MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 12, No. 1(2020), pp 10-14 DOI : 10.18860/mat.v12i1.8320
Received : January25th 2020; **Accepted** : March 9th 2020 ; **Avalaible Online** : April 5th 2020

Query Answering System of Shahih Hadith Muttafaqun ‘Alaih Using Indonesian Thesaurus Based on Query Expansion and Naïve Bayes Classifier

Muhammad Fairuz Zumar Rounaqi, Cahyo Crysdian, Roro Inda Melani

Abstract— **Hadith are all the words, deeds and provisions of the Prophet Muhammad SAW that are used as the second of Islamic law after Al-Quran. The purpose of this study is to make an Information Retrieval system called the Query Answering System is expected to facilitate users in searching and finding the hadith documents as the user's needs. This study implements the Naïve Bayes Classifier method combined with Indonesian thesaurus as a query expansion to find the hadith documents that relevant to the input query. Based on the testing of 50 query data, the test results show that the use of query expansion gives better results than without using query expansion. Where based on testing of the top 1 data without using query expansion obtained an average recall value of 62%, an average precision value of 62%, an average accuracy value of 92.4% and an average value of the f-measure of 62%, while testing using query expansion obtained an average recall value of 66%, an average precision value of 66%, an average accuracy value of 93.2% and an average f -measure value of 66%. Based on the test results, the use of query expansion shows an improvement in the average recall value of 4%, an improvement in the average precision value of 4%, and an improvement in the average accuracy value of 0.8% and an improvement in the average f-measure value of 4% compared on without using query expansion.**

***Index Terms*—hadith, information retrieval, query expansion, naïve bayes.**

I. INTRODUCTION

HADITH is the second source of Islamic law after Al-Qur’an. Al-Hafidz Ibnu Hajar Al Asqalany explained in the book of Bulughul Maram that the hadith are all the words, deeds and decrees of the prophet Muhammad SAW which are used as provisions or Islamic law [1]. In general, the Hadith is in line with Al- Qur’an, where the hadith explains the mubham, details the mujmal limits the absolute, specifies the general and elaborates on the laws and the aims of its purpose.

The hadith that can be used as a way of life is hadith that claims it is truth by the Ulama’ of hadith experts, the one type of it is the shahih hadith narrated by Imam Bukhari and Imam Muslim. Shahih hadith is widely used by Ulama to determine the law of certain disputes. One of the books of the collection of authentic hadith written by Imam Bukhari and Imam Muslim is the book of Al-Lu’Lu’ Wal Marjan by Muhammad Fuad Bin Abdul Baqi. To facilitate the users for finding hadith documents from that book and accommodate with the user needs, it is important to develop an Information Retrieval system as a simple data search system where users only posed a query to the system, then the system will look for hadith that match from the database and displays search results to the users.

To find a relevant hadith, it is necessary to measure the similarity between queries and hadith documents. This study implement the naïve bayes method to measure the probability of queries on hadith documents. Then the results of probability testing will be ranked and sorted from the highest probability value, and then the system will displays the hadith data based on the rank results as a feedback from the query posed by the users.

II. RELATED RESEARCH

Ginting & Trinada [2] examines the implementation of the Naïve Bayes Classifier method as a technique for constructing a classification model based on documents contained in libraries. In their research, the Naïve Bayes Classifier method is used to classify several titles and categories of documents that are already in the library database, then the search process is carried out by involving descriptions of each document so that it can display more references as search results. The results of this study, the information retrieval system that has been made can display more documents based on user queries so that users get more document references. However, this research does not measure the amount of recall and precision of the documents retrieved, so it is not known how much the percentage of relevant documents to the

Manuscript received - Desember, 2019. This work was supported in part by Computer Science Department of Maulana Malik Ibrahim Islamic State University.

Muhammad Fairuz Zumar Rounaqi, Cahyo Crysdian, Roro Inda Melani are the researcher of Computer Science Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia

user query.

