LEONARDO

Leonardo Labs

Ital-IA

ITALIA INTELLIGENZA ARTIFICIALE

cini National Lab AIIS

An adaptable method for developing an Open-domain Question Answering system

Sara Piscitelli, Sara Cardarelli, Alessandro Nicolosi

Lab of Applied Artificial Intelligence, Leonardo Labs, Leonardo S.p.A.

Pisa, Italy, 30/05/2023

Electronics

Helicopters

Aircraft

Cyber & Security

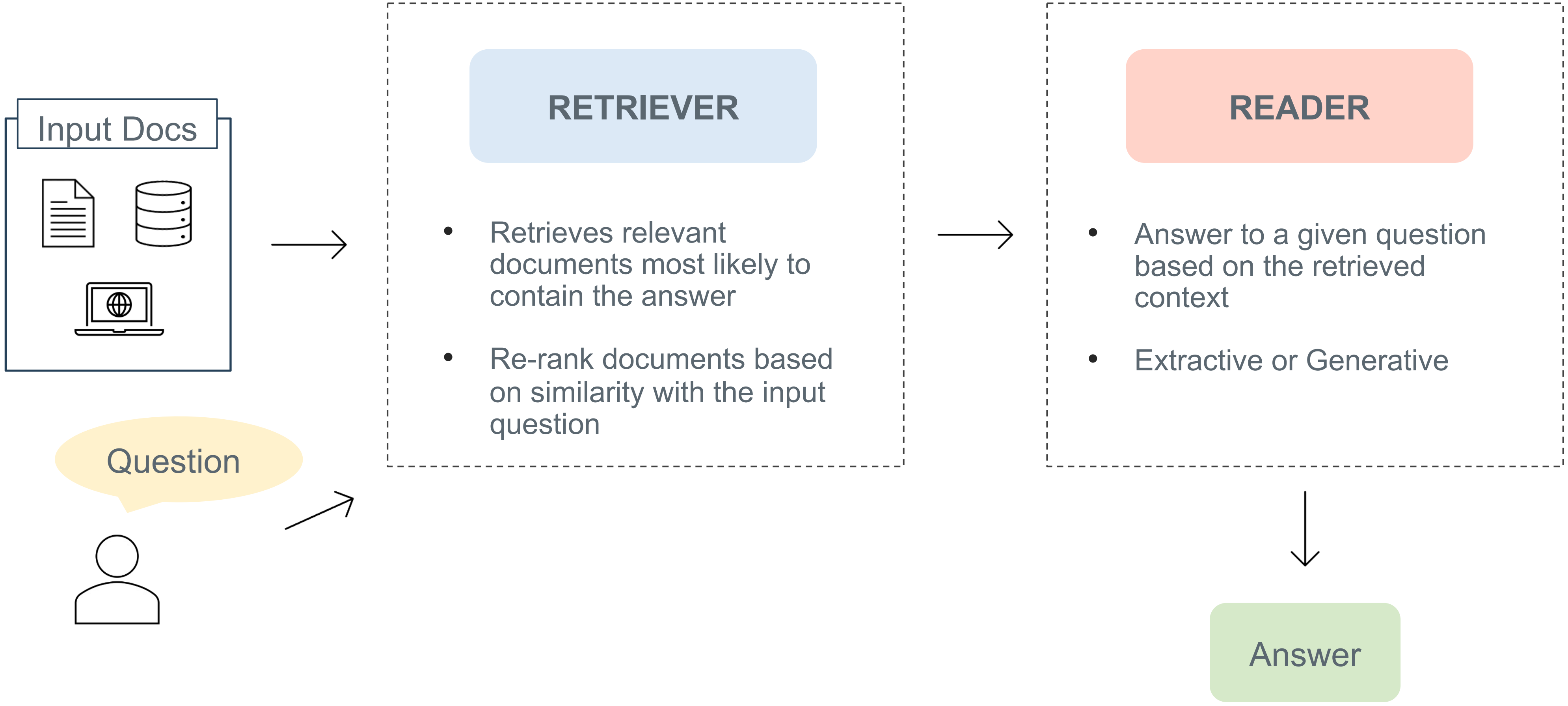
Space

Unmanned Systems

Aerostructures

An Open-Domain Question Answering System

- **Goal**: Answer user queries in natural language by searching vast unstructured text collections



