UDC: 004.932 https://doi.org/10.15407/jai2023.01.013

AMAZON TEXTRACT AND ARTFICIAL INTELEGENCE SYSTEM AT BANKING DOCUMENT MANAGMENT SYSTEM

S. Ustenko[1], **T. Ostapovych**[2]
Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine
54/1, Peremogy Ave., Kyiv, 03057
stanislav.ustenko@kneu.edu.ua
ostapovych@meta.ua
[1]https://orcid.org/0000-0001-9356-4742
[2]https://orcid.org/0000-0001-6742-3575

Abstract. Electronic document management systems have a great prospect of use in the banking sector, all information stored in electronic document management systems requires further analysis and processing, this involves the use of a machine learning service to build a semantic search result, which implies the presence of a search service with the thinking of artificial intelligence and the ability provide links to clearly reasoned answers.

Such a service that satisfies the needs of semantic search is the Amazon Kendra service, the question of using such a service is more relevant than ever for the construction of modern banking products.

Under such conditions, an important area of research is the assessment of the efficiency of Amazon Kendra in the banking sector, which necessitates the development of a conceptual model for assessing the efficiency of banks for making management decisions aimed at improving the efficiency of individual banks and the banking system as a whole.

Objectives: The purpose of this work is to improve the work of electronic document flow in the banking sector using Amazon Kendra and Amazon Textract to design an innovative banking product and develop the banking sector of Ukraine.

Methods/Approach Scientific research methods – both comparative and analytical – is used in the process of drawing up of this article.

Results: A semantic search system based on the bank's electronic document flow system was designed.

Keywords: information and communication technologies, innovative technologies in the banking sphere, digitalization processes, bank activity, electronic document circulation system, innovative banking products, Amazon Kendra.

Introduction

Innovative development of the banking sector in the direction of modeling the implementation of information technologies to support innovative bank products and services is extremely important. In this direction, the Scientific Research Institute "Institute of Information Systems in the Economy" of Vadym Hetman KNEU conducts scientific research (R&D), in particular on the topics: "Development of methods and technologies of intellectual management of organizational structures in the conditions of the digital economy" (state registration number 0119U002604) and "Modeling of processes of implementation of information technologies supporting innovative products and services of banks" (R&DKR is registered by the State scientific institution "Ukrainian Institute of Scientific and Technical Expertise and Information", State registration number: 0122U001987, date of registration: 11-03-2022), scientific supervisor, Doctor of Economics, prof. Ustenko S.V. According to the current results of these research works, scientific articles have been published in international monographs (Ustenko, 2020), (Ustenko, 2021), (Ustenko, 2020), (Ustenko, 2019). The relevance of the research topic is due to the fact that in market conditions banking products and services play a key role in the functioning of the financial system and the market. This leads to the urgent need to build intelligent information systems for the interaction of banking institutions with the user, the involvement of artificial intelligence, in particular neural networks. The main feature and innovation of such systems is that they have the property of machine learning and with each new training the system improves its performance. In information and communication systems and technologies supporting the information security of banking

activities and conceptual approaches to the sustainable development of Ukrainian banks based on the general principles of banking education, the main ones of which are the principles of integrity, stability, digitalization and structural-logical connections of elements and the banking system as a whole, which requires a generalization of approaches to model studies and technologies for using banking systems (Ustenko, 2019). The work is devoted to the study of the conceptual basis of the processes of information provision of digital educational activity, which does not take into account the production (operational) sphere of activity of enterprises and organizations (Ustenko, 2022). Publications provide approaches, trends and factors of economic growth in the most technologically developed countries (Tew, 2017), (Hussaini, 2020), (Dusange, 1994), (Millier, 2011). Technological development is one of the important factors of economic growth and includes the use of a set of production technologies and scientific methods that must be taken into account for a reasonable analysis and assessment of banks' activities. At the same time, there is an urgent need to develop a general (conceptual) model for evaluating the effectiveness of bank activity, which can take into account key performance indicators of a number of bank subsystems, including operational, economic, financial, managerial, information technology, etc. (Dusange, 1994), (Millier, 2011). The implementation of the conceptual model in each bank will allow at the system level to conduct model experiments to assess the effectiveness of the bank's functioning and development, develop practical recommendations and ways to increase the efficiency of Ukrainian banks, take into account the introduction of banking services and provide banking services to clients.

Since the beginning of 2014, the banking system of Ukraine has experienced one of the strongest crises in its history. In terms of banking assets as a percentage of gross domestic product, Ukraine's banking sector was similar to Poland's. However, by 2016, bank closures and reduced lending led to a sharp reduction in the role of banks in the economy. Today, Ukraine lags far behind many European banks. As of October 2020, out of 180 banks operating at the beginning of 2014, the National Bank of Ukraine declared 104 insolvent or liquidated, which is almost 60% of the country's banks. It should be noted that the assets of some Ukrainian banks in 2014 were overstated due to the concealment of loans granted to related parties, but many banks, unfortunately, did not have the opportunity to model and forecast the impact of internal and external destabilizing factors on the activities of a financial institution, which led to the search tools and approaches for strategic analysis, performance evaluation and development of banks. Banks are at the epicenter of these changes. Technological developments and social changes have a deeper and more immediate impact on the financial industry than on most other sectors, as its primary raw materials are information and money. And money, in turn, can be dematerialized and turned into accounts, in other words, into data that can be stored, processed and transmitted in real time with little cost (Ustenko, 2021), (Ustenko, 2020).

In 2022, under the conditions of the global energy crisis, which arose as a result of missile attacks on critical infrastructure facilities of the state of Ukraine, from the territory of russia, the banking sector faced a new problem, namely the migration of infrastructure to Amazon's data centers, and the provision of new functionality to smartphone users.

