Qstack: Multi-tag Visual Rankings

Phi Giang Pham[1], Mao Lin Huang[1, 2*]
[1] University of Technology, Sydney, New South Wales, Australia.
[2] Tianjin University, Tianjin, China.

* Corresponding author. Tel.: +61 2 9514 1858; email: mao.huang@uts.edu.au
Manuscript submitted January 30, 2016; accepted February 20, 2016.
doi: 10.17706/jsw.11.7.695-703

Abstract: Multi-tag-based search is quite popular on collaborative websites and sharing-online-content systems. For this kind of search results, the challenge is how to compare grouped-tag values of tag collections on heterogeneous alternatives. This paper introduces a new visualization approach named Qstack for dealing with the challenge. Qstack purpose is to help users to visually rank multi-tags based on grouped-score combination within and across the categorized alternatives. The methodology applying interactive stacked bars, dynamic queries and adaptive focus+context techniques enables users to easily create and adjust grouped-tag rankings of a large number of heterogeneous alternatives. A case study on Flickr photo award allocation will be presented for Qstack demonstration. We conducted a qualitative study for evaluating Qstack effectiveness, and the result indicates that our approach is useful for multi-tag rankings.

Key words: Multiple tags, visual rankings, interactive stacked bars.

1. Introduction

A long with development of social networks, multi-tag-based queries are becoming popular for meta-data construction and information seeking such as photo search on Flickr and book search on LibraryThing [1], [2]. The characteristics of a multi-tag-based search might lead to a heterogeneous result which includes both the totally satisfying items and partially satisfying ones. Normally, the total satisfaction is concerned; however, for complex analytics, the remaining ones might be valuable as well. For instance, we are looking for books by terms hacker, internet, security and secret and receive a number of diverse books. Besides of the four-tag-satisfying books, we might be still interested in the three-or-two-tag-satisfying ones. Furthermore, multi-tag values could be used to support decision making. For instance, in a photo association, using the scores of tag values towards topics to rank members is essential for the judges to make decisions on annual award allocation. The topics could be formed by single tags and groups of tags, which would lead to heterogeneous alternatives. Our motivation is to deal with the challenge of analyzing heterogeneous items given by flexible sets of considered tags.

For generalization of the topic, multi-tag comparison is referred to the multi-attribute comparison due to the similarity of score computation algorithms. We have studied the multi-attribute visual ranking techniques and believe they would be useful to display given items in visual patterns for analysis. Current visual ranking tools well support combinations of single or individual attributes [3]; however, they have not aided to the rankings of grouped-attribute combinations.

This paper introduces a new visualization technique named Qstack which enables to flexibly rank and analyze grouped-tag combination within and across categories of given items. Our methodology is to develop visual rankings of multi-tags based on a chain of multiple stacked bars and dynamic queries. By adaptive focus+context features, users are able to easily navigate combined tags and clustered items during exploration.

The contributions of this paper include:

- Qstack, a new visual ranking technique and its prototype supporting grouped-tag rankings inside and across categorized alternatives,
- A qualitative study assessing usability of Qstack for multi-tag visual comparison.

2. Related Work

2.1. Tag Visualization

A tag is a word or a group of words assigned to online items or extracted from text collections. While multi-tag layouts are commonly represented in lines, circulars and clusters, tag frequencies are visualized by text font sizes and colors [4], [5]. Tag cloud visualization is a well-known approach for analyzing text collections at a single time point and over time series. Combination of trend charts and tag clouds is a typical method proposed for overtime tag comparisons [6], [7]. The charts show the change of tag usage, and the clouds display the representative key words via force layouts. In addition to the trend charts, parallel coordinating visualization is offered for faceted text collection analysis [8]. The tag clouds are represented in multiple columns, associated items are represented in stacked bars, and the tag relations are displayed in colorful lines. Tag relationships could be analyzed by hierarchical networks with a semantic model as well [9]. For tag-based-search, the clustering algorithms are mostly used for improving this kind of search results [10]-[12]. For dealing with different kinds of online information, tag clouds are combined with multi-graph interaction to improve navigation processes [13].

