The Study of Religion and History
Volume 3 issue 2 (2024)
www.srhjournal.com
ISSN(Print): 3006-3329 ISSN(Online): 3006-3337

# A CORPUS BASED LEXICAL AND SEMANTIC MAPPING OF HERMAN MELVILLE'S MOBYDICK

***Zunaira Rasheed***
*M.Phil Scholar, NCBA&E Alhamra University Bahawalpur*
*zainijoyia9292@gmail.com*
***Humna Fatima***
*M.Phil Scholar, NCBA&E Alhamra University Bahawalpur*
*humnayasin597@gmail.com*
***Sidra Malik***
*M.Phil Scholar, NCBA&E Alhamra University Bahawalpur*
*za7450228@gmail.com*
***Dr. Muhammad Arfan Lodhi****
*Higher Education Department, Punjab*
*samaritan_as@hotmail.com*

***Abstract:***

*This study explores how digital tools, specifically Voyant Tools, can enhance literary analysis through computational approaches. By applying Digital Humanities (DH) methods to Herman Melville's Moby Dick or The Whale (1851), the research combines quantitative text analysis with qualitative interpretation to uncover linguistic and thematic patterns. Using Voyant's modules such as Cirrus (word cloud) Trends (frequency graph) and Contexts (keyword in context) the study examines recurrent words and ideas to identify the novel's central motifs including whale, sea, captain, man, go and queequeg. These frequently occurring terms highlight major themes such as the struggle between man and nature, obsession and fate, faith and divinity and cultural diversity aboard the Pequod. The research employs a mixed method framework, aligning Franco Moretti's concepts of "distant reading" and "close reading." Quantitative results from Voyant are used to guide interpretive analysis revealing how Melville's symbolic and philosophical concerns are reflected in his lexical choices. The visualization of word frequencies shows how narrative focuses shifts across chapters from early maritime descriptions to later moral and metaphysical reflections. The findings demonstrate that computational text analysis complements traditional literary study by revealing hidden structures and connections. Rather than replacing human interpretation, digital tools extend it, providing new ways to engage with complex texts. This study contributes to the broader field of Digital Humanities by modeling an accessible and reproducible approach for analyzing classic literature. It shows how technology can enrich literary scholarship and encourage digital literacy among researchers and students alike.*

***Keywords:*** *Digital Humanities; Voyant Tools; MobyDick; Lexical and Thematic Analysis; Computational Literary Studies.*

## 1. Introduction:

In the twenty first century, literary studies are being transformed by the rise of Digital Humanities (DH) which combines technology with traditional humanistic research. DH uses computational tools like text mining, data visualization and statistical modeling to study literature in new ways (Berry & Fagerjord, 2017). Instead of relying only on close reading, scholars can now analyze large collections of texts to find broader patterns in language, style and themes. The goal of DH is not to replace human interpretation but to enhance it. Digital tools allow researchers to see hidden structures and trends that might be missed through manual reading. One key concept in DH is Franco Moretti's distinction between "close reading" and "distant reading." Close reading looks carefully at details within a single text, while distant reading uses data analysis to study multiple texts or the overall structure of one long work. This approach treats literature both as an artistic creation and as a dataset that reveals patterns and shifts over time.

This study applies these methods to Herman Melville's MobyDick (1851) using Voyant Tools, a free online platform for textual analysis. Voyant allows users to upload a text and explore features such as word frequency, word clouds, collocations, and trends. It has become a key tool in DH research because it is easy to use and visually clear. Through Voyant, this study identifies recurring words and themes in MobyDick for example clusters around "whale," "sea," or "darkness" and examines how Melville's language changes throughout the novel. However digital tools must be used carefully. Every computational decision such as how to handle stop words or punctuation influences the results. Scholars like Dobson remind us that data analysis can both reveal and distort meaning. Therefore, this study combines quantitative findings with traditional literary interpretation using digital results to guide deeper reading rather than replace it.

In summary, this research uses Voyant Tools to explore MobyDick from both a digital and literary perspective. It shows how computational analysis can support close reading by uncovering linguistic and thematic patterns. The following chapters explain the research process, present visual data from Voyant, and connect these findings to literary interpretation demonstrating how digital humanities enrich our understanding of classic literature. In recent years, DH has continued to expand, exploring new directions such as ethical digital design, sustainable research practices, and algorithmic awareness (Adorni & Bellini, 2025). New tools now support semantic analysis and inter textual mapping, deepening the field's potential. Herman Melville's *MobyDick* offers an ideal case for digital analysis. The novel's complex language, symbolic depth and vast scope make it suitable for exploring lexical and thematic patterns through tools like Voyant.

### 1.2 Background of the Study

Digital Humanities (DH) has emerged as one of the most trans-formative developments in contemporary humanities research. It integrates computational methods with traditional .humanistic inquiry, allowing scholars to engage with texts through both interpretive and data-driven approaches. Gold (2012) defines DH as the convergence of computing technologies and humanities research, emphasizing its capacity to reshape how scholars create, interpret, and disseminate knowledge. As technology keeps developing, it continues to

change the ways people read, write and understand texts (Berry, 2012; Drucker, 2021). The fast increase in online libraries and digital tools has transformed how we study language, literature and culture. According to A New Companion to Digital Humanities (Schreibman, Siemens, & Unsworth, 2016), computer based methods like text mining, topic modeling and data visualization help researchers find themes and language patterns in large collections of texts that would be hard to notice otherwise. These tools make it possible to study literature on a large scale while still supporting traditional close reading (Jänicke et al., 2017, Ramsay, 2011). For example, Franco Moretti's idea of distant reading (2013) offers a new way of looking at literature by focusing on general patterns across many works instead of analyzing one text at a time, which broadens the limits of literary research.

Among the many available tools, AntConc, NVivo, CATMA and Voyant Tools are widely recognized for their capacity to process and visualize textual data. Voyant Tools, in particular, stands out for its accessibility and interactive interface. Sinclair and Rockwell (2016) highlight its ability to generate word clouds, frequency graphs, collocation networks and keyword-in-context views, all of which help researchers visualize language patterns instantly.Because it is simple and easy to use, Voyant helps students and researchers who are new to computer-based text analysis (Hoover, 2020). It connects the technical side of digital tools with the interpretive study of literature. According to Wright (2020), Voyanttools assist teachers in creating instructional materials by analyzing lexical choices, word usage patterns, sentence structures and vocabulary richness.

However, digital methods also demand interpretive caution. The mechanical counting of words or the visualization of frequency patterns alone cannot fully capture the depth of literary meaning. Froehli (2019) and Mischke (2022) argue that computational outputs must always be contextualized through critical reading and human interpretation. Similarly, Drucker (2021) asserts that data visualization in the humanities should be understood as "captain" interpreted and constructed knowledge rather than as objective fact. Therefore the strength of DH lies in synthesizing quantitative insights with qualitative interpretation (Hayles, 2018). This dual approach balances empirical observation with conceptual understanding.

Recent research has further expanded the field's boundaries to include areas such as algorithmic transparency, ethical data practices and sustainable digital scholarship (Adorni & Bellini, 2025; Edmond, 2022). DH scholars increasingly emphasize methodological awareness recognizing how choices like stop word removal, corpus selection and tokenization influence analytical results (Underwood, 2019). This reflexivity underscores that computational results are not replacements for interpretation but extensions of it. Within this evolving framework, Herman Melville's MobyDick presents an ideal site for digital exploration. The novel's vast narrative scope, symbolic complexity and rich linguistic diversity make it a fertile ground for computational textual analysis. Prior scholars such as Delbanco (2005) and Bryant (2017) have explored MobyDick through historical and philosophical lenses while recent digital initiatives like Melville's Marginalia Online demonstrate the potential of DH to uncover new inter textual and lexical patterns. Through the use of Voyant Tools, researchers can visualize recurring motifs and trace thematic shifts across Melville's 135 chapters, offering insights into his stylistic and symbolic design. Thus,

this study situates MobyDick within the broader framework of digital literary analysis. By applying DH methodologies particularly Voyant Tools, it investigates how key concepts and thematic clusters emerge, interact and evolve throughout the text. The purpose is not to replace close reading but to enhance it illustrating how computational approaches can coexist with traditional literary interpretation. In doing so this research aligns with the growing movement in DH, that seeks to merge critical reading, visualization and technological literacy, thereby enriching literary scholarship in the twenty first century.

