

Article

Toward a Model to Evaluate Machine-Processing Quality in Scientific Documentation and Its Impact on Information Retrieval

Diana Suárez López ^{1,*}, José María Álvarez-Rodríguez ¹  and Marvin Molina-Cardenas ²

¹ Computer Science and Engineering Department, Carlos III University of Madrid, 28911 Leganés, Spain; josemaria.alvarez@uc3m.es

² Engineering Department, Libre University, Barranquilla 08002, Colombia; marvin.molinac@unilibre.edu.co

* Correspondence: 100394339@alumnos.uc3m.es

Abstract: The lack of quality in scientific documents affects how documents can be retrieved depending on a user query. Existing search tools for scientific documentation usually retrieve a vast number of documents, of which only a small fraction proves relevant to the user's query. However, these documents do not always appear at the top of the retrieval process output. This is mainly due to the substantial volume of continuously generated information, which complicates the search and access not properly considering all metadata and content. Regarding document content, the way in which the author structures it and the way the user formulates the query can lead to linguistic differences, potentially resulting in issues of ambiguity between the vocabulary employed by authors and users. In this context, our research aims to address the challenge of evaluating the machine-processing quality of scientific documentation and measure its influence on the processes of indexing and information retrieval. To achieve this objective, we propose a set of indicators and metrics for the construction of the evaluation model. This set of quality indicators have been grouped into three main areas based on the principles of Open Science: accessibility, content, and reproducibility. In this sense, quality is defined as the value that determines whether a document meets the requirements to be retrieved successfully. To prioritize the different indicators, a hierarchical analysis process (AHP) has been carried out with the participation of three referees, obtaining as a result a set of nine weighted indicators. Furthermore, a method to implement the quality model has been designed to support the automatic evaluation of quality and perform the indexing and retrieval process. The impact of quality in the retrieval process has been validated through a case study comprising 120 scientific documents from the field of the computer science discipline and 25 queries, obtaining as a result 21% high, 39% low, and 40% moderate quality.

Keywords: information retrieval; metrics of quality; Open Science; document retrieval; machine-processing quality



Citation: Suárez López, D.; Álvarez-Rodríguez, J.M.; Molina-Cardenas, M. Toward a Model to Evaluate Machine-Processing Quality in Scientific Documentation and Its Impact on Information Retrieval. *Appl. Sci.* **2023**, *13*, 13075. <https://doi.org/10.3390/app132413075>

Academic Editor: Yosoon Choi

Received: 17 October 2023

Revised: 23 November 2023

Accepted: 5 December 2023

Published: 7 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The latest advances in technologies such as natural language processing [1], image recognition [2], neural networks [3], etc., have extensively enriched the theory and methods of information retrieval. More specifically, large language models such as BERT [4,5], BARD [6], Llama [7,8], GPT-x [9], and RoBERTa [10] are now being fine-tuned to perform traditional natural language processing tasks such as information extraction, name entity recognition, text summarization, text classification, and, in most cases, building question/answer systems. However, the need of providing precise retrieval systems remains challenging to fulfill since these large language models can hardly trace the results of a user query to a data source. In the context of scientific documentation, this fact is even more relevant when someone is looking for specific research artifacts such as an equation,

a model, a table, etc. That is why traditional retrieval systems are still relevant in the field of scientific documentation.

In this context, the process of extracting information from documents hosted in scientific repositories on a specific subject is partially efficient [11] due to the massive volumes of documents and their structural differences. It is not easy to efficiently extract and retrieve information from these documents due to the variety and diversity of the formats in which documents are published, which implies the necessity of employing information extraction techniques for each distinct format, resulting in complications during their processing. Hence, the automatic extraction of content and metadata from scientific repositories remains a challenge for researchers [12]. This challenge also implies issues regarding document classification and machine-processing quality, meeting information needs, and providing precise query results. Additionally, there is a need for metrics to measure the quality of text-based contents and other research artifacts, as well as specific metrics tailored to scientific documentation to ensure that the Open Science principles are fulfilled. Models for predicting or measuring quality are required, alongside considerations of accessibility and usability to ensure reproducibility.

The fields of application of information retrieval are increasingly extensive, due to changes in behavior and dissemination of science [2], making it difficult to search and access due to the heterogeneity of the information source such as tables, equations, and algorithms, among others. The quality, visibility, and accessibility [13,14] of scientific documents also have impact in the satisfaction of the user's need, because several objects may coincide with the query with different degrees of relevance [15].

Information retrieval methods all have in common features to query and classify unstructured data (documents) through the use of computational tools, so that the documents of interest to users can be quickly and accurately located, thus achieving greater efficiency and effectiveness [1], which are usually measured using the metrics of precision and recall. In this sense, the focus of information retrieval is to evaluate the sensitivity and importance of the keywords or concepts within contents, analyze the potential search intent and the user's target range, and accurately feed the results back [16].

On the other hand, the Open Science initiative is looking for bringing research closer to society through a set of good research practices such as open access to data, results, and scientific articles with the objective of making research properly available for validation, reuse, etc. [17]. Through these practices, the contribution of researchers becomes more visible by applying the principle of sharing academic information. That is why researchers are increasingly seeking more forms of collaboration, which has led to the practices related to Open Science being used more frequently.

The objective of this research is to assess the machine-processing quality of scientific documents to demonstrate their impact on user query responses in information retrieval systems for scientific documents. This evaluation focuses on the relevance of the output documents in three domains: (1) accessibility, visibility, metadata, and editorial policies; (2) comprehension and readability of English abstracts, as well as their ease of interpretation; (3) research reproducibility, verified through access and utilization of resources hosted in repositories. In this manner, it contributes to enhancing the relevance of search results in information retrieval systems by considering only those documents that meet high-quality standards.

The structure of the document is organized as follows: Section 2 delves into the analysis of related works. Section 3 presents the proposed method in detail. Section 4 provides the results of the conducted experiments. Finally, the conclusion is given in Section 5.

2. Related Work

Information retrieval is an interdisciplinary field, as it encompasses areas such as computer science, information sciences, linguistics, and artificial intelligence, among others. It has been significantly influenced by technological advancements that have introduced novel approaches, procedures, storage, and information access methods, contributing to the

evolution of the information society. These developments have facilitated the creation of new tools for accessing information in various formats and have led to the implementation of databases and advances in the natural language processing domain. From web search engines to recommendation systems, the ability to efficiently find relevant information remains a relevant task. The use of techniques such as natural language processing with large language models and machine learning techniques has transformed how we access, search, filter, and retrieve information, thereby optimizing the accuracy and effectiveness of retrieval systems.

Sidi and Gunal [18] propose a semantic entity-based search approach to enhance document retrieval. This method enables an improved document retrieval and classification system, enhancing the precision of the retrieval system. The method's evaluation used the TREC 2004 and MSMARCO document collections. In their research, Nagumothu et al. [19] demonstrated that Linked Data Triples in document relevance classification can significantly enhance the accuracy of classification in information retrieval systems based on deep learning techniques. To achieve this, they suggest constructing additional semantic features from natural language processing elements, such as named entity extraction, topic modeling, and linking these elements through Linked Data Triples.

Frihat et al. [20] apply natural language processing techniques and regression to assess and predict the readability and technicality of abstracts extracted from PubMed documents. The authors propose that these evaluative aspects can be integrated into the information retrieval process to facilitate search results and classify documents relevant to healthcare professionals. Al Sibahee et al. in [21] propose an innovative and effective solution for information retrieval in documents that addresses privacy and efficiency challenges in retrieving similar documents from encrypted data. They utilize a fingerprint algorithm to compare documents without exposing their content.

The information retrieval system presented by Yeshambel, Mothe, and Assabie in [22] identifies optimal representations for documents and queries in the Amharic language. It focuses on selecting term structures and stop words based on the language's morphological characteristics. Similarly, Novak, Bizjak, Mladenić, and Grobelnik [23] propose a supervised multilingual information retrieval system based on machine learning, capable of calculating document relevance according to the query. This system exhibits high precision and provides additional insights into why a particular document is deemed relevant. Lechtenberg et al. [24] introduce an innovative approach to information retrieval from citation and scientific abstract databases such as Scopus and Web of Science. Their approach makes use of a Monte Carlo sampling to construct query chains, reducing the necessity of reviewing irrelevant documents and increasing the likelihood of uncovering potentially overlooked relevant documents. The method is applicable across a broad spectrum of research domains and can alleviate the need for costly and scarce human intervention.

