Advances and Applications in Mathematical Sciences
Volume 20, Issue 11, September 2021, Pages 2749-2765

A SURVEY ON AUTOMATIC SUBJECTIVE ANSWER EVALUATION

MADHAVI B. DESAI, VISARG D. DESAI, RAHUL S. GUPTA, DEEP D. MEVADA and YASH S. MISTRY

Computer Science and Engineering Department
R.N.G. Patel Institute of Technology
Gujarat, India
E-mail: desaimadhavi30@gmail.com
visargdesai@gmail.com
guptaji048@gmail.com
deepmevada0@gmail.com
yashmistry143@gmail.com

Abstract

This paper presents different approaches that can be used for automatic subjective answer evaluation. The subjective answer evaluation task consists of two parts: extraction of answers from handwritten answer sheets and finding similarity in answer. Handwritten recognition consists of handwritten character recognition and handwritten word recognition. Answer similarity is done by finding word similarity and then sentence similarity. The goal of this paper is to present the evolution of Natural Language Processing (NLP) and Optical Character Recognition (OCR) techniques for automatic subjective answer evaluation. This paper reviews various Natural Language Processing techniques on popular datasets such as SICK dataset, STS benchmark, Microsoft Paraphrase Identification. Optical character Recognition techniques can be evaluated on MNIST dataset, EMNIST dataset, IAM dataset etc.

I. Introduction

Each educational and non-educational organization conducts examinations. The question papers that evaluate a student's performance consist of descriptive questions along with objective questions. Schools usually use descriptive or subjective questions whereas Competitive

2010 Mathematics Subject Classification: 68T50.
Keywords: natural language processing optical character recognition sentence similarity dataset, word similarity answer evaluation handwritten text recognition.
Received October 13, 2020; Accepted November 6, 2020

examinations consist of objective or multiple choice questions. The objective answers can be easily evaluated by machines which are very useful in saving time and resources. However, the schools and colleges face challenges while evaluating descriptive questions as there is no automatic system or machine to evaluate student's answers. This leads to manual evaluation which involves lot of time and effort by the evaluators. There is also the possibility of bias as the quality of evaluation depends on the emotion of the evaluator, which will reflect on the student's performance score.

In machine learning based answer evaluation system each result depends on the input data provided by the user and pre-trained models. The machine learning models are trained with datasets which contains scores given to right answers. The trained model knows what score to be given based on the input given by the user. Natural Language Processing is used to decrease the task of extraction of useful data from the user input. Some of the steps of natural language processing are Tokenizing of Words and Sentences, Part-of-Speech Tagging and Lemmatizing Words. Students' answers can either be collected digitally, i.e., answers written on the website for the questions provided, or from handwritten answer sheets. This is where the part of Optical Character Recognition or handwritten character recognition (HCR) comes in. The main challenge of handwritten character recognition is variations in handwritings of different students. This makes handwritten answer recognition and automatic subjective evaluation system an essential for education institutes to reduce time and efforts of evaluation process and ensure a transparent evaluation without any bias in evaluating the answers.

NLP is the first building block of answer evaluation system. The keywords, sentences, grammar, and Question-specific words/things are extracted from the handwritten answers for matching. A Similarity Matching Algorithm will calculate the score by comparing the students' answers and the evaluators' correct answers. Once the answers are evaluated they are analysed and detailed report of score card is generated which help the students to score better next time.

The accuracy of answer evaluation system depends on the accuracy with which handwritten answers are recognized. The main painstaking task in Optical Character Recognition is to get data from the answer sheets with at least 80-90 % accuracy. As every student writes different answers for the same question with different handwritings it is very difficult to train OCR model. Different datasets used to train such an OCR model is discussed further in this paper.

This paper provides detailed information on the limitations and techniques used for automatic answer evaluation systems. The enhancements in the capabilities of NLP provide a solid basis for the extraction and evaluation of students' answers. Despite many improvements in OCR systems, they still lack the ability to easily process any other language rather than English. Most OCR systems only extract data from printed out text, whereas in general the aim is to get data from handwritten answer sheets.

