Interactive Curating of User Tags for Audiovisual Archives

Yi-ling Lin
University of Pittsburgh
Information Sciences
135 North Bellefield Avenue
Pittsburgh, PA 15260
yil54@pitt.edu

Lora Aroyo
VU University Amsterdam
Computer Science
The Netherlands
l.m.aroyo@cs.vu.nl

ABSTRACT
With the rapidly increasing popularity of social sharing sites, the traditional manual indexing techniques are no longer feasible to cope with the growing amount of multimedia content. Emerging folksonomies of user tags through crowdsourcing provide a potential for the collaborative annotation of various types of online multimedia resources. However, the shortcomings of folksonomies still present researchers with challenges to effectively use the collected user tags in professional or public collections. Examples of such challenges are determining how to tackle the quality of tags, to understand tags' meaning and relevance to the resource material, and to define quality parameters of the final (targeted) annotations of multimedia resources. This work addresses such challenges in a concrete use case – the crowdsourcing video annotation game called *Waisda?*. This game is used to collect user tags for videos from the Dutch National Audiovisual Archive 'Sound and Vision'. In this paper we explore the interactive aspects of a post-game crowdsourcing tool called *'Tag Gardening'* for curating user tags. We tackle the challenges of bringing out quality and extracting meaning from the user tags in order to finally achieve satisfactory video annotations.

Categories and Subject Descriptors
H.5.2 [Information Interface and Presentation]: User Interfaces – *graphical user interfaces (GUI), interaction styles, and user-centered design.*

General Terms
Design, Human Factors.

Keywords
Social tagging, Folksonomy, Tag gardening, Interactive Interfaces, Cultural Heritage.

1. INTRODUCTION
Currently, we observe a rapidly increasing amount of daily digital material entering audiovisual archives, as well as an enormous stream of multimedia content from the Social Web, e.g. Flickr, YouTube, or Delicious. The traditional (mostly manual) indexing techniques are no longer feasible to cope with the demands from this growing amount of multimedia content. Social tagging systems allow users to upload diverse resources with arbitrary tags (keywords), creating a so-called folksonomy. This folksonomy provides the potential for the collaborative organization of various types of online resources, such as articles, references [5], pictures [4,14], videos [8], blog posts, and discussions. Although several shortcomings of folksonomies, such as noise, shallowness, ambiguity, inconsistency, have been criticized in the literature [12,13], folksonomies are still valuable to obtain and to be utilized for different enrichment and refinement purposes [7,13].

This paper is a follow-up study of [8,11]. The crowdsourcing experiment called Waisda?(In Dutch, it means "What is that?") is a video-labeling game used to collect time-related metadata (user tags) as a result of the mutual agreement between online amateurs (gamers) while tagging audiovisual material. The Netherlands Institute for Sound and Vision (one of the European's largest audiovisual archives) aims at increasing the findability of their collection consisting of more than 700,000 videos. The Waisda?, a game-based tagging system, invites participants to annotate a video with a freely formed tags describing what they see or hear while watching videos from their archive. The participants receive points for matching a tag with one or more of their online opponents (in a window of ten seconds around each video fragment). In the period of the first Waisda? pilot from March till November 2009, there were over 400,000 tags collected through Waisda? by 2,400 users for a collection of about 650 videos [3]. During the second pilot in 2011, which only lasted 3.5 months, there were more than 2,500 videos tagged by only 530 users. The total amount of the tags was impressively high - above 300,000. The main conclusion, a comparison between the two pilots, is that there was a significant increase in the number of the collected tags, improved efficiency in terms of number of users and time, and in overall significant increase in the quality of tags in terms of their matches with other opponents (e.g. 51% matched tags compared to the 37% from the first pilot). Despite the success in terms of the efficiency of the tagging game, we are still puzzling over the problem of the actual usefulness and effectiveness of the collected tags, i.e. (1) are the collected user tags good enough for video fragment search, and (2) is their quality good enough to increase the findability of video search in the Sound and Vision collection.

