# Online Examination with Short Text Matching

[1]Pooja Kudi          [2]Amitkumar Manekar          [3]Kavita Daware          [4]Tejaswini Dhatrak

[1]poojakudi00@gmail.com
[2]amitkumar.manekar@sitrc.org
[3]kavitadaware100@gmail.com
[4]tejudhatrak25@gmail.com

**SITRC, Computer Engineering Department, Sandip Foundation, Nashik.**
**Affiliated to Pune University**

*Abstract*— In traditional Online Examination System, only objective type questions are assessed and according to that marks are given to the student. However, this technique lacks the capability of evaluating descriptive answers. In university examinations, there are many types of question included for evaluation of the students. Therefore, the automated system must be capable of evaluating the descriptive answers. The online examination system checks the student answer by matching the answer with predefined set of answer. The predefined answers are saved on the server and evaluation is done automatically using the automatic assessment tools. Here the machine learning approach is used to solve this problem using text mining. Measuring the similarity between, sentences, words, documents and paragraphs is an important component in various tasks such as text summarization, information retrieval, automatic essay scoring, document clustering, and machine translation and word-sense disambiguation. In this system JSON is used for transferring data between web application and server, serving as an alternative to XML.

*Keywords*— *Text mining, automatic assessment tools, Machine learning, JSON.*

## I. INTRODUCTION

Large number of Candidates attends the exams for that evaluation is required manual effort. In some cases student having the poor hand-writing which is not clearly understandable by evaluator. Sometimes quality of evaluation may change according to mood of evaluator. The evaluation work is very lengthy and time consuming. So, to avoid such problems the automated examination environment is developed. Hence, the solution made for automating the work using computers .The novel approach is to describe an automated descriptive answer marking system that can be utilized to improve teaching and learning of the technical subjects. Question paper is prepared by the expert teacher in consultation with the technical staff with the system who assists in uploading the same in the desired format. Candidates appear for the test online within the exam centre. User authentication has been provided. On successful login, each student gets the question paper and his/her answer book. They are required to answer the questions by typing in the blank space provided for the answer. The software facilitates saving of the typed work as per student convenience. In case of a system failure, the saved work is available to the student for continuation. The problems like document representation, classifier construction and classifier evaluation are solved by using machine learning approach. In Online Examination System *JSON* [5] is used for transferring data between web application and server, serving as an alternative to *XML* [8]. Data mining consists of many techniques such as clustering, neural networks and classification and decision trees. Data mining is a process, before applying any technique preprocessing of data is required. All textual based information are stored, may be in computers or on web. Any computer or laptop can easily stored large amount of data because of advance hardware storage devices. Accumulating information is easy but finding related information on demand can be difficult.

## II. RELATED WORK

An overview of automated short-answer marking systems is presented in this section.

### A. C Rater

C rater [1] is educational testing service which is technique used for the content scoring. This content scoring is based on model building which makes various answer model for candidate's short answer [6]. ETS (Education Testing Service) used for checking short answer of candidate in range of 100 words approximately [7]. It uses analytical approach & rubrics item which specify correct & important terms which should be present in candidate's answer.

The problem of different answer structure of candidate is solved using c-rater. It saves 0-12 hour work of humans.

*a. Steps of c-rater model are* [1]:
    1. Model building.
    2. C Rater automatically processes.
    3. Maintaining algorithm gold map.
    4. Apply student answer.

*b. Task performs by c-rater* [1]:
    1. Processes for spelling correction.
    2. Part of speech tagging & parsing.
    3. Parse tree passed through feature extractor.

## B. Automark

*Automark* [6] system is developed to provide short answer marking. The system processes a number of modules specifically worked for finding errors in typing, spelling, syntax and semantics. *Automark* [6] looks for specific content within answers, the content being specified in the form of a number of marking scheme designed [7]. Each structure represents all possible valid or an invalid answer. Using an offline custom written configuration interface, templates are developed. The *Automark* [6] system was used for marking the spelling, semantic errors and syntax in answer text.
Every marking goes through various stages.

