CROSS-DISCIPLINARITY IN COGNITIVE SCIENCE: A DOCUMENT SIMILARITY ANALYSIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

OĞUZHAN ALAŞEHİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

January 2024

Approval of the thesis:

CROSS-DISCIPLINARITY IN COGNITIVE SCIENCE: A DOCUMENT SIMILARITY ANALYSIS

Submitted by OĞUZHAN ALAŞEHİR in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics** ____________________

Prof. Dr. Altan Koçyiğit
Head of Department, **Information Systems** ____________________

Assoc. Prof. Dr. Murat Perit Çakır
Supervisor, **Cognitive Science Dept., METU** ____________________

Assoc. Prof. Dr. Cengiz Acartürk
Co-Supervisor, **Cognitive Science Dept., Jagiellonian University** ____________________

Examining Committee Members:

Prof. Dr. Altan Koçyiğit
Information Systems Dept., METU ____________________

Assoc. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU ____________________

Prof. Dr. Erhan Eren
Information Systems Dept., METU ____________________

Assoc. Prof. Dr. Erol Özçelik
Psychology Dept., Çankaya University ____________________

Asst. Prof. Dr. Murat Ulubay
Business Administration Dept., YBU ____________________

Date: **22.01.2024**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Oğuzhan Alaşehir

Signature : ____________________

ABSTRACT

CROSS-DISCIPLINARITY IN COGNITIVE SCIENCE: A DOCUMENT SIMILARITY ANALYSIS

Alaşehir, Oğuzhan

Ph.D Department of Information Systems

Supervisor: Assoc. Prof. Dr. Murat Perit Çakır

Co-Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

January 2024, 117 pages

Systematic quantification of cross-disciplinarity necessitates bibliometric and spatial analysis, socio-institutional aspects, or text-based techniques. Especially, with the advancement in bibliometric methods, a variety of measures have been developed. Yet, while these measures capture a snapshot of the concept of cross-disciplinarity, they overlook the content itself. With the rise of Data Science and Natural Language Processing (NLP) techniques, analyzing and evaluating vast volumes of documents has become technically possible. Our study introduces a methodology using text-based techniques, offering valuable insights into the relationship between publications and their specific research fields, showing potential as a robust measure of cross-disciplinarity. This approach utilizes Doc2Vec for vectorization and cosine similarity for measuring the similarity among the articles. We designed and developed models utilizing the Doc2Vec method for analyzing cognitive science and related fields. Cognitive science was chosen as a case study due to its inherent cross-disciplinarity. Cognitive science was established as a cross-disciplinary domain of research in the 1970s. Since then, the domain has flourished, despite disputes concerning its cross-disciplinarity. Our findings reveal that this methodology is applicable to quantify cross-disciplinarity. Furthermore, we observed that cognitive science collaborates closely with most constituent disciplines. For instance, we found a balanced engagement between several constituent fields—including psychology, philosophy, linguistics, and computer science—that contribute significantly to cognitive science. In our analysis, we find that the scholarly domain of cognitive science has been exhibiting overt cross-disciplinary collaboration for the past several decades.

Keywords: cross-disciplinarity, interdisciplinarity, text similarity analysis, natural language processing, Doc2Vec modeling

ÖZ

BİLİŞSEL BİLİMDE DİSİPLİNLERARASILIK: DOKÜMAN BENZERLİK ANALİZİ

Alaşehir, Oğuzhan
Doktora, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Doç. Dr. Murat Perit Çakır
Ortak Tez Yöneticisi: Doç. Dr. Cengiz Acartürk

Ocak 2024, 117 sayfa

Disiplinlerarasılığının sistematik olarak nicel bir şekilde değerlendirilmesi, bibliyometrik ve mekânsal analizleri, sosyo-kurumsal yönleri veya metin tabanlı teknikleri gerekli kılmaktadır. Özellikle, bibliyometrik yöntemlerdeki ilerlemelerle birçok ölçüm geliştirilmiştir. Ancak bu ölçümler, disiplinlerarasılık kavramının bir boyutunu sunsa da, içeriğe bakmayı göz ardı etmektedirler. Veri Bilimi ve Doğal Dil İşleme (NLP) tekniklerinin yükselişiyle, büyük miktarda dokümanın analizi ve değerlendirmesi teknik olarak mümkün hale gelmiştir. Çalışmamız, yayınlarla spesifik araştırma alanları arasındaki ilişki hakkında değerli bilgiler sunarak, disiplinlerarasılığın bir ölçütü olarak potansiyelini gösteren metin tabanlı teknikleri kullanarak bir yöntem sunmaktadır. Bu yaklaşım, vektörleştirme için Doc2Vec'i ve makaleler arasındaki benzerliği ölçmek için kosinüs benzerliğini kullanmaktadır. Bilişsel bilim ve ilgili alanları analiz etmek için Doc2Vec yöntemini kullanan modeller geliştirdik. Disiplinlerarasılığın doğuştan gelen özelliği nedeniyle bilişsel bilim, bir durum çalışması olarak seçilmiştir. Bilişsel bilim, 1970'lerde bir disiplinlerarası araştırma alanı olarak kurulmuştur. O zamandan beri, disiplinlerarasılığı hakkındaki tartışmalara rağmen, bu alan büyümeye devam etmiştir. Bulgularımız, bu metodolojinin disiplinlerarasılığı nicel olarak değerlendirmek için uygulanabilir olduğunu göstermektedir. Ayrıca, bilişsel bilimin çoğu bileşen disiplinle yakın bir işbirliği içinde olduğunu da gözlemledik. Örneğin, bilişsel bilime önemli katkıda bulunan birkaç bileşen alan - psikoloji, felsefe, dilbilim ve bilgisayar bilimi - arasında dengeli bir etkileşim bulduk. Analizimizde, bilişsel bilimin bilimsel alanının son birkaç on yılda açıkça disiplinlerarası bir yaklaşım sergilediğini bulduk.

Anahtar Sözcükler: disiplinlerarasılık, metin benzerlik analizi, doğal dil işleme, Doc2Vec modelleme

To My Family

ACKNOWLEDGMENTS

First and foremost, I would like to sincerely express my deepest gratitude to Assoc. Prof. Dr. Cengiz Acartürk, for his guidance, encouragement, and patience throughout all phases of this academic research. His assistance and positive approach have been invaluable, and I will never forget them.

I also wish to express my sincere appreciation to Assoc. Prof. Dr. Murat Perit Çakır for his valuable academic contributions and encouragement, especially during times when I lost my motivation.

I owe my deepest gratitude to Prof. Dr. Altan Koçyiğit and Asst. Prof. Dr. Murat Ulubay for their insightful contributions and suggestions during the thesis committee meetings. Additionally, I extend my thanks to Prof. Dr. Erhan Eren and Assoc. Prof. Dr. Erol Özçelik for their valuable comments and constructive feedback during the final jury.

I am also grateful to the editors and reviewers of Cognitive Science journal for their valuable feedback during the article submission process, which has significantly contributed to this thesis.

I want to express my gratitude to my parents, Fatma Alaşehir, and Cengiz Alaşehir, for their best wishes and endless support.

Lastly, but most importantly, I owe a tremendous thank you to my wife, Remziye Alaşehir, and my son, Yusuf Görkem Alaşehir. I could never have achieved this accomplishment without their support, affection, and belief in me. My wife's patience and encouragement, especially during the most challenging phases of my research, have been invaluable. To my beloved Remziye, and my dear son Görkem, thank you for bringing happiness and light into my life. Your both presence in my life is an immeasurable gift, filling it with constant joy, inspiration, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANOVA	ANalysis Of VAriance
API	Application Programming Interface
Bert	Bidirectional Encoder Representations from Transformers
BoW	Bag of Words
CBOW	Continuous Bag-of-Words
CSS	Cognitive Science Society
Elmo	Embedding from Language Model
FN	False Negative
FP	False Positive
GloVe	Global Vectors for Word Representation
IEEE	Institute of Electrical and Electronics Engineers
LDA	Latent Dirichlet allocation
LSA	Latent semantic analysis
NLP	Natural Language Processing
OECD	Organization for Economic Co-operation and Development
P	Precision
PMI	Pointwise Mutual Information
PV-DBOW	Paragraph Vector – Distributed Bag of Words
PV-DM	Paragraph Vector Distributed Memory
R	Recall
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
WoS	Web of Science

CHAPTER 1

INTRODUCTION

People of Ancient Greece were among the first to systematically accumulate knowledge. They are credited with pioneering a scientific approach from the perspective of Western thought. Like their predecessors, such as the Egyptians, they focused on knowledge in specific disciplines. Although this knowledge was categorized into various disciplines during the era of Plato and Aristotle, such as mathematics, politics, agriculture, and medicine, it is not known whether there was a high level of specialization of scholars within specific disciplines. The key figures of the *academy*[1]*and lyceum*[2] were known as polymaths. This was particularly evident in the case of Aristotle, who adopted a holistic approach to knowledge acquisition. He was not only a philosopher but also a mathematician, physicist, and biologist and he integrated his understanding of these various disciplines in his exploration of the world. Therefore, from today's perspective, cross-disciplinarity can be attributed to scholars as an inherent characteristic of people in ancient Greek.

Throughout the centuries, the breadth and complexity of human knowledge have evolved with exponential growth. This evolution led to human knowledge being organized, compartmentalized, and divided into distinct disciplines, each with its own unique techniques, methodologies, epistemologies, terminologies, and frameworks. This resulted in the development of expertise and specialization in research fields. Consequently, a shift was observed from the comprehensive understanding

---

[1] An educational institution founded by Plato, focused on abstract ideas and philosophical teachings.

[2] An educational institution founded by Aristotle, emphasizing empirical approaches and observational methods.

characteristic of an Aristotle-like polymath to more concentrated and specialized areas of study. Especially after the 17th century, traditional academic disciplines like physics, biology, and mathematics emerged. The contemporary structure of distinct disciplines was established and solidified after the beginning of the 18th century (Turner, 2017). The 19th century faced widespread institutionalization of this disciplinary diversification (Eykens, 2022).

This specialization trend resulted in the formation of knowledge silos, which limited collaboration (Mclevey et al., 2018). The siloed nature of these disciplines hindered the integration of knowledge and the development of comprehensive solutions to complex problems. Besides, the industrial revolution, took place in the late 18th century, resulted in a gap between science and technology due to rapid advancements in technology. Cross-disciplinary approach was one way to bridge the gap between science and technology. As a result, there has been a rising trend towards pursuit of more integrative research that goes beyond the boundaries of the existing disciplines for complex issues of the 20th century such as space exploration, nuclear energy, sustainable development, climate change, global health crisis, and social inequality.

These motivations for bridging disciplinary gaps have led to new cross-disciplinary terminologies such as interdisciplinarity, multidisciplinarity, and transdisciplinarity. Although their importance has been persistently addressed in higher education, their connotations lacked consensus among researchers (Frodeman, 2010). They are generally defined as the antithesis of unidisciplinarity, which refers to single discipline with established methodologies, theories, and paradigms. In essence, these terms represent the combination of multiple disciplines, albeit in differing ways. For instance, multidisciplinarity typically involves combination of perspectives from different disciplines without integration. Interdisciplinarity, on the other hand, refers to the more integrative approach for combination of disciplines. Transdisciplinarity goes a step further by synthesizing and extending discipline-specific models. Although these are some commonly used definitions in literature, there are alternative ways of describing the terms. The definitions may change due to context, or they may be used interchangeably. Thus, these are controversial concepts that have been under debate for the past several decades in scientific research. In this study, we prefer to use the term "cross-disciplinarity" as an umbrella term for all kinds of approaches. This study does not delve into the nuanced differences between these terms.

Certain developments, such as the increase in multi-author studies in scientific publications (Katz & Martin, 1997; Glänzel & Schubert, 2005; Huang, 2015) and the rise of hyper-authorship (Cronin, 2001) have accelerated the discussion on the topic of cross-disciplinarity. Over the past several decades, there has been a remarkable increase in interest in cross-disciplinary research and the evaluation of cross-disciplinarity (Morillo et al., 2003; Porter & Rafols, 2009; Silva et al., 2013). An exploratory analysis of articles with *TITLE (interdisciplin* OR multidisciplin* OR transdisciplin* OR crossdisciplin*)* within the Scopus database reveals a significant growth trajectory. For instance, the number of such articles was 238 during 1960-1969. This figure saw a dramatic increase of 1,900%, reaching 4,763 in 1990-1999. The growth continued with a 481% rise, totaling 27,692 in 2010-2019. Relatively, the total number of English articles published during these periods were 1,535,499 for 1960-1969. By 1990-1999, there was an increase of 383%, bringing the count to 7,417,276. The next decade saw a growth of 131%, culminating in 17,167,973 for 2010-2019.

This shift towards cross-disciplinarity has also been fueled by funding organizations, which started to devise strategies to evaluate the cross-disciplinarity of studies. These entities have been in search of refined criteria and evaluation frameworks to measure the cross-disciplinarity of proposals. Their aim is to favor research going beyond single disciplines and provides a more holistic approach to problem-solving. Similarly, policy makers expressed a growing preference towards supporting studies which address complex societal problems in a comprehensive and integrative manner. By formulating policies that favor comprehensive solutions to complex societal challenges, they have indirectly increased the demand for evaluating cross-disciplinarity. The increasing request for evaluating cross-disciplinary research has created demand for methodologies designed for quantification. At this juncture, the growing body of knowledge within the field of science studies offered solutions to meet these demands. The idea of evaluating scientific activities finds its roots in Campbell's 1890s research on the scattering of subjects within publications (Sengupta, 1992). Over the years, the evolution of domains such as library science, information science, scientometrics, bibliometrics, and informetrics has further enhanced these evaluative capabilities. With the growing focus on these areas, different approaches have been developed to quantify scientific studies. Consequently, a systematic evaluation of cross-disciplinarity required analyses on multiple fronts, including the quantification of bibliometric and spatial approaches, socio-institutional aspects, and text-based techniques. Recent years have witnessed a growing body of literature dedicated to exploring methods for measuring cross-disciplinarity, particularly facilitated by the availability of comprehensive and reliable publisher databases such as Web of Science (WoS) by Clarivate Analytics Inc., and Scopus by Elsevier Inc. These databases have simplified the extracting of long-term and credible content for scholars.

The academic publisher databases categorize journals based on subject matter experts' decisions[3]. Such evaluations undergo systematic periodic reviews to ensure the categorizations remain current and relevant. Every journal covered by collections is assigned to at least one of the subject categories which means that a journal to be cross-listed across multiple categories. Journals in Scopus are classified under four broad subject clusters: life sciences, physical sciences, health sciences and social sciences &

[3] For Scopus, there is a Content Selection and Advisory Board (CSAB) consisting of 17 subject chairs. For Web of Science, there are many criteria are considered like subject matter, scope, affiliations of authors and editorial boards, grant-supporting funding agencies, citation relationships, and its recognition by other entities and databases.

humanities. These clusters are further delineated into 27 major subject areas (e.g.: agricultural and biological sciences, computer science, energy) and 334 minor subject areas (food science, computer vision and pattern recognition, energy engineering and power technology). In comparison, the WoS systematizes its content into 254 subject categories, aligned with the granularity of Scopus's minor subject areas. To illustrate, computer science is segmented into 13 minor subject areas within Scopus, whereas WoS designates 7 corresponding categories. Artificial intelligence, information systems, hardware & architecture and software are the matching areas which are consistently represented in both databases. The approach employed by these two academic databases in cataloging cross-disciplinary areas exhibits marked differentiation. Scopus introduces a distinct major subject area named "multidisciplinary" whereas WoS extends categories such as different subject categories such as "computer science, interdisciplinary applications", "mathematics, interdisciplinary applications" and "social sciences, interdisciplinary". The dynamism in categorization reflects the evolving landscape of indexed journals for both databases. For instance, the coverage of WOS was initially around 700 journals in 1964 (Singh et al., 2021). As of today, there are more than 24,000 journals covered while this number is around 28,000 for Scopus.

This enhanced access to scholarly publications has facilitated quantifying scientific activities at various levels, including individual researchers, teams, departments, universities, regions, countries, journals, and disciplines, with a focus on mainly metadata. These databases have served as valuable resources for metadata analysis such as citations, keywords, authors, affiliations, departments. Therefore, quantification studies mainly focused on metadata analysis (some examples: Boyack, 2004; Leydesdorff & Rafols, 2011; Rafols, 2014; Deng & Xia, 2020). The quantification of cross-disciplinarity is mainly aimed to create new measurement indexes or adopt existing ones to define existence or the level of interactions between academic fields. Although these measures provide a snapshot of the concept of cross-disciplinarity, they overlook an important point, focusing on content itself. Although content has played a limited role in the quantification studies of cross-disciplinarity, it carries significant potential. Especially, with the emergence of Data Science and NLP techniques in Information Systems (IS), it became technically possible to analyze and evaluate the contents of massive number of documents.

In the present study, we employ semantic text similarity analysis which is a component of NLP within IS domain to offer a broad picture of the concept of cross-disciplinarity in cognitive science and the contributing subfields. The main motivation behind the reason to choose this methodology is related to the definition of cross-disciplinarity which is generally defined as the process of solving problems that a single profession cannot solve. In that case, a cross-disciplinary field is expected to employ other, related fields' tools and terminologies. In other words, contextual and semantic relationships should be observed between constituent fields and the new cross-disciplinary field. Those relationships can be investigated through text-based analysis, specifically, through text similarity measures, as supported by Kallens and Dale (2018). Accordingly, the main aim of this study is to investigate whether analyzing text similarity offers consistent insights into the relationship between academic documents and their respective research fields. If this is the case, what does such text similarity analysis reveal about the interdisciplinarity of these fields? The techniques included Doc2Vec (for vectorization) and cosine similarity (for measuring the similarity among the articles). This investigation takes the form of a case study that focuses on cognitive science and related subfields.

Cognitive science is chosen as a case study due to its inherent interdisciplinarity and discussions around it. More recently, cross-disciplinarity has been debated in cognitive science through various types of analyses to support or refute the claims that favor interdisciplinarity (e.g., Núñez et al., 2019; Gray, 2019; Oey et al., 2020). In cognitive science, cross-disciplinarity has been conceived as merging or combining different fields' methods, models, and languages rather than sequencing them (Thagard, 2010, p. 243). Thagard describes cross-disciplinary research in cognitive science in three forms: individual, collaborative, and inspirational. In the individual form of cross-disciplinary research, the researcher employs methodologies from multiple fields. In a collaborative team, individuals from different backgrounds combine skills. Finally, the research may be based on ideas inspired by various fields, despite a lack of explicit individual or collaborative cross-disciplinarity. On the other hand, Núñez et al. (2019) presented a contrastive approach, arguing that cognitive science has failed as a cross-disciplinary field, and followed this claim with a series of commentaries by several researchers evaluating the cross-disciplinary characteristics of the field (Gray, 2019).

The methodology requires making operational assumptions that allow the quantification of concepts through simplification. For instance, we broadly use the term cross-disciplinarity, covering interdisciplinary, multidisciplinary, and transdisciplinary research. Moreover, we assume that conference proceedings comprise an acceptable dataset, representative enough for cognitive science research. After specifying the scope of the study, the central research question of the article, investigating cross-disciplinarity in cognitive science through text-similarity analyses, relies on quantifiable measures.

1.1. Research Questions

Academic publications serve as repositories of information, covering the intricate details, methods, and findings of research endeavors. Such information can be used as an input for quantitative studies of science. Through methods like bibliometric analysis and text-based measures, these academic works can be transformed into a distinct form of information. In this study, our goal is to process this information in a manner that aligns with the core principles and objectives of the Information Science domain.

Building on this foundation, we employ text-based analysis to investigate relationship information between fields by assuming that a cross-disciplinary field is expected to employ other related fields' tools and terminologies, contextual and semantic relationships are expected to be observed between fields in relation. This dissertation presents step-by-step approach for a case study examining the relationship between academic documents (structured and unstructured ones) and their respective subfields, with a specific focus on cognitive science and its given six related disciplines (philosophy, anthropology, linguistics, neuroscience, computer science, psychology). The research questions guiding this study are as follows:

RQ1: Does text-similarity analysis provide consistent information about the relationship between academic documents and their research fields.

To find an answer to the first question, field-specific journal articles and field-specific conference proceedings will be investigated. The model is expected to find significant relationships between given field's data sets. If text-similarity analysis provides consistent information about the relationship between academic documents and their respective fields, the subsequent question will be:

RQ2: What does text similarity analysis explain regarding the cross-disciplinarity of fields?

The subsequent analysis will aim to interpret the findings of the model which will be run on the proceedings of Cognitive Science Society (CSS) conferences. Understanding how text similarity relates to cross-disciplinary research can shed light on the interconnections and knowledge flow between different subfields and cognitive science.

Overall, this dissertation aims to contribute to the literature on text-similarity analysis and its implications for cross-disciplinarity of cognitive science.

1.2. Contributions

Semantic text similarity, which is a core concept of NLP within IS domain, can be basically defined as an analysis to compare two or more texts based on their shared content. In this study, we apply this concept to the question of how to measure cross-disciplinarity of academic disciplines. The methodology and analyses designed to answer the research questions in this research have the potential to offer contribution to the field of Information Science by enhancing the understanding and application of semantic text similarity analysis, particularly in the context of measuring the relevance and cross-disciplinarity of academic research.

The first contribution of this study is about how text similarity analysis can support the creation of valuable information for defining relations between academic documents and their research fields. The analysis in this dissertation presents that text similarity analysis can predict an academic document's relevant research field by measuring cosine similarities of the vectors created by Doc2Vec method.

The other contribution of this dissertation is providing a method to quantify the cross-disciplinarity of cognitive science. The methodology and the analyses contribute to a better understanding of the cross-disciplinarity of cognitive science. It is a novel study in terms of the approach named document similarity employed for quantification of a discipline's cross-disciplinarity, in this case cognitive science. Cognitive science was founded as an inherently cross-disciplinary field with relevant disciplines including philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. Since the birth of it, the effect of the contributing subfields has been in question and various attempts have been made to evaluate cognitive science's cross-disciplinary nature. Especially, in the last 3 years beginning with Núñez et al. (2019), different perspectives have been provided to analyze it. It is found that psychology, philosophy, linguistics, and computer science are contributing fields. Additionally, anthropology and neuroscience are found to be limited contributions.

The findings of this research may carry potential policy and practical implications which will help advancement of cognitive science research. It is recommended that policy makers and funding organizations allocate a considerable portion of funding and resources towards the four major contributing fields. Although the contributions from anthropology and neuroscience have been found to be limited, they provide a variety of perspectives. Policies should thus support cooperation between all these fields. Practically, this could mean creating joint research projects, research programs, educational programs and platforms that encourage collaboration. Furthermore, to provide a comprehensive understanding of the field, cognitive science education should be planned to provide a solid foundation in psychology, philosophy, linguistics, and computer science as well as cover ideas from anthropology and neuroscience.

The study presented in this dissertation has been published in the journal of Cognitive Science (Alasehir, O., & Acarturk, C., 2022)

1.3. Organization

The rest of the dissertation is organized in the following chapters. The chapter titled 'Literature Review and Background' starts with definitions of fundamental concepts such as academic discipline and cross-disciplinary research. This is followed by a comprehensive review of the existing body of literature related to the quantification of cross-disciplinary studies. The last section of this chapter provides an analysis of literature specifically dedicated to the quantification of cognitive science which is highly related to this dissertation's research questions.

The 'Methodology' chapter first introduces basic concepts such as text similarity, vectorization, and similarity measures that will form the basis of this study. Then, a set of operational assumptions that are required to implement the methodology are provided. The flow of methodology starting from data sources, data itself, and pre-processing of it to the creating model and measuring similarities are presented.

In the 'Results' chapter, outputs of analysis are provided in three major stages in accordance with data source definition of methodology. In the first two stages, the model is validated by creating similarity matrices for journal articles and field specific proceedings. The subsequent stage evaluates cross-disciplinarity of cognitive science by running the model for proceedings of the Cognitive Science Society (CSS) Meetings.

The ‘Discussion’ chapter addresses limitations and challenges encountered during the research. The insights into our understanding of cross-disciplinarity characteristics of cognitive science over the time period are also included in the section. The findings are compared with the quantification research in the literature. The final chapter, ‘Conclusion’ summarizes the overview of the dissertation covering research questions, methodology, and results. This chapter also includes suggestions for potential future research.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

This chapter starts with the introduction of relevant basic concepts; academic discipline and cross-disciplinary research. A brief history of how academic disciplines are classified, and the definition of the term "academic discipline" will be provided from different perspectives. The idea of cross-disciplinary research will be discussed with examples in the literature by emphasizing the distinctions between multidisciplinarity, interdisciplinarity, transdisciplinarity and other relevant terms. This chapter will continue with studies on the approaches to quantify cross-disciplinary research. The last section is the "Cross-disciplinarity in Cognitive Science" where cognitive science's history and the attempts to quantify the cross-disciplinarity character of it will be examined.

2.1. Basic Concepts

2.1.1. Academic Discipline

In parallel to the increase in knowledge production, organization for the body of knowledge became a necessity in history. The organization has been achieved by specialization in narrow study of areas and institutionalization of those areas. As a consequence of these knowledge organization activities, academic disciplines have emerged. Although they were relatively superficial when compared to 18^{th} century disciplines, the history of disciplines can be traced back to the times of Plato in 360 BCE. His academy classified the science in various groups as visualized in interactive historical atlas of the disciplines in the Figure 1.

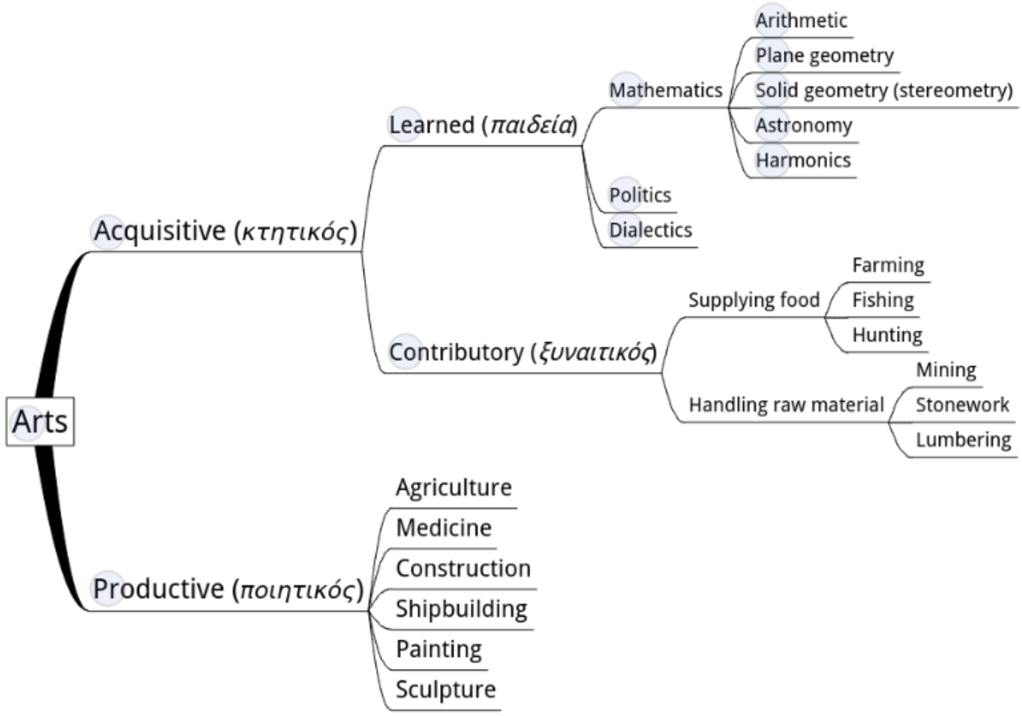


Figure 1: Plato's classification of disciplines in Interactive Historical Atlas of the Disciplines (Sandoz, 2022)

Plato mainly distinguished between acquisitive and productive disciplines. Productive ones consisted of disciplines that helped people to get insight into the world and develop practical applications such as agriculture, medicine, and construction. Acquisitive disciplines included abstract or eternal studies like mathematics, politics, and dialectics as well as some contributory disciplines which were applicable in real life like fishing, hunting, mining. After Plato, Aristo made a distinction between disciplines based on theoretical (mathematics, natural science, and metaphysics), practical (ethics and politics), and productive (agriculture, mining, painting, poetry) categories (Sandoz, 2022). By following them, there were various classifications in Ancient Greece and Ancient Rome to define the borders of disciplines. By their nature, they have been changed over a long period of time. Although academic discipline's history goes back to ancient times, they have been institutionalized with universities in the Middle Ages. The rise of modern academic disciplines was in the 18th century.

Academic discipline is a controversial concept that different perspectives explain in different terms. Researchers have described it in different ways in the literature. According to Moran (2002: p2) the term's broader definition suggests a meaning that covers a specific area of study or knowledge. Hirst (2010: p4) defines disciplines as a logically separated body of knowledge. Huber and Morreale (2002) emphasizes that disciplines have their own history and traditions. It means each has its academic community, journals, and other platforms to exchange ideas. According to institutional theory of Jacobs (2014: p27), which was based on Turner (2010), the term refers to a broadly recognized area of study. The acceptance of a field as a discipline requires institutionalization in higher education systems which provide degrees. Stehr and Weingart (2000: p xi) mentions roles of disciplines in the modern age. The main proposal for scientific disciplines is that they create "social order of knowledge". This role is based on mainly three functions of disciplines: knowledge transfer between generations by education system, affecting structure of occupations and being a social organization, which contribute to moral, legal and economical conflicts. Bordons et al. (2004) also defines disciplines in two dimensions; social and intellectual. They define it as a framework for organization of contemporary knowledge on these two dimensions.

