

Comment exploiter un corpus à l'aide des technologies du Web sémantique ? Le cas de la correspondance d'Henri Poincaré.

Colloque Humanistica 2021

Nicolas Lasolle et Pierre Willaime

10 mai 2021



Archives Henri Poincaré
Philosophie et Recherches sur les
Sciences et les Technologies



IMPACT
OLKi

Contexte

Contexte

Le corpus de la correspondance d'Henri Poincaré

- ▶ un projet collectif depuis 2002
- ▶ une édition numérique de la correspondance de Henri Poincaré (1854-1912)
- ▶ un appareil critique pour chaque lettre

une méthodologie un peu à part

Au lieu d'encoder la transcription (en TEI par exemple), on renseigne des triplets RDF pour représenter les connaissances.

Contexte

Omeka S et le site henripoincaré.fr

- ▶ Omeka S : un formidable outil libre et modulable, (<pub>rejoignez l'association AUFO (<https://omeka.fr/>)</pub>
- ▶ qui permet à chacun de pouvoir décrire le corpus sans connaissance technique.
- ▶ pas de barrière entre publier et exploiter
- ▶ adapté à un corpus dont la description se fait au fil de l'eau



LETTRE : Henri Poincaré à Eugénie Poincaré - janvier 1876

Transcription

Métadonnées

Citer ce document

[Janvier 1876]

Ma chère **maman**,

Hier soir j'ai eu trois invitations à dîner.

1° Chez **M^{me} Malgaigne**, faite depuis dimanche.2° Chez **M^{me} Barthélémy** ; j'ai dû refuser, j'ai été chez elle samedi, point ne l'ai trouvée.3° Chez **M^{me} Brice** chez qui se trouvait l'oncle Charles¹. Envoyé dépêche.

Chez **M^{me} Malgaigne** se trouvaient outre les **Olleris**, les **Lefort**, un **taupin** et sa mère, papillotée² et Pont Lévêquoise³ et enfin un jeune homme très gentil aspirant je crois au conseil d'État. Le dîner n'a pas présenté l'aspect lugubre redouté de **M^{me} Olleris**, l'absence des **Pilastre** ayant empêché la conversation traditionnelle sur l'enfantépatancologie Pilastro-Lefortique comparée.

M. Lefort a raconté ses voyages en découpant son poulet comme à l'ordinaire, mais les réussites traditionnelles n'ont pas paru.

Ce matin tentative infructueuse de verglas.

M. Daubrée, inspecteur général des mines, et **Madame Daubrée** resteront chez eux les samedis 22 janvier et 19 février.

1. Il s'agit probablement de **Charles Comon**.↩

2. Il s'agit d'une allusion à la coiffure de cette dame qui avait sans doute les cheveux frisés ou qui portait de petits morceaux de papier (des papillotes) pour faire friser ses mèches de cheveux.↩

3. Une habitante de Pont-l'Évêque ? Le terme exact serait « pontépiscopoise ».↩

Noms cités ▾

Personne citée dans le document primaire.

ahpo:citeName

Eugénie Launois 

 Contenus  Collections  Médias

Françoise Malgaigne 

 Contenus  Collections  Médias

Marie-Anne Barthélémy 

 Contenus  Collections  Médias

Mme Brice 

 Contenus  Collections  Médias

Famille Olleris 

 Contenus  Collections  Médias

famille Lefort 

 Contenus  Collections  Médias

Marie Olleris 

 Contenus  Collections  Médias

Famille Pilastre 

 Contenus  Collections  Médias

Léon Le Fort 

 Contenus  Collections  Médias

Gabriel Auguste Daubrée 

 Contenus  Collections  Médias

Marie Laure Rosalie Daubrée 

 Contenus  Collections  Médias

pros

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables

cons

- ▶ aplatissement de la structuration de l'ontologie (ex : dcterms:created sous-propriété de dcterms:date)
- ▶ pas d'attribut ou de domaine d'application pour une métadonnée

⇒ Omeka S reste très utile comme un outil d'alimentation et d'exposition.

L'utilisation de technologies du Web sémantique

L'utilisation de technologies du Web sémantique

Introduction

- ▶ Extension du Web qui vise à décrire des ressources de façon formelle pour qu'elles soient exploitables par des agents logiciels

L'utilisation de technologies du Web sémantique

Introduction

- ▶ Extension du Web qui vise à décrire des ressources de façon formelle pour qu'elles soient exploitables par des agents logiciels
- ▶ Le terme regroupe un ensemble de pratiques et standards pour représenter, lier, interroger et raisonner avec des données

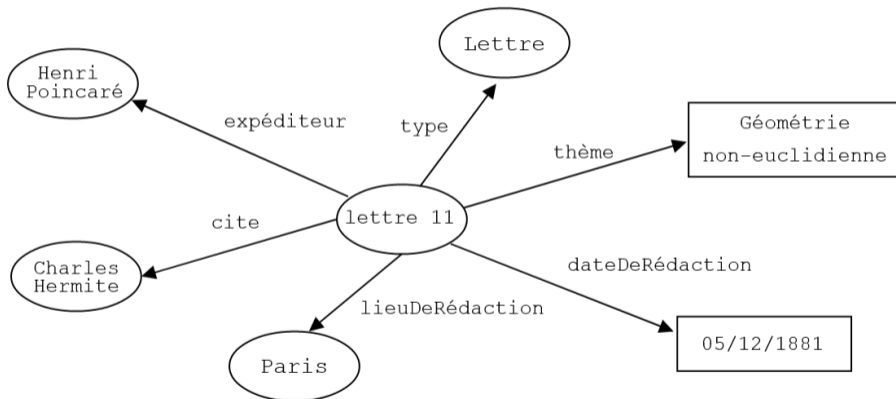
L'utilisation de technologies du Web sémantique

Introduction

- ▶ Extension du Web qui vise à décrire des ressources de façon formelle pour qu'elles soient exploitables par des agents logiciels
- ▶ Le terme regroupe un ensemble de pratiques et standards pour représenter, lier, interroger et raisonner avec des données
- ▶ Certaines de ces technologies sont utilisées pour représenter et valoriser les données du corpus d'Henri Poincaré

L'utilisation de technologies du Web sémantique

Resource Description Framework (RDF)



SPARQL

Langage permettant de formuler des requêtes afin d'interroger des graphes RDF.

Exemple de requête (informelle)

$Q =$ | Donner les lettres envoyées par Henri Poincaré
à Gösta Mittag-Leffler entre 1880 et 1890
et ayant pour thème l'optique.

Des problématiques de représentation de connaissances

Des problématiques de représentation de connaissances

Temporalité

Un des manques dans la plupart des représentations du Web sémantique est de ne pas considérer la validité temporelle des faits représentés.

Exemple

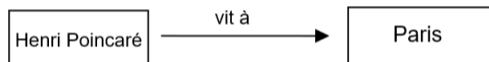
Le fait indiquant qu'Henri Poincaré avait Paris pour lieu de sa résidence principale doit être considéré comme valide pour la période allant de 1881 à 1912.

Comment lier une temporalité à un fait en RDF ?

Des solutions existent, chacune avec des avantages et inconvénients techniques à considérer lors du choix d'un modèle.

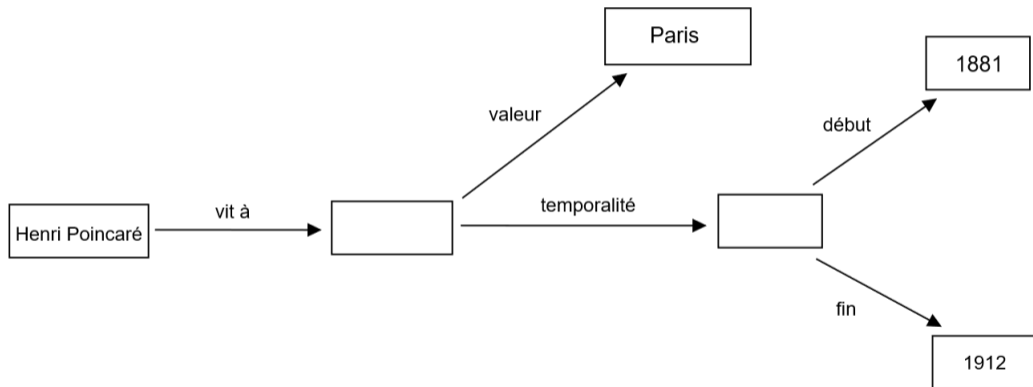
Des problématiques de représentation de connaissances

Temporalité



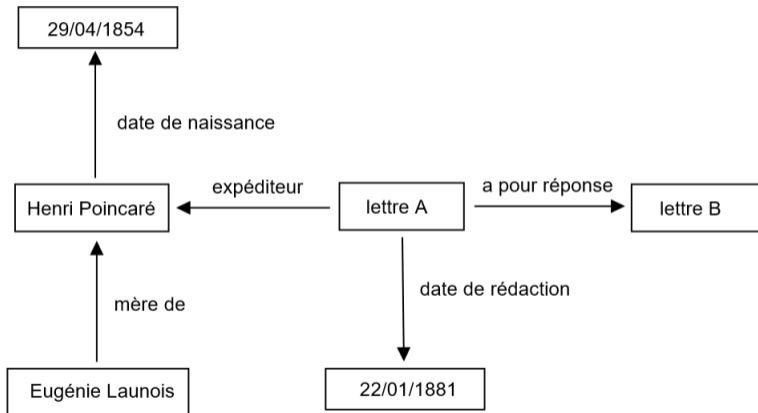
Des problématiques de représentation de connaissances

Notre choix : le modèle *n-ary*



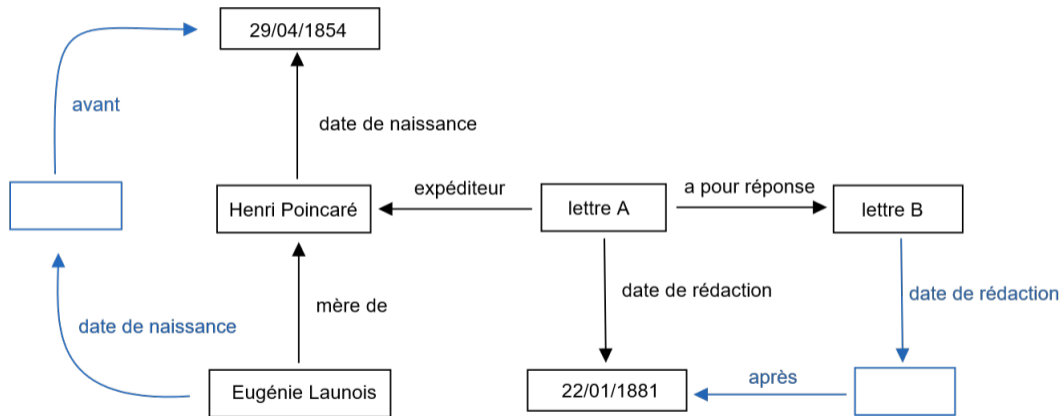
Des problématiques de représentation de connaissances

Temporalité et règles d'inférences



Des problématiques de représentation de connaissances

Temporalité et règles d'inférences



Des problématiques de représentation de connaissances

Incertitude et imprécision

Incertitude

- ▶ Il est fort probable que la lettre ait été rédigée par Henri Poincaré juste après sa prise de fonctions à la faculté des sciences de Paris au début de l'année 1881...

Des problématiques de représentation de connaissances

Incertitude et imprécision

Incertitude

- ▶ Il est fort probable que la lettre ait été rédigée par Henri Poincaré juste après sa prise de fonctions à la faculté des sciences de Paris au début de l'année 1881...
- ▶ ... mais peut-être que sa rédaction a eu lieu quelques mois plus tôt à Caen.

Des problématiques de représentation de connaissances

Incertitude et imprécision

Incertitude

- ▶ Il est fort probable que la lettre ait été rédigée par Henri Poincaré juste après sa prise de fonctions à la faculté des sciences de Paris au début de l'année 1881...
- ▶ ... mais peut-être que sa rédaction a eu lieu quelques mois plus tôt à Caen.

Imprécision

- ▶ L'événement s'est produit à la fin du *XIX*^{ème} siècle.
- ▶ La lettre a été rédigée dans une ville de l'est de la France.

Des outils pour explorer le corpus

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Exemple de requête

$Q =$ | Donner les lettres envoyées par Henri Poincaré en 1880
| traitant de topologie algébrique
| et citant Charles Hermite.

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Problématique

Des recherches flexibles peuvent être utiles pour :

- ▶ Remédier au problème de l'absence de résultats suite à une recherche
- ▶ Affiner certaines requêtes
- ▶ Identifier des liens entre certains éléments du corpus

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Problématique

Des recherches flexibles peuvent être utiles pour :

- ▶ Remédier au problème de l'absence de résultats suite à une recherche
- ▶ Affiner certaines requêtes
- ▶ Identifier des liens entre certains éléments du corpus

Mécanisme de transformation de requêtes

Proposition d'un outil s'appuyant sur la définition de règles de transformation de requêtes SPARQL :

- ▶ Des règles générales ;
- ▶ Des règles dépendantes d'un contexte.

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Requête initiale

$Q =$ | Donner les lettres envoyées par Henri Poincaré en 1880
| traitant de topologie algébrique
| et citant Charles Hermite.

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Requête initiale

$Q =$ | Donner les lettres envoyées par Henri Poincaré en 1880
traitant de **topologie algébrique**
et citant Charles Hermite.

Requête générée (en généralisant le thème)

$Q_1 =$ | Donner les lettres envoyées par Henri Poincaré en 1880
traitant de **mathématiques**
et citant Charles Hermite.

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Requête initiale

$Q =$ | Donner les lettres envoyées par Henri Poincaré en 1880
traitant de topologie algébrique
et citant Charles Hermite.

Requête générée (en étendant les bornes temporelles)

$Q_2 =$ | Donner les lettres envoyées par Henri Poincaré
entre 1878 et 1882 traitant de topologie
algébrique et citant Charles Hermite.

Des outils pour explorer le corpus

Un mécanisme pour les recherches flexibles

Requête initiale

$Q =$ | Donner les lettres envoyées par Henri Poincaré en 1880
traitant de topologie algébrique
et citant Charles Hermite.

Requête générée (en remplaçant une personne citée)

$Q_3 =$ | Donner les lettres envoyées par Henri Poincaré en 1880
traitant de topologie algébrique
et citant Paul Appell.

Des outils pour explorer le corpus

Utilisation de ce mécanisme

Plusieurs applications

- ▶ Un éditeur de données RDF qui propose une auto-complétion s'appuyant sur les données déjà éditées.
- ▶ Un système de recherche par formulaire qui facilite l'accès aux informations temporelles.
- ▶ Un outil d'exploration du corpus qui, à partir d'une lettre initiale, permet de naviguer dans le corpus en recherchant des lettres liées.

Quelques liens utiles

- ▶ Vidéo de présentation de l'outil d'édition disponible <https://videos.ahp-numerique.fr/videos/watch/0d544e5b-b4be-423e-9497-216f29ab44f3>
- ▶ Vidéo longue de présentation de l'outil d'exploration disponible <https://videos.ahp-numerique.fr/videos/watch/0d544e5b-b4be-423e-9497-216f29ab44f4>

Merci de votre attention !

`nicolas.lasolle@univ-lorraine.fr`

`pierre.willaime@univ-lorraine.fr`