JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 30, 687-700 (2014)

DOI:10.1688/JISE.2014.30.3.9

A Novel Spatial Tag Cloud Using Multi-Level Clustering

JEN-WEI HUANG[1], KUAN-YING CHEN[1], YUAN-CHANG CHEN[1], KAI-NING YANG[2],
I-SHYAN HWANG[2] AND WEI-CHE HUANG[2]

[1]*Department of Electrical Engineering*
National Cheng Kung University
Tainan, 701 Taiwan
E-mail: jwhuang@mail.ncku.edu.tw
[2]*Department of Computer Science and Engineering*
Yuan Ze University
Chungli, 320 Taiwan
E-mail: ishwang@saturn.yzu.edu.tw

Conventional tag cloud systems can only present the frequency of tags and the connection of tags and tag clusters, but cannot properly express the strength of the relationships. In this paper we tackle this problem by improving the representation of tag clustering. We combine the advantages of conventional systems to create a multi-level interactive tag cloud system using clustering scheme. In our tag cloud system, tags are mapped onto a two-dimensional space. A force-directed algorithm is applied to make the tags move in 2D space. The tags will move until equilibrium is reached. As a result, the tags with high correlation are clustered in the close vicinity, and so are the tag clusters with high correlation. In addition, we let the users to interact with the tag cloud by zooming and panning for different perspectives and let user to select a tag than the tag will locate at the center of window and show the correlations of the user selected tag with other tags. We believe that our system can improve the user browsing experience, and can also be served as a great basis for social network analysis and other related research.

Keywords: tag cloud, interactive, multi-level, tag clustering, spatial

1. INTRODUCTION

Social bookmark sites have become phenomenal. The major players such as Facebook, Twitter, Flicker, *etc*. let users annotate original and shared contents of their interests and further share the across their social network. Other users in the same network can retrieve these contents through keyword search by user-defined tags. The same content might be annotated with different user-defined tags from different user perceptions. The collection of all these collaboratively created and user-defined tags is called Folksonomy [20]. Folksonomy allows users to classify all tags, and the tag with the most occurrences can serve as the tag representative of the content. Folksonomy also enhances the overall accuracy of search results and lays the foundation of our study.

Data visualization plays a very important role to analyze and understand tags. Data visualization is able to help to convey messages behind the data through presentations in graphics, animations, or charts [17]. Researchers can not only analyze data but also capture the embedded messages more easily. Among all the data visualization approaches, Tag cloud is a common model to show tags through different visual arrangements, such as var-

Received February 28, 2013; accepted June 15, 2013.
Communicated by Hung-Yu Kao, Tzung-Pei Hong, Takahira Yamaguchi, Yau-Hwang Kuo, and Vincent Shin-Mu Tseng.

ying font sizes or colors, to reveal the characteristics, importance, correlations of tags [7].

A tag cloud is composed of tags with frequent occurrences to inform users about popular information. It also helps users to attain new messages and discover patterns. Conventional tag clouds are built in the alphabetical order or by frequencies while the font size of a tag represents its frequency. In Figs. 1 and 2, we can see examples of tag clouds in the alphabetical order and by frequencies respectively [10]. However, in these two conventional tag clouds, we can only perceive the importance and the ranking of the individual tag to a certain degree, but not the correlation between these tags. The major problem is that similar tags are not grouped in the same cluster. The following work proposed a method to solve this problem, called tag clustering.

In order to express the correlation between tags, some clustering algorithms are applied to build tag clusters to form a tag cloud for the representation. The purpose of the cluster analysis is to extract information, characteristics of data and trends by data dependency. In Fig. 3, similar tags are grouped by a cluster algorithm to show the correlation between tags and clusters [19].

Fig. 1. Tag cloud in the alphabetical order.

Fig. 2. Tag cloud in the frequency order.

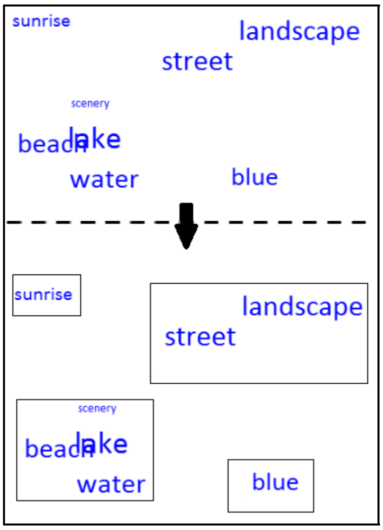

Fig. 3. Tag grouping.

