Recommendation Systems for Online Learning Materials Using Cosine Similarity and Simple Additive Weighting

Iska Yanuartanti*, Imam Much Ibnu Subroto**, Arief Marwanto*
*Electrical Engineering Department, Universitas Islam Sultan Agung, Indonesia
**Informatics Department, Universitas Islam Sultan Agung, Indonesia

Abstract

This study focuses on searching for teaching materials in order to obtain relevant teaching material information appropriately for further use as material for recommendation of course material to students using the Cosine Similarity method and calculating weighting using the Simple Additive Weighting (SAW) method. With the SAW method, 3 criteria and weight values are determined for each attribute, followed by a ranking process. So that in the end the search results that are ranked in the order of similarity and most relevant can be displayed and then selected and used as recommendations in the student's e-learning learning system. From the results of the study, the Cosine Similarity and SAW methods have provided a fairly good/effective recommendation with an average precision of 0.7867 and a recall of 0.766 so that this method is appropriate to be placed in campus e-Learning.

Keywords : *Recommendation System; e-Learning; Cosine Similarity; Simple Additive Weighting (SAW)*

1. Introduction

Information technology and systems are currently developing very rapidly, both web-based technology and desktop application-based technology. With the technology and information systems will provide convenience and help humans solve problems every day.

The e-learning system is one of the fields in the development of technology, especially in the field of education that can be touched. With this system, technology is not only used to deliver teaching materials electronically and to automate the evaluation process, but also in the selection of teaching materials that must also be considered. The selection of the right teaching materials for students will depend on several things, such as the results of evaluating student abilities or teaching materials that have never been given or read by students. In the conventional learning process, the ability of each student is considered the same so that the teaching results are considered less effective, because there will be students who really understand and understand the material but there are also students who do not understand or even understand the material given. With this information technology, it is hoped that every student will get special attention so that the teaching materials provided can suit the needs of each student. Student needs are defined as the achievement of the objectives of the learning system in which students' understanding of the subject in learning is achieved. An adaptive e-learning system will further help lecturers and students to make the learning process more comfortable[1][2][3].

The e-learning system that is currently developing requires each lecturer/teacher to design their own and upload their own teaching materials so that it requires no light effort even though on the other hand the same course material has been uploaded to the system both from the same lecturer and other lecturers who teach the subject. similar college. This is less effective in the preparation of teaching materials for lecturers. On the other hand, students are also limited to reading material from class lecturers while other relevant material and already available in the database cannot be accessed. From the problems of lecturers and students, the researcher proposes a teaching material recommendation system using the cosine similarity method and Simple Additives Weighting (SAW) as a solution.

This study focuses on the search for teaching materials in order to obtain information on relevant teaching materials appropriately for further use as material for recommendation of course material to students using the cosine similarity method and calculating the weights using the Simple Additive Weighting (SAW) method. This method determines the weight value for each attribute, which is followed by a ranking process[4]. So that in the end the search results that are ranked in the order of similarity and most relevant can be displayed and then selected and used as recommendations in the e-Learning learning system of their students. Cosine Similarity is a method used to calculate similarity (level of similarity) between two objects. The main advantage of the Cosine Similarity method is that it is not affected by the short length of a document[5],

In the Cosine Similarity Method, similarity is calculated between two document objects (eg D1 and D2) expressed in two vectors using keywords from a document that is used as a measure. The advantage of this conformity measurement method is the normalization of the length of the document. This will reduce the effect of the length of the document. The Euclidean distance (length) of the two vectors is used as a normalization factor. This normalization is necessary because long documents tend to get large values compared to shorter documents[5].

So to overcome some problems related to the relevance of the results of searching for teaching material document information, it is necessary to optimize the process of the search system. The search system displays documents that have been sorted.

This research is a combination of cosine similarity and simple additive weighting (SAW) methods. The cosine similarity method will produce data based on the criteria for similarity descriptions of qualifications entered by the user, so it is necessary to add a simple additive weighting method in order to obtain recommendations for teaching materials documents in the form of ranking with a weighting model against the criteria that have been determined at the beginning which are involved in stages. Then, a simple additive weighting method was applied to process the ranking stage of the cosine similarity data retrieve process.

