ISSN 2395-1621

Online Subjective Test Result Analysis

Prof. A. G. Said, Mr. Devanand V. Kale, Mr. Aniket B. Shinde, Mr. Avinash S. Ugale, Mr. Kaustubh K. Wagh,

devanandkale18@gmail.com
andy471998@gmail.com
ugaleas28@gmail.com
wagh2898@gmail.com

Department of Computer Engineering
AISSMS Institute Of Information Technology

ABSTRACT

Now days, universities are taking the theoretical exams on the papers, after exams the answer sheets are to be checked by the professors. The evaluation might be differ as per the professor and also the answers written by the student, so it will affect the result of student but the evaluation should be accurate and as per the intellectual of student the marks are given. We are implementing the system evaluate the theoretical answers written by the students using NLP (Natural Language Processing). NLP helps to perform semantic analysis from large text collection. With help of NLP we are going to check the subjective answer step by step using the natural language processing steps the step are elaborated further. This can lead to minimize the human errors and efforts while checking the answers manually. This can give us fair evaluation of answers written by the applicant. This can also reduce the time of releasing results.

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

Key words: Artificial Intelligence, Machine Learning, Cloud Storage, Preprocessing, Employment.

I. INTRODUCTION

Nowadays ,when it comes to subjective checking we all know that humans can be fallible so to lest that from happening we are going to introduce our system which can check the subjective answers automatically evaluated as per the content written by the applicant. We are implementing this project using NLP (Natural Language Processing).NLP helps to perform semantic analysis from large text collection. With help of NLP we are going to check the subjective answer step by step using the natural language processing steps the step are elaborated further. There are system which are currently using the system for only essays and paragraph writing, we are developing behavior system which can more accurate than the existing system. This can give us fair evaluation of answers written by the applicant.

II. MOTIVATION

Focusing on making less use of paper which will lead a big hand is saving Green culture, in other words making use of Green computing.

III. NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) refers to AI method of communicating with intelligent systems using a natural language such as English. Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

The field of NLP involves making computers to perform useful tasks with the natural languages humans use. The input and output of an NLP system can be Written Text.

WORKING OF NATURAL LANGUAGE PROCESSING:

Syntax and semantic analysis are two main methods used with natural language processing. Syntax is the presentation of words in a sentence to make well form sense. NLP uses syntax to evaluate meaning from a language based on grammatical rules. Syntax technique used include parsing (syntactic analysis for a sentence), word segmentation (which split a large section of text to units), sentence divide (which places sentence boundaries

in large texts), morphologic segmentation (which divides words into category) and stemming (which split words with inaction in them to root forms). Semantics include the use and meaning behind words. NLP applies algorithms to understand the meaning and formation of sentences. Techniques that NLP uses with semantics involve word sense summarize (which gets meaning of a word based on context), named entity identification (which discover words that can be classify into groups), and natural language generation (which will use a database to determine semantics behind words).

IV. TF-IDF

TF-IDF TF-IDF is a very important algorithm which is used in information retrieval. It can be termed as a numerical statistic which is used to calculate how important a word is in a document.it is frequently used as a weighting factor in searches of information retrieval, text mining and user modeling. the tf-idf value is directly proportional to the number of times a word appears in the given documentation and is set by the number of documents in the collection that carry the word. this is helpful to adjust for the fact that some words occurs more frequently in general.TF-IDF is a very common and mostly used algorithm for information retrieval. search engine uses this algorithm to fetch the information from the databases.it can be used for classification of data and figuring out the databases.

TF: frequency which calculates how frequently a term or a word occurs in a document. since each document is not of same length, it is possible that a term would occurred much more time in a large documents rather than in a short documents.so the term frequency is calculated as TF(t))= (Number of times term t occurred in a document) / (Total number of terms in the document.).

IDF: some words occurs for a lot of times.eg ,the preposition, articles like for, a, an, on etc. this words has less importance thus the algorithm weighs down the frequent terms while scaling up the rare ones.it is calculated as IDF (t) = log E (Total number of documents / Number of documents with term t in It.)

V. SYSTEM DESIGN

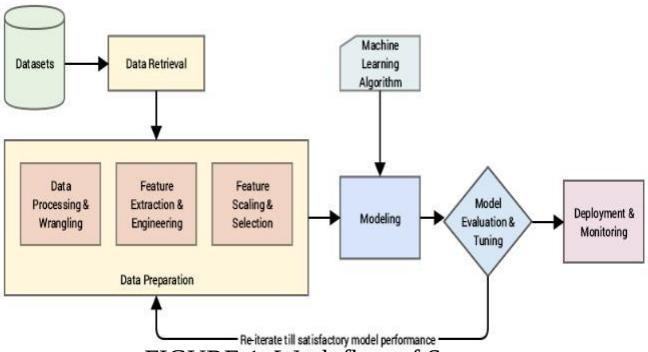

FIGURE 1: Work flow of System

DATASET:
We have created our own dataset, By collecting the question and answers from different books and available online PDFs.

