

ALGERIAN DEMOCRATIC AND POPULAR REPUBLIC MINISTRY  
HIGHER EDUCATION AND SCIENTIFIC RESEARCH



ECHAHID HAMMA LAKHDAR UNIVERSITY OF EL-OUED



FACULTY OF: EXACT SCIENCES

Computer Science department

End of Study Thesis Presented for obtaining the Diploma of

## ACADEMIC MASTER

**Field:** Mathematics and Computer Science

**Sector:** Computer science

**Specialty:** Distributed Systems and Artificial Intelligence

### Theme:

Computational Processing for Concept Network Analysis in  
Educational Programs - A Comparative Study between  
Arabic and Foreign Programs

**Supervisor:** Mohammed Charaf Eddine Meftah

**Presented by:**

- Lalmi Sadia
- Boukhalfa Nouha Yasmine

**Supported on: //2024. Before the jury :**

Retima Farida .....President

Berdjough Chafik..... Examiner

**Academic Year: 2023/2024**

# REMERCIEMENT

*Praise be to God who enlightened us on the path of science and knowledge He helped us to perform this duty, and enabled us to complete this work, and may blessings and peace be upon his Prophet. Above all, we thank God Almighty for granting us success from the first stage of our academic journey until now. We now extend our sincere thanks, gratitude and appreciation to Supervising Professor Mohammed charaf Eddine Meftah for his good supervision and guidance. We are very proud of him as the professor responsible for this report, and of everyone who contributed to helping us from far or near, even with a small idea. We also extend our sincere thanks to the members of the discussion committee who devoted their time to reading and discussing this thesis, and we do not forget to extend our sincere thanks to our esteemed professors who participated with us in our academic journey during our university career, and we do not forget all our colleagues and loved ones without exception*

# Dédicace

*I present the fruit of my humble efforts To the one who, thanks to the purity of his heart and his continuous struggle, was able to establish the foundation of our home. The one who taught me what I did not learn from books, schools, and university. Giving, self-confidence, and patience. My dear father, may God have mercy on him. To the one who planted hope in my heart and raised me with a passion for learning and knowledge. Who stayed up all nights, illuminating my path. And the one with the giving hand to my beloved mother. To those whom God gave me the blessing of their presence in my life, to the solid contract Those who helped me in my research journey : my brothers and sisters.*

*To those who supported me as we paved the way together towards success in our scientific journey, to my companion Boukhalifa, Noha Yasmine.*

*Finally, to everyone who helped me and had a role, from near or far, in completing this study*

*Asking the Lord to reward everyone with the best reward in this world and the hereafter. Then to every seeker of knowledge who strives with his knowledge to benefit Islam and Muslims With all the knowledge and knowledge God has given him*

*sadia*

# *Dédicace*

*To God above all, praise be to God for you as it should be for your face and your  
great authority*

*Whoever said, "I got it." The journey was not short, nor should it be. The dream  
was not close, nor was the path fraught with ease, but I did it and got it.*

*To the wall on which I lean in my fatigue and sadness, to the shoulder on which  
I put my weight, to my dear and beloved, whom I love as much as this world, to  
the paradise of this world and the hereafter. my dad*

*To my pure angel and my strength after God. My first and eternal supporter, to  
the one who supported me without limits and gave me for nothing in return, I  
dedicate to you this achievement that without your sacrifices would not have  
existed. My mom*

*To whom he said about them (We will support you through your brother)*

*To the one who stretches out his hand tirelessly in times of my weakness, my  
brothers, may God keep you as a steadfast side*

*To the comrades of the first step and the penultimate step, to those who during  
the lean years were rain clouds, I am grateful.*

*To my scientific friend Sadia.*

*And to every hand I once shook by heart*

*nouha*

# Résumé

*Le traitement informatisé du langage et du texte (PN(L)) a confirmé la capacité de cette approche à analyser des textes et à extraire des termes et des concepts de base pour comprendre le message cognitif du texte. L'étude vise à identifier les termes les plus importants et les relations sémantiques entre eux dans les programmes éducatifs arabes et européens, et à utiliser la technique d'analyse des réseaux de textes pour déterminer le niveau de présence de ces termes et la nature des relations entre eux, avec la objectif d'extraire le message cognitif ciblé de ces programmes. Dans cette étude, nous soulèverons les problèmes suivants dans le domaine du traitement automatique des langues naturelles : Quels sont les termes les plus importants qui habitent les programmes éducatifs arabes et européens ? Quelle est la nature des relations sémantiques entre eux ? Par conséquent, quels sont les concepts les plus importants qui y sont déposés ? Quel est le message cognitif visé par ces programmes éducatifs ?*

*Mots clés : Traitement informatisé, texte, programme éducatif, terme, concept, relation sémantique .*

# *Abstract*

*Computerized language and text processing (NLP) has confirmed the ability of this approach to analyze texts and extract basic terms and concepts to understand the cognitive message of the text. The study aims to identify the most important terms and the semantic relationships between them in Arab and European educational programs, and to use the text network analysis technique to determine the level of presence of these terms and the nature of the relationships between them, with the aim of extracting the targeted cognitive message from these programs. In this study, we will raise the following problems within the field of automatic processing of natural languages : What are the most important terms that inhabit Arab and European educational programs ? What is the nature of the semantic relationships between them ? Therefore, what are the most important concepts deposited in them ? What is the intended cognitive message of these educational programs ?*

*key words : Computerized processing, text, educational program, term, concept, semantic relationship.*

## المخلص

أكدت المعالجة المحوسبة للغة والنص ( بع ) قدرة هذا النهج في تحليل النصوص واستخلاص المصطلحات والمفاهيم الأساسية لفهم الرسالة المعرفية للنص. تهدف الدراسة إلى تحديد أهم المصطلحات والعلاقات الدلالية بينها في البرامج التربوية العربية والأوروبية، واستخدام تقنية تحليل شبكة النص لتحديد مستوى حضور هذه المصطلحات وطبيعة العلاقات بينها، بهدف استخلاص الرسالة المعرفية المستهدفة من هذه البرامج. في هذه الدراسة سنطرح الإشكاليات التالية ضمن مجال المعالجة الآلية للغات الطبيعية : ماهي أهم المصطلحات التي تستوطن البرامج التربوية العربية والأوروبية؟ وماهي طبيعة العلاقات الدلالية بينها؟ وبالتالي ماهي أهم المفاهيم المودعة فيهم؟ وماهي الرسالة المعرفية المستهدفة من هذه البرامج التربوية؟

الكلمات المفتاحية : معالجة محوسبة، النص، برنامج تربوي، مصطلح، مفهوم، علاقة دلالية

---

# TABLE DES MATIÈRES

<b>Contents</b>	<b>i</b>
<b>List of Figure</b>	<b>vi</b>
	<b>1</b>
<b>1 General definitions of automated natural language processing and data mining</b>	<b>2</b>
1.1 Introduction . . . . .	2
1.2 Definition of computational linguistics . . . . .	2
1.3 Topic Computational Linguistics . . . . .	3
1.4 Computational linguistics approach . . . . .	3
1.5 Fields of computational linguistics . . . . .	3
1.6 Objectives of computational linguistics . . . . .	4
1.6.1 Simulating human performance . . . . .	5
1.6.2 Modeling (simulating human thinking) . . . . .	5
1.6.3 Insight into the linguistic processes that occur within the human mind . . . . .	5
1.7 Methodologies used in linguistic processing . . . . .	6
1.8 Natural language processing task paths . . . . .	6
1.8.1 Tokenization . . . . .	6
1.8.2 Sentence splitting . . . . .	7
1.8.3 Part-of-speech tagging . . . . .	7
1.8.4 Morphological analysis . . . . .	8

1.8.5	Grammatical analysis	8
1.8.6	Text segmentation	9
1.9	Non-digital techniques	11
1.9.1	Mechanical processing	11
1.9.2	Electromechanical processing	12
1.9.3	Analog processing	12
1.10	Digital techniques	13
1.10.1	Computational processing	13
1.11	Automated language processing applications	14
1.11.1	Understanding linguistic context	14
1.11.2	Linguistic statistics	14
1.11.3	Writing processing	14
1.11.4	Morphology or word structure	15
1.11.5	Sounds and audio systems	15
1.11.6	Grammar and systems	15
1.11.7	Developing linguistic dictionaries	15
1.12	DATA MINING	16
1.13	Conclusion	17
<b>2</b>	<b>Comparison system design</b>	<b>18</b>
2.1	Introduction	18
2.2	Example 1 : The book Semantic Analysis and Graphical Description of Ibn Khaldoun Introduction.	18
2.2.1	Computerized processing of Ibn Khaldoun introduction	18
2.2.2	Summary and conclusions	22
2.3	Example 2 : A computerized processing book to analyze the network of meanings and concepts in books on the expression of vision	23
2.3.1	Summary and conclusions	27
2.4	Exemple 3 : Review of Research on Digital Supply Chain Management Using Network Text Analysis	28
2.4.1	Subject Analysis	28
2.4.2	Network Text Analysis on Articles Published until 2019	30
2.4.3	Network Text Analysis on Articles Published from 2020	32
2.4.4	Conclusions	37

2.5	example 4 :paths study on knowledge convergence and development in computational social science : data metric Analysis Based onWeb of Science . . . . .	38
2.5.1	Data Sources and Research Methods . . . . .	38
2.5.2	Analysis of Evolutionary Dynamics . . . . .	39
2.6	example 5 : Infranodus : generate insight with ai and network thinking . . . . .	48
2.6.1	EXPLORING AN AI TOOL . . . . .	48
2.7	Conclusion . . . . .	54
<b>3</b>	<b>Explaining the work and presenting the results</b>	<b>55</b>
3.1	Itroduction . . . . .	55
3.2	Plan for workflow steps . . . . .	55
3.3	Computer study of textbooks (statistics for comparison) . . . . .	56
3.4	A comparison between school books (reading books) for the primary stage in Qatar and America . . . . .	57
3.4.1	Computerized processing of the books of the State of Qatar . . . . .	57
3.4.2	Analysis of the books network for the State of Qatar . . . . .	59
3.4.3	Computer processing of American state books (English time) . . . . .	60
3.4.4	Analysis of the books network for the State of American (English time) : . . . . .	61
3.5	comparison between school books (reading books) for the primary stage in Egypt and America : . . . . .	63
3.5.1	Analysis of the books network for the State of Egypt . . . . .	64
3.5.2	Computer processing of American state books (Every body) . . . . .	66
3.5.3	Analysis of the books network for the State of American (Every body) : . . . . .	67
3.6	A comparison between textbooks (reading books) for the primary stage in Arab countries and foreign countries : . . . . .	69
3.6.1	Forming a word cloud for arab books : . . . . .	69
3.6.2	Analysis of the books network for the Arab countries : . . . . .	70
3.6.3	Forming a word cloud for foreign books . . . . .	72
3.6.4	Analysis of the books network for the foreign countries . . . . .	73
3.7	results . . . . .	75
3.8	Conclusion . . . . .	75
<b>4</b>	<b>Environment and offices used for development</b>	<b>76</b>
4.1	Introduction . . . . .	76

4.2 Python . . . . .	76
4.3 NLTK . . . . .	77
4.4 Matplotlib . . . . .	78
4.5 Word Cloud . . . . .	78
4.6 Networkx . . . . .	79
4.7 Arabic reshaper . . . . .	79
4.8 Bidi algorithm . . . . .	80
4.9 pycharm . . . . .	81
4.10 Some parts of the codes used . . . . .	82
4.11 Conclusion . . . . .	84

<b>Bibliographie</b>	<b>86</b>
----------------------	-----------

---

## TABLE DES FIGURES

1.1	Simple fragmentation of a nominal phrase - The old man bought a hat[1]. . . . .	10
1.2	Compound segmentation of a nominal phrasal verb that does not include prepositional and prepositional phrasal verbs - The Old Man[1]. . . . .	10
1.3	Compound segmentation of a nominal phrasal verb that includes prepositional and prepositional phrasal verbs[1]. . . . .	10
1.4	Compound segmentation of phrasal verbs[1]. . . . .	11
1.5	Architecture of a typical Data Mining System [2]. . . . .	16
1.6	Map of Data Mining domains[3]. . . . .	17
2.1	Vocabulary cloud of Ibn Khaldoun introduction before the purification process was carried out from linguistic impurities[4]. . . . .	19
2.2	Vocabulary cloud of Ibn Khaldoun introduction after practicing purification processes from linguistic impurities[4]. . . . .	19
2.3	The most common vocabulary words in the vocabulary clusters of the Introduction book[4]. . . . .	20
2.4	Table The most important characteristics of the selected text network fabric[4].	21
2.5	The grid for the content of Ibn Khaldoun introduction[4]. . . . .	22
2.6	Word cloud of the chosen book of dream interpretation before and after the treatments[5]. . . . .	23
2.7	The word cloud of the book Tatir al-Anam in the expression of vision and dream Before and after treatment[5]. . . . .	24
2.8	The network structure of the text material of the selected book from Interpretation of Dreams before and after treatment[5]. . . . .	25

2.9	The network texture of the text material of the book Tatir al-Anam in the expression of vision and dream before and after processing[5]. . . . .	25
2.10	Network of concepts of expressions of the viewer’s story of his vision[5]. . . . .	26
2.11	The conceptual network of the meaning of the vision of a man, a woman, a child, and money and its reflection on its expression according to Imam Ibn Sirin [5]. . . . .	26
2.12	Articles by year [6]. . . . .	28
2.13	Table Major journals publishing digital supply chain articles[7]. . . . .	29
2.14	Major authors publishing digital supply chain management articles [7]. . . . .	29
2.15	Word cloud (until 2019) [6]. . . . .	30
2.16	Keyword network (until 2019) [7]. . . . .	31
2.17	Table Degree centrality analysis (until 2019)[7]. . . . .	32
2.18	Table Betweenness centrality analysis (until 2019)[7]. . . . .	32
2.19	Word cloud [7]. . . . .	33
2.20	Keyword network (from 2020)[7]. . . . .	33
2.21	Table Degree centrality analysis (from 2020)[7]. . . . .	34
2.22	Table Betweenness centrality analysis (from 2020)[7]. . . . .	35
2.23	Cluster analysis (from 2020) [7]. . . . .	36
2.24	Main keywords by cluster (from 2020) [7]. . . . .	37
2.25	Keyword clusters [7]. . . . .	37
2.26	Bibliometric analysis process[8]. . . . .	38
2.27	Academic papers published related to computational social science, 2001–2020 [8].	39
2.28	Table Valid literature and high frequency keywords in different periods[8]. . . . .	40
2.29	Average annual citations in computational social science, 2001–2020[8]. . . . .	40
2.30	Top 20 cited literature [8]. . . . .	41
2.31	Table Top 20 authors with significant impact in computational social science [8].	42
2.32	Collaborative network of scholars related to computational social science [8]. . . . .	43
2.33	Annual number of publications by lead authors [8]. . . . .	44
2.34	Word cloud in computational social science [8]. . . . .	45
2.35	2001–2005 cooccurrence network [8]. . . . .	45
2.36	2006–2010 cooccurrence network [8]. . . . .	46
2.37	2011–2015 cooccurrence network [8]. . . . .	46
2.38	2016–2020 cooccurrence network [8]. . . . .	47
2.39	Keywords cooccurrence network[8]. . . . .	48
2.40	Abbott, D. (2019). Game-Based Learning for Postgraduates[9]. . . . .	49

2.41	Text analysis network based on Abbott’s (2019) article created by the AI[9]. . .	49
2.42	Question generated by InfraNodus based on Abbott’s (2019)[9]. . . . .	50
2.43	Additional questions generated [9]. . . . .	51
2.44	The topical network based on fifteen abstracts[9]. . . . .	52
2.45	Most influential concepts offered by the tool through "Reveal underlying ideas" feature. Previously this feature was called "Reveal Non-obvious" [9]. . . . .	52
2.46	The topical network based on fifteen full papers. Note that ‘ct’ and ‘item’ are identified as ‘nodes’[9]. . . . .	53
3.1	workflow steps . . . . .	56
3.2	Statistics table . . . . .	57
3.3	Word cloud before processing. . . . .	58
3.4	Post-processed word cloud. . . . .	58
3.5	A table representing the frequency of the words that appear most frequently in Qatar . . . . .	59
3.6	Text content grid . . . . .	60
3.7	Word cloud before the cleaning process . . . . .	60
3.8	Word cloud after the cleaning process . . . . .	61
3.9	A table representing the frequency of the words that appear most frequently in American books (English time) . . . . .	62
3.10	Word network . . . . .	63
3.11	Word cloud before filtering . . . . .	63
3.12	Word cloud after filtering . . . . .	64
3.13	A table representing the frequency of the words that appear most frequently in Egypt . . . . .	65
3.14	Concept network . . . . .	66
3.15	Word cloud before the cleaning process . . . . .	67
3.16	Word cloud after the cleaning process . . . . .	67
3.17	A table representing the frequency of the words that appear most frequently in American (Every body) . . . . .	68
3.18	Concept network . . . . .	68
3.19	Word cloud before the cleaning process . . . . .	69
3.20	Word cloud after the cleaning process . . . . .	70

3.21 A table representing the frequency of the words that appear most frequently in Arab . . . . .	71
3.22 Concept network . . . . .	71
3.23 Word cloud before the cleaning process . . . . .	72
3.24 Word cloud after the cleaning process . . . . .	72
3.25 A table representing the frequency of the words that appear most frequently in foreign . . . . .	73
3.26 Concept network . . . . .	74
4.1 python logo[10]. . . . .	77
4.2 nltk logo[11]. . . . .	78
4.3 matplotlib logo[12]. . . . .	78
4.4 Word Cloud logo [13]. . . . .	79
4.5 networkx logo[14]. . . . .	79
4.6 Arabic reshaper logo [15]. . . . .	80
4.7 Bidi algorithm logo[16]. . . . .	80
4.8 pycharm logo [17]. . . . .	81
4.9 Read a file in English . . . . .	82
4.10 Read a file in arabic . . . . .	82
4.11 List of offices . . . . .	82
4.12 Post-processing word cloud formation . . . . .	83
4.13 Display the post-processed word cloud in addition to word repetition . . . . .	83

---

# GENERAL INTRODUCTION

Computer processing and data analysis are among the most important technologies that contribute to developing the field of education and improving the quality of educational programs. This memorandum deals with the study of applying computer processing to analyze the network of concepts in educational programs, with a focus on making a comparison between Arab and European programs. This research aims to understand how to use modern technologies in the field of data analysis to improve the learning process and develop educational curricula. The methods used in data processing and network analysis will be discussed. In addition to reviewing Arab and European educational programs and evaluating their effectiveness and quality. This memorandum is an attempt to understand the challenges and opportunities in applying computer processing in concept network analysis in the context of educational programs, with a focus on the differences and similarities between Arab and foreign programs and the extent of their impact on the quality of education. We will follow the following structure to achieve our goal :

chapitre 1 :Automated natural language processing and text mining :This chapter presents general definitions of automated natural language processing and data mining for analyzing and understanding human texts automatically.

chapitre 2 :Similar examples support the topic :In this chapter, we will discuss previous cases or examples that support the topic we are addressing, enhancing our understanding of it.

chapitre 3 :Explaining the work and presenting the results :This chapter aims to explain the work, present the results, comparisons, and conclusions that will be discussed in this study.

chapitre 4 :Environment and offices used for development :The work environment and utilized offices are displayed, along with the various steps of code writing.

