



# **Rapport hackathon de la Banque de France**

**Pichon Raphaël**  
pichon.rafael@icloud.com

**2020**

## Introduction

Dans le cadre de ce hackathon, l'ensemble du travail a été réalisé en python. Nous nous arrêterons ici sur un descriptif de la démarche répondant à la problématique du hackathon.

<b>1 - web scraping .....</b>	<b>p.3</b>
<b>1.1 - Profil détaillé d'un économiste et de ses publications .....</b>	<b>p.3</b>
<b>1.2 - Liste d'économistes par compétence et institution .....</b>	<b>p.4-6</b>
<b>2 - UI .....</b>	<b>p.7</b>
<b>3 - Limites et atouts .....</b>	<b>p.8</b>

## Bibliothèques utilisées

**UI** : Tkinter

**Web scraping** : Requests, bs4

**OS** : string, sys, os

**PDF** : PdfFileReader, Tabula

## Utilisation

***python3 UI.py ou chmod +x UI.py puis ./UI.py ou python3 UI.py -update***

Ajouter l'option -update pour mettre à jour les données utilisées dans le moteur de recherche

## Descriptif de la problématique

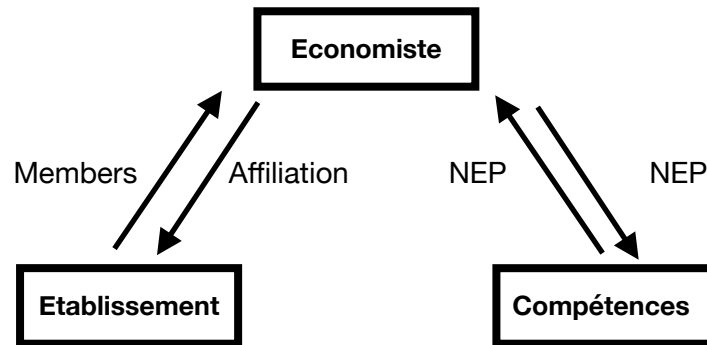
La problématique étant de créer un moteur de recherche avec UI. Ce moteur de recherche devra répondre aux critères suivants :

- 1 - Rechercher un nom d'économiste, et afficher un bref descriptif de l'économiste
- 2 - Rechercher un économiste par domaine de compétence (ex : inflation) et/ou par institution (ex : Banque de France, FMI ...)
- 3 - Recherche et accès aux articles de la sélection

# 1 - Web scraping

Le web scraping se fait uniquement sur le site web suivant : <https://ideas.repec.org/>

Afin de créer le moteur de recherche il faut délimiter les connexions qui existent entre les pages qui nous intéressent (économiste, institution et des compétences). Ci-dessous un schéma résumant les connexions que je vais utiliser pour le moteur :



Connecter les institutions et les compétences n'est pas ici nécessaire, le moteur étant centré sur l'économiste. Une recherche par institution et/ou compétence se fera par 2 requêtes et une jonction (ou non) des économistes trouvés. On gagne donc du temps en évitant de recroiser des informations.

## 1.1 Profil d'un économiste par son nom

La recherche d'un économiste par nom se fait à l'aide d'un fichier regroupant l'ensemble des noms et leurs liens scrapés en amont du programme. Ce fichier peut être mis à jour au choix à l'aide de la fonction **update\_economist()**. Une fois l'adresse obtenue, deux fonctions permettent d'obtenir les noms des articles et leurs liens ainsi que les informations personnelles de l'économiste (respectivement **economist\_articles(url = False,soup = False)** et **economist\_personal\_information(url = False,soup = False)** )

Je viens ainsi scraper les informations des onglets « personal details », « affiliation » et « research ». Ces informations feront office de bref descriptif de l'économiste. L'ensemble des publications de l'économiste est également affiché à l'aide du scraping de l'onglet « Research ».



