

## EMAS Framework For Text Plagiarism Detection (Evolutionary Multi-Agent System)

<sup>1</sup> S.N.Autade

<sup>1</sup> Research Scholar, Department of  
Information Technology,  
Bharati Vidyapeeth College of  
Engineering, Pune, India.

<sup>2</sup> Prof.S.Z.Gawali

<sup>2</sup> Professor and Head, Department of  
Information Technology,  
Bharati Vidyapeeth College of  
Engineering, Pune, India.

<sup>3</sup> Prof. Dr. D. M. Thakore

<sup>3</sup> Professor and Head, Department of  
Computer Engineering,  
Bharati Vidyapeeth College of  
Engineering, Pune, India.

### Abstract

Research ultimate goal remains to Enhance Science and Technology. Scientists, Research scholars and teacher are dedicated to research. But It has been Observed that in order to achieve success research methodology is been plagiarized. Investigating and Identifying Genuine Research innovation is demand of Today's research domain. Idea Innovation and Invention are vital for today's research domain. In context to Computer Technology algorithmic procedures, architecture Design are been commonly plagiarized without proper citation and referencing. Text Plagiarism is preliminary stage to evaluate any research work. Research Analysis Question Observed in Text Plagiarism is large unknown patterns and ever increasing unstructured data. Patterns change interestingly from research domains and no precise technique exists to retrieve them. Existing plagiarism detection system lack to sense information and merely work on keyword sense. Further Architecture of existing plagiarism detection system in not scalable. A decision Support System is been required which would learn over time as data grows in volume. Proposed research is implementation of Multi-agent system (MAS) with evolutionary Feature. Multi-Agent system assists in decision making and interestingly detect new Patterns of plagiarism adopted by scholars.

**Keywords** Text Plagiarism Detection, Multi-Agent System, Software Agents, Natural language Processing, Preprocessing, Pearson correlation.

### INTRODUCTION

Plagiarism has been termed as stealing, theft of concept, idea writings of other research scholar and presenting as inventor of it [1]. In context to research today scholars are aware that plagiarism should be avoided but lack to understand what why and how plagiarism occurs. Hand book of sterling university defined plagiarism as "process of compelling others research idea, innovation or invention as own work" [2]. Plagiarism is simply defined as incorporating scholarly work of innovation without proper acknowledgment. In context to Computer Science it is act of stealing Algorithm design or complete research methodology.

Plagiarism has been well pictured on Albright college website [3] as shown in figure 1. Plagiarism comprises of any of following act

- ❖ Paraphrasing author writings with citation
- ❖ Using other Authors ideas without acknowledging source.
- ❖ Embedding pictures images with referencing
- ❖ Submission of mathematical equations Formulas as own with citations.
- ❖ Misguiding with alteration of documented idea or even spoken talk without compelling inventor of innovation.
- ❖ Twisting and resubmitting innovation that has already been credited.



Figure 1: Act of Plagiarism

Focus of invention remains to enhance system with better automation but tampered research nullifies time and resources. As of Easy way to gain faster results Plagiarism has been common act in academics, in poetry movies, scientific research and even in medicinal domain. Figure 2 represents domain of plagiarism that have been commonly observing higher rate of plagiarism.

