Journal of Positive School Psychology
2022, Vol. 6, No. 5, 8765–8782

Artificial intelligence as a support for librarians who buy scientific documents: what factor do they take into account?

Mokeddem Allal

University of Algiers 3, Department of Management (Algeria)
Email: allalmokeddem@gmail.com

Abstract:

With the arrival of advanced technologies, the search for good documentation has become a very sophisticated practice in terms of tools and methods. In this context, a librarian seeks to improve his process of investigation and evaluation of pre-existing studies before starting a purchase operation. For this reason, three factors were discussed relying on artificial intelligence as an integral practice: how to index a document in order to identify the most important information, how to summarize a paper or a group of papers by integrating the sentences most representative of each other and finally, how to classify the citations in order to clearly identify the most popular studies over the others.

Keywords: Artificial intelligence (AI); librarian buyers, indexing and tagging; content summary; citations.

1. INTRODUCTION

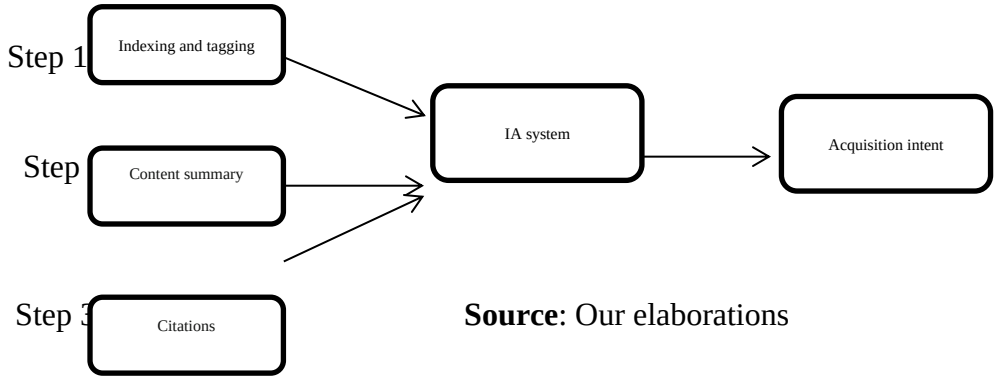
Nowadays, technology is no longer considered as an IT tool, but beyond that is considered as an integral practice. According (Cave, 2019, p. 4), some studies observe that advanced technologies such as artificial intelligence (AI) will become a cornerstone of the global economy here with a global investment rate of $150 trillion by 2025. In this regard, this new intention towards technology opens horizons for users to exploit the different work environments presented by the different characteristics of specific tasks (Spies and al, 2020, pp. 397-408). At this level, the technology can also identify the scientific achievements made by researchers via indicators, which can thus adapt the source of information to the benefit of the user.

Faced with this new trend, librarianship has become a discipline framed by approaches and models that manages all scientific and academic resources. In this context, the adoption of AI can improve library services and provide access to accurate information (Yusuf and al, 2022, p. 4). At this stage, having access to rich and relevant information is the primary mission of a librarian buyer. A set of practices can be used to develop programs for effective reference services, appropriate digitization of scholarly literature, and identification of appropriate subject categories. These operations are based on a strong exploitation of artificial technology which allow a better analysis of the metadata linked to the new scientific collection; its level of indexing, its reference services, its capacity to search and find information and finally to know how to ensure a very precise delivery to the whole of the community in particular the authors classified at the first rank.

In this article, we seek to discover how AI can help librarian buyers to conduct a scholarly resource buying operation by studying past research done by faculty.

2. RESEARCH MODEL

Fig N^0(01): Conceptual model



Source: Our elaborations

3. THEORETICAL BACKGROUND

The popular question of how libraries can become a scientific and academic support institution for universities and their scientific bodies. The answer to this question is how we envision the profound changes that AI can bring as a powerful integrated tool for robust documentation, making scientific resources discoverable, researchable, and peer-analyzed (Massis, 2018, p. 5).

In order to increase the achievement of this primary objective of a library, the activity of finding good documentation has become a very specific task. In this respect, according to (Salau and Oluwade, 2017, p. 5) with the growth of information and communication technologies (ICT), many other avenues for the flow of information are opening up. Knowing how to find the path of acquisition of generally electronic scientific resources is a major concern for librarian buyers. At this stage, a librarian must think about the criteria that guide the process of acquiring good documentation while ensuring the best quality dissemination of information to the public. The following table illustrates the main factors and AI methods used to approve a scientific reference before the start of a purchase operation.

Table N^0(01): The factors and AI methods used to approve a scientific reference

Main analysis factors	Objectifs	IA methods	References
Indexing and tagging	Identify the most relevant documents based on AI tools that improve consistency and quality, as concepts will be identified and corresponding keywords assigned.	Keyword augmentation and screening (NKAS); Natural language processing (NLP); Geoparsing; TimeBank corpus. TimeML; ISO-TimeML; THYM; RelEx algorithm	(Ma, J and al., 2022, p. 5); (Saart, and Tuominen., 2020, p. 6); (Ng, and al , 2020, pp. 186-191); (Mogali, 2014, p. 198); (Bischoff and al., 2008, pp. 193–202); (Voorbij, 2012, p. 12); (Furner, 2007, pp. 47-51)
	The ultimate goal is	DQN (Deep Q-	(Widyassari and al.,

