

# BEEI Ichsan 2020

*by* Jurnal Beei Ichsan

---

**Submission date:** 07-Dec-2020 09:54PM (UTC+0700)

**Submission ID:** 1467482260

**File name:** Revision\_1570639064\_BEEI\_2020\_Ichsan\_fix\_latest.doc (1.46M)

**Word count:** 3286

**Character count:** 16517



## 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 Jun x, 20xx

Revised Nov x, 20xx

Accepted Dec x, 20xx

#### Keywords:

Hadith

Science and technology

Searching

Qur'an

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 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](mailto: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 kauniah (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.

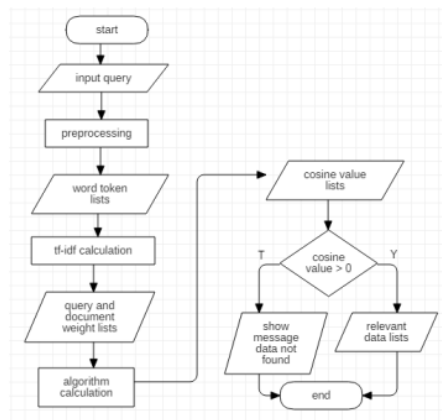


Figure 1 Vector Space Model Flowchart

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's true, then the HTML page of the templates and data from the Model in the database will be sent to the webservice and displayed.

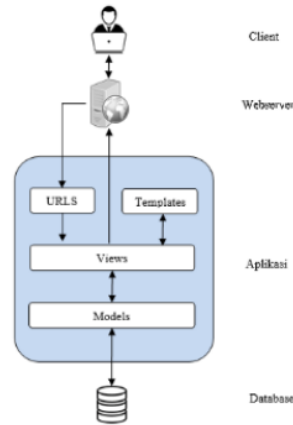


Figure 2 Application Architecture

### 3. RESULTS AND DISCUSSION

#### 1 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]. Examples of calculations are as follows.

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

##### a. Case folding

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

D1 = sesungguhnya manusia berada dalam kerugian.

##### b. 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.

##### c. 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

##### d. Tokenizing

2

The last stage is tokenizing which functions to separate each word into tokens [26] as follows: 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 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].

Table 1 TF-IDF Weighting

Term	TF					DF	IDF (log(N/DF))	W (TF * IDF)				
	Q	D1	D2	D3	D4			Q	D1	D2	D3	D4
Sungguh	0	1	0	0	1	2	0,30103	0	0,30103	0	0	0,30103
Manusia	1	1	0	0	0	1	0,60206	0,60206	0,60206	0	0	0
Rugi	1	1	0	0	1	2	0,30103	0,30103	0,30103	0	0	0,30103
Langit	0	0	1	0	0	1	0,60206	0	0	0,60206	0	0
Gugusan	0	0	1	0	0	1	0,60206	0	0	0,60206	0	0
Bintang	0	0	1	0	0	1	0,60206	0	0	0,60206	0	0
Sembah	0	0	0	2	0	2	0,30103	0	0	0	0,60206	0
Orang	0	0	0	0	1	1	0,60206	0	0	0	0	0,60206

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 equation two [28]. After that, we will get the term weight by multiplying the term frequency value and the IDF[29].

### 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 equations 3 and 4 on the theoretical basis of the algorithm by first squaring the weight value (Table 2 columns Wq2 through WD42). After that, the square value of each document and query is added up (Table 2 rows 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 WQ \* WD1 to WQ \* WD4) then add up the values for each document (Table 2 red rows). To get the document similarity values according to our equation 7 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.

Table 2 VSM Algorithm Process

Term	W <sub>q</sub> <sup>2</sup>	W <sub>D1</sub> <sup>2</sup>	W <sub>D2</sub> <sup>2</sup>	W <sub>D3</sub> <sup>2</sup>	W <sub>D4</sub> <sup>2</sup>	W <sub>Q</sub> *W <sub>D1</sub>	W <sub>Q</sub> *W <sub>D2</sub>	W <sub>Q</sub> *W <sub>D3</sub>	W <sub>Q</sub> *W <sub>D4</sub>
	0	0,09062	0	0	0,090619	0	0	0	0
sungguh	0,36248	0,36248	0	0	0	0,362476	0	0	0
manusia	0,09062	0,09062	0	0	0,090619	0,090619	0	0	0,090619
Rugi	0	0	0,362476	0	0	0	0	0	0
langit	0	0	0,362476	0	0	0	0	0	0
gugusan	0	0	0,362476	0	0	0	0	0	0
bintang	0	0	0	0,362476	0	0	0	0	0
sembah	0	0	0	0	0,362476	0	0	0	0

