



RAPPORT DE STAGE

Développeur Java

4 Janvier 2021 au 26 Février 2021

Tuteur de stage : Monsieur Eric LANCELEUR

Enseignant référent : Madame Béatrice CARAMIGEAS

Etablissement : Lycée Paul Louis Courier

Entreprise d'accueil : EXAEGIS - 6 boulevard Alfred Nobel 37540 SAINT CYR SUR LOIRE

Table des matières

Présentation de l'entreprise.....	4
Présentation du service.....	5
Présentation de l'environnement matériel et logiciel.....	5
Présentation du contexte de la situation professionnelle.....	6
Présentation des missions.....	6
Présentation des tâches.....	6
Liste des besoins en formation.....	7
Semaine 1.....	8
Listes des tâches prévues au cours de la semaine.....	8
Les trois premiers jours.....	8
Les deux derniers jours.....	8
Liste des tâches réalisées.....	9
Les trois premiers jours.....	9
Les deux derniers jours.....	9
Analyse des écarts entre le prévu et le réalisé.....	10
Liste des tâches prévues pour la semaine suivante.....	10
Semaine 2.....	11
Listes des tâches prévues au cours de la semaine.....	11
Liste des tâches réalisées.....	11
Premier jour.....	11
Deuxième et troisième jour.....	11
Derniers jours.....	11
Analyse des écarts.....	12
Liste des tâches prévues pour la semaine prochaine.....	12
Semaine 3.....	13
Listes des tâches prévues au cours de la semaine.....	13
Liste des tâches réalisées.....	13
Premier jour.....	13
Deuxième jour.....	13
Troisième et Quatrième jours.....	13
Dernier jour.....	13
Analyse des écarts.....	14
Liste des tâches prévues pour la semaine prochaine.....	14
Semaine 4.....	15
Listes des tâches prévues au cours de la semaine.....	15
Liste des tâches réalisées.....	15
Premier et deuxième jour.....	15
Troisième jour.....	15
Quatrième jour.....	15
Dernier jour.....	16
Analyse des écarts.....	16
Liste des tâches prévues pour la semaine prochaine.....	16
Semaine 5.....	17
Listes des tâches prévues au cours de la semaine.....	17
Liste des tâches réalisées.....	17
Premier et deuxième jour.....	17
Derniers jours.....	17
Analyse des écarts.....	18
Liste des tâches prévues pour la semaine prochaine.....	18
Semaine 6.....	19
Listes des tâches prévues au cours de la semaine.....	19

<u>Liste des tâches réalisées.....</u>	<u>19</u>
<u>Premier et deuxième jour.....</u>	<u>19</u>
<u>Troisième jour.....</u>	<u>19</u>
<u>Derniers jours.....</u>	<u>19</u>
<u>Analyse des écarts.....</u>	<u>20</u>
<u>Liste des tâches prévues pour la semaine prochaine.....</u>	<u>20</u>
<u>Semaine 7.....</u>	<u>21</u>
<u>Listes des tâches prévues au cours de la semaine.....</u>	<u>21</u>
<u>Liste des tâches réalisées.....</u>	<u>21</u>
<u>Premier jour.....</u>	<u>21</u>
<u>Derniers jours.....</u>	<u>21</u>
<u>Analyse des écarts.....</u>	<u>21</u>
<u>Liste des tâches prévues pour la fin de semaine.....</u>	<u>21</u>
<u>Semaine 8.....</u>	<u>22</u>
<u>Listes des tâches prévues au cours de la semaine.....</u>	<u>22</u>
<u>Liste des tâches réalisées.....</u>	<u>22</u>
<u>Jusqu'à jeudi.....</u>	<u>22</u>
<u>Vendredi.....</u>	<u>22</u>
<u>Ce que m'a apporté le stage.....</u>	<u>23</u>
<u>Conclusion.....</u>	<u>24</u>
<u>Annexes.....</u>	<u>25</u>

Pr esentation de l'entreprise

ex egis : la confiance pour le financement et l'investissement du num erique Bas e sur une quadruple comp etence, vision strat egique des march es, expertises op erationnelles, continuit e d'exploitation et financi ere, ex egis d eveloppe une Offre compl ete pour faciliter le financement et l'investissement des acteurs du march e du num erique et de leurs investisseurs et bailleurs :

