End-to-End Entity extraction, Entity linking, and Relationship extraction for Question Answering Using Knowledge Graphs

Presented by :
Edwin Thomas (300278402)
Saranya Krishnasami (300321456)

CSI 5180 – Virtual Assistants – Winter 2023

Project Summary

Knowledge Graph Question Answering System - KGQA

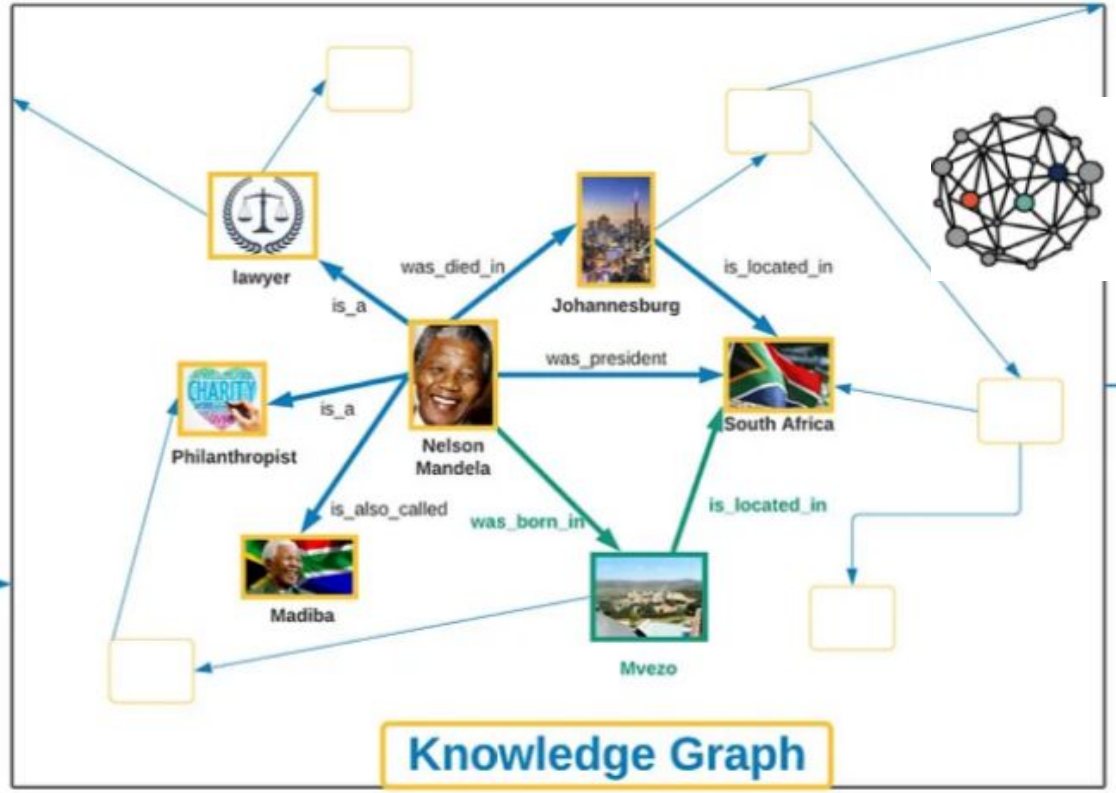
Question: Which town was Nelson Mandela born in South Africa?

Natural Question

Query Parsing

```
SELECT ?town
WHERE { ?town is_located_in "South Africa"
"Nelson Mandela" was_born_in ?town
}
```

