International Journal of Education, Information Technology and Others (IJEIT), August 2023, 6 (3), 263-276
DOI: https://doi.org/10.5281/zenodo.8294360
p-ISSN: 2654-2528 e-ISSN: 2623-2324
Accredited by Directorate General of Strengthening for Research and Development
Available online at https://jurnal.peneliti.net/index.php/IJEIT

Semantic-Based Search Design For Staffing Data Using Cosine Similarity Method

Rifqi Firdaus Alfalakhi

Information Systems Study Program, Sunan Ampel State Islamic University Surabaya

Received: 9 July 2023
Revised: 16 JJuly 2023
Accepted: 10 August 2023

Abstract

PT. Kenari Ketchup is one of the suppliers and distributors of the Kenari Ketchup brand. Employees at PT. Kecap Kenari has different tasks, educational backgrounds, and expertise. Researchers made observations at PT. Kecap Kenari, it was found that the search for employee data was still manual based on keywords only. To overcome this, a feature is needed that allows data search besides being based on keywords, also based on the semantic similarities/meanings of keywords. So that it can support data search with a more complete search scope. Some of the main stages in searching for data in this study are the text preprocessing stage, the search for keyword semantic meanings, and the string matching stage. At the preprocessing stage the text utilizes the PHP Literary library to normalize the text, the semantic meaning search stage uses the PHP Satria/Synonim Antonym library, while the employee data search uses the Cosine Similarity method as a string matching technique to determine the level of similarity between employee data and keywords and the semantic meaning of the keywords used . The results of this study, that the system succeeded in normalizing the text, finding the semantic meaning of the keywords, and finding the level of similarity of the data sought with cosine similarity, as well as presenting data sorted from the highest level of similarity, so as to make it easier for users to read system output related to searching employee data.

Keywords: Semantic Data Search, Cosine Similarity

(*) Corresponding Author: firdausrifky670@gmail.com

How to Cite: Alfalakhi, R. (2023). Semantic-Based Search Design For Staffing Data Using Cosine Similarity Method. *International Journal of Education, Information Technology, and Others*, *6*(3), 263-276. https://doi.org/10.5281/zenodo.8294360

INTRODUCTION

Nowadays, information technology has an important and massive role in human work. Information technology is generally divided into two, namely software (information processors) and hardware (input and output devices). Hardware can be interpreted as all physical computer devices, both as input and as information output, while software is a computer program as a data processor / processor from the input results which will be displayed on the output device (Abdillah et al., 2020).

In connection with current software, the more data entered in the database can be a problem in itself, namely the difficulty in searching for related data (Firman Nurdiyansyah, Samsul Arifin, 2018). More and more data can make it difficult to search if there are no groupings or categories of data in it, so it requires other methods in searching for related data. One method in software that can help in finding related information is Semantic Web (Arifin & Prasetyo, 2021).

Semantic Web is a technology where computer machines can understand the input entered (Pratama & Ropianto, 2020). One method for data search is cosine similarity. The Cosine Similarity algorithm is an algorithm used to calculate similarity based on vector space similarity measure or the degree of similarity between two objects (Prasetiya & Et.all, 2019).

PT Kecap Kenari is one of the distribution supply companies for the Kecap Kenari brand. Employees at PT Kecap Kenari have different tasks, educational backgrounds, and expertise. Researchers conducted observations at PT Kecap Kenari, found that employee data searches are still manual with Excel searches on one employee data worksheet based on keywords only, so that the scope of the search is limited according to the keywords used without seeing the semantic meaning of the keywords. To overcome this, a feature or application is needed that allows searching data apart from keywords also based on the equation / semantic meaning of the keywords entered by the user. Determination of the semantic meaning of keywords is generated from interaction with other systems that have a database of word equations according to the KBBI (Interoperability), so that this can support data searches with a more complete search scope.

Research has been conducted by Rais Hafiyyan in 2019, with the title of applying the SWFilter System method to the data search function. The test results show that the SWFilter method can be one solution to the problem of semantic search on the web and has good accuracy in search results (Hafiyyan et al., 2019).

From the description above, the problem that arises is that there is no semantic data search in companies that allows users to find related information with a broader scope than just the keywords entered. It is hoped that this semantic web technology can be a solution and bring convenience in information search. The object of this research is a system specifically designed for searching company employee data.

The purpose of this research is to find out information with a wider scope, namely from keywords and from keyword word equations according to their semantic meaning. and to find out the value of semantic-based string matching with the cosine similarity method.

RESEARCH METHOD

The research was conducted from August 2021 to October 2022, aiming to produce an accurate employee data search with a semantic-based employee search application with the cosine similarity method. The data obtained amounted to 99 permanent employee data from a total of approximately 400 permanent and outsourced employees. This is due to company policies related to providing outsourced employee data.

