



**COMSATS University Islamabad,
Abbottabad Campus**

Project Proposal

(SCOPE DOCUMENT)

for

**Urdu Captioning For Visually Impaired
People of Pakistan**

Version 1.0

By

Student Name 1 CIIT/SP09-BCS-xxx/ATD

Student Name 2 CIIT/SP09-BCS-xxx/ATD

Sardar Badar Saghir CIIT/SP19-BCS-022/ATD

Supervisor

Sir Faize Ali Shah

***Bachelor of Science in Computer Science
(2019-2023)***

No.	Comment	Action

SCOPE DOCUMENT REVSION HISTORY

Supervisor Signature

Date:

Table of Contents

Abstract.....	5
Introduction.....	6
Problem Statement.....	6
Problem Solution for Proposed System.....	6
Related System Analysis/Literature Review.....	6
1. Be My Eye.....	7
2. Sullivan+.....	7
3. Lookout Assisted Vision.....	7
Advantages/Benefits of Proposed System.....	8
Scope.....	8
Modules.....	8
1. Module 1: Data Collection.....	9
2. Module 2: Data Prepossessing.....	9
3. Module 3: Features Extraction.....	9
4. Module 4: Model Training.....	9
5. Module 5: Model Testing.....	9
6. Module 6: Model Deployment.....	9
7. Module 7: Android App Development.....	9
8. Module 8: Software Testing.....	9
System Limitations/Constraints.....	9
Software Process Methodology.....	10
Tools and Technologies.....	10
Project Stakeholders and Roles.....	11
Team Members Individual Tasks/Work Division.....	12
Data Gathering Approach.....	12
Concepts.....	12
Machine Learning.....	12
Computer Vision.....	12
Data Science Methodology.....	13
Cross platform development.....	13
Gantt chart.....	13

Mockups.....	14
Conclusion.....	15
References.....	15
Plagiarism Report.....	15

Project Category: (H, C, E)

☐ A-Desktop Application/Information System ☐ B-Web Application/Web Application based Information System
☐ C- Problem Solving and Artificial Intelligence ☐ D-Simulation and Modelling ☐ E- Smartphone Application ☐
☐ F- Smartphone Game ☐ G- Networks ☐ H- Image Processing ☐ Other (specify category)

Abstract

In Pakistan, people with visual impairment face many hurdles in performing day to day to task. They must have to estimate the state of the object. The issue can be addressed by using a computer vision, but the lack of Urdu images captions dataset makes it much harder. The project aims to manually develop the image dataset with Urdu captioning and then create a model on it. The dataset focuses on some of the common scenarios faced by the visually impaired people. After the dataset is created, a convolutional neural network (CNN) is used to extract image features. These features are used in a recurrent neural network (RNN) to generate a valid Urdu caption. The model is used with the android application to develop a computer vision app that can create and read Urdu captions at real time for visually impaired peoples.

Introduction

The people with visual impairment face difficulties in estimating the state of the object. Thus such a system is needed to that create the caption by analysing the state object in real time. The advancement in computer vision makes it possible to develop software's that can detect the objects in real time. The most of algorithm requires great amount of data for high accuracy result. The datasets that have been created in the past are not particularly focus on the population of Pakistan. The lack of Urdu dataset is one of the core problems for creating the system that can help the visual impaired people. Thus, project design to create images datasets with Urdu captions dealing with common scenario face by visually impaired people and than develop a app that can visualize the states of object in real time by creating caption's using computer vision models. Then user is inform about the caption through TTS engines by converting text into speech.

Problem Statement

According to research in 2017 that about 2.21 million people in Pakistan either are blind or have sever visual impairment [1]. The aim of project is to design a mobile application that help the visually impaired people of Pakistan in their routine tasks. The android app like "Be My Eye" [2], or "Sullivan"[3] try to solve this problem by either hiring volunteer which can be expensive for end user or by training models that do not support Urdu language. Thus this project aim is to fulfilled the needs of impaired and blind users of Pakistan to develop such system. During the development of the system it is highly likely that we will encounter the technologies that required the skills related to machine learning and computer vision. Thus it is expected that we will polish our skills related to convolutional neural network (CNN) and recurrent neural network (RNN). In the end of the project, it is anticipated that we will learn technique related to data gathering, feature extraction and training the model on cloud platforms.

