

## The search for science and technology verses in Qur'an and hadith

Ichsan Taufik<sup>1</sup>, Mohamad Jaenudin<sup>2</sup>, Fatimah Ulwiyatul Badriyah<sup>3</sup>, Beki Subaeki<sup>4</sup>, Opik Taupik Kurahman<sup>5</sup>

<sup>1,3,4,5</sup>Informatic Department, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia

<sup>2</sup>Islamic Education Department, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia

### Article Info

#### Article history:

Received Sep 18, 2020

Revised Nov 16, 2020

Accepted Dec 5, 2020

#### Keywords:

Hadith

Qur'an

Science and technology

Searching

Vector space model

### ABSTRACT

Currently, the vector space model algorithm has been widely implemented for the document search feature because of its reliability in retrieving information. One of them in the search for verses of the Qur'an based on the translation. However, if the phrase or word used is different (even though it has one meaning) with the word in the document in the database, the system will not display the verse. As we know that the Qur'an has a very deep meaning, so an interpretation of the verse is needed. Therefore, this research focuses on implementing the vector space model (VSM) algorithm for searching verses and hadiths in science and technology by using the discussion parameters of these verses or hadiths. The test results obtained with 20 keyword samples using metric recall were 81% with an average time of 2.24 seconds.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Ichsan Taufik

Informatic Department

Universitas Islam Negeri Sunan Gunung Djati Bandung

Jl. A. H. Nasution No. 105, Cibiru, Bandung 40614, Indonesia

Email: ichsan@uinsgd.ac.id

## 1. INTRODUCTION

The Qur'an is the essence of all science. However, the knowledge contained in the Qur'an is still in the form of seeds and principles. The Qur'an contains the principles of all science, including technology and knowledge of the universe. The Qur'an contains various levels of definition and layered meaning for all readers [1]. Sunnah or hadith according to El-Naggar [2] is everything that comes from the Prophet Muhammad, in the form of words, behavior, determination of the Prophet, nature, sirah (biography of the Prophet SAW), both before and after he was sent as a prophet and apostle. The sunnah of the Prophet is the second life guide after the Qur'an.

Information technology facilitates all human needs in almost all aspects. Smartphones, Search engines and television are real examples of the application of information technology [3]. Search engines like Google, Bing, Yahoo and Ask in their search process adopted a system called the information retrieval system [4]. This system accepts keywords or input from the user to the document based on its suitability to the keywords [5].

The vector space model (VSM) algorithm is an information retrieval model that presents documents and keywords into vectors in multidimensional spaces. The similarity of both can be measured by calculating the angles formed by document vectors and queries [4]. With this algorithm, we can use words, phrases, or sentences for the keyword [6]. The structure of phrases and sentences does not have to be exactly the same as the document in the database. VSM adopts similarity measure for matching between documents and user

query, and assign scores from the biggest to smallest. The documents and query are assigned with weights using term frequency and inverse document frequency method [7]. There are many variants TF-IDF model to get the weights including term frequency, classical TF-IDF, normalized TF-IDF and sub-linear normalized TF-IDF [8].

Currently, it has many applications hadith or the Qur'an which include search features. However, the technique is still limited to the search for a match search word (string matching) so that the results often do not appear when a query is given in the form of a sentence or phrase [8]. In addition, until now there has been no application or system that specifically provides information on the verses of the Qur'an and hadith relating to science and technology [9]. Whereas in the interpretation of Shaykh Thantawi Al-Jawahir, it is said that in the Holy Qur'an there are more than 750 verses of kauniyah (verses about the universe [10]).

