

It is widely accepted that, with large databases, the key to good performance is effective data-clustering. In any large document database clustering is essential for efficient search, browse and therefore retrieval. Cluster analysis allows the identification of groups, or clusters, of similar objects in multi-dimensional space [1]. Conventional document retrieval systems involve the matching of a query against individual documents, whereas a clustered search compares a query with clusters of documents, thereby achieving efficient retrieval. In most document databases, periodic updating of clusters is required due to the dynamic nature of a database. Experimental evidence, however, shows that clustered searches are substantially less effective than conventional searches of corresponding non-clustered documents. In this paper, we investigate the present clustering criteria and its drawbacks. We propose a new approach to clustering and justify the reasons why this new approach should be tested and (if proved beneficial) adopted. © 1998 Elsevier Science B.V. All rights reserved.