

Chapter 23

Plagiarism Detection in Marathi Language Using Semantic Analysis

Ramesh Ram Naik

Dr. B.A.M. University, India

Maheshkumar B. Landge

Dr. B.A.M. University, India

Namrata Mahender C.

Dr. B.A.M. University, India

ABSTRACT

In this article, the authors have proposed a method to detect plagiarism in the Marathi language by using semantic analysis. Nowadays, plagiarism is a challenging task in educational and research fields. Currently, there are some tools available to detect the plagiarism on the basis of similarity of words. But there is no tool available to detect the plagiarism semantically. In this article, the authors have applied preprocessing to a database i.e. tokenization, removed stop words and punctuations, for the goal of calculating the frequency of words. Then searching the same word or synonyms of words in wordnet to detect the semantic plagiarism. It is useful for many researchers who are working in this domain.

1. INTRODUCTION

Plagiarism is defined as “the re-use of someone else’s prior ideas, processes, results, or words without explicitly acknowledging the original author and source” (Barrón-Cedeno & Rosso, 2010). There are mainly two methods of plagiarism detection i) Extrinsic or External plagiarism detection and ii) Intrinsic or internal plagiarism detection. Intrinsic plagiarism detection analyses the input document only to find some parts which are not written by the same author without performing comparisons to external corpus. External plagiarism detection needs a reference collection of documents which are assumed to be genuine. A suspicious document is compared to all the documents in this collection to find duplicates or near duplicates fragments in source documents (Mahdavi et al., 2014). Semantic similarity plays an

DOI: 10.4018/978-1-5225-8057-7.ch023

important role in natural language processing, information retrieval, text summarization, text categorization, text clustering and so on. Many semantic similarity measures have been proposed. In general, all the measures can be grouped into four classes: path length-based measures, information content-based measures, feature based measures, and hybrid measures (Meng, Huang & Gu, 2013).

1.1. Types of Plagiarism

- **Copy and Paste Plagiarism:** This refers to directly picking up content from a viable source and put it in one's own research paper with or without citing the source appropriately or providing credit to the original author and declaring the work to be one's own (Weber-Wulff, 2010).
- **Disguised Plagiarism:** It subsumes practices intended to mask copied segments (Lancaster, 2003).
- **Contractive Plagiarism:** It describes the summary or trimming of copied material (Lancaster, 2003).
- **Expansive Plagiarism:** It refers to the insertion of additional text into or in addition to copied segments (Barnbaum, 2002).
- **Mosaic Plagiarism:** Patchwork paraphrasing refers to obtaining content from a various sources catering to the same topic of interest and rephrasing the sentences, switching words, using synonyms and improvising on the grammar styles to finally producing one's own research paper without citing the sources (Weber-Wulff, 2010, Lancaster, 2003)
- **Paraphrasing Plagiarism:** Paraphrasing generally refers to using the idea from only one specific source but switching words, changing sentence constructions, improvising on grammar styles and using synonyms for the words wherever possible for the work to look one's own or legitimate (Lancaster, 2003).
- **Metaphor Plagiarism:** Metaphors are used either to make an idea clearer or give the reader an analogy that touches the senses or emotions better than a plain description of the object or process. Metaphors, then, are an important part of an author's creative style (Barnbaum, 2002; Liles & Rozalski, 2004).
- **Idea Plagiarism:** If one copies an innovative idea or a solution provided by another author in a source document, whilst one cannot provide a solution or an idea of his own, the idea plagiarism is said to have occurred. The research paper authors have a hard time distinguishing the ideas and/or solutions provided by the author of the source paper from public domain information. Public domain information is any idea or solution about which people in the field accept as general knowledge (Maurer, Kappe & Zaka, 2006).
- **Self-Plagiarism:** Here the author of the research paper reuses his own previous work to produce a new work (Bretag & Mahmud, 2009).

1.2. Approaches of Plagiarism Detection

General plagiarism detection Approaches can be classified into following:

1. **Structural-Based Methods:** These methods focused on lexical features of the text in the document. Many studies that proposed different methods in the field of plagiarism detection focused on

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/plagiarism-detection-in-marathi-language-using-semantic-analysis/222322

Related Content

Teaching Materials for Civil Engineers in a Slightly Different Way

Jonathan Oti (2019). *Technology-Supported Teaching and Research Methods for Educators* (pp. 111-127).

www.irma-international.org/chapter/teaching-materials-for-civil-engineers-in-a-slightly-different-way/213061

Diversity, Equity, and Inclusion for Learners With Diverse Needs: An Action Research Approach

Lloyd Dan Tlale (2018). *Cross-Disciplinary Approaches to Action Research and Action Learning* (pp. 271-285).

www.irma-international.org/chapter/diversity-equity-and-inclusion-for-learners-with-diverse-needs/190343

Comparing the Behaviour of Two Topic-Modelling Algorithms in COVID-19 Vaccination Tweets: LDA vs. LSA

Jordan Thomas Bignell, Georgios Chantziplakis and Alireza Daneshkhah (2022). *International Journal of Strategic Engineering* (pp. 1-20).

www.irma-international.org/article/comparing-the-behaviour-of-two-topic-modelling-algorithms-in-covid-19-vaccination-tweets/292445

How Can Human Resource Management Help the Theory of Constraints

Brian J. Galli (2019). *International Journal of Strategic Engineering* (pp. 1-13).

www.irma-international.org/article/how-can-human-resource-management-help-the-theory-of-constraints/219320

Avoiding Project Failure and Achieving Project Success in NHS IT System Projects in the United Kingdom

Carol Matirangana Verner and Dilshad Sarwar (2021). *International Journal of Strategic Engineering* (pp. 33-54).

www.irma-international.org/article/avoiding-project-failure-and-achieving-project-success-in-nhs-it-system-projects-in-the-united-kingdom/269716