

# A Narrative-Based Alternative to Tagging

Nuno Tomás, Tiago Guerreiro, Joaquim A. Jorge, Daniel Gonçalves  
INESC-ID / DEI, IST-TULisbon  
Av. Rovisco Pais, 1000 Lisboa, Portugal  
nmrt@ist.utl.pt, tjvg@vimmi.inesc-id.pt, {jaj,daniel.goncalves}@inesc-id.pt

## ABSTRACT

The enormous dissemination of multimedia information over the past few years has led to mechanisms to support its organization, cataloging and search through descriptions or keywords. A popular way of associating such descriptions to content is tagging as can be found in popular sites such as Flickr (for images) or Delicious (bookmarks). This method allows users to associate tags to media, richly describing its content and may help in its retrieval at a later time. However, the process is mostly unstructured, leading to several problems. Nothing guarantees that the tags used are the most appropriate or the same tags are used in similar situations, making retrieval difficult.

Our approach relies on narrative-based interfaces which use stories as an organizing principle for tagging media. Given that humans have used stories to communicate since the dawn of time, narrative is a natural form of interaction. By inter-relating bits of information into a coherent whole, stories convey data in a rich, structured way. A study carried out with 40 users over a period of three months shows that users convey almost six times more information when using narratives to describe their media than what is typical of traditional methods. Furthermore, our pilot study saw narratives increasing tag reuse to 94%. Finally, other problems found in tagging such as synonyms and polysemy were notably absent from story-generated tags.

## ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. H.3.3 Information Search and Retrieval:

## General Terms

Design, Experimentation, Human Factors

## Keywords

Tagging, Narrative-based interfaces, Digital Media

## INTRODUCTION

The amount of information we store nowadays, especially

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'10, June 13–16, 2010, Toronto, Ontario, Canada.

Copyright 2010 ACM 978-1-4503-0041-4/10/06...\$10.00.

on online services such as flickr and delicious, makes it hard to retrieve, especially for non-textual media at a later time. Thus, some kind of classification scheme must be used to organize the information. One of the most common indexing schemes is tagging, which allows users to freely associate keywords to content. Most sites provide some form of tagging which is relatively easy to use and relatively effective. Tagging allows users to assign several words to the same object, overcoming one of the greatest problems with traditional hierarchy-based schemes. Also, tagging is often collaborative, as in most sites users can search for content of others, and sometimes tag such content. From this behaviour a *folksonomy* [7] might arise: a set of popular tags shared by a community, used by general consensus to describe certain objects.

Although very popular, these tagging systems suffer some problems [2]. These problems combined lead to *low tag reuse*. While ideally the same tags would be used by everyone in a consistent way, in practice that is not the case. Not only do different people use different tags to describe the same objects, but the same user might change the tags in use as time goes by. This causes problems when retrieving the tagged objects.

To overcome some of these limitations, we propose a semi-structured way to describe media objects in general and digital photos in particular. Our approach elicits stories from users to describe their media. In this way people will convey more information to the system. Additionally, the sequence and structure enforced by narratives should yield more consistent descriptions. Narratives are a natural and structured approach [4] for humans to communicate. Storytelling is innate to humans, who grew up listening to stories told by their parents and grandparents. The three key elements in narrative, structure, sequence and rhythm can overcome problems found in free association mechanisms characteristic of conventional tagging such as ambiguity, polysemy, and synonyms. This more consistent association of common terms should result in higher tag reuse.

To validate our approach, we started out by studying how users tell stories about their media, especially photos. From that study we found the most relevant elements emerging in those stories, and how the different elements are related, allowing us to infer archetypical story patterns. In a second phase, we developed a website, using the lessons learned from the aforementioned study, where users could store their personal photos and then associate one or more stories

to each photo. An alternate version of the site resorting to traditional tagging was used by a different set of users. We measured their behaviours over the course of three months. Most notable among our findings is the fact that narrative-based tagging lead to a tag reuse of 89%, vs 36% for traditional tagging.

In Section 2, we give an overview of tagging systems and their major problems. Section 3 presents the study on photo-describing stories. The website created for the user studies is explained in Section 4. In Section 5 we present the results from user evaluation. Section 6 concludes this paper presenting possible avenues for future work.

## RELATED WORK

Tagging enables users to easily describe their content easily and provides a relatively good basis for retrieval [6]. However, it is not without its problems [2, 7], including cognitive problems associated to having to remember which tags were associated to which specific content [5].

*Polysemy* occurs when a word has multiple meanings and its meaning can only be disambiguated in context of the sentence where it appears. One example is the word "position", which can be applied in various contexts such as: "Put yourself in a correct position for the picture" (body posture) and "achieved a good position in the company" (rank). Superficially, polysemy is similar to *homonyms*, where a word can have multiple unrelated meanings. However, homonyms can be largely circumvented by adding additional tags to the search query, which refer to related concepts to filter out unwanted search results.

*Synonyms* present another challenge because many words can have the same or very similar meaning. They pose complicated problems in a collaborative system due to the inconsistency of the terms used in tagging, making the task of searching more complex. One solution is either to have all users standardize on a consensus, which is often impossible. Another approach is to have users enter multiple combinations of synonym tags during searches in order to account for many possible outcomes.

Another problem associated with tagging is related to the *basic levels of language*. In fact, the level of detail with which each user describes the same content can vary, depending on their experience or degree of expertise in the area. For example, an animal lover and particularly a dog specialist can tag a new picture using the breed ("beagle", for example), while the average user would tag it as "dog". Both variations in the degree of knowledge, as changes in social and cultural from the users, may cause variations on the basic level of description.

All these problems combined explain the *low tag reuse* identified in the literature. While some very common tags are indeed reused very often, that is not the case for the large majority of them. A study of tag use in delicious [2] has shown that most tags are used by a very small number of people, following an inverse power-law distribution,

both for a single user and for the entire community. We aim to alleviate these problems by following a narrative approach to tagging, described below.

## STORIES ABOUT PHOTOS

A story is more than a series of facts strung together without apparent logic. Rather a good story contains elements that, together, convey a consistent structure obeying well-defined rules. As we shall see, it is possible to identify certain elements that appear more or less often in the stories told by people. When designing the narrative-based interface we take into account not only each story element individually, but also the relationships between them and how they fit in the overall structure. This structure must be recreated so that people can tell a story as naturally as possible, ideally as they would to other people. To find out exactly what to expect in photo-describing stories, we conducted a set of interviews where people told stories describing their photos.

Twenty interviews were conducted in order to get a good sample of possible stories. Each user told three different stories, thus yielding a total of 60 stories to analyse. We used a semi-structured interview to ensure that the interviewees felt free to tell their stories while the interviewer retained the possibility to guide the subjects, should they somehow stray from the objectives of the interview. The interviewer had at his disposal a number of guide questions for different elements of the stories. They were used only on extreme cases, when the interviewee was at a loss about what to say next. The interviews were recorded and transcribed for further analysis.

The subjects were asked to choose three personal photos, and then to tell a story for each photo. The only requirement was that the interviewee had to remember the photo, so that a story could be told. It would not be correct to show them a predetermined set of photos from somebody else's, as only personal pictures would be likely to elicit useful stories. Indeed, stories often contain autobiographic and contextual elements that may be important to a particular user, but irrelevant to others.

## Analyzing the interviews

Below is an example of a story told by a user, confirming the large quantity of information present in a small snippet:

*"This first picture was taken in April this year, the wedding of my older cousin ... I'm in the photo, my sister and my grandmother also... ah, I am a little better dressed than usual, jacket, beige trousers. This picture was taken in Loures... in the afternoon. A lot of people were there. It was the marriage of cousins here in the father's side, was the marriage of the oldest daughter, who is 35 years old. "*

The content analysis was performed according to a guide containing 11 different categories specified below, that describe what possible elements might be found. These categories were fine-tuned in two preliminary interviews.

The elements are: **Time** (when the photograph was taken, including specific holidays or seasons); **Location** (whether geographic locations or landmarks); **Author** (a name or a degree of kinship); **Purpose** (reason why the photo was taken); **Photo Type** (portrait, landscape, urban, etc.); **Size** (large, medium, normal, small, etc.); **Event** (in some way connected to the photo or the holder's photograph: trip to Italy, his grandfather's birthday, etc.); **Device** (Digital / analogue machine, cameraphone, etc.); **Description** (content that describes the picture); **People** (people who are both in picture as those in the story behind the photo); **Quality** (good, bad, etc. ).

We collected information about not only whether an element occurred in a story or not, but also about the frequency of the elements (some appeared more than once per story) to give us an idea of which were the most easily remembered by users and therefore the most relevant.

We distinguished two types of elements: the induced, ie, those that were elicited by some intervention by the interviewer to help interviewees to resume their stories, and the spontaneous were people didn't need any reminder and naturally reported the element. Negative responses to questions from the interviewer were also considered. For instance, mentioning "there were no more people in the group" is different from not knowing if there were more people or not. In the first case the subjects know something, in the second they do not. This distinction was also taken into account in the analysis of the interviews.

Besides analysing occurrences of each element we also studied their relations. We counted the number of times that a specific element precedes another in stories. This can reveal the order in which the elements have been referred, giving an idea about the relations between them. We account only for transitions where the second element was not induced, as only then would an actual connection between the elements exist in the mind of the subject.

## Results

Of the 20 users, 13 (65%) were male, and 7 (35%) female, with ages from 18 to 57 years. Most had contact with a computer and were regular users of the Internet (90%). Also, many (85%) had an account in websites that store personal photos. Two belonged to a different age group and had little or no experience with computers. These users were added to the interviewee group to explore whether more experienced people, with different life experiences could tell stories with different content. We did not find any such differences. This group of people, although not very diverse in terms of age, is representative of most users of the services we were considering in our research.

### Story Elements

We can observe the frequency of elements in Figure 1. Elements such as "Location" and "Time" are the ones used with greater frequency and therefore, in general, are more easily remembered. Figure 1 also shows the frequencies of

the elements were induced by interviewers in the stories. It can be seen that the elements most often induced were "Author", "Device" "Type" and "Dimension". But as we have seen, for example the element "Author" or "Device" are elements with high frequency, indicating that although important, they are not easily remembered by users without some outside help.

Elements less frequently induced include "Event", "Description" and "People". While they do not show among the top elements referred, they appear in stories mostly spontaneously. This may indicate that these elements are important and type of elicitation is needed to remember them. "Event"s, and more understandably "Local"s and "Time"s, are important elements as they occur easily and most often spontaneously in stories.

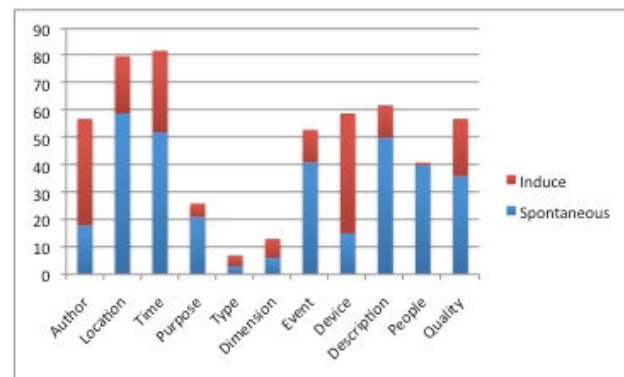


Figure 1 – Frequency of Elements

The more stories an element appears in, the more important it becomes when describing a photo. The elements "Time" and "Local" have a frequency well above their number of occurrences. This redundancy may mean that either these elements are complicated to describe, or users make several references in order to clarify the information. Even so, they were present in almost all 60 stories analyzed.

### Story Structure

After taking into account and normalizing the results for the different element frequencies, we find that the most recurrent transitions are "Device→Quality", "Time→Location" and "Location→Event". Using frequency and transition probability data, we trained Hidden Markov Models [9] to infer an archetypical story structure: Location, Time, Device, Quality, Size, Author, People, Description, Purpose, Type and Event. That was the basis for our prototype, described below.

### Design Guidelines

The results presented in previous sections allow designers to infer some basic principles for narrative interface design. Overall, little need for customization is necessary, as we found no relevant differences in stories for users of different gender or age group. Some elements appeared more often as induced elements, but still seemed to be important to interviewees, showing it is important to

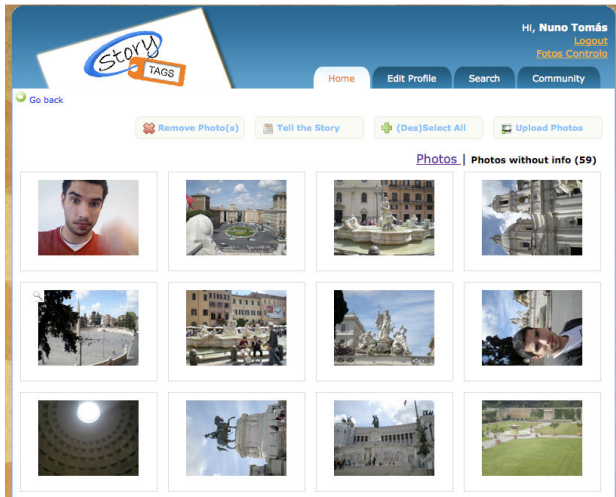


Figure 2 – Application Interface

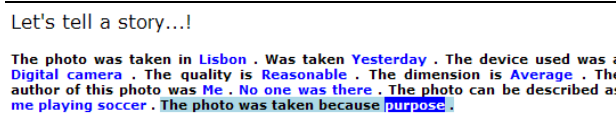


Figure 3 – Story example

dialogue with users to get as much information as possible. Also, ambiguity and inaccuracy is common in stories and should be taken into account. A possibility is to disambiguate using other story elements to provide context. Finally, resorting to the knowledge of which elements are the most frequent and relevant, and also what kind of information to expect when the users mention those elements. we can craft interfaces that help the users to tell their stories in natural and effective ways.

## INTERFACE

Figure 2 depicts the application interface after the user login. This is a full-fledged site where users can upload their photos and retrieve those of other users in the community as well as their own.

The storytelling process was based on Quill, an application for personal documents retrieval using narratives [3]. Each element is suggested, in turn, to the user, in the archetypical order found before. The user has the freedom to mention another in its stead, however. As the different elements are presented to the user, different options, selected from the interviews made, are presented to them. Those options can be a set of pre-defined answers or a box that can be filled with text. The pre-defined possible answers were chosen based on the results of the interviews, when it became clear that, for some elements, most users would mention only a fairly small set of things. Users have the option to select “can’t remember” when the element is not relevant to the story. Furthermore, it is possible to edit any element at a later time, simply by clicking on the corresponding text, and then changing it. To recreate as faithfully as possible the act of storytelling, users may reuse the same element several times during the dialog, as happened in several

interviews, adding repeat elements, to add more information to the story. Figure 3 shows an example of a narrative description of a photo.

As previously stated, users search for a photo among those in the community by performing a query the same way as previously described. As users fill different elements into a story, photos will show up as the result for the partial search thus formulated. We defined a “story-difference” metric, accounting both the structure and the values of different elements. Thus, by calculating this value to compare a story being told as a query to those stored when describing the photos, we are able to find possible matches.

## Alternate Version

We also created an alternate version of the site where photos are described using traditional tagging processes. It mimics Flickr’s tagging process at the time of the conception of the system. We did this since we want to prove that more and better information can be conveyed by narratives than just by tagging. By creating an alternate version of the site, we could perform our study with just one variable: the way tagging is done. All other features are exactly the same. Thus, if some differences are found, we can be confident they are due to the tagging/narratives annotation processes.

## EVALUATION

Forty people used the system over a three month period. They were divided into two groups, in which one inserted tags using a traditional method, while the other group used the system based on stories presented and studied in this paper. The data was analyzed according to the information entered into the system, the time it took to annotate each picture and the time users needed to retrieve photos.

The analysis considered a set of 1229 different photos entered into both systems. Of those, 52% were entered in the narrative-based version of the site, and 48% using the traditional tagging system.

For the narrative-tagging system, we see that each story contained on average 7.96 elements ( $\sigma = 1.16$ ). As each element is in itself a tidbit of information describing the photo, we considered each element as equivalent to a tag in the traditional system for comparison purposes. However, it should be noted that this does skew the comparison towards traditional tagging, since several story elements contain more than one word, and if separated into individual tags, the actual number of those would be greater. We follow this conservative approach, though, as no deeper analysis of the story elements’ contents has been performed yet. Using traditional tagging, users submitted an average of 1.35 tags per photo ( $\sigma = 0.47$ ), taking 45.6 seconds on average to associate tags to each photo ( $\sigma = 10.79$ ). We see, thus, that *by using narrative tagging users are able to convey 5.9 times more information about their photos*. That extra information might make the difference between being able to retrieve that photo at a later time or not.

In terms of tag reuse, taking into account the number of tags that were entered by users in both systems and the distinct tags in each system, we conclude that the narrative-based approach led to a greater reuse of tags. Using stories, the users, overall, entered 5363 elements in their stories. Of those, only 310 were distinct. This corresponds to 94% of story elements having been reused at least once. On the other hand, for traditional tagging, the users entered a total of 793 tags, of which only 36% were used more than once. This shows that our approach is able to lead the users to far greater reuse of information than traditional approaches.

It might be argued that, since in the narrative-based approach some story elements provided the user with pre-determined values to be chosen from, this would artificially increase tag reuse. That information is nevertheless relevant and was found to be spontaneous in the interviews. Removing those elements from the analysis does indeed lower the tag reuse percentage for narratives. Considering only free-form entry elements, we find a reuse of 89%, instead of the original 94%. Those elements did influence the results, but even accounting for that, narrative-based tagging still is, by far, able to elicit more consistent information from the users.

Analyzing the data, we could see that problems characteristic of traditional tagging such as polysemy and synonyms appeared as expected in the traditional system. However, in the narrative-tagging system these problems *were not observed*. This is a direct result that each tag is considered in the context of the element to which corresponds in the story. The tags presented in that context do not show signs of ambiguity and thereby problems such as synonyms and polysemy do not show up.

## CONCLUSIONS

This paper presents a new methodology that helps users in the process of tagging a photo. Instead of associating individual words to a photo, the user tells a story describing it. Furthermore, the narrative tagging method encompasses traditional tagging. Even if users are at a loss to enter relevant information that stories would allow them to enter, they can still enter freeform text in fields such as "Description", placing there what otherwise would be isolated tags. As we found out oftentimes the semi-structured nature of storytelling will elicit users to enter more information to the system.

To assess the merits of our approach we performed a user study. Users of the narrative-based system **added 5.9 times more information on average** than users of traditional tagging systems. Most notably, tag reuse **reached an average value of 94%** in the narrative-based system (89% after removing preset elements from stories), substantially higher than the 36% obtained in the traditional system. Interestingly enough, common problems like polysemy and synonyms do not appear in the narrative-tagging system. This is because in our approach tags are associated with a story element, which both provides context and removes

possible ambiguities. We conclude that narrative-based semi-structured tagging of media can indeed, help users to convey more and better information, alleviating some of the problems found in traditional tagging.

In terms of future work, we plan to perform a new user study where we will look closely at the photos entered by the users and study how particular tagging problems such as synonymy are affected by narrative tagging. We will also conduct tests using pre-defined sets of photos, to see to what extent people can describe photos that are not their own and about which they have little or no contextual information to relate to. Finally, we'll measure how narrative tagging impacts photo retrieval

## ACKNOWLEDGMENTS

Tiago Guerreiro was supported by the Portuguese Foundation to a Science and a Technology grant. SFRH/BD/28110/2006.

## REFERENCES

- [1] Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. The vocabulary problem in human-system communication. *Commun. ACM* 30, 11 (1987), 964-971.
- [2] Golder, S. and Huberman, B. (2006). "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science*, 32(2). 198-208. Available at <http://www.hpl.hp.com/research/idl/papers/tags/>
- [3] Gonçalves, D., Jorge, J. Evaluating Stories in Narrative-Based Interfaces, In Proc. Int. Conference on Intelligent User Interfaces (IUI'2006), In Proceedings International Conference on Intelligent User Interfaces (IUI'2006), ACM Press, 29 Jan - 1 Feb 2006, Sydney, Australia.
- [4] Gonçalves, D., Jorge, J. "Tell me a Story": Issues on the Design of Document Retrieval Systems. In Proceedings DSV-IS'04, Lecture Notes on Computer Science, Springer-Verlag, July 2004, Hamburg, Germany.
- [5] Halpin, H., Robu, V, Shepherd, H., The Complex Dynamics of Collaborative Tagging. WWW 2007, May 8-12, 2007, Banff, Alberta, Canada. ACM 978-1-59593-654-7/07/0005.
- [6] Hayman, S., Folksonomies and Tagging: New developments in social bookmarking. Ark Group Conference: Developing and Improving Classification Schemes, 27-29 June, 2007, Rydges World Square, Sydney
- [7] Mathes, A. Folksonomies – Cooperative Classification and Communication Through Shared Metadata. UIC Technical Report, 2004.
- [8] Marlow, C., Naaman, M., Boyd, D., Davis, M., HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read. August 2006 HYPertext '06: Proceedings of the seventeenth conference on Hypertext and hypermedia
- [9] Rabiner, L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, vol 77, no 2, February 1989