3. Materials and Methods

3.1. Background

This work focuses on building a quality model for scientific documents based on three of the eight policies of Open Science (OS) with the objective of applying the values of openness, transparency, collaboration, and reuse to the entire cycle of scientific research from design to publication [25]. These values, especially the reuse of data, can have a direct and immediate impact on innovation and the way research is conducted. Below, the three policies upon which the research is based are described:

1—Open information. Based on the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [26], OS promotes that data should be findable, accessible, interoperable, and reusable; although not all scientific documents can be open due to certain restrictions, access must be ensured through authentication procedures. In that sense, OS promotes that research artifacts should be more open for sharing, collaboration, reviewing, refutation, improvement, and reuse so that both the research community and the society can benefit [27]. To do so, it is necessary to provide the proper metadata and ontologies

in public repositories, making these scientific databases interoperable where machines can access publications and their metadata and contents autonomously, without human intervention [28]. The exchange of open research data should become a common outcome of scientific research.

For instance, the set of data stored in the research repositories should be properly described using metadata so that they can be automatically processed, understood, found, and reused by other researchers. Such metadata should include attributes to describe the provenance information, license, level, and conditions of access and a semantic categorization [29]. In this sense, metadata facilitate effective retrieval, and they should be available in a readable format. It is also important to define metadata using a common representation knowledge framework, being a cornerstone for information retrieval processes.

2—Generation of new quality indicators to complement the conventional ones as the H-index (Hirsch), i10-index, citations' IF (Impact Factor), SJR (Scimago Journal Rank), JCR (Journal Citation Reports), or Ranking Core: assuming that traditional metrics to measure scientific impact have been proven to be problematic since they mainly focus on publications only at the journal level, Open Science also seeks metrics that can evaluate some additional elements that allow retrieving, reproducing, and replicating the information, for which the following metrics have been defined [30] to

- Evaluate the licenses to reuse the data in a legal context.
- Evaluate whether the dataset is contained in the metadata.
- Determine whether metadata remain even if the data are no longer available.
- Determine the level of access to publications and public or restricted data, and the conditions of access.
- Evaluate the standard and machine-readable format to describe metadata.
- Measure domain-independent core metadata.

3—Reproducibility of scientific results. Reproducibility is a term used to refer to the set of attributes that Open Science should possess such as quality, reliability, and efficiency. According to [31], this should be carried out in three phases of research, reproduction, replication, and reuse, which is used by third parties; using raw data, methodologies, and other elements of the original research should also be available and properly described, allowing other researchers to run again and compare the results with those of other studies and disciplines that reach similar conclusions [32].

Replicability and reproducibility are the fundamental principles by which scientific claims are judged [26]. Reproducibility and reusability of research can be improved by encouraging transparency of the search process and products through an open scientific culture [33,34]. Without testing, reproducibility, and replicability, it is almost impossible to repeat the research design and obtain the same or similar results [35]. In a study conducted by the authors of [36], only 25% of the published results are replicable due to the unavailability of data and the lack of platforms to validate the dataset.

The reproducibility of the results has value as a mechanism to ensure good science based on truthful statements and as a driver of discoveries and innovations that generate a change in traditional research processes [37]. In this sense, there is interoperability between repositories and the corresponding portals collecting scientific production, and such repositories are an essential part of the technological infrastructure for Open Science.

Therefore, the documents to be processed must comply with a series of characteristics [38] that make it possible to quantitatively determine their quality, whose value will make it possible to group them into good, regular, and low documents [4,39]. Depending on the level of quality, the effectiveness of their retrieval will depend on the quality level; therefore, the higher the quality, the better the retrieval.

Building on these definitions, the first step is to establish a set of metrics focused on measuring the accessibility, content, and reproducibility of each document. In that sense, retrieval can be improved through a set of quality indicators measured through a set of metrics in such a way that they allow assigning a quality value to scientific documents. Then the improvement process will allow us the extraction of the best terms from the

information units that the user considers relevant or irrelevant, so the weights of a term extracted from a title and another extracted from a section must be different.

To meet the growing demand for accurate text retrieval in scientific documentation, a set of indicators and metrics based on Open Science principles is proposed to measure the quality of scientific documents, assigning a quality value according to the weighting of each of the elements that can be retrieved.

3.2. Applying the AHP Method to Prioritize Quality Indicators in Scientific Documents

The implementation of the multi-criteria decision technique Analytic Hierarchy Process [40] was carried out since it adjusts to a great extent to the requirements of the problem posed (establish a quality value for a scientific document). The method looks for weighting and prioritizing quality metrics, considering a quantitative and qualitative approach, which allows transforming multi-criteria decision problems into simple hierarchical structures to evaluate the different alternatives of criteria. The first level is the highest, where the goal is located, the following levels represent the general objectives and criteria that affect the problem, and the lower levels represent the alternatives to be evaluated [41]. The relevance of the AHP method lies in the assignment of relative weights to establish a ranking or qualification of each of the alternatives. For this purpose, pairwise comparison matrices are used to find the priority vectors, with the elements of these vectors commonly presented as constant values [42]. The method was carried out following the following steps:

1. Definitions of the indicators and metrics that allow us to evaluate the quality of the document and thus improve its retrieval; they were grouped into three areas:

Accessibility reflects how easy it is to locate and access information resources for the creation of knowledge. According to [43], it is necessary to implement metrics and indicators that measure the practices of open access to publications and data, and to re-evaluate research that produces positive and tangible benefits for society. In the context of this article, accessibility can be understood as the aspect of content from a machine's perspective, meaning that the content should be machine-processable for the purpose of indexing and retrieval.

On the other hand, publishers have determined the way to access scientific publications according to their policies, thus having hybrid models that combine articles accessible by subscription or pay-per-publication; pay-per-publication models in which the articles are all open access funded by the author or sponsoring entity; and finally, open access without payment per publication. In a study carried out by the authors of [44], they state that full open access journals are more accessed than fee-paying ones.

The content addresses the structuring and presentation of information, emphasizing how text, tables, images, and equations are displayed to facilitate their comprehension and processing by machines. In this sense, the paper must be easy to understand and must be organized clearly and succinctly, without lexical and grammatical errors. However, it is important to take into account readability, which measures the complexity and difficulty of the text, linguistic quality, semantic complexity, and fluency [45], and there are currently very few techniques to evaluate the quality of the machine's processable content.

Reproducibility: At the data level, elements were identified that can be retrievable and, in some cases, executable or at least have access and availability for reuse. The researcher needs to be clear about the conditions of use from the legal point of view, for which a standard machine's readable license must be available.

2. Building the hierarchical model of quality indicators and metrics to be machine-processable. As shown in Figure 1, two hierarchical levels have been identified: the first (red) corresponds to the scopes defined in the previous step that cover the quality indexes and the second level (green) corresponds to each of the sub-scopes (nine in total) into which the scopes of the first level are divided to classify each of the previously defined indicators. The AHP method allows us to group the different indicators to facilitate and to measure the influence on the general objectives.