Our Open-Domain Question Answering System

INFORMATION RETRIEVAL SYSTEM

Retrieve a corpus of documents

- open web sources (web scraping)
- text documents (e.g., internal documents, emails, online reviews, and other sources)

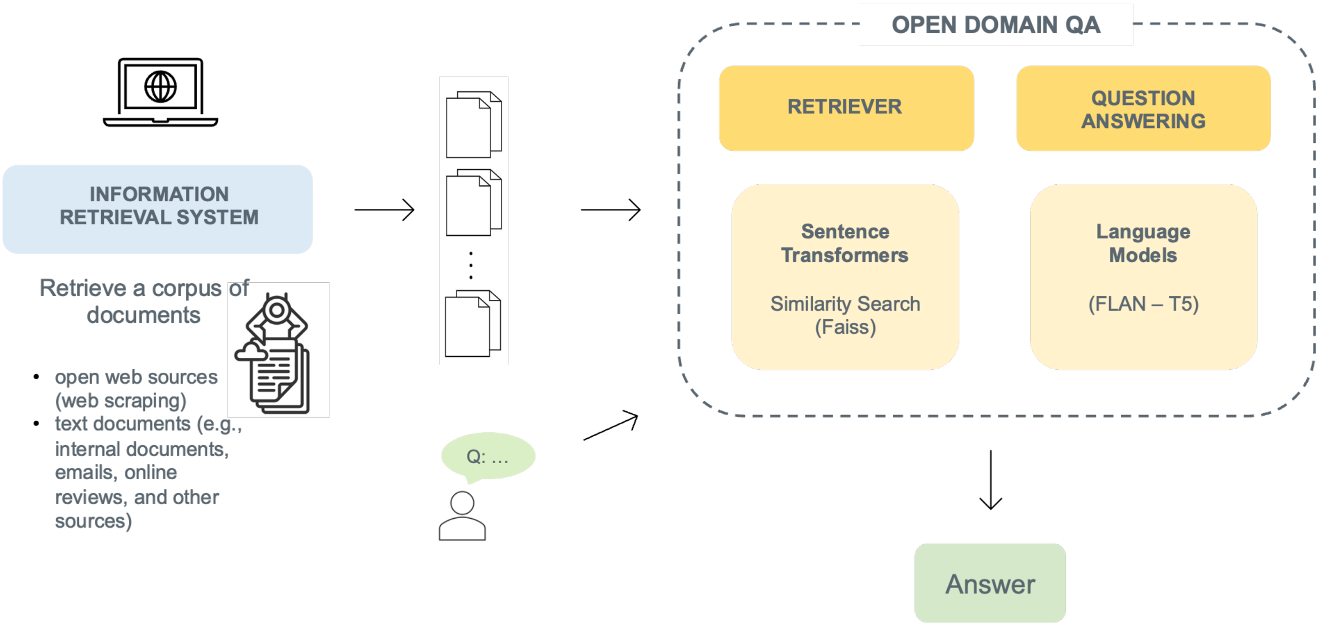
Q: …

OPEN DOMAIN QA

RETRIEVER	**QUESTION ANSWERING**
Sentence Transformers Similarity Search (Faiss)	**Language Models** (FLAN – T5)

