



Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du 19ème siècle: application aux métiers de la photographie

Solenn Tual ⁽¹⁾, Nathalie Abadie ⁽¹⁾, Bertrand Duménieu ⁽²⁾, Joseph Chazalon ⁽³⁾, Edwin Carlinet ⁽³⁾

Conférence Ingénierie des Connaissances, Strasbourg, 3 juillet 2023

Licence : CC BY 4.0



(1)



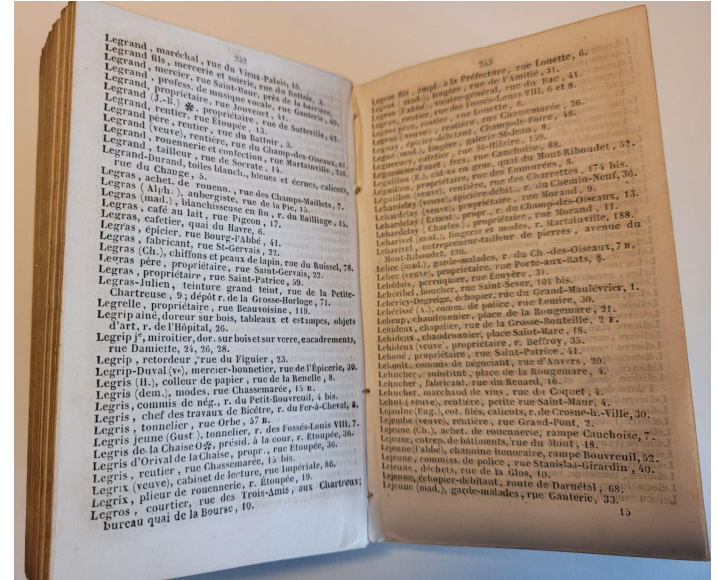
(2)



(3)

Annuaire du commerce de Paris

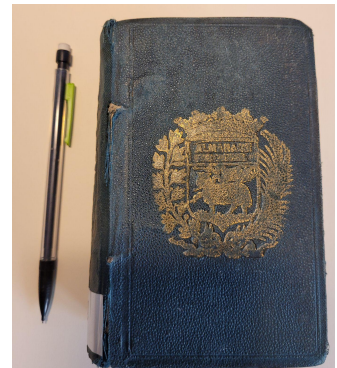
- Édités à partir de la fin du XVIIIème siècle jusqu'au milieu du XXème siècle par différents éditeurs.
- Organisés en différentes listes (alphabétique, par profession, par adresse).
- Chaque entrée contient :
 - Nom/Raison sociale
 - Activité/Profession/Statut
 - Adresse(s)
 - Distinctions militaires/professionnelles



Cherot, <i>joaillier</i> , S. Martin, 51.	Chevalier, <i>opt. du Roi</i> , Tour de l'Horl. du Palais, 1.
Chéroux, <i>orf.</i> Ste Avoye, 42.	Chevalier (L.), <i>opt. q. Horl. Pal.</i> 65.
Cherre, <i>layet</i> , Caire, 7.	Chevalier (Vinc.), <i>opt. q. Horl. Pal.</i> 69.
Cherré, <i>tap.</i> Moulins, 12.	Chevalier (Victor), <i>opt. q. Horl. Pal.</i> 77 b.
Cherrier, <i>tabl.</i> St Denis, 277.	

Liste ordonnée par nom - Annuaire Cambon Almgène de 1839.

Source : Gallica BNF



Annuaire du commerce de Paris

- Redondance importante des entrées d'une année ou d'une édition à une autre
- Snapshots temporels

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl que protestante (Société), Moulins, 16.
 Bibron, aide-natural., au Muséum d'hist. nat.
 Bibus, tailleur, Roule. 21.
 Bibus, tailleur, Richelieu, 31.
 Bical et Dorre, fab. de socques, Vertbois, 14.
 Bican (Mme), fondeur en cuivre, cour de la
 Corderie-du-Temple, 26.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl que protestante (Société), Moulins, 16.
 Bibron, aide-natural., au Muséum d'hist. nat.
 Bibus, tailleur, Roule. 21.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 Bical et Dorre, fab. de socques, Vertbois, 14.
 Bican (Vve) et fils, fondeur en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, marché d'Aguesseau, 15.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1842a - page 117

*Bibliothèque Ste-Geneviève, rue des Sept-Voies
 et place du Panthéon.*
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 Bical et Doiro, fab. de socques, Vert-Bois, 14.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, Marché-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1843a - page 129

Source : Gallica BNF

Des sources géographiques anciennes à l'analyse géohistorique

Reconnaissance

Classification

Géo-référencement

Structuration

Scan → Texte

Ravrio et comp., fabr. de bronzes et curiosités, r. Richelieu, 93; la fabrique rue Montmartre, 161.

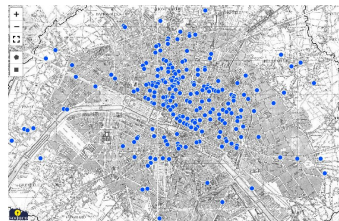
Segmentation des pages
+
Optical Character
Recognition (OCR)

Texte →
Entités nommées spatiales

Ravrio et comp. PER, fabr. de bronzes et curiosités ACT,
r. Richelieu LOC, 93 CARDINAL, la fabrique FT,
rue Montmartre LOC, 161 CARDINAL.

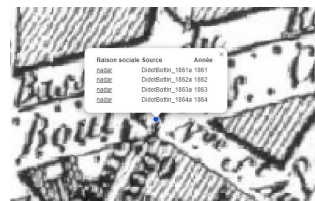
Reconnaissance des
Entités Nommées (NER)

Entités nommées spatiales →
Données géographiques



Géocodage

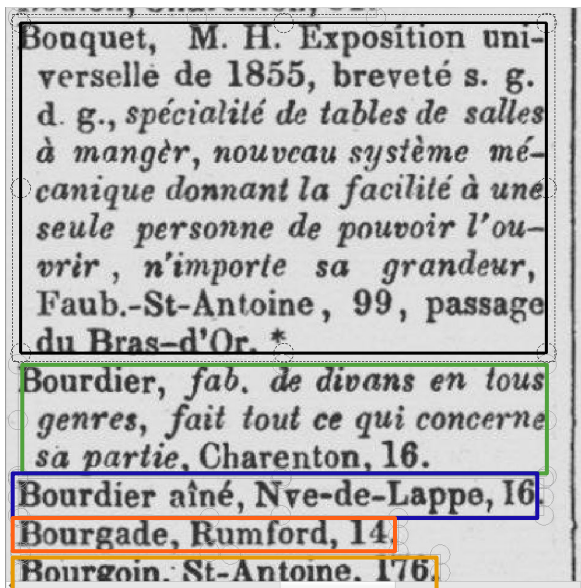
Données géographiques →
Données géohistoriques



Filtrage et liage des
entrées

Segmentation des entrées et reconnaissance optique des caractères (OCR)

Segmentation des pages en colonnes, paragraphes (XY cut), lignes (watershed) regroupées en entrées.

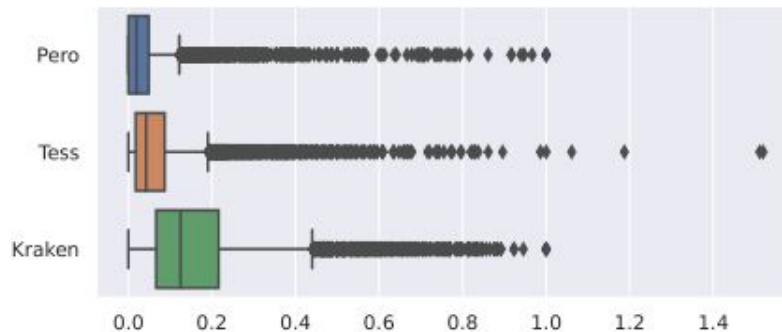


Source : Gallica BNF

Comparaison de 3 modèles d'OCR

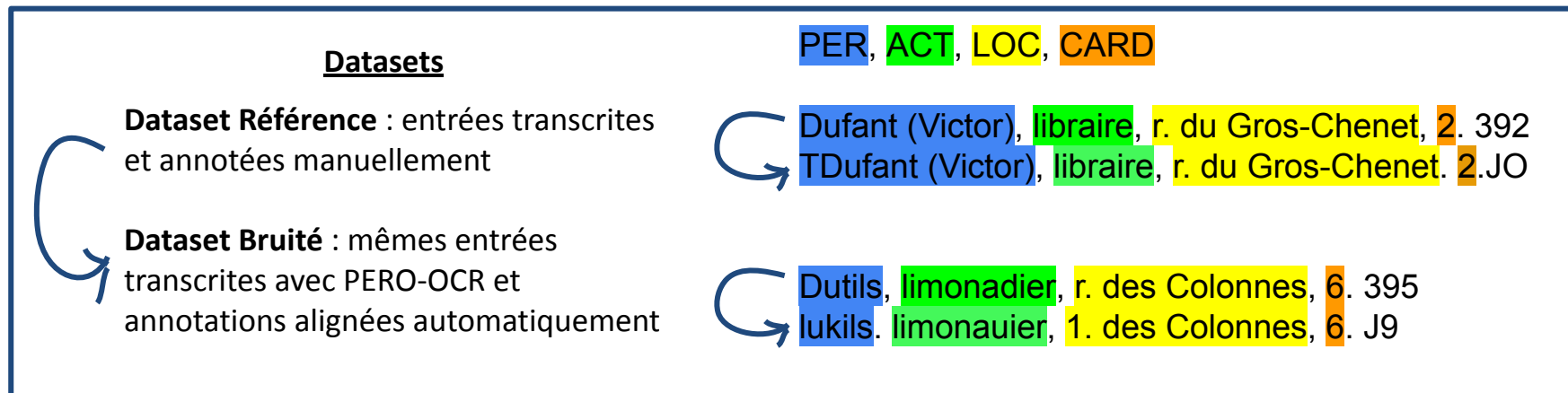
	PERO OCR	Tesseract	Kraken
CER	3.78%	6.56%	15.72%

CER : Character Error Rate



Abadie et al. (2022)

Reconnaissance des entités nommées (NER)



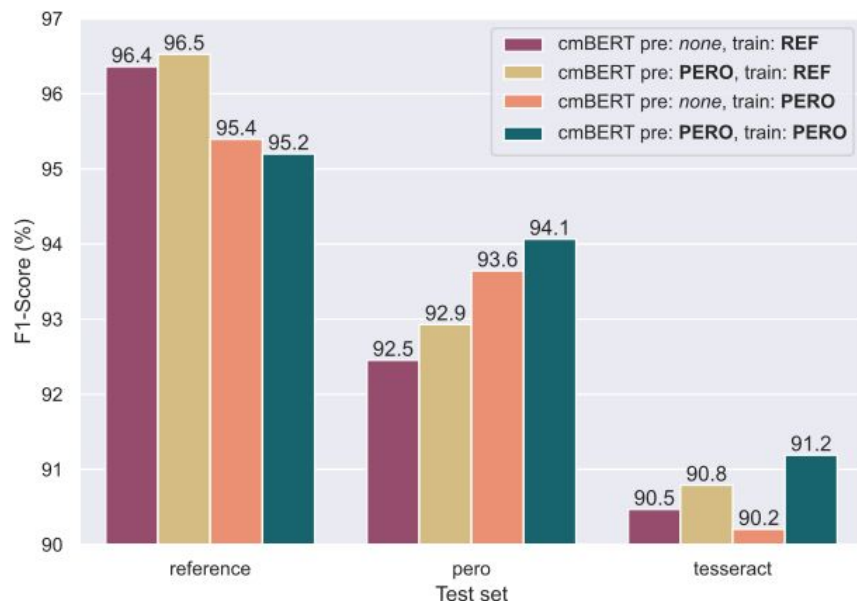
Modèles

- CamemBERT
- CamemBERT pré-entraîné avec des entrées d'annuaires OCrisées

8341 entrées annotées :

- Train : 6004 entrées
- Dev : 668 entrées
- Test : 1669 entrées

Reconnaissance des entités nommées (NER)



Abadie et al. (2022)

Mettereau, prop., quai d'Anjou, 7.\n
 Mettemberg, élig., méd., St-Thomas-d'Enf., 5.\n
 Metz (de), rentier, St-Guillaume, 30.\n
 Metzinger, avocat, Rameau, 6.\n
 Metzmacher, peint. sur émaux, St-Martin, 124.\n
 Meurgey, épicier-herboriste, Dragon, 33.\n
 Meurice, Chaussée-d'Antin, 3.\n
 Meurice (Eug.), tapissier, Vivienne, 12.\n
 Meurillon, marbrier-sculpteur, butte Mont-\n
 Parnasse, 15.\n



NER

Mettereau PER, prop. ACT, quai d'Anjou LOC, 7 CARDINAL.
 Mettemberg PER, élig. TITRE, méd. ACT, St-Thomas-d'Enf. LOC, 5 CARDINAL.
 Metz (de) PER, rentier ACT, St-Guillaume LOC, 30 CARDINAL.
 Metzinger PER, avocat ACT, Rameau LOC, 6 CARDINAL.
 Metzmacher PER, peint. sur émaux ACT, St-Martin LOC, 124 CARDINAL.
 Meurgey PER, épicier-herboriste ACT, Dragon LOC, 33 CARDINAL.
 Meurice PER, Chaussée-d'Antin LOC, 3 CARDINAL.
 Meurice (Eug.) PER, tapissier ACT, Vivienne LOC, 12 CARDINAL.
 Meurillon PER, marbrier-sculpteur ACT, butte Mont-Parnasse LOC, 15 CARDINAL.



Base de données contenant près de 10 millions d'entrées structurées

Géocodage

Sources géohistoriques des données d'adresses et de rues utilisées pour localiser les annuaires :



Rues de l'Atlas de Verniquet et adresses "Paris artistique, 1800-1820"



Rues et adresses de l'atlas Jacoubet, 1827-1836



Rues et adresses de l'atlas municipal, 1888

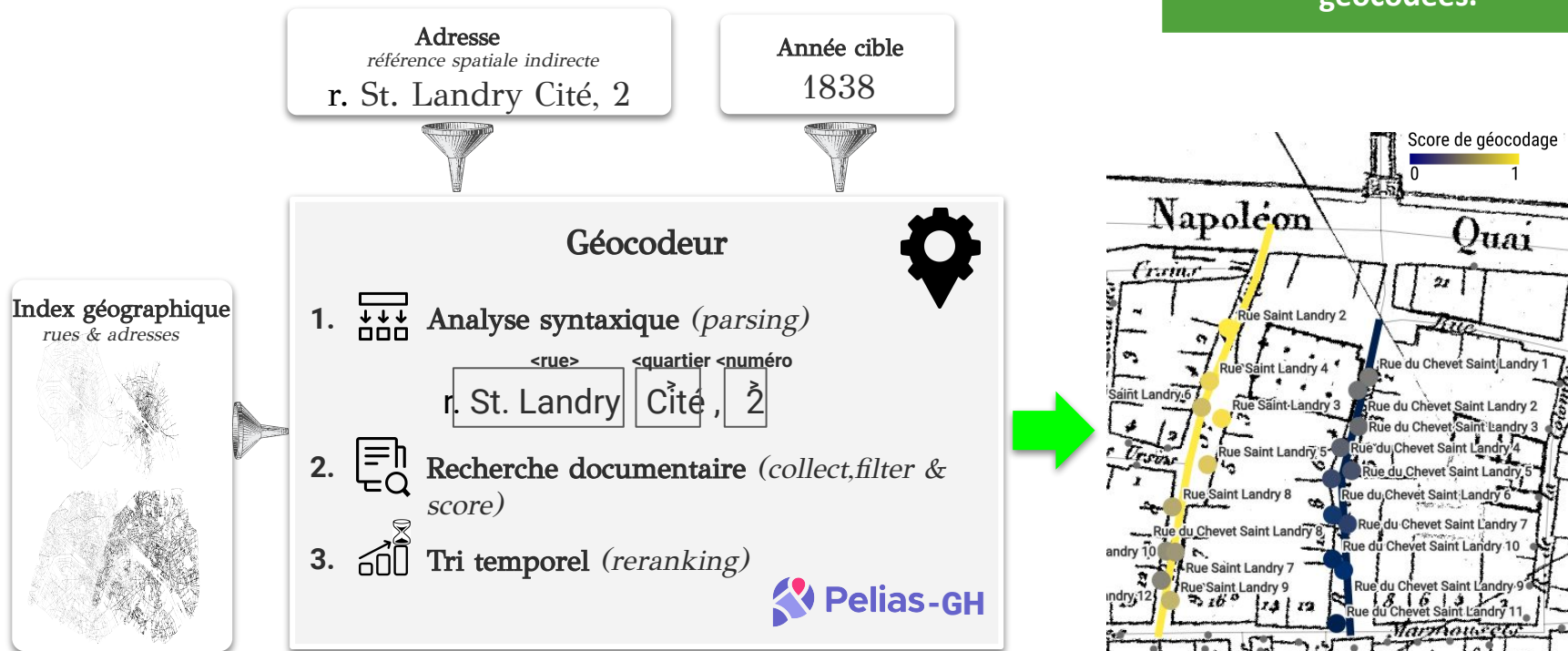


Rues et adresses OpenStreetMap, 2020

Duménieu (2023)

Géocodage

Géocodeur sensible à la dimension temporelle



Environ $\frac{2}{3}$ des entrées sont géocodées.

Outil open-source : https://github.com/GeoHistoricalData/historical_geocoding

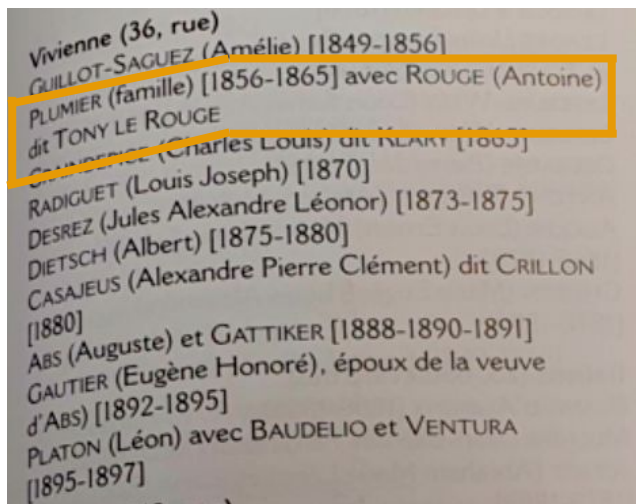
- ★ CQ1. Quelle est l'adresse du commerce X en 1861 ?
- ★ CQ2. Combien y a-t-il de commerces de ce type localisé *rue de Rivoli* en 1856 ?
- ★ CQ3. Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?
- ★ CQ4. Quels commerces ont déménagé au cours de leur existence ?
- ★ CQ5. Quels commerces ont été repris par un autre commerçant exerçant la même activité ?

Sélection des entrées pertinentes

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude



Recherche des photographes et ateliers (~230) listés par des historiens de l'art dans la base de donnée des extractions



```
SELECT *  
FROM directories.elements AS e  
WHERE e.persons ILIKE '%plumi%'  
ORDER BY e.published
```



Exemple de requête SQL
Toutes les entrées dont l'entité
"Nom de personne/commerce"
contient "plumi"

Plumier (Victor), **portraits** sur
plaques et sur pap., Vivienne, 36.

Didot_1856a

Exemple de résultat

Marc Durand (dir.). « De l'image fixe à l'image animée : 1820-1910.

Tome 2 : actes des notaires de Paris pour servir à l'histoire des photographes et de la photographie ». Archives nationales (2015).
Pierrefitte-sur-Seine.

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude

Réduction de la liste de mots-clés utilisés aux trois mots les plus couramment associés aux photographes listés dans la référence.

PHOTO	DAGUER	OPTI
Photographe	Daguerréotype	Optique
Photographie		Opticien



34 062 entrées

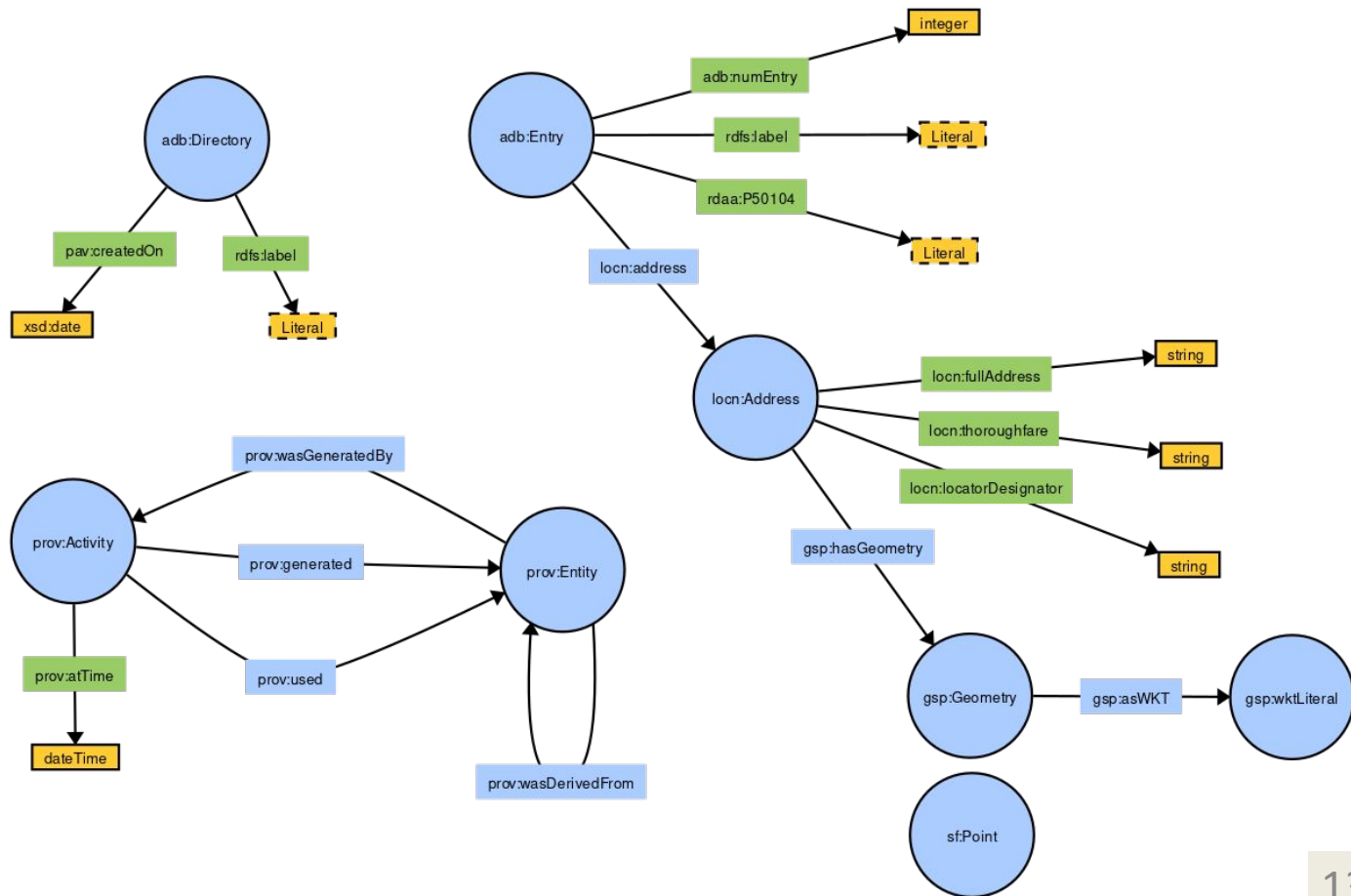


Création de ressources RDF à partir de la base de donnée relationnelle contenant ces entrées (avec *r2rml*)

• Ontologie

Vocabulaires utilisés :

- **RDFS** : label des entrées
- **LOCN** : description des adresses
- **Geosparql** : description des objets géographiques
- **RDA Registry** : description des activités
- **PROV** : description des données



- Liage des entrées relatives aux mêmes personnes ou commerces

Méthode fondée sur les connaissances

Appariement des entrées par **comparaison stricte** de clés



Clé = combinaison de propriétés, découvertes avec *Sakey*

- Numéro de l'entrée (si plusieurs activités ou adresses dans l'entrée)
- Nom et Activité
- Nom et Adresse
- Adresse et Activité



Création des liens par inférence

- Liage des entrées relatives aux mêmes personnes ou commerces

Méthode fondée sur les données

➔ Appariement par **comparaison numérique** des propriétés

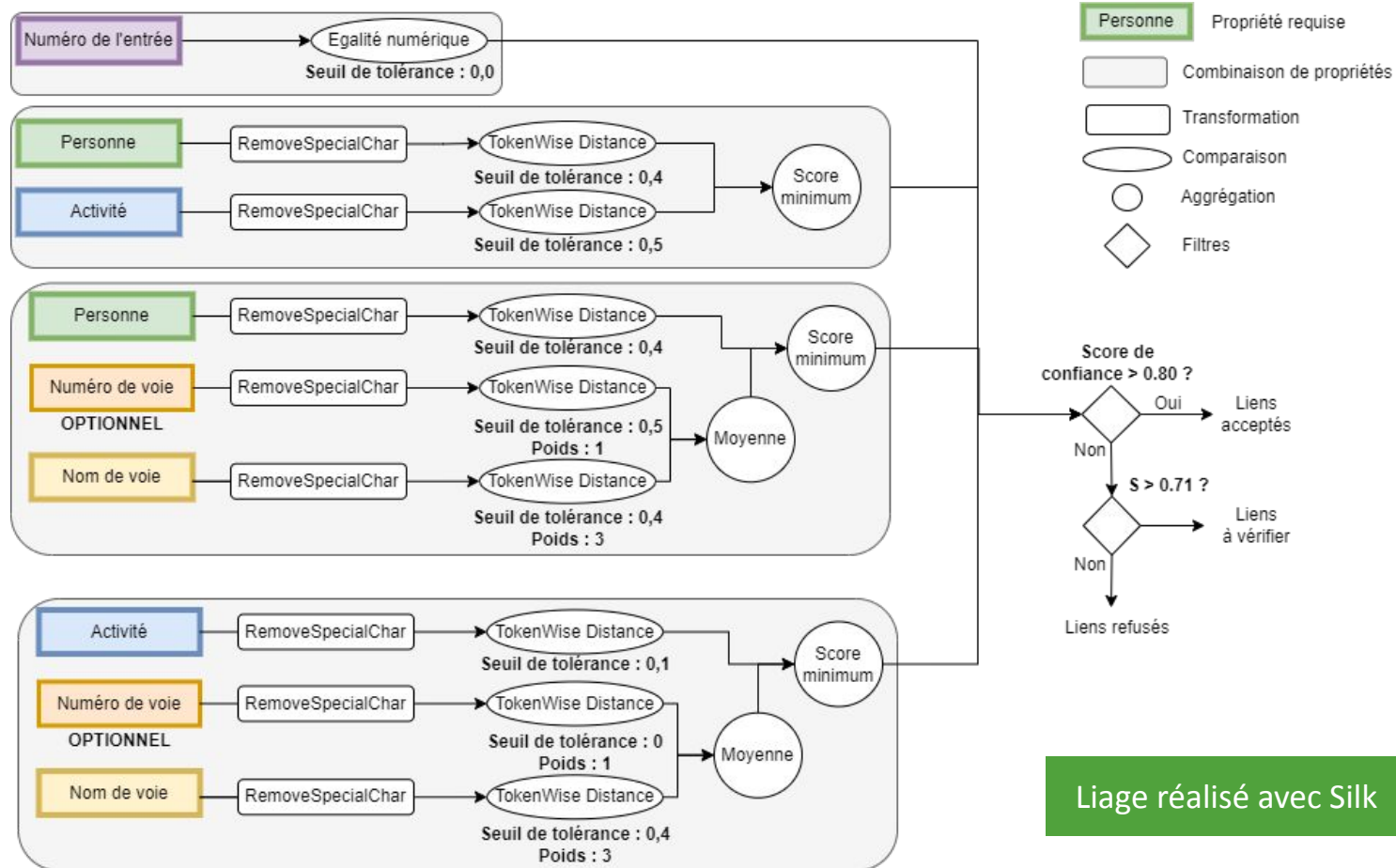
➔ Distance d'édition TokenWise

- Calcul de la distance de Levenshtein* entre tous les mots du texte
- Score final normalisé (entre 0 et 1)

* Distance de Levenshtein : compte le nombre d'insertions, de remplacements et de suppressions de caractères pour passer d'un mot à un autre

ex : **CHAT** <=> **PLAT** 2 remplacements

Création du graphe spatio-temporel



- Liage des entrées relatives aux mêmes personnes ou commerces

Méthode logique

250 622 liens

Paramétrage simple = clés

Comparaison très stricte

Méthode numérique

& Propagation des liens sameAs

357 130 liens

Paramétrage complexe :

identifier les seuils de tolérance pertinents

Comparaison plus adaptée aux chaînes
de caractères résultant de l'OCR

BILAN

401 852 liens d'équivalence distincts entre les ressources du graphe

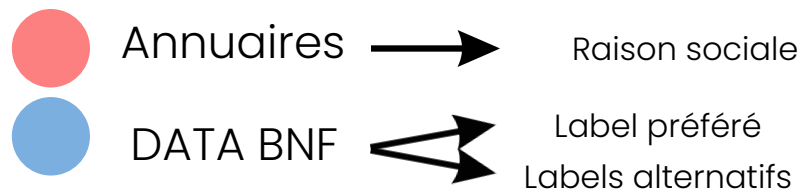


Graphes géohistorique

<https://dir.geohistoricaldata.org/sparql>

Liage avec d'autres ressources

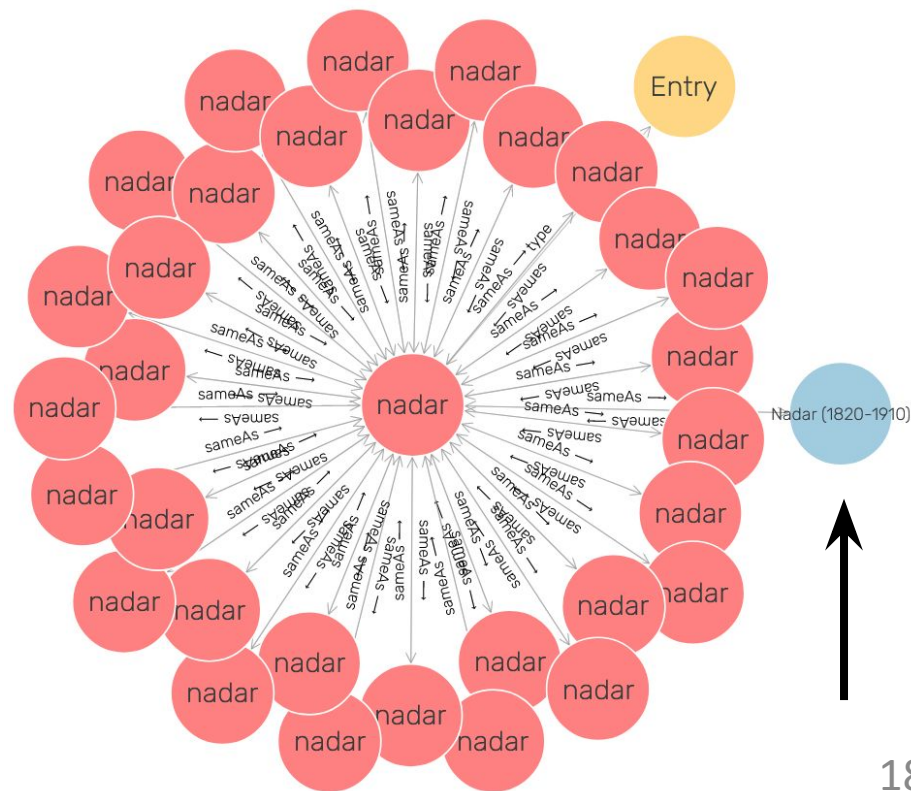
➔ Appariement par **comparaison numérique** des propriétés relatives au nom des photographes entre :



Nadar (1820-1910)



Pays :	France
Langue :	Français
Sexe :	Masculin
Naissance :	Paris, 06-04-1820
Mort :	Paris, 21-03-1910
Note :	Photographe, journaliste, caricaturiste. - Aéronaute. - Son frère, Adrien Tournachon, peintre de formation et collaborant à son atelier de photographie, se servait du même pseudonyme
Domaines :	Photographie
Autres formes du nom :	Gaspard-Félix Tournachon (1820-1910) Gustave-Félix Tournachon (1820-1910) Félix Tournachon (1820-1910) Félix Tournachon-Nadar (1820-1910) Félix Nadar (1820-1910) Nadar (1820-1910) Nadar père (1820-1910)
ISNI :	ISNI 0000 0001 2141 8535 (Informations sur l'ISNI)



- 76 ressources "Photographes" sur data.bnf.fr
- 161 liens validés manuellement
- ➔ 29 ressources data.bnf.fr liées aux ressources issues des Annuaire

Visualisation spatiale...



DEMO

bit.ly/soduco-photographes



Entrées extraites des annuaires du commerce de Paris (XIX^{ème} siècle) relatives aux photographes et aux professions associées

Filtres

Propriétés

Raison sociale

nadar

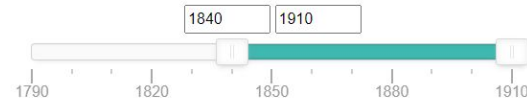
Description

Ex : photo

Adresse

Ex : rivoli

Période



Le filtre temporel permet de faire varier l'affichage des points préalablement chargés sur la carte sans lancer une nouvelle recherche. Données chargées pour la période 1840-1910.

Localisation

Dessinez l'emprise de votre zone de recherche avec l'outil de dessin disponible sur la carte.

Lancer la recherche

Plan de Paris contenant l'enceinte des fortifications - Andriveau-Goujon - 1849 © GeoHistoricalData

... et temporelle



DEMO

bit.ly/soduco-photographes



1 km
3000 ft

Leaflet | Plan de Paris contenant l'enceinte des fortifications - Andrieux-Goujon - 1849 © GeoHistoricalData

ANNÉE - 1861

NADAR

homme de lettres, artiste photographe
boul. des capucines 35
Source : DidotBottin 1861a - 1204351
Nombre d'entités liées : 6



ST-LAZARE

nadar

nadar

BOUL. DES CAPUCINES 35

nadar

nadar

nadar

nadar

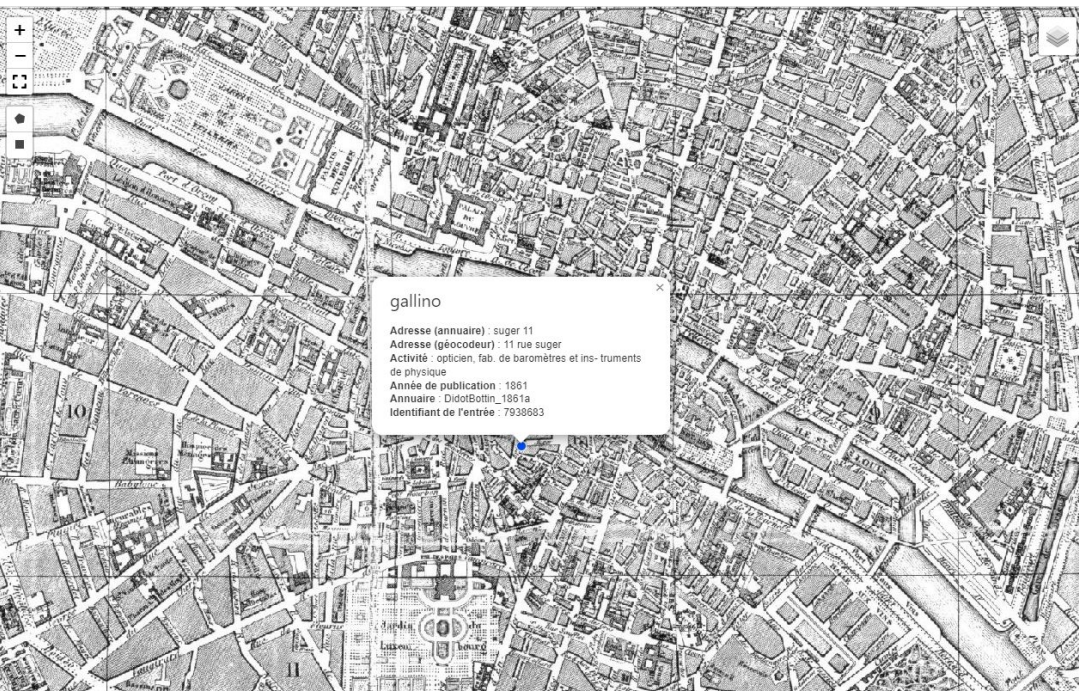
Carte

Questions de compétences

Quelle est l'adresse du commerce de Gallino en 1861 ?



11 rue suger



```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
select * where {
  ?e a ont:Entry.
  ?e rdfs:label ?label.
  ?e prov:wasDerivedFrom ?directory.
  ?directory pav:createdOn 1861.
  ?e locn:address ?add.
  ?add locn:fullAddress ?fullAdd.
  Filter regex(?label, "gallino").
}
```

Entrées
(XIXème
profession
})
Filtres

Propriétés

Raison sociale

gallino

Description

Ex : photo

Adresse

Ex : rivoli

Période



? Le filtre temporel permet de faire varier l'affichage des points préalablement chargés sur la carte sans lancer une nouvelle recherche.
Données chargées pour la période **1860-1862**.

Localisation

Dessinez l'emprise de votre zone de recherche avec l'outil de dessin disponible sur la carte.

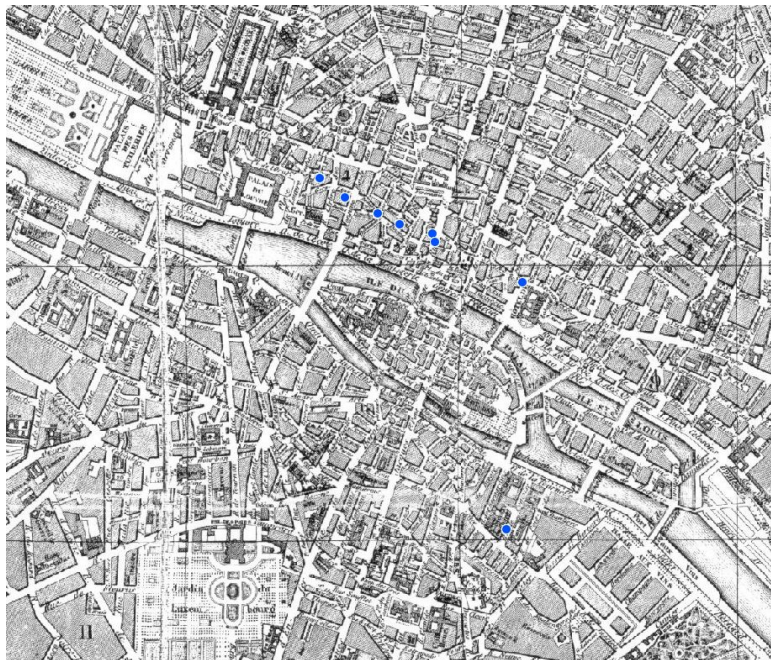
Lancer la recherche

Questions de compétences

Combien y a-t-il de commerces de ce type localisés rue de Rivoli en 1856 ?



9



```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
select (count(?e) as ?nombre) where {
  ?e a ont:Entry.
  ?e prov:wasDerivedFrom ?directory.
  ?directory pav:createdOn 1856.
  ?e locn:address ?add.
  ?add locn:thoroughfare ?voie.
  Filter regex(?voie, "rivoli").
}
```

Questions de compétences

Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX gsp: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
select ?label where {
  ?e a ont:Entry.
  ?e rdfs:label ?label.
  ?e prov:wasDerivedFrom ?directory.
  ?directory pav:createdOn 1875.
  ?e locn:address ?add.
  ?add gsp:hasGeometry ?geom.
  ?geom gsp:asWKT ?wkt.
  FILTER (geof:sfIntersects(?wkt, "<http://www.opengis.net/def/crs/OGC/1.3/CRS84> Polygon((2.3360 48.8625,2.3360 48.8663,2.3360 48.8663,2.3360 48.8625))")
}
```

Questions de compétences

Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?



Exemple de la zone située autour du Palais royal :

foucher

legros

crillon

reignier

vantier

gatel

bureau (s)

Questions de compétences

Quels commerces ont déménagé au cours de la période 1860-1870 ?

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select distinct ?label ?voie ?labelo ?voieo where {
  ?e a ont:Entry.
  ?e rdfs:label ?label.
  ?e prov:wasDerivedFrom ?directory.
  ?directory pav:createdOn ?date.
  ?e locn:address ?add.
  ?add locn:thoroughfare ?voie.
  ?e owl:sameAs ?o.
  ?o rdfs:label ?labelo.
  ?o locn:address ?addo.
  ?addo locn:thoroughfare ?voieo.
  Filter ((?date > 1860) && (?date < 1870)).
  Filter (!sameTerm(?voie, ?voieo))
}
```



Exemple : Déménagement de Nadar de la Rue St-Lazare au Boulevard des Capucines vers 1860-1861

Questions de compétences

Quels commerces ont déménagé
au cours de la période 1860-1870 ?

ANNÉE – 1861

NADAR

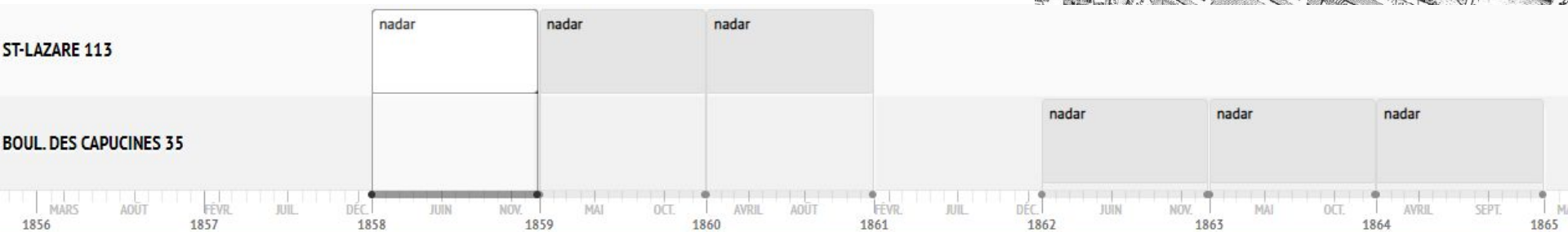
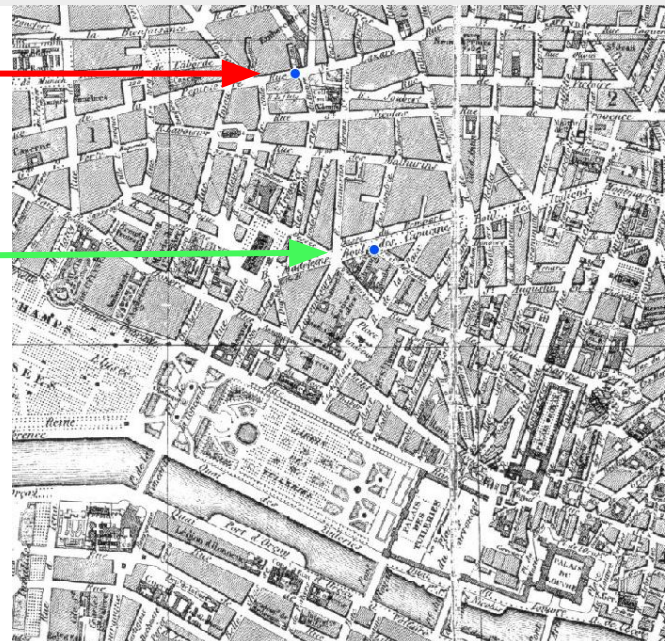
homme de lettres, artiste photographe
boul. des capucines 35

Source : DidotBottin_1861a - 1204351

Nombre d'entités liées : 6

St-Lazare


Boulevard des
Capucines



Questions de compétences

Quels commerces ont été repris par un autre commerçant exerçant la même activité ?

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select distinct ?label ?voie ?labelo ?voieo where {
  ?e a ont:Entry.
  ?e rdfs:label ?label.
  ?e locn:address ?add.
  ?add locn:thoroughfare ?voie.
  ?e owl:sameAs ?o.
  ?o rdfs:label ?labelo.
  ?o locn:address ?addo.
  ?addo locn:thoroughfare ?voieo.
  Filter (!sameTerm(?label, ?labelo) && sameTerm(?voie, ?voieo) )
}
```

 Exemple : Succession des photographes Bureau puis Mouret puis Baccard puis Courrier (A.) au 59 rue de Rivoli entre 1858 et 1904

- ➔ Démarche fonctionnelle d'exploitation des entrées issues des annuaires
 - extraites en masse,
 - structurées, souvent bruitées.

- ➔ Travaux à venir :
 - évaluation quantitative des appariements,
 - mise en oeuvre de l'approche pour d'autres types de commerces,
 - explicitation des relations spatio-temporelles entre les entrées (succession, déménagement).

- ➔ Méthode, code, application et données accessibles sur le web.

Merci pour votre attention

Ontologie, ressources, code et questions de compétences

https://github.com/soduco/ic_2023_photographes_parisiens

Application

bit.ly/soduco-photographes

Endpoint

<https://dir.geohistoricaldata.org/sparql>

Nathalie Abadie, Edwin Carlinet, Joseph Chazalon, et Bertrand Duméniéu. **A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories**. Number 13237 in Document Analysis Systems. DAS 2022., La Rochelle, France, May 2022. Springer, Cham.

Bertrand Duméniéu. **Localiser les adresses anciennes ? Plans historiques & annuaires parisiens au XIXe siècle**. Journée d'étude scientifique "Pour un traitement scientifique automatisé des cartes anciennes", Festival Printemps des Cartes 2023

Marc Durand (dir.). **De l'image fixe à l'image animée : 1820-1910. Tome 2 : actes des notaires de Paris pour servir à l'histoire des photographes et de la photographie**. Archives nationales (2015). Pierrefitte-sur-Seine.

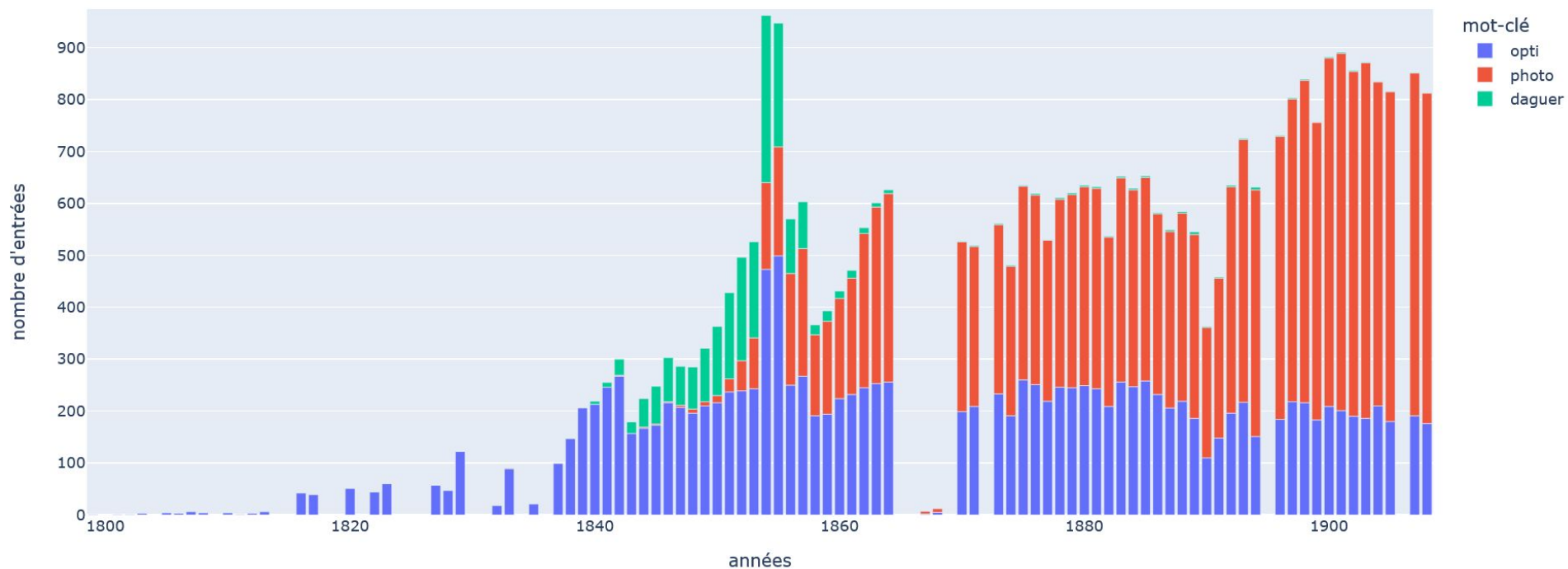
Danai Symeonidou, Vincent Armant, Nathalie Pernelle, and Fatiha Saïs. **SAKey : Scalable Almost Key Discovery in RDF Data**. In proceedings of the 13th International Semantic Web Conference, ISWC 2014, volume Lecture Notes in Computer Science of The Semantic Web – ISWC 2014, pages 33–49, Riva del Garda, Italy, October 2014. Springer.

Silk Linked Data Integration Framework : outil de liage de données, <https://github.com/silk-framework/silk>

Pero OCR, <https://pero-ocr.fit.vutbr.cz/>

Filtrage des entrées

Evolution du nombre d'entrées contenant les mots-clés 'photo', 'daguer' et 'opti' par année (table : par_activites)



- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude



Liste de 19 mots-clés => 227 323 entrées.

Mots et expressions visés	Mot-clé retenu	Mots et expressions visés	Mot-clé retenu
Photographe, photographie, ...	photo	Instrument/outil/appareil de mathématiques, mathématicien	math
Daguéréotype	daguer	Physique	physique
Opticien, optique	opti	Produits chimiques	chimi
Lentille	lentil	Ingénieur	ingenieur
Lunettes, lunettier	lunet	Image	imag
Graveur	grav	Portrait	portr
Lithographe	litho	Prisme	prisme
Artiste	artiste	Appareil/instrument de mesure	mesure
Chambre noire	chamb AND noir	Camera obscura	camera
Cinéma, cinématographe	cinem		

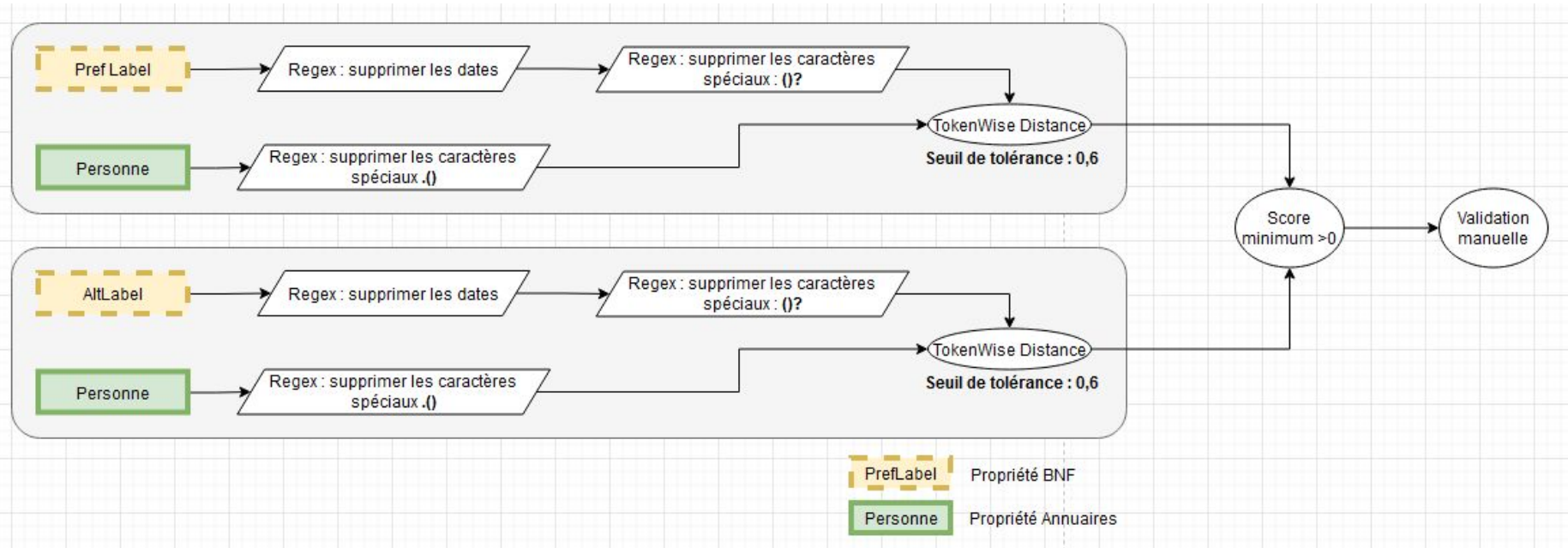

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX bnf-onto: <http://data.bnf.fr/ontology/bnf-onto/>
CONSTRUCT {
  ?s a foaf:Person;
  bnf-onto:firstYear ?fy ;
  bnf-onto:lastYear ?ly ;
  skos:prefLabel ?pf;
  skos:altLabel ?al.
}
WHERE {
  ?c a skos:Concept; skos:prefLabel ?pf; skos:altLabel ?al; foaf:focus ?s.
  ?s a foaf:Person ;
  bnf-onto:firstYear ?fy ;
  bnf-onto:lastYear ?ly ;
  rdagroup2elements:countryAssociatedWithThePerson <http://id.loc.gov/vocabulary/countries/fr> ;
  rdagroup2elements:fieldOfActivityOfThePerson <http://dewey.info/class/770/>, "Photographie" .
  Filter (?fy > 1760 && ?fy < 1885)
}
```



<https://data.bnf.fr/sparql>

76 fiches

Annexe : Appariement Annuaire – Data BNF



161 liens validés manuellement vers 29 fiches