The targets of current approaches almost focus on visual comparison of tag values for query refinement and text collection mining; however, the challenge of how to use tag values to support choosing a right item in a heterogeneous collection has not been addressed. With the motivation, our study aims to deliver a visualization technique which enables to analyze, compare and rank the items against groups of tag values. These operations could help users to choose a suitable alternative for their needs.

2.2. Multi-attribute Visual Rankings

While stacked charts are a well-known visualization approach for multi-attribute data summarizing [14]-[16], stacked bar charts are a popular technique for comparing and ranking such data. The stacked bar layouts could be formed from multi-bars aligned in a same baseline, multiple baselines and stacked positions. While the same baseline layouts are suitable for a small number of items and attributes due to the horizontal space limitation, the latter methods optimize both vertical and horizontal space usage for enhancing scalabilities. Table Lens is a basic idea of multi-baseline bar charts where the single bars of each item are allocated in the cells of a table or a spreadsheet [17]. The structure is associated with Fish-eye technique which maximizes focused cells and minimizes the remaining ones. This approach enables flexible comparison of attributes within and across items. LineUp, another stack-bar-based approach, is a task-driven visualization for mapping and ranking combined attributes [3]. By applying Fish-eye and slope charts, the technique enables users to track the changes of ranks in same collections overtime.

The current works would rather support single-attribute or single-score combination rankings than consider the rankings based on grouped attributes or grouped scores. The grouped-score rankings are essential when analytic criteria concentrate on sets of attributes instead of single ones. This paper addresses such challenge via the case of multi-tag rankings with grouped-tag comparison and proposes a new visualization method for multi-attribute rankings of grouped-score combination.

3. Ranking Visualization Technique

3.1. Basic Design and Interaction

Qstack is an interactive visualization technique developed for multi-tag rankings by grouped-score comparison. The technique constructs chains of stacked bar charts with focus+context features which are synchronously displayed in a single screen for analysis. By the design, users are able to analyze both inside and across clusters of given items according to the groups of tags.

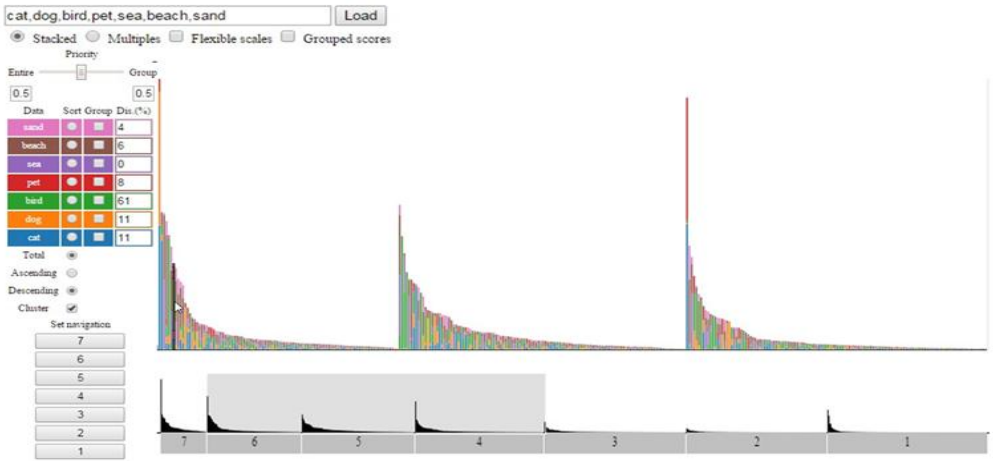

Fig. 1. Basic design of Qstack.

The general design of Qstack is a serial stacked bar representation controlled by a query panel (Fig. 1). The visual representation includes a context side as the sub view and a focus side as the main view. In context side, the heights of bar charts encode total scores of all grouped tags as the ranks of all items while stacked bar charts in focus view visualize original distributions of tag values for single tags and grouped tags. The total heights of the stacked bars encode grouped scores of tags (Fig. 2).

