

077 BCT

PLAGIARISM CHECKER

Final Defense

Avahan Tamrakar (015)

Javed Ansari (033)

Kripesh Nihure (037)

• • • •
• • • •
• • • •
• • • •

Content

01

Introduction

02

Literature Review

03

Methodology

04

Experimental Setup & System Design

05

Results & Output

06

Limitations & Future Enhancements



The background image shows a modern office space with a focus on greenery. Large, leafy plants are integrated into the ceiling structure and placed throughout the room. The furniture includes light-colored wooden desks and chairs. In the background, there are glass-walled rooms labeled 752, 753, and 756, and a sign for 'REIDIN'.

INTRODUCTION

Problem statement:

- Obviously, students try to cheat
- Teachers understandably can't check everything

Objectives:

- Create a web based UI for checking plagiarism
- Mathematical based implementation to check similarity
- Check for plagiarism within the class and if detected, point out the extent of similarity
- show files side-by-side highlighting similarities

The background image shows a modern office environment. Large, lush green plants are suspended from the ceiling and integrated into the furniture, creating a natural and airy atmosphere. The room features light-colored wooden desks and chairs. A glass partition wall is visible in the background, with a sign that reads "752 Digital Pals". The ceiling has exposed pipes and a series of circular pendant lights.

LITERATURE REVIEW

Related work:

- Traditional method: manually compare with eye
- Automated tools:
 - **Turnitin** and **Grammarly** (Algorithm based)
 - MOSS (Measure Of Software Similarity)
 - Draftable: It is an enterprise-grade comparison software tool. It compares documents, side-by-side highlighting differences between documents

Challenges in existing solutions:

- There exist similar software to ours but they do not achieve exactly what we aim to.
- our requirement: check **intra-class** plagiarism in a google classroom-like environment. We deal with uploaded assignments.

Related Theory:

- Text extraction
- Tokenization
- Vectorization
 - N-grams {'ai or', 'or artificail'} [n=2]
 - when [n=1] it is essentially Bag of Words (**BoW**)

Cosine similarity:

measure of similarity between two non-zero vectors. cosine of the angle between them. focuses on the direction, not the magnitude, of the vectors. vectors A and B are of **n dimension**

Related Theory:

Cosine similarity:

measure of similarity between two non-zero vectors. cosine of the angle between them. focuses on the direction, not the magnitude, of the vectors. vectors A and B are of **n dimension**

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

The background image shows a modern office space with a high ceiling featuring exposed pipes and ductwork. Large windows on the right side provide natural light. The walls are covered in lush green plants and vines. In the foreground, there are several wooden desks with black office chairs, some of which have small potted plants on them. A large, comfortable sofa is visible in the background. The overall atmosphere is bright and airy.

METHODOLOGY

Requirements

Functional Requirements:

- User Authentication
- User friendly interface
- Plagiarism detection
- Assignment upload

Non-Functional Requirements:

- Security
- Scalability
- Usability
- Performance

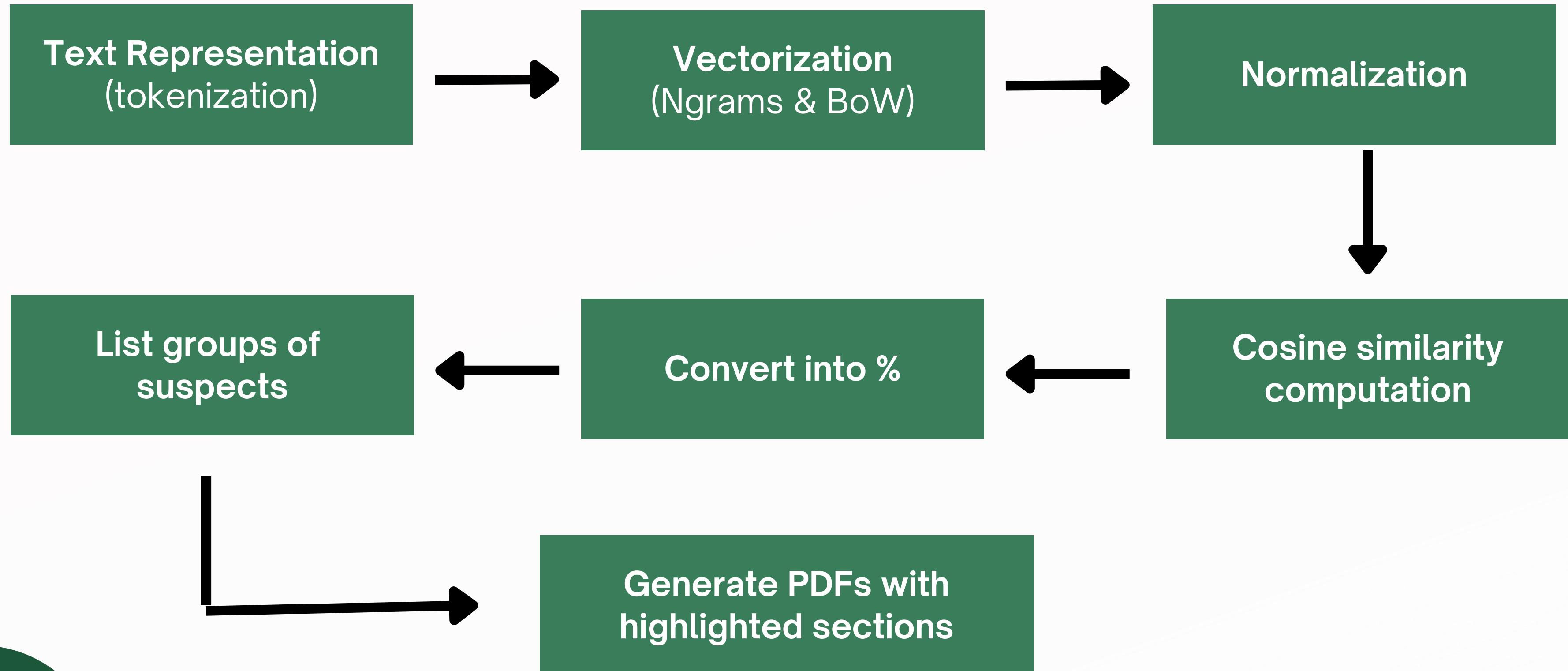
system design

- Frontend
 - create & join class, create assignment, upload assignment, view results
- Backend
 - document processing, **similarity analysis**, and data storage
 - user authentication

system design implementation

- Frontend
 - React, JS, HTML, CSS
- Backend
 - Python , Flask, Firebase
 - various python libraries like PyPDF2, PyMuPDF, itertools etc
 - API: Axios
 - Database: firestore, local storage

