Computing Time Reduction Using Personalized Web Search Applications

Syaharulla Disa[1], Sitti Arni[2], Lili Sugiarti[3]

[1,3]Department of Information Engineering, Sekolah Tinggi Manajemen Informatika Komputer AKBA Makassar, Indonesia. 90245
[2]Department of Information Systems, Sekolah Tinggi Manajemen Informatika Komputer Profesional Makassar, Indonesia. 90233

{syahruldisa181073@gmail.com[1], arnist1903@gmail.com[2], lilykimss99@gmail.com[3]}

Abstract. The aim of the study was to reduce the search time and to display appropriate result by using personalized web application. Search engine is used for searching based on keyword entered but the indicated result are not sometimes appropriate and it take a long time. K-Means algorithm is used for classifying the data obtained from search engines by dividing into several clusters with the same radius. Genetic algorithm are used for optimizing the cluster result in the form of ranking URLs in sequence. The result indicated that searching through a web search application using the K-Means algorithm and genetis can improve the information quality and the faster process. The use of web search application can reduce the search time of 0.00044183 seconds. The use of web search application can improve the accuracy and the speed of searching process.

Keywords: *Web Search, K-Means, Genetics.*

1 Introduction

Search Engine is a tool to find a variety of information available on the web with the help of an internet connection. In search of a document, users sometimes find results that do not conform to what is expected. This causes the user to spend more time to search due to the accuracy of the data shown is not maximized.

Personalized web search is a technique that aims to improve the accuracy by customizing web search based on the information required by the user. Web data is not always labeled, incomplete, inaccurate, heterogeneous and dynamic therefore the techniques of soft computing such as Fuzzy Logic, Neural Networks and Genetic Algorithms have been applied in the web mining with the aim to improve the effectiveness of Personalized Web Search (Chawla, 2016).

Methods to increase the effectiveness of Personalized Web Search have been proposed by researchers such as Fuzzy Logic (Akhlaghian, Arzanian, and Moradi 2010) were used for Personalized Web Search based on user profile, Neural Networks (Chih et al., 2009) that implement Personalized Web Search in commercial web and Genetic algorithm (Boughanem, Chrisment and Tamine 2002).

Genetic algorithms are used for optimization problems and search is an evolutionary computational methods to solve optimization problems. Genetic Algorithm is very useful when applied to applications that are in desperate need of robustness and global optimization. Personalized Web Search will be made in this study using a genetic algorithm to optimize the ranking of a URL. In a very large search scope, the completion of the optimization problem would be to use a high computational time. This leads to the optimization problem will be more difficult to resolve (Chawla, 2016).

This study proposes a solution to solve the problems on Genetic Algorithm by adding a K-Means clustering method. K-Means method will divide the data into several clusters. Genetic Algorithm calculations will be performed on the data that has been divided into several parts.

2 Research Methods

This type of research conducted in this study is a quantitative research experiment. Ekpserimen research is a form of research by quantitative or objective approach.

The data used in this research is divided into two parts, namely the primary data and secondary data. Primary data is data obtained directly results from Google Search. While the secondary data obtained from the research that has been done before through the literature study conducted by researchers. The data will be used as a dataset obtained online. Data processing is performed offline using web-based applications created using programming language PHP and MySQL database. The data set in this study is a text-based search results that do use Google Search interface. Search Results obtained by including a wide range of keywords (keywords) into the Google search box. The data set used is the data set that is in the domain of academic, economic, and entertainment. The dataset will be used in this research were 30 datasets. The dataset was obtained by performing a search using Google Search.

Research conducted by Chawla (2016) proposed method is divided into two phases, namely Phase I and Phase II. In phase I, the process that occurs when you're offline. Vectors keywords from Query Sessions will be generated from the Web Query Sessions using Information Scent and the content of the URL that was clicked. Vectors keywords from Query Sessions will be clustered into groups qualified URL similarity with the need information. Genetic algorithms are applied to the URL that has been clustered to identify optimal ranking of the URL associated with each cluster.

Gradually description of the proposed method is based on the representation of chromosomes is divided into two phases:

Phase I: Offline Preprocessing
1. The data set collected from the internet in preprocess to get Query Sessions.
2. For each URL that was clicked on Query Sessions, Information Scent calculated metric is a measure of the relevance of the URL that was clicked with the information needs of Query Sessions.
3. Sessions in Vector Keyword Query-generated from query sessions by using Information Scent and content of the URL is clicked.
4. Algorithm K-Means clustering vectors used for keyword query sessions.
5. Each cluster associated with mean vector j clust_meanj keyword.
6. For each cluster j is a list of URLs are clicked in List Lj where Information Scent> = threshold.
7. For the individual cluster j Genetic Algorithm applied to obtain optimal ranking of URLs that have been through the process of the cluster in the List Lj related to cluster j to

determine an optimal top rank m of the URL is clicked in the list Lj associated with each cluster and is represented by ORj (optimal Ranking j).