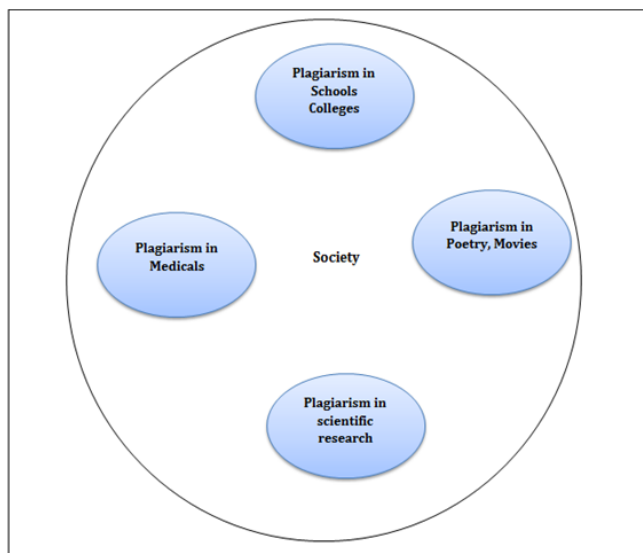


Figure 2: Plagiarism in society

Innovation is been burnt up with Plagiarism and upholds scientific progress of society. Innovation is been required to harvest knowledge and years of decades are required to achieve this. Plagiarist with Intelligence copy Tampere and spread it as own. In current research scenario with availability of web and large library resources Plagiarist copy from single or multiple resources on web. Offline resources like Library documents .Investigating plagiarism at such levels remains one open challenge. Change in pattern of writing or language is most common technique employed for restructuring sentence to Misguide scientific community and detecting such work is also major challenge. Scientific ethics urge that pattern phrasing could be done provide with correct acknowledgement and citation. Plagiarism has been observed at in various domains but in general context to scientific research and computer technology Plagiarism refers to nature of content as categorized as:

1. Text –Image Plagiarism
2. Source Code plagiarism
3. Chart and Equation Plagiarism.

Numerous sub or cross categories of plagiarism exists like in ideas songs and stories etc. this research article focus on majorly solving issues related text Plagiarism. Information of web consist of both structured, unstructured information investigation is as such challenge. Text Plagiarism stands as initial step to analyze research hard work and merit of work. Information retrieval from unstructured data is challenge as no standard patterns can be defined and nature of patterns depend on domains. Computational challenge remains to architect a scalable and adaptable algorithmic procedure to handle ever growing information.

## LITERATURE SURVEY

Literature analysis is effectively first step in order to find open issues and challenges .Tag up unresolved research issues related to text Plagiarism. Objective of this section remains to

understand existing research methodology, Limitations of this techniques and scope of work which would enhance System. Define a Simplified problem statement to major objective. Survey Analysis has been done in around text plagiarism and previously scholarly work presented in [18].

### A. Survey Technique

Systematic Literature Survey breakdown has been performed on articles extracted from IEEE Explorer ACM and portals like Google Scholar. Major Review articles have been taken from IEEE contributing 50% 20% articles from ACM 15% from international Journal 15% from other source

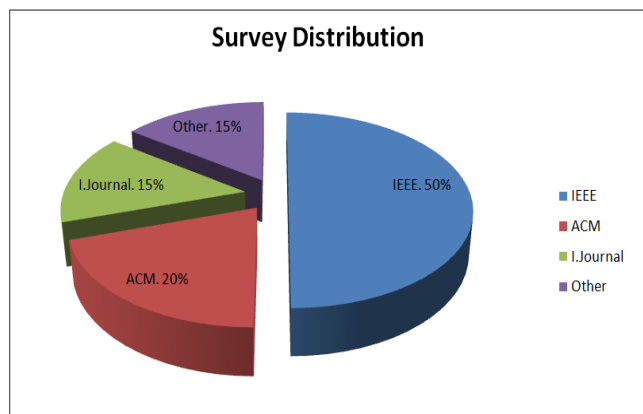


Figure 4: Survey Distribution

### B. Survey String

Survey string “s” has been used to extract Manuscript from portals

S= “Plagiarism Detection”+ text plagiarism+ citation based plagiarism detection.

### C. Survey Pattern

Each and every manuscript has been surveyed and breakdown in pattern <core Technique, Limitation, Scope>. Outcome of survey is Research analysis Questions (RAQ’s) which need to be addressed by research

Measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

### D. Survey Analysis

First IEEE addressing student plagiarism is been presented by [1]. Numerous algorithms have been Analyzed simplest ne like string matching, correlation between characters. Level based plagiarism has been introduced. Five levels of plagiarism have identified major challenge is time complexity in all algorithms. Existing techniques are time complex and cannot perform accurately. Fully automated plagiarism detection System for source code and text based assignment examination.

