Thesis Proposal

Automatic Subjective Answer Sheet Checker by Machine Learning

Thesis Supervisor

Dr. Md Ashikur Rahman Khan

Professor

Chairman, Dept of ICE

Dean, Faculty of Engineering

Noakhali Science & Technology University

Submitted by

Md. Nurul Amin

ID: ASH1811021M

Dept of ICE

Noakhali Science & Technology University

Date: 29th August, 2022

Department of Information & Communication Engineering

Noakhali Science & Technology University, Noakhali-3814, Bangladesh

DECLARATION

This research proposal is submitted to the Department of Information and Communication Engineering of Noakhali Science & Technology University in partial fulfillment of the requirements for having the Bachelor of Science (BSc) Engineering in Information and Communication Engineering (ICE). So, I hereby declare that this research proposal has not been submitted elsewhere for the requirement of any kind of degree, diploma, or publication.

Md. Nurul amin
ASH1811021M

ACCEPTANCE

This Thesis Proposal is submitted to the Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali- 3814 in partial fulfillment of the requirements for having the B.Sc. degree in ICE. This proposal report will be evaluated under the Course: “Project and Thesis” with the Course Code ICE-4110.

Dr. Md Ashikur Rahman Khan

Professor

Chairman, Dept of ICE.

Dean, Faculty of Engineering.

Noakhali Science & Technology University.

Abstract

The current way of checking subjective paper is adverse. A subjective paper evaluation is a tricky and tiresome task to do by manual labor. When a human being evaluates anything, the quality of evaluation may vary along with the emotions of the person. In Machine Learning, all result is only based on the input data provided by the user. The main objective of the journal is to especially reduce manpower and time consumption. Besides not affect the evaluation process by human emotion. Our proposed system uses machine learning, image processing, and NLP to solve this problem. This system is divided into two modules. The first one is extracting the data from the scanned images and organizing it properly and the second is applying ML, image processing, NLP, and tools such as Wordnet, Word2vec, word mover's distance (WMD), cosine similarity, multinomial naïve Bayes (MNB), and term frequency-inverse document frequency (TF-IDF) to evaluate descriptive answers automatically. The system will let students give exams online, calculate the results automatically as well as store a record for the administrator. Moreover, this system brings some revolutionary changes to the offline examination system.

TABLE OF CONTEXT

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

A student's performance and abilities can be evaluated in an open-ended way using subjective questions and replies. Naturally, there are no restrictions on the answers, and students are allowed to construct them following their perspectives and conceptual understanding. Having said that, there are still several crucial distinctions between subjective and objective solutions. They are longer than the objective questions, for starters. Second, writing them requires more time. Additionally, they require a lot more focus and neutrality from the teacher grading them because they contain a lot more context.

Checking responses demands intense concentration for an extended period, which frequently results in mistakes. The automated increase in the effectiveness of answer evaluation by completing this activity in a significant way. It was clarified during a quick conversation that the response sheet is assessed with consideration for a specific keyword that moderators use to find the solution assessing a response. Our suggested algorithm will call for using keyword inputs. These keywords will be supplied by an expert in the field. Our suggested algorithm will correspond to these derived from the keywords with identified words utilizing a supervised learning technique. The learning phase of the model will need a handwritten dataset for phase alphabets used in English. These datasets can be obtained from a variety of web sources in order to train the model.

A student needs to answer any figure (graph or anything) on a separate page. One has to scan the paper. Then the system extracts the words. For evaluation, these words need to be matched with the correct answer's words. The correct answer for each question will be provided by the user. Besides, the correct answer should have both positive and negative forms. If a student's answer matches one form, the answer will be considered as correct. On the other hand, figures will be compared with the correct figure provided by the user by the image processing technique. Mark from the text part and mark from the figure part will be combined for producing the final mark.

There are numerous reasons why incorporating artificial intelligence into the systems used for online exams would be very beneficial. First off, as tests are now marked by examiners, this causes fatigue and boredom as they have to review several answer sheets. However, with the online approach, this problem is instantly resolved. Additionally, a human would need hours to complete what a computer can do in minutes with accuracy and speed. The proposed system would also produce unbiased results which would further make everything more transparent.

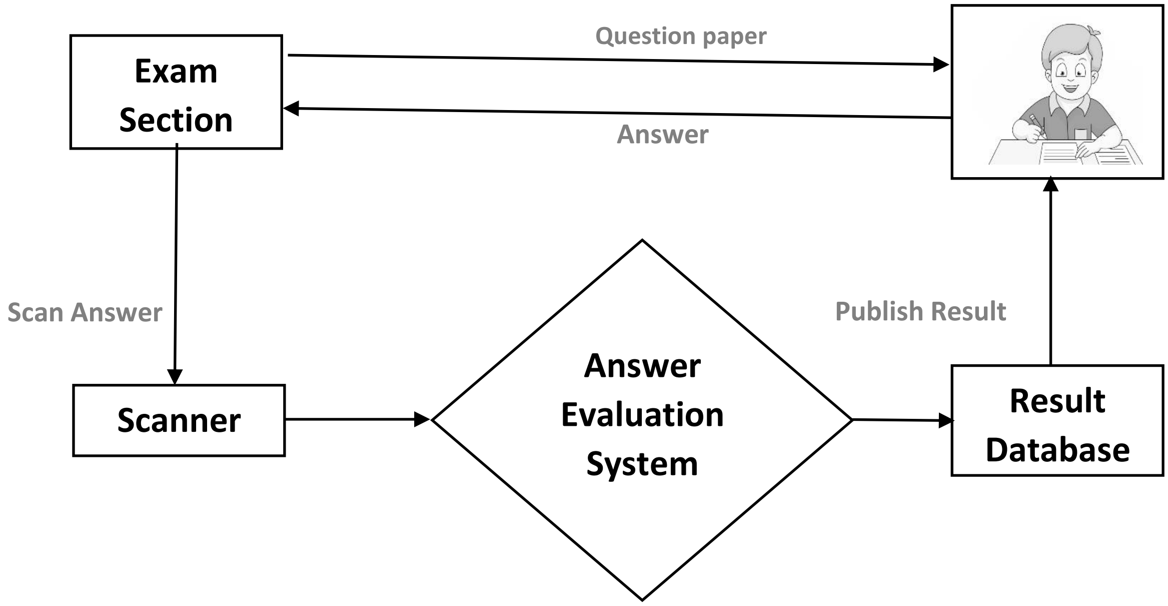


Figure 1.1: Overall architecture of the system

1.2 Problem Statement

Subjective exams are considered more complex and scary by both students and teachers due to their one fundamental feature, context. A subjective answer demands the checker check every word of the answer for scoring actively, and the checker's mental health, fatigue, and objectivity play a massive role in the overall result. Therefore, it is much more time and resource-efficient to let a system handle this tedious and somewhat critical task of evaluating subjective answers.

