

Recall, Precision, and F-Measure for Evaluating Information Retrieval System in Electronic Document Management Systems (EDMS)

Annisa Fajriyati Sadeli¹ & Ike Iswary Lawanda²

¹Departemen Ilmu Perpustakaan dan Infomasi, Fakultas Ilmu Pengetahuan Budaya,
Universitas Indonesia, Indonesia

Correspondence email: annisa.fajriyati@ui.ac.id

Notes

Submitted: 23-08-2023

Revised: 19-10-2023

Accepted: 28-10-2023

Available online: 17-11-2023

How to cite: Sadeli, A. F., & Lawanda, I. I. (2023). Recall, Precision and F-Measure for Evaluating Information Retrieval System in Electronic Document Management Systems (EDMS). *Khizanah Al-Hikmah : Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 11(2). <https://doi.org/10.24252/kah.v11i2a8>

DOI: [10.24252/kah.v11i2a8](https://doi.org/10.24252/kah.v11i2a8)

Copyright 2023 ©the Author(s)

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



ABSTRACT

The Legal Division of PT. ABC has implemented EDMS using the Integrated Legal Document Management Application, Centralized Legal Administration System (CLAS), to make the management of documents more accessible. The urgency is the system's accuracy in retrieving data or document input. The study employed a quantitative method with an experimental approach. Data collection techniques were observation, application search tests, documentation, and literature. Data analysis regarding the recall, precision, and F-measure calculation formulas will be compared with the standard information retrieval system effectiveness. The calculation results show that the average recall value for the search equals 1 or 100%. For the precision value that uses quotation marks, it is equal to 0.9713, while for searching without quotation marks, it is equal to 0.3920. From these results, the precision value is lower than the recall value. The precision value of 0.9713 is in the range of 0.68–1.00, which is placed as high precision, meaning that information retrieval is effective. The F-measure value is 0.98, close to 1, so the tracing system is sound. CLAS is quite effective in retrieving information according to user needs.

Keywords: Information retrieval, recall, precision, F-measure

1. INTRODUCTION

Daily, an organization is required to produce substantial documentation, each holding distinct value. This phenomenon occurs mainly in organizations involved in law-related cases, The significance of documents within these organizations is heightened, given their utility as evidence during decision-making processes in legal matters. Consequently, disarray may ensue if proper organization is lacking, and swift retrieval of documents becomes challenging. The evolution from traditional hardcopy formats to an entirely digital paradigm over time has underscored the growing importance of Electronic Document Management Systems (EDMS).

This transition has proven increasingly crucial in facilitating efficient document handling. The primary function of an EDMS is to facilitate the flow of documents in an organized manner and to ensure the availability of information on demand (Alshibly et al., 2016).

The Legal Division of PT. ABC has implemented an Integrated Legal Document Management Application, known as CLAS (Centralized Legal Administration System). It was built in 2018 by PT ABC, assisted by an IT Consultant, CLAS is currently ongoing feature development. This application is used to store CLAS data and softcopy in a structured manner according to the classification of legal documents produced by the legal division. Apart from that, CLAS also focuses on a system for retrieving information from documents that have been input so that users can easily find them again. The purpose of creating this application is to make it possible for administrative management to be recorded centrally in one door and one storage place, allowing for quick and accurate administration processes, document processing, document retrieval, and document monitoring.

