

PROJECT

Engineering 4 - ISI

INF4268 - Advanced algorithms and data structures - 5 crédits

Dedline : Thursday 16 January 2024 at 4 p.m.

ADJEW A Frédéric frederic.adjewa@institutsaintjean.org

KACFAH Cheikh cheikh.kacfah@institutsaintjean.org

School year 2024-2025 ; 1st Term

The aim of this project is to apply the concepts studied in course INF4268-Advanced Algorithmics and data structures to solve a concrete problem. It is therefore not permitted to use *Machine learning* algorithms. The work is to be carried out in groups of 2 or 3 students. **The project will count for 40% of the final mark.**

Description

Plagiarism is a growing problem in various fields, particularly in academia, where students and researchers sometimes appropriate the work of their peers. This act can include a variety of sources, such as books, scientific articles, journals, source code, blogs, online publications and peer assignments. Detecting plagiarism manually is a complex and unreliable task, requiring the development of automated methods capable of carrying out this detection efficiently.

The aim of this project is to design an application enabling any user to identify whether, among a set of documents (two **or more**) submitted to the system, any of them have similarities that could be considered plagiarism.

Goals

As part of this project, you will :

- ✓ Propose a structured method for solving the plagiarism detection problem, providing clear and precise explanations.
- ✓ Write the detailed pseudo-code for each proposed algorithm and rigorously analyse their performance.
- ✓ Develop a hosted application that allows the solution to be tested in a practical and user-friendly way.

It should be noted that we are dealing here with two main cases :

- 2 documents as input : which allows you to tell to what extent one document is 'inspired' by the other
- n input documents : This will allow documents to be grouped by similarity.

Your solution must not only focus on detection accuracy, but also take into account aspects such as speed of execution, scalability and adaptability to various document formats.

This work must be submitted in a report written in \LaTeX . The deadline is Thursday 16 January 2024 at 4 p.m. sharp. Any report handed in after the deadline will be given a mark of 0/20. Please take this seriously. Reports must be submitted exclusively via Moodle.

PROJET

Ingénierie 4 - Informatique et Systèmes d'Information INF4268 - Algorithmique et structures de données avancées - 5 crédits

Délai : Jeudi 16 janvier 2024 à 16 heures

ADJEWA Frédéric frederic.adjewa@institutsaintjean.org

KACFAH Cheikh cheikh.kacfah@institutsaintjean.org

Année académique 2024-2025 ; Semestre I

Le but de ce projet est de mettre en application les concepts étudiés dans le cadre de l'UE INF4268-Algorithmique et structures de données avancées pour résoudre un problème concret. Il n'est donc pas permis d'utiliser des algorithmes de *Machine learning*. Le travail est à réaliser par groupe de 2 ou 3 personnes. **Le projet comptera pour 40% de la note finale.**

Description

Le plagiat constitue un problème croissant dans divers domaines, en particulier dans le milieu académique, où étudiants et chercheurs s'approprient parfois les propriétés intellectuelles de leurs pairs. Cet acte peut inclure des sources variées, telles que des livres, articles scientifiques, journaux, codes sources, blogs, des publications en ligne ou encore des devoirs de camarades. Détecter le plagiat manuellement est une tâche complexe et peu fiable, nécessitant le développement de méthodes automatisées capables d'effectuer cette détection de manière efficace.

Ce projet vise à concevoir une application permettant à tout utilisateur d'identifier si, parmi un ensemble de documents (deux **et** plus) soumis au système, certains présentent des similarités susceptibles d'être considérées comme du plagiat.

Objectifs du projet

Dans le cadre de ce projet, vous devez :

- ✓ Proposer une méthode structurée pour résoudre le problème de détection de plagiat, en fournissant des explications claires et précises.
- ✓ Rédiger le pseudo-code détaillé de chaque algorithme proposé et analyser rigoureusement leur performance.

- ✓ Développer une application hébergée qui permet de tester la solution de manière pratique et conviviale

Notons qu'il est question ici de gérer deux principaux cas :

- 2 documents en entrée : Qui permet de dire à quel point un document est "inspiré" de l'autre
- n documents en entrée : Qui permettra grouper les documents par similarité.

Votre solution doit non seulement se concentrer sur la précision de la détection, mais également prendre en compte des aspects tels que la rapidité d'exécution, l'évolutivité, et l'adaptabilité à divers formats de documents.

Ce travail doit être remis dans un rapport rédigé en \LaTeX . Pour rappel, le délai est fixé au jeudi 16 janvier 2024 à 16h heures précises. Tout rapport rendu hors délai vaudra une note de 0/20. Mettez-y tout le sérieux nécessaire. Le rendu se fait exclusivement par Moodle.