

인터넷 서비스 검색을 위한 정보 공유 디렉토리 PISD 설계

오경희

한국전자통신연구원 무선인터넷보안연구팀
e-mail : khoh@etri.re.kr

Design of Public Internet Service Directory for Internet Service Search

Kyung-Hee Oh

Wireless Internet Research Team, ETRI

요 약

인터넷에서 제공되는 각종 서비스들의 목록으로 구성되는 공개되고 공유되는 디렉토리 시스템과 이들로 구성된 네트워크를 통하여 정보를 검색할 수 있는 환경을 제안하여, 기존의 인터넷 검색 도구들이 가지고 있는 검색의 비효율성과 개인정보 침해의 단점을 보완하며, 정보의 생산자가 능동적으로 정보를 제공함으로써 정보 공유의 목적을 이루는 시스템을 설계한다.

1. 서론

현재 인터넷에서 원하는 서비스를 검색하는 방법으로 가장 널리 사용되고 있는 것은 야후와 같은 주제별 디렉토리나 구글과 같은 키워드 검색도구가 있다.[1] 이 두 가지 방법은 인터넷에 존재하는 방대한 정보들을 검색할 수 있는 가장 널리 사용되는 방법들이다. 그러나, 몇 가지 점에서 사용자에게 불편한 부분이 있고, 정보를 자유롭게 공유함에도 장애가 있다.

주제별 디렉토리 방식을 사용하는 검색 서비스의 경우, 서비스 제공자가 각각의 주제별 디렉토리 시스템에 자신의 서비스를 등록하거나, 디렉토리 관리자가 직접 해당 서비스를 찾아 분류하여야 등록하여야 하며, 따라서 어쩔 수 없이 검색 대상이 되는 서비스의 수가 제한될 수밖에 없다. 그리고, 키워드 검색도구의 경우, 주제별 분류가 이루어지지 않아 검색자가 원하는 분야의 검색결과를 포함하는 경우가 많아 검색결과의 질이 떨어지며, 서비스 제공자가 공개되길 원하지 않는 개인 정보들이 검색 로봇에 의하여 자동으로 수집되어 유출될 가능성이 있다. 실제 검색 사이트를 통한 개인 정보의 유출 사례가 알려져 있다.[2]

이러한 두 가지 검색 방법을 통합한 검색 서비스도

제공되고 있으나, 앞에서 언급된 문제들을 해결하지는 못한다. 또한, 두 가지 방식 모두, 웹사이트 목록 DB가 디렉토리 서버간 혹은 검색엔진 서버 사이에 공유되지 못하며, 검색 대상 서비스의 변경 사항에 대한 갱신이 지연되며, 검색 대상이 WWW 등으로 제한된다는 문제점을 갖고 있다.

서비스 제공자가 직접 자신의 서비스 목록을 디렉토리에 제공하고, 그 디렉토리들을 네트워크로 연결함으로써, 위에서 제기한 문제점들을 해결하고자 하는 것이 정보공유디렉토리 PISD(Public Internet Service Directory)의 목적이다.

기존의 디렉토리 서비스로서 LDAP[3]을 사용하는 디렉토리 시스템이 있다. 그러나 LDAP의 경우, 검색 대상이 되는 정보들이 계층적인 구조로 여러 시스템에 분산되어 있어, 광범위한 영역의 내용을 한꺼번에 빠르게 검색을 수행하여야 하고, 찾은 디렉토리 내용의 갱신이 이루어지는 환경에서 인터넷 검색도구로 사용하기에는 적합하지 않다.

서비스 제공자가 직접 디렉토리 구성에 참여함으로써 얻는 다른 이점은 서비스 제공자가 직접 자신의 서비스 내용을 기술함으로써, 기존에 널리 사용되고 있는 인터넷 서비스들뿐만 아니라 표준화되지 않은

서비스들이나 새롭게 등장하는 서비스들에 대한 디렉토리 구성도 가능하다는 것이다.

