

# Rapport 1 : Scraping de TheFork

---

## Choix Technologiques

Nous avons choisi de développer ce scraper en JavaScript/Node.js pour plusieurs raisons :

- Facilité d'inspection et de réplication des appels API (les requêtes sont nativement en JS dans le navigateur)
- Excellente gestion asynchrone avec async/await
- Large écosystème de bibliothèques pour le web scraping
- Compatibilité naturelle avec notre future application web

## Méthodologie

### Tentative Initiale : Scraping Traditionnel

Notre première approche consistait à scraper directement le site web de TheFork. Cependant, nous avons rapidement rencontré des limitations :

- Détection immédiate par les systèmes anti-bot
- Redirection systématique vers des CAPTCHAs
- Impossibilité d'accéder au contenu de manière fiable

### Solution Adoptée : Interception API

Face à ces obstacles, nous avons opté pour une approche alternative :

1. Inspection du trafic réseau sur TheFork via les outils développeur
2. Identification des endpoints API utilisés par l'application web
3. Réplication des appels API avec les mêmes paramètres et headers
4. Stockage structuré des données dans Redis

## Structure des Données

Le scraper collecte et organise les données suivantes pour chaque restaurant :

- Informations générales (nom, slug, cuisine)
- Données tarifaires (niveau de prix, prix moyen)
- Évaluations (note moyenne, nombre d'avis)
- Localisation (adresse complète, coordonnées GPS)

Les données sont indexées dans Redis selon plusieurs critères :

- Index par prix
- Index par note
- Collections par type de cuisine
- Collections par localité

## Perspectives

Ce premier scraper se concentre uniquement sur la collecte des données des restaurants de TheFork. Les prochaines étapes incluent :

1. Développement d'un scraper complémentaire pour les avis TripAdvisor
2. Création d'un chatbot utilisant le RAG (Retrieval-Augmented Generation) pour interroger intelligemment ces données
3. Développement d'une interface web pour accéder à ces informations

La combinaison de ces données permettra de créer un assistant conversationnel capable de fournir des recommandations personnalisées basées sur une large base de données d'avis et d'informations sur les restaurants.

## Auteurs

Alex Szpakiewicz, Léonard Roussard et Océan Spiess