Content summary	to present one or more text documents while retaining the main information content using a machine.	Network) ; systematic literature review (SLR); ROUGE-L uses Longest Common Subsequence (LCS).	2019, pp. 491-496); (Rajasekaran, and Varalakshmi., 2018, pp. 456-460); (Day and Chen., 2018, pp. 478-484); (Lin, 2004, pp. 74-81)
Citations	Citation is a rich source of information for researchers and can be used to create exciting new ways to browse data.	Stanford CoreNLP	(Iqbal and al., 2021, pp. 6551-6599); (Manning and al., 2014, pp. 55-60); (Di Iorio and al., 2013, pp. 63-74)

Source: Our elaborations

3.1 Indexing and tagging

Librarian buyers seek to increase the efficiency and effectiveness of access to rich and relevant information for the research community. In order to achieve this goal, an intelligent system for retrieving and using information resources is highly recommended. According to (Cao and al, 2018, pp. 811-825) scientists and professionals have created systems that can be thought and decided instead of the librarian. In this regard, a librarian buyer performs a routine and repeated operation to reach the desired source of information. For this reason, according to (Gerolimos, 2013, pp. 36-58) indexing documents by keywords helps librarians and users during the documentary research phase. This operation consists of identifying concepts, translating these concepts into verbal descriptions and selecting and assigning controlled vocabulary terms that are conceptually equivalent to verbal descriptions.

According to (Ma and al, 2022, p. 5) the process of word indexing involves one in three distinct phases. The first phase emphasizes generic word searches and usually starts with two or three user-defined keywords to reach a large list of references. In the second phase, the search operation is more rigorous than the first. In this regard, correlated words, synonyms, close words, closest words, related words will also be considered to access the second level of indexing. This indexation taxonomy is based on a database collected on the basis of a large number of research articles with open access information sources. To achieve this level of involvement, a source code based on a natural language (NLP) introduces the basics of research by a librarian. According to the NKAS approach based on the technical NLP, two operations were followed:

1. Abstracts were limited to sentences; even the title was considered a single sentence.
2. The abbreviations were recorded, at the same time; the original words related to the field have been codified in the database.

The last phase called screening phase consists of two distinct steps. The first is based on a general elimination according to a set of criteria are the impact factor of the journal, the type of article, the research methods (simulated study or experimental study) and the structure of the article (such as the presence or lack of a requested abstract). The second is based on a specific elimination dedicated to a particular subject. In this regard, the indexing system can be based on scores that value the

relevance and non-relevance of one article over another. This classification is done in a systematic way where each word presents a score either positive or negative. At the end of the operation, the relevant articles come to light following preconditions established by the library community.

In addition, the study conducted by (Ng and al, 2020, pp. 186-191) showed how to identify the places from which the events related to the field of research come from how they will be reported in order to help locate the desired studies. Geoparsing is a combined process of geolocation and geocoding based on geographical coordinates such as cities and the country where this resource comes from (GeoNames, 2020). Then, the geoparsing approach consists in characterizing the toponyms by a set of characteristics such as the name of the toponym, the position of the first and the last character in the text, the length of the characters and so on. From there, all the information retained will then be processed by calculation methods linked to each toponym stored in a location database. This is done in order to assign the corresponding coordinates to each spatial classification (Santos and al, 2015, pp. 375-392) and therefore know how to link each resource with the nearest neighboring toponym (DeLozier and al, 2015, p. 451)

The chronological determination of the event is necessary to fix the chronological order and the coherence between the events. It is very important to make the distinction between an article reporting a recent event and an earlier article known to the scientific community. In this context, the best-known temporal identifiers are not limited to the date of publication of a scientific article, and the date of first reception, but beyond, two analysis operations can be exploited:

1. First, temporal relation mining focuses on classifying temporal relations between extracted events and temporal expressions. We quote a passage from the article defining the time factor as the first introduction of AI which dates back to the 1960s, the global death rate from the pandemic has reached 6,315,932 deaths since March 2020 (Worldometer, 2020).
2. Second, temporal reasoning focuses on the chronological order of events by inference. For this reason, several temporal extraction systems have been delivered, notably TimeML (developed for the temporal extraction of news articles in finance); ISO-TimeML (a revised version of TimeML); and THYM (developed for temporal extraction in patient records).

In addition, counted the number of cases in a scientific paper and strongly requested by researchers. Everyone is looking for statistics on a situation and an event, whether economic, cultural, environmental and even health. Based on this information, the artificial intelligence techniques that are used by our librarian to determine the number of cases of an event such as the number of cases of foodborne diseases, the number of innovative entrepreneurial projects, the number of patents already deployed by companies... in all of its information will help librarians track and predict how an event will unfold and the importance of acquiring its resources. As an example the RelEx algorithm, an algorithm aims to identify sentences in news articles that report the number of cases of foodborne illnesses (Nasheri and al , 2019).

3.2 Content summary

With the increasing number of publications covering the entire field of research, the task of evaluating pre-existing work is quite a complicated task. In this regard, the arrival of advanced technologies serves to promote the analysis of scientific documents in an automatic way so that information arrives more quickly and more precisely without

losing the original meaning of the original documents.

To successfully carry out a documentary investigation task, the basic principles of documentary analysis desired by a librarian must respect the following constraints:

a- Number of targeted documents such as: single document and multi-document.
b- Opted analysis techniques such as: extractive and abstract.
c- Desired classification such as: supervised and unsupervised.
e- Its use such as: informative, indicative and critical summaries.

The number of targeted documents is a major factor in the process of investigating scientific resources published by professors. For this purpose, we either focus on a single paper from a reputable researcher or publisher while building on the key concepts of the article (Sarkar, 2013, pp. 602-620). Therefore, the multiple documents are used to determine a detailed overview of the research area either based on the publication period, or the same subject from one document, or information comes from several sources (Widjanarko and al, 2018, pp. 520-524).

