International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Artificial Intelligence in Natural Language Processing

Zahraa Raji Mohi Al-zobaidy *

Department of invitation, rhetoric and thought, AlImam AlAdham college, Baghdad, Iraq.

International Journal of Science and Research Archive, 2024, 13(01), 370–377

Publication history: Received on 25 July 2024; revised on 05 September 2024; accepted on 08 September 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1638

Abstract

Natural language processing is one of basic topics used by artificial intelligence techniques. realized it is role in content analysis. In this study we introduce artificial intelligence role in content analysis and data revealing process by using natural language processing in libraries specifically that is powerful in content analysis domain. applied to one of natural language processing systems used in content analysis as a resource of information for aim investment depends on analytic description theory.

The most important result from this study : the ability of using natural language processing in libraries for getting resources. content processing and give visitors the resources by answering their explication. applied on IBM Watson knowledge studio platform to construct machine learning model and tested on Watson discovery machine that have high quality in find and returns the English resource's, while in Arabic needs more improvement.

Keywords: Artificial intelligence (AI); Natural Language Processing (NLP); Machine Learning (ML); Content analysis (CA)

1. Introduction

Artificial intelligence used in : management. business. military work. medical and etc. intelligent systems placed instead of traditional systems in libraries since 1990's that offers services depending on customer, workers and specialists knowledge in libraries to prepare quick and efficient service by enhanced it is technical service using AI. This includes special system for reference services. reading robots. virtual learning and NLP to extract knowledge based on content [2].

Last year's some universities works on construction of AI for increasing chances of enhance their services like University of Rhode Island (URI) to help students and teacher's members to recognize AI cases and development of intelligent algorithms in artificial intelligence domain and give better services in inside libraries. AI is a new technology developed with a huge application in libraries by exploring these techniques and it's applications in a proper way like : development of AI systems for helping in technical operations. references services. resources management and content analysis. natural language processing is one of the important technique realized it's big role in CA, by adding a lot of improvement on it like web indication techniques. This study aims to discovering applications of artificial intelligence in content analysis and it's applications on natural language processing in general and in libraries as special. counting on programming libraries in AI that is beneficial in CA for data resources to investigate powerful help [4].

In general all libraries want to increase the value of data resources through selecting various valuable resources for their readers by representing high quality data description and data extraction to improve research and exploring operations. huge data accumulated and technology development make it difficult for analysis creation of needs to system's for help with access to resources and makes it available for how need's. One of best AI technology is NLP and

* Corresponding author: Zahraa Raji Mohi

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

it's big role in content analysis and data release operations for discovering NLP uses to observe platform for AI libraries that is benefit in CA [1].

In this study we offers NLP system application in CA for data resources through gaining a set of aims represented as follows:

- Explain AI applications and machine learning in libraries.
- Analyzing NLP systems in general. libraries specifically.
- Studying NLP system uses in CA and their protocols to work.
- Observes platforms and libraries uses AI in CA domain.
- Applying one of NLP systems in CA operations and construct machine learning model.
- Testing machine learning constructed and search for the ability of developing it.

We have three boundaries for this study first is objective boundaries : studying AI systems and machine learning applications and describing its role in CA by applying one of NLP systems and construction of machine learning model. second language boundaries: using NLP system for CA to both English and Arabic resources, third grade boundaries :the study contain content analysis for a group scientific articles in knowledge management topic. to execute ML model we use Watson knowledge studio tools.

2. Artificial intelligent and Natural language processing relationships

AI based on tow principles are:

- **Data representation :** means the way of represent data to the computer, so it can understand and present proper solution for it, the special languages used for data representation are RDF and OWL used in description web.
- **Searching:** where the computer searches in available choices and evaluate it according to previous criteria or what the computer concluded by choosing the right solution.

AI is the newest technology used in digital libraries management and its final results to develop systems or machines that thinks or acts like human beings. this automated systems for digital libraries depends on AI technique to present services based on libraries users and employees knowledge, which is works on decreasing human intrusion in doing usual libraries tasks and saving time data specialist to interact with beneficial directly [6].

AI uses area in libraries can be determined as follows :

- Expert systems: that simulate making decisions in human brains. it can design expert systems in libraries operating like: offering. referencing and feedback services.
- NLP: it can be used in CA which leads to improvement of search and retrieve operations.
- Pattern recognition: focus on the interactive between human and applications friendly to the user and easy to use and interact.
- Robotics : deals with moving and releasing tasks that can investment in libraries to doing many activities like: retrieving and organization for resources and counting operations [9].

