



Text Visualization Using NLP Tools

Roshan Sanap¹ | Akanksha Bhute¹ | Vibhavari Tayade¹ | Lalita Khairnar¹ | Prof. G. Dashmukhe²

¹ Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India

² Assistant Professor, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India

To Cite this Article

Roshan Sanap, Akanksha Bhute, Vibhavari Tayade, Lalita Khairnar and Prof. G. Dashmukhe. Text Visualization Using NLP Tools, International Journal for Modern Trends in Science and Technology, 2023, 9(11), pages. 38-42. <https://doi.org/10.46501/IJMTST0911008>

Article Info

Received: 19 October 2023; Accepted: 11 November 2023; Published: 15 November 2023.

Copyright © Roshan Sanap et al. This is an open access article distributed under the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

This paper presents a study of implementing computational methods like Natural Language Processing (NLP) to perform Text Analytics and Visualization. With an increasing number of Data produce each year, there is need to synthesize and obtain insights across ever-growing volumes of literature. Text Analysis and Visualization are very important concepts when comes to Data reading and when these two technologies are combined this makes the user possibly understand any data within few constraints. The Corporate world values time more than money. The data is sometimes available in unstructured textual format and thus they are a part of big data requiring analytics to derive insights from it. In this experiment, a significantly large volume of Text is analyzed and graphical visualizations are generated such as Word cloud, Mendenhall Curve, Tokenization, Graph, Processed Text and Name Entity Recognition (NER) using various Python Libraries. So far, the linguists have tried to perform analysis using manual linguistic approaches which are extremely time-consuming and complex to understand the Data. Our experiment of applying NLP based text analytics proves to be a very efficient technique for understanding the Data. Therefore, this paper is an attempt to provide detailed assessment about data, also to make proper Data-driven decisions and Reduce Time Wastage and Efforts. This system provides user to learn and revise his/her Grammar Learnings by describing Nouns, Verbs, Pronouns in the data. A System such as this could be very useful for Teachers, Employees, Students, Analyst and Meetings.

KEYWORDS: Text Visualization, Natural Language processing, Name Entity Recognition.

1. INTRODUCTION

Keeping track of conceptual and methodological developments in any scientific discipline is imperative to advance research. Data is the lifeblood of any business. It's what fuels your decisions and helps you stay on top of your competitors. But it is also one of the most difficult things to track, since it usually happens behind the scenes, in databases or spreadsheets that are difficult

to access or interpret. That is where Natural Language Processing comes in. With NLP, you can use computer algorithms to read text and extract useful information from it that would otherwise be impossible to understand on its own. Natural Language Processing (NLP) is an umbrella term that refers to a set of techniques used to recognize and understand human language. Natural Language Processing (NLP) is a field

that applies computational methods to the analysis of natural human language. It has applications in many different fields, including information retrieval and machine translation. Natural language processing can help you take your data analytics to the next level by enabling you to extract information from unstructured text. For example, if you want to understand the sentiment of a message or an email, then NLP can help you with it. A sentiment analysis tool will automatically identify positive and negative emotions in messages and emails and give you insights into their meaning. You can also use NLP to create automated customer service interactions on your website or mobile app by asking users questions about their products or services.

Natural Language Processing (NLP) is the process of analyzing and understanding human language. It allows you to extract important insights from large amounts of data, which can help make better business decisions. Text Analysis is the process of extracting meaning from text. You can use NLP to understand what your audience wants and needs, so that you can provide them with the information they need to make informed decisions. Text analysis can help to see deeper into the data and use the text in a business context.

Text Visualization is the process of using graphics or images to show readers how their data looks like in a new way. This will help them visualize the data more clearly, so they can better understand it and make better decisions with it. Text Visualization uses data to create images that tell a story to clarify what is going on, bringing out important highlights.

We make it easy for you to make smart business decisions, by providing data insights and context into the trends that matter. Also, we Maximize productivity and reduce time wastage using our NLP based product. Tell our product what you want to do and it will analyze your data. It will find the most important elements in a quick, but thorough way for you to make informed decisions. Powered by an intelligent platform, this tool helps you gain understanding of complex data and make better business decisions.

2. HISTORY AND BACKGROUND

Text visualization technologies, as forms of computer-supported knowledge discovery, aim to improve our ability to understand and utilize the wealth

of text-based information available to us. While the term "text visualization" has been used to describe a variety of techniques for graphically depicting the characteristics of free-text data (Havre et al. 2002; Small 1996), it is most closely associated with the so-called semantic clustering or semantic mapping techniques (Chalmers and Chitson 1992; Kohonen et al. 2000; Lin et al. 1991; Wise et al. 1995).

