IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF IAS INTERVIEW TRANSCRIPT USING WORD DATA VISUALIZATION

[1]Rajat Savdekar, [2]Darshan Patil, [3]Tushar Chaudhari,[4]Shivam Jadhav,[5]Mr.Atul Kachare
[1,2,3,4]Student,[5,Assistant] Professor
Computer Science and Engineering,
[1,2,3,4,5] Shah and Anchor Kutchhi Engineering College, Mumbai, India,

Abstract : In this project we are analyzing the interview transcript of the Indian civil service exam which is called MPSC and UPSC. Given the importance of a government job in Indian culture, understanding how to become a government employee plays an important role in the given context. We will analyze the third stage of the UPSC exam which is the interview process of the year 2020 which will be useful for all the upcoming candidates and provide an insight into the mind of the interviewer when conducting the interview.

IndexTerms **- UPSC, Interview, Board member, IAS, Candidate, Word cloud, Data visualization, Image generation, Analysis, .**

INTRODUCTION

In this project which is analysis of IAS interview transcript using word data visualization gives a basic idea of how the interview process is conducted in this exam. First we get to know about the different methods of the word data visualization technique. There are many basic visualizations like bar chart, pie chart, Histogram, Area chart. We also have many advanced and detailed methods for the visualization of word data Network diagram, Pictogram chart, Word cloud and many more.

Words clouds, which are visual representations of a document's text (Kaptein, Hiemstra, & Kamps, 2010), emerge as a potential strategy to support both critical thinking and engagement within the context of online discussions. A word cloud takes the most frequently used words in a particular text and randomly displays them by size, based on their frequencies (DaPaolo & Wilkinson, 2014). Word clouds also display variation in color, typography, and composition, offering an aesthetically pleasing look. Because of this, word clouds are used in a myriad of creative and playful ways by a broad array of users (Viégas, Wattenberg, & Feinberg, 2009)
In this project we have made an effort to make an educated analysis by the help of one method of word data visualization technique called word cloud. A word cloud is generated for the interviews of the students with the background in different domains and thus with the help of this specific word cloud we can analyze what type of questions are mostly asked to students from certain domains. There are also other ways to analyze text which are

- Phrase Net
  - Connecting key words in a text using lines to show linkages
- Word Tree
  - Displaying of the words in qualitative dataset, where frequently used words are connected by branches to the other words that appear nearby in the data.

RELATED WORK

"Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data" by Concetta Deapaolo, Kelly Wilkinson, using word clouds, instructors can quickly and easily produce graphical depictions of text representing student knowledge. By investigating the patterns of words or phrases;Teacher can evaluate if the students has grasped the concept or missed on key points.

"Word Cloud Explorer: Text Analytics Based on Word Clouds" by Florian Heimerl; Steffen Lohmann, in this work, they explore the usefulness of word clouds for general text analysis tasks. They have developed a prototypical system called the Word Cloud Explorer that relies entirely on word clouds as a visualization method.
"Word cloud analysis of the BJGP" by Rob Atenstaedt, the word cloud generated from the analysis was measured against the stated editorial policy of the BJGP Journal. It was seen that the two most prominent words highlighted are 'care' and 'patients'.
"Word Clouds as a Learning Analytic Tool for the Cooperative e-Learning Platform NeuroK" by Fernando Calle-Alonso, Carlos J. Pérez, Miguel A. Vega-Rodríguez, in this work, word clouds are built upon a cooperative e-learning platform, Neurok, to provide an overview of the main concepts. The tool under consideration consists of three different word clouds: the target cloud, the real cloud, and the mixture of both, with the target concepts according to their use by the students

PROBLEM STATEMENT
Clearing an IAS aka Indian Administrative Service exam is a crucial part of a student's life to become an IAS officer. To become an IAS officer the candidate needs to crack the Civil Service Examination (CSE) conducted by the Union Public Service Commission (UPSC). After passing the UPSC-CSE exam only an aspiring candidate can become an IAS officer. UPSC CSE exam takes place in three stages namely:-
1. Preliminary exam
2. Mains exam
3. Interview process
We aim with the help of this paper to give an insight to the students about the last stage of the exam which is the interview with the panel of board members from various senior government positions.

