

Exploration d'une collection de chansons à partir d'une interface de visualisation basée sur une analyse des paroles

Rémy Kessler, Audrey Laplante, Dominic Forest

Université de Montréal
C.P. 6128, succursale Centre-ville, Montréal H3C 3J7, Canada
{remy.kessler, audrey.laplante, dominic.forest}@umontreal.ca

Résumé. Dans cet article, nous présentons une approche de fouille de textes ainsi qu'une interface de visualisation afin d'explorer une large collection de chansons françaises à partir des paroles. Dans un premier temps, nous collectons paroles et métadonnées de différentes sources sur le Web. Nous utilisons une approche combinant clustering et analyse sémantique latente afin d'identifier différentes thématiques et de déterminer différents descripteurs significatifs. Nous transformons par la suite le modèle afin d'obtenir une visualisation interactive permettant d'explorer la collection de chansons.

1 Introduction

Une grande quantité d'information textuelle sur la musique peut être extraite du Web. On y trouve notamment des données générées par les utilisateurs finaux (p. ex. : tags, critiques), des métadonnées (p. ex. : date de sortie, nom du parolier) et, enfin, les paroles des chansons. Ces informations peuvent souvent être moissonnées au moyen des API qu'offre un nombre croissant de services musicaux sur le Web, de même qu'avec l'aide d'outils développés par la communauté de chercheurs dans le domaine de la recherche d'information musicale. Cependant, les paroles de chansons ont reçu relativement peu d'attention de la part des développeurs des systèmes de repérage pour la musique. Or, la recherche à partir des thèmes abordés dans les paroles peut être pertinente dans certains contextes, pour les chercheurs s'intéressant à la musique populaire ou pour toute personne souhaitant trouver une musique pour un événement particulier (mariage, funérailles). Nous avons donc construit un système d'exploration d'une collection de chansons à partir des paroles. Après la constitution du corpus à partir de données provenant de diverses sources sur le Web, nous avons utilisé des algorithmes de fouille de textes pour en détecter les structures thématiques puis développé une interface de visualisation afin de naviguer dans la collection. Dans cet article, nous expliquons comment les données ont été recueillies et décrivons les différents traitements ayant été appliqués. Nous présentons également l'interface de visualisation qui en résulte.

2 Travaux connexes

Étant donné l'abondance d'informations musicales disponibles en accès libre sur le Web, il n'est pas surprenant de constater qu'un grand nombre de chercheurs ont développé des outils afin de collecter ces informations afin de faciliter le repérage de la musique. Logan et al.