For exploration control, Qstack features a dynamic query panel and interactive bars. The query panel supports users to arbitrarily create and adjust the visual analysis by the operations as follows: *Multi-tag Load* to load the relevant items against all considered tags; *Stacked, Multiples, Flexible scales* to change the layouts between stacked bars and multi-baselines with flexible scales; *Entire, Group Priority* to assign priorities as the weights on entire tags and grouped tags; *Grouped-scores* to switch between grouped-score comparison and single-score comparison; *Data column* to list all tags and its encoded colors; *Sort column* to sort the results in order by single or total scores of tags; *Group column* to combine the tags for grouped scores; *Dis.(%)* to show the distribution of tag values on each item and *Cluster* to cluster the results against all groups of tags.

Users can watch the detail of tag values by moving mouse pointer over corresponding bars. The activated bars, then, will be highlighted in both focus and context view for item tracing. The identified information of items is displayed only when horizontal space is sufficient.

3.2. Group-Score Rankings

Strength of Qstack is the ability to flexibly compare and rank grouped scores of tags. Users are able to visually combine tags to a group, assign group weights and analyze ranking patterns of given items (Fig. 2). The ranking patterns in focus view are double-layer stacked bars consisting of the original scores of entire tags in bright transparent background and the grouped scores of combined tags in front. The design enables users not only to keep tracking on total rankings but also to observe the causes from the sets of tags.

Additionally, ranking component allows users to simply cluster and quickly navigate the given items towards groups of tags. All the clusters are aligned in a serial order by user demands. Therefore, the comparison of multi-tags could be analyzed both inside and across clusters.

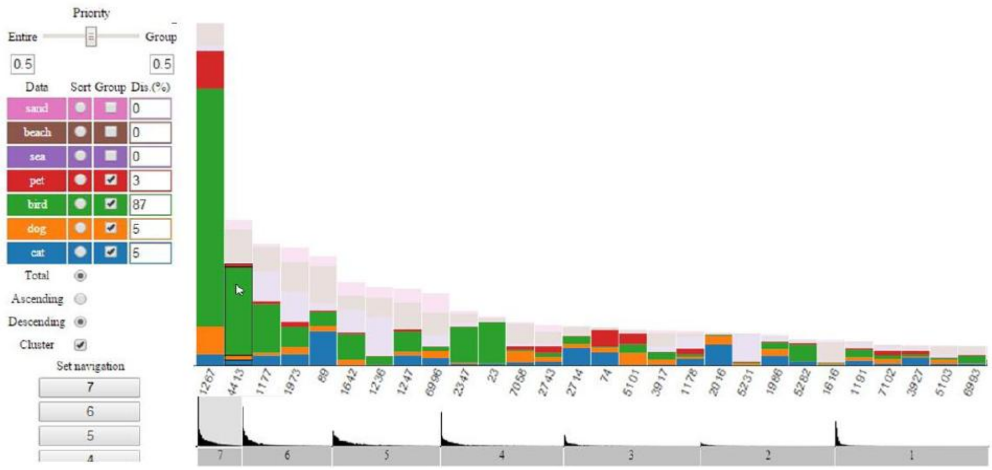

Fig. 2. Grouped-score visual ranking.

The formal definition of the rankings is based on the grouped scores of tag sets A and A' on each item with W and W', where $A=\{A_i|i=1,2...n\}$ is the entire considered tag values, $A'=\{A_j|j=1,2...n'\}$ is a sub set of A grouped by users, and W and W' are the given weights on A and A'. It is supposed that C and D are sets of given items where C satisfies all tag values, D satisfies partial tag values. So, we compute the total scores $G(C_x)$, $G'(D_y)$ for $C_x \in C$ and $D_y \in D$ by

$$G(Cx) = W * \sum_{i=1}^{n} S(Cx, Ai) + W' * \sum_{j=1}^{n'} S(Cx, Aj) \text{ and } G'(Dy) = W * \sum_{i=1}^{m} S(Dy, Ai) + W' * \sum_{j=1}^{m'} S(Dy, Aj)$$

where S is the single score of a tag value on one item, $m < n$ and $m' <= n'$.

