Search Title

Intelligent processing of documents using artificial intelligence

An analytical Study

Preparation

Dr. Ayman Abdel Rahman Ismaeil

Assistant Professor of Information Sciences College of Business Administration / Al Sharqiyah University

Sultanate of Oman

Prof. Dr. Ashraf Abdel Mohsen Al-Sharif

Professor of documents and electronic archives

Faculty of Arts / Beni Suef University / Egypt

2024

Abstract

The study addresses the role of artificial intelligence in the intelligent processing of documents by using natural language processing (NLP) to review and analyze documents. The more documents the artificial intelligence scans and analyses, the more end users interact with the documents, and the more intelligent it becomes in identifying and processing information.

The problem of the study is to demonstrate the role and ability of artificial intelligence in intelligent document processing (IDP) and the process of automating data extraction from unstructured and semi-structured documents and converting them into structured and usable data in companies. The study used the descriptive analytical method, which is based on collecting information about the subject under study and analyzing it to arrive at facts and results that can be generalized.

The study reached several results, the most important of which is that large institutions and emerging companies resort to Intelligent Document Processing (IDP) technology as one of the important solutions in dealing with unstructured data and storing it in the form of structured data that enables the institution to benefit from it in decision-making.

key words

Artificial intelligence - intelligent document processing - document understanding technology - data mining

Methodology

Artificial intelligence aims to understand the complex mental processes that the human mind performs while practicing thinking, and then translate these mental processes into equivalent accounting processes that increase the computer's ability to solve complex problems.

The concept of artificial intelligence is related to the intelligence associated with digital or electronic devices such as; Computers, cellular devices, or robots, and artificial intelligence expresses the ability of these digital devices to perform tasks associated with intelligent beings. The term artificial intelligence applies to systems that have human intellectual processes, such as; The ability to think, discover meaning, and learn from past experiences. Examples of operations performed by digital devices that are due to the presence of artificial intelligence include: Discover proofs of mathematical theorems, medical diagnosis, web search engines, voice or handwriting recognition.

AI is not only redefining how organizations use digital documents today, it is reinventing them. By using natural language processing (NLP) to review and analyze documents, AI can unleash the full power of enterprise data to enable better engagement for both employees and customers. The more documents AI scans and analyses, the better it can discover how end users interact with documents. , the smarter it becomes in identifying and processing information. As organizations collect data from diverse sources, the amount of unstructured data in their databases grows exponentially, which ultimately hurts efficiency and costs money.

AI can mitigate these problems by identifying similarities between data through classification and extraction based on pre-defined criteria. This allows a deeper look to classify different document types and extract relevant information in an organized way. AI also enables analysis of the work of the most productive employees and specific patterns in each employee's performance to extract actionable insights. In terms of a customer-oriented approach, some organizations are already using machine learning (ML) to sift through emails, reviews, and other forms of customer interaction to understand the sentiment behind it. These models created by AI are then connected to departments such as sales, marketing, and support. Customers, where data is collected together to predict customer behavior and increase their purchasing experience.

Study Problem:

The twenty-first century brought about a technological revolution in various business sectors, including the document management sector, whose operations were affected by this revolution and its applications; Rather, the profession of document specialists was forced to change its job specifications and requirements. To comply with modern documentary technological needs. Despite ongoing digital transformation, many organizations today still spend a significant amount of time manually processing information from countless documents.

Due to the nature of digital files such as PDF files, images, spreadsheets and even multimedia such as video, images, etc. Hence the problem of the study, which works to clarify the role and ability of artificial intelligence in intelligent document processing (IDP) and the process of automating data extraction from unstructured and semi-structured documents and converting them into organized and usable data in companies.

Study questions:

Based on the above, the study raises the following questions:

What is artificial intelligence and its characteristics?

2- What is the Intelligent Document Processing (IDP) technology?

3- What are the benefits and characteristics of intelligent document processing?

4- What are the areas that benefit from the application of artificial intelligence in the intelligent processing of documents?

5- How does smart document processing technology extract data from documents?

6- What is the mechanism used to understand documents?

7- What are the digital platforms used to understand documents?

Objective of the study: The study aims to:

1- Defining the concept of artificial intelligence and its importance.

2- Explain what intelligent document processing.

3- Explaining the benefits and characteristics of intelligent document processing.

4- Explaining the areas that benefit from the application of intelligent document processing in understanding documents.

5- Explaining the mechanism used in the intelligent processing of documents.

6- Explaining the method of automatically extracting data from documents using artificial intelligence.

7- A statement of the digital platforms used to understand documents.

Importance of study

Due to the large number of documents that result from digital transformation or that are generated digitally on computers, and due to the many types of textual data, statistical tables, maps, videos, and images; Therefore, the importance of the research is clear in highlighting the importance of using smart document processing technology to understand documents by using the technology of automated data extraction from documents, classifying them, and storing them in independent databases that can be used to make decisions in various organizations and companies.

Study Methodology

The research relied on the descriptive analytical method, which is based on collecting facts about the subject of the study and analyzing these facts to reach results that can be generalized.

Search terms:

- **Artificial Intelligence (AI)** It is one of the modern computer sciences that searches for advanced methods for carrying out actions and conclusions that are similar, albeit within narrow limits, to those methods attributed to human intelligence

- **Intelligent Document Processing (IDP):** It is a technology-based approach that automates document processing and extracts valuable information. Compared to document data mining, IDP includes a broader set of capabilities because it combines optical character recognition (OCR), artificial intelligence (AI), natural language processing (NLP), and machine learning (ML) algorithms to process documents. It also enables devices to

understand the data content, context and meaning of various documents, regardless of their format or structure.

- **Document Understand:** It is an artificial intelligence service that allows developers to extract texts, tables, and other key data from document files through application programming interfaces. Through it, business processing tasks can be automated using pre-prepared artificial intelligence models and document extraction can be customized to suit the needs of the organization or company.

- **Automated data extraction:** Document data extraction refers to the process of extracting relevant information from different types of documents, whether digital or printed and includes identifying and retrieving new data points such as tables, graphics, numbers, names, addresses, etc.

Previous studies:

1- A study on Sardouk (2020) entitled "The Use of Smart Robots in University Libraries: Global Experiences and the Current Reality in the Maghreb Countries," The Journal of Information Studies & Technology, Hamad Bin Khalifa Publishing House, Qatar.

This study aimed to highlight the importance of artificial intelligence technology in general and intelligent robots specifically in library service, by presenting the most important global experiences in exploiting this technology within the library. To focus on the Maghreb reality, a survey was conducted on a sample of Maghreb university libraries to determine the extent of their use of artificial intelligence technology and the prospects for that in the future. The study concluded that some European, American, and Asian libraries have taken a distinct step towards exploiting artificial intelligence in some library activities, while Maghreb libraries do not attach importance to this technology as a strategic tool for developing knowledge management methods.

2- Noura Abdullah Al-Azzam (2020) entitled The Role of Artificial Intelligence in Raising the Efficiency of Administrative Systems for Human Resources Management at the University of Tabuk, Journal of the College of Education, Sohag University, April Issue - Part 1, 2021.

The study aimed to identify the role of artificial intelligence in raising the efficiency of administrative systems for human resources management at the University of Tabuk. The study concluded that it is necessary to conduct more studies on artificial intelligence and its relationship to the efficiency of administrative systems for human resources management so that they include larger samples of universities across the Kingdom of Saudi Arabia.

3- Graham A. Cutting (2021) study entitled Intelligent Document Processing - Methods and Tools in the real world, available at: https://arxiv.org/ftp/arxiv/papers/2112/2112.14070.pdf

This paper deals with the intelligent document processing of classifying and extracting information and later integrating it into business processes about forms and invoices. This study conducted a brief review of the market in commercial tools for optical character recognition, document classification, and data extraction. The study provided brief definitions of the main terms that They are encountered in computer science publications and trade bulletins in order to demystify the language for the layman. The researcher conducted a practical application on a typical large industrial company where the requirement is to extract 100% of the information with 100% reliability to integrate it into the ERP system. Finally, a brief description of the latest research conducted by large companies that push the boundaries of deep learning techniques is presented. Further and further through massive computing and financial power.

