LSBU Course Finder Chatbot

Module Title: Future Internet Technologies (CSI_7_FIT)

Team Members:

Syed Jafrul Husen (4037556)
Shourob Ahmed Khan (4108471)
Syed Ishmam Mahmud (4103190)
Md Mehedhi Hasan (3626236)

What is chatbot?

Chatbot is a software application that allow humans to interact with digital device(text or speech)

LSBU course finder chatbot

A user, ideally student, who want to know details about a specific course it will gives proper response in chatbot as a text format.

Use case scenerio

- New student
- Parents/ guardian
- University admission team

Tool and Technology

- Rasa 2.0
- Python 3.8.10
- Visual studio as a code editor
- Slack for deployment

Development model

To develop this project, we followed **waterfall model.**

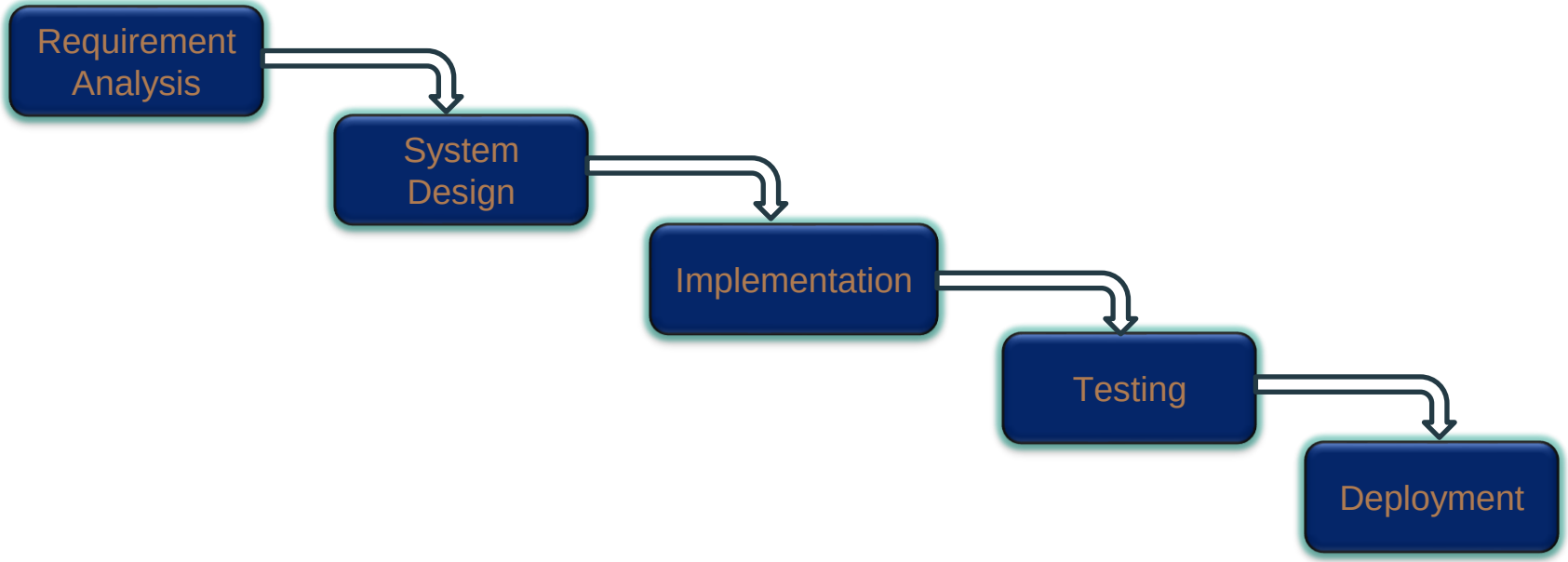


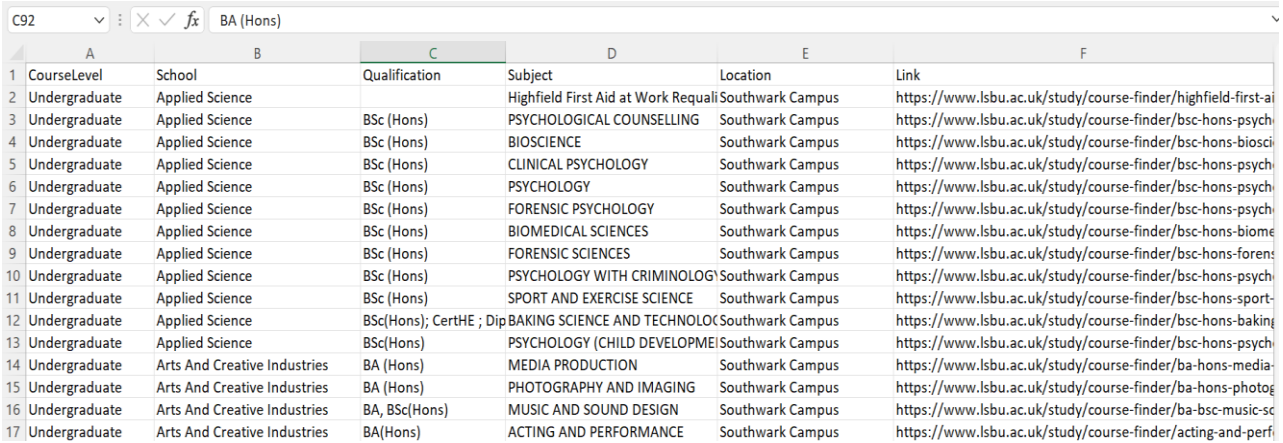
Fig 1: Waterfall Model

Why used waterfall model?

- Easy to use and manage due to rigid model
- Works well where requirement are specified and the project is small
- Easy to arrange task
- Clearly defined stages

Data collection

- LSBU website
- Collected details information about 150 subjects
- Data collected into csv file (comma delimited)



C92 | BA (Hons)

	A	B	C	D	E	F
1	CourseLevel	School	Qualification	Subject	Location	Link
2	Undergraduate	Applied Science		Highfield First Aid at Work Requali	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/highfield-first-ai
3	Undergraduate	Applied Science	BSc (Hons)	PSYCHOLOGICAL COUNSELLING	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
4	Undergraduate	Applied Science	BSc (Hons)	BIOSCIENCE	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-biosci
