

Using clustering in web classification

Nguyen Thi Hien

Abstract: In the context of the increasingly large and complex amount of information on the Internet, the problem of information management and organization becomes urgent. This topic focuses on applying clustering techniques to perform automatic web classification. The goal of the research is to develop a system based on machine learning methods to improve accuracy in the process of classifying web pages. The method combines communication network theory with experiments using unsupervised clustering algorithms. The expected outcome of this topic is not only to provide a useful solution for information management on the Internet but also to contribute to data science and artificial intelligence.

Keywords: clustering, web, machine learning, information management, communication networks