

불리언 질의어 재구성에서 초기 입력자료의 변화에 따른 의사결정나무의 학습 성능 검토 분석

(Sensitivity Analysis of Boolean Query Reformulation using Decision Tree with different initial input sets)

윤정미*, 김남호**, 권영식*

ABSTRACT

인터넷상에 의한 정보검색환경의 변화로 많은 양의 정보 축적의 용이해짐에 따라 사용자가 원하는 정보를 효과적으로 검색하는 것이 중요한 문제로 제기되고 있다. 정보검색시스템을 사용하는 사용자는 불명확한 정보요구문을 갖고 여러 번의 시행착오를 거쳐 질의어를 만들어 나간다. 그러나 이러한 모든 과정에는 사용자가 질의어를 직접 재구성해야하는 어려움, 사용자와 시스템간의 용어 불일치 등 많은 문제점이 대재되어 있다.

기존연구들은, 시소러스를 이용하여 질의어를 확장하거나 축약이나 질의어에 가중치를 부여하여 보다 적합한 문헌을 검색하는 방법이 제기되어 왔다. 이런 방법은 보다 적합한 용어를 선택할 수 있지만 사용자가 직접 질의어를 구성해야 하므로 사용자가 초기에 지정한 질의어를 구성하는 것이 어려운 단점이 있다. 저널에는 하이퍼텍스트를 이용한 네비게이션을 통해 사용자가 원하는 정보를 찾는 연구도 행해지고 있지만, 네비게이션과정에서 발생하는 단절적 과부담과 정보에 접근하는 과정이 길다는 단점이 있다.

이런 문제점을 해결하기 위해서는 사용자가 직면하는 비합리적 혹은 비합적하지 않은 정보들을 기초로 질의 부분을 학습하여 사용자에게 편리하고 효율적인 지능적 질의어 재구성 기법을 제공하는 것이 바람직하다. 본 연구에서는 질의어 재구성을 비합적한(비합적하지 않은) 문서를 분류하는 분류문제로 보고 간단한 학습 알고리즘인 ID3을 적용한다. 정보검색문제에서는 원하는 정보를 단 한번에 검색하기 어렵기 때문에 여러 번에 걸쳐 정보를 검색하게 된다. 이 경우 사용자의 패드백을 학습할 수 있는 ID3 알고리즘이 효과적이다.

실험은 불리언 정보검색 시스템 하에서 사용자의 패드백에 의한 예제를 중심으로 ID3 알고리즘을 이용하여 행해지며, 초기 입력 자료와 사용자의 패드백 자료에서 나온 단계 질의어 재구성을 위한 입력 자료의 수를 변화시킬 때의 학습 성능 검토 분석을 한다.

이 연구를 통하여 사용자의 직합성 패드백에 의한 예제를 학습하는 지능형 정보검색시스템을 구현함으로써, 사용자가 시스템은 이용함에 있어서 그 편의성의 증가와 보다 적합한 문서 검색을 기대할 수 있다.

* 동국대학교

** 전북대학교