1. S ecurisation du financement : audit, surveillance, s ecurisation, conseil « risk policy » .
Principalement   destination des bailleurs du num erique, ex egis pr esente une large gamme de services visant   encadrer les risques de performance en amont et en aval des contrats de financement. ex egis intervient pour s ecuriser les offres financ ees « as a Service » (Software, Infrastructure et Platform), les redevances d'infog erance d'infrastructure, les contrats de location int egrant des prestations de services dont l'ex ecution d epend d'un prestataire.
2. Audit et Labellisation : audit,  valuation, surveillance Pour toutes les entreprises du Digital, ex egis a mis en place un syst eme de notation et d'audit de nature   faciliter la recherche d'investisseurs ou de partenaires financiers. Selon la notation obtenue, le label TRUXT r ecompense les entreprises offrant les meilleures performances de gestion op erationnelles et financi eres.
3. Etudes : sectorielles, benchmarks, cartographie Pr ecieuses aides   la d ecision tactique et strat egique, les  tudes s'appuient sur des m ethodologies  prouv ees et transparentes. Elles d elivrent des r esultats objectifs et pr ecis selon des segmentations reconnues. Leur spectre est large (technologies, processus et m etiers, offre et demande, sourcing et financement), pour mieux mettre en lumi ere les orientations des march es.
4. Conseil et M&A : d efinition de trajectoire, recherche de cibles, ex ecution M&A. Gr ace   une connaissance approfondie du secteur du num erique, ex egis accompagne les entreprises dans leur strat egie de croissance externe ou de d esinvestissement.

L'entreprise se situe   Saint-Cyr-Sur-Loire (Si ege Social). Elle est dirig ee par M. Laurent Briziou. Il y a entre 15 salari es dans l'entreprise.

Pr esentation du service

Le service informatique est charg e de programmer les applications de l'entreprise Exeagis.

Actuellement, le service informatique est charg e de programmer une application permettant de r ecup erer les d eliberations publiques des collectivit es (r egions, d epartements, m etropoles et villes) sur les sites web des collectivit es. Une fois r ecup er es, les documents sont index es dans une sorte de moteur de recherche cr ee par Exaegis. Dans le service informatique il y a 4 employ es et un apprenti. Mon ma tre de stage a la fonction de Responsable informatique.

Pr esentation de l'environnement mat eriel et logiciel

Pour travailler j'ai acc es  a une machine distante via le logiciel « Connexion Bureau Distant » de Microsoft afin de faciliter le t el etravail.

Je peux m'y connecter via mon PC  a la maison et je m'y connecte aussi via le PC du bureau quand je vais sur le lieu de l'entreprise. Ainsi c'est comme si j'avais toujours le m eme ordinateur peu importe o u je suis.

L'environnement qu'on utilise dans l'entreprise est NetBeans avec JDK 8.

Nous utilisons la librairie Jsoup et/ou Selenium pour connecter l'application aux sites web sur lesquels nous r ecup erons les documents  a indexer.

Jsoup ne fait que des requ etes http aux URLs qu'on lui passe.

Selenium simule un navigateur (Firefox) pour se comporter comme un utilisateur physique se connectant au site.

Tout est cod e en Java.

Présentation du contexte de la situation professionnelle.

Exaegis indexe les documents des délibérations des collectivités pour que les entreprises puissent se préparer à l'avance aux appels d'offres des collectivités. Cela permet aux entreprises de répondre à l'appel d'offre le plus vite possible en ayant toutes les cartes en main pour obtenir la charge du projet.

Présentation des missions

Chaque site de chaque collectivité est différent. Nous devons coder un scraper (Le **scraping** définit de façon générale une technique permettant d'extraire du contenu d'un ou de plusieurs sites web de manière totalement automatique.) pour chaque collectivité afin de s'adapter à chacun des sites. Nous nous occupons actuellement des départements car les autres collectivités ont pratiquement toutes été faites.

Présentation des tâches

Tout d'abord nous devons choisir une collectivité sur laquelle travailler.

Ceci étant fait, il faut étudier son site web pour savoir comment accéder aux documents :

- Est-ce qu'il faut remplir un formulaire de recherche ? (voir annexe 1)
- Est-ce qu'il y a des cookies et des headers particuliers quand on accède à la page des documents ? (voir annexe 2)
- Est-ce que l'affichage des documents est paginé ? (Si oui, peut-on faire en sorte d'afficher assez de documents pour empêcher la pagination ?) (voir annexe 3)
- Y a-t-il des pop up (style « accepter les cookies ») à valider ?

