TagClusters: Semantic Aggregation of Collaborative Tags beyond TagClouds

Ya-Xi Chen[1], Rodrigo Santamaría[2], Andreas Butz[1], Roberto Therón[2]
[1] *Media Informatics, University of Munich, Germany*
[2] *Department of Informatics and Automatics, University of Salamanca, Spain*

ABSTRACT
TagClouds are a popular visualization for collaboratively generated tags. However, they have some distinct problems such as linguistic issues, high semantic density and poor communication of hierarchical structure and semantic relationships among tags. In this paper we investigate ways to support semantic understanding of collaboratively generated tags beyond TagClouds. Following the results of a survey, we propose an improved visualization named TagClusters. Based on a semantic analysis, tags are clustered into different semantic groups. Their visual distance depends on the semantic similarity between tags, and thus the visualization offers a better semantic understanding of collaboratively generated tags. We conducted a comparative evaluation with TagClouds and TagClusters based on the same tag set. We received overall positive feedback on TagClusters and the results indicate that it has advantages in supporting efficient browsing, searching, impression formation and matching. In our future work, we will explore the possibilities of tag recommendation and tag-based Information Retrieval based on TagClusters.

Keywords: Improvement of TagClouds, collaborative tagging, user-contributed tags, visualization of tags, semantic analysis.

INTRODUCTION
With the rapid growth of the next-generation Web, many websites allow their users to make contributions by tagging digital items. This collaborative tagging has become a fashion on many websites, the most representative of which are the social bookmarking site Del.icio.us (http://delicious.com/), the photo sharing site Flickr (http://www.flickr.com/) and the music community Last.fm (http://www.last.fm). The low technical barrier and easy usage of tagging have attracted millions of users. The tags contributed by these users are not only an effective way to facilitate personal organization, but also allow searching or discovering new information.

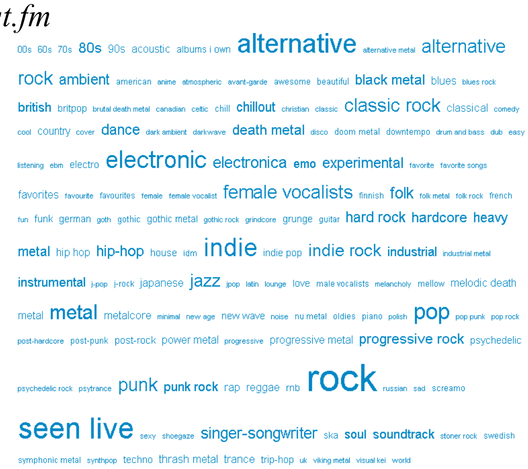
Currently, there are two main methods being used for tag-based music retrieval. The first category is keyword-based search, which is the most common method to seek information on the Web. The system will simply return all information related to the respective keyword. The second one is a visualization-based method called TagClouds (as shown in Figure 1). Due to their easy understandability and aesthetical presentation, TagClouds have become a fashion on many websites. However, they still have some intrinsic disadvantages and many researchers have been dedicated to improve their aesthetical presentation or provide a better semantic understanding. In this paper, we take Last.fm as our experiment platform and explore how to improve the visualization of tags in this website. We are specifically interested in the semantic understanding of structures and relationships among tags, which might lead to more successful tag recommendation and visualization-based music retrieval.

In the remaining sections of this paper, TagClouds and their related work will first be discussed. Based on a survey among Last.fm users, some possibilities to improve semantic understanding beyond TagClouds will be explored. Finally, a prototype named TagClusters and its evaluation will be presented. The discussions on the evaluation results and future work will conclude the paper.

TAGCLOUDS

TagClouds are a visual presentation of the most popular tags, in which tags are usually displayed in alphabetical order and attributes of the text such as font size, weight or color are used to represent certain features, for example, font size for prevalence and color brightness for recency. As a result of the collaborative tagging, the popular tags shown in TagClouds have a more accurate meaning than those assigned by a single person. The visualization draws attention to more important items and thus provides an overall impression of popularity, reflecting the general interests among a broad demography (Viégas & Wattenberg, 2008), (Hearst 2008), (Hearst & Rosner, 2008). With TagClouds, a keyword-based search can be conducted by selecting a tag as input.

Figure 1. TagClouds in Last.fm



A lot of existing work focuses on the factors influencing the visualization quality of TagClouds. Halvey and Keane (2007) assessed tag presentation by evaluating different factors which ease the finding of tags. They discovered that font size, position and alphabetization can aid to find information more easily and quickly. Similar conclusions can be found in (Golder & Huberman, 2006). An interesting discovery of (Rivadeneira, Gruen, Muller & Millen, 2007) was, that large fonts are not necessarily to be found easily.

LIMITATIONS OF TAGCLOUDS

As revealed in (Golder & Huberman, 2006), after 100 users indexed the same resource, each tag’s frequency is a nearly fixed proportion of the total frequency of all tags assigned to the resource. This stability was also verified in (Halpin, Robu & Shepherd, 2007). These results, can, however, not entirely be transferred to uncontrolled collaborative tagging. Although there is also a relative stability in collaborative tagging, because of its free nature there are intrinsic problems with inevitable noise and redundancies (Li, Bao, Yu, Fei & Su, 2007), (Hassan & Herrero, 2006).

Linguistic problems with free tagging

Nielsen (2007) found that different educational and cultural background might lead to tag inconsistency, which was also mentioned in (Begelman, Keller & Smadja, 2006). Specifically, there are two common problems with free tagging systems, which are difficult to avoid from the user's side: synonymy and ambiguity. Synonymy is also defined as "inter-indexer inconsistency" (Nielsen, 2007) and it happens when different indexers use different terms to describe the same item. Ambiguity means that one term may have several different meanings (Mathes, 2008), which will bring noise into the retrieval results. Although social collaborative tagging could alleviate these problems, as pointed out in (Wu, Zhang & Yu, 2006), such problems still widely exist.