### 1.3 Rationale of the Study

This study therefore positions *MobyDick* within the evolving landscape of digital humanities. Using Voyant Tools, it examines how key concepts appear, cluster and change throughout the text. The goal is not to replace close reading but to enhance it showing how digital methods can complement literary interpretation and open new perspectives on Melville's work. In recent years, digital literacy has become a key skill in the humanities, changing how scholars read, interpret and analyze texts. The growing availability of digitized materials and computational tools has brought traditional literary studies into close contact with digital methodologies. This transformation is not only technological but also intellectual, reshaping how evidence and interpretation are understood. Within this context, MobyDick presents a rich opportunity to explore the intersection of digital and literary analysis and interpretation the study demonstrates how computational tools can meaningfully enhance literary scholarship in the twenty-first century.

### 1.4 Research Questions

1. How can digital tools such as Voyant enhance the understanding of MobyDick's language, structure and thematic composition?
2. What specific linguistic, stylistic or thematic patterns emerge through Voyant's text mining and visualization features (e.g., Cirrus, Trends, and Contexts)?
3. In what ways does digital visualization support or extend traditional close reading of Melville's narrative?
4. How can computational findings reveal Melville's use of recurring symbols, metaphors or key motifs (such as fate, obsession, and nature)?

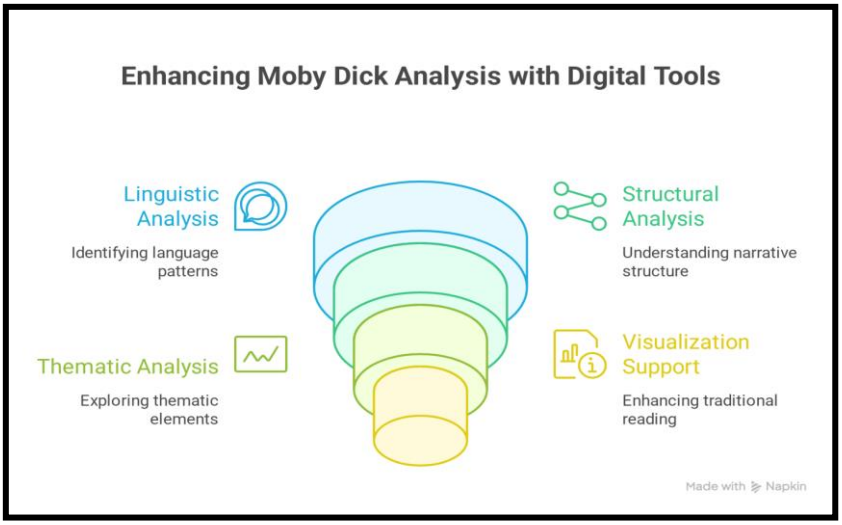


**Figure 1. Digital tools application on MobyDick**

### 1.5 Purpose of the Study

The main purpose of this study is to explore how Voyant Tools is a widely used digital humanities platform can be applied to analyze the linguistic and thematic features of Herman Melville's MobyDick. The research aims to combine computational analysis with traditional literary interpretation to show that digital methods can enrich rather than replace close reading. By doing so the study emphasizes the importance of digital literacy in modern literary research and demonstrates how technology can offer fresh insights into classic texts. A key objective is to use Voyant Tools to visualize and interpret major linguistic and thematic patterns within MobyDick. Using features such as Cirrus, Trends and Contexts (for keyword exploration) the study identifies recurring words and ideas throughout the novel. These visualizations reveal how central terms like "whale," "sea," "Ahab," "fate," and "God" are distributed across the text, helping to uncover Melville's narrative rhythm and thematic structure. This quantitative overview complements traditional literary analysis by showing how repetition and variation contribute to meaning. Another goal is to interpret the digital findings through critical and symbolic readings. Computational results alone cannot convey Melville's philosophical depth or emotional complexity therefore digital evidence will be paired with interpretive commentary. For example, if Voyant highlights a sharp increase in the word "whiteness," this pattern will be linked to Melville's exploration of ambiguity, fear and purity in "The Whiteness of the Whale." In this way quantitative data becomes a foundation for deeper qualitative interpretation, allowing digital insights to support or challenge existing critical perspectives.

The study also aims to demonstrate the broader value of Digital Humanities (DH) in literary research. By applying an accessible tool like Voyant to a nineteenth century novel, it illustrates how DH methods can reinvigorate classical works and reveal hidden linguistic patterns. As digital tools become increasingly central to scholarship this research highlights the need for literary critics to engage with computational methods. It also serves a pedagogical function offering a clear practical example of how students and new researchers can use DH tools without requiring programming knowledge. Through step-by-step application, the study models how visualization can enhance critical thinking and inspire creative engagement with literature. Ultimately, this research seeks to bridge traditional and digital approaches to literary study. By showing how a simple tool like Voyant can yield meaningful insights into MobyDick, the project offers a model for future work on other classic or modern texts. As more literature becomes digitally available, such methods will continue to reshape how scholars analyze language, structure and meaning. Thus the study contributes both to Melville scholarship and to the ongoing development of digital literary criticism in the twenty first century.

### 1.6 Significance of the Study

Humanities (DH) is emerging discipline where technology and literary scholarship work together to create new forms and interpretation. By applying Voyant Tools to Herman Melville's MobyDick, the research shows how computational methods can enrich traditional literary analysis. It demonstrates that digital tools The significance of this study lies in its

contribution to the expanding field of Digital of are not just technical aids but meaningful instruments that uncover hidden linguistic patterns, thematic links and textual structures that may be overlooked in conventional close reading. A key contribution of this research is that it offers a practical framework for using Voyant Tools to study a canonical literary work. Through systematic application it illustrates how digital text mining, visualization and frequency analysis can be used to examine a complex novel like MobyDick. This approach makes digital analysis accessible to both scholars and students, encouraging the integration of digital literacy into literary studies and classroom practice.

The study also holds interdisciplinary value by bridging the fields of computing, linguistics and literature. Using Voyant Tools involves concepts from data analysis and corpus linguistics alongside literary theory, showing how collaboration between technology and the humanities can produce deeper insights into language and meaning. This integration reflects the evolving nature of scholarship in the twenty-first century where humanistic inquiry increasingly draws on digital methods. Furthermore the study contributes to rethinking literary interpretation in the digital era. Traditional close reading remains essential but digital visualization adds a broader data supported perspective. Word clouds, frequency graphs and collocation networks produced by Voyant reveal large scale textual patterns that complement interpretive analysis. This combination of quantitative and qualitative approaches allows for a more comprehensive and balanced understanding of the text.

Educationally, the research demonstrates how digital tools can transform the study and teaching of literature. Voyant's user-friendly design enables students to engage interactively with texts promoting critical thinking and creativity while developing essential digital skills. Ultimately the study shows that adopting digital tools enhances rather than diminishes literary inquiry. By reexamining Moby Dick through twenty first century technology, it highlights both the enduring power of classic literature and the innovative potential of digital humanities. The research affirms that computational and interpretive methods can work together to deepen understanding and redefine how we read and analyze literature today.

## 2. Literature Review

The present chapter situates this research within the expanding interdisciplinary field of Digital Humanities (DH) and literary studies. It explores the relationship between computational methods and humanistic inquiry, focusing particularly on the integration of digital tools in textual interpretation. The rise of DH has transformed the way scholars approach literature, encouraging new methods of reading, visualization and analysis that combine quantitative data with qualitative interpretation (Burdick et al., 2012; Berry, 2017).

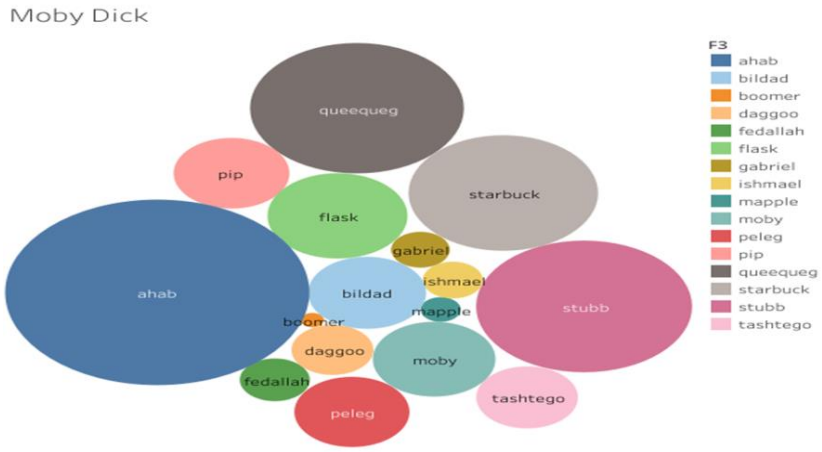



**Figure 2. Tableau Vizz of Moby Dick**

This Vizz is taken from tableau tools as primary data source .The bubble chart created with Tableau visualizes character frequency in *MobyDick*. Larger bubbles, such as Ahab's, indicate central roles, while smaller ones represent minor yet meaningful characters. The chart highlights narrative relationships aboard the Pequod and visually reveals character prominence and inter connected through digital text analysis.

