HAL open science

Assister l'édition de données Omeka S par un mécanisme de suggestion

Nicolas Lasolle, Geoffrey Douilly

To cite this version:

Nicolas Lasolle, Geoffrey Douilly. Assister l'édition de données Omeka S par un mécanisme de suggestion. Journées Omeka, Oct 2020, Nancy, France. hal-02964423

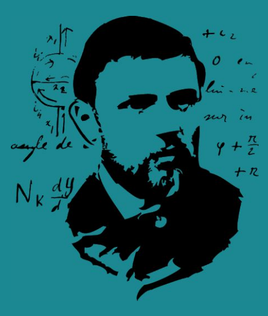
HAL Id: hal-02964423

https://hal.univ-lorraine.fr/hal-02964423v1

Submitted on 12 Oct 2020

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.



Assister l'édition de données Omeka S par un mécanisme de suggestion

Nicolas Lasolle[1,2], Geoffrey Douilly[1]

[1] Université de Lorraine, CNRS, Université de Strasbourg, AHP-PReST, F-54000 Nancy, France
[2] Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Contexte et problématique

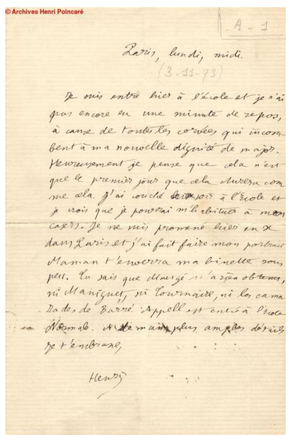


FIGURE 1 – Henri Poincaré à Aline Poincaré, 1873.

Le corpus de la correspondance active et passive d'Henri Poincaré (1854-1912) est une source importante d'informations pour les historiens des sciences de la période. Composé de plus de 2000 lettres, il regroupe des échanges d'ordres scientifique, administratif et privé. L'édition des données via l'interface d'Omeka S est un processus qui peut sembler long et fastidieux pour les utilisateurs concernés mais qui nécessite néanmoins leur concentration pour éviter des erreurs diverses. Un outil utilisant des technologies du Web sémantique a été mis en place pour assister l'édition de données du corpus lié à Henri Poincaré.

Exemple d'édition

Omeka S permet la mise en place de liens entre des ressources dans la logique du modèle RDF, qui est le modèle standard pour représenter les données du Web sémantique. La figure 2 donne un exemple de graphe pour une lettre de la correspondance. Cet ensemble de données peut s'exprimer sous la forme d'une liste de triplets de la forme ⟨`sujet prédicat objet`⟩.

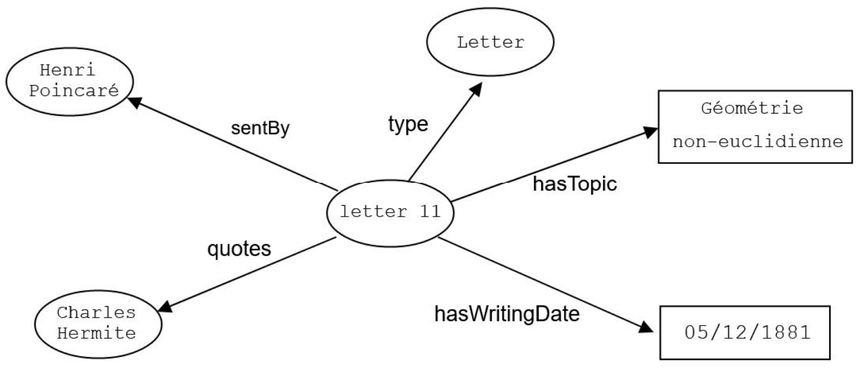


FIGURE 2 – Exemple de graphe pour une lettre de la correspondance.

Imaginons qu'un utilisateur souhaite poursuivre l'édition de cette lettre par l'ajout d'un nouveau triplet pour définir l'expéditeur. L'objectif de l'outil est de fournir une liste ordonnée de suggestions pour l'objet :

$$\texttt{Problème} = \langle \texttt{letter11 sentTo ?o} \rangle$$

Méthodologie

Méthode 1 : exploitation des connaissances de l'ontologie

Le langage de représentation de connaissances RDFS permet d'ajouter des informations sur les relations entre les classes et les propriétés de l'ontologie. En particulier, les notions de domaine et co-domaine sont introduites et permettent d'ajouter des contraintes pour les ressources qui sont associées à une propriété dans un triplet. Dans l'ontologie utilisée pour le corpus d'Henri Poincaré, la propriété `sentTo` a pour domaine la classe `Letter` et pour co-domaine les classes `Person` et `Institution`. Pour aider à résoudre le problème d'édition courant le co-domaine est utilisé par l'outil afin de favoriser les instances des classes `Person` et `Institution` dans la liste de suggestions.

