Plagiarism Checking and Document Editor System

[1]Mr. Shubham Chaudhary, [2]Mr. Akash Bankar, [3]Mr. Nilesh Gavande, [4]Mr. Rohit Patil, [5]Prof. (Dr) Amit Gadekar

[1,2,3,4]B.E (Computer Engineering), [5]Guide
Department of Computer Engineering,
Sandip Institute of Technology and Research Centre, Nashik, India

Abstract: Plagiarism of digital documents seems a serious problem in today's era. Plagiarism refers to the use of someone's data, language and writing without proper acknowledgment of the original source. Plagiarism can be of different types. This paper presents a different approach for measuring semantic similarity between words and their meanings. English is a language that is spoken by around 380- 420 million people on this planet and understanding it is not at all easy. The meaning of a sentence varies according to the context and the tone of the speaker. To convey the thoughts efficiently, the knowledge of the language and its various rules is very important as thoughts take the form of words and the words take the form of action. Existing systems are based on the traditional approach. For detecting plagiarism, traditional methods focus on grammatically correctness according to keywords and sentence. We have suggested new strategies for detecting the plagiarism in the user document using the semantic web. In paper we have proposed architecture and algorithms to better detection of grammatically mistakes using semantic search, it can improve the performance of grammar detection system. It analyses the user typed document. After the implementation of this technique, the accuracy of plagiarism detection system will surely increase. One should aim to minimize the errors while using the language. Lesser is the number of mistakes, better will be the communication. To aid in achieving this goal, we are creating a frequency-based spell checker and a rule-based grammar checker for English language. The grammar checker focuses on detecting and correcting tense related mistakes. In this system, user have some difficulties that are not solve by the system then he also chats with our expert and try to solve their problem

Keywords: Plagiarism; plagiarism detection, SemanticWeb, Grammar Checker, Tense, Suggestion Prediction, Expert Chat.

I. INTRODUCTION

Nowadays with the rapid use of internet makes it easy to collect information as a matter of mouse click over the web. With this enormous opportunity it becomes easy to make copy and use of someone else's works or ideas without giving the proper credit to the original owner. Normally this king of practice is known as plagiarism. It has already become a serious issue for the different publishing house. Considering this issue several tools have been developed to check for plagiarism mostly based on syntactical or word-based matching. In this paper a new conceptual model is proposed to generate a decent tool to identify the plagiarized content based on grammatical, syntactical and semantic matching.

II. RELATED WORK

Intelligent Plagiarism Detection Mechanism using Semantic technology: A Different Approach.
Juhi Agarwal, R H Goudar, Pratik Kumar, Nishkarsh Sharma: Plagiarism of digital documents seems a serious problem in today's era. Plagiarism refers to the use of someone's data, language and writing without proper acknowledgment of the original source. Plagiarism of another author's original work is one of the biggest problems in publishing, science, and education. Plagiarism can be of different types.
This paper presents a different approach for measuring semantic similarity between words and their meanings. Existing systems are based on the traditional approach.
For detecting plagiarism, traditional methods focus on text matching according to keywords but fail to detect intelligent plagiarism using semantic web.

We have suggested new strategies for detecting the plagiarism in the user document using the semantic web. In paper we have proposed architecture and algorithms to better detection of copy case using semantic search, it can improve the performance of copy case detection system. It analyses the user document. After the implementation of this technique, the accuracy of plagiarism detection system will surely increase. Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods Salha M. Alzahrani, Naomie Salim, and Ajith Abraham: Plagiarism can be of many different natures, ranging from copying texts to adopting ideas, without giving credit to its originator. This paper presents a new taxonomy of plagiarism that highlights differences between literal plagiarism and intelligent plagiarism, from the plagiarist's behavioural point of view. The taxonomy supports deep understanding of different linguistic patterns in committing plagiarism, for example, changing texts into semantically equivalent but with different words and organization, shortening texts with concept generalization and specification, and adopting ideas and important contributions of others. Different textual features that characterize different plagiarism types are discussed. Systematic frameworks and methods of monolingual, extrinsic, intrinsic, and cross-lingual plagiarism detection are surveyed and correlated with plagiarism types, which are listed in the taxonomy. We conduct extensive study of state-of-the-art techniques for plagiarism detection, including character n-gram based (CNG), vector-based (VEC), syntax-based (SYN), semantic- based (SEM), fuzzy-based (FUZZY), structural- based (STRUC), stylometric based (STYLE), and cross-lingual techniques (CROSS). Our study corroborates

