Université de Technologies de Troyes

_

MASTÈRE SPÉCIALISÉ EXPERT BIG ANALYTICS ET MÉTRIQUES

Projet Neo4j : Application marmiton

Auteurs: Jean-Baptiste Berry Arthur Frémond Camille André

Intervenant : Jérôme Baton

31 janvier 2018





Table des matières

1	\mathbf{Mis}	se en place de la base de données
		Collecte des recettes de marmiton.org
	1.2	Mise en place de la base de donnée orientée graph
2	Dév	veloppement de l'application
	2.1	Py2Neo et développement web
	2.2	Pour aller plus loin: le machine learning
		2.2.1 Déroulement
		2.2.2 Modèle de deep learning
	2.3	Modèles choisis
	2.4	Rendu Final
		2.4.1 Utilisation sans Machine Learning
		2.4.2 Utilisation avec Machine Learning

Partie 1

Mise en place de la base de données

1.1 Collecte des recettes de marmiton.org

Pour la collecte des informations sur les recettes de cuisine via le site marmiton.org, nous avons dans un premier temps développé un scraper permettant de récupérer des .json de ces recettes. Le détail du scraper est consultable dans le notebook Jupyter ¹ dédié et joint à ce mémoire.

Après récupération de toutes les recettes, nous obtenons un fichier .json dont chaque élément se présente comme ceci :

```
{"@context": "http://schema.org", "@type": "Recipe",
2
    "aggregateRating": { "@type": "AggregateRating", "bestRating": 5,
    "ratingValue": 4, "reviewCount": 13, "worstRating": 0},
3
    "author": "Am\u00e9ly", "cookTime": "PT25M",
4
    "datePublished": "2006-10-23T11:43:00+02:00",
5
6
    "description": "haricot plat, tomate cerise, ail, huile d'olive",
7
    "image": "https://image.afcdn.com/recipe/20130719.jpg",
    "name": "Haricots plats (cocos) aux tomates cerises",
8
    "prepTime": "PT10M"
9
    "recipeIngredient": ["1 haricot plat", "15 tomate cerise", "3 gousse
10
11
     ail", "2 cuill\u00e8re \u00e0 soupe huile d'olive"],
    "recipeInstructions": "Laver et \u00e9cosser les haricots plats les
12
     mettre de c\u00f4t\u00e9. Couper en morceaux les gousses d'ail et
13
14
     les faire revenir dans une cuill\u00e8re \u00e0 soupe d'huile d'olive.
     Ne pas les faire griller. Ajouter les haricots plats et mettre sur feux
15
16
     doux (important) pendant 15 mn en remuant de temps en temps.
     Quand les haricots sont ramollis (pas trop), ajouter les tomates
17
     cerises pour le reste de la cuisson.", "recipeYield": "4 personnes",
18
     "totalTime": "PT35M"}
```

On pourra noter un petit souci d'encodage qui sera réglé pendant le traintement sous python. C'est donc ce fichier *.json* que nous importerons sous Neo4j et qui définira notre base de données.

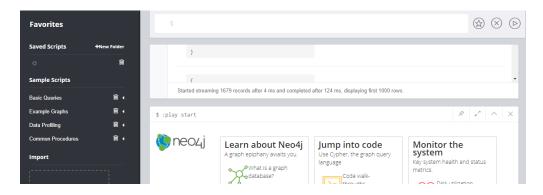
1.2 Mise en place de la base de donnée orientée graph

La première étape est d'installer NEO4J en local, pour y avoir accès dans votre navigateur (http://localhost:7474/browser/). On pourra trouver la dernière version ici : https://neo4j.com/download/other-releases/#releases. Dans un souci de praticité, on nommera le fichier racine NEO4_JHOME. Ensuite, avec la console windows, on se place dans l'emplacement du dossier et on exécute la commande bin\neo4j console.

^{1.} Scraper Jupyter.ipynb

Ainsi, nous avons:

Et on peut donc se connecter à Neo4j à l'adresse suivante : http://localhost:7474/. On ne touche pas au *Bolt* par défault ni au nom d'utilisateur et mot de passe qui sont également définis par défaut comme neo4j. Puis, nous arrivons sur l'interface utilisateur de neo4J :



On va à présent effectuer une modification dans neo4j.conf et plus précisément dans le dossier NEO4J_HOME/conf. De cette manière, nous pouvons utiliser une APOC (Awesome Procedure on Cypher) qui nous permet d'importer nos données au format .json dans Neo4j:

```
# This setting constrains all `LOAD CSV` import files to be under the `import` direc
# allow files to be loaded from anywhere in the filesystem; this introduces possible
# `LOAD CSV` section of the manual for details.
dbms.directories.import=import
apoc.export.file.enabled=true
apoc.import.file.enabled=true
apoc.import.file.use_neo4j_config=true
apoc.trigger.enabled=true
apoc.ttl.enabled=true
dbms.security.procedures.unrestricted=apoc.*
dbms.security.procedures.white list=apoc.load.
dbms.security.allow_csv_import_from_file_urls=true
# Whether requests to Neo4j are authenticated.
# To disable authentication, uncomment this line
#dbms.security.auth_enabled=false
# Enable this to be able to upgrade a store from an older version.
```

Ensuite, nous téléchargeons le .jar d'APOC correspondant à notre version de neo4j qui est déposé dans NEO4J_HOME/plugins, comme par exemple, apoc-3.3.0.1.jar pour la version 3.3.2 de neo4j. Ensuite, nous pouvons redémarrer neo4j et accéder à l'APOC. Nous pouvons également faire : CALL apoc.help("apoc") afin de s'assurer que tout s'est déroulé correctement. Puis, nous allons créer notre base de données avec la commande suivante :

Listing 1.1 – Import de la base de donnée dans neo4j.

```
WITH apoc.text.urlencode('marmiton.json') AS url
CALL apoc.load.json('file:///'+url) YIELD value AS recette
MERGE (r:Recette{name: recette.name}) ON CREATE
SET r.Instructions = recette.recipeInstructions,
```

```
5 r.Quantite = recette.recipeIngredient
6 FOREACH (value IN SPLIT(toLower(recette.description),', ') |
7 MERGE (e:Element{ingredient: value})
8 CREATE (e)-[:COMPOSE]->(r))
```

Une fois les noeuds et les liens implémentés dans neo4j, nous obtenons une base de données contenant toutes les recettes. Ci-dessous, un élément de la base de données *recette* tel que présenté dans neo4j.

```
"name": "Haricots plats (cocos) aux
tomates cerises",
  "instructions": "Laver et écosser
les haricots plats les mettre de côté.
Couper en morceaux les gousses d'ail
et les faire revenir dans une cuillère
à soupe d'huile d'olive. Ne pas les
faire griller. Ajouter les haricots
plats et mettre sur feux doux
(important) pendant 15 mn en remuant
de temps en temps. Quand les haricots
sont ramollis (pas trop), ajouter les
tomates cerises pour le reste de la
cuisson.",
  "quantite": [
   "1 haricot plat",
   "15 tomate cerise",
   "3 gousse ail",
    "2 cuillère à soupe huile d'olive"
```

FIGURE 1.1 – Extrait d'une recette sous neo4j avec ses instructions et la quantité d'ingrédients nécessaires.

Dans un souci de clarté et en limitant l'affichage à 3 noeuds, nous obtenons le graph suivant :

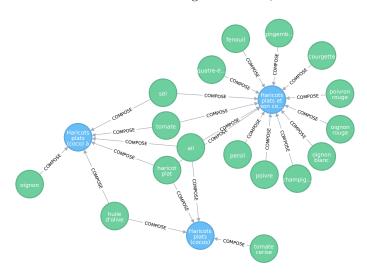


FIGURE 1.2 – Extrait de la base de données Neo4j créée.

Chaque noeud est représenté par un nom de recette, auquel est lié les noeuds des ingrédients nécessaires à sa préparation. Par ailleurs, les noeuds correspondant aux recettes sont composés de 3 propriétés :

• name : le nom de la recette

• Instructions : les étapes de la recette

• quantite : quantité d'ingrédients nécessaire

Parallèlement, les noeuds d'ingrédients seront uniquement composés du nom de ceux-ci. Enfin, plusieurs recettes partagerons bien sûr les mêmes ingrédients.

On gardera à l'esprit pour la suite qu'avec les relations entre les différents noeuds de la base de données, le nombre d'informations générées croît très rapidement. A titre d'exemple, on affiche ci-dessous les noeuds et liens pour 10 recettes de notre base données.

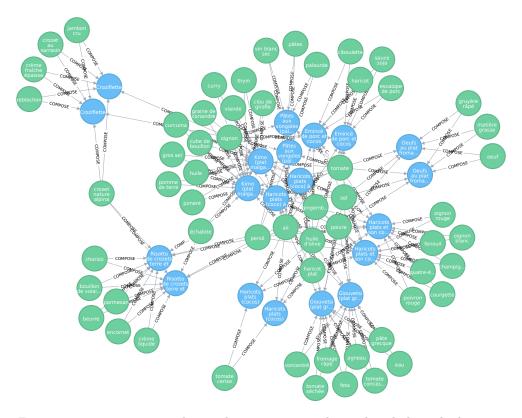


FIGURE 1.3 – 10 recettes et leurs relations aux ingrédients dans la base de données.

Partie 2

Développement de l'application

2.1 Py2Neo et développement web

Pour pouvoir maintenant effectuer des requêtes cypher depuis python, nous utilisons le package py2neo.

```
1 from py2neo import authenticate, Graph
2 authenticate("localhost:7474", "user", "pass")
3 sgraph = Graph("http://localhost:7474/db/data/")
```

Dans un premier temps, on initialise les paramètres d'authentification pour ensuite pouvoir se connecter à la base de données orientée graph ainsi authentifiée. Pour les requêtes que nous effectuerons via l'application, nous préparons des fonctions que nous pourrons utiliser plus tard. Ainsi, nous crééons une fonction recette(), prenant en argument une liste d'ingrédients qui sera rentrée par l'utilisateur sur l'application dans les champs mis à disposition, et qui renvoie toutes les recettes correspondant à ces ingrédients.

Pour le développement web sous python, nous avons utilisé Flask (framework open-source). Nous avons choisi d'ordonner notre dossier du projet de la manière suivante :

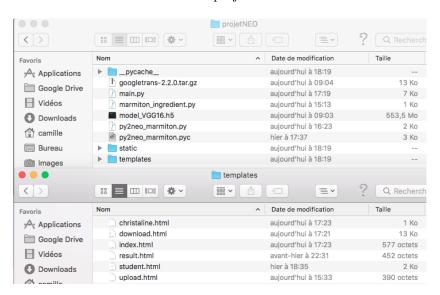


FIGURE 2.1 – Organisation des fichiers de l'application.

Il y a un dossier static qui contient notre fichier css, les photos utilisées sur le site et une police de FontSquirrel 3D pour le titre de l'application sur la page d'accueil. Il y a un dossier templates qui contient tous les fichiers HTML utilisés autour de ce projet et les images téléchargés dans la

page d'accueil pour le Machine Learning. Notre algorithme de ML ira chercher des photos dans ce dossier. Nous y reviendrons plus tard. Enfin, on retrouve tout en haut du projet nos scripts python, qui vont nous permettre de lancer notre application et d'exécuter nos requêtes sur la base de données neo4j.

A présent, regardons le programme principal, main.py:

```
7 #librairies
8 from flask import Flask, render_template, request
9 from flask_uploads import Uploadset, configure_uploads, IMAGES
0 from py2neo_marmiton import recette, etapes
```

On définit deux strings / html que l'on va utiliser plus tard :

```
html_str =""
<!dottype htm>
<!dottype htm>
<!dottype htm>
<!dottype htm->
<!meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<!link rel="stylesheet" href="{{ url_for('static', filename='style.css') }}" type="text/css" charset="utf-8"/>
cbody>
cp>
Dans votre frigo, vous avez les ingrédients suivants : {{result}}.cbr/>cbr/>cles plats que nous vous proposons sont : {0} 
cybeuillez choisir une recette parmi les résultats :
cform method="POST" action="http://localhost:5000/index">
cybelect name="nom" size="l">
{puces}

cyselect name="nom" size="l">
cylect parmi les résultats :
cyselect name="nom" size="l">
cylect>
cinput type = "submit" value = "submit" />

cylody>

cylody>

cylody>

cylous avez choisi la recette suivante : {{camille}}.cbr/>cbr/>cbr/>colic les étapes de la recette : dbr/>cbr/>cylous style="list-style-type:square">
cylous les étapes de la recette : dbr/>cbr/>cylous cylous cyl
```

On définit notre application de la manière suivante, puis on lui indique le dossier dans lequel les photos téléchargées depuis la page d'accueil seront sauvegardées.

```
32
33 app = Flask(_name__)
54
55 photos = UploadSet('photos', IMAGES)
56
57 app.config['UPLOADED_PHOTOS_DEST'] = 'static/img'
58 configure_uploads(app, photos)
```

Puis, on définit, les différentes pages, paramétrées avec Flask. On commence par la première page :

```
60 @app.route('/')
61 def student():
62    return render_template("student.html")
63
```

Regardons le fichier HTML, student.html:

```
1 chtml>
2 <meta charset="utf-8">
2 <meta charset="utf-8">
3 chead>
4 clink rel="stylesheet" href="{{ url for('static', filename='style.css') }}" type="text/css" charset="utf-8">
5 cimg class="right" src="/static/neoCook.png" border="0" height="252px" width="256,5px"/>
6 cdiv id="main">
7 chl align="center">MARMIBON APP</hl>
8 / cdiv>
9 </head>
10 chody>
11 cimg class="left" src="/static/frigoMB.png" border="0" height="376px" width="328px"/>
12 cform action = "http://localhost:58000/result" method = "POST">
13 cdiv align="center" class="recherche">
14 ch2>Ici, rentrez manuellement vos ingrédients</h>
15 cp>ingrédient !<input type = "text" name = "ingredient" />
16 cp>ingrédient 3<input type = "text" name = "ingredient" />
17 cp>ingrédient 3<input type = "text" name = "ingredient" />
18 cp>ingrédient 5<input type = "text" name = "ingredient" />
19 cp>ingrédient 7<input type = "text" name = "ingredient" />
20 cp>ingrédient 7<input type = "text" name = "ingredient" />
21 cp>ingrédient 7<input type = "text" name = "ingredient" />
22 cp>ingrédient 9<input type = "text" name = "ingredient" />
23 cp>ingrédient 9<input type = "text" name = "ingredient" />
24 cp>ingrédient 9<input type = "text" name = "ingredient" />
25 cp>sinput diput type = "text" name = "ingredient" />
26 c/div>
27 c/form action = "total characteries" /> characteries = "total characteries = "total characteries" /> characteries = "total characterie
```

Pour ce qui est du FrontEnd, nous nous organisons comme suit. On appelle notre CSS grâce à la balise link tout en haut. Sa syntaxe est particulière à Flask. On insère une photo et on va définir une fiche d'action sur la page http://localhost/5000/result et on va récupérer les ingrédients dans une liste python. Puis, on définit une zone pour télécharger une photo d'ingrédient, l'enregistrer dans le dossier static/img où elle pourra être utilisée par un algorithme de Machine Learning pour être labélisée (par exemple, on telecharge une photo de tomate et l'algorithme reconnait que c'est une tomate). Nous pouvons voir ci-dessous la définition de la page http://localhost/5000/result:

Le script sur Flask pour la page d'index est le suivant (Voir les commentaires du script pour plus de détails) :

2.2 Pour aller plus loin: le machine learning

2.2.1 Déroulement

De manière avant-gardiste, nous avons souhaité intégrer dans notre application du Machine Learning notamment à travers la mise en place d'un réseau de neurones. Le but est de permettre à l'utilisateur de charger une photo de l'intérieur de son réfrigérateur pour que l'application lui propose des recettes automatiquement.

Pour se faire, nous devons décomposer le processus en plusieurs étapes clés en termes de Machine Learning :

- Localisation des différents ingrédients présents dans l'image de l'intérieur du réfrigérateur.
- Découpage de l'image en multiples images composées chacune d'un unique ingrédient.
- Analyse des images et reconnaissance (labélisation) des ingrédients.

Une fois ces étapes faites, les ingrédients peuvent être regroupés dans une liste et transférés sous forme d'une requête à la base de données Neo4j. Cette étape de requêtage est identique à celle implémentée lorsque l'utilisateur spécifie ses ingrédients manuellement dans les champs laissés à sa disposition.

2.2.2 Modèle de deep learning

Les modèles les plus performants en termes de reconnaissance et traitement d'images sont les modèles à convolution (CNN). En effet, les réseaux de neurones convolutifs (CNN) sont principalement utilisés pour classer des images. Ils comportent deux parties bien distinctes :

- Une première étape composée de couches convolutionnelles
- Une deuxième étape composée de couches connectées ordinaires

La première partie d'un CNN, partie convolutive à proprement parler, fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image.

Puis, les cartes de convolutions sont mises à plat et concaténées, constituant l'entrée de la 2eme étape. La deuxième partie est simplement faite de plusieurs couches de perceptrons connectées, exactement comme un réseau de neurones ordinaires. La sortie est une dernière couche comportant un neurone par catégorie pour classifier l'image.

Il existe de multiples modèles CNN bâtis sur ce principe. Chacun possède une architecture différente par le nombre de filtres, leurs tailles, etc. Chaque CNN est conçu dans un but bien précis.

2.3 Modèles choisis

Il est possible sur les bases décrites ci-dessus d'entrainer son propre modèle CNN. Cependant plusieurs problèmes se posent :

- Une base de données importante doit être utilisée pour entrainer ces réseaux de neurones.
- Il faut une base de données par ingrédient (ou classe) que le modèle doit apprendre.
- Notre problématique comprend un nombre de classes égal au nombre d'ingrédients que notre modèle est censé connaitre, soit un nombre très important.

Des débuts de modèles sont actuellement développés sur la base de la reconnaissance alimentaire. Cependant, rien n'est encore abouti pour le moment. Développer un modèle « from scratch » serait un projet à part entière vu la complexité des éléments à reconnaitre. Nous avons donc décidé d'utiliser des modèles pré-entrainés disponibles sur keras, une librairie de deep learning sur Python fonctionnant avec Tensorflow. Ces modèles ont été développé pour la compétition imagenet1000 challenge, où les modèles sont testés sur la reconnaissance de 1000 classes différentes. Il est possible de récupérer les poids des modèles entrainés et de les utiliser pour ses propres besoins.

Dans notre cas, nous avons choisi et utilisé le modèle VGG16. Une explication de son implémentation dans notre projet est disponible sous forme d'un fichier ¹ jupyter dédié, joint à ce mémoire. Nous avons dû ajouter un traducteur en sortie du réseau de neurones pour traduire la prédiction en anglais du label, en français. En effet, notre base de données Neo4j du site Marmiton est en français et toute requête avec des termes anglais aboutit à des résultats nuls. Une étape est manquante : celle de la localisation des différents ingrédients dans la photo de l'intérieur du réfrigérateur. Apres recherches, il existe un modèle performant de localisation d'éléments dans une image, le modèle YOLO (You Only Look Once). Cependant ce modèle, ne fonctionne que sur de la vidéo. A nouveau, des modèles sont en cours de développement mais aucune solution n'est utilisable pour le moment.

Face à ces limites, nous avons donc décidé de limiter l'implémentation du deep learning à la reconnaissance d'un unique ingrédient présent sur une seule image, et non sur une image de réfrigérateur composée de multiples ingrédients.

Pour résumer, l'utilisateur peut uploader sur le site une image ayant un unique ingrédient puis l'application est capable de reconnaitre l'ingrédient présent sur la photo et enfin propose à l'utilisateur les recettes possibles avec cet ingrédient.

 $^{1. \ \} Deep_Learning_Jupyter.ipynb$

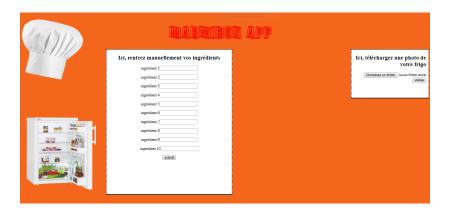
2.4. RENDU FINAL

2.4 Rendu Final

2.4.1 Utilisation sans Machine Learning

Maintenant que l'application est prête, nous obtenons les résultats suivants lorsque nous la lançons.

Pour la page d'accueil :

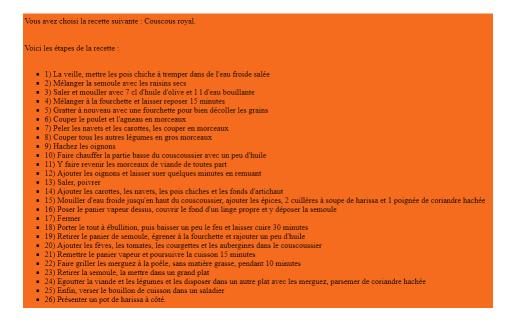


Par exemple, on rentre les ingrédients sel, poivre, tomate, aubergine, ail, courgette, poulet et on obtient pour la page de résultats :



2.4. RENDU FINAL

Puis, on sélectionne la recette que l'on souhaite suivre. Par exemple, on choisit "Couscous royal" et la page index nous renvoie les étapes de préparation associées à cette recette :



2.4.2 Utilisation avec Machine Learning

L'utilisateur peut charger une image depuis son bureau grâce à la fenêtre prévue à cet effet à droite de la page d'accueil.

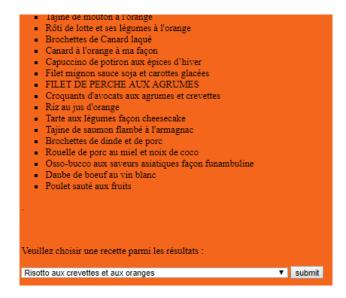


Ici, nous décidons d'uploader une image d'orange. Après avoir télécharger l'image, la machine détecte l'ingrédient présent et la labélise. Dans notre exemple, le programme détecte bien une orange, et nous propose ainsi toutes les recettes utilisant des oranges.

2.4. RENDU FINAL



De la même manière que dans la partie précédente, nous pouvons selectionner la recette d'intérêt dans le menu déroulant.



Et enfin, l'application nous montre les étapes à suivre pour la recette sélectionnée.

