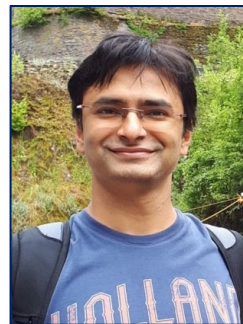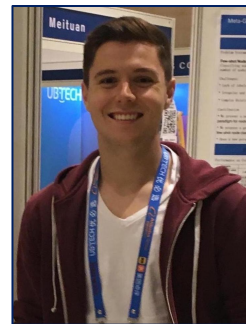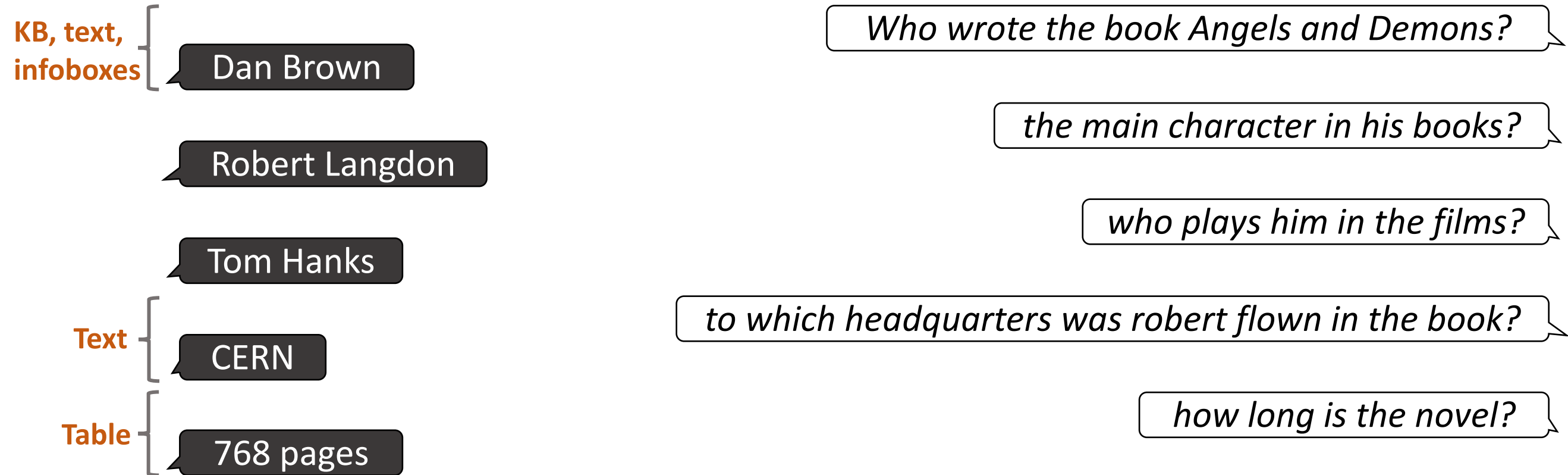# Explainable Conversational Question Answering over Heterogeneous Sources

Philipp Christmann, Rishiraj Saha Roy, Gerhard Weikum

Max Planck Institute for Informatics, Saarbrücken, Germany

Conversational Question Answering

KB, text, infoboxes

Who wrote the book Angels and Demons?

Dan Brown

the main character in his books?

Robert Langdon

who plays him in the films?

Tom Hanks

Text

to which headquarters was robert flown in the book?

CERN

Table

how long is the novel?

768 pages

⇒ **Single** information source (KB / text corpora / infoboxes / tables) **not sufficient**

⇒ By **combining heterogeneous sources** the **answer coverage** is **enhanced**

⇒ **Information redundancy** can help improve **accuracy**

Desiderata and Contributions

Leverage **heterogeneous** information sources for conversational QA

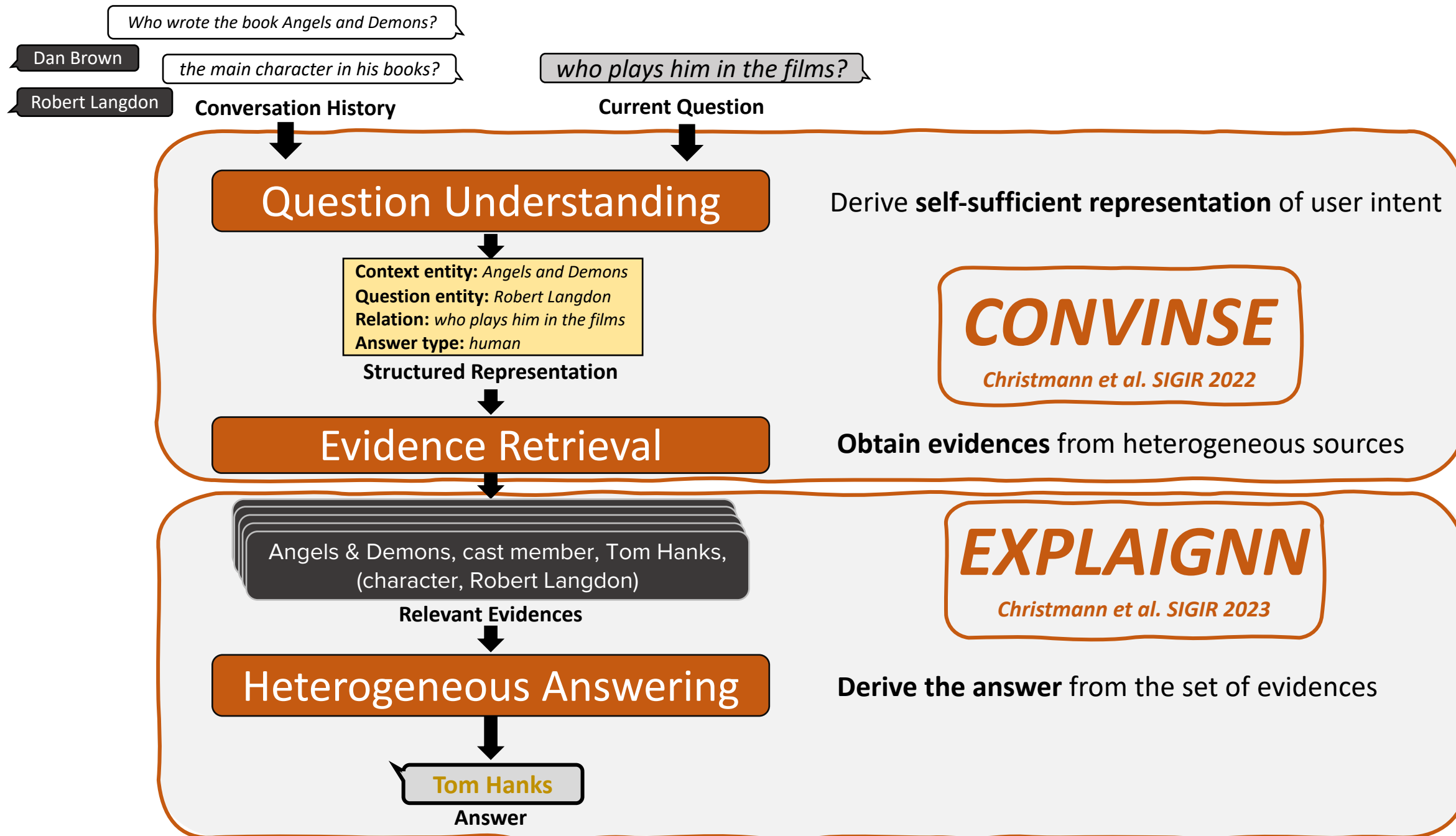**No restriction** to **specific** information **source**

Make process **explainable**
★ **Intermediate** representations in **symbolic** space

Avoid expensive annotations
★ **Distant supervision** from **raw** sequences of **QA pairs**

Overview

Structured Representation (SR)

$\Rightarrow$ Capture information need in a **structured** way

$\Rightarrow$ Relaxed **categorization** into

    $\Rightarrow$ Context entity, question entity, relation, expected answer type

Who wrote the book Angels and Demons?

