

Description de PFE de Licence

Intitulé du projet : Un système intelligent d'organisation des fichiers sur son ordinateur

Encadrant du projet : Pr. Ahmed Guessoum, USTHB

Co-encadrant du projet :

Motivation du Projet :

Organiser son « desktop » c'est comme l'organisation de son bureau (livres, articles, etc.) : plus le contenu grandit, plus cela devient un calvaire de l'organiser et, surtout, de s'y retrouver. Des solutions existent, comme « Mendeley Desktop » qui organisent les permettent d'indexer les différents fichiers du « desktop » (bureau sur ordinateur/laptop) et, ainsi, retrouver facilement des documents par recherche avec des mots clés.

Le grand problème qui existe aujourd'hui avec nombre d'applications que nous pouvons installer sur nos machines et smart phones est celui de la confiance : comment peut-on être sûr qu'un logiciel installé sur nos machines, qui de surcroît a accès à tous nos documents (fichiers, photos personnelles, vidéos, etc.) est tout à fait innocent. Il n'y a contre cela d'autre solution, en l'absence d'outils/de logiciels de sécurité très sophistiqués et que nous n'avons pas encore, que de développer ses propres compétences... tout en ajoutant sa propre innovation.

Description du Projet :

Le but de ce projet est de développer un système intelligent de gestion des documents du desktop de l'utilisateur. Le système peut être semi-automatisé initialement puis, pourrait dans le temps (ce qui n'est pas exigé dans ce PFE) être entièrement automatisé. Ainsi, le système fera ce qui suit :

- Une indexation de tous les documents qui existent sur une machine
- Une recherche de documents pertinents par mots clés, domaines, etc.
- Une suggestion de réorganisation du desktop, un document à la fois, sur la base d'une classification par apprentissage automatique qui prend en considération les contenus (mots) de ce document.

D'autres opérations pourraient être suggérées selon le développement du système.

Option : l'utilisateur peut avoir des documents en arabe (en plus de l'anglais et du français).

Mots-clés: Traitement Automatique du Langage Naturel ; apprentissage automatique ; classification; indexation de fichiers.

Profil requis:

- L3 Académique ou ISIL ; bons programmeurs ; binôme motivé et sérieux ; bonne expression écrite en Français ; moyenne en L2 relativement bonne.

Plan du travail :

1. Recherche bibliographique
2. Familiarisation avec Python et la programmation d'outils TAL
3. Conception de l'interface et des fonctions de la plateforme
4. Implémentation, tests et évaluation.
5. Rédaction du mémoire. (Celle-ci se fera au fur et à mesure)

Bibliographie:

1. Daniel Jurafsky & James H. Martin, 2008, "*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*", 2ème Edition, Prentice Hall, 2008.
2. Nizar Y. Habash, "*Introduction to Arabic Natural Language Processing*", Morgan & Claypool publishers, 2010.
3. W Bruce Croft, Donald Metzler, Trevor Strohman, 2010, Search engines: Information retrieval in practice, Addison-Wesley
4. Tomek Strzalkowski (editor), 1999, Natural Language Information Retrieval, Kluwer Academic Publishers (Text, speech and language technology series, edited by Nancy Ide and Jean Véronis, volume 7).
5. www.nltk.org
6. Natural language Processing with Python; Steven Bird, Ewan Klein and Edward Loper; O'Reilly, 2009.