Article de recherche **Rev. Cong. Sci. Technol., Vol. 05, No. 03, pp. 1533-1540 (2026)**

OPEN ACCESS

Revue Congolaise des Sciences & Technologies

ISSN: 2959-202X (Online); 2960-2629 (Print)

https://www.csnrdc.net/

REVUE CONGOLAISE DES SCIENCES ET TECHNOLOGIES

Développement d'une plate-forme intelligente d'aide à la détection du plagiat et au choix de sujets de mémoire : cas de la Faculté des Sciences et Technologies de l'Université de Kinshasa

[Intelligent Platform for Plagiarism Detection and Thesis Topic Selection: Case of the Faculty of Sciences and Technologies at the University of Kinshasa]

Moïse KASOMBO TSHIBANGU*

Université de Kinshasa, Faculté des Sciences et Technologies, Mention Mathématiques, Statistiques et Informatique, Kinshasa, République démocratique du Congo

Résumé

Le recensement manuel des redondances entre sujets de mémoire et la vérification des similitudes dans les travaux académiques demeurent difficiles à la Faculté des Sciences et Technologies de l'Université de Kinshasa, en raison de l'absence d'outils adaptés au contexte local. Cet article présente AcademiaCheck, une plate-forme Web articulant deux fonctions : la validation intelligente de sujets de mémoire et l'aide à la détection de similitudes dans les travaux déposés. Le moteur, implémenté nativement en TypeScript, combine une vectorisation TF-IDF, la similarité cosinus et l'indice de Jaccard. Cette preuve de concept évalue principalement la validation de sujets sur un corpus anonymisé de 83 dossiers (57 à 110 mots) comparés à un catalogue réel de 114 sujets (instantané figé à la date de l'évaluation), comportant 27 cas de redondance documentés selon quatre niveaux de difficulté. Au calibrage opérationnel (poids de la composante TF-IDF = 0,70 ; seuil = 0,20), la précision atteint 1,00 (IC à 95 % : 0,86–1,00) et le rappel 0,89 (IC à 95 % : 0,72–0,96) ; le F1-score est de 0,94 (IC bootstrap à 95 % : 0,86–1,00). Une recherche systématique sur une grille de 198 couples (poids, seuil) situe le calibrage retenu sur le plateau optimal. La comparaison avec un modèle sémantique multilingue (Sentence-BERT, paraphrase-multilingual-MiniLM-L12-v2) montre un léger gain de rappel (0,93), avec un temps de calcul supérieur d'environ 50 %. Le module de recommandation de sujets alternatifs est considéré comme un prototype fonctionnel. Ces résultats, obtenus sur un corpus synthétique, caractérisent les performances du système en conditions contrôlées et justifient une évaluation ultérieure sur des soumissions réelles.

Mots-clés : plagiat académique ; validation de sujets ; TF-IDF ; indice de Jaccard ; Sentence-BERT ; preuve de concept ; Université de Kinshasa.

Abstract

Manual screening of thesis-topic redundancy and verification of similarities in academic work remain difficult at the Faculty of Sciences and Technologies of the University of Kinshasa because of the lack of tools adapted to the local context. This paper presents AcademiaCheck, a web platform combining two functions: intelligent validation of thesis topics and computer-assisted similarity detection in submitted work. The engine, implemented natively in TypeScript, combines TF-IDF vectorisation, cosine similarity, and the Jaccard index. This proof-of-concept study primarily evaluates topic validation using an anonymised corpus of 83 dossiers (57–110 words) compared with a real catalogue of 114 topics (snapshot at the evaluation date), including 27 documented redundancy cases across four difficulty levels. At the operational calibration (TF-IDF component weight = 0.70; threshold = 0.20), precision reaches 1.00 (95% CI: 0.86–1.00) and recall 0.89 (95% CI: 0.72–0.96); the F1-score is 0.94 (95% bootstrap CI: 0.86–1.00). A systematic grid search over 198 (weight, threshold) pairs places the retained calibration on the optimal plateau. Comparison with a multilingual semantic model (Sentence-BERT, paraphrase-multilingual-MiniLM-L12-v2) shows a slight recall gain (0.93), with approximately 50% higher computation time. The alternative-topic recommendation module is considered a functional prototype. Because the corpus is synthetic, these results characterise system performance under controlled conditions and support subsequent evaluation on real submissions.

Keywords: academic plagiarism; topic validation; TF-IDF; Jaccard index; Sentence-BERT; proof of concept; University of Kinshasa.

