

An approach for in-text citation of manuscript with computational Linguistic tools

¹ Devendra Das, ² Shika Pandey

¹ Research Scholar, M-Tech (Software Engineering), ² Assistant Professor,

¹ Computer Science & Engineering,

¹Rungta College of Engineering and Technology, Bhilai, India

Abstract— Due to improvements in the field of computer technology and due to the extensive use of Internet and increasing accessibility of resources the authenticity of the contents is a major issue. With the accessibility of the information on WWW and digital libraries, the authenticity of the content became one of the most important issues for universities, schools and researcher's fields. Earlier various methods have performed to check the authenticity but not a perfect way. The need of citations is extremely helpful to anyone who wants to find out more about your ideas and where they came from not all sources are good or right your own ideas may often be more accurate or interesting than those of your sources. In this paper we are going to analyze the cited text in the manuscript were actually used from the references or not. The cited texts are original or not

KEYWORDS: *In_text-citations, Type dependencies, Tagging, Relation Generation, String matching, cited text.*

I. INTRODUCTION

The aim of this paper is to study and analyze how to cite the spots in manuscripts from its references. The main idea behind this research work is to analyze and check the use of references in the main manuscript. Also to study the description of the grammatical relationships in a sentence that can easily be understood and effectively used by people without linguistic expertise who want to extract textual relations [1]. In text Citation analysis is a new technique which is used to compute quantitatively the value of document through arranging the citations in some kind of relationship and then check the similarity between two or sentences. It is also used to study and analyze the authenticity of research work is done on a particular subject or topic. "Citation analysis" refers to references in one text to another text with information on where that text can be found. Citation analysis reflects on citation practices.

More recent studies have shown, however, that this assumption is oversimplified. The reason or motivation for a citation matters

NEED OF CITATIONS IN MANUSCRIPT

- To cut and permit off (the ideas or words of another) as one's own.
- To use the information without accepting the source.
- To constraint literary theft.
- To present as fresh and original an idea or product derived from an existing source.
- Turning in somebody else's work as your own.
- Copying words or ideas from someone else without giving credit [3].
- Fading to put a citation in quotation marks.
- .Giving incorrect information about the source of a quotation.
- Changing words but copying the sentence structure of a source without giving credit
- Copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not.

DEFINITION OF CITATION, COMPUTATIONAL LINGUISTIC TOOLS

Citation is a way by which we tell the readers that certain materials in your work came from another source [2]. It also gives the essential information to find that source again.

A reference citation is the documentation needed to make your paper acceptable for academic purposes.

Computational linguistics is an interdisciplinary field concerned with the statistical or rule-based modeling of natural language from a computational perspective.

Development and need of Citation Analysis.

The development of citation analysis has been marked by the invention of new techniques and measures, the exploitation of new tools, and the study of different units of analysis [3].

1. Citation of a document implies use of that document by the citing author. This assumption actually has two parts: the author refers to all, or at least to the most important, documents used in the preparation of his work; and all documents listed were indeed used, i.e., the author refers to a document only if that document has contrib -buted to his work.

2. Citation of a document (author, journal, etc.) reflects the merit of that document. The underlying assumption in the use of citation counts as quality indicators is that there is a high positive correlation between the number of citations which a particular document (author, journal, etc.) receives and the quality of that document (author, journal, etc.).
3. A cited document is related in content to the citing document; if two documents are bibliographically coupled, they are related in content; and if two documents are cited, they are related in content.

METHODOLOGY

The methodology is based on the following techniques and methods:

Grammar-based method

The grammar-based method is one of the important techniques used for citing of contents of manuscripts in systematic way [1]. It focuses on the grammatical structure of documents, and this method uses a string-based matching approach to detect and to measure similarity between the documents. The grammar-based methods is suitable for detecting exact copy without any modification, but it's not suitable for detecting modified copied text by rewriting or switching some words that has the same meaning.

Tagging method

The term tagging or grammatical tagging is a process in which marking of word is done in text to its corresponding parts of speech with its position. Tagging is very important in the context of research work done as to check the word or phrase its position in the sentence with its relationship with adjacent and related words in an expression, sentence, or article. Following is the example how the tagging of a sentence is done.

