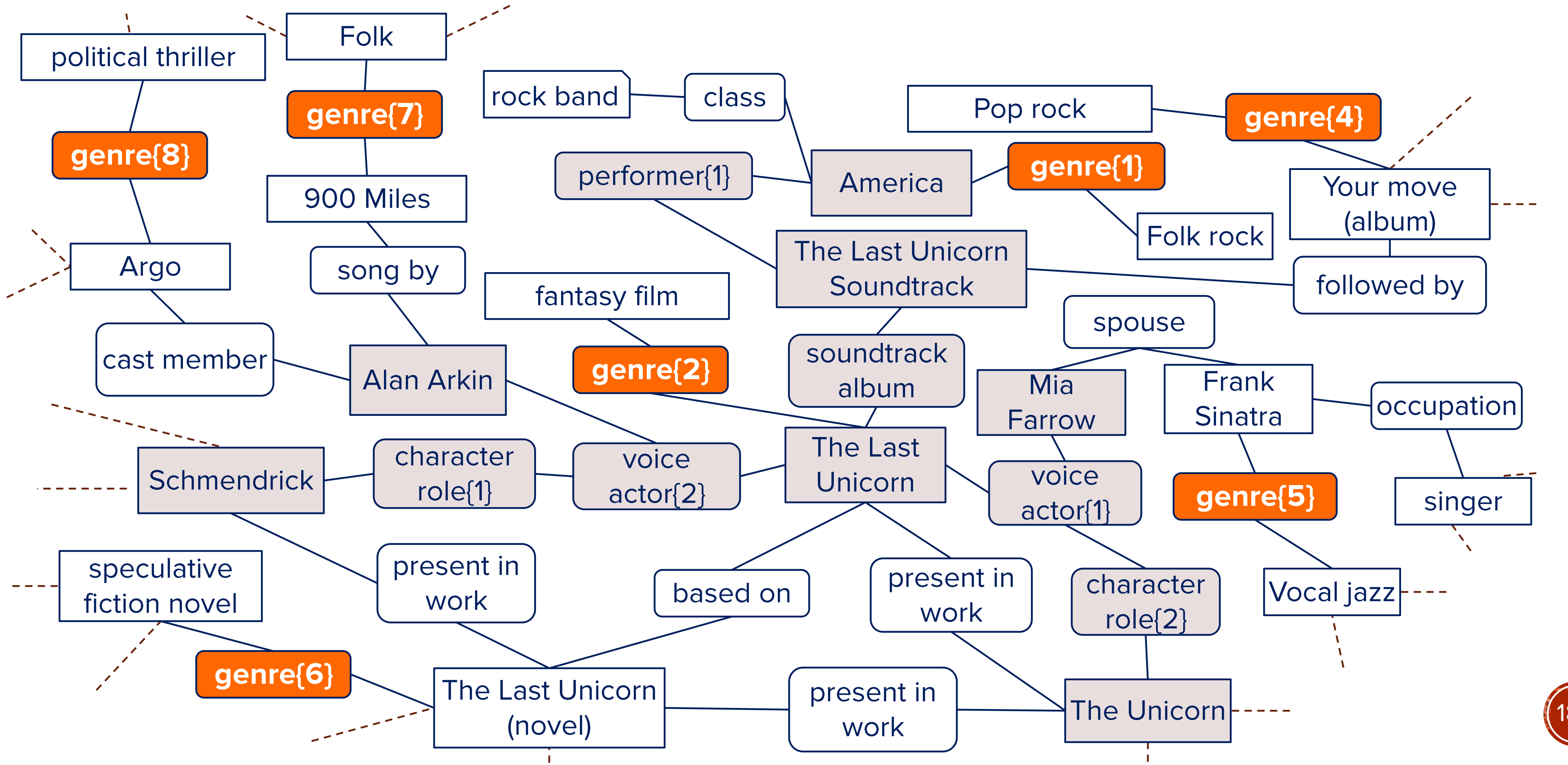
Matching similarity	Context relevance	KG priors
match (candidate c)	*prox (candidate c)*	*prior (candidate c)*

$$frontier\_score(candidate\ c) = h_1 \cdot match(c) + h_2 \cdot prox(c) + h_3 \cdot prior(c)$$

With hyperparameters h_1, h_2, h_3

The great disambiguation

Genre of this band?



Frontier nodes

Matching similarity

Candidate	*Match*
genre{1}	1.00
genre{2}	1.00
…	…
folk rock band	0.89
RSH-Gold for Cult Band	0.87
fantasy film	0.36
…	…

Context relevance

Candidate	*Prox*
genre{1}	0.91
folk rock band	0.86
RSH-Gold for Cult Band	0.86
…	…
genre{2}	0.34
fantasy film	0.36
…	…

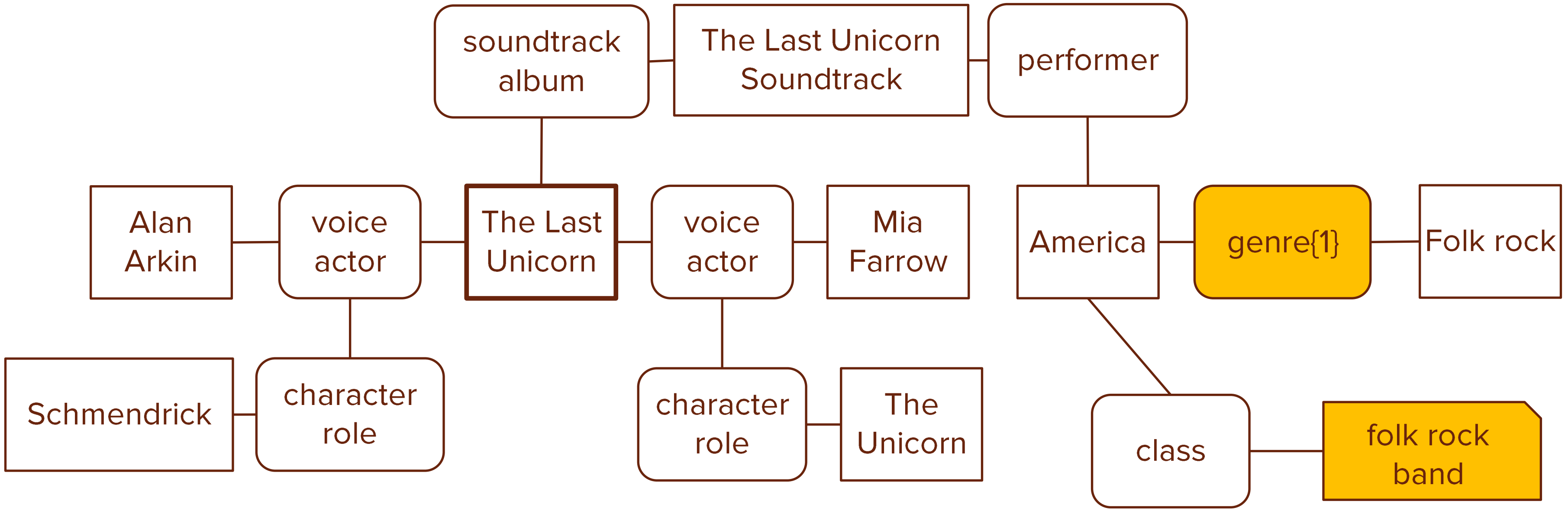
KG priors

Candidate	*KG priors*
…	…
genre{1}	0.56
genre{2}	0.56
…	…
folk rock band	0.34
…	…
RSH-Gold for Cult Band	0.01

Fagin's Threshold Algorithm (FTA) to retrieve top-*k* ranked nodes according to frontier score

Frontier nodes

Genre of this band?



Frontier nodes

Answer to the question

Genre of this band?