Problem Solution for Proposed System

The project is design to create images datasets with Urdu captions dealing with common scenarios face by visually impaired people and than develop a app that can visualize the states of object in real time by creating caption's using computer vision models. Then user is inform about the caption through TTS engines by converting text into speech. It provide the simple way for visually impaired people to interact with their surroundings. The simple UI design with the real time caption with voice, help the visually impaired to identified state of objects near him. The project not only help the blind people but it will also create the datasets which provide the basis for Urdu image datasets for the blind people which can be used in the future for other problem having similar characteristics.

Related System Analysis/Literature Review

Many systems has try to solve the problem of similar nature but they have some of the limitation that is trying to solve by this project

1. Be My Eye

The system is develop to support visually impaired people. The premises behind the app is to provide the platform where call can be arrange between visually impaired and volunteers who want to assist [4]. The only issue with this app is that vastly dependent upon availability of volunteers thus it is highly likely that volunteer will be unavailable when user is looking for the assistant[5].

2. Sullivan+

The app is available in both android and IOS users that can help in visual details. The premise of app to recognize the face, text, image description and colour [6]. The issue is that, app has complex UI that might be useful for that low vision person but can not work with blind persons. It lack the voice feature thus not useful in real life for scenarios for blind peoples. The another issue is that its lack the support of the Urdu language thus not good for illiterate peoples of Pakistan.

3. Lookout Assisted Vision

This app can recognize simple document in English and try to identify plants. But this app is not useful for common use as it does not deal with the common scenarios like identifying the objects and the situation context. Thus issue has been tried to encounter in our project.

Table 1 Related System Analysis with proposed project solution

Application Name	Weakness	Proposed Project Solution
Be My Eye	The issue with this app is that vastly dependent upon availability of volunteers and has no advantage has been taken of computer vision.	Introduce computer vision for explaining scenario to eliminate the need of volunteers.
Sullivan+	The issue is that, app has complex User Interface for blind persons. It lack the voice feature. The another issue is that its lack the	Our aim is to create user friendly by that can be used by blind people's. The Urdu language will be their for

Lookout Assisted Vision	support of the Urdu language. App is not useful for common scenarios due to limited capability of the object detection	Pakistan Peoples. Create general purpose data set with support of common house hold scenario, currency, vehicle , and common sign recognition
--------------------------------	-------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------

Advantages/Benefits of Proposed System

The system offers certain benefit's that include

1. It covers the more general scenario as VizWiz dataset is use for reference.
2. It support Urdu Language.
3. The real time Urdu captioning creates the opportunity to target the untouched market.
4. The application is scaleable which can be extended to further languages.
5. Integrated Pakistani currency identification for impaired people of Pakistan is unique to this application.

Scope

The main function of app is to detect and create Urdu caption for Vehicles, Furniture, Appliances, Persons, Bench's, Street Sign's, Stop Sign's, Kitchen object's and Currency of Pakistan. The model dataset is limited to above scenarios. The captions of image will be converted to voice with the help of TTS engines. By studying datasets like VizWiz it is estimated that number of image require for developing the project could be between 8k-31k. The

The project does not cover all scenario face by the visually impaired person due to constrains of time and resources. The object caption will be generated when user capture the image. It will be quite difficult to generate real time captions due technical issues.

The project is mainly dealing with both machine learning and data science aspects. The project will be in a pipeline instead of sequential with data collection, data preprocessing, feature extraction, machine learning and deployment can be run in parallel. The most appropriate software development for our project is agile methodology.

Modules

The project can be divided into Data Collection, Data Preprocessing, Features Extraction, Model Training, Model Testing, Model Deployment And Android App Development.

1. Module 1: Data Collection

The raw data is collected by capturing manual images or extracting relevant images from the internet. Besides that, relevant images will also be extracted from publicly available datasets.

2. Module 2: Data Preprocessing

The raw data is then processed into relevant categories, and images that are not relevant to the required data set will be removed and then images will be labeled with relevant captions. The number of captions can vary depending upon the image.