The second most important factor in generating a report on the work performed is the type of extraction technique used. In this regard, the first is to extract keywords and paragraphs to generate summaries. The generated summary is entirely composed of extracted content (Khan and Salim, 2014). The second technique based on abstract reasoning is the one that generates summaries by creating new texts or using words that are not in the original text. More explicitly, (Rajasekaran and Varalakshmi, 2018, pp. 456-460) see that the extractive technique seeks to select the main important sentences of the document using different statistical methods. On the other hand, the abstractive technique creates a semantic representation of the input text to generate a summary.

The third factor is presented by the desired type of classification based on supervised and unsupervised learning. The first aims to the synthesis of the text by identifying events related to a research area such as “Covid 19” and its impact on society. As a result, a graphical classification will be deployed based on the set annotations. We cite AdaBoost as a very sophisticated algorithm using to summarize text in Arabic (Belkebir and Guessoum, 2015, pp. 227-236). On the other hand, unsupervised learning techniques do not rely on clear instructions and guidelines (Yousefi-Azar and Hamey, 2017, pp. 93-105). On the contrary, data mining techniques are in great demand in this field of investigation using statistical methods such as K-Means, clustering machine learning, etc. Therefore, this classification can help to group similar sentences and thus have the possibility to choose the most appropriate sentence from each group used in order to compile the summary.

The last factor presents the three types of text summary provided following a compilation of the various parameters presented above namely; informative summaries, indicative summaries, and critical summaries. According to (Wibisono and Hendratmo, 2007, pp. 454-470), informative summaries provide details of the main information or a summary of the text in a few lines of summaries, indicative summaries these summaries only provide an indication of the text and contain only partial information about the text. Finally, critical summaries are also called evaluative summaries because they capture the author's summary according to a given topic.

Once the text is pre-selected and the analysis parameters are set, the technology is used to produce summaries according to the AI techniques used. The technique proposed by (Lin, 2004, pp. 74-81) based on graphs and machine learning to evaluate each sentence, then chooses the

best set of sentences to generate summaries. ROUGE-L uses Longest Common Subsequence (LCS) technique is one of the artificial intelligence techniques to calculate the similarity score between the selected abstract and the reference abstracts. In this respect, this technique offers the possibility to generate word pairs in sentence order, allowing calculates the deviations of the percentage of matched pairs between the candidate's summary and the reference summaries. Therefore, this AI technique does not assign any score to a sentence if the selected sentence does not have a pair of similar words with that of the references.

In this respect, the extraction of a synthesis grouping together a set of studies is in fact based on a set of statistical methods. For this reason, the most commonly used methods for determining the importance of sentences in literature are listed in the following table:

Table N^0(02): Statistical methods that support the development of summaries

Statistical method	Objective
Term Frequency-Inverse Document Frequency(TF-IDF)	Refers to the number of times a term (T) occurs in the provided input document (D), while the inverse document frequency refers to the number of times a word occurs in the given text corpus at through which it measures the salience of a word in the document.
Cue Phrases	This method assigns a weight to sentences based on the presence of certain pragmatic words (classified as positive or negative) such as : “develop”, “meaningful”, “goal”, “hardly”, “goal”, “impossible ". ', 'believe' etc. These words present the extent of the sentence in the text.
Location	Weights are assigned to sentences based on where they appear at the beginning or end of the document, such as the conclusion or summary, or in the first and last sentences of a paragraph.
Sentence centrality (Si)	Sentence Centrality: The centrality of a sentence is derived based on which words overlap or appear more frequently in the given sentence (If) in a document with the other sentences in a document (Others).
Topic word	Terms that appear frequently in a document may be more likely to be related to that topic and these topic words in a sentence contribute more to the calculation of sentence scores as follows: Sentence Score (S) = [Total Number of Topic Words in Sentence(s) / (Total Length of Sentence(s))&
	Graph theories really revolve around the

Linguistic methods	relationship that unites the sentences illustrated using graphs where the sentences in the document are represented in the form of nodes and the edges represent the connection between these sentences. Grouping methods: This method groups the similar sentences or paragraphs into different groups to detect a common theme or sub-topic between them, and then selects the textual unit (a representative sentence) from these groups one by one to summarize them. The Singular Value Decomposition (SVD) method is applied to capture the most recurrent and relevant word combination pattern from the input document and represents it as singular vectors with the sentences containing this pattern. Fuzzy logic: This approach uses a fuzzy analyzer to calculate the rank of each sentence in the input source text from its statistical parameters. The relationship between statistical parameters is described using fuzzy rules (if-then rules). High ranking sentences are chosen for the final summary
Others	Other characteristics such as sentence length, pronouns in sentences, named entities (proper nouns), numeric data, fonts (bold, italics, underlined words) and biased words (domain-specific words) , etc. are also considered more important in calculating sentence scores.