Natural Language Processing (NLP) caters those users who do not have enough time to learn new languages or get perfection in it. In fact, NLP is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. It came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language, and can be classified into two parts i.e. Natural Language Understanding or Linguistics and Natural Language Generation which evolves the task to understand and generate the text. Linguistics is the science of language which includes Phonology that refers to sound, Morphology word formation, Syntax sentence structure, Semantics syntax and Pragmatics which refers to understanding. Noah Chomsky, one of the first linguists of twelfth century that started syntactic theories, marked a unique position in the field of theoretical linguistics because he revolutionized the area of syntax (Chomsky, 1965).

Further, Natural Language Generation (NLG) is the process of producing phrases, sentences and paragraphs that are meaningful from an internal representation. The first objective of this paper is to give insights of the various important terminologies of NLP and NLG. In the existing literature, most of the work in NLP is conducted by computer scientists while various other professionals have also shown interest such as linguistics, psychologists, and philosophers etc. One of the most interesting aspects of NLP is that it adds up to the knowledge of human language. The field of NLP is related with different theories and techniques that deal with the problem of natural language of communicating with the computers. Few of the researched tasks of NLP are Automatic Summarization (Automatic summarization produces an understandable summary of a set of text and provides summaries or detailed information of text of a known type).

As the Text visualization using Natural Language Processing Techniques in in 1960s efforts in

computational linguists focused on basic tasks like machine translation and syntactic analysis. And then onwards for the research on text analysis and information lead to development of the field of NLP began.

Then in early 2000s text mining techniques such as topic modeling and sentiment analysis began to gain attraction with the information retrieval systems.

In 2010s, the introduction of Word Embeddings such as Word2Vec revolutionized how words were represented computationally. Which then approached to the techniques like deep learning (RNNs), visualization to specific text data.

3. MOTIVATION

To explore the field of Machine Learning and promote NLP study. Data in the organization is at bulk and unstructured it gets difficult to get data abstract quickly. This project will help the Organizations to understand the data quickly. MNC's work on very little time constraint. They must understand what the data is about very quickly. This project will help them to extract important highlights from large data. Visualization allows user to comprehend vast amount of data briefly and in better way.

To Read and understand text for Business insights.

MNC's work on very little time constraint. They must understand what the data is about very quickly. This project will help them to extract important highlights from large data. Visualization allows user to comprehend vast amount of data briefly and in better way.

Most of the time when there is a Time Constraint and a Important meeting is about to start the reader reads the data quickly and is unable to match the points and understand relation between sentences if not formatted quickly. Here, Text analyzer will make the communication quicker and will match the sentences with different colors to understand it quickly.

Anyone can understand the data easily without any prior knowledge. Text analysis and visualization can help the user make sense of qualitative data, quickly, easily, and at a scale. Through charts, Word cloud, which will turn plaintext into eye-catching stories. 8. Often People forget their learnings about Noun, Verbs, Pronouns etc. Text Analyzer is a perfect tool for them as

it not just only summarizes the data but also it marks the Nouns, Verbs, Pronouns from the text which provides learnability and readability at the same time.

4. METHODOLOGY

As in text visualization using NLP techniques are powerful ways to represent textual data. However, are several design issues and considerations that need to be addressed to create effective and meaningful visualizations. Here are some key design issues in text visualization using NLP: Word frequency, Stop words and common terms sorting, Color and Font Selection, Dynamic and Interactive Visualization, Data preprocessing, User-Centered design, and the most important contextual understanding. So, these were some basic design issues considerate which would be conquered by the below methods and diagram represented Figure 1.

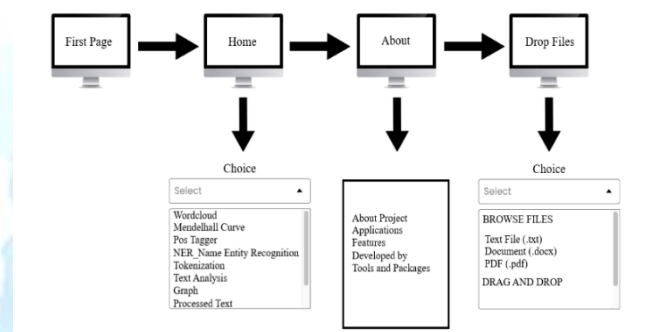


Figure 1: A Proposed System Model

A. Software selection

Unlike elementary statistics that one can compute by hand or with a calculator for small datasets, this level of complexity requires technology, not because of the size of a single dataset, but because modeling and simulation involve the creation of very large numbers of datasets. Many specialized software applications perform statistical analysis. In addition, many general math applications and even some specialized programming languages contain libraries of statistical programs. Teachers must familiarize themselves with the computational resources available to them before they plan lessons to teach modeling and simulation, and then plan them in ways supported by those resources. In addition, instructors must allow time for students to learn how to use these computational tools.