High semantic density

As discussed in (Hassan & Herrero, 2006), (Begelman, Keller & Smadja, 2006), if visible tags are selected only by their usage frequency, there might be a problem of high semantic density, which means that very few topics and related prominent tags tend to dominate the whole visualization and less important items fade out (Hearst & Rosner, 2008). Therefore, a more reasonable selection method should be designed.

Poor understandability of structure and relation

It was claimed in (Hassan & Herrero, 2006) that the alphabetical arrangements neither facilitate visual scanning nor provide semantic relations between tags. In the evaluation of (Hearst & Rosner, 2008) a significant proportion of interviewees did not even realize that tag clouds are regularly organized alphabetically. They also discovered that the users had difficulties to compare tags in small print and to derive semantic relations. Spatial proximity was also misinterpreted as a semantic relation between tags. Therefore, TagClouds do not seem very suitable for conveying structure and relations.

IMPROVEMENTS OF TAGCLOUDS

Recent research mainly investigates improvements of TagClouds to enhace their aesthetic appearance and their semantic understanding.

Aesthetic enhancements

Since several factors influence the effectiveness of a TagCloud visualization, some systems have already allowed the user to adjust these parameters. One representative of these systems is PubCloud (Kuo, Hentrich, Good & Wilkinson, 2007).

Tight coupling (Ahlberg & Shneiderman, 1994) improved the quality of TagClouds by introducing a spatial algorithm to pack the text in the visualization tighter. Kaser and Lemire (2007) used Electronic Design Automation (EDA) to improve the display of TagClouds by avoiding large unused white space. A series of algorithms was proposed in (Seifert, Kump & Kienreich, 2008) which can display tags in arbitrary convex polygons with a dynamically adapting font size. A circular layout was developed in (Bielenberg & Zacher, 2005), in which both font size and the distance to the center represent tag importance. It was claimed in (Lohmann, Ziegler & Tetzlaff, 2009) that sequential layout with alphabetical sorting is suitable for finding a specific tag, while a thematically clustered layout can facilitate finding tags which belong to a certain topic.

Supporting semantic understanding

In (Shaw, 2008) tags were visualized as a graph, in which length of edge represents similarity. TagOrbitals (Kerr, 2008) presented related tags and summary information in an atom metaphor, in which each primary tag was placed in the center, and other related tags were placed in surrounding bands. The main problem with this visualization is the varying orientation of texts.

Clustering algorithms were applied to gather semantically similar tags. In (Hassan & Herrero, 2006) the k-means algorithm was applied to group semantically similar tags. Similar work can be found in (Provost, 2008). Li, Bao, Yu, Fei and Su (2007) supported browsing large scale social annotations based on an analysis of semantic and hierarchical relations. The user's profile and the temporal dimension can be integrated for personalized or time-related browsing (Nielsen, 2007).

Most of the methods discussed above are static visualizations and lack interaction. Furthermore, the low level sub-structures should be deeper explored which will help to form a better understanding of hierarchical structure and relations.

PRACTICAL TAG USAGE IN LAST.FM

As discussed above, TagClouds have distinct linguistic problems and poorly support the understanding of structure and relationships among tags. In this paper, we have chosen the tags in Last.fm as the experimental source. Last.fm allows users to tag each track, album and artist with free form texts, which can then be used for tag-based visualizations and search. Last.fm offers TagClouds for the visualization of the top tags generated by users and most of these popular tags are genre-related (see Figure 1). Based on these collaboratively generated tags, the user can conduct tag-based search and Last.fm will return a page with the retrieval results of the respective tag, in which related tags and artists will be displayed. For example, in the retrieval result for the tag "*rock*", the user can see the related tags such as *alternative*, *indie* and *punk*.

In order to obtain a better understanding of the practical tag usage in Last.fm, we conducted a user interview followed by an online survey. Based on these, we derived a number of design guidelines, which are essential for the development of our semantic-based tag visualization.

User interview

In order to gain more insights on the effective use of tags contributed by the users, we need to explore the features and benefits of music-related tags, as well as the users' search and tagging behavior. In order to explore these, we conducted interviews with 13 Last.fm users, who were recruited in the Last.fm online forums. Three of them were female and 10 male, and their age varied from 18 to 26 with an average age of 23 years. Among these participants, there were 7 undergraduate students, 2 graduate students, 2 researchers and 2 employees in business. All the participants had a common knowledge about computers and the Internet. They had a relatively long listening history: all of them registered more than 2 years ago, 2 users even more than 3 years and one user more than 4 years ago.

During the interview, the participants were equipped with a PC, a keyboard and a mouse. They were given a list of all the questions to be discussed. While they were formulating their answers, the Think-Aloud protocol was applied, which helped them to express their answers in a more natural and flexible way. Participants could also freely browse their personal profiles and other services of Last.fm. The key aspects we explored were the tagging and searching behavior and relevant user-generated tags. Example questions for searching were: "How often do you search for music in Last.fm?", "How often do you use tags for searching?" and "What do you think about the TagClouds of Last.fm?". About the tagging behavior, some example questions were: "How often do you tag music in Last.fm?", "Which kind of tags do you use for tagging?" and "Do you think tagging music is difficult?".

Most of the participants used the search functionality frequently, with the exception of one, who found music by browsing the charts of popular artists. Besides the standard keywords such as the name of the artist, album and song, tags were less used for searching and the average score for their usage frequency was 2.2 (1 for never, 5 for daily). The top three types of tags used for search were genre, mood and artist biography. These tags represent quite diverse aspects, but currently in Last.fm, the user cannot combine multiple tags for more specific searching.

User 3: It is a pity that I cannot use more than one tag as keywords. So it is hard to find a tiny part between punk and indie electronic.

All participants felt that the too general tags might make the user getting lost among too many results and thus find nothing specific.

User 5: Tags are too subjective and heavily depend on the personal musical taste. They cannot represent the essence of music well enough. For example, others might feel awful with your favorite song.

User 12: "seen live" (the second top tag) doesn't help me at all. It's like asking for the way to the Eiffel Tower and someone tells you "in Europe".

When asked for comments about the top tags shown in Figure 1, one prominent comment was their redundancy. It might be caused by different reasons. For example, the language variation between British and U.S. English leads to the different tags for *favorite* and *favourite*. Since music is difficult to express verbally and there is no standard category for genre, people have different definitions of genres and even have different understandings of the same genre, which leads to remarkable redundancy and even errors with genre-related tags.

User 4: I noticed that some people think IDM (Intelligent Dance Music) and electronic are the same so these two tags always appear in a pair. But actually they are different.

The participants did not tag very often themselves, and the average tagging frequency was 1.1 (1 for never). Similarly to the description of personal musical taste and tags used for search, most of their actual input tags were also related to genre, mood and artist biography. Some other participants also use personalized tags for quick relocating, such as "listen again" and "Sunday morning". The majority of participants thought that tagging music is hard and one participant even postponed it as the last task because he thought it was a hard and serious job and he needed more time to think about it.

User 1: someone said that talking about music is just like dancing with a poem. It is hard to describe music with words.

Online survey

In order to verify the results of our interviews, we conducted an online survey with a larger population. The survey lasted for two months and 228 Last.fm users joined in it, 93 of which were female and 133 male (two gender identifiers were left blank). Their age varied from 12 to 54 with an average age of 22 years. The participants were mainly students and employees from North America and Europe. Participants reported themselves to be enthusiastic about music with an average score of 3.8 (1 for not at all, 5 for very much). Their registration histories varied from 0.13 to 5.72 years with an average duration of 1.8 years.

The questions mainly covered the general experience with search and tagging. Participants looked for information quite often in Last.fm with an average score of 4.0 (1 for never, 5 for daily), and a significantly larger proportion of this time (with an average percentage of 64%) is spent in browsing-like filtering rather than specific search activities. Concerning keyword-based search, participants mostly search music-related information (4.0/5), such as names of artists, albums, songs, and less about groups, users and events. Consistently with the interview result, tags are used less for search (2.0/5).

The same Last.fm TagCloud (as shown in Figure 1) was offered and participants thought that it offered an efficient overview for the most popular items (3.6/5), but they also noticed the apparent linguistic problems. They commented that a too general selection of tags is counterproductive for tag-based search.

Participants did not tag very often (1.8/5) and they mainly tagged music in their own libraries. In the questionnaire, they were required to tag the same artist and the top tags were genre-related, such as *pop*, *funk* and *soul*. The top motivations for tagging were easing browsing and searching, facilitating personal organization, and helping to understand music.

Implications

Based on the results of the interview and the online survey, we derived some design guidelines regarding the users' searching and tagging behavior. People's information seeking behavior is more alike to browsing than to specific keyword-based search. People rarely tag music. Some people take music very seriously and want others to know more about their favorite music through tags. Some users annotate music with special tags for personal use. Others simply enjoy making a contribution or offering knowledge by tagging.

In Last.fm, most of the top tags are related to genre, mood or artist biography. There is less chance for users to be 'educated' since the personal understanding of genre and emotion is subjective and according to different musical experiences, users might come up with different tags for the same music. Therefore, search by tags is not common in Last.fm because freely entered tags are normally too general to help users narrowing down the retrieval results. More neat and organized tags with less redundancy would be more useful and the option of combining multiple tags in the searching process might help the user to harness the searching direction.

We believe that if all the tags are organized in a more understandable semantic way, they will be more helpful for tag recommendation and tag-based music retrieval. We therefore explored the semantic aggregation of tags to support efficient hierarchical browsing and understanding the relations between tags.

Semantic aggregation

Based on the text analysis, the synonym issue can be resolved by grouping semantically similar tags into one cluster, for example, *favorite* and *favorites*, *rock and roll* and *rock n roll*. The semantic aggregation also helps to alleviate the problem of ambiguity. For example, fewer users know that "*electronic*" and "*IDM*" represent roughly the same genre. By aggregating tags in the visualization, the users can see that these two tags are grouped into the same cluster, suggesting that they have roughly the same meaning. This is also an efficient way to help the users gain some music knowledge.

Hierarchy exploration and relation visualization

We also explored the implicit hierarchical structure hidden inside the freely formulated tags. By unraveling the structure it is possible to support a better understanding of tags, especially genre-related categories. In a top-down manner, the user can search more specifically with less ambiguity problems. From highlighting of overlapping areas, the user can tell the relation between tags at a higher semantic level.

