

# Web Document Clustering using Genetic Algorithm Based Fuzzy Clustering

**May Tha Zin, Khin Mar Soe**

*University of Computer Studies, Yangon,  
maythazin.mtz15@gmail.com*

## **Abstract**

*Clustering is a typical unsupervised learning technique for grouping similar data points. In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. Fuzzy clustering algorithm can be optimized by genetic algorithms, which are popular evolutionary algorithm and can be used to search for large search space. This paper used genetic algorithm based fuzzy clustering for news web page clustering. Web documents are preprocessed and features are selected from the web page content, and they are clustered by the genetic algorithm based fuzzy clustering algorithm. The experimental result with various features is presented.*