

실시간 웹 게시판 모니터링 및 모바일웹을 이용한 알람 서비스 개발

김종근*, 심근호**, 이요셉***, 임영환****

요약

웹사이트의 게시판을 각종 정보를 올리고 공유할 수 있는 정보소통의 장이다. 지금 이 순간에도 수많은 정보들이 각종 웹사이트의 게시판을 통해 공유되고 있다. 웹 게시판에 등록되어 공유된 정보는 누군가에게는 득이 될 수 있고 또 누군가에게는 실이 될 수 있다. 이러한 이유로 웹 게시판에 등록되는 정보에 대하여 실시간으로 모니터링하고 그 결과를 알려주는 기술에 대한 연구가 오래 전부터 있어 왔다. 기존의 방법들은 DB에 직접 접근하는 방법과 오픈API를 이용하는 방법으로 게시판을 모니터링 하였다. 또한 이를 사용자에게 알려주기 위해 이메일을 이용하는 방법을 사용하였다. 이러한 방법에는 두 가지 큰 문제가 있다. 오픈되지 않은 게시판에 접근할 수 없는 문제와 사용자에게 실시간으로 알려주기 어렵다는 문제이다. 특히나 모니터링과 같은 서비스는 실시간성이 매우 중요하다. 왜냐하면 모니터링의 목적은 그에 대한 대응을 기본적으로 하고 있기 때문이다. 모바일 웹은 언제 어디서나 접속할 수 있다는 장점으로 인하여 실시간적인 요소가 필요한 서비스에 적합하다. 하지만 휴대폰 단말기에 따라 콘텐츠의 규격이 다르다는 단점으로 인하여 콘텐츠를 미리 생성해야 하기 때문에 이러한 실시간적인 서비스에 잘 이용하지 못하였다. 본 논문은 비 규격화된 모바일 웹의 단점을 극복하는 기술을 제시함으로써 모니터링과 같은 실시간성이 필요한 서비스에 모바일웹을 이용하여 그 효과를 극대화하는 모습을 보여주고자 한다.

Development of real-time monitoring web BBS and the alerts service using mobile web

Jongkeun Kim*, Kunho Sim**, Yoseb Lee***, Younghwan Lim****

Abstract

The BBS of Web Site is a space of informational communication to share information. At the moment, countless information is shared through various kinds of BBS. The shared information through the Web BBS can be gains to someone and losses to the other. Thus, for a long time, there have been many researches about technologies, which monitor information uploaded to Web BBS in real-time and inform the results to subjects. For BBS monitoring, existing methods have utilized direct approaches to DB or open API. Also, e-mail addresses have been utilized to inform results to subjects. However, there are two main problems for these methods. The one is a problem on closed BBS, which cannot be accessed, and the other is impossibility of supporting real-time alarm to e-mail. Especially, the real-time support for services, such as monitoring, is extremely important, since the main purpose of monitoring is based on the counter action for the result. The mobile Web has a merit to be accessed in anywhere, so that it is quite suitable for services requiring real-time support. However, the real-time support of mobile device has not been utilized enough, because each mobile phone has different standards of contents so that the contents should be pre-created. This paper suggests a technology that overcomes pitfalls of mobile Web while it is out of standard, and shows how the effect can be maximized through services such as monitoring, which requires the real-time support.

Keywords : web BBS, monitoring, mobile web, notify, alerts service

※ 제일저자(First Author) : 김종근
접수일:2011년 12월 22일, 수정일:2012년 01월 31일
완료일:2012년 02월 21일
* 송실대학교 대학원 미디어학과
jongni@ssu.ac.kr

** 송실대학교 대학원 미디어학과
*** 코원시스템(주)
**** 송실대학교 글로벌미디어학과(교신저자)

1. 서론

1.1 연구 배경 및 관련 연구

인터넷의 보급화가 급속도로 이루어지면서 서로간의 정보를 공유하기 위한 웹페이지가 활성화 되었다. 이러한 활성화를 통해 웹페이지의 제작이 확산 되었고 이를 통한 게시판의 의존도가 높아졌다. 게시판은 웹페이지에 없어서는 안 될 콘텐츠로 존재 하게 되었고 게시판을 통해서 각종 정보를 올리고 공유함으로써 사용자들은 게시판을 활용하게 되었다. 이러한 현상은 자동화로 연결이 되었다. 사용자들은 자신의 원하는 게시판의 정보를 알람해 줄 수 있는 시스템이 없는지 연구하기 시작하였다[1]. 이 연구는 두 가지 측면의 연구가 필요한데 한 가지는 게시판의 정보를 수집하여 모니터링하기 위한 연구이고 또 한 가지는 모니터링 된 결과를 원하는 사용자에게 알람하기 위한 연구이다.

먼저 게시판의 정보를 수집하여 모니터링 하기 위한 기존의 연구는 크게 두 가지 방법으로 진행 되었다. 초기 연구 방식은 데이터베이스(DB)접근방식이다. 이 방법은 사이트의 개발자가 DB를 오픈함으로써 DB의 오픈 권한을 가진 모듈이 DB에 직접 접근하여 정보를 모니터링하는 방식이다. 또 하나의 연구 방식은 오픈 API(Application Programming Interface)나 RSS와 같은 규격화된 정보공유를 통한 방법이다. 게시물의 정보에 접근할 수 있는 인터페이스 규격을 오픈하여 정보를 원하는 사용자가 열람할 수 있도록 제공하는 방식이다[2].

첫 번째 방식을 사용하기 위해서는 DB에 대한 권한이 필요하며 두 번째 방식을 사용하기 위해서는 오픈 API나 RSS를 지원하는 사이트 이어야만 한다. 결국 모니터링 할 수 있는 게시판에 제한이 생기게 된다.