* Auteur Correspondant : Moïse Kasombo Tshibangu, (Tshibangumoise16@gmail.com). Tél. : (+243) 817597387

https://orcid.org/0009-0003-5496-4155 ; Reçu le 23/07/2026 ; Révisé le 14/08/2026 ; Accepté le 08/09/2026

DOI : https://doi.org/10.59228/rcst.026.v5.i3.342

1. Introduction

L'enseignement supérieur congolais connaît une croissance soutenue des effectifs qui s'accompagne d'une pression accrue sur la qualité des travaux académiques. Parmi les difficultés signalées par les enseignants, la redondance des sujets de mémoire plusieurs étudiantes et étudiants travaillant, parfois sans le savoir, sur un sujet déjà traité et le plagiat occupent une place particulière. À la Faculté des Sciences et Technologies de l'Université de Kinshasa (FST/UNIKIN), la vérification de l'originalité d'un sujet repose encore sur la mémoire des enseignants et sur des consultations manuelles des archives ; la littérature montre cependant que de telles pratiques manuelles atteignent rapidement leurs limites lorsque le volume de travaux augmente (Park, 2003 ; McCabe & Trevino, 1993).

Les logiciels commerciaux de détection de plagiat existent, mais leur adoption dans les universités congolaises se heurte à trois obstacles documentés : le coût des licences en devise, la dépendance à une connexion Internet stable et l'inadéquation des corpus de référence aux productions académiques locales (Belter & Du Pré, 2009 ; Sutherland-Smith, 2011). En outre, ces outils traitent le plagiat après coup, alors qu'une partie du problème peut être prévenue plus en amont, dès le choix du sujet de mémoire.

C'est dans cette perspective que la plate-forme AcademiaCheck a été conçue et développée pour la FST/UNIKIN. Elle articule deux fonctions complémentaires : (i) la validation intelligente d'un sujet de mémoire, qui compare la proposition de l'étudiante ou de l'étudiant à un catalogue de sujets déjà déposés et signale les redondances avant l'engagement du travail ; (ii) l'aide à la détection de similitudes dans les travaux déposés, fondée sur un pipeline d'analyse automatique. La plate-forme est opérationnelle et déployée : elle tourne en production sur un serveur dédié et son code source est disponible en dépôt ouvert.

La présente étude est conçue comme une preuve de concept. Elle vise à évaluer, dans des conditions contrôlées, la faisabilité technique et les performances du module de validation des sujets d'AcademiaCheck, sans prétendre à une généralisation immédiate aux usages réels. L'évaluation repose sur 83 dossiers comparés à un catalogue réel de 114 sujets, avec estimation des intervalles de confiance à 95 %, calibrage par recherche systématique sur grille et courbes précision-rappel, comparaison à une référence lexicale fondée sur l'indice de Jaccard, comparaison à Sentence-BERT multilingue et évaluation exploratoire du module de recommandation.

La contribution de l'article est triple : (i) décrire une architecture et une implémentation légères adaptées au contexte institutionnel étudié ; (ii) proposer un protocole reproductible d'évaluation du module de validation des sujets, intégrant intervalles de confiance, calibrage et systèmes de comparaison ; (iii) expliciter les limites de validité écologique afin de définir les conditions d'une future évaluation sur des soumissions réelles. La suite de l'article présente le matériel et les méthodes (section 2), les résultats (section 3), la discussion (section 4), puis la conclusion (section 5).

La détection automatique de plagiat a fait l'objet de travaux nombreux depuis les années 2000. La compétition internationale PAN a structuré le champ en définissant des tâches, des corpus et des métriques de référence (Potthast et al., 2010) ; les approches fondées sur la mesure de proximité lexicale et sémantique y ont été systématiquement comparées (Stein & zu Eissen, 2006 ; Gipp & Beel, 2010). Ces travaux ont toutefois été menés pour l'essentiel sur des corpus anglophones de grande taille, éloignés des conditions des universités africaines francophones.

L'évaluation des logiciels commerciaux a par ailleurs révélé des limites importantes : sensibilité variable selon les corpus, faux positifs préjudiciables aux étudiants, opacité des algorithmes propriétaires (Belter & Du Pré, 2009 ; Weber-Wulff, 2014). Ces constats plaident pour des systèmes transparents, explicables et adaptés au contexte local, dont le fonctionnement peut être audité par l'institution elle-même.

Sur le plan sociologique, les travaux consacrés à l'intégrité académique soulignent que la dissuasion ne suffit pas : la perception que les étudiants ont de la règle, la qualité de l'encadrement et la disponibilité d'outils institutionnels conditionnent les comportements (Ashworth et al., 1997 ; Guibert & Michaut, 2011 ; Bergadaà, 2015). Cette perspective justifie le choix de conception d'AcademiaCheck, qui associe le contrôle (détection de redondances) à l'accompagnement (suggestions de sujets alternatifs, rapports explicites).

Enfin, les modèles de plongements de phrases de type Sentence-BERT permettent de représenter la similarité sémantique au-delà du simple recouvrement lexical (Reimers & Gurevych, 2019), avec des

variantes multilingues couvrant le français. Leur coût mémoire et de calcul peut toutefois être plus important pour une infrastructure légère. La comparaison empirique entre l'approche TF-IDF/Jaccard native et un modèle Sentence-BERT multilingue, sur le même corpus et avec mesure des temps de calcul, constitue donc un élément du protocole d'évaluation présenté ici.