Answer

Our Open-Domain Question Answering System



- **Main contributions**

  1. Transformer models in both the retriever and question answering components

  2. Not limited to a specific domain and can be readily adapted to various use cases and models

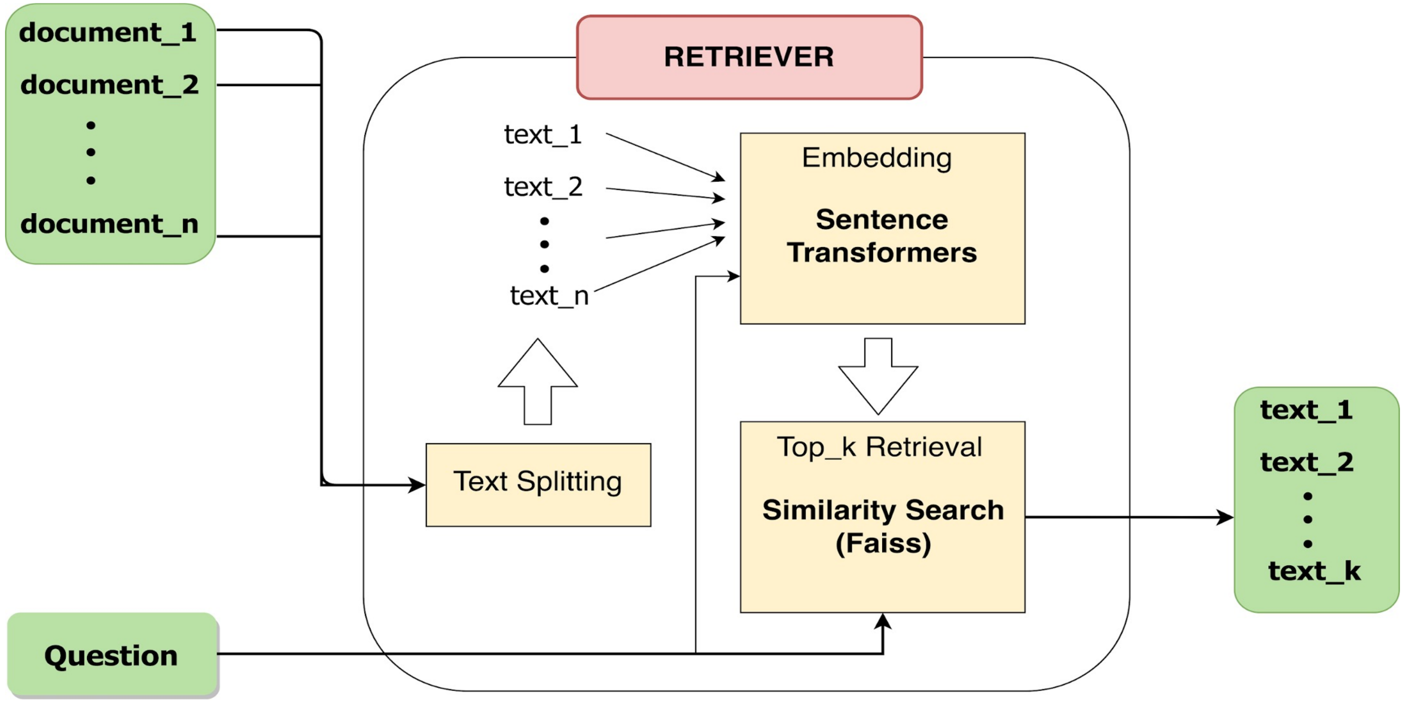
  3. Instruction fine-tuned language model to generate and extract an unique answer

Retriever

1. **Text Splitting**
   - Overcome maximum input length constraint (512 tokens)
   - Chunking: divide text into smaller parts
   - Overlap between chunks to mantain context

2. **Sentence Embedding**
   - Pre-trained sentence transformers
     - *Msmarco-distilbert-base-v4*[1]

3. **Similarity Search**
   - Retrieve top_k best candidates
   - Calculate cosine similarity between question and each text chunk
   - Efficient similarity search using *Faiss*[2] library



