Knowledge Graph – Deep Learning: A Case Study in Question Answering in Aviation Safety Domain

Ankush Agarwal†, Raj Gite†, Shreya Laddha†, Pushpak Bhattacharyya†, Satyanarayan Kar◇, Asif Ekbal‡, Prabhjit Thind◇, Rajesh Zele†, Ravi Shankar◇

†IIT Bombay, ◇Honeywell, ‡IIT Patna

Motivation

- Post every flight, need to address problems (if any), faced and recorded by pilots.
- The maintenance people need to **consult manuals** which are **voluminous**.
- This process consumes a **lot of time**. Sometimes manuals are **overlooked**, and the incorrectly detected faulty part is replaced (false positive). A more dangerous situation is a false negative.
- A search-engine-like system would quickly bring up relevant information from aircraft manuals, identify the cause, save time & effort, and reduce maintenance costs.

Introduction

- We contribute an **Aviation Knowledge Graph**, constructed from accident reports using the domain knowledge and information extraction techniques to the community.
- We propose a **Question Answering system** that answers the questions asked in natural language from two modules:
  - (a) **KG-based QA (KGQA)** module, which uses a pipeline to convert Natural Language Queries to SPARQL queries and fetches responses from the Aviation Knowledge Graph.
  - (b) **DL-based QA (DLQA)** module that extracts answers from the plain text in the documents. The DLQA module has been tested using two different QA models, BERT-QA and GPT3-QA.
- We show that a **combined Question Answering system**, such as ours, outperforms the individual Knowledge Graph and Deep Learning methods on our curated test set.
- The combined system **KGQA+BERT-QA** attains **7%** and **40.3%** increase in accuracy over KGQA and BERT-QA modules respectively. **KGQA+GPT3-QA** attains **29.3%** and **9.3%** increase over KGQA and GPT3-QA modules respectively.

Problem Statement

Given a query, retrieve relevant information from a collection of documents using KG and DL. The question(input) is in natural language, and the output should be the solution to the input. The result can be a direct answer from KG or a fragment of text from the DL model.

Data Repositories

- The **NTSB** store investigation reports of civil **aviation accidents** in the US. It captures all the aspects of an accident, namely aircraft specifications, pilot details, environmental state, and a comprehensive description of the suspected cause of the accident.
- The **ADREP taxonomy** contains a complex multi-level hierarchy of factual descriptors (time, place, aircraft models, engine, component manufacturers, etc.).

Aviation Knowledge Graph

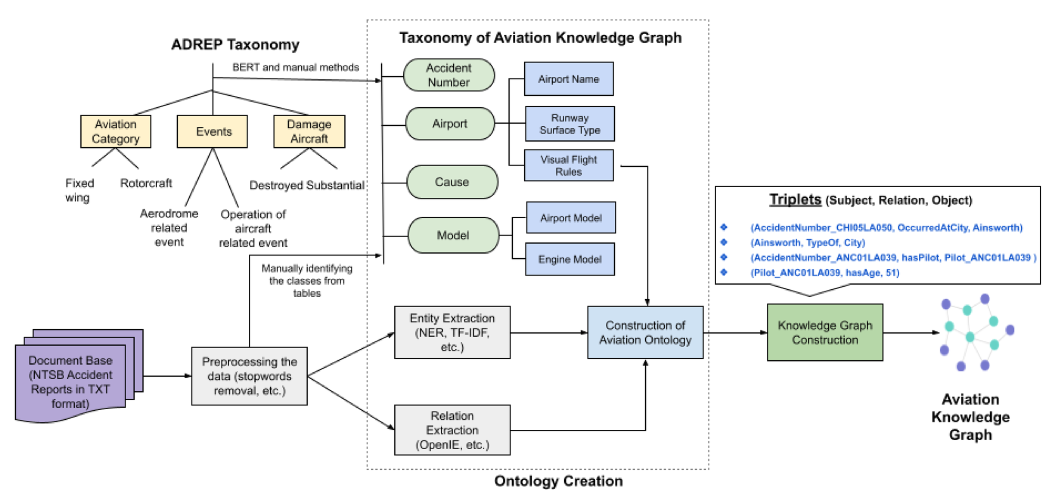

Figure 1:Aviation KG Construction Process from NTSB reports with assistance from ADREP taxonomy

Metrics	Count(#)
Entity Class	239
Individual	8894
Object Property	300
Data Property	71
Axioms	97879
Part of Relation among Classes	494
Property of Relation among Classes	353

Table 1:Properties of Aviation Knowledge Graph

A total of 120 SPARQL queries with gold answers of different categories were tested for the KG evaluation, where our KG answered 83 questions.

Analysis

- **GPT3-QA** is the best performing model among the **individual models** owing to its large size and extensive pre-training.
- **KGQA+GPT3-QA** is the **best model** for the QA task. It achieves an increase of **9.3%** in **answer accuracy** and a **5.3%** increase in **passage retrieval accuracy** over GPT3-QA.