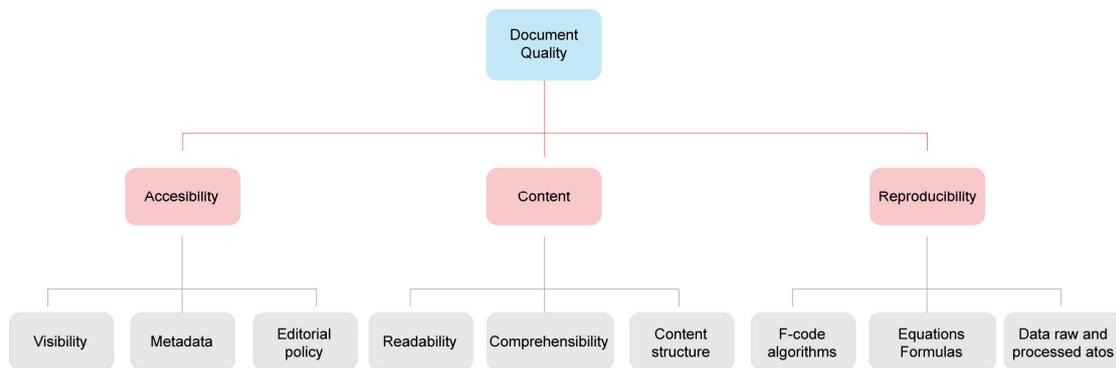


Figure 1. Hierarchy of quality indicators and areas.

For the area of accessibility, the following set of indicators and metrics [46] has been defined in Table 1, to determine the degree of completeness:

Table 1. Metrics of accessibility.

Indicator	Metrics	Description	Value
Visibility	Open access	Identify that there are no restrictions for users to access digital resources freely.	0 1
	Restricted access	Validate whether a username and password or payment are required for access or download.	0 1
	Embargoed access	Verify if the resources are available for a limited time.	0 1
	Access to metadata only	Limited access to only metadata implies that the resources are not available in this case.	0 1
	Full text	Verify full access to the metadata of the document or a portion of it.	0 1
	Dataset	Verify access to the structured set of information such as images, videos, numbers, text, tables, etc.	0 1
Content Metadata	Author’s name	Check for the existence of metadata; if found, a value of 1 is assigned; otherwise, 0 is assigned. This assignment is conducted for each of the metadata elements.	
	Title		
	Year		
	Keywords		
	Classification codes		
	Abstract		
	Multimedia objects		
	From location		
	DOI		
	URI		
	URL		
Format and versions			
Multimedia objects			
From location			
Links			
Editorial Policy	Full open access	Verify that the documents are available for free.	0 1
	Pay per download	Validate that articles can be individually downloaded for a fee without a subscription.	0 1
	Partial access	Confirm the existence of a partial access model for accessing the content.	0 1
	Subscription	Identify if the content can be accessed via subscription through a regular fee.	0 1

Regarding the content domain, Table 2 presents the following set of identified metrics and indicators:

Table 2. Metrics of the machine-processable content.

Indicator	Metrics	Description	Value
Comprehensibility	Lexical density	A greater number of different words per text results in an increased difficulty for comprehension.	0–1
	Frequency of use	The more frequent a word is, the fewer cognitive resources it will demand for perception, recognition, and integration into text processing. As the words in a text become less frequent, reading becomes more burdensome, and the process slows down.	0–1
	Sentence complexity	Measure the number of words per sentence, thus obtaining the sentence length index, and the number of complex clauses per sentence, yielding a complex clause index.	0–1
	Syntactic complexity	Measure sentence length and the quantity of modifiers.	0–1
	Punctuation marks	The average number of punctuation marks is used as one of the complexity indicators.	0–1
Readability	SSR index (measures vocabulary)	The focus is on measuring vocabulary and sentence structure to predict the relative readability difficulty of a text.	0–100 [47] <ul style="list-style-type: none"> • 0–40: very easy • 40–60: easy • 61–80: moderate difficulty • 81–100: difficult
	Readability index	Calculate the number of words, the mean number of letters per word, and its variance.	0–100 [48] <ul style="list-style-type: none"> • 80–100: very easy • 60–80: easy • 50–60: preuniversity • 30–50: selective courses • 0–30: university, scientific
	Text analysis metrics	Represent the grammatical structure of a text in the form of an abstract syntax tree to facilitate the measurement of its depth and density. In this structure, each node represents a word or phrase, and the connections between them symbolize grammatical relationships.	0–1
Content structure	Degree of compliance with a standard structure	Measures compliance considering the following: <ul style="list-style-type: none"> ■ Title ■ Abstract ■ Keywords ■ Introduction ■ State of the art, development ■ Related jobs ■ Methodology ■ Results ■ Conclusions ■ Discussions 	According to the number of items found: <ul style="list-style-type: none"> 1–3: low 3–6: half More than 6: high
	Depth of sections	Measure the levels of depth of the sections within the document, according to their importance: main sections, subsections, and sub-subsections.	Add one point (1) for each section level identified

Finally, in the area of reproducibility, another set of indicators and metrics were defined to measure the degree of completeness. Regarding indicators at this level, presented in Table 3, it should be noted that this set applies to all metrics.

Once the hierarchical structure was defined, the AHP method was executed using the open access application developed in [49].

Evaluation of the metrics: The assignment of weights was performed using the technique of expert judgment, considered as a reliable source of a topic, technique or skill, or authority on a specific subject [50], for which we proceeded to the selection of people who met the required profile, choosing 3 professionals with knowledge in computer science, software engineering, and information retrieval, aiming to establish what is most relevant for them. Similarly, algorithms were implemented to validate and verify both the documents and their machine-processable contents, to measure the completeness of each area,

taking into account the set of indicators. Once the individual judgments of the experts had been compiled, the relative weights of each metric were calculated.

Table 3. Metrics of Reproducibility.

Resource Type	Metrics	Description	Value
<ul style="list-style-type: none"> • Algorithms • Equations • Formulas • Theorems • Raw data • Processed data 	Repository	Verify the existence of a digital asset management system in which digital resources such as documents, software, multimedia files, etc., are stored and controlled.	0 1
	Platform	Verify if it provides services or resources that are useful for algorithms or source code.	0 1
	Site	Confirm access to the location where the information is hosted, typically personal websites or blogs.	0 1
	License of use	Ensure that the resources can be used without restrictions by the scientific community.	0 1
	Authorization license	Combine copyright with non-commercial use of the resources.	0 1
	Dataset	Verify the existence of a structured set of information resulting from analysis and studies, such as images, videos, numbers, text, etc.	0 1
	Text format	Formats may vary depending on the repository, so it is necessary to identify whether they are in plain text, i.e., those without formatting; with programming language extensions such as Java or Python; or structured in JSON or XML format.	0 1

To implement the AHP technique, a process involving several steps was followed: Firstly, the number of criteria to be evaluated was determined, which in this case was three. Subsequently, an expert-participation voting system was established. The criteria were evaluated in pairs, considering their priority as either A or B, using an underlying scale with values ranging from 1 to 9, where 1 represented the minimum value and 9 the maximum. Finally, the linear approach was selected with a consistency ratio of 0.1.

After the voting by the experts, the following results were obtained.

In Table 4, the results obtained for the first-level indicators analyzed can be observed; these are calculated concerning 100%, showing that accessibility contributes 31.9%, content 13.8%, and reproducibility 54.3%, this being the highest priority, thus defining the values established for calculating the weight of quality in a document.

Table 4. Hierarchy of consolidated priorities.

Decision Hierarchy			
Level 0	Level 1	Global Priorities	Rank
Quality papers	Accessibility	31.9%	2
	Content	13.8%	3
	Reproducibility	54.3%	1

Each of the indicators is broken down according to the global consensus as shown in the Table 5. In the case of the accessibility metric, the indicator with the greatest weight corresponds to metadata, followed by visibility, and the editorial policy improving the retrieval process is considered less relevant.

Concerning the content metrics, it is evident that how the content of the document is structured is considered relevant, as well as its writing, since it allows a better understanding of the document; as for readability, it is considered less relevant.

Finally, the reproducibility indicators of algorithms, source code, models, and raw data were assigned greater weight; on the other hand, equations and processed data were considered less relevant when calculating the quality of the document.

In this sense, and considering the above, the contribution made by each indicator within the sub-area is calculated, leaving the hierarchy with the defined weights (see Table 6), resulting from the normalization generated with the AHP method.

Table 5. Hierarchy with consolidated priorities level 2.