Provide simple reference metadata due to the time and skill required to create more sophisticated metadata .
 Corresponding Tag
 provide_VB simple_JJ reference_NN metadata_NN due_JJ to_TO the_DT time_NN and_CC skill_NN required_VBN to_TO create_VB more_RBR sophisticated_JJ metadata_NP

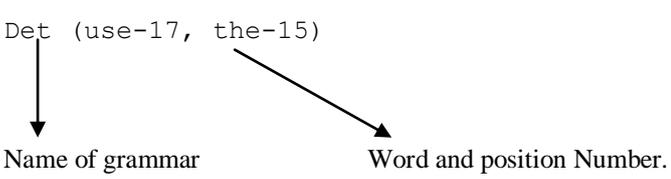
Type Dependency method

The typed dependencies are a way of representation of sentences and paragraphs [1]. It was designed by Stanford and they provided a simple description of the grammatical relationships in sentence. These relationships were easy to understand and effectively used by people without expertise knowledge to extract textual relationships. It displays the phrase representations that have long conquered in the computational linguistic community. It actually represents all sentence relationships consistently as typed dependency relations. It also represents as triples of a relation between pairs of words that may try to fill in next to the graphic .The dependencies map directly to a directed graph representation, in which words available with the sentence are nodes in the graph and grammatical relations are edge labels. The typed dependency given by Stanford group was designed to deliver a simple description of the grammatical relationships in words or phrases that could easily be understood and effectively used by people without linguistic expertise who wanted to extract textual relations. The representation was not designed for the purpose of parser evaluation but for proper knowledge illustration. The Stanford dependency presentation is an effective representation for the tasks proposed. It is perhaps closer to an appropriate task based evaluation than some of the alternative dependency representations available. The main purpose of representation was not to evaluate the parser but for other purposes also. The dependency-based evaluation of parsers provides a better understanding that how it actually works. The Stanford dependency representation is an effective representation of the sentences which displays the position of the word corresponding to the parts of speech.

Typed Dependency Representation Style
 POS (Word-Position, word-position)

Example: det(use-17, the-15)
 amod(use-17, widespread-16)

Det (use-17, the-15)



↓
Name of grammar
↘
Word and position Number.

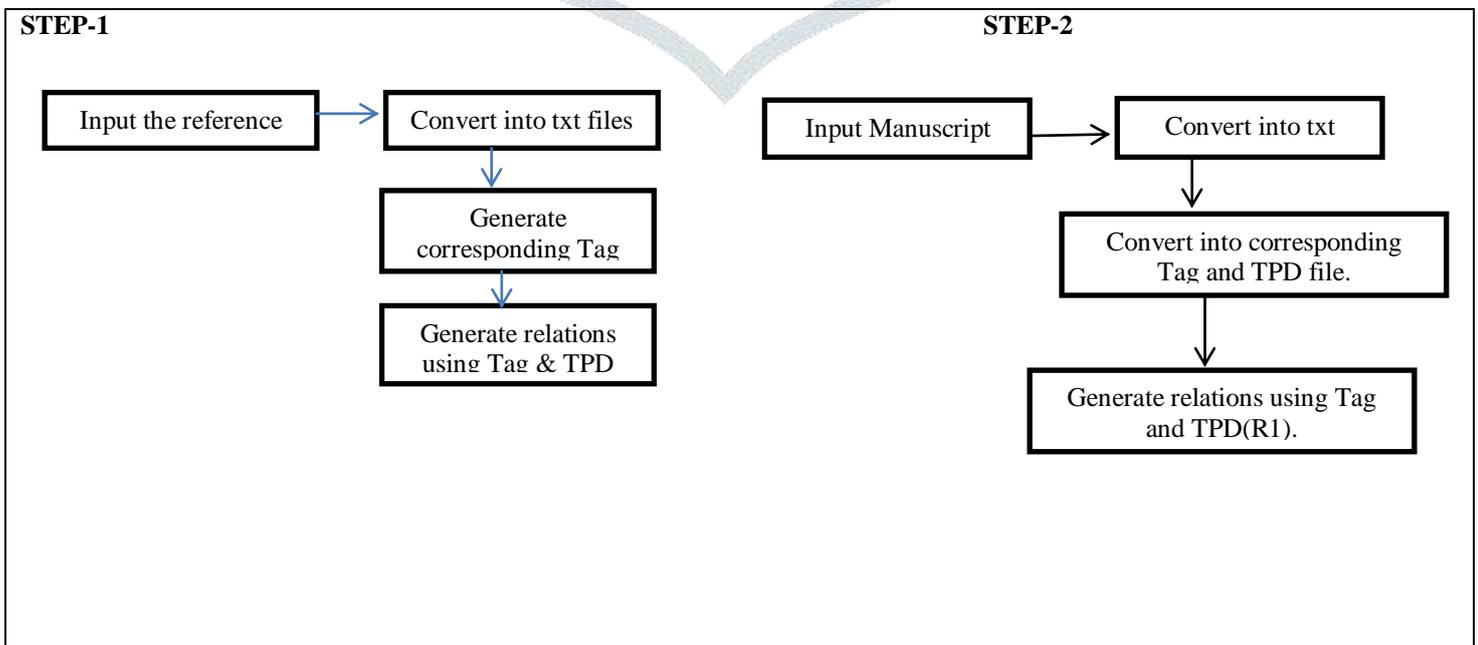
Here is an example sentence:

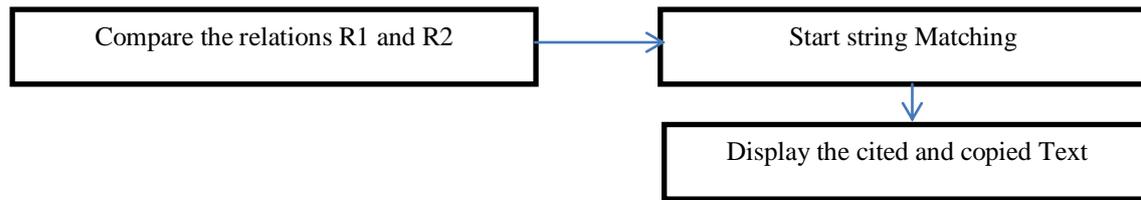
Sample Sentence
Introduction: The aim of this paper is to study and analyze how to cite the spots in manuscripts from its references.
 Corresponding Type dependency-
 (INTRODUCTION-2, I-1)
 nsubj(is-8, INTRODUCTION-2)
 det(aim-4, The-3)
 dep(INTRODUCTION-2, aim-4)
 prep(INTRODUCTION-2, of-5)
 det(paper-7, this-6)
 pobj(of-5, paper-7)
 root(ROOT-0, is-8)
 aux(study-10, to-9)
 xcomp(is-8, study-10)
 cc(study-10, and-11)
 conj(study-10, analyze-12)
 advmod(cite-15, how-13)
 aux(cite-15, To-14)
 ccomp(study-10, cite-15)
 det(spots-17, the-16)
 dobj(cite-15, spots-17)
 prep(cite-15, in-18)
 pobj(in-18, manuscripts-19)
 prep(manuscripts-19, from-20)
 poss(references-22, its-21)
 pobj(from-20, references-22)

OVERALL METHODOLOGY

We have used Stanford parser to get tagged values of each sentence and generate their associate tpd files. Our application generates a relation between necessary phrases by using this tag and tpd. Further we have implemented a simple text matching via pattern matching approach in this generated relationship, to check copied contents .

Firstly we take the master documents that are reference papers and generate their tag and tpd and relations. After that we input the manuscript papers that have to be checked. We perform same series of tasks given above with these documents. After getting the relations generated we perform a matching operation that checks each word with their associated relationship in series. Final output is generated using the comparative study of relationships between cited text of the references paper and the input manuscript document.



STEP-3**RESULT**

The final outcome of the research work is that we get the cited text which is taken from the references mentioned in the input manuscript. The cited text is grammatically checked and is very useful to the reviewers who are able to check the amount of genuine research work is done.

In this study the citation detection was considered as it is one of the most publicized forms of text reuse and checks the authenticity of the contents around us today.

LIMITATIONS AND FUTURE WORK

. In particular, it has been shown in this study how the citations can be handled using different techniques and tools. However, there are still some weaknesses and shortages in these techniques and tools which will affect the success of citation detection significantly. Following are some limitations.

1. All pdf files should be in standard format.
2. All the reference files should be available.
3. The initial cited sentences should be in the form ... [1].

The future scope of this project is that the in-text citations can be implemented for all the formats of citrated text. To check the authenticity of the contents it can be implemented various other fields also. As the grammatical nature of the sentence can't be modified therefore it can be a useful way to check originality of the contents in the other fields also.

REFERENCES

- [1] Marie-Catherine de Marne, Christopher D. Manning, "Stanford typed dependencies manual I Software, September 2008, Stanford Parser v. 3.3.
- [2] Dr. K.Kumar, T.Raghunadha Reddy," Citation Analysis of dissertations submitted to the department of library and information science "venkateswara university,,triputi International Journal of Digital Library Services,vol 1.4,april-june 2012,IssN:2250.1142.
- [3] Xiaozhong Liu,Jinsong Zhang,Chun Guo" Full-Text Citation Analysis: A New Method to Enhance Scholarly Network",School of Library and Information Science, Indiana University Bloomington
- [4] Michael A. Covington," A Fundamental Algorithm for Dependency Parsing", Artificial Intelligence Center The University of Georgia Athens, GA 30602-7415 U.S.A...
- [5] Michael A. Covington," A Free-Word-Order Dependency Parser in Prolog", Artificial Intelligence Center The University of Georgia Athens, GA 30602-7415 U.S.A
- [6] Michael A. Covington," ET: an Efficient Tokenizer in ISO Prolog", Artificial Intelligence Center the University of Georgia Athens, Georgia 30602-7415 U.S.A.
- [7] Asim M. El Tahir Ali, Hussam M. Dahwa Abdulla, Vaclav Snase," Survey of Plagiarism Detection Methods", Department of Computer Science,2011 Fifth Asia Modelling Symposium.
- [8] Michael A. Covington,"Important Additional Notes about Dependency Parsing",April 15, 2004,Artificial Intelligence Center The University of Georgia Athens, Georgia 30602-7415,