2. PISD 망

PISD 는 인터넷에서 제공되는 각종 공중 서비스들의 목록을 제공하는 디렉토리로서, PISD 를 공유하는 망을 통하여 사용자들이 쉽게 각종 인터넷 서비스들을 공유하고 검색할 수 있게 한다. 그림 1 은 PISD 를 공유하는 네트워크의 예를 보여준다.

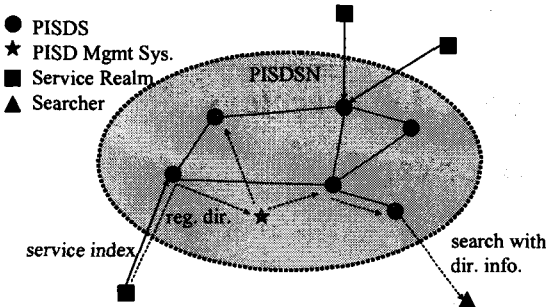


그림 1. PISD 망구조의 예

PISD 망은 Service Realm, PISDS (PISD Server), Pisd Management System, Searcher 들로 구성된다.

Service Realm 은 인터넷을 통하여 WWW, FTP 등 각종 서비스를 제공하는 시스템이다. 이 서비스들은 현존하는 인터넷 서비스들뿐만 아니라 앞으로 등장할, 아직은 알지 못하는 인터넷 서비스들일 수도 있다. 또한 표준화된 인터넷 서비스들뿐만 아니라, 각 Service Realm 에서 고유의 형태로 제공하는 서비스일 수 있다. 그리고, Service Realm 은 물리적으로 하나의 호스트일 수도 있으며, 여러 호스트에 분산되어 제공되는 서비스일 수도 있다. 그러나, 하나의 관리 권한에 의하여 관리되는 시스템을 전제한다.

PISDS 는 PISD 망에 소속된 모든 Service Realm 들이 제공하는 서비스 디렉토리를 저장하고 있는 시스템이다. 이들은 그 서비스 디렉토리를 공유하는 PISDSN (PISD Sharing Network)을 구성한다. Service Realm 은 자신이 제공하는 서비스 디렉토리를 PISDS 에 제공하기 위하여, 어느 하나의 PISDS 에 등록되어 있어야 한다. 그리고, 자신이 제공하는 서비스 목록을 그 PISDS 에 전해주게 되면, PISDS 는 전달 받은 서비스 디렉토리를 이웃한 PISDS 들로 flooding 시켜 PISDSN 전체로 전파되도록 한다. 이렇게 함으로써 PISDSN 내의 PISDS 들은 모두 동일한 서비스 디렉토리를 유지하게 된다.

PISDSN 은 공중 권한에 의하여 관리되며, 그 권한에 의하여 PISDSN 의 세부 운영에 관한 규칙들이 규정될 것이며, 이러한 규칙들이 PISDSN 의 특성을 규정하게 된다.

Service Realm 은 자신의 서비스 디렉토리를 배포하기 위하여 특정 PISDS 에 연결되어 있을 뿐만 아니라, PISDSN 내의 PISDS 가 공유하는 주제별 계층적 디렉

토리에서 특정 하위디렉토리에 소속된다. PISD 의 계층 디렉토리 구조는 PISD Management System 에 의하여 유지되며, PISDSN 내의 PISDS 에 의하여 공유된다. 검색자는 서비스 디렉토리 전체에 대한 검색을 수행할 수 있을 뿐만 아니라, PISD Management System 에서 제공하는 디렉토리 구조를 이용하여 특정 주제에 한정된 Service Realm 들에 대한 서비스 디렉토리에 대한 검색을 수행할 수 있다.