5	Undergraduate	Applied Science	BSc (Hons)	CLINICAL PSYCHOLOGY	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
6	Undergraduate	Applied Science	BSc (Hons)	PSYCHOLOGY	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
7	Undergraduate	Applied Science	BSc (Hons)	FORENSIC PSYCHOLOGY	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
8	Undergraduate	Applied Science	BSc (Hons)	BIOMEDICAL SCIENCES	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-biome
9	Undergraduate	Applied Science	BSc (Hons)	FORENSIC SCIENCES	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-forens
10	Undergraduate	Applied Science	BSc (Hons)	PSYCHOLOGY WITH CRIMINOLOGY	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
11	Undergraduate	Applied Science	BSc (Hons)	SPORT AND EXERCISE SCIENCE	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-sport-
12	Undergraduate	Applied Science	BSc(Hons); CertHE ; Dip	BAKING SCIENCE AND TECHNOLOG	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-baking
13	Undergraduate	Applied Science	BSc(Hons)	PSYCHOLOGY (CHILD DEVELOPME	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/bsc-hons-psych
14	Undergraduate	Arts And Creative Industries	BA (Hons)	MEDIA PRODUCTION	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/ba-hons-media-
15	Undergraduate	Arts And Creative Industries	BA (Hons)	PHOTOGRAPHY AND IMAGING	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/ba-hons-photog
16	Undergraduate	Arts And Creative Industries	BA, BSc(Hons)	MUSIC AND SOUND DESIGN	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/ba-bsc-music-sc
17	Undergraduate	Arts And Creative Industries	BA(Hons)	ACTING AND PERFORMANCE	Southwark Campus	https://www.lsbu.ac.uk/study/course-finder/acting-and-perf

Fig 2: Sample dataset

Technical details

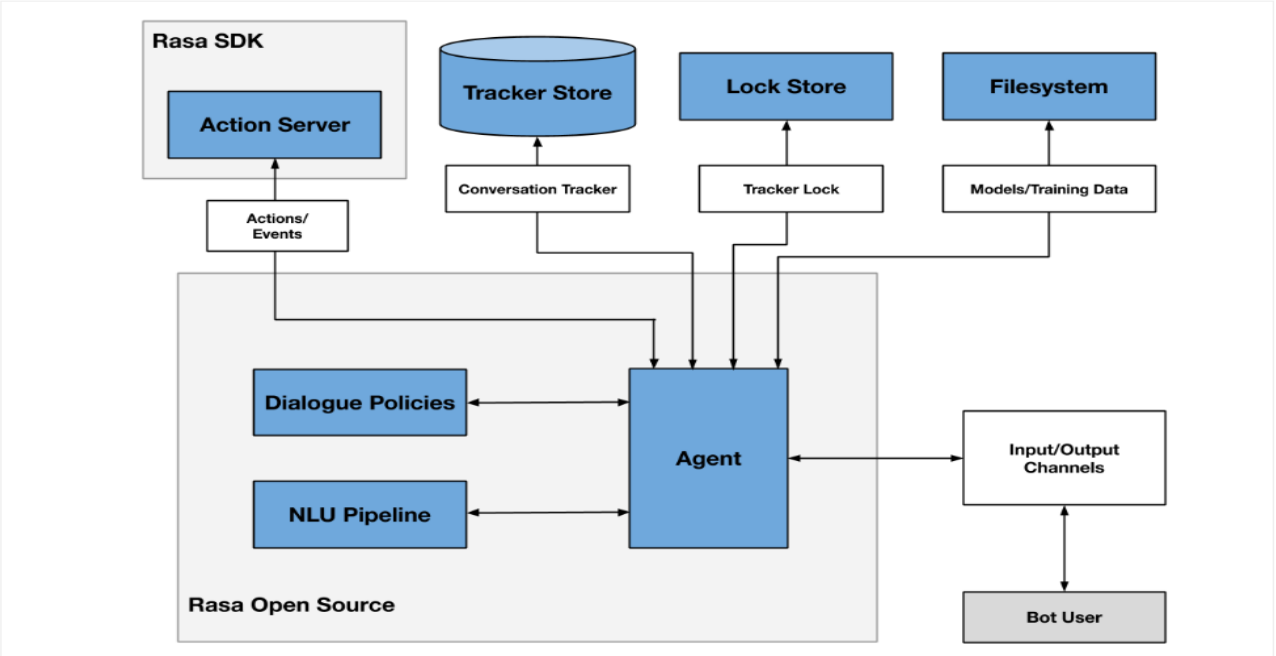


Fig 3: RASA architecture [2]

Pipeline

```yaml
pipeline:
  - name: WhitespaceTokenizer
  - name: RegexFeaturizer
  - name: LexicalSyntacticFeaturizer
  - name: CountVectorsFeaturizer
  - name: CountVectorsFeaturizer
    analyzer: char_wb
    min_ngram: 1
    max_ngram: 4
  - name: DIETClassifier
    epochs: 100
  - name: EntitySynonymMapper
  - name: ResponseSelector
    epochs: 100
  - name: FallbackClassifier
    threshold: 0.3
    ambiguity_threshold: 0.1
```

Fig 4: Pipeline

Policy

```yaml
policies:
  - name: MemoizationPolicy
  - name: TEDPolicy
    max_history: 5
    epochs: 100
  - name: RulePolicy
```

Fig 5: Policy

Technical details cont…

- Entity:
  - Course_level
  - Course
- Custom Action:
  - Action_course_search

Spelling corretor

- Algorithm: Peter norvigs
- Used Baye's theorem: $\text{argmax}_{c \in candidates} P(c)\, P(w|c) / P(w)$
- Four parts of this function:
  - Selection mechanism: argmax
  - Candidate model: $c \in candidates$
  - Language model: $P(c)$
  - Error Model: $P(w|c)$