that existing systems for plagiarism detection focus on copying text but fail to detect intelligent plagiarism when ideas are presented in different words. Semantic Similarity/Relatedness for Cross language plagiarism detection Hanane EZZIKOURI, Mohamed OUKESSOU, Mohammed ERRITALI Generally utterances in natural language are highly ambiguous, and a unique interpretation can usually be determined only by taking into account the context in the utterance occurred. Automatically determining the correct sense of a polysemous word is a complicated problem especially in multilingual corpuses. This paper presents an application programming interface for several Semantic Relatedness/Similarity metrics measuring semantic similarity/distance between multilingual words and concepts, in order to use it after for sentences and paragraphs in Cross Language Plagiarism Detection (CLPD); using WordNet for the English-French and English-Arabic multilingual plagiarism cases.

III. SCENARIO IN WHICH MULTI- CORE, EMBEDDED AND DISTRIBUTED COMPUTING USED

In our proposed technologies, multi-core is usually the term used to describe two or more CPUs working together on the same chip. Also called multicore technology, it is a type of architecture where a single physical processor contains the core logic of two or more processors. This system will support the multi- core functionality.

Our system is embedded system this is controller with a dedicated function within a larger mechanical or electrical system, often with real- time computing constraints. It is embedded as part of a complete device often including hardware and mechanical parts. Embedded systems control many devices in common use today.

V. SYSTEM IMPLEMENTATION AND RESULT

IV. PROPOSED SYSTEM

Most of techniques for copy case detection are based on ideas of sub-string matching. In paper Sub- string-matching approach basically identifies maximum matches in pairs of strings, which will be used as plagiarism indicators later but this is a traditional technique which is limited to the good accuracy. In paper and paper of documents keyword similarity mechanism is used. In this paper the idea was to weight words and to compare them to the keywords. Spell Checking and Grammar Checking are very important tasks in the process of writing.

When people are in a hurry to submit an essay or abstract or any written piece in any language these aspects are recklessly abandoned without thinking of the consequences. Considering the fact that if you are given a material that has various spelling and grammatical mistakes and due to which the entire document does not deliver the idea that it must, then would you be convinced about the idea the writer wants to convey.
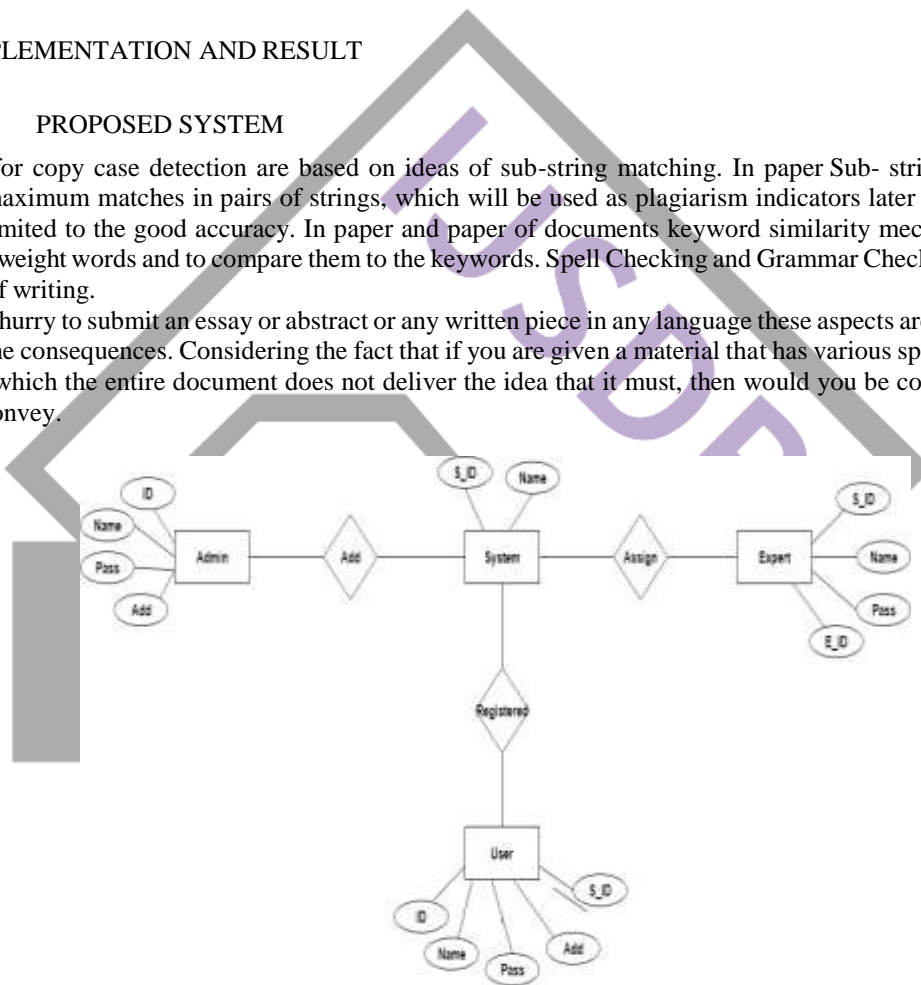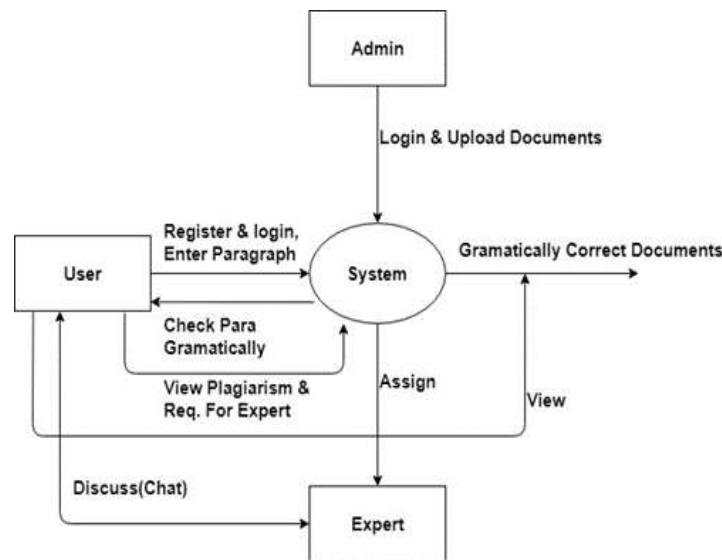

