University of Kirkuk Journal For Administrative and Economic Science. A special issue of the 1st Scientific Conference of the College of Administration and Economics - University of Kirkuk - Information technology, digitization, and their impact on sustainable development - 8-9, Oct- 2025.

IRAQI Academic Scientific Journals

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521
University of Kirkuk Journal For Administrative and Economic Science

Hussein Yildirim Salahaldin Hussein & Albayati Ali Ibrahim Mohammed. Search: AI, Statistics, and Digital Administration for Intelligent Management of Unstructured Data. ***University of Kirkuk Journal for Administrative and Economic Science*** **Volume (15) Issue (4) Part (2) Supplement (1) A special issue of the 1st Scientific Conference of the College of Administration and Economics - University of Kirkuk - Information technology, digitization, and their impact on sustainable development - 8-9, Oct- 2025, p-p:380-388.**

Search: AI, Statistics, and Digital Administration for Intelligent Management of Unstructured Data

Yildirim Salahaldin Hussein Hussein [1], Ali Ibrahim Mohammed Albayati [2]

[1] *University of Kirkuk–College of Administration and Economics/Department of Statistics, Kirkuk, Iraq*
[2] *Imam Ja'afar Al-Sadiq University – College of Financial Administrative Sciences – Department of Oil and Gas Economics, Kirkuk, Iraq*

yildirimsalahaldin@uokirkuk.edu.iq [1]
ali.ibrahim@ijsu.edu.iq [2]

Abstract: The academic and administrative domains are currently witnessing a massive growth in the volume of unstructured data, which includes textual and scanned documents such as PDF, DOCX, JPG, and JPEG files. This rapid increase renders traditional search systems incapable of meeting the speed and accuracy requirements for effective information retrieval. To address this challenge, the Search program was developed as an innovative platform that integrates Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques, combined with statistical indicators to evaluate performance and retrieval efficiency.

This paper aims to highlight both the statistical and administrative aspects of the program through an applied study conducted on a large-scale archive consisting of 61,777 files organized into 7,061 main and subfolders. The experiment involved two main tests: the first used a frequent keyword found in 226 files, where the program achieved a 96% success rate within 38 minutes and 23 seconds. The second used a rare keyword found in 3 files, where the program achieved a 100% success rate within 4 minutes and 41 seconds.

The findings demonstrate the program's effectiveness in managing unstructured archives and its potential to enhance digital administration by accelerating access to information and improving decision-making quality. Moreover, the results underscore the importance of future development, particularly through the integration of High-Performance Computing (HPC) to increase response speed and scalability. It is expected that the program will contribute to establishing a robust technological infrastructure to support academic and administrative sectors in Iraq and the wider region.

Keywords: Artificial Intelligence; Natural Language Processing; Archive Search; Optical Character Recognition; Digital Administration; Statistical Analysis; High-Performance Computing.

SEARCH: AI, الإحصاء والإدارة الرقمية للإدارة الذكية للبيانات الغير المنظمة

م.م. يلدرم صلاح الدين حسين [١]، م.م. علي إبراهيم محمد البياتي [٢]

[١] جامعة كركوك-كلية الإدارة والاقتصاد/ قسم الإحصاء، كركوك، العراق
[٢] جامعة الامام جعفر الصادق (ع) – كلية العلوم الادارة والمالية/قسم الاقتصاديات النفط والغاز، كركوك، العراق

المستخلص: يشهد العالم الأكاديمي والإداري نمواً هائلاً في حجم البيانات الغير المنظمة، والتي تشمل وثائق نصية وصورية مثل ملفات (PDF, DOCX, JPG, JPEG). هذا التزايد السريع يجعل أنظمة البحث التقليدية عاجزة عن تلبية متطلبات السرعة والدقة في الوصول إلى المعلومات. وانطلاقاً من هذا التحدي، تم تطوير برنامج Search كمنصة ذكية تعتمد على مزيج من تقنيات التعرف الضوئي على الحروف (OCR) ومعالجة اللغة الطبيعية (NLP)، إضافة إلى مؤشرات إحصائية لقياس الأداء وكفاءة الاسترجاع.