---

---

# CHAPITRE 1

---

## GENERAL DEFINITIONS OF AUTOMATED NATURAL LANGUAGE PROCESSING AND DATA MINING

### 1.1 Introduction

Attempts to link language with modern technological means have been the birthplace of many linguistic technical branches, including computational linguistics, as well as automated processing of natural languages, data mining, and linguistic engineering. This chapter aims to provide a general overview of these fields and their importance in the world of modern technology. Emphasis will be placed on understanding how computing can be used to analyze and process human language automatically.

### 1.2 Definition of computational linguistics

The term "computational linguistics" is a combination of two words : "linguistics" and "computational". The first is located in the field of humanities, and the second is located in the field of computer science. This interaction indicates that the researcher is in this natural field of knowledge Interdisciplinary study of language in its various aspects using computers. In other words, the research is concerned with the study of language using computational knowledge, with the need to represent natural language in a way that is flexible for application to compu-

ters, as this requires the use of mathematical knowledge and symbols. Thus, the primary goal of computing is to measure linguistic knowledge in human brains to guide how computers work [18].

### 1.3 Topic Computational Linguistics

Computational linguistics is a branch of applied linguistic studies that combines linguistic and information aspects, and is concerned with studying the linguistic phenomena of natural languages in a computational manner. This includes developing and creating computer systems to process natural language at various linguistic levels, such as phoneme, morphology, lexicon, syntax, and semantics. Natural language processing requires formal and logical stages before the computational stage, which relies on mathematical approaches and techniques. Pragmatic aspects cannot be formalized because they relate to culture and external reality rather than linguistic aspects, making them more difficult to process [18].

### 1.4 Computational linguistics approach

The Computational Linguistics approach is based on the intersection of linguistics, mathematics, and information sciences, aiming to develop computer systems for natural language processing. This approach follows basic steps that are as follows [18] :

- Determine the level of the language to be studied
- Representing it mathematically
- Creating an information or computer system to process it.

### 1.5 Fields of computational linguistics

The most important areas of computational linguistics[19] :

- Using computer capabilities in teaching the Arabic language.
- Providing useful programs for its students.
- Developing programs dedicated to the Arabic language to facilitate scientific research.
- Create programs to meet the needs of speakers of other languages.

- Creating electronic dictionaries covering the fields of criticism, rhetoric, style, grammar, and morphology, as well as the linguistic field in general.
- Using computers to translate texts to and from Arabic.
- Spelling, stylistic and linguistic checking.
- The weight of old and new Arabic poetry.
- Defining the types of similes, metaphors, and wonderful colors.
- Adjust and change punctuation marks.
- Automatic formation of texts.
- Supporting speech recognition and automatic synthesis systems.
- Creating rich linguistic dictionaries.
- Identify and analyze speech patterns.
- Generating speech and texts.
- Automatic writing and reading of texts.
- Understand and answer questions.
- Retrieving and extracting information.
- Perform automatic summarization and indexing.
- Text mining.
- Extracting information.
- Perform quick statistics of texts.
- Developing smart education systems.
- Electronic publishing.
- Developing programs to serve the Arabic language or downloading programs from the Internet and modifying them to meet the needs of the Arabic language.

## 1.6 Objectives of computational linguistics

Computational linguists aspire to study language and its computer-mediated representation to achieve a set of goals. They have defined their paths, aspirations, and hopes, and motivated themselves to achieve these goals. As a result, they formulated their approaches, principles, and tools based on these goals. These goals vary among themselves, and include[20].

### 1.6.1 Simulating human performance

They seek to create machines that imitate the performance of humans in understanding and producing language, but they face challenges due to the complexity of the linguistic process and its interaction with social and cultural factors and previous experiences. Therefore, they are working on developing specific models that represent the linguistic and cognitive levels. At the cognitive level, for example, they have created a model that represents the child's knowledge, and their focus is on building computer systems specialized in specific fields such as retrieving or classifying information. As for the linguistic level, they have paid attention to aspects of language such as the lexicon and phonetics. Grammar, semantics, morphology, and linguistic usage[20].

### 1.6.2 Modeling (simulating human thinking)

In an attempt to model the human mind, computational linguists are turning to understanding mental processes to design a model of a machine that works in the same way. This exemplary work reflects the relational model of the human brain in a computing-based model, and focuses specifically on natural language processing. Despite continuous attempts, researchers have been unable to develop an ideal model that mimics this thinking from its cognitive and psychological aspects, due to the complexity and fragmentation of the problem. Human cognition is complex, involving aspects such as cognitive linguistics, psychology, artificial intelligence, and life science, which need to be studied separately to comprehensively understand how human cognition works[20].

### 1.6.3 Insight into the linguistic processes that occur within the human mind

Computational linguistics sought to investigate linguistic ability and understand involuntary processes in the human mind, with the aim of enabling computers to understand, interpret and reproduce natural language, and to bring part of the unconscious into consciousness - language consciousness, which is a prerequisite for achieving machine linguistic consciousness. Achieving machine language awareness requires a computer to perceive language in a context that allows the user to interact in their natural language, which requires accuracy and objectivity from the linguistic researcher. The ultimate goal of computational linguistics is to enable direct human-computer interaction in natural language without an intermediary, while trying to

provide the computer with enough science and knowledge to understand and reproduce natural language[20].

## 1.7 Methodologies used in linguistic processing

There are two main types of approaches to linguistic processing : knowledge-based and [3] learning-based, and they can be combined together. Each methodology has advantages and disadvantages. Knowledge-based methodology relies on hand-written rules, requires language skills, provides easy understanding of results but consumes time in writing and updating the rules. Learning-based methodologies require sufficient training data and may be difficult to adapt to new changes, but they allow systems to be built quickly and accurately. Machine learning can be combined with unsupervised methods to overcome data availability issues[1].

## 1.8 Natural language processing task paths

pre-NLP task paths typically consist of the following elements[1] :

- Tokenization Sentence splitting .
- Part-of-speech tagging .
- Morphological analysis.
- Parsing and chunking .

### 1.8.1 Tokenization

Breaking text into linguistic units, known as “tokens,” is a fundamental step in the field of natural language processing. It involves dividing text into simple units such as words, numbers, and symbols, often separated by white space in English. Complex algorithms such as segmentation of speech typically analyze these units as input, so it is important to use a high-quality linguistic unit segmenter to ensure the accuracy of the results. Common types of linguistic units include numbers, symbols, punctuation marks, and words in their various forms. Programs specialized in segmenting linguistic units usually add properties that describe the unit, such as spelling style and type of unit, which helps improve the quality of the linguistic processing

process. Some tools offer customization to ensure they are compatible with different languages and different types of text [1].

### 1.8.2 Sentence splitting

Sentence marking, or sentence splitting, is the process of dividing text into individual sentences, and includes identifying punctuation such as periods, commas, question marks, abbreviations, and more. Sentence breakers use lists of abbreviations to mark the end of sentences, with a full stop being the end of a sentence except in certain cases, such as Mr. (Mr), or when inside quotation marks. Determining syntax depends on factors such as line breaks in headings or bulleted lists. Cases get more complicated with tables, headings, or formatting tags, and some sentence breakers ignore these elements entirely, while others rely on commas and line breaks. Sentence breakers provide several alternatives for defining boundaries between sentences, and the appropriate method depends on the type of text. Unlike rule-based ANNIE, the OpenNLP sentence segmenter is machine learning-based, and can handle unsegmented texts well, but may cause some errors in defining boundaries between sentences. NLTK uses the Punk sentence segmenter, which relies on letters and ordinals to define boundaries between sentences. The accuracy of sentence segmenters varies depending on the type of text, and performance depends on the correct learning and grammar used. Sentence segmenters are important tools in natural language processing, and each of them has strengths and weaknesses related to text characteristics, which makes choosing the right tool important to ensure the accuracy of the results [1].

### 1.8.3 Part-of-speech tagging

Part of speech (POS) classification involves labeling words to determine their linguistic classification, such as nouns, verbs and adjectives. These basic categories are divided into precise categories that distinguish between singular nouns, plural nouns, and verb tenses, and gender can also be included in the classification. Using a range of possible classifications is crucial and varies depending on the tools used. Common classifications include PTB, Brown's Thesaurus, and LOB's Thesaurus. Determining a segment of speech classification depends on the context in which the word appears as well as the word itself, so classification tools use machine learning methods such as HMMs, Brill classifier, and others. The accuracy of these classifiers depends on the type and quality of text they are trained on, however, accuracy can decrease significantly when dealing with different types of text or noisy data[1].

### 1.8.4 Morphological analysis

Identifying and classifying the linguistic units within a word forms the core of morphological analysis, where the word is broken down into the root, antecedents and suffixes, facilitating the identification of the forms of verbs and nouns. Morphological analysis is usually applied to English for verbs and nouns, where these words appear in different forms resulting from inflectional morphology. Morphology in English depends on the addition of suffixes or internal modifications such as changing vowels. In other languages, antecedents may be used to add at the beginning of a word or suffixes to add at the end of a word, with other internal changes. Morphological analysis tools mainly provide processing of inflectional morphology and follow rule-based methods, allowing them to deal with different inflectional forms based on grammatical rules. In addition, part-of-speech classification cards can be selected for words, which helps in determining the grammatical classification of words. The use of tools such as GATE, Stanford and WordNet is based on linguistic rules and standards, which enable them to deal with morphological analysis effectively, in addition to their ability to process words within the lexical database[1].

#### Derivation of the word stem

Word stem derivation tools produce the stem form of each word, making it easier to identify the root that words share. For example, the words "driving" and "drivers" share the root "drive." In contrast, morphological analysis tends to produce the root forms of words plus their antecedents and/or suffixes, as in the case of "drive" and "driver" of the previous examples, plus the suffixes "ing" and "s" respectively. There is a large discrepancy between stem derivation tools in how they work and the data they produce, as these tools do not attempt to parse the stem or word stem, but rather simply remove the suffix and return the stem. In contrast, morphological analysis tools take into account the context surrounding the word and look at the word as part of the entire text[1].

### 1.8.5 Grammatical analysis

Grammatical analysis means studying the structure of sentences and analyzing them, by extracting their grammatical structure according to known grammatical rules. Syntactic analysis explains how different elements in a sentence relate to each other, such as the connection between the subject and the object in a given sentence. Syntactic theories vary in the field of computational linguistics, as these theories provide different understandings of sentence struc-

tures. Therefore, analysis tools vary in the methods and results they provide, depending on the syntactic theory they follow.

Several parsing tools are freely available covering a wide range of syntactic theories, such as the Minipar dependency parser, the RASP statistical parser, the Stanford statistical parser, and the versatile SUPPLE parser.

These tools are embedded within the GATE working platform, allowing the user to try them out and choose the appropriate tool according to their needs.

The Minipar parser is a dependency parser that determines the dependency relationships between words in a sentence. This analyzer processes the text sentence by sentence, and identifies the linguistic structure and parts of speech, such as the addition, relative clauses, subject and object in verbs, in addition to the determiners and the way to connect elements in the sentence. In contrast, component parsers are based on the concept of component relationships, and include a variety of different syntactic theories.

For example, Shift-Reduce Constituency Parser is a good example of a component parser, as it uses a parsing process optimization algorithm that is based on dynamic code graphs.

This parser aims to improve the performance of traditional component parsers, which relied on graphs to search for the best result among available grammars, which were accurate but slow [1].

### 1.8.6 Text segmentation

Parsing algorithms are often computationally expensive, and are usually at their best when they are trained on texts similar to the ones they are dealing with. Due to the complexity of the text segmentation task, its performance is often less effective compared to low-level processing tasks such as sentence segmentation. This may affect the efficiency of other tasks such as recognizing name entities and analyzing relationships. So, sometimes it may be better to sacrifice accuracy for a lighter, less expensive tool, such as text segmentation tools that perform superficial linguistic analysis, recognizing strings of related words without analyzing their internal structure.

Text segmenters can be divided into noun phrase segmenters and phrasal segmenters. The differences between these types of segmenters are smaller compared to the differences between parsing algorithms, because the parsing process generally takes place at the level of principal components analysis, where sentence segmenters identify relevant parts of the text, but do not seek to analyze those parts. However, there are differences between these tools regarding what

is considered relevant to the piece of text under analysis. For example, a simple noun phrase may contain a succession that includes an optional determiner, one or more optional adjectives or adjectives, as well as one or more nouns, as shown in figure 1.1.

On the other hand, more complex nominal phrases may include, in addition to the above, an adjective or relative clause that modifies the noun phrase. Some text segments include these things as part of the noun phrase, while others do not (Figure 1.3). Decisions about whether to include an adjective or relative clause in a noun sentence depend largely on the intended use of the text fragments later. For example, if text fragments are expected to be used as input for a tool that recognizes terms, one should consider whether or not there are possibilities of a phrase containing a prepositional and prepositional phrase. In the case of generating ontologies, this statement may not be necessary, but it may be useful when used as the goal of the sentiment analysis process[1].

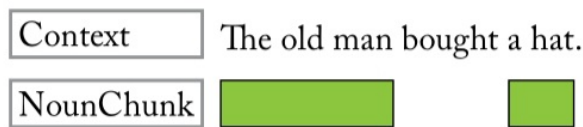


FIGURE 1.1 – Simple fragmentation of a nominal phrase - The old man bought a hat[1].

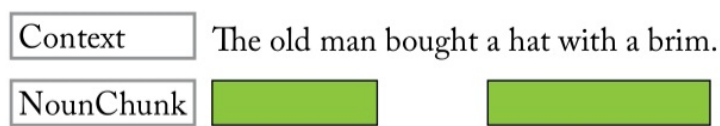


FIGURE 1.2 – Compound segmentation of a nominal phrasal verb that does not include prepositional and prepositional phrasal verbs - The Old Man[1].

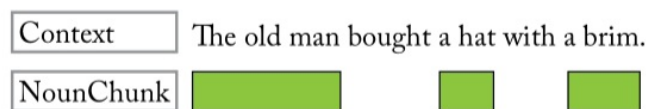


FIGURE 1.3 – Compound segmentation of a nominal phrasal verb that includes prepositional and prepositional phrasal verbs[1].

Phrasal segmenters define boundaries between verbs, where verbs can consist of a single word such as 'bought', or of a more complex group including infinitive verbs, auxiliary modal verbs and so on, such as 'he probably bought' or 'to buy'. It may also include negative items such as "he probably did not buy" or "he did not buy." Figure 1.4 shows an example of one of the results of a sentence segmentation program, which combines the processes of segmenting nominal phrases and segmenting verbs.

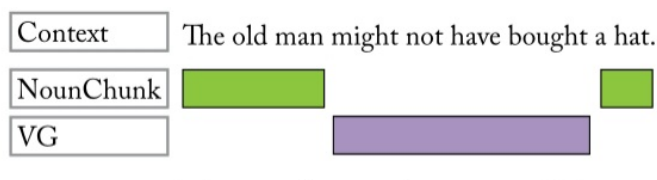


FIGURE 1.4 – Compound segmentation of phrasal verbs[1].

## 1.9 Non-digital techniques

### 1.9.1 Mechanical processing

Mechanical processing generally refers to the system that directly manages the machine and regulates its movement to carry out its tasks automatically and sequentially. This is the basic principle of mechanization and the workings of mechanical machines developed by man in the first wave of the Industrial Revolution at the end of the eighteenth century. Although there were several factors facilitating innovations and inventions in that period, technologies were simple and limited in the first wave of the Industrial Revolution. In fact, mechanical machines at this stage required direct control and operation without an intermediary. At the same time, machines with mechanical processors appeared with the beginning of the Industrial Revolution, with England being a leader in this field, despite French progress in the natural and social sciences.

Several factors contribute to stimulating innovations and inventions in this period, including government support for economic development, capitalist thought, and a culture of mechanized production, along with expanding markets and encouraging migration to industrial cities. Therefore, the political will and the social and economic environment were available to support that revolution. In fact, mechanical processors were not necessarily new at that time, rather they had existed before, but the Industrial Revolution contributed to the improvement and development of these technologies[21].

### 1.9.2 Electromechanical processing

Electromechanical language processors appeared in the nineteenth century with the expansion of the study of electricity and the discovery of its phenomena. Experiments and theories in that period led to tremendous achievements in the field of using electricity. These achievements allowed the development of multiple industries such as motor manufacturing, electroplating, lighting, refrigeration and heating, opening the door to a new wave of the industrial revolution. As scientists became more ambitious, the desire to develop language processors increased with the progress of successive electrical discoveries. The goal was to find a faster means of communication and data exchange that meets the requirements of human society and is an alternative to slow traditional means of communication. Scientists found what they were looking for in electromechanical processors for transmitting and managing linguistic data.

The idea of electromechanical language processors depends on representing language with a certain mechanical movement and sending it through a transmission medium in the form of electrical pulses. Two systems - transmission and reception - contribute to achieving this goal. The first system organizes data and represents it in the form of electrical impulses, while the second system interprets the electrical impulses and converts them into audio or visual symbols based on a linguistic coding model.

Electric telegraph plays a prominent role as a means of transmitting data and linguistic messages via wires or electromagnetic waves. The telegraph uses a specific mechanism to encode data and send it over the carrier, which is then decoded and transformed into audio or visual images, sound beats, written symbols, light shapes, or chemical effects based on the model of linguistic coding[21].