Plagiarism Detection Algorithm



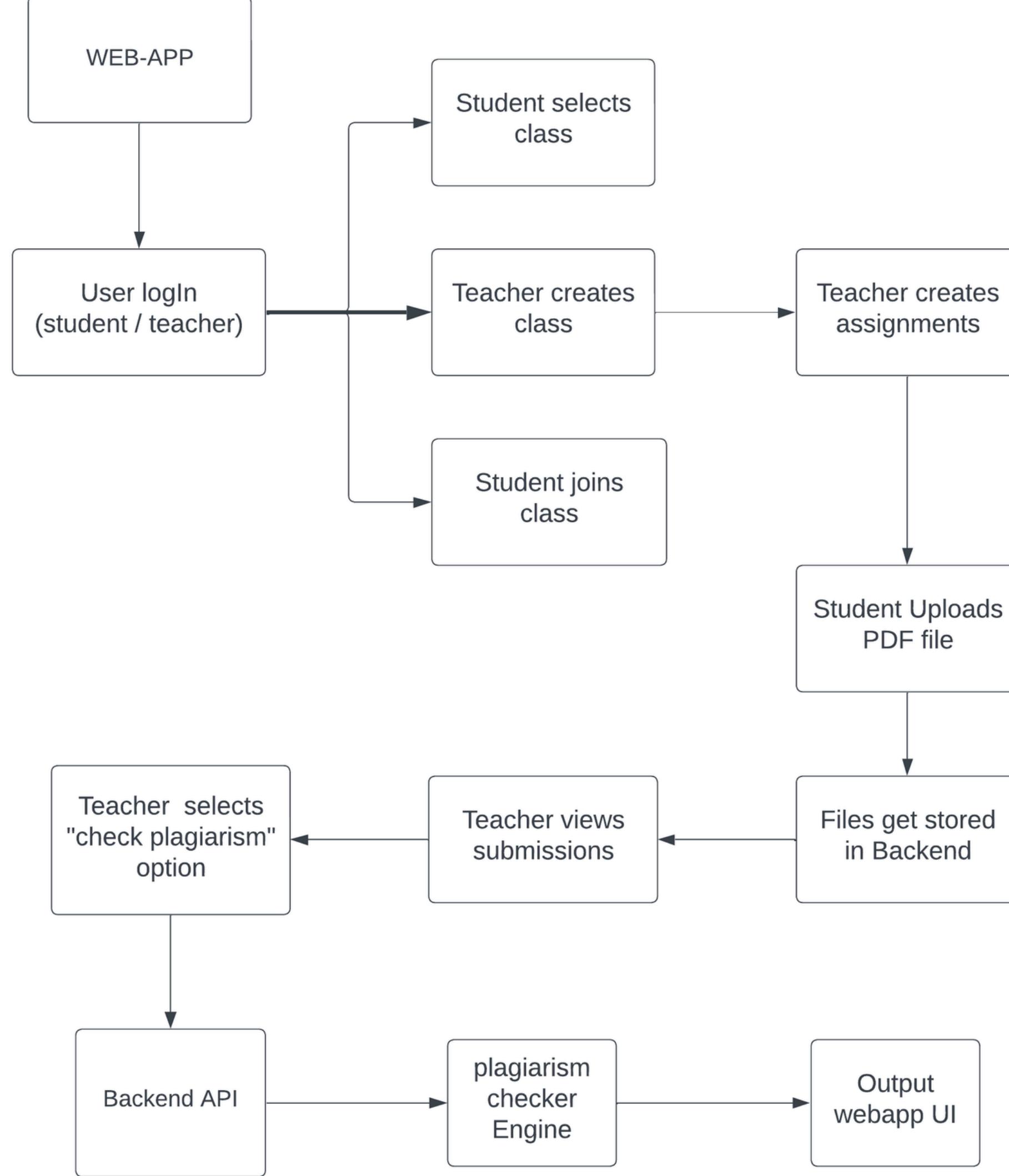


EXPERIMENTAL SETUP & SYSTEM DESIGN

Experimental setup

- Data selection
 - set of known plagiarized and non-plagiarized document pairs
- Testing
 - True Positive (TP): no. of document pairs correctly identified as plagiarized by the system.
 - False Positive (FP): no. of document pairs incorrectly identified as plagiarized by the system.
 - True Negative (TN): no. of document pairs correctly identified as non-plagiarized by the system.
 - False Negative (FN): no. of document pairs incorrectly identified as non-plagiarized by the system.

system design



The background image shows a modern office space with a high ceiling featuring exposed pipes and ductwork. Large green plants are integrated throughout the room, including hanging plants from the ceiling and large potted plants on the floor. There are several wooden desks with black office chairs, and a large sofa area in the background. The overall aesthetic is bright and natural.