이를 해결하기 위해 HTML을 파싱하여 정보를 모니터링 하고자 하는 기존 연구도 있었다[3]. 하지만 이러한 방식도 문제점이 존재하는데 웹 게시판들이 다양한 레이아웃을 가지고 있기 때문에 시스템이 HTML을 파싱하였을 때 게시판 영역을 자동으로 추출하기 어렵다는 것이다. 이 때문에 사용자가 직접 게시판의 정보를 넣어 줘야만 했다.

정보수집 및 모니터링 결과에 대한 알람 시스템은 SMS서비스나 e-mail을 이용하는 사례가 많다[4]. SMS는 사용자에게 즉각적으로 알람을 해줄 수 있지만 구체적인 내용을 확인하기 위해서는 웹으로 확인해야 한다. 또한 e-mail은 사용자가 e-mail을 확인하기 전까지는 알람 내용을 알 수 없어 즉각적으로 알려줄 수 없는 문제가 있다.

1.2 문제점 및 해결방안

본 논문은 앞의 기존연구들에 대한 분석을 바탕으로 웹 사이트 내의 특정 게시판 게시물에 대한 모니터링 및 알람을 위한 문제점을 3가지 측면에서 제시하고 이를 해결하고자 한다.

첫째는 웹 페이지 내에서 게시판 영역의 추출이다. 웹 사이트 내에 사용자가 원하는 키워드가 발견되었다고 할 때 이 키워드가 게시글에 해당하는 내용인지 알기 위해서는 웹 페이지 내에서 게시판 영역을 알 수 있어야만 한다. 예를 들어 (그림 1)은 특정 페이지의 게시물에서 “신제품 인증”이라는 키워드를 찾고자 할 때 그 키워드가 게시판 뿐 아니라 페이지의 다른 위치에서도 존재할 수 있음을 보여준다.



(그림 1) 게시판 영역 추출의 필요성

이렇듯 문서 내에서 실제 게시판인 테이블을 찾아 낼 필요가 있으며, 테이블 식별에 대한 연구는 여러 방향으로 진행되어 왔다[5-8]. 그 중 본 논문에서는 테이블의 속성을 이용해 게시판 영역을 추출하는 테이블 식별 알고리즘을 사용하고자 한다. 웹 사이트를 개발하기 위해서는 테이블의 사용이 많은데 특별히 게시판의 레이아웃을 구성하기 위해서는 테이블 요소를 주로 사용하게 되고 여기에는 여러 가지 속성이 들어

갈 수 있다. 테이블 식별 알고리즘은 이러한 테이블 내의 속성을 판단해서 페이지 내의 게시판 영역을 판별한다[9].

두 번째는 게시판 내용의 실시간 모니터링이다. 해당 게시판에 사용자가 원하는 키워드와 관련된 내용이 등록 되는지 실시간으로 모니터링 해야 한다. 이 때, 단순히 키워드를 검색하는 것으로는 부족하다. 해당 키워드 관련 내용이 이미 게시판에 존재 할 수 있기 때문이다. 따라서 모니터링 기법은 키워드 검색과 함께 게시물의 변화를 판단할 수 있어야 한다.

이를 위해 본 논문에서는 키워드 기반 실시간 검색 프로그램을 설계하였다. 사용자가 지정한 페이지 URL에 대하여 주기적으로 페이지를 파싱하고 내용을 모니터링하는 프로그램이다. 이에 대하여 2장에서 자세히 설명할 것이다.

마지막 세 번째는 모니터링을 통해 사용자가 원하는 키워드의 내용이 게시판에 등록되었을 때 이를 실시간으로 사용자에게 알려주는 알람 방법이다. 관련 연구에서 얘기한 것처럼 SMS를 통해 내용을 알려주는 것은 즉시성은 보장되지만 해당 내용을 자세하게 전달할 수 없다. 또한 E-mail은 자세한 내용을 전달 할 수 있지만 즉시성이 보장되지 않는다. 이에 대해 본 논문은 해당 내용을 자동으로 모바일 페이지로 변환하여 SMS에 연결하여 보내는 모바일 페이지 변환 및 Callback SMS 기법을 제안한다. 이를 3장에서 자세히 설명할 것이다.

2. 키워드 기반 실시간 검색 기법

앞장에서 언급한 바와 같이 웹 페이지 내에 테이블을 식별한 후에 테이블 내의 정보에 대하여 키워드 검색을 실시한다. 이를 위해서 크게 2가지 단계를 거친다. 첫 번째는 식별된 테이블 엘리먼트의 DOM-Tree를 분석하여 최 하단 셀에 접근하여 내부 html을 리스트화 한다. 두 번째는 내부html에서 텍스트 정보와 alt 속성 등 정보로 판단될 수 있는 문자열을 정리한 후 키워드 기반 검색을 실시한다.

여기에서 중요한 것은 실시간 검색을 위해서는 키워드 검색이 주기적으로 실행되어야 한다는 것이다. 주기적인 검색 실행과 페이지 내의

콘텐츠 변화를 확인하여 특정시간에 웹 페이지에서 사용자가 지정한 키워드의 게시글이 등록되었는지 확인 할 수 있다. 검색 주기의 설정은 간격 및 회수로 지정될 수 있으며 정확한 검색 결과를 얻기 위해서는 이전의 키워드 검색 결과를 저장하고 있어야 한다.

