ISEA2015
Proceedings of the 21st International Symposium on Electronic Art
ISSN: 2451-8611
ISBN: 978-1-910172-00-1

Identifying community resources using data mining, crowdsourcing, and networked co-curation

Daragh Byrne and Aisling Kelliher

Carnegie Mellon University
Pittsburgh, PA, USA
daraghb@andrew.cmu.edu, aislingk@andrew.cmu.edu

Abstract

The curatorial process is typically an expert led endeavor that requires extensive content review and judicious selection to assemble an archive of shared cultural value. Recent innovations in digital social curation open up new opportunities for non-expert participation in assembling collections, although challenges remain in terms of maintaining quality, straddling expert and amateur goals and integrating disparate and related efforts. In response, we present a flexible online web application designed to computationally support collective curatorial decision-making across diverse communities of interest. Findings from a 10-week deployment with a technology-arts community point to the utility of the system in accurately identifying and recommending useful content.

Keywords

Crowdsourcing, digital curation, network co-curation, community resources, art/science

Introduction

The act of curation is typically associated with expert identification and interpretation of critical cultural artifacts experienced in museums and galleries. The process of collecting, and explaining cultural archives presents implications for memory making [1], identify formation [2], and individual and institutional power [3]. Recent innovations in social computing, the digital humanities, and online curation have opened up new opportunities for expanding the remit, roles and activities of curators, communities and audiences.

Of particular current interest are social bookmarking platforms such as Delicious and Pinterest, which have given rise to the notion of online crowd curation [4]. These platforms serve to diversify the spectrum of individuals adopting curatorial practices, leading to the emergence of previously unknown arbiters of taste and cultural value. However, these platforms typically empower individuals to self-select and identify content of relevance and interest to them. While these amassed lists of content may have general or broad relevance to a community of interest, they are not typically authored explicitly for them. Additionally, these online environments lack collaborative features to allow a group to collectively identify resources of relevance and only allow informal opportunities for a community to highlight those resources which are of use and relevance to them.

Yet communities of interest regularly identify, disseminate and exchange online resources and such practices are visible and apparent in the regular interchange between community members in online social platforms such as Twitter. While these platforms are not intended as a breeding ground for curatorial practice, there is a wealth of information, as well as social metadata, that can be co-opted for such an endeavor.

In this work we explore the use of a mixed method approach combining crowdsourcing and automatic computational methods to augment and support traditional expert led judgments for digital resource curation. We propose mining Twitter to identify potential resources on an ongoing basis, the adoption of crowdsourcing techniques in combination with machine learning approaches to recommend resources for inclusion, and finally computer mediated review to aid in the organization of accepted content into a structured archive. Through this process we seek to distribute the curatorial effort across a diversity of community opinions, while creating value and reward for participation.

We present the design and implementation of our approach, developed in collaboration with a small and emerging art/science community with a strong interest in creating a shared repository of useful and diverse resources. Beginning with a review of related work, we next introduce our approach integrating community-led activities and algorithmic processes. Our system is implemented as an online web application and we describe findings from a 10-week study with fifty members of our collaborating community. We conclude our paper with a discussion of our results and details of ongoing and future work.

Background

There is a recognized need for curated community archives to enrich a shared understanding of the nature and practice of art/science integration [5, 6]. Several past efforts have explored the development of a reference space to coordinate this emerging interdisciplinary community. Notably, the late Stephen Wilson maintained an extensive online

repository of art/science projects, which was additionally compiled into several books [6, 7] while organizations such as Rhizome offers an online database of new media works and artist profiles. While valuable, these initiatives present some limitations. In particular several of them have been driven by a small number of individuals with a leadership role in their preparation and maintenance. Thus, the potential scale and scope of the archive is limited by the availability of their time. Additionally, as these archives grow, so does their complexity, placing significant burden on those gatekeepers to successfully sustain and maintain them.