Une fois qu'on a étudié le site, il faut coder le scraper en l'adaptant au site.

Liste des besoins en formation

J'ai besoin de conna tre les fonctionnalit s de Jsoup et de Selenium pour optimiser le code au mieux afin d'acc der aux documents de la mani re la moins lourde possible pour ne pas surcharger le serveur. Il faut aussi que je comprenne les erreurs pour adapter au mieux le code (par exemple, si le script va trop vite pour le site, il faut le comprendre pour savoir qu'il faut mettre une temporisation pour laisser le temps au site de charger les pages).

Pour nous (les stagiaires) former   Selenium et Jsoup, c'est Jean-Marc (chef de projet) qui nous a expliqu  les fonctions dont on aurait le plus besoin en nous montrant des exemples de script. Ensuite on a eu une journ e pour  tudier les scripts d j  existants et essayer de corriger ceux qui ont  t  signal s pour mieux assimiler ces fonctions. Ensuite, quand on a besoin d'aide nous utilisons la documentation ou nous posons des questions sur le groupe de l' quipe (sur Teams).

Semaine 1

Listes des tâches prévues au cours de la semaine

Les trois premiers jours

- Mise en place de l'environnement de développement (NetBeans)
- Mise en place du projet
- Mise en place des outils collaboratifs :
 - Office 365 comporte tous les outils de partage :
 - Teams pour les réunions à distance
 - Outlook pour les mails
 - OneDrive pour le partage de fichiers
 - Excel pour le partage du tableau des collectivités à faire.
 - Gitlab pour la gestion de version et partage du code
- Étude du projet

Les deux derniers jours

- Commencement d'un script pour une collectivité

Liste des t aches r alis ees

Les trois premiers jours

- La machine distante  tait d ej a mise en place, il ne restait plus qu' a installer Netbeans avec le bon JDK
- Le projet a  t e clon e depuis le repository GitLab, puis build e sur Netbeans afin que toutes les librairies soient install ees automatiquement gr ace  a la configuration pr e- crite par l' equipe de d eveloppement
- Les outils collaboratifs  taient d ej a mis en place, l' equipe nous a juste expliqu e comment tout  a fonctionnait
- Apr es quelques explications sur le fonctionnement du projet nous avons lu les scripts d ej a  crits pour mieux les comprendre, et essay e de corriger les scripts signal es comme ayant un probl eme (car c' est en pratiquant qu'on apprend le mieux).
 - Probl eme sur lequel j'ai d ecid e de travailler : Le script pour la r egion Corse pr esentait   priori un probl eme de lecture du DOM. On pensait que c' etait le site qui avait chang e.
 - Les solutions envisag ees : La seule solution  tait donc de tester le code pour voir   quel moment le script plantait afin de savoir quel  l ement du site g enait.
 - La d emarche suivie : J'ai donc mis quelques « `System.out.println();` » un peu partout dans le script pour d eboguer le code (on ne peut pas utiliser l'outil de d ebogage de Netbeans puisque le script communique avec des sites web et si on met le programme en pause, la connexion plante parce que le code met trop de temps   r epondre). Une fois fait, j'ai lanc e le script pour voir si toutes mes lignes de d ebogage affichaient quelque chose.
En fait, non seulement elles ont toutes affich e quelques chose, mais en plus le programme fonctionnait compl etement sans aucune erreur. J'en ai donc parl e   l' equipe et ils en ont conclu que mes lignes de d ebogage cr eaient une micro latence du script qui fait que le site r epond mieux.

