

Les humanités numériques, une science "plug and play" ?

Antoine Blanchard, Elifsu Sabuncu

► **To cite this version:**

Antoine Blanchard, Elifsu Sabuncu. Les humanités numériques, une science "plug and play" ?. Valérie Carayol; Franc Morandi. Le Tournant numérique des sciences humaines et sociales, Maison des sciences de l'homme d'Aquitaine, 2015, 978-2-85892-441-7. halshs-01248081

HAL Id: halshs-01248081

<https://halshs.archives-ouvertes.fr/halshs-01248081>

Submitted on 23 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les humanités numériques, une science “plug and play” ?

Antoine Blanchard et Elifsu Sabuncu

antoine@deuxieme-labo.fr et elifsu@deuxieme-labo.fr

Deuxième labo, 43 bd Auguste Blanqui, 75013 Paris

Comme le constatait en 2011 le “Manifeste des digital humanities” (sur lequel nous reviendrons par la suite), le champ des humanités numériques s’est constitué ces dernières années autour de pratiques, d’outils et d’objets transversaux divers. Le manifeste cite pêle-mêle “encodage de sources textuelles, systèmes d’information géographique, lexicométrie, numérisation du patrimoine culturel, scientifique et technique, cartographie du web, fouille de données, 3D, archives orales, arts et littératures numériques et hypermédiatiques” (collectif, 2011). Nous avons choisi de détailler quatre de ces concepts clés qui nous tiennent à cœur :

- la numérisation
- la participation
- la sémantique
- l’interopérabilité.

Une fois ces notions introduites, nous examinerons si la profusion d’outils “prêts à l’emploi”, doublée d’une communauté de pratiques très active favorisant la reproduction, ne risquent pas d’entraîner les humanités numériques vers un “prêt à penser”. Nous montrerons alors par un retour d’expérience à quel point la construction de nouveaux outils numériques peut elle-même être un moteur heuristique de la pensée en sciences humaines et sociales.

Quatre concepts clés des humanités numériques

#numérisation

La numérisation peut concerner tout type de document ancien mais quand on s’arrête sur la numérisation du patrimoine écrit, il est difficile de ne pas penser à Google Books. La firme américaine s’est lancée en 2004 dans une grande campagne de numérisation des fonds des plus grandes bibliothèques publiques et

universitaires. Ils sont allés plus loin en 2010 avec l'outil Ngram Viewer qui permet de rechercher dans une partie importante de la base Google Books (représentant 4% de tous les livres jamais publiés) des mots ou paires de mots, afin d'observer leur fréquence d'apparition dans ce corpus et leur évolution au fil du temps (Michel, 2011). Voilà qui intéresse doublement les humanités numériques : côté pile, les chercheurs qui ont proposé cet outil voulaient lancer par là un nouveau terrain de recherche appelé la *culturomics*, potentiellement fertile pour les humanités numériques (études textuelles, sociologie de la culture, histoire des idées...) comme pour tous les scientifiques. Quel géologue ou chimiste n'a pas souhaité suivre la popularité de son objet de recherche au fil du temps ? Côté face, voilà un outil qui s'appuie sur des connaissances en lettres et en sciences humaines ! Les paléographes savent par exemple que la langue évolue et que l'on peut parfois se tromper en cherchant dans un tel corpus. Jusqu'en 1800, la lettre "s" était représentée par le "s long" qui ressemble au "f", donc en cherchant "prefque" on s'aperçoit qu'elle est la forme dominante avant d'être remplacée par "presque" aux alentours de 1790. Voilà le type de pièges dans lesquels il ne faut pas tomber et où les archivistes et paléographes peuvent apporter de la valeur. La numérisation par et pour les spécialistes des textes anciens et des grands corpus (qu'on appellerait aujourd'hui "big data").

#participation

Un autre mot-clé pour les humanités numériques, c'est la participation. Il en va ainsi du projet PhotosNormandie mené par des historiens du débarquement et de la Seconde Guerre Mondiale en Normandie, professionnels ou amateurs. Il consiste à créer les descriptions les plus complètes possible de jeux de photographies dans le domaine public (malheureusement des clichés américains et anglo-saxons puisqu'en France le domaine public s'applique moins largement). Elles sont mises à disposition sur la plateforme de partage de photos Flickr, où chacun y va de ses indices pour identifier un lieu ou un personnage. Par exemple, une maison du village du Tronquay a pu être retrouvée grâce à l'outil cartographique Google StreetView, qui permet d'explorer des villages via internet à la recherche d'endroits figurant sur les prises de vue anciennes — aidés en cela par un carrefour, une église, une entrée de village. D'autres indices peuvent être recoupés pour identifier par exemple le fameux photographe Robert Capa, qui apparaissait sur une photographie connue sans qu'on sache que c'était lui. Ce projet montre comment une communauté de contributeurs peut se lancer dans de longues quêtes et venir à bout d'énigmes que les seuls professionnels n'auraient sans doute jamais pu résoudre.

#sémantique

Qui dit sémantique en SHS aujourd'hui dit Isidore. Vous connaissez ce moteur de recherche et portail qui agrège un ensemble de bases documentaires (collections secondaires, sources primaires...), moissonnées pour être mises en relation et augmentées grâce à des ontologies et des thésaurus. L'objectif est de pouvoir interroger à partir d'une interface unique ce vaste ensemble documentaire et filtrer ensuite de façon intelligente les résultats : par période historique, par discipline, par type de source, par type de document, par collection, etc. Cependant, le web sémantique concerne aujourd'hui essentiellement quelques types de données structurées sur le web (événements, auteurs d'articles, recettes de cuisine...) et des corpus bien délimités (comme DBpedia, Wikidata ou le Centre Pompidou virtuel). Pour de nombreux projets en humanités numériques qui extraient des informations du web, le texte libre reste donc la norme et le web sémantique l'exception. Notre expérimentation [vocabulari.se](#), décrite plus loin, illustre cette situation.

#interopérabilité

Mon dernier mot-clé est celui de l'interopérabilité : comment créer des outils qui soient utilisables par tous à travers la planète pour favoriser notamment la reproductibilité de la recherche ? Comment mutualiser des outils, faire le plus possible des économies d'échelles ? L'interopérabilité est une condition *sine qua non* du développement de systèmes d'information efficaces, et la base à la fois de l'internet et du web. Mais qu'entend-on par là exactement ? Pour l'*European Interoperability Framework* (EIF) publié en 2004 par la Commission Européenne, ceci comprend l'interopérabilité organisationnelle, l'interopérabilité sémantique, et l'interopérabilité technique :

- **l'interopérabilité organisationnelle** doit permettre à des organisations de collaborer à la fourniture de services communs, ou de procéder de manière régulière à des échanges automatisés d'informations, malgré les différences de leurs structures et processus internes
- **l'interopérabilité sémantique** vise à rendre les informations échangées compréhensibles par toute autre application même non développée initialement à cet effet pour pouvoir les combiner entre elles et en exploiter tout le sens. Elle implique généralement la définition et la normalisation de données et métadonnées, le choix de référentiels ou ressources de référence à mettre en œuvre par tous (répertoires d'identification, bases de données, nomenclatures...)
- **l'interopérabilité technique** couvre la mise en relation des systèmes et des services informatiques grâce aux interfaces ouvertes, l'interconnexion des services, l'échange de données, l'accessibilité et la sécurité.

Il va de soi que l'EIF, transposé dans le Référentiel Général d'Interopérabilité (RGI), s'impose aux établissements d'enseignement supérieur et de recherche. Son application est évidente quand on parle d'infrastructures numériques et de grands projets — grâce notamment à l'emploi d'API, de moissonneur OAI-PMH ou de *SPARQL endpoint* — mais à l'échelle des laboratoires de recherche c'est la tradition du "bricolage" qui l'emporte. Cette position est de moins en moins tenable et les laboratoires (chercheurs et ingénieurs) vont devoir apprendre à respecter des pratiques standard qui permettent l'interopérabilité, et inscrivent chaque projet dans un écosystème plutôt que dans un silo étanche.

Une grande cour de récréation ?

Communauté en ligne et hors ligne

Ce survol purement technique des humanités numériques ne doit pas faire oublier que ce sont aussi des hommes et des femmes — qui aiment se retrouver en ligne. Cornélius Puschmann, post-doctorant à l'université Humboldt de Berlin, faisait remarquer (sur Twitter !) que le sens de la communauté des *digital humanities* est basé sur les liens et la communication que les chercheurs entretiennent via Twitter et les blogs plutôt que via les citations d'article à article qui sont le mode traditionnel et le plus courant de la reconnaissance par les pairs et de la formation de "collège invisible" de recherche.

Il est indéniable que la communauté des humanités numériques aime se retrouver en ligne, commenter sur les blogs, converser sur les réseaux sociaux numériques. Mais c'est aussi une communauté qui a une présence physique, AFK (*away from keyboard*), comme en 2010, 2012 et 2013 avec les trois éditions du THATCamp français. En mode "non-conférence" (*barcamp*), les acteurs des humanités numériques viennent avec leurs questions, leurs expérimentations et pendant deux journées échangent au sein d'ateliers qu'ils auront eux-mêmes proposés en fonction de leurs intérêts. Chacun est participant et pas seulement spectateur de ces ateliers, qui se déroulent le plus souvent dans une ambiance studieuse mais bon enfant — *a fortiori* quand l'emplacement du THATCamp renforce le sentiment d'entre soi, que ce soit sur une péniche parisienne en 2010 ou dans un château près de Saint-Malo en 2013.

Il y a donc dans les humanités numériques un mélange de communautés virtuelles et de rencontres physiques. Un mélange qui se produit en simultané : au moment où se tenaient les Journées d'étude de l'Institut des Humanités Digitales de Bordeaux, des participants connectés répétaient les propos (par exemple sur Twitter), les commentaient, les enrichissaient. Ceci participe d'un espace de discussion alternatif plutôt convivial, où se marient réflexion et plaisir.

Du fastidieux au ludique

Après avoir passé en revue ces quelques “marqueurs” des humanités numériques, regardons dans le rétroviseur. Souvenez-vous de la recherche d’information la plus basique, dans une bibliothèque. Il fallait trouver son chemin à la main dans des armoires d’index papier, en partie “au petit bonheur la chance” et en partie guidé par une grande expertise. Aujourd’hui, grâce à des applications logicielles comme Mendeley sur iPad, on peut promener toute sa bibliothèque d’articles scientifiques avec soi, où qu’on aille. L’interface est propre, l’ergonomie est magnifiquement pensée, il devient plaisant de travailler et de lire des articles académiques.

Le plaisir, ou au moins l’excitation et la fierté, c’est aussi ce qui a présidé à la naissance du “Manifeste des digital humanities” qui commence ainsi : “Nous, acteurs ou observateurs de *digital humanities*, nous sommes réunis à Paris lors du That Camp 2010...” Un manifeste qui définit et commence à structurer ceux qui se reconnaissent dans les humanités numériques, en attendant une association professionnelle francophone qui devrait voir le jour en 2014.

Enfin, comment être plus heureux qu’en créant de l’emploi quand on est une communauté enthousiaste qui aime ce qu’elle fait ? “Nous recrutons”, s’exclame sur son site web le Médialab de SciencesPo Paris, l’air de dire qu’ici se prépare la recherche en sciences humaines et sociales de demain ; que c’est un champ en expansion ; et que l’avenir se construira ensemble.

“Plug and play”

Il suffit d’assister à des journées d’études ou de s’essayer à une mini anthropologie de laboratoire pour s’apercevoir qu’en pratique, les humanités numériques ressemblent à une “foire aux outils” constituée par un ensemble d’applications développées par la communauté et dans laquelle chacun va pouvoir puiser pour constituer sa trousse à outils personnelle de chercheur. Omeka, Pajek, R, Gephi, Processing, Zotero, Wordpress, Lodel, TXM... sont autant d’outils de publication, d’analyse et visualisation de données ou de corpus, prêts à l’emploi, qu’il suffirait de “brancher” pour lancer un nouveau projet de recherche ou jeter les bases d’une nouvelle publication. voire simplement pur “jouer”, avec le blog comme terrain de jeu. Sur les carnets de recherche de la plateforme Hypotheses.org, un article peut se résumer à quelques graphes produits à partir d’un jeu de données suivis d’une petite analyse rapide “ce que ça m’évoque”. Un schéma presque ludique, à répéter autant de fois qu’on veut : un jeu de données (souvent en lien avec l’actualité), un outil d’analyse, et voilà des graphes suffisamment parlants pour inciter à la discussion.

Prenons un exemple emprunté à Hypotheses.org : en décembre 2010, Josquin Debaz (historien et sociologue des sciences à l'EHESS) s'était livré à une petite analyse des liens de citations entre les carnets de la plateforme. Il publie son "graphe de l'Hypothesosphère", qui donne naissance à une discussion autour de ce thème (Debaz, 2010). Quelques mois plus tard, Pierre Mercklé (sociologue à l'ENS Lyon) a accès à d'autres données, non pas les liens de citations entre les blogs mais les habitudes de visite et de navigation des internautes. Il propose alors une autre "cartographie de l'Hyposphère" et tente évidemment de comparer les deux (Mercklé, 2011). Quels sont les points communs ? Quelles sont les différences ? On retrouve dans cet enchaînement à la fois une part d'exploration, voire de jeu, et une part de confrontation et reproduction des analyses : qu'est-ce que j'obtiens un avec autre jeu de données ou le même jeu de données, le même outil ou d'autres outils, *etc.* ?

Si l'on marie le *plug* (la foire aux outils) et le *play* (le plaisir de jouer avec les outils et confronter ses analyses à la volée), on arrive à la notion de science *plug and play*. Poussée à l'extrême, ce serait une science déroulant une méthode bien rôdée (Question -> Données -> Outil -> Réponse) avec, au centre, un outil tout puissant. Est-ce ceci les humanités numériques ?

"Out of the box" ou "outside the box" ?

Dans le schéma de la science *plug and play*, les outils ne sont pas loin de constituer des boîtes noires pour les chercheurs-utilisateurs, avec les risques que cela comprend. Les anglo-saxons ont deux expressions parfois confondues en français, "out of the box" et "outside the box". Un logiciel "out of the box" est un outil prêt à l'emploi, qui impose une forme de "pensée packagée" ; alors qu'une pensée "outside the box" sort de son carcan et tente de s'affranchir des schémas préconçus. Toute science est en tension entre ces deux tendances, qui rappellent l'alternance kuhnienne entre science normale ("out of the box") et science révolutionnaire ("outside the box"). Dans la humanités numériques, la place de l'outil doit donc être pensée pour ne pas trop enfermer, ou savoir libérer quand c'est nécessaire.

"Build it and they will come"

Un autre risque guette la profusion des outils : en même temps que ces outils sont relativement faciles à mettre en œuvre, ils sont fragiles. C'est presque une contradiction entre le sentiment que tout devient possible et l'incertitude intrinsèque aux nouveaux outils, à l'instar du grand mythe "construisez et lancez l'outil — *build it* — et vous aurez des utilisateurs en masse — *and they will come*". Une des institutions britanniques qui s'intéresse à la diffusion de l'information en milieu académique a produit un rapport

posant précisément cette question : “*If you build it, will they come? How researchers perceive and use web 2.0*”. Ils concluent évidemment que ce n’est pas aussi simple, et qu’il ne suffit pas de lancer un outil ou un service web pour qu’automatiquement les chercheurs ou le public l’adoptent (Research Information Network, 2010).

Empire ou rébellion ?

La troisième note discordante des humanités numériques est sa position fluctuante entre la marge et le *mainstream*. Le *Manifeste des digital humanities* est un marqueur d’appartenance à une communauté qui se revendique à la marge, hors des cadres institutionnels. Nous avons déjà mentionné l’enthousiasme du Médialab de Sciences Po Paris recrutant trois profils techniques. Or cette page web est ambivalente : elle est illustrée par une photo de Lego représentant des personnages de “La Guerre des étoiles” surmontés de l’inscription “The Empire wants you”.



Les photos de personnages Lego sont un monument la culture *geek*, à laquelle le Médialab fait ici un clin d’œil. Or les vrais geeks auront remarqué que l’Empire ce sont les méchants de “La Guerre des étoiles”, et la rébellion les gentils. N’est-ce pas ironique que quand le Médialab recrute, il le fasse en se mettant à la place des méchants de “La Guerre des étoiles” ? Et au-delà du clin d’œil, il y a de quoi être interpellé : ce recrutement se fait dans le cadre du projet Equipex DIME SHS, financé par les Investissements

d'avenir (grand emprunt). Difficile d'être plus institutionnel. Alors, les humanités numériques, Empire ou rébellion ?

Nous situant plutôt du côté de la rébellion, nous avons voulu à notre tour rejoindre cette grande foire aux outils des humanités numériques en développant au sein de Deuxième labo une application web libre et gratuite. Il est temps maintenant de partager un retour d'expérience sur cet outil et le semi-échec qu'il constitue : vocabulari.se <<http://vocabulari.se>>.

Retour d'expérience vocabulari.se

Le principe de vocabulari.se était de construire un réseau social des objets de recherche. Non pas des personnes, mais des objets de recherche, partant du principe que chacun travaille sur son propre objet — le calcul distribué, le réchauffement climatique, les écoulements diphasiques, la constitution française, l'addiction au tabac... Grâce à vocabulari.se, nous voulons vous permettre de chercher le mot ou les deux mots qui forment votre objet de recherche pour, d'un coup, convoquer les autres objets de recherche auxquels il se rattache. Ainsi, nous révélons son réseau, ce dans quoi il s'inscrit, et tout son poids politique. Un des effets de ce dévoilement pourra être de renvoyer le chercheur à son vocabulaire, de donner un autre relief aux termes qu'il emploie lorsqu'il écrit un article, rédige un abstract ou soumet une demande de financement — et finalement d'explicitier la dimension politique des objets de recherche, au sens de Bruno Latour.

Cette idée ne vient pas de nulle part : la fréquentation de Bruno Latour et ses écrits nous a conduit à penser ainsi, et la référence que vous venez de lire n'est pas fortuite. Il y a également en toile de fond une réflexion sur l'actualité médiatique où la science est presque toujours présentée à travers les objets qu'elle étudie — et par là on entend aussi bien des objets "physiques" que des objets plus abstraits, plus conceptuels : la fonte des glaces, la disparition des abeilles, la susceptibilité génétique au cancer du sein, la mondialisation... Et c'est également par ces objets multiples et hybrides que la science fait irruption dans nos journaux télévisés et dans les rubriques "Société", "Politique", "Planète" ou "Santé" de nos magazines. Ces objets ("non-humains" dirait Latour) sont donc incontournables aujourd'hui et nous souhaitons leur donner l'importance qu'ils méritent — tout en bousculant l'idée selon laquelle ils seraient monolithiques, univoques, et réductibles à une vérité présentée comme universelle.

Pré-étude

Une fois l'objectif fixé, comment s'y prendre ? Nous avons commencé par défricher le terrain lors d'une pré-étude confiée à des étudiants en informatique de Polytech Lyon dans le cadre d'une UE de veille technologique. Nous leur avons demandé de rechercher parmi les API (des interfaces machine/machine permettant de puiser dans des sources de données hétérogènes) lesquelles donnent accès à l'information scientifique et technique, en l'occurrence aux bases de données d'articles. Et lesquelles de ces API sont réutilisables, avec des licences plutôt souple.

La conclusion à laquelle ils sont arrivés, conforme à notre intuition, était qu'il y a d'une part des archives ouvertes où les articles sont réutilisables et qui possèdent une API mais qui sont extrêmement pauvres quantitativement (par exemple HAL en France — on pourrait s'attendre à ce que toute la production des chercheurs français payés par nos impôts soit disponible sur HAL, mais non seulement ce n'est pas le cas, et en plus cela semble une utopie hors de notre portée aujourd'hui...). D'autre part il y a des éditeurs scientifiques privés, comme Elsevier, qui sont assis sur leurs bases de données d'articles et qui en ferment pratiquement l'accès, et qui restent organisés en silos de spécialités, ce qui rend une interrogation pluridisciplinaire compliquée. Il ne reste qu'un petit groupe d'acteurs, les outils du web 2.0 comme CiteULike, Zotero ou Mendeley qui permettent aux chercheurs d'organiser en ligne leurs références d'article. En mettant en commun les bibliothèques de chacun des utilisateurs, ils donnent une très bonne image de ce que sont les publications scientifiques aujourd'hui, pratiquement toutes disciplines et tous pays confondus.

Déclencheur

Arrivés à ce stade, nous avons choisi l'API de Mendeley comme source de références en raison de son excellente couverture bibliographique (50 millions d'articles déclarés, soit 10 millions de plus que le Web of Science de Thomson Reuters qui fait pourtant référence) d'une part, et de son API et ses données placées sous licence libre Creative Commons d'autre part. Juste au moment (était en 2011) où Mendeley et PLoS lançaient un concours de la meilleure application bâtie avec leurs API, doté de 10 000 \$. (À la lumière de cette expérience, tous les producteurs d'API donnant accès à des données, notamment en France, seraient bien inspirés de lancer ce type d'initiative qui fonctionne assez bien, pour justement inciter des communautés de développeurs à s'y intéresser et proposer de nouveaux services.) Nous voilà donc inscrits au concours.

Ingrédients

La recette de vocabulari.se telle qu'on l'avait composée était assez simple : nous nous appuyons d'abord sur des corpus issus de la science contributive (issus de la mise en commun des informations de nombreux utilisateurs), c'est-à-dire Mendeley mais aussi Wikipédia que nous avons voulu ajouter à la recette pour apporter une couche de connaissances plus "grand public". Puis nous ajoutons quelques algorithmes d'analyse de grand corpus, c'est-à-dire de cette masse de données. Enfin nous ajoutons l'ingrédient du design de l'information, qui concerne la façon de présenter le résultat, ce qu'on retourne à l'utilisateur de vocabulari.se, l'expérience qu'on veut lui faire vivre.

Design

Le design de l'information a consisté à imaginer comment se présenterait la page de résultats de vocabulari.se. Après plusieurs itérations, nous sommes tombés d'accord sur le fait que l'utilisateur devrait entrer un ou deux mots-clés et que le moteur de recherche lui retournerait trois colonnes de mots, tous liés au mot-clé cherché.

La première colonne correspond aux **mots inattendus** ("unexpected"), classés ainsi car ils apparaissent peu fréquemment en compagnie du mot-clé mais fonctionnent bien lorsque c'est le cas (reçoivent beaucoup de vues dans Mendeley).

La deuxième colonne concerne les **mots controversés** ("controversial") : des mots qui apparaissent dans l'article Wikipédia du mot-clé recherché et dont la page Wikipédia est elle-même très controversée (page de discussion très longue, où les wikipédiens débattent longuement de ce que signifie ce concept, son histoire, ses ramifications, son importance etc.).

La troisième colonne est celle des **mots qui agrègent d'autres disciplines** ("aggregating"). Ce sont des mots apparaissant aux côtés du mot-clé de départ dans des documents appartenant à des bibliothèques Mendeley de chercheurs de disciplines différentes. Utilisez ces mots plutôt que votre mot de départ et vous pourriez vous ouvrir tout un champ de collaboration ! Voilà ce que nous avons en tête comme page de résultat pour vocabulari.se.

Un autre niveau de navigation et d'information est disponible en cliquant sur chacun des mots associés : s'affiche alors la "preuve" du lien entre le mot-clé cherché et ce mot, c'est-à-dire une référence bibliographique dans Mendeley, un article Wikipédia ou une page discipline dans Mendeley.

Pour la page d'accueil, nous avons voulu épurer au maximum en concentrant l'attention sur le champ de recherche (façon Google), accompagné quand même de quelques lignes d'explication (dont certaines, cachées, ne s'affichent qu'au survol de la souris) et de l'injonction "vocabulari.se your research!".

Collaborations

Il ne restait plus qu'à réunir l'équipe. L'idée originale revient aux auteurs du présent article et notre collaborateur Axel Meunier. Pour le développement nous avons fait appel à des développeurs passionnés par ce type de challenge technique et intellectuel : Gnuside. Le design graphique est signé Jérémie Elalouf, qui a créé tout l'habillage et toute l'identité visuelle du projet. Enfin, nous nous sommes appuyés sur les conseils de Pierre Maumont pour l'infogérance, c'est-à-dire la gestion des serveurs et de l'hébergement.

Résultat

Le résultat, malheureusement, n'est pas utilisable : actuellement `vocabulari.se` ne fonctionne pas. Quand on lance une recherche, elle s'exécute mais ne retourne pas de résultat. C'est un constat décevant.

Le problème provient du difficile compromis entre nos exigences (pour que le résultat soit extrêmement pertinent et s'appuie sur des données assez vastes) et les contraintes, au premier rang desquelles l'API Mendeley qui a un plafond de requêtes. Au final il manque un soupçon d'optimisation dans notre recette pour qu'on puisse réellement produire les résultats attendus par l'utilisateur.

Et pourtant, nous avons tout mis en œuvre pour que le résultat soit convaincant. Par exemple, nous avons pensé à mettre en place un cache afin de pouvoir afficher plus rapidement les résultats d'un mot-clé déjà cherché (ou proche d'un autre mot-clé déjà cherché). Nous avons même envisagé d'utiliser le cache pour, à chaque répétition d'une recherche, augmenter le nombre de documents analysés et la profondeur du réseau des objets de recherche, trouver d'autres degrés de relations et retourner des résultats encore plus pertinents. Sauf que pour constituer un cache il faut qu'il y ait des utilisateurs qui fassent des requêtes, et que si l'outil ne marche pas alors il n'y a pas d'utilisateurs, donc il n'y a pas de cache, et donc ça marche encore moins bien, etc. : un cercle vicieux.

Améliorations

Aujourd'hui, maintenant que nous sommes entrés à l'ère du big data, certains de ces problèmes ont trouvé leur réponse. Aussi bien pour la gestion du cache que de la "file des opérations", des technologies existent qui seraient faciles à mettre en œuvre et amélioreraient vraisemblablement notre outil. C'est prévue dans la prochaine mise à jour de `vocabulari.se`.

Plus que dans le cas d'un franc succès, ce retour d'expérience d'un semi-échec met en lumière le travail de construction d'un outil numérique et leur fragilité quand on entre dans le domaine de la recherche. Néanmoins c'est un échec qui est assez fructueux dans le sens où il nous incite à aller plus loin. D'abord parce que nous allons tout faire pour corriger ces défauts et faire en sorte que *vocabulari.se* fonctionne rapidement. Ensuite, parce qu'il ouvre également des perspectives : comme *vocabulari.se* met en relation des mots selon trois critères, on pourrait monter à partir de là au niveau des documents (puisque les documents sont des ensembles de mots), puis déduire des relations entre auteurs à partir de relations entre documents. Ainsi, plutôt que de donner des affinités entre auteurs par le simple fait que les auteurs sont en relation (ils publient ensemble, ils appartiennent à la même institution...), comme c'est si souvent le cas, on pourrait qualifier leurs affinités par le fait qu'ils publient des documents dont les mots relèvent du même champ sémantique ou possèdent les liens de controverse, de fertilité, de pluridisciplinarité dont il était question précédemment. Une autre idée pourrait être d'utiliser *vocabulari.se* comme rouage d'une chaîne de traitement de l'information : à partir d'un mot clé, on pourrait obtenir son champ lexical et requêter ensuite des bases de données de publications pour obtenir des articles connexes pertinents — bien qu'ils ne contiennent pas le mot clé lui-même. À votre imagination, voire à votre code puisque le code source de *vocabulari.se* est publié sous la licence GNU Affero GPL. Invitation à contribuer ou s'en inspirer.

Avec l'exemple de *vocabulari.se*, nous espérons avoir donné un aperçu de la position qu'occupe non pas l'utilisateur de l'outil, mais celui qui le crée. Il mène un travail radicalement différent de celui de l'utilisateur, et ô combien enrichissant. Les questions sans réponse, les ratages... font partie de l'aventure. Si les humanités numériques sont une science centrée sur les outils, alors il est important de s'intéresser au processus de création de ces outils, aux questions nouvelles qu'ils posent. Bien loin de stériliser le champ, ils deviennent alors des moteurs heuristiques pour la recherche, associée à la création. Les humanités numériques ont finalement de beaux jours devant elles !

Bibliographie

- Collectif, 2011, "Manifeste des digital humanities", *THATCamp Paris*, mis en ligne le 26 mars 2011, consulté le 25 octobre 2013. URL : <http://tcp.hypotheses.org/318>
- Debaz Josquin, 2010, "Hypothesosphère", *Socio-informatique et argumentation*, mis en ligne le 8 décembre 2010, consulté le 25 octobre 2013. URL : <http://socioargu.hypotheses.org/1921>
- Mercklé Pierre, 2011, "L'Hyposphère", *pierremerckle.fr*, mis en ligne le 7 juin 2011, consulté le 25 octobre 2013. URL : <http://pierremerckle.fr/2011/06/lhyposphere/>

Michel Jean-Baptiste et coll., 2011, “Quantitative Analysis of Culture Using Millions of Digitized Books”, *Science*, vol. 331 n° 6014, pp. 176-182

Research Information Network, 2010, *If you build it, will they come? How researchers perceive and use web 2.0*, consulté le 25 octobre 2013. URL :

http://www.rin.ac.uk/system/files/attachments/web_2.0_screen.pdf