

Parseur d'articles scientifiques en format texte (carnet du produit)

*** Sprint 3

Suite à une discussion à l'IRISA, nous avons décidé que les fichiers parsés plein texte ne suffisent plus à nos besoins. Nous avons entendu dire que les fichiers structurés XML possèdent plusieurs avantages et ils sont largement utilisés en TAL (traitement automatique de la langue). Nous préférons donc comme sortie un fichier XML utf8 avec la structure suivante:

```
<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>
    </auteur>
    <auteur>
      <name>Auteur2</name>
      <mail>auteur2@mail.com</mail>
    </auteur>
  </auteurs>
  <abstract> Le résumé de l'article </abstract>
  <biblio> Les références bibliographiques du papier</biblio>
</article>
```

Cette nouvelle sortie .xml **complète** la sortie texte .txt du sprint précédent.

Ainsi, au moment de lancer votre parseur, il faudra ajouter un argument (**-t** | **-x**) pour bien choisir le type de sortie : soit des fichiers dans une version texte (**-t**) , soit au format XML (**-x**) :

En complément des différentes sections extraites précédemment, vos sorties devront inclure une section contenant les adresses courriels des auteurs ainsi que la section bibliographie.

Le système final, et les versions intermédiaires doivent être placés sur Github. Le fichier README.md doit contenir une explication générale du système et la procédure nécessaire pour lancer les programmes. Ainsi il faut documenter tous les différents événements de la méthode Scrum en ajoutant sur ENT des images de votre travail (photos de réunions de planification de sprint, mêlée quotidienne, revue de sprint, rétrospective du sprint) ainsi que captures d'écran et/ou des documents écrits à la main (dessins, idées, etc.).

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
- Dépôts : sur ENT et GIT (vous devez m'avoir ajouté comme membre dans votre git)
- Soyez agiles !



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