The existing research about tag grouping [1] puts more emphasis on different presentations of tag cloud to explore the messages carried behind the tags and the correlation among them. There is a lack of higher level correspondence about tag clusters and without considering the relation of each tag. Therefore, we developed a presentation of tag cloud to facilitate researchers with data visualization to gather correlation information. The degree of the correlation among the tags can be explored and obtained by putting the tags onto two-dimensional space with coordinates. The font size of the tags means the importance of the tags. Users can explore the tag cloud from different angles by switching views. We leverage the law of gravity, which used to be Force-Directed Algorithm [14], to establish the tag clusters and use the correlation of the tags to generate tag cloud automatically. With our presentation, users can capture essential information in micro and macro perspectives of the interactive tag cloud at ease.

2. LITERATURE REVIEW AND DISSCUSSION

2.1 Tag Clustering

Tag clustering not only helps us to analyze data but also improves Internet browsing experiences [4]. The correlation of the tags must be found before tag clusters are established. There are numerous methods proposed for correlation calculation. They can be approximately divided into two categories. Tag co-occurrence, on the one hand, is to identify the terms that are semantically close or are common expressions [8]. Tags with more frequent co-occurrence have higher correlation. We could then calculate the correlation to identify the group of tags and build the tag clusters accordingly.

Self-Organizing Map, known as SOM, on the other hand, maps the input data from higher dimension to lower dimension and uses the topological map to express the distribution of each output value [15]. Self-organizing map also adopts neighborhood function to maintain the topological properties of the input data [16]. Thus the tags with similar properties will be mapped to in the same neighborhood.

2.2 Tag Representation

Conventionally tags are sorted in alphabetical order or frequency and the font size is used for the weight of tags. Layouts of this type are called tag clouds with inline text [12]. Another type of layout is called tag clouds with arbitrary placement. Tags are placed arbitrarily in the tag cloud, as shown in Fig. 4.

Tag clouds with inline text and tag clouds with arbitrary placement are great options for the tag representation. There is a huge disadvantage that this representation presents the frequency of tags only but cannot present the correlation of each tag correctly. In the next section, we will conduct two case studies of tag clustering based on these representations.

2.2.1 Case 1

First, we dived in Yusef's work by applying visual information retrieval to improve the representation of the tag cloud. The result is shown in Fig. 5 [21].

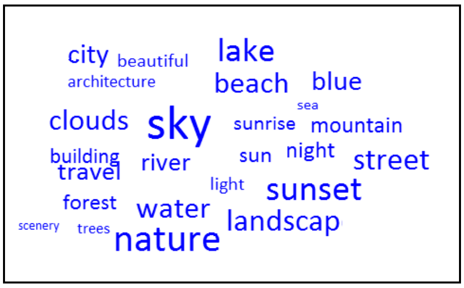

Fig. 4. Tag clouds with arbitrary placement [6].

Fig. 5. Tag clustering with tag co-occurrence.

Tag clustering with tag co-occurrence, mentioned in Section 2.1, is implemented to find the tag similarity by Jaccard coefficient and then the tags are clustered by *K*-means algorithm. Jaccard coefficient is commonly adopted to calculate the similarity and diversity of samples in statistics [13], and *K*-means algorithm can help to distribute *N* elements to *K* clusters based on attributes or correlation of the elements [19]. This algorithm requires the number of the clusters to be defined in advance, which is difficult to predict. The authors presented a tag in a row and tag clusters with high correlation will appear in the adjacent rows. In the same tag cluster, *i.e.*, in the same row, the tags with high correlation would be neighbors. We can identify the relationship between tags and clusters through horizontal and vertical views in a convenient way. In this presentation, tag clustering can be observed directly, but the exact correlation of each tag cannot obtain from this presentation. We just know the approximate correlation of each tag. If users need to obtain the exact relation of each tag then this presentation still cannot present all information.