RESULTS & OUTPUT

Performance:

File1: Javed_crypt_arithmatic.pdf

File2: output.pdf

Similarity Between Them: 61.59%

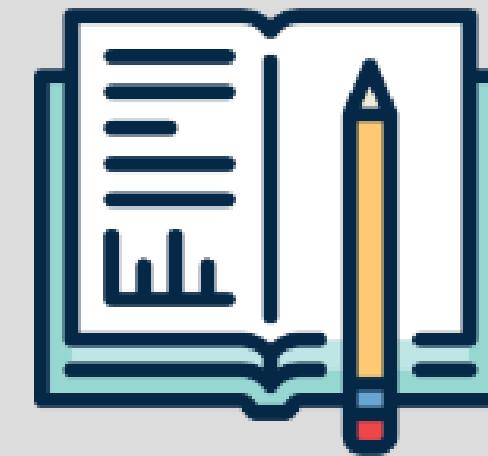
Time Taken For The Comparision: 1.95 Seconds

Please wait. Highlighting "Javed_crypt_arithmatic.pdf" Document.

Time Taken to Highlight: 18.253 Seconds.

Please wait. Highlighting "output.pdf" Document.

Time Taken to Highlight: 16.069 Seconds.



admin@gmail.com

.....

Login

Login with Google

Don't have an account? [Register](#) now.



Classroom

AI

Basanta Joshi



DBMS

Aman Shakya



OS

Santosh Giri



AI

Basanta Joshi

Class Id: DErfYoniftXifT1USY92

Announcement Id: NLsnoRpsSHal8PVAqqW39aV8uRS2

check Plagiarism

List of groups of plagiarized

(077BCT037)Ram_Krishna_Subedi.pdf

(077BCT033)Krishna_Kattel.pdf

[View pdfs](#)

Filename: (077BCT033)Krishna_Kattel.pdf

[View submission](#)

Filename: (077BCT037)Ram_Krishna_Subedi.pdf

[View submission](#)

Class Id: DErfYoniftXifT1USY92

Announcement Id: 9X2BU29ihaM7zFpLZqHOA4ZwiUD3

 No Plagiarism found.

(077BCT033)Krishna_Kattel.pdf

[View pdf](#)

(077BCT037)Ram_Krishna_Subedi.pdf

[Close X](#)

[◀-- Previous](#)

[Next -->](#)



1 / 1

-

+



What is Artificial Intelligence (AI)?

In simple words, Artificial Intelligence (AI) refers to the development of intelligent machines that can perform tasks that usually require human intelligence. Some examples include visual perception, speech recognition, decision-making and language translation.

Importance of Artificial Intelligence

◀-- Previous

Next --►



files/9X2BU29ihM7zF...

1 / 1



124%



What is Artificial Intelligence (AI)?

In simple words, Artificial Intelligence (AI) refers to the development of intelligent machines that can perform tasks that usually require human intelligence. Some examples include computer vision, natural language processing, perception, speech recognition, decision-making and language translation.

Importance of Artificial Intelligence

The significance of Artificial Intelligence (AI) cannot be overstated in today's world. AI has immense potential to revolutionize industries, enhance efficiency, and improve lives. Advanced AI technologies enable automation of tasks that were once time-consuming or impossible for humans to accomplish, leading to increased productivity and cost savings across various sectors. AI-driven insights derived from analyzing massive datasets can inform decision-making processes in fields such as healthcare, finance, and business, leading to more informed and efficient operations.

List of groups of plagiarized

(077BCT037)Ram_Krishna_Subedi.pdf

(077BCT033)Krishna_Kattel.pdf

[View pdfs](#)

Submitted by: Ram Krishna Subedi (077BCT037)

What is Artificial Intelligence (AI)?

In simple words, Artificial Intelligence (AI) refers to the development of intelligent machines that can perform tasks that usually require human intelligence. Some examples include visual perception, speech recognition, decision-making and language translation.

