

순차적 전달방식(STM)을 이용한 분산정보검색시스템의 설계 및 알고리즘 구현

윤 희 병* · 김 용 한** · 김 화 수***

요 약

분산검색시스템은 중재자 또는 메타검색엔진에 의한 중앙통제로 인하여 중앙 시스템에 트래픽이 집중된다. 또한 중앙통제를 위한 복잡한 알고리즘의 설계와 별도의 하드웨어 설치 등의 비용 증가 문제가 발생한다. 따라서 이러한 문제를 해결하기 위해 여러 검색시스템이 서로 독립적인 검색기능을 가지면서도 중앙의 통제를 받지 않으면서 상호 협조할 수 있는 방식이 필요하다. 본 논문에서는 분산검색시스템과 관련된 여러 연구를 살펴보고 다중의 정보검색시스템이 중앙의 통제를 받지 않는 순차적 전달방식(STM)을 이용한 분산정보검색시스템의 Framework를 설계하며 이와 관련된 알고리즘을 구현한다. 이를 위해 먼저 웹을 논리적으로 분할하여 관리하는 웹 분할 정책을 제시하며, 검색시스템의 수에 따른 순차적인 질의 처리방식을 각각 예제를 통해 설명한다. Framework에 적합한 정보검색시스템의 3 Layer구조와 각 Layer별 기능과 모듈을 또한 제시한다. 마지막으로, STM 알고리즘의 효율적 구현을 위하여 모듈 구조를 분석하고 이의 pseudocode 표현을 제시하며, 서버간 순차적 질의 전달 시연과정을 통해 제안한 STM 알고리즘이 효율적으로 동작함을 보여준다.

Design and Algorithm Implementation of a Distributed Information Retrieval System using Sequential Transferring Method(STM)

HeeByung Yoon* · YongHan Kim** · HwaSoo Kim***

ABSTRACT

The distributed Information Retrieval System centrally controlled by mediator or meta search engine result in congestion of heavy traffic and int he problem of increment of cost for the reason of the design of complicated algorithm for central control and installation of hardware. So to figure out this problem, the way is needed that has independent retrieval functionality and can cooperate each other without dependency. In this paper, we overview a few works involved in distributed information retrieval system, then, implement algorithm and design the framework of distributed information retrieval system using sequential transferring method(STM) including multiple information retrieval system separated from central control. For this first of all, we present a web partition policy which devide and manage web logically and we present the sequential query processing way by means of illustration through changing numbered information retrieval system. Then, we also present 3-layered structure of framework and function and module of each layer suitable for information retrieval system. Last of all, for effective implementation of STM algorithm we analysis module structure and present description of pseudocode of this, and show that the proposed STM algorithm works smoothly by demonstration of sequential query transfer process between servers.

키워드 : 분산(Distributed), 정보검색시스템(Information Retrieval System), 순차적 전달방식(Sequential Transferring Method), 3 Layer, STM, 웹 분할(Web Partition), 유효시간(Valid Time), 협력적(Cooperative)

1. 서 론

1969년 미국방성의 ARPANET에서 시작된 인터넷은 Tim Berners Lee에 의해 WWW으로 변하게 되었고 이 변화는 '정보의 공유'와 '정보의 홍수'라는 새로운 경향을 만들어 냈다. 이러한 경향에 따라 사용자들은 웹상에 있는 수많은 정보 중에서 적절하고 자신의 요구에 맞는 정보를 찾기 위

해서 많은 시간을 소비해야 했고 이를 해결하기 위해 정보 검색시스템에 대한 수많은 연구가 행하여졌다.

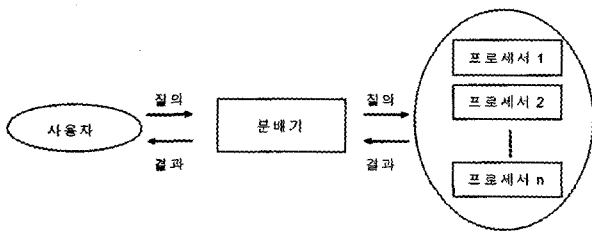
인터넷의 정보검색시스템은 일반적으로 클라이언트-서버 구조를 이루고 있다. 이러한 형태의 대표적인 구조로 ALTA-VISTA 구조를 들 수 있다. ALTA-VISTA 구조는 웹상에서 공유되고 있는 대량의 문서를 수집하여 자신의 서버에 저장하고, 이를 사용자의 질의에 적절히 응답할 수 있는 형태로 색인한 뒤, 사용자의 질의에 결과를 제시하는 구조를 가지고 있다[1].

최근에는 웹 규모의 증가와 광역화로 인하여 더 이상 한

* 종신회원 : 국방대학교 전산정보학과 교수
** 준 회 원 : 국방대학교 대학원 전산정보학과
*** 정 회 원 : 아주대학교 정보통신대학원 교수
논문접수 : 2004년 5월 17일, 심사완료 : 2004년 8월 11일

곳에 모든 정보를 모아두기 어려워지게 되자 분산정보검색에 대한 연구가 활발히 수행되고 있다. 분산정보검색의 한 가지 방법으로 병렬정보검색이 있다. 병렬정보검색은 단일 문제 해결을 위한 멀티프로세서의 동시적인 응용을 말한다. (그림 1)은 병렬정보검색시스템의 구조를 보여준다. 사용자의 질의를 분배기가 받아서 여러 프로세서들에게 나누어 주고 다시 종합하는 구조로 이루어져 있다.

