

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسيدي بلعباس
Ecole Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



Mémoire de Fin d'étude

Pour l'obtention du diplôme d'ingénieur d'état

Filière : Informatique

Spécialité : **Système d'information et Web (SIW)**

Thème

**Fouille de Web basé sur les méthodes d'apprentissage automatique et
apprentissage profond**

Présenté par :

- Melle Bousba Fadia Nouna
- Mr Felouh Brahim

Soutenu le : **16/09/2019**

Devant le jury composé de :

- | | |
|----------------------|-----------|
| - Keskes Nabil | Président |
| - Elouali Nadia | Examineur |
| - Allal Lamia | Examineur |
| - Bennabi Rim Sakina | Encadreur |

Année Universitaire : 2018 / 2019

Résumé

L'apprentissage en profondeur est une technique d'apprentissage automatique qui vise à extraire des représentations de haut niveau à partir de données brutes en utilisant plusieurs couches neuronales. Le succès récent de l'apprentissage en profondeur a montré qu'il surpassait les systèmes conventionnels en traitement d'image. La fouille du Web utilise les algorithmes de fouille de données pour détecter les modèles à partir de la toile. Il est divisé en trois axes : fouille du contenu du web, fouille de l'usage et fouille de la structure du web. L'usage de l'approche d'apprentissage profond peut offrir un potentiel considérable pour améliorer la fouille du Web, et ce en exploitant les grandes masses de données disponibles sur la toile d'une manière efficace. L'utilisation des méthodes conventionnelles en fouille de données sur ces grands volumes de données se trouve souvent limitée, n'exploitant pas au mieux ces bases d'information.

Ceci nécessite donc un apprentissage en profondeur afin d'extraire de la connaissance nouvelle et non évidente. SONATRACH est la Société nationale pour la recherche, la production, le transport, la transformation et la commercialisation des hydrocarbures. GTP (Grands Travaux Pétroliers) est une filiale du Groupe Sonatrach spécialisée dans l'ingénierie et la construction d'installations industrielles destinées à la production, transformation, transport et distribution des hydrocarbures.

Dans ce travail, on dresse un état de l'art des travaux existants dans l'usage des algorithmes d'apprentissage automatique et apprentissage profond dans l'exploration du Web. Ceci permet de comparer ces travaux et de les synthétiser pour avoir une vision globale sur cet axe de recherche. Dans notre application on va analyser l'usage d'un site Web à partir des fichiers journaux (log files) fournis par le serveur web du groupe GTP dans le but de La détection des failles sécuritaires de sites web de l'entreprise, on se focalise sur l'implémentation de la technique de la fouille de l'usage du web avec les trois méthodes suivantes ANN profond, KNN, et SVM.

Mots clés : la fouille de données, l'apprentissage profond, l'exploration de web, ANN(réseau de neurones profond), SVM, KNN.

Abstract

Deep learning is an automatic learning technique that aims to extract high-level representations from raw data using multiple neural layers. The recent success of deep learning has shown that it outperforms conventional image processing systems. Web search uses data mining algorithms to detect models from the web. It is divided into three areas : searching the content of the web, searching the usage and

searching the structure of the web. The use of the deep learning approach can offer considerable potential for improving Web browsing by leveraging the vast amount of data available on the web in an efficient manner. The use of conventional methods in data mining on these large volumes of data is often limited, not making the best use of these databases. This requires deep learning to extract new and unseen knowledge. SONATRACH is the National Society for Research, Production, Transport, Processing and Marketing of Hydrocarbons. GTP (Grands Travaux Pétroliers) is a subsidiary of the Sonatrach Group specializing in the engineering and construction of industrial facilities for the production, processing, transport and distribution of hydrocarbons.

In this work, we draw up a state of the art of existing works in the use of machine learning algorithms and deep learning in the exploration of the Web. This allows to compare these works and to synthesize them to have a global vision on this line of research. In our application we will analyze the use of a website from the log files provided by the web server of the GTP group for the purpose of the detection of security breaches of the company's websites. focuses on the implementation of the technique of the excavation of the use of the web with the following three methods ANN deep, KNN, and SVM.

key words : data mining, deep learning, Web mining, ANN(deep neural network), SVM, KNN.