Phase II: Online Processing

1. Search query entered by the user is used to select the cluster j to which one has a lot of similarities in the information needs of a keyword based query input using the cosine similarity measure.
2. Ranking list of URLs associated with a cluster j ORj will be selected.
3. ORj represented on the user selected.
4. If the user makes a request for further results page:
    a. Modeling part of the information needs of the user profile that ongoing use of Information Scent and URLs in the content section and a user profile containing the keyword vector current_usersessionvektort session.
    b. Select the cluster j to which one has the most in common with the information requirements associated with current_user vektort session.
    c. Return to step 2.

Broadly speaking, this research can be made into a fishbone diagram form as in Figure 1 below.
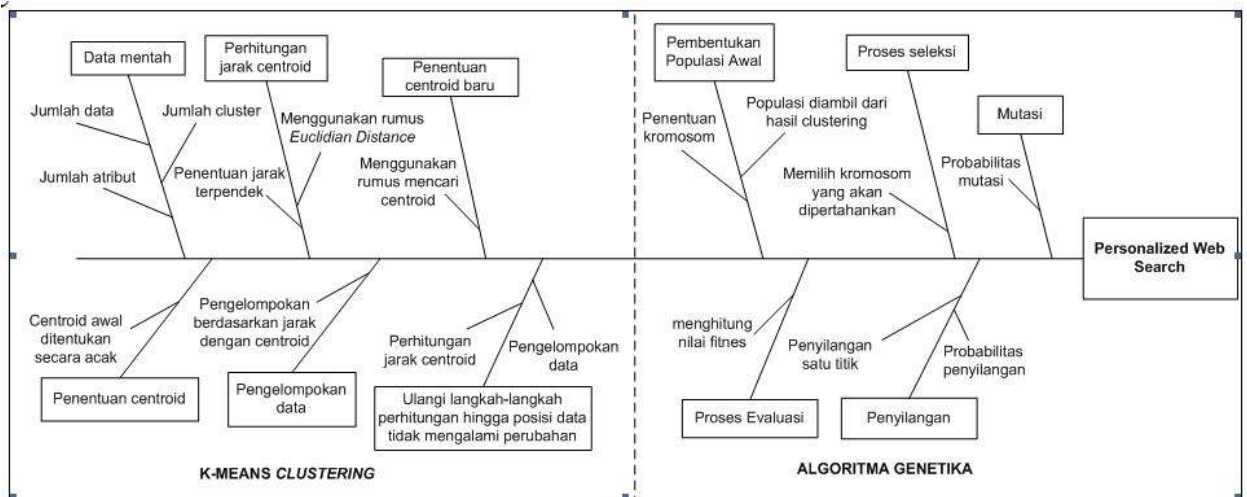
Figure 1. Proposed Method Fishbone

Figure 1 shows the process undertaken to get a Personalized Web Search proposed in this study. Personalized Web Search proposed using the K-Means clustering method and Genetic Algorithm.

3 Results and Discussion

The page was first performed at the time the program starts is the main page contains a menu that presents all program access. Tgersedia menu Personalized pages including Web Search, Genetic Algorithm, K-Means, Data Personalized Web Search page URL will appear when the button is clicked Personalized Web Search. Personalized web search page there is a text box to enter a keyword to search then click the search button. Figure 2 displays one of the search results.

Figure 2. Search Results Page

Genetic Algorithm Selection button will display a page of search optimization and genetic algorithm. At the prompt enter the keyword to be searched and click the Run button next GA optimization results will appear along with captions. Figure 3 displays the results of the optimization.

Figure 3. Results Optimization

When the button labeled with the K-Means on the main page is clicked, it will show a page Personalized Web Search with K-Means. Search results appear when a keyword has been entered and the search button is clicked. Figure 4 displays the search results with K = Means.

Figure 4. Results with K-Means

When the button labeled with the data on the main page URL is clicked, it appears data input page URL. Figure 5 displays the URL data that has been entered into the database through a data input form the URL. Figure 5 displays the URL data.