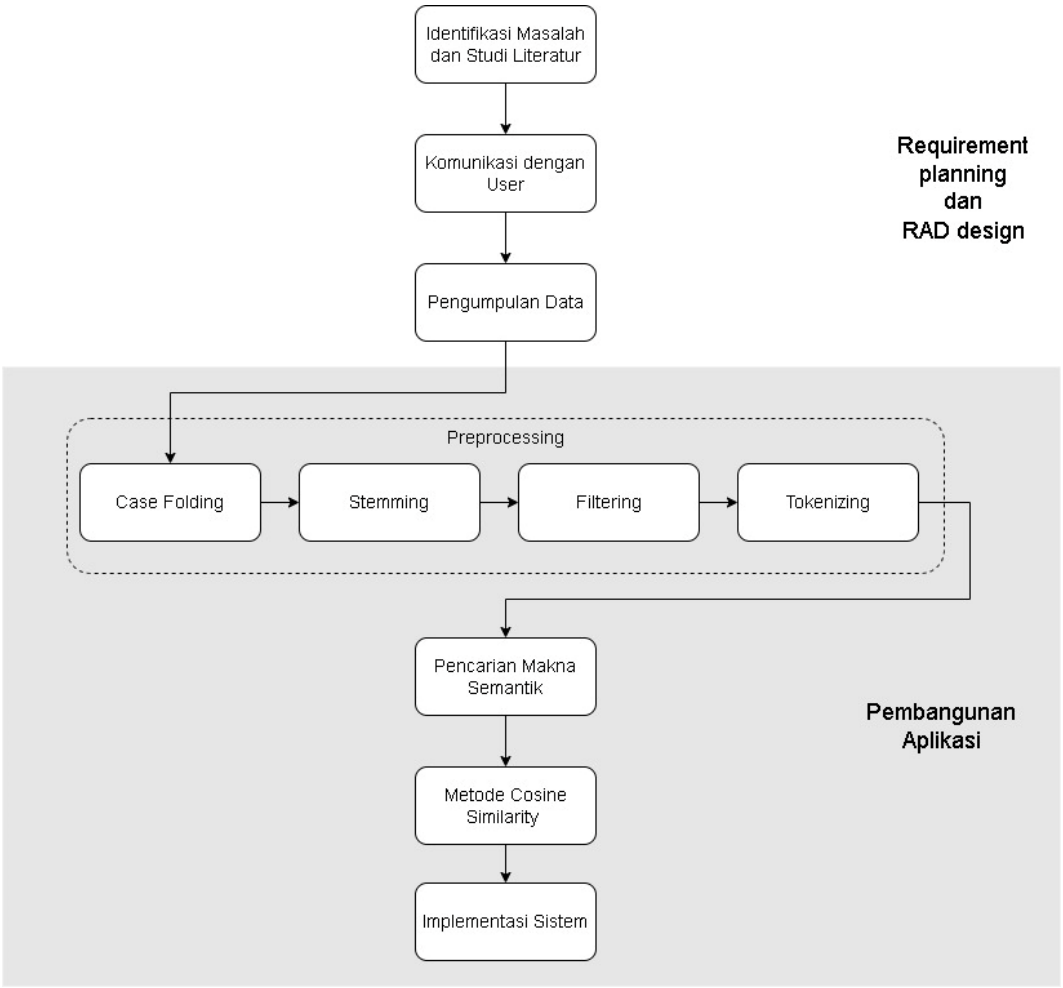

Figure 1. Research Methodology

RESULTS AND DISCUSSION

Research Results

Application Implementation

This section describes the parts of the system interface both from the frontend and backend, along with an explanation of each part.

User Interface Implementation

At this stage is the implementation stage of the application for employee search from the user side. With the user interface as follows:

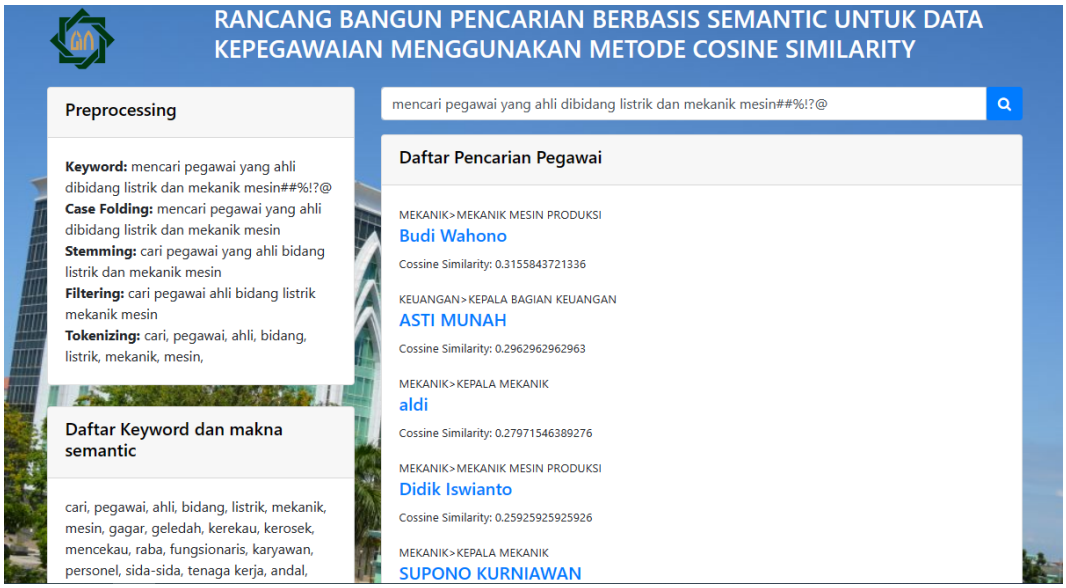

Figure 2. User Display (search page)

The user interface is composed of several parts in order, namely:

1. The top header contains the title of the program / application.
2. Header Body contains input text fields for keyword input.
3. Top-left sidebar contains the data preprocessing stage which holds the results of keyword input preprocessing.
4. The bottom-left sidebar contains semantic meaning data which is the result of searching the semantic meaning of the keyword preprocessing results.
5. Body contains a list of employees related to the list of preprocessed keywords and the resulting semantic meaning.