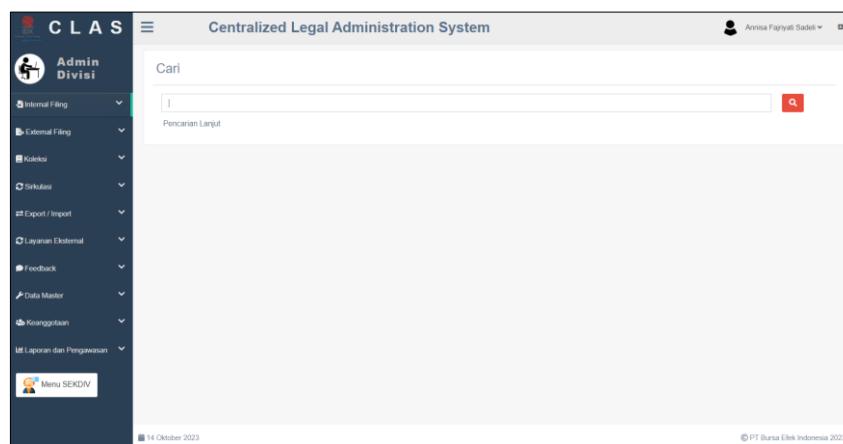


Figure 1. User interface of CLAS
Source: Institution, PT. ABC (2023)

A prominent aspect of CLAS is its rapid document retrieval capability, illustrated in Figure 1. Users can efficiently locate the documents they require by utilizing the provided search field, entering relevant keywords to initiate the search process. Information retrieval is a method for retrieving structured data stored in a collection of documents and then providing the required information (Martin & Nilawati, 2019). In short, the information retrieval system contained in the CLAS application can be a bridge between the user's information needs and the information they need. However, sometimes, users need to be met due to search results documents that are not relevant to the content in question. The values that are generally used as determinants in considering the relevance of search results are recall and precision values. These two things are considered essential for calculating the effectiveness of a search engine. The ease in retrieving information and documents input into the CLAS application will provide efficiency to the Legal Division's services in general. Sukma et al., (2015) said that apart from the timeliness of retrieving a document, it is easier for us with an increasingly large memory capacity as more and more documents are archived in a system. The greater the memory capacity used, the more files and information will be stored in this EDMS; this also encourages people to continue to develop and improve existing systems.

Several previous studies discuss recall and precision in a system. Martin & Nilawati, (2019) wrote an article entitled "Recall and Precision in the Online Public Access Catalog (OPAC) Information Retrieval System in Libraries." They found that the SLiMS application used by the Komnas HAM Library was effective in that the acquisition rate recall of 99.8% and precision of 99.6% indicates that the system can perform searches using a specific set of keywords.

Additional research indicates that there are indeed two essential steps to enhance the effectiveness of document searches and yield relevant results. These steps involve (a) effectively classifying all documents and (b) selecting representative documents within each document category. The outcome is purportedly more efficient compared to expanding keyword searches within existing search results (Song et al., 2019). Technically, the algorithm will find what it is looking for more quickly, increasing a system's recall and precision values.

In 2016 Bramer and colleagues compared recall and precision results between three databases, Embase, MEDLINE, and Google Scholar. According to the results of their research, Google Scholar has a reasonably low recall; calculations produce 72.8%, Embase/Medline produces 81.6%, and Medline alone produces 72.6%. For precision, Google Scholar earns 1.9%, Embase 2.8%, and Medline 2.8% (Bramer et al., 2016).

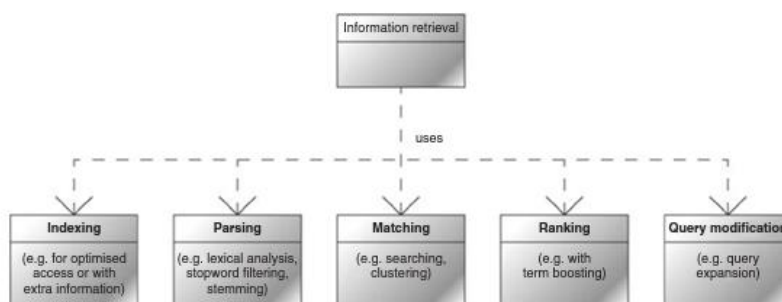


Figure 2. Information retrieval process
Source: Goker & Davies, (2009)

