

Research Effort Evaluation Framework A Prototype Model (Text-Code-Image-Analysis)

Kamlesh Patil¹, Dr. S.B. Vanjale²

^{1,2}Department of Computer Engineering,
Bharati Vidyapeeth deemed to be University, College of Engineering
Pune, Maharashtra, India.

Abstract

This Research work presents work from idea to prototype model implementation for Research effort Evaluation. In this time of readily available information from diverse sources judging original efforts contributing to research work is challenging task. Moto of research remains to facilitates ease to human life which requires original work from research scholars. Existing software focusing are Text Analyzers which predict based on textual and image features. Currently where few integrated system exist, but no single framework for scholar effort estimation.

This work presents prototype model with integrated approach on text, image, code feature analyzes to predict scholarly efforts. Identification and sensing this diverse pattern is challenge . Paraphrasing and synomus replacement are common pattern of diversifying, this work puts forward the prototype for idea based Research falsifying. Many research scholars just to hype research work include in cross domain terms , irrelevant terms which are confusing and taking research in incorrect direction.

This Research work efforts to facilitate framework for academic scholars at graduate and post graduate level. This Prototype Model is Three layers architecture where layer focuses on text analysis , layer two focuses on image analysis and layer 3 focuses on code analysis finally generating a report of scholar person estimating this research efforts.

Keywords: Text Analyzer, Plagiarism detection, Pattern Analysis, code Analyzer, Image Features.

INTRODUCTION

Research Falsifying is Major Challenge Face by academic and Institutions which leads to unnecessary time and effort wastage. Software systems like Text Plagiarism detection exist which try to find in patterns of false work. Although such systems are not sufficient as they have only one dimension to detect limited to either of one domain. Challenge 1: diverse Patterns with singular dimension. Challenge 2: cross domain and multilingual another unfocused challenge.

Real life scenario: The problem statement behind this research work remains from a real life scenario where really why need to have this Framework comes up.

In post graduate research two scholars have different domains of research. Research scholar “A” has search engine design whereas scholar “S” has data mining concept in software engineering. Research ideas of scholar A are been stolen and effort is been made to prove that “S” has work which has all subset of “A” work. This act has been done by simply including terms which are not relevant to research from other scholars work like “document in question” , Natural language processing , era of web in formation. Search engine of future. Cross lingual methodology etc.

This scenario mainly leads to motivation of having a integrated system for real effort and detect idea stealing act , additionally as a software system it would facilitate data mining research .

SURVEY

A. Survey Technique

Simpler Survey technique has been Integrated here only latest research articles focusing on terms like research effort Plagiarism , text analysis code Analyser have been integrated here in finding only ten best articles for surveying existing system . Moto of surveying remains to find challenges what where been taken , methodology been deployed and limitations with scope of improvement

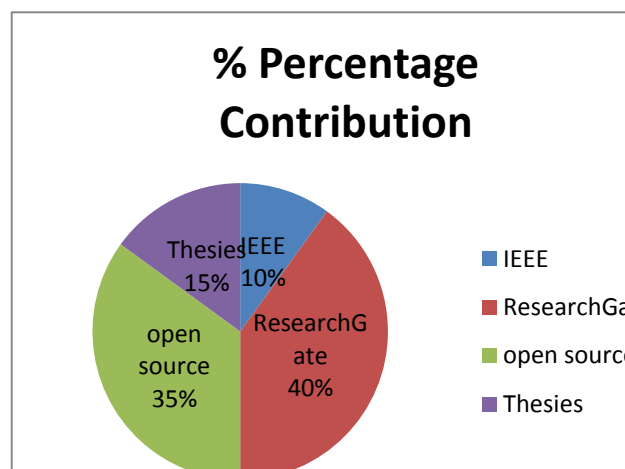


Figure 1: Survey distribution

B. Literature Survey

Ever growing information on www and ease access has given rise to unaided work without proper efforts and justifying references [1]. Incorporating some else idea and presenting as own is termed as act of falsifying research with intent to gain credit. Students been unaware of such is commonly observed at institution level. As such awareness should be maid among students. As such work such be evaluated in different way.