Implementation of Admin Interface / Backend Display

At this stage is the implementation stage of the backend page as an admin page to manage employee data to be entered in the system. With the following display:

a. Admin Login

Before entering employee data management, it is necessary to log in as an admin first. The admin login display is as follows.

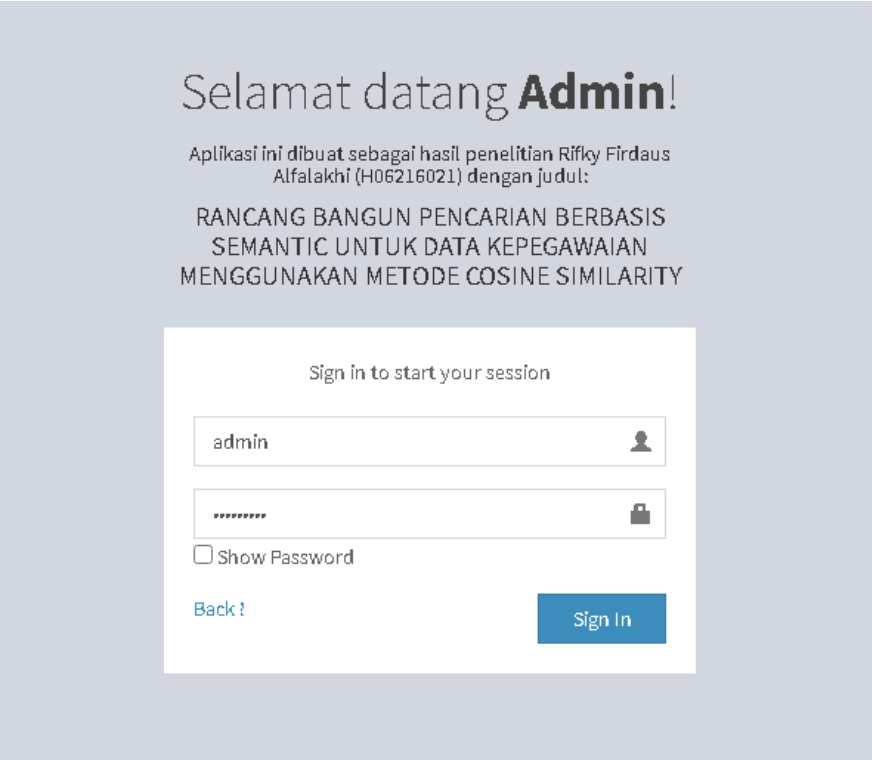

Figure 3. Admin login page

Admin enters username and password to manage application data which will then be stored in the system database.

b. CRUD (Create, Read, Update, Delete) Division Data

This page functions to manage division master data with the following display.