Dan Brown

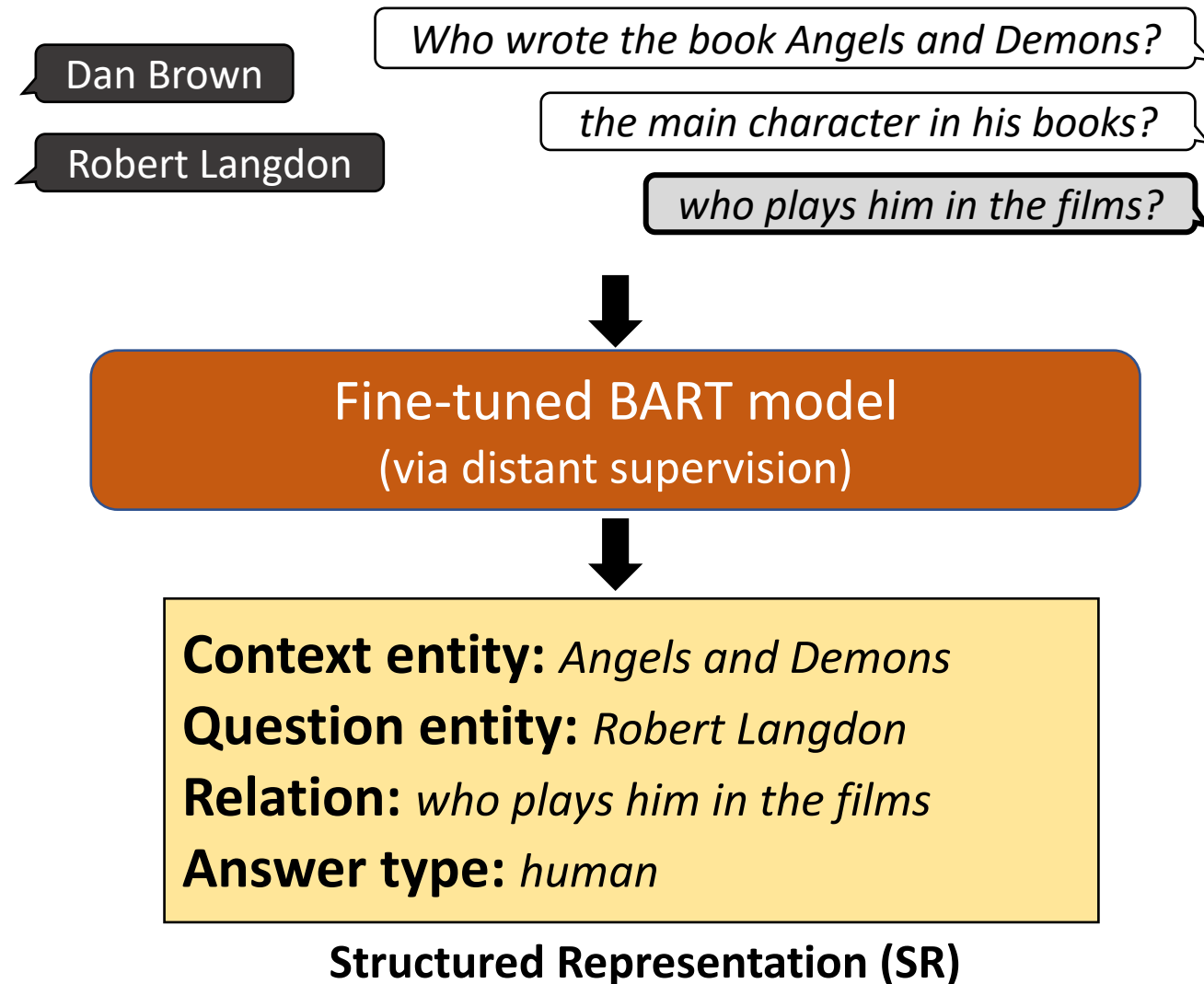the main character in his books?
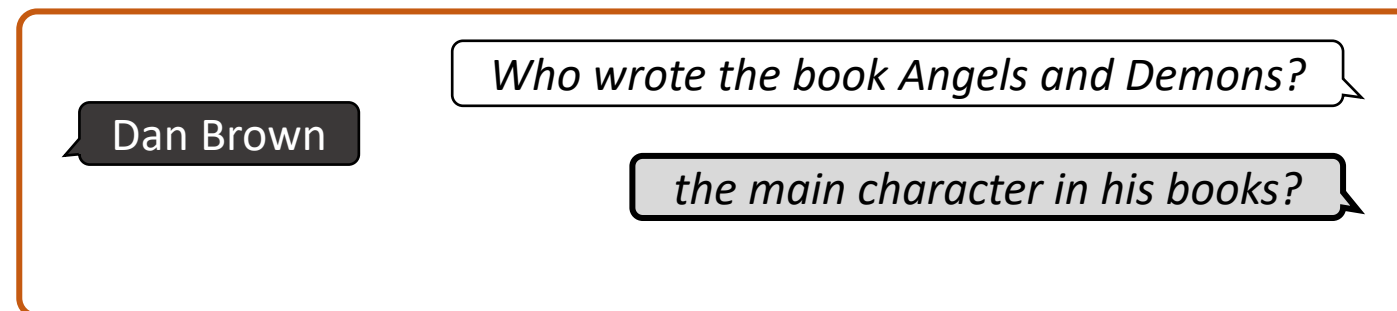
Context entity: *Angels and Demons*
Question entity: *Dan Brown*
Relation: *the main character in his books*
Answer type: *fictional character*

Question Understanding

Evidence Retrieval

Structured Representation (SR)

Context entity: *Angels and Demons*
Question entity: *Robert Langdon*
Relation: *who plays him in the films*
Answer type: *human*

CLOCQ
[Christmann et al. WSDM 2022]

KB-facts

Question entities

Angels & Demons (film series)
Angels & Demons (novel series)
Angels & Demons (soundtrack)
Robert Langdon (film series)
Robert Langdon (character)

Entity-centric Wikipedia retrieval

Text-snippets

Table-records

Infobox-entries

Angels & Demons, cast member, Tom Hanks, (character, Robert Langdon)

Graph Construction

who plays him in the films?

Evidences

KB Angels & Demons, cast member, Tom Hanks, character, Robert Langdon

...

TEXT Professor Robert Langdon is a fictional character created by author Dan Brown for his Robert Langdon book series.

...

INFO Robert Langdon, Portrayed by, Tom Hanks (film series), Ashley Zukerman (The Lost Symbol)

...

Angels & Demons, cast member, Tom Hanks, (character, Robert Langdon)

Plain textual forms...
⇒ **Incorporate structure** using shared entities!

9

Graph Construction

Graph Construction

Evidences

who plays him in the films?

KB Angels & Demons, cast member, Tom Hanks, character, Robert Langdon

"*Angels & Demons*" → Angels & Demons (film),
"*Tom Hanks*" → Tom Hanks (human),
"*Robert Langdon*" → Robert Langdon (fictional character)

...

TEXT Professor Robert Langdon is a fictional character created by author Dan Brown for his Robert Langdon book series.

"*Robert Langdon*" → Robert Langdon (fictional character),
"*Dan Brown*" → Dan Brown (human),
"*Robert Langdon book series*" → Robert Langdon (book series)

...

INFO Robert Langdon, Portrayed by, Tom Hanks (film series), Ashley Zukerman (The Lost Symbol)

"*Robert Langdon*" → Robert Langdon (fictional character),
"*Tom Hanks*" → Tom Hanks (human),
"*Ashley Zukerman*" → Ashley Zukerman (human),
"*The Lost Symbol*" → The Lost Symbol (tv series)

...

Angels & Demons, cast member, Tom Hanks, (character, Robert Langdon)

Professor Robert Langdon

Prof. Robert Langdon portrayed by Tom Hanks in *Angels & Demons*

First appearance   *Angels & Demons*

Last appearance   *Origin*

Created by   Dan Brown

Portrayed by   Tom Hanks (film series)
                             Ashley Zukerman (*The Lost Symbol*)

Leverage href anchors!

Obtain entity mappings for evidences!

Graph Construction

Heterogeneous Graph

who plays him in the films?

⇒ Heterogeneous graphs often have **hundreds** of **nodes**
⇒ Showing such large graphs to **end users** is **impractical**
⇒ **Idea:** Train **GNN** to predict **node relevance**
⇒ **Iteratively** apply to **reduce size** of graph

13

Iterative Graph Reduction

who plays him in the films?
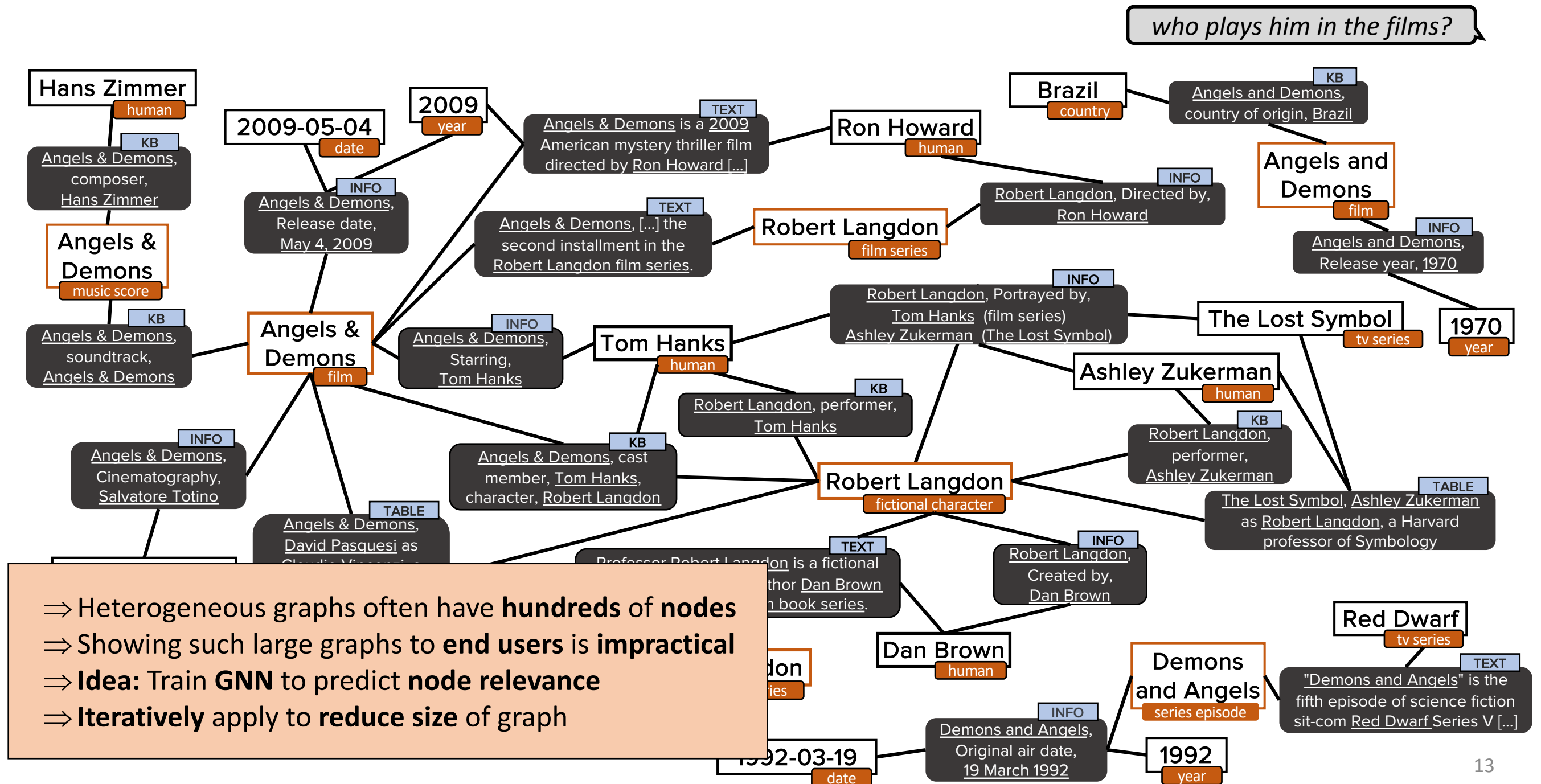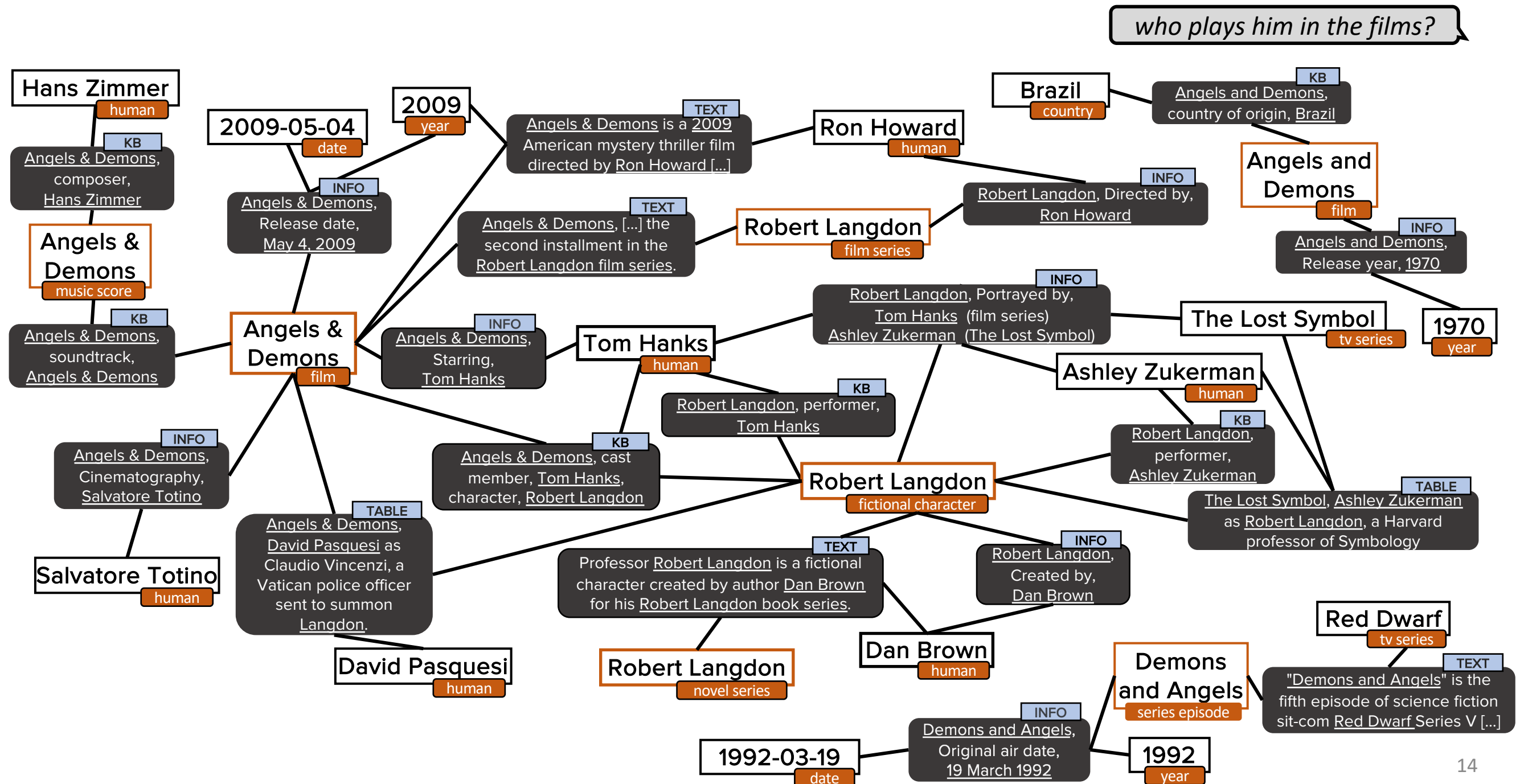
Iterative Graph Reduction