이러한 특성을 따르는 망의 응용사례에 대하여 예를 들어보겠다. 특정한 주제에 관련된 정보를 제공하는 Service Realm, R 이 있다고 하자. R 은 정보를 제공하는 방법으로 WWW 홈페이지뿐만 아니라, FTP 를 통하여 자료실의 파일을 전송 받을 수 있게 하며, mailing list archive 도 가지고 있다. R 은 자신의 서비스 목록을 PISD 에 제공하기 위해서는 먼저 PISDSN 에 소속되어있는 PISDS 들 중 하나에 등록되어야 한다. 또한 PISD 가 유지하고 관리하는 계층적인 디렉토리 구조 내에서 자신이 속하게 될 하나의 하위디렉토리에 가입을 하는 것이 권장된다. 따라서, R 의 관리자는 자신의 서비스 R 을 통하여 주제가 불분명한 다양한 정보를 제공하기 보다는 특정 주제에 중점을 맞추어 서비스를 제공할 것이다. 이제 R 이 제공하는 서비스들은 WWW 의 각 페이지들, FTP 파일들, mailing list archive 의 mail 들이 될 것이다.

R 은 자신이 가입한 PISDS 에 자신의 서비스 목록을 등록한다. 목록에는 각 서비스의 제목, 색인어, 접속방법 등이 포함된다. 그리고, 제공되는 서비스의 변동사항이 있을 때마다, 또는 정기적으로 등록된 PISDS 에 서비스 목록을 갱신한다.

PISDS 는 R 의 서비스 목록 갱신 내용을 PISD 전체로 flooding 시킨다. 갱신된 내용이 PISDSN 전체로 전파되는데 걸리는 시간은 PISDSN 의 운영방침에 따라 수십 초에서 수 일이 걸릴 수 있다. 실시간 정보들은 갱신 즉시 flooding 시키는 것이 좋을 것이며, 다소 정적인 정보들은 갱신 내용을 일정기간 모았다 주기적으로 전파하는 것이 효율적일 것이다.

사용자는 PISD 로부터 원하는 검색을 수행할 수 있다. 사용자가 특정 주제에 대하여 검색하기를 원한다면, PISD 의 디렉토리 구조를 먼저 찾아본 다음, 색인어를 이용하여 검색을 수행한다. 이렇게 함으로써 주제별 디렉토리 방식과 키워드 검색 방식을 동시에 활용하는 검색이 이루어진다.

3. PISD 프로토콜

PISD 메시지는 Service Realm 과 PISDS, 또는 PISDS 상호간에 서비스 디렉토리를 전달하기 위하여 주고 받는 메시지이다. PISD 메시지는 다음 네가지의 메시지 유형으로 나뉜다.

Update-Start

Service Realm 에서 PISDS 로 서비스 목록을 등록하기 위하여, 또는 PISD 에서 PISD 서비스 목록을 flooding 하기 위하여 갱신을 요청하는 시스템이 연결을 맺기 위하여 보내는 메시지이다. 이 메시지에는 갱

신하고자 하는 Service Realm 의 식별자와 Service Realm 의 최종 갱신 시간이 포함되어 있다. Service Realm 의 식별자는 그 Service Realm 이 관리되고 있는 시스템의 도메인 이름과 그 시스템 내에서 유일한 Service Realm 식별자로 구성된다. Service Realm 에서 PISD 로 보내는 Update-Start 메시지의 경우, 시스템의 도메인 이름은 생략될 수 있으며, 생략된 경우에는 메시지를 보낸 시스템의 도메인 이름이 그 값이 된다. PISDS 는 Update-Start 메시지에 포함된 최종 갱신 시간 값을 통하여, 갱신 메시지가 보내는 정보가 자신의 데이터베이스 갱신에 이미 반영되어 있는지 확인할 수 있다.

Difference-Request

Update-Start 메시지에 대한 응답으로, 또는 PISDS 가 서비스 목록의 갱신을 요청한 Service Realm 또는 PISDS 로 갱신 내용을 요청하는 메시지이다. 이 메시지에는 갱신을 원하는 Service Realm 식별자와 자신의 데이터베이스에서 해당 Service Realm 이 갱신된 최종 시간, 그리고 한번의 메시지에서 전송 받을 수 있는 서비스 목록의 수가 포함되어 있다. 이 메시지는 Update-Start 메시지의 응답으로 이 메시지가 보내어 지거나, 정기적으로 PISDS 가 Service Realm 의 서비스 목록을 갱신하는 경우에는 Update-Start 메시지 없이 보내어 질 수도 있다.