Information retrieval is a process where users find information contained in the system according to their needs. More specifically, information retrieval is the process of matching the query against the information objects that are indexed (Goker & Davies, 2009). Systems for retrieving information frequently include several tasks. The following processes are indexing, parsing, matching, ranking, and query modification (Goker & Davies, 2009). The Information Retrieval System has a fundamental aspect: "The relevance of the results is assessed relative to the information need, not the query" (Ceri et al., 2013). This statement is enough to explain that the documents found are relevant if they meet users' needs, not just because the documents found contain all the words in the keywords entered. However, in the end, the suitability of the search results with the content of the keywords used will be considered a good indicator of relevance as the following retrieval model according to S. Ceri (2013):

$$IRM = \{ D, Q, F, R(qk, dj) \}$$

Where :

- D = Documents in the collection
- Q= User information needs → Query
- F = search strategy used
- R = order of relevance of documents

The recall (R) is part of a document that is "really" relevant and is effectively found (Ceri et al., 2013). In information retrieval, recall relates to the ability to retrieve items of information that have been stored. To reach one hundred percent recall, we must continue to retrieve documents until both relevant documents are retrieved (Grossman & Frieder, 2004). Precision (P) is the fraction of retrieved documents relevant to the query or keywords (S. Ceri et al., 2013). Precision relates to the system's ability not to retrieve irrelevant documents. In the other opinion, precision is the ratio of the number of relevant documents retrieved to the total

number retrieved (Grossman & Frieder, 2004). We employ the following function to integrate phrase frequency:

$$P = \frac{|TP|}{|TP| + |FP|}$$

$$R = \frac{|TP|}{|TP| + |FN|}$$

Meanwhile Lancaster (1979) has long categorized assessments to measure the level of precision in assessing the effectiveness of an information retrieval system with numerical measurements and is explained as follows:

1. Value 0 - 0.33: Low Precision is said to be Ineffective
2. Value 0.34 - 0.67: Moderate Precision is said to be Less Effective
3. Value 0.68 - 1.00: High Precision is said to be Effective

Calculation of recall and precision values alone can still be developed to consider the effectiveness of information retrieval of a system. Precision, recall, and the F-measure are set-based measures (Manning et al., 2008). It can combine recall and precision to find a measure that displays reciprocity or mediation between the two. This measurement is called the F-measure, and the calculation results produce the harmonic means of recall and precision. (S. Ceri et al., 2013). The F-measure can be viewed as a compromise between recall and precision that will be high only when both recall and precision are high (Zhang & Zhang, 2018).

$$F_{\beta} = \frac{(1 + \beta^2) \times P \times R}{(\beta^2 \times P) + R}$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

(Koehrsen, 2023)

The implementation of retrieval also applies to electronic document management systems. In the legal division of PT ABC, document management is carried out using CLAS. The current document flow is as follows:

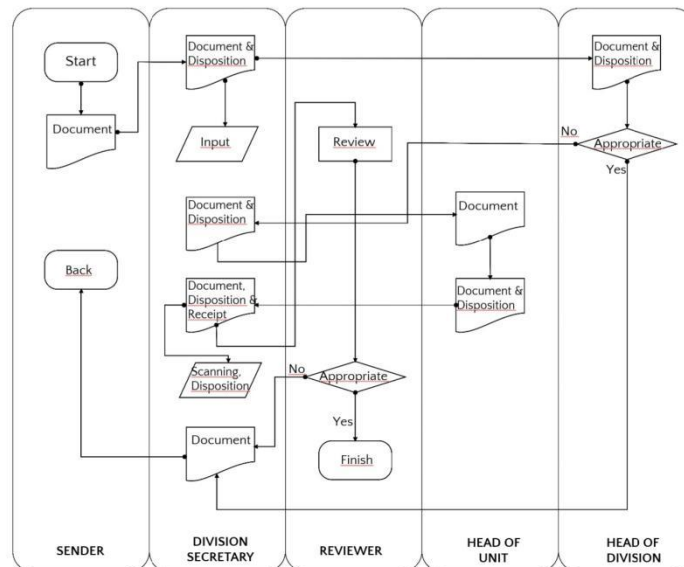


Figure 3. Flowchart Document Process
 Source: Institution, PT ABC (2018) with adaptation

