**Tribhuvan University**

**Institute of Engineering**



**Himalaya College of Engineering**

**Chyasal-09, Lalitpur**

**A THIRD-YEAR MINOR PROJECT REPORT**

**ON**

**QUESTION PAPER GENERATOR**

**AND MODERATOR**

**Submitted by:**

**Ashok Ghimire (HCE075BCT009)**

**Resarch Paija Pun (HCE075BCT024)**

**Sabal Thapa (HCE075BCT027)**

**Kishor Humagain (HCE075BCT044)**

**Submitted to:**

**Department of Computer and Electronics Engineering**

**Date: 2078/11/29**

**A THIRD –YEAR MINOR PROJECT**

**ON**

**QUESTION PAPER GENERATOR**

**AND MODERATOR**

**Tribhuvan University**

**Institute of Engineering**

**Himalaya College of Engineering**

**Chyasal-09, Lalitpur**

**Submitted by:**

**Ashok Ghimire (HCE075BCT009)**

**Resarch Paija Pun (HCE075BCT024)**

**Sabal Thapa (HCE075BCT027)**

**Kishor Humagain (HCE075BCT044)**

**Submitted to:**

**Department of Computer and Electronics Engineering**

**Date of submission: 2078/11/29**

**ABSTRACT**:

Examination is a medium of knowing the capability of a person and how far their understanding of certain subjects. Today education has become a part of life. In academics, level of understanding is mostly determined by how well the learner performs in their examinations. Making examination to be a very import part of the learning system. Examination prepare students in their quest for knowledge. Therefore, having an up to standard examination paper and format is very important. With the way examinations are prepared it becomes difficult to eliminate all malpractices and inefficiency.

We therefore propose an automated process of Question Paper Generation, which is fast, randomized, streamlined and secure. Every task performed by this system is automated so that storage space, bias and security is not a concern anymore. This system can be helpful to many educational institutes. Traditionally, the questions are prepared by the teachers manually writing each question on the sheet of paper. This paper describes a method of auto generating a new set of exam questions. Questions that have been frequently asked in the previous questions are already stored in the database and if the teachers wish to add new questions or ask new questions in examination they can add it into the database of questions by using OCR.

**Keywords: Examination, Questions, OCR, Regular Expression, Automation, Randomization, Secure**

**TABLE OF CONTENTS**

1. Introduction……………………………………………………………1

1.1 Background and motivation..............................................................6

1.2 Problem Statement.............................................................................6

1.3 Objective............................................................................................7

1.4 Features..............................................................................................7

1.5 Applications……………………………………………………..….8

2. Feasibility Analysis.................................................................................8

2.1 Technical Feasibility..........................................................................8

2.2 Economic Feasibility…......................................................................9

2.3 Operational Feasibility…...................................................................9

2.4 Legal feasibility.................................................................................9

3. Requirement Anaslysis..........................................................................10

3.1 Software and Hardware Requirement…......................................... 10

3.2 Functional Requirements………………………………………….11

3.3 Non-Functional requirements……………………………………..11

4. Literature review....................................................................................11

5. Methodology..........................................................................................12

5.1 Incremental Model.........................................................................12

5.2. Process…………………………………………………………..14

5.3 Flow Charts………………………………………………………15

5.4. Control Flow Diagram…………………………………………..17

5.5 Data Flow Diagram (DFD) ...........................................................18

5.6 Use Case Diagram..........................................................................19

5.7. E-R Diagram…………………………………………………….20

6.Gannt-chart……………………………………………………………..21

7. Expected Output……………………………………………………….22

9.References...............................................................................................23

1. **INTRODUCTION**

Technology today has become the driving force for everything. Governments invite a lot of money in institutions to update and improve the technological world, the major contributors of this technology movement are educational institutions. Education has become the key to national development. Universities and colleges are being opened and established in all parts of the country, with this number of new universities the critical question remains to be what is the quality of students being produced at these universities. The most traditional way of determining students understanding has been examinations. Examinations play a very important role in the evaluation of students and determining students’ progress. Therefore, the quality of examination is the key to the credibility of the institution. The traditional way of preparing exams has always been that the lecturers, professors, instructors, set the question paper and them submits it for review by the examining board or senate. This method introduces a lot of human errors and high possibility of paper leakages. Since the papers are done in hard copy, storage also becomes a problem to those institutions wishing to maintain a database for past examination papers.