Jean-Charles Bricongne

Personal details Affiliation **Research** Citations More Corrections

Research output

Export references 85 HTML

Jump to: Working papers Articles

**Working papers**

1. Gigout, Timothee & Bricongne, Jean-Charles, 2019. "Explaining the Persistent Effect of Demand Uncertainty on Firm Growth," MPRA Paper 94228, University Library of Munich, Germany.
2. Jean-Charles BRICONGNE & Alessandro TURRINI & Peter PONTUCH, 2019. "Assessing House Prices: Insights from HouseLev, a Dataset of Price Level Estimates," LEO Working Papers / DR LEO 2675, Orleans Economics Laboratory / Laboratoire d'Economie d'Orleans (LEO), University of Orleans.

Other version(s):

1. Jean-Charles Bricongne & Alessandro Turrini & Peter Pontuch, 2019. "Assessing House Prices: Insights from

A l'aide de l'interface utilisateur, on peut donc obtenir les résultats d'une recherche par nom. Ci-dessous la recherche d'un nom contenant « bric » :

41 économistes trouvés Affichage de 100 au maximum						
Nom de famille	Prénom	Page principale	Adresse	Téléphone	Affiliations	
Hubrich	Stefan				Institut für Volkswirtschaftslehre	Universität Rostock
Perez	M. Fabricio	<a href="http://www.fabricioperez.com/">http://www.fabricioperez.com/</a>			School of Business and Economics	Wilfrid Laurier Univ
Lumineau	Fabrice	<a href="https://www.krannert.purdue.edu/fa">https://www.krannert.purdue.edu/fa</a>			Krannert School of Management	Purdue University
Briceno Avalos	Hernan Ricardo				N/A	
Somogyi	Fabrizius				Institut für Versicherungswirtschaft	School of Finance U
Zubrickas	Robertas	<a href="https://researchportal.bath.ac.uk/en">https://researchportal.bath.ac.uk/en</a>			Department of Economics	University of Bath
Valognes	Fabrice				Centre de Recherche en Économie et Management (C	
Briceno Vicente	Wendy Carolina				Department of Policy Studies	Mount Royal University
Lebraty	Jean-Fabrice	<a href="http://www.gestion-sic.net">http://www.gestion-sic.net</a>	IAE LYON Université Jean Moulin Ly		Institut d'Administration des Entreprises (IAE)	Université
Stinebrickner	Ralph				Department of Economics	University of Western Ontar

Le temps de réponse d'une requête est d'environ 0,2 seconde pour obtenir les informations personnelles des économistes. Il n'est pas possible d'envisager un temps de réponse qui évolue de manière linéaire avec le temps (pour n économistes, le temps de réponse sera de n\*0,2 secondes). De plus pour avoir un programme léger, nous devons optimiser les requêtes et éviter de télécharger toute la base de données. En revanche nous téléchargeons les liens et les descriptions permettant d'accéder rapidement aux catégories sans requêtes. Nous allons alors utiliser une liste de thread et effectuer les requêtes en parallèles. Le temps d'une recherche est alors considérablement réduit.

## 1.2 Profils d'économistes par institution et compétence

Les institutions et compétences viennent, tout comme les économistes, être scrapées et stockées dans les fichiers institutions.txt et NEP.txt.

### 1.2.1 Institutions

Les institutions seront cherchées par mots-clés dans le fichier institutions.txt avec leur lien respectif. Les institutions sont mises à jours avec la fonction **update\_all\_institution(categories\_urls)** à l'aide de la page les triant par catégories JEL. Afin d'optimiser le temps de recherche, nous ne ferons des requêtes que sur les institutions où figurent des économistes. J'impose donc que l'icône « Has members registered with the RePEc Author Service » soit présente lorsqu'on ajoute une institution au fichier texte. Exemple :

#### Belgium

- [European Corporate Governance Institute \(ECGI\), Bruxelles/Brussel](#) 




La recherche des économistes se fait parmi toutes les institutions trouvées avec l'onglet « members » dans une institution.