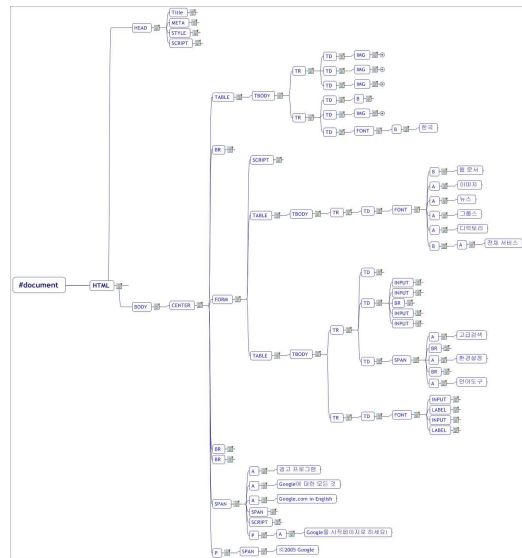
2.1 DOM-Tree 분석을 통한 게시물 정보 추출

2.1.1 웹 페이지와 DOM

PC 웹 브라우저 상에 브라우징 된 PC 웹 페이지(HTML Document)는 DOM 으로 표현 할 수 있다. DOM(Document Object Model)이란 함은 문서 객체 모델이란 의미로서 XML 문서를 트리구조 표현으로 명시하는 것이다.



(그림 2) 구글 웹 페이지



(그림 3) 구글 웹 페이지의 DOM 구조

이 트리에서 최상위 문서 인스턴스는 루트이며, 하나의 최상위 요소 인스턴스를 갖는다. 이 요소는 내용을 표현하는 자식 노드나, 추가적인 자식을 가질 수 있는 하위 요소를 갖는다. 함수들은 결과 트리를 탐색하거나, 요소와 속성 값으로 접근하거나, 노드의 추가와 삭제 및 트리를 다시 XML로 변환하는 등의 작업을 제공한다.

PC에서 브라우징 된 DHTML 문서는 XML의 규칙을 따르기 때문에 이를 DOM으로 표현할 수 있다. 예를 들어 (그림 2)와 같은 구글 사이트는 (그림 3)과 같은 DOM구조를 가지게 된다.

2.1.2 DOM-Tree를 통한 게시물 내용 분석

앞에서 테이블 영역을 확인하기 위한 방법으로 테이블 엘리먼트와 그에 대한 속성을 사용한다고 하였다. 이렇게 테이블 영역이 확인되면 다음으로는 테이블 내에 있는 게시물의 내용을 확인하여야 한다.

테이블 내에 존재하는 여러 내용 중에 게시물 목록을 판단하기 위하여 본 논문에서는 DOM-Tree를 이용하고자 한다. (그림 3)에서 보는 것처럼 DOM-Tree로 분석된 웹 페이지는 각 엘리먼트 하위에 어떤 요소가 있는지 계층적인 구조로 볼 수 있다.

게시물 내용의 DOM-Tree에서의 특징을 2가지로 살펴볼 수 있는데, 첫 번째는 같은 구조가 반복된다는 것이다. TR아래의 TD요소가 같은 구조와 속성을 가지고 있는 것이다. 두 번째 특징은 게시물 내용은 최하위 레벨에 존재한다는 것이다. 실제 우리가 필요한 것은 테이블 내의 텍스트 정보이며 이는 DOM-Tree의 최하위 레벨에 존재한다.

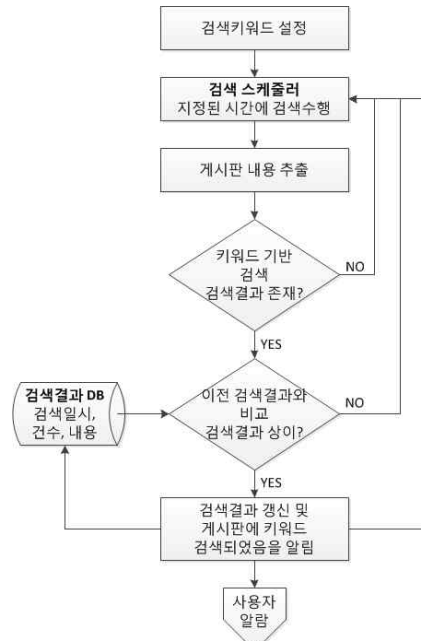
결론적으로 테이블 영역의 DOM-Tree를 분석하여 반복되는 구조 속의 최하위 레벨의 텍스트를 모으게 되면 해당 게시판의 게시물 내용을 하나의 리스트로 구성 할 수 있게 된다.

2.2 키워드 기반 검색과 게시물 갱신여부 판단

앞 절의 과정을 통해 게시물의 내용을 리스트로 구성한 후 (그림 4)와 같은 검색 흐름으로

게시물 내용을 가지고 키워드 검색을 실시한다.

리스트 내에서 사용자가 입력한 키워드가 존재한다면 그 결과를 저장한다. 사용자에게 중요한 것은 이 게시판에 사용자가 입력한 키워드를 포함하는 게시물이 있는지의 여부이기 때문에 키워드 검색 결과가 있다는 것 자체가 중요한 요소가 된다.



(그림 4) 키워드 기반 실시간 검색 흐름도

하지만 이와 더불어 또 한 가지 중요한 것이 있는데 그것은 앞서서도 얘기한 바와 같이 게시물의 변화이다. 똑같은 게시물에 대하여 사용자에게 반복적으로 알려줄 필요가 없기 때문이다. 이에 따라 게시판의 내용 변화를 감지할 수 있도록 결과를 저장함에 있어서 2가지 중요한 요소를 함께 저장한다. 한 가지는 키워드가 검색된 결과 건수이며 또 한 가지는 키워드가 검색된 html 요소의 전체 텍스트 내용이다.