Stage 1: In this stage the answer text is preprocessed to standardize the input with respect to punctuation and spellings.

Stage 2: In the second stage syntax analyzer analyzes the main syntactic of the input text.

Stage 3: This is the last stage where pattern matching module finds all the possible matches for marking.

## C. IndusMarker

*IndusMarker* matching the structure of answer text in terms of predefined structure which is developed by structure editor which built for purpose. The examiner specifies the required structure of an answer in a simple purpose designed language. The language was initially called *QAL* [7] but later on it is redefined it as a sublanguage of *XML* and named it Question Answer Markup Language *(QAML)* [6].

## III. PROPOSED SYSTEM

Many approaches have been proposed for descriptive answer evaluation. The approaches are mainly based on text mining technique which involves keyword matching, sequence matching and quantitative analysis and semantic analysis.

### A. Text Mining

Text mining is used to extract important information or data or pattern or knowledge from various sources which are in the unstructured form [6]. Unstructured-form means which is not in predefined form if the data or information is in an unstructured form then it is difficult to handle them. The main purpose of text mining is to find valuable information from natural language text. The text mining consists of two components:

### 1. Text refining:

*Text refining translates free-form text data into an intermediate form* [3].

### 2. Knowledge distillation:

*Knowledge distillations retrieve the knowledge or patterns from intermediate form* [3]. Intermediate form can be converted into structure like graph and relational data representation. The IF is based on the documents where each entity of document is concept based. Each entity represents the concept or object in particular interested domain.
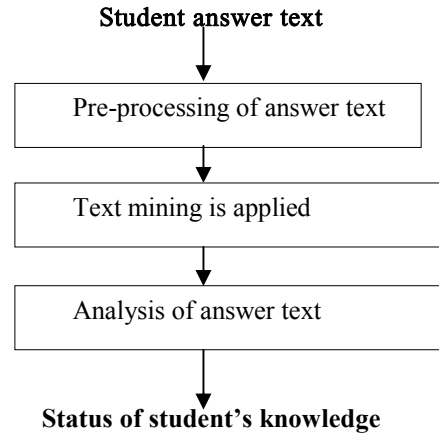


**Figure 1.Process of text mining**

The Figure (1) shows process of text mining and it follows following steps:

**Stage-I: Pre-processing Text**
Pre-processed text is easy to mining as compare to natural languages documents. So, before applying any text mining technique pre-processing of documents that are from various sources is an important task during text mining process.

**Stage II- Text Mining Technique is applied**
At this stage to process the text one of selected algorithm is applied on answer text. The different algorithms are used such as clustering, summarization, classification, information extractions and visualizations.

**Stage III - Analysis of Text**
At this stage the outputs are analyzed for knowing the status of student's knowledge.

### B. QAL (Question Answer Language) and Indus Marker

The short-answer marking system presented is called IndusMarker [6]. IndusMarker performs structure matching, i.e., matching the written answer with the predefine answer structure, developed via a structure editor. The examiner specifies the required format of an answer in a simple purpose designed language. The language was initially called QAL but later on redefined it as a sub language of XML [23] and named it Question Answer Markup Language (QAML) [6].

The question classifier is part of QAL which determines the question and answer type. After the question type is analyzed, the system typically uses several techniques that apply increasingly complex *NLP* [4] techniques on a summarized text. Then search engine is used by the information retrieval module to identify keywords and sentences from the document which contain the answer text. After that filter filters small text that contain text like correct answer.

If candidate can indeed the answer the question then extraction module uses the further clues to determine the answer text and get the overview of expected answer written in QAL. For example, if asked question is 'What is function overloading?' the expected answer for this question is the definition of function overloading. If candidate can indeed the answer the question then extraction module uses the further clues to determine the answer text and get the overview of expected answer written in QAL, consider the following example:

Test question: Describe the idea of "function overriding" in one line (2 marks).