4- **Daniel J. Chenok (2018) study entitled Delivering Artificial Intelligence in Government: Challenges and Opportunities, IBM Center for The Business of Government, available at:**

https://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf

This paper reviews recent progress in applying artificial intelligence to public sector service delivery, drawing on lessons learned from commercial experience as well as burgeoning cognitive computing activity by federal, state, local, and international governments. The study develops a framework for agencies to plan, develop, and deploy artificial intelligence systems. It then presents a set of challenges for government leaders and innovators in this field along with opportunities for US government agencies to address these challenges. Finally, the study develops a model for US government agencies to use in applying artificial intelligence to improve task performance.

5- **Martin Geletka et al.'s (2022) study entitled Information Extraction from Business Documents (Case Study), at:**

https://nlp.fi.muni.cz/raslan/2022/paper18.pdf

This paper reviews the topic of the use of artificial intelligence in the field of documents, which is a relatively new research topic that refers to techniques for automatically reading, understanding and analyzing business documents. Nowadays, many companies extract data from business documents through manual efforts that are time-consuming, expensive, and require manual customization or configuration. This paper describes techniques to address these issues, apply them to real-world data, and implement them as a comprehensive solution for automatically extracting information from business documents.

6- **Ephesoft study (2024) entitled Elements of Intellegent Document Processing-IDP Essential, available at:**

https://www.tungstenautomation.com/-/media/files/e-books/en/eb_ephesoft-the-6-essential-elements-of-intelligent-document-processing.pdf

This paper discusses how to process documents such as paper documents, e-mail, or electronic documents such as PDF, Microsoft Word, and image files. In fact, more than 80% of information is trapped in unstructured and semi-structured content. This means that only about 20% of data is organized and can easily be searched and retrieved from relational databases. The paper also discussed the method of mining unstructured and semi-structured data and the way it has evolved from manual data

entry to the correct format for Optical Character Recognition (OCR) to now being able to use artificial intelligence, which is referred to as Intelligent Document Processing (IDP). The industry has advanced innovative tools and functions that allow this.

7- **Docbyte study (2024). Entitled Capturing Data Intelligently, available at:**

https://www.docbyte.com/wp-content/uploads/2019/11/Docbyte-whitepaper-Capturing-data-intelligently.pdf

This paper discussed the reasons why companies still use paper in doing business, and showed how the method of capturing or extracting data from documents in an automated manner uses artificial intelligence technology, and that this saves companies a lot of time and money. It also explained the steps by which this is done. The process of capturing data from documents, whether paper, using Optical Character Recognition (OCR) technology or using Natural Language Processing (NLP), and its role in extracting data from documents.

The concept of artificial intelligence (AI) refers to the way in which the capabilities of human intelligence are simulated, and it is a part of computer science that deals with the process of designing intelligent systems, which exhibit a set of characteristics that are linked to deep intelligence in many human behaviors. It is based on collecting knowledge and information that is in-depth with various sectors in the world and working to process it and disseminate it to benefit from it in the form of practical intelligence. [1]

To know what is the intelligence of industrial intelligence, the intelligence of human intelligence is required, as it is linked to mental solidarity, such as the ability to adhere to the conditions of life, and to benefit from the experiences, the previous experiences, thinking, analysis, The solutions, the problems and the sound

[1] **Kevin C. Desouza (2018) . Delivering Artificial Intelligence in Government: Challenges and Opportunities , IBM Center for The Business of Government , At:** https://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf#

production and the sensitivity in the other, in addition to the speed of learning and the use of what has been learned in the healthy and useful form, Artificial intelligence is almost everywhere around us, starting with self-driving cars and drones, translation or investment software, and many other widespread applications In life.

Artificial intelligence:

It is a branch of computer science concerned with simulating intelligent behavior in humans, and in it, we need:

• **Data system**: used to represent information and knowledge.

• **Algorithms**: We need them to plan how this information is used.

• **Programming language**: used to represent both information and algorithms.[1]

Types of Artificial Intelligence:

Artificial intelligence falls into two main categories, which are as follows:

1- Narrow AI:

Narrow AI is also known as weak artificial intelligence, and it is a type of intelligence that mimics human intelligence but is specialized in one and limited type of intelligence. Narrow AI focuses on performing one type of task, but very well., so it focuses on carrying out one task professionally, but it operates

under far more restrictions than human intelligence. Examples of narrow artificial intelligence include the Google search engine, image recognition programs, and self-driving cars.[2]

---

[1] **João Reis (2019) . Artificial Intelligence in Government Services: A Systematic Literature Review , Springer Nature Switzerland , At:** https://comum.rcaap.pt/bitstream/10400.26/30168/1/10.1007%40978-3-030-16181-123.pdf

[2] **Tanja Sophie Gesk (2022) . Artificial intelligence in public services: When and why citizens accept its usage , At:** https://www.researchgate.net/profile/Tanja-

2- Artificial General Intelligence

General artificial intelligence is also known as strong artificial intelligence, and it is a type of intelligence found in smart machines and devices. General artificial intelligence is characterized as a type of intelligence found in a machine that gives it general intelligence like a human so that this intelligence is used to solve any problem. Examples of devices that have general artificial intelligence include: Robots that are used to accomplish multiple tasks and that make decisions based on the situation.[1]

Government institutions adopt a digital transformation strategy and increase their investments made in information technology infrastructure, the Internet of Things, machine learning and artificial intelligence. John Reese points out that the Italian government, which is the second largest investor in government digitization in the world, is increasing online access points and services to increase the spread of systems that do not require direct interaction with the government.[2]

Artificial intelligence must be viewed as an extension of digital transformation, and although scientific research focuses on business economics, computer science, or operations research, it must also focus on social issues and how public administration can effectively contribute to improving the lives of citizens through... Use of artificial intelligence. [3]

(ML) for document processing. It also enables devices to understand the data

content, context, and meaning of various documents regardless of their format or structure.

Gesk/publication/359738929_Artificial_intelligence_in_public_services_When_and_why_citizens_accept_its_usage/links/62501f61cf60536e234cc832/Artificial-intelligence-in-public-services-When-and-why-citizens-accept-its-usage.pdf?origin=publication_detail

[1] Ibid, pp 10-14

[2]) Hila ,Mehr (2017) . Artificial Intelligence for Citizen Services and Government , Harvard AshCentreTechnology At:
https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

[3] Slava Jankin Mikhaylov (2020) . AI for the Public Sector: Opportunities and challenges , 1Institute for Analytics and Data Science , University of Essex , At:
https://pubmed.ncbi.nlm.nih.gov/30082303/

Document processing can also be defined as a systematic method for converting raw, unstructured data from various sources, such as invoices, contracts, and purchase orders, into structured, actionable information. It involves using software tools to manage document-based workflows and improve the efficiency, accuracy, and security of document-related tasks. Here are three important components of document processing:[1]

• **Data mining**: This involves carefully identifying and extracting basic information from a wide range of documents.

• **Data transformation**: The extracted data is converted into a structured format, making it machine-readable or more useful for analyzing reports or other applications.

• **Data Validation**: Ensures that the transformed data being processed or analyzed is accurate, complete, consistent, and meets specified standards or requirements. **Intelligent Document Processing (IDP)** automates the process of manual data entry from paper documents or document images to integrate with other digital business processes. For example, imagine a workflow that automatically issues orders to suppliers when inventory levels are low.

Although the process is automated, no order is shipped until the supplier receives payment. The supplier sends an invoice via email, and the accounts team manually enters data before payment is completed, resulting in manual checkpoints that cause bottlenecks or errors. Instead, intelligent document processing systems automatically extract the invoice data and enter it into the required format into the accounting system. You can use document processing to automate document management using machine learning (ML) and various AI techniques.[2]

Intelligent Document Processing (IDP) solutions transform unstructured and semi-structured information into usable data.