			Daftar Data URL Pencarian			
			Jumlah Data : 23			
ID	Judul	Deskripsi	URL	Kata Kunci	Nilai	
1	Membuat Fasilitas Buku Tamu dengan PHP, MySQLi dan Bootstrap	Cara membuat buku tamu di website dengan bahasa pemrograman PHP dan database MySQL dan Bootstrap	http://tutorialweb.net/membuat-fasilitas-buku-tamu-dengan-php-mysqli-dan-bootstrap	buku tamu/guestbook dengan php, mysql dan bootstrap	91	[edit] [delete]
3	5 htaccess Trik Yang Harus Anda Ketahui	Trik terbaik dengan menggunakan htaccess	http://tutorialweb.net/5-htaccess-trik-yang-harus-anda-ketahui	trik htaccess	89	[edit] [delete]
4	Membuat Halaman Dinamis Dengan PHP	Bagaimana membuat halaman dinamis dengan PHP	http://tutorialweb.net/membuat-halaman-dinamis-dengan-php	halaman dinamis dengan php	88	[edit] [delete]
4	Membuat Countdown Timer Redirect dengan jQuery	Tutorial bagaimana cara membuat count down timer dengan jQuery	http://tutorialweb.net/membuat-countdown-timer-redirect-dengan-jquery	count down timer jquery	85	[edit] [delete]
6	Simple & Pure CSS Dropdown Menu	Membuat dropdown menu dengan CSS murni yang sederhana	http://tutorialweb.net/simple-pure-css-dropdown-menu	dropdown menu css	45	[edit] [delete]
7	Cara Installasi Joomla	Step by step installasi CMS Joomla	http://tutorialweb.net/cara-installasi-joomla	installasi cms joomla	98	[edit] [delete]
8	Bersama Membangun Negeri - Dengan Indonesia Mengajar	Menjamin pendidikan generasi muda. Ayo bergabung Indonesia Mengajar!	http://indonesiamengajar.org	Pendidikan Indonesia	88	[edit] [delete]
9	Hardiknas 2017 : Pendidikan Indonesia Semakin Baik	Kita harus menghargai bagaimana pendidikan di Indonesia ini semakin hari semakin baik Pendidikan Indonesia	http://belmawa.ristekdikti.go.id/2017/05/04/hardiknas-2017-pendidikan-indonesia-semakin-baik	Pendidikan Indonesia	67	[edit] [delete]
10	KILAS BALIK DUNIA PENDIDIKAN DI INDONESIA - USAID	Dunia pendidikan di Indonesia masih memiliko beberapa kendala yang berkaitan dengan mutu pendidikan	http://www.prestasi-iief.org/index.php/id/feature/65-kilas-balik-dunia-pendidikan-di-indonesia	Pendidikan Indonesia	77	[edit] [delete]
11	Pendidikan Indonesia di Bawah Ethiopia, Kemendikbud: Kita Perbaiki	Hasil penelitian Jaringan Pemantau Pendidikan Indonesia (JPPI) menyatakan indeks kualitas pendidikan di Indonesia berada di	https://news.detik.com/berita/d-3454747/pendidikan-indonesia-di-bawah-ethiopia-kemendikbud-kita-perbaiki	Pendidikan Indonesia	90	[edit] [delete]
12	Pendidikan di Indonesia - Wikipedia bahasa Indonesia	Pendidikan di Indonesia adalah seluruh pendidikan yang diselenggarakan di Indonesia, baik itu secara terstruktur maupun tidak terstruktur	https://id.wikipedia.org/wiki/Pendidikan_di_Indonesia	Pendidikan Indonesia	81	[edit] [delete]
13	Macam Tradisi & Budaya Jepang, Budaya Tradisional Hingga Modern	http://uskajepang.com/macam-macam-budaya-jepang	Sebelum menelaah lebih lanjut macam-macam budaya Jepang, ada baiknya terlebih dahulu kita ketahui apa itu makna budaya.	Budaya Kebudayaan Jepang	95	[edit] [delete]
14	11 Budaya Jepang yang pasti tak banyak kamu tahu	Budaya-budaya itu merupakan adat istiadat masyarakat Jepang sejak dulu dan ditambah dengan kebiasaan masyarakat modern	https://www.brilio.net/life/11-budaya-jepang-yang-pasti-tak-banyak-kamu-tahu--150325n.html	Budaya Kebudayaan Jepang	79	[edit] [delete]
15	Jepang Jadi Negara Maju Karena Kebiasaan. 8 Budaya Positif	Jepang terkenal dengan kegigihannya dalam membangun negara maju, negara-negara di seluruh penjuru dunia pun mengakuinya	http://www.hipwee.com/sukses/jepang-jadi-negara-maju-karena-kebiasaan-8-budaya-positif-mereka-ini-layak-kita-jadikan-teladan/	Budaya Kebudayaan Jepang	66	[edit] [delete]
16	Kebudayaan Jepang	Kebudayaan Jepang menjadi daya pikat pariwisata Jepang. Kebudayaan tradisional ini tercermin dari aneka pertunjukan seni dan budaya	http://www.jnto.or.id/aktivitas/budaya/kebudayaan-jepang	Budaya Kebudayaan Jepang	78	[edit] [delete]
17	Kategori Budaya Jepang - Wikipedia bahasa Indonesia	Halaman dalam kategori "Budaya Jepang". Kategori ini memiliki 82 halaman, dari total 82	https://id.wikipedia.org/wiki/Kategori:Budaya_Jepang	Budaya Kebudayaan Jepang	91	[edit] [delete]
19	5 Budaya Jepang yang Misterius Bahkan Untuk Orang Jepang Sendiri	Jepang memiliki berbagai macam budaya yang unik, biasanya budaya tersebut mengutamakan kecantikan dan keindahan dengan segala	https://japanesestation.com/5-budaya-jepang-yang-misterius-bahkan-untuk-orang-jepang-sendiri	Budaya Kebudayaan Jepang	93	[edit] [delete]
20	17 Peraturan Permainan Sepak Bola Terbaru yang Harus Kamu Tahu	Kamu penasaran, apa saja peraturan dalam permainan sepak bola? Inilah peraturan permainan sepak bola terbaru dan terlengkap resmi dari FIFA.	https://olahraga.pro/peraturan-permainan-sepak-bola	Peraturan Sepak Bola	99	[edit] [delete]

