

Groupe de Travail EGC TextMine



emvista



EGC

Groupe de Travail EGC TextMine

Le but de ce groupe de travail est de réunir des acteurs (chercheurs, ingénieurs, doctorants ...) sur la thématique large de la **fouille de textes**.

Ce groupe 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.

Ce groupe de travail est **ouvert** à toute personne s'intéressant à la fouille de textes sous toutes ses formes (TAL, ML, Web Sémantique...), venant du monde académique ou industriel, de France ou d'autres pays.

Groupe de Travail EGC TextMine

Créé en 2022

Membres du groupe de travail :

Kévin Cousot, Emvista kevin.cousot@emvista.com

Pascal Cuxac, CNRS-INIST pascal.cuxac@inist.fr

Vincent Lemaire, Orange labsvincent.lemaire@orange.com

Cédric Lopez, Emvista cedric.lopez@emvista.com

> 60 inscrits à cette journée de lancement

LIRMM - univ Montpellier / Institut National de l'Audiovisuel / CEGEDIM / INRIA / Univ Paris 8 / CIRAD / LASTI - CEA / Nextino / IRIT - Berger-Levrault / Akio / ACS / Geerico - Univ Lille3 / INIST CNRS / Univ de La Réunion / Universität Tübingen / INRAE / CEFAS-CNR / Univ Paris-Nanterre / NLP - Dassault Systèmes / INIST-CNRS / Université de Strasbourg / Loria / INIST-CNRS / INRAE / Univ. Paris 1 / CSTB / Nextino / TETIS - INRAE / DDOR CNRS / BIMBAMJOB / SNCF / LaTICE - ESSTT - University of Tunis, Tunisia. / INRAE / MIRACL FSEGS / Université de Franche-Comté / Service de la Recherche - Institut National de l'Audiovisuel / LS2N Nantes Université / Aix-Marseille Univ / EDF r&d / CNRS / DEPE - INRAE / LIUM, Université du Mans / LS2N Nantes Université / CEA LIST / LIUM, Université du Mans / Univ Lorraine / AGORA - CY Cergy Paris Université / Orange / LIG / ELICO - Univ. Lyon 1

Activités du GT TEXTMINE

- **Fédérer** la collectivité scientifique francophone autour de la fouille de textes au sens large (TAL, Machine Learning, Web sémantique, Traduction automatique...)
- **journée d'échanges** autour d'invités (monde académique et industriel) avec des conférenciers invités et des présentations d'étudiants,
- journée **hackathon, démo** outils / jeux de **données**,
- Organisation de **challenges**,
- Construire et animer l'**atelier TextMine** en conjonction avec la conférence EGC.
- Être une pépinière pour la préparation de soumissions à **appels à projets**,
- **Communiquer** autour de la fouille de textes et de nos activités,

Atelier TextMine - EGC

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 soit :

- sous forme **d'actes en ligne...**
- dans une **revue électronique** (Journal of Data Mining & Digital Humanities <https://jdmdh.episciences.org>)
- dans une collection **HAL**

Atelier TextMine - EGC

Chaque année, depuis 2017

Site Web : <https://textmine.sciencesconf.org/resource/page/id/4>

Actes téléchargeables ici :

<https://textmine.sciencesconf.org/resource/page/id/5>

Prochaine édition lors de la conférence EGC à Lyon le 17 janvier 2023 à Lyon

Le groupe de travail TextMine

Informations sur notre site : <https://textmine.sciencesconf.org/>

Venez nous rejoindre pour :

- **animer** le GT et proposer des actions (ex : organisation défi, groupe de lecture, session démos, ...)
- **participer** au CP de l'atelier TextMine
- **proposer** des communications à l'atelier TextMine

Programme de la journée

9h30 – 10h00 : Accueil - café

10h00 – 10h30 : Ouverture de la journée

10h00 : Prise de parole représentant EGC (Arnaud Martin)

10h10 : Présentation du groupe de travail TextMine (P. Cuxac, V. Lemaire, C. Lopez)

10h30 - 11h00 : Présentation du défi TextMine 2022-2023 (prix décerné) par Kévin Cousot (Emvista)

11h00 – 11h45 : Logiciel CorText Manager - Extraction d'information et analyse socio-sémantique pour les sciences humaines et sociales par Lionel Villard (IFRIS-LISIS)

11h45 – 12h10 : Réflexions en cours pour dynamiser la recherche en TAL et TDM autour des données ISTEEX par Mathieu Constant (ATILF)

12h10 – 13h30 : Déjeuner offert aux participants

13h30 – 14h00 : Du text-mining aux macroscopes sociaux par David Chavalarias (ISC-PIF)

14h00 – 14h45 : Analyse de données textuelles à la SNCF : cas d'usage et difficultés rencontrées par Coralie Reutenauer (Direction Risques Audits Sécurité Sûreté à la SNCF)

14h45 – 15h30 : Extraction de l'information en santé végétale par Mariya Borovikova (INRAE),

15h30 – 16h15 : Apprentissage actif pour la détection de classes arbitraires dans les contenus multimédia. par Paul Guélorget (AIRBUS)

16h15 – 16h30 : Clôture de la journée (P. Cuxac, V. Lemaire, C. Lopez)