Table of Contents

TagCloudPlugin

Renders a tag cloud given a list of terms

Description

This plugin helps rendering tag clouds. From Wikipedia:Tag_cloud

A Tag Cloud is a text-based depiction of tags across a body of content to show frequency of tag usage and enable topic browsing. In general, the more commonly used tags are displayed with a larger font or stronger emphasis. Each term in the tag cloud is a link to the collection of items that have that tag.

Tag clouds give a very quick overview of the distribution of terms within a document base. This can be used to support navigation or just to display the characteristics of the given data. Tag clouds are quite common in blog archives where you can click on a tag in the cloud and list all postings that are tagged that way. Tag clouds can be generated on any list of terms that don't need to be "tags" in the more stricter sense, though the name of the figure is still "tag cloud".

How tag clouds are computed

Computing the tag cloud is done by counting the occurrences of the terms in the input data.

First, the input data has to be tokenized and normalized to match each word in the input to a term by

1. splitting up the data, removing special characters,
2. mapping different terms on one term, like handling synonyms (e.g. twikiapps=TWikiApplication) (optional)
3. filtering unwanted words; there's a predefined list of (english) stop words that can applied
4. mapping plural to singular word forms (done rudimentarily for english only)

Once all terms have been counted they are mapped into a fixed set of *buckets*. For example given 102 different terms have been counted and ordered by frequency, they are mapped into a set of -- let's say -- 30 buckets that are containing those terms. Each term is assigned a *weight*, that is the ID of the bucket it has been sorted in.

In general the tag cloud renders the more frequent terms bolder and/or more colorful than the less frequent. See the exmaple below. But you are free to configure any variation of the appearance of a tag depending on its frequency.

Syntax Rules

Syntax:

```
%TAGCLOUD{(terms=)"<term-list>" ... }%
```

renders a tag cloud given a list of terms from which the term frequencies are extracted. There are a couple of options that influence the way the list of terms is tokenized and processed, as well as the appearance of the resulting tag cloud.

Parameters:

- `terms`: input data, a list of terms; a term can have the form `term:weight` whereas the occurrence of this term is counted with the given weight
- `split`: regular expression used to split up the input data (default: `[/,\.?\s]+`)
- `stopwords`: switch on/off filtering common (English) stop words (default: off)
- `filter`: regular expression of characters to be filtered out; possible values are 'on': to remove a predefined set of special chars, 'off' don't filter at all or an arbitrary string of characters to be replaced with a single space; default value is 'off'
- `exclude`: regular expression of terms to exclude from the tag cloud (default: undefined)
- `include`: regular expression of terms that should be included (default: undefined)
- `buckets`: number of buckets which to sort terms into (default: 10)
- `sort`: sort terms in the tag cloud alphabetically or by weight (alpha, count or weight) (default: alpha)
- `group`: format string to distinguish groups within the sortion; if the tag cloud is sorted alphabetically terms are groupd according their first letter; terms sorted by weight are grouped by their weight collected in tenner chunks. That is terms with a weight 1-10 are put in group "10", terms with a weight 11-20 are in "20" and so on; the current group displayed by replacing the pseudo variable $group in it; example: group="$group" (default: <empty>)
- `reverse`: reverse the sorting (on or off) (default: off)
- `min`: minimum times a term must occur to be included in the tag cloud
- `offset`: integer value that is added to the weight number of a term (default: 10)
- `lowercase`: switch on/off converting terms to lowercase (default: off)
- `plural`: switch on/off counting plurals, if set to off plural terms are matched to their singular form (caution: very rudimentary and assuming the input data is all English, use stopwords="on" to get good results)
- `map`: list of terms to be normalized as synonyms; the format of this is: "from1=to1,from2=to2,..." Each occurrence of `from1` will be mapped to `to1` etc, and counted as such
- `header`: format string to precede the output
- `format`: format string for each tag in the tag cloud
- `sep(arator)`: format string to be put between each tag in the cloud
- `footer`: format string appended to the output
- `warn`: switch on/off warning if no term was found (default: on)
- `limit`: number of terms to display, defaults to 0 - meaning display all

Format strings (`terms`, `header`, `format`, `sep` and `footer`) might contain the following pseudo variables:

- `$index`: the number of the tag in the (sorted) tag cloud
- `$term`: the term itself
- `$weight`: the term weight (that is: bucket + offset)
- `$count`: the number the term has been counted
- `$fadeRGB(startRed,startGreen,startBlue,endRed,endGreen,endBlue)` color of the term given its current weight in a color gradient defined by the given 6 integer values; returns a string `rgb(red,green,blue)` that can be used in a css color property
- `$percnt`: expands to `%`
- `$dollar`: expands to `$`
- `$n`: expands to a linefeed
- `$nop`: is removed from the format string
- `$group`: grouping format string, see above

Example

Cloud for the above text:

You type:

```
%TAGCLOUD{"$percntINCLUDE{\"%WEB%.%TOPIC%\"}$percnt"
  header="<div style=\"text-align:center; padding:15px;line-height:180%\">"
  format="<span style=\"font-size:$weightpx;line-height:90%\"><a style=\"color:$fadeRGB(104,144,1
  footer="</div>"
  buckets="40"
  offset="12"
  lowercase="on"
  stopwords="on"
  plural="off"
  min="2"
  map="bucket=pail"
}%
```

alpha appearance cloud color common counted counting data default different english example exclude expression filtering form format frequency frequent general given included input integer list lowercase mapped mapping normalized number occurrence offset output pail plural regular render set singular sort sorted stop string switch synonym syntax tag term tokenized undefined used value way weight word

You get (faked):

You get (if installed): %TAGCLOUD{"$percntINCLUDE{\"TWiki.TagCloudPlugin\"}$percnt" header="

" format="$term" footer="
" buckets="40" offset="12" lowercase="on" stopwords="on" plural="off" min="2" map="bucket=pail" }%

Plugin Installation Instructions

Note: You do not need to install anything on the browser to use this plugin. The following instructions are for the administrator who installs the plugin on the TWiki server.

- Download the ZIP file from the Plugin Home (see below)
- Unzip it in your twiki installation directory. Content:

File:	Description:
`data/TWiki/TagCloudPlugin.txt`	
`lib/TWiki/Plugins/TagCloudPlugin/Core.pm`	
`lib/TWiki/Plugins/TagCloudPlugin.pm`	
`pub/TWiki/TagCloudPlugin/TagCloud.jpg`	

- Run the configure script to enable the plugin

Plugin Info

Plugin Author:	TWiki:Main.MichaelDaum
Copyright:	© 2006-2008 Michael Daum http://michaeldaumconsulting.com ; © 2006-2010 TWiki:TWiki.TWikiContributor
License:	GPL (GNU General Public License)
Plugin Version:	19266 (2010-07-26)
Change History:	
2010-07-25:	TWikibug:Item6530 - doc fixes, changing TWIKIWEB to SYSTEMWEB
03 Jan 2008:	added limit parameter; added sorting according to term frequency (count)
13 Sep 2007:	don't remove numericals from terms
05 Jun 2007:	

Example 3

	better default values, e.g. filter is off by default now; fixed expansion of standard escapes
31 Aug 2006:	added filter parameter to customize special chars to be excluded; added NO_PREFS_IN_TOPIC
10 Mar 2006:	added grouping
07 Mar 2006:	added escape chars to the term list parameter
03 Mar 2006:	added warn parameter; fixed use of uninitialised value; if tags are sorted by weight tags of the same weight get sorted alphabetically now; sorting by weight is desending by default now
01 Mar 2006:	added docu, added more pre-defined english stop words, added map and plural parameters, reworked order of tokenization. added fadeRGB format string variable
24 Feb 2006:	Initial version
Perl Version:	5.8
TWiki:Plugins/Benchmark :	GoodStyle nn%, FormattedSearch nn%, TagCloudPlugin nn%
Plugin Home:	TWiki:Plugins/TagCloudPlugin
Feedback:	TWiki:Plugins/TagCloudPluginDev

This topic: TWiki > TagCloudPlugin
Topic revision: r6 - 2011-10-28 - TWikiAdminUser