



**International Conference on Latest Trends in Engineering,  
Management, Humanities, Science & Technology (ICLTEMHST -2022)  
27<sup>th</sup> November, 2022, Guwahati, Assam, India.**

**CERTIFICATE NO : ICLTEMHST /2022/C11221026**

**A COMPARATIVE STUDY ENHANCING XML DOCUMENT  
CLUSTERING PERFORMANCE WITH DELTA COMPRESSION**

**WANJARI RAVINDRA SHANKAR**

Research Scholar, Department of Computer Science  
Kalinga University, Naya Raipur, Chhattisgarh, India.

**ABSTRACT**

In order to update clustering solutions without completely decompressing documents, this work offers a time-efficient method for document clustering that uses compressed delta representations. Using known distances before modifications and a set of changes stored in the compressed delta, the proposed approach reassesses pairwise distances between documents. The technique is tested using XML documents of different sizes from a data source with an average depth of four layers, and it is implemented in Java. Rather of decompressing the changes that are responsible for document versions, compressed delta saves them. Test findings demonstrate that the suggested method outperforms FDC in terms of time efficiency, particularly when evaluating new distances between document versions, and that the compression methodology considerably decreases document sizes. According to the findings, the suggested approach significantly reduces processing time without sacrificing clustering precision.

***Keywords: Documents, Clustering, Time, Compression, Efficiency***