

Projet de Développement

INF1603 réalisé _

3 avril 2021

Table des matières

Table des matières	2
1 Menu	3
2 Explications sur l'analyse des articles	3
2.1 Les adresses email	3
2.2 Les affiliations	4
2.3 Autres éléments	4
3 Les formats de sortie	4

Sprint 4 - Rapport

3 avril 2021

1 Menu

Comme demandé dans le sprint4, nous avons choisi de faire un menu en ligne de commande plutôt qu'un menu graphique. Cela nous permet de garder la compatibilité avec toutes les plateformes. De plus, ce script devant s'exécuter en ligne de commande, nous avons préféré une interface épurée et facile d'utilisation.

Pour sélectionner les fichiers, il suffit d'écrire la liste des numéros de fichiers séparés par un espace, ou de mettre une étoile pour analyser tout le répertoire.

2 Explications sur l'analyse des articles

2.1 Les adresses email

Les adresses email étant déjà récupérés lors du sprint3, nous vous avons déjà expliqué notre méthode de récupération. Ici nous n'avons que fait l'association entre les adresses email et les auteurs.

Pour la majorité des articles, cela s'est bien passé, mais pour quelques articles vous verrez apparaître la balise <mail> alors que pour d'autres articles vous ne la verrez pas.

On peut expliquer cela facilement, tout simplement car nous ne pouvions d'aucune manière associer, et savoir (sauf visuellement) que ces emails correspondaient à ces auteurs. Nous avons donc choisis de faire apparaître tout de même les adresses mail sans les faire correspondre aux auteurs.

2.2 Les affiliations

Pour la totalité des PDFs, nous avons pu récupérer les affiliations en fonction de chaque auteur. Cependant, pour un ou deux articles scientifiques, il y avait au moins deux auteurs avec d'une part, les mêmes affiliations et d'autre part, un minimum de deux affiliations par auteur. Tout ceci, fait que notre fonction de comparaison n'a pas permis d'afficher les affiliations correspondantes aux auteurs. C'est pourquoi à certains endroits, les balises affiliations seront vide.

2.3 Autres éléments

Pour l'analyse de l'introduction, du corps du texte, de la conclusion et/ou de la discussion, nous avons créé des sections.

A noter que pour certains articles vous verrez apparaitre la balise conclusion (respectivement discussion) vide si la section n'existe pas dans le PDF.

3 Les formats de sortie

À travers les options `-t` ou `-x`, nous pouvons générer des sorties TXT ou XML comme demandé.

L'option `-a` permet d'avoir les deux sorties simultanément.