Crowdsourcing image descriptions using gamification : a comparison between game-generated labels and professional descriptors

Ivanjko, Tomislav

Crowdsourcing image descriptions using gamification: a comparison between game-generated labels and professional descriptors

Tomislav Ivanjko*
* Faculty of Humanities and Social Sciences, University of Zagreb / Department of Information and Communication Sciences, Zagreb, Croatia
tivanjko@ffzg.hr

***Abstract* – The traditional approach to achieve high metadata quality in image description is to use subject experts. However, cultural heritage institutions often lack the human resources to handle the amount of material that is in need of description. One of the possible solutions to this problem is applying the gamification approach in the process of description. Many studies have shown that applying game design features outside traditional game environments can increase the motivation and productivity, and that those games can be particularly effective in invoking intrinsic motivations and overall enjoyment. However, there is a need to explore the quality of such game-generated tags in comparison with using controlled vocabularies and traditional approaches. In this paper, we compare game-generated image labels and professional descriptors. First, a subject expert using controlled vocabulary added descriptors for each image. Then, by using a gamified platform for collecting semantic annotations of digitized images, game-generated tags were collected. In the final stage, game-generated labels were evaluated by the subject expert in the context of appropriateness of using them as descriptors within a standardized system. Results have shown that game-generated labels can serve as a basis for high quality labels suitable for including a standardized description in order to enhance description and retrieval.**

Keywords - gamification; image labelling; photographs; subject acess

I. INTRODUCTION

While the technology to search text is widely available and familiar, the technology to effectively search images and video without any descriptive metadata is still in its early stages. Most web search engines find images by indexing the text that surrounds any image and matching it with the given query. Naturally, this means that someone has to add annotations to every single image in order for the image to be retrieved. Any such process of manually adding descriptive metadata to a large image archive would entail large expenses and a large amount of time. Since algorithmic solutions for image retrieval are currently not on par with human performance, there is a need to investigate how the process of adding descriptive annotations could potentially be transformed from a tedious and labour-intensive task into an enjoyable experience. In this paper, we discuss different game-based approaches in harnessing collective efforts of online communities, with a special emphasis on the development of solutions aimed at image labelling and descriptive metadata.

II. THE RISE OF GAMIFICATION

With the development of different digital technologies in the Web 2.0 era, people have moved from a passive position to a more participative role and now have the spare time to use their *cognitive surplus* (spare processing power of the brain) to engage in different online activities harnessing collective intelligence [1]. One such activity that has become a main source of using that cognitive surplus is gaming, not only for children or teenagers, but for the entire age spectrum. The latest Entertainment Software Association [2] annual report on the essential facts about gaming habits in America states that 60 percent of Americans play video games daily, with an average of two gamers in each game-playing US household. According to the report, adult women represent a greater portion of the video game-playing population (33%) than boys under 18 (17%) and the average female video game player is 36, whereas the average male video game player is 32. According to the research, today, the average gamer is a woman of 36 years who plays casual games on her mobile phone as a favourite leisure activity [2].

Following the popularity of games as a leisure activity among all age groups, organisations across the globe began exploring games as a way to outsource various tasks related to their products and services to the crowds gathered under the notion of gamification [3]. A seminal paper on the topic of gamification [4] defined gamification as “the use of game design elements in non-game contexts.” In a later paper [5], authors differentiate between gamification and gameful design, stating that gamification is the application of gameful design:

gameful design is defined by the end of affording ludic qualities or gamefulness (the experiential qualities characteristic for gameplay) in non-game contexts. In contrast, gamification describes the means of using game design elements in non-game contexts, typically for the end of gameful design [5]. One of the most commonly used frameworks for game design is the MDA framework consisting of *mechanics, dynamics, and aesthetics*, and it is used as a tool to help designers, researchers and scholars achieve gamified design [6]. *Mechanics* describes the particular components of a game, at the level of data representation and algorithms. It includes various actions, behaviours and control mechanisms afforded to the player within a game context. Together with the game's content (levels, assets, etc.), mechanics support the overall gameplay dynamics. *Aesthetics* describes the desirable emotional responses evoked in the player when he interacts with the game system [6].

There is a different number of terms emerging in literature when these types of games that have some greater purpose or goal are discussed, using the collective participation of human players to support computational activities [7]. Games that focus on scientific problems have been called "citizen science games", while others have more broadly used the terms "games with a purpose," (GWAP) or "human computation games." Reference [7] also adds a new term, "knowledge games," defining them as: "…games that seek to invent, create, and synthesize new understandings of the world, solve real-world problems big and small, and help us reconsider, reframe, and reflect on humanity and our universe. "[7].

Various empirical studies indicate that gamification is an effective approach to increase user motivations [8], participation and long-term engagement [9]. Thus, the use of gamification in crowdsourcing is drawing an increasing attention, in both academia [10] and practice [11].

III. GAMIFYING IMAGE LABELLING

A large amount of research has been carried out on image retrieval (IR) in the last two decades. Image retrieval systems can be broadly categorized into two main categories: context-based and content-based. Context-based image retrieval systems use text to describe the image, whereas content-based image retrieval (CBIR) systems employ visual features such as colour, shape, texture and object position for image description [12]. Context-based image retrieval systems have been used since the late 1970s and are still the predominant method used for image search. They are known to be more efficient and accurate, and are based on assigning metadata to images. The metadata could be a title, natural language description, author, date and time of creation, and assigned keywords, either with the help of controlled vocabulary, professional natural language description, or through social tagging [12].

Within the content-based image retrieval systems, there are two main approaches: one that focuses on automatic methods and one that focuses on human-oriented image metadata creation. Automatic approaches aim to identify semantics relevant to the content of static images via the identification of visual features. Various approaches that use machine learning for image or image region categorization can perform well, but are limited to a small number of categories and the lack of training sets to be used effectively for the acquisition of more specific metadata [13]. Generally, automated approaches introduce certain inaccuracy, which makes them difficult to apply to heterogeneous resources. More or less, they also need large training sets of already annotated images, which stress the initial requirement of human labour needed to create them [13]. The traditional approach to achieve high metadata quality is carried out by using dedicated experts, who are aware of the purpose of their activity and who annotate resources as their primary job. However, such experts must be properly paid for their work, and cultural heritage institutions often lack the resources to handle the amount of material that is in need of cataloguing. Therefore, in the image domain, image annotation cannot yet be effectively performed via automated approaches and it requires manual image annotation, which does provide quality metadata, but does not scale due to the lack of time, resources and motivation of humans to do it [14].

One of the possible solutions to this problem is applying the gamification approach in the process of annotation. First significant developments in this area came from the field of human computation with the works of Luis von Ahn [15] and the related papers which explored how using the collective intelligence can solve problems that are hard or still impossible to do by using computer programs or algorithms [16], especially interested in the field of image labelling [17]. In order to address this problem, he introduces the term *human algorithm games* as a "paradigm for utilizing the human processing power to solve problems that computers cannot yet solve" (von Ahn, 2005). Von Ahn applied this idea by developing one of the first, as he called it "game with a purpose", the *ESP Game* [16]. The general idea was to gamify the process of image labelling in the online environment. From the player's perspective, the goal of the *ESP Game* is to guess what their partner is typing for each image. Once both players have typed the same string, they move on to the next image. The game does not ask the players to describe the image: all they are told is that they have to

"think like each other" and type the same string (thus the name "ESP"). The game was very successful, and almost 1.3 million labels were collected with only 13 630 players, some of whom spent over 50 hours playing the game. Manual evaluation of the added labels showed that more than 85% of labels were relevant, thus providing the evidence that this game-based approach produces high-quality labels [17]. However, despite its influence and novelty, there were some shortcomings in the way people labelled images in the *ESP Game*. Firstly, there was a lot of redundancy in the tag sets with a number of synonyms appearing as labels, secondly, there was a tendency to match based on colours and thirdly, players tended to add more generic labels [18].

Another development in the field of metadata crowdsourcing games aimed at describing different digitized collections was the development of the Metadata Games [19] project, a free and open source software system that uses computer games to collect information on archival images. Metadata Games is designed to be a free, open source, customizable software package available to a wide range of cultural heritage institutions without expensive licensing fees or contracts, helping cultural heritage institutions gain useful data for their collections, assisting scholars in learning how to interact and utilize collections in new and possibly unexpected ways, and providing a host of opportunities for the public to interact with cultural heritage institutions [20]. The project currently offers two mobile and four online games aimed at collecting metadata, and is used with over 45 collections worldwide.

The usability testing of the Metadata Games platform compared crowdsourced metadata generated by the users of the Metadata Games platform against the traditional metadata provided by the staff of the Boston Public Library. The findings of the study demonstrated that crowdsourced metadata, when used in tandem with traditional metadata, increases findability, corrects preventable search failures, and is by and large accurate [21]. It was concluded that games can be particularly effective in invoking intrinsic motivations and overall enjoyment, and that given the same tagging conditions, librarians and non-librarians produce a surprisingly similar distribution of useful metadata [21]. Other studies have also shown that applying game design features outside the traditional game environments can increase motivation and positively influence the behaviour of individuals [10].

In order to examine game-generated tags and their usefulness for subject description, a comparison of game-generated labels and professional descriptors was undertaken. The research question examined whether game-generated image labels can produce useful semantic annotations for visual resources in comparison with the traditional approach carried out by using dedicated experts.

IV. GAME-BASED LABELS VS. PROFESSIONAL DESCRIPTORS: A CASE STUDY

The research was structured into three complimentary stages. In the first stage, 20 digitized photographs were thematically selected by a subject expert in charge of the collection of the photographs used in the "Croatian Homeland War" exhibition authored by the Croatian Historical Museum. After the process of selection, the subject expert added his professional descriptors to each item, following relevant standards for assigning subject metadata**.** The subject expert assigned a total of 117 descriptors on 20 photographs with an average number of six descriptors assigned to one item. (Table 1). After that, a game-based application was developed and used for gathering image labels on the same 20 items. The application was based on the tools developed by the Metadata Games (metadatagames.org) project, a free and open source software system that uses computer games to collect information on archival images (Figure 1).

FIGURE I. APPLICATION INTERFACE FOR GAME-BASED LABEL COLLECTING

By using this platform and customizing it in local language, we had the tool to compare game-generated metadata against the traditional metadata assigned by the subject expert. The application was active during a period of 30 days and it attracted a total of 147 active players contributing a total of 1811 different tags describing a total of 20 different images. (Table 1)

TABLE I. GAME-GENERATED VS SUBJECT EXPERT DESCRIPTORS

	Game-generated labels	Subject expert descriptors
No. of labels	1811	117
Max. no. of tags assigned	129	9
Min. no. of tags assigned	56	4
Avg no. of tags assigned	90.5	5.85

After this first stage of data collection, each item in the collection was now described with two types of data: (1) descriptors assigned by the subject expert and (2) game-generated labels.

The second stage was aimed at analysing the collected corpora. By comparing these two sets of labels on a syntactical level, there was an overlap of 52.87%, meaning that little less than half of all game-generated labels was not present as a descriptor assigned by the subject expert. This left us with 850 game-generated labels for which we wanted to know their potential as possible index terms. When this set was analysed on a linguistic level, it was shown that a typical game-generated label consists from either one or two words (91%), is a noun (82%), in singular (78%) and in its nominative form (99%). In that sense, the analysis showed that a typical game-generated label does not differ from the linguistic characteristics of standard descriptors used for subject indexing. This has led us to believe that such game-generated labels can provide a basis for including the user warrant criteria in the description as a basic prerequisite for any subject indexing process, and that these labels could be useful as additional indexing terms.

For this reason, the final part of the research analysed the usefulness of game-generated labels for subject description from the point of view of the subject expert who assigned the professional descriptors in the first stage.

In order to make the set manageable for the subject expert, a frequency threshold for each label was introduced. It was decided that any label that has a frequency of three or higher would be subjected to subject expert evaluation. Since other game-based collection platforms such as the ESP game considered the label validated only if two players agreed on the same label, it was considered that if three independent players entered the same label describing one image, this label should be considered relevant within the system. By using this approach, the initial set of 850 game-generated labels was reduced to a manageable 178 labels subjected to expert evaluation. This meant that on average, for each item, the expert would evaluate eight or nine game-generated tags for each image. The research idea was to examine what percentage of these game-generated labels (with a frequency of three or more) would the subject expert include in the standard description within a controlled environment such as a library or a museum catalogue. The expert used a 5-point Likert scale ranging from 1 (would not include the label in any form), across 3 (would include the label with some normalization added) to 5 (would include the label as it is). Figure II shows the grades for each of the 20 images.

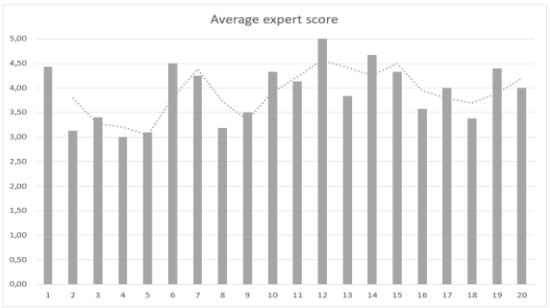

FIGURE II. AVERAGE EXPERT SCORE OF GAME-GENERATED LABELS

As we can see, the expert valued the game-generated tags very high with an average score of 3.91, and no tags scored lower than 3. This effectively shows that game-generated tags, from an expert point of view, provide high quality labels suitable for inclusion in a standardised description, such as museum or library catalogues.

V. CONCLUSION

The right use of gamification in crowdsourcing can significantly increase enjoyment and participation, and can transform tedious and time-consuming activities into fun and exciting experiences.

Gamification is not just a method of getting things done by the user, but an approach that can engage them to contribute, collaborate and co-create. Many successful projects in the heritage sector, often gathered under the notion of citizen science or crowdsourcing, have shown that gamification can be a viable solution to a number of tasks, and many projects started to explore the potential of that kind of gamification approaches in order to make different tasks more appealing, fun and productive.

When applied to the field of image description, gamification has also shown many benefits. Since image annotation cannot yet be effectively performed via automated approaches and since it requires manual image annotation, cultural heritage institutions often lack the resources to handle the amount of digitized material in

need of indexing. Many studies have shown that applying the gamification approach in the process of annotation can increase motivation and positively influence the behaviour of individuals. Furthermore, it has been deduced that games can be particularly effective in invoking intrinsic motivations and overall enjoyment. The key issue when gathering crowdsourced image descriptions lies in the achievement of the semantic richness of labels in terms of search and retrieval.

This paper has presented a case study for the potential of including game-generated image labels into standard description, comparing them with standard descriptors assigned by a subject expert. It was shown that a typical game-generated label does not differ from the linguistic characteristics of standard descriptors used for subject indexing. Furthermore, when game-generated labels (with a frequency of three or higher) were subjected to evaluation by a dedicated subject expert, it was shown that game-generated tags provide high quality labels suitable for inclusion in a standardised description, such as museum or library catalogues.

Following this results, we can conclude that this research confirms the earlier results [21], and that game-generated image labels can serve as a basis of including the user warrant perspective into the standardized description in order to enhance description and retrieval.

REFERENCES

[1] C. Shirky, Cognitive surplus: creativity and generosity in a connected age. London: Penguin, 2010.

[2] ESA, "Essential facts about the computer and video game industry," 2018. Available at: http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf (Accessed 31 January 2019)

[3] A. Eveleigh, C. Jennett, S. Lynn, and A. L. Cox, "I want to be a captain! I want to be a captain!: gamification in the old weather citizen science project," in Proceedings of the first international conference on gameful design, research, and applications, L. E. Nacke, Ed. New York: ACM, 2013, pp. 79-82.

[4] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining 'gamification'," in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. New York: ACM, 2011, pp. 9-15.

[5] S. Deterding, "The lens of intrinsic skill atoms: a method for gameful design," Human-Computer Interaction, vol. 30, no. 3-4, 2015, pp. 294-335.

[6] R. Hunicke, M. LeBlanc, and R. Zubek, "MDA: a formal approach to game design and game research," in Proceedings of the AAAI Workshop on Challenges in Game AI, 2004. Available at: http://www.cs.northwestern.edu/~hunicke/MDA.pdf (Accessed 31 January 2019)

[7] K. Schrier, Knowledge games: how playing games can solve problems, create insight, and make change. Baltimore, MD: JHU Press, 2016.

[8] R. Tinati, M. Luczak-Roesch, E. Simperl, and W. Hall, "An investigation of player motivations in Eyewire, a gamified citizen science project," Computers in Human Behavior, vol. 73, 2017, pp. 527-540.

[9] N. R. Prestopnik, and J. Tang, "Points, stories, worlds, and diegesis: comparing player experiences in two citizen science games," Computers in Human Behavior, vol. 52, 2015, pp. 492-506.

[10] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? - a literature review of empirical studies on gamification," in 47th Hawaii international conference on system sciences (HICSS). Washington: IEEE Computer Society, 2014, pp. 3025-3034.

[11] K. Huotari, and J. Hamari, "A definition for gamification: anchoring gamification in the service marketing literature," Electronic Markets, vol. 27, no. 1, 2017, pp. 21-31.

[12] E. Konkova, A. S. Goker, R. Butterworth, and A. MacFarlane, "Social tagging: exploring the image, the tags, and the game," Knowledge Organization, vol. 41, no. 1, 2014, pp. 57-65.

[13] J. Šimko, M. Tvarožek, and M. Bieliková, "Human computation: image metadata acquisition based on a single-player annotation game," International Journal of Human-Computer Studies, vol. 71, no. 10, 2013, pp. 933-945.

[14] D. Paraschakis, and M. G. Friberger, "Playful crowdsourcing of archival metadata through social networks," 2014. Available at: https://www.researchgate.net/publication/299978917_Playful_crowdsourcing_of_archival_metadata_through_social_networks (Accessed 31 January 2019)

[15] L. von Ahn, Human computation: doctoral dissertation. Pittsburgh, PA: Carnegie Mellon University, 2005.

[16] L. von Ahn, "Games with a purpose," Computer IEEE, vol. 39, no. 6, 2006, pp. 92-94.

[17] L. von Ahn, and L. Dabbish, "Labeling images with a computer game," in Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM, 2004, pp. 319-326.

[18] S. Robertson, M. Vojnovic, and I. Weber, "Rethinking the ESP game," in Extended Abstracts on Human Factors in Computing Systems (CHI'09). New York: ACM, 2009, pp. 3937-3942.

[19] Metadatagames.org, "Metadata games," 2018. Available at: http://www.metadatagames.org/about/ (Accessed 31 January 2019)

[20] M. Flanagan, and P. Carini, "How games can help us access and understand cultural artifacts," American Archivist, vol. 75, no. 2, 2012, 514-537.

[21] C. Manzo, G. Kaufman, S. Punjashitkul, and M. Flanagan, "By the people, for the people: assessing the value of crowdsourced, user-generated metadata," DHQ: Digital Humanities Quarterly vol. 9, no. 1, 2015, p. 6.