분산정보검색은 단일 문제 해결을 위해 다중 컴퓨터들을 네트워크로 연결한 것으로 이는 다시 검색엔진 수준에서의 통합과 메타검색 방식의 통합으로 구분된다.

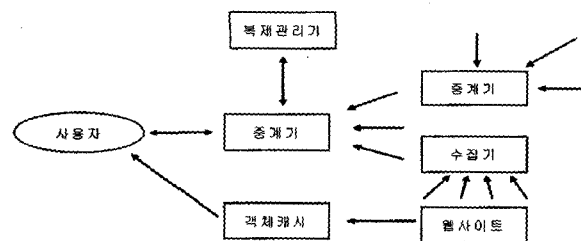


(그림 1) 병렬정보검색시스템의 구조

검색엔진 수준에서의 통합이란 정보검색엔진이 제공하는 고유의 검색 API를 각 검색엔진이 공유하여 시스템 수준에서 연결시키는 것이다. 고유 검색 API는 서로 다른 구분형식을 갖고 있기 때문에 이를 통합하기 위해서는 Z39.50과 같은 표준 프로토콜이나 자체 내부 규약을 이용하게 된다[3].

메타검색엔진을 이용하면 전 세계 웹 페이지의 42%까지 검색할 수 있다는 연구보고가 발표되었다[4]. 메타검색엔진은 자신의 데이터베이스가 없고 다른 여러 검색엔진에 사용자의 질의를 의뢰하여 그 결과를 종합하는 서비스를 말한다. 서로 다른 검색엔진의 결과를 종합하기 때문에 이질적 결과의 종합을 위한 알고리즘 개발이 중요하다.

분산정보검색시스템의 대표적인 형태인 Harvest 구조가 (그림 2)에 나타나 있다. Harvest 구조는 중계기와 수집기의 두 가지 중요한 요소를 사용하였다[5].



(그림 2) Harvest 구조

위의 구조에서 데이터는 중계기와 수집기에 의해 중앙 관리되는 형태를 취하고 있다는 것을 알 수 있다. 이런 시스템에 대한 중앙 집중적인 관리는 트래픽 집중과 별도의 서버 설치 등의 문제가 발생한다.

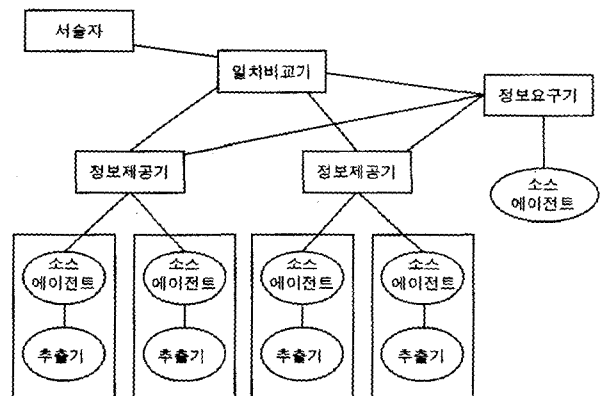
따라서 본 논문에서는 분산정보검색시스템과 관련된 여러 구조에 대해 살펴보고, 분산정보검색시스템이 중앙 관리되지 않으면서 서로 독립적으로 동작하고 상호 협력할 수 있는 협력적 Framework를 설계한다. 그런 다음 이 Framework에 적합한 검색시스템의 3 Layer 구조와 기능을 제시한다. 분산정보검색시스템간의 협력 작업을 위한 질의 전달 방식으로 순차적인 전달방식인 STM(Sequential Transferring Method)을 제안하고 이에 대한 모듈구조 분석 및 알고리즘을 구현하였다.

2. 관련 연구

본 장에서는 분산정보검색시스템의 구조와 관련된 여러 연구를 살펴본다. 분산검색시스템에서 질의를 어떻게 분배하고 결과를 종합하는지에 대해 살펴보고, 구조적 특징과 기능에 대해 알아본다.

2.1 분산검색을 위한 멀티에이전트 구조

분산검색을 위한 한 가지 방법으로 멀티에이전트를 사용하는 방법이 있다. (그림 3)은 일반적인 멀티에이전트 구조를 보여준다[6]. 에이전트간 통신은 Knowledge Query Manipulation Language(KQML)을 이용하였다[7]. 사용자 에이전트와 정보소스 에이전트를 두고 서로 협력하면서 정보를 찾는 방식이다. 분산정보검색 에이전트는 사람이나 에이전트 클라이언트로부터 요청 수신, 정보소스가 이해할 수 있는 언어로 요청 번역, 요청 관련 정보를 갖고 있는 정보소스 식별, 정보소스로 요청 요구, 부합하는 결과 수집, 돌아온 결과 처리, 그리고 클라이언트에 결과를 제시하는 기능을 가지고 있다.



