

# Automated Score Evaluation of Unstructured Text using Ontology

Badar Sami  
Department of Computer  
Science, University of  
Karachi,  
University Road, 75270

Huda Yasin  
Department of Computer  
Science, University of  
Karachi,  
University Road, 75270

Mohsin Mohammad Yasin  
Department of Computer  
Science, National University  
of Computing and Emerging  
Sciences, Karachi

## ABSTRACT

One of the important endeavors of Computer Science is its dealing with data and performing different responsibilities regarding analysis. In this paper, an ontology based automated score evaluation of unstructured text in the domain of text mining is presented. The use of ontologies in this respect is not old. For this research, we have dealt with different approaches and have also represented those methods which provide less optimized score as compared to our finally opted method. For our experimental work, we have collected real answers of students and then compared them with the model answer. We have found that our ultimate approach gives much more optimized end result as compared to other approaches which were carried out throughout our delve process. Moreover, the efficiency of result depends on the ontologies stored in the dataset.

## General Terms

Text Mining

## Keywords

Automated score evaluation, ontology, text data mining, unstructured text.

## 1. INTRODUCTION

At any level, assessment is a tough job. On closed queries, computerized assessment can be carried out in a triumphant manner. Table 1 enlightens some of such questions [1] [2] [3].

For many mentors, scoring of subjective questions is an exigent matter. Ranking for a descriptive question is generally based on the personal observation, acuity and understanding of the mentor. And explanation of a specified answer and the essential expected (standard) terminologies tinted in the answer scheme organized by the teacher. During major assessments, teachers are overloaded with a large number of answer sheets, so especially at that moment, marking can be an arduous process. If the answers are written manually then they have to read dissimilar forms of writing styles. The writing modes can diverge from excellent to horrible. After a lengthy period of the ranking process, both the physical fortitude and mental steadiness can be exaggerated [4] [5].

Free-text questions were assumed to be very knotty to score devoid of human intervention. So, they have conventionally been missing from automated tests. With the initiation of innovative technology, for example, enhancement in the area of natural language processing and information extraction [6], it is feasible to incorporate specific categories of free-text questions in automated tests because their trustworthy computerized scoring is now achievable. Foremost advantages of computerized free-text scoring comprise of time and price savings, and the idyllic shrinking in mistakes and injustice due to prejudice, exhaustion or deficiency of steadiness or reliability [7].

As compared to multiple-choice questions, descriptive questions want students to write their own answer, it also permits students to put across and prop up their thoughts in response to the question. Because of this, students can exhibit their various capabilities and talents like explaining analysis, describing his or her individual responses, producing their own assumptions, or sustaining vital estimations. But on the other hand, the grading of such descriptive questions is pricey and protracted. Moreover, it includes probable measurement fault to check outcomes because of discrepancies in the grading process [8].

**Table 1. Automatic assessment and closed questions**

➤ Best choice answers
➤ True or False.
➤ Diminutive rejoin.
➤ Numerical, a precise figure or a digit inside a range.
➤ Harmonizing, a two-column concept equivalent query.
➤ Calculated Queries: hit and miss values are placed into an equation; upshot varies each time the query is seen

The paper is organized as follows. In section 2, we have given a brief introduction of related work done by others. Section 3 presents the experimental results of different methods. In section 4, we have presented the result and section 5 concludes the paper with future studies.

## 2. RELATED WORK

Raheel Siddiqi and Christopher J. Harrison worked on the computerized evaluation of short free-text answers. They appraised two presently existing systems in order to classify their potentials and restrictions, and to draw attention to the domains in which upcoming associated research may be suitably directed. Moreover, they suggested that a general warehouse of standardized data sets be formed and made accessible to investigators and as well as to the implementers of the system, probably by means of some administrative influence, so as to make the advancement in this domain scrutinized and quantified [7].

Victor Gonzalez and Martin Llamas worked on the electronic appraisal of open questions based on the viewpoint of a teacher. They analyzed the kinds of questions appropriate for computerized grading and their effectiveness in the field of Engineering. They investigated some representative accessible computerized grading systems. They also stated the necessities of educators, stated the art potential and suggested some opinions for the design of an automated assessment of open question grading system based on a restricted lingo for answer scripting along with the comparison of comprehension representation for appraisal [9].

