Interactive Online Platform For Discussing Yelp Reviews Grouped By Topics Using Topic-Modeling

Group Member: Chan Ki Yung, Lam Yuen Ting, Li Chun Ho

Supervisor: Prof. Nevin L. Zhang

Introduction

Nowadays, there are many reviews website providing useful information for life style. Our topic is on a famous reviews website Yelp and uses its dataset to demonstrate. However, readers do not have time to look all the reviews. So, now, we are providing a **SOLUTION** aiming to achieve the followings:

1. Automatically detect the key topics in the reviews and use those topics to organize them
2. Get a clear overall picture and to quickly locate the reviews of the greatest interest
3. Let people to comment and interact with the reviews

To achieve the above, We USE:

1. LDA algorithm as a base to automatically generate topics
2. Design how to organize the data of 80000 reviews to database to have quickest retrieval time
3. Design the criteria to retrieval the most relevant result to users

Fig.1 A yelp overview article

QR code for our product

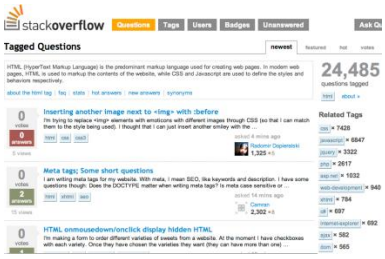
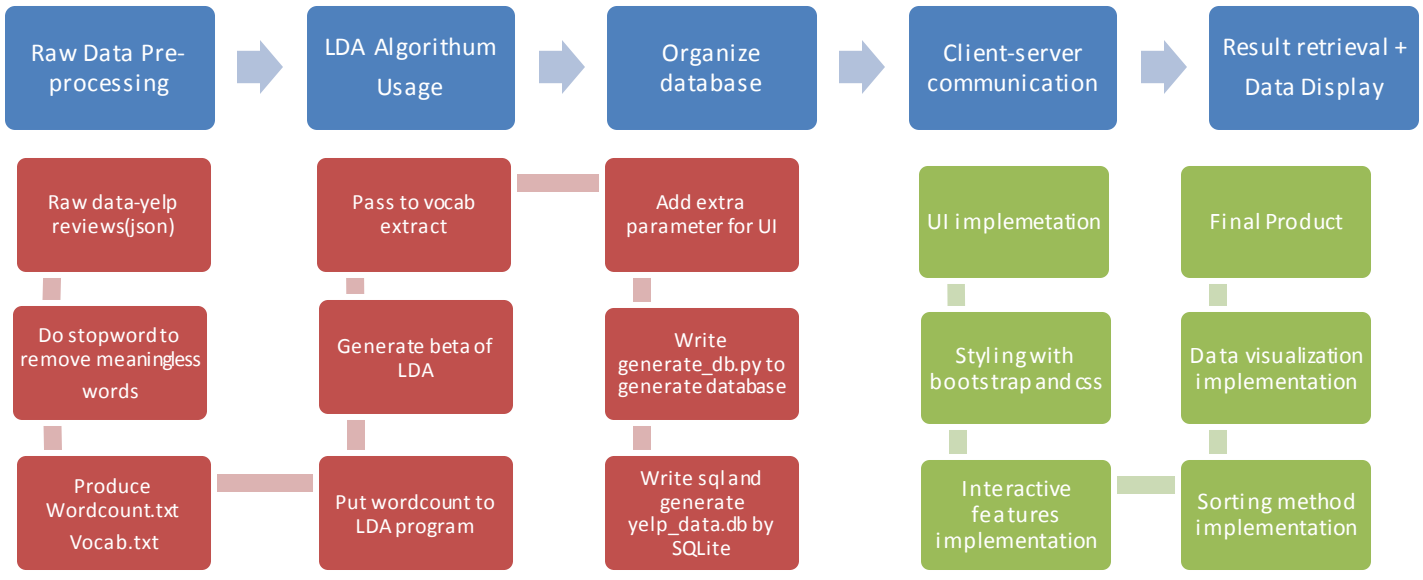


Fig2. Other interactive platform(stackoverflow)

Implementation



Database/Server side implementation (core)

UI and information retrieval implemtation(Display)

