

이동 통신 환경에서 효율적인 웹 문서 변환에 관한 연구

°강성천, 정광수

광운대학교 전자공학부 컴퓨터통신연구소

The Study of Efficient WWW Contents Conversion in a Mobile Environment

°Sunchun Kang, Kwangsue Chung

Computer Communications Lab., Kwangwoon Univ.

요 약

최근 인터넷과 이동 통신의 발달로 다양한 이동 통신 단말기 환경에서도 인터넷에 접근하는 것이 가능하게 되었다. WWW의 표준인 HTML로 작성된 많은 웹 문서들을 그대로 이동 통신 환경에 적용시키는데는 여러 가지 문제점이 있다. 본 논문에서는 그에 따른 문제점을 지적하고 이동 통신 환경에 기존의 HTML기반의 웹 문서를 변환하여 새로운 환경에 적용시키는 효율적인 문서 변환 방법을 제시하며, 또한 이동 통신 단말기를 위한 HTML의 부분 집합을 제안한다.

1. 서론

최근 WWW(World Wide Web)의 발전과 함께 이동 통신 망의 발전으로 이동 통신 환경에서도 인터넷상의 웹 문서에 접근하는 것이 가능해 졌다. 하지만 기존의 웹 환경을 이동 통신 환경에 그대로 적용시키는 것은 처리 능력이 부족한 단말기나 좁은 대역폭을 가지는 이동 통신 망에 적합하지 않다. 따라서 기존의 웹 문서를 이동 통신 환경에 적용하기 위해서는 적절한 처리를 통해 사용자의 요구와 망 상황에 적합한 문서를 제공 할 필요가 있다.

한편 사용자에게 만족할 만한 이동 통신 서비스를 제공하기 위해 HDML(Handheld Device Markup Language), WML(Wireless Markup Language), Compact HTML등의 마크 업 언어(markup language)가 개발되었다. 하지만 HDML과 WML은 기존의 웹 문서들과 호환성이 없으며, HTML과 호환성을 가지는 Compact HTML 역시 기존의 웹 환경을 적용시키기 위해서는 문서의 변환을 필요로 한다. 또한 일반적으로 문서를 처리하는 방법에 있어 문서 내의 태그(tag)관계로 직접적인 문서를 처리하는 방법은 여러 가지의 문제점을 가지며, HTML 문서 내의 마크 업 오류들은 문서를 다루는 것을 어렵게 한다.

본 논문에서는 이러한 문제점을 개선시켜 문서를 효율적으로 처리 하여 이동 통신 환경에 적합한 문서를 제공하는 방법과 이를 위한 HTML의 부분 집합을 제안한다.

2. 이동 통신 환경에서의 웹 접근 시 요구 사항

이동 통신 환경은 기존의 웹 환경과 다음과 같은 차이가 있다.

이동 통신 단말기 :

- 제한된 메모리와 전원을 가진다.
- 소형의 화면을 가진다.
- 처리 능력이 낮다.
- 제한적인 사용자 인터페이스와 입력 방법을 가진다.

이동 통신 망 :

- non-IP 기반이다.
- 느린 속도를 가지며 비용이 비싸다.
- 좁은 대역폭을 가지며 신뢰성이 낮다.

이러한 점들은 데스크탑 PC와 워크스테이션 기준으로 제작된 웹 문서들을 이동 통신 환경에 그대로 적용 할 수 없음을 의미한다. 웹 문서는 이미지 및 다양한 기능을 위한 부가적인 데이터를 포함하기 때문에 전송 해야할 파일의 크기가 크다. 따라서 이동 통신 망에 인터넷상의 웹 문서를 그대로 전송하는 것은 전송지연 문제를 가진다. 만일 이러한 문서를 이동 통신 망에 그대로 전송하기 위해 대역폭을 늘린다면 단말기에 소비 전력을 증가 시켜 전원의 수명을 단축시키게 된다. 따라서 웹 문서를 이동 통신 망에서 효과적으로 전송해주기 위

해서는 적절한 문서의 처리가 필요하다. 문서의 처리는 게이트웨이(gateway)를 망에 설치하여 이동 통신망에 적합한 문서 구조로 변환할 수 있다. 그림 1은 일반적인 게이트웨이의 구조를 나타낸 것이다.

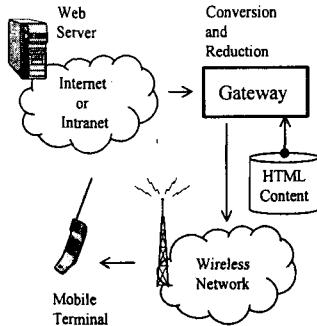


그림 1. 문서 처리를 위한 게이트웨이 구조

3. 이동 통신 환경에서의 웹 문서 변환

본 논문에서의 문서 변환 과정은 마크 업 오류 제거 과정, 태그 단순화 과정, 표준 문서 변환의 3가지 과정을 거치게 된다. 그림 2는 문서의 변환 단계를 나타내고 있다.

