IN2901 - Software Development Project

Project Proposal

Department of Information Technology

Faculty of Information Technology

University of Moratuwa

2023

<table>
<tr><td>Group Name</td><td colspan="2">Divergent</td></tr>
<tr><td>Project Name</td><td colspan="2">Name
(Natural Language Based Data Visualization Tool)</td></tr>
<tr><td>Client’s name and address</td><td colspan="2"> 
105, Bauddhaloka Mawatha,
Colombo 4</td></tr>
<tr><td rowspan="6">Group Members</td><td>Index No</td><td>Name</td></tr>
<tr><td>214139D</td><td>Neeshan M.I</td></tr>
<tr><td>214177P</td><td>Sahla A.S.F</td></tr>
<tr><td>214076F</td><td>Hamdha N.A</td></tr>
<tr><td>204235J</td><td>Wijethunge L.S.K</td></tr>
<tr><td>194079M</td><td>Kavishka S.H.D</td></tr>
<tr><td>Supervisors’ names</td><td colspan="2">Dr. Supunmali Ahangama</td></tr>
</table>

Contents

1. Introduction

Problem: Today, companies have lots of data. However, without the expertise in database querying and technical analysis, many stakeholders find it hard to understand this data as the traditional tools are tricky to use, especially if you're not a tech expert.

Importance: It's essential for everyone in a company to easily access and understand their data. Making good decisions relies on this. If data is hard to access, analyze and visualize, companies might miss out on organizational growth, strategies and success.

Technology: Our solution will harness the capabilities of Natural Language Processing (NLP) through OpenAI's API, dynamic data visualization tools, and user-friendly web technologies to bridge this gap. Furthermore, secure and efficient backend technologies like Python will be utilized to ensure quick data retrieval and processing.

Proposed Solution: Name is a robust and innovative data visualization tool designed to meet this challenge head-on. It allows users to effortlessly pose questions in natural language and, in return, presents visualized data-driven answers. By leveraging NLP, OpenAI's capabilities, and cutting-edge visualization techniques, Name promises an intuitive, efficient, and user-friendly experience, transforming the way stakeholders engage with data.

2. Problem in Brief

Problem:

The issue addressed by the Name tool is rooted in the complexity and inaccessibility of current data exploration systems. As data continues to grow rapidly in the modern business world, there's a pressing need for easy-to-use tools to understand this data. However, most tools available today demand a good understanding of technical languages like SQL, making them unsuitable for non-tech users. This disconnect between intricate data systems and everyday users has created a wide gap, where valuable insights remain locked behind a barrier of technical complexity.

Importance:

- **Ease of Access:** Simplifying data access means that more people, regardless of their technical skills, can understand and use data. This can lead to a broader participation in data-driven decision-making processes.
- **Decision Making:** With more intuitive tools, organizations can make better decisions. This is because clearer data visualizations help people spot trends, patterns, and anomalies more easily.
- **Empowerment:** When non-tech users can access and interpret data without always relying on IT departments or data specialists, it empowers them to be more proactive in their roles.
- **Efficiency and Time-Saving:** A tool that translates complex data into understandable visuals can save a lot of time and effort that might otherwise be spent in lengthy explanations or training sessions. It streamlines the data interpretation process. Users just input data and pose a question, significantly reducing the time traditionally spent on selecting and crafting visualizations.

In summary, the Name tool seeks to address the longstanding challenge in the data industry by offering a platform that makes data interpretation as simple as asking a question. By reducing the technical barriers, it promises a more inclusive, informed, and efficient approach to data management and understanding.

3. Aim and Objective

Aim: Develop a product that can visualize data from a database when users ask Questions.

Objectives:

• To create a user-friendly interface allowing secure login and query input.

• To integrate OpenAI's API for advanced natural language processing capabilities.

• To connect seamlessly with major databases, ensuring quick data retrieval.

• To design a data processing module that translates NLP results into meaningful visualizations.

• To implement feedback and continuous learning mechanisms for system improvement.

4. Proposed Solution

The proposed solution is a Natural Language Based Data Visualization Tool designed to enable users to seamlessly visualize data from databases by asking questions in natural language. This solution encompasses various key components and technologies to deliver a robust and user-friendly experience.

4.1. Technology to be Adapted

Login: Utilize OAuth 2.0 or OpenID Connect for secure social login, integrating SDKs and APIs.

Frontend Development: Create a responsive React-based web app for user interactions and data presentation, communicating with the Python backend.

Backend Development: Python serves as the core, equipped with libraries like SQLAlchemy for robust database interaction.

Database: Seamlessly connect with various databases and employ the Langchain NLP engine for enhanced data processing and user queries. Also usage of an internal database to store user information, system configurations etc.

Tables: Utilize HTML, CSS, and JavaScript (React) for tabular data presentation with sorting, searching, and pagination. Python manages data handling for efficiency.

Dynamic Data Visualization: Implement JavaScript libraries (D3.js, Chart.js, Highcharts), data-driven chart selection, and customizable visualizations using HTML, CSS, and JavaScript frameworks like React or Angular.

