

A tag cloud (or text cloud) offers a visual sense of the key terms in a piece of writing at a glance. Each cloud visually represents the number of times a writer uses certain words in a particular piece of text. The words are listed alphabetically and weighted to allow easy identification of the use frequency for each key word. The larger the font size, the more frequent the key word.

Evaluating Written Texts

Here are two examples generated from George W. Bush's first (2001) and last (2007) State of the Union addresses. (By Chirag Metha, they can be found at <http://chir.ag/phernalia/preztags/>.)

2001-02-27: State of the Union Address

George W. Bush (2001-)

abusive administrator affordable agreement bipartisan bless capitol challenges chamber charities china civility
classroom college commission commitment compassion confront congressman conservation crime debates debt
dedicates deficit democrat deserve disabilities diseases earn economy education enforce equality
expand failing fairness families foundation freedom funding global god ideals
independence inflation internet invest invitation john josefina judged loved lowest math mayor
medicare mentor neighborhoods nuclear overtime owe partners philadelphia pledge prescription presidential
priority privilege prosperity punish recruit refundable regardless repaid restore reward science senate seniors steven
strength students surpluses targeted tax teachers teaching terrorists threats transform trillion triple
uncertainty values violence war washington weapons welfare

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable affordable afghanistan africa aided ally anbar armed army baghdad bless challenges
chamber chaos choices civilians coalition commanders commitment confident confront congressman constitution
corps debates deduction deficit deliver democratic deploy dikembe diplomacy disruptions earmarks economy
einstein elections eliminates expand extremists failing faithful families freedom fuel funding god haven
ideology immigration impose insurgents iran iraq islam julie lebanon love madam marine math medicare
moderation neighborhoods nuclear offensive palestinian payroll province pursuing qaeda radical regimes resolve
retreat rieman sacrifices science sectarian senate september shia stays strength students succeed sunni tax
territories terrorists threats uphold victory violence violent war washington weapons wesley

Using Cloud Tags for Analysis

As you can see from the examples, a shift in focus from domestic to international affairs is easily visible with this graphic representation of word frequency. Creating a tag cloud of an existing document can often yield simple information about its main themes and rhetorical techniques.

Using Cloud Tags for Revision

Tag clouds are useful not only for dissecting others' writings but also for analyzing and revising your own. Since it provides a picture of word-frequency, a cloud can give you a quick and easy picture of what you are actually writing and which terms you use most often. Try turning a paper into a tag cloud, and see what words turn up most frequently:

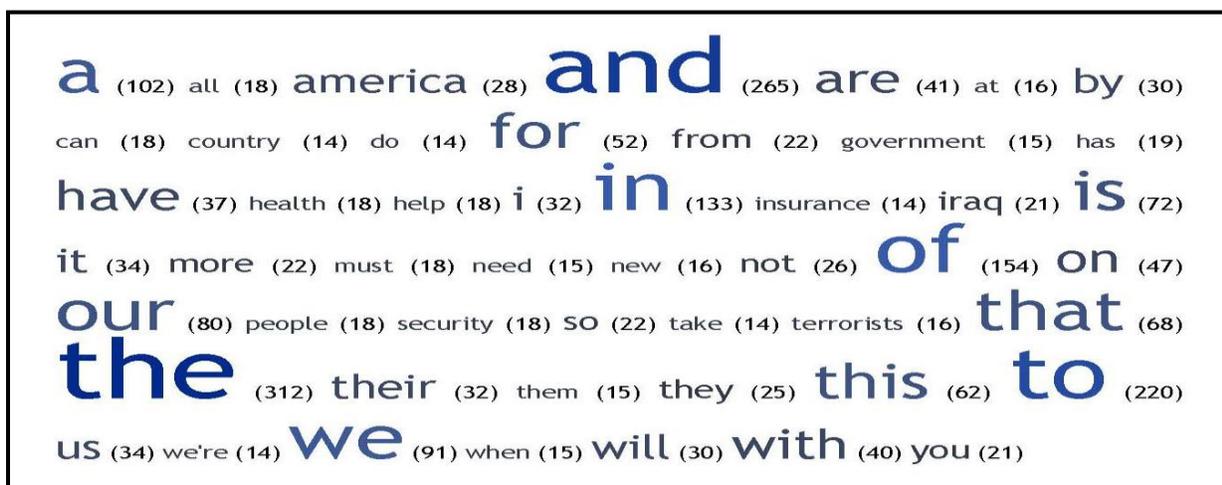
- Are they the ones you would expect?
- Are they the ones you want your reader to associate with the paper?
- Do you know *why* these words are the most frequent?
- Are you using unnecessary or unclear jargon, or would your topic be identifiable to anyone who saw this cloud?

You can also try writing up your own keyword list before you create the cloud. If the key concepts you identify do not show up strongly on the tag cloud, you may want to think of ways to emphasize them more.

The Devil Is in the Details

Often we end up picking up short stock phrases that slip below our revising radar. A tag cloud is one quick way of finding out how often we use certain little words, which often turn into writing tics. The default setting for most tag cloud generators filters out prepositions like “to,” “of,” “for,” and “by,” as well as transition words like “however,” “thus,” “therefore,” and “although.” These words are often the prime offenders of wordiness and redundancy, and you can use a tag cloud to identify which words you are using most and how often you use them. Most tag cloud generators will let you turn the filter off, and at least some will give you a numerical as well as visual representation of frequency. With that filter off, the cloud will look very different. (A useful test is to see if you can still tell what the paper is about from the unfiltered cloud.)

Here is the same 2007 State of the Union speech, now with the common-word filter off and the frequencies listed. This one was generated by www.tagcrowd.com.



Some words, like “and” and “the,” will almost always be at the top, but others may tell you something about your writing style. Bush’s speech is well honed, and was subject to much revision. There is, for example, a reason “we” and “our” show up so many more times than “they” and “their.”

Evaluating Your Own Writing

You probably don’t have the staff President Bush does to go over your papers, so you may find some surprises when you look at an unfiltered word count of your own work.

Once you know which words you are using most often, you can revise with those words in mind. Follow this two-step process:

1. Do a simple search for the words you use most often in the file.
2. Examine how you use the words in context. Is the repetition necessary, or can you make the same point using other words?

This type of quantitative analysis can give you a sense of your writing crutches: do you tend to refer to things “as” something else, or often write that something happens “in order to” enable something else? Do you begin many sentences with “yet” or “for”? None of these are inherently wrong, but if you find recurring habits, you might want to try to wean yourself off of them.

The same goes for clichés and superlatives: if you find yourself pronouncing something “the most [anything] in the world,” declaring repeatedly that something “gives up the ghost” or shows the “triumph of the human spirit,” or using vague and insubstantial words like “interesting” or “precisely,” you will want to flag those words or phrases when you write in the future.

Useful Links

<http://tagcrowd.com/>

David Steinbock has created a handy website that allows anyone to create a tag cloud from a file or pasted text (up to 3MB of data). The controls allow you to add a numerical frequency display, group similar words, and turn off common word filters.

<http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices/>

“Tag Clouds Gallery: Examples and Good Practices” in *Smashing Magazine*.

If you are interested in learning more about the details of and theory behind different tag cloud models, this is a fine place to start. It is focused on web use, but provides practical insight into design and function differences in various types of text cloud.

<http://chir.ag/phernalia/preztags/>

If you find the State of the Union comparison above interesting, this could be the website for you. Chirag Mehta has generated the top 100 tags for official US Presidential documents and speeches between 1776 and 2007. The interface is intuitive and the site demonstrates the analytic potential of tag clouds. Mehta also describes how the tag clouds and timeline slider were created and provides a downloadable “Tagline Generator” to allow users to create their own chronological tag clouds.