From Web 1.0 to Web 2.0 and Back – How did your Grandma Use to Tag?*

Sheila Kinsella†‡, Adriana Budura‡, Gleb Skobeltsyn‡,
Sebastian Michel‡, John G. Breslin†, Karl Aberer‡

† Digital Enterprise Research Institute (DERI)
National University of Ireland, Galway
firstname.lastname@deri.org

‡ Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
firstname.lastname@epfl.ch

ABSTRACT

We consider the applicability of terms extracted from anchortext as a source of Web page descriptions in the form of tags. With a relatively simple and easy-to-use method, we show that anchortext significantly overlaps with tags obtained from the popular tagging portal del.icio.us. Considering the size and diversity of the user community potentially involved in social tagging, this observation is rather surprising. Furthermore, we show by an evaluation using human-created relevance assessments the general suitability of the anchortext tag generation in terms of user-perceived precision values. The awareness of this easy-to-obtain source of tags could trigger the rise of new tagging portals pushed by this automatic bootstrapping process or be applied in already existing portals to increase the number of tags per page by merely looking at the anchortext which exists anyway.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Experimentation

Keywords

social tagging, Web 2.0, anchortext, tag prediction

*The work presented in this paper was partially supported by the Lion project (Science Foundation Ireland, SFI/02/CE1/I131) and by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322 and by the European project OKKAM No 215032.

1. INTRODUCTION

We live in exciting times in which almost every few months we can enjoy another successful facet of Web 2.0 applications, the number of which is growing at an immense rate. A clear example of this trend are tagging portals like del.icio.us[1] and Flickr[2] in which millions of users publish and annotate resources on the Web. Social Web communities like tagging systems are perceived as the new generation Web and have triggered a new branch of research with already numerous publications in the last few years. Commenting on these novel and interesting features of Web 2.0 seems to be a standard ingredient of current publications. With this study we show that the basic principles that are behind the process of "tagging" in modern Web 2.0 social portals are not much different from creating simple anchortext on the previous Web 1.0.

In Web 1.0, people implicitly annotated resources by putting links to interesting Web sites onto their personal homepages. Hence, tags in the form of HTML anchortext were spread across the entire Web and not gathered in a portal to allow for tag-based browsing or community based aspects. Moreover, anchortext has long been extensively used by Web search engines as a rich source of Web page annotations to improve the search quality [4]. However, such "Web 1.0 tagging" has always been limited to those few users that create their own Web pages, which substantially restricts the user community.

The interesting aspect of Web 2.0 tagging portals is the fact that everybody can participate in the tagging process since it is an extremely easy task, which is certainly one of the reasons behind the success of these portals. Due to the large scale of the tagging community, portals like del.icio.us have accumulated decent annotations in the form of tags for numerous resources. These tags are used for search and navigation and form easy-to-read summaries for the described resources.

However, in this paper we show that many of the tags carefully produced by del.icio.us users have already been freely available on the Web in the form of anchortext. We observe that in the same way as anchortext often repeats the contents of the Web pages that it describes, user generated tags often repeat anchortext. Furthermore, anchortext tag extraction could potentially be used to help distinguish gen-

[1] http://del.icio.us/
[2] http://www.flickr.com/

eral tags from more personal social annotations on Web 2.0 portals.

In summary, this paper makes the following contributions:

- We introduce an approach for large-scale automatic tag generation for Web resources based on anchortext.
- We compare tags extracted from anchortext using our approach to those from the del.icio.us tagging portal and confirm a high overlap between the two.
- We report on a user study performed to measure the precision of tags acquired using our approach and compare it to the precision of del.icio.us tags. Surprisingly, both approaches deliver very similar results which supports our claim on the significant degree of equivalence between social tagging and anchortext annotation.

The remainder of this paper is organized as follows. We first discuss related work in Section 2. In Section 3 we present our approach to extract tags from anchortext. We present our experimental results in Section 4, discuss possible future work in Section 5, and conclude with Section 6.

2. RELATED WORK

In this paper we present a comparison of two methods of annotating a URL: the Web 1.0 approach of anchortext, confined to web authors, and the Web 2.0 method of tagging, accessible to a much larger number of users. In the following, we discuss previous work concerning both anchortext and tagging as resource annotations.

The use of anchortext in web search is well established [4] and the effectiveness of this approach has been proven [10]. A study by Craswell et al. [8] found that in their experiments, ranking based on link anchortext was twice as effective as ranking based on document content. In [10], Eiron and McCurley argue that the reason why anchortext is so effective for web search is because it provides a short and precise summarization of the target pages. The InCommonSense system [2] uses snippets of text surrounding hyperlinks to automatically generate summaries for Web sites. Other applications where anchortext has been exploited include translation of Web queries [19], classification of Web queries by user intent [12], generation of query refinements [18], and entity resolution [11].

The practice of tagging has become increasingly widespread on the Web since 2004 [15], and in the meantime several notable studies of this phenomenon have been published. Mika [21] proposed a tri-partite model of tagging composed of relationships involving resources, users and tags. Other works have studied various aspects of tag usage including the behavior of users of different types of tagging systems [20], tag distribution, tag dynamics and tag-tag correlations [14], and the changes in user activity in tagging systems over time [13].

Various systems have been proposed to solve the problem of tag prediction, using very different approaches. TagAssist [24] automatically generates tags for blog posts based on existing tagged posts. P-TAG [7] suggests personalized tags for Web pages based on both the content of the page and the data on the user's desktop. PicShark [9] is a community-based self-organizing system in which semi-structured metadata is inferred through decentralized instance and schema matching in an image sharing scenario. Sigurbjörnsson et al. [23] proposed a method which uses tag co-occurrence to predict additional tags for tagged images. Recent work by Budura et al. [6] considered the problem of tag inference by leveraging a graph relating resources and introducing a scoring model to rank tags based on their expected relevance w.r.t. a particular resource. The scoring model comprised of tag co-occurrence measures and other IR style means to assess the relevance of tags according to the resource's neighborhood. The same problem was tackled by Heymann et al. in [17] where the authors predicted tags for Web pages in del.icio.us (bookmarks) based on the anchortext of incoming links, Web page content and on the surrounding hosts. The approach predicted tags from the set of 100 most frequent tags by training a classifier which runs a binary classification task for each tag. The classifier is trained on a set of bookmarks which have a large number of attached tags. As opposed to this method, we only consider anchortext and in addition metadata extracted from the Web page header and we use an easy-to-apply method which does not rely on classification. We do not have any initial conditions on the number of postings for our bookmarks, our only requirement being that the bookmark is present in both the datasets we use to run our experiments: a crawl of del.icio.us and a crawl of the Web. In addition, we do not restrict ourselves to a predefined set of tags to consider. By doing so we could expect a gain in precision; however, this would drastically restrict our general approach to only a small number of tags.

In [26], Zhou et al. proposed a framework for modeling document content and annotations, detecting topical information in tags, and integrating this topic-level information into traditional information techniques. Noll and Meinel in [22] compared the metadata supplied by authors of documents (title and keywords of a HTML page) to the metadata created by the readers of documents (del.icio.us tags). They found that at less than 60% of tags are found in the document content or metadata. They did not however investigate anchortext. Bao et al. [3], Yanbe et al. [25], and Heymann et al. [16] have explored the use of social annotations to improve Web search. In [16] the authors also analyzed the occurrence of tags in the pages they annotate, and the text of the pages linked to or from the annotated page. They found that at least 80% of the time a tag was present in at least one of these places, indicating that a substantial amount of tags are redundant for search as the terms would be discovered in the page content regardless. They did not investigate the probability of tag occurrence in anchortext, and they did not attempt to recover tags from anchortext.

3. DERIVING "ANCHORTAGS"

In the old Web 1.0 style, people implicitly annotate resources by putting links to interesting Web sites onto their personal homepages. It is only natural to assume that there is a substantial overlap between these short text snippets and explicitly assigned tags in portals like del.icio.us. However, the intended usage of anchortext is different from social tagging, i.e., anchortext quite often contains navigational information like *click here*, or *read more* which are not very useful as tags, while tags often express personal annotations which might not correspond to the general description of a Web page like *toRead*. The top 20 terms found in anchortext (with common stopwords removed) and del.icio.us tags are shown in Table 1. While there is some overlap between the two lists, in general they have quite different characteristics.

Anchortext Terms		Del.icio.us Tags	
phpbb	website	web	free
xhtml	uk	software	blog
national	adobe	tool	art
valid	reader	design	google
dems	download	reference	opensource
lib	power	program	technology
web	site	music	internet
css	free	css	fun
online	blog	webdesign	science
new	acrobat	new	photography

Table 1: Top 20 anchortext terms and del.icio.us tags ordered by number of occurrences

In order to compare the two types of annotations we propose a method for extracting tags for a Web page from the available anchortext and metadata.

We select all anchortext from incoming links to a page, and from the page header we also include text from the TITLE element and text from the "keywords" and "description" properties within the META element, if present. Since the dataset we use is not a complete crawl of the web, we only have a subset of the total set of anchortexts pointing to any page. Also, many of the URLs we study are not present in the dataset themselves, so we have no metadata for them but instead rely solely on the anchortext pointing to them.

We convert the extracted text to lower-case and remove punctuation. All strings are omitted except those which are at least two characters long, contain one or more letters, and consist only of alphanumeric characters. We remove any words which are on a widely-used list of stopwords[3] containing frequently-occuring words in the English language (e.g. the, for) and also remove words which are commonly found specifically on the Internet (e.g. www, click). Additionally, we apply a stemmer so as to ensure that words with the same root (such as environments and environmental) are considered to be the same.

In standard IR document indexing, a document is described by its content, which is usually a multi-set of terms, then indexed to the form of (docId, score)-pairs where score can be computed based on various scoring models (see below). In our scenario, a document is described by a multi-set of terms (tags) extracted from the anchortext which refers to that document. From the indexing point of view, these two scenarios do not differ much. In order to rank tags, we experiment with two standard IR measures, TF and TF-IDF, to determine those terms which are most strongly relevant to a particular URL.

The term frequency (TF) of a term w.r.t. a particular document describes the number of times a tag occurs for that document. Inverse document frequency (IDF) is simply the inverse of the number of documents (DF) that are tagged (or contain) a particular term. The IDF is used to decrease the influence of overly popular terms. Both TF and TF-IDF can be applied to terms retrieved from anchortext and to tags obtained from del.icio.us.

3.1 Example

Figure 1 shows three example URLs and the top del.icio.us tags and top anchortext tags for each URL ranked by TF,

[3] ftp://ftp.cs.cornell.edu/pub/smart/english.stop

Figure 1: Example tag clouds for URLs (Logos are trademarks of Virgin Radio, BBC and Ecotricity)

which is the default ranking used in del.icio.us. On the first and second glance, it is hard to see a difference in the tags derived from del.icio.us and those obtained from anchortext. On closer inspection, we can, however, observe slight differences in the type of words used. There are some strong indicators that a tag was derived from anchortext and we can observe the same for del.icio.us tags.

We would like to invite the interested reader to participate in a little quiz: for each of the Web sites displayed in Figure 1, deduce which set of tags belong to which source, i.e., which tags have been obtained from anchortext and which from del.icio.us. The solution can be found in the footnote[4] on the following page.

We hope it was not too easy to come up with the correct solution for the quiz. Here are some hints and explanations. As mentioned earlier, tags derived from anchortext as well as tags from del.icio.us reflect their usage pattern. Del.icio.us contains organizational tags like "toRead" while tags extracted from anchortext are terms like "listen" which is often used for web radio stations. Another example is that names of people and organizations are usually present in anchortext but not too often in del.icio.us tags. However, guessing the origin of the two sets of tags is not easy and this is one of the points we want to make in the remainder of this paper.

4. EXPERIMENTAL STUDY

In this section we compare the tags that we can extract from anchortext and metadata on the Web to the tags assigned to the corresponding webpages in del.icio.us. In order to perform our analysis, we required a large collection of Web documents and corresponding tags. We therefore used the following two datasets in our study, one from a Web crawl and one from a crawl of del.icio.us:

Web Collection: In our experiments we use the WEBSPAM-UK2007 [1] dataset as a source of anchortext and metadata. For calculating the indegree distribution we use the full Web collection of approximately 106M webpages from the *.uk* domain. For all other experiments we use the summary version containing around 12M pages from the same domain.

Del.icio.us Dataset: The tags which we use in our analysis are taken from a crawl of del.icio.us which was carried out in 2007. This dataset contains tags for approximately 4.5 million URLs.

For our analysis we are only interested in URLs which occur in both the del.icio.us dataset and the Web collection summary version. These amount to 192,489 URLs. For each URL, we have the complete set of del.icio.us tags assigned at the time. However the *.uk* domain represents only a small fraction of the Web, hence for most URLs we only have a small fraction of the corresponding anchortext. Therefore the quality of the tags which we infer is probably lower than what we could achieve with a complete crawl, however we aim to show that even with this Web dataset we can still report interesting results. In order to avoid confusion, in the remainder of this paper we will refer to two types of tags:

Anchortags: Tags that we have extracted from anchortext using the method described in Section 3.

Del.icio.us tags: Tags that users have assigned to URLs in del.icio.us.

Firstly, we study properties of the two datasets to get an overview of the general behaviour of users in tagging and linking. Next, we directly compare the del.icio.us tags and anchortags assigned to documents, in order to ascertain how similar the tags extracted from the anchortext are to those explicitly assigned by users of del.icio.us. Finally we conduct a user study where we ask evaluators to judge the relevance of tags, both for del.icio.us tags and anchortags.

4.1 Dataset Characteristics

We first compare the distribution of inlinks on the Web to the distribution of tags in del.icio.us. Figure 2 shows the cumulative indegree distribution of our Web collection dataset, and Figure 3 shows the cumulative distribution of the number of times URLs are tagged in del.icio.us. It appears that parts of these curves follow a power law distribution. Therefore in each figure we have fitted a line to the curve proportional to the power law function $k^{-\alpha}$. For a cumulative distribution, the exponent is $\alpha - 1$ rather than α. The slope of the line fitted in Figure 2 in the range of $y \in [0.01, 10^{-7}]$ is 0.99, indicating a power law exponent α of 1.99. This is comparable to previous exponent of 2.1 calculated for the whole World Wide Web [5]. The slope of the line fitted in Figure 2 in the range of $y \in [0.1, 10^{-6}]$ is 1.16, indicating a power law exponent α of 2.16. From these two figures we can see that both datasets share a quite similar distribution; i.e. user patterns in tagging follow user patterns in creating hyperlinks to pages. Specifically, many pages are linked to or tagged very few times, but a small number of pages attract many links or tags.

Figure 4 shows the relationship between the indegree of URLs and the number of tags which they receive in del.icio.us. We average the tag counts across documents ordered by indegree to smooth the distribution. For example, the first point indicates that for URLs with an indegree between 1 and 10, the average number of tags assigned in del.icio.us is approximately 9. The average number of tags received by documents increases with indegree, showing that the same pages which receive many inlinks also tend to receive many tags.

[4]Here is the solution of the quiz: For the Virgin Radio website the tags on the right side are derived from anchortext. For BBC and ecotricity.co.uk the tags on the left side are derived from anchortext.

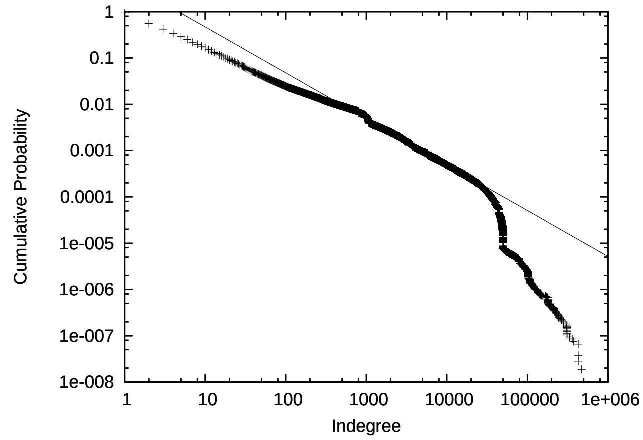

Figure 2: Cumulative distribution of the inlinks of documents from 106M UK Web pages.

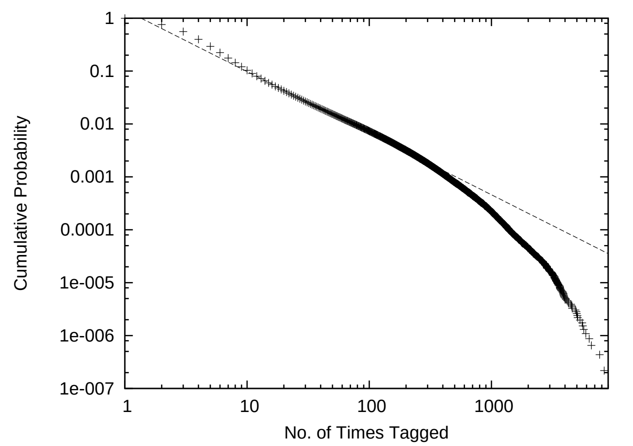

Figure 3: Cumulative distribution of number of times documents are tagged in del.icio.us.

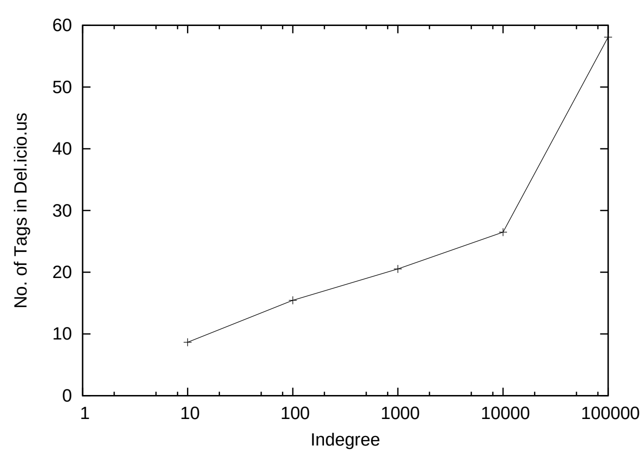

Figure 4: Number of del.icio.us tags per document, averaged by indegree.

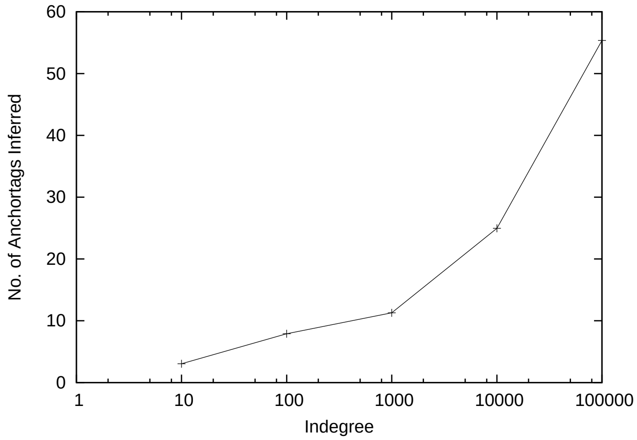

Figure 5: Number of anchortags inferred per document, averaged by indegree.

Figure 5 shows the relationship between the indegree of URLs and the number of anchortags that we can infer. Again we smooth the distribution by averaging the tag counts across documents ordered by indegree. For example, the first point indicates that for URLs with an indegree between 1 and 10, the average number of anchortags we can infer is approximately 3. Since a higher number of inlinks generally results in a greater amount of anchortext, more inlinks to a document means that more anchortags can usually be inferred.

From these simple analyses, we can conclude that there are similarities in the way that users tag and the way that they link to pages, and also that the more popular a page is on the Web, the more tags we should be able to extract. These results are encouraging and give weight to the idea that the practices of linking and tagging share common properties. However it remains to be seen whether the tags we can extract from anchortext share any significant overlap with the tags in del.icio.us, and whether or not they are considered relevant by users. We attempt to answer these questions using the measures described in the next section.

4.2 Measures of Interest

In order to measure the quality of our anchortags, we use two benchmarks. Firstly, we compare the anchortags extracted from the Web collection against the tags from del.icio.us, in order to assess the extent to which our method generates tags that users have also provided. We assume that the user-provided tags are always an appropriate annotation for the relevant URL, and therefore are a good standard to measure the automatically-generated annotations against. Secondly, we conduct a human evaluation on the predicted tags in order to gauge the success of our method in predicting tags regardless of whether they are already present in del.icio.us. For a random selection of URLs, we ask human judges to decide whether or not the top ranking anchortags are relevant to the corresponding URL. We also evaluate del.icio.us tags assigned to the same URLs for comparison. The evaluators are not told how the tags that they assess are derived.

We use the following metrics in our evaluation:

Deliciousness at rank k (D@k) We suggest the term Deliciousness as a measure of how the predicted tags compare to the tags assigned in del.icio.us. Deliciousness at rank k measures the average proportion of the top-k predicted tags that are also assigned to the corresponding document in del.icio.us, averaged over all documents. In other words, Deliciousness represents a measure of relative precision where the ground truth is considered to be the set of tags extracted from del.icio.us.

Anchortextness at rank k (A@k) We use the term Anchortextness as a measure of how the tags assigned in del.icio.us compare to the predicted anchortags. Anchortextness at rank k measures the average proportion of the top-k tags in del.icio.us that can also be predicted from anchortext, averaged over all documents. Anchortextness represents a measure of relative recall where the ground truth is taken to be the set of del.icio.us tags.

Precision at rank k (P@k) Precision at rank k measures the average proportion of the top-k predicted tags that are deemed relevant by users and can be considered our absolute precision.

We do not report on absolute recall since the total number of relevant tags for a document is potentially unlimited.

Ranking	D@1	D@2	D@3	D@4	D@5
TF	0.44	0.43	0.42	0.39	0.36
TF-IDF	0.38	0.39	0.38	0.35	0.32

Table 2: Deliciousness@1-5.

Ranking	A@1	A@2	A@3	A@4	A@5
TF	0.40	0.37	0.34	0.32	0.30
TF-IDF	0.38	0.37	0.34	0.32	0.30

Table 3: Anchortextness@1-5.

4.3 Anchortags vs. del.icio.us tags

Table 2 and Table 3 show the Deliciousness and Anchortextness results respectively for values of k from 1 to 5. Hence we only test for URLs which have at least 5 del.icio.us tags to compare against. We employed both the TF and TF-IDF ranking methods. We compare anchortags ranked by each method to the del.icio.us tags ranked by the same method. Surprisingly, even with our naive approach we find that approximately 40% of the top anchortags extracted are also found in del.icio.us. This raises the question whether a large portion of the tags in del.icio.us could be considered redundant for search purposes as search engines would already have located them from anchortext or metadata. However there is still the issue of the 60% of predicted anchortags which have not been entered in del.icio.us, but may still potentially serve as useful annotations. In order to evaluate these tags that are not in del.icio.us but may be relevant nevertheless, we conducted a user study, which is described in the next section.

Figure 6 shows the relationship between Deliciousness@5 and the total number of anchortags which we were able to infer from anchortext. A first observation is that we achieve a high Deliciousness@5 value even when the number of available anchortags is relatively low. The results of our method are consistently good as the amount of available anchortext

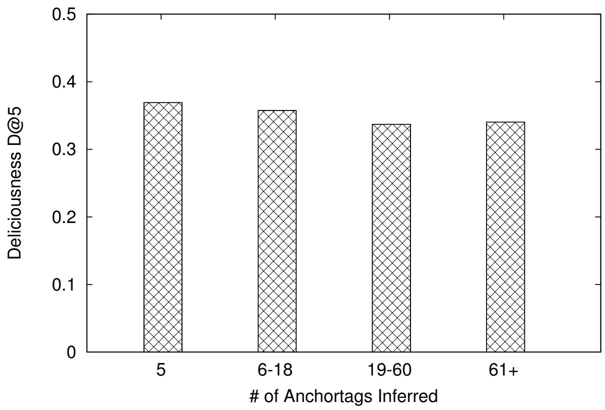

Figure 6: Deliciousness@5 for varying numbers of anchortags inferred.

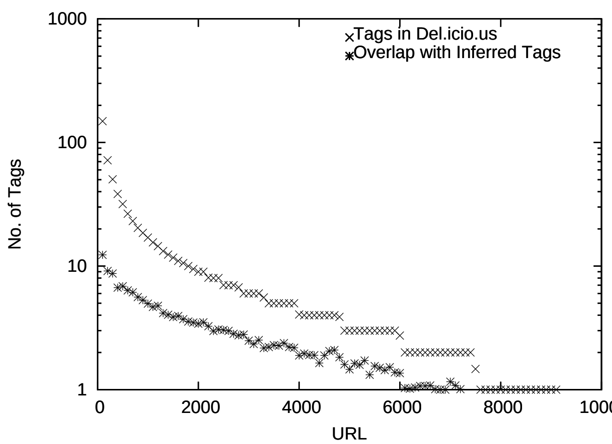

Figure 7: Number of del.icio.us tags per URL, ordered by tag count, and overlap with anchortags

varies, however they are not increasing. We believe that with a more sophisticated ranking method, these results could improve as the amount of available text increases. Another explanation (as we will see later) is that with an increasing amount of anchortext we infer new, relevant tags which have not (yet) been assigned in del.icio.us.

Figure 7 shows the distribution of del.icio.us tags, and the number of anchortags which can be inferred for the corresponding URLs (as in the previous tables, we only include URLs for which we can extract at least 5 tags). It can be seen that an overlap exists, although it is quite small. However, the fact that our results for Deliciousness and Anchortextness are relatively good, shows that the anchortags which are relevant must rank fairly highly. Considering that the Web collection dataset contains only a tiny fraction of the anchortext and metadata available on the Web, whereas the del.icio.us dataset covers a sizeable portion of del.icio.us, we would expect that in larger Web datasets such as that indexed by search engines, the overlap would be much greater.

4.4 User Evaluation

In order to conduct a user study we have randomly selected 80 URLs from del.icio.us and for each we generated

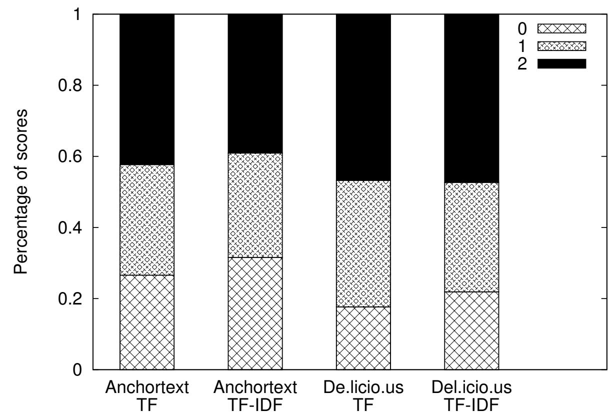

Figure 8: Distribution of 0, 1 and 2 scores for each method of tag extraction and ranking.

Tags	Ranking	P@1	P@2	P@3	P@4	P@5
Anchortags	TF	0.48	0.46	0.39	0.39	0.36
Anchortags	TF-IDF	0.54	0.41	0.38	0.35	0.33
Del.icio.us	TF	0.61	0.52	0.50	0.45	0.42
Del.icio.us	TF-IDF	0.55	0.49	0.47	0.44	0.43

Table 4: Precision@1-5 when we consider tags which were scored on average 1.5 or above to be relevant.

a set of corresponding tags by selecting the top-5 tags from del.icio.us and the top-5 tags which were predicted from anchortext. Since for some URLs we have very many inferred anchortags, and for others less, we wanted to ensure that we had an even spread of sample URLs across this range. We therefore choose 20 Web pages from each of the ranges shown in Figure 6. We then asked 25 users to evaluate the relevance of these tags w.r.t. the URLs. In order to distinguish between different degrees of relevance, we asked the users to grade each tag with a score of 0, 1 or 2, where 0 indicates a non-relevant tag, 1 is given to the tags which are somewhat relevant, and 2 denotes a high degree of relevance. In order to lower the degree of subjectivity in our study we asked three different users to annotate each URL and later aggregated their scores.

Figure 8 shows the distribution of 0, 1 and 2 scores for each method of tag ranking and extraction. The anchortags columns have a greater number of 0 scores, however the gap in 2 scores assigned is much lower. The two ranking methods TF and TF-IDF yield similar results from the user's point of view, with TF delivering slightly less 0 scores.

Table 4 shows the precision results when we consider tags scored 1.5 or above to be relevant, and Table 5 shows the results when we consider tags scored 1.0 or above to be relevant. In both tables we can see that the del.icio.us tags are considered more relevant than the inferred tags; however the gap is not extremely large. In Table 5 the rating of del.icio.us tags for P@5 is 78%, compared to 66% for anchortags. This also indicates that many of the tags that users assign in del.icio.us are not considered relevant by other people which points out that we are dealing with subjective data. We also see that the usual del.icio.us method of ranking, TF, outperforms TF-IDF. However for ranking anchortags it seems that TF-IDF or some other method should also be considered.

Tags	Ranking	P@1	P@2	P@3	P@4	P@5
Anchortags	TF	0.78	0.76	0.69	0.67	0.66
Anchortags	TF-IDF	0.80	0.70	0.66	0.61	0.60
Del.icio.us	TF	0.86	0.84	0.82	0.80	0.78
Del.icio.us	TF-IDF	0.75	0.75	0.75	0.75	0.74

Table 5: Precision@1-5 when we consider tags which were scored on average 1 or above to be relevant.

Tags	Ranking	P@1	P@2	P@3	P@4	P@5
Anchortags	TF	0.85	0.80	0.73	0.69	0.67

Table 6: Precision@1-5 for those URLs for which we can extract at least 19 anchortags. Tags which were scored on average 1 or above are considered to be relevant.

For cases where we have a large amount of anchortext, our accuracy improves considerably. In Table 6, we show the precision results for when we rank by TF only URLs which occur in the two highest ranges shown in Figure 6, i.e. for which at least 19 anchortags can be inferred. In these cases P@1 rises from 78% to 85%. With a large Web dataset of the scale that search engines possess, and a more refined ranking method, even better results should be possible.

An important aspect of tagging is its inherent subjectivity which we wanted to measure in terms of user disagreement when rating tags. In order to quantify the amount of disagreement between users, we use the average deviation measure. This is the average of the absolute deviation of each score from the mean for that (URL, tag) pair. We found that the average deviation is 0.34 and does not vary much between anchortags and del.icio.us tags. This gives us an interesting insight, namely that a similar amount of subjectivity can be found in both types of annotations. In Table 7 we take a closer look at the levels of agreements between users. While the ratings assigned tend to be similar, discrepancies are quite frequent if we consider that users agree to some extent only in $\sim$50% of the cases. This indicates that tag evaluation is highly subjective and opinions on the relevance of a tag vary significantly from person to person.

5. FUTURE WORK

As future work we plan to look at ways to combine the two types of annotations (anchortags and social tags) and to gain some insight into their differences. These differences could later on be leveraged to derive tag semantics in terms of personal or more general tags, or tags which describe the content of a resource as opposed to tags which capture the contextual information surrounding a resource.

Another important problem which can be addressed is a meaningful ranking of tags according to their relevance as opposed to the simple popularity-based mechanisms used at the moment. This problem is especially relevant in the context of new approaches [17, 6, 23] for automatic tag inference where proper identification of most relevant tags will become a central issue.

6. CONCLUSIONS

With this paper we have contributed a study on how close Web 1.0 anchortext usage is related to Web 2.0 usage of tags. Can we now answer the initial question on how your Grandma used to tag? The interesting result of our study is that tags derived from anchortext substantially overlap with tags within the popular tagging portal del.icio.us. Furthermore, with a rather simple and easy to deploy method to extract tags from anchortext, we show that the quality of those “anchortags” is quite comparable to tags created explicitly in del.icio.us.

This work is not meant to compete with state of the art tag inference mechanisms that target particularly high precision values. The point we wanted to make here, is that without much of sophisticated processing a substantial amount of relevant tags can be generated from anchortext. This can be used either for bootstrapping tagging portals or for enriching the set of tags already present in portals. While conducting experiments we found the tags generated from anchortext to be of surprisingly good quality, often fully comparable to the del.icio.us tags for the same resource. Indeed our user study revealed the precision achieved with both approaches to be very similar, leading to the somehow provocative title of the work. However, the social annotation of Web 2.0 is much more than simple anchortext as it is highly user-dependent with millions of people actively participating to form a personalized information source overlayed on top of content-rich communities.

Extent of agreement	Frequency
All 3 scores agree	34.97%
All 3 scores almost agree	49.90%
All other cases	15.13%

Table 7: Breakdown of user agreement. Almost agree means that 2 scores are equal and the other score is just one point higher or lower.

7. REFERENCES

[1] Yahoo! Research: “Web Collection UK-2007”. http://research.yahoo.com/ Crawled by the Laboratory of Web Algorithmics, University of Milan, http://law.dsi.unimi.it/. URL retrieved 05 2008.

[2] E. Amitay and C. Paris. Automatically summarising web sites: is there a way around it? In *CIKM ’00: Proceedings of the ninth international conference on Information and knowledge management*, pages 173–179, New York, NY, USA, 2000. ACM.

[3] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW ’07: Proceedings of the 16th international conference on World Wide Web*, pages 501–510, New York, NY, USA, 2007. ACM.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.

[5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer Networks*, 33(1-6):309–320, 2000.

[6] A. Budura, S. Michel, P. Cudré-Mauroux, and K. Aberer. To tag or not to tag - harvesting adjacent metadata in large-scale tagging systems. In *SIGIR ’08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in*

information retrieval, pages 733–734, New York, NY, USA, 2008. ACM.

[7] P. A. Chirita, S. Costache, W. Nejdl, and S. Handschuh. P-tag: large scale automatic generation of personalized annotation tags for the web. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 845–854, New York, NY, USA, 2007. ACM.

[8] N. Craswell, D. Hawking, and S. Robertson. Effective site finding using link anchor information. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 250–257, New York, NY, USA, 2001. ACM.

[9] P. Cudre-Mauroux, A. Budura, M. Hauswirth, and K. Aberer. Picshark: Mitigating metadata scarcity through large-scale p2p collaboration. In *VLDB Journal, Accepted for publication*, 2008.

[10] N. Eiron and K. S. McCurley. Analysis of anchor text for web search. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 459–460, New York, NY, USA, 2003. ACM.

[11] E. Elmacioglu, M.-Y. Kan, D. Lee, and Y. Zhang. Web based linkage. In *WIDM '07: Proceedings of the 9th annual ACM international workshop on Web information and data management*, pages 121–128, New York, NY, USA, 2007. ACM.

[12] A. Fujii. Modeling anchor text and classifying queries to enhance web document retrieval. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pages 337–346, New York, NY, USA, 2008. ACM.

[13] S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *J. Inf. Sci.*, 32(2):198–208, 2006.

[14] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 211–220, New York, NY, USA, 2007. ACM.

[15] T. Hammond, T. Hannay, B. Lund, and J. Scott. Social Bookmarking Tools (I). *D-Lib Magazine*, 11(4):1082–9873, 2005.

[16] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 195–206, New York, NY, USA, 2008. ACM.

[17] P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008.

[18] R. Kraft and J. Zien. Mining anchor text for query refinement. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 666–674, New York, NY, USA, 2004. ACM.

[19] W.-H. Lu, L.-F. Chien, and H.-J. Lee. Anchor text mining for translation of web queries: A transitive translation approach. *ACM Trans. Inf. Syst.*, 22(2):242–269, 2004.

[20] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006. ACM.

[21] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *International Semantic Web Conference*, Lecture Notes in Computer Science, pages 522–536. Springer, 2005.

[22] M. G. Noll and C. Meinel. Authors vs. readers: a comparative study of document metadata and content in the www. In *DocEng '07: Proceedings of the 2007 ACM symposium on Document engineering*, pages 177–186, New York, NY, USA, 2007. ACM.

[23] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pages 327–336, New York, NY, USA, 2008. ACM.

[24] S. Sood, K. Hammond, S. Owsley, and L. Birnbaum. TagAssist: Automatic Tag Suggestion for Blog Posts. In *Proceedings of the 1st International Conference on Weblogs and Social Media*, 2007.

[25] Y. Yanbe, A. Jatowt, S. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *JCDL '07: Proceedings of the 7th ACM/IEEE joint conference on Digital libraries*, pages 107–116, New York, NY, USA, 2007. ACM.

[26] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles. Exploring social annotations for information retrieval. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pages 715–724, New York, NY, USA, 2008. ACM.