



Offre n°2021-03253

Stage TAL : Paraphrases technologiques par apprentissage profond.

Type de contrat : Stage

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Stagiaire de l'ingénierie

Niveau d'expérience souhaité : De 3 à 5 ans

Contexte et atouts du poste

Le Startup Studio Inria propose un stage pour le projet qatent au sein de l'équipe projet ALmanaCH en TAL. Il s'agit de : **Paraphrases technologiques par apprentissage profond.**

Mission confiée

L'essentiel pour réussir

- Il est important d'être opérationnel en python et outils TAL d'apprentissage profond (NLTK, Spacy, Pytorch, ...)
- Aussi : Vous êtes passionné(e) par l'étude de la langue et par le traitement automatique de la langue

Début du contrat idéal : début mars.

FAQ

- Sur quelles données entraîner ?
 - Lire : Aaditya Prakash, Sadid A Hasan, Kathy Lee, Vivek Datla, Ashequl Qadir, Joey Liu, and Oladimeji Farri. 2016. Neural paraphrase generation with stacked residual lstm networks. arXiv:1610.03098. <https://arxiv.org/abs/1610.03098> (MSCOCO, Quora Duplicates, WikiAnswers Duplicates, PPDB)
- Quel type de modèle neuronal utiliser ?
 - lire : <https://proceedings.neurips.cc/paper/2019/file/5e2b66750529d8ae895ad2591118466f-Paper.pdf> <https://opendata.stackexchange.com/questions/6094/paraphrase-data-sets>
 - V. aussi: Wang, S., Gupta, R., Chang, N. and Baldrige, J., 2019, July. A task in a suit and a tie: paraphrase generation with semantic augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 7176-7183). http://suwangcompling.com/wp-content/uploads/2018/10/AAAI_2019_Draft_3-1.pdf

Plus d'informations voir :

<https://qatent.com/jobs>

Principales activités

Stage 2: paraphrases spécifiques aux brevets (environ 6 mois)

- Entraînement d'un système de génération de paraphrases spécifiques au corpus des brevets par apprentissage croisé avec des corpus génériques de paraphrases.
- Spécialisation du modèle de langage par domaine technologique (e.g. tels que définis par les sections CPC/IPC).
- Évaluation de mesures de similarités entre phrases.
- Utilisation des ontologies (Wordnet, stage 1, ...) pour obtenir une mesure / ranking de spécificités des paraphrases proposées.
- Intégration dans une API Django.

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle

Informations générales

- **Ville** : Paris
- **Centre Inria** : [Centre Inria de Paris](#)
- **Date de prise de fonction souhaitée** : 2021-04-01
- **Durée de contrat** : 6 mois
- **Date limite pour postuler** : 2021-02-28

Contacts

- **Équipe Inria** : INCUB-PRO (DGD-I)
- **Recruteur** :
Sagot Benoit / Benoit.Sagot@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.