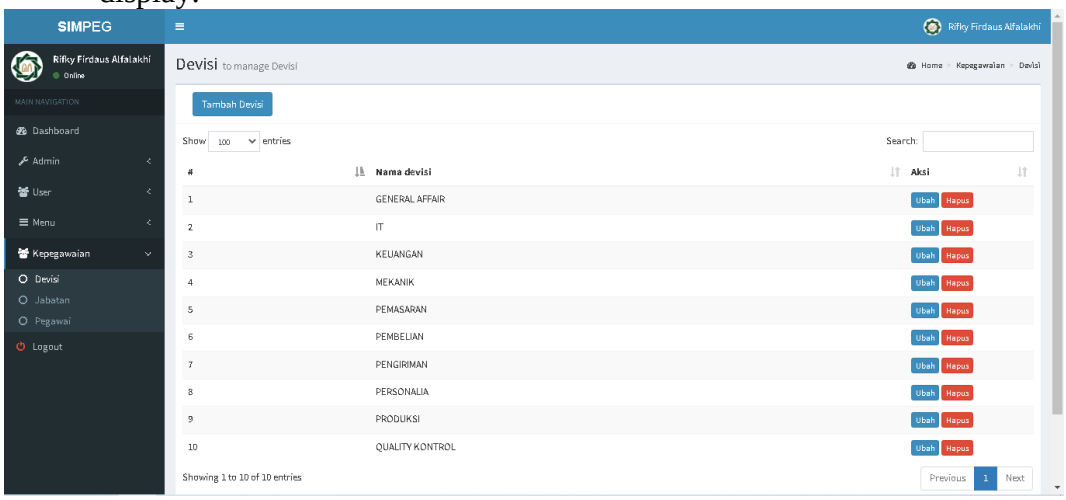

Figure 4. Division submenu page

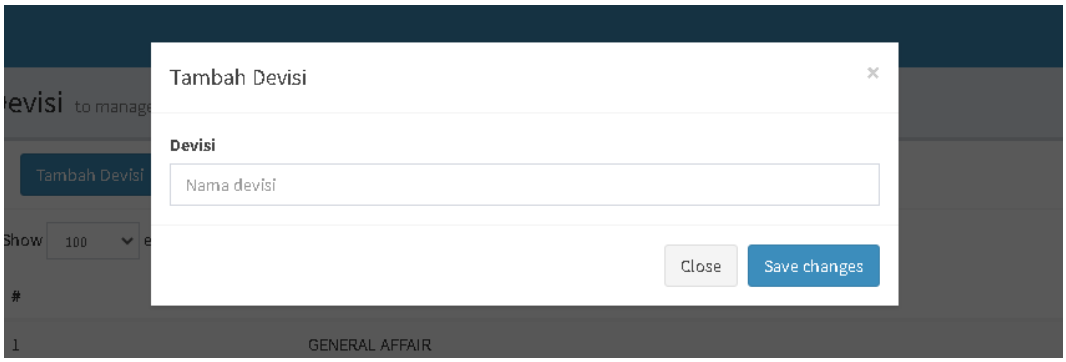

Figure 5. Add division data

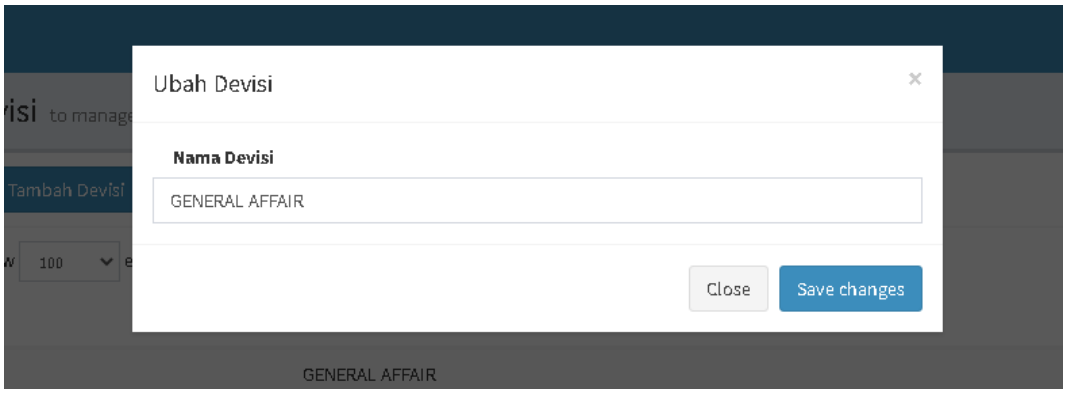

Figure 6. Change division data

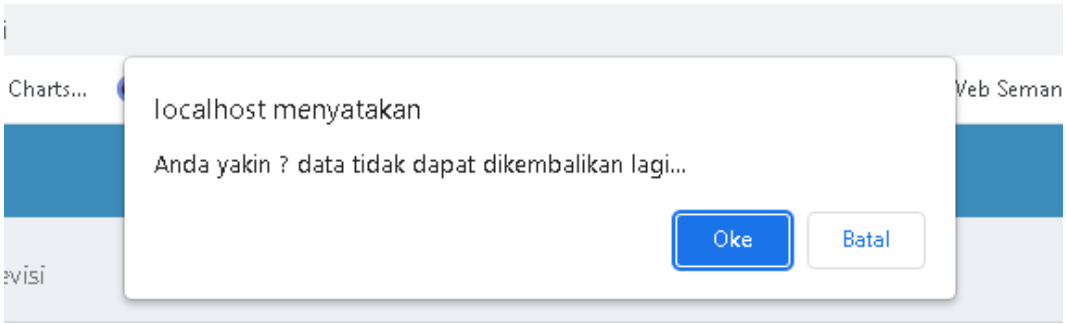

Figure 7. Delete division data

On the backend page of the division submenu staffing menu, Admin can manage division master data both adding, changing, and deleting data.

c. CRUD (Create, Read, Update, Delete) Position Data

This page functions to manage position master data with the following display.

