

# Comparative Study for Better Result on Query Suggestion of Article Searching with MySQL Pattern Matching and Jaccard Similarity

Komang Rinarta<sup>1</sup>, Wayan Suryasa<sup>2</sup>  
Departement of Computer System  
STIKOM Bali  
Bali, Indonesia

<sup>1</sup>[komangrinarta@gmail.com](mailto:komangrinarta@gmail.com), <sup>2</sup>[suryasa@stikom-bali.ac.id](mailto:suryasa@stikom-bali.ac.id)

**Abstract-** Article searching is the process to find the appropriate article based on user description or keywords. For the traditional article searching, user types keywords then press the button and the result presents, but unfortunately the process needs more time and the search result is not accurate based on user keywords. Query suggestion makes it easy to search the data without submitting the data in a traditional way. There are several methods that can be used in query suggestion thus it is necessary to know which method should be used. This paper describes the comparison between MySQL pattern matching and Jaccard similarity for query suggestion with the parameters, they are processing time, the proximity of keywords with the suggestion, the proximity ratings and sorted from nearest data, the strength of data retrieval based on the number of words entered, and the strength of data retrieval from a number of data that can be suggested. MySQL pattern matching and Jaccard similarity can be used as query suggestion with the processing time is 2.5 milliseconds for MySQL pattern matching and 7 milliseconds for Jaccard similarity with 492 articles. Query suggestion using MySQL pattern matching takes an average of 1.8 milliseconds for 100 articles, 2 milliseconds for 200 articles, 2.175 milliseconds for 300 articles, while Jaccard similarity 3.175 milliseconds for 100 articles, 4.025 milliseconds for 200 articles, 5.05 milliseconds for 300 articles. Proximity ratings for MySQL Pattern Matching cannot be calculated since there is no counting process, but in Jaccard similarity the system displays 0.8 for jaccard index and display 1 for jaccard index if the keywords are 100% similar to the data on database. The result of this study is Jaccard similarity query suggestion produces more accurate search result but has a longer processing time than MySQL pattern matching.

**Keywords:** Article searching, query suggestions, MySQL Pattern Matching, Jaccard Similarity

## I. Introduction

Searching the data through the Internet expand along with the hardware developments support. Data searching can be done using some keywords based on previous searches. In addition, the search process can also use reference data in the database so that the data sought in accordance with the data in the database.

Dynamic query suggestions have been widely applied to information retrieval website to facilitate the process of finding the desired information. Dynamic query suggestion cannot be separated from the form autocomplete on a input form on the web. Autocomplete is a pattern which first appeared in the help function of desktop applications, where users enter text into the box and then suggestions will appear automatically; [1] Form autocomplete is featured in an input form to display data in accordance with the data typed by the user. Autocomplete query suggestion feature is often found in the search engines and also the input form is found in the website.

Google, Bing, and Yahoo! have consistently been reported as the three major web search engines. Together, they share around 96% of the search market, namely Google (66.9%), Bing (18.1%) and Yahoo! (11.1%). Over the past few years, these three search engines have introduced dynamic term/query suggestions functionalities to support user in query formulation, reformulation or query expansion; [2] Several methods can be used in a query suggestions, they are MySQL Pattern Matching, Jaccard Similarity and many more methods that can be implemented. From a few simple concept of dynamic query suggestion, need to know a better method used in dynamic query suggestion.

Jaccard similarity is an algorithm to compare one document to another based on string that they have. Jaccard similarity is usually used in document comparison to find the similarity value of the documents. Based on this ability, Jaccard similarity is used in dynamic query suggestions to give some suggestions to user in article searching process.

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

with :

J(A,B) = jaccard similarity

A = document/set of words

B = document/set of words

This paper will discuss about the introduction on the first chapter, then the related works in the second chapter, research method in the third chapter, result and analysis in the fourth chapter and in the fifth chapter will discuss about conclusion.