SPARQL Query



Answer

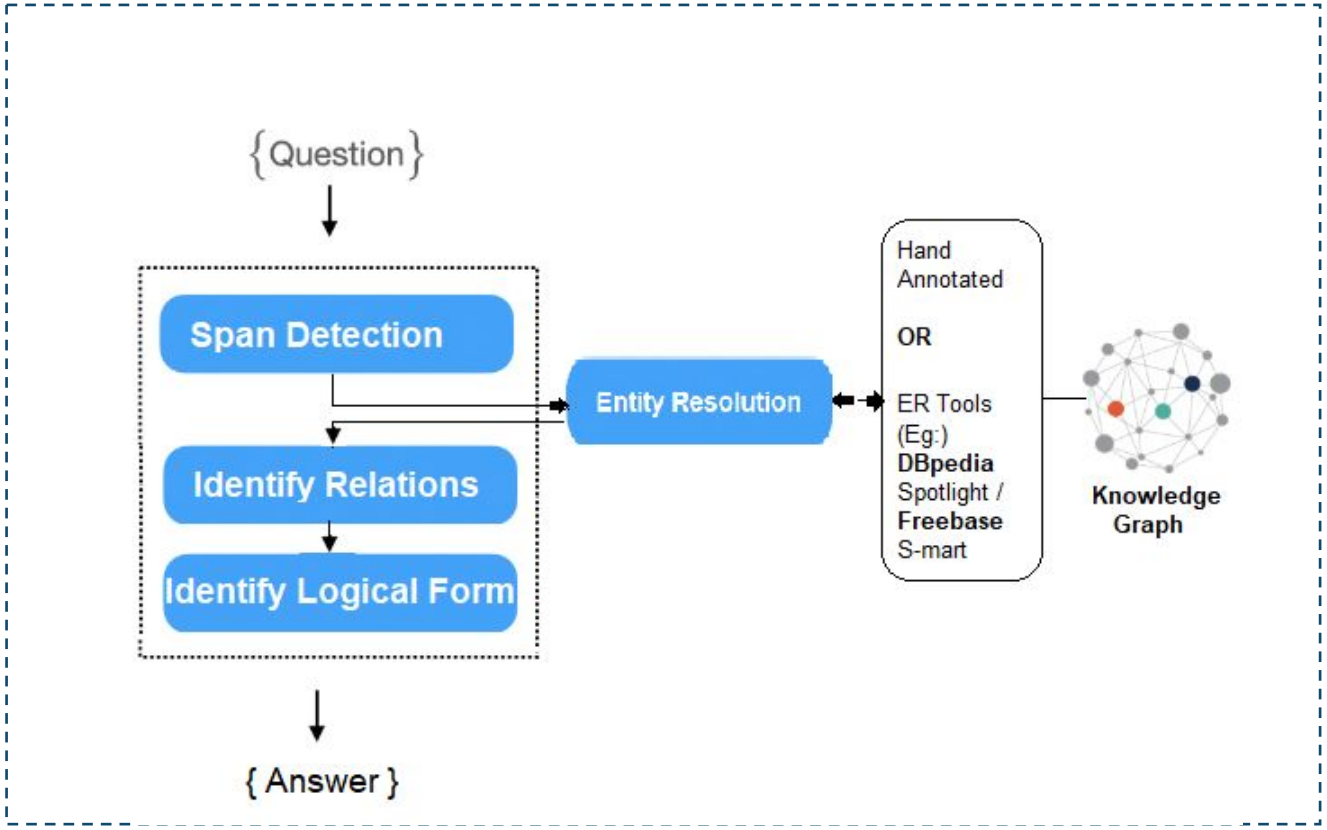
Mvezo

Answer: Nelson Mandela was born in the town of Mvezo

Project Summary

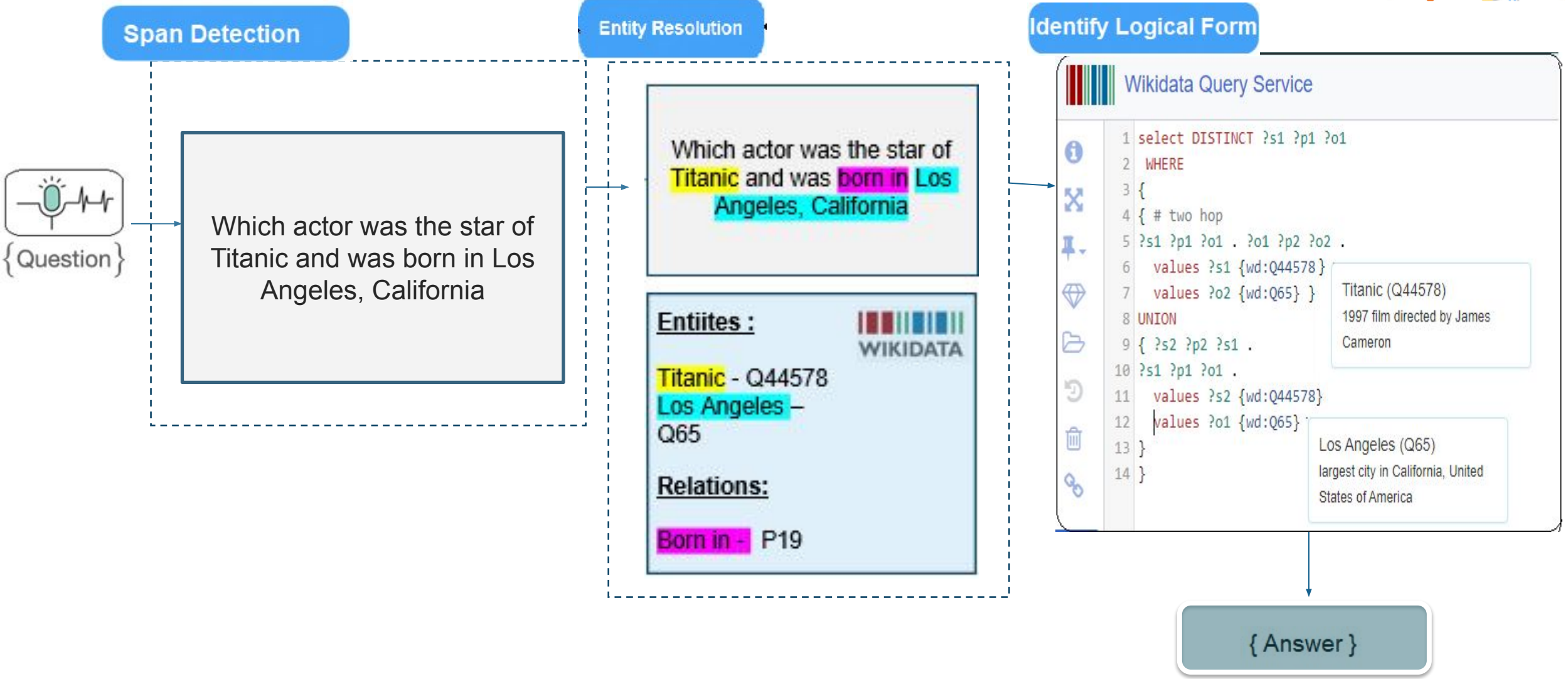
Knowledge Graph Question Answering System - KGQA