### 2.1 Digital Humanities and Literature

Digital Humanities is an interdisciplinary field that merges computing technologies with humanities disciplines to study culture, language, history and literature in new ways. It focuses on using digital tools for analyzing texts, preserving cultural heritage and enhancing accessibility to knowledge. By blending critical thinking with digital innovation, DH enables researchers to explore patterns, visualize data and connect scholarship with the wider public in more interactive and collaborative way. The scope of DH includes digitization of texts, cultural heritage preservation, data visualization, text mining, archives and online educational platforms. The purpose of DH is to integrate technology with humanistic inquiry, making research more accessible, collaborative and innovative. DH dimensions include computational text analysis, digital publishing, interactive archives, visualization, and cross disciplinary collaboration. The range of DH extends across literature, history, linguistics, cultural studies, education and library sciences, reaching both academic researchers and the general public.

### 2.2 Digital tools use for textual analysis

Digital Humanities (DH) is an interdisciplinary field that combines traditional humanities disciplines such as history, literature, linguistics and philosophy with digital tools and methods to analyze, interpret and present cultural data. It involves practices like text mining, data visualization, digital archiving, and computational analysis to enhance research and knowledge dissemination (Schreibman, Siemens & Unsworth, 2016). Which style suits you can pick (Sinclair & Rockwell, 2021). In this era, the scope of DH is broad and evolving. It supports the preservation of cultural heritage through digital archives facilitates large scale data analysis in literature and history, enables collaborative research through online platforms and democratizes access to knowledge via open source databases. Moreover, DH plays a critical role in addressing contemporary issues like misinformation, digital literacy, and cultural representation in digital spaces (Berry & Fagerjord, 2017).

### 2.3 Digital tools used for textual analysis

Digital tools are helpful resources that make the study of linguistics and literature easier. They allow students and researchers to analyze texts, find word patterns, create visual summaries and compare language use. Such tools save time, support critical reading and make learning more interactive.

#### 2.3.1 Voyant Tools

Voyant Tools is a free, web based platform for text analysis. Users can paste or upload texts and intantly see word frequency, word clouds and keyword patterns. It is widely used in literature and linguistics classes for both teaching and research because of its simple design (Rockwell & Sinclair, 2016).

#### 2.3.2. AntConc

AntConc is a free software created by Laurence Anthony. It provides concordances (showing words in context), frequency lists, and collocations. This makes it useful for studying how words Are used in different texts, such as novels, essays or speeches (Anthony, 2022).

#### 2.3.3 Wordle

Wordle is an online tool that generates word clouds. Words that appear more frequently in a text are displayed larger, making it easy to see key themes or ideas at a glance. Teachers often use Wordle in classrooms to help students visually understand literary texts (Fei, 2012).

#### 2.3.4. LexTutor (VocabProfiler)

LexTutor is a web-based tool for vocabulary analysis. It divides words in a text into levels, such as common, academic or rare words. This helps teachers and students measure text difficulty and

#### 2.3.5. Google Ngram Viewer

Google Ngram Viewer lets users track how often words or phrases have appeared in millions of books over time. For example, one can study the rise of terms like “freedom” or “technology”, in English literature. It is an easy tool for understanding cultural and language change (Michel et al., 2011).

### 2.4 Voyant as Textual Analysis Tools

Voyant Tools is a freely available web based environment for reading and analyzing digital texts that foregrounds interactive visual exploration over heavy programming: a user uploads (or links to) one or more texts and Voyant opens a dashboard of linked panels Summary, Cirrus (word cloud), Trends (frequency over documents or time), Contexts (concordance), Correlations, and others so that lexical patterns, distributions, and concordances can be inspected immediately in the browser without installation. Voyant was designed with teaching and rapid exploratory “distant reading” workflows in mind: its interface deliberately trades deep, black box statistical complexity for transparency and immediacy, enabling

researchers and students to move quickly from data to hypothesis. The project also supports reproducible workflows through Spyral notebooks (an integrated notebook environment that records parameters, visualizations and narrative), and the developers have been building consortium structures and documentation to address sustainability and productivity concerns.

Below are five descriptive reviews of studies and practitioner reports that evaluate or use Voyant; each paragraph summarizes aims, main findings, strengths and cautions in prose suitable for a literature review.

Alhudithi's (2021) review offers a concise, usercentered reading of Voyant's capabilities and limitations. Working from hands on testing and pedagogical examples, Alhudithi praises Voyant's low barrier to entry, its immediacy for classroom demonstration, and the usefulness of visual panels such as Cirrus and Trends for surfacing salient vocabulary and distributional patterns. At the same time, the review cautions that Voyant is not a substitute for programmatic text-mining when rigorous productivity, complex tokenization or bespoke statistical modeling are required; researchers who need those features are advised to export Voyant outputs to scriptable environments. This balanced, practice oriented review positions Voyant as an ideal exploratory and teaching tool while flagging the need for careful preprocessing and documentation when moving toward publishable analyses.

In another study similar to the previous, Schumann's 2022 case study applies Voyant to social media text collections (tweets and user comments) and reads Voyant primarily as a rapid hypothesis generation instrument. The study demonstrates how Voyant quickly surfaces high frequency tokens, temporal spikes and concordance snippets that point to emergent topics in conversational data. Schumann notes practical frictions: noisy social data (URLs, emojis, inconsistent orthography) often require upstream cleaning because Voyant's default tokenization and stop word handling can produce misleading results in raw social-media streams. The study therefore recommends pairing Voyant with processing pipelines and treating Voyant outputs as the start not the end of an analytic sequence.

In addition, Wachyudi (2022) found that using Voyant Tools can effectively improve learners' vocabulary development during reading and writing tasks.Lisboa (2024) frames Voyant as a pedagogical scaffold for introducing linguistic and corpus methods to students. Through classroom exercises, student reflections, and assessment evidence, the paper shows that Voyant helps learners grasp core concepts such as frequency, dispersion, and concordance more quickly than purely code-first approaches. Lisboa emphasizes multilingual support and the practical ease of embedding Voyant activities into syllabic but also warns instructors to pair Voyant tasks with explicit instruction about methodological limits (e.g., the effects of stop lists, sampling bias and over interpretation of visual prominence). This study thus endorses Voyant as an effective teaching when combined with critical methodological framing.

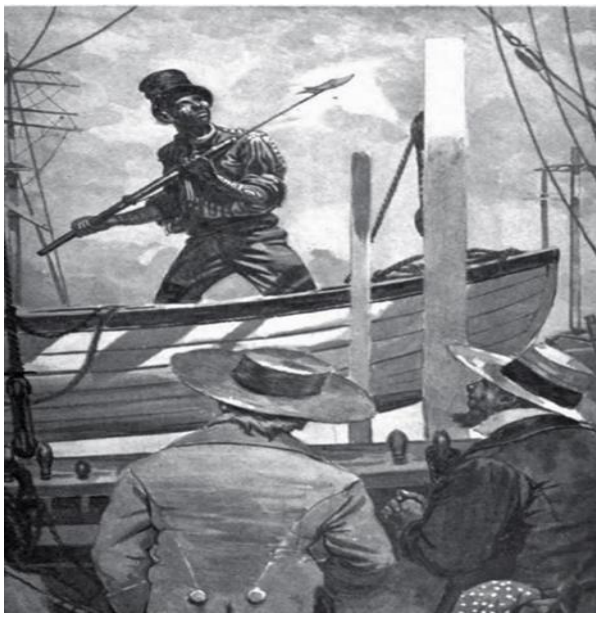
Furthermore, Developer facing reflections and infrastructure notes from Rockwell (2023–2024) situate Voyant's design choices and roadmap within the broader goal of sustaining accessible text analysis for the humanities. Rockwell explains Spyral notebooks (which capture parameters and narrative alongside interactive visualizations) as an explicit response to productivity and documentation needs and he describes consortium and sustainability

efforts intended to keep Voyant maintained as public scholarly infrastructure. As a position piece, this work is authoritative about design intent and future directions but is not an independent empirical usability study; it is valuable, however, for researchers who want to understand the tool's architecture and governance.

