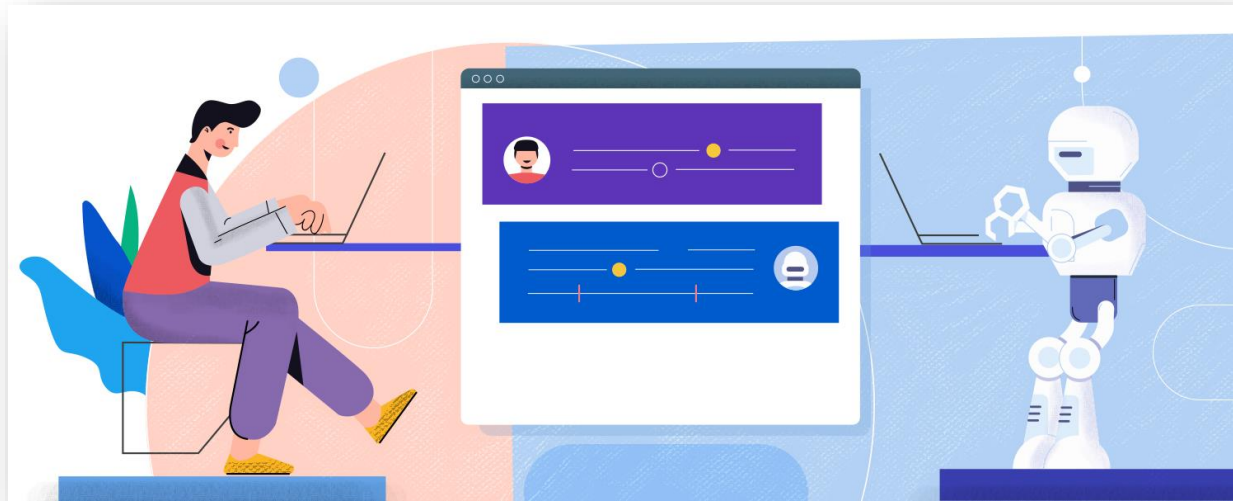# H.E.R.A

HELPFUL INFORMATION RETRIEVAL ASSISTANT

Bulelani Nkosi and Caryn Pialat
Supervised by Ebrahim Noormahomed

Introduction to Chatbots

▶ Applications that use AI and natural language processing (NLP) to automate responses to user queries, simulating human conversation, in place of a human agent.

Where are chatbots used?

▶ Estimated that up to 85% of customer interactions will be managed by chatbots by the end of 2021 (https://www.smallbizgenius.net/by-the-numbers/chatbot-statistics/#gref)

▶ Chatbots can be used in any environment where client interaction is necessary (IBM)

▶ Pros:

   ▶ Manage multiple users thereby avoiding long wait times

   ▶ Reduce costs of maintaining a 24-hour support centre

   ▶ Improve sales rates by rapidly answering product-related questions and guiding customers towards purchasing

   ▶ Boosting customer satisfaction by creating efficient and positive user experiences

Problem Statement

▶ Explore Data Science Academy is an amazing company helping South Africas youth do amazing things. This repository is a testament to that. Explore is an educational institution in the information systems development field.

▶ Inquiries are an administrative burden on companies. A substantial amount of time is spent on drafting responses to each one. Further, These responses are often repetitive.

Hera aims to address this issue by creating an information retrieval assistant to act as the first trouble shooting step before contacting a member of staff.
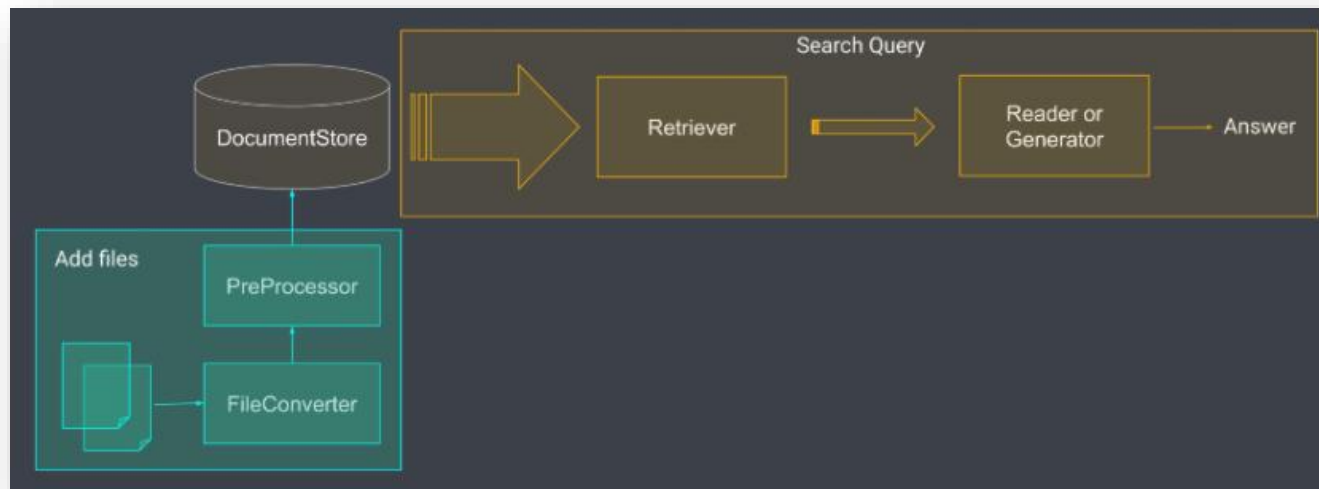
Introducing H.E.R.A.