Decision Hierarchy			
Metrics	Level 2	Global Priorities	Rank
Accessibility	Visibility	33.3%	2
	Metadata	45.2%	1
	Editorial policy	21.5%	3
Content	Compressibility	43.7%	2
	Readability	11.9%	3
	Content structure	44.4%	1
Reproducibility	Algorithms/source	42.8%	1
	Equations/theorems	13.4%	3
	Raw data	40.2%	2
	Processed data	3.6%	4

Table 6. Indicator weights and metrics.

Indicator	Metrics	Normalized	Weight
Accessibility $w = 0.138$	Visibility	0.333	0.1062
	Metadata	0.452	0.1442
	Editorial policy	0.215	0.0686
Content $w = 0.319$	Compressibility	0.437	0.0603
	Readability	0.119	0.0164
	Content structure	0.444	0.0613
Reproducibility $w = 0.543$	Algorithms/source	0.4280	0.2324
	Equations/models	0.1340	0.0728
	Raw data	0.4020	0.2183
	Processed data	0.0360	0.0195

3.3. Technological Implementation: Indexing and Retrieval Process

Once the quality model was defined, a software system, see Figure 2, was designed to provide an implementation to the required set of metrics and to automatically gather information, calculate quality and the index, and retrieve scientific documents. For the indexing and retrieval process, we used the Elasticsearch platform, an open-source text search and analysis server that offers stable and reliable real-time retrieval services [51]. Additionally, Elasticsearch facilitates high-speed document stream processing and indexing [52]. Furthermore, Elasticsearch utilizes the Query DSL (Domain-Specific Language) for information retrieval based on the JSON format, enabling the definition of each search parameter. Elasticsearch consistently provides real-time retrieval services [53]. In this context, inverted indexes were employed as a fragmentation technique due to their high scalability, which enhances search speed through a distributed architecture.

To evaluate the machine-processable quality of scientific documents, an algorithm (source code available at <https://github.com/jtvans/Algorithm-Dataset> accessed on 4 July 2023) was implemented to automatically calculate the values of the metrics. The objective is to assess the quality of machine-processable content for the purpose of indexing and retrieving information, with a particular emphasis on its document structure rather than the content per se. This methodology enables the algorithm to carry out the process more efficiently. For this purpose, natural language processing techniques were implemented such as an automatic text classification, extraction, and analysis, as well as the reuse of the Crossref algorithm through which the metadata of the collection of documents are automatically extracted by connecting the API with the different scientific databases where the documents are hosted. This is possible thanks to the services and features provided

with Crossref, which are based on the principles of 20/20 metadata [52], which issues a response in the JSON format with the information of each document. This algorithm consists of three modules:

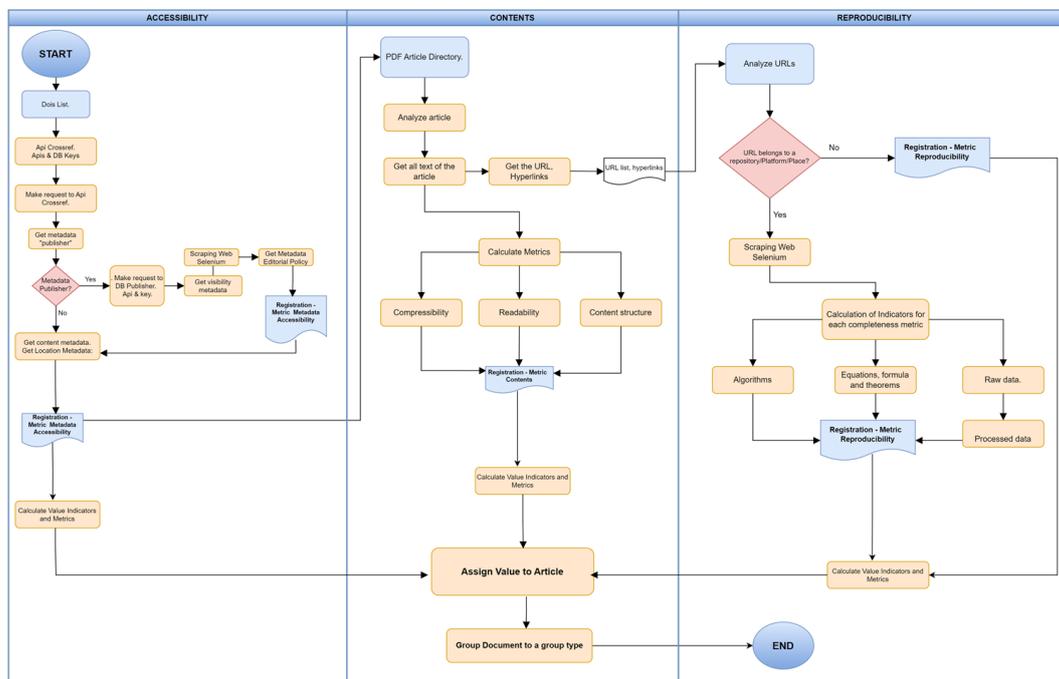


Figure 2. Flowchart of the evaluation process.

Accessibility module: In which the level 2 metrics corresponding to visibility were evaluated, comprising 7 indicators; the set of metadata contains 14 elements and the editorial policies are with 7. This module is responsible for verifying whether each of the indicators is found in the document.

Content module: It is in charge of evaluating the textual content that will be processed by the machine. The test is performed through the abstract extracted from the document using Python’s NLTK open library.

Reproducibility module: It is in charge of identifying if there are repositories in the document and access to them; it automatically redirects the corresponding links to these repositories, and once inside them, it makes a tour to verify if it has the resources such as algorithms, equations, and data; in the same way, it verifies if it complies with the reproducibility principles defined in the model.

3.4. Applying the Quality Model to a Set of Documents

To carry out the execution of the algorithm, a sample of 120 documents (see Table 7) was selected from various scientific databases, primarily academic journals and conferences, all in PDF format. Subsequently, they were classified into three categories: high quality, moderate quality, and low quality, based on the machine-processing quality assessment result.

For this experiment, the documents have been classified into three groups as shown in Figure 3, where 40% are moderate documents, 38% are of low quality, and 22% are high documents. Thus, it is determined that 78% of the sample does not meet the high-quality criteria as determined by the algorithm. The complete results are available at the following Github repository.

Table 7. Machine-processing quality results for the dataset of documents.