Yadav[3] applies the documents similarity matrix and Naïve Bayes classification to do web information retrieval. In this research, the document-document similarity matrix was used after performing pre-processing and feature construction. Then, a Naïve Bayes Classifier was used to find the relevant category of information that is required for the user. The purpose of this research is to analyze the proposed algorithm with sensitivity, specificity, and accuracy. The result of this research is the proposed algorithm can increase 3% of sensitivity as compared with the existing algorithm. Also, it increased the specificity and accuracy as compared with the existing algorithm.

Shahabadkar, *et al*. [4] proposes a new Enhanced Query Expansion based Classifier (EQC) technique for web document retrieval. Where expansion of demand is used to improve document retrieval. Original user requests are reformulated and feedback is given to the dataset to find relevant documents. The results of this study showed an increase in the value of precision by 1%, an increase in the value of recall by 3% and an increase in the value of the f-measure by 3.5%. That is measurements of precision, recall, and f-measure show an increase in document retrieval schemes relevant to reduced computational complexity.

III. Naïve Bayes Classifier

This study proposed the Naïve Bayes Classifier Method. That method will implemented on the searching process to found the relevant hadith. According to Ginting & Trinada [2], Naïve Bayes Classifier is one of machine learning method that utilizes the probability and statistical calculations proposed by Thomas Bayes. According to Suyanto [5], probability or conditional probability is expressed as comparison 1:

$$P(H|X) = P(X|H) * P(H) / p(X) \qquad (1)$$

Where X is proof, H is the hypothesis, P(H|X) is the probability that the hypothesis H is true for the proof of X in other words P(H|X) is the posterior probability H with terms X, P(H|X) is the probability that the proof of X is true for the hypothesis H or probability X with the condition H, P(H) is the probability of the prior hypothesis H, and P(X) is the probability of prior proof of X.

IV. Research Methodology

A. Data Collection

The data used in this study was taken from the book Al Lu 'Lu' Wal Marjan which contains a collection of Sahih hadith narrated by Imam Bukhari and Imam Muslim by Muhammad Fuad Bin Abdul Baqi. The hadith topics taken from the book has the follow Indonesian term such as “*Iman, Thaharah, Haidh, Shalat, Mushala, Shalat Orang Musafir, Al-Jum'ah, Shalat Dua Hari Raya, Shalat Istisqaa ', Salat Gerhana, Janazah*”.

B. Process Design

The process design purposes to create a flow system that would be implemented to this study, an overview of the process design shown in Figure 1.

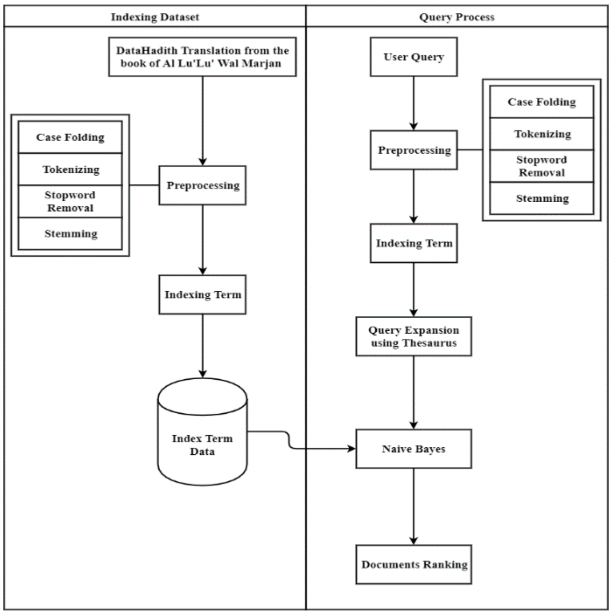

Figure 1. Design Process

Based on Figure 4.1, the searching process started from preprocessing query posed by the users, the preprocessing steps are contains case folding process, tokenizing process, stopword removal process, and stemming. After that, the terms of query as the results of preprocessing will be expanded based on Indonesian thesaurus, then the system will match the terms of query with the terms of documents. The result of matching terms will be build into the training set, it used to facilitate the system for calculate the probability value. Then the results of probability calculation will be ranked and showed to the users as the result of searching process.