2. Matériel et méthodes

2.1. Architecture de la plate-forme

AcademiaCheck est une application Web de type client-serveur développée avec le framework Next.js 16.1.1 et React 19 (exécution Node.js 20.20.2). L'interface du personnel académique compte 31 pages de back-office, dont huit pages d'administration du système, et l'interface étudiante (portail) 16 pages ; les captures des principales pages d'administration figurent en annexe B. L'interface de programmation expose 89 points d'accès REST, regroupés par domaine fonctionnel (authentification, sujets, documents, analyses, statistiques, administration) ; la liste exhaustive en est donnée en annexe A. La persistance repose sur un magasin JSON instrumenté d'écritures atomiques et de sauvegardes automatiques, choix adapté au volume actuel et aux contraintes d'exploitation du serveur. Le contrôle d'accès repose sur un modèle à quatre rôles (super-administrateur, administrateur de faculté, enseignant, étudiant) et sur l'authentification par jetons JWT ; la journalisation d'audit enregistre les actions sensibles. L'ingestion des travaux accepte les fichiers DOCX (bibliothèque mammoth) et les documents numérisés par reconnaissance optique de caractères (tesseract.js 7).

La plate-forme est déployée en production sur un serveur Ubuntu 24.04 géré par le superviseur de processus PM2 derrière un mandataire nginx, avec chiffrement HTTPS. L'environnement logiciel détaillé (versions des bibliothèques et des outils) est consigné en annexe C afin de garantir la reproductibilité des mesures.

2.2. Moteur de détection de similitudes

Le moteur natif (modèle « TF-IDF v2.0 ») applique un pipeline en cinq étapes. (1) Normalisation : suppression des adresses Web et électroniques, des marqueurs de citations numériques et des caractères non latins. (2) Segmentation : découpage du document en segments de 5 à 60 mots, par fusion de phrases. (3) Vectorisation TF-IDF : la pondération associe à chaque terme t d'un segment s le produit $(1 + \ln tf) \times idf$, où tf est la fréquence du terme dans le segment et $idf = \ln((1 + N) / (1 + df)) + 1$, N étant le nombre de segments indexés et df le nombre de segments contenant le terme ; les vecteurs sont normalisés en norme L2. (4) Scoring : la proximité vectorielle entre segments est mesurée par le cosinus des vecteurs TF-IDF, et la proximité lexicale par l'indice de Jaccard sur les ensembles de termes ; pour chaque segment de la requête, les trois meilleurs correspondants au-dessus du seuil de correspondance $\theta = 0{,}15$ sont retenus. (5) Classification et agrégation : chaque correspondance est typée en cinq classes (copie-coller, paraphrase, reformulation, traduction, similarité faible) selon des bornes sur les scores vectoriel et lexical, et un score global est calculé comme la moyenne pondérée de la couverture (part des segments associés, poids 0,7) et du score moyen des correspondances (poids 0,3).

Cette implémentation ne dépend d'aucune bibliothèque externe d'apprentissage automatique : elle est exécutée entièrement côté serveur Node.js, ce qui garantit des temps de réponse courts et une empreinte mémoire minimale.

2.3. Validation intelligente de sujets

La fonction de validation de sujets compare un dossier de candidature — titre, description, domaine, mots-clés, objectifs et problématique concaténés — à l'ensemble des sujets du catalogue. Chaque sujet du catalogue reçoit un score combiné $S = w \cdot \cos(q, s) + (1 - w) \cdot jac(q, s)$, où cos désigne la similarité cosinus entre vecteurs TF-IDF, jac l'indice de Jaccard lexical et w le poids accordé à la composante vectorielle TF-IDF. Le dossier est déclaré redondant si le score combiné maximal atteint ou dépasse le seuil de décision τ ; le calibrage opérationnel retenu est $w = 0{,}70$ et $\tau = 0{,}20$. Un rapport explicite est produit : sujet le plus proche, score, mots-clés partagés, justification textuelle. En cas de redondance détectée, un module de recommandation génère jusqu'à six formulations alternatives par gabarits à partir des mots-clés du dossier soumis ; ce module, de conception simple, fait l'objet d'une évaluation spécifique (section 3.5).