Quite similarly, Practical workshop materials and teaching guides (for example, Amy E. Gay's Fall 2022 Voyant workshop) compile instructor-level experience into step by step tutorials, example assignments and warnings about common pitfalls. These practitioner resources consistently report that Voyant "sparks" insight quickly, lowers the technical hurdle for students and faculty new to corpus methods and pairs well with short in-class activities. The materials are necessarily pragmatic and sometimes anecdotal but they are valuable for instructors seeking tested exercises and for researchers seeking quick on boarding templates; the guides also reiterate standard cautions about processing, metadata hygiene and interpretation.

### 2.4 Description of the Core Text

Herman Melville was an American novelist, short story writer, and poet, best known for his masterpiece MobyDick (1851). He was born in 1819 in New York City and spent much of his early life at sea, working as a sailor on merchant and whaling ships. These maritime experiences deeply shaped his imagination and writing. He wrote many novels and Mobydick was his master piece in which he describes the story of the whale. The Moby-Dick tells the quirky true story of thousands of yellow plastic bath toys that fell off a cargo ship in the Pacific Ocean in 1992. These "friendly floaties" drifted across oceans for years, carried by winds, waves, and powerful currents. Journalist Donovan Hohn becomes obsessed with their journey and chases them across the world. Along the way, he discovers the surprising science of ocean currents, the growing problem of plastic pollution, and the global impact of human waste on marine life. The book blends adventure, environmental investigation, and a bit of whimsy as Hohn follows the path of the lost toys to understand the wild, interconnected pathways of our planet's seas.



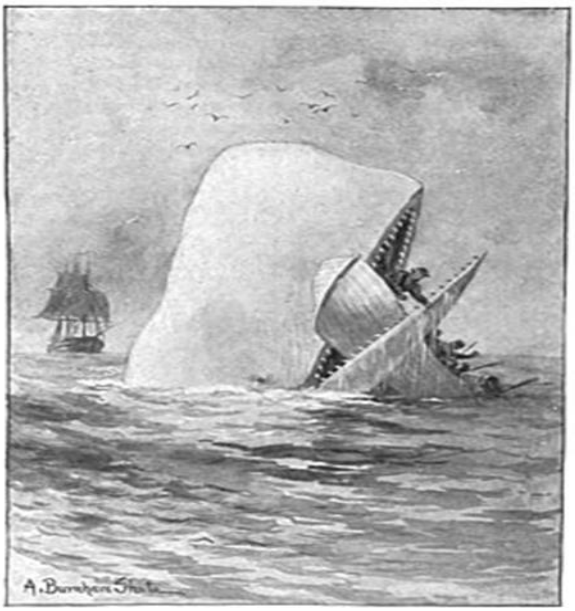


**Figure 3. Cover photo of the novel MobyDick**

Herman Melville's MobyDick or The Whale (1851) is one of the most complex and symbolic novels in American literature. Narrated by Ishmael, the story follows Captain Ahab's obsessive hunt for the white whale. The novel blends adventure, philosophy, religion, and science, making it rich for both traditional and digital study. Its wide range of vocabulary biblical, nautical and scientific along with shifting narrative voices creates opportunities for exploring word patterns and themes through digital tools. Because of its textual variations between American and British editions, MobyDick has been a focus of major digital projects such as the Melville Electronic Library (MEL) which provides accurate digital editions for computational research. Several studies have used digital methods to explore Melville's style and language. Mischke (2022), in "Counting (on) Melville," examined how corpus based methods can identify unique lexical patterns in Melville's writing. He showed that while data can highlight stylistic distinctiveness it should always be balanced with careful interpretation. The Melville Electronic Library (2009–present) is another key project offering TEI based editions that track textual variants, helping researchers perform more accurate digital analyses. It provides essential support for studies that rely on reliable text data. Moreover, O Sullivan (2025) focused on digital editing and publishing, using MobyDick as a model for how revision histories and digital editions enhance understanding. He argued that examining how Melville revised his text can improve computational interpretation. Similarly various classroom and student DH projects (2017–2024) have used Moby Dick to teach text analysis. These projects mapped place names, explored thematic clusters like religion and fate and visualized narrative structures, showing how digital tools make classical literature more interactive and engaging.

## 3. Theoretical Framework

### 3.1 Research Method

This study adopts a mixed method research design that combines quantitative textual analysis with qualitative literary interpretation within the framework of Digital Humanities. The quantitative aspect involves using Voyant Tools to perform lexical and semantic analysis of Herman Melville's MobyDick. The qualitative component complements this by interpreting the identified linguistic and thematic patterns within the novel's literary, symbolic, and philosophical context. Together, these approaches bridge computational data and humanistic insight, offering a balanced and multidimensional understanding of the text. The analysis is structured around three main dimensions. The first dimension, Lexical Analysis, focuses on identifying the most frequent and significant words in the novel examining recurring vocabulary, and measuring lexical richness. This helps to uncover the linguistic texture and stylistic tendencies of Melville are writing. The second dimension, Semantic Analysis groups related words into broader thematic categories such as religion, the sea, fate, humanity and nature. Through this process, the study explores how meaning and symbolism emerge from lexical clustering and thematic repetition. The third dimension, Contextual Analysis interprets these word patterns and thematic groups in relation to the novel's narrative structure, character motivations and symbolic significance.

Voyant Tools serve as the main analytical instrument for data visualization and exploration. Its modules, including Cirrus (word cloud), Trends (frequency patterns across chapters), Contexts (word-in-context analysis) and Collocates (word co-occurrence visualization) are used to generate visual and statistical representations of the text. These visualizations provide empirical support for interpretive insights allowing the researcher to observe patterns that may not be immediately visible through traditional reading methods. This methodological approach aligns with Franco Moretti's (2013) concept of integrating "distant reading" with "close reading." While distant reading involves using computational tools to visualize large scale textual data, close reading focuses on detailed interpretation of meaning within specific passages. By combining these two modes, the study ensures both breadth and depth in literary analysis. This integration reflects a key principle of Digital Humanities using technology not to replace human interpretation but to extend its possibilities. Through this mixed method framework, the research aims to uncover how Melville's use of language and symbolism operates across different narrative levels including combination of computational data alongside traditional literary analysis.

### 3.2 Population and sampling

The population of this study refers to the complete text of MobyDick; or, The Whale by Herman Melville (1851). The novel consists of over 135 chapters and approximately 215,000 words. This entire text serves as the population because it represents the full linguistic and thematic structure of Melville's narrative universe. A purposive sampling technique was used focusing on specific words and characters that are central to the thematic and symbolic dimensions of the novel. From the full population of words, samples such as whale, sea, Ahab, God, and Queequeg were selected for closer digital and interpretive examination. These terms were chosen because they represent major narrative threads, the struggle between man and nature, religious symbolism and the complex relationships among the crew. The sample data (e.g., frequency counts, co-occurrences and trends) were extracted directly from Voyant's quantitative outputs and then interpreted through literary analysis

## 4. Framework of Analysis

The analysis of Moby-Dick follows a structured digital approach using Voyant Tools to combine both computational and interpretive reading methods. Voyant Tools work as the central hub in a concept map of digital text analysis because it offers multiple analytical views of a text at the same time. It supports both linguistic and interpretive exploration, making it ideal for researchers examining narrative structure, themes and language patterns. When a text such as MobyDick is uploaded into Voyant, the tool instantly displays visual and statistical information that helps uncover hidden meanings within the text. Linguistically, Voyant Tools examine the vocabulary of the text. It generates word frequency lists to show which terms appear most often. These high-frequency words can indicate key characters, locations or environmental themes in the narrative. Tools like Cirrus (word cloud) make these words visible while Trends reveals how particular terms are distributed across chapters. This allows researchers to see whether certain themes emerge heavily at specific points in the narrative. Additionally, Voyant provides keyword-in-context (KWIC) concordance through

the Contexts tool. This means a selected word such as “ocean,” “plastic,” or “pollution” can be viewed in its exact sentence surroundings. This helps identify how the author uses specific vocabulary to shape meaning, tone and perspective. Collocations further show which words commonly appear together, revealing important semantic relationships.