Many search techniques are currently used by researchers. One of the most popular search techniques is word matching or string matching. The concept of searching for this technique is to match a certain word pattern to a long sentence or text [11]. However, if the wording is not the same as in the database, the search will not be found. Another alternative technique for finding information that is also currently popular is the vector space model algorithm that adopts word weighting. With this technique, the wording does not have to be exact. Even if the user enters the words upside down or in the wrong order in the database, the system will still return the information. In previous studies related to the implementation of the VSM algorithm for searching Al-Qur'an verses [12], the parameters used were through the translation of Al-Qur'an verses and expansion queries to improve keywords to be based on data in the database. However, if the user enters a different query but has the same meaning and is not in the database, the system will not be able to return the information it is looking for. Therefore this research in its implementation tries to use the discussion that comes from the books of commentaries of the Qur'an and the Hadith whose language structure is closer to daily life as its parameters.

## 2. RESEARCH METHOD

In this paper, research is conducted with the steps that can be seen in Figure 1. The first step taken is to enter keywords (can be in the form of words, phrases or sentences), then do the preprocessing to get a list of word tokens which then do the weight calculation tf-idf after it obtained a list of queries and documents weights used in the computation algorithm VSM [13-16] if the cosine value  $> 0$  then shown a list of relevant data otherwise found no data is displayed.

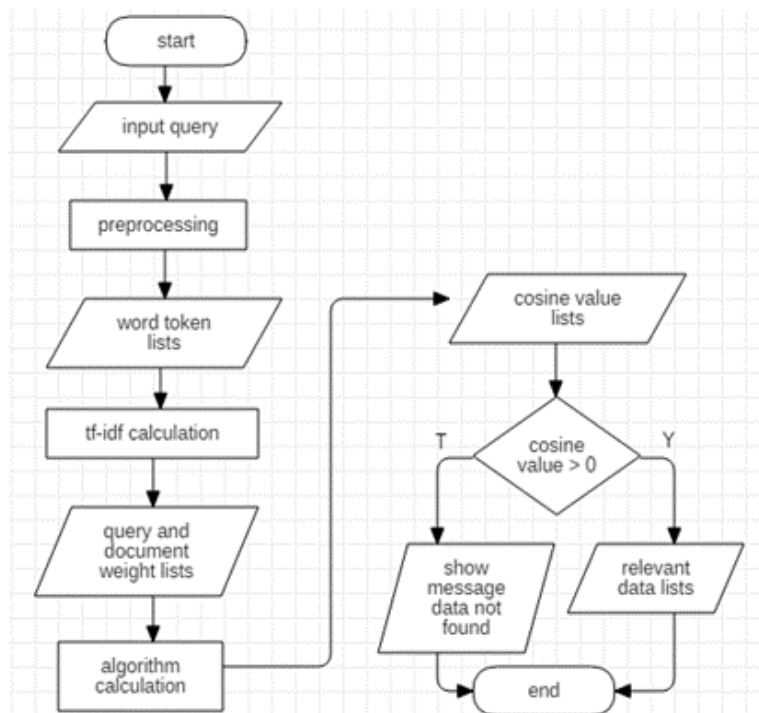


Figure 1. Vector space model flowchart

This research uses a software development method called the prototype model [17-19]. The application to be built is a website based application that is built using the Django framework and Python. To be able to access it, users need a browser [20]. The application architecture developed is illustrated in Figure 2. First, the client will send a URL request through the browser. This URL contains the application pathname. If it is true, then the HTML page of the templates and data from the Model in the database will be sent to the webserver and displayed.