The aim of this research isto build a recommendation system for relevant lectures using cosine similarity and Simple Additive Weighting (SAW). Relevant documents displayed using the Cosine Similarity will be further enhanced by the Simple Additive Weighting (SAW) method, which will then be used as recommendations.

2. Method

Lecturers will get teaching materials from the recommendation system database, with cosine similarity relevant relevant content will be displayed and if it is deemed necessary to add course materials to the database, uploading lecture materials can also be done. From the lecture material that will be used as a recommendation, a decision support system analysis is then carried out to formulate a more precise decision making from several criteria. One of the methods in the decision support system is the SAW method which is quite familiar which supports decision making by weighting all the criteria and alternatives and getting the right reference value[6][7][8].
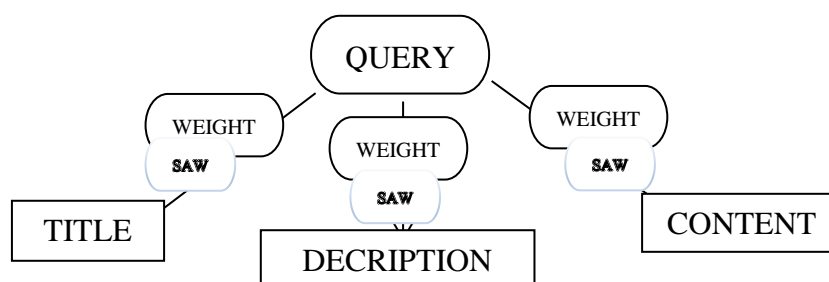
Figure 1 Component of Recommender System

In Figure 1 an overview of the system is presented, where in this system, users (Lecturers) can compose/input course material to be taught, then through this search system, teaching lecturers can search for files that will be recommended as teaching materials in online learning. After all the processes have been successful, the next step is calculating the similarity of the searched word with the documents in the database, in this case the Cosine Simmilarity method is applied to calculate the similarity. After calculating the similarity, then the system will

make an assessment based on the suitability of each SAW criteria. Furthermore, the system can display search results and sort by ranking, where the ranking is sorted by the highest similarity value.
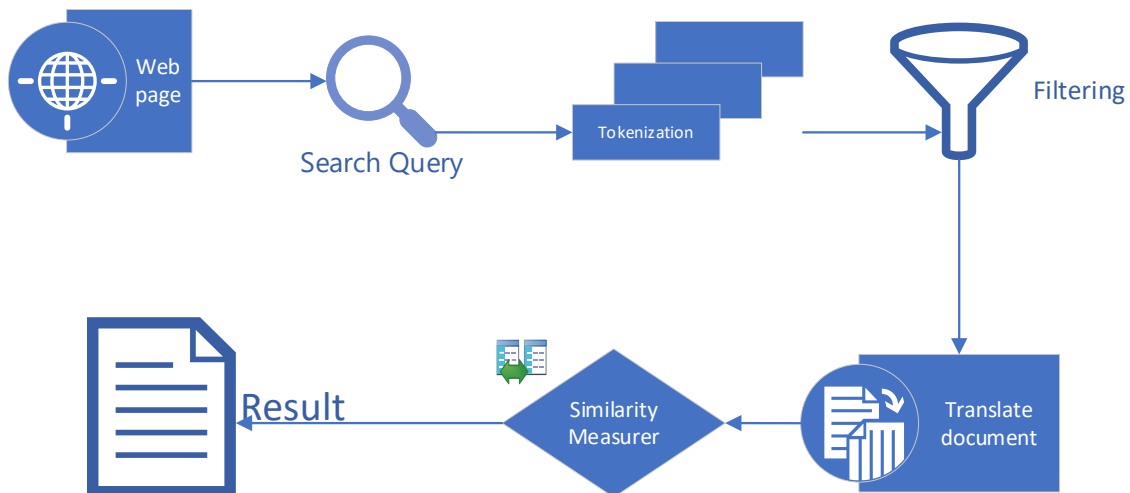

