An Approach to Second Generation Tag Cloud for Assessment of Business Search

Preeti Mundada[#], Archana Ghotkar[*]

[#]*ME-Computer Student,*
PICT, Pune, MH, IN.
[1]preetimundada@hotmail.com

[*]*Associate Professor,*
PICT, Pune, MH, IN.
[2]asghotkar@yahoo.com

***Abstract*— With the growth of Information technology, a very efficient visualisation is highly needed. One of the way to achieve this is by the use of tag cloud. Tag clouds are text-based visual representations of a set of tags usually depicting tag importance by font size and color. Tag clouds are used not only for visualisation but also for navigation and retrieval of data. Various techniques and algorithms for tag presentation are already there. This paper presents models for creation of a dynamic 2nd generation tag cloud in circular fashion. This implies that the most relevant tag is displayed at the centre with highest size. And others surrounding it. Tag cloud is generated by using specific search criteria, with additional functionality's of 2nd generation, such as refinement of tag cloud on the basis of time and location.**

***Keywords*— Information Visualisation, Information Search and Retrieval, Information System Applications, Tag Cloud.**

I. INTRODUCTION

Tag clouds are visual displays of set of words (or 'tags') in which attributes of the text such as size, color or font weight are used to represent relevant properties e.g. Popular search result,etc,. Tag clouds have become a popular method to support navigation and retrieval of tagged data. This paper seeks to optimise the display of tag clouds, without concern for the origin of the tags. Designers have to balance aesthetic considerations, for example with regard to font size and tag arrangement, against the capabilities of algorithms to produce optimised layouts. Tag clouds have become a frequently used interaction technique in the web. The popularity of tag clouds can be explained by their ability to help users in getting a fast overview of a certain area.

Tag clouds are typically presented in alphabetical order, but also different approaches such as random or importance based tag arrangements have been used[3]. Tag clouds have been popularised by social sites where users annotate shared resources using short textual labels, namely tags. In these sites, a tagcloud typically summaries a collection of resources by showing the most popular tags.

Tag clouds are becoming increasingly popular visualisation and interaction techniques used on the web. They have been shown to be useful for tasks such as helping the users of a website to gain a quick overview or summary of the information available[11] , clustering semantically related topics from a search query [10], or even for use as navigation interfaces [3].

In our work we are using tag cloud for assessment of business search. We are adding more functionality's to the tag cloud, such as the different visual effects, different searching and display options, like search by time, search by location, and many more.

Section 2 describes previous work. Proposed work and the system model is explained in section 3. Section 4 is conclusion and section 5 is references.

II. PREVIOUS WORK

Most existing studies evaluate tag clouds at a general level by comparing them with other types of user interfaces compared the tag cloud with tables[9]. An experiment in which the subjects could either use a tag cloud or a traditional search interface to answer questions, concluded that tag clouds are not a replacement but a valuable extension[10].

Visual features of tags have a very vast effect on tag cloud selection. Features have to be used very effectively to draw attention of viewers on appropriate tag [7]. Again there are various experiments on the layout of tag presentation [5].

Alphabetization makes user to find information more quickly, Font size and position of tag are very important in tag cloud presentation[3] explained three task layout matching:(a)Finding a specific tag: Sequential layout with alphabetical sorting.(b)Finding the most popular Tag: Circular layout with decreasing popularity.(c)Finding task that belong to a certain topic:Thematically clustered layout.

Various algorithms for drawing tag cloud are presented by different researchers. [6] explained two algorithms:(a) tag clouds with inline text.(b)tag cloud with arbitrary placement. Ranking on tag, based on frequency, diversity and aggregation is explained in [8], experiments concluded that increasing size of tag cloud improves its performance. The second generation functionality's of tag cloud [2]. In this paper we are presenting the use of all this papers to generate tag cloud , which will be helpful for business analysts to find various information in business domain.

III. PROPOSED IDEA

A. General description

This paper presents models for creation of a dynamic 2nd generation tag cloud in circular fashion. Circular placement implies that the Most relevant tag is displayed at the centre with highest size. And others surrounding it. Tags are placed in such a way to maximise space utilisation and to minimise overlap and space wastage. Circular placement requires sorting of tags in Descending order of weights. Each Tags in the tag cloud have weight according to which size and color to the tag is assigned. This weight is calculated according to the number of accesses to the particular object. Other characteristics of tag such as intensity, color and size are calculated form the weight itself. Sizes are considered in the relative form, in such a way that size of most highly weighted tag is 100. Other tag sizes can vary from 0 to 100, depending on their relative weight with respect to most weighted tag. Size of the tag at centre is highest at size of tag at boundary is smallest. Hence the color of the inner most element will be dark. Outer most set of element will be brightest one. Colors can be assigned to tags according to 256 bit pattern or 64-bit gray scale. So it will be very easy to find the most relevant tag as it is at the centre only. Figure shows an example of circular placement of tags, this figure is taken from[12]