The flow depicted in Figure 3 illustrates that a document is collaboratively developed with support from various stakeholders, including the divisional secretary, reviewers, heads of units, and heads of divisions. The document production flow process must be distinct from the storage method, both final and ongoing documents. Therefore, the EDMS, in this case, is one of the essential parts in assisting document management so that existing document flow processes run more effectively. The use of an EDMS-like application in a corporate environment provides convenience in the business processes that occur, especially in the case of the CLAS application as a database containing documents; the goal is to facilitate document retrieval.

In addition to the provided explanation, the ultimate urgency lies in the system's accuracy in retrieving data or documents that have been input into the system. Very important to know because it can make it easier for users to find information that is relevant to their needs precisely and accurately, so evaluating the effectiveness of information retrieval until it is as detailed as possible is very necessary (Miao & Zhu, 2022).

The main object of recall and precision research is generally carried out by librarians at OPAC (Online Public Access Catalog), which is related to library automation systems in their organization. On the other hand, research related to document management applications generally discusses management information systems or the development of document management applications. Meanwhile, in this research, apart from recall and precision calculations, the measurement concept is supplemented by F-measure calculations, and the research object is a document management system. Based on the conditions above, this research will discuss how the effectiveness of information retrieval in document management applications can help evaluate the document management system that has been built by the organization. Based on this formulation, this research aims to identify the effectiveness of recall and precision in the information retrieval system via CLAS.

2. METHODS

This study used a quantitative method with an experimental approach. Data collection techniques are observation, data search tests on applications, documentation, and literature as support (Musvida, 2017).

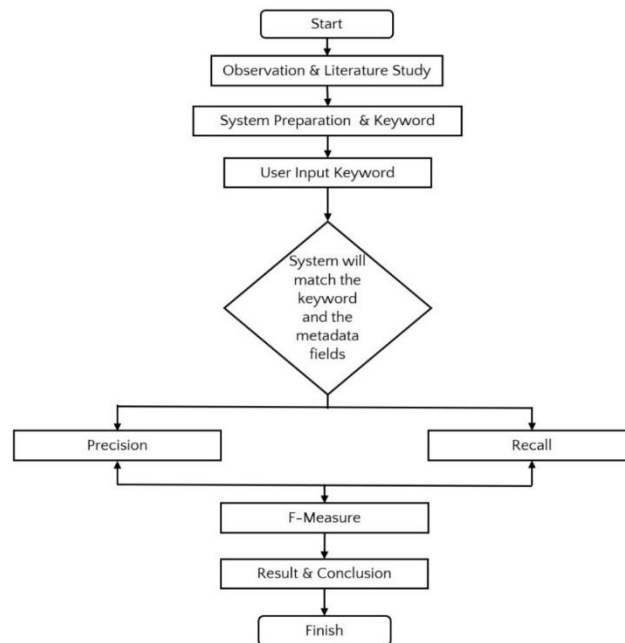


Figure 4. Research Design Method
Source: Researcher, (2023)

Figure 4 delineates the research flow, encompassing the experimental steps. The calculations involving recall, precision, and F-measure are conducted to assess the effectiveness of the CLAS application's information retrieval system. The procedure for calculating it is by conducting experiments, namely entering keywords related to the information users need to search for information through the general search feature in the CLAS Application. Users enter keywords into the available search box; then, the system will match the keyword and the metadata fields that have been input into the system. If metadata matches the keywords entered, then the data will be displayed by the application retrieval system. Each search result displayed by the system is in the form of a list of metadata that is linked to the softcopy document.

The data analysis technique used refers to the calculation formula that has been described previously, and once the results are known, they will be compared with the effectiveness standard of the information retrieval system.