2.4. Corpus d'évaluation

Conformément à la demande d'extension du corpus d'évaluation, un corpus de référence a été constitué. Il comprend : (i) le catalogue réel de la plate-forme, soit 114 sujets académiques effectivement enregistrés à la FST/UNIKIN à la date de l'évaluation, couvrant 61 domaines (intelligence artificielle et apprentissage automatique, bases de données réparties, exploration de données, génie logiciel, gestion, sécurité informatique, etc.) ; (ii) un corpus d'évaluation de 83 dossiers de sujets (identifiants S001 à S083), longueur 57 à 110 mots (moyenne 88 mots), conformes au

format réellement soumis par les utilisateurs. L'effectif de 114 sujets correspond à l'instantané figé de la base au moment de l'évaluation : le catalogue de production étant alimenté en continu, il compte désormais environ 1 000 sujets, et l'évaluation a été volontairement réalisée sur cet instantané afin de garantir la reproductibilité des mesures.

Les 27 cas de redondance ont été construits par réécriture contrôlée de sujets du catalogue, selon quatre niveaux de difficulté croissante : huit copies adaptées (reprise quasi conforme avec variations cosmétiques), dix paraphrases légères (réorganisation des phrases, synonymie partielle), huit paraphrases profondes (même objet et même problématique, vocabulaire différent) et un assemblage de deux sources (patchwork). Les 56 dossiers restants sont des sujets originaux élaborés hors du catalogue, dans des domaines variés du contexte congolais (santé, agriculture urbaine, mobilité, énergie, administration, éducation). La vérité terrain est donc établie par construction : chaque dossier est étiqueté redondant ou original, avec référence au(x) sujet(s) source(s) le cas échéant. Les textes ont été générés par réécriture assistée par un modèle de langage, puis revus manuellement ; le corpus anonymisé est documenté en annexe D et sera déposé dans un entrepôt ouvert (Zenodo) accompagné du code d'évaluation.

Limite de validité écologique : ce corpus est synthétique. Il ne provient pas de soumissions réelles d'étudiantes et d'étudiants, et la distribution des difficultés y est imposée, avec une proportion de redondances plus élevée que celle qui pourrait être observée en usage réel. Les résultats doivent donc être interprétés comme une caractérisation contrôlée du comportement du moteur, et non comme une estimation de ses performances en production. Une évaluation ultérieure sur des soumissions réelles anonymisées est nécessaire pour apprécier sa validité écologique.

2.5. Protocole d'évaluation

La tâche évaluée est la classification binaire de chacun des 83 dossiers : « redondant » si le score combiné maximal sur le catalogue atteint le seuil, « original » sinon. Les métriques rapportées sont la précision, le rappel, le F1-score et l'exactitude. Cinq instruments complètent ce socle :

- Intervalles de confiance à 95 %, calculés par la méthode de Wilson pour la précision et le rappel, et par bootstrap (2 000 rééchantillonnages) pour le F1-score ;
- Recherche systématique sur grille de 198 couples ($w \in \{0{,}0 \ldots 1{,}0\}$, pas de 0,1 ; $\tau \in \{0{,}05 \ldots 0{,}90\}$, pas de 0,05), afin de justifier empiriquement le calibrage ;
- Courbes précision-rappel obtenues par variation continue du seuil de décision ;
- Test exact de McNemar sur les décisions discordantes entre le système hybride et la référence lexicale Jaccard pure ($w = 0$), complété par un bootstrap apparié de la différence de F1 ;
- Comparaison avec Sentence-BERT multilingue (paraphrase-multilingual-MiniLM-L12-v2, 384 dimensions, similarité cosinus pure), sur le même corpus, avec mesure des temps de calcul par requête.

Les temps de calcul du moteur natif ont été mesurés sur une réimplémentation fidèle du pipeline en Python (mêmes règles de segmentation, de pondération et de scoring), exécutée sur un poste de travail standard ; ils sont comparés à ceux du modèle Sentence-BERT dans la section 3.4. Le module de recommandation a été évalué sur dix cas de redondance détectés : pour chaque alternative générée, la similarité maximale vis-à-vis du catalogue a été calculée, une alternative étant considérée comme non redondante selon le moteur lorsqu'elle demeurait sous le seuil de décision.

2.6. Considérations éthiques

Le dispositif est conçu comme un outil d'aide à la décision et de pédagogie de l'intégrité académique, non comme un dispositif de décision automatisée : tout signalement reste soumis à l'appréciation des enseignants. L'évaluation décrite dans cette étude utilise des dossiers synthétiques anonymes (S001 à S083) et un catalogue de référence constitué d'énoncés de sujets académiques. La transparence du dispositif repose sur des rapports explicables (sujets les plus proches, scores, mots-clés partagés), la description du protocole d'évaluation et la disponibilité annoncée du code source. Le règlement intérieur de la Faculté et les recommandations de l'UNESCO sur l'éthique de l'intelligence artificielle sont cités comme cadres de référence (UNESCO, 2023).

