Blogging September 11; Transforming Personal Expressions
Into Data Visualization in Wordle and Tagxedo

A Master’ s Thesis
Presented to
School of Arts and Science

In Partial Fulfillment
of the Requirements for the
Master of Science Degree

State University of New York
Institute of Technology

By

Andrzej Lata
May 2012
DEPARTMENT OF INFORMATION DESIGN AND TECHNOLOGY

Approved and recommended for acceptance as a thesis in partial fulfillment of the requirements for the degree of Master of Science in Information Design and Technology

DATE

________________________________________

Dr. Steven Schneider
Thesis Advisor

________________________________________

Dr. Ibrahim Yucel
Second Reader
Abstract

September 11, 2001 is one of the most recognizable dates in world’s modern history. Much has been said, published and reported about the tragic events and more may still be revealed and understood about that unforgettable day. So much information has been collected about September 11th, and many have studied and analyzed this information.

Many educational, governmental and other organizations and institutions have worked and contributed toward a better understanding of the events of that tragic day. Under the guidance and leadership of Professor Steven Schneider, teachers and students of SUNY IT had an opportunity to participate in several projects directed towards a deepening understanding and reflection upon 9/11. I had the opportunity to participate in two of these projects during the Spring and Summer of 2011. My participation sparked my choice of a topic for my final project. I investigated content of blog entries posted immediately after the terrorist acts of September 11th. The research questions I sought to answer were centered on the personal expression of bloggers: What did they write about? How can this data be captured and analyzed? How can the researcher best present this data and findings?

Analyzing blog content, identifying key words and transferring text of blogs into images called word clouds, is the solution I have modeled. The project was finalized by creating 42 word clouds, 21 in Wordle and the same number in Tagxedo, both are word cloud generation software. Sifting through the blogs and examining data in different forms enriched my understanding of the variation in the personal expression of the bloggers and helped me to see what they reported and contributed in a totally new light. There are still many questions to be answered about personal expression of people who reported their stories on the Internet in September, 2001 and many more projects and studies that could contribute much more towards understanding of content of 9/11. My hope is that the findings and methodology reported here will add a small piece in understanding of what happened in September, 2001.
Table of Contents

List of Tables

Introduction 1

Literature Review 2
Terminology and data visualization 2
Similarity of conducted research 3
Color in word and tag clouds 5
Data visualization in the clouds and their limitations 5
Clouds as research tools 6
Students of foreign languages and word clouds 7
New development in data visualization – Sparkclouds 7
Conclusions 8

Research Questions 9
Theoretical context 10

Methods 11
Goals and object of the study 11
Influence of literature review 12
Process of data selection 14
Conclusion 19

Project Development and Application 21
Differences in Word Clouds generators 21
Focusing on Differences ob blog transferred into Wordle and Tagxedo 24
Conclusions 57

Future Research 58

References
**Introduction**

People have been communicating from the beginning of humanity. Sharing words, emotions, warning of danger signs and expressing joys as well as disappointments constantly evaluated throughout the millennia. Human beings need medium for communication and others to exchange information. They also need a place, gathering area where information can be share and where connection and community can exist.

Individuals as well as cultures and societies tell stories. They all are unique, special, and often contain elements of emotions. Stories have potentials of transforming people and history. What thousands of years ago once was a gathering place in the middle of jungle, park or city to serve as a place of meeting others now can be created quite differently. The Internet does not provide opportunity for physical gathering but it can help to build community and connection among people.

Bloggers, can “gather” online, share their stories and express themselves in many ways. They can create communities in the blogosphere and continue developing language of communication.

Scholars and researchers studied communication in Web Sphere, analyzing various aspects of users’ engagement, “*Many individually produced blog sites displayed postings in which the blogger reported his/her reactions to the attacks*” (Foot and Schneider, 2004).

Project “Blogging September 11, transforming personal expressions into data visualization in Wordle and Tagxedo” focuses on asking and answering questions regarding content of weblogs. It concentrates on the content of 21 blogs and presents it form of word clouds.
The project tries to answer questions about bloggers’ expressions and bring a new light into understanding of what happened in blogosphere in the first hours and days after the terrorist attacks of 9/11. “Blogging September 11” project tries to tell stories in a new way applying technology for data visualization. Word clouds created in Wordle and Tagxedo that are the final result of this project and they continue exploring new ways of how we communicate and express ourselves.

LITERATURE REVIEW – WORD CLOUDS, BLOGGING AND SEPTEMBER 11

Word clouds are still relatively new phenomena and as with any new technology they are evolving. While researching this subject I found limited resources on the topic. This literature review contains material about word clouds and similar areas such as tag clouds that have many things in common with my main subject. Word clouds can be classified in a broad category of infographics. There are many applications of infographics and the number of available tools to create visual representation of data is growing. Data in infographics can be represented in charts, graphs, timelines and other visual formats Troutner (2011).

Terminology and data visualization - Tag clouds, content clouds, word clouds

Words can be classified and displayed in a variety of ways. In many applications of Web 2.0, users can add tags to content that they create and publish or find in other sources. This process requires users to manually assign tags as they see appropriate. As Cosh et al. (2008) accurately point out: “The benefits of the content cloud approach, are firstly that it can be fully automated, not requiring the efforts of a tagging community. This ensures its neutrality taking a standardized text for comparisons” (p.728). Tags can be
assigned to different kind of objects such as pictures, graphics, graphs and text. On the other hand, content clouds are user generated visual representation of text. They can apply to a single text as well as collection of written materials, Cosh et al. (2008).

The authors also mention that tag clouds and content clouds are based on different linguistic bases, “Content clouds are the result of a detailed analysis of lexemes found directly in the text” (2008, p.728) and “Tag clouds are more abstract in comparison to single words or text.” There is a distinction in the way that content clouds are designed. As Cosh, Burns and Toby (2008) indicated, content clouds apply “natural language processing (NLP) tools to automatically extract the content of some text, visualizing the results in a content cloud “(p.722).

Hassan-Montero and Herrero-Solana (2006) promote terms that help to understand tag cloud environment. One of the words they use is “Folksonomy” which “allows anyone to access any web resource that was previously tagged, based on two main paradigms of information access: Information Filtering (IF) and Information Retrieval (IR)” (p.25). Tagging becomes a crucial element of creating social indexing of the resources on the web.

Appelo (2008) states that “some text” may not be interesting to users and should be omitted from the tag cloud. This is the case against articles and other small words that are considered to be “noise” by search algorithms” (p.21).

**Similarity of conducted research**

Lohmann, Ziegler and Tetzlaff (2009) compared several tag clouds in their research. They focused the study of tag clouds with special emphasis on perception and performance in relationship to user goals. In
addition, authors conducted analysis of eye tracking data which contribute to gain insights that can be applied in creation of the tags. Lohmann et al.(2009) invited 36 participants to take part in their study. Their research group consisted mainly of students with an average age of 26. On a scale 1 to 10, the average level of familiarity with tag cloud was estimated at 4.

Their summary of their findings, a comparison to other sources on the subject, was presented mainly in a table (p.403). They supported authors of other publication in the area of similar research and through their experimental project came up with the following conclusions:

Tag size - the bigger the better, size affect ease of finding. In addition they stated that number of characters as well as closeness to other clouds helps enhance the viewer’s experience.

Scanning - rather than reading, users scan tags. This was been proven thanks to eye tracking data.

Centering - authors stated that “tags in the middle of the cloud attract more user attention then tags near the borders” (p.12). For most tag clouds, user attention increases with Circular layout and decreases in Clustered layouts

Position, authors confirm other findings and literature sources that the upper left quadrant is better recalled and is easier and quicker to find.

Layout, eye tracking data supports the idea that the layout of the tag cloud plays a role in user’s perception.

Exploration, Lohmann et al. (2009) stated that, “tag clouds provide suboptimal support when searching for specific tags” (p.12).

Evaluation studies of tag clouds conducted by Rivadeneira, Gruen, Muller and Millen (2007) provided similar finding to Lohmann et al. (2009), especially in the area of the effect of font size.
A statistically significant, but moderate, effect of layout formation was found. In addition, authors concluded that “There was no effect of layout on recognition; the layout of the tag cloud does not assist or hinder the recollection of the tags presented” (p.998), although shorter presentation times may affect it.

Knowledge of psychology in the area of perception theory may be helpful to tag creators. Low-level perception can be affected by such factors as font size and location. Authors suggested that focusing on layout may help designers to affect users’ ability to identify categories in tag clouds (2007).

**Color in word and tag clouds**

Batemans and Gutwin (2008) in their research work on tag clouds have not found a clear indication about to what degree color and which colors draw user’s attention the most. They stated that,” If color is to be used, an appropriate color palette should be displayed with clear mappings for the viewer, and colors should be distributed evenly to avoid undesired pop-out effects”(p.201).

Color is important in the generation and perception of word clouds. Background color may improve user experience and increase the chance for better comprehension of information presented in the cloud. When words are presented in darker colors that increase with word frequency then it creates a visual clue that help users to look for information and better recognize frequent and often key words. One of the applications for color in word clouds can be found in using *Concordle* software (*Concordle*).
**Data Visualization in the Clouds and their limitations**

Word clouds are not free from limitations. There are issues that need to be taken into consideration while using word clouds for various tasks. Andy Ramsden (2008) advocates for proper spelling to create more accurate word clouds. Among other suggestion he proposed proper punctuation and also stated that data entered into creation of the clouds can not be cross referenced. In word clouds common words are usually ignored “otherwise the word cloud would be dominated by "and", "to", "the" and other uninformative words” (Ramsden 2008, p.3). Additional issues pointed out by the author are that frequency and importance may not go together and assuming so may be misleading. He also stated that “ambiguous questions could lead to a wider variety of answers, none of which are particularly frequent. “(p.3) He suggested that improvement needs to be made toward grouping similar words, e.g. “learned” and “trained”.

Miley and Read (2011) in ‘Using word clouds to develop proactive learners” advocate that word clouds can be used as learning tool. Their research presented a group of students that benefited from flexibility of using word clouds that were consistent with their learning style preferences. Miley and Read concluded that using word clouds made students more engaged and motivated in learning and gave them sense a of ownership of their work. Their work on the subject suggests that word clouds provide a useful adjunct to other learning strategies but they cautioned that using word cloud must be done carefully. As the main argument for this caution they stated that “(clouds) summarize word frequency and this may not align with word relevance “(p.104).

**Clouds as research tools**
McNaught and Lam (2010) shared their experience in using word clouds and concluded that they can be helpful as a research tool in supporting educational research. By providing quick visualization of some general patterns that can be generated from text, researchers are able to see the common themes in the text and usually find main differences between sets of responses.

Authors suggest that when text needs to be analyzed it should include the full text of each informant. On the other hand, shorter text samples derived from memos and summaries may not provide adequate results (2010).

Retrieving words out of context is seen by the research as another limitation of word clouds. Since each individual word is object of analysis it may be inadequate because “Word clouds treat each word as the unit of analysis. This mechanical manipulation of text is fast but at the same time it can be misleading because, “it neglects the semantics of the words and also the phrases and even sentences the words are composed of “(p.641).

The conclusion delivered by McNaught and Lam recommends that word clouds may best be a supplemental research aid rather than a stand-alone research tool.

**Students of foreign languages and word clouds**

Student of foreign language can also benefit from using word clouds. At the same time teachers have opportunities to improve their teaching methods and effectiveness of their work using the clouds as instructional tools. Work conducted by Baralt,Pennestri and Selvandin (2011) proved these statements. Both students and instructors benefit by using Wordle in the classroom. As the result of applying Wordle in the classroom, students used “more varied vocabulary, more verb tenses, and more accurate grammar
in their writing” (p.20). Instructors of foreign language that used Wordles in writing workshops resulted in a higher level of teaching effectiveness and a more student-centered approach.

Using text and generating it into word clouds provided many opportunities to explore students’ creativity. Transferring speeches and articles into wordles can add value to the learning process. According to Edyburn (2010) using word clouds can enhance process that helps users to gain proper vocabulary and conceptual understanding of text.

**New development in Data Visualization - Sparkclouds**

Another interesting tool that can be used to present information is a relatively new program called “Sparkclouds”. It is combination of tag cloud and sparklines.

Lee, Riche, Karlson, and Carpendale (2010) explored how efficiently people could learn from Sparkclouds in comparison to more traditional Parallel Tag clouds. Lee et al. (2010) found that participants were more efficient with Sparkclouds in “three different types of tasks: specific data tasks; topic trends tasks; and overview tasks” (p.1186). They also liked Sparkclouds better than alternative tag clouds generators

**CONCLUSIONS**

In cited resources for the literature review readers encounter various terms including; tag clouds, tagclouds, tagcloud, word cloud, content cloud. Terminology regarding visual representation of text needs to
be more clearly defined and standardized among authors. Tag clouds and content/word clouds are not the same even if some sources mixed the terms. They have similar characteristics but the crucial difference is in the way they are created. Tag clouds are usually manually assigned names of object but word/content clouds are generated by software. Both can be very useful in content analysis but it is important to know the difference.

Most of the authors agree that word/tag clouds can be successfully used in various forms to accomplish assigned tasks. Generating most frequently used words and showing patterns help users to grasp the concept of analyzed text, provide a better overview and lead to quicker comprehension of text generated into word clouds. In brief: CLOUDS can add value to the process of understanding text.

There are limitations that suggest that word/content clouds should be used in combination with other research tools. Some limitations are: finding and grouping similar words and synonyms; treating words as the data unit units rather than part of something else; word frequency does not always representing importance.

Improvement of word / tag cloud generators could include more features, improved functionality, less “noise” and more effective usage of “stop words”. Those improvements would likely contribute to more precise and accurate representation of text, increased comprehension and clarity of data presented.

**Research questions**
The main purpose of this project was to investigate how people reacted to the tragic events of the September 11, 2001 terrorist attacks. Media reported the events on many different levels: press, television, radio or through the internet. People were overwhelmed by the amount of information. Yet they realized, despite the chaos, that this event had unpredictable global consequences.

What was less accessible and available were reactions of individuals who shared their contribution through their personal blogs. It is valuable to look back and analyze the content of their shared materials. The following key questions address matters connected with personal expression of individuals immediately after September 11. The platform for the reflection and collection of data was the blogosphere:

- How did individuals react to the September 11 terrorist attacks on the as evidenced by their web activity?

- What was happening in the blogosphere in the first hours through 2 days after the attacks?

- What did the bloggers share? What expressions/idioms/words did they use the most?

- How can one best present the data visually to represent the content of blogs?

Time invested into reflection, group and individual work on Summer 2011 Web Sphere Analysis class and previous analytical work on materials collected on September 11 Web Archive were crucial for this project.

Theoretical Context
There are many ways that metadata can be analyzed. Secondary questions dealt with presenting data already collected through blogs. Looking for ways of sharing the wealth of information and deepening understanding of the tragic events plus using ideas and available tools resulted in visualization of content of the blogs.

Blogs can be categorized as a text based computer mediated communication which is defined as:

“Synchronous or asynchronous electronic mail and computer conferencing, by which senders encode, in text, messages that are relayed from senders’ computers to receivers.” (Walther, 1992, p. 52)

Grounded theory is applicable to blog content and the collected data came from real life situations. As stated by Merriam (2009): “Data in grounded theory studies can come from interviews, observation, and a wide variety of documentary materials. “ (p.31). Applicable methods were used to compare one segment of data with another one (Merriman 2009). Visual representation of bloggers’ content provided an opportunity to compare data generated in word clouds using two different visualization tools: Wordle and Tagxedo. These visualization tools allow us to see the content of blogs from different perspectives and to analyze data with a starting point in text which is transferred into word cloud – images for presentation and evaluation.

Data driven learning (DDL) is also applicable to this project. It is important to mention that: “The most common kind of concordance is called KWIC - Key Word In Context”( from cte319.pbworks.com)

Concordance is applicable to word clouds and provides opportunities to enhance learning.

Perception theory, user’s attention, motivation, learning styles and personal preference play roles in cognition and other processes employed in perceiving word clouds. This project was limited and more
study could reveal how those factors affect individuals and how they contribute to a viewer’s comprehension of visualized data compared to text-based resources of the same content.

Blogs presented in Wordle and Tagxedo provided alternative ways of seeing information and approaching content of analyzed blogs from a different perspective.

Methods

Goals and object of the study

My main goal was to analyze content of personal weblogs to answer the question: “How did bloggers react to the attacks of September 11, 2001? Blogs were getting more popular in the late 90’s and into the new millennium, but there is no comparison to the explosion of the blogosphere we have experienced in the last few years. The relatively few blogs surrounding 9/11 presented a manageable number of unique perspectives on the events of those dramatic hours, days and weeks after the attacks.

I had an opportunity to participate in the Web Sphere Analysis class in the summer of 2011 directed by Professor Steven Schneider. That experience affected how I looked at September 11th. Throughout the class, students had an opportunity to analyze content of some of the web sites from the collection of 25,000 web pages archived between September 11th and December, 2001 by the Library of Congress and the Internet Archives.

Analyzing blogs after 9/11 brought new insights of the terrorist act. In studying archived materials the class explored many aspects of web content and learned how to analyze archived materials through various techniques and available tools. Not the least was the need to familiarize ourselves with arcane
terminology such as: object action, producer type, actor type, web action, action object, and content action.

Having some general idea about the web after September 11, enabled me to narrow my area of interest and focus my project on exploring personal blogs and their content in the immediate aftermath, up to 2 days after, of the terrorist attacks. In the Fall of 2011, I applied my research criteria and identified 21 weblogs from the September 11 Web archives. Qualitative research using personal documents are valuable sources of data which historically have included include diaries, letters, scrapbooks; they “give a snapshot into what the author thinks is important, that is, their personal perspective” (Merriam, p.142). Blogs are modern, personal online diaries that fall into this category. Like their pen and paper ancestors, they are a personal expression of their author and may reveal the motivations and ideas of people writing them to a researcher.

**Influence of literature review**

In the process of collecting resources for the literature review limited scholarly materials on the subject were found. Literature in which I was interested in was about data visualization, especially in terms of content clouds. Several articles were identified about word clouds and tag clouds. Insights from the resources for literature review influenced my analysis and provided perspective on application, benefits and limitations of using the clouds in data analysis. Information gathered through process of reviewing literature about data visualization and object of my study suggested that some limitation of word clouds include stop words and noise words. Word cloud generators reflected on and used for this project created clouds that showed those stop words.
Process of data selection

After selecting blogs, I needed to extract data for analysis. Organizing data is an important step in qualitative research. Care at this step is especially crucial, indeed foundational. My literature review as to the limitations of the algorithms used by cloud generators indicated that they are particularly sensitive to the “cleanliness” of the data, e.g. misspellings.

The blogs were in spreadsheet format in a Google document that included links to the archives. Then all blogs were saved in Scrapbook, a Firefox extension that allowed preserving entire weblog pages and then cleaning blogs of content that was not object of my study. Removed data included mainly external links, tables, side bars, and bloggers’ posts before 9.11 and after 9.13.

By no means are cloud generators are alike. The processing and display mechanisms they use are often unique. However there is considerable overlap in features. A second question I sought to examine is: “What is the effect of tool (generator) selection on the analysis and conclusions drawn from a given set of data?” I chose two word clouds generators, Wordle and Tagxedo to compare. I entered identical text from each of the 21 individual blogs as input to the two generators. Wordle and Tagxedo have many similar features, my intention was to experiment with the unique features of each programs in order to conduct my comparison of the word clouds generated. Those unique features included: shaped “clouds” vs. classic cloud; preset images as well as images uploaded and used by me; custom colors; word counts; option for word orientation; available fonts; and rounded or straight edges.

Even something as basic as word counts varied between the two generators, i.e. an unequal number of words was displayed in the clouds when the same data (text) was entered into the two generators.
Counting is not a parameter that the operator can manipulate. This difference in word counts give the reader a sense of how biases and preferences in setting parameters for features noted above could substantively impact the “look and feel” of the cloud generated.

In some cases stop or noise words also referred to as stoplists, were trimmed to represent data and compare content as clearly as possible. Words less than 4 letters were trimmed, with the exception of a few words that were crucial to represent data closely connected to 9/11 events. For example: WTC (World Trade Center), CNN, ABC, FOX, “hit”, “ash”, NYPD, FDNY. They represented direct result of the terrorist attacks of September 11 and were part of description of places, emotions, situations, objects affected by 9/11.
As analysis of content of blog showed (Figure 1) that many people were watching TV during and after the events.

Table below (Figure 1) presents results of entering text of all 21 blogs into *Tag Crowd* cloud software. I marked TV stations in red for better visibility of the data.

![Tag Cloud Image]

**Figure 1**

TV stations were mentioned 76 times total in weblogs in this project.

<table>
<thead>
<tr>
<th>Station</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>21</td>
</tr>
<tr>
<td>CNN</td>
<td>20</td>
</tr>
<tr>
<td>FOX</td>
<td>23</td>
</tr>
</tbody>
</table>

16
Although my goal for the project is to study bloggers personal expression and to find way to present data in a new way that could add value to reflection and understanding of September 11 events, Figure 1 shows that some data elements can rely on quantitative data analysis.

Several tables included in this chapter provide insights about content, word frequency and other factors that help to gain additional insights about the blogs.

As mentioned in the literature review, one of the limitations of the visual text representation in content clouds is often lack of grouping similar words or synonyms. Table 2 gives an example of this.

| Table 2 Concordle software words: “attack” and “attacks”. |
Both examples included here Figure 2, (Tagcrowd) and Table 2(Concordle) contain text of all content from 21 blogs. Table 3 was created by Textalyzer tool available on line. As in other cases, content of all 21 blogs was entered into this tool. Online tools such as Textalyzer
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total word count</td>
<td>15922</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of different words</td>
<td>4426</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity factor (Lexical Density)</td>
<td>27.8%</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Readability (Gunning-Fog Index): (6-easy 20-hard)</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of characters</td>
<td>156827</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of characters without spaces</td>
<td>90686</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Syllables per Word</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence count</td>
<td>2349</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Average sentence length (words)</td>
<td>12.46</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Max sentence length (words)</td>
<td>128</td>
</tr>
</tbody>
</table>
(42 pm =| i think i was in shock earlier watching the news passing it on to friends that didn t have access to the info took about 3 hrs before it really hit me though i m not related to anyone in nyc i still feel for everyone there that s one thing that kind of sucks about being even slightly empathetic started breaking down and crying every few minutes i ve heard a few stories from coworkers one stated that his friend had prepped all last night for a meeting today his meeting was cancelled but his coworkers meeting was not he did not get on the plane from boston but his coworkers did i think he s having a nervous breakdown or at the very least freaking out)

**Conclusion**

Literature review and personal experience with utilizing various forms of data clouds for this project resulted in a different level of understanding and presenting data. As the literature review indicated
data clouds possess limitations. Focusing on differences of two applicable tools using the same data set will be presented in the next section: Analyzing Blogs in Word Clouds.

Summary word clouds generated by both Tagxedo and Wordle deal with a larger data set and therefore present a higher possibility of showing trends and summarizing content of the studied blogs.

The limitations of visual representation of data will be reflected in the differences between the tools that we will see. But isn’t one of the purposes of qualitative analysis to allow the analyst and her audience to experience the “art” embedded within the data? We value the differences in the output of these tools, precisely because the alternative view each generates allows a deeper reflection upon, and understanding of, the blogosphere immediately after September 11.
Project Development and Application

Differences in Word Clouds generators

Word clouds represent a visualization of text data. They can serve as useful tools for data analysis. Web based services that create word clouds operate on the Java platform.

This is how the creator of Wordle, Jonathan Feinberg, describes the way that a word cloud is created:

"Count the words, throw away boring words, and sort by the count, descending. Keep the top N words for some N. Assign each word a font size proportional to its count. Generate a Java2D Shape for each word, using the Java2D API.

Each word "wants" to be somewhere, such as "at some random x position in the vertical center". In decreasing order of frequency, do this for each word:

Place the word where it wants to be
While it intersects any of the previously placed words
Move it one step along an ever-increasing spiral"
Tools that generate “word clouds” differ in how they allow users to display any chosen text. The clouds vary in terms of fonts, word orientation, colors or color themes and also in content that is displayed; especially numbers, common or rare words, their frequency and layout of clouds.

People who use services such as: Wordle, Tagxedo, Tagul, Tagcrowd, Wordsift, Worditout, AbCya, to mention the most well known word cloud generators, are presented with various levels of text visualization. Generally, visitors to those sites paste text into a designated area that is immediately converted into a word cloud.

Word Clouds have many applications. They can provide a valuable tool for data analysis that empowers and encourages their creators and viewers to gain better understanding of any text. Powerful examples are famous presidential speeches that reflect important ideas, key words and data in a graphical way.

This method of text visualization can enhance and nourish deeper comprehension, provide insights and stimulate thinking process. Word Clouds, for example, can be used in academic environments and they can be applicable in business presentations. They can also help to increase creativity and can be used in many other ways such as literacy classes, art programs, illustrations, slideshows and are limited in use primarily by one’s creativity.
In analyzing content of a number of blogs selected for my project I focused on two Word cloud generators; *Wordle* and *Tagxedo*. Both services offer tools that enable a level of modification that allows creative ways of choosing features which can increase understanding of content.

In order to better represent differences between *Wordle* and *Tagxedo* I created a table below (Table 1). It shows major features and differences between *Wordle* and *Tagxedo*.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Wordle</th>
<th>Tagxedo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Font</td>
<td>32 preset fonts, no option to add fonts</td>
<td>25 preset fonts plus options to add more</td>
</tr>
<tr>
<td>Color/Color Theme</td>
<td>14 color themes to choose from plus custom pallets and palette color variation</td>
<td>Over 160 color themes available custom limited to choosing background color</td>
</tr>
<tr>
<td>Orientation</td>
<td>vertical, mostly vertical, horizontal, mostly horizontal, Half and Half, Any Which Way</td>
<td>vertical, horizontal, H/V, orthogonal</td>
</tr>
<tr>
<td>Shape</td>
<td>No option to use shapes. Forms word clouds that differ depending on orientation of words and rounded or straight edges</td>
<td>Several shapes to choose from plus custom shapes available</td>
</tr>
<tr>
<td>Setting up image options</td>
<td>After text is pasted / loaded</td>
<td>Before text is pasted / loaded</td>
</tr>
<tr>
<td>Alphabetical order</td>
<td>Can put words in alphabetical order in word clouds</td>
<td>No alphabetical order</td>
</tr>
<tr>
<td>Saving options</td>
<td>Print, public gallery</td>
<td>Print, Image, web, thumbnail, other</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Number of words</td>
<td>Available</td>
<td>Available, but not for all numbers (skips every few numbers)</td>
</tr>
<tr>
<td>Word count</td>
<td>Word count available</td>
<td>Not available</td>
</tr>
<tr>
<td>Other / special</td>
<td>Option to remove common words in many languages, leave words as spelled, all words in lower or upper case, advanced options available but do not work well</td>
<td>History, zoom, full page display, several advanced options available, words are displayed larger when users will roll over them</td>
</tr>
</tbody>
</table>

Both, Wordle and Tagxedo can be used as helpful tools in text analysis by visual representation of content. This basic comparison provides a few conclusions. Wordle is especially useful when a user needs to organize words in clouds in an alphabetical order, choose custom colors for background and letters appearing in clouds, count words or remove some elements of text such as foreign words. On the other hand Tagxedo enable users to choose shapes, including uploading their own images, offer many more color themes and saving options. Presenting the “Clouds” generated by Tagxedo and Wordle from the same text can provide useful insights into how one can enhance understanding and address individualized viewer preferences and understanding of the text data.

**Focusing on Differences on blogs transferred into Wordle and Tagxedo**

In order to compare Wordle and Tagxedo I decided to use the same content and paste it in both word cloud generators. There are variables used to illustrate how the clouds differ. Several elements may play a role in the appearance of word cloud creations. Images presented below resulted mainly from choice of features such as font, shapes, color, orientation, frequency of words used. It is important to realize
that there are differences in algorithm used by programmers who created them. Therefore even if both services used Java outcomes, most likely, outcomes would not be identical.

A great number of images that are the products of word cloud generations have limited numbers of colors. The reason behind it was to provide more clear examples of two different word tag generators.

The image below, Figure.1, contains content of weblog posted by Ian Miller, with words limit set to 101 in Wordle and used font Lucida Sans.

![Image of word cloud](image)

Figure 1

Viewers may immediately recognize layout of the page and its characteristics. Wordle tends to illustrate words more compacted in the cloud, generally smaller and there are no additional shorter words written in letters as it happened in Tagxedo (example letter “p”, “b” or “e”).
In addition Tagxedo showed more words in larger sizes. Examples of this dynamics can be exemplified by in the cloud such as “loved”, “close” but also in “Ian Miller”. That also showed a difference in how information is presented and what kind of word hierarchy is displayed.

Figure 2

Tagxedo, in the example above, Figure 2, may present more key words. Wordle’s example of Ian’s blog content helps viewers to easily see fewer key words but recognize most frequently used. They are likely more crucial to analysis of the blog. Wordle displayed numbers but they weren’t presented in Tagxedo cloud.

More detailed analysis of these two clouds could also show which words appear in one image but not in the other.
To bring the comparison under closer examination, next two clouds were set up to display 124 words.

This time content was used from Anil’s Danish weblog, Figure 3, in black letters on white background using same fonts as in previous comparison.

Wordle

![Wordle Cloud](image)

Figure 3

Tagxedo

There are no numbers in Tagxedo. Words are more even size wise in terms of numbers of enlarged words in both clouds. Interestingly the word “just” exists in Wordle’s cloud only. Similarly to the previous comparisons numbers are only presented in Wordle. Looking at both clouds viewers can see that “friends”, “work”, “city”, “think”, “know”, “want”, “kind”, “pretty” tend to be placed in similar position in both visual representation of Danish’s blog content, Figure 4.
Figure 4

Vertical layout of Wordle (left) and then Tagxedo with limit of 151 words and shades of blue can serve as an example of what was of creation required to attempt to close, as much as possible, a gap between the two.
This time custom color option, Figure 5, was selected in order to provide a closer comparison example in Tagxedo, Figure 6. Ian’s Miller blog that had been pasted into word tools required more setting up and modifying in order to show similarities. Once more, despite of higher number of allowable words, viewers can see more clearly differences in color, sizes, and layout. Words “September”, “Miller”, “occurred”, “practice”, “light”, “commentary” just to point out a few notable discrepancies in the way how information is presented. Words, as in the previous example, differ in wider spectrum of font sizes.
One of the features that set apart Wordle and Tagxedo is option to form word cloud into shapes. 

“Classic Cloud Vertical” and “Classic Cloud horizontal” are available in Tagxedo. They are available to create basic clouds as in Wordle. Advantage of Tagxedo comes into play when user wants to create word shape. In addition to two classic clouds, Tagxedo provided 113 preset shapes in addition to two classic clouds. Since images can be uploaded by users, there are many opportunities of creating various shapes. It is important to acknowledge some limitation users may encounter while using Tagxedo software. Not all images will result in clouds that will provide equal opportunity and clarity for viewers.

In this part of my work I will present several shape based word clouds, process of creating them, possible issues with text visual representation. I will emphasize opportunity associated with using shapes in Tagxedo as well as show example of feature present in Wordle that other service lacks.

After creating a simple Microsoft Word document that only contained “9/11” in very large size of font I saved it as image which later was uploaded to Tagxedo, Figure 7.

9 / 11

Figure 7
Quoting author of Xkot.net may help to comprehend why words such “whatsoever” appears in larger font or notion of several words:

“"Finally, brethren, whatsoever things are true, whatsoever things are honest, whatsoever things are just, whatsoever things are pure, whatsoever things are lovely, whatsoever things are of good report; if there be any virtue, and if there be any praise, think on these things.""........

“I am going to try not to make any more entries about the horror of Sept. 11th for a while. I am going to try to focus on the good things in life. The things that we have to be thankful for.” posted Xkot.net blogger.

Figure 8   From content of Xkot.net blog into “9/11“image, various orientation, limit 100 words

Figure 9   Blog Xkot.net. limit 200 words, combination of fonts, variable orientation
Quantity of words used in shape based cloud does play a role in presenting information where a creator wants to see the importance of words and image in the same visual representation of text.

Words orientation in shape based cloud generated by Tagxedo also contributes to viewers overall perception. Next two examples also based of the same text. Xkot.net blog illustrated the idea. Larger, more frequent used words combined with horizontal orientation created a cloud with distorted image. Viewers have a chance to connect content of the cloud with what the cloud’s creation is trying to represent. Their perception can, most likely, be improved by using both vertical and horizontal orientation as shown in the next image. This experience also shows that orientation of text in cloud can play an important role in final generation of a cloud.

Figure 10 Blog xkot.net, not fully shaped, 200 words White on Black, font: Constantia, horizontal
Figure 11 Blog – Xkot, 9/11, 200 words white on black, font: Constantia, horizontal plus vertical
The process of creating word clouds in the same a session in Tagxedo can be tracked thanks to an option called “history” located on the left side of the page, Figure 12.

Figure 12

Occasionally, leaving particular concept or intention of how the creator wants to present information and trying something completely new may bring something meaningful and useful. On the other hand it may prove that lack of contrast, poor combination of colors or lack pf proper background to particular color of letters may result in giving up new experiments and returning to previously created cloud.

Word cloud creators can use text to illustrate and understand image. On the other hand images can help to understand text.
After using “9/11” text converted to an image that I used to generate a word cloud I decided to experiment with more sophisticated graphics or photographs.

First one is taken from “Likenorb” blogger that is one of the web blogs analyzed for this project.

Since the blog did not contain too much text but many very interesting images I used it to present a great opportunity by combining image and text and converted it to a word cloud.

Figure 13
To the Left – Xkot.net now ad for always 150 words, Xkot.net shape 150 words, theme: Tee Hause Color theme; “Elephant in the room”

Both clouds contain combination of three fonts; Paper Cuts, Lane A and License Plate. Cloud on the right may provide better contrast against the black background. In both cases, knowing what the source of information is may likely contribute to understanding reason behind the shape. In “9/11” word cloud viewers are more likely to find meaning behind the image.

In most of the blogs selected for this project “Word”, “Trade”, “Center” (WTC) was mentioned often.
Summary clouds in both text word generators showed how present they are, Figure 16 and 17. These summary clouds were created by copying and pasting all content of all analyzed blogs combined together.

Words presented in summary clouds; Trade, Center, Tower, Towers, WTC, World,

Figure 16
Figure 17
The terrorist attacks of 9/11 are predominately symbolized by the image of the World Trade Center. The crippled towers represent so many ideas, opinions, actions, feelings that are very well reflected in the weblogs. An image of the towers taken before 9/11, found on the web, has been used for number of world clouds associated with this project.

Figure 18
Image from europe-virtual.com
As in previous clouds generated by Tagxedo I uploaded image this time of WTC, Figure 18. Both word clouds, Figure 19 and 20, were created based on Xkot.net blog. Whether words are wrapped around the towers or they filled up the towers those images may well be suited to inspire people, motivate them to reflect about the tragic events or symbolize many ideas and thoughts connected with September 11th, 2001.
There is one more Tagxedo cloud that I decided to present here, Figure 21. This time I chose to combine words that are in various orientations.
Xkot.net’s two towers, street view W on B, various orientation and fonts

On September 11 and after the tragic events many people expressed their patriotism. American Flag unifies Americans and it is one of the most recognized symbols of the United States of America, Figure 22.
Using features of Tagxedo allowed creating world cloud based of content of Xkot.net blog.

White words on black background in the word cloud shaped in American Flag generated two images that are presented below.

Figure 23

After several attempts I finally decided to keep horizontal orientation of words, Figure 23. That provided, in my opinion, the best representation of the cloud and words used in Xkot.net’s blog.
Some variation of colors helped to present information differently, Figure 24. Ideally, I would have created a visual representation of the cloud using colors that are present in American Flag. Since it was not something I could custom made in Tagxedo I decided to use a comparison word cloud generator in Wordle.

Next three clouds are created with custom made colors based of colors present in the United States Flag. Using opposite background colors give creators additional opportunity to address users preferences. Some of the clouds in patriotic colors have rounded edges.

Figure 25 is presented below, Xkot.net, Patriotic, Wordle, alphabetical order, while background
Next, Figure 26, contains of Xkot.net, Patriotic, Wordle, alphabetical order, no rounded edges, black background.

Figure 26

Figure 27, Xkot.net, Patriotic, alphabetical order, no rounded edges, white background.
Changing any element in the cloud and setting process for generating world cloud affect the final creation.

Exploring various shape options either those preset ones or custom to be uploaded differentiate most significantly Tagxedo from Wordle. Using various images in this project seems to be the best way to show differences in both word cloud generators.

On September 11th the whole world had seen a new “face” of terrorism and the word “terrorist”. Terrorists generated many reactions, emotions and changed their meaning.

Bloggers expressed, often angrily, their feelings and thoughts about those who attacked America on that well remembered day in 2001. This is expressed on Simpson’s Homer blog.

A picture represented a terrorist, Figure 28, served as material for further exploration of my project.
Although the image’s impact speaks volumes, adding words to the image brings the message to another level.

Using “Chunk Five” font or combination of fonts, in addition to varied words orientation or horizontal one affected how the terrorist shaped by the word clouds can be perceived.

This is how Homer Simpson himself describes the situation: “a car bomb just went off in front of the US Senate...shit...i’m on edge now.

posted by Homer Simpson at 8:06 AM” ...

damn man...that's a lot of fucking people. the second tower collapsed too. this is a fucking shame.

posted by Homer Simpson at 7:48 AM“.

Figure 28

Original image taken form the web at www.securityofficertraininghq.com/
Despite of lack of shape based word clouds; Wordle does offer several benefits and features that can contribute to text analysis.

Ian Miller’s blog, one of which was a focus of analysis for this project, was converted into a word cloud.

Custom setting of Wordle allowed me to put words in alphabetical order, exclude the numbers, and apply custom picked colors and background.

It all had a purpose to generate a cloud that would bring order and clarity. Looking at the cloud, viewer can see that words’ sizes are proportional to the frequency in which they appear in Ian’s blog. Applying
only three colors had a goal to make the cloud more easily readable, Figure 31. Having words in an alphabetical order makes looking for words of interest much easier that in doing it in Tagxedo where this is not available.

Ian Miller, Wordle, alphabetical order, no numbers, custom colors

Figure 31
I transferred Wordle’s world cloud, Figure 32, into an image and uploaded it to Tagxedo to use it as shape.

It is up to viewers and their individual preference to determine which image speaks to them the most or whether one presents information in a way that helps to understand it better. Wordle offers a unique way of words orientation called “Any Which Way” that Tagxedo does not provide. Wordle word cloud was transferred into an image in my experiment and that way a new creation includes both: shaped cloud in Tagxedo where words are in “Any Which Way” and cloud image. Viewers can still see differences in words’ sizes and inequal appearance of words that still comes from the same source: text from “Xkot.net” blog.

Wordle’s visual representation was created in “Blue meets orange” theme. Colors may influence viewers experience as well as words’ sizes. To see which would result in a better viewing experience and the presentation of a blog in a world cloud would require further studies which is not part of this
project. Main purpose of this exercise was to explore capability of both software programs and
explore ways how they can intersect, Figure 33.

![Figure 3](image)

Events of September 11, 2001 caused many unexpected reactions from the public. The effects
of September 11th changed lives forever. For some the world changed completely, for others it
made them pause or leave of whatever they were doing to watch how things went unexpected and unpredictable. Even simple tasks, daily routines needed to be completely abandoned.

Maganut’s blog left a mark of one of those changes and showed how somebody’s live was impacted by little things. In her blog we can read, “I’m going to wash my coffee press and brew that cup of coffee I never had yesterday”.
Maganut’s blog left a mark of one of those changes and showed how somebody’s live was impacted by little things. In her blog we can read, “I’m going to wash my coffee press and brew that cup of coffee I never had yesterday”.

Finding a cup shape in Tagxedo resonated in creating new word cloud. Another sentence from her blog describes an impact of the attack and a way some people dealt with the tragic events. 24 hours later, “I’m heading back into the kitchen to finish up the dishes, to pick up the spatula that still sits in the sink where I dropped it”. The author watched TV for many hours. In general words that have less than 4 letters are candidates to be avoided. On many occasion they take space and do not bring anything new into a unique “conversation” or word representation that happened in the cloud. In Maganut’s blog “TV” presented an important piece of information therefore it is in the cloud.

Figure 34 Tagxedo, cup shaped word cloud.
Next cloud based on the same text as used in cup shaped Tagxedo but it is created with Wordle, Figure 35. I chose colors that somehow can be associated with coffee. Color of napkins or coffee cups, color of coffee beans plus background that may be associated with coffee helped in color selection.

Both clouds represent Maganut’s blog but a cup brings it to a different level, Figure 34.
Two clouds show Andy Chest’s blog content in Wordle and Tagxedo clouds. Both have similar shapes and numbers of colors. The clouds provide method of presenting words sizes accordingly to the frequency in text. Horizontal orientation of words in Tagxedo may help in reading process. To give an example of different between the word cloud generators I wanted to point out a word “e-mail” that appears only in Wordle. The same text was loaded into both software but only Wordle presented in. A few example of content of Andy’s blog:

“First and foremost I want to thank everyone for their kind and concerned phone calls and e-mails. It’s wonderful to hear from so many of you. Thanks and thanks again.” Another instance where e-mail is mentioned and it seems important where the word is placed in the contest;

“Secondly: I don't know why, but on the "L" Train back to brooklyn, I had a small panic attack.
It makes no sense. But I was afraid of being trapped. That's never happened before. I need to get back into the city, back with people I know. I'm just catching up on e-mails and phone calls to make sure everyone's okay. “

This analysis of content of Andy Chest’s content based word clouds suggest that on occasion it may be helpful to use more than word cloud to present information.

Andy Chest Tagxedo, Figure 36
Last two blogs that I decided to mention in this section are characterized by showing differences between regular clouds in Wordle and shaped based clouds in Tagxedo.

Anil Dash’s cloud tree may appear to viewers in many ways. Often branches in trees symbolize connection, relation. This is what Anil writes in his blog,
“I have had so many of the same conversations over and over. I want so badly to think of something else, but I just can't get past this”. Another sentence may be also well represented by symbol of tree,” Those of you who know me well know that I’m passionate about this city, that I love the architecture and the grandeur and the dimension of it all. That, actually, may be the hardest part for me, as I don't have anyone I'm close to who I've lost. I have just lost innocence about the city I love, the city I spent my whole life trying to get to.”

A cloud of thoughts in Wordle also gives viewers reasons to make connections of what they see in the cloud and another window of opportunity to analyze what they see.

Anil Dash Tagxedo tree, Figure 38
How do people act in the face of terror? How do they respond? What actions do they take to deal with the situation and to help to make things better?

Blogs that are part of analysis for this project provided interesting insights to the questions above.

Nichole writes in her blog,

“9.11.2001
I did not get to give blood today.

I'm sure they will have a blood drive at work….”

“I feel I've done a good thing

I have not yet given blood

they're having a blood drive at the hospital near my house one Friday - I had planned on going down then

but then I had heard that they only want o+ & o- blood.

so I don't think they'd want mine =|

but I did a good thing anyway.

I had visited endeffect and saw the links that precurer had posted - one of them being the Red Cross Relief Fund.

I took advantage of the fact that there are ~7k people here at work - let alone the campuses around the country, and forwarded the email to over a dozen people - one of which passed it on to all of his techs that work for him, the other lead techs, and management.

thanks Dave.

I'm sure that the url will be passed around quickly and will help raise money to help.

I feel a bit better.

Thanks precurer.

posted by Nichole Smith at 11:42 AM” …
This post touched on, in my opinion, some of very crucial aspects of blogging. Blogging is the result of interconnected people who all have the desire to share information, ideas, and opinions. But it also used the information and transferred it into concrete actions. Giving blood, asking others to give blood, informing others how and where to do it can be a life saving effort. Blogging was something more than sharing personal expression. It was a way of using relatively new web platform to change reality, to connect, to create places of belonging and to report among many other things.

The word “posted” appeared on many occasion in world cloud and it happened on purpose. “Posted” was and is a contribution to many aspects of September 11 events. “Posted” resulted in actions, expression and provided opportunity for deeper understanding of the events and people expressions of those tragic moments. Contributions of bloggers definitely changed how I see and understand September 11 and I believe it can help others as well.

Plane became another symbol of September 11. It carries so many words, Figure 40. It represents so much terror, pain, destruction, death. At the same time it carries many gestures of solidarity of human race, sacrifice, heroism, hope and many others.

Nichole Smith’s blog transferred into Tagxedo.
Nichole’s Wordle representation of her blog shows clearly key words crucial for what she wanted to express.

Having some insights about what she posted in the blog helps to comprehend messages expressed by her. “Need”, “blood”, “think”, “coworkers”, “reports”, “towers”, “hit”, “cancelled”, “meeting” are among other words that create a unique story, her perspective and perception of what happened.

Nichole’s Wordle, Figure 41
Conclusions

Entering the same data, in this project, content of selected blogs in first 2 days after terrorist attacks created an opportunity to conduct analysis of differences of both Wordle and Tagxedo. Even analysis of limited data, provided a base to see what they can offer and how they can help in the process of receiving, organizing and comprehending information.

Tagxedo advances over Wordle by generating shape based word clouds. It enables users to utilize feature to further modify cloud characteristic that, after some practice, may be insightful to viewers and provide new perspective that enhance learning. Saving and sharing generated clouds provide additional advance for users as well as viewers.

Wordles’ features that contribute to better information comprehension of text in word clouds include alphabetical order of words, options of choosing custom made background and word colors as well as word count.

Both services offer visual representation of text that can be used successfully and benefit recipients of word cloud images by stimulating thinking process. Using both software programs for the same text data helped me gain more insights and come up with more educated conclusions.
Future Research

Working on the project and analyzing content of the blogs, exploring what other researchers have studied about content clouds as well as using various software tools was an inspiring experience and generated new ideas. There are many levels on which subject of word clouds and visualization data could be further explored.

Comparing more tools that offer generation of the clouds and using the same content could provide a broader perspective and possibly new application for visualization of data. Comparison of various algorithms and applying knowledge of computer science to expend this project could definitely extend this project. It could also provide better applications and solutions for dealing with “noise” and “stop words” in visualization.

More studies on eye tracking data and role of cognitive psychology could result in practical and theoretical implication that could be used by scholars as well creators of data clouds.
REFERENCES


Concordle, Concordle - Not so pretty cousin of Wordle, http://folk.uib.no/nfylk/concordle/


europe-virtual.com, Retrieved on April 15, 2012


Tagxedo, word cloud generator, [http://www.tagxedo.com](http://www.tagxedo.com)

Troutner, J. (2011). QR codes, infographics, and a host of cool tools to impress your colleagues and
improve your work. *Teacher Librarian*, 39(1), 43-45,47.  
http://search.proquest.com/docview/902627381?accountid=26969