The role of ML, AI and NLP in content analysis is to optimize and increase the functionality of analyzing texts to benefit of NLP specifications that transfer unorganized text into data and usable vision. usually NLP algorithms depends on ML algorithms. in spite of encoding big set of rules manually NLP can depend on ML to learn those rules automatically through analyzing a set of resources like: books and articles and make statistical conclusions to design ML model, where this model is the summation of learning output from the training set data, whenever data analyzed increases the model precision increased [10].

NLP located in many specialized like computer science. mathematics, engineering. AI, robotics and etc.. It's application includes automatic translation, text natural language processing. user interface. data retrieve. speech recognition. AI, expert systems. NLP importance in helping computers to communicate with human languages as the ability of computers to read texts. hearing words and explain it and emotion measuring. since human has a lot of languages so NLP is very important in AI because it asset to solve mystery in language that adds infrastructure benefit for data inside many application as : speech recognition and text analysis. NLP is one of the used way for text mining for testing many

resources to extract new knowledge and data organization tools like dictionaries needs NLP systems to do their tasks and jobs [12,11].

3. Natural language processing systems in content analysis operations

CA contains merging many techniques are:

- AI: the ability of computer systems to execute activities that needs human intelligence includes sound recognition and making decision that is used in analyzing huge text and classify it automatically.
- ML: one of AI content focus on computer algorithms ability to effectively learning from testing to improve performance without human programming, using text analytics to determine how classify new text piece according to the preprocessed text, evaluate the used groups to classify these text pieces in specific textures.
- Deep learning : is a set of specialized secondary from AI includes the ability of computer system to analyze data and make decisions in CA, also deep learning can be used to understand the context better in unorganized comments and enforcement analysis accuracy text automatically [4].

NLP sometimes begins with morphological analysis then stemming of terms, in all queries and documents including dictionaries and grammatical to specify properties of words and recognition part of speech. sentence analysis. figure 1 describes the mechanism of NLP systems in CA [7].