C. Experiment

The experiment done by evaluating the results of testing. According to Hasugian [6], there are two things that can be used as a reference assessment to measure the effectiveness of an information retrieval, that is precision and recall. The confusion matrix of recall, precision, and accuracy is given by Table 1.

Table 1. Confusion matrix of recall, precision, and accuracy

	Relevant	Irrelevant	Total
Retrieve	a (*hits*)	b (*noise*)	a + b
Not Retrieve	c (*misses*)	d (*rejected*)	c + d
Total	a + c	b + d	a + b + c + d

Description of Table above:

Recall = [a/ (a+c)] x 100% (2)

Precision = [a/ (a+b)] x 100% (3)

Accuracy = [a+d/(a+b+c+d)] x 100% (4)

To evaluate the information retrieval system also necessary to measure the f-measure. Where the f-measure calculation is used to evaluate the information retrieval system by combining the results of precision and recall calculations. The F-measure represents the relative effect between precision and recall. F-measure is the mean harmonic weight of precision and recall. Below is the formula of the f-measure calculation.

$$\text{F-Measure} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \quad (5)$$

V. Results and Discussion

The first test is system testing that has been done by posting the 50 queries into the system, and the system provides feedback in the form of relevant hadith documents. The 50 query testing is given by Table 2.

Table 2. List of query test

Table of Query Test		
No	Query in Indonesian term	Query expansion in Indonesian term
1	*Rukun Iman*	*Rukun Iman Tiang Religiositas*
2	*Rukun Islam*	*Rukun Islam Tiang*
3	*Menghormati Tetangga*	*Hormat Tetangga Segan Jiran*
4	*Tanda Orang Munafik*	*Tanda Munafik Ciri Munafiq*
5	*Meratakan Saf Saat Shalat*	*Saf Shalat Jajar Salat*
6	*Tanda Orang Beriman*	*Tanda Iman Ciri Religiositas*
7	*Hukum Bunuh Diri*	*Hukum Bunuh Atur*
8	*Syafaat Nabi*	*Syafaat Nabi Mediasi*
9	*Ahli Surga*	*Ahli Surga Kaum Janah*
10	*Ahli Neraka*	*Ahli Neraka Kaum*
11	*Anjuran Bersuci Sebelum Shalat*	*Suci Shalat Murni Salat*
12	*Anjuran Bersiwak*	*Siwak*
13	*Hukum Jilatan Anjing*	*Hukum Jilat Anjing Atur*
14	*Hukum Kencing Bayi*	*Hukum Kencing Bayi Atur Ompol*
15	*Hukum Menggauli Istri Saat Sedang Haidh*	*Hukum Gaul Istri Haidh Atur*
16	*Tata Cara Mandi Janabat*	*Tata Mandi Janabat Siram Junub*
17	*Menjaga Aurat*	*Jaga Aurat Tutup*
18	*Cara Tayammum*	*Tayammum*
19	*Fadilat Shalat Jama'ah*	*Fadilat Shalat Jamaah Rahmat Salat*
20	*Larangan Kencing dalam Air yang Menggenang*	*Larang Kencing Air Genang Ompol*
21	*Larangan Berjalan didepan Orang Shalat*	*Larang Jalan Shalat Salat*
22	*Dosa Besar*	*Dosa Maksiat*
23	*Bacaan dalam Shalat*	*Baca Shalat Salat*
24	*Kewajiban Beriman kepada Allah*	*Wajib Iman Allah Kudu Religiositas*
25	*Kewajiban Beriman kepada Rasulullah*	*Wajib Iman Rasulullah Kudu Religiositas*
26	*Adab Buang Air*	*Adab Buang Air Akhlak*
27	*Cara Membersihkan Kulit Bangkai*	*Bersih Kulit Bangkai Suci*
28	*Amal yang Utama*	*Amal Utama Kebaji Baik*
29	*Mengangkat Kedua Tangan pada saat Takbiratul Ihram*	*Angkat Tangan Takbiratul Ihram Naik*
30	*Cara Berwudhu*	*Wudhu Suci*
31	*Larangan Bicara Ketika Shalat*	*Larang Bicara Shalat Salat*
32	*Bacaan Dalam Ruku' dan Sujud*	*Baca Ruku Sujud Sembah*
33	*Shalatnya Orang Musafir*	*Shalat Musafir Salat*
34	*Fadilat Menghafal Qur'an*	*Fadilat Hafal Quran Rahmat*
35	*Shalat Sunnat*	*Shalat Sunnat Salat*
36	*Shalat Dua Hari Raya*	*Shalat Raya Salat*
37	*Shalat Malam*	*Shalat Malam Salat*
38	*Menjamak Shalat*	*Jamak Shalat Salat*
39	*Cara Sujud Tilawah*	*Sujut Tilawah Sembah*
40	*Perubahan Arah Qiblat*	*Rubah Arah Ganti*
41	*Shalat Sunnat yang dilarang*	*Shalat Sunnat Larang Salat*
42	*Hukum Shalat Memakai Sepatu*	*Hukum Shalat Sepatu Atur Salat*
43	*Imam Shalat*	*Imam Shalat Pimpin Salat*
44	*Meminta Hujan*	*Hujan*
45	*Membersihkan Najis*	*Bersih Najis Suci*
46	*Wajib Mandi*	*Wajib Mandi Kudu Siram*
47	*Shalat Gerhana*	*Shalat Gerhana Salat*
48	*Shalat di Kendaraan*	*Shalat Kendara Salat*
49	*Memandikan Orang Mati*	*Mandi Mati Siram*
50	*Ingat Mati*	*Mati Maut*

