# Unstructured Data Analysis Final Project

**Name: Tian-Ze Liu CID: 02261234**

**GitHub Website: https://github.com/tomanick/UDA-Final-Project**

## Statement

I confirm that the work presented in this assignment is entirely on my own.

## Final Project

In this final project, I embark on an exploratory journey through the literary landscapes crafted by the eminent Victorian author, Charles Dickens. My focus is centered on a comprehensive Unstructured Data Analysis (UDA) of six of his celebrated works: "A Tale of Two Cities", "The Pickwick Papers", "David Copperfield", "Great Expectations", "A Christmas Carol", and "Nicholas Nickleby". These novels, renowned for their vivid characterizations and intricate narratives, present a fertile ground for UDA, offering rich insights into Dickens' thematic explorations and stylistic nuances. Our analysis spans seven distinct topics, each delving into various facets of the texts – from Introduction, Bi-grams, (TF, IDF, TF-IDF), Emotion Analysis, Topic Modeling (LDA) to Lexical Diversity Analysis and Cross-Document Analysis (WordClouds). This project aims to blend traditional literary analysis with modern computational techniques, thereby unraveling the complexities of Dickens' prose and uncovering the underlying threads that weave these classic novels together.

## Introduction and Visualization of the Datasets

For Introduction and Visualization, let us delve into the linguistic essence of five acclaimed novels by Charles Dickens, revealing the depth of his narrative artistry. "A Tale of Two Cities" unfolds with 138,965 words, enriched by 10,538 unique terms, illustrating Dickens' intricate storytelling with an average word length of 4.32 characters. Regarding "The Pickwick Papers", his earliest novel, encompasses a voluminous 303,110 words, with a diverse vocabulary of 16,529 unique words, and a slightly longer average word length of 4.49 characters. "David Copperfield", the most extensive with 357,861 words, reflects autobiographical nuances through its 16,096 unique words, averaging 4.16 characters each. "Great Expectations" captures 187,517 words and 10,992 unique terms, balancing narrative complexity with a 4.11-character average word length. "A Christmas Carol", in its succinct 32,430 words, efficiently uses 4,717 unique words, averaging 4.36 characters, showcasing Dickens' ability to convey profound stories concisely. Finally, "Nicholas Nickleby" spans 325,340 words with 17,371 unique terms, averaging 4.45 characters each, showcasing Dickens' vivid storytelling and linguistic depth. This analysis not only highlights the quantitative aspects of Dickens' work but also underlines his linguistic diversity and storytelling prowess.

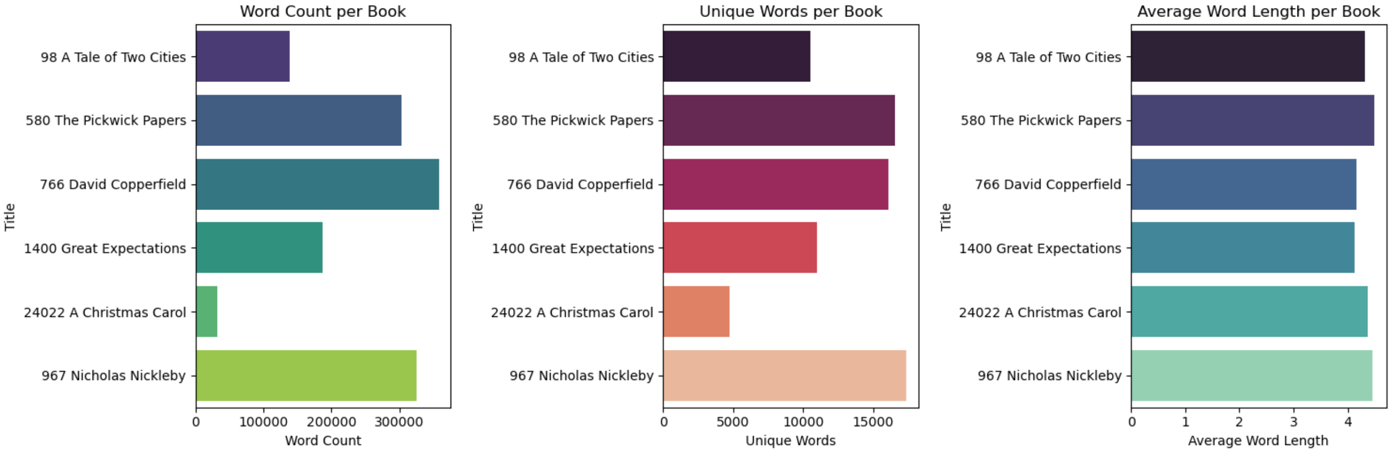


**Figure 1. The Visualization of Three Topics of the Six Books**

**Bi-grams**

In a compelling exploration of Charles Dickens' literary mastery, a bigram analysis of six seminal works—"A Tale of Two Cities", "The Pickwick Papers", "David Copperfield", "Great Expectations", "A Christmas Carol", and "Nicholas Nickleby"—uncovers the intricate layers of his narrative craft. Bigrams, sequential pairs of words, serve as insightful portals into the thematic depths, character dynamics, and the distinctive narrative styles that hallmark Dickens' storytelling. "A Tale of Two Cities" resonates with historical richness, as bigrams such as 'Madame Defarge' and 'Doctor Manette' not only illuminate pivotal characters but also embed the narrative within the turbulent fabric of the French Revolution. Bigrams like 'Saint Antoine' and 'Monsieur Marquis' further enrich the historical context, while 'Mender of Roads' subtly reflects on societal roles and destinies amidst historical upheaval. In the more lighthearted "The Pickwick Papers," the bigrams 'sir replied' and 'replied Sam' typify the novel's vibrant dialogue, painting a lively picture of Victorian society. Characters like 'Bob Sawyer' and phrases such as 'fat boy' and 'old gentleman' emerge vividly, contributing to a satirical yet affectionate portrayal of the era.

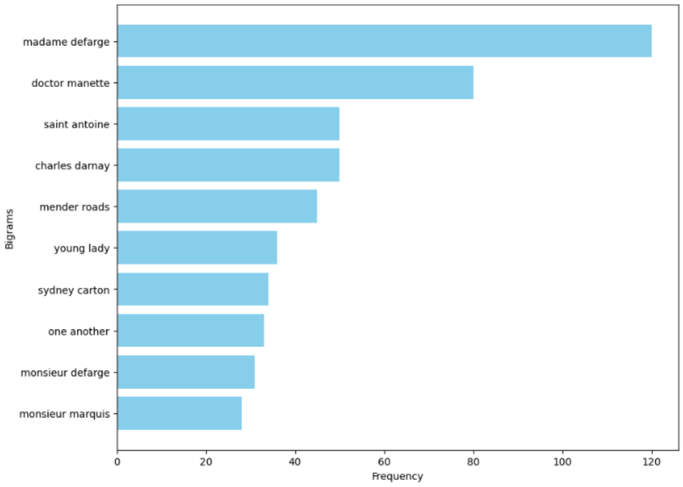


**Figure 2. The bi-grams of A Tale of Two Cities**

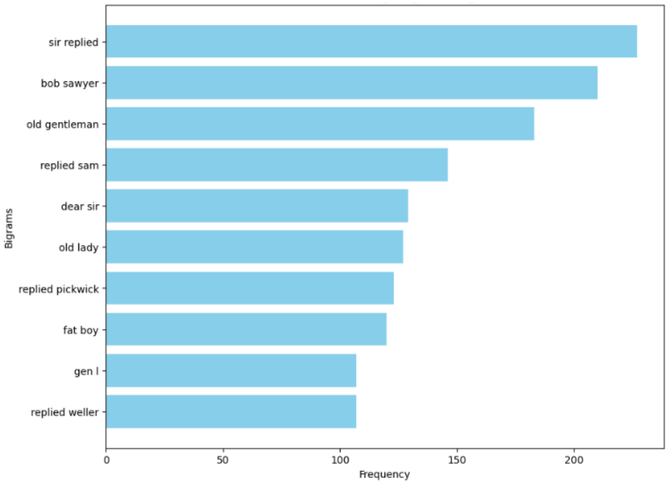
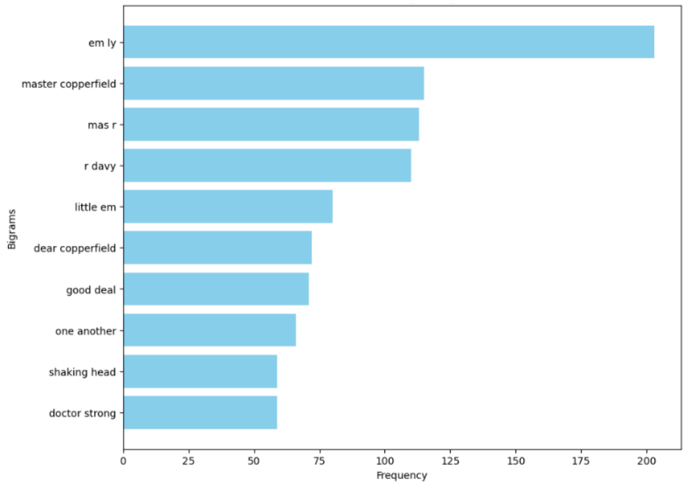


**Figure 3. The bi-grams of The Pickwick Papers**

"David Copperfield" takes a turn towards the personal, with bigrams like 'Em'ly' and 'Master Copperfield' hinting at intimate relationships and emotional depth. This narrative intimacy is further encapsulated in bigrams such as 'little Em'ly' and 'dear Copperfield', painting a poignant picture of the complexities of human relationships. "Great Expectations" employs bigrams like 'dear boy' and 'old chap', encapsulating the protagonist Pip's developmental journey. Bigrams such as 'young man' and 'Pip Joe' mirror the evolving nature of the characters' relationships and identities, embodying the novel's coming-of-age essence.



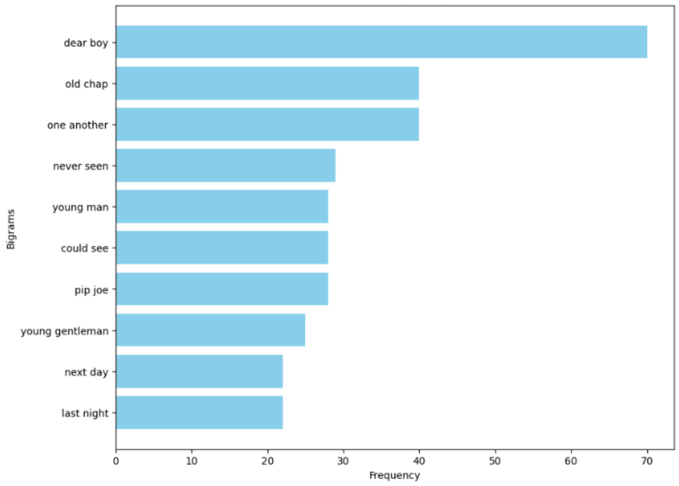
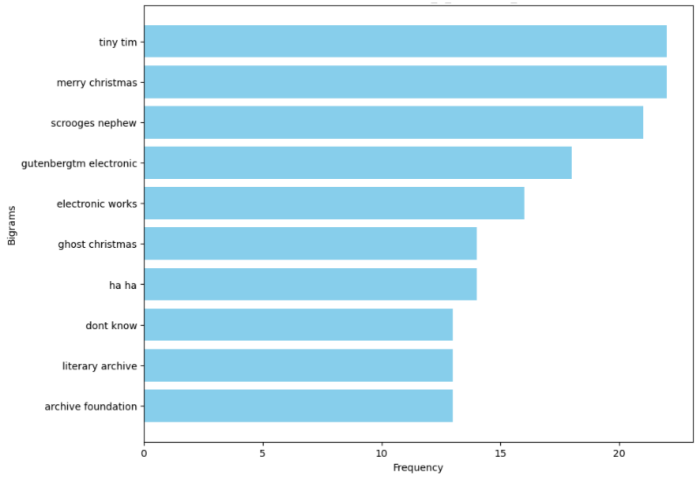


**Figure 4. The bi-grams of David Copperfield** **Figure 5. The bi-grams of Great Expectations**

"A Christmas Carol" distinctly showcases its festive spirit through bigrams like 'Tiny Tim' and 'merry Christmas', while 'Scrooge's nephew' and 'ghost Christmas' encapsulate the novel's themes of redemption and the supernatural. Lastly, in "Nicholas Nickleby," the top bigrams illuminate key characters and themes. 'Sir mulberry' and 'la creevy' highlight central figures, emphasizing Dickens' focus on character development. Frequent use of 'dont know' and 'young lady' in dialogue showcases Victorian societal interactions. Bigrams like 'replied nicholas' and 'ralph nickleby' reflect the narrative's depth, while 'madame mantalini' and 'newman noggs' emphasize Dickens' skill in creating memorable, diverse characters.



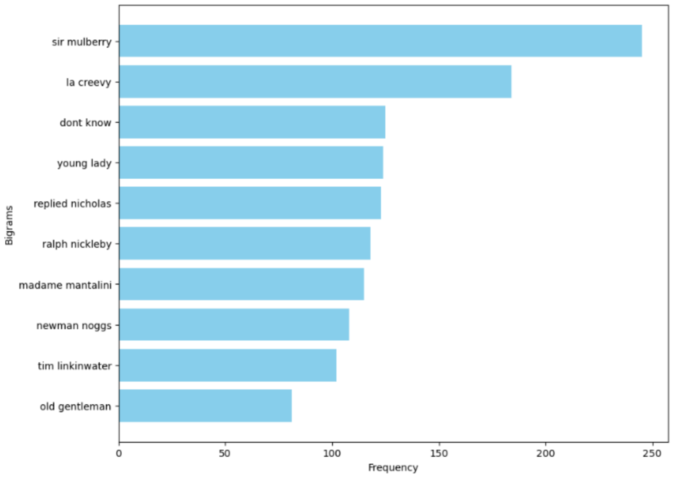


**Figure 6. The bi-grams of A Christmas Carol** **Figure 7. The bi-grams of Nicholas Nickleby**

**Insightful Findings and Conclusion of Bi-grams**

This bigram analysis of Charles Dickens' novels vividly demonstrates the unique aspects of his narrative style, characterized by rich dialogues, intricate character portrayals, and complex thematic structures. Utilizing computational text analysis, this study enhances our appreciation of Dickens' classic works, revealing the intricate interplay of language and narrative that underpins his enduring literary creations. Despite the diverse settings and tones, there's a common thread in Dickens' use of bigrams to foreground relationships and societal interactions, whether through humor, struggle, or historical context. These bigrams consistently highlight his focus on character development and his skill in crafting complex social narratives. This analysis sheds light on the unique qualities of each novel while weaving together the stylistic elements that define Dickens' literary legacy, a legacy characterized by vibrant character portrayals, dynamic dialogues, and a profound understanding of human and societal dynamics.

