Snippet Visualization System

Kishori K. Kamble[1], Kavita S. Koli[2], Prajakta S. Chothe[3], Prajakta P. Angre[4] , Mr.Vishal. V. Mangave[5]
[1, 2, 3, 4, 5] D.Y.Patil College of Engineering and Technology,Kolhapur.

Abstract- *Normally people search query to get desired results as they want. Most users relay on what they get results from internet. So that Results need to be highly ranked as user get most appropriate and accurate outcomes. Each search results show URL, text, and description about query searched. Search engine show a content based on how many times user accessed this particular sites counts, to get search relevant to user query. Sometimes it may happens that user will not get result for particular web page and also not as expected. To overcome this limitation of the search engine we need to visualize the search results. We propose a system that will help to get results according to their needs. By showing how contents are related to each other by forming a cluster of similar words. We also try to rank search results according to user visits particular page.*

The system display results in snippet format, to avoid overlapping within this snippet we optimize the placement of snippet .The system help user to get results in visualization approach ,placing snippet in colors formats.

Keywords- Google, Snippet, Web Search Visualization, Text mining, cluster , optimization.

I. INTRODUCTION

Many of the internet users search for some query and In result they get ranked list of URL. User could possibly get some disappointment with this related search . Indeed it has limitations in particular situations as it fails to provide an overview of the document collection retrieved. The challenge lies in ordering retrieved pages and presenting them to users in line with their interests.

We are proposing a system is focusing on retrieving search results in thumbnail form. So user can have interaction with more results in single view. This display has many advantages e.g.it affords easy navigation and is straightforward to interpret with user. It gives listing results according to user needs. Therefore it is a real time necessity. It utilizes page rank algorithms to analyze and re-rank search results according to the relevance of the user's query by estimating the importance of a web page.

Based on the nature of the query it may be too general, ambiguous, results may be poorly ranked or it may time consuming . Several search tasks would be easier if users were shown an overview of the returned documents, organized so as to reflect how related they are content-wise.

An energy function is required which considers both the overlapping between snippets and the neighborhood structure provided by a multidimensional projection. Multidimensional projection techniques may be employed to generate visualizations that gives the perception of groups of similar documents.

II. RELATED WORK

Visualization techniques to support textual searching can be split into following major tasks:

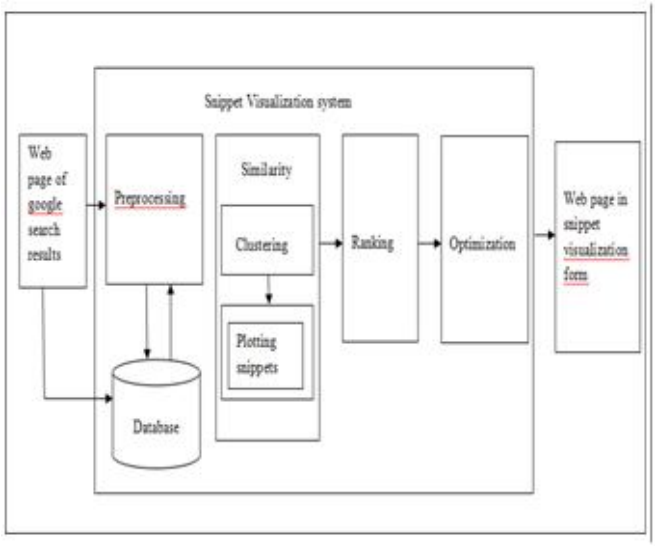
Arrangement by similarity. A major requirement is to easily identify documents with similar content. A straight-forward way to accomplish this is to build layouts where similar documents recovered by a search engine are placed close to each other. Such a layout may be naturally obtained with k-means clustering.

Ranking identification. The success and effectiveness of search engines rely on a ranking mechanism that sorts documents according to their relevance. Therefore, any search result visualization technique must convey the document ranks.

Optimization. It avoids overlapping within placement of snippet.

For getting google search results we need to add google api key in javascript it enable to access google results[3] It gives search results for a specific query.Still there are mix of keywords in search results.It need to remove all stopwords and all unnecessary symbols.After preprocessing of results we need to make cluster of similar words[4]

III. SYSTEM ARCHITECTURE



In this system we retrieve Google search result as a input. Then preprocessing on web search results and collected results are stored in database. After that the system finds similarity by using clustering technique which collects similar content and group them into single cluster. After applying clustering technique snippets are plotted and ranked according to their relevance.The overlapping within snippet is removed by using energy function and final results are displayed.

