

단순 조회 기능 서버 기반 P2P 시스템의 검색 성능 향상을 위한 Ptops 인덱스 서버 설계

김분희⁰ 김영찬
중앙대학교 컴퓨터공학과
(bhkim⁰, yckim)⁰@sslslab.cse.cau.ac.kr

Design of Ptops Index Server to Advance Search-Performance of P2P with a simple discovery server System

Boon-Hee Kim⁰ Young-Chan Kim
Dept. of Computer Science and Engineering, Chung-Ang University

요 약

기존 P2P 시스템의 성능 향상에 대한 연구가 다양한 주제로 그 해결책을 모색하고 있는 가운데, 본 논문에서는 정보에 의미성을 부여하여 검색 성능 향상을 이루는 P2P 인덱스 서버 프레임워크를 제안한다. 본 논문의 제안 기법은 검색 성능의 속도 면에서 시스템 초기화 과정의 소요 시간을 감안한 비교 결과 기존 시스템에 준하는 성능을 보였으며, 해당 피어의 요구사항에 대해 신뢰성 있는 결과 데이터 리스팅의 관점에서 타 시스템과의 의미 있는 평가치를 나타내고 있다.

1. 서 론

최근 웹 서비스에 대한 사용자 요구가 늘어나고 있는 추세이며, 개인 컴퓨터의 속도 및 처리능력이 향상되면서 사용자의 컴퓨터에서 직접적인 자원 제어, 교환, 공유의 필요성이 대두되었다. 이에 피어들 사이에서 정보의 양방향 흐름을 직접적으로 연결해 줄 수 있는 P2P (Peer-to-Peer) 모델이 제안되었다. 클라이언트/서버 모델과는 달리 P2P는 서버에 대한 의존도가 낮아 서버로부터의 중앙 집중 제어를 받지 않는 강점을 지닌다.

P2P 모델은 일반적으로 순수한 P2P(Pure Peer-to-Peer), 간단한 조회 기능 서버를 가진 P2P(P2P with a simple discovery server), 조회 서버와 룩업 서버를 가진 P2P(P2P with discovery and lookup servers), 조회/룩업/컨텐츠 제공 기능의 서버를 가진 P2P(P2P with discovery, lookup and content servers) 모델로 분류되는데, 찾고자 하는 자료가 존재하는 피어를 검색하는 시간과 서버의 과부하 관계에서 성능상의 효율성을 지닌 "조회 서버와 룩업 서버를 가진 P2P"를 본 논문의 기반 모델로 하고자 한다.

P2P 어플리케이션에서 서버 설계 시 부하 감소형 서버 이면서, 각 피어의 최소한의 검색 시간을 보장하기 위한 인덱스 리스팅 기능은 보장되어야 한다. 일반적으로 서버는 피어들의 목록과 해당 자원의 리스트를 가지며, 이에 대한 요구가 들어올 때 자원의 이름과 부합하는 전체 목록을 전달하는 역할을 함으로 이를 인덱스 서버라 한다. 이렇듯 P2P 어플리케이션들은 단순히 찾고자하는 파일의 이름과 그에 대응되는 다른 피어로 연결된 정보로써 검색의 과정을 거친다. 그러나 제공되는 방대한 양의 정보의 단순한 연결만으로 사용자가 원하는 그 자료를 찾기에는 역부족이다. 이에 정보의 디스플레이에 그쳤던 기존 P2P 어플리케이션의 효율적인 정보 검색 방법의 대안이 제시되어야 한다.

분산 어플리케이션의 일례로써 1989년 이래 가장 널리 알려진 개념은 팀버너스리에 의해 제안된 웹일 것이다.

기존의 웹은 구조의 단순성으로 성장 가도를 달리고 있지만 방대한 정보 가운데 원하는 정보를 찾기 위해 고군분투하는 검색기법들은 그 한계를 드러내고 있다. 이러한 문제점의 발생 원인으로 HTML 언어의 용도를 들 수 있다. 문서의 내용과 의미를 나타내기보다 디스플레이 하는 것이 목적인 HTML 언어로 구성된 웹은 다양한 검색 기법의 성능 제한성의 원인이 된다. 이에 시맨틱 웹은 웹 문서에 의미성을 부여하여 정보의 자동추출, 확장, 공유 등의 분야의 연구에 기여하고 있다. 이러한 시맨틱 웹 개념을 P2P 어플리케이션에 도입하여 정보 검색의 성능 향상의 기여 여부를 판단해 볼 수 있겠다.

P2P 관련 성능 향상에 대한 연구가 많은 주제로 그 해결책을 모색하고 있는데, 본 논문에서는 이러한 주제들 가운데 정보에 의미성을 부여하여 검색 성능 향상을 이루는 P2P 인덱스 서버 프레임워크인 Ptops(Peer-to-Peer for the Semantic Web)를 제안한다. 본 논문의 구성은 다음과 같다. 2장에서 관련연구, 3장에서는 Ptops 인덱스 서버, 마지막으로 4장에서 결론을 맺는다.

2. 관련 연구

2.1 P2P

현존하는 P2P 모델은 피어간의 네트워킹 방법과 역할에 따라 다음과 같이 분류할 수 있다.

- 순수한 P2P(Pure Peer-to-Peer),
- 간단한 조회 기능 서버를 가진 P2P(P2P with a simple discovery server)
- 조회 서버와 룩업 서버를 가진 P2P(P2P with discovery and lookup servers)
- 조회/룩업/컨텐츠 제공 기능의 서버를 가진 P2P (P2P with discovery, lookup and content servers)

순수한 P2P 모델은 중앙 서버 개념이 배제된 순수 클라이언트 대 클라이언트 간 네트워킹으로써 이루어진다. 진정한 의미에서의 P2P 모델이라 볼 수 있다. 이 모델은 P2P 어플리케이션의 동작과 함께 각 피어들은 네트워킹

에 접속된 다른 피어들을 동적으로 찾고, 연결된 피어들 사이의 통신에 있어서 어떠한 서버도 개입되지 않는다. 통신방법에 있어서 기존의 클라이언트/서버 모델의 관습에서 벗어난 대등한 구조이다. 그러나 접속하고자 하는 다른 피어들에 대한 검색이 네트워크 상에서 이뤄지므로 로그인한 피어들의 목록을 등록해주는 중앙 관리자의 역할이 없어 피어 스스로가 다른 피어의 위치를 파악해야 함으로 검색 시간 및 피어 자체의 성능상의 제약이 야기될 수 있다.

간단한 조회 기능 서버를 가진 P2P 모델은 실제 클라이언트/서버 모델에서처럼 서버를 포함하고 있는 것은 아니지만, 이미 접속된 피어의 정보를 제공하는 최소한의 서버의 역할이 포함되어 있다. 그러므로 실제 피어들 간의 접속을 형성하고 통신을 수행하는 것은 여전히 피어 자신의 몫이다. 이는 순수 P2P 모델에 비해 다른 피어를 조회할 가능성이 높은 장점이 있다. 그러나 자원을 다운로드 받기 위해 각 피어에 개별적인 접근이 이뤄져야 하므로 요청을 보내는데 걸리는 처리시간이 길어질 수 있다. 이러한 면에서 클라이언트/서버 기반 모델은 서버 내에 모든 자원을 보유함으로써 다른 피어를 찾는 데 소모되는 시간을 줄여주는 상반된 역할이 있다.

조회 서버와 룩업 서버를 가진 P2P 모델은 앞선 모델에서 나타난 서버의 역할을 보완한 모델로써 접속된 피어들의 목록을 제공함과 동시에 각 피어들의 이용 가능한 자원을 함께 함께 제공한다. 이 모델에서 피어들은 정보를 얻기 위해 피어들을 각각 방문해야 하는 어려움을 덜 수 있다. 물론 서버의 역할 면에서는 위의 두 모델에 비해 두 피어들 사이의 통신 개시의 역할이 더해졌다.

조회/룩업/컨텐츠 제공 기능의 서버를 가진 P2P 모델은 전통적인 클라이언트/서버 모델에서와 같은 역할 배분이 이뤄진다. 즉 모든 자원이 중앙의 서버에 위치하게 되며 클라이언트/서버 모델에서 나타나는 단점들을 그대로 안고, 이로써 직접적인 피어간의 통신이 단절을 초래하며 위의 모델과는 달리 중앙 서버의 고장으로 인한 전체 시스템의 영향력이 매우 크다.

현존하는 P2P 어플리케이션으로 특정 확장자를 가진 파일들을 공유하는 냅스터(Napster), 순수 P2P 모델과 작동 방식이 가장 유사한 그누텔라(Gnutella) 등이 있으며, 이밖에도 파일 공유 서비스와 소형 검색 엔진을 갖춘 클라이언트 기반 프로그램들이 다수 존재한다. 본 논문에서는 정보 검색 성능의 효율성을 부가하기 위해 위에서 분류한 P2P 시스템 모델 가운데 세 번째 모델인 “조회 서버와 룩업 서버를 가진 P2P” 구조를 채택했다.

3. Ptops 인덱스 서버

3.1 Ptops 구성 요소

정보에 의미성을 부여하여 검색 성능 향상을 이루는 P2P 인덱스 서버 프레임워크인 Ptops는 다음과 같은 모듈로써 구성되어 있다.

- 스텝(User Stub)
- 스켈레톤(User Skeleton)
- SD(Semantic Document Producer)

- SDP(Semantic Document Parser)
- 바인더(Binder Modules)
- 요청자(Requester Modules)
- SC(Server Contact)
- PC(Peer Contact)

Ptops 시스템의 피어는 그 역할에 따라 스텝과 스켈레톤으로 이뤄진다. 스텝(User Stub)은 자원을 요청하는 피어이고, 스켈레톤(User Skeleton)은 요청에 대한 자원을 제공하는 응답자 피어이다. 스켈레톤의 바인더(Binder Modules)는 인덱스 서버에 IP 주소, 로그인 이름, 공유 자원 정보로써 로그인한 후 실제 스텝의 요청을 기다리는 역할을 한다. 스텝의 요청자(Requester Modules)는 인덱스 서버에 원하는 정보를 보유한 스켈레톤의 리스트를 요청하고, 서버로부터 바인더 목록을 받게 된다. 요청자는 통신 수립의 두 단계를 SC(Server Contact)와 PC(Peer Contact) 과정을 거친다. SD(Semantic Document Producer)는 사용자의 요구에 따라 요청과 응답을 생성하는 역할을 수행하고, SDP(Semantic Document Parser)는 생성된 요청과 응답을 파싱하는 역할을 담당한다.

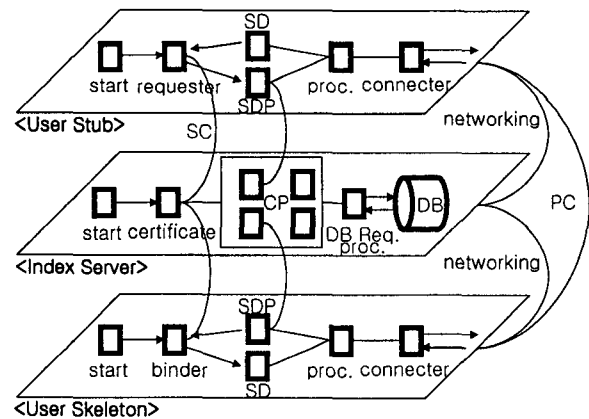


그림. Ptops 스텝의 요청 처리

Ptops의 핵심 구성요소인 인덱스 서버(Index Server)는 수행됨과 동시에 검증 모듈(Certificate Module)과 연계되고, 이는 바인더의 로그인 요청을 검증, 허가, 취소의 기능을 지원한다. 또한 접속한 바인더를 바인더 리스트 항목에 추가하여 관리하게 되며, 요청자의 경우는 따로 목록화 하지 않는다. 이는 Ptops 인덱스 서버는 요청자의 요구에 응답할 수 있는 스켈레톤과 보유 정보만 유지하면 되기 때문이다.

인덱스 서버의 핵심 모듈인 CP(Central Process)는 인덱스 서버에서 사용자의 요구자 모듈(requester)에 의해 전달된 요구사항을 처리하는 중심 모듈로써 스켈레톤 피

어와 해당 피어의 보유 자원을 인덱싱하고, 스텝 피어의 요구에 따라 의미적 조화를 검토하여, 스텝 피어의 요구에 부합하는 스켈레톤 피어와 자원의 조합 정보를 해당 SDP에 전달하는 역할을 담당한다. CP 모듈에서 시멘틱 데이터의 분석 및 적용 관계는 다음 절 Ptops CP 모듈 부분에서 다루겠다.

3.2 Ptops CP 모듈

Ptops의 CP는 그림. 과 같이 구성된 모듈간의 상호작용으로 동작된다.