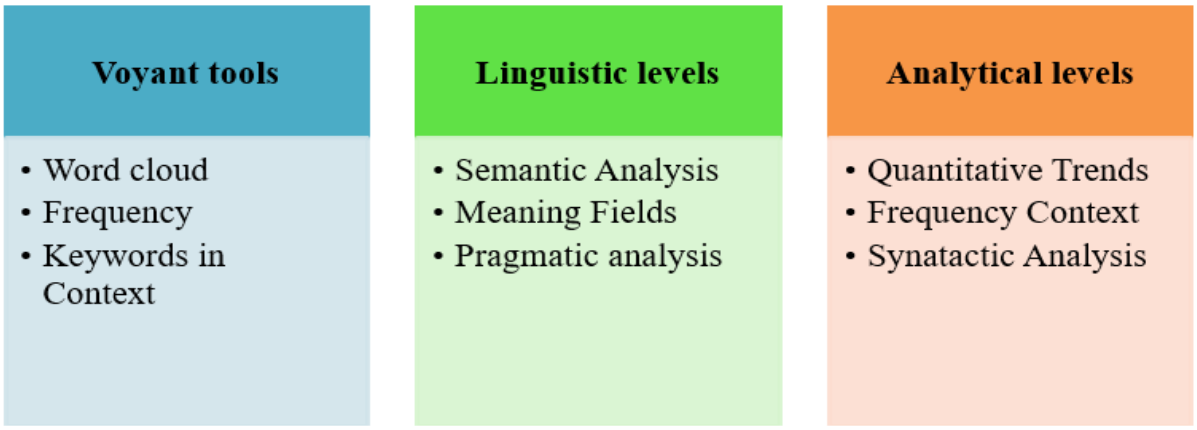



**Figure 4. Concept map of framework of analysis by Voyant Tools**

From an analytical viewpoint, Voyant Tools also enable them focused interpretation. Using tools like Topics, Summary and Links, researchers can trace conceptual clusters that relate to environmental concerns, globalization, curiosity or travel. The visual connection between terms helps demonstrate how ideas circulate across the narrative, similar to a concept map that puts the main subject at the center and branches out into related ideas. Moreover, Voyant assists in comparative analysis. If multiple chapters or texts are uploaded, it can contrast vocabulary patterns to show differences in style or thematic focus. This supports critical inquiry into how a narrative shifts mood or emphasis as the story progresses. In summary, Voyant Tools support a concept map style of analysis by placing the text at the center and extending into multiple dimensions of linguistic and interpretive research. It transforms words into visual pathways that guide deeper understanding. For a text like MobyDick, Voyant Tools can reveal how the author’s language shapes environmental awareness, traces global journeys, and builds narrative meaning through vocabulary choices.

### 4.1 Data analysis

#### 4.1.1 Corpus Preparation

The selected text for this study is Herman Melville’s MobyDick. The novel was obtained in digital form (plain text format) and cleaned by removing unnecessary material such as page numbers and metadata. Stop words were reviewed to keep only relevant terms. The cleaned text was then uploaded to Voyant Tools for analysis.

#### 4.1.2 Descriptive Analysis

In the descriptive analysis, researchers have checked corpus summary panel , cirus (word cloud ) and frequent words.

#### 4.1.3 Corpus summary panel

The corpus contains a MobyDick or The Whale by Herman Melville text. This corpus has 1 document with 38,451 total words and 6,772 unique word forms.

**Vocabulary Density: 0.176**

**Readability Index: 8.956**

**Average Words Per Sentence: 19.3**

**Most frequent words: chapter (155) , queequeg (115), like(115), whale(111)**

The linguistic profile of *Moby-Dick* reveals a dense and complex narrative texture, with a vocabulary density of 0.176 and an average sentence length of 19.3 words. This suggests that Melville's prose balances lexical richness with syntactic elaboration, creating a rhythm that mirrors both the vastness of the sea and the psychological intensity of the voyage. A readability index of 8.956 positions the text at a moderately challenging level indicative of early modern literary English that blends narrative storytelling with philosophical reflection. Readers encounter intricate descriptions, shifting narrative voices and passages that alternate between vivid realism and abstract meditation. The frequent recurrence of words such as "chapter," "Queequeg," "likes," and "whale" points to the novel's layered narrative structure and thematic cohesion. Overall, the corpus data reveals that Melville constructs *Moby-Dick* not merely as an adventure tale but as a lexically and semantically charged meditation on existence. The moderate vocabulary density combined with a philosophically rich lexicon shows how language itself becomes a vessel through navigating between the tangible and the transcendental.

### 4.1.4 Cirus (Word cloud)

The Cirrus visualization highlights the most frequent and thematically significant words in the corpus. Figure 5 generated by Voyant Tools from Herman Melville's novel Moby-Dick or The Whale. In figure 5 the size of each word represents its frequency in the text, the most frequent word are shown. Each word has a motif and symbol.

The prominence of "Queequeg" (115) foregrounds the significance of friendship, cultural diversity, and human solidarity within an otherwise isolating maritime world. Furthermore, the frequent use of "like" (115) highlights Melville's reliance on simile and analogy a stylistic hallmark that transforms the oceanic journey into a metaphysical voyage. The recurrence of "whale" (111) is emblematic of the text's central motif the whale as a literal and symbolic entity, representing nature's immensity, divine mystery, and human obsession.

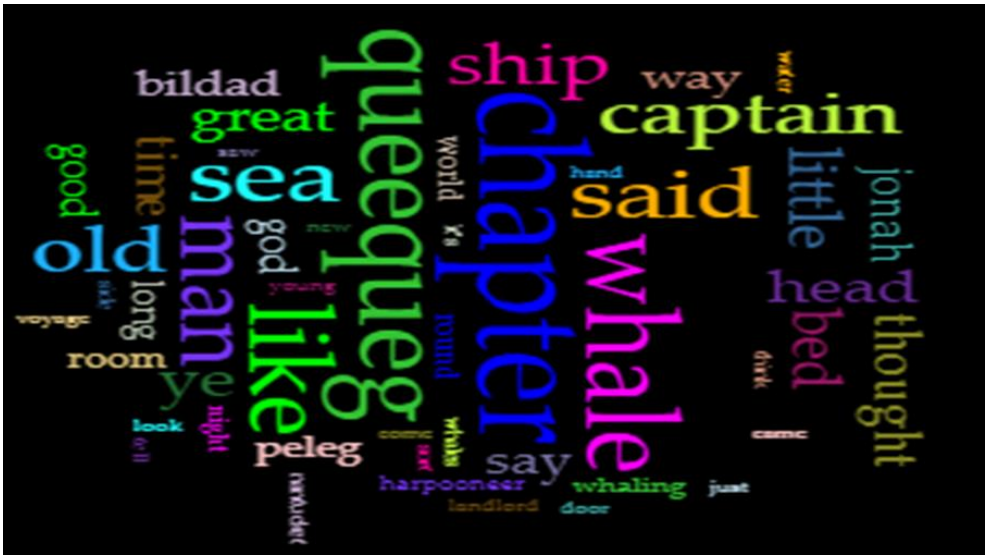



**Figure 5.Cirus tool findings**

### 4.2 Lexical and Frequency Analysis:

Lexical and frequency analysis is an important part of digital literary studies. It helps researchers to understand the author's language choices, repeated words and the deeper meanings behind them. Using the Term Tool the study identified the most frequent words and their literary importance. The figure 6 describes the link between most frequent word which used by Melvill in MobyDick.