Evaluating objective answers with machines is very easy and feasible. A program can be fed with questions and one-word answers that can quickly map students' responses. Nevertheless, subjective answers are much more challenging to tackle. They are varied in length and contain a vast amount of vocabulary. Furthermore, people tend to use synonyms and convenient abbreviations, which makes the process that much more tricky.

Besides students may write Mathematical expressions. It is a great challenge to evaluate mathematical terms like formula or full solve. Sometimes students add a relatable figure or questions demand some figure (Ex- graphs). In that case, it becomes difficult to evaluate figures combined with the main textual answer. We are going to overcome all those walls with our proposed system.

1.3 Motivation

This type of automated review will let the educational sector carry out its other responsibilities more effectively and eliminate human labor for menial chores such as contrasting the responses with the ideal response. This results in more time being spent by teachers preparing lessons for students. improved instruction and less-human testing evaluation more transparency and mistakes.

Traditional answer evaluation manual checking of answers takes a lot of time and energy. However, till now multiple choice questions can be evaluated with the use of computers. When it comes to the theoretical evaluation of answers, there is a need for a teacher to check the answer sheet.

Hence teacher has to put his effort into answer sheet evaluation rather than providing knowledge to the students. Also, the traditional answer evaluation takes more time than that of the machine. There is a limitation when it comes to a human. He can work for a limited amount of time. After that, he must take a rest. But that is not the case with the machine. A machine can work 24*7 nonstop to give the output. The machine also eliminates human error and hence it can be reliable.

That's why, we came up with the solution for this problem so that, it will save time and resources.

1.4 Objectives

1. Reduce Human effort
2. Brings figures (graph/ table) and mathematical terms in the evaluation process which may be untouched in previous work.
3. Make the evaluation process transparent.
4. Train and build a system capable of evaluating any type of subjective answer sheet.
5. Work with hand-written answer sheets by the student.
6. Show results using different algorithms and methods.

Chapter 2

LITERATURE REVIEW

In this Section, we review some closely related work based on some different basis.

2.1 Typed answer evaluation

There are several works that deal with the typed document. In that case, the Student submits their answer directly in the answer section for the corresponding Question. The author of [17] collected the data set in the very first step which consists of answers to the questions in the question paper. • Upon collecting the data, all the text in the data is converted to lowercase. • After the conversion to lowercase, word tokenization is performed on the text. Word tokenization is the process of splitting a large sample of text into words. This is a requirement in natural language processing tasks where each word needs to be captured and subjected to further analysis like classifying and counting them for a particular sentiment etc. The Natural Language Tool kit (NLTK) is a library used to achieve this. • Moving forward, the next important step that is performed is the removal of stop words and punctuations. A stop word is a commonly used word (such as “the”, “a”, “an”, or “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. We would not want these words to take up space in our database, or take up the valuable processing time. Author [18,19,20] approaches almost the same process.

Some other author works with only one sentence answer evaluation. The author of [20] thought about only some 1 line answers for some online quizzes. He used the processed data is then converted into vectors using Embedding and is then passed through cosine formula which provides the similarity between the two. On the basis of the value of similarity achieved (between 0 and 1), marks are rewarded for each provided answer and the final result is calculated by adding them all. The author of [10,22] also thought the same thing.

In [7,21,24] author deals with some typed documents of laws to check their similarity and check plagiarism.

2.2 Handwritten Answer evaluation

2.2.1 Semantic and Synonym-based evaluation

The author of [7,30] presents an automated system that addresses the following problems with assessing subjective questions: synonymy, polysemy, and trickiness. Latent semantic analysis (LSA) and the ontology of a subject are introduced to solve the problems of synonymy and polysemy. A reference unit vector is introduced to reduce the problem of trickiness. The system consists of two databases: a science knowledge library and a question- and reference answer library. The science knowledge library stores the ontology of a subject as text documents. The question- and reference-answer library stores questions as text documents and reference answers as a text document matrix. When a teacher adds new questions, a system using this science knowledge library will search for related points of knowledge and keywords and give them to the teacher. Then, the teacher will submit the reference answer to the system. It will process the reference answer using Chinese automatic segmentation, which produces text document vectors and sends them to the teacher. Then, the teacher detects the terms and their weights for each vector and sends them back to the system. Weights of the terms in the reference answer are computed using the term-frequency and inverse-document-frequency functions. In the questions and reference answers, the library will save the vector of the reference answers and questions as text documents. To compute the similarity between a student's answer and the reference answer, the former is sent to the system, which assesses the answer using Chinese automatic segmentation and produces a text vector projected into k-dimensional LSA space. This LSA is formed by a vector using the mathematical technique of singular value decomposition (SVD), which represents terms and documents that are correlated with each other. The system computes the cosine similarity of student and reference answer vectors projected into k-dimensional LSA space in the reference unit vector.

Similar types of processes are followed by the author [2,3,5,28]. But they used different algorithms and methods for getting better outcomes.

2.2.2 Keyword matching-based Evaluation

The author of [25] Provide that, answer sheet to the system in jpeg (.jpg) format. Provide keywords, maximum marks, and minimum length required for the answer. The system will separate words from the given answer. The given words will be stored in a .csv file. The length of the answer will be calculated by counting words from the CSV file. Check the percentage of keywords matched. Check the percentage of words written compared to that of minimum length. Check the percentage of marks allotted for the given percentage of keywords matched from the graph. Check the percentage of marks allotted for the given percentage of word length from the graph. Multiply both the percentage of the maximum marks for the answer. Display the marks obtained.

The author of [25] has got the most accurate result I have studied ever. Some other authors [26,27] also proceed with a similar method, but they use different algorithms.