2.1.2. Cross-disciplinary Research

Over the past several decades, a combination of factors together has played a significant role in shaping academic structure and organizational definitions. Looking within academia, researchers have been seeking ways to collaborate with other researchers in different research areas to overcome the barriers of limited knowledge. Meanwhile, strategic encouragements as a result of current needs have been supporting researchers to work with different teams having different perspectives. Fundings, policy decisions, and strategic planning at various levels, from educational institutions to governmental bodies, have facilitated this trend. As an example, The National Academies KECK Futures Initiative (NAKFI) started a program to support interdisciplinary activities in 2003 (Porter et al, 2006). One of the current initiatives started with Horizon Europe which is designed to facilitate collaboration and innovation in the region. According to the European Commission (n.d.), the program aimed to support interdisciplinary projects for the years 2021-2027. General factors such as economic reasons, global problems or technological improvements played a role in the popularity of these concepts. More specifically, in a report to the European Union in 1994, Schmoch et al. (1994) summarized the factors that supported the emergence of interdisciplinary research. These included the need to find alternative solutions for critical social and environmental problems, the quest for enhanced efficiency in knowledge production by cooperating with researchers in other fields dealing with similar problems, and the need to pursuit of new paradigms in established fields as a common nature of scientific improvement. Another study on motivations behind cross-disciplinary research was published by National Academies of Sciences, Engineering, and Medicine (2005). The attempts to solve social problems, technological improvements with transform capability, and the need to understand highly complicated nature and social issues around humanity were mentioned as major driving forces for researchers to focus on going beyond limits of disciplines. These pursuits and motives led to the creation of new phrases such as collaborative research or Team Science which is defined as scientific activity or joint research conducted by two or more researchers in a collaborative way (National Academies of Sciences, Engineering, and Medicine. 2015; Fiore 2008; Wuchty et al., 2007). Policy makers have been seeking to remove challenges in creating a collaborative environment for researchers by enhancing the effectiveness of Team Science.

These motivations of crossing discipline boundaries have also resulted in new cross disciplinary terms such as inter-, multi-, and trans-disciplinarity. They are controversial concepts that researchers have been discussing for the past years. There are different definitions or different approaches to define them in the literature which makes it hard to provide an exact meaning of each term. This ambiguity is further compounded by the tendency to use these terms interchangeably. For instance, the term interdisciplinarity is perceived as a generic definition covering all types of cross-disciplinary research. Graff (2015) mentions that there exists a disagreement on the meaning of the term and on whether it is perceived positively or negatively.

To bridge the gap between these conceptual discussions, an initial attempt to differentiate and define the terms, in other words the attempts for creating typology of cross disciplinary phenomena, goes back to a conference sponsored by the Organization for Economic Co-operation and Development (OECD) in 1970. According to Berger (1972) typology was a major, systematic categorization of cross-disciplinary research. It identified cross-disciplinary research within the context of interdisciplinary, pluridisciplinary, multidisciplinary, and transdisciplinary research. Further studies proposed various definitions of each of those terms. For example, Klein and Newell (1996) defined interdisciplinary studies as the process of finding answers to complex questions, or of dealing with broad problems that cannot be handled by a single discipline. Similarly, interdisciplinary research was described as integrative or synthesis research, aiming to solve complex problems by integrating theories, techniques, or tools used by two or more disciplines. Multidisciplinary research was described as combining more than one field by additive contributions of specific research fields, rather than from their integration (National Academy of Sciences et al., 2005). Stokols et al. (2008) emphasized four aspects of scientific orientation in disciplines: the concept of unidisciplinarity addresses researchers from a single discipline, usually working as a team. Multidisciplinarity is a sequential research process, wherein discipline-specific perspectives of individual researchers are preserved. The main characteristic of interdisciplinarity research is its interactive and collaborative nature. So, interdisciplinarity requires researchers to focus on a topic jointly by preserving their individual, disciplinary perspectives. Finally, transdisciplinarity is an integrative process that aims to establish an original model to address a research question by synthesizing and extending discipline-specific models. Graff (2015) claimed that the difference between multidisciplinary and interdisciplinary research is the research team's approach to a given problem. While multidisciplinary teams approach a problem solely through their disciplinary methods, interdisciplinary teams approach the problem by integrating the methods of the specific disciplines. Accordingly, interdisciplinarity can be achieved only when the researchers learn the language of the constituent disciplines, their theories, and their assumptions.

Through this section, we have established the core discussions around concept of academic disciplines and the nature of cross-disciplinary research. We have traced the historical development of academic disciplines to see how they were first organized and explained from various viewpoints. We also discussed cross-disciplinary research, distinguishing among multidisciplinarity, interdisciplinarity, and transdisciplinarity. Moving forward, the next section will delve into the quantification research of cross-disciplinary studies.

2.2. Quantifying Cross-disciplinary Research

As provided in the previous part, the discussions on different aspects of cross-disciplinary research have been ongoing. The existence, types, and differentiations between the terms, aims and structures have been in question. Moreover, depending on how a researcher defines these categories and understands that specific field based on experiences, the same research area is classified in various cross-disciplinary terminology. In other words, perspectives and employed criteria may change how a field is classified. As an example, ecology, which can be basically defined as the science of the relationships between living organisms, has interactions with different fields such as biology, physics, chemistry, economics, sociology. One may define ecology as an interdisciplinary field since it is integrating various fields to solve ecological problems around us. On the other hand, another researcher considers it to be a transdisciplinary field which synthesizes and extends models of biology, sociology, and others. Terminology differences and perspectives of researchers may result in such results. Another example for this confusion might be environmental science. It is a research area that some scholars classify as an interdisciplinary field and others as pure discipline. Similarly, cognitive science has also been in question about its cross-disciplinarity.

In literature, both qualitative and quantitative investigations have been used to ground these arguments in more tangible evidence. As a result of these initiatives, evaluation, or measurement of cross-disciplinarity of fields evolved. There is a growing literature dedicated to how to measure cross-disciplinarity of fields in the past years. Library science and information science (or a combination named Library and Information Science) cover these topics. Especially, with the help of scientometrics, which can be defined as science of science studies or quantitative evaluation of science, much research has been devoted to this aim. The emergence and scope improvements in academic publisher databases such as Web of Science by Clarivate Analytics Inc. and Scopus by Elsevier Inc. supported researchers to access reliable and long-term data of both contents and metadata. Additionally, domain-specialized databases such as Medline, National Library of Medicine’s (NLM) have facilitated accessing bibliometric data for publications. These advancements have streamlined the process of quantifying the scientific activities in different levels such as researcher, team, department, university, region, country, journal, or discipline.

The scope of quantifying cross-disciplinary studies is multi-faceted, examination of publication contents or analyzing metadata. The former involves evaluating the core of the research like full text, abstract, theories, methodologies. On the other hand, evaluating metadata explores authorship details, citation patterns, keywords, departments, and other contextual information. By considering both content and/or metadata, researchers can gain a different perspective on the view of cross-disciplinarity of research.

Regarding these aspects, the methodologies employed to quantify cross-disciplinarity may include a combination of bibliometric and spatial approaches, the analysis of socio-institutional aspects and text-based analyses as will be detailed in the following section.

2.2.1. Bibliometric and Spatial Approaches

Recent improvements in academic publisher databases and domain-specialized databases have facilitated accessing bibliometric data for publications. Document types, citations (i.e., citing publications or cited publications), co-authorships, and co-words comprise a list of primary inputs for bibliometric studies. For the analyses, various features of academic publications have been used, such as a part of the article (abstract, title, keywords), its metadata (author, journal, institutional information), or the research field assigned by the publisher (domain, subject area, subfield, discipline). The spatial measures include the measures of cross-disciplinarity in terms of relational attributes among authors, papers, journals, institutions, or disciplines. For instance, social network analysis measures, such as centrality measures (degree, betweenness, or closeness), have been used as spatial measures. Diversity and coherence are the two key concepts that may be employed to measure the spatial structures of the parties (Rafols, 2014). Bibliometric and spatial measures are usually accompanied by visualizations, animations, and time-series analyses (Rafols et al., 2010). Due to their close relationship, we integrated bibliometric and spatial measures in a single category in the present study.

There are numerous studies employing bibliometric and spatial approaches in cross-disciplinarity quantification research in literature. Boyack (2004) focused on import and export mapping of citation analysis for The Proceedings of the National Academy of Sciences (PNAS), a peer reviewed journal of the National Academy of Sciences (NAS). He evaluated the diffusion between PNAS topics based on normalized citation counts and created an index of independence. Porter et al. (2007) offered two metrics to measure researcher level interdisciplinarity: integration and specialization. While the integration metric was calculated by counting the number of papers published in different subject categories, specialization was derived from citations counts. Leydesdorff (2007) merged bibliometric and spatial measures for journal level analyses using citation matrix. He proposed that betweenness centrality of a citation network can be used as an indicator of interdisciplinarity. In other words, the higher betweenness centrality value meant higher intermediary role between different fields, hence more interdisciplinary journal. In their 30 years period investigation for six research domains degree of interdisciplinarity, Porter & Rafols (2009) analyzed the citations and combined integration score (Rao-Stirling diversity) as in Porter et al. (2007) with a visualization method. Rafols & Meyer (2010) also focused on a combination of approaches to explore the bionanoscience field’s interdisciplinarity based on references in article. Disciplinary diversity and network coherence measures were used to decide whether a field is a specialized interdisciplinary, specialized disciplinary, potential interdisciplinary integration or potential integration within discipline. Huang & Chang (2011) analyzed interdisciplinary changes in the field of information sciences by using citation and co-authorship data of journal articles. They employed Brillouin's Index to shed the light on the degree of interdisciplinarity. Leydesdorff & Rafols (2011) was another research which employed citations networks. They suggested three indicators to be considered as possible candidates: Shannon entropy, betweenness centrality and Rao–Stirling diversity. The principal finding of the study suggested that different indicators might encompass distinct interpretations of complex concept of interdisciplinarity. Silva et al. (2013) also employed entropy-based measurement for journal citation networks to present the diversity of the subject categories. Yegros-Yegros et al. (2015) analyzed the effect of interdisciplinarity on citation impact of publications. During this research, interdisciplinarity was defined as diversity of disciplinary categories cited in a publication. The diversity meant to be variety (number of distinctive WoS categories cited in an article.), balance (Shannon diversity of evenness of the distribution of categories) and disparity (degree to which the categories are different/similar by averaging distance between WoS categories within the reference list). Bark et al. (2016) contributed to literature by providing evaluation principles for interdisciplinary research. They created an interdisciplinary index based on WoS subject area differences among the papers of a project team. In their comparison research, Abramo et al. (2018) analyzed disciplinary diversity approach to compute variety, balance, disparity, and integrated diversity index of fields. The comparison was based on disciplinary diversity of authors and reference list of publications. In a current study, Deng & Xia (2020) explored information behavior research field's interdisciplinarity by focusing on network analysis and diversity measure.

2.2.2. Analysis of Socio-institutional Aspects

The second method of quantifying cross-disciplinarity is the analysis of socio-institutional aspects. The approach focuses on descriptive statistical studies on researchers and research environments, rather than on academic publications. Researchers' backgrounds, affiliations, institutional curriculums, and courses offered by the departments are investigated via primary statistical analyses to assess the cross-disciplinarity of research fields.

One of the earlier proposed socio-institutional measure was Urata's (1990) analysis on researchers in Japan. It was based on a survey to collect the migration of scholars among disciplines data. It was used to evaluate how the disciplines found closer based on migrations of researcher from one to other. In his nanoscience and nanotechnology focused research, Schummer (2004) defined a multidisciplinarity index and interdisciplinarity index separately, both of which were based on co-authorship of the field. The distinction was based on the number of disciplines (refers to multidisciplinarity) and interactions between disciplines (refers to interdisciplinarity). Rijnsoever & Hessels (2011) investigated the factors affecting the research collaborations by asking perceptions of respondents in a university. In that study, they defined interdisciplinary collaborations by measuring co-authorships of the researchers based on their self-report of worked discipline and the number of collaborations with other researchers in different disciplines. Carr et al. (2018) presented a program evaluation framework based on social learning processes, social capital outcomes and knowledge and human capital outcomes. They classified publications as cross-disciplinary if the authors were affiliated with different research fields according to WoS classification.

2.2.3. Text-based Analysis

The third method is the text-based analysis of cross-disciplinarity, which mainly focuses on publication content. Titles, abstracts, topics, keywords, or full texts are input data for those analyses. Natural Language Processing (NLP) approaches play an important role since it is necessary to analyze unstructured text data. Text similarity measures (e.g., Jaccard Similarity, Cosine Similarity, Jensen-Shannon Distance) and vectorization methods (e.g., TF, TF-IDF, word embeddings) are commonly employed.

In literature, there have been numerous studies conducted which employed text-based analysis. Nichols (2014) quantified interdisciplinarity by employing topic model using Latent Dirichlet Allocation (LDA) algorithm to identify latent topics in the documents. The research focused on the National Science Foundation (NSF)'s content of grant proposals and awards. The decision was made based on the number of disciplines related to topics of each award. Evans (2016) created a weighted measure of interdisciplinarity score by calculating cosine similarity between words of a researcher and corpus of disciplines. The higher similarity score was based on the higher overlapping terms in both corpora. Xu et al. (2016) combined text-based measures with social network analysis to introduce an indicator of interdisciplinarity named topic terms interdisciplinarity (TI). TI was calculated by multiplying distribution of topic terms with term frequencies. In the Information Science & Library Science (LIS) case study, TI was found an indicator representing interdisciplinarity of topics. One of the current studies on text-based measures was Dias et al. (2018)'s language-based analysis of nearly 20M scientific articles. The methodology was described as measuring the dissimilarity with (generalized and normalized) Jensen–Shannon divergence between discipline vectors. The vectors were generated by calculating the frequency of words in each field. Chakraborty (2018) also combined citation networks with keyword-based text measures. In the computer science domain-oriented research, he explored diversity of article keywords with Keyword Diversity Index. Eykens et al. (2022) used topic modelling for titles and abstracts of social sciences and humanities academic outlets like journal papers, proceedings, book chapters, and monographs. Their topic modelling approach was named Top2Vec which created topic vectors from documents and words.

2.3. Cross-disciplinarity in Cognitive Science

One of the challenging questions of science is what the mind is and how it works. Throughout history, philosophers, psychologists, biologists, and other relevant field researchers have proposed perspectives from their own point of view. However, it is still a phenomenon that remains unclear despite the development of different approaches over the years. This complex question brought different researchers from various fields to collaborate and contribute to a comprehensive understanding. This practice has led to the emergence of a new field named cognitive science.

From the perspective of Miller (2003) the history of cognitive science goes back to the 1950s which can be addressed to breakthroughs in psychology, linguistic and artificial intelligence. Shifting from behavioral to cognitive approaches in psychology, invention of artificial intelligence terms, use of computers in modeling cognitive processes and Chomsky's definition of linguistics formed pillars of cognitive science. The first officially named cognitive science program was declared by Alfred P. Sloan Foundation in 1976 (Miller, 2003). After the program started, the committee prepared an unpublished report. According to the State of the Art Committee, in 1978, cognitive science was established as an inherently cross-disciplinary field of research.

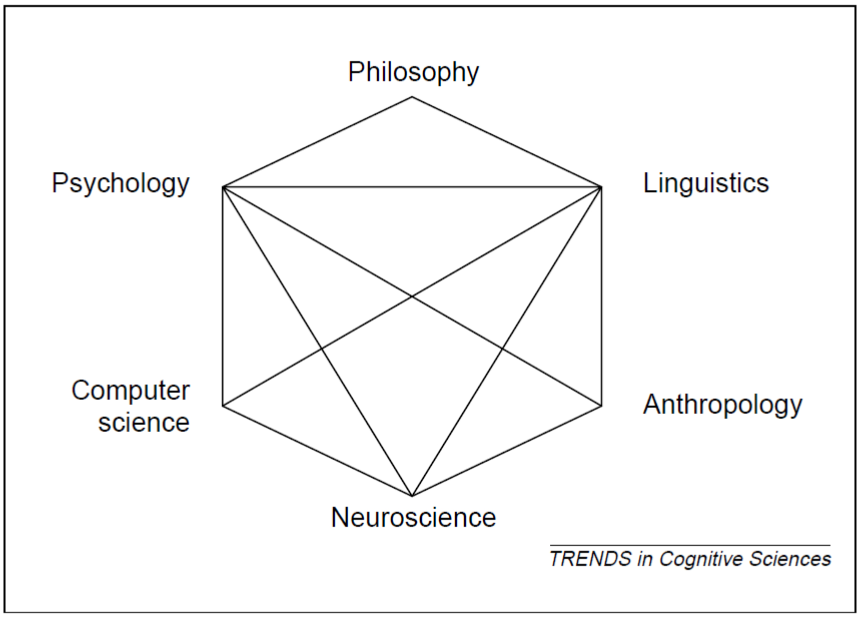


Figure 2: Reproduced version of hexagon presenting cognitive science related fields in Alfred P. Sloan Foundation report (Miller,2003)

As given in Figure 2, the relevant disciplines included philosophy, anthropology, linguistics, neuroscience, computer science, and psychology (see Gardner, 1985; Thagard, 2005; Miller, 2003; Boden, 2006; Serrano et al., 2014 for a history of the birth of cognitive science as an independent field of research and its relation to other disciplines). Since then, there have been numerous attempts to assess the cross-disciplinarity of cognitive science.

For instance, Schunn et al. (1995) investigated socio-institutional aspects by conducting questionnaires at the 16th Annual Meeting of the Cognitive Science Society (CSS), collecting data about participants' training backgrounds, professional status, and roles as article authors. The results revealed significant interdisciplinary collaborations. Schunn et al. (1998) reported an evaluation of the articles of the Cognitive Science journal and the 17th Annual Meeting of the CSS. They analyzed a combination of dimensions, including a set of socio-institutional aspects (e.g., authors' affiliations and academic backgrounds), text-based analyses (primary methodologies applied) and bibliometric parameters (e.g., citations of their published work). The results revealed that psychology and computer science were the two dominant fields at the time, and the authors also categorized 30-50% of the studies as multidisciplinary research. Von Eckardt (2001) investigated the multidisciplinarity of cognitive science by emphasizing two types of notions: The localist multidisciplinarity and the holist multidisciplinarity. A localist multidisciplinary field is established on researchers' research capabilities, whereas a holistic multidisciplinary field reflects the contribution of various disciplines. The study emphasized the need for adopting a holistic multidisciplinary approach in cognitive science. The study also extended Schunn et al. (1998) by reporting results of a text-based analysis, investigating the content of articles in the journal Cognitive Science. Psychology (especially cognitive psychology) and computer science were the dominant fields.

Thagard (2005a) evaluated the interdisciplinarity of cognitive science by a five-fold analysis of the trading zones (or interactions): People (the interdisciplinary interests of the founders), Places (the effect of important institutions on fostering interdisciplinary research), Organizations (the role of societies and journals in exchanging ideas), Ideas (the cross-disciplinary intellectual contents, like mental representation) and Methods (the cross-disciplinary methods, such as computer simulation). The approach was mainly based on text-based and socio-institutional approaches. As a result, Thagard stated that cognitive science had a successful interdisciplinary characteristic, due to fruitful interactions between cognitive science and six constituent disciplines (i.e., the trading zones).

Goldstone and Leydesdorff (2006) reported an analysis of citation interactions in the journal Cognitive Science by employing bibliometric citation measures (i.e., the cited articles and the citing articles) and spatial measures (centrality measures). In general, they pointed out the binding position of cognitive science across psychology, computer science, neuroscience, and education, as well as the relatively minor roles of philosophy and linguistics. They found that psychology was dominant in cognitive science when the import profiles were considered, whereas computer science dominated from the perspective of the export profiles. Leydesdorff et al. (2008) extended their previous study by adding dynamic perspectives on centrality measures to capture the picture for different periods. They observed fluctuations in the interaction between 1994-2006 publications in cognitive science and social psychology, business, human-computer interaction, linguistics, and decision science. A major finding was the intermediary role that the journal Cognitive Science played between cognitive psychology and education research.

The dominance of a single research discipline in cognitive science, or a decrease in contribution from a specific discipline, are factors that also attracted researchers' attention. For instance, Gentner (2010) reported an assessment of psychology in cognitive science for the past 30 years, making a projection of developments for the next 30 years (encompassing a time frame from 1978-2038). Gentner employed socio-institutional analysis (disciplines of the authors) for Cognitive Science publications and text-based analysis (concepts and methods) for the field in general. She emphasized the risks of total domination of cognitive science by psychology. Leydesdorff and Goldstone (2014) supported Gentner's findings of the dominant role psychology has played in the field of cognitive science. Their principal methodology was bibliometric and spatial analysis of Cognitive Science, from 1980-2011. The findings included the dominant role of psychology in cognitive science, neuroscience's central role in the 2000s, and an accelerated split between cognitive science, philosophy, and artificial intelligence. Bergmann et al. (2017) approached the interdisciplinarity of cognitive science by employing a novel metric, based on co-authorship networks. The publications in Cognitive Science from 2005 to 2010 were investigated to detect authors' collaboration patterns, and their findings suggested that the journal was highly interdisciplinary.

Research in cognitive science has also gone beyond the contribution of the primary constituent fields. For instance, Youtie et al. (2017) investigated the interaction between cognitive science and educational research by employing bibliometric and spatial approaches (citation analysis of journal articles for 1994 - 2014). They found an increasing trend, beginning in the 2000s, in favor of the interaction between the two fields, as indicated by the number of citations. Moreover, they found that the border-field journals (such as journals in educational psychology and applied linguistics) had played a bridging role. Further studies by Kwon et al. (2017), Porter et al. (2019), and Solomon et al. (2019) also reported close relationships between cognitive science and education by employing bibliometric and spatial approaches.

On the other hand, researchers have also presented contrasting opinions about the cross-disciplinarity of cognitive science. For instance, Núñez et al. (2019) argued that cognitive science became a "failed interdisciplinary coherent field" (p. 788), in contrast to the main goals of the field's founders. They used two bibliometric and spatial approaches (the authors' affiliations and journal citations) and two socio-institutional indicators (the Ph.D. background and curriculum) to support their position. Their findings revealed an unbalanced research contribution, with psychology as the dominant discipline. A series of response papers were published, which mainly examined the methods and inferences of Núñez et al. (Gray, 2019). According to Schunn (2019), the methods and findings of Núñez et al. (2019) were not applicable to claim the failure of the field. Instead, they opened the doors to discuss redefining the field as plural science ("cognitive sciences").

In parallel to Schunn (2019), Cooper (2019) claimed that the field did not fail. He rejected the idea that cognitive science was established with the intention of developing it as a coherent interdisciplinary field. He also criticized the Lakatosian approach to evaluating the field, agreeing with Núñez et al.'s (2019) findings concerning the changing balance among different disciplines. Cooper investigated article submission statistics in Cognitive Science articles, stating that computer science and logic were indeed vanishing fields in cognitive science research. Gentner (2019) also underlined the founders' intentions for the field. In personal communication with two of the founders, they stated no goals for a single, coherent theory at the beginning. Gentner used the "multilingual (bi- or tri-lingual) people" metaphor rather than creating "combined language" (p. 886) to define the field. Although Gentner (2019) had a similar view about the success of the field, they nonetheless confirmed Núñez et al.'s (2019) findings of unbalance in the field due to the dominance of psychology. Gentner provided a set of suggestions to reduce the dominant effect of psychology on cognitive science. Broude et al. (2019) also rejected the idea that cognitive science is a "dead field" (p. 864), highlighting the mismatch of the Lakatosian criterion and the nonexistence of a single framework. They emphasized the presence of diversity frameworks in the field as indicators of its dynamic, evolving nature, rather than an indication of failure. Moreover, they argued that the institutional identity problem was a political and managerial situation, rather than a problem regarding a field's nature.

Rosenbloom and Forbus (2019) also highlighted the domination of psychology in cognitive science. However, they explained this alignment between cognitive science and psychology as a desirable outcome of their common goals. They did not conceive the lack of cognitive science departments as a failure since most interdisciplinary fields do not have their departments. However, they emphasized the need for a higher share of artificial intelligence and computation in the field. French (2019) pointed out that the empirical findings did not necessarily mean a failure by highlighting that the intentions of the founders were different than Núñez et al. (2019) assumed. According to French, cognitive science had been constituted from diverse but intersecting disciplines.

Similarly, McShane et al. (2019) objected to Núñez et al.'s assumptions and methods (2019), stating that cognitive science did not aim at being a coherent field. Instead, they defined cognitive science as an "integrated science" (p. 915). Their methodological objection was twofold: the limited representative potential of citation information and the effect of historical and sociological factors on the remaining metrics. Bender (2019) presented a critical approach from the perspective of anthropology. She criticized previous analyses regarding the elusiveness of author affiliations investigated, the limited coverage of publication types, and the evaluation citations. Although she noted that anthropology had been less represented in cognitive science, her in-depth text-based analysis of titles/keywords for the journal topiCS, pointed out successful interdisciplinary collaborations. Goldstone (2019) also maintained that psychology was dominant in the field but disagreed with expectations about representing the disciplines equally. Like French (2019), Goldstone suggested cognitive science encompassed more than six fields, including economics, literary studies, medicine, biology, and others, thus supporting an integrative science perspective. In general, Goldstone opposed the criticisms of the expectations from a traditional academic department and the failure to bridge the constituent fields.

Goel (2019) agreed with Núñez et al. (2019) that cognitive science had been dominated by a mono-disciplinary (psychology) approach, in contrast to the expected evolution of a multidisciplinary and interdisciplinary field. He corroborated this thought with a set of conversations with the community. Specifically, Goel described the field by referring to Kuhn's (1962) definition of "pre-science," considering it normal that different paradigms strive for power.

In response to the commentaries above, Núñez et al. (2020) elaborated on the findings and discussions, emphasizing the power of empirical data. First, they stated that there were no claims in their previous study about the failure or death of the field. Instead, there was a failure in the transition from multidisciplinarity to interdisciplinarity. Secondly, Núñez et al. (2020) objected to the use of diversity and plurality as a counterargument to the claims in their previous work, Núñez et al. (2019), maintaining that those terms could not indicate the evolution from multidisciplinarity to interdisciplinarity. The authors also proposed that the response was due to a misunderstood expectation for an interdisciplinary field achieving coherence (not unified monolithic theory) at different levels. Additionally, they emphasized the missing core, fundamental concepts of cognitive science, not core inquiry, as the criticisms claimed. Finally, the intention of the founders to create a coherent, interdisciplinary field was another item for which Núñez et al. (2019) were criticized. Núñez et al. (2020) replied to this denial with historical records presenting the field's goal, stating that some commentators used the term integrative science to specify the nature of cognitive science.

More recently, Oey et al. (2020) (a revised version was published as DeStefano et al. (2021)) focused on the Cognitive Science Society's past 19 years (2000-2019) of studying the level of interdisciplinarity. They combined bibliometric and spatial approaches (edge density, transitivity, maximum subgraph size for co-authorship networks) and text-based metrics (topic similarity). The analysis showed that the interdisciplinary structure of cognitive science had been rising. Similarly, Kallens et al. (2022) employed Latent Semantic Analysis (LSA) as a semantic text similarity measure to compare cognitive science related journals and other topical journals. Their concept of interdisciplinarity was based on "mixture of expertise" for abstracts of multi author publications for the years between 2005 and 2018. The results suggested that cognitive science had higher interdisciplinarity structure with diversity in mixture of expertise when compared to topical fields.

CHAPTER 3

METHODOLOGY

The primary objective of the present study is to address whether the semantic text similarity analysis, a key component of Information Systems (IS) applications, provides consistent similarity scores for the relevance of academic documents to their research fields. In this context, consistency refers to reliability of the scores. If semantic text similarity analysis can accurately categorize academic documents into their relevant fields, it suggests a potential for quantifying cross-disciplinarity of fields. This conclusion stems from the understanding that a cross-disciplinary field employs other related fields' tools and terminologies. Therefore, we expect to find contextual and semantic relationships between fields. So, the subsequent question this study aims to answer is: What does text similarity analysis explain regarding the cross-disciplinarity of fields? This investigation takes the form of a case study that focuses on cognitive science and related fields. Specifically, the present study employed the semantic text similarity approach to evaluate the cross-disciplinarity of cognitive science. The techniques included Doc2Vec (for vectorization) and cosine similarity (for measuring the similarity among the articles). The methodology and analyses detailed in the subsequent sections aim to fulfill the dual objectives of this study: to illustrate a novel application of text similarity analysis within the IS domain, and to enhance our understanding of the cross-disciplinarity of cognitive science.

This section will be organized in the following order. First, we define semantic text similarity and discuss its general applications. Next, we delve into the method used to vectorize texts, specifically Doc2Vec. Following this, we focus on similarity measures, with a particular emphasis on cosine similarity, which is used in this research to compare different vectors. After establishing these general definitions and descriptions, we outline the operational assumptions in the implementation of the methodology. Finally, we describe the processing pipeline of the research methodology, explaining the three major stages designed to answer our research questions.