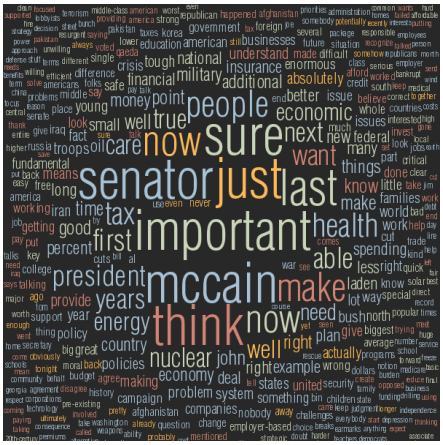

Fig 1:Tag cloud in circular style.

This tag cloud will automatically change according to selection of search criteria. This will include implementation of a set of tag selection algorithms for selecting tags from the hierarchical searches and using those tags for tag cloud generation. Various factors like ranking, relevance, overlap, coverage are considered, while creating tag cloud.

Ranking is the number of occurrence of object. For our tag cloud we are considering relative ranking of object.

Coverage is the space which is covered by the specific tag.

Overlap ensures that no two tags are overlapping in the space in tag cloud. Avoiding overlap is very important to support the proper navigability between web pages.

Relevance depicts how much the specific tag is related to the search criteria.

Similarity of tags is also needed, so as to avoid the duplication.

This factors are explained in details in the System model section. We will now see how an effective tag cloud contribute to the business search.

B. Business Implications

In any business environment we have set of reports, spreadsheets, dashboards, artifacts, etc,. Business analyst want to know how frequently a report or spreadsheet is being used, Which is the most popular search according to day and location. Which are the related searches in particular criteria, and for many more other details. We are generating tag cloud in the circular fashion, Which will encompasses the most relevant search item at the centre of the cloud and others circumference to it. This Tag cloud will help in rating the results in business environment, getting the list of frequently accessed reports, getting list of most popular searches in particular area and in many other similar applications.

Analyst will specify the criteria according to which tag cloud will be refined. Again we are providing two drop down box for specifying the time and location. Tag cloud will produce set of all tags which satisfy the criteria specified by user. For example, user want to search the report which is in high demand, The tag cloud was firstly and foremost implemented as a way to visualise which tags are most popular at a given moment. But, popularity shifts and something that is popular today may not be popular in a week or two. The first generation tag cloud is very static, and does not visualise the development in the tag space over time.

In 2nd generation Tag cloud, User will be able to refine the tag cloud by choosing year, month and day. So that user will be able to see which report is in more demand from last n number of days.

This type of tag cloud will also be useful in news highlighting. A temporal tag cloud is a tag cloud that implements utilisation of time to allow the user to choose the time frame for the tag cloud [2].A tag cloud provides its users with a snap-shot of the most used tags in a given time frame. Also the same thing we are applying for the location. User will be able to check or refine the result according to the specified location. Hence in this case we will need to use the concept of geographic location. For each access we will need to keep information such as, date and time of access and location of access and the number of access. Hence our tag cloud is helpful to all business components.

1) Executives can explore a timely and reliable view of their business , by a quick view of reports and take action to drive better business outcomes and decision-making.

2) Business managers have a single and reliable view of information at their fingertips that can help them focus on issues that need attention and action.

3) Business analysts can effectively conduct in-depth analysis and share results easily with broader business communities with the help of tag cloud.

C. SYSTEM MODEL

A set C of objects.(reports, spreadsheets, BI artifacts).A set T of tags. Each object c in C is associated with a tag t from T. Given a criteria Q, the set of results is a set C_q. $C_q \subseteq C$. T_q the set of tags that are associated to C_q, $T_q \subseteq T$.

For any two objects Ci, Cj, s.t.

$$Ci \in C_q \&\& C_j \in C_q,$$

$$\text{if, } s(q; C_i) > s(q; C_j),$$

then C_i is ranked higher than C_j .

1) Ranking:

Ranking is nothing but the number of occurrence of objects. Ranking will be incremented every time the object is used. R_i is the ranking of object i.

