 Iraqi Journal of Science, 2019, Vol. 60, No.1, pp: 143-156
DOI: 10.24996/ijs.2019.60.1.16

ISSN: 0067-2904

Location Aspect Based Sentiment Analyzer for Hotel Recommender System

Alia Karim Abdul Hassan[1], Ahmed Bahaa Aldeen Abdulwahhab*[2]
[1]Department of Computer science, University of Technology, Baghdad, Baghdad, Iraq
[2]Department of Informatics, Middle Technical University, Baghdad, Baghdad, Iraq

Abstract

Recently personal recommender system has spread fast, because of its role in helping users to make their decision. Location-based recommender systems are one of these systems. These systems are working by sensing the location of the person and suggest the best services to him in his area. Unfortunately, these systems that depend on explicit user rating suffering from cold start and sparsity problems. The proposed system depends on the current user position to recommend a hotel to him, and on reviews analysis. The hybrid sentiment analyzer consists of supervised sentiment analyzer and the second stage is lexicon sentiment analyzer. This system has a contribute over the sentiment analyzer by extracting the aspects that users have been mentioned in their reviews like (cleanness, service, etc.) by using accurate parsing system built on latent semantic analysis results. The accuracy measurements of the proposed sentiment analyzer were perfect.

Keywords: sentiment analysis, recommender systems, location-based, natural language processing, ABSA

نظام توصية جغرافي للفنادق معتمد على تحليل الشعور لتعليقات الزبائن

علياء كريم عبد الحسن[1] ،احمد بهاء الدين عبد الوهاب[2]*
[1]قسم علوم الحاسبات ، الجامعه التكنلوجيه ، العراق ، بغداد
[2]قسم المعلوماتيه ، الجامعه التقنيه الوسطى، العراق ، بغداد

الخلاصة

انتشرت في السنوات الاخيرة انظمة التوصية الشخصية لما لها من دور في مساعدة مستخدميها في اتخاذ قراراتهم. وبرز احد انواعها انظمة التوصية الجغرافية التي تستقرئ موقع الشخص لتقدم له التوصية حول افضل الخدمات المحيطة به. انظمة التوصية المعتمدة على تقييم الزبون الرقمي منشرة ولكنها تعاني من مشاكل قلة او اهمال تقييم الزبائن للسلع و الخدمات او مايسمى cold start و الذي يولد فراغات في البيانات مما يعيق عملها بدقة اضافه الى ان التقييم الرقمي لا يعكس ملاحظات الزبون و شعوره عن جوانب السلعة المتعددة. يقترح هذا البحث نظام توصية يعتمد على معرفة موقع الشخص الحالي لاقتراح افضل فندق له. هذا النظام يعتمد بالاضافه للموقع على تحليل توصيات الزائرين السابقين لذلك الفندق. تم اقتراح نظام هجين لتحليل شعور المستخدمين السابقين من تعليقاتهم ، ويعتمد على مرحلتين الاولى بتقنية تعلم الماكنة الموجه، و المرحلة الثانية يعتمد على تقنية القاموس. كذلك تم تطوير هذا النظام ليكون قادر على استخلاص الجوانب التي ركز عليها الشخص في تعليقه مثل النظافه، سلوك الكادر، الخدمات و جوانب اخرى ، وذلك بنظام اعراب دقيق يستخلص الجملة الاسمية و يصنفها حسب مجموعة من الكلمات المحددة بمرحله تحليل

*Email: ahmed80.ab@gmail.com

المعاني الضمنيه latent semantic analysis ، كانت نتائج الدقه لنظام تحليل المشاعر من النصوص جيده جدا.

1. Introduction

The rapid growth of various e-commerce and social network services, such as Amazon, Foursquare, and Gowalla have caused in sheer a massive amount of data collected by service's providers 1. Because of this enormous amount of data, the need for designing of more intelligent information retrieval system known as recommender systems 2. Most of the recently used location-based recommender systems use rating's customer feedback to help customer's decision-making process, but the reliability of these rating is questionable3. So the reviews provide a better insight about the hotel services. While most travelers do not have the time or patience to read all the reviews, the need to summarize the aspect of these reviews arise 4. Many researchers have been worked on location opinion-based recommender systems using many paradigms. For example, najeefa Ch.5 has worked on sentiment analysis for tweets about tourism location in Bangladesh. Najeefas' design of sentiment analyzer used to reveal the polarity of tweets about places to either positive or negative using lexicon approach. That SA did not extract aspects from reviews to understand what the reviewers focus on their comments. Zhang et al.3 handle in thier research a method to analyze text comments using latent Dirichlet analysis (LDA), then extract the rate of each aspect in the review. Zhang did not apply aspect-based sentiment analysis on geo-location applications. Patit et al. 6 used the supervised model of naïve bays to classify the polarity of the reviews into three (positive, negative and neutral, but this system is not a recommender system it is more close to being a support decision system or intelligent business system. These support decision systems are used to find out the rends of public people about service or a product in social networks. Mohit et al. 7 built a geo-notification system. This system gives a user a notification to avoid a specific venue based on opinions of previous visitors by analyzing the sentiments of their reviews. Mohit paper is close to this proposed work, but the main difference is that Mohit did not extract aspect to be more specific about the reasons to avoid such a venue, while this work bases on aspect based SA which is more complicated and need more advanced aspect extraction, aspect categorization processes before SA

2. Problem statement

The traditional location-based RS that are depending on rating scores suffers from three issues, first, cold start problem. This problem appears when the system has to suggest a venue for a new user, who has insufficient check-ins. This insufficiency in new users' ratings results in the absence of similarity with other users. The second issue is data sparseness; the sparseness is coming from a few numbers of ratings to venues because of a little number of places visited by the users. This sparseness may lead to difficulty in finding similar uses to the recommendation process. Lastly, the scalability issue that resulted from the fast expansion of users check-ins and ratings, causing in difficulty in administrating the check-ins dataset [1].

This research tries to overcome the mentioned problems by using user's locations and analyzing reviews of the previous visitors of the specific venue or site. Cold start issue is resolved by depending on other users' opinions rather than targeted users check-ins. The effect of data sparseness will also be illuminated by using previous visitors' opinions, so the lack in check-ins and rating of the targeted user will be not important. The proposed system consists of three parts. The most crucial part is aspect extraction part that responsible for aspect phrase extraction from reviews, the second part is the SA part that specifies the polarity of each aspect phrase to either positive or negative. The third part is location-based RS that use knowledge from extracted aspects about each venue.

3. Contributions

The contributions of this paper are first, build a location-sentiment based recommender system that depends on user location and previous visitor reviews, which give more details about hotel quality and services. Reviews would be adequate than rating system to reflect the opinion of users about service. The second contribution is building hybrid sentiment analyzer that consists of two parts, supervised sentiment analyzer that is first part learned from reviews, the second part is unsupervised or lexicon based that handle new reviews that are coming with new idioms and words not used in the training dataset. Our sentiment analyzer also controls the issue of negations before training the model. The third contribution is summarizing the reviews by extracting the essential aspects from user review to reach an excellent understating about the positive and negative aspects of the targeted venue/hotel.