키워드가 검색된 건수가 변화되었다는 것은 게시물의 변화가 거의 확실하다. 특히나 건수가 증가한 경우에는 100% 새로운 게시물이 발견된 것으로 판단 할 수 있다. 건수가 변화하지 않아도 게시물의 변화가 있을 수 있는 경우는 키워드를 포함한 기존 게시물이 삭제되고 새로운 게시물이 올라왔을 경우이다. 이를 위해 키워드검

색 건수의 변화가 없더라도 키워드가 검색된 html 요소를 비교하여 이전 결과와 차이가 있는지를 확인 할 필요가 있는 것이다.

이렇게 키워드가 포함된 게시물의 등록이 검색이 된다면 검색결과와 게시물의 내용을 사용자에게 알려준다.

3. 모바일페이지 변환 및 Callback SMS

3.1 웹 게시판의 모바일페이지 생성

2장의 과정을 통해 사용자가 원하는 웹사이트에 원하는 키워드로 정보가 게시되었다고 판단되었다면 이를 사용자에게 알려주어야 한다. 서론에서도 얘기했지만 사용자에게 알려주는 방법은 여러 가지가 존재한다. 본 논문에서는 모바일 웹과 Callback SMS를 통해 즉시성과 자세한 내용을 보여주는 2가지 모두를 만족시키고자 한다. 이를 위해서는 먼저 웹에 게시된 게시물을 모바일 웹으로 변환하여야 한다. 휴대전화는 일반 PC보다는 작은 화면을 가지고 있기 때문에 PC웹을 휴대폰으로 보여주기 위해서는 레이아웃의 변환이 반드시 필요하다. 또한 휴대전화는 통신사와 단말기에 따라 다른 마크업 언어를 사용하기 때문에 각 단말기에 맞도록 콘텐츠를 변환하는 것도 필요하다.

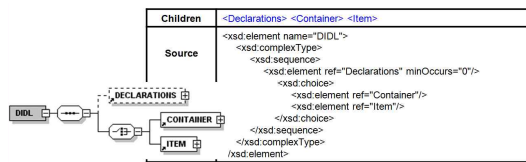
이를 위해 본 연구에서는 PC웹에서의 단위 콘텐츠를 DI(Digital Item)으로 정의하고 이를 표현하는 DIDL을 이용하여 PC웹의 게시판 콘텐츠를 모바일 웹 페이지로 변환하였다.

3.1.1 DIDL

DIDL(Digital Item Declaration Language)을 이해하기 위해서는 먼저 MPEG-21과 DI (Digital Item), DID(Digital Item Declaration)을 알아야 하는데, MPEG-21은 광범위한 네트워크와 장비에서 멀티미디어 자원을 투명하고 안전하며 다양하게 제공함으로써 멀티미디어의 효과적인 전달과 소비가 가능하도록 하는 멀티미디어 프레임워크이다. Digital Item(DI)은 MPEG-21의 Part2에 기술방법이 명시되어 있으며, 멀티미디어 콘텐츠를 생성, 공급, 거래, 인증, 소유, 관리, 보호, 사용과 관련된 사용자들 간에 전달되는

과정에서 디지털 형태로 표현되어 네트워크나 터미널에서 이용되는 멀티미디어 콘텐츠의 기본 단위를 말한다. DID는 멀티미디어 리소스(비디오, 오디오, 이미지, 텍스트, 그래픽, eBook 등)인 CDI(Content Digital Item)와 리소스에 대한 정보(저작권, 콘텐츠 ID, 종류 등)를 기술한 메타데이터인 XDI로 구성되어 있다.

DIDL(Digital Item Declaration Language)은 XML 형식을 따르는 언어로서 DID를 XML 형태로 작성한 언어이다. DIDL은 기본적인 DI의 정보를 표현할 뿐 아니라 필요한 속성을 추가로 정의할 수 있게 설계되어 있으며 기본적인 구성요소는 Container, Item, Component, Anchor, Descriptor, Choice 등이 있으며 루트 요소는 (그림 5)와 같다.[11]



(그림 5) DIDL의 루트요소

3.1.2 PC웹 게시판의 DIDL 제작

PC웹 게시판 내에는 여러 개의 이미지와 텍스트뿐 아니라 여러 종류의 오브젝트들이 존재한다. 각각의 오브젝트는 HTML 요소내의 속성으로 추출되어지며 각각 DI로 정의할 수 있다. 또한 DI로 정의된 정보는 DIDL로 표현할 수 있다. 즉 텍스트, 이미지 등과 같은 오브젝트들이 DIDL의 아이템 단위로 존재하는 것이다. 단위 오브젝트로 정의된 DI은 다양한 속성을 가질 수 있으며 이러한 속성을 이용하여 단말 화면에 적합하도록 레이아웃을 변형하는 것이 용이해진다. <표 1>은 모바일웹 페이지의 콘텐츠를 정의하고 있는 DIDL페이지의 전체적인 모습이다. 모든 아이탬은 유일한 아이디를 가지고 구별되어 있으며, <표 2>는 아이탬이 가지고 있는 공통속성을 보여준다. <표 1><표 2>와 같이 DIDL은 페이지가 가지고 있는 아이탬과 아이탬이 가지고 있는 속성을 기술함으로써 모바일 페이지가 어떠한 내용을 가지고 어떻게 구성되어 있는지 표현하고 있다. 앞 장에서 설명한 키워드 기반 실시간 검색 알고리즘을 통하여

추출된 게시판의 내용은 자동으로 DIDL파일로 생성된다.

