MODUL UNIVERSITY
PRIVATE UNIVERSITY VIENNA

# UNVEILING DIMENSIONS OF WISDOM LEADERSHIP THROUGH TEXTUAL ANALYSIS OF ELOQUENT SPEECHES BY INFLUENTIAL FIGURES

---

Master Thesis submitted in fulfillment of the Degree

Master of Business Administration (MBA)

in Entrepreneurship, Leadership and Innovation

Submitted to Dr. Daniel Dan

Mohammadjavad
Khadjehalichaleshtari

62003825

Vienna, 19 June 2023

# Affidavit

I hereby affirm that this Master's Thesis represents my own written work and that I have used no sources and aids other than those indicated. All passages quoted from publications or paraphrased from these sources are properly cited and attributed. The thesis was not submitted in the same or in a substantially similar version, not even partially, to another examination board and was not published elsewhere.

Date Signature

# Abstract

This research expands the understanding of wisdom leadership across diverse professional fields - Academia, Arts and Literature, Business, Entertainment, and Law and Politics - by incorporating an analysis of 32 commencement addresses. By employing a mixed-method approach, this study combines quantitative text mining techniques such as Topic modeling, word frequency analysis, and sentiment analysis, with a qualitative, interpretive perspective to decipher the nuances of wisdom leadership in different professional contexts.

Our examination of the 32 commencement speeches illuminated a universality of certain values and themes across professions, underscoring the pervasive appeal of wisdom concepts. Nonetheless, the emergence of unique terms within each profession hinted at distinct priorities and values, revealing profession-specific perspectives on wisdom leadership. Sentiment analysis portrayed an overall positive emotional valence, albeit with variances across professions, and discovered a common narrative arc within the speeches, suggestive of a shared understanding of wisdom leadership. Topic modeling techniques identified key themes pervading the speeches, with each profession reflecting unique thematic prominences.

While recognizing potential biases and limitations, this research, with its focus on the analysis of 32 commencement addresses, yields rich insights into wisdom leadership across a variety of professions. It suggests that wisdom leadership manifests both universal principles that are applicable across all fields, and profession-specific elements that adapt to different contexts. This study thereby contributes to a nuanced understanding of wisdom leadership, its commonalities, and its variations.

# Acknowledgements

I would like to express my profound gratitude to my supervisor, Professor Daniel Dan. His faith in my capabilities encouraged me to delve into the complex realm of text-mining, despite my initial lack of coding experience. His regular meetings and consistent guidance served as a lighthouse, keeping me motivated and on the right path throughout this endeavor.

To my wife, Masoumeh Abolhasani, who provided unwavering and wholehearted support, I am forever grateful. Her assistance was invaluable and her belief in me was a constant source of encouragement.

My deepest appreciation goes to my parents and siblings, whose continued mental and financial support has always been a cornerstone in my academic pursuits. Their love and encouragement have been invaluable to me during this journey.

I am also fortunate to have had an insightful colleague, Verena Foppa. Her intellectual input and constructive feedback on my drafts greatly contributed to the quality of this thesis.

I would like to extend my appreciation to Modul University. The institution has provided a number of proficient professors who have positively impacted my understanding and perspective in various courses, thereby indirectly contributing to this research.

I consider myself lucky to have such a supportive circle. This work would not have been possible without your contributions.

**Dedication**

*I dedicate this work to my sons, Saleh and Sadra. During the most stressful periods of this work, it was your smiling faces that brought me relief and motivation to continue. Your joy was a constant reminder of what truly matters, and for that, I am eternally grateful.*

# Contents

**AFFIDAVIT** iii

**Abstract** v

**Acknowledgements** vii

**Contents** ix

**List of Tables** xiii

**List of Figures** xv

**List of Abbreviations** xvii

**1 Introduction** 1

1.1 Background and context of the research . . . . . 1

1.2 Research problem and research questions . . . . . 2

1.3 Objectives and scope of the study . . . . . 3

1.4 Significance of the research . . . . . 3

1.5 Thesis structure . . . . . 4

**2 Literature Review** 5

2.1 Introduction . . . . . 5

2.2 The Origin and History of Wisdom . . . . . 7

2.3 The Interdisciplinary Evolution of Wisdom Research: From Philosophy to Psychology . . . . . 7

2.3.1 Philosophy: Tracing Wisdom from Ancient Thinkers to Contemporary Perspectives . . . . . 8

2.3.2 Psychology: The Multifaceted Nature of Wisdom . . . . . 9

2.4 Understanding Wisdom Leadership: Origins, Principles, and Applications 11

2.4.1 Navigating the Paradoxes: The Interplay of Wisdom and Leadership 14
2.5 Wisdom: A Mysterious and Multidimensional Concept . . . . . 14
2.5.1 Limitations of Individual Insight and the Need for Interdisciplinary Collaboration . . . . . 15
2.5.2 The Complexity and Subjectivity of Wisdom . . . . . 15
2.6 Text Analysis . . . . . 18
2.7 Text Analysis vs. Text Mining: Convergence and Divergence . . . . . 18
2.8 Text Mining: A General Overview . . . . . 19
2.8.1 Definition and Key Concepts . . . . . 19
2.8.2 Common Text Mining Tasks and Techniques . . . . . 19
2.9 Preprocessing Techniques . . . . . 21
2.9.1 Tokenization . . . . . 21
2.9.2 Stemming and Lemmatization . . . . . 22
2.9.3 Stopwords Removal . . . . . 23
2.9.4 Text Representation Techniques . . . . . 24
2.10 Topic Modeling . . . . . 25
2.10.1 Latent Dirichlet Allocation (LDA) . . . . . 26
2.10.2 Top2Vec . . . . . 27
2.10.3 STM (Structural Topic Model) . . . . . 28
2.10.4 A Synopsis of Supervised Methods . . . . . 28

**3 Methodology** **31**
3.1 Introduction to Methodology . . . . . 31
3.2 Research Design: Selection of Methodology . . . . . 31
3.3 Research instrument: Data Collection and Selection . . . . . 32
3.3.1 Speech Selection Criteria . . . . . 33
3.3.2 Data Source Evaluation . . . . . 35
3.3.3 Data Preparation . . . . . 36
3.3.4 NLP Techniques for Data Preprocessing . . . . . 37
3.4 Data analysis . . . . . 37
3.4.1 Word Frequency Analysis . . . . . 37
3.4.2 Topic Modeling . . . . . 38
3.4.3 Sentiment Analysis . . . . . 38
3.5 Data Analysis Process . . . . . 39
3.5.1 Software and Tools . . . . . 40
3.6 Ethical Considerations . . . . . 41

3.7 Limitations and Assumptions . . . . . 42

**4 Results and Discussion** **45**

4.1 Introduction . . . . . 45

4.2 Word Frequency Analysis (Unigrams, Bigrams, Trigrams) . . . . . 45

4.2.1 The Most Frequent Words . . . . . 46

4.2.2 The Most Frequent Unique n-grams . . . . . 48

4.2.3 Common unigrams, bigrams and trigrams . . . . . 51

Common unigrams . . . . . 51

Common bigrams . . . . . 52

Common trigrams: . . . . . 54

4.3 Sentiment Analysis . . . . . 56

4.3.1 Profession and Sentiment . . . . . 57

4.3.2 Timing and Sentiment . . . . . 63

4.4 Topic Modeling . . . . . 65

4.4.1 LDA . . . . . 65

Heatmap . . . . . 66

Topics-Profession Network . . . . . 67

4.4.2 STM . . . . . 70

Number of topics selection . . . . . 70

Topics correlation . . . . . 71

Heatmap . . . . . 78

4.4.3 Top2Vec . . . . . 81

Terms Network . . . . . 81

The Most Semantically Relevant Paragraph to Wisdom . . . . . 86

Top2Vec Topics . . . . . 88

Heatmap . . . . . 90

4.5 Conclusion . . . . . 91

**5 Conclusion** **93**

5.1 Summary of Key Findings . . . . . 93

5.1.1 Word Frequency Analysis Findings . . . . . 93

5.1.2 Sentiment Analysis Findings . . . . . 95

5.1.3 LDA Findings . . . . . 96

5.1.4 STM Findings . . . . . 97

5.1.5 Top2Vec and Word2Vec Findings . . . . . 98

5.2 Wisdom Leadership Reinterpreted: The Pursuit of Passion and Achievement 99
5.3 Acknowledging Limitations and Potential Biases . . . . . 100
5.4 Final Thoughts and Future Research . . . . . 102

**Appendix I: Speeches Information** **105**

**Appendix II: Dataframe Making Codes** **111**

**Appendix III: The Complete List of the Packages in R** **117**

**Appendix IV: The extremely negative or positive paragraphs in each profession** **121**

**Appendix V: The Top Documents in STM Topic Modeling** **123**

**Appendix VI: The Top Documents in Top2Vec Topic Modeling** **129**

**Bibliography** **133**

# List of Tables


1 List of Abbreviations . . . . . xvii

2.1 Comparing LDA, Top2Vec, and STM . . . . . 29

3.1 The number of speakers associated with each profession . . . . . 33
3.2 The list of commencement addresses . . . . . 34
3.3 The version details of the R . . . . . 40
3.4 Version details of RStudio . . . . . 40

4.1 Total number of unique n-grams . . . . . 46
4.2 The Window for don_know . . . . . 53
4.3 Trigrams appeared in at least three professions . . . . . 54
4.4 The Window for make_difference_life phrase . . . . . 55
4.5 Top document for each topic . . . . . 69
4.6 Exclusivity and semantic coherence of topics considering number of topics . 70
4.7 Words associated with each topic . . . . . 73

.1 Information Summary of Speeches. . . . . 106
.1 Information Summary of Speeches (continued) . . . . . 107
.2 Pictures of Speakers. . . . . 108
.2 Pictures of Speakers (continued) . . . . . 109
.3 The extremely negative or positive paragraphs in each profession . . . . . 121

# List of Figures


4.1 Most Frequent Unigrams in Corpora . . . . . 47
4.2 Top 10 Frequent unigrams in each Field . . . . . 48
4.3 Top 10 unique unigrams in each field . . . . . 49
4.4 Top 10 unique bigrams in each field . . . . . 49
4.5 Top 10 unique trigrams in each field . . . . . 50
4.6 Top Common Unigrams . . . . . 51
4.7 Top Common Bigrams . . . . . 52
4.8 The distributions of sentiment across each field . . . . . 58
4.9 Emotional spectrum across professions . . . . . 61
4.10 Sentiment of the speeches in each profession, from start of speech to end . 64
4.11 The optimum number of topics determined by `FindTopicsNumber` function. . 66
4.12 Top words that represent each topic. . . . . 67
4.13 Distribution of topics over different fields . . . . . 68
4.14 network of topics and professions . . . . . 68
4.15 Lack of convergence for the model with 5 topics . . . . . 71
4.16 Convergence of the model with 7 topics . . . . . 72
4.17 Correlation between topics . . . . . 72
4.18 Distribution of topics across professions . . . . . 79
4.19 Word network generated using Word2Vec . . . . . 82
4.20 Distribution of topics over different fields . . . . . 90

# List of Abbreviations

Table 1: List of Abbreviations

| Abbreviation | Full Form |
|---|---|
| 3D-WS | Three-Dimensional Wisdom Scale |
| AI | Artificial Intelligence |
| BCE | Before Common Era |
| BERT | Bidirectional Encoder Representations from Transformers |
| BPE | Byte Pair Encoding |
| CSV | Comma-Separated Values |
| DIKW | Data, Information, Knowledge, Wisdom |
| GDR | German Democratic Republic |
| GPS | Global Positioning System |
| HDBSCAN | Hierarchical Density-Based Spatial Clustering of Applications with Noise |
| IoT | Internet of Things |
| KDT | Knowledge Discovery from Text |
| LDA | Latent Dirichlet Allocation |
| LOESS | Locally Estimated Scatterplot Smoothing |
| ML | Machine Learning |
| NA | Not Available |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| OS | Operating System |
| PMW | Polyhedron Model of Wisdom |
| QDA | Qualitative Data Analysis |
| RQ | Research Question |
| SARS | Severe Acute Respiratory Syndrome |
| SAWS | Self-Assessed Wisdom Scale |
| STM | Structural Topic Model |
| TF-IDF | Term Frequency-Inverse Document Frequency |
| VUCA | Volatility, Uncertainty, Complexity, Ambiguity |
| WICS | Wisdom, Intelligence, and Creativity Synthesized |

CHAPTER 1

# INTRODUCTION

## 1.1 Background and context of the research

Long-standing research interests include how companies might thrive in VUCA-type circumstances and what CEOs need to succeed in a world of uncertainty and quickly changing technology. It is suggested that wisdom and wise leadership are essential to obtaining success.

Given its crucial role in achieving organizational, social, and international goals, leadership has been a subject of considerable scholarly attention for decades. Traditional leadership theories, however, frequently fail to address contemporary leadership's complexity in the face of dynamism and rapid change (Schoemaker et al., 2018). As a result, academics are becoming increasingly interested in investigating alternate viewpoints on leadership that consider the problems of the twenty-first century. One such viewpoint is wisdom-based leadership or simply wisdom leadership.

Although wisdom leadership and information/knowledge management are not two fundamentally different ideas, it is crucial to distinguish between them from the beginning. The development of moral character, the ability to use knowledge and experience in challenging real-world circumstances, and personal qualities are the main focuses of wisdom leadership. This strategy necessitates gaining practical wisdom, which includes outstanding insight, judgment, and decision-making that serves the greater good. On the other

hand, information and knowledge management focus on gathering, storing, and sharing data inside an organization to aid decision-making.

This study attempts to further the developing field of wisdom leadership by identifying its dimensions and the features of wise leaders using text analysis and text mining of speeches delivered by successful persons.

## 1.2 Research problem and research questions

By utilizing text-mining techniques on the speeches of great leaders, the current exploratory research intends to examine the possibility of discovering similar characteristics among them to relate these characteristics to aspects of wisdom leadership. Although the topic of wisdom in leadership has been explored in earlier research(Vanharanta & Markopoulos, 2020; Karami et al., 2020; Vanharanta et al., 2021), the ever-evolving nature of the leadership role in our society necessitates continuous investigation and re-evaluation of these attributes. In addition, while wisdom scholars emphasize the importance of practical wisdom, very few have tried to understand wisdom from the perspective of those deeply engaged in practical endeavors. Leaders are often seen as eminently practical individuals who must make challenging decisions in real-world contexts, yet there is little research examining their speeches for manifestations of wisdom. This study aims to fill that gap, exploring how practical wisdom manifests in the words and ideas expressed by renowned figures.

The research questions guiding this study are:

*RQ1. What are the most frequently occurring ngrams in the speeches across different professions, and how do they provide insights into the concept of wisdom leadership?*

*RQ2. How can text mining techniques and sentiment analysis be used to uncover the traits of wise leaders from different professions, and how do these traits align with the dimensions of wisdom leadership?*

*RQ3. What principal themes and topics can be recognized across the speeches of leaders from various professions using topic modeling techniques?*

*RQ4. How can the common elements in speeches, identified through text-mining techniques and varying by profession, be linked with the aspects of*

*wisdom leadership?*

*RQ5. Are there noticeable differences in the themes and common elements of speeches delivered by leaders from diverse professions, and what can these variations reveal about wisdom leadership in different professional contexts?*

A thorough assessment of the relevant literature on wisdom leadership will be conducted to answer these research questions. A selection of speeches delivered by successful great leaders will be submitted to text analysis techniques. The study's findings are supported by empirical data thanks to the emphasis on text analysis, which guarantees a methodical and impartial approach to discovering similar traits. The results of this research may provide a more thorough and nuanced knowledge of wisdom leadership and wise leaders. Moreover, it could shed light on how text-mining tools are used to examine leadership literature. Ultimately, this study aims to further knowledge about wisdom leadership and how it may be developed in contemporary businesses by utilizing the experiences and perceptions of seasoned leaders.

## 1.3 Objectives and scope of the study

The main goal of this exploratory study is to see whether text-mining techniques may be used to identify attributes that well-known leaders have in common and to analyze the relationship between these traits and the dimensions of wise leadership. To achieve this, the study will thoroughly review the literature on wisdom leadership, identify the key themes and dimensions related to this idea, and use text-mining techniques to analyze the transcripts of speeches of people widely regarded as outstanding and highly successful. To be more precise, the analysis will involve locating relevant keywords, unigrams, bigrams, and trigrams, then using statistical and computational methods to assess their frequency, co-occurrence, sentiment analysis, and distribution throughout the texts.

## 1.4 Significance of the research

While wisdom in leadership has been studied before, this study brings a new methodological perspective by using text-mining techniques. This might improve our theoretical understanding of wisdom leadership and contribute to the existing body of knowledge.

By exploring wisdom from the standpoint of individuals who often engaged in intensely practical endeavors, the research could broaden our understanding of practical wisdom,

or provides an opportunity to reassess the importance of previously proposed dimensions of wisdom, but through the lens of influential figures. This might give a more concrete and relevant understanding of wisdom leadership.

## 1.5 Thesis structure

The thesis is organized into five chapters, each with a specific focus. Chapter 1 introduces the research problem and the purpose of the study. It also highlights the significance of the research, particularly because text analysis is used.

Chapter 2 reviews the relevant literature on definitions of wisdom, wisdom leadership models, and text analysis techniques. The purpose of this chapter is to provide a comprehensive understanding of the research area.

Chapter 3 describes the methodology used in this study, including the data collection process, data pre-processing, and text analysis techniques employed. This chapter aims to provide a clear understanding of the research methodology used in the study.

Chapter 4 presents the study's findings, including the identified dimensions of wisdom leadership and the characteristics of wise leaders using qualitative and quantitative techniques.

Chapter 5 provides a conclusion to the research, summarizing the essential findings and their implications for the field of leadership. It also identifies the study's limitations and provides suggestions for future research.

CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

Several recent occurrences and technical developments have highlighted wisdom's value in navigating the VUCA (Volatility, Uncertainty, Complexity, and Ambiguity) (Whiteman, 1998) environment marked by fast change, unanticipated difficulties, and complex interdependencies. The well-being of our world is under threat from significant challenges - September 11 attacks and conflicts in the Middle East pose a severe challenge to global security. Likewise, recurring pandemics (such as SARS) or financial crises such as the one in 2008, undermine various aspects, including socio-economic-political systems. Not only that but instances where nature unleashes its wrath upon us include natural disasters (like the Tohoku earthquake and the subsequent tsunamis), exposing weaknesses in our preparedness for dealing with them. In the meantime, social movements such as Occupy Wall Street or Black Lives Matter, along with political upheavals like Brexit, contribute to increasing societal polarization and discontent.

Rapid technological advancement is another aspect of this VUCA environment. It has altered how we live, work, and communicate. Machine Learning techniques, in particular deep learning, have allowed us to analyze vast amounts of data, leading to advances in various fields (Nguyen et al., 2019). Smartphones and social media platforms have revolutionized communication, while the IoT and cloud computing have initiated new opportunities and challenges in privacy and security (Stergiou et al., 2018). The development

of blockchain technology has the potential to reshape businesses (Y. Chen & Bellavitis, 2020), and the rise of renewable energy and autonomous vehicles are crucial steps toward a more sustainable future (Chehri & Mouftah, 2019).

In this context, the importance of wisdom in navigating the VUCA environment cannot be overstated (Brooks, 2021). In front of complexity and uncertainty, wisdom brings discretion, clarity, and understanding (McKenna, 2013). Leadership has been studied for centuries to identify the key characteristics that make a leader successful. Nonetheless, in recent years, there has been a growing pursuit in the contemporary trends in the field of leadership (Bennis, 2007). Wisdom leadership, an old concept but a relatively new trend (Bachmann et al., 2018), focuses on the idea that successful leaders possess unique qualities beyond traditional leadership attributes such as intelligence and charisma.

Cambridge dictionary of philosophy defined wisdom as "an understanding of the highest principles of things that functions as a guide for living a truly exemplary human life" (Audi, 1995). However, upon delving deeper into wisdom's definitions, its multifaceted complexity becomes apparent. Descriptions of wisdom are most likely based on individuals' unique experiences, including the types of complex problems and best-perceived solutions they have encountered (Staudinger & Glück, 2011). Wisdom involves identifying our knowledge's limits (Taranto, 1989) and the future's unpredictability (Ardelt, 2004b). It encourages humility, curiosity, and a commitment to learning from our mistakes and the experiences of others (Edmondson, 2011). In a world characterized by volatility, uncertainty, complexity, and ambiguity, wisdom is essential for individuals, organizations, and societies to thrive and overcome challenges.

In an effort to contribute to the field of wisdom leadership research, this study seeks to shed light on the dimensions of wisdom leadership and the qualities of wise leaders. To do this, in this chapter, we review the literature on text analysis techniques and their application, while also exploring the literature on wisdom and its relationship with leadership. Through the examination of various text-mining strategies, such as topic modeling and sentiment analysis, and the study of significant articles and books published by leading scholars in the field, we delve into the dimensions of wisdom and its evolution over time. The research underscores that wisdom is neither a fixed nor unique condition. Instead, it is viewed as a multi-tool kit, a collection of behaviors, skills, and principles that enable individuals to deliberate on life's important aspects. Subsequently translating those considerations into choices and actions that simultaneously enhance their own well-being and that of others.

## 2.2 The Origin and History of Wisdom

According to "Online Etymology Dictionary" (n.d.), the Germanic languages frequently utilized the composite term "wisdom" in their vocabulary. The word's initial component, "wis," is a derivative of the adjective "wise" and denotes "learned, sagacious, cunning, sane, prudent, discreet, experienced, and having the power of discerning and judging rightly." The word's second component, "-dom," is an Old English suffix that is used to denote quality, state, or condition.

People have tried to pass on their knowledge to future generations throughout human history through various storytelling techniques, such as tales, poetry, pictographs, and paintings. It demonstrates how wisdom has been cherished and passed down for centuries and is considered a crucial part of human civilization. So, the idea of wisdom is not brand-new and has been around for a while. The Sumerians are believed to have first accepted the idea of wisdom around 2,500 BCE (Mitchell et al., 2017).

Even while wisdom has long been a part of human culture, defining it has proven to be complicated. Wisdom has been an ongoing topic of philosophical debate, yet a precise and concise definition has been elusive. Psychologists have recently started empirically investigating wisdom to pinpoint the psychological traits underpinning wise behavior (Bergsma & Ardelt, 2012). The following section combines psychological and philosophical concepts to offer a multifaceted definition of wisdom.

## 2.3 The Interdisciplinary Evolution of Wisdom Research: From Philosophy to Psychology

Some knowledge management academics have defined wisdom simply in terms of the "Knowledge Hierarchy" or "Knowledge Pyramid," also referred to as the "Data, Information, Knowledge, and Wisdom Hierarchy (Ackoff, 1989; Rowley, 2007)," the DIKW chain. However, the concept of wisdom appears to be more nuanced and expansive than this.

Scholars from various disciplines have investigated and studied wisdom throughout history (McKenna, 2013). These academic fields are neuroscience, sociology, theology, and education, with philosophy and psychology being notable contributors. The philosophical perspective on wisdom examines it as a virtue, knowledge, or way of life. In contrast, psychologists investigate it as expert knowledge (Ardelt, 2004a; Baltes & Kunzmann, 2004), a personality type (Ardelt et al., 2019), or a cognitive and emotional process

(Bluck & Glück, 2005). The expansion in thought due to this multidisciplinary approach increases our grasp on how wisdom plays into social relationships, decision-making (W. S. Brown, 2005), and well-being (Kunzmann & Baltes, 2005).

### 2.3.1 Philosophy: Tracing Wisdom from Ancient Thinkers to Contemporary Perspectives

The word "philosophy" is derived from the combination of two Greek words: "philo" ($\phi\acute{\iota}\lambda o$), meaning "love," and "sophia" ($\sigma o\phi\acute{\iota}\alpha$) , meaning "wisdom" or "knowledge." When combined, "philosophy" can be understood as "the love of wisdom" or "the pursuit of knowledge." Unsurprisingly, it has a long history of being associated with wisdom (Kenny, 2012). Durant (1961) stated that "Science without philosophy, facts without perspective and valuation, cannot save us from havoc and despair. Science gives us knowledge, but only philosophy can give us wisdom." Leading thinkers like Plato, Aristotle, Confucius, and the Stoics had their own conceptions of wisdom and guided how to lead morally upright lives. In his Nicomachean Ethics, Aristotle (Ross et al., 2009) introduced the idea of "practical wisdom, emphasizing the value of moral and intellectual growth. Contrarily, Confucius strongly stressed forging close friendships and leading a morally upright life (Yuan et al., 2023).

For 600–700 years, Western philosophy was dominated by Plato and his students' writing (Sternberg & Jordan, 2005). According to Sternberg and Jordan (2005), Plato, Socrates' disciple, had the same belief that wisdom was theoretical, abstract, and the preserve of a select few. In his work, Nicomachean Ethics (Ross et al., 2009), Aristotle describes the idea of practical wisdom. According to him, traits like loyalty, truthfulness, gentleness, fairness, friendliness, self-control, courage, and generosity must be learned and developed. Aristotle referred to these qualities as "excellences" or "virtues." Still, he thought that "practical wisdom," which is the ability to decide how to behave in particular situations with particular people, is the most crucial virtue. Because none of the other qualities can be effectively used without it, he thought that a lifetime of moral and intellectual growth might lead to wisdom.

In Nicomachean Ethics VI.6, he identified two types of wisdom: metaphysical wisdom and practical wisdom. He wrote that Phronèsis is "a true and reasoned state of capacity to act with regard to the things that are good or bad for man." Hughes (2001) asserts that Aristotle distinguished between intellectual and moral virtues by highlighting two essential intellectual qualities. He refers to "Sophia" as the capacity to think clearly on

scientific topics and “Phronèsis” or “practical wisdom” as the practice of doing so. While the DIKW hierarchy does not exactly translate into Aristotle’s divisions between Sophia and Phronesis, we may argue that Sophia correlates more closely to the knowledge chain and appears more in line with the upper stage of it. The DIKW hierarchy is a simplistic and linear paradigm, but Aristotle’s Phronèsis notion is more complicated and interrelated.

McKenna (2013) described “practical wisdom” as the capacity to behave morally under pressure. Later, Vaccarezza and De Caro (2021) collected fresh interpretations of the centuries-old virtue of Phronèsis, highlighting practical wisdom’s significance in moral philosophy.

Other viewpoints can be derived from Meacham (1983). He discussed the nature of wisdom and knowledge, especially in Socratic thoughts. Socrates believed that genuine wisdom derived from knowing one’s own limitations. His statement, “what I do not know, I don’t think I do,” implies that admitting one’s ignorance is a type of wisdom. Knowledge inevitably creates uncertainty since knowing something means questioning it. In a classical concept of knowledge, where knowing and not knowing are mutually incompatible, this may appear inconsistent. The confluence of knowledge and doubt is not a problem but rather an essential component of wisdom (Meacham, 1983).

Another important school of thought, with its own notion of wisdom, began to flourish in another corner of the planet, not at the same time in ancient history. Confucius, the son of a low government official in China, began to lay the groundwork for a very practical curriculum of public behavior that is still used today as a guide to wisdom (Hall, 2011). Confucius underlined practical wisdom’s value and believed it comes from leading a morally upright life and developing healthy interpersonal interactions (Yuan, 2013). Confucianism defined ethics in terms of five fundamental qualities, or the “five constants.” According to Yuan et al. (2023), it is composed of “benevolence and compassion (ren), righteousness (yi), ritual propriety (li), trustworthiness (xin) and wisdom (zhi).”

### 2.3.2 Psychology: The Multifaceted Nature of Wisdom

Since the publication of Clayton and Birren’s study in 1980, the psychology of wisdom has advanced significantly (McKenna, 2013). The Berlin School, Robert Sternberg’s methodology, and US Positivists are only a few of the distinct “schools” of wisdom study that have since arisen (McKenna, 2013). Each of these schools has contributed significantly to our knowledge of wisdom by providing distinctive viewpoints and models.

Wisdom, according to the Berlin School, is “expert knowledge in the fundamental pragmatics of life that permits exceptional insight, judgment, and advice about complex and uncertain matters” (Baltes & Staudinger, 2000). The Berlin school holds that wisdom is said to be involved with life’s crucial and challenging issues and associates five facets: “rich factual knowledge, rich procedural knowledge, lifespan contextualism, relativism of values, and awareness and management of uncertainty (Pasupathi et al., 2001).”

Renowned wisdom researcher Robert Sternberg created the Balance Theory of Wisdom and strongly emphasizes values as an integral part of wisdom. According to Sternberg and Glück (2022), wisdom involves applying practical intelligence or tacit knowledge mediated by values aimed at realizing a common good through a balance of different interests and perspectives; “a person is wise to the extent that they use their skills and knowledge to (1) achieve a common good, by (2) balancing intrapersonal (their own), interpersonal (others’), and extrapersonal (larger) interests over (3) the long term as well as the short term, through (4) the utilization of positive ethical values, by (5) adapting to, shaping, and selecting environments (Sternberg & Glück, 2022).”

Sternberg and Glück (2022) state three types of practical action. One can adapt to the environment as it is, sometimes recognizing that it is unsatisfactory but that there is nothing better available. Alternatively, one can choose to alter the setting for the betterment of both themselves and others. That also aligns with studies of Law and Staudinger (2016) that “Wisdom is heavily dependent on an understanding of the good life as involving self-transcendence and a concern for the good of others”. Then again, one can decide that the environment in which they live does not promote a widespread benefit and then look for another one. Sometimes situations are unchangeable, and one must choose whether to stay or leave.

Based on these concepts, Karami et al. (2020) carried out a systematic evaluation of 50 papers from the psychology, management, leadership, and education domains to look for areas of agreement among conceptions of wisdom. They put out the Polyhedron Model of Wisdom (PMW), which includes several elements, such as “the adequate use of knowledge, intelligence and creativity, self-regulation, openness and tolerance, altruism and moral maturity, and sound judgment to solve critical problems.”

Igor Grossmann and his associates proposed a single model of wisdom (Grossmann et al., 2020), which incorporates components of earlier models. According to Grossmann et al. (2020), wisdom is “morally-grounded excellence in social-cognitive process-

ing" (Grossmann et al., 2020). While excellence in social-cognitive processing requires considering various contexts, perspectives, short- and long-term effects, thinking reflectively and dialectically, and being aware of limitations and subjectivity of thought, morally grounded wisdom balances self-interest with others, values truth, and cares for humanity. Additionally, Sternberg and Karami (2021) tried to develop a thorough and organized model of the so-called 6Ps: **p**urpose, environmental/situational **p**ressures, **p**roblems requiring wisdom, traits of wise **p**eople, psychological **p**rocesses, and **p**roducts of wisdom (Sternberg & Karami, 2021).

The study of wisdom has evolved over time, with contemporary moral philosophy and psychology engaging in an interdisciplinary dialogue that has brought new insights into the concept of practical wisdom. This dialogue, as Vaccarezza and De Caro (2021) point out, has been fueled by the resurgence of virtue ethics and its recent engagement with psychology. Despite the extensive discussions on the topic, they argue that there is still room for further exploration, primarily due to "the resurgence of virtue ethics within contemporary moral philosophy and its recent dialogue with psychology (Vaccarezza & De Caro, 2021)."

In conclusion, the understanding of wisdom has advanced via the shift from philosophical research to psychological analysis. The interdisciplinary discussion between philosophy, psychology, and other areas has shed light on the value of wisdom in our lives. To help us understand this lasting quality as we seek wisdom in our continuously changing environment, ongoing collaboration across various disciplines is essential.

## 2.4 Understanding Wisdom Leadership: Origins, Principles, and Applications

Numerous scholars have investigated the concept of wisdom leadership and the characteristics of wise leaders, contributing to our understanding in diverse ways. Sternberg and Karami (2021), Riggio et al. (2010), Grossmann and Brienza (2018), Ardelt (2004a), Nonaka and Takeuchi (2019), and McKenna (2013) all discussed the multifaceted concept of wisdom and especially "practical wisdom," which involves adapting to new situations, solving complex problems, and balancing competing interests while maintaining one's values. Practical wisdom has its roots in the concept of phronèsis which was explained before.

Nonaka and Takeuchi (2019) explained that wisdom is cultivated through practice,

transforming knowledge into wisdom. They argue that wisdom persists through generations and is essential for businesses, communities, and society's sustainability. Adams (2007) identified wisdom leadership as a key component of success in fast-paced and complex environments. Schwartz and Sharpe (2010) stated that wise leaders balance short-term goals with long-term values and engage in ethical decision-making while inspiring and empowering others.

McKenna et al. (2009) presented five principles of wisdom that can be used to measure and evaluate wise leadership. The principles include the use of reason and careful observation, allowing for non-rational and subjective elements when making decisions, valuing humane and virtuous outcomes, being practical and oriented towards everyday life, being articulate, understanding the aesthetic dimension of work, and seeking intrinsic personal and social rewards.

Grossmann and Brienza (2018) proposed that wise reasoning, distinct from intelligence, helps solve societal problems. Sternberg's WICS (Wisdom, Intelligence, and Creativity Synthesized) model (Sternberg, 2003) emphasizes the integration of practical, creative, and analytical abilities and the application of wisdom in decision-making. He also identified five patterns among unwise leaders. According to the model, effective leaders are those who can integrate these different components, such as the ability to communicate effectively, build relationships, and inspire others in a balanced way, depending on the specific situation they are facing (Sternberg, 2003).

In their studies, Nonaka and Takeuchi (2011) showed that practical wisdom is crucial for prudent judgments and actions. They outlined six abilities of wise leaders, highlighting the importance of practical wisdom such as decision-making for the good of the organization and society, perception and understanding of people, things, and events quickly, creating contexts for meaningful interactions, using metaphors and stories to convey tacit knowledge, employing political power to mobilize action, and mentoring and cultivating practical wisdom in others.

"Prudence" is the actual translation from the word "Phronèsis" that Aristotle used to describe the ability to find the balance between two extremes and make the appropriate decision that both minimizes harm and maximizes the good. Riggio et al. (2010) argued that prudence is essential for ethical leadership. Rooney et al. (2021) proposed a model that helps leaders develop skills and habits needed for wise decision-making in complex and uncertain environments.

