

분산 다중 에이전트 웹 검색 시스템 설계 및 구현

김기용*, 한현구
한국의외국어대학교 컴퓨터공학과

Design and Implementation of a Distributed Multi-Agent Web Search System

Gi-Yong Kim*, Hyungoo Han
Dept. of Computer Science and Engineering, Hankuk University of Foreign Studies

요약

본 논문에서는 사용자를 대할 수 있는 분산 다중 에이전트 기술을 이용한 웹 문서 검색 시스템을 설계 및 구현한다. 분산환경에서의 에이전트기술은 문서의 근접도를 평가할 때 기존의 웹 검색 엔진 보다 복잡한 알고리즘 적용이 가능하며, 이로써 좀더 정확한 문서들을 사용자에게 제공 할 수 있다. 또한 분산환경으로 네트워크의 부하를 분산시킴으로써, 빠르게 자료검색 결과를 얻을 수 있는 장점이 있다. 분산된 시스템에서 에이전트들의 효율적인 통신을 위하여 프로토콜 레이어를 구현하였으며, 에이전트들의 상호 협동과 의사 소통 및 중재를 위한 하부구조(infrastructure)를 구축하여 웹 문서검색을 효과적으로 하고, 사용자 이동성을 가능하게 하였다.

1. 서론

정보 검색은 급성장 하는 인터넷에서 어떻게 하면 정보의 보고인 인터넷으로부터 원하는 정보를 빠르고 정확하게 그리고 손쉽게 찾아낼 수 있는가 하는 문제로 초점이 맞추어지고 있다. 그러나, 정보의 바다라 불리는 인터넷 상에서 이러한 요구를 충족시킨다는 것은 쉬운 일이 아니다. 원하는 정보를 쉽게 찾기위한 방법중 하나인 웹 검색 시스템들은 대부분 사용자에게 제목이나 요약 등의 리스트를 제공함으로써 사용자가 각 문서들을 하나하나 확인하고, 방문하여 각각의 문서가 적절한가 평가하는 방법을 사용하고 있다. 그리고 웹 검색 엔진이 일반적으로 낮은 정확도를 갖기 때문에 사용자가 적합한 문서를 찾는 데 많은 시간과 노력이 필요하다. [1]

현재 위와같은 사용자 요구를 바탕으로 검색 톨과 검색 엔진이 개발되어 상용화되고 있다. 검색 엔진이나 디렉토리 서비스는 정확한 검색 요구와 사용의 불편함, 많은 정보처리 요구 등으로 인해 좀 더 지능적인 검색 방법이 필요하게 되었다. [2]

본 논문에서 제시하는 정보검색 방법은 사용자를 대리할 수 있는 에이전트 기술과 서버를 분산시킴으로써 네트워크 환경에 대한 제약성을 줄임으로써, 인터넷 상의 광대한 정보 바다에서 사용자가 원하는 문서를 빠르고 정확하게 찾는 것을 목적으로 시스템 설계 및 구현하였다.

2. 관련연구

2.1 에이전트 시스템

에이전트 개념은 오래전부터 인공지능에서 연구되어온 분야로 1990년대 초부터 독립된 분야로 인식되기 시작한 개념이다. 에이전트는 여러 학자들에 의해 개념이 정립되었다. 대표적으로 S.Russell과 P.Norving은 에이전트 시스템이 자신의 감각기관을 통해 환경을 인지하여 작용기를 통하여 그 환경에 대해 반응하는 시스템이라고 하였다. [3]

에이전트는 약한 의미와 강한 의미 두 가지 형태로 나누어 정의 할 수 있다. 약한 의미로서의 에이전트는 자체의 행위 및 내부 상태에 대한 제어를 하는 자치성, 다른 에이전트 또는 사람과 에이전트 통신언어를 통하여 상호작용을 하는 사회

성, 환경의 변화에 대하여 적절한 반응을 하는 반응성, 그리고 목적 지향적인 행위를 하는 주도적 능동성이란 속성을 가지고 있는 하드웨어 또는 소프트웨어 컴퓨터 시스템을 의미한다.