3. Résultats

3.1. Performances au calibrage opérationnel

Au calibrage opérationnel (w = 0,70; τ = 0,20), sur les 83 dossiers du corpus, le système identifie correctement 24 des 27 cas de redondance et ne signale aucun dossier original à tort. La précision atteint 1,0 (IC à 95 %: 0,862–1,000), le rappel 0,889 (IC à 95 %: 0,719–0,961), le F1-score 0,941 (IC bootstrap à 95 %: 0,857–1,000) et l'exactitude 0,964. Les trois cas non détectés sont des paraphrases profondes, conformément à la difficulté attendue de cette catégorie.

Tableau I. Performances au calibrage opérationnel

Métrique	Valeur	IC à 95 %	Effectifs
Précision	1,000	0,862 – 1,000	VP = 24 ; FP = 0
Rappel	0,889	0,719 – 0,961	FN = 3 (paraphrases profondes)
F1-score	0,941	0,857 – 1,000 (bootstrap)	n = 83
Exactitude	0,964	—	VP = 24 ; VN = 56

Note : VP, FP, VN, FN : vrais et faux positifs, vrais et faux négatifs. Calibrage opérationnel : w = 0,70 ; τ = 0,20. IC calculés par méthode de Wilson (précision, rappel) et bootstrap 2 000 (F1).
Source : corpus d'évaluation AcademiaCheck (83 dossiers, catalogue de 114 sujets), 2026 ; calculs des auteurs.

Type de cas	Effectif	Détectés	Rappel
Copie adaptée	8	8	1,000
Paraphrase légère	10	10	1,000
Paraphrase profonde	8	5	0,625
Assemblage (patchwork)	1	1	1,000

Note : lecture — sur les 8 paraphrases profondes, 5 sont détectées au seuil opérationnel.
Source : corpus d'évaluation Academia Check, 2026 ; calculs des auteurs.

3.2. Recherche sur grille et justification du calibrage

La recherche systématique sur les 198 couples (w ; τ) définis dans le protocole montre que le F1-score maximal (0,941) est atteint à w* = 0,1 et τ* = 0,15. Un plateau de dix couples atteint le même F1-score maximal, pour des valeurs de w allant de 0,1 à 1,0 et des seuils voisins de 0,10 à 0,25. Le couple opérationnel (0,70 ; 0,20) appartient à ce plateau. Sur ce corpus, l'augmentation du poids de la composante vectorielle TF-IDF n'améliore donc pas le F1-score maximal par rapport aux autres couples du plateau. Le calibrage opérationnel est ainsi compatible avec la zone de performance maximale observée sur la grille, tandis que les figures 1 et 2 décrivent sa sensibilité au seuil.

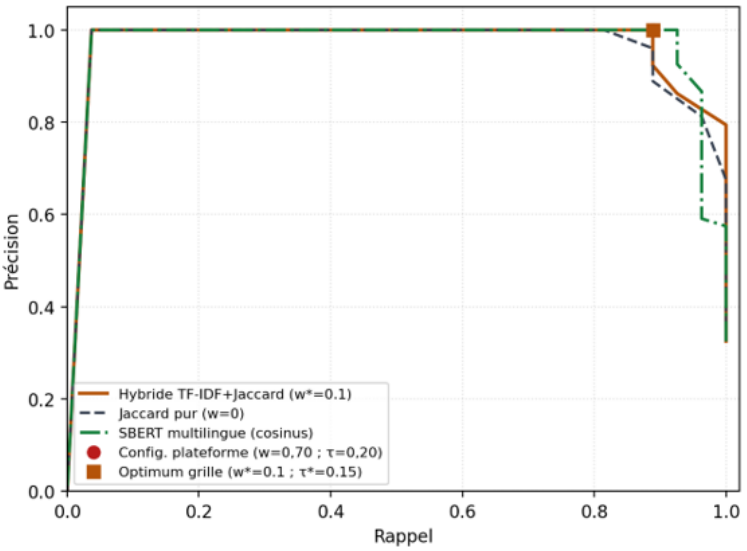

Figure 1. Courbes précision-rappel du système hybride (w = 0,1), de la référence Jaccard pure (w = 0) et de Sentence-BERT multilingue ; points de fonctionnement du calibrage opérationnel et de l'optimum de grille.*
Source : corpus d'évaluation AcademiaCheck, 2026 ; calculs des auteurs.

