

HASIL

CEK_60040497_Implementation of Jaccard Coefficient Method

by Dewi Soyusiawaty Dewi Soyusiawaty_indriyani Putri Utami

Submission date: 09-Mar-2022 08:07AM (UTC+0700)

Submission ID: 1779839955

File name: soyusiawaty-2020-ijca-920954_1.pdf (1.02M)

Word count: 4185

Character count: 19741

Implementation of Jaccard Coefficient Method for Searching Report Findings of Internal Quality Audit in Ahmad Dahlan University

Dewi Soyusiawaty
Department of Informatics Engineering
State Ahmad Dahlan University
Yogyakarta, Indonesia

Indriyani Putri Utami
Department of Informatics Engineering
State Ahmad Dahlan University
Yogyakarta, Indonesia

ABSTRACT

Currently, the implementation of the Internal Quality Audit (AMI) at Ahmad Dahlan University has been facilitated by the Quality Assurance Support System (QASS) of Ahmad Dahlan University. However, the implementation still takes a long time because the search for the results of AMI findings are still done manually by reading one by one the hardcopy report on the implementation of the previous AMI period. Quality Assurance Agency (BPM) still finds difficulties to carry out the audit process. Therefore, BPM requires AMI findings report searching system based on the finding criteria that the user wants. The purpose of this paper is to implement the Jaccard Coefficient method in the AMI findings report search system at Ahmad Dahlan University. Based on the research conducted by the writers, the AMI findings report search system using the Jaccard Coefficient method according to the keywords by the user can ease in finding the AMI Findings report. The test results in this study obtained recall value 1 which states that the system can find the relevant AMI findings and the average precision value of 0.46 indicates that there are still other documents besides the relevant documents found (retrieved) by the system

General Terms

Jaccard Coefficient, Information Retrieval, Searching System

Keywords

AMI Finding Reports, Internal Quality Audit, Information Retrieval, Jaccard Coefficient, Searching System

1. INTRODUCTION

The quality of higher education is the harmony between the implementation of higher education and the National Education Standards and the standards set by higher education institutions themselves based on the vision-mission and the needs of stakeholders [1]. The higher education quality assurance system consists of an Internal Quality Assurance System (SPMI) and an External Quality Assurance System (SPME). Ahmad Dahlan University (UAD), as the Best Private University in Yogyakarta and Central Java based on the version of Webometrics July 2016 and that has excellent academic quality, participated in the development of higher education quality in Indonesia including the dissemination of the implementation of the university's internal quality audit. The implementation of the Internal Quality Audit (AMI) in each unit for quality assurance assessment is held twice a year, which is controlled by the Quality Assurance Agency (BPM) of UAD. AMI aims to maintain mutualism and to pay more attention to small things that may get less attention. Besides, AMI is also a requirement for certification of quality assurance implementation in UAD to observe its activities in

the field.

At this time, the implementation of AMI at Ahmad Dahlan University has been facilitated by the existence of Ahmad Dahlan University Quality Assurance Support System (QASS). This information system is addressed to the auditor in entering findings data until making AMI implementation reports. Even so, this activity still takes a long time in its implementation because the search for AMI findings is still done manually by reading one by one the audit report in the form of hardcopy on the implementation of the previous AMI period.

Based on the results of interviews with the Auditor, the BPM conducts an internal quality audit process in 55 units each period consisting of Faculties, Study Programs D4, S1, S2, and Post Graduate. Reading the reports of AMI findings one by one in each unit in the form of hardcopy made it difficult for the BPM to carry out the audit process. Therefore, it is common to make mistakes in reading the findings during the audit process.

Jaccard Similarity or Jaccard Coefficient is an algorithm that is usually used to compare documents based on the words they have and calculate the similarity value [2]. Unlike other similarity methods, the Jaccard Coefficient does not consider how many times a term appears in a document. This method also uses slices between documents and queries.

2. METHOD

2.1. Information Retrieval

Retrieval of information stored in database can be implemented using natural language. Several techniques and approaches used and developed with the aim that computers are better able to measure / demand humans use natural language [3]. Information Retrieval aims to find unstructured documents that meet the information needs in an extensive collection of documents [4]. Information Retrieval System is a system for representing, storing, organizing, and processing information [5]. The essence of information retrieval is a system that automatically finds information according to user needs from a collection of information.

Technically, the purpose of the information retrieval system is to match the terms that are built (queries) with the terms or indexes in the document, so that the relevant documents will be retrieved from the database. The relevant documents taken are the results of the information retrieval system. The objective of the Information Retrieval System is the retrieval of documents based on the user's request, so the content of the document taken is relevant to the information that the user needs.

It needs to be noted that searching for information in an information retrieval system does not necessarily return all relevant documents. It can be only partial or not at all. The information retrieval system may not provide any results if the relevant documents are not found [6].

2.2 Internal Quality Assurance System

The higher education institution system must run a quality assurance system, especially the Internal Quality Assurance System (SPMI), which is developed following the characteristics and vision-mission of each higher education institution. The implementation of SPMI aims to guarantee the National Standards of Higher Education (SN Dikti) held by each higher education institution through the implementation of the Tridharma of Higher Education in realizing the vision and meeting the needs of higher education institution stakeholders, both internal and external. SPMI is also a preparation for each higher education institution to implement an External Quality Assurance System (SPME), which will determine the accreditation of the higher education institution.

As one of the subsystems of the Higher Education Quality Assurance System (SPM-PT), internal quality assurance that has been implemented by higher education institutions in Indonesia needs to be evaluated in the term of its success in continuously improving the quality of higher education [7].

For AMI implementation, the quality assurance management unit needs to develop the required audit instruments. The instruments are then used by internal auditors who have received SPMI and AMI training. The results of AMI are then followed up by controlling the implementation of standards and improving the standards continuously.

2.3. Jaccard Coefficient

Jaccard Similarity or well known as the Jaccard Coefficient, is a method to measure the level of similarity (Similarity) between two documents [8]. In this method, the parameter is

the number of words in the document that will be compared to determine the similarity [9]. In this study, the parameters compared were the words in the query and documents in the database.

Measuring the Jaccard Coefficient between two datasets is the result of dividing the same amount of data from the two datasets by the sum of all data in the dataset as the following formula:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$J(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

The length of words strongly influences the Jaccard Coefficient method in each document. This method does not consider which terms rarely and which terms often appear in documents. Therefore, this method utilizes the Term Frequency Binary in determining the length of words in the document.

Normalization of the length of the dividing vector becomes an alternative in calculating the value of the Jaccard Coefficient. The following is the Jaccard Coefficient formula with long normalization.

$$J(A, B) = \frac{|A \cap B|}{\sqrt{(|A| + |B| - |A \cap B|)}} \quad (3)$$

2.4 Recall and Precision

Recall is the proportion of the number of documents that the system can retrieve. Precision is the proportion of the number of documents found and considered relevant to the needs of the user. The following Fig 1 is a representation of the Recall and Precision set along with the calculation formula.

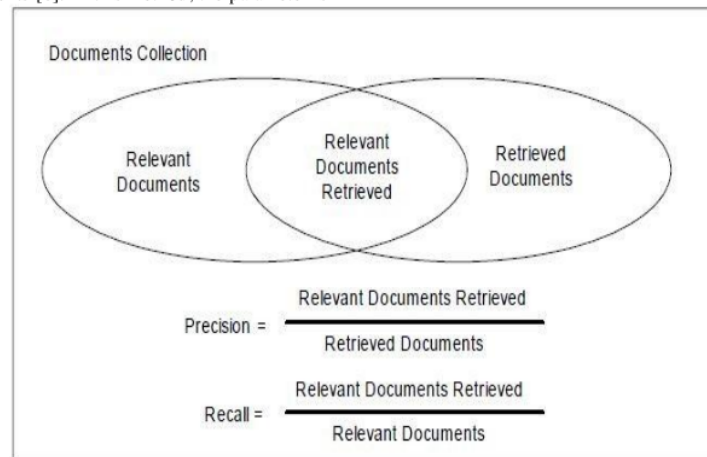


Fig.1. Representation of the Recall Set and Precision [10]

The ideal condition of the information retrieval system's capability is if the ratio between recall and precision is equal or 1: 1 [11]. However, the ratio of recall is difficult to measure because the number of all relevant documents in the database is enormous. Therefore, precision is one of the benchmarks used to assess the effectiveness of an information retrieval

system [12].