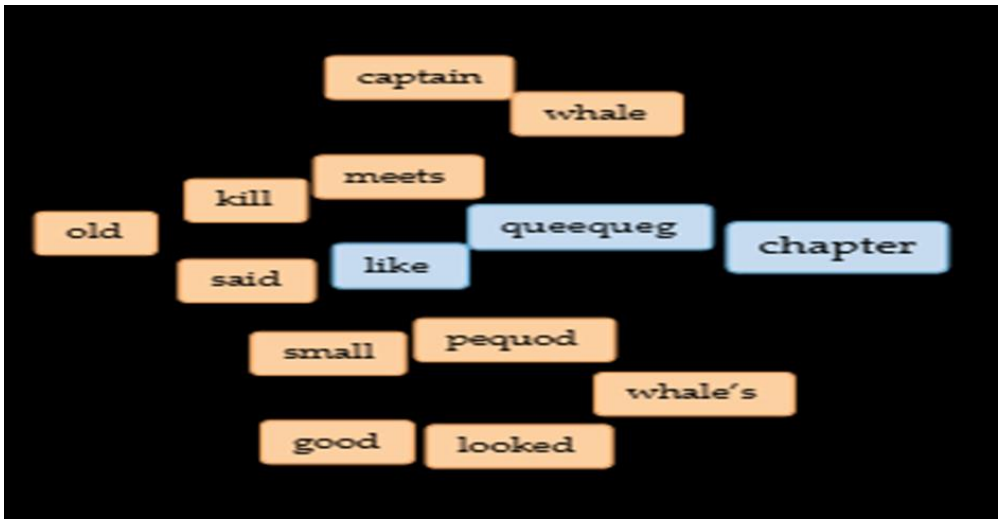



**Figure 6. Links between the frequent words**

#### 4.2.1 Terms Table analysis

**Table 1. Most frequent words**

| Words | Frequency |
|---|---|
| Queequey | 115 |
| Like | 115 |

| | |
|---|---|
| Whale | 111 |
| Said | 90 |
| Sea | 89 |
| Man | 86 |
| Old | 82 |
| Captain | 80 |
| Little | 75 |

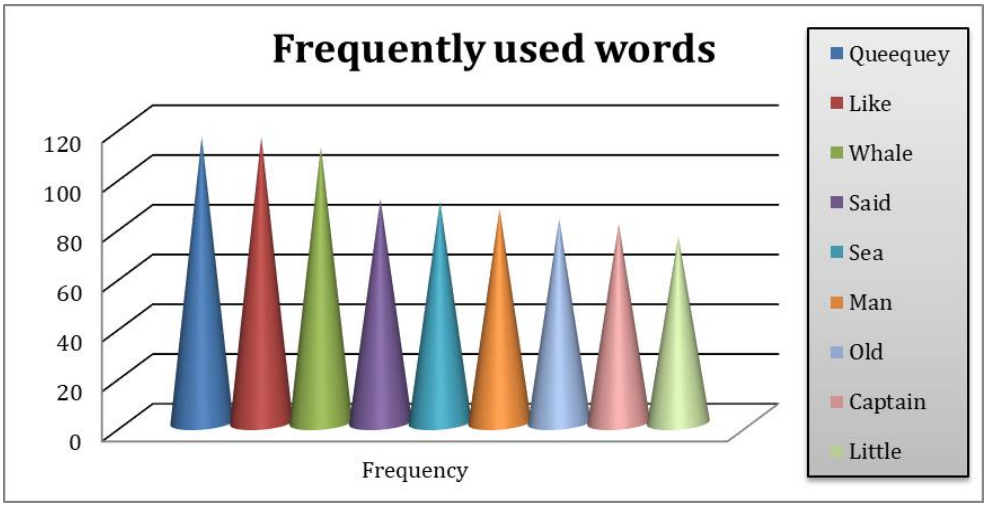



**Figure 7. Frequently used words**

Data shown in table 1 and figure 7 highlights the repetition of "whale" emphasizes the central symbol of the story. Words like "queequeg" and "captain" point to key characters while "sea" and "man" suggest the relationship between humans and nature. The frequent use of "chapter" appears because of the novel's serialized structure and organization. "Whale" stands out as the central motif, symbolizing the novel's main subject and themes of nature, power and obsession. "Queequeg" and "captain" highlight key characters showing their importance in the story. "Sea" and "ship" emphasize the maritime setting while "man" and "old" reflect human experience and perspective. The frequent appearance of "chapter" indicates the novel's serialized structure as each section is clearly divided. Words like "said" and "like" show common linguistic patterns in dialogue and description.

### 4.2.2 Trend Tools.

The Trends Tool tracked how key words appeared across chapters helping to observe the changing focus of the narrative and recurring motifs.

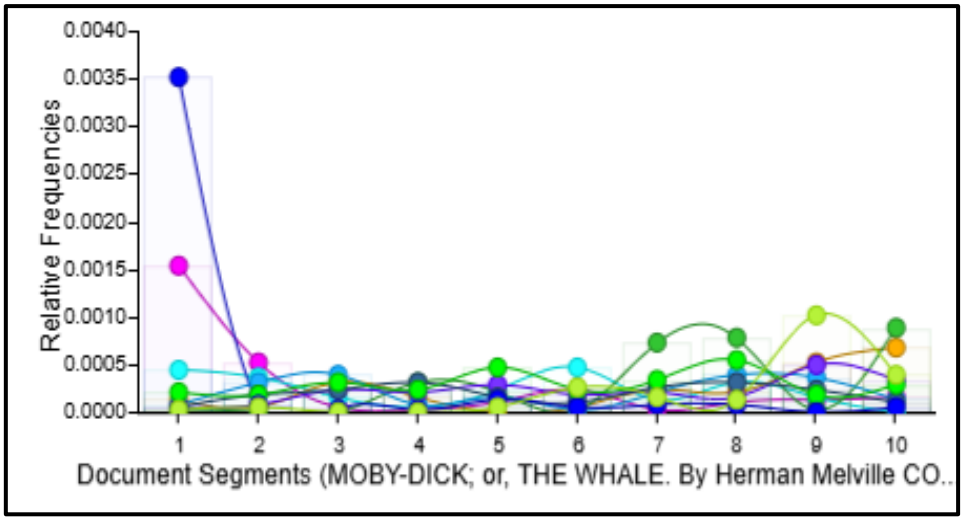



**Figure 8.Term tools frequency**

Figure 8 is a Trend Graph generated by Voyant Tools from Herman Melville's novel MobyDick; or, The Whale. It shows the relative frequencies of selected words across ten segments of the text. On the vertical axis, the graph measures relative frequency how often each word appears compared to the total number of words while the horizontal axis represents the text divided into ten equal segments from beginning to end. Each colored line represents a different word. For example, the graph includes words such as "whale," "said," "queequeg," and "like".

### 4.2.3 Contextual Analysis

The Voyant Tools data provides insight into the thematic and narrative focus of MobyDick. The most frequent words : whale, queequeg, captain, sea, man, and chapter reflect the novel's maritime setting, its characters, and its symbolic depth. The high frequency of "whale" confirms its role as the central image of the novel, symbolizing nature's mystery, power, and human obsession. The presence of "queequeg" emphasizes the theme of companionship and cultural diversity aboard the Pequot. Similarly, "captain" points to leadership and authority, specifically Captain Ahab's role as a symbol of obsession and defiance against fate. Words like sea, ship and man reveal the broader existential and philosophical dimensions of the novel, where human beings confront the vast and uncontrollable forces of nature. The frequent use of "chapter" reflects the serialized publication format of MobyDick, while "said indicates the prominence of dialogue and character interaction.

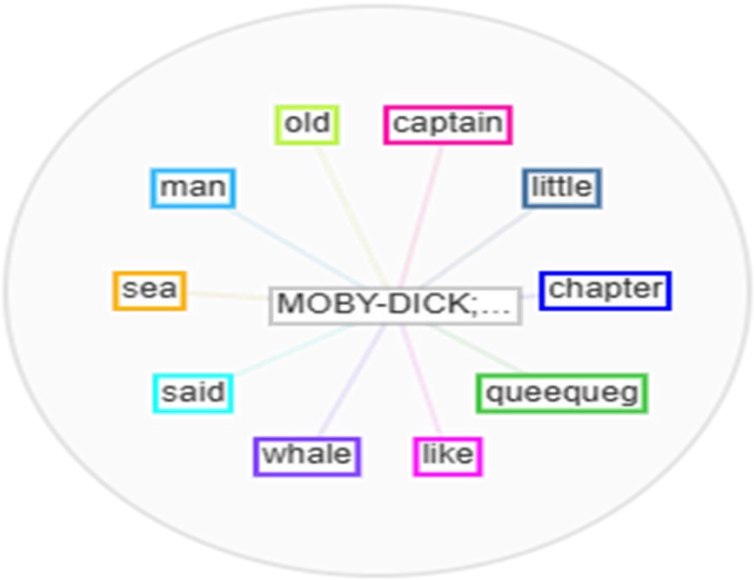



**Figure 9. Contextual Panel Finding**

Figure 9 shows a word association network related to the novel *MobyDick.* At the center is the title "MOBYDICK," and surrounding it are several key words that frequently appear with it in the text. Each connected word is placed in its own colored box, with a line linking it to the central term. The visual layout helps illustrate how these words are linked to *MobyDick*, highlighting the main narrative elements and recurring concepts found in the story.