<표 1> 모바일 UCC 폰피 DIDL페이지 구조

```

<DIDL xmlns:mpeg21="urn:mpeg:mpeg21:2002:01-didl-NS" xmlns:mbd =
"http://medialab.ssu.ac.kr/ObjectAttribute">
  <CONTAINER ID="Page">
    <DESCRIPTOR>
      <STATEMENT Type="text/xml">
        <mbd:PAGEID>index</mbd:PAGEID>
        <mbd:TITLE></mbd:TITLE>
        <mbd:BACKIMAGE></mbd:BACKIMAGE>
      </STATEMENT>
    </DESCRIPTOR>
    <ITEM ID="0">
      <DESCRIPTOR>
        <STATEMENT Type="text/xml">
          <mbd:OBJECTTYPE>IMAGE</mbd:OBJECTTYPE>
          <mbd:SUBTYPE>JPG</mbd:SUBTYPE>
          <mbd:RESOURCE>./IMAGE/15_f.jpg</mbd:RESOURCE>
          <mbd:POSITION>0,20,240,24</mbd:POSITION>
          <mbd:Z-INDEX>0</mbd:Z-INDEX>
          <mbd:ALT></mbd:ALT>
          <mbd:PHONENUMBER></mbd:PHONENUMBER>
          <mbd:HYPERLINK><![CDATA[]]></mbd:HYPERLINK>
          <mbd:SOFTKEYTYPE></mbd:SOFTKEYTYPE>
          <mbd:NUMKEYTYPE></mbd:NUMKEYTYPE>
        </STATEMENT>
      </DESCRIPTOR>
    </ITEM>
    <ITEM ID="2">
      <DESCRIPTOR>
        <STATEMENT Type="text/xml">
          <mbd:OBJECTTYPE>TEXT</mbd:OBJECTTYPE>
          <mbd:SUBTYPE>NULL</mbd:SUBTYPE>
          ... 종략
        </STATEMENT>
      </DESCRIPTOR>
    </ITEM>
    ... 종략
  </CONTAINER>
</DIDL>
    
```

<표 2> DIDL 아이탬의 속성태그

요소명	설명
<mbd:OBJECTTYPE>	현재 객체의 타입을 정의(이미지, 텍스트, 버튼 등)
<mbd:SUBTYPE>	이미지의 종류나 입력박스의 종류 등의 하위 타입을 정의
<mbd:RESOURCE>	객체의 실제 리소스의 경로
<mbd:POSITION>	모바일 페이지에서의 객체의 위치와 크기를 정의
<mbd:Z-INDEX>	여러 객체 중에서의 우선 순위를 정의
<mbd:ALT>	객체에 대한 부가 설명
<mbd:PHONENUMBER>	전화번호 링크에 대한 정의
<mbd:HYPERLINK>	객체에 대한 하이퍼 링크에 대한 정의
<mbd:SOFTKEYTYPE>	객체에 평선키가 연결 될 경우 이를 정의
<mbd:NUMKEYTYPE>	객체에 숫자 키가 연결 될 경우 이를 정의

3.2 CallBack SMS

앞의 과정을 통해 생성된 DIDL 페이지는 각 사용자의 계정공간에 저장된다. 이렇게 저장된 DIDL을 사용자가 확인하기 위해서는 휴대폰을 통해 접속할 수 있는 경로 및 방법이 필요하다. 최근의 스마트폰에서는 모바일 주소 입력이 편리하지만, 휴대폰에서는 URL을 입력하는 것이 어렵다. 또한 본 시스템의 특징상 실시간으로 만들어지는 DIDL 기반의 모바일 페이지는 그 생성 시점을 사용자가 알 수 없다. 따라서 사용자가 휴대폰의 브라우저에서 직접 URL을 입력하여 생성된 DIDL을 확인하는 것은 불가능하다. 이를 위해 CallBack SMS를 이용하고자 한다. 실시간으로 생성되어 각 사용자의 계정공간에 저장된 DIDL은 사용자에게 따른 접속 URL을 가질 수 있다. 이 URL을 DIDL이 생성되어 저장될과 동시에 사용자에게 SMS로 보내주는 것이다.

현재 3개 통신사 모두 SMS에 CallBack URL을 연결하여 보낼 수 있다. 간단한 메시지 내용과 앞에서 생성한 URL을 포함한 SMS가 사용자에게 발송되면 사용자의 SMS 화면에 URL이 보이지는 않지만, 사용자는 접속 버튼을 눌러 SMS에 연결된 URL로 휴대폰의 브라우저를 이용하여 접속할 수 있게 된다. 이렇게 되면 사용자는 직접 URL을 입력하여 접속하지 않아도 실시간으로 생성된 DIDL에 접속하여 게시판의 내용을 확인 할 수 있게 된다.

다만 이 때 SMS사용과 휴대폰의 웹 접속에 대한 서비스 이용 비용이 발생하는 단점이 존재한다. 하지만 웹 게시판의 의미 없는 내용에 대한 검색 결과를 알려주는 것이 아니라 특정 상품 및 브랜드에 대한 네거티브 게시글에 대하여 즉각적으로 대응할 수 있도록 활용한다면 SMS 발생 비용에 비교할 수 없을 만큼 큰 효과를 기대할 수 있다.

3.3 모바일 웹 변환 및 서비스 호환성

3.3.1 디바이스 분석

PC웹 게시판의 내용을 DIDL로 정의한 모바일웹 콘텐츠가 생성되고 휴대폰에서 접속한다고 해도 이를 휴대폰에 서비스하기 위해서는 휴대폰에서 볼 수 있는 콘텐츠로 변환이 필요하다.

이는 앞에서도 언급한 바와 같이 휴대폰마다 또한 이동 통신사마다 지원하는 콘텐츠가 상이하기 때문이다. 이처럼 생성된 콘텐츠를 휴대폰이 지원하는 콘텐츠로 변환하기 위해서는 현재 서비스를 요청하는 휴대폰이 어떤 종류의 콘텐츠를 지원하는지, 또한 어느 통신사에 가입되어 있는지를 가장 먼저 확인하여야 한다.