Solution

- Hera is built on a Haystack backend
    - Haystack is a tool that provides a pipeline for Closed Domain Question Answering using the latest pretrained models
    - Full details available at https://github.com/BNkosi/odin

Solution

- ▶ H.E.R.A works by scraping your website and loading this data into an ElasticSearch Document Store

  - ▶ Questions can be answered immediately

  - ▶ Training a model boosts accuracy

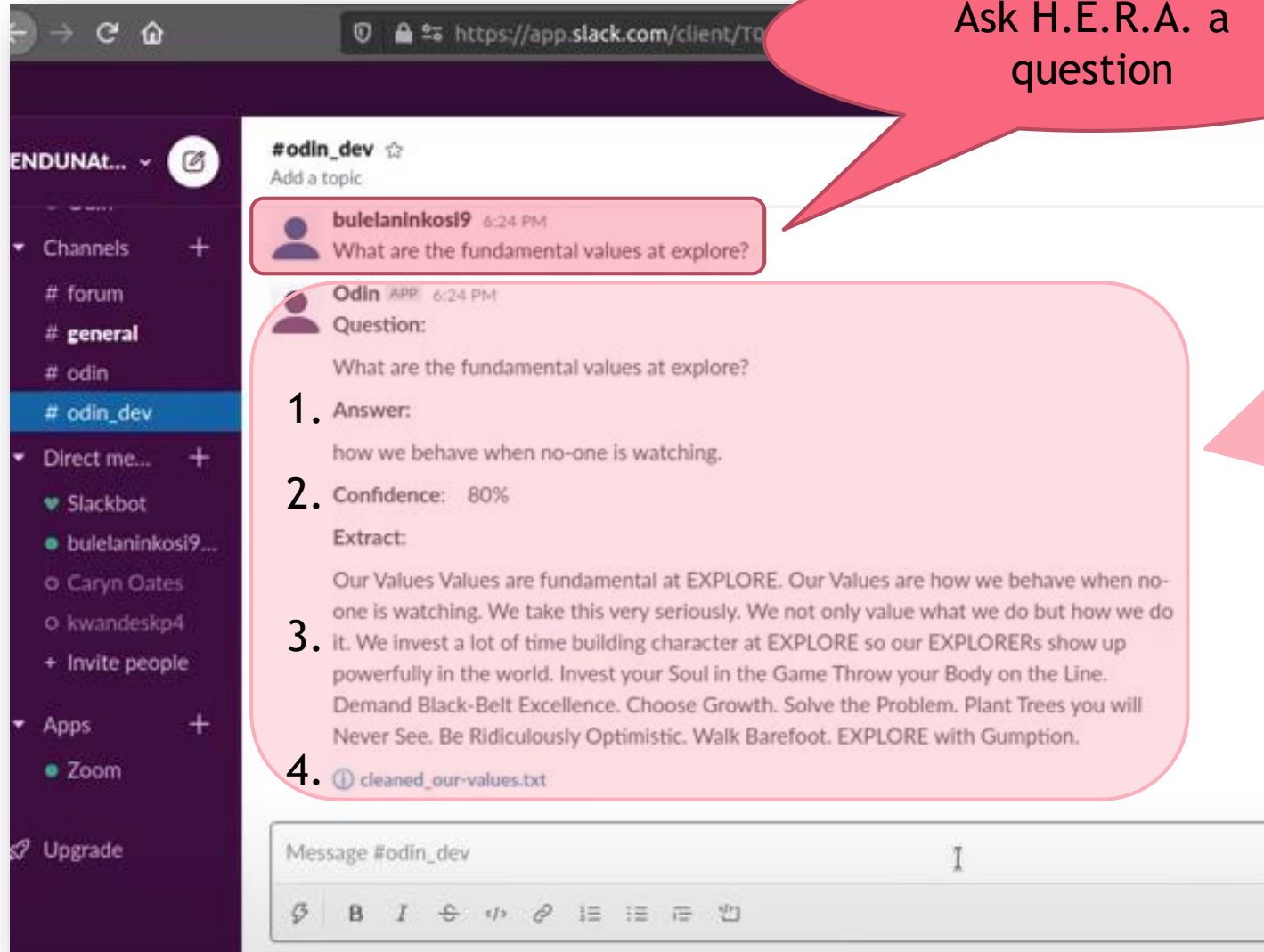| Scrape website | Clean and load data | Create tunnel between Slack and your local server | Invite the bot to your slack channel | Start asking questions |

Solution

Ask H.E.R.A. a question

The bot will generate 4 pieces of information in response:

1. The answer to the question
2. A confidence level (ie. How confident the bot is about the answer)
3. The excerpt of text where the answer was found
4. The name of the file where the answer was found

Future work

▶ Deployment - Chatwoot has been identified as the prefered tool to integrate the Slack window into websites.

▶ Retrieval-Augmented Generation - Currently the model works by selecting the most appropriate span of text and presenting it as the answer (extractive QA). The next step is the generation of novel answers from the same documents. This makes the bot more human-like and thus more trust-worthy.

▶ Generative Pretraining - The annotation tool is labour intesive and costly. Generative Pretraining aims to simulate a human asking and answerinig (annotating) documents.

▶ Context Management - Users ask ambiguous questions, such as: "How much is this course?". Context management allows the bot to know which course the user is talking about or ask for clarity where there is uncertainty.

▶ Feedback - Logging of questions still needs to be built in as well as notification when answers aren't found. Notifications must be balanced between always sending notifications and always answering the question.

Acknowledgements

▶ Our project supervisors:

▶ Ebrahim Noormahomed and Joanne Moonsamy

▶ Explore Data Science Academy