Introduction

Folksonomy, or social tagging, is a relatively new and powerful way to organize information. We’ve had keywords for some time, but only with the distributed nature of the Internet, more powerful computers, and enough people to provide bandwidth supplying a rich vocabulary to a variety of resources have we been able to see our collective opinion and observe it in interesting ways.

We can see what many people think of things in near-real-time with the ability to search through the data. As tagging technologies spread into more applications and resources, the population doing the tagging, we will continue to uncover new ways to see and discuss extremely interesting data.

But because folksonomy is so new, all the tags and tag datasets we are currently analyzing are still relevant in the time domain. We haven’t been doing this long enough to have our tags become stale. But this will surely change.

Background

What happens when we’ve been tagging for many years—for many decades? Are all the tags that have ever been described an object relevant? Should the searher be able to determine what an old tag should matter to his investigation? How did it all begin? And who decides?

This is a struggle that has existed for some time in other fields besides library and information science—to determine how to retire old terms from official usage. Doctors must decide the current terminology for certain diseases; biologists decide how organisms are classified; chemists must decide how nomenclature is used. Each of these fields has its own way to determine how long should terms be active before they fall out of use.

What is new here is that this decision-making is getting pushed to the searcher themselves instead of being made by the curator or the panel of credentialed experts. The awareness that terms age out and become less relevant means what it used to mean (Coates, 2005). Deciding which one of these has occurred when a term becomes stale is a subtle distinction. We need better tools to help us decide what is actually happening.

Timeline

The recent collective, or aggregate, use of a tag suggests it is still relevant and its usage drops off over time, we can assume that it no longer has as much relevance to the community doing the tagging. If this trend is being observed for a particular tag, it can be argued that this term is of less importance than it was used to be.

Two graphing techniques may give us insight into how the fluctuation over time is occurring. Both involve a timeline with five key points defined, the researcher can see the direct feedback when investigating any tool to allow for the most interactive and useful, as it is the best general balance between not counting tags used only once (misspellings, special use-cases, personal tagging techniques) and finding the greatest variance before converging at 100%.

Tag Profile

The first technique will allow us to see how an aggregated tag profile for either an item or a person has changed over time. The tag profile is projected from left to right between points (4) and (5) of Figure 1. The y-axis is the percentage of tag events that used this tag since aggregation began at point 5. By inspecting this graph for diagonal motion, interesting tags can be identified and investigated further (Russell, 2006).

Tag Decay

The second technique allows us to see the aggregate change in a tag profile over time. The y-axis is the percentage of all tags seen since point (2) on the x-axis. And we graph a line each for tags used once, twice, three times, etc. The line formed for common tags (used two or more times) seems to be the most useful, as it is the best general balance between not counting tags used only once (misspellings, special use-cases, personal tagging techniques) and finding the greatest variance before converging at 100%.

Another interesting aspect of this view of a tag cloud is the way long ago the common tags reached certain thresholds (50%, 75%, 90%, 100%). For example, in Figure 2c below, 50% of the “common” tags (second line from the bottom) have been used in the last seven months. 90% of the “common” tags have been used in the last 21 months.

Commons, or populations, with more static vocabulary might go much longer before 90% of its tagging space has decayed through lack of recent use. It also seems fair to assume that the most recently used tags are actually at the head of the taglist of tags most used. The us could be suggested that we’re no longer seeing recently are in the long tail and might be good candidates for trimming from an official curated vocabulary.

Future Work

These techniques should allow for interesting comparisons across genres of sites as well as different groups of individuals (taggers). Do sports-related web sites have a shorter half-life in terms of the tags being used to describe them than web sites discussing the latest hits of movies produced in the 21st century? Both techniques could be used for single items and taggers (as above), but could also be used for multiple aggregated items or multiple aggregated taggers at once. This could also be used to determine similarity of a single object or person to a group.

References


Tag Decay and Tag Activity for the Web Site "Jaiku". There are lots of diagonal lines in this tag profile. This tells us that the tag space is very unstable and still in flux. After the large number of tagging events March 2007, the “presence” tag has been steadily losing ground while “blog” has been gaining. This is most probably due to the different demographic using Jaiku—the people have changed, so the vocabulary has changed. We can see the effect of the March boost as an elbow in the decay graph as well. The activity has been more fevered since March and we note this as a more rapid churn of older/ original items (we refer to this as the “common” tags) used in the last 21 months. 90% of the “common” tags (used two times or more) have been used in the last six months.

Tag Decay and Tag Activity for a web page discussing what would later be termed “Ajax”. The x-axis on the timeline can be represented as both time and as tagging events. The two are directly correlated, but will be varied depending on the amount of tag activity being investigated. Popular items and prolific people will have more tagging events within a shorter amount of time than those that are not active. Showing both time and tagging events on the x-axis allows the searcher more context in which to understand what is happening.