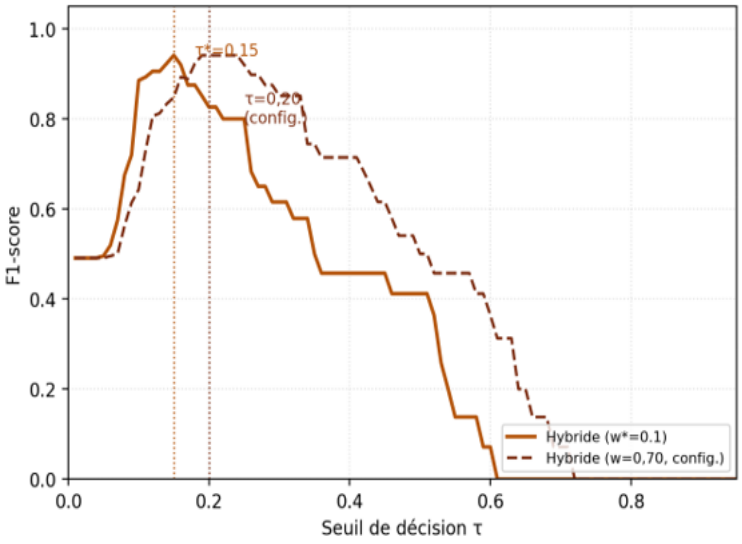

Figure 2. F1-score en fonction du seuil de décision τ, pour w = 0,1 et pour le poids opérationnel w = 0,70.*
Source : corpus d'évaluation AcademiaCheck, 2026 ; calculs des auteurs.

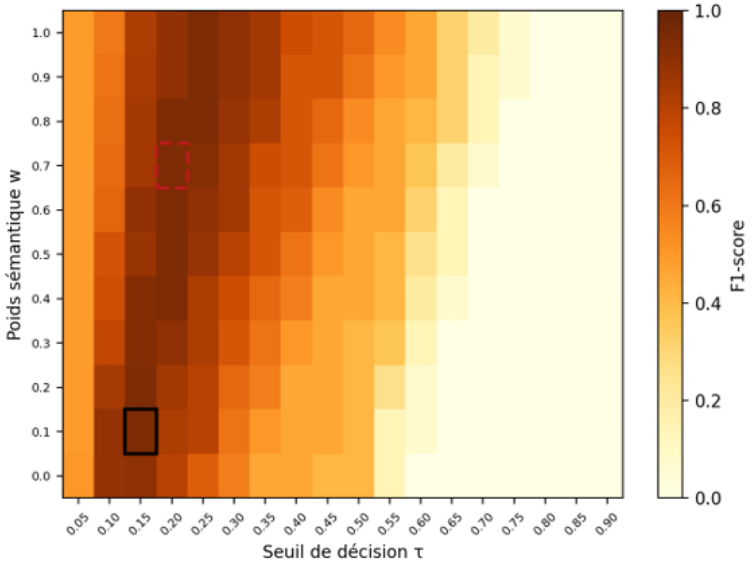

Figure 3. Carte de chaleur du F1-score sur la grille (w ; τ). Le rectangle plein marque l'optimum de grille ; le rectangle pointillé, le calibrage opérationnel, situé sur le plateau.

Source : corpus d'évaluation AcademiaCheck, 2026 ; calculs des auteurs.

3.3. Comparaison avec la référence lexicale Jaccard

À son propre optimum (τ = 0,13), la référence lexicale Jaccard pure atteint un F1-score de 0,923 (précision 0,960, rappel 0,889), contre 0,941 pour le système hybride à son optimum. Sur les décisions discordantes entre les deux systèmes, le test exact de McNemar donne une valeur p de 1,000 (un seul couple discordant) ; le bootstrap apparié de la différence de F1 donne un intervalle à 95 % de [0,000 ; 0,061]. Sur ce corpus à fort recouvrement lexical, la combinaison TF-IDF/Jaccard n'apporte donc pas de gain statistiquement démontré du F1-score par rapport à la référence Jaccard pure. Ce résultat limite l'interprétation du bénéfice du système hybride et devra être réévalué sur des corpus écologiquement plus représentatifs.

3.4. Comparaison avec Sentence-BERT multilingue

Le modèle Sentence-BERT multilingue (paraphrase-multilingual-MiniLM-L12-v2, 384 dimensions, similarité cosinus pure) atteint, au seuil optimal τ = 0,78, un F1-score de 0,962 (précision 1,000, IC à 95 % : 0,867–1,000 ; rappel 0,926, IC à 95 % : 0,766–0,979). Il détecte 25 des 27 cas de redondance, contre 24 pour le moteur natif. Sur ce corpus, ce résultat correspond à un gain absolu de rappel d'environ 0,037. Le temps moyen rapporté est d'environ 63 ms par requête, contre 42 ms pour la réimplémentation de référence du pipeline natif, soit un temps environ 1,5 fois plus élevé. S'y ajoutent une empreinte mémoire du modèle annoncé à plusieurs centaines de mégaoctets et un temps d'indexation initial de 3,6 s pour le catalogue. Ces résultats suggèrent un compromis entre performance et ressources qui devra être confirmé dans l'environnement de production.