- Distance to Frontier nodes
  - Weighted by the frontier score
  - distance_F

=> Explicit part

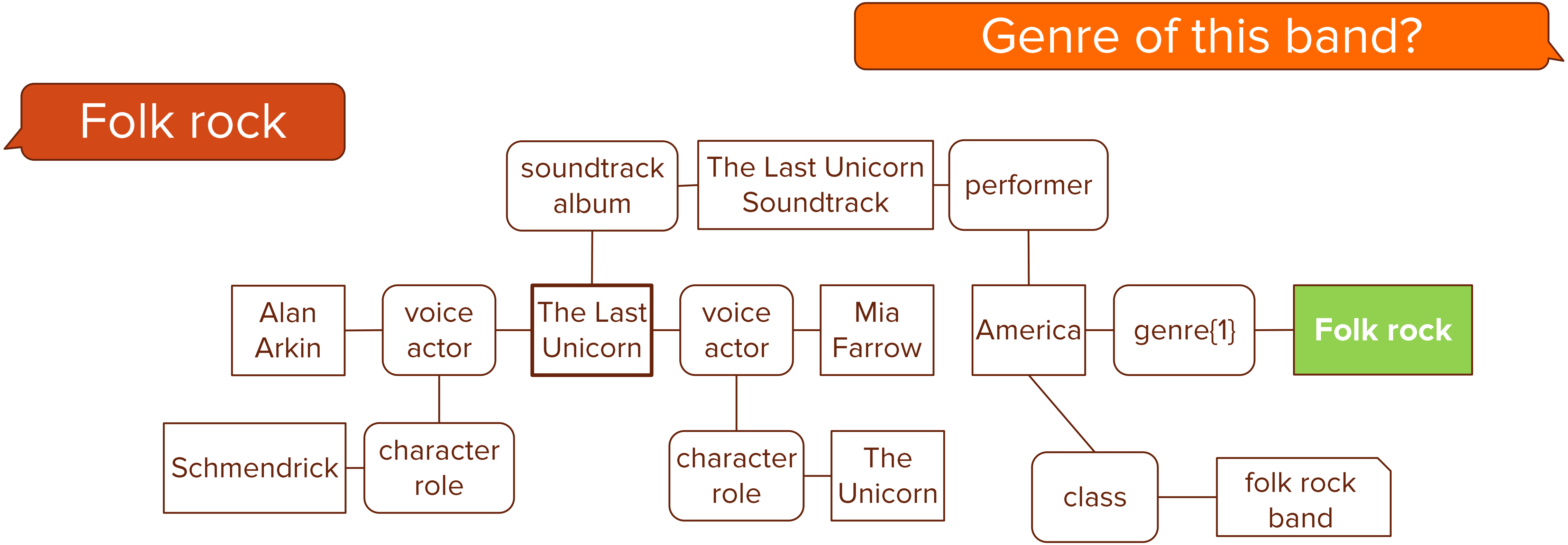
- Distance to all nodes in context graph X
  - Weighted by the turn they occurred in
  - distance_X

=> Implicit part

$$answer\_score(candidate\ c) = h_4 \cdot distance\_F + h_5 \cdot distance\_X$$

Christmann et al., Look before you Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion, CIKM 2019.

Answer detection



Top-ranked node according to *answer_score*

Conversational QA: Sequence-to-sequence modeling

- The **MaSP** model (Shen et al. 2019)
- **Shared supervision** for tasks: Entity detection and answering
- **Grammar-based** semantic parsing model
- Designed to resolve **coreference** in conversations
- **Type-aware** entity detection
- Uses **transformers** for sequence encoding

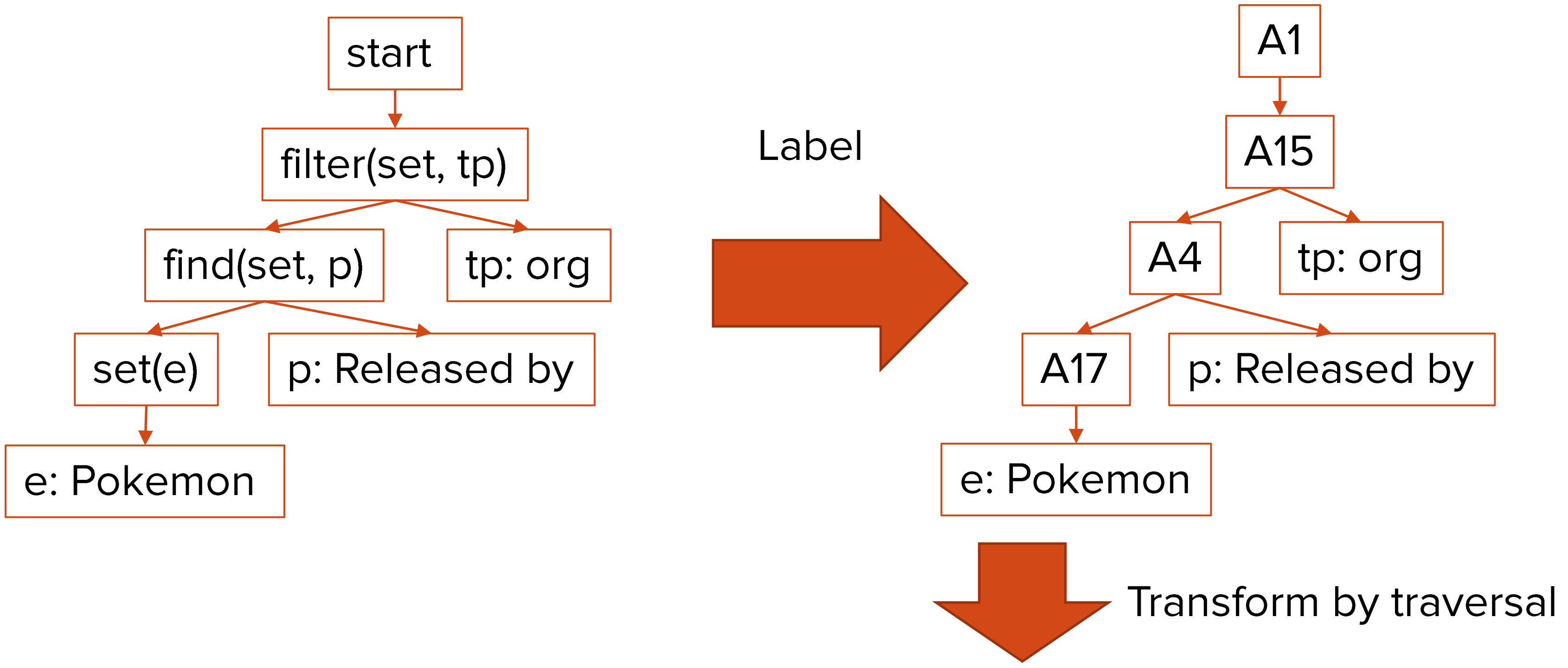
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

Sequence-to-sequence model

NL question as sequence

Question: Who released Pokemon?