DOI	Accessibility	Content	Reproducibility	Weight	Category
10.1007/s11831-020-09496-0	11.98	10.6064	0	22.1044	Low
10.1007/s11277-020-07108-5	12.5376	13.713	0	26.2506	Moderate
10.1109/JIOT.2017.2683200	11.0576	12.8324	0	23.89	Low
10.1007/s42979-021-00521-y	12.5376	12.849	0	25.3866	Moderate
10.1007/s11277-021-08439-7	13.1876	9.8987	0	23.0863	Low
10.1109/ACCESS.2019.2930345	12.5778	13.9471	0	26.5249	Moderate
10.1109/ACCESS.2018.2842034	13.2278	12.6112	0	25.839	Moderate
10.1109/ACCESS.2018.2842034	13.2278	12.6112	0	25.839	Moderate
10.1109/ACCESS.2019.2908684	12.5778	9.6833	0	22.2611	Low
10.1109/ACCESS.2018.2877293	12.5778	10.931	0	23.5088	Low
10.1109/JTEHM.2018.2822681	12.5778	11.653	0	24.2308	Low
10.1109/ACCESS.2018.2864675	12.5778	13.365	0	25.9428	Moderate
10.22430/22565337.1485	6.3678	13.2353	0	19.6031	Low
10.1109/ACCESS.2020.3024066	13.2278	12.9525	0	26.1803	Low
10.1007/s40860-020-00116-z	12.5376	15.7444	0	28.282	Moderate
10.1109/ACCESS.2020.3004486	12.5778	12.3583	0	24.9361	Low
10.1109/ACCESS.2020.2998983	12.5778	12.3886	0	24.9664	Low
10.1109/ACCESS.2020.2986381	12.5778	14.1379	0	26.7157	Moderate
10.1007/s11036-018-1085-0	12.5376	16.4839	0	29.0215	Moderate
10.1109/ACCESS.2019.2951164	12.5778	12.9714	0	25.5492	Moderate
10.1109/TASE.2020.3004313	13.0976	12.5862	0	25.6838	Moderate
10.15446/esrj.v24n2.87441	6.3678	12.8262	0	19.194	Low
10.1007/s11227-021-03653-3	12.5376	12.8196	0	25.3572	Moderate
10.1109/MS.2017.2	11.0576	9.6594	0	20.717	Low
10.1016/j.jnca.2016.10.013	11.0576	16.9416	0	27.9992	Moderate
10.1109/ACCESS.2020.3022641	12.5778	12.7838	0	25.3616	Moderate
10.1109/ACCESS.2019.2893445	6.3678	8.6874	0	15.0552	Low
10.19053/01211129.v26.n46.2017.7326	6.3678	10.5453	0	16.9131	Low
10.1109/ACCESS.2019.2956980	12.5778	14.2777	0	26.8555	Moderate
10.1109/ACCESS.2019.2910411	12.5778	13.4773	0	26.0551	Moderate
10.1109/ACCESS.2019.2906265	12.5778	10.4612	0	23.039	Low
10.1109/ACCESS.2019.2905017	12.5778	13.3965	0	25.9743	Moderate
10.2991/ficaset-18.2018.20	3.1188	5.2043	0	8.3231	Low
10.1007/s11277-020-07446-4	12.5376	13.3678	0	25.9054	Moderate
10.1109/ACCESS.2019.2932609	12.5778	13.2871	0	25.8649	Moderate
10.1007/s11227-018-2288-7	12.5376	12.4159	0	24.9535	Low
10.1109/CCAA.2016.7813916	10.018	13.9867	0	24.0047	Low
10.1109/ACCESS.2020.2988059	13.0976	16.497	0	29.5946	Moderate
10.15517/eci.v8i1.30010	5.848	14.3093	0	20.1573	Low
10.1109/ACCESS.2020.2986681	13.0976	12.4434	0	25.541	Regular
10.1007/s12525-020-00405-8	13.2278	12.6494	0	25.8772	Moderate
10.1109/ACCESS.2019.2941978	12.5778	13.8413	0	26.4191	Moderate
10.1109/ACCESS.2019.2958257	12.5778	13.4266	0	26.0044	Moderate
10.1007/s10270-020-00785-7	13.7476	16.165	23.2404	53.153	High
10.11144/Javeriana.iyu21-1.iprc	5.848	10.9357	0	16.7837	Low
10.1109/ACCESS.2018.2793280	12.5778	10.9511	0	23.5289	Low
10.1109/ACCESS.2019.2895368	12.5778	9.9894	0	22.5672	Low
10.1109/JIOT.2020.2988321	13.0976	12.6334	0	25.731	Moderate
10.1186/s13635-020-00111-0	13.7476	11.9492	0	25.6968	Moderate
10.1109/ACCESS.2019.2946400	12.5778	14.5298	0	27.1076	Moderate
10.15446/dyna.v85n204.68264	6.3678	14.245	0	20.6128	Low
10.1007/s00521-020-04874-y	12.5376	15.6602	0	28.1978	Moderate
10.1109/ACCESS.2019.2951168	12.5778	13.2015	0	25.7793	Moderate

Table 7. Cont.

DOI	Accessibility	Content	Reproducibility	Weight	Category
10.1109/ACCESS.2020.2998739	12.5778	13.2083	0	25.7861	Moderate
10.1109/ACCESS.2019.2902865	12.5778	10.6249	0	23.2027	Low
10.1007/s11227-019-02928-0	12.5376	9.9435	0	22.4811	Low
10.1109/ACCESS.2020.2997761	12.5778	13.1285	0	25.7063	Moderate
10.1109/ACCESS.2019.2899828	12.5778	10.6818	0	23.2596	Low
10.1007/s40860-016-0027-5	8.8376	15.2348	0	24.0724	Low
10.1109/ACCESS.2017.2692247	12.5778	15.162	0	27.7398	Moderate
10.1109/TCC.2019.2902380	13.0976	10.0035	0	23.1011	Low
10.1109/ACCESS.2017.2717818	12.5778	9.7447	0	22.3225	Low
10.1109/TVT.2019.2944926	13.0976	13.2034	0	26.301	Moderate
10.1109/ACCESS.2018.2872799	12.5778	13.018	0	25.5958	Moderate
10.1109/ACCESS.2020.2987749	12.5778	13.4614	0	26.0392	Moderate
10.1109/ACCESS.2019.2933014	12.5778	12.8661	0	25.4439	Moderate
10.1109/ACCESS.2020.3034466	12.5778	12.3171	0	24.8949	Low
10.1109/JIOT.2015.2483023	11.0576	11.3964	0	22.454	Low
10.1109/ACCESS.2016.2607786	12.5778	9.7146	0	22.2924	Low
10.1109/ACCESS.2018.2871271	12.5778	13.7031	0	26.2809	Moderate
10.1007/s11227-016-1684-0	12.5376	12.5583	0	25.0959	Moderate
10.1007/s11277-020-07649-9	12.5376	9.5646	0	22.1022	Low
10.1016/j.jnca.2016.08.007	11.0576	13.1881	0	24.2457	Low
10.1109/loTDI.2015.22	10.018	11.8547	0	21.8727	Low
10.1109/ACCESS.2019.2927394	12.5778	12.1005	0	24.6783	Low
10.1109/JCN.2019.000049	13.0976	11.5436	0	24.6412	Low
10.1007/s10916-019-1158-z	12.5376	10.9914	0	23.529	Low
10.1109/ACCESS.2019.2929915	12.5778	12.5362	0	25.114	Moderate
10.1109/TCSI.2020.2973908	13.0976	12.3713	0	25.4689	Moderate
10.1109/ACCESS.2019.2931868	12.5778	13.0582	0	25.636	Moderate
10.1016/j.softx.2022.101218	13.7476	17.1149	23.2404	54.1029	High
10.1016/j.softx.2022.101081	13.7476	15.689	47.0238	76.4604	High
10.1016/j.iot.2022.100677	13.7476	13.5239	40.725	67.9965	High
10.1016/j.comnet.2020.107673	13.1876	14.5269	47.0238	74.7383	High
10.1016/j.softx.2022.101089	13.7476	16.6358	47.0238	77.4072	High
10.1016/j.softx.2021.100661	13.7476	16.4539	23.2404	53.4419	High
10.1016/j.iot.2020.100255	13.1876	14.0631	47.0238	74.2745	High
10.1016/j.softx.2023.101390	13.7476	14.0765	47.0238	74.8479	High
10.1016/j.dib.2023.109248	13.7476	12.45	47.0238	73.2214	High
10.1016/j.dib.2022.108400	13.0976	10.6664	35.5	59.264	High
10.1016/j.dib.2021.107530	13.7476	11.0984	47.0238	71.8698	High
10.1016/j.simpa.2022.100282	13.7476	12.1564	47.0238	72.9278	High
10.1016/j.dib.2022.108026	13.7476	12.4982	47.0238	73.2696	High
10.1016/j.dib.2021.106826	13.0976	11.3757	28.154	52.6273	High
10.1016/j.dib.2021.107453	13.7476	11.0544	47.0238	71.8258	High
10.1016/j.simpa.2020.100029	13.7476	12.9906	47.0238	73.762	High
10.1016/j.comnet.2021.108627	13.1876	12.7975	54.3	80.2851	High
10.1016/j.softx.2022.100991	13.7476	12.6896	47.0238	73.461	High
10.1016/j.softx.2022.101180	13.7476	13.7839	47.0238	74.5553	High
10.1016/j.dib.2022.108366	13.7476	13.0354	45.2500	72.0330	High
10.1016/j.softx.2022.100991	13.7476	17.1149	23.2404	54.1029	High