4.Opinion mining and sentiment analysis

Opinion mining means extracting and analyzes people's opinions about an aspect of an entity, while sentiment Analysis (SA) specifies the sentiment or expression behind the text and interpret it 8. The problem of opinion mining can be viewed as natural language processing issue integrate to recommender systems. While recommender system analyzes ratings to predict user's needs, opinion mining analyzes text to guess users like dislike and sentiments 9. SA can be classified into three levels: document-level sentiment analysis which its goal is to discover whether the document expresses positive or negative about an issue10. Sentence –level SA. This type will determine whether conviction show positive or negative. The third class is aspect-level SA which is aims to classify the sentiment concerning specific aspects of entities. Specifying aspects can be accomplished by firstly identifying the entities and their aspects. The opinion holder can give a different opinion about different aspects of some object like "the hotel's restaurant was good, but the staff was rude." That is talking about two aspects of the restaurant and staff11. Aspect-based sentiment analysis needs two operations to be performed on reviews statements; these are; aspect extraction process that can be performed using one of the following methods :

1-extraction using frequent nouns and frequent phrases

2-extraction by using syntactic relation inside the noun phrase, these relations take two forms (syntactic dependencies, and lexico-syntactic patterns)

3-extraction of aspects using supervised learning

4-extraction of aspects using topic modeling methods like (latent semantic analysis or latent Dirichlet analysis).

After specifying and extracting aspects, the process of sentiment analysis coming to determine the positive aspects and negative aspects of an item (service, product, event, etc.) 11. There are two approaches to implement sentiment analysis, the machine learning approach (supervised) and the lexicon-based approach (unsupervised) which can be divided into dictionary based SA and lexicon-based approach8. These categories are clear in Figure-1.

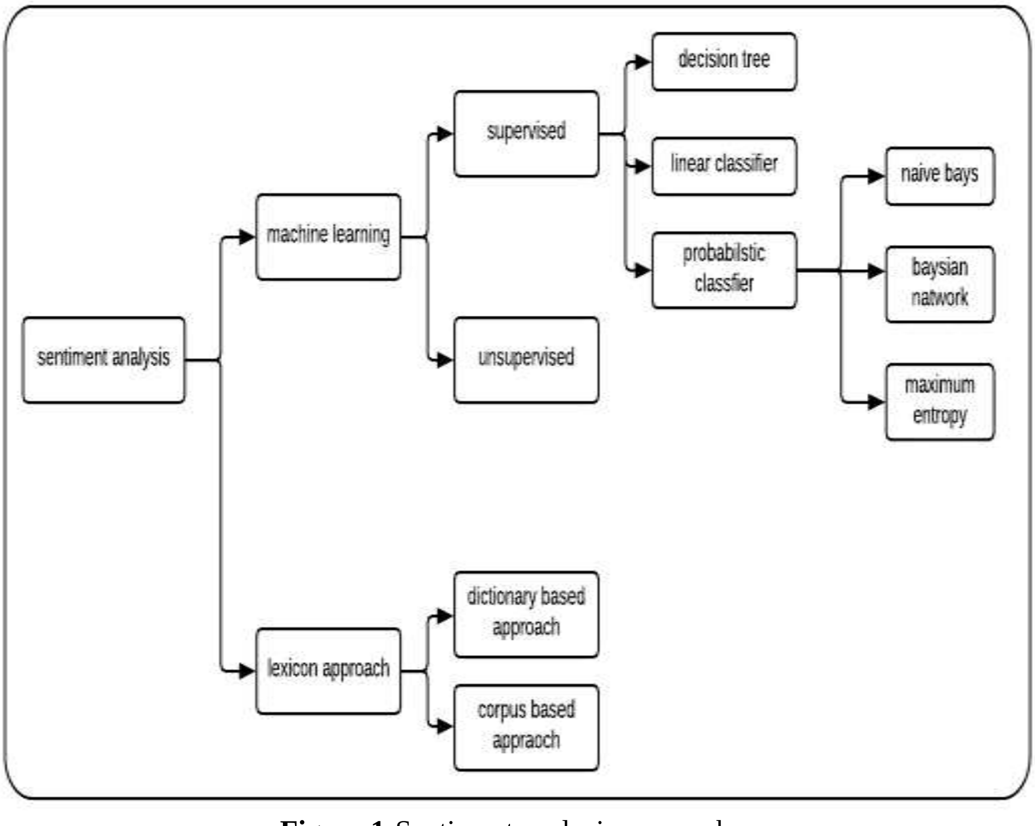


Figure 1-Sentiment analysis approaches

These can be presented in more details as follows:-

4.1 Machine learning approach

Machine learning approach depends on the known machine learning algorithms to solve sentiment analysis text as a classification issue. The classification model use features in the underlying records to one of the class labels, the model use this functions to predict the class for given reviews text 8. The machine learning approach used for sentiment analysis has to be supervised, these models depend on existing of labeled reviews dataset, which is marked as either positive or negative. Many researchers used different machine learning models for sentiment analysis. There are three commonly used classifiers in SA (naïve bays, maximum entropy, and support vector machine). The experiments showed that feature presence (presence of specific words) is more important than feature frequency. Naïve bays perform better than SVM, while SVM performs better when the dataset is big. Maximum entropy may have better performance than SVM, but it may suffer from overfitting that may come from training by using many features or many examples 10.

4.2. Lexicon based approach

This approach is unsupervised, SA is done by computing features of a given review against sentiment lexicon whose sentiment values are determined before. Opinion words exploited in sentiment classification tasks. Positive words use to express preferred states, otherwise negative words used to represent undesired states 11. There are three methods to collect opinion word. The manual manner which is not practical regarding time and two other ways are:-

4.2. A. Dictionary based approach

A simple process to build a dictionary. This method begins with the small seed sentiment words with known positive or negative orientation collected manually. The algorithm then expands this set of words by searching the wordnet or another online dictionary for synonyms and antonyms of the harvested seed. The words added to the dictionary gradually until no words found in the dataset 11.

4.2. B. Corpus-based approach

This approach adopts a general purpose sentiment lexicon to a new one using a domain corpus for an SA application. The corpus-based helps to solve the issue of finding opinion word with context-specific orientation. This method depends on the syntactic pattern. The original way described by Hazivassiloglou and McKeown in 1997. They begin with the seed of adjectives and use it with a set of linguistic constraints. The constraints are conjunction like (AND) for example conjured adjective have some orientation. There is also opposite expression like (BUT) which are an indicator of opinion change. So this method needs parsing reviews with a lexicon to determine the sentiment polarity11.

5. Latent semantic analysis and aspect extraction

One of the contributions of this paper is extracting aspects. Before extracting aspects, this work suggests specifying those aspects mentioned in reviews by using latent semantic analysis (LSA) to define essential topics and the words that represent it. LSA is a mathematical method used to determine the relationship between terms and concept in content. This method used in author recognition, search engines, detecting plagiarism, and comparing text similarities. The central principle behind the LSA is that similar terms tend to be used in the same context and hence tend to co-occurrence more. This technique can uncover latent hidden terms which correlate

Semantically to form topics [9]. Consider a data matrix contains m documents and n terms, the terms columns represent terms and words mentioned in these documents. This matrix can be described as a composition of three matrices as in equation (1) 12:

$$A = U \sum V^T \qquad \text{--------------- (1)}$$

A m×n is input data matrix (m documents, n terms), U m×r (left singular vector) consist of (m documents, r concept), $\sum$ r×r singular value diagonal matrix represent the strength of each concept, and V or right singular vectors are an n×r matrices (n terms and r concept). Each column of U represents a concept while $\sum$ representing the strength of concept and V^Tmap the document or a review to the concept 912.

6. Proposed system

The location-based sentiment recommender system is built and experimented over standard dataset from kaggle.com*. The dataset consists of hotel name, review, and the coordinate location of each hotel. This data is provided by DATAFINITI to do experiments in sentiment analysis in kaggle.com. This data set is chosen because it contains hotels names, and users' reviews for 1000 hotels with the longitude and latitude of each hotel to assist in mapping the hotel location in the map of the RS. The proposed system in Figure-2:

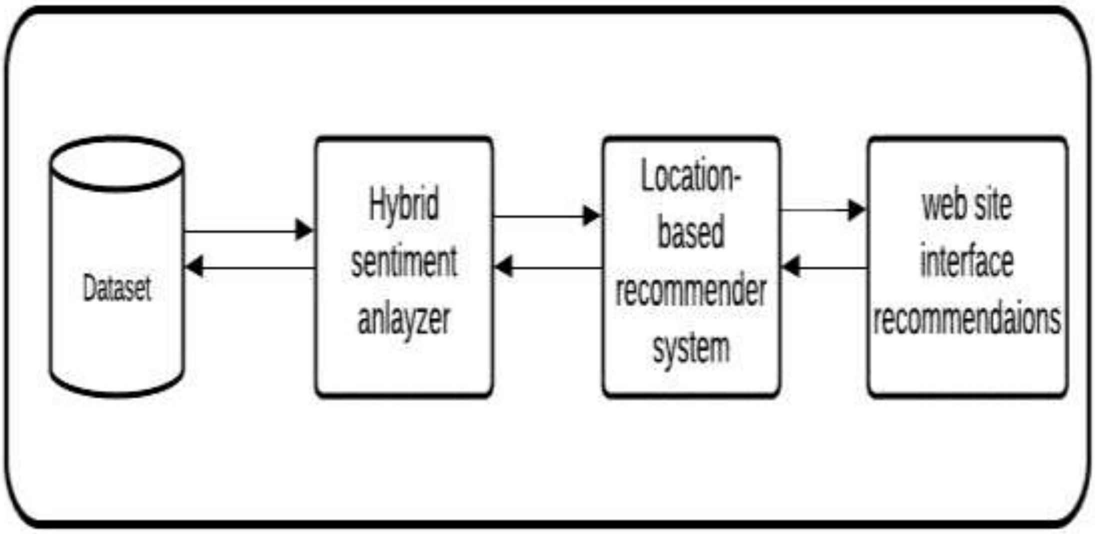


Figure 2-Proposed system parts

From Figure-2 the reader can notice the following elements:
1. Dataset: - from kaggle.com contains 3000 records for hotels in USA region only.
2. Sentiment analyzer:- Input:- reviews in the dataset
   Output: - polarity of each review (positive or negative)

The sentiment analyzer is the most crucial part of the proposed system, as it computes the sentiment to positive or negative. This recommender system depends on the attitude of reviews to decide if the venue or hotel is an excellent choice for the user. The sentiment analyzer consists of two parts (hybrid), the first part is supervised and the second one is lexicon based (unsupervised). The philosophy of this SA is reading the review, then calculate the percentage of the review's words in TF/IDF of the reviews dataset, if the majority of review's words exist in TF/IDF then use the supervised SA, otherwise use the lexicon based SA. See flowchart in Figure-3.

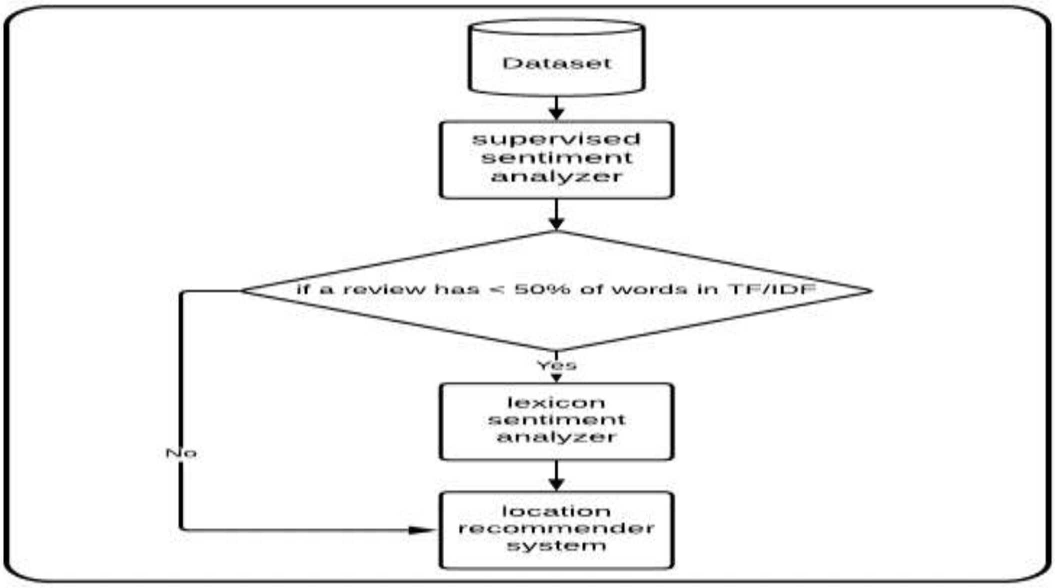


Figure 3- Hybrid sentiment analysis approaches

The review also refined from negation before entering the SA. system to increase the accuracy. The negation case like (not good) substituted by (bad). See algorithm (1) below to understand the hybrid sentiment analyzer

Algorithm 1: Hybrid sentiment analysis algorithm
Input: user's reviews
Output: - the sentiment polarity of each review (positive or negative)
1. Start **2. For I =0 TO the total number of reviews :** **a. if a review (i) contain negation then:** **• Negation handler** **• Else: go to step 4** **3. Convert the reviews to bigram representation** **4. Send bigram of review (i) to supervised sentiment analyzer** **5. If review (i) words in TF/IDF dataset < 0.50 then :** **a. Else use a lexicon sentiment analyzer to calculate sentiment for review (i).** **6. End**

Now, this part explains the algorithm to build the supervised SA. Mentioned in step 5 of algorithm1. To construct a supervised machine learning model the sentiment analyzer need to preprocess the reviews as in algorithm (2) : -

Algorithm 2: reviews the preprocessing algorithm
Input: n reviews dataset
Output: Corpus of n normalized reviews
i=0; n=number of reviews in dataset; **1. Start** **2. While (i< n) do** **3. input review (i)** **4. keep English characters [A-Z,a-z] in review(i)** **5. convert review(i) to lower case** **6. for all words in review(i) do :** **Stem (word) in review (i)** **Remove stop word in review (i)** **7. add review to list corpus[ ]** **8. i=i+1** **9. end**

The next step is to build the bag of words (bigrams), then dividing the dataset into 0.75 training set and 0.25 test set. The machine learning model naïve bays is ready to be trained to predict the sentiment of each review in the dataset to either positive or negative, and test its accuracy using test set, look up algorithm (3)

Algorithm 3: reviews SA machine learning model
Input: a corpus of normalized reviews [ ]
Output: trained machine learning system to predict sentiment review to (positive, or negative)
1. Start **2. Build a bag of the word (bigram) from cleaned normalized corpus reviews dataset** **3. Divide bag of the word into 0.75 training set and 0.25 training set** **4. Train the machine learning model using training dataset** **5. Test the sentiment analysis ML model using testing dataset** **6. End**

If the percentage of words in the entered new review is less than 0.5 in TF/IDF of the reviews dataset, the review is sent to the lexicon SA. This type of SA. Depends on the parsing sentence to extract features, then compare features to the lexicon to specify the polarity of these features, and from the sum of features polarity decide whether the phrase has positive or negative emotion, as in algorithm (4) below:-

Algorithm 4: lexicon based sentiment analyzer
Input: a review
Output: sentiment polarity (positive or negative)
1. Start **2. Find part of speech POS for review (i)** **3. For each word in POS :** **a. If POS (word) in review(i) in [NNS,NN,NNPS,JJ,JJR,JJSRB,RBR,RBS,VB,VBSD,VBG,VBN,VBP] AND NOT IN stop-words:** **b. Lemmatize POS(word)** **c. # unify all verb tenses to verb** **# unify all noun cases to noun** **# unify all adjectives adverb cases to a, r respectively** **If POS (word) in set [VB, VBD, VBG, VBN, VBP, CVZ] THEN return V** **If POS (word) in [JJ,JJR,JJS] THEN return a** **If POS(word) in [RB,RBR,RBS] THEN return r** **If POS(word) in set [NN,NNS,NNP] THEN return n** **Else: return a** **4. Find score of each feature of review(i) using WORDNET lexicon** **5. If score > 0 then score = sum of scores L length (score)** **6. End**

This lexicon based SA. is domain independent, which means it can handle the cases of new reviews which is talking about topics out of the dataset scope and have a little number of words or idioms in the training dataset in the supervised SA part.

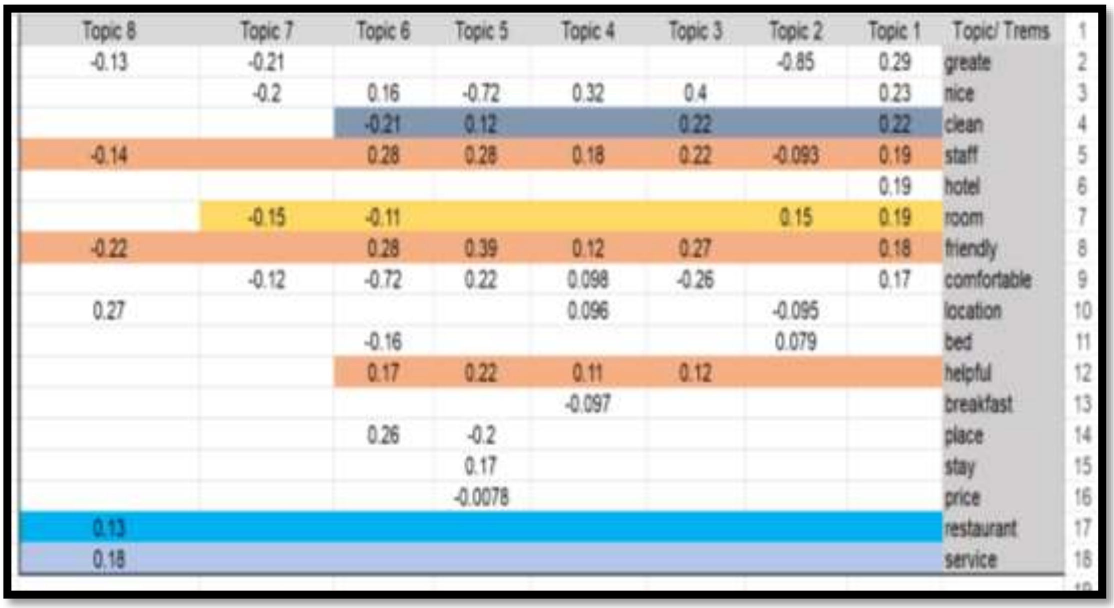
The following algorithm catches the cases of negations (not, isn't, etc.) and turns it into its synonyms, read algorithm (5):

Algorithm 5: negation handling algorithm
Input : part of speech of a review (i) **Positive_word_list [ ]= WORNET corpus ,negative_word_list [ ]= WORNET corpus** **Negation_list = [not, no, never, nothing, nowhere hardly, barely , n'\t']**
Output: negation reflected in a reviews(i)
1. Start **2. For each word in review(i):** **For word in nagtion_lis:** **if associative_verb in positive_word_list.words( ) then :** **Negative_sentiment + = 1** **Score= - score** **Else if associative_ver in negative_word_list .words( ) then :** **Positive_sentiment + = 1** **Score= score** **3. End**

The last part is the aspect extraction procedure, and how we could extract aspect. First, we perform latent semantic analysis to specify the essential aspects in the reviews of the users. The algorithm of implementing the first needs TD/IDF matrix to be built to compute term frequency in each review, then analyze the TF/IDF using LSA as in the following algorithm (6) :

Algorithm 6: Latent semantic analysis algorithm
Input: a dataset of reviews
Output: topic modeling for reviews contains topics and the degree of each word inside
1. Start **2. For each review in reviews dataset** • **Preprocess reviews** • **The stemming operation for a review** • **Stopword removing for a review** **3. Build frequency matrix (TF/IDF)** **4. Apply frequency matrix** **5. Decompose the frequency matrix into U, S, V** **6. Specify the number of topics** **7. Execute LSI** **8. Present analysis result** **9. End**

The result was as in Figure-4:



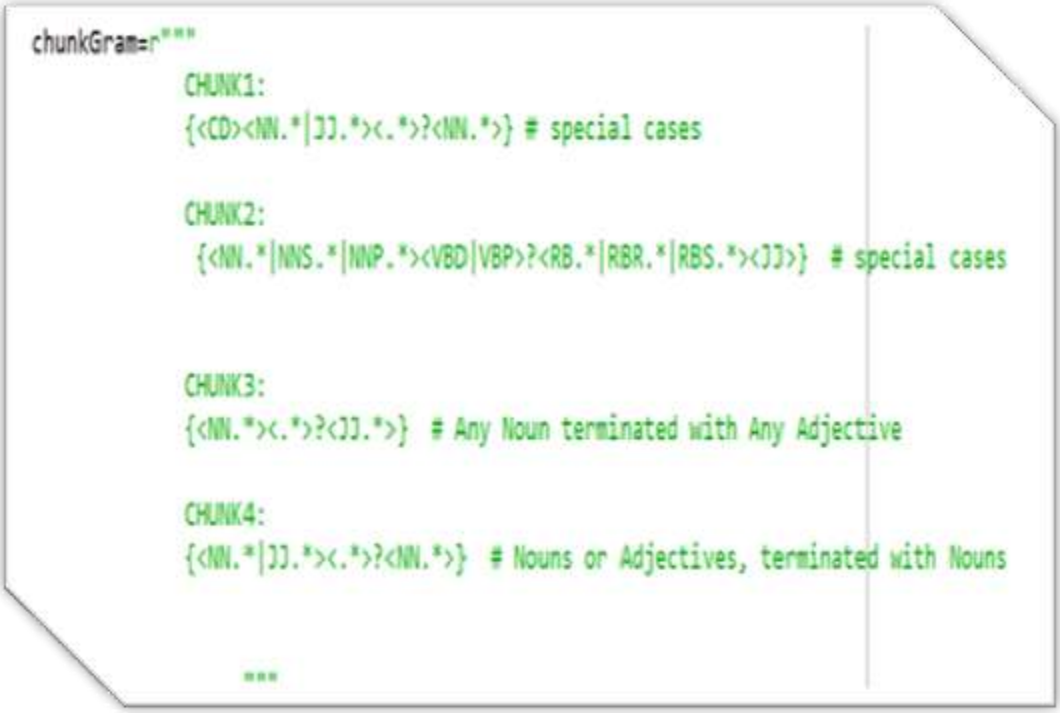
Topic 8	Topic 7	Topic 6	Topic 5	Topic 4	Topic 3	Topic 2	Topic 1	Topic/ Trems	1
-0.13	-0.21					-0.85	0.29	greate	2
	-0.2	0.16	-0.72	0.32	0.4		0.23	nice	3
		-0.21	0.12		0.22		0.22	clean	4
-0.14		0.28	0.28	0.18	0.22	-0.093	0.19	staff	5
							0.19	hotel	6
	-0.15	-0.11				0.15	0.19	room	7
-0.22		0.28	0.39	0.12	0.27		0.18	friendly	8
	-0.12	-0.72	0.22	0.098	-0.26		0.17	comfortable	9
0.27				0.096		-0.095		location	10
		-0.16				0.079		bed	11
		0.17	0.22	0.11	0.12			helpful	12
				-0.097				breakfast	13
		0.26	-0.2					place	14
			0.17					stay	15
			-0.0078					price	16
0.13								restaurant	17
0.18								service	18

Figure 4-Result of the Latent semantic analysis process

From Figure-4 the latent semantic analysis reveals the essential aspects mentioned by the users in the reviews dataset (location, hotel, staff, room, restaurant, clean, service), each aspect is represented by a set of words are:

- Location: location, position, close, near
- Staff: staff, helpful, friendly, warm, perfect
- Room: room, bed, rooms, beds, comfy, comfortable.
- Restaurant: breakfast, meal, lunch, restaurant.
- Clean: clean
- Service: service, bathrooms, reception, bath, spa, gym, Wi-Fi, TV. , air conditioner, everything.

The aspect sentence extracted using a chunking system that chunks the noun sentence cases and two other specific cases. Chunking process is building a parser according to patterns or rules specified by linguistic experts. First, we have to seek about particular cases, then catch the general form of a noun phrase, the chunk patterns are in Figure-5.



```
chunkGram=r"""
        CHUNK1:
        {<CD><NN.*|JJ.*><.*>?<NN.*>} # special cases

        CHUNK2:
         {<NN.*|NNS.*|NNP.*><VBD|VBP>?<RB.*|RBR.*|RBS.*><JJ>}  # special cases


        CHUNK3:
        {<NN.*><.*>?<JJ.*>}  # Any Noun terminated with Any Adjective

        CHUNK4:
        {<NN.*|JJ.*><.*>?<NN.*>}  # Nouns or Adjectives, terminated with Nouns

           """
```

Figure 5-Chunking parser to extract aspect from reviews

Chunk1: for example, "24 hours business trip."
Chunk 2: example, "staff were so friendly," and "staff was quite bad."
Chunk 3: for example, "building romantic."
Chunck4: for example, "nice location."
Lastly, the recommendation operation is done according to the procedure that depending on specifying the user his location; then the system gathers the hotels in that location to calculate the best place according to the mean of all aspects, as in algorithm (7):

Algorithm 7: Recommendation process algorithm
Input: aspect and sentiment dataset of hotels
Output: a recommendation of a hotel in a city
1. Start **2. City = Read city location from the user** **3. For each hotel in the city :** **4. Calculate mean of sentiment degree for** **Location, staff, room, restaurant, service, price, hotel** **5. Choose the hotel with the highest summation of aspect's degrees** **6. Drop duplicates by hotel name** **7. The present result to the user using table and gauge** **8. End**

7. System operation

The web interface has been implemented based on Flask framework for websites, written with python language, and tested on local server machine installed on windows 10 machine.
The system operations in steps are: -
1. The user enters the system site specifying his position (city) as in Figure-6)

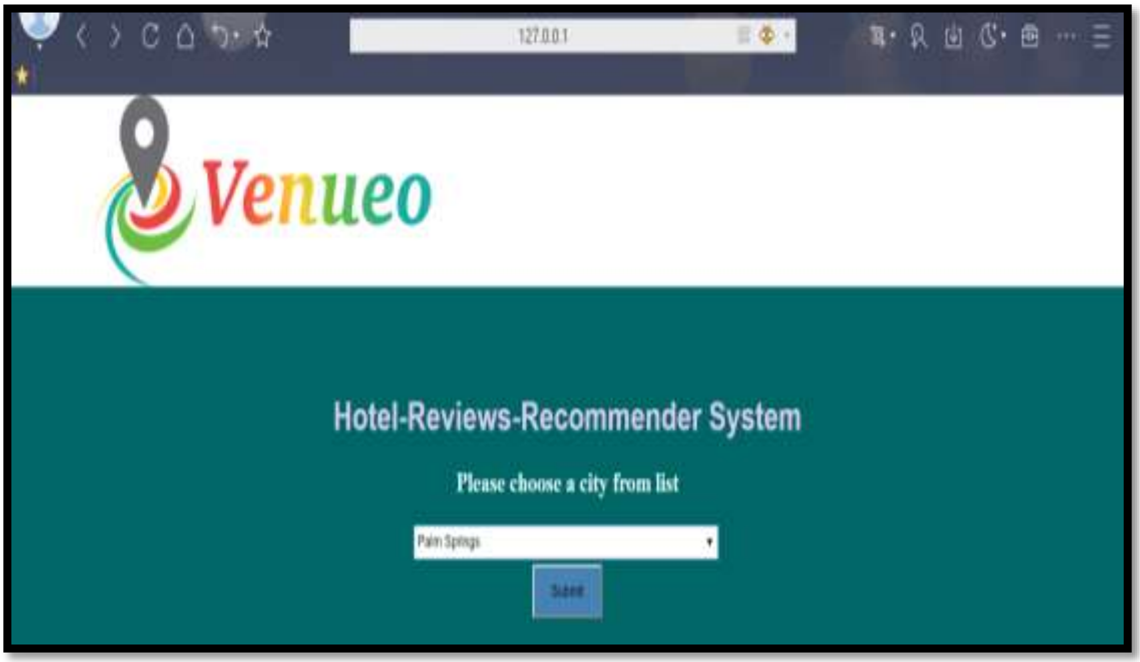


Figure 6-the main web interface for the proposed system

2. The system calculates the closest hotel according to user location and presents the sentiment of the hotel aspects with a summary of the latest review; also our system provides a gauge present the sentiment of aspects as in the Figure-7 below :

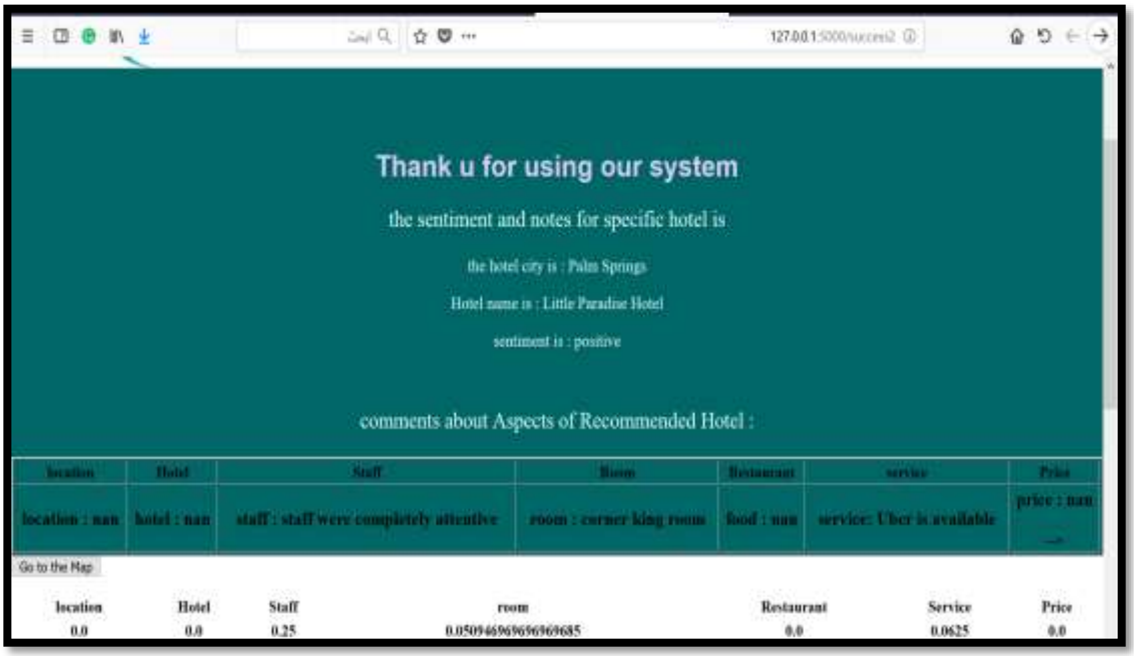


Figure 7-the result recommendation presented to a user

We can notice that aspects extracted for the hotel (little paradise) were about the staff with a score (0.25), the aspect extracted from the review was "staff were completely attentive." The second aspect was service "Uber is available," Uber is a taxi reservation service with a score (0.0625), and the third aspect is about the room (corner king room) with the score (0.050), while nothing about other aspects. This system can give a gauge for the aspects has been extracted form review as in Figures-(8,9).

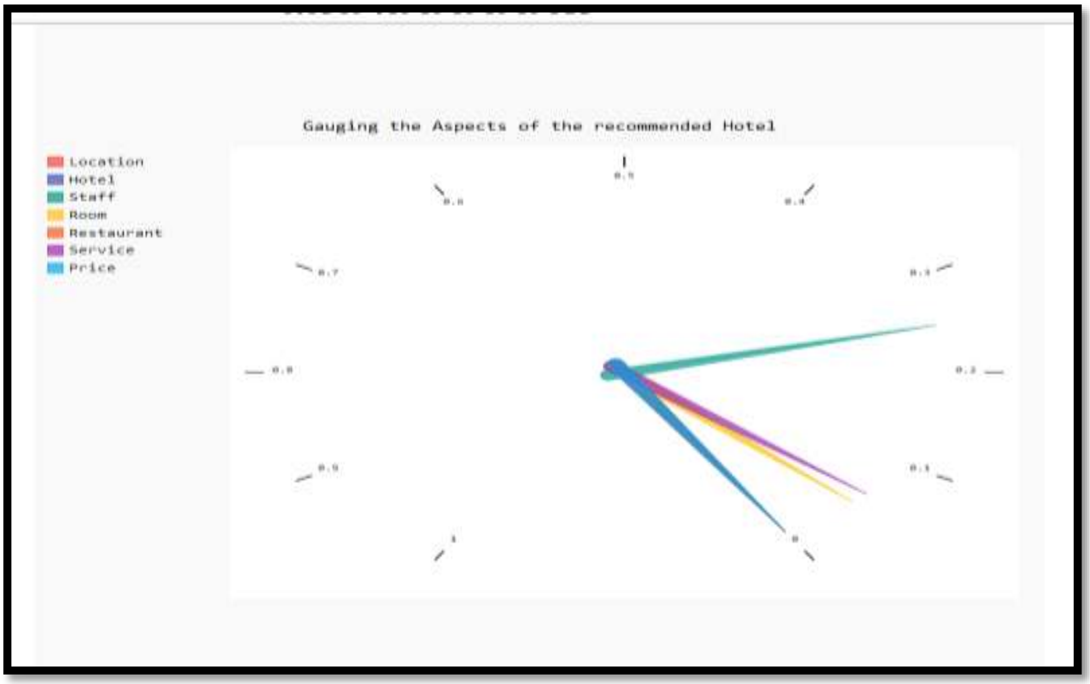


Figure 8-Gauge representation for scores of aspects for a specific hotel

3. The user can choose the map button to find the hotel on the geolocation map.

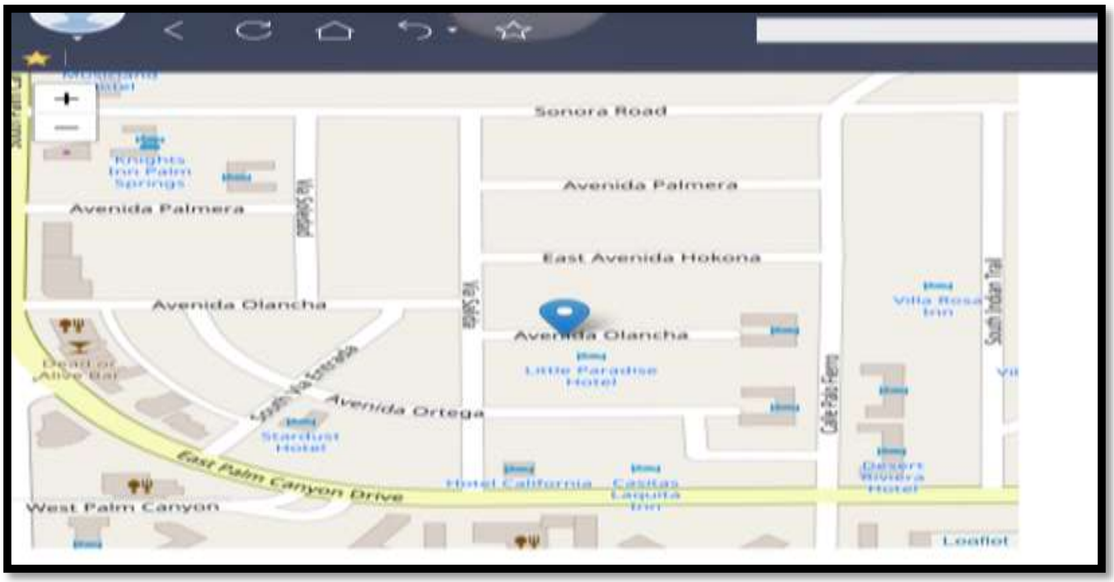


Figure 9-the location of the recommended hotel in the geo-web sitemap

7. Results

The location sentiment recommender system is depending on sentiment analysis operation, which means the accuracy of sentiment analyzer should show good accuracy. The accuracy of any sentiment analyzer have to measured using confusion matrix and then calculate three measurements that are

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{---------- (1)}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{-------------------- (2)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{-------------------- (3)}$$

Where TP is true positive that is the number of right positive polarity for reviews that are predicted by the sentiment analyzer, while TN or true negative is the number of right negative polarities that were predicted by the sentiment analyzer. FP represent false positive that is the number of wrong

positive predictions that are made by the machine learning system and FN for false negative that is the number of wrong negative cases predicted by the sentiment analyzer system.
The accuracy of the supervised sentiment analyzer which the first part of the sentiment system is apparent in the confusion matrix in Figure-10:

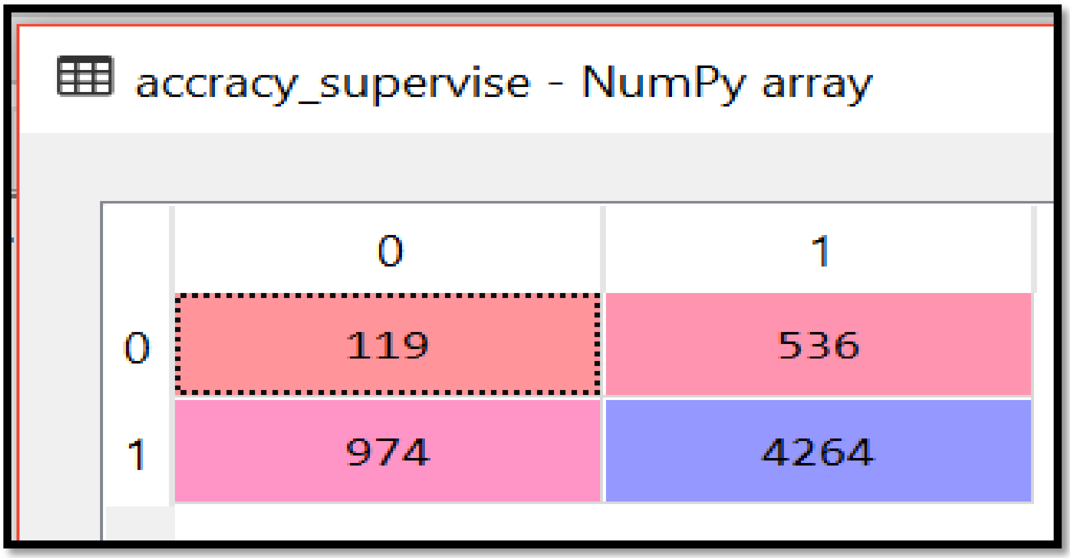


Figure 10-supervised sentiment analyzer confusion matrix

In the above confusion matrix (0) represent negative reviews and (1) describe the positive ones. From the above confusion matrix, three measures are computed, accuracy, precision, and recall presented in the Table-1 below:

Table 1-the accuracy table for supervised sentiment analyzer

Accuracy	0.74
Precision	0.81
Recall	0.88

Moreover, the accuracy measure of lexicon-based sentiment analyzer part is evident in the matching matrix below which is calculated by finding the sentiments of reviews and match this results with the labeled sentiment in the dataset. Read the Figure-11.

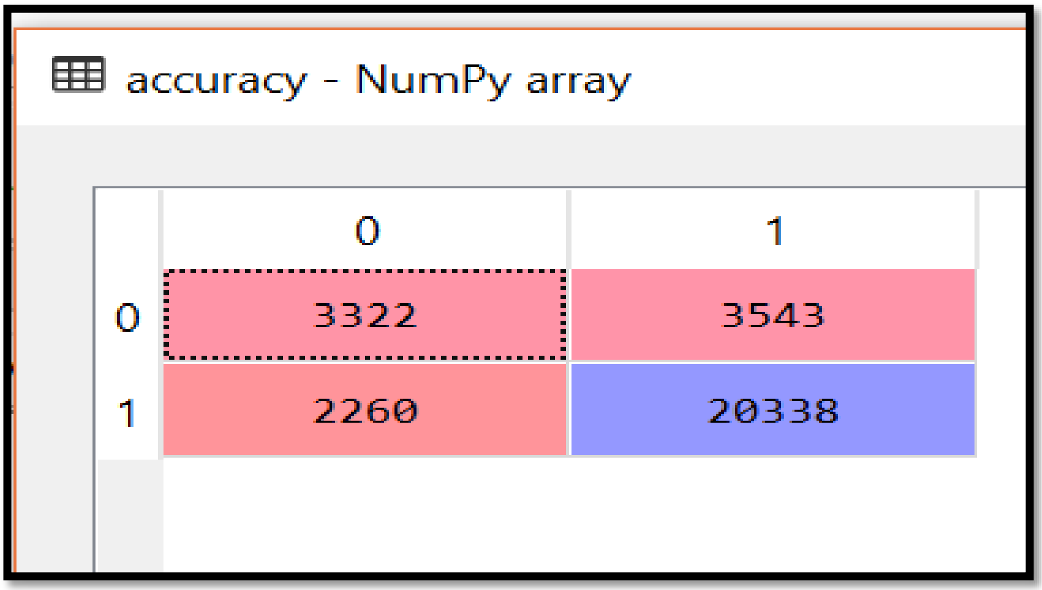


Figure 11-matching matrix for lexicon based sentiment analyzer

From the matching matrix, we derive accuracy, recall, and precision that are listed in the Table-2 below:

Table 2-the accuracy measures of lexicon-based part of SA

Accuracy	0.80
Precision	0.85
Recall	0.89

The aspect extraction part is based on the latent semantic analysis which result is in the Figure-12 below. The reader can notice the most important aspects that the user has focused on, and the interference between terms in each topic. Latent semantic analysis results are showing that most reviewers have focused on (cleanness, staff, room, restaurant, and service), and they used terms like (friendly, helpful, comfortable). For example term (clean) appear in most of the extracted topics. Aspects that spread over all reviews represent each topic. The spreading of aspect over most reviews is an excellent indication to focus on obtaining this aspect from all reviews, and also indicates that all users intersect in the same aspect when choosing a hotel, this also is clear in the graph in Figure-12 below:

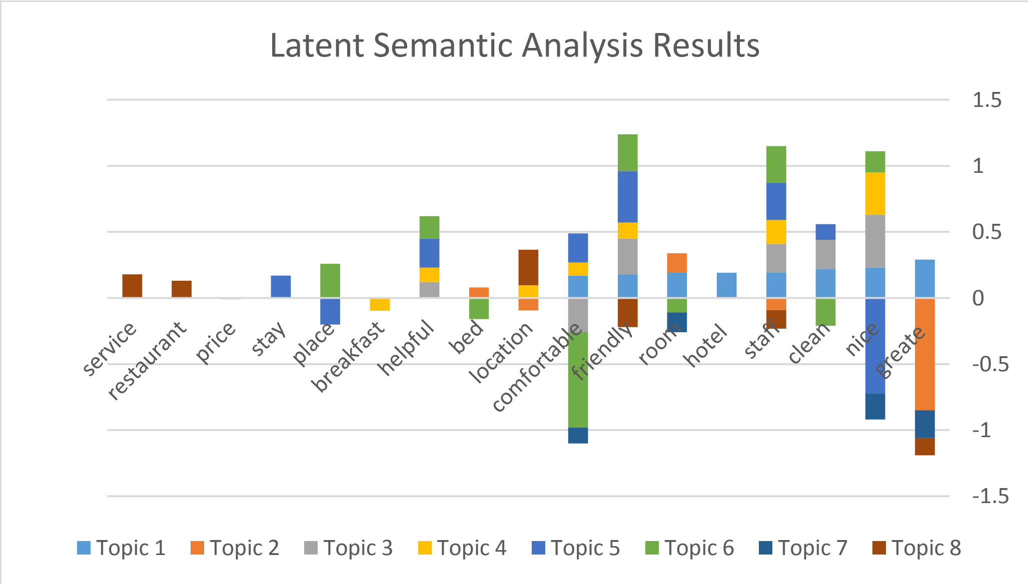


Figure 12-latent semantic analysis topics aspects and essential idioms

In Figure-12 the sign (positive or negative) for term inside the topic mean that the similar correlated idioms and terms in the topic will have the same sign or direction.

Finally, the proposed system here does not need to measure the accuracy of the recommendation process, because it does not work as collaborative filtering or on machine learning model. As a compression with some previous SA, SVM has been taken to compare with the hybrid sentiment analyzer; the accuracy reach to 82.2, this shows that hybrid SA is better as it manages negation issue and uses supervised naïve bays SA empowered with lexicon SA.

8. Conclusion

The location opinion-based RS plays a crucial role because it can understand user feelings about a service or a product. The first contribution is that this system also exploits current user location to suggest a nearby venue to him that got a good number of positive opinions. This paper proposed a hybrid SA that empower the supervised SA by a lexicon SA to have a wide range of vocabulary, which was the second contribution. The supervised part has accuracy about 0.74 while it has precision and recall, and that indicates that the system is far from overfitting. The lexicon SA part has an accuracy of 0.80 because it depends on dominant WORDNET lexicon with negation handler. This work also proposed an efficient chunking parser that extracts aspects depending on noun phrases patterns and probabilistic aspect classifier that relies on LSA reviews dataset analysis. Lastly, this type of RS is participating in many social applications like crime map informing systems, traffic congestion alert system; the suggested future work is trying to implement aspect sentiment analysis in the Arabic language.

References

1. Irfan, R., Khalid, O., Usman, M. and Khan, S. **2015**. Mobicontext: A context-aware cloud-based venue recommendation framework. *IEEE transactions on cloud computing*, **PP**(99).
2. Akhtar N., Zubair N., Kumar A. and Ahmad T. **2017.** Aspect-based sentiment oriented summarization of hotel review. 7th international conference on an advance in computing and communications, pp: 563-571.
3. Zhang, Z. and Morimoto, Y. **2017**. Collaborative hotel recommendation based on topic and sentiment of review comments. In proceeding of the 9th forum for information and engineering, DEIM.
4. Geetha, M., Singha, P. and Sinha, S. **2017**. The relationship between customer sentiment and online customer ratings for hotels. *An empirical analysis tourism management journal*. **61**: 43-54.
5. Nikhat, N. **2016**. *Sentiment analysis of Twitter data for a tourism recommender system in Bangladesh.* Master thesis. Aalto University.
6. Ptil, A. and Kin, S. **2016**. Location-based sentiment analysis of products or events over social media. *International Journal on recent and innovation trends in computing and communication*. **4**(7): 50-55.
7. Mohit, P., Nair, A., Shaik .N. and Kjedkar. S. **2017**.Geo-notification location based service using sentiment analysis. *Internal research journal of engineering and technology*. : 1501-1503.
8. Walla, M., Hassan A. and Korachy H. **2014**. Sentiment analysis algorithms and applications: A survey. *Ain-shams engineering journal*. DOI: https://doi.org/10.1016/j.asej.2014.04.011
9. Aggarwal, C. **2018**. *Machine learning for text*. Springer. EBook. DOI: https://doi.org/10.1007/978-3-319-73531-3
10. Karim, A. and Abdulwahhab, A. **2017**. Reviews sentiment analysis for collaborative recommender system. In Proceedings of international conference on engineering and computing, DOI https://doi.org/10.24017/science.2017.3.22
11. Liu, Bing . **2012**. S*entiment analysis and opinion mining.* Morgan and clayfoal publishers, E-book.
12. Sarkar, D. **2016**. *Text Analytics with python.* Apress. http://doi.org/10.1007/978-1-4842-2388-8.