3.3. Scalability

For handling a large number of items displayed in horizontal direction, Qstack applies zooming by brushing, a powerful focus+context navigation technique. There are two brushing components including navigation menu and customized brushing (Fig. 2). Users can either click on navigation menu to quickly jump to a zoomed cluster or directly customize the width of the brushing rectangle and drag it to a focus area. While the former method is appropriate for individual cluster access, the latter one is useful for flexible focus to the details across clusters.

The entire tags and grouped tags are represented in double-layer stacked bars with same baseline, which supports to save vertical space of focus display (Fig. 2). For dealing with a large number of tag values represented in stacked bars, Qstack provides multi-tag filtering function. The selected tag bars are kept visible while the remaining bars are converted to a transparent layer (Fig. 3).

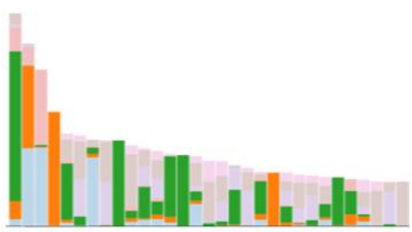
Fig. 3. Multi-tag filtering.

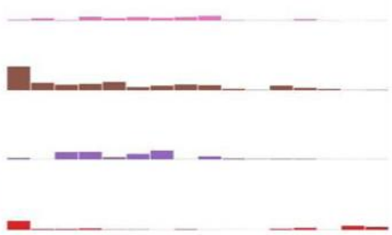
Fig. 4. a. Normal scales.

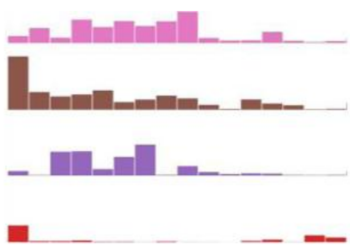
Fig. 4. b. Flexible scales.

In addition, in order to support multi-tag comparison across items, the stacked bars can be transformed to a multi-baseline chart with flexible scales (Fig. 4). The flexible-scale function improves the vertical space usage by adjusting the height of each bar against the highest one of the same baseline. Besides, this component is helpful for users to easily recognize the missed tag values in an item and a cluster.

3.4. Data Mapping

Qstack extracts data from a database where tag information is stored in dimensions of user identification (ID), tag names and the number of tag frequencies. Once normalized from tag frequencies, tag values are organized in a two-dimension matrix appropriately designed for stacked bar visualization. One of two dimensions contains user ID, and the other contains single tag values, total tag values and tag set indices. The tag set indices are used to categorize given items and serve the navigation feature. The number of tag set indices is the number of sub sets of n tags defined by $p=\sum_{k=1}^{n} C(n,k)$. In order to reduce the item clusters, the navigation function is implemented according to k instead of p.

The data used for the instances throughout this paper are a set of Flickr data [18]. Only user ID is shown in Qstack prototype instead of other real information due to the anonymous characteristics of the records.

4. Case Study

This section demonstrates the ability of Qstack via two use cases of Flickr photo award allocation. We assume that Emma, a president of a photo association on Flickr, would perform annual award allocation to the members. The allocation criteria are mainly about the tag values which a member tagged in his or her collection in term of overall and specific contributions.

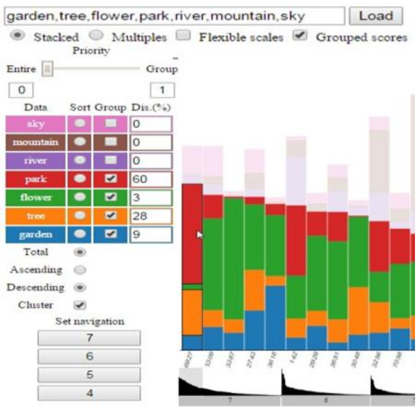

Fig. 5. Overall contribution ranking for price *A*.