Preprocessing as algorithmic impact on plagiarism detection has been analyzed by [2]. Preprocessing techniques like tokenization, stemming, numeric replacement, synonym identification, word generalization, First meaning selection

that have been commonly used in plagiarism detection. Research presents best technique with combinational approach for preprocessing as it highly influences text plagiarism detection. Synonymy recognition (SYR), word generalization (WG) are only effective techniques that increase in performance of system. Generalization technique up to level of 4<sup>th</sup> level would enhance system performance.

Literal and intelligent plagiarism has been answered in [3]. Pattern analysis on linguistics like semantic term replacement, changing to short form and adoption of idea. Article is exhaustive analysis on Plagiarism Detection System. Each and every pattern in plagiarism detection and techniques have been put forward. Idea plagiarism is been unaddressed and incorporating interesting patterns in plagiarism detection is scope of work. Semantic and fuzzy logic have future scope in plagiarism detection software's.

Keyword plagiarism detection system is been presented in [4]. Proposed system extracts document section as image, extracting keywords. Comparative analysis is been done on synonyms from background research papers. Plagiarism report is been generated mapping similarity to other documents. Methodology is entirely dependent on scanning approach which is major limitation. Integration of web based dictionary would enhance system performance.

[5] Approach is Natural language processing with Fuzzy Semantics for plagiarism detection. Methodology integrates preprocessing at initial step. Secondly N-gram and semantics are been performed. Thirdly boundary for input set of passage is been computed. Finally plagiarism is been evaluated. Comparative examination has been done on PAN 2012 dataset. NLP is iterative research domain which requires better Processing techniques. Adopted NLP techniques have limitations. Boundary detection techniques could be enhanced with soft computing techniques.

Information retrieval Vector Space Model has been applied to detect plagiarism in text based assignments [6]. Proposed system implements trigram analysis with cosine similarity and Jaccard coefficient. Cosine similarity with tri-gram analysis has been found to be best. Scholars have presented tool which minimizes plagiarism in assignments. Tri-gram analysis is bit time consuming process and requires optimization. Optimization techniques development is future scope of work.

Fingerprint analysis FTPDS has been applied in plagiarism detection [7]. Suspicious documents are been detected with fingerprint analysis. System has been designed as agile software which compares given pdf file to set of files. System has three phase detection process, document analysis, comparison, reporting with results. Major limitation observed is speed of comparison which could enhance. Graph and table plagiarism has not been detected. Memory and internet speed is observed as limitation in online plagiarism detection. Eliminating duplicate fingerprints. Online search could be enhanced.

Hybrid methodology in plagiarism detection is been implemented by scholar [8]. Existing approach are lesser accurate and time complex. Citation based methodology reduces retrieval space and scalable methodology for large database search. A prototype system Citeplag is been designed which is in preliminary stage of development. Proposed

system could be future be evaluated for large and real time scenario.

[9] Focuses on keyword based plagiarism detection system. Suspicious documents are been detected based on terms and phrases. TF-IDF and W-TF-IDF are two parameters which assist in suspected document detection. In TF\_IDF with Vector space model has for detecting plagiarized paragraphs has been found best. Keyword extraction system don't highly impact detection process. Every plagiarism detection system based on word match could adopt correct methodology.

Machine with enhanced architecture would certainly uplift plagiarism detection process. Research [10] focuses on implementing parallel algorithm for plagiarism detection. GPU graphics processing platform enhances algorithm speed 6 times existing one and optimal choice for real time plagiarism detection process. Hardware is costlier as compared to other system. Future enhanced graphics systems would uplift system performance.