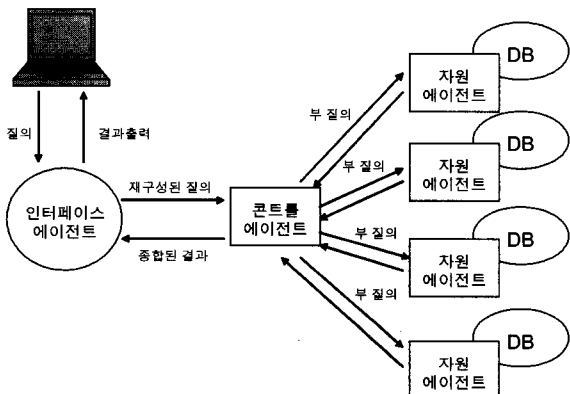
(그림 3) 웹에서 분산검색을 위한 멀티에이전트 구조

2.2 분산시스템을 위한 정보수집 에이전트

정보수집 에이전트는 정보 수집 및 종합을 위해 멀티에이전트를 이용한 또 다른 변형으로 인터페이스 에이전트, 컨트롤 에이전트 그리고 다수의 자원 에이전트로 구성되었으

며 (그림 4)에 그 구조가 도시되어 있다.

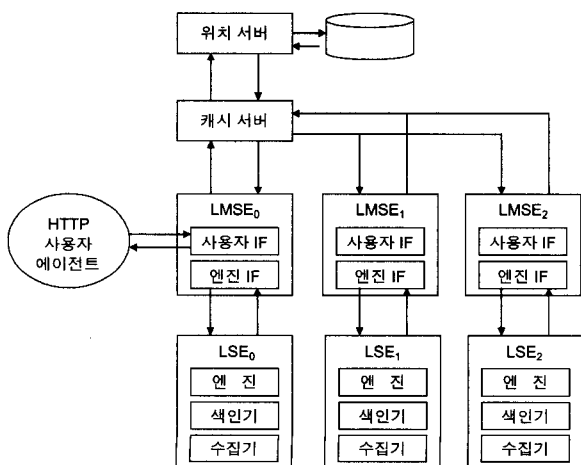
인터페이스 에이전트는 사용자의 질의를 분석하고 분산 DB를 위한 부 질의로 변환한다. 변환된 부 질의는 컨트롤 에이전트로 보내진다. 컨트롤 에이전트는 에이전트간 중재자 역할을 수행하며 자원 에이전트로부터 정보를 받아 메타데이터를 구축한다. 메타데이터를 이용하여 어떤 자원 에이전트가 질의와 관련된 정보를 갖고 있는지 판단하여 부 질의를 해당 자원 에이전트로 보낸다. 에이전트간 통신은 KQML을 이용하였다[8].



(그림 4) 정보수집을 위한 멀티에이전트 구조

2.3 협동적인 메타검색 엔진 구조

분산검색을 위한 또 다른 방법으로 메타검색엔진을 이용할 수 있다. (그림 5)는 메타검색엔진의 구조를 보여주며, LMSE(Local Meta Search Engine)와 LSE(Local Search Engine)로 구성되어 있다[9]. LMSE는 협동적인 검색엔진으로 LSE의 프로토콜을 이해하고 사용자 IF와 엔진 IF로 구성되어 있다. LSE는 HTML문서를 수집하는 수집기, 색인기, 검색엔진으로 구성되어 있다. 위치서버는 배타적으로 지식을 관리하고 캐시서버는 최근 사용질의 저장 및 연속적 병렬 검색을 수행한다.



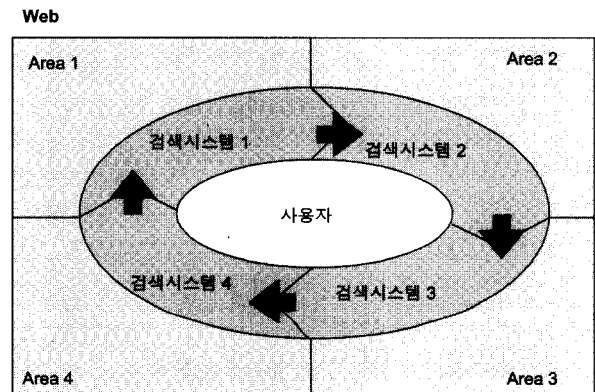
(그림 5) 협동적인 메타검색엔진의 구조

3. STM 분산정보검색시스템의 Framework 설계와 질의 처리방식

트래픽의 집중과 중앙 시스템의 추가 설치 등 구조적 제한사항을 갖고 있는 중계기 기반의 분산정보검색시스템의 문제점을 해결하기 위하여 순차적 전달방식(STM)을 이용한 분산검색시스템의 Framework를 설계하고 이와 관련된 질의 처리방식을 제안한다.