Méthode 2 : application du raisonnement à partir de cas

Une autre méthode permettant d'affiner la liste de suggestions est d'utiliser le raisonnement à partir de cas : des cas d'édition similaires au problème courant sont utilisés afin de faire des suggestions pertinentes. Dans l'exemple courant, l'objectif est de rechercher les lettres éditées présentant des similitudes avec la lettre en cours d'édition (`letter11`).

Cette recherche s'effectue grâce à l'utilisation du langage SPARQL qui permet d'interroger des graphes RDF. Cependant, il est rare d'avoir deux ressources ayant exactement les mêmes caractéristiques, et c'est pourquoi un mécanisme de recherche approchée a été mis en œuvre. Celui-ci entraîne la génération de requêtes SPARQL à partir d'une requête initiale, d'une base de données RDF et d'un ensemble de règles de transformation. Par exemple, la requête $\mathcal{Q}$, exprimée de façon informelle, peut être transformée en la requête $\mathcal{Q}_1$.

$\mathcal{Q} =$ Donne-moi les lettres envoyées par Henri Poincaré ayant pour thème la géométrie non-euclidienne et citant Charles Hermite.

$\mathcal{Q}_1 =$ Donne-moi les lettres envoyées ou reçues par Henri Poincaré ayant pour thème la géométrie et citant un homme de sciences.

En exécutant les requêtes générées, il est alors possible de récupérer et classer les valeurs associées à la propriété `sentTo` pour des lettres similaires à la lettre en cours d'édition.

Outil d'édition

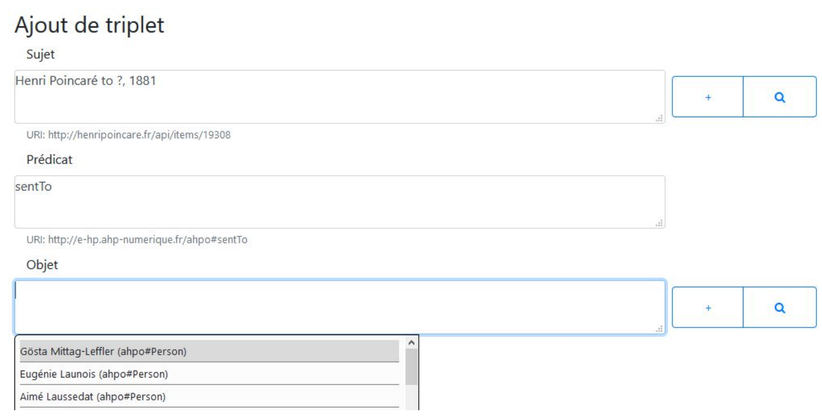


FIGURE 3 – Interface d'édition de données RDF.

Évaluation des différentes approches

TABLE 1 – Score moyen, sur une échelle de 1 à 7, lié à la pertinence des suggestions telle que perçue par un humain.

	Sans assistance	Méthode 1	Méthode 2	Combinaison
Score moyen	3.4	5.7	5.3	7

TABLE 2 – Mesures du rang de la valeur attendue dans la liste de suggestions en simulant des problèmes d'édition à partir de 100 lettres éditées de la correspondance.

	Sans assistance	Méthode 1	Méthode 2	Combinaison
$rang \leqslant 15$	11.3 %	22.11 %	49.0 %	49.5 %
$rang \leqslant 10$	7.1 %	21.15 %	43.2 %	43.2 %
$rang \leqslant 5$	2.7 %	19.23 %	33.6 %	34.7 %

Perspectives

Un travail en cours concerne l'adaptation de ce mécanisme de suggestion pour une intégration à l'interface d'édition d'Omeka S. Des premiers essais ont apporté des résultats convaincants et il serait pertinent de tester cet outil dans d'autres contextes que celui de la correspondance d'Henri Poincaré.

Remerciements

Ce travail a bénéficié d'une aide de l'État, gérée par l'Agence Nationale de la Recherche, au titre du projet Investissements d'Avenir Lorraine Université d'Excellence, portant la référence ANR-15-IDEX-04-LUE. Merci à Olivier Bruneau, Jean Lieber, Emmanuel Nauer, Siyana Pavlova et Mickael Smodis de leur participation à ces travaux de recherche.