تهدف هذه الورقة إلى استعراض الجانبين الإحصائي والإداري للبرنامج، من خلال دراسة تطبيقية أجريت على أرشيف ضخم يضم (٦١,٧٧٧ ملفاً) موزعة على (٧,٠٦١ مجلداً رئيسياً وفرعياً). شملت التجربة اختبارين رئيسيين: الأول باستخدام كلمة مفتاحية شائعة ظهرت في (٢٢٦ ملفاً)، حيث بلغت نسبة النجاح (٩٦٪) بزمن (٣٨:٢٣ دقيقة)، والثاني باستخدام كلمة مفتاحية نادرة ظهرت في (٣ ملفات)، حيث حقق البرنامج نسبة نجاح (١٠٠٪) بزمن (٤:٤١ دقيقة).

تكشف هذه النتائج عن كفاءة البرنامج في التعامل مع الأرشيفات العشوائية، وقدرته على تعزيز الإدارة الرقمية من خلال تسريع الوصول إلى المعلومات وتحسين جودة القرار. كما تسلط الضوء على أهمية التوسع المستقبلي في تطوير البرنامج، لاسيما عبر دعم حوسبة عالية الأداء (High Performance Computing) بما يرفع من سرعة الاستجابة ويزيد من شمولية النتائج. ومن المتوقع أن يسهم البرنامج في إرساء بنية تحتية تقنية قوية تخدم المجالات الأكاديمية والإدارية في العراق والمنطقة.

الكلمات المفتاحية: AI, معالجة اللغة الطبيعية, الأرشيف, التعرف الضوئي على الحروف, الإدارة الرقمية, التحليل الإحصائي, الحوسبة عالية الأداء.

Corresponding Author: E-mail: yildirimsalahaldin@uokirkuk.edu.iq

Introduction

The contemporary world is undergoing a profound transformation in the nature of data generated and stored within academic and administrative institutions. Statistics indicate that more than **80% of organizational data is unstructured**, encompassing textual documents, scanned images, and multi-format files such as PDF, DOCX, JPG, and JPEG (Smith, 2020; Kitchin, 2014). This rapid growth of unstructured data poses significant challenges for researchers and decision-makers, particularly in environments that rely on vast archives accumulated over many years (Rowley, 2007).

From an **administrative perspective**, Knowledge Management represents a cornerstone in enhancing institutional efficiency and competitiveness. However, traditional cataloging and search systems often fail to cope with the enormous volume of unindexed documents, leading to delays in accessing information and weakening the decision-making process (Dalkir, 2017; Davenport & Prusak, 1998). Consequently, there has been a growing need for more intelligent tools that integrate **digital administration**, **statistical analysis**, and **artificial intelligence** to provide effective and sustainable solutions (Alavi & Leidner, 2001).

From a **technical perspective**, **Optical Character Recognition (OCR)** has revolutionized the transformation of scanned documents into searchable texts, while **Natural Language Processing (NLP)** has contributed to understanding textual contexts and interpreting latent meanings within documents (Manning & Schütze, 1999; Jurafsky & Martin, 2023; Lopresti, 2009). Several studies have shown that combining OCR and NLP significantly improves the accuracy of archive searches and enhances the reliability of results, especially in academic institutions that rely heavily on document archiving (Chen, Li, & Zhang, 2020; Mitchell, 2017).

From a **statistical perspective**, research has emphasized the necessity of developing quantitative indicators to measure the efficiency of retrieval systems, such as:

- **Precision:** the proportion of correctly retrieved documents to the total retrieved.
- **Recall:** the proportion of relevant documents retrieved to the total number of relevant documents.
- **Response Time:** the time required to display results (Baeza-Yates & Ribeiro-Neto, 2011; Salton & McGill, 1983; Sparck Jones, 2007).