Author [11] presents "preprocessing" an effective post data mining technique reducing information to actionable and meaningful information. Web log analysis has been done to extract frequent patterns. Stop word removal, special symbol elimination and stemming three step process has been implemented. Common Pattern Analysis has not been done for predefined ones only .observed as limitation of research. Generalized and more complex patterns needs to be analyzed. Specialized stemmer algorithm could assists in more accurate results.

Scholar [12] implements effective clustering procedure to Map user interest levels. Developed system map interest of user in real time. Clustering technique has been evaluated based on Manhattan, Euclidean, and Pearson correlation. Clustering reduces search space and enhances pattern mining process. System lack in Decision Support and requires human intervention for accurate results. Algorithmic procedures like genetic algorithms would uplift system performance and remains scope of research.

Research Analysis of [13] presents application of reinforcement learning a software agent Approach. Temporal alteration in text writing pattern has been used to investigate plagiarized documents. Author misconduct has been analyzed with machine learning. POS, Dynamic programming and carlo procedure have been major components in reinforcement learning system development. System has not been tested for large dataset. System has been evaluated only for documents consisting of ten sentences and overall test scenario of 107 sentences. Graph based Matching or decision tree support with Multi-threading are future scope of enhancement

Scientific research by [14] focuses on analyzing linguistic patterns, features and investigative method in context. Plagiarists have used common patterns like Synonym replacement. Short form reduction of text is second most commonly used plagiarism detection technique. Current research methodologies like Semantic analysis, vector matching and latest n-gram fail to detect intelligently plagiarized patterns in text. Existing plagiarism detection techniques only perform document to document, sentence to sentence and finally word to word similarity computation only. Pattern analysis is need to be done in better way with new Patterns identification. Semantic Match (SEM) and fuzzy

techniques have not been implemented. Design and Development of scalable framework with dynamic component add up is yet to be achieved. Idea based plagiarism detection is open highlighted question

Three layer architecture has been implemented by [15] for multi-document comparison. System detects in alteration substitution in text plagiarism. System firstly examines structure of document with preprocessing, Levenshtein distance measure is been used multi-document comparison. Recursive programming has been used to enhance system. Algorithmic procedure is just a prototype and needs to be tested for generalized scenario. In future work System needs to be fine-tuned to obtain better results.

Innovation of [16] focuses on analyzing text based plagiarism with citation match with GuttenPlag. Existing plagiarism detection focuses on fingering and string matching. Research work focuses on paraphrasing. Integrated approach of text and citation match enhances system performance. Similar citation pattern has been detected and tested on PAN-PC corpus for 31 pages input. System has been successful to detect only small fragment of 14 sentences and remains major limitation of work. Citation based investigation has additionally detect 13 to 16 more segment as plagiarized

Innovation of scholar [17] is extension of previously designed system and focuses to analyze intelligently disguised plagiarism patterns. Proposed system implements bibliographic Coupling, recognizing documents with similar citations. Three procedures have been used longest citation chunking and tiling. System is prototype and works only for small amount of disguised citation references. Largely disguised plagiarism detection is yet to be achieved and remains major scope of work.

Doctoral research work of [18] presents plagiarism detection based on analysis of article title, author and content pattern examination. Internet and easy to access information in form of pdf doc has raised plagiarism levels. Existing system use open source API for investigation, but this is complex and time consuming process. Work focuses to find intelligently plagiarism with levels of plagiarism.

Limited to small file formats. Inner deep crawling is not ben done. Documents structure needs to be examined. Design and development of deep web crawlers is scope of work. Offline dataset would enhance system speed. Additionally format of document and font needs to be examined.

Innovative idea of [19] has developed in learning system. Current research challenge is to integrate e-learning system. Proposed work with ontology and sharable object model implements MAS. A multi-agent framework for real time scenario. System is based on co-operative nature of agents. Content controller, presenter organizer is four agent framework used. Rule based i.e. supervised learning has been implemented. Complete learning based system needs to be implemented. Design and development of generalized multi-agent system with evolutionary algorithmic procedure is future of MAS system.

