

멀티에이전트 시스템 기반 학술 정보 검색 시스템에 관한 연구

용한마로*, 박윤정*, 구민오*, 조나연*, 경민기*, 민덕기*⁽¹⁾
*건국대학교 컴퓨터공학과

e-mail : {maro1076, sm6280p, happykus, nycho, moonend,
dkmin}@konkuk.ac.kr

Study on Academic Information Retrieval System based on Multi Agent System

Hanmaro Yong*, Yunjung Park*, Mino Ku*, Na-yun Cho*,
Mingi Kyung*, Dugki Min*⁽¹⁾

School of Computer Science and Engineering, Konkuk University*

요 약

학술 정보의 결정체라 할 수 있는 논문 정보에 대한 검색을 할 때 기존의 학술 정보 검색 시스템에서는 키워드와의 일치성을 기반으로 학술 정보 검색 시스템을 구축할 수 있지만, 검색 사용자가 지정한 키워드와 유사한 논문 정보에 대해서는 검색 할 수 없었다. 본 논문에서는 온톨로지 서버를 이용하여 각 학술 분야에 대한 의미를 바탕으로 검색시 특정 학술 분야 검색에 대한 “깊이와 넓이”의 연관성을 충족하는 검색 시스템에 대해 설명한다.

1. 서론

인터넷의 등장으로 인해 정보로의 접근도는 극도로 높아졌다. 이러한 정보 접근도의 상승은 전자적 속도로 인해 지리적 한계를 뛰어넘는 넓이와 빠른 접근 속도에 대한 기술적 업적을 이루게 하여 학술 분야 간의 융합을 이루며 새로운 차세대 시장을 형성하며 인간의 사회를 더욱더 의미 있는 수준을 가능하게 한다. 그러나 이러한 지리적 한계를 초월한 범위적 측면과 특정 학술 분야에 대한 빠른 접근을 가능하게 하는 지금의 인터넷 기술은 유사 및 동일한 의미의 다른 여타 지식을 형성하지 않을 경우 접근할 수 없어 학문의 깊이를 형성하는 학생과 학자들에게 보다 많은 주제와 연관어에 대한 검색 작업을 요구 하고 있다.

특히 학술 정보의 결정체라 할 수 있는 논문 정보에 대한 검색을 할 때 논문 저자가 지정한 키워드는 해당 논문을 함축하고 있는 연관도가 밀접한 키워드임은 분명하다. 따라서 기존의 학술 정보 검색 시스템에서는 이러한 키워드와의 일치성을 기반으로 학술 정보 검색 시스템을 구축할 수 있지만 검색 사용자가 지정한 키워드와 유사한 논문 정보에 대해서는 검색 할 수 없는 치명적인 약점을 가지고 있다.

본 논문에서는 키워드간의 밀접한 연관성을 가지고 있는

모든 키워드를 추출하여 연관성을 가진 논문 정보를 추출하여 접근도가 높은 논문 정보를 찾는 방법에 대한 설계를 설명한다. 또한, 본 논문에서는 온톨로지 서버를 이용하여 각 학술 분야에 대한 의미의 고리를 바탕으로 검색시 특정 학술 분야 검색에 대한 “깊이와 넓이”의 연관성을 충족하는 검색 시스템을 구축했다.

2. 관련연구

현재, 상용화된 검색 시스템들은 질의어를 구성할 때 검색자가 입력한 키워드를 기반으로 처리하고 있다. 질의 키워드의 빈도수가 높은 문서를 검색하기 때문에 관련된 문서를 다수 찾아 주고 있지만, 검색된 문서가 실제로 사용자의 요구에 맞는지에 대해서는 보장 할 수 없다[1][2]. 예를들어, 검색 사용자가 “분산 컴퓨팅”이라는 주제를 바탕으로 검색을 할 때, 검색 엔진은 “분산”, “컴퓨팅”, “분산 컴퓨팅”이라는 3가지 키워드를 형성하여 검색을 하지만, 사실 “분산 컴퓨팅”은 그 역사적 기술적 배경으로 봤을 때 “클러스터링”, “LVS”, “MultiLayer Switch”, “하드웨어 분산 컴퓨팅”등 수많은 키워드와 밀접한 연관성을 가지고 있어 이 모든 키워드를 내포한 질의어를 구성하여 “분산 컴퓨팅”에 접근도가 높은 논문 정보를 찾을 수 있어야 한다.

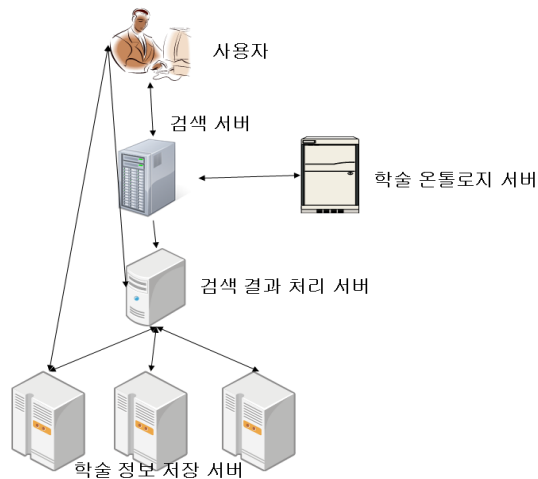
(1) 교신저자 : 민덕기 (e-mail : dkmin@konkuk.ac.kr)

- 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과 (NIPA-C1090-0902-0026) 및 IT성장동력기술개발사업[2008-S-007-01, 차량 전장용 통합제어 SW플랫폼 개발]의 일환으로 수행하였음.