Tree-structured KB-query



Query as sequence

A1 A15 A4 A17 e:Pokemon p:ReleasedBy tp:org

Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

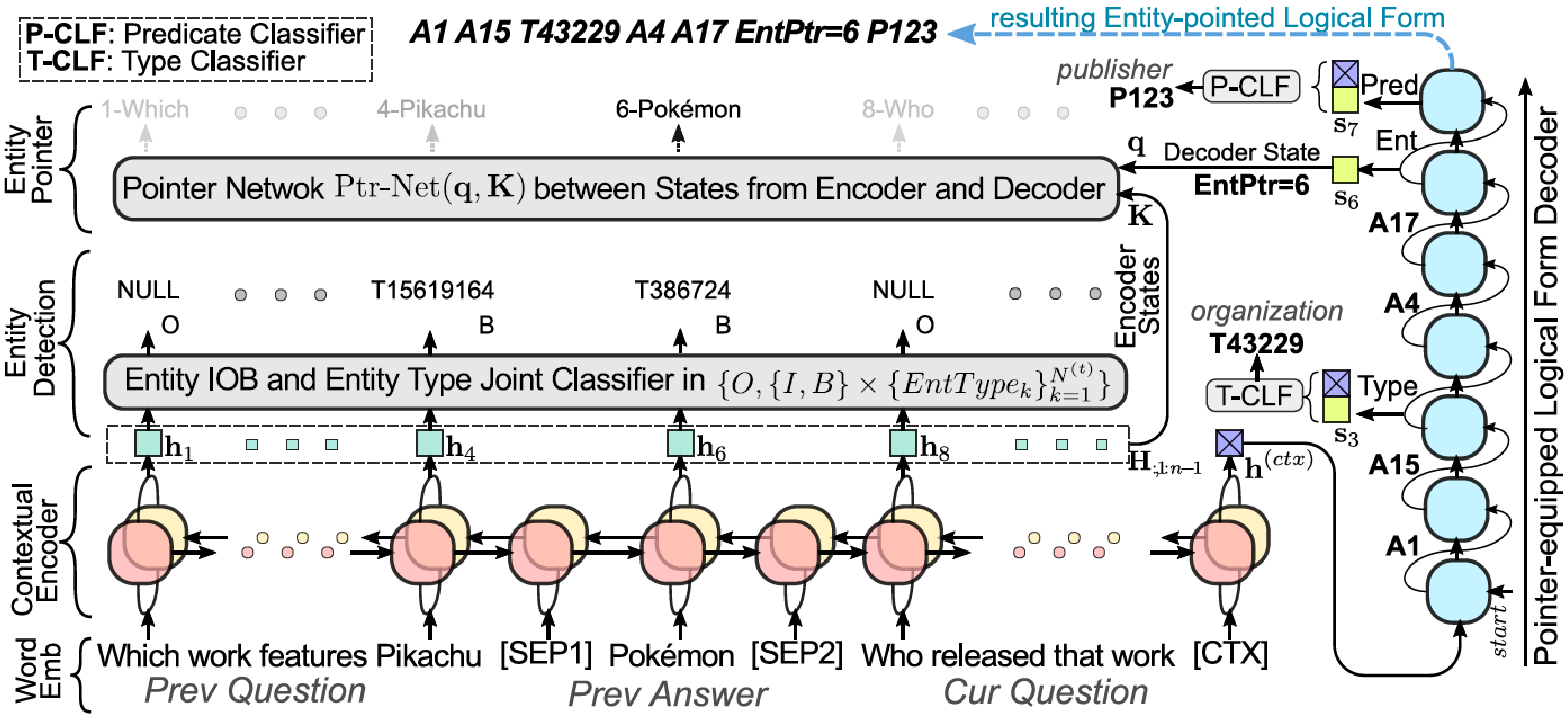
Sequence-to-sequence model

NL question as sequence

Question: [CONTEXT] Who released that work?

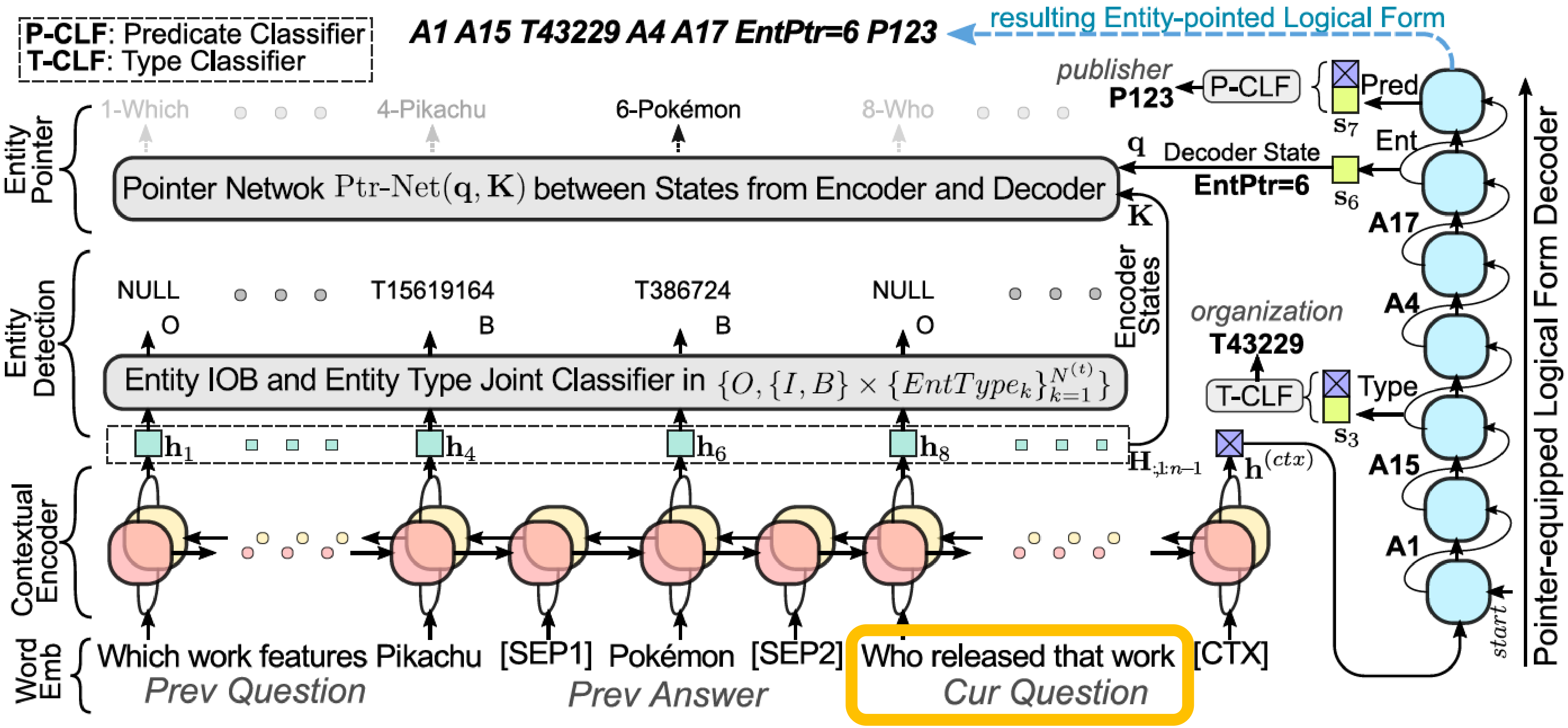
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

Seq-to-seq: The MaSP model



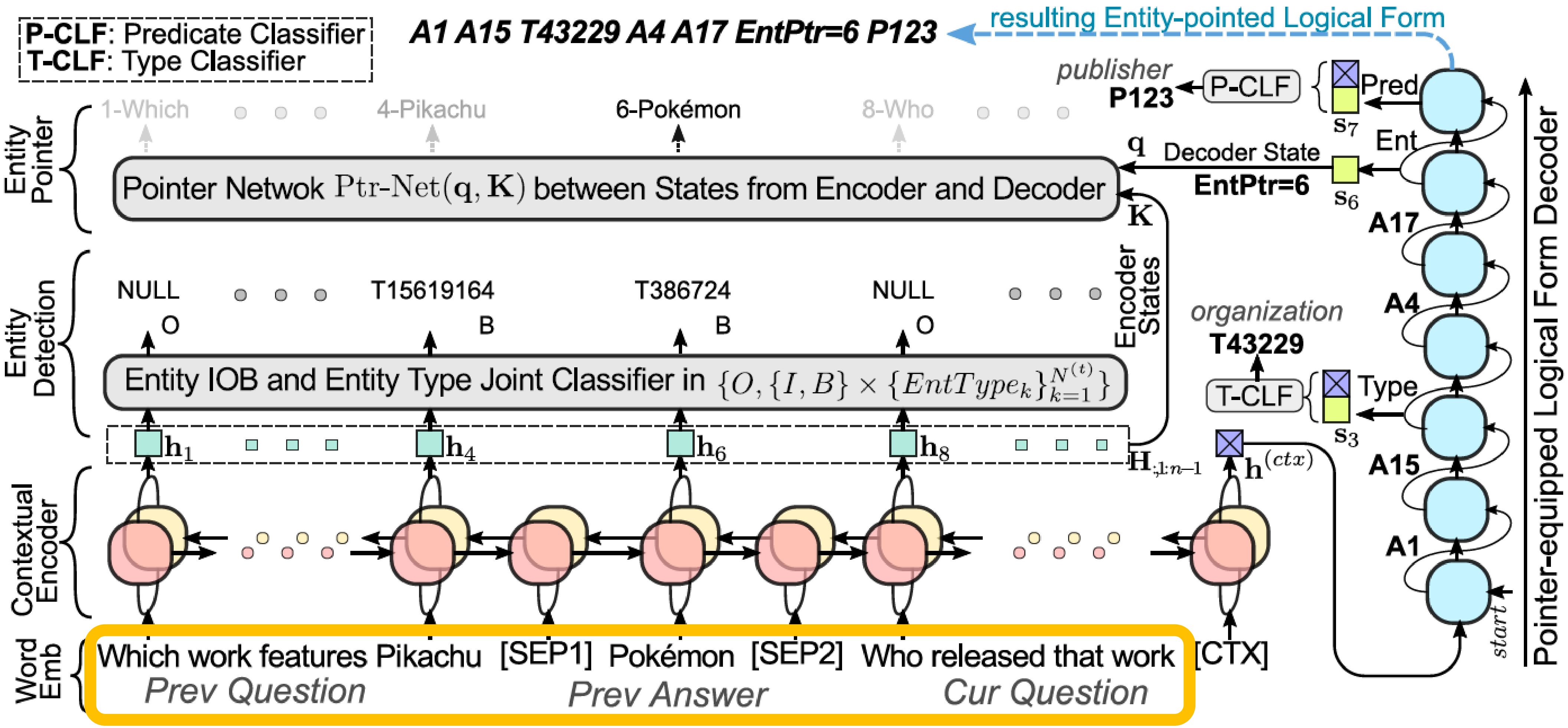
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