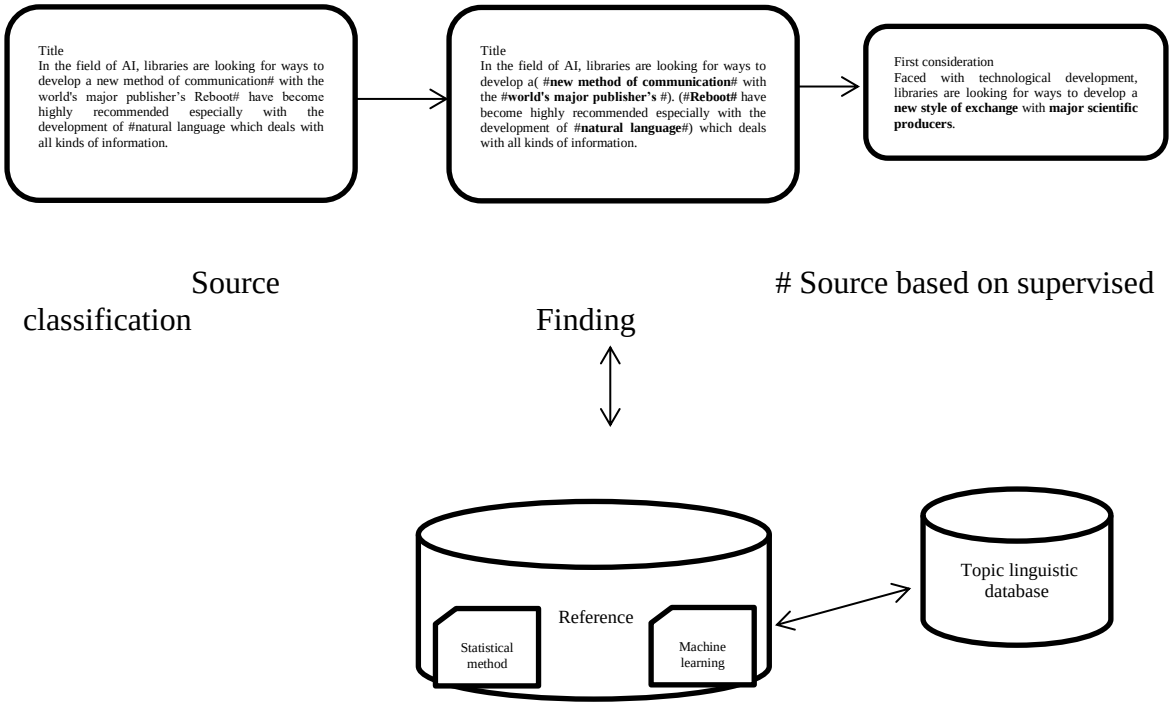
Source: (Rajasekaran and Varalakshmi, 2018, pp. 456-460)

We can illustrate this demonstration via an example (see Fig1) that emulates how a linguistic sentence can be transformed into a summary all based on the technique of abstraction with a statistical method that seeks to locate topic words by pairs.

The formula is as follows:

Sentence Score (S) = [Total Number of Topic Words in Sentence(s) / (Total Length of Sentence(s))]

Fig N^0(02): Summary production based on ROUGE technique



Source: Our elaborations

Sentence Score (1) = [2 / (128)] = 0, 0156
Sentence Score (2) = [2 / (132)] = 0, 0151

In the first sentence two domain-related words were picked up by the machine as a result of the linguistic database: #new method of communication# and #world's major publisher's. The detection of the words was made following the process of indexation made in order to produce a reference list adapted to each domain. Once the sentence scores are calculated, the sentences are ranked in descending order of their scores, starting with the highest score at the top to the lowest at the bottom. Sentences with the highest scores are selected first to generate the summary.

The operation of producing a summary of a scientific article is carried out in a cluster, bringing together all the characteristics of extracting a homogeneous and significant summary. The characteristics of the most frequently used extraction methods are:

1. Know how to identify words related to the search field and therefore create associations between synonyms, close words, closest words, related words all based on an indexing taxonomy based on a neuro-informatics approach to know the DQN (Deep Q-Network) as a technique that suggests potential actions from the sentence located in the selected document (Yao and al, 2018, pp. 52-62).
2. Suggested potential actions presented by new concepts will be used to paraphrase the original text without losing the original meaning of the sentence.

3.3 Citations

The task of a librarian is not limited to ensuring access to information, but

seeks to evaluate the most relevant information in order to carry out a purchase operation. Relevant information is generally the publications that influence the research in order to explain the framework in which the research took place. In this context, citation analysis is one of the most frequently used methods in research evaluation (Iqbal and al, 2021, pp. 6551-6599). In this regard, whenever a researcher writes an article, he seeks the degree of endorsement of this research through bibliographic references such as pointers to related works, sources of experimental data, background information, standards and methods related to the solution under discussion, etc. The citation analysis is subdivided at several levels presented in the following table:

Table N^0(03): The different dimensions of the citation analysis

Analysis Dimension	Definition	Objectifs	References
Context	When a reference is mentioned in the body of the citing article, the text that appears next to the mention is called the context of the citation.	Context is established by the location of the citation in the citing text, the surrounding words, and its semantic context, each of which has implications for all parties involved in the research.	(Zhu and al,, 2015, pp. 408-427); (Bornmann and al., 2018, pp. 427-437); (Bertin and Atanassova, I. , 2018 , pp. 127-138)
Classifications	lexical, structural, arithmetical, and sentimental classifications	Determine the lexical field of a citation; Determine the distribution rate of a citation; Determine the frequency rate of a citation; Determine the sentimental polarity of a citation	(Abu-Jbara and al, 2013, pp. 596-606); (Iqbal and al, 2021, pp. 6551-6599); (Di Iorio and al, 2013, pp. 63-74); (Small, 2018, pp. 461-480); (Yang, and al. , 2018, pp. 52205-52217); (Agarwal and al., 2010, p. 11); (Wang, and al, 2012, p. 21) (Mifrah and al., 2018, pp. 145-156)

Source: Our elaborations

3.3.1 Context of citation

According (Hu and al, 2015, p. 21) the contextual analysis of a citation is done according to two parameters: the context of the citation and the location of the citation. In this regard, according to (Abu-Jbara and al, 2013, pp. 596-606), a citation context is identified based on a set of machine-supervised sequence labeling. The calculation is performed according to six new characteristics by applying the technique:

#Demonstrative determiners: the current sentence contains a demonstrative determiner (this, that, those, etc.).
#Conjunctive adverbs: the current sentence begins with a conjunctive adverb (however, accordingly, moreover, etc.).
#Position: position of the current sentence relative to the citation.
#Contains the closest noun phrase: the current phrase contains the closest noun phrase (method, corpus or tool).
#Contains the mention of the target reference: the sentence contains a mention (explicit or anaphoric) of the target reference.
#Multiple references: the quoting sentence contains multiple references.