그리고, 강한 의미의 에이전트는 상기의 특성 외에도 인간에게 적용될 수 있는 자가학습 등의 여러 개념들이 구현된 컴퓨터 시스템을 의미한다. [4]

2.2 언어 분석 기술

웹 정보는 다양한 멀티미디어 데이터로 되어있지만 정보검색의 초점은 주로 문자로 표현된 정보에 맞추어져 있다. 일반적으로 정보검색은 사용자의 질의를 분석하여 사용자가 원하는 정보를 찾아주는 것을 말한다. 정보 검색 시스템은 질의어를 분석하고 사용자의 의도와 검색 대상이 되는 정보가 얼마나 일치하는지 그 유사성을 판단해야 한다. 그러나, 투자에 비해 효과가 적을 뿐만 아니라 사용자 질의에 대한 응답시간 제약이 있기 때문에 분석 수준이 낮은 단순한 패턴 매칭을 사용하고 있다

본 논문에서 제안하는 다중 에이전트를 이용한 정보 검색은 기존의 정보검색 시스템의 부하를 다수의 에이전트가 분산하여 수행함으로써 시간적인 제약을 줄일 수 있기 때문에 복잡한 알고리즘의 적용이 가능하며, 더욱 정확한 문서 검색이 가능하다.

2.3 검색 엔진

검색 엔진은 로봇 검색엔진, 주제별 검색 엔진, 메타 검색엔진과 같이 세가지로 크게 구분할 수 있다. [5]

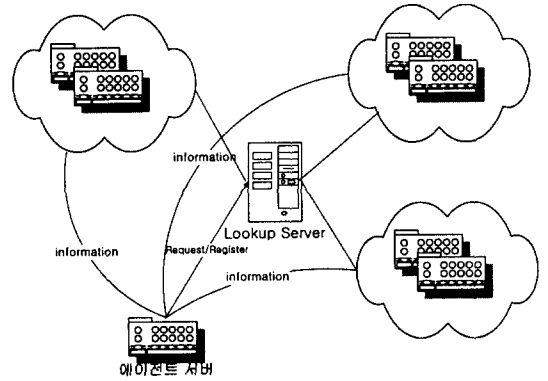
로봇 검색 엔진은 로봇 에이전트라 불리는 검색 프로그램이 웹을 자동적으로 돌아다니면서 자료를 수집하여 데이터 베이스를 구축한다. 사용자는 로봇이 구축한 데이터 베이스를 검색을 하게 되는데, 다량의 자료를 검색할 수 있는 잇점이 있으나, 정확도가 떨어지는 단점으로 사용자가 원하는 자료가 아닌 전혀 다른 결과를 보여 주기도 한다.

주제별 검색엔진은 자료들을 주제어나 카테고리별로 구분하여 분류하고, 설명 및 평가를 덧붙여 데이터 베이스를 구축한 검색 엔진을 말한다. 사람이 직접 주제어나 카테고리를 찾아

입력을 하므로 검색에 대한 신뢰도가 높으나 많은 시간과 인력이 요구되며, 정보의 양이 로봇 검색 엔진에 비하여 상대적으로 적다.

메타 검색 엔진은 사용자의 질의를 받아서 이를 여러 다른 검색 엔진에 병렬적으로 질의하여 검색된 내용을 취합한후 사용자에게 보여주는 방식의 검색 엔진이다. 다양한 결과를 보여 줄 수 있으나, 각 검색 엔진에 의존적이며, 검색 시간이 더 길어지는 단점을 가지고 있다.

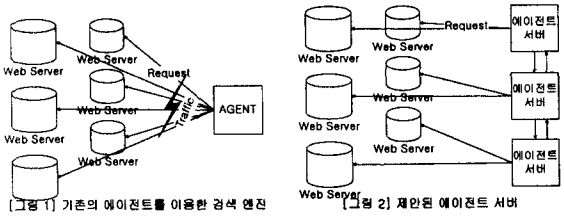
에이전트를 이용하는 로봇 검색엔진이나 메타 검색엔진은 고정되고 한정된 시스템에서 에이전트를 구동하는 방식을 사용하고 있다. 이는 에이전트를 구동하는 시스템 주위의 네트워크 송수신량을 크게 증가시키며, 또한 고정된 대역폭으로 인하여 한정된 에이전트만이 활동할 수밖에 없는 환경을 야기시키는 문제점이 있다.