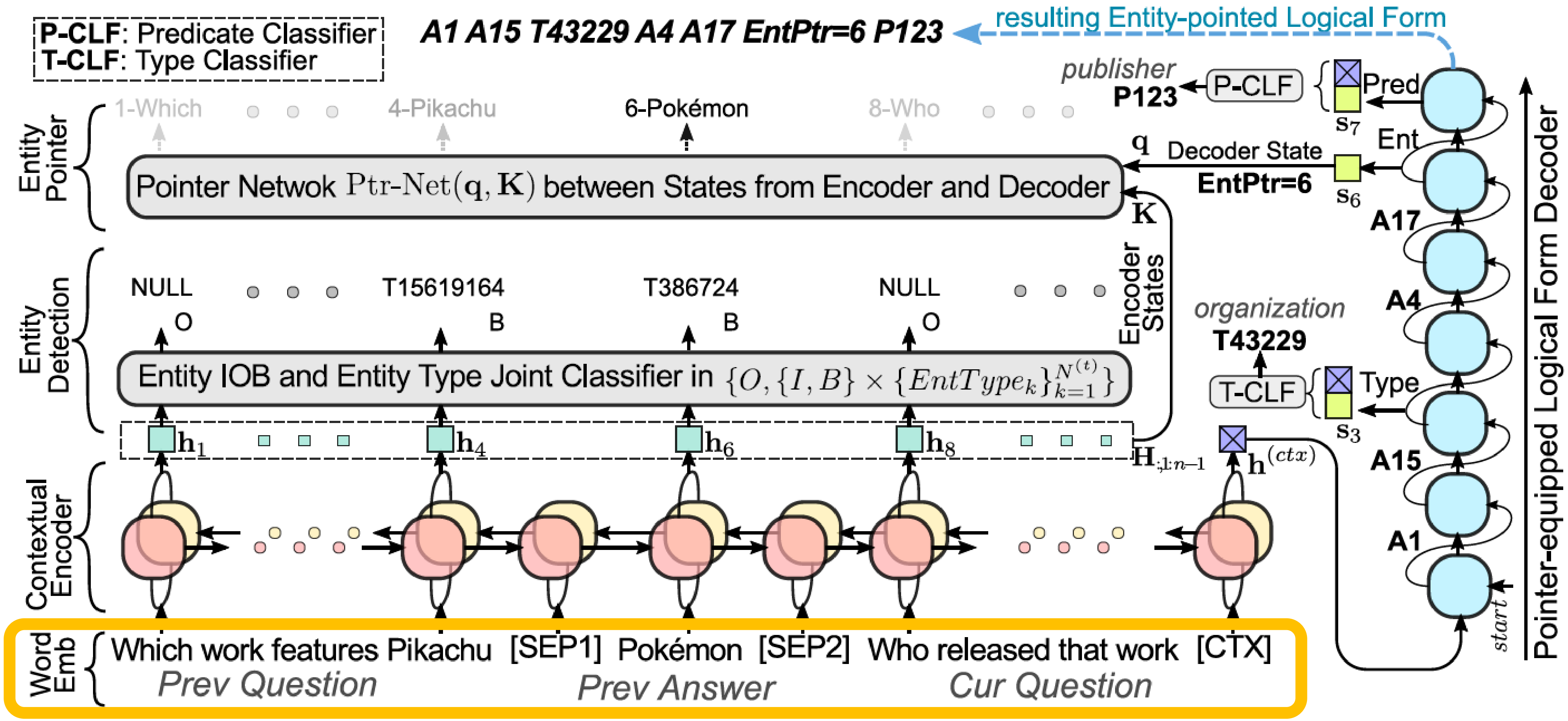
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



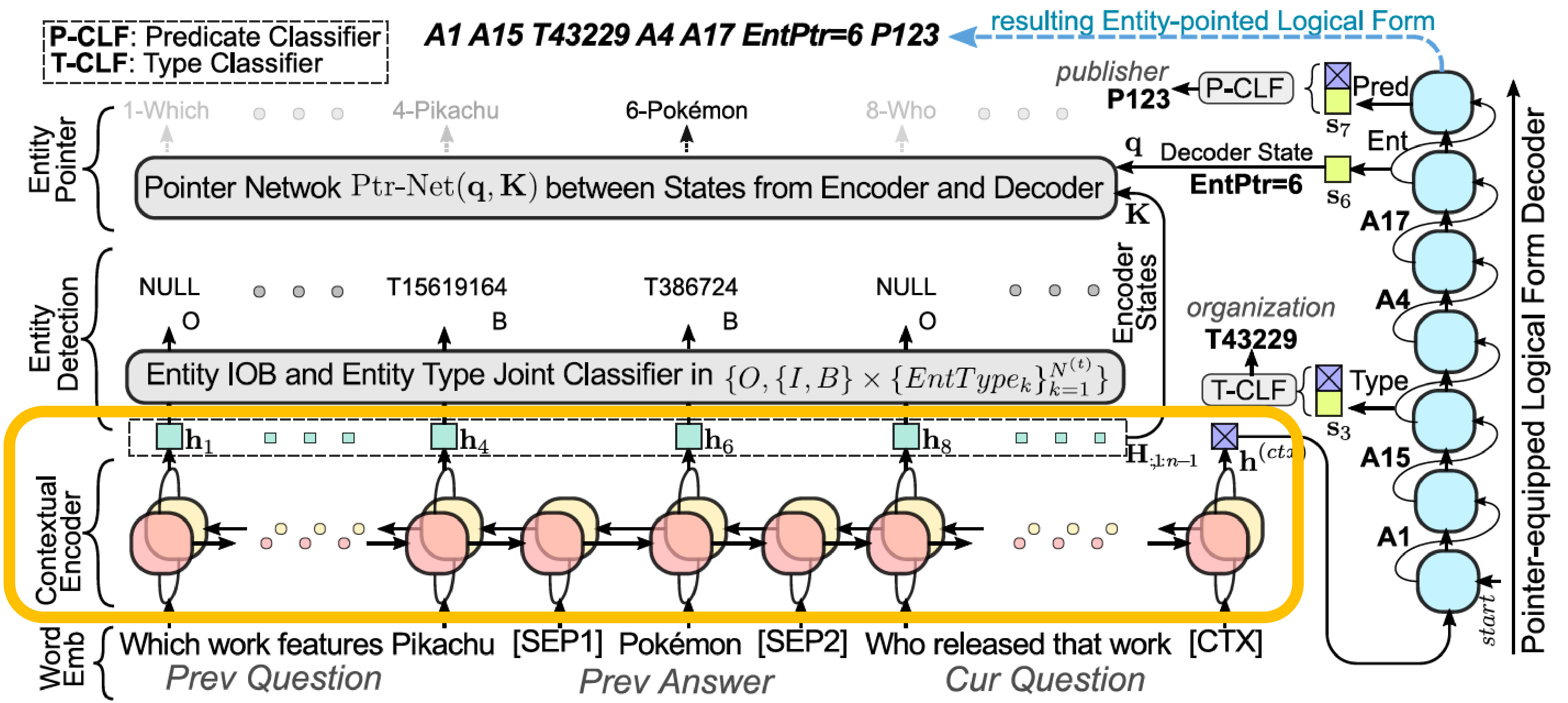
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