3. Module 3: Features Extraction

Features from images are extracted using a convolutional neural network (CNN), so that we can use those features to train the model.

4. Module 4: Model Training

The algorithm like Long Short Memory recurrent neural network (LSTM-RNN) will be used for training the model from features extracted during the features extraction process.

5. Module 5: Model Testing

The testing data is used for testing the model. In this process, the model is tested with both development and deployment conditions. If the accuracy is less than the required amount, the model will be trained again.

6. Module 6: Model Deployment

The model will be deployed on the cloud services like Azure or Amazon cloud services. APIs for the respective cloud service will be used to communicate with the applications and model.

7. Module 7: Android App Development

The project will be packaged into a cross-platform app that has basic permissions of internet and camera. The app will utilize the model and then create the captions on the basis of model prediction and finally the TTS engine will convert it into Urdu voice.

8. Module 8: Software Testing

The final module will be the testing of both app and model from different users to get an unbiased result. And in this phase, bugs that will be found will be fixed. And performance will be optimized in the end of the testing phase.

System Limitations/Constraints

The app is limited to data set used is created by group thus it can not cover every aspect of visual impaired people. Train model accuracy is uncertain. The application will not deal at real time the model require processing on the cloud. The application is internet dependent and can not work at offline mode.

Software Process Methodology

The Procedural methodology is adopted for training model and object oriented approach for developing the application. Agile model is selected for the project. As individual interaction is more important for us than the process. Flexibility and adoption to change make it more relevant for our projects.

Tools and Technologies

Table 2Tools and Technologies for Proposed Project

Tools And Technologies	Tools	Version	Rationale
	Google Collab	2.x	Online Editor
	Android Studio	2022	IDE
	Figma		Design Work
	MS Word	2015	Documentation
	MS Power Point	2015	Presentation
	Pencil	2.0.5	Mock ups Creation
	Technology	Version	Rationale
	Python	6.0	Programming language
	Java	2013	Programming language
	Tensor Flow	5	Library

Project Stakeholders and Roles

Table 3Project Stakeholders for Proposed Project

Project Sponsor	COMSATS University, Islamabad, Abbottabad
Stakeholder	<p>The stack holder in our project</p> <ul style="list-style-type: none">• Sardar Badar Saghir• Muhammad Shaid Ali• Irtisam Ali• Project Supervisor Name: Sir Faize Ali Shah• Final Year Project Committee: Evaluation of project

Team Members Individual Tasks/Work Division

Table 4Team Member Work Division for Proposed Project

Student Name	Student Registration Number	Responsibility/ Modules
Sardar Badar Saghir	SP19-BCS-022	Module 1, module 2 and module 3 and module 5. Also Responsible for Grantt
Muhammad Shaid Ali		He is responsible for module 1, module 2, and module 6 and module 8. Also responsible for UML Digrams
Irtisam Ali		Module 1, Module 2 , Module 4 and Module 7. Also responsible making Mockups diagrams

Data Gathering Approach

Data is collected by Interview, Questionnaire, scrapping from online resources and by modifying data set resources.

Concepts

- **Machine Learning**

The project is base upon the machine learning concepts. Thus we will have to deal lot concept of machine learning on developing the model.

- **Computer Vision**

Our Machine Learning require the knowledge of computer vision domain, as our datasets are base upon the images.

- **Data Science Methodology**

As dataset are build from scratch and thus it require work, it is require for us to apply various data science to create the quality of dataset.

- **Cross platform development**

Final Product will in the form of both IOS And Android thus it will help us in learning cross platform development

Gantt chart

Create the Grant Chart and provide estimated start and end dates of all proposed modules/tasks for each team member. Also identify the dependencies (which tasks cannot be started/completed, until the dependent task is completed). Gantt chart can be created using MS Project.