Figure 2 System business processes to be used as recommendations

Figure 2 describes the business processes of the system for further recommendations. First the user (Lecturer) visits the specified url. First, the user (Lecturer) visits the specified url (http://[IP Address]:43000/mk/) where the IP Address can change because it is placed on the internet server with a period of time limited use. Then the user enters the word to be searched and then the system will perform several processes. The processes that occur include tokenization, filtering and stemming, where all of these processes have used literary libraries, nltk and others, then the system calculates the similarity of words from the inputted words to the documents in the database. After the calculation is complete, then the system displays the search results based on the similarity calculation results.

Cosine Similarity Model Design
Cosine Similarity is a method used to calculate similarity (level of similarity) between two objects. The main advantage of the Cosine Similarity method is that it is not affected by the short length of a document. In general, the calculation of this method is based on the vector space similarity measure.

Designing SAW Criteria
In the use of the SAW method to assess the relevance of a document, a weighted value is used to assess the relevance[8][8][10][11].

Table 1 SAW Criteria and Weighting

No	Criteria	Description	Weight
1	Judul (C1)	Match the title with the word you're looking for	0.7
2	Deskripsi (C2)	Match Short Description containing the word you are looking for	0.2
3	Konten (C3)	File content	0.1

Table 1 presents three criteria that can be weighted in the system for relevance assessment, including Title, Description, and Content.

Weighting

The technique of weighting the criteria can be done in various ways and the correct method. This stage is known as pre-processing. But it can also be done in a simple way by assigning a value to each directly based on the percentage of its weight value. At this stage we fill in the weight of the value of an alternative with the criteria that have been described earlier. Please note that the maximum value of this weighting is '1'.

Table 2. Lecture Material Database Table and File Name.pdf

1	Doc 6	Doc 6 merupakan doc 6 xxx	pdf	True(2019-11-20 21:39:28)	Re-Index	Edit	Hapus
2	Managemen Sistem Informasi	Berisi Modul Perkuliahan Sistem Informasi Manajemen	pdf	True(2020-02-14 23:48:41)	Re-Index	Edit	Hapus
3	Sistem Mikrokontroller	Berisi Penjelasan terkait materi perkenalan mikrokontroller dan Konsep Mikrokontroller,	pdf	True(2019-11-27 15:28:15)	Re-Index	Edit	Hapus
4	Image Processing	Berisi Materi Tentang Pengenalan Citra Digital	pdf	True(2019-11-27 15:28:30)	Re-Index	Edit	Hapus
5	Data Mining Dan Business Intelligence	Berisi Dasar dasar Managemen Database dan Konsep Data Mining	pdf	True(2019-11-27 16:09:51)	Re-Index	Edit	Hapus

SAW score will be displayed, obtained by comparing: SAW score of course title with document content, SAW score of discussion description with document content, SAW score of suitability of document title with query, SAW score of suitability of document description with query, SAW score of suitability of document content with query[12]. So that if you do a search / search file it will display a score based on the title, query and description.

In the initial hypothesis, it can be stated that the researcher is interested in analyzing the results of using Simple Additive Weighting (SAW) in document search with the cosine similarity method to be used as a recommendation for lecture material documents for students.

This system uses cosine similarity to generate relevant documents. Cosine similarity only displays documents based on the number of words found according to the query, which is not always relevant to the user's wishes. Therefore, additional assessment is needed to determine the relevance of the search results to the wishes of the user. In the use of the SAW method to assess the relevance of documents, a weighted value is used for assessing relevance with multiple criteria, whereas in previous studies only 1 (one) criterion was used.

Cosine Similarity Analysis with SAW

Cosine Similarity is a method of calculating the similarity of the entered query with the document content[13][14][15][16]. After all processes have been successful, the next step is to calculate the similarity of the searched word to the document in the database, in this case the Cosine Simmilarity method is applied to calculate the similarity. After calculating the similarity,

then the system will make an assessment based on the suitability of each of the Simple Additive Weighting (SAW) criteria. Furthermore, the system can display search results and sort by ranking, where the ranking is sorted by the highest similarity value. It can be explained, the Lecturer will add the Microcontroller Practice course material and then search for the course material by pressing search and add then the Advanced Microcontroller query can be inputted. After pressing search it will display the search results sorted by the highest SAW rank. Also displayed the suitability score of the criteria. The criteria scores displayed include: Conformity of Discussion Title with document content, Conformity of discussion description with document content, Conformity of Document Title with query, Conformity of Document Description with query. The advantage of the SAW method compared to other decision support system methods lies in its ability to make a more precise assessment because it is based on the criteria value and the weight of the level of importance required.

System testing

In this stage, the system will be run and tested to find out whether the system is running in accordance with the results of the analysis and the expected goals. Information search engines also bring the issue of relevance because no matter how sophisticated a machine is, it will be difficult to understand the complexity of the human mind. So that to measure the effectiveness of the system in meeting requests for information and to be able to find out information that is relevant to the needs of users in a system, recall and precision formulas are used to measure the effectiveness of the information retrieval system. The effectiveness of the information retrieval system is assessed based on the theory from Lancaster in Pendit (2008) which is relevant and irrelevant. The theory also explains that the effectiveness of the information retrieval system is categorized into two, namely effective if the value is above 50% and ineffective if the value is below 50%. The two measures above are assessed in the form of a percentage, 1-100%.

The effectiveness of the system can also be measured only based on the high or low accuracy of the document with the query. In addition, it was also explained that recall is actually difficult to measure because the number of all relevant documents in the database is very large. Therefore, precision is usually one of the measures used to assess the effectiveness of an information retrieval system.

To determine the capability of the IR system that has been built, a test will be carried out by measuring the retrieval quality, namely by performing precision and recall tests.

Evaluation using Precision and Recall

Table 3. Precision and Recall Test Results

No	Query Q	Number of Documents	Relevant Data Found	Relevant	Not Relevant	Total	Not found	Total	Recall (%)	Precission (%)
				a	b	(a+b)	c	(a+c)		
1	Mikrokontroller Lanjutan	20	16	13	3	16	4	17	76,47	81,25
2	Sistem Mikrokontroller	20	15	12	3	15	5	17	70,59	80,00
3	Kualitas Daya Listrik	20	15	11	4	15	5	16	68,75	73,33
4	Sistem Informasi Management Basis Data	20	16	13	3	16	4	17	76,47	81,25
5	Demand Side Management	20	15	13	2	15	5	18	72,22	86,67
6	Pengolahan Citra Digital	20	18	13	5	18	2	15	86,67	72,22
7	Desain Sistem SCADA	20	18	15	3	18	2	17	88,24	83,33
8	Listrik	20	16	12	4	16	4	16	75,00	75,00
9	Informasi	20	16	12	4	16	4	16	75,00	75,00
							AVERAGE		**76,60**	**78,67**

Recall is a comparison between relevant search results with all relevant data in the database collection. While precision is the comparison between the relevant search results to all searches that were successfully retrieved. Recall and precision are used to measure the success rate of the search. The higher the precision and recall, the better the search strategy[17][18]. In addition, a system is declared effective if the search results are able to show high precision even though the recall acquisition is low.

3. Conclusions

From the experimental results in this study, it can be concluded that the teaching material recommendation system using the Cosine Similarity and SAW methods in this study has provided recommendations or showed good / effective results with an average Precision value of 0.7867 (78.67%) and the recall value is 0.766 (76.60%} from the scale (0 − 1) or (0%-100%), so this method is feasible to be applied in e-Learning.

REFERENCES

[1]   L. McGinty and B. Smyth, "Adaptive selection: An analysis of critiquing and preference-based feedback in conversational recommender systems," *Int. J. Electron. Commer.*, 2006, doi: 10.2753/JEC1086-4415110202.

