

Atelier

Des outils conviviaux pour la recherche en humanités numériques

Que faites-vous si votre ordinateur personnel n'est plus suffisant pour vos besoins de recherche? Avez-vous besoin d'exécuter des analyses Web successivement pendant des mois? Certains outils en humanités numériques ne sont disponibles qu'en services Web et beaucoup fonctionnent bien quand ils sont autohébergés dans des nuages. Dans cet atelier, nous examinerons certains outils uniquement Web, ainsi que la manière d'héberger les vôtres dans un environnement infonuagique. Après une brève introduction à quelques outils et ressources, nous travaillerons sur quelques exemples, une analyse de texte et une capture Web avec visualisation.

Ressources

Voyons d'abord comment trouver les outils pour la recherche en humanités numériques. En fait, c'est simple : faites une recherche avec votre moteur de recherche préféré.

Il y a plusieurs listes de ces outils dont beaucoup se trouvent sur des sites de bibliothèques universitaires ou de programmes d'études en sciences humaines et sociales. Souvent, ces listes mentionnent des listes de listes. Vous trouverez rapidement de nombreuses listes, sans trop d'efforts.

- [TAPoR](#)
 - Ce portail a été reconstruit en 2018 et en est à sa troisième version. Initialement conçu pour l'analyse de texte, l'incorporation de la grande liste DiRT y a ajouté des outils pour d'autres domaines comme des systèmes d'information géographique, le traitement photo/vidéo/audio, etc.
- [Digital Humanities Tools](#)
 - Préparée et maintenue par Alan Liu, cette liste comprend surtout des outils gratuits (ou presque) dont la plupart sont en ligne ou peuvent être installés sur un ordinateur personnel sans devoir passer par un serveur. Autrement dit, c'est pour tout le monde!

En faisant une recherche Web avec *social sciences tools*, on trouve les listes suivantes:

- [SAGE Ocean research tools directory](#)
 - Cette liste se trouve dans le portail de recherche SAGE Ocean de la compagnie SAGE Publishing.
- [Data analysis tools & training](#)
 - Cette liste est maintenue par la bibliothèque de Bodley de l'Université d'Oxford.

À première vue, il peut sembler qu'il y ait moins de listes de ressources pour les sciences sociales que pour les sciences humaines, mais ce n'est que pour des raisons historiques. Néanmoins, c'est quand un chercheur en géographie m'a demandé d'installer les outils suivants dans un nuage de Calcul Canada que je les ai découverts :

- [médialab SciencesPo Outils](#)
 - Ce groupe fondé par Bruno Latour écrit beaucoup de code pour des outils libres (*open tools*), ce qui est merveilleux! C'est par médialab que j'ai découvert :
- [DMI Tools](#)
 - Le groupe de recherche Digital Methods Initiative offre plusieurs outils en ligne qu'on peut utiliser à partir de son site Web. Dans certains cas, il n'est pas nécessaire d'être membre pour y avoir accès.

Il ne se fait pas beaucoup d'analyse de texte en sciences sociales; on fait plutôt des travaux comme des *cartographies d'influence* ou des *cartographies des controverses*. Certains outils peuvent cependant servir toutes les spécialités, je pense à Gephi, Voyant, OpenRefine, RAWGraphs, etc.

Exemples

Les trois phases de base de la recherche sont la collecte, l'analyse et la publication/présentation. Certes, il ne s'agit que de quelques-unes des étapes de tout processus de recherche, mais ça suffit pour découvrir les outils qui sont disponibles. Je vais vous présenter deux exemples : un exemple réel tiré de la recherche en sciences sociales et un exemple fictif d'analyse de texte.

Exemple réel : Déchets électroniques transfrontières

Un projet de recherche pluriannuel dirigé par [Josh Lepawsky](#), professeur de géographie à l'Université Memorial de Terre-Neuve, a produit un livre physique et un livre en ligne

avec des outils Web. Le livre physique a été publié par MIT Press en 2018 sous le titre [Reassembling Rubbish](#) et voici le livre en ligne disponible sur la plateforme Scalar :

- [Reassembling Rubbish](#)

[Scalar](#) est une plateforme numérique de publications savantes bien connue. Comme vous pouvez le constater, le résultat est proche d'un livre physique avec des chapitres, des index, des notes en bas de page, etc. Ceci est un exemple d'ouvrage complémentaire. La plateforme ne produit toutefois pas que du contenu sous forme de livres, mais aussi des expositions numériques, même si on reconnaît Scalar dans l'interface graphique.

Josh s'est adressé à Calcul Canada en 2016 pour de l'aide avec les outils Web dont il avait besoin pour son projet. L'Université Memorial avait contribué en hébergeant des fichiers média que Josh avait déjà assemblés sur Scalar en vue de publier. Il avait par contre d'autres besoins que l'université ne pouvait pas satisfaire, en particulier l'hébergement de Hyphe, une application infonuagique, comme moteur de collecte et d'analyse de données :

- [Hyphe](#)

C'est un *scraper* du Web, mais il est conçu pour la recherche en sciences sociales; il permet la curation et la conservation d'ensembles de pages Web de manière ordonnée. À ce moment, Hyphe fonctionnait sous Linux, ce que Josh n'avait pas, et il devait pouvoir lancer des tâches sur son propre ordinateur de façon indépendante. Nous lui avons fourni des ressources dans l'environnement infonuagique de Calcul Canada, puis lui avons construit une instance virtuelle et installé Hyphe. C'est plus facile aujourd'hui parce que Hyphe est maintenant disponible sous la forme d'un conteneur qui fonctionne avec Docker. Il reste que le fait de travailler dans un nuage a l'avantage de laisser plus d'indépendance.

Avant de parler de l'outil, voyons un des résultats du travail de Josh :

- Cartographie des controverses sur le mouvement transfrontières des déchets électroniques; site [Transboundary Movement of Electronic Waste: mapping a controversy](#)
- [Fichier source pour le graphique](#)

Le corpus extrait à l'aide de Hyphe a été raffiné et examiné avec Gephi. Pour produire ensuite une vue présentable du réseau, ils ont utilisé une extension Gephi qui permet l'affichage interactif dans un navigateur Web :

- [Sigma Exporter](#)

Voici quelques guides d'utilisation pour Hyphe :

- [document de recherche préparé par les développeurs](#)
- [Un cas d'usage de l'outil](#)

Avec seulement quelques pages initiales suggérées dans ce second guide, on peut voir Hyphe en action.

Après seulement quelques minutes d'exploration, Hyphe produit des résultats finaux qui demandent beaucoup de travail à raffiner et extraire de l'information utile.

Exemple fictif : Analyse de texte

Je vais maintenant importer un texte de *ibiblio* (Project Gutenberg) que je vais soumettre à quelques outils Voyant et afficher les résultats dans un projet sur le site Web du Collaboratoire scientifique des écrits du Canada (CSÉC). En accompagnement à la liste TAPoR, l'équipe de TAPoR a aussi produit ce site qui est une collection de méthodes et de techniques pour l'analyse de textes.

- [Methodica Commons](#)

Parmi leurs ressources, on trouve une grande collection de recettes pour réaliser des analyses comme l'analyse de sentiments, d'auteurité d'un texte, etc. En cherchant le terme *Voyant*, j'ai obtenu une liste de recettes et choisi [On Liberty](#) de John Stuart Mill.

- [Voyant](#)

Outils à regarder : Termes, Cirrus, Liens, Tendances, Contextes, Collocations, Documents

Fonctions intéressantes : Options => Mots d'arrêt, Exporter => Exporter une vue, Exporter la visualisation (outils graphiques), Exporter => Exporter une vue, Exporter les données actuelles (tableaux)

Que feriez-vous si vous aviez du matériel numérique d'un projet de recherche ou d'une archive et que vous vouliez le présenter en ligne? Si votre intention est de produire quelque chose qui ressemble à un livre, Scalar serait une option. Vous pourriez aussi considérer une plateforme comme [Omeka](#) pour les présentations/collections ou une plateforme se rapprochant plus d'un dépôt de données, comme [Islandora](#). Connaître ces plateformes et apprendre à les utiliser est l'objet d'au moins un autre atelier, mais je

vais vous donner un avant-goût et vous montrer comment vous familiariser avec l'une d'elles, Islandora.

Le [Collaboratoire scientifique des écrits du Canada](#) (CSÉC) est une infrastructure pour la recherche sur la littérature canadienne et la littérature au sujet du Canada. Le Collaboratoire veut relever les défis de la technologie numérique et explorer ses possibilités. Par la diversité des projets actifs, on constate qu'il s'agit de recherche littéraire dans le sens large qui inclut du matériel numérique d'origines et de sources plus traditionnelles. Explorez Collaboratoire => CWRC Commons => CWRC Videos.

Islandora est la technologie sous-jacente; elle est très employée pour les collections numériques des bibliothèques. Le dépôt source en arrière-plan est sur la plateforme [Fedora](#), alors que [Drupal](#) sert à l'avant-plan pour présenter le matériel. Le rôle d'Islandora est de rassembler un groupe d'extensions Drupal qui permettent les fonctionnalités CRUD à l'arrière-plan, incluant la gestion des schémas de métadonnées, des types d'objets, des collections d'objets, etc. en plus de l'indexation Solr et triplestore. CRUD est un acronyme pour *create*, *read*, *update*, *delete*, soit toutes les fonctions pour gérer le cycle de vie des objets dans le dépôt. Cette terminologie vous est peut-être inconnue, mais ne vous en faites pas, il y a un autre atelier pour clarifier cet aspect. Il faut toutefois mentionner deux choses importantes en rapport avec le Collaboratoire :

- Il s'agit d'un site qui est hébergé, ce qui veut dire que vous pouvez travailler sans vous soucier de la gestion du système d'exploitation, de l'application et de la sauvegarde de vos données. Ce type de plateforme présente un avantage considérable pour l'intégrité de vos données et la flexibilité de vos métadonnées, en plus de prendre en charge tous les détails techniques que vous ne connaissez peut-être pas.
- La puissance de Drupal peut, dans une large mesure (entre guillemets), être utilisée pour développer des présentations. Contrairement à Islandora, vous n'avez pas de restriction comme d'obtenir un résultat ayant l'apparence d'un dépôt moderne. Il faut par contre porter attention à comment l'accès se fait aux objets et aux données en arrière-plan pour afficher le résultat à l'avant-plan. Cependant, il est très possible de développer de belles expositions numériques.

Comme nous allons le voir, un autre avantage de cette plateforme est qu'elle permet l'intégration d'outils comme Voyant et le CWRC-Writer du CSÉC.

Autres ressources

- [UofT Collections](#) - collections de l'Université de Toronto avec des exemples d'utilisation d'Islandora pour la présentation de collections
- [NYAM Collections](#) - collections de la New York Academy of Medicine avec d'autres collections Islandora.