Les deux derniers jours

- J'ai choisi de commencer par le d epartement de l'Aisne car le site  tait plus simple. Je me suis dit que c' est mieux de commencer par un site simple pour le d ebut. Le script n' est pas encore fini.
 - Probl eme rencontr e : Il y avait un formulaire de recherche sur la page des d eliberations. J'ai remarqu e que dans les informations envoy ees par le formulaire il y avait les positions « x » et « y » du clique sur le bouton « Rechercher » (voir annexes 4 et 5). L'image du bouton fait 199 pixels sur 75 pixels. Quand on clique dessus il y a un programme JavaScript qui envoie la position du clique sur l'image.
 - Les solutions envisag ees : J'ai donc pens e que c' etait une sorte captcha. D'ailleurs j'ai pu trouver le code JavaScript de ce captcha dans le code source de la page. Il n' tait pas obfusqu e (le code  tait  crit en clair) je pense donc que c' est le d eveloppeur du site qui

Brunschwig N emo

l'a cod e. Si je le pr ecise, c'est parce que  a me fait penser que le captcha est de mauvaise qualit e. Je me suis alors dit qu'il suffisait d'envoyer des valeurs al eatoires pour lui faire croire que le script avait bel et bien cliqu e sur le bouton.

- o La d emarche suivie : J'ai donc cherch e sur internet comment cr eer des valeurs « pseudo » al eatoire (rien est vraiment al eatoire en informatique). J'ai cliqu e dans le coin inf erieur droit du bouton « Recherche » puis regard e quelles  taient les valeurs envoy ees afin de d eterminer un plafond   ne pas d epasser pour mes valeurs al eatoires (« x » ne peut pas avoir une valeur plus grande que la largeur du bouton et « y » ne peut pas avoir une valeur plus grande que la hauteur du bouton). Une fois fait j'ai test e, et  a fonctionne. (voir annexe 6)

Analyse des  cartes entre le pr evu et le r ealis e

Tout ce qui  tait pr evu a  t e fait.

Liste des t aches pr evues pour la semaine suivante

- Finir le script pour le d epartement sur lequel je travaille
- Une fois fini, passer   un autre d epartement

Semaine 2

Listes des t aches pr evues au cours de la semaine

- Finir le script du d epartement sur lequel je travaillais (le d epartement de l'Aisne)
- Commencer le script d'un nouveau d epartement

Liste des t aches r ealis ees

Premier jour

- Le script pour le d epartement de l'Aisne est termin e.

Deuxi eme et troisi eme jour

- Tentative d'utilisation d'un OCR (Optical Character Recognition = Reconnaissance Optique de Caract eres) pour « lire » les textes sur des images (sans r esultats)
 - Probl eme rencontr e : Des documents PDF ne contenaient pas de texte. Parfois c' etait des scannes des d eliberations. Exaegis a besoin de r ecup erer les textes dans les PDF pour son application mais quand c'est des images c'est compliqu e d'en tirer le texte.
 - Les solutions envisag ees : Utiliser un OCR pour reconna tre les textes dans les images.
 - La d emarche suivie : J'ai donc propos e d'essayer moi m eme d'impl ementer l'OCR. J'ai cherch e quel  et e l'OCR le plus utilis e en Java et je suis tomb e sur Tesseract. J'ai essay e de l'utiliser en suivant des exemples mais c' etait beaucoup trop compliqu e pour moi. Eric (ma tre de stage) m'a alors dit d'abandonner et de reprendre les scripts des d epartements.

Derniers jours

- Commencement d'un script pour le d epartement des Hautes-Alpes
 - Probl eme rencontr e : A premi ere vue, le site semblait  tre exactement comme le site du d epartement de l'Aisne. Mais plus je regardais et plus il y avait des diff erences . L'affichage des documents  tait contenu dans un « frame » et la pr esentation change   partir d'une certaine date (probablement que ce n'est plus la m eme personne qui g ere les documents).
 - Les solutions envisag ees : Adapter le script du mieux qu'on peut aux contraintes.
 - La d emarche suivie : J'ai d u trouver la bonne URL dans le « frame » pour faire mes requ etes http directement dessus. J'ai d u trouver les dates des documents dans les URL des documents car elles n' etaient plus dans les descriptions.

Analyse des  carts

J'ai fait plus que ce qui  tait pr vu (  savoir l'essai sur l'OCR) ce qui m'a retard  sur le nouveau script...

Liste des t ches pr vues pour la semaine prochaine

- Finir le script pour le d partement des Hautes-Alpes
- D marrer un nouveau script

Semaine 3

Listes des tâches prévues au cours de la semaine

- Finir le script pour le département des Hautes-Alpes
- Démarrer un nouveau script

Liste des tâches réalisées

Premier jour

- J'ai fini le script pour les Hautes Alpes.