Current E2E KGQA System

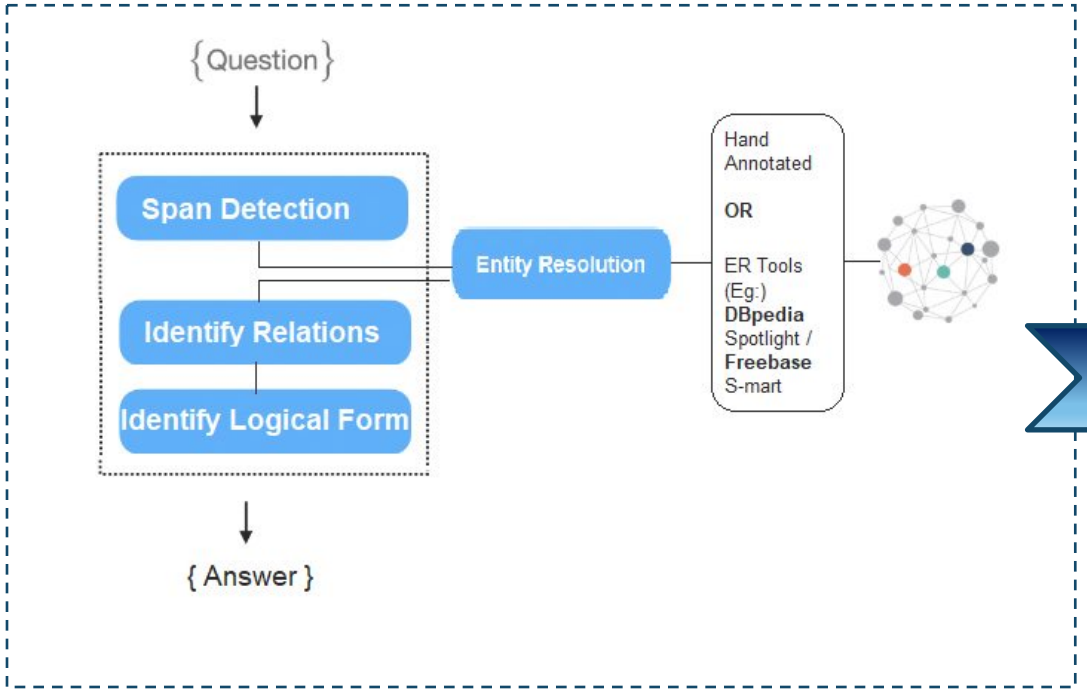


Project Summary

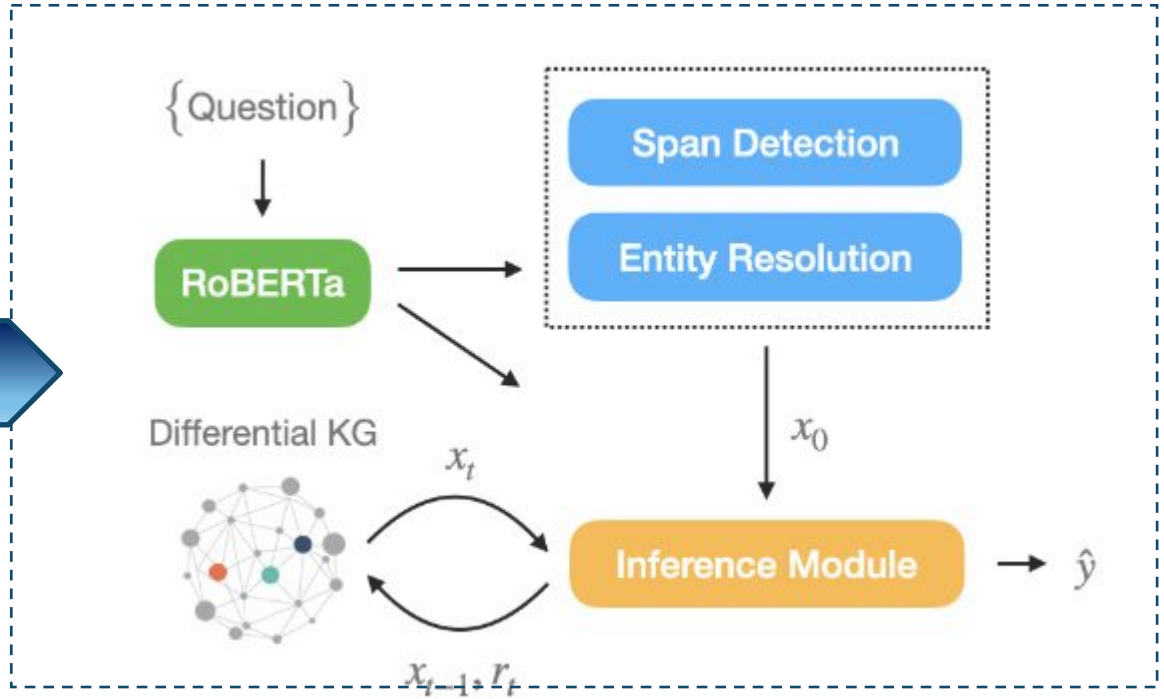
Current End-To-End KGQA Systems



Current E2E KGQA System



Proposed E2E KGQA System



Current E2E KGQA Systems :

1. Loosely coupled components
2. Dependency on external tools for Entity Resolution
3. Challenges to obtaining training datasets for individual components

Proposed E2E KGQA Prototype:

1. Complete standalone End to End Pipeline
2. Train E2E jointly
3. **Question -> Answer**
4. No intermediate annotations

Methodology

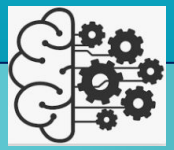
Stage 1

Data Engineering

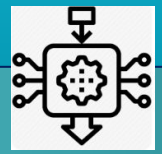
Stage 2

Model Development

Model Training



Model Inference



Methodology – Data Engineering

Stage 1

Data Extraction

WIKIDATA

Wikidata KG

MINTAKA

Mintaka Dataset

Data Cleaning

1. Filter Movies Category
2. Remove Boolean type questions
3. Retain questions that are only of type ‘Entities in Wikidata KG

Data Preparation

Dataset

Training Dev Test

Data Source

Stage 2

Question: Hey Alexa, provide a record of the birthplace of every actor who appeared in the Titanic movie

