Jurnal Mantik
Volume 4, Number 3, November 2020, pp. 2304-2310 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Using of Exact Queries and Expansion Queries in Searching for Indonesian Translated Al-Quran Verses

Rusydi Umar[1], Ihyak Ulumuddin [2]

[1,2] Master Program of Information Technology
Ahmad Dahlan University Yogyakarta Campus 1 (Head Office) Jl. Kapas 9, Semaki, Umbulharjo, Yogyakarta 55166

E-mail: rusydi_umar@rocketmail.com[1], ihyak1689048035@webmail.uad.ac.id[2]

ARTICLE INFO

Article history:
Received: 01/11/2020
Revised: 10/11/2020
Accepted: 30/11/2020

Keywords :
stemming, quran, query expansion

ABSTRACT

This paper presents a search technique for translating verses in the Al-Quran into Indonesian by developing a query expansion of the hot themes currently being discussed in social media. The hope of finding information in the verses of the Koran is very high as a step to respond and find a holistic religious perspective on the theme being discussed. The method of searching for verses using Indonesian translations is still deemed inadequate. The availability of dictionaries and indexes for the Indonesian translation of the Al-Quran is still very little, most of the verse search indexes use the Arabic book language including the very popular ones are Al-Mu'jam al-Mufahras lialfazhi Al-Qur'an and Fathurrahman li Thalibi Ayatil Qur'an. Researchers have carried out many studies to make certain techniques in searching to find related words in the Al-Quran, including from exact matching, matching according to major themes in the Koran (topic matching), matching with synonyms and a search for explanatory books (tafsir matching). This is done to find relevant verses and provide comprehensive information on the search terms. This research was conducted on 11 random keywords to the entire translation of the Al-Quran and the query results were tested based on the Al-Quran Index Book How to Find Al-Quran Verses compiled by N.A. Baiquni. HE. S Compoundqi, and R.A. Azis. The results of this study indicate that the use of query expansion gets a value of 44% and the exact query gets a value of 33% to find related themes, however the use of query expansion is better than the use of exact queries in the search for the translation of the Indonesian Al-Quran.

1. Introduction

Al-Quran is the highest guide for Muslims in guiding their lives, placing the Qur'an as a guide for human life, so someone must be able to make the Al-Qur'an as solution to problems that occur in their life. Arkoun (1990), a Muslim scientist, tried various readings of the Qur'an which could produce a new interpretation that had never been done by other interpreters. According to him, understanding the text of the Koran is not separated from three elements, namely language, thought and history. He distinguishes between two text classification models, namely first, Arkoun terms the forming text (an-nash al-mu'assis) on the one hand, the second classification, he terms the forming text or hermeneutical text (an-nash al-tafsiri).

A person must be able to search for text, either forming text or text that describes it. Muslims in Indonesia who don't know much about Arabic vocabulary can search based on the Indonesian translation of the Al-Quran. Since the passage of the last revelation, the verses of the Al-Quran have certainly not changed, let alone the addition of words, while Indonesian words continue to experience new vocabulary additions that cannot be found with the exact maching method. From here a new problem arises to find a new words that are not in the Koran.

This research is motivated by the *first* challenge to find a search word that is not exactly stated in the Al-Quran translation. The *second* social media phenomena are on the rise. The impact of social media is extraordinary, the positive impact is that social media can be used as a very effective da'wah media. A 2016 survey on the use of social media for da'wah activities conducted by the Indonesian Internet Service Providers Association (APJII) stated that 82% said respondents agreed, 15% disagreed, and the remaining 3% did not answer. The *third* method of searching for verses using Indonesian translations is still deemed

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

inadequate. The availability of a dictionary and index for the Indonesian translation of the Koran is still very little, most of the verse search indexes are use the Arabic book language

Based on these three backgrounds, the researchers tested the extraction of verses from the Al-Quran that match the related words that were available in the Indonesian thesaurus semantic algorithms in the form of synonyms, hyponyms, hypernimes, and derivative words in Indonesian.