2.2.2 Case 2

Another tag presentation is the clustering based on the contents as shown in Fig. 6. The Self-Organizing Map was utilized in [22]. Projecting the high-dimensional data to a low-dimensional view is an essential characteristic of SOM, which avoids repetitive calculation of correlation and therefore enhances the efficiency of the system. The authors presented the layout of the tag cloud by a 12×12 two-dimensional matrix, and each square in the matrix represented a tag cluster. SOM makes tags with higher correlation mapping closer to each other to build tag clusters. Tags with high correlation will be put in the same square, and the tag clusters with high correlation will be in the vicinity. Thus we can understand the relationship of tags and clusters in the two-dimensional matrix easily.

2.2.3 Summary

The above two cases we discussed are good examples to represent the relationship between tags and clusters. However, neither of cases shows the strength of the relation-

	0	1	2	3	4	5	6	7	8	9	10	11
0	apple osx mac	software freeware windows	howto Ubuntu linux	Java performance	tech computer hardware		Email library	flex flash		web word press		portfolio inspiration illustration
1	programming Python_net development	type	security					reference	tutorials	ajax	typography	design art
2	3d graphics open source			audio			rails ruby					fashion
3					videos video youtube					image	photoshop	Photo shopping photography
4	games game mobile					tool		tv movies	cool			diy
5	science database			Google search politics	news				interesting			funny
6	visualization math				seo jobs	work		economics finance	travel	green home		fic au humor
7	architecture					2008		articles		environment		language english
8	lifehacks					community		culture				health
9	Iphone			research	online	Internet						food cooking recipes recipe
10	collaboration socialnetworking php	tools		resources		twitter	advertising					
11	flicker productivity firefox wiki photos	tutorial		education teaching learning	technology blogging web2.0	social business socialmedia	marketing media		music mp3 download blogs blog	article	free to read writting	books book fun history webdev

Fig. 6. Tag clustering with SOM [22].

ships of tags and clusters. In Section 3, we will put tags onto two-dimensional space with coordinates and apply the force-directed algorithm to explore and present the strength of the relationship properly, which is the missing piece in previous cases.

3. MULTI-LEVEL INTERACTIVE TAG CLOUD

In this section, we develop a multi-level interactive tag cloud using a hierarchical clustering algorithm which is capable of presenting the strength of relationship between tags and use the forced-directed algorithm to adjust the distance between tags. The tag cloud presentation also provides an interactive browsing experience to allow users to explore the tag cloud from different perspectives.

3.1 Tag Representation

Now we will introduce the hierarchical clustering algorithm and forced-directed algorithm in our work. We want the tags with high correspondence to be in the same cluster in our tag cloud and place the tags in different cluster which the tags are not similar. We discussed folksonomy in the introduction about how users classify the user-defined tags and the tag with highest occurrences would apply as the tag representative for the cluster. The correlation value of any two tags T_u and T_v is given and stored in a tag association matrix. The diagonal line of association matrix is the frequency of tags. The association matrix is an upper (lower) matrix. Because the value which is located on upper is same as lower. With any tag association matrix, we apply the force-directed algorithm and the agglomerative hierarchical clustering algorithm on the matrix to explore the correlation between tags and tag clusters. The following is the concept of the force-directed algorithm. We define two n by n square matrixes, R and D, for n tags:

$$D_{uv} = \frac{k}{R_{uv}}, R, D \in Z^{n\times n} \tag{1}$$

where k is the distance constant. The distance constant used to adjust the distance between tag u and tag v. R_{uv} is the tag association matrix recording the correlation of any two tags T_u and T_v and D_{uv} is the ideal distance matrix between tags, which is the equilibrium after force-directed algorithm is applied. If two tags have the shorter distance, they must have the correlation of these two tags. They possess the higher correlation to each other. By mimicking the attraction and repulsion forces in physics, the force-directed algorithm can be utilized to represent the data in figures [14]. In our design, we regard the relationship between tags in the tag association matrix as the force in the force-directed algorithm. We define a function $D(T_u, T_v)$ for the distance between two tags. If the current distance $D(T_u, T_v)$ is greater than the ideal distance, D_{uv}, a force of attraction will incur to draw two tags closer. On the other hand, if the current distance $D(T_u, T_v)$ is shorter than the ideal distance, D_{uv}, the force of repulsion will separate two tags further. The concepts of the force-directed algorithm are as follows.

$$\begin{cases} \text{if } D(T_u, T_v) > D_{uv}, \text{ draw two tags closer} \\ \text{if } D(T_u, T_v) < D_{uv}, \text{ separate two tags futher} \end{cases}$$

When the force field of all tags in the tag cloud reaches equilibrium, *i.e.*, no tags move further, the initial tag cloud is built. We will get a graph Fig. 9 when the force-directed algorithm is complete. In this graph, the distance between tags means the correlation of two tags. The location of the similar tags will be neighbor. Some of tags which are not closed with other tags will be separated. We can get the first result from force-directed algorithm. If users need to change the views of the window *i.e.* zoom-in or zoom-out, we need to user hierarchical clustering to accomplish this work.

