

Application BASTRI

Fiches Equipes

ALMANACH (SR0776AR)

Modélisation et analyse linguistique automatique et humanités computationnelles
ALPAGE (SR0098OR) □ ALMANACH

Statut: Décision signée

Responsable : Benoit Sagot

Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" : *Aucun mot-clé.*

Mots-clés de "B - Autres sciences et domaines d'application - 2024" : *Aucun mot-clé.*

Domaine : Perception, Cognition, Interaction
Thème : Langue, parole et audio

Période : 01/01/2017 -> 30/06/2025
Dates d'évaluation : 02/10/2019 ,

Etablissement(s) de rattachement : <sans>
Laboratoire(s) partenaire(s) : <sans UMR>

CRI : Centre Inria de Paris
Localisation : Centre de recherche Inria de Paris
Code structure Inria : 021145-0

Numéro RNSR : 201722248N
N° de structure Inria: SR0776AR

Présentation

L'équipe-projet ALMANACH (*Automatic Language Modelling and Analysis & Computational Humanities*) est une équipe pluridisciplinaire en intelligence artificielle (IA) dédiée aux domaines du Traitement Automatique des Langues (TAL) et des Humanités Numériques (HN), se situant à la croisée de l'informatique théorique, de l'apprentissage automatique et de la linguistique. Les recherches de l'équipe couvrent un large éventail de sujets, et notamment les modèles de langue neuronaux, la traduction automatique, la modélisation de dialogues, le développement de ressources linguistiques (corpus monolingues, parallèles et annotés, lexiques, etc.), l'IA interactive, les stratégies d'évaluation, l'extraction d'informations, la reconnaissance optique de caractères et la reconnaissance de texte manuscrit. L'équipe s'intéresse aux données provenant de domaines variés, notamment des contenus produits par les utilisateurs, des données biomédicales, des brevets, ainsi que des documents historiques. Au-delà des données purement textuelles, l'équipe s'intéresse également à des scénarios multimodaux impliquant la parole et les images. Un défi transversal à l'ensemble des recherches de l'équipe est la variation linguistique dans toute sa diversité (en genre, style, niveau de langue, et en variation dialectale et diachronique), à la fois comme un défi pour les systèmes actuels et comme un objet d'étude.

Axes de recherche

- Analyse linguistique automatique améliorée par les informations contextuelles
 - Analyse linguistique automatique en contexte à tous les niveaux: morphologie, syntaxe, sémantique
 - Intégration du contexte dans les systèmes de traitement automatique des langues
 - Extraction d'informations et de connaissances
- Modélisation computationnelle de la variation linguistique
 - Linguistique théorique synchronique
 - Variation sociolinguistique
 - Variation diachronique
 - Variation en termes d'accessibilité
- Modélisation et développement de ressources linguistiques
 - Construction, gestion et annotation automatique de corpus textuels
 - Développement de ressources lexicales
 - Développement de corpus annotés

Contact

- **Responsable :** Benoit Sagot
- **Tél :** 01.80.49.43.14
- **Secrétariat Tél :** 01.39.63.52.62

En savoir plus

- Site de l'équipe
- Site sur inria.fr
- Site du [responsable](#)
- Derniers Rapports d'Activité : [2017](#) , [2018](#) , [2019](#) , [2020](#) , [2021](#) , [2022](#) , [2023](#)

Documents sur la structure

- [Intranet](#)
- [Privés](#)

Décisions

- [12049](#) (01/01/2017) : création
- [12544](#) (11/12/2017) : prolongation
- [12916](#) (21/06/2018) : prolongation
- [13694](#) (01/07/2019) : création
- [14340](#) (30/07/2020) : prolongation
- [16995](#) (26/04/2024) : prolongation
- [17498](#) (02/12/2024) : prolongation

Localisation

- **Adresse postale :** Centre Inria de Paris 48, rue Barrault CS 61534 75647 PARIS CEDEX
- **Coordonnées GPS :** 48.826, 2.346

Relations industrielles et internationales