3. RESULTS AND DISCUSSION

Recall, Precision, and F-Score calculations are conducted using legal document-related subjects as keywords to assess the information retrieval effectiveness within the CLAS Application. The CLAS Application offers two search types: general searches and advanced searches. The distinction lies in the diversity of search columns, with advanced searches providing more options. In contrast, general searches offer a single column, and the system retrieves data already entered into the application based on numbers, subject, and other metadata serving as browsing information sources. This experiment involved utilizing the general search feature in the CLAS Application and assessing the relevance of the search results. The relevant and irrelevant category analysis itself is stated in the following criteria:

Documents that are called relevant in this case are documents that have sequential keyword elements in the document metadata; A document is relevant if it addresses the stated information need, not because it just happens to contain all the words in the query (Manning et al., 2008).

Due to the application having two keyword search methods, namely 1) with quotation marks, for example: "Securities Trading," and 2) without quotation marks, for example, "Lease Agreement", therefore the Recall and Precision calculations are divided into two types. The following are the results of the search trials that have been carried out:

Table 1. Search results with quotation marks

Test No	Keywords	Be Discovered	Relevant	Not Relevant	Time (Second)
1	Securities Trading	209	205	4	3,14 s
2	Common Core Partners	11	11	0	0,752 s
3	Lease Agreement	652	650	2	2,08 s
4	Arbitration	16	15	1	2,07 s
5	Suspension and Unsuspension	34	32	2	1,46 s

Source: Data Result, 2023

Table 2. Search results without quotation marks

Test No	Keywords	Be Discovered	Relevant	Not Relevant	Time (Second)
1	Securities Trading	8556	205	8351	3,62 s
2	Common Core Partners	740	11	729	1,57 s
3	Lease Agreement	15200	650	14550	3,08 s
4	Arbitration	16	15	1	2,29 s
5	Suspension and Unsuspension	34	32	2	1,46 s

Source: Data Result, 2023

The data explains that the collections stored in this application database for documents that have the term "Perdagangan Efek" total 209 documents, and collections related to "Trading" and "Securities" total 8556 documents. There are 11 documents have the terms "Common Core Partners" and documents have the terms "Partners" "Core" and "Common"; there are 740 documents. Collections related to the term "Lease Agreement" contained 652 documents, while collections related to the terms "Agreement" "Lease" and "Rent" totaled 15200 documents. The number is the same for collection related to the term "Arbitrase" totaling 16 documents, both in quotation marks or not. Because the word entered as a search keyword only consists of 1 word. Last keyword, there are 34 documents relating to "Suspension and Unsuspension" while "Suspension" and "Unsuspension" are the same, namely 34 documents.

The average time needed to get information from the 5 keywords tested is 1.9004 seconds. However, this also depends on the internet connection, but the average number below this second is considered fast. Referring to what was stated by Faila Saufa & Wahyu, (2017) the less time it takes to search, the better the quality of the search engine; therefore, CLAS already has good search time.

The experimental results show that the difference between searches using quotation marks and without quotation marks raises different values, where many irrelevant documents are found in search results that do not use quotation marks. The researcher manually searched

the papers in the Master data to check relevant documents missed by the search method. Meanwhile, the author states the relevant and irrelevant criteria based on the results of subjective interpretation according to the current search results. The authors process the experimental data by referring to the previously mentioned formulas to find the recall and precision values in the retrieval system. The following are the results of the calculations and the averages obtained after the calculations:

Table 3. Recall and Precision Calculations

	No Relevant (a)	Not Relevant (b)	Total a+b	Not Found (c)	Total (a+c)	Recall a/(a+c)	Precision a/(a+b)
1	205	4	209	0	205	1,00	0,980861244
2	11	0	11	0	11	1,00	0,1
3	650	2	652	0	650	1,00	0,9969325153
4	15	1	16	0	15	1,00	0,9375
5	32	2	34	0	32	1,00	0,9411764706
Average						1,00	0,971294046
6	205	8351	8556	0	205	1,00	0,239597943
7	11	729	740	0	11	1,00	0,1486486486
8	650	14550	15200	0	650	1,00	0,4276315789
9	15	1	16	0	15	1,00	0,9375
10	32	2	34	0	32	1,00	0,9411764706
Average						1,00	0,3920528575