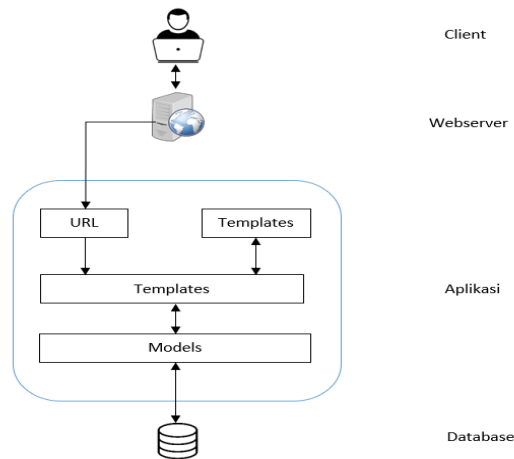


Figure 2. Application architecture

### 3. RESULTS AND DISCUSSION

#### 3.1. Collecting text data of Qur'an and hadith

The verses data used in this research is in the form of Indonesian translation Qur'an and hadith text data that obtained from the Mushaf Al-'Alim edition of science [1], Book of scientific miracles on earth and space [5] and three series of books proving science in the sunnah El-Naggar [2]. In addition, until now there has been no application or system that specifically provides information on the verses of the Qur'an and hadiths relating to science and technology [9].

#### 3.2. Preprocessing

As has been described in the flowchart that this algorithm uses the weight of the query and document to find the resemblance [21]. Before entering the algorithm calculation, the document and query must go through the preprocessing (text processing) to get the weight value. This stage is divided into 4 parts, namely case folding, filtering, stemming, and tokenizing [22].

There are 4 known documents and Q (input query):

D1 = *Sesungguhnya manusia berada dalam kerugian*

D2 = *Demi langit yang mempunyai gugusan bintang*

D3 = *Aku tidak akan menyembah apa yang kamu sembah*

D4 = *Sesungguhnya mereka adalah orang-orang yang merugi*

Q = *Manusia yang rugi*

- Case folding

The first stage in preprocessing is leveling letters (case folding) into a lowercase [23].

D1 = *Sesungguhnya manusia berada dalam kerugian*

- Filtering

This stage removes the stopword (meaningless word) and punctuation in each document such as the word 'adalah, yang, apa, dan sebagainya' [24].

D1 = *Sesungguhnya manusia kerugian*

- Stemming

The stemming stage is the process of converting words into their basic form, by removing affixes to the word [25].

D1 = *Sungguh manusia rugi*

- Tokenizing

The last stage is tokenizing which functions to separate each word into tokens [26] as: *sungguh, manusia, rugi, langit, gugusan, bintang, sembah, orang.*

**3.3. Word weighting (TF-IDF)**

In Table 1 the term column is words that had previously gone through the preprocessing stage. Next, we calculate the frequency of terms that appear in each document. If the document contains the term sought, then the value is one, if not, then the value is zero. Column Q is the keyword sample that you want to find, that is, human loss. After that, we calculate the frequency of occurrence of each term in the whole document (DF) [27]. Based on calculations on the theoretical basis, to determine how important a term is to an entire document, we must calculate the value of the IDF (inverse document frequency) of each term with (1) [28]. After that, we will get the term weight by multiplying the term frequency value and the IDF [29].

$$idf_i = \log \frac{N}{df_i} \tag{1}$$

Table 1. TF-IDF weightings

Term	TF					DF	IDF (log(N/DF))	W (TF*IDF)				
	Q	D1	D2	D3	D4			Q	D1	D2	D3	D4
<i>Sungguh</i>	0	1	0	0	1	2	0.30103	0	0.30103	0	0	0.30103
<i>Manusia</i>	1	1	0	0	0	1	0.60206	0.60206	0.60206	0	0	0
<i>Rugi</i>	1	1	0	0	1	2	0.30103	0.30103	0.30103	0	0	0.30103
<i>Langit</i>	0	0	1	0	0	1	0.60206	0	0	0.60206	0	0
<i>Gugusan</i>	0	0	1	0	0	1	0.60206	0	0	0.60206	0	0
<i>Bintang</i>	0	0	1	0	0	1	0.60206	0	0	0.60206	0	0
<i>Sembah</i>	0	0	0	2	0	2	0.30103	0	0	0	0.60206	0
<i>Orang</i>	0	0	0	0	1	1	0.60206	0	0	0	0	0.60206