Similarly, location of citation is determined based on the recurrent citation distribution using the IMRaD structure. According to studies conducted by (Hu and al, 2015, p. 21), the distribution of citations between sections was based on 350 publications in XML format. Thus, the distribution and density of citations are as follows: (41.8%) in the introduction, (25.2%) in the methods, (25.9%) in the results and (7%) in the discussion. We find that the introductory part has more citations than the other parts. This is due to the fact that the subject is being treated as a topical issue. On the other hand, if the study results suggest that if a publication has more citations in the methodology section, the focus of the article is on the methodology. If an article has an equal distribution in terms of citation between sections, there is a strong possibility of falling into a reflection that tends towards criticism.

The study conducted by (Bertin and Atanassova, 2018 , pp. 127-138) browsing full-text articles seeks to identify the sources of information grouping a set of multiple in-text references (MIR) and their locations. A considerable number of databases based on the MIR-PNL technique, approximately 80,000 research articles published by the Public Library of Science in 7 journals were analyzed. In this respect, two characteristics were considered by the analysis, namely: the position of the reference in the IMRaD structure for scientific articles in the experimental sciences and the number of in-text references that make up an MIR in the different journals.

In order to identify sentences containing MIR, the following steps were taken using the Stanford CoreNLP (Manning and al, 2014, pp. 55-60), as a basic NLP tool:
1- Segment all paragraphs into sentences;
2- Identify all references in the text;
3- Consider the number of in-text references in each sentence.

We focus on the part-of-speech (POS) tagger, the named entity recognition (NER) system and the coreference resolution system to show the relevant information located in the MIR contexts. This function in conjunction with the annotators provided with StanfordCoreNLP which can work with any character encoding, using the JAVA-based Uni-code language or support for other languages (French, German and Arabic) is strongly assured. The following table presents which annotator with their natural meanings:

Table N^0(04): Article splitting markers

Annotator dimension	Role
Tokenize	Tokenizes the text into a sequence of tokens. The tokenizer saves the character offsets of each token in the input text.
Cleanxml	Removes most or all XML tags from the document.

Split	Splits a sequence of tokens into sentences.
Truecase	Determines the likely true case of tokens in text (that is, their likely case in well-edited text), where this information was lost, e.g., for all upper case text
Pos	Labels tokens with their part-of-speech (POS) tag, using a maximum entropy POS tagger
Lemma	Generates the lemmas (base forms) for all tokens in the annotation.
Gender	Adds likely gender information to names
Ner	Recognizes named (PERSON, LOCATION, ORGANIZATION, MISC) and numerical (MONEY, NUMBER, DATE, TIME, DURATION, SET) entities.
Regexner	A default list of regular expressions that we distribute in the models file recognizes ideologies (IDEOLOGY), nationalities (NATIONALITY), religions (RELIGION), and titles (TITLE)
Xref	Tagged all references in the text that were identified by the previous processing with tags

Source: (Manning and al, 2014, pp. 55-60)

A sequence of tags embedded in the body. This allowed us to examine the most frequent types of contexts from the point of view of their syntactic structures. The results indicated that: a) MIR appears frequently in all sections (approximately 41% of sentences with citations); b) MIRs appear fairly often in the introduction, discussion and results sections (around 20% of sentences), and less often in the methods section (only 15% of sentences); c) MIR are mostly found near verbs in a sentence.

3.3.2 Classification of citation

According to the studies of (Abu-Jbara and al, 2013, pp. 596-606), revealed the basic characteristics of better organization of citations. This will help librarians find the most relevant citations in order to launch the buying process. In this respect, lexical and structural, arithmetical, sentimental characteristics are essential for the classification of a citation. All the studies agreed that the classification of a citation is made according to lexical characteristics. The authors observed that features such as proper nouns as well as preceding and following phrases are helpful in classification (Agarwal and al, 2010, p. 11). In addition, (Small, 2018, pp. 461-480) study revealed the importance of the linguistic dimension in order to classify references according to the rate of lexical coverage of words. The model deployed based on the frequency of cover words such as "May", "Show", "Suggest" and "Using" that was used as pointers in the citation classification process. (Small, 2018, pp. 461-480) concluded that the predictive ability of the word “using” in the classification of citation contexts in the methodology section and other sections was superior to that of other cover words, with an accuracy of 89.5%.

The structural approach of a classification is highly recommended during a resource acquisition process. According to (Wang and al, 2012, p. 21)

proposed a classification approach comprising four structural categories describing the structure of an article: extend, criticize, improve and compare. In this framework, out of a set of 40 selected quotes, the results showed that more than 50% belong to the extension class, the rest are distributed as follows: "criticize": (30.14%) , " compare": (13.88%) and "improve": (3.83%).

