KANTIPUR ENGINEERING COLLEGE

(Affiliated to Tribhuvan University)

Dhapakhel, Lalitpur

NEPALI VOICE ASSISTANT

Submitted by:

Ankit Kafle   [KAN076BCT012]

Dikshyant Giri [KAN076BCT026]

Jenith Rajlawat [KAN076BCT034]

Nawaraj Shah [KAN076BCT044]

A MAJOR PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR IN COMPUTER ENGINEERING

Submitted to:

Department of Computer and Electronics Engineering

June, 2023

NEPALI VOICE ASSISTANT

Submitted by:

Ankit Kafle   [KAN076BCT012]

Dikshyant Giri [KAN076BCT026]

Jenith Rajlawat [KAN076BCT034]

Nawaraj Shah [KAN076BCT044]

A MAJOR PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR IN COMPUTER ENGINEERING

Submitted to:

Department of Computer and Electronics Engineering

Kantipur Engineering College

Dhapakhel, Lalitpur

June, 2023

ABSTRACT

With the advancement in Machine Learning models, machines are being made capable of doing tasks that otherwise require human intelligence. Artificial Intelligence (AI) technology has paved the way for the development of AI voice assistants, transforming the way humans interact with computers and digital devices. In today's world, Automatic Speech Recognition(ASR) can be an ideal thing to do with the development of technology. So our proposed system Nepali Voice Assistant is a voice assistant which recognizes Nepali speech, processes it and then executes the required operation. Our system explores the development of a Nepali AI voice assistant using a combination of Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), Connectionist Temporal Classification (CTC) network, and Mel-frequency cepstral coefficients (MFCC) feature extraction. The feature extraction process begins with the extraction of MFCCs from the audio input, which captures the spectral characteristics of the Nepali speech. The CNN component processes the MFCCs, employing convolutional layers to extract local patterns and spatial features. The GRU network, a type of recurrent neural network (RNN), is then employed to model the temporal dependencies within the speech data. To enable accurate transcription and understanding of user queries, the CTC network is utilized. The CTC network, trained on labeled speech data, performs sequence-to-sequence mapping, aligning the input speech with the corresponding text outputs. This aids in speech recognition and enables the voice assistant to convert spoken Nepali commands into textual representations. The combination of these neural network models, including CNN for feature extraction, GRU for temporal modeling, and CTC for transcription, allows the voice assistant to understand and respond to user commands accurately.

Keywords *− Artificial Intelligence, Automatic Speech Recognition, Convolutional Neural Networks, Gated Recurrent Units,Connectionist Temporal Classification, Mel-frequency cepstral coefficients, Recurrent Neural Network.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI:Artificial Intelligence

ASR:Automatic Speech Recognition

CNN:Convolutional Neural Network

CTC: Connectionist Temporal Classification

GRU: Gated Recurrent Units

MFCCs: Mel-frequency cepstral coefficients

RNN:Recurrent Neural Network

CHAPTER 1
INTRODUCTION

1.1 Background

Speaking and writing are the two important things that help us to communicate among us. Deficient either writing or speaking affects our daily activities. Most of the people in rural areas are able to speak properly but not able to write properly. Most communication technology (gadgets, mobiles, computers, etc) needs text as an input for their operation. To make familiar with the technology Automatic Speech Recognition (ASR) can play a significant role[1]. Voice assistants have become an integral part of our daily lives, revolutionizing the way we interact with technology and access information. The concept of voice assistants traces back to the early days of AI research, but it wasn't until recent years that they truly gained widespread popularity and advanced functionality.

The need for a Nepali voice assistant arises from the growing importance of personalized and efficient interactions with technology for Nepali-speaking individuals. As technology becomes increasingly integrated into our daily lives, a voice assistant tailored to the Nepali language offers numerous advantages and benefits. It provides a natural and intuitive means of communication, allowing users to perform tasks, access information, and receive assistance seamlessly through voice commands in their native language. By understanding and responding to Nepali speech accurately, a dedicated voice assistant can empower individuals, simplify their daily routines, and foster a more connected digital experience that is specifically designed to meet their needs.