---

[1] Esposito , Floriana (2014) . Intelligent document processing , at:
file:///C:/Users/ADMIN/Downloads/Intelligent_document_processing%20(3).pdf

[2] **Graham Cutting and Af Cutting–Decelle (2021) . Intelligent Document Processing – Methods and Tools in the real world , at: https://arxiv.org/ftp/arxiv/papers/2112/2112.14070.pdf**

During the digital transformation of an organization's business, more than 80% of business data is available in unstructured formats, such as business documents, emails, images, and PDF documents. Intelligent document processing represents the next generation of automation, through which data from a variety of document formats can be captured, extracted, and processed. It uses artificial intelligence techniques such as natural language processing (NLP), computer vision, deep learning, and machine learning (ML) to classify, organize, and extract relevant information, as well as verify the extracted data.

The importance of intelligent document processing is evident if we know that by 2025, data around the world will exceed 175 zettabytes. With most of this information found in emails, texts, PDFs, and scanned documents, imagine for a moment how much data is in the email alone; This constitutes an obstacle to automation and digital transformation.[1]

The importance of using intelligent document processing systems is due to:

1-While dealing with structured data directly, processing and analyzing unstructured data is tedious and difficult.

2- Intelligent Document Processing provides users with the ability to process many types of documents, including PDF files, spreadsheets, Word documents, and more.

3- IDP Intelligent Document Processing platforms offer a powerful and accessible solution to extract data from these documents by eliminating the need for any manual intervention.

4- The extracted data, when combined, enables beneficiaries to make reliable decisions and improve business efficiency.

5- Intelligent document processing is an integral part of the digital transformation journey within the organization. By adopting IDP intelligent document processing technology, you move from manual and paper processes to workflow processes. Automated work. This transformation enables the organization to benefit from the advantages of emerging technologies in document management.[2]

---

[1] **Astera Company (2023) . Leverage the Data Trapped in Unstructured Sources with Data Extraction , at: https://www.astera.com/wp-content/uploads/2020/07/Leverage-the-Data-Trapped-in-Unstructured-Sources-with-Data-Extraction.pdf**

[2] **Astera Company , op.cit , p.4**

The difference between intelligent document processing and automated document processing

Automated Document Processing is a related technology but very different from Intelligent Document Processing (IDP).

Automated document processing is primarily used to digitize paper documents. Creating digital copies of paper documents allows them to be indexed and searchable within a database.

This is critical for organizations like the DMV that rely on paper documents but need digital data. Intelligent document processing is business process management software capable of digitizing and indexing paper documents, but it It also goes several steps further in separating valuable information and providing insights from data contained in paper documents (or any other type), rather than just computerized access to it.

Benefits of Intelligent Document Processing

Intelligent document processing offers many attractive benefits. The technology's ability to extract and process data from documents provides significant benefits in terms of efficiency, accuracy, and cost savings.

It also enables organizations to streamline their document-driven workflow, enabling employees to focus on more value-added tasks. The best part about intelligent document processing is that it is constantly learning and improving by adapting to new document requirements and variables. Ultimately, IDP makes document management less labor-intensive and more accurate. Here are some benefits of intelligent document profiling:

1- Save time and money:

Implementing intelligent document processing significantly reduces the time spent on manual document processing tasks, allowing your team members to focus on higher-value activities. They no longer need to sift through piles of documents or manually enter data. Instead, they can invest their time in critical tasks, such as dealing with customers, and feel more productive. Moreover, intelligent document processing software can bring significant cost savings to your business. By automating document processing, you also reduce the need for manual labor and reduce expenses associated with hiring and training additional resources. It can

reduce the accuracy and efficiency you obtaining financial risks associated with errors, such as incorrect billing errors.

You can then redirect these cost savings into other strategic initiatives that improve business growth and competitiveness.

2- **Enhance accuracy**: No matter how accurate your team is, human error can and will continue to occur. However, IDP software significantly reduces the risk of human error in document processing. Automated data extraction ensures the accurate capture of information from documents and eliminates inaccuracies that can occur during manual data entry.

3- **Integrating existing systems**: IDP software can integrate with existing document management systems, such as enterprise resource planning (ERP) software, or other business applications. This integration enables you to leverage IDP capabilities across multiple departments and processes within your company. It also ensures smooth data flow and reduces manual data transfer, enhancing overall workflow efficiency.

4- **Achieve compliance and security**: You can leverage the IDP platform to adhere to regulatory requirements and maintain audit trails. These solutions simplify compliance processes by creating a reliable digital trail that proves invaluable during audits and enforcement of data protection regulations.[1]

In a world characterized by relentless data growth, automated document processing that relies on artificial intelligence has emerged as a beacon of efficiency and accuracy, through which the company can:

• **Increase efficiency and reduce manual work**: Break the constraints of time-consuming manual work and increase the efficiency of your organization.

AI-driven systems handle document processing at lightning speed, freeing up valuable human resources for more strategic endeavors.

[1] **Amazon Web Services –AWS . (2024) . Intellegent Document Processing , at:**

https://aws.amazon.com/ar/textract/

• **Improved data accuracy and reduced errors**: reduce errors! AI-powered document processing systems dramatically improve data accuracy, eliminating the risks associated with manual data entry and extraction.

• **Streamline business processes and decision-making**: With data flowing like a well-oiled machine, your company can make informed decisions based on actionable insights. AI-based document processing enables organizations to adapt, innovate, and outperform the competition.

• **Enhanced security and scalability**: Automated document processing systems enhance security by protecting sensitive information from unauthorized access and data leakage. In addition, AI-based solutions easily adapt to increasing document volumes, allowing organizations to scale operations without additional resource constraints or delays.

• **Facilitating digital transformation**: Adopting automated document processing powered by artificial intelligence is a strategic step towards digital transformation.

As organizations evolve, they can harness the power of AI to improve operations, enhance customer experiences, and open new opportunities. Companies can chart a path to success and propel themselves into a future characterized by agility, accuracy, and innovation by using intelligent document processing technology.[1]

Challenges of intelligent document processing

in today's fast-paced business environment, manual document processing has become a significant obstacle to productivity, growth, and competitive advantage. Let's take a closer look at the challenges associated with manual document processing:

• **Inefficiency and imprecision**: The labor-intensive nature of manual processing makes it a time-consuming and error-prone process. Employees often struggle to maintain consistency in their work, resulting in inaccuracies and inconsistencies in the extracted data.

[1] **Amazon Web Services –AWS, op.cit , p5**

• **Resource constraints**: Employees tasked with manual document processing are often bogged down by the sheer volume of documents they have to process, diverting valuable human resources from more strategic and essential tasks.

• **Security risks**: Manual processing can expose sensitive data to security risks, as human intervention may inadvertently lead to data leakage or unauthorized access to confidential information.

• **Expansion limitations**: Relying on manual document processing can limit an organization's ability to scale its operations, as increasing document sizes require additional resources and time for processing.

Uses of intelligent document processing:

Intelligent document processing can be used in the following areas:

• **Healthcare:** Intelligent document processing improves healthcare records management; In the healthcare sector, pristine patient records must be maintained at every point of contact with a hospital or medical institution. Healthcare businesses and businesses use intelligent document processing to extract data from patient records and better organize medical documents. The health insurance sector is also using intelligent document processing to verify claims and reduce manual paperwork in this area.

• **Finance:** The financial sector uses intelligent document processing to automate many aspects of expense management and invoice processing. Businesses can simplify creating expense reports by extracting data from expenses, forms, and business receipts. Finance departments can manage employee and contractor payments quickly and efficiently. For example, intelligent document processing solutions can extract numbers from financial documents and process the data for future payments.

• **Legal:** Companies in the legal sector can use intelligent document processing to analyze contracts. Legal teams use natural language processing (NLP) to analyze the terms and obligations of legal contracts. It can extract data from legal documents and court records to build more powerful legal claims.

• **Logistics:** Logistics companies need to track shipments, transit permits, and other important documents. Companies use intelligent document processing in document processing to reduce the possibility of human error leading to a critical error.

Intelligent document processing helps extract, validate and classify data, so that companies operating in the logistics sector can accelerate logistics functions.