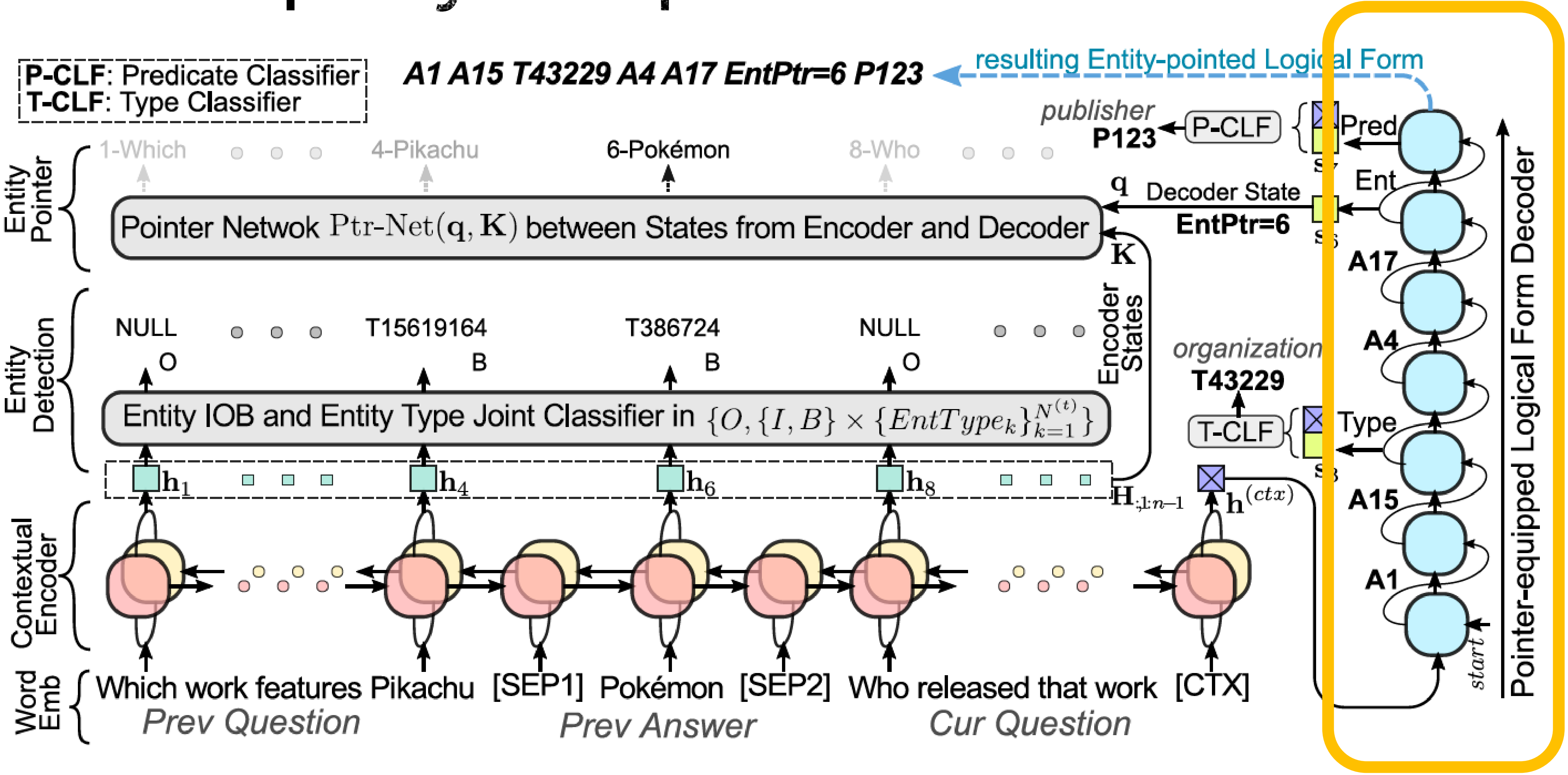
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



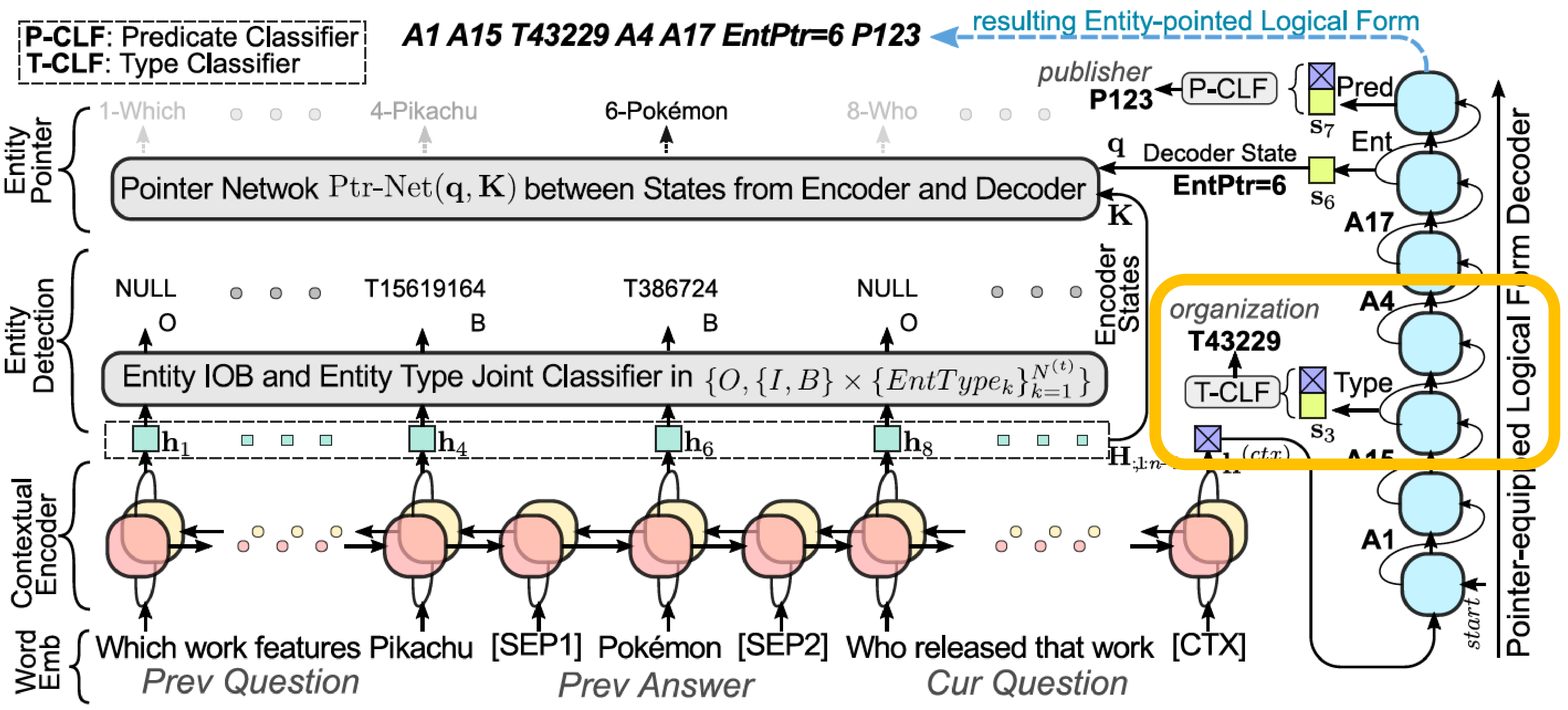
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