## II. Related Work

Banowosari et al in 2014 in his research, [3] stating that the feature autocomplete suggestion is a feature that is implemented on a web browser and a search engine that allows web browsers or search engines to provide search suggestions when just a few words typed in the search field or the address bar. The features in the Google autocomplete display some search suggestions based on: the location of the geographic and language used in the search, popular keywords that are often searched, and the search history of the user. Google Autocomplete also can correct the incorrect spelling when a user enters a keyword in the presence of spelling correction feature.

Yang JM et al in 2008 in his research, [4] stated that the query suggestion feature has been considered to be possessed by the search engines. Existing query suggestion approaches can be classified into two categories based on the data they used, the first one is log based and the second one is search result-based. Query suggestion based on query log approach and search result approach have their own merits and weaknesses, which make them suitable for different kind of queries.

S Plansangket and J Q Gan in 2015 in his research, [5] stated that query expansion is a technique to expand the query with the related words and is widely used for query suggestion. Query expansion aims to improve the overall recall of the relevant documents. Query reformulation or dynamic query suggestion is more complex than query expansion, which forms a new query using specific models.

Sugiyanto, B Surarso, A Sugiharto in his research, [6] stated that Cosine and Jaccard, indicating that both have high performance on average, but when compared to the level of accuracy, the Cosine method has higher accuracy than Jaccard method with the accuracy rate 0.949808 for Cosine method and accuracy rate 0.949077 for Jaccard method.

S. Niwattanakul, et al in 2013 in his research, [7] stated that Jaccard similarity used in information retrieval and calculation of similarity coefficient by using the prologue language using a level of precision, recall and F-measure. The results showed that the test method by Jaccard coefficient can perform well in measuring the similarity of words when comparing with each letter of the word. Particularly, each letter can switch positions and counted as the same words. Nevertheless, this method is not able to detect the over-type the words in the data sets. In conclusion, Jaccard similarity coefficient is suitable sufficiently to be employed in the word similarity measurement. In efficiency measurement, the program performance can deal appropriately with high stability and when the failure occurred spelling mistake.

## III. Research method

Two article search applications are created using MySQL pattern matching and Jaccard similarity then the result of the application is compared with each other. Parameters to be compared are in terms of processing time, the proximity of keywords with the suggestion, the proximity ratings and sorted from nearest data, the strength of data retrieval based on the number of words entered, and the strength of data retrieval from a number of data that can be suggested.

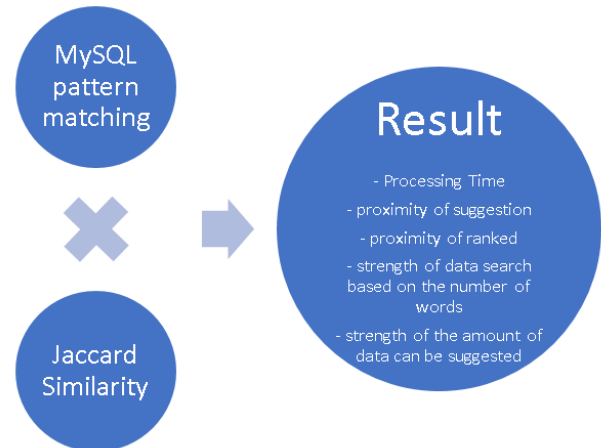


Figure 1. The concept of study comparison

The use of MySQL pattern matching uses clause "like" while Jaccard similarity changes the mathematical formula Jaccard into a query on MySQL. The MySQL pattern matching that is used is ;

```
SELECT * FROM makalah where judulMakalah like '%signal%' LIMIT 0, 10
```

And the Jaccard similarity is shown as follow ;

```
SELECT *, (if (locate('signal', judulMakalah)>0,1,0)+0)/(LENGTH(judulMakalah) - LENGTH(REPLACE(judulMakalah, ' ', '')) + 1 +if (locate('signal', judulMakalah)>0,0,1)+0) as score FROM makalah order by score desc LIMIT 0, 10
```

Implementation of these two algorithms is in accordance with the flow diagram below:

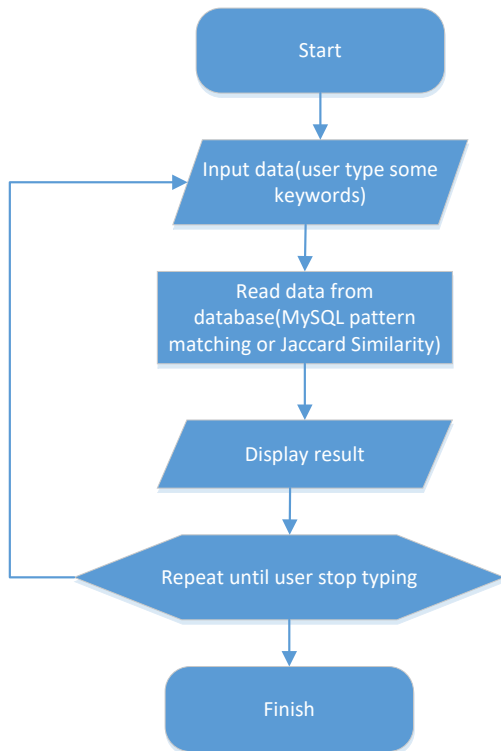


Figure 2. Simulation method

#### IV. Result and analysis

When using MySQL pattern matching and Jaccard similarity, users only need to enter a few letters or words to search and then the system will provide search suggestions that can be used by the user automatically. System will stop providing search when the user stops typing in the input field area or when the user selects the suggestion given.

`SELECT * FROM makalah where judulMakalah like '%signal%' or judulMakalah like '%processing%' LIMIT 0, 10`

When a user enters the word "signal processing", then the system looking for a title of the paper containing the word "signal" both the signal in front of, behind, or in the middle or "processing" whether it be in front, middle and behind. As for the Jaccard similarity, it will search for words that have a member of the set in accordance with the "signal processing" by the similarity values derived from calculating the Jaccard similarity index as follows

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

$$= \frac{|signal\ processing|}{|SIGNAL\ PROCESSING\ OF\ RADAR\ INDERA|} = \frac{|2|}{|5|} = 0,4$$

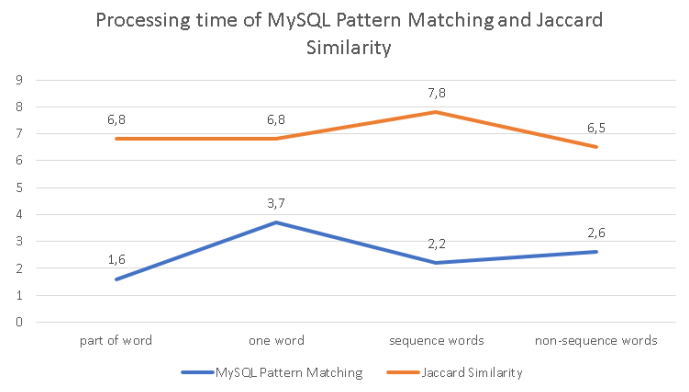


Figure 3. Graph of processing time

Tabel I.  
Proximity of searching suggestions based on keyword entered

Input	Top 5 result	
	MYSQL pattern matching	Jaccard similarity
Sig	5 suggestions with "sig"	5 suggestions with "sig"
Signal	2 suggestions with "signal"	2 suggestions with "signal"
Signal processing of radar	5 suggestions with "of"	1 suggestion with "Signal processing of radar" and 4 suggestions with "of"
Processing signal radar of	5 suggestions with "of"	1 suggestion with "Signal processing of radar" and 4 suggestions with "of"

Tabel II.  
Strength of suggestion rank based on keyword entered

Input	Top 5 result	
	MYSQL pattern matching	Jaccard similarity
Sig	-	0.2
Signal	-	0.2
Signal processing of radar	-	0.8
Processing signal radar of	-	0.8