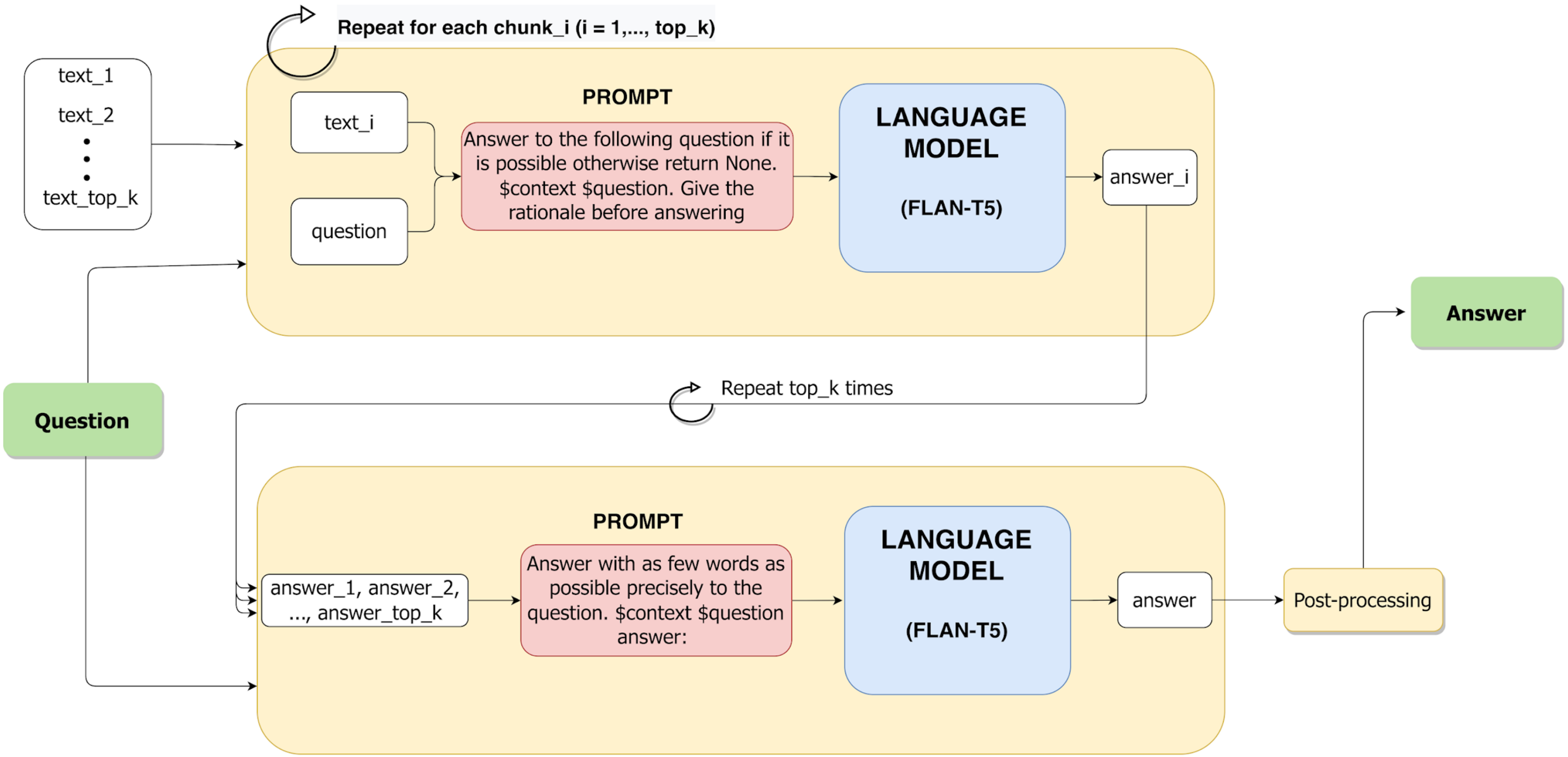
1. *https://huggingface.co/sentence-transformers/msmarco-distilbert-base-v4*
2. *https://github.com/facebookresearch/faiss*

