

Compte rendu de la session PFE du 19/09

Nous avons pu discuter lors de la séance, en compagnie de Mr Rouvoy et Mr Vastel, des objectifs à atteindre pour ce PFE : il s'agit ainsi d'observer la problématique de la détection de bot du point de vue du bot et de mettre en place une solution utilisant le machine learning afin de contourner les défenses que les sites web mettent en place.

Après discussions sur les possibles moyens d'atteindre ces objectifs, nous avons conclu que l'implémentation d'algorithmes du reinforcement learning était adapté au sujet, en raison du fait que nous désirons que le bot puisse apprendre de ses différents tests.

Il était ainsi question lors de cette séance de se documenter sur les bots et les différents types d'implémentation possible. J'ai ainsi pu observer que les bots classiques, ou basés sur PhantomJS ou CasperJS n'étaient plus d'actualité, étant entièrement remplacé par Chrome Headless, qui présente un nouveau challenge dans la détection de bots.

J'ai ainsi pu me documenter sur Chrome Headless et son fonctionnement, ainsi que sur la librairie officielle associée, Puppeteer.

J'ai également commencé à me documenter sur la théorie des algorithmes de reinforcement learning, afin de comprendre en profondeur l'outil que j'utiliserai dans le but de réaliser ce projet.

Enfin, j'ai pu rassembler une série de papiers de recherche décrivant l'état de l'art en terme de détection de bots. J'ai ainsi commencé à lire le contenu de quelques uns de ces papiers, qui m'ont permis de dresser une liste des paramètres utilisés dans la détection de bots, que j'ai noté dans la documentation mise en place, que je joins au compte rendu.

L'objectif pour la prochaine séance serait de renforcer mes connaissances en termes de bots et approfondir les méthodes de détection possibles, et surtout, me documenter dans le domaine du réinforcement learning.

Naif Mehanna