Jurnal Mantik
Volume 5, Number 1, May 2021, pp. 41-45 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Information Retrieval System for Online Food Ordering Using the Vector Space Model Method

Saifulloh Rifai

Program Studi Ilmu Komputer, STMIK Nusa Mandiri
Jl. Kramat Raya No.18, RT.5/RW.7, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450

Email : 14002275@nusamandiri.ac.id

ARTICLEINFO

Article history:
Received: 10/03/2021
Revised: 20/03/2021
Accepted: 21/03/2021

Keywords:
vector space model, Information retrieval

ABSTRACT

The development of start-up in Indonesia continues to grow along with increasingly sophisticated technology developments, this is one of the business opportunities from sharing start-up in Indonesia, and everyone is always given the convenience of every activity that is being experienced by millennial. Speed and accuracy in the information search process have become mandatory. In order to facilitate the search process or find the required information, the aim of this study is to provide a general description of the restaurant search process from online food ordering applications. The implementation of the theory applied in this study uses the vector space model method, with the token and indexing process so that results are found according to the keywords inputted in the application, and will be compared with the data contained in the application, so that it can produce the correct information.

1. Introduction

The implementation of the Vector Space Model (VSM) can be used in various scientific fields such as Computational Linguistics Expert Systems, Medical VSM method was chosen because the way this model works is efficient, easy to represent and can be implemented in document matching[1]. using a search engine or Information Retrieval System, the user enters a query and the search engine displays the search results.

The online public access catalog (OPAC) retrieval tool has been around since 1970. Since it was first created, the creation of an information retrieval system has undergone a process of change according to its development[2]. Information has grown very rapidly in various types of content, in order to provide ease of user activity[3].

This study aims to provide a broad overview of the token process, indexing, calculation, Term Frequency, Inverse Document Frequency until the end of the Ranking process, with sample data from one of the online food ordering service providers, Vector Space Model (VSM) research with 27 documents 3 highest ranking from the most relevant calculation results.

2. Method

2.1 Information Retrieval System

Information retrieval system is a system that finds (retrieves) information according to user needs from a collection of information automatically[1]. Retrieval of information shows the information retrieval process needed. Information retrieval (IR) is generally related to information retrieval (IR) information retrieval [4]. One application of the principle of relevance that has long been used in system development (Lestari, 2016) The Information Retrieval System finds information which is usually in the form of documents from unstructured data in text form to meet the information needs of very large data collections, generally stored in databases computer[5].

2.2 Vector Space Model

The vector space model provides a partial matching framework.This is achieved by assigning non-binary weights to index terms in queries and documents[5]. Not only for text search, information search can also query multimedia elements such as images, sounds[6],this method looks at the level of closeness or similarity (smilarity) of terms by weighting terms. The document is seen as a vector that has a magnitude (distance) and direction (direction). In the Vector Space Model, a term is represented by a dimension of

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

vector space. The relevance of a document to a query is based on the similarity between the document vector and the query vector. [7]documents and queries are represented as vectors and angles between the two. vector calculated using the cosine of similarity function. The effectiveness of VSM largely depends on the weighting term applied [8] allowing the calculation results to be ranked according to the similarity measure.

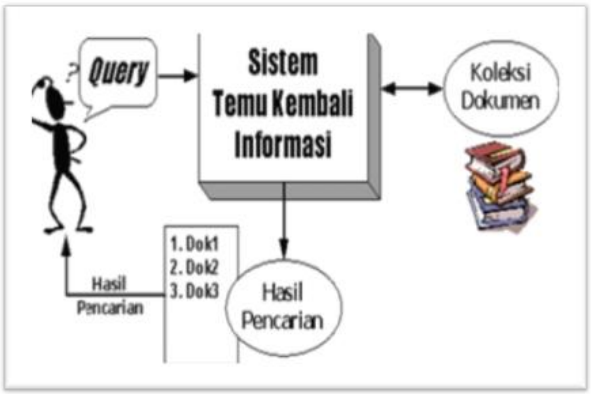

Fig 1. Illustration of Retrieval System Model Information

Steps of the vector space model method :

a. Calculating document weight with tf-idf Idf = log (D / df)
b. Calculating the distance of each document and query Sqrt (Q) = Sqrt (∑) Sqrt (D) = Sqrt (∑ = nj 1 QJ 2)
c. Calculating Dot Product Sum (Q * Di) = (∑ = nj DJ, 2 J)
d. Calculating Cosine Similarity Θ $Q*D$ | Q | ∗ | D

This research was carried out in several stages which are illustrated in Figure 1. The data processed in this system is a collection of digital documents in the form of the name of the restaurant contained in the online food ordering application. Kramat Raya District area kusus. Senen, City of Central Jakarta

3. Results and Discussion

The relevant document is the closest to the given query. In the same way, two documents will be deemed relevant if they are in neighboring territories with each other (R.K.Makhijani1, I.N.Bharambe2) e. Ranking. After calculating the cosine value, a ranking is made of these documents, so that the accuracy of the information generated is obtained.

Example: Query (Q)= Soto Betawi H.Maruf Food Colony Plaza Atrium
1 (D1) = NasiKupuBarehSolokUni Emi Ukm Food Street
2 (D2) = CotoMakasarsenenDaeng Awing KramatSoka
3 (D3) = Sop KonroSartikaSenen
4 (D4) = KedainasiKapu Bukit Tinggi
5 (D5) = Watt CoffeSenen
6 (D6) = AyamGeprekSeksi
7 (D7) = the Eatery Bowl
8 (D8) = Thai Street Plaza Atrium
9 (D9) = Sour Sally Mini Atrium
10 (D10) = Sharetea Plaza Areium
11 (D11) = Mm Juce Plaza Atrium
12 (D12) = EmpalGentong Food Colony Plaza Atrium
13 (D13) = NasigorengKebuliKambing Food Colony Plaza Atrium
14 (D14) = BuntutKomplit Food Colony Plaza Atrium
15 (D15) = Cakwe Master Food Colony Plaza Atrium
16 (D16) = Soto Betawi H Maruf Food Colony Plaza Atrium
17 (D17) = KuetiauMabes 77 Food Colony Plaza Atrium
18 (D18) = Burger King Atrium Senen
19 (D19) = Hokben Plaza Atrium
20 (D20) = Pezzo Atrium Senin
21 (D21) = Bakso Malang Cak Kris
22 (D22) = WaroengBakudapaMenado Food Colony Plaza Atrium

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

23 (D23) = GadoGado Java Plaza Atrium
24 (D24) = PisangGorengRoa Plaza Atrium
25 (D25) = RumahMakanPaman
26 (D26) = EsTeler 77 Plaza Atrium
27 (D27) = AyamGeprekSeksi

TABLE 1
CALCULATION OF TF (TERM FREQUENCY)

	Document																		
Token	Q	1	2	3	4	5	…..	27	DF	Kris	Q	1	2	3	4	5	…..	27	DF
77	0	0	0	0	0	0	…..	0	2	Kuetiau	0	0	0	0	0	0	…..	0	1
Atrium	1	0	0	0	0	0	…..	1	17	Kupu	0	1	0	0	0	0	…..	0	1
Awing	0	0	1	0	0	0	…..	0	1	Mabes	0	0	0	0	0	0	…..	0	1
Ayam	0	0	0	0	0	1	…..	0	1	Makan	0	0	0	0	0	0	…..	0	1
Bakso	0	0	0	0	0	0	…..	0	1	Makasar	0	0	1	0	0	0	…..	0	1
Bakudapa	0	0	0	0	0	0	…..	0	1	Malang	0	0	0	0	0	0	…..	0	1
Bareh	0	1	0	0	0	0	…..	0	1	Mango	0	0	0	0	0	0	…..	0	1
Betawi	1	0	0	0	0	0	…..	0	1	Maruf	1	0	0	0	0	0	…..	0	2
Bos	0	0	0	0	0	0	…..	0	1	Master	0	0	0	0	0	0	…..	0	1
Bowl	0	0	0	0	0	1	…..	0	1	Menado	0	0	0	0	0	0	…..	0	1
Bukit	0	0	0	1	0	0	…..	0	1	Mini	0	0	0	0	0	0	…..	0	1
Buntut	0	0	0	0	0	0	…..	0	1	Nasi	0	1	0	1	0	0	…..	0	3
Burger	0	0	0	0	0	0	…..	0	1	Paman	0	0	0	0	0	0	…..	0	1
Cak	0	0	0	0	0	0	…..	0	1	Pezzo	0	0	0	0	0	0	…..	0	1
Cakwe	0	0	0	0	0	0	…..	0	1	Pisang	0	0	0	0	0	0	…..	0	1
Coffee	0	0	0	0	0	1	…..	0	1	Plaza	1	0	0	0	0	0	…..	1	14
Colony	1	0	0	0	0	0	…..	0	7	Rice	0	0	0	0	0	0	…..	0	1
Coto	0	0	1	0	0	0	…..	0	1	Roa	0	0	0	0	0	0	…..	0	1

Before doing TF calculations, it is necessary to do indexsing and filtering first of all existing documents, so that the data generated from each document has meaning that has meaning. D1, D2, D3, D4, D5, D6… D27 = Document tf = many words searched in a document. D = total documents, df = Many documents containing searched words.