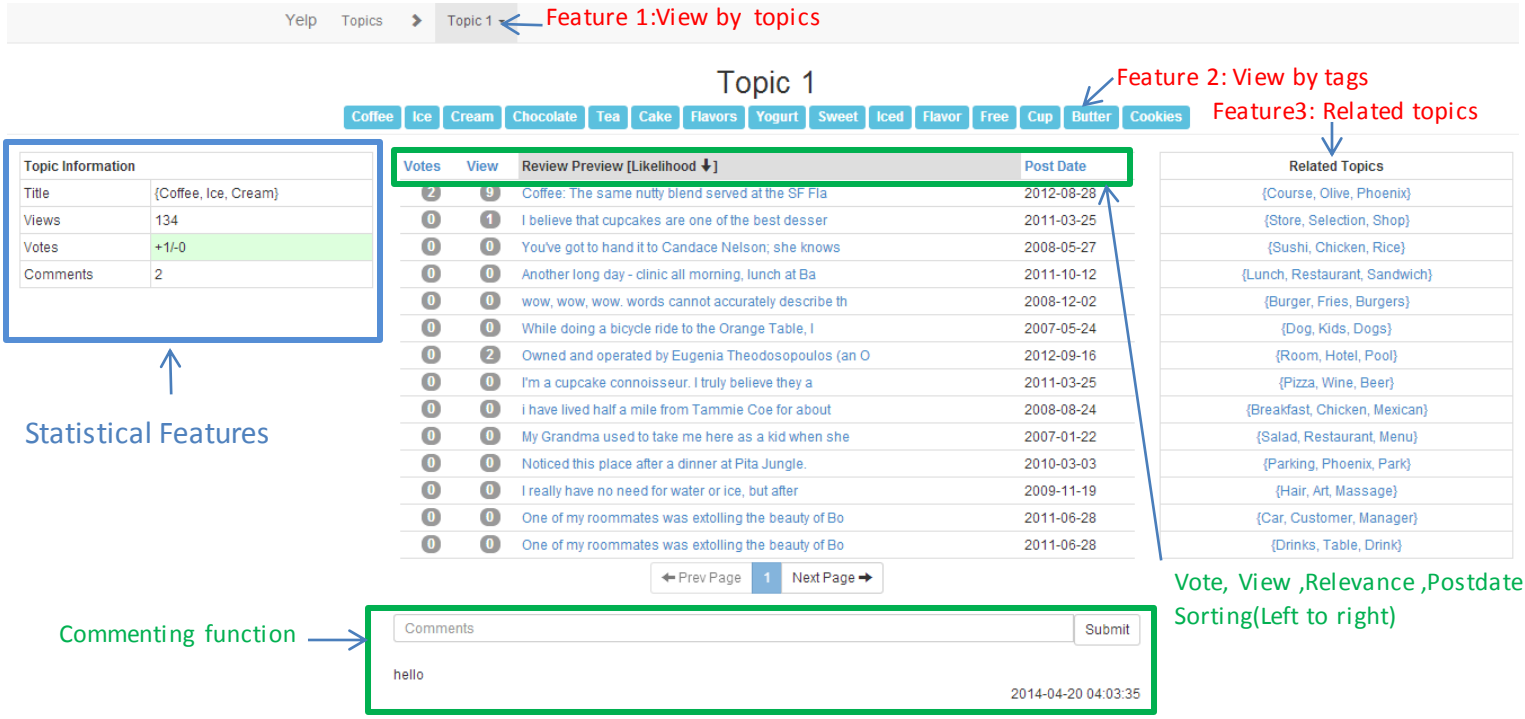


Fig3. Overview of UI in Topic 1

Features:

Our goal:

1. Grab the first sight idea of the topic
2. Get the statistic of overview of the dataset
3. To let views to interact with reviews

Result retrieval Features: (in red)

-"Topic" Features (Deep blue box)
15 large categories of topics to let viewers grab big picture

-"Tag" Feature

-"Relative topic" Feature

Interactive Features: (in green)

1) Vote: +/- vote
2) Number of views
3) Comments (comment on reviews)
4) Sort by postdate/relevance

