

Personalized Nichesourcing: Acquisition of Qualitative Annotations from Niche Communities

Chris Dijkshoorn¹, Mieke Leyssen², Archana Nottamkandath¹, Jasper Oosterman³, Myriam Traub², Lora Aroyo¹, Wan Fokkink¹, Geert-Jan Houben³

¹ The Network Institute, VU University, Amsterdam, The Netherlands
{c.r.dijkshoorn, a.nottamkandath, l.m.aroy, w.j.fokkink}@vu.nl

² Centrum Wiskunde en Informatica, Amsterdam, The Netherlands
{leyssen, traub}@cwi.nl

³ Web Information Systems, Delft University of Technology, The Netherlands
{j.e.g.oosterman, g.j.p.m.houben}@tudelft.nl

Abstract. Diversity and profundity of the topics in cultural heritage institutions' collections make experts from outside the institution indispensable to acquiring qualitative annotations. We define the concept of nichesourcing and present challenges in the process of obtaining qualitative annotations from people in these niches. We believe that experts provide better annotations if this process is personalized. We present a framework, called Accurator, that allows to realize and evaluate strategies and applications for personalized nichesourcing.

Keywords: cultural heritage, nichesourcing, annotation framework, qualitative annotations, user interaction

1 Introduction

Access and retrieval mechanisms for archives and museums rely on a rich description of the collection. Most cultural heritage institutions employ professional experts to describe their collections by manually compiling metadata for each item. For large and diverse collections the knowledge of experts from other domains is indispensable. Cultural heritage institutions therefore seek to understand whether and how they can make use of external users to produce these annotations.

Our demo aims at understanding which strategies and techniques lead to high-quality annotations by (crowds of) external experts. The first challenge of the project is to identify the niche of relevant experts and to motivate them to contribute to the annotation of artworks. As a next step, the personalization mechanisms must make sure that the experts are shown items that correspond to their expertise. The quality of the annotations and annotators will be evaluated using trust algorithms. All these aspects must be presented in an appropriate interface.

To evaluate our hypotheses, we develop a framework designed to support crowd annotation processes, called Accurator.

2 Research Challenges

One of the challenges of nichesourcing is finding candidate annotators that will produce high quality annotations for collection items. We believe that persons part of a topical community have an active interest in the topic and might be willing to help and share knowledge related to that topic. We call these topical communities niches and they manifest themselves, among others, on the Social Web. We will analyze social data and perform user studies using Accurator to understand what identifies a niche community, whether a person is part of such a community and which properties identify a good candidate to provide qualitative annotations.

The challenge for recommender strategies in Accurator is twofold: keep the expertise needed to annotate the artwork in the range of the experts knowledge and yet diversify the suggestions to get high quality annotations for as many distinct artworks as possible. Our aim is to develop recommender strategies that use content patterns in the Linked Data cloud, resulting in a list of recommendations consisting of diverse artworks. We hypothesize that encountering diverse artworks to annotate will help keep the expert motivated.

We address issues of determining trust in the users and their contributed annotations by modeling the user reputation and tracking their expertise across various topics over time. We intend to use Subjective logic to model the reputation of users and semantic similarity measures to track and update the users expertise. Since there is no gold standard for evaluating the annotations, we must rely on peer reviewing process and other mechanisms such as provenance of the annotation process.

Since external users are not familiar with classification schemes and (art-)historical expert knowledge, the annotation process must be broken down into facile tasks that can be solved with little effort and without this kind of expert knowledge (suggested in [1]). The interface for such a system has to present the task in a straightforward way while motivating the users to contribute their knowledge and time. We will investigate which design aspects and underlying mechanisms are responsible for the quality and quantity of tags added by users and how to visualize trust and personalization aspects.

3 Accurator framework

Our main assumption is that making use of niche sourcing increases the quality of annotations. We believe the we can use specific techniques to identify niches and create user profiles. Based on the user profiles we can recommend relevant tasks to the user and apply trust mechanisms to improve the recommendations. In Figure 1, we show a diagram that represents the workflow of Accurator.

The process starts (see Figure 1a) with searching the social web for user generated content that is relevant for a specific topic. We calculate the relevance of the content creators in respect to the topic and exploit social relations to identify a topical niche. When a person starts using Accurator, a user profile (see Figure 1b) is build based on available data

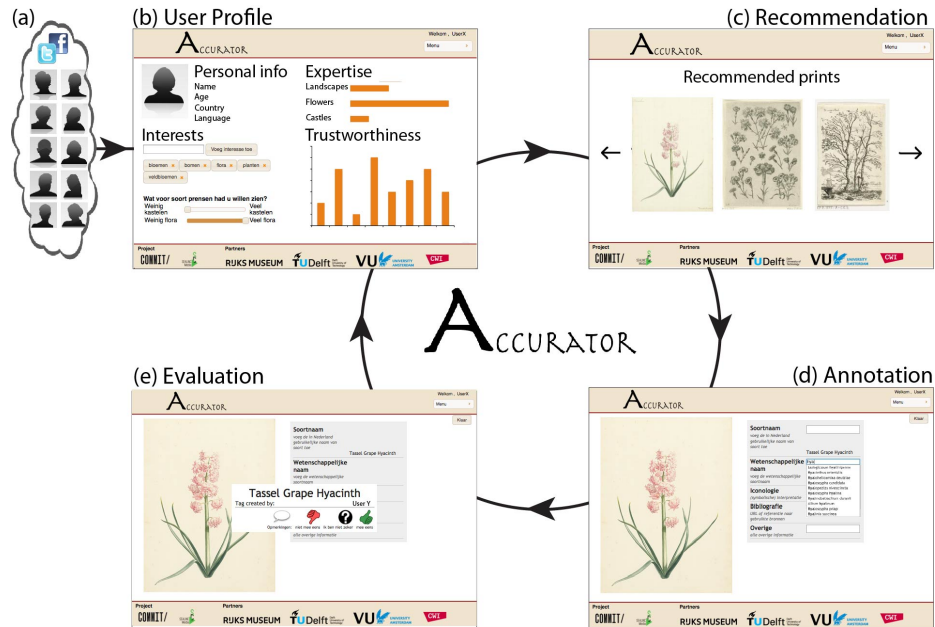


Fig. 1. Diagram representing the workflow of Accurator

Figure 1c shows the recommendation of collection items for a user. The recommendation strategy is based on specific patterns in the data, the user profile and the current annotation quality of an item. Accurator allows to easily change between different strategies to cater for different users. The choice of recommended item will affect the calculated interest of that user.

Figure 1d shows the interface where users add their annotations to an item. The presented fields depend on the topic and the expertise of the user on that topic. Accurator can also be configured to use a vocabulary for a field to support the user. Figure 1e shows the interface in which users can evaluate the annotations of other users. This task is only available to users who are trustworthy and have a certain level of expertise. The result of a review affects 1) the quality of an annotation, 2) the expertise level of a user and 3) the trustworthiness of another user.

Accurator is built using Clupatria to store RDF, GWT for the user interface and GAE for hosting. Accurator is now used for experimentation with data from the Rijksmuseum in the Netherlands and a demo is available at <http://rma-accurator.appspot.com>.

Acknowledgements. This publication was supported by the Dutch national program COMMIT. We like to thank the members of the SEALINCMedia worktable and in particular the Rijksmuseum for their support.

References

1. He, J., van Ossenbruggen, J., de Vries, A.: Do you need experts in the crowd? a case study in image annotation for marine biology. In: Open Research Areas in Information Retrieval (OAIR 2013), the 10th International Conference in the RIAO series (2013), to appear