Web 2.0

Phil Bradley

Word Clouds and Word Cloud Software

This month's column looks at word clouds: software that reveal frequently used terms from a page or site – and some surprising applications for word clouds.

Introduction

When I run courses about Web 2.0 technology one of the resources that always proves very popular are word cloud packages, so I thought that I'd share some ideas with you, together with some packages to try out, and of course – things to do with them! I mentioned Tagxedo last time around which people liked, so I thought I'd develop the theme.

If you're unfamiliar with the idea of a word cloud it's actually quite straightforward. Whichever resource you use will want to have some text – either a website URL, a blog address or just some text that you post into a dialogue box. The application then runs through the data, basically adding up the number of times that words have been used. It then uses this knowledge when producing the word cloud by ensuring that the more times a word is repeated, the larger the word appears in the cloud. Stop words such as 'the' are of course excluded. You can then decide on the layout, colours used, font and so on, and keep exploring until you're happy with the end result, which you can then save.

Before looking at the applications you can try, let's start with a look at what you can do with word clouds.

Word clouds – practical uses

What are you really saying? I was listening to a particular politician give a speech during the last election and he was a good orator – gave the speech well, sounded encouraging and upbeat and left people cheering. I got a copy of his speech from the BBC and ran it through a word cloud generator, and the words that leapt off the screen were things such as 'cannot', 'failure', 'disappointment' and so on. Now, it could be argued that he was actually saying things like 'we cannot have failure, or provide the country with a disappointment' (context is

after all everything), but on the other hand, negative words do have a subliminal importance and value. Word cloud generators (from here on in I'm referring to them as "Generators" for ease of typing and reading) do give you an opportunity to look at what you're saying in a rather different way. If you have a CV ready to hand, run it through a Generator; what are you actually saying? Take speeches from US presidential state of the union addresses at different time periods – what key words come out, and how do they reflect the times?

- Identifying works. Within an educational environment you can have a lot of fun with Generators. Grab the text from a play, or the first chapter of a book, or even visit Wikipedia and take the titles of a well-known author and turn the content into a word cloud. Can you identify the book, play or novelist? I created one of these for the works of Agatha Christie and 'murder', 'dead' and 'death' were the words that leapt off the screen at me, unsurprisingly.

- Make posters. Once you've got one of these word clouds, turn it into a poster, which is going to look much more fun and interesting than straightforward pictures. Alternatively, get really jazzy and print them onto T-shirts; I did this for a friend of mine who really likes the works of Jane Austen, and the resulting item of clothing was a real hit on the schools next 'dress down Friday'. Instead of the usual Christmas or birthday cards you could create an interesting and personalised card by using appropriate words or phrases that relate to your mother, sibling or friend, perhaps.

- Feedback. This is something I've seen quite often – take the evaluation forms for a course or conference and turn it into a cloud to see what people really thought about the event. If you send delegates a 'what do you want to get out of the course' questionnaire you could do the same thing and produce an interesting 'before and after' word cloud.

- Make a report more interesting. Use the text from a company annual report, a syllabus, a piece of legislation or any other 'dry' material and jazz it up.

- Run student essays through a generator, anonymise them, display them and discuss which words are common across them, and which are not.

- Pro- and Anti-clouds. Take an event, find two people who have written about it from different angles and compare the clouds. Which words or concepts are common to both approaches, and what are the differences?

- Run a poll and take the answers and create a word cloud of out them. It could be something as simple as 'what is your favourite colour?' to 'what ethics should information professionals hold dear?'

Some generators allow you to choose specific shapes for a word cloud. I did a really interesting session with some school librarians where we took the Wikipedia text of the Elizabeth I entry and turned it into a key shape. We were then able to discuss how 'marriage' as a concept works in conjunction with keys and locks. We then did the same thing again, only this time by turning the word cloud into the shape of a heart.

If you're trying to decide what to do first, put off the dreaded moment for a bit longer by writing down all your tasks, run the text through a Generator and see what comes out. Or do the same thing, but this time do it for a syllabus or curriculum.

Highlight your own blog or website. Create a word cloud based on your news feed and use it as an image header perhaps. Or fade the words out and use them as the background to a business card or invitation.

Word cloud generators

Tag Cloud at http://www.tagcrowd.com/ is a free alternative that can be used without any kind of registration. It's simple to use, though more limited than Wordle (see below), since users can't fine control the output. However, results can be saved, embedded onto blogs and websites or kept as a PDF. This is fine, but not that exciting. Users can't control the colours or positioning, but they can choose the maximum number of words to use, include a stop word list, and can group similar words together.

Wordsift is at http://wordsift.com/. It does give you more options than a simple word cloud. Words can be dragged around the screen, for example, so if you don't like the way in which something is displayed you can always change it. Sorting is also possible – common to rare, rare to common, A-Z, Z-A. The tag cloud words can also be marked up, either custom, or according to categories such as language, science, maths or social studies. This simply pops the words into a different colour, but it's a very useful training aid in the classroom.

WordSift does more than this, however: there is much more detail available to users. Under the word cloud there are images appropriate to the words in the cloud, and these can be clicked and dragged into a sandbox on the same page. Users can also search for a particular word in the Visual Thesaurus that's on the site, in order to see the relationships between different terms.

If you're a teacher, WordSift is a tremendous resource, and there's no registration required. It's also worth saying that it's been produced by Stanford University as well, so it has a good pedigree. The main disadvantage that I can see is that it's not easy to save the word cloud unless you use a 3rd party application. That's disappointing.

ABCya! at http://www.abcya.com/word_clouds.htm is fun. The cloud can be randomised, the fonts changed, the colours altered, the layout can also be changed. No registration required and it's free.

Tagul at http://tagul.com/ is a similar resource, but with significant differences. Mouseover words and they will flick into horizontal and will be zoomed, so it's quite exciting visually. Clicking on the term you've chosen will then lead to a Google search for it, but it's possible to link to other sites or pages instead. In order to use the service registration is required and you're limited to ten clouds. Words can be pasted in to be used, or users can provide a URL

instead. Users can choose the shape of the cloud, font, or angles of words. Clouds can be embedded, but the resource does require some work and thought to get the most out of it. If you're a teacher you'd be best advised to prepare the cloud prior to a lesson, unlike other resources that will just pull up a cloud for you.

WordItOut at http://worditout.com/ is another resource that you can just use directly, without registration. Words can be pasted, or you can provide a URL to the application. Colours can be changed, as can font, size, number of colours, but layout doesn't appear as an option.

ImageChef Word Mosaic at http://www.imagechef.com/ic/word_mosaic/ allows you to create a cloud in a specific shape – a heart, exclamation mark, envelope, skull and so on. It's quite a 'pop' type site, with options for changing shapes, colours and fonts. Very easy to save, share on Twitter, email and so on. If you're working with children this is certainly worth exploring.

TagCloud builder at http://tagcloud.oclc.org/tagcloud/TagCloudDemo is a project of OCLC Research. Another fairly basic resource, it arranges words alphabetically, so it's not very exciting visually. You can however change colours, group like words and save the results.

If you're aware of any word cloud generators, the one you'll be most familiar with is Wordle at http://www.wordle.net as it was probably the first of the generators (though I could be wrong!). It's simple, easy to use and very friendly.

However, my favourite is Tagxedo at http://www.tagxedo.com which you'll recall I mentioned in the last column, so I won't repeat myself. I did however use it to create a very simple cloud (see figure 1 below), which you can view at http://www.flickr.com/photos/39329798@N00/4982196103/ or visit the Tagxedo pool itself at http://www.flickr.com/groups/tagxedo/pool/with/4982196103/ and if you create any yourself using Tagxedo do please think of adding them!

Figure 1 Wordle word cloud

Summary

On the face of it word clouds are a simple, fun and enjoyable way to waste a few minutes. However, they can be so much more than that – really useful tools to explore content in entirely new and fresh ways. They really do make you consider a subject an entirely different fashion. Moreover, they're easy, quick and simple to create, so give them a go.