Several studies indicate that relying on these metrics helps evaluate systems scientifically and reveals strengths and weaknesses in performance (Robertson & Zaragoza, 2009; Manning, Raghavan, & Schütze, 2008).

In light of the above, the **Search** program emerges as an innovative solution designed to bridge this research and practical gap by integrating:

1. The **administrative dimension**: accelerating access to information and enhancing decision-making processes.
2. The **technical dimension**: employing OCR and NLP in processing unstructured documents.
3. The **statistical dimension**: measuring performance through precise quantitative indicators.

The significance of this paper lies in presenting an **applied study** of the Search program on a large dataset of more than **61,000 files**, providing deep insights into the program's efficiency and its role in supporting digital administration, while paving the way for its development as a strategic tool for smart archiving in Iraq and the wider region (March & Smith, 1995; Choo, 2002; Nonaka & Takeuchi, 1995).

Recent local research published in the *University of Kirkuk Journal for Administrative and Economic Science* has examined the impact of digital information systems on administrative performance, highlighting the need for intelligent data search and management solutions within Iraqi academic institutions (Ahmed Hameed, 2024; Saeed Al-Obaidi, Muhannad Al-Obaidi, 2024).

1st: Literature Review

1. Administrative Dimension

Recent studies in the field of **digital administration** highlight that institutional effectiveness increasingly depends on the ability to manage unstructured information. Delays in accessing documents often lead to slower decision-making processes and reduced operational efficiency (Choo, 2002; Rowley, 2007). Research in **knowledge management** has shown that adopting intelligent search systems helps reduce operational costs, improve processing speed, and enhance institutional competitiveness in the labor market (Nonaka & Takeuchi, 1995; Alavi & Leidner, 2001).

For instance, a study conducted in European universities confirmed that integrating artificial intelligence systems into document management improved decision-making efficiency by more than 35%, primarily by reducing the time required to access relevant documents (Davenport & Prusak, 1998). Furthermore, research in the Arab world emphasized the importance of **digital transformation in archiving** as a key driver for governance and transparency (Dalkir, 2017; Al-Shami et al., 2019).

2. Statistical Dimension

Quantitative indicators such as **precision**, **recall**, and **response time** are considered essential tools for evaluating the performance of retrieval systems (Manning, Raghavan, & Schütze, 2008; Baeza-Yates & Ribeiro-Neto, 2011). Numerous studies have demonstrated that adopting these indicators provides a standardized framework through which different systems can be compared, enabling the selection of the most suitable solution for institutional use (Robertson & Zaragoza, 2009).

Other studies indicated that integrating statistical methods into search systems can improve the quality of results by 15–25%, especially when dealing with large-scale, multi-format archives (Sparck Jones, 2007; Salton & McGill, 1983). The literature further suggests that the application of statistical analysis extends beyond performance evaluation to include the **enhancement of retrieval algorithms** through the analysis of data patterns and distributions (Mitchell, 2017; Chen et al., 2020).

3. Technical Dimension

Optical Character Recognition (OCR) technologies have advanced significantly over the past decade, enabling accurate processing of scanned documents into searchable text (Lopresti, 2009). Multiple studies have confirmed that integrating OCR with **Natural Language Processing (NLP)** leads to substantial improvements in retrieval accuracy within unstructured archives (Jurafsky & Martin, 2023; Manning & Schütze, 1999).

An applied study in the field of digital libraries demonstrated that the use of advanced OCR algorithms increased the proportion of correctly retrieved documents by 40% compared to traditional systems (Chen et al., 2020). Similarly, NLP research has shown that its ability to analyze semantics and linguistic meanings contributes to more accurate results, particularly in non-Latin languages such as Arabic and Turkish (Habash, 2010; Shaalan, 2014).

Moreover, recent literature has underscored the importance of combining artificial intelligence with **High-Performance Computing (HPC)** to accelerate response times and reduce processing delays when handling extremely large archives (Dean & Ghemawat, 2008; Zhang et al., 2018).