Possible applications

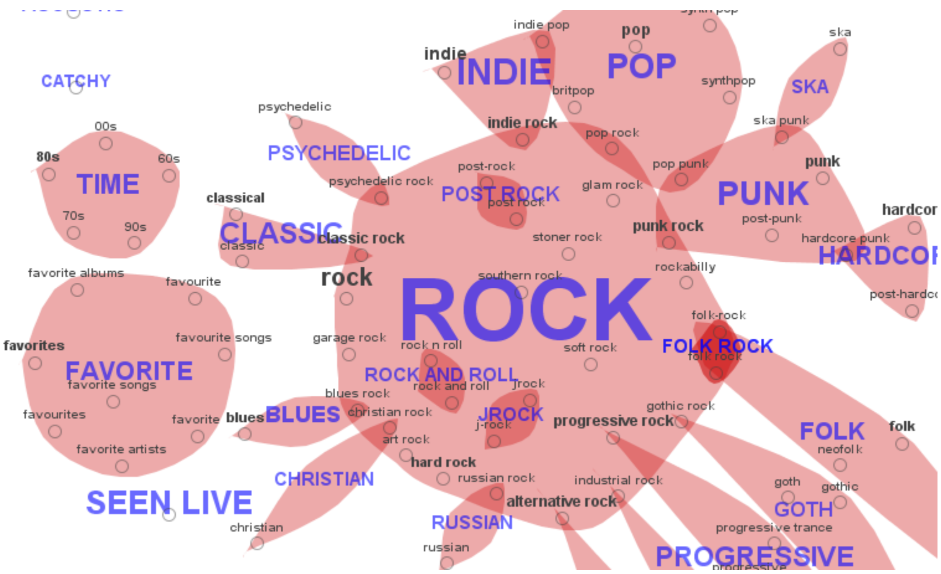
Based on the hierarchical structure, potential usages include tag recommendation and tag-based music retrieval. Once the hierarchical structure of tags is derived, the user can get useful tag suggestions while effectively avoiding spelling errors and redundancy. Genre is one of the most common criteria for music organization and retrieval (Cunningham, Bainbridge & Falconer, 2006), however, there is no standard definition of genres. Thus, system recommendations for genre-related tags might be convenient. For example, if the user types in "*electronic*", the system could prompt possible relevant tags such as "*IDM*". When semantically similar tags are grouped hierarchically, this also facilitates a more controlled tag-based search and the system can return richer retrieval results.

TAGCLUSTERS: SUPPORTING THE SEMANTIC UNDERSTANDING OF TAGS

We investigated the problems with user-generated tags in Last.fm and explored possibilities to improve the semantic understanding of tags. Based on the aforementioned design guidelines, we built a prototype named TagClusters to assist the semantic understanding of tags (as shown in Figure 2).

TagClusters is implemented based on Overlapper (Santamaría & Therón, 2008), a graph-based visualization tool that highlighs the connections and overlaps among entities (nodes) in data. TagClusters is an interactive interface in which tags are drawn as labels with different sizes (size representing tag popularity, as in TagClouds), and tag groups are drawn as transparent colored areas (see Figure 2). TagClusters use the underlying visualization of Overlapper, based on a force-directed layout (Fruchterman & Reinhold, 1991) that does not use a typical node-link approach, but a Venn-diagram approach, in order to represent groups and group relationships. On a force-directed layout, two forces are typically computed within an iteration loop: a spring force is applied to connected nodes, keeping them close, while a repulsion force is applied among nodes, separating non-connected nodes. In our case, each node in the visualization corresponds to a tag, and two tags clustered in the same group are connected by an edge. In this schema, each group is a complete sub-graph joining every node in it, and it is represented by a transparent hull that wraps all of its nodes. This is achieved by using the outermost nodes of the groups as anchoring points for a spline curve which is drawn in a solid color to improve group traceability. Edges are only used to compute attraction forces among nodes in the same groups, but not drawn to avoid edge cluttering (Gansner & North, 1998).

Figure 2. TagClusters visualization with the same tag set in Figure 1.



The initial placement of tags is not random, but coherent with tag co-occurrence (which will be discussed in the next section). This placement reduces the time required for the stability of forces in the layout, also minimizing possible misplacements. Moreover, edges are weighted according to the co-occurrence between the tags they connect, so the more similar they are, the higher the attraction force represented by the edge. In this way, nodes will be separated proportionally to their similarities even if they are in the same group. Finally, as in TagClouds, the label size is proportional to the number of occurrences of the tag. Group labels are drawn using capital letters and with a different color to distinguish them from tag labels. Group label size is proportional to the sum of the occurrences of all the tags the group contains. This is done on a logarithmic scale to avoid very large tags. Therefore, TagClusters can be seen as TagClouds in which position is relevant and based on tag co-occurrence and tag groups. The visualization is further supported by group wrappers. These characteristics exploit human perception for traceability and group detection, improving the visual analysis.

In addition, the final user interface offers several interactions, such as panning and zooming without losing context (provided by an overview in the top-right corner), hiding or showing tags, groups and labels, searching tags by text, modifying the underlying forces, modifying node and group locations, changing color settings, and exporting the current visualization to an image. The system also provides multiple options for tag selection, which facilitate tag-based music retrieval. For example, the user can choose multiple tags or groups using both keyboard and mouse. He/she can also draw a shape manually and all the tags included in this shape will be selected. As pointed out in (Hearst & Rosner, 2008), it is difficult to compare or discover tags in small print, so our interface also allows customized label font sizes. Although infrequently, the force-directed technique may misplace nodes, positioning them inside groups they do not belong to (this may be the case, for example, for elongated hulls connecting two other groups, such as the *goth* and *folk* groups in Figure 2). This particular issue is resolved by allowing the user to hover over a node. This highlights all nodes connected to the respective node, thereby resolving any possible ambiguity. The user can also manually adjust the positions of groups and tags, and relevant changes can be saved automatically.

Since most of the popular tags are genre-related and the relevant groups overlap each other, these genre-related groups are placed in the center and other semantically less related tags and groups are scattered in the periphery. Within a group, related tags will be further grouped into sub-groups. Since groups are represented as transparent colored areas, overlapping groups (groups that share one or more common tags) will have intersecting, more opaque areas, thus highlighting the overlapping tags. With such a visualization, the user can obtain a better understanding of structure and relation between tags. For example, in Figure 2, we can see that *rock* is the most relevant category and related to several others, such as *pop*, *indie* and *punk*. Also we find several genres at the bottom right of the figure relating to both *metal* and *rock* groups, such as *progressive* and *goth* (see Figures 2 and 3).