Source: Data Result, 2023

After obtaining the recall and precision values with the average of the two types of searches, we proceed with the F-measure calculation as follows:

$$F_1 = 2 \times \frac{97,13 \times 100}{97,13 + 100} = 0,9854$$

$$F_2 = 2 \times \frac{39,2 \times 100}{39,2 + 100} = 0,5632$$

Based on the collected data and performed calculations, the average recall value for the search method with quotes and without quotes is the same, which is equal to 1 or 100%. The precision value using quotation marks equals 0.9713, while for searches without quotation marks, it equals 0.3920. According to these findings, the precision value is lower than the recall value. However, the precision value for searches without quotation marks is deficient compared to the precision values for search results with quotation marks; this indicates that searches with quotation marks provide more specific search results with relatively high relevance.

It shows that the precision value resulting from a search using quotation marks worth 0.9713 is in the range of 0.68-1.00, categorized as high precision, meaning that information retrieval is said to be effective. However, it differs from the precision value resulting from a search without quotation marks, which is worth 0.3920, which is in the range of 0.34-0.67, which is categorized as medium precision, meaning that information retrieval is less effective.

Based on a statement from Bama & Ahmed, (2015): "F1 score reaches its best value at 1 and worst score at 0," means that the F1 value of the search results uses quotation marks which results in 0.98, and the value is close to 1. Meanwhile, the F2 number from search results without quotation marks shows the number 0.56, and the value means that it is still far from the number 1 and tends to be close to the middle. It means that the relevant information

retrieval system in the CLAS system tends to be effective and is good enough when using quotation marks.

From a technical standpoint, the key distinction lies in the presence or absence of quotation marks in the search keywords. If quotation marks are included, the search engine will specifically look for documents containing the words in the exact sequence entered. This means that regardless of the number of words entered, the engine will prioritize finding documents with the specified sequential order. Conversely, without quotation marks, the engine will search for documents containing the individual words, regardless of their consecutive arrangement. This distinction significantly impacts the relevance of the documents presented in search results and their alignment with user needs. The use of quotation marks ensures a more stringent and precise search, while their absence allows for a broader search that may encompass documents with the specified words in varied sequences.

Discussion

Based on the calculation results, searching with quotation marks is more effective than searching without quotation marks. From the user's point of view, this EDMS system can effectively meet user needs. However, there are special notes, namely writing the right keywords, for example, using quotation marks so that search results are correct. The next question is how exactly a system produces search results, especially with the differences seen from the results of this study where the value of Precision and F-measure with keywords using quotation marks as big as to produce a value that is far greater than the value of Precision and F-measure without quotes.

The search engine has settings for showing search results from the stored database, as seen in the search results list. The search engine brings up documents where one of the words contained in the keywords is in the title. EDMS system owners can consider how to regulate how the machine should be able to display search results according to user needs because, in this case, it is a bridge between users and the data they need.

4. CONCLUSION

The calculation is done by conducting search experiments for several keywords and then finding the number of documents, relevant documents, and the total number of existing documents. The Recall precision and F-measure formulas are then applied to the calculations. It is known that the recall value is at number 1, meaning that all documents contained in the database and related to keywords have been raised. The precision value of 0.9713 is in the range of 0.68 – 1.00, which is categorized as high Precision, meaning that information retrieval is effective. The F-measure value is 0.98, close to 1, so the tracing system is considered good. Based on these results, it can be concluded that CLAS is quite effective in retrieving information per user needs.

The keywords entered when using quotation marks without quotation marks have a significant difference in results, so this provides a new perspective for EDMS application owners to be able to consider search methods and settings in selecting search results for future EDMS application development.