• **Human Resources:** Human Resources employees use intelligent document processing to extract important information from the candidate's CV. The advantage of the intelligent document processing system is that it saves time and ensures that HR teams focus on selecting among the best candidates. The HR sector also uses intelligent document processing when managing payroll, allocating leave, and other HR tasks

• **Education:** IDP can be used to automate administrative processes in educational institutions, such as student registration, word processing, and financial aid applications.

Technologies used in intelligent document processing:

Intelligent document processing uses a set of technologies to process different types of documents:

First: Optical Character Recognition (OCR) In the past, optical character recognition (OCR) was performed primarily using flatbed scanners. Recently, mobile phone photo cameras have become a convenient way to scan physical documents. Optical character recognition (OCR) is a process that converts an image of text into a machine-readable text format. You can use OCR technology to scan paper documents and turn them into images with searchable text data.

OCR technology is vital to document processing because it turns paper forms, receipts, invoices, contracts, legal documents and more into digital documents. The process is subject to many pitfalls and errors, and even under ideal conditions it is rarely 100% accurate.[1]

There are several types of OCR, each with different applications: • Simple OCR programs use matching algorithms to compare text images with patterns of font images and text.

[1] **Graham Cutting and Af Cutting–Decelle (2021) . Intelligent Document Processing – Methods and Tools in the real world , at: https://arxiv.org/ftp/arxiv/papers/2112/2112.14070.pdf**

• Intelligent Character Recognition (ICR) programs use machine learning software to process different image features, such as curves and lines, to process text.

• Intelligent Word Recognition technology uses principles similar to Intelligent Character Recognition but focuses on processing entire words rather than working on individual letters.

• OCR uses a matching algorithm to identify text schemes, logos, and watermarks. • Optical character recognition technology.[1]

Second: Natural Language Processing (NLP)

Machine learning is the study and application of computer algorithms that can automatically improve through experience and the use of data. It is a machine-learning technique that allows computers to analyze, interpret, and understand human language. Natural language processing software processes text and audio data to analyze sentiment, content, or intent. Natural language processing uses a range of techniques - including machine learning, computational linguistics, and deep learning models - to process human language. Here are some of these techniques.[2]

• Computational linguistics involves semantic and syntactic analysis to create frameworks that capture the essence of human language.

• Machine learning technology allows natural language processing models to improve their understanding of metaphors, sentence structure changes, grammar rules, slang expressions, sarcasm, and other elements of human speech.

• Deep learning neural networks allow computers to recognize, classify, and identify complex patterns in sample data, and natural language processing is especially useful when working with unstructured documents and unstructured data, such as live recordings or human speech.

[1] **Ibid , p.3**

[2] **Graham Cutting and Af Cutting–Decelle , Op.cit , p.4**

Third: Robotic Process Automation Robotic process automation (RPA)

It is a form of technology that makes it easier to create and deploy software that automates human actions. You can automate workflow processes with robotic process automation software. For example, a user can record how he or she processes a document. Robotic process automation software then repeats the same steps, eliminating the need for manual document processing work. You can use Robotic Process Automation to automate any process, from data extraction to data capture and more.[1]

Steps to Understanding Documents [2]

Intelligent Document Processing can interpret, classify, and extract data from a variety of document types, from structured data to unstructured text such as emails or reports.

1- **Document classification**: The first step in intelligent document processing is document capture and classification. This includes exporting both paper and digital documents into the system. Document processing tools use artificial intelligence to recognize and classify various types of scanned documents, such as invoices, purchase orders, or legal contracts. This classification is necessary to determine the post-processing steps for each type of document.

2- **Data extraction**: After classification, the system extracts relevant data from the documents. Using optical character recognition (OCR) and natural language processing (NLP), intelligent document processing systems accurately identify specific information, such as dates, amounts, or names. After extraction, the system also performs data validation to ensure accuracy. For example, the system may perform a comparison of extracted data with existing databases or employ the use of pre-defined rules for error checking.

3- **Data processing** : After validation, the extracted data is processed according to its purpose. For example, billing data may be directed to payment processing, and contract details may be sent to a legal platform. An intelligent document processing system integrates with other business systems, such as

---

[1]) **Ibid , p.4**

[2]) **KOFAX Company (2023) . The 6 Essential Elements of Intelligent Document Processing (IDP) , at:** https://www.tungstenautomation.com/-/media/files/e-books/en/eb_ephesoft-the-6-essential-elements-of-intelligent-document-processing.pdf

enterprise resource planning (ERP) and customer relationship management (CRM) for seamless data flow and automation of actions based on processed data.

4- **Continuous learning**:

A key advantage of intelligent document processing systems is their ability to learn and improve over time. Using machine learning algorithms, systems learn from previous errors and adapt to changes in document formats to improve accuracy.

A continuous learning process ensures that the system remains effective even as business needs and document types evolve.

5-Preparing reports and analyses:

Intelligent document processing systems can track metrics, such as processing time, error rates, and throughput rates. They can be further processed through business analytics to extract insights that help identify choke points, improve workflow, and make data-driven decisions for overall efficiency[1].

Document understanding and automated data extraction technology

The modern world is data-driven. There is a huge amount of data running through an organization with every process, regardless of the size of the business. From invoicing and payroll to customer surveys and call logs, every business area has valuable insights that can lead to better results. Taking data from these various, often unstructured, sources, extracting useful information and ultimately analyzing it is essential for making key business decisions. Thanks to modern technology, it has become relatively easy to process existing documents in the form of structured data

[1] **KOFAX Company, op.cit, p.5**

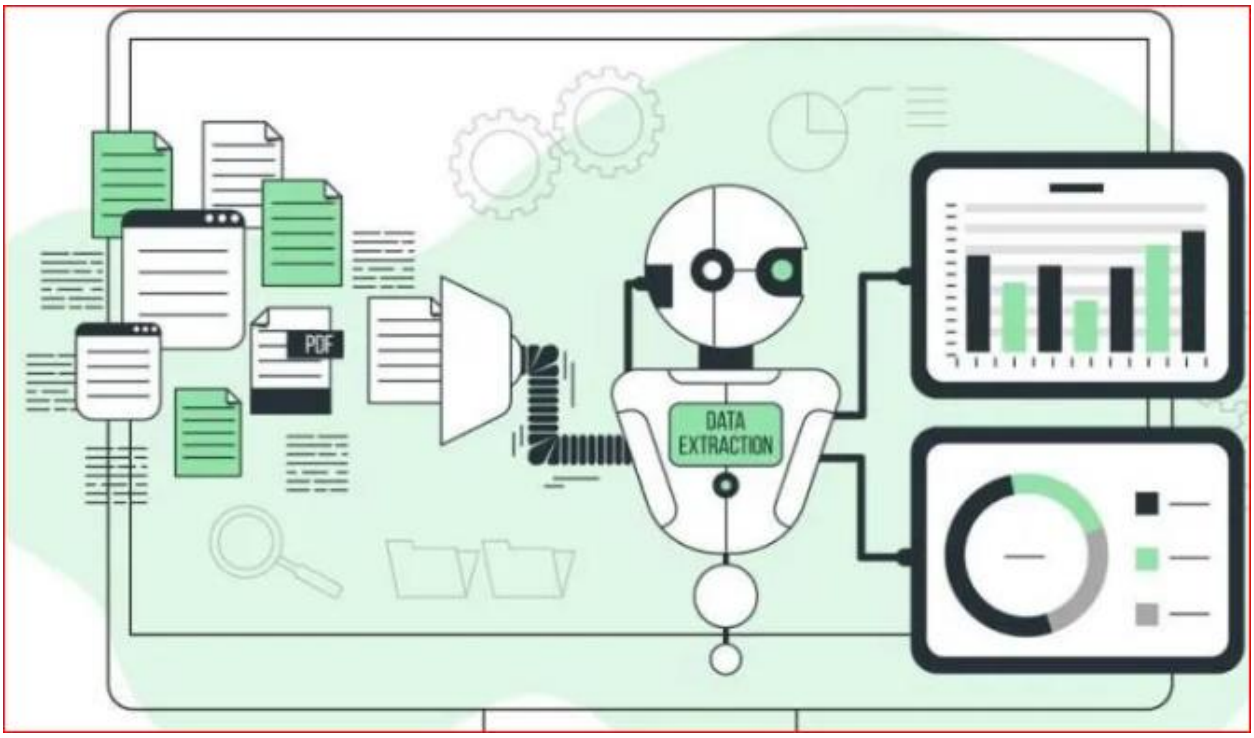

Figure (1) shows the mechanism for automatically extracting data from documents

Challenges of manual data extraction

More than 40% of employees spend nearly a quarter of the week manually performing repetitive tasks that could be automated. Most organizations still rely on manual data extraction to retrieve relevant data from unstructured sources for business use. However, the process is not only resource-intensive but also has many downsides that can affect business growth. Manual data extraction has a high error rate compared to automated tools.