Answer

Model Check point

Model Development

Model Training

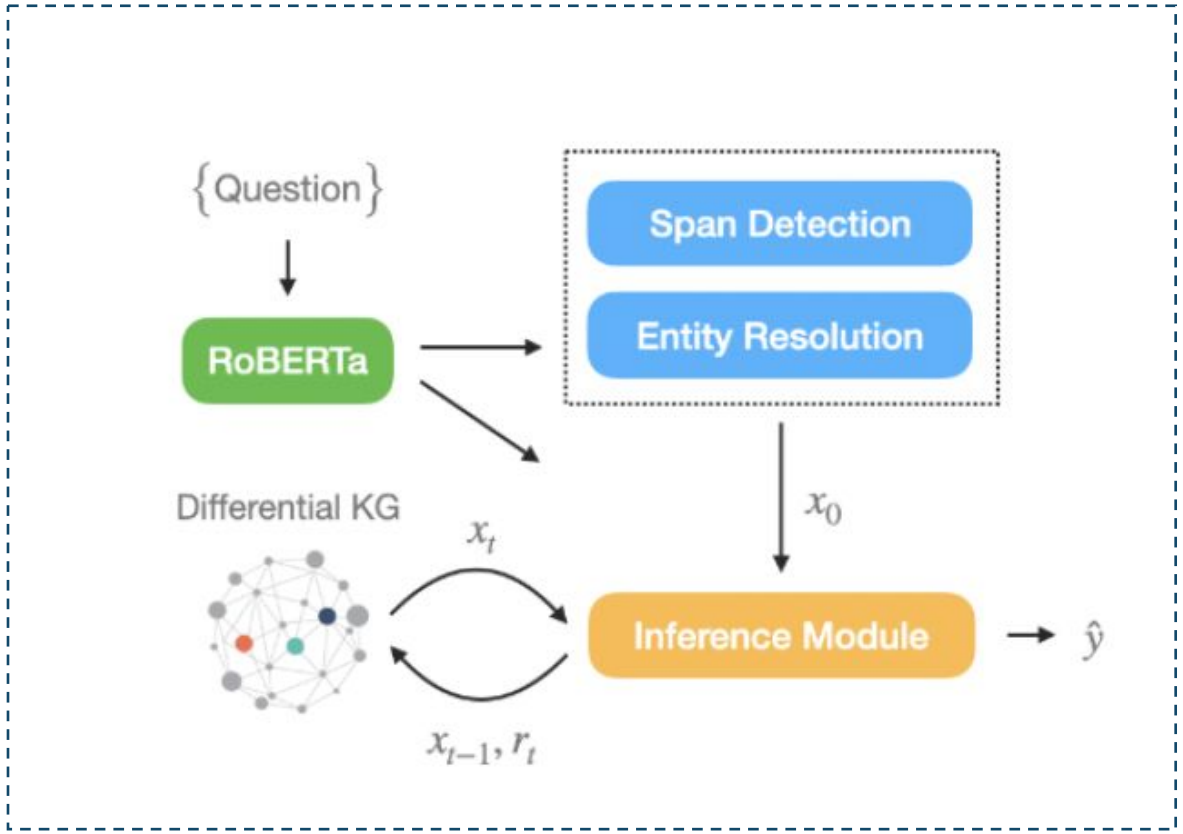
Methodology – Data Engineering

- **Data Engineering**
- Model Training

- Dataset - Mintaka
  - A Complex, Natural, and Multilingual Dataset for End-to-End Question Answering
  - 20000 QA pairs
  - Translated into 8 languages
  - 9 Question complexity types [Count, Comparative, Superlative, Ordinal, Multi-Hop, Intersection, Difference, Yes/No, Generic]