3.1. Semantic Text Similarity Concept

Semantic text similarity, which is a core concept of Natural Language Processing (NLP), is a measure to compare two or more texts based on their semantic relationship. It transcends traditional methods that rely solely on the simple lexical matching[4] (Islam & Inkpen, 2008). Unlike traditional methods that focus on exact word matches, semantic text similarity goes farther, investigating words within their context, which usually provides a more accurate comparison (Mihalcea et al., 2006). Semantic text similarity is assumed to uncover semantic relationships in various textual units, ranging from a single word or n-gram to a sentence, paragraph, article, or even an entire book. This relationship is defined in terms of percentage of similarity or dissimilarity.

The emergence of text similarity measures was supported by growth of textual data in various formats from social media posts to academic books. Additionally, the necessity for efficient processing, classification and interpretation of textual data also led to development of text similarity measures. Initially, these measures focused on lexical matching, which counted words and phrases that were often used. However, this approach ignored the context and meaning. As the variety and volume of data increased, the focus shifted towards semantic text similarity, which examines how textual elements are similar in terms of semantic relations. This advancement has enhanced the capabilities of NLP applications. Among these applications, semantic text similarity analysis has found a wide range of uses. For instance, in text categorization, Ko et al. (2004) utilized it to classify sentences important or unimportant based on similarity to the title of a document. Another sentence classification study was conducted by Kim (2014) by employing convolutional neural network and word embeddings. Mohamed & Oussalah (2019) proposed a graph-based text summarization framework for single and multi-document with semantic similarity. Park et al. (2005) used short text similarity to improve the recall and precision of a search engine in retrieval effectiveness tasks. Kim et al. (2017) also focused on information retrieval effectiveness by measuring query-document similarity in the cases when there were no direct matches between a query and a document. In semantic comparison, Chen et al. (2018) used question generation and answering mechanisms to discover content differences between the original and new text passages. Zhu & Iglesias (2018) proposed a method for entity disambiguation that employs semantic similarity between contextual words and informative words of entities to clarify the meaning of a word in a specific context. Hirst & Budanitsky (2005) used semantic text similarity in spelling error detection to identify and correct real-word spelling errors by looking for words that don't fit contextually with the rest of the text. Tien et al. (2019) applied semantic similarity between sentences in textual entailments to determine if one sentence logically follows from another. Lastly,

[4] Simple lexical matching means calculating similarity by counting the number of common words or phrases present in both compared texts.

Nguyen et al. (2019) improved the performance of paraphrase identification by using interdependent representations between short texts.

Chandrasekaran & Mago (2021) classify semantic text similarity methods into four categories: knowledge-based, corpus-based, deep neural network-based and hybrid methods. Knowledge-based methods compute semantic similarity between two texts relying on external sources like lexical databases (e.g., WordNet, Wiktionary). Corpus-based methods utilize large corpora and employ statistical techniques to compute the degree of similarity between texts. They are grounded in the principle that similar words occur together, thus their vector representations in high-dimensional space are also close. The vector representations are created by word embeddings, meaning that converting text into high dimensional vector. Techniques such as Word2vec, GloVe and fastText and BERT exemplify this approach, where texts are converted into vector forms. Deep neural network-based methods have evolved with advancements in neural networks and include techniques like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Recursive Tree LSTM. Although these methods also use word embeddings (which are generated from large text corpora), the key distinction is that deep neural networks are employed to estimate the similarity between these embeddings. Hybrid methods combine techniques from aforementioned semantic similarity approaches. The idea is to leverage the strengths of each method while mitigate the weaknesses. For instance, they might combine the structural efficiency of knowledge-based methods (which use structured data like ontologies) with the versatility of corpus-based methods (which use statistical analysis of large text corpora).

Applying semantic text similarity typically requires following common steps as outlined in Figure 3.

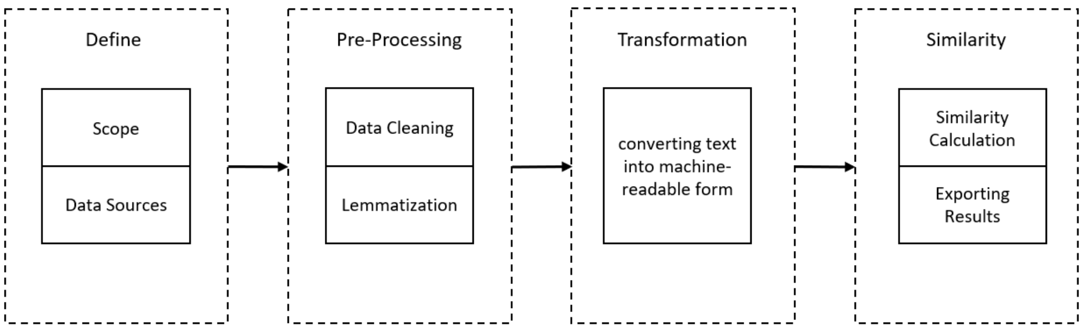


Figure 3: Common steps of semantic text similarity application

The first step typically involves defining the scope of analysis, identifying appropriate data sources, and gathering the necessary data for analysis. As the data is produced and published in a variety of forms ranging from strictly structured academic articles to an unstructured tweet, the subsequent step, often referred to as preprocessing, requires contextual processing. At this step, based on data structure and research objectives, various operations are applied based on the data structure and research objectives, including removal of special characters or numbers, elimination of stop words (like ‘a’, ‘the’, ‘and’ etc.), case conversion, tokenization, and lemmatization. These operations convert raw data into a format suitable for the next step of analysis. The following step involves converting text into machine-readable form, typically by transforming it into numerical values or vectors, especially for corpus-based semantic text similarity models. There are various models in the literature that outline different methods of this vectorization process. It could be a straightforward method which calculates the number of term occurrences or could be a complex deep learning algorithm. In the final step, the vector values are compared using similarity measures, which calculate distances or similarities between two vectors. This step is crucial in determining the degree of similarity between the analyzed texts.

3.2. Vectorization

Corpus-based text similarity requires transformation of texts into machine readable format, such as vectors. There are various methods of creating vectors from the text. The most fundamental of these is the Bag-of-Words (BoW) model (Manning et al., 2009, p. 117), which quantifies frequency of each unique word in a document and represents the document in terms of word frequencies. In this model, unique words are assigned to numbers and their frequencies in the corpus are computed to create vectors. For instance, consider two sentences with the following word frequencies: ‘cognitive’ (the 1st term) appears twice, ‘science’ (2nd term) four times, ‘field’ (3rd term) three times. These sentences can be vectorized as {(1,2), (2,4), (3,3)}. By definition, the dimensions of the vector will equal the total number of unique words in the text.

Building on the BoW concept, Term Frequency-Inverse Document Frequency (TF-IDF) emerged as one of the earliest and most widely used methods for vectorization. The significant contribution of TF-IDF lies in its calculation of vector values, considering the relative importance of words (Jones, 1972). The 'TF' in TF-IDF stands for 'term frequency', which assigns a weight to a word based on its frequency in a specific document. Conversely, 'IDF' refers to 'inverse document frequency,' which assigns a weight based on the rarity of the word across the entire corpus. In essence, a word that appears frequently in a document but is uncommon in the entire corpus will have a higher weight.

As an alternative to the TF-IDF model, numerous corpus-based vectorization methods have been developed to capture relationships in text. For instance, Latent Semantic Analysis (LSA) has been proposed to reveal hidden semantic structures within a collection of documents by reducing the dimensionality of the TF-IDF vector to a smaller topic vector (Deerwester et al., 1990). Similarly, Latent Dirichlet Allocation (LDA), a generative probabilistic model that assumes each document is a mixture of a certain number of topics, has been developed for topic modeling tasks, where the topic of a document is represented as a vector (Blei et al., 2003). It can identify the main themes in a large corpus of documents. Mikolov et al. (2013a, 2013b) developed Word2vec which is a neural network-based word embedding method that creates vector representations of words in a high dimensional vector space based on the context of words in the corpus. Global Vectors for Word Representation (GloVe) has been developed as an unsupervised model which creates contextual word embeddings by combining both matrix factorization methods and context window methods to capture semantic relationships (Pennington et al., 2014). FastText, an extension of Word2vec, has been designed to address the problem of words that appear in a test corpus but were unseen in the training corpus by using sub-word models (Bojanowski et al., 2017). Embeddings from Language Model (Elmo) have been developed to produce deep context sensitive word representations. This means that the same word can have different embeddings depending on its context. Lastly, Bidirectional Encoder Representations from Transformers (BERT) is a Transformer-based model pre-trained on a large corpus of text and then fine-tuned for specific tasks (Devlin et al., 2019). It considers the context from both left and right sides of a word in all layers. This makes it particularly effective for tasks that require a deep understanding of context.

Researchers select the most suitable corpus-based vectorization method according to their specific goals and the nature of data they are working with, as different methods are designed for distinct applications. For instance, TF-IDF is commonly used in straightforward tasks such as keyword extraction or information retrieval. It is particularly useful at defining important words in the context of individual documents within corpus. Word2vec, on the other hand, is better suited for semantic analysis research such as semantic similarity measurement. GloVe is often employed in sentiment analysis or text classification tasks. BERT is effective for tasks that need a deep understanding of context like question answering tasks. FastText is particularly effective in scenarios where dealing with out-of-vocabulary words is crucial.

This dissertation focuses on the idea of measuring similarities between documents. It requires employing a document-level embedding method that can efficiently handle documents of varying sizes and capture semantic similarities. Doc2Vec, as an extension of Word2vec, is employed for this purpose due to its ability to produce vectors from documents based on their semantic relationships. The Doc2Vec algorithm performs better than its alternatives, such as the bag of words and the TF-IDF (Term-frequency-inverse document frequency), since Doc2Vec is assumed to preserve semantic relations (Kim et al., 2017).

Doc2Vec is an unsupervised learning algorithm that uses a three-layered neural network (input-hidden-output) to vectorize input documents (Le & Mikolov, 2014). It is a modified version of the Word2vec model used for word vectorization (Mikolov et al., 2013a, Mikolov et al., 2013b). Doc2Vec adds a document identifier to Word2vec to capture contextual information. This means that a solid understanding of Word2vec operates can help us to understand how Doc2Vec works.

Word2vec is a neural network-based word embedding method (Mikolov et al., 2013a, Mikolov et al., 2013b). It has gained popularity due to its ability to create vector representations of words within a high-dimensional vector space, using a large training corpus. The output of Word2vec is representative vectors for words. The values of these vectors are defined according to semantic similarity of words, with more similar ones being mapped in neighborhood points in the vector space. This is achieved by training words in their natural context in which they appear, resulting in similar vector representations for words that occur in similar contexts. This functionality of Word2vec supports researchers working with various text-related problems such as sentiment analysis, classification, synonym detection, entity recognition. There are two main models that are used in prediction mechanisms: continuous bag-of-words (CBOW) and Skip-gram, as shown in Figure 4.

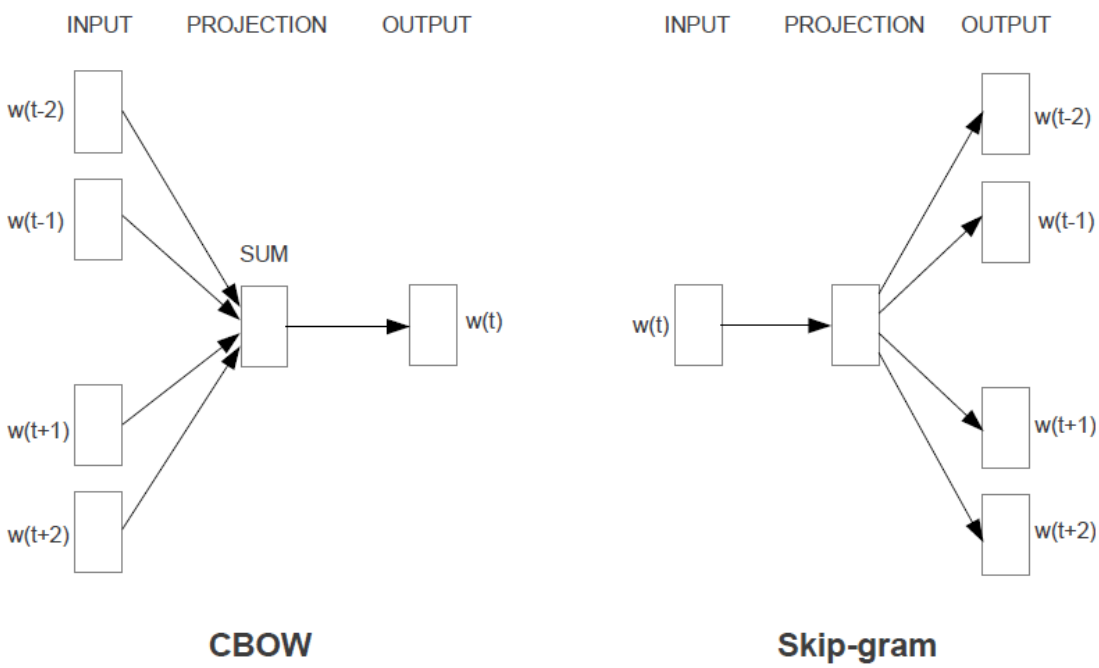


Figure 4: CBOW and Skip-gram architectures used in Word2vec model (Redrawn by the author, based on the original source Mikolov et al., 2013b)

In the CBOW architecture, probabilities of words are computed by predicting a selected word based on its surrounding words within window size at each step. Window size defines the number of nearby words which is specified by the user. For example, let's say we set the window size as 1 and use the following sentence to train the model: "cognitive science attracts attention from different disciplines". As illustrated in the Figure 5, the sliding window covers one word on either side of the central word at each step. In this case, the center word corresponds to output layer whereas the surrounding words represent the input layer within a three-layer neural network.

Conversely, the Skip-gram architecture operates in the reverse manner when compared to CBOW. It predicts nearby words with a selected center word. This means that there is only one word (the central word) in the input layer and two words (surrounding the central word) in the output layer.

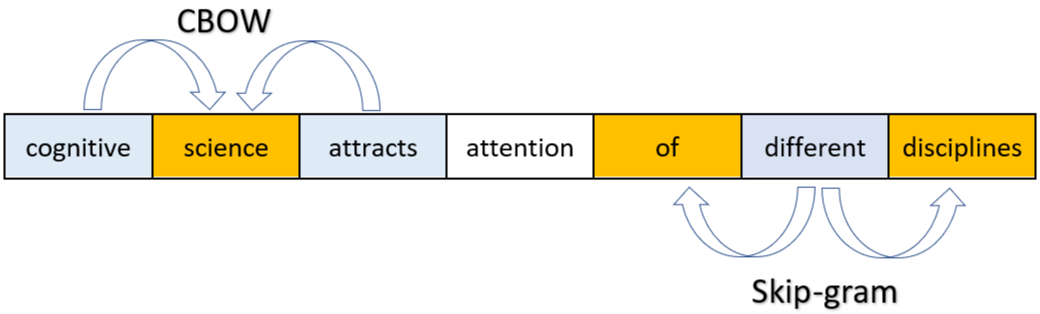


Figure 5: How CBOW and Skip-gram work with nearby words in window size

As previously noted, Word2vec utilizes a neural network to create vector representations of words. The architectures of CBOW or Skip-gram specify the input and output layers of this network. However, the hidden layer requires more mechanisms to be explained. Regardless of whether CBOW or Skip-gram architecture is applied, several techniques are employed to optimize the training process. These include *gradient descent with backpropagation*, *negative sampling*, *Hierarchical SoftMax based on the Huffman tree*. The training process in Word2vec relies heavily on *Gradient descent with backpropagation*. It forms the backbone of Word2vec, and it optimizes the model's performance. Its primary role is to minimize the discrepancy between the predicted and the actual word vectors. It refines the model's parameters using error feedback. The process starts from the output layer and the error is propagated backwards through the network. This procedure helps in determining the contribution of each neuron in the hidden layer to the total error. Therefore, it allows the model to fine-tune its weights and bias values during training to minimize the error. *Negative sampling* serves to expedite the training process and improve the quality of the generated word vectors. As noted in *Gradient descent with backpropagation*, training the neural network requires adjusting the weights to accurately predict word given a specific context. However, this implies a substantial computational cost due to the large number of weights in the model. *Negative sampling* addresses this issue by updating only a small percentage of the weights at each step of the training process. Another technique used to enhance computational efficiency of the training process in Word2vec is *Hierarchical SoftMax based on the Huffman tree*. This technique is helpful when dealing with a large vocabulary set. It simplifies the issue into a series of binary decisions rather than calculating the probability for every individual word in the vocabulary. The decisions form a tree, and each leaf node of the decision tree represents a word from the vocabulary. As a result, fewer decisions are required to reach the most frequent words since they are placed closer to the tree's root. These techniques enable Word2vec to learn high-quality word embeddings from varying size of datasets, thereby supporting its prevalent use in various natural language processing tasks.

Expanding upon the foundational principles of Word2vec, this section delves into the extended capabilities offered by Doc2Vec. It works in a similar way to Word2vec, with an added feature known as document vector. Within this model, document vectors contribute to prediction tasks. During the training process, a unique vector is assigned to each document which encapsulates a larger contextual scope than the individual words. The model has the capability of learning "fixed-length feature representations" from pieces of text of varying lengths (Le & Mikolov, 2014). Like Word2vec, Doc2Vec has two architectures. The first one is called Paragraph Vector-Distributed Memory (PV-DM), operates like Word2vec's CBOW architecture as explained previously. The second architecture is Paragraph Vector-Distributed Bag of Words (PV-DBOW), functions similarly to the Skip-gram model in Word2vec. To present the differentiation over an architecture, PV-DM is illustrated in Figure 6.

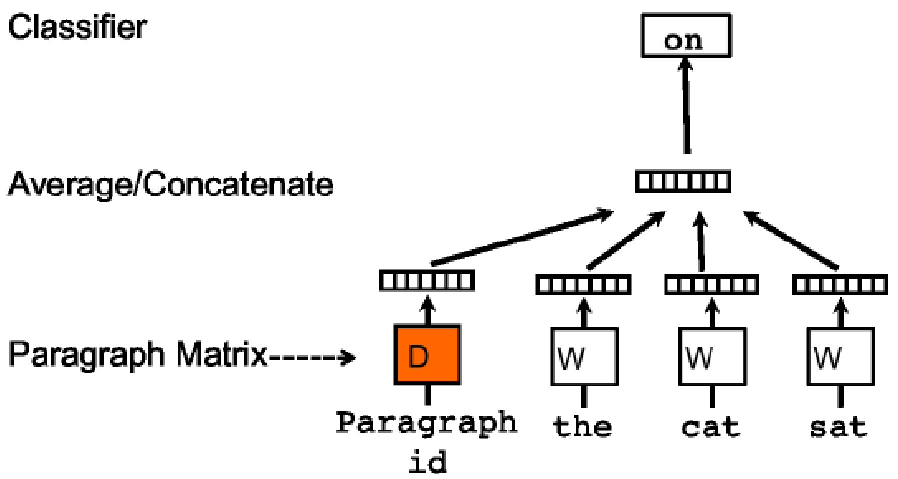


Figure 6: PV-DM architecture in Doc2Vec (Redrawn by the author, based on the original source: Le & Mikolov, 2014)

PV-DM is visualized as a three-layered neural network. Context words and paragraph vectors represent input layer. The hidden layer consists of operations that either average or concatenate the input vectors. The output layer represents the prediction of the next word in the text. While the word vectors are shared across paragraphs, a key point in this architecture is the uniqueness of the paragraph vector to each specific paragraph. During training, the paragraph and word vectors are learned by trying to predict or correctly classify the following word in a text.

3.3. Similarity Measures

In NLP applications, text data is converted into vectors (numerical representations) for analysis. This vectorized textual data is then compared using similarity measures. These measures of text similarity are an important part of text relevant applications notably in the domains of information extraction, text generation, recommender systems, text classification, clustering, and topic detection.

Several metrics for measuring text similarity have been developed. These include Euclidean Distance, Cosine Similarity, Jaccard Similarity, Jenson-Shannon Divergence. Among these, cosine similarity is one of the widely used measures in vector space due to its simplicity and ignorance of vector's magnitude. The basic idea behind it is to calculate the cosine angle between two vectors (in this research they are vectors of academic documents) as provided in (Eq. 1.

$$\cos(v1, v2) = \frac{v1.v2}{\|v1\|\|v2\|} = \frac{\sum_{i=1}^{n} v1i.v2i}{\sqrt{\sum_{i=1}^{n}(v1i)^2}\sqrt{\sum_{i=1}^{n}(v2i)^2}} \qquad \text{(Eq. 1)}$$

In this formula, v1 and v2 are two vectors of different documents. The sign of · denotes the dot product of the two vectors, ||v1|| and ||v2|| indicates the magnitudes (i.e., lengths) of the two vectors. The division in the formula results in a range between 1 and -1. If the result is near to 1 it means these two vectors have high similarity. The value around 0 shows no significant similarity. On the other hand, the values near to -1 indicate opposed vectors. The following example will present basically how cosine similarity works for two vectors V1 (1,1) and V2 (0,1) consisting of only two dimensions.

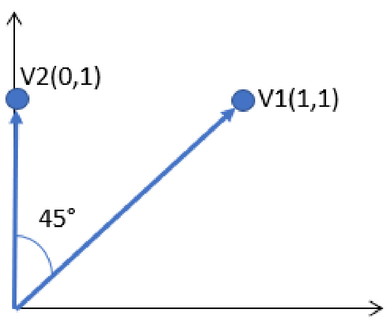


Figure 7: Cosine similarity of vectors in two-dimensional space

In Figure 7, the angle between two vectors is 45° which means cos(45)=0.71. The formula will also provide the same result.

$$\cos(v1, v2) = \frac{\sum_{i=1}^{n} v1i.v2i}{\sqrt{\sum_{i=1}^{n}(v1i)^2}\sqrt{\sum_{i=1}^{n}(v2i)^2}} = \frac{(1x0) + (1x1)}{\sqrt{1^2 + 0^2}\sqrt{1^2 + 1^2}} = 0.71$$

3.4. Operational Assumptions

The process of implementing the methodology requires making a set of operational assumptions. This research is based on several assumptions, especially in defining scope and data sources.

The first assumption pertains to our choice of the term "cross-disciplinarity". Although there are different terms like interdisciplinarity, multidisciplinarity, or transdisciplinarity, the methodology applied in this research is not entirely comprehensive in capturing the differences between those concepts. Moreover, these concepts are subject to intense debate in the literature as provided in the Literature section. Therefore, the term cross-disciplinarity is used to cover any kind of interaction between fields.

Another major assumption concerns the target of analysis. Cross-disciplinarity may be evaluated in various dimensions, such as through author profiles, courses, funding agencies, or organizational features. A major operational assumption of the present study is that the outputs of academic activities are limited to academic publications to make them quantifiable to reveal the domain's cross-disciplinary characteristics. Therefore, we limit the scope of our investigation of cross-disciplinarity to the measurement of academic publications (journal articles and conference proceedings, as described below). Accordingly, we quantified the contribution of relevant research domains to cognitive science in terms of domain-specific publications. We also assumed that the leading target venue for cognitive scientists (and researchers who aim to contribute to the field of cognitive science) is the Annual Meetings of the Cognitive Science Society (CSS), conducted since 1979. For our analysis, we looked closely at the articles published in journals related to cognitive science and at the proceedings of the CSS meetings, assuming both would offer a representative picture of cross-disciplinarity in the field. Moreover, the titles and the abstracts of articles published in relevant journals were used to create the contributing disciplines (philosophy, anthropology, linguistics, neuroscience, computer science, and psychology) corpus. This limitation is required since the retrieval of full text documents is not practical. The titles and abstracts can be defined as the minimum representative unit of an academic document.

Another assumption at this stage pertains to the selection of highly cited articles for each domain. By definition, a field's highly cited articles are assumed to be best representatives of that field since they attract the attention of others who are also studying similar subjects.

The last assumption concerns the six contributing fields as listed in the reproduced version of hexagon presenting cognitive science-related fields in Alfred P. Sloan Foundation report (Miller,2003). These fields are philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. Although these fields are important contributors, it should be emphasized that cognitive science is an ever-evolving field, therefore new contributing fields may appear over time.

3.5. Processing Pipeline

To answer the research questions within the defined scope, this study systematically follows the steps as illustrated in Figure 8. The first step is data source definition and data collection for the train and test sets which will be used throughout all steps of the study. The subsequent step is pre-processing, which encompasses text tokenization, lemmatization, and corpus creation. This step ensures that the raw data is cleaned and structured in a way that is suitable for creating corpus. Following pre-processing, we embark on the modeling step. We set (and refine based on the classification success of the model) specific parameters for Doc2Vec, and we built the vocabulary to train the model. After the model is trained, we transition to the similarity analysis step. In this step, the trained model is used to generate vectors for the train and test sets. These vectors are then used to calculate cosine similarity scores, which provide input for discussions regarding the cross-disciplinarity of the field under investigation.

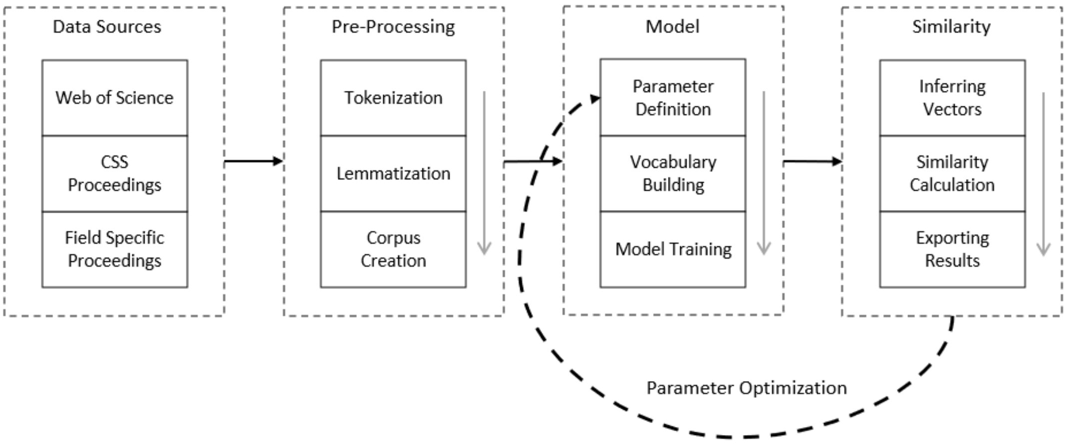


Figure 8: Processing pipeline of the systematic steps followed in the study - from data collection to similarity analysis.

Throughout each of these steps, we document our findings, building a solid base that can be used by other researchers in the future. It also facilitates a comprehensive understanding of our subject matter.

3.5.1. Data Sources

In line with our operational assumptions, we utilized academic publications as input data. Specifically, these were the titles and abstracts of journal articles and the full texts of conference proceedings. Journal articles are mostly accessible from scholarly databases (such as WoS in our case), making it simple to export this structured data. On the other hand, conference proceedings are generally published in a format that requires more effort to extract relevant data, and hence, we refer to them as unstructured data.

Upon defining our major data sources, it is important to mention that there will be variations of data sources across different stages of analysis. For clarity, we have divided our research into three main stages, which we will follow in sequential order. The results of the research will also be presented in the same structure in the Results chapter.

As we designed and developed our models, the first stage was a validation study in which we used only journal articles. This initial stage aimed to analyze how successful the model is in providing information about the relationship between well-defined academic articles and their respective research fields. In the second stage, we shifted our focus to field-specific proceedings and randomly selected conference proceedings, serving as test sets. This stage allowed us to observe the model's performance when working with non-standardized documents. Finally, in the third stage, we analyzed CSS proceedings to investigate what a text-similarity analysis may reveal about the field's cross-disciplinarity as introduced in the following sections. Each stage required developing field-specific or general training and test data sets, as presented in Table 1.

Table 1: The use of data sets in the three stages of the pipeline.

Datasets	**Stage 1: Creating the Similarity Matrix for Journal Articles**	**Stage 2: Running the Doc2Vec Model for Field-Specific Proceedings**	**Stage 3: Running the Model for the CSS Proceedings**
Training	A subset of field-specific journal articles (70%)	A complete set of field-specific journal articles (100%)	A complete set of field-specific journal articles (100%)
Test	A subset of field-specific journal articles (30%)	Field-specific conference proceedings	CSS Meeting proceedings

This table outlines the breakdown of the training and testing datasets used during each stage of the research. In Stage 1, 70% of the field-specific journal articles were used for training, while the remaining 30% were used for testing. In Stages 2 and 3, the entire set of field-specific journal articles was used for training. The test datasets in Stage 2 were field-specific conference proceedings, while in Stage 3, the CSS Meeting proceedings were used. The next part describes the details of training and test data.

3.5.1.1. *Field-specific Journal Articles*

The first and main dataset is field-specific journal articles, used for model training and testing. To construct this dataset, we selected titles and abstracts of the top 18,000 most-cited articles for each relevant field of cognitive science (namely philosophy, anthropology, linguistics, neuroscience, computer science, and psychology), assuming that the highly cited articles in the specific fields are acceptable representatives in each field. More specifically, for each relevant field, the titles and the abstracts of the most-cited articles were extracted from the Clarivate Analytics Web of Science (WoS) database for the period between 1990 and 2019. The relevant WoS article categories and queries used for this process are presented in Table 2. It is important to note that we excluded journals assigned to multiple research field categories to create a homogeneous field corpus.