Author [20] presents research work for identification of context and sensing a word meaning in document. Existing approach's include dictionary based supervised or unsupervised disambiguate elimination but come up with shortcomings. Intelligent agent based extraction system is

been developed. Only smaller length word have been considered. Multi-value classification is problem. Combined agent system would assist in system enhancement development of multi-agent system is scope of work.

Author [21] presents research work for identification of context and sensing a word meaning in document. Existing approach's include dictionary based supervised or unsupervised disambiguate elimination but come up with shortcomings. Intelligent agent based extraction system is been developed. Only smaller length word have been considered. Multi-value classification is problem. Combined agent system would assist in system enhancement development of multi-agent system is scope of work

Research work of [22] focus on intelligent text analysis. Agent based Approach is been adopted for concept extraction from text. Agent is software program which analysis document based on input keyword. Proposed work presents Architecture of agent based system. Singular agent based approach is been adopted which is major limitation. Multi-threading and parallel programming remains future implementation scope.

Scholars [23] have designed a small search tool with innovative ranking algorithm. Proposed Methodology innovates agent program which understand sense of query and extract concept based information rather than just keyword based. Ranking of information is challenge which is been addressed with functionality of TF\_IDF and background document list. Retrieved results have been tested on Reuters dataset-21758 .observations conclude that simplified ranking is approach is been innovated. Proposed system is idea, a protocol implementation and tested only on single dataset.

Automated indexing is future scope of work. Research of [24] compares citation based approach with text based approach. Online easily availability of resources like articles, many research scholars reuse others work with acknowledging original writer. Effective research methodology is been devised based on integrated approach of text and citation matching. Distance metric like CF score, Jaccard have been used in system evaluation. Larger space and time complexity. Better information representation techniques are been required for better comparative analysis. Hybrid research methodology is future scope of work.

Textual plagiarism has been evaluated majorly on available dataset which is major limitation. [25] Work focuses on intrinsic plagiarism considering authors writing style to identify suspicious documents. Proposed algorithm is simple procedure dividing text in blocks of tokens, analyzing Stylographic features, future POS is been examined. Predictive score is been given to identify suspicious documents. Word list is been used that is limited to 200-400, failing system to perform in real time scenario. Analysis of overall word frequency function would be enhanced. Considering real time scenario is scope of work.

### ***E. Research Analysis Question(RAQ's)***

Summarizing above survey major vital research Questions have been found as:-

**RAQ1:** Pattern Analysis is major Research Question in Text based Plagiarism.

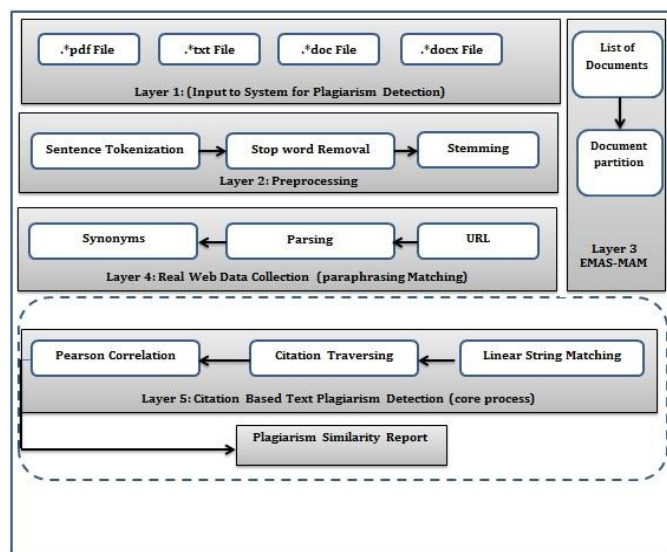
**RAQ2:** Detection text Plagiarism with citation referencing and context mapping would enhance existing plagiarism system.  
**RAQ3:** Identifying and recognizing patterns as exact, partial or web match with reinforcement approach in machine learning.  
**RAQ4:** Enhancing MAS (Multi-Agent System) with Evolutionary Algorithmic process, Design and Development of EMAS Framework.