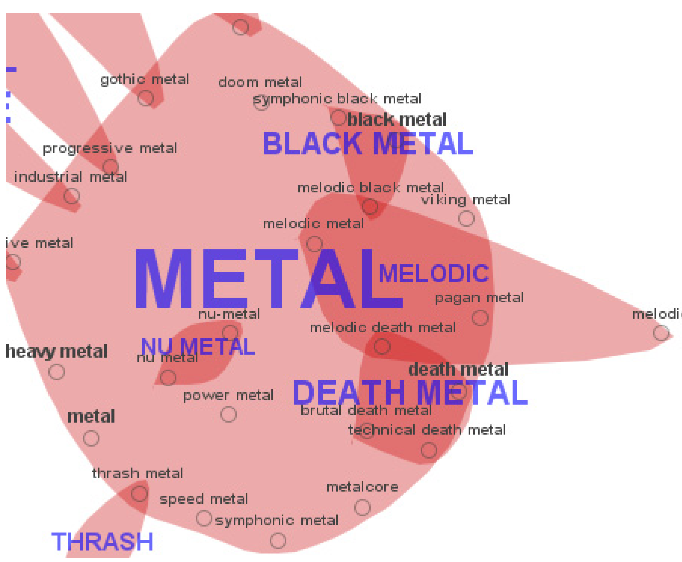
UNDERLYING SEMANTIC ANALYSIS

The organization of tags in TagClusters is based on a semantic analysis which determines the structure and position for each tag. First we apply text analysis to create the hierarchical structure while excluding redundancy. Then we determine the initial position for each tag based on the calculation of semantic similarity.

Text analysis based clustering

After an observation of the existing TagClouds (as shown in Figure 1), we found that synonymy is the most prominent problem with tags. This is mainly caused by single/plural such as "*female vocalist*" and "*female vocalists*" or different spellings, such as "*favorite*" and "*favourite*". Besides, people tend to add different separations between the same words, for example, "*post-rock*" and "*post rock*", or "*rock and roll*" and "*rock n roll*".

Figure 3. Examples of text analysis result



In Last.fm, many tags, but especially the genre-related tags share a characteristic property: The tag at the lower semantic level almost always contains the tag at the higher level and the length of tag is roughly proportional to its semantic level, for example, "death metal" and "brutal death metal". This property helped us to derive the hierarchical structure.

In our system, after removal of different separators, such as "_" and "&", the Porter algorithm (Porter, 2006) is applied to detect the stem for each tag. Tags with the same stemmed words will be clustered into the same group. Within one group, tags with similar semantic meanings will further be clustered into sub-groups. For example, all the tags containing "*metal*" will be grouped in the *Metal* group and related tags such as "*death metal*" and "*brutal death metal*" will be further placed into a sub-group (as shown in Figure 3). All the tags related to gender will be clustered in a *Vocal* group, and a similar thing is done for the time-related tags, such as "*80s*" and "*00s*" (see left part of Figure 2). After the text analysis, tags can be effectively grouped into relevant clusters. We also found that this basic technique has some limitations and should be further enhanced to distinguish the literally similar but musically different tags such as *classic* and *classic rock*.

Calculation of semantic similarity

After deriving the hierarchical structure of tags, the semantic similarity among tags is calculated based on their co-occurrence, which is widely used in the field of music retrieval to determine the semantic relationship between information items (Nielsen, 2007), (Begelman, Keller & Smadja, 2006). In our case, this semantic similarity (*Sim*) between tag A and B equals to the ratio between the number of resources in which these two tags co-occur and the number of resources in which any of the two tags appears (Nielsen, 2007), (Hassan & Herrero, 2006), as equation 1 shows.

$$Sim(A,B) = |A \cap B| \,/\, |A \cup B| \qquad (1)$$

With this semantic analysis, all the tags will be well organized: the initial location of each tag is assigned by means of a 2D projection based on a non-metric multidimensional scaling of co-occurrences (Sammon, 1969). The genre-related tags, which might be the most useful category for tag-based searching, become prominent in the visualization. Other categories such as time- or emotion-related categories are scattered because of the less semantic relationship with the genre category. Instead of excluding them from the visualization, these categories still remain visible and can be inspected by browsing or keyword-based search.

USER STUDY

In order to evaluate our visualization tool, we conducted a comparative evaluation of TagClouds and TagClusters. We were specifically interested in the performance of our tool in supporting the perception of semantic relationships among tags.

User study design

TagClouds and TagClusters were evaluated using a repeated measures within participants factorial design. The independent variable was sysType (TagClouds and TagClusters). The order of sysType was counterbalanced between participants to minimize learning effects.

Participants

We recruited 12 participants at the University of Munich with different majors, 7 German and 5 foreigners, 4 female and 8 male. These 12 participants allowed a perfect counterbalance of sysType to minimize learning effects. Their age varied from 24 to 29 with an average age of 27 years. All the participants had a common knowledge about computers and the Internet. Participants reported themselves to be familiar with TagClouds with an average score of 3.58 (1 for unfamiliar at all, 5 for very familiar).