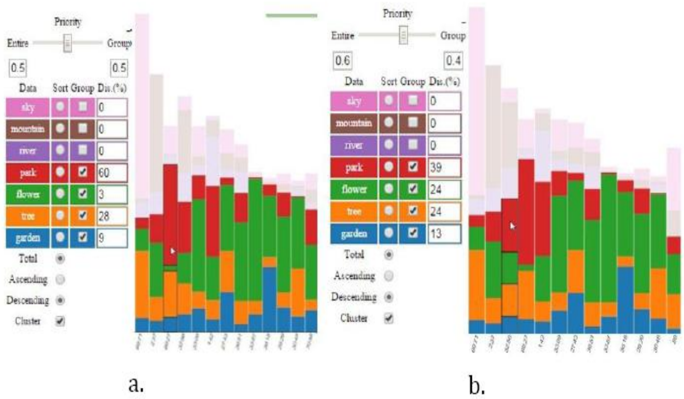

Fig. 6. Overall contribution ranking for price *B*.

4.1. Overall Contribution Rankings

This use case demonstrates the ability of Qstack for interactive rankings of multi-tags based on grouped-score combination. For overall contribution award, there are two types of prices including Best grouped-topic contribution *A* and Best overall contribution *B*. While price *A* considers grouped-topics with a highest priority, price *B* takes both grouped-topics and overall contribution into account. Eligible candidates are the members who had the contribution to all of considered tags. Firstly, Emma enters seven considered tags of the year into the Qstack for data load including *garden, tree, flower, park, river, mountain* and *sky*. There are two grouped topics: *garden, tree, flower, park (T*1*)* and *river, mountain, sky (T*2*)*. For *T*1 of price *A*, once reviewing the tag value distribution, she uses the query panel to visually combine *garden, tree, flower* and *park* and ranks the members against this group with a highest priority. She enables grouped-score calculation, clustering and descending order to make the view focused on the cluster where the members satisfy all seven tags. Emma customizes brushing area to zoom in on the cluster and takes note three highest-ranked members *6927, 3309* and *3387* for *T1* (Fig. 5). Then, she goes through the similar tasks with *river, mountain* and *sky* for *T*2 allocation.

About price *B*, the final score of each member is the combination of the score on entire contribution and a grouped topic. At the beginning, Emma makes it simple by giving the same priority (0.5) on the entire tags and grouped tags for *T*1 (Fig. 6.a). The update shows that three highest-ranked members involve *6927* who received a price *A*. Emma carefully watches the visual pattern to analyze the scores, and she recognizes the score of *6927* on entire contribution is lower than a few others. To encourage more members to access the prices, Emma adjusts a little higher priority (0.6) on entire contribution than the grouped topic (0.4) (Fig. 6.b). The result, then, includes member *3256* instead of *6927*. She takes note the result and does a similar procedure for *T*2 allocation.

4.2. Topic Contribution Rankings

This use case demonstrates the ability of Qstack for flexible rankings of multi-tags across categorized alternatives. Topic contribution awards are evaluated for the members who at least toke four of seven considered tags (over fifty percent) of the year, are not eligible for overall contribution prices and had the highest score of tag values for single topics. In order to filter the eligible members for the price, Emma makes brushing area cover all categories 4, 5 and *6* (Fig. 8). Once the members are navigated, she uses multi-baseline and flexible scale mode to make the view suitable for separate topic comparison. Emma adjusts query options to sort the result by each topic tag in descending order and selects the highest ones for the awards (Fig. 7).

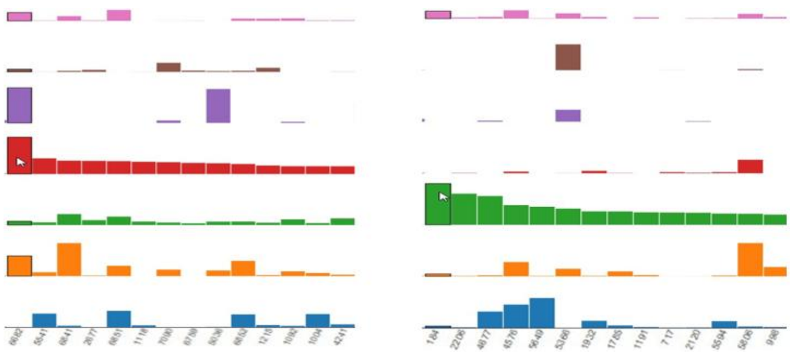

Fig. 7. Sorting by tag *park* (red bars) and tag *flower* (green bars) with multi-baselines and flexible scales.