New technology always present a challenge for developers of innovative banking services, such as semantic search it actually became built in feature of new smartphone to recognize humane voice and type text over the voice. But services of electronic document management system up to this days doesn't support the semantic search capabilities. It is important for banking sector to stay updated and bring new services to the market for better quality of services and fulfillment the requirements of customers. It is obvious that person with new smartphone will ask for new capabilities of banking services such as semantic search, other words, the banking client will ask for possibility to search information inside the document which is stored in electronic document management

system, by asking question to the smartphone, that brings a challenge to recognize the question from voice and recognize the meaning of the question in such a way that every words from the question should bring the value for the overall meaning of the question and don't just match by character but instead match the meaning of the question, that was asked. Those challenges should be overcame with new functionality of semantic search it s obvious that technology for voice recognition is already at the place. New smartphones could recognize the words and type those words into the sentence, but from technology point of view those words would be only the character without the meaning, whats a pity. For further words and sentence recognition new way of understanding the words should be put into place of banking services. Such a technology that will understood the meaning of the sentences in whole not just the meaning of the words. That will bring the value for banking sector and banks in particular and bring new functionality for the client of banking sector. Due to obvious necessity to update the technology of clients it is a time to think ahead and update the software technology of the banking sector. New functionality don't just bring the value but also bring new possibility for the client and open those possibility that will be key for clients success. Information, and structured information always was a challenge for the industry but it is possible to make information not only structured but also find the meaning of information that is stored in data centers. In case the meaning of information will be unlocked new way of working with information should be presented for banking client as key advantages of the bank that use electronic document management system. Faster and more accurate search result will reduce the amount of time which is necessary for searching of documents, that is especially helpful if we take into account that amount of information always and constantly grow from year to year, that bring additional challenge for the technology that store documents such as electronic document management systems. The bank who will implement new semantic search will have key advantages over the competitors, due to the reason that capitalism without competition is exploitation, it is time to think carefully about the banks institutions that doesn't accept new technology. Such competition, for better future and better services already force banking industry to change the attitude to data centers. Previously there was a requirement that data center should be physically located inside Ukraine motherland. That attitude already changed, for the last eight month the biggest banks of Ukraine migrate their software from Ukrainians data center to Amazon data center, it really helps to provide the banking services during the war period, russia was not able to stop the blood of Ukrainian economy and even the challenges such implementing new technology shouldn't be the challenge for banking industry. The information of Ukrainians banks already stored in cloud data center, that will help to implement new technology of semantic search of electronic document management system. Capabilities of semantic search is an advantages, the advantages that will play a key role in competition between the biggest banks of Ukraine.

New clients of banking institution will definitely have new way of thinking, as an example they will use voice search for searching information and will ask question instead of matching the information new banking client will expect the result with the meaning. Such result could provide service Amazon Kendra with semantic search capabilities in combination with service Amazon Textract which have artificial intelligence capabilities working together they will provide necessary capabilities for semantic search of new banking client with new smartphone. Such approach of using semantic search will definitely speed up search of information, which is necessary part of innovative banking product. Also it is right to mention that new smartphone increase the load on existing document management system, due to the reason that new smartphone had new faster processor they speed up the search of the information, that increase load on document management system. In order to keep up with new client who expect information to be search even faster it is necessary to migrate existing infrastructure to the cloud and start using semantic search which could be provided with

service Amazon Kendra and Amazon Textract. The main benefits of using Services Amazon Kendra and Amazon Textract is speed of searching documents and also accuracy due to the reason that information which are searched via Amazon Kendra is searching with meaning of the search query instead of just matching the worlds, it is obvious that all future innovative banking services will use semantic search and now days semantic search capabilities will advantages over the competitors. Such useful advantages should be constructed with the meaning behind that time especially time that people spend searching information are spending in huge amount in bigger organization, that is why advantages that provide new capabilities that require to spend less time is especially useful.

We live in a world that is far from ideal and our effort to make it better will be evaluated by millions of users. New technologies solve some old issue but open even bigger challenges of our species. Young children receive new technologies even faster than previous generation. That is why semantic search actually help resolve issue with search results in such a way that reduce time that is necessary to find an answer faster and more accurate than ever before, for so long time the search result was inaccurate and banking manager spend a lot of time to find an answer for such an obvious for human understanding question. All of this challenges could be over come with AWS Kendra service and AWS Textract service in such a way that out come of search result will be accurate and could be presented for new smartphone users of innovative banking services.

Methodology

Methods and materials: to achieve the goals of the study, comparative and analytical methods of scientific research were used. Research materials consist of the analysis of documentation texts, scientific articles and publications, as well as practical experience gained on the research topic.

Results and discussion

Amazon Textract at banking document management system

Banking has not yet undergone the transformation that other information sectors have undergone. This is largely due to the fact that banking has historically been a highly regulated industry subject to close supervision and control by government authorities. However, the transformation of the industry is not only inevitable, but also gaining momentum every day. The main reason is that the technological revolution introduces new ways of doing business every day and increases the potential to reduce costs, and the number of users who resort to non-traditional methods of banking continues to grow. Another reason for the transformation is that the current crisis is causing changes in different directions. Banks are perceived as the "culprit" of the recession, and rightly so, because many institutions made very serious mistakes and chose to ignore the basic principles of banking: prudence, transparency and even honesty. As a result of these mistakes, many banks faced serious difficulties, with some banks failing and others undergoing complete restructuring, usually financed with public funds. The colossal amount of taxpayers' funds invested in savings banks caused serious damage to the reputation of financial institutions and the entire industry in the eyes of ordinary citizens. The crisis also triggered a process of radical changes in banking regulation: credit limits, increased capital and reserve requirements, the need for large investments to improve risk and compliance systems, etc. All this comes down to a decrease in income and an increase in expenses, in other words, to a decrease in the current and future profitability of financial institutions. Banks must respond to the new demands of their customers and society, meet this challenge with a damaged reputation, lower profits and slower growth rates of traditional banking business. Such a situation requires a radical transformation: banks must radically revise the way they interact with clients and make a qualitative leap in the direction of the efficiency of their activities. To a certain extent, the increase in efficiency will be achieved due to the sharp consolidation of the banking sector, which has already begun. But the true transformation of the industry will be achieved through the broad and above all,

intelligent use of technology as part of a continuous process of innovation.

In recent decades, banks have been among the most important users of information and communication technologies, which they have adopted with two main goals: to reduce costs and optimize processes to increase profits, and to develop communication channels that are different from the usual ones. With the development of banking, the Internet has become a leading source of information, an indispensable business communication and even a forum for personal relationships: now more than a billion people around the world use various social networks. The Internet also contributes to the fragmentation of banks' production chains, facilitating the outsourcing of services. Banking services offered by cloud computing are a major breakthrough in universal access to datastorage and processing at very low costs and will have far-reaching consequences. The use of the Internet has also increased significantly due to the development of mobile phone technology. Thanks to these new devices, almost 4.5 billion people are online and have almost universal access to some level of information services, which has a huge impact on productivity (Ustenko, 2019), (Ustenko, 2018). Mobile phones are equipped with more and more powerful and diverse functions, which will gradually be included in other devices, additional services and services of banking systems ("Internet of Things", "Internet banking"). The methodology of researching the processes of functioning and development of banks is based on a general analysis and principles of bank development and takes into account a comprehensive approach to researching the processes of effective development of banks (Ustenko, 2019). A comprehensive approach to the study of bank development processes is focused on the holistic development of all processes, not individual processes, which contributes to the comprehensive development of the bank. This approach allows taking into account the information technology aspects of banking services, developing new banking products and using modern information technologies and banking systems. The basis of the information technology support of banks is the process of implementing digitalization as a tool for bank development and scaling. Digitization is the direction of development of banks in the sense of the introduction of modern digital technologies, aimed at the transition to automated digital technologies, controlled by real-time intelligent systems in constant interaction with the external environment outside the boundaries of one bank, with the prospect of unification on a global scale of the Internet of Things and Services network. Today, the first steps in the implementation of digitalization are the introduction of such technologies as machine learning, blockchain systems, blockchain systems, AR technologies (augmented reality), AWS cloud technologies (cloud technologies), data processing systems (Ustenko, 2021), (Ustenko, 2020), (Tew, 2017).

To improve the work of the bank manager, it is suggested to use the Amazon Textarct system to obtain semantic text recognition.

Simply put, AWS Texttract is a deep learning-based service that converts different types of documents into an editable format. Consider that we have hard copies of invoices from various companies and keep all important information from them in Excel/spreadsheets. We typically rely on data entry operators to enter data manually, which is stressful, time-consuming and error-prone. But using Texttract, all we need to do is load our invoices into it, and in turn, it returns all the text, forms, key-value pairs, and tables in the documents in a more structured way. Below is a screenshot of how AWS performs intelligent information extraction.

Not only printed text, AWS Texttract also identifies handwritten texts in documents. This makes information extraction more useful because in some cases handwritten text is more difficult to extract than typed text. Now let's look at some typical use cases for Texttract. Reliable and standardized data collection: Amazon Texttract allows you to extract text and tabular data from a variety of documents, such as financial documents, research reports, and medical records. However, these are not dedicated APIs, but they learn on a huge amount of data every day, and with this continuous learning, extracting unstructured

and structured data from your document will be much easier.

Key-Value Pair Extraction: Key-value pair extraction has become a common problem for document processing, but Amazon Texttract can easily solve it. We can create pipelines for extracting key-value pairs using Texttract, which automates document processing, from scanning documents to sending data to an Excel sheet, etc.

Create an intelligent search index: Amazon Texttract allows you to create libraries of text found in images and PDFs.

Using Intelligent Text Extraction for Natural Language Processing (NLP) – Amazon Texttract allows you to extract text into words and strings. It also groups text by table cells if Amazon Texttract document table parsing is enabled. Amazon Texttract gives you control over how text is grouped as input to NLP.

We will discuss how AWS Texttract works. We know that powerful AI and ML algorithms are behind them; however, there are no open source models to go into detail. But I will try to decipher the work by summarizing the documentation that can be found here (Fig. 1).

Fig. 1. Amazon Textract

First, whenever a new or scanned document is submitted to Texttract, it creates a list of block objects for all detected text. For example, say an invoice today is a hundred words long, AWS creates a hundred block objects for all the words. These blocks contain information about the detected item, its location, and Amazon Texttract's confidence in the accuracy of the processing.

Usually, most documents consist of the following blocks:

Page

Lines and words of the text

Form data (key-value pairs)

Tables and cells

Selection of elements

Below is an example and AWS Texttract data block structure:


```
{
 "Blocks":[
    {
"Geometry": {
 "BoundingBox": {
    "Width": 1.0,
    "Top": 0.0,
    "Left": 0.0,
    "Height": 1.0
 },
 "Polygon": [
    {
       "Y": 0.0,
       "X": 0.0
    },
    {
       "Y": 0.0,
       "X": 1.0
    },
    {
```

```
                    "Y": 1.0,
                    "X": 1.0
                  },
                  {
                    "Y": 1.0,
                    "X": 0.0
                  }
                ]
              },
              "Relationships": [
                {
                  "Type": "CHILD",
                  "Ids": [
"2602b0a6-20e3-4e6e-9e46-3be57fd0844b",
"82aedd57-187f-43dd-9eb1-4f312ca30042",
"52be1777-53f7-42f6-a7cf-6d09bdc15a30",
"7ca7caa6-00ef-4cda-b1aa-5571dfed1a7c"
                  ]
                }
              ],
              "BlockType": "PAGE",
              "Id": "8136b2dc-37c1-4300-a9da-
6ed8b276ea97"
            }.....

        ],
        "DocumentMetadata": {
          "Pages": 1
        }
      }
```


However, the contents inside the blocks change depending on the operation we call. For a text detection operation, the blocks return the pages, lines, and words of the detected text. If we use document parsing operations, the blocks will return detected pages, key-value pairs, tables, selections, and text. However, this only explains how Texttract works at a high level, in the next section let's dive into the OCR behind Texttract.

There are no specifics about the type of OCR that Amazon Texttract uses because it is a commercial product. However, we can compare it with one of the most popular open source OCRs, "Tesseract", to understand its accuracy and ability to extract different types of documents.

Tesseract OCR is based on LSTM, a deep learning-based neural network architecture that works extremely well with text data. Below are the document formats supported by tesseract: plain text, hOCR (HTML), PDF, PDF with invisible text only, TSV. It supports Unicode (UTF-8) and supports over 100 languages out of the box. However, since the entire code is open source, it can be trained to recognize other languages, but this requires deep learning and computer vision expertise. When it comes to table and key-value pair extraction, tesseract fails. However, we can create our own pipelines to solve this problem.

Texttract OCR is also a deep learning-based neural network architecture, but it cannot be fully customized or trained on a custom dataset. Its task is to analyze and extract all the data contained in the document. However, Texttract automatically adjusts to your data and achieves higher accuracy on the fly if a human verifies the extracted information (human in the loop). For tasks like table extraction and key-value pair extraction, Texttract does a good job, achieving higher accuracy than Tesseract. But it is limited to only a few languages and document formats.

Below are some of the document types that can be processed with AWS Texttract:
Regular Accounts/Accounts
Financial documents
Medical documents
Handwritten documents

Payment information or documents of the employee.

The Amazon Texttract API can be used in a variety of programming languages. We'll look at a key-value extraction code block using Python's Texttract. To learn more about the language and API support, check out the documentation here.

This code snippet is an example of how we can perform key-value pair extraction in documents using Texttract's Python API. For this to work, we will also need to configure the API keys in the AWS dashboard. Now let's dive into the code snippet. First, we import all the necessary packages to send documents to AWS and process the extracted text.

```
import boto3
import sys
import re
import json
```

Next we have a function called get_kv_map, here we use boto3 to communicate with the Amazon Texttract API, load the document and get the block response. Now we get all the key-value pairs by checking for "BlockType" and returning it to the dictionaries.

```
def get_kv_map(file_name):
with open(file_name, 'rb') as file:
img_test = file.read()
bytes_test = bytearray(img_test)
print('Image loaded', file_name)
# process using image bytes
client = boto3.client('textract')
response = client.analyze_document(Document={'Bytes': bytes_test}, FeatureTypes=['FORMS'])
# Get the text blocks
blocks=response['Blocks']
# get key and value maps
key_map = {}
value_map = {}
block_map = {}
for block in blocks:
block_id = block['Id']
block_map[block_id] = block
if block ['BlockType'] = "KEY_VALUE_SET":
if 'KEY' in block['EntityTypes']:
key_map[block_id] = block
else:
value_map[block_id] = block
return key_map, value_map, block_map.
```

After that, we have a function that gets the relationship between the extracted key-value pairs using the block elements. Essentially, using the relationships found in the block information (JSON), this function links keys and values in a document.

```
def get_kv_relationship(key_map, value_map, block_map):
kvs = {}
for block_id, key_block in key_map.items():
value_block = find_value_block(key_block, value_map)
key = get_text(key_block, block_map)
val = get_text(value_block, block_map)
kvs[key] = val
return kvs
def find_value_block(key_block, value_map):
for relationship in key_block['Relationships']:
if relationship['Type'] == 'VALUE':
for value_id in relationship['Ids']:
value_block = value_map[value_id]
return value_block
```

Lastly, we return the text present in the saved key-value pairs.

```
def get_text(result, blocks_map):
text = ''
if 'Relationships' in result:
for relationship in result['Relationships']:
if relationship['Type'] == 'CHILD':
for child_id in relationship['Ids']:
word = blocks_map[child_id]
if word['BlockType'] == 'WORD':
text += word['Text'] + ' '
if word['BlockType'] == 'SELECTION_ELEMENT':
if word['SelectionStatus'] == 'SELECTED':
text += 'X'
return text
def print_kvs(kvs):
for key, value in kvs.items():
print(key, ":", value)
def search_value(kvs, search_key):
for key, value in kvs.items():
if re.search(search_key, key, re.IGNORECASE):
return value
def main(file_name):
key_map, value_map, block_map = get_kv_map(file_name)
# Get Key Value relationship
kvs = get_kv_relationship(key_map, value_map, block_map)
print("\n\n== FOUND KEY : VALUE pairs = \n")
print_kvs(kvs)
# Start searching a key value
while input('\n Do you want to search a value for a key? (enter "n" for exit) ') != 'n':
search_key = input('\n Enter a search key:')
print('The value is:', search_value(kvs, search_key))
if __name__ == "__main__":
file_name = sys.argv[1]
main(file_name)
```

So we can use the AWS Texttract API to perform various information extraction tasks. The functions/approach are similar to most

programming languages. We can also customize the approach based on our use cases if we want to use the API.

Amazon Texttract is a machine learning (ML) service that automatically extracts text, handwritten text, and data from scanned documents. It goes beyond simple optical character recognition (OCR) to identify, understand and extract data from forms and tables.

Amazon Kendra is a document search and indexing interface. Amazon Kendra can be used to create an updatable index of various types of documents, including plain text, HTML files, Microsoft Word documents, Microsoft PowerPoint presentations, and PDF files. It has a search API that can be used from a number of client applications, including websites and mobile applications. Other services are integrated with Amazon Kendra.

For example, you can use Amazon Kendra search to run Amazon Lex chatbots and provide answers to user queries. Amazon S3 can be used as a data source for your Amazon Kendra index. AWS Identity and Access Management can also be used to manage access to Amazon Kendra resources.

Amazon Kendra consists of the following elements:

Index provides a client-side search API. The index consists of source documents.

Documents to be indexed are stored in the source repository.

The data source synchronizes the documents of your source repositories with the Amazon Kendra index. You can synchronize your data source with the Amazon Kendra index to update the index with new, updated, and deleted files from the source repository.

A document addition API that directly adds documents to the index.

Benefits of using Amazon Kendra:

- Get answers in natural language: We can use simple keywords to search. It will return the best answers to your query, whether your answer is in a document, FAQ, or PDF. It will also provide suggested answers rather than going through a long list of documents. In the image below, we can see the difference in how Amazon Kendra returns results after a search.
- Content Access: With Kendra, we can easily access content from various repositories like SharePoint, Amazon S3, ServiceNow, and Salesforce into a centralized index that allows you to search all questions in your data and find the exact answer.
- Fine-tuning search results: We can fine-tune search results by manually adjusting the importance of data sources or by using custom tags.
- Deploy with just a few clicks: Just a few clicks. We can set up the index, connect the appropriate data sources, and start using Kendra to find answers to our questions.

Amazon Kendra users can ask the following types of questions or requests. Factual questions are simple who, what, when, and where questions whose answers are based on facts that can be given in a single word or phrase. Descriptive questions are questions with a single line, section, or full text as the answer.

Search by keywords - when the purpose and scope of the question are unclear. Amazon Kendra can determine user intent from a search query and return results that match the user's expected value.

Amazon Kendra is a widely used service defined as an intelligent search (ML) service powered by machine learning. Amazon Kendra redefines business search for user websites and applications so that their employees and customers can quickly find the information they need, even if it is located in multiple locations and content repositories within the company. With Amazon Kendra, users can stop sifting through reams of unstructured data and instead find relevant answers to their queries when they need them. Because Amazon Kendra is a fully managed service, there is no need to configure servers and train or install machine learning models. Use natural language queries in addition to basic keywords to get the information you need. Whether it's a text snippet, an FAQ, or a PDF document, Amazon Kendra will provide the exact answer from it. Instead of searching for exact answers in huge lists of documents, Amazon Kendra offers suggestions in advance. Amazon Kendra is also defined as a service that offers intelligent search capabilities for websites and applications. With this service, employees can easily identify the material they need, even if the data is stored in multiple

locations, and get the right answers to their queries when they need them.

Amazon says goodbye to browsing through long lists of links and browsing through articles in the hope of finding something that will help users. Natural language search capabilities, unlike traditional search technologies, provide the answers users are looking for quickly and accurately, regardless of where the content is stored in their company, so they find relevant answers quickly. Amazon Kendra easily aggregates content from content repositories such as Microsoft SharePoint, Amazon Simple Storage Service (S3), ServiceNow, Salesforce, and Amazon Relational Database Service (RDS) into a centralized index using Amazon Kendra. It allows users to quickly search all of your enterprise data and find the most accurate answer, thus centralizing access to knowledge. The deep learning models used by Amazon Kendra have been pre-trained in 14 industries, helping to produce more accurate answers in a variety of business use cases. Users can also fine-tune search results by directly prioritizing data sources, authors, or relevance, or by applying custom tags, thus customizing search results. Compared to traditional search solutions, Amazon Kendra is quick to configure, allowing users to access Amazon Kendra's advanced search capabilities more quickly. Without any programming or machine learning skills, users can simply create an index, link relevant data sources, and launch a fully functional and customizable search interface with just a few clicks of the mouse, and thus it deploys with just a few clicks of the mouse.

As with any data discovery tool, metadata is key. We will use the S3 databases and tables available in the AWS Glue data directory. To make this information searchable through Amazon Kendra, I needed to prepare the metadata (ie, the database and table names in the AWS Glue data catalog) in a format that could be indexed in Amazon Kendra. It's very easy with boto3's AWS Python SDK. See the example below (see Figure 1):

```
def get_all_glue_tables():
"""
Function to get all tables in AWS Glue Data
Catalog
"""
glue_tables = []
kwargs = {}
response = glue.search_tables(**kwargs)
glue_tables.extend(response['TableList'])
while 'NextToken' in response:
token = response['NextToken']
kwargs['NextToken'] = token
response = glue.search_tables(**kwargs)
glue_tables.extend(response['TableList'])
return glue_tables
```

Fig. 1. AWS Python SDK від boto3

With metadata added as documents to Amazon Kendra, it's time to experience data discovery. Our first query was to find user session data. To do this, Amazon Kendra returned the correct results along with a suggested answer that matched what we were looking for. Additionally, based on metadata and Facet configuration in Amazon Kendra, I can filter the columns I'm interested in or the types of tables (views or external tables, see Figure 1).

After examining the session data, our task is to review the data available for conversion. So we just ask Amazon Kendra, "Where's the conversion data." Voila, the result, as seen in fig. 2.

Finally, we want to see the tables with the eventId column so we know which tables or views to join for analysis (see Figure 4).

Search allows you to ask questions in natural language. eg "where is eventid used?" or "where is the conversion data?". This capability makes it easy for anyone to find the relevant data they need for analytics. Thus, the time required to search for data is reduced.

Amazon Kendra document attributes can be used as filters, in this case column names, providing an intuitive user interface for filtering.

The architecture of the electronic document management system using Amazon Kendra and Amazon Textract is presented in Fig. 5.

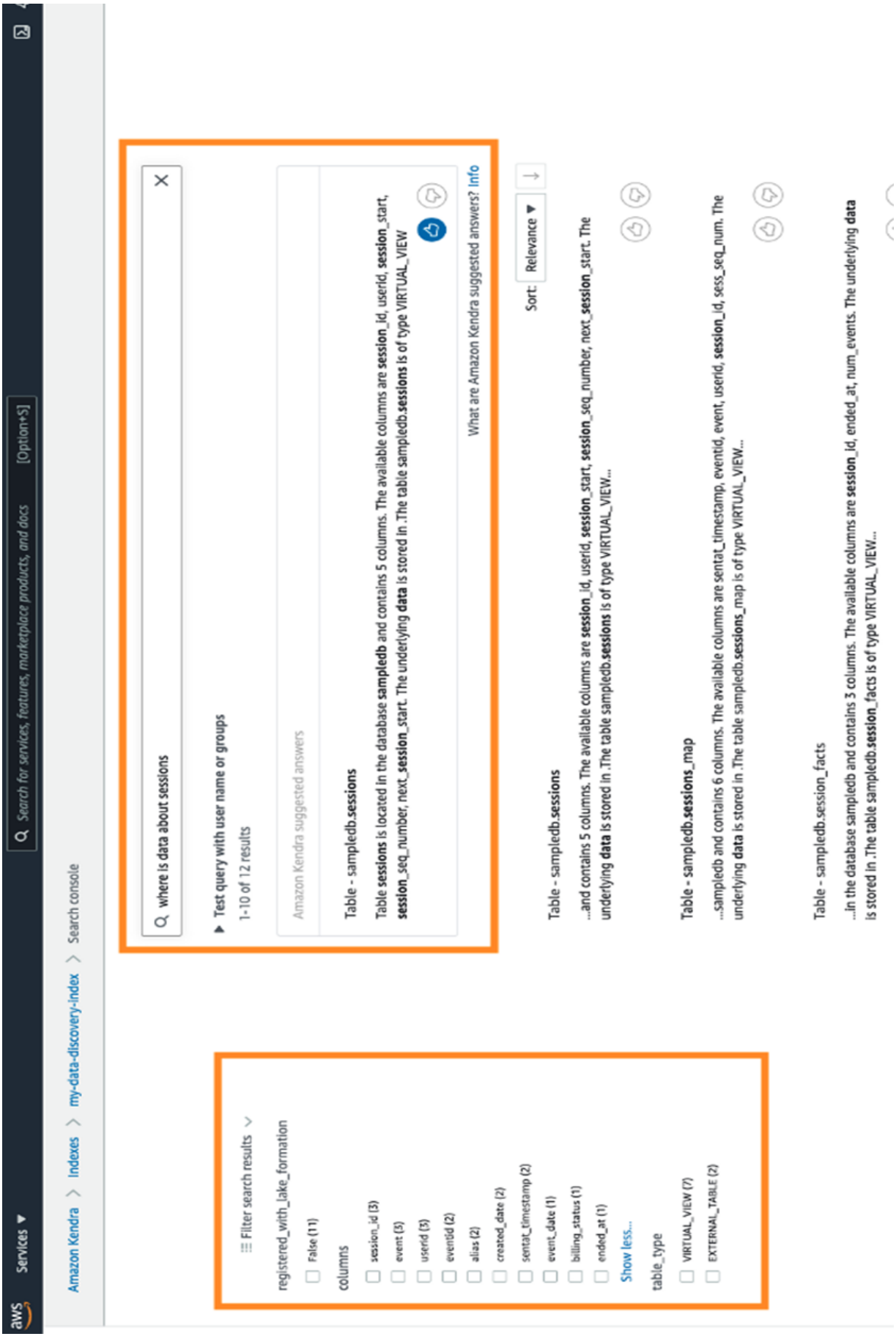

Fig. 2. Search for user session data

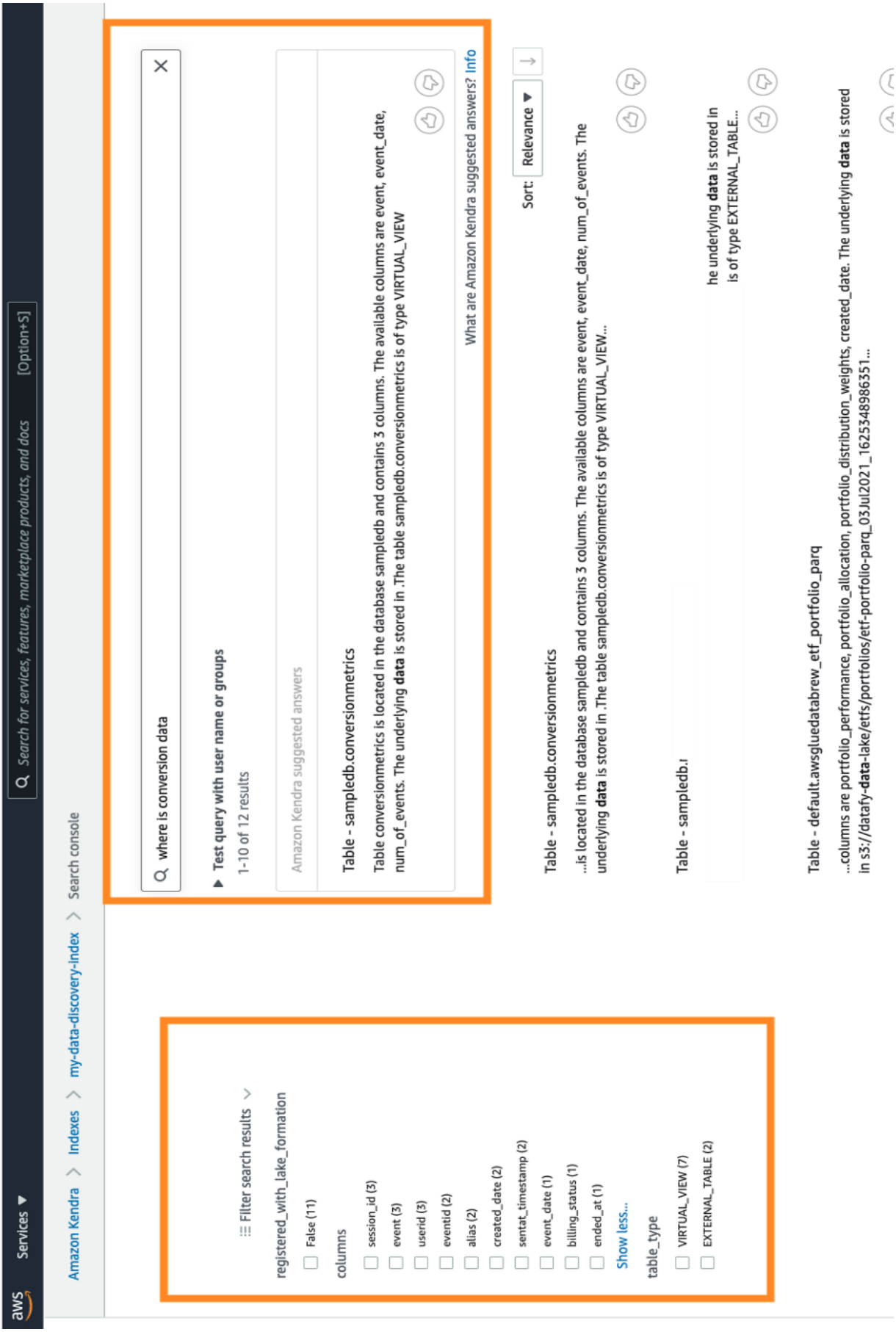

Fig. 3. Data available for conversion

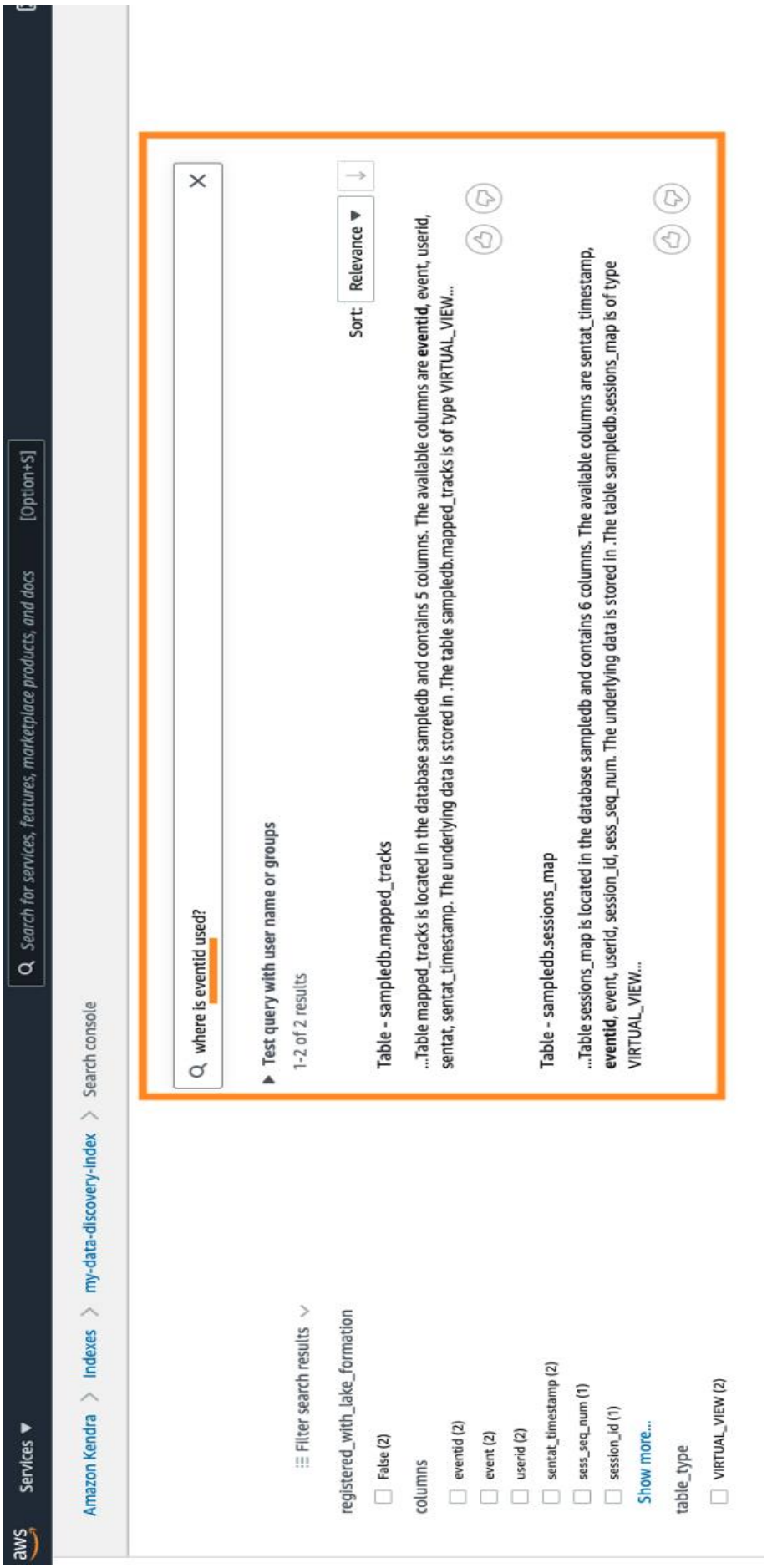

Fig. 4. Tables with an eventid column

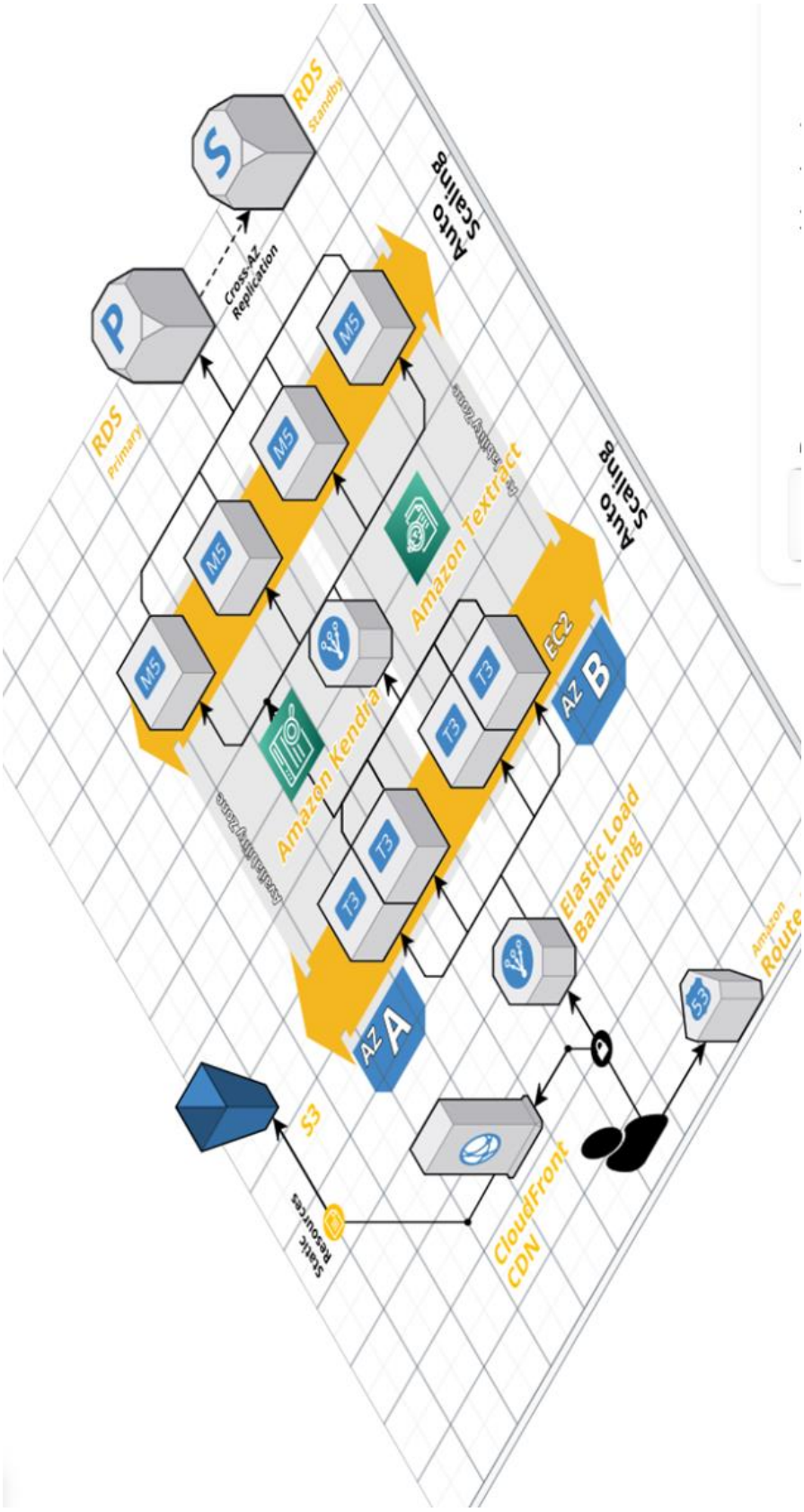

Fig. 5. Architecture of the electronic document management system Amazon Kendra and Amazon Textract

Conclusions

The new economy of Ukraine must find answers to challenges called energy crises, the state in a smartphone, a bank in a smartphone. This is the new reality of Ukraine. The main challenge of our generation and our time. If the challenge is answered properly, the economy will be stronger than ever. Finance is the lifeblood of our economy and we must see banks as a key tool to support the economy so that payments can be made and key banking services can be obtained. The article describes one of the key factors that will enable a smartphone to provide semantic search and artificial capabilities. Obviously, a person with an Iphone 14 will use the voice search function and ask questions in fluent language, these features of the banking product can be achieved by using Amazon Kendra and Amazon Texttract.

Author Contributions:

Conceptualization, T.O S.U methodology, S.U; formal analysis, T.O; investigation, T.O; project administration, S.U; data curation, T.O; resources, T.O; supervision, S.U; validation, T.O; writing—original draft preparation, T.O S.U; writing—review and editing, S.U.

All authors have read and agreed to the published version of the manuscript.

Data Availability Statement:

The data presented in this study are available on request from the corresponding author.