## CORE METHODOLOGY

Proposed Methodology Implements layered system with five layer Approach.

**Layer I:** Accepts Input to system. File formats like PDF, DOC/DOCX or Text File are accepted. On layer I Initialization complete file is been read by software agent as string. NLP processing protocols assist in separation of sentences, generating sentence vector. Algorithmic procedure like preprocessing and data cleaning have been applied at this layer.

Figure 5 represent blueprint of proposed system Table Styles



**Figure 5:** Proposed Architecture

**Layer II:** Implements Evolutionary Multi-Agent system. Algorithmic procedure of Map reduction is been implemented. Here dynamically and according to situation agents are been initialized, breaking down Document list for future processing. Multi-threading procedure is been implemented for parallel processing.

**Layer III:** Implements preprocessing data transformation and data cleansing process. This process consist of three sub process special symbol elimination, stop word removal and stemming procedure.

**Special Symbol Elimination:** discarding every special symbol is been done. List of known symbols is been stored in vector for comparison. Symbol characters like {"!", "\$", "%"}.....} are been eliminated.

**Stop word Elimination:** NL conjunctions are been discarded. A vector list of stop word is been used to eliminated known

conjunctions. Even though words have been eliminated the sense of sentence and content remains undistorted.

**Stemming Process:** A process which help in part of speech recognition by riming word to root form. Rule based stemmer has been implemented which is lighter and lesser in time complexity. Rule used for stemming are {"ing", "ization"}

**Layer IV:** Paraphrase Matching is been done at layer 4 where real web data is been collected from input URL and synonymous word a list is been generated for future matching. Automated web crawler has been implemented at this layer parsing web information to semi-structured format.

**Layer V:** Linear string matching is been implemented at this layer. For given query document every sentence is been matched to dataset file and N-gram technique is been implemented. Algorithmic procedure 2 represents overall procedure. Citation based Analysis is done citation agent which looks for known referencing patterns in background dataset. Correlational similarity is been computed with equation 1 which retrieves accurate results.

$$R = \frac{\sum m_i n_i - \frac{\sum m_i \sum n_i}{10}}{\sqrt{(\sum m_i^2 - \frac{(\sum m_i)^2}{10})} \sqrt{(\sum n_i^2 - \frac{(\sum n_i)^2}{10})}} \quad \text{--- (1)}$$

Here

m=vector fits to query string

n=vector in dataset

r=correlation factor

Correlation is nearer to 1 means the vector are correlated, so that sentences are also correlated. If they are nearer to 0 means they are not correlated.

### Algorithm I: Evolutionary Multi Agent System (EMAS)

Input: List of Documents

Output: Input of Document to Agent

Process:

Step I: Start

Step II: Decision Analysis of Size N

Step III: Initialize List L

Step IV: Initialize Name of all documents in a List

Step V: Map Stream (M)= D /N

Step VI: Set Count=0

Step VII: For i=0 to size of D

Step VIII: count++

Step IX: Add Di into Temp List T

Step X: IF count=M

Step XI: Add T to L

Step XII: Reset T

Step XIII: END FOR

Step XIV: return L

Step XV: Stop

**Algorithm II:** Pattern Analysis Algorithm

Input: Vector Q[] and D[] Sentences  
 Output: Similarity report  
 Step I: Start Process  
 Step II: Set FLAG=FALSE, Set Y[ ]  
 Step III: FOR i=0 TO size of total size of Q  
 Step IV: QW=Q[i], Initialize TF=FALSE  
 Step V: FOR j=0 TO size of D  
 Step VI: DW=D[j]  
 Step VII: IF QW=DW  
 Step VIII: TF=TRUE, BREAK  
 Step IX: END FOR  
 Step X: IF TF=TRUE  
 Step XI: X[i]=1  
 Step XII: ELSE  
 Step XIII: X[i]=0  
 Step XIV: END FOR  
 Step XV: FOR i=0 TO SIZE OF X  
 Step XVI: IF X[i]==1 AND X[i+1]=1 AND x[i+2]=1  
 Step XVII: FLAG=TRUE  
 Step XVIII: END FOR  
 Step XIX: Return FLAG  
 Step XX: Stop