Figure 5. Data URL

Figure 6 is a page created to provide the data editing features. If there is an error in the input data, then this feature can be enabled to handle it.

	Pengeditan Data
Judul	Membuat Fasilitas Buku Tamu dengan PHP, MySQLi dan Bootstrap
Deskripsi	Cara membuat buku tamu di website dengan bahasa pemrograman PHP dan database MySQL dan Bootstrap
URL	http://tutorialweb.net/membuat-fasilitas-buku-tamu-dengan-php-mysqli-dan-bootstrap/
Kata Kunci	buku tamu/guestbook dengan php, mysql dan bootstrap
Nilai	91
	update

Figure 6. Edit Data URL

Analysis and Evaluation System

Personalized Web Search Results Without K-Means

30 datasets from a URL and a description (title, description, and keywords) that comes from the search results using Google, do a keyword search in turn. Figure 7 shows the results using the Personalized Web Search without K-Means with keywords Pendidikan Indonesia. Time spent in the search for 0.00172070 seconds.

Figure 7. Results of Search by Keyword Pendidikan Indonesia

Figure 8 shows the results meunggunakan Personalized Web Search without K-Means with keywords Japanese culture. Time used for 0.003364801 seconds.

Figure 8. Results of Search by Keyword Japanese Culture

Figure 9 shows the results meunggunakan Personalized Web Search without K-Means with keywords Rules Football. Time used by 0.00363302230 seconds.

Figure 9. Results Search by Keyword Rules Football

Data obtained from the search results every keyword that is input can be seen in Table 2.

Table 2. Results with Personalized Web Search

Kata Kunci	Waktu (detik)	Kategori	Akurasi
Pendidikan Indonesia	0,00172070	Akademik	100%
Budaya Jepang	0,00363302	Hiburan	100%
Peraturan Sepak Bola	0,00336480	Olahraga	100%

Searches were conducted with three different keywords. 100% accuracy was obtained through the results of an assessment based on the similarity of data desired by the user from the data displayed search results Personalized Web Search. One of the advantages Personalized Web Search is to provide personalized search tailored to the needs of the user so that the level of accuracy can reach 100%.

Genetic Algorithms

Genetic Algorithm optimization on keywords with different combinations of letters that exist in order to establish an optimal solution. Figure 10 shows that the optimal solution search results using keywords Pendidikan Indonesia made within 23.64 seconds.
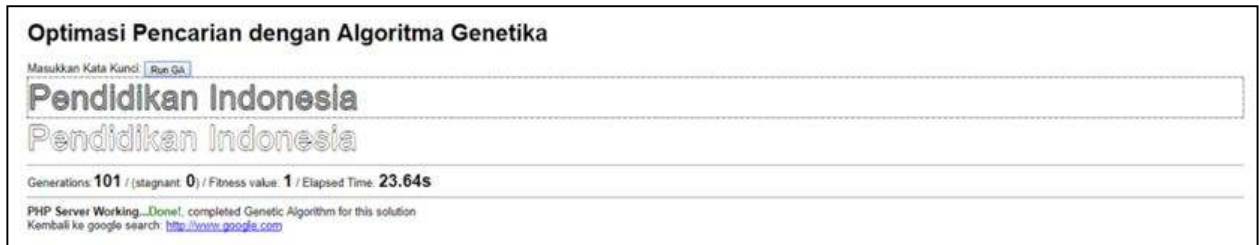
Figure 10. Optimization Results with Keyword Pendidikan Indonesia