Deuxième jour

- J'allais commencer un nouveau script mais on m'a renvoyé le script des Hautes Alpes car j'avais oublié une partie. Du coup j'ai codé la partie du script manquante en une journée.
 - Problème rencontré : J'avais oublié de faire le script pour récupérer les budgets. Sur la page « d'accueil » des budgets il y avait 2 catégories : Les budgets avant 2010 ; les budgets après 2010.
 - Les solutions envisagées : Je me suis dit qu'il faudra faire deux scripts pour récupérer tous les budgets (un script par catégorie).
 - La démarche suivie : Après avoir réfléchi quelques instants je me suis rendu compte qu'on n'avait pas besoin des délibération datant d'avant 2010, j'ai alors fait qu'un seul script.

Troisième et Quatrième jours

- J'ai commencé et terminé le script pour le département du Calvados.

Dernier jour

- J'ai commencé le script pour le département de la Corrèze.
 - Problème rencontré : L'interface du site s'avère être plus complexe que prévu. D'habitude on pioche les éléments dont on a besoin en les sélectionnant par leur class, leur id ou leur tag (dans le code source HTML de la page web). Cette fois la sélection la plus pertinente que je puisse faire pour sélectionner tous les documents est une sélection sur la class. Sauf que les class des balises contenant les documents ne leur sont pas propres et sont réutilisés un peu partout dans la page du site. Du coup, quand je sélectionne avec leur class, je sélectionne également tout un tas d'autres éléments dont je n'ai pas besoin.

Brunschwig N mo

- Les solutions envisag es : je n'ai pas encore trouv  de solution.

Analyse des  carts

J'ai r ussi   terminer 2 scripts en une semaine et   en commencer un troisi me. J'ai fait plus que ce qui  tait pr vu pour la semaine.

Liste des t ches pr vues pour la semaine prochaine

- Finir le script pour le d partement de la Corr ze.
- Commencer un nouveau script.

Semaine 4

Listes des t aches pr evues au cours de la semaine

- Finir le script pour le d epartement de la Corr eze.
- Commencer un nouveau script.

Liste des t aches r ealis ees

Premier et deuxi eme jour

- Script pour le d epartement de la Corr eze fini.
 - La solution envisag ee (par rapport au probl eme de la semaine pass ee) : Filtrer les  el ements qu'on r ecup ere.
 - La d emarche suivie : J'ai remarqu e que les  el ements qui m'int eressaient avaient une combinaison de class particuli ere par rapport aux autres  el ements. J'ai donc r ecup er e les class de chaque  el ement, les ai analys ees, et ainsi j'ai pu filtr e les  el ements pour ne garder que ce qui m'int eressait.

Troisi eme jour

- Commencement du script pour la C ote D'Or.
 - Probl eme rencontr e : le certificat SSL du site  etait trop r ecent, et n' etait pas reconnu par ma JVM.
 - Les solutions envisag ees : Jean-Marc (chef de projet) a trouv e une solution et a propos e d'attendre qu'on se voit le lendemain dans l'entreprise pour nous montrer car Sarah avait le m eme probl eme.
- Commencement d'un nouveau script pour les C otes D'Armor (en attendant d'avoir r egl e le probl eme pour le script pr ec edant).

Quatri eme jour

- Jean-Marc nous montre la solution pour le probl eme de la veille.
 - La d emarche suivie : on t el echarge le certificat du site gr ace  a Mozilla FireFox (voir annexe 7). On tape la ligne de commande suivante dans le cmd « `.\.\bin\keytool -import -trustcacerts -keystore cacerts -storepass changeit -noprompt -alias [le nom du d epartement] -file [chemin vers le certificat du site]` »
- J'ai continu e le script pour le d epartement des C otes D'Armor.

Dernier jour

- J'ai continué le script pour le département des Côtes D'Armor.
 - Problèmes rencontrés : Le site n'est pas très bien fait, c'est difficile de naviguer dedans via le code html.
Le site est hébergé sur un vieux serveur (voir annexe 8), quand on fait quelque chose d'un peu trop lourd l'onglet plante.
 - Solutions envisagées : Filtrer les éléments comme pour le script pour le département de la Corrèze.
Faire des temporisations pour éviter de surcharger le serveur.

Analyse des écarts

J'ai fait ce qui été prévu pour la semaine. J'aurais aimé finir un script de plus mais les sites que j'ai décidé de faire sont plus compliqués que ce que je pensais.