3.1 Framework 설계

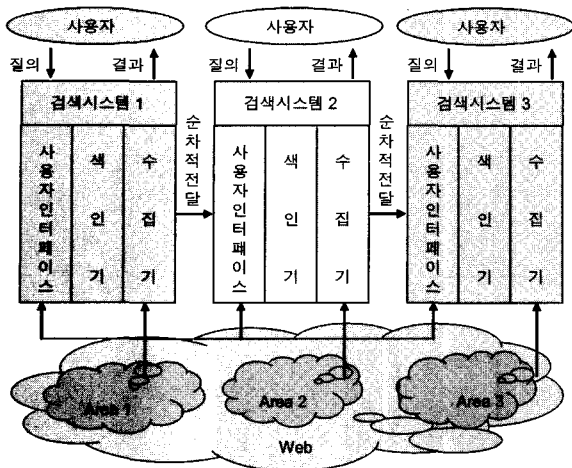
순차적 전달방식(STM)이란 다중의 분산검색시스템이 일정한 순서를 가지고 질의나 메시지를 전달하며 이에 따른 결과는, 최초 질의를 제기한 시스템으로 직접 전달하는 방식이다. 이러한 순차적 전달방식의 개념적인 Framework이 (그림 6)에 나타나 있다.



(그림 6) 순차적 전달방식의 개념적 Framework

검색시스템 1부터 4까지 순차적으로 질의를 전달하며 모든 검색시스템의 검색결과는 바로 검색시스템 1로 보내진다. 각 검색시스템은 웹 영역의 일부를 할당받아 관리한다. 만약 사용자가 검색시스템 1에 질의를 요청하면, 이 시스템은 사용자의 질의를 받아 처리하고 이 질의를 다시 다른 검색시스템으로 보내며 또한, 그 결과를 종합하여 사용자에게 제공하는 원천 검색시스템이 된다. 검색시스템 2는 검색시스템 1의 차순 검색시스템이 되며 검색시스템 1은 검색시스템 2의 전순 검색시스템이 된다. 즉, 순차적 전달방식은 원천 검색시스템에서 출발하며, 전순 검색시스템에서 차순 검색시스템의 순서대로 질의가 처리되는 협력적 전달방식이다.

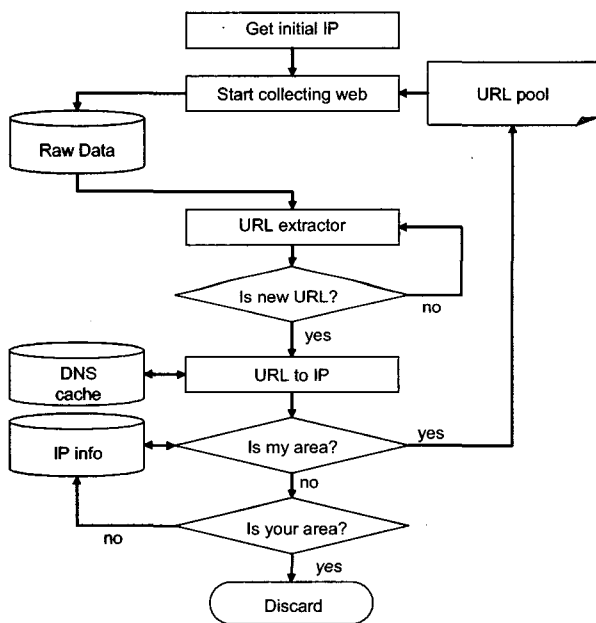
(그림 6)에 대한 구체적인 구조가 (그림 7)에 나타나 있다. 각 검색시스템은 분할된 웹 영역에 있는 웹 문서들을 수집하고 색인한 뒤 사용자의 질의에 순차적 전달방식을 이용하여 질의를 처리한다. 처리된 결과는 원천 검색시스템으로 바로 전달되며, 원천 검색시스템은 각 검색시스템으로부터 전달된 결과를 종합하여 사용자에게 제공한다.



(그림 7) 순차적 전달방식의 Framework

3.2 웹 분할(web partition) 정책

분산정보검색시스템이 동일 자료를 중복하여 관리하는 것을 방지하기 위해 웹을 분할하는 알고리즘이 필요하다. 이를 위해 우리는 웹을 논리적으로 분할한다. 분할 단위는 IP를 기준으로 하며 같은 IP내의 모든 페이지는 1개의 검색시스템이 담당한다. 각각의 분산검색시스템이 각각 다른 초기 IP를 임의적으로 할당받아 문서 수집을 시작한다. 새로운 URL이 발견되면 DNS로부터 IP변환 후 자신의 담당 IP인지 확인한다. 자신의 담당 IP가 아니라면 차순 검색시스템에게 문의를 하고 그 결과 다른 검색시스템이 담당하고 있지 않다면 자신의 담당 IP정보에 추가하고 작업대기열에 URL을 추가한다. 만약 다른 검색시스템의 담당이면 이를 무시한다. 웹 분할 정책알고리즘은 (그림 8)에 나타나 있다.



