Can social tags help you find what you want?

Khasfariyati Razikin; Goh, Dion Hoe-Lian; Chua, Alton Yeow Kuan; Lee, Chei Sian

2008

Khasfariyati, R., Goh, D. H. L., Chua, A. Y. K., & Lee, C. S. (2008). Can Social Tags Help You Find What You Want? In Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries ECDL 2008, pp. 50-61.

https://hdl.handle.net/10356/94588

https://doi.org/10.1007/978-3-540-87599-4_6

Can Social Tags Help You Find What You Want?

Khasfariyati Razikin, Dion Hoe-Lian Goh, Alton Y.K. Chua, Chei Sian Lee

Wee Kim Wee School of Communication & Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718, Singapore
{khasfariyati, ashlgoh, altonchua, leecs}@ntu.edu.sg

Abstract. One of the uses of social tagging is to associate freely selected terms (tags) to resources for sharing resources among tag consumers. This enables tag consumers to locate new resources through the collective intelligence of other tag creators, and offers a new avenue for resource discovery. This paper investigates the effectiveness of tags as resource descriptors determined through the use of text categorisation using Support Vector Machines. Two text categorisation experiments were done for this research, and tags and web pages from del.icio.us were used. The first study concentrated on the use of terms as its features. The second study used both terms and its tags as part of its feature set. The results indicate that the tags were not always reliable indicators of the resource contents. At the same time, the results from the terms only experiment were better compared to the experiment with terms and tags. A deeper analysis of a sample of tags and documents were also conducted and implications of this research are discussed.

Keywords: Social tagging, Resource Descriptors, Resource Discovery, Support Vector Machines

1 Introduction

Social tagging has a variety of uses [1], one of which is the assigning of freely selected terms (tags) to resources, such as web pages, online videos, digital photographs and bibliographies, among tag consumers for the purposes of sharing [5, 13, 14]. This enables tag consumers to locate new resources through the collective intelligence of tag creators, and offers a new avenue for resource discovery apart from search engines and Web subject directories. Tags function both as content organizers and discoverers. As content organizers, tags enable tag creators to annotate and categorize a resource so that it can be retrieved subsequently with ease. Tag consumers will use those same tags to locate that resource. As content discoverers, tags could be used as a means to tap into the collective intelligence of tag creators to make serendipitous discoveries of additional relevant resources. Furthermore, through tags, a tag consumer is able to find like-minded tag creators with resources that meet his or her information needs, potentially leading to the creation of social networks [14]. Examples of popular social tagging systems include del.icio.us, Flickr, YouTube, Cite-U-Like and Last.fm.

Social tagging differs from conventional methods of resource categorisation based on taxonomies, controlled vocabularies, faceted classification and ontologies. The creation of systems utilizing such methods requires experts with domain knowledge and often adds on to the costs of implementation. Conventional categorisation methods are also invariably rule-bound to ensure consistency in their classification schemes [15]. As the system gets bigger, the rules tend to be more complicated, leading to possible maintenance and accessibility issues. Lakoff [10] explains that the classification done by ordinary people are defined by tacit knowledge. This is in turn dependent on a person's language and culture. Based on this argument, a conventional categorisation system suffers from the lack of precision as it is not able to provide a gamut of contextual information a user needs [13]. In contrast, social tagging systems make use of the knowledge from a (possibly large) community of tag creators instead of relying on (a few) experts. These systems have a flat hierarchy [5], doing away with the need for defining classes and subclasses. At the same time, social tagging systems do not have prescribed rules to govern the choice of tags for a given resource. Instead, tag creators can exercise discretion to decide what tags to use. The Wisdom of Crowds theory [22] postulates that the knowledge that comes from a large group of users will be more reliable than that from an individual. As such, a resource which attracts different tags contributed by multiple users is conceivably more meaningful described than one which attracts a few tags from a single user.

As social tagging systems become increasingly popular, there is also a growing amount of research that focus on the role of tags as resource descriptors. For example two studies compared the reliability of tags against that of manually indexed terms for academic papers [9, 12]. Another has examined the similarities between blogs sharing the same tags [21]. Still, it remains to be seen if tags can be used as effective means for discovery of information.