Liste des tâches prévues pour la semaine prochaine

- Finir le script pour les Côtes D'Armor (car il n'est toujours pas fini).
- Finir le script pour la Côte D'Or.
- Commencer un nouveau script.

Semaine 5

Listes des tâches prévues au cours de la semaine

- Finir le script pour les Côtes D'Armor (car il n'est toujours pas fini).
- Finir le script pour la Côte D'Or.
- Commencer un nouveau script.

Liste des tâches réalisées

Premier et deuxième jour

- Finissions du script pour les Côtes D'Armor.

Derniers jours

- Reprise du script pour la Côte D'Or.
 - Problèmes rencontrés :
 - On ne peut pas afficher les délibérations par date, on est obligé de tout afficher (avec pagination, il y a un maximum de 50 délibérations à l'écran). De plus, les délibérations s'affichent dans l'ordre chronologique (les plus anciens en premiers et les plus récent tout à la fin).
 - Il y a un fichier qui est appelé par la page et qui renvoie toutes les informations des délibérations (titre, lien de téléchargement, etc...) au format JSON. Problème, je ne peux pas accéder à ce fichier directement car il fonctionne avec la session et donc quand on y fait appel directement il renvoie un JSON vide (même en lui donnant des données en méthode POST, tous les headers qu'il faut et les cookies de session).
 - Solutions envisagées :
 - Parcourir les documents en sens inverse. Il suffit d'aller à la dernière page et de commencer par la fin de la page. Dans la pagination on peut marquer à quel page on peut se rendre (et combien de pages il y a en tout). Avec Selenium, je récupère le nombre de pages en tout, puis je l'inscris dans l'endroit où on peut indiquer la page à laquelle on veut se rendre. Grâce à une boucle for, je décréméte le chiffre que j'ai récupéré puis inscris le résultat pour me rendre à la page précédente.
 - Soit ouvrir 2 fenêtres simultanément, soit essayé de ré-injecter les cookies de session, soit abandonner le JSON et récupérer ce qu'il faut directement dans la page web.
 - La démarche suivie :
 - La boucle est fait, on peut parcourir dans le sens inverse.

Brunschwig Némó

- On ne peut pas ouvrir 2 fenêtres simultanément car au lieu d'afficher le contenu il télécharge le fichier (avec un JSON vide). Ré-injecter les cookies reviendrait au même, on ne peut tout simplement pas traiter ce fichier.

Analyse des écarts

Je n'ai réussi qu'à finir le script des Côtes D'Armor, je ne pensais pas que le script pour la Côte D'Or serait aussi compliqué... Donc je suis en retard sur ce qui était prévu.

Liste des tâches prévues pour la semaine prochaine

- Finir le script pour la Côte D'Or.
- Passer à un autre script.

Semaine 6

Listes des t ches pr vues au cours de la semaine

- Finir le script pour la C te D'Or.
- Passer   un autre script.

Liste des t ches r alis es

Premier et deuxi me jour

- Finissions du script pour la C te D'Or :
 - La d marche suivie :
 - Finalement r -injecter les cookies dans une requ te avec Jsoup fonctionne. Je peux donc traiter le JSON ce qui me facilite la t che pour r cup rer toutes les infos dont j'ai besoin.

Troisi me jour

- Un probl me sur le script de la C te D'Or est signal  :
 - Probl mes rencontr s :
 - C'est le m me document qui est trait  plusieurs fois. En fait, j'ai mal g r  les listes dans le script. J'avais besoin de cr er des listes pour pouvoir les parcourir   l'envers (cf : « Derniers jours » de la semaine pr c dente) et au lieu de recrer des objets « List » je rempla ais leur valeur. Comme je stockais ces listes dans un tableau, toutes les listes du tableau pointaient vers la m me liste. Du coup, quand je changeais la valeur de ma liste, je changeait la valeur de toutes les listes du tableau.
 - La d marche suivie :
 - Au lieu de changer directement la valeur de la liste, je recrer un objet « List » pour que les listes du tableau ne pointent plus dessus.

Derniers jours

- Commencement du script pour le d partement de la Somme :
 - Probl mes rencontr s :
 - A chaque fois qu'on fait une action sur le site, la page se rafra chit et cr e une session en g n rant un token (= « jeton » en anglais) dans l'URL. Le token est g n r  gr ce   du JavaScript. On ne peut donc pas faire de requ te directement sur la page d'affichage des d lib rations via Jsoup en m thode POST, car l'URL change constamment.

