Abstract

In today’s world internet is the answer to every question. So at any time one can easily copy the content from web and use it. This is known as plagiarism. It is growing now days. Usually in plagiarism people reword the documents, copy them, do not give references. It is difficult to detect plagiarism as people rephrase the text do not copy it directly. To detect plagiarism apache lucene have been used. Firstly indexing of the original document is done and then used cosine similarity to compare the plagiarised document with set of documents which are there saved previously.

References

3. Asim M. El Tahir Ali, Hussam M. Dahwa Abdulla, and V´aclav Sn´aˇsel ,"Overview and
4. Daniele Anzelmi, Domenico Carlone, Fabio Rizzello, Robert Thomsen, D. M. Akbar
Hussain,"Plagiarism Detection Based on SCAM Algorithm", Proceedings of the International
MultiConference of Engineers and Computer Scientists, March 2011
5. Bela Gipp Norman Meuschke ,"Citation Pattern Matching Algorithms for Citation-based
Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation
Sequence", Mountain View, CA, USA, September 2011
7. Romans Lukashenko, Vita Graudina, Janis Grundspenkis, "Computer-Based Plagiarism
Detection Methods and Tools: An Overview", International Conference on Computer Systems
and Technologies - CompSysTech’07, 2007
8. Reena Kharat, Preeti M. Chavan, Vaibhav Jadhav, Kuldeep Rakibe,"Semantically
Detecting Plagiarism for Research Papers", International Journal of Engineering Research and
Applications (IJERA), May-Jun 2013

Index Terms

Computer Science Pattern Recognition

Keywords

Apache Lucene, Indexing, Cosine similarity, Plagiarism