The error rate can be attributed to:

- Inadequate training of people involved in data entry
- Human error
- Misinterpretation of data

• Incomprehensible information.[1]

These issues can cause serious delays in various business processes, especially those that deal with large amounts of data on a daily basis. Another major problem with manual data extraction is the cost and effort involved in verifying the extracted data.

Data quality represents an essential step in the data mining process. It prevents problems arising due to inaccurate data being sent for analysis. Verifying quality using the manual data process requires hiring additional employees who spend hours ensuring the accuracy of the data entered by reviewing each file and checking for errors or discrepancies. This ends up costing the organization time and money, and by using automated data extraction tools, this can be done easily at a lower cost and in less time.[2]

Finally, manual data extraction can divert employees from completing more value-added tasks, which ultimately hurts the organization and prevents it from achieving its strategic goals that require timely delivery of specific data.

Extracting Information from Text

The term Neuro-Linguistic Programming (NLP) refers to methods of interpreting data, i.e. spoken or written by humans. To process human languages using NLP, many tasks such as machine translation, question-answering system, information retrieval, information extraction, and natural language understanding are considered high-level tasks, and the process of information extraction (IE) is considered one of the important tasks in extracting and analyzing data, and mining that extracts information.

Organized from unstructured data. IE is defined for relevant information that meets different criteria. IE data mining techniques efficiently analyze text in free form by extracting the most valuable and relevant information in a structured format. Thus,

---

[1] **Astera Company (2024) . Data Extraction Document , at:** https://www.astera.com/ar/knowledge-center/document-data-extraction/#:~:text=Astera%20ReportMiner%20%D9%87%D9%8A%20%D8%B9%D8%A8%D8%A7%D8%B1%D8%A9%20%D8%B9%D9%86,%D9%85%D9%86%20%D8%A7%D9%84%D9%85%D8%B3%D8%AA%D9%86%D8%AF%D8%A7%D8%AA%20%D8%A7%D9%84%D9%85%D8%B9%D9%82%D8%AF%D8%A9%20%D9%88%D8%BA%D9%8A%D8%B1%20%D8%A7%D9%84%D9%85%D9%87%D9%8A%D9%83%D9%84%D8%A9.

[2] **Astera Company, Op.cit , p2-4.**

the ultimate goal of IE techniques is to identify salient facts from text to enrich databases or knowledge bases.[1]

Named entity recognition (NER)

Named entity recognition is one of the important tasks of IE systems used to extract meta-entities. It helps to identify generic or domain-independent entities such as location, people, organization and domain-specific entities such as disease, drugs, chemicals, proteins, etc. In this process, entities are identified and semantically classified into pre-defined categories.

Traditional NER systems used rule-based methods (RBM), learning-based methods (LBM), or hybrid methods. IE with Neuro-Linguistic Programming (NLP) plays an important role in modeling language and context using morphological, syntactic, phonological, and semantic analysis of languages. Morphologically rich languages such as Russian and English make the process of extracting information easier, and the information extraction technique is difficult for morphologically poor languages because these languages require additional effort for morphological rules to extract nouns due to the lack of a complete dictionary for them. [2]

Automated data mining applications Automated data extraction is rapidly becoming used in many industries, whether it is to gain more accurate insights, faster access to data, or reduce data processing costs. With features such as optical character recognition (OCR) and table extraction offered by advanced automated data extraction tools, companies can Identify and extract data from unstructured documents with high accuracy and minimal manual force.[3]

A good example of an industry that can benefit from data mining automation is logistics. Logistics providers handle a lot of incoming data. From invoices and purchase orders to customer information, people need to process these documents

[1] **Kiran Adnan and Rehan Akbar (2019) . An analytical study of information extraction from unstructured and multidimensional big , , data , Journal of Big Data , Article number: 91 , Springer open , at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0254-8**

[2] **Kiran Adnan and Rehan Akbar , Op.Cit , pp3-20**

[3] **Martin Geletka.. ect..(2022) . Information Extraction from Business Documents , at: https://nlp.fi.muni.cz/raslan/2022/paper18.pdf**

instantly. Any delay or error can lead to serious problems for service providers. Using data mining tools to automate the process, logistics service providers can create templates. These templates easily identify key fields in various unstructured documents, such as PDF files, text files, etc. This allows the program to collect the required information from the documents based on the completed form

Created, where employees can easily extract information from all incoming documents with the same layout, and the user can do this without any coding or recalibration of the program.

The future of automated data extraction

Four trends are expected to witness significant development in the field of data mining in the coming years: [1] First: from **the automation side**

Advances in data mining make it possible to automate the entire process, from the time a document is selected until it is uploaded at the destination. Intelligent data mining tools now automatically scan and pull data from disparate sources with the help of AI and NLP and transform it into the desired output in real time. This eliminates the need for manual scripting. These tools also allow documents to be processed in batches.

Once the template is selected, the tool identifies and performs necessary actions including extraction, conversion, and data quality checking on the document itself to make it usable for business users.

Second: **Capture mobile data In today's dynamic marketplace**, consumers demand not only advanced technology but also portability. Mobile data capture allows cell phones and other portable devices to be used as portable scanners on the go. They upload data to extract from a variety of documents, such as checks and receipts.

Technologies such as Optical Character Recognition (OCR), Intelligent Character Recognition (ICR), and Optical Mark Recognition.[2]

---

[1] **Astera (2023) . Leverage the Data Trapped in Unstructured Sources with Data Extraction , at: https://www.astera.com/wp-content/uploads/2020/07/Leverage-the-Data-Trapped-in-Unstructured-Sources-with-Data-Extraction.pdf**

[2] **) Martin Geletka.. ect , Op.cit , pp3-10**

Third: **extracting data on the cloud**

The papers often reign supreme, with more than 73% of business owners and decision-makers printing information at least four times. Capturing customer data from inboxes or messaging channels and account creation is critical in many businesses. However, it is surprising that a very large amount of data is still recorded on paper or sent to and from companies in an unstructured manner. This makes entering this data into correct systems more difficult, often resulting in longer times for document processing, maintenance, classification, and information retrieval, as well as higher processing and material costs. In short, your business processes are operating less than optimally, affecting the speed at which you can help customers and your bottom line.[1]

The cloud is one of the biggest trends of this decade and has impacted the field of data mining as well. With cloud data mining tools, users can manage data from anywhere. The user just needs a login ID and password for extraction. Automated incoming documents.

These tools are easier to set up because they do not require local drivers In some cases, cloud-based data extraction tools can eliminate the need to set source files for extraction as partners can upload the files themselves, and the data can be extracted according to business requirements. File validation and error checking can also be done in real-time.[2]

Intelligent document processing programs – foundations of choice

Steps for choosing smart document processing programs:

1. The first step in choosing an IDP is to determine the needs of intelligent data processing, that is, determining structured or unstructured data and determining the format and size of the data to be processed within the organization.

2. Next, it must be evaluated which data would be ideal for smart document processing, by determining how long it takes to process each data set, and which

---

[1] **Docbyte Company (2023) . Capturing data intelligently ,p.3 , at:**
https://www.docbyte.com/wp-content/uploads/2019/11/Docbyte-whitepaper-Capturing-data-intelligently.pdf

[2] **Ibid , p.p 3-5**

would benefit the organization the most through automation. Documents that take longer to process manually may be good candidates for smart processing. For documents

3. Once the document processing workflow to be automated is determined, the organization is ready to decide on intelligent document processing software.

This decision is made based on the demand for automation and the organization's requirements for automated document processing. Below are some questions that help in choosing the appropriate program for the smart document processing process:

• How quickly can an intelligent document automation solution be implemented?

• How easy is the software to maintain?

• What is the level of support provided by the software company for intelligent document automation services?

• Can the software's intelligent document recognition technology read all the documents the organization wants to process?

• Will software accuracy significantly improve your error rate?

Examples of the most popular intelligent document processing programs

First: Microsoft Power Automate.[1]

Microsoft Power Automate provides a low-code experience to improve enterprise productivity by automating time-consuming, repetitive tasks and creating intelligent automation based on generative AI.

Power Automate provides a better way to complete processes within the organization through cloud flows, automated process automation (RPA), and data mining, and this program integrates with the integrated Microsoft 365 ecosystem and with the rest of the Microsoft Power platform programs. Features of Microsoft Power Automate

[1] **Microsodt (2024) . Power Platform , at**: https://www.microsoft.com/en-us/power-platform/products/power-automate

1-Easier experience to create, troubleshoot, and manage automation: Power Automate allows an organization to apply the power of generative AI to create richer automation faster, and make it easier to manage automation ownership.

2- Ensure scalability and privacy: Power Automate continues to simplify management and enhance management and privacy controls.

3- Ease of data mining: through artificial intelligence to recommend actionable insights, innovative additional capabilities, and deep integration with computer systems.

Support for using the cloud in Power Automate

Cloud flows are a core automation capability in Power Automate that uses the latest AI technology to allow employees to focus on what matters and automate the rest using natural language and low code. Cloud streaming capabilities include:

1- Experience creating and managing automation processes across applications such as Microsoft Teams, SharePoint, and OneDrive.

2- Integration with the rest of the Microsoft Power platform software, including the ability to invoke workflows from an application created using Power Apps, and when a data alert is triggered in Power BI.

3- Support intelligent processing of documents and include intelligence through the possibility of creating artificial intelligence.

4- Mobile applications help customers receive notifications, perform approvals, and invoke and monitor data automation processes.[1]

Desktop software support Power Automate provides automation for Windows desktop apps, services, and experiences for orchestration and deployment within the Power Platform.

[1] **Microsoft Co. Power Automatic Guided Tour , at:** https://guidedtour.microsoft.com/en-us/guidedtour/power-platform/power-automate/1/1

Supporting the data mining process Power Automate Process Mining empowers organizations to unleash the full potential of their data and discover valuable insights that can accelerate their journey to data operational excellence by:

1-Use ready-made templates for rapid deployment from data ingestion to pre-defined custom reports

2- Gain deep insights using powerful features including rework, process comparison, and custom metrics.

3- The ability to create, view, and publish Power BI reports and share them across the organization.[1]

Below is a simplified explanation of Microsoft Power Automate First: the main menu of the program The main menu of Microsoft Power Automate contains the following items:

[1] **Ibid , p. 3**

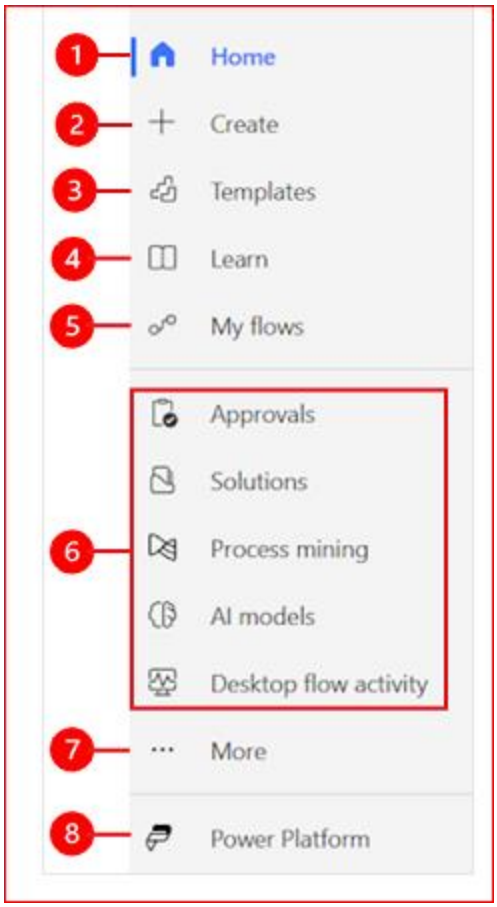

Figure (2) shows the main menu of Microsoft Power Automate

First: The home page contains:

• Create: To create models using the Copilot program supported by artificial intelligence technology.

• Templates: To display templates that you can use to create and search for data.

• Learn: The learning experience takes you to the Power Automate documentation.

• Flows My If a process has been created, or someone else has created a process and shared it with you, it can be viewed or edited.

• Most frequently used pages: When you first log in, items such as Approvals, Solutions, Process Extraction, AI Models, and Desktop Flow Activity appear in the

left navigation menu by default. Use the "More" menu item to uninstall any of these items and install something else.

• More: Used to pin frequently used items to the left navigation pane, such as schedules, cloud streaming activity, and connections.

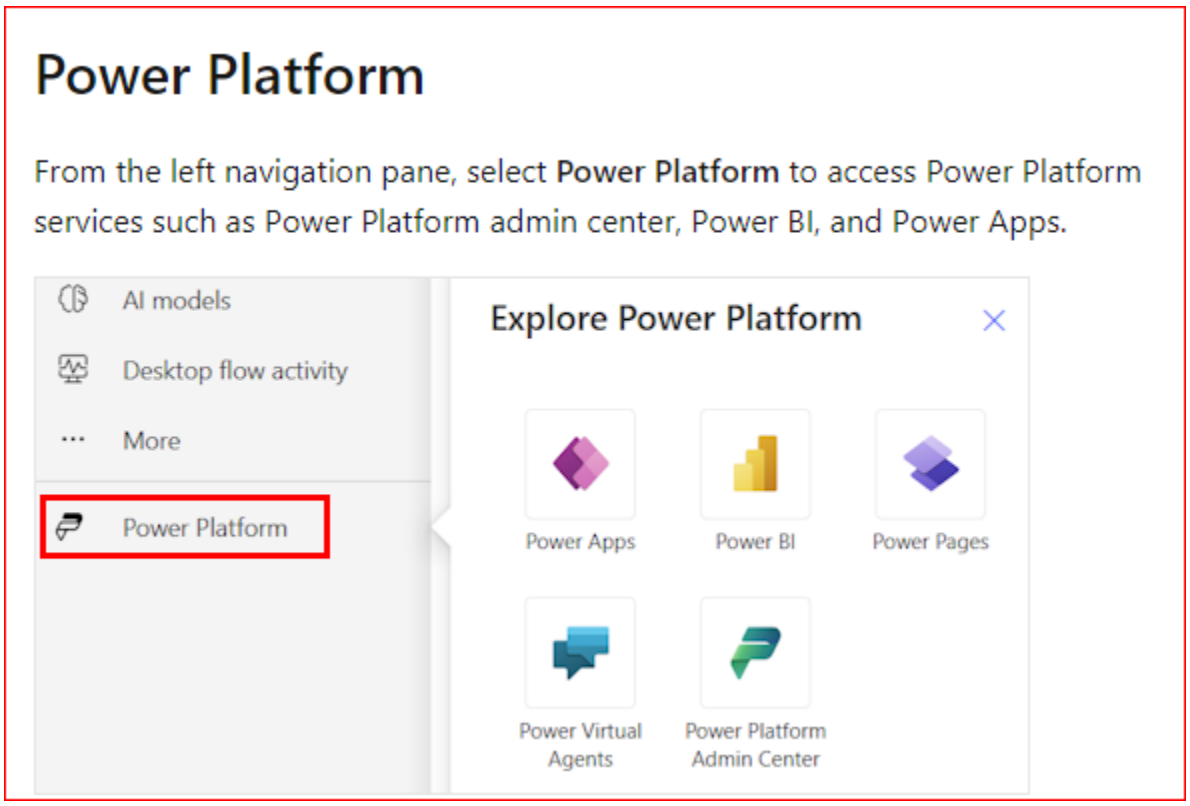

Figure (3) shows the Power Platform

Second: DocAcquire platform [1]

Smart document processing solutions are an essential factor in the digital transformation process, and achieve many benefits, including enhancing efficiency and effectiveness, accelerating operations, reducing expenses, reducing the risk of human errors, simplifying compliance and data security, and scalability - and these solutions can be easily integrated into the technologies used. In institutions

. DocAcquire is an intelligent cloud platform for document processing that automates document-centric business processes by taking advantage of digital transformation technologies and other means that help extract key data from

[1] **DocAcquire (2024) . DocAcquire Platform , at:** https://www.docacquire.com

business documents intelligently to pass them on to the business flow. The company is committed to enabling organizations in various sectors to benefit from... Smooth workflows.