Brunschwig N mo

- Solutions envisag es :
 - Il faut utiliser Selenium pour simuler un navigateur afin d'ex cuter le JavaScript de la page.
- La d marche suivie :
 - Sur la page de recherche il y a 3 menu d roulant : le premier est pour l'ann e, le 2  pour le type de s ance et le dernier est pour la date pr cise. J'ai donc fait 2 boucles (le 2  menu d roulant je choisi toujours le m me type de s ance). Mes 2 boucles sont bien g r es, je peut maintenant passer au traitement des donn es et la r cup ration des documents.

Analyse des  carts

J'ai fait tout ce qu'il y avait   faire,   savoir finir le script de la C te D'Or et commencer un nouveau script.

Liste des t ches pr vues pour la semaine prochaine

- Finir le script pour la Somme.
- Commencer un nouveau script.

Semaine 7

Listes des tâches prévues au cours de la semaine

- Finir le script pour la Somme.
- Commencer un nouveau script.

Liste des tâches réalisées

Premier jour

- Script pour le département de la Somme fini.

Derniers jours

- Commencement et finition du script pour le département de Tarn et Garonne
- Initialisation du script pour le département des Deux Sèvres.

Analyse des écarts

J'ai fait tout ce qu'il y avait à faire, j'ai même commencé un script supplémentaire.

Liste des tâches prévues pour la fin de semaine

Au moins finir le script pour le département des Deux Sèvres et si possible, en faire un autre.

Semaine 8

Listes des t aches pr evues au cours de la semaine

Au moins finir le script pour le d epartement des Deux S evres et si possible, en faire un autre.

Liste des t aches r ealis ees

Jusqu'  jeudi

- Finissions du script pour le d epartement des Deux S evres.
 - Probl emes rencontr es :
 - Il y avait des menus d eroulants contenant des ann ees (voir annexe 9). Quand on cliquait sur une des ann ees, cela faisait dispara tre les autres ann ees. D'autres menu apparaissaient contenant cette fois les types de s ances (voir annexe 10). Quand on clique sur une des s ances cela fait dispara tre les autres et ainsi de suite jusqu'  le bouton « d eliberations » apparaisse (voir annexe 11). On ne peut pas revenir en arri re pour acc der aux autres ann ees/s ances/dates.
 - Solutions envisag es :
 - Recharger la page   chaque fois qu'on a fini de r cup rer les documents d'une page.
 - La d emarche suivie :
 - J'ai d u recharger la page   chaque d ebut de boucle et j'ai d u faire une boucle pour chaque menu (une pour les ann ees, une pour les types de s ances, une pour parcourir les documents). J'avais donc un total de 4 boucles imbriqu es et je n'ai pas pu faire autrement que recharger la page   chaque d ebut de boucle.

Vendredi

- J'ai pu faire un dernier script, celui pour le d epartement de la Haute Marne. Je n'ai rencontr  aucun probl eme et j'ai pu le terminer en une journ ee.

Ce que m'a apporté le stage

- D'un point de vu « compétences » : à présent je maîtrise beaucoup mieux Java. Je l'ai dit à mon maître de stage lors d'un entretien de « clôture » du stage : avant je pouvais écrire sur mon CV que je connaissais les bases de Java. Maintenant je peux marqué que je le maîtrise assez bien. Même si il est vrai que, quand on maîtrise un langage, on sait à peu près coder dans tous les langages. Ce que je voulais dire c'est qu'à présent je connaît plus ou moins les particularités spécifiques à Java.
- D'un point de vu « relationnel » : comme on faisait un point tous les midi sur ce qu'on avait fait la veille et le matin même, on devait forcément un moment ou un autre prendre la parole devant tout le monde pour expliquer les difficultés qu'on avait rencontrer. Ça m'a permis de me décoincer un peu et je me suis senti plus à l'aise devant les autres à force de prendre la parole.
- D'un point de vu « méthodologique » : on utilisait Gitlab comme gestionnaire de version. Utiliser les différentes fonctionnalités de Gitlab (commit, issues, etc...) m'a permis de mieux comprendre comment gérer correctement un programme. Il y avait la branche « master » qui correspondait au programme en production, la branche « dev » qui est la branche sur laquelle on travail avant que nos scripts soient envoyés en production, Sarah et moi avions nos propres branches pour éviter tout problème. Quand on avait fini un script, Jean-Marc vérifiait que notre script fonctionnait bien puis faisait un « merge » pour mettre ce qu'on avait fait sur la branche « dev ».

