International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Automated System for Frequently Occurred Exam Questions

Manjusha sanke[1], Nihar Pednecar[2], Viraj Padwalkar[2], Raghavendra Poojary[2], Vikas Vishwakarma[2]

[1]Assistant Professor, Department of Information Technology, Shree Rayeshwar Institute of Engineering and Information Technology Goa, India

[2]Department of Information Technology, Shree Rayeshwar Institute of Engineering and Information Technology Goa, India

---

Abstract - *Education has become an integral part of our society today. This is the reason why examinations play an important role in testing the performance of the student. Generating an effective question bank is a task of great importance for any educational institute. The traditional method where lecturers and students manually prepare question bank, is very challenging, time consuming and mostly inaccurate. In this paper, we present the solution in the form of Automated System for Frequently Occurred Exam Questions.*

This web based application will enable lecturers to automatically generate question bank from existing question papers in the database in a very short time and also help students.

Key Words: component, formatting, style, styling, insert (key words)

1. INTRODUCTION

There is a change from manual to automated systems for various fields of education system by student as well as teachers. However the main problem is a low quality of question banks generated because of some human factors such as instability and relatively narrow scope of topics. Teachers need to invest a lot of time and energy in composing question banks. Also, the concern remains is how the current technologies would also help the lecturers automatically generate the different sets of questions from time to time without being concern about repetition. Using this system, the same action can be accomplished in minutes.

According to the need, a system named as Automated System for Frequently Occurred Exam Questions is proposed, so as to make the system more efficient, reliable, improve its quality, and also to reduce the time taken by lecturers in setting the question bank manually. This web based application will enable lecturers to automatically generate question bank from existing question papers in the database. The system performs all tasks related to question bank, starting from preparing question bank to printing it. This system is very useful for students as well as lecturers. The system will help the individuals to generate question bank in a very short time period thus saving a lot of their precious time.

2. LITERATURE SURVEY

2.1. Information Retrieval System

Information Retrieval is a field of computer science that deals with the representation, storage and access of information. Information Retrieval deals with organization and retrieval of information from database collections [6]. IR is the process by which a collection of data is searched, represented and stored for the purpose of knowledge discovery as a response to a user's request.

IR is a component of an information system. An information system must make sure that everybody it is meant to serve has the information needed to accomplish tasks, solve problems and make decision, no matter where the information is available. An information retrieval system must actively find out what users need, acquire data resulting in a collection and match data with needs. Figuring out the information that the user needs to solve a problem is important for successful retrieval.

2.2. Information Retrieval Model

IR models[8] specify the details of the document representation, the query representation and the retrieval functionality. Set-theoretic, algorithm, probabilistic, feature based models are the different models in IR.

2.2.1. Set-Theoretic model

Set-theoretic models represent documents as sets of words or phrase. Similarities are usually computed from set-theoretic operations on those sets. Commons models are:

- Standard Boolean model
- Extended Boolean model
- Fuzzy retrieval

2.2.2. Probabilistic model

Probabilistic models[3] make the process of document retrieval as a probabilistic inference. Similarities are obtained as probabilities that a document is related for a given query.

Probabilistic theorems like the Bayes' theorem are often used in these models.

- Probabilistic relevance model on which it is based the okapi (BM25) relevance function.
- Uncertain inference
- Language models
- Divergence-from-randomness model
- Latent Dirichlet allocation
- Binary Independence model

2.2.3. Algebraic model

Algebraic models represent document and queries usually as vectors or tuples. The similarity between the query vector and document vector is represented as a scalar value.

- Vector space model
- Generalized vector space model
- (Enhanced) Topic-based Vector Space Model
- Extended Boolean model
- Latent semantic indexing

Vector Space Model

The fundamental IR model used in our system is vector space model. A vector space model is also an algebraic model, involving two steps:

- In first step we represent the text document into vector of words.
- In second step we change to number format so that we can apply any text mining techniques such as information retrieval.

In a Vector Space Model

- Document and queries are M-dimensional term vectors.
- Non binary weights to index terms.
- A query is similar to document if their vectors are similar.
- Retrieved documents are sorted by decreasing order.

2.3. Implementation Techniques

2.3.1. Data Pre-processing Algorithm

2.3.1.1. Stemming

Removing suffixes by automatic manner is an operation which is especially useful in the area of information retrieval. In a typical IR environment, one has a huge collection of documents, each described by the words in the document title and by words in the document abstract. [9]

Ignoring the problem of precisely where the words originate, we can say that a document is represented by a vector of words or terms. Terms with a common stem will usually have same meanings[1], for example:

CONNECT

CONNECTED

CONNECTING

CONNECTION

CONNECTIONS

The performance of an IR system will be enhanced if term groups such as this are conflated into a single term. This may be done by removing the suffixes

-ED, -ING, -ION, -IONS to leave the single term CONNECT. In addition, the suffix stripping process will reduce the total number of terms in the IR system and hence the size and complexity of the data in the system, which is always advantageous.

2.3.1.2. Stop word Removal

Stop-words[5] are the words which do not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence.

You should remove these tokens only if they don't add any new information to the sentence. Classification problems normally don't need stop words because it is possible to talk about the general idea of a text even if you remove stop words from it.

The algorithm is implemented as below given steps:

Step 1: The document from which stop words have to be removed that target document has to be tokenized and than individual words are stored in array.