Difference-Response

Difference-Request 메시지에 대한 응답으로 갱신된 서비스 목록을 전달한다. 이 메시지에는 Difference-Request 메시지에서 요청한 시간 이후에 추가 또는 수정되거나 삭제된 서비스들의 목록이 포함되어 있다. 또한 갱신하여야 할 서비스의 목록 수가 Difference-Request 메시지에서 허용한 최대 목록 수보다 큰 경우에는 더 갱신해야 할 서비스 목록이 남아 있음을 알리는 식별자를 포함하고 있다.

Difference-Response 메시지에 대한 성공적인 Status-Notify 응답 메시지를 받고, 갱신하여야 할 서비스 목록이 남아있다면, Difference-Response 메시지를 계속하여 보낼 수 있다.

Status-Notify

Update-Start 메시지에 대한 실패 응답, Difference-Request 메시지에 대한 실패 응답, Difference-Response 메시지에 대한 성공 또는 실패 응답 등의 상태정보를 나타내는 메시지이다.

그림 2 는 이러한 메시지들이 Service Realm 및 PISDS 사이에서 전송되는 순서를 보여준다.

Service Realm 에서 갱신된 서비스 목록을 등록하고자 한다면, Update-Start 메시지를 PISDS 로 보낸다. PISD 서버는 Update-Start 메시지를 보낸 Service Realm 이 자신에게 등록되어 있는 지 여부와 권한을 확인하여, Service Realm 등록되어 있지 않거나, 갱신이 PISDS 에 의하여 수동적으로만 갱신될 수 있도록 설정되어 있다면, Status-Notify 메시지를 통하여 갱신을

거부한다. 그렇지 않다면, Difference-Response 메시지를 보내어 갱신을 허가할 것이다.

PISDS 에서 Service Realm 에 대한 서비스 목록을 정기적으로 갱신하도록 설정되어 있다면, Service Realm 으로부터 Update-Start 메시지 없이 PISDS 가 Difference-Response 메시지를 Service Realm 으로 보낼 것이다.

Service Realm 은 Difference-Response 메시지의 요청에 따라 갱신된 서비스 목록을 Difference-Response 메시지에 담아 전송한다.

PISDS 는 Difference-Response 메시지에 포함된 서비스 목록을 자신의 데이터베이스에서 갱신한 후, Status-Notify 메시지를 통하여 갱신 결과를 Service Realm 에 알려준다.

PISDSN 에서 서비스 목록을 flooding 하기 위한 PISDS 사이의 서비스 목록 갱신 과정은 Service Realm 에서 PISDS 로의 서비스 목록 갱신 과정과 거의 동일하다. 단, Update-Start 메시지가 항상 사용된다. 그리고, 이미 다른 PISDS 로부터 갱신 받은 서비스 목록에 대한 Update-Start 메시지를 받는 경우에는 Status-Notify 메시지를 통하여, 이미 갱신되었음을 알려줌으로써 갱신과정을 종료하게 된다.

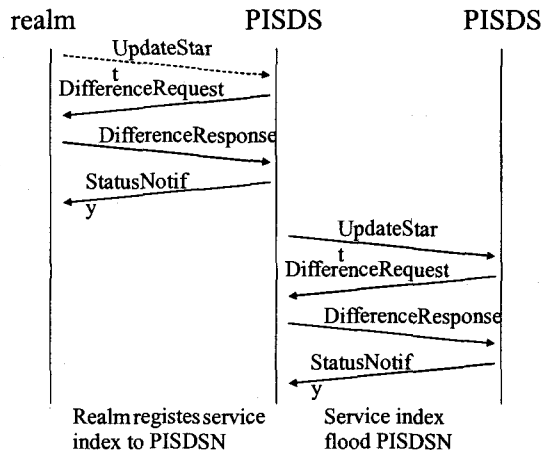


그림 2. PISD 프로토콜을 통한 서비스 목록 갱신 과정

4. 실험

PISD 프로토콜에 사용되는 메시지들을 ASN.1[4]으로 정의하였으며, snacc[5]을 사용하여 컴파일하였다. 실험을 위하여 Service Realm 클라이언트, PISDS 클라이언트, PISDS 서버를 구현하였으며, MySQL 을 각 시스템의 데이터베이스로 사용하였다. 개발환경은 Debian Linux 시스템이다.