The project Nepali Voice Assistant is a groundbreaking development aimed at revolutionizing the way people interact with technology by providing an intelligent voice-controlled assistant specifically designed for Nepali-speaking individuals. This innovative system enables users to execute commands, access information, and perform tasks effortlessly through the spoken Nepali language. By leveraging advanced technologies such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Neural Networks, our Nepali Voice Assistant offers an intuitive and personalized

user experience. It is designed to capture the essence of the Nepali language, enabling natural and context-aware conversations that enhance user engagement and satisfaction.

1.2   Problem Statement

In today's time, communication plays a vital role in human development. We learn through communication but most of the people living in the rural areas of Nepal are only able to speak properly, when they have to write a command to a gadget through text they are unable to do it due to the unawareness of the technology. Also, existing voice assistants primarily support major languages, leaving a gap in providing personalized assistance for Nepali users. This limits their ability to access information, perform tasks using voice-based interactions. Therefore, there is a need to develop a user-friendly and accurate Nepali voice assistant that understands and responds to Nepali speech, enabling a more efficient digital experience for Nepali-speaking individuals.

1.3   Objectives

  I To develop an Automatic Speech Recognition system.
 II To enable the execution of various tasks and commands through voice inputs.
III To create a user-friendly interface for Nepali voice interactions.

1.4   Application Scope

The application scope of the Nepali Voice Assistant project is vast where seamless voice interactions can greatly benefit Nepali-speaking individuals. In todays world, the voice assistant can serve as a virtual companion, aiding users in managing their daily tasks, organizing schedules, and setting reminders, all through spoken commands in Nepali. Additionally, the assistant can facilitate hands-free communication, allowing users to make calls, send messages, and dictate emails or text messages without needing to manually operate their devices.

1.5 Features

1. Provides proper execution of commands.
2. Provides support for the Nepali Language.
3. Provides an effective user-friendly environment.

1.6 System Requirements

The system requirements for the project are as follows:

1.6.1 Development Requirments

1.6.1.1 Hardware Requirments

- PC with a minimum specification of 8GB RAM and a sixth generation i5 processor

1.6.1.2 Software Requirement(Minimum)

- Microsoft Windows 7/8/10 (32 or 64-bit) or Mac OS

1.6.2 Deployment Requirments

1.6.2.1 Hardware Requirments

- More than 1.5 GHz clock speed
- Minimum 4GB RAM

1.6.2.2 Software Requirement(Minimum)

- Visual Studio Code
- Pycharm

1.7 Feasibility Study

The feasibility study is one of the main important things to be considered for the project development. The feasibility study must be done for different factors affecting the project. Here are some factors whose feasibility study should be done for our project.

1.7.1 Economic Feasibility

Economic feasibility attempts to weigh costs of developing and implementing a new system, against the benefits that would accrue from having the new system in a place. This feasibility study gives the top management the economic justification for the new system. A simple economic analysis which gives the actual comparison of costs and benefits is much more meaningful meaning in this case. In addition, this proves to be a useful point of reference to compare actual costs as the project progresses. There could be various types of intangible benefits on account of automation. These could include increased customer satisfaction, improvement in product quality, better decision making, and timeliness of information, expediting activities, improved accuracy of operations, better documentation and record keeping, faster retrieval of information, better employee morale

1.7.2 Technical Feasibility

Since the proposed system uses software technologies and tools which are freely available and technical skills required can be tricky but are manageable. There are many free machine learning libraries available for data analysis and predictions with proper documentation and courses. The hardware system in the project need not be highly computing but requires normal computing and the system server must be adequate and 4 manageable in the future. so it is seen that the hardware and software meet the needs of the system. So it's clear that the proposed project is technically feasible.

1.7.3 Operational Feasibility

Though the advancement in technologies, any kind of system software or application is no longer hard to operate. Thus the user needs to be only a bit familiar with the software system backed with graphical explanations and that can be easily understood faster in time with usage. This system highly focuses on design-dependent parameters like reliability, maintainability, supportability, usability, predictability, sustainability, affordability, etc. So, the project is feasible in operation.