2.3 Related works

- Hu and Xia [1] proposed a Latent Semantic Indexing approach for the assessment of subjective questions online. They used Chinese automatic segmentation techniques and subjective ontologies to make a k-dimensional LSI space matrix. The answers were presented in TF-IDF embedding matrices, and then Singular Value Decomposition (SVD) was applied to the term-document matrix, which formed a semantic space of vectors. LSI played the role of reducing problems with synonyms and polysemy. At last, the Similarity between answers was calculated using cosine similarity. Dataset consisted of 35 classes and 850 instances marked by teachers, and the results showed a 5% difference in grading done by teachers and the proposed system.
- Kusner et al. [2] presented a novel concept of using Word Mover's Distance (WMD) to find the dissimilarity between two texts. The system used no hyper-parameters and used a relaxed WMD approach to loosen up the vector space bounds. Dataset included eight real-world sets, including Twitter sentiment data and BBC sports articles. Word2vec model from google news was used, and two other custom models were trained. KNN classification approach was used to classify the testing data. As a result, relaxed WMD reduced the error rates and led to 2 to 5 times faster classification.
- Kim et al. [3] proposed a method to grade short descriptive answers lexico-semantic pattern (LSP) due to its good performance with the morphologically complex Korean language. LSP can structure the semantics of the answer to help understand the user's intentions. A synonym list was also utilized to help expand the keywords so they match various answer styles. Dataset was obtained from 88 students and converted to LSP, which was later compared with the solution LSP to score the answer. As a result, the system performed better than the existing system by 0.137.
- Oghbaie and Zanjireh [4] proposed a pair-wise Similarity measure to measure the similarity between two documents based on the keywords which appear in at least one of the documents. The work proposed a new similarity measure called PDSM (pair-wise document similarity measure), a modified version of the preferable properties approach. The proposed similarity measure was applied to text mining applications such as document detection, k Nearest Neighbors (KNN) for single-label classification, and K-means clustering. An evaluation measure of accuracy was used, and as a result, the PDSM method produced better results than other measures like the Jaccard coefficient by 0.08 recall.
- Orkphol and Yang [5] used the word2vec approach to represent words on a fix-sized vector space model and then measured the Similarity of sentences using a cosine similarity measure. Word2vec from google was used, and the sentence vector was obtained as a result of an average of words in the sentence. The score was accepted if it passed a specified threshold for similarity results, between 0 and 1. Evaluation measure of recall and accuracy was used, and as a result, the system's performance was 50.9% with and 48.7% without the probability of sense distribution.
- Xia et al. [6] combined the word2vec approach with the legal document corpus to identify similarities between different law documents. Cosine similarity was used to measure the Similarity between different sentence vectors. As a result, word2vec improved the accuracy by 0.2 compared to the Bag of Words approach, which could further be increased by 0.05-0.10 by training the word2vec model on law documents.

- Wagh and Anand [7] proposed a multi-criteria decision-making perspective to find the Similarity between legal documents. The work included using Artificial Intelligence and aggregation techniques such as ordered weighted average (OWA) for obtaining the similarity value between different documents. Dataset was obtained from Indian Supreme Court case judgments from years ranging from 1950 to 1993. Evaluation measures of score and recall were used. As a result, a concept-based similarity approach such as the one proposed in the work performed better than other techniques such as TF-IDF, getting an F1 score of up to 0.8.
- Alian and Awajan [8] studied various factors affecting sentence similarity and paraphrasing identification using different word embedding models, clustering algorithms, and weighting methods to find the context of sentences. Pre-trained embeddings included AraVec and FastTex, both trained for the Arabic language. The Arabic training dataset included around 77,600,000 tweets. As a result, pre-trained embedding with labeled data from experts provided better recall and precision of 0.87 and 0.782 for K-means and agglomerative clustering.
- Muangprathub et al. [9] proposed a novel approach to plagiarism detection using formal concept analysis (FCA). The work showed formal context in FCA, starting with two sets containing elements with some attributes that somehow relate the element to its set. The documents and their shared keywords formed a group set in FCA whose values are typically but not limited to 0 and 1. The approach used a many-valued context. The work also introduced a new similarity concept that uses both the object extent and attribute intent. The approach used is not normally utilized in similarity analysis and ranks similar documents because they have similar object and attribute intents. The proposed system detected plagiarism in documents with 94% accuracy.
- Jain and Lobiyal [10] proposed a novel approach for subjective evaluation using concept graphs. Concept graphs were created for both the solution and the answer, and the score was evaluated using various graph similarity techniques. Montes et al. [38] explained various techniques to find Similarities between concept graphs and information retrieval from such graphs.
- Bahel and Thomas [11] presented an architecture for the evaluation of subjective questions using text summarization, text semantics, and keywords summarization and compared the results with existing approaches. The results showed an error of 1.372 compared to a 1.312 error from Jaccard's similarity approach. The approach, however, failed to compute non-textual data such as diagrams, images, and other formats.
- Sheeba Praveen [12] says that in recent years it has been seen that several government and semi-government examinations have gone online, for example [IBPS Common Written Examination (CWE)]. According to the professor, in recent years, we can make online test provisions for objective-type questions. As the online system has many more advantages over the offline examination. Online exam saves time to evaluate the answers. As there was no provision for offline paper evaluation, he designed an algorithm that automatically evaluate the single sentence descriptive answer
- B Vanni, M. shyni, and R. Deepalakshmi [13] OCR refers to translating handwritten text to a format that is machine-readable and can be used for searching, editing, and indexing. This paper is using the artificial neural network to achieve high accuracy for optically recognizing

the character. The proposed approach is tested and implemented on a character database consisting of English characters, digits, and special characters.

- Another approach by Baoguang Shi, Xiang Bai, and Cong Yao [14] combines the capabilities of CNN (Convolutional Neural Networks) and RNN (Recurrent Neural networks). The convolutional layers in this case are constructed by considering only the convolutional and max-pooling layers from a standard CNN model (fully-connected layers are removed). As CNN can identify the individual characters in the answer. But the variable length of answers is a problem for CNN to identify as the length of inputs is fixed. To solve this problem RNN is used with LSTM (Long-Short Term Memory) to save a long sequence of characters in the string. This model was termed CRNN (Convolutional Recurrent Neural Network). In CRNN, deep features are represented in sequential representations to be invariant to the length variation of sequence-like objects.
- Yusuf Perwej and Ashish Chaturvedi [15] proposed a neural network for handwritten English alphabet recognition. In this paper, they proposed and developed a scheme for handwritten character recognition by representing each English alphabet in binary values that can be used as input to the system and the output produced by the system then can be used as input to the neural network system. They have explained two phases namely pre-processing phase and the neural network-based recognition phase. In processing, only 25 bits are used because the alphabets are divided into 25 segments. They have achieved an accuracy of 82.5% but the system found less accuracy for similar alphabets.
- J.Pradeep , E.Srinivasan, and S. Himavathi [16] worked on an off-line handwritten alphabetical character recognition system that uses a multilayer feed-forward neural network. A new method other than the horizontal and vertical method was introduced called diagonal-based feature extraction. Fifty datasets were taken to train the network. Each dataset contained 26 alphabets written by different people. The system performed quite well with higher levels of accuracy in recognizing the alphabet.

Table 2.1

The summary of the Related work.