For this reason, the objective of this paper is to investigate the effectiveness of tags in assisting tag consumers discover relevant content. Del.icio.us was selected as our dataset for two main reasons. One, it is one of the earliest and more popular social tagging sites. Its main function is to store, organize and share bookmarks [14] among a community of users, and it provides an authentic context appropriate for this study. Two, it has a large and diverse set of tags and web resources for analyses.

Pages together with tags mined from del.icio.us are analyzed by determining if the tags are indeed accurate descriptors of the resource. This is done by techniques drawn from text categorisation [19], and more specifically from past studies [7] that have looked into the automatic assignment of documents to pre-defined categories. In our work, the Web documents in our dataset are fed to the classifier which will determine the category to which the documents belong. Tags, in our case, serve the same purpose as the category labels in text classification experiments. Here, we define an effective tag as one that is able to categorize a resource with high precision, recall and F-measure scores as determined by the classifier. Apart from conducting text categorisation experiments, we conduct detailed manual analyses to study the relationship between the application of a tag on a document and the document's terms to better understand how tags are created and used. To the best of our knowledge, there are a limited number of studies that have been done employing these two approaches. Our work can therefore be used as a basis for future work in this area as well as for designing techniques that better harness social tags for resource discovery.

The next section will expound on related studies, followed by a section that will illustrate on the methodology employed. The three subsequent sections elaborate on our findings and the paper will conclude with a section on discussion and conclusion.

2 Related Studies

There has been a steady stream of research done in the area of social tagging. These studies concentrate mainly on the architecture of systems [6, 16], usage patterns in these systems [5, 14], visualization of tags [3], spamming in tagging systems [9] and encapsulation of tags in search systems [23]. Despite the popularity of such research, there are a limited number of studies that focus on the effectiveness of tags as resource descriptors and organizers. Here, we highlight a few of them.

Comparing tags with controlled vocabularies provides a basis for evaluating how tags are similar to or differ from keywords provided by experts and content creators. Lin [12] compared tags with indexing terms to determine characteristics which could improve searching and browsing. Tags from Connotea were compared with Medical Subject Heading terms (MeSH terms). Their comparison found only 11% of similarity between MeSH terms and tags supplied by the tag creators. This is because MeSH terms function as descriptors while tags are selected based on tag creators' area of interest. It is evident from these results that there are differing views between an expert and the common user.

Related to [12], [8] compared tags with author supplied keywords and indexing terms to determine usage overlap in scholarly articles. Author-supplied keywords were compared against tags from Cite-U-Like and either INSPEC or Library Literature terms. Approximately 35% of the tags were found to be related to the keywords and indexing terms. The relation between the tags and supplied terms were more on the conceptual level as opposed to those relationships which were formally defined in the created thesaurus. One such example would be when a Cite-U-Like tag creator used the methodology in the article as a tag which differs from the keywords. The findings in this study were consistent with those found in [12] where the tag creators would tag with descriptors that indicate their focus of collecting such articles.

Pioneering work done on automatic text categorisation in social tagging systems was done by [2] in the blogosphere. The authors used 350 popular tags from Technorati and 250 of the most recent articles of the collected tags. Using TF-IDF [18] to cluster documents and pairwise cosine similarity to measure the similarity of all articles in each cluster, they found that tags categorize articles in the broad sense. It was implied that the tag consumer would be able to find articles that are related but not entirely about the topic. A similar study was done by [21], who concentrated their efforts on classifying whole blogs with tags. Their aim was to determine if tags were effective in classifying blogs and, at the same time, investigate the usefulness of including tags in classification. They studied 52709 blogs and 161 tags mined from BlogFlux and used their blog descriptions and tags. Automatic text categorisation using Support Vector Machines (SVM) was adopted. They compared the classification results of blogs based on tags only, and tags and the description of blogs, and descriptions only. Tags and descriptions had the best classification results

and tags alone were a more effective classifying feature than blog descriptions alone. In short, the results suggest that that tags can help tag consumers find relevant information.