The result of system without query expansion is given by Table 3.

Table 3. Result of system testing without query expansion

RESULT OF SYSTEM TESTING WITHOUT QUERY EXPANSION										
Query	**Ranked result in term of hadith number**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Q-1	36	97	29	28	22	21	47	23	30	34
Q-2	9	6	25	24	77	75	63	64	62	89
Q-3	29	30	53	88	471	370	400	509	23	403
Q-4	38	47	28	48	37	49	194	370	471	23
Q-5	248	282	251	506	403	400	342	294	508	510
Q-6	47	28	48	36	97	29	49	22	21	23
Q-7	95	61	70	69	175	55	85	73	96	163
Q-8	122	121	126	125	120	66	215	118	119	115
Q-9	8	114	72	71	102	115	113	127	369	33
Q-10	72	116	127	114	71	117	133	115	16	102
Q-11	269	420	419	421	134	268	192	386	270	422
Q-12	143	144	490	142	269	422	419	491	492	421
Q-13	160	175	96	370	471	163	290	291	403	23
Q-14	163	164	190	155	161	442	175	162	96	370
Q-15	169	168	171	173	174	170	172	253	179	197
Q-16	183	171	176	177	178	181	197	179	198	196
Q-17	195	422	369	312	367	194	368	104	533	370
Q-18	207	206	208	396	209	14	11	10	566	7
Q-19	381	387	383	350	380	353	377	351	447	388
Q-20	161	151	163	543	154	186	164	153	150	204
Q-21	285	284	317	283	286	282	400	370	403	475
Q-22	118	76	81	536	62	120	284	389	53	80
Q-23	265	260	266	264	263	450	341	224	262	256
Q-24	12	94	11	19	18	93	29	8	10	134
Q-25	12	94	11	18	27	19	10	50	29	83
Q-26	149	148	150	154	157	158	172	153	186	185
Q-27	205	189	471	370	400	23	509	403	186	294
Q-28	50	51	25	24	75	429	77	52	369	381
Q-29	217	218	516	317	244	349	277	370	471	143
Q-30	135	136	437	140	176	204	159	178	181	141
Q-31	323	312	311	317	475	473	351	476	474	350
Q-32	220	275	272	521	234	224	232	217	341	522
Q-33	400	398	402	399	401	342	362	508	403	354
Q-34	453	452	460	454	472	304	461	451	93	259
Q-35	480	447	413	474	420	399	362	414	421	477
Q-36	508	509	507	506	510	511	278	505	342	362
Q-37	436	432	428	433	439	374	431	404	406	438
Q-38	411	409	410	342	508	362	400	403	398	354
Q-39	341	338	340	339	277	246	276	335	272	220
Q-40	303	302	407	406	250	31	304	291	320	150
Q-41	473	474	475	351	476	477	350	310	420	421
Q-42	325	155	403	400	370	342	294	374	508	509
Q-43	235	482	239	503	270	354	243	268	238	233
Q-44	517	46	518	405	515	404	5	384	11	566
Q-45	189	210	166	349	162	167	100	471	370	23
Q-46	487	199	196	489	180	488	175	197	490	198
Q-47	530	527	526	529	522	528	524	523	525	342
Q-48	406	407	279	408	313	282	232	342	301	362
Q-49	546	544	545	487	489	58	85	534	555	185
Q-50	72	85	560	537	530	58	71	60	66	65