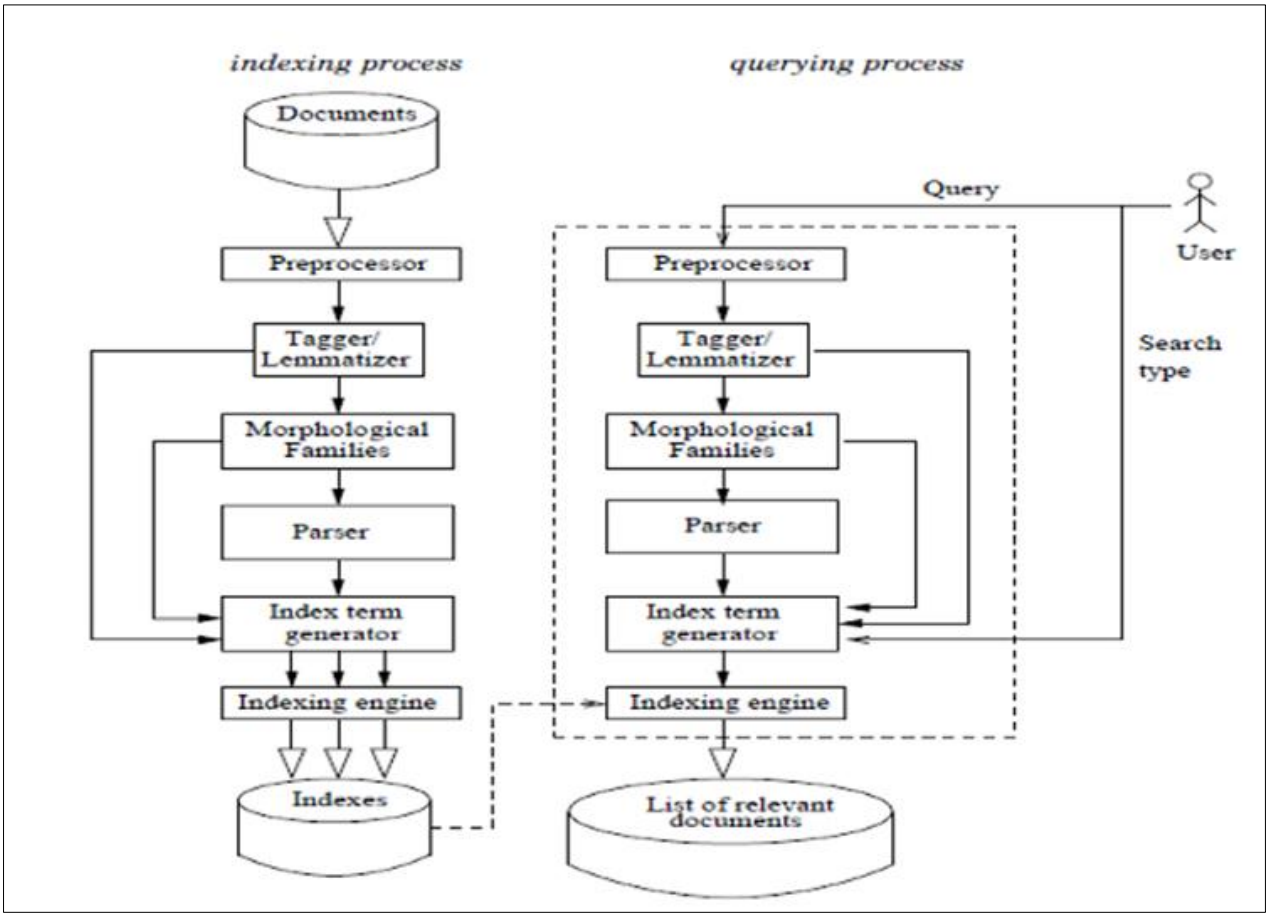

Figure 1 NLP system mechanism in CA [4]

Inputs to the NLP system is a group of documents applying a set of operations as follows

- The preprocessor : including filtering where transform text from source context like HTML or XML into plain text and removing spaces, then encoding by partitioned the sentences to words separated by set of codes, then sentence partitioned through specific limitations like dot followed by capital letter, then morphological process by analyzing the word to find its root and weight and its prefix and index.

- The tagger: using Hidden Markov Model to recognize patterns like: part of speech and hand writing. through putting signs on text to extract content words (nouns, verbs, adjectives) to reveal.
- Morphological Families : determine set of words that obtained from the same root through derivation methods, expected basic indicative relation between specific family words.
- Parser: define grammatical infrastructure of text by analyzing words formed it assisted on basic grammar rules.
- Indexing: when making a queries from the user the system search on previous steps from user queries to retrieve documents related to the search topic [2].

4. Data implementation and used application tool

Applying one of NLP systems to analyze data resources content:

4.1. Used tool in application

IBM Watson free plan is the choose that is one of the famous AI platform supports NLP to work on CA known by its ability of understanding natural language. allowing determination and extraction of master words. groups. emotion. entities and etc. depends on PYTHON language in AI used for the easiest and flexibility that it has.

Applying tow from IBM platform tools are

- IBM Watson knowledge studio: Used to find entities and relationships in the texts and construction of ML model that applied on content analysis on set of articles in Arabic and English in knowledge management topic.
- IBM Watson Discovery tool: used in search. retrieve results and CA through applying ML model constructed by IBM Watson Knowledge studio on set of other articles in the same domain to improve CA operation, search and retrieve [3].

4.2. Implementation steps

4.2.1. First stage: applying Watson Knowledge Studio Tool

Work on platform divided in two stages are: first stage: applying Watson Knowledge Studio tool by constricting workspace where work formulas tools and article CA inside it and making type system file that defines the entities and relationships in Json formula then uploaded on the system by defining 15 words in the file in addition to the available words as shown in table 1 [5].

Table 1 Entity File in Json Formula

Author	E-mail
Title	Citation
Place	Persons
Publisher	Library
Organization	Date
System	Time period
Information resource	Database
Keywords	

After that construction of dictionary to help CA human annotator to initials of CA tasks then designing simple dictionary supported by set of terminology in knowledge management domain, adding a group of articles for examining and content analysis for ML training model. articles represent domain content must added that is knowledge management to the workspace then uploading chosen articles in word formula by annotation acceptance system file in type ZIP, HTML, DOCX, DOC, PDF, TXT, CSV preferred volume not more than 2000 word in the single file.

Building ML model by uploading four articles in.docx type tow in Arabic and tow in English and performing pre annotating for all articles using constricted dictionary for easiest CA operation. creating an annotation task and

distribution work on existing files separately then dividing each article file to task menu for analyzing as shown in figure 2.