Model answer: "Function overriding means the two methods would have the same signature, but different implementation." The examiner must specify the required structure so that all the expected paraphrases are elaborated. There are two important components in the model answer for the above question. The first required region is "same signature" and the second region is for "different implementation." The given answers both keyword have similar importance. The following is the regions specification in QAL for the above question:

Begin_regions;
Begin_region(marks = 1);
"same signature"
End_region;
Begin_region(marks = 1);
"different implementation"
End_region;
End_regions;

There are multiple possibilities in each region of its own required structure as given below:
Region: "same signature"
Possibility #1:
<main> = ("same";"signature")*2:1:;
Possibility #2:
<main> = ("same";"parameter")*2:1:;
Possibility #3:
<main> = ("same";"argument")*2:1:;
Possibility #4:
<main> = ("same";"data type")*2:1:;
Region: "different implementation"
Possibility #1:
<main> = ("different";"development")*2:1:;
Possibility #2:
<main> = ("different";"construction")*2:1:;

Possibility #3:
<main> = ("different";"implementation")*2:1:;

## IV. MATHEMATICAL MODULE

The question and question type has a distinct difference that is mandatory to remember. Multiple questions were used for a particular question type and there are multiple possibility of answers. For example, two questions were used for the "single term generation" question type. Length of answer text to find the question type is calculated by following formula:

$$z = \frac{\sum_{n=1}^{Q} Wn}{Q \times A}$$

Where,
$Z$ is the average answer length (in words)
$Wn$ is the total number of words in all the answers for nth question
$Q$ is the total number of questions used for the question type
$A$ is the total number of answers per question.

Average human system agreement rate for each question belonging to the question type is calculated by using following formula:

$$r = (a/b) \times 100$$

Where,
$r$ is the system agreement rate for the question
$a$ is the number of judgments where the system agree
$b$ is the total number of judgments.

For each question type average system agreement rate is calculated using following formula:

$$X = \frac{\sum_{n=1}^{Q} rn}{Q}$$

Where,
$X$ is the average system agreement rate
$rn$ is the system agreement rate for the nth question
$Q$ is the total number of questions used for the question type.

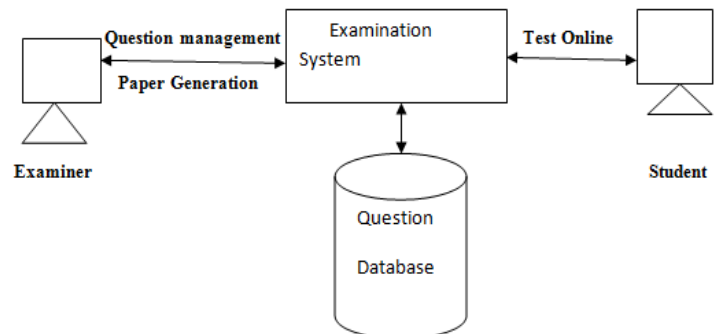## V. SYSTEM ARCHITECTURE



**Figure 2 .Architecture of online descriptive examination**

The above Figure (2) shows various functions of Online Descriptive Examination System [12]:

1) **Question Management:** This module performs following functions such as querying, modifying, deleting, and adding the questions.

2) **Paper Generating:** According to specified requirements this module randomly generates exam paper. It was the core function of online examination system. The system extracts random questions from the question database according to three restrictions, difficulty, including question type, and chapter to generate an exam paper.

3) **Test Online:** Student user could use the test online function for selecting random paper or use a paper designated by examiner, and login the system at any time within the campus network, self-test online, and understand their learning level.

