Crowdsourcing architectural knowledge: Experts versus non-experts

Patrick Alexander Brouwer
2523300
Bachelor Informatie, Multimedia, & Management.
p.a.brouwer@student.vu.nl
Supervisor: Victor de Boer

Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Abstract. A growing number of archive and heritage organizations are digitalising their collections, moving their respective knowledge into the public domain. Often only limited metadata about these collections is available. This data, while useful, does not provide a way to search through these vast collections with descriptive keywords, such as house or chimney. The ArchiMediaL project[1] aims to solve this problem by using a number of Artificial Intelligence solutions. This paper looks at crowdsourcing as an alternative solution. A group of architecture experts and a group of non-experts were asked to annotate several objects. A team of independent evaluators provided data supporting the fact that crowdsourcing can be seen as a viable option. The data also suggests that the expert annotations were of a higher quality than those of non-experts.

Keywords: crowdsourcing, annotation, linked data, expert knowledge, architecture

1 Introduction

In the current age of information, legacy archive and cultural heritage organizations are using linked data to make their collections available through the internet [1]. Digitalisation of these collections is mostly done by hand and usually only involves adding metadata of the object to the linked data store. This data, while relevant, often fails to describe the actual content of each object in the collection. Because of this, a search engine, which has indexed the collection, is capable of running searches for "painting", "photo", "created in 1864", "Picasso", etcetera. But what if you want to find all objects which contain a depiction of, for example, a house or a horse? For these kinds of searches, a whole new level of information is required. Data which describes the content of an art piece, instead of data about the object itself, descriptive data. The ArchiMediaL project is a collaboration between the TU Delft and the VU Amsterdam, which

[1] http://archimedial.eu/

aims to find a way to add this data to the Colonial Architecture database[2]. The project is mostly interested in information pertaining to architectural depictions.

The problem of generating descriptive metadata, can be approached in several ways, the most simple solution being: have experts assess each individual object by hand to generate the descriptive data required. This method has some major flaws. Namely, the time consuming nature of the work, as well as the fact that the assessor requires sufficient knowledge about the objects in order to add useful data. The latter being a problem because there are only so many "experts" in the population. To circumvent these issues the ArchiMediaL project has set out to investigate a number of Artificial Intelligence solutions which can identify and correctly label descriptive data in pictures and paintings. This approach removes a large part of the time required to generate the data, but it introduces the problem of teaching an Artificial Intelligence expert knowledge.

The approach described in this paper aims to solve the problem the other way around, by removing the need for experts. In this context an expert is someone who currently works, or has received a degree, in an Architecture related field. Most modern countries provide people with relatively broad schooling, as well as the opportunity to practice a vast range of hobbies. This provides a basis for assuming, that posing a question to a large crowd of randomly selected people, should lead to a correct answer after aggregating their responses. This means that even if the people who respond are in no way certified experts in the specific field of science, together they still know enough to answer complex questions. Using crowds to solve complex problems has been shown to work on several accounts, in different fields of study [2][3]. However, no one has taken a look into the viability of crowdsourcing architectural knowledge.

Since the ArchiMediaL project is interested in finding a working solution, we will also try to find out if crowdsourcing to only experts is a potential solution. This approach allows for the acceptance or rejection of crowdsourcing as a viable option for use in the ArchiMediaL project. This has led to the following research question:

"Does crowdsourcing from either an expert or a randomly selected population provide data which is rich enough for use in the ArchiMediaL Project?"

To find out if crowdsourcing to a wider audience, and aggregating their responses, can be seen as equivalent to asking solely experts, the question above has been complemented with the following:

"Can an expert's annotation be identified as such for more than 50% of the instances when tested blindly by ArchiMediaL project members?"

Answering these two questions will give insight in both the viability of crowdsourcing architectural knowledge, as well as the necessity of experts in this process.

To find the answers to these questions, we will first take a look at previous work related to annotating and crowdsourcing in section 2. In section 3, the

[2] http://colonialarchitecture.eu/

research method used to create this experiment will be described. Afterwards, in section 4, we take a look at the results of this experiment, followed by a discussion based on these results in section 5. Finally, section 6 will comprise of the conclusion of this paper.