And the result of system testing with query expansion is given by Table 4.

Table 4. Result of system testing with query expansion

RESULT OF SYSTEM TESTING WITH QUERY EXPANSION										
Query	Rank Result in Term of Hadith Number									
	1	2	3	4	5	6	7	8	9	10
Q-1	23	28	22	47	83	471	370	29	32	36
Q-2	9	24	25	63	64	6	75	77	89	479
Q-3	29	30	471	370	509	23	400	403	294	67
Q-4	38	47	48	37	28	370	471	400	23	403
Q-5	510	400	403	370	282	342	294	248	508	509
Q-6	47	28	23	48	22	83	471	370	29	32
Q-7	95	55	175	69	70	61	96	471	370	163
Q-8	122	121	126	125	294	370	471	215	66	400
Q-9	519	395	72	8	146	370	393	471	369	113
Q-10	72	127	116	115	40	395	117	71	519	394
Q-11	134	403	400	370	342	294	192	386	508	509
Q-12	143	144	142	490	493	492	491	375	376	379
Q-13	160	370	471	400	509	175	23	403	294	186
Q-14	163	164								
Q-15	169	168	171	174	170	173				
Q-16	177	183	171	181	210	179	176	197	182	546
Q-17	549	295	422	195	369	471	370	112	312	23
Q-18	207	206	208	396	209	14	11	10	566	7
Q-19	381	400	403	350	370	383	377	342	294	380
Q-20	161	151	543	186	163	154	155			
Q-21	284	282	317	285	475	388	403	400	370	473
Q-22	80	53	364	81	324	389	435	284	231	229
Q-23	260	265	264	225	263	522	450	341	222	504
Q-24	12	29	83	11	134	94	23	344	19	93
Q-25	12	27	83	197	23	490	509	19	28	549
Q-26	149	148	150	154	157	186	153	158	172	185
Q-27	205	189	154	152	370	471	134	400	403	23
Q-28	50	24	25	77	51	467	75	370	471	550
Q-29	217	218	516	317						
Q-30	135	136	437	140	176	204	159	178	181	141
Q-31	317	323	312	475	311	400	403	370	473	342
Q-32	272	275	220	234	521	341	273	232	340	217
Q-33	400	398	402	401	399	510	403	366	342	294
Q-34	451	453	471	370	509	403	23	400	452	294
Q-35	477	420	421	414	480	474	422	413	399	445
Q-36	510	509	508	507	506	403	400	366	342	294
Q-37	432	439	433	428	436	431	374	441	404	427
Q-38	411	409	410	342	362	508	398	400	403	354
Q-39	338	341	340	339	277	276	246	272	293	318
Q-40	303	302	407	237	250	406	488	471	370	31
Q-41	477	475	473	474	476	351	350	420	317	421
Q-42	155	403	370	400	342	294	279	265	439	295
Q-43	503	269	268	502	566	400	403	354	270	370
Q-44	517	46	518	405	515	404	5	384	11	566
Q-45	189	162	154	152	134	370	471	163	166	386
Q-46	487	489	197	488	490	199	175	183	198	196
Q-47	527	530	529	526	528	522	523	510	525	400
Q-48	406	279	407	510	282	408	400	403	366	342
Q-49	546	183	544	487	186	485	184	181	549	171
Q-50	58	85	534	555	530	65	556	549	59	546

The second test is done by the expert user, where the expert user is asked to determine the relevance of query with the hadith documents that has been retrieved by the system. The result of expert testing is given by Table 5.