Procedure

The user study consisted of a pre-questionnaire, an interview and a post-questionnaire. Participants were first asked to fill out a pre-questionnaire with demographic information and their general experience with tags. After a brief introduction of TagClouds and TagClusters, participants were asked to execute 6 tasks concerning searching, browsing, impression formation and matching. Each task consisted of two similar sub-tasks and the following is a brief description of the representative tasks:

Task 1: Locate one single item: Find a tag named "german".
Task 2: Tag sorting: List the top 5 popular tags.
Task 3: Tag comparison and filtering: List the top 5 genre-related tags.
Task 4: Derive group structure: Give a hierarchical structure for the *Metal*-related tags.
Task 5: Find relation between tags: Is there an overlap between *Indie* and *Classic*?
Task 6: Judge the tag similarity: Is Alternative more similar to *Rock* or *Electro*?

After completing each task, participants scored how easy the task was, and how helpful the system was in supporting the task. After completing all the tasks, participants filled out a post-questionnaire concerning the overall impression of both sytems. The study was conducted in our lab and participants were equipped with a PC, a keyboard and a mouse. On average the user study lasted about 35 minutes per participant. It was recorded on video and the Think-Aloud protocol was applied.

Hypotheses

Based on the main features of both visualizations (for example, tags are ordered alphabetically in TagClouds and semantically grouped in TagClusters), the following hypotheses were stated:

H1: TagClouds will outperform TagClusters in Task 1 regarding completion time, answer precision, task easiness and system usefulness.

H2: TagClouds will outperform TagClusters in Task 2 regarding completion time, answer precision, task easiness and system usefulness.

H3: TagClusters will outperform TagClouds in Task 3 regarding completion time, answer precision, task easiness and system usefulness.

H4: TagClusters will outperform TagClouds in Task 4-6 regarding completion time, answer precision, task easiness and system usefulness.

Results

The results are reported based on the 24 sessions performed by the participants. We analyzed the questionnaires, answers and the screen activities and discovered the following results. The comparison of the completion time, the answer precision, the task easiness and the system usefulness for both systems are shown in Figure 4.

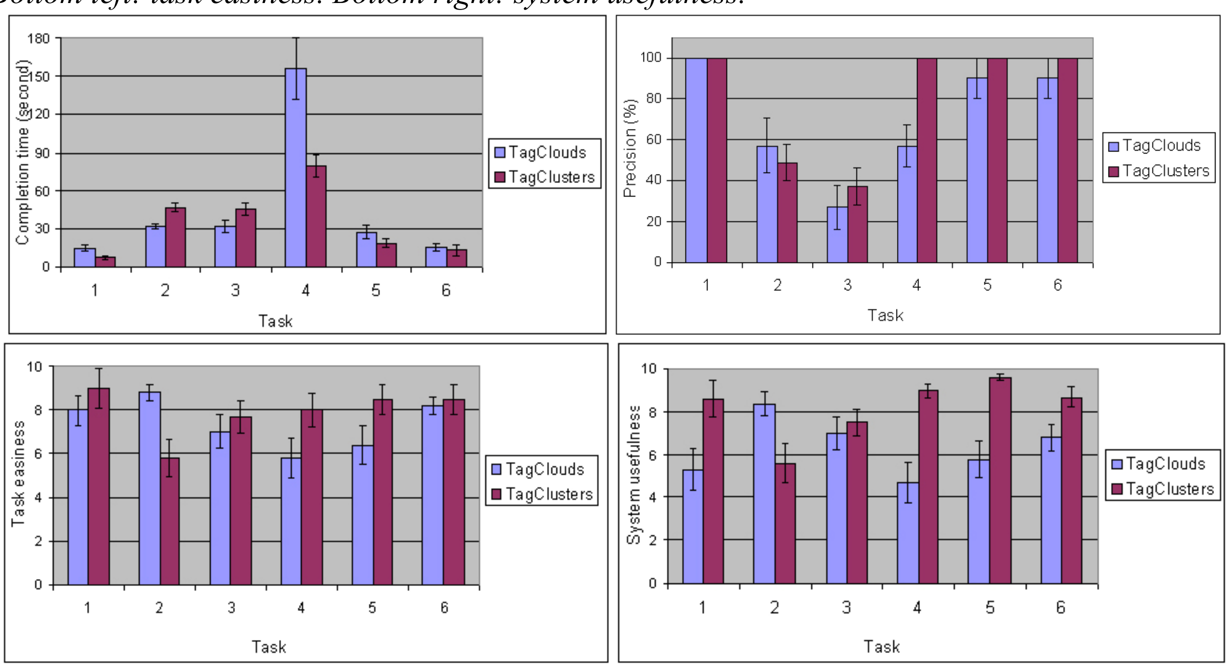
For task 1, although the tags in TagClouds can be located by their alphabetical order, locating the first character still needed some time. A dependent t-test showed that TagClouds cost significantly more time with task 1 than TagClusters (t(11)=3.131, p<0.05). Furthermore, 25% of the participants did not realize that TagClouds are ordered alphabetically, thus they spent more time on tag locating. Participants claimed that the search functionality in TagClusters helped to locate the item quickly. Both systems received the same answer precision but TagClouds was rated lower in the aspects of task easiness and system usefulness. Thus hypothesis (H1) was rejected.

For task 2, TagClouds performed better concerning the time efficiency and answer precision, and participants spent significantly less time with TagClouds (t(11)=3.92, p<0.05). The scores for the task easiness and system helpfulness were also significantly higher with TagClouds (t(11)=3.354, p<0.01; t(11)=3.139, p<0.05). Then Hypothesis (H2) was supported.

For task 3, although TagClusters performed better in the aspects of answer precision, task easiness and system usefulness, it cost more time than with TagClouds. Thus Hypothesis (H3) was rejected.