A design of suitable automated system for generating question papers and managing related data may prove vital in an Educational Institute. In this paper, we have proposed an integrated automated system that stores questions related to a particular course and prints a question paper based on its syllabus and curriculum which will be based on website and done from it. We have implemented a role-based hierarchy which restricts access to the users. The system also deploys security mechanisms that prohibit duplication of question papers. There are provisions to enter and edit data suitable to any educational organization with complete freedom for specifying courses, semesters, syllabus and pattern. This enables an educational institute to generate question ensuring security and non-repetitiveness of question papers and is a boon for organizations with limited staff and resources. Our system aims to provide fast operations, data storage and high security for all its tasks. The evolution of traditional and existing Question Paper Generation systems and the need for an automated system is most essential for this present time. In this project, we have been working to make our system of Question Paper Generation. This system will be using OCR to scan the documents or questions and add it into the database.

**1.1 Background and Motivation**

As most human working processes, this system suffers due to bias. There might be some questions which are repeated in many question papers as the professor has a personal inclination towards them. So there is no guarantee of pure randomly generated question paper. Other problems that may plague this system are non-availability of staff and resources, systematic errors, natural calamities and accidents. Also, the security of the system can be easily compromised if leverage over the person responsible for generating question papers is obtained.

Other limitations include: -

a) Lack of storage space

b) Prone to damage

c) Inefficient document transportation

d) Supply costs

e) Poor environmental credentials

f) Limited collaboration

g) Editing problems

**1.2 Problem Statement**

The examination question paper generation method currently used is Zambia method. In Zambia method, all government intuitions and most private intuitions use the manual way of examination paper preparation. The professors and lecturers are asked to prepare the examination questions then submit to an examinations board or department. The process requires a lot of time and the paper setters need to divide the time between lecturing/teaching and setting examinations. This compromises the quality of the paper due to human errors and the many stages and number of officers involved in the preparation of the paper, the risk of leakages are increased.

**1.3 Objectives:**

* To generate the question paper in certain amount of time that saves the consumption of time as occupied by manual paper generation.
* To moderate the paper i.e. to add any new questions into the database such that new questions can be provided for examinations.

**1.4 Features:**

* Following are some of the features of the Question Paper Generation and Moderation System:-
* a) Simple interface which enhances the ease of updating data.
* b) Generates and prepares the Question Paper in specific time.
* c) This system will be website-based.
* d) Questions can be easily edited.
* The Question Paper Generation and Moderation System provides various advantages to the user when compared to the traditional system. Listed below are some of the advantages of the system: -
* a) User can generate test papers randomly and instantly, saving a lot of time.
* b) The algorithm enables randomization of questions so avoids the often repetition of questions.
* c) A new question can be added to the database at any instant and different sets of test papers could be generated.

**1.5 Applications:**

This Question Paper Generation and Moderation System has certain specific applications. Firstly, it is applicable to generate the papers in very short period of time which will ultimately save the time that is consumed during the traditional way of generating the papers. Similarly, if the teacher wishes to add new questions, then, it can be instantly added into the database and the question can be given in the next examination. Teachers often tend to repeat a lot of questions that had been asked before. So, with this system the paper can be generated in short period of time and there is no need to frequently type the same questions always. Also, new questions can be added into the database and new questions may also appear in the paper.

Other applications are:

* can be used by all educational institutions for the quicker generation of papers.
* can be used to moderate the papers and add new questions instantly.

1. **REQUIREMENT ANALYSIS**

**2.1 Software and Hardware Requirements**

As it is web-based application, the external software requirement for the client side is not specific. Client can access from any web browsers after launched. For developers, certain software are used to develop the web page. Simple developer tools with low memory space can be used to develop the project.