Iterative Graph Reduction
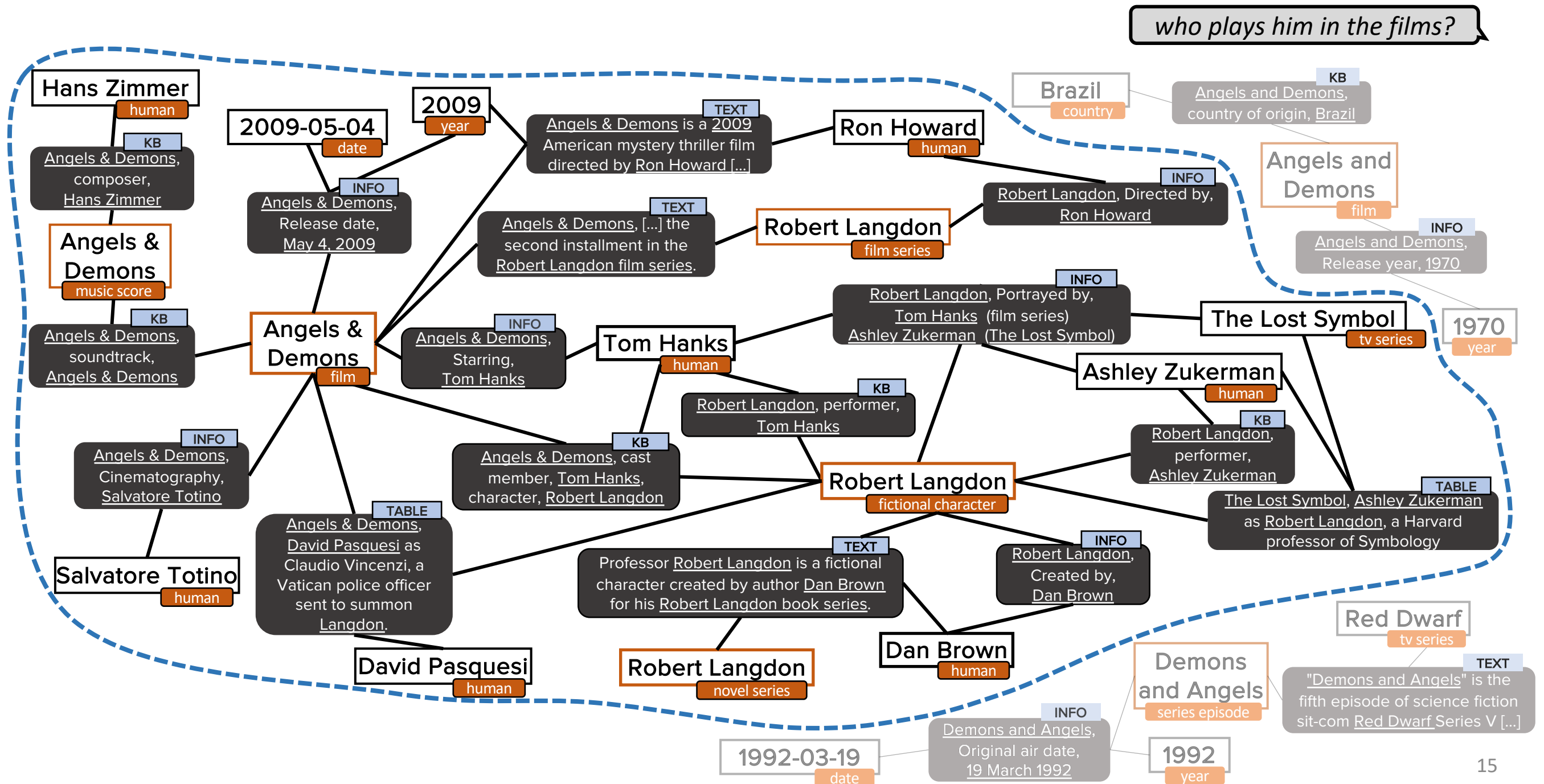
who plays him in the films?
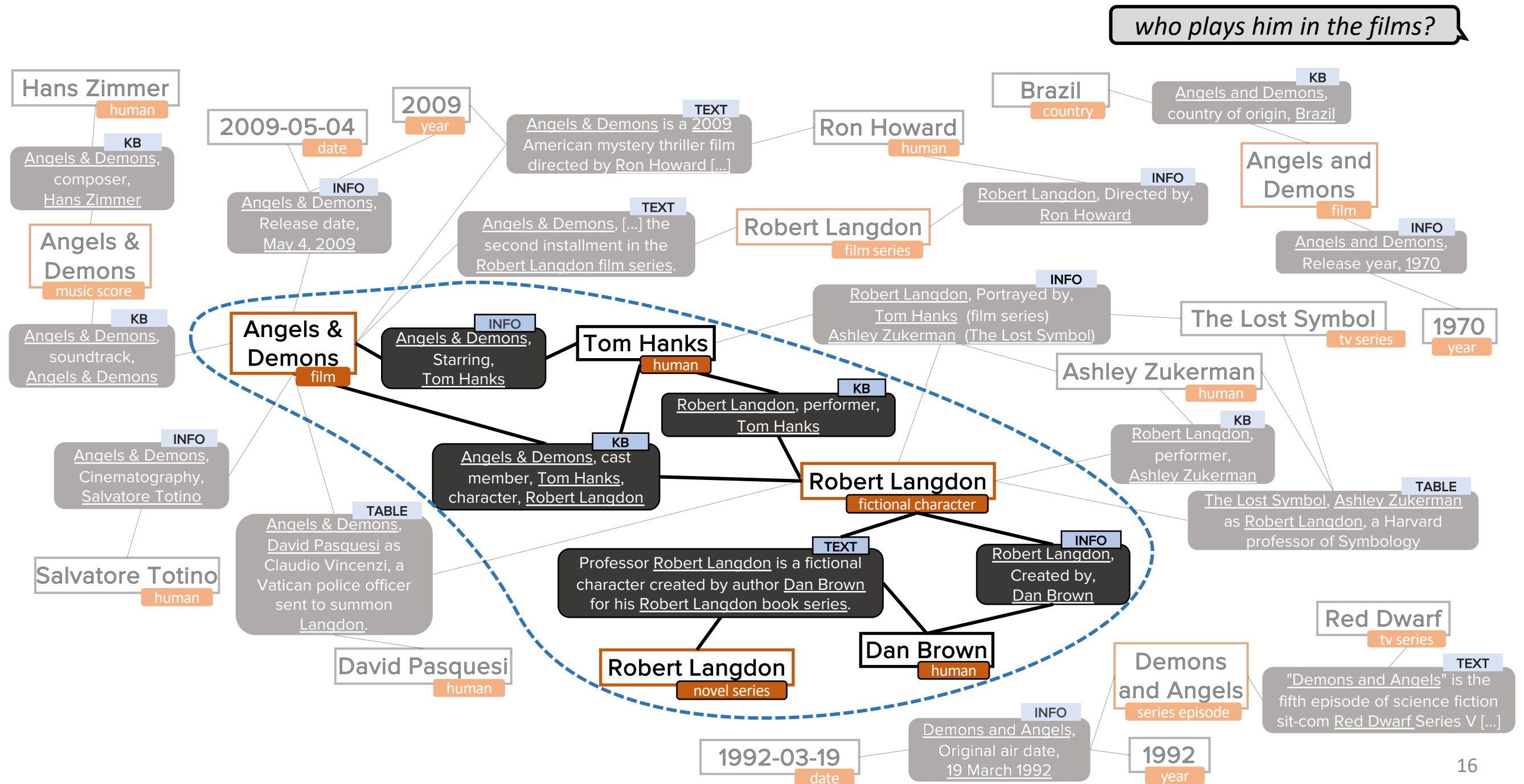
Final Answer Prediction
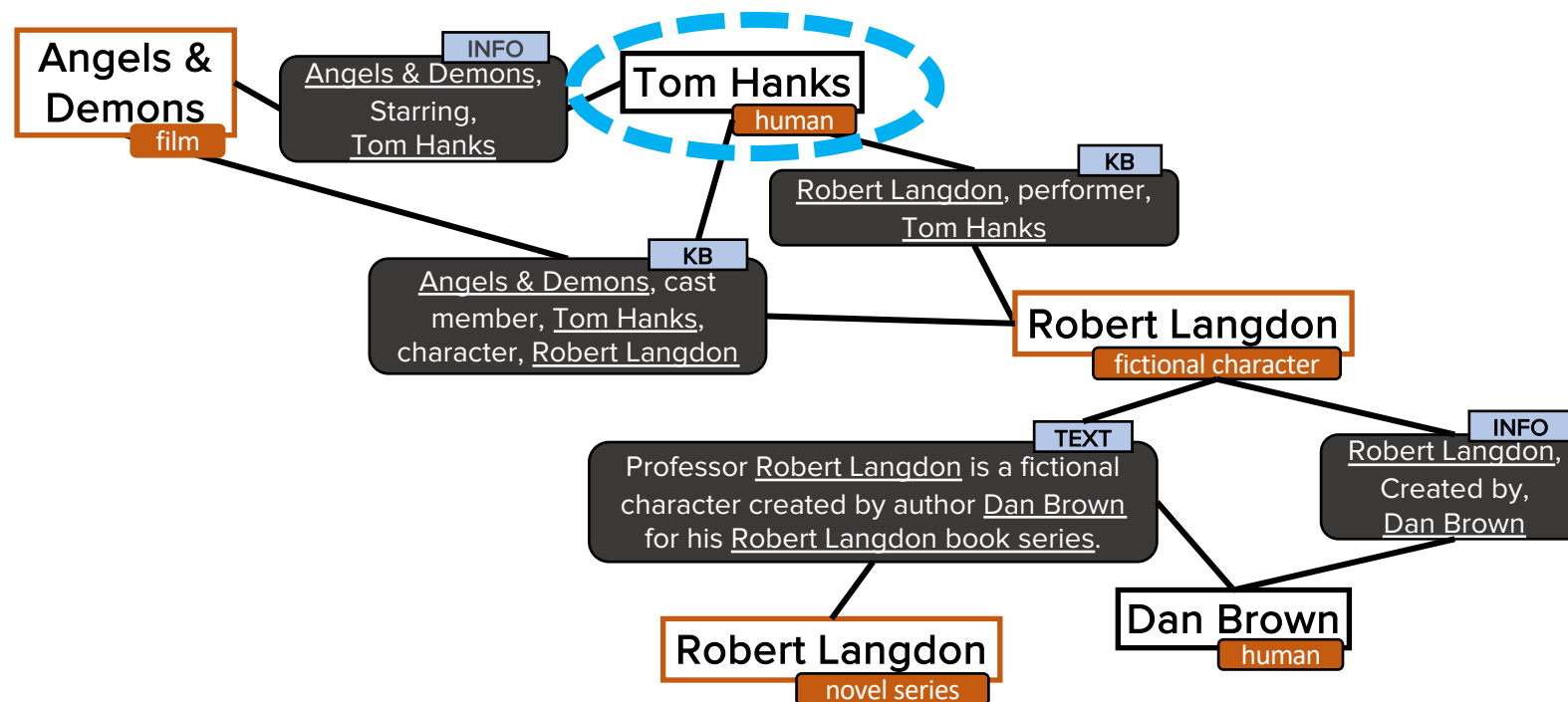
GNN Encoder
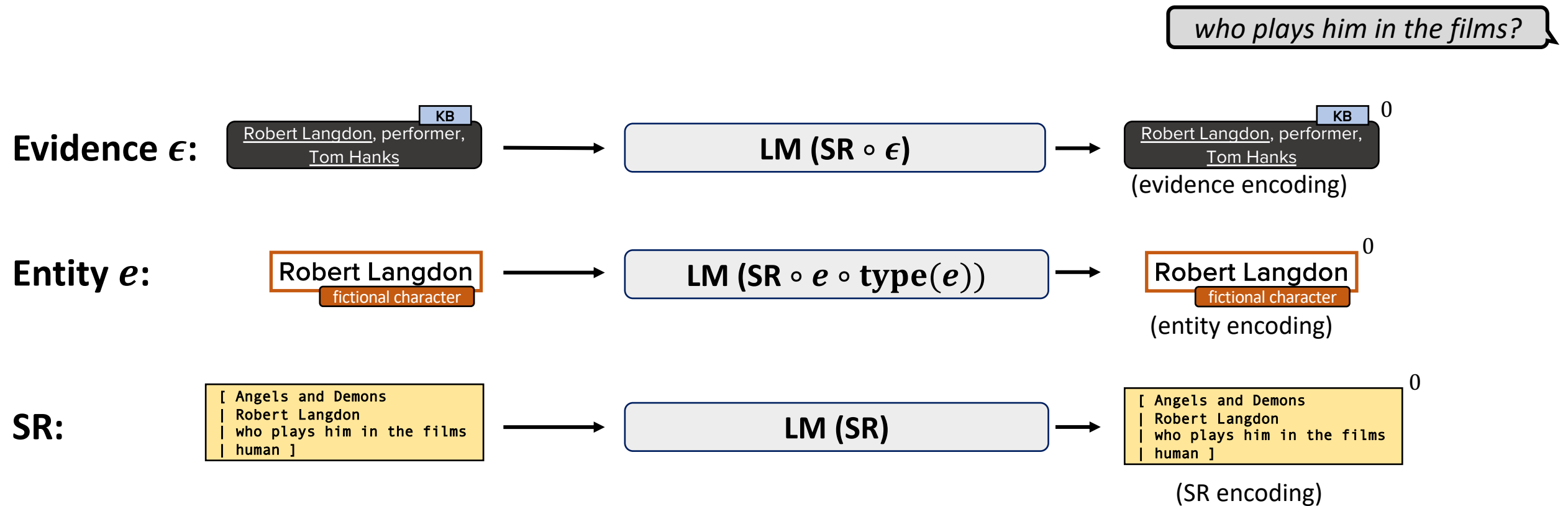
LM: Pre-trained language model (**DistilRoBERTa**) encodings + mean pooling

$\Rightarrow$ Leverage **cross-attention with SR** in encoder

$\Rightarrow$ Obtain **question-relevant encodings**

Message Passing

⇒ **Message passing** in local neighborhoods
⇒ Weight messages by **SR attention**
⇒ Distribute **only question-relevant** information

who plays him in the films?