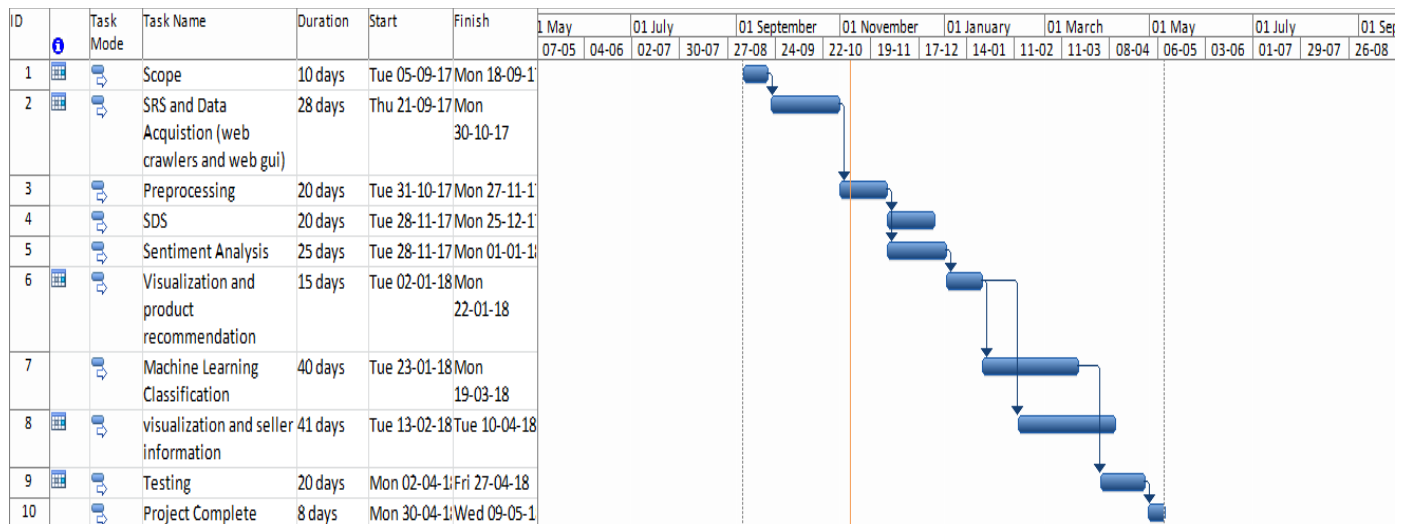
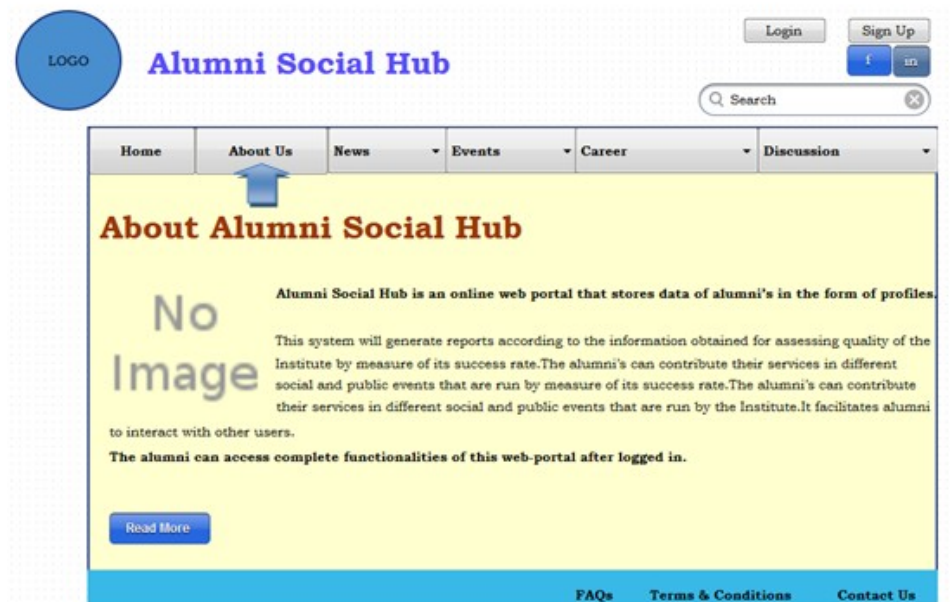
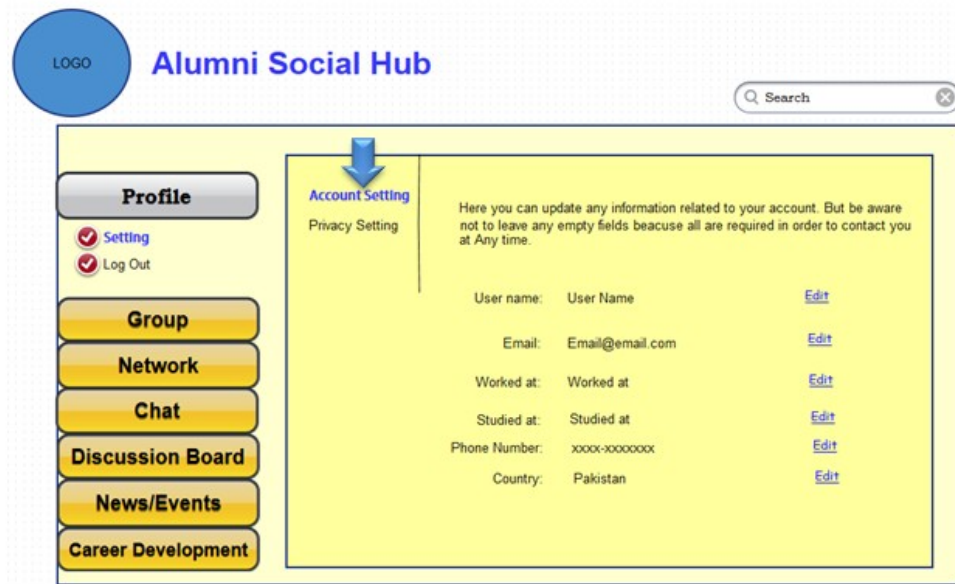