Due to the low processor load utilized by our project, high-end hardware requirement is not an issue for the client. But for the developer to build the overall project from scratch to complete, hardware requirement of mid-range can be good pick. Server can be the PC of developer if it is hosted in localhost and the specification of the server should be increased for the online hosted server.

|  |  |  |
| --- | --- | --- |
|  | Hardware | Software/Language/Framework |
| Client | Mobile, Computer  Any PC above 2GB RAM  Any mobile above 540MB RAM | Web Browser  Windows 7 and above |
| Developer | Computer  Above (I5 7-Gen, CPU 1.8GHZ, 8-GB RAM) | Visual Studio Code  HTML, CSS, JS,  Db.sqlite3, Django  Windows 7+ OS |
| Server | 3.6 GHZ or fast CPU  10 GB HDD or 5GB SSD  **(**Developer specification for local host server) | Django Live server  (Can be implemented in online server) |

**Functional Requirements:**

Login:

Admin and Teachers have unique username and password. They can login to our system simply entering correct login username and password.

Interface:

Our user interface is simple, responsive and interactive. User will have easy idea on how to use our system.

Uploading raw questions:

Admin has right to upload the raw data (group of questions) in pdf format to save to the database. Only admin can modify the questions in database.

Dashboard:

Teachers will have a dashboard where they can fill up the fields to generate the question paper.

Viewing Question paper:

The generated question paper can be viewed in web page after proceeding to generate. The question paper can also be downloaded for further use.

**Non-Functional Requirements**:

Time Availability:

Currently the project is run in localhost Django-server. So, we can access the web-app as long server is run.

Performance:

OCR scanning of a pdf and storing of questions in database can be executed in less than 30 second and the generation of the question paper can be done in less than 3 second.

Login Security:

Only admin can add teachers to the system by providing login credentials.

System Security:

A user trying to access unauthorized webpages will be redirected to login page if the use isn’t authenticated. Database access isn’t given to teacher so the teacher doesn’t unintentionally messes up with the database.

Maintainability:

Our project is easy to extend with new functionality.

Individual table for Subjects:

Questions can be fed to OCR system as much as admin can and they will be stored subject-wise in the tables of database.

Uniqueness:

The contents of tables in database are unique to each other. This will generate unique questions.

