

# Text-retrieval based Chatbot

Oliver Foth, Ole Moering, Ranjiraj Rajendran Nair

Group - Q

Faculty of Computer Science

Otto-von-Guericke-University Magdeburg, Germany

January 17, 2022

# About our use case

- ▶ "Topic Explainer"- for querying files present in the system.
- ▶ Data: Text files
- ▶ Bot is expected to reply with greetings, answers to questions
- ▶ Chatbot API: Apache OpenNLP (1.5 series)
- ▶ Used few trained models
- ▶ 'en-sent.bin', 'en-token.bin', 'en-pos-maxent.bin', 'en-lemmatizer.bin'

# About OpenNLP API

- ▶ An open-source Java library
- ▶ Pre-process natural language text
- ▶ Tokenization, Sentence Segmentation, POS tagging, NER, chunking, parsing etc.

# NLP Pipeline

- ▶ Trains categorizer model with latest samples data
- ▶ Take input from the user
- ▶ Partition the sentences
- ▶ Tokenize each sentences into words
- ▶ Find POS tags for each word
- ▶ Lemmatize each word using tokens & POS tags
- ▶ Categorize lemma tokens & then find answer for detected category

# Training Data

- ▶ **greeting**: Basic greetings that user will to use to start chat

# Training Data

- ▶ **greeting**: Basic greetings that user will to use to start chat
- ▶ **conversation-continue**: Words like “ok”, “hmm” that user might use in between of conversation

# Training Data

- ▶ **greeting**: Basic greetings that user will use to start chat
- ▶ **conversation-continue**: Words like “ok”, “hmm” that user might use in between of conversation
- ▶ **question**: Questions that user will ask questions in a different way for various topics

# Training Data

- ▶ **greeting**: Basic greetings that user will use to start chat
- ▶ **conversation-continue**: Words like “ok”, “hmm” that user might use in between of conversation
- ▶ **question**: Questions that user will ask questions in a different way for various topics
- ▶ **conversation-complete**: Words or sentences that user might end to end conversation



# References

- [1] <http://opennlp.sourceforge.net/models-1.5/>
- [2] <https://opennlp.apache.org/models.html>
- [3] <https://opennlp.apache.org/docs/1.8.1/manual/opennlp.html>
- [4] <https://self-learning-java-tutorial.blogspot.com/2015/10/opennlp-tokenization.html>
- [5] <http://www.java2s.com/example/java-src/pkg/opennlp/tools/util/trainingparameters-98536.html>

Thanks for your attention!

Now Demo