

분산환경에서의 WebView 구체화 모델을 위한 사용자 서비스 확장

서진형*, 김경창**

*경인여자대학 컴퓨터정보기술학부

e-mail : jhsuh@yoda.kic.ac.kr

** 홍익대학교 공과대학 정보컴퓨터공학과 컴퓨터공학과

e-mail : kckim@hongik.ac.kr

User service extension for WebView materialized model in distributed environment

Jin Hyung SUH*, Kyung Chang KIM**

*School of Computer & Information Technology, Kyungin Women's College

**Dept. of Computer Engineering, Hong Ik University

요 약

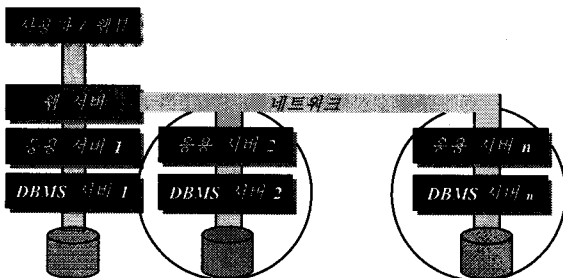
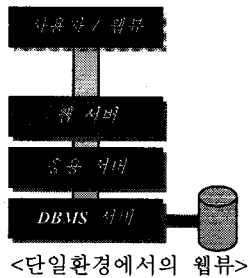
분산환경에서의 웹뷰 (WebView)는 기본적으로 3 단계 구조 (3-tier architecture)를 기본으로 하고 있으며, 이를 통하여 사용자가 요구하는 정보를 구체화시켜 제공하게 되나 현재 인터넷 환경에서는 사용자가 요구하는 정보가 한곳에만 있다고 볼 수 없다. 즉, 적절히 분산되어 있는 각 웹 서버의 정보를 다양하게 활용할 수 있는 환경을 제공하여야 사용자가 요구하는 정보를 얻을 수 있으며, 이를 위한 적합한 데이터 저장 및 웹뷰 유지 방법이 제시되어야 한다. 그러나 일반적인 3 단계 구조를 적절히 분산되어 있는 환경에 적용할 경우 사용자의 질의에 해당하는 정보를 추출할 수 있는 방법, 네트워크 연결시의 지연과 네트워크의 장애등으로 인하여 많은 시간적인 문제와 상대방에서의 잘못된 정보의 전달등으로 인하여 사용자가 요구하는 정보를 전달할 수 없으며, 이에 따른 추가적인 비용 모델이 산정하여야 한다. 이러한 이유로 본 논문에서는 기존의 단일 3 단계 모델을 확장한 분산환경에서의 사용자 서비스와 가장 관계가 있는 계층으로 OSI 7 레이어에서 응용 계층으로 사용자 서비스와 응용 계층의 사이에서는 제공하고 있는 분산 정보 서비스에 대한 조건의 확장을 통하여 추후 고려하여야 할 분산된 네트워크 웹뷰 모델에서의 성능 및 비용 분석에 대비한다.

1. 서론

데이터베이스에서의 뷰 (view)는 분산되어 있으면서 서로 연관성이 있는 (loosely coupled) 데이터베이스에 저장되어 있는 관계와 유추된 데이터베이스 테이블 등을 이용하여 원하는 데이터를 요구하는 사용자에게 보여주는 것으로 보여줄 때 마다 반복하여 계산을 하게 된다. [gupta95] [nick01] 이에 반하여 웹뷰 (WebView)는 사용자가 요구할 것으로 예상되는 결과

를 데이터베이스에 저장되어 있는 자료 또는 기본 데이터 테이블을 이용하여 미리 계산하여 저장하고 있다가 사용자에게 제공되는 것으로 특히 HTML 형태의 페이지로 보여주는 것을 의미한다. 즉 기본적으로 가지고 있는 자료 (base data)로부터 사용자가 작성한 질의어를 통하여 얻어낸 결과 (query result)를 요구한 사용자에게 볼 수 있는 웹의 형태로 보여주는 페이지 [alex00], [alex02], 웹 페이지에 대한 재사용을 도와주

