

# 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" :** A3.1.1. Modélisation, représentation, A3.1.7. Données ouvertes, A3.1.8. Données massives (production, stockage, acheminement), A3.1.11. Données structurées, A3.2.2. Extraction de connaissances, nettoyage, A3.2.5. Ontologies, A3.3.2. Fouille de données, A3.3.3. Analyse de données massives, A3.4. Apprentissage et statistiques, A3.4.1. Apprentissage supervisé, A3.4.2. Apprentissage non supervisé, A3.4.3. Apprentissage par renforcement, A3.4.4. Optimisation pour l'apprentissage, A3.4.5. Méthodes bayésiennes, A3.4.6. Réseaux de neurones, A3.4.8. Apprentissage profond, A3.5. Réseaux sociaux, A3.5.2. Systèmes de recommandation, A5. Interaction, multimédia et robotique, A5.1. Interaction Homme-Machine, A5.1.1. Ingénierie des systèmes interactifs, A5.1.2. Evaluation des systèmes interactifs, A5.1.7. Interfaces multimodales, A5.1.8. Interfaces 3D, A5.1.9. Analyses perceptives et études utilisateurs, A5.6. Réalité virtuelle, réalité augmentée, A5.6.1. Réalité virtuelle, A5.6.3. Simulation et incarnation d'avatars, A5.7. Modélisation et traitement audio, A5.7.3. Parole, A5.7.4. Analyse, A5.7.5. Synthèse, A5.8. Traitement automatique des langues, A9. Intelligence artificielle, A9.1. Connaissances, A9.2. Apprentissage, A9.3. Analyse de signaux (vision, parole, etc.), A9.4. Traitement automatique des langues, A9.7. Algorithmique de l'intelligence artificielle, A9.8. Raisonnement, A9.10. Approches hybrides de l'IA

**Mots-clés de "B - Autres sciences et domaines d'application - 2024" :** B1. Sciences du vivant, B1.1. Biologie, B1.2. Neurosciences et sciences cognitives, B1.2.2. Sciences cognitives, B1.2.3. Neurosciences computationnelles, B2.2.6. Maladies neuro-dégénératives, B2.5. Handicap et assistances à la personne, B2.5.2. Handicaps cognitifs, B9.5.1. Informatique, B9.5.6. Science des données, B9.6. Sciences humaines et sociales, B9.6.1. Psychologie, B9.6.2. Droit, B9.6.5. Sociologie, B9.6.6. Archéologie, Histoire, B9.6.8. Linguistique, B9.6.9. Sciences politiques, B9.6.10. Humanités numériques, B9.7. Diffusion du savoir, B9.7.1. Accès ouvert, B9.7.2. Données ouvertes, B9.8. Recherche reproductible, B9.9. Ethique, B9.10. Confidentialité, vie privée

**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-

### 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](http://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

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

### Relations industrielles et internationales