COSC2669/COSC3186 - Case Studies in Data Science

Group Work: *Milestone 1*

Group ID: *WIL Project 41*

Contribution Sheet

For each of member in your team, please write your name, student number, contribution percentage to the assignment and your signature. Marks are awarded to the individual team members, according to the contributions made towards the final work. Please submit what percentage each member made to this assignment and submit this sheet in your submission.

The contributions of your group should add up to 100%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Student ID** | **Contribution Percentage** | **Specific Work on Assignment** | **Signature** |
| Patricia Celon | s4119966 | 1/6 = 16.67% | Project Idea & Aim Statement. Documentation and Editing |  |
| Natchaya Chanbanchee | s4134044 | 1/6 = 16.67% | Review and reproduce Walert |  |
| Justin Nguyen | s3943743 | 1/6 = 16.67% | Milestone 1 Template and Draft |  |
| Khwar Nyo Thinn | s4133194 | 1/6 = 16.67% | Appendix & Walert Reproducibility |  |
| Brandon Kok Zheng Tan | s4144857 | 1/6 = 16.67% | Review of report against Structure/Marking Criteria |  |
| Joseph Khai Zhe Lee | s4057112 | 1/6 = 16.67% | Reproduce Walert and explore deeper about it |  |

**Note: Mark calculation according to contributions.**

Assume the mark for the presentation is m,

If the contribution percentages are even (assuming there are n members, 1/n contribution for each member is considered as even), then each member receives m mark;

Otherwise, if the contribution is not even:

Students with the **highest** contribution (e.g., X%) receives m marks;

For other students, e.g., let another Student who has contribution Y% (Y<X), his achieved mark will be calculated as (Y/X)\*m

PROJECT STATUS REPORT

**PROJECT STATUS REPORT**

*WIL Project Milestone 1*

## Group ID

WIL Project 41

## Project Aim Statement

As an RMIT Student I want to use a chatbot to quickly get answers from RMIT Connect services So that I do not have to wait long for emails or spend time searching for information

## Table of Group Members

|  |  |  |
| --- | --- | --- |
| Student ID | Name | Skills/Strengths |
| s4119966 | Patricia Celon | Data Visualization  (Python, R, Power BI), Data Scraping, EDA, Cloud Computing (Azure, AWS), SQL (T-SQL, Spark SQL) |
| s4134044 | Natchaya Chanbanchee | EDA, Data Visualisation (Python, R, Power BI), SQL |
| s3943743 | Justin Nguyen | Machine learning (Python – scikit-learn, NLP, embeddings, classification), Data visualisation (Python – matplotlib, seaboard; R – ggplot2, plotly), SQL (Azure, T-SQL) |
| s4133194 | Khwar Nyo Thinn | - Data Wrangling and cleaning  - Data Visualization |
| s4144857 | Brandon Kok Zheng Tan | Data Visualisation/Analysis (Python, R,) |
| s4057112 | Joseph Khai Zhe Lee | SQL, Machine learning (Python), Data visualisation(Python), JavaScript |

## Current Progress

* Held initial team meetings (scheduled recurring time: Tuesdays 9am)
  + Drafted team availability schedule for coordination
* Agreed on potential project directions (e.g., RMIT Connect Chatbot, Healthcare/Mental Health)
  + Started exploring how to get data from RMIT Connect Chatbot
* Conducted a literature review of RAG approaches, focusing on the Walert framework
  + Explored Github and tried reproducing Walert
* Assigned preliminary tasks and shared OneDrive folder access for collaboration

## Next Steps (3 Week Plan)

**Week 1**

* Finalize project domain selection
* Conduct reproducibility study of Walert RAG pipeline

**Week 2**

* Collect and clean preliminary dataset(s) for chosen project
* Distribute roles for EDA

**Week 3**

* Begin EDA and summarise findings
* Draft evaluation framework for measuring chatbot performance (effectiveness, coverage, correctness).
  + Tested the existing VAL chatbot with sample queries (e.g., course prerequisites) and confirmed its limitations.

## Issues or Risks

* Dataset availability for chosen domain
* Access limitations for scraping RMIT website content
  + RMIT website uses dynamic content (via JavaScript), which makes automated scraping more complex.
  + If we can’t extract FAQs, **resort to the Walert Datase**t, and combine with partially scrapped data to expand dataset
* Team scheduling conflicts (different availability)
  + Using Microsoft teams to update group with tasks and notes
* Technical complexity of RAG pipeline setup
  + Could be fixed by beginning with small-scale Walert replication before scaling.

**Declaration of AI Use:** All core content (project aim, steps, reproduction of walert study) was completed solely by group members.

*- Content below this line does not count for the page limit -*

## References

[1] D. Spina. 2024. Walert: Your Open Day FAQ Buddy. Retrieved August 22, 2025 from <https://www.damianospina.com/project/walert/>

[2] RMIT IR Group. 2024. Walert: Code and data for the Walert large language model-based chatbot. Retrieved August 22, 2025 from <https://github.com/rmit-ir/walert>

## Appendices - Walert Reproduction

**Progress**

* Enviromental setup:
* Indexing: Built the BM25 index on the FAQ dataset, all passages were handled without error
* Evaluation: Ran eval.py script on both known and inferred topics using qrels.txt and target/runs/. Outputs were stored in target/trec\_eval\_results/ and combined into a full report text file.
* Baseline reproduction: Obtained consistent NDCG scores that matched the expected Walert baselines for intent-based, dense FAISS and BM25 models.

**Appendix A – Reproducibility of Walert**

**Environmental Setup**

- OS: Mac OS

- IDE: Visual Studio Code

- Environment: conda create –n walert python=3.10

- Installed dependencies with pip install –r requirements.txt

**Indexing**

./index-bm25.sh

**Evaluation on Known Topics:**

python src/retrieval/eval.py known data/qrels.txt \

target/runs/rag-bm25.txt \

target/runs/rag-dense-faiss.txt \

target/runs/walert-intent.txt\

>> target/trec\_eval\_results/full\_report.txt

**Evaluation on Inferred Topics:**

python src/retrieval/eval.py inferred data/qrels.txt \

target/runs/walert-intent.txt \

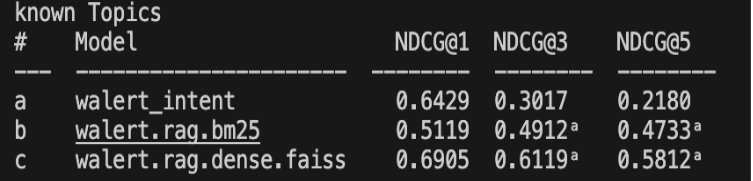
target/runs/rag-bm25.txt \

target/runs/rag-dense-faiss.txt \

> target/trec\_eval\_results/full\_report.txt

**Preliminary Results**

The reproduced evaluation results for known topics:



The reproduced evaluation results for inferred topics:

