

LinkBoy : CORBA 기반 링크 정보 검색 엔진

김재훈 손주영

부산시 영도구 동삼동 1번지 한국해양대학교 컴퓨터공학과

LinkBoy : A CORBA-based Link Information Search Engine

Jaehoon Kim Jooyoung Son

Dept. of Computer Engineering, Korea Maritime University

1 Dongsam-dong, Youngdo-Gu, 606-791, Korea

요약

전세계를 연결하는 인터넷은 웹의 출현으로 급속히 퍼져나가 하나의 거대한 정보망으로 형성되었다. 이러한 정보망에서 필요한 정보를 찾기 위한 검색 엔진이 인터넷에서의 필수 요건이 되고 있다. 그러나 정보가 저장되어 있는 여러 서버들 간의 플랫폼의 이질성의 극복, 언어 독립적인 플랫폼의 필요성이 대두된다. 그 해결책으로 이질적 분산 환경의 통합 환경을 위한 CORBA가 제시되고 있다. 이에 본 논문에서는 CORBA 기술을 이용하여 인터넷에서의 정보 제공 서버간의 투명한 상호 작용과 정보 검색이 가능하게 하며, 특히 일반 검색 엔진과는 달리 웹의 링크를 통해 정보를 검색할 수 있게 하기 위한 서버 엔진과 클라이언트 사용자 인터페이스를 설계, 구현하였다.

1. 서론

흔히 인터넷 상에서 정보를 검색할 때는 altavista, yahoo, infoseek, 심마니 등과 같은 검색 엔진을 이용한다.

검색 엔진을 이용하거나 검색 엔진을 개발하는 측면에서 보면 계속 증가하는 정보의 양과 다양성, 그리고 분산성으로 인해 원하는 정보를 얻는 데 더욱 많은 시간을 들이게 되고, 급속히 증가하는 정보를 저장하는 서버 또한 로드 과부하 문제가 생기게 된다.

기존의 검색 엔진에서는 필요한 질의어를 주게 되면 그 질의어를 가지고 있는 홈페이지를 모두 표시하여 사용자 입장에서 검색된 모든 페이지들을 순차적으로 찾아 보아야 원하는 정보를 얻을 수가 있다. 또한 같은 URL을 여러 번 참조하게 되거나 웹 서핑에서 길을 잃게 되는 경우도 있다.

본 논문에서 제시하는 LinkBoy는 하나의 URL 정보나 질의어에 의한 검색을 한다. 검색의 결과는 질의어를 담고 있는 웹 페이지들뿐만 아니라 각 페이지에 링크된 페이지 정보도 함께 제공된다는 점이 큰 특징이다. 이로써 정보검색에 있어서의 비순차적 접근(random access)이 가능하게 된다. 또한 LinkBoy는 하나의 주제어에 해당하는 URL을 제외한 같은 이름의 URL을 제거했으며, 링크정보를 나타내는 사용자 인터페이스를 통해서 웹 서핑에서의 미아가 되는 현상을 방지하였다.

클라이언트와 서버 사이에 CORBA를 기반으로 통신이 이루어지므로 서버들간 로드 밸런싱, 위치투명성 등이 가능하다. 이는 일반적으로 분산되어 있는 여러 검색 엔진 서버에 의해 검색 서비스가 제공되는 점을 고

려하면 검색 성능을 크게 높이는 효과를 가져다 준다.

현재 홈페이지의 가치 측정의 기준으로 평균 접속률을 이용한다. LinkBoy가 실용화되면 더 나아가 한 웹 페이지가 다른 페이지들에 의해 얼마나 많이 링크되어 있는가 하는 점이 계산되기 때문에 이점도 웹 페이지의 중요한 가치 측정의 기준으로 이용할 수 있게 된다.

CORBA는 OMG에서 제안한 OMA(Object Management Architecture)표준의 일부분이다. OMA는 분산 환경 하에서 필요로 하는 전반적인 객체 솔루션을 제공한다[1].

본 논문의 구성은 다음과 같다. 클라이언트와 서버 사이의 통신은 CORBA 기반으로 이루어지므로 2장에서는 CORBA 기반의 LinkBoy에 대한 기능, 구조, 서비스의 유용성, 사용자 인터페이스에 대하여 설명하고, 3장에서는 구현 환경을 통하여 이를 검증한다. 마지막으로 4장에서는 결론을 맺고, 향후 연구 방향을 제시하였다.

2. CORBA 기반의 LinkBoy

2.1 LinkBoy의 기능과 구조

기존의 검색 엔진들은 웹 서버내의 홈페이지에 대하여 명사 위주의 주제어를 찾아서 데이터베이스에 저장하고, 질의어를 통해 저장된 주제어를 보여주는 구조로 되어있다. LinkBoy는 주제어와 URL 외에 사용자가 찾는 정보를 한 눈에 알 수 있게 하기 위한 링크 정보를 추가로 데이터베이스에 저장한다. 추가적으로 제공되는 링크 정보를 통해 검색된 웹 페이지가 자신이 찾고자 하는 정보를 얼마나 충실하게 가지고 있는지 그리고 그

페이지에 의해 주변 정보 또는 심화 정보를 찾아갈 수 있는지를 그 웹 페이지를 굳이 브라우징하지 않더라도 바로 알 수 있게 된다.

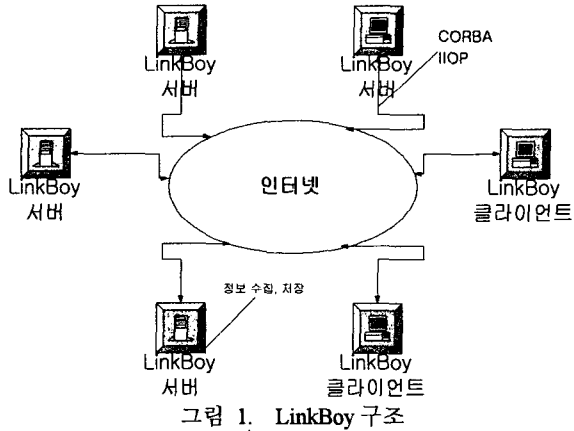


그림 1. LinkBoy 구조

그림 1은 LinkBoy의 구조를 나타낸 것이다. LinkBoy 서버와 LinkBoy 클라이언트로 이루어져 있고, 인터넷과의 통신은 CORBA IIOP를 사용하였다[2]. LinkBoy 서버들은 정보를 검색해서 분산 저장을 하고, Linkboy 클라이언트에서는 저장된 정보를 보여준다.

2.2 검색 엔진 서버 기능

일반적으로 검색 엔진은 Robot-driven search engines, Web-Directory-Service search engines, Meta-Search engines 등이 있다[3]. 링크 검색 엔진 서버는 일반적인 사례 중 공개된 가치네 소스를 수정하여 구현하였다[4]. 우리의 관심사가 검색 엔진 자체의 새로운 구현이 아니라 링크 정보 제공 방식, 그리고 그의 유용성을 연구, 검증하는 것이므로 실험적인 구현을 하였다.

가치네 엔진은 어떤 한 서버의 계층적인 디렉토리 구조를 참고하여 자료를 모으고 패턴 매치기법으로 구현한 Robot-driven 검색 엔진이다.

서버 프로그램은 html 문서를 인덱싱하여 데이터베이스에 URL 과 주제어 그리고 링크 정보를 저장하는 구조로 되어있다. 여기서 중복되는 URL 은 제거된다. 서버에 저장되어 있는 링크 정보들은 CORBA 에이전트가 서버 구현 객체 호출이 오면 클라이언트에게 전달해 준다. 또한 CORBA 를 기반으로 하여 정보를 분산해서 로드 밸런싱을 가능하게 하였다[5,6]. 그러므로 정보가 계속 증가하고, 서버 환경이 서로 달라도 투명한 상호 운용이 가능하게 된다.

2.3 제공 서비스의 유용성

기존의 검색 엔진은 하나의 질의어에 대한 모든 사이트들을 순차적으로 참조해야만 자신이 원하는 정보를 얻을 수가 있었다[7]. 또한 검색 후 서평을 계속하게 되면, 같은 URL 을 들어가게 되는 경우가 많고, 나중에는

자신이 어디에 있는지도 모르게 되는 경우가 많다. 정보 검색에서는 자신이 필요로 하는 정보만을 찾는 것이 가장 중요하다.