는 그래픽화된 도움방법 [andy99] 또는 다양한 검색 방법을 지원하기 위해 웹 데이터에 대한 뷰를 정의하는 것으로 정의할 수 있다. 그러나, 이 웹뷰에서 질의 성능 향상을 위해서는 구체화된 웹뷰 (materialized WebView)를 저장 공간의 효율성을 위해서는 가상 웹뷰(virtual WebView)로 구현하는 것으로 정의되고 있으나, 새로운 데이터가 계속하여 동적으로 생성되어 생성된 데이터를 저장하기 위하여 저장되는 장소가 계속하여 확장이 되어야 한다는 문제점을 내포하고 있어 이를 해결하기 위하여 웹뷰 구체화가 가장 효율적인 방법으로 간주된다. 다음으로 동적으로 데이터를 계속하여 생성하기 때문에 데이터에 대한 신뢰성 또는 확실성 들을 따지는 데이터의 질 (Quality of Data; QoD)에 있어 많은 문제점을 내포하고 있다. 이러한 상황에 대하여 해결할 수 있는 방안으로 제시한 웹뷰 구체화 방법으로 구체화를 하지 않은 경우 (no materialization), 데이터베이스 서버에서 구체화를 시키는 경우 (materialize inside DBMS) 그리고 마지막으로 웹/응용 서버에서의 구체화를 시키는 경우 (materialize inside web/application)가 있으며 이중 웹/응용 서버에서 구체화 방법이 서버에 무리를 주지 않으면서 사용자가 요구하는 데이터를 제공할 수 있다. [alex00] 그리 제시된 구체화 방법은 일반적인 3 단계 구조에 적용된 것으로 실제 분산 네트워크 환경에서 구체화 방법을 적용하기 위하여 분산된 환경에 대한 변수 들을 적용하여야 하며, 이를 위한 조건을 확인하여야 한다.

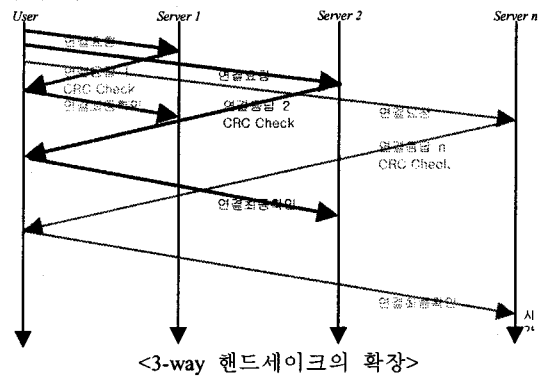


2. 분산된 네트워크 환경에서의 웹뷰 구체화를 위한 확장 조건

분산된 네트워크 환경에서의 웹뷰 구체화를 위하여 웹뷰 사용자 서비스와 관계가 있는 응용 계층을 통하

여 사용자 서비스와 응용 계층의 사이에서 분산 정보 서비스를 제공하고 있다. [이병기 00] 또한 OSI 7 레이어의 응용계층에서 적용되어야 하는 기능인 이름의 부여 (naming), 자원의 공유 (sharing), 자원의 가용성 및 신뢰성 (availability and reliability), 데이터의 복사 (replication), 비밀 유지와 보안성 (privacy and security), 프로세서 또는 서비스 간의 통신 (communication), 프로세서 또는 서비스 간의 협력과 동시성 (concurrency & synchronization), 시간성과 데이터 일치성 (time and coordination), 데이터의 구성과 회복 (fault tolerance and recovery) 그리고 데이터의 확장성 (scalability) [crichlow00]이 분산 네트워크 환경에서 웹뷰를 적용할 경우에 사용자의 질의에 대한 서비스에 대하여 해결하여야 하는 항목으로 볼 수 있다.

