Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

SimpleText CLEF 2026 : Simplifier les textes scientifiques (et rien de plus) *

Liana Ermakova[1] Hosein Azarbonyad[2] Jan Bakker[3] Gautam Kishore Shahi [4]
Benjamin Vendeville[1,5] Jaap Kamps[3]
(1) Université de Bretagne Occidentale, HCTI, Brest, France
(2) Elsevier, The Netherlands
(3) University of Amsterdam, Amsterdam, The Netherlands
(4) University of Duisburg-Essen, Essen, Germany
(5) Bretagne INP, Lab-STICC CNRS UMR 6285, Brest, France

RÉSUMÉ

Au cours des dernières années, SimpleText a fédéré une communauté active de chercheurs en TALN et en recherche d'information (IR) travaillant à améliorer l'accès aux textes scientifiques. Ses jeux de données de référence sur la recherche de passages scientifiques, la détection et l'explication de la terminologie scientifique, ainsi que la simplification de textes scientifiques, sont devenus des références dans le domaine. Après avoir conservé une structure similaire entre 2021 et 2024, nous avons introduit des modifications substantielles dans l'organisation et les tâches en 2025. Nous prévoyons de poursuivre cette configuration réussie en 2026, tout en ajoutant une nouvelle tâche pilote consacrée à la classification des articles scientifiques par domaine de recherche. SimpleText CLEF 2026 comprendra ainsi les trois tâches suivantes. Tâche 1 – Simplification de texte : simplifier des textes scientifiques. Tâche 2 – Créativité contrôlée : identifier et éviter les hallucinations. Tâche 3 – Classification par domaine de recherche : classer des articles scientifiques selon leur domaine de recherche.

ABSTRACT

CLEF 2026 SimpleText Track : Simplify Scientific Text (and Nothing More)

Over the last few years, the SimpleText Track has created an active community of NLP and IR researchers collaborating to improve access to scientific text. Its benchmarks on scientific passage retrieval, scientific terminology detection and explanation, and scientific text simplification have become standard references. Following a similar track design from 2021 to 2024, we introduced substantial changes to the track's structure and tasks in 2025. We plan to continue this successful setup in 2026, plus add a new (pilot) task on research area classification of scientific papers. Hence, the CLEF 2026 SimpleText track will contain the following 3 tasks. Task 1 (Text Simplification) : simplify scientific text. Task 2 (Controlled Creativity) : identify and avoid hallucination. Task 3 (Research Area Classification) : classification of scientific articles by research area.

MOTS-CLÉS : Recherche d'information, simplification de textes, documents scientifiques.

KEYWORDS: Information retrieval, text simplification, scientific documents.

ARTICLE ACCEPTÉ À : The 48th European Conference on Information Retrieval (ECIR 2026).
URL : https://doi.org/10.1007/978-3-032-21321-1_31

∗. Cette recherche a été financée en tout ou partie, par l'ANR au titre du projet ANR-22-CE23-0019-01.

Cette œuvre est mise à disposition sous licence Attribution 4.0 International.