주제어를 통한 일반적 검색에서 추가하여 링크를 이용한 검색은 유사한 정보들이 서로 링크되어 있다는 생각에서 비순차적 접근을 통해 더욱 신속, 정확하면서도 더 깊이 있는 검색을 가능하게 해준다. 사용자는 이러한 링크 정보를 한 눈에 알 수 있어 출발지, 방문한 곳, 그리고 현재 위치에 대한 정보도 알 수가 있어서 중복 방문 경우, 위치 상실 경우 등이 없게 된다. 또한 사이트의 새로운 가치 척도로 일반적인 평균 접속률이 아닌 링크 접속률을 제공한다.

2.4 클라이언트 사용자 인터페이스

클라이언트 프로그램은 Java applet 으로 구현이 되어 있어 인터넷을 통한 ORB 끼리의 통신은 TCP/IP 상위의 IIOP 를 이용한다. ORB 는 서버측의 ORB 와 상호 작용을 통하여 객체의 생성, 소멸을 관리하게 된다.

클라이언트에서는 기존의 검색 엔진과 같이 질의어를 통해 검색을 하여 나타난 주제어를 가지고, 링크를 통한 검색을 할 수 있다.

본 논문에서는 LinkBoy 의 클라이언트 사용자 인터페이스의 두 가지 형태를 설명한다. 기본적으로 모두 DAG 구조로 되어 있다.

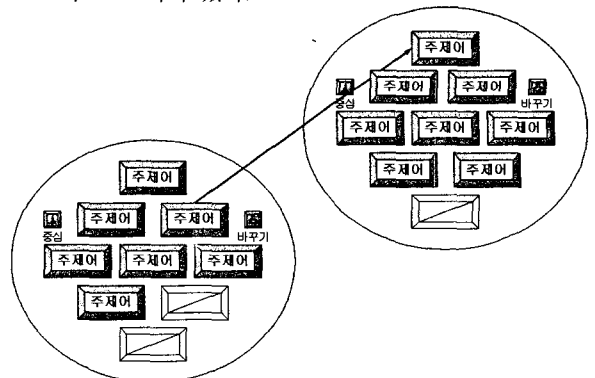


그림 2. 기본적 사용자 인터페이스

그림 2에서 하나의 타원은 하나의 사이트를 나타내고, 왼쪽 타원이 주제어에 의해 검색된 기본 사이트가 된다. 주제어 위에 마우스 포인터를 위치시키면 해당 URL 을 볼 수 있으며 클릭으로 해당 사이트의 링크 정보가 오른쪽 타원 내에 나타난다. 화살표는 2 단계 검색을 나타내며, 각 단계별로 다른 색깔의 화살표가 나타난다. 이는 연결된 사이트가 기본 사이트로부터 얼마나 깊이 들어간 것인지를 구별할 수 있게 한다.

하나의 타원 내에는 모두 9개의 링크를 표시할 수 있고, 9개가 안 되는 링크를 가지는 사이트는 빈 페이지로 표현하였고 “바꾸기” 버튼을 통해 그 외의 링크 정보를 볼 수 있다. 그리고, “중심” 버튼을 클릭함으로써 해당 사이트를 화면의 중앙에 위치시켜 기본 사이트가

되고 그 사이트를 중심으로 한 새로운 링크 검색을 할 수 있다. 활성화 되지 않은 사이트는 희미하게 나타나는 계층적 구조를 이룬다. 그림 2에서는 메뉴가 표현하지 않았지만 해당 URL의 홈페이지를 새로운 창을 통해 볼 수 있으며 지난 링크 정보도 볼 수 있는 기능을 가진다.

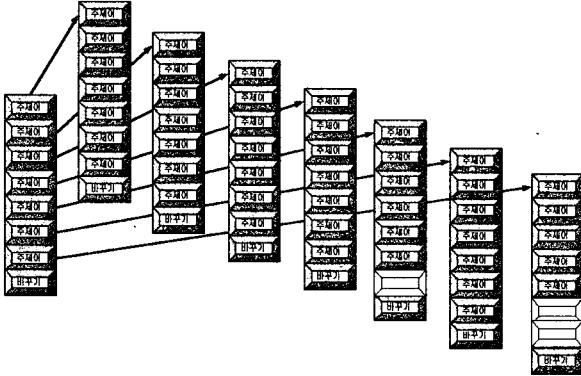


그림 3. 2단계 사용자 인터페이스

그림 3은 2단계 사용자 인터페이스를 나타낸다. 하나의 검색하려는 주제어에 대하여 2단계까지의 링크 정보를 한 눈에 알 수 있다. 기본적으로 7개의 링크 정보를 담고 있다. “바꾸기” 버튼의 클릭으로 추가 링크 정보를 알 수 있으며 7개가 안 되는 사이트는 빈 페이지로 나타난다. 기본 사용자 인터페이스에서 제공해 주는 기능을 그대로 유지하면서 주제어 위에 마우스 포인터를 위치시키면 해당 URL을 보여주고, 클릭한 사이트는 제일 왼쪽에 위치되며 2단계 검색의 뿌리가 된다.