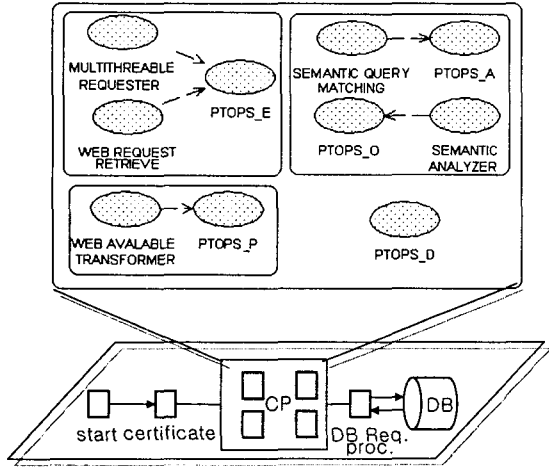


그림. Ptops CP모듈

PTOPS_E는 많은 수의 검증된 피어들의 요청을 받아들이기 위해 멀티쓰레딩 기능의 지원을 받으며, 입력된 시멘틱 정보는 PTOPS_O 모듈의 분석 결과를 토대로 이해되고 처리한다. 이와 더불어 PTOPS_A는 인덱스 서버의 데이터베이스에 저장되어 있는 피어 리스트 가운데, 요청 정보에 의미적으로 부합하는 피어를 파싱하여 PTOPS_P 모듈에 전달한다. PTOPS_D는 데이터베이스 요청 중재의 역할을 전달한다. 스텝의 요청에 부합하는 정보를 전달받은 PTOPS_P는 웹 환경에 적합한 형태의 리스트 정보를 구성하고, 해당 스텝에게 인코딩된 정보를 보내게 된다.

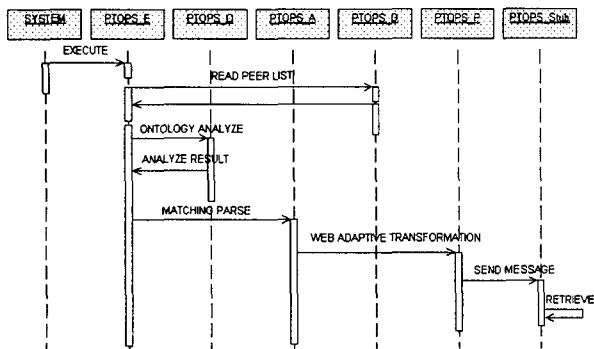


그림. Ptops CP 모듈 수행 절차

Ptops 시스템의 CP 모듈의 수행 절차는 그림. 과 같다. 우선 CP 모듈의 수행 시점에서 시작된 PTOPS_E는 PTOPS_D에 일차적으로 피어 리스트를 읽어온다. 다음 스텝의 의미적 요청을 PTOPS_O 모듈에서 분석하고 PTOPS_A 모듈에서 매칭되는 결과를 파싱한다. 이렇게 생성된 정보를 웹 환경에 적합한 형태로 변환하는 과정이 PTOPS_P 모듈을 통해 이뤄지고, 해당 스텝에 구성된 리스트 의미 정보를 보냄으로써 CP의 역할을 마무리한다.

3.3 구현 및 비교 평가

Requirements	JXTA	OceanStore	Pastry	Ptops
File Sharing	Only binary message		yes	
Automatic Knowledge base creation	Not included			Semi-automatic
Semantic Information	Not included			yes
Extensibility	Java allows own extensions			
Interoperability	Not integrated, small version thinkable	Needs JVM	Difficult, requires Java and RMI	Good
Scalability	yes	yes	Routing scales very good	yes

그림. JXTA, OceanStore, PASTRY와 Ptops 비교 그림. 에서도 나타난 바와 같이 Ptops는 기존 P2P 시스템과 비교해 봤을 때 확장성, 상호운용성 등의 측면에서 유사한 가능성을 보여줬고, 특히 시멘틱 정보 처리 부분에서의 초보적인 시도 단계의 기존 시스템에 비해 관련 요구사항의 충분한 해결점을 제시하고 있다.

4. 결론 및 향후연구

기존 P2P 시스템의 성능 향상에 대한 연구가 많은 주제로 그 해결책을 모색하고 있는 가운데, 본 논문에서 제안한 Ptops 시스템은 관련 주제들 가운데 정보에 의미성을 부여하여 검색 성능 향상을 이루고자 하였다. Ptops 구현물의 실험결과 검색 성능의 속도면에서는 시스템 초기화 과정의 소요 시간을 감안했을 때 기존 시스템에 준하는 성능을 보였으며, 전문가 피드백에 의한 평가 결과 나타난 데이터의 신뢰성은 타 시스템의 성능에 비해 의미 있는 결과를 보여줬다. 이러한 결과의 당위성을 높이기 위해 다음 연구에서 덩컨 와츠(Duncan Watts)와 스티븐 스트로게츠(Steven Strogatz)의 실험된 방법론을 적용하고, 이러한 적용이 용이한 환경의 구성을 위해 실험값의 시작점인 평균치에 이를 수 있도록 평가 수준을 높여야겠다.

참고문헌

[1] Brian McBrids, "Jena: A Semantic Web Toolkit", IEEE Internet Computing, November, 2002, pp.55-59.
 [2] Brian McBrids, "Jena: Implementing the RDF Model and Syntax Specification", <http://www-uk.hpl.hp.com/people/bwm/papers/20001221-paper>.
 [3] Aditya Kalyanpur, and so on, "SMORE-Semantic Markup, Ontology, and RDF Editor", <http://www.mindswap.org/papers/SMORE.pdf>.