

계층 구조의 Deck를 갖는 HTML Filter 구현 방안

김 대건, 이 승진, 강 경용, 최 린, 강 철희

고려대학교 전자공학과

전화 : 02-929-5625 / 핸드폰 : 018-353-9023

Dae-Geon Kim, Seung-Jin Lee, Kyung-Young Kang, Lynn Choi, Chul-Hee Kang
Dept. of Electronic Engineering, Korea University
E-mail : dkim@widecomm.korea.ac.kr

Implementation of HTML Filter with structured Decks

Abstract

This paper has been studied a implementation of HTML Filter with structured decks. In wireless environment, conventional HTML Filter converts HTML document to WML by once. However, it is more efficient to access a WEB site by transmitting the minimal capacity of core data in WAP environment. Therefore this paper suggests new HTML Filter that automatically analyze the contents of HTML WEB and transmits it by selecting the one of many Decks that have the structured architecture. Also the protocol between wireless terminal and WAP Gateway is considered. Finally, We test the proposed HTML Filter with structured Decks for some particular web site.

I. 서론

일반적으로 인터넷의 웹 페이지를 액세스하면 다른 불필요한 정보로 인해 웹 페이지가 느리게 디스플레이 되는 경우가 자주 발생된다. 이러한 현상은 WAP 환경에서는 더욱 빈번히 발생되고 너무 지연되면 연결이 끊어질 수 있다. 따라서 WAP 환경에서는 꼭 필요한 정보만 전송하도록 하여 데이터 전송량을 최소화하는 것이 무선인터넷을 효율적으로 액세스할 수 있는 한 가지 방법이다. 또한, 기존의 HTML Filter는 많은 정보를 담고있는 복잡한 웹 페이지의 HTML 문서를 한

꺼번에 WML 문서로 변환하므로 WAP환경의 대역폭, 전송속도, 디스플레이 화면, 메모리, CPU 성능의 제한으로 효율적이지 못하다. 따라서, 본 논문에서는 HTML Filter가 HTML 웹 페이지 내용을 자동으로 분석하여 WML로 변환하여 계층 구조의 Deck을 갖는 방식으로 적절한 Deck을 선택적으로 전송하는 방안에 대해 논의하고 무선이동단말기와 WAP Gateway 간에 이 계층구조의 Deck을 액세스하기 위한 프로토콜을 제시하고 특정 웹 콘텐츠를 WML로 변환하여 계층구조의 Deck의 구현 예를 제시한다.

II. 계층구조 HTML 구현

그림 1은 계층구조의 Deck을 갖는 HTML Filter 모델을 나타낸다. 그림1에서 Client에서 Web server를 액세스하기 위한 WML Encoded request를 하면 WAP Gateway에서는 관련 Deck이 있는지 여부를 확인한 후 없으면 WEB server에 HTML request한다. WEB server에서는 HTML request가 요구하는 데이터 처리를 하여 HTML response을 Gateway로 송신한다. Gateway에서는 HTML response을 수신하여 새로운 Contents 내용을 HTML Filter에서 계층구조 Deck으로 생성하여 저장한다. 이 저장된 계층구조 Deck 정보를 Gateway가 Encoder를 거쳐 Client로 송신한다. 이러한 Client와 WAP Gateway 사이의 프로토콜은 WSP(Wireless Session Protocol)에서 처리하는 데 계층구조 Deck을 갖는 HTML Filter을 구현하기 위해서는 기존의 WSP(Wireless Session Protocol)의 cache-directive를 이용하여 계층구조 Deck을 액세스

할 수 있다. WAP Gateway의 HTML Filter는 복잡한 HTML문서를 WML 문서로 변환할 때 계층 구조를 갖는 여러 개의 Deck으로 생성하고 Client가 맨 처음 해당 사이트를 액세스시 Client쪽에 최상위의 Deck의 정보를 준다. 해당 웹사이트에서 Client가 하위 Deck을 request하면 Gateway의 에서는 request에 해당하는 하위 Deck이 있으면 바로 response를 해주고 해당 하위 Deck이 없으면 server에 request하여 하위 Deck을 생성한다. 즉, 상위계층의 Deck에서 적절한 목록을 선택하면 하위계층의 Deck을 액세스할 수 있다. 계층구조 Deck의 장점은 다음과 같다. 첫째, 복잡하고 많은 콘텐츠 내용을 효율적으로 짧은 여러 개의 Deck로 변환하여 불필요한 데이터를 전송하지 않는다. 둘째, WAP 환경의 특성상 Client 단말기의 메모리와 CPU 용량을 많이 필요로 하지 않는다. 셋째, WAP 단말기의 특성을 고려하여 짧은 메시지로 큰 화면의 단말기를 필요로 하지 않는다. 반면, 다음과 같은 문제점이 나올 수 있다. 짧은 메시지로 인한 빈번한 Client의 request로 Client와 Gateway 사이의 signal의 overhead가 증가할 수 있고, Gateway와 server사이에서는 계층구조를 갖는 Deck을 저장하는 데이터베이스의 신뢰성을 보장하기 위해 여러 번의 Refresh request/response로 유선망에서의 대역폭 사용량이 증가할 수 있다. 그러나 이러한 문제점은 계층구조 Deck의 HTML Filter가 자주 액세스하는 웹사이트에 대해서만 주기적으로 계층구조 Deck을 Refresh해 주어 대역폭 사용량 증가를 최소화할 경우 기존 HTML Filter에서의 액세스 지연시간을 감소시켜 향상된 기능을 제공하므로 전체 성능은 보다 효율적이라 할 수 있다. 따라서, 향후 signaling overhead와 delay time을 서로 비교 검토해 볼 필요가 있다.