Figure 11 shows the results of the optimization is done by using the keyword Japanese culture. The time needed for the search process is 2.08 seconds.
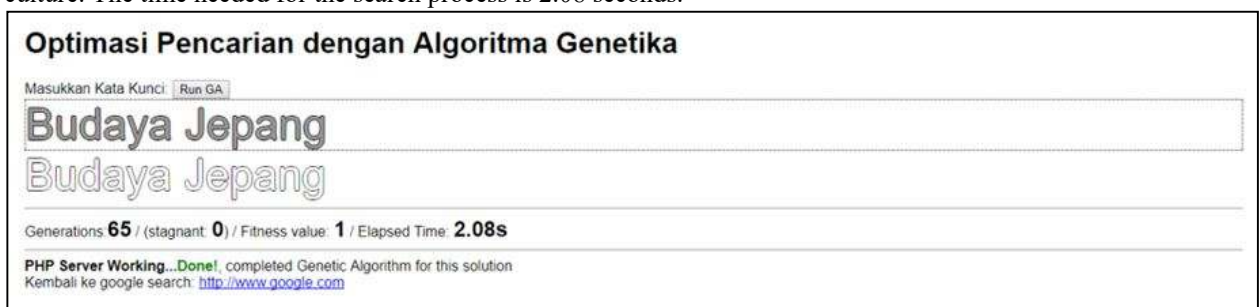

Figure 11. Results with Keyword Optimization Japanese Culture

Figure 12 shows the results of the optimization is done by using keywords Rules Football. The time needed for the search process is 5.36 seconds.
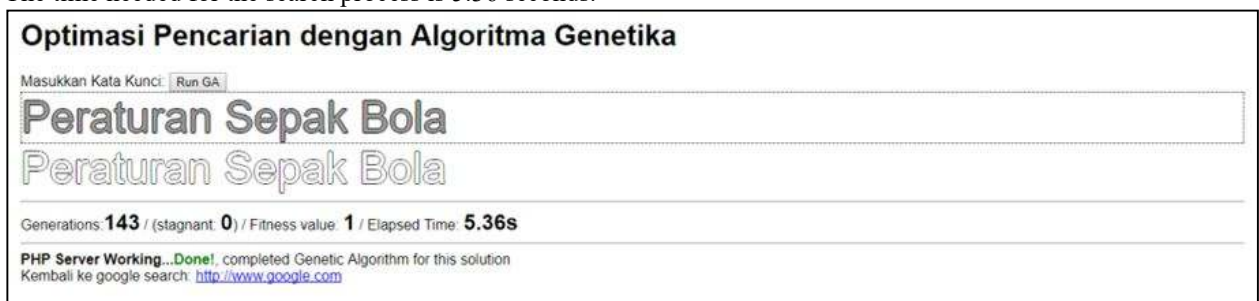

Figure 12. Results with Keyword Rules Football

Overall search result data optimal solution using Genetic Algorithms using three different keywords can be seen in Table 3.

Table 3. Results of Genetic Algorithms

Kata Kunci	Waktu (detik)	Kategori	Akurasi
Pendidikan Indonesia	23,64	Akademik	100%
Budaya Jepang	5,36	Hiburan	100%
Peraturan Sepak Bola	2,08	Olahraga	100%

Table 3 shows that for keywords Pendidikan Indonesia, time optimization process takes 23.64 seconds with an accuracy of 100%. For keywords Japanese culture process ends in 5.36 seconds with an accuracy of 100%. For keywords Rules Football, the time required for optimization is 2.08 seconds with an accuracy of 100%.

K-Means

Of the 30 datasets such as URLs and other information (title, description, and keywords) that comes from the search results using Google, do a keyword search done alternately.