Conclusion

Pendant 8 semaines j'ai fait la m me chose. Enfin dans le principe c' tait la m me chose mais comme chaque site est diff rent il fallait faire un script diff rent   chaque fois. J'ai d  trouver des m thodes adapt    chaque interface pour parvenir   aller chercher les documents de mani re automatis .   force de faire des recherches sur les meilleurs m thodes   mettre en  uvre afin de parvenir   mes fins, j'ai beaucoup appris sur le langage Java et ses particularit s. Je me suis perfectionn  dans ce langage mais aussi dans ma logique de d veloppement. Le stage m'a aussi permis de m'am liorer dans l'organisation de mes programmes et depuis j'organise beaucoup mieux mes projets personnels.

Annexes

Rapports et délibérations du Conseil départemental de l'Aisne

Recherche multi-critère optimisée pour le navigateur Internet Explorer

Du au

Séance

Thèmes

N° Définitif (Exemple : 305)

Dans le texte

 LANCER UNE RECHERCHE  NOUVELLE RECHERCHE

Annexe 1 : Exemple de page de recherche pour des délibérations, ici les délibérations de l'Aisne.

```
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image
webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9
Accept-Encoding: gzip, deflate, br
Accept-Language: en-US,en;q=0.9,fr;q=0.8,fr-FR;q=0.7
Cache-Control: max-age=0
Connection: keep-alive
Content-Length: 443
Content-Type: application/x-www-form-urlencoded
Cookie: ASPSESSIONIDSQTDSTR=BPMHCGKADGCEBKPPiHOALDOB
Host: deliberations.aisne.fr
Origin: https://deliberations.aisne.fr
Referer: https://deliberations.aisne.fr/
Sec-Fetch-Dest: document
Sec-Fetch-Mode: navigate
Sec-Fetch-Site: same-origin
Sec-Fetch-User: ?1
Sec-GPC: 1
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML
L, like Gecko) Chrome/88.0.4324.152 Safari/537.36
```

Annexe 2 : Exemple de headers, ici les headers pour la page des délibérations de l'Aisne

Dossiers 1 à 50 des 356 correspondant à la requête - Page 1 sur 8 - [Nouvelle requête](#)

Annexe 3 : Exemple de pagination, ici la page des délibérations de l'Aisne



x: 133
y: 37

Annexe 4 : Bouton recherche (image venant du site)

Annexe 5 : Coordonn ees du cliqu e envoy ees en m ethode POST

```
int Xmin = 1;  
int Xmax = 199;  
  
int Ymin = 1;  
int Ymax = 75;  
  
int random_x = (int) (Math.random() * (Xmax - Xmin + 1) + Xmin);  
int random_y = (int) (Math.random() * (Ymax - Ymin + 1) + Ymin);
```

Annexe 6 : script g en erant les valeurs « pseudo » al eatoires pour les coordonn ees du bouton « recherche »

Divers	
Num�ero de s�erie	12:B3:2F:20:FD:C1:8A:A8:BB:9D:63:69
Algorithme de signature	SHA-256 with RSA Encryption
Version	3
T�el�echarger	PEM (cert) PEM (chain)

Annexe 7 : Pour t el echarger le certificat SSL il faut cliquer sur « PEM (cert) » (l' autre n'est pas au bon format)

Annexe 8 : Informations du serveur trouv ees sur le site

Recherche d'une catégorie...

- 2021 >
- 2020 >
- 2019 >
- 2018 >
- 2017 >
- 2016 >
- 2015 >
- 2014 >
- 2013 >

Annexe 9 : Menus contenant les années

Recherche d'une catégorie...

- ▼ 2021
 - Commission Permanente >
 - Conseil départemental >

Annexe 10 : Menus contenant les type de séances (les autres menus ont disparut)

Recherche d'une catégorie...

- 2021
 - | Commission Permanente
 - ▼ || 2021-01-25
 - DELIBERATIONS

Annexe 11 : Bouton « DELIBERATIONS »