Analyzing the Effect of Tagging by Combining Automatic and Manual Annotation Processes

Rodrigo Laiola Guimarães
CWI: Centrum Wiskunde & Informatica
Science Park 123
1098 XG Amsterdam, The Netherlands
+31 20 592 93 33

rlaiola@cwi.nl

ABSTRACT

A key research question in multimedia is how to provide mechanisms for adding semantic annotations to given media assets. Content analysis focuses on automatic tools, social networks can help in describing the media, and professionals might provide high-quality annotations. In general, tagging is a very critical process, and all these solutions have benefits and drawbacks based on parameters such as required effort, quality, validity, and usefulness of the annotations.

In the literature, the tagging process is normally situated in one of the following extremes:

- It is carried out with considerable effort, manually, by people dedicated to this job (e.g. museum catalogues); on the one hand this approach allows the development of richer services on top of such repositories, however, it has the drawback of requiring significant effort;
- It is carried out automatically, but normally with very simplistic and limited sets of tags (ontology), thus having limited value in providing for richer services;
- It is assumed a resolved issue in the presentation of services that are based on comprehensive (rich) tagging; and
- More recently, it is opened to the "general public", and is based on "folksonomies" – i.e. not relying on a controlled vocabulary; the success of this approach in different domains is still to be evaluated.

While less automatic approaches have tackled different aspects of tagging on the multimedia workflow [1], a number of research efforts also have addressed the potential of automatic processes in content analysis based on events, semantic understanding from aggregated end-user comments, media composition based on narrative constructs, and content adaptation depending on the end-user context.

In my work I discuss the effect of tagging in the context of a pan-European project called Together Anywhere, Together Anytime (TA2 - http://www.ta2-project.eu). This project studies new forms of computer-mediated social communications. In particular, I concentrate on how tagging affects automatic and manual authoring processes in an asynchronous Web-based social communication system called MyVideos. MyVideos is a community-based video sharing environment in which users can combine video assets contributed by other members to create and share personalized videos within a restrict social group.

With regards to the manual (or directed) authoring approach, tagging is essential in finding the relevant footage for a particular edit. End-users have particular editing aims and tagging might help them find, as quickly as possible, footage that is relevant to meet those aims. Therefore, we need to understand the editing needs of an end-user with regards to both the overall end product (the video compilation) and the current task at hand (the propose now).

If the editing needs are very limited, it is possible that simple ontologies and not too robust annotations could be sufficient. If the editing needs are ambitious, then more complex ontologies and more precise annotations might be required. I regard as essential here the issue of editing intent. This issue is further nuanced by the size of the repository. For example, if the number of media objects is not too high, then simple annotations might suffice even for ambitious editing aims.

With regards to automatic authoring, tagging is crucial in the automatic construction of the video compilation. Without appropriate tagging, certain compilations could simply not be achievable. Nevertheless, in the automatic editing process, a professional editor is assumed (to construct the narrative structures and provide some of the required annotations). Then, any investigation of the annotation process uses to consider that, at one end, annotation could be entirely carried out by the professional editor, whereas at the other end, it could be entirely carried out by the end-users themselves.

Here, too, tagging depends on the narrative intent. Complex compilations, more certainly, require more comprehensive tagging. An obvious issue for investigation regards the tradeoff between the efforts required in providing rich annotations and the satisfaction of the final compilation. This requires the development of a number of narrative structures, with different levels of ambition, and their evaluation with repositories of different complexities of annotations.

In my work I take a hybrid approach and try to define the thresholds and boundaries of what we can expect from automatic annotation processes, and how manual approaches can be used in annotating community-contributed media assets within MyVideos.

REFERENCES

[1] Marlow, C., Naaman, M., Boyd, D., and Davis, M. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In Proceedings of the conference on Hypertext and hypermedia (Odense, Denmark, August 22 - 25, 2006). HT '06. ACM Press, New York, NY, 526-531. DOI= http://doi.acm.org/10.1145/1149941.1149949