Case Study

1. What discrepancy was noted due to which flight landed at La Belle Municipal Airport?
   Gold Answer: problem with fuel gauge
   KGQA model prediction: NA
   GPT3-QA model prediction: problem with fuel gauge
2. Which accidents involved aircraft operated by Johnny Thornley and manufactured by Subaru?
   Gold Answer: FTW02LA110
   KGQA model prediction: FTW02LA110
   GPT3-QA model prediction: SEA02FA036 and LAX02LA123

KGQA+GPT3-QA model answers both the above questions correctly.

Results

Model	Answers		Answers		Passages	
	Exact Match	Exact Recall	Semantic Accuracy	Semantic Recall	Semantic Accuracy	Semantic Recall
BM25-BERT (baseline)	0.013	0.153	0.113	0.143	0.333	0.253
KGQA	0.347	0.345	0.560	0.545	-	-
BERT-QA	0.040	0.157	0.227	0.217	0.393	0.311
GPT3-QA	**0.600**	0.547	0.760	0.620	0.813	0.734
KGQA + BERT-QA	0.349	0.403	0.630	0.588	0.788	0.715
KGQA + GPT3-QA	0.500	**0.628**	**0.853**	**0.680**	**0.866**	**0.784**

Table 2:Evaluation results of 'Answers' and 'Passages' predicted by our Question Answering models. The models are the results of our curated test set created from the NTSB reports.

QA System

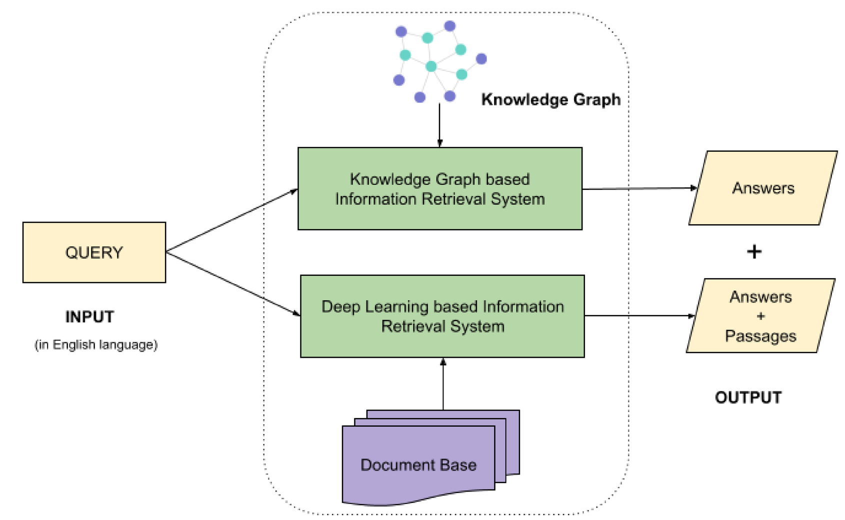

Figure 2:Knowledge Graph guided Deep Learning based Question Answering system

We combined KG and DL because KGQA has the advantage of domain knowledge, and DLQA has broader coverage over text.

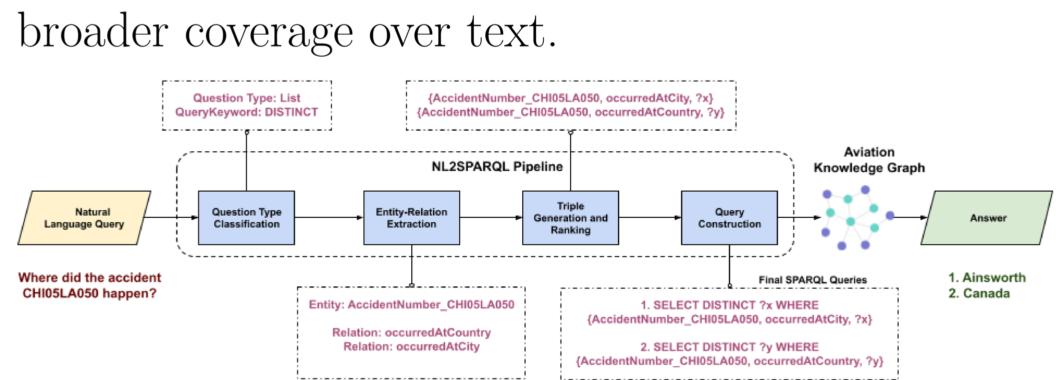

Figure 3:NL2SPARQL pipeline for KGQA system

We adopted BERT-QA and GPT3-QA for DLQA system because BERT-QA extracts a text span from the relevant passage as an answer (extractive QA), while GPT3-QA generates text from the relevant passages (abstractive QA).

Conclusion

- We have successfully created an Aviation Knowledge Graph from NTSB aircraft accident reports to help experts in their study.
- We have developed the KGQA model on our Aviation KG and DLQA model on the NTSB reports.
- We have built a Knowledge Graph guided Deep Learning based Question Answering system, which outperforms the individual KGQA and DLQA systems.
- The dominance of the combined QA system is proved theoretically and empirically by evaluating it on our handcrafted test set.

Future Work

We plan to expand our knowledge scope in aviation safety and make our KGQA + DLQA system robust.

Resource & Code Repository

https://github.com/RajGite/KG-assisted-DL-based-QA-in-Aviation-Domain