<b>Sum</b>	0,4531	0,54371	1,087429	0,362476	0,543714	<b>Sum <math>W_0 * W_{Dk}</math></b>			
<b>SQRT</b>	0,67312	0,73737	1,042798	0,60206	0,73737	0,453095	0	0	0,090619
					<b>Rank</b>	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.	Bintang	00.02.76	√		$\frac{9}{9} \cdot 100 = 100\%$
2.	Perbintangan	00.03.32	√		$\frac{9}{9} \cdot 100 = 100\%$
3.	Gunung sebagai pasak	00.02.43	√		$\frac{5}{6} \cdot 100 = 83\%$
4.	Ilmu genetika	00.03.10	√		$\frac{7}{7} \cdot 100 = 100\%$
5.	Komputer	00.01.80	√		$\frac{1}{1} \cdot 100 = 100\%$
6.	Manfaat siwak	00.02.15	√		$\frac{1}{1} \cdot 100 = 100\%$
7.	Matahari yang menggerhana	00.02.86	√		$\frac{9}{11} \cdot 100 = 81\%$
8.	Gerhana matahari	00.01.78	√		$\frac{10}{11} \cdot 100 = 90\%$
9.	Besi berasal dari langit	00.02.33	√		$\frac{13}{16} \cdot 100 = 81\%$
10.	Pengaruh gen keturunan	00.01.42	√		$\frac{5}{7} \cdot 100 = 71\%$
11.	Tanaman	00.02.22	√		$\frac{1}{1} \cdot 100 = 100\%$
12.	Tujuh lapis bumi	00.01.67	√		$\frac{9}{11} \cdot 100 = 81\%$
13.	Angin yang berhembus	00.02.48	√		$\frac{1}{1} \cdot 100 = 100\%$
14.	Laut yang mendidih	00.02.19	√		$\frac{4}{6} \cdot 100 = 66\%$
15.	Meteor	00.02.36	√		$\frac{2}{2} \cdot 100 = 100\%$
16.	Orbit bumi	00.02.15	√		$\frac{10}{12} \cdot 100 = 83\%$
17.	Tulang-belulang	00.02.54	√		$\frac{1}{1} \cdot 100 = 100\%$
18.	Gravitasi bumi	00.02.08	√		$\frac{20}{24} \cdot 100 = 83\%$
19.	Nebula	00.01.89	√		$\frac{1}{1} \cdot 100 = 100\%$
20.	Asam nukleat	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 as follows:

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

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," *BASA, EAI*, pp. 1–7, 2020.
- [2] Z. El-Naggar, "Scientific Precision in the Sunnah," *London: Dar Al-Marefah*, 2010.
- [3] F. Faiz, et al., "A Literature Review: Women's Equality in the E-Commerce Information Technology Industry in the Digital Age in Developing Countries," *EBCG, 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*, vol. 19, no. 1, pp. 349–356, 2020.
- [5] S. Amudha, E. Shanthi I. "Phrase Based Information Retrieval Analysis in Various Search Engines Using Machine Learning Algorithms. In: Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing", vol 1016, 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," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 2, p. 1004-1014, 2020.
- [7] E. Wahyudi et al., "Information Retrieval System for Searching JSON Files with Vector Space Model Method," *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 Issue 7, pp. 415-421, 2019
- [9] I. O. Jimoh, "Qur'an Addresses Science and Technology," *Int. J. Of Res. And Innov. In Soc. Sci.*, vol. III, no. VIII, 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, 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.*, pp. 10–27, 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.*, pp. 165–169, 2019.
- [16] C. D. Manning, P. Ragahvan, and H. Schutze, *An Introduction to Information Retrieval*, no. c. 2009.
- [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," *AASEC 2019*, vol. 1402 no. 7, 2019.
- [20] S. Aisa and T. R., "Private Cloud Implementation Using Raspberry Pi for Accessing Personal Data," *Jppi*, 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," vol. 288, no. 1, pp. 1–7, 2018.
- [22] L.M.Q. Abualigah, "Applying Genetic Algorithms to Information Retrieval Using Vector Space Model," *IJCSEA*, Vol. 5 No. 1, pp. 19–28, 2015.
- [23] A. Jain, G. Kulkarni, V. Shah, "Natural Language Processing," *Int. J. Of Comp Sci. And Eng.*, Vol. 6 no.1, pp. 161–167, 2018.
- [24] Bordignon et al, "Natural Language Processing in Business Process Identification and Modeling: A Systematic Literature Review," *SBSI*, pp. 191–198, 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, D.S. Maylawati, T. Mantoro, "Indonesian News Classification Using Convolutional Neural Network," *IJECS*, 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.*, 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. Alaqtash, M.A. Fadhil, and A. F. Al-Azzawi, "A Modified Overlapping Partitioning Cluster Algorithm for Categorical Data Clustering," *Bulletin of Electr. Eng. and Inf.*, vol. 7, no. 1, pp. 55–62, 2018.
- [30] M. Md. Saad, N. Jamil, R. Hamzah, "Evaluation of Support Vector Machine and Decision Tree for Emotion Recognition of Malay Folklores," *Bulletin of Electr. Eng. and Inf.*, vol. 7, no. 3, pp. 479–486, 2018.
- [31] W. Darmalaksana et al, "Latent Semantic Analysis and Cosine Similarity for Hadith Search Engine," *TELKOMNIKA*, vol. 18, no. 1, pp. 217–227, 2020.