### 1.9.3 Analog processing

Scientists have worked to improve and develop the linguistic processing system for nearly a century and a half, with the aim of meeting the growing ambitions to improve the quality of communication and increase efficiency and speed in transferring linguistic information and data. New technologies have emerged that use what is known as analogue processors, a technology that converts successive physical quantities with time-varying values into a series of analogue electrical signals.

Analog processors allow natural quantities such as sound, light, pressure, heat, and velocity to be represented electrically, mimicking and replicating natural waves. Analog natural energy is converted into an electrical voltage representing a natural wave, and these electrical signals are

sent over transmission media over long distances, where they are decoded and restored to their original form.

Through this technique, language can be represented electrically and transmitted quickly and effectively. Analog signal data is analyzed using axes representing time and voltage, where the signal is variable and continuous over a period of time. The telephone is considered one of the most prominent inventions that use this technology, as it allows the transmission of linguistic sounds from one place to another via electricity wires[21].

## 1.10 Digital techniques

### 1.10.1 Computational processing

The use of digital technology in automated processors has brought about a major transformation in the world, leading to the emergence of the massive revolution in information technology that we enjoy today. The idea of digital processors depends on converting commands and data into a digital form that can be understood and executed accurately and effectively.

The binary number system is the core of digital processors, as it consists of only two numbers : zero and one. This system was discovered by the scientist Gottfried Leibniz in the seventeenth century, and with the development of electrical discoveries, the binary system was linked to the control of electricity.

Dual-processor machines appeared in the third decade of the twentieth century, and as they developed, they turned into digital computers. In the middle of the twentieth century, digital processors witnessed a real technical revolution, making them the primary means of dealing with and exchanging information.

Regarding digital language processors, these processors convert linguistic symbols into a digital representation using the binary system. Audio codes are represented digitally by converting them into an analogue electrical signal, which is then converted into a digital signal. This digital representation is based on the machine's ability to understand and interpret linguistic symbols and perform required tasks.

Representing linguistic symbols in digital form allows a variety of operations and algorithms to be performed on them, which contributes to achieving the required performance and desired purpose[21].

## 1.11 Automated language processing applications

Applications of automated language processing are diverse and span a range of different fields. Computational linguistics has adopted a theoretical and practical approach, directing its capabilities towards exploring the energies of the human mind in creating and understanding linguistic knowledge. Linguistic principles, ideas and symbols are transformed into a mathematical symbolic representation that is compatible with the nature of computers, enabling them to simulate human thinking in using and understanding language And its production.

Automated language processing applications require dealing with language in various aspects of life, whether spoken or written and involving words, texts, analysis and synthesis. These applications vary according to different language levels, e.g Morphology, grammar, semantics, lexicon, and writing..

Automated language processing applications are divided into a group of applied fields, including grammatical, morphological, lexical, semantic, written, orthographic studies, machine translation, and other fields.

Applications of automated natural language processing appear in a wide range of fields, including[20] :

### 1.11.1 Understanding linguistic context

The main goal of human language processing is to achieve a system that can understand linguistic context, whether in its spoken or written form[20].

### 1.11.2 Linguistic statistics

It covers multiple areas, including :

- A : A quantitative estimate of some linguistic features, such as the frequency of use of letters and words, and patterns of inflection in various texts.
- B : A quantitative description of some linguistic relationships or relationships between texts, such as relationships between subject type and verb type, sentence length, and ease of understanding[20].

### 1.11.3 Writing processing

Computing can be applied in automated reading systems for printed or handwritten texts. As for Displaying and printing texts has come a long way[20].

#### 1.11.4 Morphology or word structure

Morphological studies address a variety of purposes, including machine language translation, which relies on computers to do the translation, and intersect with other fields such as computer science, linguistics, and artificial intelligence. In addition, it is used in text compression and spell checking, as the morphological analyzer can perform the analysis and synthesis of language vocabulary, as the analytical part breaks down the word into its primary elements, such as derivation, morphology, parsing, prefixes and suffixes, while the synthetic part does the opposite[20].

#### 1.11.5 Sounds and audio systems

This field is concerned with developing software for automatic speech recognition and conversion. This software analyzes the sound and converts it into symbols that can be understood by computers and identify its source[20].

#### 1.11.6 Grammar and systems

These technologies are used in many applications such as automatically interpreting and producing human texts, whether for querying or machine translation purposes[20].

#### 1.11.7 Developing linguistic dictionaries

In understanding human speech, words are the intelligible units. Since any program designed to deal with human languages, it needs to use dictionaries of all kinds. These dictionaries require a set of characteristics such as comprehensiveness, regularity, consistency, clarity, accuracy, scalability, expansion, and modification.

In addition to the traditional applications of computational linguistics and others, the electronic researcher can perform tedious and tedious work such as building vocabulary, analyzing its meanings, and matching the semantic and grammatical structure of the mother tongue with the foreign language.

Computational linguistic studies have demonstrated new applications as a result of the interaction of language and computers, and new interactive, creative linguistic practices have emerged, such as linked text, which relies on linking internal and external text structures.

To achieve this interactive literature, the writer must go beyond the traditional stereotype and acknowledge the role of the recipient in constructing the literary text, and seek to present a

lively text in which true interactivity is achieved[20].

## 1.12 DATA MINING

Data mining is the process of exploring interesting knowledge from a large amount of data stored in databases, warehouses or other information sources. Based on this belief, a typical system architecture includes the following major components [3] . as shown in Figure 1.5

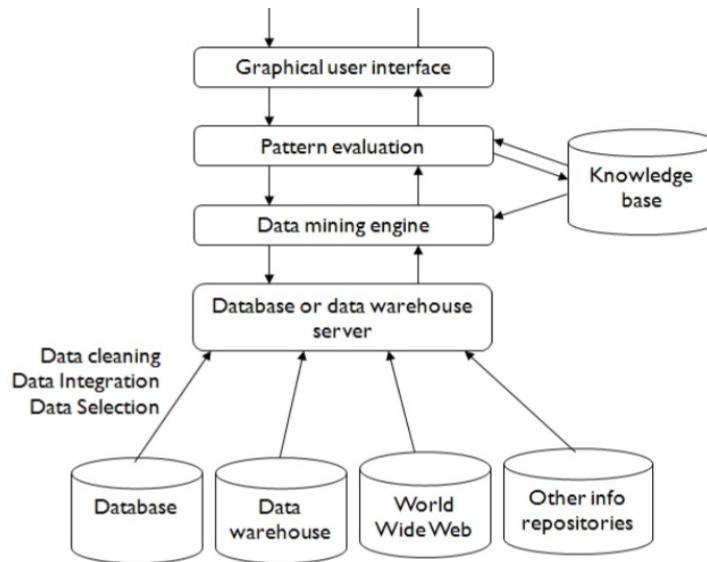


FIGURE 1.5 – Architecture of a typical Data Mining System [2].

- Data mining techniques provide a wide range of tools, techniques, and combinatorial algorithms to extract the required data, as well as pre- and post-processing of data, and metrics of interest and complexity.
- Data mining algorithms allow useful patterns to be revealed within huge amounts of data.
- Data mining algorithms provide precise and precise analysis results.
- Data mining contains various tools for classification, clustering, anomaly detection, and pattern extraction.

Data mining includes a variety of tools for classification, clustering, anomaly detection, and pattern extraction, as shown in Figure 1.6 [22].

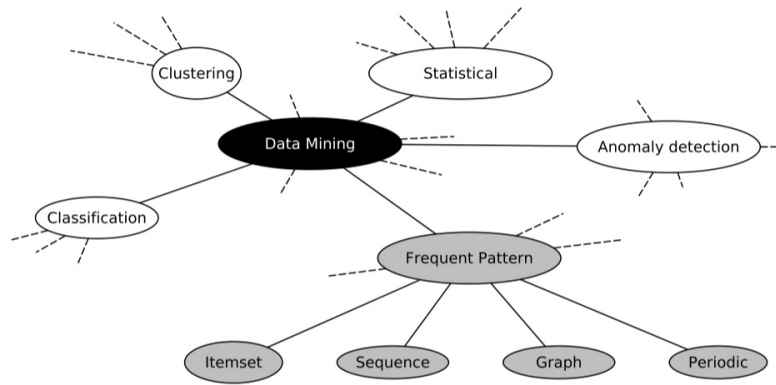


FIGURE 1.6 – Map of Data Mining domains[3].

### 1.13 Conclusion

In this chapter we learned about important concepts in computer science and linguistics, including computational linguistics, automated natural language processing, and data mining. We discussed the basics of these fields and their importance in understanding and analyzing human language with the help of computers, which opens the way to exploring how modern technologies can be used to extract knowledge from linguistic data and interact with it more effectively.

---

---

# CHAPITRE 2

---

## COMPARISON SYSTEM DESIGN

### 2.1 Introduction

Reviewing examples of a specific task is considered one of the fundamental steps to understand its nature and effects. In every challenge faced by individuals or organizations, leveraging past experiences and works can lead to success and overcoming obstacles. This chapter aims to review several similar examples to our work, analyzing how these tasks were executed and identifying their outcomes.

### 2.2 Example 1 : The book Semantic Analysis and Graphical Description of Ibn Khaldoun Introduction.

#### 2.2.1 Computerized processing of Ibn Khaldoun introduction

At first, the book's text underwent refinement to remove linguistic flaws that could impact its meaning, resulting in the creation of a Word Cloud, a visual representation of the text's content based on the frequency of words used - see [Figure 2.1](#)

The dominance of prepositions and pronouns over the concepts presented in [Figure 2.1](#) is evident due to their frequent usage in constructing sentences and linking their components. Therefore, in the next step, intelligent software algorithms are utilized to remove the surplus "Stop Words" that lack independent meaning. This process results in the emergence of concepts after elimi-

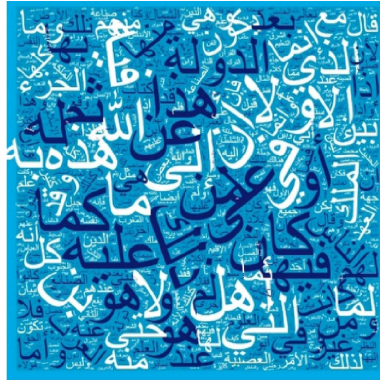


FIGURE 2.1 – Vocabulary cloud of Ibn Khaldoun introduction before the purification process was carried out from linguistic impurities[4].

nating the polluted words from the text, as illustrated in Figure 2.2.

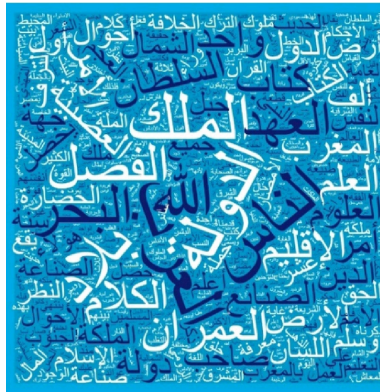


FIGURE 2.2 – Vocabulary cloud of Ibn Khaldoun introduction after practicing purification processes from linguistic impurities[4].

The new graphical description clearly highlights the emergence of concepts that formed part of Ibn Khaldoun cognitive framework. We observe the prominence of terms in the graphical representation, such as : Arabs, Sultan, State, Land/Industries/Nervousness, Sharia, Science, Language, Maghreb, among others, with variations in the size of vocabulary, which is correlated with their frequency levels within the text[4].

**Analysis of selected text network data**

Analysis of text network data is considered the first stage in identifying distinctive features, such as statistical properties, levels of association and grouping between terms, and the meanings inferred from them. In Ibn Khaldun’s book, the analysis revealed four main clusters of terms in Figure 2.3., where some of them, such as urbanization and people, stand out at

a rate of 25%, followed by other terms, such as the state and king, at a similar rate. These findings highlight the author’s interest in these terms and their connection to the rest of the discourse. Upon deeper examination of Table 2.4, it appears that the network consists of 3,000 basic nodes connected to each other, and content analyzes confirm that the most influential elements revolve around basic concepts such as God, the state, and the Arabs, which indicates the strength of interconnection and cohesion in the cognitive network of the text[4].

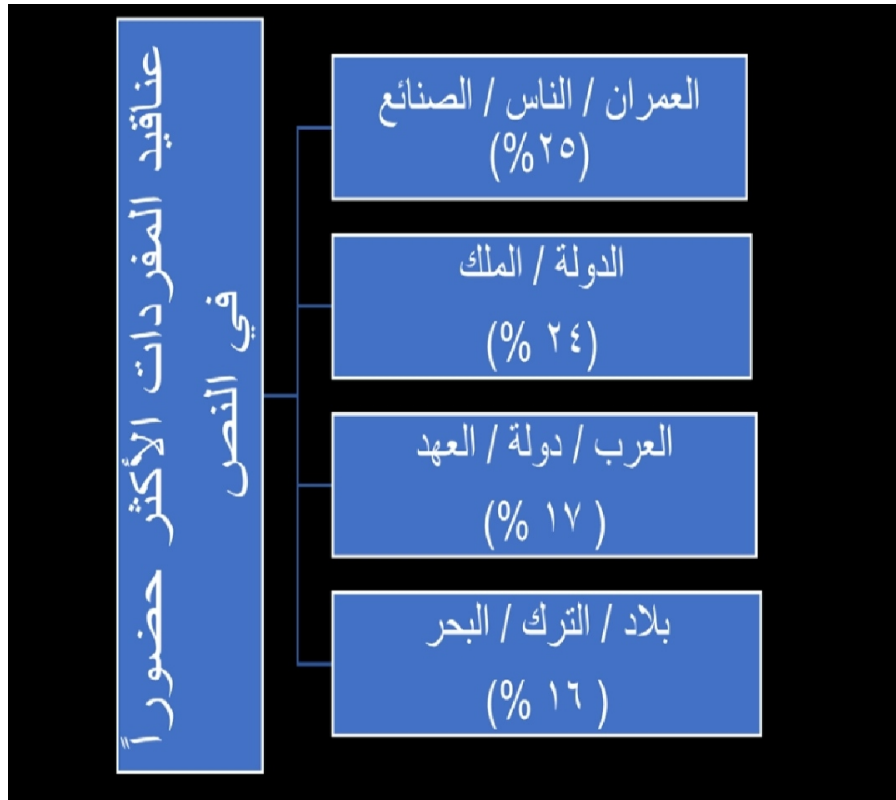


FIGURE 2.3 – The most common vocabulary words in the vocabulary clusters of the Introduction book[4].

العقدة	الدرجة	التكرار (مستوى الحضور)	الرابط البيني للعقدة
مجموع عقد الشبكة الأساسية.	3000	15,314	2.4567
المجموع لكل 150 عقدة شبكية	20	102.09	0.0164
الله سبحانه وتعالى	126	936	0.4790
الدولة	85	499	0.2369
العرب	62	296	0.1526
بلاد	33	255	0.1124
المُلك	77	350	0.0932
العمران	49	187	0.0731
العلوم	25	145	0.0387

FIGURE 2.4 – Table The most important characteristics of the selected text network fabric[4].

### Graphical representation of the elected text network

After the previous text processing stage, it becomes possible to create a graphic network that directly represents the terms and concepts found in the text, linking and arranging them according to their locations and relationships in the text. Figure (2.5) By studying this network, it is possible to identify the vocabulary that was the focus of the writer's attention, which indicates patterns And complex relationships between these terms in the content of the book.

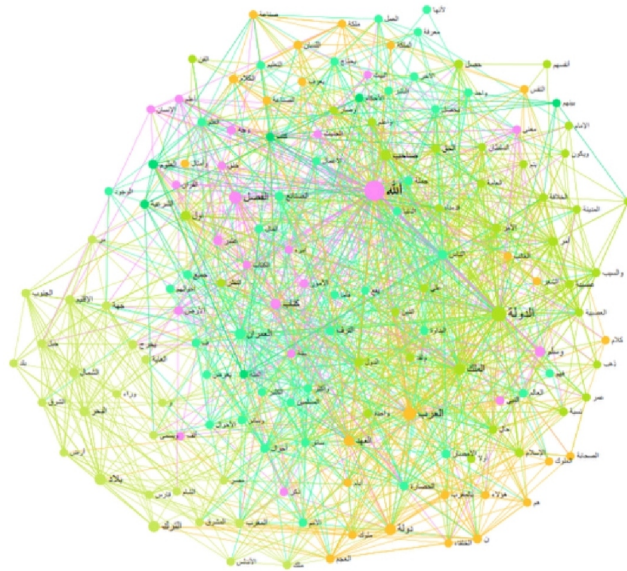


FIGURE 2.5 – The grid for the content of Ibn Khaldoun introduction[4].

### 2.2.2 Summary and conclusions

Through a deep exploration, we’ve analyzed the terminology and relationships within Ibn Khaldun’s work. Utilizing text network analysis with smart computing processors, we’ve uncovered patterns and constructed networks of concepts across various fields of knowledge in the book.

This approach offers new insights into historical texts, revealing terminology usage, deployment methods, and intimate relationships between terms. It provides comprehensive descriptions and analyses beyond traditional reading methods.

However, while this method enhances understanding, it remains one tool among many, complementing traditional reading rather than replacing it. It aids in interpretation and analysis without hindering comprehension of the texts[4].

## 2.3 Example 2 : A computerized processing book to analyze the network of meanings and concepts in books on the expression of vision

### Forming a word cloud of the two books

The term "word cloud" refers to a graphical representation used to display the vocabulary present in a text. Words are shown in sizes proportional to their frequency in the text, offering an initial view of the most significant words and terms. To illustrate the impact of text purification and subsequent processing, we created word clouds for both books before and after these treatments. 2.6 shows the word cloud of the book "Muntakhab al-Kalam fi Interpretation of Dreams" before and after processing.

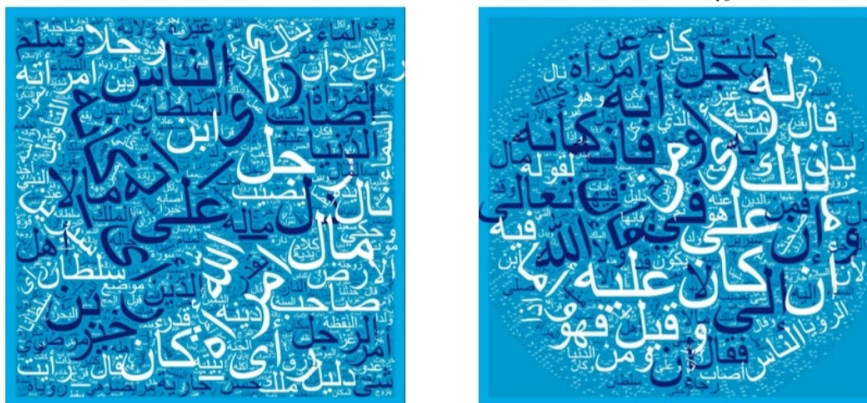


FIGURE 2.6 – Word cloud of the chosen book of dream interpretation before and after the treatments[5].