IV. METHODOLOGY

1. Collecting query results data from google search engine.
2. Store the data in database.
3. Apply preprocessing on raw data.
4. Identification of similar content
5. Relevant documents are sorted and ranked using
6. Performing Optimization to avoid overlapping in positioning of snippet by using Energy function
7. Display output using visualization technique.

Algorithm K-means++

k-means++ clustering aims to find the set of k clusters such that every data point is assigned to the closest center.

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select ‘c’ cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, ‘ci’ represents the number of data points in ith cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step 3).

Energy Function

The energy functional E comprises two components, one that considers the overlap of snippets, denoted by EO, and a second component related to the neighborhood relations resulting from the projection step, denoted by EN. In mathematical terms, the energy E is written as

$$E=(1-\alpha)EO + \alpha EN$$

where the parameter α €[0,1] balances the relative contributions of both EO and EN in the total energy.

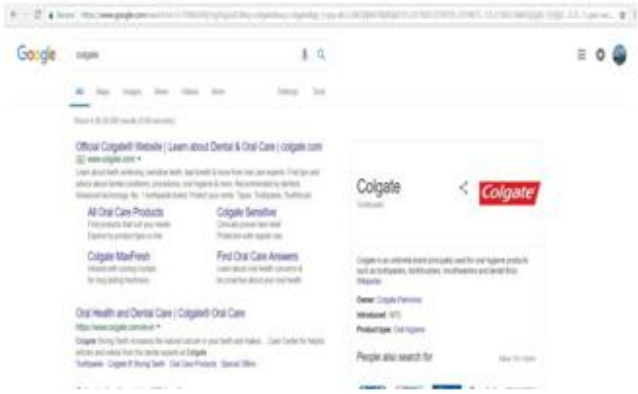


Fig 1-Traditional system’s search results

Fig 2 –Snippet visualization system's results

V. CONCLUSION

We introduced Snippet visualization technique to visualize the collection of textual result returned from web a query. The methods builds meaningful layouts that optimize the placement of snippets that considers both overlapping removal and preservation of neighborhood structures. We showed results illustrating how the snippet visualization layout convey a global view of result query while allowing for identifying similar content through a clustering mechanism.
Since snippet visualization system relays only on information extracted from web result, it can be plugged into search engines in a straightforward manner.

There are number of future research directions to extend and improve this work. One direction that this work might continue on is to improve on the accuracy of similarity calculation between documents by employing different similarity calculation strategies. Although the current scheme proved more accurate than traditional methods, there are still room for improvement.

REFERENCES

[1] Similarity Preserving Snippet-Based Visualization of Web Search Results‖ Erick Gomez-Nieto, Frizzi San Roman, Paulo Pagliosa, Wallace Casaca, Elias S. Helou, Maria Cristina F. de Oliveira, Member, IEEE, and Luis Gustavo Nonato, Member, IEEE.IEEE transactions on visualization and computer graphics, vol. 20, no. 3, MARCH 2014.

[2] Visualization on Web Snippets: A Survey ,Shefali Kedia1, Prof. K. P. Wagh, Dr. P. N. Chatur P.G. StudentInternational Journal of Innovative Research in Science, Engineering and Technology.Volume 6, Special Issue 1, January 2017.

[3] A Heuristic method for overlap removal of 2D data representatives
H.strobelt, M.Spicker, D.Keim and O. Deussen. Computer Graphics Forum, vol. 31, no.3,pp,1135-1144, 2012

[4] Visualizing the web search results with web search visualization using scatter plot
M.Nizamee and M.Shojib,Proc.IEEE Second Symp.Web Soc.,pp.5-10-2010

[5] Web document clustering using proposed similarity measure‖
PH Govardhan,K.P.Wagh,P.N.Chatur,National conference on Emerging trends in Computer technology.,pp.15-18-2014

[6] The visual exploration of web search result using hotmap
O.Hocber and X.D.Yang Proc.10th Int conf .Information Visualization,pp.157- 165,2006