

(12) PATENT APPLICATION PUBLICATION

(21) Application No.202211060995 A

(19) INDIA

(22) Date of filing of Application :26/10/2022

(43) Publication Date : 04/11/2022

(54) Title of the invention : SYSTEM TO VERIFY INGENUITY OF MULTIPLE DOCUMENTS USING TERM FREQUENCY AND COSINE SIMILARITY

(51) International classification :G06F0040194000, G06F0016951000, G06F0016310000, G06N0020000000, G06F0040300000

(86) International Application No :NA  
Filing Date :NA

(87) International Publication No : NA

(61) Patent of Addition to Application Number :NA  
Filing Date :NA

(62) Divisional to Application Number :NA  
Filing Date :NA

(71)Name of Applicant :

**1)Chitkara University**

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India. Patiala -----

**2)Bluest Mettle Solutions Private Limited**

Name of Applicant : NA

Address of Applicant : NA

(72)Name of Inventor :

**1)SHARMA, Rajnish**

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India. Patiala -----

**2)MISHRA, Rahul**

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India. Pune -----

**3)PANDEY, Sakshi**

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India. Pune -----

(57) Abstract :

The present disclosure relates to system (100) and a method (400) to verify ingenuity of one or more documents using term frequency and cosine similarity. The system includes an input device (106) configured to provide one or more documents to at least one processor (104). The processor, receive a signal, indicative of the one or more documents from the input device, divide each sentence of the received one or more documents into component words, and checks similarities between the component words included in the received one or more documents with a database of crawled content. Further, the processor evaluates plagiarism value using a term frequency and cosine similarity technique, and display the evaluated plagiarism value on the input device, and the evaluated plagiarism value pertains to an amount of plagiarism in one or more documents.

No. of Pages : 20 No. of Claims : 8