**RESEARCH EVALUATION**

Any information retrieval system is been commonly evaluated based on precision and recall. Above system has been evaluated with precision recall and correlation parameter. Plagiarism software has been designed in java under web deployment. Overall considering hardware limitations system has been tested for 157 PDF /DOC/DOCX/Txt File. System has been evaluated for detecting known patterns. Overall relative efficiency has been determined with equation (2) and (3)

$$Precision = \frac{R}{R + P} \times 100 \dots \dots (2)$$

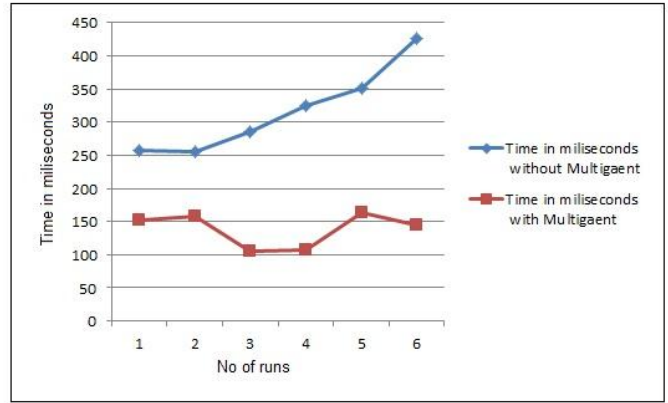
$$Recall = \frac{R}{R + Q} \times 100 \dots \dots (3)$$

Here

P = sentences identified as plagiarized

Q= sentences not identified as plagiarized.

R= irrelevant sentences identified as plagiarized.



**Figure 6** Research evaluation of system with EMAS and Without EMAS Approach

| Sr.No | Time in milliseconds without EMAS | Time in Milliseconds with EMAS |
|-------|-----------------------------------|--------------------------------|
| 1     | 257                               | 153                            |
| 2     | 256                               | 157                            |
| 3     | 286                               | 105                            |
| 4     | 324                               | 107                            |
| 5     | 352                               | 163                            |
| 6     | 426                               | 145                            |

**ACKNOWLEDGMENT**

I Acknowledge Each and every research scholar whose research article has been used by me for knowledge gaining. All acknowledge my guide Prof. S.Z. Gawali for selecting me for this research work. I thank all of my friends who assisted me directly or indirectly in research work.

**REFERENCES**

- [1] Parker, Alan, and James O. Hamblen. "Computer algorithms for plagiarism detection." *IEEE Transactions on Education* 32.2 (1989): 94-99.
- [2] Ceska, Zdenek, and Chris Fox. "The influence of text pre-processing on plagiarism detection." Association for Computational Linguistics, 2011.
- [3] Alzahrani, Salha M., Naomie Salim, and Ajith Abraham. "Understanding plagiarism linguistic patterns, textual features, and detection methods." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.2 (2012): 133-149.
- [4] Dutta, Sandipan, and Debotosh Bhattacharjee. "Plagiarism Detection by Identifying the Keywords." *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on.* IEEE, 2014.
- [5] Gupta, Deepa, K. Vani, and Charan Kamal Singh. "Using Natural Language Processing techniques and fuzzy-

