A Semantic Web Navigation Tool for Exploring the Henri Poincaré Correspondence Corpus

Nicolas Lasolle[1,2], Olivier Bruneau[1], Jean Lieber[2], Philippe Nabonnand[1], Laurent Rollet[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

September, the 20th

UNIVERSITÉ DE LORRAINE

SWODCH 2021

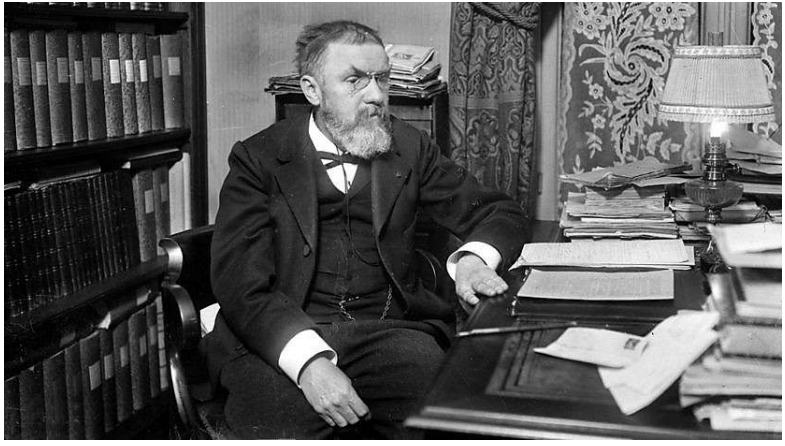
Context

A navigation tool for the Henri Poincaré correspondence corpus

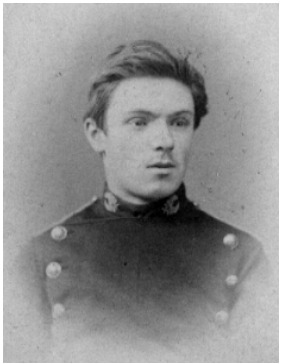
- Exploitation of a historical corpus with Semantic Web technologies
- Presentation of a navigation system designed for exploring the corpus graph
- System which relies on the use of a flexible search mechanism to enhance similarities between resources

System available for exploring other RDF corpora

The Henri Poincaré correspondence corpus

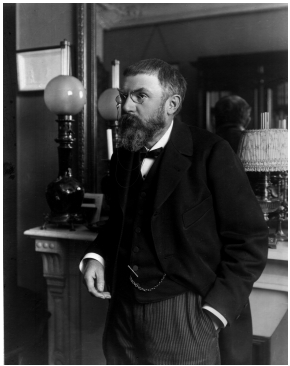


Henri Poincaré (1854-1912)



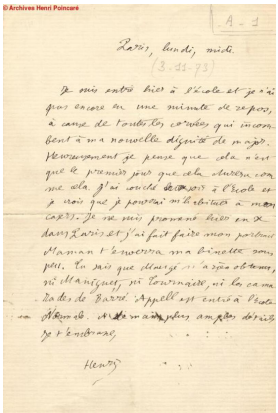
1873

1887



1908

The Henri Poincaré correspondence corpus

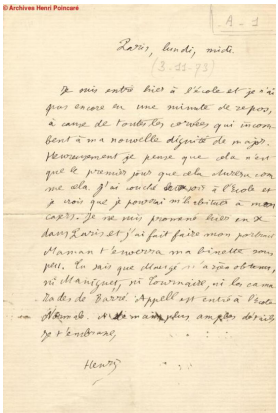


Henri Poincaré to Aline Poincaré, 1873

Online publishing

- Around 2000 letters (private, administrative and scientific exchanges)

The Henri Poincaré correspondence corpus

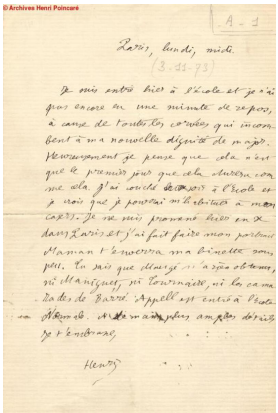


Henri Poincaré to Aline Poincaré, 1873

Online publishing

- Around 2000 letters (private, administrative and scientific exchanges)
- Rich source of historical information
  - Scientific theories
  - Academies and administrations
  - Social, political and cultural context