To make informed decisions, managers must understand a company's purpose and pursue the common good to ensure sustainability. Japanese companies practicing "Wise Capitalism," as described by Nonaka and Takeuchi (2011), prioritize the common good, create both economic and societal value, and align with social entrepreneurship principles. This approach represents a shift away from the narrow, profit-maximizing view of capitalism and towards a more holistic and sustainable model of business that recognizes the interdependence of business and society. Rocha and Pinheiro (2021) explored the role of business education in addressing gaps in leaders' awareness of organizational practical wisdom through 23 interviews. The research emphasizes the importance of Phronèsis in creating wise organizations.

Rooney and McKenna (2022) propose a unique application of wisdom in practical leadership scenarios, suggesting that wisdom operates both as a personal attribute and a collective duty. It is crucial to consider one's part in preventing the projection of toxic leaders into influential roles within commercial or political spheres. Such harmful leaders frequently overemphasize their achievements and abilities. However, the responsibility for their placement in power principally lies with others. Regardless of one's ability to embody the characteristics of a wise leader, one can play a crucial role in shaping an environment that discourages the rise of toxic leaders. This environment promises to yield the common good.

Various researchers have developed tests to measure wisdom, such as Ardelt (2003) and Webster (2003), emphasizing the importance of reflection and understanding life's dialectical and uncertain nature as fundamental to wisdom. Ardelt (2003) described wisdom as a set of personality attributes that allow people to see things from other people's viewpoints, overcome prejudices and blindspots, learn from life, and care for people. She suggested wisdom involves three personality dimensions incorporating cognitive, affective, and reflective(3D-WS). Ardelt (2003) defined the cognitive dimension as obtaining knowledge and its application, critical thinking, and the capacity to make informed judgments through problem-solving and decision-making. Compassion, empathy, and emotional control are examples of affective traits that allow a wise person to show concern and sensitivity for the well-being of others. Finally, the reflective component stresses self-awareness, introspection, and the ability to learn from previous experiences, which includes knowledge of one's opinions, beliefs, and values. In his Self-Assessed Wisdom Scale (SAWS), Webster (2003) recognized five aspects of wisdom. It includes "Experience, Emotional Regulation, Reminiscence and Reflectiveness, Openness, and Humor."

He has developed a Self-Assessed Wisdom Scale (SAWS). Both researchers highlight the significance of critical life experiences and reflection in developing wisdom.

### 2.4.1 Navigating the Paradoxes: The Interplay of Wisdom and Leadership

Sheppard (2022) investigates the leadership attributes of extraordinary people who have effectively navigated complicated situations. These leaders exhibit an unusual mix of apparently contradicting traits. He refers to them as the Six Leadership Paradoxes, which require leaders to successfully balance contrasting personalities, abilities, and beliefs to manage complex and contradictory situations.

Concerning wisdom, the Six Leadership Paradoxes align with several scholarly definitions of wisdom. As observed in these paradoxes, an effective leadership is not straightforward or linear. Instead, it entails the integration and balance of disparate elements, such as being “tech-savvy yet humanistic”, or having “high-integrity” while navigating “political” environments. The integration of practical and analytical abilities, context-specific knowledge, developing healthy interpersonal interactions, prudent judgments and an understanding of uncertainty are crucial in balancing this paradoxes.

These connections underscore the significance of wisdom in leadership. Leaders who demonstrate wisdom can skillfully navigate these paradoxes, effectively balancing contrasting traits and navigating complex situations. Therefore, studying wisdom in leadership can offer valuable insights into how leaders can cultivate these essential attributes, ultimately enhancing their effectiveness in their roles.

Wisdom involves a nuanced understanding of one’s goals and values and the ability to navigate complex social interactions in ways that align with those goals and values. These scholars suggest that wisdom is complex and multifaceted construct and context-dependent. Hence, any attempt to define it precisely and unambiguously is likely to fall short and requires further investigation and empirical testing. The following section suggests two reasons why wisdom has remained undefinable for ages. It also emphasizes the value of ongoing investigation and cooperation in the search for a deeper and more comprehensive understanding of this complicated concept.

## 2.5 Wisdom: A Mysterious and Multidimensional Concept

As mentioned, the idea of wisdom has endured throughout human history, drawing scholarly interest from many fields, including philosophy and psychology. A universal agree-

ment on the definition of wisdom is elusive despite much investigation. There are two explanations for this. First, the constraints of individual insight necessitate more interdisciplinary collaboration. Second, the complex nature and inherent subjectivity of wisdom.

### 2.5.1 Limitations of Individual Insight and the Need for Interdisciplinary Collaboration

Rumi's parable of the elephant in the dark chamber is a helpful metaphor for examining the difficulties associated with the concept of wisdom and the limitations of personal insight.

In Rumi's metaphor, numerous men touch various portions of an elephant as they try to explain it in a dark space. Because each man's description is constrained by his or her own perception and background, they only partially understand the elephant's true character. This narrative highlights the importance of embracing openness to diverse points of view to develop a more thorough awareness of the world by showing how individual opinions and experiences can limit one's sense of reality.

When this metaphor is applied to the study of wisdom, It becomes clear that the different viewpoints of academics from various fields may be the reason for the lack of a generally agreed-upon definition. Each academic discipline may give its own distinctive insights and knowledge of wisdom, much like the elephant in the room, but a thorough understanding necessitates the confluence of diverse viewpoints. The introduction of light into the dark room signifies the potential of different types of perspectives and research methodologies. If light - in the form of new theories, interdisciplinary studies, or innovative analytical tools - were introduced, the true form of the elephant could be perceived. Scholars may advance a more comprehensive and nuanced understanding of wisdom by encouraging interdisciplinary collaboration and discussion, transcending the constraints imposed by unique experiences and academic boundaries.

### 2.5.2 The Complexity and Subjectivity of Wisdom

Marcel's view on the distinction between problems and mysteries is discussed in his book "The Mystery of Being" (Le Mystère de l'être), which was originally published in French in 1951 (Knepper, 2020). In this book, Marcel argues that there are two distinct ways of approaching reality: a *problem* to be solved through scientific inquiry or a *mystery* to be encountered through personal reflection and existential questioning.

Marcel believed that problems are situations that can be solved through rational in-

quiry, whereas mysteries are situations that resist rational investigation and can only be understood through intuition and personal experience. Problems are situations that we can solve through our own efforts and resources, while mysteries are situations that require us to look beyond ourselves and rely on something greater than ourselves. Marcel argued that recognizing the difference between problems and mysteries is crucial for understanding our place in the world and cultivating a sense of wonder and awe.

The lack of a clear and concise definition of wisdom could be due to a combination of factors, including the complexity of the concept and the subjective nature of its interpretation. While there have been numerous attempts to define and understand wisdom, it is possible that some aspects of wisdom may be inherently mysterious or elusive.

Gabriel Marcel's distinction between problems and mysteries suggests that certain facets of the human experience could be inherently enigmatic and not entirely amenable to logical explanation. Even if this could be the case for some parts of wisdom, further study and multidisciplinary discussion may eventually result in a more precise understanding and more thorough description of wisdom.

By examining commencement speeches from successful individuals, this study aims at exploring wisdom in leadership. It acknowledges the difficulty in defining wisdom, as shown by Rumi's parable and Marcel's distinction between problems and mysteries. The research uses text-mining tools to find characteristics of smart leadership that are present in many different professions. The study adds to a more varied and nuanced understanding of wisdom by bridging various points of view. The study emphasizes the significance of multidisciplinary collaboration in generating a more thorough knowledge of wisdom while acknowledging the inherent subjectivity in defining wisdom. The study aims add to our understanding of wisdom and its role in effective leadership by examining transcripts of commencement addresses given by successful people. We increase our appreciation for the complexity of wisdom as we continue to discern between issues and mysteries. This fosters curiosity that leads us to a deeper comprehension of the idea.

In conclusion, the notion of wisdom is a broad and complicated construct that has been investigated in numerous branches of literature. It has cognitive, moral, and interpersonal components that represent the complexities of human thought and action. To summarize all of the numerous definitions presented above, we may cluster them into three major groups:

1. Personal (Intrapersonal) Dimensions:

- Expert knowledge acquisition, critical thinking, and awareness of ignorance
- Integration of intelligence and creativity
- Dialectical thinking and reflection
- Integration of rational and non-rational elements
- Emotional control

2. Relational (Interpersonal) Dimensions:

- Empathy, compassion
- Communication, storytelling, and humor
- Adaptability and environmental shaping

3. Ethical and Contextual Dimensions:

- Moral grounding
- Focus on the common good
- Long-term and short-term balance
- Contextualism and relativism of values

These clusters emphasized the varied character of wisdom and the significance of integrating distinct components with practical knowledge, moral grounding, and interpersonal skills in order to promote the common good and navigate complicated circumstances. Thus far, we have examined the notion of wisdom, but in order to analyze our dataset, which contains transcripts of speeches delivered by notable individuals, we must enter an entirely new realm. The next part describes the text mining techniques and algorithms that will aid us in our research.

## 2.6 Text Analysis

Due to the exponential expansion of digital data, it is crucial for academics and organizations to create techniques for effectively deriving insightful information from massive volumes of text data. A powerful method for analyzing and processing unstructured textual data to uncover relevant patterns and correlations is text mining, a branch of natural language processing (NLP). Topic modeling, a collection of primarily unsupervised algorithms that seek to reveal the hidden theme organization contained within a corpus of texts, is an essential component of text mining. To better grasp the underlying themes and attitudes, these strategies have been used in a variety of contexts, including news stories (K. Chen et al., 2023; Rajasundari et al., 2017), scholarly publications (Y. Wang et al., 2017; Glazkova, 2021), consumer evaluations (F. Wang et al., 2019), behavioral marketing (Dan, 2023), and political speeches (Müller-Hansen et al., 2021; Atiq et al., 2022).

This section of the literature review's main goal is to examine the volume of text mining research that has already been done, with an emphasis on topic modeling and the discovery of themes in particular. Some prominent topic modeling techniques, such as Latent Dirichlet Allocation (LDA), Top2Vec, and Structural Topic Model (STM), will be discussed in this study, along with their guiding principles, advantages, and disadvantages. The review will also go over text representation and preprocessing methods that are essential for successful topic modeling and look at the relationships between text mining findings and wisdom leadership aspects. This literature review intends to shed light on the possibilities of text mining and topic modeling approaches in enhancing the knowledge of wisdom leadership via the analysis of speeches by combining the findings from prior research.

## 2.7 Text Analysis vs. Text Mining: Convergence and Divergence

Text mining and text analysis have distinct roots but are converging as their methods and applications overlap. Text analysis has its origins in the social sciences and humanities, dating back long before the concept of computers. It involves the systematic analysis of word use patterns in texts, combining formal statistical methods with humanistic interpretive techniques. Major approaches to text analysis include Frame Analysis (Goffman, 1974), Grounded Theory Methodology (Strauss & Corbin, 1994), Discourse

Analysis (Johnstone, 2017), (Qualitative) Content Analysis (Mayring, 2014), and conversation analysis (Sacks, 1992). These techniques establish procedures for analyzing textual data in the social sciences and may also fall into the area of qualitative data analysis (QDA).

Text mining, on the other hand, originates from computer science and involves information retrieval, natural language processing, part-of-speech tagging, syntactic parsing, named entity recognition (NER), and sentiment analysis. Knowledge discovery from text (KDT) or text mining was initially described by Feldman and Dagan (Girju, 2002). At the core of text mining lies natural language processing. Machine learning (ML) and natural language processing (NLP) algorithms are utilized in these techniques to automatically extract insights, patterns, and relationships from large volumes of unstructured text data (Moreno & Redondo, 2016). Applications of these techniques span various domains, such as business, healthcare, social media, and education (Deng & Liu, 2018).

## 2.8 Text Mining: A General Overview

### 2.8.1 Definition and Key Concepts

Although several researchers have given different explanations of text mining, Hotho et al. (2005) defined it clearly and concisely as "the application of algorithms and methods from the fields machine learning and statistics to texts with the goal of finding useful patterns."

### 2.8.2 Common Text Mining Tasks and Techniques

In text mining, there are several typical tasks and techniques, including but not restricted to:

*Text preprocessing* is the first stage of text mining in the bag-of-words approach (Harris, 1954), which entails preparing raw text data for subsequent analysis by cleaning and putting it into a structured format. Tokenization, stopword elimination, stemming, and lemmatization are a few examples of this (Nayak et al., 2016). All of which can be done in various ways and algorithms (Vijayarani et al., 2015).

*Sentiment analysis* identifies the emotions present in a text and classifies them as positive, negative, or neutral. In many circumstances, Sentiment analysis helps understand public opinion, client feedback, or general mood (Medhat et al., 2014).

*Named entity recognition (NER)* is the process of locating and categorizing named entities—such as individuals, groups, places, and dates—in texts. NER can assist in the extraction of structured data from unstructured data (Goyal et al., 2018).

*Text classification* is the process of classifying texts into predetermined groups or labels based on their content (Kobayashi et al., 2018). Supervised machine learning techniques, including Naive Bayes, Support Vector Machines, Nearest Neighbor, Decision Tree Induction, Centroid/ Association based Classifications, and Neural Network deep learning strategies, are frequently used for text categorization (Desai et al., 2015).

Discovering latent topics or themes within a collection of texts is the goal of the mostly unsupervised learning approach known as "*topic modeling*" (Kherwa & Bansal, 2019). Popular topic modeling techniques include Latent Dirichlet Allocation (LDA) (Blei et al., 2003), Top2Vec (Angelov, 2020), and Structural Topic Model (STM) (Roberts et al., 2019).

Text mining is a flexible and effective technology that may be used in a variety of fields to unearth important insights and guide decision-making. Because of developments in NLP, machine learning, and artificial intelligence, text mining techniques and approaches are continually changing.

Most researchers do not have the time to read thousands of pages. Reading alone is deemed too vulnerable to biased judgments to be considered scientifically reliable since it relies on the subjective interpretation of texts (Mayring, 2014). As mentioned above, while the debate between social researchers (text analysis) and computer scientists using (text mining) continues, there has been an increasing adaptation of text mining tools for social scientists' research. Two continuous trends have altered the way social scientists do qualitative data analysis: first, the rapid rise of the 5Vs (Volume, Velocity, Variety, Veracity, and Value) of big data (Anuradha et al., 2015) in digital text worthy of investigation, and second, the advancement of text analysis technology that enables a range of tools and new capacities (Wiedemann & Wiedemann, 2016). Some researchers want to narrow the gap between these techniques by addressing theoretical, methodological, and empirical elements of topic modeling and researching strategies to enhance the creation and deployment of these models in understanding social phenomena (Doogan, 2022).

It is critical to evaluate the preprocessing and text representation procedures involved in preparing the textual data for analysis and knowing the different topic modeling strategies. The quality of preprocessing and text representation can have a great influence on

topic model performance and interpretability. In the next section, we will go through some of the most prevalent preprocessing and text representation approaches available in the literature.

## 2.9 Preprocessing Techniques

Text preprocessing is a vital step in text-mining because it transforms raw, unstructured data into a format that can be effectively analyzed and interpreted. It involves various processes like tokenization, removal of stop words and punctuation, and stemming or lemmatization. Without this crucial step, the analysis may be tainted by irrelevant information or overwhelmed by data variability, thereby negatively affecting the accuracy of subsequent tasks like sentiment analysis or topic modeling.

### 2.9.1 Tokenization

The practice of breaking down a text into separate words or tokens is known as tokenization (Vijayarani, Janani, et al., 2016). Because it converts unstructured text into a more organized format, this is frequently the initial step in preparing text data for analysis. Tokenization methods include Lucene Analyzer, Byte Pair Encoding (BPE), whitespace-based, rule-based, and more complex algorithms that employ machine learning models or language-specific resources (Rai & Borah, 2021). The type of tokenization used might be influenced by the unique properties of the text data as well as the desired level of granularity (Boyd-Graber et al., 2014).

Although some researchers (Rahmoun & Elberrichi, 2007) classify n-grams and skip n-grams (Goodman, 2001) under text representation techniques, they may be better considered an extension of tokenization because they represent continuous sequences of n-words in a given text (Žižka et al., 2019). While simple tokenization frequently entails breaking text into individual words (unigrams), n-grams can aid in the capture of local context and word order information by generating tokens comprising two or more consecutive terms. Bigrams (n=2), for example, represent pairs of successive words, trigrams (n=3) represent three-word sequences, and so on. The following example shows how the sentence “This is an example of tokenization.” can be tokenized.

```
## [1] "Unigram:"
```

```
## # A tibble: 6 x 1
```

```
##   word
##   <chr>
## 1 this
## 2 is
## 3 an
## 4 example
## 5 of
## 6 tokenization
```

```
## [1] "Bigram:"
```

```
## # A tibble: 5 x 1
##   bigram
##   <chr>
## 1 this is
## 2 is an
## 3 an example
## 4 example of
## 5 of tokenization
```

```
## [1] "Trigram:"
```

```
## # A tibble: 4 x 1
##   trigram
##   <chr>
## 1 this is an
## 2 is an example
## 3 an example of
## 4 example of tokenization
```

Text preprocessing activities that rely on word order or contextual information, such as sentiment analysis, machine translation, or language modeling, can benefit from using n-grams and even combining n-grams (Dashtipour et al., 2019). However, as the value of n and the window size of skip grams rises, so does the number of potential n-grams and the size of the feature space, which can result in increasing computational complexity and memory needs (Li et al., 2005). As a result, selecting a suitable n value is critical based on the work type and available resources.

### 2.9.2 Stemming and Lemmatization

To shrink words to their base or root form, stemming and lemmatization processes are performed. This can aid in improving the consistency of word representations and reducing data dimensionality. Stemming leads to deleting word affixes (such as suffixes or

prefixes), whereas lemmatization uses morphological analysis and linguistic resources, such as dictionaries or morphological databases, to turn words into their base forms. The decision between stemming and lemmatization is determined by the language's complexity and the resources available for analysis. While both lemmatization and stemming can be useful methods for reducing words to their basic form, it is not suggested to utilize both procedures simultaneously since they may yield contradictory or duplicate findings. The following example shows the results of stemming and lemmatization for the words "running", "ran", "jumps", "leaves", "better".

```
## Stemmed words:  'run' 'ran' 'jump' 'leav' 'better'
```

```
## Lemmatized words:  'run' 'run' 'jump' 'leave' 'good'
```

Schütze et al. (2008) compare and contrast stemming and lemmatization in the context of information retrieval. The authors point out that stemming is faster but less accurate than lemmatization and that the decision between the two approaches is dependent on the application's unique demands (Schoemaker et al., 2018).

By considering the context and grammatical structure of the word, lemmatization is often more precise than stemming. As a result, lemmatization can yield more relevant and interpretable findings (Balakrishnan & Lloyd-Yemoh, 2014), which is especially crucial for applications like topic modeling and sentiment analysis (May et al., 2016).

Stemming, on the other hand, is often faster and easier than lemmatization since it simply requires the application of a set of predetermined rules (Schütze et al., 2008). However, stemming can occasionally yield mistakes - over-stemming and under-stemming - (Jivani et al., 2011) because it does not examine the context or grammatical structure of the word. As a result, the decision between lemmatization and stemming is determined by the job or application's unique needs. Lemmatization may be better suited in some circumstances since it generates more accurate and interpretable results. In other cases, stemming may be preferable because of its speed and simplicity.

### 2.9.3 Stopwords Removal

Stopwords are common words with minimal significance that can be eliminated from the text in order to decrease noise and computational complexity. Stopwords include articles, prepositions, and conjunctions (for example, "the," "and," and "in") (Asghar et al., 2014).

Stopword removal can be done with predetermined lists, frequency-based approaches, or more complex systems that take into account word distribution over the full document collection. The individual qualities of the data and the required level of text filtering might impact the choice of stopword removal strategy (Kaur & Buttar, 2018).

### 2.9.4 Text Representation Techniques

Text representation approaches are ways of converting text input into a structured format that machine learning algorithms can readily understand and interpret. These approaches are critical for computers to be able to deal with human language and accomplish tasks like topic modeling, sentiment analysis, and information retrieval. Here are a few examples of typical text representation techniques:

The Bag-of-Words (BoW) concept, initially described by Harris (1954), is one typical approach. This approach depicts text as an unordered set of words that ignores syntax and word order while tracking word frequency. Each document is represented as a vector in feature space, with dimensions matching vocabulary words.

Term Frequency-Inverse Document Frequency (TF-IDF) (Jones, 1972; Jones, 1973; Jones, 2004) is a text analysis technique used for document comparison, text categorization, and information retrieval. It uses statistical assumptions to determine the significance of a token in a corpus. The method is divided into two parts: Term Frequency (TF) and Inverse Document Frequency (IDF). The term frequency (TF) reflects how frequently a token appears in a text in comparison to all other tokens. IDF counts the number of times a token appears in a corpus of documents. The product of TF and IDF yields a score or weight for each token, evaluating its importance in relation to other tokens. The TF-IDF principle states that the more frequently a token occurs, the more essential it is, and the less frequently a token appears, the more important it is as well. This aids in the identification of key tokens inside a corpus document, with TF representing the first half of a Bag-of-Words method and TF-IDF determining which tokens describe an individual paper.

Word Vectors, also known as Word Embeddings, are dense vector representations that encapsulate semantic meaning and word relationships within a continuous mathematical space (Salim & Mustafa, 2022). These representations have become crucial in the field of NLP, mainly due to the pioneering work of researchers such as Tomas Mikolov (Lopez-Martinez & Sierra, 2020).

Mikolov et al. (2013) introduced Word2Vec, a technique that ignited a wave of research and development in word embeddings. The approach involves training shallow neural networks on large text corpora to yield word vectors that capture complex semantic and syntactic relationships. Following Word2Vec's success, Mikolov, along with Le, further extended this idea with the development of Paragraph2Vec (Le & Mikolov, 2014), which focused on learning representations for larger pieces of text, like sentences and paragraphs.

Tomáš Mikolov's contribution to the field did not stop there. Mikolov also played a significant role in the development of FastText (Joulin et al., 2016). Like Word2Vec, FastText is designed to capture the syntactic and semantic meaning of words, but with several improvements. It also considers the internal structure of words, allowing it to better handle rare words and languages with complex morphology. Another notable method for generating word embeddings is GloVe (Global Vectors for Word Representation), developed by Pennington et al. (2014).

These models have significantly contributed to the development and refinement of word embedding techniques. They continue to play a substantial role in the way we represent and understand textual data in machine learning.

## 2.10 Topic Modeling

Topic modeling emerged in the 1980s as a part of the "generative probabilistic modeling" field (Liu et al., 2016). Researchers often use topic models as a time-saving and unbiased alternative to reading extensive text collections in various domains (Jelodar et al., 2019). Latent Dirichlet Allocation (LDA), Top2Vec, and Structural Topic Model (STM) are the most widely used models for topic modeling. Because of its interpretability, topic modeling has found enormous applications in a variety of areas (Abdelrazek et al., 2022). Topic models see texts as a "bag of words," recording word co-occurrences independent of syntax, narrative, or placement within a text. A subject may be regarded as a group of words that often appear in a discussion and hence co-occur more frequently than they would otherwise, anytime the issue is discussed. The TF-IDF reduction strategy was the first approach designed for this goal.

### 2.10.1 Latent Dirichlet Allocation (LDA)

Blei et al. (2003) established Latent Dirichlet Allocation (LDA) as an unsupervised topic modeling approach. LDA is a probabilistic generative model that seeks to uncover latent themes within a set of texts. The essential premise of LDA is that texts contain mixes of subjects, and each topic is defined by a distribution of words.

The number of topics in LDA is fixed, and the technique uses the Dirichlet distribution (Minka, 2000) to estimate the topic-word distribution and document-topic distribution. As a consequence, each topic is represented by a list of words with associated probabilities (Blei et al., 2003).

Zvornicanin (2022) provided a more extensive but simple description of how LDA works, which may be found here:

1. Initialization: LDA asks you to specify how many topics (K) you wish to extract from the texts. The model begins by allocating each word in the texts to one of the K subjects at random. Topic allocations are haphazard at this point and do not reflect relevant themes.

2. Iterative Process: The LDA model then iteratively refines the topic assignments using Gibbs sampling or variational inference. The topic allocations for each word are updated using two probabilities:

a. Document-Topic probability (P(topic | document)): The likelihood of a topic based on the document. This reflects the percentage of words in a document that are related to a specific topic.

b. Topic-Word probability (P(word | topic)): The likelihood of a particular word given the topic. This is the percentage of a certain term in a subject in all documents in the corpus.

3. Convergence: The model iteratively updates the topic assignments based on the previously stated probabilities until the assignments become stable and the model converges. A specified threshold for the change in the maximum number of repeats (iterations) is often used to establish the convergence criteria.

4. Topic Extraction: Once the model has converged, the resultant topic assignments offer information on the words that comprise each subject as well as the topic distributions for each document. Each subject is represented by a collection of words and their associated probabilities, with higher probability indicating phrases that are more important to the topic. These high-probability terms are then used to interpret the themes.

5. Topic Assignment: LDA gives a probability distribution of topics for each document. The proportion of words in the document that pertain to each topic is shown by this distribution, and the document may be allocated to the topic with the highest proportion.

It is vital to remember that LDA is an unsupervised learning approach (Blei, 2012), which means that it does not learn subjects using labeled data. Instead, it finds subjects entirely based on word patterns and co-occurrences in the texts. As a result, the quality and interpretability of the retrieved themes are significantly dependent on parameter selection (Griffiths & Steyvers, 2004), text preparation, and the structure of the underlying data (Blei et al., 2003).

It is critical to understand the difference between Linear Discriminant Analysis and Latent Dirichlet Allocation. Despite having the same abbreviation, Linear Discriminant Analysis (LDA) and Latent Dirichlet Allocation (LDA) are two completely different approaches utilized in separate fields of research. In a larger statistical framework, Linear Discriminant Analysis is employed for class separation or dimensionality reduction (Balakrishnama & Ganapathiraju, 1998).

### 2.10.2 Top2Vec

Top2Vec is another unsupervised topic modeling approach proposed by Angelov (2020) that makes use of the capabilities of contemporary word embedding and density-based clustering techniques. Top2Vec begins by creating a dense document and word embeddings via methods such as Word2Vec (Mikolov et al., 2013), Doc2Vec (Le & Mikolov, 2014), or BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018). The HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) (Campello et al., 2013) technique is then used to cluster the dense document embeddings based on their semantic similarity.

The resultant clusters indicate subjects, with each topic defined by the terms that are most semantically comparable to it. Top2Vec is remarkable for its ability to estimate the ideal number of topics based on the data's intrinsic structure (Angelov, 2020). This method has demonstrated promising results in a variety of applications and provides an alternative to established approaches such as LDA (Egger & Yu, 2022).

### 2.10.3 STM (Structural Topic Model)

Roberts et al. (2019) developed Structural Topic Model (STM) as an extension of LDA that adds document-level metadata into the topic modeling process. This enables STM to capture the links between document characteristics and the found topics, which may be valuable for investigating how topics differ across contexts or time periods.

STM may simulate a wide range of information, including authorship, publication date, and categorization designations. Incorporating topical prevalence covariates - metadata that explain topical prevalence - into the modeling process can improve topic interpretation and give new insights into the underlying structure of the document collection (Roberts et al., 2019).

### 2.10.4 A Synopsis of Supervised Methods

While unsupervised approaches such as LDA, Top2Vec, and STM dominate the topic modeling field, supervised strategies that use labeled data to assist the topic-finding process are also available. LDA, for example, is extended by integrating response variables linked with each document, allowing the model to predict these variables based on the learned topic distributions.

Supervised approaches can be beneficial when labeled data is available, or certain subjects of interest are known ahead of time. However, this literature review focuses primarily on unsupervised techniques, as they are more generally applicable when labeled data is insufficient, or the goal is to uncover latent subjects without any prior assumptions.

Table 2.1 summarizes the advantages and disadvantages of Latent Dirichlet Allocation (LDA), Top2Vec, and Structural Topic Models (STM). This table gives a broad summary of the benefits and drawbacks of each algorithm. Specific benefits and drawbacks may vary depending on the nature of the text data and the application's requirements.

Table 2.1: Comparing LDA, Top2Vec, and STM

| | LDA | Top2Vec | STM |
|---|---|---|---|
| **Advantages** | | | |
| | Generative probabilistic model | Combines document & word embeddings | Incorporates document-level metadata |
| | Scalable to large corpora | Automatically determines number of topics | Flexible modeling approach |
| | Interpretable topics (with proper tuning) | Provides topic, document, and word similarities | Can model relationships between topics |
| | Wide range of applications | Faster than traditional topic modeling techniques | Directly models covariates |
| **Disadvantages** | | | |
| | Difficulty in selecting number of topics | Sensitive to the choice of word embeddings | Assumes linear effects of covariates |
| | Parameter selection is crucial | Difficult to interpret dense embeddings | Requires proper metadata |
| | No word-order or context information | Limited research on algorithm performance | Less interpretable than LDA (in some cases) |
| | Assumes topic independence | Not as established as LDA or STM | Sensitive to tuning and preprocessing |

CHAPTER 3

# METHODOLOGY

## 3.1 Introduction to Methodology

A thorough understanding of the data collection, research design, and analysis procedures is involved to ensure the validity and reliability of the study's conclusions. As a result, this chapter describes the study's methodology, which attempts to analyze the similarities and contrasts linking leaders' speeches and features of wisdom leadership.

A mixed-methods strategy is used in this study, integrating qualitative and quantitative text analysis tools to assess a broad collection of commencement addresses. This method allows us to derive insights from the speeches, discover common themes and patterns, and investigate their relevance to wisdom leadership. We may triangulate our findings and overcome the limits of individual approaches with the mixed-methods design, resulting in a more robust knowledge of the subject matter. This chapter is divided into parts that describe the methods of the study, such as data collection and selection, research design, text-mining techniques, data analysis procedure, ethical issues, and limits.

## 3.2 Research Design: Selection of Methodology

In this study, we use a mixed-methods research strategy that combines qualitative and quantitative text-mining tools to examine a collection of commencement speeches. This

technique enables us to identify commonalities across these addresses and extensively investigate their relationships to dimensions of wisdom leadership in a comprehensive manner.

Our analysis will follow a systematic method that includes data collection, preparation, analysis, and result interpretation to achieve the study objectives. Our study design is intended to ensure that the findings are credible and valid.

The *Quantitative Analysis* of our study focuses on discovering and measuring trends within the speeches. Word frequency analysis, for example, provides objective and quantifiable statistics by emphasizing the most frequently used words and phrases across the speeches. Furthermore, we can quantify topic occurrence and sentiment scores, allowing us to compare and contrast speeches and identify statistically significant trends. The quantitative analysis provides a more systematic and objective view on the similarities between the speeches, reinforcing the qualitative research's findings.

Typically, many text-mining techniques operate on the bag-of-words model. As explained in Chapter 2, this approach essentially simplifies text into a collection of n-grams. However, the bag-of-words model has its limitations. Primarily, it ignores the context and order of words, focusing instead on the frequency or presence of words. This results in the extraction of n-grams might not make much sense or represent the text accurately. Thus, interpreting these findings without accounting for the broader context in which these n-grams are found can indeed be unrealistic and even impractical.

A more robust approach would be to couple this with *qualitative analysis*, examining the context in which these words and phrases appear to grasp the nuances of the text. For example, sentiment analysis can help determine the emotion or opinion conveyed in a text, while topic modeling can uncover the underlying themes. Incorporating a *mixed-methodology* approach indeed enhances the depth of speech assessment as it blends the strengths of both qualitative and quantitative techniques.

## 3.3 Research instrument: Data Collection and Selection

The data collection and selection procedure constitute a critical phase in our study as it forms the foundation for our future analysis and interpretation of the speeches and topic modeling outcomes. In this section, we outline the criteria employed in selecting the speeches for our study, the sources of our data, and the data preprocessing techniques

we deployed to ensure data consistency and quality.

### 3.3.1 Speech Selection Criteria

To ensure a diverse and representative sample of speeches, we selected 32 commencement addresses (summarized in table 3.2) from a range of influential figures spanning the following fields: law and politics, business, arts and literature, entertainment, and academia. Table 3.1 shows the number of influential figures within each profession. The number of speeches was limited by time constraints but was deemed sufficient for the study's purposes. More information about the speeches can be found in Appendix I.

Table 3.1: The number of speakers associated with each profession

| profession | same_pro |
|---|---|
| Business | 9 |
| Law and Politics | 8 |
| Entertainment | 7 |
| Arts and Literature | 5 |
| Academia | 3 |

Initially, we contemplated analyzing biography books of renowned leaders such as Gandhi, Mandela, and Steve Jobs. However, we pivoted towards commencement addresses for several reasons. Firstly, the text preparation process for books, especially those not available in digitized format, necessitates substantial time and resources. Extracting text from physical books involves manual transcription or scanning and optical character recognition, all of which prolong the data collection phase of our research. This is in stark contrast to commencement speeches, which are readily accessible online on university websites, enabling us to expedite the data collection process significantly. Secondly, commencement addresses are typically more concise and focused, with speakers endeavoring to convey their wisdom to graduates, while books often encompass extensive non-insightful information that could potentially confound our study, given that text-mining largely operates with the bag-of-words approach and frequency counts.

The selection of speakers for this study was based on their established recognition as influential characters within their respective fields, demonstrating exceptional vision and impacting their communities or the world significantly. We relied on the evaluations of various universities in choosing these leaders, under the assumption that such institutions would be highly discerning in selecting individuals possessing valuable qualities and wisdom to address their graduates.

The choice of speakers was also influenced by the recognition and reputation of the universities from which the commencement speeches were sourced. For instance, 12 speeches were derived from Harvard, and 9 from Stanford, both of which are globally renowned for their academic excellence and high standards. It was our assessment that prestigious universities, ranked highly in global education rankings, frequently invite distinguished individuals of profound wisdom and accomplishment to deliver commencement addresses.

Furthermore, the frequency of speeches from certain institutions also arose from our intent to gather as comprehensive a dataset as possible. Given the prestige of these universities and the high-profile nature of their chosen speakers, their commencement addresses were often readily available and well-documented, allowing for easier access. Therefore, the selection of speakers for this research was not only based on their individual merits as leaders but also on the prestige of the inviting universities. This dual criterion aimed to ensure that the chosen speeches would provide insightful, high-quality data for our analysis of wisdom leadership across various professional fields.

