

LMO 정보검색용 지능형 Web Agent 개발

민홍기*

인천광역시 남구 도화동 177 인천대학교 정보통신공학과

Objectives

오늘날의 시대를 정보화시대라고 부르고 있다. 정보화 사회란 무수한 정보가 대량으로 범람하는 정보홍수의 피해를 겪고 있는 현대사회에서 우리는 특정 주제에 있어서 의미의 틀에 평가받을 수 있는 조직화된 정보의 접근을 원하고 있다. 특정정보의 입수과정에서 사회가 고도화될수록 생산되는 정보량도 많아지기 때문에 필요한 지식을 필요한 시기에 필요한 량만을 입수한다는 것은 쉬운 일이 아니다.

정보검색(Information Retrieval)이란 수집된 정보 또는 정보자료의 내용을 분석한 뒤 적절히 가공하여 축적해 놓은 정보화일로부터, 이용자의 정보요구에 적합한 정보를 탐색하여 찾아내는 일련의 과정을 의미한다. 즉 기하급수적으로 생성, 증가되는 정보를 정보전문가, 주제분석가, 색인작성가 등에 의하여 분석, 가공, 축적되어 데이터베이스와 같은 축적 매체에서부터 이용자의 요구에 적합한 정보를 탐색하는 일련의 과정으로서 축적된 정보에서 이용자가 만족하는 특정 속성을 가진 정보를 선택하는 것을 의미한다.

정보검색시스템은 수요가 예측되는 정보를 수집, 분석, 축적하고 찾기 쉬운 형태로 가공해 두었다가 이용자의 정보요구가 발생하면 적합한 정보를 검색하여 제공하는 시스템이다.

그러나 특정분야의 정보검색시스템은 다양한 정보를 검색하기 보다는 특정자료만을 검색하게 되며, 대부분 범용의 정보검색시스템을 사용하기가 어렵기 때문에 별도로 개발된 정보검색시스템을 사용한다. 본 고의 주제인 LMO 정보검색시스템도 이런 부류에 속한다.

범용 LMO 정보검색시스템은 현재 몇 개의 시스템이 개발되어 운용되고 있다. 이중에 하나는 Agbios사에서 개발한 시스템이다. Agbios는 생명공학 제품의 관리와 위험평가 전문기술을 제공하기 위하여 설립된

캐나다 회사로서, 이 회사의 검색사이트의 URL은 <http://www.agbios.com/dbase.php> 이다. 질의어 입력창에 전부 혹은 일부의 검색어를 입력하여 원하는 정보를 검색할 수 있다. 질의어의 정확성 및 개수에 따라 검색 결과는 매우 다양한 형태를 보인다.

가장 보편적으로 사용하는 검색 사이트는 BioTrack Database이다. 이 검색 사이트의 URL은 <http://webdoninal.oecd.org/biotrack.nsf>이며 처음에는 OECD국가에서 보고되는 LMO field test 결과를 저장한 데이터베이스이지만 현재는 UNIDO에서 개발한 BINAS를 포함하고 있다. 일반적으로 범용 LMO정보검색시스템을 사용하는 경우 개발된 사이트에 따라서 접근 방식, 질의어의 종류, 인터페이스 방식 및 디스플레이 등이 다양하고 또한 검색 결과의 양이 매우 많아질 수 있다. 따라서 가장 적절한 질의어를 결정한 후 이 질의어를 모든 사이트에 적용 가능한 전용 LMO 정보검색시스템, 즉 Web Agent의 개발이 필요하다.

개발한 전용 LMO 정보검색시스템은 색인어 추출부와 질의어 검색부로 구성되어있다. 색인어 추출부는 야외방출 사이트에서 주제어에 해당하는 낱말을 자동으로 판단하여 색인어로 결정하는 부분이며 질의어 검색부는 질의어를 분석하여 해당문서를 검색하는 부분이다. 색인(Index)은 지식 기록물인 문서의 지적내용과 물리적 위치에 대한 안내이며 지시자나 안내자라 할 수 있다. 따라서 해외 LMO DB 검색에서는 색인을 통하여 지적내용에 접근한다. 검색방법에는 단일검색방법과 다중검색방법이 있다.

단일검색방법은 LMO DB의 내용 중 Country/ Organism Common Name/ Gene/ Trait/ Year/ Location 등의 키워드중 하나를 이용하여 검색하는 방법이다. 다중검색방법은 여러 개의 키워드를 이용하여 검색이 가능한 시스템이다.