

Tag Cloud in Chinese websites



By: Rex Wong

Published – December 2006

Tag cloud displays tags in a website which emphasize some of the tags by showing them with larger font sizes, and/or in darker colors. Moreover, tags in a tag cloud are usually arranged in alphabetical order. Tag cloud seems to work in the English world as a means of visualization as well as an extra means of navigation - what about in the Chinese world or more specifically, what about in Hong Kong?

The Language Environment in the Chinese World

Before answering this question, let's have some basic understanding on the language environment in the Chinese world. In Mainland China and Taiwan, people read and write simplified and traditional Chinese characters respectively, yet they both speak Hanyu*. Almost all foreign vocabularies have Chinese translations, except some of the new technological terms like "MP3", "QQ" (a popular instant messenger in China), etc.

In Hong Kong, the language environment is a bit more complicated. Hong Kong people write and read Chinese, yet we speak mainly Cantonese, a dialect of the Chinese language. Being influenced by such dialect as well as the English language, which is still spoken in business settings, the Chinese that Hong Kong people write can be different from what mainlanders and Taiwanese are familiar with. Moreover, Hong Kong people use more English vocabulary than mainlanders or Taiwanese do, and do not bother much to localize the foreign terms, e.g. "buffet", "tiramisu", "proposal", "(to) confirm", "(to) cancel", etc. These nouns and verbs have Chinese translations, yet Hong Kong people simply prefer saying these words in English, even though the rest of the sentence is in Cantonese.

Implications on Tag Cloud Design

What are the implications on visualizing tags with a tag cloud then?

First of all, alphabetical sorting does not work at all for Chinese characters. The usual Chinese counterpart of A-Z sort is sorting by number of strokes, but number of strokes is not a readily comparable properties of Chinese characters. Surely I can easily say "一" has fewer strokes than "國" does, but how about "國" versus "條"? These two characters have same number of strokes, then which among them should go first when sorted? Normal users would not have an answer.

Some websites in mainland China try another approach - sorting Chinese characters by "pinyin" (Romanized Chinese) in alphabetical order. This might work in mainland China as mainlanders learn Chinese characters together with "pinyin", though the Roman alphabet used in pinyin does not exactly map to the A-Z order. However, such way surely won't work in Hong Kong at the moment, as the Cantonese pronunciation system is quite different from that of the standard Hanyu. Although we Hong Kong people sort our family names in Romanized Cantonese by alphabetical order, it is not a usual practice to sort other daily life vocabularies by how they are pronounced.

Secondly, it can be challenging to show Chinese characters of different font sizes in a proper way. The reason is that the most common Chinese font types for displaying texts on web pages do not look as good when they size up. As shown in the figure below, the same character in larger font size could look "amateurish" or "code-like" to Chinese web

users (Visual designers usually get rid of this by using prettier fonts and convert it into graphic instead). You might think this is just a matter of taste or aesthetics, but I've actually had an experience of hearing users saying that they thought the tag cloud indicated website malfunctioning - "why is there a mess showing the code-like fonts in irregular font sizes?" This also hints that the larger-size-more-popular association may not be as guaranteed as expected.

工具 收藏2006 两性 美食 音乐 【软件下载】 电脑 数码 杂类
搜索 中国国民党硬盘 音乐网站 网站 下载 【电脑技术】 软件 收藏夹 软
件下载 blog 视频 生活 爱情 电影 影视 专业网站 虚拟空间 学习
英语 web2.0 游戏 免费电影 it 手机 新闻 论坛 图片 博客 电脑网
络 文学 社区论坛 搜索引擎 感动 焦点 tea 网络硬盘 美女 网络存
储 其他 热贴 娱乐

Thirdly, a tag cloud on Chinese website would not consist of only tags in Chinese. There would be English tags, and I guess the proportion of English tags would be highest on Hong Kong website, followed by Taiwan and then mainland websites. This is really just a guess because as far as I know there tagging is rare on Hong Kong websites, but I would not be surprised if such a tag cloud, if exist in the coming future, is interwoven by almost equal number of Chinese and English tags. This point goes back to the sorting order discussion again - if tags are sorted alphabetically, English tags and Chinese tags would be visually separated into two groups. Do you think it is a good way to let users visualize the "hot issues"?

How would a tag cloud in a Chinese website look like?

Given the three implications mentioned above, how can we design a tag cloud displaying tags in a Chinese website?

The font-size problem would be comparably easier to tackle. One way is to use flash or graphical text to show Chinese characters in a prettier as well as more professional manner. Another way is to use color instead of font size to convey differences in popularity.

The sorting-order problem discussed here is actually not a new problem, people who have developed localized Chinese versions for English software would have experienced the same problem before, and there has not been a good way to solve the problem. Yet, if we are referring to tag cloud design, my thought is that perhaps we can forget about sorting at all.

Why do we have to sort the tags? You can superimpose different orders of sorting on a tag cloud, not just alphabetical or by number of strokes, but also by latest added or by latest modified. However, if you consider these orders meaningful, you are actually making an assumption that people would scan tags in a tag cloud from left to right, and from top to bottom, as they were reading a paragraph. Is this assumption valid? Or are tags in larger font size "popping-up" and being better able to get users' attentions, rendering the order of sorting unnoticed at all? As long as tag cloud is not the sole navigation aid on the website, it does not have to be an exhaustive list of the tags. Perhaps the best way to use a tag cloud is to emphasis visualization instead of navigation.

* What foreigners learn as "Hanyu" (the Chinese language) is actually referred as "Putonghua" which means "the common language" in Hong Kong and the mainland. In Taiwan, people call it "Mandarin". Putonghua and Mandarin can almost be regarded as the same language, although there are some style differences between them.