## TF, IDF and TF-IDF

The exploration of Charles Dickens' literary works through Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) analysis offers a unique lens to examine his narrative style and thematic intricacies. This study encompasses "A Tale of Two Cities", "The Pickwick Papers", "David Copperfield", "Great Expectations", "A Christmas Carol", and "Nicholas Nickleby". By removing common stopwords and specific words like 'Project' and 'Gutenberg' for precise data preprocessing, we delve into the linguistic essence of these novels.

In "A Tale of Two Cities", the TF analysis reveals frequent use of words like 'lorry', 'defarge', and 'man', indicating a focus on specific characters and the human element within the historical backdrop. The TF-IDF results, highlighting 'lorry', 'defarge', 'manette', and 'pross', further accentuate these characters' unique roles in the narrative. This suggests a complex interplay of personal stories amidst the turmoil of the French Revolution, showcasing Dickens' ability to weave individual human experiences into a broader historical context.

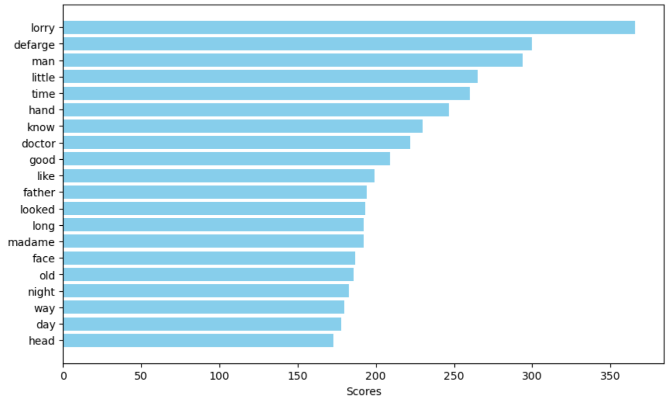


**Figure 8. The Top 20 TF of A Tale of Two Cities**

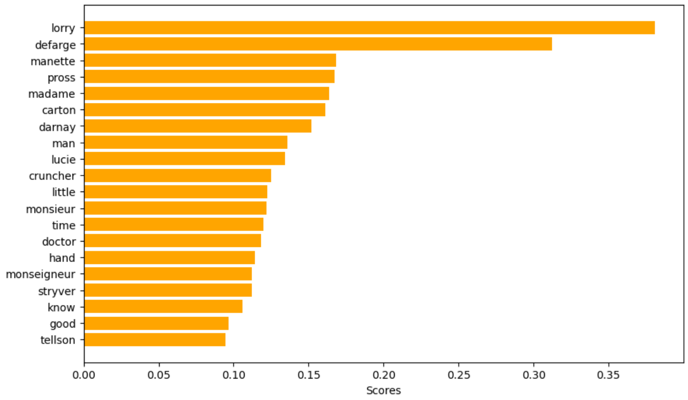


**Figure 9. The Top 20 TF-IDF of A Tale of Two Cities**

"The Pickwick Papers" displays a distinct character-driven narrative, evidenced by the predominance of 'pickwick', 'sir', 'sam', and 'weller' in the TF analysis. The TF-IDF results reinforce the central importance of 'pickwick' and 'sam', suggesting their pivotal roles in the novel. The prevalence of these character names, coupled with words like 'old' and 'gentleman', reflect the novel's exploration of social interactions and Victorian society, offering a blend of humor and a light-hearted view of English life.

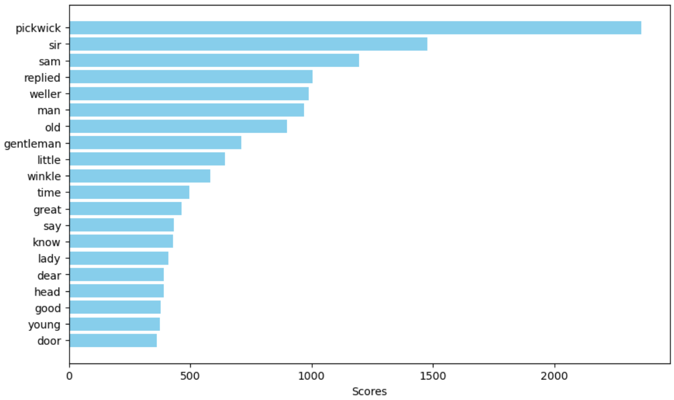


**Figure 10. The Top 20 TF of The Pickwick Papers**

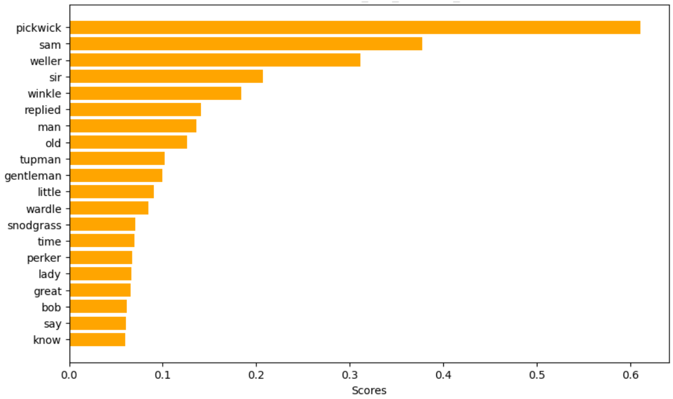


**Figure 11. The Top 20 TF-IDF of The Pickwick Papers**

In "David Copperfield", the frequent appearance of 'little', 'micawber', 'aunt', and 'know' in the TF analysis underscores the novel's emphasis on personal relationships and introspection. The TF-IDF analysis, stressing 'micawber', 'peggotty', and 'aunt', highlights these characters as central to the narrative, emphasizing their significant influence on the protagonist's life. This suggests Dickens' focus on character development and emotional depth, characteristic of this semi-autobiographical novel.

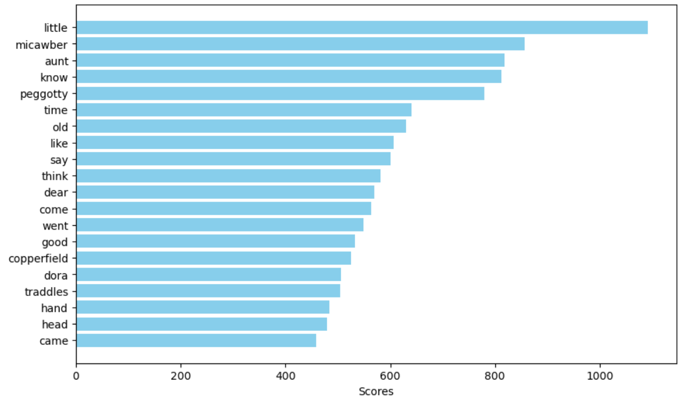


**Figure 12. The Top 20 TF of David Copperfield**

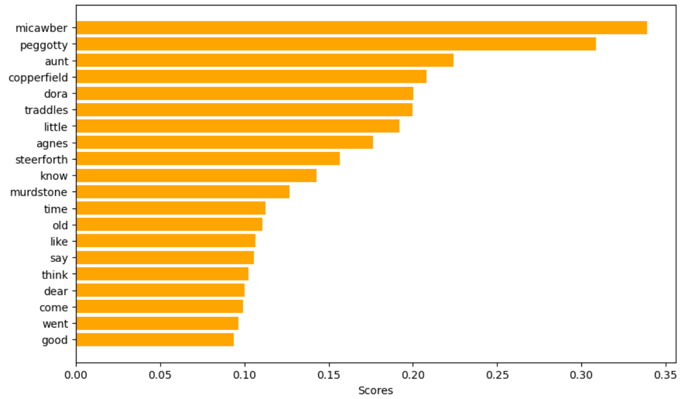


**Figure 13. The Top 20 TF-IDF of David Copperfield**

"Great Expectations" shows a high frequency of 'joe', 'know', 'come', and 'little' in the TF analysis, with 'joe', 'pip', and 'havisham' being prominent in the TF-IDF results. This indicates the novel's focus on character growth and the protagonist's journey. The blend of names and common words like 'time' and 'like' reflects the novel's exploration of time, identity, and personal evolution, underlining its status as a coming-of-age story interwoven with social critique.