Acknowledgement

The paper was supported by Institute Information Technologies in Economics in the implementation of the research work "Modeling of the processes of implementation of information technologies supporting innovative products and services of banks".

Conflict of interests

The authors declare no conflict of interest.

References

1. Dusange, P., & Ramanantsoa, B. (1994). Technologie Et Stratégie D'entreprise, Édition International. Ediscience International. No. 1. P. 248. Paris, France.

2. Hussaini, N. (2020). Economic Growth and Higher Education in South Asian Countries: Evidence from Econometrics. Int. J. High. Educ. No. 9. P. 118–125.

3. Millier, P. (2011). Stratégie Et Marketing De L'innovation Tech-nologique. 3ème Édition: Lancer Avec Succès Des Produits Qui N'existent Pas Sur Des Marchés Qui N'existent Pas Encore. Dunod: Paris, France. Available online: https://www.dunod.com (accessed on 25 April 2022).

4. Tew, J.H., & Lee, K.J.X., & Lau, H.C., & Hoh, Y.C., & Woon, S.P. (2017). Linkage between the Role of Knowledge and Economic Growth: A Panel Data Analysis: Ph.D. Thesis. UTAR, Kampar, Malaysia.

5. Ustenko S., & Ostapovych T. (2020). Systema upravlinnia kiberbezpeky bankiv z vykorystanniam zasobiv shtuchnoho intelektu [Bank cybersecurity management system using artificial intelligence]. Systemnyi analyz y modelyrovanye protsessov upravlenyia – System analysis and modeling of control processes. V. Ponomarenko, T. Klebanova, L. Guryanova (Eds.). P. 209–217. Bratislava-Kharkiv. [in Ukrainian].