Table 3.2: The list of commencement addresses

| speaker | location | profession | year | title | SpeechLength |
|---|---|---|---|---|---|
| Chimamanda Ngozi Adichie | Harvard | Arts and Literature | 2018 | Cultivating a sense of empathy | 1823 |
| Jacinda Ardern | Harvard | Law and Politics | 2022 | The Fragility of Democracy | 1777 |
| BarackObama | Notre Dame | Law and Politics | 2009 | The World as It Should Be | 1692 |
| Martin Baron | Harvard | Law and Politics | 2020 | Imperfect though | 1678 |
| Jeff Bezos | Princeton | Business | 2010 | Two key questions | 1647 |
| BillGates | Harvard | Business | 2007 | Global inequality, technology and innovation | 1603 |
| Mike Bloomberg | Johns Hopkins | Business | 2021 | Communication in the Digital Age | 1541 |
| Sterling K. Brown | Stanford | Entertainment | 2018 | Let Your Light Shine | 1538 |
| Ken Burns | Stanford | Arts and Literature | 2016 | Reflections on History | 1507 |
| Tim Cook | Stanford | Business | 2019 | Responsibility and Building | 1436 |
| France Cordova | Stanford | Academia | 2020 | Navigating a Changing World | 1429 |
| Mariano Florentino Cuellar | Stanford | Entertainment | 2017 | The Gift of Progress | 1353 |
| Ellen DeGeneres | Tulane | Entertainment | 2009 | Follow Your Passion, Stay True to Yourself | 1334 |
| Richard Engel | Stanford | Law and Politics | 2015 | Taking the Leap | 1323 |
| Neil Gaiman | University of the Arts | Arts and Literature | 2012 | Make good art | 1319 |

| speaker | location | profession | year | title | SpeechLength |
|---|---|---|---|---|---|
| Atul Gawande | Stanford | Academia | 2021 | Finding Purpose in Life | 1304 |
| Reed Hastings | Stanford | Business | 2022 | Keys to Progress and Change the World | 1283 |
| Steve Jobs | Stanford | Business | 2005 | Commencement Address at Stanford | 1248 |
| John Lewis | Harvard | Law and Politics | 2018 | Importance of equality | 1189 |
| Angela Merkel | Harvard | Law and Politics | 2019 | Anything can change | 1170 |
| MichelleObama | Oberlin | Law and Politics | 2015 | Engage with the world around | 1102 |
| Elon Musk | Caltech | Business | 2012 | Think Big and Dream Even Bigger | 1099 |
| Conan OBrien | Dartmouth | Entertainment | 2011 | Failure and Invention | 1091 |
| Natalie Portman | Harvard | Entertainment | 2015 | Embracing Inexperience | 1076 |
| J.K. Rowling | Harvard | Arts and Literature | 2008 | The Fringe Benefits of Failure | 1074 |
| Sheryl Sandberg | BERKELEY | Business | 2016 | Finding Gratitude and Appreciation | 1022 |
| Ruth J. Simmons | Harvard | Academia | 2021 | Fight inequality | 999 |
| Steven Spielberg | Harvard | Entertainment | 2016 | A villain to vanquish | 992 |
| David Foster Wallace | Kenyon | Arts and Literature | 2005 | This is Water | 884 |
| Oprah Winfrey | Harvard | Entertainment | 2013 | An internal emotional G.P.S | 837 |
| Fareed Zakaria | Harvard | Law and Politics | 2012 | We live in an age of progress | 742 |
| Mark Zuckerberg | Harvard | Business | 2017 | Purpose and Community | 591 |

### 3.3.2 Data Source Evaluation

The speeches were collected from various sources, including official university websites. Although we initially collected over 40 speeches, we wanted to lessen any potential bias from our personal perceptions of the speeches by listening to speeches or reading transcripts. Consequently, we narrowed down our selection to 32 speeches obtained directly from university websites and reliable sources to ensure authenticity and accuracy.

The data collected from these sources presents several advantages:

Rich and Varied Content: The speeches cover a wide range of topics, providing ample material for identifying common themes and analyzing their connection to dimensions of wisdom leadership.

Authenticity: The speeches represent the speakers' genuine thoughts and experiences, providing an authentic and reliable source of information.

Transfer of Life Experience and Wisdom: Commencement addresses provide prominent speakers with an important platform to share their life experiences, wisdom, and insights with a new generation of graduates eager to learn and grow. These speeches offer valuable insights and advice drawn from the speakers' personal and professional journeys, which can inspire and guide the graduates as they embark on their own life paths. This transfer of wisdom broadens the graduates' worldview and lays the groundwork for their future success.

Importantly, the dimensions of wisdom leadership that emerge from our analysis should be understood as reflecting the perspectives of these speakers and may not necessarily be objective. Further discussions on the biases and limitations of this study can be found in the subsequent section.

By providing this in-depth overview of our data collection and selection procedure, we aim to establish a transparent approach to our study.

### 3.3.3 Data Preparation

To prepare the data for analysis, we performed several steps to ensure data consistency and quality:

*Data cleaning:* We removed any irrelevant information from the transcripts, such as annotations or stage directions, and tried to retain only the speaker's words.

*Formatting and standardization:* We standardized the transcripts by converting them to a consistent `.txt` format, ensuring uniformity in the text representation.

*Language verification:* We verified that all speeches were delivered in English.

*Metadata Extraction:* We extracted the metadata, such as speaker name, year, and location, from the text.

Following these data collection and preprocessing steps, we obtained a high-quality dataset of speeches that reflects the diversity of leadership perspectives and is suitable for our text-mining analysis.

### 3.3.4 NLP Techniques for Data Preprocessing

Natural Language Processing (NLP), as indicated in chapter 2.8, is concerned with the development of algorithms that allow computers to interpret and produce human language (Kalyanathaya et al., 2019). NLP plays a crucial role in text mining and analysis, as it provides the necessary tools to preprocess, manipulate, and analyze textual data. In this study, to prepare the speeches for subsequent topic modeling and sentiment analysis, we employed various NLP techniques that we described in Chapter 2. The NLP preprocesses methods are the following: text cleaning, stop words removal, tokenization, and lemmatization. In this study, we opted for lemmatization over stemming, as it provides a more accurate representation of the speeches' content and ensures a higher quality analysis of the commonalities among great leaders' speeches and their connection to dimensions of wisdom leadership.

The complete code for preparing our data frame can be found in Appendix II. By employing these NLP techniques, we were able to preprocess and transform the speeches into a structured format suitable for subsequent topic modeling and sentiment analysis. The application of NLP techniques not only facilitated the extraction of meaningful insights from the speeches but also contributed to a more robust and accurate analysis of the commonalities among great leaders' speeches and their connection to dimensions of wisdom leadership.

## 3.4 Data analysis

In our study, we have chosen word frequency analysis, topic modeling, and sentiment analysis as the primary text-mining techniques to extract commonalities among great leaders' speeches and analyze how these commonalities connect to dimensions of wisdom leadership. These techniques and the process through which they will help achieve our research objectives are as follows:

### 3.4.1 Word Frequency Analysis

Word frequency analysis is a fundamental text-mining technique that involves counting the occurrence of words or word combinations in a text corpus (Rayson, 2015). In this study, we employed word frequency analysis to identify the most commonly used unigrams or n-grams in the selected speeches, providing valuable insights into the leaders' communication patterns, focus areas, and ideas. We used unigrams, bigrams, and tri-

grams to capture different levels of word relationships and context. As such, we were able to capture a comprehensive view of the language used in the speeches, highlighting the most prevalent words or n-grams. This approach provided us with a better understanding of the speakers' communication models, the key topics they addressed, and the linkages between the different aspects of Wisdom Leadership.

### 3.4.2 Topic Modeling

Topic modeling is an unsupervised machine-learning technique that aims to discover hidden patterns or themes in a collection of documents. In this study, we employed three popular topic modeling algorithms, Latent Dirichlet Allocation (LDA) (Blei et al., 2003), Top2Vec (Le & Mikolov, 2014), and STM (Roberts et al., 2019), to identify and analyze commonalities among the selected speeches. These algorithms were chosen for their ability to effectively extract meaningful topics from large text corpora and their complementary strengths in revealing thematic structures within the speeches.

By incorporating LDA, Top2Vec, and STM topic modeling techniques in this study, we were able to effectively identify and analyze commonalities among the selected speeches. This approach not only allowed for a more robust extraction of topics but also provided complementary perspectives on the thematic structure of the speeches, eventually contributing to a greater understanding of the connection between great leaders' speeches and dimensions of wisdom leadership.

### 3.4.3 Sentiment Analysis

As mentioned in section 2.8, Sentiment analysis (opinion mining) or emotion AI, is a domain of natural language processing that aims to determine the sentiment or emotional tone behind a series of words. In the context of this research, sentiment analysis was employed to identify and quantify the emotional tone present in the selected speeches, offering insights into how great leaders convey their messages and the emotions they evoke in their audiences.

The sentiment analysis process involved several steps:

*Sentiment Scoring:* Using the trained machine learning models and the lexicon-based methods, each word or phrase in the speeches was assigned a sentiment score. These scores typically range from negative sentiment to positive sentiment, with 0 representing a neutral sentiment.

*Aggregation and Visualization:* The sentiment scores were aggregated for each speech to provide an overall sentiment score, which was then visualized using bar charts or other graphical representations. This allowed for the comparison of sentiments across different speeches and professions.

*Interpretation:* The sentiment analysis results were interpreted in the context of broader research objectives, examining how the identified sentiments relate to the dimensions of wisdom leadership and the commonalities among great leaders' speeches.

*The Sentiment of Words Associated with Topics:* To assess the sentiment of each topic, the sentiment scores of the words assigned to each topic were aggregated. This allowed for the determination of the average sentiment score for each topic, reflecting the general emotional tone associated with the underlying theme. By comparing the sentiment scores across topics, we could identify which themes were more positively or negatively charged.

*Sentiment of Documents in Topics:* For each document classified under a specific topic, a sentiment analysis was performed to determine the overall sentiment score of the document. This information was then used to calculate the average sentiment score for all documents within each topic. This approach provided a more granular understanding of the sentiment distribution within each topic and highlighted potential variations in sentiment among the speeches associated with a particular theme.

By analyzing the sentiment of both the topic words and the documents classified under each topic, this study offers a deeper understanding of the emotional content of the speeches and their relationship with the identified themes. This additional layer of analysis can help illuminate potential connections between the emotional tone of the speeches and the dimensions of wisdom leadership, as well as inform our understanding of how great leaders effectively communicate their messages and engage their audiences.

## 3.5 Data Analysis Process

In this section, we outline the data analysis process, describing the software and tools used, the model selection and parameter tuning, and the validation and reliability measures employed to ensure robust and accurate findings.

### 3.5.1 Software and Tools

The data analysis process employed various software tools and programming languages to conduct the text-mining techniques described in the methodology. The primary software and tools used in this study include:

1. R: R is a widely-used open-source programming language and software (R Core Team, 2021). Table 3.3 shows the details of the R version we used.

Table 3.3: The version details of the R

| Property | Value |
|---|---|
| platform | x86_64-w64-mingw32 |
| arch | x86_64 |
| os | mingw32 |
| crt | ucrt |
| system | x86_64, mingw32 |
| status | |
| major | 4 |
| minor | 2.2 |
| year | 2022 |
| month | 10 |
| day | 31 |
| svn rev | 83211 |
| language | R |
| version.string | R version 4.2.2 (2022-10-31 ucrt) |
| nickname | Innocent and Trusting |

2. RStudio: RStudio is an R and Python integrated developing environment (IDE). It comes with a console, a syntax-highlighting editor with direct code execution, and tools for graphing, history, debugging, and workspace management (Posit team, 2023). Table 3.4 shows the details of the RStudio version we used.

Table 3.4: Version details of RStudio

| | |
|---|---|
| Version | 2023.03.1+446 |
| Release Name | Cherry Blossom |
| OS (Mode) | desktop |

R was used for Latent Dirichlet Allocation (LDA), word frequency analysis (including unigrams, bigrams, trigrams, and skipgrams), and semantic analysis. Several packages were employed to facilitate the analysis, including tidytext (0.4.0.9000) (Silge & Robinson,

2021) for text manipulation and analysis, quanteda (3.3.0) (Benoit et al., 2021) for text processing and analysis, tm (0.7-11) (Feinerer et al., 2021) for creating and manipulating text documents, text2vec (0.6.3) (Selivanov & Feuerriegel, 2021) for text modeling and analysis, ggplot2 (3.4.0) (Wickham, 2021) for data visualization, tidyverse (1.3.2) (Wickham & RStudio, 2021) for data wrangling and manipulation, and memoiR (1.2-2) (Marcon, 2022) offers templates for publishing properly organized publications in HTML and PDF. The complete list of the packages in R is in Appendix III.

The use of these software tools and programming languages, along with their associated packages, enabled the efficient and accurate execution of the text-mining techniques outlined in the methodology, ultimately providing valuable insights into the commonalities among great leaders' speeches.

## 3.6 Ethical Considerations

In conducting this research, it was imperative to carefully consider and address ethical concerns to ensure that the study was conducted responsibly and with integrity. The following ethical considerations were considered during the research process:

*Privacy and Confidentiality:* The speeches and commencement addresses used in this study are publicly available and delivered by well-known public figures. Nonetheless, it was crucial to respect the privacy of the individuals whose speeches were analyzed. There is no personal information or sensitive content in the data, analysis and subsequent discussions.

*Data Source Attribution:* To maintain transparency and give proper credit, all speeches and commencement addresses were appropriately cited, with clear references to the original sources. This practice ensures the acknowledgment of the intellectual property of the authors and respects their rights to their work.

*Data Manipulation and Bias:* The text-mining techniques employed in this research were designed to be objective and unbiased. We carefully chose our methods and validated the results to minimize the potential for bias or misinterpretation in the analysis. Furthermore, we acknowledged and discussed any limitations in the dataset and the potential biases that could arise from the selection of speeches and commencement addresses.

*Research Integrity:* Throughout the research process, we adhered to principles of

research integrity, ensuring that the study was conducted honestly and transparently. The methodology and data analysis process were clearly outlined, and any assumptions or potential limitations were acknowledged and discussed.

## 3.7 Limitations and Assumptions

In conducting this research, several limitations and assumptions were identified that may have influenced the results and interpretations. Recognizing and addressing these factors contributes to the transparency and validity of the study. The following limitations and assumptions were considered:

*Limited Sample Size and Dataset Accessibility:* This research was conducted based on a dataset of 32 speeches, which limits the scope of the analysis and the generalizability of the findings. During the research process, a larger dataset containing 300 commencement addresses was discovered but could not be accessed despite efforts to obtain it from the dataset's owner. Consequently, the results and conclusions drawn from this study are constrained by the relatively small dataset and may not be representative of commencement addresses at large.

*Speech Selection:* The speeches and commencement addresses selected for this study represent a diverse group of influential public figures. However, the selection process may have unintentionally introduced biases or excluded other important speeches. The results and conclusions should be interpreted with the understanding that they are based on the specific dataset utilized in this study and may not be generalizable to all leaders' speeches.

*Language and Cultural Context:* The speeches included in this research are exclusively in English, which may limit the ability to detect commonalities in speeches delivered in other languages or cultural contexts. Furthermore, cultural nuances and rhetorical styles may differ across various countries and regions, which could affect the interpretation of the commonalities identified in this study.

*Text-mining Techniques:* The text-mining techniques employed in this research, including natural language processing, topic modeling, sentiment analysis, and word frequency analysis, have inherent limitations. The accuracy and interpretability of the results depend on the quality of the data, the choice of algorithms, and the parameter tuning. Additionally, the techniques may not always capture the subtlety and complexity of human

language, leading to potential misinterpretations or oversimplifications.

*Subjectivity in Interpretation:* While the text-mining techniques used in this study were designed to be objective and unbiased, some level of subjectivity may still be present in the interpretation of the results. Our preconceptions and biases could potentially influence the understanding and presentation of the findings.

*Temporal Context:* The speeches and commencement addresses analyzed in this research span a range of time periods, and the societal context and events at the time of their delivery may have influenced their content. The analysis may not fully account for the impact of temporal context on the commonalities identified in the speeches.

By acknowledging and addressing these limitations and assumptions, the study aims to enhance the transparency, validity, and reliability of the research. Moreover, recognizing these factors can help inform future research in this area, guiding the development of more robust and comprehensive methodologies to explore the commonalities among famous individuals' speeches and their connection to dimensions of wisdom leadership.

CHAPTER 4

# RESULTS AND DISCUSSION

## 4.1 Introduction

In this chapter, we will go on a multifaceted examination of commencement addresses, using the capabilities of Natural Language Processing (NLP) and text mining tools combined with qualitative analysis. The goal is to find patterns, similarities, and unique insights concealed within these rich narratives. We use complex modeling approaches to delve further into underlying issues after discovering repeated words and themes through frequency analysis. We want to create a more nuanced perspective of the dialogue by connecting these subjects to the speakers' occupations. It provides an opportunity to present the findings, discuss the relation to the research questions, and draw conclusions based on the data.

## 4.2 Word Frequency Analysis (Unigrams, Bigrams, Trigrams)

In this section, we will extract the words that exist in each and every corpus. Then we extract the most frequent words within each corpus (each field of profession) and also within all the corpora. Next, we extract the top *common words* defined as the words that existed in every corpus. After that, the same process is conducted for bigrams and trigrams.

To do that, after loading the required packages and importing our data frame, we split

Table 4.1: Total number of unique n-grams

| Unigrams | Bigrams | Trigrams |
|---|---|---|
| 6237 | 34669 | 38403 |

the text into single words using `unnest_token` function that strips all punctuation and converts each word to lowercase for easy comparability. We had done more than these text cleanings in our `preprocess_f()` function when we made our data frame. However, we will do it again just to tokenize our text. We follow the same process to tokenize bigrams and trigrams. Notice that in the `unnest_tokens` function, we apply a `token="ngrams"` argument to state we want n-grams, and the `n = 2` is used for bigrams and the `n = 3` for trigrams. Table 4.1 presents the total number of unique n-grams. Here, by unique, we mean that we did not count an n-gram more than once. It represents the total number of unique combinations.

Now we have a table with columns: speaker names, profession (field of activity), doc_id, year, and words. We can count the frequency of each word in each field. Since speeches vary in length, the frequency of words increases as the length increases. Therefore, we compute the relative frequency to get a fairer comparison. We have also defined a new column that counts the number of professions that each word appeared in. This column will help us to filter words that appeared in only one profession or in all the five professions: “Arts and Literature”, “Law and Politics”, “Business”, “Entertainment”, and “Academia”.

### 4.2.1 The Most Frequent Words

Figure 4.1 shows the word cloud of our most frequent words in all the corpora:

In examining the most frequent words in these speeches, it seems that a clear pattern of emphasis on community, ethical behavior, and personal growth emerges. The speakers may try to emphasize the potential and capacity each individual holds (“can”, “one”), highlighting the power of action proactively (“get”, “go”, “make”, “work”).

Simultaneously, they underscore the importance of knowledge and understanding (“know”) throughout life’s journey (“life”, “year”). They also stress the interconnectedness of individuals within broader communities and humanity (“people”, “us”), and acknowledge the larger global context we are part of (“world”).


Figure 4.1: Most Frequent Unigrams in Corpora

While analyzing the frequency of individual words (unigrams) provides a valuable first glimpse into these speeches' main themes and topics, this approach has its limitations. It treats every word in isolation, ignoring the context in which words appear and the relationships between them.

Analyzing the most frequent words within each profession group may provide insights into the themes and subjects most relevant to each field. Each profession might have its own unique language that they tend to focus on more. Figure 4.2 shows the most frequent words in each field of activity of speakers.

In the speeches across all professions, the words "can", "people", "good", "one", "life", "year", "know", "make", "think", and "us" are the most common. These terms consistently appear in the top lists for each profession, which indicates their wide-ranging use in different fields. The prevalence of these words suggests that despite the diversity of professions, common themes resonate in the speeches made by individuals in these fields. These include the ability to do something ("can"), interactions and relations with others ("people", "us"), quality and value assessments ("good"), individuality and singularity ("one"), time and experience ("year", "life"), understanding ("know"), creation or action ("make"), and cognitive processes ("think").

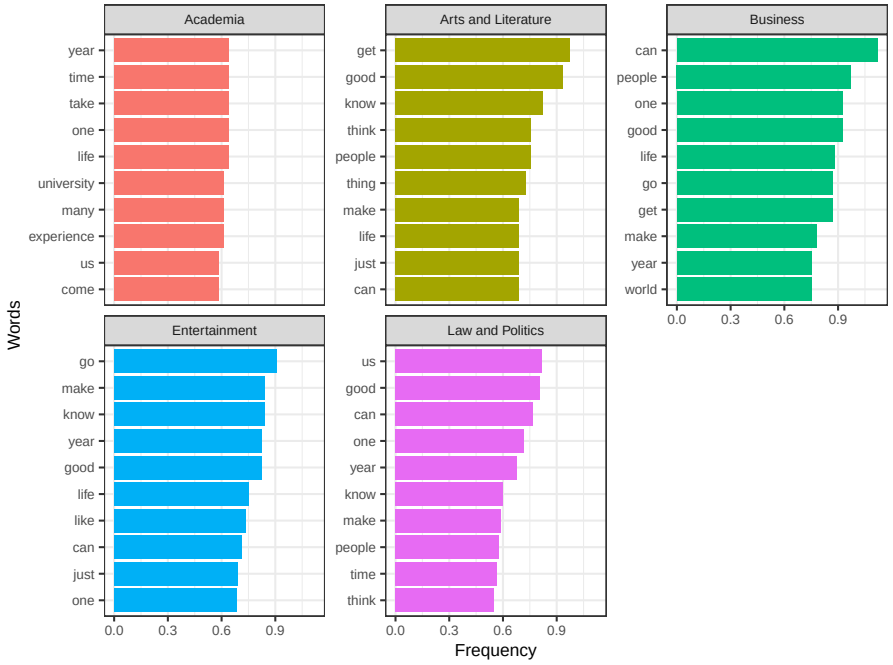



Figure 4.2: Top 10 Frequent unigrams in each Field

As we can see, numerous common unigrams are shared across different fields. However, to delve deeper into the specific linguistic nuances of each profession, we will next focus on words unique to each field, providing a more distinctive perspective on the language used within a particular domain.

### 4.2.2 The Most Frequent Unique n-grams

We will extract the most common unigrams, bigrams, and trigrams that appeared in only one of the professions in this section. By “unique,” we mean “exclusive within a domain.” Indeed, analyzing unique n-grams within each profession allows us to capture the specific linguistic patterns and themes that characterize the discourse in each field.

Figures 4.3, 4.4, and 4.5 show unique unigrams, bigrams, and trigrams, respectively.

Words like “oberlin”, “dartmouth”, and “notre dame”, are locations where addresses were delivered. “new zealand” was probably used multiple times by Jacinda Ardern, the 40th prime minister of New Zealand.

Key terms in Academia, such as “hbcus”, “underfund”, and “minority serve institution” indicate the frequent discussion of historically black colleges and universities (HBCUs), the challenges of funding in academic institutions, and the importance of educational

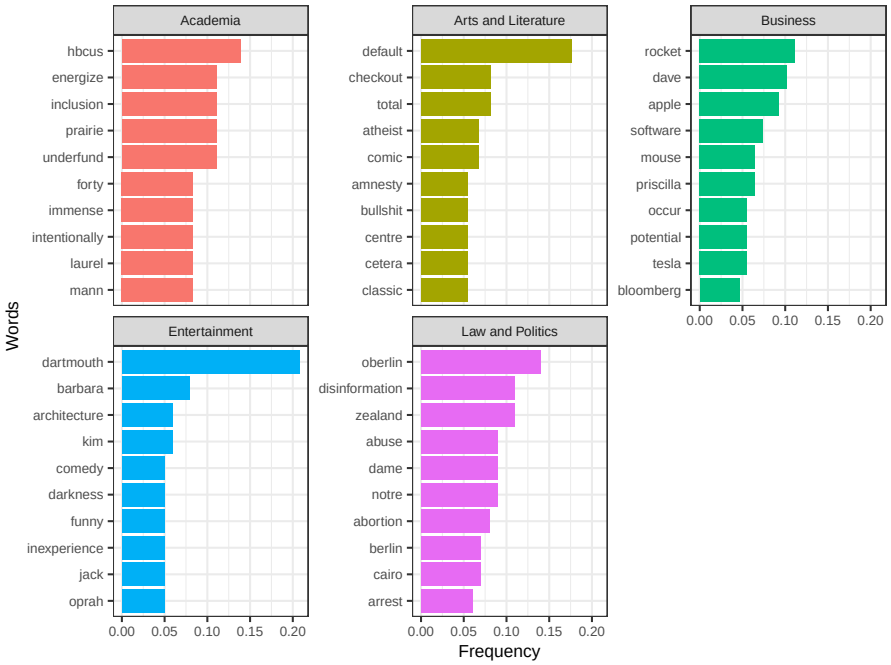



Figure 4.3: Top 10 unique unigrams in each field

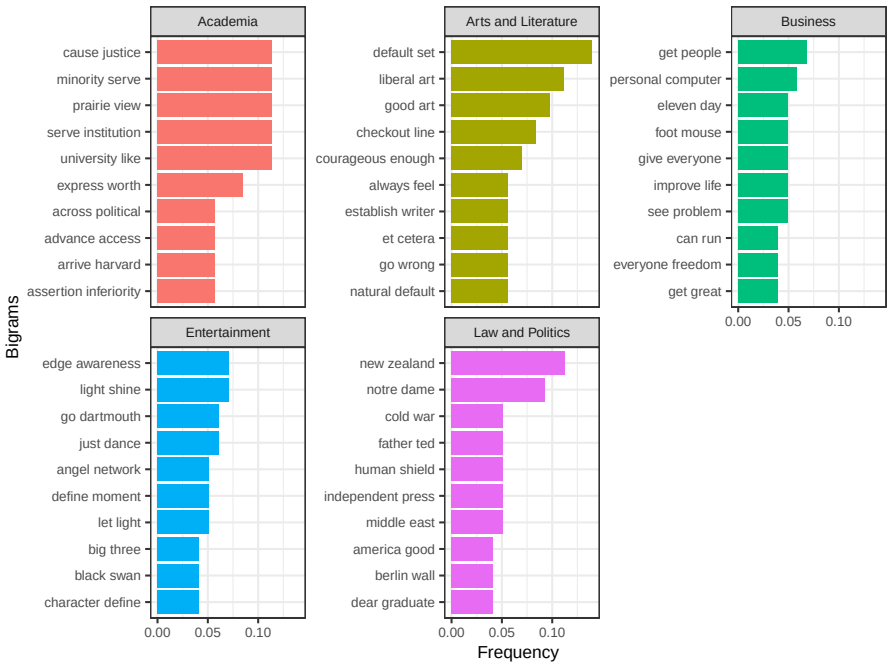



Figure 4.4: Top 10 unique bigrams in each field

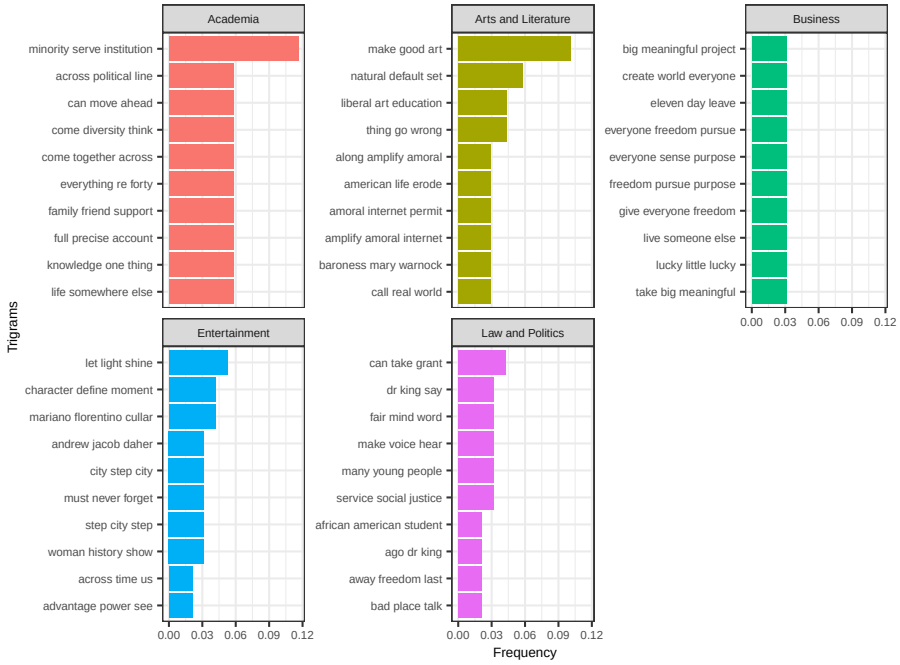



Figure 4.5: Top 10 unique trigrams in each field

services for minority students.

The phrase “make good art” and the word “default” in Arts and Literature may suggest a strong emphasis on the creation of quality artwork and challenging the standard or conventional ways of creating art.

Words like “rocket” and phrases such as “big meaningful project” or “give everyone freedom” in the Business field suggest that speeches in the business sector often focus on innovation (symbolized by “rocket”), the importance of undertaking significant projects, and individual freedom.

Speakers in the field of Entertainment used phrases like “edge awareness”, “let light shine”, and “character define moment”. It shows that speeches in the entertainment sector often revolve around the importance of self-awareness, personal enlightenment, and the influence of defining moments on a person’s character.

Phrases like “can take grant”, “dr king say”, and terms like “cold war” in Law and Politics, show that political speeches often refer to historical events and figures, like Dr. Martin Luther King Jr., to discuss civil rights, freedom, and historical context, including international relations.

While the unique phrases and words have given us a detailed view of the distinct topics discussed in each profession, it is equally interesting to observe their common ground. Some words and phrases cut across all fields, revealing universal themes and concerns. The following section will delve into these shared elements that connect Academia, Arts and Literature, Business, Entertainment, and Law and Politics. The exploration of these commonalities will offer us insight into the collective narrative.

### 4.2.3 Common unigrams, bigrams and trigrams

#### Common unigrams

We can select the words that appeared in all the professions by filtering out the unigrams that appeared in every profession. These unigrams are shown in Figure 4.6.


Figure 4.6: Top Common Unigrams

Words like “get”, “go”, “make”, “work”, “know” and “people” which were in the top frequent words, also appeared in the top common unigrams in all corpora; this observation is interesting because it suggests that despite the diversity in backgrounds and professions of the speakers, there is a commonality in the language used to express their thoughts. Words like “get”, “go”, “make”, “work”, and “know” are action-oriented and reflect a shared focus on practicality, progress, and understanding. The speakers from diverse professions and backgrounds may focus on different aspects of the human experience, but the

term “people” brings universality to their messages.

### Common bigrams

The same process generates the common bigrams that appeared in all professions. The common bigrams are depicted in figure 4.7.

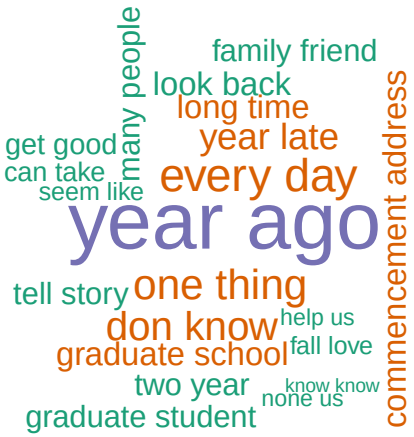



Figure 4.7: Top Common Bigrams

Common bigrams are pairs of words that appear in speeches from all professional sectors. These shared bigrams can help us understand the shared vocabulary and ideas that cross professional boundaries in these talks. Given the context of commencement speeches, it is not surprising that the terms “commencement_address”, “family_friend”, “graduate_school” and “graduate_student” appeared frequently. They could be used when addressing the audience directly or discussing the speakers’ experiences.

Storytelling (“tell_story”) is a typical style in speeches to make the content more engaging and relatable. “year_ago”, “long_time”, “look_back” and “year_late” suggest that the speeches frequently reflect on past experiences and possibly discuss changes and developments over time.

“help_us” and “don_know”[1] seem more related to the concept of wisdom. However,

[1]The original phrase is “don’t know”. The ” ’ ” and “t” were removed in remove punctuation and stopwords removal preprocessing stages.

Table 4.2: The Window for don_know

| Document_id | word_context |
|---|---|
| Adichie_54 | Be courageous enough to say I don't know. This might be harder to do with everyone calling you 'Harvard'. But ignorance acknowledged is an opportunity; ignorance denied is |
| Adichie_54 | ignorance denied is a closed door, and it takes courage to admit to the truth of what you do not know. |
| Engel_10 | I didn't let it bother me. Sometimes you have to roll with it, put yourself in situations where you don't know what's going on around you and let your brain sort it out. That's the fun part: the constant learning, the |
| Portman_14 | be trusted. But I think that contradiction can be reconciled and is in fact instructive. Achievement is wonderful when you know why you're doing it. And when you don't know, it can be a terrible trap. |
| Portman_14 | reconciled and is in fact instructive. Achievement is wonderful when you know why you're doing it. And when you don't know, it can be a terrible trap. |

the actual meaning of these bigrams can depend heavily on their context within each individual speech. The further qualitative analysis would help to confirm and expand upon these initial interpretations. To do this, we may extract all instances of these bigrams and choose some of the most relevant contents to analyze. We create a 40 words window around keywords of interest in order to grasp the context and meaning of how the term is being utilized.

Table 4.2 shows the context where "don_know" bigram were used. Chimamanda Ngozi Adichie's lecture highlights the need to accept ignorance, which, according to Meacham (1983), is a vital part of wisdom. She encourages the audience to be brave enough to state, "I don't know", implying that wisdom frequently entails acknowledging one's own limitations.

