SparkClouds: Incorporate Spark Lines into a Tag Clouds

G.Lakshmi	K.Rajendar	J.Pratap
M.Tech(C.S.E).	M.Tech(C.S.E) .	M.Tech(P.hd.).
VCET,Karimnaga,Ap.	JITS ,Karimnagar,Ap.	VCET,Karimnagar,Ap.

Abstract

The aim of this project is to introduce Spark Clouds, which integrate spark lines into a tag cloud to convey trends between multiple tag clouds. Spark Clouds ability is to show trends compares favorably to the alternative visualizations. In the Existing System, Tag clouds are used to display the relative tag frequency, popularity, or importance by font size. They serve as a visual summary of document content. Tag clouds seem to invite exposure of their evolution over time; they do not explicitly represent them. In the proposed System, we introduce Spark Clouds, a new breed of tag cloud that incorporates spark lines with more typical tag cloud features to convey evidence of change across multiple tag clouds. Spark Clouds retain the advantages of tag clouds while incorporating minimal but sufficient indication of trends for a reasonable number of related tag clouds.

Index Terms: Tag Clouds,Trend visualization,Multiple line graphs,Evaluation,Staked bar charts.

1 INTRODUCTION

Tag clouds have proliferated over the web over the last decade. They provide a visual summary of a collection of texts by visually depicting the tag frequency by font size. In use, tag clouds can evolve as the associated data source changes over time Interesting discussions around tag clouds often include a series of tag clouds and consider how they evolve over time. However.since tag clouds do not explicitly represent trends or support comparisons, the cognitive demands placed on the person for perceiving trends in multiple tag clouds are high. In this paper, we introduce SparkClouds, which integrate sparklines into a tag cloud to convey trends between multiple tag clouds. We present results from a controlled study that compares SparkClouds with two traditional trend visualizations—multiple line graphs and stacked bar charts—as well as Parallel Tag Clouds . Results show that SparkClouds' ability to show trends compares favourably to the alternative visualizations.

Tag clouds are a text-based visual depiction of tags (or words), typically used to display the relative tag frequency, popularity, or importance by font size. They can also serve as a visual summary of document content. In the last decade, tag clouds have proliferated over the web. They are now a common visualization in news sites for displaying the most active news story themes , photo sharing sites for conveying the distribution of image content , and social bookmarking sites for showing popular tags . In fact, several online programs are available that help you create your own tag clouds from different types of text sources .

Tag clouds can evolve as the associated data source changes over time. For example, the US Presidential Speeches tag cloud shows the popularity, frequency, and trends in the usages of words within speeches, official documents, declarations, and letters written by the Presidents of the US between 1776 and 2007]. Other sources of highly dynamic content include online news and photo-sharing sites which serve freshly uploaded and tagged material every day. Interesting discussions around tag clouds often include a series of tag clouds and consider how they evolve over time. However, while tag clouds seem to invite exposure of their evolution over time, they do not explicitly represent them. This results in a significant cognitive demand on people who want to Under stand how a tag cloud is evolved In this paper, we introduce SparkClouds , a new breed of tag cloud that incorporates sparklines with more typical tag cloud features to convey evidence of change across multiple tag clouds. We also present a controlled study that we conducted to compare SparkClouds with Parallel Tag Clouds (PTCs) (the only previous tag cloud visualization designed for understanding multiple tag clouds), as well as two traditional trend visualizations—multiple line graphs and stacked bar charts. We compared these four visualizations in terms of speed and accuracy in supporting three types of tasks (specific data, topic trends, and overview). We found that SparkClouds' ability to show trends compares favourably to the alternative visualizations. Participants also preferred SparkClouds to stacked bar charts and PTCs.

We organize this paper as follows. In the next section we outline the related work to provide context for our description of SparkClouds, which follows in Section 3. Section 4 describes the design and results of the controlled study along with the alternative visualizations we used in the study. We then conclude the paper with a discussion of the

lessons learned from the study and suggestions for future work.

2. RELATED WORK

The origin of tag clouds goes back to 1976 when an experiment was carried out by Stanley Milgram . A collective "mental map" of Paris was created using font size to show how often each place was mentioned as a landmark in the city. In 1997, Search Referral Zeitgeist was created by Jim Flanagan as a way to visualize the number of times a term was used to find a given website by font size. Among high-profile websites, Flickr used tag clouds first, followed by other Web 2.0 sites . For more details about the history of tag clouds, see . Due to their astonishing popularity, there have been many efforts in exploring various properties of tag clouds. Several websites now enable people to create their own tag clouds from different types of text sources . One interesting variation, showing two-word phrases, provides a quite different perspective of the textby revealing themes in the content . There has been considerable research to improve tag cloud layouts. Kaser and Lemire organized tags in nested tables for HTML based sites by using an Electronic Design Automation (EDA) packing algorithm . Seifert et al. proposed a new algorithm to address several issues found in the traditional layouts . It creates compact and clear layouts by reducing whitespace and featuring arbitrary convex polygons to bound the terms. Tree Cloud arranges words on a tree to reflect their semantic proximity according to the text .

Tag Maps employs a unique layout based on real geographical space . Wordle provides remarkably distinctive layouts by utilizing typography, color, and composition to balance various aesthetic criteria . Research efforts that attempt to understand the effectiveness and utility of tag clouds generally fall into one of two categories; those which investigate the visual features of tag clouds and those which compare tag clouds with different layouts. Bateman et al. compared nine visual properties of tag clouds for their effects on visual search for tags . Their results show that font size and font weight have stronger effects than others such as color intensity, number of characters, or tag area. Rivadeneira et al. conducted two experiments . In the first study, they examined the effect of font size, location, and proximity to the largest tag, asking participants to recall terms (for 60 seconds) that were previously presented in tag clouds (for 20 seconds). In the second study, they investigated the effect of bothfont size and word layout on users' abilities to form an impression (gist). From both studies, in accordance with previous research, they observed a strong effect of font size.
Lohmann et al. compared four tag cloud layouts for

three types of search task; one of them was a standard list using uniform font size with wrapping . All but the traditional tag cloud worked best for
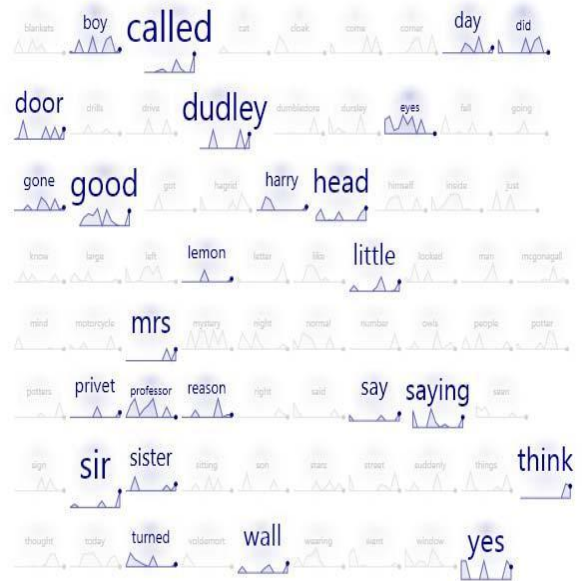

Fig. 1. SparkClouds showing the top 25 words for the last time point
(12th) in a series. 50 additional words that are in the top 25 for the
other time points can be (top) filtered out or (bottom) shown in gray
at a smaller fixed-size font. (bottom) is used in the study.

one task. For the task of finding a specific tag, the list performed better than the tag cloud. Halvey and Keane compared tag clouds with traditional lists (horizontal and vertical), each with regular vs. alphabetical order by asking participant to find a specific tag . They found that lists perform better than tag clouds and that alphabetical order further accelerates the search speed. Sinclair and Cardew-Hall conducted an experiment to investigate the preference between a tag cloud and a traditional search interface both for general browsing and for information seeking tasks They found that participants preferred the search interface for specific information retrieval tasks whereas the tag cloud was preferred for more open-ended browsing tasks. In use, tag clouds can evolve as the associated data source changes over time. Interesting discussions around tag clouds often include a series of tag clouds and consider trends of their tags over time. This desire to study trends and understand how text content or topics evolve over time has been the purpose of other visualizations such as the commonly used line graphs and bar charts. However, despite the significant amount of research on tag clouds, there has not been much research on how to visualize trends in tag clouds.
Many Eyes allows people to compare two texts in a single tag cloud . It uses two colors (one

per tag cloud) and pairs the tags that appear in both clouds. While it enables easy comparison between two tag clouds, this technique does not offer help in understanding trends over time as it is limited in the number of tag clouds visualized. Tagline Generator allows people to generate a sequence of tag clouds that are associated with time, from a collection of documents ; a dynamic slider control is used to navigate the time points, but only one tag cloud is shown at a time. Parallel Tag Clouds (PTCs) is designed to provide an overview of a document collection by incorporating graphical elements of parallel coordinates with the text size encoding of traditional tag clouds . While PTCs do show multiple clouds simultaneously, they do not explicitly represent trends, and thus comparing multiple tag clouds to ascertain trends places the cognitive demands on the person.

Cloudalicious is an online tool specifically designed to visualize how tag clouds develop over time . For a given website, it downloads the tagging data from del.icio.us and then graphs the collective users' tagging activities over time using multiple line graphs. While Cloudalicious clearly shows some trends, such as decay of the collective usage of tags, it may suffer from overlapping of lines and does not retain the visual appearance of tag clouds. Dubinko et al. presented a new approach to visualize the evolution of tags in Flickr using an animation via Flash in a web browser . While they allow people to observe and interact with the tags, their main contributions were not focused on the visualization but rather on algorithms and data structures to generate the list of "interesting" tags for a specific time period. To better convey the evidence of change across multiple tag clouds, we developed a new breed of tag cloud called SparkClouds that integrates sparklines into a tag cloud. We also conducted a controlled study to explore the efficacy of SparkClouds by comparing it with two traditional trend visualizations, multiple line graphs and stacked bar charts, as well as with PTCs.

3 DESIGNING SPARKCLOUDS

The basic idea behind SparkClouds is to retain the advantages of tag clouds while incorporating minimal but sufficient indication of trends for a reasonable number of related tag clouds. In particular, we focused on these advantages of tag clouds: Compact use of space that can be flexibly reorganized into different aspect ratios without negatively impacting the readability of the cloud as a whole. Tag (or word) readability in that the importance or frequency of a tag is encoded directly in the size of the word. Since we based our design of SparkClouds on two usage scenarios, we first describe these scenarios to provide the setting for our design development.

3.1 Scenarios

3.1.1 Keeping track of different non-familiar tag clouds Ramu is a stock market analyst. Every day, he has the same morning ritual. He spends about an hour on the web to absorb the information required to keep up-to-date with the market. Is the new product everyone is talking about out already? What is the next big application that mobile users are talking about? What are the reviews on the latest phone? Ramu has dozens of websites bookmarked, but in fact, the majority of them are irrelevant today. Indeed, he already knows a great deal of information and may already have consulted them in recent days. The real challenge for him is to identify and select which sites to dive into, to find where the new information is, and to locate any deeper information he may need. Currently, Ramu's strategy is to read RSS feed titles and browse a dozen or so websites, many of which conveniently present the current topics in a tag cloud form. Ramu likes these simple representations as they give him a gist of the content of the website. He often remembers the large tags, but he still has trouble spotting the new tags and topics, as well as keeping track of the less popular ones, especially because he sees many different tag clouds every day.

3.1.2 Monitoring familiar tag clouds

Anjali is a researcher working as part of a 16-person team that is rapidly increasing in size. Right now, she works closely with 5 of her team members and is relatively well-aware of their activities. But, as she juggles increasing numbers of projects, she realizes that she cannot keep up-to-date with the whole team. For example, it is quite challenging for her to keep track of what is happening from the monthly meetings and weekly status reports. Indeed, she recently learned from her manager that she spent several days surveying a topic that one of the team members had already been working on for the past few weeks. To maintain awareness of the team activity, she generates tag clouds from the weekly status reports. While they help her remember the key projects, she still has a difficult time noticing what has changed week-to-week, much less over longer spans of time. Being able to compare these project tag clouds as they evolved might have helped highlight her colleague's shift in their work focus.

3.2 Design Rationale Summary

A tag cloud is a visual representation that a broad range of people can easily decode, and makes effective use of display space. Our primary goal with SparkClouds is to preserve these two characteristics, while incorporating the ability to convey trends. We also aim at supporting the two usage scenarios described above. Ramu is dealing with a large number of tag clouds; these may hange radically and Ramu may not remember all topics in

previous tag clouds. He needs to have an *overview* of the trends to understand the market at a high level. Anjali is familiar with the data. She has a good memory of the simplified line graphs in the sense that axes are implied rather than explicitly drawn or labelled. They can be very compact and still provide an indication of a trend. As was also shown in , this property made them attractive to use as they can be inserted adjacent to each tag without cluttering the entire presentation. previous tag clouds *time-point-by-time-point* and compare them to get a deeper understanding of what each person did over time. A SparkCloud encodes the popularity of tags by font size, as do standard tag clouds. To show the trends in popularity of each tag over time, we introduce a second visual element adjacent to each tag: a sparkline, *.i.e.* a minimal simplified line chart

3.2.1   Tag Clouds as Used in SparkClouds
The tag cloud aspect of SparkClouds has two design parameters: the font size encoding and layout. whether one is slightly more frequent than the other. This may be useful when precisely comparing trends but we thought it was not needed for our current usage scenarios.

Representing Zero: To help people make comparisons between trends, we depict the horizontal axis of each sparkline. This transforms the sparkline into a sparkarea by filling the space between the sparkline and the horizontal axis using a light gradient blue color .

Font Size Encoding: In traditional single tag cloud present. In the second method, the font size is used to encode the frequency of tags at a given time point. In this view, Anjali is able to identify at a glance her team members' most popular topics for each time point. This view is also more appropriate for routine review, as it most clearly depicts the activity at a given time point. With this method, the font size is stable for each tag in that it does not change whether the visualization shows either one or several tag clouds.   However, this encoding makes tags between tag clouds comparable only in rank not in frequency.   While SparkClouds, by default, use per time point font size encoding, they allow people to select a more appropriate encoding (or normalization) method according to their task as PTCs do.

Layout: As described above, there has been much research on laying out tag clouds . Even though we present the design of SparkClouds as having an alphabetical order throughou the paper, it can support any existing tag cloud layout because SparkClouds retains the same structure as conventional tag clouds.

3.2.2 Sparklines
The most common visual encodings of trends over time are the traditional line graphs and more recently stacked graphs . We selected the simpler one of the two, sparklines, for use in visualizing trends over time in SparkClouds. Sparklines are simplified line graphs in the sense that axes are implied rather than explicitly drawn or labelled. They can be very compact and still provide an indication of a trend. As was also shown in , this property made them attractive to use as they can be inserted adjacent to each tag without cluttering the entire presentation

Data Encoding: In SparkClouds, a sparkline depicts the popularity of the tag (vertical axis) over time (horizontal axis). To maintain consistency in the representation, the vertical axis of sparklines does not encode the raw popularity of tags, but instead uses a linear transformation function based on the relative popularity of tag. A potential refinement to this approach could be to take advantage of the extra precision offered by sparklines—if two word functionally map to the same font size, the sparkline can indicate whether one is slightly more frequent than the other. This may be useful when precisely comparing trends but we thought it was not needed for our current usage scenarios.

Representing Zero: To help people make comparisons between trends, we depict the horizontal axis of each sparkline. Thistransforms the sparkline into a sparkarea by filling the space between the sparkline and the horizontal axis using a light gradient blue color (Fig. 1). This visual encoding helps in comparing sparklines that are not horizontally aligned with each other and gives additional visual assistance in identifying the periods during which a tag was not By glancing at the sparkline below a given tag, Ramu and Anjali can assess if it is new, if its popularity has been stable or when it experienced a spike in popularity.

3.2.3 Unifying Tag and Sparkline:
Given that we introduced an additional visual element for each tag, we broke the "homogeneity" of the tag cloud. To help people perceive the two visual elements (tag and sparkline) as a single unit of information, we considered several possibilities .Alignment: Sparklines can be placed before. We experimented with using a script (handwriting) font to provide visual continuity, but we thought that the legibility of the tag was compromised . We also tried using an explicit baseline and aligning both text and sparkline on the baseline, but decided against this idea as it introduced too much clutter.

Overlays: We explored overlaying the tag over the sparkline . These representations appeared too cluttered and did not work well in practice since words vary widely in length; this made it difficult to compare points in time between two words whose

sparklines spread across different scales. In addition, we thought that text overlays, even with light font and white outlines, made the words less readable (especially when the font size is small).

Mirror: To avoid the clutter caused by the overlays, we tried to mirror the line chart and place it under the word . We decided against this option as we believed it may compromise the decoding of the sparkline and thus be misleading. Indeed, even with various background and line-color options, we tended to interpret the mirrored sparkline as standard non-mirrored ones.

Foreground and background colors: We first explored using two foreground colors, which alternated between adjacent terms (but coloring the tags and associated sparkline with the same color). This solution worked quite well, but it looked more cluttered than the one with a single color. We also thought that we could use color more effectively for other purposes. In Swapna's scenario, for example, we

(a) align before
(b) align after
(c) align above (d) overlay (e) mirror

(a)   (b)

(c)   (d)

(e)
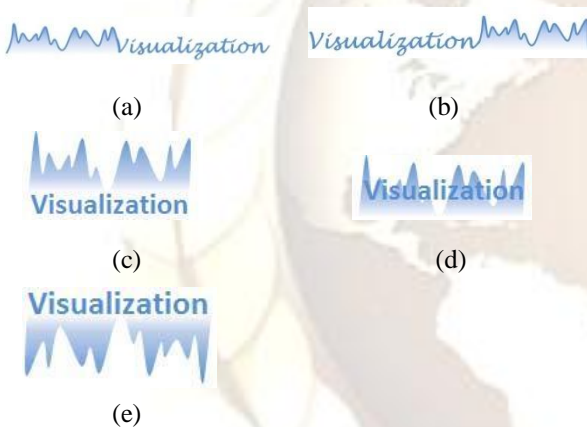
4.3   Visualizations

We compared four visualizations: SparkCloud, Multiple Line Graph, Parallel Tag Cloud, and Stacked Bar Chart (see Figs 1, 3, and 4). We chose multiple line graphs and stacked bar charts because they are commonly used in many tools to show trends . We chose Parallel Tag Clouds because they are the only tag cloud specifically designed to support cross tag cloud comparison for more than two tag clouds.

25 word While each of these visualizations is inherently different from one another, when possible we implemented the visualizations based on the same visual guidelines. For example, words are presented in an alphabetical order in all four visualizations. SparkClouds and Parallel Tag Clouds share the same range of font sizes (from 10 to 34), and Multiple Line Graphs and Stacked Bar Charts use the same font size (15). For all

visualizations, all 75 top words were shown, clearly indicating which were the top 25 words for each month. In Parallel Tag Clouds, since the top s for each month are displayed simultaneously in the column for their month, all words are shown and some words are shown in multiple columns. The three other visualizations display all the top 75 most frequent words and highlight only the words that were included in the top 25 for the selected month.

4.3.1   Multiple Line Graph (MultiLine)

A line graph explicitly shows trends (i.e., how a value changes) by connecting a series of successive data points; usually the x-axis represents time. While a multiple line graph helps people compare trends between multiple variables, it often suffers from overlapping when many lines are displayed at once. To alleviate this problem, we implemented the multiple line graph in the following way. While all the top 75 most frequent words are displayed, only the words that were included in the top 25 for the selected month are highlighted . Participants can select a time point by clicking on the label shown at the bottom of the visualization. The currently selected time point is marked with blue background and a thick border. When participants move the mouse over a line or a word in the legend list box at the top, we highlight only the focused word and the line. When the values overlap, we slightly shift the data point diagonally (2 pixels to the right and below). We used 5 distinctive colors to help participants better differentiate the lines.
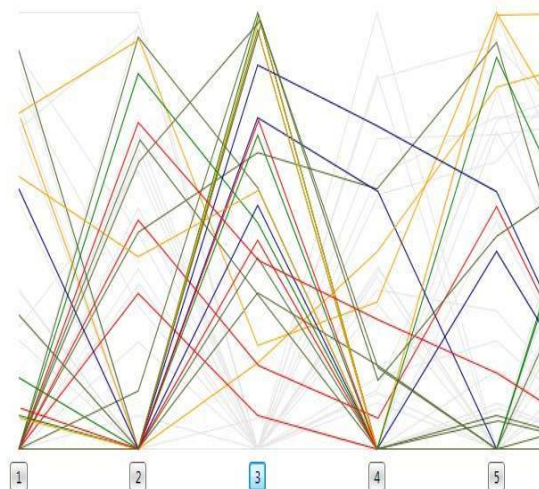
4.3.2 Stacked Bar Chart (StackedBar)

Stacked bar charts, in which bars are divided into nominal variables, are commonly used to show trends. ThemeRiver is a timeline indicating the flow of document themes. It uses width of the river to show the number of documents and the river is sub-divided by topics, which ebb and flow over time. The Name Voyager visualizes a graph of the popularity of baby names over the past century]. Both of these examples use smoothly connecting lines between the data points. In our implementation we kept the adjacent bars discrete to make individual data values more readable. While stacked bar charts are particularly good for conveying at cumulative and overall trends, we identified two major issues with standard stacked bars. First, they may be deceiving for evaluating tends across time for a single term because at any given time point the placement of neighboring terms interferes with (or rather, has a displacement effect on) the placement of the series of interest. Second, label placement is non-trivial because each tag requires a position that is large enough to ensure the label is readable, which is not always possible if a tag generally has a low popularity; furthermore, labels are not guaranteed to be vertically or horizontally aligned, making scanning difficult. Both of these issues are problematic for tag clouds since readability is of critical importance to investigate trends in tags over time. To address these issues, we modified the standard stacked bar in the following way. The tags are shown as a vertical list on the left, each of which is accompanied with a (horizontally) stacked bar (Fig. 3b). This stacked bar for each tag is created by horizontally stacking individual bars for each month. To help people compare bars for a particular month, stacked bars can be interactively aligned to the left side of the bars for the selected month. We also drew a baseline starting from the time point label.
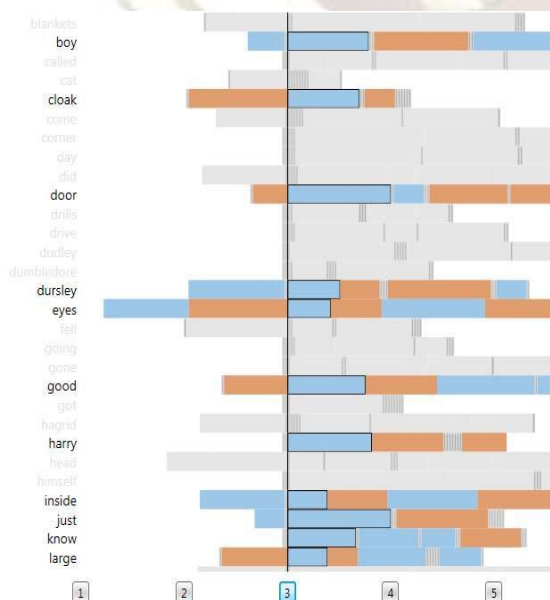
Conclusion

The disadvantage of the tag clouds could be overcome by using spark clouds. In this paper, we have described SparkClouds, a novel visualization that incorporates sparklines into a tag cloud to represent trends across a series of tag clouds; SparkClouds inherit advantages from both sparklines and tag clouds. First, with sparklines, SparkClouds effectively provide people with an overview of trends using very little additional space. Second, because it is still in the form of tag clouds, SparkClouds offer a compact and aesthetically pleasing layout and can be used in place of traditional tag clouds. We have also described the design of SparkClouds along with usage scenarios.

REFERENCES

1) Visualizing Translation Variation: Othello : A Survey of Text Visualization and Tools Zhao Geng, Robert S.Laramee, Tom Cheesman, Andy Rothwell, David M. Berry,   Alison Ehrmann, Swansea University, 2011
2) ManiWordle: Providing Flexible Control over Wordle Kyle Koh, Bongshin Lee, Bohyoung Kim, and Jinwook Seo
3) Marketspace Envisioning the Cloud:The Next Computing Paradigm March *Jeffrey F. Rayport ,Andrew Heyward.*
4) ABC News – Tag Cloud – News topics organized to show what's in today's headlines, http://www.abc.net.au/news/tag/cloud.htm
5) S. Bateman, C. Gutwin, and M. Nacenta, "Seeing things in the clouds: the effect of visual features on tag cloud selections," *Pro*
6) P. Bausch and J. Bumgardner, *Flickr Hacks.* O'Reilly Press, pp. 82-86, 2006.