Figure 1Sample Gantt chart

Mockups

Insert minimum mockups (Usually 4-6 mockups) which show the major modules mentioned in the scope section of the document. Do not include mockups for Login, Signup, Forgot Password, Contact Us, About Us etc. If the project is a Web or a Smartphone Application, then include at-least three mockups from each part of the project. You can design mockup in any design tool for example pencil tool (<https://pencil.evolus.vn/>) or Balsamiq (<https://balsamiq.com/>)





Conclusion

Project is to creating a solution that can help the blind people of Pakistan in daily life . In future this datasets can be extended, thus making it possible to improve application accuracy and deal the scenario. The project is one of its kind that is dealing with the Urdu speaking population of Pakistan and addressing problem of visual impairment.

References

- [1] "As of 2017, out of 207.7 million people in Pakistan, an estimated 1.12 million (95% Uncertainty Interval [UI] 1.07–1.19)."
 - [2] "Be My Eyes - See the world together." <https://www.bemyeyes.com/> (accessed Mar. 16, 2022).
 - [3] "Sullivan plus for visually impaired and low vision | Blind Help Project." Accessed: Mar. 16, 2022. [Online]. Available: <https://blindhelp.net/software/sullivan-plus-visually-impaired-and-low-vision>
 - [4] "Our story." <https://www.bemyeyes.com/about> (accessed Mar. 16, 2022).
- [1] "As of 2017, out of 207.7 million people in Pakistan, an estimated 1.12 million (95% Uncertainty Interval [UI] 1.07–1.19)."
 - [2] "Be My Eyes - See the world together." <https://www.bemyeyes.com/> (accessed Mar. 16, 2022).
 - [3] "Sullivan plus for visually impaired and low vision | Blind Help Project." Accessed: Mar. 16, 2022. [Online]. Available: <https://blindhelp.net/software/sullivan-plus-visually-impaired-and-low-vision>
 - [4] "Our story." <https://www.bemyeyes.com/about> (accessed Mar. 16, 2022).
 - [5] "Our story." <https://www.bemyeyes.com/about> (accessed Mar. 16, 2022).

- [6] “Sullivan+ (blind, low vision) - Apps on Google Play.”
<https://play.google.com/store/apps/details?id=tuat.kr.sullivan&hl=en&gl=PK> (accessed Mar. 16, 2022).

Plagiarism Report

Attach the Plagiarism report of your project scope document from library staff of turnitin tool (<http://turnitin.com>)