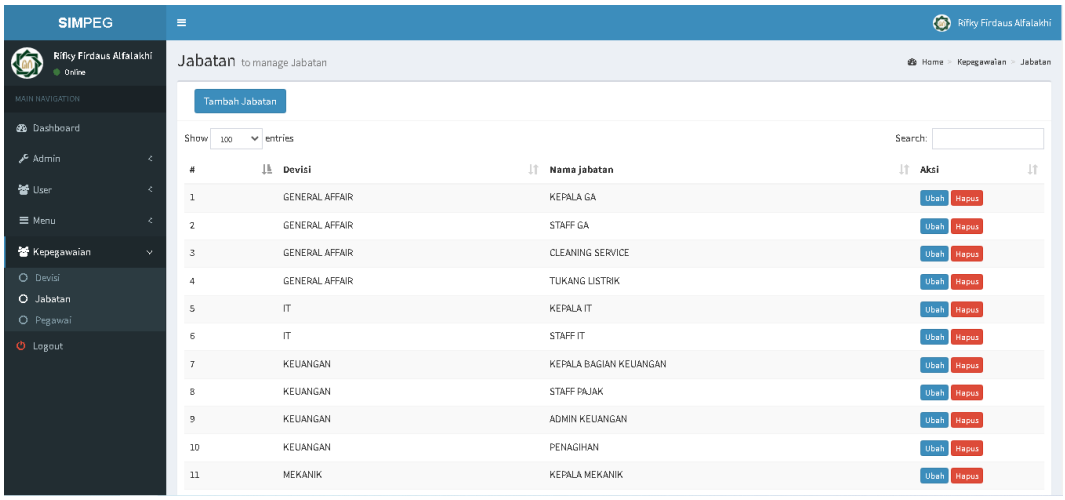

Figure 8. Position submenu page

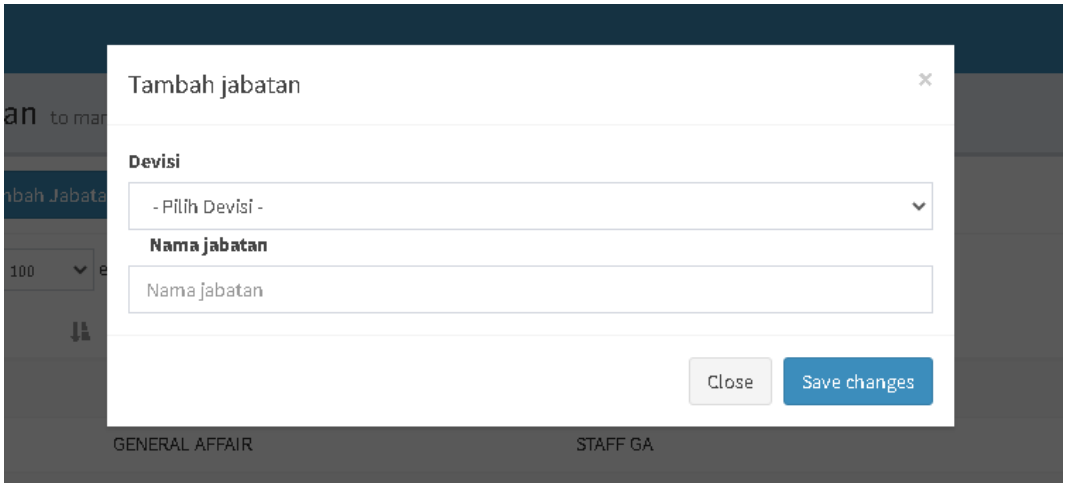

Figure 9. Add job title data

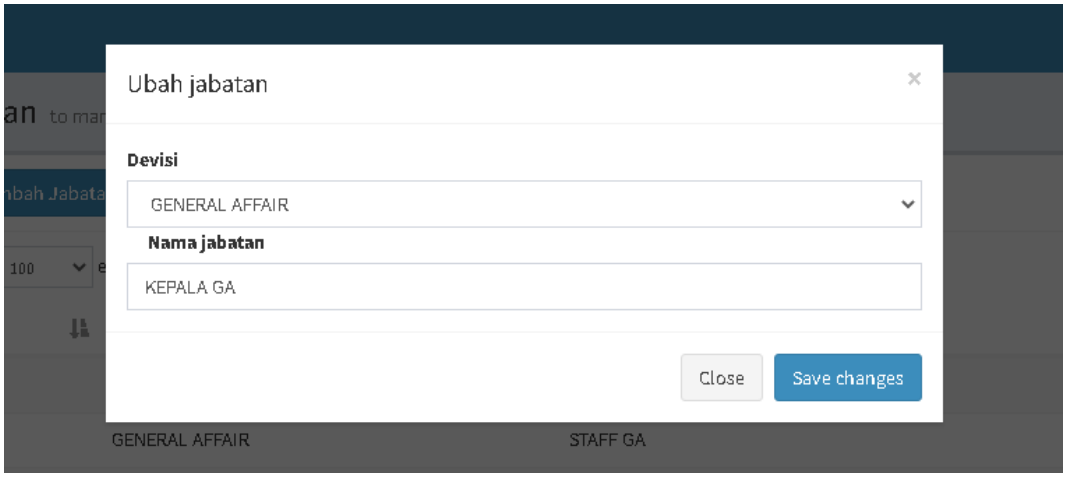

Figure 10. Change job title data

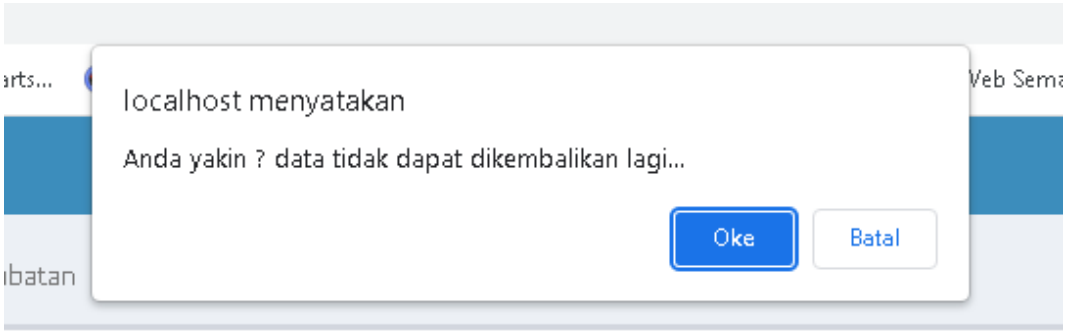

Figure 11. Delete job title data

On the backend page of the staffing menu of the position submenu, Admin can manage position master data both adding, changing, and deleting data.

d. CRUD (Create, Read, Update, Delete) Data Pegawai

This page functions to manage employee data with the following display.

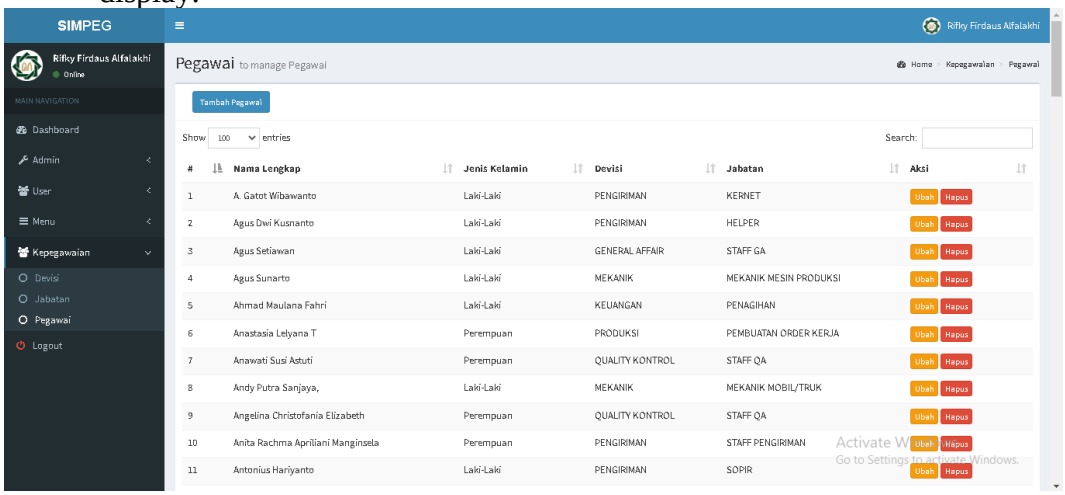

Figure 12. Employee submenu page

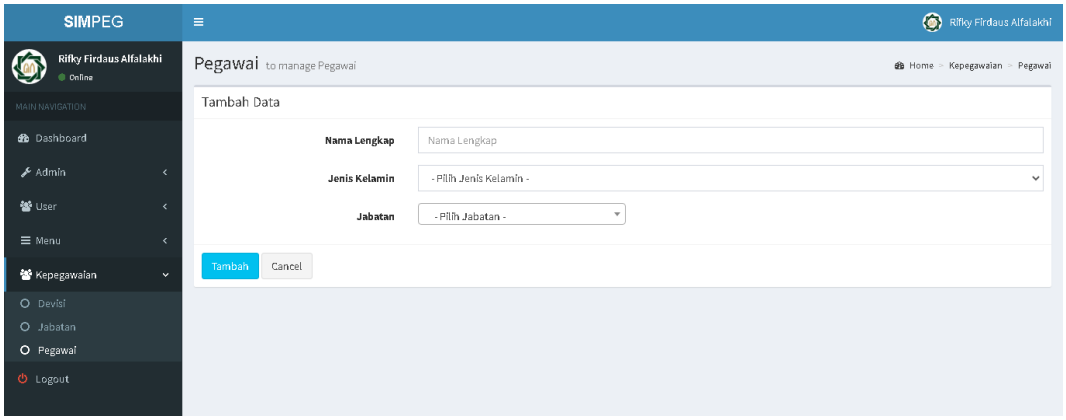

Figure 13. Add employee data

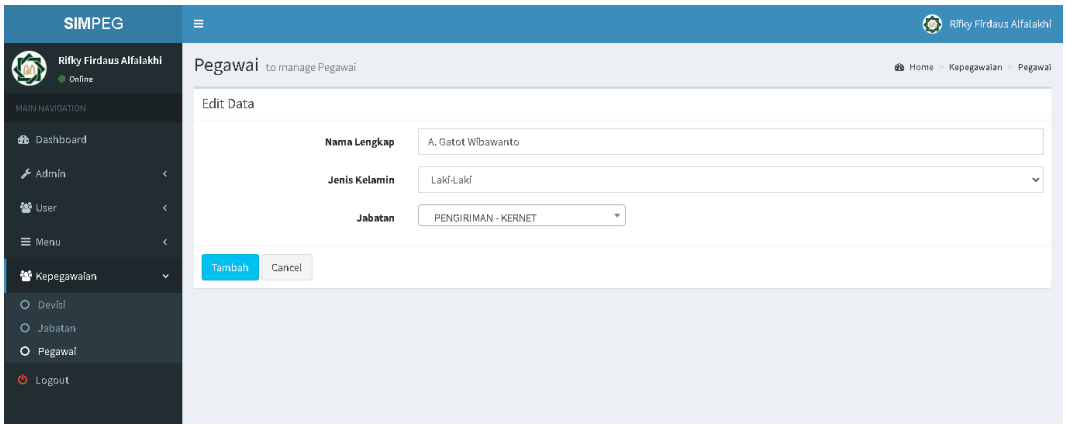

Figure 14. Change employee data

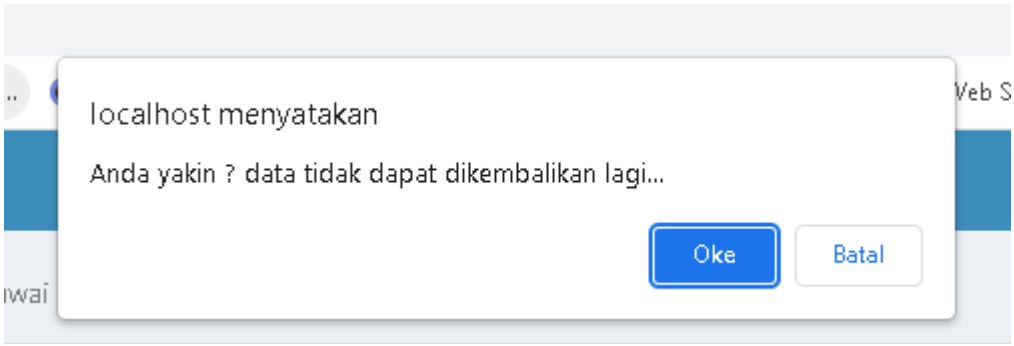

Figure 15. Delete employee data

On the backend page of the employee submenu staffing menu, Admin can manage employee data both adding, changing, and deleting data.

System Testing Scenario

After implementing the user's needs and the response given regarding the system requirements is appropriate, then enter the data testing stage. Data testing is done with blackbox testing, which will test each component in the application. The results of blackbox testing are as follows.

Table 1. System functional testing

No	Name	Use Case	Expected Result	Testing Result
1	Login process	Login as admin	Enter the employee page	OK
2	Input division data	Login as admin	Division master data increased	OK
3	Edit division data	Login as admin	Division master data changed	OK
4	Delete division data	Login as admin	Division master data deleted	OK
5	Input position data	Login as admin	Position master data is added	OK
6	Edit position data	Login as admin	Position master data changed	OK
7	Delete position data	Login as admin	Position master data deleted	OK

8	Input employee data	Login as admin	Employee data increased	OK
9	Edit employee data	Login as admin	Employee data changed	OK
10	Delete employee data	Login as admin	Employee data deleted	OK
11	Search data input	User Features	Keywords are successfully entered to be processed in data preprocessing.	OK
12	Case folding search	(without login)	Entered keywords can be removed unique characters (other than numbers and letters).	OK
13	Stemming search data	User Features	Keyword case folding results can be made in the form of basic words (eliminated affix words if any).	OK
14	Filtering search data	(without login)	Stemming results can be filtered by removing conjunctions or words that have no meaning.	OK
15	Tokenizing search data	User Features	Filtered keywords can be split or separated per word in the form of an array.	OK
16	Semantic meaning search	(without login)	Each preprocessing result keyword can be found the meaning or synonym of the word.	OK
17	Employee search	User Features	The list of employees is displayed according to the keywords that have been processed	OK