The Henri Poincaré correspondence corpus



Henri Poincaré to Aline Poincaré, 1873

Online publishing

- Around 2000 letters (private, administrative and scientific exchanges)
- Rich source of historical information
  - Scientific theories
  - Academies and administrations
  - Social, political and cultural context
- Semantic Web technologies used to structure and to exploit data

A website dedicated to the work and correspondence of Henri Poincaré[1]

LA CORRESPONDANCE D'HENRI POINCARÉ

Accueil | Lettres | Recherche | Index

LETTRE : Henri Poincaré à Aline Boutroux - 3 novembre 1873

Modifier l'item

Transcription | Métadonnées | Citer ce document

[3 novembre 1873[1]]

Paris, lundi, midi.

Je suis entré hier à l'école[2] et je n'ai pas encore eu une minute de repos, à cause de toutes les nouvelles corvées qui incombent à ma nouvelle dignité de major. Heureusement je pense que cela n'est que le premier jour que cela durera comme cela. J'ai couché hier soir à l'École et je crois que je pourrai m'habituer à mon casern. Je me suis promené hier en X[3] dans Paris et j'ai fait faire mon portrait[4]. Maman t'enverra ma binette sous peu. Tu sais que Maugé[5] n'a rien obtenu, ni Maniguet[6], ni Tournaire, ni les camarades de Barré[7]. Appell est entré à l'École Normale[8]. À demain plus amples détails.

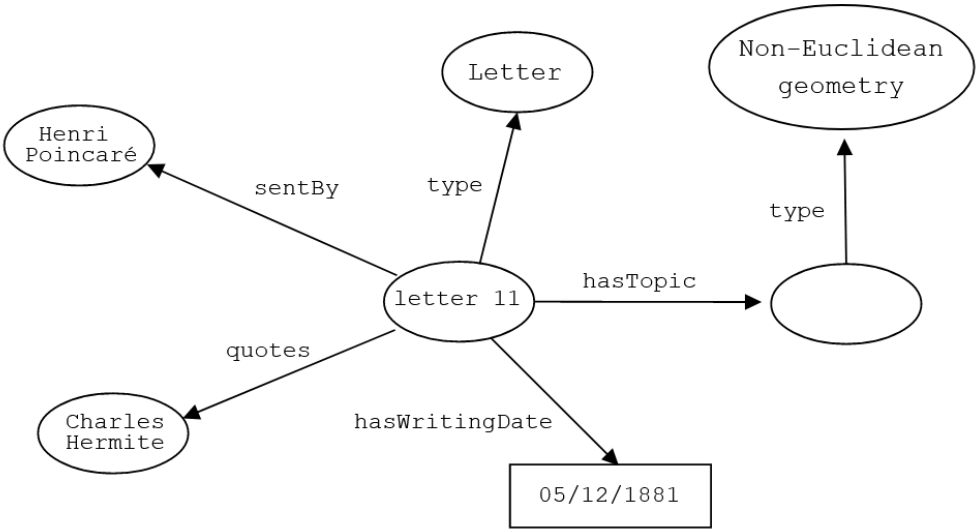
Je t'embrasse,

Henri

1. Comme de nombreuses autres, cette lettre n'avait pas de destinataire précis. La référence à Eugénie Poincaré laisse supposer qu'elle s'adressait plutôt à la sœur de Poincaré, Aline. On sait que toute la famille avait accompagné le jeune polytechnicien à Paris pour sa rentrée. Il est est probable qu'Eugénie Poincaré était restée quelques jours à Paris tandis qu'Aline et Émile-Léon Poincaré étaient retournés à Nancy.↩

[1] `henripoincare.fr`

Resource Description Framework (RDF)



An ontology based on RDF Schema (RDFS)

RDFS

The RDFS language is used for representing domain knowledge

The Archives Henri-Poincaré ontology (AHPo)

- An ontology created for corpus resources representation
- which relies on several standards vocabularies (dcterms, foaf, rel, bibo, etc.)
- which is still evolving but which should be published soon

Corpus querying

SPARQL query language used for querying the corpus.

$\mathcal{Q} =$ Give the letters sent by `Henri Poincaré` to `Gösta Mittag-Leffler` between 1880 and 1895 having `optics` as a topic.

$\mathtt{Q} =$

```
SELECT ?l
WHERE { ?l a letter .
        ?l sentBy henriPoincaré . ?l sentTo göstaMittagLeffler .
        ?l hasTopic optics . ?l hasWritingDate ?y .
        FILTER(YEAR(?y) >= 1880 AND YEAR(?y) <= 1895)}
```

A navigation system for exploring the RDF graph

Issue and general idea

Querying the corpus with SPARQL?

This approach leads to three difficulties

Issue and general idea

Querying the corpus with SPARQL?

This approach leads to three difficulties

1. Knowledge of SPARQL syntax necessary

Issue and general idea

Querying the corpus with SPARQL?

This approach leads to three difficulties

1. Knowledge of SPARQL syntax necessary
2. Good understanding of the ontology schema required

Issue and general idea

Querying the corpus with SPARQL?

This approach leads to three difficulties

1. Knowledge of SPARQL syntax necessary
2. Good understanding of the ontology schema required
3. Exact results sometimes do not encourage the exploration aspect when studying the corpus

Issue and general idea

Querying the corpus with SPARQL?

This approach leads to three difficulties

1. Knowledge of SPARQL syntax necessary
2. Good understanding of the ontology schema required
3. Exact results sometimes do not encourage the exploration aspect when studying the corpus

Our proposal

A dynamic tool proposing a visual exploration of the corpus which allows to navigate through resources

Presentation of the navigation system

General idea

- Start with an initial resource (e.g. a letter)
- Extract relevant conditions for creating filters
- Generate queries using these filtering conditions
- Results are presented in a chronological-based view
- Focus can be changed from a resource to another

Presentation of the navigation system

General idea

- Start with an initial resource (e.g. a letter)
- Extract relevant conditions for creating filters
- Generate queries using these filtering conditions
- Results are presented in a chronological-based view
- Focus can be changed from a resource to another

For the Henri Poincaré correspondence corpus

Easily identify letters sent during the same temporal period, having the same correspondent, quoting the same people or institutions.

Demonstration

SQTRL

A tool associated with a language, named **SQTRL (SPARQL Query Transformation Rule Language)** has been designed to enable flexible querying.

SQTRL

A tool associated with a language, named **SQTRL (SPARQL Query Transformation Rule Language)** has been designed to enable flexible querying.

Transformation rules

Different kind of rules can be created:

- Application-independent (generalization, specialization, etc.);
- Application-dependent (switching sender/recipient of the letter, replacing someone by a colleague, etc.).

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Generated query (by generalizing the topic)

$$\mathcal{Q}_1 = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{mathematics} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Generated query (by extending the temporal bounds)

$$\mathcal{Q}_2 = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1898 and 1907,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Generated query (by switching the sender and the recipient)

$$\mathcal{Q}_3 = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Giovanni Battista Guccia} \text{ to} \\ \texttt{Henri Poincaré} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Generated query (by replacing Guccia)

$$\mathcal{Q}_4 = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{an Italian} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

A flexible querying system

Initial query

$$\mathcal{Q} = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{Giovanni Battista Guccia} \text{ between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Generated query (by replacing Guccia)

$$\mathcal{Q}_5 = \left| \begin{array}{l} \text{Give the letters sent by } \texttt{Henri Poincaré} \text{ to} \\ \texttt{a member of the } \textit{Circolo Matematico di Palermo} \\ \text{between 1900 and 1905,} \\ \text{having } \texttt{geometry} \text{ as a topic and quoting a } \texttt{scientist}. \end{array} \right.$$

Reusability of the navigation system?

System reusability

Conditions for reusing the tool with another corpus

1. RDF data exposed through a public or local SPARQL endpoint

System reusability

Conditions for reusing the tool with another corpus

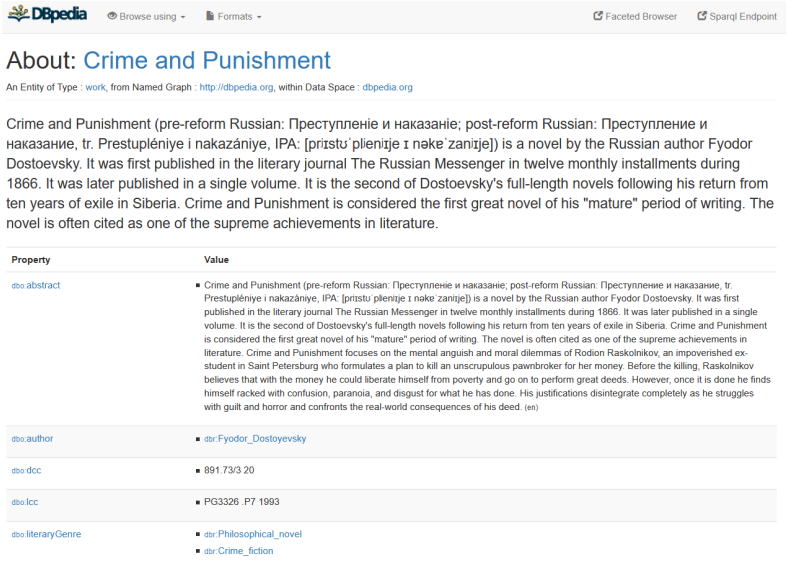
1. RDF data exposed through a public or local SPARQL endpoint
2. Results which can be presented using a chronological-based view

System reusability

Conditions for reusing the tool with another corpus

1. RDF data exposed through a public or local SPARQL endpoint
2. Results which can be presented using a chronological-based view
3. Resources described with labels:
   e.g. `http://e-hp.ahp-numerique.fr/ahpo#/843` associated to `Henri Poincaré`

Several DBpedia use cases

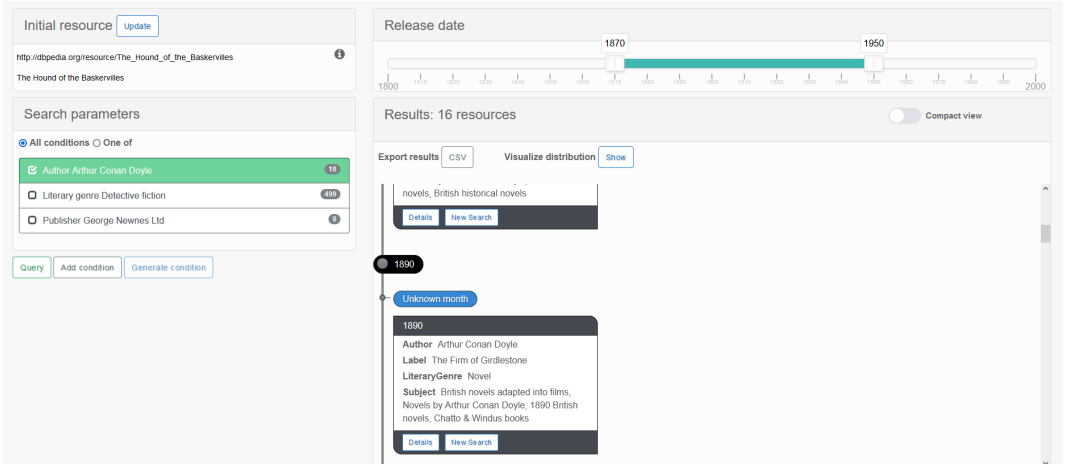


Several DBpedia use cases

Literary works

Works by the same author, associated with the same movement, dealing with the same themes, written in the same language, etc.

The Hound of the Baskervilles, by Sir Arthur Conan Doyle



Several DBpedia use cases

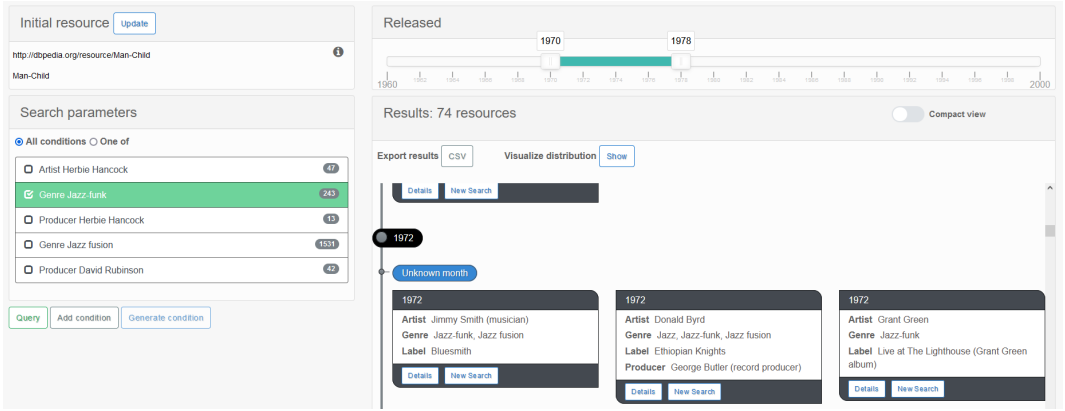
Literary works

Works by the same author, associated with the same movement, dealing with the same themes, written in the same language, etc.

Musical albums

Works of the same artist, associated with the same musical genre, produced by the same individual, etc.

Man-child, by Herbie Hancock



Several DBpedia use cases

Literary works

Works by the same author, associated with the same movement, dealing with the same themes, written in the same language, etc.

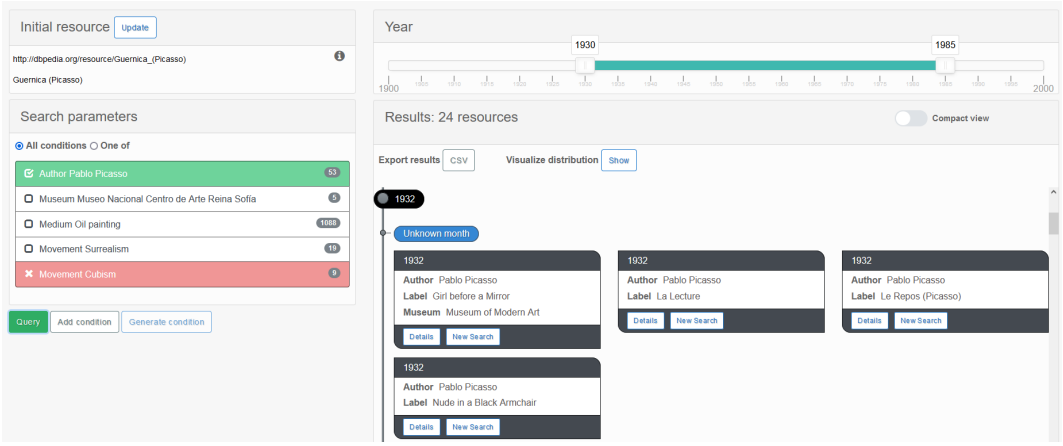
Musical albums

Works of the same artist, associated with the same musical genre, produced by the same individual, etc.

Painting

Paintings by the same artist, associated with the same movement, preserved in the same institution, etc.

Guernica, by Pablo Picasso



System reusability

Configuration file

- SPARQL endpoint URL
- Bounds of the time interval linked to the filter
- Ontology properties to use for:
  - chronological presentation and filtering;
  - generating conditions to search for similar resources;
  - finding the labels associated with the resources ;
  - specifying information to be displayed for each result.
- Main language to be used for labels

System reusability

GitHub[2] repository containing

- The application source code (Java backend and Web frontend);
- A deployed version of the backend application;
- Three versions of the configuration file;
- A user guide and technical documentation.

Any corpus suggestion for testing or collaboration?

[2] https://github.com/nlasolle/rdf_navigation_tool

Thank you for listening!

nicolas.lasolle@univ-lorraine.fr

SQTRL

```
<rule iri="http://sqtrl-rules/generic/1"
   label="Generalize object class">
   <context>?C rdfs:subClassOf ?D</context>
   <left>?x ?p ?C</left>
   <right>?x ?p ?D</right>
   <cost>5.0</cost>
   <exception>?C rdfs:subClassOf ?X . ?X rdfs:subClassOf ?D</exception>
   <exception>FILTER(?C != ?D)</exception>
   <explanation>Generalizing ?C in ?D</explanation>
</rule>
```

SQTRL