3.2 Agglomerative Hierarchical Clustering

After applying the force-directed algorithm to create an initial tag cloud, we continue working on multi-level interactive functions, *i.e.*, zoom-in, zoom-out, and the translation on the tag cloud. In this paper, we use the hierarchical clustering algorithm to complete the works which the clusters combine or separate. Zooming out the tag cloud would cause the presentation of tag cluster to shrink to a small range and created an unfriendly view of individual tags. The agglomerative hierarchical clustering algorithm would help to select the tag representing the cluster for a better view. The algorithm proposed by Jang [3] states that the hierarchical clustering method splits and groups the data repetitively through hierarchical structure. When the clustering is done, a tree would be generated as shown in Fig. 7. The concept of hierarchical clustering algorithm is to merge two clusters, C_i and C_j with the shortest distance among all clusters and repeat this step until the total number of clusters reduces to the pre-defined ideal number of clusters. The distance between two clusters is defined by single-linkage agglomerative algorithm, which is the distance between closest two points from each cluster as follows:

$$D(C_i, C_j) = \min\{d(a, b)\},\ a \in C_i,\ b \in C_j. \quad (2)$$

Based on this concept, we apply the agglomerative hierarchical clustering on our initial tag cloud. As a result, we can identify the tag representative of each cluster by the hierarchical clustering when users switch their views, say zoom-in, zoom-out, and the translation.

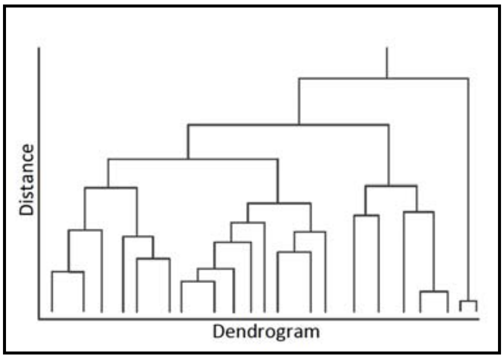

Fig. 7. Agglomerative hierarchical clustering.

4. DEMONSTRATION AND DISSCUSSION

In this section, we show the results and demonstrate the interactive features of our tag cloud. The tag cloud is implemented on Canvas by HTML5 and drew our final view by KineticJS from Canvas JavaScript library [2].

4.1 Testing Data

The proposed multi-level interactive tag cloud applies to any tag system given any tag association matrix. In this work, we adopt the tag association matrix from a progressive image search and recommendation system, PISAR system, proposed by Huang [11], to achieve this purpose. The tag association matrix in PISAR system is a sparse matrix, which means that only non-zero value is stored. This approach works great on large data sets [18]. For the testing data, we select 30 most popular tags in PISAR system. In order to render correlation in the tag cloud automatically through the tag association matrix, the layout of the initial tag cloud was established by tag clouds with arbitrary placement mentioned in Section 2.2. The tags were also mapped to a two-dimension space with coordinates embedded, as shown in Fig. 8. This layout facilitates the verification of the algorithm later. Then, in order to represent the strength of relation between tags and

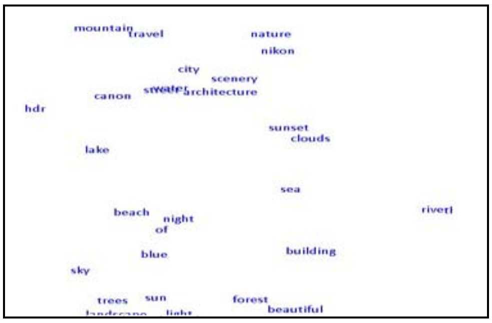

Fig. 8. Original tag layout.