18	Calculation of cosine similarity level	(without login)	Each list of employees displays the cosine similarity value based on the calculation of the keyword array matrix in each employee who is indicated to have a match with the keyword.	OK

Based on the test results, it is known that the system can be used properly. Starting from employee data management to processing keyword data to find semantic meaning and search employee data by calculating the cosine similarity level.

Requirements Traceability Matrix

Requirements Traceability Matrix (RTM) As an addition related to testing needs with the linkage of documents or systems built. document traceability related to system testing coherently is as follows.

Table 2. Requirements Traceability Matrix

Project Name	**Semantic-based Employee Search System with Cosine Similarity**
Project Description	Employee search based on the similarity of keyword meaning with the level of similarity (cosine similarity)

No	**Main Requirement**	**Sub-Requirements**	**Category**	**Description**	**System Relationship**	**Use Case**	**Expected Result**
1	System Backend	Login process	Needed	Account verification process for backend page access	Association	Login as admin	Enter the employee page
2	System Backend	Input division data	Needed	Process to manage (insert) employee division data	Composition	Login as admin	Division master data increased
3	System Backend	Edit division data	Needed	Manage (update) employee division data	Composition	Login as admin	Division master data changed
4	System Backend	Delete division data	Needed	Manage (delete) employee division data	Composition	Login as admin	Division master data deleted
5	System Backend	Input position data	Needed	Manage (insert) employee position data	Composition	Login as admin	Position master data is added
6	System Backend	Edit position data	Needed	Manage (update) employee position data	Composition	Login as admin	Position master data changed
7	System Backend	Delete position data	Needed	Manage (delete) employee position data	Composition	Login as admin	Position master data deleted
8	System Backend	Input employee data	Needed	Manage (insert) employee data	Composition	Login as admin	Employee data increased
9	System Backend	Edit employee data	Needed	Manage (update) employee data	Composition	Login as admin	Employee data changed
10	System Backend	Delete employee data	Needed	Employee data manage (delete) process	Composition	Login as admin	Employee data deleted