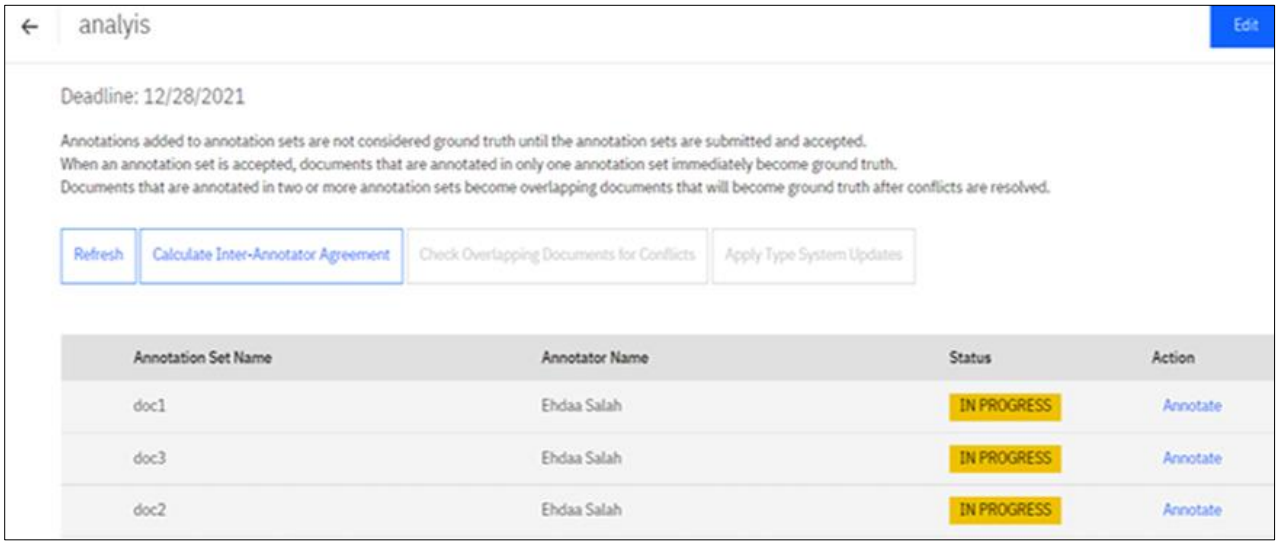

Figure 2 Distribution of content analysis task and articles

Annotator works by choosing first article and opened from the system noticed that words from dictionary are shades as shown in figure 3 below. its task easier the human work in CA and discover entities and relationships. CA and main entities determinate like books name, location. important dates, citation and keywords.