A l'aide de l'interface utilisateur on peut constater que la recherche par institution est fonctionnelle :

Economics Departments, Institutes ar

Centre for Health Policy  
University of Melbourne

Coordinates Members Corrections

 People who have registered with RePEc an institution:

Their publications  
Find all publications of all authors listed below conveni

Their author profiles  
Authors with publications are listed below with . We h  
count towards the ranking of this institution. We would z

1. Carvalho, Natalie
2. Devlin, Nancy Jay 
3. Rice, James Mahmud 



## 1.2.1 Compétences

Dans cette base de données, les économistes sont classés par NEP. Je vais donc utiliser la classification pour faire une recherche par compétence. Une brève description des catégories étant disponible, la recherche se fera par code NEP et mots-clés. Je viens scraper la page recensant les noms à partir de la page NEP. Exemple ci-dessous avec la catégorie ACC Accounting & Auditing :




### Authors in Accounting & Auditing

This list attempts to categories authors by fields. The procedure is to look at all their papers announced in a [NEP report](#). If 5 or more have appeared in a report, authors are considered to be working within that field. For authors whose first paper is less than 10 years old, having 25% of the papers announced in NEP be announced in this report is sufficient. Note that a paper may appear in several reports.

Below are authors identified to be in the following field: [NEP-ACC \(Accounting & Auditing\)](#). See the [top institutions and authors](#) in this field. ? signals authors with whom we have [lost](#) contact and are seeking a current email address.

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

#### A

Abad, Jorge 	Akcigit, Ufuk	Angrist, Noam
Aboutajdine, Reda	Akgun, Oguzhan	Anwar, Chowdhury Mohammad Sakib
Adhikari, Bibek 	Allen, Franklin	Apps, Patricia Frances
Adu, Kofi Osei	Alm, James	Assa, Jacob
Advani, Arun N. 	Alstadsaeter, Annette	Atanasov, Atanas
Afonso, Sergio Pires	Altshuler, Rosanne	Auerbach, Alan
Agostini, Marisa	Alvarez Martinez, Maria Teresa	Azimi, Mohammad Naim
Ahsina, Khalifa	Amat, Oriol	

#### B

On constate que tous les économistes sont listés sur une seule page, avec cette méthode une requête suffit pour obtenir tous les économistes. La recherche par compétence est donc très rapide.

On peut constater ci-dessous une recherche par compétence :

ATRICA

Agricultural Economics

Banking

**Big Data**

Business Economics

Business, Economic & Financial History

Central & South America

Central & Western Asia

Big Data

Recherche

Exporter la recherche

Quitter

1444 économistes trouvés

Affichage de 100 au maximum

Nom de famille	Prénom	Page principale	Adri	Téléphone	Affiliations
Ali	Searat				Department of Accounting, Finance and EconomicsGriffith Business SchoolGriffith Univer
Arthi	Vellore	http://www.vellorear			Department of EconomicsUniversity of California-Irvine
Abbring	JaapH.	http://jaap.abbring.o	De +31 13 466 8749		https://www.tilburguniversity.edu/research/economics-and-management/graduate-scho
Artunc	Cihan	https://cihanartunc.c			Department of EconomicsMiddlebury College
Adda	Jerome	https://sites.google.c			Dipartimento di Economia "Ettore Bocconi"Università Commerciale Luigi Bocconi
Artiach	Miguel				Departamento de Fundamentos del Análisis EconómicoFacultad de Ciencias Económicas

Une recherche avec les deux conditions (compétence et institution) est réalisée en superposant tous les économistes ou en testant les économistes présents dans les deux sets de données. Ci-dessous un exemple comprenant deux conditions :

Compétence  
 None  
 Accounting & Auditing  
 Africa  
 Agricultural Economics  
 Banking  
 Big Data  
 Business Economics  
 Business, Economic & Financial History  
 Central & South America  
 Central & Western Asia

  
  
  
  