2 Related Work

2.1 Annotation

The need for adding descriptive data to images has been around for a long time, as shown by a 2003 paper on creating file formatting which is capable of supporting additional metadata [4]. Over the years there have been many improvements on this concept. One of which is the Resource Description Framework (RDF), which is capable of creating vast graphs of semantic data [5]. De Boer, et al. [6], have developed a framework for the storage of cultural heritage metadata which uses RDF datastores. The application used for this paper uses RDF to store its data.

Research has not only focussed on how to store this new type of metadata, they have also investigated how to simplify the annotation process for users. For example, Hollink, et al. [7] have looked into the possibility of connecting multiple ontologies, to support the user in finding the right annotation for an image. This research shall be taken into account by the addition of an architecture related ontology to the experiment.

Most closely related is the work by Dijkshoorn, et al. [8], in which an application for the annotation of cultural heritage objects is developed. This application is called Accurator and allows users to easily tag images with annotations. This application will be used directly in our experiment. Dijkshoorn, et al. continue their work by looking into the viability of improving annotations by personalising the annotation environment [9].

2.2 Crowdsourcing

While the term crowdsourcing was first coined in an article by Jeff Howe[10], he only gave a broad overview of what crowdsourcing was capable of. In later years, Schenk & Gittard have summarised a solid foundation of crowdsourcing techniques which have been proven effective [11]. Whereas Erickson, et al. have come up with a way to match the needs and wants of a data collector, with the right way to approach crowdsourcing [12].

Using their Accurator application, Oosterman, et al. have expanded their research to the difference between experts and non-experts [13]. Their focus lays on floral art available at the Rijksmuseum in Amsterdam[3]. While the subject is different, there is some useful overlap with this research. This usefulness is mostly related to the research method and its evaluation.

[3] https://www.rijksmuseum.nl/

3 Research Method

To find an answer to the research questions, a crowdsourcing experiment was conducted with both a group of experts as well as a group of non-experts. To create this experiment we have followed Dijkshoorn, et al. [8] in their use of the Accurator application. Because of the existence of Accurator, it was not necessary to create a new crowdsourcing application.

While Accurator comes with its own test database, this did not include objects which were useable for this experiment. To acquire objects which were equal to the ones ArchiMediaL uses, we have turned to the Colonial Architecture database. This database contains objects related to colonial era buildings and sites located in non-European countries which were controlled by European countries at the time.

Finally, we have chosen for an experimental setup which closely resembles the one used by Oosterman, et al. [13].

3.1 Accurator

Accurator is an annotation application which is currently used to provide users with an easily (online) accessible way to add their annotations to the art owned by the Rijksmuseum[4]. It is the best solution available, because of its accessible user interface and the ability to run and modify a private instance of it.

Accurator's annotations are stored in a RDF datastore, which means that all data is stored in triples consisting of a subject, a predicate, and an object. This way of storing data enables the use of the SPARQL Protocol and RDF Query Language (SPARQL) [14], which allows users to perform complex searches that span multiple datastores throughout the internet.

Because of the flexibility which comes with the use of SPARQL, Accurator could be used for this experiment without any modifications. For example, keeping the annotations created by the expert population separated from those made by the non-expert population. This was easily done by creating usernames containing either "expert" or "crowd", SPARQL's partial string matching was then used to identify which annotations belonged to each population. The SPARQL queries used for this paper can be found online[5].

3.2 Objects

The four objects used in this experiment, figure 1, have been taken from the Colonial Architecture database. This was done mainly because it is the same database the ArchiMediaL team uses for their work. Because this is the case, this experiment closely resembles the real world application it tries to mimic. To prevent any bias in the selection of these images, they were selected randomly out of all the images depicting buildings.

[4] http://accurator.nl/

[5] `https://figshare.com/articles/Queries_docx/6725216`

(a)

(b)

(c)

(d)

Fig. 1: Objects used in experiment.

3.3 Experiment

As was the case for Oosterman, et al. [13] this experiment targeted two populations: experts and a crowd of non-experts. The non-expert participants were found through social media and several student associations. This provided a crowd which was diverse in age, as well as fields of knowledge. The expert population was found through acquaintance with the members of the ArchiMediaL team. To ensure the division between participant and evaluator, these experts were not part of the ArchiMediaL team.

Both populations were allowed to annotate the four objects, seen in figure 1, in a randomised order. This was done to prevent any bias regarding a particular object. The participants were instructed to provide annotations regarding the architectural entities depicted in these objects. To encourage participants to annotate everything they recognised, no limit was placed on the number of annotations each of them was allowed to create.

As can be seen in table 1, 865 annotations were provided by the non-expert population. To improve the quality of the crowd's annotations, a solution similar to the majority voting of Oosterman, et al.[13] was implemented. Before an annotation would be considered for use it should be mentioned at least twice by the participants. The result of this step was the discarding of 743 unique annotations.

Due to the difficulty of finding experts, only three have participated in the experiment. The 54 annotations they have provided have not been culled to improve quality because of the smaller number of participants. The fact that the annotations provided by experts are expected to be of a higher quality has aided in this particular decision.

	Expert	**Non-expert**
Population	3	28
Annotations	54	865
Verified Annotations	*n.a.*	122

Table 1: Gathered Data

After gathering the annotation data, a method for evaluation was required. Since the research pertains to the viability for use in the ArchiMediaL project, a way to evaluate was to ask the project's members for their opinion. Due to their involvement with the project they have a clear understanding of which annotations are useful and which are not. To do this a questionnaire was constructed which was send to all ArchiMediaL members.

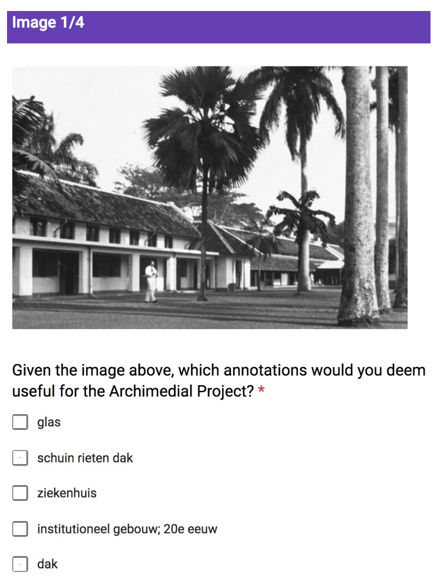

Fig. 2: ArchiMediaL questionnaire example question

This questionnaire consisted of the four objects used (figure 1) and their corresponding randomised list of annotations. These lists consisted of the verified annotations of the crowd mixed with the annotations made by the experts. This list was randomised to preclude any insight as to which annotation originated from a particular population. For every object the evaluators were asked two questions namely: "*Given the image above, which annotations would you deem useful for the Archimedial Project?*" and "*Given the same image and annotations, which do you think have been made by an expert?*". Figure 2 shows an example of the top part of the first page of this questionnaire.

The evaluators were first asked to indicate which annotations they deemed useful for the ArchiMediaL project. Afterwards they were asked to indicate which annotations they believed, to have originated from the expert population. To prevent an individual's opinion from influencing the results to the questionnaire in an oppressive way, majority voting was implemented as a filter. Thus an annotation was deemed useful only if at least 50% of the ArchiMediaL team agreed on its usefulness. The same was the case for the expert recognition part of the questionnaire.

4 Results

The final results of the questionnaire on expert recognition and usefulness are shown in figure 3 and figure 4 respectively. The corresponding dataset is available online[6].

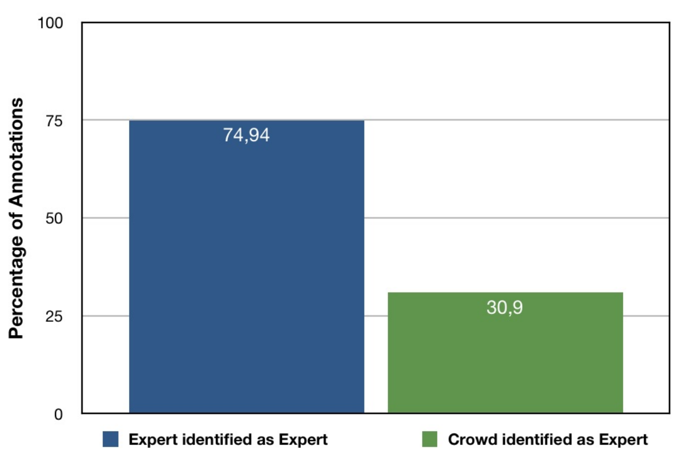

Fig. 3: ArchiMediaL questionnaire: Expert recognition results

[6] `https://figshare.com/articles/Experiment_Data_xlsx/6725129`

After the majority voting done by the ArchiMediaL members, 74,94% of all annotations created by the expert population were identified as such. This is in stark contrast to the 30,90% of crowd annotations, which were incorrectly identified as expert annotations.

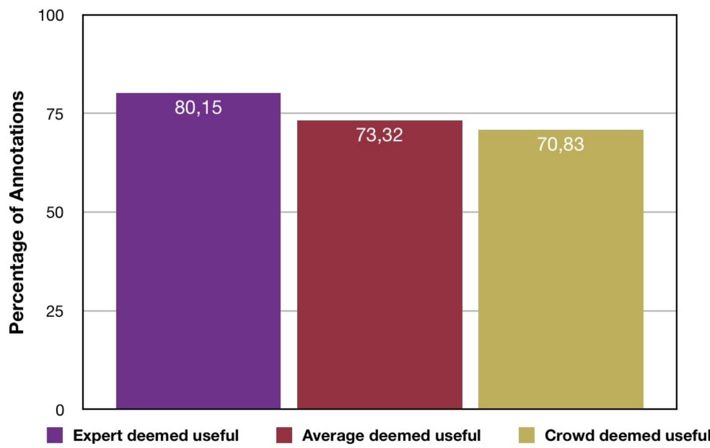

Fig. 4: ArchiMediaL questionnaire: Usefulness results

On the usefulness side of the questionnaire, there is only a minor difference between the expert and crowd populations. Of the annotations made by experts 80,15% were deemed useful, whereas the crowd was useful in 70,83% of the cases. This means that a total of 73,32% of all annotations provide meaningful data, which can be used in the ArchiMediaL project.

5 Discussion

With over 73% of all the gathered annotations being seen as an addition of meaningful data, it is clear that crowdsourcing architectural knowledge is a viable way of adding descriptive data for the ArchiMediaL project. When combined with the results found by Oosterman, et al. [13] we can safely say that crowdsourcing knowledge intensive tasks is viable. Whether this can be said for all tasks remains unclear due to the differing fields of knowledge.

While looking at the usefulness metric, the crowd and expert populations seem alike. This suggests that there is little reason to take the extra effort of locating and contacting experts when finding participants for such a crowdsourcing project. However, the fact that almost 75% of the annotations created by experts were identified as such gives reason to re-evaluate this statement. A difference in annotation quality between crowd annotations and expert annotations is implied. Experts provide higher quality annotations but the crowd's annotations are useful as well. This poses the question, does the difficulty of finding experts justify the gains from higher quality annotations? As well as the question, for which purpose can the annotations created by the crowd be used? While the allure of greater quality annotations is present for experts in the field, it might not be something which the main audience of linked open data requires. If an application only requires the simplest of annotations, crowdsourcing with non-experts is a fitting solution.

What can be derived from all this, is the fact that crowdsourcing architectural knowledge for the ArchiMediaL project is viable. But to increase the quality of the annotations, a system which prefers expert's annotations over non-expert annotations should be implemented. Furthermore, alterations to the (semi-) majority voting system for the crowd's annotations, should be made to accommodate for scaling of the system.

The conducted experiment was prone to some limitations, or things which can be improved upon. To provide better insight in the viability of crowdsourcing, the crowd should consist of more participants. This is also the case for the expert's side of the experiment, even though it is difficult to find them. Furthermore, the questionnaire which was send to the evaluators should include a wider array of questions pertaining the annotations. This could include things such as: "Rate the usefulness/quality of these annotations on a Likert scale.", or "Please provide insight into why you think this annotation was made by an expert.".

6 Conclusion

Even though there has been prior research into the topic of crowdsourcing, none of it specifically looked into architectural knowledge. This paper set out to fill this gap in knowledge by conducting a crowdsourcing experiment involving experts and non-experts alike. The results of this experiment have shown that crowdsourcing architectural knowledge is a viable option. While this is the case, one should still take into account the fact that experts seem to provide data of a higher quality, when compared to non-experts.

Possible future work in this field could be conducted by taking a look into the cost of experts compared to the increase in quality they provide.

7 References

1. Jett, J., Cole, T., Kinnamen, A., Kudeki, D., Han, M., & Szylowicz, C. (2017). Extending Legacy Metadata with Linked Open Data. International Conference on Dublin Core and Metadata Applications, 0, 29-38. Retrieved from: `http://dcpapers.dublincore.org/pubs/article/view/3853/2038`
2. Savage, N. (2012). Gaining wisdom from crowds. Commun. ACM 55, 3 (March 2012), 13-15. Retrieved from: `https://dl-acm-org.vu-nl.idm.oclc.org/citation.cfm?id=2093553`

3. Kittur, A., Smus, B., Khamkar, S., & Kraut, R.E. (2011). CrowdForge: Crowdsourcing Complex Work. UIST'11, 43-52. Retrieved from: http://smus.com/crowdforge/crowdforge-uist-11.pdf
4. Lux, M., Becker, J., & Krottmaier, H. (2003). Semantic Annotation and Retrieval of Digital Photos. CAiSE Short Paper Proceedings. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.2711&rep=rep1&type=pdf
5. Miller, E. (2005). An Introduction to the Resource Description Framework. Bulletin of the American Society for Information Science and Technology, 25-1,15-19. Retrieved from: https://onlinelibrary-wiley-com.vu-nl.idm.oclc.org/doi/full/10.1002/bult.105
6. De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., van Ossenbruggen, J., & Schreiber,G. (2012). Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. ESWC 2012: The Semantic Web: Research and Applications, 733-747 Retrieved from: https://link-springer-com.vu-nl.idm.oclc.org/chapter/10.1007%2F978-3-642-30284-8_56
7. Hollink, L., Schreiber, G., Wielemaker, J., & Wielinga, B. (2003). Semantic Annotation of Image Collections. Retrieved from: http://dare.ubvu.vu.nl/bitstream/handle/1871/24868/222149.pdf?sequence=2
8. Dijkshoorn, C., De Boer, V., Aroyo, A., & Schreiber, G. (2014). Accurator: Nichesourcing for cultural heritage. Human Computation, 1-1, 101-131. Retrieved from: https://arxiv.org/pdf/1709.09249.pdf
9. Dijkshoorn, C., Leyssen, M. H. R., Nottamkandath, A., Oosterman, J., Traub, M.C., Aroyo, L., Bozzon, A., Fokkink, W., Houben, G., Hövelmann, H., Jongma, L., Ossenbruggen, J., Schreiber, G., & Wielemaker, J. (2013). Personalized Nichesourcing: Acquisition of Qualitative Annotations from Niche Communities. Retrieved from: https://www.researchgate.net/profile/Wan_Fokkink/publication/261098150_Personalized_Nichesourcing_Acquisition_of_Qualitative_Annotations_from_Niche_Communities/links/5523ee940cf24f1609442f16.pdf
10. Howe, J. (2006). The Rise of Crowdsourcing. Wired Magazine, 14-06. Retrieved from: https://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf
11. Schenk, E., & Gittard, C. (2011). Towards a characterization of crowdsourcing practices. Journal of Innovation Economics & Management, 7, 93-107. Retrieved from: http://www.cairn.info/revue-journal-of-innovation-economics-2011-1-page-93htm
12. Erickson, L., Petrick, I., & Trauth, E.M. (2012). Hanging with the right crowd: Matching crowdsourcing need to crowd characteristics. AMCIS 2012 Proceedings, 3. Retrieved from: https://pdfs.semanticscholar.org/1633/b09cbe208c1a5e1562882147c5635f3e0375.pdf
13. Oosterman, J., Bozzon, A., Houben, G.J., Nottamkandath, A., Dijkshoorn, C., Aroyo, L., Leyssen, M.,& Traub, M. (2014). Crowd vs. Experts: Nich-

esourcing for Knowledge Intensive Tasks in Cultural Heritage. Retrieved from: https://vu.on.worldcat.org/oclc/6893714435
14. Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL Query Language for RDF. W3C Recommendation. Retrieved from: https://www.w3.org/TR/rdf-sparql-query/