Bridging the Motivation Gap for Individual Annotators: What Can We Learn From Photo Annotation Systems?

Tabin Hasan[1] and Anthony Jameson[2] ⋆

[1] University of Trento
Trento, Italy
[2] Fondazione Bruno Kessler
Trento, Italy

Abstract. The importance of incentives and socially based motivation for metadata generation should not distract attention entirely from the need to design tools for metadata generation that use every means available to maximize the efficiency and intrinsic motivation of the individual annotator. The popular application domain of (individual) photo management has recently given rise to a number of strategies and methods that can serve as a source of inspiration for the design of metadata generation support for the semantic web. This position paper offers a brief synthesis of relevant work that is intended to serve as a basis for the representation of this perspective at the Insemtive 2008 workshop.

1 Why Photo Annotation Is a Relevant and Instructive Scenario

The problem of motivating contributions to a community-supported resource (of which the semantic web can be seen as an especially ambitious example) is often framed in terms of a contrast between the interests of an individual contributor and the interests of the group as a whole (see, e.g., [1]): If only people were as motivated to contribute to the semantic web as they are to their own personal knowledge bases, it would seem, the creation of metadata for the semantic web would thrive.

While this perspective is valid and important, we would like to call attention to the fact that there can also be a major "motivation gap" when individuals are making similar contributions for their own benefit. Consequently, we also need to examine ways of closing the motivation gap that arise even when individuals are working for their own benefit. These methods can in turn also benefit the community-supported semantic web indirectly.

More concretely, consider the familiar problem of adding metadata to photos: Since photos form a natural part of the semantic web as well as of many Web 2.0 systems, improving people's motivation to add metadata to photos would constitute a contribution to the goals of this workshop. But even when an individual is managing their own personal photo collection, there is a challenging motivation gap: Having good metadata would make it much easier for the user to accomplish common tasks such as searching

⋆ The research described in this paper is being conducted in the context of the KnowDive project (http://disi.unitn.it/∼knowdive). The participation of the second author was supported by Vulcan, Inc.

for photos that fit a particular description; but as has often been noted (see, e.g., [2]), few users get very far in adding such metadata, largely because of the time-consuming and tedious nature of the work that is involved.

Because of the rapidly growing popularity and practical importance of digital personal photo collections, a good deal of research has been devoted in recent years to the problem of motivating and/or supporting untrained end users in adding metadata to their photos. Despite—or indeed because of—the differences between this scenario and the more general scenario of adding metadata for the semantic web, it is worthwhile to look closely at the successes that have been achieved in this area and to consider how they might be generalized.

2 Overview of Determinants of Successful and Motivating Photo Annotation

Types of metadata that users often want to add to photos include (a) persons, objects, locations, and events depicted in the photos; and (b) information about the context in which the photo was taken (e.g., "just before sundown" or "just after the end of the championship football game"). It is often assumed that the photos already have accurate time and location stamps that can serve as input to automatic processing (though in fact such automatically generated metadata may be missing or incorrect for various reasons and may therefore need to be supplied by the user—a problem to which some of the metadata creation approaches discussed below can be applied).

Figure 1 summarizes a number of the ideas that have emerged from recent work on interfaces that help users to add such metadata. Before discussing these points individually and illustrating them with reference to recent research, we will comment on them briefly.

In terms of motivation, the overall approach taken in photo annotation systems for individuals is not based on external incentives or social mechanisms but rather on the provision of an intrinsically motivating experience for the individual user. Somewhat more concretely, the strategy is to optimize the relationship between (a) the cost to the user in terms of work done (in particular, tedious work) and; (b) the benefits in terms of enjoyable experiences, successful task performance, and visible improvements to the collection of items.

In some ways, the most straightforward approach is to exploit *external resources* (see the bottom left-hand corner of the figure) that can straightforwardly generate new metadata on the basis of existing metadata (e.g., supplying the name of a town on the basis of GPS coordinates). But external resources may also serve as input to sophisticated *algorithms* that analyze the content of items, either suggesting metadata or at least grouping together items that appear (to the system) to belong in the same category. Since such algorithms do not in general perform perfectly, there is generally a *user interface* that is designed to enable the user to supply the necessary manual input with minimal effort and maximal enjoyment. The *user input* itself can be seen as a valuable resource, which includes both explicit *annotation* actions and *naturally occurring* actions that provide useful information although the user does not perform them specifically for the purpose of adding metadata.

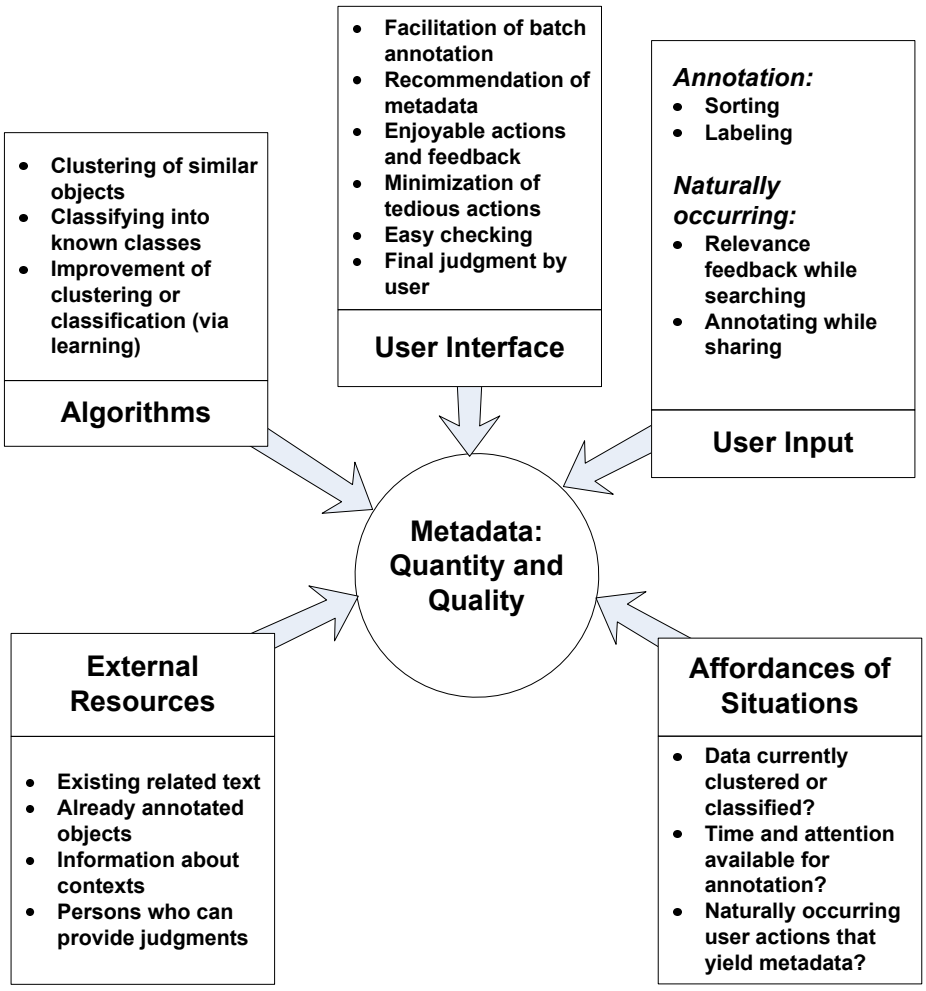

Fig. 1. Overview of factors that can contribute to the quality and quantity of metadata added in a sophisticated system for the individual annotation of resources such as photos.

Finally, some systems take into account and exploit the *affordances of situations*, taking into account the fact that people use their photo management systems in a variety of situations, each of which offers certain possibilities and limitations in terms of metadata generation.

As we will see in the next sections, these five contributors to metadata generation do not contribute independently in an additive manner. Often, a favorable combination of two or three contributors is required to achieve good results. For example, a classification algorithm may work well only on the basis of information in an external database; and it's output may be manageable only with a cleverly designed user interface that

elicits the necessary user input with minimal effort in an especially favorable situation. One objective of this position paper is to encourage this holistic view of the various contributing factors, whereas most of the primary research literature understandably focuses on one or two factors.

We will now briefly discuss some representative examples of systems that illustrate the contributing factors shown in Figure 1.

3 External Resources

Naaman et al. ([3]) provided a relatively early demonstration of how a variety of types of contextual metadata can be added to geo-referenced digital photos with the use of off-the-shelf and web-based data sources. The types of metadata added included the local daylight status and the local weather conditions. In addition to showing the feasibility of automatically adding contextual metadata, the authors showed how such metadata can be useful for searching and browsing, despite the fact that they may seem at first glance not to be especially important. For example, when searching for a given photo people may have a hard time characterizing the content of the photo itself yet find it easy to characterize the weather and daylight status—which may together narrow down the search space dramatically. A lesson for semantic web metadata creation is that the intrinsic importance of the metadata should not be the only criterion for deciding whether they are worth adding.

Another well-known system that uses this approach is PHOTOCOPAIN ([4]). This system also illustrates how an external resource can be used to support a sophisticated algorithm: Tagged photos on flickr.com serve as training data for the system's image analysis algorithms.

4 Algorithms and User Interfaces

A compelling example system in which algorithms play a central role is SAPHARI ([5]). One of the algorithms uses the clothes worn by people in photos for the heuristic clustering of photos that presumably depict the same person. This approach is an example of the clever exploitation of the strengths of the computer and the human, respectively: The computer does the tedious work of putting into a single place all of the photos that show a person wearing a particular set of clothes; all that remains for the user is to check whether these photos do in fact depict the same person and to supply the identity of that person. Note that the output of the algorithm would be useless if it were not combined with a suitable user interface.

Automatic photo clustering is also done in the EASYALBUM system ([6]), here on the basis of the similarity of faces or scenes. The results of the clustering are exploited in subtle ways throughout the interface—for example, in order to minimize the amount of scrolling that is required.

Some systems that provide clustering or classification algorithms also provide machine learning mechanisms that boost the performance of the algorithms over time for a particular user or collection. For example, whenever EASYALBUM (mentioned above)

receives new user input indicating the correct annotation of a given photo, the performance of the clustering algorithm is adapted accordingly. An approach that is apparently still new with regard to photo annotation systems for end-users is *active learning* ([7]; [8]): The system attempts to minimize the amount of input required of the user by determining at each point in time which additional training examples would be most helpful.

5 User Input

We have already seen several strategies for minimizing the number of explicit annotation actions required of the user by allowing the system to make maximal use of each such action. A different approach to optimizing the use of the user input is to interpret actions that involve no (or minimal) additional effort on the part of the user beyond the effort that they would normally exert in performing non-annotation tasks with their photo management system.

For example, in MIALBUM ([9]), a search algorithm for photos is made available that includes the opportunity for the user to supply relevance feedback by explicitly indicating which of the photos returned for a given query in fact satisfy the query. This relevance feedback is then used as input for enhancing the metadata associated with the photos in question. Given that relevance feedback is in principle worthwhile even just in terms of improving the results of the current search, its exploitation for metadata enhancement can be seen as not requiring additional user effort.[3]

Other types of natural user action that can be exploited include actions that occur when the user communicates with other persons about the photos in the collection—for example, when sending photos to another person ([10]) or or when discussing photos with other persons face-to-face (see, e.g., [11]).

6 The Affordances of Situations

The examples just mentioned illustrate the more general points that (a) photo annotation systems are used in a variety of settings and (b) each such setting typically offers some particularly good opportunities for metadata generation (as well as being limited with respect to other types of metadata generation). It therefore makes sense to design an annotation system so that it can exploit the specific potential (or *affordances*, to use the term from the HCI literature) of each situation. To take a simple example: When a user is uploading photos from their camera's memory chip, there is a good chance that many or all of the photos concern a single event (e.g., a wedding or a vacation). Moreover, at this point in time the user is likely to have a relatively precise recollection of the relevant facts. This is therefore an especially favorable time to encourage the user to make bulk annotations: Once these photos have flowed into the ocean of already stored photos and the relevant events have faded in the user's mind, adding the same metadata would present more of a challenge for both the system and the user.

[3] The authors point out that, in reality, getting users to supply relevant feedback is still a partly unsolved interface design problem, despite the immediate utility of such feedback.

7 Concluding Remarks

If you want to motivate a person to mow their lawn every week, you can offer some material incentive or set up a social mechanism by which they earn approval if they mow their lawn and perhaps disapproval if they fail to do so. A different approach is to take away their clumsy mechanical lawn mower and give them a well-designed and -engineered electric mower that makes it fun and intrinsically rewarding to mow the lawn in just a few minutes.

Strategies of the first type will presumably attract the most attention in the Insemtive 2008 workshop, and they certainly are important for the semantic web. Our position is that such approaches work best when combined with approaches of the second type; and that many generalizable ideas along these lines have recently emerged that have not yet made it into the mainstream literature on metadata generation for the semantic web.

References

1. McDowell, L., Etzioni, O., Gribble, S.D., Halevy, A., Levy, H., Pentney, W., Verma, D., Vlasseva, S.: Mangrove: Enticing ordinary people onto the semantic web via instant gratification. In: Proceedings of ISWC 2003, Sanibel Island, Florida (2003) 754–770
2. Rodden, K., Wood, K.R.: How do people manage their digital photographs? In Terveen, L., Wixon, D., Comstock, E., Sasse, A., eds.: Human Factors in Computing Systems: CHI 2003 Conference Proceedings. ACM, New York (2003) 409–416
3. Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H., Paepcke, A.: Context data in geo-referenced digital photo collections. In: Proceedings of the Twelfth International Conference on Multimedia, New York (2004) 196–203
4. Tuffield, M.M., Harris, S., Dupplaw, D., Chakravarthy, A., Brewster, C., Gibbins, N., O'Hara, K., Ciravegna, F., Sleeman, D., Shadbolt, N., Wilks, Y.: Image annotation with Photocopain. In: Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia, held at the World Wide Web Conference. (2006)
5. Suh, B., Bederson, B.B.: Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. Interacting with Computers **19** (2007) 524–544
6. Cui, J., Wen, F., Xiao, R., Tian, Y., Tang, X.: EasyAlbum: An interactive photo annotation system based on face clustering and re-ranking. In Begole, B., Payne, S., Churchill, E., Amant, R.S., Gilmore, D., Rosson, M.B., eds.: Human Factors in Computing Systems: CHI 2007 Conference Proceedings. ACM, New York (2007) 367–376
7. Zhang, C., Chen, T.: An active learning framework for content-based information retrieval. IEEE Transactions on Multimedia **4**(2) (2002) 260–268
8. Cord, M., Gosselin, P.H., Philipp-Foliguet, S.: Stochastic exploration and active learning for image retrieval. Image and Vision Computing **25** (2007) 14–23
9. Wenyin, L., Dumais, S., Sun, Y., Zhang, H., Czerwinski, M., Field, B.: Semi-automatic image annotation. In: Proceedings of Interact 2001, Eighth IFIP TC.13 Conference on Human Computer Interaction. (2001)
10. Lieberman, H., Rosenzweig, E., Singh, P.: Aria: An agent for annotating and retrieving images. IEEE Computer (2001) 57–61
11. Barthelmess, P., Kaiser, E., McGee, D.: Toward content-aware multimodal tagging of personal photo collections. In: Proceedings of the Ninth International Conference on Multimodal Interfaces. (2007) 122–125