Automatic and Manual Processes in End-User Multimedia Authoring Tools: Where is the Balance?

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

ABSTRACT
This thesis aims to analyze, model, and develop a framework for next-generation multimedia authoring tools targeted to end-users. In particular, I concentrate on the combination of automatic and manual processes for the realization of such framework. My contributions are realized in the context of a pan-European project called Together Anywhere, Together Anytime (TA2). More specifically in a community-sharing environment in which users can combine video assets contributed by other community members to form personalized mini-stories that can be shared within their (probably restricted) social groups. The expected outcome of my thesis work is contributing for the design of authoring and sharing tools that better fit end-users' needs.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Audio, Video*. H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *User-centered design*.

General Terms
Design, Experimentation, Human Factors.

Keywords
Automatic processes, Multimedia authoring tools.

1. INTRODUCTION
The success of rich media and social networking on the Web has fundamentally changed the media landscape. Now, end-users play an active role in the media creation chain, imposing a set of new requirements for multimedia authoring tools. Traditional solutions [2] need to be revisited paying special attention to the authoring workflow, to media description, to multimedia composition, and to the social factors behind multimedia authoring. My doctoral thesis aims at the development of a framework that comprises the underlying technological infrastructure for realizing multimedia authoring tools that fit the current media landscape. In particular, I am interested in evaluating and comparing automatic and manual processes on the authoring workflow.

Automatic processes aim at minimizing the burden of the multimedia authoring processes. Media content analysis based on events [5], semantic understanding from aggregated end-user comments [7], and media composition based on narrative constructs [8] are just some examples. My thesis acknowledges all these previous efforts, and further evaluates how they affect the multimedia authoring process. With special focus on sociological theories and human factors, I compare them to less automatic approaches where the end-user can influence the end result. My expectation is that a balanced combination of manual and automatic processes will be the basis of the so needed multimedia authoring tools for the *prosumers* of tomorrow.

The main focus of my research is on empowering users for creating and commenting on personalized multimedia presentations. It is important to clarify that I am not talking about personalization in the sense of recommending content a user might like, but that of sharing an experience with strong ties. In particular, the research questions addressed in this thesis include:

i. *Media description:* a key research question in multimedia is to provide automatic mechanisms for adding semantic annotations to given media assets. Content analysis focuses on automatic tools, social networks can help in describing media, and professionals might provide high-quality annotations. All these solutions have benefits and drawbacks based on parameters such as required effort, quality and validity of the annotations, and usefulness of the annotations. In my thesis 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 media assets;

ii. *Workflow:* because of the dynamic nature of the new media landscape, an essential topic for further investigation is the multimedia creation workflow. Theoretical proposals [4] define a number of canonical processes and interfaces. More human-centered studies [6] identify new roles of end-users in the creation process. While concentrating in the media creation process, they tend to neglect that user comments and reactions are as well part of the authoring workflow [1]. My intention is to update current proposals to the new media landscape;

iii. *Multimedia composition:* there are some classes of authoring tools that generate multimedia presentations as a self-contained media file (early binding). Because media resources might be distributed - and very large in size - other tools produce multimedia presentations as reference documents (late binding). In both approaches automatic processes play a major role; and

iv. *Togetherness:* sociologists have provided a number of theories for better understanding of human relationships. It is my intention to pay special attention to such theories and to use them as an input for my thesis – in the form of

MM'10, October 25–29, 2010, Firenze, Italy.
ACM 978-1-60558-933-6/10/10.

requirements. For example, the tie strength [3] imposes a number of requirements normally forgotten by technical research. An important challenge is to evaluate the impact of social theories on the design of tools for automating the production of tailored presentations.

2. METHODOLOGY

In order to measure and validate the contributions of my thesis I intend to use different methodologies: literature review, prototype implementation, laboratory user studies, field trials, and focus groups/interviews.

Evaluating and defining the boundaries and thresholds between automatic content analysis and laborious manual approaches require asking users. From field trials I expect to find out how much users are willing to contribute with the tagging process depending on the task at hand. Moreover, I intend to explore how different levels of metadata – in terms of quality and correctness – affect the workflow. In particular, the intention is to perform comparisons by using a number of prototype implementations, some of them with advanced content analysis algorithms and some of them with primitive ones.

To validate the multimedia workflow and multimedia composition, I investigated existing authoring services intended for end-users and I reviewed and analyzed related literature. This groundwork resulted in the development of a number of ad hoc prototypes. In particular, I developed a captioning system for videos on the Web and a pan-and-zoom utility that helped users to tell stories around their photos. The implementation effort combined with the literature review provided me a deep understanding on authoring models and on the state-of-the-art. In order to fully validate the resulting workflow potential end-users and professional video creators will evaluate it.

The sociability aspects need a more human-centered approach. In this study I intend to capture users expectations, learning experience and evaluation via focus groups, questionnaires and individual interviews. Moreover, I already took advantage of existing literature on social science in order to better define the requirements for the prototype work (e.g., tie strength).

My thesis is funded by the pan-European project TA2 (http://www.ta2-project.eu). The project studies new forms of computer-mediated social communications. In TA2, my contribution is in an asynchronous social communication system called MyVideos. MyVideos is a community-based video sharing environment that supports end-user creation of highly personalized music videos from a collection of independent user-made recordings. It allows as well for asynchronous sharing among heterogeneous participants within a restricted social community. My participation in this project provides me the funding for completing the thesis. But more importantly, it allows me to work in an international environment, where both the theoretical and business aspects are taken into consideration. Finally, it gives me the opportunity to work with researchers from different fields, including social scientists. They will take care of the user-centered methodologies described above.

3. CURRENT STATUS AND NEXT STEPS

In the beginning of 2009, we started a long field trial within the TA2 project, and it will last until the summer of 2011. The first goal was to identify requirements and to better understand the user needs. The final intention is to involve potential users in all the stages of the PhD process: requirements gathering, prototype implementation, and user testing. During the second year we conducted a number of interviews across Europe – in three different countries [8]. In addition, two focus groups took place in the UK, summer of 2009, and in the Netherlands, December 2009. The results of the interviews and the focus groups lead to a first version of the new multimedia lifecycle. The lifecycle is not definitive, but I have the intention of further iterate it based on the results from the field trial.

I have been actively implementing a first prototype that fits the identified requirements and the multimedia lifecycle. The implementation provides automatic processes such as content analysis and identification, automatic recommendation of content for its inclusion in an authored video, authoring templates for lowering the authoring burden, clip boundary suggestion, and automatic presentation generation based on narrative aspects. Using my tool, the user can navigate, organize, and annotate – or improve automatic annotations of - media assets. The user also can create his own version of the video helped by automatic processes and can contribute with more assets if needed.

The next steps of my work consist on a new round of trials in July 2010 in Amsterdam. The intention is to collaborate with human-centered groups – TNO (in the Netherlands) and Interactive Institute (in Sweden). Potential users will use my system for the creation of multimedia presentations based on a school concert that took place in April 2010, where several parents were given video cameras. The intention is to validate and to revise the prototype and the first lifecycle proposal. The expectation is that parents will provide me with enough data for further assessing the benefits and drawbacks of automatic processes, and how to combine them with more manual approaches. Another round of end-user evaluations are scheduled for 2011.

4. ACKNOWLEDGMENTS

This work was supported by EU FP7-ICT project TA2. My thanks to my advisors Dick Bulterman and Pablo Cesar.

5. REFERENCES

[1] Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J. and Ross, K. 2009. Video interactions in online video social networks. In ACM TOMCCAP, 5, 4, 30.

[2] Bulterman, D. C. A. and Hardman, L. 2005. Structured multimedia authoring. In ACM TOMCCAP, 1, 1, 89-109.

[3] Gilbert, E. and Karahalios, K. 2009. Predicting tie strength with social media. In Proceedings of CHI, 211-220.

[4] Hardman, L. 2005. Canonical processes of media production. In Proceedings of the ACM workshop on Multimedia for human communication: from capture to convey, 1-6.

[5] Kennedy, L. and Naaman, M. 2009. Less talk, more rock: automated organization of community-contributed collections of concert videos. In Proceedings of WWW, 311-320.

[6] Kirk, D., Sellen, A., Harper, R. and Wood, K. 2007. Understanding videowork. In Proceedings of CHI, 61-70.

[7] Shamma, D. A., Kennedy, L. and Churchill, E. F. 2009. Tweet the Debates: Understanding Community Annotation of Uncollected Sources. In Proceedings of WSM, 3-10.

[8] Williams, D., Ursu, M. F., Cesar, P., Bergström, K., Kegel, I. and Meenowa, J. 2009. An emergent role for TV in social communication. In Proceedings of EuroITV, 19-28.