

Dominic Forest
Université de Montréal, Montréal, Québec, Canada

Valérie Rioux
Université de Montréal, Montréal, Québec, Canada

Sabine Mas
Université de Montréal, Montréal, Québec, Canada

Cartographier l'évolution de l'archivistique québécoise : la revue *Archives* (1994-2016) analysée au moyen de la fouille de textes

1. Introduction

Publiée par l'Association des Archivistes du Québec (AAQ) depuis 1969, la revue *Archives* contribue à faire rayonner l'expertise professionnelle et scientifique de ses membres dans toute la francophonie. À cet égard, elle rend disponible sur son site Web le contenu des articles publiés depuis 1994. Ces articles témoignent des nombreux changements ayant eu cours dans le paysage archivistique québécois et international et rendent compte de ses différentes relations avec les disciplines des sciences de l'information.

2. Problématique et objectif

À l'occasion du 50^e anniversaire de l'AAQ, nous proposons une analyse du contenu de la revue *Archives* à partir d'une approche originale basée sur des techniques de fouille de textes. Ces dernières permettent de traiter de grandes quantités de données afin d'en extraire automatiquement certaines caractéristiques. L'intérêt pour ces techniques s'est fortement développé depuis une dizaine d'années. Ainsi Graham et al. (2015) ont clairement démontré que l'utilisation des techniques de modélisation thématique basée sur des techniques de fouille de textes permet d'enrichir les perspectives de recherche. En outre, la cartographie thématique basée sur des techniques de fouille de textes a fait l'objet de plusieurs initiatives dans plusieurs domaines (Kim et al., 2008 ; Ravikumar et al., 2015 ; Moro et al., 2015).

La revue *Archives* a été l'objet de quelques analyses par le passé. Louise Gagnon-Arguin a publié à trois reprises les résultats d'une analyse du contenu de la revue *Archives* respectivement en 1988, 1994 et 2009. Dans le cadre d'un mémoire de maîtrise en sciences de l'information, déposé en 1989, Louis-René Dessureault s'est intéressé à l'état du discours de la profession d'archiviste au Québec sur la base d'une analyse bibliométrique de la revue *Archives*. Les travaux de Gagnon-Arguin et de Dessureault reposent toutefois sur une démarche entièrement manuelle.

Notre contribution offrira l'occasion de revenir sur les thématiques abordées dans la revue depuis les vingt dernières années en rendant compte des changements survenus sur

les objets de recherches et sur la profession archivistique. Enfin, notre communication visera aussi à mettre de l'avant les techniques de fouille de textes permettant de cartographier automatiquement l'évolution thématique d'un corpus et d'identifier certaines relations entre les chercheurs et les différents volets d'une discipline scientifique à partir d'une collection de données textuelles.

3. Méthodologie

La première étape de ce projet a consisté à recueillir les documents que nous souhaitons soumettre au processus d'analyse. Notre corpus est composé des volumes 26 (1994-1995) à 46 (2016). Nous avons retenu les articles scientifiques et professionnels (bilan d'expériences, mémoires présentés au gouvernement, etc.), laissant ainsi de côté les comptes rendus de lecture et les notes de la rédaction. Le corpus ainsi constitué est composé de 238 documents. Le corpus a été divisé en quatre périodes : 1995-1999, 2000-2004, 2005-2009 et 2010-2016 (tableau 1).

Période	Nombre d'articles
1995-1999	61
2000-2004	55
2005-2009	57
2010-2016	65

Tableau 1. Répartition des articles par période.

Par la suite, différents traitements statistiques (calcul de fréquences, pondération par tf-idf, etc.) et linguistiques (normalisation des formes, suppression de mots fonctionnels, etc.) ont été appliqués au corpus de documents afin de pouvoir le soumettre aux algorithmes de fouille. Ces derniers nous ont permis d'identifier plusieurs caractéristiques déterminantes, comme l'évolution des thèmes abordés en fonction de la période chronologique ainsi que les proximités thématiques reliant les auteurs. Voici les statistiques de base du corpus :

Données brutes : 37 929 formes et 1 791 514 occurrences

Données normalisées et filtrées : 20 676 formes et 1 785 420 occurrences

Période	Nombre de formes	Nombre d'occurrences
1995-1999	8 761	432 196
2000-2004	9 052	408 828
2005-2009	11 504	533 593
2010-2016	10 181	410 803

Tableau 2. Fréquences des formes et des occurrences normalisées et filtrées par période

nombre restreint d'auteurs. Quelques observations peuvent ressortir vis-à-vis ces résultats, notamment en prenant en compte le contexte de publication de la revue (membres du comité éditorial, numéros spéciaux, etc.) et la période de publication en lien avec l'émergence du Web et la publication des RDDA. La figure 2 présente un échantillon des résultats obtenus (pour la période 2000-2004).

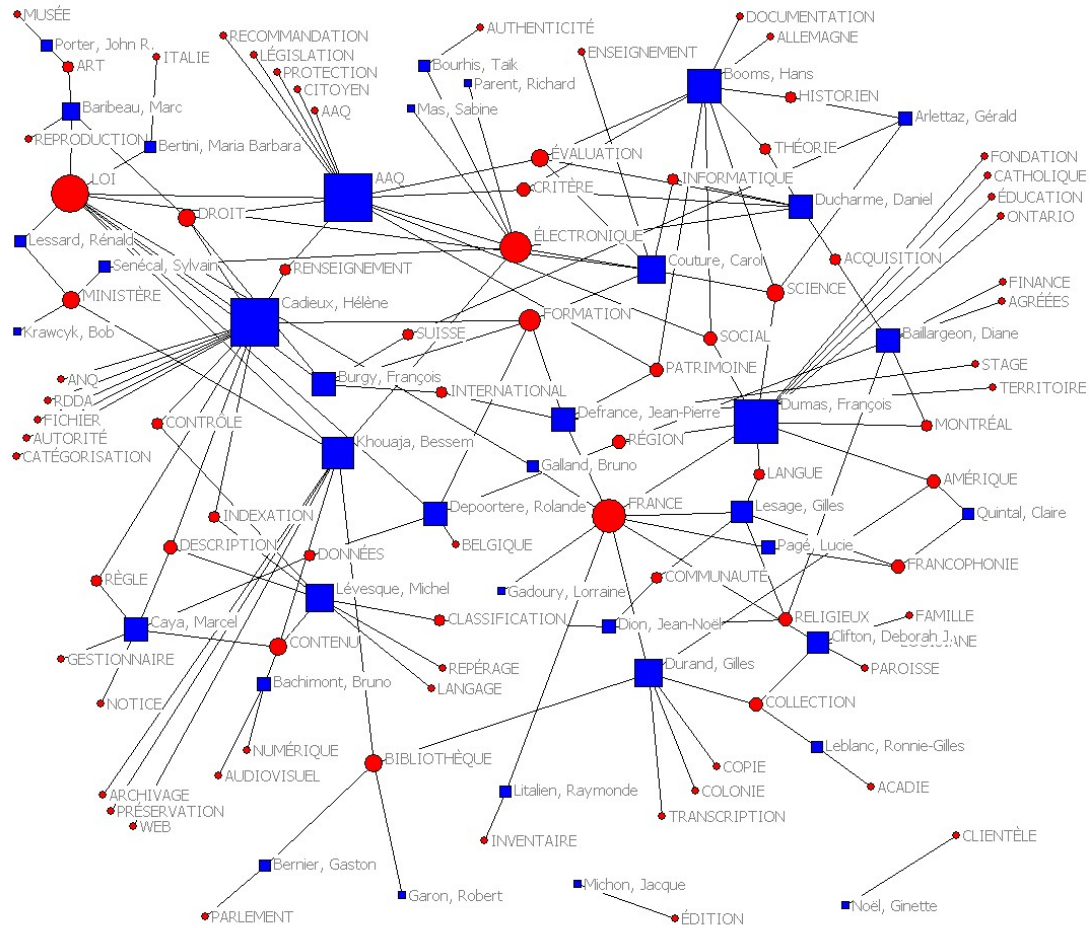


Figure 2. Les principales thématiques et les auteurs de la période 2000-2004.

Dans le cadre de nos expérimentations, il a été possible de quantifier l'évolution de la fréquence des mots-clés en fonction de la période de publication (années). Bien que la revue ait publié moins de numéros dans les dernières années, il est tout de même possible de remarquer les tendances thématiques qu'a prises la revue au courant des 20 dernières années. Alors que l'intérêt pour la défense de la profession et de la gestion des documents électroniques ne se dément pas d'une période à l'autre, il est intéressant de remarquer une augmentation de l'attention portée aux enjeux touchant particulièrement l'usager des archives. Par ailleurs, certains choix éditoriaux semblent avoir accordé une plus grande place aux archives en lien avec la production artistique dans les dernières années. Enfin, les mutations au sein des institutions et des techniques sont également reflétées dans les résultats que nous avons obtenus.

Lors de notre communication, nous insisterons sur notre méthodologie et les techniques utilisées dans le cadre de nos expérimentations. En outre, nous présenterons en détail les résultats que nous avons obtenus. Nous accorderons une importance particulière à la représentation visuelle de nos résultats afin de faciliter la compréhension et le partage de nos recherches à l'ensemble de la communauté canadienne des sciences de l'information. Enfin, nous discuterons des limites de notre projet de recherche et aborderons les travaux à poursuivre sur notre objet d'étude.

Références

Dessureault, L-R. (1989). L'état du discours de la profession d'archiviste au Québec: une analyse de contenu de la revue *Archives*. Montréal: Université de Montréal.

Gagnon-Arguin, L. (1988). Les 20 ans de la revue *Archives* : analyse des articles et des auteurs de 1969 à 1988. *Archives*, 20(1), 3-29.

Gagnon-Arguin, L. (1994). La revue *Archives* depuis 1988 : étude de son évolution. *Archives*, 25(3), 3-22.

Gagnon-Arguin, L. (2009-2010). La revue *Archives* à travers ses vingt dernières années d'existence, 1988-2008. *Archives*, 41(2), 22-44.

Graham, S., Milligan, I., & Weingart, S. (2015). *Exploring Big Historical Data: The Historian's Macroscope*. Londres, Royaume-Uni: Imperial College Press.

Kim, H., & Lee, J. Y. (2008). Exploring the emerging intellectual structure of archival studies using text mining: 2001—2004. *Journal of Information Science*, 34(3). doi:10.1177/0165551507086260

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 42(3), 1314-1324.

Ravikumar, S., Agrahari, A., & Singh, S. N. (2015). Mapping the intellectual structure of scientometrics: a co-word analysis of the journal *Scientometrics* (2005–2010). *Scientometrics*, 102(1), 929-955. doi:10.1007/s11192-014-1402-8