(1) 지정 착신지의 확인 및 호출 가능성 확인
 분산된 네트워크 환경에서의 웹뷰는 사용자가 요구한 데이터가 있는 저장 장소를 확인하고 그 저장 장소에서 원하는 데이터를 가져올 수 있는 지에 대하여 확인할 수 있어야 하는 것으로 송신 및 착신을 위하여 일반적인 네트워크 상에서의 연결설정 및 해제 기본인 3-way 핸드셰이크를 사용할 수 있으나, [배시규 02] 3-way 핸드셰이크의 경우 요청, 허락, 그리고 연결허락에 대한 응답만이 있어, 연결 허락을 통한 데이터의 확인 작업과 오류의 확인 과정을 수반 하여야 하며, 특히 단순한 클라이언트-서버 환경이 아닌 다수의 클라이언트와 서버 환경에서의 호출 가능성을 확인하여야 한다. 즉, 3-way 핸드셰이크의 확장에 추가적인 작업에 대한 귀환 작업을 수행할 수 있어야 하며, 이 귀환 작업으로 인한 다양한 네트워크 연결이 필요로 하며, 특히 전체 연결 사이트를 확인할 것인지 아니면 일부 사이트만을 확인할 것인지 역시 검토하여야 할 사항이다. 이때 제시할 수 있는 방법으로는 각 사이트를 일일이 확인하는 폴링 (polling) 방법을 사용할 수 있으나, 이 경우 사용 가능한 모든 사이트를 점검하여야 하므로, 제한적으로 사용하는 것이 효율적이라 할 수 있다.



(2) 통신 권한 설정
 원하는 장소에서의 데이터를 가지고 올 수 있다 하더라도 통신권한의 설정이 되지 않으면 원하는 데이터를 가지고 올 수 없다. 이런 경우에 대하여 데이터를 수신 받으려는 사이트에 대한 권한 즉, 인증에 대한

처리가 필요로 하다. 인증이후 데이터를 전송받게 되는 중간에 제 3 자에 의한 정보의 절취등으로 문제가 발생할 수도 있으나 이는 본 연구에서는 제외한다.

(3) 지정된 착신지 진위 여부 입증

일단 통신 권한이 설정되어 원하는 데이터를 가지고 올 수 있는 상황이 되어도 주어진 데이터의 진위여부를 확인할 수 없게 된다. 분산 네트워크 상에 접속되어 있는 다양한 사이트에 대한 진위여부 입증이 필요하며, 진위 여부의 입증이 되지 않음으로 인하여 이에 대한 재검증을 필요로 한다. 실제 이로 인하여 많은 지연 시간이 발생하게 되어 전체 시스템의 성능에도 영향을 미칠 수 있다. 또한 데이터의 복사, 자원의 가용성 및 신뢰성 등이 영향을 미치게 된다. 이때 발생하는 네트워크 상의 왕복지연시간 추정은 일반적인 네트워크 상황에서는 송신 측에서 검침 (probe) 패킷을 전송하고 난 이후 송신시간, 수신시간, 응답시간, 응답 도착 시간을 기반으로 하여 이루어지며, 이를 일반적인 수식으로 표현하면 다음과 같다.

$$\begin{aligned} \tau_{arrive} &= \tau_{send} + \delta + \tau_{offset} \\ \tau_{loopback} &= \tau_{sendback} + \delta - \tau_{offset} \\ rtt &= \tau_{loopback} - \tau_{send} - \tau_{sendback} + \tau_{arrive} \end{aligned}$$

그러나 이 경우에는 일반적인 상황에서의 수신시간, 응답시간, 그리고 응답 도착 시간을 기준으로 작성한 것이며, 이를 네트워크 기반의 웹뷰에 적용할 경우에는 웹뷰의 수 또는 웹뷰를 구성하는 네트워크의 수에 따라 수신시간, 응답시간, 그리고 응답 도착 시간이 달라지게 된다.