Figure 13 shows the results of searches using Personalized Web Search with K-Means and keywords of education in Indonesia. The time needed for the search process by 0.00170207023 seconds.
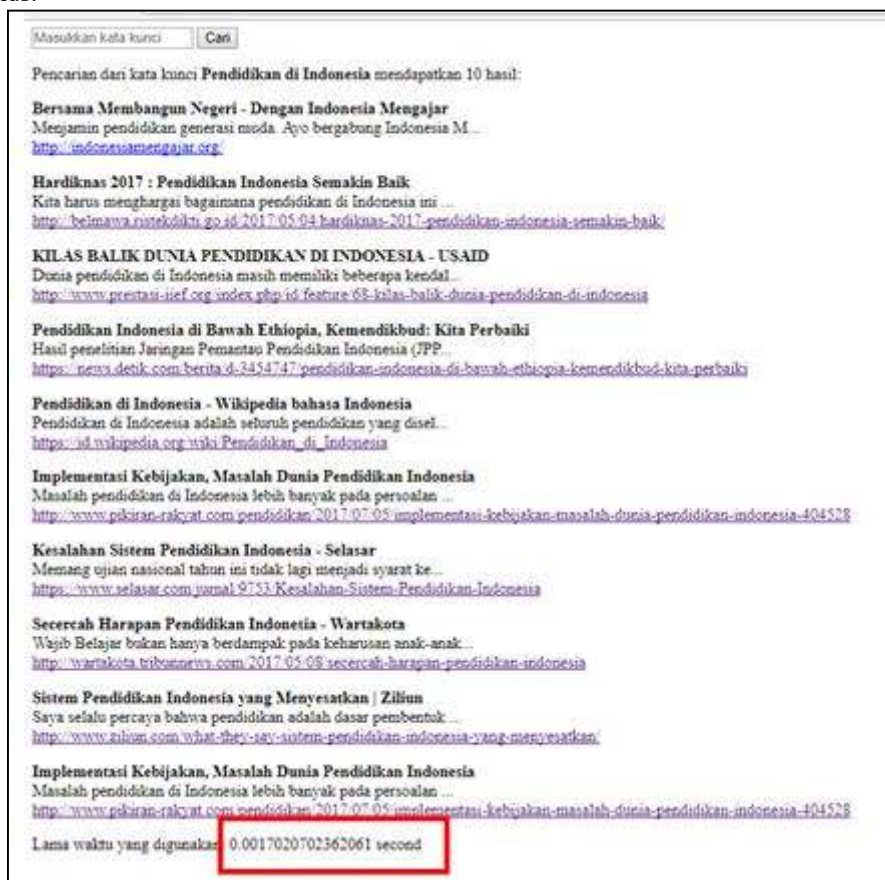

Figure 13. Results with Keyword Education in Indonesia

Figure 14 shows the results of searches using Personalized Web Search denganK-Means and keywords Culture of Indonesia. The time needed for the search process for 0.0026361942 seconds.

Figure 14. Keyword Search Results with Japanese Culture

Figure 15 shows the results of searches using Personalized Web Search with K-Means and keywords Rules Football. The time needed for the search process by 0.00260305404 seconds.
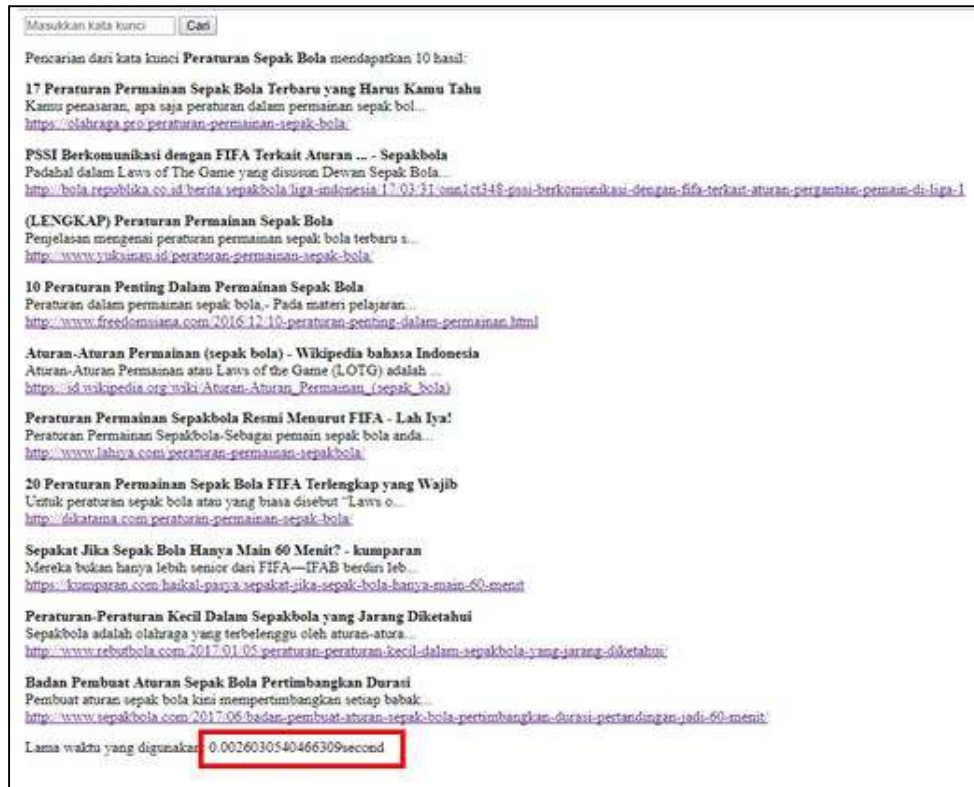
Figure 15. Results with Keyword Rules Football

Overall the data obtained from the results of each keyword that is input can be seen in Table 4.