INFO
Robert Langdon, Portrayed by,
Tom Hanks (film series)
Ashley Zukerman (The Lost Symbol)

KB
Robert Langdon, performer,
Tom Hanks

KB
Angels & Demons, cast
member, Tom Hanks,
character, Robert Langdon

KB
Robert Langdon,
performer,
Ashley Zukerman

Robert Langdon
fictional character

TABLE
The Lost Symbol, Ashley Zukerman
as Robert Langdon, a Harvard
professor of Symbology

TABLE
Angels & Demons,
David Pasquesi as
Claudio Vincenzi, a
Vatican police officer
sent to summon
Langdon.

TEXT
Professor Robert Langdon is a fictional
character created by author Dan Brown
for his Robert Langdon book series.

INFO
Robert Langdon,
Created by,
Dan Brown

19

SR Attention

⇒ **Message passing** in local neighborhoods
⇒ Weight messages by **SR attention**
⇒ Distribute **only question-relevant** information

who plays him in the films?

SR
Context entity: *Angels and Demons*
Question entity: *Robert Langdon*
Relation: *who plays him in the films*
Answer type: *human*

attention (ε, SR)

0.21

INFO
Robert Langdon, Portrayed by,
Tom Hanks (film series)
Ashley Zukerman (The Lost Symbol)

0.18

KB
Robert Langdon, performer,
Tom Hanks

0.12

KB
Robert Langdon,
performer,
Ashley Zukerman

0.24

KB
Angels & Demons, cast
member, Tom Hanks,
character, Robert Langdon

Robert Langdon
fictional character

TABLE
The Lost Symbol, Ashley Zukerman
as Robert Langdon, a Harvard
professor of Symbology

0.11

TABLE
Angels & Demons,
David Pasquesi as
Claudio Vincenzi, a
Vatican police officer
sent to summon
Langdon.

TEXT
Professor Robert Langdon is a fictional
character created by author Dan Brown
for his Robert Langdon book series.

INFO
Robert Langdon,
Created by,
Dan Brown

0.09

0.02

0.03
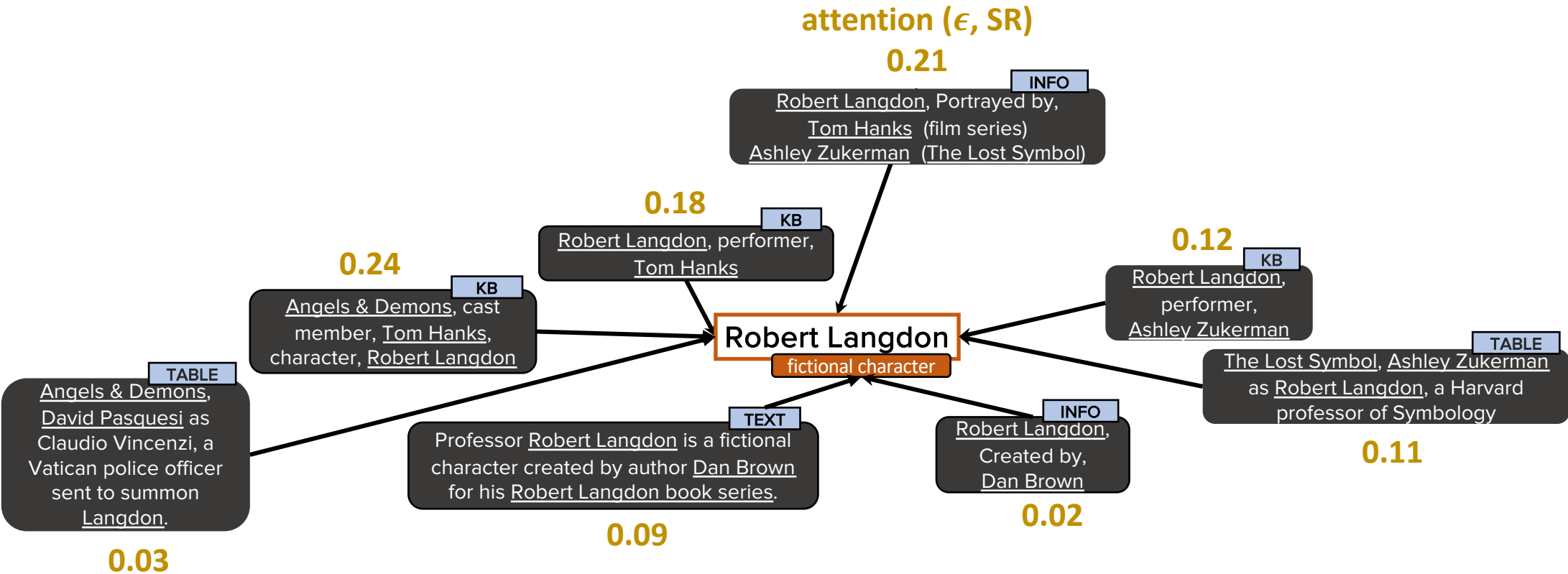
Message Passing
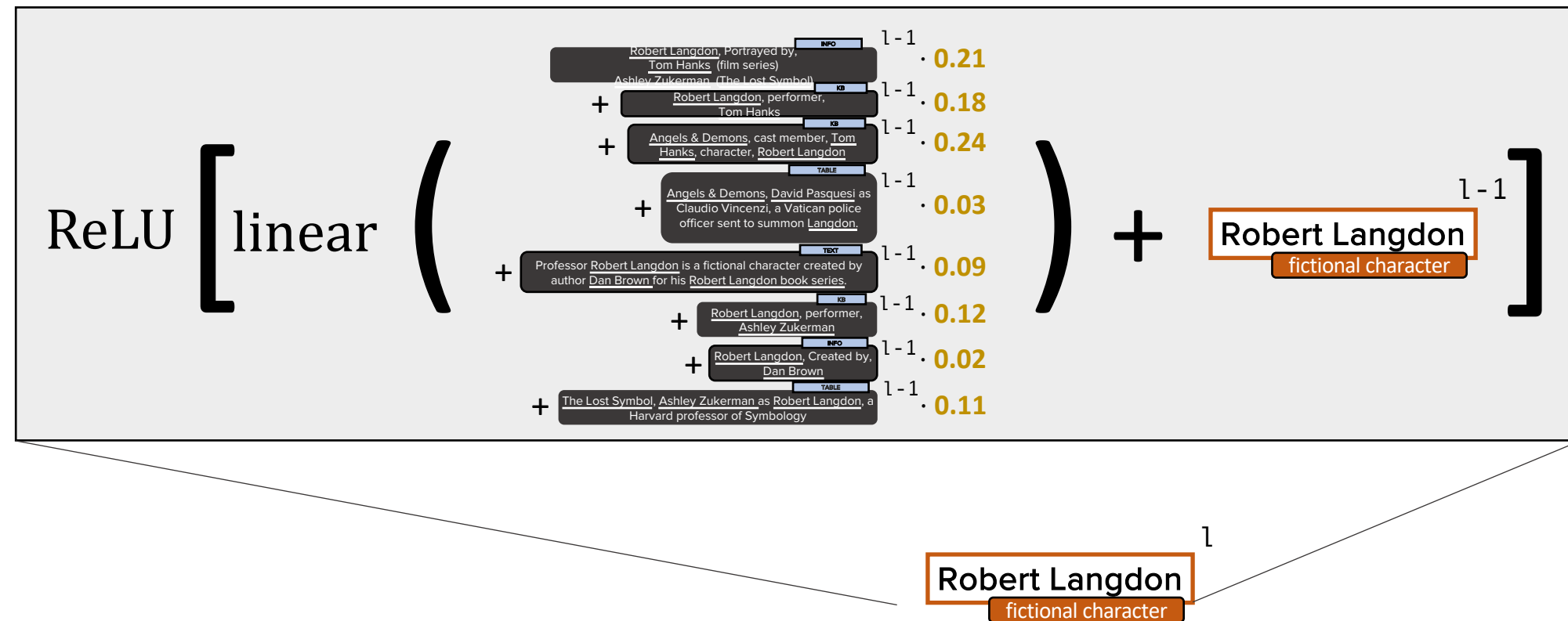
Score Prediction

who plays him in the films?

$$\text{score}\left(\boxed{\begin{array}{c}\text{Robert Langdon}\\ \text{fictional character}\end{array}}\right) = \underset{\text{(all entities)}}{\text{softmax}}\left[\text{linear}\left(\boxed{\begin{array}{c}\text{Robert Langdon}\\ \text{fictional character}\end{array}}^L\right) \cdot \begin{array}{l}[\text{ Angels and Demons}\\ |\text{ Robert Langdon}\\ |\text{ who plays him in the films}\\ |\text{ human }]\end{array}^0\right]$$

$$\text{score}\left(\boxed{\begin{array}{c}\text{Robert Langdon, performer,}\\ \text{Tom Hanks}\end{array}}^{KB}\right) = \underset{\text{(all evidences)}}{\text{softmax}}\left[\text{linear}\left(\boxed{\begin{array}{c}\text{Robert Langdon, performer,}\\ \text{Tom Hanks}\end{array}}^{KB\,L}\right) \cdot \begin{array}{l}[\text{ Angels and Demons}\\ |\text{ Robert Langdon}\\ |\text{ who plays him in the films}\\ |\text{ human }]\end{array}^0\right]$$

$\Rightarrow$ After **L** layers, **compute** entity and evidence **scores**

$\Rightarrow$ **Binary-cross-entropy** loss for both (classification) tasks

$\Rightarrow$ **Entity labels: gold answers** are labeled with **1, other entities** with **0**

$\Rightarrow$ **Evidence labels: evidences** directly **connected** to **gold answers** labeled with **1**

$\Rightarrow$ **Multi-task learning:** loss is **weighted combination**

GNN Scoring

who plays him in the films?