2. Research Methods

Fadi A. (2013) explains that studying the retrieval of the Koran cannot completely depend on finding out the verses that match the exact query word or synonym and stem. The topics of the verses may not be mentioned in the verse exactly as synonyms; Therefore, extracting information from the Koran cannot be considered completely successful if it is unable to retrieve all the relevant verses. Najadat assesses the exact match of the default options and adds a suitable synonym as optional which users can manually add synonyms. After that, a new search option was added which is a suitable topic which classifies the verses of the Quran based on their topic.

The technique of searching for verses in the Koran in a classical way, namely by using a dictionary is sufficient in terms of word accuracy, but does not yet meet the elements of speed, ease and scope of meaning of Fadi A. (2013). If what you want is a specific verse or theme that is in accordance with the current trending topic, then a system is needed that can identify, find and classify these verses of the Qur'an as references and solutions. The process of identifying keywords that are inputted by the user is processed and recognized first, the search technique starts with searching for the basic word then the next stage is looking for related meanings, this process is what researchers mean as Text Mining to find related words. The method for searching the verses of the Quran is as follows:

2.1 Query Expansion

Query Expansion as the stage where the user's initial query statement is enhanced by adding search terms to improve data discovery. Query expansion is rationalized by the fact that the initial query formulations do not necessarily reflect the user's information needs. Application of language dictionaries such as word synonyms, hyponyms, derivatives and others related to the first query. The process of expanding this query is for reformulation of the data discovery intended by the user.

Three types of query expansion are discussed in the literature: Manual, automatic, and interactive (that is, user mediated, semiautomatic, or user assisted). This approach uses multiple sources of search terms and various expansion techniques (Efthimiadis, 1996) Beaulieu and Robertson (1996) argue that the distinction between manual and interactive methods in query formulation is difficult because they both involve human intervention. The difference is that the manual approach does not include gathering consultation while in the interactive approach the request is modified through a feedback process; However, in both cases, help can be sought from other sources, including a dictionary or thesaurus. The results reported here are clearly based on expanding user requests; Thus, this research lies in the "interactive" category.

2.2 Stemming

Stemming is the act of reducing a word to its semantic stem or root (Meadow et al., 2000). Stemming is one of the many ways to improve the quality of query results or what is known as Information Retrival (IR) by rearranging words or sentences in a text document to their root word. Meadow explained that each language has its own algorithm for stemming it, finding a root word in Arabic is not the same as finding a root word in Indonesian. The process of stemming between languages has its own complex characteristics and characteristics.

Stemming is the process of reducing the morphological variant of a word to the usual stem form. Jelita Asian (2007) in her dissertation Jelita explained that Stemming is the process of reducing the morphological variant of a word to an ordinary stem form. Previous research has shown that stemming is language dependent. Although several stemming algorithms have been proposed for Indonesian, there is no consensus that gives better performance. We empirically explore these stemming algorithms, showing that the best algorithms still have scope for at least a five percentage point increase.

2.3 Kateglo

Kateglo is a web-based application that can be accessed on the page: http://kateglo.com. This application was developed by Ivan Lanin, Romi Hardiyanto and Arthur Purnama. The latest Kateglo version is v1.00.20131128 while the first version was released on May 12, 2009

The discovery of Kateglo comes from an abbreviation ka, te, and glo, namely ka stands for dictionary, te from thesaurus, and glo from glossary. The source code license from Kateglo is GPL that is CC-BY-NC-SA. This (still) very simple application programming interface (API) was created to allow developers to take advantage of the data provided by Kateglo. For the initial stages accessible with the following APIs:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

```
http://kateglo.com/api.php?format=[xml|json]&phrase=[lema_yang_dicari]
```

Fig 1 *Kateglo API Request*

In the http://kateglo.com page all contents can be copied, used, distributed, and even adapted freely provided that the source of the content must be included, and is not used for commercialization purposes, and under the same or similar license as the CC-BY license. NC-SA. Data from the Language Center of the Indonesian Ministry of National Education (currently the Language Center of the Ministry of Education and Culture) Pen - marked with "Pusba" or "Language Center" - is copyright of the Language Center and is used in Kateglo with the permission of Pusba.

