GANDELIN Antoine

Un Zéro Un



Compte rendu n°2

Période du 24 janvier au 04 février 2022

Stage suivi par Monsieur **GIARELLI** pour l'organisme et par Monsieur **PASQUALINI** pour le lycée Roosevelt

Table des matières

Objectifs de la période	3
Déroulement de la période	4
Compétences abordées	
Bilan de la période	7
Annexes	8
Vue globale de mon programme pour la ville de Clermont-Ferrand (sous	
PHPStorm)	8
Extrait du fichier JSON de la ville de Clermont-Ferrand (sous PHPStorm)	9
Extrait de mon programme pour la ville de Clermont-Ferrand (Utilisations de	
fonctions PHP + Expressions régulières sous PHPStorm)	.10
Extrait du fichier CSV de la ville de Clermont-Ferrand (sous LibreOffice Calc).	.10
Vue d'un « var_dump() » de la variable « \$matches » (sur le terminal)	.11
Liens externes	
Extrait d'une expression régulière (vue sur le site « regex101 »)	.12

Objectifs de la période

Vous indiquez ici les attentes de votre tuteur de stage concernant les objectifs à atteindre sur cette période, les missions confiées. Le cas échéant, vous pouvez découper ce chapitre en plusieurs parties correspondant aux différentes missions.

Les objectifs que j'ai eus à atteindre sur cette période sont :

- · Poursuivre la mission qui m'a été confiée
- Enrichir et modifier mon programme
- Sélectionner les données à partir d'une page web (web scraping)
- Respecter la conformité du format CSV qui m'a été donnée
- Trouver des solutions/méthodes afin de résoudre mes difficultés
- Mettre en place une méthode pour scinder une chaîne de caractères en segments et la manipuler correctement
- Convertir correctement le fichier CSV
- Produire le programme attendu

Déroulement de la période

Vous indiquez ici ce que vous avez fait pour répondre aux besoins exprimés, les difficultés que vous avez rencontrées, les solutions que vous avez proposées ... Le cas échéant, vous pouvez découper ce chapitre en plusieurs parties correspondant aux différentes missions.

Lors de ma troisième semaine de stage, j'ai poursuivi la mission que l'on m'avait confié au début. En effet, j'ai amélioré mon programme et je l'ai adapté afin que le CSV puisse correspondre à chacun des fichiers. Cependant, mon maître de stage m'a fait constater que je n'avais pas respecté le format de CSV qui était attendu. En effet, j'étais confronté à deux problèmes : soit il manquait des colonnes soit il fallait que je modifie ces colonnes afin de séparer les données.

Il m'a expliqué que c'était indispensable de corriger cela car sinon, mon fichier CSV ne servirait à rien et il ne pourrait pas être utilisé par l'application que l'agence souhaite mettre en œuvre. Il fallait donc que je respecte impérativement le format des colonnes, qui sont les suivantes : « Insee / Code postal / Commune / Nombre habitants / Rue / Ordures / Horaires ordures / Récurrence / Tri / Horaires tri / Récurrence ».

Avant de poursuivre ma mission, une visite de stage avec mon professeur référent, mon maître de stage et moi même a eu lieu le mercredi 26 janvier en visioconférence. Cela a permis à mon maître de stage d'échanger avec mon professeur concernant l'avancée de mon travail ainsi que mon comportement et mon investissement au sein de l'organisme d'accueil. Lors de cet échange, mon professeur m'a conseillé de choisir, comme sujet de classe inversée, le format de données « JSON ».

Pour corriger mon traitement de données de la ville de Clermont-Ferrand, j'ai dû récupérer les données à partir de la page web du site. Pour ce faire, j'ai réalisé du « web scraping » : technique d'extraction du contenu de sites web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte. J'ai recueilli toutes les informations concernant la collecte des déchets de cette ville. Une fois que j'étais en possession de toutes les données, j'ai pu commencer à programmer.

J'ai repris mon programme initial mais en restructurant le CSV afin qu'il puisse correspondre au format souhaité. Ensuite, j'étais confronté à un nouveau problème. En effet, afin que mon CSV soit au format stipulé, j'étais obligé de séparer certaines données. Par exemple, dans le fichier JSON de Clermont-Ferrand, le jour et l'horaire de collecte des déchets étaient renseignés dans la même colonne. Étant donné que le format (évoqué plus haut) indiquait qu'il fallait une colonne pour le jour et une autre pour les horaires, je n'avais pas d'autre choix que de dissocier ces données en deux.