First we take the raw data through csv file. There will be 2 columns question and answer. Ex, if we are working on the topic operating system. We like store some questions and its answer in csv file.

Example: (sample question from csv file Table 1)

FIG. Table 1

what is namespaces in c++?	A:to avoid nameclases concept of namespace should be used.Namespace concept is introduced in c++.while it is as it is used in PHP.in C programming we can use header file.in c++ we can use header files as well as namespaces and in JAVA we can use concept of packages.Generally c++ is used in team development approach where multiple developer works on the same project.due to which,there may be a possibility that names of classes or names of functions can be same.Due to this same name ,compiler will generate error and to avoid it we can use the concept of namespace. The are two types of namespace 1)predefined namespaces by c++ such as std 2)userdefined namespaces written by the programmer.

DATA PROCESSING & WRANGLING:
Initially we will have a deliberation on what data we have. In this we will have discussion on what kind of data we want to retrieve on basis of some factors like data should not include stop words (articles, preposition).

FEATURE EXTRACTION AND ENGINEERING:

Unwanted data is removed and the important data is preserved. In feature extraction and Engineering, stop words are words which are prepositions filtered out after processing of natural language data. Short function words, such as the, is, at, which, and on. By this step, remove those stop words from your answer.
Example: Filtered Data(Table 2)

FIG. Table 2

Text To Filter	Filtered Text
A:to avoid nameclases concept of namespace should be used.Namespace concept is introduced in c++.while it	Avoid nameclass namespace should used namespace concept introduced while used

is as it is used in PHP.in C programming we can use header file.in c++ we can use header files as well as namespaces and in JAVA we can use concept of packages.Generally c++ is used in team development approach where multiple developer works on the same project.due to which,there may be a possibility that names of classes or names of functions can be same.Due to this same name ,compiler will generate error and to avoid it we can use the concept of namespace. The are two types of namespace 1)predefined namespaces by c++ such as std 2)userdefined namespaces written by the programmer.	PHP C programming we use header file in c++ java we can use concept of package genrally team development approach work same project which possibility names classes names function same name compiler generate error avoid concept namespace Two types namespaces predefined namespaces c++ std userdefined namespaces written programmer

FEATURE SCALING AND SELECTION:

We standardize the important data present in a data which are independent in a fixed range. If scaling is not performed then the algorithm will consider greater value as high and lower value as low irrespective of the unit of values.

MODELLING:
Machine learning algorithm:

Machine learning algorithm TF IDF is used in this application for test mining.

MODEL EVALUATION & TUNING:

In this the model is evaluated.TF IDF will work on the important data. Classification of data will take place. The data is arranged in key value pair. Semantic, syntax analysis and all will be carried out

DEPLOYMENT AND MONITORING:

The model will be deployed and the model will be analyzed. The performance of the model will be calculated and analyzed. If the performance of a model is not satisfactory then again the input data will be increased and all the steps will be carried out

VI. RESULT

In the whole process using NLP & TFIDF algorithms the question written by the student is compared with the standard answer given in the csv file. And the result displayed on the screen is in the percentage of total marks given to the question.
Accuracy of the system depends upon the dataset provided. Accuracy is directly proportional the size of dataset. About the accuracy of our system it can reach up to 80% to 85% on available dataset.

VII.FUTURE SCOPE

1. For future work, we are planning to extend the proposed architecture to make it.

2. Compatible with Big Data analysis for all types of exams.

3. We are also looking for analysis of diagrams in the exams in our system.

VIII. CONCLUSION

We have worked on TFIDF algorithm for carrying out test mining, the algorithm works on dataset created by our self. We have developed the framework which can be used by many organizations to organize the theoretical testing which can provide the accurate results of your test.

REFERENCES

1. Diellza Nagavci Mati, Mentor Hamiti, Jaumin Ajdari, Besnik Selimi, Bujar Rau, “A Systematic Mapping Study of Language Features Identification from Large Text Collection” IEEE(2019).

2. Anjana.J. S, Poorna S. S, “Language Identification from Speech Features Using SVM and LDA”. IEEE (2018).

3. Atzeni M and Atzori M. “Translating Natural Language to Code: An Unsupervised Ontology-Based Approach”. IEEE (2018).

4. Caranica A, Cucu H, and Buzo A, “Exploring an unsupervised language independent, spoken document retrieval system”. IEEE (2016).

5.Carbajal M. J, Dawud A, Thiolliere R and Dupoux E, “The languagelter hypothesis A feasibility study of language separation in infancy using unsu-previsedclustering of I-vectors”. IEEE (2016).

6.Celikyilmaz A, Sarikaya R, Jeong M and Deoras A. “An Empirical Investigation of Word Class-Based Features for Natural Language Understanding”. IEEE (2016).

7. Colbert R, “A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning”. IEEE (2016).