Engel's lecture highlights the significance of confronting unknown situations in order to learn and grow. This demonstrates a wise perspective on life, recognizing the value of experience and learning even in confusing or challenging situations, which links with Webster's concept of grasping life's unpredictable nature as a key to wisdom.

This part of Natalie Portman's speech is connected with the balance theory of wisdom. It discusses the importance of understanding why you are pursuing achievement, indicating the wisdom in aligning actions with values. Therefore, acting and not knowing

the values, might not be considered a wise action, even if one gains excellent success.

**Common trigrams:**

Table 4.3: Trigrams appeared in at least three professions

| Trigrams |
|---|
| president tessier lavigne |
| world war ii |
| deliver commencement address |
| every day life |
| board overseer member |
| happen right now |
| many year ago |
| take deep breath |
| can still remember |
| every single day |
| make difference life |

All professions shared no trigram. However, the common trigrams in at least three professions from our dataset, represented in Table 4.3, provide another layer of insight into the recurring themes and specific phrases that the speakers tend to use. Some possible interpretations are as follows:

“Deliver commencement address”: This phrase directly refers to the act of delivering the speeches that are being analyzed. It is not surprising to find this as a common trigram among addresses. “President Tessier Lavigne” and “Board overseer member”: These phrases are related to the administration of the universities where the speeches are given. Particularly Marc Tessier-Lavigne, the president of Stanford University.

“World War II”, as the historical event, might be used as a reference point for lessons learned, historical context, or comparisons to current events.

“Many years ago,” typically introduces a reflection on the past, whether personal memories or historical events.

“Every day life” & “Every single day”: These trigrams suggest a focus on daily routines or habits, implying that the speakers might be providing advice or insights that can be applied regularly.

“Happen right now” may suggest a focus on current events or recent developments, indicating that the speeches are responsive to contemporary issues.

Table 4.4: The Window for make_difference_life phrase

| Document_id | word_context |
|---|---|
| BarackObama_37 | as we wish to be treated. the call to love. to serve. to do what we can to make a difference in the lives of those with whom we share the same brief moment on this earth. |
| Jobs_9 | — your gut, destiny, life, karma, whatever. this approach has never let me down, and it has made all the difference in my life. |
| Winfrey_6 | would like to be a journalist. i would like to tell other people's stories in a way that makes a difference in their lives and the world." and as those words were coming out of my mouth i went whoa! this |
| Winfrey_6 | mouth i went whoa! this is pretty good! i would like to be a journalist. i want to make a difference. well i was on television by the time i was 19 years old. and in 1986 i launched my own |

"Take deep breath" could be a piece of advice given to the audience, often suggesting a moment of pause or reflection.

"Can still remember" is frequently used before the speaker shares a personal experience or reminiscence.

"Make difference life" is likely part of a larger discussion about impact, purpose, and the value of contributing positively to the world or to others' lives.

Similar to common bigrams, context is crucial when interpreting these trigrams, especially "Make difference life". While these interpretations can provide a general idea of the shared language and themes, the actual meaning will depend on each speech's surrounding text. Therefore, we can take a look at the context around this specific trigram.

The windows around the trigram "Make a difference in life" that appeared in "Law and Politics", "Business", and "Entertainment" (table 4.4) provide some good material for qualitative interpretation. Looking at the context can help us understand how this phrase is used across speeches.

Obama uses the phrase to speak about love, service, and making a difference in others' lives. This is framed as a call to action for his audience, encouraging them to positively impact the world during their time on Earth. Obama leverages the phrase to emphasize altruism and effecting a change in the lives of others. This sentiment aligns with the views of McKenna, Karami, and Nonanka.

In his famous Stanford commencement address, Steve Jobs refers to his own life and how following his intuition and passion has made a difference in it. Here, the phrase is more inwardly focused, reflecting on personal growth and fulfillment. He underscores intuition's role, reflecting McKenna's definition of wisdom, which values non-rational and subjective elements when making decisions. However, based on the excerpts we are analyzing, Jobs does not explicitly advocate for a balanced approach to addressing intrapersonal, interpersonal, and extrapersonal interests, as Sternberg's definition of wisdom would suggest. Jobs' primary focus seems to lie more on personal passion.

Oprah Gail Winfrey speaks about her aspiration to make a difference in people's lives aligns with Nonaka's perspective that wisdom is crucial for employing political power to initiate action; Oprah Winfrey discusses her ambition to positively influence people's lives and the world via journalism as her [political] power.

We can see from these examples that the term "make a difference in life" is used to refer to making a difference in the lives of others (in the sense of service or influence) as well as one's own (in the sense of personal growth or fulfillment). This implies that the speakers are pushing graduates to pursue both personal fulfillment and a more significant societal effect in their future endeavors.

Having completed our analysis of word frequencies and common and unique n-grams, we have gained a valuable understanding of our corpus's keywords and phrases. However, the frequency of certain words or phrases does not necessarily reveal the emotional tone or sentiment of the content. Therefore, it is essential to expand our analysis to investigate the underlying sentiments carried by speakers. This leads us to the next crucial aspect of our text analysis: sentiment analysis.

## 4.3 Sentiment Analysis

Before the sentiment analysis, we must create a customized sentiment lexicon that suits our data. This is an important step in sentiment analysis, allowing us to refine our tools to fit our needs better. We have identified words that potentially influence the sentiment score in a way that might not be accurate or useful for our specific analysis. Terms such as "obama", "musk", "gates", "jobs", "harvard", and "stanford" might be associated with positive sentiments due to the prestige of these individuals or institutions. But in the context of our specific analysis of speeches, these are likely just neutral references to individuals or places. Thus, should not be considered as influencing the sentiment of the

speech. So, we have set their sentiment scores to neutral (0) in our customized lexicon list.

We are now ready to execute sentiment analysis after customizing the lexicon for our data. We do a sentence-level sentiment analysis on the original transcripts using a mix of the `get_sentences()` and `sentiment_by()` functions. That is, we do not employ n-grams to generate sentiment scores. We preserve the speeches' original context and semantic structure by not using n-grams to produce sentiment scores, which can be critical for properly identifying sentiment. Sentence-level sentiment analysis is frequently more accurate than n-gram-level sentiment analysis because it captures the context in which the words are used.

Valence shifters like negators (e.g., "not", "n't", "never") can change the polarity of a phrase, and `sentimentr`'s algorithm handles this by default. For example, the phrase "I am not afraid" would typically be given a positive sentiment score, even though "afraid" is a negative word, because the "not" negates it. In the `sentimentr` package, the `sentiment_by` function is designed to calculate text polarity sentiment at the sentence level quickly and considers negation.

The outcome may be divided into two sections. Analyzing the distribution of sentiment in each profession and investigating how sentiment varies during the course of speeches in every field.

### 4.3.1 Profession and Sentiment

To visualize distributions of data across different categories, we use The `geom_jitter()` function. It provides a little bit of random variation to each point's position. This can make the plot easier to read by reducing overlapping. On top of that, we overlay box plots to show the median, upper and lower quartiles. The result is shown in Figure 4.8. Each point on the plot represents a document, colored according to whether its sentiment is negative (red for less than -0.1), neutral (between -0.1 and 0.1, Gray), or positive (green for greater than 0.1).

The visualization presented in Figure 4.8 also illuminates the extreme sentiment values extracted from each profession. These values represent the documents associated with the most positive and negative sentiment scores. As mentioned above, the `geom_jitter()` function has been employed to improve the visibility of individual data points within each profession category. Consequently, the position of the labels might

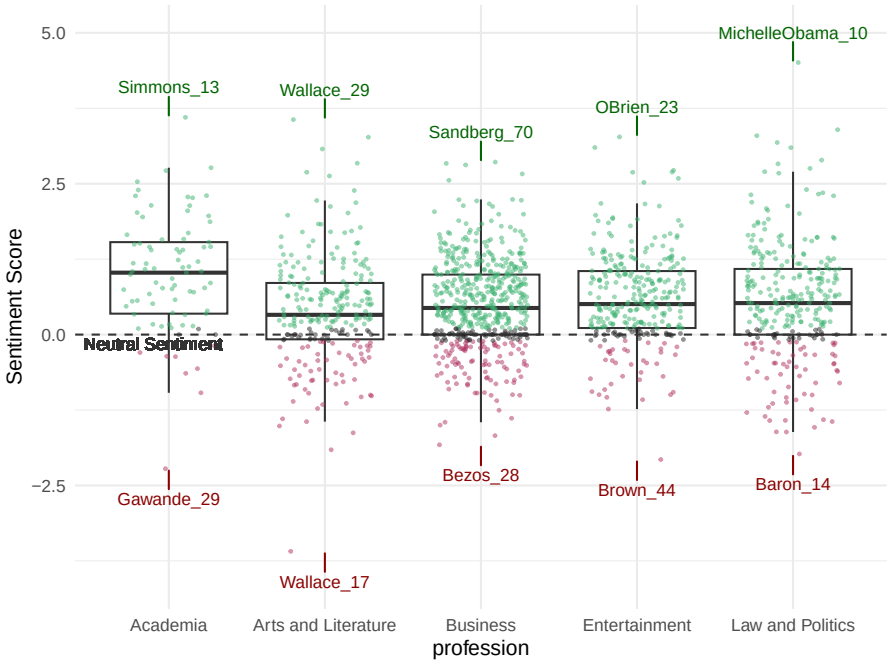



Figure 4.8: The distributions of sentiment across each field

not perfectly align with the respective points horizontally. However, given that the labeled points correspond to the extreme values within each profession, their identification with the corresponding labels should be straightforward.

Speeches from all professions show both negative and positive sentiments. However, the sentiment analysis of the addresses across many professional sectors demonstrates a positive trend. There does not seem to be a clear correlation between profession and sentiment score. Within the scope of the investigation, the academic discourse was the least negative in comparison to the other professional sectors. This outcome aligns with expectations, as the realm of academia often addresses various life facets, encompassing psychological elements like personal growth, self-esteem, and life management, and economic factors such as operational skills and the pursuit of economic success. The entertainment industry followed closely behind, with a lower level of negative sentiment. In contrast, professions such as arts and literature, as well as law and politics, were shown to have higher levels of negativity.

To enhance our comprehension of the context, we can extract the initial paragraphs that display intense positive or negative sentiments. Appendix IV showcases the original paragraphs corresponding to extremely positive and negative expressions across various professions.

In academia, the most positively rated paragraph (Simmons_13) celebrates Harvard's impact, highlighting the university's influence achieved through scholarly and professional contributions. This positivity for academia's role in societal advancement stands in contrast with the negative sentiment found in Gawande_29, which expresses uncertainty and anxiety about the impact of a human-made crisis. It captures the inherent uncertainty of the real world.

The very positive sentiment (Wallace_29) in the Arts and Literature area stems from a critique of modern civilization and a demand for real freedom. It aligns with several definitions of wisdom, emphasizing the importance of self-awareness, empathy, and a broader perspective that seeks the common good. For instance, the passage's emphasis on being 'truly able to care about other people and to sacrifice for them', aligns with Grossmann's definition of wisdom that wisdom balances self-interest with others values truth and cares for humanity.

This capacity is not simply to act based on one's self-interest but to consider the interests of others, a quality Sternberg suggests. However, when referring to "sacrificing for others", the passage seems to strongly emphasize interpersonal interests, potentially at the expense of intrapersonal ones. Therefore, the alignment proposed in the previous explanation indeed requires further clarification. The key to resolving this apparent contradiction lies in adopting a broader, more nuanced understanding of "sacrifice."

Historically, numerous individuals have demonstrated such selflessness and commitment to a cause greater than themselves, such as the sacrifices made during civil rights movements worldwide. In these contexts, "sacrifice" takes on its most profound meaning, often involving the ultimate sacrifice of one's life. The decisions they made and the actions they took required an understanding of the complex dynamics they faced. Much like many other aspects of human experience, wisdom is inherently subjective and multifaceted. It is not a one-size-fits-all concept but rather a deeply individual process that relies on personal experiences, values, and circumstances.

In addition, it is essential to note that while these extreme sacrifices are illustrative of certain aspects of wisdom, they are not a prerequisite for wisdom. Wisdom can be presented in numerous ways and does not always require such extreme demonstrations. It can be displayed in daily actions and decisions. Wisdom does not mandate an always future-oriented perspective or a preference for the interests of future generations over one's own life. Rather, it requires awareness and consideration of these longer-term

effects, even if the final decision ultimately prioritizes immediate or personal needs.

Therefore, “sacrifice” does not necessarily imply a severe loss to oneself (intrapersonal interest). Instead, it can be perceived as acts of kindness, empathy, and understanding towards others (interpersonal interests) that might require one to deprioritize immediate personal desires occasionally. This balance reflects the essence of phronesis. While these acts may seem to be “sacrifices” in the short term, they could potentially contribute to long-term intrapersonal interests by fostering relationships, personal growth, and a sense of purpose.

Still, a more explicit discussion of how one balances these differing interests in the context of wisdom is necessary. It is a complex task, and, as noted by Sheppard, leaders often need to navigate paradoxes and make complex decisions that balance various aspects of these interests. While not explicitly discussed in this section of Wallace’s speech, this delicate balance is inherent in many definitions of wisdom and is an integral part of navigating life in a wise manner.

On the other hand, the paragraph with the most negative feeling (Wallace_17) criticizes consumer society’s greed, particularly those who recklessly utilize fossil fuels without regard for future generations. He voiced his displeasure at wasteful conduct that contributes to climate change. It is interesting how he criticizes the unwise actions of people. The reckless consumption of fossil fuels may satisfy short-term intrapersonal interests. Still, it neglects interpersonal interests (current societal needs) and extrapersonal interests (the well-being of future generations and the planet).

In business, Sandberg_70, characterized by positive sentiment, reflects on the importance of gratitude, even in challenging times. The concept of practical wisdom as presented by Aristotle, involving the capacity to determine suitable behavior in specific circumstances, can be recognized in Sandberg’s adaptation to the situation of losing her spouse. In contrast, paragraph 35 of Bill Gates’ speech is a series of infectious diseases such as Malaria and hepatitis-B.

In entertainment, OBrien_23 offers practical advice to graduates in a positively scored paragraph, aiming to equip them for life after graduation. Brown_44, showing a negative sentiment, illustrates a critique of society’s fear of enlightenment and change.

In Law and Politics, the positive sentiment of MichelleObama_10 stems from an appreciation of community, commitment to service, and social justice. Contrasting this is

the negative sentiment in Baron_14, which expresses concern over threats to free expression and the importance of an independent press for a functional democracy. This depiction starkly contrasts with several elements of wisdom. For instance, both Sternberg's and Karami's definitions emphasize achieving a common good and positive ethical value. Suppressing free expression for the sake of personal power directly conflicts with these principles.

After having completed a comprehensive sentiment analysis of the speeches, evaluating positive, negative, and neutral sentiment scores, as well as identifying the most emotionally charged paragraphs, it is important to delve deeper into the emotional content of these dialogues. We have seen how different sentiments are expressed, but understanding the specific emotions may help us better understand the speaker's messages. The next phase of our investigation will focus on a detailed emotion analysis, exploring the presence and impact of particular emotions such as fear, anger, and trust.

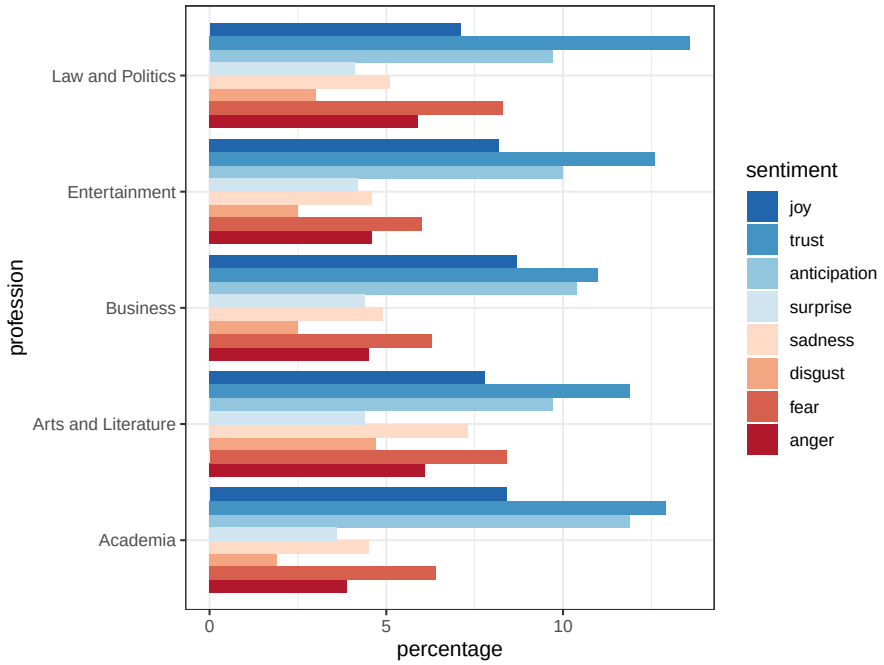



Figure 4.9: Emotional spectrum across professions

Figure 4.9 shows the spectrum of emotions across different professions.

Fear is often considered a negative feeling in the general human experience, with its relevance varying greatly from individual to individual based on unique circumstances, experiences, and cognitive-emotional frameworks. This subjectivity is vital to emphasize since emotions do not have universally set valences. However, fear appeared as a highly

prominent sentiment in our text-based study. Fear often signifies a threat or potential harm. It can be related to the wisdom dimension of “awareness and management of uncertainty” mentioned by the Berlin School, as recognizing fear and addressing it effectively requires an understanding of uncertainties and potential risks.

Trust, on the other hand, is among the most positive emotions and is highest in all professions. Trust can be associated with the “relativism of values” dimension of wisdom, as it often involves understanding and respecting diverse perspectives and values. It also relates to the “interpersonal” dimension outlined by Sternberg, emphasizing the importance of fostering healthy interpersonal relationships. In addition, trust implies a certain level of “moral maturity,” as mentioned by Karami, as it requires ethical behavior and a commitment to honesty and integrity.

Anticipation, another positive emotion, shows a forward-looking perspective. This is indicative of the “lifespan contextualism” facet of wisdom outlined by the Berlin School, as it requires a long-term perspective and understanding of the evolving context of life. Anticipation also suggests elements of Sternberg’s “adaptation to, shaping of, and selection of environments” dimension of wisdom, as it involves envisioning future scenarios.

Sadness, prevalent in Academia, Arts and Literature, and Business, often stems from experiences of loss or failure, which are inherent uncertainties in life. In this context, wisdom might involve acknowledging these feelings and learning to cope with such disappointments, enabling resilience and growth.

In Entertainment, the presence of both anger and sadness suggests a mix of emotional challenges. Anger can connect to the wisdom dimension of “relativism of values,” as it often arises from perceived injustices or personal or social norms violations. In this case, wisdom might involve understanding diverse perspectives and finding constructive ways to express and manage this strong emotion.

In Law and Politics, the prominent emotion of anger may arise from injustices or conflicts. Sternberg’s dimension of wisdom regarding achieving a common good by balancing interests also resonates here, as it may imply striving for justice and the greater good.

This emotion-wisdom connection offers insights into how different professional contexts might cultivate wisdom, demonstrating the links between emotional experiences and the various dimensions of wisdom. In the next section, we try to understand how the speakers conveyed their wisdom to graduates.

### 4.3.2 Timing and Sentiment

As we continue our investigation of wisdom leadership in speeches, it is critical to understand what is said and how it is presented. The tone of a speech might change, creating a cadence that keeps the audience's attention. In this part, we will look at the timing of emotions in speeches from various professions. We can get insights into how leaders plan their discourse for optimum impact by evaluating the sentiment evolution from the beginning to the end of a speech, perhaps adding new dimensions to our knowledge of wisdom leadership. The generated plot in 4.10 shows the results.

In this graph, the y-axis represents the average sentiment of each profession. Positive values indicate a generally positive sentiment, while negative values denote a negative sentiment. The x-axis symbolizes the sequence of paragraphs in percentage. It allows us to understand how the sentiment of speeches changes over the speeches' timeline.

The graph has been plotted using `geom_smooth` function. A smooth line (a line created by local regression using LOESS method (Cleveland, 1979)) is drawn for each profession to show the general trend of sentiment throughout the speeches. The span parameter in the LOESS function controls the degree of smoothing. Values of the span parameter typically range from 0.1 (indicating a highly local model where each regression is based on 10% of the data) to 1 (indicating a global model where each regression uses all data points). We have chosen `span = 0.8` to reduce noise and highlight larger patterns over capturing smaller, local variations.

The findings from the sentiment analysis indicate commonalities in the sentiment path across speeches from diverse professions. Generally, each discourse appears to begin with a positive sentiment. This pattern could potentially be attributed to the typical structure of commencement speeches, which often start with congratulations aimed at the graduating students and their families. However, subsequent to this initial phase, there is a noticeable reduction in sentiment. Except for speeches from the academic domain, a gradual decline in average sentiment can be observed towards the midpoint of these discourses. Interestingly, this decline is followed by an increase as the speakers approach their conclusion. This trend suggests a characteristic sentiment structure in commencement speeches, despite the variation in the professional backgrounds of the speakers.

When these trends are interpreted through the storytelling framework outlined by O'Hara (2014), it is clear that the speakers have effectively used storytelling tactics in their talks. These commencement addresses appear to follow a narrative arc consisting

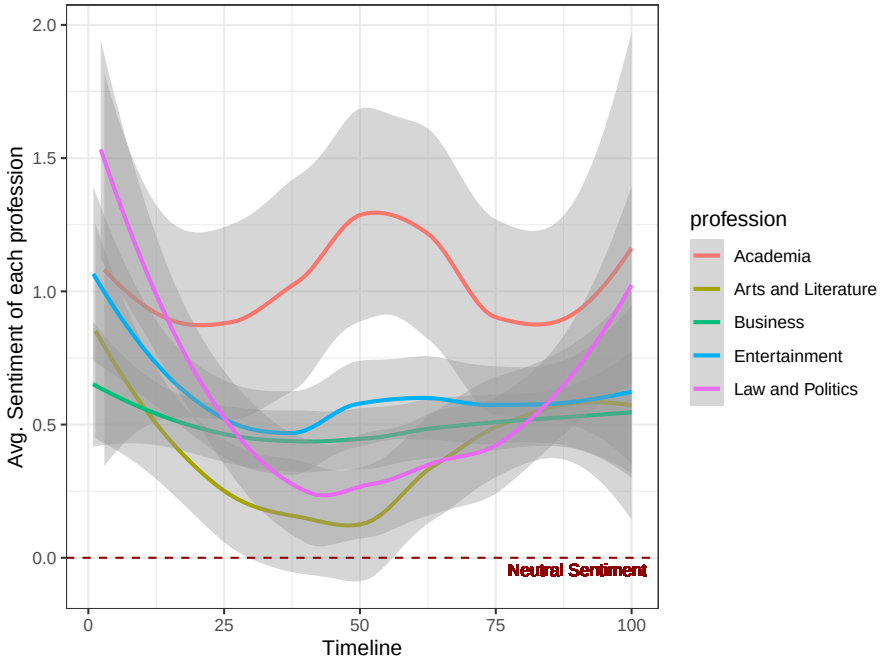



Figure 4.10: Sentiment of the speeches in each profession, from start of speech to end

of an introduction, a challenge, and a resolution, much like well-crafted stories.

In contrast to the other professional domains, speakers within the academic field demonstrate a distinct sentiment path. There appears to be a peak in positive sentiment around the midpoint of the speeches. A conservative interpretation of this anomaly could suggest that it might be related to the inclusion of more stories within the speech content. However, this is a preliminary interpretation that would require further investigation for substantiation.

The commencement speech, much like a story, is initiated with a message. This phase, marked by highly positive sentiments, often contains greetings, congratulations, and the establishment of a shared sense of achievement. As the speeches move into the challenging phase, sentiments become less positive. This could be reflective of the speakers sharing their personal experiences, struggles, and the lessons they learned. Thus, the speakers make the narrative more authentic and relatable to the audience.

Finally, the narrative arc is completed with the growing feelings near the end of the talks. This is the resolution phase, in which the initial message is reinforced, typically in conjunction with a call to action or transmission of wisdom to the graduates, therefore increasing engagement and connection between the speaker and the audience. This

narrative format, which mirrors the arc of a story, has the ability to increase audience engagement by allowing them to relate to the speaker's experiences, appreciate their insights, and retain the wisdom conveyed.

Whereas sentiment analysis helped us comprehend the emotional dynamics of the speeches, the following section will employ topic modeling techniques to delve deeper into their content, uncovering the principal themes that characterize these discourses.

## 4.4 Topic Modeling

After analyzing unique and popular terms across many professional disciplines and sentiment analysis, we now investigate further the thematic structure of the speeches. In this part, we use topic modeling approaches to find latent patterns in the talks. We begin with Latent Dirichlet Allocation (LDA), a popular topic modeling approach. We then go on to more sophisticated approaches, such as Structural Topic Modeling (STM), which add metadata to the study. Finally, we will employ `Top2Vec`, a newer approach using word embedding techniques to determine topics accurately. By taking these steps, we try to achieve a better understanding of the complexity of the discourse inside commencement addresses from various professions.

### 4.4.1 LDA

LDA is a popular topic modeling approach that assumes each text is a collection of topics, and each topic is a collection of words. As mentioned in 2.10.1, it uses an iterative process to assign topics to words and documents. We want to discover key topics in our collection of commencement addresses by using LDA, offering an essential understanding of the major themes.

There is no definite approach for determining the ideal number of topics for Latent Dirichlet Allocation (LDA). Several approaches, such as Perplexity and Topic Coherence, have been proposed and are widely utilized. First, we utilize the `FindTopicsNumber` function from `ldatuning` package (Murzintcev, 2021), which computes several metrics to assess the quality of LDA topics. Figure 4.11 depicts the results of this calculation. The results of the `FindTopicsNumber` function are displayed in Figure 4.11. This method determines the optimal number of topics by pinpointing where the curves of Griffiths2004 and Deveaud2014 reach their maximum values and where the curves of CaoJuan2009 and Arun2010 reach their minimum values. Upon analysis of all four curves, it is discernible

that five, six, and seven topics present the most favorable choices.

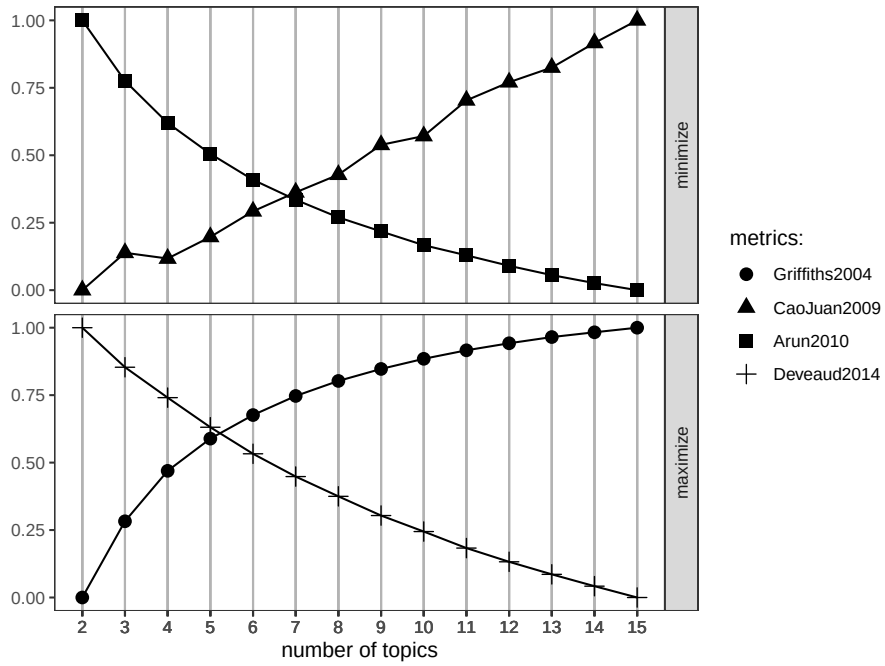



Figure 4.11: The optimum number of topics determined by `FindTopicsNumber` function.

Human judgment is still a significant aspect in determining the number of subjects for LDA topic modeling. The appropriate number of topics might be a judgment call depending on what makes the most sense for a certain application. As a result, we personally examined the results of various number of topics and selected five topics that give us the most meaningful results.

Figure 4.12 illustrates the top terms that define each topic. While the results show some themes in our corpora, interpreting the results is not very straightforward. To understand topics more, we apply NLP techniques like heat-map to demonstrate the distribution of topics over different professions, the relation between documents and topics, and sentiment analysis.

### Heatmap

For each profession in our data set, we may compute the mean of the document-topic probabilities (`posterior` probabilities from the LDA model). The values show the profession's average topic distribution. For example, according to the LDA model, topic 1 relates to 96% of business speeches. Figure 4.13 illustrates these proportions that represent the average degree to which each topic is included in the speeches of the respective

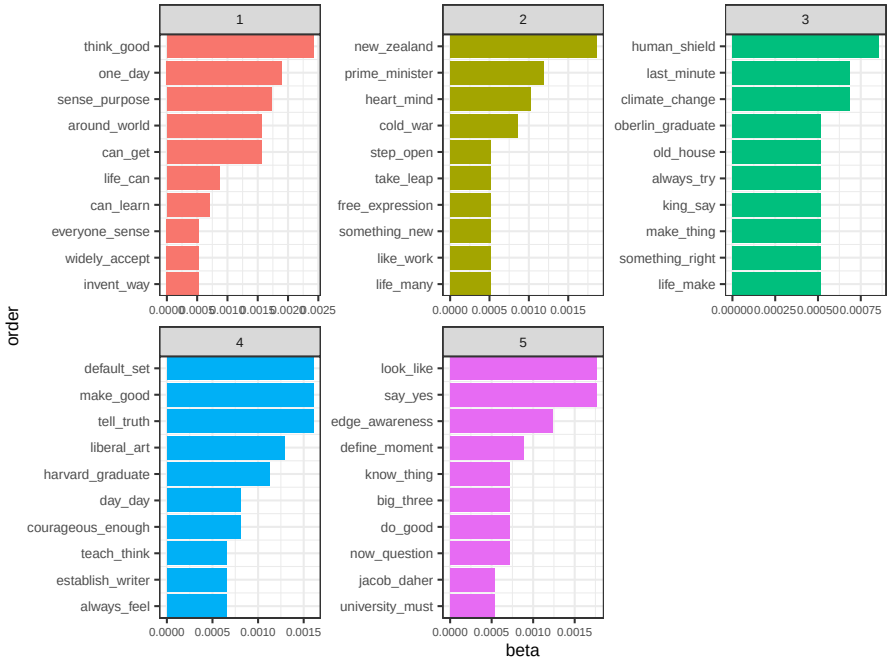



Figure 4.12: Top words that represent each topic.

profession. As it is seen, each profession covers every issue to some level, yet one or two themes dominate each group.

### Topics-Profession Network

For more in-depth analysis, we can extract the Topics Over Professions network. Or, in other words, professions-topics relations. Figure 4.14 reveals the relational network of topics and professions.

Topic 5, serving as a universal theme, appears in all professions except business, showing its broad relevance. In contrast, Topics 2 and 4 are distinctly associated with Law and Politics and Arts and Literature, respectively, revealing a more exclusive focus on their respective fields. The Business profession is remarkably influenced by Topic 1, suggesting that the key bigrams defining other topics may not significantly contribute to our interpretation in the business context. In the Entertainment field, Topic 5 predominates. Law and Politics professions demonstrate a close relationship with Topics 2 and 3, with almost equal representation of these themes. Arts and Literature speeches predominantly reflect Topic 4, but also incorporate elements of Topics 1 and 5. Within the Academic profession, the speeches seem to span a diverse range, with a stronger emphasis on Topic 5, followed by Topics 1 and 3, demonstrating the nature of academic

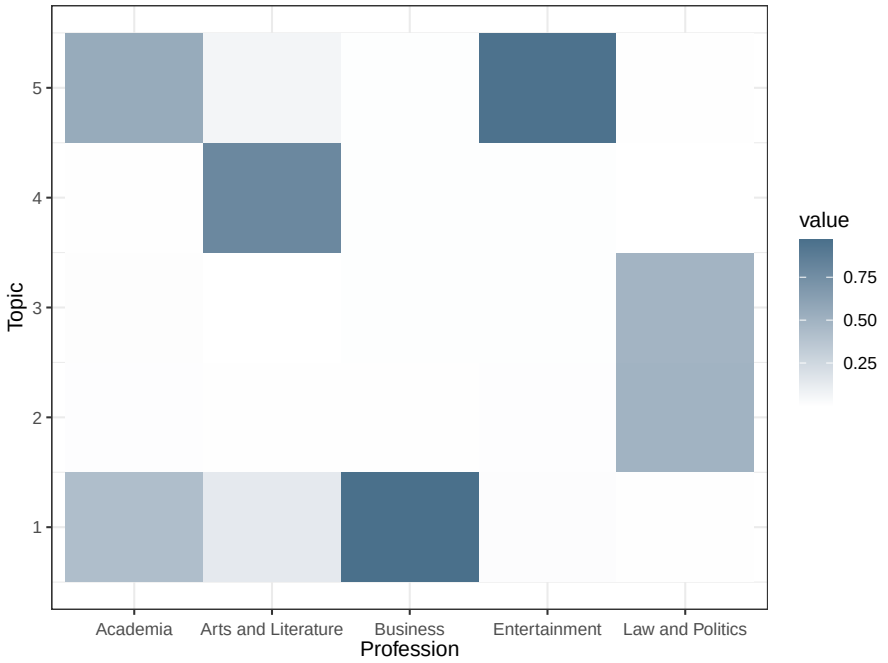



Figure 4.13: Distribution of topics over different fields

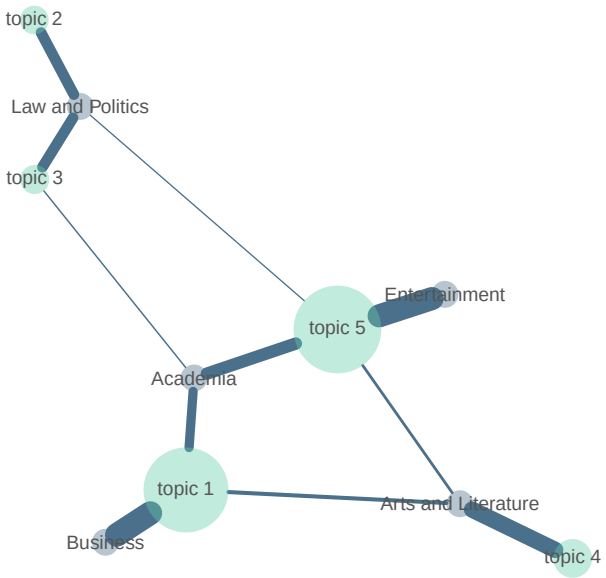



Figure 4.14: network of topics and professions

discourse. Table 4.5 displays the most probable document associated with each topic.