이를 위해 휴대폰이 서비스를 요청 시 해당 단말기의 헤더를 분석하게 된다. 단말기의 헤더 또한 통신사마다 상이하며 <표 3>에 휴대폰의 헤더정보를 보여주고 있다. 헤더정보의 많은 정보 중에서 서비스에 필요한 요소는 단말기의 제조번호를 나타내는 USER-AGENT, X-UP-SUBNO와 단말기에 내장되어 있는 브라우저의 종류 및 버전을 나타내는 BROWSER-TYPE, 단말기의 하드웨어 정보를 나타내는 WAP-HARDINFO, HTTP_DEVICE_INFO 이며 이를 분석하여 <표 4>와 같은 정보를 파싱한다. 표에서 보는바와 같이 가입되어 있는 통신사, 휴대폰이 가지고 있는 브라우저 종류 및 지원 마크업, 이미지 종류, 액정화면 크기 등 서비스에 필요한 모든 정보들을 분석하게 된다.

<표 3> 통신사별 브라우저 헤더 정보

통신사	헤더정보
SKT	Connection : Keep-Alive Accept : application/x-skt-lbs, /* Accept-Encoding : gzip, deflate Accept-Language : ko Cookie : MIN=6268****; PHONE_TYPE=LGD0; CARRIER=010; BROWSER_TYPE=40; HTTP_REQ_TIME=20080904124213NGP151; Host : www.winplus.com User-Agent : 0106CLGD04011240320362216*****;10;1;18;0 Browser-Type : 40; Version=v5.0.SKT071226r
KTF	Accept : /*/* Accept-Encoding : chunked Accept-Language : ko, * User-Agent : Mozilla/1.22 (compatible;KUN/2.2.1; IM-S240K; CellPhone) HTTP_PROXY_INFO : PNAME:pasgw2;PTIME:20081219221410 COUNTER : 4 HTTP_PHONE_NUMBER : 82010***** HTTP_PHONE_SYSTEM_PARAMETER : WCDMA,MCC:450, MNC:08, LAC:1c02, RAC:01, Cell_ID:0e20165 HTTP_DEVICE_INFO : LX:240,LY:320,CL:18 HTTP_DRIVER_INFO IMG:SIS2 NBMP JPEG MCARD MCOUPON,SN D:MA5 SMAF KMP MSGR VIB,VOD:HWVOD H.264,CIF:CIF1 HTTP_PLATFORM_INFO : PNAME:KTFWIPI,PVER:V2.0.2-1.0B,PID:1225 HTTP_PLATFORM_INFO2 : PNAME:BREW,PVER:V3.1.5,PID:29028

	Proxy-Connection : Keep-Alive HTTP_MM_INFO : MSS:VGA,MBR:MPEG4=0450,FPS:MPEG4=30,
LGT	Cache-Control : no-cache Connection : Keep-Alive Accept : application/x-skt-lbs, /*/* Accept-Language : ko Host : www.winplus.com User-Agent : 152PCIMS250085002402963017161618010*****13VBBJ08E06 x-wap-profile : http://cas.ez-i.co.kr/cgi-bin/gettrfall.cgi WAP-UserInfo : Min=010*****13; SubID=EB10-20080712-703808801; InitUrl=http://wap2.ez-i.co.kr/wireless/charge_noti.asp; BrowserType=EmbiderTM 5; WapVer=2.0; WmlVer=2.0; WmlScriptVer=; XhtmlVer=1.1 Mobile; Adult-Info=10; MIC=00000000 WAP-HardInfo : DeviceType=IM-S250L; ScreenSize=240x320; ScreenSizeChar=10X22; BitsPerPixel=18 WAP-SoftInfo : PhoneAppInfo=? WAP-JavaInfo : JavEnabled=M; JavaPlatform= X-UP-SUBNO : EB10-20080712-703808801_airnet019.ezweb019.com

<표 4> 헤더정보 설명

정보	설명	값
PhoneID	휴대폰의 제품코드	IM-240k, LGD0, ...
Provider	통신사	SKT, KTF, LGT
Browser Type	설치되어 있는 브라우저 종류	KUN, NMB, LION, ...
Markup	지원하는 마크업 언어	mhtml, xhtml, sk-wml, wml 2.0
Image Type	지원하는 이미지 파일 타입	jpg, gif, sis, bmp, wbmp...
Display Width	액정화면 넓이	120, 176, 240, ...
Display Height	액정화면 높이	176, 240, 320, ...
Color Depth	액정화면이 지원하는 색정보	ICD4, ICD8, ICD16, ICD24

3.3.2 모바일 마크업 생성

단말기에 대한 정보 분석이 끝나면 휴대폰에 지원 가능한 마크업 언어로 합성된 DIDL을 변환하게 된다. 휴대폰이 지원하는 마크업 종류는 mhtml, xhtml, sk-wml, up-wml, wml2.0, html, hdml, html5이며 단말기 분석에 따라 변환할 마크업 언어가 결정된다.

PC 웹 게시판에서 추출된 각각의 아이템은 PC웹 게시판 레이아웃의 순서대로 정렬된다. 정렬된 아이템은 XML 문서의 기본단위를 노드로 취급하는 DOM(Document Object Model)을 이용해서 아이템 하나하나가 휴대폰이 지원하는

마크업의 오브젝트 태그 형태로 변환된다. 이렇게 변환된 오브젝트에 각각의 마크업의 헤더를 연결하여 마크업 언어의 생성을 완료하게 된다.