The paper is organized as follows. Section I starts with the introduction of the problem. Section II provides different techniques and current approaches to NLP and OCR on different datasets. This section helps to figure out the limitations and problems with current approaches. Section III presents a brief conclusion of this survey paper.

II. Literature Survey

A. Natural Language Processing. In today's covid-19 situation, each institute is trying to impart knowledge to learner digitally. Automation in subjective answers evaluation is key research area as immediate manual assessment is not practical when number of students are large. It is a big challenge for recognizing answers in natural language and extraction of precise meaning to appropriately evaluate the knowledge of student's answers. In 2015, Burrows et al. [1] provided and summarized analysis of steps required for automatic grading of short essays. It contains two steps. (i) Preparing datasets (ii) building grading models and evaluation of models using Natural language processing (NLP) techniques. Hence, we can say NLP perform key role for automatic grading of essays. In 2017 automated grading system of short answers was proposed by Pribadi et al. [2]. In this research, author compared and analysed the different methods to measure the degree of similarity between student's answer and expert answers. Author analysed through experiment that cosine coefficient performed better than Jaccard and Dice coefficient methods. V. Nandini et al. [3] proposed to use semantic relational features for automatic assessment of descriptive answers in an online examination system in the year 2018. Author implemented proposed solution in four stages: (i) Question classification (ii) Answer Classification (iii) Answer Evaluation (iv) Grade/Score allocation. Syntactical relation based feature extraction technique was proposed in this paper for automatic Evaluation of descriptive type answers. Student's answers were scored on the bases of correctness of phrases used for answering the question. Score and feedback both system were implemented by author for awareness of subject knowledge of students.

Machine learning approach based automated Subjective answer grading system was proposed by Sakhapara et al. [4] in 2019. Author analysed performance of latent semantic analysis (LSA) and information gain (IG) for automatic grading system by experiments. For better performance author proposed to use Word Net to add synonyms and improved the results. For the experiment author used biology dataset from Kaggle. In 2020 Tamim et al. [5] proposed a keyword based technique for broad question answer sheet evaluation. Researcher proposed a solution that automatically examines and evaluates the handwritten answer sheets of students by finding keywords and compares it with parsed keywords from open and closed domains. Researcher has implemented approach to find grammatical mistakes and spelling errors from answer sheets and tested it on 100 students answer sheets and achieved precision score of 0.91. In 2020, Leila et. al. [6] introduced an AR-ASAG Arabic dataset for automatic grading of short answers and also explored corpus based approach for the automatic grading of Arabic language. To create a sematic space of word distribution author used correlated occurrence analogue to lexical semantic algorithm. To check the similarity between student and teacher answer summation vector model has been used. Various experiments domain specificity, stemming technique, semantic space dimension is performed to check the validity of dataset and algorithm performance. To measure the comprehensive ability of student, Sadhu Prasad Kar et al. [7] proposed intelligent assessment using N-gram technique. N-gram tuple has been used to eliminate the dependency on pre-defined keywords. In 2020 Sonakshi Vij et al. [8] proposed automatic evaluation of short answers using machine learning approach. Author has used Word Net graphs for finding similarity between student's answer and expert answer. Author incorporated semantic relation of answers text for better performance. Author proved through experiment on 400 answer sheets that approach gives promising result compare to state-of-art methods.

The language which humans use to communicate with each other is known as natural language. Humans can easily understand and communicate using this language. However, for machines, it is very difficult to interpret the language. In order to enable the machine to understand this language, preprocessing of data is required. This preprocessing consists of a series of steps including processes like Sentence Segmentation, Word Tokenization, POS tagging, Stemming and Lemmatization and finding n-grams.

The data for natural language is available in a very large quantity on different mediums such as World Wide Web, newspaper, magazines, books, etc. This data is very useful in the field of deep learning to perform human-like sophisticated tasks. Once we perform the above stated preprocessing steps, we will get a rich set of features that can be used as input for machine learning models or a neural network. Recommended font sizes are shown in Table 1.

The preprocessing steps used to extract the useful information from different sentences of an English language are as follows.

