

TextMine - Atelier sur la Fouille de Textes

Journée du 22 janvier 2019

<http://www.vincentlemaire-labs.fr/TM2019>

Dates importantes (dates prévisionnelles) :

- Date limite de soumission des articles : **03/12/2018**
- Notification aux auteurs : **18/12/2018**
- Version finale : **08/01/2019**
- Atelier : **22/01/2019**
- **Site de soumission**
- Les soumissions sont à faire sur : <https://easychair.org/conferences/?conf=textmine19>

Résumé :

Après le succès des ateliers TextMine en 2017 et 2018, nous proposons un troisième numéro TextMine'19 lors de la conférence EGC'19 qui se tiendra à Metz. 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 toute 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 media 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 créées 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 démonstration 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.

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
- ...

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
- ...

Modalités et déroulement de l'atelier :

- Introduction par les organisateurs
- Exposé invité
- Exposés selon articles reçus et sélectionnés
- Table ronde

Comité d'organisation :

Pascal Cuxac

DPI-SRDE

INIST – CNRS

2 allée du Parc de Brabois

CS 10310

54519 Vandoeuvre les Nancy Cedex

pascal.cuxac@inist.fr

<https://sites.google.com/site/pascalcuxac>

Vincent Lemaire

ORANGE-LABS

2 avenue Pierre Marzin

22300 Lannion

vincent.lemaire@orange.com

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

Comité de programme (au 07/10/2018) :

Guillaume	Cabanac	IRIT	Toulouse
Mariane	Clausel	LJK	Grenoble
Vincent	Claveau	IRISA	Rennes
Guillaume	Cleuziou	LIFO	Orléans
Dominique	Gay	U. Réunion	St Denis de la Réunion
Natalia	Grabar	STL – Lille3	Lille
Mustapha	Lebbah	LIPN	Paris
Denis	Maurel	LIT	Tours
Patrick	Paroubeck	LIMSI	Orsay
David	Reymond	i3N	Toulon – Nice
Julien	Velcin	ERIC	Lyon