

PROPOSITION DE PROJET ISIMA

Filière F2 : Génie Logiciel et Systèmes Informatiques

Projet proposé en : ☒ 2^{ème} année ☐ 3^{ème} année

Intitulé du projet : Nuage de mots

Personne responsable : David Hill

Email : david.hill@univ-bpclermont.fr

Téléphone : 04 73 40 50 19

Entreprise ou Laboratoire : LIMOS UMR CNRS 6158

Adresse : <http://www.isima.fr/~hill>

Mots clés décrivant le projet : Extraction de connaissance, C, C++, Big Data, Evaluation des performances.

Description du travail :

Le travail principal consistera à réaliser en C et en C++ un logiciel d'analyse de texte capable de produire un compte rendu statistique sur les mots utilisés d'un texte. Les textes étudiés utiliseront les caractères français.

Le développement sera réalisé tout d'abord de façon simple séquentielle, les codes C et C++ seront évalués pour leurs performances sur de gros fichiers.

On prévoira d'exclure une liste de petits mots récurrents inutiles (le, les, un, de, des,...)

En fonction de l'avancement du projet, une sortie graphique simple est envisagée sans être obligatoire.

Un travail en binôme du style eXtreme Programming peut être réalisé pour, d'une part développer en étant tiré par les tests, et d'autre part, échanger les rôles de développeurs testeurs en alternant par exemple les rôles tous les quinze jours.

L'utilisation d'outils logiciel adaptés aux développements collaboratifs est recommandée, tout en étant capable de livrer un archive simple à un client potentiel.

Une attention sera portée à la qualité du code source en fonction d'un style que vous aurez choisi.

Environnement de travail : (Unix/Linux) :

Outils / langages utilisés :

- | | | | |
|---|--|----------------------------------|--|
| <input checked="" type="checkbox"/> C/C++ | <input type="checkbox"/> Android | <input type="checkbox"/> Scala | <input type="checkbox"/> Objective C |
| <input type="checkbox"/> Java | <input type="checkbox"/> IOS | <input type="checkbox"/> Fortran | <input type="checkbox"/> XCode / Eclipse |
| <input type="checkbox"/> C# | <input type="checkbox"/> Windows Phone 7 | <input type="checkbox"/> OpenCL | <input type="checkbox"/> OpenGL |

Autre langages et/ou environnement : _____

Le projet est soumis à une clause de confidentialité : ☐ OUI ☒ NON

Le travail demandé pourra se faire en partie en dehors de l'ISIMA (Entreprise ou laboratoire) :

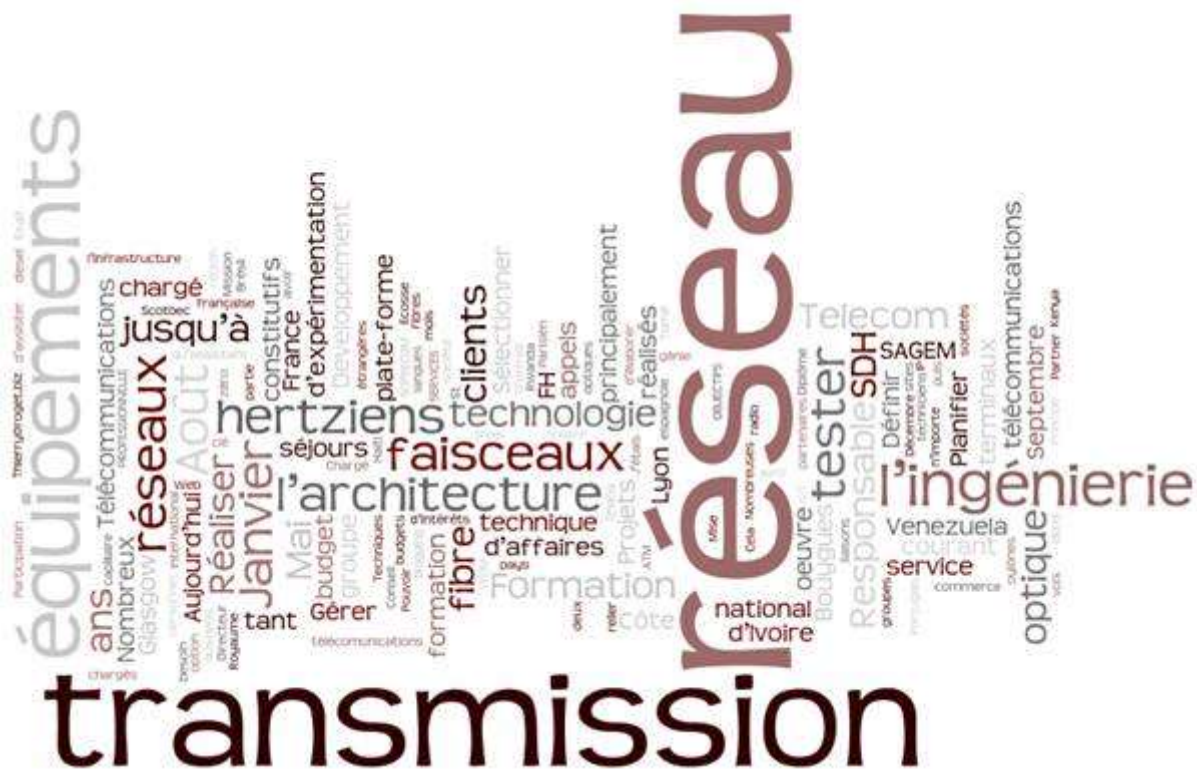
☒ OUI ☐ NON

Pour des raisons propres à votre entreprise/labο, souhaitez-vous qu'une convention ou un accord de confidentialité soit établi ?

☐ OUI ☒ NON

Autres remarques :

Une étude sera faite des logiciels et sites Web existants et proposant ce service.



Document à retourner au plus tard le 1 octobre 2015 à :

Christophe Duhamel (christophe.duhamel@isima.fr)
cc : **Claude Mazel** (mazel@isima.fr)