2.4 Al-Quran Index

The existence of Al-Quran Index function that can make it easier to find verse searches in various ways and verse indexation techniques, either based on textual content or based on contextual meaning. This is done to support efforts to make it easier for a Muslim to find instructions in the Koran. The works of the Al-Quran Index are arranged alphabetically. The index book is very well known and widely spread in Indonesia is Mu'jam al-Mufahras li Alfaz al-Quran al-Karim by Muhamad Fuad Abdul Baqi, Fath al-Rahman li Talib Ayat al-Quran by Ilmi Zadeh Faidullah, these two books are index Al-Quran In Arabic. Apart from this index, there is also an index presented in Indonesian, namely the Al-Quran Index How to Find Al-Quran Verses compiled by N.A. Baiquni. HE. S Compoundqi, and R.A. Azis.

In the introduction, N.A. Baiquni (1996) explains that the preparation of the Al-Quran Index book is intended to facilitate the search for translations of Al-Quran verses published by the Al-Quran Translator / Interpreter Organizing Foundation of the Ministry of Religion of the Republic of Indonesia (now the Indonesian Ministry of Religion). The source of the compilation of the book is Mu'jam al-Mufahras li Alfaz al-Quran and Fath al-Rahman li Talib Ayat al-Quran. This book is also a tool for measuring the accuracy of querying verses of the Koran with the stemming method and expanding the query by searching for word relations using the Kateglo service.

2.5 Testing

It is important to measure the performance of the query results. The data generated from the query will describe how well a technique is in searching for data in a table in the database. Confusion matrix is a technique that can be used to measure the performance of a query. According to E. Prasetyo (2012) the results of the confusion matrix calculation are information that compares the classification results obtained from the query with the classification results that should be.

M. Sokolova (2009) explains that based on the number of class confusion matrix outputs, the classification system is divided into four types, namely binary classification, second multi-class, third multi-label and fourth hierarchical. In binary classification, input data is divided into one of these two classes, the four types of classification in the confusion matrix can be described in 4 quadrants as in the following Fig:

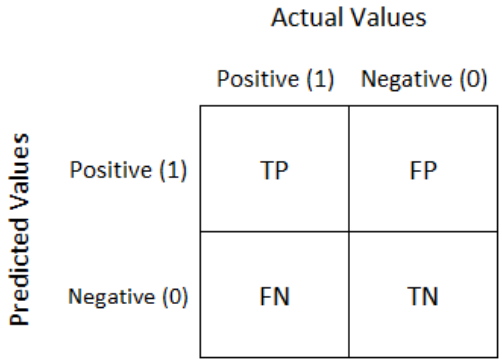

Fig 2 Confusion Matrix Classification

To measure the accuracy and retrival value of each search result from each step the author uses the equation algorithm as follows:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$Presisi = \frac{TP}{FP+TP} * 100\% \quad (2)$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad (3)$$

Where:

- TP (True Positive): the number of positive data results from the correct classification.
- TN (True Negative), the number of negative data results from the correct classification.
- FN (False Negative, the number of negative data is wrongly classified.
- FP (False Positive), the number of positive data results from wrong classification.

3. Implementation

In this research , added a new search option that can improve the search for verses of the Qur'an that have the same topic by adding related queries from the synonym word thesaurus service regardless of whether the contents contain the search term correctly. This study only focuses on testing the Search for Al-Quran Verses and Jalalin interpretations in Indonesian using word relations and Query Expansion with the following algorithm:

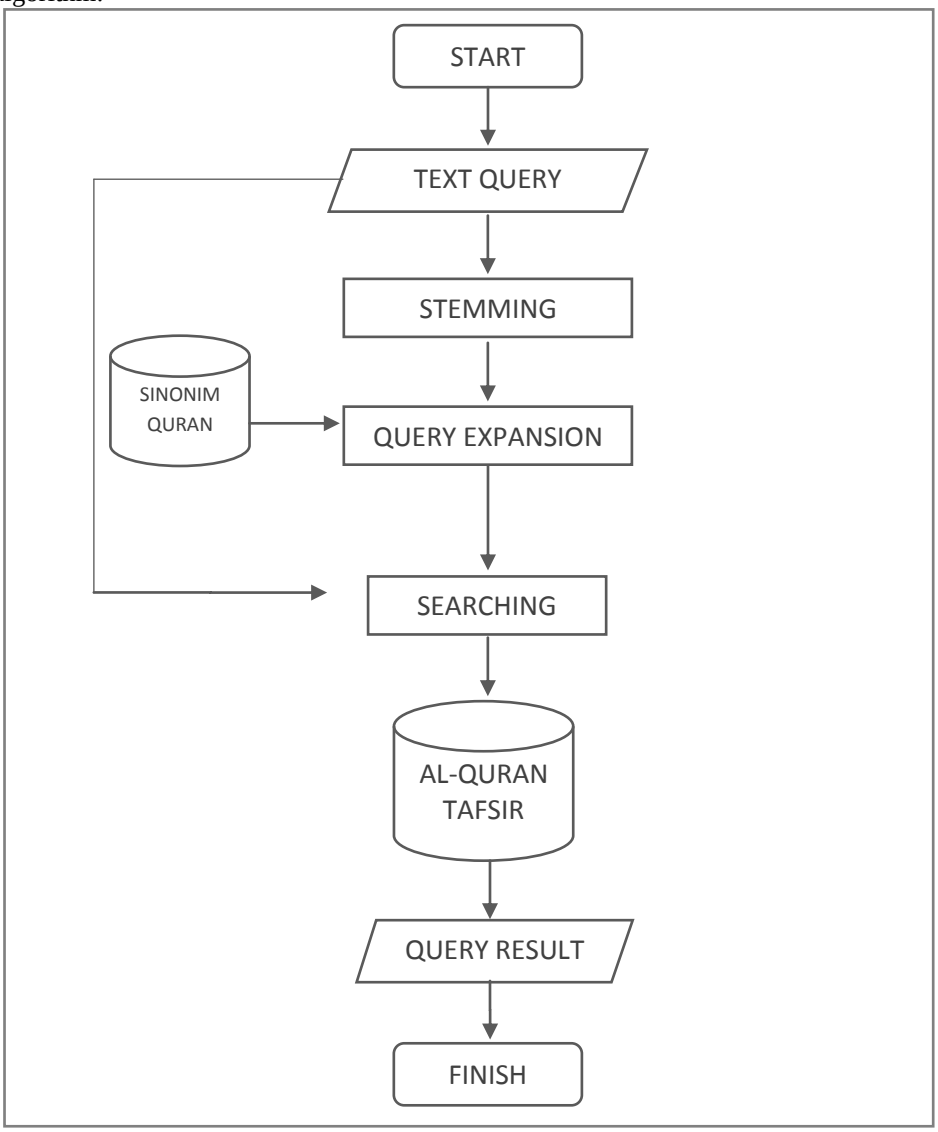

Fig 3 *Diagram Flow Query Expansion Al-Quran*

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The system development software that will be used in this research is PHP version 5.6 using the MySql version 10.1.21-MariaDB data base. As for the process of stemming search words and translating the Indonesian Al-Quran, the author uses the PHP library Sastrawi, this PHP library is arranged in an Object Oriented Design based on the Nazief and Adriani Algorithms, Asian J. 2007, Arifin, A.Z., I.P.A.K. Mahendra and H.T. and A. D. Tahitoe, D. Purwitasari. 2010.

Stemming is used to solve the problem of suffixes and prefixes not being considered properly by traditional search users. The next implementation stage of the research is after entering each query individually and the stemming process, then applying the cumulative search options as follows: Exact maching, synonym matching, Google treds matching topics against Al-Quran translations.

The testing is applied using 11 keywords as shown in table I: The results of query expansion are as in the following table:

Table 1
Synonym Word Results From Kateglo.com

User input	Related Words (from Kateglo)	Number of verses
haid	bercemar kain|bocor|datang bulan|datang kotor|kedatangan tamu| melihat bulan|membawa adat|membawa bulan|membawa cemar| mendapat bulan|mendapat kain camar|mendapat kain kotor|mens| menstruasi|merah|sakit bulan|uzur	17
hijrah	memindahkan|mengungsikan|menyingkirkan|eksodus|imbit|hijrah	6
islam	selam	1
judi	spekulasi	1
junub	janabah	1
kawin	bersetubuh|mengawin|menikah|perkawinan|berjunjungan|naik nobat|mawin|berbaur|baur|berbakak|berkeluarga|berjodoh|duduk	13
nasib	takdir|tuah|keberuntungan|bintang|peruntungan|tulisan nasib| perputaran dunia|hoki|nyawa|kismat|garis hidup|bilangan| suratan|garis bujur hidup|untung|suratan takdir	17
pasukan	bak|baskom|bejana|belanga|jambang|capah|pasu	7
puasa	saum|siam|ifah|upawasa	4
rencana	-	0
teladan	contoh|turutan|kaca|cermin|ideal|arketipe|tepa|iktibar|acuan| panutan|ikutan|suri teladan|patron|terdahulu|idaman|anutan	16

Based on table 1 the basic words that have the most synonyms are menstruation, namely 17 words, while the word " rencana " has no synonyms. Expansion of the query using the synonym word above with the following results:

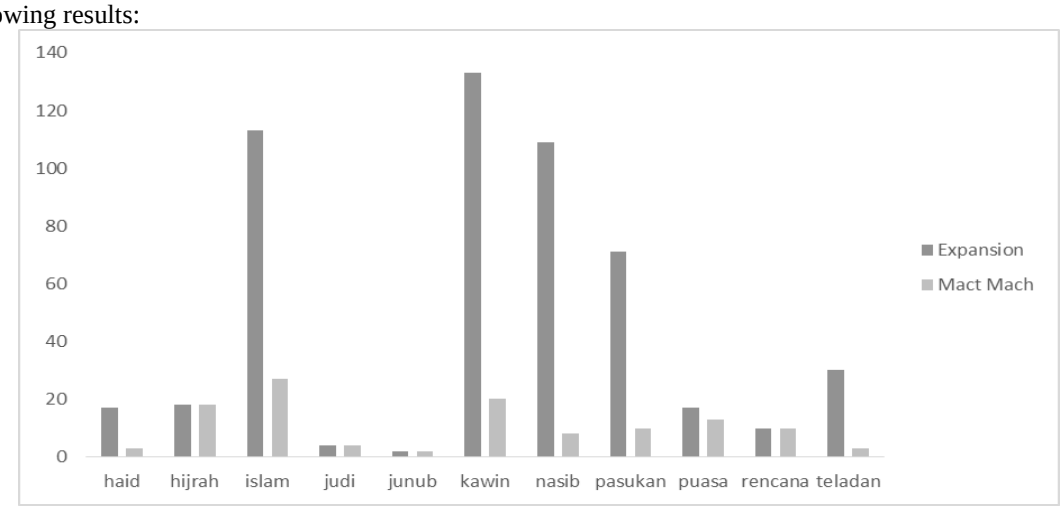

Fig 4 Number of Verses Query Results Including Synonyms

Based on Fig 1.4, the average number of verses found using the Expansion Query is more than the exact query with a difference of 37 verses. Like the word "nasib" which has 17 synonyms, found only 10 verses with exact query and 109 verses with query expansion. The number of verses found is not directly proportional to the similarity of the themes presented in the Al-Quran Index Dictionary, the number of verses produced from synonym words is almost not recorded in the Index Dictionary. This also happened to the *exact query* of several verses which were found not all of them were listed in the Index Dictionary.

Testing is done by grouping the following paragraphs:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

a. TP: the number of verses of the Koran is predicted to exist, indeed the query results were also found
b. TN: the number of verses of the Koran is predicted not to exist and in fact they do not exist
c. FP: the number of Al-Quran verses that are predicted to be positive, the query results are not found
d. FN: the predicted number of Al-Quran verses does not exist, but is / is in the query results

The testing result which using exact query the lowest f-score is 11% to the highest is 100%, while with query expansion the lowest f-score is 1.8% to the highest result is 100% as in the following table:

Table 2
Exact Query Test Results

Keywords	Index Dictionary	TP	TN	FP	FN	total	Precission	Recall	F-Score
haid	1	1	6233	0	2	3	100%	33%	50.0%
hijrah	16	11	6218	5	7	18	69%	61%	64.7%
islam	19	4	6209	15	23	27	21%	15%	17.4%
judi	4	3	6232	1	1	4	75%	75%	75.0%
junub	2	2	6234	0		2	100%	100%	100.0%
kawin	20	9	6216	11	11	20	45%	45%	45.0%
nasib	3	1	6228	2	7	8	33%	13%	18.2%
pasukan	8	1	6226	7	9	10	13%	10%	11.1%
puasa	4	4	6223	0	9	13	100%	31%	47.1%
rencana	2	2	6226	0	8	10	100%	20%	33.3%
teladan	3	2	6233	1	1	3	67%	67%	66.7%
Total	**82**	**40**		**42**	**78**	**118**	**66%**	**43%**	**48%**

Table 3
Test Results for Expansion Queries

Keywords	Index Dictionary	TP	TN	FP	FN	total	Precission	Recall	F-Score
haid	1	1	6219	0	16	17	100%	6%	11.1%
hijrah	16	11	6218	5	7	18	69%	61%	64.7%
islam	19	4	6123	15	109	113	21%	4%	6.1%
judi	4	3	6232	1	1	4	75%	75%	75.0%
junub	2	2	6234	0		2	100%	100%	100.0%
kawin	20	9	6103	11	124	133	45%	7%	11.8%
nasib	3	1	6127	2	108	109	33%	1%	1.8%
pasukan	8	2	6165	6	69	71	25%	3%	5.1%
puasa	4	4	6219	0	13	17	100%	24%	38.1%
rencana	2	2	6226	0	8	10	100%	20%	33.3%
teladan	3	2	6206	1	28	30	67%	7%	12.1%
Total	**82**	**41**		**41**	**41**	**524**	**67%**	**28%**	**33%**

4. Conclusion

The performance of Exact Query and Query Expansion is depends on the keywords entered by user, if the keywords are unique and contained explicitly in the Indonesian translation of the Al-Quran, the performance of the query results can reach 100% otherwise if the keywords are classified as explicit, then performance queries are so low that they even reach 1.8%. The use of Query Expansion to search for Al-Quran verses using Indonesian translation is better than using Exact Query. The average F-score of Query Expansion is 48% while the average F-score of Exact Query is only 33%.

5. References

[1] Arkoun, *Al-Fikr al-Islam Naqd wa Ijtihad, Terjemah Hasyim Salih* (London : Dar al Saqi, 1990), hlm. 234
[2] Asosiasi Penyelenggara Jasa Internet Indonesia, 2016, *Survey Internet APJII*
[3] Muhammad Fuad 'Abd Al Baqi, *Al Mu'jam Al Mufahras Li Alfazh Al Quran Al Karim* Darul Fikr, Mesir.
[4] Efthimiadis, E.N. (1996). *Query expansion. In M.E. Williams (Ed.), Annual review of information science and technology (pp. 121–187).* Medford, NJ: Information Today
[5] Beaulieu, M., & Robertson, S. (1996). *Evaluating interactive systems in TREC. Journal of the American Society for Information* Science, 47(1), 85–94.
[6] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1 ed. Yogyakarta: Andi Offset, 2012.
[7] https://kateglo.com/
[8] Dr. Fadi A. O. Najadat (2013). *Approaches to Retrieve Verses of the Holy Quran Based* on Full Meaning Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences

[9] B. Hammo, "*Towards enhancing retrieval effectiveness of search engines for diacritisized Arabic documents*," Information Retrieval, vol. 12, 2009, pp. 300-323.
[10] C. T. Meadow, B. R. Boyce, and D. H. Kraft. *Text information retrieval systems.* Academic Press, San Diego, California, second edition, 2000.
[11] Jelita Asian, *Effective Techniques for Indonesian Text Retrieval*, A thesis submitted for the degree of Doctor of Philosophy School of Computer Science and Information Technology, Science, Engineering, and Technology Portfolio, RMIT University, Melbourne, Victoria, Australia, 2007
[12] Sokolova dan G. Lapalme, "*A systematic analysis of performance measures for classification tasks*," *Inf. Process. Manag.*, vol. 45, no. 4, hal. 427–437, 2009

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.