

Report on System created to Analyze Resumes

Introduction

The Intention of this analysis is to search into the complexities of resumes through a multifaceted text analysis approach. The analysis encompasses wide range of techniques, including full text cleaning, bi-grams and tri-grams generation, topic detection, word cloud creation, main statistics calculation, and generating networks.

Data Loading

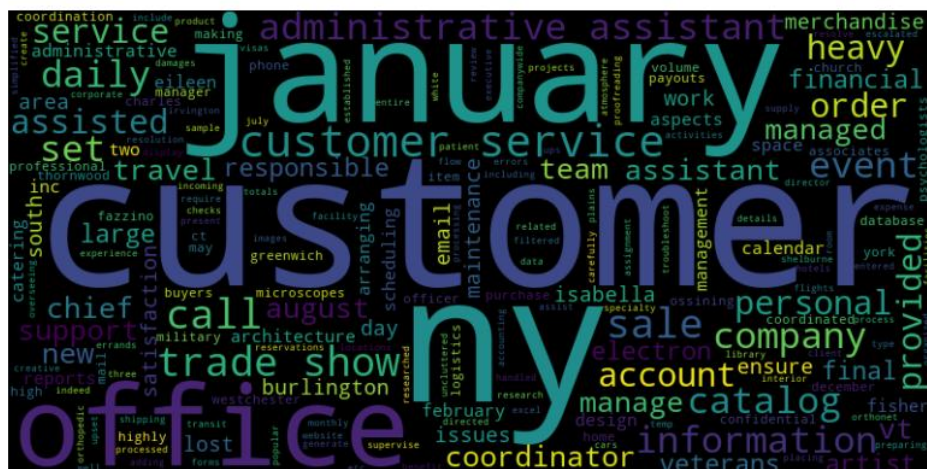
Resumes, sourced from the 'resumes.csv' file, represent a diverse collection of professional narratives. I have used two resumes from the dataset given and have derived information from them and compared with each other. This dataset forms the foundation of our exploration, allowing us to traverse through varied experiences, skills, and expressions. Resumes were meticulously loaded from a CSV file 'resumes.csv' containing textual data for extensive analysis. This diverse dataset enables a thorough exploration of the characteristics and patterns within the resumes.

Resumes Under Analysis

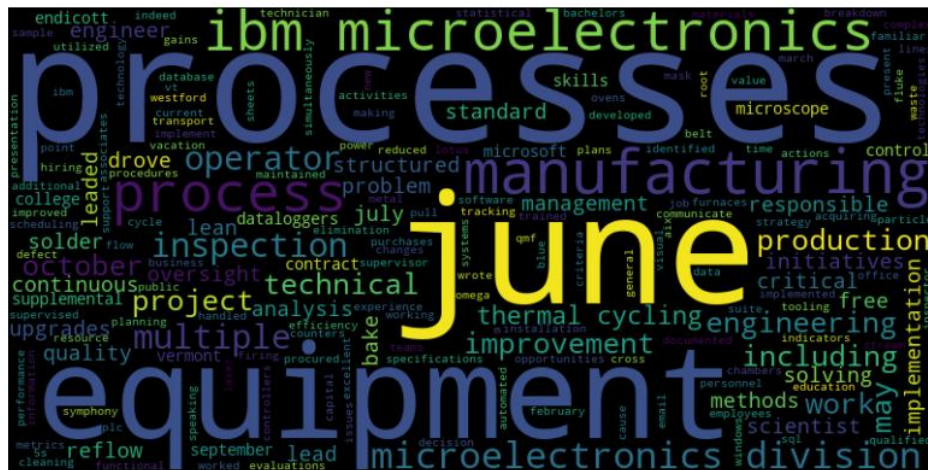
- **Resume 1:**

In the first resume, a detailed analysis reveals three prominent topics: "Customer service," "NY," and "Assistant." The predominant focus on customer service aligns with the individual's experience and expertise. The geographical association with New York (NY) suggests a localized professional background. The role of an assistant signifies a specific job function within the customer service domain.