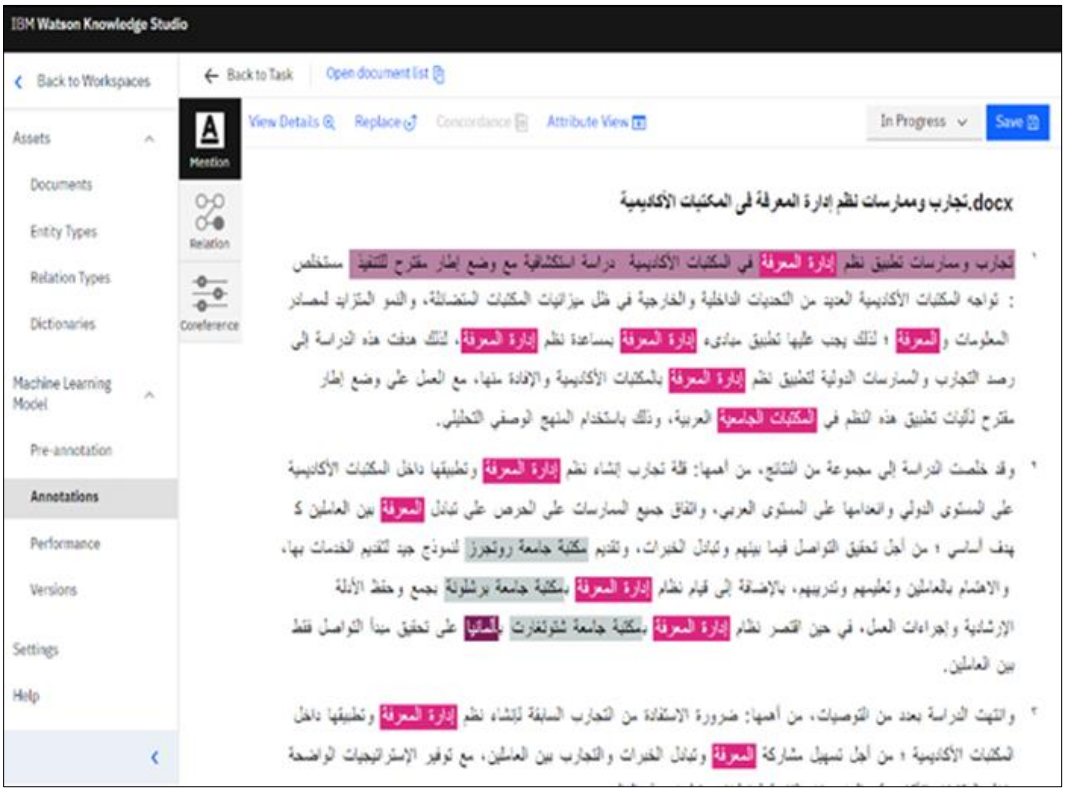

Figure 3 CA task and determine main entities

then trying to find entities relationships like : university libraries. its location. keywords and text related and multiple relationship available inside the system that described in table 2 below. Also we can add more relationships if needed.

Table 2 Relationship types between entities in the Watson platform

Relative	partOf	partOfmany	ownerOf	memberOf	locatedAt
EducatedAT	capitalOf	Before	basedIn	autherOf	AgentOf
PlaysRoleOf	patricipantIn	spokespersonFor	Overlaps	affiliatedWith	AffectedBy

After finishing CA operations to available articles design train and evaluate model system applied on a set of available data and evaluate it, when the result appears then move to the next stage using ML model that constructed for CA of articles from the same topic and extract entities and its relationship.

4.2.2. Second stage: using ML model in CA operations

At this stage using ML model constructed before to rise entities and relationships extracting operations through investigate connections and integrity between the Watson Knowledge Studio machine learning and Watson Discovery model through the following steps :

- Loading files for set of articles in knowledge discovery to Watson discovery tool, 4 article 2 in Arabic and 2 in English publish ML model constructed by knowledge studio.
- Additional improvement added to the ML model keyword extraction, entity extraction and relation extraction.

By analyzing the articles content knowing the entities that has a result to predetermined in knowledge studio service. for the keyword entity the imported keywords in the text and its relation to the text and reputation. testing search result on set of new documents depending on ML model improvement and making a query on the famous import entity in the article in the following formula :

Nested(enriched_text_entities).term(enriched_text.type.count:5).term(enriched_text.entities.text)

The result is recognition of entities determined before in the ML model as shown in figure 4 below.

