Mining Tool for Educational Purposes

[1]Niyati Shah, [2]Pooja Shah, [3]Ankita Malani, [4]Neepa Shah, [5]Stevina Dias
Department of Information Technology,
D. J. Sanghvi College of Engineering, Mumbai, India

Abstract— In Current Educational System, teachers follow manual checking process for assessing the students' performance. They search for the key points in the answers and based on their relevance in the text, grade the students. Checking each and every student's assignment and paper manually is a very tiresome, lengthy and prolonged process. Also there are chances that they may overlook some of the points which the student might have mentioned in their answers and it may reduce the student's grade. This paper presents a mining tool to highlight the sentences that contains the relevant key terms of extracted questions and proposes its use in educational applications. The concepts of pattern matching have been employed to analyze texts and build graphs to represent the overall performance. The mining tool will help evaluate texts from qualitative perspective and ease out teachers' tasks to a great extent.

Keywords— text mining, highlight, education, assignment evaluation

I. INTRODUCTION

In recent years, data mining and text mining have become more popular in the field of education because of the need for employing methods to better evaluate (or introduce) the learning process and to concentrate more on specific problems in education field which need to be tackled. Text Mining refers to the process of deriving high-quality information from the text. It usually involves the process of structuring the input text, deriving patterns within the structured data, and finally evaluation and interpretation of the output. Structuring the input text usually involves parsing, along with the addition of some derived linguistic features and subsequent insertion into a database. The purpose of education is to digitally automate the manual work being practiced by the professionals, to ease out and speed up their work so that they can give more attention to the specific area which actually needs their concern. The main objective is to achieve the qualitative view for the information rather than being concerned for quantitative view. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness. Hence, we thought to design and develop a text mining tool that could be used for Educational purposes.

The tool will search for the important terms required by the teacher within the documents and highlight [3] them for quick reference. It will also generate the report and graph based on the search result. In other words, the tool will help teachers evaluate student's text from a qualitative point of view by determining whether all the required points are addressed or not. Thus, teachers who have to check the assignments and papers written by student can make use of this tool to quickly search the important points in e-documents and grade the students accordingly. The next section will give you brief description about current education system and problems with them.

II. EXISTING SYSTEM

The teachers usually check and evaluate students' answers manually. They search for the key points in the answers, check its correctness and based on their relevance in the text, grade the students. It is very tiresome, lengthy and prolonged process to check each student's answer and evaluate their performance.

Pitfalls in Existing System:
The current scenario of checking assignments and evaluating students' performance suffers from following pitfalls:
A. *Work Overload*: Manually checking and evaluating each and every student's performance increases a lot of work on teacher's part.
B. *Time Consuming*: Since teacher has to read full text of the student, it takes them a lot of time to check assignments and grade students.
C. *Human Errors*: There may be chances that in a hurry, teacher may overlook some points which the students may have addressed in their writings.

III. PROPOSED MODEL

The development of this model has been inspired by an actual need of university teachers who work with distant education and who have to review a large number of assignments produced by students. The idea is that teachers will already have a list of questions related to their subject and the important terms they expects in the answers. They will just have to select the digital text document written by the students that he/she wants to evaluate. The tool will quickly search

important terms in the text and highlight the sentences containing those words. So instead of going through the entire document, teacher can simply refer the highlighted points to evaluate the text. On the basis of the search result, report and graph will be generated that will help the teacher to grade the students appropriately. Architecture of our tool is shown in Figure 1. Detailed description of various modules is given as below.
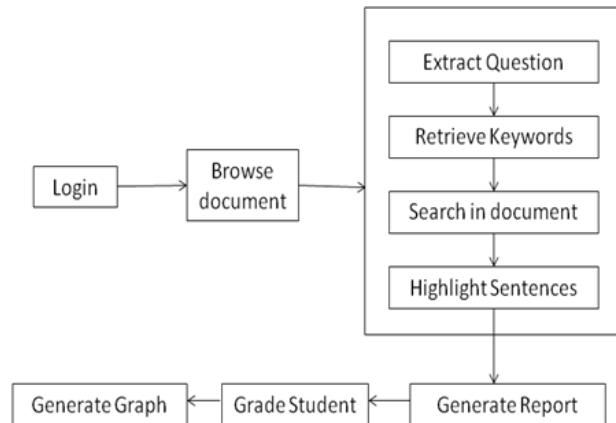

Figure 1 : System Architecture

A. **Login:** The Teachers can login into their account for access to the application. If they don't have an account, they can register for the same.
B. **Browse Document:** This module will allow the Teachers to browse the document they want to evaluate. They will gain access only to the documents that they are authorized to.
C. **Highlight Module:** It will extract all the questions contained in the document and retrieve the keywords required in the corresponding answers. After this processing is done, sentences pertaining to those keywords will be highlighted for quick review and evaluation.
D. **Report Generation:** Based on the evaluation, a performance report will be generated in the form of text file. The report will contain the statistics of the document specifying the frequency of the keywords and appropriate grades. The reports will be generated for each student who has submitted their assignments. These reports will get stored in related subject folder [5].
E. **Graph Generation:** Since graphical representation provides a better view than textual information, this module will depict the overall performance of the entire class in one particular subject. The graph will be generated based on the grades of the students. Various types of graphs such as bar graph, pie chart and line graph will be generated for graphical representation of student's performance.

