**Cartographie du Web littéraire numérique francophone**

Introduction

Bonjour à tous, nous sommes enchantés de pouvoir partager avec vous notre retour d’expérience au sein du projet CartoWeb. Il s’agira d’aborder au sein de cette communication à plusieurs voix les cheminements que nous avons éprouvés afin de répondre à nos missions, d’exposer le modèle de données ainsi que les premiers résultats obtenus.

Nous sommes tous les trois étudiants en double cursus d’humanités numériques et avons comme vous pourrez le constater des profils complémentaires. Riche d’une formation d’infographiste, Inès Burri est actuellement en Master 2 d’histoire ancienne. Célian Ringwald est quant à lui diplômé d’un master en science des données. Il travaille depuis trois ans en tant que *Data scientiste* chez *Data Observer*. Enfin, quant à moi, Anaïs Chambat, je suis titulaire d’un master en histoire des mondes germaniques. J’étudie actuellement les applications du numérique et des méthodes quantitatives à l’histoire.

Les missions qui nous ont été confiées dans le cadre de Cartoweb sont « la mise en place d’une base de données et des outils pour verser automatiquement les données issues d’un nombre limité de pages Web au format clairement défini ainsi qu’une réflexion sur les visualisations les plus adaptées à un utilisateur non expert et l’implémentation d’une ou plusieurs de ces visualisations. » Pour effectuer ce travail, une liste d’adresses Web répertoriant des sites et des pages d’écrivains francophones publiant en ligne ont été mises à notre disposition au format RDF. Les sites et pages peuvent être associés à un ou plusieurs auteurs et ont été recueillis à partir de métadonnées relativement limitées (localisation, date, auteur, type de média utilisé).

* Cartographier et visualiser le web littéraire francophone, oui, mais comment ?
  + Avec des cartes thématiques et géographiques ?
  + Par une visualisation de liens de collaborations ? de contenus ?
  + Mais, encore par le biais d’une lecture distante de contenus et de parcours d’auteurs physico-numériques.

Si notre mission principale était la visualisation des données, nous avons été contraints de compléter les métadonnées existantes et de collecter de nouvelles informations afin de constituer une base de données augmentée.

IV. Résultats

Grâce au balisage, il a été possible d’observer les possibilités et les limites de l’extraction. D’un point de vue concret, concernant les informations extraites à partir des répertoires d’autorités, nous pouvons établir un premier constat :

* Sur 220, 7 entités auteurs ne possèdent qu’un identifiant à savoir celui de la BNF.
* 15 possèdent deux identifiants à savoir ceux de la BNF et du VIAF.
* 35 possèdent à la fois des identifiants au sein des répertoires d’autorité mais aussi une entrée au sein de la base de données Wikipédia.
* 163 demeurent ne pas être directement renseignés. On peut ici supposer que ce sont précisément ces auteurs, encore méconnus que cherchent à mettre au jour les projets CartoWeb et Lifranum.

Les auteurs « traditionnels » francophones sont en effet rendus visibles par leur notice BNF. Les auteurs bénéficiant une certaine visibilité européenne voire internationale possèdent également un identifiant VIAF. Enfin, les auteurs prolifiques et jugés représentatifs une entrée au sein de la base de données de Wikipédia. L’ISNI pourrait ainsi constituer un point d’entrée supplémentaire.

Les notices existantes ne sont d’ailleurs pas nécessairement complètes. Prenons l’exemple du sexe qui est renseigné pour 67 notices soit 30,7% des entités auteurs.

Concernant les données liées à la langue et à la profession, le site de référence d’auteurs SPLA a été d’une aide précieuse.

* 75% des données extraites automatiquement proviennent de cette source,

le quart restant de la BNF.

Concernant la langue maternelle, 98% des notices mentionnent le français (BNF, SPLA) et 2% le créole (BNF). Les pays de résidence des entités auteurs extraits à partir du SPLA sont francophones. Ainsi, Haïti est représenté à hauteur de 90%, la France 5% et le Canada 3%. On remarque ainsi une faible mobilité des auteurs originaires d’Haïti. Ceci est à mettre en regard de la langue de traduction des ouvrages des entités auteurs présents sur *île-en-île*. On observe ici que si l’anglais domine à 38%, le français et l’espagnol ne sont pas loin derrière avec 17 et 16%. 4% des ouvrages ont même été traduits en créole. Par le biais de la traduction, la littérature francophone semble donc ainsi gagner en mobilité.

Nous avons également pu extraire les professions et dresser un top 5 en effectifs des professions extraites depuis le SPLA. La première place revient ainsi aux écrivains, la deuxième aux poètes, la troisième aux auteurs, la quatrième aux universitaires et la cinquième aux journalistes. On remarque ici le paradigme sémantique entre l’auteur et l’écrivain, qu’est-ce qui les différencient ? Pourquoi revendiquer l’un et pas l’autre ? Un individu peut par ailleurs avoir exercé plusieurs activités ou professions au cours de sa vie. L’extraction ne nous permet pas d’adopter une approche diachronique et successive des entités auteurs et ne peut qu’en rendre compte quantitativement. Nous pourrions à termes envisager une représentation physico-numérique des auteurs où il s’agirait de partir de leurs supports de diffusions ou de productions numérique pour définir leur « parcours de vie et d’activités professionnelles ».

Enfin, grâce au processus d’enrichissement des données, il est possible d’identifier de nouveaux auteurs. Ceci peut être exploré à travers une analyse textuelle des notices biographiques ou encore des visualisations en réseaux :

* De liens de collaborations identifiés au sein des notices biographiques et des répertoires d’autorités
* Ici, de liens entre les entités auteurs et les personnes qui rédigent leurs notices biographiques

Conclusion

* Notre cas d’étude sur Haïti est généralisable du local au global.
* En effet, des connexions avec d’autres aires géographiques se dessinent déjà.
* Le modèle proposé se veut ainsi ouvert et flexible.
* Partant du noyau de 200 auteurs haïtiens, nous avons pu constater qu’il était possible d’identifier près de 1000 autres auteurs potentiels.