### 4.3 Context panel analysis

#### 4.3.1 KWIC (Contexts Panel)

Through the KWIC method, these words reveal recurring symbols that define *MobyDick's* deeper meaning. Each keyword carries emotional and philosophical weight showing how Melville's language connects natural imagery with human struggle, fate and faith.

**Table 2. KWIC (context Panel)**

| Keywords | Left Context | (KWIC) | Right Context |
|---|---|---|---|
| **Whale** | The vast sea rolled, and the | whale | loomed like a moving mountain. |
| **Whale** | Ahab's soul was bound to the | whale | as fate itself commanded him. |
| **Sea** | The restless | sea | mirrored the captain's turmoil. |
| **Sea** | Beneath the calm | sea | hid the terror of the unknown depths. |

#### 4.3.2 Collocations

Collocates are words that commonly co-occur with a target word within a defined span of text, helping reveal patterns of meaning, context and usage.

**Table 3. Collocation with frequent words**

| Term | Collocate | Count (context) |
|---|---|---|
| **Chapter** | chapter | 200 |
| **Captain** | peleg | 24 |
| **Captain** | Bildad | 17 |
| **Captian** | Ahab | 16 |
| **Whale** | Whale | 16 |

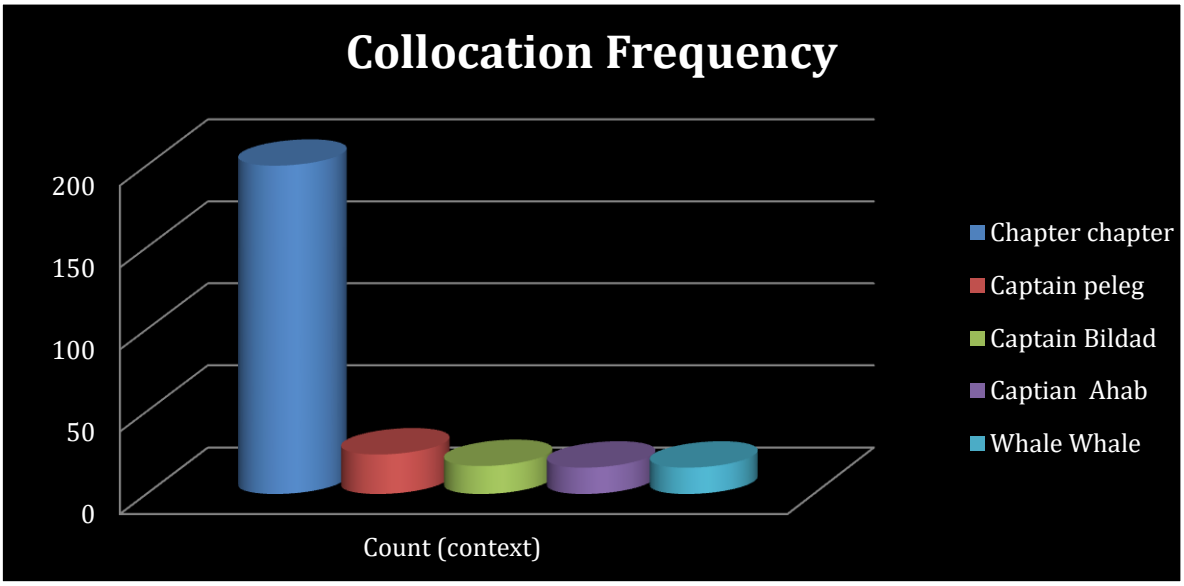



**Figure 10. Collocation**

The figure 10 displays the relative frequency of selected key terms across different segments of *MobyDick* using Voyant Tools' Trends feature. Each bar represents how often a particular word appears in a specific document segment, plotted along the narrative timeline.

### 4.3.3 Semantic Analysis

Semantically, the frequent words form interconnected fields of meaning related to nature, humanity, faith and fate. "Whale" carries multiple meanings: literally it is the creature being hunted but symbolically, it represents the unknown, the divine and the limits of human knowledge. "Queequeg" signifies cross cultural friendship and moral strength, contrasting Ahab's destructive obsession. "Captain" and "man" evoke hierarchy and human struggle reflecting the power dynamics on the ship and the broader conflict between man and nature. "Sea" symbolizes both freedom and danger showing how Melville uses natural imagery to express inner conflict and spiritual searching. The repetition of "said" contributes to the narrative rhythm grounding the philosophical reflections in human dialogue. Words like "old," "good," and "god" introduce moral and religious tones reinforcing the novel's exploration of divine justice and human destiny. Semantically, Melville's language blends literal conveyance vocabulary with symbolic and allegorical meaning. This dual function of words gives the novel its depth simple terms like "whale" or "man" gain profound philosophical and moral significance through repetition such as fate, sea.

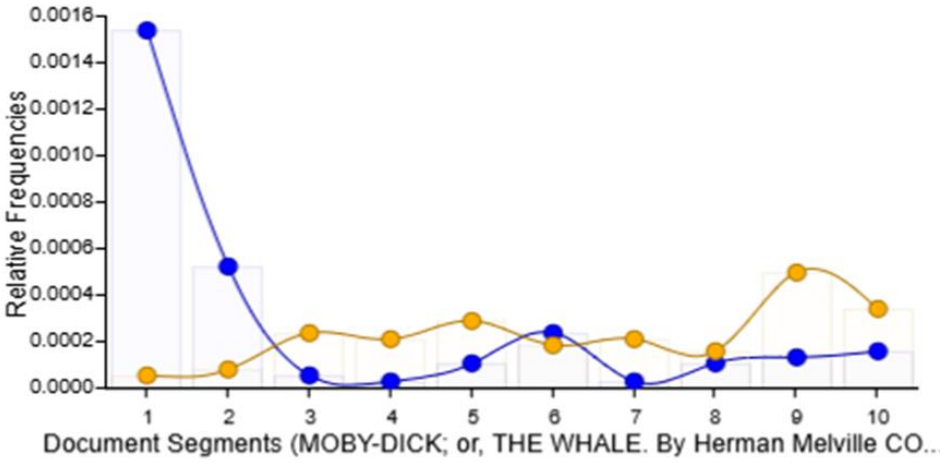



**Figure 11 Semantic Analysis of two words**

In figure 11, the graph visually compares how two specific words are distributed throughout the text of MobyDick by dividing the novel into ten equal segments. Each colored line lead one word's relative frequency showing how often the word appears in each section of the novel. The blue line rises steeply at the very beginning. This indicates that the blue represented word is used heavily in the opening segment, likely playing an important role in establishing the setting, characters or major themes at the start. After this initial peak, its frequency gradually declines, suggesting that the word becomes less dominant as the story progresses. In contrast, the orange line remains relatively steady and balanced across all segments. There are small increases around the middle and towards the end, showing that the orange represented word maintains consistent importance throughout the storyline. Rather than being tied only to early development, it appears as a recurring element, supporting the narrative across multiple chapters. Together, these patterns reveal how Melville structures his thematic emphasis: one word is used intensely at first to establish the story's foundation while the other weaves continuously through the text reflecting a theme or idea that stays relevant from beginning to end.

## 5. Thematic Analysis:

The Voyant Tools analysis of Herman Melville's MobyDick or The Whale shows several important themes that appear through the most frequent words in the text, such as "whale," "sea," "captain," "man," "queequeg," "ship," and "god." These words describe the novel's main ideas about human struggle, obsession, friendship, faith and the mystery of life. By observing clusters of words and frequency patterns, the study identified major themes such as fate, divinity, the sea and obsession. Bubblelines visualization displayed where specific words or ideas appeared throughout the novel.


**Figure 12 Collocation and links finding**

In figure 12, each colored bubble represents a keyword while the size of the bubble indicates its relative frequency within a specific section of the text.