0.05 Hans Zimmer — human

0.03 KB Angels & Demons, composer, Hans Zimmer

0.01 Angels & Demons — music score

0.04 KB Angels & Demons, soundtrack, Angels & Demons

0.02 INFO Angels & Demons, Cinematography, Salvatore Totino

0.01 Salvatore Totino — human

0.01 2009-05-04 — date

0.01 INFO Angels & Demons, Release date, May 4, 2009

0.07 Angels & Demons — film

0.01 2009 — year

0.04

0.07 INFO Angels & Demons, Starring, Tom Hanks

0.06 TABLE Angels & Demons, David Pasquesi as Claudio Vincenzi, a Vatican police officer sent to summon Langdon.

0.03 TEXT Angels & Demons is a 2009 American mystery thriller film directed by Ron Howard [...]

0.05 TEXT Angels & Demons, [...] the second installment in the Robert Langdon film series.

0.02 Ron Howard — human

0.11 Robert Langdon — film series

0.17 Tom Hanks — human

0.12 KB Angels & Demons, cast member, Tom Hanks, character, Robert Langdon

0.10 KB Robert Langdon, performer, Tom Hanks

0.10 Robert Langdon — fictional character

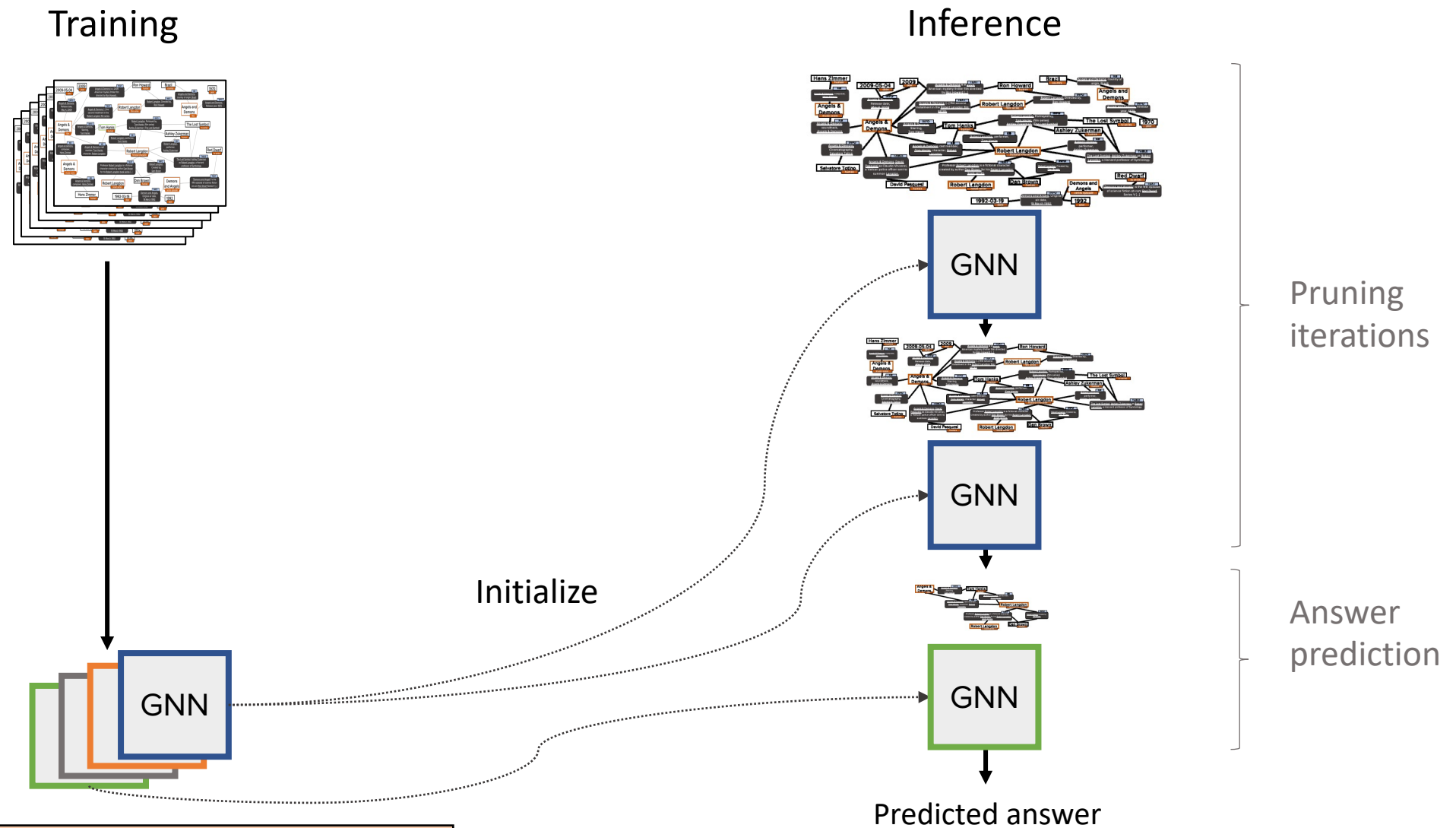**0.07** TEXT Professor Robert Langdon is a fictional character created by author Dan Brown for his Robert Langdon book series.

0.01 Brazil — country

0.01 KB Angels and Demons, country of origin, Brazil

0.04 INFO Robert Langdon, Directed by, Ron Howard

0.05 Angels and Demons — film

0.14 INFO Robert Langdon, Portrayed by, Tom Hanks (film series) Ashley Zukerman (The Lost Symbol)

0.01 INFO Angels and Demons, Release year, 1970

0.08 The Lost Symbol — tv series

0.01 1970 — year **0.01**

0.19 Ashley Zukerman — human

0.10 KB Robert Langdon, performer, Ashley Zukerman

0.03 TABLE The Lost Symbol, Ashley Zukerman as Robert Langdon, a Harvard professor of Symbology

0.04 INFO Robert Langdon, Created by, Dan Brown

Langdon — novel series

0.15 Dan Brown — human

0.03 Demons and Angels — series episode

0.02 Red Dwarf — tv series

0.01 TEXT "Demons and Angels" is the fifth episode of science fiction sit-com Red Dwarf Series V [...] **0.01**

0.01 INFO Demons and Angels, Original air date, 19 March 1992

0.01 1992-03-19 — date

0.01 1992 — year **0.01**

⇒ Obtain **scores** for **every evidence** and **entity**
⇒ Use scores to identify **subgraph** for **next iteration**
⇒ **Iteratively** apply this scheme

23

Iterative Graph Reduction

who plays him in the films?

