

Des annuaires du commerce parisien du XIX^e siècle à des graphes de connaissances géohistoriques professionnels.

N. Abadie, S. Baciocchi, E. Carlinet, J. Chazalon, P. Cristofoli, B.
Duménieu, J. Gravier, J. Perret, **S. Tual**
Lastig (IGN), CRH (EHESS), LRE (EPITA)

Journée “Reconnaissance automatique de caractères et extraction
d’information spatiale à partir de sources anciennes numérisées”

10 décembre 2025

CC BY 4.0

Social dynamics in urban context (SoDUCo - 2019-2023)

Etude de l'évolution de la ville de Paris au 19ème siècle à travers deux corpus de sources historiques:

- Les plans et atlas à grande échelle;
- Les annuaires du commerce qui recensent les commerçants, les notables, les organisations et institutions.

Objectifs méthodologiques:

- Proposer des approches et des modèles pour étudier la co-évolution des structures urbaines de la ville avec les pratiques économiques et sociales des habitants.
- Produire et publier des ressources ouvertes (codes, modèles, données) pour la communauté scientifique.

<https://soduco.geohistoricaldata.org/>

<https://anr.fr/Projet-ANR-18-CE38-0013>



Social Dynamics in Urban Context

Des sources géographiques anciennes aux graphes de connaissances géohistoriques

Reconnaissance

Image → Géométries



Scan → Texte

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

Classification

Géométries →
Annotations



Texte →
Entités nommées spatiales

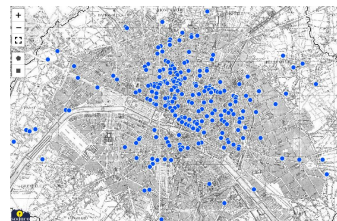
Ravrio et comp. fabr. de bronzes et curiosités
r. Richelieu 93 CARDINAL la fabrique
rue Montmartre 161 CARDINAL

Géoréférencement

Données géométriques →
Données géographiques



Entités nommées spatiales →
Données géographiques

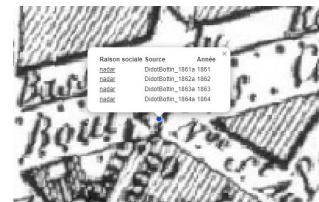


Structuration

Données géographiques →
Données géohistoriques



Données géographiques →
Données géohistoriques



Les annuaires du commerce

Publiés de la fin du 18ème siècle au milieu du 20ème siècle: 144 annuaires du commerce, publiés par différents éditeurs, de 1797 à 1914, couvrant 88 années.

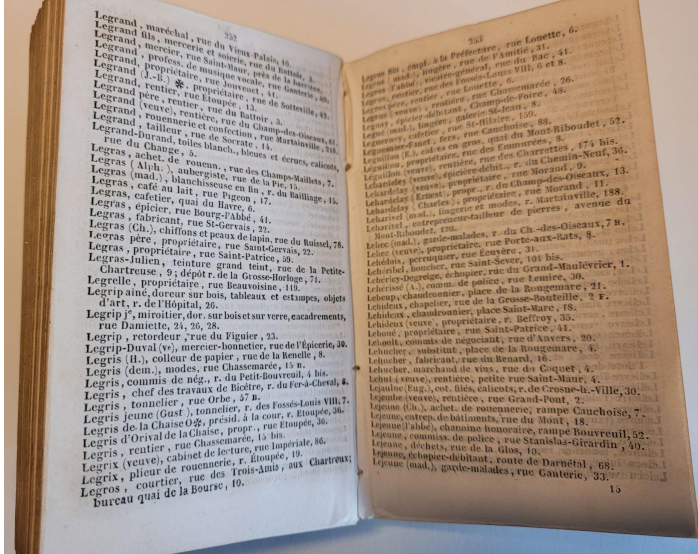
Organisés selon différents index: noms, rues et activités.

Chaque entrée contient:

- (Raison sociale) Nom
- Activité ou statut social
- Adresse
- Titres et décorations militaires ou professionnels

Différences entre annuaires:

- Longueur, richesse et organisation des informations
- Mise en page, police
- Qualité : papier, impressions, résolution du scan, taches sur le document...



Non-Commerçans. (Paris).		269
Chardin, R. Pavée, 16. — R. C.	Cheviron, R. Chapon, 13.	
Chardin, R. Michel Lepelletier, 21.	Chimay, (Mme.) R. de Varennes, 31.	
Chardon, (Ve.) R. S. Marc, 15.	Choart-Duplessis, R. de Turenne, 31.	

AMADOU ET ALLUMETTES. — Pour les ALLUMETTES OXYGÉNÉES. Voyez BRIQUETS PHYSIQUES.

DARRAS (Thomas), r. de la Vieille-Monnaie, 10.
Briquets et veilles, mèches à quinquets, à quinquet, veilles mèches, souffrées; mèches souffrées, pierres, agaric de chène, pierres, agaric, bouchons, liège.
LEROY, r. Aubry-le-Boucher, 43.

BAUDoyer (place).	26* Lougré aîné, bijoutier en et argent.	Bourguille, fabr. de pressés.	École communale de jeunes filles.	et tapissières.
IX Arr. Hôtel-de-Ville. — Rue Tixeranderie, pourtour St-Gervais, Saint-Antoine et Renaud-Lefèvre.	Saint-Omer, émailleur. Cellier (A.), graveur-ciseleur.*	Vaudain, passementier. Fininojnc, bronze doré. Rabé aîné, fabr. de bal-lons.*	Berthelot, vins.	10 Lainé jeune, vins. Jumelles omnibus et entrepise générale des Omnibus.
1 Lissoty (Vve), vins. 2* Privé, distillateur. Lemoine-Cluzel et Leroy, nouveautés. Chantrier, court.-gourm.	Rousseau (J.), bijoutier en or.* Benoit, orfèvre-fabr. Lérôys, doreur.	49* Cendrier aîné, prop. Desmarests, fab. bottes d'emballage. Ferrand, lapidaire.	6 Verstaen, serrurier-mécanicien.	11 Melouzy, vins en gros, et à Bercy, ferrurier.
	30 Bouton, fab. de cuir ver-nis.*	31 Pardon, vins.	8 Michel, brossier.	12 Combaud, colporteur. Monmain (P.), vins en gros.
			13 Lejard, clouteries et cré-pins.	13 Dufailly, sculpt. fabr. de carton-pierre.
			14 Baduel (Vve), fab. de	

Des instantanés des activités économiques à une date donnée

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 Doire, 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

Bibliothèque Ste-Geneviève, rue des Sept-Voies
 et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
 Bibolet, relieur, passage Sainte-Marie-Saint-
 Germain, 10.
 Bibonne, architecte, Magasins, 12.
 Bibron, aide-naturaliste au Jardin-des-Plan-
 tes, Cuvier, 29.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 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 1844a - pages 125-126

Des instantanés des activités économiques à une date donnée

Avec un haut niveau de redondance d'informations au sein d'un annuaire et annuaires.

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.
Richard (Mme), Nve-de-Luxembourg, 17.
Richard, 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.
Bicel, épiciier, marché d'Aguesseau, 15.
Richard (Mme), Nve-de-Luxembourg, 17.
Richard, 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 Doire, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épiciier, Marché-d'Aguesseau, 15.
Richard, tabac et eau-de-vie, Faub.-St-Mar-
tin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, rue des Sept-Voies
et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33
(provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-
Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plan-
tes, Carier, 23.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épiciier, Marché-d'Aguesseau, 15.
Richard, tabac et eau-de-vie, Faub.-St-Mar-
tin, 45.

Didot 1844a - pages 125-126

Des instantanés des activités économiques à une date donnée

Certaines entrées disparaissent au cours du temps.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
(provisoirement).
Bibl'que protestante (Société), Moulins, 16.
Bibron, aide-naturaliste, au Muséum d'Hist. nat.
Bibus, tailleur, Roule, 21.
Bibus, tailleur, Richelieu, 31.
Bical et Dorre, fab. de socques, Vertbois, 14.
Bicant (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-naturaliste, 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.
Bicel, épiciers, 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 Doire, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épiciers, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Mar-
tin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, rue des Sept-Voies
et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33
(provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-
Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plan-
tes, Cuvier, 29.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épiciers, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Mar-
tin, 45.

Didot 1844a - pages 125-126

Des instantanés des activités économiques à une date donnée

D'autres entrées apparaissent.

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-Martin, 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.
Bicel, épicier, marché d'Aguesseau, 15.
Bichard (Mme), Nve-de-Luxembourg, 17.
Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 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 Dorre, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épicier, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, rue des Sept-Voies et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plantes, Cuvier, 29.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épicier, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1844a - pages 125-126

Des instantanés des activités économiques à une date donnée

Et enfin d'autres entrées évoluent.

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-Martin, 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.
Bicel, épicier, marché d'Aguesseau, 15.
Bichard (Mme), Nve-de-Luxembourg, 17.
Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 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 Doire, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épicier, Marche-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, rue des Sept-Voies
et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33
(provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plantes, Cuvier, 29.
Bibus, tailleur, Richelieu, 31.
Bicel, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place
de la Corderie-du-Temple, 26.
Bicel, épicier, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1844a - pages 125-126

Chaîne de traitement: des annuaires à un graphe de connaissances géohistorique

Reconnaissance

Classification

Géoréférencement

Structuration

Scan → Texte

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

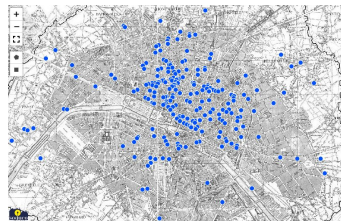
Segmentation des
pages/ Reconnaissance
de la mise en page +
Reconnaissance Optique
de Caractères (OCR)

Texte →
Entités nommées spatiales

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

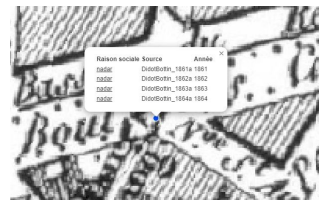
Reconnaissance
d'Entités Nommées
(NER)

Entités nommées spatiales →
Données géographiques



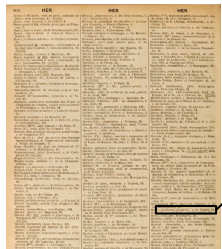
Géocodage

Données géographiques →
Données géohistoriques



Filtrage et Liage

Reconnaissance des entrées, du texte et des entités nommées



Lemonnyer, *plomberie*, r. de Bondy, 86, et
r. Bouchardon, 1.



Lemonnyer, plomberie, r. de Bondy, 86, et r. Bouchardon, 1.



PERSONNE	ACTIVITÉ	RUE	NUMÉRO		RUE	NUMÉRO
Lemonnyer,	plomberie,	r. de Bondy,	86,	et	r. Bouchardon,	1.

Chaîne de traitement proposée

1

Pré-traitement des images
(correction géométrique,
débruitage, ...)

2

Segmentation et
classification des éléments
de mise en page

3

Transcription

4

Annotation
sémantique
(NER)

Lemonnyer, plomberie, r. de Bondy, 86, et
r. Bouchardon, 1.

Lemonnyer, plomberie, r. de Bondy,
86, et r. Bouchardon, 1.

PERSONNE ACTIVITÉ RUE
Lemonnyer, plomberie, r. de Bondy,
NUMÉRO RUE NUMÉRO
86, et r. Bouchardon, 1.

Pré-traitement des images

□ Améliorer la qualité des images pour faciliter les traitements à venir

- ✓ Amélioration du contraste/ retrait du fond
- ✓ Rotation / Cisaillement dû à la pliure

Publication sur la détection de lignes (rapide et fiable):

P. Bernet, J. Chazalon, E. Carlinet, A. Bourquelot et E. Puybareau, "Linear Object Detection in Document Images using Multiple Object Tracking", in Proc. ICDAR 2023

<https://github.com/EPITAResearchLab/bernet.23.icdar>



Source image



Detected lines



Corrected image

2 Segmentation et classification de

Approche ad hoc qui tire profit de la relative régularité des mises en pages des annuaires:

1. Séparation des blocs (XY cuts) et classification des régions (en-tête, titre, etc.)
2. Reconnaissance des lignes (watershed)
3. Groupement des lignes en entrées (HMM)

Advantages:

- Très rapide (fraction de seconde par image)
- Mêmes performance que les approches SOTA comme LayoutParser **sur ce corpus**

Inconvénients:

- L'information textuelle n'est pas prise en compte
- Les entrées à cheval sur deux pages ne sont pas détectées
- Sensible au bruit: pré-traitement indispensable

[illegible]

Transcription du texte (OCR)

Benchmark des systèmes OCR sur étagère sur notre corpus
“A Dataset of French Trade Directories from the 19th Century (FTD)”

DOI [10.5281/zenodo.6394464](https://doi.org/10.5281/zenodo.6394464)

Les systèmes évalués:

- [Tesseract v4](#) (v5 non évaluée)
- [PERO OCR](#) (code Github + modèle fourni par les auteurs en 2020)
- [Kraken OCR](#) (modèle générique pour l'anglais imprimé)

Aucun entraînement spécifique n'a été fait sur notre corpus.

Bottin 1820

**Dufort, bottier, Palais-R., gal. vitrée, 215.
295**

Bottin 1827

Baleste, chef aux domaines, S.-Georges, 17.

Bottin 1837

Cattois, pharmac., Bretagne, 46.

Bottin 1854

Fontaine, draperies, Neuve-des-Petits-Champs, 2.

Cambon Almgene 1841

**Aron Javal (L.) art. de Paris, r. des Bour-
donnais, 17.**

Deflandre 1828

**DEVILLERS, r. Croix-des-Pet.-Champs, 25.
Cordonn.**

Deflandre 1829

Huguenin, épici., r. de Valois, 8, Pal.-Royal.

Didot 1851

**Viéville, fab. de boutons, Aumaire, 48, et place
St-Nicolas-des-Champs, 2.**

Transcription du texte (OCR)

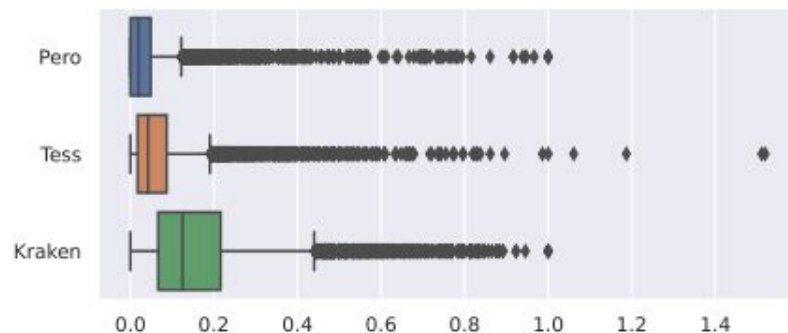
Résultats:

- PERO OCR fournit les meilleures performances
- Kraken manque de modèles ouverts (peut-être résolu par le projet [HTR-United](#)?)

Quelques tests supplémentaires ont été faits depuis:

- [Microsoft UniLM/TrOCR](#)
- [Mindee docTR](#)

	PERO OCR	Tesseract	Kraken
CER	3.78%	6.56%	15.72%



Annotation sémantique (*Reconnaissance d'Entités Nommées*)

Reconnaître et classer des tokens selon des classes prédéfinies : personne, lieu, organisation, date, ...

Les entrées d'annuaires contiennent des entités nommées particulières.

Exemple:

PERSON
Ravrio et comp., fabr. de bronzes et curiosités,
r. Richelieu, 93; la fabrique rue Montmartre, 161.
LOCATION CARDINAL FEATURE TYPE LOCATION CARDINAL

Difficultés : le texte OCRisé est sujet à erreurs

Gavarret ✱, prof. de physique à la faculté de
médecine, Grenelle-St-Germain, 49.

ravarret #, prof. de physique à la faculté de
médecine, Grenelle-St-Germain, 49.

Sortie de Tesseract v4

Duffaut, chaudronnier, r. de la Sourdière,
314

Daflant, c'audronnier, v, de la RARES
Ge OO a x

Sortie de Tesseract v4



= erreurs

Besoin d'une approche de NER robuste au bruit OCR, ou
alors de procéder à une post-correction du texte à traiter.

Difficultés : l'organisation des entrées peut varier

Nom	Activité	Adresse
-----	----------	---------

Durand jeune; pour bas, Charenton, 12 ancien. 18. *

Numéro de rue

Prévost-Guillaume, f. ta., r. N.-St.-Mart., 28.

Lefranc Méquignon et co., satins turcs, pruneaux, satins et draps de soie, gros de Naples, toiles, coutils, galons, rubans, coulisses et lacets pour chaussures de dames, sommières, flanelles et molletons de soie pour fourrures, r. des Prouvaires, 32.

Planche , R. de Poitou, 9.-H. Armé.

Baronnat frères, soies teintes et écries, fil-Denis, 257, passage du Renard ; maison à Lyon, r. Cen-

Jamain, orangiste, @S. H.1831, Fossés-St-Marcet, 12.

Appert fils, verres et cristaux, 21-23 Jour.

Mabire, Lourcine, 124.

Besoin d'une approche de NER robuste aux variations d'organisation, de syntaxe et au bruit, utilisant de l'apprentissage par l'exemple.

Une approche de NER fondée sur de l'apprentissage profond

Un modèle de langage de type *Transformer* : **CamemBERT**, conçu pour le français et déjà entraîné pour une tâche de NER générique.

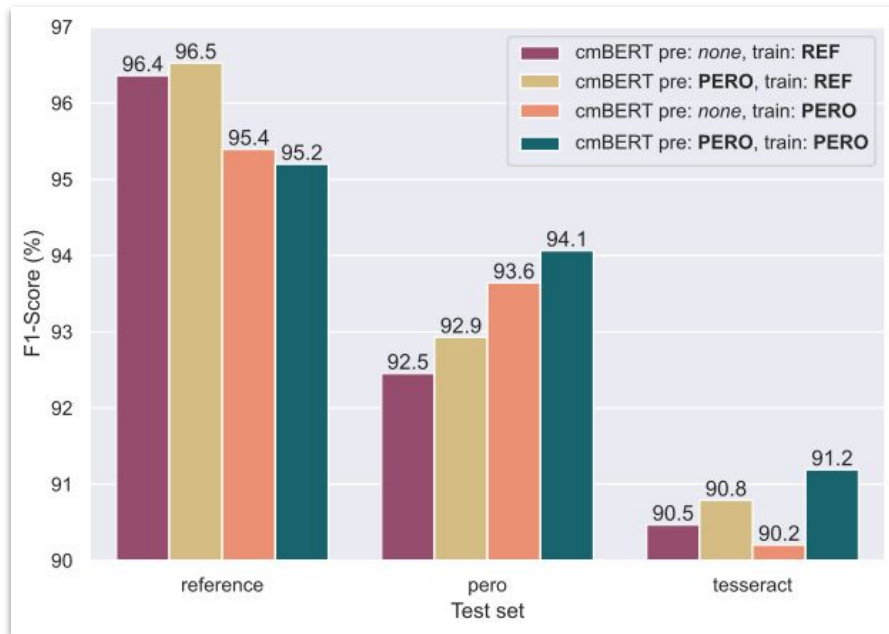
Adaptation pour répondre à nos besoins:

- Pré-entraînement non supervisé sur du texte OCRisé masqué (avec du bruit)
- Fine-tuning supervisé sur du texte OCRisé, annoté manuellement pour reconnaître les différents éléments constitutifs des entrées d'annuaires.

⇒ Nette amélioration des performances F-score > 94% sur le jeu de test

Plus de détail dans :

N. Abadie, E. Carlinet, J. Chazalon et B. Duménieu, "A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories", in *Proc. DAS 2022*, <https://github.com/soduco/paper-ner-bench-das22>



Un développement itératif...

Extraction “v2”, ~110 annuaires, livrée en janvier 2023

- 9 825 018 d’entrées extraites
- ~ 74% des entrées extraites sont “bien formées” :

PERSON [ACT] LOC [CARDINAL]

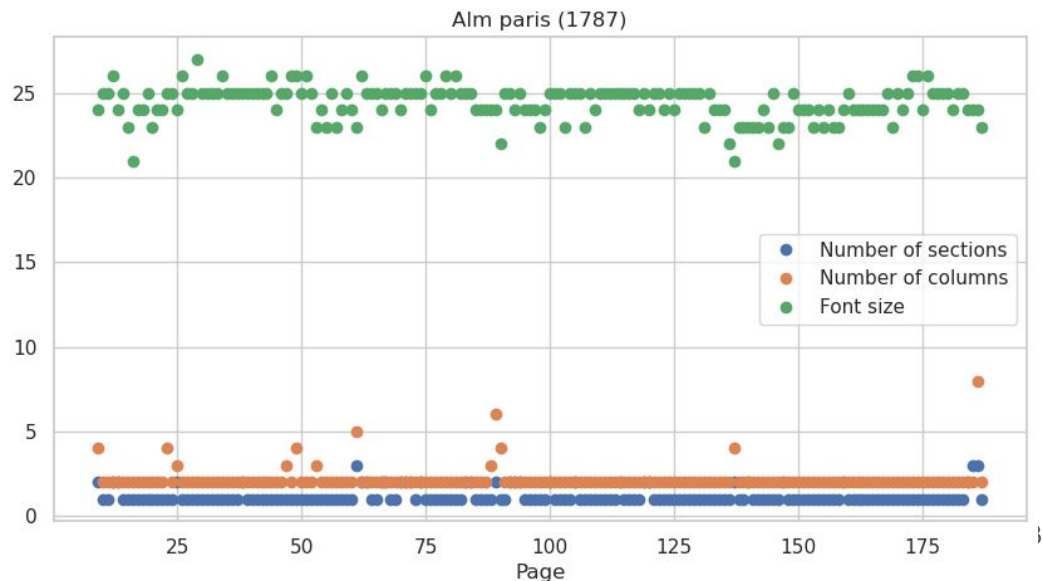
Jeu de données publié en Open Data:

<https://nakala.fr/collection/10.34847/nkl.abe0gxah>

Amélioration 1: Validation de la cohérence interne des pages

2 étapes :

1. Choisir et calculer des indicateurs clés pour chaque page d'annuaire : police, taille, nombre de colonnes, nombre d'entrées, etc.
2. Identifier les pages incohérentes pour adapter leur prétraitement ; relancer la chaîne de traitement pour les pages concernées.



Amélioration 2 : Traitement du texte en flux - et plus par page

Répond aux défis suivants:

1. Comment mettre à profit à la fois l'information visuelle et textuelle pour reconnaître les entrées des annuaires ?
2. Comment éviter de perdre des entrées à cause des sauts de pages et de colonnes ?
3. Est-il possible de segmenter les entrées et de reconnaître les entités nommées en même temps?

Approche proposée : un modèle de langage fondé sur le texte et la typographie

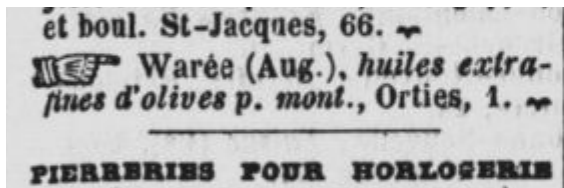


Image source

```
<page_break><space> et boul. St-Jacques, 66. <space>  
<line_break> <space>☞ Warée (Aug.), huiles extra-<space>  
<line_break> <space> fines d'olives p. mont., Orties, 1. <space>  
<line_break> <space> — <space>  
<line_break> <space> PIERRERIES POUR HORLOGERIE <space>
```

On entraîne un modèle de NER sur un flux de texte enrichi avec les espaces et les sauts pour qu'il apprenne à reconnaître à la fois les entrées et les entités.

Au bout du bout, qu'est-ce qui a été fait ?

Extraction “v4”, 144 annuaires soit 192 fichiers PDF:

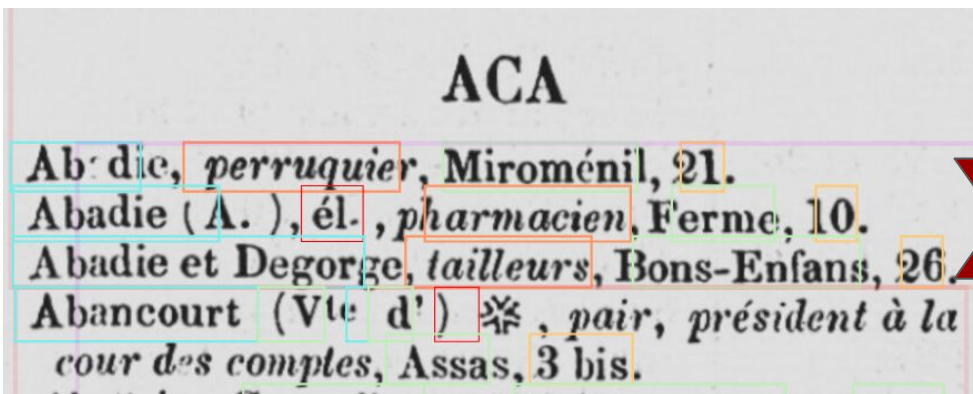
- 22 743 928 entrées
- 23 728 378 entités PER
- 2 106 808 entités TITRE
- 14 663 426 entités ACT
- 27 618 637 adresses (CARDINAL + LOC ou seulement LOC)
96.3% sont géocodées → **27 millions de points à placer sur une carte !**

Le jeu de données est publié en Open Data:

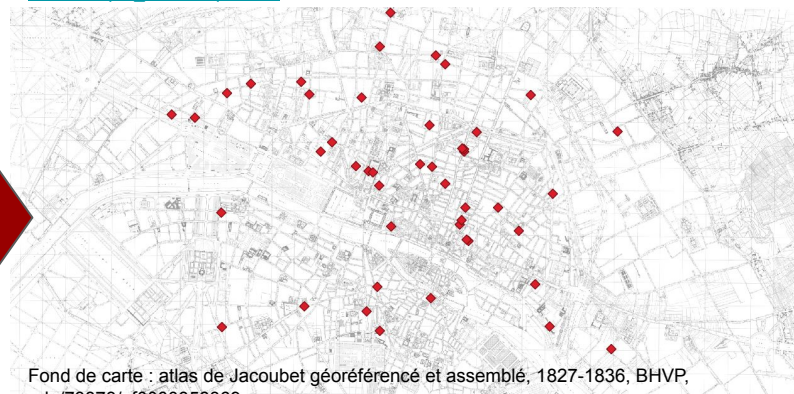
<https://nakala.fr/collection/10.34847/nkl.abe0gxah>

Géocodage historique

(Lamy_1840) Annuaire général du commerce, judiciaire et administratif de France et des principales villes du monde, v. 128.



https://api.geohistoricaldata.org/directories/export_geojson?source_pdf_id=eq.Lamy_1840&source_pdf_view=eq.0128&



Post-traitement des adresses

#1 OCR

> Aaron, bronzes et pendules, passage Choiseul, 72 et 74.

#2 NER

> Aaron, bronzes et pendules, passage Choiseul, 72 et 74.

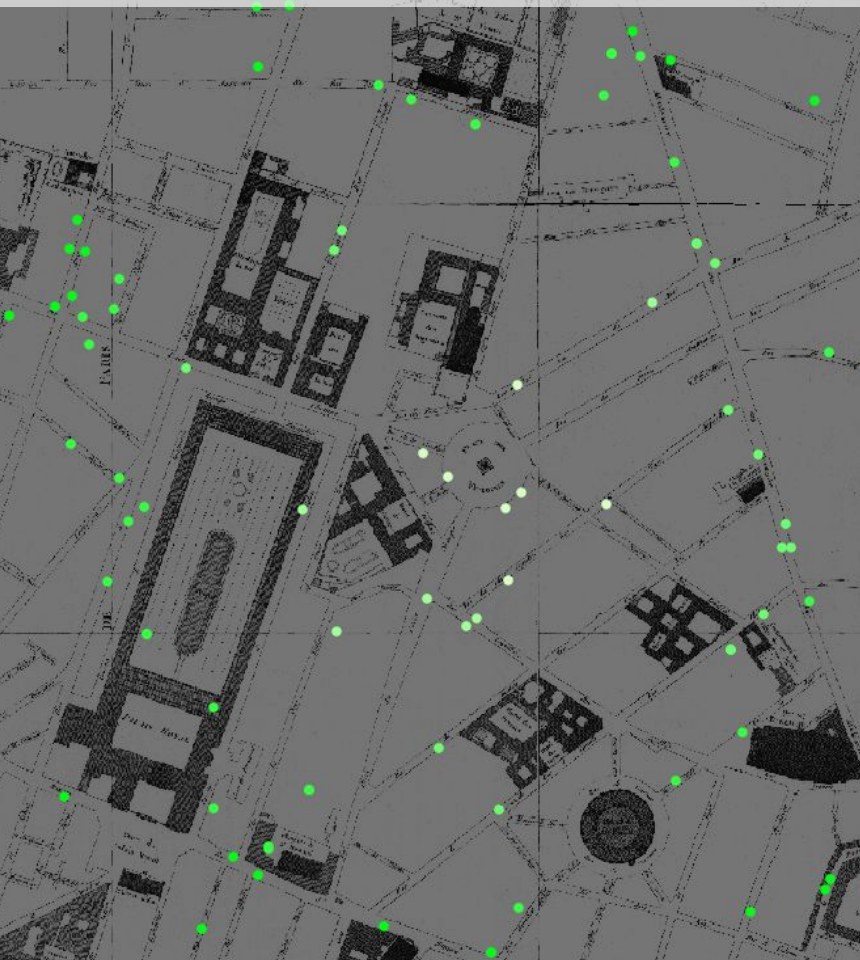
#3 Structuration

Address #1 : passage Choiseul 72

Address #2 : passage Choiseul 74



Vérité (101 points)



Géocodeur moderne (67 points)

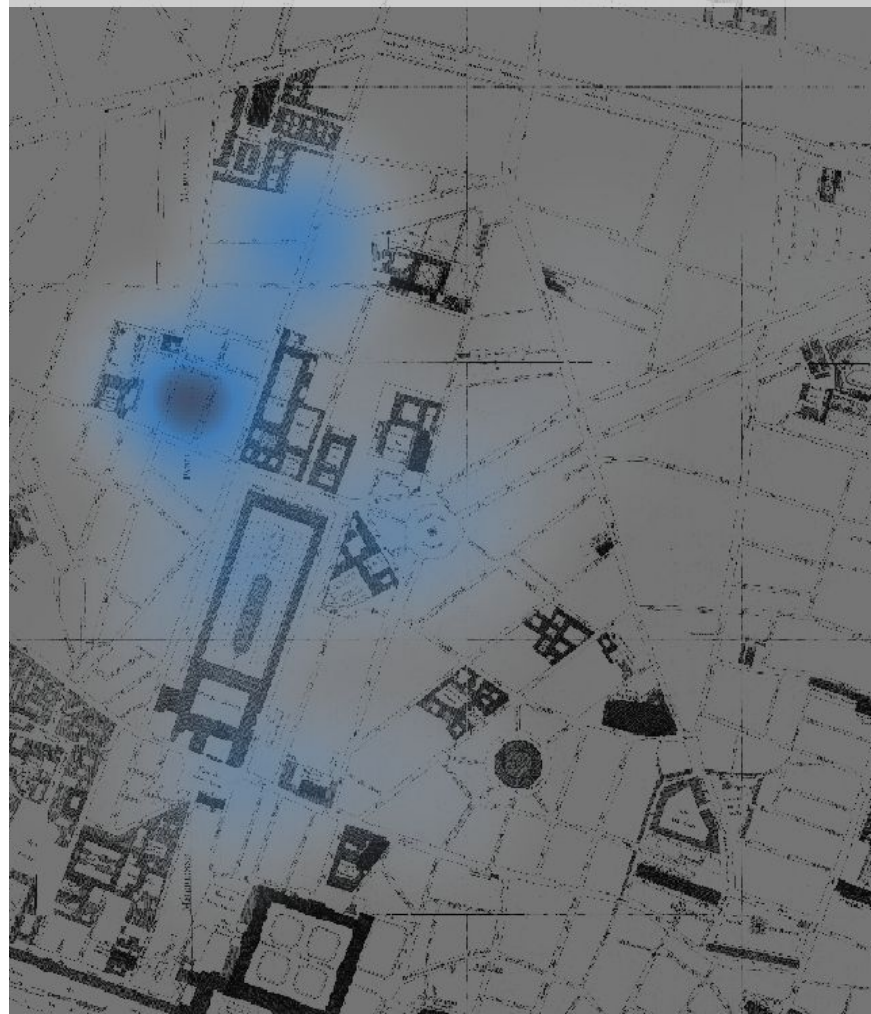


Données : tables de succession 1812-1827

Vérité (101 points)



Géocodeur moderne (67 points)



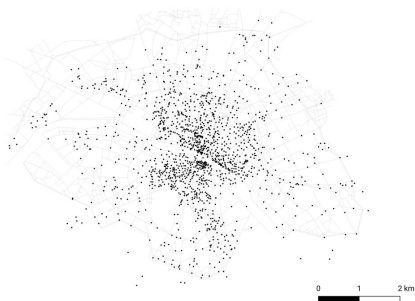
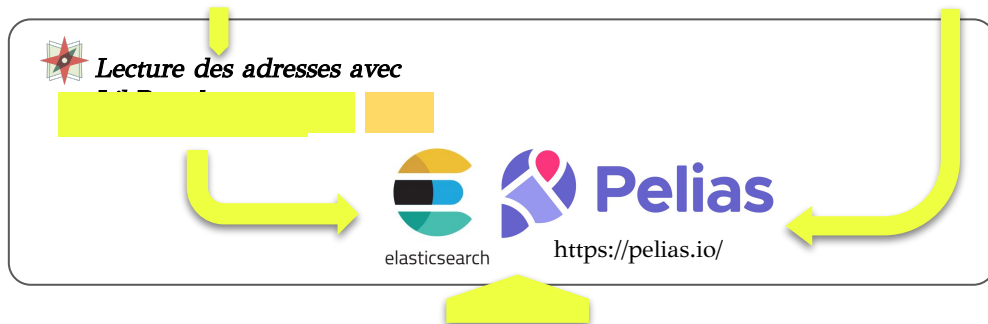
Le géocodeur historique : un moteur de recherche utilisant des données d'adresses issues d'atlas parisiens anciens

Recherche multicritère dans les données d'adresses saisies à la main sur des atlas parisiens anciens :

- Appariement flou des adresses et des numéros
- Distance temporelle

adresse
"passage Choiseul, 72"

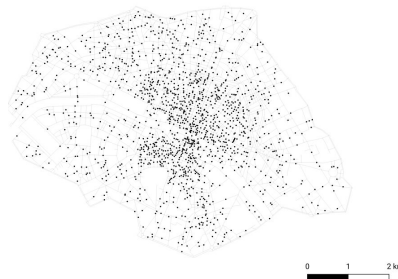
Année cible (date ou intervalle)
[1850,1852]



Atlas de Verniquet, 1784-1795 ⌚
Filaire des rues



Atlas de Jacoubet, 1827-1836 ⌚
Filaire des rues & points adresse



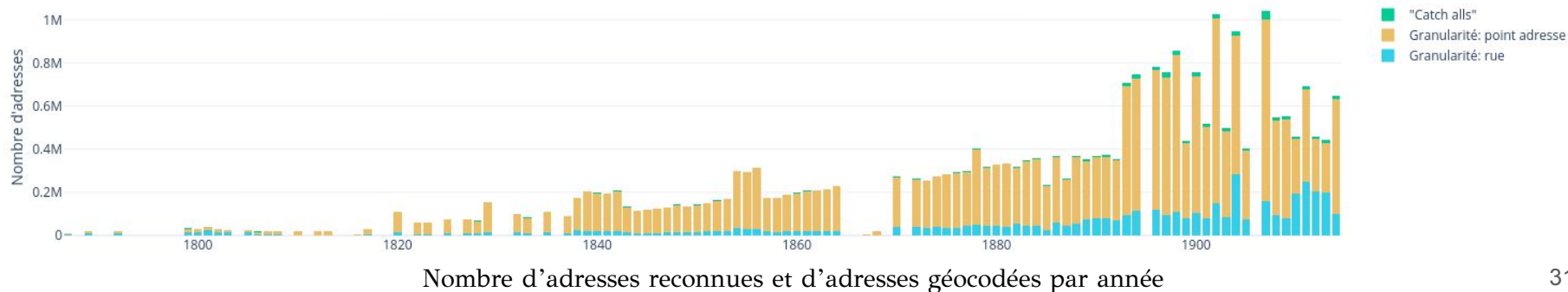
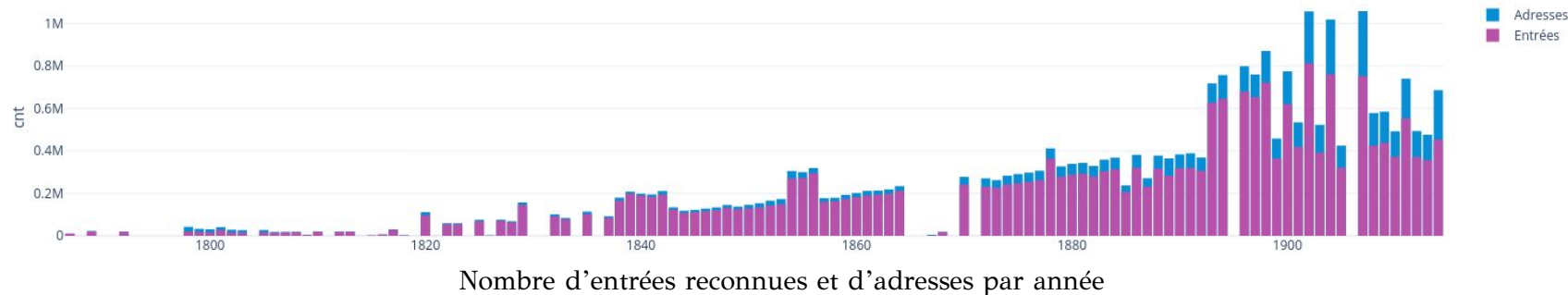
Plan Andriveau-Goujon, 1849 ⌚
Filaire des rues



Atlas municipal des 20 arrds., 1888 ⌚
Filaire des rues & points adresse

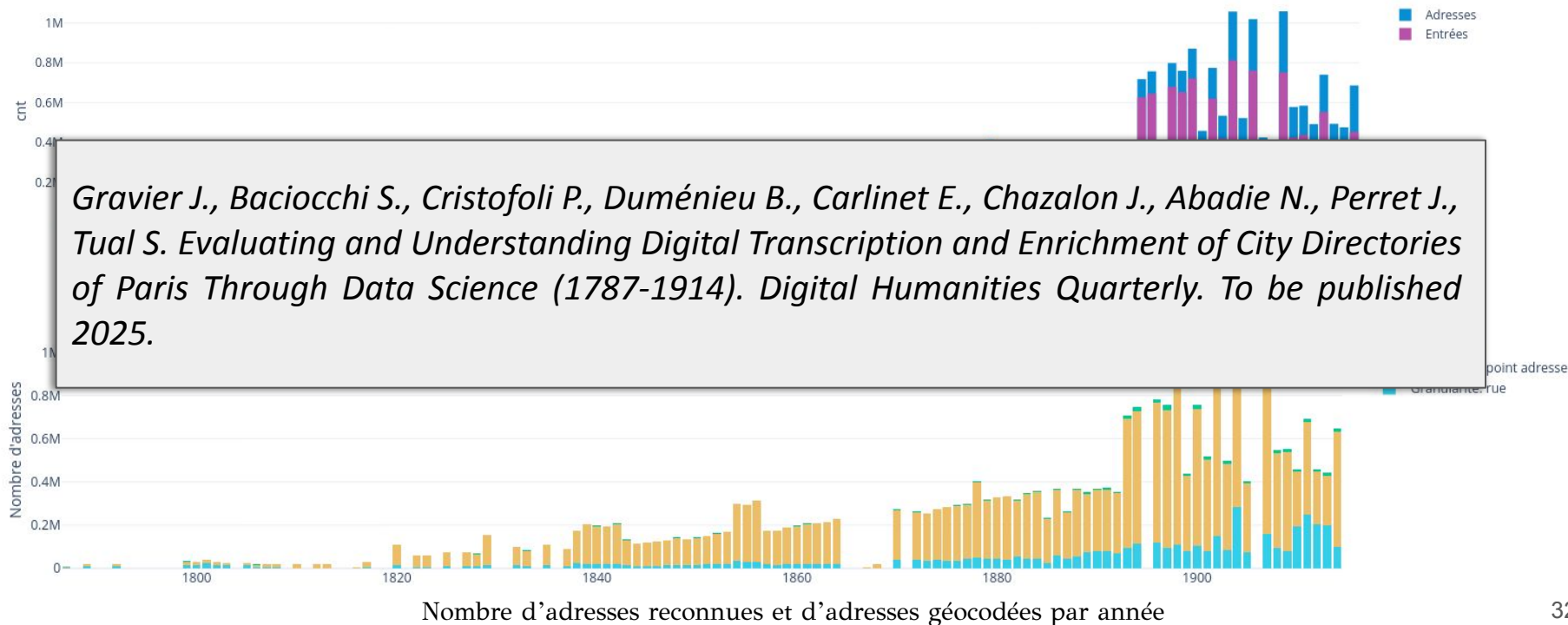
Géocodage des annuaires : aperçu des résultats

V4, octobre 2023 : 96% d'entrées géocodées (V2 septembre 2022 : 66%)



Géocodage des annuaires : aperçu des résultats

V4, octobre 2023 : 96% d'entrées géocodées (V2 septembre 2022 : 66%)



Structuration et liage

Même commerce

Succession

Déménagement

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-Martin, 45.

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.
 Bical, épicier, marché d'Aguesseau, 15.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

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 Doire, fab. de socques, Vert-Bois, 14.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie-du-Temple, 26.
 Bical, épicier, Marche-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Bibliothèque Ste-Geneviève, rue des Sept-Voies
 et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
 Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
 Bibonne, architecte, Magasins, 12.
 Bibron, aide-naturaliste au Jardin-des-Plantes, 10.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie-du-Temple, 26.
 Bical, épicier, Marché-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Les annuaires sont des instantanés décrivant les commerces à un instant t ...

...à partir desquels on peut découvrir de nouvelles connaissances!

Même commerce

Succession

Déménagement

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-Martin, 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.
Bicel, épiciier, marché d'Aguesseau, 15.
Bichard (Mme), Nve-de-Luxembourg, 17.
Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 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 Doire, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épiciier, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, rue des Sept-Voies et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plantes, Carier, 20.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épiciier, Marché-d'Aguesseau, 15.
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1844a - pages 125-126

Mêmes noms, mêmes activités, mêmes adresses → le commerce continue d'exister

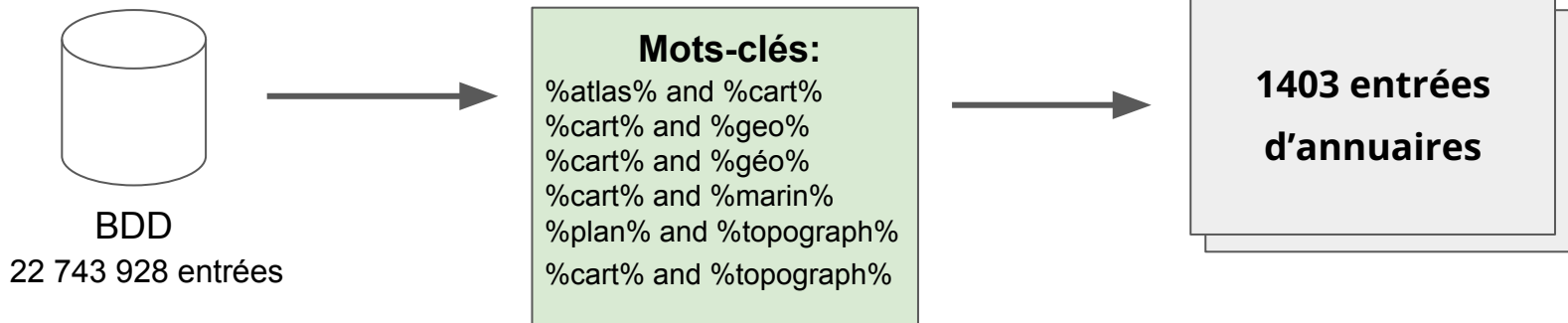
Noms différents, mêmes activités, mêmes adresses → le commerce a changé de propriétaire ?

Mêmes noms, mêmes activités, adresses différentes → le commerce a déménagé ?

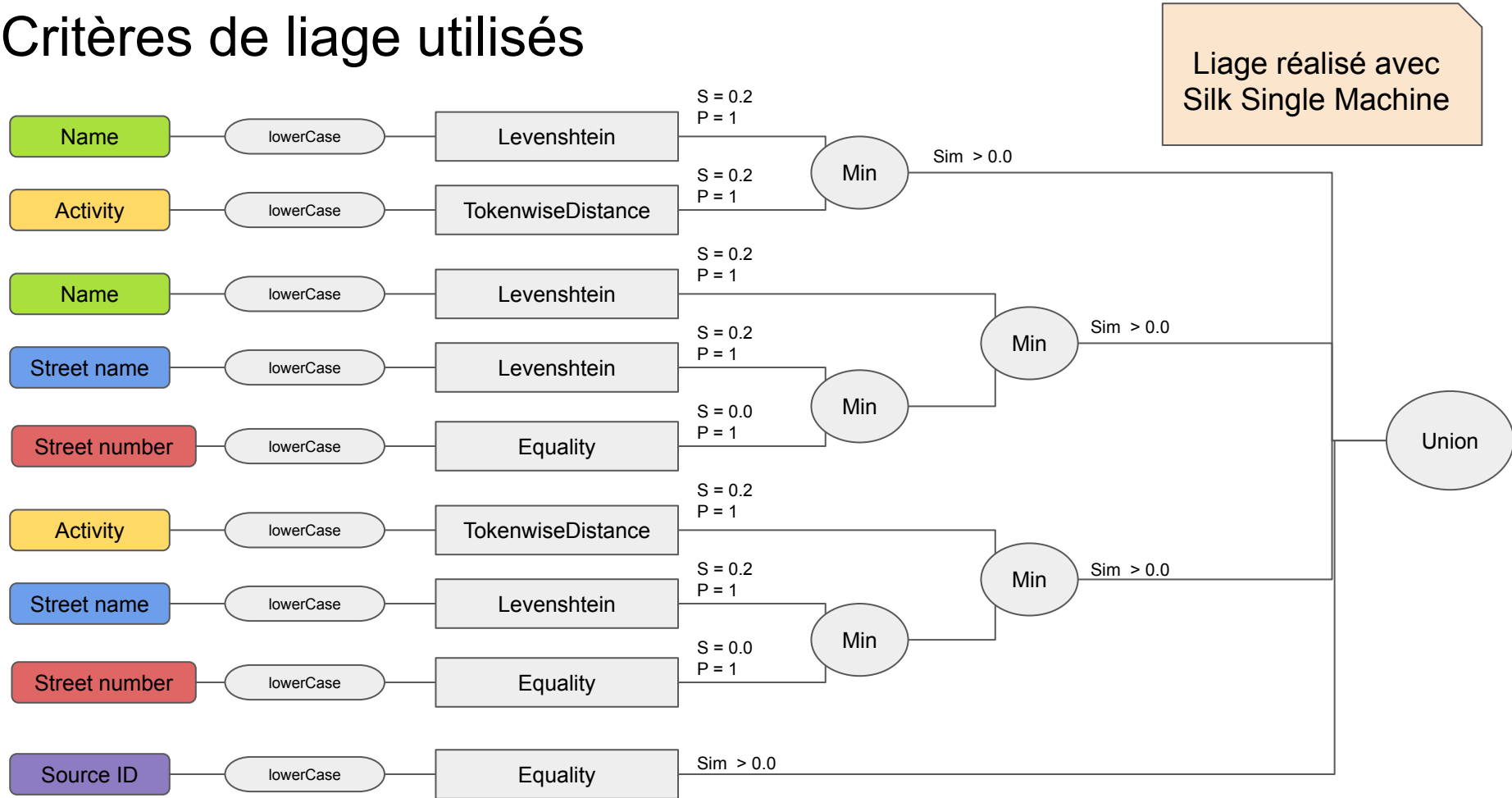
Une approche pour construire des graphes de connaissances géohistoriques professionnels

1. *Sélection des entrées d'annuaires relatives à un type d'activité à l'aide de mots-clés.*
2. *Création de ressources RDF à partir des entrées de la base de données des annuaires sélectionnées.*
3. *Utilisation d'une approche de liage numérique à base de clés pour lier les entrées d'annuaires successives de sorte que :*
 - *Mêmes noms, mêmes activités, mêmes adresses → lien sameAs*
 - *Mêmes noms, mêmes activités, adresses différentes → lien sameAs interprétable comme un déménagement*
 - *Différents noms, mêmes activités, mêmes adresses → lien sameAs interprétable comme un changement de propriétaire*
4. *Visualisation cartographique du graphe de connaissances créé et évaluation qualitative des liens.*

Ex. Graveurs et marchands de cartes



Critères de liage utilisés



Résultats du liage des entrées

Graveurs et marchands de cartes

Approche de liage numérique

20 362 liens

Besoin de définir des seuils

Utilisation de mesures de similarité de chaînes de caractères → compatible avec du texte OCRisé

+

Liens créés par raisonnement

16 046 liens

Propagation de liens
owl:sameAs par transitivité

Au final



36 408 liens owl:sameAs entre les entrées du graphe

Le graphe est publié sur le Web et requêtable : <https://dir.geohistoricaldata.org/sparql>

Visualisation et analyse du graphe



Visualisation des graphes géohistoriques construits à partir des entrées d'annuaires du commerce de Paris (XIX^{ème} siècle)

[Aide ?](#)

Dataset

Graveurs et marchands de cartes et plans ▼

Statistiques du dataset

Filtres

Propriétés

Raison sociale

Ex : nadar

Description

Ex : photo

Adresse

Ex : rivoli

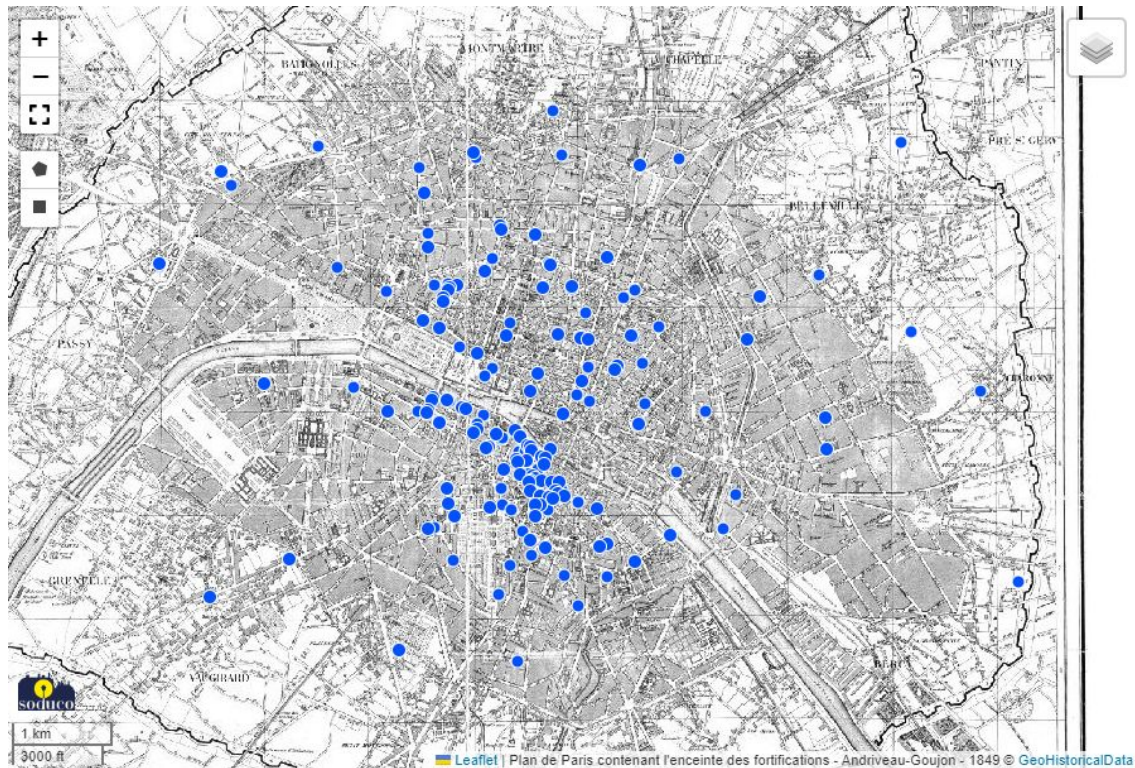
Période

1 km 3000 ft

Leaflet | Plan de P * Sans titre - Bloc-notes is fortifications - Andriveau-Goujon - 1849 © GeoHistoricalData

https://soduco.geohistoricaldata.org/atelier_graphes_geohistoriques_annuaires/

Visualisation et analyse du graphe



Ex : nadar

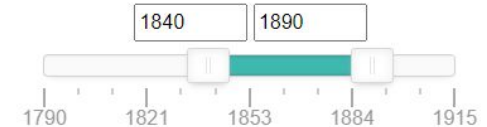
Description

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 **1840-1890**.

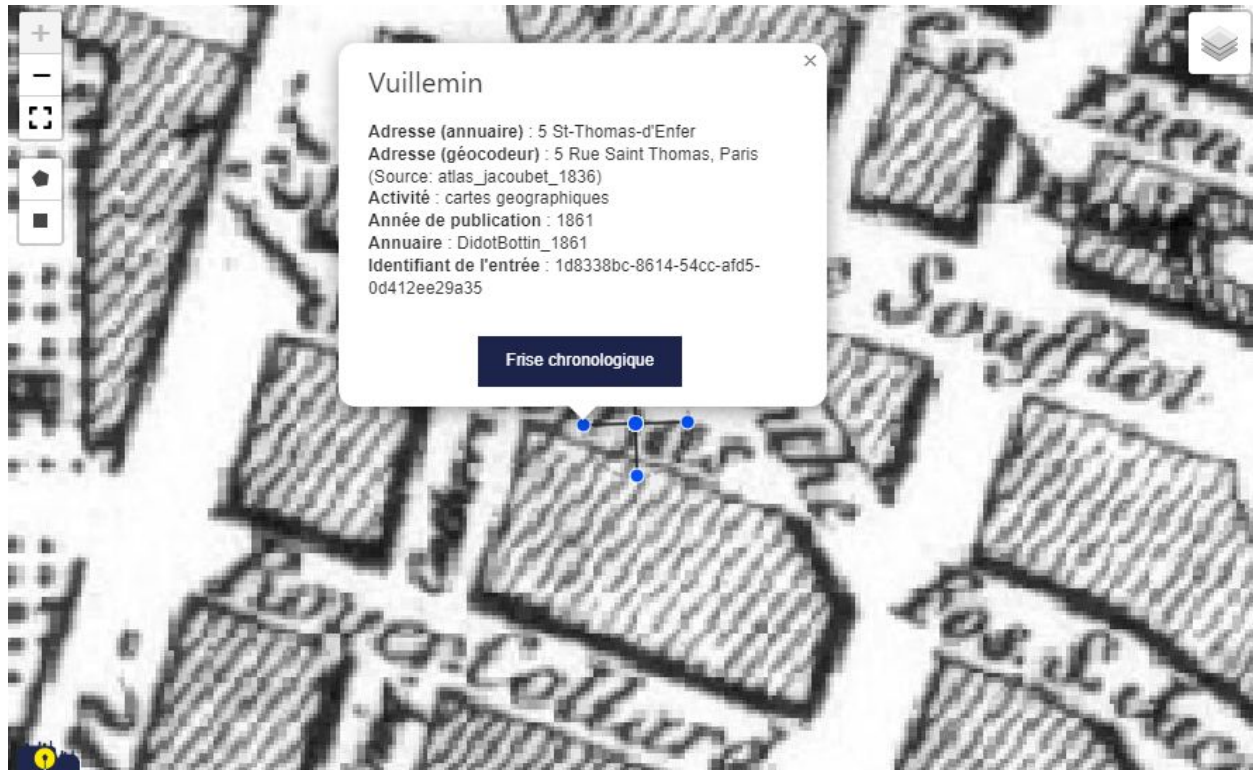
Localisation

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

Lancer la recherche

https://soduco.geohistoricaldata.org/atelier_graphes_geohistoriques_annuaires/

Suivi individuel des commerces dans le temps



The background is a historical map of Paris, showing street names like 'Rue Saint-Thomas', 'Rue de la Harpe', and 'Rue de la Vierge'. A pop-up information box is centered on the map, displaying details for a specific location.

Vuillemin

Adresse (annuaire) : 5 St-Thomas-d'Enfer
Adresse (géocodeur) : 5 Rue Saint Thomas, Paris
(Source: atlas_jacoubet_1836)
Activité : cartes géographiques
Année de publication : 1861
Annuaire : DidotBottin_1861
Identifiant de l'entrée : 1d8338bc-8614-54cc-afd5-0d412ee29a35

[Frise chronologique](#)

Ex : photo

Adresse

Ex : rivoli

Période

1790

1915

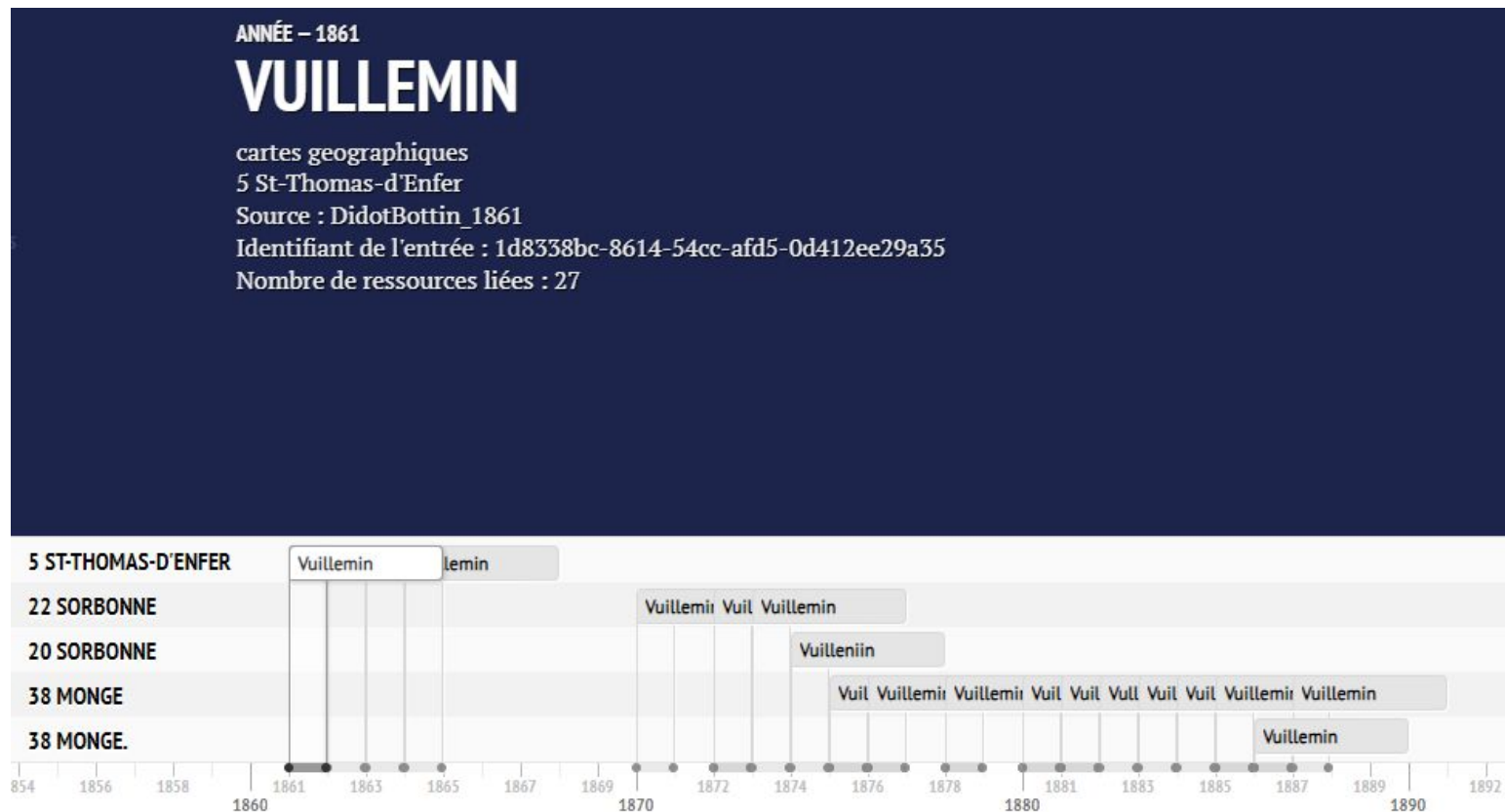


? Le filtre temporel permet de faire varier l'affichage des points préalablement chargés sur la carte sans lancer une nouvelle recherche. Données chargées pour la période **1790-1915**.

Localisation

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

[Lancer la recherche](#)



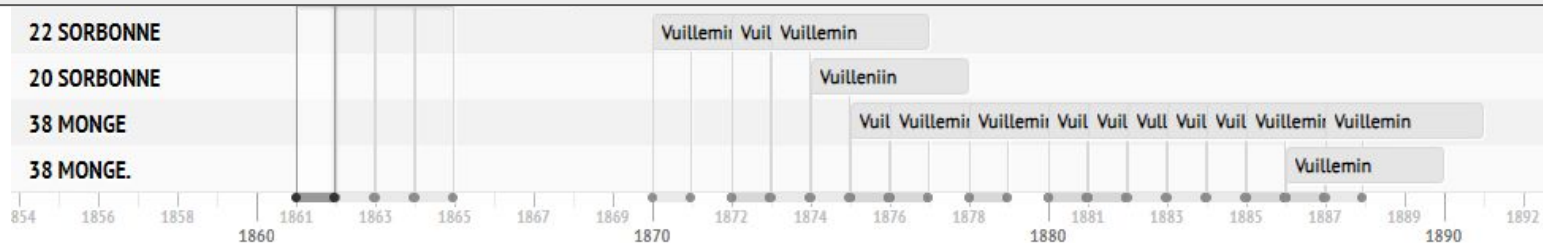
VUILLEMIN

5 St-Thomas-d'Enfer

Identifiant de l'entrée : 1d8338bc-8614-54cc-afd5-0d412ee29a35

Nombre de ressources liées : 27

Solenn Tual, Nathalie Abadie, Bertrand Duménieu, Joseph Chazalon and Edwin Carlinet. Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du XIX^e siècle: application aux métiers de la photographie. *Revue Ouverte d'Intelligence Artificielle*. Numéros 1 et 2, p. 179-200, 2025.



SoDUCo, et après...

→ Tutoriel: Créez votre graphe de connaissances géohistorique professionnel!

https://github.com/solenn-tl/photographies_soduco

→ Nouvelle livraison de données début 2026

→ **Mezanno** : un outil d'extraction de données structurées à partir de corpus sériels (<https://mezanno.xyz/>)

→ **EPHEMER**: Patrimoine imprimé éphémère : des métadonnées d'éphémères à l'exploration des relations professionnelles

(<https://www.pamir.fr/projets-soutenus/ephemer/>)

?