OBJECTIVE
The objective of analysis of IAS interview transcript using word data visualization is that it is a statistical method of extracting the frequently occurring word in the text and thus assist us in the finding the importance of word in the current text context. In this project we are going to create a word cloud of the interview transcripts to visualize the data. The most frequent word will appear as a larger font compared to other words with less frequency.

METHODOLOGY
A word cloud is generated for the interviews of the students with the background in different domains and thus with the help of this specific word cloud we can analyze what type of questions are mostly asked to students from certain domains. We have divided the interviews on the basis of the background education of the students. We also explored other ways to divide the transcript but nothing was exhaustive to be considered independently, for example if we take the optional subject chosen by the student for the analysis it was not mentioned in most of the transcripts, if we considered the board members for the decision we would have to make a large amount of decision since the board members retires and new board member replaces them. In conclusion we found the students' education background to be the most effective and most reliable function for the analysis. We have divided the interview transcripts into the following given seven education backgrounds of the students.
1.Mechanical engineering
2.Computer science engineering
3.Civil engineering
4.Biotechnology / Zoology and Forestry /dental
5.Electronics and telecommunication engineering
6.Aerospace engineering
7.Chemical engineering

IMPLEMENTATION

- **Main GUI for creating the word cloud**

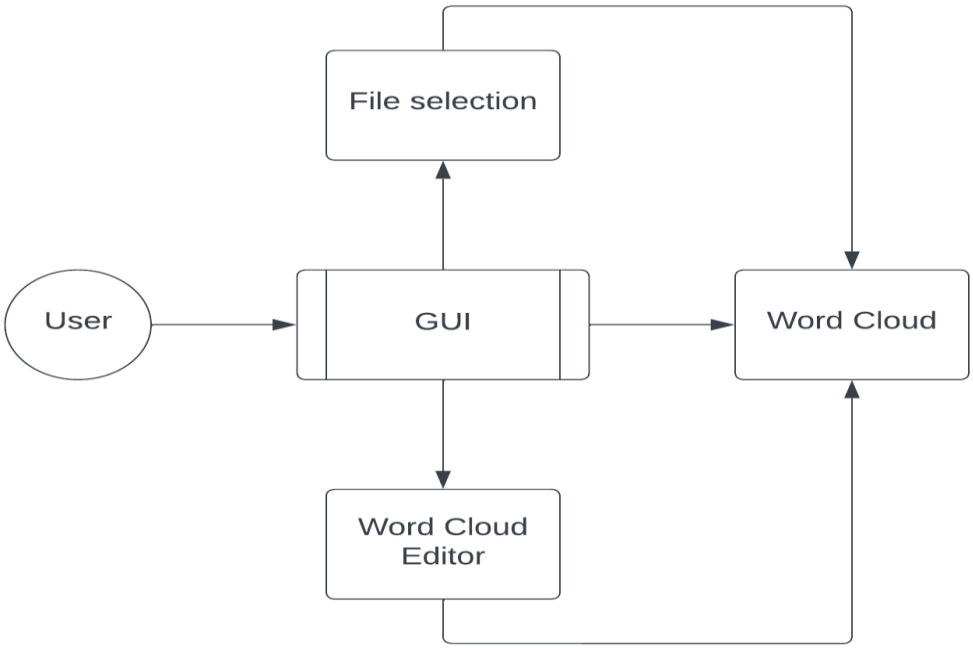
The GUI has a select file button to give the user an option to choose a text file to insert the word data in our case interview transcripts. Also an additional select folder folder button is created for if the user wants to input all the files in a certain folder. We can directly add plaintext by using the add plaintext button . Users can delete any of the selected files which are displayed on a list in the left side of the window. We have to give a project name to each of our word clouds.

- **Generate Word Cloud**

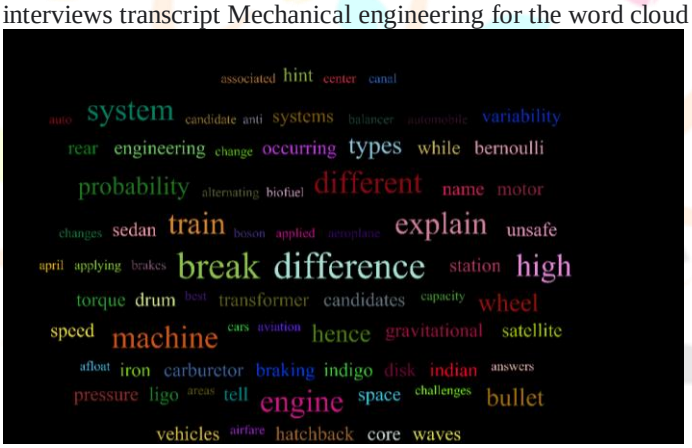
After pressing the Generate Word Cloud button the app will process all the input files given in the last step and then display the output in a new window.

