**PLAGIARISM CHECKER**

**Documents :** Documents can be checked with our database and with web.

Two documents can also be checked against each other by uploading both the documents at the same time.

**PROCEDURE/FEATURES :**

1. Document will be directly uploaded or can be copy pasted to our editor via python libraries.
2. Type of document will be identified eg: pdf,doc,docx,htm,xml .
3. Then according to the type of document,file will be converted to text file eg: for pdf ,Python provides PDFMiner to convert pdf to text file.
4. Then by applying suitable algorithms(to be discussed), user file will be tested against web and database to find the plagiarism.
5. A bar will be shown indicating the amount of plagiarism in document via different danger colors and percentage.
6. An option will be provided to the user to save the document to cloud or personal PC. Google API’s will be used to save document to cloud.
7. A check box will be provided to enable/disable whether the document will be send to specified email with comments or not. Python provides the library to send the email. Voice comments can also be sent as attachments through google API.
8. A final analysis will be shown to the user about the document i.e. the percent of document copied from different websites using bar graph.
9. Logged in user can also see complete history of the documents and can also access his/her documents via cloud.

**Codes :** Codes will be checked for different formats(.cpp,.java etc) according to the above procedure but by applying different algorithms due to its different structures(like instead of full stop ,a semi colon is used to identify line break etc.)

**Techniques:**

1. Web Crawling
2. Python libraries
3. Database for storing files
4. API’s for Cloud Storage
5. API’s for Voice Comments
6. Algorithms