Fig 1. ER Diagram

Fig 2. Flow Chart Diagram

1. First the admin Logins and Uploads the documents.
2. Then the user Registers and Login and then enter the paragraphs.
3. System then automatically corrects the grammatical errors.
4. System assigns the Experts if necessary.
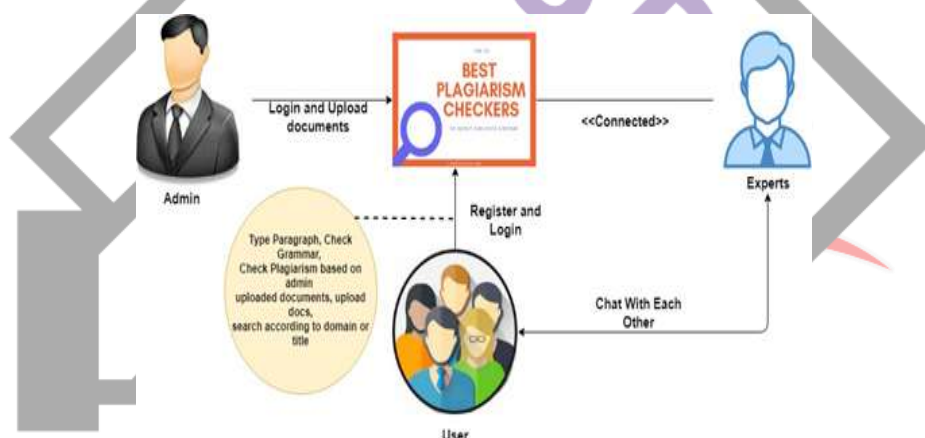5. View the corrected documents.

Fig 3. Architectural Diagram

For each word, in our dictionary we maintain the usage frequency. For the words, which are within a threshold edit distance of the given incorrect word, we look at the frequency of each suggested word. The word with higher frequency is presented higher in the suggestion list. That it gives efficient and accurate results. Plagiarism in text documents can be in several forms like plagiarized text may be entered one-tone, passages may be modified to a greater or lesser extent or they may be translated or it is act of claiming of information that actually user wrote. So, the focus of this paper is to give a plagiarism detection technique using semantic technology that will better catch the plagiarism. Till now several techniques for plagiarism analysis have been proposed. Match with the Tense Rules Using the noun type select the root in the rules, start matching the words in sentence against the next allowed values. If matched, proceed to next word and continue the process until we reach the leaf node in our rules tree. If we reach to a leaf node, the sentence is correct. If the current word does not match with any of the rules in next allowed values, the sentence is incorrect. Dictionary Lookup A dictionary is maintained, where each word should be matched with it. If the word is there in the dictionary then it is considered as correct and we simply increase its frequency by 1, but if not there in the dictionary then we will further proceed.