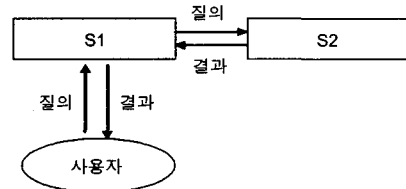
(그림 8) 웹 분할 정책 알고리즘

3.3 순차적인 질의 처리방식

순차적인 질의 처리란 사용자가 입력한 질의를 원천검색시스템이 분석하고 재구성하여 차순 검색시스템에 전달하는 것을 말한다. 차순 검색시스템은 전달받은 질의를 확인하고 선전달 후조치하는 순서로 작업을 수행한다. 원천 검색시스템은 전달 후 자신의 데이터를 검색하고 다른 검색시스템의 보고를 기다린다. 다른 검색시스템의 결과가 모두 모이면 이들을 종합하여 사용자에게 제공한다.

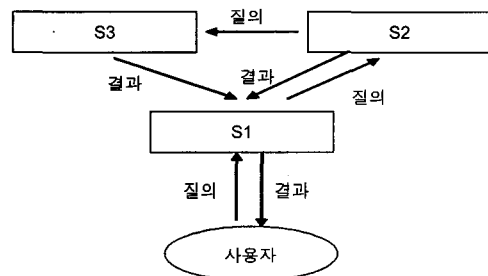
3.3.1 질의 처리의 예

예 1) 시스템이 2개인 경우(그림 9) : 사용자가 질의를 하면 검색시스템 S1이 질의를 분석하고 S2에게 질의를 전달한다. S1은 자신이 갖고 있는 정보를 검색하고 S2가 처리한 검색결과를 수신한다. 그런 다음, S1은 두 개의 검색결과를 종합하고 이를 사용자에게 제공한다.



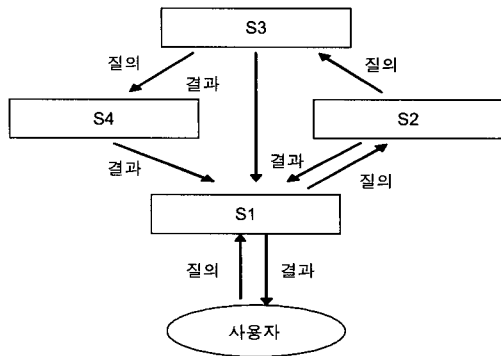
(그림 9) 시스템이 2개인 경우

예 2) 시스템이 3개인 경우(그림 10) : 사용자가 질의를 하면 S1은 질의를 분석하여 S2에게 질의를 보낸 후 자신의 정보를 검색한다. S2는 질의를 받아 확인한 후 S3에게 먼저 전달하고 자신의 정보를 검색한다. 여기서 S2와 S3는 각각 자신이 검색한 결과를 바로 S1으로 보내며 S1은 종합된 결과를 사용자에게 제공한다.



(그림 10) 시스템이 3개인 경우

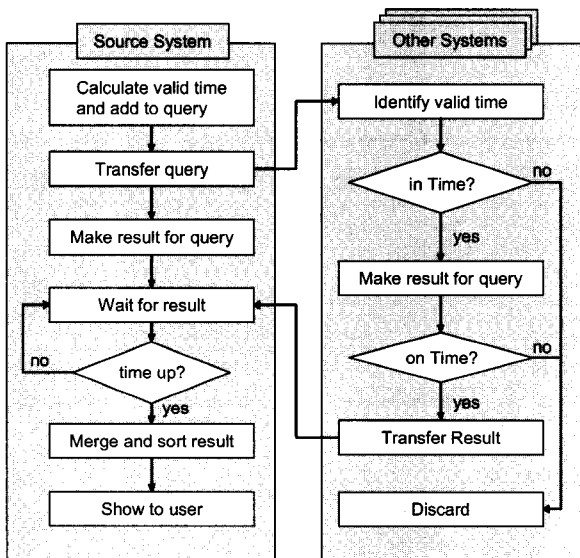
예 3) 시스템이 4개 이상인 경우(그림 11) : 사용자가 질의를 하면 S1은 질의를 분석하여 S2에 보낸 후 자신의 정보를 검색한다. 이후 질의를 수신 받은 시스템 Si는 먼저 질의를 Si+1로 보내고 검색결과를 만들어 S1으로 보낸다. S1은 자신의 검색결과와 다른 시스템의 검색결과를 종합한 후 사용자에게 제공한다.



(그림 11) 시스템이 4개 이상인 경우

3.4 유효시간(valid time) 설정

순차적인 검색방식에서 특정 검색시스템의 작동중지나 트래픽 집중으로 인한 혼잡지연, 데이터의 분실 등으로 인하여 전체 검색시스템의 무한대기, 전달지연 등의 상황이 발생할 수 있다. 이러한 상태는 전달되는 질의에 유효시간을 첨부함으로써 해결이 가능하다. 원천 검색시스템은 질의 결과 수집을 완료하는 유효시간을 설정하고 질의에 첨부하여 전송한다. 유효시간까지 원천 검색시스템은 다른 시스템으로부터의 결과를 대기하고 유효시간까지 수신된 결과를 분석하여 재작업이나 작업종료를 결정한다. 차순 검색시스템은 전달받은 질의의 유효시간을 확인하여 이미 지나가거나 결과를 전송할 만큼의 시간이 남지 않았다면 질의를 무시한다. 이러한 유효시간 판단 알고리즘이 (그림 12)에 나타나 있다.