## VI. ANALYSIS OF JSON AND XML

| Parameter | JSON | XML |
|---|---|---|
| **Extended From** | JavaScript. | SGML: "Standard Generalized Markup Language". |
| **Purpose** | JSON is one type of text-based format for interchanging data | having format that contains set of rules for the encoding the documents which is readable for both human & machine. |
| **Syntax** | syntax is lighter than XML | is not as lighter as JSON |
| **Speed** | it has serialized format and so faster. Also its lighter weighted | XML is not so light weighted as JSON. |
| **Support of data type & array** | Supports data type including integer and strings array. | XML does not support for array and data type also. |
| **Object Support** | It has support of native objects. | XML can get support of objects through mixed use of attributes &elements. |
| **Namespace** | JSON does not have support for Namespaces. | XML supports Namespaces. |
| **Mapping** | JSON is data oriented and can be mapped more easily. | XML is document oriented and needs more effort for mapping. |
| **Parsing** | evel ()is used for parsing i.e. for interpreting the JavaScript code & returns the result. It does not need any additional code for parsing. | XML needs XML Document Object Model (DOM) |

**Table 1. Comparative study of JSON and XM*L***

The above Table (1), indicate that JSON is faster and uses fewer resources than its XML counterpart; thus providing significant evidence to refute the null hypothesis [8].

## VII. CONCLUSION

This work has tremendous potential for implementation at large scale. With subsequent research, the system may be adopted by all universities as well as level organizations. It would save enormous amount of man hours for evaluation of examinations. It would also save effort, time and paper and help to overcome other limitations of manual evaluation. Therefore, the automated system must be capable of evaluating the descriptive answers. The online examination system checks the student answer by matching the answer with predefined set of answer. The predefined answers are saved on the server and evaluation is done automatically using the automatic assessment tools. Here the machine learning approach is used to solve this problem using text mining. Measuring the similarity between, sentences, words, documents and paragraphs is an important component in various tasks such as text summarization, information retrieval, automatic essay scoring, document clustering, and machine translation and word-sense disambiguation. In this system JSON is used for transferring data between web application and server, serving as an alternative to XML. For the mathematical answering students need not to write intermediate step he/she can directly write final answer. Because this concept focusing on short answer marking.

### REFERENCES

[1] C. Leacock and M. Chodorow, "C-Rater: Automated Scoring of Short-Answer Question," Computers and the Humanities, vol. 37,no. 4, pp. 389-405, 2003.

[2] Mahesh T R, Suresh M B, M Vinayababu "Text Mining: Advancements, Challenges and Future Directions" International Journalof Reviews in Computing 2009-2010.

[3] Vishal Gupta, Gurpreet S. Lehal "A Survey of Text Mining Techniques and Applications" Journal of Emerging Technologies in Web Intelligence, vo!.1, no. 1, august 2009.

[4] The Stanford Natural Language Processing Group, "The Stanford StatisticalParser,"http://nlp.stanford.edu/software

[5] Dunlu PENG, Lidong CAO, Wenjie XU "Using JSON for Data Exchanging in Web Service Applications" School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China,2011.

[6]RaheelSiddiqi,ChristopherJ.Harrison,andRosheenaSiddiq i"Improving Teaching and Learning through Automated Short-Answer Marking"IEEE transactions on learning technologies, vol. 3, no. 3, july-september 2010.

[7]Salvatore Valenti, Francesca   Neri and Alessandro Cucchiarelli,"An Overview of Current Research on AutomateEssayGrading "DIIGAUniversita 'Politecnicadelle Marche, Ancona, Italy.Journal of Information Technology Education Volume 2, 2003.

[8] NurzhanNurseitov, Michael Paulson, Randall Reynolds, Clemente Izurieta "Comparison of JSON and XML Data Interchange Formats",A Case Study Department of Computer ScienceMontana State University –Bozeman, Montana, 59715, USA.

[9]Official Website of Higher Education Commission Pakistan,http://www.hec.gov.pk, 2010.

[10]Official Website of Pakistan Eng. Council, http://www.pec.org.pk, 2010.

[11] J.H. McMillan, Classroom Assessment: Principles and Practice forEffective Instruction, pp. 40 and 117. Allyn and Bacon, 1997.

 [12]Zhao     Qiao-fang,Li     Yong-fei"Research     and Development of Online Examination System"North China Institute of Science and Technology Beijing, China,2012