11	System Frontend	Search data input	Needed	Data search keyword input process	Association	User Features (without login)	Keywords are successfully entered to be processed in data preprocessing.
12	System Frontend	Case folding search	Expected	Text preprocessing process of data search keywords (case folding stage)	Association	User Features (without login)	Entered keywords can be removed unique characters (other than numbers and letters).
13	System Frontend	Stemming search data	Expected	Data search keyword preprocessing text process (stemming stage)	Association	User Features (without login)	Keyword case folding results can be made in the form of basic words (eliminated affix words if any).
14	System Frontend	Filtering search data	Expected	Data search keyword text preprocessing process (filtering stage)	Association	User Features (without login)	Stemming results can be filtered by removing conjunctions or words that have no meaning.
15	System Frontend	Tokenizing search data	Expected	Data search keyword preprocessing process (tokenizing stage)	Association	User Features (without login)	Filtered keywords can be split or separated per word in the form of an array.
16	System Frontend	Semantic meaning search	Expected	Keyword semantic meaning search process	Association	User Features (without login)	Each preprocessing result keyword can be found the meaning or synonym of the word.
17	System Frontend	Employee search	Expected	The process of finding a list of employees from matches with keywords	Association	User Features (without login)	The list of employees is displayed according to the keywords that have been processed
18	System Frontend	Calculation of cosine similarity level	Expected	The process of determining the level of similarity	Association	User Features (without login)	Each list of employees displays the cosine

	with cosine similarity	similarity value based on the calculation of the keyword array matrix in each employee who is indicated to have a match with the keyword.

DISCUSSION

Based on the results that have been described, it shows that searching employee data based on keywords and semantic meaning with the cosine similarity method can be used to find recommendations for related employees. This is in line with research (Atikah Azhari, 2022) on finding recommendations for accompanying lecturers for students based on research fields. Which from the results of the study found that the cosine similarity method can be applied to find recommendations for accompanying lecturers for students who need accompanying lecturers.

Meanwhile, this research seeks to find employee recommendations that can be an option when urgently needing personnel who have competence in the fields needed by the company. Which can be a temporary alternative job before recruiting new employees if needed.

CONCLUSION

Based on the formulation of the problem and the results of the analysis and design of the Semantic-Based Search System for Personnel Data Using the Cosine Similarity Method, it can be concluded as follows:

1. The system can produce information in the form of semantic meaning of keyword input used to search for related employee data. Semantic meaning information is generated through several stages of the process starting from the preprocessing stage (case folding, stemming, filtering, and tokenizing) which utilizes the Sastrawi PHP library, and translates the preprocessing result keywords to the synonym list in Indonesian by utilizing the Satria / Synonym Antonym PHP library.
2. The system can find the calculation value of the level of similarity between keywords and semantic meaning with employee data attributes (employee name, division, and position) using the cosine similarity method. The similarity level value is obtained from matching strings from each keyword word with employee data attributes using cosine similarity by utilizing the PHP library Mlwmlw/php-cosine-similarity which has previously been preprocessing text to get keywords and semantic meaning of keywords.

So it can be concluded that employee search based on the semantic meaning of keywords using the cosine similarity method can be used to find employee recommendations according to related fields.

Based on the research conducted, the data attributes used in this study are only about employee names, divisions, and positions. So only about this data is done string matching using the cosine similarity method. So that it can be an option for further research in the implementation of the cosine similarity method using broader data attributes such as profile, hobbies, educational background, work history, and so on, so that the search system can produce recommended results that can see individual competencies from a broader scope.

BIBLIOGRAPHY

Abdillah, G. Y., Andryana, S., & Iskandar, A. (2020). Augmented Reality Sebagai Media Pembelajaran Perangkat Keras Komputer Dengan Fast Corner Dan Natural Feature Tracking. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *5*(2), 79. https://doi.org/10.29100/jipi.v5i2.1767

Arifin, N. Y., & Prasetyo, E. (2021). Perancangan sistem web semantik database dokumen qa. *Engineering and Technology International …*, *3*(1), 46–54. https://www.mand-ycmm.org/index.php/eatij/article/view/65

Azhari, A., Buulolo, E., & Sialalhi, N. (2022). Sistem Rekomendasi Dosen Pendamping Skripsi Berbasis Text Rank menggunakan Metode Cosine Similarity. *Pelita Informatika: Informasi dan Informatika*, *10*(3), 119-122.

Firman Nurdiyansyah, Samsul Arifin, F. M. (2018). Clustering algorithm untuk pengelompokan pelanggan dalam bidang usaha server reload. *Seminar Nasional Sistem Informasi*, 1043–1047.

Hafiyyan, R., Kusumo, D. S., & Sardi, I. L. (2019). Penerapan Metode Semantic Web Filtering System (SWFilter) pada Fungsi Pencarian Data: Studi Kasus pada Website telkomtesthouse. co. id. *EProceedings …*, *6*(1), 2227–2244. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/8666

Prasetiya, E. D., & Et.all. (2019). Pengembangan Sistem Aplikasi Pencarian Dosen Pembimbing Skripsi dengan Teknologi Web Semantik (Studi Kasus: Fakultas Ilmu Komputer Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, *2548*(9), 964X.

Pratama, A. L., & Ropianto, M. (2020). Perancangan Database Dosen Berbasis Semantic Web. *Academia.Edu*. https://www.academia.edu/download/63977890/Jurnal_Semantic_Web20200720-5979-bhzx3u.pdf