[그림 3] 에이전트 서버와 Lookup Server구성도

3. 연구 개발

3.1 분산환경



[그림 1] 기존의 에이전트를 이용한 검색 엔진

[그림 2] 제안된 에이전트 서버

기존의 검색엔진의 문제점을 해결하기 위하여 에이전트 서버를 구성하고, 이 에이전트 서버를 여러 시스템에 분산 시킴으로써 네트워크 부하를 줄일수 있다. 또한 각각의 에이전트 서버는 다른 네트워크 상에 위치 함으로써 대역폭으로 인한 활동 에이전트 수에 대한 제약성을 줄일 수 있도록 구성하였다. 그러나 분산된 에이전트들이 상호 협동 및 의사교환을 통한 공동작업 하기 위해서는 프로토콜과 하부구조에 관한 연구가 필요하다. 본 논문에서는 효율적인 통신을 위하여 KQML 및 스크립트를 이용한 프로토콜을 정의, 구현 하였으며, 에이전트들의 상호 협동 및 의사교환을 위한 하부구조를 구축 함으로써 가능하게 하였다.

3.2 시스템 구성

[그림 3]은 시스템의 전체적인 구성을 나타낸다. 새로운 에이전트 서버가 실행되면, 먼저 Lookup Server를 참조하여 다른 에이전트 서버들을 인식하여 하부구조를 구축한다. 인식된 에이전트 서버들간에는 프로토콜을 통하여 정보들을 주고 받는다.

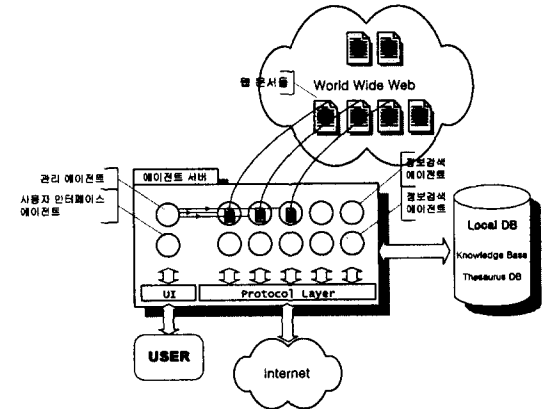
1) Lookup Server

초기 에이전트 서버가 기반구조를 구축하기 위하여 다른 에이전트 서버들의 위치를 참조하게 되는데 이때 참조하는 서버가 Lookup Server이다.

Lookup Server는 에이전트 서버들의 위치 정보를 저장하고 있으며, 새로운 에이전트 서버가 생성되면, 위치 정보를 기록하고, 다른 에이전트 서버들의 위치 정보를 선택적으로 전달한다.

2) 에이전트 서버

에이전트 서버는 자신의 서버내의 에이전트의 생성과 소멸, 작업의 진행을 직접적으로 제어한다.[그림 4]는 에이전트 서버의 구성을 나타낸다. 하나의 에이전트 서버에는 목적에 따른



[그림 4] 에이전트 서버 구성도

여러 에이전트들과 사용자가 검색을 할 수 있는 인터페이스 그리고 Local Database 들을 가지고 있다.

본 논문에서 에이전트 서버는 각각의 에이전트들의 상호 협동과 통신을 하기위한 프로토콜 레이어를 설계 및 구현 하였다. 다른 에이전트 서버들과 프로토콜 레이어를 통하여 지역이나 원거리 통신에 상관없이 통신에 대한 투명성을 제공하며, 에이전트간 하부구조를 구축하게 된다.

3) 하부구조와 에이전트의 통신

Lookup Server를 통하여 다른 에이전트 서버들을 인지 할수 있다. 본 논문에서는 하부구조를 구축 함으로써, 인지된 여러 개의 에이전트 서버들의 정보검색 에이전트는 모두 하나의 에이전트 서버 안에서 동작하는것과 같은 환경을 제공하며, 인지된 에이전트 서버들의 관리 에이전트들이 협동과 상호작용을 통하여 작업을 분배한다.