1.7.4 Schedule Feasibility

Schedule Feasibility is defined as the probability of a project to be completed within its scheduled time limits, by a planned due date. If a project has a high probability to be completed on-time, then its schedule feasibility is appraised as high. Schedule feasibility ensures that a project can be completed before the technology becomes unnecessary. Since there are many features in our project but can be implemented in a quality way it has a very high probability to be completed on time.
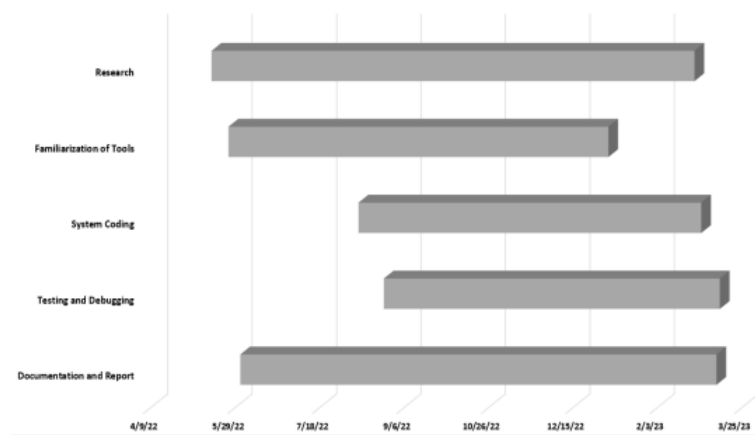

Figure 1.1: Gantt Chart

CHAPTER 2
LITERATURE REVIEW

2.1 Related Projects

- IOE SYLLABUS : IOE Syllabus is a website in which students can see the notices, faculty syllabus, faculty notes, etc. Also students can get collection of past questions from this website.

- Onlineocr.net : A user friendly and free online OCR converter tool service to extract text from picture and that is available free of cost at the internet, help in doing Optical character recognition online.

2.2 Related research

In recent years, the use of technology and the popularity of e-commerce have led to a significant increase in the number of online shopping websites available. While this makes it convenient for people to shop online, it can also be time-consuming and require a lot of effort to search for the best deals and offers on these websites. In order to address this issue, the authors of the paper "Exploiting Filtering approach with Web Scraping for Smart Online Shopping" have developed a web scraping framework that allows users to efficiently gather product information and deals from multiple ecommerce websites. The framework is implemented using front-end technologies such as HTML (Hypertext Markup Language) and CSS (Cascading Style Sheet) as well as a back-end language, PHP (Hypertext Preprocessor). It uses Python libraries and HTML tags to write the scraping scripts, and the results are displayed dynamically to the user rather than being stored in a local database. According to the authors, this framework has a high accuracy rate of 93time required. The goal of the framework is to make it easier for people to find the best deals and offers on e-commerce websites and save them time and effort in the process

The paper "Barcode Character Defect Detection Method Based on Tesseract-OCR" aims to address the increasing quality requirements for barcodes as their use becomes more widespread with the advancement of information technology. However, during 6

the printing process, various defects can occur, such as flying ink, missing print, wrong print, black spots, and improper registration, which can be caused by factors such as poor typography, inadequate printing equipment, and imperfect printing technology. The traditional method of manually sorting defective barcodes is inefficient and prone to error due to the influence of various factors, leading to low precision in detection. In order to address these issues, the authors propose a method for detecting defects in barcodes using Tesseract-OCR (Optical Character Recognition) software. This method involves segmenting the barcodes using horizontal projection, recognizing the characters in the barcodes using Tesseract-OCR, and applying the Levenshtein Distance algorithm to detect character defects. The authors conducted experiments using 1000 barcode images, and the results showed that the proposed method had an accuracy of 94.3detecting barcode defects.