구현된 시스템은 그림 3 과 같다. Service Realm 에서 제공하는 HTML 문서정보를 추출하여 데이터베이스로 구축하였으며, Service Realm 클라이언트가 데이터

베이스로부터 이 서비스 목록을 읽어와 PISDS 서버로 전송한다. PISDS 서버는 이로부터 데이터베이스를 갱신하고, 동일한 시스템의 PISDS 클라이언트가 데이터베이스의 갱신된 내용을 다른 PISDS 서버로 전달하여 flooding 시킨다.

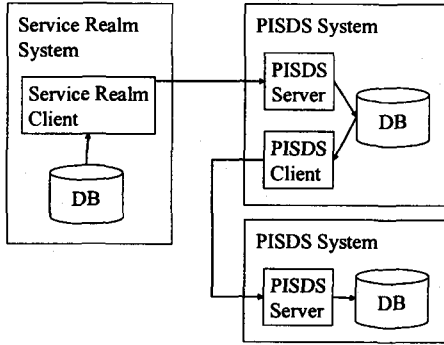


그림 3. 실험을 위하여 구현된 PISD 망

5. 결론

PISD 를 완성하기 위해서 우선 PISD 에서 사용되는 프로토콜이 정의되어야 한다. 실험에서는 서비스를 기술하는데, 식별자, 제목, 서비스에 접속하기 위한 URL 만을 사용하였다. 그러나 서비스의 검색을 위해서는 서비스의 갱신 버전이나 검색어, 저자 등 서비스를 기술하기 위한 보다 자세한 구조가 정의되어 사용되어야 할 것이다. 또한 이 구조는 다양한 인터넷 서비스를 기술하는데 적합해야 한다. 그리고, PISD 프로토콜에 따른 서비스 목록을 제공할 수 있는 인터넷 서비스들이 필요하다. 모든 인터넷 서비스들은 서비스 목록을 제공할 잠재성을 가지고 있기에 기존의 인터넷 서비스들로부터 서비스 목록을 추출하는 작업이 필요하다. 그리고, 운영자의 운영원칙에 따라 service realm 의 서비스 목록과 다른 PISDS 의 목록을 PISDSN 에 배포하는 범용 PISDS 와 PISDSN 전체를 관리하기 위한 PISD Management System 의 개발도 필요로 한다.

그러나 이러한 시스템의 개발뿐만 아니라, 서비스 제공자들이 PISD 를 통하여 서비스를 제공하도록 유도하는 환경과, PISD 를 어떻게 민주적으로 운영하는 것도 중요한 과제가 될 것이다. PISD 에서 정보의 흐름은 Usenet[6]과 유사하다. 운영방식에서도 많은 참고를 할 수 있을 것이다.

PISD 는 아직 구현된 시스템이 아니다. PISD 는 앞으로 구현될 시스템이며, 그 목적이 공공영역에서의 정보 공유에 있으므로, 개발과정도 공공영역에서 많은 이들의 의견을 바탕으로 이루어지는 것이 바람직하리라 생각된다.

참고문헌

[1] 임영이, 이동일, “웹에서의 정보 검색”, 주간기술동

향, 통권 874 호, 1998 년 11 월.

[2] “검색사이트 통한 개인정보 노출 심각”, 연합뉴스, 2003 년 1 월 14 일.

[3] Mark Wilcox, ‘Implementing LDAP’, Wrox Press, 1999.

[4] Olivier Dubuisson, ‘ASN.1 – Communication Between Heterogeneous Systems’, Academic Press, 2001.

[5] ‘Snacc Homepage’,

<http://www.fokus.gmd.de/ovma/freeware/snacc/>

[6] Mark Moraes, Gene Spafford, ‘What is Usenet?’,

<http://isc.faq.org/faq/usenet/what-is/part1/>