분산 환경하에서 에이전트들의 통신상 투명성을 보장하기 위해 에이전트 서버가 제공하는 프로토콜 레이어를 통하여 에이전트는 서로 통신을 한다. 프로토콜 레이어는 TCP/IP 계층 상위 개념에 추상화된 레이어이다. 본 논문에서는 소프트웨어 에이전트통신을 위한 규격인 KQML[6]을 수용하였으며, 웹 검색을 위한 스크립트를 추가하였다. 이러한 확장된 프로토콜을 이용하여 에이전트끼리 효율적인 상호 대화나 협력을 할 수 있다.

4) 사용자 인터페이스 에이전트

사용자 인터페이스 에이전트는 사용자와 대화형식으로 질의를 받으며, 정보 검색 에이전트가 제시하는 자료검색 결과 정보를 사용자에게 보여준다. 또한, 사용자 인터페이스 에이전트

는 다른 관리 에이전트와 정보 검색 에이전트의 요구에 의한 사용자와의 상호 작용을 하는 에이전트이다. 만약 사용자가 제시한 질의가 너무 평이한 경우 정보 수집 에이전트는 사용자 질의에 의하여 수집된 문서들의 범위를 축소하기 위한 후보 질의어(candidate query)들을 설정하며, 사용자 인터페이스 에이전트는 이를 받아 사용자에게 제시하여 준다. 사용자가 쿼리 선택한 후보 질의어는 검색 결과를 가지고 있는 에이전트 서버들에게 보내지며, 후보 질의어에 의하여 검색된 자료들은 추출된다.

5) 정보 검색 에이전트

정보 검색 에이전트는 관리 에이전트의 명령을 받아 문서를 웹 서버로부터 읽어들인다. 여기서 문서에 대한 의미있는 링크, 핵심어 또는 링크상에 요약된 문장들을 추출하고, 이를 이용하여 Local Database 에 자료를 저장한다.

또한, 정보검색 에이전트는 문서의 근접도 알고리즘을 적용하여 각 문서가 사용자가 원하는 문서에 얼마나 가까운지를 평가하여 Local Database 에 저장한다.

6) 관리 에이전트

관리 에이전트는 정보 검색 에이전트들에게 작업을 분배 및 할당하고, 필요한 웹 문서를 정보 검색 에이전트를 이용하여 가져오게 한다. 이렇게 함으로써 불필요한 링크들을 방문하지 않아도 되며, 적절한 수준의 깊이로 링크를 따라 방문할 수 있다. 또한, 정보 검색 에이전트가 돌아온 Local Database 상의 정보를 필요한 경우 다시 사용자 조건에 의하여 추출하며, 추가로 필요한 문서를 검색하게 한다.

사용자가 초기 질의 입력시 설정한 우선순위에 따라 정보 검색 에이전트의 일을 분배하며, 사용자가 입력한 시간 제약 조건에 따라 검색을 중단시키는 역할을 한다.

7) 데몬(Demon) 에이전트

데몬 에이전트는 다른 에이전트 서버의 관리 에이전트의 질의에 대한 적합한 에이전트를 찾아 연결 시켜 주는 역할을 하는 에이전트이다. 이는 특정 작업을 수행할 수 있는 에이전트에게 자료를 요청하거나 새로운 자료를 갱신할 때 두 에이전트간 통신을 성립시켜 준다. 데몬 에이전트는 자신의 에이전트 서버 네트워크 대역폭을 인지하며, 다른 에이전트 서버와 네트워크 부하 정보를 가지고 있다. 이 정보는 관리 에이전트가 작업을 배분할 때 참고하게 된다.

또한 사용자 이동성을 지원하기 위하여 데몬 에이전트는 서로 사용자 정보를 교환하는 역할을 한다.

8) Local Database

각 에이전트 서버는 독립된 데이터 베이스를 갖는다. 이 데이터베이스는 다음의 3 가지 주된 데이터로 구성된다.

• 웹 문서 데이터

정보 검색 에이전트가 웹 문서를 검색하여, 임시적으로 저장하며, 문서의 내용, 문서의 근접도, 문서의 요약정보와 기타 정보들이 저장된다. 이 데이터는 나중에 사용자 인터페이스 에이전트에 의해 재 평가되거나 추출되어 사용자에게 제공된다.

