Process Meeting #1

23 October 2019 / 12:00-13:30 / ROOM JHL 3304

ATTENDEES

All.

AGENDA

Project board creation and task formalization

- ❏ Github Project board creation
- ❏ Basic task construction on GitHub Projects board.

NOTES

- ❏ Created tasks and added to To-Do list
- ❏ Gave a high level story points to each of the tasks.

ACTION ITEMS && NEXT MEETING'S AGENDA

- ❏ Coming up with more tasks
- ❏ Assigning tasks to individuals

Process Meeting #2

25 October 2019 / 14:00-16:00 / ROOM JHL 4321

ATTENDEES

All.

AGENDA

Task Formalization and Assignment

- ❏ Finalizing the tasks and assigning it to the team.

Data Investigation

- ❏ To Download and investigate StackOverflow data.
- ❏ Determine relevant/irrelevant components of the dataset.

Algorithm Discussions

- ❏ Algorithm and logic behind text analysis and matching questions with Stackoverflow answers.

NOTES

- ❏ **Task Formalization and Assignment** Added more tasks and completed task construction on GitHut Projects board.
- ❏ **Data Investigation** Found the data to be huge, so to apply some form of splitting of questions and answers.
- ❏ **Algorithm Discussions** To read and come up with ideas for algorithms for text analysis and tokenization.

ACTION ITEMS

- ❏ **Data set creation** Rishal and Michael to begin dataset creation and formulation.
- ❏ **Algorithm Analysis** Rohit and Prathamesh to find algorithms to match questions with their respective answers from the dataset.

NEXT WEEK'S AGENDA

Start to find algorithms to handle the logic and creation of datasets as per tasks assigned.

Process Meeting #3

29 October 2019 / 14:00-15:00 / JHL Graduate Commons

ATTENDEES

All.

AGENDA

Refining Extraction of Questions

- ❏ Discuss doubts with refining the previous BOT phase code.

Data Investigation - II

- ❏ How to break down the StackOverflow data.
- ❏ Determine format and subsections of Example dataset.

Algorithm Finalizing

- ❏ Discuss the available algorithms for matching questions with Stackoverflow answers.

NOTES

- ❏ **Refining Extraction of Questions** Decided to modularize the code and break it down into queries and examples.
- ❏ **Data Investigation-II** Found that data needs to be filtered first and then can be put into different tables as per the GIT tag.
- ❏ **Algorithm Finalizing** Discussed two approaches of matching Stackoverflow Question-Answer pair - one for queries and other for generating examples.

ACTION ITEMS

- **Example handling** Rohit to start building code around the sample dataset built by Rishal.
- **Query handling** Prathamesh to try out tokenization algorithms and Michael to split StackOverflow dump as per tags.

NEXT MEETING'S AGENDA

Finalizing the dataset structure and testing it with the algorithm.

Process Meeting #4

01 November 2019 / 10:00-12:30 / JHL Graduate Commons

ATTENDEES

All.

AGENDA

Dataset final structure

- ❏ Finalize the structure of both the Example and StackOverflow datasets.

Query Processing Issue

- ❏ Discussing logic behind matching of NC State StackOverflow questions with StackOverflow data dump.

NOTES

- ❏ **Dataset final structure** Decided and agreed on the json structure of the Example dataset.
- ❏ **Query Processing Issue** Need to decrease dump size and then implement and test vectorization of answers.

ACTION ITEMS

- ❏ **Example handling** Fully integrate logic of mapping questions to its associated examples.
- ❏ **Query handling** Find where to store dataset and implement vectorization.

NEXT MEETING'S AGENDA

Discussion on where to store the datasets and check status of tasks.

Process Meeting #5

04 November 2019 / 17:00–19:30 / JHL Graduate Commons

ATTENDEES

All.

AGENDA

Data Management and Online Storage

- ❏ Select appropriate platform for data storage/access.
- ❏ Investigate remote hosting for bot for more consistent uptime.
- ❏ Discuss large-scale example data generation.

Task Refinement

- ❏ Cleanup, move, and reorganize tasks.
- ❏ Designate new tasks for remote storage/deployment.

NOTES

- ❏ **Bot Hosting:** App Engine vs. Cloud Functions vs. Cloud Run vs. Kubernetes.
- ❏ **Pair Programming:** Worked together in Google Collab with a restricted subset of our data to try different NLP methods for text similarity.

ACTION ITEMS

- ❏ Extract data from local database (ETL), upload to remote storage; share.
- ❏ Compare Google execution and code hosting platforms that support Pyppeteer/headless browsing to find most appropriate.

NEXT MEETING'S AGENDA

Check-in on remote deployment and data storage options and example generation.

Process Meeting #6

07 November 2019 / Time 15:00 / Room JHL 4315

ATTENDEES

All.

AGENDA

Data Update

- ❏ Discuss final database format.
- ❏ Integrate DatabaseInterface with existing codebase.

NLP

- ❏ Run NLP through full dataset as accessed by bot through the interface.

Finalization

- ❏ Discuss progress made during iterations.
- ❏ Compare finalized bot to our speculative initial specifications.

NOTES

Data Update

- ❏ Local database broken down with ETL and uploaded to BigQuery.
- ❏ Format is similar, but had to be reduced to simple tag-based tabling.
- ❏ All added as admins on Google Console: should be available for access.
- ❏ DatabaseInterface completed and integrated with bot.

NLP

- ❏ Applied the previously investigated NLP approaches within the context of our existing bot workflow.
- ❏ Informally analyzed outcomes of each.

ACTION ITEMS

- ❏ Update PROCESS.md document and put usage instructions into README.md.
- ❏ Complete code documentation and clean up repo.
- ❏ Ensure that no secret information has been left in any files.