Training and Iterative Inference

Training

Inference

Pruning iterations

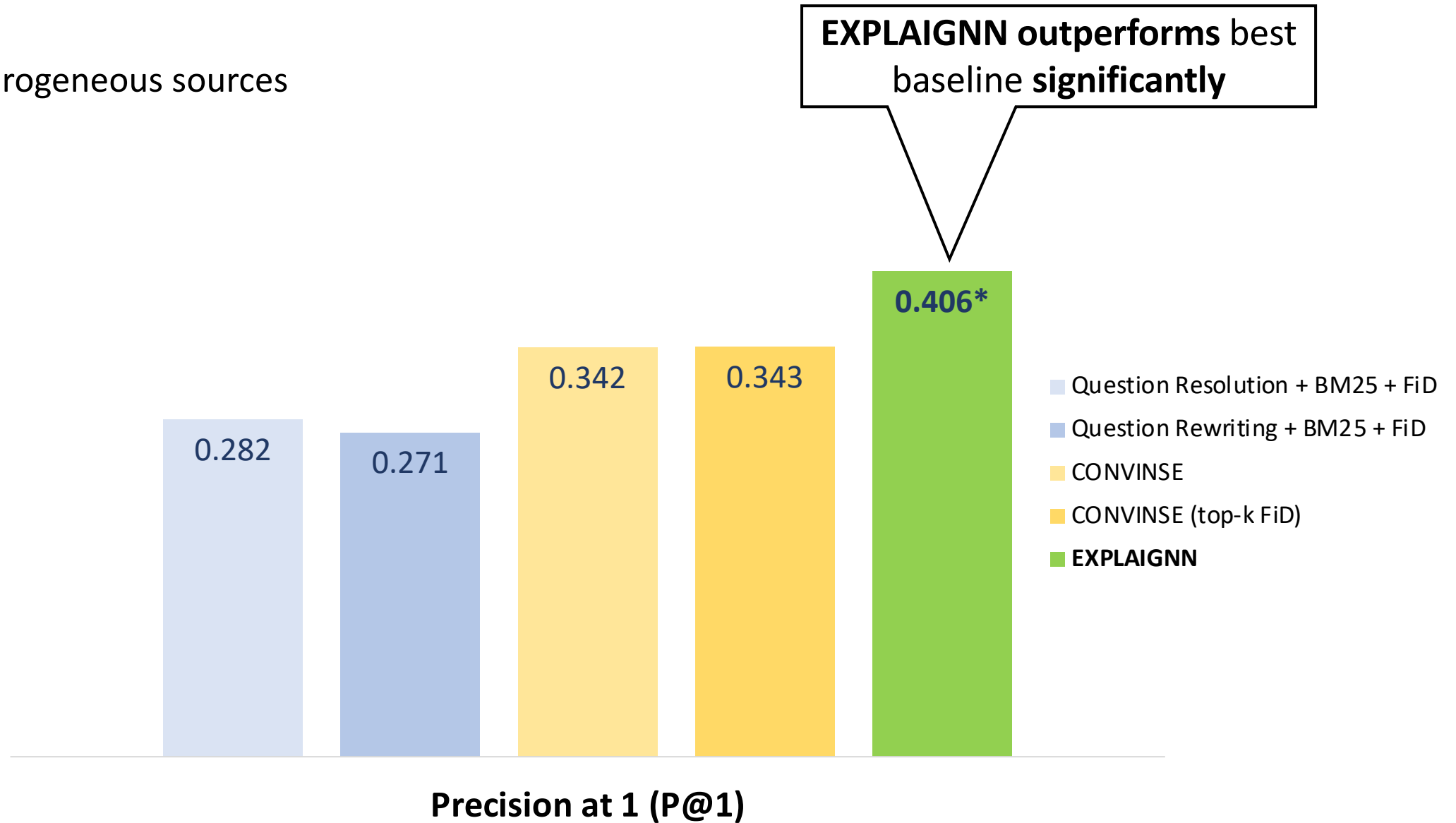Answer prediction

Initialize

GNN

Predicted answer

⇒ **Train** GNNs on **full graphs**
⇒ **Apply iteratively** at **inference** time
⇒ Helps to **improve** performance (+ more robust)
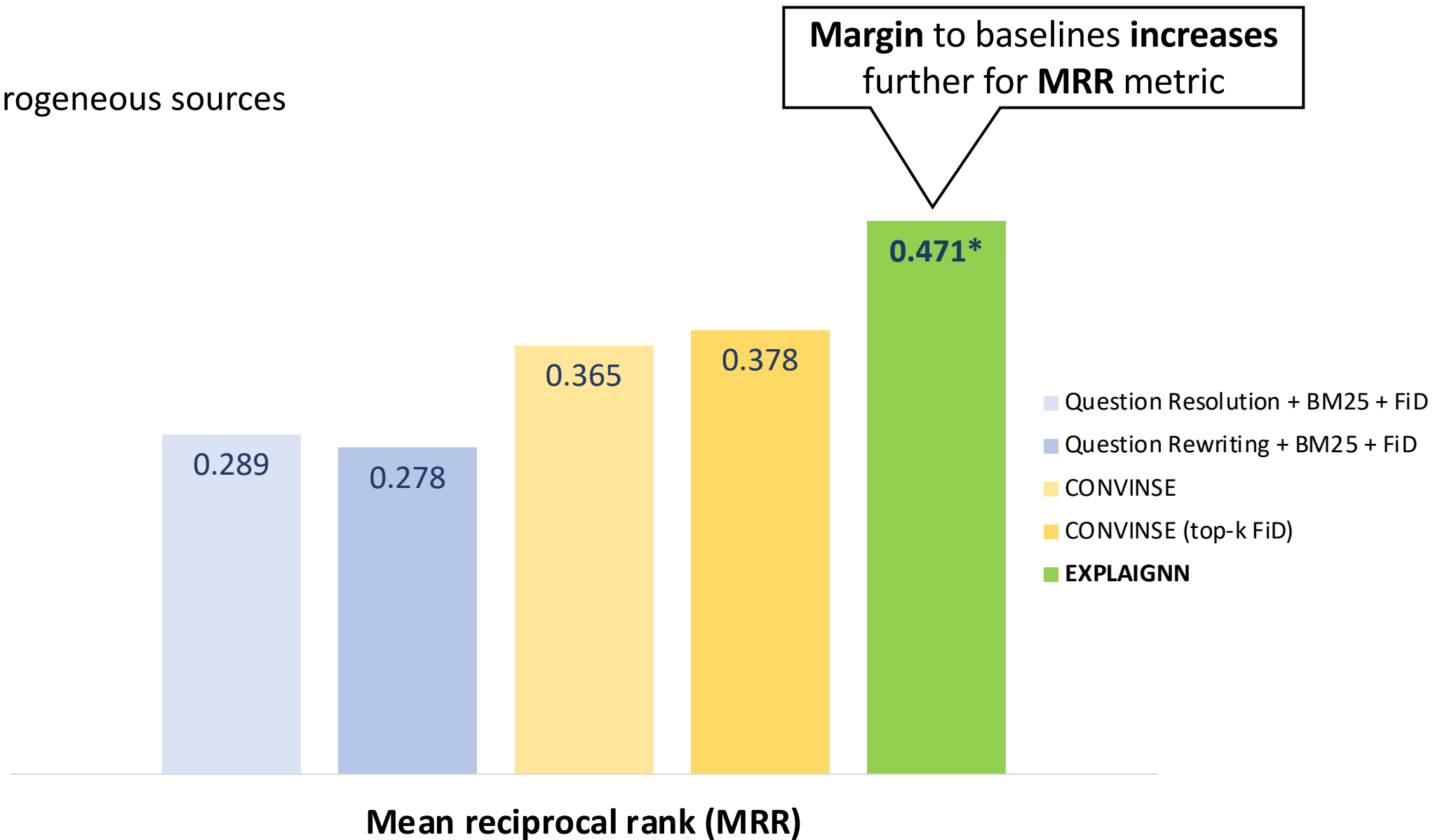
Main Results – P@1

ConvMix dataset:
ConvQA over heterogeneous sources

EXPLAIGNN **outperforms** best
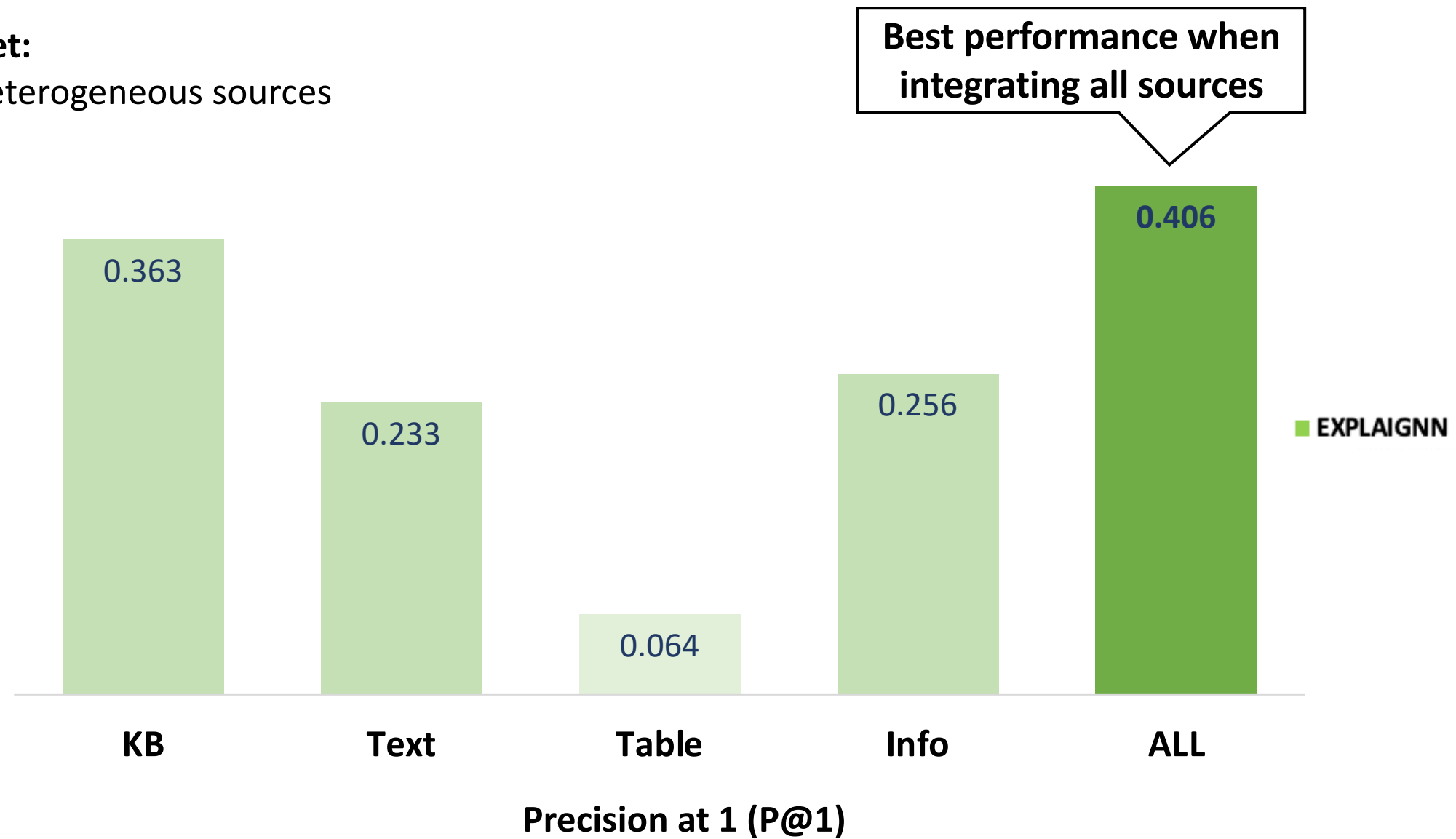baseline **significantly**

