CAIRO

Coffee Machine Q&A System

Esther Ademola, Marius Benkert, Lennard Rose, Jochen Schmidt

Ask Questions about your **coffee machine** to our Q&A System and get the answers you are looking for!

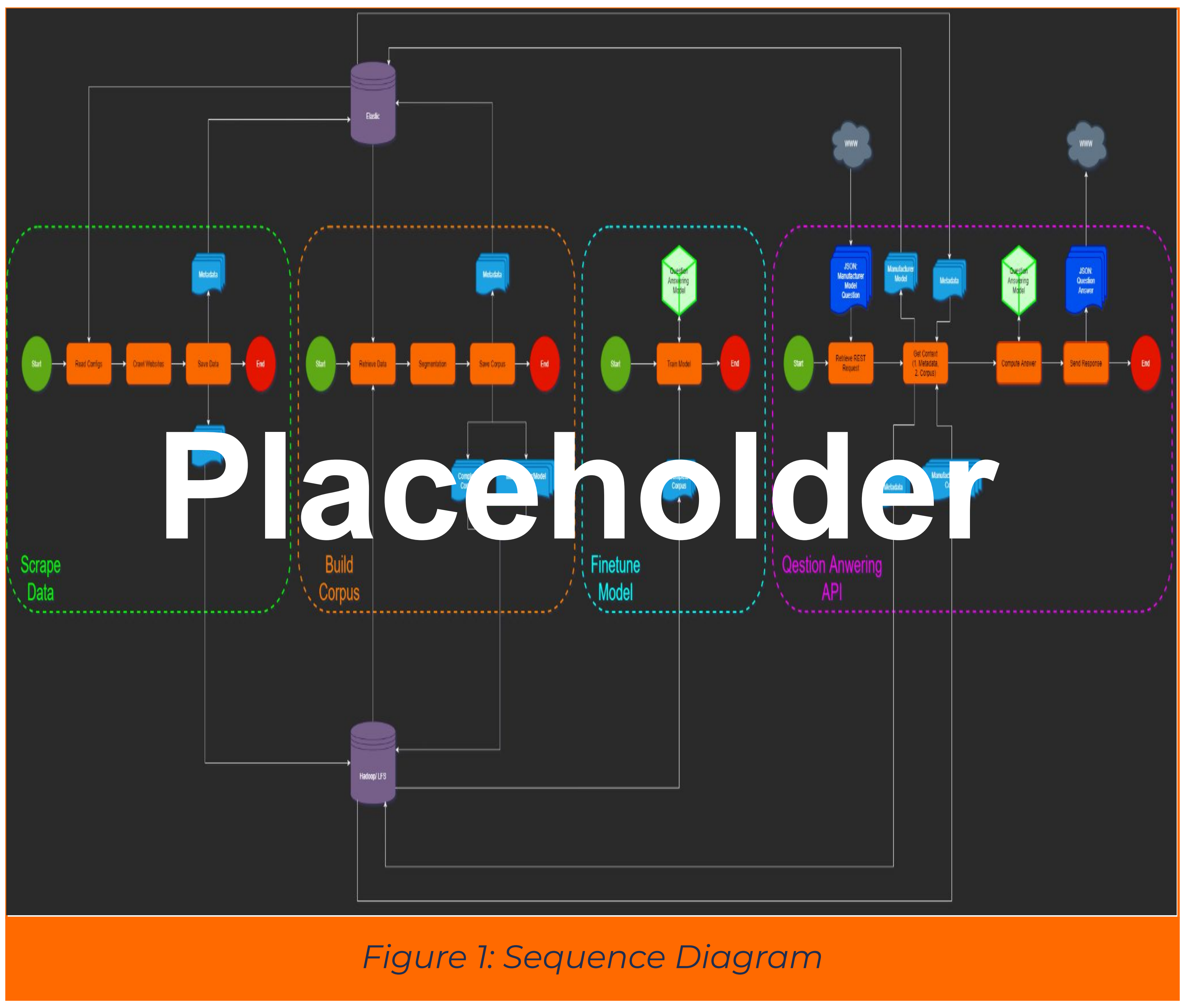

Figure 1: Sequence Diagram

Motivation

Problem:
Searching for an answer within technical documents such as Coffee Machines can be tiresome and time consuming.

Solution:
We've built a system that is able to answer technical questions for a user provided question.

Web Scraping

- Defined 10 manufacturers as our sources
- Automatically query web pages with BeautifulSoup and Selenium for documents and metadata
- Store documents and metadata to the data warehouse

Data Warehouse

- Hadoop serves as data storage for the manuals
- As a distributed file system this ensures availability
- The metadata to each manual are stored in Elasticsearch (ES)
- ES serves as a search and management tool and helps finding the relevant documents for each question

Corpus

- Separation of every manufacturer and their products
- Segmentation of individual manuals into Header, Subheader and Paragraphs
- Segmented corpus stored in Elasticsearch for fast retrieval

Question Answering

- Created training Q&A Dataset from our Corpus
- Trained NLP-Models via Huggingface
- Integrated Sentence Similarity Search to get relevant Documents to the Question
- Query the deployed model to retrieve answers for your questions from the relevant Documents

Model	With Fine-Tuning		Without Fine-Tuning	
	F1	Exact-Match	F1	Exact-Match
bert-base-uncased	61.7	54.6	10.4	0.3
distilbert-base-uncased-squad	66.4	57.0	49.9	21.1
bert-base-cased-squad2	67.6	55.8	52.3	22.0
roberta-base	66.7	57.2	14.6	0.15
roberta-base-squad2	**68.3**	**59.8**	**58.3**	**27.5**

Figure 2: Table comparing Models

thws
Technical University of Applied Sciences Würzburg-Schweinfurt

References:

1. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., and others 2019. Huggingface's transformers: State-of-the-art natural language processing.
2. Gormley, C., and Tong, Z. 2015. . Elasticsearch: the definitive guide: a distributed real-time search and analytics engine. O'Reilly Media, Inc..

Esther Ademola
Esther.Ademola@study.thws.de

Marius Benkert
Marius.Benkert@study.thws.de

Lennard Rose
Lennard.Rose@study.thws.de

Jochen Schmidt
Jochen.Schmidt.1@study.thws.de

Scan the **QR-Code** for more vital **information**