Rr_i is relative ranking of object. Relative ranking for object is defined as ratio of rank of object to the sum of rank of all objects.

$$Rr_i = R_i / \Sigma R_i$$

2) Relative Font:

Rf of tag is directly proportional to its Rr.

$$Rf \ \alpha \ Rr$$

Also the probability of C_i to be users choice is $P(C_i) > P(C_j)$. Probability of object is its relative ranking.

3) Extent of S:

Intuitively, the larger the extent of S is, the more topics, S potentially covers.

Ext(S) should be equal to Tq, since we are not considering limited number of tags for tag cloud.

$$ext(S) = | S | = | Tq |$$

4) Coverage:

Before actually putting the tag in the tag cloud all tags are need to be sorted according to their relevance.So the most relevant tag will be at the highest position in the list and placed at the center.

Coverage is the area which is covered by tag, i.e. A rectangle,with following coordinates.

$$C(T_q) = (x1 , y1 , x2 , y2).$$

Radius of the innermost circle can be calculated as,

$$r = (x2-((x1+x2)/2))$$

The next tag can be placed in the circumference $c = 2\pi r$.

5) Overlap:

All tag are placed in tag cloud in such a way that there is no overlap between any two tag.

Overlap can be expressed as, for all i, j

$$C(Tq_i) \cap C(Tq_j) = 0$$

6) Relevance:

Relevance of S with criteria Q, how much tag T_q is relevant to C_q.

$$rel(t,q) = ((t \cap Q) / Q)$$

7) Similarity function:

Sim(;) : C x C -> [0 ; 1], takes as inputs two objects out of our set C and outputs a number between 0 and 1. The higher the value of sim(;) is, the more similar the two objects are. For any two objects if sim(;) is 1 then two objects are merged in c.

$$P(c) = P(1) + P(2) - P(1)P(2).$$

IV. CONCLUSIONS

This paper presented a model for generation of 2nd generation circular tag cloud. Tag cloud is categorised in three different modules, one is the search criteria which is the basis for all other modules. Second is the time frame and the third one is the location based. In future we want to focus more on circular placement of tags. Also, the exact impact of this tag cloud on the business is need to be calculated.

REFERENCES

[1]P. Venetis, G. Koutrika , H. Garcia-Molina, "On the selection of tags for tag clouds," *in proc. fourth ACM international conference on Web search and data mining*, Feb. 09, 2011.

[2]M. Nielsen, "Functionality in a second generation tag cloud," *thesis of masters, Gjovik University College,* 2007.

[3]S. Lohmann, J. Ziegler and L. Tetzlaff, "Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration," *in proc, INTERACT,* pp 392-404, 2009.

[4]J. Schrammel, M. Leitner and M. Tscheligi, "Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches," *in proc. twenty-seventh international conference on Human factors in computing systems,* pp 2037-2040, 2009.

[5]M. Halvey and M. Keane 2007," An Assessment of Tag Presentation Techniques," *in proc. sixteenth World Wide Web Conference,* pp 1313–1314, 2007.

[6]O. Kasser, and D. Lemire, "Tag-Cloud Drawing: Algorithms for Cloud Visualization," *in proc. Tagging and Metadata for Social Information Organization Workshop,* 2007.

[7]M. Bateman, C. Gutwin, M. Nacenta, "Seeing Things in the Clouds: The Effect of Visual Features on Tag Cloud Selections," *in proc. nineteenth ACM conference on Hypertext and Hypermedia,* pp 193–202, 2008.

[8] D. Skoutas and M. Alrifai, "Ranking tags in resource collection," *SIGIR'11, (978-1-4503-0757-4/11/07),* Jul 24, 2011.

[9]J. Oosterman and A. Cockburn, "An empirical comparison of tag clouds and tables," *in proc. Twenty-second Conference of the Computer-Human Interaction,* Nov 22, 2010.

[10]I. Berlocher , K. Lee and K. Kim, "TopicRank: bringing insight to users," *in proc. Thirty-first annual international ACM SIGIR conference on Research and development in information retrieval,* July 20, 2008.

[11]B. Kuo, T. Hentrich, B.Good, and M. Wilkinson, "Tag clouds for summarizing web search results," *in WWW, (978-1-59593-654-7/07/0005.)*, pages 1203–1204, 2007.

[12]M. Krzywinski, (1999) [online]. Available: http://mkweb.bcgsc.ca/debatesdebate=mccain-obama-combined

[13]A. Rivadeneira, D. Gruen, M. Muller and D. Millen, "Getting our Head in the Clouds: Toward Evaluation Studies of Tagclouds," *in proc. SIGCHI Conference on Human Factors in Computing Systems,* pp. 995–998, 2007.