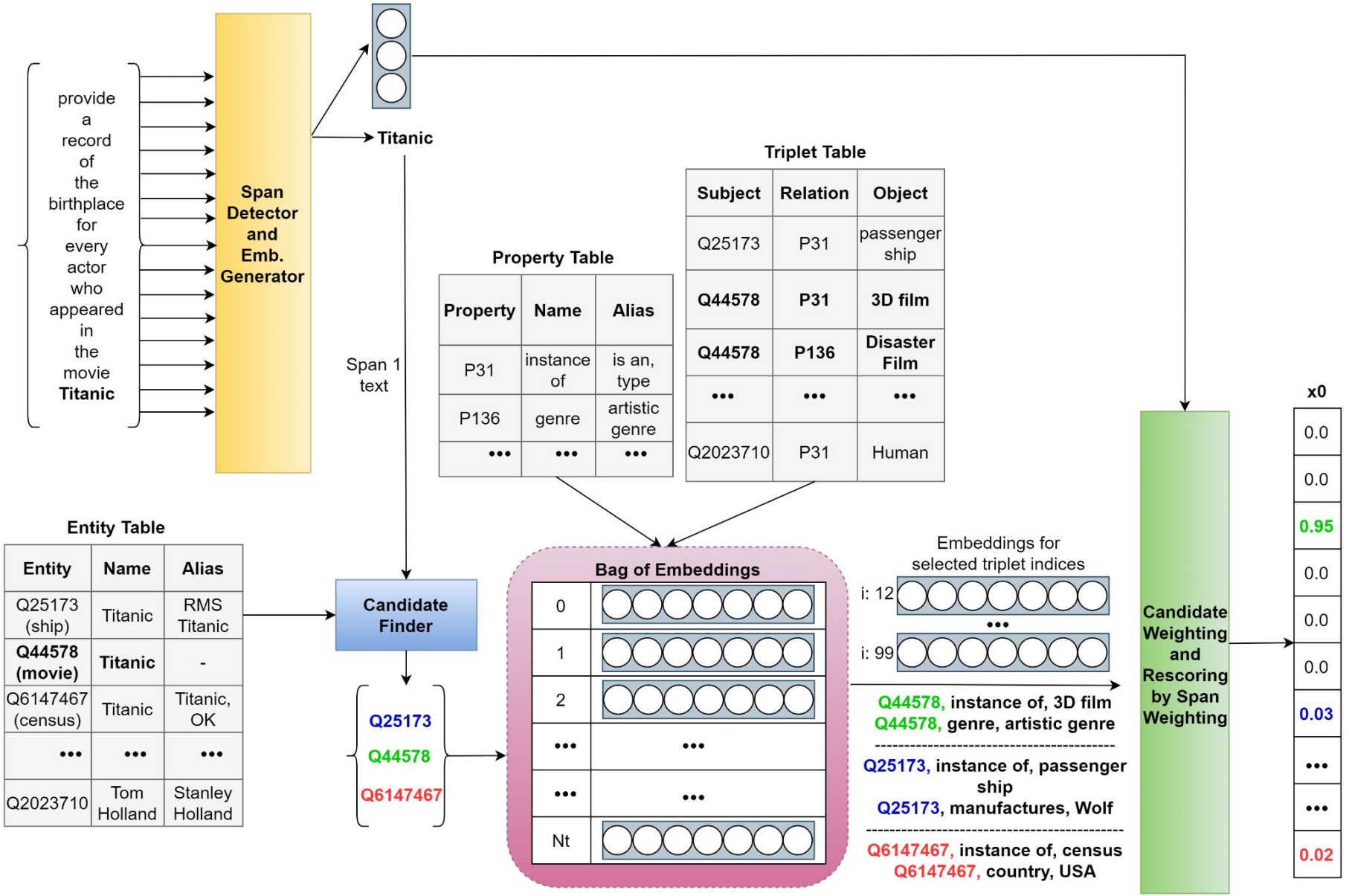
Methodology - Model Development

- Data Engineering
- **Model Development**
- Model Training
- Results
  - Qualitative
  - Quantitative



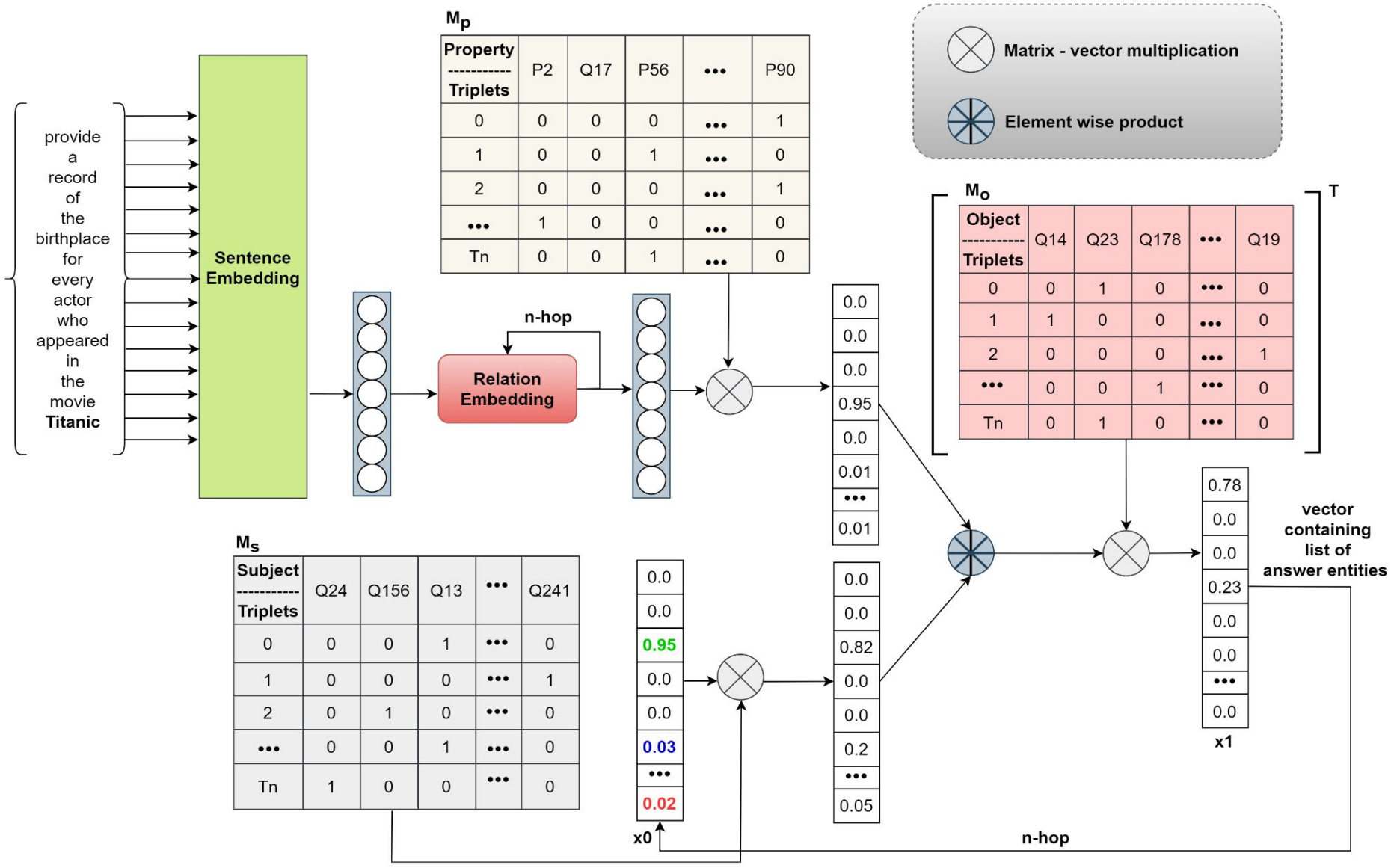
Methodology - Entity Resolution Module

- Data Engineering
- **Model Development**
- Model Training
- Results
  - Qualitative
  - Quantitative



Methodology – Inference Module

- Data Engineering
- **Model Development**
- Model Training
- Results
  - Qualitative
  - Quantitative



M_p

Property Triplets	P2	Q17	P56	•••	P90
0	0	0	0	•••	1
1	0	0	1	•••	0
2	0	0	0	•••	1
•••	1	0	0	•••	0
Tn	0	0	1	•••	0

M_o

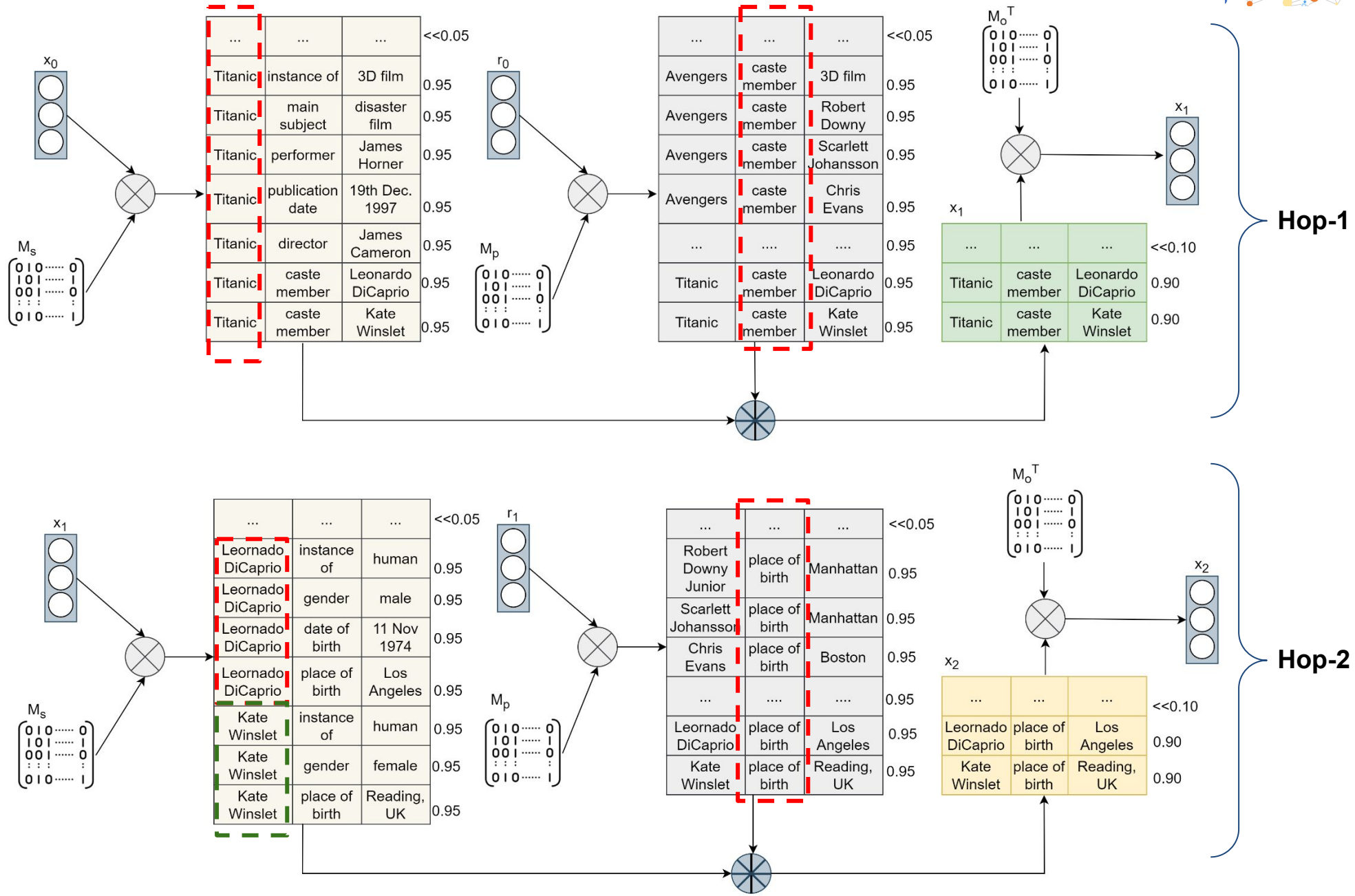
Object Triplets	Q14	Q23	Q178	•••	Q19
0	0	1	0	•••	0
1	1	0	0	•••	0
2	0	0	0	•••	1
•••	0	0	1	•••	0
Tn	0	1	0	•••	0

M_s

Subject Triplets	Q24	Q156	Q13	•••	Q241
0	0	0	1	•••	0
1	0	0	0	•••	1
2	0	1	0	•••	0
•••	0	0	1	•••	0
Tn	1	0	0	•••	0

Methodology – Inference Module

- Data Engineering
- **Model Development**
- Model Training
- Results
  - Qualitative
  - Quantitative



Methodology – Model Training

- Data Engineering
- Model Development
- **Model Training**
- Results
  - Qualitative
  - Quantitative

- Adam Optimizer (lr: 1e-03)
- Exponential Learning Rate Scheduler (gamma: 0.96)
- Loss Function (Binary Cross Entropy - Multi-label support)
- Epochs: 8 (Save on best Validation micro-f1 score)
- Maximum Candidates: 3
- Maximum Spans: 3
- Maximum Properties: 5
- Maximum Hops: 3
- Batch size: 4
- Embedding size (sentence and BOE): 768
- Span Extraction Model: [dslim/bert-base-NER]
- Sentence Embedding Model: [sentence-transformers/all-distilroberta-v1]

```
Epoch: 0, Loss: 0.003942910116165876, LR: 0.0096
100%|                                                                                | 46/46 [00:38<00:00,  1.20it/s]

100%|                                                                                | 8/8 [00:05<00:00,  1.35it/s]
Train f1:  0.1520190023752969
Validation f1:  0.15873015873015872

Epoch: 1, Loss: 0.002761147217825055, LR: 0.009215999999999998
100%|                                                                                | 46/46 [00:38<00:00,  1.20it/s]

100%|                                                                                | 8/8 [00:05<00:00,  1.35it/s]
Train f1:  0.2956521739130435
Validation f1:  0.1875

Epoch: 2, Loss: 0.00221159216016531, LR: 0.008847359999999999
100%|                                                                                | 46/46 [00:39<00:00,  1.17it/s]

100%|                                                                                | 8/8 [00:06<00:00,  1.29it/s]
Train f1:  0.35244161358811044
Validation f1:  0.28169014084507044
```