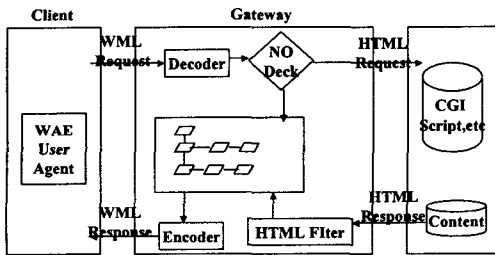


그림1. 계층 구조 Deck을 갖는 HTML Filter 모델

그림 2는 본 논문에서 구현한 계층 구조 Deck을 갖는 HTML Filter의 구조를 보여준다. HTML Filter의 기능 블록을 보면 HTML/WML 웹 콘텐츠를 HTTP를 이용하여 서버로부터 수신한 후 수신한 문서가 WML이면 Filtering을 하지 않고 WAP Gateway를 통해 WAP Browser에 전송한다. 만일, HTML 문서이면

HTML Filter는 Parsing Engine을 통해 HTML Tag를 분석하고 미리 정의된 변환 규칙 Rule Set Database를 참조하여 Markup Language Translator를 통하여 WML 문서로 변환하고 structured Deck 구조를 생성한다.

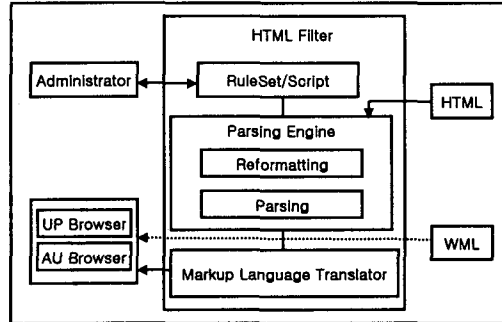


그림2. 계층구조 Deck을 갖는 HTML Filter 기능 블록도

2.1 Rule set

Rule Set Database에는 관리자가 정의한 변환규칙과 스크립트(scripts)가 저장되어 있으며, 여기에는 클라이언트(Client) Device의 지원 기능과 사용자의 입력에 의한 추가 데이터가 입력될 수 있다. Rule Set Database는 각 HTML 태그에 대해 Tag Type을 정의하고 있으며 이는 HTML Filter에서 수행되는 각 태그에 대한 변환 처리 Rule을 명시하고 있다. Tag Type은 다음과 같다.

- valid : HTML Filter의 변환 처리가 가능하며 단순히 TagSet만을 이용한 변환을 한다.
ex)
 ->
, ->
- validAttributes : HTML Filter의 변환 처리가 가능하며 valid와는 다르게 HTML 태그 내부의 특정 Attributes를 함께 변환한다.
ex) ->
- validAttributesData : validAttributes와 마찬가지로 HTML 태그내부의 Attribute를 함께 이용하여 변환하지만 변환된 결과가 WML 태그가 아니고 Content Data로 사용된다.
ex) -> [IMG] - xxx
- findEndTag : 해당 HTML 태그에 대해서는 Close Tag를 찾아서 TagRole에 해당하는 function을 수행한다.
ex) <script>xxx</script> -> 모두 삭제
<pre>xxx</pre> -> 모두 Content Data로 변환
- discard : 변환 처리가 불가능하여 태그 삭제

계층구조의 Deck을 갖는 HTML Filter 구현 방안

2.2 Parsing Engine

Script Engine은 HTTP를 이용하여 서버로부터 Stream 형태로 읽어 들인 웹 콘텐츠를 적절한 Reformatting 과정을 거쳐 주석 부분과 WML에서 지원하지 않는 문자 Set등을 Filtering한 후 HTML 문서에서 태그부분과 Data 부분을 Parsing하는 기능을 담당한다

2.3 Markup Language Translator

Markup Language Translator는 유선용 콘텐츠와 클라이언트가 지원하는 마크 업 언어가 다를 경우 클라이언트에 맞는 언어로 변환시켜 준다. 즉 WML, HTML, HDML 등의 문서를 Content의 Markup Language를 인식하여 클라이언트 Device 정보와 RuleSet, Script 데이터 베이스를 사용하여 클라이언트 화면에 꼭 맞는 콘텐츠가 생성될 수 있도록 한다.