The arithmetic dimension is also considered an important factor in the ranking of a journal. Over the past few years the number of citations has been seen as an important factor in evaluating the performance of faculty members, research institutes and universities. Journals can be ranked according to their impact factor as an indicator of the performance of one journal relative to another in the same field of research. In this context, according to (Yang and al, 2018, pp. 52205-52217) indicated that the weight of a citation is affected by the prestige of the citing articles. In this regard, library buyers seek to select the most cited journals in terms of H- INDEX. From there, the parameters for calculating citation weights are shown as follows:

Table N^0(05): The formulas for calculating the weight of citations

Dimension	Mathematical formula
Calculate citation weights taking into account the time factor	$C_P^{(W)} = \frac{1}{N_P}\sum_{i=1}^{n}\frac{\log(C_i+1)}{\Delta t}$ $C_P^{(W)}$ Is the citation weight of the publication P. Δt Is the time span of a paper from the published year to 2022. N_P Is the number of papers in publication P. C_i Is the citation count of paper i.
Calculate the citation weights of every journal from yyyy to 2022	$C_{P_y}^{(W)} = \sum_{i=1}^{n}\frac{\log(C_i+1)}{(\quad - y_i)N_{cj}^{(P_y)}}$ $C_P^{(W)}$ Is the citation weight of the publication P. Yi Is the published year of paper i. $N_{cj}^{(P_y)}$ The number of journals or conferences in the category to which the publication P belongs in year y.
The weighted cited credits of each journal	$W_P^{(C)} = \frac{1}{N_P}\sum_{i=1}^{N_P}\log(C_w^{(P_y)}C_i+1)$ $C_P^{(W)}$ Is the citation weight of the publication P. N_P Is the number of papers in publication P. $C_w^{(P_y)}$ Is the citation weight of publication P

	in year y
Calculate the mean citation weight of various categories of journals	$C_c^{(W)} = \sum_{i=1}^{n} \frac{\log(C_i + 1)}{(\quad - y_i) N_{cj}^{(P_y)}}$
Calculate the popularity score of the journals	$S_P = \frac{1}{N_P \sum_{y=} \quad N_y^{(P)}} \sum_{i=1}^{n} \log(C_i + 1)$ S_P is the popularity score of publication P.

Source: (Yang and al, 2018, pp. 52205-52217)

The sentimental dimension seeks to determine the different poles of a field of research. A quote can be categorized into a taxonomy that measures the level of sentimental contribution of a study. A citation is marked as positive if it highlights the strength of a cited article, negative if it highlights the weakness of a cited article, and neutral if the citation does not give any judgment towards the context studied. According to the study conducted by (Mifrah and al, 2018, pp. 145-156), sentiment analysis approaches are classified into two categories: Lexicon-based approaches and Corpus-based approaches.

The first relies on an external lexical field and dictionaries to extract the polarity of feelings. These lexical concepts are stored in a repository such as SentiWordNet, aims to catalog, classify and relate in various ways the semantic and lexical content of the English language. In this respect, data processing is based on semi-supervised quantitative and qualitative learning techniques. SentiStrength uses a dictionary of sentiment words with strong articulation between terminologies. Each terminology offers a higher or lower contextual coverage rate than the others. Other researchers like (Taboada and al, 2011, pp. 267-307) use an approach called The Semantic Orientation CALculator (SO-CAL) based on a dictionary of words annotated with their semantic orientation (polarity and strength). From there, the machine can locate the desired quote and therefore put it in a place as close as that of its neighbor. In the end, the set of citations closest in terms of score can promote the development of a very distinctive managerial current in terms of thought and practice. On the other hand, the second approach acquires the information necessary to define the feeling from a large corpus based on a repository of knowledge that covers all the phases that formulate a field of research.

Finding the best type of citations sought is a very delicate operation, because it requires regularly adjusting the advanced search parameter. With technological advances, this operation has become accessible to all actors involved in research. Based on the study by (Di Iorio and al, 2013, pp. 63-74) proposed an algorithm, called Citation Typing Ontology (CiTalO), to automatically deduce the function of citations using semantic web technologies and NLP techniques. This solution offers sophisticated citation network viewers and powerful interfaces for users to filter, search and aggregate data.

This tool seeks to classify citation in the form of a decision tree, taking into account the sentimental dimension of a citation. This is to identify whether a particular act of citation was done with positive intentions (eg, praising previous work on a certain subject) or negative (eg, criticizing the results obtained by a particular method). But how CiTalO works?

The first phase consists of deriving a logical representation of the sentence

containing the citation. According to (Presutti and al, 2012, pp. 114-129), Ontology Extraction is based on Representation Theory and Linguistic Frameworks (FRED) as a framework that serves to present discourses, thus transforming sentences into a logical form that facilitates recognition of scientific heuristics and therefore detect possible types of the citation.

After transforming each into a logical form following the FRED approach, the next step is to extract the different types of authors from a citation or co-citation. Author types have been grouped as follows:

a- Extend references: Serves to extend by dealing with interdisciplinary problems.
b- Outline references: Used to describe the main guidelines of a research area.
c- EarlyWork: proactive studies that seek to explore context.
d- Work and Research: Classical research that seeks to address a well-defined problem.

The final stage is the alignment dimension on CiTalO. In this step the machine will assign CiTalO types to the citations. For this, two ontologies have been developed using Web Ontology Language (OWL): CiTOFunctions and CiTO2Wordnet. The first is used by researchers to determine the rhetorical and sentimental function of a citation (positive/neutral/negative). The second defines citations with the appropriate Wordnet synsets as expressed in OntoWordNe. Accordingly, this classification will not only help researchers to determine the appropriate source of information, but also to determine the polarity of sentiments emerging from the text in which the citation is included.