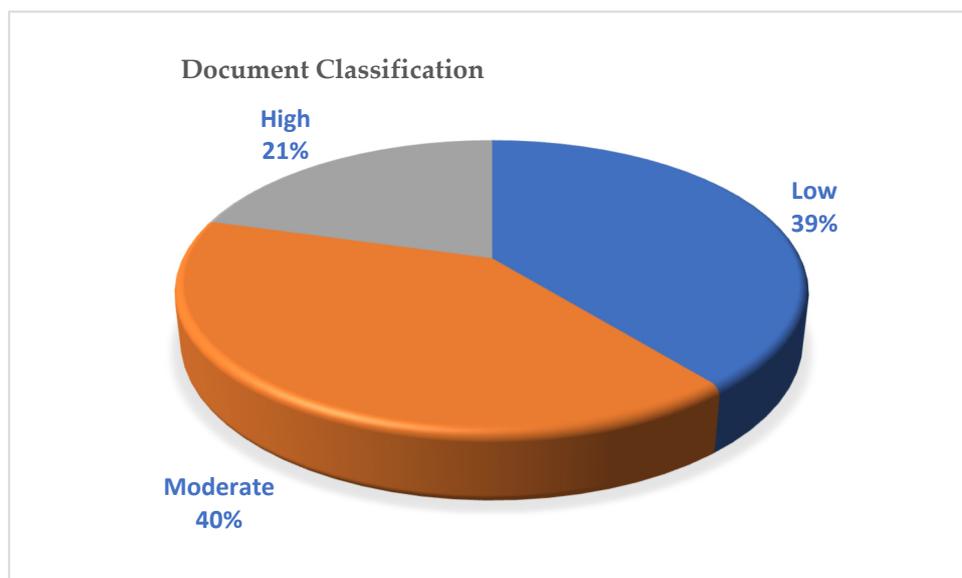


Figure 3. Classification of documents by groups according to the quality value processable machine.

4. Experimentation

The validation of the work carried out was conducted through experiments in which the implemented approach can be quantified, starting from the formulated hypothesis and all the elements implemented for its identification. In order to assess the effectiveness of the model and algorithm, the experiment was executed conducting the following steps:

Goal: measure the evaluation metrics of the information retrieval system to determine the model's effectiveness and verify whether quality has a positive impact in the information retrieval process in terms of the evaluation metrics.

Evaluation metrics: below, the common evaluation metrics of precision, recall, and the F1-score for information retrieval processes are specified.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

The formula to calculate F1 combines the precision and recall metrics into a single value.

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

True positive (TP) = retrieved relevant documents

False Positive (FP) = retrieved not relevant documents

False negative (FN) = did not retrieve relevant documents

Dataset selection: For this experiment, a balanced sample of 40 scientific documents was taken, 20 with high quality and 20 with low quality, from different information sources (see Table 7). In this case, the scientific documents were gathered from databases such as IEEE Explore, Springer, and Elsevier. Furthermore, twenty-five potential user queries (see Table 8) were designed and used for the two groups.

Execution of the experiment: Table 9 shows the queries performed in the information retrieval system created and the values obtained for each of the precision, recall, and F1 metrics used for their evaluation.

Table 8. Sample of user queries.

ID	Queries
Q1	Smart home IoT
Q2	Security system protocol
Q3	Protocol access control
Q4	Arduino base
Q5	Standalone device to internet
Q6	Smart energy management
Q7	Global positioning system
Q8	IoT hardware service
Q8	Blockchain network IoT
Q9	Management protocol
Q10	Big data for IoT
Q11	Sensor measurement
Q12	Industrial control system
Q13	Security and privacy data
Q14	Data analytics for IoT
Q15	Communication architecture for IoT
Q16	Information systems
Q17	Machine learning for IoT
Q18	Biometric data authentication
Q19	Storage data management
Q20	Smart agriculture
Q21	Remote IoT users
Q22	Wearable sensor
Q23	Secure IoT framework
Q24	Data quality of service
Q25	Smart city

Analysis of results: The results of the 25 queries are presented, and divided into two groups: the first group corresponds to the documents that have obtained a low evaluation; therefore, the precision was 42%, the recall was 17%, and the F1 was 20%, which is in the range of the expected result. In the second group, made up of the documents that obtained a high evaluation, hence a pre-accuracy of 74%, recall of 71%, and F1 of 66%, this result is within the expected range, achieving, in general, excellent accuracy and a good recall. In this sense, it can be observed that the metrics of the second group are higher than those of the first group (Figures 4–6), thus evidencing that retrieval with high-quality documents shows better results.

In addition, a two-sample Student's T-test was performed, assuming that the variances are different to compare the processes in each of the metrics, using the F-test with an alpha of 0.05 to affirm or reject the hypothesis: the quality of scientific documentation impacts information retrieval. The results are shown in Tables 10–12, respectively.

Table 9. Results of queries in the SRI and their metrics.

ID of Query	Low Documents			High Documents		
	P	R	F1	P	R	F1
Q1	1.00	0.40	0.50	0.67	0.67	0.67
Q2	0.00	0.00	0.00	0.67	0.50	0.00
Q3	1.00	0.25	0.29	1.00	0.50	0.67
Q4	0.00	0.00	0.00	0.50	1.00	0.67
Q5	1.00	0.50	0.67	0.50	0.50	0.50
Q6	0.00	0.00	0.00	1.00	0.50	0.67
Q7	0.00	0.00	0.00	1.00	1.00	1.00
Q8	1.00	0.50	0.67	0.80	0.57	0.67
Q9	0.50	0.20	0.20	0.50	0.50	0.50
Q10	0.50	0.50	0.50	0.60	1.00	0.75
Q11	1.00	0.33	0.40	0.50	1.00	0.67
Q12	1.00	0.50	0.67	0.67	0.25	0.36
Q13	0.00	0.00	0.00	1.00	1.00	1.00
Q14	1.00	0.33	0.40	0.75	1.00	0.86
Q15	0.00	0.00	0.00	0.50	0.50	0.50
Q16	0.00	0.00	0.00	1.00	0.33	0.50
Q17	1.00	0.25	0.29	0.67	0.67	0.67
Q18	0.00	0.00	0.00	0.60	1.00	0.75
Q19	0.00	0.00	0.00	0.50	1.00	0.67
Q20	0.00	0.00	0.00	0.67	1.00	0.80
Q21	0.00	0.00	0.00	1.00	1.00	1.00
Q22	1.00	0.20	0.22	1.00	0.50	0.67
Q23	0.00	0.00	0.00	0.75	0.75	0.75
Q24	0.00	0.00	0.00	1.00	0.50	0.67
Q25	1.00	0.25	0.00	0.75	0.50	0.60
AVG	0.42	0.17	0.20	0.74	0.71	0.66

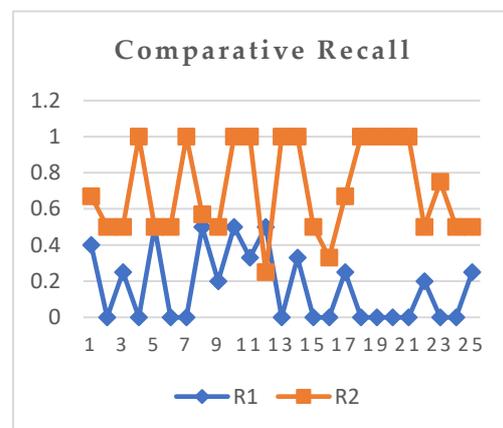


Figure 4. Comparison of precision metric results.

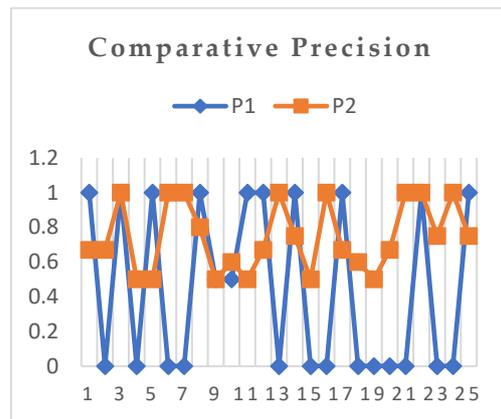


Figure 5. Comparison of recall metric results.