(4) 시작과 종료 통신절차를 포함한 대화방법의 선택
통신 절차의 시작과 종료는 기본적으로 데이터의 전송과 관련이 있는 부분으로 언제 데이터 전송이 시작되는 지 그리고 종료되는 지에 따라서 전체 성능에 영향을 미칠 수 있다. 이 대화방법은 기본적으로 네트워크에서의 흐름제어와 같은 방법으로 통신 절차를 구성하게 되는데, 먼저 모든 것을 개방한 상태에서 미리 전송할 데이터의 종류와 특성 들을 통보하고 연결이 유지되는 동안에는 계속하여 통신을 지속하는 방법 또는 네트워크의 상태에 따라 대화를 할 것인가 말 것인가를 결정하여야 한다. 마지막으로서는 이를 혼합한 형태가 있는데, 실제 분산된 네트워크에서의 웹뷰는 대부분 혼합된 형태의 대화 방법을 선택하는 것으로 간주하여야 하며, 이 경우 명시적으로 자체 네트워크의 현재 상황을 감지하여 상황 정보를 근거로 하여 대화를 시작하는 경우와 묵시적인 방법으로 데이터를 전송하여야 하는 서버 또는 응용측에서 네트워크의 상황을 판단하여 데이터를 전송할 것인가 또는 말 것인가를 결정하게 된다. 후자의 경우가 웹뷰의 근간이기는 하나 실제 네트워크의 상황 판단의 경우 전자는 현재 시스템에서 제공하는 상황 정보를 기반으로 하고 후자의 경우는 데이터를 전송하여야 하는 전송자에 의한 상황정보이기 때문에 이에 따른 시스템 자체 오버헤드가 커지게 된다.

(5) 오류 복구 책임소재 협상

통신이 시작되고 나서 발생한 오류 또는 통신 불능으로 발생한 오류에 대한 책임 소재를 확인하고 이에

대한 오류 복구를 시행한다. 일반적으로 오류 복구와 제어에 대하여는 응답과 재전송을 기본으로 하는데, 실제 일반적인 TCP/IP 환경에서의 응답과 재전송은 1 대 1 상황에서의 응답과 재전송을 기본으로 하고 있으나, 이를 분산된 웹뷰 상태에 적용할 경우에는 응답과 재전송이 "1 대 1"이 아닌 "1 대 다"로 이루어진다는 것이 특징으로 이에 대한 제어와 더불어 책임소재의 확인도 복잡해지게 된다. 특히 데이터의 폭주로 인하여 발생하는 오류의 경우에는 누구에게도 이 경우에 대한 책임 소재를 가릴 수 없어 대표적인 시스템 관리자에 의한 오류 복구가 요청된다.

