Personalized Nichesourcing: Acquisition of Qualitative Annotations from Niche Communities

Chris Dijkshoorn[1], Archana Nottamkandath[1], Mieke Leyssen[2], Jasper Oosterman[3], Myriam Traub[2]

[1] The Network Institute, Dep. of Computer Science, VU University, Amsterdam, The Netherlands {c.r.dijkshoorn, a.nottamkandath}@vu.nl
[2] Centrum Wiskunde en Informatica, Amsterdam, The Netherlands {leyssen, traub}@cwi.nl
[3] Web Information Systems, Delft University of Technology, The Netherlands

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 the challenges in the process of obtaining qualitative annotations from persons in these niches. Our assumption is that if this process is personalized, we get better annotations from the experts. We present a framework for nichesourcing, 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 therefore 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 exploit the efforts of external users to produce these annotations.

This 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 with user studies, we develop a framework designed to support crowd annotation processes, Accurator.

2 Research Challenges

One of the challenges of nichesourcing is finding candidate annotators that will produce good quality annotations for collection items. Besides topical knowledge, properties like availability, willingness to help and being able to share or transfer knowledge are also important. 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. These topical communities we call niches and manifest themselves, among others, on the Social Web. We will analyze social data and perform user studies using the Accurator framework to understand what identifies a niche community, whether a person is part of such community and which properties identifies 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. To address these challenges we will investigate the use of content patterns in the Linked Data cloud. Our aim is to develop recommender strategies that use these patterns, resulting in a list of recommendations consisting of diverse artworks. Using the alternative paths created by the patterns, items can be reached which reside in the long-tail. When experts are able to annotate these long-tail items, they will become more accessible in general. From a user perspective diversity is also important, 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 like tracking provenance of the annotation process such as usage of terms from vocabularies by the user, typing speed etc. We also investigate the different metrics which will help in identifying good behavior of the users.

The professional annotation of artworks is a complex process that requires familiarity with the used classification schemes and (art-)historical expert knowledge. Since both will mostly not be available in candidate users for nichesourcing projects, this process must be broken down into facile tasks that can be solved with little effort and without expert knowledge of classification schemes as suggested in [**?**]. 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. To gain a better understanding of how to design such an interface, we investigate what 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 personalizing the niche sourcing process 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 motivate users, provide feedback and improve the recommendations. In Figure 1 we show the interfaces of Accurator that map onto the the aforementioned challenges.

Insert figure here.

The process starts, see Figure 1a, with searching the social web for user generated content that is relevant for a topic. We calculate the relevance of the content creators in respect to then topic and exploit social relations to identify a topical niche. We motivate the niche to use Accurator and when a person starts using Accurator a user profile, see Figure 1b, is build based on available data and shown to the user. The user can specify additional social web accounts.

Figure 1c shows the recommendation of tasks for a user. The recommendation strategy is based on specific patterns in the data, the user profile of the user and the current annotation quality of an item. Accurator allows to easily change between different strategies to cater for different users and a future task is to automatically adapt the choice of strategy based on that user profile. The choice of recommended item can affects the calculated interest of that user.

Figure 1d shows the interface where users add their annotations to a collection item. Choice of fields dependent on the topic and the expertise of the user on that topic. Users with more expertise on that topic are allowed to enter more difficult fields. Accurator can also be configured to use a vocabulary for a field to support the user. Figure 1e shows a similar interface where users review the annotations of other users. Reviewing tasks are 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 a user.

Another aspect that holds for all interfaces is that they should be intuitive and helpful. Users who encounter difficulties with the interface will not return.

Accurator is build using Cliopatria 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.