3. ALGORITHM

The stages of system development applied in this study begin with input in the form of load text data. Then the text data will be processed through the preprocessing stage, which consists

of tokenization, stopword (word filtering), and stemming (return of root words). After that, the stemming data results are entered in the calculation process of the Coefficient as

illustrated by the flowchart in Fig. 2. Flowchart of System Development Stages.

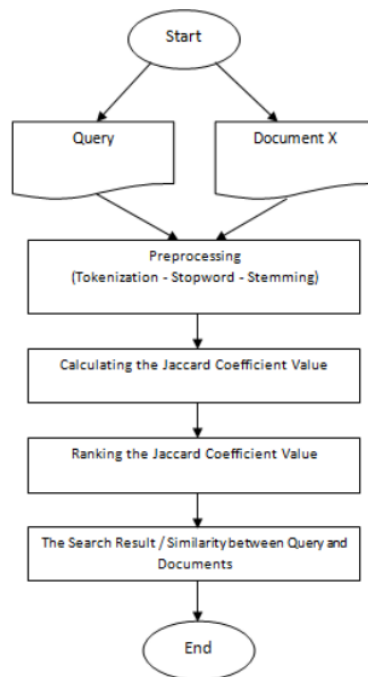


Fig. 2 : Flowchart of System Development Stages

3.1 Collecting Data

The data collected is a report on AMI findings in the form of excel files obtained through Ahmad Dahlan University's Quality Assurance Support System (QASS).

3.2 Preprocessing

a) Tokenizing

This stage is done by detecting the existence of spaces and punctuation that separate one word from another so that it produces a stand-alone token. The example is "Pengukuran kesesuaian bidang kerja belum dilakukan secara periodik (The measurement of field suitability has not been done periodically)" which is broken down into the form of tokens: Pengukuran, kesesuaian, bidang, kerja, belum, dilakukan, secara, periodik.

b) Stopword

In the stopwords stage, the words (token) are experiencing the process of removing words that are considered to have no meaning or non-essential words such as conjunctions. The example is "Pengukuran kesesuaian bidang kerja belum dilakukan secara periodik" to be: pengukuran, kesesuaian, bidang, kerja, periodik, dilakukan

c) Stemming

After the stopwords process, the next step is the normalization of words to restore the words to their root word [13]. The example is "Pengukuran (Measurement)" to ukur (measure).

3.3 TF Binary

In the Binary TF stage, in the queries and documents that have been preprocessed, it is determined whether or not there is a term in the query or document. If it exists, it will be a value of 1, while if the term is not found in the query or the document, it will be a value of 0.

3.4 Database Design

The database used in this program is MongoDB 3.4. Database design is used to determine what data will be stored in the database (database) and used to support various system designs. The data includes:

- Data file, for storing uploaded files.
- Findings data, to save every AMI finding to be processed.
- Tokenization data, to store the tokenization results of the findings data.
- Stopword data, to save stopwords results taken from tokenization data.
- Data stemming, to store stemming results from stopwords data.

4. RESULT AND DISCUSSION

Implementation of the Jaccard Coefficient Method for Searching Reports on the Findings of Internal Quality Audit Results at Ahmad Dahlan University used the Python programming language, combined with the Flask Framework. Preprocessing process in this study utilizes Literature libraries.

The data taken was the reports of AMI findings. Each finding

was considered as one document, which was then given a document ID, as shown in Table 1. Data Collection.

Table 1. Data Collection

ID	Findings
D0	Visi Misi tertulis namun belum mengarah ke tujuan dan sasaran (<i>output</i> prodi) sehingga belum terukur.
D1	Tidak ditemukan dokumen sasaran mutu yang realistik.
D2	Belum ada matriks keterkaitan antara visi misi tujuan dan sasaran prodi
D3	Tidak ditemukan pengukuran sasaran mutu di prodi BISMA
D4	Tidak ditemukan dokumen sasaran mutu yang resmi dan terukur

D5	Instrumen pengukuran pemahaman visi misi tidak ditemukan
D6	Di prodi BISMA tidak ditemukan prestasi di tingkat internasional
D7	Tidak ditemukan publikasi dosen yang <i>published</i> dalam bentuk apapun.
D8	IKD di Prodi Bisma tidak muncul dengan benar
D9	Tidak ditemukan struktur organisasi prodi BISMA yang resmi

The findings document was calculated based on input queries to be searched. The query on the manual calculation using the Jaccard Coefficient method was "Sasaran Mutu Prodi (Study Program Quality Targets)". The findings and query documents were preprocessed until the TF Binary process could be carried out as in Table 2.

