

## **Abstract**

Topic directories such as the Open Directory Project (dmoz.org) and About.com rival search engines as some of the most popular websites. The ultimate goal is to organize the entire web hierarchically by topic. Such a structure allows for more focused searching. The process of classifying web documents and associating them with topics is called 'tagging'. Brown et al explored the use of machine learning and data mining tools for automatic tagging [1]. Specifically, their work looked into performance of classification algorithms including the nearest neighbor algorithm and decision tree learning. In this paper, we compare the effectiveness of three clustering algorithms including k-means clustering, EM clustering algorithm, and COBWEB for automatic tagging of web documents. We will compare the performance of these clustering algorithms as well as present their performance relative to the classifiers.