Project Demo

Performance analysis

- Accuracy of spell corrector algorithm: 80% to 90%

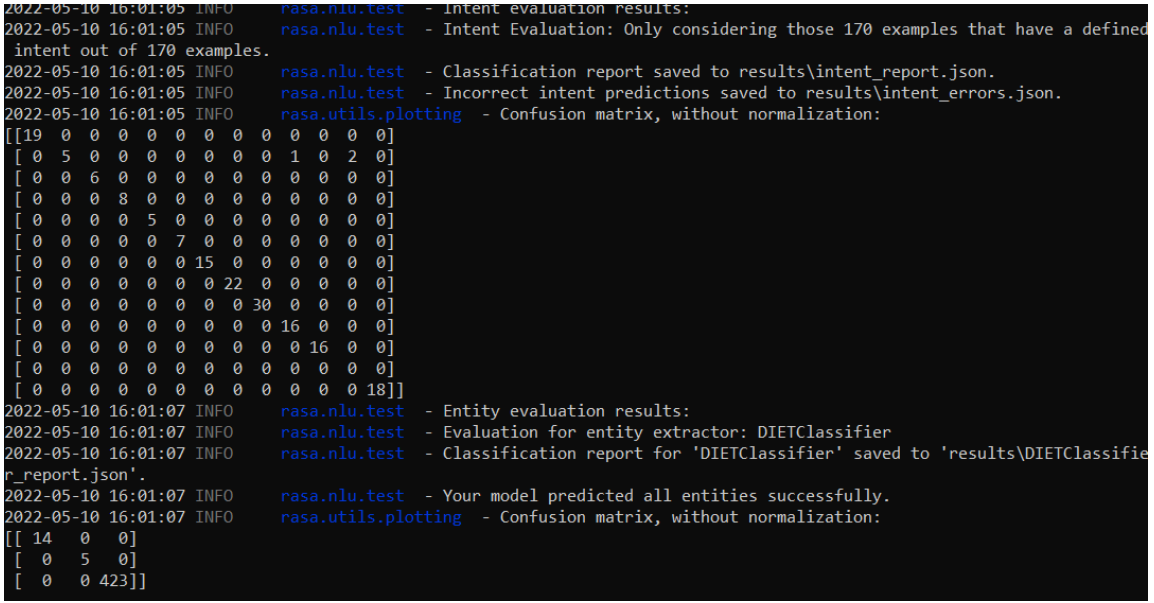


Fig 6: Confusion Matrix

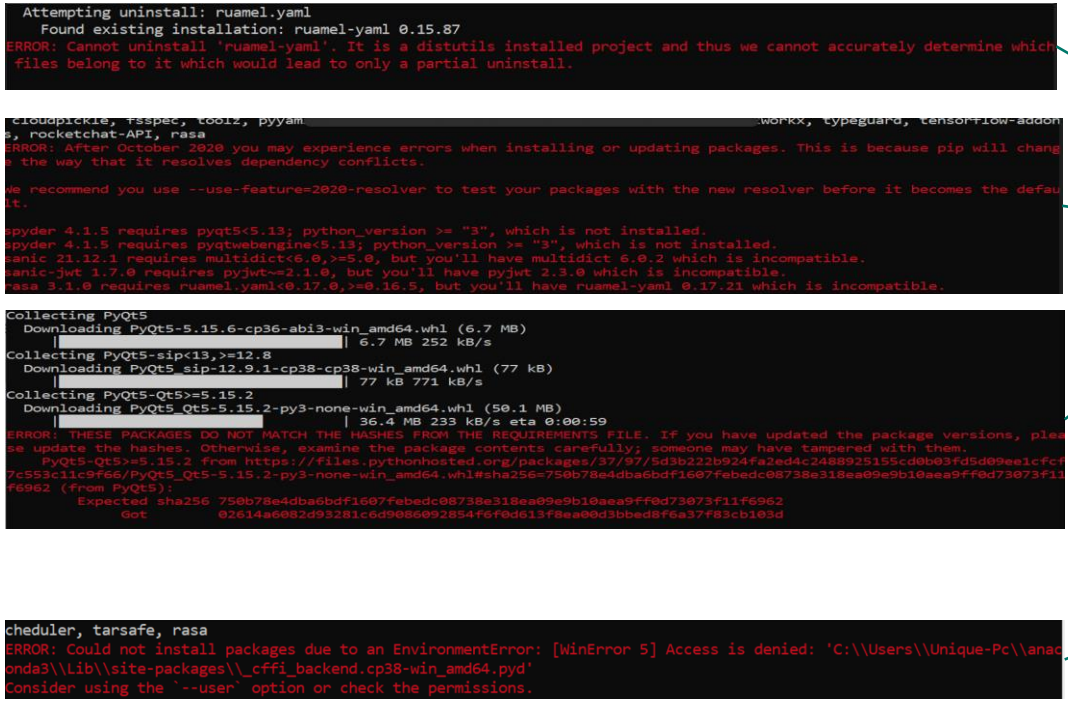
Testing

- Unit test:
  - Test separately python code as in rasa it take time to run and hard to integrate large code
    - Every unit of code tested separately (whitespace remove, edit distance calculation, custom actions, entity, slots, domain etc.)
- Integration test:
  - Evaluate how a piece of code work with the project
- Functional testing:
  - Check How the entire system works together

Teamwork

Task	Details
Requirement Analysis	As a team we collected the requirement and do analysis to make the project effective.
Data collection	Ishmam and Shourob collected data.
Create story	Jafrul and Mehedi created rasa story
System Design	Mehedi and Shourob did the requirement specification, specify the hardware and software requirements and architecture.
Implementation	We together implement the project and did coding using rasa framework. Jafrul implemented spelling correction and rasa custom actions.
Testing and troubleshooting	Ishmam and Jafrul did testing and troubleshooting.