_ In proposed system, there are following modules;

_ Admin:
Admin can login on system and uploads documents of domain; Artificial Intelligence, Machine Learning, etc.

_ User:

User can register itself on system and then after get login. On system the user can type paragraph and check grammar of that paragraph. Check plagiarism based on uploaded documents by user. User can also upload the documents.

Also search papers or documents according to domain or title. User can also chat with expert according to grammatical suggestions.
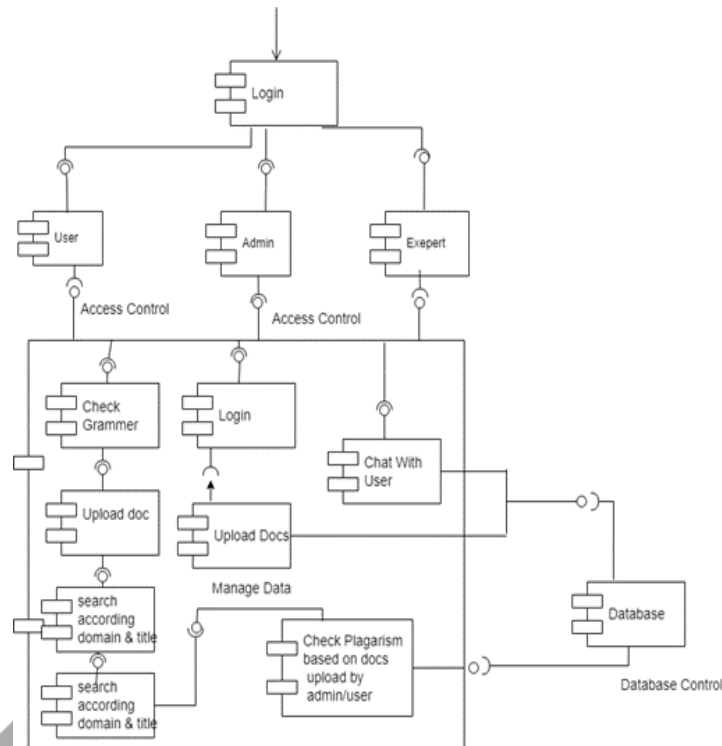

Fig 4. Component Diagram

A Component Design is a design specification for one of these Adaptable Components. Each component must be designed to satisfy relevant aspects of the Product Requirements and all design structures of the Product Architecture. A component diagram, also known as a UML component diagram, describes the organization and wiring of the physical components in a system Component diagrams are often drawn to help model implementation details and double- check that every aspect of the system's required function is covered by planned development. In the first version of UML, components included in these diagrams were physical: documents, database table, files, and executables, all physical elements with a location. In the world of UML 2, these components are less physical and more conceptual standalone design elements such as a business process that provides or requires interfaces to interact with other constructs in the system. The physical elements described in UML 1, like files and documents, are now referred to as artefacts. A UML 2 component may contain multiple physical artefacts if they naturally belong together.

VII.　　　　Summary and Conclusion

We have developed an English language processing system which comprises of a Grammar Checker. The grammar checker works on the rule-based approach whereas the checker is based on frequency-based approach. A smartest grammar checker can be made which along with Error Detection, Suggestion Prediction and Automatic error correction, also helps user to learn spellings. For instances, there are certain patterns that can't come together, so on observing the patterns of mistakes committed buy the users the application can give them appropriate suggestions so that user can learn spelling rules and can reduce spelling mistakes.

References

[1]　　Amit Sharna,P.J,(2013) Hindi Spell checker.

[2]　　Naber, D. (2003).A Rule-Based Style and Grammar Checker.

[3]　　Pratistha Mathur, N. G. (2012).Spell Checking Techniques in NLP: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering.

[4]　　Amanjot Kaur, Dr. Paramjeet Singh, Dr. Shaveta Rani(2012). Spell Checking and Error Correcting System for text paragraphs written in Punjabi Language using Hybrid approach.

[5]　　http://www-rohan.sdsu.edu/ ling354/n- grams.html

[6]　　http://www.json.org/ [7]https://en.wikipedia.org/wiki/English- speakingworld=E:K:Park;D:Y:Ra;andM:G:Jang; "Technique