The paper "NewsOne — an aggregation system for news using web scraping method" is an online platform that aggregates and summarizes the latest news updates from multiple national and international sources. It presents this information in a concise and easy-to-read format, and is designed to allow users to quickly and easily access the news without wasting time searching or waiting for content to load. To achieve this, NewsOne employs web scraping and crawling techniques to extract news content from various websites, which it then categorizes based on user interest. The platform is also service-oriented, allowing users to interact with each other from across the web. NewsOne uses a bot that dynamically extracts content from the stored URL's RSS feeds at set intervals, which are added to a database by the platform's administrators or sub-administrators. This model of categorization is designed to extract useful information for classifying news articles into specific categories, such as Just-In, Technology, Health, Science, Sports, Business and Economics, and Entertainment. NewsOne's user experience is flexible, allowing readers to choose the categories of news that interest them and read the news for free and as quickly as possible. In addition, users have access to up-to-the-minute daily news coverage and headlines from over 100 fully licensed and trusted news sources from around the world. Finally, the platform also provides recommendations and thoughts for future development [4].

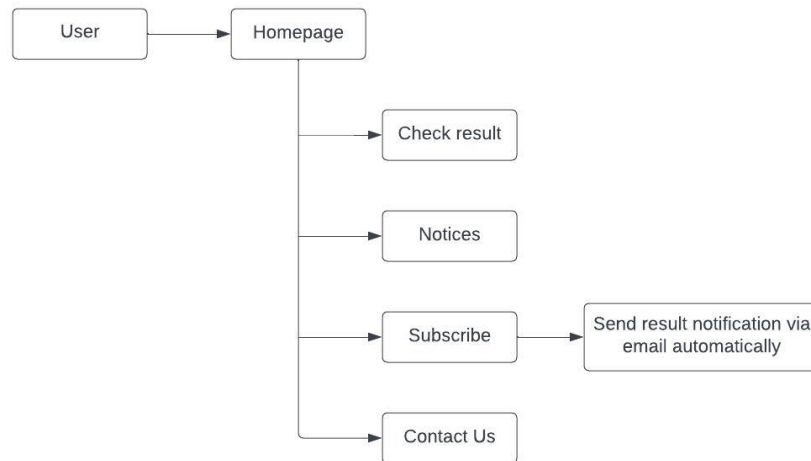# CHAPTER 3
METHODOLOGY

3.1 Working Mechanism

Figure 3.1: Block diagram of the system

1. Homepage: At first, the user interacts with the homepage of our system where they will be able to select options such as Check Result, Notices, Subscribe and Contact Us section.

2. View Notices: One major feature of the website is to show Notices published by IOE. This feature now allows the user to view notices in real time after the IOE has published it. This feature can be implemented by the use of web scraping in which the website exam.ioe.edu.np will be scraped for notices. To view notices it is not necessary that the user must be subscribed.

3. View Results: Viewing results immediately after the publication with a proper interface is a hard thing to do. So with the webpage, the user can verify the result and see if they have passed the exams. To use this feature, the user doesn't require to be subscribed in into the 8 system. But to use the feature of getting automatically notified via email regarding the result status, the user must be subscribed using the email provided by KEC. The webpage admin automatically sends activation email to the subscriber and only after the subscription is activated, then after the publishment of result, the system automatically sends them an email regarding if they have passed or not. This can be achieved by using web scraping

and OCR scan.

4. Subscribe: Viewing results immediately after the publication with a proper interface is a hard thing to do. So with the webpage, the user can verify the result and see if they have passed the exams. To use this feature, the user doesn't require to be logged in into the system. But to use the feature of getting automatically notified via email regarding the result status, the user must be logged in using the email provided by KEC. The webpage automatically verifies the user and sends them an email regarding if they have passed or not. This can be achieved by using web scraping and OCR scan.

5. Contact Us:If any problem arises then the user can directly contact the developer to solve the issue. The user must fill up the form where they have to insert their Name, Email, Subject and the message they want to send. The mail is now sent to the developers team email address

Web Scraping

Web scraping is the process of collecting structured web data in an automated fashion. It's also called web data extraction. Some of the main use cases of web scraping include price monitoring, price intelligence, news monitoring, lead generation, and market research among many others. In general, web data extraction is used by people and businesses who want to make use of the vast amount of publicly available web data to make smarter decisions. For our project we will use web scraping to extract the IOE notices and result from their website [5].

