Can we have trustworthy digital curators?

Archana Nottamkandath
VU University Amsterdam
1081HV, De Boelelaan 1105
+31-616-221337
a.nottamkandath@vu.nl

1. INTRODUCTION
Crowd sourcing is widely employed on the web to accomplish a wide variety of tasks within reasonable amount of time. Cultural heritage institutions such as Steve.Museum, BBC Your Paintings, Brooklyn Museum, Rijksmuseum Amsterdam etc. are also enriching their collection using general or more specific information about artefacts and paintings through wisdom of the crowds. Traditional tag analysis techniques such as calculating frequency of users who provided same tags for a particular image, presence of tags in particular vocabularies etc. cannot be used independently in the cultural heritage institutions, since such methods cannot effectively filter out the spammers. The tags obtained through crowd sourcing techniques by these cultural heritage institutions are used for information retrieval purposes thereby helping the users to effectively search for items through the museum collections. Thus it is very important that the information contributed from the crowds have good quality before being put into any use, since the museums have a high reputation to uphold to. The quality threshold for the crowd sourced information varies across museums. Employing trusted people for manual assessment of all the tags provided by all users across the Web seems to be an next to impossible task. The goal of the work described below is to show how it is possible to automate the process of evaluation of the tags obtained through crowdsourcing. This is done by first having the museum authorities manually evaluate a small part of user contributed tags in order to train our system about the categories used for evaluation. We later apply our supervised learning algorithms to automatically evaluate the quality of tags contributed in the future. Using our algorithm we were able to semi-automatically evaluate tag quality with an accuracy of up to 80%.

2. OUR APPROACH
In our work we use Subjective Logic to represent user opinions about topics. To our knowledge this is a novel research in the direction of automating tag evaluations on crowdsourcing systems by coupling Subjective logic opinions with measures of semantic similarity. Our goal is to reduce as much as possible the need for human intervention for evaluating annotation in crowd sourcing systems. Our hypothesis is that if a user is an expert on a particular topic such as flowers, he/she might also be an expert on semantically similar topics such as trees or plants. In our evaluations, we validate our hypothesis through our experiments on Steve Museum dataset (see [2] [3]). We also employed provenance techniques to track the annotation process. In this approach, we track the user behaviour such as typing time, time of the day etc to determine the quality of the contributed information. Our hypothesis is that user behaviour is a good indication of the quality of the tags. For e.g.: we hypothesize that if a user types the information faster, then he/she might be more confident, annotations provided at certain times of the day are more trustworthy than others etc. We prove the validity of our hypothesis using Waiscla? dataset, which is a video tagging game. The provenance is represented using PROV-model where we can represent the user and their provenance metrics as graphs. This method is quite useful in scenarios where the user is anonymous, and very less information is available to build the user reputation. We then combined the previous two approaches for determining user reputation and tracking provenance and studied its effects on our algorithms to determine the quality of user contributed tags (see[1]). When no information about the user is available, we solely rely on provenance techniques to evaluate the quality of user contributed information. But when sufficient information about the user is available to build the reputation, we take into account both the reputation of the user and the provenance measures to evaluate the user contributed information. The combined approach works better than the individual cases and provides us better results. The research design, implementation, evaluation and results of our work will be presented in detail in our poster.

3. REFERENCES