The first word cloud clearly indicates that the text of the book before processing was dominated by prepositions, pronouns, and words unrelated to the field of dream interpretation. As for the second word cloud, terms related to interpretation started to emerge after processing. The first word cloud clearly shows that the text in the book before treatment was dominated by a lot of prepositions, pronouns and words not related to the field of dream interpretation. As for the second word cloud, the terminology and vocabulary related to the interpretation of dreams began to appear more clearly. The same applies to the content of the book “Perfuming the Dreams in the Interpretation of Visions and Dreams,” as shown in Figure 2.7. After prepositions, pronouns, and words lacking inherent meaning were predominantly positioned on the right side (e.g., in, to, that, for, his, and others), they seem to have vanished after processing, giving prominence to important terms like "vision," "evidence," "perceived,"

and vocabulary related to dreams such as "people," "woman," "money," and others[5].



FIGURE 2.7 – The word cloud of the book Tahir al-Anam in the expression of vision and dream Before and after treatment[5].

**Visual representation of the two books’ text networks**

The network representation of the text illustrates the semantic and conceptual structure, identifying links between terms and the pattern of word repetition. The text is represented as a network of nodes (words and terms) and their connections (semantic and conceptual relationships).

The texts of the two books were processed as networks of nodes representing vocabulary, and the relationships connecting these words form the expression of vision in the texts. Figure 2.8 shows the complexity of the network in "The Selected for the Interpretation of Dreams" before and after text purification, where the content was obscure before purification and became clearer afterward. The same applies to the network in "Attar Al-Anam," as shown in Figure 2.9 [5] .

Selected conceptual networks from the book "Muntakhab al-Kalam fi Tafsir al-Ahlam" : We attempted to select three conceptual elements for the purpose of forming networks that would visually represent the conceptual description. Notice in Figure 2.10 the approach associated with the narrative style of the dreamer’s perception.

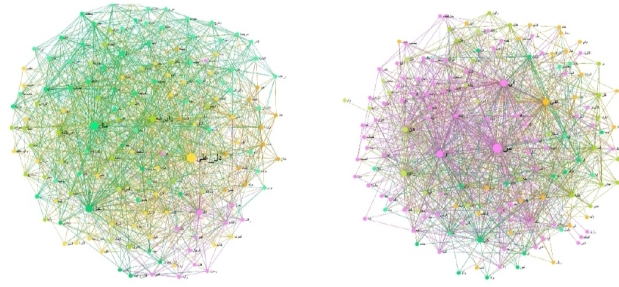


FIGURE 2.8 – The network structure of the text material of the selected book from Interpretation of Dreams before and after treatment[5].

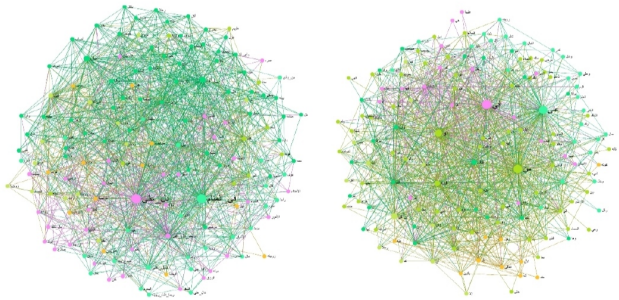


FIGURE 2.9 – The network texture of the text material of the book Tatir al-Anam in the expression of vision and dream before and after processing[5].

It seems that the most important terms used by the author in presenting the visions are : "saw, saw that, saw as if, saw as, saw that". By tracking these terms, patterns of dream narratives and their terms can be identified. As for Figure 2.11, we have depicted the network of concepts related to interpreting the significance of the dreamer's vision for men, women, children, and wealth in Imam Ibn Sirin interpretations, and the implications of that on his handling of these terms in the expression process [5].

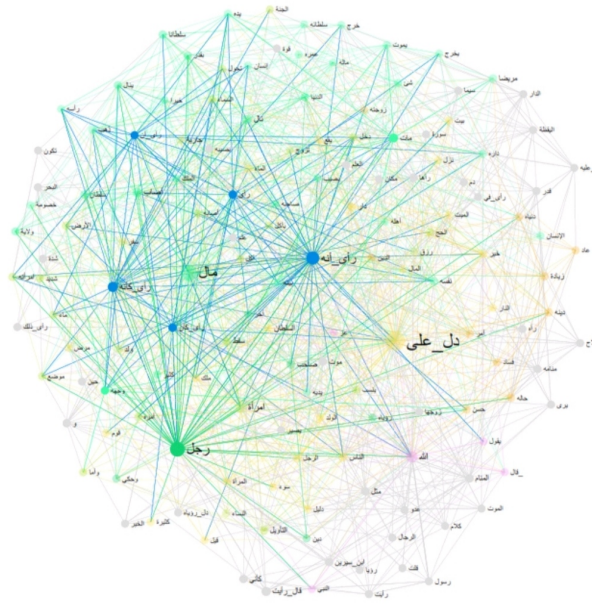


FIGURE 2.10 – Network of concepts of expressions of the viewer’s story of his vision[5].

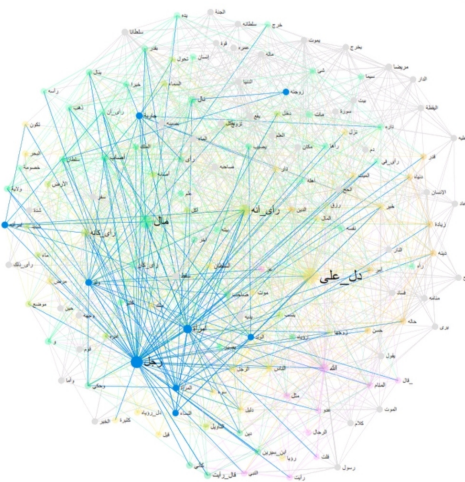


FIGURE 2.11 – The conceptual network of the meaning of the vision of a man, a woman, a child, and money and its reflection on its expression according to Imam Ibn Sirin [5].

### 2.3.1 Summary and conclusions

Employing computational analysis on Arabic and Islamic texts, and visually representing their discourse, advances computational methods in analyzing Arabic and Islamic cognitive discourse. Dream interpretation books contain valuable resources for understanding dreams and their meanings.

Our analysis has produced graphical representations of these books' vocabulary, aiding comparison and revealing conceptual networks related to dream interpretation. This sheds light on connections between recurring terms in contemporary Muslim dreams and explains their semantic field.

Analyzing conceptual networks helps understand how interpreters link dream terms across different time periods. Computational processing highlights variations in term recurrence and analysis styles in dream interpretation books.

Expanding this approach to more books could deepen our understanding of dream terms and their expression[5].

## 2.4 Exemple 3 : Review of Research on Digital Supply Chain Management Using Network Text Analysis

### 2.4.1 Subject Analysis

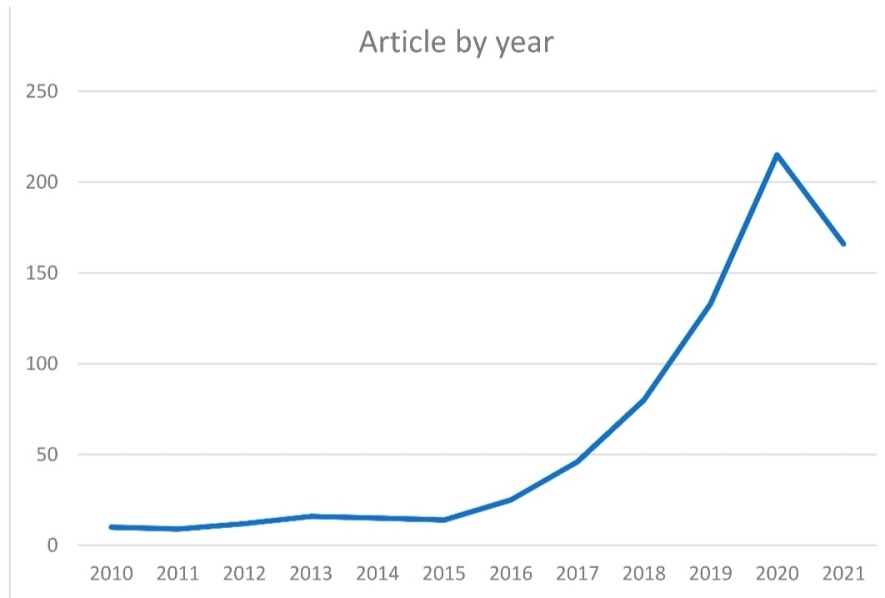


FIGURE 2.12 – Articles by year [6].

Scopus, a well-known academic database, was used to find articles on digital supply chain management. Using keywords like "digital supply chain," "supply chain digitalization," and others, articles from journals in business, management, decision sciences, and social sciences were searched. Only English-language articles were analyzed, resulting in 815 articles. Figure 2.12 shows a growing trend in studies on digital supply chain management since 2010, with a notable increase after 2015, coinciding with discussions on digital transformation. However, as the study's search ended in May 2021, the actual number of articles in 2021 may be higher. Table 2.13 lists the major journals publishing these articles, and Table 2.14 identifies prominent authors[7].

Source Title	Number of Articles
International Journal of Supply Chain Management	60
Sustainability	40
International Journal of Production Economics	31
Journal of Cleaner Production	30
International Journal of Production Research	29
Production Planning and Control	26
Supply Chain Management	17
International Journal of Logistics Management	11
Supply Chain Forum	11
Technological Forecasting and Social Change	11
IEEE Transactions on Engineering Management	10
Industrial Management and Data Systems	10
Benchmarking	9
International Journal of Information Management	9
Journal of Manufacturing Technology Management	9
Annals of Operations Research	8
European Journal of Operational Research	8
International Journal of Physical Distribution and Logistics Management	8

FIGURE 2.13 – Table Major journals publishing digital supply chain articles [7].

Author	Number of Articles
Ivanov, D.	10
Kumar, S.	7
Gunasekaran, A.	6
Dolgui, A.	6
Sarkis, J.	6
Graham, G.	5
Tsolakis, N.	5
Queiroz, M.M.	5
Kolesnikov, A.V.	4
Ukolov, V.F.	4
Barykin, S.Y.	4
Luthra, S.	4
Srai, J.S.	4
Holmström, J.	4
Rai, A.	4

FIGURE 2.14 – Major authors publishing digital supply chain management articles [7].



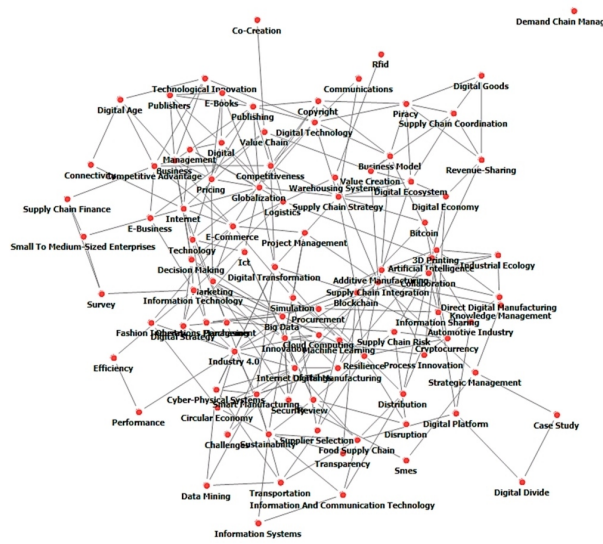


FIGURE 2.16 – Keyword network (until 2019) [7].

### Centrality Analysis

Table 2.17 displays the results of the degree centrality analysis, revealing keywords such as "big data," "Internet of Things," "globalization," "Industry 4.0," and "innovation" with high centrality. This indicates a predominant focus on technology like big data analysis and the IoT in digital supply chain studies, driven by innovation and influenced by Industry 4.0 and supply chain globalization. Similarly, Table 2.18 outlines the results of the betweenness centrality analysis, where "supply chain integration" emerges as a prominent mediator among keywords. This suggests that digitalization is widely regarded as a strategy to enhance supply chain integration and overall performance in digital supply chain management studies [7].

Keywords	Degree Centrality
Big Data	0.145833
Internet of Things	0.135417
Globalization	0.135417
Industry 4.0	0.125
Innovation	0.125
Resilience	0.104167
E-Commerce	0.104167
Supply Chain Integration	0.104167
Information Technology	0.104167
Internet	0.104167
Digital Technology	0.09375
Supply Chain Strategy	0.09375
Business	0.09375
Publishing	0.09375
Blockchain	0.083333
Digital	0.083333
Digital Transformation	0.083333
Sustainability	0.083333
Additive Manufacturing	0.083333
Marketing	0.083333
Smart Manufacturing	0.083333
Competitiveness	0.083333

FIGURE 2.17 – Table Degree centrality analysis (until 2019)[7].

Keywords	Node Betweenness Centrality
Supply Chain Integration	0.130157
Big Data	0.106023
Information Technology	0.089664
Innovation	0.080355
Internet	0.078466
Globalization	0.076832
Internet of Things	0.067257
E-Commerce	0.067026
Supply Chain Strategy	0.062592
Industry 4.0	0.053164
Resilience	0.052187
Digital Transformation	0.052162
Food Supply Chain	0.050802
Sustainability	0.048462
Additive Manufacturing	0.043211
Digital Ecosystem	0.038675
Blockchain	0.037176
Collaboration	0.037022
Digital Technology	0.035983
Value Chain	0.035457
Logistics	0.033814
Marketing	0.033204

FIGURE 2.18 – Table Betweenness centrality analysis (until 2019)[7].

### 2.4.3 Network Text Analysis on Articles Published from 2020

#### Building the Keyword Network

From 2020 to May 2021, 381 articles on the digital supply chain were published, featuring a total of 1250 keywords. Figure 2.19 presents a word cloud reflecting keyword frequency. Notably, terms like "COVID-19," "Resilience," "Supply Chain Risk Management," and "Digital Twin" saw increased frequency during this period. This reflects the heightened interest in supply



increased studies on digital supply chains as a strategy to bolster supply chain resilience. Table 2.22 illustrates the outcomes of betweenness centrality analysis, indicating a surge in studies on supply chain management facilitated by digital technology[7].

Keywords	Degree Centrality
Digital Technology	0.256757
Industry 4.0	0.243243
Resilience	0.22973
Innovation	0.202703
COVID-19	0.202703
Digital Transformation	0.189189
Digital Twin	0.175676
Sustainability	0.162162
Blockchain	0.162162
Disruption	0.162162
Circular Economy	0.148649
Coronavirus	0.135135
Simulation	0.121622
Artificial Intelligence	0.121622
Big Data	0.121622
Risk Management	0.121622
Data Analytics	0.121622
Business Model	0.108108
Pandemic	0.108108
Resource-Based View	0.108108
Dynamic Capabilities	0.108108
Fashion Industry	0.108108
Smart Contracts	0.108108
Ripple Effect	0.108108
Internet of Things	0.094595
Information Systems	0.094595
Performance	0.094595
E-Commerce	0.081081

FIGURE 2.21 – Table Degree centrality analysis (from 2020)[7].

Keywords	Node Betweenness Centrality
Digital Technology	0.157578
Industry 4.0	0.145707
Digital Transformation	0.086762
Blockchain	0.08134
Innovation	0.079363
Sustainability	0.076611
Resilience	0.070143
COVID-19	0.052188
Digital Twin	0.047308
Business Model	0.046842
Circular Economy	0.046083
Internet of Things	0.038434
Big Data	0.037876
Traceability	0.034921
Supply Chain Finance	0.03312
Supply Chain Performance	0.032956
Logistics	0.028666
Additive Manufacturing	0.027834
Artificial Intelligence	0.026272
Resource-Based View	0.025894
Systematic Literature Review	0.02343
Data Analytics	0.022839
Supply Chain Risk	0.022554
Dynamic Capabilities	0.022045
Fashion Industry	0.021602
Smart Contracts	0.020747
Disruption	0.020117

FIGURE 2.22 – Table Betweenness centrality analysis (from 2020)[7].

## Cluster Analysis

A cluster analysis was conducted to identify research topic groups. As depicted in Figure 2.23, five clusters were formed, with the primary keywords included in these clusters outlined in Table 2.24.

The first cluster includes keywords related to supply chain integration and new business models using digital technology. Fatorachian and Kazemi noted that digitalization enables process integration in the supply chain, leading to flexibility and agility, thus enhancing the overall performance of the supply chain. Zeng et al. presented a case study on small and medium-sized enterprises in Italy to demonstrate that new business models can be created using digital technologies such as blockchain in the food supply chain.

The second cluster highlights COVID-19's impact on supply chain resilience, emphasizing the need for digital solutions to detect and address emerging risks. Adopting digital supply chains is crucial for sustainability. The third cluster underscores key digital technologies like blockchain,

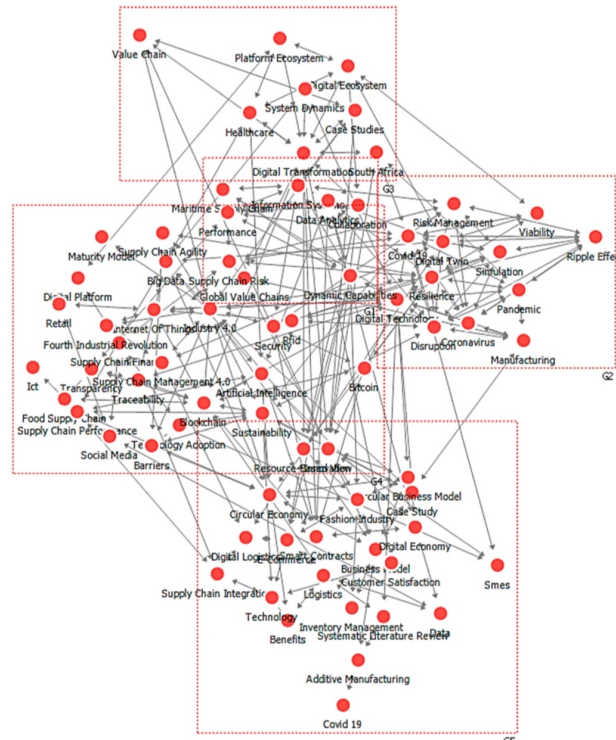


FIGURE 2.23 – Cluster analysis (from 2020) [7].