(그림 12) 유효시간 판단 알고리즘

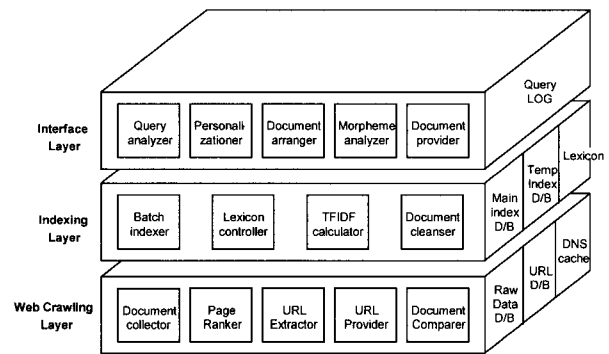
4. STM 분산정보검색시스템의 구조

본 장에서는 순차적 전달방식에 적용가능하고 독립적인

기능을 수행할 수 있는 분산정보검색시스템의 구조로 3 Layer 구조를 제안한다. 각 layer가 가지고 있는 모듈을 식별하고 모듈별 기능을 정의한다.

4.1 분산정보검색시스템의 Layer 구조

분산정보검색시스템은 Interface Layer, Indexing Layer, Web crawling Layer로 구성한다. (그림 13)는 제안한 분산 검색시스템의 3 Layer 구조를 나타낸다. 각 Layer는 고유의 기능을 수행하는 3~5개의 모듈로 구성되어 있고 자신의 DB 정보를 가지고 있다. 타 검색시스템과의 협동은 5장의 알고리즘을 이용한다.



(그림 13) 검색시스템의 3 Layer

4.2 Interface Layer

Interface Layer는 사용자와 밀접한 관계가 있는 계층으로 사용자의 질의를 분석하고 결과를 보여주기 위한 Layer이다[10]. 5개 모듈과 1개 DB로 구성되어 있다. Query analyzer는 사용자의 질의를 받고 분석한다. 사용자의 접속에 따라 멀티쓰레드의 형태로 구현된다. Personalization은 사용자 질의에 대한 로그를 유지하며 개인화와 선호도 분석, 추천 기능, 유사 키워드 제공 기능을 수행한다[11]. Document arranger는 결과에 대해 가중치를 이용하여 문서의 우선순위를 찾아낸다. Morpheme analyzer는 사용자의 질의어를 분석하여 여러 문서에서 동일하게 반복적으로 사용되는 색인어로서 의미가 없는 단어(불용어)와 매우 적게 사용되는 단어(오타 등)를 제거한다. Document Provider는 최종적으로 정렬되고 종합된 문서를 사용자에게 보여준다. 페이지당 20개의 문서를 관련된 내용의 일부와 URL, 유사도 등과 함께 제공한다.

4.3 Indexing Layer

Indexing Layer는 Web으로부터 가져온 Raw Data를 사용자의 질의에 신속히 응답할 수 있도록 색인하는 Layer이다. 4개의 모듈과 3개의 DB로 구성되어 있다. Batch Indexer는 고속으로 동작하는 Temp Index의 내용을 일정한 주기나 용량에 도달했을 때 Main Index에 반영시키는 역할을 수

행한다. Lexicon controller는 키워드를 관리한다. 새로운 신조어와 합성어 등을 식별하고 Lexicon에 추가하여 색인화에 사용한다. TFIDF calculator는 문서 내에서 용어의 반복되는 빈도를 이용하여 용어 가중치를 계산한다[12]. Document cleanser는 Raw Data를 색인하고 용어 가중치 계산에 불필요한 정보(HTML Tag 등)를 제거한다.

4.4 Web Crawling Layer

Web Crawling Layer는 새로운 사이트를 찾아 문서를 수집하고 저장하는 Layer로 Web에 직접 연결되어 작동한다. 자신에게 최적화된 DNS캐시를 가지고 있어 URL을 IP 변환할 때 발생하는 병목현상을 제거한다. 5개의 모듈과 3개의 DB로 구성되어 있다. Document collector는 URL 주소를 받아서 IP로 변환한 뒤 멀티쓰레드 형태로 문서를 가져온다. DoS 보호를 위해 한 사이트에 집중하여 문서를 가져오지 않고 여러 사이트에 분산하여 가져오게 된다. PageRanker는 하이퍼링크의 구조를 분석하여 그 페이지가 얼마나 많이 인용되었는지 파악하여 그 페이지의 중요도를 계산한다[13].

URL Extractor는 HTML문서에서 Tag정보중 <a href>의 정보를 이용하여 URL을 추출한다. 추출된 URL은 새로운 것인지 확인되고 IP 변환을 통해 자신의 담당인지 확인된다. URL Provider는 Document collector에게 문서를 수집해야 할 대상 URL을 중복되지 않도록 적절히 분산시켜 제공한다. Document comparer는 새로 가져온 문서와 보유하고 있는 문서와의 비교를 수행한다. 수정되지 않았다면 Raw Data에 저장할 필요가 없으며 인덱스의 갱신도 불필요하다.