Table 2. TF Binary

Term	TF										
	Q	D0	D1	D2	D3	D4	D5	D6	D7	D8	D9
visi	0	1	0	1	0	0	1	0	0	0	0
misi	0	1	0	1	0	0	1	0	0	0	0
tulis	0	1	0	0	0	0	0	0	0	0	0
arah	0	1	0	0	0	0	0	0	0	0	0
tuju	0	1	0	1	0	0	0	0	0	0	0
sasar	1	1	1	1	1	1	0	0	0	0	0
output	0	1	0	0	0	0	0	0	0	0	0
prodi	1	1	0	1	1	0	0	1	0	1	1
ukur	0	1	0	0	1	1	1	0	0	0	0
temu	0	0	1	0	1	1	1	1	1	0	1
dokumen	0	0	1	0	0	1	0	0	0	0	0
mutu	1	0	1	0	1	1	0	0	0	0	0
realistik	0	0	1	0	0	0	0	0	0	0	0
ada	0	0	0	1	0	0	0	0	0	0	0
matriks	0	0	0	1	0	0	0	0	0	0	0
Kait	0	0	0	1	0	0	0	0	0	0	0
tuju	0	0	0	0	0	0	0	0	0	0	0
bisma	0	0	0	0	1	1	0	1	0	1	1
resmi	0	0	0	0	0	0	0	0	0	0	1
instrumen	0	0	0	0	0	0	1	0	0	0	0
paham	0	0	0	0	0	0	1	0	0	0	0
prestasi	0	0	0	0	0	0	0	1	0	0	0
tingkat	0	0	0	0	0	0	0	1	0	0	0
internasional	0	0	0	0	0	0	0	1	0	0	0
publikasi	0	0	0	0	0	0	0	0	1	0	0
dosen	0	0	0	0	0	0	0	0	1	0	0
published	0	0	0	0	0	0	0	0	1	0	0

bentuk	0	0	0	0	0	0	0	0	1	0	0
apa	0	0	0	0	0	0	0	0	1	0	0
ikd	0	0	0	0	0	0	0	0	0	1	0
muncul	0	0	0	0	0	0	0	0	0	1	0
benar	0	0	0	0	0	0	0	0	0	1	0
struktur	0	0	0	0	0	0	0	0	0	0	1
organisasi	0	0	0	0	0	0	0	0	0	0	1
TOTAL	3	9	5	8	6	6	6	6	6	5	6
$ID_n \cap Q$	-	2	2	2	3	2	0	1	0	1	1

As the rules of TF-binary which states that TF only uses binary numbers, namely 1 and 0, table 2 TF-Binary explains that the term found in a document is 1 while the term not found in the document is 0.

In table 2, the TF Binary also explains the slices ($ID_n \cap Q$) or the number of queries found in a document that will be used in the process of calculating the Jaccard Coefficient value. The following is the calculation of the Jaccard Coefficient value between queries and documents calculated according to the Jaccard Coefficient formula in formula (3).

1. In D_0 it is found *query* ($ID_0 \cap Q$) = {sasar, prodi} = 2

$$J(Q, D0) = \frac{|Q \cap D0|}{\sqrt{|Q|+|D0|-|Q \cap D0|}} = \frac{|2|}{\sqrt{|3|+|9|-|2|}} = 0,632456$$

2. In D_1 it is found *query* ($ID_1 \cap Q$) = {sasar, mutu} = 2

$$J(Q, D1) = \frac{|Q \cap D1|}{\sqrt{|Q|+|D1|-|Q \cap D1|}} = \frac{|2|}{\sqrt{|3|+|5|-|2|}} = 0,816497$$

3. In D_2 it is found *query* ($ID_2 \cap Q$) = {sasar, prodi} = 2

$$J(Q, D2) = \frac{|Q \cap D2|}{\sqrt{|Q|+|D2|-|Q \cap D2|}} = \frac{|2|}{\sqrt{|3|+|8|-|2|}} = 0,666667$$