The corresponding word cloud visually represents the key terms associated with these topics. Notably, terms such as "customer service," "NY," and "assistant" stand out, providing a visual summary of the resume's main themes.



Word cloud for Resume 1



Word cloud for Resume 2

Main statistics derived from the analysis spotlight a total *word count of 319*, featuring *208 unique words*. The *entropy*, a metric gauging information content, is computed at *7.44*. These numerical insights offer a quantitative perspective on the composition and complexity of the resume, providing a measure of the richness of information presented.

The analysis done provides a detailed and insightful breakdown of the identified topics within the second resume. Employing advanced topic modeling techniques, this output goes beyond a mere enumeration of terms by delving into the underlying thematic structures of the text. By quantitatively elucidating prevalent themes, the topic modeling output serves as a powerful tool for uncovering the latent patterns and subject matter nuances embedded within the resume.

It not only identifies key terms but also discerns the relationships and associations between them, contributing to a more profound understanding of the content and highlighting the thematic threads that weave through the narrative. This comprehensive breakdown adds a layer of sophistication to the analysis, offering a nuanced perspective on the underlying structure and thematic composition of the resume.

The network visualization presented below, illustrates word connections in the resume, portraying co-occurrences and relationships. This visual representation unveils the structural nuances of the language used, emphasizing the interconnected nature of key terms such as "test," "IBM," "applications," and "software." The network graphs offer a visual exploration of semantic connections between words. This graphical representation provides an additional layer of insight into the structural patterns and relationships embedded in the resumes.

Main Statistics

To quantitatively characterize the textual data, key statistics were calculated. These include the number of words, total unique words, and entropy. These metrics offer a quantitative lens through which to analyze the textual diversity and complexity of each resume.

Network Generation

Intricate network graphs were constructed to visually illustrate the relationships and co-occurrences between words. This network analysis offers insights into the semantic connections and structural patterns present in the resumes.

Comparison of both Resumes:

In a comparative analysis, Resume 1 demonstrates a focus on customer-centric roles in a specific geographical location, while Resume 2 emphasizes engagement in production-related endeavors, particularly with a notable association with IBM. The word clouds visually highlight these thematic distinctions, offering immediate insights into the resumes' core themes. Statistical measures provide a quantitative lens, displaying the varying word counts, unique words, and entropies, reflecting the diverse content structures. Network visualizations further underscore the structural nuances, illustrating the interconnectedness of key terms within each resume. In essence, the comparative analysis enhances our understanding of the distinct professional trajectories and thematic emphases captured in Resume 1 and Resume 2.

Conclusion

In conclusion, the comprehensive analysis of multiple resumes using the developed tool, encompassing full text cleaning, bi-gram and tri-gram generation, topic detection, word cloud visualization, main statistical metrics, and network generation, has provided complex and interesting insights into the diverse professional backgrounds and thematic orientations of individuals. The tool's application on the provided resumes highlighted distinct focal points in each, shedding light on key topics and contextual nuances. The comparative analysis of two specific resumes unveiled unique thematic emphases: Resume 1 emphasizing customer service roles in New York and Resume 2 displaying a focus on processes, manufacturing, and IBM, with a temporal dimension.

With the help of bi-grams and tri-grams, analysis was enriched by capturing nuanced word combinations, contributing to a more refined understanding of the resume content. The word clouds provided extensive summaries, accentuating prominent terms associated with each resume's central themes. Essential statistical measures, including word counts, unique words, and entropy, quantitatively gauged the richness and complexity of the content. Network visualizations visually represented word relationships, offering structural insights into the interconnected nature of key terms.

Overall, the tool's multifaceted approach facilitated a holistic exploration of resume content, uncovering latent patterns and thematic structures. This analytical framework not only identified key terms but also discerned their relationships and associations, contributing to a profound understanding of the content. The generated reports, accompanied by relevant graphs and screenshots, offer a comprehensive and visually engaging overview of the resumes, enhancing the comparative analysis. This endeavour aligns with the initial objective of creating a robust tool for resume analysis, providing valuable insights for both practitioners and learners.