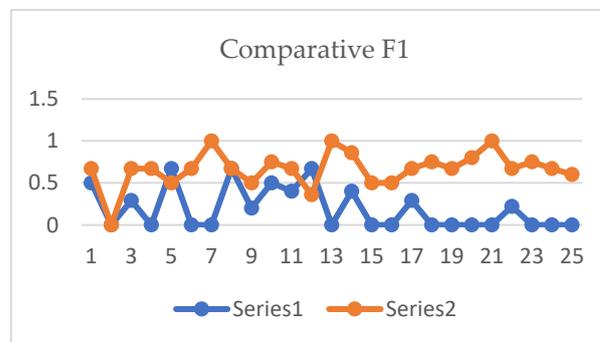


Figure 6. Comparison of F1 metric results.

Table 10. Statistical test precision metrics.

	Low Documents	High Documents
Mean	0.440	0.744
Mean standard error	0.097	0.039
Standard deviation	0.485	0.199
Observational sample	20	
Variance	0.235	0.003
Student <i>t</i>-test		
	<i>t</i> -test–one tailed	0.0033
	<i>t</i> -test–two tailed	0.0067

Table 11. Statistical test recall metrics.

	Low Documents	High Documents
Mean	0.158	0.709
Mean standard error	0.039	0.051
Standard deviation	0.199	0.259
Observational sample	20	
Variance	0.039	0.067
Student <i>t</i>-test		
	<i>t</i> -test–one tailed	0.043
	<i>t</i> -test–two tailed	1

Considering the results, it is confirmed that the quality of scientific documentation significantly impacts information retrieval. Consequently, it can be inferred that by evaluating document quality, one can predict or anticipate which documents will appear in the results of an information retrieval process.

Table 12. Statistical test F1 metrics.

	Low Documents	High Documents
Mean	0.192	0.662
Mean standard error	0.049	0.042
Standard deviation	0.249	0.210
Observational sample	20	
Variance	0.062	0.044
Student <i>t</i>-test		
	<i>t</i> -test—one tailed	0.059
	<i>t</i> -test—two tailed	0.042

5. Conclusions

This study provides an assessment of the algorithm's performance, verifying its effectiveness in processing various types of content and resources intended for the development of an information retrieval system, without delving into the evaluation of its underlying meaning. Therefore, it is essential that all published elements, such as text, images, equations, pseudocode, source code, and tables, among others, be appropriately described to enable machine processing.

During the evaluation process, documents with different sources were found. When applying the presented method, their values were different due to the nature of the databases in which they had been indexed, implying a significant impact on the quality value. Furthermore, it was observed that not all databases provide access through APIs to their repositories so a manual curation of data is sometimes required. In this sense, some documents could not be evaluated and were consequently excluded from the quality assessment.

In the future, it is expected that assessments of document quality will be conducted using text analysis and machine learning technologies, and that the measurement of publication impact will be based on the frequency of reproduction, ensuring that data availability to the scientific community and open research practices are encouraged.

Author Contributions: Conceptualization, D.S.L. and J.M.Á.-R.; methodology, D.S.L. and J.M.Á.-R.; software, D.S.L. and M.M.-C.; validation, D.S.L., M.M.-C. and J.M.Á.-R.; formal analysis, D.S.L. and J.M.Á.-R.; investigation, D.S.L. and M.M.-C.; resources, writing—original draft preparation, D.S.L.; writing—review and editing, J.M.Á.-R.; visualization, D.S.L.; supervision, J.M.Á.-R.; project administration, D.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The algorithm code is available at <https://github.com/jtvans/Algorithm-Dataset> (accessed on 4 July 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhang, X.; Li, X.; Jiang, S.; Li, X.; Xie, B. Evolution Analysis of Information Retrieval based on co-word network. In Proceedings of the 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 18–20 October 2019; IEEE: Xiamen, China, 2019; pp. 1837–1840.
- Tan, J.; Tian, Y. Fuzzy retrieval algorithm for film and television animation resource database based on deep neural network. *J. Radiat. Res. Appl. Sci.* **2023**, *16*, 100675. [[CrossRef](#)]
- Wang, Y.; Chen, L.; Wu, G.; Yu, K.; Lu, T. Efficient and secure content-based image retrieval with deep neural networks in the mobile cloud computing. *Comput. Secur.* **2023**, *128*, 103163. [[CrossRef](#)]
- Bhopale, A.P.; Tiwari, A. Transformer based contextual text representation framework for intelligent information retrieval. *Expert Syst. Appl.* **2023**, *238*, 121629. [[CrossRef](#)]

5. Thakur, N.; Reimers, N.; Rücklé, A.; Srivastava, A.; Gurevych, I. BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models 2021. *arXiv* **2021**, arXiv:2104.08663. [[CrossRef](#)]
6. Koga, S.; Martin, N.B.; Dickson, D.W. Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders. *Brain Pathol.* **2023**, e13207. [[CrossRef](#)] [[PubMed](#)]
7. Zhang, R.; Han, J.; Zhou, A.; Hu, X.; Yan, S.; Lu, P.; Li, H.; Gao, P.; Qiao, Y. LLaMA-Adapter: Efficient Fine-tuning of Language Models with Zero-init Attention. *arXiv* **2023**, arXiv:2303.16199. [[CrossRef](#)]
8. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv:2302.13971. [[CrossRef](#)]
9. Sánchez-Ruiz, L.M.; Moll-López, S.; Nuñez-Pérez, A.; Moraño-Fernández, J.A.; Vega-Fleitas, E. ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Appl. Sci.* **2023**, *13*, 6039. [[CrossRef](#)]
10. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2021**, arXiv:2108.07258.
11. Feilmayr, C. Optimizing Selection of Assessment Solutions for Completing Information Extraction Results. *Comput. Y Sist.* **2013**, *17*, 169–178.
12. Zaman, G.; Mahdin, H.; Hussain, K.; Atta-Ur-Rahman; Abawajy, J.; Mostafa, S.A. An Ontological Framework for Information Extraction from Diverse Scientific Sources. *IEEE Access* **2021**, *9*, 42111–42124. [[CrossRef](#)]
13. Suárez López, D.; Alvarez Rodriguez, J.M. Quality in Documentation: Key Factor for the Retrieval Process. In Proceedings of the Information Technology and Systems, Bogota, Colombia, 5–7 February 2020; Rocha, Á., Ferrás, C., Montenegro Marin, C.E., Medina García, V.H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 67–74.
14. Rodríguez Leyva, P.; Delgado Mesa, Y.; Viltres Sala, H.; Estrada Sentí, V.; Febles, J.P. Modelo computacional para el desarrollo de sistemas de recuperación de información. *Rev. Cuba. Cienc. Informáticas* **2018**, *12*, 173–188.
15. Tamrakar, A.; Vishwakarma, S.K. Analysis of Probabilistic Model for Document Retrieval in Information Retrieval. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; IEEE: Jabalpur, India, 2015; pp. 760–765.
16. Li, X.; Li, K.; Qiao, D.; Ding, Y.; Wei, D. Application Research of Machine Learning Method Based on Distributed Cluster in Information Retrieval. In Proceedings of the 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, 5–7 July 2019; IEEE: Haikou, China, 2019; pp. 411–414.
17. Taylor, S.J.E.; Anagnostou, A.; Fabyi, A.; Currie, C.; Monks, T.; Barbera, R.; Becker, B. Open science: Approaches and benefits for modeling & simulation. In Proceedings of the 2017 Winter Simulation Conference (WSC), Las Vegas, NV, USA, 3–6 December 2017; IEEE: Las Vegas, NV, 2017; pp. 535–549.
18. Sidi, M.L.; Gunal, S. A Purely Entity-Based Semantic Search Approach for Document Retrieval. *Appl. Sci.* **2023**, *13*, 10285. [[CrossRef](#)]
19. Nagumothu, D.; Eklund, P.W.; Ofoghi, B.; Bouadjeneq, M.R. Linked Data Triples Enhance Document Relevance Classification. *Appl. Sci.* **2021**, *11*, 6636. [[CrossRef](#)]
20. Frihat, S.; Beckmann, C.L.; Hartmann, E.M.; Fuhr, N. Document Difficulty Aspects for Medical Practitioners: Enhancing Information Retrieval in Personalized Search Engines. *Appl. Sci.* **2023**, *13*, 10612. [[CrossRef](#)]
21. Al Sibahee, M.A.; Abdulsada, A.I.; Abduljabbar, Z.A.; Ma, J.; Nyangaresi, V.O.; Umrans, S.M. Lightweight, Secure, Similar-Document Retrieval over Encrypted Data. *Appl. Sci.* **2021**, *11*, 12040. [[CrossRef](#)]
22. Yeshambel, T.; Mothe, J.; Assabie, Y. Amharic Adhoc Information Retrieval System Based on Morphological Features. *Appl. Sci.* **2022**, *12*, 1294. [[CrossRef](#)]
23. Novak, E.; Bizjak, L.; Mladenčić, D.; Grobelnik, M. Why is a document relevant? Understanding the relevance scores in cross-lingual document retrieval. *Knowl.-Based Syst.* **2022**, *244*, 108545. [[CrossRef](#)]
24. Lechtenberg, F.; Farreres, J.; Galvan-Cara, A.-L.; Somoza-Tornos, A.; Espuña, A.; Graells, M. Information retrieval from scientific abstract and citation databases: A query-by-documents approach based on Monte-Carlo sampling. *Expert Syst. Appl.* **2022**, *199*, 116967. [[CrossRef](#)]
25. Abadal Falgueras, E.; Anglada Ferrer, L. Ciencia Abierta: Cómo han evolucionado la denominación y el concepto. *An. Doc.* **2020**, *23*, 1. [[CrossRef](#)]
26. Hasselbring, W.; Carr, L.; Hettrick, S.; Packer, H.; Tiropanis, T. FAIR and Open Computer Science Research Software. *arXiv* **2019**, arXiv:1908.05986.
27. Bezjak, S.; Clyburne-Sherin, A.; Conzett, P.; Fernandes, P.; Görögh, E.; Helbig, K.; Kramer, B.; Labastida, I.; Niemeyer, K.; Psomopoulos, F.; et al. *Open Science Training Handbook*; Zenodo: Genève, Switzerland, 2018.
28. Mons, B.; Neylon, C.; Velterop, J.; Dumontier, M.; da Silva Santos, L.O.B.; Wilkinson, M.D. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Inf. Serv. Use* **2017**, *37*, 49–56. [[CrossRef](#)]
29. FAIR-Aware Online Assessment Tool. Available online: <https://fairaware.dans.knaw.nl> (accessed on 28 July 2021).
30. FAIRsFAIR Data Object Assessment Metrics: Request for Comments; FAIRsFAIR: Calgary, AB, USA, 2020.
31. DG for Research and Innovation. *Reproducibility of Scientific Results in the EU Scoping Report*; DG for Research and Innovation: Brussels, Belgium, 2020; p. 32.