4. In D_3 it is found *query* ($ID_3 \cap Q$) = {sasar, prodi, mutu} = 3

$$J(Q, D3) = \frac{|Q \cap D3|}{\sqrt{|Q|+|D3|-|Q \cap D3|}} = \frac{|3|}{\sqrt{|3|+|6|-|3|}} = 1,224745$$

5. In D_4 it is found *query* ($ID_4 \cap Q$) = {sasar, prodi} = 2

$$J(Q, D4) = \frac{|Q \cap D4|}{\sqrt{|Q|+|D4|-|Q \cap D4|}} = \frac{|2|}{\sqrt{|3|+|6|-|2|}} = 0,755929$$

6. In D_5 it is found *query* ($ID_5 \cap Q$) = { } = 0

$$J(Q, D5) = \frac{|Q \cap D5|}{\sqrt{|Q|+|D5|-|Q \cap D5|}} = \frac{|0|}{\sqrt{|3|+|6|-|0|}} = 0$$

7. In D_6 it is found *query* ($ID_6 \cap Q$) = {prodi} = 1

$$J(Q, D6) = \frac{|Q \cap D6|}{\sqrt{|Q|+|D6|-|Q \cap D6|}} = \frac{|1|}{\sqrt{|3|+|6|-|1|}} = 0,353553$$

8. In D_7 it is found *query* ($ID_7 \cap Q$) = { } = 0

$$J(Q, D7) = \frac{|Q \cap D7|}{\sqrt{|Q|+|D7|-|Q \cap D7|}} = \frac{|0|}{\sqrt{|3|+|6|-|0|}} = 0$$

9. In D_8 it is found *query* ($ID_8 \cap Q$) = {prodi} = 1

$$J(Q, D8) = \frac{|Q \cap D8|}{\sqrt{|Q|+|D8|-|Q \cap D8|}} = \frac{|1|}{\sqrt{|3|+|5|-|1|}} = 0,377964$$

10. In D_9 it is found *query* ($ID_9 \cap Q$) = {prodi} = 1

$$J(Q, D9) = \frac{|Q \cap D9|}{\sqrt{|Q|+|D9|-|Q \cap D9|}} = \frac{|1|}{\sqrt{|3|+|6|-|1|}} = 0,353553$$

Based on the results of the calculation above, the Jaccard coefficient value for each document will be ranked, as shown in Table 3. Jaccard Coefficient Rank.

Table 3. Jaccard Coefficient Rank

Dn	$J(D_n, Q)$ normalization	Dn	Ranking
D0	0,632456	D3	1,224745
D1	0,816497	D1	0,816497
D2	0,666667	D4	0,755929
D3	1,224745	D2	0,666667
D4	0,755929	D0	0,632456
D5	0	D8	0,377964
D6	0,353553	D6	0,353553
D7	0	D9	0,353553
D8	0,377964	D5	0
D9	0,353553	D7	0

Table 3 shows that D3 has the most significant Jaccard value coefficient, followed by documents D1, D4, D2, D0, D8, D6, and D9. This means that the D3 document has a significant degree of similarity to the query entered compared to other documents.

4.1 Implementation

In fig. 3 the Preprocessing page shows the preprocessing stages, which consist of Tokenization, Stopword, and Stemming. This process is done after uploading AMI data and will be saved to the database.

	T1	T2	T3	T4	T5	T6	T7	T8	T9
0	visi	misi	tertulis	namun	belum	mengarah	ke	tujuan	dan
1	tidak	ditemukan	dokumen	sesaran	mutu	yang	realistik	None	None
2	belum	ada	matrix	keterkaitan	antara	visi	misi	tujuan	dan
3	tidak	ditemukan	pengukuran	sesaran	mutu	di	prodi	bisma	None
4	tidak	ditemukan	dokumen	sesaran	mutu	yang	resmi	dan	terukur
5	instrument	pengukuran	pemahaman	visi	misi	tidak	ditemukan	None	None
6	di	prodi	bisma	tidak	ditemukan	prestasi	di	tingkat	internasional
7	tidak	ditemukan	publikasi	dan	yang	publikasi	dalam	kehidupan	manusia

Fig. 3 : Preprocessing page

In Fig. 4 Search Results Page of the Jaccard Coefficient shows the search results of AMI findings by implementing the Jaccard Coefficient method. The search results consist of the Index column, Study Program, Findings, and Score of the

Jaccard Coefficient calculation. The most significant calculation results will be in the top row because the similarity value with the inputted query is high and vice versa with the lowest similarity value.

	Prodi	Temuan	Score
491	S2 Pendidikan Vokasi	Tidak ada sasaran mutu prodi	1.732051
98	S1 Teknologi Pangan	Sasaran mutu prodi ada 2 versi	1.341641
298	S1 Pendidikan Agama Islam	prodi belum melakukan pengukuran sasaran mutu	1.341641
411	S1 Bahasa dan Sastra Arab	Sudah ada Sasaran Mutu Prodi tetapi capaiannya belum diukur	1.341641
154	S1 Sistem	tidak ada dokumentasi tentang capaian sasaran mutu prodi	1.341641

Fig. 4 : Jaccard Coefficient Search Results Page

4.2 Testing System

In this test, 545 findings of documents in the database were used. Testing was done using 32 different queries on each test. Recall and Precision Test Results with the calculation formula as follows:

$$\text{Recall} = \frac{R_a}{R_s} = \frac{27}{27} = 1$$

$$\text{Precision} = \frac{R_a}{R_t} = \frac{27}{27} = 1$$

Explanation :

Rt: Number of documents retrieved (found)

Rs: Number of relevant documents in the collection

Ra: Number of relevant retrieved documents

Table 4. Jaccard Coefficient Calculation

No	Keyword	Term	Rt	Rs	Ra	Recall	Precision
1	Kurikulum	1	27	27	27	1	1
2	Akademik	1	15	15	15	1	1
3	Mutu	1	73	73	73	1	1
4	Monev	1	18	18	18	1	1
5	Visi	1	53	53	53	1	1
6	Misi	1	48	48	48	1	1
7	Penelitian	1	14	14	14	1	1
8	Kerjasama	1	16	16	16	1	1
9	Matriks	1	8	8	8	1	1
10	Lulusan	1	47	36	36	1	0,766
11	Persentase Lulusan	2	51	4	4	1	0,078
12	Publikasi Ilmiah	2	19	4	4	1	0,21
13	Notulensi Monev	2	19	1	1	1	0,053
14	Deskripsi Jabatan	2	7	4	4	1	0,571
15	Umpan Balik	2	28	28	28	1	1
16	Visi Misi	2	55	46	46	1	0,836
17	Peta Kurikulum	2	27	7	7	1	0,259
18	IPK Rata-rata	3	25	4	4	1	0,16
19	Kelulusan Tepat Waktu	3	44	19	19	1	0,432
20	Sasaran Mutu Prodi	3	147	19	19	1	0,129
21	Dosen Pembimbing Akademik	3	78	6	6	1	0,077
22	Umpan Balik Dosen	3	84	7	7	1	0,083
23	Pengukuran Sasaran Mutu	3	121	17	17	1	0,14
24	Implementasi Jenis Kerjasama	3	22	3	3	1	0,136
25	Ketercapaian Sasaran Mutu	3	112	12	12	1	0,108
26	Pengukuran Pemahaman Visi Misi	4	99	7	7	1	0,071
27	Visi Misi Tujuan Sasaran	4	112	24	24	1	0,214
28	Dokumen Sasaran Mutu Prodi	4	215	5	5	1	0,023
29	Strategis Pengembangan Program Studi	4	33	2	2	1	0,061
30	Deskripsi Jabatan Kepala Laboratorium	4	9	2	2	1	0,222
31	Rata-rata Dosen Pembimbing Akademik	5	90	4	4	1	0,044
32	Rencana Strategis Pengembangan Program Studi	5	38	2	2	1	0,053
Average						1	0,46

Based on the test results data in Table 3, Recall and Precision testing show that the results of recall 1 in each test prove that the system can find all relevant documents. While the value of precision in testing between 0.023 and 1 shows that there are still other documents besides relevant documents that are also

found by the system.

The test data shows the comparison of the average recall value and the average precision value of 1:0.46 which proves that the average accuracy between the queries inputted with the search results document is still less relevant since the ideal

condition of the recall and precision ratio is 1:1.