3.3.3 이미지 변환

현재 한국 내 이동통신사에서 사용되고 있는 무선 마크업언어별로 지원하는 미디어 포맷중 이미지 포맷은 각 통신사 별로 다양한 특성을 갖고 있다. 무선 인터넷에서 일반적으로 사용되는 이미지 포맷은 wbmp, bmp, jpg, gif, png이며 그 외의 이미지 포맷은 해당 저작권을 개발 업체에서 소유한 이미지 포맷들이다. 따라서 모바일 UCC 폰피 콘텐츠에 포함되어 있는 이미지 및 광고콘텐츠의 이미지를 각각의 휴대폰에서 지원하는 이미지 포맷으로 변환할 수 있어야 하며 이를 위해 시스템은 다양한 형식(JPEG, GIF, BMP 등)의 그림파일들을 읽어 들여 서비스를 요청하는 휴대폰이 지원하는 이미지 포맷으로 자동 변환(WBMP, BMP, JPEG, GIF)한다.

이 때 파일 타입의 변환 뿐 아니라 크기의 변환도 필요하며, 이는 휴대폰의 화면 크기나 일정치 않기 때문이다. 즉 디바이스 분석을 통하여 얻어진 휴대폰의 화면 사이즈와 모바일 UCC 폰피가 편집되었을 때의 편집화면 사이즈의 비율에 맞추어 이미지의 크기를 변환하게 된다. 마지막으로 각각의 휴대폰이 지원하는 이미지 파일의 컬러 depth(흑백, 2bit gray, 4bit gray, 8bit color, 16bit color, 24bit color) 또한 다르기 때문에 디바이스 분석에서 얻어진 정보에 맞추어 컬러 depth를 변환한다.

4. 실험 및 결과

본 연구에서는 사용자가 지정한 특정 게시판에 설정된 키워드를 포함하는 게시물을 등록한 후 이를 검색시스템이 검색하여 모바일 페이지로 생성하고 이를 SMS로 보내어 접속하는 실험을 진행하였으며, 실험환경과 실험결과는 다음과 같다.

4.1 실험환경

실시간 검색된 게시판을 휴대폰에 보이기 위한 검색시스템 및 모바일 웹 시스템의 환경은

<표 5>와 같으며 실험에 사용된 휴대폰 단말기는 <표 6>과 같다.

<표 5> 게시판 검색 및 모바일 웹 생성 시스템 환경

구분	내용
OS	Microsoft Windows Server 2003 Standard Edition SP2
CPU	Intel Xeon E5405 2.40 GHz
Memory	3GB RAM
Web Server	IIS 6.0

<표 6> 실험에 사용된 휴대폰 단말기

단말기 명칭	이동통신사	브라우저 종류	Markup
SCH-W420	SKT	NMB 50	WML 2.0
SPH-V9850	LGT	LION	WML 2.0
SPH-W2400	KTF	KUN	XHTML
SS74	SKT	NMB	SK-WML
IM-S240	KTF	KUN	XHTML

4.2 실험결과

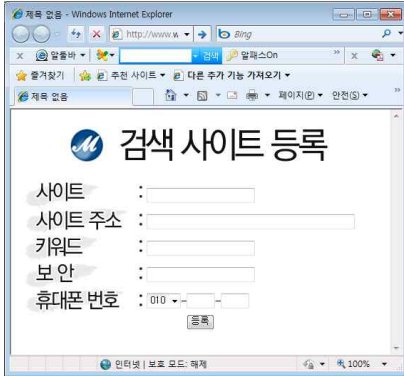
특정 문자 실시간 검색 프로그램은 사용자가 등록해둔 내용을 바탕으로 웹사이트 게시판을 특정 문자로 검색을 진행한다. (그림6)은 검색할 게시판, 키워드 등을 설정하여 검색 정보를 등록하는 화면이며, 등록에 필요한 정보는 다음과 같다.

- 사이트 주소 : 알람을 받기 원하는 웹사이트 게시판을 전체 URL을 입력한다.
- 키워드 : 사용자가 검색되기 원하는 검색어를 입력한다. 문자의 형식이 아닌 단어의 형식으로 입력한다.
- 보안 : 웹사이트의 SSL 보안 적용이 되어 있는지 확인한다.
- 휴대폰 번호 : 사용자가 휴대폰으로 SMS를 받을 번호를 입력한다.

위와 같이 정보를 입력한 후 등록 버튼을 누르면 사용자의 입력한 정보가 검색 데이터베이스로 등록이 된다.

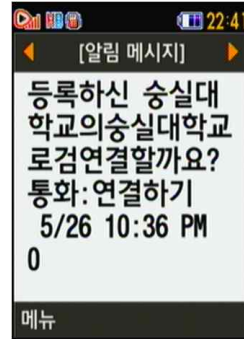
키워드기반 실시간 검색 프로그램은 사용자가 등록한 정보에 의하여 게시판을 검색하게 된다. (그림 7)은 검색 프로그램의 구동 화면이다. 여기서 검색 주기를 선택할 수 있으며 선택한 검색 주기대로 사용자가 등록한 키워드로 웹사

이트의 게시판 검색을 시작한다. 키워드 기반 검색 알고리즘에 따라 새로운 게시물이 발견되면 이를 실시간으로 DIDL로 저장하고 Callback URL이 연결된 SMS를 발송하게 된다.

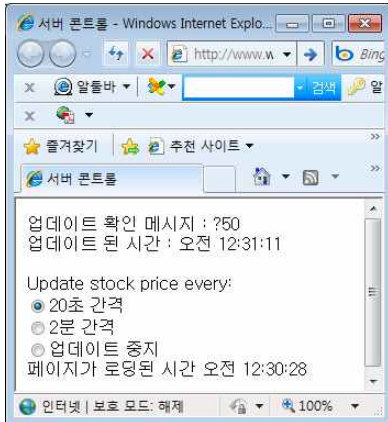


(그림 6) 검색 사이트 등록 화면

동-사용자의 게시판 및 키워드 등록-검색 시스템에서의 게시판 검색-모바일 웹 생성 및 Call back SMS 발송-사용자확인 의 일련의 과정을 거쳐 사용자는 실시간으로 등록되는 게시물을 보다 쉽게 모니터링하고 관리할 수 있게 된다.