1. Sentence Segmentation. Sentence Segmentation is the process of extracting each sentence present in a large text or corpus. For example, consider a text,

Input: “My name is Visarg. Rahul, Deep and Yash are in my project group. I live in Surat.”

Output: “My name is Visarg. Rahul, Deep and Yash are in my project group.”, “I live in Surat.”

2. Word Tokenization. Word Tokenization is the process of extracting all the words present in a large text or corpus. This output can also be considered as a set of tokens. By default, this process extracts every individual word including punctuation marks. We can then filter this output according to our needs. For example,

Input: “I live in Surat”

Output: “I”, “live”, “in”, “Surat”

3. Part of Speech Tagging. Part of speech helps us by identifying the role of words in the sentence. It classifies the word into nouns, adjectives, verbs, etc.

For example,

Input: “I live in Surat”

Output: (‘I’, ‘Possessive Pronoun’) , (‘live’, ‘Verb’) , (‘in’ , ‘Adverb’) , (‘Surat’, ‘Noun’)

4. Stemming and Lemmatization. The number of words in a document is considered as features and if the numbers of words are significantly high then it takes a load of computational resources to train a model. Stemming and Lemmatization are used to convert the different words to the base form.

- Stemming: It does not consider the actual meaning of words but focuses on removing suffix and prefix to convert the given words into a base form. For example,
  Input: [‘lists’, ’list’, ’listing’, ’listed’]
  Output: [‘list’, ’list’, ’list’, ’list’]

- Lemmatization: It compares the actual meaning of words and then converts the given words to a base form. You can see the difference in the output.
  For example,
  Input: [‘lists’, ’list’, ’listing’, ’listed’]
  Output: [‘list’, ’list’, ’listing’, ’listed’]

5. Stop Words. They are the most common words in a language and are generally removed before or after pre-processing of data. Stop words like ‘a’, ‘an’, ‘in’, ‘the’ and more such are needed to understand the dependency between various tokens, but they increase unnecessary data while doing statistical analysis. The list of stop words varies and depends on what kind of output you are expecting.

6. Word Similarity. There are three main approaches that can be used to measure similarity between words. The first approach is based on the notion that similar words will occur in similar patterns and this approach uses statistical analysis. The second approach is Knowledge based which depends on human crafted semantic networks. The third approach is Terminological or String based which considers the word as a sequence of characters [9].

- Corpus based Word Similarity
- Corpus Based Words Similarity approach tries to extract information from large corpus. Degree of similarity between word pairs is calculated from this extracted information. The Latent Semantic Analysis (LSA) and the Hyperspace Analogue to Language e (HAL) are two major corpus based measures for word similarity.
- Latent Semantic Analysis (LSA)
- Latent Semantic Analysis uses the Bag of Words model. In this approach, a term-document matrix is created which shows the occurrence of each word in different documents. The rows in this matrix represent different words and the columns represent different documents. Latent Semantic Analysis is widely used for topic modelling. [10]
- Hyperspace Analogue to Language (HAL)
- The HAL method produces a high-dimensional semantic space. In a semantic space, all the words are considered as point and the position of these points represent meaning of the word. [10]
- Knowledge based Word Similarity
- WordNet is used in Knowledge Based Words Similarity. WordNet can be considered as a collection of words along with their semantic relation with each other. By using this WordNet, we can create a list of synsets. Synsets are collections of semantically similar words which can be used to replace each other. [9]
- String based Word Similarity
- String based similarity is also known as lexical based similarity and it considers the notion that word is a sequence of characters. It makes comparison between two different sequences of characters. There are different methods to measure similarity between words based on string matching such as levenshtein distance, q-gram, and jaccard

distance.

- Levenshtein distance
- It is the distance between two words/strings where the minimum number of single character edits (insertion, deletion or substitution) is required to change one word/string to another. For example, one 4 operation is needed to convert “abcde” to “abcd” (delete). It gives us a number that tells how different two strings/words are from each other.
- Q-gram distance
- Q-gram distance estimates the string similarity based on occurrences of common substrings of length q in both strings. For example, the distance between “abcde” and “acdeb” when $q = 2$ is calculated as the sum of absolute differences between n-gram vectors of both strings. [9]

The similarity of the sentence pairs can be calculated using many different approaches. The first approach is the word based approach which takes into count the frequency of words occurring into a sentence. The second approach is Structure based which takes the structure of sentence in consideration and finds the POS (Parts of Speech) tags for each word present in the sentences. And the third approach is distance based similarity in which sentences are considered as vectors.