IoT, AI, big data, and cloud computing essential for sustainable digital supply chains. As these technologies improve in security and accessibility, more companies are expected to adopt digital supply chains.

The fourth cluster discusses creating a new ecosystem and value chain through digital supply chain transformation, driven by the digital era’s need for interconnected ecosystems. Authors like Seyedghorban et al. suggest that companies are adapting their strategies, making digital supply chain management a viable option. The final cluster highlights collaboration and dynamic capabilities among firms for effective digital supply chain implementation. Dynamic capabilities are essential for promptly adopting new technologies. Data sharing is crucial for efficient operation, necessitating strategic collaboration based on trust, enhancing relational performance within the ecosystem[7].

Cluster	Main Keywords	Related Topics
1	Business Model, Supply Chain Integration, Digital Economy	Studies on supply chain integration and new business models using digital technology with the emergence of the era of digital economy
2	COVID-19, Disruption, Risk Management, Resilience, Viability	Studies on digital supply chain as a strategy to promote supply chain resilience as COVID-19 has increased the supply chain risks
3	Sustainability, Blockchain, Internet of Things, Artificial intelligence, Big Data	Studies on various digital technologies necessary for building a sustainable digital supply chain
4	Digital Transformation, Digital Ecosystem, Value Chain	Studies on building a new ecosystem and value change due to digital transformation of supply chains
5	Collaboration, Dynamic Capability, Performance	Studies on collaboration and dynamic capabilities among firms, which are necessary for building and operating a digital supply chain

FIGURE 2.24 – Main keywords by cluster (from 2020) [7].

### 2.4.4 Conclusions

This study analyzed 815 articles on digital supply chains in Scopus to uncover trends. Findings emphasized sustainable management and the role of digital transformation. Technologies like blockchain and IoT were highlighted. Interest surged during COVID-19, prompting digital transformation for resilience. Cluster analysis identified major research topics. Findings are summarized in Table 2.25[7].

Main Clusters of Keywords	Related Studies
Industry 4.0, Digital Transformation	The mega trends such as that of Industry 4.0 and digital transformation demand digital transformation of supply chains. It is necessary to establish general business strategies for the digital transformation of supply chains and reconsider existing business models and value chains.
Supply Chain Integration	Digital supply chains reinforce supply chain integration into a complicated global network beyond simple and linear supply chains. There must be viability as an organic ecosystem beyond simply connecting trade networks among firms with digital technology.
Resilience	A digital supply chain ultimately enables real-time end-to-end visibility and improves resilience. Supply chain risk management must be done with a focus on digitalization, especially in view of COVID-19.
Digital Technologies	A digital supply chain is not built at once in a short time but must be built by applying, step by step, various digital technologies. Due to the nature of supply chains, performance cannot be easily improved when only focal firms are digitalized. The supply chain must be digitalized overall by digitalizing small and medium-sized firms in the supply chain as well.

FIGURE 2.25 – Keyword clusters [7].

## 2.5 example 4 :paths study on knowledge convergence and development in computational social science : data metric Analysis Based onWeb of Science

### 2.5.1 Data Sources and Research Methods

The study utilizes the Web of Science (WoS) database to collect data, focusing on articles related to computational social science and related fields. After refining the search criteria, 5856 relevant articles were selected. The research methodology involves standard bibliometric analysis, including research design, data collection, analysis, visualization, and results interpretation. Biblioshiny and SciMAT software packages are used for comprehensive bibliometric analysis and knowledge mapping, exploring the evolution and integration of computational social science literature. These tools enable visualization and network analysis of the literature, providing insights into knowledge integration and development paths within the field. Figure 2.26). [8].

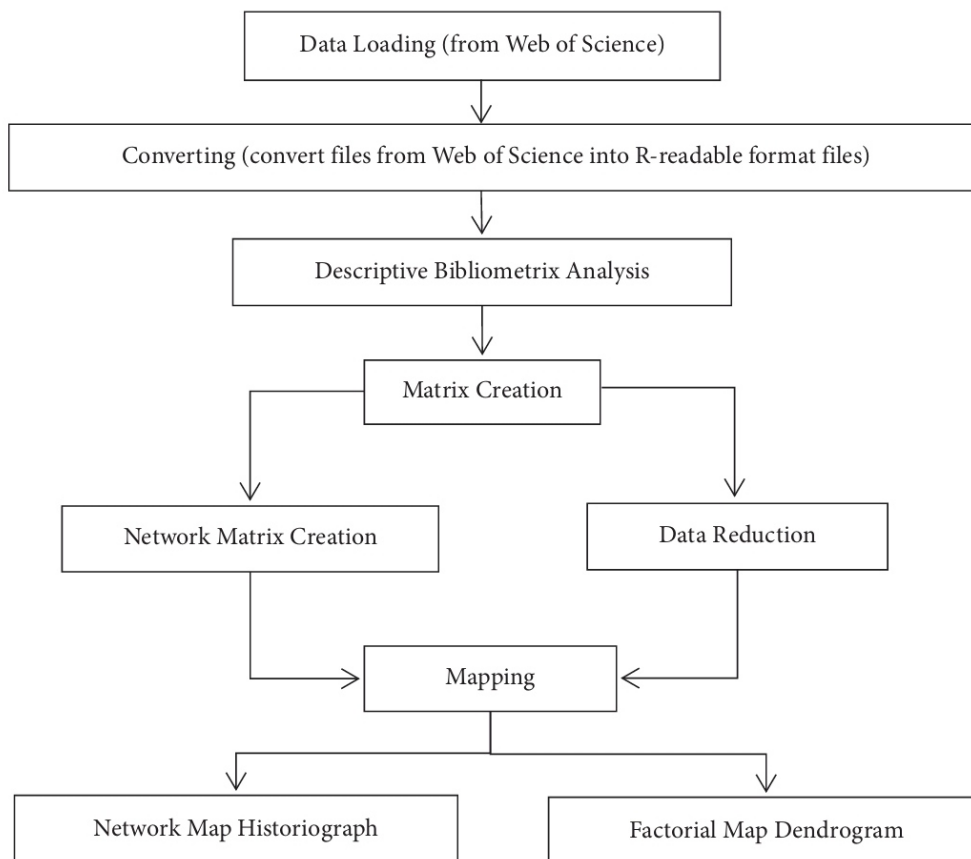


FIGURE 2.26 – Bibliometric analysis process[8].

## 2.5.2 Analysis of Evolutionary Dynamics

### Statistical Description of Literature in Computational Social Sciences

#### 1. Analysis of Published Article Volume :

This section examines the dynamic developments in scientific knowledge production within computational social science since its inception. It employs both annual time series and phased time series analysis to gain a comprehensive understanding.

From 2001 to 2020, (as depicted in Figure 2.27 and Table 2.28) there is a general upward trend in the number of articles despite minor fluctuations. The period is divided into four phases : 2001-2005, 2006-2010, 2011-2015, and 2016-2020.

During 2001-2005, there is an exploratory phase with low productivity due to the early stage of computational social science. The period from 2006 to 2010 marks initial growth, notably spurred by increased interest following the proposal of the concept in 2009.

Rapid growth occurs during 2011-2015, particularly evident in 2014 and 2015. The period from 2016 to 2020 signifies an active and highly productive phase, reaching its peak in 2020.

In summary, retrospective analyses indicate continuous growth in the number of articles in computational social science research, suggesting ongoing development in the field[8].

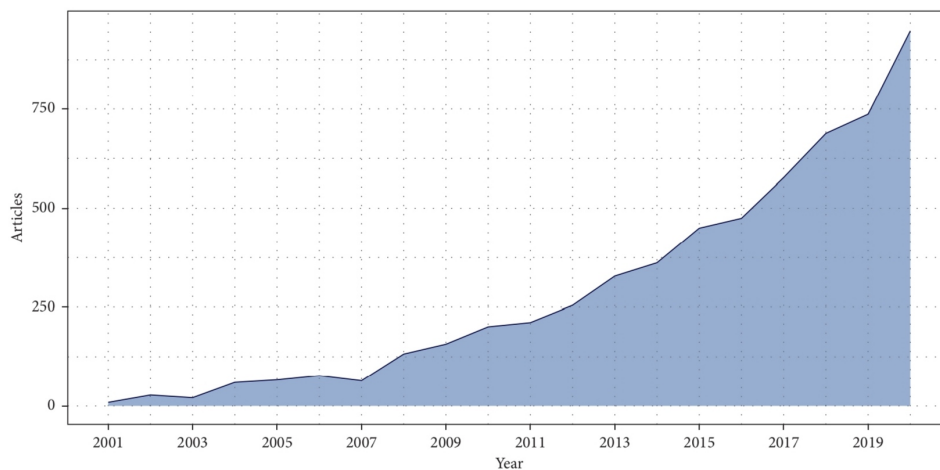


FIGURE 2.27 – Academic papers published related to computational social science, 2001–2020 [8].

TABLE 1: Valid literature and high frequency keywords in different periods.

Periods	2001–2005	2006–2010	2011–2015	2016–2020	Total
Valid literature	185	632	1608	4331	5856
High-frequency keywords	7	29	97	204	337

FIGURE 2.28 – Table Valid literature and high frequency keywords in different periods[8].

2. Analysis of Annual Citation Trends :

Figure 2.29 shows that the average number of citations per research paper peaked around 2012, followed by a slower growth trend since 2014. Despite fluctuations, the overall number of citations has continued to increase. The highest average number of citations per paper was in 2012, reaching 7.8. This peak coincided with the publication of a manifesto for computational social science by 14 European and American researchers, including R. Conte from the Italian National Council for Scientific Research, in the European Journal of Physics. From 2008 to 2014, the average citation rate of articles increased, with a focus on integration and development in social sciences and algorithmic models. Based on the analysis of published articles, this research suggests that research in this field will maintain a high growth trend, particularly with technological advancements. [8].

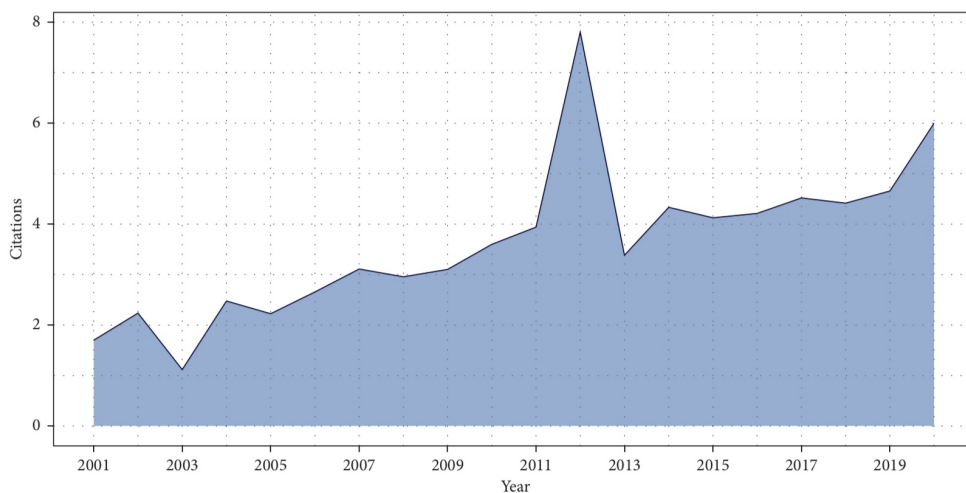


FIGURE 2.29 – Average annual citations in computational social science, 2001–2020[8].

3. The highest cited articles :

Figure 2.30 shows that, among 5856 articles, clustering by fast search and finding of density peaks, with up to 2180 citations, proposes an analysis method based on the idea

of clustering that can effectively eliminate outliers. Next, private traits and attributes are predictable from digital records of human behavior, with 942 citations. Kosinski et al. found that digital behavioral records like “Facebook Likes” can be used to predict a range of highly sensitive personal attributes, including sexual orientation, ethnicity, religious and political views, and personality traits, thus proposing a demographic model that can predict individual psychology from preferences. As mentioned, computational social science is an emerging interdisciplinary discipline that is the result of long-term knowledge accumulation in traditional social sciences and is a multidisciplinary and problem-solving oriented research[8].

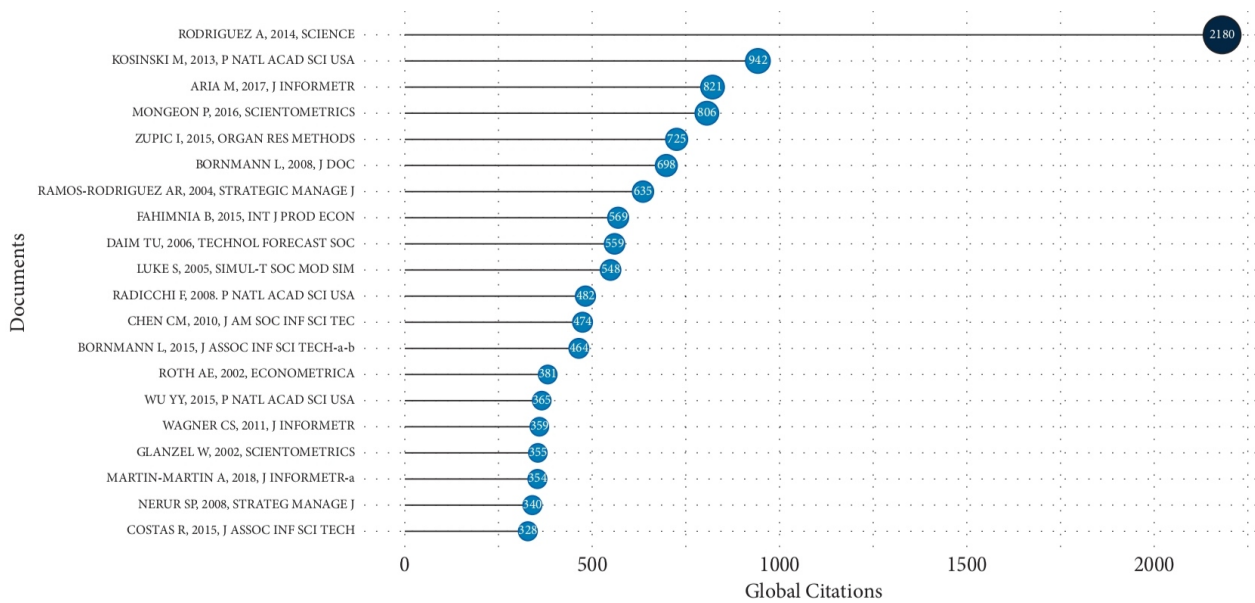


FIGURE 4: Top 20 cited literature.

FIGURE 2.30 – Top 20 cited literature [8].

4. Analysis of the Most Relevant Authors :

Moat et al. possess H-index, G-index, and citation frequency of 10, 13, and 444, respectively, among computational social science authors (Table 2.31) Furthermore, several author collaboration subnetworks, notably six, have been identified (Figure 2.32), predominantly centered on Moat, Preis, and Lazer, forming intricate relationships. These findings underscore the representation of relationships between nodes in computational social science networks, with the formation of new structures between subnetworks, enhancing overall complexity.

Moat et al. leveraged social media behavioral data to model human mobility patterns and utilized big data to quantify crowd size. In 2009, Lazer et al. published "Computational

Social Science," a seminal work marking the field's emergence. Their 2020 publication, "Opportunities and Challenges in Computational Social Science," reflects on the field's evolution. Despite not leading in publication count, Lazer et al. have made substantial contributions to computational social science's advancement.

The interdisciplinary nature of the field is evident through contributions from scholars of diverse backgrounds, such as physics, engineering, medicine, computer science, information systems, sociology, and political science. This highlights the applicability of various disciplines, including mathematics, statistical mechanics, complexity science, network science, natural language processing, artificial intelligence, and social science theories, in computational social sciences [8].

Author	h_index	g_index	m_index	TC	NP	PY_start
Moat HS	10	13	1.111	444	13	2013
Preist	8	11	0.889	407	11	2013
Cloff-Revillac	5	8	0.294	626	8	2005
Bail CA	6	7	1.000	279	7	2016
Stanley HE	6	7	0.545	392	7	2011
Burnap P	6	6	0.857	240	6	2015
Carley KM	5	6	0.250	459	6	2002
Cheny	6	6	0.375	319	6	2006
Juergens P	6	6	0.600	237	6	2012
Jungherr A	5	6	0.500	229	6	2012
Squazzoni F	5	6	0.357	114	6	2008
Brandt F	4	5	0.308	64	5	2009
Kosinski M	5	5	0.556	1402	5	2013
Lepri B	2	3	0.500	10	5	2018
O'Brien DT	4	5	0.571	94	5	2015
Shah DV	3	5	0.429	127	5	2015
Conte R	4	4	0.400	202	4	2012
Ferrarae	4	4	0.400	189	4	2012
Garcia D	4	4	0.444	77	4	2013

FIGURE 2.31 – Table Top 20 authors with significant impact in computational social science [8].

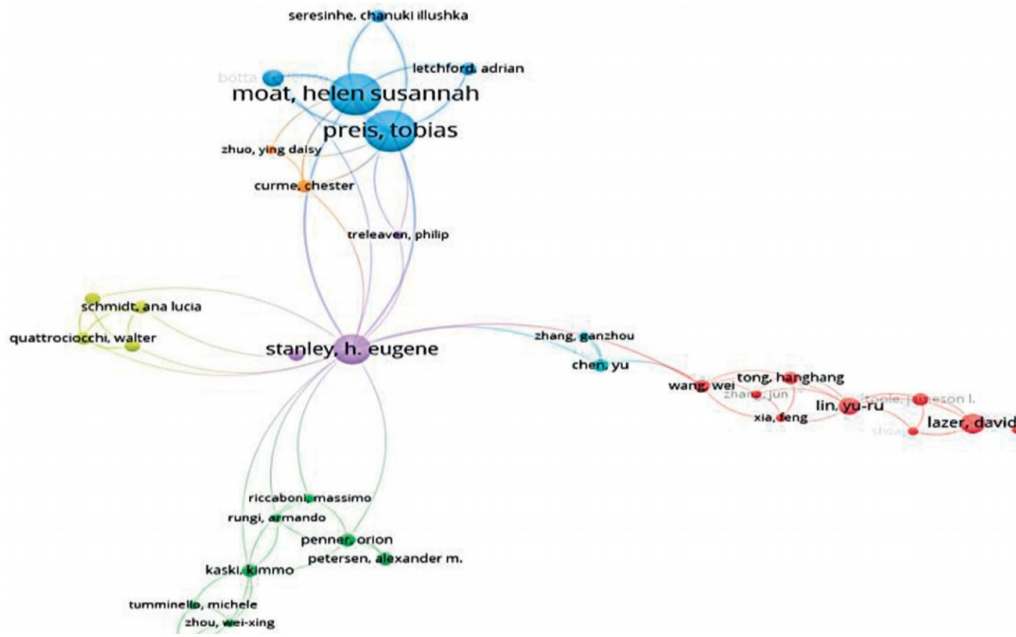


FIGURE 2.32 – Collaborative network of scholars related to computational social science [8].

