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-02964423
Submitted on 12 Oct 2020
Contexte et problématique


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).

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 approché 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 Q1, exprimée de façon formelle, peut être transformée en la requête Q2.

\[
Q_1 = \text{Donne-moi les lettres envoyées par Henri Poincaré ayant pour thème la géométrie non-euclidienne et citant Charles Hermite.}
\]

\[
Q_2 = \text{Donne-moi les lettres envoyées ou reçues par Henri Poincaré ayant pour thème la géométrie et citant un homme de sciences.}
\]

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

Évaluation des différentes approches

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

<table>
<thead>
<tr>
<th>Score moyen</th>
<th>Sans assistance</th>
<th>Méthode 1</th>
<th>Méthode 2</th>
<th>Combinaison</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>5.7</td>
<td>5.3</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Tableau 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.

<table>
<thead>
<tr>
<th>rang ⩽ 15</th>
<th>rang ⩽ 10</th>
<th>rang ⩽ 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.3 %</td>
<td>22.11 %</td>
<td>49.0 %</td>
</tr>
<tr>
<td>7.1 %</td>
<td>21.15 %</td>
<td>43.2 %</td>
</tr>
<tr>
<td>2.7 %</td>
<td>19.23 %</td>
<td>33.6 %</td>
</tr>
</tbody>
</table>

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