Tableau II. Comparaison avec Sentence-BERT multilingue

Système	F1 (opt.)	Précision	Rappel	Temps par requête	Contraintes
Moteur natif TF-IDF + Jaccard (w* ; τ*)	0,941	1,000	0,889	≈ 42 ms	Empreinte minimale ; sans dépendance externe
Référence Jaccard pure (τ = 0,13)	0,923	0,960	0,889	≈ 40 ms	Pas de typologie des correspondances
Sentence-BERT multilingue (τ = 0,78)	0,962	1,000	0,926	≈ 63 ms	Modèle de 470 Mo ; indexation 3,6 s

Note : temps mesurés sur la même machine, corpus identique (83 requêtes contre 114 sujets) ; chaque système est évalué à son seuil optimal.
Source : mesures des auteurs, 2026.

3.5. Module de recommandation: statut de prototype fonctionnel

Dix cas de redondance détectés ont été soumis au générateur d'alternatives, qui a produit 60 formulations, soit jusqu'à six par cas. La vérification interne, fondée sur la similarité maximale de chaque alternative vis-à-vis du catalogue, montre que 8 formulations sur 60 (13,3 %) demeurent sous le seuil de décision. Ainsi, 52 formulations sur 60 sont encore classées comme redondantes par le moteur. Le module doit donc être considéré comme un prototype fonctionnel: ses suggestions sont indicatives, nécessitent une validation humaine et ne garantissent pas l'originalité du sujet reformulé. L'amélioration de la génération d'alternatives, associée à une boucle de vérification automatique de non-redondance, constitue une perspective de développement.

3.6. Validation fonctionnelle en production

Au-delà de l'évaluation contrôlée, la plate-forme est déployée sur un serveur de production (https://plagiat.hpph.net) avec supervision PM2 et sauvegardes automatiques du magasin de données. Un parcours fonctionnel de bout en bout est rapporté

comme vérifié: inscription, connexion, validation d'un sujet avec production du rapport et dérivation du sujet validé en projet de mémoire. Le tableau de bord d'administration affiche l'état du moteur (modèle TF-IDF v2.0, seuil actif de 0,15, classification en cinq types, temps d'analyse affiché d'environ 120 ms) et les statistiques d'usage (annexe B). La période d'observation et la méthode de calcul d'une estimation chiffrée de la disponibilité du service ne sont toutefois pas encore documentées; une telle évaluation en production reste à mener.

4. Discussion

4.1. Interprétation des résultats

Trois enseignements se dégagent. Premièrement, aucun faux positif n'est observé parmi les 56 dossiers originaux au calibrage opérationnel, ce qui conduit à une précision estimée à 1,00; compte tenu de la taille de l'échantillon, son IC à 95 % (0,86–1,00) doit cependant être pris en compte. Deuxièmement, les trois faux négatifs appartiennent à la catégorie des paraphrases profondes, dont 5 cas sur 8 sont détectés ; Sentence-BERT détecte au total un cas positif supplémentaire. Troisièmement, plusieurs couples de paramètres atteignent le même F1-score maximal sur la grille étudiée, ce qui indique que le calibrage opérationnel appartient à une zone de performance stable sur ce corpus contrôlé, sans démontrer pour autant une robustesse identique en conditions réelles.

4.2. Limites

Les limites de l'étude en restreignent la portée. Premièrement, le corpus est synthétique, avec une proportion de redondances imposée et des textes produits par réécriture assistée puis revus manuellement ; les distributions réelles de difficulté et de style peuvent donc différer. Deuxièmement, avec 27 cas positifs, les intervalles de confiance à 95 % demeurent relativement larges (précision 0,86–1,00; rappel 0,72–0,96), ce qui limite les sous-analyses par type de cas.

Troisièmement, l'avantage du système hybride sur Jaccard pur n'est pas statistiquement démontré pour le F1-score sur ce corpus. Quatrièmement, les temps de calcul du moteur natif proviennent d'une réimplémentation Python fidèle du pipeline et non d'une mesure directe du code TypeScript en production ; ils doivent donc être interprétés comme des mesures comparatives dans l'environnement expérimental. Enfin, l'étude n'évalue pas la performance diagnostique du module d'analyse de plagiat sur des documents Couplets; ses résultats quantitatifs portent principalement sur la détection de redondance entre propositions de sujets.

4.3. Implications opérationnelles

Pour l'institution, les résultats permettent d'envisager un usage du système comme aide au tri et à la revue des propositions de sujets, sous supervision humaine. Le seuil de décision peut être ajusté en fonction du compromis recherché entre faux positifs et faux négatifs, mais les effets précis d'un abaissement ou d'une élévation du seuil doivent être lus directement sur les courbes précision-rappel du corpus étudié. L'association d'un signalement explicite — sujets les plus proches, scores et mots-clés partagés — et d'un module de recommandation encore expérimental s'inscrit dans une démarche d'accompagnement plutôt que de sanction (Sutherland-Smith, 2011; Löfström et al., 2017). L'activation éventuelle de Sentence-BERT devra tenir compte du gain observé et du coût en ressources dans l'environnement réel de déploiement.