Table 5. Result of expert testing

EXPERT TESTING RESULT					
Query	Ground Truth from Expert				
	1	2	3	4	5
Q-1	28	29	47	22	
Q-2	9	6	25	24	
Q-3	29	30			
Q-4	38	47	48		
Q-5	248	282			
Q-6	47	28	29	22	48
Q-7	69	70	55		
Q-8	122	121	126	215	125
Q-9	8	369	113		
Q-10	72	127	115	116	71
Q-11	134				
Q-12	142	143	144	490	
Q-13	160				
Q-14	163	164			
Q-15	169	168	170		
Q-16	183	181	182		
Q-17	195	194			
Q-18	207	206	208	209	
Q-19	381	380			
Q-20	161				
Q-21	284				
Q-22	53	364			
Q-23	260	264	265	266	263
Q-24	29	94	11	19	
Q-25	27				
Q-26	149	148	150	154	153
Q-27	205				
Q-28	50	51	24	25	
Q-29	217	218			
Q-30	135	136	176	159	141
Q-31	312	311	351		
Q-32	220	275	234	217	
Q-33	400	398	402	401	403
Q-34	451	452	453		
Q-35	480	420	447	414	477
Q-36	510	508	509	507	506
Q-37	432	433	428	431	439
Q-38	411	409	410		
Q-39	338	339	341	340	
Q-40	303	302	304		
Q-41	473	474	475	476	477
Q-42	155				
Q-43	269	268	270	566	503
Q-44	517	515			
Q-45	162	166	189	154	163
Q-46	197	199	196	198	
Q-47	527	529	530	526	528
Q-48	406	407	408		
Q-49	546	545	544		
Q-50	58	85	534	59	

To evaluating the performance of applying the method to the system that has been made, the test is starting with comparing the result of system testing with the expert testing to measure the recall, precision, accuracy and f-measure. The test is carried out in 3 stages, the first stage is testing on top 5 data, then testing on top 3 data and testing on top 1 data. The test is carried out on the system without using query expansion and by using query expansion. The results of testing is given by Table 6.

Table 6. Result of testing without query expansion

Results of testing without query expansion				
	Recall	Precision	Accuracy	F-Measure
TOP 5 DATA	79,81	51,2	68	59,12
TOP 3 DATA	68,34	58,01	79,2	61,14
TOP 1 DATA	62	62	92,4	62

Based on Table 6, in the top 5 data testing obtained an average recall value of 79.81%, the average precision value of 51.2%, the average accuracy value of 68% and the average f-measure value of 59.12%, whereas in the top 3 data testing obtained an average recall value of 68.34%, an average precision value of 58.01%, an average accuracy value of 79.2% and an average f-measure value of 61.14%, and in the top 1 data testing obtained an average recall value of 62%, the average precision value of 62%, the average accuracy value of 92.4% and the average f-measure value of 62%. The comparison graph based on the table above can be seen in Figure 2.

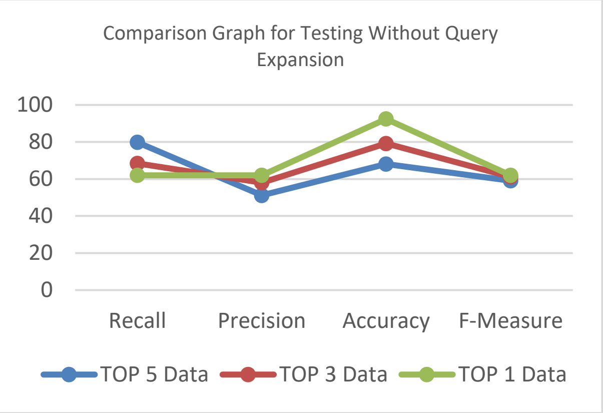

Figure 2. Comparison graph for testing without query expansion

And the result of testing with query expansion is given by Table 7.