3. 시스템 설계

본 연구에서 구성된 시스템은 “검색서버”, “학술 온톨로지 서버”, “검색 결과 처리 서버”, “학술 정보 저장 서버” 총 4개의 서버가 필요하다.

검색서버에서는 사용자의 검색 요청 받아들이는 개체로써, 사용자가 검색을 요청시 사용자가 요청한 키워드를 획득하여 이를 기반으로 “검색 요청 키워드군”을 형성하여



(그림 1) 시스템 아키텍처

온톨로지 서버에게 요청하면, 검색 결과로 받은 “검색 대상 키워드군”을 기반으로 “검색 결과 처리 서버”에게 검색 작업 요청시 활용한다.

학술 온톨로지 서버에서는 검색 서버가 넘겨 준 키워드를 기반으로 연관어를 검색할 때 기본 깊이를 3으로 하여서 연관된 단어들의 목록을 추출하여 “검색 대상 키워드군”을 형성한다. “검색 요청 키워드군”과 “검색 대상 키워드군”을 기반으로 “검색 요청 키워드군”을 “핵심 검색어”, “검색 대상 키워드군”을 “연관 검색어”로 구분을 하여 각각의 검색어군에 가중치를 부여하여 검색식을 형성한다. 그리고 이를 다시 “검색 서버”에게 그 결과를 넘겨준다.

검색 결과 처리 서버에서는 검색 서버로부터 넘겨 받은 검색식을 기반으로 “학술 정보 저장 서버”들에게 검색을 요청 및 그 결과를 취합하여 사용자에게 논문 목록을 제공하는 역할이 본 서버의 주된 역할이다.

학술 정보 저장 서버에서의 목적은 논문 정보를 저장하고 있으며, “검색 결과 처리 서버”로부터 넘겨 받은 검색식을 기반으로 검색을 수행함으로써 사용자에게 제공할 목록을 제공해 주는 역할을 한다.

현 시스템에서 키워드 추출과정은 사용자는 “분산 컴퓨팅”이라는 검색 문구를 기반으로 “검색서버”에게 검색을 요청한다. “검색 서버”는 사용자로부터 받은 “분산 컴퓨팅”이라는 문구를 “분산”, “컴퓨팅”이라는 키워드로 분리하여 “학술 온톨로지 서버”에게 연관 키워드 목록을 요청한다. “학술 온톨로지 서버”는 본 서버의 온톨로지를 기반으로 “검색서버”로부터 넘겨 받은 “분산”과 “컴퓨팅”이라

는 키워드를 기반으로 온톨로지 검색을 시도한다. 그 결과 “분산”의 경우 “클러스터링”, “로드밸런싱”, “L4 Switch”, “그리드 컴퓨팅”이라는 연관 키워드를 “컴퓨팅”의 경우 “시스템”, “소프트웨어”, “아키텍처”의 연관 키워드를 확보한다.

연관 키워드 역시 Sector를 형성하여 “분산”을 통해서는 Sector1, “컴퓨팅”을 통해서는 Sector2와 같이 구분한다. “학술 온톨로지 서버”는 학술 온톨로지 과정을 통해서는 얻은 2개의 핵심 키워드(분산, 컴퓨팅)와 연관 키워드(클러스터링, 로드밸런싱, L4 Switch, 그리드 컴퓨팅, 시스템, 소프트웨어, 아키텍처)를 검색 서버에게 전달을 한다. 검색 서버는 학술 온톨로지 서버를 통해 확보한 핵심 키워드와 연관 키워드 개수에 대해 경우의 수 조합을 통해 형성하여 검색식을 확보하여, “학술 정보 저장 서버”에 전달하여 검색 작업을 요청한다. “학술 정보 저장 서버”는 “검색 결과 처리 서버”로부터 전달된 검색식을 기반으로 자신의 학술 DB를 검색한 뒤, 연관도를 포함한 결과를 목록화 하여 “검색 결과 처리 서버”에 전달한다. “검색 결과 처리 서버”는 “학술 정보 저장 서버”로부터 넘겨받은 검색 결과를 다시 서버의 IP, 포트 번호 등의 “서버 정보”와 함께 결합하여 사용자에게 검색 결과를 제공한다. 사용자는 “검색 결과 처리 서버”로부터 전달받은 검색 결과를 바탕으로 특정 논문에 대해 “전송 요청”을 하고 “학술 정보 저장 서버”는 이에 상응하는 논문 전송 작업을 시행한다.

4. 결론 및 향후 과제

본 연구결과 질의어를 의미적으로 처리함으로써 질의 요구를 반영한 검색을 할 수 있어 키워드 기반의 검색보다 더욱 높은 접근도로 검색 할 수 있지만, 자연어 질의의 요구를 정확하게 반영하지는 못한다.

따라서, 향후 과제로는 현재 구축되어 있는 연관 키워드 검색뿐만 아니라, 자연어 질의에 대한 요구를 반영하여 보다 높은 접근도로 검색을 연구할 계획이다.

참고문헌

- [1] 한용진, 박세영, 이영화, 김권양, 이벤트 온톨로지 기반의 의미 정보 검색, 정보과학회논문지 : 컴퓨팅의 실제 및 레터 제 14권 제 1 호(2008.2)
- [2] 이인근, 서석태, 권순학, 온톨로지 구축 프로세스와 시스템 : 퍼지 및 지능시스템학회 논문지 2006. Vol. 16. No. 6. pp. 721~729
- [3] 경북대학교, 온톨로지 검증 및 온톨로지 기반 인스턴스 생성에 관한 연구
- [4] FIPA (The Foundation for Intelligent Physical Agents) <http://www.fipa.org/>