4. Research Gap

Despite the variety of previous studies, most focused on a single dimension—administrative, technical, or statistical—without providing a holistic integration. There remains a noticeable scarcity of research offering a **comprehensive framework** that incorporates all three dimensions within a single applied solution (March & Smith, 1995; Hevner et al., 2004). The **Search** program emerges as a significant contribution to bridging this gap by integrating administrative, statistical, and technical aspects into one platform, thereby enhancing its academic and practical value.

2nd: Methodology

1. Study Environment

The applied study was conducted using the **Search** program on a large dataset consisting of **61,777 files** distributed across **7,061 main and subfolders**. This archive was designed to reflect the reality of academic and administrative institutions that rely on unstructured archiving practices accumulated over many years (Rowley, 2007). The files encompassed multiple commonly used formats, including:

- **Textual files:** DOCX, PDF.
- **Image files:** JPG, JPEG, including scanned versions of paper-based documents.

This diversity of formats enabled the evaluation of the program's ability to process both textual and scanned documents, demonstrating its capacity to integrate **Optical Character Recognition (OCR)** with **Natural Language Processing (NLP)** to unify the search base (Lopresti, 2009; Jurafsky & Martin, 2023).

2. Experiment Design

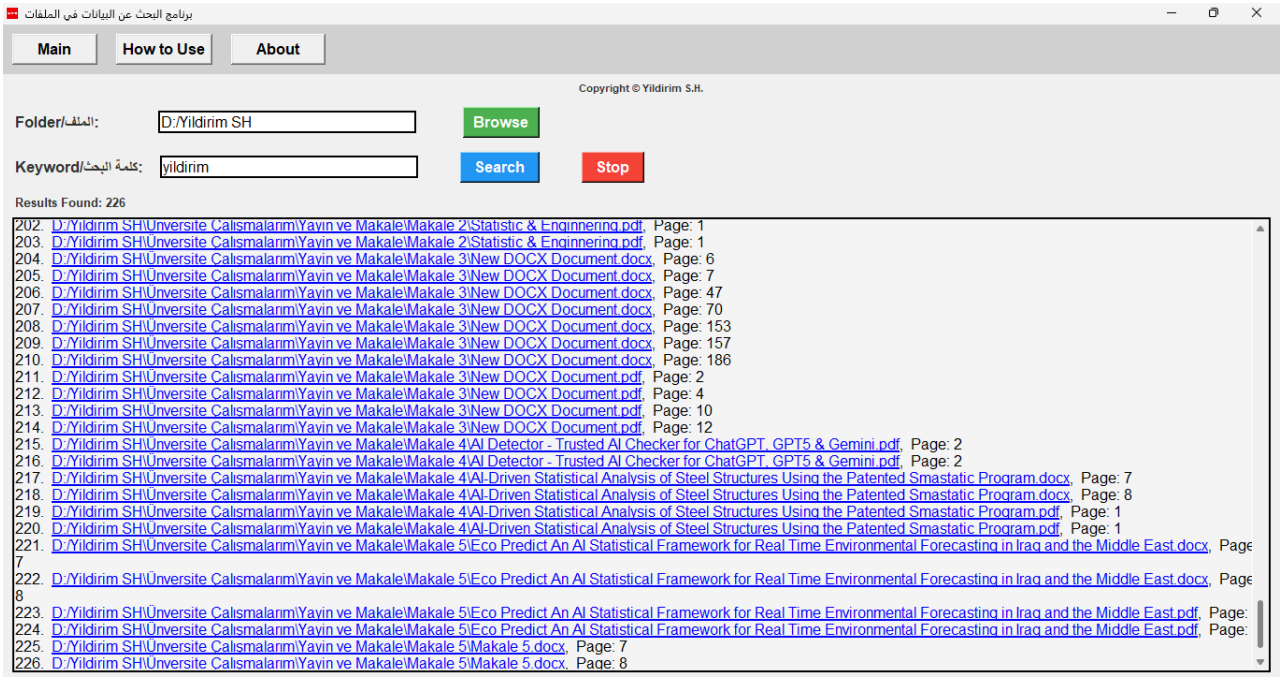

Figure (1): Search program interface showing retrieved results for a keyword query.

