

[OC]

Ofèrta d'emplec



Lo Congrès permanent de la lenga occitana recruta un/una

Desvolopaire/Desvolopaira

Lòc de trabalh : trabalh en ret a partir d'un dels burèus del Congrès o de las institucions sòcias.

Natura e durada del contracte de trabalh : CDI de projècte.

Durada de trabalh setmaniera : 35 òras.

Debuta del contracte : setembre de 2024.

Salari : segon experiéncia.

/ Presentacion del projècte

Lo Congrès permanent de la lenga occitana a enregat un programa de creacion de **plataforma lexicala, textuala e toponimica** dins l'encastre del Contracte de plan Estat-Region novèl (CPER). S'agís d'un esplech innovant e interactiu d'accès unic a l'ensemble de las donadas lingüísticas (diccionaris, gramatica, fonologia, còrpus textuels, toponimia, etc.), sa creacion comprendrà la redaccion del *Diccionari general informatizat de la lenga occitana* e mai un plan de restauracion toponimica en associacion amb l'IGN (dont lo Congrès es d'ara enlà lo referent per la restitucion e la normalizacion de las formas toponimicas occitanas). Es previst tanben un programa dedicat a la constitucion massisa de ressorsas TAL (Tractament automatic de las lengas) amb lo formatage e l'anotacion dels còrpus lexicografics e textuels.

Amb lo sosten de las regions Auvèrnhe-Ròse-Alps, Novèla Aquitània e Occitània, aital coma de l'Estat (ministèri de la Cultura, DRAC Novèla Aquitània e Occitània), aquel programa pluriannal (2024-2027) permetrà lo desenvolopament d'una infraestructura numerica d'excelléncia - inspirada del portal lingüistic del francés, lo CNRTL (<http://www.cnrtl.fr>), una referéncia en la matèria - que beneficiarà a totes los usatgièrs de la lenga, al public grand, als cercaires, als escolans e estudiants, als ensenhaires e formators, aital coma a las collectivitats.

Coma per sas lengas vesinas - basca, espanhòla, catalana e francesa - l'occitan, gràcias a aquel programa, serà atal dotat d'un **diccionari monolingüe descriptiu**, cobrissent totes sas varietats, amb descripcion fonologica, accepcions, definicions, exemples, dichas, provèrbis, citacions literàrias, etc. Es un trabalh de mantuna annada, e i aurà mestièr d'una equipa fornida dins lo domeni de la lexicografia, qu'aurà una edicion numerica permetent una difusion larga gràcias als estandards d'interoperabilitat. Se pòt pensar tanben a una edicion papièr en mantun volume mai tard. Aquela òbra academica de prestigi es indispensable a la

coneissença e a la difusion de la lenga. En efiech, dempuèi lo *Tresor dóu Felibrige* de Frederic Mistral, cap d'òbra lexicografica panoccitana d'aquel nivèl scientific e d'aquela dimension es pas estada produsida fins ara. Se parla donc d'un obrador istoric mobilizant l'ensemble de las coneissenças, de las ressorsas e de las competéncias scientificas e tecnicas en lenga d'òc.

/ Descripcion del pòste

Jos la direccion de la responsabla del pòl informatic e dins una equipa de lexicografs, prendretz part a las òbras de :

- Formatage de lexic e dictionaris del Congrès (XML, nòrma TEI-P5).
- Formatage de còrpus textuals (monolingües, bilingües, audio transcriches).
- Generacion automatica o semiautomatica de lexic (extraccion de sinonims, generacion de formas flechidas, correccion e crosament de donadas).
- *Data mining* (extraccion de donadas per tractament lexicografic).
- Participacion a la construccion d'espleches TAL, web o mobils (règlas de desambiguïzacion, segon perfil realizacion de sits web, plug-ins e aplicatius).
- Redaccion de quasèrn de las cargas e seguit dels prestataris.
- Redaccion de notícias e produccion des tutorials

Seretz sostengut dins vòstra mission per l'equipa tecnica del Congrès.

/ Descripcion del perfil

De formacion superiora, mestrejatz la programacion Python orientada sul tractament de las donadas textualas (sustot XML). Sètz a vòstre aise dins un environament *Linux* aital coma amb l'utilizacion de las linhas de comanda.

Mestrejatz la gestion de las donadas SQL aital coma los formats CSV, HTML e XML.

Segon lo perfil, una coneissença de la nòrma TEI-P5, de framework de tip modèl-vista-controlaire (de preferéncia Django) o de PHP, JavaScript e/o React Nativa serián un mai.

Sètz metodic e rigorós, avètz una exigéncia pertocant la qualitat de vòstras òbras.

Vòstre esperit d'iniciativa e vòstra autonomia, doblats d'un solid engatjament personal son de reals atots. Sètz fòrça de proposicion.

Sètz a vòstre aise en reünion aital coma dins los grops de trabalh. Mestrejatz los espleches burotics e collaboratius, avètz una apeténcia per l'innovacion e lo numeric.

Avetz enveja de rejónher una còla dinamica, dins una ret scientifica e tecnologica europèu, desvolopant d'aplicacions de poncha per la lenga

occitana. Sètz prèst a vos engatjar per mantuna annada sus un obrador TAL istoric per l'occitan.

/ Le Congrès permanent de la langue occitane

Lo Congrès permanent de la lenga occitana es l'organisme interregional de regulacion de l'occitan. A tres sites (Novèla Aquitània, Occitània e Auvèrnhe-Ròse-Alps), e amassa las institucions e federacions istoricas del territòri occitanofòne. Lo sostenon lo ministèri de la Cultura e las collectivitats territorialas. Institucion academica, a per mission de contribuïr a la vitalitat e al desvolopament de l'occitan - apelat tanben lenga d'òc - en trabalhant a sa coneissença e a sa codificacion per la produccion dels espleches pertocant los diferents aspèctes de la lenga (la lexicografia, la lexicologia, la terminologia, la neologia, la fonologia, la grafia, la gramatica e la toponimia).

Lo Congrès es l'editor de la plataforma numerica - dicodoc.eu - amassant diferents espleches lingüistics de referéncia : un multidiccionari occitan, un conjugador, una basa terminologica, una basa toponimica, de sinonimes, un còrpus textual, aital coma d'aplicacions mobilas.

Amb son equipa de permanents, son Conselh lingüistic e sa ret de partenaris, a desvolopat una expertesa dins lo domeni del TAL (tractament automatic de la lenga). Baileja a aquel títol la *Fuèlha de rota de desvolopament numeric de l'occitan*, un document-quadre de planificacion e de creacion de ressorsas e d'espleches en tecnologies lengatgièras. Sòci del consòrci europèu LINGUATEC, es editor de la primièra sintèsi vocala (votz.eu) e del traductor automatic (revirada.eu) en occitan, e desvolopa d'aquesta passa la reconeissença vocala (ReVÒc).

DEPAUS DE LAS CANDIDATURAS :

Son obèrtas las candidaturas fins al 15 de junh. Lo dossièr de candidatura es compausat d'un CV e d'una letra de motivacion a enviar per via desmaterializada a : b.dazeas@locongres.org

[FR]



Offre d'emploi

Le Congrès permanent de la langue occitane recrute un/une

Développeur/Développeuse

Lieu de travail : travail en réseau depuis un des bureaux du Congrès ou des institutions partenaires.

Nature et durée de contrat de travail : CDI de projet.

Durée de travail hebdomadaire : 35 heures.

Début du contrat : septembre 2024.

Salaire : selon expérience.

/ Présentation du projet

Le Congrès permanent de la langue occitane a engagé dans le cadre du nouveau Contrat de plan de État-Région (CPER) un programme de création de **plateforme lexicale, textuelle et toponymique**. Il s'agit d'un outil innovant et interactif d'accès unique à l'ensemble des données linguistiques (dictionnaires, grammaire, phonologie, corpus textuels, toponymie, etc.), sa création comprend également la rédaction du *Dictionnaire général informatisé de la langue occitane* ainsi qu'un plan de restauration toponymique en partenariat avec l'IGN (dont le Congrès est devenu référent pour la restitution et la normalisation des formes toponymiques occitanes). Un volet dédié à la constitution massive de ressources TAL (Traitement automatique des langues), avec le formatage et l'annotation des corpus lexicographiques et textuels, est également prévu.

Soutenu par les régions Auvergne-Rhône-Alpes, Nouvelle-Aquitaine et Occitanie ainsi que l'État (ministère de la Culture, DRAC Nouvelle-Aquitaine et Occitanie), ce programme pluriannuel (2024-2027) permettra le développement d'une infrastructure numérique d'excellence - inspirée du portail linguistique du français, le CNRTL (<http://www.cnrtl.fr>), une référence en la matière - qui bénéficiera à tous les usagers de la langue, au grand public, aux chercheurs, aux élèves et étudiants, aux enseignants et formateurs ainsi qu'aux collectivités.

Comme pour ses langues voisines - basque, espagnole, catalane et française - l'occitan sera ainsi doté grâce à ce programme d'un **dictionnaire monolingue descriptif complet**, embrassant toutes ses variétés, avec description phonologique, acceptions, définitions, exemples, dictons, proverbes, citations littéraires, etc.. Il s'agit d'un travail de plusieurs années, nécessitant une équipe fournie dans le domaine de la

lexicographie, qui aura une édition numérique permettant sa plus large diffusion grâce aux standards d'interopérabilité. Une édition papier en plusieurs volumes pourra être ultérieurement envisagée. Cette œuvre académique de prestige est indispensable à la connaissance et à la diffusion de la langue. En effet, depuis le *Tresor dóu Felibrige* de Frédéric Mistral, aucune œuvre lexicographique panoccitane de ce niveau scientifique et de cette envergure n'a été produite à ce jour. Il s'agit donc d'un chantier historique mobilisant l'ensemble des connaissances, des ressources et des compétences scientifiques et techniques en langue d'oc.

/ Descriptif du poste

Sous la direction de la responsable du pôle informatique et au sein d'une équipe de lexicographes, vous participerez aux travaux de :

- Formatage de lexiques et dictionnaires du Congrès (XML, norme TEI-P5).
- Formatage de corpus textuels (monolingues, bilingues, audio transcrits).
- Génération automatique ou semi-automatique de lexiques (extraction de synonymes, génération de formes fléchies, correction et croisement de données).
- *Data mining* (extraction de données pour traitement lexicographique).
- Participation à la construction d'outils TAL, web ou mobiles (règles de désambiguïsation, selon profil réalisation de site web, plug-in et applicatifs).
- Rédaction du cahier des charges et suivi des prestataires.
- Rédaction des notices et production des tutoriels

Vous serez appuyé dans votre mission par l'équipe technique du Congrès.

/ Description du profil

De formation supérieure, vous maîtrisez la programmation Python axée sur le traitement des données textuelles (notamment XML). Vous êtes à l'aise dans un environnement Linux ainsi qu'avec l'utilisation des lignes de commande.

Vous maîtrisez la gestion des données SQL ainsi que les formats CSV, HTML et XML.

Selon le profil, une connaissance de la norme TEI-P5, de frameworks de type modèle-vue-contrôleur (de préférence Django) ou de PHP, JavaScript et/ou React Native seraient un plus.

Vous êtes méthodique et rigoureux, vous avez une exigence quant à la qualité de vos travaux.

Votre esprit d'initiative et votre autonomie, doublés d'un solide

engagement personnel sont de réels atouts. Vous êtes force de proposition.

Vous êtes à l'aise en réunion ainsi que dans les groupes de travail. Vous maîtrisez les outils bureautiques et collaboratifs, vous avez une appétence pour l'innovation et le numérique.

Vous souhaitez rejoindre une équipe dynamique, au sein d'un réseau scientifique et technologique européen, développant des applications de pointe pour la langue occitane. Vous êtes prêt à vous engager pour plusieurs années sur un chantier TAL historique pour l'occitan.

/ Le Congrès permanent de la langue occitane

Le Congrès permanent de la langue occitane est l'organisme interrégional de régulation de l'occitan. Installé dans trois sites (Nouvelle-Aquitaine, Occitanie et Auvergne-Rhône-Alpes), il rassemble les institutions et fédérations historiques du territoire occitanophone et il est soutenu par le ministère de la Culture et les collectivités territoriales. Institution académique, il a pour mission de contribuer à la vitalité et au développement de l'occitan - appelé aussi langue d'oc - en travaillant à sa connaissance et à sa codification par la production des outils concernant les différents aspects de la langue (la lexicographie, la lexicologie, la terminologie, la néologie, la phonologie, la graphie, la grammaire et la toponymie).

Le Congrès est éditeur de la plate-forme numérique - dicodoc.eu - rassemblant différents outils linguistiques de référence : un multidictionnaire occitan, un conjugueur, une base terminologique, une base toponymique, des synonymes, un corpus textuel ainsi que des applications mobiles.

Fort d'une équipe de permanents, de son Conseil linguistique et de son réseau de partenaires, il a développé une expertise dans le domaine du TAL (traitement automatique de la langue). Il pilote à ce titre la *Feuille de route de développement numérique de l'occitan*, un document-cadre de planification et de création de ressources et d'outils en technologies langagières. Membre du consortium européen LINGUATEC, il est éditeur de la première synthèse vocale (votz.eu) et du traducteur automatique (revirada.eu) en occitan, et développe actuellement la reconnaissance vocale (ReVÔc).

DÉPÔT DES CANDIDATURES :

Les candidatures sont ouvertes jusqu'au 15 juin 2024. Le dossier de candidature est composé d'un CV et d'une lettre de motivation à envoyer par voie dématérialisée à : b.dazeas@locongres.org