Yigal Attali et al. worked on the computerized grading of short answer open ended GRE subject assessment. They explained the progression, management, and grading of open ended modifications of GRE subject assessment objects in the field of biology and psychology. These questions were supervised in a Web-based testing in order to record exams of the particular subject. 1-3 sentences were requisite by these questions, and answers were marked without human intervention by natural language processing approaches, c-rater grading engine was utilized, directly after applicants presented their answers. Instantaneous feedback was given to the applicants so as to tell them regarding the appropriateness of their responses; chance was also given to amend their responses [8].

### 3. EXPERIMENTAL WORK

This section provides fundamental understanding regarding stop or noisy words, connectivity with lexical database and elucidates different research based subjective algorithms along with their domino effects that have been scrutinized with the progression of research.

#### 3.1 Removal of Noisy Words

Noisy words like is, an, the, or etc have no significance or impact on unstructured text. We have detached such words in order to acquire optimized end result. In Japanese, noisy/stop words identification is based on grammatical information. As an exemplar, project search makes out whether the utterance is a noun or a verb, whereas the other languages deal with specific lists [10].

#### 3.2 Brief Introduction To Word Net

Word Net is a huge lexical database of English. Different parts of speech like nouns, verbs, adjectives and adverbs are clustered into series of cognitive synonyms which is also referred to as cognitive synsets, each conveying a diverse perception. Synsets are interconnected via conceptual-semantic and lexical associations. The resulting set of connections of meaningfully associated words and thoughts can be directed through browser. Configuration of Word Net makes it a constructive tool for both computational linguistics and natural language processing [11].

#### 3.3 Initial Research Based Subjective Algorithm

Here, we have described different research based algorithms applied for subjective marks estimation.

##### 3.3.1 Utterances Matching On The Basis Of POS

In this method, we extracted noun, verb, adverb and adjective from both model answer and student's answer. For doing so, each and every term from both model and student's answer goes into the lexical database in order to obtain its parts of speech (POS) i.e. noun, verb, adverb and adjective. On the basis of POS, the terms put into their specified vectors. At a time, a single term can be a noun and a verb as well. For instance, "convert" is a noun and a verb too. In this case, both noun and verb vectors would get it. After completing this procedure, percentages will be calculated according to given formula:

$$\text{Noun} = \frac{\text{Matched Noun}}{\text{Total number of nouns in model answer}} \times 100$$

$$\text{Verb} = \frac{\text{Matched Verb}}{\text{Total no. of verbs in model answer}} \times 100$$

$$\text{Adverb} = \frac{\text{Matched Adverb}}{\text{Total no. of adverbs in model answer}} \times 100$$

$$\text{Adjective} = \frac{\text{Matched Adjective}}{\text{Total no. of adjectives in model answer}} \times 100$$

The average percentage of the above four percentages give us the overall percentage of the student's answer.

##### 3.3.2 Comparing Student's Sentences with All the Sentences of Model Answer

To understand this method more clearly, let's consider that there are 4 sentences in student's answer and 4 sentences in a model answer. The initial step of the procedure can be shown with the help of Figure 1.

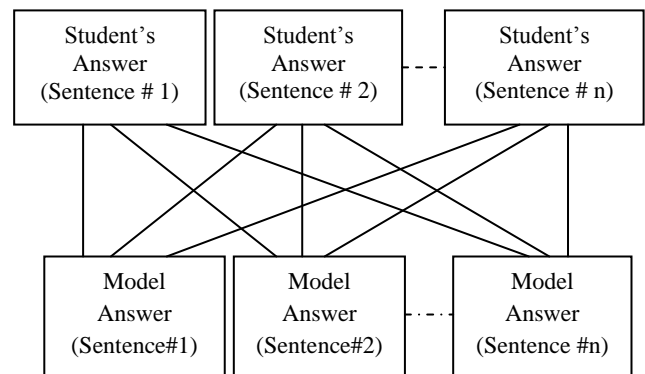


Fig 1: Comparison between sentences

In this method, percentage is obtained by:

$$\frac{\text{Matched Words}}{\text{Total number of content words in model answer}} \times 100$$

All sentences of student's answer compares with all the sentences of model answer. Through this, we would acquire four different percentages. The highest percentage would be chosen and this percentage will be placed in a vector. After comparison, all percentages of each and every sentence are shown in Table 2.

