

Gamesourcing Expert Painting Annotations

Myriam C. Traub
Centrum Wiskunde &
Informatica
Science Park 123,
Amsterdam
myriam.traub@cwi.nl

Jacco van Ossenbruggen
Centrum Wiskunde &
Informatica
Science Park 123, Amsterdam
jacco.van.Ossenbruggen@cwi.nl

Jiyin He
Centrum Wiskunde &
Informatica
Science Park 123,
Amsterdam
jiyinhe@gmail.com

Lynda Hardman
Centrum Wiskunde &
Informatica
Science Park 123,
Amsterdam
lynda.hardman@cwi.nl

ABSTRACT

Online collections provided by museums are increasingly opened for contributions from users outside the museum. These initiatives are mostly targeted at obtaining tags describing aspects of artworks that are common knowledge. This does not require the contributors to have specific skills or knowledge. Museums, however, are also interested in obtaining very specific information about the subject matter of their artworks. We present a game that can help to collect expert knowledge by enabling non-expert users to perform an expert annotation task. This is achieved by simplifying the expert task and providing a sufficient level of annotation support to the users. In a user study we could prove the usefulness of our approach.

Author Keywords

Crowdsourcing; expert task; annotation; wisdom of the crowd.

ACM Classification Keywords

H.5.2. User Interfaces; H.5.3. Group and Organization Interfaces; K.4.3. Organizational Impacts

INTRODUCTION

The Rijksmuseum Amsterdam makes large parts of its collection available online and continuously adds new items. Visitors of the website cannot only look at the images, they are also invited to interact with the items by annotating or downloading them in high resolution to use them in a creative way [4]. The annotations that users contributed so far, however, show that mechanisms are needed to detect and remove incorrect annotations and to help users provide more qualitative annotations.

The SEALINCMedia project [5] aims to support the Rijksmuseum Amsterdam in this endeavor. We develop an online platform called Accurator that combines functionalities to manage the administration and the execution of the annotation process. The predominantly art-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author(s).

Published in: van Leeuwen, JP, Stappers, PJ, Lamers, MH, Thissen, MJMR (Eds.) *Creating the Difference: Proceedings of the Chi Sparks 2014 Conference*, April 3, 2014, The Hague, The Netherlands.

historical knowledge available from museum professionals is extended by involving people from outside the museum.

To maximize not only the coverage, but also the level of detail of annotations, we collect common knowledge as well as expert knowledge. Common knowledge, i.e. information that can be provided by any user (e.g., “the painting shows a blue bird and a branch with red leaves”) can be collected via crowdsourcing platforms without knowing the skills of the users. Obtaining expert knowledge, (e.g., “the bird depicted is a Daurian Redstart”) from experts is in comparison much more difficult. Finding experts of the targeted niche (i.e. ornithologists), persuading them to help in an annotation task and keeping them engaged are considerable challenges. While overcoming these difficulties promises to lead to high quality annotations we decided not to solely rely on the traceability and good will of experts.

Therefore, to bridge the gap between the availability of lay people on crowdsourcing platforms and the scarcity of expert knowledge, we have developed a game that transforms an expert annotation task in a way that it can be carried out by non-experts. This adds another source for expert knowledge to our platform (see Fig. 1) and thereby enhances our chances of obtaining qualitative and specific descriptions of paintings.

We conducted a study [2] to investigate whether our approach is suitable to support crowdsourcing of expert annotations. We investigated

- whether the task is actually feasible for non-experts at a level that is comparable to experts;
- whether non-experts improve while playing;
- how the partial absence of the correct answer influences the users’ performance; and
- whether the aggregation of user judgements improves the agreement with experts.

During the conference, we will give conference participants the chance to test their knowledge of art-history with the “imperfect” version of the game and enjoy the competition with others while – hopefully – learning more about classification of paintings.

THE ART GAME

Our game is based on the online tagging game used for the Fish4Knowledge project [1]. For comparability reasons it was largely left unchanged. The game is available online [6].

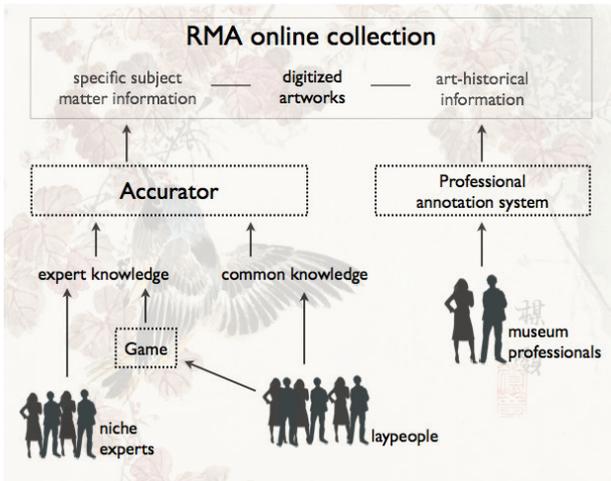


Figure 1: How laypeople and experts can contribute to a more detailed and varied description of artworks.

The Task

The task we selected for the game is the classification of paintings into 17 selected subject types (see Table 1) from the Art & Architecture Thesaurus (AAT) [7]. This task is usually performed by museum professionals. We chose the subject type classification because the collection items of the Rijksmuseum Amsterdam lack this information and because we were able to re-use data collected for a graduation project [3].

The Interface

On the start page, users are given instructions on how to perform the annotation task (if needed, they can go back to the instructions at any point in the game). Users are asked to be as *specific* as possible in their judgements. This means, that when they can choose between a general subject type (e.g. *figures*), and a specific subject type (e.g. *half figures*), they should select the more specific one.



Figure 2: Interface of the art game with the large query image on the upper left. The five candidate subject types are shown below, together with the *others* candidate.

After creating an account, users can start playing the game. They are presented a succession of images (referred to as *query images*) of paintings (see Figure 2) taken from the Steve Tagger [8] data set. Apart from the image, we provide no further information about the painting. Within the first ten images that are presented to the user, there are no

repetitions. Afterwards, images may be repeated with a 50% chance.

The query images are presented prominently in the upper left corner of the interface. Below, we present a pre-selection of six *candidates*. Five of these candidates represent subject types and one of them (labeled “others”) can be used if the assumed correct subject type is not presented. A candidate consists of an image, a label (AAT subject type) and a description. For each subject type we selected one representative image from the corresponding Wikipedia page, e.g., [9].

These images are intended to give users a first visual indication of which subject type might qualify and it makes it easier for users to remember it. If further information is needed to judge the image, users can display short descriptions taken from the AAT by moving the mouse cursor over the candidate, for example:

Marines:

“Creative works that depict scenes having to do with ships, shipbuilding, or harbors. For creative works depicting the ocean or other large body of water where the water itself dominates the scene, use ‘seascapes’.” [10]

The descriptions of the subject types are important, as the differences between some subject types are subtle.

Subject type	#
full-length figures	40
landscapes	33
half figures	13
allegories, history paintings, portraits, animal paintings, genre, kacho, figures	8
townscapes	6
flower pieces	5
marines, cityscapes, maesta, seascapes, still lifes	3

Table 1: Used subject types and the number of expert annotations.

The Feedback for Users

To motivate users to annotate images correctly and to give them feedback about the “correctness”¹ of their judgements, they are awarded ten points for correct choices and one point for the attempt (even if incorrect). On the top right corner of the interface, users can see their current and overall score.

After finishing a round of 50 images, users are directed to the dashboard, where they can compare their scores to those of the other players.

The feedback for the users is based on a comparison of their judgements to annotations given by experts of the Rijksmuseum Amsterdam collected by [3]. From this data set, we selected 168 expert annotations for 125 paintings (Table 1). The number of annotations per painting ranges from four (for one painting) to one (for 83 paintings). These multiple classifications are considered correct: a painting showing an everyday scene on a beach [11] can be classified as *seascapes*, *genre*, *full-length figure* and *landscapes*.

EVALUATION

We evaluated the suitability of our approach with two experiments. The first experiment simulated “perfect” data, meaning that the correct candidate was always presented to the user. The results of this showed us that the users are able to perform the task and improve over time. We used this data as a baseline to compare the results of the second experiment to. Here, we deliberately removed the correct candidate in 25% of the cases to simulate an imperfect data set. The analysis of the results showed that while the agreement between users and experts is higher in the first experiment, it is still acceptably high in the second setting. The learning performance, however, was much lower, which suggests that users should be given a training phase before they can successfully play on imperfect data.

The application of majority vote on the users’ judgements leads to a noticeable improvement of the agreement with the experts. For some paintings, however, the agreement remained very low. We consulted another expert from the Rijksmuseum Amsterdam and identified two main reasons: The classification done by the experts was either incomplete or incorrect or the correct classification was only possible if context information about the painting was known.

The setup of the experiments, the analysis and our findings are described in detail in [2].

CONCLUSIONS

We present a game for classifying paintings into categories of a professional vocabulary. We have simplified the task by reducing the number of categories and added assistance for the users to perform the task. In a user study we could show that non-expert users are able to perform the task and improve over time. A “Wisdom of the Crowd” effect was observed when we aggregated the users’ votes. In some cases of very low agreement with experts we could identify incompleteness or incorrectness on the expert’s side to be a reason.

This approach allows us to enlarge the circle of potential contributors of high quality annotations.

REFERENCES

1. Jiyin He, Jacco van Ossenbruggen, and Arjen P. de Vries. “Do you need experts in the crowd? : a case study in image annotation for marine biology.” In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, OAIR ’13, pages 57–60, Paris, France, France, (2013). Le Centre De Hautes Etudes Internationales D’Informatique Documentaire.
2. Myriam C. Traub, Jacco van Ossenbruggen, Jiyin He, and Lynda Hardman. “Measuring the Effectiveness of Gamesourcing Expert Oil Painting Annotations.” In *Proceedings of the 36th European Conference on Information Retrieval*, volume 8416 of *Advances in Information Retrieval*, pages 112–123. Springer International Publishing Switzerland, (2014.)
3. Sjaak Wouters. Semi-automatic annotation of artworks using crowdsourcing. Master’s thesis, Vrije Universiteit Amsterdam, The Netherlands, (2012).
4. <https://www.rijksmuseum.nl/en/rijksstudio>
5. <http://sealincmedia.wordpress.com>
6. <http://sealincmedia.project.cwi.nl/artgame/accounts/login/>
7. <http://www.getty.edu/research/tools/vocabularies/aat/>
8. <http://tagger.steve.museum/>
9. <http://en.wikipedia.org/wiki/Maesta>
10. <http://www.getty.edu/vow/AATFullDisplay?find=marines&logic=AND¬e=english%3DN&subjectid=300235692>
11. <http://tagger.steve.museum/steve/object/280>

¹ By “correct” we mean that a given judgement is in line with the expert.