3. 구현 및 평가

본 논문의 구현은 서버와 클라이언트 사이의 통신을 CORBA를 기반으로 구현하였다. 클라이언트는 윈도우 98 환경을 기반으로 JAVA 언어로 구현이 되었으며, 자바 개발 툴 킷으로는 JDK 1.2와 Visual Café를 사용하였다. Server는 Solaris 2.6 환경을 기반으로 C++ 언어로 구현이 되었다. 이는 다른 프로그래밍 언어, 다른 OS 환경에서의 신뢰성 있는 통신을 보여주기 위한 예를 보여준다. CORBA 클래스를 포함하는 CORBA 제품으로는 Visigenic사의 Visibroker V3.0을 사용하였다.

서버와 클라이언트 사이의 통신은 linkboy 라는 인터페이스를 통해 IDL로 작성되었다. 그림 4는 그 예이다. pInfo는 주제어에 대한 Keyword이고, pName은 URL을 나타낸다.

IDL to C++과 IDL to Java를 통해서 각각의 stub과 skeleton을 생성하여 서로의 Name 객체를 통해서 통신이 이루어지게 된다. 이로써 이식성, 단순성(네트워크 프로그램)의 이득을 얻을 수 있다. 즉, 위치 투명성과 상호 운용성의 이득을 얻을 수가 있는 것이다[8,9,10].

```
// This is Linkboy.idl
module Linkman
{
    interface
    {
        void dbCon(); //DBMS 에 연결.
        Void dbClose(); //DBMS 연결 종료
        void getData(out string pInfo,
                    out string pName);
    }
}
```

그림 4. 구현 IDL 예

4. 결론 및 향후 연구방향

본 논문에서는 서버와 클라이언트와의 이질적인 플랫폼과 이질적인 프로그래밍 언어 사이의 상호 운용성을 개선하기 위한 방법으로 CORBA를 사용하였고, 링크 정보를 통한 검색이라는 검색에 대한 새로운 인터페이스를 제공하였다.

링크 검색은 관련이 있는 정보는 서로 링크되어 있다는 관점에서 출발하여 사용자 측면에서 보다 빠르고 정확한 검색과 더욱 깊이 있는 검색을 가능하게 해 준다. 또한 접속률이 아닌 얼마나 많이 링크되어 있느냐 하는 링크율을 웹 페이지의 새로운 가치 기준으로 이용할 수 있다.

향후에는 계속 증가하는 정보에 대한 에이전트를 이용하여 링크 검색을 보다 쉽게 할 필요성이 있다[11]. 수많은 링크정보를 효과적으로 시각화 표현할 수 있는 사용자 인터페이스에 대한 연구가 이루어질 예정이다.

참고문헌

- [1] 박재현, Core CORBA, 영한 출판사, January, 1998.
- [2] Jon Siegel, CORBA Fundamentals and Programming, John Wiley & Sons, 1996.
- [3] <http://searchenginewatch.internet.com/webmasters/features.html>
- [4] <http://kachi.com/>
- [5] R. Orfali and D. Harkey, Client/Server Programming with JAVA and CORBA, John Wiley & Sons, 1997.
- [6] George M. Doss, "CORBA Networking with Java," Wordware Publishing, Inc, 1999.
- [7] 황호정, "심마니", <http://dmlab.chungbuk.ac.kr/~kwnam/search-engine/national/simmany.html>
- [8] Bryan Morgan, "CORBA meets Java," JavaWorld, October 1997, <http://www.javaworld.com/javaworld/jw-10-1997/jw-10-corbajava.html>
- [9] Bryan Morgan, "Applied CORBA: Integrating legacy code with the Web," JavaWorld, January 1998, <http://www.javaworld.com/javaworld/jw-01-1998/jw-01-corbalegacy.html>
- [10] "Programmer's Guide v3.2," Visigenic, February 27, 1998.
- [11] 이강찬, "로봇에이전트", <http://dmlab.chungbuk.ac.kr/~kwnam/search-engine/national/robotagent.html>