Deployment	We together and made final discussion and deployed it using slack.

Troubleshooting



Solved:
pip install rasa –ignore-installed ruamel.yaml --user

Solved the raised another error and changed the device

Fig 7: Screenshot of some error during implementation of the project

Troubleshooting cont…

- Started coding into rasa version 3.0 then migrate the version and change the coding format because get to know rasa 3.0 don’t support deployment using rasa which is our future work.

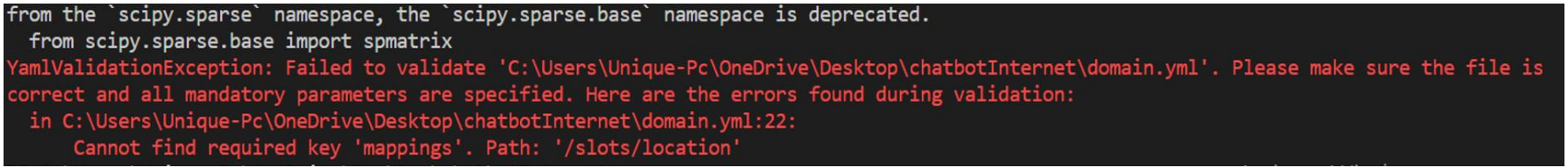


Fig 8: Domain.yml file error

Challenges, Limitation and Future Work

- Challenging task to make the project workable as there has many dependencies between python and rasa.
- Need very clear and sophisticated data and story to satisfy user
- Future work, Improve the story and increase the dataset
- Future work, make the application more dynamic and interactive with human.

References

[1]https://www.lsbu.ac.uk/study/course-finder?collection=LSBU_Courses_Meta&query=!nullsearch&start_rank=1&sort=relevance&f.Level_new|courseLevel=undergraduate

[2] https://rasa.com/docs/rasa/2.x/arch-overview

[3] https://norvig.com/spell-correct.html

[4] https://rasa.com/blog/unpacking-the-ted-policy-in-rasa-open-source/

Thank You