PROGRESS IN APPLIED SCIENCE AND TECHNOLOGY
Faculty of Science and Technology
Rajamangala University of Technology Thanyaburi
Prog Appl Sci Tech. 2024; 14(1):11-19
ISSN (Print): 2730-3012
ISSN (Online): 2730-3020
https://ph02.tci-thaijo.org/index.php/past

Research Article

Received: November 19, 2023
Revised: March 06, 2024
Accepted: March 07, 2024

DOI: 10.60101/past.2024.251684

Web-based Cooking Recipe Recommender System based on Stocked Groceries

Suvil Chomchaiya[1], Burasakorn Yoosooka[2*], Chomanong Rodmee[1], Darunee Naowaratphunsina[1] and Wannaporn Kaichid[1]
[1] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand*
[2] *Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, Bangkok 10800, Thailand*
*E-mail: burasakorn.y@rmutp.ac.th

Abstract

Due to the severity of the recent PM 2.5 and COVID-19 pandemic situations, the work-from-home lifestyle has been widely adopted as a new normal. Consequently, it's necessarily preferable to cook at home instead of dining out as usual. However, the common problems are the unplanned and overstocked grocery items which are usually unrecognized and improperly managed. To address this issue, the "What To Cook" web application was developed with the theoretical application of Term Frequency-Inverted Document Frequency (TF-IDF) calculations to search for recipes based on the stocked groceries and then applied Cosine Similarity to calculate the similarity between each recipe and the stocked grocery items. Users can input the list of own stocked grocery items into the application and then apply the content-based filtering system to recommend recipes to utilize the stocked grocery items.

Additionally, the application supports the image capturing using Google Cloud Vision API. The application also stores the user's cooking history and saves the under interested recipes for future reference. After testing the application in real-world scenarios, it was found to be easy to use with satisfiable results.

Keywords: Web Application, Recommender System, Cooking

1. Introduction

The residential grocery stocking has becoming necessarily popular in Thailand due to the recent PM 2.5 and COVID-19 pandemic situations which result in the recent work-from-home trend as new normal. This is to facilitate the home cooking due to that the people are encouraged to stay home for better safety. Hence, the consequences are the overstocking of groceries and the repetition of cooking menus due to the lacking of proper menu suggestions. Hence, some of rarely used grocery items are unintentionally forgotten and eventually spoiled and wasted.

With the stated condition, any home cooking menu suggestions with cooking instructions as web-based application would be really beneficial. This is also the motivation to develop an application for cooking menu recommendation based of the in-stock grocery items. The main objectives are as the follows.

1) Add more cooking menu selection based on the stocked grocery items to anticipate the needs of homemakers to utilize the stored items before being spoiled and wasted.

2) To extend the consumption of all stocked grocery items with developed web-based application.

3) To enable user-generate contents via user-added cooking menus as desired.

2. Literature Reviews

Research in recommender system has received more attention and more diversified from the past decades (12, 13). The related theoretical background of this paper is described as the follows.

2.1 Related Theories

Research in recommender system has recently received more attention and more diversified from the past (14). Academic research in recommender was recently diversified in terms of techniques and methods used. However, the major theoretical methods found among academic publications are as the follows.

2.2.1 Term Frequency-Inverted Document Frequency (TF-IDF)

TF-IDF is the weighting of terms in order to determine their significance in a document which was widely applied in many recent researches for making relevant decisions (18). TF-IDF is widely used in computer-based information retrieval and text mining (1, 5, 8, 22), which consists of two factors as the follows.

- TF (Term Frequency) is the frequency of a term that is existing in a document which can be determined as the following formula:

$$TF = \frac{f(t,d)}{\sum f(t,d)} \tag{2.1}$$

Where TF = Term frequency
f = Frequency
t = Term
d = Document

Invert Document Frequency or IDF is the measurement of importance of a term in all documents. IDF can be calculated as the following equation.

$$IDF(term, allDocument) = log\frac{N}{df(t)} \tag{2.2}$$

Where
IDF = Importance of a term in all documents
N = Total number of documents
df(t) = number of document with term t

The combination of TF and IDF can be calculated to determine the weighting of a term in documents. The higher weighting means the higher frequency that term is existing in documents or vice versa. The TF-IFD is as the following equation.

$$TF - IDF = TF \times IDF \tag{2.3}$$

Where TF-IDF = Weighting of a term
TF = Term frequency
IDF = Invert document frequency

2.2.2 Cosine Similarity

The Cosine Similarity is the measurement of whether two vectors are conforming in the similar direction or not. If A and B are two vectors to be measures, the Cosine Similarity are as the following equation.

$$similarity = \cos(\Theta) = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2.4}$$

Cosine Similarity are in between 1 to -1 where:

- If approaching 1, directions of both A and B vectors are CORRELATED.
- If approaching -1, directions of both A and B vectors are ANTI-CORRELATED.
- If approaching 0, directions of both A and B vectors are NOT CORRELATED.

Cosine similarity is one of the most popular methods applied in researches involving with recommender systems. Pertaining to Singh et al. (18) (the application of Cosine similarity in combination with KNN) and that of Khatter et al. (9) (the application of Cosine similarity in combination with sentiment analysis) for movie recommender system was discovered as beneficial in finding and recommending similar objects.

2.2.3 Recommender systems

The Recommender system will suggest the alternative information for more selections in deciding to purchase goods or services which may also be interested (8-11). This is based on the historical information or reference from users with similar behavior or characteristics. Recommender systems are classified into three conceptual types which are:

1) Content-based Filtering

Content-based Filtering system will suggest the information base on the filtering of contents that may be interested by users based on the historical data and the behavior of an individual user (3, 10, 13-15). Base on the characteristics of objects, the database of individual user will be retrieved and evaluated for similarities to recommend the information that similar to what user may need. Content-based recommender system was recommended to combine with the Machine Learning (ML) and matrix factorization techniques as the modern focus (16).

2) Collaborative Filtering

Collaborative Filtering system will suggest the alternative information based on user reviews and scorings toward goods or services. This is based on not only an active user's preferences towards the others' preferences (13, 14), but also historical interaction patterns (2). This is to investigate and compare the behavior similar consumers. The system will match the similarities among users, based on the scoring to suggest the information.

3) Hybrid Filtering

Hybrid Filtering system is the combination of two previous types to merge the advantages of each type to solve the cold-star problem and to enhance the better efficiency. The two previous types above can be adopted with combination of the other approaches, such as the fuzzy expert system (20), knowledge-based system (6), and so on. This is to be beneficial from not only the complementary advantages of both systems above, but also the combination with the other techniques as well (7).

2.2.3 Web Application

Web Application is the application that is accessible via web browsers through either the internet or intranet. Web application enhance the better ubiquity which composes of Client-side Technology (web browsers, plug-in, adds on, or extensions) and Server-side Technology (web application, Front-End and Back-End Technologies, and web server). Regarding the application with recommender system, the possibility is even more to offer personalized and context-sensitive recommendations (4, 11).

2.2.4 Related Work

Base on the popular search keyword on Google (Thailand) in the past Covid-19 epidemic period, most searching keywords about food menu and cooking instructions are historically increased. There is new development of web applications to anticipate home user during the stay-home-for-better-safety period as the examples below.

1) SuperCook (www.supercook.com)

With the concept as being a "Zero Waste Recipe Generator", menus from the stocking grocery items can be selected, based on user input or pre-set menus. However, most available cooking menus are westernized and most user interfaces are inconvenient since they are not designed for Thai cooking context as seen in the homepage demonstrated in Figure 1.

Figure 1 The Homepage of supercook.com (www.supercook.com)

2) Recipeland (www.recipeland.com)

Recipeland.com is the website to determine the cooking menu from the stocking grocery items. Beyond the list of stocking groceries, specific condition can also be added such as the cooking time and so on. However, web pages must be refreshed at each time and also no web instruction in Thai language, as seen in the homepage demonstrated in Figure 2.

Figure 2 The Homepage of recipeland.com (www.recipeland.com)

3) Allrecipes Dinner Spinner (www.allrecipes.com)

Allrecipes Dinner Spinner is the cooking menu recommendation website base on the stocking groceries. The advantage is that both main dishes and dessert menus are available. However, the searching of cooking menu must be manually typed in without auto fill-up. The homepage of this site is shown in Figure 3.

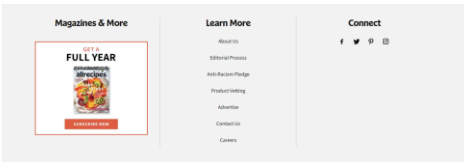

Figure 3 The Homepage of Allrecipes Dinner Spinner (www.allrecipes.com)

3. Methodology

This section will demonstrate the system framework and operations, consisting of input, processing, and output. Beginning from the use case diagram as shown below (in Figure 4), users can be both registered users and guest users. The registered users can be logged into the system via Facebook API with more privilege to utilize the system more that of guest users (visiting users who are not willing to sign up).

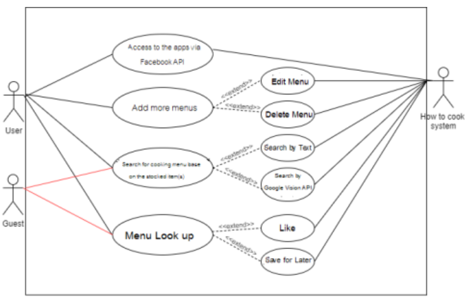

Figure 4 System's Use Case Diagram

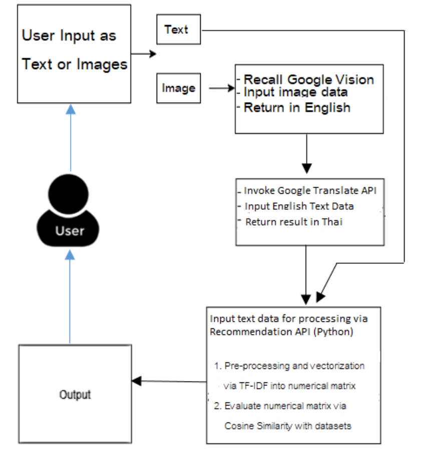

Figure 5 System Framework

The system framework as shown in Figure 5 above illustrates how the system operate beginning from data input to output as described below.

1) Text Input

Input data about the stocked grocery item(s) can be entered as text via web page and then processed by API Recommendation.

2) Image Input

Image as input can be imported via uploading image files from the devices or via web camera. Then, the Google Cloud Vision API will be employed as image processor and deliver output as text in English. Consequently, the Google Translate API (12) will translate the output text into Thai and will eventually process via Recommendation API.

Both input text and images will be processed for similarity and will thus compare to the data about the stocked groceries in order to recommend the suitable menus to be cooked. TF-IDF and Cosine similarity were implemented in the recommender system (to find the similarity) with Training Set and Test Set. The sub processes in the entire text processing procedures can be described as the follows.

Beginning with the text pre-processing, system will pre-process the input text data to be ready for TF-IDF vectorization (19). The common pre-processing procedures include:

• tokenization - where the input text is separated into individual words of tokens

• elimination of stop words - the deletion of common words that may provide none meaningful information including articles (a, an, the, this, that, etc.), prepositions (in, on, at, ect.), and so on.

• lowercasing – all the input text from the previous steps will be lowercased to ensure consistency

After pre-processing of text input, TF-IDF Vectorization is the consequence. This technique is to convert the text descriptions into numerical vectors to enable the representing of text data into numerical format that is suitable for calculating the cosine similarity.

In this context, there is the calculation of cosine similarity between the TF-IDF vectors of menu items (as pre-processed text input) where the angle between two vectors and is a common metric for text similarity would be measured. Coding for TF-IDF and Cosine are coded with Python on Jupyter, as shown in Figure 7. This is to process to determining the weighting of term (grocery items). Cosine similarity is the calculation of similarities in each possible menu. The 360 Training data sets for training the system to evaluate the accuracy is also demonstrated in Figure 6.

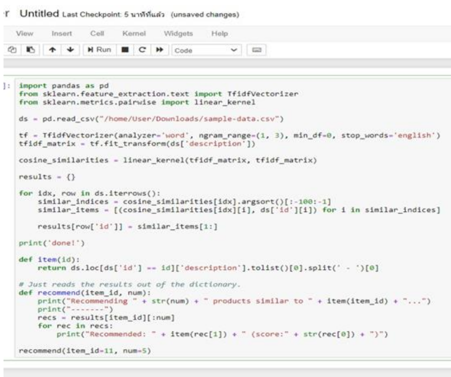

Figure 6 Code for TF – IDF and Cosine

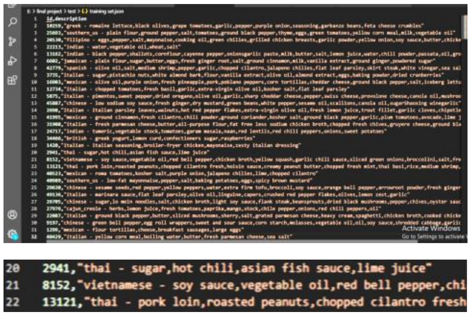

Figure 7 Example of Training Data Set

The Test Sets as data that are imported into the system to for accuracy test in TF-IDF and Cosine calculation (as shown in Figure 7).

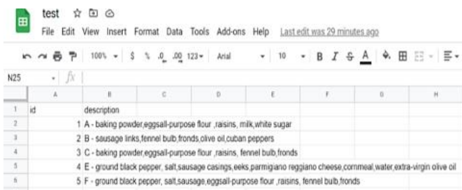

Figure 8 Example for the Test Set data

Regarding similarity test for recommender system, each command implemented in finding similarity of cooking menus is accurate as needed. In this test, B is assumed to be the input for searching and the outcome as closely similar menus are C and F (in Figure 8).

After the overall processing, the system outcome will be displayed as cooking menus which have at least three input grocery items as cooking part(s) of ingredients, as pre-set condition.

3.1 Results and Discussion

As the development outcome, the completed recommender system is included with the following features.

3.1.1 Textual Input

Textual input can be manually type in via text box and confirmation pop-up box will be displayed prior to proceeding the textual input data processing. Then, the output as recommended cooking menus with the input as cooking ingredient(s) will be display (as shown in Figures 9-10).

Figure 9 Textual Data Input

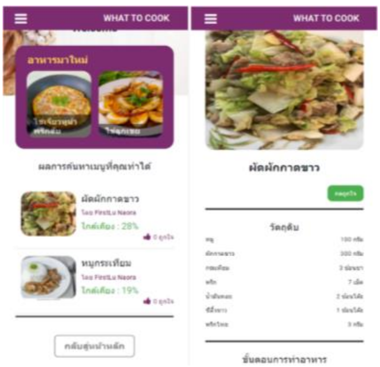

Figure 10 The Searching Result as Outcome

3.1.2 Image Input

As stated above that pictorial input can be imported to the system via techniques: uploading from the devices and from camera, the outcome was similar as those of textual input method stated previously (as exhibited in Figures 11-12).

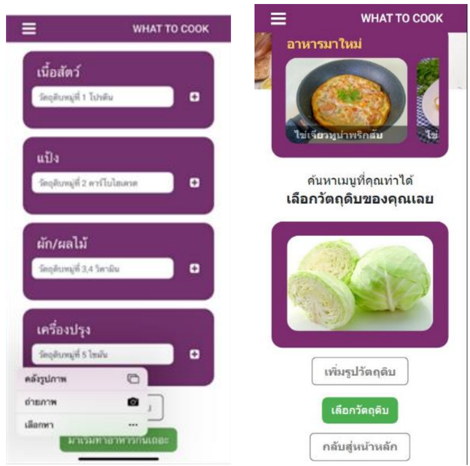

Figure 11 Image Data Input

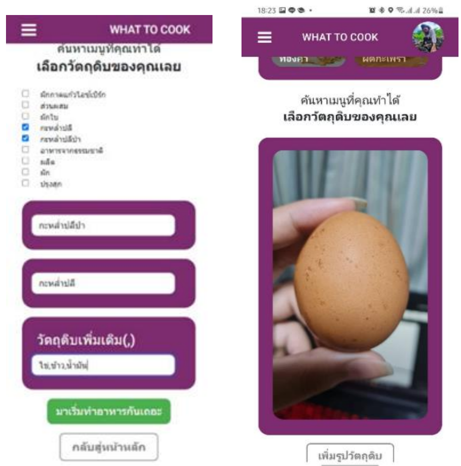

Figure 12 Image Data Input

3.1.3 Adding Cooking Menu(s)

Through "add menu" page, extra cooking menus can be added into the recommender system. This includes grocery ingredients, cooking procedures and illustration image(s). The newly added menu can be previewed after finish adding, via user profile page (Figures 13-15).

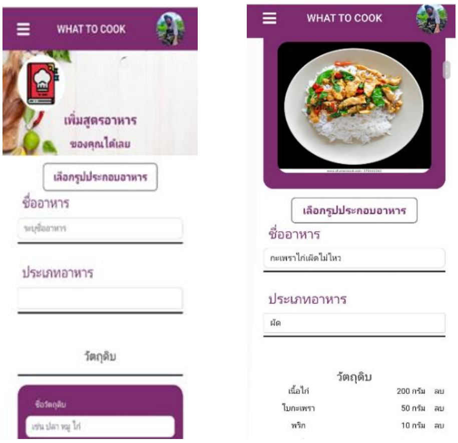

Figure 13 New Menu Adding

Figure 14 New Menu Adding

Figure 15 New Menu Adding

The results are supported and strengthened by the study of (22) in that the cosine similarity method gives the best value of proximity or similarity compared to Jaccard similarity and a combination of both.

The developed recommender system based on the TF-IDF and Cosine Similarity yielded the desirable results but however, there are system limitations as the following:

1) The pre-set menu within the system are rather few that may result to the fewer recommended menus per one searching output.

2) The image data input via Google Cloud Vision API yielded under expected result in that the processed output were diversified pertaining to one image data input. The counter solution to this limitation was to allow users to checkbox the co-ingredients.

The above limitations may affect the system's actual efficiency in some degree but the continuous improvement of Google Cloud Vision with the more pre-set menu list would enhance the better overall system efficiency in the future. Also, the image processing devices and technologies those are continuously improved would also enhance the better quality of input image data and thus improve the overall system efficiency as well.

4. Conclusions

The developed Web-based Cooking Recipe Recommender System, based on the Stocked Groceries, where input data can be via either textual or pictorial. As web-based design, the local application installation is not required. The data processing yielded the expectable output as the recommended cooking menus that included the input grocery items as ingredients. Regarding image data input, user checkbox can allow the better accuracy and thus partially resolve the limitations of Google Cloud Vision API. System operates based on the following theoretical methods and techniques:

- TF-IDF as processor to determine the weighting of term (herein as grocery items)
- Cosine as similarity calculator for each menu
- Google Could Vision API for image data input

This developed system utilizes the combination of TF-IDF and cosine similarity yielded the desirable accuracy which is supported by the study of (22) which confirmed that the cosine similarity method yields the best value of proximity.

5. Recommendation for Future Development

Since the system is initially developed as web-based application with the combination of TF-IDF and Cosine Similarity, there are recommendations for future improvement as the follows.

1) The pre-set menus should be added beyond the current 150 menus which will enhance the better searching results and outcomes.

2) The further development should enable the multichannel system accessibility, such as via Google account, LINE account, and typical membership application.

3) The application-particular image processor should be further developed for the better accuracy and to support the further development.

Acknowledgements

It is the deep appreciation to the research project sponsorship via the annual revenue budget of the Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT) and of the Faculty of Engineering, Rajamangala University of Technology Phra Nakorn (RMUTP), 2023 fiscal year. The KMUTT and RMUTP executive supports are also deeply appreciated pertaining to the accomplishment of this research project.

Declaration of conflicting interests

The authors declared that they have no conflicts of interest in the research, authorship, and this article's publication.

References

1. Abid MA, Mushtaq MF, Akram U, Abbasi MA, Rustam F. Comparative analysis of TF-IDF and loglikelihood method for keywords extraction of twitter data. Mehran Univ. Res. J. Eng. Technol. 2023;42(1):88–94.
2. Alharbe N, Rakrouki MA, Aljohani A. A collaborative filtering recommendation algorithm based on embedding representation. Expert Syst Appl. 2023;215: 1-11.
3. Amara S, Subramanian RR. Collaborating Personalized Recommender System and Content-based Recommender System Using TextCorpus. Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS); 2016 Jan 22-23; Coimbatore, India. p. 105-9.
4. Azizi M, Do H. A collaborative filtering recommender system for test case prioritization in web applications. Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18). Association for Computing Machinery; 2018; New York, USA. p. 1560-7.
5. Bafna P, Pramod D, Vaidya A. Document clustering: TF-IDF approach. Proceeding of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT); 2016 March 3-5; Chennai, India. p. 61-6.
6. Burke R. Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter: Springer; 2002. (12). p. 331–70.
7. Çano E, Morisio M. Hybrid Recommender Systems: A Systematic Literature Review. Intell Data Anal. 2017;21(6):1487 – 1524. doi:10.3233/IDA-163209
8. Gomes L, Torres RDS, Côrtes ML. BERT- and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: A comparative study. Inform Software Tech. 2016;(160),107217:1-12.

9. Khatter H. Goel N, Gupta N, Gulati M. Movie Recommendation System using Cosine Similarity with Sentiment Analysis. Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA); 2021 Sep 2-4; p. 597-603. doi: 10.1109/ICIRCA51532.2021
10. Lops, P, Gemmis MD, Semeraro G. Content-based Recommender Systems: State of the Art and Trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds) Recommender Systems Handbook: Springer; 2011. https://doi.org/10.1007/978-0-387-85820-3_3.
11. Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: A survey. Decis Support Syst, 2015;(74):12-32.
12. Neves A, Lopes D. A practical study about the Google Vision API. Proceedings of the 22nd Portuguese Conference on Pattern Recognition; 2016; Aveiro, Portugal. Available from https://www.researchgate.net/publication/309642837_A_practical_study_about_the_Google_Vision_API
13. Papadakis H, Papagrigoriou A, Eleftherios K, Panagiotakis C, Markaki S, Fragopoulou P. Content-Based Recommender Systems Taxonomy. Found. Comput. Decis Sci. 2023;48(2):211-41.
14. Papadakis H, Papagrigoriou A, Panagiotakis C, Kosmas E, Fragopoulou P. Collaborative filtering recommender systems taxonomy. Knowl Inf Syst. 2022;64:35–74.
15. Park DH, Kim HK, Choi IY, Kim JK. A literature review and classification of recommender systems research. Expert Syst Appl. 2012;39(11):10059-72.
16. Pérez-Almaguer Y, Year R, Alzahrani AA, Martínez L. Content-based group recommender systems: A general taxonomy and further improvements. Expert Syst Appl. 2021;(184),115444:1-21. https://doi.org/10.1016/j.eswa.2021.115444
17. Qaiser S, Ali S. Text Mining: Use of TH-IDF to Examine the Relevance of Words to Documents. Int. J. Comput Appl. 2018; 181(1):25-29.
18. Singh RH, Maurya S, Tripathi T, Narula T, Srivastav G. Movie Recommendation System using Cosine Similarity and KNN. Int J of Eng Adv Technolo. 2020;9(5):556-9.
19. Uthaisuri T. Keyword extraction from English abstracts [master's thesis]. Nakhon Pathom (TH), Silapakorn University; 2013.
20. Walek B, Fajmon P. A hybrid recommender system for an online store using a fuzzy expert system. Expert Syst Appl. 2023; (212):1-16. https://doi.org/10.1016/j.eswa.2022.118565.
21. Wu HC, Luk RWP, Wong KF, Kwok KL. Interpreting TF-IDF term weights as making relevance decisions, ACM Trans. Inf. Syst. 2018;26(3):1-37.
22. Zahrotun L. Comparison Jaccard similarity, Cosine Similarity and Combined Both of the Data Clustering with Shared Nearest Neighbor Method. Computer Engineering and Applications. 2016;5(1):11-8.
23. Zhang W, Yoshida T, Tang X. A comparative study of TF*IDF, LSI and multi-words for text classification. Expert Syst Appl. 2011;38(3):2758-65.