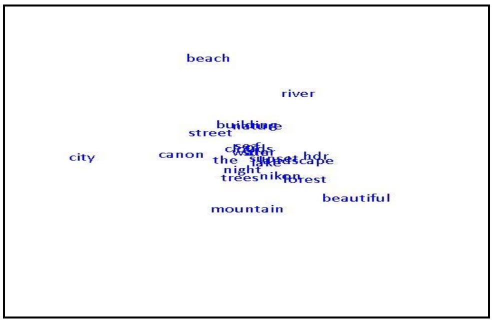

Fig. 9. Initial tag layout.

clusters, we applied the force-directed algorithm to the tags. These tags will be gradually shifted to the equilibrium position shown as in Fig. 9. In Fig. 9, the representation still exists a few questions. First, the users cannot recognize the importance of the tags clearly. The representation has only presented the correlation with each tag but has not presented the frequency of each tag. Second, there are many insignificant tags in the association matrix from PISAR, For example, "the", "of", "hdr"...*etc*. These insignificant tags will cause many problems in our program. The insignificant tags will appear lots of times in many image tags, but these tags have not any relation with any images and the insignificant tags will cause the error when the hierarchical clustering algorithm and force-directed algorithm are working. So, we need to use some methods to solve these two problems. First, we will use the font size to recognize the importance of all the tags. Second, we need to prune the insignificant tags from association matrix before we use the hierarchical clustering algorithm and force-directed algorithm to analysis the association matrix. Therefore, the users can recognize the importance of each tag directly and the insignificant tags will not effect to the result of the hierarchical clustering algorithm and force-directed algorithm. The users can get the clear and pure result from this program.

In the steady (equilibrium) state, the tags with high correlation will dominate the positions of those with low correlation. Unfortunately, in the current condition, we cannot know if each tag is popular or not. To remedy this, we can reveal this information by using their font size to recognize the popularity. If the font size is bigger, it means the tag is more popular. In Fig. 10, we sort the frequency of tags and print the top ten frequent tags. The users can gain the information from our result directly. In this example, the most popular tag is "sky." Generally speaking, it is used widely everywhere. The followings are "nature" and "sunset," *etc*.

In Fig. 9, initial Tag Layout, the tags of this image are the data which we do not filter out yet. We use the strategy to filter the tags which are insignificant such as "The", "of", "Nikon", and so on. After the filter we will get the new image shown in Fig. 10. In the left-upper corner, we can see the top 10 popular tags. In order to further see some specific tag clearly, we can click on "sky" or search for "sky" directly to move it in the center of the window. The consequence is shown as Fig. 11.

In addition, this bird-eye view makes the tags look crowded and it is difficult to explore and identify details. In order to further analyze the relationship between tags and clusters, we can zoom in the center of the cloud to group the tags by the mark, as shown in Fig. 12. From the tag association matrix, we can see how the tags are attracted and

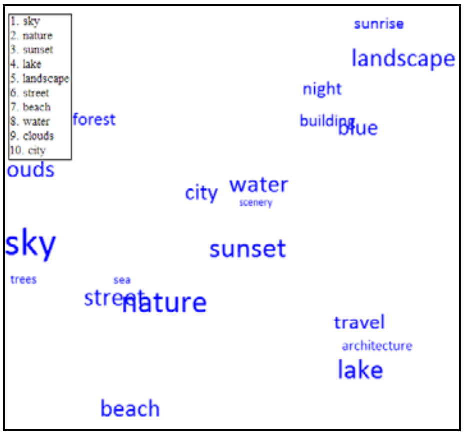

Fig. 10. Popular tags with font size.

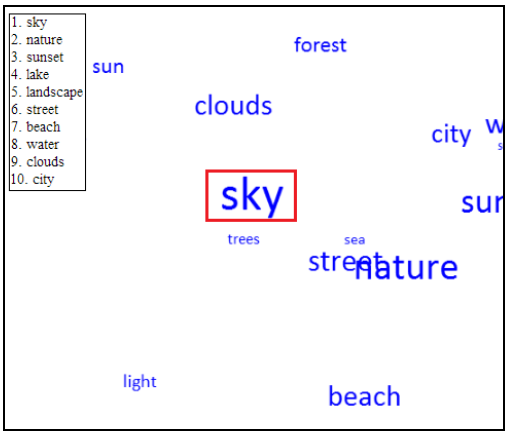

Fig. 11. Central tag cloud.

