

# **Rapport de fin de Sprint 5**

---

**Sprint numéro : 5**

**Durée : 1 semaine**

**Dates : 25/03 au 31/03**

---

## **Introduction**

Cette semaine, nous avons exploré un site qui nous a fourni des résultats de précision pour chaque section ainsi qu'une précision totale, calculée comme la somme des résultats de précision par section divisée par le nombre de sections. Ces mesures de précision ont été réalisées sur le Corpus que nous avons déjà à notre disposition. L'objectif principal de la semaine était donc de se familiariser avec le site et d'apporter des ajustements pour rendre le code aussi dynamique que possible en vue de préparer le dernier Sprint.

## **Tâches Accomplies**

Nous avons amélioré les tests et continué à perfectionner le code (plus particulièrement la discussion et la conclusion en remplaçant « Pas de discussion » et « Pas de conclusion » par « N/A »). Maintenant, le calcul de la précision global du fichier est égal à la moyenne des pourcentages par sections. Notre équipe a également amélioré l'aspect visuel des résultats des tests dans le terminal. Plus précisément, au lieu d'obtenir un résultat visuel pour chaque fichier, nous avons ajouté un rendu visuel pour chaque section de chaque fichier. Enfin, nous avons mis à disposition de nouveaux artefacts sur GitHub.

## **Visualisation des Tests dans le Terminal**

Nous avons continué à améliorer la présentation visuelle des tests dans le terminal. Cette amélioration vise à rendre le processus de test plus convivial et informatif pour l'équipe de développement, contribuant ainsi à renforcer la qualité globale du code et à accélérer le cycle de développement. Nous avons modifié la manière dont les résultats sont affichés pour chaque fichier en fournissant un rendu pour chaque section (voir l'image « schéma\_final.png » dans le dossier « Artéfacts Sprint 5 »), ce qui nous permet de détecter plus précisément, les lacunes dans notre analyse.

## Résultats et analyse de ceci

Bien que le site nous ait permis de détecter certaines erreurs, il n'était pas parfaitement codé et ne prenait pas en compte de nombreux paramètres, comme illustré par exemple sur la Figure 1 ci-dessous :

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<article>
  <preamble>Fichier original : jing-cutepaste.pdf</preamble>
  <titre>Titre : Cut </titre>
  <abstract>identifies </abstract>
  <references>upon </references>
</article>
```

Figure 1 - Exemple d'erreur

Sa fiabilité étant assez faible, nous avons obtenu quelques résultats, comme ceux ci-dessous. Ils sont donc à prendre avec du recul :

```
=>xml jing-cutepaste.xml extension validecommande : python3 compare2.py -s OutputXML/jing-cutepaste.xml OutputReferenceXML/jing-cutepaste.xml
Voici tous les resultats
Precision choisie : souple (90%)
section Precision
-----
Preamble 0%
Titre 15.69%
Auteurs 0.0%
Introduction 0%
abstract 3.88%
Discussion 0%
Conclusion 0%
```

Figure 2 - Résultat d'un des fichiers XML

Après avoir signalé au client les différentes erreurs rencontrées lors de nos tests avec le site, celui-ci a décidé de prendre avec lui l'un de nos fichiers XML afin de le tester et de nous rendre un résultat correct. Voici les résultats obtenus :

Precision choisie : souple (90%)	
section	Precision
Preamble	100.0%
Titre	100.0%
Auteurs	100.0%
Introduction	100.0%
abstract	100.0%
Discussion	100.0%
Conclusion	100.0%

*Figure 3 - Résultat pour le fichier XML jing-cutepaste*

Par rapport au résultat final, nous pouvons conclure que notre programme fonctionne parfaitement et que nous avons réussi à extraire toutes les informations.

### **Conclusion et objectifs pour les prochains Sprints**

En conclusion, le Sprint 5 a été un succès total. Notre équipe a su maîtriser le site fourni et a apporté des corrections en réponse aux erreurs détectées. Les résultats sont très concluants, démontrant que nous avons réussi à extraire toutes les informations de manière très précise pour chaque section.