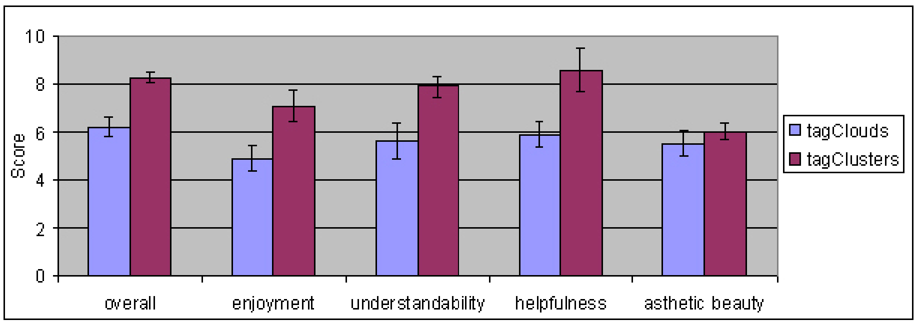
TagClouds contain all the tags in a small graph and it is easier to scan and locate tags without panning or zooming. To present all the tags in groups and describe the similarity between tags as spatial distance, TagClusers need more space and thus create a larger graph in which the participants had to keep panning and zooming to get a complete overall impression. To form the correct impression, the participant also needed to mentally compare and memorize the relevant information which might slow down the response time and lead to answers with lower precision. Results for task 2 and 3 imply that we should make better usage of the space in TagClusters in order to create a smaller and more efficient visualization.

Figure 4. Comparison of both systems. Top left: completion time. Top Right: answer precision. Bottom left: task easiness. Bottom right: system usefulness.



The results for task 4-6 were rather consistent: TagClusters received higher scores in all aspects. A dependent t-test showed that TagClusters worked significantly better with task 4 concerning both time efficiency (t(11)=2.752, p<0.05) and answer precision (t(11)=4.077, p<0.01). Since the tags are hierarchically organized and semantically similar groups are placed near to each other, it is easy to find the structure. With TagClouds, the semantically similar tags might be placed scattered all over the graph, and the participants had to scan all the tags to form a mental structure, which cost much more time and led to lower precision. To derive the complex structure for genre-related tags in task 4, the participants spent much more time with TagClouds but received lower precision in their answers. Since the semantically similar tags are hierarchically grouped and the overlapping part is visually highlighted, it is easier to determine the relation between tags. Thus TagClusters worked better with task 5 and 6 which require understanding of semantic relation among tags. TagClusters also performed significantly higher in task 4-6 regarding the system usefulness ($t(11)=4.872$, $p<0.01$; $t(11)=4.451$, $p<0.05$; $t(11)=2.526$, $p<0.05$). Therefore, hypothesis (H4) was verified.

Figure 5. Overall impression of both systems



After completing all tasks, the participants filled out a post-questionnaire concerning the overall impression of both systems in the aspects of enjoyment, understandability, helpfulness and aesthetics (see Figure 5). TagClusters was scored overall scientifically higher ($t(11)=4.358$, $p<0.01$), and specifically significantly higher in the aspects of enjoyment ($t(11)=2.905$, $p<0.05$), understandability ($t(11)=3.446$, $p<0.01$) and helpfulness ($t(11)=8.060$, $p<0.01$).

Discussion

Visualization issue

For TagClouds, the alphabetical order is useful when the user has a specific tag in mind. However, users who are unfamiliar with this visualization tend to ignore this feature. Although the tags with bigger font size are easier to be noticed which is helpful to get an overall impression of popular tags, tags with a smaller print are likely to be ignored. The position is also a crucial factor to draw the user's visual attention. Some users claimed that the top half part of TagClouds seemed to be more prominent and that they tended to ignore the bottom half. In order to get a compact view and make effective use of the available space, the system truncated long tags into separate lines which might have caused misunderstandings. For example, in the first line of Figure 1, *alternative rock* is separated into 2 lines, which confused some participants.

By grouping semantically similar tags, TagClusters help to discover small tags. For example, the *Rock*-related tags with small font size, which might be ignored in Figure 1, still remain relevant in Figure 2 since they are clustered into the same group with the prominent *Rock* tag.

Semantic understanding

Without the indication of semantic relationships in TagClouds, some participants wrongly interpreted a near position or similar font size as semantic similarity. There is no semantic organization in TagClouds and sometimes the user has to scan and understand all tags line by line and might have problems with locating multiple tags at the same time. Some participants even used the mouse to locate a viewed tag or were staring at the screen while writing down the answers. Another problem is that users with less music knowledge might meet difficulties with judging the relation between uncommon tags. This was a prominent problem for the majority of the foreigner participants. With the illustration of semantic structures in TagClusters they could conduct the same tasks easier.

The participants also came up with some aesthetical suggestions for TagClusters, such as stronger highlighting, color coding for different tag categories and the desire for a more compact graph.

CONCLUSIONS AND FUTURE WORK

TagClouds are a popular visualization for collaboratively generated tags. However, they have some distinct problems. In this paper we investigate ways to support semantic understanding of collaboratively generated tags. We conducted a survey on practical tag usage in Last.fm, an online music community. Based on the results, we propose a visualization named TagClusters, in which tags are clustered into different semantic groups and the visual distance represents the semantic similarity between tags. We compared the performance of traditional TagClouds and TagClusters in a user study. We received overall positive feedback on TagClusters and the results imply that our tool has advantages in supporting semantic browsing and better understanding of hierarchical structure and relationships among tags. The semantic organization of tags can exclude redundancy effectively and might also facilitate tag recommendation and tag-based music retrieval, which will be explored in our future work. Besides, similarly to ThemeRiver (Westerman & Cribbin, 2000), the temporal dimension can be integrated to offer a time-based visualization which may indicate the hidden trends among general musical interests over time.