This study intends to evaluate search engines, but related to measurement, it is only limited to Recall, Precision, and F-measure as a reference. Based on previously entered documents, 5 keywords are randomly chosen. In addition to the studies already mentioned, numerous other studies measuring the effectiveness of search engines can be considered for future research. These studies include calculating accuracy, sensitivity, specificity, and other metrics. In the future, it is necessary to pay more attention to the search method used because

we can also add Boolean methods or advanced search to calculate effectiveness to be more accurate.

Acknowledgment

I express my thanks to Diah Lutifiani from the Graduate Student Library of Information Science, Universitas Indonesia, for being a partner in this research in peer review and discussion about framework and calculation.

REFERENCES

- Alshibly, H., Chiong, R., & Bao, Y. (2016). Investigating the Critical Success Factors for Implementing Electronic Document Management Systems in Governments: Evidence From Jordan. *Information Systems Management*, 33(4), 287–301. <https://doi.org/10.1080/10580530.2016.1220213>
- Bama, Ss., & Ahmed, Msi. (2015). A Survey on Performance Evaluation Measures for Information Retrieval System. *International Research Journal of Engineering and Technology*. www.irjet.net
- Bramer, W. M., Giustini, D., & Kramer, B. M. R. (2016). Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: A prospective study. *Systematic Reviews*, 5(1). <https://doi.org/10.1186/s13643-016-0215-7>
- Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., & Quarteroni, S. (2013). An Introduction to Information Retrieval. In *Web Information Retrieval* (pp. 3–11). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39314-3_1
- Faila Saufa, A., & Wahyu, J. (2017). Evaluasi Sistem Temu Kembali Informasi Koha di Perpustakaan Universitas Muhammadiyah Surakarta (UMS). *Khizanah Al-Hikmah: Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 5(2), 140–151. <https://doi.org/10.24252/kah.v5i2a1>
- Goker, A., & Davies, J. (2009). *Information Retrieval: Searching in the 21st century*. Laserwords Private Ltd.
- Grossman, D. A., & Frieder, O. (2004). *Information Retrieval*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-3005-5>
- Koehrsen, W. (2023, March 8). *Precision and Recall: How to Evaluate Your Classification Model*. BuiltIn.Com.
- Lancaster, F. W. (1979). *Information Retrieval Systems: Characteristics, Testing, and Evaluation* (2nd ed.). John Wiley.
- Manning, C. D., Raghavan, Prabhakar., & Schütze, Hinrich. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Martin, :, & Nilawati, L. (2019). Recall dan Precision Pada Sistem Temu Kembali Informasi Online Public Access Catalogue (OPAC) di Perpustakaan. *Paradigma: Jurnal Komputer Dan Informatika Universitas Bina Sarana Informatika*, 21(1), 77–84. <https://doi.org/10.31294/p.v20i2>
- Miao, J., & Zhu, W. (2022). Precision–recall curve (PRC) classification trees. *Evolutionary Intelligence*, 15(3), 1545–1569. <https://doi.org/10.1007/s12065-021-00565-2>
- Musvida, E. (2017). *Recall Dan Precision pada Sistem Temu Kembali Informasi Bidang Ilmu Fikih di Perpustakaan UIN Ar-Raniry Banda Aceh*.
- Song, J. J. S., Lee, W., & Afshar, J. (2019). An effective High Recall Retrieval method. *Data and Knowledge Engineering*, 123. <https://doi.org/10.1016/j.datak.2017.07.006>
- Sukma, A., Zaman, B., & Purwanti, E. (2015). Klasifikasi Dokumen Temu Kembali Informasi dengan K-Nearest Neighbour Information Retrieval Document Classified with K-Nearest Neighbor. In *RECORD AND LIBRARY JOURNAL* (Vol. 1, Issue 2).

Zhang, E., & Zhang, Y. (2018). F-Measure. In: Liu,L., Ozsu,M.T. In L. Liu & M. T. Ozsu (Eds.), *Encyclopedia of Database System*. Springer. https://doi.org/https://remote-lib.ui.ac.id:2075/10.1007/978-1-4614-8625-9_483