- **Word Cloud Editor**

We have two different options to edit the word cloud which is generated in the last step. They are colouring options and text options. In colouring options we can change the background, select a colour map, select colouring mode, select text colour, to make the cloud more appealing to the needs of the user. In the text option user can change the number of word of which the word cloud is generated, choose different font styles, adjust text size, and also be able to adjust text cluster width.



We take the **first** category of the interviews transcript Mechanical engineering for the word cloud as an input.

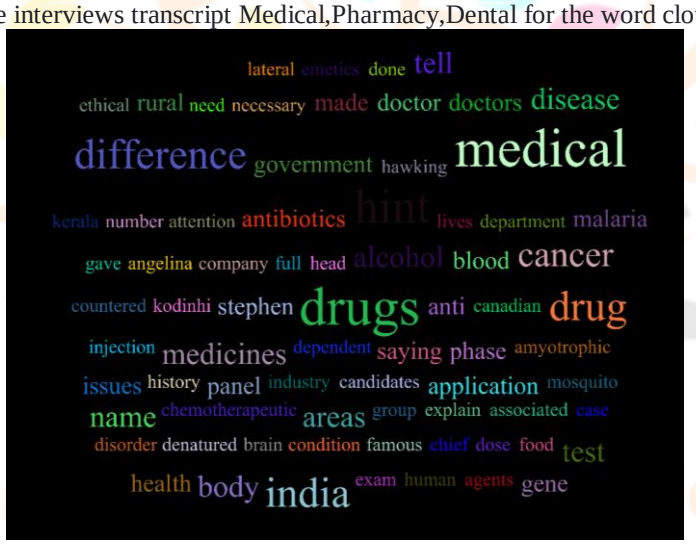


In this word cloud, we can see that ‘break’, ‘high’, ‘machine’, ‘engine’, ‘gravitational’, and ‘probability’ are among the words which occurred several times in the interview transcript. Let us take the word ‘break’, it is one of the most frequently occurring words in the word cloud. Questions such as types of breaks, the difference between different types of breaks, the drum break, and the disk brake were asked in the interview. Several questions on machines, gravitational waves, and events in probability were also asked. So, from this, we can conclude that more questions according to the mechanical background were asked in the interview.

We take the **second** category of the interviews transcript Computer science engineering for the word cloud as an input.

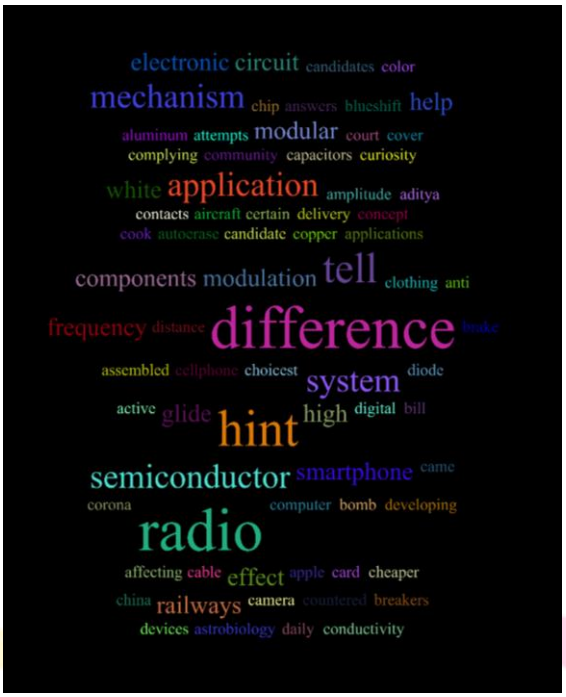
Here in this Word Cloud, it can be seen that the words which are related to computer technology and other computer terms such as ‘Server’, ‘System’, ‘data’, ‘assembly’, ‘WordPress’, ‘Software’, and ‘cloud’ were repeated and that's why highlighted in the word cloud. Here it is clearly showing the interest of the interviewer in the computer field. In this, different questions were asked such as different types of servers i.e.cloud server and dedicated server, different types of systems, hacking of systems, different questions on data, database, data warehouse, assembly language questions, etc.

