

# Fouille de textes et visualisation de l'information pour l'exploitation de grands corpus documentaires en sciences humaines numériques

Dominic Forest, Ph.D.

École de bibliothéconomie et des sciences de l'information

Université de Montréal

✉ [dominic.forest@umontreal.ca](mailto:dominic.forest@umontreal.ca)

🏠 [www.ebsi.umontreal.ca](http://www.ebsi.umontreal.ca)

🐦 [www.twitter.com/dforest](https://www.twitter.com/dforest)

Usages et cultures numériques

2 mai 2014

# Contexte

- Avec l'essor des technologies numériques, les pratiques de recherche dans les sciences humaines sont en constante mutation.
- Les chercheurs tentent de tirer profit des avantages qu'offre la documentation numérique.
- La quantité d'informations disponibles en format numérique soulève d'importants enjeux auxquels aucune solution définitive n'a été proposée.

# Objectif

- Développer des stratégies (outils et méthodes) de fouille de textes (*text analytics*) et de visualisation pour assister des tâches d'extraction et d'organisation d'informations à partir de gros corpus documentaires en sciences humaines
- Tâches :
  - Analyse conceptuelle
  - Analyse thématique
  - Attribution d'auteur
  - Datation
  - Description documentaire
  - Fouille d'opinions
- Approche centrée sur l'utilisateur

# Méthodes

- Algorithmes d'apprentissage-machine supervisés et non supervisés
  - Catégorisation
  - Classification (*clustering*)
- Prétraitement
- Transformation (description) numérique des documents
- Extraction de motifs
- Visualisation
- Évaluation



# Projets en cours

- *Application de techniques de fouille de textes et de visualisation de l'information pour l'exploitation de grands corpus documentaires en sciences humaines numériques (CRSH, 2014-2018)*
- *Comprendre le domaine de l'environnement textuellement et linguistiquement (CRSH, 2013-2018, M.-C. L'Homme)*
- *L'indexation de musique à grande échelle : exploration du potentiel de nouvelles méthodes d'indexation pour accroître la visibilité de la chanson francophone sur le Web (CRSH, 2012-2014, A. Laplante)*