

Using Word Clouds in Teaching and Learning

Andy Ramsden
Learning & Teaching Enhancement Office
e-learning Team
University of Bath
www.bath.ac.uk/learningandteaching
a.ramsden@bath.ac.uk

Andrew Bate
Learning & Teaching Enhancement Office
e-learning Team – Student Intern 2008
University of Bath
www.bath.ac.uk/learningandteaching
amb28@bath.ac.uk

Information Sheet

This working paper was written by members of the e-learning community at the University of Bath. If you'd like to access other working papers, presentations or posters given by the e-learning team at the University of Bath then see our Online Publications Store, <http://www.bath.ac.uk/opus/handle/10247/450>.

If you have any questions about this paper then please contact the authors directly.

© University of Bath, e-learning, August 2008

Acknowledgements

We'd like to thank the internal and external reviewers who took the time to read and comment on earlier drafts of this paper. We've taken on board many of the points raised, and side stepped a few others. However, all were greatly appreciated, thanks ☺

Introductions

The aim of this paper is to introduce the potential of using word clouds within your teaching and learning. In particular, it will address what a word cloud is, what we should take into consideration when using them and finally, compare different word cloud software.

What is a word cloud?

A word cloud is a "visual depiction of words. The more frequent the word appears within the text being analysed the larger the word becomes." In essence a word cloud "plots" word frequency by the size of the word.

The following example of a word cloud was generated using TagCrowd (www.tagcrowd.com). It is generated from 50 responses on a paper survey to a group of educators at a Podcasting for Pedagogical Purposes workshop in June 2008 (PPP SIG). The question was "*How will members work and learn together?*".



You very quickly get a sense of key words. For instance, if you collate the top five (or in this case with the tied frequency, the top six) words, then in response to the question "how will members learn and work together", the most frequently used words are;

- wiki
- sharing
- meetings
- online
- face to face
- blogs

The results clearly identify two tools of preference (wiki and blog), and the key process of sharing via meetings, which could be online and/or face to face. This illustrates the effectiveness of this type of analysis.

A question from you might be, how might they be used in an educational context, outside of survey evaluation? One potential use could be in gathering informal feedback during large group teaching. For instance, you might use the SMS EduTxt Service to encourage students to answer questions set during the lecture. This might even be in the form of a multiple choice question or one/two word answers. Then you can access the results via the web interface, cut and paste these into the word cloud generator and you have a visual depiction of the responses within a minute. A worked example of this approach is discussed at Andy Ramsden's Mobile Learning Blog (www.mobile-learning.blog-city.com).

We've demonstrated the potential of word clouds. However, there are a number of caveats that you need to consider when deciding if word cloud analysis is appropriate for you.

What issues do we need to consider when using word clouds?

Firstly, the data needs to be in a meaningful state. In this case the survey was completed on paper and was then converted into an electronic format. As would be expected this process led to significant problems with deciphering handwriting. Even more problematic is the issue associated with working with unstructured data. For instance, some responses included cross referencing of answers between questions. This made the digitisation of the content difficult. After the data was digitised then the data needed to be adapted for the word cloud software.

Secondly, you need an appreciation on how word cloud software works in terms of spelling and punctuation. This meant making sure that the spelling was correct, and standardising so every word only has one spelling. Also, punctuation can have an impact on outcomes. For instance, typically hyphens and apostrophes are included and interpreted as linking words together. The software normally ignores common words by default or as an option. This would be recommended otherwise the word cloud would be dominated by words like "and", "to", "the" and other uninformative words.

Not only do you need to account for this when reviewing the data before the analysis. But in some cases there is something specific to the survey which means you might need to change the data. For instance, in the PPP SIG case terms like IT and HE occurred quite frequently, however they were indistinguishable to common words and thus ignored. To resolve this, IT was replaced with "ICT" and HE with "Higher-Education" (and for consistency FE became Further-Education). Overall, if you are using a particularly large survey then a significant resource is required to standardise the data.

Thirdly, there are also some generic issues to consider when using Word Clouds. For instance, there is an assumption that the frequency and the importance of a word are one and the same thing. This is not necessarily the case and could depend on the application and context, if questions in a survey were easy to understand and not particularly open ended, then you would expect the more important points to occur more frequently, whereas ambiguous questions could lead to a wider variety of answers, none of which are particularly frequent.

In addition, word clouds do not necessarily portray the context of the word. For example, "I did not find the lecture interesting" will add another use of the word "interesting" to the text. This will result in "interesting" was greatly used to describe the lecture, when in this case an antonym like boring or uninteresting was the context, "I found the lecture boring". On top of this, word cloud programs normally fail to pick up phrases as they tend to only focus on single word frequency, again, reducing context, especially if the phrases are idioms or oxymoronic.

Finally, word clouds often fail to group similar words. This means words that all have the same or similar meaning or connotations. For example, learning, learn, learnt, and learned would appear as separate words in the word cloud. If all these words were grouped together, then the "group base word" "learn" would appear once but larger in the "plot", giving the word more prominence. Some software have developed grouping functionality, however, the grouping around the term "learn", would not encompass synonyms like "be trained" and other related words. This means that you still need to group up similar meaning by looking at the word cloud and seeing the prominence of each individual word then sum up all the words with similar meanings. For instance, in the previous example the software might group people & person, however it is not likely to include the word, member.

A method to minimise the effect of these issues would be to use Word Clouds in conjunction with other ways of expressing qualitative data, to form a triangulation. The triangulation process might include a summary of the textual responses from the survey. This method is a little time consuming, but is very likely to get the context of the text to help interpret the word cloud. Taking several complementary methods into consideration improves the reliability of the results.

To date the discussion has highlighted that it is evident that a significant amount of work needs to be allocated to prepare the data to give a meaningful outcome. However, it is clear that when appropriately used word cloud analysis offers benefits. For instance, another question asked on the PPP SIG survey was "*What is the purpose of the community?*" The word cloud is below;



It is evident from the 50 responses that the purpose of the community is to share (20) ideas (24) and practice (15).

What software to use?

So in practical terms how might you use word clouds at the University of Bath? At the moment there is no word cloud software service hosted at the University of Bath. Therefore, after you have collected and manipulated the data, you will need to use one of the many freely available externally hosted word cloud generators. The following table compares four of the most popular. You are strongly advised to try them all, to find the one that best meets your needs. A good starting point would be TagCrowd or Wordle. Please note that this is a rapidly developing area, so software functionality will change and other software providers will emerge into the market. As a result, the comparison table is likely to date quickly.

	TagCrowd	MakeCloud	ToCloud	Wordle
URL	www.tagcrowd.com	www.makecloud.com	www.tocloud.com	www.wordle.net
Final look: Simple(S)/ Complex(C)	S	S	S	C
Input: a webpage(W), cut & paste text(T), File upload(U), RSS Feed (F)	W, T, U	W, T, F	W, T	T, F
Functions: Shows Frequency(F), Phrases(P), Ignores Full- Stops(I), No. of words in plot(N), Minimal Frequency (M), Ignores Common Words(C), Stoplists(S), Ignore Capital Letters(L)	F, I, N, M, C, S, L	I, C, L	F, P, N, M, C, S, L (unless only caps spelling used)	N, C
Output format: Webpage Display(D), Save as Image file(S), Printable (P) embed into blog/website(E)	D, P, E	D, E	D, E	D, P, E
More details	This is fairly sound and simple. Can show frequency and group similar words. This is the yardstick which others are compared.	Fairly poor both with a lack of options and lack of number of words plotted, even from large text. Can however take from RSS feed (although so can wordle, although you will not get tags and the design will be more complicated.)	This is fairly sound and simple like tagcrowd. Has the ability to give the plot a title, show frequencies and to include phrases (or even phrases only), giving better context. This is a little flawed since it ignores common words before making phrases, like	Put in text or RSS feed, then once plotted, can edit the setting like use of caps, and no. of words and ignore common words from a choice of many languages. You can also change the font, colour and general direction of the

"not", so is more applicable to names. In plot, phrases are red and the more frequent words black, rest are green. words in a straightforward manner, making it a lot more personal and visual than the others. It has good use of punctuation. It is a new site and is under development. Plots can be saved into a gallery for public use, and from there can be embedded into blog/webpage.

Conclusion

In conclusion, the discussion has highlighted the potential of word clouds, it has elaborated on some of the issues that need to be accounted for in their use and it has outlined some available software solutions.

An interesting point raised by one of the external reviewers was, “is it technology that is easily transportable into class, with the effort required over and above what people already do?” I’d suggest that there we need to be aware of this rapidly evolving technology for our teaching and learning. It is likely that word clouds will evolve inline with other web services. It will become simpler to use and more sophisticated in how it analyses the data. Therefore, the changes in the technology will mean that overtime less effort will be needed, so it will become more readily transportable into class.