

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MAKING-OF : LES PROFESSEURS D'UNIVERSITÉ

Travail présenté à Jean-Hugues Roy

Dans le cadre du cours

EDM5240 : Technologies de l'informations appliquées au journalisme

Groupe 30

par

Nicolas St-Germain STGN26029703

Mercredi 17 avril 2019

À la base, l'idée avec ce projet était d'avoir une représentation visuelle des enseignants dans les différentes universités du Québec et de les comparer selon le sexe et l'origine ethnique dans le but d'infirmer ou non différentes hypothèses.

Certes, un coup que cette liste est complète, plusieurs autres outils peuvent être créés. Premièrement, il est possible de concevoir une carte du Québec ayant comme point chacune des universités et leur campus. Cette carte interactive permettrait à un étudiant du Cégep ou du secondaire, désirant savoir où il doit étudier, d'y entrer son métier ou un mot-clé et une liste d'expert apparaît sur la carte. Il peut donc déterminer en un clic où sont offerts les cours et savoir le nombre d'experts que compte le domaine. Ce qui lui permet de savoir aussi quelle université donne une importance au domaine d'étude qui l'intéresse.

Deuxièmement, ces données permettraient de créer une banque d'experts accessibles aux différents journalistes. Cela permettrait du même coup de varier les experts reçus sur les plateaux ou ceux cités lors de reportages.

Pour mener à bien ce moissonnage, nous avons utilisé les répertoires de professeurs qu'offrent les universités avec comme objectif de passer au travers des 19 établissements. Cependant, nous avons rapidement remarqué des irrégularités dans la façon dont les universités présentaient leurs données. Nous avons donc diminué notre nombre d'universités pour plusieurs raisons. Premièrement, les établissements n'usent pas tous de la même méthode pour rendre accessibles les données de leurs enseignants. Par exemple, l'UQAM met à la disposition de quiconque une banque de données séparée par département — comportant toutefois certains professeurs dans deux départements différents —, ce qui rend le moissonnage plus facile. À l'opposé, l'Université McGill — ayant finalement été retiré de nos recherches — compte plusieurs sites différents pour chacune de ses disciplines, ce qui rend le moissonnage très long. Il aurait fallu utiliser Selenium comme ce fut le cas avec l'Université Laval<sup>1</sup> qui avait aussi réparti ses données un peu partout. Deuxièmement, la tâche était ardue même si nous étions trois à faire le moissonnage, chaque site comptait son lot de surprise qui a nécessité un approfondissement de

---

<sup>1</sup> Script fait par l'étudiant de la polytechnique Antoine Lefebvre-Brossard

nos connaissances par l'entremise de recherches sur *Stackoverflow* ou avec l'aide des étudiants de la polytechnique.

Bref, après avoir recueilli toutes nos données, nous avons encore une fois éliminé une école — l'École Nationale d'administration publique (ENAP) —, car il était difficile de créer une expertise commune qui ressemblait aux autres universités. Les universités avaient aussi leurs propres termes pour décrire leur programme rendant parfois difficile l'association entre universités. C'est pour cela que nous avons créé les catégories suivantes : sciences appliquées, affaires, sciences sociales, éducation, sciences santé, sciences pures, arts, lettres, droit et autres. Par contre, comme les noms d'expertise ou même de département variaient, cela devenait difficile d'assurer avec 100 % de certitude la corrélation entre ses expertises. Ce qui veut dire que nos données ne sont peut-être pas assez fiables pour en faire un comparatif solide. Il est possible que la catégorie *sciences sociales* de l'UQAM ne compte pas les mêmes expertises que celui de l'Université Laval, ce qui ne permet pas réellement de les comparer, car le contenu n'est pas le même.<sup>2</sup>

En moissonnant les données, un autre problème est survenu alors qu'à l'intérieur même des fiches de professeurs, certaines données étaient manquantes. L'outil « *if {value} is None : etc*<sup>3</sup> », a été extrêmement utile pour permettre au *code* de continuer son travail même si la donnée n'est pas inscrite. L'outil « *try & except*<sup>4</sup> » était aussi essentiel lorsque certaines données ne figuraient pas toujours au même endroit dans les fiches. Cela m'a permis de créer des exceptions dans mon code selon la position des éléments.<sup>5</sup> Ces deux outils ont été les plus utiles. J'ai aussi tenté de comprendre *Selenium* pour faire le moissonnage de l'Université du Québec à Rimouski, mais en vain. L'Université permettait d'avoir accès à la liste de tous les professeurs, en cliquant

---

<sup>2</sup> Voir graphique «distribution» sur le site <https://quebec-professors-235619.appspot.com>

<sup>3</sup> Stackoverflow. (2016). Python `if x is not None` or `if not x is None`?. Récupéré de <https://stackoverflow.com/questions/2710940/python-if-x-is-not-none-or-if-not-x-is-none>

<sup>4</sup> W3Schools. (s. d.). Python Try Except. Récupéré de [https://www.w3schools.com/python/python\\_try\\_except.asp](https://www.w3schools.com/python/python_try_except.asp)

<sup>5</sup> Se référer au *script* de la Télug.

sur « rechercher », certes l'URL demeurait le même qu'avant de cliquer dessus ce qui ne permettait pas au *script* de trouver les données.<sup>6</sup>

D'autres problèmes sont survenus dès que nous avons obtenu tous les professeurs désirés. L'API *NamSor*<sup>7</sup>, utilisé par les étudiants de la Polytechnique pour associer le nom des professeurs à leur origine ethnique, fonctionne bien, mais il ne donne pas un résultat toujours précis comme il est possible de le constater en annexe 1. En voyant ces informations-là, l'équipe de journaliste s'est demandé s'il était possible de faire un article fiable. C'est pour cela que l'angle de notre papier est plutôt axé sur la fiabilité des API et la façon de contre vérifier les résultats obtenus.

Dans l'ensemble, nous avons visé pas mal trop gros pour ce travail, ce qui nous a compliqué la tâche. Nous aurions clairement pu décider de comparer que deux établissements qui se ressemblent. Cependant, la tâche était plutôt difficile dû à l'éclatement des modes de diffusion des données par les universités et le manque d'uniformisation.

---

<sup>6</sup> UQAR. (s. d.). Répertoire des chercheurs et expertises. Récupéré de <http://annuaire.uqar.ca/AnnuaireWeb/RechAnnuaireSpec/RechAnnuaireSpec>

<sup>7</sup> NamSor. (s. d.). Home page. Récupéré de <https://www.namsor.com/>

## ANNEXE 1

Response body

```
{
  "id": null,
  "name": "Nourhene Chaalia",
  "score": 5.893867595391429,
  "country": "TN",
  "countryAlt": "DZ",
  "region": "Africa",
  "topRegion": "Africa",
  "subRegion": "Northern Africa"
}
```

Download

*Ceci est le résultat obtenu lorsqu'un nom entre dans l'API NamSor. Il donne un score, dans ce cas-ci, 5,90 sur 10 pour noter son degré de certitude. Les étudiants de la polytechnique ont simplement gardé la section « country ».*