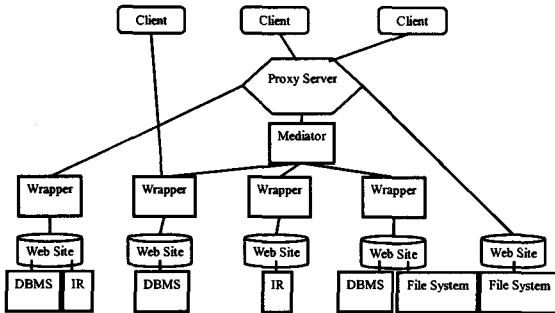
(6) 데이터 형식 제약 사항

실제 Web 상에서 입력되는 데이터는 HTML과 같은 미리 정의된 형식이 없는 문서의 형태로 저장되며 데이터의 크기는 기존의 데이터와는 달리 매우 크기가 큰 형태이다. 또한 기존의 데이터베이스와는 달리 관리자에 의하여 관리를 하는 형태가 아니어서, [barry01] 데이터 형식의 제약 사항을 일반적으로 규정하는 것에는 문제가 있다. 즉, 웹 또는 웹뷰를 위하여 사용되는 비정형적인 스키마 (semistructured schema)를 갖는 비정형 데이터 (semistructured data)에 관한 연구와 Lorel [serge97]과 UnQL [buner96]과 같은 비정형데이터에 대한 질의어의 확장이 요구되는 상황이다. 이를 위하여는 비정형적인 데이터에서 구조 정보를 추출하여 데이터 모델에 적합한 데이터로 변환을 하여야 하며, HTML 문서에서 비주요 도구를 이용하여 구조를 추출하는 NoDoSE [adel98]과 비정형 데이터 모델인 OEM 데이터로부터 스키마를 추출하는 DataGuide [gw97]이 있다. 그러나 이들은 비정형 데이터에 관련된 데이터 모델을 통한 데이터의 처리만을 다루고 있으나, 웹뷰와 같이 휘발성이 높은 데이터의 구조를 손쉽게 재설정할 수 있어야 한다.

(7) 데이터 통합

데이터 통합 (data integration)은 기본적으로 분산된 네트워크 환경에서 이산적인 웹 사이트 (discreted web site)들에 대한 단일화된 웹뷰를 사용자에게 제공하고, 효과적인 분산 질의의 처리하기 위한 것으로 웹뷰의 유지보수 전략을 추. 이에 대한 기존의 연구로는 각 웹 사이트 고유의 질의 방식과 질의 결과 형태를 공통의 것으로 변환하기 위한 래퍼 (wrapper) [ak97], 여러 웹 사이트에 대해 질의를 수행하고, 질의 결과를 통합하기 위한 메디에이터 (mediator) 등이 있으나, 여기에서의 데이터의 통합은 일반적으로 어떤 방식으로 데이터를 데이터 저장장소로부터 가지고 와서 어떻게 데이터를 갱신할 것인가를 중요시 하는 관점으로서의 데이터 통합으로 볼 수 있다. 그러나, 웹뷰에서의 데이터 통합이란 서로 다른 개발자들이 구축한 2 개 이상의 웹/응용들이 서로 정보를 교환할 수 있도록 만들어야 하는 일련의 과정을 의미하고, 이 과정엔 거의 예외없이 추출 (extraction), 전송 (transport), 변환 (transformation), 활용 (loading)이란 4 단계를 포함한다. 또한 단순한 정보 중개 (information-brokering)이라는 개념이 아닌 얻어낸 정보를 하나의 보기 좋은 사용자 뷰 (user view)로 제공하고 이를 유지하는 것이 그 목

적이라고 할 수 있다. 그러나 어떠한 경우에서 라도 일단 웹뷰에서의 데이터 통합에서 발생하는 가장 큰 문제는 비용 부분이다. 우선 데이터를 성공적이고 비용효과적으로 통합, 제공하려면 단순히 데이터 또는 래퍼등을 통하여 제공된 데이터를 통합하는 것 이상의 작업이 필요하다는 점을 인식해야만 한다. 일례로 정보 증개 방식을 사용하는 경우에도, 인간에 의한 개발 및 지속적인 관리 작업이 필요하다. 다시 말해 이 방식도 코드 관리, 버전 통제, 데이터 변환 도구의 사용자화 등 작업을 추가하여야 한다. 그리고 어떤 데이터가 필요한 데이터인가를 파악하는 일, 이 데이터가 새로운 응용에 문제없이 활용될 수 있도록 데이터를 변환하고 청소하는 방법 등은 많은 경우 상당히 어렵고 비용이 많이 들어가는 작업이며, 이는 기존에 언급하였던 모든 문제점이 복합적으로 작용하게 된다는 점도 다른 문제점보다 처리하는 데 있어 복잡할 수 있다.



