Abstract

Digital Forensic Investigation is the branch of scientific forensic process for investigation of material found in digital devices related to computer crimes. Digital evidence analogous to particular incident is any digital data that provides hypothesis about incident. The essential part of Digital forensic Process is to analyze the documents present on suspect's computer. Due to increasing count of documents and larger size of storage devices makes very difficult to analyze the documents on computer. To overcome these problems, a subject based semantic document clustering algorithm along with bisecting-kmeans has been proposed that allows the examiner to analyze and cluster the documents based on particular subject and also the terms that does not belong to any subject. The accuracy of clustering of documents has been improved by means of this hybrid approach.

References

- J. Becker, D. Kuropka, Topic-based Vector Space Model, Proceedings of the 6th
- R. T. Ng, J. Han, Efficient and effective clustering methods for spatial data mining, Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc. , San Francisco, CA, USA, 1994.
- Gaby G. Dagher, Benjamin C. M. Fung, Subject-based semantic document clustering for digital forensic investigations, Data & Knowledge Engineering 86 (2013) 224–241
- M. F. Porter. The Porter Stemming Algorithm. www. tartus. org/martin/PorterStemmer

Index Terms

Computer Science  Information Sciences
**Keywords**

Digital Forensic  Stemming  Term Importance  Subject –based semantic clustering

Document-subject similarity