

디렉토리 서비스를 위한 문서분류 알고리즘에 관한 연구 <A Study on the Document Classification Algorithm for Directory Service >

박 동 진* , 김 인 환**

mispdj@kongju.ac.kr , voca@kongju.ac.kr

*공주대학교 산업시스템공학과 부교수 , **공주대학교 산업시스템공학과 대학원

* 011-9874-2281 , ** 011-9414-9376,

요 약

대부분의 검색 엔진은 특정 분야만을 위한 것이 아니므로 대용량의 데이터베이스라 할 지라도 분야별 정보의 양은 절대적으로 부족할 뿐 아니라 제공되는 문서 자체도 구체적이지 않다. 특정 기관 및 대학에서 관련 분야의 정보만을 위한 디렉토리형 검색 엔진이 제공되고 있으나, 단지 관련기관의 URL을 연결해 주는 서비스거나 제한적인 내용밖에 포함되어 있지 않다. 또한 분야의 전문가가 웹 문서를 일일이 수작업으로 분류를 해서 저장해야 하므로 폭발적으로 증가하는 문서의 분류가 적절하게 되지 못한다. 따라서 특정 분야만을 지원하는 검색시스템이라 할지라도 문서의 일부만 그리고 오래된 문서만이 분류되어 서비스되는 곳이 많다

본 연구에서는 특정 분야를 위한 디렉토리형 검색 엔진(specialized search engine) 설계의 프레임워크를 먼저 제시하였다. 이러한 검색 엔진은 웹사이트로부터 수집한 관련 분야의 웹 문서를 즉시 분류하고, 데이터베이스에 저장하여, 주제어 및 주제별로 검색이 가능하게 한다. 아울러 본 연구에서는 검색엔진의 핵심적 역할을 하는 자동화된 문서분류 서비스 시스템의 구조 및 세부 설계를 중점적으로 다룬다. 그리고 본 연구를 통해서 제안된 Feature Weighting TFIDF 방식을 농업정보 서비스를 위한 문서들의 분류에 적용하여 테스트 해 보았다.

발표희망분야

지식경영