Results – Qualitative Analysis

- Data Engineering
- Model Development
- Model Training
- **Results**
  - **Qualitative**
  - Quantitative

Type I: Paraphrases and Synonymy

Qn: Who were the *actors* that appeared in the Batman movies?
Ans ✅: ['Christian Bale', 'Robert Pattinson', 'Michael Keaton']

Qn: Can you provide a roster of some *performers* who starred in the Batman movies?
Ans ✅: ['Christian Bale', 'Robert Pattinson', 'Michael Keaton']

Type II: Multihop follow operations

Qn: Where is the director of Rockstar from?
Ans ✅: [India]
(Rockstart → director → Imtiaz Ali → country of citizenship → India)

Results – Qualitative Analysis

- Data Engineering
- Model Development
- Model Training
- **Results**
  - **Qualitative**
  - Quantitative

Type III: Equality and Comparison (Multi-hop)

Qn: Which movie had a *bigger budget*, Lord of The Rings or Ghostbusters?

Ans ❌: ['oos']

Type IV: Min-Max operations (Multi-hop)

Qn: Which Studio Ghibli movie scored the *highest* on Metacritic?

Ans ❌: ['Spirited Away', 'Grave of the Fireflies', 'Castle in the Sky']

(Studio Ghibli→ notable work→ Spirited Away→metascore→96)
(Studio Ghibli→ notable work→ Grave of the Fireflies→metascore→94)
(Studio Ghibli→ notable work→ Castle in the Sky→metascore→78)

Results – Quantitative Analysis

- Data Engineering
- Model Development
- Model Training
- **Results**
  - Qualitative
  - **Quantitative**

Our Results (Hits @ 1)

Split	Precision (%)	Recall (%)	F1 (%)
Train	82.77	39.83	53.79
Dev	77.27	29.31	42.50
Test	65.21	16.30	26.09

6 Baselines

Model	Hits@1
LANGUAGE MODELS	
T5	0.28
T5 for CBQA (zero-shot)	0.20
T5 for CBQA (fine-tuned)	**0.38**
KGQA MODELS	
KVMemNet	0.12
EmbedKGQA	0.18
Rigel	0.20
RETRIEVER-READER MODELS	
DPR (zero-shot)	0.15
DPR (trained)	0.31

Results of RIGEL published in [Mintaka Dataset Reference Paper]

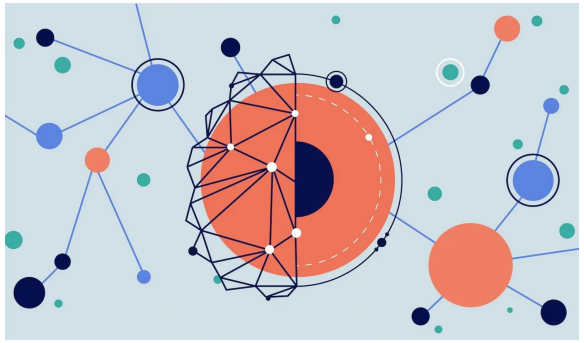
label	precision	recall	f1
brad pitt	1	1	1
tom holland	1	1	1
johnny depp	1	1	1
mark wahlberg	1	1	1
avengers: age of ultron	1	1	1
back to the future part ii	1	1	1
the godfather	1	1	1
ian mckellen	1	1	1
lady gaga	1	1	1
ben affleck	1	1	1
rachel weisz	1	1	1
tim allen	1	1	1
kirk douglas	1	1	1
antonio banderas	1	1	1
jada pinkett smith	1	1	1
the avengers	0.5	1	0.66666667
micro avg	0.772727273	0.29310345	0.425
weighted avg	0.284482759	0.29310345	0.28735632

label	precision	recall	f1
tom hanks	1	1	1
harrison ford	1	1	1
clint eastwood	1	1	1
richard attenborough	1	1	1
thx 1138	1	1	1
roy scheider	1	1	1
jude law	1	1	1
the hurt locker	1	1	1
reese witherspoon	1	1	1
amal clooney	1	1	1
paul walker	1	1	1
glenn close	1	1	1
avengers: age of ultron	0.5	1	0.666666667
the avengers	0.5	1	0.666666667
avengers: endgame	0.5	0.25	0.333333333
micro avg	0.652173913	0.163043478	0.260869565

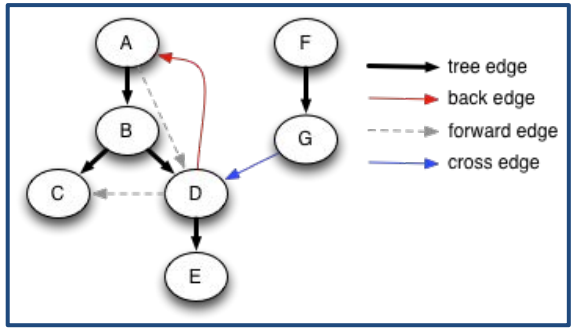
Class-wise results for *few selected entity types* from Movies Domain.