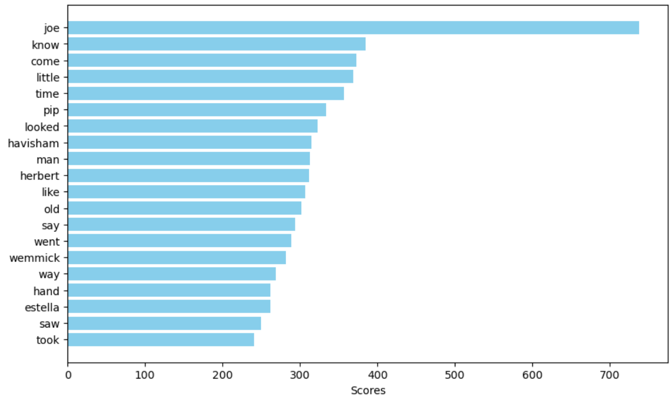


**Figure 14. The Top 20 TF of Great Expectations**

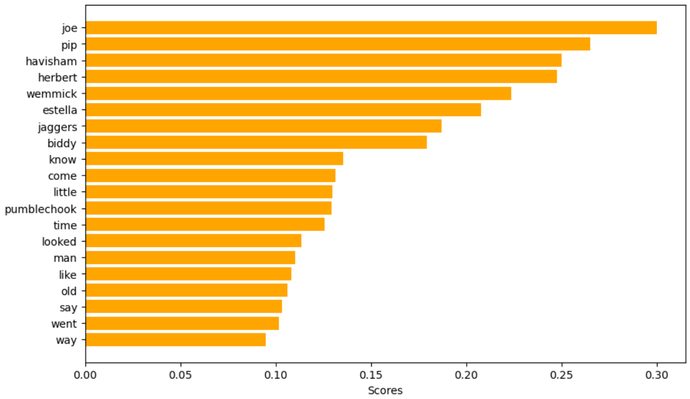


**Figure 15. The Top 20 TF-IDF of Great Expectations**

"A Christmas Carol" is characterized by a high frequency of 'scrooge', 'ghost', 'christmas', and 'spirit' in the TF analysis, with the TF-IDF results also highlighting 'scrooge', 'ghost', and 'christmas'. This underlines the novel's focus on transformation, redemption, and the essence of the Christmas spirit. The presence of words like 'man', 'time', and 'good' in both analyses emphasizes the novel's moral and temporal themes, encapsulating the story's depth and complexity.

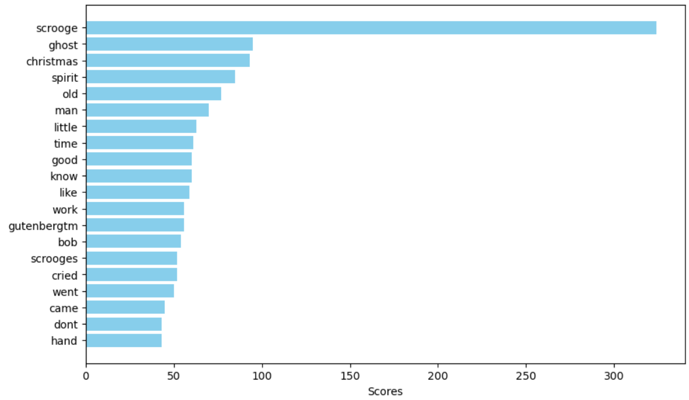


**Figure 16. The Top 20 TF of A Christmas Carol**

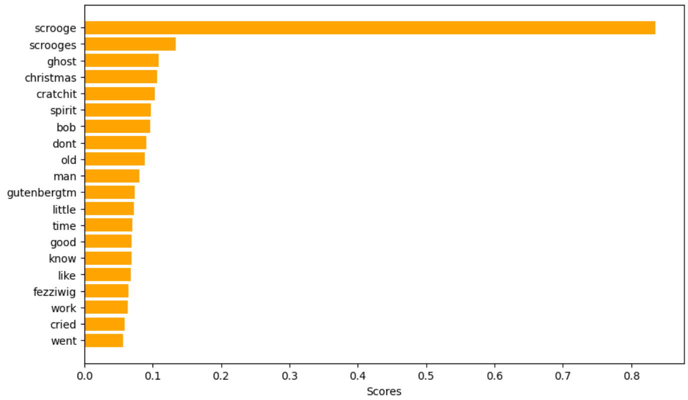
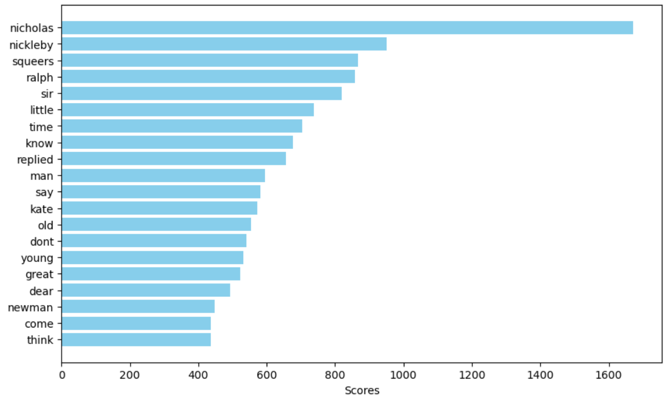


**Figure 17. The Top 20 TF-IDF of A Christmas Carol**

"Nicholas Nickleby" features 'nicholas', 'nickleby', and 'squeers' as key words in the TF analysis, with TF-IDF results also emphasizing 'nicholas' and 'nickleby'. This suggests a narrative centered around the protagonist's experiences and the challenges he faces. The presence of character names and words like 'little', 'time', and 'know' in both analyses reflects the novel's blend of personal narrative and social criticism, showcasing Dickens' evolving narrative style and his commitment to addressing societal issues through character-driven stories.