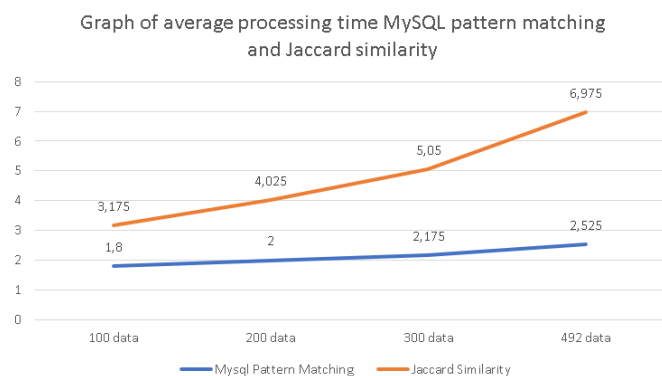


Figure 4. Strength of searching process based on numbers of data

## V. Conclusion

This research concludes that MySQL pattern matching and Jaccard similarity can be used as query suggestion with the processing time is 2.5 milliseconds for MySQL pattern matching and 7 milliseconds to Jaccard similarity with 492 articles. The application tested with part of word will produce results that correspond to the data sought for both algorithms. The application tested for one word has been approached with the data that want to be searched. Testing to find sequence words will produce data that is not relevant to the MySQL pattern matching and generates the appropriate data on Jaccard similarity. This is due to the MySQL pattern matching only sorts data by the data index while Jaccard data is sorted by the Jaccard similarity index. To search article using MySQL pattern matching, it takes an average of 1.8 milliseconds for 100 articles, 2 milliseconds for 200 articles, 2.175 milliseconds for 300 articles. While the time required for Jaccard similarity is 3.175 milliseconds for 100 articles, 4.025 milliseconds for 200 articles, 5.05 milliseconds for 300 articles. Processing time on MySQL pattern matching is affected by the position of the data in the database while Jaccard similarity affected by the number of keywords and the number of words contained in the input field. Suggestion rank for MySQL Pattern Matching cannot be calculated since there is no counting process, but in Jaccard similarity the system displayed 0.8 for jaccard index and will display 1 if the search text is 100 % similar to the data on database. The entire results of the Jaccard similarity query suggestion produce more accurate search result but has a longer processing time than MySQL pattern matching.

## References

- [1] Y. Primadani, "Simulasi Algoritma Levenshtein Distance Untuk Fitur Autocomplete Pada Aplikasi Katalog Perpustakaan," Universitas Sumatera Utara, Medan, 2014.
- [2] A. Shiri and L. Zvyagintseva, "Dynamic Query Suggestion in Web Search Engines: A Comparative Examination," in *42nd Annual Conference of the Canadian Association of Information Science & Inaugural Librarians' Research Institute Symposium Brock University. St. Catharines, Ontario*, 2014.
- [3] L. Y. Banowosari, A. Darmawan, K. Kurniawan and M. Mitchell, "Analisis Pada Fitur Autocomplete Suggestion Dan Semantik Pada Pencarian Di Mesin Pencari Google," in *Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2014)*, Jakarta, 2014.
- [4] J.-M. Yang, R. Cai, F. Jing, S. Wang, L. Zhangy and W.-Y. Ma, "Search-based Query Suggestion," in *Proceedings of the 17th ACM conference on Information and knowledge management. ACM 2008*, California, 2008.
- [5] S. Plansangket and J. Q. Gan, "A query suggestion method combining TF-IDF and Jaccard Coefficient for interactive web search," *Artificial Intelligence Research* 4, vol. 4, p. 119, 2015.
- [6] Sugiyamto, B. Surarso and A. Sugiharto, "Analisa Performa Metode Cosine dan Jaccard Pada Pengujian Kesamaan Dokumen," *Jurnal Masyarakat Informatika*, vol. 5, pp. 1-8, 2014.
- [7] S. Niwattanakul, J. Singthongchai, E. Naenudorn and S. Wanapu, "Using Jaccard Coefficient For Keywords Similarity," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I*, Hong Kong, 2013.
- [8] D. Hawking and K. Griffiths, "An enterprise search paradigm based on extended query auto-completion. Do we still need search and navigation?," in *ADCS '13 Proceedings of the 18th Australasian Document Computing Symposium (pp. 18-25)*, New York, 2013.