Table 4. Results of Personalized Web Search with K-Means

Kata Kunci	Waktu (detik)	Kategori	Akurasi
Pendidikan Indonesia	0,00127887	Akademik	100%
Budaya Jepang	0,00263619	Hiburan	100%
Peraturan Sepak Bola	0,00260305	Olahraga	100%

100% accuracy was obtained through the results of an assessment based on the similarity of data desired by the user with the data displayed on the search results Personalized Web Search (PWS). One of the advantages of Personalized Web Search is to provide personalized search tailored to the needs of the user so that the level of accuracy can reach 100%.

Personalized Web Search with Genetic Algorithm and K-Means

After the experiment, both with Genetic Algorithm and K-Means, the obtained results are shown in Table 4.5. When Personalized Web Search with K-Means and Genetic Algorithm compared with Personalized Web Search without K-Means will be obtained by the time difference 0.0004183 where Personalized Web Search with K-Means faster. Comparison of the two data search results can be concluded that the Personalized Web Search with K-Means will help reduce the time of the search process by Genetic Algorithm when compared with

Personalized Web Search without K-Means. Time data obtained by adding gentika algorithm processing time with processing time Personalized Web Search.

Table 5. Personalized Web Search with K-Means and Genetic Algorithm

Kata Kunci	Waktu (detik)	Kategori	Akurasi
Pendidikan Indonesia	23,64127887	Akademik	100%
Budaya Jepang	5,36127887	Hiburan	100%
Peraturan Sepak Bola	2,08127887	Olahraga	100%

Table 6. Personalized Web Search with Genetic Algorithms

Kata Kunci	Waktu (detik)	Kategori	Akurasi
Pendidikan Indonesia	23,6417207	Akademik	100%
Budaya Jepang	5,3617207	Hiburan	100%
Peraturan Sepak Bola	2,0817207	Olahraga	100%

Comparison search results Personalized Web Search without ClusteringK-Means method with Personalized Web Search with additional methods ClusteringK-Means can be seen in Figure 16.
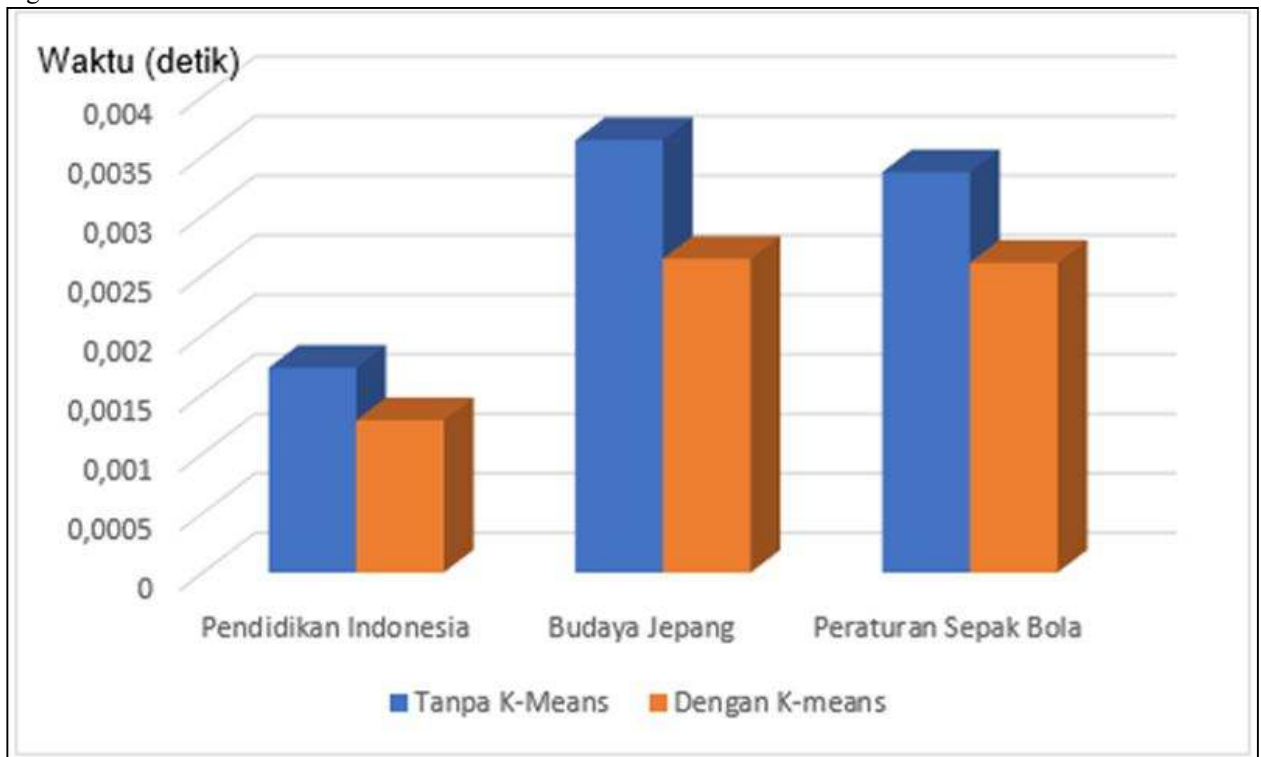