Curation typically describes the expert led process of identifying, organizing, and explaining content of cultural or communal value. As a professional practice it is well understood within the context of a museum, gallery, or in the art world. It is however far less established in the digital domain and there are many challenges present in this nascent practice. Specifically, Botticelli notes that digital curation is marred by a lack of specificity as compared with its more traditional counterpart, owing to "lack of established standards and best practices" and that there are "significant gaps in [the needed] skill set relative to the demands of curating data collections" [8]. This motivates the development of new digital techniques for this context and in response, Sabharwal offers "*networked co-curation*" [9] as a means to decentralize the curatorial process and to collaboratively amass content of shared value on an ongoing basis. The heart of this is a democratized, decentralized and collaborative approach to digital curation. While it raises issues of inconsistency and quality, it provides mechanisms to enrich public discourse, identify emerging knowledge, and increase access to cultural volumes.

While distinction is often drawn between the traditional practice of curation and its social online counterparts, networked co-curation clearly overlaps with many online strategies for information management and connoisseurship found on the social web. While these 'lightweight shared spaces' may be more 'tastemaking' rather than actual 'authorial act' [10], the social web still has much to offer. In particular, it provides technical platforms whereby institutions can leverage 'amateur' interest and 'extend the reach, use, and usefulness of their own collections' [11].

Networked Co-Curation

Our approach combines human and machine decision making within an emerging form of curatorial practice. First, we propose a flexible technical solution for recommending prospective content blending content mining, content weighting and collective action. This technical approach is implemented as an online web delivered application and is readily adaptable to a variety of use-cases, domains and contexts of use. Secondly, we ground the design of our solution within a clear user-centered experience, with direct support for outreach, feedback, and distributed verification tasks.

Using our web application, we created a Twitter account and selected a variety of high-profile art/technology accounts to follow. On an hourly basis, the public timeline for this account was reviewed and any tweet with a URL was retained. In order to rank the collected content, we used two particular measures to determine the likelihood that each shared link was a good resource – the number of unique sources sharing the link and the total number of shares. Assuming that multiple individuals within the network shared a single link, there was increased value ascribed to it based on the explicit action to disseminate it. Where more sources of community content had distributed a single link, it was assumed that it had increased relevance for the community. Once the content was collected and ranked, the online web application then began the community review component. The system periodically contacted participating community members asking them to perform a targeted review of the discovered content. Community members were notified of the set of new review assignments by email, where they could click on any of the assignment links to bring them to the application login and subsequent review page.

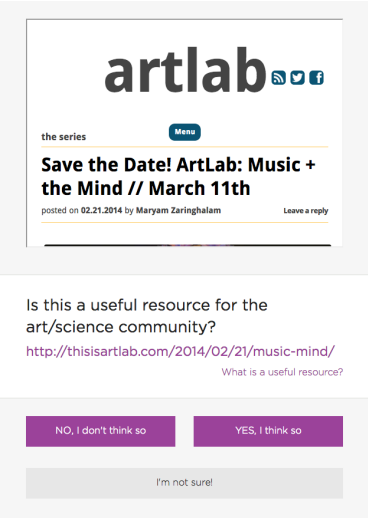

Figure 1. Review Interface

Figure 1 depicts an example review screen for an assignment for a community member helping to curate arts/science resources. Here, the link to be assessed was displayed within an iFrame, allowing the actual link content to be viewed alongside the rating functionality. The member could also chose to load the content in a new window if they wanted to examine the content in depth and in full context. Community members were asked to rate the content according to the review criteria developed by the curatorial community of interest (e.g. in Figure 2, the art/science community curators requested yes/no/unsure rankings on the 'usefulness' of the linked content, where additional instructions were provided via the "What is a

useful resource link?"). After completing an assigned review or set of reviews, the option to review additional content was presented. In this way, members were encouraged to contribute above and beyond the set number of reviews assigned to them each week.

The individual decisions by community members contributed to a combined score for the content. Recommendations for inclusion incremented the score by 1; 'unsure' did not alter the score; while an exclusion recommendation decremented the score by 1. The cumulative score was then used to provide a community driven recommendation to guide the community curator's ultimate decision. Community curators were expert members in leadership positions as defined by the participating community. A minimum of three community reviews was required before the rated link content was submitted to a community curator. Community curators arbitrated and guided the decision making process, relying on input from the community along with their own domain expertise and knowledge. To this end, the system separated community reviewed items into two categories for curator review: 'clear cut' decisions, and 'controversial' content. Community curators were notified daily with a list of the controversial community decisions that required their attention. The community curators review constituted a final decision on inclusion or exclusion to the resource database. However, the curator had to include a brief rationale for his or her decision to enhance decision making transparency and support ongoing community training.

System Implementation and Evaluation

Our system is currently implemented as part of the NSF sponsored XSEAD initiative (http://xsead.org), providing an online collaboration and presentation platform for a multidisciplinary community of art/science practitioners, critics, and researchers. Between February 2014 and March 2014, our system was deployed over a ten-week period with forty-eight participating community members and two community curators. These participants were opportunistically recruited through known networks of relevant individuals and through social media. During the ten-week study deployment, the system emailed community members on a Monday and Wednesday asking them to review and make judgments (yes, no, unsure) on three suggested art/science resource links. The two community curators received daily notifications asking them to verify or resolve disagreements based on community recommendations. Included items were made immediately available in online archive that was publically assessable both as a HTML page (see Fig. 2) and as an RSS feed. On Friday evenings, community members received an email thanking them for their participation, as well as giving them feedback on their decisions as verified or resolved by the curators. At the end of the ten-week period, a questionnaire was distributed to all community participants investigating their perception and experience of the task. In total, ten participants completed the distributed questionnaire.

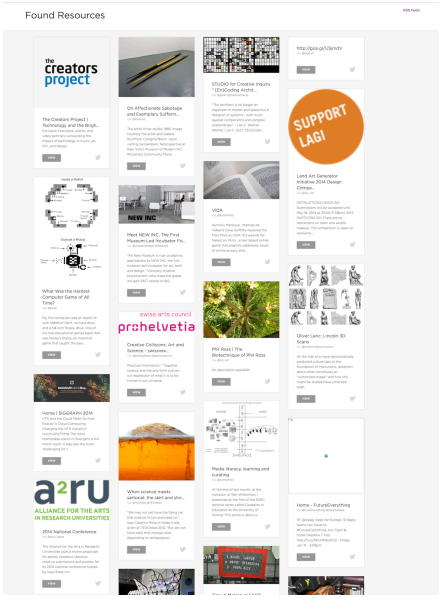

Figure 2. Archive of included resources

During the initial system rollout, 81 Twitter accounts with broad relevance to the designated art/science community (primarily US focused) were identified and followed by the project's account. These represented well regarded art/science institutions, initiatives, creative hubs, and individuals. These accounts shared a total of 13,366 distinct links representing an average of 196.14 per source. A total of 1202 assignments were sent to community members, requesting ratings for a total of 310 distinct art/science links. Of those 752 (62.56%) were completed by the community members.

Although assignments were generally completed within the same day as the email request, some users preferred to complete several sets of reviews at a later stage and en masse. Review assignments were typically completed within 4.52 days of the original request. This is reflected in the community members questionnaire responses, with one member noting *"I completed the reviews for the first email that got sent out and then sort of "binge completed" the rest of them up to a point."* A total of 1,766 ratings were completed during the study period for 811 unique links. While community members had the option to indicate if they were unsure of a rating decision they rarely did so. This constituted only 85 of the reviews made, or 4.8% of total ratings. 955 reviews recommending exclusion were made and 726 advocated for inclusion, with each user contributing an average of 33.4 ratings.

In total, the two community curators made 982 decisions on content links, selecting 234 links for inclusion and rejecting a further 748 links. While this represented a rela-

tively shallow inquiry into the total pool of socially shared links (7.4% of the 13,336 mined archive was reviewed), the process was effective in identifying useful content and supporting continuous growth of a shared archive. With a reasonably small pool of curators and community reviewers (50 total), an average of 23.4 resources were found per week. For judiciously selected content, this represents a solid growth rate.

Responses collected via the community member questionnaires highlighted some useful findings on the community feedback approach. Seven of the ten respondents agreed that the review was interesting to them and presented content relevant to the task. Five respondents indicated that the presented content was relevant to their field (and commented that it offered them an enriching perspective on art/science integration: "*I gained new perspective on science/art connections and found new resources for my own research."* However, participants indicated mixed views on the task understanding with responses well distributed with one participant remarking: *"I didn't really know who the end user of the platform was intended to be at first"*. Participants were similarly mixed in their perceived self-efficacy. While half of the respondents felt they were confident in their decisions, the other half were unsure or unconfident.

Discussion and Future work

In this work, we have explored techniques for digital curation to support the assembly of scalable archives that are responsive to emerging knowledge. It is important to remember that curation is an end-to-end process that encompasses content identification, synthesis, organization, preservation, explanation and communication. As such, the technical model and user-experience we describe in this work represents a preliminary step in a larger research endeavor.

We acknowledge that curation is a complex problem space that goes beyond content discovery and recommendation. It requires not just careful selection of content and meticulous organization, but also the inclusion of explanation so that the value and significance of the assembled content may be recognized. As such, curators must not only prepare the archive but also need to provide a context to the items they curate and synthesize structures so that value can be found at multiple levels. This offers a particular challenge for the next stage of our research.

Within this work, we have explored the opportunity to leverage social multimedia, social web mining and collective action in tandem to facilitate the co-curation of shared archives. This work also reflects how the solicitation of 'amateur' expertise and computational support can greatly assist expert led curatorial practice. Although, we have presented mechanisms to assist the pro/amateur curator, there are still several open questions on the best mechanisms to engage, motivate and continuously educate community members in co-curating shared archives.

As part of the ongoing work with this project, we are now preparing a significant revision to the online platform based on these findings. While we will continue to explore mechanisms for improved resource identification, as the archive begins to scale new challenges and opportunities emerge for research in this space. In particular, the complexities of managing, maintaining and organizing continually growing community archives will become increasingly important. As part of this ongoing work, we expect that new strategies for continued participation, motivation and engagement of community participants will become particularly important in fostering community stewardship for shared co-curated archives.

References

1. Jacques Derrida. *Archive Fever: A Freudian impression*. (Chicago: University of Chicago Press, 1996)
2. Erving Goffman. *The presentation of self in everyday life*. (New York: Doubleday, 1959)
3. Michel Foucault. *The order of things*. (New York: Random House, 1970)
4. Catherine Hall and Michael Zarro. (2012) "Social curation on the website Pinterest.com" in *Proceedings of the American Society for Information Science and Technology*. 49(1)
5. Clare Hooper, David Millard, Jill Fantauzzacoffin, Joseph. Kaye (2013) "Science vs. science: the complexities of interdisciplinary research", in *Proc. CHI EA 2013*. pp 2541-2544
6. Stephen Wilson. Information Arts. (Cambridge: MIT Press, 2002)
7. Stephen Wilson. Art + Science Now. (New York: Thames & Hudson, 2010)
8. Peter Botticelli, Bruce Fulton, Richard Pearce-Moses, Christine Szuter, and Pete Watters. (2011). "Educating Digital Curators: Challenges and Opportunities" in *International Journal of Digital Curation*, Vol. 6, No 2, pp146-164.
9. Arjun Sabharwal. (2012). "Networked co-curation in virtual museums: Digital humanities, history, and social media in the Toledo's Attic project" in *International Journal of Heritage in the Digital Era* 1(4), 587–610.
10. Susan Cairn and Danny Birchall, "Curating the Digital World: Past Preconceptions, Present Problems, Possible Futures", in *Proc. Museums and the Web 2013*, N. Proctor & R. Cherry (eds). Silver Spring, MD: Museums and the Web. Published February 6, 2013.
11. Melissa Terras. (2010) "Digital curiosities: resource creation via amateur digitization" in *Literary and Linguistic Computing*, Vol. 25, No. 4. P 425 - 438