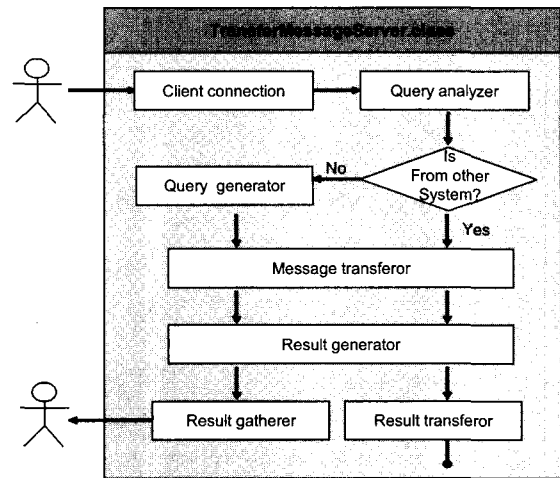
5. 순차적 전달방식(STM) 알고리즘 구현

본 장에서는 3장과 4장에서 설명한 순차적 전달방식을 이용한 검색시스템간의 협력적 Framework와 3 Layer 정보검색시스템의 구조에 적용할 수 있는 순차적 전달방식(STM)의 모듈 구조를 분석하고 알고리즘을 제시한다. 2대의 PC를 서로 LAN으로 연결한 로컬 네트워크에서 Java를 이용하여 구현한다.

5.1 STM의 모듈 구조

순차적 전달방식(STM)의 모듈구조는 (그림 14)와 같다. Client connection 모듈은 클라이언트가 연결할 때까지 대기한다. Client로부터 질의가 입력되면 Query analyzer가 질의의 종류와 적합한 질의인지 확인한다. 질의가 다른 system으로부터 입력된 것인지 확인하고 다른 system으로부터 입력되었다면 차순 검색시스템으로 전달하고 결과를 생성시켜 전송한다. 만약 다른 system으로부터 입력되지 않았다면 전달을 위한 질의 포맷을 생성한다. 여기에는 유효시간과 자신의 IP정보가 추가된다. 질의를 전달하고 자신의 결과를 처리하

고 다른 system의 결과를 종합해서 사용자에게 보여준다.



(그림 14) STM 모듈의 구조

5.2 STM의 pseudocode 표현

pseudocode로 표현한 순차적 전달방식(STM)의 모듈구조가 (그림 15)에 나타나 있다. 이 그림에서는 통신포트 개방, 클라이언트 연결 대기, 질의 수신 등의 공통 부분과 다른 시스템의 질의 처리부분 그리고 사용자의 질의 처리 부분 등 세 개의 큰 부분으로 구성된다. 다른 시스템의 질의 처리부분은 질의의 차순 검색시스템 전송, 질의에 대한 결과 검색, 그리고 원천 검색시스템으로의 결과 전달 등으로 구성된다. 사용자 질의 처리부분은 질의에 대한 IP 정보 추가, 차순 검색시스템으로 질의 전달, 질의 검색, 타 시스템으로 결과 수신, 결과 종합 및 사용자 제시 등으로 구성된다.

5.3 질의 전달과정 구현 결과

순차적 전달방식에 따른 서버간의 질의 전달 및 검색결과 처리과정에 대한 시연결과가 (그림 16)에 나타나 있다. (그림 16)의 좌측부분은 사용자의 질의를 받은 System 1이며 우측부분은 System 1의 요청을 받아 자신의 정보를 전송해주는 System 2이다. (그림 16)의 두 부분에 나와 있는 내용을 비교해 볼 때 System간의 질의 전달이 우리가 제안했던 순차적 전달방식에 따라 원활하고 효과적으로 이루어지고 있음을 알 수 있다.

```

    Algorithm name : Sequential Transferring Method (STM)

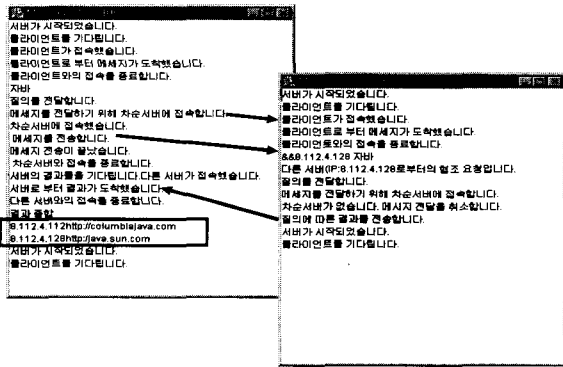
    Open Communication port
    do while(user want to quit) {
        wait for Client's connection
        Receive Query
        if Query is from other systems)
        {
            /* 다른 시스템의 질의 처리 */
            Transfer Query to next system
            Make result for Query
            Send back to the source system
        }
    }
    
```

```

} else { /* 사용자의 질의 처리 */
  add my IP information to Query
  Transfer Query to next system
  Make result for Query
  Wait for results from other systems
  Sum results
  Show result to user }
}
    
```

(그림 15) Pseudocode로 표현한 순차적 전달방식

(그림 16) 좌측하단의 사각형 박스는 양쪽 서버의 자료가 종합된 결과를 나타낸다. 이것은 검색어 '자바'가 system 1에서 system 2로 보내진 다음 그 결과가 다시 system 1에서 종합되었다는 것을 알 수 있다.