Table 4.5: Top document for each topic

| doc_id | profession | topic |
|---|---|---|
| Lewis_11 | Law and Politics | Political Considerations(2) |
| MichelleObama_11 | Law and Politics | Civil Rights and Global Issues(3) |
| Wallace_10 | Arts and Literature | Art and Truth(4) |
| Winfrey_10 | Entertainment | Personal Development and Awareness(5) |
| Zuckerberg_73 | Business | Inspiration and Purpose(1) |

The results of the Latent Dirichlet Allocation (LDA) indicate the presence of five distinctive topics. Using visualization techniques, we presented the relationship between the topics and professions.

Topic 1 has terms like “think_good”, “one_day”, “sense_purpose”, and “can_get”. These phrases suggest the encouragement of positive thinking, envisioning a better future, and having a clear purpose. This topic is particularly prevalent in speeches from the Business and Academia professions. We can label this topic as “Inspiration and Purpose”.

Topic 2 includes terms like “new_zealand”, “notre_dame”. The noticeable emphasis on “new_zealand” is because it was the focus of Jacinda Ardern as the prime minister of New Zealand. And Barck Obama’s speech was delivered in “notre_dame”. “prime_minister”, “last_year”, and “independent_press” are other bigrams in this topic. Obviously, this topic is about the political and social context of the time when the speeches were delivered. Not surprisingly, this topic is dominant in speeches within the field of Law and Politics. We call this topic “Political Considerations”.

Topic 3 features terms such as “civil_right”, “dr_king”, “lead_us”, “human_shield”, and “climate_change”. It hints at the discussion of civil rights, leadership, and pressing global issues. This topic is equally represented in speeches from Law and Politics and Academia fields. We label it as “Civil Rights and Global Issues”.

Topic 4 includes terms like “tell_truth”, “make_good”, “default_set”, “liberal_art”, and “good_art”. It suggests discussions around the importance of truth-telling, making good art, and the value of liberal arts. Unsurprisingly, this topic is most prevalent in speeches from the Arts and Literature profession. “Art and Truth” can be a good label for topic 4.

Finally, Topic 5, with terms such as “say_yes”, “look_like”, “pay_attention”, “light_shine”,

can hold the label of “Personal Development and Awareness”. It predominantly appears in speeches from the Entertainment field. This topic seems to encourage personal development, paying attention to the world, and allowing one’s “light” to shine.

In conclusion, each topic offers unique insights into the themes of different professions. To continue our investigations, we will analyze the sentiment of speeches to understand how the speakers convey their wisdom to the audience.

### 4.4.2 STM

As mentioned in section 2.10.3, metadata can be entered in the topic model in STM. “Metadata covariates for topical prevalence allow the observed metadata to affect the frequency with which a topic is discussed (Roberts et al., 2019).” The powerful `stm` function is used for topic modeling and “profession” as the topical prevalence.

#### Number of topics selection

Table 4.6: Exclusivity and semantic coherence of topics considering number of topics

| Topics | Exclusivity | SemanticCoherence |
|---|---|---|
| 4 | 9.04430321559016 | -254.863395848544 |
| 5 | 8.9973350140803 | -258.137172796541 |
| 6 | 9.22461123413171 | -246.623344229249 |
| 7 | 9.22655682811675 | -254.913690543226 |
| 8 | 9.23430467468784 | -257.221076094765 |

*Exclusivity* and *semantic coherence* are key components of evaluating the quality of topics generated by STM. In the context of topic modeling, exclusivity is a measure that helps identify words that are most exclusive or unique to each topic. High exclusivity shows that a word is used frequently in a given topic but rarely used in other topics. This helps to distinguish one topic from another. A topic is considered semantically coherent if its top words are related or tend to co-occur in the same context, which usually makes the topic easier to interpret.

When applying STM to our dataset, we try to achieve high exclusivity and semantic coherence for each number of topics. This would indicate that the topics are distinct from each other, and that each topic represents a clear, understandable concept or theme. However, there’s often a trade-off between exclusivity and semantic coherence in topic modeling. Increasing one may decrease the other, so it is important to find a balance that optimizes the interpretability of the topics. Table 4.6 shows the exclusivity and semantic

coherence of topics considering the number of topics. These considerations lead to the decision to pick a seven-topic model.

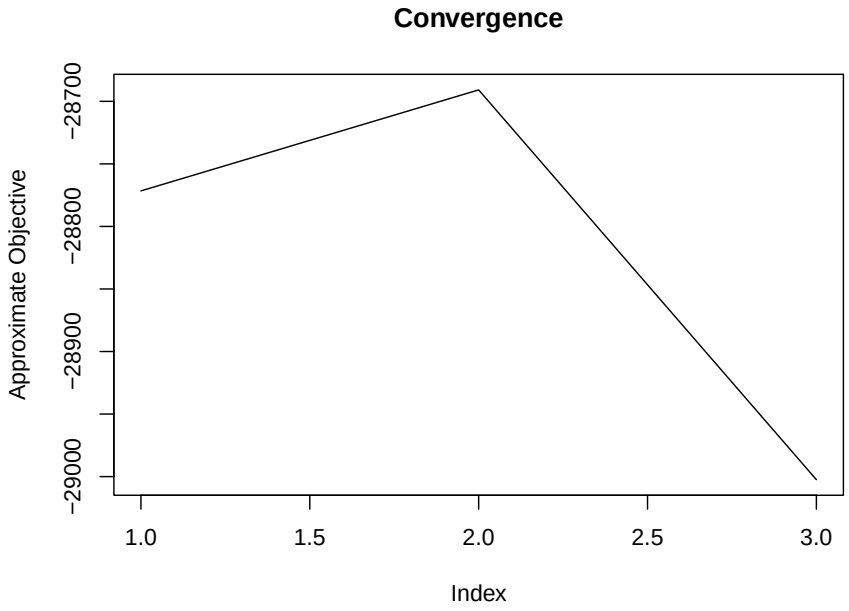



Figure 4.15: Lack of convergence for the model with 5 topics

Convergence is another essential element in the context of topic modeling. When the model converges, it means that it has found a stable solution where subsequent iterations in the estimation process do not change the outcome. If a model does not converge, it suggests that it might not be a good fit for the data or could still be improved with additional iterations. For example, Figure 4.15 depicts the reality that if we choose five topics for our model, even with a high number of iterations, convergence will not happen. A non-converging model can produce unreliable and unstable topic assignments, undermining our findings' validity. Figure 4.16 shows the fact that the seven-topic model converged. Therefore, it is a stable and reliable model for our data.

### Topics correlation

The inter-topic correlation for the derived topics was explored through the `topicCorr()` function in the STM package, and the results were visualized using the `plot.topicCorr()` function. The resulting plot (4.17) revealed no significant correlation between the seven distinct topics, suggesting that each topic captured a unique theme in the dataset.

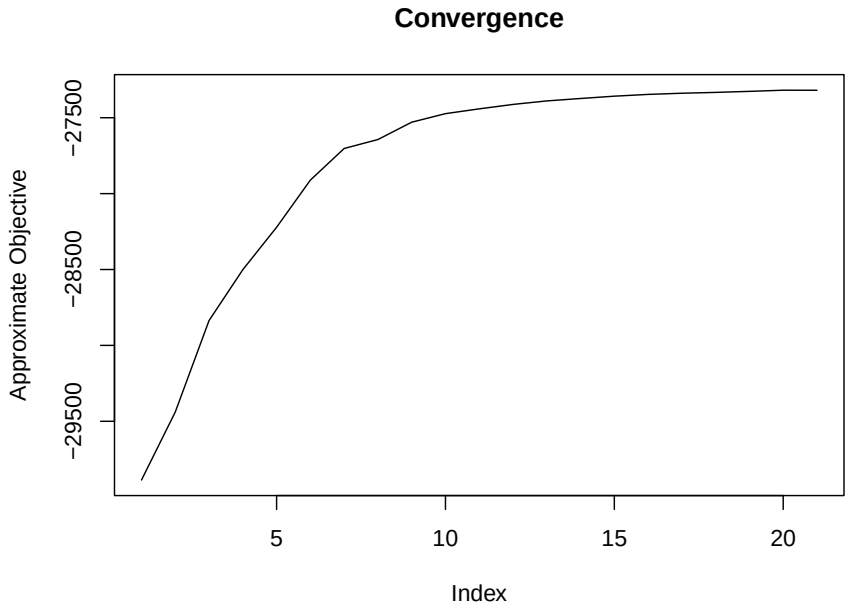



Figure 4.16: Convergence of the model with 7 topics

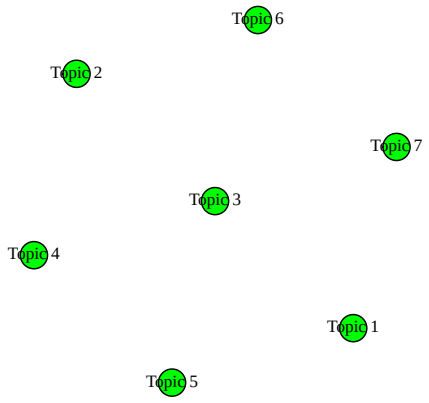



Figure 4.17: Correlation between topics

Table 4.7: Words associated with each topic

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| unite_state | heart_mind | year_ago | make_good | also_read | tell_truth | foot_mouse |
| year_late | human_being | last_year | high_school | take_grant | social_medium | hungry_stay |
| one_day | outside_world | human_right | every_single | say_yes | new_zealand | stay_foolish |
| want_make | don_know | let_us | didn_know | graduate_class | thank_much | stay_hungry |
| harvard_graduate | change_world | berlin_wall | go_college | can_take | today_want | go_back |
| get_one | president_faust | dr_king | look_back | silicon_valley | default_set | president_tessier |
| commencement_address | years_old | climate_change | even_though | full_transcript | think_good | tessier_lavigne |
| tell_story | find_someone | don_get | teach_think | along_way | notre_dame | personal_computer |
| president_unite | florentino_cullar | independent_press | thing_happen | world_need | want_talk | know_go |
| make_world | know_know | make_mistake | sense_purpose | take_responsibility | public_service | try_figure |

These topics provide insights into wisdom and leadership through themes such as learning from past experiences, addressing global challenges, fostering collaboration, personal growth, and resilience. Each topic emphasizes different aspects of wisdom and leadership. However, an interpretation of the topics extracted from STM topic modeling of commencement addresses is more valuable when we consider the context of defined topics. Consequently, we have extracted the top two documents associated with each topic in Appendix V.

Topic 1:

It is evident that Oprah Winfrey effectively incorporates various dimensions of wisdom into her narrative. One notable aspect is her wisdom of learning from failure. She persuasively argues that failure is not an end but a redirection. This notion aligns with Grossmann's idea of wisdom as recognizing and managing uncertainty. She acknowledges the presence of uncertainty, setbacks, and disappointments but encourages learning and growth from these experiences, an essential component of wisdom, according to Webster.

Winfrey also emphasizes the importance of an internal moral and emotional GPS, essentially advocating for self-regulation and moral maturity. This idea resonates strongly with Karami's wisdom dimension, which highlights the application of successful intelligence and virtues in pursuit of a common good. This moral compass serves as a guiding principle for ethical behavior and decision-making, reflecting Karami's emphasis on self-regulation and moral maturity as elements of wisdom.

The story of the Angel Network exemplifies Nonaka's dimension of wisdom that in-

volves the altruistic use of power. Oprah Winfrey employed her influence not for personal gain but to mobilize resources for the public good, manifesting the principle of altruism and the benevolent use of political power that Nonaka underscores.

Topic 2:

Zakaria's address includes various dimensions of wisdom while he emphasizes the integration of heart and mind, self-reflection, ethical understanding, and societal transformation.

To begin with, the appeal to "heart and mind" echoes Karami's view of wisdom as involving both affective (heart) and cognitive (mind) components. According to Zakaria, the qualities honored by human beings are those that integrate intelligence and emotion. "Intelligence, hard work, discipline, courage, loyalty and, perhaps above all, love and a generosity of spirit" are the qualities, he proposes, that can lead not just to a successful life (rewarded by the 'outside world') but also a fulfilled life (appreciated by those who know us best). It resonates with McKenna's studies that have noted the importance of "seeking intrinsic personal and social rewards."

Moreover, Zakaria's advice to "Trust yourself; you know what you should do" reflects Grossmann's view of wisdom as a form of self-reflection and self-understanding. The notion that one "doesn't need an ethics course to know what you shouldn't do" underscores that wisdom includes an innate understanding of one's ethical and moral boundaries. Another possible explanation for this is that he advises graduates to consider "non-rational and subjective elements when making decisions," parallel to McKenna's point of view.

Further, the final encouragement, where he urges the listeners that by living such a life, they will "change the world", resonates with Nonaka's idea of wisdom as a catalyst for societal transformation. The assumption here is that by embodying these virtues, individuals can influence the world positively, reinforcing Nonaka's understanding of wisdom as a force for the common good.

Topic 3:

Merkel_1 is Angela Merkel's short story about East Germany during the Cold War. However, this simple story is indeed a profound reflection of her life experiences and showcases various dimensions of wisdom when analyzed carefully. It is not only about her personal journey but also about her learning, adapting, and using her wisdom for the

betterment of her country and society.

As she reflected on her past experiences, Merkel exhibited the ability to recognize her own limitations and the uncertainty of her situation - an essential facet of wisdom, according to Meacham and Taranto. She had to acknowledge that she did not know when the oppression would end or when she would be able to live in freedom.

Her narrative illustrates her empathy for the people who were suffering, which aligns with Ardelt's definition of wisdom involving affective traits like empathy and concern for the well-being of others.

Furthermore, her experiences in East Germany during the Cold War, and her subsequent reflections upon them, have shaped her into a leader who prioritizes the common good, human rights, and freedom, all aligning with Nonaka's concept of wisdom.

In terms of the topic modeling words, phrases such as "year_ago", "last_year", "human_right", "berlin_wall", "don_get", and "make_mistake" are all interconnected with Merkel's recounting of her experiences during the Cold War. Her reference to the "Berlin Wall" and the violation of "human rights" in East Germany aligns with the values she upheld and the challenges she faced. Her phrase, "I don't know how often I thought that I just couldn't take it anymore" captures the sentiment of "don_get", highlighting the emotional and psychological strain she endured.

Topic 4:

Spielberg_4 presents several dimensions of wisdom. Spielberg talks about life as a series of character-defining moments, illustrating the concept of "reflection," as emphasized by Webster in understanding life's dialectical and uncertain nature. He reflects on his own journey, which is a critical aspect of Ardelt's "reflective" dimension of wisdom that underscores self-awareness.

The narrative also aligns with Meacham's interpretation of Socratic wisdom, where acknowledging one's ignorance or uncertainty is seen as a type of wisdom. Spielberg admits that while he knew what he wanted to do at the age of 18, he did not know who he was, suggesting an understanding of his own limitations and the acceptance of his evolving identity.

Spielberg's speech intertwines with the topic-associated words from your STM topic

modeling. The narrative shares themes of self-discovery and evolution (“didn’t know”, “go_college”), reflective moments (“look_back”, “every_single”), recognition of unique perspective (“teach_think”), and the importance of making morally grounded decisions (“make_good”), which underpin the various dimensions of wisdom.

Topic 5:

Oprah Winfrey’s address carries the essence of several dimensions of wisdom as defined by various scholars. Winfrey encourages Harvard graduates to utilize their skills and knowledge to achieve a common good, aligning with Sternberg’s concept of wisdom as the balance of interpersonal and extrapersonal interests.

Her storytelling about Michael Stolzenberg is an example for the wisdom dimension defined by Karami et al. – the adequate use of knowledge, intelligence, creativity, self-regulation, and moral maturity to solve critical problems. Michael’s resilience and determination to help others after overcoming his personal tragedy exemplify these characteristics, while his actions underline Grossmann’s definition of wisdom as morally-grounded excellence in social-cognitive processing. This story also resonates with McKenna’s viewpoint of wisdom as valuing humane and virtuous outcomes in decision-making, and “being practical and oriented towards everyday life.”

The narrative also matches the topic-associated words from our STM topic modeling. For instance, the phrase “along_[the ]way” ties in with Winfrey’s mention of challenges or setbacks that the graduates may encounter. “World_need” aligns with her encouragement to pursue what makes them come alive, as the world needs such individuals. The act of Michael taking responsibility to help fellow amputees resonates with “take_responsibility”.

Topic 6:

Simmons’s address displays some dimensions of wisdom as discussed by various scholars. She underscores the importance of institutional responsibility, historical awareness, and societal engagement, all vital elements of wisdom.

She speaks about the responsibility of Harvard, not just to be relaxed in its achievements but to address the disparities. This is close to Sternberg’s definition of wisdom, where he posits that wisdom involves using skills and knowledge to achieve a common good by balancing different interests over the long term. Simmons’s call for Harvard to

use its influential status to address these disparities also aligns with Nonaka's conceptualization of wisdom as prioritizing the common good and creating societal value. This suggests an understanding of the importance of decisions for resource distribution and societal engagement, which mirrors the Berlin School's facet of rich procedural knowledge of life.

Simmons's recognition of the work of historically black colleges and universities (HBCUs) and other minority serving institutions, despite the systemic underfunding and isolation they have faced, aligns with Grossmann's view of wisdom. Grossmann regards wisdom as morally grounded excellence in social-cognitive processing, requiring an understanding of various contexts and perspectives and care for humanity.

Topic 7:

In Jobs' context, "Stay Hungry" does indeed suggest a continual thirst for learning and improvement, while "Stay Foolish" encourages risk-taking and embracing the unconventional or unexpected paths that life may present. Ardelt's wisdom dimensions include a cognitive component (knowledge application, critical thinking) and a reflective component (self-awareness, introspection). In "Stay Hungry. Stay Foolish.", there's an emphasis on continual learning (cognitive) and embracing unconventional paths or challenging the norm (reflective).

However, the phrase "Stay Hungry. Stay Foolish" inherently implies a focus on one's personal aspirations and continuous growth. This emphasis, while potentially fostering individual achievement, creativity, and resilience, doesn't explicitly address the balance of personal pursuits with the consideration and understanding of others' needs and the broader societal context, a dimension prominently featured in many of the wisdom definitions.

According to McKenna's, Sternberg's, Karami's, and Grossmann's definitions, wisdom necessarily involves moral judgment, altruism, an orientation toward a common good, and balancing self-interest with others' values and the care for humanity. Such wisdom facets advocate for an empathetic and ethical approach in one's actions, suggesting a concern for the needs and interests of others, not just oneself.

In contrast, the message "Stay Hungry. Stay Foolish" doesn't directly refer to these aspects of ethical decision-making, care for others, or an orientation toward a common good. The focus on personal goals and continual learning, while crucial in many respects,

may potentially overshadow the necessity of empathy, altruism, and moral consideration in one's actions. As such, it could be argued that while the phrase "Stay Hungry. Stay Foolish" inspires elements of wisdom, it may not contain the full spectrum of wisdom's dimensions as defined by various scholars.

So far, our analysis has been limited to examining the highest-scoring paragraph for each topic in our structural topic modeling (STM). However, our analysis can be expanded beyond this point using the `stmBrowser` package. This handy tool allows us to delve deeper into our data set by providing insight into other aspects of our documents. For example, we can evaluate each document's scores for all topics, not just the highest one. This can help us better understand the distribution of topics across our corpus. Furthermore, `stmBrowser` enables us to explore the semantic polarity of the documents, providing an understanding of the positive or negative sentiment associated with each topic. We can also explore the distribution of documents across different professions or speakers, providing us with richer and more nuanced insights into our data. The `stmBrowser` package is an instrumental tool in STM that allows us to enhance our topic modeling by offering a multifaceted examination of our documents.

The detailed results and interactive visualizations of the Structural Topic Modeling analysis can be found at the following link: https://khadjehali.com/javad/vis/stm, providing a comprehensive exploration of topic distribution, semantic polarity, and document classifications across various professions, speakers, and even locations.

### Heatmap

Given the varying number of commencement addresses delivered by individuals from each field, before we plot the distribution of topics across different professions, the normalization of the topic distribution is a necessary step. Without normalization, a profession with a higher number of addresses, like Business with 9, could falsely appear to have a higher representation of certain topics merely because of its larger number of transcripts. Normalization allows us to truly compare the topic focus across different professions, irrespective of the size of their representation in the dataset.

The distribution of topics across different professions and throughout the entire corpus, depicted in Figure 4.18, varies quite significantly. When examining the corpus as a whole, it becomes evident that Topic 6 dominates discussion across all professions, with the highest representation found in business, followed closely by arts and literature.

Meanwhile, Topic 4 is also significantly represented across all fields, particularly in arts and literature, business, and entertainment. Topics 1 and 7 have the least representation overall. This indicates that certain subjects, namely Topics 6 and 4, tend to resonate across various professions, while others are more specific to specific fields.

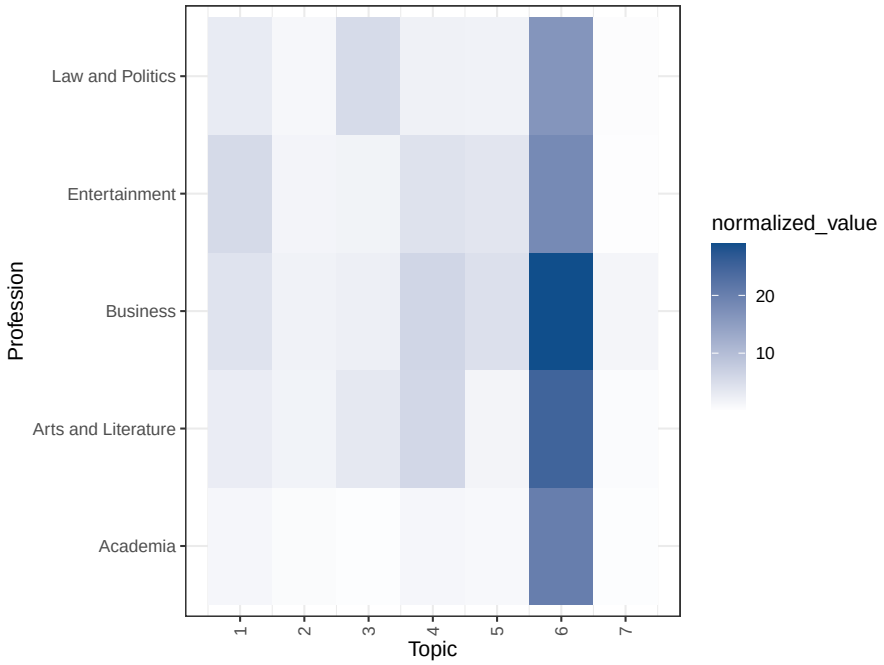



Figure 4.18: Distribution of topics across professions

We may now summarize our analysis of the seven topics spread among five professions, each of which offered different aspects of wisdom.

Analyzing the topic distributions across different professions sheds light on the unique characteristics and dominant themes evident in these discourse communities. The following discussion will interpret the topic distributions and consider how they reflect various dimensions of wisdom.

Topic 1, prominently present in all professions, was especially prevalent in the Entertainment sector, followed by Business and Academia. This topic is characterized by an emphasis on learning from failure, moral maturity, and the altruistic use of power, reflecting wisdom dimensions such as Grossmann's recognition and management of uncertainty, Karami's emphasis on self-regulation and moral maturity, and Nonaka's focus on altruism and benevolent use of political power. These characteristics suggest a universally recognized set of wisdom dimensions cutting across various fields, with Entertainment professionals potentially demonstrating a heightened sensitivity towards them

due to their broad audience and influential public roles.

Topic 2 appears to be most strongly associated with the field of Law and Politics, followed by Academia and Entertainment. This topic underscores the integration of heart and mind, self-reflection, ethical understanding, and societal transformation, reflecting wisdom dimensions from scholars such as Karami, Grossmann, and Nonaka. The strong representation in Law and Politics may stem from the inherent necessity of these wisdom dimensions in positions of governance and policymaking. The bigram "don-know" also features prominently within this topic, a common bigram that has been thoroughly investigated in the word frequency section of this chapter.

Topic 3, where Angela Merkel's narrative of resilience and empathy in East Germany exemplifies various wisdom dimensions, including recognizing one's limitations, empathy, and commitment to the common good, is dominated by Law and Politics. This result is unsurprising as such wisdom dimensions, captured by scholars like Meacham, Taranto, Ardelt, and Nonaka, are often deemed essential for leaders in this field.

Topic 4, which is dominated by Business and Arts and Literature, reflects Spielberg's journey of self-discovery, acceptance of uncertainty, and reflective life moments. This topic aligns with wisdom dimensions from Webster, Ardelt, Meacham, and Sternberg, possibly reflecting these professions' need for personal creativity and adaptability in rapidly evolving fields.

Topic 5, most prevalent in Business and Entertainment, emphasizes Winfrey's narrative of fulfilling the highest expression of oneself, utilizing skills for common good, and appreciating personal and social rewards. This topic encapsulates wisdom dimensions from scholars such as Sternberg, Karami, Grossmann, and McKenna, highlighting the need for personal development and societal contribution in these fields.

Topic 6, overwhelmingly represented in Business, resonates with Simmons' emphasis on institutional responsibility, historical awareness, and societal engagement, reflecting wisdom dimensions from Sternberg, Nonaka, and the Berlin School. This finding underscores the critical role of wisdom in leading and managing organizations effectively in a business context.

Topic 7, while comparatively underrepresented across all professions, demonstrates the wisdom dimensions inherent in Jobs' "Stay Hungry. Stay Foolish." Despite its emphasis on personal aspirations and continuous growth, it arguably lacks a broader con-

sideration of societal context, underscoring a potential limitation in representing the full spectrum of wisdom.

Overall, the topic distribution suggests that different wisdom dimensions may be more prominent or relevant in certain professional contexts. However, it also highlights the universal nature of wisdom, with all professions reflecting its various aspects to differing extents.

### 4.4.3 Top2Vec

Under the umbrella of topic modeling, we employ the innovative algorithm Top2Vec. This unique approach automatically extracts topics from the text data, not just by clustering words, but by leveraging their shared semantic meanings captured in word embeddings. These topics are semantically coherent and robust to the scale of the dataset.

In this section, we begin with the creation of Word2Vec embeddings, which lay the foundation for semantic relationships among words. Next, we turn to the construction of a 'Terms Network', a visualization tool that allows us to inspect these relationships. Following this, we delve into the extraction of Top2Vec topics, clusters of words that represent the unique themes in the corpus. Finally, we close this section by analysing the distribution of these topics across documents, using 'Top2Vec Document Vectors', providing a comprehensive view of the prevalence and significance of each theme. Each subsection represents a crucial step in our journey to uncover hidden thematic structures in relation to wisdom.

#### Terms Network

In the impending network analysis, we leverage the capabilities of word2vec to investigate the notion of wisdom as conveyed in commencement addresses. The word2vec algorithm, at its core, builds upon the premise that words frequently appearing together in the text (co-occurring) are semantically related. By converting these transcripts into high-dimensional vectors, word2vec deciphers hidden semantic relationships, thereby permitting us to depict the interconnections of these terms that condense the concept of wisdom.

Our chosen words are aimed at reflecting the multidimensional nature of wisdom as defined in various scholarly discourses. Our goal is to consolidate existing wisdom-related themes found in the literature while also being motivated to unveil potential new dimen-

sions that might expand our understanding of wisdom. Figure 4.19 serves as a visual representation of these connections, offering a visual platform for exploring the concept of wisdom within our data set.

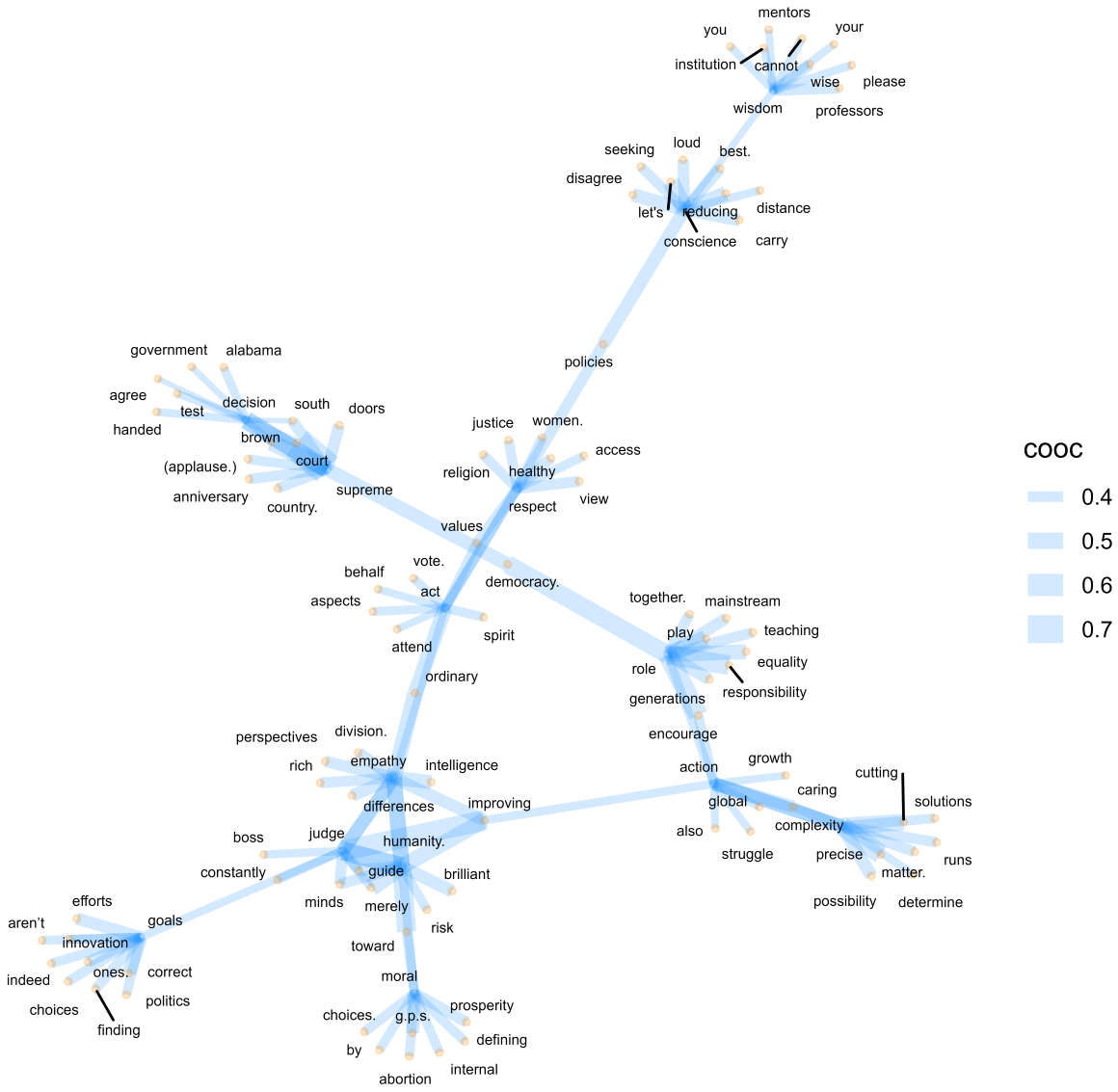



Figure 4.19: Word network generated using Word2Vec

This graph is a way of visualizing word embeddings (from the doc2vec model) and their similarity to each other. It is not a correlation graph in the statistical sense, and while it uses the concept of similarity (like cosine similarity), the exact similarity measure is determined by the doc2vec model's internal calculations. This kind of visualization can help to understand the semantic relationships among a set of words. The nodes (points) represent words and the edges (lines) represent the similarity between these words. The

similarity in this context is calculated by the doc2vec model.

We can analyse the results that might uncover additional insights about the nature of wisdom, by exploring the context (i.e., the specific paragraphs where these words appear in our data) of these nodes. Let's delve into the specific contexts of selected nodes to discern how they further contribute to our understanding of wisdom.

**Action:**

BillGates_64: "The complexity makes it hard to mark a path of action for everyone who cares and makes it hard for that caring to matter. Cutting through complexity to find solutions runs through four predictable stages: Determine a goal. Find the highest impact approach. Deliver the technology ideal for that approach and in the meantime use the best application of technology you already have."

**Moral:**

Hastings_4: "The changes to the world since Stanford was founded are breathtaking. The change rate over the rest of your lives will be exponentially higher, creating opportunity – as well as risk – for you and humanity. As the world speeds up, will our wax wings melt? Or will we bend the arc of the moral universe toward justice? To find answers, let's look to the past."

Rowling_43: "Of course, this is a power, like my brand of fictional magic, that is morally neutral. One might use such an ability to manipulate, or control, just as much as to understand or sympathise."

**Judge:**

BillGates_116: "I hope you will judge yourselves not on your professional accomplishments alone but also on how well you have addressed the world's deepest inequities. On how well you treated people a world away, who have nothing in common with you but their humanity."

**Conscience:**

Spielberg_15: “But make sure this empathy isn’t just something that you feel. Make it something you act upon. That means vote. Peaceably protest. Speak up for those who can’t and speak up for those who may be shouting but aren’t being hard. Let your conscience shout as loud as it wants if you’re using it in the service of others.”

BillGates_113: “And with that awareness, you likely also have an informed conscience that will torment you if you abandon these people whose lives you could change with modest effort.”

**Improving:**

Simmons_10: “I believe that each of us has a solemn duty to learn about and embrace that difference. That undertaking takes not a month or a year but a lifetime of concerted action to ensure that we are equipped to play a role in caring for and improving the world we inhabit together. This responsibility should encourage us to commit to our individual as well as professional role in advancing access, equality and mutual respect.”

Interpreting the results of the graph and the specific context in which these words are embedded reveals fascinating patterns that enrich our understanding of wisdom. The selected terms form a dense network of meanings depicting wisdom’s complexity and multifaceted nature, corroborating the richness and depth of our data selection. The network is organized around several central nodes, including “wisdom”, “moral”, “conscience”, “action”, and “judge.” These nodes are interconnected via paths that seem to suggest these concepts’ dynamic and interdependent nature in the discourse of wisdom.

The key terms and their interrelations in the semantic network clearly exhibit the components of Sternberg’s Balance Theory of Wisdom. The term “judge”, connected to both “humanity” and “goals”, embodies the idea of achieving a common good by balancing various interests. The ability to judge implies analytical and practical abilities and the capability to balance short and long-term goals while considering the broader implications for humanity. Furthermore, the term “moral”, associated with “humanity”, underscores the centrality of positive ethical values in wisdom, a primary principle of the Balance Theory.

The use of “moral” in the context of the collected speeches implies a value system that considers the welfare of others and the collective good of humanity.

Moreover, the term “action”, associated with “complexity” and “improving”, mirrors the Berlin School’s idea of wisdom as a pragmatic concept and judgment about complex and uncertain matters. The term, as used in the speeches, implies a practical understanding of complex situations and an active pursuit of improvement—elements that resonate strongly with Sternberg’s understanding of practical intelligence or tacit knowledge.

For instance, “moral” directly connects to “humanity”, which further highlights the ethical implications of wisdom as underscored by McKenna, who argues that practical wisdom involves leading a morally upright life and developing healthy interpersonal interactions. Similarly, “wisdom” directly connects to “conscience”, and “conscience” further connects to “empathy” through the intermediary concept of “act.” This pattern appears to resonate with the definition by Grossmann, who describes wisdom as morally grounded excellence in social-cognitive processing and emphasizes the importance of empathy, reflection, and a sense of moral balance.

“Judge” stands as a vital node in the network, connecting to both “goals” and “action” through the concept of “improving.” This aligns with Sternberg’s definition that a person is wise to the extent that they use their knowledge to achieve a common good by balancing various interests over different temporal scales through ethical means. Notably, “action” also emerges as a crucial node in the network. This association is consistent with Karami’s assertion that wisdom involves adequate knowledge, intelligence, and creativity for problem-solving, again emphasizing the pragmatic aspects of wisdom.

Furthermore, the presence of the terms “democracy” and “vote” is also noteworthy. In its ideal form, democracy supports the selection and development of leaders who exemplify honesty and the best interests of the people. The concept of “democracy” being associated with “decision” underpins the societal dimensions of wisdom. As a result, it is analogous to Rooney and McKenna’s notion of wisdom as a collective obligation that prevents the development of toxic leaders, which accords with democratic norms.

Lastly, the direct link between “wisdom” and “conscience”, which is further connected to “respect” via “policies”, accentuates the importance of intrapersonal, interpersonal, and extrapersonal considerations in wisdom. The excerpts highlight that conscience motivates actions that respect the interests of others and confront global inequalities. These

features resonate profoundly with the Balance Theory's emphasis on wisdom as the ability to balance personal, others', and broader societal interests.

In conclusion, the semantic network and contexts drawn from the speeches provide a robust illustration of Sternberg's Balance Theory of Wisdom. The prominence of judgment, moral considerations, practical actions, and a conscientious approach towards balancing varied interests, all in service of addressing complex challenges, aligns remarkably well with Sternberg's nuanced definition of wisdom.

**The Most Semantically Relevant Paragraph to Wisdom**

We can find the paragraph that exhibits the closest semantic connection to definitions of wisdom by using the capability of the word2vec model. This sophisticated model allows us to analyze the semantic distance between paragraphs of our corpus and the fed paragraph – the new paragraph that we prepared as a conceptual essence of wisdom defined in literature. The paragraph that most concisely embodies the underlying principles of wisdom (found paragraph) would be the paragraph with highest cosine similarity to the fed paragraph. The paragraphs, which have been fed and found, are displayed in the following grey boxes.

**Fed Paragraph:**

"moral and intellectual development virtue practical rich factual knowledge, rich procedural knowledge, lifespan contextualism, relativism of values, and awareness and management of uncertainty. common good through a balance of different interests and perspectives, long term as well as the short term ethical values, humility, and concern for others. the adequate use of knowledge, intelligence and creativity, self-regulation, openness and tolerance, altruism and moral maturity, and sound judgment to solve critical problems Morally grounded wisdom balances self-interest with others, values truth, cares for humanity. excellence in social-cognitive processing involves considering different contexts, perspectives, short and long-term effects, thinking reflectively and dialectically, and being aware of limitations and subjectivity of thinking. inspiring and empowering others, the use of reason and careful observation, allowing for nonrational and subjective elements when making decisions, valuing humane and virtuous outcomes, being practical and oriented towards everyday life, and being articulate. un-

derstanding the aesthetic dimension of work and seeking intrinsic personal and social rewards. ability to communicate effectively, build relationships, and inspire others in a balanced way, depending on the specific situation they are facing. prudent judgments lead to decision-making for the good of the organisation and society. perception and understanding of people, things, and events quickly. creating contexts for meaningful interactions. using metaphors and stories to convey tacit knowledge. employing political power to mobilise action. mentoring and cultivating practical wisdom in others. emotional regulation, humor, critical life experiences, reflectiveness/reminiscence, and openness to experience."

**Found Paragraph:**

Simmons_11: "Thus, I believe that the task of a great university is not merely to test the mettle and stamina of brilliant minds but to guide them toward enlightenment, enabling thereby the most fruitful and holistic use of their students' intelligence and humanity. That enlightenment suggests the need for improving upon students' self-knowledge but it also means helping them judge others fairly, using the full measure of their empathy and intelligence to do so. In an environment rich in differences of background, experience and perspectives, learning is turbo charged and intensified by the juxtaposition of these differences. Those open minded enough to benefit fully from the power of this learning opportunity are bound for leadership in this time of confusion and division. The Harvard model intentionally and successfully provides to students a head start in understanding how to mediate difference in an ever more complex reality in which some exploit those differences for corrupt purposes."

In Ruth Simmons' perspective, several key dimensions of wisdom are apparent:

*Self-Knowledge and Self-Regulation:* Simmons identifies the university's role as guiding students towards enlightenment, which entails an enhanced understanding of one's own capacities and limitations. This self-knowledge, along with the ability to regulate one's own behavior, is a central aspect of wisdom. It facilitates the development of personal integrity and the ability to learn from past experiences and mistakes.

*Empathy and Altruism:* According to Simmons, universities must encourage students to judge others fairly using empathy and intelligence. This capacity to empathize, under-

stand, and care for others is a significant dimension of wisdom. It nurtures an individual's altruistic tendencies and promotes interpersonal relationships.

*Sound Judgment and Decision-Making:* Simmons posits that through exposure to diverse backgrounds, experiences, and perspectives, students are equipped to make fair judgments and mediate differences. This ability to make informed, balanced, and ethical decisions in complex situations is a crucial aspect of wisdom.

*Understanding and Appreciating Diversity:* The paragraph stresses the importance of being open-minded and appreciating differences in backgrounds, experiences, and perspectives. Understanding and appreciating diversity, supercharges the learning process and prepares students for leadership in a time of confusion and division.

*Leadership and Social Responsibility:* Simmons highlights the potential for students who embrace this diverse learning environment to become leaders. Leadership, in this context, implies the capacity to use one's wisdom to promote a common good, balancing various interests, and making decisions that consider both immediate and long-term consequences. This emphasis on leadership underscores the social responsibility aspect of wisdom.

*Ethical Values:* Simmons mentions the corrupt purposes for which differences can be exploited. This suggests the importance of "relativism of values" as a dimension of wisdom. It entails making morally upright decisions and behaving ethically even under pressure.

These dimensions of wisdom, as outlined by Simmons, are not only crucial in the academic setting but also apply broadly to personal and other professional contexts. They provide a solid foundation for nurturing wise individuals capable of effectively navigating an increasingly complex and diverse world.

### Top2Vec Topics

Finally, we concentrate on Top2Vec's key output: semantically coherent subjects. Each topic is a semantic space cluster of words that conveys a distinct theme in the corpus. We determine the major themes in our dataset by analyzing these topics. This allows us to move away from individual words and toward a linked web of themes. The complete result of Top2Vec is shown in Appendix VI.

Topic 0 appears to emphasize the notion of 'Development and Improvement' with words such as "distance", "gradually", "end", "enable", and "reached", possibly indicating the trajectory of personal and societal progression. Moreover, the documents related to this topic touch on the urgency of responsible algorithm development, the fluidity of plans and goals, and the importance of persisting in connecting communities. Concerning wisdom, these narratives echo elements of cognitive and reflective dimensions of wisdom. They suggest an understanding of a broader context, recognition of uncertainty, and an ability to adapt and respond to change, aligning with Sternberg's "Wisdom as a form of Reasoning" perspective.

Topic 1 seems to center around 'Personal Choices and Attitudes' with words such as "compare", "follow", "happiness", "courageous", and "defined". The documents in this topic invite reflections on personal choices, attitudes toward challenges, the essence of authenticity, and the balance between personal goals and pragmatic constraints. This topic likely resonates with the reflective and affective dimensions of wisdom. It captures reflections on self and others, emotions, and values, consistent with Ardelt's 'Three-Dimensional Wisdom Model'.

The issue of 'Communal Responsibility and Global Challenges' is highlighted in Topic 2. with words like "stronger", "potential", "teams", "governments", and "diversity". The associated documents underscore the power of relationships, the sense of community with words such as "we", "our", "teams", global responsibilities, and the pursuit of social justice. This topic resonates with wisdom's reflective and cognitive dimensions, suggesting a broad understanding of social matters and a capacity for empathy and compassion, particularly relevant to 'Berlin Wisdom Paradigm'.

Topic 3 focuses on "Personal Experiences and Challenges" featuring words like "finished", "wondered", "kept", "reading", and "launched". The documents reveal personal loss stories, critical self-reflection, challenges, and aspirations. This topic, while not directly tied to a particular wisdom model, may represent a form of experiential wisdom, suggesting that wisdom could stem from personal experiences and the reflective process they prompt.

Topic 4, distinctly apart from the wisdom framework, emphasizes "Formality and Gratitude," as illustrated by the words "thank", "inviting", "honored", and "graduates". The associated documents are largely expressions of gratitude and honor, customary to commencement addresses. Nonetheless, they offer a valuable insight into the contextual

nature of the analyzed speeches.

These findings support the multifaceted character of wisdom, which encompasses personal decisions, communal obligations, personal narratives, and social betterment. While not every issue clearly ties to wisdom, their overall perspective of wisdom as a multidimensional construct highlights the wisdom buried in commencement addresses.

### Heatmap

For each profession in our dataset, we can calculate the mean of the document-topic similarities generated by the Top2Vec model. These values represent the average topic distribution for each profession. The heatmap in Figure 4.20 visually depicts these proportions, which represent the average extent to which each topic is present in the speeches of the corresponding profession. As can be observed, each profession addresses every topic to a certain degree, but one or two themes predominantly characterize each group.

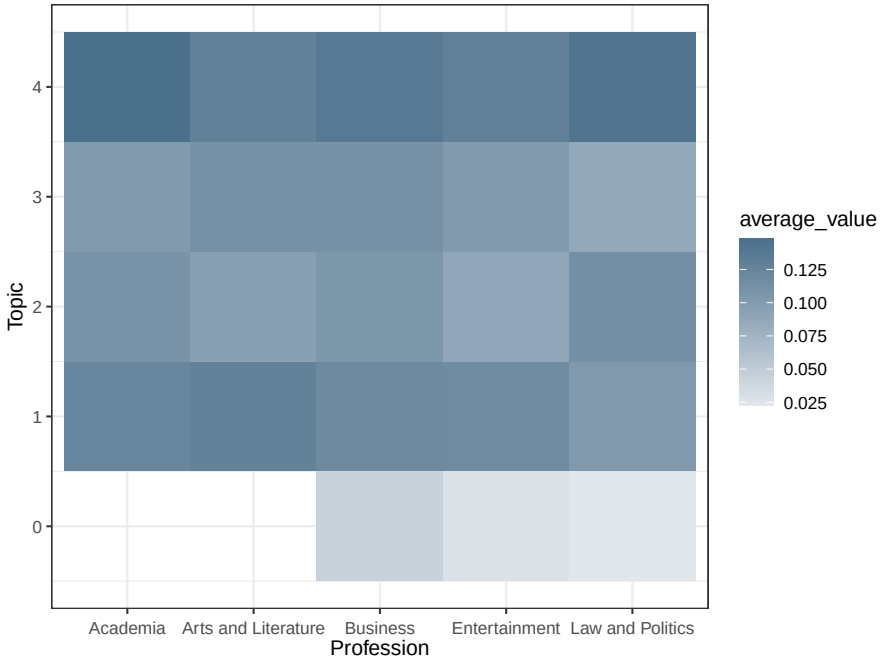



Figure 4.20: Distribution of topics over different fields

This heatmap provides an insightful visualization of the distribution of topics across various professions. The stronger colors in specific areas indicate a higher frequency of that topic within speeches given by the professionals of the corresponding field.

One of the most salient features of the heatmap is the strong prevalence of Topic 4

('Formality and Gratitude') across all professions. This may be attributable to the ceremonial nature of commencement addresses, which often start with acknowledgments and expressions of gratitude.

Additionally, Topic 1 ('Personal Choices and Attitudes') appears to have a significant presence across the sectors of academia, arts and literature, business, and entertainment, suggesting these fields emphasize personal growth, decision-making, and individuality in their addresses.

In contrast, Topic 2 ('Communal Responsibility and Global Challenges') is notably more prominent in speeches from law and politics professionals. This reflects the societal and global focus inherent in these fields.

Overall, this heatmap serves as a visual summary of how different thematic elements are distributed across speeches from various professional sectors, revealing unique patterns that characterize each profession's discourse.

## 4.5 Conclusion

In conclusion, text mining techniques such as word frequency analysis, sentiment analysis, and topic modeling have been instrumental in deriving meaningful insights from the corpus of commencement speeches purportedly without the necessity to read the whole texts. Each technique offers its unique perspectives and strengths, although they each have their limitations. Furthermore, the results of each technique have been analyzed through the lens of various dimensions of wisdom, drawing correlations between the findings and the wisdom concepts proposed by various scholars.

Word frequency analysis offers insights into the most frequently discussed terms and ideas within the corpus. It identifies the dominant narratives in different domains like Academia, Arts and Literature, Law and Politics, and others. However, it operates at a very surface level and lacks the ability to capture complex semantic and syntactic structures. The results from this technique often need to be interpreted with caution due to its context-agnostic nature. Therefore, the context of some of the most frequent words was extracted and analyzed.

Sentiment analysis unveils the emotional arcs of the speeches, revealing how the sentiment changes throughout the narrative. This gives us an understanding of how speak-

ers create engagement, express empathy, and deliver wisdom throughout their stories. However, sentiment analysis, much like word frequency analysis, can sometimes fail to capture nuances, especially when sentiments are expressed subtly or indirectly.

Topic modeling, through methods such as LDA, STM, and Top2Vec, provides more nuanced and in-depth insights. It helps to identify latent themes across the speeches, bringing out the essence of the messages delivered by speakers from various fields. Yet, topic modeling can sometimes produce less intuitive results, demanding interpretation from the analysts.

We observed that each analytical approach provides distinct yet interconnected insights, therefore complementing one another when applied in combination. Topic modeling provides various comprehensive insights, while sentiment analysis offers the context of emotional dynamics, and word frequency analysis delivers a quick scan of the key themes. The combination of these techniques provides a holistic perspective, reinforcing and complementing each other's results. For instance, while topic modeling might discover the themes around 'civil rights' and 'global issues', sentiment analysis can further expose the emotional nuances when such topics are discussed. Similarly, word frequency analysis can quickly verify these themes by indicating the high occurrence of relevant terms.

The analysis is enriched by interpreting the text mining results through the lens of various dimensions of wisdom proposed by scholars like Ardelt, Grossmann, Karami, Sternberg, and others. This approach bridges the gap between raw textual data and meaningful insights about wisdom, showcasing how commencement speeches encapsulate various aspects of wisdom as defined by these scholars.

The limitations of these techniques, however, underline the importance of combining multiple techniques and incorporating human interpretation to derive the most accurate and comprehensive understanding from the data. While these techniques can capture various elements of the speeches, they cannot fully grasp the depth of wisdom conveyed. However, as a starting point, they provide us with a robust framework to systematically analyze a large volume of text and filter valuable insights. The interplay of these techniques, coupled with the integration of wisdom dimensions, offers an intriguing avenue for exploring wisdom within commencement speeches.

CHAPTER 5

# CONCLUSION

## 5.1 Summary of Key Findings

The universality of some values and themes across professions was shown by our word frequency analysis, indicating the broad appeal of wisdom principles. However, various phrases within each field revealed diverse priorities, implying the presence of profession-specific viewpoints. Sentiment analysis revealed an overall upbeat tone with differences between professions, as well as a similar narrative arc within the speeches. Key themes were found using topic modeling approaches throughout talks, with each profession having distinct thematic prominences.

### 5.1.1 Word Frequency Analysis Findings

The research employed a robust word frequency analysis technique to systematically explore and unravel the rich content of commencement speeches across various professions - Academia, Arts and Literature, Business, Entertainment, and Law and Politics. This quantitative method was used to identify the most frequent words, bigrams, and trigrams across the speeches.

Firstly, the analysis revealed that despite the diversity of professions, there are shared words that indicate common themes. Terms such as “can”, “people”, “good”, “one”, “life”, “year”, “know”, “make”, “think”, and “us” were found to be pervasive across the speeches. This underscores the presence of shared values and themes across the professions re-

lating to ability, relationships, quality of life, and cognition.

In specific professional domains, unique terms depicted the distinct issues discussed. For instance, in Academia, phrases like “hbcus”, “underfund”, and “minority serve institution” highlight the emphasis on diversity, equity, and the challenges faced by academic institutions. In Arts and Literature, “make good art” and “default” emphasize the importance of creativity and challenging conventional norms. Business speeches with words like “rocket”, “big meaningful project”, and “give everyone freedom” suggest themes of innovation, significance, and liberty. Entertainment-related speeches contained terms like “edge awareness”, “let light shine”, and “character define moment”, reflecting themes of self-awareness, enlightenment, and personal character development. In the field of Law and Politics, references to civil rights, historic figures, and international relations were evident.

The analysis also identified common bigrams and trigrams across all professions, providing a more nuanced understanding of shared themes and narratives. Terms such as “tell_story”, “help_us”, “don_know”, and “make difference life” highlight the focus on storytelling, help-seeking, acknowledgment of uncertainty, and the urge to make a difference in life. These findings emphasize the shared wisdom in acknowledging the unknown and encouraging transformative actions across different fields.

Several phrases’ usage was found to be context dependent. For instance, “make difference life” was used differently by various speakers. Barack Obama emphasized service and altruism, while Steve Jobs focused on personal growth and fulfillment. Oprah Winfrey discussed using her influence to make a difference in people’s lives. These varying usages demonstrate the multi-layered, diverse meanings a phrase can carry across different speeches and contexts.

This analysis, built on frequency counts of words, bigrams, and trigrams, provides valuable insights into the commonalities and uniqueness within commencement speeches across various professions. It underlines the universal appeal of some themes, such as striving for progress, acknowledging uncertainty, and the ambition to make a difference. On the other hand, the unique words and phrases within each profession reveal the distinct emphasis and priorities in different fields. Thus, the word frequency analysis creates a rich initial perspective of wisdom within commencement speeches, bridging the gap between textual data and meaningful insights.

Concerning the first research question, analyzing the most frequently occurring unigrams, bigrams, and trigrams has provided insights into the common themes in the speeches across different professions. We have discovered that despite the diversity in professions, there are shared words and themes that recur across all fields. This could indicate a universal understanding of wisdom leadership that includes ideas of ability, relationships, quality of life, and cognition. Furthermore, unique terms within each profession revealed the distinct emphasis and priorities in different fields.

With respect to the fourth research question, it was found that the common elements in speeches across all professions, such as “make difference life”, “help us”, “don’t know”, etc., suggest a shared focus on acknowledging the unknown and encouraging transformative actions, key aspects of wisdom leadership. The different usage of some of these common phrases across speeches also was investigated.

The fifth question in this study sought to determine differences among professions. We have identified differences in the themes and most frequent terms of speeches delivered by leaders from diverse professions. The unique words and phrases within each profession reveal the distinct focus and priorities in different fields, shedding light on how wisdom may manifest differently in various professional contexts.

### 5.1.2 Sentiment Analysis Findings

Firstly, a sentiment analysis was performed on the speeches from different professional sectors. The analysis found that all professions exhibited a mix of negative and positive sentiments, with the overall trend being positive. It was observed that Academia was least negative, closely followed by entertainment, while sectors like arts, literature, law, and politics tended to have a higher degree of negative sentiment.

The analysis went deeper by exploring paragraphs that displayed extreme positive or negative sentiments. Positive sentiments often praised the profession or society’s accomplishments, while negative sentiments were typically critique or expressions of concern over societal issues. For example, in Academia, positive sentiment celebrated Harvard’s impact, while the negative sentiment expressed anxiety about the impact of a human-made crisis.

The analysis also found a common narrative arc among the speeches, similar to the arc of a story - it begins positively, dips as the speeches move into discussing challenges, and finally rises back as it nears the end, marking the resolution phase. Furthermore, the

study identified unique characteristics in the speeches of each profession. In Academia, there was a peak of positive sentiment around the midpoint of the speeches, which might suggest more stories were included in the speech content.

Moreover, key insights were derived from connecting sentiment with the concept of wisdom leadership. For instance, the emotion of trust, which was the most positive and prevalent across professions, was linked to the wisdom dimensions of "relativism of values" and "interpersonal," underlining the importance of understanding and respecting diverse perspectives and fostering healthy interpersonal relationships. On the other hand, fear, being the most negative emotion, was tied to the wisdom dimension of "awareness and management of uncertainty." We also analyzed the very positive and negative sentiments in each field and made connection to wisdom. For instance, paragraph "Wallace_29" in the Arts and Literature field, underscores the significance of self-awareness, empathy, and a broader perspective that seeks the common good through balancing self-interest the interests of others and cares for humanity.

Overall, the sentiment analysis and subsequent narrative arc interpretation provided valuable insights into the traits of wise leaders across professions, how their traits aligned with the dimensions of wisdom leadership, and how wisdom leadership manifests differently in various professional contexts.

This sentiment analysis provides a foundation to answer *RQ2*, which sought to understand how these traits align with the dimensions of wisdom leadership. Similarly, by uncovering principal themes and their association with sentiments across speeches, we are one step closer to answering *RQ3* and *RQ4*. It also further extends our understanding of *RQ5*, by offering a more nuanced comprehension of differences in the themes and elements of speeches delivered by leaders from different professions and the implications of these differences for wisdom leadership in various professional contexts.

### 5.1.3 LDA Findings

In our research, we employed topic modeling to identify the main themes that emerged from a set of commencement speeches across various professions, including Academia, Arts and Literature, Business, Entertainment, and Law and Politics. Through Latent Dirichlet Allocation (LDA), a probabilistic model for collections of text corpora, we identified five main topics that featured these speeches.

Our analysis indicated that every profession covered each topic to some extent, but

usually, one or two topics dominated each group. Topic 5, which we labeled as “Personal Development and Awareness”, was universal and appeared in all professions except business. This theme promoted personal growth, attention to the world, and encouraging one’s “light” to shine, reflecting a broader focus on self-development and perception.

We found that the business profession was mainly influenced by Topic 1, “Inspiration and Purpose”, which encompassed terms such as “think_good”, “one_day”, “sense_purpose”, and “can_get”. This focus indicated the profession’s inclination towards positive thinking, a clear sense of purpose, and future vision.

Law and Politics, on the other hand, closely correlated with Topics 2, “Political Considerations”, and topic 3, “Civil Rights and Global Issues”. This reflected the profession’s inherent emphasis on political and social contexts, civil rights discussions, and pressing global issues.

Arts and Literature speeches predominantly reflected Topic 4, “Art and Truth”, indicating the profession’s focus on truth-telling, making good art, and the value of liberal arts. The academic profession presented a diverse topic range, but there was a stronger emphasis on Topic 5, followed by Topics 1 and 3.

These findings suggest that each profession has its unique thematic emphasis while sharing some common themes across the board. It underscores the importance of context in discourse and highlights how various professions address their specific audience’s needs and expectations.

### 5.1.4 STM Findings

Our STM analysis revealed a diverse range of wisdom dimensions and traits embedded in these speeches (RQ2). For instance, we noticed themes of self-regulation and moral maturity, empathy and altruism, self-reflection and ethical understanding, and societal transformation. Interestingly, we observed that these wisdom dimensions align well with the definitions provided by various scholars in the field of wisdom studies, such as Grossmann, Karami, Nonaka, Sternberg, McKenna, and others. This alignment underlines the utility of our text-mining technique in extracting meaningful insights about wisdom leadership from textual data.

Moreover, by examining the distribution of topics across different professions, we connected the themes and common elements identified through our STM analysis with wis-

dom leadership aspects relevant to each profession (RQ4). For example, the entertainment sector showed a prevalent theme of learning from failure and the altruistic use of power. On the other hand, law and politics were associated with wisdom dimensions such as the integration of heart and mind, self-reflection, and ethical understanding. These findings highlight each field's unique characteristics and dominant themes.

Furthermore, our results revealed noticeable differences in topic distributions among speeches delivered by leaders from various professions (RQ5). Specific topics, such as Topic 6 on institutional responsibility and historical awareness, resonated strongly in the business sector, while other topics, such as Topic 3 on resilience and empathy, were more dominant in law and politics. These variations suggest that different wisdom dimensions may be more or less relevant depending on the professional context, reflecting the multifaceted nature of wisdom leadership.

### 5.1.5 Top2Vec and Word2Vec Findings

In response to RQ1 and RQ4, the word2vec results highlighted the complexity and multifaceted nature of wisdom through key interconnected terms such as "wisdom," "moral," "conscience," "action," and "judge." These terms and their interrelations were linked to the components of Sternberg's Balance Theory of Wisdom, the Berlin School's idea of wisdom, McKenna's and Grossmann's definitions of wisdom, and Karami's assertion about wisdom. Thus, the analysis not only surfaced the most frequent ngrams in the speeches but also illuminated how they provide insights into the concept of wisdom leadership.

For RQ2, these results demonstrated how text-mining techniques could be used to reveal the traits of wise leaders. For instance, the interconnected terms "judge," "humanity," "moral," and "action" highlighted the importance of ethical values, balancing various interests for a common good, and pragmatic intelligence. These traits align with the dimensions of wisdom leadership as they suggest a wise leader would need to be analytical, morally grounded, and practically intelligent.

Addressing RQ3, the Top2Vec analysis recognized principal themes across the speeches, ranging from 'Development and Improvement' to 'Communal Responsibility and Global Challenges.' These topics revealed wisdom's reflective, cognitive, and affective dimensions. This is crucial, as it shows that wisdom is not only a personal attribute but also a social one that extends to community interactions and global concerns.

Finally, for RQ5, a heatmap of topics across various professions revealed variations in

the emphasis on different themes in different professional contexts. For instance, 'Communal Responsibility and Global Challenges' was prominent in speeches from law and politics professionals, indicating their societal and global focus. These variations can reveal unique characteristics of wisdom leadership in different professional contexts, showing how wisdom may manifest differently depending on the context and profession.

In the process of our analysis, the themes of pursuit of passion or achievement emerged across various professional speeches. The following section serves as an interpretive space where we critically engage with these prominent themes and explore their relationship with wisdom leadership.

## 5.2 Wisdom Leadership Reinterpreted: The Pursuit of Passion and Achievement

In our analysis of commencement addresses, a recurring theme emerged centered around the pursuit of passion and the drive for achievement. At first glance, this appears to be a clear directive: follow your passions, strive for success, achieve great things, or Jobs' "Stay Hungry. Stay Foolish." Yet, when considering this theme in the context of wisdom, the interpretation deepens and evolves.

Traditionally, wisdom is not solely defined by the pursuit of personal passions or achievements. It encompasses a broader spectrum of life experiences, ethical considerations, and a balance between personal fulfillment and collective well-being. Wisdom extends beyond goal-oriented thinking to include the process of understanding why we choose specific goals, how we aim to achieve them, and the possible implications of these decisions.

Pursuing passions and achievements can indeed be seen as a facet of a wise life, but they do not associate with wisdom in themselves. They are potential components of a life enriched by wisdom, but not definitive markers of it. The reason lies in the distinction between the goal (the what) and the journey (the why and the how).

Wisdom comes into play in determining whether these passions and goals align with broader ethical considerations and contribute positively to both personal growth and societal well-being. Wisdom guides us in how we strive to fulfill these ambitions, considering not just the intended result, but also the means, the impacts, and the adjustments required along the way.

Passion is beneficial when it fuels engagements that lead to personal fulfillment and positive societal impacts. Achievements are commendable when they result from ethical actions, wise decision-making, and when they contribute to individual and societal development. Therefore, wisdom is not about achieving the goal at any cost, but rather achieving it in a way that is considerate of the broader context and implications.

In essence, the true wisdom lies in the ability to balance passion with responsibility, achievement with ethics, and personal success with collective well-being. It involves recognizing and acknowledging the interconnectedness of our actions and aspirations with the world around us. It requires the humility to continue learning, the courage to adapt and change, and the generosity to consider the greater good.

Therefore, the repeated emphasis on pursuing passion and striving for achievement in the commencement addresses, when viewed through the lens of wisdom, can be seen as a call for a balanced, thoughtful, and responsible approach to one's life journey. It encourages graduates not only to aspire and strive but also to do so with wisdom, understanding, and consideration of the broader impacts of their pursuits.

## 5.3 Acknowledging Limitations and Potential Biases

While the present study provides valuable insights into the concept of wisdom in commencement speeches, it also has a number of limitations and potential biases that should be taken into account when interpreting the findings.

Firstly, the commencement speeches represent a specific context, marked by institutional traditions and expectations. As such, they might not fully capture the breadth and depth of wisdom leadership across different professional fields. Speakers may adjust their message to fit the occasion and the expectations of their audience, which could affect the themes and wisdom-related concepts that are highlighted in their speeches.

Secondly, our findings are tied to the cultural and institutional settings in which the speeches were given. They may not necessarily generalize to other settings or countries, and additional research would be needed to explore potential cultural differences in wisdom and leadership concepts.

Thirdly, the study focused on a specific conceptualization of wisdom, as captured by the Balance Theory and other theoretical frameworks mentioned in the study. Wisdom is

a complex construct, and different theoretical perspectives might highlight other aspects of wisdom that were not covered in this study.

The interpretation of the topics generated by different algorithms also involves a certain level of subjectivity. While we tried to provide clear and consistent interpretations, other researchers might interpret the same topics differently. Additionally, topic modeling is an exploratory technique, and its findings should be seen as hypotheses that need to be confirmed through additional research.

Regarding the profession-specific insights, our sample might have been biased towards certain professions, and our findings might not necessarily generalize to all individuals within these professions. The results related to profession-specific topics should therefore be interpreted with caution and confirmed through further studies.

While we used advanced text mining techniques to analyze the speeches, these techniques have their own limitations. They are not able to fully capture the nuanced meanings of the speeches, and they might miss some wisdom-related concepts that are expressed implicitly rather than explicitly. Validation of our findings through other methods or data sources would provide stronger evidence for the identified patterns.

Lastly, the interpretation and application of wisdom is also indeed subjective and can vary greatly among individuals based on numerous factors such as personal values, cultural background, life experiences, and cognitive abilities. Moreover, not everyone can readily adopt a long-term perspective, often due to immediate pressures, limited resources, or a lack of foresight and experience. The challenge, then, is not only to develop wisdom but also to recognize and respect its diverse expressions. Wisdom does not demand grand sacrifices or historical achievements from everyone. Instead, it invites us all to grow in self-awareness, nurture empathy, develop our understanding of the world, and strive for balance in our decisions and actions. It invites us to see beyond our immediate circumstances and to consider the broader implications of our actions—something that requires practice, patience, and, indeed, wisdom itself. Future research could address these limitations by using more diverse data sources, exploring other theoretical perspectives on wisdom, employing complementary analysis methods, and conducting more detailed profession-specific analyses. Despite these limitations, we believe that our study provides a valuable starting point for exploring the concept of wisdom in leadership discourse and highlights the potential of text-mining techniques for analyzing large text corpora.

## 5.4 Final Thoughts and Future Research

It appears that wisdom leadership can be largely considered as a universal construct, as it manifests across different professions. The common elements identified, such as embracing complexity, ethical decision-making, and acting for and valuing the common good, reflect broad principles that transcend specific professional boundaries.

However, while the core principles of wisdom leadership are universal, the application of these principles may vary according to the profession. Certain themes or topics might be more pertinent in one profession than another, reflecting the unique contexts and challenges of different fields. For example, speeches from the field of Law and Politics might emphasize civil rights, while those from Arts and Literature might focus more on creativity and challenging conventional norms.

Therefore, it can be concluded that wisdom leadership is a combination of both universal and profession-specific elements. The universal aspects form the principles that are applicable across all fields, while the profession-specific elements allow these principles to be applied effectively in different contexts.

Looking ahead, there are several avenues for future research. A deeper exploration of how universal principles of wisdom leadership are interpreted and applied within each profession could be undertaken. This could involve conducting interviews with leaders from different fields or more detailed case studies, providing a richer, more nuanced understanding of wisdom leadership in context.

An expansion of the dataset is another promising direction for future research. While commencement speeches offer valuable insights, including other forms of public discourse such as professional talks, interviews, or written works by leaders could provide a more comprehensive perspective on wisdom leadership. Furthermore, incorporating more commencement addresses from different cultures could illuminate potential cultural variations in wisdom leadership.