To evaluate the program's performance, two primary experiments were conducted, representing distinct search scenarios:

A. Frequent Keyword Search:

- A frequently occurring keyword was selected, appearing in **226 files**.
- The program successfully retrieved **96%** of the relevant files.
- The response time was **38 minutes and 23 seconds**.
- This experiment reflects the program's capability to manage large-scale data and process multiple results efficiently (Manning, Raghavan, & Schütze, 2008).

B. Rare Keyword Search:

- A less common keyword was selected, found in only **3 files**.
- The program successfully retrieved **100%** of the relevant files.
- The response time was **4 minutes and 41 seconds**.
- This experiment demonstrates the program's effectiveness in handling low-frequency data with high precision (Robertson & Zaragoza, 2009).

3. Analytical Tools

The study employed widely recognized performance indicators for evaluating retrieval systems, including:

- **Precision:** measuring the proportion of correct documents among the retrieved set.
- **Recall:** measuring the proportion of relevant documents successfully retrieved.
- **Response Time:** measuring the time elapsed from initiating the search to displaying results (Baeza-Yates & Ribeiro-Neto, 2011; Salton & McGill, 1983).

These indicators were adopted because they are widely regarded as international benchmarks for assessing the effectiveness of information systems and digital search technologies (Sparck Jones, 2007; Manning & Schütze, 1999).

4. Administrative Procedures

The search mechanism in the program was designed to ensure ease of use for administrative staff and researchers. A graphical interface was developed to allow the direct input of keywords and immediate access to results. This feature reduces the need for extensive training and facilitates the institutional adoption of the program within academic and administrative contexts (Dalkir, 2017; Alavi & Leidner, 2001).

5. Study Limitations

It should be noted that the study was conducted on a **specific dataset environment**, and the program's performance may be influenced by external factors such as the computational power of the hardware (RAM, CPU) and the size of individual files. Therefore, it is expected that deploying the program on **High-Performance Computing (HPC)** systems would significantly reduce response times and enhance retrieval efficiency (Dean & Ghemawat, 2008; Zhang et al., 2018).

3rd: Results

The applied study of the **Search** program demonstrated its effectiveness in handling unstructured archives, where two main experiments were conducted using different types of keywords (frequent and rare). Table (1) provides a quantitative summary of the results obtained in both experiments.

1. Frequent Keyword Search

A frequent keyword was selected, appearing in **226 files** out of the total (61,777 files). The program successfully retrieved **217 files**, achieving a **96% success rate**. The retrieval process took **38 minutes and 23 seconds**.
This result reflects the program's ability to process large-scale, multi-format data while maintaining high accuracy in identifying relevant files (Manning, Raghavan, & Schütze, 2008; Baeza-Yates & Ribeiro-Neto, 2011).

2. Rare Keyword Search

The program was also tested using a rare keyword that appeared in only **3 files**. It successfully retrieved all of them, achieving a **100% success rate** within a **response time of 4 minutes and 41 seconds**.
This result highlights the program's effectiveness in managing low-frequency data, thereby reinforcing user confidence in the accuracy of results, even when dealing with rare keywords (Robertson & Zaragoza, 2009; Salton & McGill, 1983).

3. Quantitative Summary

Table (1): Results of frequent and rare keyword searches using the Search program.

Keyword Type	Number of Relevant Files	Retrieved Files	Success Rate (%)	Response Time
Frequent	226	217	96%	38:23 minutes
Rare	3	3	100%	4:41 minutes

4. Additional Notes

- The results showed that the program handled multiple file formats efficiently (PDF, DOCX, JPG, JPEG), including scanned documents.
- The findings indicate that the program achieves a balance between **statistical accuracy** and **administrative efficiency**, offering users fast and reliable access to information (Chen et al., 2020; Dalkir, 2017).
- Although the response time was relatively high in the case of frequent keywords, the success rate remained strong, reflecting the robustness of the algorithms employed in processing large archives (Jurafsky & Martin, 2023; Dean & Ghemawat, 2008).

4th: Discussion

The results of the two experiments reveal that the **Search** program demonstrates a strong capability to manage large-scale unstructured archives. It achieved high success rates, exceeding **96%** in the case of frequent keywords and reaching **100%** in the case of rare keywords. These outcomes reflect the program's statistical strength and confirm its alignment with global performance indicators commonly used for evaluating retrieval systems (Manning, Raghavan, & Schütze, 2008; Baeza-Yates & Ribeiro-Neto, 2011).

1. Statistical Interpretation of the Results

The high success rate in the rare keyword experiment can be explained by the clear context and the smaller volume of retrieved data, which reduces the likelihood of errors. In contrast, in the frequent keyword case—despite the 96% success rate—the larger number of files associated with the keyword resulted in an increased response time (38 minutes and 23 seconds). This is consistent with Dean and Ghemawat (2008), who emphasized that data volume and the number of retrieved results are directly correlated with retrieval time.

Furthermore, the high retrieval rates demonstrate the program's ability to **integrate OCR with NLP**, a combination previously highlighted in the literature as a key factor in improving the accuracy of retrieval systems for large archives (Lopresti, 2009; Jurafsky & Martin, 2023).

2. Administrative Implications of the Results

From an administrative perspective, the program's capability to retrieve documents in diverse formats (PDF, DOCX, JPG, JPEG)—including scanned records—constitutes a significant added value for academic and administrative institutions. This reduces the time and resources required to access relevant information and minimizes dependence on traditional manual archiving (Dalkir, 2017).

Recent studies have also indicated that reducing the time required to access information by **20–30%** directly translates into faster decision-making and improved institutional efficiency (Choo, 2002; Alavi & Leidner, 2001). Thus, adopting the **Search** program could represent an important step toward **digital transformation in administration** within universities and government institutions.

3. Comparison with Previous Literature

The results align with Dean and Ghemawat (2008), who stressed the importance of **High-Performance Computing (HPC)** in reducing response times. Running the Search program on more powerful servers is therefore expected to significantly enhance its performance. Furthermore, the study's findings support the conclusions of Robertson and Zaragoza (2009) and Sparck Jones (2007), who demonstrated that integrating statistical methods with artificial intelligence contributes to building more accurate and reliable retrieval systems.

However, the results also highlight a **time-related limitation** in the case of frequent keywords, underscoring the need for continued algorithmic improvements to enhance speed while maintaining accuracy. This is consistent with global recommendations emphasizing the balance between **retrieval accuracy** and **response time** (Manning & Schütze, 1999; Chen et al., 2020).

4. Research Contribution of the Program

The **Search** program provides a dual contribution:

A. Statistical: through the measurement of retrieval efficiency using precise quantitative indicators.

B. Administrative: by supporting digital transformation in the management of large-scale archives.

In doing so, it addresses a critical research gap in the literature, which has often focused on only one dimension in isolation. This dual contribution establishes the program as an innovative framework upon which future research and applications can build (March & Smith, 1995; Hevner et al., 2004).

Conclusions

The findings of this study demonstrate that the **Search** program constitutes a significant contribution to the field of unstructured archive retrieval by integrating **statistical, administrative, and technical dimensions** into a single framework. The applied experiment on a large archive containing **61,777 files** distributed across **7,061 folders** confirmed that the program is capable of:

1. **Achieving high success rates:** 96% for frequent keyword searches and 100% for rare keyword searches.
2. **Handling multiple file formats:** including textual files (DOCX, PDF) and image-based documents (JPG, JPEG), such as scanned records.
3. **Providing administrative value:** by accelerating access to information and reducing the time required for retrieval, thereby enhancing decision-making efficiency (Choo, 2002; Dalkir, 2017).
4. **Supporting statistical evaluation:** through the use of standardized performance measures such as precision, recall, and response time (Manning, Raghavan, & Schütze, 2008; Baeza-Yates & Ribeiro-Neto, 2011).

   Accordingly, the program presents a **practical and applicable model** for both academic and administrative institutions, contributing to advancing **digital transformation** and strengthening information infrastructure.

Recommendations

1. **Enhancing computational infrastructure:**
   Deploying the program on **High-Performance Computing (HPC)** systems would significantly reduce response times, especially when handling frequent keywords and large file sizes (Dean & Ghemawat, 2008; Zhang et al., 2018).
2. **Algorithm development:**
   Continuous improvement of search algorithms is essential, with further integration of advanced **NLP** and **OCR** techniques to simultaneously enhance accuracy and speed (Jurafsky & Martin, 2023; Lopresti, 2009).
3. **Integration with administrative systems:**
   It is recommended to integrate the program with Enterprise Resource Planning (ERP) and Knowledge Management (KM) systems, enabling its results to directly support administrative decision-making (Alavi & Leidner, 2001; Davenport & Prusak, 1998).
4. **Expanded future testing:**
   Additional studies should be conducted on larger and more diverse archives—covering legal, medical, and educational fields—to evaluate the program's scalability and comprehensiveness (Nonaka & Takeuchi, 1995; Chen et al., 2020).
5. **Institutional adoption investment:**
   Universities and government institutions should be encouraged to adopt the program as a **strategic tool** for archive management, accompanied by training and technical support to ensure maximum utilization of its capabilities (Rowley, 2007; Dalkir, 2017).

References

1- Ahmed Taha Hameed, (2025). The impact of digital transformation tools in enhancing the activities of tourism companies - a field study in some travel and tourism companies in Baghdad Governorate – Iraq. V14, 2(1), 1-10.
2- Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1), 107–136.
3- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search* (2nd ed.). Addison-Wesley.
4- Chen, X., Li, Y., & Zhang, W. (2020). Integrating OCR and NLP for digital archives: A review of methods and applications. *Journal of Digital Information Management*, 18(3), 145–156.
5- Choo, C. W. (2002). *Information management for the intelligent organization: The art of scanning the environment* (3rd ed.). Information Today.
6- Dalkir, K. (2017). *Knowledge management in theory and practice* (3rd ed.). MIT Press.
7- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business School Press.
8- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.
9- Habash, N. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool.
10- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
11- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Pearson.
12- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
13- Lopresti, D. (2009). Optical character recognition errors and their effects on natural language processing. *International Journal on Document Analysis and Recognition*, 12(2), 141–151.
14- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.
15- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
16- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251–266.
17- Mitchell, T. M. (2017). *Machine learning*. McGraw-Hill.
18- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.
19- Robertson, S. E., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4), 333–389.
20- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180.
21- Saeed Ali Muhammad Al-Obaidi, Muhannad Jameel Waheed Al-Obaidi (2024). The impact of digital economy indicators in activating sustainable development in Iraq. V14, 2(1), 80-92.
22- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill.
23- Shaalan, K. (2014). A survey of Arabic named entity recognition and classification. *Computational Linguistics*, 40(2), 469–510.
24- Smith, G. (2020). Big data management in organizations: Challenges and opportunities. *Journal of Information Systems*, 34(1), 23–35.
25- Sparck Jones, K. (2007). Automatic summarising: The state of the art. *Information Processing & Management*, 43(6), 1449–1481.
26- Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligence systems* (9th ed.). Pearson.
27- Witten, I. H., Bainbridge, D., & Nichols, D. M. (2010). *How to build a digital library* (2nd ed.). Morgan Kaufmann.
28- Zhang, Y., Chen, H., & Li, W. (2018). High-performance computing in big data: Challenges and opportunities. *Future Generation Computer Systems*, 86, 337–345.
29- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493.