TABLE 2.
CALCULATION OF TERM FREQUENCY - FREQUENCY OF INVERSE DOCUMENTS

idf	tf*Idf								
log(D/df)	Q	D1	D2	D3	D4	D5	D6	…	D27
1,1303338	0	0	0	0	0	0	0	…	1,4
0,2009148	1,4	0	0	0	0	0	1,4	…	1,4
1,4313638	0	0	1,4	0	0	0	0	…	0
1,4313638	0	0	0	0	0	1,4	0	…	0
1,4313638	0	0	0	0	0	0	0	…	0
1,4313638	0	0	0	0	0	0	0	…	0
1,4313638	0	1,4	0	0	0	0	0	…	0
1,4313638	1,4	0	0	0	0	0	0	…	0
1,4313638	0	0	0	0	0	0	0	…	0
1,4313638	0	0	0	0	0	1,4	0	…	0
1,4313638	0	0	0	1,4	0	0	0	…	0

From the results of tf calculations, the sample data from the number of existing documents produced 60 tokens from 6 documents and one query, to get the distance between documents and queries, the need for IDF calculations is generated from the tokenization of the calculation results in table 2 below:

TABLE 3.
CALCULATION OF DISTANCE Q-D

	Jarak Q-D
	4,0485
	4,2941
	3,787
	4,0485
	2,0243
	3,787
SQRT Q	2,4792
	2,8627
	2,4792
	2,4792
	3,5061
	4,2941
	3,2006

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

Jarak Q-D
3,5061

The document is seen as a vector that has a magnitude (distance) and direction (direction).In the Vector Space Model, a term is represented by a dimension ofvector space. The relevance of a document to a query is based on the similarity between themvector documents and queries, the length of the document tends to have the frequency of occurrence of wordsAfter knowing the calculation of the distance between Q-D using the formula Sqrt (D)= Sqrt ($\sum = nj\ 1\ QJ2$).

TABLE 4.
CALCULATION OF DISTANCE Q-D

	Dot Product
	4,1976
	0
	0
	0
	0
	4,198
	4,19759
	8,395
Sum Q*D	8,395
	16,79
	16,79
	16,79
	16,79
	33,581
	16,79
	4,1976
	8,3952

To get the TF-IDF weight value and the distance between the document and the query (Q-D) for each word, the dot product will be calculatedto find the degree of similarity between each available document. Results ofdot product calculations can be seen in table four, the results will affect the cosine levelsimilarity.Calculation of similarity The next step is to calculate the cosine value of the anglebetween the vector query with each document with formulas$\text{Cosine } \emptyset \text{ Di} = \frac{D*Q}{|Q|*|D|}$

From the Vector Space Model analysis, the results for the three documents above are as shown in Table 3 below:

TABLE 5.
RANKING CREATION

Document	Score	Rank
D13	2,3658	1
D9	1,6728	2
D21	1,6728	3

4. Conclusion

This study uses a vector space model to perform the information retrival system concept search processis done more quickly and precisely and computes a similarity score using a weighted average ofEach item of cosine size then calculates the size of the similarity and to determine the angle between the document vectorsand query vectors because VSM is based on a geometry where each term has its own dimension in a multi-dimensional space,Of the 27 documents that were used as test materials and included 1 query, three document rankings were most similar to the querywhich is entered into the system.

5. References

[1] F. Amin, "Dinamik vol 21 Dice Similarity," vol. 21, 2020.
[2] N. P. Lestari, "Uji Recall and Precision Sistem Temu Kembali," *Univ. Airlangga*, vol. 5, no. 3, pp. 45–46, 2016.
[3] A. Fauzi and G. Ginabila, "Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model," *J. Pilar Nusa Mandiri*, vol. 15, no. 1, pp. 41–46, 2019, doi: 10.33480/pilar.v15i1.61.
[4] V. Kumar, R. Kumar, and A. Sharma, "Applying Neuro-fuzzy Approach to build the Reusability Assessment Framework across Software Component Releases - An Empirical Evaluation," *Int. J. Comput. Appl.*, vol. 70, no. 15, pp. 41–47, 2013, doi: 10.5120/12041-8047.
[5] H. Harafani, "Optimasi Parameter Pada Metode Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan," *J. Intell. Syst.*, vol. 1, no. 2, pp. 82–90, 2015.

[6] B. Yulianto, W. Budiharto, and I. H. Kartowisastro, "The Performance of Boolean Retrieval and Vector Space Model in Textual Information Retrieval," *CommIT (Communication Inf. Technol. J.*, vol. 11, no. 1, p. 33, 2017, doi: 10.21512/commit.v11i1.2108.

[7] G. Yulianta, Y. Setiawan, and D. Puspitaningrum, "Information retrieval system pada pencarian dokumen digital pada smartphone menggunakan algoritma boyer moore," vol. 4, no. 2, pp. 236–244, 2016.

[8] E. E. Ogheneovo and R. B. Japheth, "Application of Vector Space Model to Query Ranking and Information Retrieval," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 5, pp. 42–47, 2016.