

Titre

Expert en modèles de données pour la transcription de ressources orales

Contexte

L'objectif du projet H2020 Parthenos est de mettre en place un ensemble de recommandations, d'outils et de données qui facilitent le déploiement des méthodes numériques en sciences humaines. Dans ce cadre, Inria coordonne les activités relatives aux modèles de données et aux standards de représentation associés, afin de favoriser la création et la réutilisation de jeux de données interopérables.

La publication en 2016 de la norme internationale ISO 24624 (Gestion des ressources linguistiques -- Transcription du langage parlé) offre l'opportunité de disposer enfin d'un cadre normatif précis pour l'échange de données orales transcrites et annotées tels que produite au sein de différents logiciels (CHAT, ELAN, EXMARaLDA, FOLKER ou Transcriber) et de permettre la mise à disposition de nombreux corpus sous une forme facilement réutilisable.

Mission

Il s'agit de, au sein du work package « standards » du projet Parthenos, d'identifier, de développer et de disséminer des ressources et des outils favorisant l'adoption de la norme ISO 24624. En particulier, il s'agira de :

- Réunir des informations sur les corpus oraux existant pour en réunir des exemples les plus pertinents représentatifs des bonnes pratiques dans le domaine de la transcription de données orales ;
- De mettre à la disposition des communautés concernées d'outils de conversion des formats existant de et vers la nouvelle norme ;
- De développer des composants permettant de mieux interroger, parcourir ou visualiser des ressources conformes à la norme (schémas, bibliothèque XSLT, API).

En particulier, la personne recrutée aura la mission de contribuer à l'espace d'information sur les standards mis en place au sein du projet Parthenos.

Descriptif du poste

La personne recrutée devra travailler étroitement avec les communautés scientifiques concernées pour en identifier les pratiques actuelles ainsi que leurs attentes dans le domaine de la représentation de données. Elle pourra à l'occasion participer ou même organiser des rencontres permettant de recueillir des besoins et de présenter les résultats obtenus.

Le travail demandé combinera donc expertise technique et capacité de communication.

Profil recherché

Les candidats doivent avoir un niveau maîtrise en sciences humaines, informatiques ou sciences de l'information et des bibliothèques et pouvoir démontrer une solide compétence en modélisation des données en lien avec une ou plusieurs disciplines en sciences humaines, si possible dans un contexte de transcription de l'oral. Une bonne connaissance des technologies autour de XML est indispensable, avec si possible une connaissance des directives de la TEI. Il est important de savoir travailler en contexte pluridisciplinaire, de savoir correctement s'exprimer en anglais à l'écrit comme à l'oral et de savoir à la fois être autonome dans son travail et en interaction avec des équipes internationales.

Conditions d'embauche

Il s'agit d'une mission de six mois comme ingénieur expert comprenant tous les avantages normalement associés à un recrutement chez Inria. Le salaire mensuel est calculé sur la base du niveau d'expérience du candidat.

La localisation de l'activité sera sur le site d'Inria Paris au sein de l'équipe Alpage dans un environnement scientifique motivant en linguistique computationnelle.

Veillez envoyer votre candidature (CV et lettre de motivation) par courriel à: laurent.romary@inria.fr . La date limite de candidature est le 31 janvier 2017.