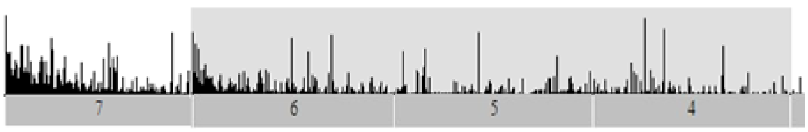

Fig. 8. Brushing across categories *4, 5, 6*.

5. Evaluation

Our goal is to measure the effectiveness of Qstack towards the tasks of comparing and ranking multi-tags based on grouped-score combination within and across categorized alternatives. As discussed in the related work section, currently, there is no visualization tool with similar purposes of Qstack; therefor, an existing tool comparison study is impossible to be carried at the presence. In term of generalization of Qstack approach, multi-attribute visualization tools including Excel, Tableau and LineUp are taken into consideration. Although Excel and Tableau are powerful tools for regular data visualization on diverse purposes, to access solutions for the addressed problem, much time, programming knowledge and tool usage experience are required [3]. Those technical activities make such tools inappropriate for casual users or novices to solve such problems. Of current studies, LineUp is a task-driven visual technique which is the closest to the generalized purpose of Qstack. Similar to the target of LineUp, Qstack is designed for novices to deal with multi-attribute comparison by simple tasks without much technical requirement. Although LineUp well supports attribute combination and rank change tracking, their approach has not considered the rankings using grouped-score combination for categorized alternatives. Due to all the reasons above, we conducted a qualitative study on Qstack usage in order to assess its usefulness instead of comparing to other techniques.

5.1. Study Setup and Procedure

We recruited six participants. Four of them are in computer science, and two is from engineering. Based on an informal survey before the experiment, it is known they all have Excel experience on basic tasks, but none of them has worked with Tableau before. The participants are familiar with stacked bar charts, but they are novices in dynamic queries and visual interaction on graphic charts. About Qstack usage, we designed nine tasks including all primary features of the tool as three types of tasks bellows: *Type 1* for data loading and basic comparison and ranking, *Type 2* for multi-tag rankings using grouped-scored combination and *Type 3* for multi-tag rankings across categorized alternatives.

For receiving feedbacks from users, a questionnaire consisting of twelve questions was constructed and based on Likert scale system with the range of *strongly disagree* and *strongly agree* in domain from 1 to 5. In addition, there was an open question for users to comment benefits and drawbacks for the tool improvement. After given a Qstack introduction and tutorial, the participants freely asked and discussed to ensure they all understood the detail of the task requirements. Once completing all the tasks following the former instruction, they used the questionnaire to express their feedbacks.

5.2. Result

Generally, most of the participants successfully and easily completed all the requests, and two participants took a short interruption. The reason of interruption is in type-3 tasks because they customized the brushing rectangle to incorrect focus areas of categorized alternatives.

The qualitative feedbacks indicated that all the participants felt confident in using the tool to solve addressed problems. The result is quite positive with 4.7(mean) for ease of use, 4.5 for usefulness and 4.2 for interesting visualization. The open comments showed constructive opinions for drawbacks regarding to zooming and animation. The vertical zooming should be equipped for stacked display mode, which would reduce user efforts to recognize the patterns in case of very small tag values. Besides, although visual patterns were updated in real time, at a few times the animation was little delayed.

6. Conclusion

This paper has introduced Qstack, a novel visualization approach for multi-tag rankings based on grouped-score combination. The approach is to organize a series of stacked bars in categories, display multi-tags in a double layer and using brushing feature for navigation on focus alternatives. This method enables users to easily analyze single tag values, grouped scores and total rankings within or across categories. Additionally, by interaction on the dynamic query panel, users are able to drive the rankings without much technical knowledge requirement. A case study on photo award allocation has been presented for Qstack demonstration, and according to the qualitative study result, Qstack is a useful tool for multi-tag rankings.