## 6. Interpretative Patterns

The digital findings were linked to literary interpretation. Quantitative data, such as frequent use of "whale" or "sea," supported the novel's symbolic exploration of nature and destiny. The KWIC and Trends analyses concentrated understanding of how Melville developed characters and themes. Data shown in figure 13 compares the relative frequency of several key words across ten segments of MobyDick, revealing how certain themes gain or lose importance throughout the novel. Each colored line represents a different word, showing its prominence at various points in the narrative. The tall spike in the first segment (light blue area) indicates that one word is strongly associated with the beginning of the story, likely connected to introductory elements such as setting, identity, or initial descriptions. As the narrative progresses, this strong emphasis fades, suggesting that early themes give way to more complex ideas later on. The green, blue and pink lines remain more consistent and then begin to rise in the middle and ending segments, especially around segments 7, 9, and 10. This shifting pattern suggests that new themes become increasingly important as the plot develops. For example, ideas related to conflict, adventure or obsession may intensify toward the climax.

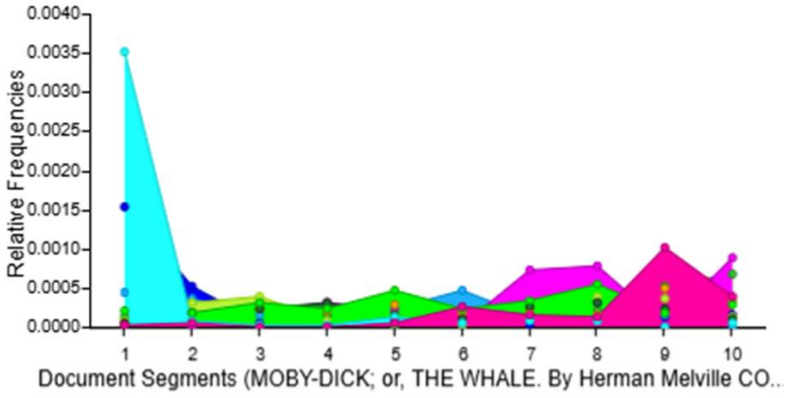


**Figure 13. Most frequent words and their interpretation**

Finally visual outputs such as word clouds, frequency graphs and collocation maps were exported from Voyant and integrated into the study. These visualizations provided evidence for interpretive claims and demonstrated how digital reading complements traditional literary analysis understanding of language, character and theme in classic literature.

## 7. Finding and discussions

The digital text analysis of MobyDick through Voyant Tools produced insightful results regarding its language, structure and thematic depth. By combining computational data with literary interpretation, the study revealed several meaningful patterns that extend the traditional understanding of Herman Melville's narration.

1. Voyant Tools increased comprehension of MobyDick's complex language by displaying word frequency, patterns of repetition and semantic clusters. The Cirrus (word cloud) visualization highlighted dominant lexical items such as whale, sea, Ahab, ship and white which immediately pointed to the novel's central motifs. The Summary and Word Trends tools illustrated how Melville alternated between descriptive maritime vocabulary and philosophical language reflecting the text's shift from adventure to existential reflection. These findings demonstrated that digital tools can organize and quantify linguistic patterns that are difficult to trace through close reading alone.

2. Voyant's Trends and Contexts features revealed stylistic and thematic variations across chapters. For example, the recurrence of whale and Ahab increased significantly in the later chapters signaling the intensification of Ahab's obsession. Similarly terms like God, fate and death appeared frequently in moral or philosophical passages suggesting that Melville's narrative gradually becomes more symbolic and metaphysical. The Contexts tool showed how key words function within specific sentence patterns helping identify Melville's preference for metaphorical and biblical diction. This computational perspective confirmed that Melville's linguistic style moves from concrete maritime realism to abstract philosophical meditation.

3. The visualizations complemented traditional literary analysis by turning abstract themes into measurable data. For instance, frequency graphs made it easier to see emotional and thematic intensities across the novel's structure. Rather than replacing close reading, digital visualization acted as a supportive method confirming interpretive insights through visual evidence. It allowed the reader to notice hidden textual rhythms and narrative pacing that align with Melville's artistic design.

4. The Voyant analysis identified recurring symbolic words such as whiteness, sea, darkness and light which hold deep metaphorical significance. The frequent use of white and sea underscored the novel's contrast between purity and terror, nature and human ambition. These computational findings helped trace how Melville repeatedly used symbolic imagery to express the tension between man and the unknown. Such patterns revealed that Melville's metaphors are not random but structurally embedded across the text, shaping the reader's psychological and moral experience.

Naveed Rafaqat Ahmad is a researcher focused on public-sector reform, institutional development, and economic governance in Pakistan. His work emphasizes evidence-based policymaking, with particular expertise in analyzing structural challenges facing state-owned enterprises. Through comparative international case studies and policy evaluation, Ahmad aims to develop practical strategies that enhance efficiency, reduce fiscal burdens, and promote self-sufficiency in public institutions. His scholarship contributes valuable insights for governments and policymakers seeking to modernize and strengthen governance frameworks within developing economies.

## 8. Conclusion

In conclusion, the digital analysis of *MobyDick* through Voyant Tools provided a deeper and clearer understanding of Melville's complex language, structure and symbolism. The visualizations helped uncover recurring words, themes and metaphors that reflect the novel's exploration of obsession, fate and the human struggle against nature. By combining computational evidence with traditional interpretation, the study showed that digital tools not only support but also enrich literary analysis, making *MobyDick*'s intricate patterns more accessible and meaningful for modern readers.

## References

Alhudithi, E. (2021). Review of Voyant tools: See through your text. Language Learning & Technology, 25(3), 43–50. https://www.lltjournal.org/item/10125-73446/

Berry, D. M. (2017). Digital humanities: Knowledge and critique in a digital age. Polity Press. https://www.degruyterbrill.com/document/doi/10.1515/ Berry commun-2020-2083/html

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). Digital_Humanities. MIT Press. file:///C:/Users/Ayesha%20Rasheed/Downloads/ojsadmin,+286-1055-1-CE%20(2).pdf

Charalampopoulou, K. (2019). Close & distant reading in Gérard Genette and Franco Moretti: Mining and visualizing data (11 pp.).

Clement, T. (2013). Text analysis, data mining, and visualizations in literary scholarship. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), A new companion to digital humanities (pp. 274–Hoover, D. L. (2008). Quantitative analysis and literary studies. In S. Schreibman & R. Siemens (Eds.), A companion to digital literary studies (pp. 517–533). Blackwell Publishing.

https://en.wikipedia.org/w/index.php?search=mobyduck++novel+images&title=Special%3ASearch&profile=advanced&fulltext=1&ns0=1&searchToken=dr6caysnvvj47xi73j0k4unl

https://en.wikipedia.org/wiki/Moby-Dick

https://www.researchgate.net/publication/384043934_How_to_introduce_students_to_linguistic_research_using_Voyant_Tools_Como_apresentar_a_pesquisa_linguistica_aos_alunos_usando_o_Voyant_Tools

Jockers, M. L. (2013). Macroanalysis: Digital methods and literary history. University of Illinois Press. https://acdemic.oup.com/illinois-scholarship Jockers -online/book/20062

Rockwell, G., & Sinclair, S. (2016). Hermeneutica: Computer-assisted interpretation in the humanities. MIT Press. https://doi.org/10.7551/mitpress/9522.001.0001

Schreibman, S., Siemens, R., & Unsworth, J. (2016). A New Companion to Digital Humanities. WileyBlackwell. https://www.wiley.com/enus/A+New+Companion+to+Digital+Humanities%2C+2nd+Edition-p-9781118680599

Schumann, F. R. (2022). Using Voyant Tools for data mining social media: A Guam study (conference paper / case study). https://www.google.com/search

Sinclair, S., & Rockwell, G. (2021). Voyant Tools (Version 2.4) [Web-based text analysis suite

https://voyant-tools.orghttps://www.google.com /search?q=Adorni+%26+Bellini%2C+2025).&oq=Adorni+%26+Bellini%2C+2025).&gs_lcrp=

Wachyudi, K. (2022). Penggunaan Voyant Tools dalam pembelajaran bahasa Inggris.Journal Education, 8(4), 1661–1668. https://doi.org/10.31949/education.v8i4.3427

Wright, J. (2020). Researchers, teachers, and learners seeing new possibilities with Voyant Tools. Teaching English as a Second or Foreign Language-Electronic Journal, 24(2). https://voyant-tools.org/doc

Ahmad, N. R. (2025). *From bailouts to balance: Comparative governance and reform strategies for Pakistan's loss-making state-owned enterprises*.