Table 7. Result of testing with query expansion

Reasults of testing with query expansion				
	Recall	Precision	Accuracy	F-Measure
TOP 5 DATA	81,97	53,2	70,08	61,18
TOP 3 DATA	72,67	60,67	80,6	64,07
TOP 1 DATA	66	66	93,2	66

Based on Table 7, in the top 5 data testing obtained an average recall value of 81.97%, an average precision value of 53.2%, an average accuracy value of 70.08% and an average f-measure value of 61.18%, whereas in the top 3 data testing obtained an average recall value of 72.67%, an average precision value of 60.67%, an average accuracy value of 80.6% and an average f-measure value of 64.07%, and in the top 1 data testing obtained an average recall value of 66%, an average precision value of 66%, an average accuracy value of 93.2% and an average f-measure value of 66%. The comparison graph based on the table above is given by Figure 3.

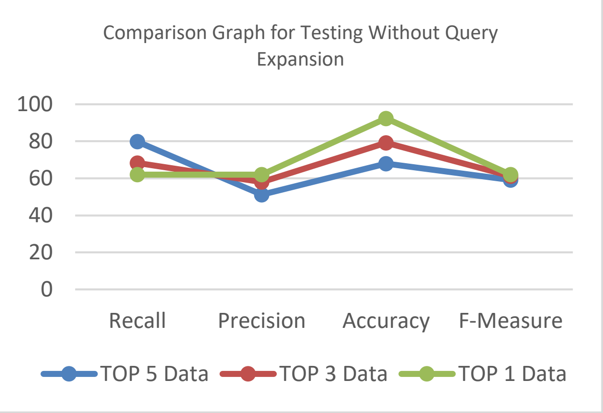

Figure 3. Comparison graph for testing with query expansion

ACKNOWLEDGMENT

Based on the results of the implementation and testing that have been conducted by the researcher, get the conclusion: based on testing of the top 5 data the use of query expansion shows an improvement in the average recall of 2.16%, an improvement in the average precision of 2%, an improvement in the average accuracy of 2.08%, and an improvement in the average f-measure of 2.06% compared by without using query expansion. Moreover, based on testing of the top 3 data the use of query expansion shows an improvement in the average recall of 4.33%, an improvement in the average precision of 2.66%, an improvement in the average accuracy of 1.4% and an improvement in the average f-measure of 2.93% compared by without using query expansion. Finally, based on testing of the top 1 data the use of query expansion shows an improvement in the average recall of 4%, an improvement in the average precision of 4%, an improvement in the average accuracy of 0.8% and an improvement in the average f-measure of 4% compared by without using query expansion.

Based on the explanation about the results of the implementation above, it can take a conclusion that the use of Indonesian thesaurus based on query expansion could give the better result than without using query expansion.

REFERENCES

[1] Baqi, M. F. (2015). AL Lu' Lu' Wal Marjan, Muttafaqun 'Alaih Shahih Bukhari. Solo: Beirut.

[2] Ginting, S. L., & Trinada, R. P. (2014). *Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan*. *JATI UNIKOM*, 14.

[3] Yadav, D. P. (2014). *Document-Document Similarity Matrix and Naive-Bayes Classification to Web Information Retrieval*. International Journal of Engineering Research and General Science Volume 2, 7.

[4] Shahabadkar, D. R., Reddy, Y. V., Khrisna, B. M., & & Devi, T. (2017). Enhanced Query Expansion For Web Information Retrieval. International Journal of Civil Engineering and Technology (IJCIET). *International Journal of Civil Engineering and Technology (IJCIET), Volume 8 , Issue 8*, 6.

[5] Suyanto, S. M. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data. Bandung*: INFORMATIKA.

[6] Hasugian. (2006). *Penggunaan bahasa alamiah dan kosa kata terkendali dalam sistem temu balik informasi berbasis teks, departemen studi perpustakaan dan informasi universitas sumatera utara*. Medan: Pustaka.