- Word Mover’s Distance
- Word Mover’s Distance uses the word embedding of the words in two texts to measure the minimum distance that the words in one text need to “travel” in semantic space to reach the words in the other text. This method is different from other methods as it does not take the frequency of words into consideration. But it takes the semantic meaning of the word into consideration. So, even if the words present in the sentence pairs are not exactly the same it will take the meaning of each word to find the similarity in the sentence pairs.

B. Optical Character Recognition

Optical Character Recognition (OCR) is a process of extracting handwritten characters or printed characters from documents by means of any specialized software and scanner. It allows the computer to read characters from image and convert them into useful data. It involves three basic steps: scanning of the document, recognizing text from the document and saving the scanned document in required format. [11]

The character recognition reads text from student's answer scripts from natural images. Ahmad Taher Azar et al. [12] analysed the performance of various machine learning algorithms for handwritten digit recognition on United States Postal Service (USPS) in 2020 i.e. k-nearest neighbour, single classification decision tree and bagged decision tree. Author proved that bagged decision tree outperformed K-nearest neighbour and single decision tree in terms of correct classification. Shrinivas R. Zanwar et al. [13] proposed to use swarm intelligence and neural network for handwritten English character recognition in 2020. Author has used independent component analysis, hybrid PSO and firefly optimization for effective feature extraction and feature selection. Backpropogation neural network was used for automatic classification. Author proved through experiment results that approach is able to get high precision rate and accuracy.

HCR has various application like data-entry for business documents, number plate recognition, information extraction, assistive technology for visually impaired people etc. [11] Figure 1 shows the basic types of character recognition systems.

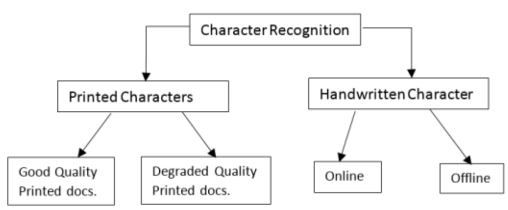

Figure 1. Types of character recognition.

Character Recognition has been classified in two categories printed and handwritten characters Recognition. Further printed characters can be divided in two parts: Good quality and Deteriorated quality of printed documents. Handwritten characters can be classified in two parts: online documents and offline documents. Online documents are captured with digital pen on electronic surface and offline document consist scanned images of text written on paper.

Stages of OCR

The process of OCR has a bunch of activities divided in different phases. The Phases are as follows:

1. Image Acquisition. Image Acquisition is the first step of OCR that includes acquiring an image in digital form and converting it into suitable form that can be processed easily. It includes quantization of an image which is also known as lossy compression technique. A Particular case of quantization is binarization that includes only two steps to process an image. In a large portion of cases, the binary image gets the job done to portray the characterized image. The compression can itself be misfortune or lossless. [14].

2. Pre-processing. Pre-Processing is the next step after image acquisition, that involves enhancing the quality of image. Technique like thresholding is used which creates binary images using some threshold values. Preprocessing also involves filters such as averaging, min and max and different operations such as disintegration, expansion and shutting can be performed.

A significant piece of pre-processing is to discover the skew in the document. Various strategies for skew assessment incorporate: projection profiles, Hough change, nearest neighborhood techniques. At times, diminishing of the picture is additionally performed before later stages are applied. At long last, the content lines present in the record can likewise be discovered as a component of the pre-handling stage. This should be possible dependent on projections or bunching of the pixels. [14]

3. Segmentation. Segmentation involves segmentation of characters from the pre-processed image before passing to the next stage. It gives a by-product for classification phase by performing some implicit and explicit operation. The other stage helps OCR by providing different factors useful for segmentation [14].

