

# **Rapport de fin de Sprint**

## **Sprint numéro : 3**

Comme convenu avec le client, le troisième Sprint consistait à extraire de nouvelles informations et à créer une nouvelle sortie de fichier au format XML.

Notre équipe a réalisé les tâches suivantes :

- Extraction des adresses courriel des auteurs,
- Extraction de la section bibliographie,
- Ajout d'un nouveau type de sortie : fichier au format XML.

Notre équipe a réussi à extraire la section bibliographie, cependant, il est à noter que nous avons pris toute la section en un seul bloc. Autrement dit, nous n'avons pas analysé la bibliographie en extrayant référence par référence.

En conclusion de l'extraction de la bibliographie, ce qui a été demandé par le client a été concrétisé, mais nous pouvons encore améliorer cette extraction.

Concernant l'extraction des adresses courriel des auteurs, il n'y a pas eu de problèmes particuliers et l'équipe a pu réaliser avec facilité cette demande du client.

D'autre part, une amélioration essentielle a été faite par notre équipe. Elle a revu ses tests unitaires en ajoutant un « score » de précision. En comparant le résultat de l'analyse de notre « parser » au résultat que nous voulons obtenir, nous avons pu établir ce « score » de précision.

Nous avons ensuite testé sur les fichiers « .pdf » du Corpus et obtenu des résultats concluants (un tableau détaillé sera joint à ce rapport).

Enfin, l'ajout d'un nouveau type de sortie fonctionne pour tous les fichiers du Corpus.

En conclusion, les demandes du client ont été satisfaites dans les délais. Le programme est efficace dans le traitement des adresses courriel des auteurs, de la bibliographie et est apte à créer des fichiers au format XML.