4. CONCLUSION:

This article has been presented in order to make library purchasers about the process of evaluating research studies performed in order to be selected for a potential purchase. For this reason, three factors have been studied based on the artificial intelligence technique.

The first factor served to index crucial concepts in order to identify studies according to a broad rating scale. Artificial intelligence techniques are present in order to handle this type of request. Indexing is a highly specialized process that depends on the words chosen, the moment that word appears and the context of its first use. All circumstances are handled using natural language processing based on highly sophisticated algorithms, such as: Keyword augmentation and screening (NKAS); Natural language processing (NLP); Geoparsing; TimeBank corpus. TimeML; ISO-TimeML; THYM; RelEx algorithm.

In order to be well documented and to have something new in the world of research, librarians are invited to find the means of carrying out a synthesis on the research work already selected in the first phase. To achieve this goal, the production of summaries that ensure access to first-degree information is the critical goal sought by a librarian. This mission can be executed through an intelligent information processing system based on the ROUGE technique as a technique for ranking the most important sentences according to their contribution score. On the one hand, the DQN (Deep Q-Network) technique suggests conceptual proposals when developing an abstract. The terminology is organized in the form of a network where the terms with the strongest meaning are considered the most significant in the search.

Finally, the citation is an absolutely essential point of reference for correctly selecting the most famous studies in the scientific world. For this reason, an intelligent recommender system is highly demanded by the library community. We

focused on two aspects in order to evaluate published research.

The first aspect is the context surrounding the citation. A citation context can be identified based on a set of machine-supervised sequence tagging. Librarians can identify targeted citations based on a set of constraints such as: #the citation that contains the closest noun phrase; #the citation that contains conjunctive adverbs; #the quoting sentence contains multiple references…etc. The second aspect focuses on the percentage of citation localization. In this respect, the librarian seeks to identify the most frequently cited studies and even which references are cited more than once. To handle this type of queries, the Stanford CoreNLP Manning has been proposed as an NPL-based tool. This instrument works in relation to a set of markers such as: part-of-speech (POS) labels; the "Regexner" tag which recognizes named entities such as character, location, organization type and even metric and temporal objects. .etc.

The second aspect of the evaluation is based on the type of classification desired before the acquisition. First, lexical classification seeks to identify the citation with respect to the linguistic field. We quote words, words relating to a field of research, pronouns and even verbs relating to time, place, proposal, demonstration, advice....etc. The second level of classification aims to identify the distribution of a citation within the framework of a contextual classification. We find that captions related to criticism of an approach or process can be represented using a linguistic marker, citations related to a linguistic field based on comparison, or even citations that suggest towards process of development and progress following the implementation of an approach or process. The third aspect of classification is displayed through the popularity rate of one quote compared to another. Buyers seek to target journals with the highest citations in research. This can be counted using a machine. Finally, the analysis of pre-existing studies according to the degree of feeling can be used by the scientific community in order to identify opinion clusters. This can be translated by algorithms, called Citation Typing Ontology (CiTalO), as a sophisticated viewer of citation networks that identify citations based on the sentimental intentions of each author.

In order to properly measure the effect of each element on the latent variable, namely "purchase intention", a quantitative study will be deployed to validate all the measures that have been presented in this theoretical phase.

5. Bibliography List:

1. Abu-Jbara, A., Ezra, J., and Radev, D. (2013). Purpose and polarity of citation: Towards nlp-based bibliometrics. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies, (pp. 596-606).
2. Agarwal, S., Choubey, L., and Yu, H. (2010). Automatically classifying the role of citations in biomedical articles. In AMIA Annual Symposium Proceedings (p. 11). American Medical Informatics Association.
3. Belkebir, R., and Guessoum, A. (2015). A supervised approach to arabic text summarization using adaboost. In New contributions in information systems and technologies, 227-236.
4. Bertin, M., and Atanassova, I. . (2018). The context of multiple in-text references and their signification. International Journal on Digital Libraries, 127-138.
5. Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. (2008). Can all tags be used for search? Proceeding of the 17th ACM Conference on Information and Knowledge

Management (pp. 193–202). doi: 10.1145/1458082.1458112.
6. Bornmann, L., Haunschild, R., and Hug, S. E. (2018). Visualizing the context of citations referencing papers published by Eugene Garfield: A new type of keyword co-occurrence analysis. Scientometrics, 427-437.
7. Cao, G., Liang, M. and Li, X. (2018). How to make the library smart? The conceptualization of the smart library". The Electronic Library, 811-825.
8. Cave, A. (2019). Can The AI Economy Really Be Worth $150 Trillion By 2025. Retrieved from Forbes:. Retrieved 05 2022, from https://www. forbes. com/sites/andrewcave/2019/06/24/can-the-ai-economy-really-be-worth-150-trillion-by-2025.
9. Committee For The Corporate Governance Of Listed Companies. (1999). REPORT CODE OF CONDUCT. MILANO: Borsa Italiana S.p.A.
10. Day, M. Y., and Chen, C. Y. (2018). Artificial intelligence for automatic text summarization. In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 478-484). IEEE.
11. DeLozier, G., Baldridge, J., and London, L. (2015). Gazetteer-independent toponym resolution using geographic word profiles. In Twenty-Ninth AAAI Conference on Artificial Intelligence.
12. Di Iorio, A., Nuzzolese, A. G., and Peroni, S. (2013). Towards the Automatic Identification of the Nature of Citations. In SePublica, 63-74.
13. Furner, J. (2007). User tagging of library resources: toward a framework for system evaluation. International Cataloguing and Bibliographic Control, 47-51.
14. GeoNames. (2020). The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge. Consulté le 05 02, 2022, sur https://www.geonames.org/
15. Gerolimos, M. (2013). Tagging for libraries: a review of the effectiveness of tagging systems for library catalogs. Journal of Library Metadata, 36-58.
16. Hu, Z., Chen, C., and Liu, Z. (2015). The Recurrence of Citations within a Scientific Article. In ISSI, 21.
17. Iqbal, S., Hassan, S. U., Aljohani, N. R., Alelyani, S., Nawaz, R., and Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: An overview of empirical studies. Scientometrics, 6551-6599.
18. Khan, A., and Salim, N. . (2014). A review on abstractive summarization methods. Journal of Theoretical and Applied Information Technology, 64-72.
19. Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out, 74-81.
20. Ma, J., Wu, X., and Huang, L. (2022). The Use of Artificial Intelligence in Literature Search and Selection of the PubMed Database. Scientific Programming.
21. Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, (pp. 55-60).
22. Massis, B. (2018). Artificial intelligence arrives in the library. Information and Learning Science.