4. Feature Extraction. In this stage, different segments of characters are separated. These features uniquely distinguish characters. The determination of the correct segments and the complete number of features to be utilized as a significant research question. Various sorts of features can be used, for example, the picture itself, mathematical highlights (circles, strokes) and measurable highlights (minutes) can be utilized. At last, methods, such as principle components analysis can be utilized to diminish the dimensionality of the image. [14]

5. Classification. It is characterized as the way toward ordering a character into its proper class/category. The basic way to deal with grouping depends on connections present in image parts. The statistical methodologies depend on utilization of a discriminant capacity to group the picture. A portion of the factual characterization approaches are Bayesian classifier, decision tree classifier, neural network classifier, closest neighbour classifiers and so forth. At last, there are classifiers dependent on syntactic methodologies that expect a grammatical approach to deal with creation of a picture from its sub-constituents. [14]

6. Post-Processing. When the character has been arranged, there are different methodologies that can be utilized to improve the precision of OCR results. One such methodology is to utilize more than one classifier for order of picture. The classifier can be utilized in falling, equal or progressive style. The results of the classifiers would then be able to be consolidated utilizing different methodologies.

In order to improve the OCR results, contextual analysis can likewise be performed. Geometric and different document factors of the picture can help in decreasing the chance of an error. Some others methods like cropping, distortion, changing colors of an image can likewise help in improving the consequences of OCR. Figure 2 shows the steps involved in OCR and various algorithms used at various stages.

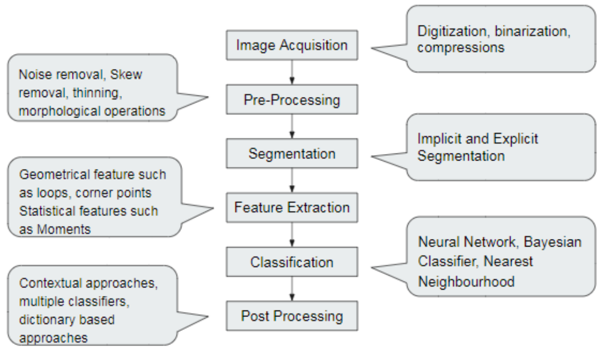

Figure 2. Stages of an OCR and various techniques

C. Datasets. Data is a very important part for Natural Language Processing and Optical Character Recognition tasks. The quality of data decides the accuracy of the model. If the amount of examples in the dataset is less than the model will not be able to generalize well to the unseen inputs. Moreover, noisy data and missing labels in the dataset can create problems. It is possible to clean the dataset and then use it but it takes a lot of time and money. So, it is advisable to use well researched and popular dataset to train our model if it fits the requirement. We reviewed different datasets such as the IAM dataset, MNIST dataset and EMNIST dataset for Optical Character Recognition tasks. And we reviewed the SICK dataset, WordSimilarity353 Test Collection and Microsoft Research Paraphrase Corpus for the Natural Language Processing tasks.

1. IAM dataset. The IAM dataset [15] contains Handwritten English text which can be used to train OCR models. The handwritten text images in this dataset are scanned in 300 dpi and are saved in PNG with 256 gray levels. The IAM 3.0 database is structured as:

- 657 writers contributed their sample handwriting.
- 1539 pages of scanned text.
- 5685 isolated and labelled sentences.
- 16353 isolated and labelled text lines.
- 115320 isolated and labelled words.

2. MNIST dataset. The MNIST dataset [16] contains a large number of examples for handwritten digits. It is a subset of a larger set available from NIST (National Institute of Standard and Technology of the US) set. The whole dataset is collected from two different populations namely the Census Bureau Employees and high school students. The handwritten digit in the image is normalized and centered. The MNIST database is structured as:

● It contains a total of 70,000 instances.

● Training set contains 60,000 instances while the remaining are included in the test set.

● Training set contains samples from more than 250 writers.

● Size of each image is 28 x 28 pixels.

● It contains digits from 0 to 9.

3. EMNIST dataset. The EMNIST dataset [17] contains handwritten English characters including digits. It was also derived from the NIST dataset. It contains 28 x 28 pixel images of over 800,000 manually checked and labeled characters from almost 3,700 writers. The dataset can be used by selecting any of the following six different splits:

● ByClass: It contains 814,255 characters including 62 classes for digits 0 to 9, lowercase alphabets a to *z* and uppercase alphabets *A* to *Z*.