**SYSTEM DESIGN:**

1. **DFD Diagram**

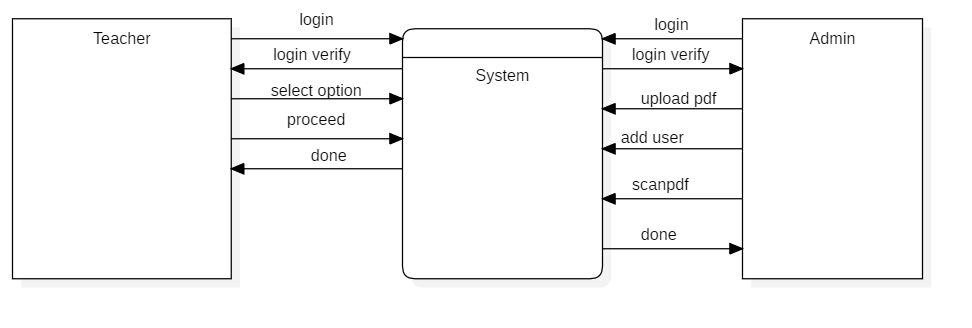


Fig: DFD-level 0

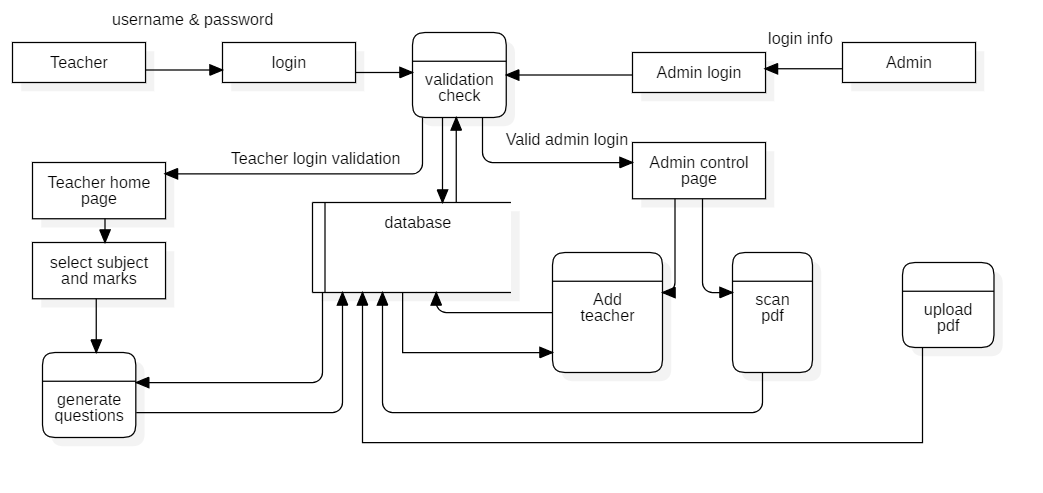


Fig:DFD-level 1

References :

1. www.youtube.com/Python Backend Web Development Course (with Django), free code camp channel
2. www.youtube.com/regular expression with Corey Schafer
3. <https://docs.djangoproject.com>
4. Software Requirements Specification for project iTest,2008
5. The International Journal of Multi-Disciplinary Research on Examination System,2014
6. International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 05 | May 2019
7. Explainthatstuff.com |OCR by Chrish Wood , 24 may 2021
8. Fonseca Cacho, Jorge Ramon, "Improving OCR Post Processing with Machine Learning Tools" (2019). UNLV Theses, Dissertations, Professional Papers, and Capstones. 3722.

**Regular Expression:**

A regular expression is a method used in programming for pattern matching. Regular expressions provide a flexible and concise means to match strings of text. Regular expressions are also known in short form as regex or re . Regular expressions can be incredibly powerful. Essentially, if the pattern can be defined, a regular expression can be created. A simple pattern might be something as simple as finding all situations where a sentence ends in "that" and is replaced with "which". The pattern could get more complex by doing the same replacement but only on the 3rd and 5th occurrence of a match. Or it could get even more complicated by using different sets of matching characters depending on the frequency and location of previous matching characters. Regular expression allows us to surf through a file containing the large amount of random words, any characters or alphanumeric characters in random order. When surfing, we can match the pattern given by our regular expression with the random group of words, characters or alphanumeric characters. We can extract the texts of our required pattern from the matched list and use them for our own purpose

As regular expression helps to find the pattern from our file. In this project we have attempted to match the question and mark pattern from the given raw question paper. The question paper consists of no of questions with question number and the marks assigned to the question. A question with a tail question can also be detected. We matched the questions with the respective marks and saved into the database of the respective subject.

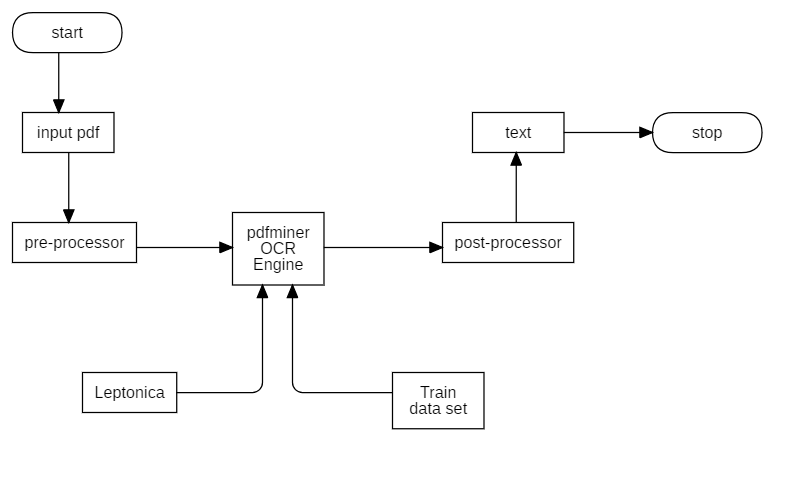
Regular expressions are used for syntax highlighting systems, data validation and in search engines such as Google, to try to determine an algorithmic match to the query a user is asking. Data is everything in our digital world and the regular expression helps us to extract the data from the unmanaged order of data.