32. Echtler, F.; Häußler, M. Open Source, Open Science, and the Replication Crisis in HCI. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; ACM: Montreal, QC, Canada, 2018; pp. 1–8.
33. Hasselbring, W.; Carr, L.; Hettrick, S.; Packer, H.; Tiropanis, T. From FAIR research data toward FAIR and open research software. *IT—Inf. Technol.* **2020**, *62*, 39–47. [[CrossRef](#)]
34. Munafò, M.R.; Nosek, B.A.; Bishop, D.V.M.; Button, K.S.; Chambers, C.D.; Percie du Sert, N.; Simonsohn, U.; Wagenmakers, E.-J.; Ware, J.J.; Ioannidis, J.P.A. A manifesto for reproducible science. *Nat. Hum. Behav.* **2017**, *1*, 0021. [[CrossRef](#)] [[PubMed](#)]
35. Shokraneh, F. Reproducibility and replicability of systematic reviews. *World J. Meta-Anal.* **2019**, *7*, 66–76. [[CrossRef](#)]
36. Sivagnanam, S.; Nandigam, V.; Lin, K. Introducing the Open Science Chain: Protecting Integrity and Provenance of Research Data. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning), Chicago, IL, USA, 28 July–1 August 2019; ACM: Chicago, IL, USA, 2019; pp. 1–5.
37. Kedron, P.; Li, W.; Fotheringham, S.; Goodchild, M. Reproducibility and replicability: Opportunities and challenges for geospatial research. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 427–445. [[CrossRef](#)]
38. Carreño, R.L.; Méndez, F.J.M. Sistemas de recuperación de información implementados a partir de COVID-19: Herramientas clave en la gestión de la información sobre COVID-19. *Rev. Española Doc. Científica* **2020**, *43*, e275. [[CrossRef](#)]
39. Roberts, K.; Alam, T.; Bedrick, S.; Demner-Fushman, D.; Lo, K.; Soboroff, I.; Voorhees, E.; Wang, L.L.; Hersh, W.R. TREC-COVID: Rationale and structure of an information retrieval shared task for COVID-19. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1431–1436. [[CrossRef](#)]
40. Lipovetsky, S. AHP in nonlinear scaling: From two-envelope problem to modeling by predictors. *Production* **2021**, *31*, e20210007. [[CrossRef](#)]
41. Willmer Escobar, J.; Willmer Escobar, J. Metodología para la toma de decisiones de inversión en portafolio de acciones utilizando la técnica multicriterio AHP. *Contaduría Y Adm.* **2015**, *60*, 346–366. [[CrossRef](#)]
42. Clinio, A. Ciência Aberta na América Latina: Duas perspectivas em disputa. *Transinformação* **2019**, *31*, e190028. [[CrossRef](#)]
43. Hernandez, D.; León, D.; Torres, D. Importancia de las revistas de acceso abierto: La indización como meta fundamental. *Dictam. Libre* **2020**, *13*, 81–98. [[CrossRef](#)]
44. Vainshtein, R.; Katz, G.; Shapira, B.; Rokach, L. Assessing the Quality of Scientific Papers. *arXiv* **2019**, arXiv:1908.04200.
45. 5.22. File Location (Ubicación del Archivo) (MA)—Documentación de Directrices Para Repositorios Institucionales de Investigación de la Red Colombiana de Información Científica (RedCol) 2020—1.0. Available online: https://redcol.readthedocs.io/es/latest/field_filelocation.html#aire-file (accessed on 27 July 2021).
46. Metadata in Science Publishing. Available online: <http://www.wis.win.tue.nl/infwet03/proceedings/8/> (accessed on 27 July 2021).
47. López-Anguita, R.; Montejo-R´aez, A.; Martínez-Santiago, F.; Díaz-Galiano, M. Legibilidad del texto, métricas de complejidad y la importancia de las palabras. *Proces. Del Leng. Nat.* **2018**, *61*, 101–108. [[CrossRef](#)]
48. Baquedano, M.M. Legibilidad Y Variabilidad de los Textos. *Boletín Investig. Educ.* **2006**, *21*, 13–25.
49. Goepel, K.D. Implementation of an Online Software Tool for the Analytic Hierarchy Process (AHP-OS). *Int. J. Anal. Hierarchy Process* **2018**, *10*, 469–487. [[CrossRef](#)]
50. Mendoza, A.; Solano, C.; Palencia, D.; Garcia, D.; Mendoza, A.; Solano, C.; Palencia, D.; Garcia, D. Application of the Analytical Hierarchy Process (AHP) for decision-making with expert judgment. *Ingeniare Rev. Chil. Ing.* **2019**, *27*, 348–360. [[CrossRef](#)]
51. Shah, N.; Willick, D.; Mago, V. A framework for social media data analytics using Elasticsearch and Kibana | SpringerLink. *Red Ina.* **2022**, 1179–1187. [[CrossRef](#)]
52. Metadata 2020 Principles. Available online: <https://metadata2020.org/resources/metadata-principles/> (accessed on 21 May 2023).
53. Yang, A.; Zhu, S.; Li, X.; Yu, J.; Wei, M.; Li, C. The research of policy big data retrieval and analysis based on elastic search. In Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–28 May 2018; pp. 43–46.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.