

Parseur d'articles scientifiques en format texte

*** Sprint 4

Suite à d'autres études de marché, nous avons conclu que la structure XML doit être enrichie avec les sections suivantes, tout en conservant les deux sorties:

```
<article>
  <preamble> Le nom du fichier d'origine </preamble>
  <titre> Le titre du papier </titre>
  <auteurs>
    <auteur>
      <name>Auteur 1</name>
      <mail>auteur1@mail.com</mail>
      <affiliation> L'affiliation de l'auteur 1 et son
adresse</affiliation>
    </auteur>
    <auteur>
      <name>Auteur2</name>
      <mail>auteur2@mail.com</mail>
      <affiliation> L'affiliation de l'auteur 2 et son
adresse</affiliation>
    </auteur>
  </auteurs>
  <abstract> Le résumé de l'article</abstract>
  <introduction> La introduction</introduction>
  <corps> Le développement du papier</corps>
  <conclusion> La conclusion du papier</conclusion>
  <discussion> La discussion du papier</discussion>
  <biblio> Les références bibliographiques du papier</biblio>
</article>
```

La version texte n'ayant pas besoin d'être enrichie puisque les informations y sont de manière non structurées. Également, une nouvelle exigence du département est d'ajouter un menu textuel au début de l'exécution du système afin de choisir les PDFs à parser (et donc ne pas parser tout le répertoire à chaque fois).

Le système final, et les versions intermédiaires doivent être placés sur Github ; chaque version dans une branche différente. Le fichier README.md doit contenir une explication générale du système et la procédure nécessaire pour lancer les programmes. Il faudra également documenter tous les différents événements de la méthode Scrum en ajoutant sur votre Github des photos de votre travail (réunions de planification de sprint, mêlée quotidienne, revue de sprint, rétrospective du sprint).

Consignes

- Équipes : fixée en séance 1
- Méthodologie agile : SCRUM
- Plateforme : le projet doit fonctionner sous GNU/Linux en ligne de commande
- Aucune conversion en ligne via navigateur n'est acceptée
- Langage : fixé en séance 2
- Soyez agiles !



Vous êtes en compétition ! Attention aux fuites des idées ou de logiciel !