**OCR:**

OCR stands for "**Optical Character Recognition**." It is a technology that recognizes text within a digital image. It is commonly used to recognize text in scanned documents and images. OCR software can be used to convert a physical paper document, or an image into an accessible electronic version with text. For example, if we scan a paper document or photograph with a printer, the printer will most likely create a file with a digital image in it. The file could be a JPG or PDF, but the new electronic file may still be only an image of the original document. We can then load this scanned electronic document it created, which contains the image, into an OCR program. The OCR program which will recognize the text and convert the document to an editable text file.

So, how does OCR work? As you read these words on your computer screen, your eyes and brain are carrying out optical character recognition without you even noticing! Your eyes are recognizing the patterns of light and dark that make up the characters (letters, numbers, and things like punctuation marks) printed on the screen and your brain is using those to figure out. Computers can do this too, but it's really hard work for them. The first problem is that a computer has no eyes, so if you want it to read something like the page of a book, you have to present it with an image of that page generated with an optical scanner or the pdf format of that book. OCR program detects the characters by the pattern recognition and the feature recognition. The pattern recognition recognizes the character or group of characters by defining the pattern of the individual character in context of how it is made or drawn in graphical like way. The character differs from the individual to individual. It would be lot easier to recognize character if the patterns are same or produced by the same machine. The feature recognition method has datasets of fonts for a character to match the character with the pre-provided fonts.

In this project, we have implemented OCR for scanning the questions provided by the teachers which are the raw data for our database. The admin uploads the pdf format questions provided by the teachers to our system. After the file is uploaded the OCR scanning link can be clicked so that the admin can scan the questions. The link redirects to the OCR module where the processing of pdf can be done and the texts are extracted from the pdf. After OCR scanning, we use regular expressions to extract the useful information from the scanned texts to store the questions in database.



General steps of OCR Working:

1 . Input pdf:

The pdf containing the questions are fed into the OCR as inputs for the further processing and manipulation of the pdf’s data.

2 . Pre-Processor:

Pre-processing is done when the input in received in OCR system. Pre-Processing involves the following techniques:

* 1. Binarization:

In layman’s terms Binarization means converting a colored image into an image which consists of only black and white pixels (Black pixel value=0 and White pixel value=255). As a basic rule, this can be done by fixing a threshold(normally threshold=127, as it is exactly half of the pixel range 0–255). If the pixel value is greater than the threshold, it is considered as a white pixel, else considered as a black pixel.

So, the crucial part of binarization is determining the threshold. The threshold can be calculated using local minima and local maxima method.

Where ,

Imax= Maximum pixel value in the image

Imin= Minimum pixel value in the image,

E = Constant value

C(i,j) is the threshold for a defined size of locality in the image (like a 10x10 size part). Using this strategy we’ll have different threshold values for different parts of the image, depending on the surrounding lighting conditions but the transition is not that smooth.

* 1. Skewness:

While scanning a document, it might be slightly skewed (image aligned at a certain angle with horizontal) sometimes. While extracting the information from the scanned image, detecting & correcting the skew is crucial.

We can use project profile method for the skewness correction. In this method, we take binary image and project it horizontally (taking the sum of pixels along rows of the image matrix) to get a histogram of pixels along the height of the image. We rotate the image at various angles to find the skew angle and we can correct the skewness by rotating the image through an angle equal to skew angle in opposite direction.

* 1. Noise Removal:

The main objective of the Noise removal stage is to smoothen the image by removing small dots/patches which have high intensity than the rest of the image. Noise removal can be performed for both Colored and Binary images.

1. . OCR engine:

OCR engine has Leptonica and group of datasets to manipulate the pre-processed data. Leptonica is a pedagogically-oriented open-source library containing software that is broadly useful for image processing and image analysis applications. Dataset is the collection of the predefined large amount of data for matching the data to given input data. NIST dataset is used in our OCR engine. NIST has 800000 characters images which are used to training our OCR algorithm and extract the meaningful data from it.