Feedback: Collect user feedback with HTML and JavaScript input forms and process it using Python for continuous learning.

Hosting: Backend hosted on Choreo, and frontend deployed on Choreo or compatible cloud platforms.

4.2. Nature of Solution

Input:

- Login information
- Database connection information
- User prompt.
- Feedback.

Process:

- NLP for user questions.
- Python connects databases.
- Dynamic data processing and visualization.
- Automatic chart selection.
- User customization.
- Continuous learning.

Output:

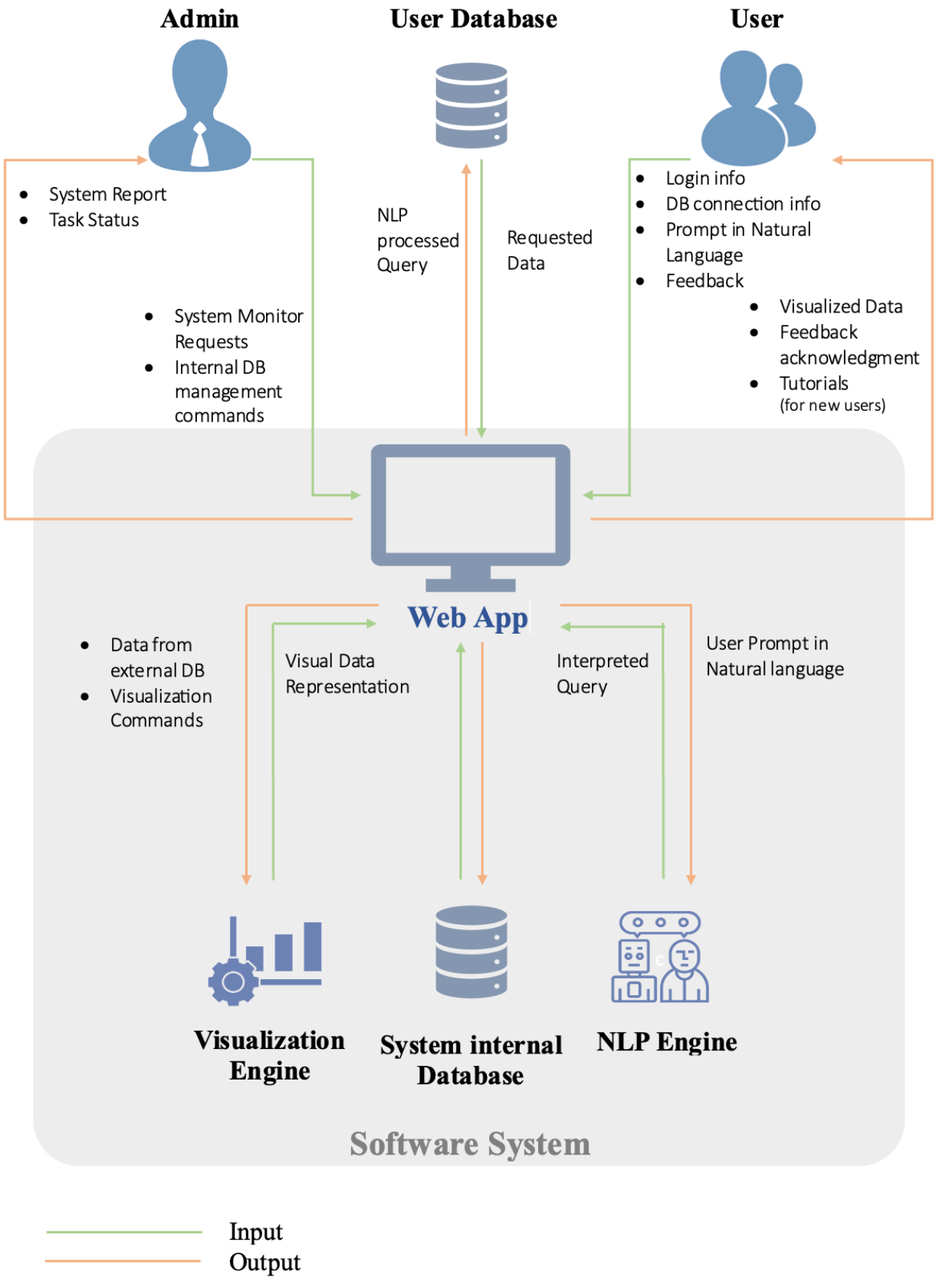
- Data visualizations.
- Customized charts.
- User feedback analysis.
- Optional tutorials.