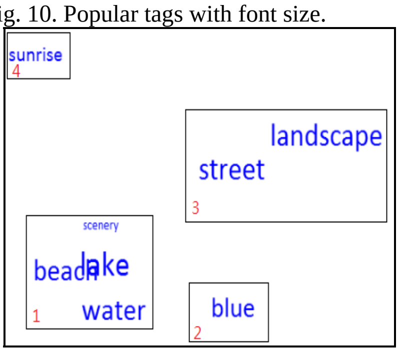

Fig. 12. Zoom-in the tag cloud.

Fig. 13. Results shown by algorithm [21].

repulsed by their correlation to generate tag clusters. We verified that force-directed algorithm did lead to the correct clustering results.

We also implement the algorithm mentioned in the case studies [21] to build the tag cloud from the same data. The result is shown in Fig. 13. We can see both methods put 8 tags into 4 groups. In our result shown in Fig. 12, we can see clearly the strength of the correlation by finding the relative distance of cluster centers. Cluster 1, 2, and 3, which are marked as red color with numbers, have higher correlation within the cluster, but the correlation within cluster 4 and between cluster 4 and all the others are lower. In addition, distance between tags also shows the strength of relationship between tags. In Fig. 12, the graph also can separate the outlier from the cluster. Clearly, the cluster 4 is the outlier in the graph. In Fig. 13, we can only see the connection of the clusters but without the strength of the relationship. Based on our result we have shown the improvement on presenting the strength of relationship which was missing in previous tag clouds and reserve the advantage in the traditional tag clouds.

As for the layout of the tag cloud, we can zoom out the tag cloud to obtain a bird-eye view of the tag cloud. However, tags on this view are overlapped and too crowded to express themselves clearly as shown in Fig. 14. In order to make the tag clusters concise and easy to recognize, we refine our layout by applying the agglomerative hierarchical clustering algorithm to simplify the representation of our tag cloud. The cluster will combine the other cluster which is the most closed cluster when the user zoom-out the window. When the users zoom-in, the clusters will separate from the original cluster and create a new cluster according the hierarchical clustering tree in Fig. 15. Moreover, we can see the tag number 5 and number 2 are high relation in this example. Also we can find the tag number 8 and number 9 are high relation. If we set tag number 5 "sky" be the central tag. We find the condition obviously that tag number 10, number 4, number 1 and number 3 are locate far place which means their relation is lowest in Fig. 15. In Fig. 16, a tag representative would be generated to mark each cluster by single-linkage agglomerative algorithm to group the tags, which results in a final simplified tag cloud.

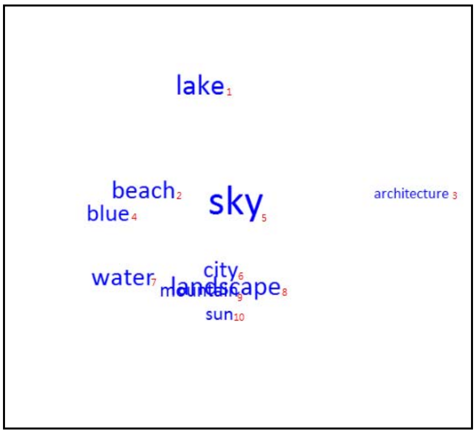

Fig. 14. Zoom-out view of the tag cloud.

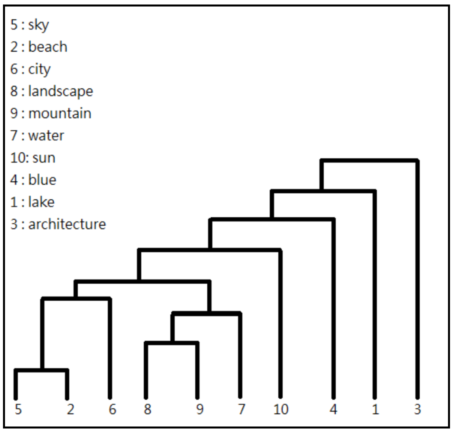

Fig. 15. PISAR with agglomerative hierarchical clustering.

