



TextMine - Atelier sur la Fouille de Textes **Conférence [EGC 2022](#)**

Journée du 25 janvier 2022

<http://www.vincentlemaire-labs.fr/TM2022>

Pascal Cuxac ; Cédric Lopez ; Vincent Lemaire

INIST- CNRS ; Envista ; Orange Labs

Depuis 2017, l'atelier TextMine rassemble un large public d'étudiants et de chercheurs du public comme du privé. Nous proposons un nouvel opus lors de la conférence [EGC'22](#) qui se tiendra à Blois du 24 au 28 Janvier 2022.

Le but de cet atelier est de réunir des chercheurs sur la thématique large de la fouille de textes. Cet atelier vise à offrir une occasion de rencontres pour les universitaires et les industriels, appartenant aux différentes communautés de l'intelligence artificielle, l'apprentissage automatique, le traitement automatique des langues, afin de discuter des méthodes de fouille de texte au sens large et de leurs applications.

Descriptif :

C'est une évidence que de dire que nous sommes entrés dans une ère où la donnée textuelle sous toutes ses formes submerge chacun de nous que ce soit dans son environnement personnel ou professionnel : l'augmentation croissante de documents nécessaires aux entreprises ou aux administrations, la profusion de données textuelles disponibles via Internet, le développement des données en libre accès (Open Data), les bibliothèques et archives en lignes, les médias sociaux ne sont que quelques exemples illustrant l'évolution de la notion de texte, sa diversité et sa prolifération.

Face à cela les méthodes automatiques de fouille de données (data mining), et plus spécifiquement celles de fouille de textes (text mining) sont devenues incontournables. Récemment, les méthodes de deep learning ont créé de nouvelles possibilités de recherche pour traiter des données massives et de grandes dimensions. Cependant, de nombreuses questions restent en suspens, par exemple en ce qui concerne la gestion de gros corpus textuels multi-thématiques. Pouvoir disposer d'outils d'analyse textuelle efficaces, capables de



s'adapter à de gros volumes de données, souvent de nature hétérogène, rarement structurés, dans des langues variées, des domaines très spécialisés ou au contraire de l'ordre du langage naturel reste un challenge.

La fouille de textes couvre de multiples domaines comme le traitement automatique des langues, l'intelligence artificielle, la linguistique, les statistiques, l'informatique...et les applications sont très diversifiées, que ce soit la recherche d'information, le filtrage de spam, le marketing, la veille scientifique ou économique, la lutte antiterroriste ...

En France, des conférences comme TALN, CORIA, JADT par exemple sont centrées sur l'analyse et le traitement des textes, mais avec des approches plus ciblées soit TAL, soit RI, soit statistiques. Cet atelier se veut plus fédérateur autour d'approches et d'applications aussi diverses que possibles.

Le but de cet atelier est de réunir des chercheurs sur la thématique large de la fouille de textes. Cet atelier vise à offrir une occasion de rencontres pour les universitaires et les industriels, appartenant aux différentes communautés de l'intelligence artificielle, l'apprentissage automatique, le traitement automatique des langues, pour discuter des méthodes de fouille de texte au sens large et de leurs applications.

Objectifs :

L'objectif de cet atelier est de favoriser des présentations et des discussions. Peuvent être soumis :

- des résumés étendus (4 pages)
- des articles longs (maximum 12 pages)
- des propositions de démonstrations logicielles (4 pages)

Les contributions seront publiées sous forme d'actes en ligne (avec ISBN). La soumission de prises de position bien articulées, d'expériences industrielles et de travaux en cours sont les bienvenus et privilégiés. Des contributions portant sur l'intérêt pratique des travaux, qu'elles viennent de l'industrie ou du monde académique, ou présentant des collaborations entre les deux seraient appréciées. Le but est le partage d'expérience et de savoir sur les problématiques liées à la fouille de textes.

Pour les démonstrations une présentation orale aménagée devra être préparée (temps de présentation plus court, et temps pour effectuer la démo sur projecteur en fin de présentation).



Un temps sera prévu dans le programme pour les démos.

Dates importantes :

Date limite de soumission des articles : **01/12/2021**

Notification aux auteurs : **13/12/2021**

Version finale : **10/01/2022**

Atelier : 25/01/2022

Le format à utiliser est le même que pour EGC ; Les soumissions sont à faire sur :
<https://easychair.org/conferences/?conf=textmine22>

Principaux thèmes (liste non limitative) :

- Méthodes de traitement automatique de la langue (TAL)
- Classifications statiques de textes
- Classifications dynamiques de textes
- Méthodes d'apprentissage
- Approches par graphes
- Recherche d'information
- Indexation
- Détection d'entités nommées
- Résumé automatique
- Détection de nouveautés
- Analyse de sentiments
- Les bots ou systèmes de dialogues
- ...

Liste des domaines d'application (liste non limitative) :

- Fouille de documents scientifiques
- Données médicales
- Brevets
- Analyse d'opinions
- Media sociaux (Twitter...)
- Veille scientifiques
- Intelligence économique
- Appui au pilotage scientifique
- Bibliométrie
- Systèmes (à base) de dialogues
- ...