● ByMerge: It contains 814,255 characters including 47 classes for digits 0 to 9, uppercase alphabets A to Z and lowercase alphabets excluding e ‘c’, ‘i’, ‘j’, ‘k’, ‘l’, ‘m’, ‘o’, ‘p’, ‘s’, ‘u’, ‘v’, ‘w’, ‘x’, ‘y’ and ‘z’.

● Balanced: It has the same classes as ByMerge but only contains 131,600 characters in which there are almost equal numbers of examples for each class.

● Letters: It contains 145,600 characters including 26 classes for alphabets.

● Digits: It contains 280,000 characters including 10 classes for digits 0 to 9.

● MNIST: This split represents the same MNIST dataset discussed previously.

4. SICK dataset. The SICK (Sentences Involving Compositional

Knowledge) dataset [18] is a large collection of sentence pairs. It includes around 10,000 pairs of English Sentences. The 8K ImageFlickr data set and the SemEval 2012 STS MSR-Video Description data set were combined to create the SICK dataset. Crowdsourcing techniques were used to collect and label these sentence pairs. Evaluation of Sentence relatedness was done on a 5-point rating scale. The categorization of entailment relation between sentence pairs was done in three classes namely entailment, neutral and contradiction.

5. Word Similarity353 Test Collection. The WordSimilarity-353 Test dataset [19] comprises two different collections of English word pairs along with their manually assigned labels.

● The similarity score of the first set was evaluated by 13 individuals and it contains 153 word pairs.

● The similarity score of the second set was evaluated by 16 individuals and it contains 200 word pairs.

The mean of every individual's score as well as all the individual scores are available in the dataset. All the word pairs in this dataset are rated on the basis of their relatedness on a scale of 0 to 10. Here, 0 shows that there is no relation between words and 10 shows that the words are strongly related or identical to each other.

6. Microsoft Research Paraphrase Corpus. The Microsoft Research Paraphrase Corpus [20] consists of 5801 pairs of sentences along with a binary judgment of whether the pair is a paraphrase or not.. SVM-based classifier was used to select likely sentence-level paraphrases from a large corpus of news data. This news data was taken from the World Wide Web. The 5801 sentence pair selected by the SVM classifier as paraphrases was then examined by the human judges. Out of 5801 pairs, 3900 pairs (67%) were actually judged as paraphrases by human judges.

There are many other well researched datasets that can be used for the tasks related to Natural Language Processing and Optical Character Recognition.

III. Conclusion

The automatic answer evaluation system aims to grade performance of the students based on the answer provided by them. These answers can be processed using Natural Language Processing and its features. Section II of this paper, presents a brief survey of various NLP based techniques used for automatic descriptive answer evaluation. This paper also gives a brief overview of processing steps involved in Natural Language Processing for text recognition and Handwritten Character Recognition. Use of Natural Language Processing coupled with robust classification techniques, checks for not only keywords but also the question specific thing along with grammar of the answer. Every question's answer must contain some question specific things, else the answer is not correct. In addition to this, the results should have a high percentage of quality, i.e., 80-90 %. Eventually the system will scale to extract answers from handwritten answer sheets using OCR or HCR. The biggest challenge is to retrieve the data from handwritten sheets with utmost accuracy. As the technicality of the subject increases, different classifiers can be employed in the system. The accuracy of the evaluation and extraction of answers can be increased by providing it a huge and accurate dataset for training. Section 3 of this paper, presents various popular datasets used for the development. It is suggested that the use of NLP along with OCR, will help to evaluate students' answers and reduce the painstaking burden of evaluators and save abandon of time in generating results.

References

[1] S. Burrows, I. Gurevych, and B. Stein, The Eras and Trends of Automatic Short Answer Grading, International Journal of Artificial Intelligence in Education, Springer, New York 25 (1) (2015), 60-117.