23. Mifrah, S., Hourrane, O., El Habib Benlahmar, N. B., and Rachdi, M. (2018). Citation Sentiment Analysis: A Brief Comprehensive Study. J. Islam. Ctries. Soc. Stat. Sci, 145-156.
24. Mogali, S. (2014). Artificial Intelligence and its applications in Libraries. In Conference: Bilingual International Conference on Information Technology: Yesterday, Today and Tomorrow, At Defence Scientific Information and Documentation Centre. Ministry of Defence Delhi.
25. Ng, V., Rees, E. E., Niu, J., Zaghool, A., Ghiasbeglou, H., and Verster, A. (2020). Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. Canada Communicable Disease Report, 186-191.
26. Nasheri, N., Vester, A., and Petronella, N. . (2019). Foodborne viral outbreaks associated with frozen produce. Epidemiology & Infection, 147.
27. Presutti, V., Draicchio, F., and Gangemi, A. (2012). Knowledge extraction based on discourse representation theory and linguistic frames. In International conference on knowledge engineering and knowledge management (pp. 114-129). Springer, Berlin, Heidelberg.
28. Rajasekaran, A., & Varalakshmi, R. (2018). Review on automatic text summarization. Inter. J. Eng. Technol, 456-460.
29. Saarti, J., and Tuominen, K. (2020). Openness, resource sharing and digitalization–an examination of the current trends in Finland. Information Discovery and Delivery.
30. Salau, S. A., and Oluwade, B. (2017). Acquisition Process for Electronic Journals in Academic Libraries. AFRICAN JOURNAL OF COMPUTING & ICT, 18.
31. Santos, J., Anastácio, I., and Martins, B. (2015). Using machine learning methods for disambiguating place references in textual documents. GeoJournal, 375-392.
32. Sarkar, K. (2013). Automatic single document text summarization using key concepts in documents. Journal of information processing systems, 602-620.
33. Small, H. (2018). Characterizing highly cited method and non-method papers using citation contexts: The role of uncertainty. Journal of Informetrics, 461-480.
34. Spies, R., Grobbelaar, S., and Botha, A. (2020). A scoping review of the application of the task-technology fit theory. In Conference on e-Business, e-Services and e-Society. Springer, Cham.
35. Taboada, M., Brooke, J., and Tofiloski, M. Voll. K., y Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 267-307.
36. Voorbij, H. (2012). The value of LibraryThing tags for academic libraries. Online information review.
37. Wang, W., Villavicencio, P., and Watanabe, T. (2012). Analysis of reference relationships among research papers, based on citation context. International Journal on Artificial Intelligence Tools, 21.
38. Wibisono, Y., and Hendratmo W, D. (2007). Generating indicative and informative summaries for search engine results. Proceedings of the International Conference on Electrical Engineering and Informatics.
39. Widjanarko, A., Kusumaningrum, R., and Surarso, B. (2018). Multi document summarization for the Indonesian language based on latent dirichlet allocation and significance sentence. In 2018 International Conference on Information and

Communications Technology (ICOIACT) (pp. 520-524). IEEE.
40. Widyassari, A. P., Affandy, A., Noersasongko, E., Fanani, A. Z., Syukur, A., and Basuki, R. S. (2019). Literature review of automatic text summarization: research trend, dataset and method. In 2019 International Conference on Information and Communications Technology (ICOIACT) (pp. 491-496). IEEE.
41. Worldometer. (2020). Real time world statistics. Consulté le 03 02, 2022, sur https://www.worldometers.info/
42. Yang, Z., Zhang, S., Shen, W., Xing, X., and Gao, Y. . (2018). Artificial intelligence related publication analysis based on citation counting. IEEE Access, 52205-52217.
43. Yao, K., Zhang, L., Luo, T., and Wu, Y. (2018). Deep reinforcement learning for extractive document summarization. Neurocomputing, 52-62.
44. Yousefi-Azar, M., and Hamey, L. . (2017). Text summarization using unsupervised deep learning. Expert Systems with Applications, 93-105.
45. Yusuf, T. I., Adebayo, O. A., Bello, l. a., and Kayode, J. O. (2022). Adoption of artificial intelligence for effective library service delivery in academic libraries in Nigeria.
46. Zhu, X., Turney, P., Lemire, D., and Vellino, A. (2015). Measuring academic influence: Not all citations are equal. Journal of the Association for Information Science and Technology, 408-427.