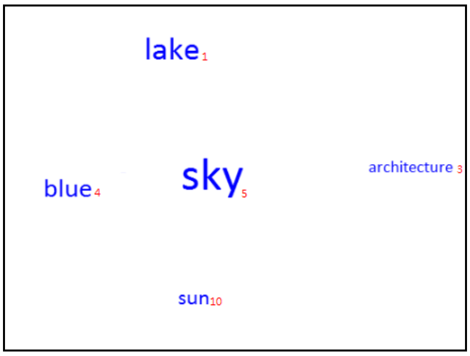

Fig. 16. Simplified representation of tag clusters.

In this paper, we have three problems in our implementation. First, we can use another method to improve the problem in Fig. 14, when the numerous of tags located on the same place. The area which contains many tags will be very crowd. This appearance will cause that users cannot recognize all the tags clearly. Users need to zoom-in or zoom-out the window when they want to understand the detail of this area. We will select some typical tags to represent the tags which are huddled and user can obtain the approximate information from this cluster. If users want to confer this cluster more detail, users can change the views of this cluster. Second, we use the technique hierarchical clustering to cluster the tags. Unfortunately, users cannot recognize the cluster directly. We will color the each cluster by different colors. Because of it, the users will distinguish the different clusters quickly and directly. Finally, when we zoom in or zoom out the window. Some tags might be out of the window. This problem will cause the users cannot find the tags which they want to examine. We have a solution to solve this problem. In the finite window size, if the location of tags is outside of the window, we use the tags and arrowheads to figure out the exact position. Thus, the users will directly and quickly know the location of tags which they want to examine. If the tags which users want to explore is out of window, users can according to the tags and arrowhead and move the window and users will find the exact location of the tags which they want to seek.

5. CONCLUSION AND FUTURE WORKS

5.1 Conclusions

In our paper, we improved the representation of tag clusters by showing the strength of tag clusters while the existing algorithms did not. We developed a new tag cloud using the clustering algorithm, a Multi-level Interactive Tag Cloud. In our work, we mapped our tags onto two-dimensional space with coordinates and used the force-directed algorithm to find the equilibrium position of the tags. With the help of the agglomerative hierarchical clustering, we can enable interactive user perspectives with different level of views. We keep the traditional works to use the font size to represent the frequency of tags and print the top ten tags on the windows. Our tag cloud focused not only the relationship but also its strength between the tag and the cluster. With the improvement of the micro and macro presentation, the users browsing experience have also been improved. This work will be applied on social network analysis and other related works.

5.2 Future Works

Although the result has been improved and become more concise, our program can still be optimized for better efficiency. Since the calculation of tag associated matrix would occupy considerable resources and take long computation time, the continuing work can adopt MapReduce or the similar type of distributed computing framework to improve the efficiency of the program. MapReduce distributes the tasks to a large number of nodes in the network, and wait for the node to complete and return status and results [5]. As a result, multiple nodes can compute multiple subtasks of the original at the same time to bring efficiency to the program. A wide range of test data can also be adopted to ensure the integrity and accuracy of our algorithm. We anticipate providing more features such as user-recommended tag, links to social network website, and the presentation of the relationship between tags and clusters. This would help to enrich the research in social network and beyond.