**3.4. VSM algorithm process**

The first step in calculating this algorithm, we calculate the size of the query vector and the document vector using (2) and (3) on the theoretical basis of the algorithm by first squaring the weight value (Table 2 columns  $W_q^2$  through  $W_{D4}^2$ ). After that, the square value of each document and query is added up (Table 2 columns in blue) and then the square root values (Table 2 rows in orange) are searched. The second step multiplies each term weight in each document by the query weight (Table 2 columns  $W_Q * W_{D1}$  to  $W_Q * W_{D4}$ ) then add up the values for each document (Table 2 red rows). To get the document similarity values according to (4) for the values that we have added in the red line with the results of the square root (in Table 2 orange lines). The final ranking of the document is indicated by a gray bar. We can see that the documents that are relevant to the query are documents one and four (the value is not equal to zero) and the document that has the highest level of similarity to the query is document one or D1.

$$w_{ij} = tf_{ij} \cdot idf_{ij} \tag{2}$$

$$|q| = \sqrt{\sum_{j=1}^t (w_{iq})^2} \tag{3}$$

$$sim(q, d_j) = \frac{q \cdot df_j}{|q| \cdot |d_j|} = \frac{\sum_{i=1}^t w_{iq} \cdot w_{ij}}{\sqrt{\sum_{j=1}^t (w_{iq})^2} \cdot \sqrt{\sum_{i=1}^t (w_{ij})^2}} \tag{4}$$

Table 2. VSM algorithm process

Term	$W_q^2$	$W_{D1}^2$	$W_{D2}^2$	$W_{D3}^2$	$W_{D4}^2$	$W_Q * W_{D1}$	$W_Q * W_{D2}$	$W_Q * W_{D3}$	$W_Q * W_{D4}$
	0	0.09062	0	0	0.090619	0	0	0	0
<i>Sungguh</i>	0.36248	0.36248	0	0	0	0.362476	0	0	0
<i>Manusia</i>	0.09062	0.09062	0	0	0.090619	0.090619	0	0	0.090619
<i>Rugi</i>	0	0	0.362476	0	0	0	0	0	0
<i>Langit</i>	0	0	0.362476	0	0	0	0	0	0
<i>Gugusan</i>	0	0	0.362476	0	0	0	0	0	0
<i>Bintang</i>	0	0	0	0.362476	0	0	0	0	0
<i>Sembah</i>	0	0	0	0	0.362476	0	0	0	0
Sum	0.4531	0.54371	1.087429	0.362476	0.543714	Sum $W_Q * W_{D_i}$			
SQRT	0.67312	0.73737	1.042798	0.60206	0.73737	0.453095	0	0	0.090619
					Rank	0.912871	0	0	0.204124

### 3.5. Testing

This test is intended to test the performance of the vector space model algorithm using 20 sample keywords through the recall method [4]. The test results can be seen in Table 3.

