

司法案例資訊檢索之研究

邱登裕, 黃小閑

資訊管理學系

資訊學院

chiuden@chu.edu.tw

摘要

Due to the huge quantity of judicial case data, it is difficult to find required information. This study focuses on document clustering and document abstract in order to help users find the required information quickly. Document clustering is used to compute the similarity to group the similar documents in a cluster. First, vector space model is used to compute the similarity of documents for document clustering. Secondly, probabilistic model is used to group similar documents in a cluster. To retrieve abstract of documents, the method is based on keywords. The number of the keywords in each paragraph is computed. The paragraph with maximum number of keywords is used to represent the document. The experimental results are valuable. The best result occurs when 700 keywords are used for the document clustering and the top 5 keywords are used to retrieve document abstract.

關鍵字：Document Clustering; Document Abstract; Document Similarity; Vector Space Model; Probabilistic Model