Apart from blogs, a pilot study of tag effectiveness in describing Web document content [17] was conducted on a del.icio.us dataset. Their corpus consisted of 20 tags with 1385 documents. Their results show that tags do help in retrieving relevant information. Despite a small scale study, their results showed positive outcome which we intend to improve upon further in our present study.

The above studies analyzed the effectiveness of tags for resource discovery using different methods and in different domains. The studies by [8] and [12] were limited to scholarly articles while [2] and [21] used blogs. The context of medium of communication used differs from our study. The purpose of an academic article is to disseminate information in a formal and objective manner, and typically caters to a limited audience. In contrast, blogs contain commentaries and sentiments, catering to a more diverse readership, and offer a wider variety of topics. The pages that are bookmarked in del.icio.us are diverse and not limited to ordinary web pages, but also includes blogs and academic articles. Although the pilot study in [17] is strongly related to our work, they took into account only specialized tags which have very specific meanings such as "Internet programming" and "Machine learning". That study did not consider the ordinary tag consumer who would not use specialize tags [5] for organization purposes. The present study is thus timely as we conduct a larger scale analysis on the effectiveness of tags in retrieving general Web documents.

3 Methodology

Del.icio.us was chosen as it is not restricted to a specific domain. Tags and web documents were harvested from the site from August 2007 to October 2007. During this period, we randomly collected 100 tags and 20210 documents that were in the English language. Consistent with the work of [2], we started mining the tags from the popular tags page as such our tags will be biased towards the more commonly used ones. Nevertheless, the popularity of a tag indicates that there are a significant number of documents related to it and these documents will provide a sufficient dataset size for our experiments. The popular tags' collection of documents is where one's resources would more likely be accessible to tag consumers [14]. Also, accessing popular tags gives a consumer a good prospect of obtaining the required information.

Two text categorisation experiments were conducted. SVM [7] was the machine learning classifier selected as it is commonly used in web-based text categorisation studies with good performance [20]. Specifically, we used the SVM^{light} [7] package. The output from the classifier, which are precision, recall and F-measure were used to determine the effectiveness of tags.

The first experiment used only the terms from the documents as features. The second experiment included tags, in addition to terms, as part of its feature set. The first experiment served as a baseline for the second experiment. It is the simplest approach as it uses the fundamental information from the documents [20]. The

performance of the tags was evaluated based on the macro-averaged and micro-averaged precision, recall and F-measure. Macro-averaged values give an indication of the overall performance of the classifier over all tag categories. Each tag is given equal weight. Micro-averaged values emphasizes on the performance on categories with larger number of documents as it measures the performance over each document.

The tags that were mined consisted of single token terms. Each tag had an average of 1331 documents, and each document on the other hand had an average of 6.66 tags each. The minimum number of tags for a document was one, while the highest number of tags for a single document was 65. There were 3167 documents with a tag each. On the other hand, there was only a single document with the largest number of tags. Figure 1 shows the distribution of the tags for the number of documents. It clearly demonstrates the power law distribution of tags. Interestingly the same was observed for blogs [21].

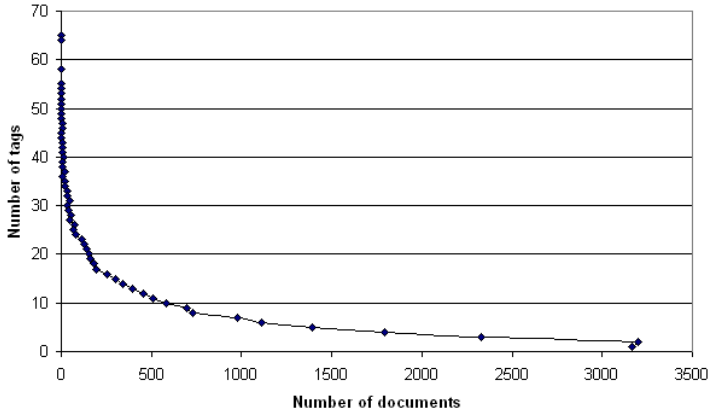

Fig. 1. The distribution of tags over the number of documents

4 Experiment 1 – Terms only

This experiment used the terms found in the content of the documents as features. The collected documents were processed by removing the HTML elements, JavaScript codes and Cascading Style Sheets elements. This was followed by stop word removal and stemming of the remaining words. TF-IDF values of the terms were then obtained. These values were used as the feature vector for the SVM classifier.