B. Developing models and algorithms

In this context, most models are mathematical representations of some real or imagined activity. These models aim to simulate reality by incorporating random behavior. Video games, for example, model all kinds of behavior, human and otherwise. Given the familiarity of many students with such games, it is probably the best way to introduce them to the concept of modeling. The model is the basic structure of the game. The algorithms are the rules that produce the behavior they observe in the game. When games record their behavior, as I imagine most do, they create datasets. Datasets from games designed to mimic natural or human phenomena may be compared to data on those actual phenomena to determine how realistic or not a game is. This is the objective of most modeling and simulation exercises. Unfortunately, not all inquiries come on the form of a game, but at their most fundamental levels, they all behave like one. The uncertainty introduced by random behavior of variables creates variability in outcomes. The accuracy of the parameters that govern the behavior of the random variables determines how realistic the models are.

C. Programming and automation

Models and algorithms must be programmed to be simulated. Unless programming itself is a learning objective, in classroom settings, it is important to focus on projects for which programming exists. Not long ago, that was not easily done, but it is relatively easy to do today with software that is often available from an open source, i.e., free.

D. Simulation

Simulation is the process of running a program that executes a model repeatedly to create many datasets that describe a phenomenon. It is from statistical analysis of these datasets that analysts can evaluate the accuracy of a model as a surrogate for the real thing it aims to describe.

II. RESULTS AND DISCUSSIONS

Evaluation

All the comments made above about static analysis apply equally to dynamic analysis. As the diagram at the top of this section implies, the process of evaluation is a step in an iterative cycle. It is important for students to

understand that the process of inquiry never ends. Results of simulations are used initially to understand how to improve models and algorithms. Even when no further improvement seems possible, it remains important to continue to test them against the natural phenomena they model to observe changes in the natural environment.

Thus, the result of the outputted programming can be tested by unit testing. And the analysis is processed with the software selection, developing model, programming and simulation and evaluation.

5. CONCLUSION

The focus of this paper was to show the usefulness of Text Scrutiny and Visualization. As we learned how important role the system plays in today's world, we can further embed this system on huge dataset with better algorithms and can be used in Businesses to make speed data driven decisions. In MNC's the system can be used by Employees to present the data using visualization and highlight important context. Schools and Universities can use this system to teach students complex data with few efforts. Students can understand the data quickly and also improve their English knowledge.

ACKNOWLEDGEMENT

We are very grateful to Prof. G. Dashmukhe (Guide and Assistant Professor, department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik) For his expert guidance and continuous encouragement throughout the project.

At last, we must express our sincere, heartfelt gratitude to all staff members and students of the Department of Information technology who helped us directly or indirectly during this course of work.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Shetty, S. J., & Ramesh, V. (2021). pyResearchInsights—An open-source Python package for scientific text analysis. *Ecology and Evolution*, 00, 1–10. <https://doi.org/10.1002/ece3.8098>.
- [2] Paritosh D. Katre "NLP Based Text Analytics and Visualization of Political Speeches" *International Journal of Recent Technology*

and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019.

- [3] Akshaya Udgave, Prasanna Kulkarni: Text Mining and Text Analytics of Research Articles-- Palarch's Journal of Archaeology Of Egypt/Egyptology 17(6). ISSN 1567-214x.
- [4] John Rischa, Shawn Bohna, Steve Poteetb, Anne Kaob, Lesley Quachb and Jason Wub: Text Visualization, Encyclopedia of Machine Learning and Data Mining DOI 10.1007/978-1-4899-7502-7_837-1.
- [5] Text mining and visualization tools – Impressions of emerging capabilities: YunYun Yang, Lucy Akers, Thomas Klose, Cynthia Barcelon Yang World Patent Information 30 (2008) 280–293.
- [6] Diksha Khurana & Aditya Koli & Kiran Khatter & Sukhdev Singh: Natural language processing: state of the art, current trends, and challenges. <https://doi.org/10.1007/s11042-022-13428-4>.
- [7] Arnold, Taylor B. 2016. *cleanNLP: A Tidy Data Model for NaturalLanguage Processing*. <https://cran.r-project.org/package=cleanNLP>
- [8] Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
- [9] Mullen, Lincoln. 2016. *tokenizers: A Consistent Interface to Tokenize Natural Language Text*. <https://CRAN.R-project.org/package=tokenizers>.
- [10] Silge, Julia, and David Robinson. 2016. "tidytext: Text Mining and Analysis Using Tidy Data Principles in r." *JOSS* 1 (3). <https://doi.org/10.21105/joss.00037>.
- [11] 2014. "Tidy Data." *Journal of Statistical Software* 59 (1): 1–23. <https://doi.org/10.18637/jss.v059.i10>.