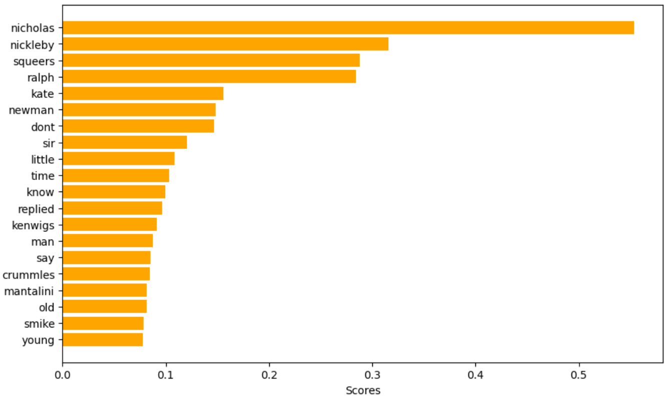


**Figure 18. The Top 20 TF of Nicholas Nickleby** **Figure 19. The Top 20 TF-IDF of Nicholas Nickleby**

### Insightful Findings and Conclusion of TF, IDF and TF-IDF

The combined analysis of TF and TF-IDF offers a rich tapestry of insights. TF highlights the most commonly used words, reflecting the surface-level focus of the narratives, such as frequent dialogues, character mentions, and thematic elements. Conversely, TF-IDF underscores words that are not only frequent but also carry a significant weight in the context of each novel, revealing unique aspects of character development, thematic depth, and narrative style. This dual approach is instrumental in literary analysis, providing a quantitative lens to assess textual elements. It allows for a nuanced understanding of Dickens' narrative techniques, showing how he weaves common linguistic elements into complex, distinctive storylines. The analysis highlights his ability to create vibrant characters and immersive worlds, using both recurring and uniquely significant terms to shape his narratives. In summary, the TF and TF-IDF analyses combined paint a comprehensive picture of Charles Dickens' storytelling mastery. They reveal his skill in balancing common linguistic patterns with distinct narrative elements, creating richly layered novels that resonate across time and literary cultures.

### Emotion Analysis

The emotion analysis of six renowned novels by Charles Dickens reveals a rich tapestry of emotional undertones that define the essence of his storytelling. This analysis uses the NRC Emotion Lexicon to quantify emotions such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, along with their positive or negative nature. It offers profound insights into the emotional landscape that Dickens crafts in each narrative.

In "A Tale of Two Cities", this novel set against the backdrop of the French Revolution, exhibits high levels of anger, fear, and sadness, resonating with the tumultuous and violent era it depicts. However, it also shows significant anticipation and joy, suggesting a balance of hope and despair in the narrative. The prevalence of positive emotions, despite the grim setting, underscores Dickens' ability to weave a tale of resilience and human spirit amidst adversity.

In contrast, "The Pickwick Papers" displays an abundance of joy, anticipation, and trust, aligning with its humorous and satirical tone. Although it has moments of anger and fear, the dominant positive emotions reflect the novel's overall light-hearted and jovial nature, capturing the adventures and misadventures of its characters with a sense of warmth and optimism.

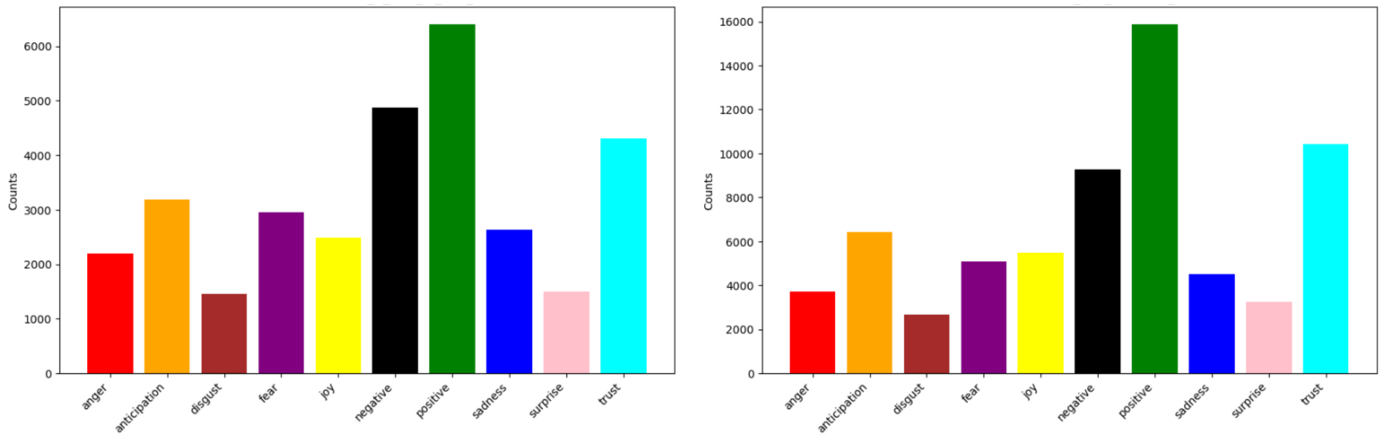


**Figure 20 - 21. The Emotion Analysis of A Tale of Two Cities and The Pickwick Papers**

"David Copperfield" shows a complex emotional range, with high levels of anticipation, joy, and trust, indicative of the protagonist's journey and personal growth. The presence of negative emotions like anger and sadness, along with fear and disgust, illustrates the challenges and conflicts faced by the characters, portraying a realistic and emotionally rich narrative. "Great Expectations" presents a balanced emotional spectrum with notable instances of fear, sadness, and anger, reflecting the protagonist Pip's struggles and conflicts. However, emotions like joy, anticipation, and trust are also prominent, depicting the novel's exploration of personal development and redemption.

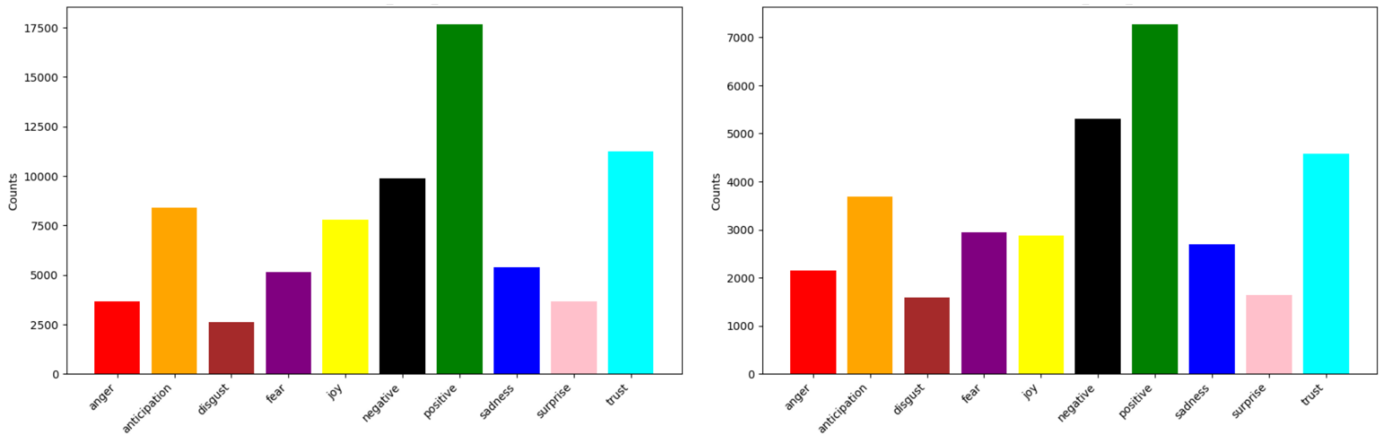


**Figure 22 - 23. The Emotion Analysis of David Copperfield and Great Expectations**

"A Christmas Carol" stands out with its focus on positive emotions like joy and trust, fitting for a story about transformation and redemption. The presence of fear and sadness captures the initial plight of the protagonist Scrooge, while the high positive scores indicate the novel's uplifting message of change and the spirit of Christmas. In "Nicholas Nickleby", the emotion analysis highlights a rich blend of feelings, mirroring the novel's multifaceted narrative. Dominant positive emotions like joy and trust reflect the story's underlying optimism and the resilience of its characters. The notable presence of negative emotions such as anger and sadness capture the characters' struggles, adding depth to the narrative's portrayal of societal challenges and personal triumphs.