Author has proposed four layer detection process {COLLECTION→ DETECTION→CONFIRMATION→INVESTIGATION}. Issue found was collection process as to term that submitted first may not be original work contributor. Scope of work is data visualization format and consistency in output.[2] Authors presents “hypothetical condition’ Methodology . The concept is to put forward question which are conceptual to answer rather than factual or readily available to answer. Indirectly authors asks the tutors to change the way question are been asked. Author claims to have found no case of plagiarism as no information was made available directly on topic. This work is limited use case and cannot be always a case but definitely could change the circumstances by reducing intensity of plagiarism.[3] presents a white paper addressing technical design of modular plagiarism detection system . Supporting cross format for essay analysis. Multiple similarity measure has been implemented presenting graphical analysis. According to nature article scientific society has been hugely affected by plagiarism. Core methodology lies in pairwise distance between doc A and Similarity matrix is been generated for plagiarism computation. Threshold recommendation and outliner are addition features developed. Lack of comparable information is one of drawback system can overcome.

Hybrid Approach combining best of capabilities of software and hardware system to achieve optimal result is commonly adopted methodology. [4] present HyPlag system for detection of paraphrases ,non contextual work in plagiarism act. Additionally system detects copying in mathematical expressions, images and citations. Limitation observed here is only retracted articles are only.

Author [5] present complete taxonomy related to Plagiarism detection tools and Their Detection Techniques. Text plagiarism has been classified in seven types by author. Common two way of plagiarizing common in code text even in equation are as :

1. Copy-paste-clone plagiarism: textual plagiarism type one of author has copied work of other author completely or partially and presented self-work knowingly.
2. Paraphrasing plagiarism: This mostly common technique used by author to rewrite others work aby changing tense and word replacement process which has been successful to disguise software in many occurrences. This type has been sub classified as
 - a) simple paraphrasing: simple copying idea and writing in one style

- b) Patch work based Phrasing: here certain vital words are been replaced to change direction of meaning

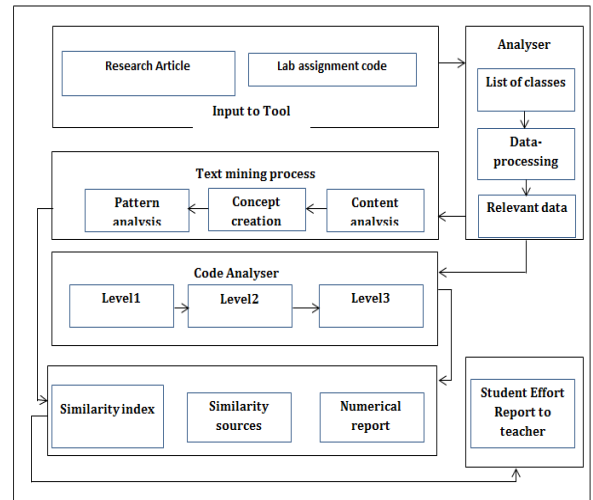


Figure 2: System Architecture

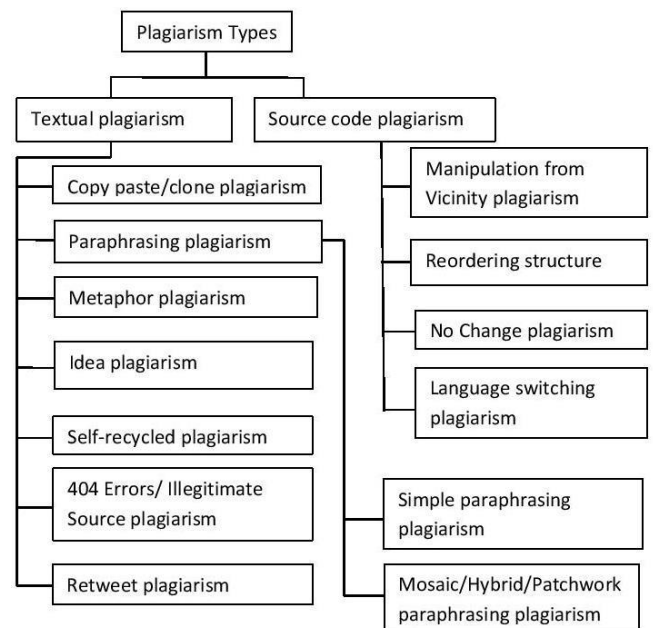


Figure 3. Types of Plagiarism [4]

Thorough analysis on plagiarism revealing Techniques are been presented here [5]. Common issue found has been no single integrated tool exists which can address all forms and check how far efforts have been taken.

Serious offense has been reported for plagiarism [6]. Article presents methodology to find similarity based detection based on section match. N gram precision recall technique has been implemented. Syntactic and grammatical resemblance Natural

language process (NLP) with supervised learning are base line methodology. Scope of work lies in implementation of block based comparison.