A crucial component of this research is the detailed results and interactive visualizations of the Structural Topic Modeling analysis, accessible at the following link: https://khadjehali.com/javad/vis/stm. This tool provides an in-depth exploration of topic distribution, semantic polarity, and document classifications across various professions, speakers, and locations. While it has been instrumental in our exploration of wisdom leadership, the potential applications of this tool extend well beyond this specific subject. The adapt-

able nature of this tool enables researchers to delve into a broad spectrum of subjects. For instance, sociologists could use it to analyze societal norms as expressed in commencement speeches, or political scientists could study the articulation of ideologies or construction of political identities. It has the potential to provide subtle insights into a wide range of study fields, from psychological studies of motivation and emotional appeals to historical analyses of social values across time.

In summary, this research represents a step in extending our understanding of wisdom leadership, highlighting both its universality and profession-specific variations. It also demonstrates the power of mixed-methods research in extracting deep, nuanced insights from text data. It affirms the potential of combining quantitative and qualitative techniques to enrich our understanding of various complex phenomena, thus making a noteworthy contribution to the broader field of text analysis research. The insights gathered here, in conjunction with potential future research directions, stand to make contribution to the ongoing discourse on wisdom leadership.

# Appendix I: Speeches Information

Table .1: Information Summary of Speeches.

| Name | Location | Title | Year | Profession | WordsLength |
|---|---|---|---|---|---|
| Chimamanda Ngozi Adichie | Harvard | Cultivating a sense of empathy | 2018 | Arts and Literature | 1823 |
| Jacinda Ardern | Harvard | The Fragility of Democracy | 2022 | Law and Politics | 1777 |
| Barack Obama | Notre Dame | The World as It Should Be | 2009 | Law and Politics | 1692 |
| Martin Baron | Harvard | Imperfect though | 2020 | Law and Politics | 1678 |
| Jeff Bezos | Princeton | Two key questions | 2010 | Business | 1673 |
| Bill Gates | Harvard | Global inequality, technology and innovation | 2007 | Business | 1603 |
| Mike Bloomberg | Johns Hopkins | Communication in the Digital Age | 2021 | Business | 1541 |
| Sterling K. Brown | Stanford | Let Your Light Shine | 2018 | Entertainment | 1538 |
| Ken Burns | Stanford | Reflections on History | 2016 | Arts and Literature | 1507 |
| Tim Cook | Stanford | Responsibility and Building | 2019 | Business | 1436 |
| France Cordova | Stanford | Navigating a Changing World | 2020 | Academia | 1429 |
| Mariano Florentino Cuellar | Stanford | The Gift of Progress | 2017 | Entertainment | 1378 |
| Ellen DeGeneres | Tulane | Follow Your Passion, Stay True to Yourself | 2009 | Entertainment | 1360 |
| Richard Engel | Stanford | Taking the Leap | 2015 | Law and Politics | 1323 |
| Neil Gaiman | University of the Arts | Make good art | 2012 | Arts and Literature | 1319 |
| Atul Gawande | Stanford | Finding Purpose in Life | 2021 | Academia | 1304 |
| Reed Hastings | Stanford | Keys to Progress and Change the World | 2022 | Business | 1283 |
| Steve Jobs | Stanford | Commencement Address at Stanford | 2005 | Business | 1248 |
| John Lewis | Harvard | Importance of equality | 2018 | Law and Politics | 1189 |
| Angela Merkel | Harvard | Anything can change | 2019 | Law and Politics | 1170 |

Table .1: Information Summary of Speeches (continued)

| Name | Location | Title | Year | Profession | WordsLength |
|---|---|---|---|---|---|
| Michelle Obama | Oberlin | Engage with the world around | 2015 | Law and Politics | 1102 |
| Elon Musk | Caltech | Think Big and Dream Even Bigger | 2012 | Business | 1099 |
| Conan O'Brien | Dartmouth | Failure and Invention | 2011 | Entertainment | 1091 |
| Natalie Portman | Harvard | Embracing Inexperience | 2015 | Entertainment | 1076 |
| J.K. Rowling | Harvard | The Fringe Benefits of Failure | 2008 | Arts and Literature | 1074 |
| Sheryl Sandberg | Berkeley | Finding Gratitude and Appreciation | 2016 | Business | 1022 |
| Ruth J. Simmons | Harvard | Fight inequality | 2021 | Academia | 999 |
| Steven Spielberg | Harvard | A villain to vanquish | 2016 | Entertainment | 992 |
| David Foster Wallace | Kenyon | This is Water | 2005 | Arts and Literature | 884 |
| Oprah Winfrey | Harvard | An internal moral, emotional G.P.S | 2013 | Entertainment | 837 |
| Fareed Zakaria | Harvard | We live in an age of progress | 2012 | Law and Politics | 742 |
| Mark Zuckerberg | Harvard | Purpose and Community | 2017 | Business | 607 |

The dataset under study have been gathered from various online sources. As these speeches are all publicly accessible online, they ensure transparency and reproducibility of the research, while also allowing anyone interested to delve deeper into the original source of wisdom (Adichie, 2018; Ardern, 2022; Baron, 2020; Bezos, 2010; Bloomberg, 2021; S. K. Brown, 2018; Burns, 2016; Cook, 2019; Cordova, 2020; Cuellar, 2017; DeGeneres, 2009; Engel, 2015; Gaiman, 2012; Gates, 2007; Gawande, 2021; Hastings, 2022; Jobs, 2005; Lewis, 2018; M. Obama, 2015; Merkel, 2019; Musk, 2012; O'Brien, 2011; B. Obama, 2009; Portman, 2015; Rowling, 2008; Sandberg, 2016; Simmons, 2021; Spielberg, 2016; Wallace, 2005; Winfrey, 2013; Zakaria, 2012; Zuckerberg, 2017).

Table .2: Pictures of Speakers.

Angela Merkel

Atul Gawande

Barack Obama

Bill Gates

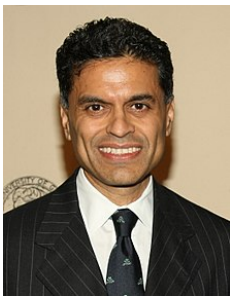
Chimamanda Ngozi Adichie

Conan O'Brien

David Foster Wallace

Ellen DeGeneres

Elon Musk



Fareed Zakaria

France Cordova

J.K. Rowling

Jacinda Ardern

Jeff Bezos

John Lewis

Ken Burns

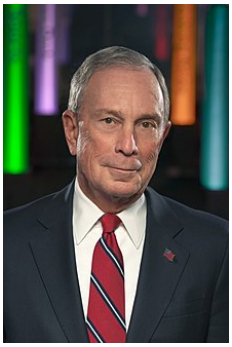
Table .2: Pictures of Speakers (continued)

Mariano Florentino Cuellar

Mark Zuckerberg

Martin Baron

Michelle Obama



Mike Bloomberg

Natalie Portman

Neil Gaiman

Oprah Winfrey

Reed Hastings

Richard Engel

Ruth J. Simmons

Sheryl Sandberg

Sterling K. Brown

Steve Jobs

Steven Spielberg

Tim Cook

# Appendix II: Dataframe Making Codes

*Load Libraries:*

Load the necessary libraries for text processing, data manipulation, and analysis.

```
library(stringr)
library(tidyverse)
library(tm)
library(quanteda)
library(textstem)
library(udpipe)
```

Read the files containing commencement speeches and store them in a list.

```
file_path <- "C:/D/R/TAWR2/data/text/Commencements/"
all_titles <- list.files(paste0(file_path, "txt/"))
speeches_l <- lapply(all_titles, function(x) scan(paste0(file_path,
    "txt/", x, sep = ""), what = "character", sep = "\n"))
all_titles <- gsub(".txt", "", all_titles)
names(speeches_l) <- all_titles
```

*Define Custom Stopwords:*

Create a custom list of stopwords, which includes standard English stopwords and additional words specific to the dataset.

```
custom.stopwords <- c(stopwords("english"), letters, "(laughter.)",
    "(applause.)", "authenticity certified text version below transcribed directly from audio")
```

*Preprocessing Function:*

Define a function to preprocess and lemmatize text. This function will clean the text by converting it to lowercase, removing punctuation, removing stopwords, removing numbers, and lemmatizing the words.

```
preprocess_and_lemmatize <- function(text) {
    text %>%
        tolower() %>%
        removeWords(custom.stopwords) %>%
        removePunctuation(preserve_intra_word_dashes = TRUE,
            ucp = FALSE) %>%
        removePunctuation(preserve_intra_word_dashes = TRUE,
            ucp = TRUE) %>%
        removeNumbers() %>%
        removeWords(custom.stopwords) %>%
        stripWhitespace() %>%
        lemmatize_strings()
}
```

Construct a data frame that includes document names, and text content.

```
doc_names <- names(speeches_l)
all_texts <- unlist(speeches_l)
docName <- rep(doc_names, sapply(speeches_l, length))
my_df <- data.frame(doc.name = docName, document = all_texts,
    row.names = NULL)
```

Add sequence numbers and percentages for each document in the data frame.

```
my_df <- my_df %>%
    group_by(doc.name) %>%
    mutate(seq_num = 1:length(document), percent = seq_num/length(document) *
        100)
```

*Clean and Preprocess Document Text:*

Clean and preprocess the text for each document in the data frame.

```
my_df$text <- preprocess_and_lemmatize(my_df$document)
my_df$text <- sapply(my_df$text, function(x) gsub("[^ _a-z]",
    "", x))
```

*Tokenize Text and Extract N-grams:*

Tokenize the text and extract bigrams and trigrams using a custom tokenizer function.

```
my_tokenizer <- function(x, n, s) {
    a <- quanteda::tokens(x, what = "word")
    b <- quanteda::tokens_ngrams(a, n = n, skip = s, concatenator = "_")
    c <- as.character(b)
```

```
    d <- paste(c, collapse = " ")
    return(d)
}
my_df$bigrams <- sapply(my_df$text, function(x) my_tokenizer(x,
    2, 0))
my_df$trigrams <- sapply(my_df$text, function(x) my_tokenizer(x,
    3, 0))
```

*Extract Years, Decades, and Speaker Details:*

Extract years, decades, speaker names, and titles from the data frame. Here we added "decade" because it can be useful in larger datasets.

```
my_df <- my_df %>%
    mutate(year = if_else(seq_num == 3, as.integer(str_extract(document,
        "\\b\\d{4}\\b")), NA_integer_), location = if_else(seq_num ==
        3, gsub("\\b\\d{4}\\b", "", document), NA_character_)) %>%
    fill(year, .direction = "down") %>%
    fill(location, .direction = "down") %>%
    mutate(decade = (year%/%10) * 10, name = if_else(seq_num ==
        1, document, NA), title = if_else(seq_num == 2, document,
        NA)) %>%
    fill(name, .direction = "down") %>%
    fill(title, .direction = "down") %>%
    mutate(speaker = str_remove(doc.name, "\\.csv"), name = str_extract(speaker,
        "^[^-]+"), title = str_extract(speaker, "(?<=-\\s)[^\\s].*")) %>%
    filter(!(seq_num %in% c(1, 2, 3))) %>%
    mutate(name = str_trim(name), speaker = str_extract(name,
        "\\w+$")) %>%
    group_by(speaker) %>%
    mutate(seq_num = 1:length(document), doc_id = paste0(speaker,
        "_", seq_num), percent = seq_num/length(document) * 100)
```

*Assign Professions to Speakers:*

Assign professions to the speakers in the dataset. Create a data frame containing speaker names and their professions.

```
professions <- c("Business", "Arts and Literature", "Entertainment",
    "Academia", "Law and Politics")
speaker <- unique(my_df$speaker)
speakerprofessions = data.frame(speaker, profession = c("Law and Politics",
    "Academia", "Law and Politics", "Business", "Arts and Literature",
    "Entertainment", "Arts and Literature", "Entertainment",
    "Business", "Law and Politics", "Academia", "Arts and Literature",
    "Law and Politics", "Business", "Law and Politics", "Arts and Literature",
```

```
    "Entertainment", "Business", "Law and Politics", "Law and Politics",
    "Business", "Entertainment", "Arts and Literature", "Entertainment",
    "Business", "Law and Politics", "Academia", "Business", "Entertainment",
    "Business", "Entertainment", "Business"))
```

*Merge Speaker Professions with Main Data Frame:*

Merge the `speakerprofessions` data frame with the main data frame, `my_df`.

```
my_df <- merge(my_df, speakerprofessions, by = "speaker", all.x = TRUE)
```

*Count Words and Filter Rows:*

Count the number of words in each document and filter the rows to keep only the ones that have more than 2 and less than 1000 words.

```
my_df <- my_df %>%
    mutate(nwords = txt_count(text, pattern = " "))
my_df <- subset(my_df, nwords < 1000 & nwords > 2)
```

*Export Clean Data Frame:*

Export the clean data frame to a CSV file for future use.

```
write.csv(my_df, file = paste0(file_path, "/df/df_cln_lem.csv"),
    row.names = FALSE)
```

*Compute Profession Statistics:*

Create a summary of the number of speakers per profession.

```
file_path <- "C:/D/R/TAWR2/data/text/Commencements/"
df <- read.csv(paste0(file_path, "/df/df_cln_lem.csv"))
# Extracting professions_stat
professions_stat <- df %>%
    group_by(profession) %>%
    dplyr::summarize(same_pro = length(unique(speaker))) %>%
    arrange(desc(same_pro))
professions_stat
```

```
## # A tibble: 5 x 2
##   profession          same_pro
```

```
##   <chr>                  <int>
## 1 Business                   9
## 2 Law and Politics           8
## 3 Entertainment              7
## 4 Arts and Literature        5
## 5 Academia                   3
```

*Compute Location Statistics:*

Create a summary of the number of speakers per location.

```
# Extracting locations_stat
location_stat <- df %>%
    group_by(location) %>%
    dplyr::summarize(same_loc = length(unique(speaker))) %>%
    arrange(desc(same_loc))
location_stat
```

```
## # A tibble: 13 x 2
##    location                  same_loc
##    <chr>                        <int>
##  1 "Harvard "                      12
##  2 "Stanford "                      9
##  3 "BERKELEY "                      1
##  4 "Caltech "                       1
##  5 "Dartmouth "                     1
##  6 "Harvard  "                      1
##  7 "Johns Hopkins "                 1
##  8 "Kenyon "                        1
##  9 "Notre Dame "                    1
## 10 "Oberlin "                       1
## 11 "Princeton "                     1
## 12 "Tulane "                        1
## 13 "University of the Arts "        1
```

*Compute Speech Length Statistics:*

Create a summary of the total speech length (in words) for each speaker.

```
# Extracting speech_length
speech_length_stat <- df %>%
    group_by(speaker) %>%
    dplyr::summarize(speech_length = sum(nwords)) %>%
    arrange(desc(speech_length))
speech_length_stat
```

*Create a Summary Data Frame for Speeches:*

Create a summary data frame containing speaker, location, profession, year, title, name, and total speech length.

```
summary_df_speeches <- distinct(df[, c("speaker", "location",
    "profession", "year", "title", "name")])
summary_df_speeches$SpeechLength <- speech_length_stat$speech_length
```

# Appendix III: The Complete List of the Packages in R

R version 4.2.2 (2022-10-31 ucrt) Platform: x86_64-w64-mingw32/x64 (64-bit) Running under: Windows 10 x64 (build 22621)

Matrix products: default

locale: [1] LC_COLLATE=English_United Kingdom.utf8 [2] LC_CTYPE=English_United Kingdom.utf8
[3] LC_MONETARY=English_United Kingdom.utf8 [4] LC_NUMERIC=C
[5] LC_TIME=English_United Kingdom.utf8

attached base packages: [1] stats graphics grDevices utils datasets [6] methods base

other attached packages: [1] text_0.9.99.2 tokenizers_0.3.0
[3] textplot_0.2.2 top2vecr_0.1.0
[5] dbscan_1.1-11 uwot_0.1.14
[7] Matrix_1.5-4.1 doc2vec_0.2.1
[9] word2vec_0.3.4 stm_1.3.6
[11] udpipe_0.8.11 flextable_0.9.1
[13] zoo_1.8-12 Hmisc_5.1-0
[15] devtools_2.4.5 usethis_2.2.0
[17] LDAvis_0.3.5 quanteda_3.3.1
[19] scales_1.2.1 Rtsne_0.16
[21] text2vec_0.6.3 sentimentr_2.9.0
[23] data.table_1.14.8 ggrepel_0.9.3
[25] qdapRegex_0.7.5 reshape2_1.4.4
[27] wordcloud_2.6 RColorBrewer_1.1-3
[29] ldatuning_1.0.2 topicmodels_0.2-14
[31] tm_0.7-11 NLP_0.2-1
[33] slam_0.1-50 magrittr_2.0.3
[35] textstem_0.1.4 koRpus.lang.en_0.1-4 [37] koRpus_0.13-8 sylly_0.1-6
[39] ggraph_2.1.0 igraph_1.4.3
[41] textdata_0.4.4.9000 drlib_0.1.1
[43] tidytext_0.4.1 knitr_1.43
[45] lubridate_1.9.2 forcats_1.0.0
[47] stringr_1.5.0 dplyr_1.1.2
[49] purrr_1.0.1 readr_2.1.4
[51] tidyr_1.3.0 tibble_3.2.1
[53] ggplot2_3.4.2 tidyverse_2.0.0

[55] kableExtra_1.3.4

loaded via a namespace (and not attached): [1] rappdirs_0.3.3 SnowballC_0.7.1
[3] overlapping_2.1 stopwords_2.3
[5] ragg_1.2.5 rpart_4.1.19
[7] hardhat_1.3.0 generics_0.1.3
[9] GPfit_1.0-8 callr_3.7.3
[11] RhpcBLASctl_0.23-42 cowplot_1.1.1
[13] future_1.32.0 tzdb_0.4.0
[15] webshot_0.5.4 xml2_1.3.4
[17] httpuv_1.6.11 viridis_0.6.3
[19] gower_1.0.1 xfun_0.39
[21] hms_1.1.3 evaluate_0.21
[23] promises_1.2.0.1 fansi_1.0.4
[25] htmlwidgets_1.6.2 stats4_4.2.2
[27] rsparse_0.5.1 ellipsis_0.3.2
[29] backports_1.4.1 fontLiberation_0.1.0
[31] bookdown_0.34 RcppParallel_5.1.7
[33] fontBitstreamVera_0.1.1 vctrs_0.6.2
[35] remotes_2.4.2 cachem_1.0.8
[37] withr_2.5.0 ggforce_0.4.1
[39] checkmate_2.2.0 prettyunits_1.1.1
[41] parsnip_1.1.0 gofastr_0.3.1
[43] svglite_2.1.1 cluster_2.1.4
[45] pacman_0.5.1 crayon_1.5.2
[47] crul_1.4.0 labeling_0.4.2
[49] recipes_1.0.6 pkgconfig_2.0.3
[51] tweenr_2.0.2 nlme_3.1-162
[53] pkgload_1.3.2 nnet_7.3-19
[55] rlang_1.1.1 globals_0.16.2
[57] lifecycle_1.0.3 miniUI_0.1.1.1
[59] fontquiver_0.2.1 httpcode_0.3.0
[61] lexicon_1.2.1 polyclip_1.10-4
[63] matrixStats_1.0.0 yardstick_1.2.0
[65] base64enc_0.1-3 processx_3.8.1
[67] png_0.1-8 viridisLite_0.4.2
[69] float_0.3-1 workflows_1.1.3
[71] tune_1.1.1 parallelly_1.36.0
[73] memoise_2.0.1 plyr_1.8.8
[75] compiler_4.2.2 cli_3.6.1
[77] urlchecker_1.0.1 DiceDesign_1.9
[79] listenv_0.9.0 janeaustenr_1.0.0
[81] ps_1.7.5 sylly.en_0.1-3
[83] htmlTable_2.4.1 formatR_1.14
[85] Formula_1.2-5 mgcv_1.8-42
[87] MASS_7.3-60 tidyselect_1.2.0
[89] stringi_1.7.12 textshaping_0.3.6
[91] yaml_2.3.7 askpass_1.1
[93] grid_4.2.2 fastmatch_1.1-3
[95] tools_4.2.2 timechange_0.2.0
[97] future.apply_1.11.0 parallel_4.2.2

[99] rstudioapi_0.14 uuid_1.1-0
[101] foreach_1.5.2 foreign_0.8-84
[103] gridExtra_2.3 prodlim_2023.03.31
[105] farver_2.1.1 digest_0.6.31
[107] shiny_1.7.4 lava_1.7.2.1
[109] gfonts_0.2.0 Rcpp_1.0.10
[111] later_1.3.1 httr_1.4.6
[113] gdtools_0.3.3 colorspace_2.1-0
[115] rvest_1.0.3 brio_1.1.3
[117] fs_1.6.2 reticulate_1.29
[119] splines_4.2.2 lgr_0.4.4
[121] graphlayouts_1.0.0 sessioninfo_1.2.2
[123] systemfonts_1.0.4 xtable_1.8-4
[125] jsonlite_1.8.5 tidygraph_1.2.3
[127] timeDate_4022.108 modeltools_0.2-23
[129] testthat_3.1.8 ipred_0.9-14
[131] R6_2.5.1 lhs_1.1.6
[133] profvis_0.3.8 mlapi_0.1.1
[135] pillar_1.9.0 htmltools_0.5.5
[137] mime_0.12 glue_1.6.2
[139] fastmap_1.1.1 class_7.3-22
[141] codetools_0.2-19 pkgbuild_1.4.0
[143] furrr_0.3.1 utf8_1.2.3
[145] dials_1.2.0 lattice_0.21-8
[147] curl_5.0.0 officer_0.6.2
[149] zip_2.3.0 openssl_2.0.6
[151] survival_3.5-5 textclean_0.9.3
[153] rmarkdown_2.22 syuzhet_1.0.6
[155] munsell_0.5.0 iterators_1.0.14
[157] rsample_1.1.1 gtable_0.3.3

# Appendix IV: The extremely negative or positive paragraphs in each profession

Table .3: The extremely negative or positive paragraphs in each profession

| Profession | doc.id | Sentiment | Native.Paragraph |
|---|---|---|---|
| Academia | Simmons-13 | + | Harvard is, in some ways, the most powerful university bully pulpit in the nation. It did not achieve that status merely through its age and wealth; it attained that status principally through the efforts of its faculty and graduates' scholarly and professional output. Through its gates have come generations of scholars with immense intelligence and passionate purpose to whom fate bequeathed the laurels of success. But it is important that universities model in their own values and actions the high purpose that they hope to see in the actions of their scholars. |
| Academia | Gawande-29 | - | For me, dealing with the uncertainty was the hardest part. How long would this plague last? How bad would it get? What troubled me most, though, was the fact that so much of the uncertainty was human-made, instead of virus-made. |
| Arts and Literature | Wallace-29 | + | And the so-called real world will not discourage you from operating on your default settings, because the so-called real world of men and money and power hums merrily along in a pool of fear and anger and frustration and craving and worship of self. Our own present culture has harnessed these forces in ways that have yielded extraordinary wealth and comfort and personal freedom. The freedom all to be lords of our tiny skull-sized kingdoms, alone at the centre of all creation. This kind of freedom has much to recommend it. But of course there are all different kinds of freedom, and the kind that is most precious you will not hear much talk about much in the great outside world of wanting and achieving…. The really important kind of freedom involves attention and awareness and discipline, and being able truly to care about other people and to sacrifice for them over and over in myriad petty, unsexy ways every day. |

Table .3: The extremely negative or positive paragraphs in each profession *(continued)*

| Profession | doc.id | Sentiment | Native.Paragraph |
|---|---|---|---|
| Arts and Literature | Wallace-17 | - | Or, of course, if I'm in a more socially conscious liberal arts form of my default setting, I can spend time in the end-of-the-day traffic being disgusted about all the huge, stupid, lane-blocking SUV's and Hummers and V-12 pickup trucks, burning their wasteful, selfish, 40-gallon tanks of gas, and I can dwell on the fact that the patriotic or religious bumper-stickers always seem to be on the biggest, most disgustingly selfish vehicles, driven by the ugliest — this is an example of how NOT to think, though — most disgustingly selfish vehicles, driven by the ugliest, most inconsiderate and aggressive drivers. And I can think about how our children's children will despise us for wasting all the future's fuel, and probably screwing up the climate, and how spoiled and stupid and selfish and disgusting we all are, and how modern consumer society just sucks, and so forth and so on. |
| Business | Sandberg-70 | + | It is the greatest irony of my life that losing my husband helped me find deeper gratitude- gratitude for the kindness of my friends, the love of my family, the laughter of my children. My hope for you is that you can find that gratitude-not just on the good days, like today, but on the hard ones, when you will really need it. |
| Business | BillGates-35 | - | Measles, malaria, pneumonia, hepatitis-B, yellow fever. |
| Entertainment | OBrien-23 | + | Besides policy, another hallmark of great commencement speeches is deep, profound advice like "reach for the stars." Well today, I am not going to waste your time with empty clichés. Instead, I am going to give you real, practical advice that you will need to know if you are going to survive the next few years. |
| Entertainment | Brown-44 | - | We can easily forgive a child who is afraid of the dark; the real tragedy of life is when men are afraid of the light. |
| Law and Politics | MichelleObama-10 | + | So I'm here today because I'm proud of you all. I really am. I'm inspired by your commitment to service and social justice. And I'm impressed by the community that you all have created here—a warm, supportive, inclusive community that embodies the values that define this school. |
| Law and Politics | Baron-14 | - | Leaders who crave more power for themselves always move quickly to crush an independent press. Next, they destroy free expression itself. Sadly, much of the world is on that worrisome path. And efforts in this country to demonize, delegitimize and dehumanize the press give license to other governments to do the same – and to do far worse. |

# Appendix V: The Top Documents in STM Topic Modeling

*Topic 1:*

Winfrey_10: This is what I want to share. It doesn't matter how far you might rise. At some point you are bound to stumble because if you're constantly doing what we do, raising the bar. If you're constantly pushing yourself higher, higher the law of averages not to mention the Myth of Icarus predicts that you will at some point fall. And when you do I want you to know this, remember this: there is no such thing as failure. Failure is just life trying to move us in another direction. Now when you're down there in the hole, it looks like failure. So this past year I had to spoon feed those words to myself. And when you're down in the hole, when that moment comes, it's really okay to feel bad for a little while. Give yourself time to mourn what you think you may have lost but then here's the key, learn from every mistake because every experience, encounter, and particularly your mistakes are there to teach you and force you into being more who you are. And then figure out what is the next right move. And the key to life is to develop an internal moral, emotional G.P.S. that can tell you which way to go. Because now and forever more when you Google yourself your search results will read "Harvard, 2013". And in a very competitive world that really is a calling card because I can tell you as one who employs a lot of people when I see "Harvard" I sit up a little straighter and say, "Where is he or she? Bring them in." It's an impressive calling card that can lead to even more impressive bullets in the years ahead: lawyer, senator, C.E.O., scientist, physicist, winners of Nobel and Pulitzer Prizes or late night talk show host. But the challenge of life I have found is to build a résumé that doesn't simply tell a story about what you want to be but it's a story about who you want to be. It's a résumé that doesn't just tell a story about what you want to accomplish but why. A story that's not just a collection of titles and positions but a story that's really about your purpose. Because when you inevitably stumble and find yourself stuck in a hole that is the story that will get you out. What is your true calling?

What is your dharma? What is your purpose? For me that discovery came in 1994 when I interviewed a little girl who had decided to collect pocket change in order to help other people in need. She raised a thousand dollars all by herself and I thought, well if that little 9-year-old girl with a bucket and big heart could do that, I wonder what I could do? So I asked for our viewers to take up their own change collection and in one month, just from pennies and nickels and dimes, we raised more than three million dollars that we used to send one student from every state in the United States to college. That was the beginning of the Angel Network.

*Topic 2:*

Zakaria_21: And turning to the graduates, I know I am expected to provide some advice at a commencement. Should you go into nanotechnology or bioengineering? What are the industries of the future? Honestly, I have no idea. But one thing I do know is that human beings will reward and honor those talents of heart and mind they have always honored for thousands of years: intelligence, hard work, discipline, courage, loyalty and, perhaps above all, love and a generosity of spirit. Those are the qualities that, at the end of the day, make you live a great life, one that is rewarded by the outside world, and a good life, one that is rewarded only by those who know you best. These are the virtues that people honor, that they built statues for 5,000 years ago. Well, nobody builds statues anymore. They build weird, modernist sculptures with strange pieces of metal falling off of them, but you get my idea. Trust yourself; you know what you should do. You know the kind of life you should live. You don't need an ethics course to know what you shouldn't do. Just trust in your instincts, be true to them, and you will make for yourself a great and a good life. And, in doing so, you will change the world.

*Topic 3:*

Merkel_1: Herman Hesse wrote, "In all beginnings dwells a magic force for guarding us and helping us to live." These words by Herman Hesse inspired me when I completed my physics degree at the age of 24. That was back in 1978. The world was divided into east and west, and it was in the grips of the Cold War. I grew up in East Germany, in the GDR, the part of my country which was not free at that time, in a dictatorship. People were oppressed and under state surveillance. Political dissidents were persecuted. The East German government was afraid that the people would flee to freedom. And that's why it built the Berlin Wall, a wall made of concrete and steel. Anyone caught trying to overcome it was arrested or shot dead. This wall, which cut Berlin in half, divided a people and it divided families. My family was also divided.My first job after college was as a physicist at the Academy of Sciences in East Berlin. I lived near the Berlin Wall. I walked towards it every day on my way home from my institute. Behind it lay West Berlin, freedom. And

every day, when I was very close to the wall, I had to turn away at the last minute in order to head towards my apartment. Every day, I had to turn away from freedom at the last minute. I don't know how often I thought that I just couldn't take it anymore. It was so frustrating.

*Topic 4:*

Spielberg_4: Well, what you choose to do next is what we call in the movies the "character-defining moment." Now, these are moments you're very familiar with, like in the last "Star Wars: The Force Awakens," when Rey realizes the force is with her. Or Indiana Jones choosing mission over fear by jumping over a pile of snakes. Now in a two-hour movie, you get a handful of character-defining moments, but in real life, you face them every day. Life is one strong, long string of character-defining moments. And I was lucky that at 18 I knew what I exactly wanted to do. But I didn't know who I was. How could I? And how could any of us? Because for the first 25 years of our lives, we are trained to listen to voices that are not our own. Parents and professors fill our heads with wisdom and information, and then employers and mentors take their place and explain how this world really works. And usually these voices of authority make sense, but sometimes, doubt starts to creep into our heads and into our hearts. And even when we think, "that's not quite how I see the world," it's kind of easier to just to nod in agreement and go along, and for a while, I let that going along define my character. Because I was repressing my own point of view, because like in that Nilsson song, "Everybody was talkin' at me, so I couldn't hear the echoes of my mind." And at first, the internal voice I needed to listen to was hardly audible, and it was hardly noticeable — kind of like me in high school.

*Topic 5:*

Winfrey_16: I know that you all might have a little anxiety now and hesitation about leaving the comfort of college and putting those Harvard credentials to the test. But no matter what challenges or setbacks or disappointments you may encounter along the way, you will find true success and happiness if you have only one goal, there really is only one, and that is this: to fulfill the highest most truthful expression of yourself as a human being. You want to max out your humanity by using your energy to lift yourself up, your family and the people around you. Theologian Howard Thurman said it best. He said, "Don't ask yourself what the world needs. Ask yourself what makes you come alive and then go do that, because what the world needs is people who have come alive." The world needs … People like Michael Stolzenberg from Fort Lauderdale. When Michael was just 8 years old Michael nearly died from a bacterial infection that cost him both of his hands and both of his feet. And in an instant, this vibrant little boy became a quadruple amputee and his life was changed forever. But in losing who he once was

Michael discovered who he wanted to be. He refused to sit in that wheelchair all day and feel sorry for himself so with prosthetics he learned to walk and run and play again. He joined his middle school lacrosse team and last month when he learned that so many victims of the Boston Marathon bombing would become new amputees, Michael decided to banish that darkness with light. Michael and his brother, Harris, created Mikeysrun.com to raise $1 million for other amputees — by the time Harris runs the 2014 Boston Marathon. More than 1,000 miles away from here these two young brothers are bringing people together to support this Boston community the way their community came together to support Michael. And when this 13-year-old man was asked about his fellow amputees he said this, "First they will be sad. They're losing something they will never get back and that's scary. I was scared. But they'll be okay. They just don't know that yet." We might not always know it. We might not always see it, or hear it on the news or even feel it in our daily lives, but I have faith that no matter what, Class of 2013, you will be okay and you will make sure our country is okay. I have faith because of that 9-year-old girl who went out and collected the change. I have faith because of David and Francine Wheeler, I have faith because of Michael and Harris Stolzenberg, and I have faith because of you, the network of angels sitting here today. One of them Khadijah Williams, who came to Harvard four years ago. Khadijah had attended 12 schools in 12 years, living out of garbage bags amongst pimps and prostitutes and drug dealers; homeless, going in to department stores, Wal-Mart in the morning to bathe herself so that she wouldn't smell in front of her classmates, and today she graduates as a member of the Harvard Class of 2013.

*Topic 6:*

Simmons_14: In that vein, Harvard has a special responsibility as both a prod and steward of the national conscience. It could sit on the hill and congratulate itself on its prowess but it could also use its immense stature to address the widening gaps in how different groups experience freedom and justice. I spoke earlier about the heroic work of HBCUs and minority serving institutions that keep our country open and advancing the cause of equality and access. Yet, many of them have been starved for much of their history by the legacy of underfunding and isolation from the mainstream of higher education.

*Topic 7:*

Jobs_25: Stewart and his team put out several issues of The Whole Earth Catalog, and then when it had run its course, they put out a final issue. It was the mid-1970s, and I was your age. On the back cover of their final issue was a photograph of an early morning country road, the kind you might find yourself hitchhiking on if you were so adventurous.

Beneath it were the words: “Stay Hungry. Stay Foolish.” It was their farewell message as they signed off. Stay Hungry. Stay Foolish. And I have always wished that for myself. And now, as you graduate to begin anew, I wish that for you.