Table 3. Vector space model testing

No.	Keyword	Time	Result		Explanation
			Success	Failed	
1.	<i>Bintang</i>	00.02.76	√		$\frac{9}{9} \cdot 100 = 100\%$
2.	<i>Perbintangan</i>	00.03.32	√		$\frac{9}{9} \cdot 100 = 100\%$
3.	<i>Gunung sebagai pasak</i>	00.02.43	√		$\frac{5}{6} \cdot 100 = 83\%$
4.	<i>Ilmu genetika</i>	00.03.10	√		$\frac{7}{7} \cdot 100 = 100\%$
5.	<i>Komputer</i>	00.01.80	√		$\frac{1}{1} \cdot 100 = 100\%$
6.	<i>Manfaat siwak</i>	00.02.15	√		$\frac{1}{1} \cdot 100 = 100\%$
7.	<i>Matahari yang menggerhana</i>	00.02.86	√		$\frac{9}{11} \cdot 100 = 81\%$
8.	<i>Gerhana matahari</i>	00.01.78	√		$\frac{10}{11} \cdot 100 = 90\%$
9.	<i>Besi berasal dari langit</i>	00.02.33	√		$\frac{11}{13} \cdot 100 = 81\%$
10.	<i>Pengaruh gen keturunan</i>	00.01.42	√		$\frac{5}{7} \cdot 100 = 71\%$
11.	<i>Tanaman</i>	00.02.22	√		$\frac{1}{1} \cdot 100 = 100\%$
12.	<i>Tujuh lapis bumi</i>	00.01.67	√		$\frac{9}{11} \cdot 100 = 81\%$
13.	<i>Angin yang berhembus</i>	00.02.48	√		$\frac{1}{1} \cdot 100 = 100\%$
14.	<i>Laut yang mendidih</i>	00.02.19	√		$\frac{4}{6} \cdot 100 = 66\%$
15.	<i>Meteor</i>	00.02.36	√		$\frac{2}{2} \cdot 100 = 100\%$
16.	<i>Orbit bumi</i>	00.02.15	√		$\frac{10}{12} \cdot 100 = 83\%$
17.	<i>Tulang-belulang</i>	00.02.54	√		$\frac{1}{1} \cdot 100 = 100\%$
18.	<i>Gravitasi bumi</i>	00.02.08	√		$\frac{20}{24} \cdot 100 = 83\%$
19.	<i>Nebula</i>	00.01.89	√		$\frac{1}{1} \cdot 100 = 100\%$
20.	<i>Asam nukleat</i>	00.01.46	√		$\frac{1}{1} \cdot 100 = 100\%$

Testing is done by calculating the percentage of recall [30] to get the algorithm's ability to reinvent information [4]. Based on the information on each test data, the average recall value of the system is:

$$Recall = \frac{\text{the number of relevant documents that are called}}{\text{The number of relevant documents in the database}}$$

The system recall results showed an algorithm testing value of 81% for the level of success in finding back information with an average time of 2.24 seconds. The failure of the system to retrieve some information can be caused by the ability of the local server and the amount of data that is processed [31].

## 4. CONCLUSION

In this paper, the vector space model algorithm is applied to the system through four stages, namely, preprocessing, calculation of document weights and queries, calculation of cosine angles of document and query vectors, and ranking of documents. Based on the test results using 20 sample data, it can be concluded that the Vector space model algorithm provides good performance in rediscovering verses in the Qur'an or Hadiths relating to Science and Technology by 81% with an average time of 2.24 seconds. Vector space

model algorithm cannot solve writing errors, so it is expected in future studies to implement expansion queries to get better search results.

## ACKNOWLEDGEMENTS

Authors wishing to acknowledge Research and Publication Centre of UIN Sunan Gunung Djati Bandung that supports and funds this research publication.

