Text Analysis of Hotel Reviews

Saurabh Nandy & Neha Singh

Oklahoma State University

Abstract

Today of all the economic sectors, there's been a significant shift in its contribution to the country GDP and that sector is Tourism which has risen from the privilege of few to a global socio economic activity. With its rise, Hotel industry is also flourishing by leaps and bounds and so are the standards of quality services by the tourists. In today's cyber age where everyone on the internet by default becomes part of the online community, word of mouth has taken bigger realms that has only steadily increased over time. Many tourists look for online hotel reviews in terms of its hospitality factors which indicates the hotel's performance. According to a recent survey, around 46% of travellers look for online reviews before traveling but what is more compelling is that only 3% of the travellers said that online reviews do no hold any significance while booking a hotel. We are aiming at text analysis on user reviews to get the interests of the customers which will help the hoteliers identify the interests and the hospitality expectations from a hotel .For this paper, we are looking at Trip-advisor's dataset of 4096 different hotel reviews with a total of 60,239 user reviews and performance data of different locations in United States.

Methodology

- Collecting and analyzing customer feedback is important because it allows organizations to learn in a continuous manner, to adapt their offerings to customer preferences (Sun and Li 2011).
- We formulated below diagram to accomplish text analysis using SAS Enterprise Miner.
- We used various text mining nodes such as text parsing ,text filter and text cluster to achieve the sense of the clustering present in the reviews and concepts linked together.
- It constructs a Singular Value Decomposition (SVD) that essentially maps all terms into a huge multidimensional space. The closer together two terms are in this space, the more similar they are.
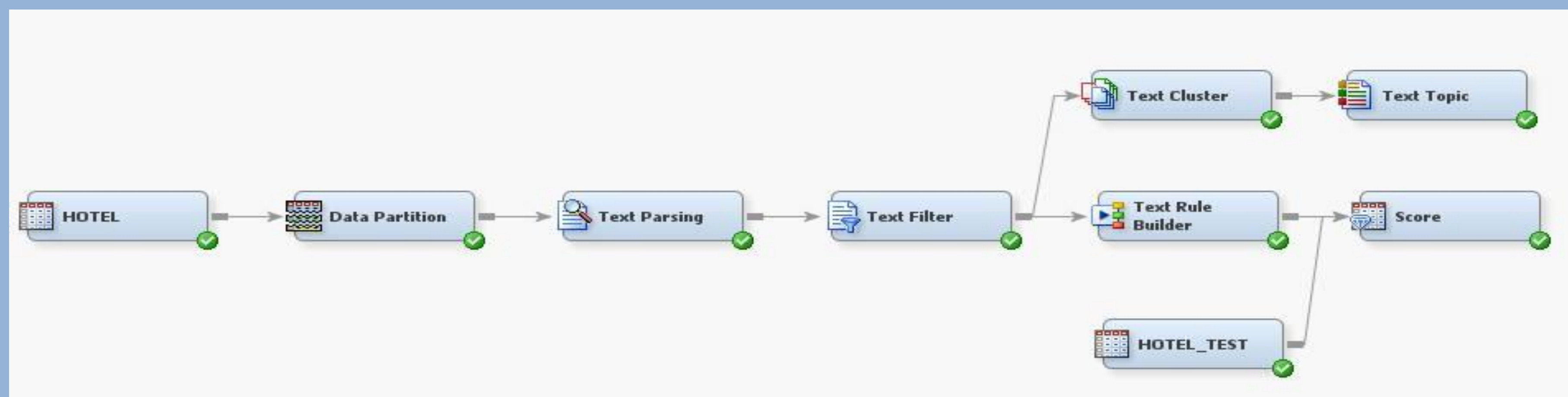
Approach