We take the **third** category of the interviews transcript Medical,Pharmacy,Dental for the word cloud as an input.



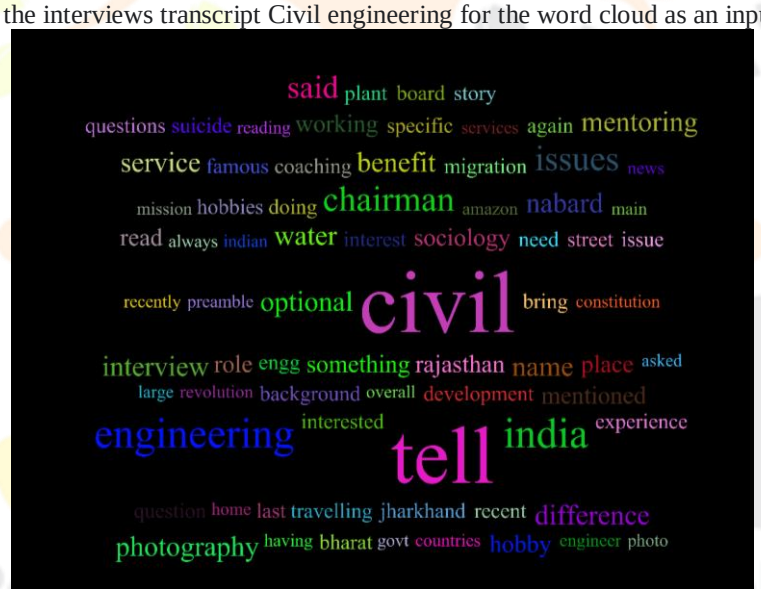
Firstly, it can be seen that the most prominent words highlighted are ‘Medical’, ‘Drugs’ and ‘Cancer’. This Word Cloud is all about the terms asked in an interview with a Student from Medical Field. This finding does conform to the aim of the Journal to focus on these areas. The Word Cloud generated is clearly showing the focus of the interviewer while asking the questions to a candidate from a medical field. It is covering the terms like blood, dose, health, body, doctors, etc. The word cloud focuses on the word Cancer which is giving the idea that questions based on cancer were asked frequently in the interview. Cross Checking with transcript records we find some questions on the future scope for cancer solutions, cancer situation in India, anticancer drugs, cancer deaths by tobacco, etc. Also, questions based on drugs are asked several times, such as clinical trial phases of drugs, the discovery of drugs, famous drugs in India, etc.

We take the **fourth** category of the interviews transcript Electronics and telecommunication engineering for the word cloud as an input.



In this word cloud, we can see that ‘radio’, ‘semiconductor’, ‘application’, ‘modulation’, ‘smartphone’, and ‘digital’ are among the words which occurred several times in the interview transcript. Let us take the word ‘semiconductor’, it is one of the most frequently occurring words in the word cloud. Questions such as types of semiconductors, the difference between register and semiconductor, and applications of semiconductors were asked in the interview. Several questions on radio and its types like community, satellite, ham radio, AM and FM radio, and questions on modulation, modular smartphones were also asked. So, from this, we can conclude that more questions according to the Electronics and Telecommunication background were asked in the interview.

We take the **fifth** category of the interviews transcript Civil engineering for the word cloud as an input.

