

Document Categorizer Agent based on ACM Hierarchy

Abstrak

As the number of research papers increases, the need for academic categorizer system becomes crucial. This is to help academicians organize their research papers into predefined categories based on the documents' content similarity. This paper presents the Document Categorizer Agent based on ACM CCS (Association for Computing Machinery Computing Classification System). First, we studied the ACM categories hierarchy. Next, based on these categories, we retrieved our corpus from ACM DL (ACM Digital Library) to train our Categorizer Agent using a popular machine learning technique called Naïve Bayes Classifier. We used two types of training data for the corpus namely, negative training data and positive training data. Next, these papers are categorized according to their content based on the same training data. We tested our Document Categorizer Agent on a number of academic papers to test its accuracy. The result we obtained showed promising results