0.406*

0.282

0.271

0.342

0.343

- Question Resolution + BM25 + FiD
- Question Rewriting + BM25 + FiD
- CONVINSE
- CONVINSE (top-k FiD)
- **EXPLAIGNN**

Precision at 1 (P@1)

Main Results – MRR

ConvMix dataset:
ConvQA over heterogeneous sources

Margin to baselines **increases** further for **MRR** metric

0.471*

0.365 0.378

0.289
0.278

■ Question Resolution + BM25 + FiD
■ Question Rewriting + BM25 + FiD
■ CONVINSE
■ CONVINSE (top-k FiD)
■ **EXPLAIGNN**

Mean reciprocal rank (MRR)

Results – Per Information Source(s)

ConvMix dataset:
ConvQA over heterogeneous sources

Best performance when integrating all sources

0.363

0.406

0.233

0.256

0.064
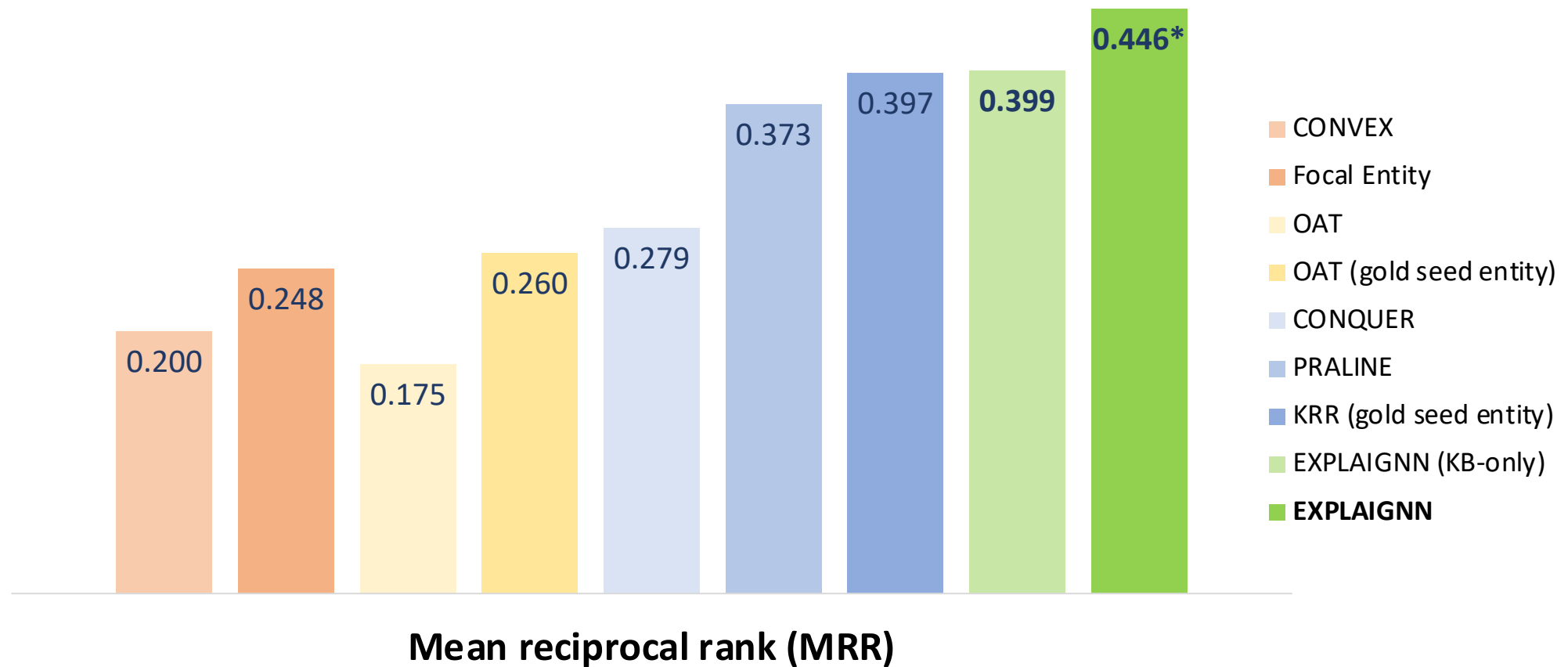
■ EXPLAIGNN

KB   Text   Table   Info   ALL

Precision at 1 (P@1)

Test Robustness

ConvQuestions dataset
ConvQA over KBs

Out-of-the-box EXPLAIGNN obtains **state-of-the-art** performance on competitive **ConvQuestions** dataset

Legend:
- CONVEX
- Focal Entity
- OAT
- OAT (gold seed entity)
- CONQUER
- PRALINE
- KRR (gold seed entity)
- EXPLAIGNN (KB-only)
- **EXPLAIGNN**

Bar values (left to right): 0.200, 0.248, 0.175, 0.260, 0.279, 0.373, 0.397, **0.399**, **0.446***

Mean reciprocal rank (MRR)

"EXPLAIGNN-ability" User Study

Claim: EXPLAIGNN can provide explanations that **end users** can **understand**

"EXPLAIGNN-ation"

Who wrote the book Angels and Demons?

Dan Brown

the main character in his books?

Robert Langdon

who plays him in the films?

Tom Hanks

System interpretation

Context entity: **Angels and Demons**
Current entity: **Robert Langdon**
Relation: who plays him in the films
Expected answer type: human

} **SR**

Supporting evidences

1. **Angels & Demons**, cast member, **Tom Hanks**, character, **Robert Langdon**
2. **Angels & Demons**, Starring, **Tom Hanks**.
3. **Robert Langdon**, performer, **Tom Hanks**.
4. **Robert Langdon**, Created by, Dan Brown.
5. Professor **Robert Langdon** is a fictional character created by author Dan Brown for his **Robert Langdon** book series.

} **Evidences in reduced graph**

"EXPLAIGNN-ability" User Study

Claim: EXPLAIGNN can provide explanations that **end users** can **understand**

$\Rightarrow$ Needs to be shown with end users

Idea:

$\Rightarrow$ **Randomly sample** answered instances for which answer is **correct or incorrect**

$\Rightarrow$ **Present user** with conversational history, answer prediction, and **explanation** (SR + evidences)

$\Rightarrow$ Let user decide **whether predicted answer is correct**

$\Rightarrow$ Ask user about their **certainty** and the **reasons** for their certainty/uncertainty

$\Rightarrow$ **Prune cases** in which user is certain/uncertain for the wrong reasons

    e.g. because of prior knowledge or commonsense

"EXPLAIGNN-ability" User Study

Collected **1,200 judgements via Amazon Mechanical Turk (AMT)**

(Masters only, >95% acceptance rate, honeypot questions)

$\Rightarrow$ **771** left **after pruning** irrelevant cases (answer known,…)

Results

    P (User certain) = 0.798

    P (User correct) = 0.761

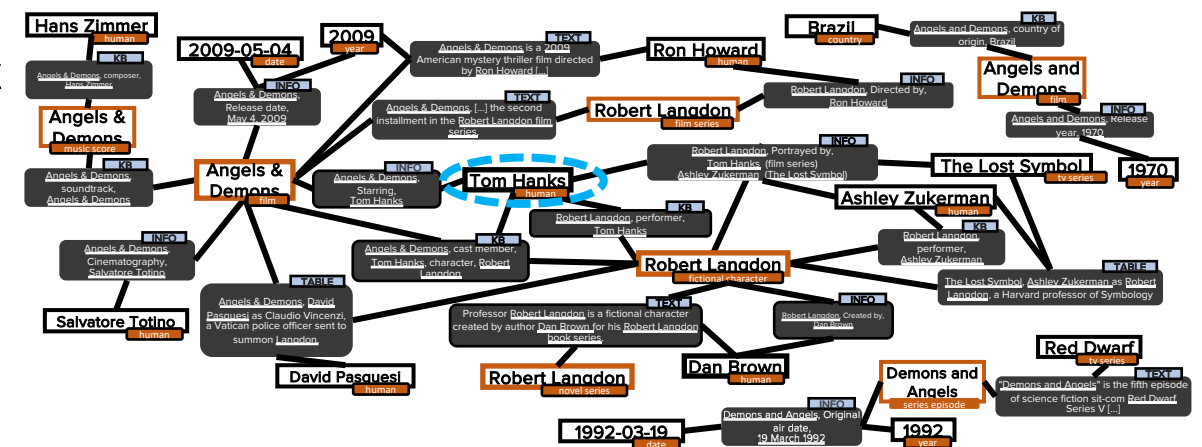   P (User correct | User certain) = 0.792

$\Rightarrow$ Indicates that **explanations** are indeed **comprehensible by end users**

Conclusion

★ We propose **EXPLAIGNN**

    ★ An **explainable pipeline** for ConvQA

    ★ Operates over **heterogeneous sources** (KB / text corpora / infoboxes / tables)

    ★ **Intent-explicit structured representation** of questions

    ★ Large answering **graph** is **iteratively reduced** in size to **obtain answer** and **supporting evidences**

> **Context entity:** *Angels and Demons*
> **Question entity:** *Robert Langdon*
> **Relation:** *who plays him in the films*
> **Answer type:** *human*

★ **EXPLAIGNN** significantly **improves** over baselines on **ConvMix**

★ Crowdsourced **user study** demonstrates **explainability**

★ **Further information:** *explaignn.mpi-inf.mpg.de*

Thank you!