The result of the proposed model is effectively represented in next section. Section 4 will give you the straight view of highlighted Document, report generation and graph generation [6] as outcome of our proposed model.

IV. RESULTS

A. **Highlighted Document:** As shown in figure 2. The highlighted Document gives concise view of important points.
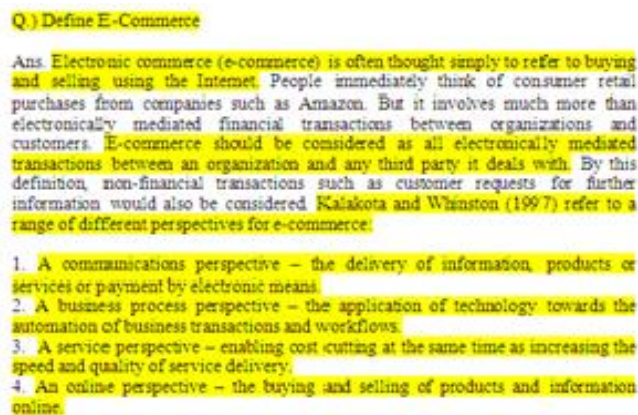

Figure 2: Highlighted Document

If the teacher does not select a document to evaluate, then he/she will be prompted to select a document. This module also works for the documents containing multiple questions. During the evaluation of document, "If Question not found in database" then teacher will be prompted to enter the keywords for that question not found in database.

B. **Report:** Figure 3 represents the report generated after evaluation of e-document. It gives the question, relevant key terms, and their frequency and performance grade in text file for each student in each subject.
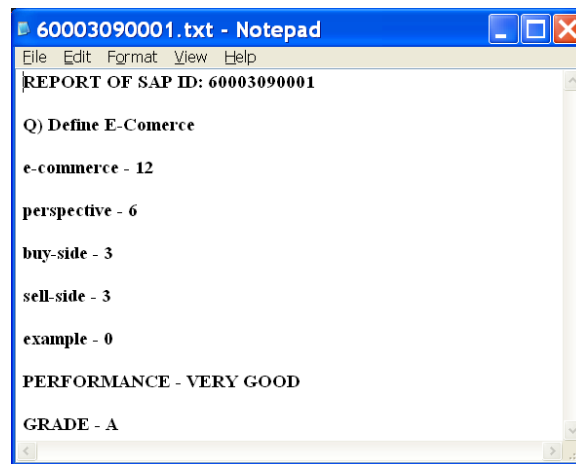
Figure 3: Report

C. **Bar Graph:** The graph shown in Figure 4 depicts the performance of the entire class. Teacher can select any type of graph such as pie chart or line chart. The numbers shown on each graph represent the count of students receiving that grade.
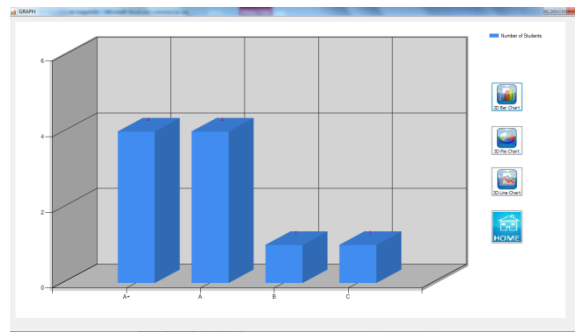

Figure 4: graphical representation of student's performance

V. CONCLUSION

Text mining is also known as Text Data Mining or Knowledge-Discovery in Text (KDT), it refers generally to the process of extracting interesting and non-trivial information and knowledge. As most information (over 80%) is stored as text, text mining is believed to have a high commercial potential value. The main contribution of this work has been to design a text mining mechanism for educational applications where we proposed a method to evaluate students' texts from quality standpoint and accelerate assessment process. Also graphical representation has been provided to exhibit a better depiction of student's performance. The results obtained from a preliminary evaluation of the system showed that the graphs obtained from the students' performance evaluation can lead teachers to further evaluate the students' work regarding certain problems, such as the need for additional development of a given subject.

FUTURE SCOPE

Educational field is very vast and can include enormous range of functions for better evaluation. Since our tool currently supports only digital documents, in future we can integrate it with OCR so that it can be used to evaluate hand-written papers as well. It can also be enhanced to assess objective-type questions. Since the tool is standalone application, it can be customized and made a client-server based application. Security can also be incorporated to meet user requirements. Since the tool is maintaining database for questions, it can randomly select question and assign different question to each students automatically. Automatic mark sheet generation can also be facilitated.

REFERENCES
[1] E. Reategui, M. Klemann, D. Epstein, A. Lorenzetti. "Sobek: a Text Mining Tool for Educational Applications". Research Paper, Data Mining International Conference, 2011
[2] Question Modification. Available at http://www.homeandlearn.co.uk/NET/nets12p1.html visited in January 2013
[3] Highlight Sentences. Available at http://visualbasic.ittoolbox.com/groups/technical-functional/visualbasic-l/fetching-the-current-line-of-a-searched-word-using-vbnet-3546681 visited in February 2013
[4] Word Counter. Available at: http://www.wordcounter.com/ visited in March 2013
[5] Report Generation. Available at http://www.homeandlearn.co.uk/net/nets8p4.html visited in March 2013
[6] Graph Generation. Available at http://www.linglom.com/2010/05/27/creating-graph-with-vb-net-part-3-data-binding-chart-to-database visited in April 2013