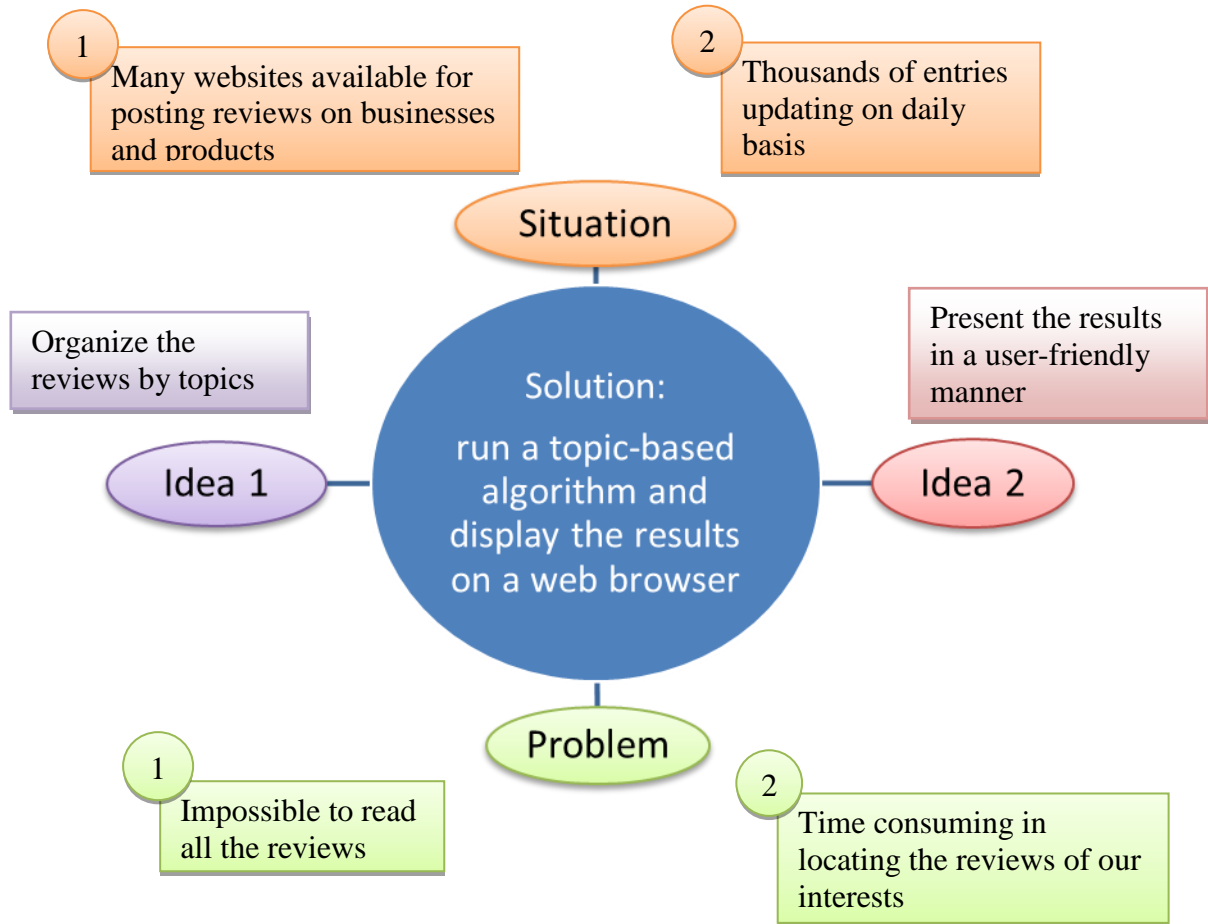


Topic-Based Browsing of Online Reviews of Businesses and Products (II)

By Ka Wai CHAN, Tsz Him KWOK

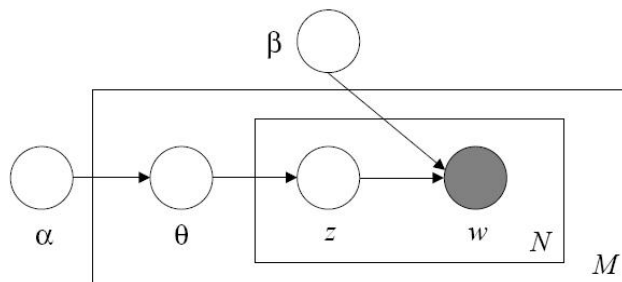
Supervised by Prof. Nevin L. Zhang

Introduction



Latent Dirichlet Allocation (LDA)

