Tagcloud

Wednesday 29 April 2009 11:51

I've implemented some sort of tagcloud on this blog. It's not really a tagcloud since I don't tag my posts, but it looks like one. The tagcloud is generated by counting the occurances of all the words in all the posts, and then stopwords and other words I think add little value to the cloud are filtered out. There are also some stemming issues resolved like putting "playing" and "player" together with "play". From the remaining words the thirty or so with the most occurances are then fed to HTML::TagCloud, and that gives a result like this:

box cool days flash friday game gta home ibook linux movie music night open perl play rotterdam run screen server site source stuff system train users weather web weekend windows work

Not using real tags means every occurance of a word adds to the significance, so it's not based on how many posts are about something, but how much content is about something. It also means the words are limited to words, so "open" is showing up, but not "open source". That's probably why "OS X" isn't in that cloud. But it's nice to see "friday" and "weekend" show up, and "Rotterdam" managed to bubble up while "Amsterdam" didn't make it.

by Roland van Ipenburg http://www.xs4all.nl/~ipenburg/blog/posts/dull/2009/04/29/tagcloud/