Figure 16. Comparison of Time Search Personalized Web Search with K-Means and without K-Means

Figure 16 shows a comparison between the Personalized Web Search searches with K-Means and Personalized Web Search without K-Means. Personalized Web Search with K-

Means faster process than Web Search without K-Means. It is seen that the time difference is not very significant process that is an average of 0.00044183 seconds. This is influenced by the number of datasets used by 30 datasets so that the search process does not require a long time.

4 Conclusion

From the results of research and discussion that has been described can be deduced that:

1. Implementation of the K-Means algorithm can reduce the computation time of genetics Personalized Web Search effective. The test results showed on Personalized Web Search without K-Means takes an average of 10.3617207 second, whereas Personalized Web Search with the help of K-Means takes an average of 10.36127887 detik. Difference in average time 0.00044183 seconds.
2. Personalized Web Search with Genetic Algorithms can be implemented with Clustering K-Means method to reduce computation time.

References

[1] Abdullah, Ma'ruf. *Metode Penelitian Kuantitatif.* Yogyakarta : Atmaja Pressindo, Indonesia (2015)

[2] Akhlaghian, F., Arzanian, B., dan Moradi, P. *2010 International Conference on Data Storage and Data Engineering*, https://doi.org/10.1109/DSDE.2010.30 (2010)

[3] Asroni dan Adrian, R. *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang*, Jurnal Ilmiah Semesta Teknika Vol. 18, No. 1, 76-82. (2015)

[4] Boughanem, M., Chrisment, C., dan Tamine, L. *On Using Genetic Algorithms for Multimodal Relevance Optimization in Information Retrieval*, *53*(2), 934–942. https://doi.org/10.1002/asi.10119 (2002)

[5] Chawla, S. .*A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective Personalized Web Search. Applied Soft Computing Journal*, *46*, 90–103. https://doi.org/10.1016/j.asoc.2016.04.042 (2016)

[6] Chih, C., Chen, P., Chiu, F., dan Chen, Y. *Expert Systems with Applications Application of neural networks and Kano's method to content recommendation in web personalization.Expert Systems With Applications*, *36*(3), 5310–5316, https://doi.org/10.1016/j.eswa.2008.06.067 (2009)

[7] Ferdian, Edward, dkk. *Penerapan Metode Latent Semantic Indexing Pada Search Engine.* Bandung : Laboratorium Ilmu dan Rekayasa Komputasi Departemen Teknik Informatika, Institut Teknologi Bandung, Indonesia (2005)

[8] Firman, A., Wowor, H. F., dan Najoan, X. *Sistem Informasi Perpustakaan Online Berbasis Web*, E-journal Teknik Elektro dan Komputer vol.5 no.2 Januari-Maret 2016, ISSN 2301-8402 (2016)

[9] Jay, P. *Review On Web Search Personalization Through Semantic Data* (2015.)

[10] Jogiyanto, Hartono. *Analisis & Desain Sistem Informasi Pendekatan Terstruktur Teori dan Praktek Aplikasi Bisnis*. Yogyakarta : ANDI, Indonesia (2005)

[11] Kusuma, W. F. *Pengembangan Halaman Web Menggunakan Xml Dalam Perkembangan Web 2.0*, Jurnal Teknik Informatika Universitas Gunadarma. (2015)

[12]   Mauluddin, Amras. *Aplikasi Web Browser Menggunakan Metode Url   (Universal Resource Locator) Pada Sistem Operasi Window.* Jurnal Informasi Vol.4 No.2 (5), http://files.stmik-im.ac.id/VOL4-2/5_amras.pdf (2011)