



Traitement Automatique des Langues Naturelles
(TALN) ¹

Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles.
Atelier DÉfi Fouille de Textes (DEFT)

Cyril Grouin, Gabriel Illouz (Éds.)

Avignon, France, 27 juin au 1^{er} juillet 2022

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

Avec le soutien de



Préface

DEFT, une campagne d'évaluation francophone

Cyril Grouin

Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400 Orsay
cyril.grouin@lisn.upsaclay.fr

Le défi fouille de textes (DEFT) a été créé en 2005, à l'image des campagnes d'évaluation internationales sur l'anglais, pour proposer à la communauté francophone du traitement automatique des langues (TAL) un ensemble de tâches et de corpus accessibles. Chaque année, DEFT propose un ensemble de tâches régulièrement renouvelées, et des corpus textuels diversifiés pour répondre aux problématiques proposées. Les dernières campagnes DEFT ont ainsi permis aux participants de traiter différents genres textuels (articles législatifs, articles de presse, articles scientifiques, nouvelles littéraires courtes, etc.), des productions individuelles de locuteurs francophones (recettes de cuisine, messages postés sur les réseaux sociaux), du contenu de langue de spécialité (cas cliniques publiés en revue spécialisée), et depuis peu, du contenu pédagogique (questions et réponses courtes d'étudiants).

Autour de ces différents corpus, nous avons proposé plusieurs tâches de base pour la fouille de texte : classification de documents, fouille d'opinion (identification de la polarité globale d'un document, détection de l'humour, de l'ironie et du sarcasme, identification des portions porteuses de sentimentopinion-émotion (OSE) et de la cible), indexation automatique, recherche d'information (RI), et repérage d'entités nommées (REN). Sur cette base, des tâches plus fines ont également existé, de manière à donner un contexte applicatif réel, notamment pour travailler sur l'origine géographique et la variation diachronique d'articles de presse, pour déterminer le profil clinique d'un patient, ou encore pour identifier la similarité sémantique entre différents contenus textuels (résumé et article scientifique complet, titre et recette de cuisine, etc.).

L'intérêt majeur de participer à une campagne d'évaluation consiste à pouvoir comparer plusieurs méthodes appliquées sur le même jeu de données, pour une même tâche, et dans un laps de temps imposé. Au-delà des effets de mode, les performances obtenues par chacune des approches proposées par les participants permet de relativiser l'importance accordée à certaines techniques. L'existence d'une baseline met en évidence les performances réelles des systèmes, en raison de l'écart calculé entre résultats de ces systèmes vis à vis de l'approche de base employée.

Enfin, nous observons avec satisfaction que les participants aux campagnes DEFT ne sont plus exclusivement des équipes académiques, mais que des centres de recherche et développement privés s'intéressent également aux problématiques scientifiques proposées. Cette nouvelle édition ne fait pas exception, alors même qu'il s'agissait de travailler sur un système de notation automatique de réponses courtes d'étudiants.

Nous remercions tous les participants actuels et passés pour l'intérêt qu'ils ont manifesté en proposant des méthodes originales pour répondre au défi fouille de textes.

Table des matières

Notation automatique de réponses courtes d'étudiants : présentation de la campagne DEFT 2022	1
<i>Cyril Grouin, Gabriel Illouz</i>	
Stylo@DEFT2022 : Notation automatique de copies d'étudiant · e · s par combinaisons de méthodes de similarité	11
<i>Ibtihel Ben Ltaifa, Toufik Boubehziz, Andrea Briglia, Corina Chutaux, Yoann Dupont, Carlos-Emiliano González-Gallardo, Caroline Koudoro-Parfait, Gaël Lejeune</i>	
Participation de l'équipe TGV à DEFT 2022 : Prédiction automatique de notes d'étudiants à des questionnaires en fonction du type de question	23
<i>Vanessa Gaudray Bouju, Margot Guettier, Gwennola Lerus, Gaël Guibon, Matthieu Labeau, Luce Lefevre</i>	
Correction automatique d'examens écrits par approche neuronale profonde et attention croisée bidirectionnelle	36
<i>Yanis Labrak, Philippe Turcotte, Richard Dufour, Mickael Rouvier</i>	
Participation d'EDF R&D à DEFT 2022	45
<i>Philippe Suignard, Xiaomin Huang, Meryl Bothua</i>	