

User Contributions on YouTube as Source of Knowledge for Classical Music

Ioannis Petros Samiotis, Andrea Mauri, Alessandro Bozzon, Christoph Lofi
i.p.samiotis@tudelft.nl, a.mauri@tudelft.nl, a.bozzon@tudelft.nl, c.lofi@tudelft.nl

Delft University of Technology

Keywords: user-generated content; youtube; classical music; knowledge bases

Knowledge Bases (KBs) are used to store and access both structured and unstructured information. Although music-related data is stored in popular and open KBs such as DBpedia (Lehmann et al. 2015) and YAGO (Suchanek et al. 2007), or in domain specific ones like MusicBrainz¹ and Discogs², the completeness and correctness of information varies greatly per entry (Galarraga et al. 2017; Zaveri et al. 2016). In this respect, classical music is no exception. Classical music is a music genre that naturally overlaps with cultural heritage, with thousands of artists throughout centuries and artworks with high complexity of information. The music piece itself has metadata and artifacts attached to it (e.g., score sheets), but also each individual performance brings its own metadata (like artists involved, time, place, special properties of that performance, etc.) and artifacts (e.g. recordings or videos).

In this study, we explore to what extent user-generated YouTube content can be used as a source for classical music knowledge extraction. We hypothesize that on YouTube, similarly to other types of Online Social Networks, knowledge generation and exploitation are taking place, due to the very nature of user-content-user interaction on YouTube (Wattenhofer et al. 2012). To do so, we collected 91,491 relevant videos, out of 184,019 retrieved³. While focusing on the comments, we distinguish between videos featuring works of famous or obscure composers. We picked 5 “famous” composers across all periods and 5 “obscure” composers based on Bachtrack list of 2017⁴. We observe clear differences in commenting behaviors between those two classes. For obscure composers, users have a higher tendency to comment on the composition (~60%) or the performance itself (~25%) compared to famous composers (~35% and ~19% respectively). Also, these users engage more frequently in discussions with the Original Poster of the video. In their discussions, they are more likely to express their opinion (~68%) than users who comment in videos of famous composers (~56%). We also witness that their tendency to make jokes related to the video content is also lower.

Our results suggest that especially for works of obscure composers, YouTube comments can indeed serve as a potential source of knowledge. Considering that knowledge bases tend to be less complete for obscure works, this is indeed encouraging: there is a high chance users have left comments on YouTube discussing the works of similar artists and linking that to the rest of their corpora or to works of other composers.

References

- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., ... & Bizer, C. (2015). DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2), 167-195.
- Suchanek, F. M., Kasneci, G., & Weikum, G. (2007, May). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web* (pp. 697-706). ACM.
- Galarraga, L., Razniewski, S., Amarilli, A., & Suchanek, F. M. (2017, February). Predicting completeness in knowledge bases. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 375-383). ACM.
- Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), 63-93. Wattenhofer, M., Wattenhofer, R., & Zhu, Z. (2012, May). The YouTube social network. In *Sixth International AAI Conference on Weblogs and Social Media*.

¹ <https://musicbrainz.org>

² <https://www.discogs.com/search/>

³ we preserved only videos where the full name of the composers was contained either in the title, or in one of the tags, or in the description.

⁴ <https://bachtrack.com/files/73896-Classical%20music%20statistics%202017-EN.pdf>