Users:

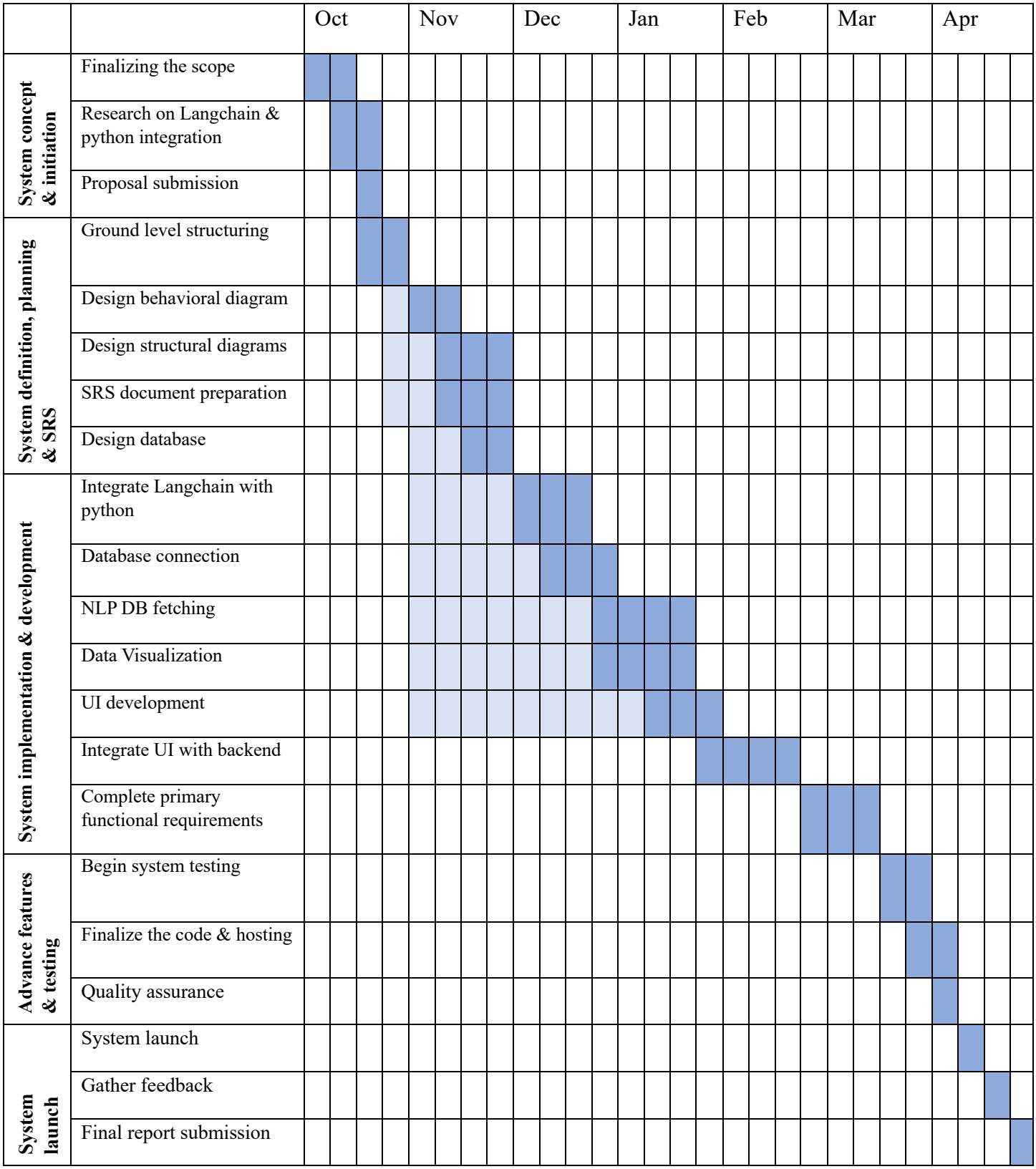
- End users.
- Admins.
- Developers.
- First-time users.

This is a challenging project because of the involvement of A.I such as NLP (Natural Learning Process) but as a team of five, one of our members will be solely concentrating on the integration of A.I into this project while others work on the rest of the technology. We hope to be able to do this project within time by putting great effort while educating ourselves along this journey. With the help of the company WSO2 and its resources and the assistance of their members, we hope this project could be done perfectly.

4.3. Conceptual overview of the proposed solution



6.Timeline


Oct
Nov
Dec
Jan
Feb
Mar
Apr
System concept & initiation
Finalizing the scope
Research on Langchain & python integration
Proposal submission
System definition, planning & SRS
Ground level structuring
Design behavioral diagram
Design structural diagrams
SRS document preparation
Design database
System implementation & development
Integrate Langchain with python
Database connection
NLP DB fetching
Data Visualization
UI development
Integrate UI with backend
Complete primary functional requirements
Advance features & testing
Begin system testing
Finalize the code & hosting
Quality assurance
System launch
System launch
Gather feedback
Final report submission

6. Signature of the Group Members

Index No	Name	Signature
214139D	Neeshan M.I	
214177P	Sahla A.S.F	
214076F	Hamdha N.A	
204235J	Wijethunge L.S.K	
194079M	Kavishka S.H.D	

7. Supervisors' declaration

I hereby declare that I have checked this project and, in my opinion, this project is adequate in terms of scope and quality.

Name of Supervisor:

Designation:

Date:

Signature:

Any further comments: