COTE Christian <christian.cote@univ-lyon3.fr>

Hier, 12:59 BURRI inès stéphanie <ines.burri@univ-lyon3.fr>; fewou@univ-tlse2.fr; Julien Velcin; achambat <anais.chambat@ens-lyon.fr>; Celian Ringwald

Bonjour

voici un premier retour sur le texte que vous m'avez fait passer, afin de commencer avant la réunion du 7 octobre à cerner le projet :

Votre schéma initial est très intéressant et particulièrement complet.

Néanmoins, si l’on se base sur ce dont nous disposons actuellement, il s’agira essentiellement de travailler sur des URI (les noms de domaines, les identifiants de page, etc.). Ces adresses donnent des indications sur les types de médis, les noms de personne, etc.

On peut utiliser des jeux de métadonnées descriptives déjà élaborés comme PRISM (<https://www.idealliance.org/prism-metadata> ) de façon à caractériser les informations que l’on a pu obtenir à partir des adresses.

Le fichier RDF récupère ces adresses et normalement des indications d’auteur et de zones géographiques.

ATTENTE N. 1.

Les adresses sont génériques au sens où elles peuvent contenir des pages spécifiques, mais cela dépend du format (template) du support et donc de son mode d’actualisation. (Elles peuvent se décliner en pages ou une seule page peut être actualisée).

**L’échantillon part de listes fournies par la BNF**, donc des auteurs aillant publier ailleurs que sur le web. Nous avons **recherché la présence de ces auteurs sur le web**, **et par inférence, de tous ceux qui leur sont liés**. De ce fait, on a pu mettre en évidence d’autres auteurs et on a fonctionné de façon incrémentale.

* Disposerons-nous de la liste donnée ? De la liste des auteurs ayant été « découverts » ?
* Comment se lien a-t-il été réalisé ?

Tous les auteurs BNF n’ont pas nécessairement de présence sur le web. Par ailleurs, des auteurs effectivement présents sur le web **peuvent utiliser différents médias pour des usages spécifiques** (infos personnelles dans Facebook, création littéraire dans un blog par exemple).

* Certaines pages servirons tantôt à renseigner la fiche auteur, tantôt à analyse les textualités de ses créations.

ATTENTE N. 2.

Le but est effectivement une **interface de consultation**.

La question suivante est un peu plus compliquée : je proposerai que l’on crée par auteur + zones géographiques des fiches décrivant les **formats de leur expression** et **les liens qu’ils peuvent avoir**. Il me semble difficile (et pas nécessairement pertinent) d’attribuer des genres et de qualifier les contenus.

* Qu’entendre par expressions ? Vocabulaire commun ? N-Grams ? Expressions ? Concepts ?

Autre question : est-ce que l’on vise à importer des documents ou pas ? Le répertoire n’implique pas la récupération des documents.

Simplement, le plus intéressant consistant à **créer un schéma XML/RDF, empruntant en partie à des schémas existant** (comme **PRISM** mais on pourrait trouver également des éléments intéressant dans les **IPTC**, ou alors partir plus clairement du côté de **FOAF** pour caractériser les personnes et leurs liens), répertoriant les informations accessibles avec les données à disposition.

* IPTC : enrichissement lieu photo ? images ? (http://peccatte.karefil.com/software/Metadata.htm)

ATTENTE N. 3.

Pour répondre à ces questions-là, il faudra mettre en place le crawling.

ATTENTE N. 4.

Au départ, c’est une visualisation **géographique** que l’on peut doubler d’une **représentation par réseau.**

Une autre représentation que l’on peut facilement obtenir, c’est par **les types de médias utilisés**. A l’aide de PRISM par exemple, c’est quelque chose que l’on peut facilement creuser.

* Une carte cloropeth filtrable permettant de filter une vue « réseaux ».
* On peux aussi exploiter une dimension temporelle

Shema de viz :

Selecteur temporel, possibilité de filtre par période de date.  
Selecteur et carte des auteurs  
Réseau social des auteurs  
Tableau d’affichage des pages et auteurs référencés.