Precision values depend on the uniqueness of the query entered. The more unique the query, the higher the value of the precision obtained. On the contrary, the more general the queries given, the lower the precision value.

5. CONCLUSION AND ADVICE

5.1 Conclusion

Based on the results of the study "Implementation of Jaccard Coefficient Method for Searching Report Findings of Internal Quality Audit in Ahmad Dahlan University", the following conclusions can be drawn:

- a) A search system for the findings of the Internal Quality Audit at Ahmad Dahlan University has been developed by implementing the Jaccard Coefficient method using the Python programming language that can search according to the keywords inputted.
- b) The Recall and Precision tests carried out produce recall value of 1 for each test, which shows that the system can find all relevant documents. It produces a precision value of 0.023 which indicates that there are still other documents besides the relevant documents found by the system.
- c) Comparison of the average recall value and average precision value is 1: 0.46 which proves that the average accuracy between the queries inputted with the search results document is still less relevant since the ideal condition of the ratio of recall and precision is 1: 1.

5.2 Advise

Research "The Implementation of the Jaccard Coefficient Method for Searching for Reports on the Findings of Internal Quality Audit at Ahmad Dahlan University" still has shortcomings so that further development is expected. Suggestions that can be used as a reference for further development are as follows

- a) The upload process is done manually with files that are limited to the .xlsx extension. It is expected to be further developed by retrieving data from a website's database.
- b) This system can only know the similarity values on each document regardless of the frequency of the terms found. The system can still be developed using other methods to search the AMI findings report so that it can be used as a comparison.

6. ACKNOWLEDGEMENTS

The researcher would like to thank to Quality Assurance

Agency Ahmad Dahlan University (BPM UAD) for being willing to provide sample data in this research and taking the time to be interviewed in the context of analysis, design and system testing.

7. REFERENCES

- [1] Putra, A. (2014) 'Sistem Audit Mutu Akademik Internal Berbasis Web pada Universitas Sriwijaya', *Prosiding Seminar Nasional Ilmu Komputer*, (19).
- [2] Rinarta, K. *et al.* (2017) 'Simple Query Suggestion Untuk Pencarian Artikel Menggunakan Jaccard Similarity', 3(1), pp. 30–34
- [3] Soyusiawaty, Dewi, and Anna Hendri Soleliza Jones. 2020. "Pemanfaatan Bahasa Alami Penelusuran Informasi Skripsi Melalui Digital Library (Studi." : 1–8.
- [4] Manning, C. D., Raghavan, P. and Schütze, H. (2008) *Introduction to Information Retrieval Introduction, Computational Linguistics*. doi: 10.1162/coli.2009.35.2.307..
- [5] Baeza-Yates, R. and Ribeiro-Neto, B. (1999) 'Modern information retrieval', *New York*, 9, p. 513. doi: 10.1080/14735789709366603.
- [6] Ingwersen, P. (1992) *Information Retrieval Interaction*. London: Taylor Graham.
- [7] Yanti, N. (2005) 'Kasim Riau (Studi Kasus : Lembaga Penjaminan Mutu)'.
- [8] Plansangket, S. and Gan, J. Q. (2015) 'A query suggestion method combining TF-IDF and Jaccard Coefficient for interactive web search', *Artificial Intelligence Research*, 4(2), pp. 119–125. doi: 10.5430/air.v4n2p119.
- [9] Ro'is, M. A. (2018) 'Implementasi Metode Jaccard Similarity Pada Aplikasi Pencarian Lirik Lagu'
- [10] Mandala, R. (2006) 'Evaluasi efektivitas metode', 2006(Snati)
- [11] Pao, M. L. (1989) *Concepts of Information Retrieval*. Englewood, Colorado: Libraries Unlimited
- [12] Lestari, N. P. (2016) 'Uji Recall And Precision Sistem Temu Kembali.
- [13] Soyusiawaty, Dewi, Anna Hendri Soleliza Jones, and Nora Lestari Lestariw. 2020. "The Stemming Application on Affixed Javanese Words by Using Nazief and Adriani Algorithm." IOP Conference Series: Materials Science and Engineering 771(1).

HASIL CEK_60040497_Implementation of Jaccard Coefficient Method

ORIGINALITY REPORT

2%

SIMILARITY INDEX

4%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Liberty University
Student Paper

2%

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%