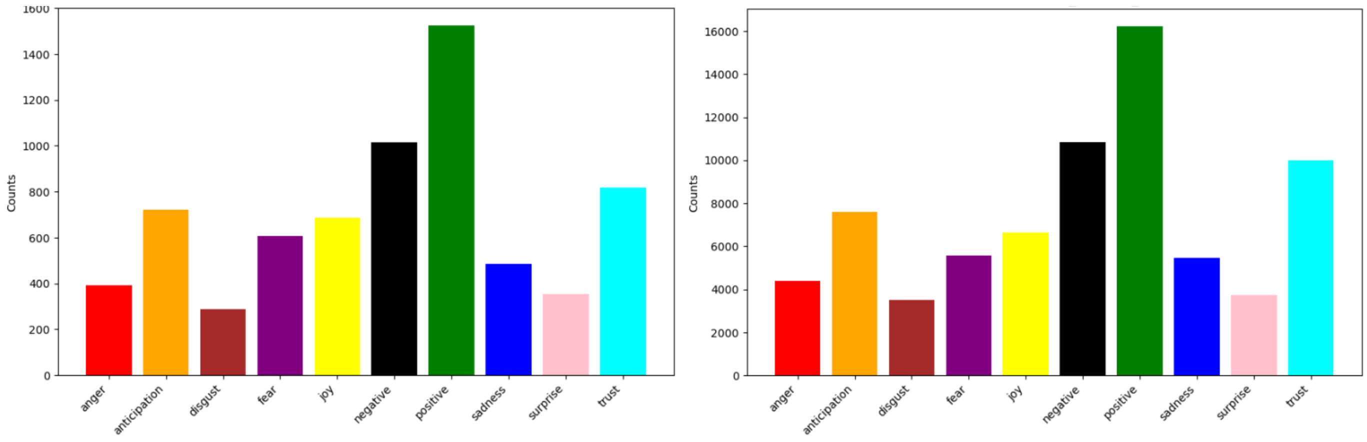


**Figure 24 - 25. The Emotion Analysis of A Christmas Carol and Nicholas Nickleby**

### Insightful Findings and Conclusion of Emotion Analysis

Through this emotion analysis, Dickens' storytelling prowess is evident in his ability to evoke a wide range of emotions that align with the themes and narratives of his novels. From the turbulent and hopeful "A Tale of Two Cities" to the heartwarming transformation in "A Christmas Carol", and the intricate exploration of societal challenges and personal resilience in "Nicholas Nickleby", Dickens masterfully uses emotional undertones to enrich his narratives, making them resonate with readers across generations. This analysis not only enhances our understanding of his work but also highlights the emotional depth and complexity that underpin his timeless classics.

### Topic Modeling (LDA)

The Latent Dirichlet Allocation (LDA) analysis of a combined corpus of six novels by Charles Dickens reveals intriguing insights into the thematic and narrative elements prevalent across his works. This analysis, which generated six topics each represented by the ten most significant words, offers a glimpse into the recurring themes and character dynamics that Dickens frequently employed. After running LDA, we could get the following results:

**6 topics, top 10 words**

Topic 0: pickwick, sir, sam, replied, weller, man, old, gentleman, little, winkle

Topic 1: bearers, swing, woods, repentance, observant, sailing, cheery, speckled, terrors, negative

Topic 2: lorry, defarge, man, little, time, hand, know, doctor, good, like

Topic 3: nicholas, nickleby, squeers, ralph, sir, little, time, know, replied, man

Topic 4: little, micawber, aunt, know, peggotty, time, old, like, say, think

Topic 5: joe, know, come, little, time, pip, looked, man, havisham, like

### Topic Interpretations

**Topic 0 - "The Pickwick Papers" Focus**: This topic is heavily centered around "The Pickwick Papers". The frequent mention of 'Pickwick', 'Sir', 'Sam', 'replied', 'Weller', 'man', 'old', 'gentleman', 'little', and 'Winkle' indicates the novel's character-driven narrative, with a focus on dialogue and social interactions among its vibrant cast of characters. This reflects the humorous and light-hearted nature of the story.

**Topic 1 - Abstract and Descriptive Elements**: This topic seems to capture more abstract and descriptive elements with words like 'bearers', 'swing', 'woods', 'repentance', 'observant', 'sailing', 'cheery', 'speckled', 'terrors', and 'negative'. These terms suggest scenes or themes involving nature, reflection, and possibly some darker or more contemplative moments in Dickens' storytelling.

**Topic 2 - "A Tale of Two Cities" Focus**: Dominated by 'Lorry', 'Defarge', 'man', 'little', 'time', 'hand', 'know', 'doctor', 'good', 'like', this topic clearly represents "A Tale of Two Cities". It highlights the novel’s key characters and themes, underscoring the intense, historical, and emotionally charged narrative of the French Revolution.

**Topic 3 - "Nicholas Nickleby" Theme**: This topic focuses on "Nicholas Nickleby", with recurring names such as 'Nicholas', 'Nickleby', 'Squeers', 'Ralph', and words like 'sir', 'little', 'time', 'know', 'replied', 'man'. It reflects the novel's character dynamics and the social issues Dickens explores, like education and societal hierarchies.

**Topic 4 - "David Copperfield" Elements**: Highlighting 'little', 'Micawber', 'aunt', 'know', 'Peggotty', and others, this topic is indicative of "David Copperfield". It emphasizes the novel's focus on personal relationships, character development, and the protagonist's emotional journey.

**Topic 5 - "Great Expectations" Focus**: With words like 'Joe', 'know', 'come', 'little', 'time', 'Pip', 'looked', 'man', 'Havisham', 'like', this topic is likely associated with "Great Expectations". It captures the essence of the novel’s coming-of-age story, the protagonist Pip’s experiences, and his interactions with other significant characters like Joe and Miss Havisham.

**Insightful Findings and Conclusion of Topic Modeling (LDA)**

The LDA analysis of Charles Dickens' six novels offers a high-level overview that encapsulates his storytelling artistry. It reveals a rich tapestry of thematic clusters, spotlighting the predominance of main characters' names, which are central to each narrative. This reflects Dickens' focus on character-driven stories, where individuals not only drive the plot but also embody the themes and societal observations he wishes to convey. The analysis also uncovers key descriptive elements, indicative of Dickens' ability to vividly paint scenes and settings, immersing readers in the time and place of his narratives. Overall, the LDA results underscore Dickens' narrative diversity, blending personal journeys with broader societal commentary, and confirming his status as a master storyteller with a distinct ability to weave complex, multifaceted literary works.

**Lexical Diversity Analysis**

| Book | Lexical Richness |
|---|---|
| A Tale of Two Cities | 0.0742 |
| The Pickwick Papers | 0.0529 |
| David Copperfield | 0.0440 |
| Great Expectations | 0.0621 |
| A Christmas Carol | 0.1525 |
| Nicholas Nickleby | 0.0541 |

**Table 1. The Lexical Richness of Six Books**

The linguistic landscape of Charles Dickens' literature presents a fascinating study in the use of language, as revealed through an analysis of lexical richness in six of his novels. Lexical richness, a measure of the diversity of vocabulary in a text, offers insights into an author's narrative style and thematic choices. In this analysis, we delve into the six novels examining the unique lexical footprint of each.