• Knowledge Base

각 사이트에 관한 정보 및, 유효한 HTML 형태에 관한 정보를 제공한다. 각 사이트의 로그인이나, CGI 구동에 관한 일정 RULE 들을 스크립트 형태로 제공하게 되며, 정보 검색 에이전트는 이를 활용하여 정보를 성공적으로 얻어낼 수 있다. 이러한 Knowledge 는 인터페이스 에이전트에 의하여 사용자에게 요청된다.

• Thesaurus Database

사용자의 검색요청에 대하여 확장하거나 일반적인 용어로 치환하기 위한 데이터베이스이다. 예를 들어 사용자가 입력한

'쿼리'라는 단어는 '쿼리', '질의' 또는 'query' 등으로 나누어져 검색함으로써 좀더 정확한 결과를 유도하도록 한다.

3.3 사용자 이동성

본 논문에서는 사용자가 질의 명령을 내린후 사용자가 다른 에이전트 서버에서 결과를 받아 볼수 있도록 구현하였다. 이것은 과거에 사용자가 한 장소에서 웹 검색으로 시간을 소비하였던것에 비하여, 에이전트를 사용하여 장소에 구애 받지 않고 사용자를 대리하여 문서를 찾게 함으로써 사용자 입장에서의 문서검색 효율성을 높였다. 사용자 이동성은 다음과 같은 두가지 경우를 살펴볼 수 있다.

- A 지점의 에이전트 서버를 이용하여 검색 명령을 내린 후 B 지점의 에이전트 서버에서 결과를 받아보는 경우
- 검색명령을 내린 에이전트 서버의 네트워크 단절이나 서버를 Shutdown 시킨경우.

에이전트 서버내의 사용자 인터페이스 에이전트는 사용자의 검색에 대하여 질의 고유 번호(QID)를 부여하게 된다. 위의 두가지 사용자 이동성의 전자의 경우 사용자는 이 고유 번호를 기억하고 있다가 인터페이스 에이전트에게 요청을 하면 결과를 얻을 수 있다. 후자의 경우 네트워크 단절이 된 시점부터 다른 서버가 검색 결과 모으는 일을 대신 수행하게 되며, 후에 사용자가 검색 결과를 요청하여 검색 결과를 넘겨받게 된다.

4. 결론 및 향후 연구계획

본 논문에서 제안한 분산 다중 에이전트 검색 시스템에서는 네트워크 부하를 분산시킴으로써 에이전트들의 서버 접근 시간을 최소화 시킬 수 있도록 하였다. 또한 에이전트 관리자의 명령 계층구조는 빠르고 효율적으로 일을 분업화 시킬수 있다. 실행 결과 에이전트 서버의 수가 많을수록 또, 여러 네트워크 클라우드에 에이전트 서버가 분산되어 있을때, 검색 속도가 증가됨을 알 수 있었다.

또한, 네트워크 전송 속도가 낮은 환경(전화 접속이나 느린 전용선)인 경우 기존의 웹 검색방법은 느린 검색 속도로 어려움이 있으나 본 논문에서 제시하는 기술은 분산된 에이전트 서버와의 협동으로 빠른 네트워크 상에 있는 에이전트들에게 웹 문서 검색 명령을 내림으로써 보다 빠르게 검색 결과를 얻을 수 있다.

앞으로 사용자가 원하는 문서를 보다 정확하게 찾기 위하여 사용자 profile, 자연어 처리등을 이용하여 더욱 정확한 문서 근접도 알고리즘 개발이 필요하다.

[참고문헌]

[1] Oren Zamir and Etizioni, "Web Document Clustering : A Fresibility Demonstration", SIGIR98, 1998
 [2] V. Gudivada and S. Tolety, "A Multiagent Architecture for Information Retrieval on the World-Wide Web", Proc of 5th RIAO conf. Paris, 1997.
 [3] R. Russell and P.Norving, "Artificial Intelligence: A Mordern Approach", Prentice Hall International Inc., 1995.
 [4] Walter Brenner, Rüdiger Zarnekow and Hartmut Wittig, "Intelligent Software Agents", Springer, 1998
 [5] Oren Zamir and Etizioni, "Moving up the Information Food Chain: Deploying Softbot on the Web". In Proc of 14th Nat. Conf. On AI, 1996
 [6] FIPA '97 Version 2, "Agent Management", 1998