

OFFRE DE STAGE

Développeur/Développeuse TAL (Traitement Automatique des Langues)

LOCALISATION : Pau (Billère, 64)

Le Congrès permanent de la langue occitane est une institution académique de régulation de la langue occitane. Il rassemble les fédérations historiques et les institutions d'étude, de valorisation et transmission de cette langue, et est soutenu par les collectivités (Régions Occitanie et Nouvelle-Aquitaine via l'Office public de la langue occitane, Auvergne-Rhône-Alpes) et le ministère de la Culture et de la Communication-DGLFLF.

Le Congrès est éditeur du portail locongres.org (300 000 visites/an), plateforme de ressources lexicales et d'outils linguistiques en langue occitane. Il pilote également la Feuille de route de développement numérique de l'occitan, dirige différents projets en technologies de la langue (programme LINGUATEC, clavier prédictif Android, correcteur orthographique pour éditeurs de texte, client de messagerie et navigateurs web) et travaille en partenariat avec le laboratoire CLLE-ERSS (UMR CNRS/Toulouse 2) autour de la constitution d'une base textuelle (BaTelÒc) et d'un lexique ouvert des formes fléchies (LOFIÒc).

Le Congrès est constitué d'une équipe de 6 personnes (directeur, développeuse TAL/Webmaster, chargé de mission linguistique, lexicographes et secrétaire-comptable).

MISSIONS

Vos missions s'effectueront dans le cadre d'un consortium européen associant universités, académies de la langue et une société de développement de logiciels : LINGUATEC (EFA227/16) « Développement de la coopération transfrontalière et du transfert de connaissances en technologies de la langue ». Il s'agit d'un programme retenu par le second appel à projets du Programme de Coopération Territoriale Espagne-France-Andorre POCTEFA (2014-2020) qui a pour objectif le transfert de technologies et le développement de ressources et d'applications linguistiques innovantes en aragonais, basque et occitan.

En fonction de la durée du stage et de vos centres d'intérêt, vous effectuerez tout ou partie des tâches suivantes :

1. Réalisation d'un lexique de formes fléchies contenant les mots français des dictionnaires bilingues du Congrès :

Dans le cadre de la refonte de son site Internet, Le Congrès souhaite fusionner ses applications en une seule multi-application. Celle-ci permettra d'accéder, pour un mot tapé par l'utilisateur, à ses traductions (ou aux traductions de son lemme), à ses flexions, aux expressions qui le contiennent... Cet outil permettra, entre autres, d'accéder aux traductions en occitan du lemme d'une forme fléchie en français.

Vous serez chargé de la réalisation de la base de formes fléchies françaises qui sera interrogée par cet outil, qui servira également à construire divers outils comme un traducteur automatique. Vous extrairez les formes françaises des dictionnaires bilingues du Congrès, recenserez les formes fléchies existantes et générerez les formes fléchies manquantes. Vous pourrez vous appuyer pour ce faire sur les règles de flexion du français et

sur la base de formes fléchies Morphalou.

2. Enrichissement des ressources en français du traducteur automatique Apertium :

Dans le cadre de Linguatéc, Le Congrès développe un traducteur automatique occitan-français et français-occitan. Afin d'améliorer son fonctionnement, vous aurez la charge d'enrichir les lexiques de formes fléchies en français d'Apertium à partir du lexique de forme fléchies décrit ci-dessus ainsi que du lexique de formes fléchies Morphalou.

Vous serez également chargé d'améliorer le PoS-tagger français d'Apertium en enrichissant sa base de règles de désambiguïsation.

PROFIL

Etudiant en Master 1 informatique ou linguistique.

Capacité à utiliser un ou plusieurs langages de programmation.

Connaissances solides en grammaire française.

Autonomie, rigueur, capacité d'analyse, maîtrise des échéances

DUREE DU STAGE

Entre un et deux mois.

REMUNERATION

Montant de la gratification obligatoire. Si stage inférieur à 44 jours, possibilité de prise en charge de l'hébergement.

Envoyer CV + Lettre de motivation à : b.dazeas@locongres.org