

Inferring relevance for tag search from user authority – Abstract

Anand Kishore
anand@semanticvoid.com

Tagging is an act of imparting human knowledge/wisdom to objects. Thus a tag, a one word interpretation/categorization of the object by the user, fundamentally represents the basic unit of human wisdom for any object. This wisdom is difficult to quantify as it is relative for every user. One approach to quantify this would be to use the wisdom of the other users to define this for us. This can be done by assuming that every tag corresponds to a topic for which every user has some authority. Also, every tag added to an object corresponds to a vote, similar to the Digg model, asserting that the object belongs to that topic (tag).

Let us consider a user U_i who has tagged object O_j with the tag T_k . Whenever other users in the system tag O_j with T_k , they are implicitly affirming U_i 's wisdom for tag T_k .

Thus, we define the function **affirmation** for the **tuple(u, d, t)** as the number of other users who have also tagged document d with tag t :

$$\mathbf{affirmation}(u, d, t) = \sum_{i=\text{All users except 'u'}} \mathbf{tagged}(u_i, d, t)$$

where,

u – the user

d – the document/object

t – the tag

tagged – 1 if the user U_i has tagged d with t

- 0 otherwise

Hence, we can proceed to define the wisdom of the user for a topic (tag) t as the sum of all such assertions by other users,

$$\mathbf{wisdom}(u, t) = \sum_{x=\text{For all documents d tagged with tag t by U}} \mathbf{affirmation}(u, d, t)$$

Likewise, we can now define the **authority** of a user for the topic t , as the ratio of the user's wisdom to the collective wisdom for t . Hence,

$$\mathbf{authority}(u, t) = \mathbf{wisdom}(u, t) / \sum \mathbf{wisdom}(u_i, t)$$

For example: Let us determine the authority of user u_1 for tag t_1

Object d_1 :

Object d_2 :

Object d_3 :

t1 by u1	t1 by u1	t1 by u2
t1 by u2	t3 by u1	t1 by u3
t1 by u3	t3 by u1	
t2 by u1		
affirmation(u1, d1, t1) = 2		affirmation(u1, d2, t1) = 0

Hence, wisdom(u1, t1) = 2

Likewise for other users,
 wisdom(u2, t1) = 3
 wisdom(u3, t1) = 3

Hence the authority of user u1 for t1 is as follows:
 authority(u1, t1) = 2 / (2 + 3 + 3) = 2 / 8 = 0.25

Whenever a user tags an object with a tag, he does so with the authority he possesses for that tag. Thus as compared to conventional methods, where the objects are usually ranked on the number of instances of the tags, in this method the measure of the relevance of a tag for an object is equivalent to the sum of all such user authorities. Thus,

$$\text{relevance_metric}(d, t) = \sum_{i=\text{all user who have tagged document } d \text{ with } t} \text{authority}(u, t)$$

This relevance score, when calculated for every tag would provide an accurate measure for ranking the objects. As compared to the conventional methods where more number of instances of a tag for an object ensured a higher relevance for that tag, here the number of authoritative users counts.

Let us consider the following example:

Object d1:	Object d2:
t1 by u1	t1 by u2
t2 by u5	t1 by u3
	t1 by u4

Let us assume that u1 has a very high authority for tag t1. Hence in the above scenario, a search for tag t1 may rank d1 higher than d2, if

$$\text{authority}(u1, t1) > \text{authority}(u2, t1) + \text{authority}(u3, t1) + \text{authority}(u4, t1)$$

This result is with the assumption that u1's authority is greater than those of u2, u3 and u4 combined.

On the other hand, d2 would be ranked higher than d1 if the combined authorities of u2, u3 and u4 exceed that of u1. If the majority of the users are

suggesting something, it indicates that their suggestion is far more valuable than that of an individual user or a subset of users.

Future Enhancements

While calculating the user assertions this algorithm currently considers all such users as equal even though they may have varying authorities for the corresponding tag. As a future enhancement, I plan to incorporate the authorities of the users as well into the affirmation calculations.