

Information Retrieval

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20528522&lokasi=lokal>

Abstrak

Machine Learning (ML) algorithms have opened up new possibilities for the acquisition and processing of documents in Information Retrieval (IR) systems. Indeed, it is now possible to automate several labor-intensive tasks related to documents such as categorization and entity extraction. Consequently, the application of machine learning techniques for various large-scale IR tasks has gathered significant research interest in both the ML and IR communities. This tutorial provides a reference summary of our research in applying machine learning techniques to diverse tasks in Digital Libraries (DL). Digital library portals are specialized IR systems that work on collections of documents related to particular domains. We focus on open-access, scientific digital libraries such as CiteSeerx, which involve several crawling, ranking, content analysis, and metadata extraction tasks. We elaborate on the challenges involved in these tasks and highlight how machine learning methods can successfully address these challenges.