Ref.	Method used	Contribution	Limitation	Difference to us
[1]	Find the Similarity between answers was calculated using cosine similarity	Dataset consisted of 35 classes and 850 instances marked by teachers, and the results showed a 5% difference in grading done by teachers and the system.	Doesn't support handwritten answer sheets. Nothing about mathematical expression and figure evaluation	Our proposed system will work with a handwritten answer. We are considering mathematical expression and figure evaluation
[2]	Using Word Mover's Distance (WMD)	Find the dissimilarity between the two texts. WMD reduced the error rates and led to 2 to 5	No Evaluation, just measures dissimilarity	Evaluates considering the similarity between the correct answer and the student's answer

		times faster classification.		
[4]	PDSM (pair-wise document similarity measure), k Nearest Neighbors (kNN) for single-label classification.	proposed a pair-wise Similarity measure to measure the similarity between two documents based on the keywords which appear in at least one of the documents.	Nothing about mathematical expression and figure evaluation	We are considering mathematical expression and figure evaluation
[6]	combined the word2vec approach with the legal document corpus. Cosine similarity was used to measure the Similarity between different sentence vectors	identify similarities between different law documents.	Doesn't support handwritten answer sheets. Nothing about mathematical expression and figure evaluation	Our proposed system will work with a handwritten answer. We are considering mathematical expression and figure evaluation. Our System will work in the education sector, not in the law sector
[7]	Evaluation measures of F1 score and recall were used	The work included using Artificial Intelligence and aggregation techniques such as ordered weighted average (OWA) for obtaining the similarity value between different documents.	Doesn't support handwritten answer sheets. Nothing about mathematical expression and figure evaluation	Our proposed system will work with a handwritten answer. We are considering mathematical expression and figure evaluation. Our System will work in the education sector, not in the law sector
[8]	using different word embedding models, clustering algorithms, and weighting methods to find the context of sentences	studied various factors affecting sentence similarity and paraphrasing identification	Doesn't support handwritten answer sheets. Nothing about mathematical expression and figure evaluation	Our proposed system will work with a handwritten answer. We are considering mathematical expression and figure evaluation. Our System will work in the education sector.
[9]	using formal concept analysis (FCA)	The proposed system detected plagiarism in documents with 94% accuracy.	Doesn't support handwritten answer sheets. Limited scope.	Our System will work in the education sector for Subjective answer evaluation.
[10]	A novel approach for subjective questions evaluation using concept graphs.	The score was evaluated using various graph similarity techniques.	Nothing about mathematical expression. Similar words may not be detected. Word sentiment not in concern	Proper statement of mathematical expression. Similar words may be detected. Word sentiment is a concern

[12]		designed an algorithm that automatically evaluates the single sentence descriptive answer	Can't deal with long descriptive answers.	Will successfully evaluate long descriptive answers.
[13]	translate handwritten text to a format that is machine readable	This paper is using the artificial neural network to achieve high accuracy for optically recognizing the character.	Nothing about mathematical expression and figure evaluation	We are considering mathematical expression and figure evaluation

2.3 What will be our contribution?

All the papers we have studied yet. There is no paper that properly evaluates figures (Ex-graphs) and mathematical terms. We know that Science and Engineering students or any other students need to draw some figures like:

- Block diagrams
- Objects
- Geometrical shapes
- Structures.

We will not get accurate marking if we can't evaluate those figures properly. Our proposed system will be capable of evaluating those figures.

Besides including math all engineering subjects contain math problems. Questions are demanding some mathematical solution for the corresponding question. We will not get accurate marking if we can't evaluate that math properly. Our proposed system will be capable of evaluating those mathematical terms.

Chapter 3

Methodology

This project is an application for automated answer evaluation using the matching keyword from a dataset based on a machine learning algorithm. Some applications are available but they are different from this and they use a different methodology. Some available application only evaluates MCQs (multiple choice questions) not the subjective question. For using this application only one has to scan the answer to that question then the system will split the answer keyword using OCR [1]. Based on keywords written in the answer and the keywords in the dataset, the application will provide marks in the range of 0 to 5.

3.1 Pre-Processing:

Get the correct answers from the teacher/ expert for each question. Do the same for all questions.

- Store all possible keywords with their synonyms.
- Keywords must be both positive and negative sentiments.
- Store possible figures as images (.jpg).
- Store all possible mathematical formula
- Store all possible final results of any mathematical problem.

3.2 Student answer format:

- **Only text:** Answer must start from e new page.
- **Only Figure:** must be on a new page
- **Only mathematical statement:** Answer must start from e new page
- **Text + figure:** Answer must start from e new page as well as the figure also start on a new page.
- **Text + mathematical statement:** Answer must start from e new page as well as the mathematical statement also start on a new page.
- **Figure + mathematical statement:** Answer must start from e new page as well as the figure statement also start on a new page.
- **Text + Figure + mathematical statement:** All three must start from a new page

3. 3 Marking standard:

The System will provide marks in the range of 0 to 5.

3.4 Step-by-step process

1. Provide answer sheet to the system in jpeg (.jpg)/ pdf (.pdf) format.
2. Provide keywords, maximum marks, and minimum length required for the answer.
3. Provide both positive and negative sentiment for all keywords and all possible synonyms.
4. Store possible figures as images (.jpg).
5. Store all possible mathematical formula
6. Store all possible final results of any mathematical problem.
7. The system will separate words from the given answer.
8. The given words will be stored in .csv file.
9. The length of the answer will be calculated by counting words from the CSV file.
10. Check the percentage of keywords matched
11. Check the percentage of words written compared to that of minimum length.
12. Check the percentage of marks allotted for the given percentage of keyword matched.
13. Check the percentage of marks allotted for the given percentage of correct grammar.
14. Check the percentage of marks allotted for the given percentage of correct math result.
15. Check the percentage of marks allotted for the given percentage of math formula.
16. Check the percentage of marks allotted for the given percentage of word length from the graph
17. Multiply both the percentage of the maximum marks for the answer
18. Display the marks obtained