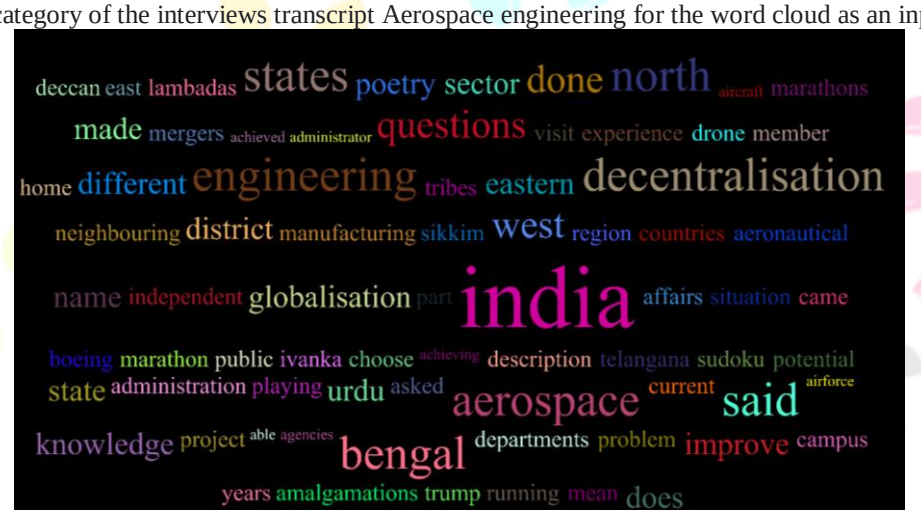


In Word Cloud generated for Civil Engineering interview questions several words like ‘civil’, ‘engineering’, ‘issues’, ‘nabard’, and ‘development’ are the most highlighted words. From this, we can conclude that the interviewer was interested in asking questions based on civil engineering concepts and terms such as development, construction, issues, nabard, etc. Questions based on ‘nabard’ were also asked. Nabard is National Bank For Agriculture And Rural Development, which infers that even if the organizations are different but the goal of the question is similar i.e.Civil Engineering.

We take the **sixth** category of the interviews transcript Chemical engineering for the word cloud as an input.

In this 'chemical', 'India', 'pesticides', 'biofuels' such kinds of words are repeated, which again showing the focus and interest of the interviewer in the chemical field. Questions on biofuels, what are the policies of India for biofuels, India's business with different countries in the chemical field, Future of different chemical companies in India, What is the problem of locust, which pesticide we use, what problem occurs when we deal with pesticides, the difference between petrochemicals and chemicals, chemical s used for different services, etc. These questions were asked by the interviewer.

We take the **seventh** category of the interviews transcript Aerospace engineering for the word cloud as an input.



In the above-given word cloud questions on highlighted words are 'engineering', 'aerospace', 'globalization', 'manufacturing', 'decentralization' and 'department problems' etc. the clear focus of the interviewer is again towards aerospace and all infrastructures related to it. Questions such as what is globalization and what is its impact on the world, what are employment opportunities in India in the aerospace field, what is decentralization, mergers, and amalgamations in the aerospace department, and how to improve the manufacturing sector in this particular field, etc. Such questions were asked in the interview which is again showing the classical pattern of field wise questions

RESULTS AND DISCUSSION

FUTURE SCOPE-WORK

This project can be further extended by performing generating semantic surfaces.

We can add a feature for Semantic Surface Scatter plot showing each separate document as a point in a 3D space. Documents that are close to each other in the plot are similar to each other. This gives you a view of how the documents are related in the project. Clicking on a point will show the document title.

Also if we can make use of different python libraries like Wikipedia to get data directly from wikipedia's api; tweepy to get all the tweets of a user of twitter; And many others like them to be able to extract data from them and to generate a word cloud. This would make the app more useful in more types of analysis while also making the tedious work of collecting and pre-processing the data to clean it for the input of the word cloud very easy.

CONCLUSION

The paper shows that the proposed method is an effective way to perform analysis of IAS interview transcripts. We presented a system in which the user can provide any sort of text data and the output will be a word cloud of the most frequent word used in the text. A word cloud is a tool used to display text data summarization in a visual way very easy to understand. There are many possibilities of use cases for a word cloud.We can use this to make students understand a concept more easily by showing them an image of the word related to the new concept. They can also be useful as a starting point or screening tool for large amounts of text data, whether related to assessment or not.

REFERENCES

[1] Florian Heimerl; Steffen Lohmann, "Word Cloud Explorer: Text Analytics Based on Word Clouds",Volume – 6 , December 2014.

[2] Concetta Deapaolo, Kelly Wilkinson, "Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data", Volume-4 Issue-3, March 2017.

[3] Rob Atenstaedt, "Word cloud analysis of the BJGP" , NIH.gov - 2017.

[4] Fernando Calle-Alonso, Carlos J. Pérez,Miguel A. Vega-Rodríguez, "Word Clouds as a Learning Analytic Tool for the Cooperative e-Learning Platform NeuroK", December 2018.