5. Conclusion

Cet article présente AcademiaCheck, une plate-forme Web de validation intelligente des sujets de mémoire et d'aide à la détection de similitudes, développée pour la Faculté des Sciences et Technologies de l'Université de Kinshasa. Dans cette preuve de concept, l'évaluation quantitative porte principalement sur la détection de redondance entre 83 dossiers de sujets et un catalogue réel de 114 sujets.

Au calibrage opérationnel, le moteur TF-IDF/Jaccard atteint un F1-score de 0,94 (IC bootstrap à 95 %: 0,86–1,00), une précision de 1,00 et un rappel de 0,89. Le calibrage appartient au plateau de performance maximale observé sur la grille de 198 couples testés. L'avantage du système hybride sur Jaccard pur n'est pas statistiquement démontré sur ce corpus, tandis que Sentence-BERT atteint un rappel de 0,93 avec un temps de calcul environ 1,5 fois plus élevé dans l'environnement expérimental. Le module de recommandation demeure un prototype fonctionnel. Ces résultats soutiennent la faisabilité du dispositif, mais ne permettent pas encore de conclure à ses performances sur des soumissions réelles ni à l'efficacité du module de détection du plagiat sur des documents complets.

Les travaux futurs devraient prioritairement porter sur l'évaluation du système à partir de soumissions réelles anonymisées, selon un protocole éthique et institutionnel approprié, puis sur la mesure directe des performances du module d'analyse de similitudes dans des documents complets. L'intégration

optionnelle de Sentence-BERT, l'amélioration du générateur de sujets alternatifs, l'extension à plusieurs établissements et l'adaptation linguistique aux corpus multilingues du contexte congolais pourront ensuite être évaluées sur des données représentatives.

Remerciements

Nous tenons à remercier l'Université de Kinshasa et la Mention Mathématiques, Statistiques et Informatique de la Faculté des Sciences et Technologies pour l'appui technique dans la réalisation de cette étude.

Considérations Ethiques

La présente étude a été conduite dans le respect des principes d'éthique.

Financement

Cette recherche n'a bénéficié d'aucun financement externe dédié; les coûts d'infrastructure ont été pris en charge par l'auteur.

Conflits d'intérêt

Il y'aucun conflit d'intérêts.

Orcid de l'auteur

Tshibangu M.K : https://orcid.org/0009-0003-5496-4155

Références bibliographiques

Ashworth, P., Bannister, P., & Thorne, P. (1997). Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment. *Studies in Higher Education, 22*(2), 187–203.

Belter, C., & Du Pré, A. (2009). A multi-institutional assessment of plagiarism detection software. *The Journal of Academic Librarianship, 35*(6), 554–558.

Bergadaà, M. (2015). *Le plagiat académique: Comprendre pour agir*. L'Harmattan.

Gipp, B., & Meuschke, N. (2011). Citation pattern matching algorithms for citation-based plagiarism detection: Greedy citation tiling, citation chunking and longest common citation sequence. In *Proceedings of the 11th ACM Symposium on Document Engineering* (pp. 249–258). ACM. https://doi.org/10.1145/2034691.2034741

Guibert, P., & Michaut, C. (2011). Le plagiat étudiant. *Éducation et sociétés, 28*, 149–164.

Löfström, E., Huotari, J., & Kupila, P. (2017). Plagiarism detection software as a pedagogical tool. *Assessment & Evaluation in Higher Education, 42*(1), 1–14.

McCabe, D. L., & Treviño, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education, 64*(5), 522–538.

Park, C. (2003). In other (people's) words: Plagiarism by university students—Literature and lessons. *Assessment & Evaluation in Higher Education, 28*(5), 471–488.

Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P. (2010). An evaluation framework for plagiarism detection. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)* (pp. 997–1005).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 2019 International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Stein, B., & Meyer zu Eissen, S. (2006). Near similarity search and plagiarism analysis. In *From data and information analysis to knowledge engineering* (pp. 430–437). Springer.

Sutherland-Smith, W. (2011). How plagiarism detection software may help learning. *International Journal for Educational Integrity, 7*(1), 23–35.

UNESCO. (2023). *Recommandation sur l'éthique de l'intelligence artificielle*. UNESCO.

Université de Kinshasa. (2023). *Règlement intérieur de la Faculté des Sciences et Technologies*. UNIKIN.

Weber-Wulff, D. (2014). *False feathers: A perspective on academic plagiarism*. Springer.

Documentation technique: Next.js (https://nextjs.org/docs) ; React (https://react.dev) ; Node.js (https://nodejs.org/docs) ; tesseract.js (https://tesseract.projectnaptha.com) ; mammoth (https://github.com/mwilliamson/mammoth.js) ; Sentence-Transformers (https://www.sbert.net) ; PAN — Plagiarism, Authorship, and Social Software Misuse (https://pan.webis.de).