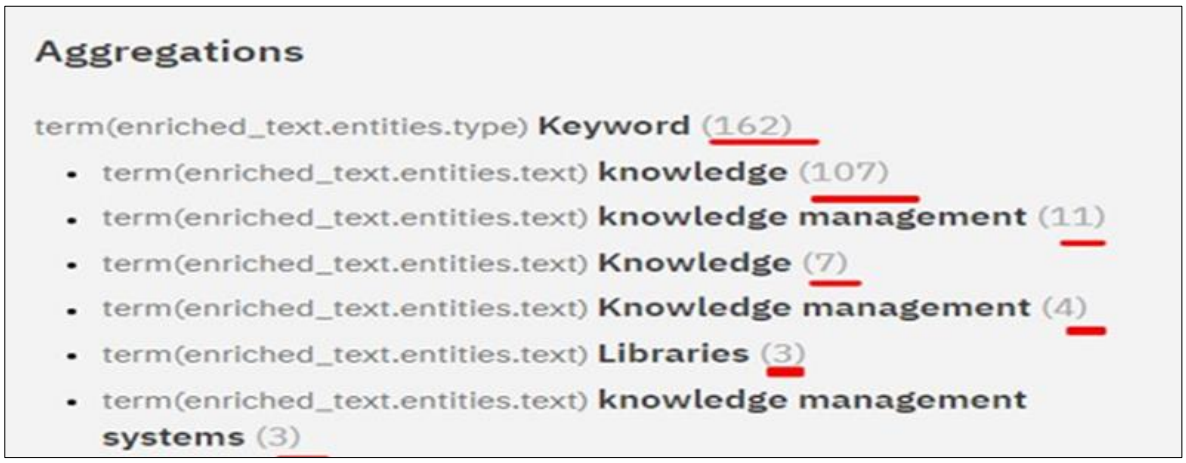

Figure 4 Entities predefined in ML model Recognition

4.3. Application Results of IBM platform:

After applying IBM Watson Knowledge Studio platform to construct ML model and apply for Watson Discovery tool its efficiency was very high in find and retrieve for both Arabic and English resources but in spite of IBM Watson Discovery supporting Arabic language but don't succeed in entity and relation between them, the determined groups in the first stage using knowledge discovery tool, but it success in finding the whole text when enquiring and searching for any word in the analyzed documents the result extracted. so the tool needs more improvement for the Arabic resources. The ML model can be improved throughout adding more studies in the specific domain knowledge discovery and the dictionary with more new words in the same topic[1,11].

5. Conclusion and future work

We can conclude a set of results grouped as below:

- The ability of using AI techniques in the libraries in many works like expert system build, resources service. NLP, pattern recognition and robotics
- NLP used in speech recognition, text classification, information extraction, text analysis. automatic translation. automated briefing text and analyzing social network
- NLP used for extracting knowledge for automatic revealing and CA as an important field in NLP research domain. that released its role earlier in CA operation with entering a lot of improvement like web indication techniques.
- NLP can be used in libraries to find resources. content processing. answering beneficiaries queries and improving processing using knowledge organize tools.
- A lot of NLP available tools on web that differ from easily to use these tools allows CA and entities and relation recognition. must of them depends on Python language since its trendy in AI, others may depends on Java and other programming languages.
- Applying IBM Watson knowledge studios to construct ML model using Watson discovery tool was efficient in find and retrieve for English resources. but for Arabic resources the tool needs more improvement.

For the future work the following recommendation are recommended :

- Working on applying AI and ML technology to introduce new services and improving jobs in libraries to investigate benefits for libraries and visitors.
- Determine specific budget from libraries for AI technique and training information specialist to manage.
- Making more interest on the studies used NLP and CA.
- Developing ML model using platforms and programming libraries available for CA and working to improve it.
- Give the priority to develop and improve NLP systems in Arabic.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Orobor I. A.. Integration and Analysis of Unstructured Data for Decision Making: Text Analytics Approach, International Journal of Open Information Technologies, 2016, vol. 4, no. 10.

[2] Hamandi A. M., wahab H. B., & Karim A.. Natural Language Processing Using Natural Language Toolkit, Iraqi Journal of Information Technology, 2016. vol. 7, No. 2.

[3] Khalid M.. Human artificial intelligence and support for learning analytics, Egyptian assembly of science and technology. 2022. vol. 10, pp:45-79.

[4] Ehdaa saleh nejy, Applications of Artificial Intelligence Systems in Content Analysis and Indexing Processes: An Applied Study of Natural Language Processing systems, journal of scientific documents for libraries and information. 2022. pp:89-216.

[5] Amjad Z..Muzzamil A.. Ioana P. & etc.. Review Artificial Intelligence-Based Medical Data Mining, journal of personalized medicine. 2022, pp:1-23.

[6] Akbar Javadi, Mohammad Rezania, Applications of artificial intelligence and data mining techniques in soil modeling, Geomechanics and Engineering, 2009. vol. 1, no. 1 pp: 53-74.

[7] Shivangi G. A..Sai S. and Ritu P.. Cyber Security Threat Intelligence using Data Mining Techniques and Artificial Intelligence. International Journal of Recent Technology and Engineering. 2019. vol.8.

[8] JiaoLong Li. E-Commerce Fraud Detection Model by Computer Artificial Intelligence Data Mining. Hindawi Computational Intelligence and Neuroscience. 2022.

[9] yassamen A.. osama A. & Layla S., Application of artificial intelligence in children libraries: scientific review, international Arabic journal of knowledge management. 2023. vol. 2, no. 3, pp:49-96.

[10] Simon Fauvel, Han Yu, A Survey on Artificial Intelligence and Data Mining for MOOCs. 2015, pp:1-46,

[11] Basallo, Y. A., Senti, V. E., & Sanchez, N. M.. Artificial intelligence techniques for information security risk assessment. IEEE Latin America Transactions, 2018. vol.3. pp: 897-901.

[12] Noh, Y., & Ji, Y. R. (2021).. A study on the service provision direction of the national library for children and young adults in the 5G era. International Journal of Knowledge Content Development & Technology, 2021, vol.2, pp: 77-105.