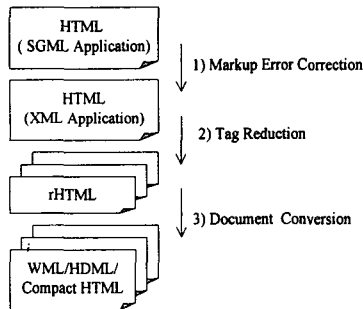


그림 2. 문서 변환 과정

3.1 마크 업 오류의 제거

인터넷의 웹 문서들은 웹 문서 개발자들의 마크 업 규칙에 대한 실수나 무시로 수많은 마크 업 오류를 포함하고 있다. 이것은 문서의 명확한 처리를 어렵게 만든다. 마크 업 오류를 가진 문서를 단순화했을 때 변환 된 문서 또한 마크 업 오류를 포함하게 된다. 이것은 변환 된 최종 문서가 변환 전 문서와는 전혀 다른 구조나 내용을 가지는 문서가 될 수 있다는 것을 의미한다. 이와 같은 문제를 해결하기 위해 HTML문서를 XML(Extensible Markup Language)의 잘 구성된(well-formed) 문서로 변환한다. 변환 규칙은 다음과 같다.

- 1) 잘 구성된(Well-formed) 문서이어야 한다.
- 2) 태그 요소와 속성은 소문자 이어야 한다.
- 3) 시작 태그와 종료 태그가 항상 대응되어야 한다.

- 4) 속성 값은 항상 인용 부호로 묶여야 한다.
- 5) 속성 값을 생략 할 수 없다.
- 6) 빈 태그 요소는 />로 끝나야 한다.
- 7) 속성 값 내의 공백 문자의 연속은 하나의 빈 문자로 취급된다.
- 8) 몇 가지 규정된 요소들은 중첩 될 수 없다.

변환 규칙으로부터 실제 마크 업 오류를 수정하는 방법을 예를 들어 설명한다. 아래는 종료태그가 올바르게 중첩되지 않은 경우를 수정하는 예이다.

```
<p>here is a para<b>bold<i>bold italic</b>bold?</i>
```

이것은 다음과 같이 수정되어야 한다.

```
<p>here is a para<b>bold<i>bold italic</i>bold?</b></p>
```

만일 이것이 이러한 수정 없이 변환되었다면 변환 된 문서 또한 올바르게 중첩된 태그가 포함된 코드를 계속 적으로 다루게 되고, 계속적으로 마크 업 오류를 포함하게되어 최종 문서는 의도하지 않았던 결과로 보여지게 된다.

이 과정으로부터 SGML(Standard Generalized Markup Language)의 응용인 HTML은 XML 응용인 XHTML(Extensible HTML) 문서가 된다. 이로부터 가장 기본적인 마크 업 규칙에 관한 오류가 제거되며, 또 다른 XML 응용으로의 변환을 용이하게 한다.

3.2 문서의 단순화

오류 제거 이후의 단계로 마크 업 오류가 제거 된 문서로부터 태그를 단순화하는 처리를 한다. 이 과정을 통해 이동 통신 망에 전송해야 할 데이터의 양을 감소시키고 낮은 처리능력을 가지는 단말기에서 처리해야 할 데이터의 양을 감소시키게 된다.

DTD	역 할
FullSet	망이 안정적이며 모든 이동 통신 서비스를 처리 할 수 있는 경우
Narrow Bandwidth	망 상황이 불안정하며 모든 이동 통신 서비스를 제공하는 것이 불가능한 경우
Text Message	망의 대역폭이 작거나 불안정하여 신뢰도가 떨어지는 경우, 혹은 사용자 단말기가 문자 수준의 서비스밖에 제공 받을 수 없는 경우

표 1. rHTML의 DTD 집합

단순화는 본 논문에 정의한 3가지의 DTD(Document Type Definition)의 집합을 참조해 이루어진다. 본 논문에서는 이 세 가지 DTD의 집합으로 정의 된 마크 업 언어를 rHTML(Reduced HTML)이라 통칭한다. rHTML은 XML의 응용이며 HTML 2.0,

HTML 3.2 과 HTML 4.0의 부분 집합이다. 표 1은 rHTML의 3가지 DTD를 나열하고 그 역할을 정리한 것이다.

rHTML은 레이아웃(layout) 구조에 영향을 주거나 단말기 측에 큰 처리능력을 요구하는 태그 요소와 속성들을 우선적으로 단순화하여 재구성하였다. 표 2는 HTML 2.0, 3.2, 4.0으로부터 우선적으로 단순화시킨 태그 요소 및 속성의 일부를 나열한 것이다.

처 리	태 그	속 성
테이블	<TABLE> <TD> <TR> <TH> <THEAD> <TFOOT>....	-
이미 맵	<AREA> <MAP>	usemap, ismap
폰트 및 스타일	 <BASEFONT> <SUB><SUP> <BIG> <STYLE>	text, class, style....
배경	-	bgcolor, background....
프레임	<FRAME> <FRAMESET> <NOFRAMES> <IFRAME>	target
기 타	<APPLET> <OBJECT> <SCRIPT> <BLOCKQUOTE> <ADDRESS>	rel, rev, title, urn, methods....