5. Rephrasing the text :

This study analyzes the top 20 most prolific scholars in terms of their yearly publication count to understand their research output pattern. Abramo G., for instance, has maintained an average of 4 publications per year since their initial publication in 2009, with a peak of 15 in 2011. Notably, one of their publications proposing a heuristic approach to author name disambiguation in bibliometric databases gained wide attention, being cited 128 times.

The analysis indicates that many prolific scholars began focusing on computational social science after its formal introduction in 2009. However, with the field being multidisciplinary, more high-quality authors have become involved, leading to an increase in related publications over time[8].



FIGURE 2.33 – Annual number of publications by lead authors [8].

### Keywords Analysis

This section aims to address the second question, which is how the main research directions in the field have changed since the emergence of computational social science. Keywords, serving as condensed summaries of the literature, are identified from among 5856 scientific documents analyzed in this study, totaling 337 high-frequency words for word cloud analysis and cooccurrence matrix visualization[8].

1. Analysis of the Application of High-Frequency Keywords : The word cloud in Figure 2.34 displays the most frequently occurring words in papers on computational social science, varying in size according to their frequency of appearance. Among these, common keywords include science, big data, model, and bibliometrics. Computational Social Science : Falling within the realm of social science, it integrates theories from various disciplines and serves as a theoretical and methodological system for deepening our understanding of society, effecting societal transformation, and addressing complex social problems in areas such as politics, economics, culture, and more[8].



FIGURE 2.34 – Word cloud in computational social science [8].

2. Keyword Co-occurrence Network Analysis : The examination of high-frequency words across different time periods enabled the construction of keyword co-occurrence matrices. Visualized using the Co-occurrence Network of Biblioshiny, these matrices facilitated the analysis of relationships among thematic words and the tracing of key term evolution, supporting related research in the field. The study constructed keyword co-occurrence networks based on four distinct literary periods : the low-yield exploration period (2001–2005), budding development period (2006–2010), rapid development period (2011–2015), and high-yield active period (2016–2020). Co-occurrence analysis generated mappings depicted in Figures 2.35– 2.38, with node and line size indicating point density, reflecting co-occurrence density and research focus.[8].

From 2001 to 2005, computational social science wasn’t distinctly proposed, with re-

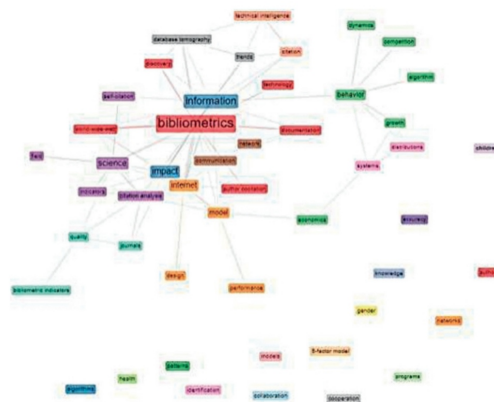


FIGURE 2.35 – 2001–2005 cooccurrence network [8].

search focusing on bibliometrics, scientific inquiries, and impact assessment. Key terms were infrequently cited, primarily concentrating on social investigations and scientific metrics, laying groundwork for computational social science.

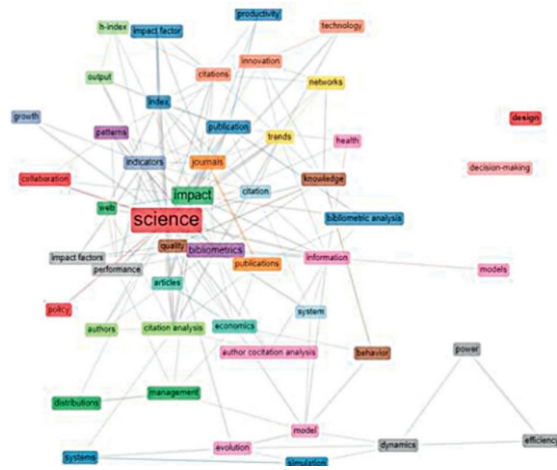


FIGURE 2.36 – 2006–2010 cooccurrence network [8].



FIGURE 2.37 – 2011–2015 cooccurrence network [8].

In 2009, Lazer et al. introduced "computational social science," alongside big data computing, gaining widespread attention since. Computational models are expected to become crucial societal infrastructure, significantly impacting from 2011 to 2015, enhancing management, policy decision-making, and economic growth evaluation.

from 2011 to 2015, with novel applications in big data-based research. Through big data analysis, simulation studies were conducted to enhance management, policy decision-making, and evaluation, thereby driving economic growth. It can be concluded that computational social science is a byproduct of the advancement in scientific, social, and economic knowledge, data collection and analysis, network and computing infrastructure, and algorithmic models.

Computational social science is a byproduct of advancements in scientific, social, and economic knowledge, data collection, network infrastructure, and algorithmic models. From



FIGURE 2.38 – 2016–2020 cooccurrence network [8].

2016 to 2020, big data computing’s frequency and impact increased, expanding research domains into interdisciplinary fields like big data sociology and political science.

The keyword co-occurrence matrix encompasses three subnetworks, with "social science" holding a central position, alongside simulation models, social impact, and research performance. Current research focuses on problem-solving, efficient knowledge organization, and effective fusion computation in the vast web big data environment. Data-driven estimation of segregation levels from noisy data is crucial in computational social science. (Figure 2.39)[8].

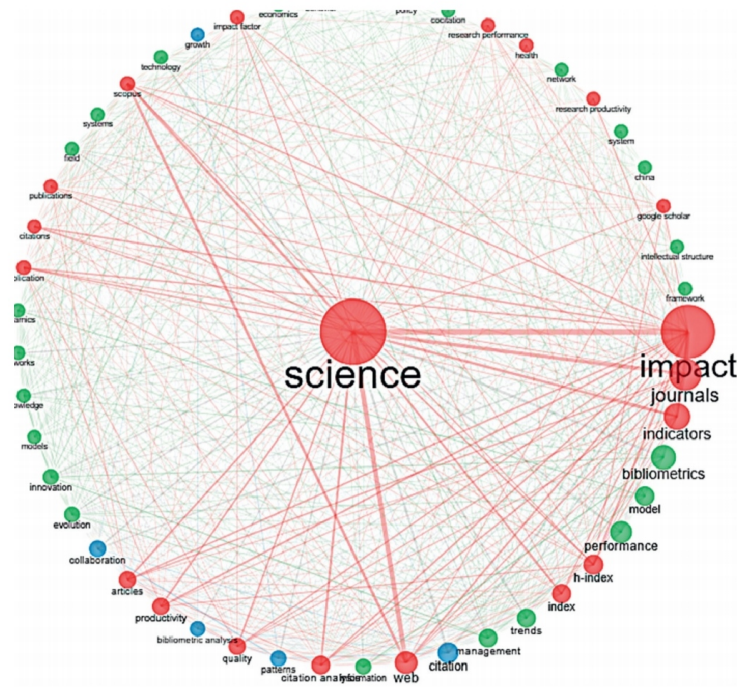


FIGURE 2.39 – Keywords cooccurrence network[8].

## 2.6 example 5 : Infranodus : generate insight wifh ai and network thinking

### 2.6.1 EXPLORING AN AI TOOL

We chose a positively reviewed AI-driven tool called InfraNodus for academic text analysis in this exploratory study. When InfraNodus first became available, its developers focused on its ability to analyze text networks and use AI to extract meaningful patterns from them. The tool provides graphical representations of relationships between topics and thematic clusters, making it central to our exploratory study on facilitating scholarly literature reviewing and research hypothesis generation[9].

Research skills are challenging to teach in a way that is meaningful to students and has ongoing impact in research practice. This paper investigates constructivist and experiential strategies for effective learning and deep understanding of postgraduate research skills and proposes a game-based learning (GBL) solution. A (non-digital) game called *How to Fail Your Research Degree* was designed and iteratively developed. Gameplay loop analysis identifies various learning and game mechanics and contextualises them in relation to GBL theory. Evaluation of gameplay (n = 127) demonstrates effective transmission of intended learning outcomes and positive game experience based on Keller's Attention-Relevance-Confidence-Satisfaction (ARCS) model. Discussion proposes that the game has high cognitive authenticity, relies heavily on tutor facilitation, can create tension between knowledge and confidence, and is applicable to multiple domains and learning situations. GBL is proposed to be an original and effective approach to teaching high-level, functional learning outcomes such as academic research skills.

FIGURE 2.40 – Abbott, D. (2019). Game-Based Learning for Postgraduates[9].

As new users of InfraNodus, we initially examined a single abstract by Abbott (2019) to understand the tool's functionalities. According to InfraNodus developers, it can visually represent an abstract as a network, with main topical groups created from conceptual "nodes" identified through AI analysis of the text. As of September 2022, abstracts were limited to 136 words, necessitating their division into two or more text boxes. Multiple texts derived from the same abstract were included as separate abstracts in the web created by InfraNodus [9]. The results are displayed in Figure 2.41.

Then, using this divided abstract, we tested the tool's ability to generate questions using the



FIGURE 2.41 – Text analysis network based on Abbott's (2019) article created by the AI[9].

GPT-3 artificial intelligence, the third-generation "Generative Pre-trained Transformer". While

GPT-4 is now widely used and gradually integrated, GPT-3 is still utilized for InfraNodus' lower subscription levels as of the writing of this text. The first question generated by the artificial intelligence was "What is the most effective approach to teaching that results in high academic outcomes?" (see Figure 2.42). To assess the appropriateness and fidelity of the AI question generator to the abstract, we tried the "more questions" feature. Figure 2.43 illustrates how the subsequent question generated differed from the initial one[9].



FIGURE 2.42 – Question generated by InfraNodus based on Abbott’s (2019)[9].

The AI feature that regenerates questions has been activated, but the resulting questions may vary significantly in meaning. During trials of the InfraNodus feature for summarizing scientific articles on play-based learning, some generated summaries did not accurately reflect the original text. Additionally, dividing long texts into smaller parts in InfraNodus could impact analysis due to the lack of connection between the last sentence of one part and the first sentence of the next[9].

The question regeneration feature has been activated in Inferencodis, resulting in generating questions with varying meanings. The feature for summarizing texts and dividing long texts has also been clarified. The developer explained how to identify and distinguish between important contracts and clear or influential contracts. In March 2023, the name of the "Detecting Unclear Contracts" feature was changed to "Identifying Key Ideas," but its function remained similar in both cases[9].

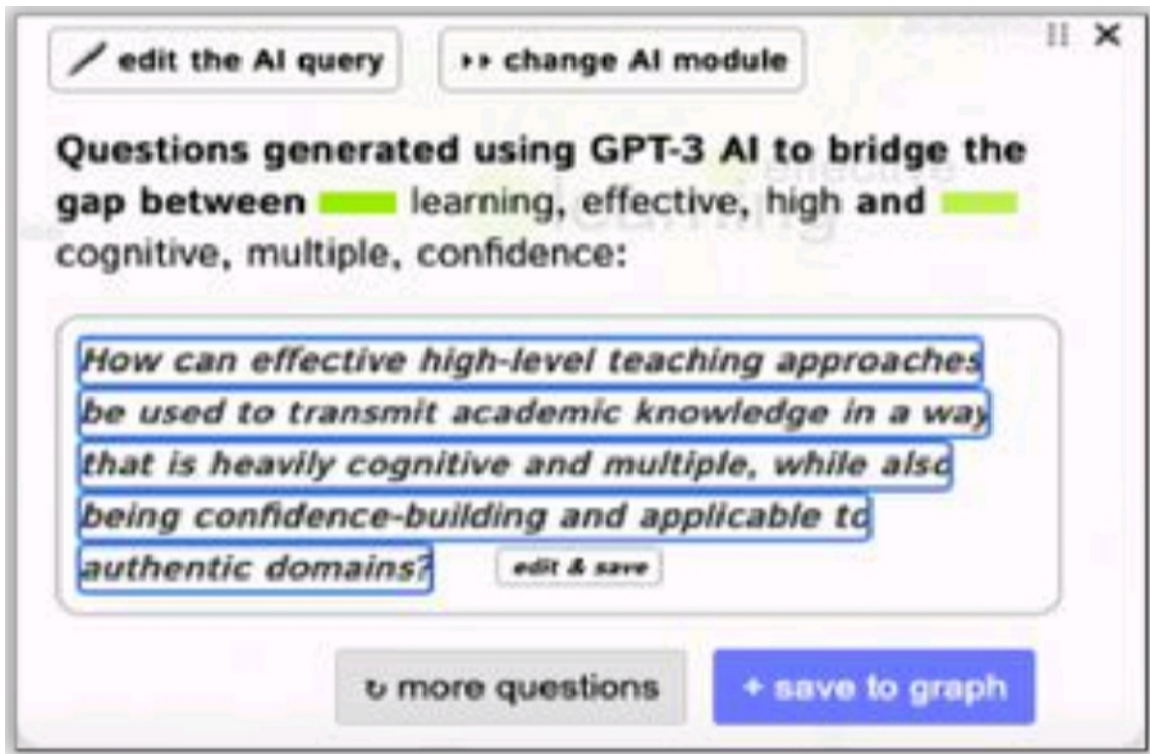


FIGURE 2.43 – Additional questions generated [9].



FIGURE 2.44 – The topical network based on fifteen abstracts[9].

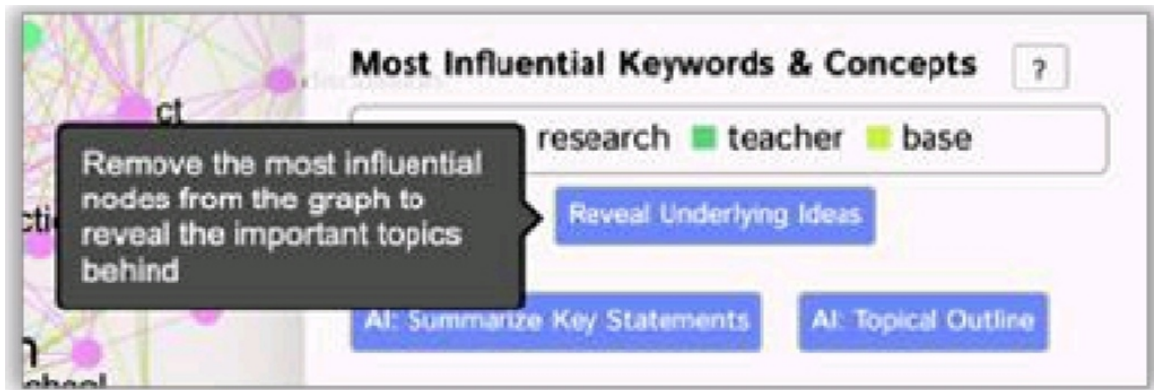


FIGURE 2.45 – Most influential concepts offered by the tool through "Reveal underlying ideas" feature. Previously this feature was called "Reveal Non-obvious" [9].

We tested the InfraNodus tool in creating and regenerating summaries of several articles and found that sometimes the regenerated summaries had different phrasing and meanings from the original articles. For example, the summary mentioned that Seg and Fligby are data collection tools, but in reality, Fligby is a serious game designed to simulate business management processes. Additionally, questions were raised to the developers about how GPT-3 is trained and how articles are transformed into summaries. The variations in summarization and question generation by InfraNodus impact the tool’s effectiveness in analyzing research materials, as the generated questions and summaries may not be deeply connected to the original content. Relying on "structural gaps" for question generation raises questions about the criteria for selection and understanding. While the developer advertises the tool’s capabilities in identifying main topics and generating summaries and research questions, there seems to be insufficient knowledge about the algorithms of GPT-3 used, raising doubts about the accuracy and quality of the generated results[9].

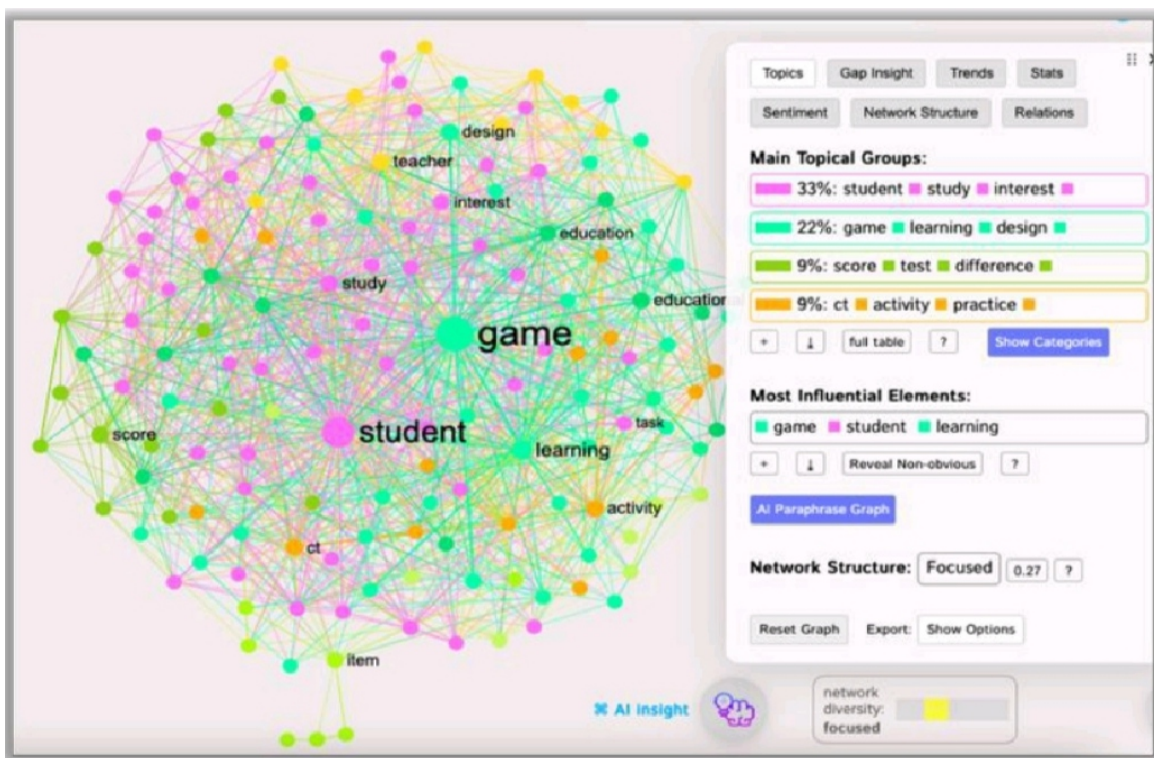


FIGURE 2.46 – The topical network based on fifteen full papers. Note that ‘ct’ and ‘item’ are identified as ‘nodes’[9].