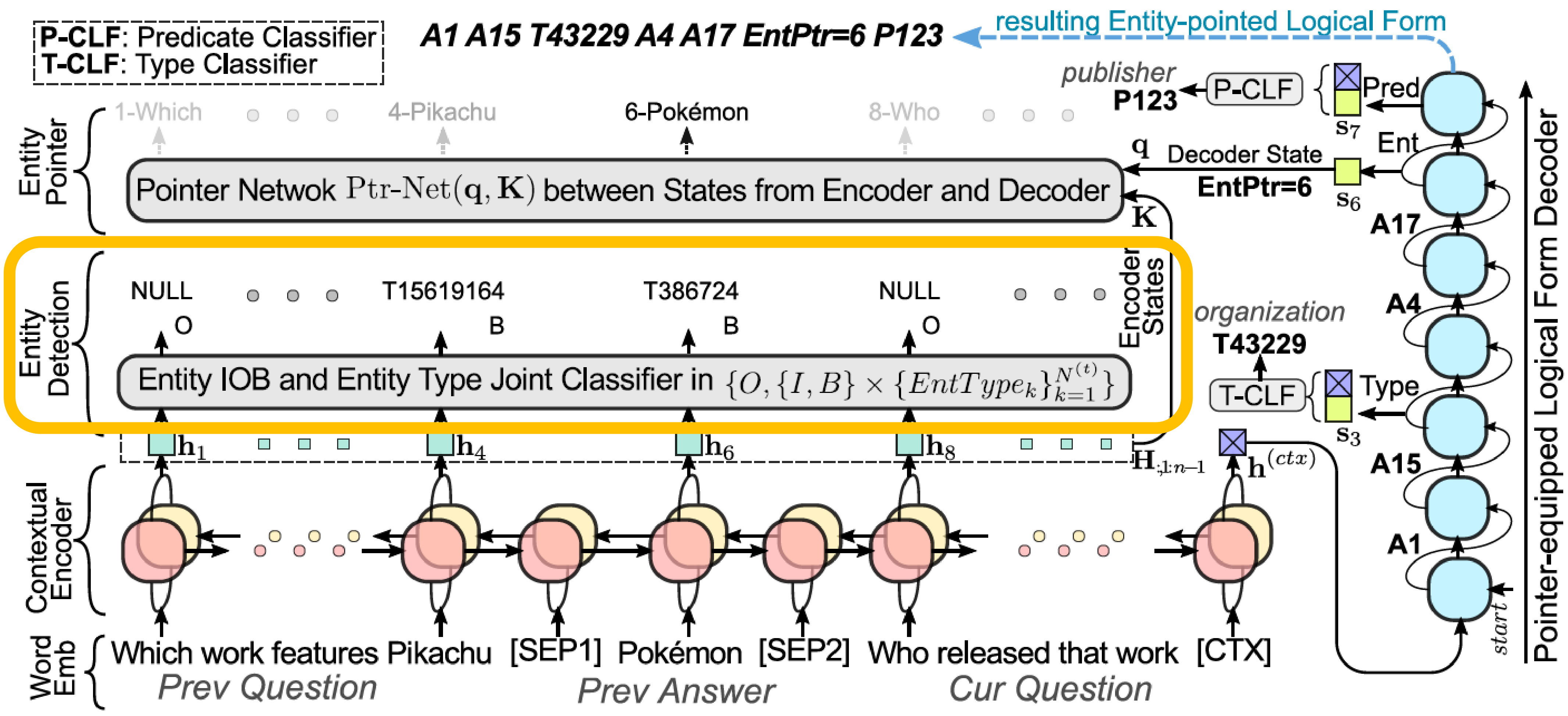
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



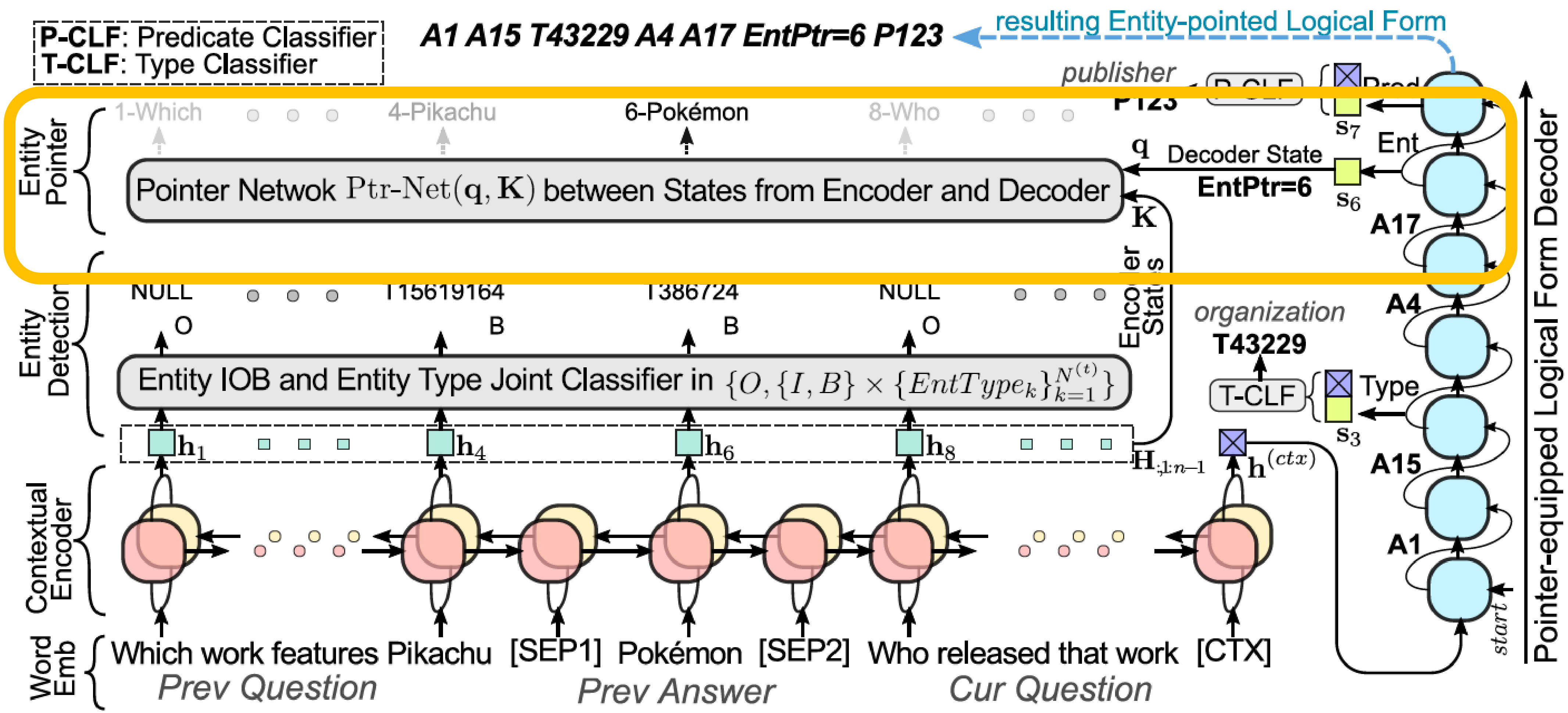
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



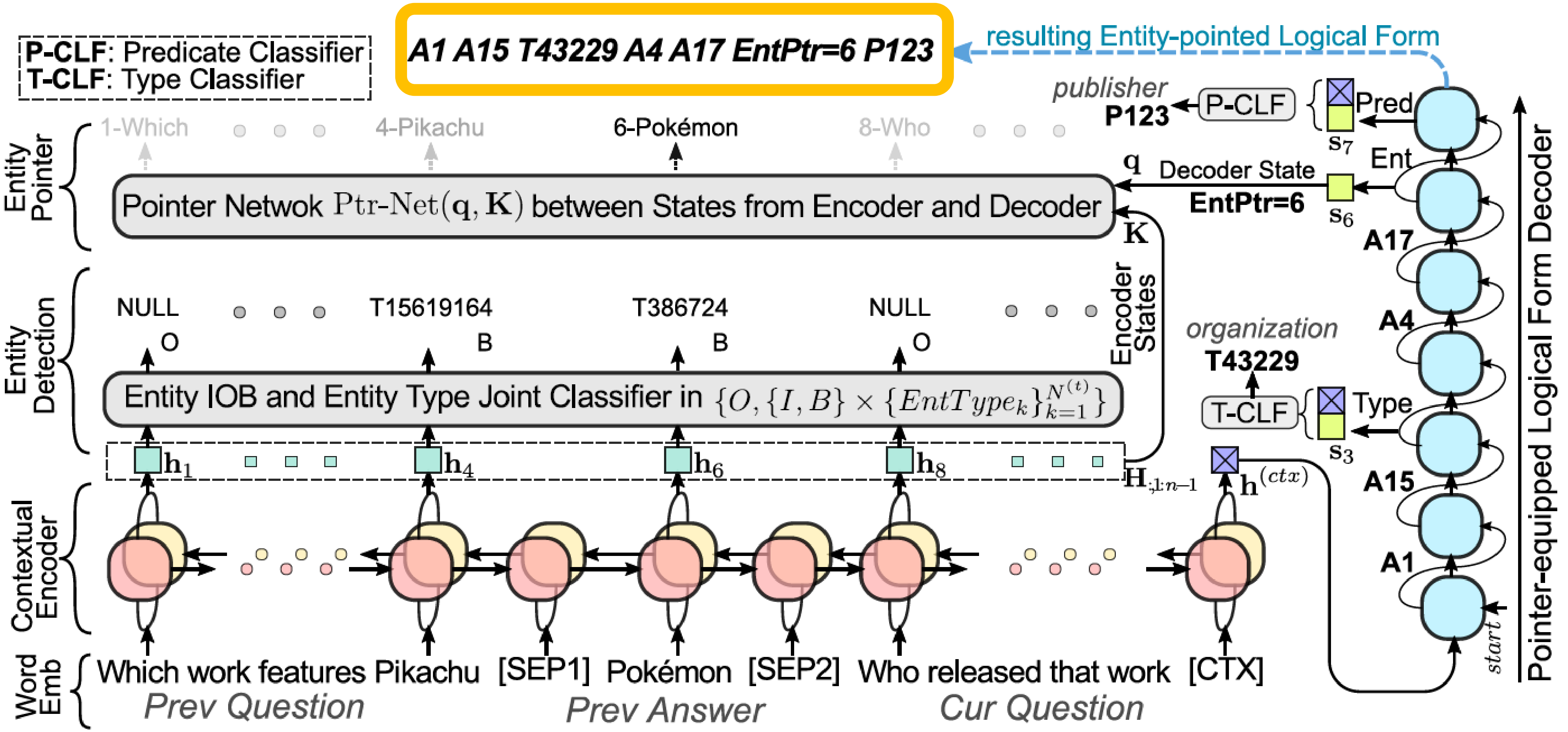
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



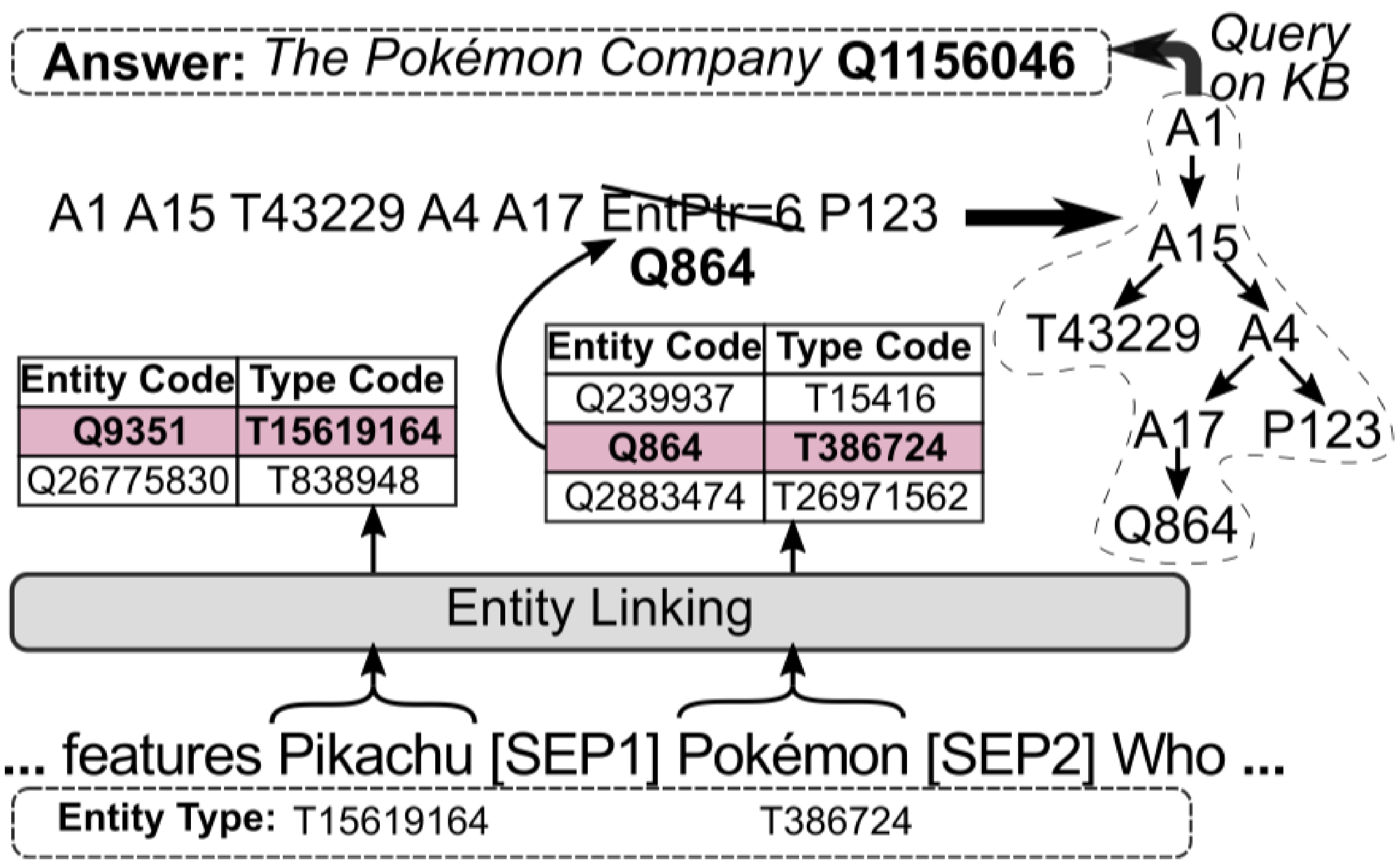
Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

MaSP: Step-by-step



Shen et al., Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base, EMNLP 2019.

Execute query obtained via sequence decoding to get answer



Conversational QA: Wrap-up

- Unsupervised graph traversal is promising
- KG connections offer vital clues for initializing and expanding context
- But limited to relatively simple information needs in utterances
- Sequence-sequence models can capture context well
- But ConvQA is much more than coreference and ellipsis resolution
- Zero-coreference / zero-anaphora utterances common ("batman actor?")
- Question completion may be intractable + overkill

Side glance: Table-QA

break duration ?x .
?x measuredIn minutes .

- Web tables also constitute a huge volume of the curated Web
- Represent canonical challenges of querying a large-scale KB
- Selected references below for the interested reader

Chakrabarti et al., Open Domain Question Answering Using Web Tables, arXiv 2020.

Zhang, CFGNN: Cross Flow Graph Neural Networks for Question Answering on Complex Tables, AAAI 2020.

Wang et al., A Neural Question Answering Model Based on Semi-Structured Tables, COLING 2018.

Iyyer et al., Search-based Neural Structured Learning for Sequential Question Answering, ACL 2017.

Jauhar et al., Tables as Semi-structured Knowledge for Question Answering, ACL 2016.

Khashabi et al, Question Answering via Integer Programming over Semi-Structured Knowledge, IJCAI 2016.

Sun et al., Table Cell Search for Question Answering, WWW 2016.

Pasupat and Liang, Compositional Semantic Parsing on Semi-Structured Tables, ACL 2015.

Outline: QA over knowledge graphs

- **Background:** Setup, benchmarks, metrics
- **Simple QA:** Templates and embeddings
- **Complex QA:** Multiple entities and predicates
- **Heterogeneous sources:** Handling KGs and text
- **Conversational QA:** Implicit context in multi-turn setup
- **Take-home:** Summary and insights

Summary and insights

Take-home messages

- Methodology
- Deployable system
- Open problems

Methodology

- Templates
- Graph embeddings
- Subgraph computations
- Belief propagation
- Graph traversal
- Sequence-to-sequence models

Methodology

- Templates: NEQA
- Graph embeddings: KEQA
- Subgraph computations: TextRay, QUEST, QAnswer
- Belief propagation: QAmp
- Graph traversal: PullNet, CONVEX
- Sequence-to-sequence models: MaSP

Methodology quad chart

Degree of supervision ↑	Explicit structured query	Without explicit query (approximate graph search)
Weakly/strongly supervised	MaSP, NEQA, TextRay	KEQA, PullNet
Unsupervised	QAnswer*	QAmp*, QUEST, CONVEX

Degree of supervision

Query generation →

* Ranker/labeler supervised

Methodology quad chart

Weakly/strongly supervised	MaSP, TextRay, NEQA, D2A, Saha et al. 2018, SEMPRE, AQQU, STAGG, PARALEX, OQA, PARASEMPRE, Cai and Yates 2013	KEQA, PullNet, GRAFT-Net, GraphParser, Bordes et al. 2014
Unsupervised	QAnswer*, DEANNA, Unger et al. 2012	QAmp*, QUEST, CONVEX
	Explicit structured query	**Without explicit query (approximate graph search)**

Degree of supervision

Query generation

* Ranker/labeler supervised

Methodology quad chart

Strongly supervised with (Q, q)	Cai and Yates 2013	-	-
Weakly supervised with (Q, A)	MaSP, NEQA, TextRay, D2A, Saha et al. 2018, SEMPRE, AQQU, STAGG	KEQA, PullNet, GRAFT-Net, GraphParser, Bordes et al. 2014	There is some interplay in current systems but largely open area Unsupervised subgraph computations with small degree of supervised neural learning..?
Weakly supervised with paraphrases	PARASEMPRE, PARALEX, OQA	-	
Unsupervised	QAnswer*, DEANNA, Unger et al. 2012	QAmp*, QUEST, CONVEX	
	Explicit structured query	**Without explicit query**	**Hybrid search methods?**

Methodology: Pros and cons

Aspect	**With explicit structured query (SPARQL-like)**	**Without explicit structured query (approx. graph search)**
Simple questions	☺	☺
Single answer	☺	☺
List answer	☺	☹?
Efficiency	☺	☹?
Complex questions	☹?	☺
Conversational questions	☹?	☺
Heterogenous sources	☹?	☺
Handling reified triples	☹?	☺

☺ Preferable ☹? Less preferable but scope for improvement

Methodology: Lessons learnt

- Templates good for simple questions, but hits hurdles for complex questions, and useless for conversational ☹
- Graph embeddings effective for simple questions ☺ , not yet clear for complex scenarios…
- Sequence models (LSTM with attention) with pre-trained word embeddings very common
- Graph models generally more flexible (scope for node/edge types/weights)

Deploying a QA system

- Templates and unsupervised graph methods great way to get off the blocks with limited complexity
- Preferably with NER/NERD systems and pre-trained word embeddings
- Need seed data + domain knowledge
- Continuous learning with similarity function and feedback vital cogs
- Level of structure and heterogeneity in data and questions indicators of follow-up modeling

Open problems

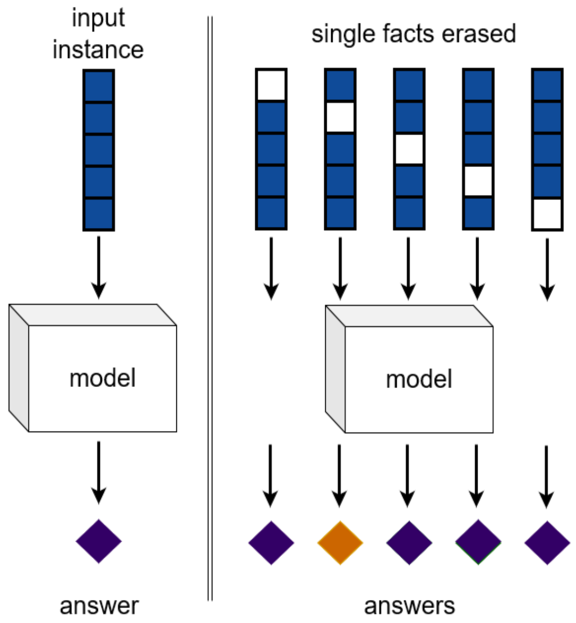
- Unanswerability
- Interpretability
- Interactivity
- Efficiency
- Robustness

Open problems: Unanswerability

- Learn when to stay quiet and prevent embarrassment ☺
  - *Where was Messi's father born?*
  - *Who was the first man on Mars?*
- Knowing when answer is:
  - Not confident
  - Not in KG
  - Null
- Open and closed world assumptions
- Learn when to consult text

Open problems: Interpretability

- Are your system's answers explainable? To the developer? What about the end user?
- Interpretability increases trust and guides user in case of mistakes
- Template- and graph-based methods construct interpretable evidence for answers - an unsolved concern for neural methods
- Sydorova et al. (2019) provide insights with **input perturbation** and **evaluation** of interpretability
- But very much an open problem!



Open problems: Interactivity

- Towards mixed initiative systems (Radlinski and Craswell 2017)
- Can your system absorb feedback?
- Positive and negative feedback?
- What kinds of feedback?
- Can your system ask clarifications?

Open problems: Efficiency

- Critical component of QA systems
- Largely unexplored
- Identify bottlenecks
- Measure trade-offs

Open problems: Robustness

- Think out of the ~~box~~ benchmark
- What is **open-domain** question answering?
- What happens for entities not seen during training?
- What about unseen predicates and vocabulary?

Take-home messages

- Overview of state-of-the-art in KG-QA and their positioning
- Families of algorithms with a few specific instantiations
- Several open problems in the key areas of focus

Simple / complex / heterogeneous / conversational questions for me ☺ ?

QA@MPII-D5: Visit qa.mpi-inf.mpg.de

- **Course** on QA systems: https://www.mpi-inf.mpg.de/question-answering-systems/
- **CONVEX:** Conversational QA over KGs [CIKM 2019]: https://convex.mpi-inf.mpg.de/
- **CROWN:** Conversational QA over passages [SIGIR 2020]: https://crown.mpi-inf.mpg.de/
- **QUEST:** Complex question answering [SIGIR 2019]: https://quest.mpi-inf.mpg.de/
- **ComQA:** QA benchmark with paraphrase [NAACL 2019]: http://qa.mpi-inf.mpg.de/comqa/
- **TEQUILA:** Temporal question answering [CIKM 2018]: https://tequila.mpi-inf.mpg.de/
- **QUINT:** Template-based question answering [EMNLP 2017]: https://quint.mpi-inf.mpg.de/
- Send an email to rishiraj@mpii.de in case of any issues!

Acknowledgements

- Gerhard Weikum for valuable feedback on slides
- Authors of several papers for sharing additional content
- Members of D5@MPII for inputs