Question Answering System

- Instruction fine-tuned language model: *flan-t5-large*[1]

1. **Answer Generation**
   Generate an answer to a given question based on a specific context

2. **Answer Aggregation**
   Combine multiple answers into a single concise response



1. *https://huggingface.co/google/flan-t5-large*

Datasets

- ***TriviaQA*[1] Dataset**
  - **487K documents** from web and Wikipedia domain
  - **650K question-answer pairs**
  - 2 versions:
    - ***TriviaQA rc web dev*:** answers are verified to be present within the corpus of documents
    - ***TriviaQA unfiltered web dev*:** no guarantees that the answer is present in the corpus of documents

→

Question: "Who was the next British Prime Minister after Arthur Balfour?"
Answer: "Sir Henry Campbell-Bannerman", "Campbell-Bannerman", "Campbell Bannerman", "Sir Henry Campbell Bannerman", "Henry Campbell Bannerman", "Henry Campbell-Bannerman ".

Excerpt: " … **Balfour** resigned as Prime Minister in December 1905, hoping the Liberal leader **Campbell-Bannerman** would be unable to form a strong government. This was dashed when Campbell-Bannerman faced down an attempt ("The Relugas Compact") to "kick him upstairs" to the House of Lords. … In 1906, the Liberal party, led by Sir **Henry Campbell-Bannerman**, won an overwhelming victory on a platform that promised social reforms for the working class. With 379 seats compared to the Conservatives' 132, the Liberals could confidently expect to pass their legislative programme through the Commons. At the same time, however, the Conservative Party had a huge majority in the Lords; it could easily veto any legislation passed by the Commons that was against their interests. … ."

1. *https://nlp.cs.washington.edu/triviaqa/*

Evaluation

- **Open-domain generative QA**
  - Retrieve the correct answers from a corpus of almost 487K documents
- **Reading comprehension**
  - Extract the correct answer from an average of six input documents, which exclusively consist of evidence documents

Reading comprehension QA	F1	EM	QA pairs
TriviaQA rc web dev	75.95	68.73	68.617

Open-domain generative QA	F1	EM	QA pairs
TriviaQA unfiltered web dev	66.32	58.13	131.993
TriviaQA rc web dev	70.85	63.14	68.617

- Reading Comprehension vs Open-domain generative QA
  - Retriever is more deceived by the large number of documents in the corpus
    - 5% decrease in F1 score

Evaluation

- **Open-domain generative QA**
  - Retrieve the correct answers from a corpus of almost 487K documents
- **Reading comprehension**
  - Extract the correct answer from an average of six input documents, which exclusively consist of evidence documents

Reading comprehension QA	F1	EM	QA pairs
TriviaQA rc web dev	75.95	68.73	68.617

Open-domain generative QA	F1	EM	QA pairs
TriviaQA unfiltered web dev	66.32	58.13	131.993
TriviaQA rc web dev	70.85	63.14	68.617

- Reading Comprehension vs Open-domain generative QA
  - Retriever is more deceived by the large number of documents in the corpus
    - 5% decrease in F1 score
- TriviaQA unfiltered vs TriviaQA filtered
  - The input document corpus is not guaranteed to contain all the answers
    - 4% decrease in F1 score

Conclusions

- An **instruction fine-tuned language** models can be succesfully used in ‘**zero-shot’** to:
  - Generate an answer for a specific question within context
  - Aggregate different answers into one
- The proposed system can be further improved and adapted to various use cases

Next Steps

1. Train a specific instruction-tuned language model
2. Enhance retriever component
3. Increase language model’s maximum input length
4. Better evaluation of the system and its limitations

LEONARDO

Leonardo Labs

THANK **YOU** FOR YOUR ATTENTION

sara.piscitelli.ext@leonardo.com

sara.cardarelli.ext@leonardo.com

alessandro.nicolosi.ext@leonardo.com