Importance of Artificial Intelligence

The significance of Artificial Intelligence (AI) cannot be overstated in today's world, as it holds immense potential to revolutionize industries, enhance efficiency, and improve quality of life. AI technologies enable automation of tasks that were once time-consuming or impossible for humans to accomplish, leading to increased productivity and cost savings across various sectors. Moreover, AI-driven insights derived from analyzing massive datasets can inform decision-making processes in fields such as healthcare, finance, and business, leading to more informed and strategic choices.

How does AI technology work?

AI technology involves the use of algorithms, computer programs and statistical models to enable machines to collect data, process information, recognize patterns, and make decisions based on that knowledge. By following this process, it is continuously learning, improving and evolving to deliver better outcomes.

Artificial Intelligence (AI) operates through complex algorithms and computational processes that mimic human cognitive functions, enabling machines to perform tasks that typically require human intelligence. These algorithms analyze vast amounts of data, identify patterns, and make predictions or decisions based on learned knowledge. Machine learning, a subset of AI, involves training algorithms on large datasets to improve their performance over time without explicit programming. Additionally, AI systems utilize techniques such as natural language processing, computer vision, and neural networks to interpret and understand information from various sources. Through continuous iteration and optimization, AI systems evolve and adapt to new challenges, demonstrating remarkable capabilities in problem-solving, decision-making, and automation across diverse domains.

Submitted by: Krishna Kattel (077BCT033)

Definition of AI:

In simple words, Artificial Intelligence (AI) refers to the development of intelligent machines that can perform tasks that usually require human intelligence. AI examples include virtual assistants like Siri, recommendation systems on platforms like Netflix, self-driving cars, healthcare applications like medical imaging analysis, and AI-powered gaming.

How does AI technology work?

AI systems employ algorithms, software codes, and statistical frameworks to empower machines in gathering data, analyzing information, identifying patterns, and rendering decisions, thereby perpetually refining, advancing, and adapting to achieve superior results.

Artificial Intelligence (AI) operates through complex algorithms and computational processes that mimic human cognitive functions, enabling machines to perform tasks that typically require human intelligence. These algorithms analyze vast amounts of data, identify patterns, and make predictions or decisions based on learned knowledge. Training algorithms within the realm of Artificial Intelligence, a subset being Machine Learning, entails enhancing their capabilities over time through exposure to extensive datasets, thus refining their performance without the need for direct programming instructions. Additionally, AI systems utilize techniques such as natural language processing, computer vision, and neural networks to interpret and understand information from various sources. Through continuous iteration and optimization, AI systems evolve and adapt to new challenges, demonstrating remarkable capabilities in problem-solving, decision-making, and automation across diverse domains.

Importance of Artificial Intelligence

The significance of Artificial Intelligence (AI) cannot be overstated in today's world, as it holds immense potential to revolutionize industries, enhance efficiency, and improve quality of life. AI technologies enable automation of tasks that were once time-consuming or impossible for humans to accomplish, leading to increased productivity and cost savings across various sectors. AI technologies facilitate the automation of



LIMITATIONS & FUTURE ENHANCEMENTS

Limitations

- Scalability: cannot handle large amount of documents well
- Algorithm sensitivity:
 - factors such as document length, language complexity, and writing style diversity.
 - The algorithms may struggle to accurately detect subtle instances of plagiarism or may produce false positives due to similarities in **common phrases** or terminology
- Plagiarism Detection speed
- Cannot handle large files well
- File format limitations (**PDF**)

Future Enhancements

- Machine Learning Implementations
 - semantic analysis
- Extension fo Google classrooms
- Local Hash maps
 - the system can improve data retrieval and storage operations
- Accept all file formats like .docx, .odt

References

- Pallets Projects. (n.d.). Flask Documentation (Version 3.0.x). Retrieved from <https://flask.palletsprojects.com/en/3.0.x/>
- Google. (n.d.). Firebase Documentation. Retrieved from <https://firebase.google.com/docs>
- React. (n.d.). React Documentation. Retrieved from <https://react.dev/>



Thank you...