그림 3 은 HTML Filter 구현에서 Tag RuleSet에 따른 각각의 태그에 대한 변환 처리를 보여주는 것이다. 그림에서 실선은 태그변환의 흐름을 나타내며 점선은 각 HTML 태그에 해당하는 WML 태그에 해당하는 정보를 TagSet Database에서 참조하여 태그 변환을 수행함을 의미한다. 최종 변환결과는 TagType에 따라 적절한 WML 태그, Attributes, Data등이 생성된다.

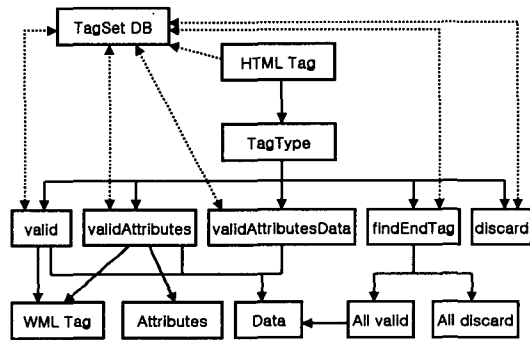


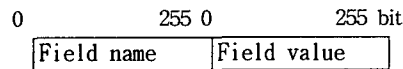
그림3. HTML 태그변환 블록도

III. HTML Filter의 구현

제안하는 HTML Filter 구조에서는 Client가 Web server를 액세스하기 위해 WML Encoded request를 하면 WAP Gateway에서는 관련 Deck이 있는지 여부를 확인한 후 없으면 WEB server에 HTML request 하여 WEB server에서 HTML response을 수신하여

HTML Filter에서 계층구조 Deck을 생성하여 Database에 저장한다. 또한, 저장된 계층구조 Deck 정보는 Gateway가 Encoder를 거쳐 Client로 송신한다. 이러한 Client와 WAP Gateway 사이의 프로토콜은 WSP(Wireless Session Protocol)에서 처리하는 데 계층구조 Deck을 갖는 HTML Filter을 구현하기 위해서는 기존의 WSP(Wireless Session Protocol)의 cache-directive를 이용하여 계층구조 Deck 을 액세스하는 방식을 이용하여 구현할수 있다. 이때 사용되는 cache-directive를 위한 WSP의 헤더 구조는 다음과 같다.

Well-known-header
Application-header
Shift-sequence



Well-known-header = well-known-field-name | wap-value

wap-value = cache-control-value
 Only-if-cached
 last-modified-value
 if-modified-since-value
 if-unmodified-since-value

cache-control-value = No-cacheNo-store
 Only-if-cached

위의 헤더 구조에서 필드명은 cache로 되어 있지만 계층구조 Deck 을 제어하기 위한 새로운 필드를 생성할 필요가 없이 기존에 있는 cache 제어 필드를 이용하여 계층구조 Deck을 제어한다. wap-value값으로 cache-control-value를 이용하여 계층구조 Deck 을 제어할 수 있다. cache-control-value를 No-cache와 Only-if-cached를 이용하여 관련 Deck이 없을 때는 No-cache값을 이용하고 관련 Deck 내용이 있을 때는 Only-if-cached값을 이용한다. 또한 오랜 시간동안 액세스를 하지 않는 사이트는 일정한 시간이 지나면 관련 계층구조 Deck의 내용도 삭제되고 관련 웹사이트가 업 데이트가 되면 관련 계층구조 Deck의 내용도 업 데이트가 되도록 계층구조 Deck 관리도 주기적으로 수행한다. 관련 wap-value는 last-modified-value와 if-modified-since-value, if-unmodified-since-value가 있는 데 이 파라미터를 이용하여 주기적으로 계층구조 Deck의 관리를 할 수 있다[4].

계층구조를 갖는 Deck을 생성하는 방법은 다음과 같다. 최상위 계층 Deck은 content의 차례를 나타내는 Deck으로 인덱스 Deck이다. 보통 웹 페이지에서는 큰 글자는 제목을 나타내므로 index tag는 글자크기를 조절하는 태그이므로 이 tag를 이용하여 계층구조 Deck을 생성할 수 있다. 이 index tag를 이용하는 방식 외에도 다른 계층구조 Deck을 생성하기 위한 방식은 여러 가지가 있을 수 있다. 본 논문에서는 기본적인 index tag를 기반으로 계층구조 Deck을 생성하여 HTML Filter구현을 하였다. 앞으로 계층구조 Deck의 향후 연구는 계층구조 Deck을 생성하기 위한 효율적인 알고리즘에 대한 연구와 저장된 계층구조 Deck의 적절한 관리에 대한 연구가 필요하다