Table 2: Web of Science (WoS) categories used to create cognitive science related fields.

Fields	WoS Categories
Philosophy (Phil)	Philosophy OR Ethics OR Religion
Anthropology (Anth)	Anthropology OR Evolutionary Biology
Linguistic (Ling)	Language & Linguistics OR Linguistics
Neuroscience (Neuro)	Neuroimaging OR Neurosciences OR Clinical Neurology
Computer Science (Comp)	Computer Science Artificial Intelligence OR Computer Science Cybernetics OR Computer Science Information Systems OR Computer Science disciplinary Applications OR Computer Science Software Engineering OR Computer Science Theory & Methods OR Computer Science Hardware & Architecture OR Logic
Psychology (Psych)	Psychology OR Psychology Applied OR Psychology Biological OR Psychology Clinical OR Psychology Developmental OR Psychology Educational OR Psychology Experimental OR Psychology Mathematical OR Psychology Multidisciplinary OR Psychology Psychoanalysis OR Psychology Social OR Psychiatry OR Behavioral Sciences OR Ergonomics

This dataset served as the primary training data throughout all stages of our study. During the first stage, we divided the dataset into two portions: 70% (amounting to 12,600 articles for each field) was utilized for training data, 30% (equivalent to 5,400 articles for each field) was allocated for test data. In subsequent stages, the model was trained with the whole set, which comprised 18,000 articles for each field. This approach ensured that we had a comprehensive and robust training dataset to help fine-tune our model.

3.5.1.2.Field-specific Conference Proceedings

After establishing the dataset for field-specific journals, we constructed a separate dataset for field-specific proceedings, which was used as a test set in stage 2. For this, we extracted proceedings from specific years of seven conferences in cognitive science-related fields, such as philosophy, computer science, and psychology. For instance, to represent field of philosophy, we downloaded Proceedings and Addresses of the American Philosophical Association Vol. 90, published by the American Philosophical Association in 2016. We also included the proceedings of Euradwaste, a conference dedicated to radioactive waste, as a non-relevant field in this set. Further information about the dataset is presented in Table 3.

Table 3: Selected field-specific conference proceedings, their fields, and full names

Conference	Field	Full Name
EURADWASTE_2013	Energy	8th EC conference on management of radioactive waste community policy and research on disposal
Philosophy_IAFOR_2016	Philosophy	The Asian Conference on Ethics, Religion & Philosophy 2016 & The International Academic Forum ACP/ACERP 2016
Philosophy_APA_2016	Philosophy	Proceedings and Addresses of the American Philosophical Association Vol. 90, NOVEMBER 2016
Philosophy_APA_2012	Philosophy	Proceedings and Addresses of the American Philosophical Association Vol. 86, No. 2, November 2012
Computer_CVPR_2016	Computer Science	2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Highly cited top 50 articles full text
Computer_FEDCSIS_2016	Computer Science	Proceedings of the 2016 Federated Conference on Computer Science and Information Systems
Psychology_InPACT_2015	Psychology	International Psychological Applications Conference and Trends (InPACT) 2015
Psychology_IAFOR_2016	Psychology	The Asian Conference on Psychology and the Behavioral Sciences 2016 & The International Academic Forum ACP/ACERP 2016

3.5.1.3.Cognitive Science Society Meeting Proceedings

This section presents our evaluation of the Doc2vec model for assessing the cross-disciplinarity of conference proceedings, usually considered timely and representative collections of documents in a field. Conferences have their own cultures, established by their constituent scientific communities. They are conducted under prevailing and traditional conventions that researchers use to present their findings and get feedback (Martens & Saretzki, 1993; Zierath, 2016). Conferences overcome the limited interaction of journal articles by providing a highly interactive environment for dissemination and collaboration (Mark Hickson, 2006; Sanders et al., 2020). According to Alberts (2013), conference meetings play an essential role in establishing new disciplinary perceptions by providing a “random collision of ideas" and incorporating other researchers’ expertise in “non-obvious ways.” In our opinion, the Annual Meetings of the Cognitive Science Society (CSS) offer an appropriate venue for the dissemination and discussion of academic activities carried out by researchers in the cognitive science community. Schunn et al. (1998) stated that annual CSS meetings had been the main “communal world” for researchers in the cognitive science field since 1979. CSS characterizes these annual conferences as primary resources for accessing new advancements in the study of the mind. They attract the attention of a huge number of researchers around the world. For instance, the number of attendees at the 44th Annual Meeting in 2022 was 1436. The average number of articles published in the proceedings is around 800, including various types of submissions like full papers, member abstracts, and poster presentation papers. We also believe that articles published in the CSS meeting proceedings provide an appropriate representation of the cross-disciplinarity of cognitive science. In the present study, we investigated the CSS meeting proceedings to understand the cross-disciplinarity of cognitive science and its relationship with other disciplines. The CSS meeting proceedings included full texts of all the Annual Meetings of Cognitive Science Society conferences published between 1981 and 2022. We crawled the proceedings to create test sets (Cognitive Science Society Past Conferences, 2019; CogSci Proceedings, 2020; Hope, 2011)

3.5.2. Pre-processing

Following the collection and consolidation of data, pre-processing is required to prepare the data for model training. This process consists of three primary operations: tokenization, lemmatization, and corpus creation. Tokenization involves segregating the text into smaller units called tokens, which are defined as individual words in this research. Tokenization forms an essential preliminary step for further cleaning operations and vocabulary construction.

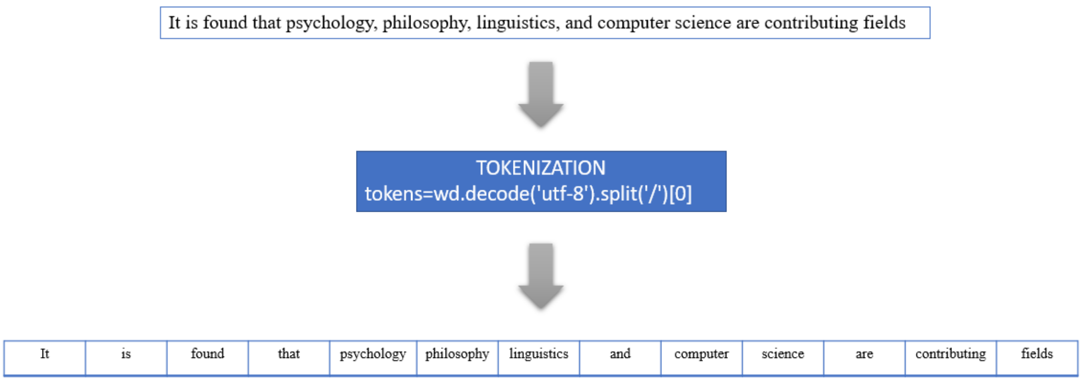


Figure 9: How tokenization works for a sample sentence

Figure 9 demonstrates tokenization applied to a sample sentence by using the provided code. The result shows that each word is recognized as a separate entity and punctuation is removed. This output will subsequently serve as input for the lemmatization phase of the pre-processing task. Lemmatization converts words into their root forms, removes stop words, and retains only nouns, verbs, adjectives, and adverbs. We employed Gensim's lemmatization function for this purpose.

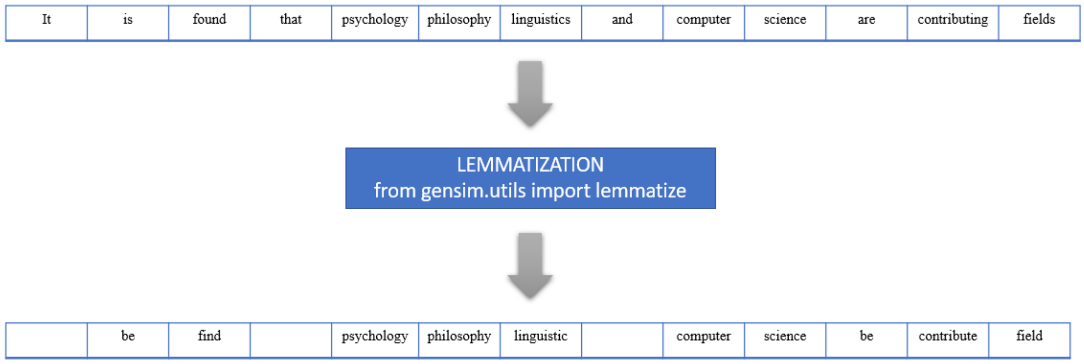


Figure 10: How lemmatization works with Gensim's lemmatization function

Figure 10 exhibits the application of lemmatization to the same sample sentence. Words like "it", "that", and "and" have been eliminated (full list of eliminated stop words are listed in Table 19 in Appendix A). There are several words transformed into their original base forms, such as "found" to "find" or "contributing" to "contribute". Additionally, plural forms have been switched to their singular counterparts, like "fields" to "field".

In the final phase of pre-processing, we construct a test and train corpus based on the tokenized and lemmatized texts. As per the Doc2Vec model's requirements, while the test set is formulated as a list of lemmatized tokens (of type <*class 'list'*>) using the code *yield list(tokens)*, the training set is established as tagged documents using the *yield TaggedDocument (tokens, [i])* line of code (of type *<class'gensim.models.doc2vec.TaggedDocument'>).*

3.5.3. Model

In this research, Gensim's Doc2Vec model is used to create vectors of academic documents. Constructing and utilizing our Doc2Vec model involves three primary steps: parameter definition, vocabulary building, and model training. As illustrated in the Figure 8, these stages form a sequential process that guides the transformation of raw text data into meaningful representations.

Before embarking on these steps, we need to ensure that we have the necessary Gensim's relevant libraries at our disposal. The Doc2Vec module within Gensim, specifically, is what we'll be utilizing for our task. To start the process, we must import these relevant libraries into our Python environment. The necessary lines of code to do this are:

import gensim

from gensim.models.doc2vec import Doc2Vec, TaggedDocument

With Gensim and its Doc2Vec module successfully imported, we can now delve into the stages of our model's construction and operation.

3.5.3.1.Parameter Definition

Doc2Vec models are built on a set of configurable parameters, variations of which can enhance or limit the performance of models. In Gensim's Doc2vec model, there are key parameters that need to be configured by the user to reach optimal results in the following way (Table 4 also provides parameters, their explanations, and values):

```
model = gensim.models.doc2vec.Doc2Vec(
                                        dm=1,
                                        vector_size=300,
                                        window=8,
                                        min_count=3,
                                        epochs=20,
                                        dbow_words=0)
```

The first parameter is called *distributed memory (dm)*. It defines which training algorithm will be employed. It is set as "*dm=1"* if the training algorithm is expected to be *Paragraph Vector - Distributed Memory (PV-DM). PV-DM* operates in a similar way to *CBOW* architecture in Word2vec with a paragraph identifier addition, as described in the vectorization section. A specific word is predicted by other surrounding words within a specified window size. Alternatively, the other option is setting "*dm=0*", which means the training algorithm will be *Paragraph Vector – Distributed Bag of Words (PV-DBOW).* In that case, the architecture is akin to Word2vec's *Skip-gram* algorithm. It predicts nearby words in a window size with a selected center word. Independent of the chosen architecture, *dm* requires a *window* parameter for prediction work. The *window* can be any integer value which specifies the maximum distance between predicted word and context word within a document. Another important parameter is the *vector size* which can be basically defined as dimensionality of the feature vector for each document. In the definition, it is mentioned that the aim of the model is to represent documents as vectors. These vectors are fixed length and can be defined by the user. The trade-off between selecting a larger or smaller size of vector is that a smaller value may result in ignorance distinctions between documents, whilst a larger value may result in long training times and disregarding of the similarities. The *min_count* parameter can be used to decrease the noise by setting a threshold value and eliminating rare words from the corpus. It ignores all words that occur less than the given number in this parameter. As a result, words that don't offer any valuable information to the features of a document are removed. The model requires iterations to increase the accuracy of the document representations. The *epochs* option specifies how many times the algorithm will run over the training data. Similar to the *vector size* parameter, the chosen number of *epochs* will have an effect on the model's accuracy. There is a risk of overfitting or underfitting problems for too high or too small numbers. The last parameter that will be explained is *dbow_words* which has two options. "*dbow_words=1*" means the model will also include word vectors in the *Skip-gram* mode. The default value is "*dbow_words=0*" in which the *DBOW* algorithm does not take words into account, it only creates representations with document level focus. The impact of varying these parameters on the model's prediction accuracy will be explored in the 'Parameter Optimization' section. Aside from these key parameters, others such as *seed, max_vocab_size,* and *workers* can also be adjusted according to user requirements.

Table 4: Key parameters configured for this research and their explanations.

Parameter	Explanation	Value
dm	type of training algorithm will be employed	1: PV-DM
vector_size	dimensionality of the feature vector	300
window	maximum distance between predicted word and context word	8
min_count	threshold value for eliminating rare words	3
epochs	number of times the algorithm will run over the training data	20
dbow_word	the level of focus	0: document level

The complete list and detailed explanations can be found in Appendix B.

3.5.3.2.Vocabulary Building

Following the parameter configuration, vocabulary of training document is generated using the following code:

model.build_vocab(train_corpus)

This method starts with generating a distinct list of every word in the corpus and assigning a unique id for each word. This is a required step to train the model since the model needs to know the frequencies to eliminate the ones lower than *min_count* value. Moreover, it helps to make an estimation of the size of the training set for the model. This estimation will be used to gradually decrease the internal alpha learning-rate in each epoch. As a result of the vocabulary building process, the model creates required objects and reserves sufficient memory based on rough estimation of the training set. The output of the log file while creating vocabulary of main train set described in stage 3 (A complete set of field-specific journal articles (%100)) is Figure 11:

```
2023-04-07 10:38:27,291 : INFO : collecting all words and their counts
2023-04-07 10:38:27,291 : INFO : PROGRESS: at example #0, processed 0 words (0/s), 0 word types, 0 tags
2023-04-07 10:38:30,696 : INFO : collected 129846 word types and 7 unique tags from a corpus of 7 examples and 12963033 words
2023-04-07 10:38:30,698 : INFO : Loading a fresh vocabulary
2023-04-07 10:38:30,840 : INFO : effective_min_count=3 retains 62920 unique words (48% of original 129846, drops 66926)
2023-04-07 10:38:30,841 : INFO : effective_min_count=3 leaves 12878632 word corpus (99% of original 12963033, drops 84401)
2023-04-07 10:38:31,039 : INFO : deleting the raw counts dictionary of 129846 items
2023-04-07 10:38:31,043 : INFO : sample=0.001 downsamples 15 most-common words
2023-04-07 10:38:31,044 : INFO : downsampling leaves estimated 12226359 word corpus (94.9% of prior 12878632)
2023-04-07 10:38:31,252 : INFO : estimated required memory for 62920 words and 300 dimensions: 1390540400 bytes
2023-04-07 10:38:31,253 : INFO : resetting layer weights
```

Figure 11: The log file of vocabulary building for train set of Stage 3's field-specific journal articles.

From the figure above, there are 62920 unique words found after eliminating less than min_count=3 frequencies in the corpus. It also indicates that 1.39 GB estimated memory required for a 300-dimensional vector space for this specific training corpus.

3.5.3.3.Model Training

The Doc2Vec model's final step is running the following line of code to start training process based on specified parameters and created vocabulary:

model.train(train_demo_corpus, total_examples=model.corpus_count, epochs=model.epochs)

The code basically starts the training of the model by using all data (total examples) in the corpus. The model runs multiple times (specified with *epoch value*) over the training data to enhance the prediction accuracy. The way how the model runs depends on the *distributed memory (dm)* parameter ("*dm=0": Paragraph Vector – Distributed Bag of Words (PV-DBOW), "dm=1": Paragraph Vector* - *Distributed Memory (PV-DM)*). In both cases, the model employs a three-layered neural network as described in the Vectorization part.

In Figure 12, a snapshot for first and the last steps in the log file for the model training process are presented. It indicates basically the used parameters, number of effective words taken into consideration during the training process and the changing values between different epochs.

```
2023-04-07 10:56:14,585 : INFO : training model with 3 workers on 62921 vocabulary and 5100 features, using sg=0 hs=0 sample=
0.001 negative=5 window=8
2023-04-07 10:56:16,213 : INFO : EPOCH 1 - PROGRESS: at 57.14% examples, 24672 words/s, in_qsize 3, out_qsize 0
2023-04-07 10:56:16,221 : INFO : worker thread finished; awaiting finish of 2 more threads
2023-04-07 10:56:16,277 : INFO : worker thread finished; awaiting finish of 1 more threads
2023-04-07 10:56:16,632 : INFO : worker thread finished; awaiting finish of 0 more threads
2023-04-07 10:56:16,633 : INFO : EPOCH - 1 : training on 12963033 raw words (70007 effective words) took 2.0s, 34280 effectiv
e words/s
2023-04-07 10:56:18,153 : INFO : EPOCH 2 - PROGRESS: at 57.14% examples, 26375 words/s, in_qsize 2, out_qsize 2
2023-04-07 10:56:18,155 : INFO : worker thread finished; awaiting finish of 2 more threads
2023-04-07 10:56:18,199 : INFO : worker thread finished; awaiting finish of 1 more threads
2023-04-07 10:56:18,579 : INFO : worker thread finished; awaiting finish of 0 more threads
2023-04-07 10:56:18,580 : INFO : EPOCH - 2 : training on 12963033 raw words (70007 effective words) took 1.9s, 36023 effectiv
e words/s
```

```
2023-04-07 10:56:47,272 : INFO : EPOCH 19 - PROGRESS: at 57.14% examples, 27728 words/s, in_qsize 3, out_qsize 0
2023-04-07 10:56:47,276 : INFO : worker thread finished; awaiting finish of 2 more threads
2023-04-07 10:56:47,278 : INFO : worker thread finished; awaiting finish of 1 more threads
2023-04-07 10:56:47,631 : INFO : worker thread finished; awaiting finish of 0 more threads
2023-04-07 10:56:47,632 : INFO : EPOCH - 19 : training on 12963033 raw words (70007 effective words) took 1.8s, 38846 effecti
ve words/s
2023-04-07 10:56:49,067 : INFO : EPOCH 20 - PROGRESS: at 57.14% examples, 27950 words/s, in_qsize 3, out_qsize 0
2023-04-07 10:56:49,071 : INFO : worker thread finished; awaiting finish of 2 more threads
2023-04-07 10:56:49,077 : INFO : worker thread finished; awaiting finish of 1 more threads
2023-04-07 10:56:49,364 : INFO : worker thread finished; awaiting finish of 0 more threads
2023-04-07 10:56:49,365 : INFO : EPOCH - 20 : training on 12963033 raw words (70007 effective words) took 1.7s, 40474 effecti
ve words/s
2023-04-07 10:56:49,365 : INFO : training on a 259260660 raw words (1400140 effective words) took 34.8s, 40258 effective word
s/s
```

Figure 12: A snapshot of the log file for Stage 3's field-specific journal articles corpus model training process

3.5.4. Similarity

This step starts with inferring vector representations for documents that will be used in similarity comparisons. The line of code used for this purpose is:

model.infer_vector.

This method creates a vector representation of a given test document based on the trained model with training corpus. An example of output for a sample sentence is given in Figure 13. The input is lemmatized text and output is vector representation consisting of 300 dimensions (the figure shows only a part of it).

```
model.infer_vector(['cognitive', 'science', 'philosophy', 'psychology', 'linguistic', 'computer', 'anthropology'])
```

```
array([ 0.0177809 , -0.08162481, -0.01101022, -0.09893596,  0.01660709,
        0.00498889,  0.13277355,  0.05022049, -0.00999397,  0.03919598,
        0.00599443, -0.02470773,  0.04776729, -0.09362443, -0.01002026,
        0.01461382,  0.1054695 ,  0.11092498,  0.07437286, -0.00853322,
       -0.01925696, -0.14720455,  0.02793587,  0.05215921, -0.14424275,
       -0.03352285,  0.09362161, -0.04495727,  0.10695815, -0.02325533,
       -0.14529814, -0.09218109, -0.08324336, -0.00357867,  0.03109528,
        0.03054421,  0.08522249, -0.11323022, -0.01103567,  0.01882504,
       -0.02323213, -0.07442391, -0.08507181,  0.00977627,  0.00636057,
        0.05218246,  0.0365843 , -0.04010918,  0.01539947,  0.04155262,
        0.04590291,  0.09339516, -0.01498471,  0.08127256,  0.06570324,
        0.03059918,  0.03765101, -0.08249575, -0.02304064, -0.01783591,
        0.03113198, -0.0314155 ,  0.03361328, -0.07440316,  0.07078947,
        0.0157912 , -0.04365136,  0.03587582, -0.131068  ,  0.00649003,
        0.04695785,  0.09266697, -0.02496654,  0.0391692 , -0.06719957,
        0.0712384 , -0.05625135,  0.02674785,  0.00332114, -0.02121778,
       -0.00525215,  0.03659535, -0.06086009,  0.00601248,  0.03427403,
       -0.04102647, -0.04779611,  0.12175003,  0.02718074,  0.11780725,
       -0.10066707,  0.06986394,  0.0249585 ,  0.08684395,  0.0219632 ,
        0.05874815, -0.11519004,  0.01911306, -0.00386848,  0.09014199,
        0.01661105, -0.02208986,  0.13169572, -0.02252489, -0.12571037,
        0.01136715,  0.14882211, -0.08997265,  0.09601379,  0.00377155,
       -0.01555311,  0.01263026,  0.06873604,  0.02497532,  0.02832546,
        0.03514043,  0.00500005,  0.08620451,  0.08152313,  0.01900124,
       -0.04554435,  0.01472464, -0.06909994, -0.1799089 ,  0.01609903,
       -0.03843017, -0.0087275 , -0.01185523, -0.00714375,  0.05753087,
        0.12691644, -0.09023814,  0.06359619,  0.02360762,  0.07677152,
       -0.06909895,  0.17109397,  0.09233737,  0.07298619,  0.00941259,
        0.01113879,  0.01711617,  0.06252424, -0.05884686, -0.02430698,
        0.07112452,  0.12300369,  0.01347483, -0.01931951, -0.06218797,
        0.01010188,  0.01855714,  0.09925374, -0.08397767,  0.0218674 ,
```

Figure 13: A snapshot of vector representation for a sample sentence

After generating all vector representations for in scope documents, the next step is calculating similarities between pairs. The measure for similarity employed in this research is cosine similarity which is described in the Semantic Similarity Measure section. The cosine similarity was calculated by the *most_similar ()* function of Gensim, which calculated the cosine of the angle between two vectors. Here is the line of code used in this specific application:

model.docvecs.most_similar([inferred_vector], topn=len(model.docvecs))

It is used to find n numbers (in this case it is equal to length of model which is 7) of most similar training documents for a given test document. It returns a list of tuples containing cosine similarity scores and document tagged numbers. Results are exported to the excel file for further analysis. The analysis based on the excel files of each stage is given in the Results section.

3.5.5. Parameter Optimization

As it is mentioned in the parameter definition part, the variations of parameters can enhance or limit the performance of models. To decide optimal values of parameters, there is a need to compute the accuracy of the model for different cases. Commonly used model performance measures are based on the basic four components given in Table 5.

Table 5: Four components of model accuracy measures.

	Positive (Prediction)	**Negative (Prediction)**
Positive (Actual)	True Positive (TP)	False Negative (FN)
Negative (Actual)	False Positive (FP)	True Negative (TN)

By definition, *True Positive (TP)* refers to the number of positive cases that are also correctly predicted as positive by the model. It is expected to be as high as possible. In our scenario, as an example, the number of philosophy articles that are found highest similar to the philosophy field corpus. *False Negative (FN)* is the number of cases that are positive but predicted as negative. In this research, there are philosophy articles computed as the highest similar to other fields. *False Positive (FP)* is the number of cases predicted as positive although they are negative. If we continue with the same example, it is the number of articles that model finds highest similar to the philosophy field but in reality, they are part of other disciplines. The last one is called *True Negative (TN)* which defines all cases that are negative and predicted as negative. In our case, it refers to non-philosophy articles predicted as a non-philosophy field.

Based on these components, there are various performance measures such as *Precision (P)* and *Recall (R)* or some derived measures like *F1 Score* are used in the applications in the literature. *Precision (P)* refers to the ratio of *TP*s over all cases that are predicted as positive (as given in (Eq. 2). It is used mostly in the cases where the cost of *FP*s is high for the system.

$$Precision\ (P) = \frac{TP}{TP+FP} \qquad \text{(Eq. 2)}$$

The other measure is called *Recall (R)* which means the ratio of *TP*s over all cases that are positive (as given in (Eq. 3). If the cost of *FN*s is high, this measure can be selected to see the performance of the model.

$$Recall\ (R) = \frac{TP}{TP+FN} \qquad \text{(Eq. 3)}$$

The *F1 Score, as defined below,* calculates the harmonic mean of *P* and *R*. This measure is crucial when a balance between *P* and *R* is expected.

$$F1\ Score = 2\ x\ \frac{P\ x\ R}{P+R} \qquad \text{(Eq. 4)}$$

In the present study, to evaluate the overall accuracy of the validation model (cf. the observed category vs. the predicted category), we focused on calculating *R* values. We adopted *R* to define the model's prediction success, as *TP*s and *FN*s are essential in our case. In other words, *R* is particularly important in contexts where missing a positive case is more critical than mislabeling a negative case as positive. Accordingly, in our analysis, *TP* represents the number of correctly classified documents for a given field, while *FN* represents the number of documents classified unsuccessfully. Thus, *R* reflects the proportion of accurately classified documents in a specific field.

We eliminated the use of *FP* and *TN*, hence the measures covering these two values such as *P* and *F1*. The main reason for this elimination was that they were affected by the number of fields we used. For instance, if we selected 10 fields instead of 6, then the values of FP and TN would have been significantly affected due to their definitions. This consideration led to our focus on measures less impacted by the number of fields, ensuring a more consistent evaluation of the model's performance.

In our case, we created a test set for each cognitive science relevant field consisting of a subset of field-specific journal articles (5400 titles and abstracts) as described in the Data Sources part. The model computes the similarities of these documents to the corpus of six fields. We assumed that if the test document is a part of philosophy and the model finds it highest similar to philosophy corpus then it means correct classification (*TP*). The details are provided in Creating the Similarity Matrix for Journal Articles part in the Results chapter.

To find the optimal hyperparameter values, we first assembled a base model (Model 1), employing the parameter values by following the implementation practice in the literature for similar-size data. We then assembled seven models (Model 2 to Model 8) by configuring the parameters to observe the effect on the *R* values. The results are presented in Table 6.

Table 6: Hyper-parameter Optimization Values and Effects on R (Recall) Value

Model	lemma	dm	vector_ size	window	min_count	epochs	dbow_words	R Value
Model 1	Yes	1	300	8	3	20	0	0.821
Model 2	Yes	0	300	8	3	20	0	0.823
Model 3	Yes	1	300	8	3	20	1	0.821
Model 4	Yes	1	300	8	3	10	0	0.818
Model 5	Yes	1	500	8	3	20	0	0.820
Model 6	Yes	1	300	8	5	20	0	0.816
Model 7	Yes	1	300	12	3	20	0	0.821
Model 8	No	1	300	8	3	20	0	0.807

*Note***: Lemma** {1,0}: Lemmatization applied or not; **dm**{1,0}: training algorithm is distributed memory or distributed bag of words; **vector size** {integer}: number of vector dimensions; **window** {integer}: distance between predicted and current word; **min_count** {integer}: minimum frequency of words that will be processed; **epochs** {integer}: iteration count for model training; **dbow_words** {1,0}: trains word vectors with document vectors or trains only document vectors.

In each step, one of the parameter values was configured to observe how it influences the average *R* score—in other words, the *TP* ratio of the model. The lowest ratio was found for Model 8, which did not include the lemmatization process. The remaining value was around 0.820. Therefore, an interim finding was that updating the parameters did not influence the model's performance, except for the lemmatization process. Therefore, we used Model 1 as the base model for initial validation. The multi-class confusion matrix for Model 1 is provided in Appendix C.

In Model 1, we selected the *PV-DM (Paragraph Vector – Distributed Memory)* method. This method predicts a center word by using the words around the center word and paragraph id (thus, *dm = 1*). We chose this method because Le & Mikolov (2014) found that the *PV-DM* performed better than its alternative (*PV-DBOW, Paragraph Vector – Distributed Bag of Words*) in the original study. We also applied the lemmatization process since it helped the model focus on the root of the words. Gensim's Lemmatization turns words into their base, removes stop words, and keeps only nouns, verbs, adjectives, and adverbs. The remaining parameters were chosen as *vector_size=300, window=8, min_count=3, epochs=20, dbow_words=0*.

Doc2Vec, cosine similarity calculations, and lemmatization processes were implemented in the GenSim library (Rehurek & Sojka, 2010). We used Python 3 and Jupyter Notebook with the Anaconda Distribution as the coding platform. It took about 96 hours to lemmatize the train set and took 31 seconds to train the model with Intel Core i7 2.00 GHz CPU and 4 GB Ram.

A caveat in Model 1 was that its input data (i.e., the journal articles) required a specific set of parameters that differed from the conference proceedings since their document structure and length differed. Therefore, we tuned the parameters for the best model fit to reduce the structural differences between document types. This process resulted in a *window size* of 12, a *min_count* value of 5, and a smaller number of *epochs*, with a value of 10.