"A Tale of Two Cities" (Lexical Richness: 0.0742) exhibits a moderate level of lexical diversity, balancing a rich vocabulary with narrative accessibility. The historical setting and the intertwining of personal and societal struggles are reflected in a varied but not overly complex language. "The Pickwick Papers" (Lexical Richness: 0.0529) shows a lower lexical richness, which aligns with its light-hearted, comedic narrative style. The simplicity in language facilitates the humorous and satirical tone of the novel. "David Copperfield" (Lexical Richness: 0.0440) has the lowest lexical richness among the six. This could be indicative of its introspective nature, focusing more on character development and emotional depth than on linguistic complexity. "Great Expectations" (Lexical Richness: 0.0621) exhibits a moderate level of lexical diversity, it reflects the blend of a coming-of-age story with social critique. The language is diverse enough to capture various characters and settings, yet remains straightforward. "A Christmas Carol" (Lexical Richness: 0.1525) stands out with the highest lexical richness. This can be ascribed to its vivid descriptive language and the intricate blend of both realistic and supernatural elements. The depth of its moral and social themes demands a varied and substantial vocabulary, contributing to the novel's distinctive linguistic richness. "Nicholas Nickleby" (Lexical Richness: 0.0541) shows a lexical richness slightly above "The Pickwick Papers", possibly due to its wide array of characters and Dickens' maturing narrative style.

**Insightful Findings and Conclusion of Lexical Diversity Analysis**

These findings illuminate the breadth of Dickens' linguistic prowess. From the focused narrative of "David Copperfield" to the engaging storytelling in "Nicholas Nickleby", Dickens' lexical choices vividly reflect his storytelling genius. Each novel's lexical richness not only echoes its thematic depth and narrative complexity but also showcases Dickens' evolution as a writer, adept at crafting diverse narratives that span a spectrum of societal, moral, and personal themes.

**Cross-Document Analysis**

The Word Cloud results for each of Charles Dickens' six novels offer a visual representation of the most prominent words used in each book. The size of each word in the cloud is proportional to its frequency in the text. These words may represent names of characters, narrative style, themes, or significant things.

In "A Tale of Two Cities", the prominence of words like 'one', 'lorry', 'would', and 'hand' emphasizes the novel's focus on individual experiences and personal narratives within the backdrop of historical upheaval. The frequent appearance of 'time', 'man', and 'little' aligns with the novel's exploration of human endurance and societal change. As for "The Pickwick Papers", the word cloud is characterized by the predominance of 'pickwick', 'sir', 'sam', and 'weller', emphasizing the novel's focus on its central characters. The presence of words such as 'man', 'old', and 'gentleman' highlights the social aspects of the story, reflecting on the various societal roles and interactions that are pivotal to the narrative. These elements combined illustrate the novel's emphasis on character development and social exploration within the framework of Victorian society.

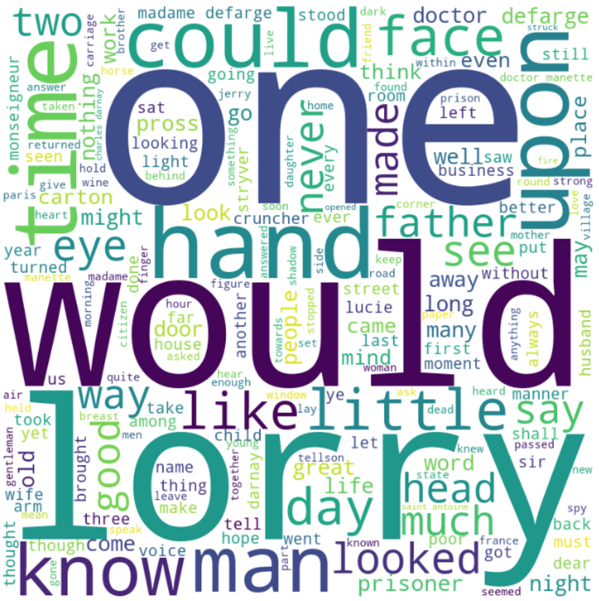


**Figure 26. The WordCloud of A Tale of Two Cities**

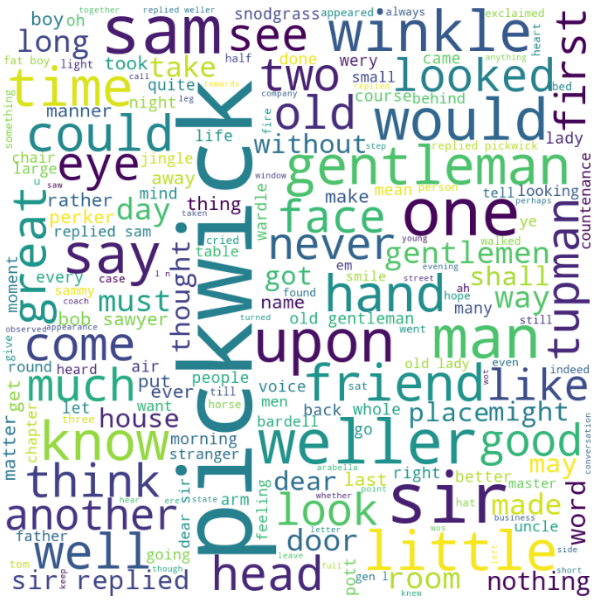


**Figure 27. The WordCloud of The Pickwick Papers**

In "David Copperfield", the prominence of words like 'little', 'know', and 'would', along with character names such as 'micawber' and 'peggotty', points to a narrative rich in personal relationships and emotional depth. This is reflective of the novel's semi-autobiographical nature, where intimate character connections and introspective insights are central to the storytelling. For "Great Expectations", the prevalence of words like 'joe', 'would', 'know', and 'one', complemented by terms such as 'time', 'come', and 'hand', eloquently encapsulates the novel's exploration of growth, the passage of time, and personal evolution. These elements are at the heart of its coming-of-age narrative, vividly portrayed through the word cloud analysis.