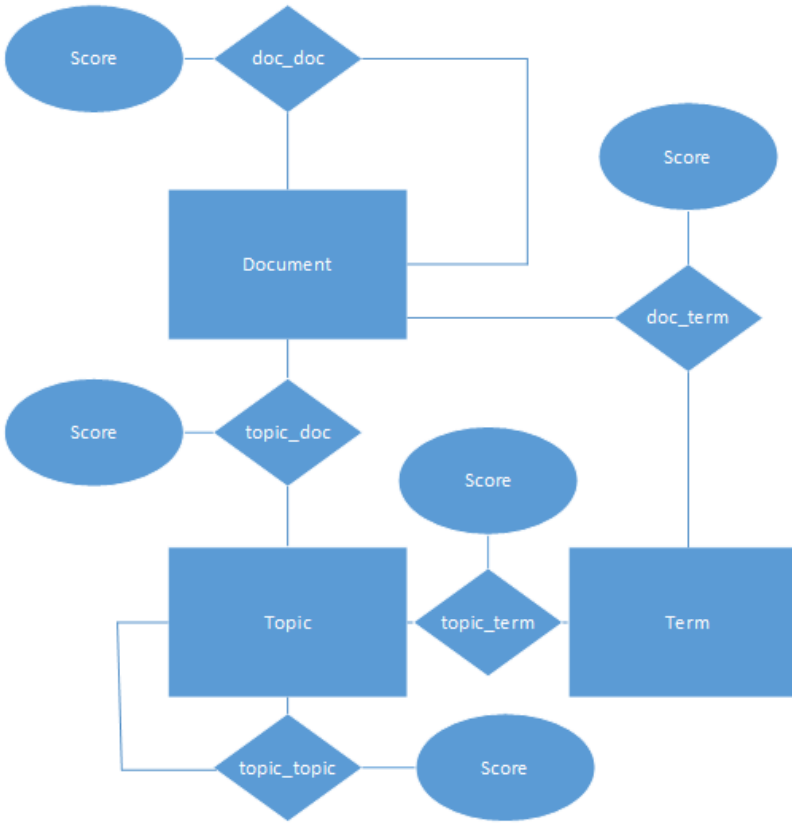
Statistics features:

1) Overall views
2) Number of article in topic/tag
3) Total +/- vote
4) Number of comments(hot topic)
5) Postdate extraction

Database Design:

One of the notable part of this project is on how to organize the database to give a quick retrieval.

Database Schema:



Evaluation

Testing Results:

Client side testing:

Item	Passing Criteria	Result	Conclusion
Data Retrieval -Correctness	100%	100%	Pass
Capacity of database	>40000 documents	80000	Pass
Server Client interaction -hyperlinks	100% to desired pages	100%	Pass

Server Side Testing:

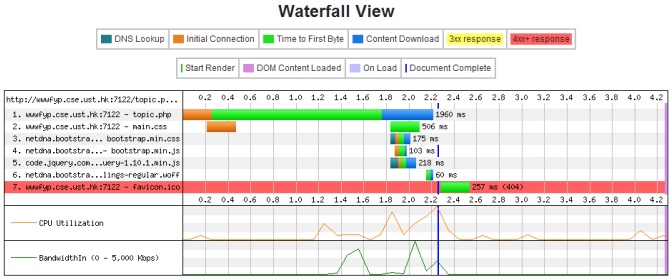
Item	Passing Criteria	Result	Conclusion
Pre-processing program	100%match	100%	Pass
Data loss betw een conversion	0%	0%	Pass
Information consistency	100%	100%	Pass

Speed Test (topic page generation):

From Dulles,VA,US under IE8
by webpagetest.org

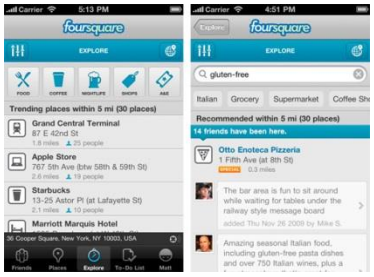
Load Time	First Byte	Start Render	Visually Complete	Speed Index	DOM Elements	Result (error code)	Time	Requests	Bytes In	Time	Requests	Bytes In
2.255s	1.756s	2.257s	2.400s	2301	184	99999	2.255s	7	184 KB	2.536s	8	184 KB

domContentLoaded	loadEvent
4.268s - 4.279s (0.011s)	4.280s - 4.283s (0.003s)



What's Next?

Search engine technique?



Social network?

forum?

Result and Future Work

Future work:

1. Continue with the project and build a community to rate on reviews
2. Use LDA to analyze the correlation between positive/negative wordings and the stars/vote of an article

Result:

The result of this project is satisfactory. The success of this project gives an alternative way for information retrieval other than search engine. The application of LDA to information retrieval site can greatly reduce the documents which are irrelevant to the topics and easier for user to search the article desired. This may be the solution to information retrieval in the era of information explosion.