- Type of topic-based modeling algorithm
- For determining the abstract “topics” in a document
- Three level hierarchical Bayesian model:



D – Corpus (Training Data)

M – Number of Documents

N_i – Number of Words in Document “i”

α –Hyperparameter of Dirichlet distribution

β –Parameters of the multinomial distribution ($K \times V$)

θ –Parameters of the multinomial distribution for topics for each document ($M \times K$)

z –Indicator variable, topic sampled

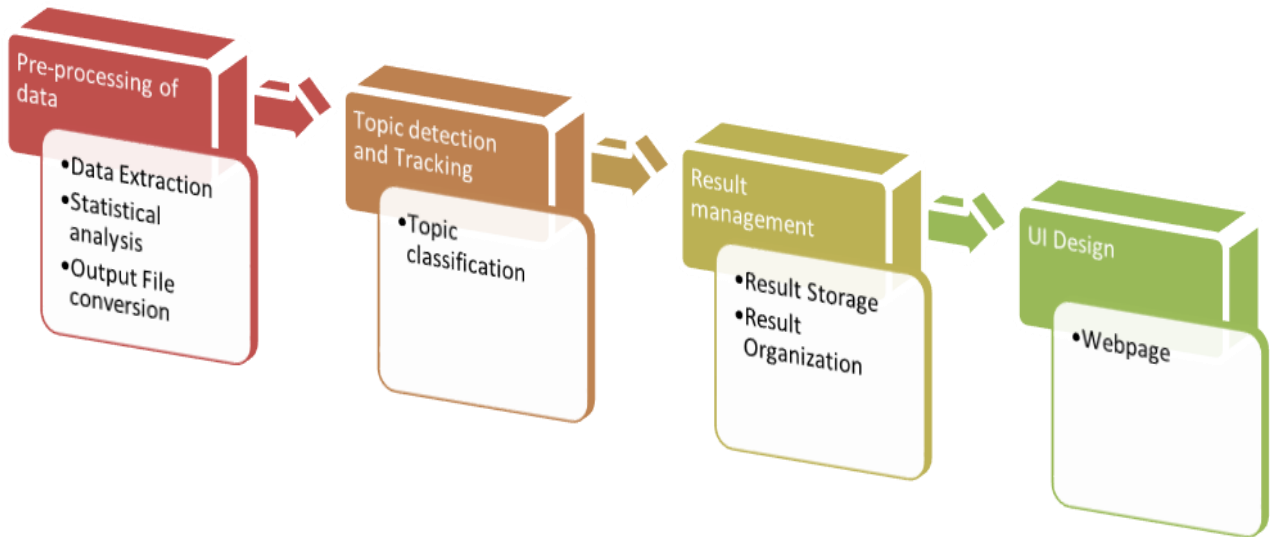
w –Indicator variable, word sampled

- Number of topics (k) is decided by users

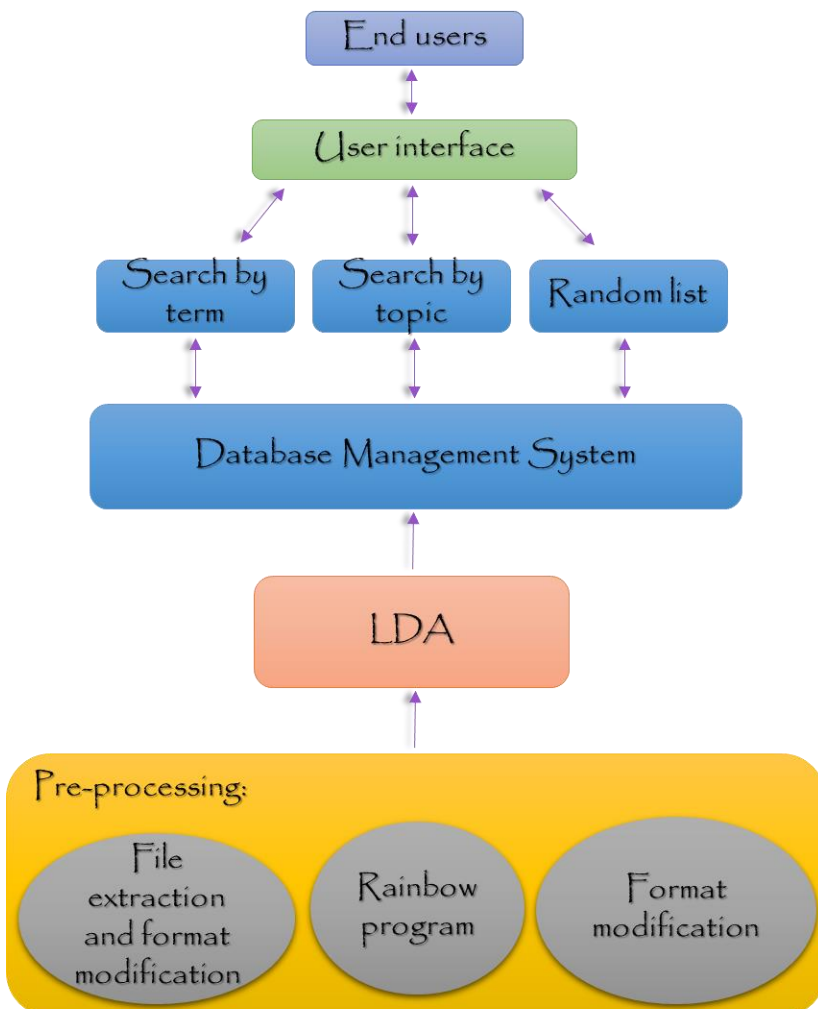
Objectives

1. Convert dataset of reviews into a file whose format is compatible with a topic detection program.
2. Use a topic detection program to assign themes to the reviews to facilitate searches by topic.
3. Provide a user-friendly graphical user interface (GUI) to display a list of topics and links to related reviews.

Design



Implementation



User interface design:

- User-friendly
- Simple
- Reusable

Search methods:

- By selecting a term
- By search engine
- By random list of topics

DBMS:

- Store the results in terms of entities and attributes

LDA:

- Determine the number of topics and other random variables for topic detection
- Execute the program and get the results

Pre-processing:

- Set of reviews stored in JSON was extracted
- Perform normalization and statistical analysis to remove common words and do the word count
- Format conversion to [M] [term_1]: [count] [term_2]: [count] ... [term_N]: [count]

Result

Users select the topic {friendly, staff, good}

sandwich	
made	{time, pretty, good}
side	{friendly, staff, good}
make	{happy, good, back}
also	

Inside the topic, users can find the list of reviews related to that topic

{friendly, staff, good}		
Terms	List-Of-Reviews	Related-Topic
	List of Business Reviews	
	Spinato's Pizzeria -review- 0	
