Traitement Automatique des Langues Naturelles (TALN) [1]

Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier TAL et Humanités Numériques (TAL-HN)

Ludovic Moncla, Carmen Brando (Éds.)

Avignon, France, 27 juin au 1er juillet 2022

1. https://taln2022.univ-avignon.fr

Avec le soutien de

Préface

Atelier TAL et Humanités Numériques (TAL&HN)

Ludovic Moncla[1] Carmen Brando[2]
(1) Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69621 Villeurbanne, France
(2) EHESS, CRH, UMR8558, Paris, France
ludovic.moncla@insa-lyon.fr, carmen.brando@ehess.fr

L'atelier TAL&HN, associé à la conférence TALN | RECITAL 2022, s'inscrit dans le cadre du travail mené avec l'action de recherche Humanités Numériques Spatialisées [2] soutenue par le GdR CNRS MAGIS. Elle fait suite à une série d'événements organisés par notre action de recherche tels que l'atelier Humanités Numériques Spatialisés associé à la conférence SAGEO 2021 et la série d'ateliers Geospatial Humanities [3] associés à la conférence internationale ACM SIGSPATIAL. Cette première édition de l'atelier associé à la conférence TALN reprend les principaux thèmes du TAL appliqués à des problématiques ou des corpus issus des travaux en sciences humaines et sociales :

— préparation de corpus et acquisition de données en SHS,
— interopérabilité de données,
— développement de ressources et Open Data,
— fouille de textes, apprentissage automatique et apprentissage profond,
— recherche d'informations, extraction de connaissances, reconnaissance d'entités nommées,
— classification de documents,
— graphe de connaissances et ontologies.

En complément, nous proposons un focus sur deux thématiques en particulier (1) l'articulation des pratiques d'encodage de corpus SHS par la norme TEI et des opportunités d'automatisation proposées par le TAL et (2) le développement de méthodes d'apprentissage (automatique ou profond) et l'entraînement de modèles de langue pour le traitement de corpus SHS (historique, littéraire, etc.).

Notre comité de programme est constitué d'une équipe pluridisciplinaire (TAL et humanités numériques) de 26 collègues chercheurs et enseignants-chercheurs issus de diverses équipes et laboratoires de recherche francophones. Nous avons reçu 13 soumissions présentant des résultats de recherche (à des niveaux différents de maturité) en lien avec les thèmes de l'atelier. Chaque article a été relu par au moins 3 relecteurs du comité de programme et suite à la phase de relecture 10 ont été acceptés pour publication. Ces 10 articles se regroupent en 3 sessions autour des thèmes suivants : 1) l'étude du genre dans des corpus de textes, 2) l'annotation sémantique en tenant compte des erreurs d'OCéRisation et 3) l'analyse de discours et sémantique des textes. Du point de vue de la science ouverte et des normes dites FAIR, les corpus utilisés dans ces travaux sont publiés en TEI (Text Encoding Initiative) et parfois déposés dans des entrepôts de données de la recherche comme Ortolang d'Huma-Num.

Les articles présentés proposent un aperçu des travaux en humanités numériques qui s'appuient sur l'utilisation et le développement de méthodes numériques pour le traitement et l'analyse de corpus textuels. Nous tenons à remercier les auteurs pour leurs contributions ainsi que les participants à l'atelier et nous espérons que ce numéro sera utile pour la communauté.

2. https ://aphns-magis.projet.liris.cnrs.fr/
3. https ://ludovicmoncla.github.io/sigspatial-geohumanities-2022/

Table des matières