## BIOGRAPHIES OF AUTHORS

	<p>Ichsan Taufik, M.T. is a lecturer in Department Informatics at UIN Sunan Gunung Djati Bandung. Ichsana has graduated from Informatics Departement at IAIN Sunan Gunung Djati. He received his M.T. Degree from Institut Teknologi Bandung.</p> <p>Research areas</p> <ul style="list-style-type: none"> <li>• Information System</li> <li>• Algorithm and Programming</li> <li>• Software Engineering</li> </ul>
	<p>Dr. Mohammad Jaenudin, Lecturer in Department Education and Teaching At UIN Sunan Gunung Djati, Bandung-Indonesia.</p> <p>Research areas:</p> <ul style="list-style-type: none"> <li>• Islamic Education</li> <li>• ICT and Islam</li> </ul>
	<p>Fatimah Ulwiyatul Badriyah <sup>1</sup> is a graduate student from Department of Informatics, UIN Sunan Gunung Djati Bandung.</p>



	<p><sup>7</sup> Beki Subaeki is a lecturer in Informatics Department at UIN Sunan Gunung Djati and Sangga Buana University. Research area:</p> <ul style="list-style-type: none"><li>• Information System</li><li>• E-Commerce</li></ul>
	<p><sup>7</sup> Opik Taupik Kurahman is a lecturer in Informatics Department at UIN Sunang Gunung Djati, graduating from the Doctoral program in 2014. Now a researcher in the Field of Islamic Education, ICT and Islam, and the Field of Halal Studies.</p>

## ORIGINALITY REPORT

12%

SIMILARITY INDEX

9%

INTERNET SOURCES

7%

PUBLICATIONS

%

STUDENT PAPERS

## PRIMARY SOURCES

- 1** **Septya Egho Pratama, Wahyudin Darmalaksana, Dian Sa'adillah Maylawati, Hamdan Sugilar, Teddy Mantoro, Muhammad Ali Ramdhani. "Weighted inverse document frequency and vector space model for hadith search engine", Indonesian Journal of Electrical Engineering and Computer Science, 2020** **3%**

Publication
- 2** **bee.i.org** **2%**

Internet Source
- 3** **eprints.whiterose.ac.uk** **2%**

Internet Source
- 4** **eprints.umm.ac.id** **1%**

Internet Source
- 5** **Wildan Budiawan Zulfikar, Mohamad Irfan, Muhammad Ghufron, Jumadi Jumadi, Esa Firmansyah. "Marketplace affiliates potential analysis using cosine similarity and vision-based page segmentation", Bulletin of Electrical Engineering and Informatics, 2020** **1%**

6 [www.ijitee.org](http://www.ijitee.org) 1%  
Internet Source

---

7 [digilib.uinsgd.ac.id](http://digilib.uinsgd.ac.id) 1%  
Internet Source

---

8 M M Chusni, A Hasanah, A M Ghazali, R Zakwandi, R F R Nurzaman, M F Nurfauzan, Y Dirgantara, R D Agustina, A Malik, A Samsudin. "Comparative study of laboratory management capability among physics students at UIN Sunan Gunung Djati and UIN Sunan Kalijaga", Journal of Physics: Conference Series, 2019 1%  
Publication

---

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On