



Analyse de la distribution des noms de rues en France

—

Cahier des charges

Étudiante

Malaurie Gomez, p1706188

Encadrants

Fabien Duchateau

Franck Favetta

Table des matières

I) Contexte.....	1
II) Présentation du projet.....	1
1) Objectifs.....	1
2) Travail à faire.....	2
III) Contraintes.....	3
IV) Livrables.....	3
V) Planning prévisionnel.....	3

I) Contexte

Historiquement, en France, les noms de rues ont été distribués en fonction de logiques différentes qui ont évolué au fil des époques. Les premiers noms de rues apparaissent au Moyen Âge par nécessité de distinguer les lieux les uns des autres. Des noms simples utilisant des repères spatiaux sont choisis : « place de l'Église », « rue du Boucher », « rue de l'hôpital » etc. Les années passant, cette logique est délaissée. Au XVII et XVIII siècles notamment, on voit des noms de personnalité « puissantes » apparaître dans les noms de rue. La révolution va amener de nouveaux noms, qui vont valoriser les idées révolutionnaires : « rue de la liberté », « rue de la nation ». On va commencer à rebaptiser les rues, puis à changer les normes de nomination, à revenir à d'anciennes normes ou à privilégier des références à des personnages locaux.

En clair, de nombreuses normes ont été suivies afin de nommer les rues en France. Aujourd'hui, nous nous retrouvons avec un mélange de noms de rues qui viennent de toutes les époques, et donc des noms de rues de différents types (repères géographiques, noms, valeurs républicaines, ...). Peu d'études existent sur le sujet de la répartition des noms des villes, notamment au niveau national, car les chercheurs se focalisent davantage sur l'évolution des odonymes dans une ville (e.g. Bruxelles ou Reims¹). De telles études permettraient pourtant de pouvoir analyser diverses hypothèses, telles que les tendances de dénomination dans une zone géographique donnée ou l'évolution des noms de rues au fil du temps.

II) Présentation du projet

1) Objectifs

L'objectif de ce projet sera l'analyse de la distribution des noms des rues en France. Comme vu dans la partie précédente, l'origine et le type des noms de rues est varié. On pourra donc s'intéresser à différentes problématiques telles que la recherche des noms de rue les plus populaires au niveau national, la recherche des catégories les plus représentées dans les noms de rue, ou encore les catégories dominantes dans une ville ou région donnée.

Si le temps le permet, on pourrait également analyser plusieurs données venant de dates différentes afin d'appréhender l'évolution des noms de rues au cours du temps.

1 Gérard Quantin. [Les noms de rues de Reims](#). Revue internationale d'onomastique, 2(3):177–192, 1950.

2) Travail à faire

Tâche 1 : Définition de la taxonomie des catégories

1.1) Rapide état de l'art sur les taxonomies utilisées en recherche, comme celle proposée par Badariotti².

1.2) Évaluation de la pertinence de la taxonomie mise en place avec l'apprentissage non supervisé (algorithme de clustering) appliqué à un jeu de données réduit. Cela permettra de voir s'il y a des catégories en trop (pas de cluster / petit cluster), ou s'il en manque (cluster ne correspondant à aucune catégorie définie).

1.3) Choix et définition des catégories et sous catégories qui seront utilisées. La taxonomie sera sur plusieurs niveaux (une catégorie « personnes » pourra être subdivisée en « scientifiques », « militaires », « politiques », etc).

Tâche 2 : Déduplication des données

2.1) Acquisition du [fichier FANTOIR](#), qui regroupe des données sous forme d'une liste des noms de rues en France, triés par commune. Il y a une mise à jour régulière des données par la direction générale des finances publiques.

2.2) Intégration des données dans une base de données plus pratique et performante afin de gérer au mieux les 9 millions d'entrées.

2.3) Création du schéma de base de données.

2.4) Détection des noms de rues identiques (par exemple, rue Louis Pasteur et rue Pasteur). Cet algorithme pourra s'inspirer d'algorithmes existants (par exemple, [algorithme de mesures de similarité appliquées aux chaînes de caractères](#)).

2.4.1) Développement de l'algorithme.

2.4.2) Expertise manuelle d'un jeu de données réduit (quelques centaines de rues) afin d'évaluer la qualité de l'algorithme.

2.4.3) Si l'algorithme n'est pas suffisamment performant, retour à l'étape 2.4.1 : Implémentation d'un nouvel algorithme, puis comparaison de la qualité de leurs résultats. Le plus performant ou une combinaison des deux sera retenu.

Tâche 3 : Classification automatique des rues

- Développement d'un algorithme capable d'associer à chaque odonyme une ou plusieurs catégories de la taxonomie.

- Utilisation soit d'un algorithme d'apprentissage (supervisé, car les catégories à assigner seront déjà connues, avec [scikit-learn](#)), soit d'une base de connaissances externes comme [Wikidata](#), soit des deux.

3.1) Développement de l'algorithme.

3.2) Expertise manuelle afin d'évaluer la qualité de l'algorithme.

3.3) Si l'algorithme n'est pas suffisamment performant, retour à l'étape 3.1 : Implémentation d'un nouvel algorithme afin d'améliorer ou d'affiner les résultats, puis comparaison de la qualité de leurs résultats. Le plus performant ou une combinaison des deux sera retenu.

² Dominique Badariotti. [Les noms de rue en géographie. plaidoyer pour une recherche sur les odonymes](#). In Annales de Géographie, t. 111, n°625, 2002., 2002.

Tâche 4 : Développement d'une application permettant de visualiser la distribution des odonymes par commune

4.1) Mise en place d'une carte choroplèthe avec des valeurs par défaut.

4.2) Interactivité de la carte.

4.2.1) Affichage d'informations au survol d'une commune avec la souris (données utilisées pour la coloration notamment).

4.2.2) Affichage de plus d'informations au clic de la souris sur une commune (villes / catégories les plus populaires de la commune, ...).

4.3) Ajout d'un panel de paramètres permettant de modifier la coloration de la carte.

Par exemple, proposer différentes options de coloration pour la carte (coloration des communes en fonction de la quantité de rues appartenant aux plus populaires en France, ou appartenant à une certaine catégorie).

4.4) Définir et implémenter d'autres paramètres.

III) Contraintes

Backend : Python, Postgres, scikit-learn, Flask (framework Python)

Frontend : HTML, CSS, Javascript, Leaflet (bibliothèque cartographique JavaScript)

Versionning : Gitlab

Organisation : réunion en début et en fin de sprint

IV) Livrables

- Cahier des charges
- Taxonomie des catégories
- Algorithme de déduplication des données et de classification automatique des rues
- Application cartographique
- Rapport final, poster, et vidéo de vulgarisation du travail

V) Planning prévisionnel

	Janvier	Février	Mars	Avril	Mai
Définition de la taxonomie des catégories					
Déduplication et intégration des données					
Classification automatique des rues					
Développement de l'application					
Finalisation + Rapport					
Tests + Débug					