# Appendix VI: The Top Documents in Top2Vec Topic Modeling

Top2vec model with 5 topics Topic distribution: 0.57 0.12 0.12 0.13 0.06

*Topic 0:*

Terms: distance, gradually, end, gates, rich, besides, food, of, sacrifice, involves, plans, enable, reached, music, intelligence.

Top docs:

(1) ardern_61: it means that there is a pressing and urgent need for responsible algorithm development and deployment.
(2) brown_69: a good traveler has no fixed plans and is not intent on arriving.
(3) zuckerberg_72: this is my story too. a student in a dorm room connecting one community at a time and keeping at it until one day we connect the whole world.
(4) ardern_59: that means upholding their own basic terms of service.
(5) wallace_15: everyone here has done this of course. but it hasn't yet been part of you graduates' actual life routine day after week after month after year.
(6) ardern_49: we found a place to experience new ways of thinking and to celebrate our difference.

*Topic 1:*

Terms: you, compare, your, follow, happiness, worship, will, courageous, defined, correct, inexperience, true, best, won't, worry.

Top docs:

(1) brown_62: when you are content to be simply yourself and don't compare or compete everybody will respect you. (right?)
(2) bezos_32: when it's tough will you give up or will you be relentless? will you be a cynic or will you be a builder? will you be clever at the expense of others or will you be kind?

(3) bezos_31: will you wilt under criticism or will you follow your convictions? will you bluff it out when you're wrong or will you apologize? will you guard your heart against rejection or will you act when you fall in love? will you play it safe or will you be a little swashbuckling?
(4) gaiman_11: sometimes the way to do what you hope to do will be clear cut and sometimes it will be almost impossible to decide whether or not you are doing the correct thing because you'll have to balance your goals and hopes with feeding yourself paying debts finding work settling for what you can get.
(5) bezos_30: tomorrow in a very real sense your life the life you author from scratch on your own begins. how will you use your gifts? what choices will you make? will inertia be your guide or will you follow your passions? will you follow dogma or will you be original? will you choose a life of ease or a life of service and adventure?
(6) zuckerberg_20: i know a lot of you will have your own stories just like this. a change in the world that seems so clear you're sure someone else will do it. but they won't. you will.

*Topic 2:*

Terms: stronger, potential, our, teams, mainstream, reduce, assured, we, governments, hbcus, requires, actions, ends, charged, diversity.

Top docs:

(1) bloomberg_70: because ultimately it is only through the power of our relationships that we can fulfill our potential and that we can build teams capable of creating a society that is more just more equal more peaceful and – in the words of our nation's founders – more perfect.
(2) zuckerberg_66: we all get meaning from our communities. whether our communities are houses or sports teams churches or music groups they give us that sense we are part of something bigger that we are not alone; they give us the strength to expand our horizons.
(3) ardern_59: that means upholding their own basic terms of service.
(4) zuckerberg_62: we get that our greatest opportunities are now global — we can be the generation that ends poverty that ends disease. we get that our greatest challenges need global responses too — no country can fight climate change alone or prevent pandemics. progress now requires coming together not just as cities or nations but also as a global community.
(5) zuckerberg_73: change starts local. even global changes start small — with people like us. in our generation the struggle of whether we connect more whether we

achieve our biggest opportunities comes down to this — your ability to build communities and create a world where every single person has a sense of purpose.

(6) winfrey_12: and we understand that most americans believe in a clear path to citizenship for the 12000000 undocumented immigrants who reside in this country because it's possible to both enforce our laws and at the same time embrace the words on the statue of liberty that have welcomed generations of huddled masses to our shores. we can do both.

*Topic 3:*

Terms: was, finished, convinced, wondered, hadn't, kept, took, kicked, caught, reading, went, seemed, wanted, novel, launched.

Top docs:

(1) degeneres_9: anyway i had no idea what i wanted to do with my life and the way i ended up on this path was from a very tragic event. i was maybe 19 and my girlfriend at the time was killed in a car accident. and i passed the accident and i didn't know it was her and i kept going and i found out shortly after that it was her.

(2) bezos_7: i have a very vivid memory of what happened next. and it was not what i had expected. i expected to be applauded for my cleverness and my arithmetic skills.

(3) portman_19: i was completely overwhelmed and thought that reading 1000 pages a week was unimaginable that writing a 50-page thesis was just something i could never do. i had no idea how to declare my intentions. i couldn't even articulate them to myself.

(4) jobs_20: i lived with that diagnosis all day. later that evening i had a biopsy where they stuck an endoscope down my throat through my stomach and into my intestines put a needle into my pancreas and got a few cells from the tumor. i was sedated but my wife who was there told me that when they viewed the cells under a microscope the doctors started crying because it turned out to be a very rare form of pancreatic cancer that is curable with surgery. i had the surgery and i'm fine now.

(5) musk_2: when i was young i didn't really know what i was going to do when i got older. people kept asking me. but then eventually i thought the idea of inventing things would be really cool.

(6) zuckerberg_77: last year i took him out to breakfast for his birthday. i wanted to get him a present so i asked him and he started talking about students he saw struggling and said "you know i'd really just like a book on social justice."

*Topic 4:*

Terms: distinguished, thank, overseers, trustees, tessier-lavigne, ladies, inviting, guests, faculty, congratulations, introduction, honored, graduates, staff, faust.

Top docs:

(1) cook_82: thank you very much. and congratulations to the class of 2019!
(2) portman_2: i am so honored to be here today. dean khurana faculty parents and most especially graduating students thank you so much for inviting me.
(3) zakaria_1: thank you so much president faust fellows of the corporation overseers ladies and gentlemen and graduates.
(4) sandberg_2: thank you marie. and thank you esteemed members of the faculty proud parents devoted
(5) simmons_1: good day and congratulations to the harvard university class of 2021.
(6) zakaria_23: thank you ladies and gentlemen and to the graduates of harvard university's class of 2012 godspeed.

# Bibliography


Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2022). Topic modeling algorithms and applications: A survey. *Information Systems*, 102131 (cit. on p. 25).

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3–9 (cit. on p. 7).

Adams, A. (2007). Developing leadership wisdom. *International Journal of Leadership in Public Services*, *3*(2), 39–50 (cit. on p. 12).

Adichie, C. N. (2018). Full transcript: Chimamanda ngozi adichie's harvard 2018 speech [accessed 2-April-2023]. https://singjupost.com/full-transcript-chimamanda-ngozi-adichies-harvard-2018-speech/ (cit. on p. 107).

Angelov, D. (2020). Top2vec: Distributed representations of topics (cit. on pp. 20, 27, 28).

Anuradha, J., et al. (2015). A brief introduction on big data 5vs characteristics and hadoop technology. *Procedia computer science*, *48*, 319–324 (cit. on p. 20).

Ardelt, M. (2003). Empirical assessment of a three-dimensional wisdom scale. *Research on aging*, *25*(3), 275–324 (cit. on p. 13).

Ardelt, M. (2004a). Where can wisdom be found? *Human Development*, *47*(5), 304–307 (cit. on pp. 7, 11).

Ardelt, M. (2004b). Wisdom as expert knowledge system: A critical review of a contemporary operationalization of an ancient concept. *Human development*, *47*(5), 257–285 (cit. on p. 6).

Ardelt, M., Pridgen, S., & Nutter-Pridgen, K. L. (2019). Wisdom as a personality type. (cit. on p. 7).

Ardern, J. (2022). Harvard commencement speech - democracy, disinformation and kindness [accessed 2-April-2023]. https://www.beehive.govt.nz/speech/harvard-commencement-speech-democracy-disinformation-and-kindness (cit. on p. 107).

Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, *4*(3), 181–186 (cit. on p. 23).

Atiq, A., Abeed, A. S., Efat, A. A., & Momin, A. (2022). *Sentimental analysis on political speeches* [Doctoral dissertation, Brac University]. (Cit. on p. 18).

Audi, R. (1995). The cambridge dictionary of philosophy (cit. on p. 6).

Bachmann, C., Sasse, L., & Habisch, A. (2018). Applying the practical wisdom lenses in decision-making: An integrative approach to humanistic management. *Humanistic Management Journal*, *2*, 125–150 (cit. on p. 6).

Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, *18*(1998), 1–8 (cit. on p. 27).

Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: A comparison of retrieval performances (cit. on p. 23).

Baltes, P. B., & Kunzmann, U. (2004). The two faces of wisdom: Wisdom as a general theory of knowledge and judgment about excellence in mind and virtue vs. wisdom as everyday realization in people and products. *Human Development*, *47*(5), 290–299 (cit. on p. 7).

Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American psychologist*, *55*(1), 122 (cit. on p. 10).

Baron, M. (2020). Washington post executive editor martin baron delivers address to harvard university 2020 graduates [accessed 2-April-2023]. https://www.washingtonpost.com/pr/2020/05/28/washington-post-executive-editor-martin-baron-delivers-harvard-university-commencement-speech/ (cit. on p. 107).

Bennis, W. (2007). The challenges of leadership in the modern world: Introduction to the special issue. *American psychologist*, *62*(1), 2 (cit. on p. 6).

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2021). *Quanteda: Quantitative analysis of textual data* [R package version 3.3.0]. https://CRAN.R-project.org/package=quanteda (cit. on p. 41).

Bergsma, A., & Ardelt, M. (2012). Self-reported wisdom and happiness: An empirical investigation. *Journal of Happiness Studies*, *13*, 481–499 (cit. on p. 7).

Bezos, J. (2010). Amazon ceo jeff bezos princeton commencement speech transcript 2010 [accessed 2-April-2023]. https://www.rev.com/blog/transcripts/amazon-ceo-jeff-bezos-commencement-speech-transcript-princeton-2010 (cit. on p. 107).

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84 (cit. on p. 27).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022 (cit. on pp. 20, 26, 27, 38).

Bloomberg, M. (2021). Mike bloomberg's remarks at johns hopkins university commencement 2021 [accessed 2-April-2023]. https://www.mikebloomberg.com/news/mike-bloombergs-remarks-at-johns-hopkins-university-commencement-2021/ (cit. on p. 107).

Bluck, S., & Glück, J. (2005). From the inside out: People's implicit theories of wisdom. (cit. on p. 8).

Boyd-Graber, J., Mimno, D., & Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. *Handbook of mixed membership models and their applications*, *225255* (cit. on p. 21).

Brooks, E. (2021). A new generation of wise thinkers and good leaders: What can character contribute? *Journal of Character and Leadership Development*, *8*(2) (cit. on p. 6).

Brown, S. K. (2018). 2018 commencement speech by stanford alumnus sterling k. brown [accessed 2-April-2023]. https://news.stanford.edu/2018/06/17/2018-commencement-speech-stanford-alumnus-sterling-k-brown/ (cit. on p. 107).

Brown, W. S. (2005). Discussion: Seven pillars of the house of wisdom. *A handbook of wisdom: Psychological perspectives*, 353–368 (cit. on p. 8).

Burns, K. (2016). Prepared text of the 2016 stanford commencement address by ken burns [accessed 2-April-2023]. https://news.stanford.edu/2016/06/12/prepared-text-2016-stanford-commencement-address-ken-burns/ (cit. on p. 107).

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17*, 160–172 (cit. on p. 27).

Chehri, A., & Mouftah, H. T. (2019). Autonomous vehicles in the sustainable cities, the beginning of a green adventure. *Sustainable Cities and Society*, *51*, 101751 (cit. on p. 6).

Chen, K., Molder, A. L., Duan, Z., Boulianne, S., Eckart, C., Mallari, P., & Yang, D. (2023). How climate movement actors and news media frame climate change and strike: Evidence from analyzing twitter and news media discourse from 2018 to 2021. *The International Journal of Press/Politics*, *28*(2), 384–413 (cit. on p. 18).

Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, *13*, e00151 (cit. on p. 6).

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, *74*(368), 829–836 (cit. on p. 63).

Cook, T. (2019). 2019 commencement address by apple ceo tim cook [accessed 2-April-2023]. https://news.stanford.edu/2019/06/16/remarks-tim-cook-2019-stanford-commencement/ (cit. on p. 107).

Cordova, F. (2020). Commencement address to class of 2020 by france cordova [accessed 2-April-2023]. https://news.stanford.edu/2022/06/11/commencement-address-class-2020-france-cordova/ (cit. on p. 107).

Cuellar, M. F. (2017). Prepared text of the 2017 stanford commencement address by mariano-florentino cuellar [accessed 2-April-2023]. https://news.stanford.edu/2017/06/18/prepared-text-2017-stanford-commencement-address-mariano-florentino-cuellar/ (cit. on p. 107).

Dan, D. (2023). Exploring the impact of negative words used in online feedback in hotel industry: A sentiment analysis, n-gram, and text network analysis approach (cit. on p. 18).

Dashtipour, K., Gogate, M., Adeel, A., Hussain, A., Alqarafi, A., & Durrani, T. (2019). A comparative study of persian sentiment analysis based on different feature combinations. *Communications, Signal Processing, and Systems: Proceedings of the 2017 International Conference on Communications, Signal Processing, and Systems*, 2288–2294 (cit. on p. 22).

DeGeneres, E. (2009). Ellen degeneres' 2009 commencement speech at tulane university [accessed 2-April-2023]. https://time.com/4336922/obama-commencement-speech-transcript-notre-dame/ (cit. on p. 107).

Deng, L., & Liu, Y. (2018). A joint introduction to natural language processing and to deep learning. *Deep learning in natural language processing*, 1–22 (cit. on p. 19).

Desai, A., et al. (2015). A review on knowledge discovery using text classification techniques in text mining. *International Journal of Computer Applications*, *111*(6) (cit. on p. 20).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (cit. on p. 27).

Doogan, C. (2022). *A topic is not a theme: Towards a contextualised approach to topic modelling* [Doctoral dissertation, Monash University]. (Cit. on p. 20).

Durant, W. (1961). *Story of philosophy*. Simon; Schuster. (Cit. on p. 8).

Edmondson, A. C. (2011). Strategies for learning from failure. *Harvard business review*, *89*(4), 48–55 (cit. on p. 6).

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7* (cit. on p. 28).

Engel, R. (2015). Prepared text of the 2015 stanford commencement address by richard engel [accessed 2-April-2023]. https://news.stanford.edu/2015/06/14/engel-commence-text-061415/ (cit. on p. 107).

Feinerer, I., Hornik, K., & Meyer, D. (2021). *Tm: Text mining package* [R package version 0.7-11]. https://CRAN.R-project.org/package=tm (cit. on p. 41).

Gaiman, N. (2012). The 134th commencement of university of the arts. [accessed 2-April-2023]. https://www.uarts.edu/makegoodart (cit. on p. 107).

Gates, B. (2007). Bill gates' commencement speech at harvard 2007 (full transcript) [accessed 2-April-2023]. https://singjupost.com/bill-gates-commencement-speech-at-harvard-2007-full-transcript/ (cit. on p. 107).

Gawande, A. (2021). 2021 commencement address by dr. atul gawande [accessed 2-April-2023]. https://news.stanford.edu/2021/06/12/2021-commencement-address-dr-atul-gawande/ (cit. on p. 107).

Girju, C. R. (2002). *Text mining for semantic relations*. The University of Texas at Dallas. (Cit. on p. 19).

Glazkova, A. (2021). Identifying topics of scientific articles with bert-based approaches and topic modeling. *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2021 Workshops, WSPA, MLMEIN, SDPRA, DARAI, and AI4EPT, Delhi, India, May 11, 2021 Proceedings 25*, 98–105 (cit. on p. 18).

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press. (Cit. on p. 18).

Goodman, J. T. (2001). A bit of progress in language modeling. *Computer Speech & Language*, *15*(4), 403–434 (cit. on p. 21).

Goyal, A., Gupta, V., & Kumar, M. (2018). Recent named entity recognition and classification techniques: A systematic review. *Computer Science Review*, *29*, 21–43 (cit. on p. 20).

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, *101*(suppl_1), 5228–5235 (cit. on p. 27).

Grossmann, I., & Brienza, J. P. (2018). The strengths of wisdom provide unique contributions to improved leadership, sustainability, inequality, gross national happiness, and civic discourse in the face of contemporary world problems. *Journal of Intelligence*, *6*(2), 22 (cit. on pp. 11, 12).

Grossmann, I., Weststrate, N. M., Ardelt, M., Brienza, J. P., Dong, M., Ferrari, M., Fournier, M. A., Hu, C. S., Nusbaum, H. C., & Vervaeke, J. (2020). The science of wisdom in a polarized world: Knowns and unknowns. *Psychological Inquiry*, *31*(2), 103–133 (cit. on pp. 10, 11).

Hall, S. S. (2011). *Wisdom: From philosophy to neuroscience*. Vintage. (Cit. on p. 9).

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162 (cit. on pp. 19, 24).

Hastings, R. (2022). Commencement address to class of 2022 by reed hastings [accessed 2-April-2023]. https://news.stanford.edu/2022/06/12/commencement-address-class-2022-reed-hastings/ (cit. on p. 107).

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, *20*(1), 19–62 (cit. on p. 19).

Hughes, G. J. (2001). *Routledge philosophy guidebook to aristotle on ethics*. Psychology Press. (Cit. on p. 8).

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169–15211 (cit. on p. 25).

Jivani, A. G., et al. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, *2*(6), 1930–1938 (cit. on p. 23).

Jobs, S. (2005). 'you've got to find what you love,' jobs says [accessed 2-April-2023]. https://news.stanford.edu/2005/06/12/youve-got-find-love-jobs-says/ (cit. on p. 107).

Johnstone, B. (2017). *Discourse analysis*. John Wiley & Sons. (Cit. on p. 19).

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, *28*(1), 11–21 (cit. on p. 24).

Jones, K. S. (1973). Index term weighting. *Information storage and retrieval*, *9*(11), 619–633 (cit. on p. 24).

Jones, K. S. (2004). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, *60*(5), 493–502 (cit. on p. 24).

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* (cit. on p. 25).

Kalyanathaya, K. P., Akila, D., & Rajesh, P. (2019). Advances in natural language processing–a survey of current research trends, development tools and industry applications. *International Journal of Recent Technology and Engineering*, *7*(5C), 199–202 (cit. on p. 37).

Karami, S., Ghahremani, M., Parra-Martinez, F. A., & Gentry, M. (2020). A polyhedron model of wisdom: A systematic review of the wisdom studies in psychology, management and leadership, and education. *Roeper Review*, *42*(4), 241–257 (cit. on pp. 2, 10).

Kaur, J., & Buttar, P. K. (2018). A systematic review on stopword removal algorithms. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 207–210 (cit. on p. 24).

Kenny, A. (2012). *A new history of western philosophy*. OUP Oxford. (Cit. on p. 8).

Kherwa, P., & Bansal, P. (2019). Topic modeling: A comprehensive review. *EAI Endorsed transactions on scalable information systems*, *7*(24) (cit. on p. 20).

Knepper, S. (2020). Gabriel marcel: Mystery in an age of problems. *Critics of Enlightenment Rationalism*, 125–137 (cit. on p. 15).

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. (2018). Text classification for organizational researchers: A tutorial. *Organizational research methods*, *21*(3), 766–799 (cit. on p. 20).

Kunzmann, U., & Baltes, P. B. (2005). The psychology of wisdom: Theoretical and empirical challenges. (cit. on p. 8).

Law, A., & Staudinger, U. M. (2016). Eudaimonia and wisdom. *Handbook of eudaimonic well-being*, 135–146 (cit. on p. 10).

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *International conference on machine learning*, 1188–1196 (cit. on pp. 25, 27, 38).

Lewis, J. (2018). 6 past harvard commencement speakers offer inspiring messages of justice, courage, resilience, empathy [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2022/05/6-memorable-harvard-commencement-speeches/ (cit. on p. 107).

Li, W.-J., Wang, K., Stolfo, S. J., & Herzog, B. (2005). Fileprints: Identifying file types by n-gram analysis. *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop*, 64–71 (cit. on p. 22).

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1), 1–22 (cit. on p. 25).

Lopez-Martinez, R. E., & Sierra, G. (2020). Natural language processing, 2000-2019 a bibliometric study. *Journal of Scientometric Research*, *9*(3), 310–318 (cit. on p. 24).

Marcon, E. (2022). *Memoir: R markdown and bookdown templates to publish documents* [R package version 1.2-2]. https://github.com/EricMarcon/memoiR (cit. on p. 41).

May, C., Cotterell, R., & Van Durme, B. (2016). An analysis of lemmatization on topic models of morphologically rich language. *arXiv preprint arXiv:1608.03995* (cit. on p. 23).

Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution (cit. on pp. 19, 20).

McKenna, B. (2013). The multi-dimensional character of wisdom. *Wise management in organisational complexity*, 13–33 (cit. on pp. 6, 7, 9, 11).

McKenna, B., Rooney, D., & Boal, K. B. (2009). Wisdom principles as a meta-theoretical basis for evaluating leadership. *The Leadership Quarterly*, *20*(2), 177–190 (cit. on p. 12).

Meacham, J. A. (1983). Wisdom and the context of knowledge: Knowing that one doesn't know. *On the development of developmental psychology*, *8*(111-134), 1 (cit. on pp. 9, 53).

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093–1113 (cit. on p. 19).

Merkel, A. (2019). 6 past harvard commencement speakers offer inspiring messages of justice, courage, resilience, empathy [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2022/05/6-memorable-harvard-commencement-speeches/ (cit. on p. 107).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26* (cit. on pp. 25, 27).

Minka, T. (2000). Estimating a dirichlet distribution. (Cit. on p. 26).

Mitchell, L. K., Knight, B. G., & Pachana, N. A. (2017). Wisdom across the ages and its modern day relevance. *International Psychogeriatrics*, *29*(8), 1231–1234 (cit. on p. 7).

Moreno, A., & Redondo, T. (2016). Text analytics: The convergence of big data and artificial intelligence. *IJIMAI*, *3*(6), 57–64 (cit. on p. 19).

Müller-Hansen, F., Callaghan, M. W., Lee, Y. T., Leipprand, A., Flachsland, C., & Minx, J. C. (2021). Who cares about coal? analyzing 70 years of german parliamentary debates on coal with dynamic topic modeling. *Energy Research & Social Science*, *72*, 101869 (cit. on p. 18).

Murzintcev, N. (2021). *Ldatuning: Tuning of the latent dirichlet allocation models parameters* [R package version 0.2.1]. https://CRAN.R-project.org/package=ldatuning (cit. on p. 65).

Musk, E. (2012). Elon musk's commencement speech at caltech [accessed 2-April-2023]. https://singjupost.com/elon-musks-commencement-speech-at-caltech-full-transcript/ (cit. on p. 107).

Nayak, A. S., Kanive, A. P., Chandavekar, N., & Balasubramani, R. (2016). Survey on pre-processing techniques for text mining. *International Journal of Engineering and Computer Science*, *5*(6), 16875–16879 (cit. on p. 19).

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P., & Hluchỳ, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. *Artificial Intelligence Review*, *52*, 77–124 (cit. on p. 5).

Nonaka, I., & Takeuchi, H. (2011). The wise leader. *Harvard business review*, *89*(5), 58–67 (cit. on pp. 12, 13).

Nonaka, I., & Takeuchi, H. (2019). *The wise company: How companies create continuous innovation*. Oxford University Press. (Cit. on p. 11).

Obama, B. (2009). Read president obama's commencement address at notre dame [accessed 2-April-2023]. https://time.com/4336922/obama-commencement-speech-transcript-notre-dame/ (cit. on p. 107).

Obama, M. (2015). Transcript of first lady michelle obama's commencement address [accessed 2-April-2023]. https://www.oberlin.edu/news/transcript-first-lady-michelle-obamas-commencement-address (cit. on p. 107).

O'Brien, C. (2011). Conan o'brien's dartmouth commencement speech 2011 [accessed 2-April-2023]. https://singjupost.com/conan-obriens-dartmouth-commencement-speech-2011-full-transcript/ (cit. on p. 107).

O'Hara, C. (2014). How to tell a great story. *Harvard Business Review*, *30* (cit. on p. 63).

*Online etymology dictionary* [Accessed: 2023-03-22]. (n.d.). https://www.etymonline.com/word/wisdom (cit. on p. 7).

Pasupathi, M., Staudinger, U. M., & Baltes, P. B. (2001). Seeds of wisdom: Adolescents' knowledge and judgment about difficult life problems. *Developmental psychology*, *37*(3), 351 (cit. on p. 10).

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543 (cit. on p. 25).

Portman, N. (2015). Natalie portman harvard commencement speech 2015 [accessed 2-April-2023]. https://singjupost.com/full-transcript-natalie-portman-harvard-commencement-speech-2015/ (cit. on p. 107).

Posit team. (2023). *Rstudio: Integrated development environment for r*. Posit Software, PBC. Boston, MA. http://www.posit.co/ (cit. on p. 40).

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/ (cit. on p. 40).

Rahmoun, A., & Elberrichi, Z. (2007). Experimenting n-grams in text categorization. *Int. Arab J. Inf. Technol.*, *4*(4), 377–385 (cit. on p. 21).

Rai, A., & Borah, S. (2021). Study of various methods for tokenization. *Applications of Internet of Things: Proceedings of ICCCIOT 2020*, 193–200 (cit. on p. 21).

Rajasundari, T., Subathra, P., & Kumar, P. (2017). Performance analysis of topic modeling algorithms for news articles. *Journal of Advanced Research in Dynamical and Control Systems*, *2017*(11), 175–183 (cit. on p. 18).

Rayson, P. E. (2015). Computational tools and methods for corpus compilation and analysis (cit. on p. 37).

Riggio, R. E., Zhu, W., Reina, C., & Maroosis, J. A. (2010). Virtue-based measurement of ethical leadership: The leadership virtues questionnaire. *Consulting Psychology Journal: Practice and Research*, *62*(4), 235 (cit. on pp. 11, 12).

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, *91*, 1–40 (cit. on pp. 20, 28, 38, 70).

Rocha, R., & Pinheiro, P. (2021). Business education: Filling the gaps in the leader's awareness concerning organizational phronesis. *Sustainability*, *13*(4), 2274 (cit. on p. 13).

Rooney, D., Küpers, W., Pauleen, D., & Zhuravleva, E. (2021). A developmental model for educating wise leaders: The role of mindfulness and habitus in creating time for embodying wisdom. *Journal of Business Ethics*, *170*, 181–194 (cit. on p. 12).

Rooney, D., & McKenna, B. (2022). Wisdom and leadership. In R. J. Sternberg & J. Glück (Eds.), *The psychology of wisdom: An introduction* (pp. 230–244). Cambridge University Press. (Cit. on p. 13).

Ross, D., Ross, W. D., & Brown, L. (2009). *The nicomachean ethics*. Oxford University Press, USA. (Cit. on p. 8).

Rowley, J. (2007). The wisdom hierarchy: Representations of the dikw hierarchy. *Journal of information science*, *33*(2), 163–180 (cit. on p. 7).

Rowling, J. K. (2008). Text of j.k. rowling's speech [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2008/06/text-of-j-k-rowling-speech/ (cit. on p. 107).

Sacks, H. (1992). Lectures on conversation (two volumes). (Cit. on p. 19).

Salim, M. N., & Mustafa, B. S. (2022). A survey on word representation in natural language processing. *AIP Conference Proceedings*, *2394*(1), 050006 (cit. on p. 24).

Sandberg, S. (2016). University of california, berkeley 2016 commencement address [accessed 2-April-2023]. https://news.berkeley.edu/wp-content/uploads/2016/05/Sheryl-Sandberg-Berkeley-commencement-speech.pdf (cit. on p. 107).

Schoemaker, P. J., Heaton, S., & Teece, D. (2018). Innovation, dynamic capabilities, and leadership. *California management review*, *61*(1), 15–42 (cit. on pp. 1, 23).

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press Cambridge. (Cit. on p. 23).

Schwartz, B., & Sharpe, K. (2010). *Practical wisdom: The right way to do the right thing*. Penguin. (Cit. on p. 12).

Selivanov, D., & Feuerriegel, S. (2021). *Text2vec: Modern text mining framework for r* [R package version 0.6.3]. https://CRAN.R-project.org/package=text2vec (cit. on p. 41).

Sheppard, B. (2022). Learning to learn: Pwc. In S. Iñiguez & P. Lorange (Eds.), *Executive education after the pandemic: A vision for the future* (pp. 191–199). Springer International Publishing. https://doi.org/10.1007/978-3-030-82343-6_19 (cit. on p. 14).

Silge, J., & Robinson, D. (2021). *Tidytext: Text mining and analysis using tidy data principles in r* [R package version 0.4.0.9000]. https://CRAN.R-project.org/package=tidytext (cit. on p. 40).

Simmons, R. J. (2021). Ruth j. simmons's harvard graduation address [accessed 2-April-2023]. https://www.harvardmagazine.com/2021/05/commencement-morning-exercises-guest-speaker-ruth-j-simmmons (cit. on p. 107).

Spielberg, S. (2016). Steven spielberg speech at harvard commencement 2016 [accessed 2-April-2023]. https://singjupost.com/steven-spielberg-speech-at-harvard-commencement-2016-full-transcript/ (cit. on p. 107).

Staudinger, U. M., & Glück, J. (2011). Psychological wisdom research: Commonalities and differences in a growing field. *Annual review of psychology*, *62*, 215–241 (cit. on p. 6).

Stergiou, C., Psannis, K. E., Gupta, B. B., & Ishibashi, Y. (2018). Security, privacy & efficiency of sustainable cloud computing for big data & iot. *Sustainable Computing: Informatics and Systems*, *19*, 174–184 (cit. on p. 5).

Sternberg, R. J. (2003). Wics: A model of leadership in organizations. *Academy of Management Learning & Education*, *2*(4), 386–401 (cit. on p. 12).

Sternberg, R. J., & Glück, J. (2022). *The psychology of wisdom: An introduction*. Cambridge University Press. (Cit. on p. 10).

Sternberg, R. J., & Jordan, J. (2005). *A handbook of wisdom: Psychological perspectives*. Cambridge University Press. (Cit. on p. 8).

Sternberg, R. J., & Karami, S. (2021). What is wisdom? a unified 6p framework. *Review of General Psychology*, *25*(2), 134–151 (cit. on p. 11).

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. (cit. on p. 18).

Taranto, M. A. (1989). Facets of wisdom: A theoretical synthesis. *International Journal of Aging and Human Development*, *29*, 1–21 (cit. on p. 6).

Vaccarezza, M. S., & De Caro, M. (2021). Practical wisdom: Philosophical and psychological perspectives (cit. on pp. 9, 11).

Vanharanta, H., Einolander, J., Kantola, J., Markopoulos, E., & Sivula, A. (2021). Phronetic leadership style evaluation with a fuzzy logic application. *TheoreTical issues in ergonomics science*, *22*(3), 317–337 (cit. on p. 2).

Vanharanta, H., & Markopoulos, E. (2020). Visualization of the wisdom cube scientific knowledge space for management and leadership. *Advances in Human Factors, Business Management and Leadership: Proceedings of the AHFE 2019 International Conference on Human Factors, Business Management and Society, and the AHFE International Conference on Human Factors in Management and Leadership, July 24-28, 2019, Washington DC, USA 10*, 14–25 (cit. on p. 2).

Vijayarani, S., Ilamathi, M. J., Nithya, M., et al. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16 (cit. on p. 19).

Vijayarani, S., Janani, R., et al. (2016). Text mining: Open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, *3*(1), 37–47 (cit. on p. 21).

Wallace, F. (2005). This is water by david foster wallace (full transcript and audio) [accessed 2-April-2023]. https://fs.blog/david-foster-wallace-this-is-water/ (cit. on p. 107).

Wang, F., Yang, Y., Tso, G. K., & Li, Y. (2019). Analysis of launch strategy in cross-border e-commerce market via topic modeling of consumer reviews. *Electronic Commerce Research*, *19*, 863–884 (cit. on p. 18).

Wang, Y., Bowers, A. J., & Fikis, D. J. (2017). Automated text data mining analysis of five decades of educational leadership research literature: Probabilistic topic modeling of eaq articles from 1965 to 2014. *Educational administration quarterly*, *53*(2), 289–323 (cit. on p. 18).

Webster, J. D. (2003). An exploratory analysis of a self-assessed wisdom scale. *Journal of Adult Development*, *10*, 13–22 (cit. on p. 13).

Whiteman, W. E. (1998). *Training and educating army officers for the 21st century: Implications for the united states military academy.* (tech. rep.). Army War Coll Carlisle Barracks PA. (Cit. on p. 5).

Wickham, H. (2021). *Ggplot2: Elegant graphics for data analysis* [R package version 3.4.0]. https://CRAN.R-project.org/package=ggplot2 (cit. on p. 41).

Wickham, H., & RStudio. (2021). *Tidyverse: Easily install and load the 'tidyverse'* [R package version 1.3.2]. https://CRAN.R-project.org/package=tidyverse (cit. on p. 41).

Wiedemann, G., & Wiedemann. (2016). *Text mining for qualitative data analysis in the social sciences* (Vol. 1). Springer. (Cit. on p. 20).

Winfrey, O. (2013). Winfrey's commencement address [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2013/05/winfreys-commencement-address/ (cit. on p. 107).

Yuan, L. (2013). Wise and virtuous leadership: The contribution of confucian values to business leadership. *Wise management in organisational complexity*, 106–121 (cit. on p. 9).

Yuan, L., Chia, R., & Gosling, J. (2023). Confucian virtue ethics and ethical leadership in modern china. *Journal of Business Ethics*, *182*(1), 119–133 (cit. on pp. 8, 9).

Zakaria, F. (2012). Text of zakaria's commencement address [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2012/05/text-of-fareed-zakarias-commencement-address/ (cit. on p. 107).

Žižka, J., Dařena, F., & Svoboda, A. (2019). *Text mining with machine learning: Principles and techniques*. Crc Press. (Cit. on p. 21).

Zuckerberg, M. (2017). Mark zuckerberg's commencement address at harvard [accessed 2-April-2023]. https://news.harvard.edu/gazette/story/2017/05/mark-zuckerbergs-speech-as-written-for-harvards-class-of-2017/ (cit. on p. 107).

Zvornicanin, E. (2022). *Topic modeling and latent dirichlet allocation (lda)* [Accessed: 2023-05-06]. https://datascienceplus.com/topic-modeling-and-latent-dirichlet-allocation-lda/ (cit. on p. 26).