(그림 16) STM의 질의 전달과정

6. 결 론

중계기 기반의 분산정보검색시스템의 트래픽 집중과 여러 문제점을 해결하기 위해 순차적 전달방식(STM)을 이용한 효율적인 분산정보검색시스템을 제안하였다. 이를 위해 웹, 검색시스템, 사용자와의 상호관계를 이용한 순차적 전달방식(STM)의 협력적 Framework를 제시하였다. 그리고 원천 검색시스템과 차순 검색시스템과의 질의 처리과정을 도식적으로 예시하였다. 제안한 Framework 설계와 질의 처리과정을 이용하여 STM 분산정보검색시스템의 3 Layer 구조를 제시하였으며, 총 14개의 기능별 모듈과 7개의 관련 DB로 구성하였다. 제안한 STM 알고리즘 구현을 위해 Transfer-MessageServer.class의 모듈구조를 분석하였으며 이를 구현하기 위한 프로그램을 작성하였다. STM 구현을 위해 PC-LAN 기반에서 JAVA를 이용하여 서버간 질의 전달과정을 시연하였으며 시연결과 시스템간 질의 전달이 원활하게 이루어짐을 증명하였다.

참 고 문 헌

[1] ALTAVISTA : Main Page. <http://www.altavista.com>, 1996.
 [2] 김영철의, 「최신정보검색론」, 홍릉과학출판사, 2001.

[3] 전우직, "디지털도서관 표준 프로토콜 Z39.50의 개요 및 구현 방안," KOSTI96, 1996.
 [4] 송광택, 이희주, 안성현, "분산통합검색을 위한 분산통합기 관리 도구," 제5회 한국과학기술정보 인프라워크샵, Dec., 2000.
 [5] C. Mic Bowman, Peter B. Danzig, Darren R. Hardy, Udi Manber and Michael F. Schwartz, "The Harvest Information Discovery and Access System," In Proc, 2nd Int. www conf., pp.763-771, Oct., 1994.
 [6] Keith L. Clark and Vasilios S. Lazarou, "A Multi-Agent System for Distributed Information Retrieval on the World Wide Web," In WETICE '97, Collaborative Agent in Distributed web Applications, IEEE Computer Society Press, 1997.
 [7] T. Finin, Y. Labrou and J. Mayfield, "KQML as an Agent Communication Language," Software Agent, pp.291-316, 1997.
 [8] Soo-Won Lee, "Information Gathering Agents for Distributed Systems," Journal of the Industrial Technology Research Institute, Soong Sil Univ, Vol.31, pp.165-172, 2001.
 [9] Nobuyoshi Sato, Minoru Uehara, Yoshifumi Sakai and Hideki Mori, "Distributed Information Retrieval by using Cooperative Meta Search Engines," 21st International Conference on Distributed Computing System Workshops, 2001.
 [10] Heebyung Yoon, Kilsup Lee and Hwa-Soo Kim, "Analysis and improvement Ranking Algorithm from Web Mining System on the Hierarchical Web Environment," International Symposium on Advanced Intelligent System, pp. 455-458, 2003.
 [11] George T. Wang, F. Xie, F. Tsunoda, H. Maezawa and Akira K., "Web search with personalization and knowledge," 4th international symposium on Multimedia Software Engineering, 2002.
 [12] V. V. Raghavan, "S.K.M.A critical analysis of vector space model for information retrieval," Journal of the American Society for Information Science, 1996.
 [13] L. Page, S. Brin, R. Motwani and T. Winograd, "The PageRank Citation Ranking : Bringing Order to the Web," Stanford University, 1998.



윤 희 병

e-mail : hbyoon@kndu.ac.kr

1983년 해군사관학교(이학사)

1986년 연세대학교(공학사)

1991년 미국 해군대학원 전산공학(석사)

1998년 미국 Georgia Institute of

Technology 전산공학(박사)

2002년~현재 국방대학교 전산정보학과 조교수

관심분야 : 웹 마이닝, 지능정보시스템, 에이전트시스템, 모바일 웹 검색



김 용 한

e-mail : airkyh@yahoo.co.kr

1992년 공군사관학교 기계공학과(학사)

2003년~현재 국방대학교 전산정보학과
석사과정

관심분야 : 정보검색, 분산시스템, 웹마이닝



김 화 수

e-mail : ajhskim@ajou.ac.kr

1976년 해군사관학교(이학사)

1984년 미국 해군대학원 전산학(석사)

1990년 미국 Case Western Reserve
University 전산학(박사)

1991년 국방대학교 전산정보학과 교수

2003년~현재 아주대학교 정보통신대학원 교수

2003년~현재 국방부장관 정책자문위원(정보화분야)

2004년~현재 감사원자문위원(행정 및 국가안보분야)

관심분야 : S/W 개발비용 산정/분석, 국방가상현실/시뮬레이션
및 모델링, 최신정보기술 관리, 인공지능, 전문가시스
템 개발