CHAPTER 4

RESULTS

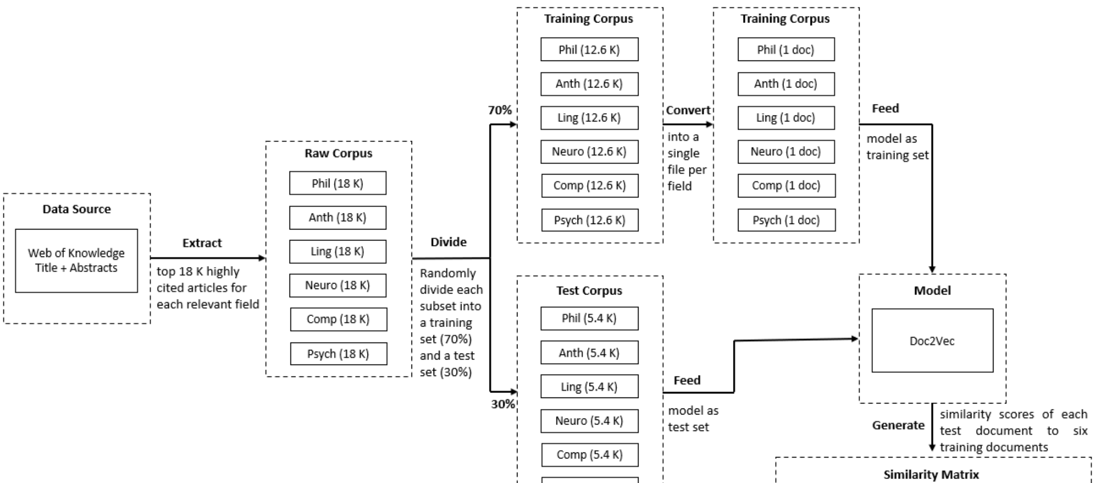
Building upon methodology designed and detailed earlier in this study, this chapter delves into the empirical findings. These findings constitute a systematic response to each of the research questions posed in Chapter 1. Specifically, we present similarity matrices to address whether semantic text similarity analysis provides consistent similarity scores for determining the relevance of academic documents to their respective research fields. This approach demonstrates the efficacy of semantic text similarity analysis in accurately categorizing academic documents into their relevant fields. Consequently, we create similarity matrices to quantify cross-disciplinarity of cognitive science.

The details of stages are provided below. In the initial stage, we evaluated journal articles from cognitive science relevant subfields to validate the model's classification accuracy. We first investigated the text-similarity analysis to determine if it provided insights into the relationship between well-defined academic documents and their respective research fields. At this point, a manual annotation study was conducted to compare the model's findings with classifications made by a domain expert. As part of these initial investigations, we also run the model for selected classical articles. In the second stage, the model was trained with field-specific conference proceedings, and random sets to assess its performance on unstandardized academic documents. Finally, the model was applied to the proceedings of Cognitive Science Society (CSS) conferences to investigate what text-similarity analysis might reveal about the field's cross-disciplinarity. This final stage of investigation was conducted at both the individual article level and the consolidated document level, including all subsequent proceeding articles published within a given year. This step-by-step approach was adopted to understand the extent to which methods like Doc2Vec and cosine similarity can explain the cross-disciplinary characteristics of cognitive science.

4.1. Creating the Similarity Matrix for Journal Articles

The primary objective of this stage is to ensure the performance of model based on its ability to reliably predict the relationship between articles and their corresponding fields with chosen parameters for journal articles in fields related to cognitive science. By presenting that the model, with its selected parameters, can accurately classify the articles into their respective fields, we would have confidence in subsequent analyses and their insights.

In the first stage, as outlined in Table 1 of the methodology section, we selected the top 18,000 highly cited articles from the Web of Science (WoS) database for each relevant field within cognitive science. These fields include philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. After extracting the titles and abstracts of these 18,000 articles for each subfield, we conducted an initial study aimed at evaluating the performance of the Doc2vec model and cosine similarity on journal articles. The expectation was for the model to accurately classify the journal articles into their respective subfields, assigning them high similarity scores.. High similarity scores in this validation study would be indicative of the model's reliability.



Doc No.	Test Category	Phil	Anth	Ling	Neuro	Comp	Psych
1	Phil	0.900	0.289	-0.140	-0.081	-0.019	0.199
2	Anth	0.324	0.941	0.013	-0.159	-0.107	-0.005
3	Ling	0.069	0.234	0.800	-0.259	-0.165	-0.192
4	Neuro	-0.199	0.005	-0.063	0.914	0.109	0.005
5	Comp	0.102	0.071	0.184	0.144	0.878	0.222
6	Psych	-0.140	-0.358	-0.262	0.194	0.084	0.833

Figure 14: The flow of creating the similarity matrix for journal articles.

For the validation study, as described in Figure 14, we randomly divided each subset into a training set (70% of the 18,000 documents, equaling 12,600 documents) and a test set (the remaining 30%, equaling 5,400 documents). This separation ensured that both datasets were a realistic representative of the original data. Subsequently, we merged the items within the training set into a single file per field for each subset. In simpler terms, we created a unified single file for the training set, which consisted of

12,600 documents for each subfield. These consolidated training sets of each subfield were then used as input data for training the model. The remaining 5,400 documents for each subfield constituted the test set. The output of the model was a similarity matrix, which showed the degree of similarity between each article and the combined documents of the respective research fields. Table 2 presents six random articles, as a partial snapshot of the extensive similarity matrix, to demonstrate the rationale behind the similarity matrix by offering a sample view.

Table 7: A snapshot of the similarity matrix for randomly selected six articles.

Doc No.	Test Category	Phil	Anth	Ling	Neuro	Comp	Psych
1	Phil	0.900	0.289	-0.140	-0.081	-0.019	0.199
2	Anth	0.324	0.941	0.013	-0.159	-0.107	-0.005
3	Ling	0.069	0.234	0.800	-0.259	-0.165	-0.192
4	Neuro	-0.199	0.005	-0.063	0.914	0.109	0.005
5	Comp	0.102	0.071	0.184	0.144	0.878	0.222
6	Psych	-0.140	-0.358	-0.262	0.194	0.084	0.833

Note: Negative numbers indicate dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology.

In Table 7, the first test document (row 1) is an article originally classified in a single research field, belonging to Phil (philosophy) according to the Web of Science (WoS) database. The Doc2vec model found a relatively high similarity score (0.900) between this article and unified document of philosophy subfield (consisting of the remaining 12,600 articles in the philosophy corpus). This result aligns with our expectation that the model would successfully identify the relevant subfield of the article. Similarly, test document number 2 showed a notable similarity score of 0.941 to its classification field (in this case, anthropology). A similar pattern was observed in the linguistics, computer science, neuroscience, and psychology documents which further supported the model's performance in classifying the articles into their respective subfields for this limited sample articles.

To uncover the comprehensive picture, we further analyzed the data by calculating the ratio of the true positives (TPs), denoted as the Recall (R) value. TPs were identified by counting the number of test articles that exhibited the highest similarity score with their original subfields. To calculate the ratio, we basically divided this number by the total number of test articles for each respective subfield. In other words, we determined the subfield with the highest similarity score for each test document and compared it with the document's original subfield categorization. We computed the ratio by counting the articles matched their original subfields in this manner. As provided in Table 8, we found that 26,592 test documents (among a total of 32,400) were categorized as TPs by exhibiting the highest similarity score with their predicted fields. This resulted in an accuracy of 0.821, an indicator in favor of the model's reliability.

When examining the performance for each subfield, the highest prediction success was observed in the computer science test set, with a remarkable R value of 0.955. This means that 5,148 randomly selected articles from the computer science corpus were computed highest similar to computer science training corpus. Conversely, the lowest prediction performance was observed in the psychology documents, with an R value of 0.674. This finding suggested that the model could have difficulty in classifying articles within psychology. We also found that the average similarity score of articles to their respective classified fields was around 0.780. This score indicated that, on average, the articles and their assigned subfields were found quite close. Overall, these analyses showed that the articles in our training dataset were consistent, showing that the model could effectively differentiate between their respective similarities to specific research fields.

Table 8: TPs and R (Recall) values based on highest similarity criteria for each test set indicating the model's success.

	Comp	Neuro	Phil	Ling	Anth	Psych	Overall
Total Documents	5400	5400	5400	5400	5400	5400	32,400
TPs	5158	4735	4574	4390	4097	3638	26,592
Total R Value	0.955	0.877	0.847	0.813	0.759	0.674	0.821

Note: Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology.

Despite the overall R value of 0.821 indicates the model's success, the previous analysis also highlighted that the model could not classify the remaining part of the articles (totally 5,808) into their expected subfields based on highest similarity scores. To gain a more detailed understanding, we further computed the similarity scores for each article to the next similar field of research, in addition to the first similar field of research. In other words, we expanded the criteria for detecting the TPs to include articles that matched either the highest or the second-highest similarity. As a result of this expanded criterion, the accuracy increased to 0.936 (30,320 test documents out of 32,400 were categorized as TPs), which shows a highly significant prediction performance for the model.

Table 9: R (Recall) values of the model based on highest similarity and 2^{nd} highest similarity cases.

	Comp	Neuro	Phil	Ling	Anth	Psych	Average
R value for the highest similarity	0.955	0.877	0.847	0.813	0.759	0.674	0.821
R value for the 2^{nd} highest similarity	0.028	0.088	0.114	0.123	0.163	0.175	0.115
Total R value	0.983	0.965	0.961	0.936	0.922	0.849	0.936

Note: Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology.

Table 9 shows the R values to represent the similarity of the articles to both their own field of research and to the next relevant field. The values in Table 9 indicate that the model's success rate was relatively high for computer science, neuroscience, philosophy, and linguistics when TPs were defined by the highest similarity scores. These fields demonstrated notable R values, as given in the table. However, R values for psychology and anthropology articles were comparatively lower, indicating relatively minor similarities to their respective classified fields of research. Nevertheless, when we computed R values considering the second-highest similarities, the model's prediction accuracy significantly increased for these subfields as well. It means that the model assigned these articles into expected categories in the second order. Despite these variances, combined R values (the total of the first and the second rows) across all fields revealed very high R values, indicating model's accurate classification capabilities.

We further investigated the closest neighbor fields for each field of research as the results summarized in Table 10.

Table 10: Closest neighbor fields based on second highest similarity criteria.

Test Category	Phil	Anth	Ling	Neuro	Comp	Psych
Phil		0.100	0.023	0.003	0.014	0.013
Anth	0.109		0.021	0.044	0.022	0.045
Ling	0.084	0.048		0.002	0.042	0.011
Neuro	0.002	0.011	0.009		0.022	0.079
Comp	0.015	0.008	0.015	0.002		0.005
Psych	0.071	0.043	0.057	0.084	0.071	

Note: Negative numbers indicate dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology.

The analyses revealed that a considerable proportion of anthropology articles (0.109) scored high in similarity to the philosophy field, and vice versa (0.100). A similar relationship was observed between linguistics and philosophy documents (0.084). Psychology documents were assigned to neuroscience (0.084), to computer science (0.071), and to philosophy 0.071). From the model-development perspective, these overlaps may be considered errors (noise) in the dataset. Nevertheless, those overlaps are indispensable, given the partial interactions among the subfields in the dataset[5]. In

[5] For example, it is likely that an article published in an anthropology journal (as categorized by the WoS database) may be more related to the philosophy subfield according to the similarity analysis. It is likely that further development of the methods of categorization in the WoS database would minimize those issues.

summary, we validated the Doc2vec model and cosine similarity for journal articles in subfields related to cognitive science. The results showed that the model with the chosen parameters and datasets reliably predicted the relationship between test articles and their fields.

4.1.1. Manual Annotation Validation Analysis

For further validation of the results and to enhance the credibility of the findings, we manually annotated a randomly selected subset of the dataset. This subset included the titles and abstracts of 108 articles from each subfield, resulting in a total of 648 articles. These articles were selected from a list where their test category and model's prediction were found to be the same in previous study. As an example, in the case of computer science, out of 5,400 articles, 5,158 were successfully predicted by the model as computer science articles, as indicated in Table 8. From this pool of 5,158, we randomly selected 108 articles, which were originally classified as computer science articles and also identified by the model as highly similar to the same field. The number of articles was determined to be optimal, considering the need for specific evaluation of each article by a domain expert.

Table 11: The results of manual annotation study for model validation.

			Expert Annotation Results (Number of articles assigned to fields)					
Test Category	Number of Articles	Model's Mean Similarity	Phil	Anth	Ling	Neuro	Comp	Psych
Phil	108	0.916	103	3	1			
Anth	108	0.894	1	106			1	
Ling	108	0.876			105		1	2
Neuro	108	0.898				107		1
Comp	108	0.887					108	
Psych	108	0.918		1		8		99

Note: Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology.

The manual annotation process was carried out by an expert with a doctoral degree from a cognitive science program, who was also unaware of the purpose of the study. The expert labeled 648 articles by assigning them to one of six contributing fields. We then compared the assigned labels with the model findings for the same subset. The results of this comparative analysis revealed a significant alignment between the model's predictions and the expert's assignments with a substantial match rate of 96.9% (628 out of the 648 articles). The details of how model's predictions and expert annotations are given in Table 11. As an example, among the 108 philosophy articles identified by the model with an average similarity score of 0.916, the expert manually assigned 103 as philosophy articles, three as anthropology, and one as linguistics. The highest compatible result was observed in computer science, where all the 108 articles were mapped to computer science by the expert. Conversely, the lowest matching ratio was in psychology, with 8 articles being assigned to neuroscience by the expert. Overall, this comprehensive comparative analysis of manual annotation findings and model's predictions supported the validity of model's performance.

4.1.2. Analysis of Cognitive Science's Classical Readings

The following part of the paper moves on to an application of mentioned model for specific selected articles referred to classicals of cognitive science. The purpose of this study is to evaluate how well the model performs in capturing the similarity between classical readings of cognitive science across cognitive science related subfields. Classical readings of cognitive science field were used as a test set to gain an understanding of text similarity approach. The term classics is highly subjective and might be individualized based on approaches and methodologies adopted by the researcher. However, there are some common articles which can be defined as classics and offered to be analyzed in cognitive science courses. The training set consisted of a complete set of field-specific journal articles (%100) defined in the Data Sources section. As a test set, randomly selected 20 cognitive science classical reading articles full texts were used after a text cleaning. As a result, we created a cosine similarity value matrix for these 20 articles and 6 cognitive science relevant fields. All results are given in Table 21 in Appendix D.

Although the overall picture shows that most articles were found to be significantly similar to more than one contributing field, we will analyze a number of articles in detail. A primary objective of this study is to discuss the potential of text similarity measures for evaluating the cross-disciplinarity of documents or fields. In line with this goal, the findings from this part suggest that there are possible signs of cross-disciplinarity.

Table 12: Similarity scores of selected classical articles to cognitive science relevant fields

Article	Phil	Anth	Ling	Neuro	Comp	Psych
Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language, 38(4), 419–439	0.002	-0.305	0.491	0.373	0.446	0.477
Christiansen, M.H., & Chater, N. (2001). Connectionist Psycholinguistics: Capturing the Empirical Data. Trends in Cognitive Sciences, 5(2), 82-88	0.023	-0.015	0.763	-0.005	0.522	0.353

A notable example illustrating this issue, as presented in Table 12, is the paper titled "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models" (1998). The content of this article integrates ideas and techniques from psychology, neuroscience, linguistics, and computer science. The mental representations involved in word recognition required contribution of psychology. Neural mechanisms underlying language processing can be captured with neuroscience. Language structure and meaning is examined using ideas and frameworks from linguistics. Additionally, the utilization of eye tracking, computational models and methods for data analysis require expertise in computer science. According to our model, this classical article is semantically similar to the fields of psychology, neuroscience, linguistics, and computer science. This finding is parallel to what the article's content mainly consisted of. The second article in Table 12 is apparently a linguistics and psychology work. However, due to the empirical nature of studies, it also demonstrates relevance to computer science. The other selected articles given in Table 21 in Appendix A also indicate that at least two fields exist together for each document. It is therefore likely that the selected documents in cognitive science field carry a cross-disciplinary characteristic. This study is important in two dimensions. First, the results are promising in that the model's prediction results are compatible with the contents of articles. Secondly, these classical papers provide preliminary intuition regarding cognitive science's cross-disciplinarity.

In this part of our study, we have validated the performance of the model using structured academic documents. The next stage involves utilizing field-specific conference proceedings as test documents. This will allow us to evaluate the model's performance on unstandardized publications, as presented in the following section.

4.2. Running the Doc2Vec Model for Field-Specific Proceedings

In the subsequent phase of our research, we aimed to assess the model's prediction capability with unstandardized academic documents, more specifically conference proceedings. This phase served as a preliminary step prior to running the model for proceedings of Annual Meetings of the Cognitive Science Society (CSS). In other words, if the model successfully assigns the given proceedings to their original fields, it would indicate its potential effectiveness in evaluating CSS proceedings as well.

To accomplish this, we extracted the proceedings of seven conferences for philosophy, computer science, psychology, and a non-relevant proceeding in the energy field as all details presented in the Field-specific conference proceedings of Data Sources section.

In Table 13, the rows list the names of the conferences used in the model's test set. The accompanying numbers represent the similarity scores, indicating the degree of resemblance to the training set presented in the previous section. In addition to the models for the six subfields mentioned above (Phil, Anth, Ling, Neuro, Comp, Psych), we developed an additional subfield model, denoted as Random, using a set of randomly selected articles from the WoS database. The training set for the Random model consisted of 9,000 documents (titles and abstracts) from diverse research fields (astronomy, business, economy, energy, mechanics, and public administration). To ensure the diversity of the dataset, we included the Euradwaste 2013 as a conference largely irrelevant to cognitive science since it was a conference about radioactive waste, and the model results showed low similarity to the subfields. Notably, the philosophy-associated conference proceedings (Phil IAFOR 2016, Phil APA 2016, and Phil APA 2012) exhibited high similarity scores to the philosophy subfield. The second highest similarity was attributed to anthropology, consistent with our previous findings (as discussed in the further investigation part of Table 10). Similarly, the two proceedings in computer science (Comp CVPR 2016 and Comp FEDCSIS 2016) exhibited high similarity scores to the computer science field. Interestingly, these also showed high similarity (more than 0.300) to the random datasets. One reason might be the use of common computer science terminology in many fields. Finally, the selected psychology proceedings (Psych InPACT 2015 and Psych IAFOR 2016) exhibited high similarity to the psychology subfield.

Table 13: Field-specific proceedings: similarity scores for relevant cognitive science fields and a random set.

Conference	Phil	Anth	Ling	Neuro	Comp	Psych	Random
EURADWASTE_2013	0.249	0.327	0.019	0.059	0.212	-0.005	0.140
Philosophy_IAFOR_2016	0.560	0.387	0.251	-0.046	0.047	0.215	-0.100
Philosophy_APA_2016	0.826	0.390	0.108	-0.140	0.057	0.082	0.006
Philosophy_APA_2012	0.724	0.182	0.105	-0.084	0.083	0.151	0.102
Computer_CVPR_2016	-0.086	-0.119	0.084	0.040	0.760	0.144	0.376
Computer_FEDCSIS_2016	0.059	0.035	0.011	0.019	0.623	0.009	0.307
Psychology_InPACT_2015	0.284	0.253	0.090	-0.075	-0.065	0.667	-0.080
Psychology_IAFOR_2016	0.079	0.063	-0.070	0.071	0.025	0.713	-0.052

Note: The negative numbers indicates dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology; Random: A random set of articles.

In summary, our analysis has revealed that the proceedings, which are unstandardized academic documents, can be reliably predicted by the model, which demonstrates their relationship to the relevant subfield. We will evaluate the CSS proceedings' cross-disciplinary characteristics in the following section using the same training set developed for field-specific conferences, aiming to shed light on their cross-disciplinary nature.

4.3. Running the Model for the Cognitive Science Society (CSS) Proceedings

Up to this point, we have evaluated how the model performs with both structured academic documents and unstandardized publications such as proceedings. The results have demonstrated the model's success in creating similarity matrices. Building upon this foundation, the next and the final stage of our study will involve employing the same approach to the CSS proceedings, which are typically regarded as timely and representative collections of documents within a specific field, to analyze the cross-disciplinary nature of cognitive science.

In accordance with the methodology outlined in the Cognitive Science Society Meeting Proceedings in methodology section, we crawled data from the proceedings of the Annual Meetings of Cognitive Science Society conferences to create test sets (Cognitive Science Society Past Conferences, 2019; CogSci Proceedings, 2020; Hope, 2011). These CSS meeting proceedings comprised full texts of all the Annual Meetings of Cognitive Science Society conferences published between 1981 and 2022. For our analysis, we utilized the same training set as in the previous section. This set comprised a complete set of field-specific journal articles (18,000 titles and abstracts for each subfield), as well as a random set (total of 9,000 titles and abstracts from irrelevant fields).

Table 14 illustrates the similarity scores generated by the model for the proceedings of each year across various subfields, including a random set. In this table, a darker cell color represents a higher degree of similarity. The “Year” column denotes the test sets documents (proceedings) of the respective year of the conference.

Table 14: Similarity scores of CSS Proceedings to the fields related to cognitive science and a random dataset.

Year	Phil	Anth	Ling	Neuro	Comp	Psych	Random
1981	0.319	0.124	0.190	-0.051	0.273	0.119	0.128
1982	0.539	0.167	0.544	-0.120	0.318	0.051	0.110
1983	0.112	0.022	0.606	0.068	0.415	0.270	0.299
1984	0.248	0.010	0.406	-0.130	0.315	0.124	0.068
1985	0.050	-0.162	0.243	0.439	0.530	0.295	0.402
1986	-0.021	-0.019	0.215	0.128	0.406	0.259	0.158
1987	0.195	-0.048	0.299	-0.019	0.439	0.025	0.192
1988	0.100	0.007	0.366	-0.015	0.506	0.116	0.150
1989	0.107	-0.073	0.420	-0.086	0.500	0.131	0.238
1990	0.219	-0.047	0.384	-0.031	0.470	0.246	0.092
1991	0.305	0.009	0.398	-0.083	0.226	0.149	0.120
1992	0.370	0.157	0.213	0.000	0.319	0.119	0.108
1993	0.450	0.113	0.336	-0.141	0.290	0.187	0.053
1994	0.586	0.015	0.382	-0.160	0.257	0.167	0.179
1995	0.010	-0.132	0.606	0.011	0.173	0.257	0.110

Table14 cont.

1996	0.110	0.011	0.341	0.079	0.285	0.099	0.133
1997	0.383	-0.028	0.481	-0.060	0.329	0.181	0.149
1998	0.381	0.052	0.287	-0.041	0.057	0.626	0.154
1999	0.235	0.138	0.290	-0.008	0.213	0.284	0.097
2000	0.371	-0.060	0.293	0.013	0.268	0.456	0.124
2001	0.057	0.019	0.410	0.128	0.461	0.227	0.159
2002	0.272	0.054	0.319	0.012	0.434	0.275	0.049
2003	0.045	-0.220	0.426	-0.006	0.480	0.367	0.280
2004	0.358	-0.008	0.300	-0.005	0.110	0.370	0.011
2005	0.221	-0.179	0.468	0.063	0.113	0.144	0.228
2006	-0.184	0.159	0.484	0.076	0.123	0.248	0.048
2007	0.140	-0.062	0.558	-0.018	0.353	0.395	0.184
2008	0.075	-0.117	0.294	0.078	0.137	0.250	0.067
2009	0.148	-0.139	0.746	0.056	0.331	0.247	0.178
2010	0.165	0.062	0.084	0.148	0.004	0.216	0.069
2011	0.248	0.080	0.262	0.009	0.259	0.215	0.016
2012	0.290	0.024	0.185	-0.081	0.241	0.231	0.052
2013	0.305	0.058	0.288	-0.017	0.367	0.190	-0.031
2014	0.298	0.050	0.254	-0.013	0.151	0.220	0.069
2015	0.247	0.096	0.277	-0.075	0.209	0.304	0.046
2016	0.329	0.176	0.474	-0.057	0.058	0.225	-0.025
2017	0.283	0.179	0.180	-0.040	0.192	0.196	-0.012
2018	0.247	-0.032	0.327	-0.069	0.226	0.339	0.064
2019	0.214	0.066	0.173	0.049	0.274	0.325	0.048
2020	0.235	0.031	0.154	-0.061	0.294	0.323	0.139
2021	0.435	0.079	0.354	-0.061	0.206	0.279	0.133
2022	0.063	0.024	0.598	-0.118	0.033	0.178	0.277

Note: The negative numbers mean dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology; Random: A random set of articles.

The results shed light on both cognitive science's relationship with relevant subfields and similarity trends within the field of study over time. Upon closer examination of Table 14 reveals that the similarity of the proceedings to the random dataset varies from -0.031 to 0.402 (M = 0.122, SD = 0.090), providing further support for the reliability of the model. The table also shows a balanced distribution effect among the four constituent fields. Based on the overall mean similarity scores, the order of the constituent fields are as follows: linguistics (M = 0.355, SD= 0.141), computer science (M = 0.277, SD= 0.136), psychology (M = 0.236 SD= 0.109), philosophy (M = 0.228, SD= 0.152), anthropology (M = 0.016, SD= 0.098) and neuroscience (M = -0.005, SD= 0.101). Over the years, linguistics demonstrates a relatively high level of similarity to CSS proceedings, with similarity scores ranging from 0.084 to 0.746. Computer Science maintains high positive similarity scores, with values ranging from 0.110 to 0.626. Especially during the1985-1990s and beginning of 2000s the outcomes present highest values of similarity scores. Psychology also exhibits notable similarity, particularly in the years 1998 and 2000, where the similarity scores reach a value of 0.627. Philosophy reached its maximum in 1994 with a value of 0.586. In the remaining years, the similarity is changed between this maximum value and minimum value -0.184. On the other hand, anthropology presents a mixture of positive and negative similarity scores, which means it is indicating both similarity and dissimilarity to CSS proceedings. Neuroscience generally demonstrates lower similarity scores when compared to other subfields, which can be an indicator of relatively lower degree of similarity. An unexpected result is the higher similarity of CSS proceedings to the random set when compared to anthropology and neuroscience.

Figure 15 presents a time course of the contributions made by cognitive science's constituent fields, providing an overview of similarity score trends for each related field between 1981 and 2022. Although it is difficult to find a regular contribution of a specific field of research into cognitive science, philosophy, computer science, linguistics, and psychology have all significantly contributed to the field, to varying degrees. On the other hand, contributions by anthropology and neuroscience have been limited so far, as much as the CSS proceedings show. When examining the longitudinal trends using moving average lines, it is important to note that some subfields like philosophy and linguistics exhibit fluctuating patterns in their similarity scores over time, while others such as psychology or neuroscience demonstrate relatively stable levels of similarity. In Figure 15, we also added linear trendline for each subfield to see the trend of similarities. The results showed that philosophy, anthropology, and neuroscience are following stable patterns over the years. On the other hand, linguistics and computer science display a declining trend while psychology appears to demonstrate an increasing trend in similarity scores.

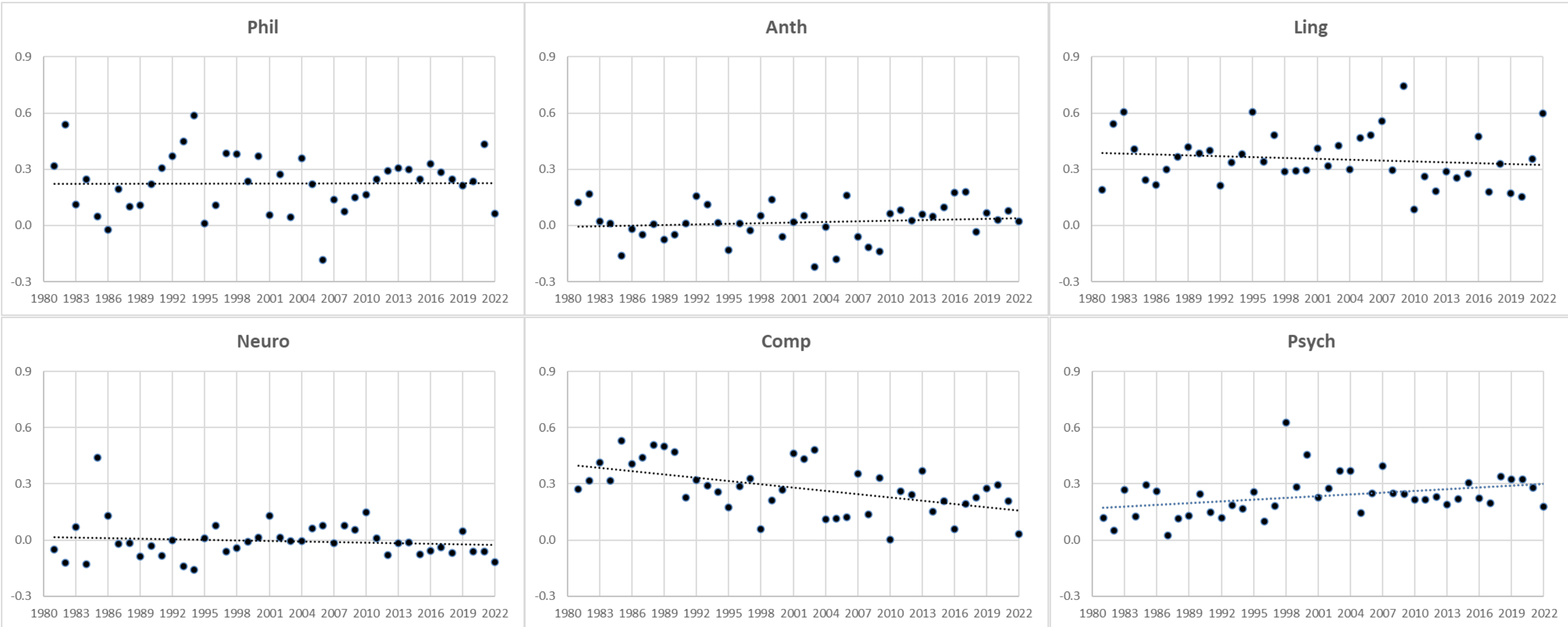


Figure 15: Similarity scores of related fields to CSS proceedings.

Finally, we conducted a case study for comparison purposes. The test included articles from a set of cognitive science journals (Cognitive Science, Topics in Cognitive Science, Trends in Cognitive Science, and WIREs Cognitive Science) from 1991-2020. The results are presented in Table 15.

Table 15: Average similarity scores for selected cognitive science journals.

	Phil	Anth	Ling	Neuro	Comp	Psych	Random
The present study (CSS proceedings)	0.228	0.016	0.355	-0.005	0.277	0.236	0.122
The case study (Journals)	0.243	0.070	0.468	-0.007	0.281	0.403	-0.049

Note: The negative numbers mean dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology; Random: A random set of articles.

The findings reveal differences between CSS proceedings and journal articles, besides the general trends among contributions from the leading constituent fields. Although the case study provides a relevant perspective for interpreting the results, the test set is limited in scope, especially in its volume (i.e., the number of articles) and non-homogeneous document types.

4.3.1. Running the model for Individual CSS Proceeding Articles

A technical challenge in designing similarity models, and more generally in NLP models, is specifying the appropriate size for the model input unit (e.g., Balikcioglu et al., 2022). Models reported in the previous sections employed document similarity scores, such that each field was represented by a single, large document that involved the proceedings of the Annual Meetings of Cognitive Science Society. Therefore, the similarity scores were computed by processing documents (as input units) for each subfield (accordingly, each cell in Table 14 shows a similarity score for one large, combined document, including all the proceeding articles for a given field, each year). Consequently, the model returned similarity scores for all the articles published in a specific year's proceedings for each field. A disadvantage of this approach, characterized as a mean similarity approach, is that it potentially overlooks distinctions and differences between specific documents published within the same proceeding. In cases where the results were predominantly monodisciplinary, there might be minimal ambiguity; the grand mean and the mean of means produce the same result. However, this limitation was particularly relevant when the results consisted of a heterogeneous mixture of subdisciplines, leading to fundamental ambiguity. Therefore, to address this limitation, we conducted a comprehensive analysis at the individual article level to investigate the contributions of subfields. By evaluating each article independently, we attempted to capture the differences and variations that might be missed by mean similarity approach.

In order to compute article-specific similarities, a new test set was developed, consisting of individual articles from CSS proceedings between the years 2010 and 2022[6]. In this period, there were 10,552 CSS proceeding articles published in total, with the number of articles per year varying between 679 and 990. The average number of articles used as a test set was 812. In the article-specific training dataset, we made a further adjustment to existing training set used in previous analysis by excluding the clinical neurology category and designing the set mainly based on neuroimaging and neuroscience categories. This refinement aimed to enhance the representativeness of neuroscience data. The resulting article-specific training dataset consisted of a matrix with 10,552 rows representing single articles and 7 columns denoting respective subfields as well a random set. Each value in the matrix represented the similarity score between a given article and a specific subfield or random set. Table 18 shows first six articles from the year 2010, as a partial snapshot of the extensive similarity matrix.

Table 16: A snapshot of the similarity matrix for randomly selected individual CSS proceeding articles.

Doc No.	Year	Phil	Anth	Ling	Neuro	Comp	Psych	Random
1	2010	0.703	0.262	0.402	0.020	0.245	0.325	0.031
2	2010	0.827	0.499	0.260	-0.037	0.240	0.461	-0.007
3	2010	0.803	0.280	0.588	0.056	0.289	0.461	0.005
4	2010	0.442	0.041	0.418	0.467	0.501	0.802	-0.063
5	2010	0.462	0.192	-0.029	0.588	0.110	0.397	0.038
6	2010	0.424	0.031	0.156	0.063	0.888	0.201	0.287

*Note***:** Negative numbers indicate dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology. Random: A random set of articles.

The present study proceeds with a high-level analysis, aiming to present average similarity scores for each year and make comparisons with outcomes of previous analysis. The

[6] The reason for the limited (2010-2022) data set was the public accessibility of the CSS proceedings. After 2010, the proceedings were stored in a structural way, which allowed us to crawl individual articles rapidly.

results of this high-level analysis are initially summarized in Table 17, where the average similarity scores of all articles published in their respective years are presented. There are noticeable differences in the similarity scores for corresponding years when comparing these findings to those in Table 14. The main reason can be attributed to the disparity between the grand mean and the mean of means. This divergence can be a result of heterogeneity of the cognitive science field. As an example, the similarity score between combined document representing the articles published in the year 2010 and philosophy subfield was found 0.165 in Table 14. However, in Table 17, where the mean value of all articles published in 2010, the similarity value was calculated as 0.315.

Table 17: Average similarity scores of the model for Individual CSS Proceeding Articles (2010-2022).

Year	Phil	Anth	Ling	Neuro	Comp	Psych	Random
2010	0.315	-0.035	0.490	0.142	0.339	0.407	0.075
2011	0.307	-0.035	0.467	0.150	0.338	0.395	0.081
2012	0.300	-0.018	0.482	0.123	0.323	0.366	0.074
2013	0.315	-0.019	0.466	0.155	0.321	0.406	0.079
2014	0.308	-0.037	0.445	0.151	0.338	0.383	0.107
2015	0.306	-0.025	0.462	0.138	0.308	0.395	0.091
2016	0.289	-0.031	0.464	0.147	0.322	0.415	0.088
2017	0.302	-0.032	0.434	0.155	0.328	0.405	0.090
2018	0.296	-0.002	0.442	0.131	0.305	0.412	0.079
2019	0.324	0.032	0.433	0.120	0.299	0.391	0.069
2020	0.324	0.052	0.453	0.130	0.296	0.439	0.072
2021	0.314	0.062	0.438	0.116	0.291	0.391	0.069
2022	0.339	0.054	0.446	0.102	0.280	0.427	0.080
Overall	0.311	-0.001	0.455	0.135	0.314	0.403	0.081

Note: The negative numbers mean dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology; Random: A random set of articles.

Although the average similarity scores were found differently when compared to previous analysis, the overall picture of the contributing fields remained the same. Linguistics exhibited the highest average similarity scores, ranging from 0.433 to 0.490 over the years with an overall mean similarity value 0.455, indicating a relatively highest level of similarity. Psychology followed closely with an overall mean score of 0.403, suggesting a relatively high level of similarity between cognitive science articles and psychology. The average similarity scores for computer science varied from 0.280 to 0.338 with an overall average similarity score of 0.314, indicating a moderate level of similarity. Similarly, philosophy's average similarity score ranges from 0.289 to 0.339 with 0.311 overall mean similarity. Neuroscience's similarity scores are found higher (overall mean similarity of 0.135) when compared to previous analyses. Anthropology is found to have the lowest level of similarity ranging between -0.037 and 0.062, which is even lower than random set.

Table 18: The number of articles having the highest similarity to the corresponding subfield (darker colors show higher similarity scores).

Year	Phil	Anth	Ling	Neuro	Comp	Psych	Random
2010	95	3	262	19	125	165	10
2011	128	9	293	29	162	193	21
2012	115	9	294	23	132	159	21
2013	159	12	348	26	177	241	27
2014	116	10	257	25	156	190	27
2015	115	9	248	17	111	153	24
2016	88	7	257	24	130	189	23
2017	135	11	299	36	160	234	35
2018	91	15	251	22	118	193	25
2019	162	24	304	24	156	206	35
2020	124	28	285	21	126	267	23
2021	131	48	278	21	126	204	35
2022	171	26	291	18	116	219	25
TOTAL	1630	211	3667	305	1795	2613	331

*Note***:** The negative numbers mean dissimilarity between the two fields. Phil: Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology; Random: A random set of articles.

We further analyzed field-specific contributions by computing the number of articles with the highest similarity score to the corresponding field (Table 18). Each row represents a specific year of CSS proceeding and the total row provides the cumulative counts for each subfield for the years 2010 to 2022. The columns represent the subfields and a random set. For example, in 2010 CSS proceeding, there were 262 articles scoring a high similarity to linguistics, 165 articles to psychology, 125 to computer science and so on. Linguistics had the highest number of articles (totally 3667 articles) with the highest similarity scores. Psychology followed it with a total of 2613 articles over the entire period. Computer science showed a notable existence as well, with a ranging scale between 111 and 177 articles and totaling 1,795 articles. Total number of articles with highest similarity scores to philosophy was found as 1630. Although it was found relatively lower in counts Neuroscience still had a presence, with a total of 305 articles over the entire period. Anthropology had the lowest counts (211 articles) among the subfields. The random set which consisted of irrelevant subfields, had counts ranging from 10 to 35 with a total of 331 articles. In conclusion, more CSS articles exhibited similarities to linguistics, psychology, computer science, and philosophy than they did similarities to anthropology and neuroscience. The overall contribution of the fields remained similar to the findings obtained in the initial base model (Table 14).

We also analyzed the weight of the subfields by calculating the scores for individual articles' similarities to each subfield, which can be considered an indicator of each article's level of cross-disciplinarity. As an operational assumption, we set a minimum threshold value of 0.299 to specify the significant contribution of a subfield to an article. The threshold value was chosen from the similarity scores of the random field dataset, after eliminating an outlier value (0.402, as revealed by an interquartile range analysis, Table 14). Our findings revealed that the most frequent pattern observed among the CSS proceedings was three contributing subfields, which was identified in 3945 CSS proceedings. There were 3141 articles that exhibited significant similarity two contributing subfields. On the other hand, there were 53 articles that did not exhibit significant similarity to any of the subfields. The overall result showed that 9588 of the 10522 analyzed articles involved two or more contributing fields, highlighting the cross-disciplinary characteristic of coginitive science field. A comprehensive overview of the distribution pattern over the number of subfields are presented in Figure 16.

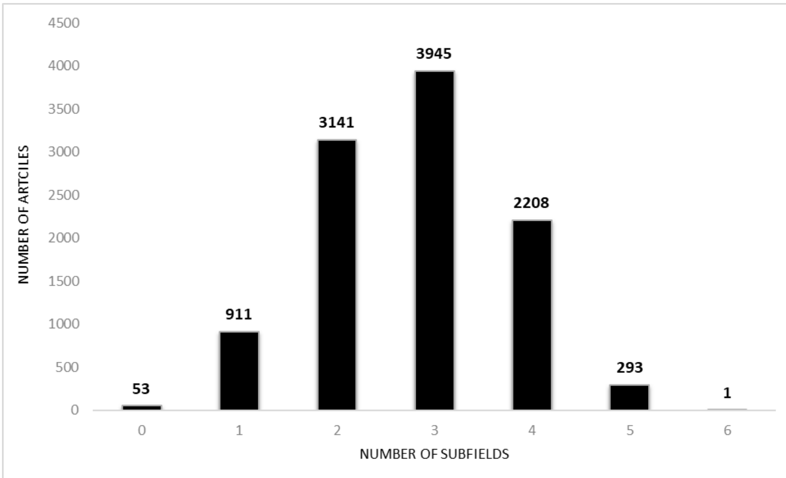


Figure 16: The distribution of cognitive science proceeding articles over number of contributing subfields

Figure 17 presents a time course of number of articles having the significant similarity to the corresponding subfields between 2010 and 2022. It highlighted the noteworthy roles played by philosophy, computer science, linguistics, and psychology. They were found as significantly contributed to cognitive science to varying degrees. The linear trendline for each subfield offered insight that linguistics, computer science, and neuroscience were following stable patterns over the years. On the other hand, philosophy, anthropology, and psychology exhibited an increasing trend in the number of significantly similar articles to cognitive science.

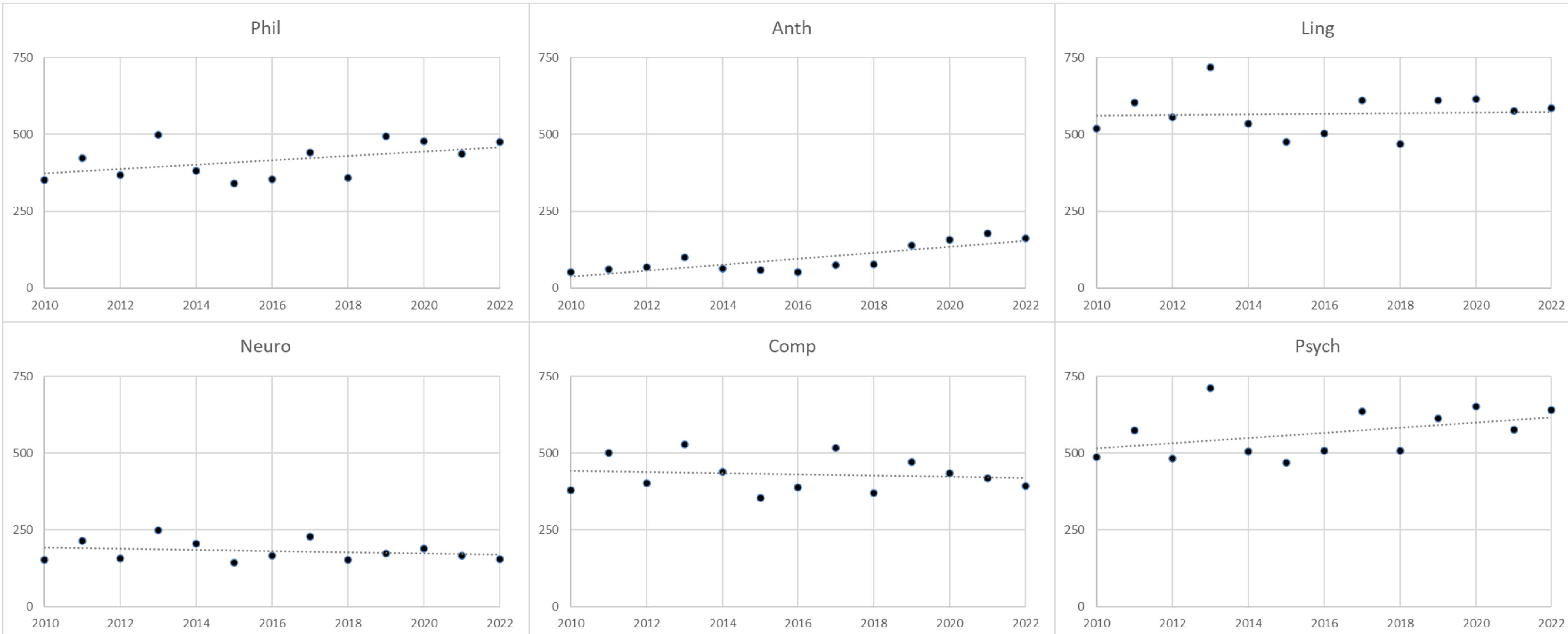


Figure 17: The number of CSS proceeding articles with similarity scores to the fields above the threshold value.

CHAPTER 5

DISCUSSION

Cognitive science has been established as an inherently cross-disciplinary field of research. Since its inception, various attempts have been made to evaluate cross-disciplinarity of the field. Recently, starting in 2019, there have been a series of publications dedicated to history, current state, and future of cognitive science's cross-disciplinarity nature from different perspectives. These perspectives have mainly included bibliometric and spatial approaches, socio-institutional aspects, and more rarely text-based techniques. In the present study, we offer a comprehensive picture of cross-disciplinarity in cognitive science by reporting analyses based on text similarity of the contributing fields. The aim of this research is to present a methodology for quantifying the cross-disciplinarity aspects of research in cognitive science and related fields. For this purpose, the study first investigated whether text similarity analysis provides valuable information about the relationship between academic documents (proceeding articles, journal articles) and relevant research fields. We then explored whether this analysis could serve as an indicator of cross-disciplinarity in cognitive science. To achieve this, we created datasets and used Doc2Vec for vectorization of the articles and cosine similarity for measuring the similarity among the datasets. We validated our approach by creating a similarity matrix for journal articles (titles and abstracts) across six constituent fields: philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. The motivation for using journal articles was the robustness of the bibliometric records, directly accessible from publisher databases, and their publisher-identified research field categorizations. The validation results were promising, as the journal articles returned sensible similarity scores relevant to their assigned constituent fields. We then created similarity matrices for field-specific conference proceedings to test the model against unstructured, full-text publications. The analyses revealed that the model was effective in generating text-similarity scores for field-specific proceedings. We applied these methods to process full-text documents, presenting an enriched methodology for evaluation. Next, we conducted a text-based analysis of cognitive science publications, focusing on their cross-disciplinary aspects, namely the contributions of respective subfields. Finally, we applied the Doc2vec model to analyze the proceedings of the Cognitive Science Society (CSS) Meetings.

The results provided supporting evidence that cognitive science is a cross-disciplinary field of research, within the limitations and operational assumptions of the present study. A major assumption that defined the scope of our analyses was that similarity measures offer a snapshot of cross-disciplinarity practices, in terms of the contributions from multiple fields to cognitive science articles. However, similarity analyses offer a limited view of cognitive science's cross-disciplinarity practices due to the challenges in measuring and classifying research outputs into their respective research fields. In literature, there are different approaches to quantify the cross-disciplinarity of fields: bibliometric methods & spatial measures, socio-institutional aspects, and text-based analysis. Each category provides its own perspective on the cross-disciplinarity of fields. For instance, citation analysis, an example of bibliometric methods, measures the influence of one field on others, which can provide meaningful insights into the dynamics of a given field. It can specify the level of cross-disciplinarity by evaluating the interactions between fields through cited articles and the citing articles and centrality measures of these relations. Co-authorship is an example of bibliometric and spatial measures. There are tools to visualize the relationships between co-authors working in different fields, helping to identify patterns or trends over time. Alternatively, there are studies conducted focusing on researchers and research environments, named as socio-institutional dimension in this dissertation. This dimension may include, but is not limited to, the backgrounds of scholars, their affiliations, and curriculums of departments. They can provide insight into one field's cross-disciplinarity characteristic by examining social and institutional factors. For example, the affiliations of researchers publishing in a field can reveal information about the profiles of active researchers in that field. Similarly, the curriculum of the departments might be an indicator of contributing fields to that specific department. As a third dimension, there are studies focusing on the content which can be evaluated with text-based evaluation metrics. For instance, keywords used in articles, abstracts of a text or full text of documents can be analyzed via NLP techniques to address the problem of how one field interacts with others. As noted, these three dimensions can be used in combination or separately to provide insights and understanding of cross-disciplinarity of fields. This dissertation employed a text-based approach, which we believe provides a complementary perspective to other methods, through substantial analysis of the content written by authors. Publicly available datasets and models offer an opportunity to develop techniques for assessment and to present researchers with robust ground for evaluating cognitive science, as it continues to develop as a cross-disciplinary field of research.

The first major limitation of the present study is that we used cross-disciplinarity in its broad definition. Our methodology was not entirely comprehensive in capturing the differences between the concepts of cross-disciplinarity, the terminology and concepts of which have been subject to intense debate in the literature. The terms are defined ambiguously in literature and sometimes used interchangeably (Choi & Pak, 2006). Based on the OECD conference in 1970 and follow-up studies, the interactions between the fields have been defined using different terms like interdisciplinarity, multidisciplinarity, and transdisciplinarity. In general, these terms imply the necessity of expertise from two or more fields to address complex problems. However, the approaches can differ based on definitions. For instance, according to National Academy of Sciences et al. (2005), interdisciplinary research was described as integrative or synthesis research, aiming to solve complex problems by integrating theories, techniques, or tools from two or more disciplines, whereas multidisciplinary research was described as combining fields through additive contributions of specific research fields, rather than integration. Stokols et al. (2008) defined multidisciplinarity as a sequential research process, preserving discipline-specific perspectives, while interdisciplinary research was characterized by its interactive and collaborative nature, focusing jointly on a topic while preserving individual disciplinary perspectives. On the other hand, he defined transdisciplinarity as an integrative process, aiming to establish an original model to address research questions by synthesizing and extending discipline-specific models. Although the vagueness, in their research, Núñez et al. (2019) mentioned multidisciplinary and interdisciplinary differences for the field of cognitive science drawing on the definitions of Choi & Pak (2006). They argued that multidisciplinarity involves using the expertise from different fields staying within boundaries. To stay within boundaries means not integrating the perspectives and findings of different fields approaching the problem. Their inference from the definition is that if a field is multidisciplinary then it is dominated by a traditional field. On the other hand, interdisciplinarity results in an “new identity” through a sophisticated level of cohesive integration of different fields. An interdisciplinary field’s unique feature prevents it depending on a traditional field. Based on these definitions, Núñez et al. (2019) presented that cognitive science field has been established as a multidisciplinary field. However, the analyses (two bibliometrics and two socio-institutional) indicated that the field has not managed to transform from multi- to interdisciplinary field. In our view, and as illustrated in this study, the vagueness of the borders between those concepts poses a major challenge, forcing researchers to adopt a broad understanding of cross-disciplinarity. Additionally, the methodology employed in our research is limited to distinguishing between these terms. In the scope of our methodology, the contributing fields and their similarities can be extracted for a document or a particular field. However, we cannot draw any conclusion about the level of integration between these fields, although our findings provide groundwork for further discussions. For instance, in Figure 16, we identified 53 articles that did not show significant similarities to any of the fields within our research scope. This anomaly could suggest two possibilities: the existence of other relevant fields beyond the scope of our research, the potential of these articles to bridge existing fields, indicating opportunities for transdisciplinary studies. In the scope of this research, we do not have a definitive explanation for this observation. Therefore, future research is needed to investigate if those connotations of interdisciplinarity, multidisciplinarity and transdisciplinarity can be analyzed through Natural Language Processing (NLP) techniques.

Another operational assumption in the present study was the unit size for the model input. In the model explained in "Running the Model for the CSS Proceedings" section, the documents were combined such that a single document, which involved many articles, representing one year of cognitive science field. For example, for the year 2000, approximately 250 conference proceedings were consolidated into a single document. Similarity scores were computed between these consolidated documents and contributing fields like philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. This approach resulted in calculating mean similarity scores for each field, rather than for each article in the proceedings. In a case where the focused field has a mono-disciplined structure, minimal ambiguity in the results could be expected. The similarity of a document or mean similarity of articles might align closely. However, in the case of cognitive science, the articles are expected to be a mixture of different fields. Therefore, the interpretation of results may bring ambiguity when a consolidated document is evaluated. The mean-similarity approach imposes a limitation on the findings of the study. That is a valid objection since the articles in the dataset may be related to a single domain or multiple domains and treating them as a member of a single set is a limitation. However, this approach was a practical requirement, due to the intensive manual labor involved in processing each article separately in the CSS proceedings appearing before 2010. We resolved this issue for the conference proceedings after 2010, due to their relatively structured design, and developed a model for calculating the similarity scores for individual articles as given in the "Running the model for Individual CSS Proceeding Articles" section. Although the average similarity scores are higher than the findings obtained in the initial model, the results revealed similar pictures between the former and the latter models. As an outcome, this specific study provided an insight into the number of contributing fields for each article in the proceedings. The results showed a normal distribution where most articles consisted of 2-4 contributing fields.

A research field may encompass numerous, relevant sub-fields, each with varying degrees of relationship to cognitive science. For instance, neuroscience spans work in a wide range of research endeavors from anatomy of giant squid axons to brain imaging of creative thinking. It is plausible that a cognitive science publication is expected to be more like the latter type of research than the former. It should be noted that the existence of various neuroscience articles that are not directly related to cognitive science is expected to artificially lower the similarity score, even if a considerable number of cognitively relevant articles exist. Similarly, computer science spans a range of sub-disciplines, including compiler design and artificial intelligence (AI). While the AI sub-field is expected to be more similar to cognitive science than compiler design, the existence of the latter sub-field within computer science may reduce the similarity between computer science and cognitive science. To address this limitation, we created six field corpora based on relevant WoS categories for each field, as detailed in Table 2. We were interested in the overlap between cognitive science and its contributing disciplines. However, the limitation persists since the categorization is done at the journal level rather than at the article level. Our analyses also indicated that this situation is partially reflected by the imperfect similarity scores. We prespecified the proposed level of analysis (i.e., the specification of six contributing fields to cognitive science) by employing a common definition for each of these contributing fields in cognitive science. This assumption allowed us to keep the number of models manageable throughout the study. However, a higher granularity level of modeling would lead to tens or hundreds of models, making the analyses hard to interpret. Future work should address those limitations by focusing on the role of further subfields.

In relation to previous limitation of research, we specifically grounded the scope of our cognitive science-related fields to those provided in the 1978 report of Sloan Foundation: philosophy, anthropology, linguistics, neuroscience, computer science, and psychology. Philosophy and cognitive science share common questions about the mind. Cognitive science borrows from the philosopher's tradition of studying the nature of the mind. Social and cultural context shaping the human mind are the overlapping focus areas of cognitive science and anthropology. Anthropologists examine how social norms and cultural practices influence human thinking. This knowledge is valuable for researchers in cognitive science aiming to understand how environment shapes the mind. Language acquisition, processing, and representation are key areas where linguistics and cognitive science collaborate. Cognitive scientists study how people parse phrases and understand meaning using language models. Neuroscience deals with the brain and nervous system, which are also important for a cognitive scientist to comprehend the biological basis of human cognition. To understand the brain mechanisms underpinning cognitive processes, cognitive science utilizes neuroscience techniques. Computer science, especially in AI, interacts intensively with cognitive science. It supports cognitive scientists with the tools for modeling and simulating cognitive processes. Psychology, which is also discussed as a dominant contributing field of cognitive science in literature, deals with human behavior and mental processes which are also a focus of cognitive scientists. They use theories and experiments of psychologists to understand how the mind works. The common overlapping topics are not limited to the given examples. In an era of increased interactions between scientific research fields, we should expect to observe more related fields for any field claiming cross-disciplinarity. Therefore, further research is needed to investigate the contribution of other subfields besides the six related fields examined here. This need has also been noted by scholars in various fields, such as business and decision science (Leydesdorff et al., 2008), education (Youtie et al., 2017), economics (French, 2019), business, law, literary studies, medicine, physics, and biology (Goldstone, 2019).

A technical challenge in developing the similarity models was establishing the field-specific datasets, such as the field category assigned by data provider. As described in Field-specific Journal Articles, WoS categories were used to define the fields. The relevant WoS categories and queries are presented in Table 2 of the same section. Each journal indexed by WoS is categorized into at least one of the 254 subject categories. To create a homogeneous field corpus, the journals assigned to more than one research field category were excluded. As an example, the International Journal of Psychophysiology was assigned to physiology, psychology, and neurosciences categories. Therefore, none of the articles published in this journal was not included in creating psychology or neuroscience training set. Similarly, IEEE Transactions on Cognitive and Developmental Systems journal's articles were excluded since it was categorized as a part of robotics, computer science, artificial intelligence, and neurosciences. Although this approach solved the problem of creating pure corpus, it is entirely dependent on WoS's categorization. A significant limitation regarding that is that WoS categorizes journals, not individual articles, meaning that unique characteristics of each article are not considered in creating field-specific corpora. For instance, an article titled "Intrusive Images in Psychological Disorders: Characteristics, Neural Mechanisms, and Treatment Implications" is a part of psychology corpus since it was published in the journal of Psychological Review which is categorized as a pure psychology journal by WoS. The authors preferred to use memory, imagery, neuroscience, psychopathology, and treatment keywords, indicating the article's cross-disciplinary structure. Another example is an article titled "A computational approach to paleoanthropology" published in a pure anthropology journal named Evolutionary Anthropology. Although we used it to create a corpus of anthropology, the author's keywords contain computer related terms like computation, computer graphics, computer tomography, signal theory and virtual reality. Therefore, our approach is limited by the cross-disciplinary scope of the journals categorized by WoS, rather than by the articles' genuine cross-disciplinarity. This operational assumption also impacted the training of the models. We observed that specific journals might be more closely related to cognitive science than to their categorized subfield (e.g., psychology). Recently, cognitive science has not been specified as a research field category in these databases (e.g., Web of Science). Our team found a definitive solution to this issue. One approach could be to expand the scope of datasets by including articles from the multiple sources employed by the present study. A manual inspection of the data suggests that these instances do not significantly affect the model's performance. Nevertheless, further research should address this challenge by filtering field-specific journals through manual, expert analysis. Another challenge in developing the corpora and the models was the need for manual cleansing and rearranging when preparing some of the proceedings as test data. Future work should address a higher granularity classification of article tokens, so that the training sets provide a better approximation for similarity to the relevant fields.

Another technical challenge in model development was the use of the titles and abstracts of journal articles for validating the prediction success of the model. Specifically, we utilized the titles and abstracts of highly cited articles, as detailed in the Field-specific Journal Articles section, to create similarity matrices for journal articles. This approach was employed to assess the model's predictive accuracy. We opted for titles and abstracts rather than full texts due to practical limitations such as data accessibility constraints and the substantial operational workload. In subsequent phases of our research, however, we incorporated full texts from conference proceedings to compute similarity scores within selected fields. For future research, it is recommended to evaluate the model using full-text data of journal articles and to compare these findings with those derived from limited content. This approach would provide a more comprehensive understanding of the model's efficacy in different contexts and with varying levels of data complexity.

The optimization of our model was constrained by a limited number of hyperparameter values. As detailed in Parameter Optimization part, our initial step was to assemble a base model (Model 1), employing parameter values that align with standard implementation practices in the literature for datasets of similar size. Subsequently, we constructed seven additional models (Models 2 to 8), each varying in its configuration to observe the impact on the Recall (R) values. In this process, we systematically adjusted one parameter value at a time to observe its influence on the average R score. This approach primarily involved running the model with a key set of parameters, modifying them individually, and then assessing the model's performance based on the R score for each variant. While this method was practical and straightforward, it lacked the rigor and systematic approach often essential for optimal model performance. It's crucial to recognize that more sophisticated and structured methods exist for conducting hyperparameter optimization. Techniques like grid search, random search, or Bayesian optimization allow for a systematic exploration of a wider array of parameter combinations, potentially leading to the discovery of more effective configurations. Our current approach, despite providing initial insights, might not have fully captured the model's optimal performance. Therefore, future research in this area would benefit from adopting a more methodical and comprehensive strategy in hyperparameter tuning. Such an approach could significantly enhance the model's accuracy and efficiency.

The scope of the present analyses is also specified by the limited variety of academic research outputs considered. In this study, journal articles and conference proceedings were chosen as representative academic outputs for training and testing the models. However, different academic output genres may be more representative of one research field than another. For instance, conference proceedings comprise the leading type of publication in computer science, whereas journal articles are more prevalent in psychology. The reason behind this selection is related to the nature of how researchers in each field work. For computer scientists, conference proceedings are regarded as a main publication venue to present their most recent findings regarding new algorithms or tools, collaborate with colleagues on new projects, and get feedback from peers about practical solutions and ideas in a timely manner. For such a rapidly evolving field, it is also the fastest way to disseminate information. Journal articles, as opposed to conference proceedings, are more prevalent in psychology. Psychological research usually requires long term investigations and experiments. This complexity makes it harder to present it in a shorter conference format. Instead, they require enough space and structure to present methods, results, and discussions on a psychology topic. Therefore, journal articles are a much preferable format for reporting the results. We excluded other outputs, such as other document types (books, book chapters, technical reports, editorial materials, letters, field notes, essays, briefs, white papers, patents), due to their limited acceptability and unstructured representation format. Future work should consider field-specific research outputs and consider their varying representative power within their respective disciplines.

Some of the findings in the present study may include ambiguities due to confounding factors—especially in the analyses of contributions from the computer science subfield. In Table 13, computer science related conferences are found a moderate degree of similarity to random selection of articles (approximately 0.3). The similarity between the scores of computer science and the random dataset indicates the presence of such a confound, likely due to the use of computer science terminology across many research fields. The question here is whether the utilization of basic standard computer tools is necessarily indicative of cross-disciplinarity in nature. For instance, using a programming language like Python in a cognitive science article will contribute to increase the similarity score between the two fields. A similar scenario could occur when comparing statistics to another field that primarily uses statistical methods. For example, employing ANOVA or any other statistical analysis in empirical research is not typically considered a contribution to statistical research. Researchers in various disciplines often employ these technical tools independently, without requiring collaboration from experts in other relevant fields. This ambiguity was unavoidable, given that software tools and computational methods have been widely used in research in many fields. On the other hand, this ambiguity may not have a significant impact on our findings. In our case, the content of the computer science articles in the dataset was broader than just descriptions of computational methods. Our data set encompassed a wide range of computer science topics, including artificial intelligence, computer architecture, computer networks, image processing, and cyber security. Furthermore, we did not observe an increasing trend in the similarity scores of computer science throughout the years. Figure 15 and Figure 17 present a trend overview of fields based on similarity scores to cognitive science sets. According to the figures, while the first one exhibits a decreasing trend over time, the second figure shows a stable trend indicating that the contribution of computer science to cognitive science has remained relatively constant. Further research is needed for a closer look at the role of computer science terminology in cognitive science, and in other (seemingly) unrelated research fields.

Over the decades, academic fields have constantly evolved in various directions due to a range of internal and external motivations (Cohen, E., & Lloyd, S. ,2014). New ideas, methods, approaches, technologies, and other novel concepts have shaped the evolution of fields. Furthermore, interactions with different disciplines may also reform the ways of thinking and refine the major concepts of a given field. In summary, academic fields are dynamic and constantly evolving throughout history. This dynamism raises a question about the necessity of examining temporal patterns in this research. We used Annual Meetings of Cognitive Science Society conferences published between 1981 and 2022 as the test set (each year separately) and the titles and the abstracts of the highly cited articles from the Clarivate Analytics Web of Science (WoS) database for the years between 1990 and 2019 as the training set (all years combined as a single document as described in methodology section). A proceeding published from the early 80s is thus compared to a field vector that includes articles published after 2010. Similarly, a proceeding published in 2019 is compared to a field corpus encompassing articles from the 90s. Clearly, one article cannot build upon work from the future, nor can it influence past research. Since it is impractical to verify the temporal compatibility of all instances, we adopted a methodology that covers long durations. We assume that the concept of temporality is an operational assumption for this study and requires the creation of new sets.

Notwithstanding these limitations, the study offers initial insights into our understanding of cross-disciplinary characteristics of cognitive science over the period in question, achieved by providing similarity scores. A major finding is that cognitive science may be conceived as a cross-disciplinary field of research since the proceedings have consistently addressed multiple, relevant subfields, rather than being focused on a single dominant area. More specifically, our models revealed that four main subfields (linguistics, psychology, computer science, and philosophy) have contributed to cognitive science more significantly than the remaining two subfields (neuroscience and anthropology). This comprehensive overview of the findings confirms that cognitive science has consistently maintained its cross-disciplinary structure over time, without being dominated by a single field, in line with the findings reported by Oey et al. (2020).

We also found that CSS proceeding had low similarity scores to neuroscience, contrary to Leydesdorff and Goldstone's (2014) findings that emphasize the neuroscience's central role in cognitive science. In their bibliometric study of the journal Cognitive Science, Leydesdorff and Goldstone (2014) analyzed long-term citation data, suggesting that a specific focus on citations might reveal a different picture of the relationships between the relevant fields. The observed low similarity in our study might also be attributed to the distinctions between the two communities. Typically, the neuroscience community publishes in its own specific venues, which could explain the lower overlap with cognitive science as reflected in CSS proceedings.

Anthropology also revealed a low contribution to cognitive science articles, as noted by Núñez et al. (2019) and French (2019). One possible explanation for this is that anthropology researchers often prefer different methods of disseminating their scholarship, such as through monographs, rather than article publications. Although Bender (2019) agreed on anthropology's limited representation in cognitive science, they found optimistic results in titles and keywords analysis of the topiCS journal between 2009 and 2018. Their findings showed that a specific focus on analyzing titles and keywords may reveal a different picture of the relationship between cognitive science and relevant research fields. Nevertheless, our findings about psychology's position in cognitive science align with Bender's findings.

Psychology, subject to intense debate as a contributing field to cognitive science, is an essential part of the present study. Contrary to the findings of several previous studies (Von Eckardt, 2001; Leydesdorff and Goldstone, 2014; Gentner, 2019; Goel, 2019; Goldstone, 2019; Núñez et al., 2019; Rosenbloom & Forbus, 2019), we found no evidence to suggest the dominance of psychology in cognitive science. The discrepancies between our findings and those of previous research appear to stem from differences in the type of bibliometric data used, the methods of analysis, and the time frames of the publications examined. For example, Núñez et al. (2019) focused on author affiliations and citation environments for their selected years in the bibliometric part of their analysis, rather than examining the content of the publications themselves.

Another divergence in our findings is the contribution of computer science to the field. We found computer science as a notable contributing field to cognitive science. Although Von Eckardt (2001) and Goldstone & Leydesdorff (2006) reported computer science as one of the dominant fields, our results present a balanced contribution to cognitive science, exhibiting a decreasing trend, aligning with the observations of Rosenbloom & Forbus (2019) and Cooper (2019). The major difference between the work of Goldstone & Leydesdorff's (2006) and our study lies in the time frame and the type of analysis conducted. Their analysis focused on the citations of the journals published in 2004.

As for the contribution of philosophy to cognitive science, our findings highlighted it as a significant contributor, in contrast to Goldstone & Leydesdorff (2006) and Leydesdorff and Goldstone (2014), both of whom pointed to a minor role for philosophy in cognitive science.

Linguistics has been defined as a contributor to cognitive science as provided in the 1978 report of Sloan Foundation. However, a linkage that has not been extensively explored in prior studies. We found that linguistics played a crucial role in cognitive science. This finding stands in contrast to the conclusions drawn by Leydesdorff et al. (2008), who using different data sets and analytical approaches, reached divergent conclusions. We emphasize that those divergences should be conceived as complementary findings, revealing different pictures of the multifaceted research in cognitive science, rather than as contradictions; there were simply many differences in our data, methodology, and time frame of the publications being considered.

CHAPTER 6

CONCLUSION

The present study aimed to quantify the cross-disciplinarity of cognitive science by employing a document similarity approach. Specifically, we investigated how a text-based measure—more specifically a document similarity approach—might be employed to reveal the cross-disciplinary structure of cognitive science. We found that Doc2vec and cosine similarity provide robust data for evaluating the trends shaping cognitive science as a cross-disciplinary field of research. Nevertheless, assessing the cross-disciplinarity of a field has multiple fronts. First, defining the scope and limits of a field is a complex task. Questions such as "What constitutes the boundaries of a field?" and "How do we draw the frame of the field?" need to be addressed. Concerning this front, the unit of analysis (author, department, journals, communities) also requires consideration. Is analyzing the cross-disciplinarity of authors sufficient to define a field's cross-disciplinarity? Do journals or conferences reflect the best representation of the field? Second, the chosen methodology—for instance, bibliometric analysis, spatial measures, text-based studies, and socio-institutional evaluations—shapes the analysis's scope. Can we develop a perfect approach or a combination of approaches to define the level of cross-disciplinarity? Different methodologies may explain various dimensions of cross-disciplinarity. The characteristics of the dataset also play a crucial role in influencing the results. For instance, a representative period, effective search queries, and reliable data sources are needed for comprehensive analyses. Multiple data sets may also be needed to address those requirements. Consequently, we believe that there is no clear-cut answer (nor is it a yes/no question) for the cross-disciplinarity of cognitive science.

Is similarity analysis an appropriate method for quantifying cross-disciplinarity, given that traditional research domains, such as computer science, may overlap highly with relevant fields, such as logic and mathematics? Analyzing a traditional field alongside its related fields may provide additional insights into their overlapping characteristics. Nevertheless, it's important to note that the relationship between similarity and cross-disciplinarity is akin to the relationship between correlation and causation in statistics. In essence, while similarity does not necessarily imply cross-disciplinarity, cross-disciplinary fields are expected to exhibit similarities with their relevant fields. In the context of present study, the cross-disciplinarity of cognitive science is given, in that the field was established as a cross-disciplinary research field. Our findings show how the relevant fields contribute to cognitive science without the apparent dominance of any one relevant field over the other. Therefore, we interpret these field-specific contributions, measured by the similarity scores, as indicators of cross-disciplinarity in cognitive science.

In further research, the definition of fields needs to be improved to prevent bias in training sets, which may require eliminating cross-disciplinary articles from the sets. We expect such an enhancement to contribute to the quality of training sets. One approach could be the removal of very highly cited articles from the dataset at the preprocessing step, as they may introduce some bias, due to potentially exhibiting more cross-disciplinary characteristics than others. Nevertheless, such adjustments during preprocessing would require justifications for identifying thresholds. Given that citation distributions may not be homogeneous across the fields, we included these highly cited articles in our datasets for the present study. Future research should explore the impact of very highly cited articles on the results. Additionally, incorporating the periods of scientific activity as an integral part of the corpora is essential. This inclusion could allow for observing varying trends in the structure of fields over time. For instance, adding a temporal dimension may enable comparisons within the same time frames by training models for selected periods and calculating similarity scores of test sets published in corresponding periods.

It is worth noting that emerging technologies like GPT-4 offer potential possibilities for exploring the research questions presented in this work. The advanced capabilities and sophisticated features of these technologies, which include deep learning and NLP abilities, have the potential to analyze and interpret complex datasets. Nevertheless, it is crucial to emphasize that, as of now, these technologies have not validated sufficiently to be reliably employed in academic research. Extensive testing, validation studies, and academic evaluation are necessary to establish the validity, reliability, and potential biases of these tools before they can be integrated into established research methodologies.

REFERENCES

Abramo, G., D'Angelo, C. A., & Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs. the disciplinary diversity of the reference list. *Journal of Informetrics*, *12*(4), 1182–1193. https://doi.org/10.1016/j.joi.2018.09.001

Alasehir, O., & Acarturk, C. (2022). Interdisciplinarity in Cognitive Science: A Document Similarity Analysis. *Cognitive Science*, *46*(12), e13222. https://doi.org/10.1111/cogs.13222

Alberts, B. (2013). Designing Scientific Meetings. *Science*, *339*(6121), 737. https: //doi.org/10.1126/science.1236324

Balikcioglu, P. G., Sirlanci, M., Kucuk, O. A., Ulukapi, B., Turkmen, R. K., Acarturk, C. (in press). Malicious code detection in Android: The role of sequence characteristics and disassembling methods. *International Journal of Information Security*. Httos://doi.org/10.1007/s10207-022-00626-2.

Bark, R. H., Kragt, M. E., & Robson, B. J. (2016). Evaluating an interdisciplinary research project: Lessons learned for organisations, researchers and funders. *International Journal of Project Management*, *34*(8), 1449–1459. https://doi.org/10.1016/j.ijproman.2016.08.004

Bender, A. (2019). The Value of Diversity in Cognitive Science. *Topics in Cognitive Science*, *11*(4), 853–863. https://doi.org/10.1111/tops.12464

Berger, G. (1972). Introduction. In L. Apostel, G. Berger, A. Briggs, & G. Michaud (Eds.), *Interdisciplinarity Problems of Teaching and Research in Universities* (pp. 23–

24). Organisation for Economic Cooperation and Development (OECD) and Center for Educational Research and Innovation (CERI).

Bergmann, T., Dale, R., Sattari, N., Heit, E., & Bhat, H.S. (2017). The interdisciplinarity of collaborations in cognitive science. *Cognitive Science*, *41*(5),1412–1418. https://doi.org/10.1111/cogs.12352.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(4-5), 993–1022.

Boden, M. A. (2006). *Mind as machine: A history of cognitive science*. Oxford University Press.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Bordons, M., Morillo, F., Gómez, I. (2004). *Analysis of Cross-Disciplinary Research Through Bibliometric Tools*. In: Moed, H.F., Glänzel, W., Schmoch, U. (eds) Handbook of Quantitative Science and Technology Research. Springer, Dordrecht. https://doi.org/10.1007/1-4020-2755-9_20

Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5192–5199.

Broude, G. J., Livingston, K. R., Leeuw, J. R. de, Andrews, J. K., & Long, J. H. (2019). Rumors of Our Death…. *Topics in Cognitive Science*, *11*(4), 864–868. https://doi.org/10.1111/tops.12466

Carr, G., Loucks, D. P., & Blöschl, G. (2018). Gaining insight into interdisciplinary research and education programmes: A framework for evaluation. *Research Policy*, *47*(1), 35–48. https://doi.org/10.1016/j.respol.2017.09.010

Chakraborty, T. (2018). Role of interdisciplinarity in computer sciences: quantification, impact and life trajectory. *Scientometrics*, *114*, 1011–1029. https://doi.org/10.1007/s11192-017-2628-z

Chandrasekaran, D., & Mago, V. (2021). Evolution of semantic similarity—a survey. *ACM Computing Surveys (CSUR)*, *54(2)*, 1–37. https://doi.org/10.1145/3440755

Chen, P., Wu, F., Wang, T., & Ding, W. (2018). A Semantic QA-Based Approach for Text Summarization Evaluation. *Proceedings of the AAAI Conference on Artificial*

Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11911

Choi, B. C., & Pak, A. W. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and investigative medicine. Medecine clinique et experimentale*, 29(6), 351–364. https://doi.org/10.1111/tops.12609

Cohen, E., & Lloyd, S. (2014). Disciplinary evolution and the rise of the transdiscipline. *Informing Science: the International Journal of an Emerging Transdiscipline*, 17,189-215. Available: http://www.inform.nu/Articles/Vol17/ISJv17p189-215Cohen0702.pdf

Cognitive Science Society. (2019). *Cognitive Science Society Past Conferences*. Retrieved October 18, 2022, from https://cognitivesciencesociety.org/past-conferences

CogSci Proceedings. (2020). Retrieved October 18, 2022, from https://cogsci.mindmodeling.org

Contreras Kallens, P., Dale, R. and Christiansen, M.H. (2022). Quantifying Interdisciplinarity in Cognitive Science and Beyond. *Topics in Cognitive Science*, 14, 634-645. https://doi.org/10.1111/tops.12609

Contreras Kallens, P. and Dale, R. (2018). Exploratory mapping of theoretical landscapes through word use in abstracts. *Scientometrics*, 116, 1641–1674. https://doi.org/10.1007/s11192-018-2811-x

Cooper, R. P. (2019). Multidisciplinary Flux and Multiple Research Traditions Within Cognitive Science. *Topics in Cognitive Science*, *11*(4), 869–879. https://doi.org/10.1111/tops.12460

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569. https://doi.org/10.1002/asi.1097

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Deng, S., & Xia, S. (2020). Mapping the interdisciplinarity in information behavior research: A quantitative study using diversity measure and co-occurrence analysis. *Scientometrics*, *124*(1), 489–513. https://doi.org/10.1007/s11192-020-03465-x

DeStefano, I., Oey, L. A., Brockbank, E., & Vul, E. (2021). Integration by Parts: Collaboration and Topic Structure in the CogSci Community. *Topics in Cognitive Science*, *13*(2), 399–413. https://doi.org/10.1111/tops.12526

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at http://arxiv.org/abs/1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Dias, L., Gerlach, M., Scharloth, J., & Altmann, E. G. (2018). Using text analysis to quantify the similarity and evolution of scientific disciplines. *Royal Society Open Science*, *5*(1). https://doi.org/10.1098/rsos.171545

European Commission. (n.d.). Funding programmes and open calls. Research and Innovation. Retrieved March 15, 2023, from https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls_en

Evans, E. D. (2016). Measuring Interdisciplinarity Using Text. *Socius*, *2*(2016), 1-18. https://doi.org/10.1177/2378023116654147

Eykens, J., Guns, R., & Vanderstraeten, R. (2022). Subject specialties as interdisciplinary trading grounds: the case of the social sciences and humanities. *Scientometrics*, *127*, 7193–7213. https://doi.org/10.1007/s11192-021-04254-w

Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research*, *39(*3), 251-277. https://doi.org/10.1177/1046496408317797

French, R. M. (2019). Missing the Forest for the Trees: Why Cognitive Science Circa 2019 Is Alive and Well. *Topics in Cognitive Science, 11*(4), 880–883. https://doi.org/10.1111/tops.12465

Frodeman, R. (2010). Introduction. In R. Frodeman, J. Thompson, & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (First Edition, p. XXXİ). Oxford University Press.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. Basic Books.

Gentner, D. (2010). Psychology in Cognitive Science: 1978–2038. *Topics in Cognitive Science*, *2*(3), 328–344. https://doi.org/10.1111/j.1756-8765.2010.01103.x

Gentner, D. (2019). Cognitive Science Is and Should Be Pluralistic. *Topics in Cognitive Science*, *11*(4), 884–891. https://doi.org/10.1111/tops.12459

Glänzel, W., & Schubert, A. (2005). Analysing Scientific Networks Through Co-Authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 257–276). Springer Netherlands. https://doi.org/10.1007/1-4020-2755-9_12

Goel, A. (2019). A Cognitive Reformation. *Topics in Cognitive Science*, *11*(4), 892–901. https://doi.org/10.1111/tops.12469

Goldstone, R. L. (2019). Becoming Cognitive Science. *Topics in Cognitive Science*, *11*(4), 902–913. https://doi.org/10.1111/tops.12463

Goldstone, R. L., & Leydesdorff, L. (2006). The Import and Export of Cognitive Science. *Cognitive Science*, *30*(6), 983–993. https://doi.org/10.1207/s15516709cog0000_96

Graff, H. J. (2015). *Undisciplining Knowledge: Interdisciplinarity in the Twentieth Century* (1st edition). Johns Hopkins University Press.

Gray, W. D. (2019). Welcome to Cognitive Science: The Once and Future Multidisciplinary Society. *Topics in Cognitive Science*, *11*(4), 838–844. https://doi.org/10.1111/tops.12471

Hirst, P. H. (2010). *Knowledge and the curriculum: A collection of philosophical papers*. Routledge.

Hirst, G., & Budanitsky, A. (2005). Correcting real-word spelling errors by restoring lexical cohesion. *Natural Language Engineering*, *11*(1), 87–111. https://doi.org/10.1017/S1351324904003560

Hope, R. (2011). *Cognitive Science Society's Journal Archive*. Retrieved October 18, 2022, from http://csjarchive.cogsci.rpi.edu/

Huang, D. (2015). Temporal evolution of multi-author papers in basic sciences from 1960 to 2010. *Scientometrics*, *105*(3), 2137–2147. https://doi.org/10.1007/s11192-015-1760-x

Huang, M., & Chang, Y. (2011). A study of interdisciplinarity in information science: using direct citation and co-authorship analysis. *Journal of Information Science*, *37*(4), 369-378. https://doi.org/10.1177/0165551511407141

Huber, M. T., & Morreale, S. P. (2002). *Disciplinary styles in the scholarship of teaching and learning: Exploring common ground*. AAHE Publications Orders

Islam, A., & Inkpen, D. (2008). Semantic text similarity using corpus-based word similarity and string similarity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *2*(2), 1-25. https://doi.org/10.1145/1376815.1376819

Jacobs, J. A. (2014). *In defense of disciplines. In In Defense of Disciplines.* University of chicago Press.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kim, H. K., Kim, H., & Cho, S. (2017). Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. *Neurocomputing*, *266*, 336–352. https://doi.org/10.1016/j.neucom.2017.05.046

Kim, S., Fiorini, N., Wilbur, W. J., & Lu, Z. (2017). Bridging the gap: Incorporating a semantic similarity measure for effectively mapping PubMed queries to documents. *Journal of biomedical informatics*, *75*, 122–127. https://doi.org/10.1016/j.jbi.2017.09.014

Kim, Y. (2014). Convolutional neural networks for sentence classification. *roceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*, 1746-1751

Klein, J. T., & Newell, W. (1996). Advancing Interdisciplinary Studies. In J. Gaff & J. Ratcliff (Eds.), *Handbook of the Undergraduate Curriculum: Comprehensive Guide to Purposes, Structures, Practices, and Change* (1st Edition, pp. 393–415). Jossey-Bass.

Ko, Y., Park, J., & Seo, J. (2004). Improving text categorization using the importance of sentences. *Information Processing & Management*, *40*(1), 65–79. https://doi.org/10.1016/S0306-4573(02)00056-0

Kuhn, T. S. (1962). *The structure of scientific revolutions* (1st Edition). Chicago, IL: The University of Chicago Press.

Kwon, S., Solomon, G. E. A., Youtie, J., & Porter, A. L. (2017). A measure of knowledge flow between specific fields: Implications of interdisciplinarity for impact and funding. *PLOS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0185583

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Proceedings of the 31st International Conference on Machine Learning*, *32*(2), 1188-1196

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*(9), 1303–1319. https://doi.org/10.1002/asi.20614

Leydesdorff, L., & Goldstone, R. L. (2014). Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal cognitive science. *Journal of the Association for Information Science and Technology*, *65*(1), 164–177. https://doi.org/10.1002/asi.22953

Leydesdorff, L., Goldstone, R. L., & Schank, T. (2008). Betweenness centrality and the interdisciplinarity of cognitive science. *Cognitive Science: A Multidisciplinary Journal*, *Supplementary Material*. Retrieved October 18, 2022, from https://cognitivesciencesociety.org/wp-content/uploads/2019/01/Leydesdorff.pdf

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, *5*(1), 87–100. https://doi.org/10.1016/j.joi.2010.09.002

Manning, C.D., Raghavan P., & Schütze, H. (2009). *An introduction to information retrieval*. Cambridge university press.

Mark Hickson, I. (2006). Raising the Question #4 Why Bother Attending Conferences? *Communication Education*, *55*(4), 464–468. https://doi.org/10.1080/03634520600917632

Martens, B., & Saretzki, T. (1993). Conferences and courses on biotechnology. Describing scientific communication by exploratory methods. *Scientometrics*, *27*(3), 237–260. https://doi.org/10.1007/BF02016941

McLevey, J., Graham, A.V., McIlroy-Young, R. et al. (2018). Interdisciplinarity and insularity in the diffusion of knowledge: an analysis of disciplinary boundaries between philosophy of science and the sciences. *Scientometrics*, *117*, 331–349. https://doi.org/10.1007/s11192-018-2866-8

McShane, M., Bringsjord, S., Hendler, J., Nirenburg, S., & Sun, R. (2019). A Response to Núñez et al.’s (2019) “What Happened to Cognitive Science?”. *Topics in Cognitive Science*, *11*(4), 914–917. https://doi.org/10.1111/tops.12458

Mihalcea, R., Corley, C. D., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. *In Proceedings of the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference (pp. 775-780)*. Boston, Massachusetts, USA: AAAI Press.

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013a). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, 3111–3119

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9

Moran, J. (2002). *Interdisciplinarity*. Routledge

Mohamed, M., & Oussalah, M. (2019). SRL-ESA-TextSum: A text summarization approach based on semantic role labeling and explicit semantic analysis. *Information Processing & Management*, *56*(4), 1 1356–1372. https://doi.org/10.1016/j.ipm.2019.04.003

National Academies of Sciences, Engineering, and Medicine. 2005. *Facilitating Interdisciplinary Research* (pp. 30–40). Washington, DC: The National Academies Press. https://doi.org/10.17226/11153.

National Academies of Sciences, Engineering, and Medicine. 2015. Enhancing the Effectiveness of Team Science. Washington, DC: The National Academies Press. https://doi.org/10.17226/19007.

Nichols, L.G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics, 100,* 741–754.

https://doi.org/10.1007/s11192-014-1319-2.

Nguyen, H. T., Duong, P. H., & Cambria, E. (2019). Learning short-text semantic similarity with word embeddings and external knowledge sources. *Knowledge-Based Systems*, *182*, 104842. https://doi.org/10.1016/j.knosys.2019.07.013.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791. https://doi.org/10.1038/s41562-019-0626-2

Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2020). For the Sciences They Are A-Changin': A Response to Commentaries on Núñez et al.'s (2019) "What Happened to Cognitive Science?". *Topics in Cognitive Science*, *12*(3), 790–803. https://doi.org/10.1111/tops.12511

Oey, L., Destefano, I., Brockbank, E., & Vul, E. (2020). Formalizing interdisciplinary collaboration in the CogSci community. In S. Denison, M. L. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (pp. 474–480)

Park, E.-K., Ra, D.-Y., & Jang, M.-G. (2005). Techniques for improving web retrieval effectiveness. *Information Processing & Management*, *41*(5), 1207–1223. https://doi.org/10.1016/j.ipm.2004.08.002

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

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics* (pp. 2227-2237)

Porter, A. L., Carley, S. F., Cassidy, C., Youtie, J., Schoeneck, D. J., Kwon, S., & Solomon, G. E. A. (2019). Measuring Interdisciplinary Research Categories and Knowledge Transfer: A Case Study of Connections between Cognitive Science and Education. *Perspectives on Science*, *27*(4), 582–618. https://doi.org/10.1162/posc_a_00317

Porter, A. L., Cohen, A. S., David Roessner, J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, *72*(1), 117–147.

https://doi.org/10.1007/s11192-007-1700-5

Porter, A.L., Roessner J.D., Cohen, A.S., & Perreault, M. (2003). Interdisciplinary research: meaning, metrics and nurture. *Research Evaluation*, *15*(3), S187–S195. https://doi.org/10.3152/147154406781775841

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719. https://doi.org/10.1007/s11192-008-2197-2

Rafols, I. (2014). Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact: Methods and Practice* (pp. 169–190). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_8

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287. https://doi.org/10.1007/s11192-009-0041-y

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, *61*(9), 1871–1887. https://doi.org/10.1002/asi.21368

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Rosenbloom, P. S., & Forbus, K. D. (2019). Expanding and Repositioning Cognitive Science. *Topics in Cognitive Science*, *11*(4), 918–927. https://doi.org/10.1111/tops.12468

Sanders, K., Kraimer, M. L., Greco, L., Morgeson, F. P., Budhwar, P. S., Sun, J.-M. (James), Shipton, H., & Sang, X. (2020). Why academics attend conferences? An extended career self-management framework. *Human Resource Management Review*.https://doi.org/10.1016/j.hrmr.2020.100793

Sandoz, R. Interactive Historical Atlas of the Disciplines, University of Geneva. Entry "Plato": URL=< https://atlas-disciplines.unige.ch/#Plato> (accessed 16.12.2022).

Schmoch, U., Breiner, S., Cuhls, K., Hinze, S., & Münt, G. (1994). Interdisciplinary cooperation of research teams in science–intensive areas of technology. *Final* *report to the Commission of the European Unit (VALUE II, Interface II, HS1). Karlsruhe: Fraunhofer Institute for Systems and Innovation Research*.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59*(3), 425–465. https://doi.org/10.1023/B:SCIE.0000018542.71314.38

Schunn, C. D. (2019). What Should Cognitive Science Look Like? Neither a Tree Nor Physics. *Topics in Cognitive Science, 11*(4), 845–852. https://doi.org/10.1111/tops.12467

Schunn, C. D., Crowley, K., & Okada, T. (1998). The growth of multidisciplinarity in the Cognitive Science Society. *Cognitive Science*, *22*(1), 107–130. https://doi.org/10.1016/S0364-0213(99)80036-6

Schunn, C. D., Okada, T., & Crowley, K. (1995). Is cognitive science truly interdisciplinary?: The case of interdisciplinary collaborations. *In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*.

Sengupta, I. N. (1992). Bibliometrics, informetrics, scientometrics and librametrics: an overview. *Libri, 42*, 75–98. https://doi.org/10.1515/libr.1992.42.2.75

Serrano, J. I., del Castillo, M. D., & Carretero, M. (2014). Cognitive? Science? *Foundations of Science*, *19*(2), 115–131. https://doi.org/10.1007/s10699-013-9323-1

Silva, F. N., Rodrigues, F. A., Oliveira Jr, O. N., & Costa, L. D. F. (2013). Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics*, *7*(2), 469-477. https://doi.org/10.1016/j.joi.2013.01.007

Solomon, G. E. A., Youtie, J., Carley, S., & Porter, A. L. (2019). What people learn about how people learn: An analysis of citation behavior and the multidisciplinary flow of knowledge. *Research Policy*, *48*(9). https://doi.org/10.1016/j.respol.2019.103835

Stehr, N., & Weingart, P. (Eds.). (2000). *Practising interdisciplinarity*. University of Toronto Press.

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The Science of Team Science: Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine*, *35*(2, Supplement), S77–S89. https://doi.org/10.1016/j.amepre.2008.05.002

Thagard, P. (2005a). Being interdisciplinary: Trading zones in cognitive science. In S. J. Derry, C. D. Schunn, & M. Ann (Eds.), *Interdisciplinary collaboration: An emerging cognitive science* (1st Edition, pp. 317–339). Erlbaum Mahway, NJ.

Thagard, P. (2005b). *Mind: Introduction to Cognitive Science* (2nd Edition). MIT Press.

Thagard, P. (2010). Cognitive Science. In R. Frodeman, J. Thompson, & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (First Edition, pp. 234–245). Oxford University Press.

Tien N.H., Le N.M., Tomohiro Y., Tatsuya I. (2019). Sentence modeling via multiple word embeddings and multi-level comparison for semantic textual similarity. *Information Processing and Management*, *56* (6), art. no. 102090

Turner, S. P. (2000). “What Are Disciplines? And How Is Interdisciplinarity Different?”. In P. Weingart and N. Stehr, eds., *Practicing Interdisciplinarity* (pp. 46–65). University of Toronto Press.

Turner, S. (2017). Knowledge Formations: An Analytic Framework. In R. Frodeman, J. Thompson, & R. Pacheco (Eds.), *The Oxford Handbook of Interdisciplinarity* (Second Edition, pp. 17). Oxford University Press.

Urata, H. (1990). Information flows among academic disciplines in Japan. *Scientometrics*, *18*(3-4), 309-319. https://doi.org/10.1007/bf02017767

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472. https://doi.org/10.1016/j.respol.2010.11.001

Von Eckardt, B. (2001). Multidisciplinarity and cognitive science. *Cognitive Science*, *25*(3), 453–470. https://doi.org/10.1016/S0364-0213(01)00043-X

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039. https://doi.org/10.1126/science.1136099

Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S. (2016). Interdisciplinary topics of information science: A study based on the terms interdisciplinarity index series. *Scientometrics*, *106*(2), 583–601. https://doi.org/10.1007/s11192-015-1792-2

Yegros-Yegros, A., Rafols, I., & D’Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal

Interdisciplinarity. *PloS One*, *10*(8). https://doi.org/10.1371/journal.pone.0135095

Youtie, J., Solomon, G. E. A., Carley, S., Kwon, S., & Porter, A. L. (2017). Crossing borders: A citation analysis of connections between Cognitive Science and Educational Research … and the fields in between. *Research Evaluation*, *26*(3), 242–255. https://doi.org/10.1093/reseval/rvx020

Zhu, G., Iglesias, C.A (2018). Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Systems with Applications*, *101*(2018), 8-24

Zierath, J. R. (2016). Building Bridges through Scientific Conferences. *Cell*, *167*(5), 1155–1158. https://doi.org/10.1016/j.cell.2016.11.006

APPENDICES

APPENDIX A

List of Stop Words

Table 19: List of stop words eliminated in analysis.

a	here	ours	some
about	hers	ourselves	such
above	herself	out	than
after	him	over	that
again	himself	own	the
against	his	s	their
all	how	same	theirs
am	i	she	them
an	if	should	themselves
and	in	so	then
any	into	some	there
are	is	such	these
as	it	than	they

At	its	that	this
Be	itself	the	those
because	just	their	through
been	me	theirs	to
before	more	them	too
being	most	themselves	under
below	my	then	until
between	myself	there	up
both	no	these	very
but	nor	they	was
by	not	this	we
could	now	those	were
did	of	through	what
do	off	to	when
does	on	too	where
doing	once	under	which
down	only	until	while
during	or	up	who
each	other	very	whom
few	ought	was	why
for	our	we	with
from	ours	were	would
further	ourselves	what	you
had	out	when	your
has	over	where	yours
have	own	which	yourself

having	s	while	yourselves
he	same	who	
her	she	whom	

APPENDIX B

'Doc2vec paragraph embeddings' API reference

class **gensim.models.doc2vec.Doc2Vec**(*documents=None, corpus_file=None, vector_size=100, dm_mean=None, dm=1, dbow_words=0, dm_concat=0, dm_tag_count=1, dv=None, dv_mapfile=None, comment=None, trim_rule=None, callbacks=(), window=5, epochs=10, shrink_windows=True, **kwargs*)

Bases: `Word2Vec`

Class for training, using and evaluating neural networks described in Distributed Representations of Sentences and Documents.

Parameters:

- **documents** (iterable of list of `TaggedDocument`, optional) – Input corpus, can be simply a list of elements, but for larger corpora,consider an iterable that streams the documents directly from disk/network. If you don't supply *documents* (or *corpus_file*), the model is left uninitialized – use if you plan to initialize it in some other way.
- **corpus_file** (*str, optional*) – Path to a corpus file in `LineSentence` format. You may use this argument instead of *documents* to get performance boost. Only one of *documents* or *corpus_file* arguments need to be passed (or none of them, in that case, the model is left uninitialized). Documents' tags are assigned automatically and are equal to line number, as in `TaggedLineDocument`.
- **dm** (*{1,0}, optional*) – Defines the training algorithm. If *dm=1*, 'distributed memory' (PV-DM) is used. Otherwise, *distributed bag of words* (PV-DBOW) is employed.

- **vector_size** (*int, optional*) – Dimensionality of the feature vectors.
- **window** (*int, optional*) – The maximum distance between the current and predicted word within a sentence.
- **alpha** (*float, optional*) – The initial learning rate.
- **min_alpha** (*float, optional*) – Learning rate will linearly drop to *min_alpha* as training progresses.
- **seed** (*int, optional*) – Seed for the random number generator. Initial vectors for each word are seeded with a hash of the concatenation of word + *str(seed)*. Note that for a fully deterministically-reproducible run, you must also limit the model to a single worker thread (*workers=1*), to eliminate ordering jitter from OS thread scheduling. In Python 3, reproducibility between interpreter launches also requires use of the *PYTHONHASHSEED* environment variable to control hash randomization.
- **min_count** (*int, optional*) – Ignores all words with total frequency lower than this.
- **max_vocab_size** (*int, optional*) – Limits the RAM during vocabulary building; if there are more unique words than this, then prune the infrequent ones. Every 10 million word types need about 1GB of RAM. Set to *None* for no limit.
- **sample** (*float, optional*) – The threshold for configuring which higher-frequency words are randomly downsampled, useful range is (0, 1e-5).
- **workers** (*int, optional*) – Use these many worker threads to train the model (=faster training with multicore machines).
- **epochs** (*int, optional*) – Number of iterations (epochs) over the corpus. Defaults to 10 for Doc2Vec.
- **hs** (*{1,0}, optional*) – If 1, hierarchical softmax will be used for model training. If set to 0, and *negative* is non-zero, negative sampling will be used.
- **negative** (*int, optional*) – If > 0, negative sampling will be used, the int for negative specifies how many "noise words" should be drawn (usually between 5-20). If set to 0, no negative sampling is used.
- **ns_exponent** (*float, optional*) – The exponent used to shape the negative sampling distribution. A value of 1.0 samples exactly in proportion to the frequencies, 0.0 samples all words equally, while a negative value samples low-frequency words more than high-frequency words. The popular default value of 0.75 was chosen by the original Word2Vec paper. More recently, in https://arxiv.org/abs/1804.04212, Caselles-Dupré, Lesaint, & Royo-Letelier suggest that other values may perform better for recommendation applications.
- **dm_mean** (*{1,0}, optional*) – If 0, use the sum of the context word vectors. If 1, use the mean. Only applies when *dm* is used in non-concatenative mode.

- **dm_concat** (*{1,0}, optional*) – If 1, use concatenation of context vectors rather than sum/average; Note concatenation results in a much-larger model, as the input is no longer the size of one (sampled or arithmetically combined) word vector, but the size of the tag(s) and all words in the context strung together.
- **dm_tag_count** (*int, optional*) – Expected constant number of document tags per document, when using dm_concat mode.
- **dbow_words** (*{1,0}, optional*) – If set to 1 trains word-vectors (in skip-gram fashion) simultaneous with DBOW doc-vector training; If 0, only trains doc-vectors (faster).
- **trim_rule** (*function, optional*) – Vocabulary trimming rule, specifies whether certain words should remain in the vocabulary, be trimmed away, or handled using the default (discard if word count < min_count). Can be None (min_count will be used, look to `keep_vocab_item()`), or a callable that accepts parameters (word, count, min_count) and returns either `gensim.utils.RULE_DISCARD`, `gensim.utils.RULE_KEEP` or `gensim.utils.RULE_DEFAULT`. The rule, if given, is only used to prune vocabulary during current method call and is not stored as part of the model.

  The input parameters are of the following types:

  - *word* (str) - the word we are examining
  - *count* (int) - the word's frequency count in the corpus
  - *min_count* (int) - the minimum count threshold.

- **callbacks** – List of callbacks that need to be executed/run at specific stages during training.
- **shrink_windows** (*bool, optional*) – New in 4.1. Experimental. If True, the effective window size is uniformly sampled from [1, *window*] for each target word during training, to match the original word2vec algorithm's approximate weighting of context words by distance. Otherwise, the effective window size is always fixed to *window* words to either side.
- **following** (*Some important internal attributes are the*) –

wv

This object essentially contains the mapping between words and embeddings. After training, it can be used directly to query those embeddings in various ways. See the module level docstring for examples.

Type: `KeyedVectors`

dv

This object contains the paragraph vectors learned from the training data. There will be one such vector for each unique document tag supplied during training. They may be individually accessed using the tag as an indexed-access key. For example, if one of the training documents used a tag of 'doc003':

```
>>> model.dv['doc003']
```

Type: `KeyedVectors`

__getitem__(*tag*)

Get the vector representation of (possibly multi-term) tag.

Parameters: **tag** (*{str, int, list of str, list of int}*) – The tag (or tags) to be looked up in the model.

Returns: The vector representations of each tag as a matrix (will be 1D if *tag* was a single tag)

Return type: np.ndarray

add_lifecycle_event(*event_name*, *log_level=20*, ***event*)

Append an event into the *lifecycle_events* attribute of this object, and also optionally log the event at *log_level*.

Events are important moments during the object's life, such as "model created", "model saved", "model loaded", etc.

The *lifecycle_events* attribute is persisted across object's `save()` and `load()` operations. It has no impact on the use of the model, but is useful during debugging and support.

Set *self.lifecycle_events = None* to disable this behaviour. Calls to *add_lifecycle_event()* will not record events into *self.lifecycle_events* then.

Parameters:

- **event_name** (*str*) – Name of the event. Can be any label, e.g. “created”, “stored” etc.
- **event** (*dict*) –
  Key-value mapping to append to *self.lifecycle_events*. Should be JSON-serializable, so keep it simple. Can be empty.

  This method will automatically add the following key-values to *event*, so you don’t have to specify them:
  - *datetime*: the current date & time
  - *gensim*: the current Gensim version
  - *python*: the current Python version
  - *platform*: the current platform
  - *event*: the name of this event
- **log_level** (*int*) – Also log the complete event dict, at the specified log level. Set to False to not log at all.

add_null_word()

build_vocab(*corpus_iterable=None*, *corpus_file=None*, *update=False*, *progress_per=10000*, *keep_raw_vocab=False*, *trim_rule=None*, ***kwargs*)

Build vocabulary from a sequence of documents (can be a once-only generator stream).

Parameters:

- **documents** (iterable of list of `TaggedDocument`, optional) – Can be simply a list of `TaggedDocument` elements, but for larger corpora, consider an iterable that streams the documents directly from disk/network. See `TaggedBrownCorpus` or `TaggedLineDocument`
- **corpus_file** (*str, optional*) – Path to a corpus file in `LineSentence` format. You may use this argument instead of *documents* to get performance boost. Only one of *documents* or *corpus_file* arguments need to be passed (not both of them). Documents’ tags are assigned automatically and are equal to a line number, as in `TaggedLineDocument`.
- **update** (*bool*) – If true, the new words in *documents* will be added to model’s vocab.
- **progress_per** (*int*) – Indicates how many words to process before showing/updating the progress.
- **keep_raw_vocab** (*bool*) – If not true, delete the raw vocabulary after the scaling is done and free up RAM.

- **trim_rule** (*function, optional*) –
  Vocabulary trimming rule, specifies whether certain words should remain in the vocabulary, be trimmed away, or handled using the default (discard if word count < min_count). Can be None (min_count will be used, look to `keep_vocab_item()`), or a callable that accepts parameters (word, count, min_count) and returns either `gensim.utils.RULE_DISCARD`, `gensim.utils.RULE_KEEP` or `gensim.utils.RULE_DEFAULT`. The rule, if given, is only used to prune vocabulary during current method call and is not stored as part of the model.

  The input parameters are of the following types:

  - *word* (str) - the word we are examining
  - *count* (int) - the word's frequency count in the corpus
  - *min_count* (int) - the minimum count threshold.
- ****kwargs** – Additional key word arguments passed to the internal vocabulary construction.

build_vocab_from_freq(*word_freq, keep_raw_vocab=False, corpus_count=None, trim_rule=None, update=False*)

Build vocabulary from a dictionary of word frequencies.

Build model vocabulary from a passed dictionary that contains a (word -> word count) mapping. Words must be of type unicode strings.

Parameters:

- **word_freq** (*dict of (str, int)*) – Word <-> count mapping.
- **keep_raw_vocab** (*bool, optional*) – If not true, delete the raw vocabulary after the scaling is done and free up RAM.
- **corpus_count** (*int, optional*) – Even if no corpus is provided, this argument can set corpus_count explicitly.
- **trim_rule** (*function, optional*) –
  Vocabulary trimming rule, specifies whether certain words should remain in the vocabulary, be trimmed away, or handled using the default (discard if word count < min_count). Can be None (min_count will be used, look to `keep_vocab_item()`), or a callable that accepts parameters (word, count, min_count) and returns either `gensim.utils.RULE_DISCARD`, `gensim.utils.RULE_KEEP` or `gensim.utils.RULE_DEFAULT`. The rule, if given, is only used to prune vocabulary during `build_vocab()` and is not stored as part of the model.

The input parameters are of the following types:

- *word* (str) - the word we are examining
- *count* (int) - the word's frequency count in the corpus
- *min_count* (int) - the minimum count threshold.

- **update** (*bool, optional*) – If true, the new provided words in *word_freq* dict will be added to model's vocab.

create_binary_tree()

Create a binary Huffman tree using stored vocabulary word counts. Frequent words will have shorter binary codes. Called internally from `build_vocab()`.

property **dbow**

Indicates whether 'distributed bag of words' (PV-DBOW) will be used, else 'distributed memory' (PV-DM) is used.

property **dm**

Indicates whether 'distributed memory' (PV-DM) will be used, else 'distributed bag of words' (PV-DBOW) is used.

property **docvecs**

estimate_memory(*vocab_size=None, report=None*)

Estimate required memory for a model using current settings.

Parameters:

- **vocab_size** (*int, optional*) – Number of raw words in the vocabulary.
- **report** (*dict of (str, int), optional*) – A dictionary from string representations of the **specific** model's memory consuming members to their size in bytes.

Returns: A dictionary from string representations of the model's memory consuming members to their size in bytes. Includes members from the base classes as well as weights and tag lookup memory estimation specific to the class.

Return type: dict of (str, int), optional

estimated_lookup_memory()

Get estimated memory for tag lookup, 0 if using pure int tags.

Returns: The estimated RAM required to look up a tag in bytes.

Return type: int

get_latest_training_loss()

Get current value of the training loss.

Returns: Current training loss.

Return type: float

infer_vector(*doc_words*, *alpha=None*, *min_alpha=None*, *epochs=None*)

Infer a vector for given post-bulk training document.

Notes

Subsequent calls to this function may infer different representations for the same document. For a more stable representation, increase the number of epochs to assert a stricter convergence.

Parameters:

- **doc_words** (*list of str*) – A document for which the vector representation will be inferred.
- **alpha** (*float, optional*) – The initial learning rate. If unspecified, value from model initialization will be reused.
- **min_alpha** (*float, optional*) – Learning rate will linearly drop to *min_alpha* over all inference epochs. If unspecified, value from model initialization will be reused.
- **epochs** (*int, optional*) – Number of times to train the new document. Larger values take more time, but may improve quality and run-to-run stability of inferred vectors. If unspecified, the *epochs* value from model initialization will be reused.

Returns: The inferred paragraph vector for the new document.

Return type: np.ndarray

`init_sims(replace=False)`

Precompute L2-normalized vectors. Obsoleted.

If you need a single unit-normalized vector for some key, call `get_vector()` instead: `doc2vec_model.dv.get_vector(key, norm=True)`.

To refresh norms after you performed some atypical out-of-band vector tampering, call *:meth:`~gensim.models.keyedvectors.KeyedVectors.fill_norms()* instead.

Parameters: **replace** (*bool*) – If True, forget the original trained vectors and only keep the normalized ones. You lose information if you do this.

Returns: The inferred paragraph vector for the new document.

Return type: np.ndarray

`init_sims(replace=False)`

Precompute L2-normalized vectors. Obsoleted.

If you need a single unit-normalized vector for some key, call `get_vector()` instead: `doc2vec_model.dv.get_vector(key, norm=True)`.

To refresh norms after you performed some atypical out-of-band vector tampering, call *:meth:`~gensim.models.keyedvectors.KeyedVectors.fill_norms()* instead.

Parameters: **replace** (*bool*) – If True, forget the original trained vectors and only keep the normalized ones. You lose information if you do this.

Figure 18: 'Doc2vec paragraph embeddings' API reference (source: https://radimrehurek.com/gensim/models/doc2vec.html)

APPENDIX C

Multi-class Confusion Matrix for Model 1

Table 20: Multi-class confusion matrix for Model 1.

		Predicted					
		Phil	Anth	Ling	Neuro	Comp	Psych
Observed	Phil	4574	541	125	14	77	69
	Anth	590	4097	114	238	119	242
	Ling	452	260	4390	12	228	58
	Neuro	13	61	48	4735	118	425
	Comp	81	43	81	10	5158	27
	Psych	381	234	310	454	383	3638

*Note***:** Philosophy; Anth: Anthropology; Ling: Linguistics; Neuro: Neuroscience; Comp: Computer Science; Psych: Psychology

APPENDIX D

Cognitive Science Classical Article Analysis

Table 21: Analysis of randomly selected 20 classical articles in the field of cognitive science.

Article	Phil	Anth	Ling	Neuro	Comp	Psych
Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language, 38(4), 419–439	0,002	0,000	0,491	0,373	0,446	0,477
Baayen, R.H. (2011). Corpus linguistics and naive discriminative learning. Submitted to Brazilian Journal of Applied Linguistics	0,195	0,000	0,677	0,141	0,675	0,425
Barrington, L., Marks, T.K., Hsiao, J.H.-W., & Cottrell, G.W. (2008). NIMBLE: A kernel density model of saccade-based visual memory. Journal of Vision, 8(14):17, 1-14	0,167	0,000	0,026	0,274	0,785	0,396
Botvinick, M., & Plaut, D.C. (2004). Doing Without Schema Hierarchies: A Recurrent Connectionist Approach to Normal and Impaired Routine Sequential Action. Psychological Review, 111(2), 395-429	0,428	0,078	0,298	0,339	0,753	0,306
Brown, G.D.A., Neath, I., & Chater, N. (2007). A Temporal Ratio Model of Memory. Psychological Review, 114(3), 539-576	0,299	0,000	0,121	0,449	0,715	0,463

Christiansen, M.H., & Chater, N. (2001). Connectionist Psycholinguistics: Capturing the Empirical Data. Trends in Cognitive Sciences, 5(2), 82-88	0,023	0,000	0,763	0,000	0,522	0,353
Christiansen, M.H. & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. Cognitive Science, 23, 157-205	0,099	0,000	0,666	0,038	0,520	0,293
Cleeremans, A., & McClelland, J.L. (1991). Learning the structure of event sequences. Journal of Experimental Psychology: General, 120, 235-253	0,120	0,000	0,565	0,265	0,511	0,379
Cowell, R.A., Bussey, T.J., & Saksida, L.M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. Journal of Neuroscience, 26(47), 12186-12197	0,000	0,000	0,247	0,462	0,621	0,280
Criss, A.H., & McClelland, J.L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). Journal of Memory and Language, 55, 447-460	0,213	0,000	0,318	0,490	0,754	0,465
Tenenbaum, J.B., Griffiths, T.L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10(7), 309–318	0,657	0,000	0,470	0,000	0,531	0,442
St. Clair, M.C., Monaghan, P., & Christiansen, M.H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. Cognition, 116, 341-360	0,000	0,000	0,717	0,000	0,012	0,119
Shi, L., Griffiths, T.L., Feldman, N.H, & Sanborn, A.N. (2010). Exemplar models as a mechanism for performing Bayesian inference. Psychonomic Bulletin & Review, 17(4), 443-464	0,386	0,000	0,296	0,365	0,757	0,544
Plaut, D.C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. Cognitive Neuropsychology, 10(5), 377-500	0,144	0,000	0,347	0,306	0,706	0,184
Monaghan, P., Christiansen, M.H., & Fitneva, S.A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. Journal of Experimental Psychology: General	0,321	0,000	0,678	0,232	0,557	0,438

Oaksford, M., & Chater, N. (2009). Précis of bayesian rationality: The probabilistic approach to human reasoning. Behavioral and Brain Sciences, 32(1), 69-120	0,807	0,048	0,273	0,008	0,401	0,227
Gao, J., Tortell, R., & McClelland, J.L. (2011). Dynamic Integration of Reward and Stimulus Information in Perceptual Decision-Making. PloS One, 6(3), 1-21	0,000	0,000	0,266	0,194	0,271	0,463
Hutchins, E. (1995). How a cockpit remembers its speeds. Cognitive science, 19(3), 265-288.	0,191	0,000	0,540	0,183	0,723	0,104
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and brain sciences, 28(5), 675-691	0,774	0,533	0,296	0,101	0,381	0,473

CURRICULUM VITAE

PERSONAL INFO

Name : Oğuzhan Alaşehir

WORK EXPERIENCE:

July 2022 – Ongoing Deloitte, Berlin (Germany)

Position: ***Senior Consultant in Risk Advisory*** (July 2022 – Ongoing)

Duty: Supporting customers in Software Asset Management (SAM) related activities.

Details: I am working as a Senior Consultant in Risk Advisory. I am supporting customers in SAM related activities. The customers are operating in various industries. My role includes but is not limited to fixing technical problems in SAM environment, revising / creating data source interfaces, creating ELP reports, following tasks of stakeholders.

June 2013 – July 2022 Turk Telekom, Ankara (Turkey)

Position: ***Software Asset Manager*** (September 2017 – July 2022)

Duty: Management of Software Asset Management (SAM) Project, License & Demand Optimization, Software Vendor and Contract Management. SAM process implementation.

Details: My role was Software Asset Manager in Turk Telekom which is one of the biggest ICT Company in Turkey. I was the project leader of software asset management project that includes implementation of a SAM tool (Aspera) in a complex environment (millions of installations on 50K+ devices) and establishment of end-to-end lifecycle practices and processes based on ITIL and ISO 19770. After accomplishing the implementation of the project, we were managing products of 200+ vendors in terms of license and software perspectives. I presented this successful project in Europe's best-known conference named SAMSEUROPE in 2019. Our project was awarded as the best project in 2020.

Position: ***Project Manager*** (June 2013 – September 2017)

Duty: Project Management throughout the entire project / product development lifecycle within pre-defined scope, budget and time by applying PMI methodology. Experienced in working under pressure, managing varying size of teams in complex environment, overcoming cultural differences. Some examples of projects: Appointment Organization for nDSL, Signature via Tablet PC, Document Archive System, Workforce Management System Log Management, 5651 Dedicated Firewall & UTM, FTTH Process Revision, Virtual PBX IVR Establishment, Single Sign On Revision Project, TTES (Turk Telecom Retail Electricity Sales) Project.

December 2007 - May 2013 METU Informatics Institute, Ankara (Turkey)

Position: ***Researcher***

Duty: Project Leader of Turkey's first university ranking system URAP - University Ranking by Academic Performance (Establishing University Ranking Methodology, Data Collection by Automation Tools, Unreliable Data Detection and Cleaning (Algorithms developed), Scoring Method Development, and Network Analysis for Citation Traffic).

Instructor of Information Technologies and Applications (IS100) Course (Online Education Coordination and Question Bank Management)

Member of TUBITAK 1001 Project named Factors Influencing Users' Adoption of Technology (Questionnaire Design, Data Collection by Online Questionnaire from more than 2.000 pharmacists, Data Cleaning, Statistical Analysis (by using SPSS software), Structural Equation Modeling (by using LISREL software))

July 2006 - December 2007 Cozum Software, Ankara (Turkey)

Position: ***Network Product Manager***

Duty: Responsible for managing network products (router, switch, wireless devices, IP cameras and IP surveillance systems) including coordination, purchasing, importing, sales and marketing guidance.

EDUCATION

PhD:

Ongoing	Information Systems Department, Graduate School of Informatics Institute, METU

MS:

2007 - 2010	Information Systems Department, Graduate School of Informatics Institute, METU

BS:

2001 – 2006	Computer Education & Instructional Technology Department, Edu. Faculty**,** METU

HONORS, SCHOLARSHIPS AND AWARDS

METU Thesis of the Year Award: Given by Prof. Dr. Mustafa N. Parlar Education and Research Foundation in 2010 for Master Thesis "*University Ranking by Academic Performance: A Scientometrics Study for Ranking World Universities*".

Project Scholarship: Given by Turkish National Scientific and Research Council (TUBITAK) between March 2010 - September 2011 for TUBITAK 1001 SOBAG Project named "Factors Influencing Users' Adoption of Technology: Empirical Investigations in Different Contexts using Structural Equation Modeling Approach".

CERTIFICATES & TRAINING

AWS: AWS Certified Cloud Practitioner, by AWS, December 2022

PMP: Project Management Professional Certificate, by PMI, PMP #1745453, August 2014

AGILE:

Professional Scrum Product Owner I (PSPO I), Scrum.org, April 2017

Professional Scrum Master I (PSM I), Scrum.org, October 2016

ITIL: ITIL Service Management and Foundations Education, by Noventum, March 2014

CMMI: Introduction to CMMI for Development, by SEI, June 2013

REQUIREMENT ENGINEERING: by N. Alpay Karagöz, Proven, June 2015

TOGAF: TOGAF Foundation, by Serkan Türkeli, TESODEV, February 2018

PYHTON: Programming with Python, by Kenan Kaan Kurt, TESODEV, July 2018

LANGUAGE:

Language	Level
Turkish	**Native**
English	**Professional working proficiency**

JOURNAL PUBLICATIONS

Alasehir, O., & Acarturk, C. (2022). Interdisciplinarity in Cognitive Science: A Document Similarity Analysis. Cognitive Science, 46(12), e13222.

Çakır P M., Acartürk C, Alaşehir, O., & Çilingir C (2015). A comparative analysis of global and national university ranking systems. Scientometrics, Volume 103, Issue 3, pp 813-848

Alaşehir, O., Çakır P M., Acartürk C, Baykal N, & Akbulut U (2014). URAP-TR: a national ranking for Turkish universities based on academic performance. Scientometrics, Volume 101, Issue 1, pp 159-178