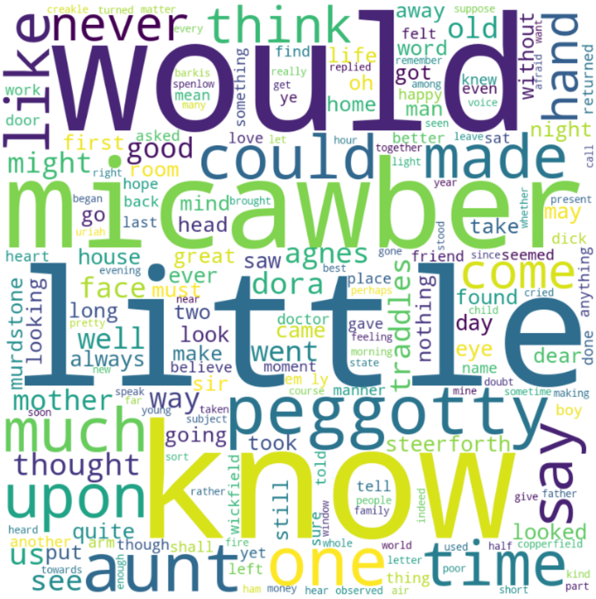


**Figure 28. The WordCloud of David Copperfield**

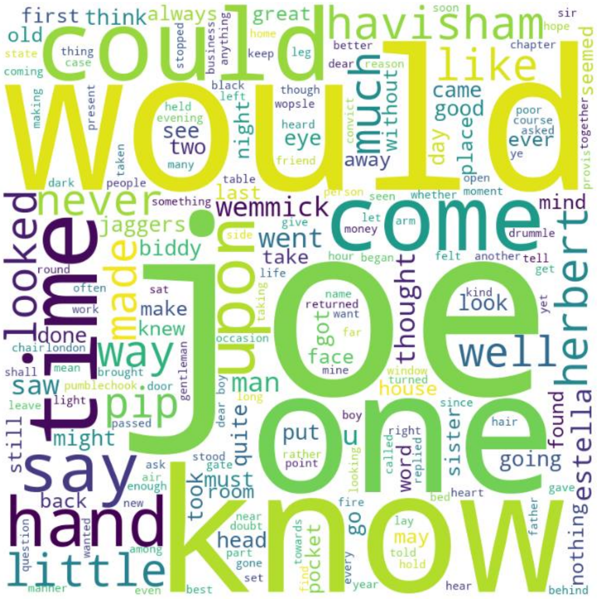


**Figure 29. The WordCloud of Great Expectations**

In "A Christmas Carol", the prominence of 'scrooge', 'spirit', 'ghost', and 'christmas' in the word cloud underscores themes of redemption and holiday spirit, enhanced by words like 'man', 'time', and 'good', reflecting its moral core. Meanwhile, "Nicholas Nickleby" is shaped by 'nicholas', 'one', and 'would', focusing on the protagonist's experiences. The inclusion of 'nickleby', 'squeers', and 'ralph' reveals its character-driven plot, with terms like 'little', 'time', and 'say' highlighting a narrative interweaving personal tales with broader social commentary.

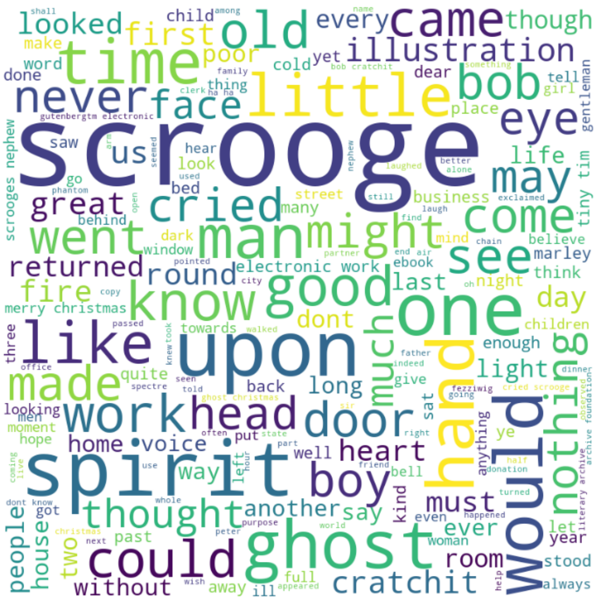


**Figure 30. The WordCloud of A Christmas Carol**

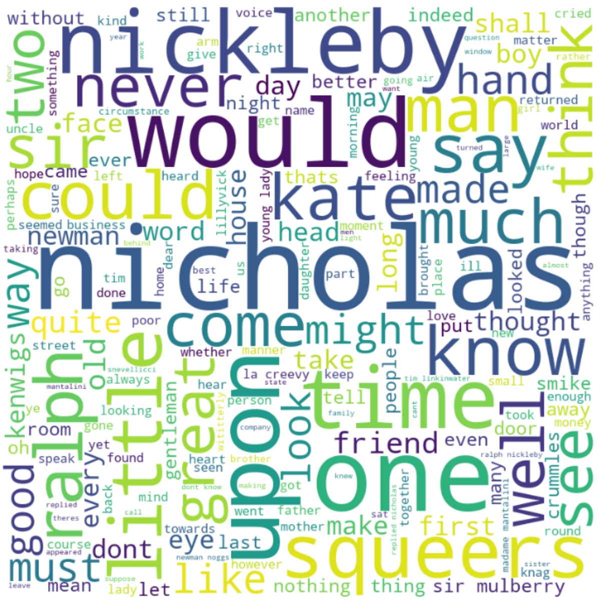


**Figure 31. The WordCloud of Nicholas Nickleby**

### Insightful Findings and Conclusion of Cross-Document Analysis

The word clouds from Charles Dickens' novels brilliantly showcase his narrative artistry. They reveal his focus on intricate character development and thematic depth, as evidenced by the frequent mention of character names and elements of daily life. These insights into actions and everyday experiences underscore Dickens' talent for crafting rich, descriptive narratives. His skillful portrayal of characters and their surroundings immerses readers in his vividly constructed worlds. This analysis highlights Dickens' ability to weave complex stories that delve into the nuances of human nature and society, solidifying his status as a master storyteller.

### Final Conclusion

The final project in the Unstructured Data Analysis (UDA) of six novels by Charles Dickens - "A Tale of Two Cities", "The Pickwick Papers", "David Copperfield", "Great Expectations", "A Christmas Carol", and "Nicholas Nickleby" - presents a comprehensive examination of Dickens' literary artistry through modern computational techniques. This analysis covers various facets of the texts, including Bi-grams, Term Frequency (TF), Inverse Document Frequency (IDF) and Term Frequency-Inverse Document Frequency (TF-IDF), Emotion Analysis, Latent Dirichlet Allocation (LDA) for Topic Modeling, Lexical Diversity Analysis, and Cross-Document Analysis using WordClouds.

**Bi-gram Analysis** vividly demonstrated the intricate narrative styles and thematic depths of Dickens' works. The bigrams effectively highlight key relationships, character dynamics, and the distinct narrative styles that hallmark Dickens' storytelling. For example, in "A Tale of Two Cities", bigrams like 'Madame Defarge' and 'Doctor Manette' illuminate pivotal characters and embed the narrative within the French Revolution's historical context.

**TF and TF-IDF Analysis** offers insights into the most commonly used words and those carrying significant weight in the context of each novel. This dual approach is instrumental in literary analysis, revealing the unique aspects of character development, thematic depth, and narrative style in Dickens' works.

**Emotion Analysis** uses the NRC Emotion Lexicon to quantify emotions such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. This analysis underscores Dickens' ability to evoke a wide range of emotions that align with the themes and narratives of his novels, enhancing the emotional depth and complexity of his timeless classics.

**Topic Modeling (LDA)** illuminates his narrative mastery, revealing thematic clusters and character-focused storytelling. It highlights Dickens' skill in blending personal journeys with societal commentary, showcasing his ability to craft multifaceted narratives that intricately weave character dynamics with vivid descriptive elements, affirming his status as a literary virtuoso.

**Lexical Diversity Analysis** illuminates the breadth of Dickens' linguistic prowess, and each novel's lexical richness reflects its thematic depth and narrative scope. This analysis demonstrates Dickens' growth as a writer, adept at weaving intricate narratives across a spectrum of societal, moral, and personal themes.

**Cross-Document Analysis** uses WordClouds to highlight the frequent mention of character names and elements of daily life, showcasing Dickens' focus on intricate character development and thematic depth.

In conclusion, this Unstructured Data Analysis project, blending traditional literary analysis with modern computational techniques, unravels the complexities of Dickens' prose. It provides a comprehensive overview of his narrative techniques, thematic explorations, and stylistic nuances. The analysis underscores his ability to create vibrant characters, immersive worlds, and complex storylines, solidifying his status as a master storyteller. Through this meticulous examination, Dickens' enduring legacy in literature is reaffirmed, showcasing his profound understanding of human and societal dynamics, and his unique narrative artistry that continues to resonate with readers across generations.