- Cropping the image: First , the image after the detection of the position of the license plate and the bounding box formulation is cropped according to the coordinates of the bounding ox so that we are resulted with the number plate from where the character extraction is to be done.
- Noise reduction and gray scaling: In order to male the image suitable for the character recognition, the image is filtered for the removal of noise and followed by the gray scaling of the cropped image so that the characters which are to be detected are separated from the background color. The conversion of image into

black and white that is, improving the contrast causes better performance of the model.

The processed image is then sent through the process of OCR where the characters that are present in the image are extracted . The extraction process or the character recognition process is done with the help of either pattern recognition or feature recognition.

Procedure for Web Scraping

1. Start
2. Create scraping template
3. Browse website
4. If the content is found, go to step 5, else repeat step 3
5. Get the link post
6. Explore the link post
7. Get the required data (Data extracted)
8. Stop

Flowchart of Web Scraping

VIPS BASED SEARCHING

1. Start
2. Construct parse tree for the HTML tags
3. Isolate all the anchor tags
4. Search for the result and notices inside the anchor tag
5. If result and notices are present, go to step 6, else go to step 7
6. Extract the content of anchor tag
7. If the entire webpage is scraped, go to step 8, else go to step 4
8. Stop

Flowchart based on VIPS

Implementation of Web Scraping

1. At first, the website of IOE is visited. After this, the site is explored.

2. The HTML content of the link is explored up until 10 pages initially to gather data into the database.

3. Now the entire webpage is converted into a parse tree through the use of an HTML parser. As from the inspection of the webpage, we found that the data of the notices and results along with the date of publishment and the link of the PDF file for the notices and results were stored by forming a table. The publishing date of the results and notices were stored in the third data cell for every table row. Also, every table row was considered for the extraction of the PDF file and the title for the notice and result. Only the notices and results of BE were taken in this stage.

4. Following the above step, the entire table rows of the table were explored and the title of the notices and results, which were present in the content of the `<span>` tag, was extracted. Also, the link for the PDF was present in the `<href>` part of the anchor tag, which was also extracted.

5. On observation of the actual PDF link for the notices and results, we encountered a problem in which the `%20` part of the link was regarded as space which caused the actual link of the PDF to break. This was easily solved by replacing the space with `%20` and joining the starting of the URL `"https://exam.ioe.edu.np"` with the link from the `<href>` tag.

6. Now the filtered title obtained from the content of `<span>` from step 4, the date of the PDF obtained from step 3, along with the link of the PDF file obtained at the end of step 5, was inserted into the built-in SQLite database with a constraint that if the link of the notices and results were already present in the top of the database, then the insertion process would stop.

Optical Character Recognition(OCR)

OCR also referred as text scan is the process that extracts and repurposes data from scanned documents, camera images and image-only pdfs. OCR systems use a combination of hardware and software to convert physical, printed documents into machine-readable text. For our project we will use OCR to extract the symbol number as a string and then we will compare the symbol number that has been obtained with the help of OCR scan to the symbol number of the students. If the symbol number 12 matches, then the backend of our project will send an email to the registered users automatically and then the students can know whether they have passed or not.

In our project the OCR scanning will be performed using a library in python.

Implementation Of OCR

1. First, extract the PDF file's link from the database.
2. Using the link from step 1, extract the PDF file.
3. Convert the extracted PDF to an image format.
4. Scan the characters from the image and save the resulting string in the database.