[2] F. S. Pribadi, T. B. Adji, A. E. Permanasari, A. Mulwinda and A. B. Utomo, Automatic Short Answer Scoring Using Words Overlapping Methods, AIP Conference Proceedings vol. 1818, issue 1, Published Online: 10 March 2017.

[3] V. Nandini, and P. Uma Maheswari, Automatic assessment of descriptive answers in online examination system using semantic relational features. J Supercomput 76 (2020), 4430-4448.

[4] A. Sakhapara, D. Pawade, B. Chaudhari, R. Gada, A. Mishra, and S. Bhanushali, Subjective Answer Grader System Based on Machine Learning, Soft Computing and Signal Processing, Springer (2019), 347-355.

[5] Tamim Al Mahmud, Md Gulzar Hussain, Sumaiya Kabir, Hasnain Ahmad, and Mahmudus Sobhan. A Keyword Based Technique to Evaluate Broad Question Answer Script. In Proceedings of the 2020 9th International Conference on Software and

Computer Applications (ICSCA 2020), Association for Computing Machinery, New York, NY, USA (2020), 167-171.

[6] Leila Ouahrani and Djamal Bennouar. AR-ASAG An ARabic Dataset for Automatic Short Answer Grading Evaluation, Proceedings of The 12th Language Resources and Evaluation Conference, European Language Resources Association (2020), 2634-2643.

[7] S. P. Kar, R. Chatterjee and J. K. Mandal, Intelligent Assessment Using Variable N-gram Technique. In: Auer M., Hortsch H., Sethakul P. (eds) The Impact of the 4th Industrial Revolution on Engineering Education. ICL 2019, Advances in Intelligent Systems and Computing, vol 1135 Springer, Cham, (2020).

[8] S. Vij, D. Tayal, and A. Jain, A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on Word Net Graphs. Wireless Pers Commun 111 (2020), 1271-1282.

[9] Farouk and Mamdouh, Measuring sentences similarity: a survey. arXiv preprint arXiv:1910.03940, (2019).

[10] Islam, Aminul, and Diana Inkpen, Semantic text similarity using corpus-based word similarity and string similarity, ACM Transactions on Knowledge Discovery from Data (TKDD) 2.2 (2008), 1-25.

[11] Guides.library.illinois.edu. 2020. Libguides: Introduction To OCR And Searchable Pdfs: An Introduction To OCR. [online]Available at: <https://guides.library.illinois.edu/OCR> [Access at 12-09-2020].

[12] A. T. Azar, A. Khamis, N. A. Kamaland B. Galli Machine Learning Techniques for Handwritten Digit Recognition, In: Hassanien AE., Azar A., Gaber T., Oliva D., Tolba F. (eds) Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020). AICV 2020. Advances in Intelligent Systems and Computing, vol 1153. Springer, Cham, (2020).

[13] S. R. Zanwar, U. B. Shinde, A. S. Narote and S. P. Narote, Handwritten English Character Recognition Using Swarm Intelligence and Neural Network, In: Thampi S. et al. (eds) Intelligent Systems, Technologies and Applications. Advances in Intelligent Systems and Computing, vol 1148. Springer, Singapore, (2020).

[14] Islam, Noman, Zeeshan Islam and Nazia Noor, A survey on optical character recognition system." arXiv preprint arXiv:1710.05703, (2017).

[15] Marti, U-V. and Horst Bunke, The IAM-database: an English sentence database for offline handwriting recognition, International Journal on Document Analysis and Recognition 5.1 (2002), 39-46.

[16] LeCun and Yann The MNIST database of handwritten digits, http://yann. lecun. com/exdb/mnist/ (1998).

[17] Cohen and Gregory, et al., EMNIST: Extending MNIST to handwritten letters 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, (2017).

[18] Marelli and Marco, et al., The SICK (Sentences Involving Compositional Knowledge) dataset for relatedness and entailment, (2014).

[19] Finkelstein and Lev, et al. Placing search in context: The concept revisited, Proceedings of the 10th international conference on World Wide Web. 2001.

[20] Dolan, B. William and Chris Brockett, Automatically constructing a corpus of sentential paraphrases, Proceedings of the Third International Workshop on Paraphrasing (IWP2005), 2005.