3.5 Flow chart

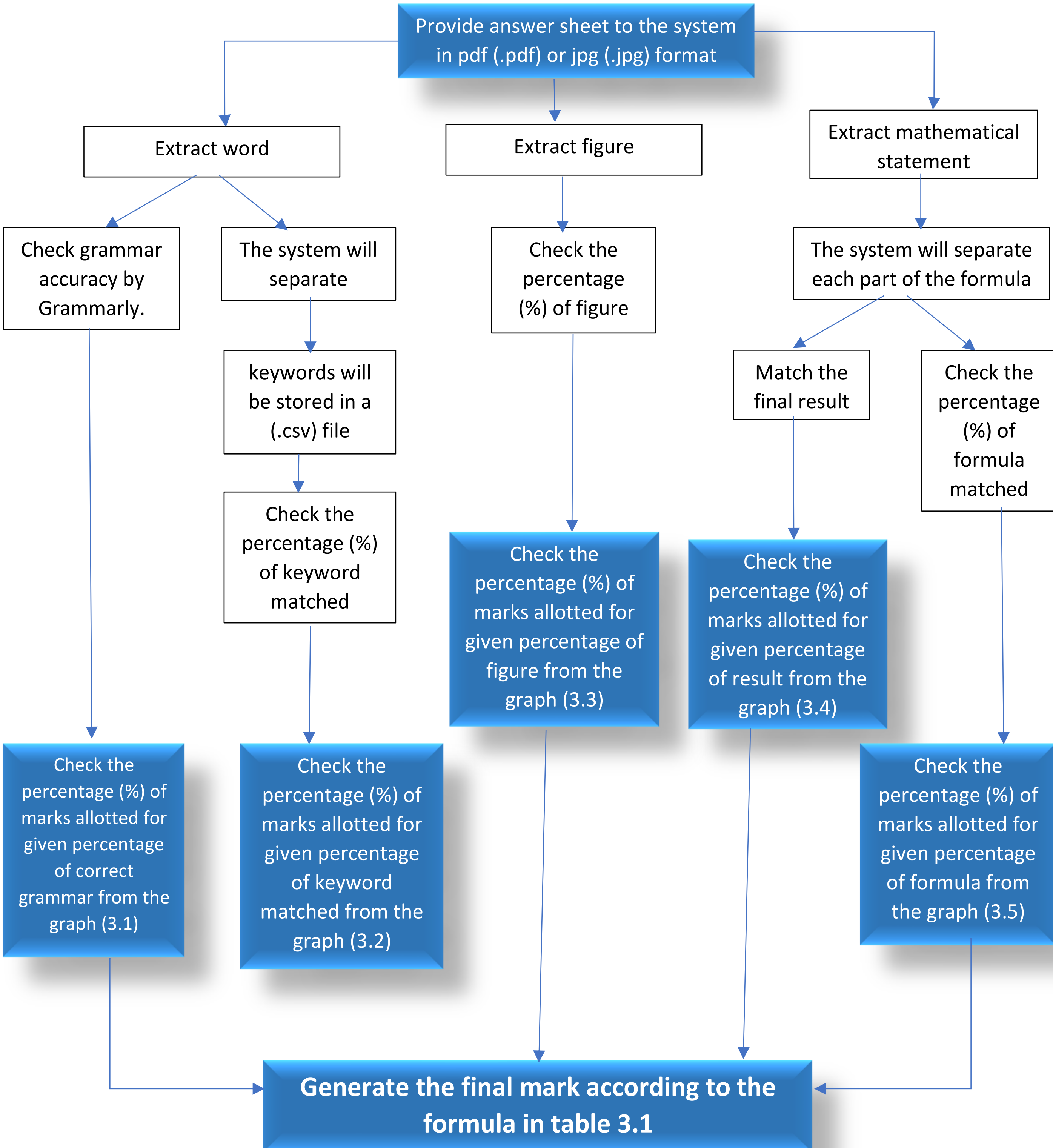


Figure 3.1: Flowchart of the proposed method

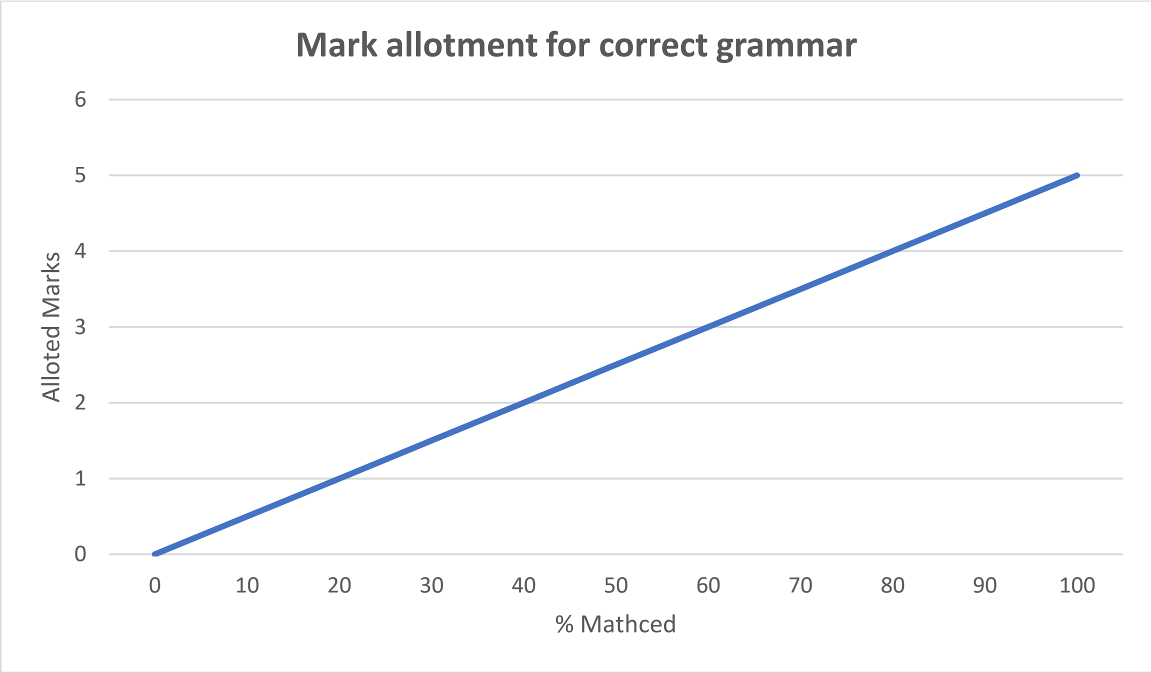
3.6 Mark allotment for each category

Here we are discussing a technique by which we can produce an appropriate mark. For this, we are using some graphs. The graph provides a mark between (0 to 5) for each category.