1. Post-Processor:

Post Processing involves data cleaning steps for documents that were digitized, such as pdf.  One step in this process is the identification and correction of spelling and grammar errors generated due to the flaws in the OCR system.

1. . Text :

Finally our OCR process is completed . The result after the post-processing is the plain text which is saved in text file in our system.

**Programming languages used:**

1. **Python**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. It can be used in many applications like front end web development, backend web developing, database management, data mining, artificial intelligence etc.

2 . HTML

HTML stands for ‘Hyper Text Markup Language’. HTML, or Hypertext Markup Language, is a markup language for the web that defines the structure of web pages. It is one of the most basic building blocks of every website, so it's crucial to learn for the front end web development. It can integrate CSS and Javascript to make more interactive websites. HTML is the skeleton of the webpages.

3. CSS

CSS stands for ‘Cascading Style Sheet’. CSS describes how HTML elements are to be displayed on screen, paper, or in other media. CSS saves a lot of work. It can control the layout of multiple web pages all at once. External stylesheets are stored in CSS files. CSS is the skin of the webpages.

4 . Javascript

Javascript is the scripting language which helps to make the webpages interactive. Javascript is the cross platform language that contains a standard library of objects, such as Array, Date, and Math, and a core set of language elements such as operators, control structures, and statements. Client-side JavaScript extends the core language by supplying objects to control a browser and its Document Object Model (DOM). For example, client-side extensions allow an application to place elements on an HTML form and respond to user events such as mouse clicks, form input, and page navigation. Basically, JavaScript is the circulatory, digestive, and respiratory systems that brings the structure and the skin to life.

5 . Django

Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and many options for free and paid-for support. Django can be used to build almost any type of website from content management systems, through to social networks and news sites. It can work with any client-side framework, and can deliver content in almost any format (including HTML, JSON, XML, etc.).

6 . VS-Code:

Visual Studio Code is the powerful and interactive code editor which is lightweight and supports various programming languages. It has powerful developer tooling, like IntelliSense code completion and debugging.  VS Code helps with syntax highlighting, bracket-matching, auto-indentation, box-selection, snippets, and more. Intuitive keyboard shortcuts, easy customization and community-contributed keyboard shortcut mappings let you navigate your code with ease. It is widely used in the programming world due to simple to learn environment for beginners and the advance toolkits for developers.

**Methodology:**

The problem with the existing Zambia examination question paper generation method has led us to develop this project which ensures the fast and secured question paper generation. The approaches we have taken to generate the question paper are through the OCR scanning of the raw questions sets, setting them in order in database by extracting the useful information from scanned file and re-accessing the database to display the questions in exam question paper format.

We start the generation of our final output (which is question paper generation) by feeding the raw data into the system. The admin has right to update the file in pdf format which is given to the OCR for scanning. On clicking the link to scan the pdf the OCR is set to process the pdf’s raw data. The raw questions stored in the pdf should have special formatting to be fed into our system. For a given question, it should start with question numbering, the second field should have the question and the third field should define the marks of respective question. We have an OCR module called pdf miner to implement the text scanning functionality. It recognizes the unique characters from the pdf and extracts the texts. The texts are stored in text file. The text file is then used by the regular expression code to extract only meaningful texts. The regular expression defines the certain rules under which the text expressions or patterns are selected from the huge number of random data. When the extracted text pattern matches with the regular expression, the information including the question and its marks is saved into database table of that respective subject.

The question paper generation is easy and fast with some fields to fill on. The teacher can login into the system with the unique login credential provided by the admin. When a teacher is validated by the system, they are redirected into dashboard. In dashboard, teacher is provided with a form to input the full marks, pass marks, no of question weighing four and two marks, and the faculty should be chosen for generating question paper of respective subject. The form data from html template is received by Django in backend to work further on accessing and taking values from database based on the input given by the user. When teacher clicks on proceed button, s/he is redirected to the output page where the random questions are generated according to the number of weight they carry satisfying the full marks given. The full layout of the normal examination question paper is supposed to displayed which can be viewed by the teacher.