Pour résoudre ce problème, j'ai utilisé, dans mon programme, la fonction « explode » qui permet de scinder une chaîne de caractères en segments (cf. Documentation PHP). Cette fonction m'a permis de séparer les jours et les horaires concernant le tri sélectif. Cependant, ce n'était pas suffisant pour séparer ceux des ordures ménagères. En effet, étant donné qu'il y avait plusieurs jours (et parfois plusieurs horaires) de collecte des déchets pour les ordures ménagères de Clermont-Ferrand, il a fallu que je trouve une autre solution pour résoudre ce problème. Ainsi, j'ai utilisé les expressions régulières (ou

expressions rationnelles). Il s'agit de chaînes de caractères, qui décrivent, selon une syntaxe précise, un ensemble de chaînes de caractères possibles.

L'expression régulière est un outil très puissant concernant la manipulation des chaînes de caractères. Cela convient donc parfaitement pour résoudre mon problème. Avant de l'utiliser, je me suis aidé du site « regex101 » qui permet de tester des expressions régulières avec une aide sous forme de coloration syntaxique. Ces recherches m'ont permis de progresser concernant la manipulation des expressions régulières. En effet, avant de les utiliser pour mon cas concret, je me suis déjà entraîné en réalisant plusieurs tests à partir du site.

Avec les expressions régulières, j'ai utilisé la fonction PHP « preg_match_all() » qui est une expression rationnelle globale (cf. Documentation PHP). Elle renvoie le nombre de correspondances (« \$matches ») d'un modèle trouvé dans une chaîne et remplit une variable avec les correspondances trouvées. Au niveau des paramètres, il y avait trois paramètres à indiquer : le « pattern », le « subject » et les « matches ». Au niveau du pattern (du masque à chercher, sous la forme de chaîne de caractères), j'ai mis l'expression régulière qui devait correspondre avec ce que je souhaitais séparer (c'est à dire, ici, les jours et les horaires).

Ensuite, pour le subject (la chaîne d'entrée), j'ai indiqué l'emplacement des données situé dans le fichier JSON (ici : « \$dataset[fields][ordures_menagers_jours] »). Enfin, je n'avais plus qu'à écrire « \$matches » à la suite afin d'avoir le tableau contenant tous les résultats.

Avant de réaliser la conversion, j'ai vérifié, en réalisant un « var_dump(\$matches) », pour voir si tout était correct. Une fois le programme exécuté, j'apercevais les données/résultats que je souhaitais obtenir (jours et horaires). Il n'y avait plus qu'à mettre ces données dans deux colonnes distinctes. Pour cela, j'ai écrit sur une ligne le premier tableau (« \$matches[1] ») afin d'avoir les jours de collecte des déchets de Clermont-Ferrand puis le second tableau (« \$matches[2] ») afin d'avoir les horaires de collecte des déchets.

Lorsqu'il y avait plusieurs jours ou horaires sur la même ligne, j'ai utilisé la fonction « implode() » qui permet de concaténer les éléments d'un tableau en une chaîne. J'ai indiqué (en paramètres) comme « separator », un tiret : « - ».

Une fois le programme terminé, la conversion pouvait donc être réalisée. Le programme s'exécutant parfaitement et le CSV étant au bon format cette fois-ci, je n'avais plus qu'à déposer mon dossier avec les fichiers correspondants (le fichier JSON, le programme et le fichier CSV de Clermont-Ferrand) sur le Google Drive afin que mon maître de stage puisse observer l'avancée de mon travail.

Compétences abordées

Vous indiquez ici les compétences que vous avez mises en œuvre au cours de cette période en reliant les compétences mises en œuvre avec les missions suivies. Vous pouvez vous aider du tableau des compétences fourni dans un fichier séparé. L'intérêt est de pouvoir remplir aisément son portfolio par la suite.

Les compétences que j'ai abordées durant cette période sont :

- B1-4: Travailler en mode projet
- B1-6 : Organiser son développement professionnel
- B2D-2 : Assurer la maintenance corrective ou évolutive d'une solution applicative
- B2D-3 : Gérer les données

Bilan de la période

Vous faites un rapide retour sur les objectifs a atteindre et la réalité de la réalisation. C'est l'occasion de prendre du recul sur votre démarche et de proposer des remédiations éventuelles.

Durant cette deuxième période, j'ai été confronté à quelques difficultés. J'ai été quelque peu déstabilisé car le programme que j'avais créé ne correspondait pas tout à fait aux attentes de mon maître de stage. Mais, grâce à nos échanges réguliers et à mes recherches personnelles, je suis parvenu à surmonter mes difficultés et à produire le programme attendu.