- The raw dataset which we had was in a .json format which we had to convert into a .csv file using Python. After converting the data into .csv file, the file had a lot of null and reviews in other language.
- For our research, we have considered the reviews written in English. Many of the improper reviews were removed for our analysis.
- On the right is the process flow of this paper.
- We parsed the data using the Text parsing node and then interactively filtered the text using the Text Filter node. We grouped all the similar words as synonyms and also dropped the unwanted terms from the analysis. The synonyms have been saved and exported for future use.
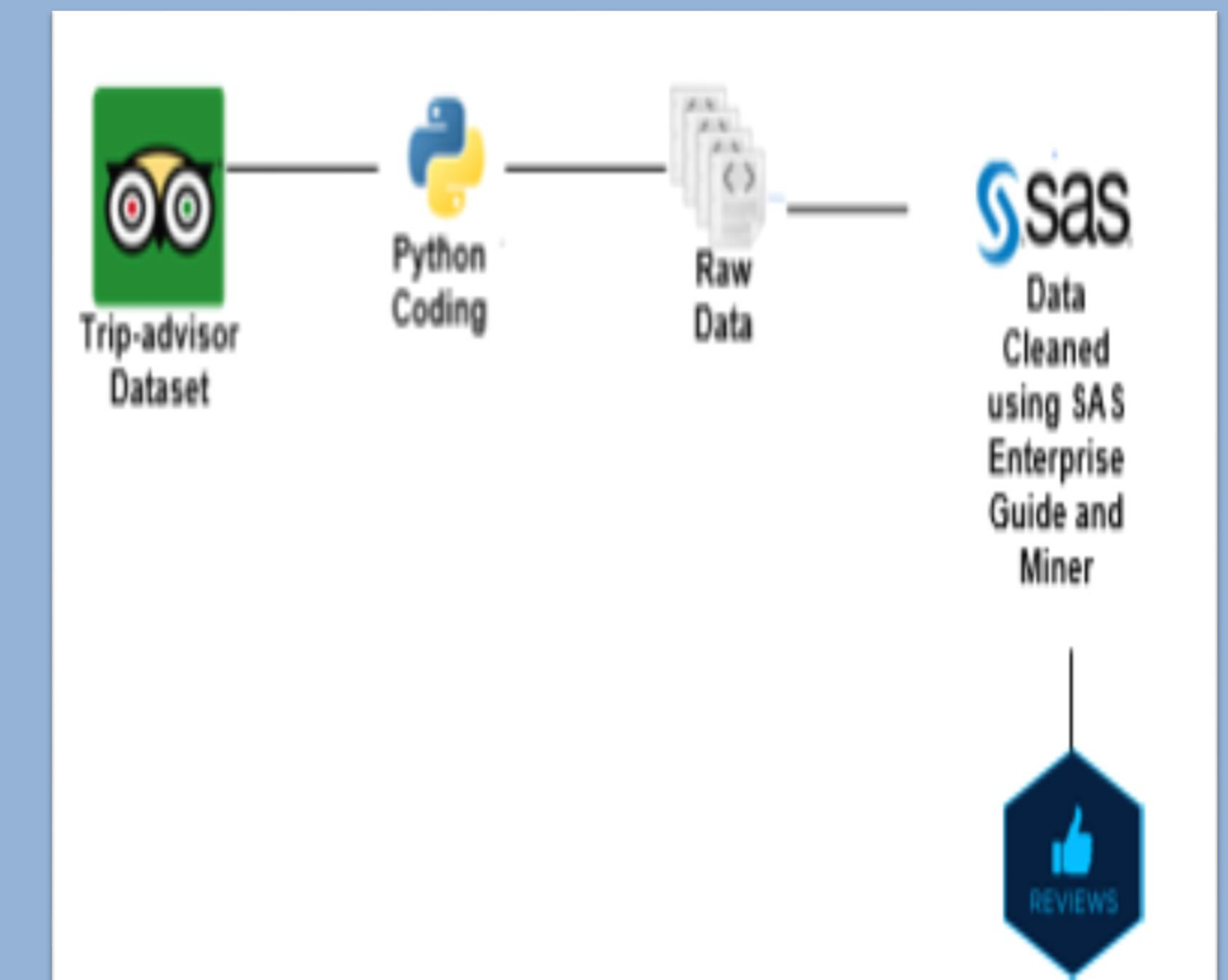
Results

- Table on the right exhibit the descriptive terms for clusters identified obtained with default setting.

| Cluster ID	Prominent factor	Descriptive Terms
1	Medical Facilities	girl kind medical +center +'medical center' +bring +choose
2	Restaurant facilities	hotel plaza +restaurant four +night +weekend +old +center +book +spend
3	Health and Fitness	gym rooms +week +'park lot' stylish + business +facility +kind +hear +place +review +park
4	Check-in/ Check-out/ Pickup Services	'front desk' +shuttle desk +flight front +airport +month +look setai +expect +arrive +late +member +price +place
5	Internet/Wi-fi	internet cheap deal
6	Location	plaza +top hotel great +hotel +stay
7	Offers and Discounts	birthday +friend few +shop +offer +recommend excellent +return +bed distance +check eventi +room +suite helpful
8	Landscape	park +lounge 'great hotel' +great +back +pretty outdoor pool +easy +food distance hilton +large +negative +good
9	Beautification	rate walk chelsea efficient +renovation +book +good brand +shop +know +offer +mile +start
10	Member facilities	boutique counter luxury pleasant +brand home +cheap +member +hotel +expect +door +renovation
11	Complimentary/ Room	room comfortable breakfast first trip +nice great +clean +small +great +large beautiful wonderful
12	Number of visits/ Time of visit	three +time +mile april san complimentary +week carlos +month +stay staff business york +pretty

Text Analysis of Hotel Reviews

Saurabh Nandy & Neha Singh

Oklahoma State University

Result (Cont)

- Figure on the left shows Hierarchal clustering results using low SVD. It provides an optic representation of the cluster distribution in the document.
- Cluster 11 , complimentary/Room has got one of the biggest share in the pie , representing the occurrence and significance of this cluster in customer reviews.
- Similarly cluster 4 and 7, check-in/check-out/Pickup services and Offers & Discounts are next in line to hold the maximum number of occurrence in customer reviews and also inferencing that they can be the deal breakers or makers in booking a hotel.
- For a more holistic approach to understand the underlying semantic structure in terms of the relationship among concepts, we used concept link graphs.
- The first vector based representation shows the prominent terms outlined in the customer reviews and their relation with the Hotel.
- Length and thickness of the line pertains to the closeness and weight of the term with the center term. Here ,Stay and Looks are the most influencing factors for a customer to write about.
- The second vector based representation shows the second layer of the reviews. Here room is the center term and the concept link terms are surrounding it. The weighted link represents the strength of the relationship. Bed has the shortest and thickest link, indicating that quality of bed is governing customers' review irrespective of it being negative or positive.
- Other major governing factors are bathroom, large and clean.

Prominent Text Clusters
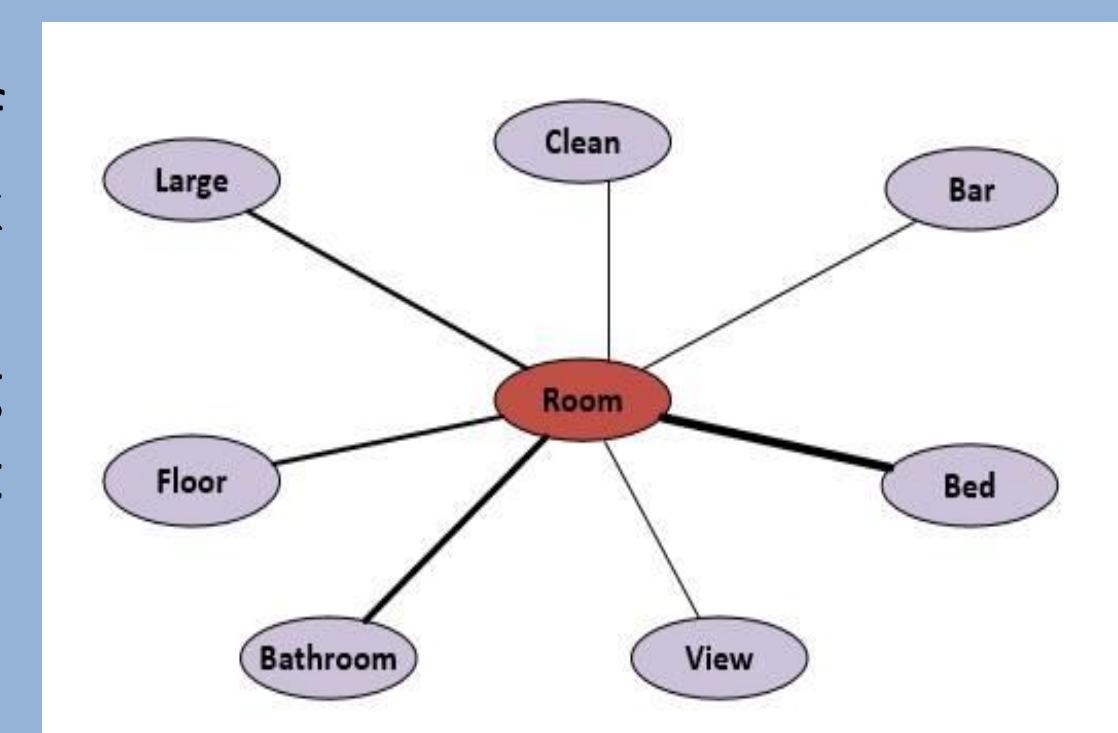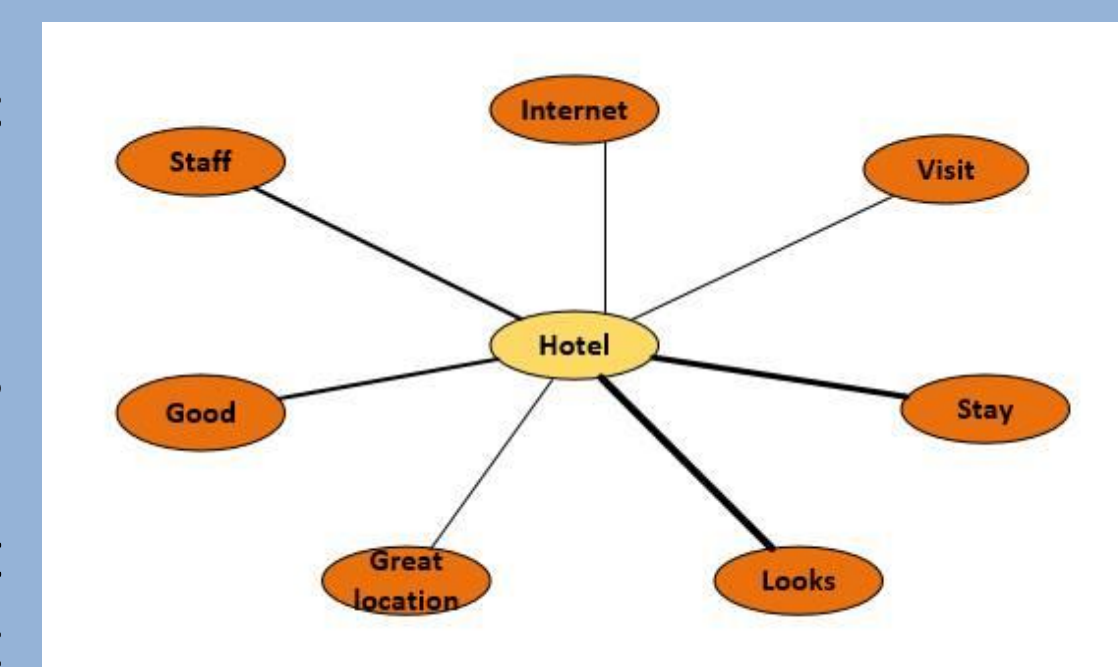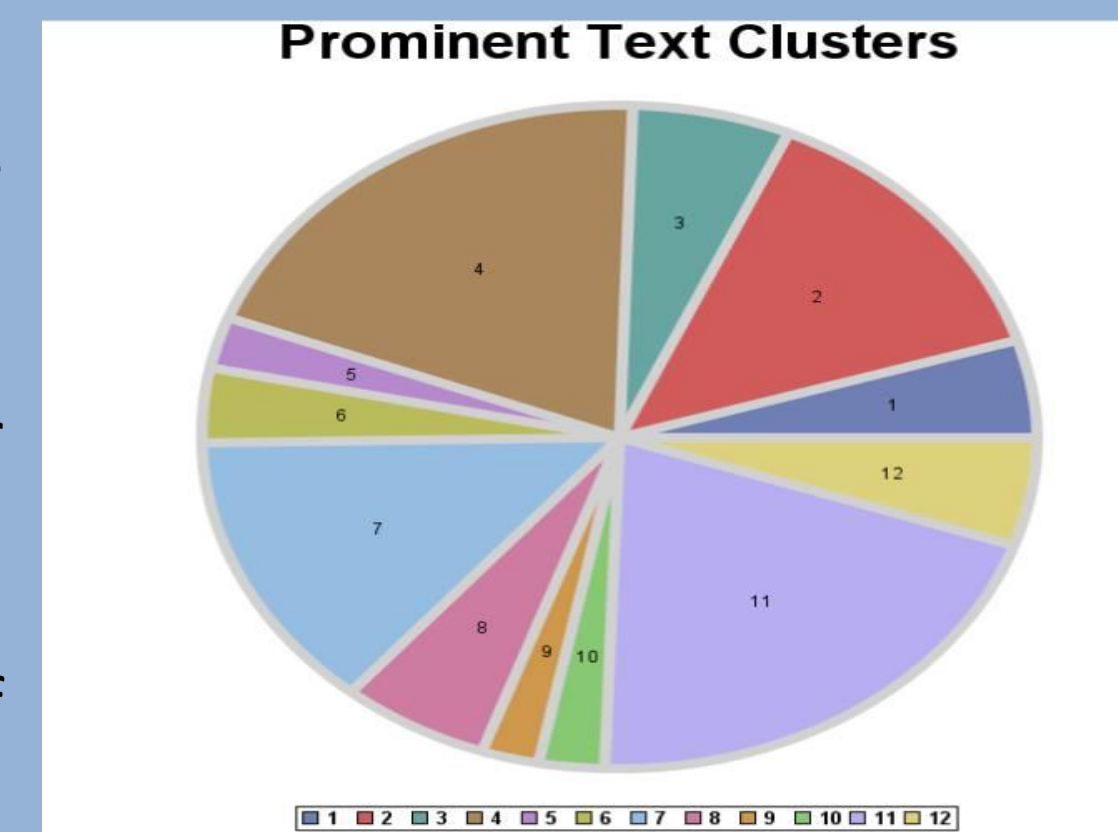
Conclusion

- While selecting a hotel either by a frequent traveler or a non frequent traveler they look up the internet for review while finding positive or negative things about a hotel helps them decide whether to go or not. Usually every hotel has some good and bad points but its mostly the good that matters.

Future Scope

- Creation of an online survey targeting the interests of the customers of those different clusters which will help the hoteliers identify the interests and the hospitality expectations from a hotel.

Limitations

- This research has several limitations which provides scope for further research and exploration.
- First would be the reduction of data. Many reviews were neglected because of presence of foreign language. Additional linguistic research is recommended.
- Second would be domain expertise.

References

1] Rowena Chau, Ah Chung Tsoi, Markus Hagenbuchner, Vincent C.S. Lee, A ConceptLink Graph for Text Structure Mining .

[2] Thorsten Gruber , Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach

[3] Dr. Goutam Chakraborty, Murali Pagolu, and Satish Garla (2013), Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS