

Sémantiser Omeka S : pourquoi et comment ?

Nicolas Lasolle^{1,2}, Pierre Willaime¹

¹ Université de Lorraine, CNRS, Université de Strasbourg, AHP-PRéST, F-54000 Nancy, France

² Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Journées Omeka, 13 et 14 octobre 2020, Nancy



Plan

Le projet

Les limites sémantiques d'Omeka S

Quelques solutions de contournement

Tirer partie du Web sémantique pour l'exploration du corpus

Le projet

Les limites sémantiques d'Omeka S

Quelques solutions de contournement

Tirer partie du Web sémantique pour l'exploration du corpus

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)
- ▶ en parallèle d'une édition papier

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)
- ▶ en parallèle d'une édition papier
- ▶ un collectif de chercheurs au départ

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)
- ▶ en parallèle d'une édition papier
- ▶ un collectif de chercheurs au départ
- ▶ petit à petit rejoint par de plus en plus d'ingénieurs

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)
- ▶ en parallèle d'une édition papier
- ▶ un collectif de chercheurs au départ
- ▶ petit à petit rejoint par de plus en plus d'ingénieurs

Le projet henripoincare.fr

- ▶ un projet de longue date (2002)
- ▶ une édition numérique de la correspondance de Henri Poincaré (incluant un apparat critique pour chaque lettre)
- ▶ en parallèle d'une édition papier
- ▶ un collectif de chercheurs au départ
- ▶ petit à petit rejoint par de plus en plus d'ingénieurs



LA CORRESPONDANCE D'HENRI POINCARÉ

Accueil Lettres Recherche Index

LETTRE : Gösta Mittag-Leffler à Henri Poincaré - 11 avril 1881

Transcription Métadonnées Citer ce document

Helsingfors 11 Avril 1881
Finlande

Monsieur,

Permettez-moi de envoyer sous bande quelques exemplaires d'un mémoire que notre maitre vénéré **M. Hermite** a voulu bien me faire l'honneur de publier ici.¹ **M. Hermite** y communique entre autres choses un résultat qui m'a paru d'un très grand intérêt.² Le résultat m'intéresse autant plus comme je me suis occupé depuis longtemps avec des fonctions d'une nature analogue avec la votre. Voyez là dessus quelques mots dans une lettre de moi à **M. Hermite**³ publiée dans les annales de Monsieur **Darboux**.⁴

Le projet

Pourquoi Omeka ?

- ▶ logiciel libre

Le projet

Pourquoi Omeka ?

- ▶ logiciel libre
- ▶ modulable et modifiable

Le projet

Pourquoi Omeka ?

- ▶ logiciel libre
- ▶ modulable et modifiable
- ▶ **prise en main très simple - les membres du projets peuvent tous modifier les métadonnées**

Le projet

Pourquoi Omeka ?

- ▶ logiciel libre
- ▶ modulable et modifiable
- ▶ **prise en main très simple - les membres du projets peuvent tous modifier les métadonnées**
- ▶ adapté à un corpus dont la description se fait au fil de l'eau

Le projet

Pourquoi Omeka ?

- ▶ logiciel libre
- ▶ modulable et modifiable
- ▶ **prise en main très simple - les membres du projets peuvent tous modifier les métadonnées**
- ▶ adapté à un corpus dont la description se fait au fil de l'eau
- ▶ ce qui n'est pas le cas des chaînes de conversion (LaTeX → HTML ou TEI → HTML)

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)

Le projet

Pourquoi Omeka S ?

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

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables
- ▶ multi-sites (ce qui correspond aux objectifs scientifiques avec l'extension du projet au delà de la correspondance).

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables
- ▶ multi-sites (ce qui correspond aux objectifs scientifiques avec l'extension du projet au delà de la correspondance).
- ▶ interrogations poussées sur le corpus (fouille) avec les outils du Web sémantique

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables
- ▶ multi-sites (ce qui correspond aux objectifs scientifiques avec l'extension du projet au delà de la correspondance).
- ▶ interrogations poussées sur le corpus (fouille) avec les outils du Web sémantique
- ▶ générer des métadonnées automatiquement à partir d'inférences

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables
- ▶ multi-sites (ce qui correspond aux objectifs scientifiques avec l'extension du projet au delà de la correspondance).
- ▶ interrogations poussées sur le corpus (fouille) avec les outils du Web sémantique
- ▶ générer des métadonnées automatiquement à partir d'inférences
- ▶ détecter des incohérences

Le projet

Pourquoi Omeka S ?

- ▶ gestion native des relations entre contenus (RDF)
- ▶ utilisation native de schémas de métadonnées interopérables
- ▶ multi-sites (ce qui correspond aux objectifs scientifiques avec l'extension du projet au delà de la correspondance).
- ▶ interrogations poussées sur le corpus (fouille) avec les outils du Web sémantique
- ▶ générer des métadonnées automatiquement à partir d'inférences
- ▶ détecter des incohérences
- ▶ créer des outils qui exploitent le caractère sémantique des données dans Omeka S (visualisation par exemple).

Plan

Le projet

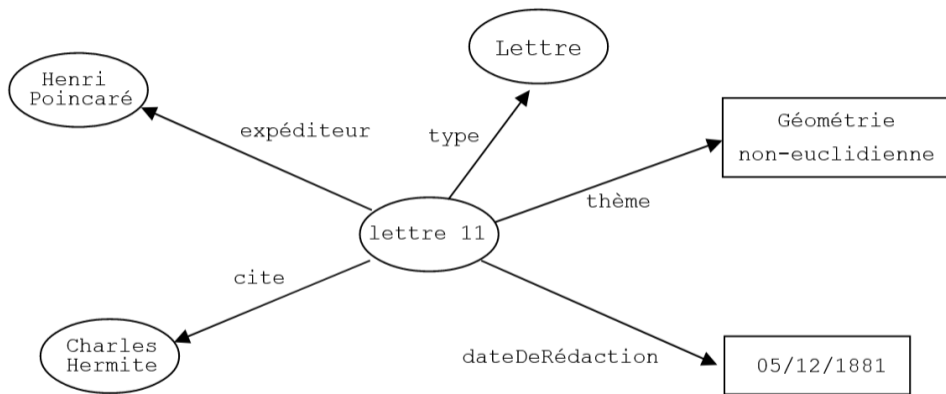
Les limites sémantiques d'Omeka S

Quelques solutions de contournement

Tirer partie du Web sémantique pour l'exploration du corpus

Les limites sémantiques d'Omeka S

Modèle fondé sur un graphe



Dans la logique du modèle *RDF*¹.

1. *Resource Description Framework*.

Les limites sémantiques d'Omeka S

Import d'ontologies























Vocabulaires

Importer un nouveau vocabulaire

1 de 1 < > 1-8 sur 8

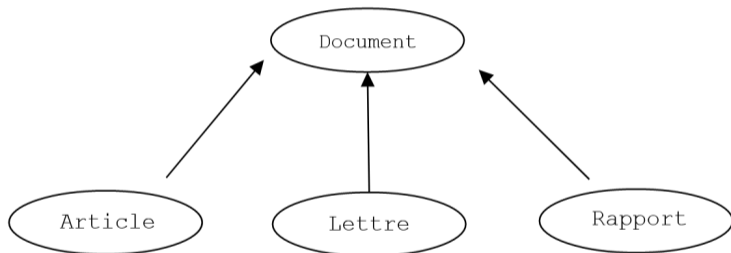
Libellé ⌵ Croissant ⌵ Trier

Libellé		Préfixe	Classes	Propriétés
Archives Henri Poincaré ontology	  	ahpo	17	60
Bibliographic Ontology	  	bibo	58	67
BIO	  	bio	42	33
Dublin Core		dcterms	22	55
Dublin Core Type		dctype	12	0
Exif format picture data	  	exif	1	160
Friend of a Friend	  	foaf	13	62
Relationship	  	rel	1	34

1 de 1 < > 1-8 sur 8

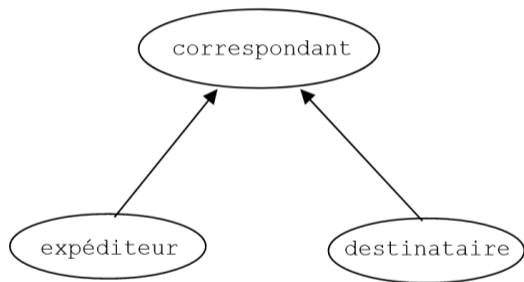
Les limites sémantiques d'Omeka S

Hiérarchie entre les classes



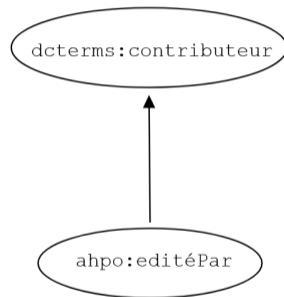
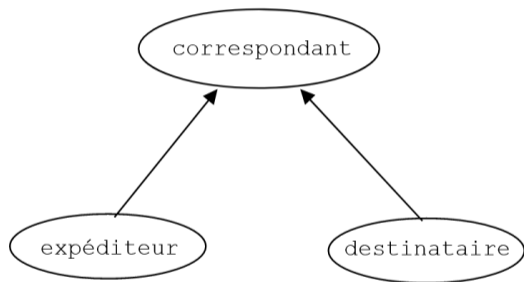
Les limites sémantiques d'Omeka S

Hiérarchie entre les propriétés



Les limites sémantiques d'Omeka S

Hiérarchie entre les propriétés



Les limites sémantiques d'Omeka S

Hiérarchie entre les propriétés

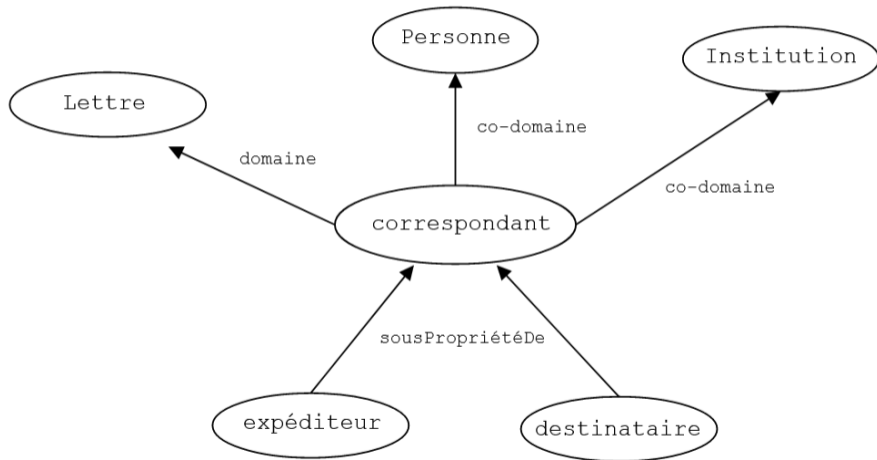
VOCABULAIRES

Dublin Core • Propriétés

Créateur	...	dcterms:creator
Date	...	dcterms:date
Date d'acceptation	...	dcterms:dateAccepted
Date de disponibilité	...	dcterms:available
Date du copyright	...	dcterms:dateCopyrighted
Date de création	...	dcterms:created
Date de publication	...	dcterms:issued
Date de modification	...	dcterms:modified
Date de soumission	...	dcterms:dateSubmitted
Date de validité	...	dcterms:valid

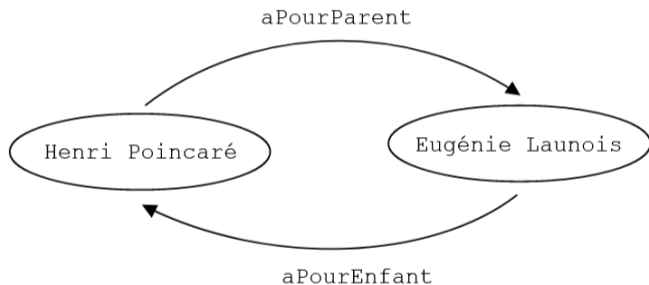
Les limites sémantiques d'Omeka S

Domaine et co-domaine



Les limites sémantiques d'Omeka S

Propriétés inverses



Recherche dans les transcriptions et l'apparat critique

Omeka S permet l'intégration d'un module de recherche fondé sur l'outil Solr².

$Q =$ | Donne-moi les lettres dont la transcription
contient le terme « espaces lacunaires ».

2. <https://lucene.apache.org/solr/>.

LETTRE : Édouard Goursat à Henri Poincaré - 27 mars 1882

[Transcription](#)[Métadonnées](#)[Citer ce document](#)

Toulouse, 27 Mars 1882¹

Monsieur,

Je reçois à l'instant un exemplaire de votre travail sur les fonctions à **espaces lacunaires**². Comme je ne suis pas encore très au courant de tout ce qui se publie en Mathématiques, surtout de ce qui se publie à l'étranger, il m'arrive quelquefois de faire des découvertes comme celle-là. Je vous prie de croire que, lorsque j'ai adressé ma note à M^r **Hermite**, je n'avais aucune connaissance de votre méthode ; je regrette que M^r **Hermite**³ ne m'ait pas prévenu immédiatement au lieu de communiquer à l'Académie des Sciences une démonstration qui me paraît, quant au fond, identique à la vôtre⁴. En terminant cette lettre, permettez-moi de vous dire que, tout en regrettant de m'être laissé devancer dans cette voie, je ne puis m'empêcher d'être fier de m'être rencontré avec vous⁵.

Veillez agréer, Monsieur et cher collègue, l'assurance de mes sentiments distingués.

E. Goursat

Les limites sémantiques d'Omeka S

Pas d'utilisation du langage SPARQL



Recherche à partir des métadonnées

Le langage SPARQL permet de formuler des requêtes plus expressives.

$Q =$ | Donne-moi les lettres envoyées par Henri Poincaré
à Gösta Mittag-Leffler entre 1880 et 1890
traitant de fonctions à espaces lacunaires
et citant Charles Hermite.

Plan

Le projet

Les limites sémantiques d'Omeka S

Quelques solutions de contournement

Tirer partie du Web sémantique pour l'exploration du corpus

Solutions de contournement

Sémantiser Omeka

1. conversion de la base de donnée d'Omeka S (JSON) en triplets RDF en filtrant au passage

Solutions de contournement

Sémantiser Omeka

1. conversion de la base de donnée d'Omeka S (JSON) en triplets RDF en filtrant au passage

Solutions de contournement

Sémantiser Omeka

1. conversion de la base de donnée d'Omeka S (JSON) en triplets RDF en filtrant au passage
↳ un script python qui interroge l'API d'Omeka S

Solutions de contournement

Sémantiser Omeka

1. conversion de la base de donnée d'Omeka S (JSON) en triplets RDF en filtrant au passage
↳ un script python qui interroge l'API d'Omeka S
2. réappliquer l'ontologie sur la nouvelle base pour restaurer les fonctionnalités sémantiques perdues avec Omeka S

Solutions de contournement

Sémantiser Omeka

1. conversion de la base de donnée d'Omeka S (JSON) en triplets RDF en filtrant au passage
↳ un script python qui interroge l'API d'Omeka S
2. réappliquer l'ontologie sur la nouvelle base pour restaurer les fonctionnalités sémantiques perdues avec Omeka S
3. mise en place d'un SPARQL Endpoint

Solutions de contournement

API Omeka S

```
{ "@context": "http://henripoincare.fr/api-context", "@id": "http://henripoincare.fr/api/items/843", "@type": "http://henripoincare.fr/api/users/9", "o:id": 9, "o:resource_class": { "@id": "http://henripoincare.fr/api/resource_templates/8", "o:id": 8 }, "o:thumbnail": null, "o:title": "Henri Poincar\u00e9", "o:created": { "@value": "2019-12-13T08:01:33+00:00", "@type": "http://www.w3.org/2001/XMLSchema#dateTime", "o:media": [], "o:parent": { "@id": "http://henripoincare.fr/api/item_sets/11", "o:id": 11 }, {"@id": "http://henripoincare.fr/api/item_sets/12", "o:id": 12 }, {"@id": "http://henripoincare.fr/api/item_sets/14481", "o:id": 14481 } ], "dcterms:title": [ { "type": "literal", "property_id": 1931, "property_label": "pr\u00e9nom", "is_public": true, "@value": "Henri" }, {"type": "literal", "property_id": 1929, "property_label": "nomDeFamille", "is_public": true, "@value": "Poincar\u00e9" }, {"type": "valuesuggest:geonames:geonames", "property_id": 1908, "property_label": "vit\u00e9", "is_public": true, "@value": "Paris" }, {"type": "resource", "property_id": 222, "property_label": "Child Of", "is_public": true, "@id": "http://henripoincare.fr/api/items/497", "value_resource_id": 497, "value_resource_name": "items", "url": null, "display_title": "\u00c9mile-L\u00e9on Poincar\u00e9" }, {"type": "resource", "property_id": 221, "property_label": "Sibling Of", "is_public": true, "@id": "http://henripoincare.fr/api/items/3545", "value_resource_id": 3545, "value_resource_name": "items", "url": null, "display_title": "Aline Poincar\u00e9" }, {"type": "resource", "property_id": 33, "property_label": "Is Part Of", "is_public": true, "@id": "http://henripoincare.fr/api/items/667", "value_resource_id": 667, "value_resource_name": "items", "url": null, "display_title": "Famille Poincar\u00e9" } ] }
```

Solutions de contournement

triplets en turtle

```
<http://henripoincare.fr/api/items/843> a ahpo:Person ;
  rdfs:label "Henri Poincaré" ;
  ahpo:familyName "Poincaré" ;
  ahpo:firstName "Henri" ;
  ahpo:liveAt <http://www.geonames.org/2988507> ;
  hp:link "http://henripoincare.fr/s/correspondance/item/843" ;
  o:item_set <http://henripoincare.fr/api/item_sets/11>,
    <http://henripoincare.fr/api/item_sets/12>,
    <http://henripoincare.fr/api/item_sets/14481>,
    <http://henripoincare.fr/api/item_sets/22>,
    <http://henripoincare.fr/api/item_sets/8> ;
  dcterms:isPartOf <http://henripoincare.fr/api/items/12603> ;
  dcterms:title "Henri Poincaré" ;
  rel:childOf <http://henripoincare.fr/api/items/365>,
    <http://henripoincare.fr/api/items/497> ;
  rel:engagedTo <http://henripoincare.fr/api/items/667> ;
  rel:siblingOf <http://henripoincare.fr/api/items/3545> .
```

Recherche SPARQL

Options

- recherche sur la base recherche sur une base rdf personnalisée

La requête SPARQL

Tapez une requête SPARQL

```
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix time: <http://www.w3.org/2006/time#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix o: <http://omeka.org/s/vocabs/o#> .
@prefix cnt: <http://www.w3.org/2011/content#> .

SELECT DISTINCT ?lien ?sujet ?year WHERE {
  ?item rdf:type ahpo:Letter .
  ?item ahpo:sentTo ?dest .
  ?dest ahpo:familyName "Poincaré" .
  ?item ahpo:sentBy ?exp .
  ?exp ahpo:familyName "Craig" .
  ?item dcterms:title ?sujet .
  ?item hp:year ?year .
  ?item hp:link ?lien .
  FILTER(?year > 1887 && ?year < 1897 ) .
} order by asc(?year)
```

Solutions de contournement

SPARQL Endpoint

Recherche par formulaire

Recherche SPARQL

Recherche par propriétés

Administration

Recherche sur la base

Options de recherche

insensible à la casse
[ici](#)

utiliser les expressions régulières (un guide sur les expressions régulières est disponible

Type de document

lettre
publication

article

livre

journal

chapitre

rapport

thèse

Options pour les lettres

Mode de recherche par correspondant

Mode de recherche directionnelle

Correspondant

Nom :

Prénom :

indifférent

Allemagne
Arménie
Autriche

indifférent

anatomiste
astronome
astrophysicien

Solutions de contournement

Interopérabilité

- ▶ le moissonnage du site est possible

Solutions de contournement

Interopérabilité

- ▶ le moissonnage du site est possible
- ▶ mais il faut que le moissonneur « connaisse » les ontologies utilisées

Solutions de contournement

Interopérabilité

- ▶ le moissonnage du site est possible
- ▶ mais il faut que le moissonneur « connaisse » les ontologies utilisées
- ▶ lorsqu'on a des besoins spécifiques, la seule solution « propre » est de créer sa propre ontologie, alignée sur d'autres et qui vient combler les « trous »

Solutions de contournement

Interopérabilité

- ▶ le moissonnage du site est possible
- ▶ mais il faut que le moissonneur « connaisse » les ontologies utilisées
- ▶ lorsqu'on a des besoins spécifiques, la seule solution « propre » est de créer sa propre ontologie, alignée sur d'autres et qui vient combler les « trous »
- ▶ et tenter de la faire reconnaître par les outils de moissonnage (Isidore et cie)

Plan

Le projet

Les limites sémantiques d'Omeka S

Quelques solutions de contournement

Tirer partie du Web sémantique pour l'exploration du corpus

Aller plus loin avec le Web sémantique

Édition de données

Édition du corpus

L'édition de données est une tâche parfois fastidieuse avec un important risque d'erreurs :

- ▶ Erreur de duplication ;
- ▶ Ambiguïté par rapport aux données ;
- ▶ Erreur de saisie.

Aller plus loin avec le Web sémantique

Édition de données

Édition du corpus

L'édition de données est une tâche parfois fastidieuse avec un important risque d'erreurs :

- ▶ Erreur de duplication ;
- ▶ Ambiguïté par rapport aux données ;
- ▶ Erreur de saisie.

L'idée : un outil exploitant les ressources déjà éditées pour assister l'utilisateur dans cette tâche.

Aller plus loin avec le Web sémantique

Recherche approfondie

Problématique

L'exécution d'une requête peut entraîner plusieurs situations :

- ▶ Trop de résultats ;
- ▶ Pas assez de résultats ;
- ▶ Des résultats qui n'apportent pas de réponse au problème de recherche : on souhaite explorer de nouvelles pistes.

Aller plus loin avec le Web sémantique

Recherche approchée

Problématique

L'exécution d'une requête peut entraîner plusieurs situations :

- ▶ Trop de résultats ;
- ▶ Pas assez de résultats ;
- ▶ Des résultats qui n'apportent pas de réponse au problème de recherche : on souhaite explorer de nouvelles pistes.

Mécanisme de transformation de requête

Un outil fonctionne en s'appuyant sur la définition de règles de transformation de requêtes SPARQL :

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

Aller plus loin avec le Web sémantique

Recherche approchée

Requête initiale

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

Aller plus loin avec le Web sémantique

Recherche approchée

Requête initiale

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

Requête générée (en généralisant le thème)

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

Aller plus loin avec le Web sémantique

Recherche approchée

Requête initiale

$$Q = \left| \begin{array}{l} \text{Donne-moi les lettres envoyées par Henri Poincaré en 1880} \\ \text{traitant de topologie algébrique} \\ \text{et citant Charles Hermite.} \end{array} \right.$$

Requête générée (en étendant les bornes temporelles)

$$Q_2 = \left| \begin{array}{l} \text{Donne-moi les lettres envoyées par Henri Poincaré} \\ \text{entre 1878 et 1882 traitant de topologie} \\ \text{algébrique et citant Charles Hermite.} \end{array} \right.$$

Aller plus loin avec le Web sémantique

Recherche approchée

Requête initiale

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

Requête générée (en remplaçant une personne citée)

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

Merci !

`pierre.willaime@univ-lorraine.fr`

`nicolas.lasolle@univ-lorraine.fr`