ACKNOWLEDGEMENTS

This research was funded by the Chinese Scholarship Council (CSC), the German state of Bavaria and the Spanish Ministerio de Educación y Ciencia (project CSD 2007-00067). We would like to thank the participants of our user study.

REFERENCES

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays. Proc. of the SIGCHI conference on Human Factors in Computing Systems (pp. 313-317). ACM Press.

Bateman, S., Gutwin, C., & Nacenta, M. (2008). Seeing things in the clouds: the effect of visual features on tag cloud selections. Proc. of the 9th ACM conference on Hypertext and Hypermedia (pp. 193-202). ACM Press.

Begelman, G., Keller, P., & Smadja F. (2006). Automated tag clustering: improving search and exploration in the tag space. Proc. of the 15th international conference on World Wide Web. Retrieved October 02, 2009, from
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.5736&rep=rep1&type=pdf

Bielenberg, K. & Zacher M. (2005). Groups in social software: Utilizing tagging to integrate individual contexts for social navigation. Unpublished master dissertation, University of Bremen, Germany.

Cunningham, S. J., Bainbridge, D., & Falconer, A. (2006). 'More of an Art than a Science': Supporting the Creation of Playlists and Mixes. Proc. of the 7th International Conference on Music Music Retrieval. Retrieved October 02, 2009, from http://waikato.researchgateway.ac.nz/bitstream/10289/77/1/content.pdf

Fruchterman, T. M. J., & Reinhold, E. M. (1991). Graph Drawing by Force-directed Placement. Software -- Practice and Experience, vol. 21(11) (Nov. 1991), 1129-1164.

Gansner, E. R., & North, S. C. (1998). Improved forced-directed layout. Proc. of the 6th Symposium on Graph Drawing (pp. 364-374). Springer Press.

Golder, S. A. & Huberman, B. A. (2006). Usage Patterns of Collaborative Tagging Systems. Journal of Information Science, 32(2) (Apr. 2006), 198-208.

Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. Proc. of the 16th international conference on World Wide Web (pp. 211-220). ACM Press.

Halvey, M. J., & Keane, M. T. (2007). An assessment of tag presentation techniques. Proc. of the 16th international conference on World Wide Web (pp. 1313-1314). ACM Press.

Hassan, M. Y., & Herrero, S. V. (2006). Improving tag-clouds as visual Music Retrieval interfaces. Proc. of the International Conference on Multidisciplinary Information Sciences & Technologies.

Havre, S., Hetzler, B., & Nowell, L. (2000). ThemeRiver: Visualizing Theme Changes over Time. Proc. of IEEE Symposium on Information Visualization (115-123). IEEE Press.

Hearst, M. A. What's up with Tag Clouds? (2008). Retrieved October 02, 2009, from http://www.perceptualedge.com/articles/guests/whats_up_with_tag_clouds.pdf

Hearst, M. A. & Rosner, D. (2008). Tag Clouds: Data Analysis Tool or Social Signaller? Proc. of the 41th International Conference on System Sciences (pp. 1-10). IEEE Press.

Kaser, O., & Lemire, D. (2007). Tag-cloud drawing: algorithms for cloud visualization. Proc. of the international conference on World Wide Web. ACM Press.

Kerr, B. (2008). TagOrbitals: tag index visualization. IBM research report. Retrieved October 02, 2009, from http://domino.research.ibm.com/library/cyberdig.nsf/1e4115aea78b6e7c85256b360066f0d4/8e1905dc8b000673852571d800558654?OpenDocument

Kuo, B. Y. L., Hentrich, T., Good, B. M., & Wilkinson, M. D. (2007).Tag clouds for summarizing web search results. Proc. of the 16th international conference on World Wide Web (pp. 1203-1204). ACM Press.

Li, R., Bao, S., Yu, Y., Fei, B., & Su, Z. (2007). Towards effective browsing of large scale social annotations. Proc. of the 16th international conference on World Wide Web (pp. 943-952). ACM Press.

Lohmann, S., Ziegler, J., & Tetzlaff, L. (2009). Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration. Proc. of the 12th international conference on Human-Computer Interaction (pp. 392-404). Springer Press.

Mathes, A. Folksonomies - cooperative classification and communication through shared metadata. Retrieved October 02, 2009, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Nielsen, M. (2007). Functionality in a second generation tag cloud. Unpublished master thesis, Gjøvik University College, Norway.

Porter, M. F. (2006). An algorithm for suffix stripping. Program: electronic library and information systems, 14(3) (2006), 211-218.

Provost, J. (2008). Improved document summarization and tag clouds via singular value decomposition. Unpublished master thesis, Queen's University, Canada.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. Proc. of the SIGCHI conference on Human factors in computing systems (pp. 995-998). ACM Press.

Sammon, J. W. (1969). A non-linear mapping for data structure analysis. IEEE Transactions on computers, vol c-18, no. 5 (May 1969), 401-409.

Santamaría, R., & Therón, R. Overlapping Clustered Graphs: Co-authorship Networks visualization. Proc. of the 9th international Symposium on Smart Graphics (pp.190-199). Springer Press.

Seifert, C., Kump, B., & Kienreich, W. (2008). On the beauty and usability of tag clouds. Proc. of the 12th international conference on Information Visualization (pp.17-25). IEEE Press.

Shaw, B. Utilizing folksonomy: Similarity metadata from the del.icio.us system. Retrieved October 02, 2009, from http://www.metablake.com/webfolk/web-project.pdf

Viégas, F. B., & Wattenberg, M. (2008). Tag Clouds and the Case for Vernacular Visualization. Interactions. 15(4) (July+August 2008), 49-52.

Wu, X., Zhang, L., & Yu, Y. Exploring social annotations for the semantic web. Proc. of the international conference on World Wide Web (pp.417-426). ACM Press.