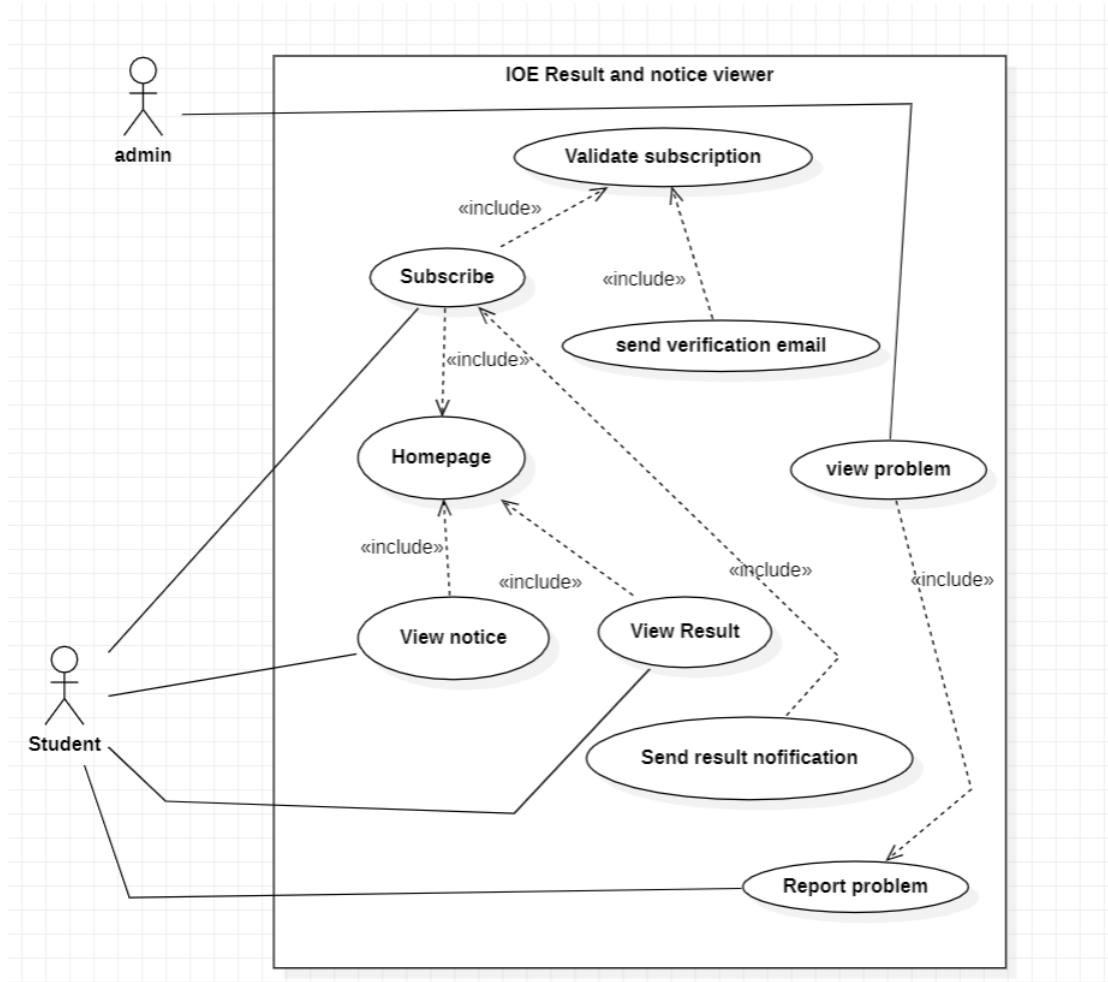## 3.2 UML Diagrams

Use Case Diagram

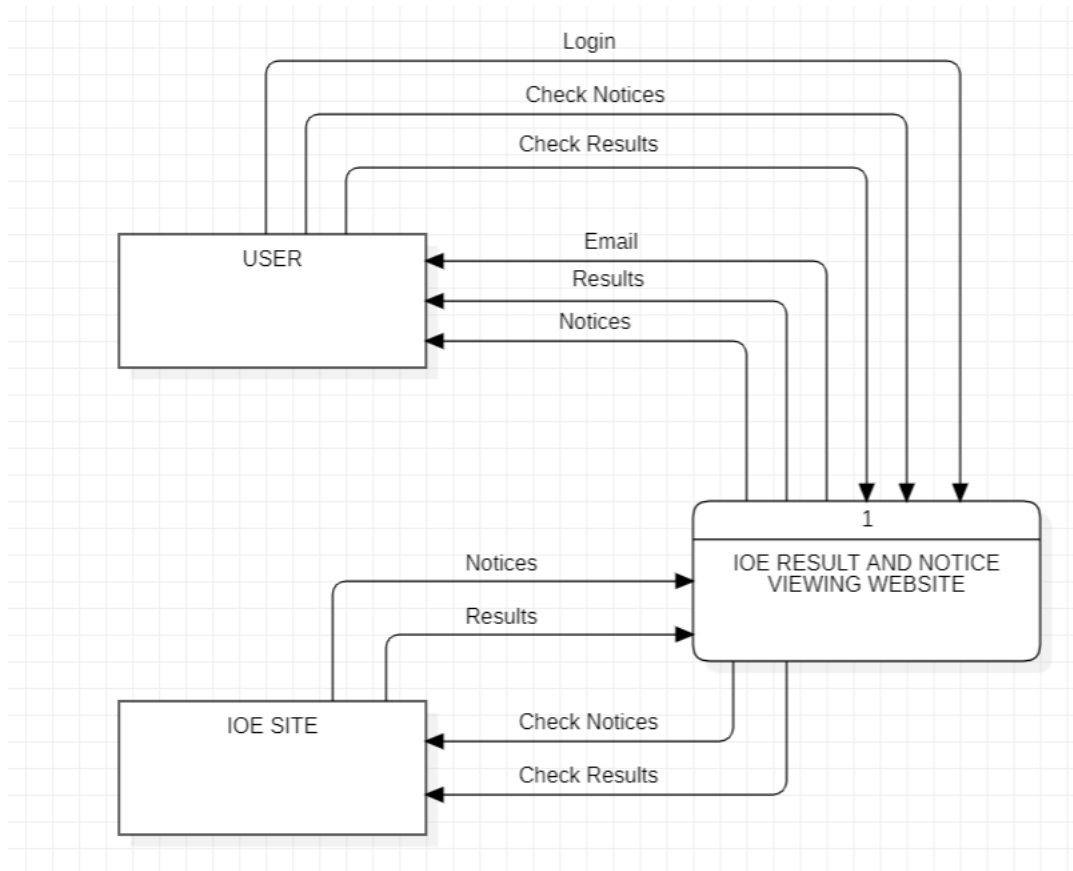Figure 3.2: Use Case Model of the System

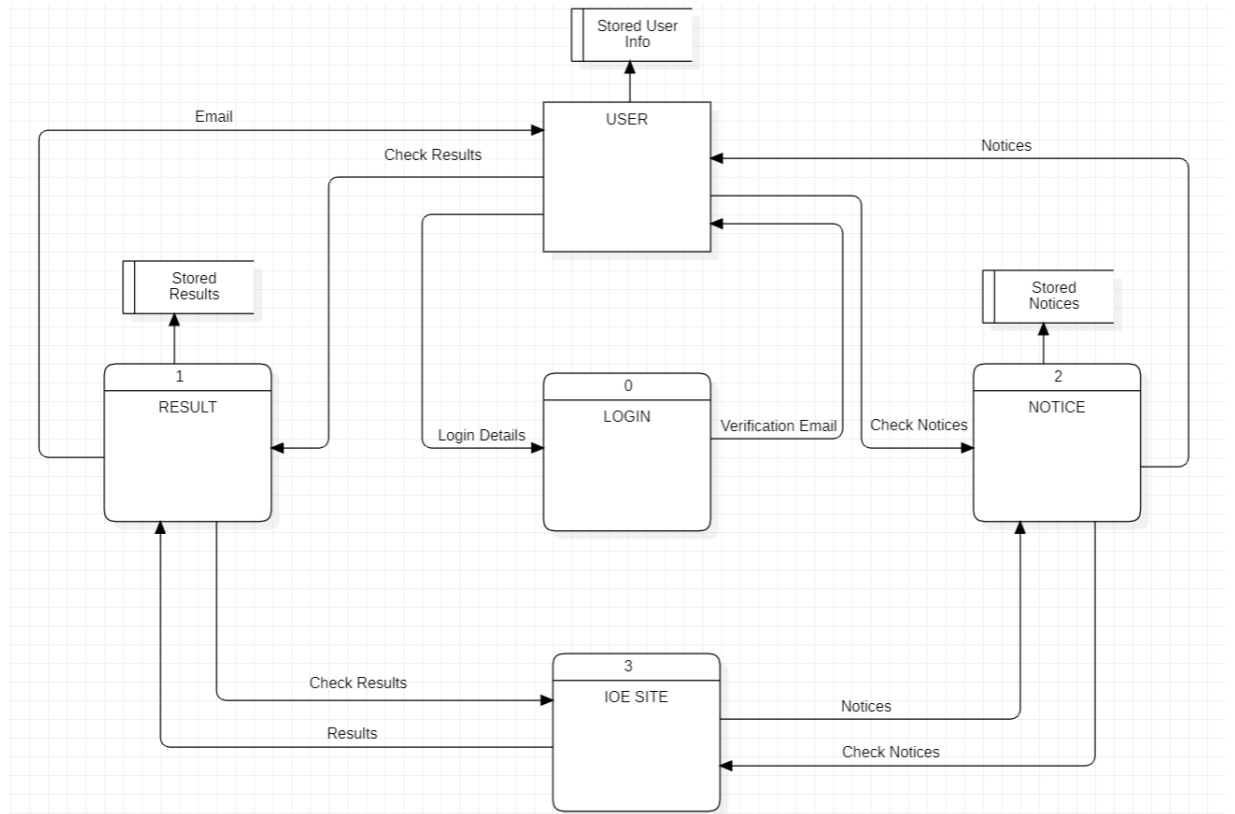**DFD Diagrams**

Figure 3.3: DFD Level 0 of the System

Figure 3.4: DFD Level 1 of the system

3.3 Software Development Model

The incremental model is one of the easiest to implement software development life cycle models. There are certain scenarios where the initial or the core software requirements are clearly defined, but the actual span or the full set of features of the project are unknown. Moreover, the development company might decide to not give the full functionality of the software in one go. Rather they prefer to give it out through periodic updates. Or the client requests some functionality enhancements during the process of development. In such cases, the incremental model is used.