## REFERENCES

- [1] N. Amalia, N. H. A. Abd. Safa, M. A. Armin, "Application of Charles Sanders Peirce's Semiotics Theory in Naming the al-Baqarah and the al-Ankabut of the Qur'an," *Proceedings of the Third International Seminar on Recent Language, Literature, and Local Culture Studies, BASA, EAI*, pp. 1-7, 2020.
- [2] Z. El-Naggar, "Scientific Precision in the Sunnah," *Quotations from the Sayings of Prophet Mohammad (PBUH). Beirut, Lebanon: Dar Al-Marefah*, 2010.
- [3] F. Faiz, T. Boyhan, W. Izaak, and I. Yuhertiana, "A literature review: Women's equality in the e-commerce information technology industry in the digital age in developing countries," *Proceedings of the 2nd International Conference on Economics, Business, and Government Challenges, EBGC 2019, EAI*, 2020.
- [4] S. H. Toman, M. H. Abed, and Z. H. Toman, "Cluster-based information retrieval by using (K-means)-hierarchical parallel genetic algorithms approach," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 19, no. 1, pp. 349-356, 2021.
- [5] S. Amudha, E. I. Shanthi, "Phrase based information retrieval analysis in various search engines using machine learning algorithms," *Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing*, pp. 281-291, 2020.
- [6] S. E. Pratama, W. Darmalaksana, D.S. Maylawati, H. Sugilar, T. Mantoro, and M.A. Ramdhani, "weighted inverse document frequency and vector space model for hadith search engine," *Indonesian Journal Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1004-1014, 2020.
- [7] E. Wahyudi, S. Sfenrianto, M. J. Hakim, R. Subandi, O. R. Sulaeman, and R. Setiyawan, "Information retrieval system for searching JSON files with vector space model method," In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT), IEEE*, pp. 260-265, 2019.
- [8] D. Yogish, Manjunath T. N., R.S. Hegadi, "Variants of term frequency and inverse document frequency of vector space model for effective document ranking in information retrieval," *Intern. J. of Inn. Tech. & Expl. Eng.*, vol. 8, no. 7, pp. 414-421, 2019
- [9] I. O. Jimoh, "Qur'an Addresses Science and Technology," *Int. J. Of Res. And Innov. In Soc. Sci.*, vol. 3, no. 8, pp. 415-424, 2019.
- [10] Y. A. Ahmad, "Scientific Wonders on the Earth & in Space," Riyadh: *Darussalam*, 2014.
- [11] M. B. Sri, R. Bhavsar, and P. Narooka, "String matching algorithms," *Int. J. Of Eng. And Comp. Sci.*, vol. 7, no. 3, pp. 23769-23772, 2017.
- [12] A. Salih, B., "Applying vector space model (VSM) techniques in information retrieval for Arabic Language," *arXiv preprint arXiv:1801.03627*, 2018.
- [13] B. Pathak and N. Lal, "Information retrieval from heterogeneous datasets using moderated IDF-cosine similarity in vector space model," *Int. Conf. on En. Comm. Da. Anal. And soft comput.*, IEEE, pp. 3793-3799, 2017.
- [14] S. M. Weiss, N. Indurkha, T. Zhang, and F. J. Damerou, "Information retrieval and text mining," Springer Berlin Heidelberg. no. *Fundamentals of Predictive Text Mining*, pp. 75-90, 2010.
- [15] U. Parida, M. Nayak, and A. K. Nayak, "Ranking of Odia text document relevant to user query using vector space model," *Int. Conf. On Appl. Mach. Learn.*, IEEE, pp. 165-169, 2019.
- [16] C. D. Manning, P. Ragahvan, and H. Schutze, "An Introduction to Information Retrieval," Cambridge University Press., vol. 39, pp. 1041-4347, 2008.
- [17] R. S. Pressman, "Software Engineering: A Practitioner's Approach," Palgrave Macmillan, 2005.
- [18] I. Sommerville, "Software Engineering, 5th ed.," Addison Wesley, 1995.
- [19] I. Taufik, Y. A. Gerhana, A. I. Ramdani, and M. Irfan, "Implementation K-Nearest Neighbour for Student Expertise Recommendation System," *Journal of Physics: Conference Series AASEC 2019*, vol. 1402, no. 7, p. 077004, 2019.
- [20] S. Aisa, "Implementation private cloud using Raspberry PI for accessing personal data," in Bahasa "Implementasi private cloud menggunakan Raspberry PI untuk pengaksesan data pribadi," *Jurnal Penelitian Pos & Informatika*, vol. 6, no. 2, pp. 137-152, 2016.
- [21] C. Slamet, R. Andrian, D. S. Maylawati, W. Darmalaksana, and M. A. Ramdhani, "Web scraping and Naïve Bayes classification for job search engine," *IOP Conference Series: Materials Science and Engineering, The 2nd Annual Applied Science and Engineering Conference (AASEC 2017)*, vol. 288, no. 1, pp. 1-7, 2018.
- [22] L. M. Q. Abualigah and E. S. Hanandeh, "Applying genetic algorithms to information retrieval using vector space model," *International Journal of Computer Science, Engineering and Applications*, vol. 5, no. 1, pp. 19-28, 2015.
- [23] A. Jain, G. Kulkarni, and V. Shah, "Natural language processing," *Int. J. Of Comp Sci. and Eng.*, vol. 6, no. 1, pp. 161-167, 2018.
- [24] A. C. de Bordignon, L. H. Thom, T. S. Silva, V. S. Dani, M. Fantinato, and R. C. B. Ferreira, "Natural language processing in business process identification and modeling: a systematic literature review," *Proceedings of the XIV Brazilian Symposium on Information Systems*, pp. 1-8, 2018.