In order to increase the quality of the collected tags and to improve the experience of the users and their motivation to play the game more frequently and for longer period of time, specific improvements were implemented in the UI from the first to the second pilots according to the result of usability testing and focus group feedback. Several changes such as interface simplification, information awareness enhancement, and challenge disclosure were addressed [2] . However, the main obstacle, which still

AVI'12, May 21–25, 2012, Capri Island, Italy.

remains, is that the game itself requires a lot of the players' attention on the video and there is no enough space for more sophisticated tag curating during this process.

Several theoretical approaches have been studied to improve the feasibility of user tags and enrich folksonomies with semantic meanings. For example, [4] applied clusters of semantically related tags to refine users' searching experience in a tag space. Similarly, [6,15] focused on semantic enrichment with ontologies and controlled vocabularies for structural enhancement of folksonomies. Further, [12,16] explored algorithms to create hierarchies for automatic restructuring of folksonomies to support users navigation. In spite of the fact that the automatic approaches did enhance the feasibility of user tags to certain degree, the lack of linguistic processing for tag enrichment and the challenge of differentiating various associative relations still fail to enhance the tag quality to a usable state [17].

Tag Gardening [10], a mainly manual activity supported by certain automatic processes, has been positioned within existing initiatives to manage and reengineer folksonomy tags in order to make them more productive and effective. Along the lines of this, Hildebrand and Ossenbruggen [9] developed a tag gardening post-game system by investigating a semi-automatic process to link user tags to concepts from the Linked Open Data cloud.

In this paper we focus on improving the Tag Gardening user experience in order to achieve an effective and efficient tool for crowdsourcing the process of User Tags curating. The input for the Tag Curating Tool is the collection of user tags and videos from the Waisda? game. We use the analysis of the collected tags across the two Waisda? pilots in order to understand better the challenges for the Tag Curating as well as to provide some 'intelligence' for this tool.

2. CURATING TAG CHALLENGES

In a typical video tagging game, one of the most common problems is that the lack of time for users to do other things than just watching, listening and tagging. They need to be fast and focused in order to tag the right thing at the right time. As we showed in section 1 we have experimented in the Waisda? pilot 1 with several interactions to help people tag with the tags that we wanted (e.g. more specific tags, people's names, locations' names, etc.). From the experiences in the pilot 1, we created a new interaction and UI for the pilot 2. Finally, we decided to focus on the quality of the tags and the overall annotations outside of the gaming environment. So, our "Tag Curating" framework is to create a second phase in which the users curate and improve the quality of tags and the video annotations created by the Waisda? users.

Below we outline the set of challenges from the User Tags and Video Annotation perspectives. In section 3, we show the corresponding user interaction to tackle those challenges. From the User Tags perspective, there are two main challenges we have to address:

- ***The quality of the individual tags***, e.g. correcting spelling mistakes, aligning multiple forms (verbal forms: participles, gerundives, etc. and singular verse plural forms), and abbreviations.
- ***The meaning of the individual tags***, e.g. disambiguating potential polysemy of each tag, aligning tags with synonyms, and identifying the tags with the same semantic meaning. A single tag might have various meanings (polysemy) such as 'bank' which could represent a steep natural incline, a business establishment, or a set of similar things arranged in a row. Besides, a tag might have the same or similar meaning with other words in the language called synonyms. For instance, "automobile" and "car" are the synonym words for the same concept. Moreover, users might apply semantic ambiguous tags to represent the same thing/person. For example, the 43rd president of the United States might be represented as "43rd" , "George W. Bush", "the president of the United States George W. Bush", "George Bush", "Bush".

From Video Annotation point of view, there are two more challenges we are trying to solve:

- ***The suitability of each tag as a descriptor of a video fragment***, a tag might have been assigned to a video fragment with different roles (depicted, associated, descriptive), or different types (person, location, building, etc.) e.g. how accurately this tag describes the selected video fragment? Is the relationship of the tag to the video fragment clear? Is the type of the tag explicit? Is it specific enough? These are the issues we are addressing here. Besides, since the temporal nature of videos, a tag might possibly be assigned to a video fragment later than the intended spot according to the user's typing speed. How to help users to position the tag in the intended spot is also a task.
- ***The suitability and completeness of a collection of tags as an overall video annotation***, e.g. how well a set of tags describes the entire video so that the video could appear as a result for diverse information needs (search queries)? Are the tags describing enough the visual and audio information in the video, including the main topics, and the background topics for its findability? Are the tags distributed throughout the entire video in a manner that describes the individual fragments as well as the entire video? Is there any way to make the tags distributed more evenly among video fragments? Since our video fragments were set to average 10 minutes, some of the tags might be related to consecutive fragments. Therefore, allocating those tags to consecutive fragments in a certain period of time is also required for findability improvement in the video collection.

3. INTERACTIVE TAG CURATING

Ahn [1] proposed a series of games to explore harnessing crowd power to assist in various research tasks related to cataloging audiovisual material in the heritage domain. With our experience with the studies of Waisda? [8,11], game model has been shown to be an engaging tool for collecting social tags. Therefore, our follow-up curating system would keep the same game format. The system not only possesses entertainment quality as a game but it also supports the extra ambition of conquering the challenges of tag curating. In designing the interactive tag curating system, we focus on (1) translating challenges into interactions, (2) validating the quality of crowdsourcing, and (3) employing game elements to engage users. In the following three subsections, we describe these three focuses in detail.

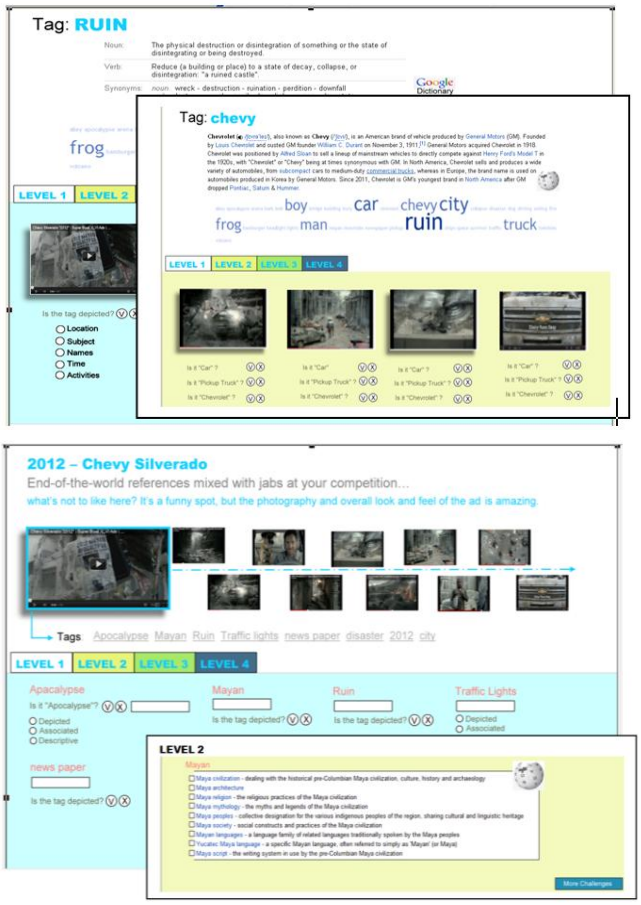

Figure 1. Interactive Curating Interface : Tag-centered view (top) , and Video-centered view (below)

3.1 Tag-centered & Video-centered Views

In accordance with the challenges for curating user tags, the whole interactive interface is designed to have a tag-centered view and a video-centered view (see Figure 1). The interface design is flexible to allow users to switch between tag and video views. The tag-centered view is designed to give users a quick overview for facilitating quality management and meaning reconciliation of tags. Four tasks are addressed in this view: (1) correcting spelling errors, (2) word form converting, (3) mapping tags to concepts, and (4) semantic recognition providing additional specificity. At the backend, we employ the approach of [9] to automatic verifying the spelling, forms, and concepts. At the front-end we provide an interactive interface that allows users to verify the automatic recognitions and judgments as well as further providing us correct and suitable tags.

In the video-centered perspective, the goal is to find the best description of a video as a whole and of video fragments based on user tags. Therefore, identifying right tags for right video fragments is critical in this view. Whether the tag is properly used in different videos/video fragments in the same or different contexts is a challenge. Whether the tag has been assigned for certain to a specific video fragment or not? Whether a set of tags sufficiently describes a video and enhances the probability of it being found by someone using common keyword search techniques? These are the essential issues with which this view is concerned. Similar to the tag-centered view, we have four targeted tasks in the video view: *(1) tag role identifying, (2) tag type identifying, (3) tag generalization-specificity identifying, and (4) tag positioning*. The following are the details of each interactive task:

Tag role identifying – In this task, we require users to identify the role of a tag from three categories: depicted, associated, and descriptive. A tag could be either expression or description related annotation. As an expression tag, it could be a central annotation related to the 'foreground' view where something has been seen or heard (depicted) in the video or it could express a non-central annotation related to the 'background' concept which users obtained (associated) from the video. As a description related (purpose-oriented) tag, the tag is related (descriptive) to the video but not actually visually or audibly presented. It could be used to describe an entire video, a series of video fragments, or a video fragment.

Tag type identifying – This task requests users to define the type of a tag. A tag could target different types of elements within the video, such as person, location, organization, object, etc. For instance, apple could be a type of organization when it has been tagged to a video about Apple Computer Company. It could also be tagged as a type of object to represent an apple in the video.

Tag generalization-specificity identifying – This task presents users possible generalization or specificity of a tag by applying existing vocabulary resources such as WordNet. Users are required to identify the most proper general or specific term of the tag by selecting system automatically generated suggestions or providing any specific input. For example, if the tag is a general one like "woman", then users might need to provide her name as input. If the tag is a specific such as "Jordan", the system will derive the general concepts such as "basketball player", "man", etc. for "Jordan".

Tag positioning – It is a task to allocate a tag to its corresponding video fragment or a series of fragments. A tag could be assigned to earlier or later fragments since players assign the tags while playing a video. A tag might be related to several fragments in a row but hard to assign to all of them due to the 10 minutes cutting points.

3.2 Validation Mechanisms

The main validation mechanism in the Tag Curating can be realized in a similar way as in the Waisda? game – by letting two or more players match their curating suggestions. In other words, the underlying assumption is that the action is valid for the purpose if there is a mutual agreement between players.

Additionally, we also design two types of progressing levels in the Tag Curating: the levels of difficulty on tag curating activities, and the levels of editing. The levels of difficulty will be explained in detail in the next section. The levels of editing are introduced in the system so that users will have different levels for the editing and approval of their proposed edits/changes. If a user has earned sufficient points of a particular task, he/she can be considered an 'expert' and can be authorized to approve others' suggested edits.

3.3 Incentive Mechanisms

Several incentives mechanisms are applied to engage and encourage users to participate in the proposed tag curating system. The following is a brief overview of our incentive mechanisms.

Incentive with flexible view of the game – Users are allowed to choose either tag-centered view or video-centered view to start the game. And it is easy to switch between views by going through

hyperlinks of tags or videos' titles. In each view, users could choose the most appropriate task to start with.

Incentive by goal setting – In order to stimulate passion for competition and challenge, we provide a game goal and the levels of difficulty in the system. A user would score more points by participating in more curating activities. With a certain score, the user could unlock the next difficulty level. They can either stay in one level or move further to explore the activities at the next level. Whenever they challenge more difficult levels, they will score more extra bonus.

Incentive with reputation – We also design a leader board to increase users' motivation level based on competition and reputation. We not only have an overall leader board to show the leading users in the whole system, but also create specific leader boards for all the individual tasks. The leading users are shown in the leader board with their identity and current scores according to their performance in the whole system or different tasks.

Incentive with rewards – Along with the leader board, we create a personal performance bar to show individual's level and scores so that users could easily aware their own accomplishment and realize how long they could reach the next level. To encourage users to try diverse videos, we would give more extra bonus to the users who curate tags in different videos than the users who only curate tags in one video. Meanwhile, we also consider users' expertise in one particular type of videos or tags. If a user performs well on curating in specific type of videos or tags, we would give him/her more points on the level of editing. While he/she reaches a certain level of editing, he/she will be authorized to become a reviewer or even an editor for the specific type of content.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we tackle the challenges of bring out quality tags and extracting the meanings for audiovisual archives. We proposed a Curating tool for crowdsourcing the process of user tags curating. In the current design stage, we focus on qualitative analysis and evaluation of this system. First an online questionnaire will be sent out to the users playing in Waisda? The questions are addressed on whether the design elements on interfaces are easy to curate tags, whether the tasks designed in two views are clear and meaningful to pursuit gradually, whether the validation mechanisms are fair, and whether the incentive mechanisms are inspiring to curate tags. Secondly, we will form a focus group with professionals from the Institution of Sound and Vision for clarifying the message delivered from the system and the meaningfulness / usefulness of the tasks of tag curating. Afterward, usability study will be conducted with the subjects who never used the system before to evaluate our system design on interface, validation mechanisms, and incentive mechanisms. For these qualitative studies, we use Cohen's kappa coefficient to evaluate the inter-rater agreement.

Quantitative evaluation will be carried on all curated tags and videos. We will not only have the descriptive statistics of the curated tags and videos, but also conduct a user study to assess the usefulness and the value of Tag Curating. Our hypotheses are that the game-based design would attract users to curate tags and build stronger relationships between tags and videos. Future work is addressed to make the connection among tag collection, tag curating, and tag retrieval to enrich the findability and usability of social tags.

5. REFERENCES

[1] Ahn, L. von. Games with a Purpose. IEEE Computer 39, 6 (2006), 92-94.

[2] Annelies van Ees. How to TAG 70,000 videos? The power of games. 2010, 66.

[3] Baltussen, L.B. Barbarians versus gatekeepers? Tagging as a way of defining the emergent living archive paradigm. 2010, 44.

[4] Begelman, G., Keller, P., and Smadja, F. Automated tag custering: Improving search and exploration in the tag space. Collaborative Web Tagging Workshop at WWW2006 Edinburgh Scotland 22, 2 (2006), 15–33.

[5] Capocci, A. and Caldarelli, G. Folksonomies and clustering in the collaborative system CiteULike. Journal of Physics A: Mathematical and Theoretical 41, 22 (2008).

[6] Damme, C.V., Hepp, M., and Siorpaes, K. FolksOntology: An integrated approach for turning folksonomies into ontologies. Bridging the Gap between Semantic Web and Web 2, (2007), 57-70.

[7] Fichter, D. Intranet applications for tagging and folksonomies. Online-Weston Then Wilton 30, 3 (2006), 43-46.

[8] Gligorov, R., Baltussen, L.B., Ossenbruggen, J. van, et al. Towards integration of end-user tags with professional annotations. Web Science, (2010).

[9] Hildebrand, M. and Ossenbruggen, J.V. Linking user-generated video annotations to the web of data. Multimedia Semantic, (2011), 261-277.

[10] James Governor. On the emergence of professional tag gardeners. 2006. http://www.redmonk.com/jgovernor/2006/01/10/on-the-emergence-of-professional-tag-gardeners/.

[11] Oomen, J., Baltussen, L.B., Sander Limonard, et al. Emerging practices in the cultural heritage domain - social tagging of audiovisual heritage. Web Science 10: Extending the Frontiers of Society On-Line, (2010).

[12] Plangprasopchok, A., Lerman, K., and Getoor, L. Growing a tree in the forest: Constructing folksonomies by integrating structured metadata. KDD2010, ACM (2010), 949-958.

[13] Rolla, P.J. User tags versus subject headings: can user-supplied data improve subject access to library collections? Library Resources & Techinical Services 53, 3 (2009), 174.

[14] Sigurbjörnsson, B. and van Zwol, R. Flickr tag recommendation based on collective knowledge. Proceeding of the 17th international conference on World Wide Web - WWW '08, ACM Press (2008), 327.

[15] Specia, L. and Motta, E. Integrating folksonomies with the semantic web. In E. Franconi, M. Kifer and W. May, eds., The Semantic Web Research and Applications. Springer, 2007, 624-639.

[16] Trattner, C., Körner, C., and Helic, D. Enhancing the navigability of social tagging systems with tag taxonomies. i-KNOW '11, ACM (2011).

[17] Weller, I.P. and K. Tag gardening for folksonomy enrichment and maintenance. Webology 5, 3 (2008).