Future Work

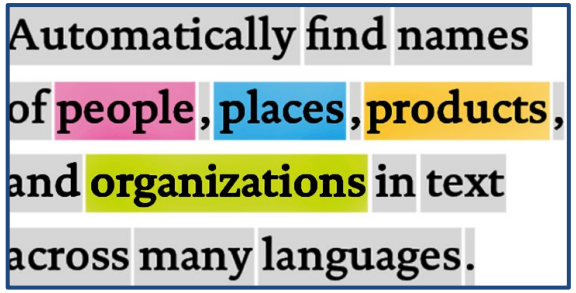
1. **Differential KG methodology that supports complex operations**
    1. Model does not support intersection, union, comparison, equality, min, max operations (as it can't be obtained using Reified KB follow operations).



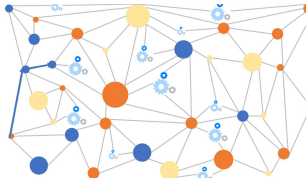
2. **Richer KB representations (Data Augmentations)**
    1. *Backward edges* to combat sparse connections.
    2. Eg: Which actress starred in Black Widow and studied in USC (University of Southern California?)
        1. It is possible to answer this query if a backward edge from USC to "Scarlett Johansson" existed in Wikidata.



3. **Better Span Detectors**
    1. As we expand to more domains, NER models will not be able to detect all entity types.

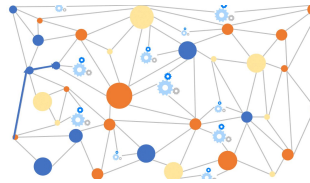


Learning



- Entity resolution process and implement it to produce a rapid working prototype.
- Build a customized dataset and pipeline
- Effectively manipulate tensors, build complex pipelines, implement DL models
- Build an E2E system using the PyTorch framework

Activity Table



Activity	Purpose	Estimated Effort	Deliverable	Team member
Literature Review	Read and understand existing approaches to KGQA such as Classification based, Neural Network ranking based and Translation based approaches.	2 Hours	Carefully understand the use-cases and limitations of these approaches and share with the team. Decide upon an approach to implement.	Edwin
E2E Model Prototype implementation	Using the research paper as reference Implement the Entity Extraction and Resolution Module using PyTorch	8.5 Hours	Torch Modules that can be invoked by the Inference Module	Edwin
	Implement the Inference Module. Based on the local KB collected through SPARQL queries, create sparse entity, relation and triplet matrices. Write PyTorch modules to replicate the Differential Knowledge Graph based traversal, relationship and answer extraction.	9.5 Hours	Torch Modules that are inference in the E2E pipeline.	Edwin
	Create an E2E Training and Inference Pipeline.	2 Hours	A pipeline that integrates the above two modules, trains the model and performs inferencing.	Edwin
Unit/System Testing	Test the developed modules.	1 hour	Tested and verified modules before integration	Edwin
Documentation	Add relevant implementation details to the final report	2 hour	Design and Implementation modules in the documentation	Edwin
Team meetings	Meetings to discuss project plan, work split, integration	-	To-Do lists, meeting minutes	Edwin

Activity	Purpose	Estimated	Deliverable	Team member
Literature Review	To find the key purpose of the article, current works, related works, gaps in the system, understanding of the method to be prototyped	3 Hours	Article notes documents and stored in Zotero app	Saranya
Background survey of dataset and Preparation of Movies dataset from Master dataset	Survey to check the different datasets and finalize the one to support our prototype. Validate the dataset and study of the distribution of data. Prepare Train, Validation and Test dataset on movies domain	4 Hours	Documentation of survey study	Saranya
Prepare SPARQL queries to fetch the entity id's, predicate id's, triples from wikidata KG	To get the list of entities, relationships and objects from the Knowledge graph	3 hours	List of all entities with their alias, predicate with their names in csv format	Saranya
Extract entities id - QID , PID , <S,P,O> from WikiData for model training	To get the id's of all movie entities ,triples and predicates	4Hours	csv of the entities Qid , List of predicate (Pid) in csv format , Fetch the triples <S,P,O> for the entities in csv format	Saranya
Integration of the error_analysis pipeline to the E2E system	To implement parts of pipeline and integrate	3 Hours	Error log, Rectified pipeline	
Unit testing/ System / integration testing	Unit and integration testing	4 Hours	Tested modules	Saranya
Documentation	Process documents, flowcharts, test cases	3 Hours	Project definition report, Final project report, testcase documents	Saranya

Conclusion

- Successfully implemented a working prototype of the State of art E2E (End to end) Knowledge graph Question answering comprising of Entity extraction, Entity Resolution, and relationship mapping

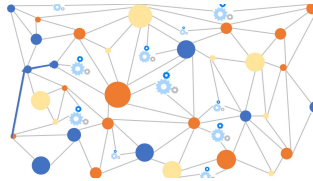
- Develop a paradigm that is currently only accessible to Amazon Alexa teams

- Contribution to the research community by a pipeline of functioning E2E Entity resolution model

•For Model coding details, kindly refer to the GitHub link :

•For Data Pipeline details, Kindly refer to the GitHub link:

References



[1] Armin Oliya, Amir Saffari, Priyanka Sen, and Tom Ayoola. 2021. End-to-end entity resolution and question answering using differentiable knowledge graphs. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 4193–4200, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[2] Priyanka Sen, Alham Fikri Aji, and Amir Saffari. 2022. Mintaka: A Complex, Natural, and Multilingual Dataset for End-to-End Question Answering. In Proceedings of the 29th International Conference on Computational Linguistics, pages 1604–1619, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

[3] William W. Cohen, Haitian Sun, R. Alex Hofer, and Matthew Siegler. 2020. Scalable neural methods for reasoning with a symbolic knowledge base. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net.

[4] https://query.wikidata.org/

[5] Chakraborty, Nilesh et al. "Introduction to neural network-based question answering over knowledge graphs." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11 (2021)