Initial attempts to upload all fifteen articles were unsuccessful due to difficulties in merging and compressing PDF files because of images and signatures within them. All articles were copied and pasted into .txt format instead, which is a time-consuming process. Converting files from

PDF to .txt requires an easy and fast tool, and it is not mentioned how references, tables, and image descriptions might affect the analysis.

The study illustrates how including some references and all graphs, tables, and image descriptions affects the outcome. InfraNodus provided several different summaries for the fifteen articles, and it was found that more input can lead to generating more questions and summaries before repetition. There is a difference between the summaries generated from abstracts and those generated from full articles, with summaries from full articles using more precise phrases. Further exploration is needed to understand how these differences impact knowledge representation in InfraNodus [9].

## 2.7 Conclusion

The examples highlighted in this chapter underscore the importance of using computer techniques in the field of conceptual network analysis. Through computing, the process of examining and analyzing large datasets can be facilitated and accelerated. With the continuous advancement in information technology, we anticipate further innovation in this field to achieve additional progress in conceptual network analysis.

---

---

# CHAPITRE 3

---

## EXPLAINING THE WORK AND PRESENTING THE RESULTS

### **3.1 Introduction**

In this chapter, we will discuss the analysis of the network of concepts in selected samples of Arab and foreign educational programs. Computing methods and techniques are used to analyze and compare these networks, allowing the differences and similarities between educational approaches in each to be identified.

### **3.2 Plan for workflow steps**

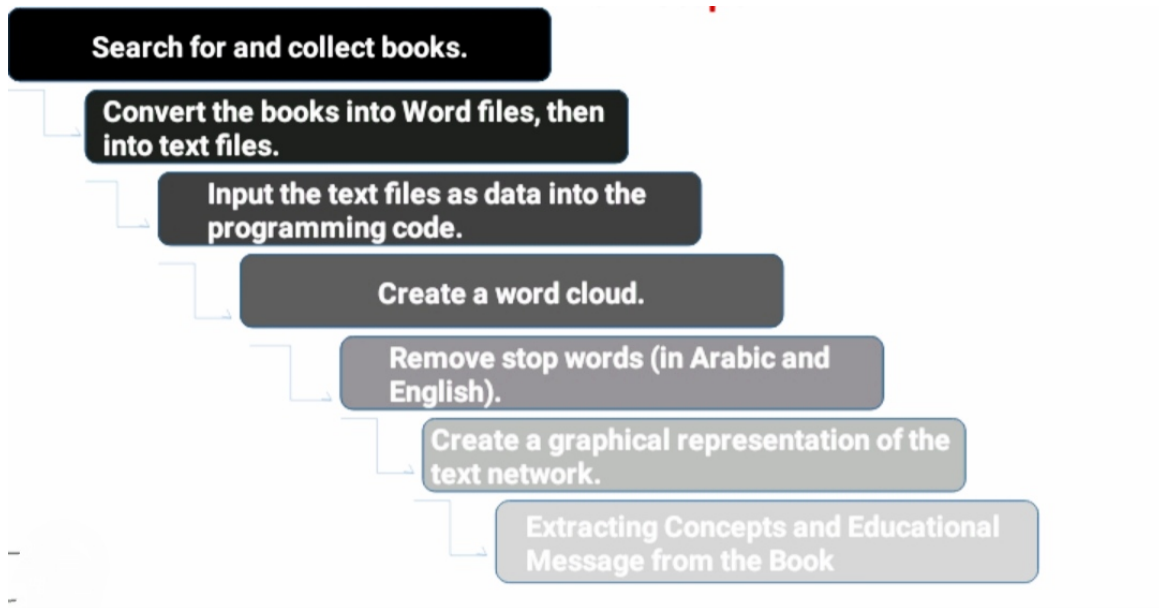


FIGURE 3.1 – workflow steps

### 3.3 Computer study of textbooks (statistics for comparison)

The number of raw words after converting to Word	the level	stage	Book title, publishing house information, and year of printing	Country	the language	Category
2181	1	primary	اشراق 2023, لغتي هويتي	Qatar	Arabic	Arabic
6304	2		اشراق 2023, لغة عربية			
6720	3		اشراق 2023, لغة عربية			
9834	4		اشراق 2023, لغة عربية			
7810	5		اشراق 2023, لغة عربية			
11264	6		اشراق 2023, لغة عربية			
853	1	primary	الديوان الوطني للمطبوعات المدرسية 2021/2020, لغة عربية	Algeria	Arabic	Arabic
3067	2		الديوان الوطني للمطبوعات المدرسية 2021/2020, لغة عربية			
3002	3		الديوان الوطني للمطبوعات المدرسية 2021/2020, لغة عربية			
6035	4		الديوان الوطني للمطبوعات المدرسية 2017/2018, لغة عربية			
2510	5		الديوان الوطني للمطبوعات المدرسية 2019/2020, لغة عربية			
4514	6		الديوان الوطني للمطبوعات المدرسية 2019/2020, لغة عربية			
2958	1	primary	دار نهضة مصر للنشر 2024/2023, لغة عربية	Egypt	Arabic	Arabic
3536	2		دار نهضة مصر للنشر 2024/2023, لغة عربية			
5841	3		دار نهضة مصر للنشر 2021/2020, لغة عربية			
9521	4		دار نهضة مصر للنشر 2024/2023, لغة عربية			
5456	5		دار نهضة مصر للنشر 2024/2023, لغة عربية			
7526	6		دار نهضة مصر للنشر 2024/2023, لغة عربية			
98932						the total
1705	1	primary	Cambridge primary English Learner's Book, Cambridge University prees 2015	United kingdom	English	foreign
3257	2	primary	Cambridge primary English Learner's Book, Cambridge University prees 2014			
2457	3	primary	CAMBRIDGE GlobalEnglish Learner's Book, Cambridge University prees 2014			
770	4	primary	CAMBRIDGE GlobalEnglish Learner's Book Cambridge University prees 2014			
6640	5	primary	CAMBRIDGE GlobalEnglish ActivityBook , Cambridge University prees 2014			
3878	6	primary	CAMBRIDGE GlobalEnglish Learner's Book, Cambridge University prees 2014			
5460	1	primary	English time , Oxford University prees	America	English	foreign
4856	2		English time , Oxford University prees			
6735	3		English time , Oxford University prees			
6123	4		English time , Oxford University prees			
9157	5		English time , Oxford University prees			
9810	6		English time			
3551	1	primary	Every body, Oxford University prees 2012	America	English	foreign
4570	2		Every body, Oxford University prees 2012			
5816	3		Every body, Oxford University prees 2012			
6801	4		Every body, Oxford University prees 2012			
10664	5		Every body, Oxford University prees 2012			
2471	6		Every body, Oxford University prees 2012			
93763						the total

FIGURE 3.2 – Statistics table

### 3.4 A comparison between school books (reading books) for the primary stage in Qatar and America

#### 3.4.1 Computerized processing of the books of the State of Qatar

First, we processed book texts to create a word cloud, a visual representation of text content based on the frequency of words used in it – see Figure 3.3.



FIGURE 3.3 – Word cloud before processing.

Figure 3.3 shows the dominance of prepositions and pronouns over the concepts contained in it, as a result of their extensive use in constructing sentences and linking words. The next step is to use intelligent algorithms to delete redundant words (Stop Words) that do not have a separate meaning, highlighting key concepts and removing noise from the text - see Figure 3.4.

The new word cloud shows the emergence of concepts that were part of the cognitive framework



FIGURE 3.4 – Post-processed word cloud.

of education in Qatar. We notice the emergence of terms such as :

اقرأ، أجب، الأفعال، القصة، قراءة، نص، بلد، الجمل، بدائل،

and others. The size of these items varies, reflecting their frequency within the text.

### 3.4.2 Analysis of the books network for the State of Qatar

The word elements that make up book texts represent the nodes that form the core of the mesh fabric. As for the edges, they describe the relationships that link the text’s vocabulary, which leads to the formation of terms and concepts through these close connections.

#### Text network data analysis

Word	Repetition
قطر	96
اقرا	86
النص	85
القصة	74
القصة	74
جملة	52
القصيدة	52
الحيوانات	50
الحروف	47
المدرسة	40
الارض	39
استخدام	35
الارشادات	34
الاطفال	32

FIGURE 3.5 – A table representing the frequency of the words that appear most frequently in Qatar

It is clear from Figure 3.5 that there is a discrepancy in the importance of the terminological vocabulary that populates Qatar’s books, where the word “قطر” occupies the first place, while the word “أطفال” comes in the last place.

These results can be used as a guide in determining the author’s level of interest in this vocabulary and treating it in more detail than the rest of the vocabulary.

### Graphical representation of the text network

The network illustrates the density of concepts used in educational programs and provides a comprehensive view of the nature of the educational content. It shows that programs in Qatar place significant emphasis on teaching reading and writing, educational morals, and poetry

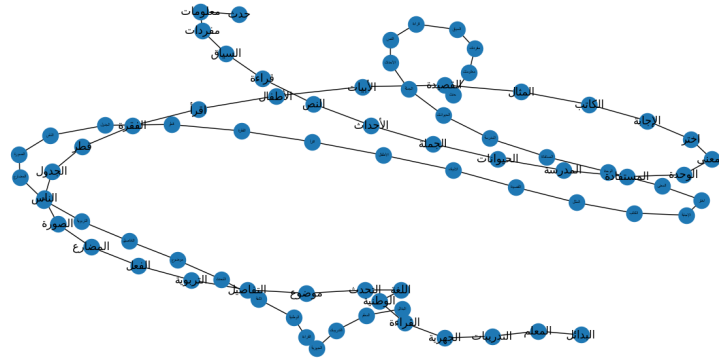


FIGURE 3.6 – Text content grid

### 3.4.3 Computer processing of American state books (English time)

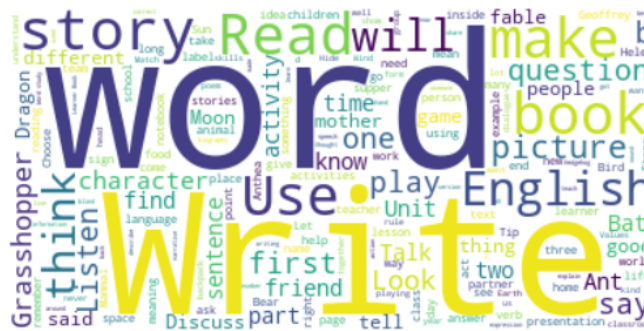


FIGURE 3.7 – Word cloud before the cleaning process

From Figure 3.7 we notice the presence of words that have no importance, in addition to individual letters of the alphabet, so we had to remove them in order to obtain concepts of greater importance 3.8.



Word	Repetition
practice	229
write	226
read	159
conversation	133
play	113
review	111
school	103
word	100
picture	96
thanks	55
reporter	47
question	26
mother	24

FIGURE 3.9 – A table representing the frequency of the words that appear most frequently in American books (English time)

**Graphical representation of the text network**

The network illustrates the density of concepts used in educational programs and provides a comprehensive view of the nature of the educational content. It shows that "English Time" programs prioritize teaching reading, writing, educational ethics, and family values

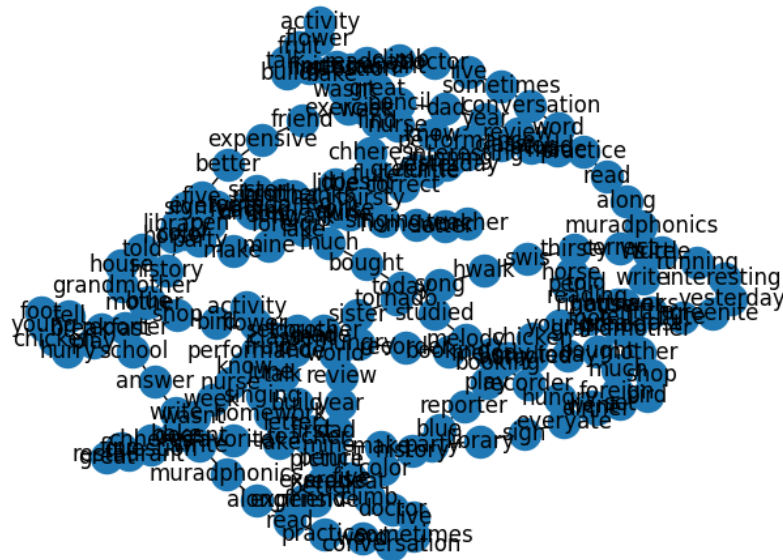


FIGURE 3.10 – Word network

### 3.5 comparison between school books (reading books) for the primary stage in Egypt and America :

#### Forming a word cloud for Egyptian books

A word cloud before processing is shown in Figure 3.11.



FIGURE 3.11 – Word cloud before filtering

What is clear from the cloud is that the texts of the books before treatment contained many prepositions, pronouns and words that do not belong to the terminological field of education.

After treatment, terms related to education began to appear more clearly. See Figure 3.12.



FIGURE 3.12 – Word cloud after filtering

After the prepositions, pronouns, and words that lacked a meaning in and of themselves were squared in Figure 3.12 , we see that they were absent after the treatment processes, so important terms emerged such as **المدرسة ، الفصل ، الجمل ، العربية ، اللغة ، المدرسة ، أكمل ، اكتب ، الأهداف**

### 3.5.1 Analysis of the books network for the State of Egypt

#### Text network data analysis

It appears clear from Figure 3.13 that there is a discrepancy in the importance of the terminological vocabulary that has settled in Egyptian books, where the word “الأهداف” occupies the first place, while the word “الجملة” comes in the last place. These results can be used as a guide in determining the author’s level of interest in this vocabulary and treating it in more detail than the rest of the vocabulary.

Word	Repetition
الأهداف	100
اكتب	86
الانسان	67
اللغة	56
العربية	52
الشاعر	48
الابتدائي	46
لاحظ	45
الحيوانات	43
امي	39
القراءة	38
الصف	33
الجملة	31

FIGURE 3.13 – A table representing the frequency of the words that appear most frequently in Egypt

### Graphical representation of the text network

The network illustrates the density of concepts used in educational programs and provides a comprehensive view of the nature of the educational content. It shows that programs in Egypt place significant emphasis on teaching reading and writing, educational morals, family values, and poetry

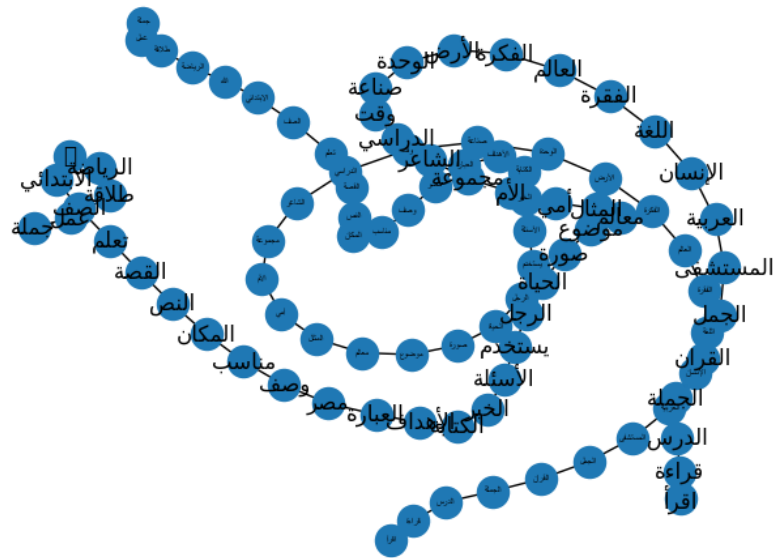


FIGURE 3.14 – Concept network

### 3.5.2 Computer processing of American state books (Every body)

From Figure 3.15 we notice the presence of words that have no importance, in addition to individual letters of the alphabet, so we had to remove them in order to obtain concepts of greater importance. See Figure 3.16

We notice the presence of terms such as : practice ,answer , read, ask ,play , School, number and others. The size of these elements varies, reflecting their repetition within the text



Word	Repetition
ask	156
answer	148
practice	100
unit	79
read	73
talk	59
school	50
thanks	45
number	40
story	39
play	36
oclock	23
color	20
family	9

FIGURE 3.17 – A table representing the frequency of the words that appear most frequently in American (Every body)

### Graphical representation of the text network

The network illustrates the density of concepts used in educational programs and provides a comprehensive view of the nature of the educational content. It shows that "Every Body" programs prioritize teaching reading, writing, educational ethics, and family values.

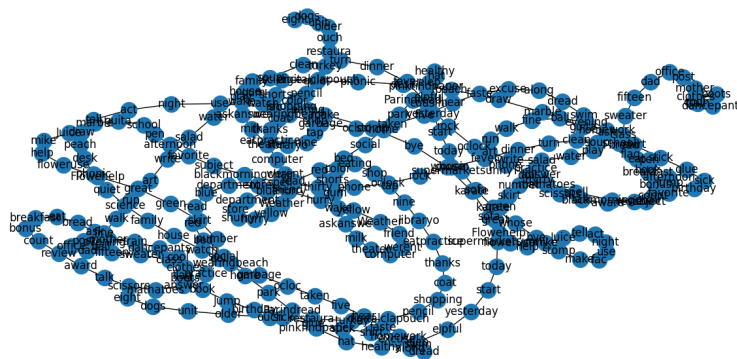


FIGURE 3.18 – Concept network





FIGURE 3.20 – Word cloud after the cleaning process

### 3.6.2 Analysis of the books network for the Arab countries :

#### Text network data analysis :

It is clear from Figure 3.21 that there is a difference in the importance of the terminological vocabulary that populated the books of arab , where the word "الوحدة" occupies the first place, while the word "التلاميذ" comes in the last place. These results can be used as a guide in determining the author’s level of interest in this vocabulary and treating it in more detail than the rest of the vocabulary.