The most important feature of the DocAcquire program is that it supports the Arabic language, which enabled it to enter companies in the Middle East and the Gulf countries, and it has great experience in dealing with the smart processing of documents written in the Arabic language. Therefore, Amnex International, a leader in integrated digital solutions and services, announced the signing of a partnership.

With DocAcquire Limited, a platform specialized in smart document solutions. As a solutions and consulting partner, Amnex will benefit from its successes and experience in the Gulf region to cooperate with institutions that are embarking on the field of digital transformation to enable them to achieve distinctive value in dealing with the huge numbers of unstructured documents that they deal with daily.

DocAcquire program features:

1-DocAcquire applies machine learning to classify, recognize, and extract key information from documents more accurately

2- DocAcquire works to ensure the guaranteed optimal flow of data into business systems. It makes it easier to process data faster.

3- DocAcquire and its open API enable integration with business systems so you can act on them faster.

4- The DocAcquire program is compatible with many applications, whether used on the desktop or on the web, as it allows documents from these applications to be uploaded to the program, such as Google Drive, Gmail attachments, Xero, SQL Server rows, and Dropbox.

5- It allows the automation of any document, as machine learning enables it to automatically recognize and extract basic data from even the most complex documents.

6- Makes document workflow run smoothly within enterprise resource planning (ERP), customer relationship management (CRM), and financial and business systems.[1]

[1] **DocAcquire (2024) . DocAcquire Platform , at:** https://www.docacquire.com

7- Automate document workflow:

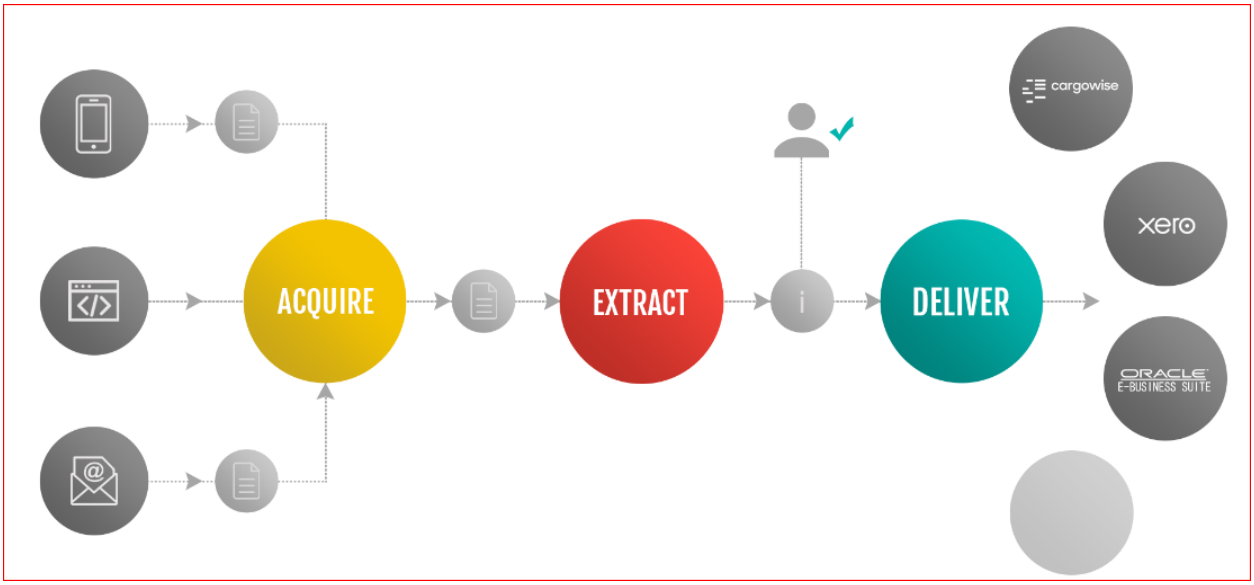

Figure (4) shows the mechanism for extracting data from various documents

Third: Amazon Web Services (AWS) program [1]

The AWS platform is one of the programs provided by Amazon Web Services, which was launched in 2007. It is one of the most comprehensive and widely adopted cloud systems in the world, as it provides more than 200 distinct services from data centers and is used by millions of customers around the world, including Fast-growing emerging companies and most of the leading institutions and government agencies due to its low cost, flexibility, and innovation.

The AWS platform was designed to be one of the most secure and flexible platforms, as a strong infrastructure was created to meet the security requirements of military institutions, global banks, and other highly sensitive institutions, and this is supported by a deep set of cloud security tools, with more than 300 services and features for security and governance. In addition to supporting 143 security standards and security accreditation certificates. AWS also has strong experience,

[1] **Amazon Web Services –AWS. (2024). Intelligent Document Processing, at:**

https://aws.amazon.com/ar/textract/

reliability, and security, and provides cloud services to millions of customers around the world, as AWS services have significant operational experience compared to any other platform.

AWS Platform Services and Intelligent

Document Processing Amazon Web Services (AWS) offers two services to support Intelligent Document Processing (IDP) requirements. They are :

1- Amazon Textract service, which makes it easy to automatically extract handwriting, layout elements, printed text, and data from any document. Amazon Textract uses machine learning to read, process, and understand a document of any type without manual interaction.

Amazon Textract is a machine learning (ML) service that automatically extracts text, handwriting, and data from scanned documents. It is superior to simple Optical Character Recognition (OCR), as it can recognize, understand, and extract data from forms and tables.

Through the Amazon Textract service, the following tasks can be accomplished: [1]

• Extract vital information from business documents with a high degree of accuracy.

• Expand document processing pipelines to have the flexibility you need to adapt to market demands •

Automate data processing in a secure environment that meets compliance standards.

2- Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to uncover valuable insights, links, and relationships in text. It's a fully managed service with continuous training, so you don't have to manage scaling resources, maintaining code, or maintaining training data.

Using Amazon Comprehend, you can accomplish the following tasks:

• Discover valuable insights from text in any document format.

---

[1] **Amazon Web Services –AWS . (2024) . Intelligent Document Processing , at:**

https://aws.amazon.com/ar/textract/

• Simplify the document processing pipeline by extracting sentiments, texts, phrases, or topics from documents.

• Identify and redact personally identifiable information (PII) from private documents.

Data extraction mechanism [1]

1- Files and documents are uploaded programmatically through the AWS Console A, where images (scanned documents, screenshots, or images of documents) are uploaded as input. You can download these files into the input folder.

2- These documents and files are sent to the Amazon Simple Storage Service (Amazon S3) cluster and this action triggers the AWS Lambda function, Textract comprehends Lambda, through event notifications.

3- The Textract Comprehend Lambda function sends the image to Amazon Textract to extract text from the image. When it gets the results, it aggregates the results and sends the text to the custom entity recognizer in Amazon Comprehend. The custom entity recognizer is a pre-trained model that identifies entities in text that are valuable to your company. This post explains how to do this in detail in the following sections.

4- The custom entity recognition tool stores the results in a separate container, which acts as a temporary store for this data. This group contains another event notification, which triggers the ComprehendA2ILambda function. This Lambda function takes the output from the custom entity recognizer, processes it, and sends the results to Amazon A2I by creating a human loop for review and verification.

5- Amazon A2I begins the human loop, providing reviewers with an interface to double-check and correct results that may not have been identified in the custom entity recognition process. These reviewers submit their responses through the Amazon A2I Worker console. When the human loop completes, Amazon A2I sends

[1] **Amazon Web Services –AWS . (2024) , Data Extract , at:**

https://aws.amazon.com/ar/what-is/etl/

an Amazon CloudWatch event, which triggers the Human Review Completed Lambda.

6- The Human Review Completed function checks whether human reviewers have added more annotations (because they found more custom entities). If human reviewers find something that the custom entity recognizer missed, the function creates a new file called Update_entity_list.txt. This file contains all entities that were not present in the previous training dataset.

7- At the end of each day, the CloudWatch alarm runs the New Entity Check function and this function compares the entity_list.txt file and the updated_entity_list.txt file to check if any new entities have been added in the last day. If so, it starts a new training job for the custom entity recognizer in Amazon Comprehend and enables the CloudWatch time-based event trigger that runs the CER Training Complete Check job every 15 minutes.

8- The CER Training Complete Check function checks whether the custom entity recognizer in Amazon Comprehend has finished training. If so, the function adds the entries from update_entity_list.txt to entity_list.txt so that it does not train the model again, unless more entities are found by human reviewers.

9- It also disables the time-based event trigger in CloudWatch, because it does not need to check the training process until it starts again. The following call to the Textract Comprehend function uses the new custom entity recognizer, which has been learned from previous reviews by humans.

Conclusion

1- Digital transformation is the primary gateway towards implementing the artificial intelligence strategy. Through digital transformation, entities can convert protective documents into digital forms to which artificial intelligence technology can be applied.

2- The Document Understand system is one of the artificial intelligence technologies. Through it, the organization is able to access the content of documents and extract data from them, such as tables and drawings, and place them in documents that are compatible with the artificial intelligence environment.

3- Companies working in the field of data management and artificial intelligence provide many programs that are used in the intelligent processing of documents and extracting data from them, such as Amazon, which owns the Amazon Web Services (AWS) program, the Microsoft Power Automate program, and the DocAcquire program.

4- Large institutions and emerging companies resort to Intelligent Document Processing (IDP) technology as one of the important solutions in dealing with unstructured data and storing it in the form of structured data that enables the institution to benefit from it in decision-making.

5- Optical character recognition (OCR) technology is one of the pillars of intelligent document processing (IDP). Through it, a text image can be converted into a machine-readable text format.

6-Natural language processing (NLP) is one of the pillars of intelligent document processing. Through it, computers can analyze, interpret and understand human language with the help of artificial intelligence.

Recommendations

1- The study recommends the need for institutions to digitally transform their documents, which allows the introduction of artificial intelligence technology to deal with data, extract important information from it, and store it in the form of structured data that can be used.

2- The study recommends implementing the Document Understand system, which is one of the applications of artificial intelligence for intelligent handling of documents in government agencies and companies.

3- The necessity of using programs specialized in intelligent document processing and data extraction technology, such as Amazon Web Services (AWS), which saves a lot of time and money spent on analyzing and interpreting data.

4- It is preferable to use the DocAcquire program because it supports the Arabic language, so many companies in the Arab world have used it.

5- The necessity of qualifying human cadres in the field of artificial intelligence and data management and analysis, so that companies can achieve administrative, industrial, commercial, and other development.

6- The study recommends the establishment of a specialized ministry for artificial intelligence in the Arab world, similar to the one in the United Arab Emirates, to work to enter the era of the Fourth Industrial Revolution based on artificial intelligence technology.

References

- Amazon Web Services -AWS. (2024). Intelligent Document Processing, at: https://aws.amazon.com/ar/textract/

- Amazon Web Services -AWS. (2024). Intelligent Document Processing, at: https://aws.amazon.com/ar/textract/

- Amazon Web Services -AWS. (2024), Data Extract, at: https://aws.amazon.com/ar/what-is/etl/

- Astera (2023). Leverage the Data Trapped in Unstructured Sources with Data Extraction, at: https://www.astera.com/wp-content/uploads/2020/07/Leverage-the-Data-Trapped-in-Unstructured-Sources-with-Data-Extraction.pdf

- Astera Company (2024). Data Extraction Document, at: https://www.astera.com/ar/knowledge-center/document-data-extraction/#:~:text=Astera%20ReportMiner%20%D9%87%D9%8A%20%D8%B9%D8%A8%D8%A7%D8%B1%D8%A9%20%D8%B9%D9%86,%D9%85%D9%86%20%D8%A7%D9%84%D9%85%D8%B3%D8%AA%D9%86%D8%AF%D8%A7%D8%AA%20%D8%A7%D9%84%D9%85%D8%B9%D9%82%D8%AF%D8%A9%20%D9%88%D8%BA%D9%8A%D8%B1%20%D8%A7%D9%84%D9%85%D9%87%D9%8A%D9%83%D9%84%D8%A9

- Carlos E. Jimenez-Gomez (2020) . Artificial Intelligence in Government, ResearchGate , At: https://www.researchgate.net/publication/345458831

- DocAcquire (2024) . DocAcquire Platform , at: https://www.docacquire.com

- Docbyte Company (2023). Capturing data intelligently,p.3, at: https://www.docbyte.com/wp-content/uploads/2019/11/Docbyte-whitepaper-Capturing-data-intelligently.pdf

- Esposito , Floriana (2014) . Intelligent document processing , at: file:///C:/Users/ADMIN/Downloads/Intelligent_document_processing%20(3).pdf

- Graham Cutting and Af Cutting-Decelle (2021). Intelligent Document Processing - Methods and Tools in the real world, at: https://arxiv.org/ftp/arxiv/papers/2112/2112.14070.pdf

- Hila ,Mehr (2017) . Artificial Intelligence for Citizen Services and Government , Harvard AshCentreTechnology At: https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

- João Reis (2019) . Artificial Intelligence in Government Services:A Systematic Literature Review , Springer Nature Switzerland , At: https://comum.rcaap.pt/bitstream/10400.26/30168/1/10.1007%40978-3-030-16181-123.pdf

- Kevin C. Desouza (2018) . Delivering Artificial Intelligence in Government:

Challenges and Opportunities , IBM Center for The Business of Government , At: https://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf#

- Kiran Adnan and Rehan Akbar (2019) . An analytical study of information extraction from unstructured and multidimensional big , data , Journal of Big Data , Article number: 91 , Springer open , at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0254-8

- KOFAX Company (2023) . The 6 Essential Elements of Intelligent Document Processing (IDP) , at: https://www.tungstenautomation.com/-/media/files/e-books/en/eb_ephesoft-the-6-essential-elements-of-intelligent-document-processing.pdf

- Martin Geletka.. ect..(2022) . Information Extraction from Business Documents , at: https://nlp.fi.muni.cz/raslan/2022/paper18.pdf

- Microsodt (2024) . Power Platform , at: https://www.microsoft.com/en-us/power-platform/products/power-automate

- Microsoft Co. Power Automatic Guided Tour , at: https://guidedtour.microsoft.com/en-us/guidedtour/power-platform/power-automate/1/1

- Slava Jankin Mikhaylov (2020) . AI for the Public Sector: Opportunities and challenges , 1Institute for Analytics and Data Science , University of Essex , At: https://pubmed.ncbi.nlm.nih.gov/30082303/

- Tanja Sophie Gesk (2022) . Artificial intelligence in public services: When and why citizens accept its usage , At: https://www.researchgate.net/profile/Tanja-Gesk/publication/359738929_Artificial_intelligence_in_public_services_When_and_why_citizens_accept_its_usage/links/62501f61cf60536e234cc832/Artificial-intelligence-in-public-services-When-and-why-citizens-accept-its-usage.pdf?origin=publication_detail