	Chaparral Dog Park -review- 2	

After selecting a review, users can know about the related topics and reviews for that review and the business information about the review

Spinato's Pizzeria -review- 0				
Review Content	Related-Reviews	Related-Topics	Review-Topic-Distribution-Pie-Chart	Business Information
I have no idea why some people give bad reviews about this place. It goes to show you, you can please everyone. They are probably griping about something that their own fault...there are many people like that.				
In any case, my friend and I arrived at about 5:50 PM this past Sunday. It was pretty crowded, more than I thought for a Sunday evening and thought we would have to wait forever to get a seat but they said we'll be seated when the girl comes back from seating someone else. We were seated at 5:52 and the waiter came and got our drink orders. Everyone was very pleasant from the host that seated us to the waiter to the server. The prices were very good as well. We placed our orders once we decided what we wanted at 6:02. We shared the baked spaghetti calzone and the small "Here's The Beef" pizza so we can both try them. The calzone was huge and we got the smallest one (personal) and got the small 11" pizza. Both were awesome! My friend liked the pizza better and I liked the calzone better. The calzone does have a sweetish sauce but that's how I like my sauce!				

Evaluation

Survey result done by volunatry users



Conclusion

The goal of this project was to ensure the accuracy of the information provided including the review content and the statistical result of the algorithm and the design of the user interface layout is simple and user-friendly to all users. We have successfully implemented a system for users to locate their interested topics of reviews in an effective and efficient way by adjusting the performance of the algorithm and doing research on the user interface layout design