Ainsi, j'ai poursuivi ma mission en améliorant et en adaptant mon programme en fonction des fichiers que j'avais à traiter. Pour ce faire, j'ai dû modifier mon programme en respectant impérativement un format de fichier CSV donné afin que celui-ci puisse être intégré dans l'application que l'agence souhaite créer.

En effet, afin de respecter le format de fichier CSV donné, je devais séparer les données. En me renseignant auprès de mon maître de stage et sur Internet (en regardant la documentation officielle de PHP), je suis parvenu à trouver les fonctions ainsi que les expressions régulières à utiliser pour résoudre mon problème. Ainsi, le fichier CSV de la ville de Clermont-Ferrand était conforme aux attentes et respectait le format donné.

J'ai également appris à réaliser du web scraping, c'est à dire, recueillir les données à partir d'une page web. J'ai utilisé cette technique afin de récupérer les données de collecte des déchets de la ville de Clermont-Ferrand.

L'objectif de ma troisième période (cinquième et dernière semaine de stage) sera de réaliser, pour d'autres villes, un fichier CSV conforme (en réutilisant, si besoin, les connaissances acquises au cours de cette deuxième période) ainsi que d'importer les données que j'ai traitées, dans l'application.

Pour conclure, cette deuxième période fut très enrichissante pour moi. J'ai pu approfondir une nouvelle fois mes connaissances ainsi que mes compétences en informatique et développement. J'ai compris l'utilité de certaines fonctions et l'utilisation des expressions régulières afin de manipuler correctement les chaînes de caractères. Cette deuxième période m'a permis de progresser dans la manipulation des données dans un fichier CSV.

Annexes

Ajoutez ici tous les documents : morceaux de code, manuels, scripts, images, diagrammes ; qui vous semblent pertinents pour étayer votre démarche.

Vue globale de mon programme pour la ville de Clermont-Ferrand (sous PHPStorm)

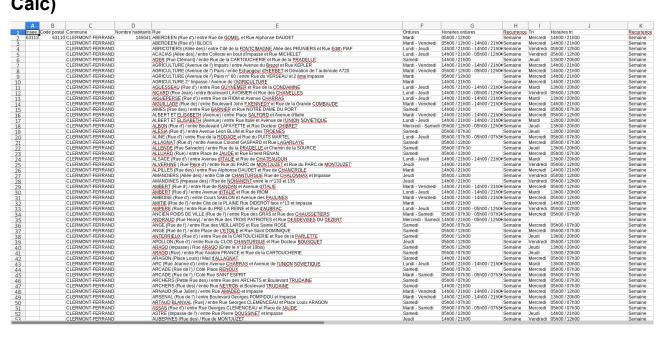
```
<?php
$content = file_get_contents($file);
foreach ($datasets as $dataset) {
    preg_match_all(pattern: '#(\w+) : (\w+ / \w+)#', $dataset['ordures_menagers_jours'], &: $matches);
        implode( separator: ' - ', array: $matches[1] ?? []),
        implode( separator: ' - ', array: $matches[2] ?? []),
```

Extrait du fichier JSON de la ville de Clermont-Ferrand (sous PHPStorm)

```
"emcombrants_jours": "",
"gravats_observations": "",
```

Extrait de mon programme pour la ville de Clermont-Ferrand (Utilisations de fonctions PHP + Expressions régulières sous PHPStorm)

Extrait du fichier CSV de la ville de Clermont-Ferrand (sous LibreOffice Calc)



Vue d'un « var_dump() » de la variable « \$matches » (sur le terminal)

```
array(3) {
  [0]=>
  array(2) {
   [0]=>
   string(21) "Lundi : 05h00 / 07h30"
   string(21) "Jeudi : 05h00 / 07h30"
  [1]=>
  array(2) {
   [0]=>
   string(5) "Lundi"
   [1]=>
   string(5) "Jeudi"
  [2]=>
  array(2) {
   [0]=>
   string(13) "05h00 / 07h30"
   [1]=>
    string(13) "05h00 / 07h30"
```

Liens externes

https://www.php.net/manual/fr/function.implode.php

https://www.php.net/manual/fr/function.explode.php

https://www.php.net/manual/en/function.preg-match-all.php

https://www.php.net/manual/fr/function.substr.php

https://regex101.com/

Extrait d'une expression régulière (vue sur le site « regex101 »)

```
*/ (\w+) •: • (\S+•\/•\S+)

TEST STRING

Mardi •: • 05h00 • / • 12h00 
Lundi •: • 05h00 • / • 12h00
```