  
 23 économistes trouvés  
 Affichage de 100 au maximum

Nom de famille	Prénom	Page principale	Adri	Téléphone	Affiliations
Castellani	Francois-Daniel				Banque de France
Couaillier	Cyril	<a href="https://sites.google.c">https://sites.google.c</a>			(50%) Banque de France (50%) Department of EconomicsSciences économiquesSciences
Breton	Regis	<a href="https://www.banque">https://www.banque</a>			(99%) Banque de France (1%) Laboratoire d'Économie d'Orléans (LEO)Faculté de droit, d'
De Bandt	Olivier	<a href="https://www.banque">https://www.banque</a>			Banque de France
Vinas	Frederic				Banque de France
Chahad	Mohammed				Banque de France
Heam	Jean-Cyprien	<a href="http://www.crest.fr/p">http://www.crest.fr/p</a>			(50%) Centre de Recherche en Économie et Statistique (CREST) (50%) Banque de France
Mesonnier	Jean-Stephane	<a href="http://pagesperso-o">http://pagesperso-o</a>			Banque de France

Le moteur de recherche comprenant la recherche d'économistes par noms ainsi qu'une recherche filtrée par compétences et/ou institutions est donc fonctionnelle. Afin de continuer et de rendre ce moteur utilisable, nous montrerons dans la prochaine partie l'interface utilisateur (UI) développée.

## 2 - UI

L'UI est faite avec la bibliothèque Tkinter. Deux barres de recherche et un menu déroulant (nom, institution et compétence), trois boutons (quitter, exporter et chercher) et les tableaux d'affichage sont les interfaces où l'utilisateur peut avoir une interaction.

La méthode d'affichage par liste de compétences permet de limiter la recherche par les compétences déjà présentes dans la base de données. Il est envisageable de rajouter la classification JEL et de fusionner l'ensemble.

L'ensemble des informations importantes sont affichées en deux parties: un tableau résumant les informations personnelles des économistes et un autre où sont affichés les articles et leurs liens. Ci-dessous un exemple de recherche d'économistes.

Nom

Institution

Compétence

None

ACC Accounting & Auditing

AFR Africa

AGE Economics of Ageing

AGR Agricultural Economics

ARA MENA - Middle East & North Africa

BAN Banking

BEC Business Economics

BIG Big Data

CBA Central Banking

Aucune compétence

Recherche

Exporter la recherche

Quitter

41 économistes trouvés

Affichage de 100 au maximum

Nom de famille	Prénom	Page principale	Adresse	Téléphone	Affiliations
Briceno Vicente	Wendy Carolina				Department of Policy StudiesMount Royal University
Ularich	HolleyHewitt				Stram Thurmond Institute of Government and Public
Zubrickas	Robertas	<a href="https://researchportal.bath.ac.uk/en">https://researchportal.bath.ac.uk/en</a>			Department of EconomicsUniversity of Bath
Darrigues	Fabrice	<a href="http://catt.univ-pau.fr/live/">http://catt.univ-pau.fr/live/</a>			Centre D'Analyse Théorique et de Traitement des don
Perez	M. Fabricio	<a href="http://www.fabricioperez.com/">http://www.fabricioperez.com/</a>			School of Business and EconomicsWilfrid Laurier Univ
Hubrich	Stefan				Institut für VolkswirtschaftslehreUniversität Rostock
Ibanez Velazaga	FabricioFernando				Instituto de Investigaciones Socio Económicas (IISEC)
Bricco ("Gleck")	Jana				(50%) International Monetary Fund (IMF) (50%) Frankfu
Herve	Fabrice				Centre de Recherche en Gestion des Organisations (C
Stinebrickner	Ralph				Department of EconomicsUniversity of Western Ontar

Date	Nom article	Lien
------	-------------	------

Les requêtes se font à l'aide des deux barres de recherche et de la liste des compétences. Les institutions étant très nombreuses, un passage de mots-clés est suffisant pour trouver l'institution et reste le plus efficace.

Pour faciliter la lecture du résultat de recherche, un clic sur le titre de la colonne classe le contenu par ordre alphabétique.

Il est possible de double-cliquer sur un économiste lorsqu'une recherche mène à plusieurs économistes et qu'on souhaite consulter ses articles. En double-cliquant sur un article, une page web s'ouvre sur l'article. La navigation est simplifiée et évite de devoir copier le lien et chercher l'article soi-même.

Ci-dessous un double clic sur un économiste de la recherche précédente. Afin de faciliter la consultation de la liste des publications, l'information de la date de publication a été ajoutée.

Quitter						
1 économiste trouvé						
Nom de famille	Prénom	Page principale	Adresse	Téléphone	Affiliations	
Bricongne	Jean-Charles				Banque de France	

Date	Nom article	Lien
2005	Les comptes financiers de la Nation en 2004 : forte poussée de l'endettement immobilier des ménages	<a href="https://ideas.repec.org/a/bfr/bulbf/200613701.html">https://ideas.repec.org/a/bfr/bulbf/200613701.html</a>
2006	Les comptes financiers de la Nation en 2005 : nouvel essor de l'endettement des ménages, reprise de celui des entreprises	<a href="https://ideas.repec.org/a/bfr/bulbf/200614902.html">https://ideas.repec.org/a/bfr/bulbf/200614902.html</a>
2006	Les statistiques financières permettent-elles d'appréhender les évolutions liées à la mondialisation ?	<a href="https://ideas.repec.org/a/bfr/bulbf/200615604.html">https://ideas.repec.org/a/bfr/bulbf/200615604.html</a>
2006	Évolution du taux d'intermédiation financière en France (1994-2004)	<a href="https://ideas.repec.org/a/bfr/bulbf/200614603.html">https://ideas.repec.org/a/bfr/bulbf/200614603.html</a>
2008	Désintermédiation financière ou diversification internationale ? Le cas des pays développés	<a href="https://ideas.repec.org/p/drm/wpaper/2008-22.html">https://ideas.repec.org/p/drm/wpaper/2008-22.html</a>
2009	Firms and the global crisis: French exports in the turmoil	<a href="https://ideas.repec.org/p/bfr/banfra/265.html">https://ideas.repec.org/p/bfr/banfra/265.html</a>
2009	Firms and the global crisis: French exports in the turmoil	<a href="https://ideas.repec.org/p/bfr/banfra/265.html">https://ideas.repec.org/p/bfr/banfra/265.html</a>
2010	De la crise financière à la crise économique. L'impact des perturbations financières de 2007 et 2008 sur la croissance de	<a href="https://ideas.repec.org/a/ors/ecstat/estat_0336-1454_2010_num_438_1_9591.html">https://ideas.repec.org/a/ors/ecstat/estat_0336-1454_2010_num_438_1_9591.html</a>
2010	Exports and sectoral financial dependence: evidence on French firms during the great global crisis	<a href="https://ideas.repec.org/p/feb/ecbwp/20101227.html">https://ideas.repec.org/p/feb/ecbwp/20101227.html</a>

Il est également possible d'enregistrer une recherche sous forme de fichier CSV en appuyant sur le bouton 'exporter la recherche'. Deux fichiers seront créés pour les articles et les économistes.

### 3 - Limites et atouts

Les limites des solutions retenues sont les suivantes :

- Un grand nombre de requêtes est lancé sur le site hébergeant la base de données. Afin de ne pas se laisser saturer par un grand nombre de requêtes, il est impossible d'en faire plus de 100 en quelques secondes. Nous sommes donc restreints à l'**affichage des données personnelles de 100 économistes maximum**. Cependant une recherche plus précise permettra à l'utilisateur d'obtenir le résultat souhaité.

Les atouts des solutions retenues sont les suivants :

- Pour une recherche relativement précise, la recherche prend très **peu de temps**
- Seul les menus et les classements du site étant sauvegardés dans notre solution, le programme est **très léger**.
- La recherche n'ayant que 100 requêtes maximum à faire, le **temps de recherche est borné** entre moins d'une seconde à une dizaine de secondes
- La mise à jour des données téléchargées est faisable rapidement (~ 30 secondes) et peut être faite quotidiennement. Ainsi le **moteur de recherche est facilement et rapidement actualisable**.