- [25] N. Bakala, "Information retrieval system by using vector space model," *Int. J. of Scient. & Tech Res.*, vol. 8 no. 10, pp. 1562-1568, 2019.
- [26] M. A. Ramdhani, *et al.*, "Indonesian news classification using convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2, pp. 1000-1009, 2020.
- [27] C. Platzer and D. Dustdar, "A vector space search engine for web services," *Proc. of th 3<sup>rd</sup> European Conf. On Web. Serv.*, p. 9, 2005.
- [28] B. Yulianto, W. Budiharto, and I. H. Kartowisastro, "The performance of boolean retrieval and vector space model in textual information retrieval," *Comm. In. Tech. J.*, vol. 11, no. 1, pp. 33-39, 2017.
- [29] M. Alaqdash, M.A. Fadhil, and A. F. Al-Azzawi, "A modified overlapping partitioning cluster algorithm for categorical data clustering," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 1, pp. 55-62, 2018.
- [30] M. Md. Saad, N. Jamil, and R. Hamzah, "Evaluation of support vector machine and decision tree for emotion recognition of Malay Folklores," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 3, pp. 479-486, 2018.
- [31] W. Darmalaksana, C. Slamet, W. B. Zulfikar, I. F. Fadillah, D. S. A. and H. Ali, "Latent semantic analysis and cosine similarity for hadith search engine," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 18, no. 1, pp. 217-227, 2020.

## BIOGRAPHIES OF AUTHORS



**Ichsan Taufik** received the S.T. degree in informatic from IAIN Sunan Gunung Djati, Indonesia, in 2005, the M.T. degree in information systems from Insitute Technology Bandung, Indonesia, in 2010. He is currently a lecturer in informatic at State Islamic University Sunan Gunung Djati, Indonesia. He is also Editorial Section of Jurnal Online Informatika (JOIN). His current research interests include information system, algorithm and programming, and software engineering.



**Mohammad Jaenudin** received the S.Ag. degree in religious court from IAIN Sunan Gunung Djati, Indonesia, in 1997, the M.Ag. degree in Islamic education from IAIN Sunan Gunung Djati, Indonesia, in 2005, and the Dr. Degree from State Islamic University Sunan Gunung Djati, Indonesia, in 2016. He is currently a lecturer in fiqh at State Islamic University Sunan Gunung Djati, Indonesia. He is also University Senate substantive specialty staff. His current research interests include islamic education, ict and islam.



**Fatimah Ulwiyatul Badriyah** received the S.T. degree in informatic from State Islamic University Sunan Gunung Djati, Indonesia, in 2019. She is currently a website developer at P.T Kerjaku and also has experience to build Rest API and Mobile apps.



**Beki Subaeki** received the S.Kom. degree in IT from STMIK AKAKOM, Indonesia, in 2008, the M.Kom. degree in information systems from Universitas Komputer Indonesia, Indonesia, in 2012. He is currently a lecturer in information system at Sangga Buana University, Indonesia His current research interest include information system.



**Opik Taupik Kurahman** received the S.Ag. degree in Islamic education from IAIN Sunan Gunung Djati, Indonesia, in 1993, the M.Ag. degree in Islamic education from IAIN Ar-Raniry Banda Aceh, Indonesia, in 1998, and the Dr. Degree from State Islamic University Sunan Gunung Djati, Indonesia, in 2014. He is currently a lecturer in ict and islam at State Islamic University Sunan Gunung Djati, Indonesia. He is also a head of university library. His current research interests include islamic education, ict and islam and the field of halal studies.