In Table 2, it can be noticed that (1, 1) and (1, 3) have same percentages. In this case, we will opt on the basis of their occurrence i.e. first come first opt. It can also be observed that the percentage of (4, 1) is 80.01 is greater than the percentage of (1, 1) which is 77.31. In this condition, percentage of (1, 1) will be discarded. This average percentage gives us the overall percentage of student's answer.

We move on towards other approaches in order to achieve more optimized domino effects.

**Table 2. Obtained percentages**

(Student's Sentence, Model Sentence)	Percentage %
(1,1)	77.31
(1,2)	31.14
(1,3)	77.31
(1,4)	21.42
(2,1)	11.2
(2,2)	33.24
(2,3)	12.3
(2,4)	18.01
(3,1)	19.23
(3,2)	16.3
(3,3)	3.23
(3,4)	5.67
(4,1)	80.01
(4,2)	20.21
(4,3)	11.2
(4,4)	10.3

**3.3.3 Similarity Measure Using Jaccard's Coefficient**

In this method, we have followed the way as mentioned in 3.3.2. But here, we have employed 'Jaccard's Coefficient' as similarity measure:

$$Jaccard's\ Coefficient = sim(H_y, H_z) = \frac{\sum_{j=1}^i H_{yj} H_{zj}}{\sum_{j=1}^i H_{yj}^2 + \sum_{j=1}^i H_{zj}^2 - \sum_{j=1}^i H_{yj} H_{zj}}$$

**3.3.4 Similarity Measure Using Dice's Coefficient**

This method too follows the same way as mentioned in 3.3.2. The difference crops up while measuring similarity. 'Dice's

Processed Points of 1st Student	Processed Points of 2nd Student
$\frac{7.28}{7.28}$	$\frac{4.61}{7.28}$
$\Rightarrow 1$	$\Rightarrow 0.6332$

'Coefficient' is used as similarity measure in this approach:

$$Dice's\ Coefficient = sim(H_y, H_z) = \frac{2 \sum_{j=1}^i H_{yj} H_{zj}}{\sum_{j=1}^i H_{yj}^2 + \sum_{j=1}^i H_{zj}^2}$$

**Comments:** From the four different techniques described in section 3.3, we have make out that the domino effects acquired from these approaches are satisfactory up to some extent. But the upshot can be more optimized by adopting some other way.

**3.4 Ontology Based Subjective Marks Evaluation**

**3.4.1 Succinct Introduction to Ontology**

The theme of 'ontology' is the study of groups or classes of things that present or may present in some field. Ontology is a list of the kinds of things that are supposed to be present in an area of interest **H** from the viewpoint of an individual who uses a lingo **L** for the purpose of talking concerning **H**. The modes in the ontology symbolize the predicates, word senses, or conception and relation kinds of the lingo **L** when used to talk about subjects in the domain **H**. An unexplained logic, for instance, predicate calculus or conceptual graphs is ontologically impartial. It inflicts no restrictions on the subject issue. Logic lonely says nothing regarding anything, however the grouping of logic by means of ontology endow with a language that can express associations concerning the entities in the area of interest [12].

**3.4.2 Final Method**

We have used this methodology in the domain of 'text mining' i.e. we maintained a database which consists of text mining based ontologies.

The steps involved in this method are as follows:

- Exact word matching
- Ontology based matching
- Nominalization based matching
- Synset based matching
- Synonym based matching

After passing the above five steps, we applied the following formula:

$$Percentage = \frac{Matched\ words}{Total\ content\ words\ in\ model\ answer} \times 100$$

**Negative Aspect:** Matched words can be greater than total contents words present in model answer. Due to this drawback, we made some alterations. Now, the formula has taken the following form:

$$Points = \frac{Matched\ words}{Sentence\ length} \times Content\ words$$

At this instant, consider that there are two students who have given the answer of any particular question. Using above formula, assume that they have obtained '7.28' and '4.61' points. It is noticeably shown that 1<sup>st</sup> point is greater than 2<sup>nd</sup> point. Now, the method proceeds in the following way:

Suppose, there are 25 content words in the answer of 1<sup>st</sup> student and 20 content words in the model answer. The total content words present in the answer of 1<sup>st</sup> student are greater than the total content words of model answer. The method will move in the following way:

Marks =

$$Processed\ Points \times \frac{Content\ Words\ of\ Student's\ Answer}{Content\ Words\ of\ Model\ Answer}$$

For this particular example:

$$\text{Marks} = 1 \times \frac{25}{20}$$

$$\text{Marks} = 1.25$$

The above score considered as the ultimate score acquired by the 1<sup>st</sup> student.

For 2<sup>nd</sup> student, consider that there are 15 content words present in his/her answer which is lesser than the content words present in the model answer. Therefore:

$$\text{Marks} = 0.6322 \times \frac{20}{15}$$

$$\text{Marks} = 0.844$$

The above score consider as the final marks obtained by the 2<sup>nd</sup> student.

**Comments:** This final method provides us an efficient and professional mechanism for the evaluation of subjective answers.

#### 4. RESULT

So as to validate the planned approach, the methodology has been analyzed via checking the real answers of large number of students in the domain of text mining. In the initial step of the testing part, a comparison is made flanked by the model answer of the question and all the students' answer of the question. We have found that this final method gives us much more optimized end result as compared to other approaches which were carried out throughout our delve process. Plus, the efficiency of result depends on the ontologies stored in the dataset.

In our proposed method, we have shown that relationship exist between the highest marks and the marks of all the students. If the mentor feels that the maximum marks of a student should be more or less then he can easily change the score. This change would also affect the marks of rest of the student.

#### 5. CONCLUSION AND FUTURE WORK

Our proposed algorithm gives an efficient way for the evaluation of descriptive answers. The accuracy of marks depends on the stored ontologies of any specific domain and as well as on the model answers. Besides, association exists between the maximum marks and rest of all the marks. In future, we will broaden this study to acquire more enhance domino effects by using different text mining algorithms. In addition, we will apply fuzzy learning models for more enhance assessment of descriptive answers.

#### 6. ACKNOWLEDGMENT

We would like to thanks Mr. Hassan Muhammad Yasin who contributed in the evaluation of our research algorithm.

#### 7. REFERENCES

- [1] Sugato Basu, Raymond J. Mooney, Krupakar V. Pasupuleti, and Joydeep Ghosh, 2011, Evaluating the Novelty of Text-Mined Rules Using Lexical Knowledge, Department of managerial economics, strategy and innovation (msi)
- [2] Berry, M. W. (Ed.), 2003, Survey of text mining. New York: Springer
- [3] Tom Magerman, Bart Van Looy, Bart Baesens, and Koenraad Debackere, 2011, Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents
- [4] Miao Chen and Klaus Zechner, 2011, Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 722–731, Portland, Oregon, Association for Computational Linguistics
- [5] Elizabeth Dalton, 2002, Beyond Multiple Choice: Computer-Mediated Practices and Assessments to Support Higher-Order Objectives,
- [6] D. Perez-Marin, I. Pascual-Nieto, E. Alfonseca and P. Rodriguez, Automatic Identification of Terms for the Generation of Students' Concept Maps
- [7] Raheel Siddiqi and Christopher J. Harrison, 2008, On the automated assessment of short free-text responses, 34th annual conference
- [8] Yigal Attali, Don Powers, Marshall Freedman, Marissa Harrison, and Susan Obetz, 2008, Automated Scoring of short-answer open-ended GRE subject test items; ETS GRE research report No. 04-02
- [9] Victor Gonzalez-Barbone, Martin Llamas-Nistal, eAssessment of Open Questions: an Educator's Perspective
- [10] List of Stop Words. Available at: <http://www.lextek.com/manuals/onix/stopwords1.html>
- [11] Word Net- A large lexical database. Available at: <http://wordnet.princeton.edu/>
- [12] S. Bloehdorn, P. Cimiano, A. Hotho, and S.Staab, 2005, "An Ontology-based Framework for Text Mining", LDV Forum – GLDV Journal for computational linguistics and language technology, Vol.20, No.1