In the first phase we performed Web Scraping and extracted all the results and notices from the website. In the second phase we used Database using MYSQL and inserted all the notices and results in the database. In the third phase we implemented the concept of OCR. In the fourth phase we performed mailing methods to send the email of result status to the subscribed students
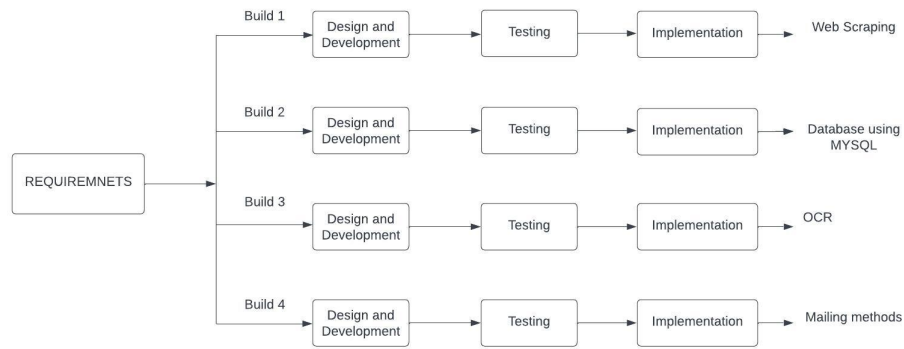
15

Figure 3.5: Incremental Model

3.4 Requirment Analysis

3.4.1 Functional Requirements

- The system should view the notices, results..
- The system should provide subscription feature for students of KEC only.
- The system should send result status email only to the subscribed accounts automatically after the publishment of result.
- The system should provide a feature to contact with admin of the website to report problems and error.

3.4.2 Non-Functional requirements

- The system should be user friendly and easy to use.
- The system should be compatible and should run on almost all the device.
- The system should be reliable and responsive.

CHAPTER 4
EPILOUGE

4.1 Expected Output

The expected output of the Nepali Voice Assistant is to provide a seamless and fun experience for Nepali-speaking users by accurately understanding and responding to their voice commands and queries. Users can expect the assistant to execute various tasks effortlessly, such as setting reminders for important events,accessing specific applications or services, all through spoken Nepali instructions.The Nepali Voice Assistant will be able to enhance user convenience, productivity, and engagement by providing accurate, relevant, and context-aware responses and actions in the Nepali language.