IV. 웹 Contents 변환 결과 및 분석

계층구조 Deck을 갖는 HTML Filter의 기능 Test를 위해 실제 HTML로 작성된 웹 콘텐츠에 대한 Test를 수행하고 각 수행 결과에 대한 분석을 통해 향후 WAP Gateway 설계 시 HTML Filter 기능에 대한 연구 방향을 모색한다. Test에서 사용된 웹 브라우저는 M3Gate에서 만든 WAP Emulator를 사용하였고 Test 웹 콘텐츠는 Yahoo Korea! 검색 사이트를 HTML Filter를 통하여 Filtering해 보았다. 그림 4. 를 보면 웹 브라우저의 왼쪽 화면은 기존 콘텐츠를 보여주고 있고 오른쪽은 웹 콘텐츠를 WML로 변환한 소스코드를 보여주고 있다. 또한, M3Gate WAP 브라우저는 웹 브라우저에 보여지는 웹 콘텐츠에 해당하는 HTML 문서가 WML로 변환되어 브라우징 되는 것을 보여주고 있다.

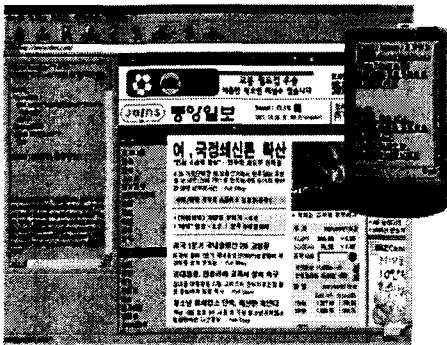


그림4. HTML Filter test

변환 과정에서 HTML 문서의 그림 파일은 단순히 [IMG]라는 텍스트로 변환이 되었다. 무선 단말기가 이미지 표현이 가능하다면 본 논문에서 구현한 HTML Filter에 WBMP 변환기를 함께 이용하여 WBMP 이미지로의 변환이 가능하다. 현재 인터넷에서 볼 수 있는

수많은 웹사이트의 대부분이 HTML을 사용해서 작성한 문서들이다. 이러한 HTML 문서를 WML을 지원하는 이동 단말기의 브라우저에 표시하기 위해 Dynamic한 HTML Filter 구현은 반드시 필요하다. 본 논문을 기반으로 한 HTML Filter의 구현에 대한 향후 연구 방향은 다음과 같다.

- 무선망에서의 Signal overhead를 최소화하는 방안
- 유선망에서의 Refresh 따른 대역폭 사용량을 최소화하는 방안
- 효율적인 계층구조의 Deck을 갖는 Database 생성 방안

V. 결론

본 논문에서는 무선이동 단말기를 통하여 웹 액세스를 가능하게 하는 무선 응용 프로토콜인 WAP에서 가장 중요하게 다루어지는 WAP Gateway의 HTML Filter 기능에서 복잡한 HTML 문서를 계층구조의 Deck을 생성하여 WML로 변환하는 방안을 제시하였다. 또한 Client와 WAP Gateway 사이의 관련 프로토콜에 대해서도 언급하였다. 이러한 계층구조를 갖는 Deck은 웹 Contents를 짧은 메시지로 구현하므로 무선 WAP 환경에서는 대역폭의 절감과 WAP 단말기에서 Contents 디스플레이를 효율적으로 할 수 있다. 그러나, 계층구조 Deck으로 인한 짧은 메시지를 여러 번 액세스하므로 빈번한 Client와 Gateway 사이의 Refresh request/response 발생으로 signaling의 overhead가 증가할 수 있는 반면, 기존 HTML Filter에서의 액세스 지연시간 보다는 효율적이다.

참고문헌

- [1] WAP Forum, <http://www.WAPforum.org>
- [2] Wireless Application Protocol: Wireless Application Environment Overview , WAP Forum, 04-Nov-1999
- [3] Wireless Markup Language Specification Version 1.2 , WMLSPEC1.2
- [4] Wireless Application Protocol Tutorial, Apion Telecommunications, viewed 05-Jun-2000
- [5] Extensible Markup Language (XML) 1.0 W3C Recommendation, W3C, 10-Feb-1998
- [6] Marcin Metter, Dr Robert Colomb, Wap enabling existing HTML applications, User Interface Conference 2000, Page(s): 49 57, 2000
- [7] 이승진, 김대건, 최린, 강철희, "확장성 있는 웹 서비스를 위한 무선 응용 프로토콜 기반의 HTML Filter 구현," 한국정보과학회, 2001년도 한국 정보과학회 학술발표 논문집 Vol. 28., No.1 p391-393