For each tag, we selected all the documents that were tagged with the keyword and these were grouped as the positive samples for the particular tag. An equal number of documents, which were tagged with a different tag, were randomly selected as negative samples. From this set of positive and negative samples, two-thirds of the documents were used as the training sample while the rest were part of the test set.

We made use of binary classifiers for the study, and one classifier was created for each tag. Default parameters of the SVM package were used. The output of the classifier was used to assess the accuracy of the classification algorithm.

Table 1 shows the top 15 tags with the highest F-measure obtained from experiment 1 while Table 2 shows 15 tags with the lowest F-measure. In both tables, the extreme right column shows the difference in the F-measure values obtained in experiments 1 and 2. The results are ranked in descending order according to the tag's F-measure values obtained in this experiment.

Table 1. The top 15 tags with the highest F-measure values obtained in experiment 1. The bold entry indicates an increase in the F-measure value for experiment 2

	Experiment 1			Experiment 2			Diff
Tag	Precision	Recall	F-measure	Precision	Recall	F-measure	
reference	58.38	87.23	69.95	57.80	62.83	60.21	-9.74
howto	56.02	86.21	67.92	61.93	54.83	58.16	-9.76
politics	55.25	87.91	67.85	52.81	90.04	66.57	-1.28
imported	58.57	79.50	67.45	56.40	52.99	54.64	-12.81
Fun	55.01	86.83	67.35	50.05	55.94	52.84	-14.51
blogs	55.07	85.74	67.06	59.14	73.92	65.71	-1.35
web	57.37	80.24	66.90	55.76	71.92	62.82	-4.08
web2.0	55.58	82.92	66.55	55.86	75.00	64.03	-2.52
inspiration	53.51	86.29	66.06	54.10	63.04	58.23	-7.83
internet	54.90	82.18	65.83	55.17	66.22	60.19	-5.64
california	57.14	76.40	65.38	55.17	66.22	60.19	-5.64
restaurants	55.43	79.69	65.38	49.07	88.76	63.20	-2.18
osx	54.07	82.58	65.35	48.00	56.25	51.80	-13.58
recipe	56.83	73.79	64.21	54.92	69.30	61.28	-4.07
news	**54.93**	**76.52**	**63.96**	**58.19**	**88.24**	**70.13**	**6.17**

Table 2. The bottom 15 tags with the lowest F-measure values. The bold entries indicate an increase in the F-measure value for experiment 2

	Experiment 1			Experiment 2			Diff
Tag	Precision	Recall	F-measure	Precision	Recall	F-measure	
templates	**49.63**	**31.60**	**38.62**	**63.27**	**43.87**	**51.81**	**13.19**
animation	46.99	31.97	38.05	52.43	22.13	31.12	-7.88
xml	47.03	31.52	37.74	51.30	28.42	36.57	-1.17
ajax	52.47	29.32	37.62	39.58	9.52	15.35	-22.27
economics	44.71	30.89	36.54	49.25	26.83	34.74	-1.80
windows	54.95	26.93	36.14	40.00	9.32	15.12	-21.02
accessories	**47.37**	**28.42**	**35.53**	**52.63**	**52.08**	**52.36**	**16.83**
cms	45.28	27.80	34.45	45.59	23.85	31.31	-3.14
journal	**51.32**	**25.83**	**34.36**	**42.74**	**35.10**	**38.55**	**4.19**
ruby	**55.56**	**24.15**	**33.67**	**55.64**	**35.75**	**43.53**	**9.86**
actionscript	43.36	26.34	32.78	49.38	21.51	29.96	-2.82
parts	**50.00**	**22.50**	**31.03**	**57.89**	**27.50**	**37.29**	**6.26**
self-improvement	43.55	23.28	30.34	44.00	18.97	26.51	-3.83
icons	**45.45**	**14.93**	**22.47**	**55.84**	**32.09**	**40.76**	**18.29**
adobe	**45.10**	**13.29**	**20.54**	**42.86**	**13.87**	**20.96**	**0.42**

On the whole, the top 15 tags shown in Table 1 had better recall than precision values indicating that the classifier was able to correctly assign the documents which actually belonged to the tag more than 75% of the time. This means that the classifier predicted a low number of true negatives correctly in comparison to false positives.

However, the bottom 15 tags shown in Table 2 paint a different picture. Here, the recall values for these tags are now lower than its precision values. This implies that the classifier tended to predict more true negatives that true positives. In other words, the number of documents that did not belong to the category was higher than the documents belonging to it.

In terms of macro-averaged values (Table 3), the precision value suggests that 52.66% of the documents which were thought to belong to the tag were correctly predicted, while the recall value indicates that 54.86% of the documents which were correctly predicted are in fact part of the document set. Additionally, the standard deviation for recall was greater than the precision's standard deviation. The reason for this could be attributed to the classifier's tendency to misclassify a page which actually belonged to the tag. This is dependent on tag itself as previously stated. The macro-averaged F-measure suggests that the classifier managed to predict at least half of the test data correctly and manages to perform this task almost equally well for all tags. Micro-averaged values shown reflect the classifier's performance for each document in the collection. The recall value shows that 54.4% of the documents that were identified to be relevant were correctly predicted, and the precision value shows that 64.76% of the documents that were predicted as part of the document set were correct. On average, 59.14% of the predictions (accuracy) were correct.

Table 3. Experiment 1 macro- and micro-averaged values for precision, recall and F-measure

	Precision (%)	**Recall (%)**	**F-measure (%)**
Macro-averaged	52.66 (s = 4.21)	54.86 (s = 19.05)	52.05 (s = 10.99)
Micro-averaged	64.76	54.40	59.14

Both the macro-averaged and micro-averaged F-measure values are quite close. However, the F-measure value suggests that the tags might not be reliable as resource descriptors as the motivation of tag creators may go beyond than just simply sharing resources with tag consumers. The outcome here is interesting. It was stated previously that tags from a large group of users would be more reliable for resource description in contrast to expert individuals. However, this is not the case as shown from our results. Documents in our dataset were tagged by people who followed no well-defined rules. This leads to inconsistency [5] with the underlying reason being that the tags can have multiple meanings attached to it. This also demonstrates that there is no agreement on a tag's usage in a social tagging system. As a result, the documents that are within the same tag cluster may not be semantically related. This in turn reduces the classifier's precision. The Vocabulary Problem [4] is another reason that contributes to the results. It was found that there is a 20% chance that a pair of random people would choose the same label.

5 Experiment 2 – Terms and Tags

The second experiment augmented the first by adding additional features with the aim to determine if these new features would improve the results. The setup for the experiment was similar to that done for experiment 1. The main difference was the addition of the document's tags to the feature set. The TF-IDF values for the tags

were used as the feature values in addition to the documents' terms. Likewise in this experiment, the default parameters of the SVM package were used.

The results obtained in this experiment are shown on the right column of Tables 1 and 2. The same tags that were selected in experiment 1 are again shown in the tables. In addition, the difference between the F-measures obtained in both experiments for the selected tags are shown. The entries in bold show an increase in F-measure values from that obtained in experiment 1. Here, only 8 tags have increased in their F-measure values. The tag "icons" has the largest gain with 18.29 indicating that the documents belonging to this tag have an increased chance to be classified correctly On the other hand, the tag "ajax" suffered the largest drop in F-measure value with a decrease by 22.27. This shows that the classifier has made more incorrect predictions.

Table 4 shows the macro-averaged and micro-averaged values for precision, recall and F-measure obtained for experiment 2. On average, the categories had precision and recall values of 50.77% and 45.24% respectively. The standard deviation for precision is 6.06 was smaller than that for recall (20.75), similar to the values obtained in experiment 1. The classifier only managed to predict 45.77% of the documents correctly for each category on average. With a standard deviation of 13.21, the classifier's performance did not vary much between categories. For micro-averaged values, the classifier managed to predict the relevance of each document with a precision and recall of 56.47% and 54.93% respectively. The micro-averaged value for F-measure is 55.69%.

Table 4. Experiment 2 macro- and micro-averaged values for precision, recall and F-measure

	Precision (%)	**Recall (%)**	**F-measure (%)**
Macro-averaged	50.77 (s = 6.06)	45.24 (s = 20.75)	45.77 (s = 13.21)
Micro-averaged	56.47	54.93	55.69

Here, we compare both the macro-averaged and micro-averaged values obtained from the experiments. Interestingly, it can be seen that the values obtained in experiment 2 are lower than those obtained in experiment 1. This implies that the addition of the tags as part of the features does not help in improving the precision, recall and F-measure values. The results concur with previous work in text categorisation [11] where the terms only approach scored better than other combinations. This is probably because words are the at the most atomic level where the "syntax and semantics meet" [7]. Although tags are words, they seem to degrade performance because of the frequency it appears in the document. This in turn causes it have an insignificant weight in the document collection. Hence, the tag here does not contribute to the grouping of documents.

6 Analysis of Selected Tags and Documents

As seen in Table 1 and Table 2, there appears to be no discernible patterns among the top and bottom 15 tags in terms of their characteristics. Furthermore, deriving conclusions solely from the precision, recall and F-measure values does not give a comprehensive finding because they do not reveal specific reasons contributing to the scores obtained. Thus, detailed analyses were done to determine trends which could

account for the performance of the SVM classifier. This section describes the methodology and discusses our findings.

A total of eight tags were selected based on the following characteristics:

1. Subjective or objective type of tags, and
2. High false negative value or high false positive value

The definitions for subjective and objective tags were built upon the definitions of intrinsic and extrinsic tags defined in [5]. Subjective tags refer to terms which could be adjectives or verbs. These tags could either describe features of the resource based on the tag creator's intent, have reference to the creator or some action that the he/she wants to take in the future with the resource. Conversely, objective tags refer to terms which are nouns. These tags describe the content of the resource, specify the context of the resource, state the owner of the resource and/or improve upon other tags that are associated with the resource. A tag with a high false negative (FN) value has the highest instances of documents being misclassified as not associated with the tag when the opposite is true. A tag with the highest false positive (FP) value has the highest number of documents being categorized as belonging to the tag when the opposite is true.

The following eight tags were selected based on the characteristics above: "interesting", "funny", "software", "3d", "re", "free", "adobe" and "dessert". For each tag, ten documents were randomly selected from the testing set and manually analyzed to uncover discernable patterns.

Of the eight selected tags, four of them, namely "software", "3d", "adobe" and "dessert" exhibit objective characteristics. It was observed that these tags appear frequently as terms in their associated documents. For example, among documents associated with the tag "software", the term "software" appears 94 times in a document on Agile software development process. Likewise, in another document offering instructions on building an Adobe AIR Application that was tagged with "adobe", the term "adobe" appears 45 times in the content. It does seem that objective terms which appear frequently in a document have prompted the tag creators to use them as tags.

The subjective tags selected for this analysis are "interesting", "funny", "re" and "free". The documents associated with these tags were found to cover a variety of topics. For instance, documents that were tagged "funny" range from documents on comics to articles tinged with sarcasm and humorous online videos. In contrast to objective tags, subjective tags reflect the tag creators' personal judgement on the content of the associated documents. Hence, if the goal is to find relevant documents within a narrowly-defined scope, then searching with subjective rather than objective tags is likely to yield better results.

Tags selected with high FP are "funny", "3d", "free" and "dessert". The documents associated with these tags do not seem to have any relation with the corresponding tag. For example, a document on CSS tutorial was classified with the tag "3d". However, a closer inspection reveals that the documents' terms are in fact transitively connected to the tag. Returning to the earlier example, the CSS tutorial document contained frequently occurring terms such as "tutorial" and "design". Furthermore, documents tagged with "3d" were also found to have high occurrences of the terms "tutorial" and "design" in their content. Hence, the overlap of commonly occurring terms in these documents appears to account for the classifier's performance.

Finally, tags selected with high FN are “interesting”, “software”, “re” and “adobe”. Among a total of 40 associated documents, only 16 were annotated with these tags. For instance, a tutorial on Python scripting language was tagged with “software”. In another case, a Wikipedia entry on one of the branches of Philosophy was tagged with “interesting”. While these tags may serve the purposes of the tag creators well, they hold broad meanings and are certainly not discriminating for the classifier to associate the documents to the tags. This suggests that tag consumers may find it difficult to access documents effectively using such tags.

7 Discussion and Conclusion

Social tagging has become a popular means of organizing web resources. Rather than to propose new techniques related to social tagging, the purpose of this paper is to investigate the effectiveness of tags in assisting the discovery of relevant content. Using a text categorisation approach, two experiments were conducted. The first examined the use of document terms only as features while the second added the document’s tags in addition to the previous feature set. The terms only experiment yielded slightly better results than the experiment with terms and tags. Our results suggest that not all the tags are useful descriptors for resource sharing. In the analysis, it was found that the performance of the SVM classifier was likely to be influenced by the tag creator’s motivations, and the appearance of the tag in the document content. Also, documents with high FP tend to have terms which are semantically connected to the tag itself. Tags with high FN have broad definitions, which in turns causes it to have diverse documents, making it hard for the classifier to predict correctly.

Our findings are similar to [21]. In that study, the range of macro-averaged F-measure obtained for description only experiments ranged from 32% to 41%. Perhaps the much lower values were a result of using a shorter length of text as descriptions. It was reported that the description contains an average of 14.8 terms for each blog. While the work of [17] was similar to ours, the results obtained in that study was better. A reason for this could be the tags chosen were not from the popular page and consisted more than one term. In addition, the documents that were associated with such tags tended to be specifically about the subject themselves.

Three main implications can be drawn from our study. First, on the basis of the Wisdom of Crowds theory [22], the quality of tags created by a community was thought to be better than that provided by an expert. However, our study shows that the theory has not been consistently supported. In particular, some tags were found to be good descriptors while some were not. Given that tags are created for a variety of purposes, the use of tags to search for relevant documents must therefore be treated with care. Second, objective tags have been found to appear frequently as terms in their associated documents. If the intention of a tag creator is to share a document with others, then objective terms that appear frequently in the document content could be used as tags. Furthermore, the more specific in meaning the tags hold, the better chance the document would be searchable by others. Third, better guidelines for tag creation could be provided by social tagging systems, although this appears to go against the spirit of free keyword assignment. One could envisage a semi-automated

tagging approach in which the system analyzes a Web resource and suggests possible tags, but leaving the user the freedom to make his/her own selections.

There are limitations to our study with regard to use of terms and tags of the documents. These might not be the only features that could be used. Additional features like the document's title and the anchor text could prove useful for classification. Other factors like the frequency of the tags being assigned to the document could also be another feature to be considered. Further, the present study used only popular tags but the number of such tags is proportionately smaller than the entire collection of tags in del.icio.us. Future work could utilize a wider variety of tags to determine if performance may be affected. For example, less popular tags may be associated with more esoteric, but more specific concepts and therefore could result in better classifier performance.

Here, we have put forward our results based on our investigation on how good tags are as resource descriptors and which feature set being used gives better results. We have shown that the terms only experiment gave better performance. At the same time, not all tags describe a document's contents sufficiently for public access. This investigation has shed some light on the characteristics of tags as resource descriptors and it will be useful for future work to be further conducted along these lines of investigation in order to understand tags better.

Acknowledgements
This work is partly funded by A*STAR grant 062 130 0057.

References

1. Ames M., Namaan, M.: Why we tag: motivations for annotation in mobile and online media. In Proceedings of CHI 2007, pp 971- 980 ACM (2007)
2. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In Proceedings of the 15th international conference on World Wide Web, pp. 625-632 ACM (2006)
3. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing tags over time. In Proceedings of the 15th international conference on World Wide Web, pp. 193-202 ACM (2006)
4. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. *Commun. ACM*, *30* (11). 964-971 (1987)
5. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. J. Inf. Sci. 32 (2) 198-208 (2006)
6. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social Bookmarking Tools (I) D-Lib Magazine (2005)
7. Joachims, T.: Learning to classify text using support vector machine. Kluwer Academic Publishers, Boston (2002)
8. Kipp, M.E.: Exploring the context of user, creator and intermediate tagging ASIS&T 2006 Information Architecture Summit, Vancouver, Canada (2006)
9. Koutrika, G., Effendi, F.A., Gyöngyi, Z., Heymann, P., Garcia-Molina, H.: Combating spam in tagging systems. In: Proceedings of the 3rd international workshop on Adversarial information retrieval on the web, pp. 57-- 64. ACM (2007)

10. Lakoff, G.: Women, fire, and dangerous things: what categories reveal about the mind. University of Chicago Press, Chicago (1987)
11. Lewis, D.D.: An evaluation of phrasal and clustered representations on a text categorisation task. In: 15th annual international ACM SIGIR conference on Research and development in information retrieval, pp 37--50 ACM (1992)
12. Lin, X., Beaudoin, J.E., Bui, Y., Desai, K.: Exploring characteristics of social classification. In: Furner, J., Tennis, J.T. (eds.) 17th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research, Austin, Texas (2006)
13. Macgregor, G., McCulloch, Emma: Collaborative tagging as a knowledge organisation and resource discovery tool. Library Review, 55 (5) 291—300 (2005)
14. Marlow, C., Naaman, M., boyd, d., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In Proceedings of the seventeenth conference on Hypertext and hypermedia, pp 31--40 ACM (2006)
15. Morville, P.: Ambient findability. O'Reilly, Sebastopol, CA (2005)
16. Puspitasari, F., Lim, E.P., Goh, D.H.L., Chang, C.H., Zhang, J., Sun, A., Theng, Y.L., Chatterjea, K., & Li, Y.Y.: Social navigation in digital libraries by bookmarking. In: Goh, D.H.L., Cao, T. H., Sølvberg, I., Rasmussen, E. M. (eds.) ICADL 2007. LNCS vol. 4822, pp. 297-306. Springer-Verlag, Heidelberg (2007)
17. Razikin, K., Goh, D.H.L., Cheong, E.K.C., Ow, Y.F.: Social efficacy of tags in social tagging system. In: Goh, D.H.L., Cao, T. H., Sølvberg, I., Rasmussen, E. M. (eds.) ICADL 2007. LNCS vol. 4822, pp. 425--426. Springer-Verlag, Heidelberg (2007)
18. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Proc. and Mgt. 24 (5) 513-523 (1988)
19. Sebastiani, F.: Machine learning in automated text categorisation. ACM Computing Survey, 34 (1). 1-47 (2002)
20. Sun, A., Lim, E.-P., Ng, W.-K.: Web classification using support vector machine. In: Proceedings of the 4th international workshop on Web information and data management, pp. 96--99 ACM (2002)
21. Sun, A., Suryanto, M.A., Liu, Y.: Blog classfication using tags: an empirical study. In: Goh, D.H.L., Cao, T. H., Sølvberg, I., Rasmussen, E. M. (eds.) ICADL 2007. LNCS vol. 4822, pp. 307--316. Springer-Verlag, Heidelberg (2007)
22. Suroweicki, J.: The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economics, societies, and nations. Doubleday, New York (2004)
23. Yanbe, Y., Jatowt, A., Nakamura, S., Tanaka, K.: Can social bookmarking enhance search in the web? In: Proceedings of the 2007 conference on Digital libraries, pp 107-116 ACM (2007)