

TP Génie logiciel - Scrum 2020



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

*** Version 4

Le Département de contrôle de qualité demande une **évaluation de la qualité du système**. Cette évaluation peut être réalisée en fonction de la **précision** obtenue sur un **corpus de référence** (découpage de sections fait par des humains). Ainsi, les fragments de texte extraits présents dans les références seront comptabilisés positivement; leur absence sera pénalisée (voir l'exemple en Annexe).

La précision sera calculée comme suit :

$$\text{Précision} = \frac{\text{Sections_correctes_trouvées_par_le_système}}{\text{Sections_trouvées_par_le_système}}$$

Une section sera considérée correcte si :

- Toutes ses phrases ont été bien détectées : frontières strictes.
- Quelques phrases peuvent chevaucher les sections (± 2 phrases) : frontières souples.

Il faut ajouter une dernière phase d'évaluation du système qui reporte le résultat de la précision.

La précision sera mesurée sur un **corpus de test** de **10 nouveaux articles PDF**. Une partie de l'équipe doit passer votre système sur les 10 articles, une autre doit établir le **corpus de référence** les frontières (souples) des sections. Enfin, vous devez calculer (à la main si vous voulez être agiles) la précision de votre système comme la moyenne sur l'ensemble des articles parsés. Finalement, un rapport doit être écrit sous forme d'article scientifique. Cet article doit avoir les éléments suivantes :

- Titre
- Auteurs
- Abstract
- Méthode (explication du système)
- Résultats
- Conclusions

L'article doit être rédigé de préférence en Latex (bonus de +1pt) et doit être rendu en format PDF sur le ENT (minimum 3, maximum 4 pages).

Groupe1 : Séance 8 (16/12), Séance 9 (6/01)

Remise : Séance 9 (06/01 avant minuit)

Groupe3 : Séance 8 (09/12), Séance 9 (06/01)

Remise : Séance 9 (06/01 avant minuit)

Alt-Grp4 : Séance 8 (16/12), Séance 9 (06/01)

Remise : Séance 9 (06/01 avant minuit)

Le rapport, le système final et les versions intermédiaires doivent être placés sur ENT et sur Github. Chaque version dans une branche différente. Il faut documenter tous les différents événements de la méthode Scrum en ajoutant sur l'ENT des photos de votre travail (réunions de planification de sprint, mêlée quotidienne, revue de sprint, rétrospective du sprint).

Consignes

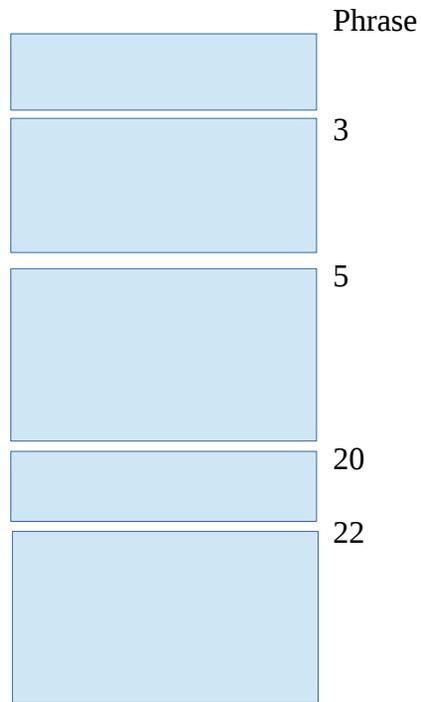
- Déposer vos rapports et codes sur ENT
- Équipes : Fixée en séance 1
- Méthodologie agile : SCRUM+XP
- Plateforme : le projet doit fonctionner sous GNU/Linux en ligne de commande.
- Choix de langage : Fixé en séance 1
- Soyez agiles !

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

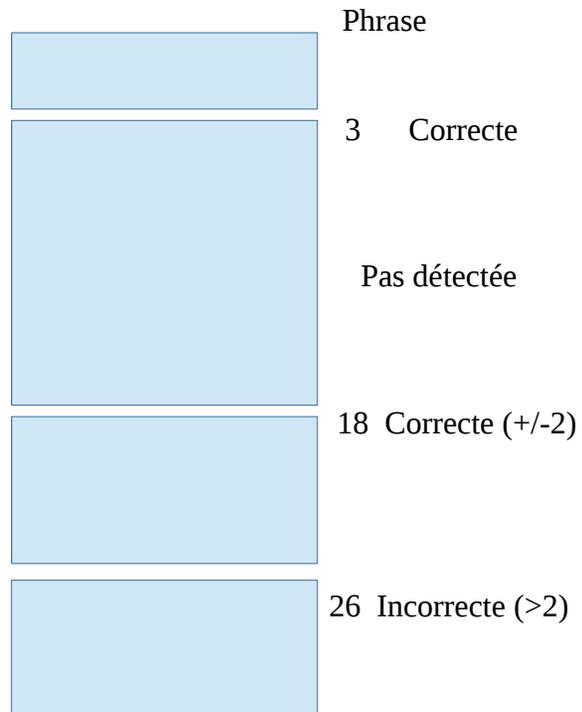


Annexe. Exemple de découpage et calcul de la précision

Découpage de référence (humain)



Découpage automatique



Frontières véritables = 4

Frontières trouvées = 3

Frontières correctes = 2

Frontières incorrectes = 1

Frontières non détectées = 1

$$\text{Précision} = \text{correctes/trouvées} = 2/3 = 0,666$$

Vous devez calculer la moyenne des précisions obtenues sur les 10 articles du corpus de test.