[2]   M. K. Kusrini, "Konsep Dan Aplikasi Sistem Pendukung Keputusan," *Penerbit Andi*, 2007.

[3]   D. Nurdiyah, "Decision Support System For Approval New Student And Majoring Selection Based On Student's Interest And Talent By Fuzzy Multiple Decision Making, Simple Additive Weighting And Buble Sort Method In SMK Telekomunikasi Tunas Harapan," *J. Transform.*, 2016, doi: 10.26623/transformatika.v14i1.388.

[4]   S. Kusumadewi and I. Guswaludin, "Fuzzy Multi-Criteria Decision Making," *Media Inform.*, 2005, doi: 10.20885/informatika.vol3.iss1.art3.

[5]   Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Longman Publishing, Boston (1989)

[6]   N. Nuraeni, "PENERAPAN METODE SIMPLE ADDITIVE WEIGHTING (SAW) DALAM SELEKSI CALON KARYAWAN," *Swabumi*, 2018, doi: 10.31294/swabumi.v6i1.3317.

[7]   A. S. Putra, D. R. Aryanti, and I. Hartati, "Metode SAW (Simple Additive Weighting) sebagai Sistem Pendukung Keputusan Guru Berprestasi (Studi Kasus : SMK Global Surya)," *Pros. Semin. Nas. Darmajaya*, 2018.

[8]   Zarnelly and N. Yusuf, "Sistem Pendukung Keputusan Penentuan Peserta Pelatihan Berprestasi Menggunakan Metode Simple Additive Weighting," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, 2018.

[9]   A. Setiadi, Y. Yunita, and A. R. Ningsih, "Penerapan Metode Simple Additive Weighting(SAW) Untuk Pemilihan Siswa Terbaik," *J. Sisfokom (Sistem Inf. dan Komputer)*, 2018, doi: 10.32736/sisfokom.v7i2.572.

[10]   H. Harsiti and H. Aprianti, "Sistem Pendukung Keputusan Pemilihan Smartphone dengan Menerapkan Metode Simple Additive Weighting (SAW)," *JSiI (Jurnal Sist. Informasi)*, 2017, doi: 10.30656/jsii.v4i0.372.

[11]   M. I. Dzulhaq, Sutarman, and S. Wulandari, "Sistem Pendukung Keputusan Penerimaan Siswa Baru dengan Metode Simple Additive Weighting di SMK Kusuma Bangsa," *J. Sisfotek Glob.*, 2017.

[12]   "Analysis and Implementation Fuzzy Multi-Attribute Decision Making SAW Method for Selection of High Achieving Students in Faculty Level," *Int. J. Comput. Sci. Issues*, 2013.

[13]   J. Hasugian, "Penelusuran Informasi Ilmiah Secara Online: Perlakuan Terhadap Seorang Pencari Informasi Sebagai Real User," *PUSTAHA*, 2006.

[14]   I. Xie, B. Edward, and H. Zhang, "How do users evaluate individual documents? An analysis of dimensions of evaluation activities," *Inf. Res.*, 2010.

[15]   F. Amin, "Sistem Temu Kembali Informasi dengan Metode Vector Space Model," *J. Sist. Inf. BISNIS*, 2012, doi: 10.21456/vol2iss2pp078-083.

[16]   R. LVN, Q. Wang, and J. D. Raj, "Recommending News Articles using Cosine Similarity Function," *Proc. SAS Glob. Forum*, 2014.

[17] F. Quinet, "Évaluation des coûts-avantages de la négociation collective et banque d'expérience en relation de travail," *Relations Ind.*, 2005, doi: 10.7202/028447ar.

[18] W. M. Shaw, R. Burgin, and P. Howell, "Performance standards and evaluations in IR test collections: Vector-space and other retrieval models," *Inf. Process. Manag.*, 1997, doi: 10.1016/S0306-4573(96)00044-1.