<일반적인 웹 구조에서의 데이터 통합>

3. 결론

분산된 네트워크 환경에서의 웹뷰 구체화를 위한 확장된 조건을 검색하여 보면 실제 일반적인 분산 시스템에서의 해결하여야 하는 문제를 확대한 것으로 파악할 수 있으며, 추가로 기존의 3 단계 모델과 이에 적합한 질의어의 확장 역시 요구된다. 차후 연구될 사항으로 확장된 조건에 대한 사용자의 질의와 사용자 서비스를 지원하기 위한 인터페이스의 제안을 필요로 한다. 다음으로 지금까지 제시된 확장 모델의 경우 모두 완전히 연결되어 있다는 가정하에 확장되어야 할 조건을 확인하였으나, 완만하게 연관되어 분산되어 있는 네트워크 환경에서 다양한 리소스로부터 데이터를 전송 받아 이를 사용자에게 제공하는 데이터베이스가 정적인 데이터를 보여주는 반면 동일한 환경에서 적용되는 웹뷰는 동적인 데이터를 사용자에게 보여주게 되어 제시된 환경에 대한 상세한 분석, 검증 그리고 이에 대한 시뮬레이션 결과와 네트워크의 단절로 인한 오류 발생 [alex02] 에 대한 연구도 필요로 한다.

참고문헌

[이병기 00] 광대역 정보통신, 이병기, 강민호, 이종희

공저, pp373-375, 2000, 교학사, ISBN-89-09-02784-3
 [배시규 02] TCP/IP 네트워크 및 프로그래밍, 배시규 저, pp106, 2002, 21 세기사, ISBN 89-8468-068-0
 [adel98] B. Adelberg. Nodose "a tool for semi-automatically extracting structured and semistructured data from text documents" Proceedings of SIGMOD'98, pp 283-294, 1998.
 [ak97] Naveen Ashish and Craig A. Knoblock, "Semi-Automatic Wrapper Generation for Internet Information Sources", In Proceedings of the Second IFICIS International Conference on Cooperative Information Systems (CoopIS), Charleston, SC, 1997
 [alex00] Alexandros Labrinidis and Nick Roussopoulos, "WebView Materialization" In Proc. of the ACM SIGMOD Intl Conf on Management of Data, Dallas, Texas, USA, May 2000.
 [alex02] Alexandros Labrinidis, "WebView Materialization for Scalable Web Servers", Colloquia, University of California Santa Barbara, May 2002
 [andy99] Andy Cockburn Saul Greenberg Bruce McKenzie Michael Jasonsmith Shaun Kaaste , "WebView: A Graphical Aid for Revisiting Web Pages", <http://www.cpsc.ucalgary.ca/projects/grouplab/papers/1999/99-WebView.Ozchi/webview-ozchi.pdf>
 [barry01] Web Database Systems, Barry Eaglestone and Mick Ridley, pp35, McGraw Hill, 2001
 [buner96] Peter Buneman and Susan Davidson and Gerd Hillebrand and Dan Suciu, "A query language and optimization techniques for unstructured data", In Proceedings of ACM SIGMOD International Conference on Management of Data, pages 505-516, Montreal, Canada, June 1996
 [crichlow00] The Essence of Distributed Systems, Joel M. Crichlow, pp14-17, Prentice Hall, 2000
 [gupta95] Ashish Gupta, Inderalp Singh Mumick, "Maintenance of Materialized Views: Problems", Techniques and Applications, Data Engineering Bulletin, 18(2):3-18, June, 1995
 [mark96] Evangelos P. Markatos and Antonis Danalis: Web Caching. Enterprise Networking: Multilayer Switching and Applications, editors V. Theoharakis and D.N. Serpanos, pp. 383-408, Idea Group Publishing
 [nick01] Nick Roussopoulos, Yannis Kotidis, Alexandros Labrinidis and Yannis Sismanis, "The Opsis Project: Materialized Views for Data Warehouses and the Web". In Proc. of the 8th Panhellenic Conf. on Informatics, Nicosia, Cyprus, Nov. 2001.
 [serge97] Serge Abiteboul and Dallon Quass and Jason McHugh and Jennifer Widom and Janet L. Wiener, "The Lorel Query Language for Semistructured Data", International Journal on Digital Libraries vol.1. no.1 pp68-88, 1997