표 2. rHTML에서 기본적으로 단순화 된 요소

하나의 웹 문서는 사용자의 요구사항이나 망 상황에 따라 rHTML의 세 가지 DTD중 한 가지를 참고해 단순화되고, 이로부터 단일한 웹 문서에 대해 다른 수준의 서비스를 제공할 수 있는 3가지의 문서가 생성된다.

3.3 표준 문서로의 변환

문서의 단순화 과정을 통해 생성 된 문서는 또 다시 WML이나 HDML, Compact HTML같은 표준 문서로 변환 될 수 있다. 결국 하나의 문서로부터 자기 다른 서비스 수준을 가지는 문서들을 생성할 수 있게 된다. 하지만 이 과정 중 표준 문서로 변환하지 않고도 rHTML 문서를 직접 서비스하는 것이 가능하다. 이것은 이동 통신 단말기의 사용자 에이전트의 기능과 서비스에 의존한다. 또한 rHTML은 Compact HTML처럼 HTML과 호환성을 가지므로 HTML을 지원하는 환경에서도 그대로 사용 할 수 있다는 장점을 가진다.

문서의 레이아웃 구조에 있어 rHTML은 문서의 화면 구조와 외양을 정의하는 요소들을 제거했기 때문에 문서를 변환하는 과정에서는 레이아웃에 대해 고려할 필요가 없다. 따라서 문서의 실제 레이아웃의 렌더링(rendering)은 사용자 단말기의 인터페이스 구성 요소들에 의존한다. rHTML의 태그 요소들은 브라우저의 사용자 인터페이스 구성요소를 사용하는 렌더링 모듈과 대응관계를 가진다. 그림 3은 자기 다른 표준 문서상의 태그 요소들이 동일한 렌더링 구성 요소에 대응되는 한 예를 보이고 있다. 이러한 구조는 문서 개발자가 레이아웃 구조를 고려하지 않고도 문서를 작성하는 것을 가능하게 한다. 또한, rHTML의 3가지 DTD 집합 중 어느 한 종류의 DTD를 참조해 변환된 rHTML문서는 다시 여러 가지의 표준 문서들로 변환이 되며,

이 문서들은 자기 다른 표준 문서임에도 불구하고 단일한 단말기에서 동일한 레이아웃 구조를 가지게 된다. 그리고 렌더링을 다루는 구성 요소가 모듈화 되어 있기 때문에 새로운 태그를 다루는데 있어 확장성을 가진다. 그림 4는 이러한 브라우저의 구조를 나타낸 것이다.

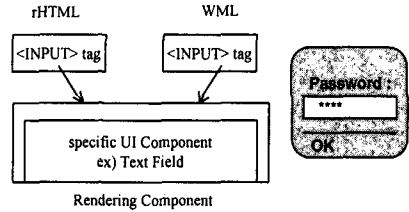


그림 3. 태그와 렌더링 구성 요소의 대응 관계

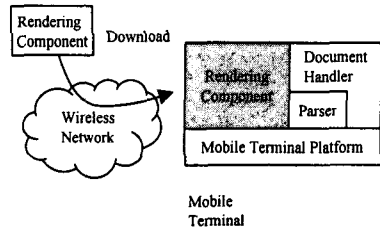


그림 4. 확장 가능한 브라우저 구조

4. 결론 및 향후 과제

본 논문에서는 이동 통신 환경에 기존의 웹 환경을 적용하기 위해 효율적으로 HTML문서를 처리하는 방법과 이를 위한 HTML의 부분 집합인 rHTML을 대해 제안하였다. 이를 통해 이동 통신 환경에 적합한 문서를 제공할 수 있다.

향후 과제로는 HTML 문서의 처리 외에도 문서와 함께 다루어지는 데이터에 대한 처리 방법과 지능형 에이전트들을 통해 망과 사용자에 적합한 문서 처리를 하도록 하는 방법이 필요하다.

참 고 문 헌

- [1] Tomihisa Kamada, Takuya Asada, Masayasu Ishikawa, Shin'ichi Matsui, "HTML 4.0 Guidelines for Mobile Access", W3C Note, 15 March 1999.
- [2] Tomihisa Kamada, "Compact HTML for Small Information Appliances", W3C Note, 9 February 1998.
- [3] Dave Raggett, Arnaud Le Hors, Ian Jacobs, "HTML 4.0 Specification", W3C Recommendation, 18 December 1997, revised 24 April 1998.
- [4] Johan Hjelm, Bruce Martin, Peter King, "WAP Forum - W3C Cooperation White Paper", W3C Note, 30 October 1998.
- [5] "WAP Forum", <http://www.wapforum.org/>
- [6] "W3C Mobile Access", <http://www.w3.org/Mobile/>