

Designing a Crowdsourcing Tool to Analyze Relationships Among Jazz Musicians: The Case of Linked Jazz 52nd Street

Hilary K. Thorsen

Stanford University
Stanford, CA
thorsenh@stanford.edu

M. Cristina Pattuelli

Pratt Institute
New York, NY
mpattuel@pratt.edu

Abstract

We discuss the design of Linked Jazz 52nd Street, a crowdsourcing tool using linked data technology to build a social network of jazz musicians through the analysis of their personal and professional relationships.

Introduction

Archives often reveal a rich history of the social relationships of the community that created them. Increasingly, as they are digitized, the data they contain is made available to a wider audience, but is not necessarily easier to retrieve. It remains challenging to share information across repositories, establish relationships between materials, and give them context. Linked Open Data (LOD) is seen as a promising web technology allowing archives, libraries, and other cultural institutions to connect and reuse their data to establish new relationships and networks (Bizer, Heath, Berners-Lee 2009). While machine processing of data is often sufficient to build basic social networks, personal relationships have subtle nuances that only a human can discern. As Simperl (2012) suggests, employing both human and computational intelligence in LOD applications often produces better results.

The Linked Jazz project¹ serves as a case study for developing best practices using LOD to construct a network of personal and professional relationships of jazz musicians using both a human and machine approach. The first phase of the project relied on machine processing of data from interview transcripts to create a social network representing the jazz community.² The second phase revolves around human computation. In order to hone in on the more granular relationships that the text of the transcripts reveals, e.g. *close friend of*, *collaborated with*, or *mentor of*, we developed a crowdsourcing tool to leverage the

knowledge of jazz scholars to assist with the analysis of the relationships in the transcripts.

Linked Jazz 52nd Street

The crowdsourcing tool, named Linked Jazz 52nd Street,³ is a web-based application that asks contributors to classify the relationships between two jazz musicians by choosing from several relationship options. On the back-end, personal names from interview transcripts are paired with names mentioned in the interview. The interface presents segments of the transcript that mention the artist in question to allow users to see the context of the conversation and choose the best option to describe the relationship (figure 1). Outputs from the participants, converted to RDF statements, are fed back into the project's LOD dataset and semantically enhance the social network.

The design of the Linked Jazz 52nd Street application was informed by a survey of existing tools and a review of the literature on crowdsourcing. Several studies demonstrated that both extrinsic factors, such as monetary compensation and recognition, and intrinsic factors, such as contributing to the greater good and learning new skills, are strong motivators to engage in crowdsourcing projects, and the most successful projects support both types of motivation (Brabham, 2010; Causser, Tonra & Wallace, 2012). A number of features were included in the design of Linked Jazz 52nd Street in order to leverage motivational factors and create a tool that would encourage user participation. For example, an ego network visualization is built in real time while a progress bar fills indicating the contributor's progress on the task of assigning relationships, which gives them immediate visual feedback with the intent of making their work recognizable and transparent.

¹ <http://www.linkedjazz.org>

¹ <http://www.linkedjazz.org>

² See <http://www.linkedjazz.org> for papers and presentations about our tools.

³ <http://linkedjazz.org/52ndStreet/>

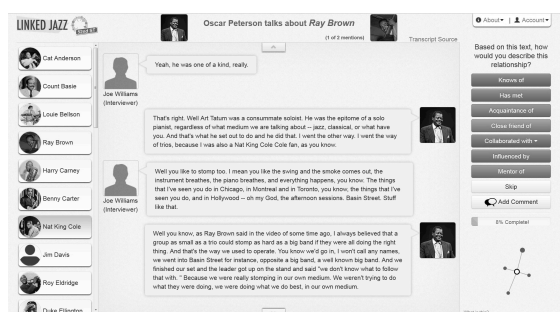


Figure 1: Linked Jazz 52nd Street relationship classifier interface

Challenges

After a first design iteration, we have been left with a number of open questions. When we first began working on Linked Jazz 52nd Street, crowdsourcing was just beginning to be used within the cultural heritage sector and most projects were in their initial stages. As crowdsourcing cultural heritage has recently grown in popularity, several open source tools are becoming available that we can relate to. For example, the Roy Rosenzweig Center for History and New Media's Scripto,⁴ and Zooniverse's Scribe⁵ used by the Bodleian Library's What's the Score⁶ and New York Public Library's Ensemble project.⁷ While we analyzed these applications to learn from their experience and avoid reinventing the wheel, neither of these tools meets all of our needs. They are designed for different purposes, such as transcribing entire documents. Instead, we require our users to analyze text and classify relationships, which present a higher level of complexity.

After we completed the development of a prototype of Linked Jazz 52nd Street, task-based usability testing was conducted. The results indicated that users generally had a positive experience with the tool; participants found it visually appealing, but were not fully engaged while using it and were unclear about how their contributions impacted the project as a whole. These findings confirmed the need to focus on the improvement of those areas that are critical for fostering motivation and user engagement. We are currently revising the interface of the tool with the goal of providing feedback on task progress after each action with visual and textual indicators to help keep users focused. We are redesigning the comment box with the goal of creating a space where people can contribute their personal knowledge and learn from their peers. The comments will be submitted in the form of structured data, such as dates, locations, and events. This information will then be converted into LOD to semantically enrich our network and enable new paths of connections to external datasets. In addition, users will be able to contribute content in a free-text format and conversational style to foster community

engagement and sharing. We are also considering incorporating reward systems as a mechanism of instant gratification through recognition of users' work. Projects, such as Transcribe Bentham⁸ and Old Weather⁹ employ badges and ranking systems to reward users whereas other projects have turned crowdsourcing tasks into games. Prestopnik and Crowston (2011) caution that games may interfere with expert users' motivation to participate. As a result, they incorporated several tasks that were not game-based into their application Citizen Sort, including rapid classification tools for experts.¹⁰ Because we envision that our target users will mainly be jazz experts and enthusiasts, we will test whether gamification elements would enhance their experience without being perceived as distracting and trite.

Future Work

Further user testing with representatives of the target audience of Linked Jazz 52nd Street is planned in order to assess this next design iteration as well as functionality and data quality. We aim to make the crowdsourcing tool fully effective for expanding and enriching the social network representing the jazz community and further linking it with open cultural datasets and discographies. We believe that our tool as well as the development methods can be transferred to other domains and will be useful for performing knowledge representation tasks in libraries, archives, and museums with collections ranging from correspondence to photographs, audio, and video. Crowdsourcing analysis and classification of relationships using LOD has the potential to reveal complex relationships between cultural heritage collections leading to increased understanding and exposure of these collections.

References

- Bizer, Christian, Heath, Tom, and Berners-Lee, Tim 2009. Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems* 5 (3): 1–22.
- Brabham, D. C. 2010. Moving the crowd at Threadless. *Information, Community, & Society*, 13(8): 1122–1145.
- Causar, T., Tonra, J. and Wallace, V. 2012. Transcription maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*. *Literary and Linguistic Computing*. 27(2): 119–137.
- Prestopnik, N. R., and Crowston K. 2011. Gaming for (citizen) science: Exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. 7th IEEE International Conference on eScience, Dec. 6–8, 2011.
- Simperl, Elena. 2012. Crowdsourcing semantic data management: challenges and opportunities. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*. New York, NY, USA: ACM.

⁴ <http://scripto.org/>

⁵ <http://github.com/zooniverse/Scribe>

⁶ <http://www.whats-the-score.org/>

⁷ <http://ensemble.nypl.org/>

⁸ <http://blogs.ucl.ac.uk/transcribe-bentham/>

⁹ <http://www.oldweather.org/>

¹⁰ <http://citizensort.org/>