Research work [7] presents a detailed article with research abstract scope , methodology, contribution and findings with future work. This research work finds real time plagiarism among college students, examination is commonly considered as way of assessment which needs to be overridden with only counting actual scholar efforts,

CORE METHODOLOGY

The system implemented modular layered design approach where each layer focuses on particular type Analysis and domain.

To develop the first Text Plagiarism Detection Model for research Articles, Secondly develop code clone Analysis Model. Finally An integrated Model for evaluating research scholar efforts

Three Phase Algorithm approach is proposed where Algorithm one works on Text Plagiarism detection. Second works on code Plagiarisms. Third works on image plagiarism

Simpler lesser complex algorithm based on Pattern matching match is been designed for text analysis

Remove all comments.

Ignore all blanks and extra lines, except when needed as delimiters.

Perform a character string compare between the two Pattern Matching

Code analysis algorithm works on similarity level . Image analysis algorithm works on image similarity matching computing RGB values.

A. Layer 1:Input to system

Input to system consist of pdf .txt .doc format research articles . Code file of .java extension and images of .jpeg format . Input layer analysis this format and sense to accept only those one

B. Layer 2:Analyzer

Analyzer Layer performs preprocessing and data cleansing process. Eliminating unwanted symbols and irrelevant terms from text document and code files

C. Layer 3: Text Minier

Data mining with pattern identification is been applied at this layer. Content analysis is been done based on N gram technique. In order to overcome paraphrasing concept extraction is been done based on synonyms generated from web information

D. Layer 4: Code Analyzer

Similar on ground of text analysis code analysis and image analysis exists which find similar images from doc .

E. Layer 5:

Integrated Approach in plagiarism detection is presented. Detection method is based on pattern matching and similarity score.

Finally report been submitted by student is been evaluated and generated sending a mail to teacher and student.

Research work should focuses on designing A prototype Model evaluating Research Scholar Effort towards research

RESULTS AND DISCUSSION

This is a prototype Model and hence has certain limitations of implementing in time as such system presents a integrated approach which would help in future and have many open direction of research. Even though this work is not comparative to any existing tool the approach taken up has been diverse and higher than other tools

CONCLUSION AND FUTURE SCOPE

This Work is purely a prototype and idea work which has been implemented at post graduation level and cannot be compared with any industrial software as such comparative examination has not been presented as it would be false judgment .

Future this System would be taken to major level with incorporation of institutes to make a complete software system

ACKNOWLEDGMENT

The would like to acknowledge my guide and all my friends who directly or indirectly helped we in research completion

REFERENCES

- [1] Culwin, Fintan, and Thomas Lancaster. "Plagiarism issues for higher education." *Vine* 31.2 (2001): 36-41.
- [2] McKillup, Steve, and Ruth McKillup. "An assessment strategy that pre-empts plagiarism." *International Journal for Educational Integrity* 3.2 (2007).
- [3] Freire, M. A. N. U. E. L., and M. Cebrian. "Design of the ac academic plagiarism detection system." *Technical report, Tech. rep., Escuela Politecnica Superior, Universidad Autonoma de Madrid* (2008).
- [4] Meuschke, Norman, et al. "HyPlag: A Hybrid Approach to Academic Plagiarism Detection." (2018).
- [5] Chowdhury, Hussain A., and Dhruva K. Bhattacharyya. "Plagiarism: Taxonomy, Tools and Detection Techniques." *arXiv preprint arXiv:1801.06323* (2018).
- [6] Imran, Shamas, et al. "An Enhanced Framework for Extrinsic Plagiarism Avoidance for Research Article." *Technical Journal*23.01 (2018): 90-97.

- [7] Jeske, Heimo J., Manoj Lall, and Okuthe P. Kogeda. "A Real-time Plagiarism Detection Tool for Computer-based Assessments." *Journal of Information Technology Education: Innovations in Practice* 17 (2018): 23-35.
- [8] Jiang, Yanyan, and Chang Xu. "Needle: Detecting Code Plagiarism on Student Submissions." (2018).
- [9] Ragkhitwetsagul, Chaoyong, Jens Krinke, and Bruno Marnette. "A picture is worth a thousand words: Code clone detection based on image similarity." *Software Clones (IWSC), 2018 IEEE 12th International Workshop on*. IEEE, 2018.
- [10] Gelbukh, Alexander, and Grigori Sidorov. "Plagiarism Detection Through Paraphrase Recognition." (2018).