Word	Repetition
الوحدة	219
النص	176
الله	129
الجملة	120
القصة	112
الاهداف	110
القراءة	103
الحيوانات	100
الدرس	95
المدرسة	83
الناس	73
العربية	70
اللغة	68
ابي	65
الكتابة	63
الارض	60
الحياة	56
الماضي	54
اكتب	48
الرياضة	47
المعلم	44
التلاميذ	32

FIGURE 3.21 – A table representing the frequency of the words that appear most frequently in Arab

### Graphical representation of the text network

The network sheds light on the density of concepts used in educational programs and provides a comprehensive view of the nature of the educational content. It appears that educational programs in Arab countries focus significantly on teaching reading, writing, and educational ethics, as well as moral, religious, and national values

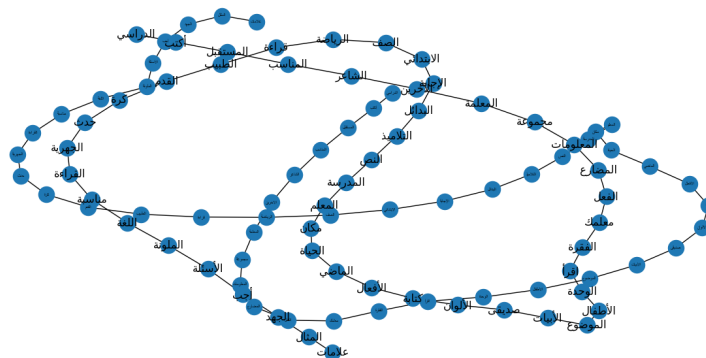


FIGURE 3.22 – Concept network



### 3.6.4 Analysis of the books network for the foreign countries

#### Text network data analysis

Word	Repetition
practice	343
write	330
read	311
answer	233
ask	221
school	170
play	199
word	156
review	149
unit	147
conversation	146
talk	142
circle	142
story	140
picture	134
book	120
help	100
use	95
friend	65
thank	63

FIGURE 3.25 – A table representing the frequency of the words that appear most frequently in foreign

It is clear from Figure 3.25 that there is a discrepancy in the importance of the conventional vocabulary that has settled in foreign books, where the word “write” occupies the first place, while the word “thank” comes in the last place. These results can be used as a guide in determining the author’s level of interest in this vocabulary and treating it in more detail than the rest of the vocabulary.



## **3.7 results**

Based on the analysis of the concept network, we can conclude that Arab and foreign educational programs share many common goals and core contents such as teaching ethics, reading, and writing. However, differences emerge in cultural and societal contexts. For example, Arab programs emphasize religious education and fostering patriotism as part of their curriculum, while foreign programs may focus on different values or diverse cultural orientations. These differences reflect the varied interests and values that each country seeks to promote through its educational system, highlighting the broad cultural and social diversity worldwide

## **3.8 Conclusion**

This is the competent educational method in education to better understand the content of educational curricula and design more effective and effective programs. It also helps in identifying the strengths and weaknesses of each educational program, thus directing work towards improving educational performance and quality of education.

---

---

# CHAPITRE 4

---

## ENVIRONMENT AND OFFICES USED FOR DEVELOPMENT

### 4.1 Introduction

In this chapter, we will delve into programming languages and the specialized tools required for implementing and realizing this project. We will utilize the Python programming language to materialize our project using the pychmar development environment. Additionally, we will provide images of the code used in the project to clarify the programming process.

### 4.2 Python

Python is a very versatile and widely used programming language, popular in a variety of fields such as web development, data science, and artificial intelligence. It is well designed and can be used for real world programming[23]. It is a very high-level, dynamic, object-oriented, compiler-based programming language that can be used in very diverse fields. Python is designed to be easy to understand and use, and has recently been considered a user- and beginner-friendly language. Python has gained popularity as a beginner's language and has taken Java's place as the most popular educational language. As a dynamic type language, Python is really flexible. Furthermore, Python allows for fault tolerance, allowing you to compile and run your program until you get to the problem part. This language can support different styles of programming including structured and object-oriented programming. Other styles can

be used as well. Python is very flexible, due to its ability to use modular components that were designed in other programming languages. For example, you can write a program in C++ and import it into Python as a module. Then add something else to it (eg design a GUI for it)[24].



FIGURE 4.1 – python logo[10].

### 4.3 NLTK

NLTK is a Python library designed for fast and easy processing of Natural Language text. Initially developed at the University of Pennsylvania, it serves as both a research and teaching tool for natural language processing. NLTK comes with an extensive collection of built-in corpora, encompassing various types of textual materials such as novels, news articles, network chat transcripts, movie reviews, and more. This includes notable corpora like Brown's Corpus, Gutenberg's Corpus, Inaugural Address Corpus, and Reuters Corpus (Kambhampati, 2019). Additionally, NLTK offers user-friendly interfaces to access over 50 corpora and lexical resources such as WordNet. Furthermore, NLTK provides a comprehensive suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning. It also includes wrappers for integrating with industrial-strength NLP libraries and an active discussion forum. By collaborating with Python's robust standard library and third-party libraries, NLTK facilitates secondary processing of its results [25].



FIGURE 4.2 – nltk logo[11].

## 4.4 Matplotlib

Matplotlib stands out as one of Python’s most widely utilized data visualization libraries. Originally crafted by John Hunter, alongside numerous contributors, it has undergone significant development to become an essential tool for scientists and researchers worldwide. Matplotlib serves as a crucial component within the Python data science ecosystem, seamlessly integrating with NumPy, Pandas, and other pertinent libraries[26].



FIGURE 4.3 – matplotlib logo[12].

## 4.5 Word Cloud

A word cloud is essentially a visual representation of text, commonly used in various contexts, particularly in text analytics. Typically, word clouds are generated from a body of text, offering a summarized view of isolated words without considering their linguistic meaning or relationships. They are often employed in text summarization and provide insights into the frequency and relevance of words within a specific context or research area. Word clouds, along with their variations like letter clouds, offer statistical analysis but usually lack interactive

capabilities[27].



FIGURE 4.4 – Word Cloud logo [13].

## 4.6 Networkx

NetworkX is a Python package that enables the creation, manipulation, and examination of complex networks. It utilizes a "dictionary of dictionary" data structure and is implemented purely in Python. Initially developed by Aric Hagberg, Dan Schult, and Pieter Swart between 2002 and 2003, it was officially released in April 2005, with its background elucidated in a research paper by Hagberg et al. (2008). NetworkX stands out as the most prominent Python package for network analysis, offering a plethora of algorithms and comprehensive documentation. Several books, including Caldarelli and Chessa (2016), Al-Taie and Kadry (2017), Fouss et al. (2016), and Tsvetovat and Kouznetsov (2011), are grounded in its capabilities. Widely regarded as the standard tool for network analysis in Python, NetworkX is particularly suited for analyzing small to medium-sized networks containing up to millions of nodes[28].



FIGURE 4.5 – networkx logo[14].

## 4.7 Arabic reshaper

The Arabic reshaper library is a Python library used for reshaping Arabic text and right-to-left (RTL) text in a proper and Arabic language standards-compliant manner. This library helps improve the representation of Arabic text on graphical interfaces and in texts relying on Arabic text encoding, making it more readable and understandable[29].



FIGURE 4.6 – Arabic reshaper logo [15].

## 4.8 Bidi algorithm

The Bidi (Bidirectional) algorithm is used to process text written in languages that are read from right to left (such as Arabic and Hebrew) and left to right (such as English) within the same text. This algorithm is essential for determining the correct order of characters and words to ensure proper reading and display of text in computer interfaces and applications

The basic principle of the Bidi algorithm is to divide the text into different levels of directionality (from right to left and vice versa), and then apply the appropriate ordering for characters, words, sentences, and paragraphs based on the direction of each level. This is achieved by using special directionality markers added to the text to specify the correct direction for each part of the text.

The goal of the Bidi algorithm is to provide a unified and effective way to display texts written in different languages correctly on digital interfaces, making it easier for users to understand the text and interact with applications more effectively[30].



VectorStock®  
VectorStock.com/2796879

FIGURE 4.7 – Bidi algorithm logo[16].

## 4.9 pycharm

PyCharm is a popular integrated development environment (IDE) for Python programming, featuring several developer-friendly features, including[17] :

- Smooth code completion for both built-in and external packages.
- Support for breakpoints and viewing SQL representations of user expressions.
- Easy checking of recent changes with highlighted differences between the last commit and the current state.
- Ability to run .py files outside the editor with code coverage tracking.
- Visual display and management of installed packages.
- Local history tracking changes similar to Git.
- Smooth refactoring support with various shortcuts.
- A comprehensive user interface for creating and importing projects.



FIGURE 4.8 – pycharm logo [17].

## 4.10 Some parts of the codes used

```
10 with open("English time.txt", "r", encoding="utf-8") as file:  
11     text = file.read()
```

FIGURE 4.9 – Read a file in English

```
12 # Open and read the text file  
13 f = codecs.open(path.join(d, "Arabic.txt"), "r", encoding="utf-8")  
14 text = arabic_reshaper.reshape(f.read())  
15 text = get_display(text)
```

FIGURE 4.10 – Read a file in arabic

```
2 from wordcloud import WordCloud  
3 import matplotlib.pyplot as plt  
4 from os import path  
5 import codecs  
6 import arabic_reshaper  
7 from bidi.algorithm import get_display
```

FIGURE 4.11 – List of offices

```

18 # Create a graph
19 G = nx.Graph()
20
21 # Add nodes to the graph
22 for word in re.findall(r'\w+', bidi_text):
23     G.add_node(word)
24
25 # Add edges to the graph
26 for word1, word2 in zip(G.nodes(), list(G.nodes())[1:]):
27     G.add_edge(word1, word2)
28
29 # Draw the graph
30 nx.draw(G, with_labels=True)
31 pos = nx.spring_layout(G)
32 nx.draw_networkx_nodes(G, pos)
33 nx.draw_networkx_edges(G, pos)
34
35 # Draw node labels
36 for node in G.nodes():
37     label = ''.join(c for c in node if c.isalpha())
38     x, y = pos[node]
39     plt.text(x, y, label, fontsize=5, fontfamily='times new roman', ha='center', va='center', rotation=0)
40
41 plt.show()

```

FIGURE 4.12 – Post-processing word cloud formation

```

68 # عرض سحابة الكلمات
69 plt.imshow(wordcloud, interpolation='bilinear')
70 plt.axis('off')
71 plt.show()
72
73 print("تكرار الكلمات")
74 for word, count in word_counts.items():
75     print(f"{word}: {count}")
76
77 def save_filtered_words_to_file(filtered_words, file_path):
78     with open(file_path, 'w', encoding='utf-8') as file:
79         for word in filtered_words:
80             file.write(f"{word}\n")
81
82 # Save the filtered words to a text file
83 save_filtered_words_to_file(filtered_words, 'filtered_wordseng.txt')

```

FIGURE 4.13 – Display the post-processed word cloud in addition to word repetition

## 4.11 Conclusion

In conclusion, this chapter provided insights into programming languages and the essential tools necessary for the execution and realization of our project. By using the Python programming language within the Visual Studio Code environment, we presented a practical application of our project. Additionally, the inclusion of illustrative images showcasing the code used in the project enhances the understanding of the programming process. In conclusion of this chapter, we emphasize the importance of proficiency in programming languages and the utilization of appropriate tools to effectively implement software projects

---

## GENERAL CONCLUSION

The analytical study that you presented to us in this memorandum highlights the importance of using computer information technologies in analyzing the network of concepts in educational programs, and highlighting the differences and similarities between Arab and foreign programs by omitting the use and results achieved. We realize that these elements are an important part of improving the quality of education and developing studies programs. Based on the analysis of the conceptual network, we can conclude that Arab and foreign educational programs share many common goals and basic contents such as teaching ethics, reading, and writing. However, differences appear in cultural and societal contexts. For example, Arab programs emphasize religious education and the promotion of patriotism as part of their curricula, while foreign programs may focus on different values or diverse cultural orientations. These differences reflect the diverse interests and values that each country seeks to promote through its education system, highlighting the vast cultural and social diversity across the world.

The results show that using computer processing techniques to extract word clouds and network textures can contribute to analyzing the structure of educational programs more effectively, and provide valuable insights into the distribution of concepts and the relationships between them. The study also shows that there are great opportunities to enhance the use and development of these technologies in the Arab context to achieve the best results.

By applying the recommendations resulting from this study, educational programs can be improved and their quality and effectiveness enhanced, which contributes to improving the learning experience for students and preparing them for a better future. With our commitment to continuous research and development in the field of computer processing for concept network analysis, we can achieve sustainable development in education and enhance learning across cultures and societies.

---

# BIBLIOGRAPHIE

- [1] D. Maynard, K. Bontcheva, and I. Augenstein, *Natural language processing for the semantic web*. Springer, 2017.
- [2] S. B. Aher and L. Lobo, “Data mining in educational system using weka,” in *International conference on emerging technology trends (ICETT)*, vol. 3. Foundation of Computer Science, 2011, pp. 20–25.
- [3] J. Han, J. Pei, and H. Tong, *Data mining : concepts and techniques*. Morgan kaufmann, 2022.
- [4] H. M. Al-Razo, *Semantic analysis and graphical description introduction to ibn khaldun computerized processing*, 2023.
- [5] H. Muzaffar, *computerized processing to analyze the network of meanings and concepts in books expressing vision*, 2023.
- [6] *Review of Research on Digital Supply Chain Management Using Network Text Analysis*. [Online]. Available : [https://www.mdpi.com/2071-1050/13/17/9929?type=check\\_update&version=2](https://www.mdpi.com/2071-1050/13/17/9929?type=check_update&version=2)
- [7] J. Pyun and J. S. Rha, “Review of research on digital supply chain management using network text analysis,” *Sustainability*, vol. 13, no. 17, p. 9929, 2021.
- [8] Y. Liu, X. Feng, Y. Zhang, Y. Kong, R. Yang *et al.*, “Paths study on knowledge convergence and development in computational social science : Data metric analysis based on web of science,” *Complexity*, vol. 2022, 2022.
- [9] I. Tursunkulova, S. de Castell, and J. Jenson, “Exploring infranodus : A text analysis tool.” *International Association for Development of the Information Society*, 2023.

- 
- [10] *The Python Logo*. [Online]. Available : <https://www.python.org/community/logos/>
- [11] *Sentiment Analysis — using NLTK Vader*. [Online]. Available : <https://medium.com/@skillcate/sentiment-analysis-using-nltk-vader-98f67f2e6130>
- [12] *File :Created with Matplotlib-logo.svg*. [Online]. Available : [https://commons.m.wikimedia.org/wiki/File:Created\\_with\\_Matplotlib-logo.svg](https://commons.m.wikimedia.org/wiki/File:Created_with_Matplotlib-logo.svg)
- [13] *Python Logo WordCloud Shirt for Programmers on Men's T-Shirt*. [Online]. Available : [https://www.spreadshirt.com.au/shop/design/python+logo+wordcloud+shirt+for+programmers+mens+t-shirt-D598b596af6c60d28beb60147?sellable=ekElAjyZXVHnrY8dvvAO-210-7&view=2\\_6285](https://www.spreadshirt.com.au/shop/design/python+logo+wordcloud+shirt+for+programmers+mens+t-shirt-D598b596af6c60d28beb60147?sellable=ekElAjyZXVHnrY8dvvAO-210-7&view=2_6285)
- [14] *NetworkX*. [Online]. Available : <https://github.com/networkx>
- [15] *Lib Arabic Reshaper*. [Online]. Available : <https://www.curseforge.com/wow/addons/lib-arabic-reshaper>
- [16] *Symbol bidirectional arrows data transfer vector image*. [Online]. Available : <https://www.vectorstock.com/royalty-free-vector/symbol-bidirectional-arrows-data-transfer-vector-27162879>
- [17] *Pycharm*. [Online]. Available : <https://www.linkedin.com/pulse/pycharm-darshika-srivastava>
- [18] F. Krieche, “Computational linguistics and automated processing of natural languages,” *The Egyptian Journal of Language Engineering*, vol. 9, no. 1, pp. 32–47, 2022.
- [19] D. Mahmoud, *Arabic Computational Linguistics*.
- [20] M. M. I. Khalil and S. F. s. Naaja, “Attributing verbs to pronouns in the light of computational linguistics,” Ph.D. dissertation, 2011.
- [21] M. Raafat Mahmoud Farag and Muhammad, “Automated language processing techniques and their development,” *Journal of the Faculty of Arts in Qena*, vol. 31, no. 54, pp. 537–594, 2022.
- [22] S. Lagraa, “New mp-soc profiling tools based on data mining techniques,” Ph.D. dissertation, Grenoble, 2014.
- [23] O. B. Elhalid, Z. Alm Alhelal, and S. Hassan, “Exploring the fundamentals of python programming : A comprehensive guide for beginners,” *SAMER, Exploring the Fundamentals of Python Programming : A Comprehensive Guide for Beginners (October 25, 2023)*, 2023.
- [24] K. Srinath, “Python—the fastest growing programming language,” *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 354–357, 2017.

- [25] M. Wang and F. Hu, “The application of nltk library for python natural language processing in corpus research,” *Theory and Practice in Language Studies*, vol. 11, no. 9, pp. 1041–1049, 2021.
- [26] A. H. Sial, S. Y. S. Rashdi, and A. H. Khan, “Comparative analysis of data visualization libraries matplotlib and seaborn in python,” *International Journal*, vol. 10, no. 1, p. 45, 2021.
- [27] A. I. Kabir, K. Ahmed, and R. Karim, “Word cloud and sentiment analysis of amazon earphones reviews with r programming language,” *Informatica Economica*, vol. 24, no. 4, pp. 55–71, 2020.
- [28] V. Batagelj, R. Alhajj, and J. Rokne, “Python packages for networks.” 2018.
- [29] *arabic-reshaper 3.0.0*. [Online]. Available : <https://pypi.org/project/arabic-reshaper/>
- [30] *UNICODE BIDIRECTIONAL ALGORITHM*. [Online]. Available : <https://www.unicode.org/reports/tr9/tr9-18.html>