REFERENCES

1. Tag Clustering – A Tutorial on Clustering Algorithms, http://home.dei.polimi.it/matteucc/Clustering/tutorial_html.
2. KineticJS – HTML5 Canvas JavaScript Library Framework, http://www.kineticjs.com/.
3. M.-R. Anderberg, *Clustering Analysis for Application*, Academic Press, University of Michigan, MI, 1973.
4. G. Begelman, P. Keller, and F. Smadja, "Automated tag clustering: Improving search and exploration in the tag space," in *Proceedings of Collaborative Web Tagging Workshop*, 2006.
5. J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, Vol. 51, 2008, pp. 107-113.
6. D. DiNucci, "Fragmented future," *Print*, Vol. 53, 1999, pp. 32.
7. M.-J. Halvey and M.-T. Keane, "An assessment of tag presentation techniques," in *Proceedings of International Conference on World Wide Web Conference*, 2007, pp. 1313-1314.
8. R.-M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, 1973, pp. 610-621.
9. J.-A. Hartigan and M.-A. Wong, "Algorithm as 136: A *k*-means clustering algorithm," *Journal of Applied Statistics*, Vol. 28, 1979, pp. 100-108.
10. M.-A. Hearst and D. Rosner, "Tag clouds: Data analysis tool or social signaller?" in *Proceedings of International Conference on System Sciences*, 2008, pp. 160.
11. J.-W. Huang, C.-Y. Tseng, M.-C. Chen, and M.-S. Chen, "PISAR: Progressive image search and recommendation system by auto-interpretation and user behavior," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 1442-1447.
12. O. Kaser and D. Lemire, "Tag-cloud drawing: Algorithms for cloud visualization," in *Proceedings of International Conference on World Wide Web-Workshop on Taggings and Metadata for Social Information Organization*, 2007.
13. B.-J. Kefford, R.-B. Schäfer, and M. Liess, "A similarity-index based method to estimate chemical concentration limits protective for ecological communities," *Environmental Toxicology Chemistry*, 2010, pp. 2123-2131.
14. S.-G. Kobourov, "Force-directed drawing algorithms," 2004.
15. T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, Vol. 43, 1982, pp. 59-69.
16. T. Kohonen, *Self-Organizing Maps*, Springer, NY, 1997.
17. F.-H Post, G.-M. Nielson, and G.-P. Bonneau, *Data Visualization: The State of the Art*, Springer, NY, 2002.
18. W.-H. Press, B.-P. Flannery, S.-A. Teukolsky, and W.-T. Vetterling, "Sparse linear system," *Numerical Recipes in FORTRAN; The Art Scientific Computing*, 1993, pp. 63-82.
19. H.-C. Romesburg, *Cluster Analysis for Researchers*, Lifetime Learning Publications, Belmont, CA, 1984.
20. T.-V. Wal, "Folksonomy," Online Information 2005, http://www.vanderwal.net/essays/051130/folksonomy.pdf.

21. H.-M. Yusef and H.-S. Victor, "Improving tag-clouds as visual information retrieval interfaces," in *Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies*, 2006, pp. 25-28.
22. A. Zubiaga, A.-P. García-Plaza, and V. Fresno, "Content-based clustering for tag cloud visualization," in *Proceedings of International Conference on Advances in Social Networks Analysis and Mining*, 2009, pp. 316-319.

Jen-Wei Huang (黃仁暐) received the BS and Ph.D. degrees in Electrical Engineering from National Taiwan University, Taiwan in 2002 and 2009, respectively. He was a visiting scholar in IBM Almaden Research Center from 2008 to 2009 and was an Assistant Professor in Yuan Ze University from 2009 to 2012. He is now an Assistant Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He majors in computer science and is familiar with data mining. His research interests include data mining, mobile computing, and bioinformatics. Among these, social network analysis, progressive mining, mining time series, and multimedia data mining are his special interests. In addition, some of his research are on data broadcasting, privacy preserving data mining, e-learning and bioinformatics.

Kuan-Ying Chen (陳冠穎) received the BS degree in Computer Science from Yuan Ze University, Taiwan in 2012. He is pursuing the master degree in National Cheng Kung University, Taiwan. His research interest is in data mining.

Yuan-Chang Chen (陳元章) received the BS degree in Information Management from Chang Gung University, Taiwan in 2012. He is pursuing the MS degree in National Cheng Kung University, Taiwan. His research interest is in data mining.

Kai-Ning Yang (楊凱甯) received the BS degree in computer science from Yuan Ze University, Taiwan in 2013. He is pursuing the master degree in National Sun Yat-Sen University, Taiwan. His research interest is in data mining.

I-Shyan Hwang (黃依賢) received the BS in Electrical Engineering and MS in Electronic Engineering from Chung-Yuan Christian University, Chungli, Taiwan, in 1982 and 1984, respectively, M.S. and Ph.D. Degrees in Electrical and Computer Engineering from the State University of New York at Buffalo, N.Y. in 1991 and 1994, respectively. Since February 2007, he has been promoted as a Full Professor in the Department of Computer Engineering and Science at the Yuan-Ze University, Chungli, Taiwan. His current research interests are Fault-tolerant Computing, High speed Fiber Communication, Mobile Computing and Heterogeneous Multimedia Wireless Networks, Integration Assessment of PON with Broadband WiMAX, Optical Networks based Infrastructure over Cloud Computing. He serves as a member of the Editorial Board for the Springer Photonic Network Communications Journal.

Wei-Che Huang (黃偉哲) received the BS degree in Computer Science from National Yuan Ze University, Taiwan in 2013. He is pursuing the master degree in National Chung Cheng University, Taiwan. His research interest is in hardware.