

# Visualization on Web Snippets: A Survey

Shefali Kedia<sup>1</sup>, Prof. K. P. Wagh<sup>2</sup>, Dr. P. N. Chatur<sup>3</sup>

P.G. Student, Department of Computer Science & Engineering, Government College of Engineering, Amravati, India<sup>1</sup>

Assistant Professor, Department of Information Technology, Government College of Engineering, Amravati, India<sup>2</sup>

Head of Department, Department of Computer Science & Engineering, Government College of Engineering,  
Amravati, India<sup>3</sup>

**ABSTRACT:** Web search engines provide a ranked list of snippets on a given web search query. These snippets make navigation easy and straightforward to interpret. But, for certain queries that are general or ambiguous (e.g. apple), these ranked list does not provide the desired solution to the search query. A visualization technique aims to overcome such limitation. A multidimensional projection of snippets followed by clustering and organization can solve these problems.

A two-dimensional layout of the snippets by multi-projection technique uses tf-idf vector to calculate similarity. Clustering on these points then make a global view on the queries.

**KEYWORDS:** snippets, clustering, visualization.

## I. INTRODUCTION

Web search engines play a vital role in today's technological world. Any information desired on a certain thing is preferable on web search engines. The procedure includes the submission of a web query to a search engine and obtaining a ranked list of web search results. These results are very efficient and straightforward for a unambiguous query. The disadvantage it has is the user has to explore to find the desired results through the pages. The variety of results are not shown as they are ranked according to most visited links. For example, if a user types "apple" in google's search engine, the first 8 pages of google web results only has links to apple products like ipads, iphones, etc., apple company but no link to apple fruit. Users then need to navigate through the pages or refine their search by typing "apple fruit".

Visualization[1] makes the navigation for user easy and more friendly. Some existing technologies enhances snippets with color glyphs and tag clouds, adding information about documents. The ranked list is also replaced with thumbnails for better understanding.

Multi-dimensional projection helps in greater visualization by plotting the snippets on a 2-D layout. The points that are close to each other represent similar type of snippets and those far represent different category. After clustering these points, they are replaced with other geometric entities for visualization. These can thus create a certain overlapping thus reducing the convenience of navigation. This problem can be reduced by reducing the overlapping on these objects through certain functions and making it more efficient.

The main techniques of this paper includes a 2-D layout of snippets preserving their similarity that uses only the information of snippets. It then clusters the snippets according to a certain technique and replaces these points by other shapes making it simple and easy to navigate.

## II. RELATED WORK

### 1. TileBars:

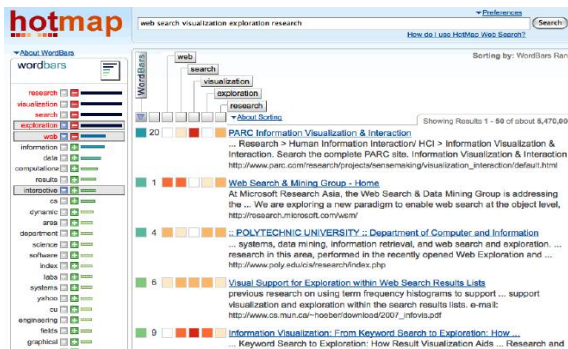
M. A. Hearst[2] first introduced a visualization technique known as TileBars, for a Boolean type query. It provides the information on term distribution of Boolean queries in the snippets by avoiding user to go through the web pages. It indicates the length of document, the term frequency of the query and its distribution. These are displayed on a tilebar column of each search result. It thus makes the judgement of the desired document or web page more easy.

**2. Query Occurrence Visualization:**

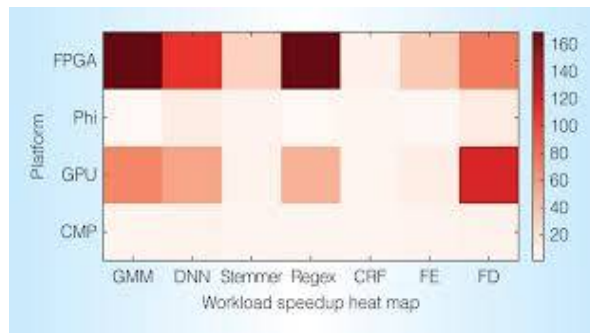
Heimonen and Jhaveri[3] introduced a scheme that represents the query result in a small icon besides the web search result. The scheme though did not reduce any time to complete the task nor made it more efficient, but also there was no change in the number of users searching the web. The ratings of the user stated that it was easy to understand and unobtrusive.

**3. Hotmap:**

Orland Hoerber and Xue Dong Yang[4] aimed to replace the simple list based search results for the users that aims to compare the results or find relevant documents, which list web results fail to provide. They introduced a method called Hotmap that gives a two level detail on the result which helps the user for interactive surfing. It provides the term frequencies of the query terms on a heat scale. Multiple occurrences of a query term represented in a dark red color, and fewer occurrences are represented by lighter shades of red and orange as shown in Figure 1(a) and 1(b).



1 (a)



1 (b)

Figure 1: (a) Hotmap screenshot (b) Magnification of heat scale

**4. Tag clouds:**

Kuo, Hentrich B.M. Good and Wilkinson[5] proposed a tag cloud method, to represent the words also known as tags for represent large collection of information. Their sizes, font and color change according to certain frequency or indexing.

**5. Resultmaps:**

Clarkson, Desai and Foley[6] introduced a hierarchical representation of web search results known as resultmaps. It uses hierarchical tree structure to map the documents into a tree map and highlights information.

**6. Rank Spiral:**

A. Spoerri[7] proposed a rankspiral method to scan large documents and their titles in single window. It uses spiral mapping to reduce occlusions and maximize density. Spiral structure thus solves the labeling problem.

**7. Scatter Plot:**

Nizamee and Shojib[8] made a search engine which displays the web results in the traditional list in the scatter plot thus making it more easy to use and compare.

**8. Clustering on Similarity measure:**

K.P. Wagh[9] proposed a similarity measuring technique that measures similarity between the snippets and cluster the results further so that similar type of documents becomes easy to navigate.

**III. DESIGN RATIONALE**

The list based approach of showing web documents is useful for many search tasks but not for ambiguous words or a broad topic. Visualization is important for more user friendly, easy way to navigate the search results. Visualization can be an add-on on the list based technique rather its replacement. It helps users to have a more comprehensive view of the query results.

**DESIGN RATIONALE:**

There are three major design goals:

**Arrangement based on Similarity:**

The similar type of snippets should be grouped and identified together. These documents are placed near to each other and dissimilar are placed far from each other. Such a 2-D layout may be obtained by a multi-dimensional projection technique. It preserves the neighbourhood similarity.

**Ranking of documents:**

The relevant documents are sorted and ranked accordingly. Hence, the visualization on search results must preserve the rankings of the documents. The size of our geometric documents can be resized according to the ranks. Users thus can visually discriminate the objects of different sizes.

**Uncluttered Layout:**

Overlapping on the structures replacing the points can create trouble for users to understand. Hence, to avoid this an uncluttered layout should be formed which is user friendly, easy and useful for the desired results. The search results should be displayed in a structured manner and colors given to different clusters for better understanding.

**IV. SYSTEM ARCHITECTURE**

The technique comprises of three steps:

1. Pre-processing the search results, and forming the tf-idf vector of each snippet obtained.
2. Multi-dimensional projection of each vector representation of snippets and then clustering them by k-means++ algorithm.
3. Re-ranking and re-organization of clustered results in a structured way using rankings and colors for comprehensive view.

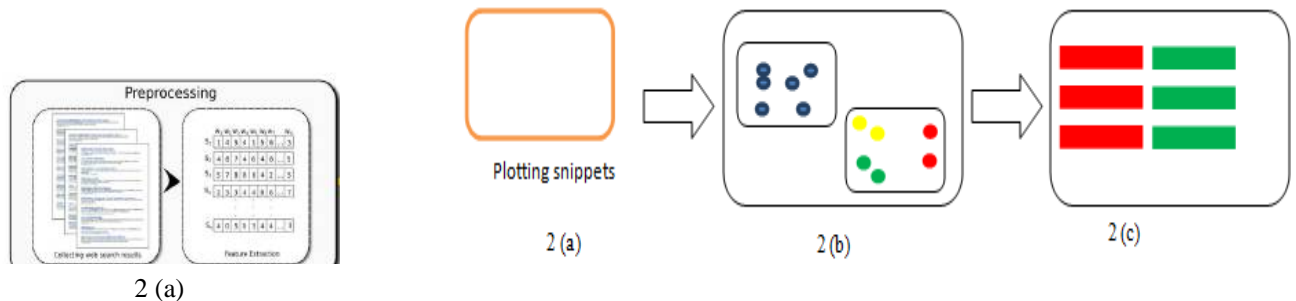


Figure 2: (a) Pre-processing (b) Multi projection and clustering (c) Re-ranking and organization

In figure 2, the first step, preprocessing of the web search results is done. The results are extracted and the stemming and stopwords removal is done. Then the tf-idf vector is formed of each snippet. Only the text in snippets are processed and not the full document.

In the next step, each vector is treated as a point on 2-D graph and plotted on by a multidimensional projection technique. Least square projection by Albeit can be used. Following this, the points are then clustered by using a k-means++[10] algorithm that uses an automatically detection of k.

Further, these clustered points are replaced by geometric shapes and organized in a structured way preserving the ranks. Thus, we get a comprehensive view of the web search results. This is shown in Figure 3.

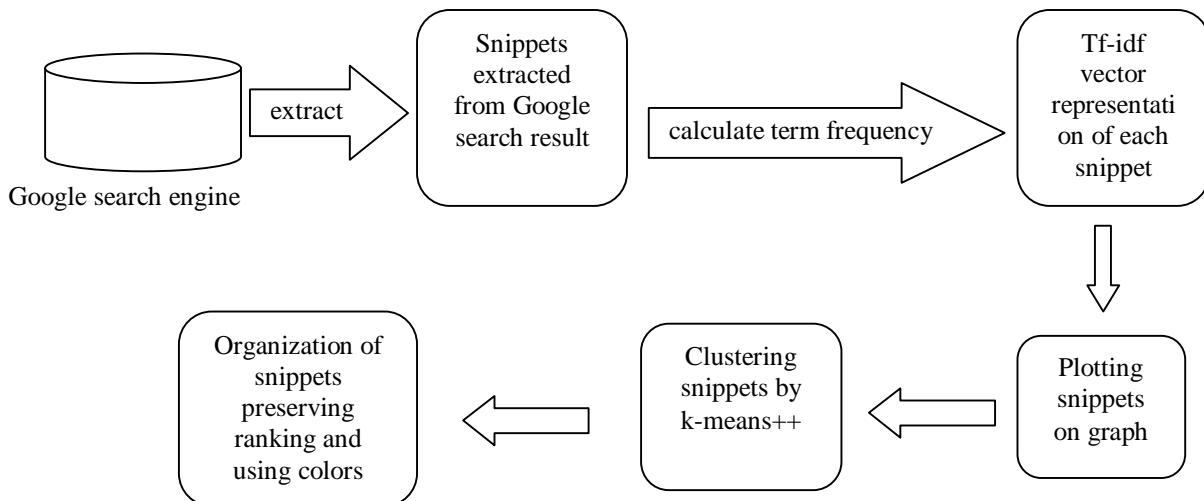


Figure 3: Architecture system

## V. CONCLUSION

We have introduced a novel approach to visualize the web search results in a comprehensive way by making it more useful, easy, user friendly. The limitations of the list based search results are overcome and new improved way of searching the web can be seen. The problem of overlapping is reduced by re organization of clustered preserving the similarity and ranking.

## REFERENCES

- [1] Erick Gomez-Nieto, Frizzi San Roman, Paulo Pagliosa, Wallace Casaca, Elias S. Helou, Maria Cristina F. de Oliveira, "Similarity Preserving Snippet-Based Visualization of Web Search Results", IEEE Transactions on Visualization and Computer Graphics, Vol. 20, pp. 457-463, 2014.
- [2] M.A. Hearst, "TileBars: Visualization of Term Distribution Information in Full Text Information Access," Proc. ACM SIGCHI Conf. Human Factors in Computing Systems, pp. 59-66, 1995.
- [3] T. Heimonen and N. Jhaveri, "Visualizing Query Occurrence in Search Result Lists," Proc. Ninth Int'l Conf. Information Visualisation, pp. 877-882, 2005.
- [4] O. Hoerber and X.D. Yang, "The Visual Exploration of Web Search Results Using Hotmap," Proc. 10th Int'l Conf. Information Visualization, pp. 157-165, 2006.
- [5] B.Y.-L. Kuo, T. Hentrich, B.M. Good, and M.D. Wilkinson, "Tag Clouds for Summarizing Web Search Results," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 1203-1204, 2007.
- [6] E. Clarkson, K. Desai, and J. Foley, "Resultmaps: Visualization for Search Interfaces," IEEE Trans. Visualization and Computer Graphics, vol. 15, no. 6, pp. 1057-1064, Nov. 2009.
- [7] A. Spoerri, "RankSpiral: Toward Enhancing Search Results Visualization," Proc. IEEE Symp. Information Visualization, pp. 208-214, 2004.
- [8] M. Nizamee and M. Shojib, "Visualizing the Web Search Results with Web Search Visualization Using Scatter Plot," Proc. IEEE Second Symp. Web Soc., pp. 5-10, 2010.
- [9] PH Govardhan, K.P. Wagh, P. N. Chatur, "Web Document Clustering using Proposed Similarity Measure", National Conference on Emerging Trends in Computer Technology, pp. 15-18, 2014.
- [10] D. Arthur and S. Vassilvitskii, "k-Means++: The Advantages of Careful Seeding," Proc. 18th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA), pp. 1027-1035, 2007.
- [11] Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr. P. N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach", International Journal of Emerging Technology and Advanced Engineering, vol. 2, pp. 110-113, 2012.
- [12] Rasika G.Charate, Dr.P.N.Chatur, Prof.K.P.Wagh, "Document Filtering: Intelligent Inference System for Web", International Journal of Management, IT and Engineering, vol. 3, pp.207-222, 2013.
- [13] S. Kolhe, K.P. Wagh, "Semantic Similarity based on Information Content", International Journal of Computer Science and Application Issue, pp. 82-86, 2010.
- [14] S. Dehankar, K.P. Wagh, "Web Page Classification using Apriori Algorithm and Naive Bayes Classifier", vol. 3, April, 2015.
- [15] P. Jadhav, K.P. Wagh, P. N. Chatur, "Integrating performance of web search engine with Machine Learning approach", Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2016.