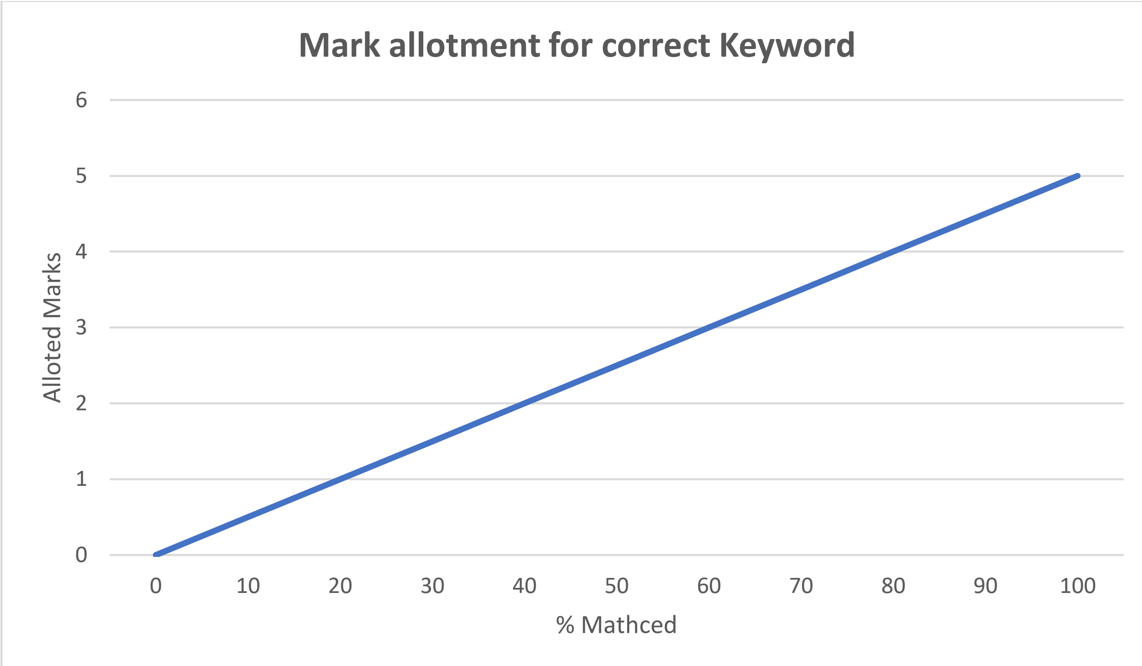
There are 5 graphs:

1. Mark allotment for correct grammar
2. Mark allotment for correct Keyword
3. Mark allotment for correct figure
4. Mark allotment for correct math result
5. Mark allotment for correct math Formula

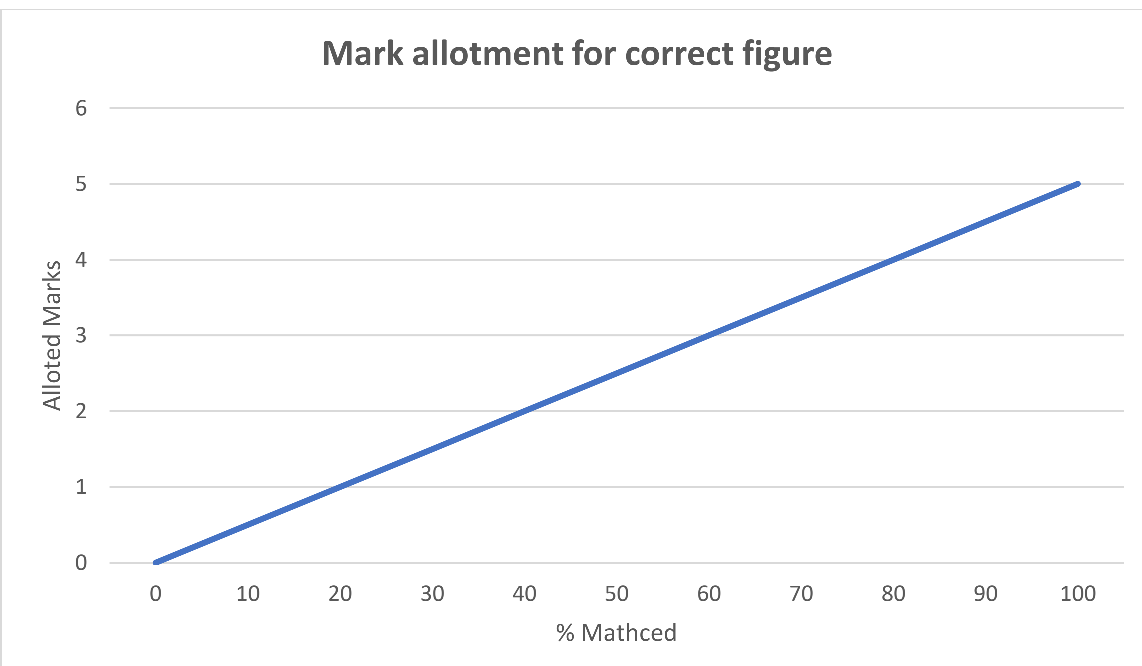
Graph 3.1



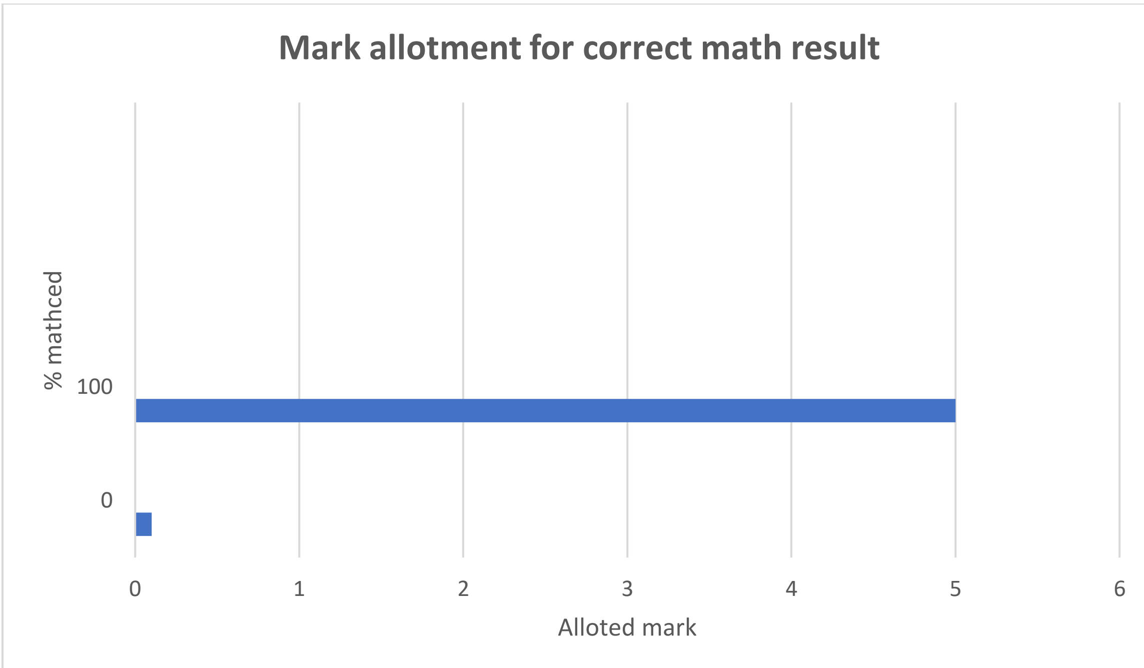
Graph 3.2



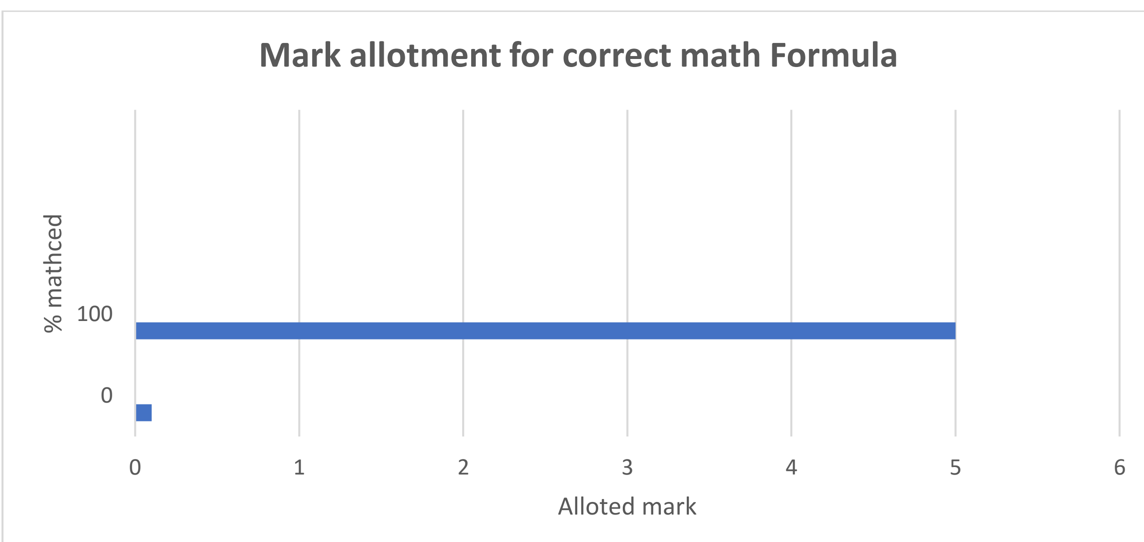
Graph 3.3



Graph 3.4



Graph 3.5



3.7 Generate final mark

For generating the final mark we need to convert the mark from the graph into the form of question mark distribution. For this, we will use some formulas shown in table 3.1

Table 3.1:

Final Mark Evaluation Formula

Category	Formula
Only text	(0.2 * CG + 0.8 * CKW) * TM / 5
Only figure	(1 * CF) * Total mark / 5
Only mathematical statement	(0.5 * CMR + 0.5 * CMF) * TM / 5
Text + figure	(0.1 * CG + 0.4 * CKW + 0.5 * CF) * TM / 5
Text + mathematical statement	(0.1 * CG + 0.4 * CKW + 0.25 * CMR + 0.25 *CMF) * TM/5
Figure + mathematical statement	(0.4* CF + 0.3 * CMR + 0.3 * CMF) * TM/5
Text + Figure + mathematical statement	(0.05 * CG + 0.25 * CKW + 0.25 * CF + 0.20 * CMR + 0.25 * CMF) * TM/ 5
*Mark for correct grammar = CG *Mark for correct keyword match = CKW *Mark for correct figure = CF *Mark for correct math result = CMR *Mark for correct math formula = CMF Total mark = TM	

Chapter 4

Expected Outcome

4.1 Expected Outcome

The current manual evaluation takes about 60 seconds to evaluate an answer whereas the proposed system takes about 15 seconds to evaluate an answer. The proposed system is 300% more time efficient as compared to the manual answer evaluation system. The proposed system completely eliminates the human effort and time to evaluate an answer.

The proposed system can evaluate 5760 answers in a day whereas a human working for 8 hours can evaluate 480 answers a day. Hence, the proposed system can evaluate 1100% more answers compared to that of the manual evaluation system.

Reference:

1. X. Hu and H. Xia, ‘‘Automated assessment system for subjective questions based on LSI,’’ in Proc. 3rd Int. Symp. Intell. Inf. Technol. Secure. Information., Apr. 2010, pp. 250–254.
2. M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, ‘‘From word embeddings to document distances,’’ in Proc. Int. Conf. Mach. Learn., 2015, pp. 957–966.
3. J. E. Kim, K. Park, J. M. Chae, H. J. Jang, B. W. Kim, and S. Y. Jung, ‘‘Automatic scoring system for short descriptive answer written in Korean using Lexico-semantic pattern,’’ Soft Comput., vol. 22, no. 13, pp. 4241–4249, 2018.
4. M. Oghbaie and M. M. Zanjireh, ‘‘Pairwise document similarity measure based on present term set,’’ J. Big Data, vol. 5, no. 1, pp. 1–23, Dec. 2018.
5. Orkphol and W. Yang, ‘‘Word sense disambiguation using cosine similarity collaborate with Word2vec and WordNet,’’ Future Internet, vol. 11, no. 5, p. 114, May 2019.
6. C. Xia, T. He, W. Li, Z. Qin, and Z. Zou, ‘‘Similarity analysis of law documents based on Word2vec,’’ in Proc. IEEE 19th Int. Conf. Softw. Qual., Rel. Secure. Companion (QRS-C), Jul. 2019, pp. 354–357.
7. R. S. Wagh and D. Anand, ‘‘Legal document similarity: A multicriteria decision-making perspective,’’ PeerJ Comput. Sci., vol. 6, p. e262, Mar. 2020.
8. M. Alian and A. Awajan, ‘‘Factors affecting sentence similarity and paraphrasing identification,’’ Int. J. Speech Technol., vol. 23, no. 4, pp. 851–859, Dec. 2020.
9. J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, ‘‘Document plagiarism detection using a new concept similarity in formal concept analysis,’’ J. Appl. Math., vol. 2021, pp. 1–10, Mar. 2021.
10. G. Jain and D. K. Lobiyal, ‘‘Conceptual graphs based approach for subjective answers evaluation,’’ Int. J. Conceptual Struct. Smart Appl., vol. 5, no. 2, pp. 1–21, Jul. 2017.
11. V. Bahel and A. Thomas, ‘‘Text similarity analysis for evaluation of descriptive answers,’’ 2021, arXiv:2105.02935.
12. Sheeba Praveen, “An Approach to Evaluate Subjective Questions for Online Examination System”, Assistant Professor, Dept. CSE, Integral University, Lucknow, U.P, India.
13. B Vanni, M. shyni, and R. Deepalakshmi, ”High accuracy optical character recognition algorithms using learning an array of ANN” in Proc. 2014 IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference.
14. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition ”, arXiv:1507.05717v1 [cs.CV] 21 Jul 2015.
15. Yusuf Perwej, Ashish Chaturvedi, “ Neural Networks for Handwritten English Alphabet Recognition ”, International Journal of Computer Applications (0975 – 8887) Volume 20– No.7, April 2011.
16. J.Pradeep , E.Srinivasan and S.Himavathi, “Neural network based handwritten character recognition system without feature extraction ”, International Conference on Computer, Communication and Electrical Technology – ICCCET 2011, 18th & 19th March 2011.
17. K Arlitsch, J Herbert “Microfilm, paper, and OCR: Issues in newspaper digitization. the Utah digital newspapers program” Microform & Imaging Review, 2004 - degruyter.com

18. Shai Shalev-Shwartz and Shai Ben-David "Understanding Machine Learning: From Theory to Algorithms", ©2014 [11] George Nagy, Stephen V. Rice, and Thomas A. Nartker, "Optical Character Recognition: An Illustrated Guide to the Frontier (The Springer International Series in Engineering and Computer Science)"
19. Batuhan Balci, Dan Saadati, Dan Shiferaw, "Handwritten Text Recognition using Deep Learning ", CS231n: Convolutional Neural Networks for Visual Recognition Spring 2017 project report.
20. Judy McKimm, Carol Jollie, Peter Cantillon, "ABC of learning and teaching Web based learning" http://www.bmj.com/
21. Yuan, Zhenming, et al. A Web-Based Examination and Evaluation System for Computer Education. Washington, DC: IEEE Computer Society, 2006.
22. Effie Lai-Chong Law, et al. Mixed-Method Validation of Pedagogical Concepts for an Intercultural Online Learning Environment. New York: Association for Computer Machinery, 2007.
23. Lan, Glover, et al. Online Annotation- Research and Practices. Oxford UK: Elsevier Science Ltd, 2007.
24. Sophal Chao and Dr. Y.B Reddy Online examination Fifth International Conference on Information Technology: New Generations, 2008
25. Hanumant R. Gite, C.Namrata Mahender "Representation of Model Answer: Online Subjective Examination System" National conference NC3IT2012 Sinhgad Institute of Computer Sciences Pandharpur.
26. "Discourse analysis" website http://www.tlumaczenia _angielski.info/linguistic/discourse.htm
27. H. Mittal and M. S. Devi, ‘‘Subjective evaluation: A comparison of several statistical techniques,’’ Appl. Artif. Intell., vol. 32, no. 1, pp. 85–95, Jan. 2018.
28. L. A. Cutrone and M. Chang, ‘‘Automarking: Automatic assessment of open questions,’’ in Proc. 10th IEEE Int. Conf. Adv. Learn. Technol., Sousse, Tunisia, Jul. 2010, pp. 143–147.
29. G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, ‘‘A two-stage text feature selection algorithm for improving text classification,’’ Tech. Rep., 2021.
30. H. Mangassarian and H. Artail, ‘‘A general framework for subjective information extraction from unstructured English text,’’ Data Knowl. Eng., vol. 62, no. 2, pp. 352–367, Aug. 2007.
31. B. Oral, E. Emekligil, S. Arslan, and G. Eryigit, ‘‘Information extraction ˇ from text intensive and visually rich banking documents,’’ Inf. Process. Manage., vol. 57, no. 6, Nov. 2020, Art. no. 102361.
32. H. Khan, M. U. Asghar, M. Z. Asghar, G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, ‘‘Fake review classification using supervised machine learning,’’ in Proc. Pattern Recognit. Int. Workshops Challenges (ICPR). New York, NY, USA: Springer, 2021, pp. 269–288.
33. S. Afzal, M. Asim, A. R. Javed, M. O. Beg, and T. Baker, ‘‘URLdeepDetect: A deep learning approach for detecting malicious URLs using semantic vector models,’’ J. Netw. Syst. Manage., vol. 29, no. 3, pp. 1–27, Mar. 2021.

34. N. Madnani and A. Cahill, ‘‘Automated scoring: Beyond natural language processing,’’ in Proc. 27th Int. Conf. Comput. Linguistics (COLING), E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Santa Fe, NM, USA: Association for Computational Linguistics, Aug. 2018, pp. 1099–1109.
35. Z. Lin, H. Wang, and S. I. McClean, ‘‘Measuring tree similarity for natural language processing based information retrieval,’’ in Proc. Int. Conf. Appl. Natural Lang. Inf. Syst. (NLDB) (Lecture Notes in Computer Science), vol. 6177, C. J. Hopfe, Y. Rezgui, E. Métais, A. D. Preece, and H. Li, Eds. Cardiff, U.K.: Springer, 2010, pp. 13–23.
36. G. Grefenstette, ‘‘Tokenization,’’ in Syntactic Wordclass Tagging. Springer, 1999, pp. 117–133.
37. K. Sirts and K. Peekman, ‘‘Evaluating sentence segmentation and word Tokenization systems on Estonian web texts,’’ in Proc. 9th Int. Conf. Baltic (HLT) (Frontiers in Artificial Intelligence and Applications) vol. 328, U. Andrius, V. Jurgita, K. Jolantai, and K. Danguole, Eds. Kaunas, Lithuania: IOS Press, Sep. 2020, pp. 174–181.
38. A. Schofield, M. Magnusson, and D. M. Mimno, ‘‘Pulling out the stops: Rethinking stopword removal for topic models,’’ in Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL) vol. 2, M. Lapata, P. Blunsom, and A. Koller, Eds. Valencia, Spain: Association for Computational Linguistics, 2017, pp. 432–436.
39. M. Çagatayli and E. Çelebi, ‘‘The effect of stemming and stop-wordremoval on automatic text classification in Turkish language,’’ in Proc. 22nd Int. Conf. Neural Inf. Process. (ICONIP) (Lecture Notes in Computer Science), vol. 9489, S. Arik, T. Huang, W. K. Lai, and Q. Liu, Eds. Istanbul, Turkey: Springer, 2015, pp. 168–176.