6. Ustenko, S., & Ostapovych, T. (2021). AI at banking services. Modeli sistemnogo analiza v upravlenii ekonomicheskimi Protsessami – Systems analysis models in the economic processes management. V. Ponomarenko, T. Klebanova, L. Guryanova (Eds.). P. 230–243. Bratislava-Kharkiv. [in English].

7. Ustenko S., & Ostapovych T. (2020). AI at Banking Infrastructure. Shtuchnyi intelekt – Artificial intelligence. #. 4. P. 7–13. Retrieved from https://doi.org/10.15407/jai2020.04.007 [in English].

8. Ustenko, S., & Hivarhizov, I. (2019). Ukrainian banks sustainable development research and management concept. Periodyk naukowy akademii polonijnej. #. 37(6). P. 35–45. Retrieved from http://pnap.ap.edu.pl/index.php/pnap/article/view/400

9. Ustenko S.V. (2018). Metodolohichni zasady modeliuvannia protsesiv rozvytku vysokotekhnolohichnykh pidpryiemstv [Methodological bases of modeling of processes of development of high-tech enterprises]. Informatsionnaya ekonomika: etapyi razvitiya, metodyi upravleniya, modeli – Information economy: stages of development, management methods, models. V.S. Ponomarenko, T.S. Klebanova, N.A. Kizima (Eds.). P.

576–586. Kharkov: VShEM-HNEU im. S. Kuznetsa. [in Ukrainian].

10. Ustenko S.V., & Vozniuk Ya.Yu. (2022). Kontseptualni zasady doslidzhennia protsesiv informatsiinoho zabezpechennia tsyfrovoi osvitnoi diial-nosti [Conceptual bases of research of processes of information support of digital educational activity]. Naukovi zapysky Natsionalnoho universytetu «Ostrozka akademiia». Seriia: Ekonomika – Scientific notes of the National University «Ostroh Academy». Economics series. No. 24(52). P. 144–148. Ostroh: Vyd-vo NaUOA. [in Ukrainian].

The article has been sent to the editors 09.12.22.
After processing 25.12.22.