

Muzeeglot : annotation multilingue et multi-sources d'entités musicales à partir de représentations de genres musicaux

Elena V. Epure Guillaume Salha Félix Voituret

Marion Baranes Romain Hennequin

Deezer Research, Paris, France

research@deezer.com

RÉSUMÉ

Au sein de cette démonstration, nous présentons Muzeeglot, une interface web permettant de visualiser des espaces de représentations de genres musicaux provenant de sources variées et de langues différentes. Nous montrons l'efficacité de notre système à prédire automatiquement les genres correspondant à une entité musicale (titre, artiste, album...) selon une certaine source ou langue, étant données des annotations provenant de sources ou de langues différentes.

ABSTRACT

Muzeeglot : cross-lingual multi-source music item annotation from music genre embeddings

In this demonstration, we present Muzeeglot, a web interface providing a visualization of multi-sources and multilingual music genres embedding spaces. We demonstrate the ability of our system to automatically infer the genres annotations of a music entity (track, artist, album...) according to some source or language, based on annotations from different sources or languages.

MOTS-CLÉS : Genres Musicaux Multilingues, Annotation Automatique d'Entités Musicales, Espaces de Représentations Sémantiques.

KEYWORDS: Cross-lingual Music Genres, Music Entity Annotation, Semantic Embedding Spaces.

1 Introduction

Pour les plateformes de streaming musical, l'exploitation des données liées aux genres musicaux est essentielle (Mandel *et al.*, 2010; Schedl & Ferwerda, 2017). Ces genres peuvent en effet refléter et résumer les préférences musicales des utilisateurs, permettre d'améliorer l'organisation d'un catalogue musical ou encore être utilisés par des systèmes de recommandation pour proposer du contenu musical personnalisé. Toutefois, relier des genres musicaux à un titre ou à un artiste est une tâche difficile, en raison de la nature subjective de ces concepts. Des études ont montré qu'il existe une grande disparité dans la façon dont les gens associent les genres à la musique, selon leurs préférences, leurs connaissances musicologiques et leur culture (Sordo *et al.*, 2008; Lee & Downie, 2013).

Les plateformes de streaming musical collectent des annotations d'entités musicales (titres, artistes, albums...) par genre auprès de nombreux fournisseurs de musique, provenant du monde entier. Homogénéiser ces différentes annotations est crucial pour l'analyse de données et la recommandation, mais est particulièrement difficile (Hennequin *et al.*, 2018; Epure *et al.*, 2019). Outre les disparités culturelles et musicologiques mentionnées précédemment, des différences d'annotations entre les

sources peuvent également apparaître suite à l'utilisation de niveaux de détails inégaux (selon la spécialisation de la source), à un manque d'exhaustivité (bien qu'il existe des milliers de genres, certaines sources n'en utilisent qu'un petit sous-ensemble pour l'annotation) ou encore suite à des incohérences orthographiques entre les différentes sources (certains genres sont annotés sous des formes différents : *d&b* vs. *drum and bass*).

Au sein de cette démonstration, nous présentons Muzeeglot, un système qui, étant données les annotations d'entités musicales par genres selon de multiples sources (ces sources étant désignées comme des *taxonomies* par la suite) pouvant être de langues différentes, prédit automatiquement les annotations équivalentes au sein d'une taxonomie cible. La solution proposée est non-supervisée, et est fondée sur l'apprentissage d'espaces de représentations sémantiques et multilingues des genres musicaux. Muzeeglot est accessible publiquement sur la page GitHub de Deezer¹.

2 Solution et démonstration

Afin d'aboutir à de tels espaces de représentations, nous calculons tout d'abord les « *word embeddings* » des genres musicaux apparaissant dans les taxonomies sources et dans la taxonomie cible. Il s'agit de représentations vectorielles de mots dans un espace de faible dimension, apprises à partir de leurs cooccurrences au sein de grands corpus textuels (Pennington *et al.*, 2014; Grave *et al.*, 2018). Dans cette démonstration, nous utilisons la version multilingue de fastText (Mikolov *et al.*, 2018; Joulin *et al.*, 2018) pour la construction de ces représentations. Nous les consolidons afin qu'elles correspondent davantage à notre domaine d'application : la musique. En effet, selon fastText, *house* est plus proche de *building* que de *music*, et *jazz* est plus proche de *folk* que de *bop*. Pour pallier ces ambiguïtés, liées à l'utilisation de corpus d'entraînement très généraux, nous créons un graphe de connaissances des genres musicaux à partir de DBpedia (Auer *et al.*, 2007). Ce graphe relie les genres, de différentes langues, selon plusieurs types de relations (sous-genre, genre dérivé, fusion...). Via l'algorithme de *retrofitting* (Faruqui *et al.*, 2015), nous ajustons les représentations des genres musicaux en y incorporant l'information sémantique du graphe. Les vecteurs des genres musicaux connectés dans le graphe auront ainsi tendance à être plus proches dans l'espace final.

Muzeeglot est une interface web, présentant une visualisation en 3D de cet espace de représentations des genres musicaux. Les Figures 1 et 2 proposent un aperçu du rendu final. Le frontend de Muzeeglot est basé sur le framework VueJS, des composants UX Vuetify, et Plotly pour le rendu 3D avec WebGL. Le backend est développé en Python3 via FastAPI, et les données sont stockées dans un datastore Redis. Sur cette interface web, l'utilisateur a la possibilité de sélectionner :

- une entité musicale, tel que le groupe Pink Floyd au sein des Figures 1 et 2 ;
- plusieurs taxonomies sources, pour lesquelles les genres correspondants s'afficheront ;
- une taxonomie cible, pour lesquels Muzeeglot devra prédire les annotations équivalentes de genres pour l'entité musicale sélectionnée.

Un Top 10 des prédictions des genres musicaux de cette entité au sein de la taxonomie cible sera alors proposé (voir Figure 2). Ces prédictions seront calculées à partir de la *similarité cosinus* moyenne entre les représentations des genres de la taxonomie cible et les représentations des genres de l'entité musicale sélectionnée au sein des taxonomies sources. Les dix genres de la taxonomie cible ayant la plus forte similarité cosinus seront ceux prédits par le système. Une comparaison des prédictions avec les annotations réelles de la taxonomie cible sera également présentée.

1. <https://github.com/deezer/muzeeglot>

Lors de la démonstration de Muzeeglot, nous nous attarderons plus particulièrement sur le sous-problème de *l'annotation multilingue d'entités musicales*. Nous présenterons un jeu de données de plus de 63 000 entités musicales (titres, artistes, albums...), collectées depuis DBpedia et annotées de genres musicaux de six langues différentes : Anglais, Espagnol, Français, Japonais, Néerlandais et Tchèque. Nous mettrons en avant l'efficacité de notre système à inférer les genres d'une langue à partir d'autres. Notons qu'il ne s'agit pas simplement d'un problème de traduction littérale des genres. En effet, les annotations intègrent les interprétations de rédacteurs de cultures et de pays différents ; par ailleurs, certains genres ne sont tout simplement pas traduits (*heavy metal* ne devient pas *métal lourd* en Français). La Figure 2 présente par exemple les prédictions de genres musicaux en Français pour le groupe Pink Floyd, à partir de ses annotations en Anglais et en Japonais au sein de DBpedia.



FIGURE 1 – Interface de Muzeeglot. Visualisation des genres musicaux de Pink Floyd, selon deux taxonomies *sources* correspondant aux annotations en Anglais et en Japonais au sein de DBpedia.

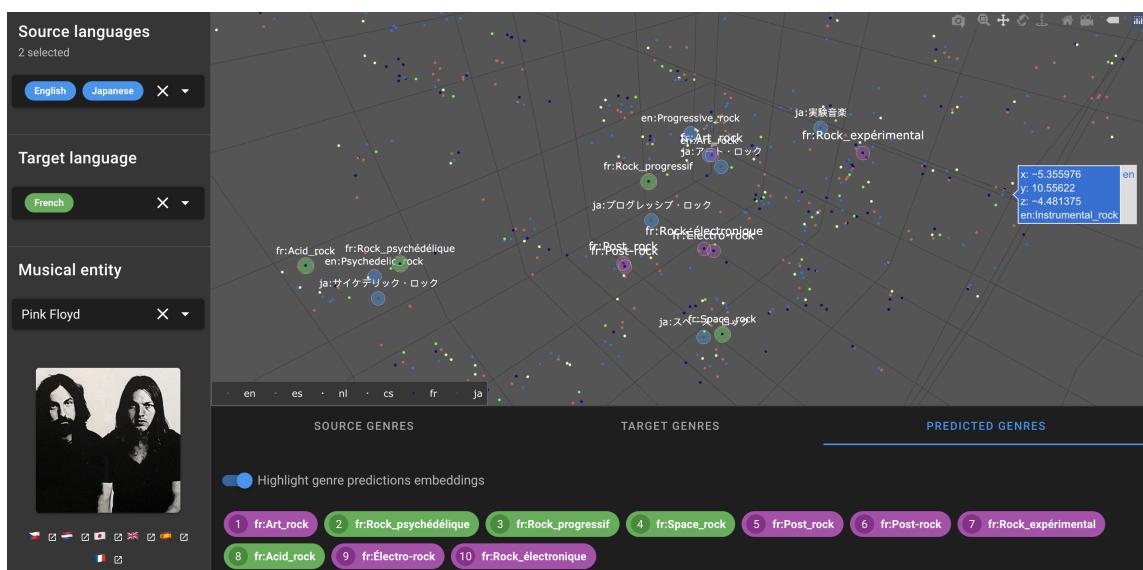


FIGURE 2 – Prédiction des genres musicaux de Pink Floyd en Français. Les prédictions représentées en vert coïncident avec les annotations de genres de Pink Floyd en Français au sein de DBpedia.

Références

- AUER S., BIZER C., KOBILAROV G., LEHMANN J., CYGANIAK R. & IVES Z. (2007). Dbpedia : A nucleus for a web of open data. In *The Semantic Web*. DOI : [10.1007/978-3-540-76298-0_52](https://doi.org/10.1007/978-3-540-76298-0_52).
- EPURE E. V., KHLIF A. & HENNEQUIN R. (2019). Leveraging knowledge bases and parallel annotations for music genre translation. In *Conference of the International Society of Music Information Retrieval, ISMIR 2019*. <http://archives.ismir.net/ismir2019/paper/000103.pdf>.
- FARUQUI M., DODGE J., JAUHAR S. K., DYER C., HOVY E. & SMITH N. A. (2015). Retrofitting word vectors to semantic lexicons. In *Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, NAACL-HLT 2015*. DOI : [10.3115/v1/N15-1184](https://doi.org/10.3115/v1/N15-1184).
- GRAVE E., BOJANOWSKI P., GUPTA P., JOULIN A. & MIKOLOV T. (2018). Learning word vectors for 157 languages. In *International Conference on Language Resources and Evaluation, LREC 2018*. <https://www.aclweb.org/anthology/L18-1550>.
- HENNEQUIN R., ROYO-LETELIER J. & MOUSSALLAM M. (2018). Audio based disambiguation of music genre tags. In *Conference of the International Society of Music Information Retrieval, ISMIR 2018*. <http://archives.ismir.net/ismir2018/paper/000163.pdf>.
- JOULIN A., BOJANOWSKI P., MIKOLOV T., JÉGOU H. & GRAVE E. (2018). Loss in translation : Learning bilingual word mapping with a retrieval criterion. In *Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*. DOI : [10.18653/v1/D18-1330](https://doi.org/10.18653/v1/D18-1330).
- LEE J. & DOWNIE J. (2013). K-pop genres : A cross-cultural exploration. In *Conference of the International Society on Music Information Retrieval, ISMIR 2013*. <http://archives.ismir.net/ismir2013/paper/000233.pdf>.
- MANDEL M., ECK D. & BENGIO Y. (2010). Learning tags that vary within a song. In *Conference of the International Society for Music Information Retrieval Conference, ISMIR 2010*. <http://archives.ismir.net/ismir2010/paper/000068.pdf>.
- MIKOLOV T., GRAVE E., BOJANOWSKI P., PUHRSCH C. & JOULIN A. (2018). Advances in pre-training distributed word representations. In *International Conference on Language Resources and Evaluation, LREC 2018*. <https://www.aclweb.org/anthology/L18-1008>.
- PENNINGTON J., SOCHER R. & MANNING C. (2014). Glove : Global vectors for word representation. In *Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*. DOI : [10.3115/v1/D14-1162](https://doi.org/10.3115/v1/D14-1162).
- SCHEDL M. & FERWERDA B. (2017). Large-scale analysis of group-specific music genre taste from collaborative tags. In *IEEE International Symposium on Multimedia, ISM 2017*. DOI : [10.1109/ISM.2017.95](https://doi.org/10.1109/ISM.2017.95).
- SORDO M., CELMA O., BLECH M. & GUAUS E. (2008). The Quest for Musical Genres : Do the Experts and the Wisdom of Crowds Agree ? In *Conference of the International Society on Music Information Retrieval, ISMIR 2008*. <http://archives.ismir.net/ismir2008/paper/000267.pdf>.