Step 2: A single stop word is read from stop word list one by one.

Step 3: The stop word is compared to target document text in the form of array using sequential search technique.

Step 4: If it matches, the word in array is removed and the comparison is continue till length of array.

Step 5: After removal of stop words completely, another stop words is read from stop word list and again algorithm follows Step 2. The algorithm runs continuously until all the stop words are compared and removed.

Step 6: Resultant text after removing stop words is displayed, also required statistics like stop word removed, number of stop words removed from target text, total count of words in target text, count of

words in resultant text, individual stop word count found in target document text is displayed.

2.3.1.3. Text Rank

Text Rank[2] is an extractive and unsupervised text summarization technique. The algorithm works as follows:

- The first step is to concentrate all the text containing in the articles.
- Then split the text into individual sentence.
- Next, we will find vector representation (word embeddings) for each and every sentence.
- Similarities between sentence vectors are calculated and stored in a matrix.
- The similarities matrix is then converted into a graph, with sentence as vertices and similarity scores as edges, for sentence rank calculation.
- At last a certain number of top-ranked sentences form the final summary.

3. Application

3.1. Existing System

- The existing system is paper based system.
- The production of question bank is done manually.
- It needs human working to create question bank .For this teachers select questions by checking syllabus and previous question paper.

Drawbacks of Existing System:-

- Manual process consumes more time.
- More time and effort require managing and retrieving from manual question papers.
- It is difficult to handle with repetition.

3.2. Proposed System

- The proposed system will overcome limitations of existing.
- Automatic generation of question bank.
- Reduce time and efforts.
- Provides more security and flexibility for generation of question bank.

Advantages of proposed system:-

- The entry of the questions is all done by facilities with syllabus and subject label.
- The user can take print of this question bank or they can save it for future use.
- This system will help us to keep our data integrated.
- The problem of data loss will be eliminated.

3.3 Working of the System

- User uploads a particular subject syllabus file.
- After the syllabus is uploaded it will be converted in the form of query. This query will be used to select the question papers of the particular subject from the corpus.
- Set of question papers will be stored in a data set in the system.
- The question papers of a particular subject will be fetched by the query for analysing the question papers.
- The system will perform the analysis on the question papers according to the syllabus.

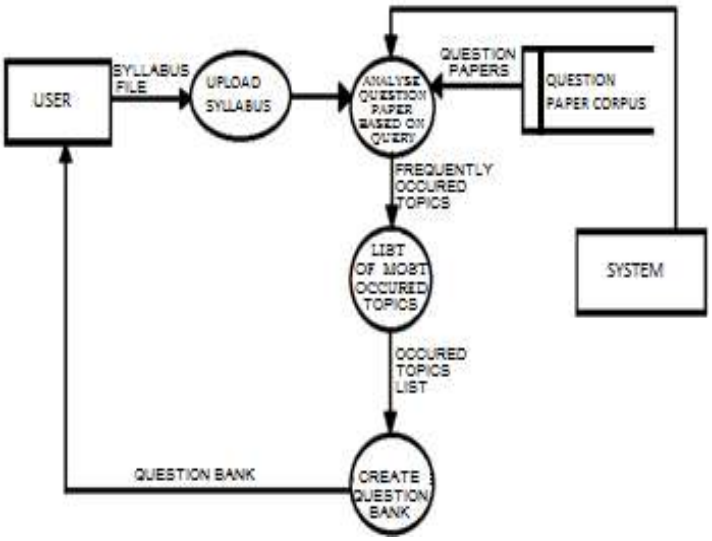

Fig.1. Working of the system

4. CONCLUSION

In this paper, a new automated system for frequently occurred exam questions has been proposed. When we compare the proposed system with the existing system, we found that the proposed system is more reliable, accurate and feasible in comparison with existing system. The proposed system saves lot of time with very less effort. In future we will enhance the scope by creating the list of important topics for semester exams or university exam.

REFERENCES

1. Ms. Anjali Ganesh Jivani, "A Comparative Study of Study Stemming Algorithm". Anjali Ganesh Jivani et al, Int. J. Comp. Tech. Appl., Vol 2 (6), 1930-1938
2. Rada Mihalcea and Paul Tarau, "TextRank: Bringing Order into Texts"
3. Akram Roshdil and Akram Roohparvar, "Review: Information Retrieval Techniques and Application". International Journal of Computer Networks and Communications Security VOL.3,no.9, September 2015, 373-377

4. Jovdip Datta, "Ranking in Information Retrieval". April 16, 2010
5. Balasubramai Ramasamy, "Survey on Pre-processing Techniques for Text Mining ", June 2016
6. Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston,USA, 1999.
7. Vaibhav Kant Singh, Vinay Kumar Singh "Vector Space Model: An Information Retrieval System"
8. Hai Dong, Farookh Khadeer Hussain,Elizabeth Chang "A Survey in Traditional Information Retrieval Models", 2008 Second IEEE International Conference on Digital Ecosystems and Technologies.
9. Martin F. Porter. An algorithm for suffix stripping. Pages 313–316, 1997.
10. B. A. Ribeiro-Neto and R. Muntz, "A brief network model for IR," in 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, 1996, pp. 235-260.
11. Vikram Singh, Balwinder Saini, "An Effective Pre-Processing Algorithm for Information Retrieval Systems", January 2015 DOI: 10.5121/ijdms.2014.6602