- semantic similarity for automatic external plagiarism detection." *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on. IEEE, 2014.*
- [6] Jiffriya, M. A. C., MAC Akmal Jahan, and Roshan G. Ragel. "Plagiarism detection on electronic text based assignments using vector space model." *Information and Automation for Sustainability (ICIAfS), 2014 7th International Conference on. IEEE, 2014.*
- [7] Elkhidir, Mohamed, et al. "Plagiarism detection using free-text fingerprint analysis." *Computer Networks and Information Security (WSCNIS), 2015 World Symposium on. IEEE, 2015.*
- [8] Meuschke, Norman, and Bela Gipp. "Reducing computational effort for plagiarism detection by using citation characteristics to limit retrieval space." *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries. IEEE Press, 2014.*
- [9] Ning, Hui, et al. "Comparisons of keyphrase extraction methods in source retrieval of plagiarism detection." *Computer Science and Network Technology (ICCSNT), 2015 4th International Conference on. Vol. 1. IEEE, 2015.*
- [10] Jiffriya, M. A. C., et al. "Accelerating text-based plagiarism detection using GPUs." *Industrial and Information Systems (ICIIS), 2015 IEEE 10th International Conference on. IEEE, 2015.*
- [11] Thakare, Sanjay Bapu, and Sangram Z. Gawali. "A effective and complete preprocessing for Web Usage Mining." *International Journal on Computer Science and Engineering 2.03 (2010): 848-851*
- [12] Thakare, Sanjay B., and Sangram Z. Gawali. "Selecting type of clusters that are most appropriate for capturing overlapping interests of different types of users in personalization tasks using Web Usage Mining." *International Journal of Computer Science and Information Technologies 1.1: 58-60.*
- [13] Salunkhe, Sudhir D., and S. Z. Gawali. "A plagiarism detection mechanism using reinforcement learning." *International Journal of Advance Research in Computer Science and Management Studies 1.6 (2013).*
- [14] Alzahrani, Salha M., Naomie Salim, and Ajith Abraham. "Understanding plagiarism linguistic patterns, textual features, and detection methods." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.2 (2012): 133-149.*
- [15] Zini, Manuel, et al. "Plagiarism detection through multilevel text comparison." *Automated Production of Cross Media Content for Multi-Channel Distribution, 2006. AXMEDIS'06. Second International Conference on. IEEE, 2006.*
- [16] Gipp, Bela, Norman Meuschke, and Joeran Beel. "Comparative evaluation of text-and citation-based plagiarism detection approaches using guttenplag." *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries. ACM, 2011.*
- [17] Gipp, Bela, et al. "Demonstration of citation pattern analysis for plagiarism detection." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013.*
- [18] Al Jarrah, Abeer, Izzat Alsmadi, and Zakariya Za'atreh. "Plagiarism Detection based on studying correlation between Author, Title, and Content." *International Conference on Information Communication System (CICS). 2011.*
- [19] Yaghmaie, Mahkameh, and Ardeshir Bahreininejad. "A context-aware adaptive learning system using agents." *Expert Systems with Applications 38.4 (2011): 3280-3286.*
- [20] Barriuso, Alberto López, et al. "JOUR-MAS: a multi-agent system approach to help journalism management." *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 4.4 (2015): 23-34.*
- [21] Suryawanshi, Ranjeetsingh S., D. M. Thakore, and Kaustubh S. Raval. "Context Based Word Sense Extraction in Text: Design Approach." *International Journal of Computer Science and Information Security 9.5 (2011): 95.*
- [22] Raval, Kaustubh S., Ranjeetsingh S. Suryawanshi, and Devendra M. Thakore. "An Intelligent Agent Based Text-Mining System: Presenting Concept through Design Approach." *International Journal of Computer Science and Information Security 9.4 (2011): 112.*
- [23] Raval, KAUSTUBH S., et al. "The Anatomy of a Small-Scale Document Search Engine Tool: Incorporating a new Ranking Algorithm." *International Journal of Engineering Science and Technology 3.7 (2011).*
- [24] Zu Eissen, Sven Meyer, and Benno Stein. "Intrinsic plagiarism detection." *European Conference on Information Retrieval. Springer Berlin Heidelberg, 2006.*