Although the initial purpose of Qstack is for tag data application, we believe our visualization approach might be generalized to multi-attribute or multi-criteria rankings for general analytic data. In future, we will improve the data mapping module in order to handle various input data and feature other navigation components for enhancing Qstack benefits. Further experiments will be conducted to test the suitability of Qstack for different kinds of data.

References

[1] Flickr. Retrieved, from https://www.flickr.com/

[2] Library Thing. Retrieved, from https://www.librarything.com/

[3] Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., & Streit, M. (2013). Lineup: Visual analysis of multi Attribute rankings. *Visualization and Computer Graphics, IEEE Transactions*, *19(12),* 2277-2286.

[4] Clark, J. (2008). Clustered Word Clouds. Retrieved January 30, 2009, from http*://www.neoformix.com/2008/ClusteredWordClouds.html.*

[5] Lohmann, S., Ziegler, J., & Tetzlaff, L. (2009). Comparison of tag cloud layouts: Task-related performance and visual exploration. *Human-Computer Interaction*.

[6] Cui, W., Wu, Y., Liu, S., Wei, F., Shou, M., & Qu, H. (2010). Context preserving dynamic word cloud visualization. *Proceedings of the IEEE Pacific Visualization Symposium on PacificVis* (pp.121-128).

[7] Lee, B., Riche, N. -H., Karlson, A. K., & Carpendale, S. (2010). Sparkclouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics, 16(6)*, 1182-1189.

[8] Collins, C., Viegas, F., & Wattenberg, M. (2009). Parallel tag clouds to explore and analyze faceted text corpora. *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology-VAST* (pp. 91-98).

[9] Di Caro, L., Candan, K. S., & Sapino, M. L. (2011). Navigating within news collections using tag-flakes. *Journal of Visual Languages and Computing, 22(2),* 120-139.

[10] Begelman, G., Keller, P., & Smadja, F. (2006). Automated tag clustering: Improving search and exploration in the tag space. *Proceedings of the Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland* (pp. 15-33).

[11] Hassan-Montero, Y., & Herrero-Solana, V. (2006). Improving tag-clouds as visual information retrieval interfaces. *Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies, Citeseer* (pp. 25-28).

[12] Zubiaga, A., Garcia-Plaza, A. P., Fresno, V., & Martínez, R. (2009). Content-based clustering for tag cloud visualization. *Proceedings of the International Conference on Advances in Social Network Analysis and Mining* (pp. 316-319).

[13] Dork, M., Carpendale, S., Collins, C., & Williamson, C. (2008). Visgets: Coordinated visualizations for web-based information exploration and discovery. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics, 14(6),* 1205-1212.

[14] Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, 8(1),* 9-20.

[15] Liu, S., Zhou, M., Pan, S., Qian, W., Cai, W., & Lian, X. (2009). Interactive, topic-based visual text summarization and analysis. *Proceedings of the 18th ACM Conference on Information and knowledge Management, ACM.*

[16] Shi, C., Weiwei, C., Shixia, L., Panpan, X., Wei, C., & Huamin, Q. (2012). RankExplorer: Visualization of ranking changes in large time series data. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*.

[17] Rao, R., & S. K. Card (1994). The table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.*

[18] Plangprasopchok, A., Lerman, K., & Getoor, L. (2010). Growing a tree in the forest: Constructing folksonomies by integrating structured metadata. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Phi Giang Pham was born in Hanoi, Vietnam. He received his bachelor degree in informatics from Cantho University, Vietnam in 2006. He obtained a MSc in information management from Asian Institute of Technology, Thailand in 2008. His research interests are human-computer interaction, information visualization and information system development.

Mao Lin Hoang is an associate professor in Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He is an active researcher and focusing on big-data visual analytics, visual network intrusion detection, information visualization and social network visualization.