

# Constraints and simplification for a better mobile video annotation and content customization process

Robert Knauf  
Chemnitz University of Technology  
Strasse der Nationen 62  
09111 Chemnitz – Germany  
+49 371 531-35793

robert.knauf  
@cs.tu-chemnitz.de

Arne Berger  
Chemnitz University of Technology  
Strasse der Nationen 62  
09111 Chemnitz – Germany  
+49 371 531-36872

arne.berger  
@cs.tu-chemnitz.de

Maximilian Eibl  
Chemnitz University of Technology  
Strasse der Nationen 62  
09111 Chemnitz – Germany  
+49 371 531-31562

maximilian.eibl  
@cs.tu-chemnitz.de

## ABSTRACT

Users have limited will and creativity for describing, tagging, and rating audiovisual content, especially when they are consuming media in a mobile setting. The approach of our research is to reduce the burden of intellectually annotating user generated videos and to simplify content rating. Thus, more relevant results and more accuracy of fit are to be gained when potential consumers are in search of quick entertainment. We propose a video upload tool for smartphones which allows users to share captured video clips with buddies and the community. The distinctive feature is an elementary, lightweight but expressive ontology that speeds up and simplifies content tagging for the producer on the one hand, and allows very quick and well-matched retrieval of potentially interesting media items on the other hand. Furthermore, an unobtrusive but required rating request should lead to more precise estimations regarding the relevance of content.

## Categories and Subject Descriptors

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *Evaluation/methodology*; H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology*, *Graphical user interfaces (GUI)*, *User-centered design*.

## General Terms

Design, Experimentation, Human Factors

## Keywords

mobile video, mobile television, video annotation, video retrieval, video search, rating, GUI

## 1. INTRODUCTION

The way of selecting and consuming audio-visual content in mobile settings varies widely from “ordinary” domestic usage. Most often the users are in a hurry and have only limited time for enjoying entertainment on their mobile smartphones while on the move. Whether they are commuting or going out with friends, users usually avoid constructing complex search queries to find suitable content to watch. In the same way, they often omit rating media items after they are consumed. Besides time constraints, factors such as a lack of concentration, obtrusive user interfaces,

and constrictions from inter-personal circumstances (e.g. commuting in a crowded train or quickly retrieving a funny video clip to entertain a clique), may prevent a search engine from gaining insight into the entertainment needs of the users.

An even greater amount of creativity and potentially disruptive interaction is required when describing content one has just captured in a rich and sufficient way for others to retrieve it. Superficially speaking, entering tags or keywords grants freedom of description. Despite having that freedom, people often describe being unable to think of any tags [8]. Nevertheless [5] wrote that user-based tags are usually derived from a connection between concept and content in the mind of the author. They assume that tagging supports solely the supplier of the metadata or just a few users because of ambiguities or even codes in closed user groups. As a result, a loss of meaning may occur beyond the context of a single user. Neither the community nor the recommendation systems take advantage of uncontrolled and chaotic tagging behavior when they look for relevant content afterwards. This is aggravated by the fact that a huge number of terms entered in tagging applications are misspelled, mistakenly encoded, or given in a plural or otherwise altered form or as compound words. [5] point out that a serious weak point in tagging audiovisual content is that most users do not give much thought to the way they tag resources. Changing tagging habits turns out to be a difficult task, because people tend to repeat behavior they have acquired in the past [8]. As an exception, observing community tags has an influence on an individual’s tagging style.

In addition, an oversaturated vocabulary is deemed less efficient for retrieving adequate content from a huge amount of information, as observed for the social bookmark tagging service, *del.icio.us*<sup>1</sup> in [4]. With an increasing level of tag saturation, users must either find new words or use multiple words to describe introduced content. But [5] illustrate that most tags are generally used by one, two, or small groups of users. The number of tags that are used by the whole community is relatively small. Over time, increasing numbers of available tags diminish the completeness of search results.

Current research mainly focuses on automatic video recommendation systems to find relevant video clips according to the users viewing preferences [7] based on the recurring

---

<sup>1</sup> “*Delicious – Social Bookmarking*”: <http://del.icio.us/>

assumption that existing tags in video repositories are quite rare [2]. Users' conscious decisions are often undermined by recommender algorithms that try to be context-sensitive, as described in [3]. Video sites like *YouTube*<sup>2</sup> try to recommend relevant video clips as soon as users finish watching a clip. The system does not communicate the basis upon which a recommendation has been made to the users. Consequently, the algorithm might carry the user farther away from his entertainment source.

With regard to this question, we proceed from the assumption that finding interesting media items quickly and matching the users' entertainment requirements at just the right moment is difficult for both the content provider and for the requesting consumers. The negative effect of the lack of both descriptive metadata and broad user feedback is difficulty in retrieving suitable content inside a video portal with as few clicks as possible. Most often, the seeking consumers are left having to refine their search query several times or to use non-customized item lists such as "most viewed". Most commonly, retrieving media items referring to a local geographic area, a special interest, or a social group can only be achieved by following the "channels" of certain users or user groups. Social media sites such as *Flickr*<sup>3</sup> enable users to keep track of their contacts' content both by harvesting activities of interconnected users and by providing interfaces that create personalized "Explore pages," offering interesting content produced by the users' contacts [6]. Feasible solutions of this kind for overcoming information overload reduce the importance of tags as a way to share multimedia content.

## 2. ANNOTATING AND RATING QUICKLY BUT RESTRICTIVELY

To improve the lopsided ratio of content and assigning metadata, or tags, we propose a tool for mobile phones, such as smartphones, with features as follows:

The basic functionality includes connecting to an online video platform. The participants create their own audiovisual content using the built-in cameras on their mobile devices for the purpose of sharing impressions with buddies or with the whole community. Suitable platforms include *Qik*<sup>4</sup> or *YouTube* to name but a few.

In addition, we focus on enriching the repository of content with descriptive metadata such as tags, categories or feedback information. Participants are to be increasingly required to make their own content more relevant in associated search or selection procedures. Concurrently, we want to increase tagging activity by providing methods of quick and easy annotation at the point of capturing content, as previously suggested by [1].

### 2.1 Mandatory but quick and easy tagging

Our proposed concept uses a balance between constraints on the one hand, and unobtrusiveness on the other: The video upload procedure permits transferring data to the server only if proper

metadata has been provided by the user. So, capturing the video clip and entering appropriate tags should not take more than three steps of screen interaction with the purpose of limiting usage interruption. The strict limitation of user interaction is achieved as follows.

#### 2.1.1 Limiting descriptions to cardinal questions

As suggested by [8], in persuading a community, it is helpful to specify only certain types of tags that benefit the system. Their results suggest that users would tend to follow a "pre-seeded" tag ontology.

For generating narrowed but important content descriptions, only metadata is considered that answers the questions *Who?*, *What?*, *How?*, *Where?* and *When?*

For the questions *Who?*, *Where?* and *When?* it is usually possible to retrieve the answers automatically. Users identify themselves to the video portal by providing a username and a proper (often locally stored) password. If personal identification is not wanted, one might optionally publish content anonymously or as a member of a user group. Furthermore, a mobile device with positioning capabilities (e.g. GPS, Wi-Fi triangulation) provides both location and time.

Only the questions *What?* and *How?* remain to be answered manually. For answering these questions we propose to answer *What?* with a descriptive noun, representing a category, and *How?* with a corresponding adjective. Figure 1 shows the dummy application we used for user interviews.

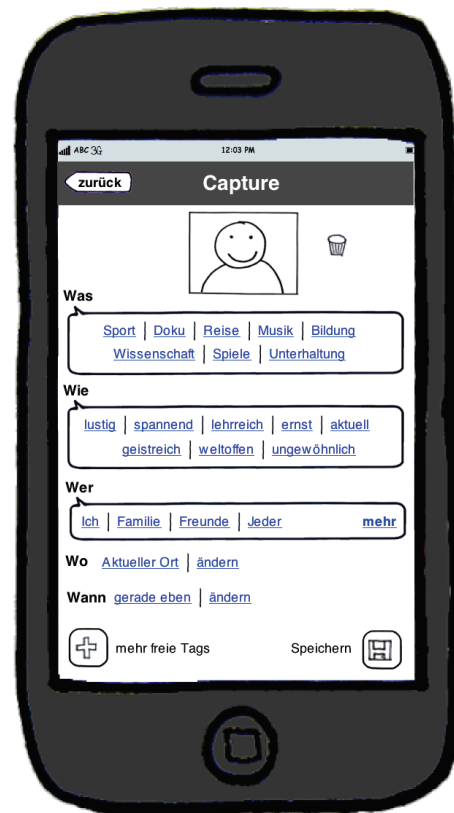


Figure 1. Dummy prototype for entering tags after video capture

<sup>2</sup> "YouTube – Broadcast yourself": <http://www.youtube.com/>

<sup>3</sup> "Flickr – Photo Sharing": <http://www.flickr.com/>

<sup>4</sup> "Qik – Record and share video live from your mobile phone": <http://qik.com/>

### 2.1.2 Limiting keyword vocabulary

As previously described, one can anticipate limitations in users' will and creativity to provide descriptive keywords when uploading content. Adverse factors such as distraction, shortage of time or unwillingness often keep them from submitting specific tags.

For these reasons, manually entering tags has been omitted, and instead, users choose appropriate keywords from a closed and narrow vocabulary. Since uploading a video clip and providing metadata should take place with no more than three steps requiring user interaction, as previously indicated, only one keyword per remaining cardinal question (*What?* and *How?*) is to be selected.

## 2.2 Rating before carrying on

Our second proposition for gaining relevant user feedback in an only slightly annoying manner is to reuse the principle of constraints and simplification. After watching a video clip, a consumer may only proceed to the next item after submitting a rating. It should only require one click to accomplish this. Additionally, a choice of "I don't care/know" is to be kept available.

## 3. REQUESTING INTERESTING CONTENT

The same policies proposed for upload functionality in 2.1.1 and 2.1.2 are to be applied to the search process as well: One can shape a simple search request with just a few steps of interaction with the graphical user interface of the software. The request is to be constructed from the terms that have already been predefined for video annotation (see 2.1.2). For this purpose, the cardinal questions described above (see 2.1.1) provide an elementary structure for creating requests to find interesting media items.

Through this process, the need for entertainment is channeled into a strongly structured query with the objective of delivering an individualized media "channel" to the user that contains all items matching the user's presumed interest. Moreover, users will be able to customize the results by adding restrictions or defining sort criteria such as geographic location, topicality or content provided by certain community members.

Both search queries and result order are to be generated by using the same metadata set. The graphical user interface of our proposed tool shall provide functionality to interrelate primary search terms from one class (e.g. *What?* or *How?*) with limiting parameters derived from one or more other classes (e.g. *Who?*, *Where?*, *When?* or popularity, see Table 1 for examples). Figure 2 again shows the corresponding dummy application used for user interviews.

**Table 1. Examples of using cardinal questions to limit or sort search results**

Question	Default terms
Who?	single person, friend, family member, anybody
Where?	current location, home, known POI, map position
When?	last <i>n</i> hours, days, weeks, etc., time span

Assuming that a user's interests remain more or less stable, an interesting potential use is to follow events related to a certain person, location or time span.

## 4. RESEARCH QUESTIONS

Our approach is to be tested first by paper-prototyping and by setting up an initial set of keywords. These frontend prototypes will be tested by potential users (see Figures 1 and 2). In the subsequent focus group interviews, we want to figure out if the participants accept the idea or not. For this purpose we will draw a comparison between our concept and the ordinary "freehand tagging."

We anticipate acceptance if the advantages of the simplification methods prevail over negative impressions resulting from the application of constraints.

With the knowledge gained from the interviews we will implement the tool on a mobile phone software platform coupled with a portal server system. In subsequent user studies we want to answer the following questions: Is the association between given keywords and actual content correct? Is the set of abstract keywords sufficient? Is the concept of constrained rating providing better and more realistic scores or is it necessary to add editorial scoring?

Finally we want to find out if and how our tool can be a worthwhile supplement to video portals such as *Qik* or *YouTube*.



**Figure 2. Dummy prototype for customizing content demand**

## 5. CONCLUSIONS

Our concept is intended to address the deficiency characteristic of ordinary video portals that provide large quantities of content but lack sufficient quantities of matched metadata. We want to hand over a tool to the users that encourages them to tag and rate audiovisual content by means of quick and easy questioning, in order to obtain requested entertainment media.

The users should be relieved of the burden of constructing excessively specific search terms to retrieve suitable content. The effort necessary to gain rapid access to an interesting video clip category should lie between that characteristic of established video portals (high effort, high level of customization: complex search by entering free terms) and that of a regular television program (low effort, low level of customization: instant access but no search capability).

## 6. REFERENCES

- [1] Ames, M. and Naaman, M. 2007. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 – May 03, 2007). CHI'07.  
DOI= <http://doi.acm.org/10.1145/1240624.1240772>
- [2] Baluja, S., Seth, R., Sivakumar, D., Jing Y., Yagnik, J., Kumar, S., and Ravichandran, D., Aly, M. 2008. Video suggestion and discovery for YouTube: Taking random walks through the view graph. In *Proceedings of the 17th international conference on World Wide Web* (Beijing, China, April 21 – 25, 2008). WWW'08.  
DOI= <http://doi.acm.org/10.1145/1367497.1367618>
- [3] Bellekens, P., Houben, G.-J., Aroyo, L., Schaap, K., and Kaptein, A. 2009. User model elicitation and enrichment for context-sensitive personalization in a multiplatform TV environment. In *Proceedings of the 7th European Conference on European Interactive Television Conference* (Leuven, Belgium, June 3 – 5, 2009). EuroITV'09.  
DOI= <http://doi.acm.org/10.1145/1542084.1542106>
- [4] Chi, E. H. and Mytkowicz, T. 2007. Understanding Navigability of Social Tagging Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 – May 03, 2007). CHI'07.
- [5] Guy, M. and Tonkin, E. 2006. Folksonomies: Tidying up Tags? *D-Lib Magazine*, 12(1), January 2006. Available at: <http://dx.doi.org/10.1045/january2006-guy> (accessed 15 April 2010).
- [6] Lerman, K. and Jones, L. 2007. Social Browsing on Flickr. In *Proceedings of the International Conference on Weblogs and Social Media* (Boulder, Colorado, USA, March 26-28, 2007). ICWSM'07.
- [7] Mei, T., Yang, B., Hua, X.-S., Yang, L., Yang, S.-Q., and Li, S. 2007. VideoReach: an online video recommendation system. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (Amsterdam, The Netherlands, July, 23 – 27, 2007). SIGIR'07.  
DOI= <http://doi.acm.org/10.1145/1277741.1277899>
- [8] Sen, S., Lam, S. K., Rashid, A., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. 2006. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada, November 04 - 08, 2006). CSCW'06. ACM, New York, NY, 181-190.  
DOI= <http://doi.acm.org/10.1145/1180875.1180904>