(그림 8) CallBack SMS 메시지



(그림 7) 키워드 기반 실시간 검색 프로그램 실행



(그림 9) 모바일 웹 서비스 모습

(그림 8)는 시스템에서 발송한 Callback URL이 연결된 SMS의 모습이다. Callback URL이 연결된 SMS는 그림에서 보는 것과 같이 연결하기 기능이 있으며 접속버튼인 통화버튼을 눌러 연결된 URL에 접속할 수 있게 된다.

이렇게 접속 요청을 하면 시스템은 사용자의 휴대폰 단말기에 적합하게 마크업 페이지를 실시간으로 변환하여 서비스를 제공한다. (그림 9)은 사용자 휴대폰에 맞게 마크업이 생성되어 서비스된 모습이다.

이렇듯 키워드 기반 실시간 검색 시스템 구

5. 결론

우리는 본 연구를 통해 모바일웹의 새로운 서비스를 창출하고자 하였다. 모바일웹은 언제 어디서나 접속 가능하다는 장점을 가지고 있지만 규격의 비표준화로 인하여 콘텐츠를 미리 생성하여야 하기 때문에 실시간 응용 서비스가 어렵다는 단점을 가지고 있다. 이를 위해 규격의 비표준화를 해소할 수 있도록 DIDL 및 실시간 마크업 생성이라는 방법을 적용하여 모바일웹을 응용서비스에 적용하고자 하였다. 본 연구에서 개발한 응용서비스는 키워드 기반의 게시판 정보 모니터링 및 알리미 서비스이다.

키워드 기반의 게시판 정보 모니터링 및 알리미 서비스는 크게 3가지 과정으로 구현하였다. 첫 째는 사용자가 설정한 게시판과 키워드에 따라 해당 웹페이지의 게시판영역을 추출하고 이를 모니터링 하는 과정. 둘째는 검색된 게시판의 내용을 DIDL을 사용하여 모바일웹 생성이 가능한 디지털 아이템으로 생성하는 과정. 마지막으로 사용자가 확인할 수 있도록 Callback SMS 로 발송하고 접속한 사용자의 휴대폰의 분석을 통해 실시간으로 콘텐츠를 변환하여 서비스를 제공하는 과정이다.

우리는 실험을 통해 이 3가지의 과정이 정상적으로 이루어지는지를 검증하였고 이러한 과정을 통하여 두 가지 기본적인 목적을 달성하였다. 첫 번째는 데이터베이스나 XML등 사용자에게 공개되지 않은 게시판의 내용을 모니터링 할 수 있다는 점이고, 두 번째는 모바일 웹과 SMS를 이용해서 사용자가 실시간으로 모니터링 된 내용을 확인 할 수 있다는 것이다.

모니터링 서비스는 실시간성이 제일 중요하며 모바일 웹은 언제 어디서나 이용할 수 있다는 장점에서 실시간성을 가장 효과적으로 달성할 수 있는 도구이다. 하지만 모바일웹은 규격의 비표준화로 인하여 실시간 응용 서비스가 어려웠던 문제가 있었다.

본 연구는 이를 DIDL 및 실시간 마크업 생성기술을 통하여 해결하였다. 이를 통해 본 연구는 모바일 웹의 단점을 극복하고 모바일 웹이 가지는 최대의 장점을 그에 맞는 서비스에 연결했다는 의미를 부여할 수 있다.

본 연구가 기술적인 방법 즉 모니터링 기법 및 모바일웹페이지로의 변환과 그에 대한 구현에 초점을 맞추고 있어 기본적인 기능 구현에 치중되어 있는 것이 사실이다. 차후 연구는 이 시스템을 SSL 보안이 적용되어 있는 게시판에 적용하는 부분과, 서비스를 상용으로 적용하여 사용자 만족도 및 실제적인 효과에 대하여 분석해 보고자 한다.

참 고 문 헌

[1] 전진욱 “정보 수집 에이전트를 사용한 어린이 온라인 교육 정보 검색 시스템의 설계 및 구현”, 대구가톨릭대학교 교육대학원 석사학위논문, 2002.

[2] 홍석주, “역 RSS 웹 크롤링 검색엔진의 설계 및 구현,” 한국통신학회논문지, vol.34, no.2, pp. 139~147, 2009.

[3] 김광식, “다중 웹 게시판에서의 정보수집 및 관리 시스템 설계 및 구현,” 한국콘텐츠학회 종합학술대회 논문집, vol.3, no.1, pp. 337~340, 2005.

[4] 최재훈, “스마트 약상자를 이용한 약복용 모니터링 시스템 개발,” 한국컴퓨터종합학술대회 논문집, vol. 36, no.1, pp. 172~176, 2009.6

[5] 김연석, “HTML 문서의 테이블 식별,” 한국정보과학회 2004년도 봄 학술발표논문집, vol.31, no.1, pp. 163~165, 2004.4

[6] 이태민, “게시판 형태 HTML테이블의 논리적 구조 분석,” 정보창의교육논문지, vol.1, no.1, pp.25~33, 2007.12

[7] 이경호, “HTML 테이블의 논리적 구조 분석을 위한 효율적인 방법,” 멀티미디어학회 논문지, vol.9, no.9, pp. 1231~1246, 2006.9

[8] 강주영, “웹 페이지의 내재 규칙 습득 과정에서 규칙 식별 역할에 대한 효과 분석,” 한국지능정보시스템학회논문지, vol.11, no.1, pp. 123~161, 2005.6

[9] 김연석, “HTML 문서의 테이블 식별을 위한 효율적인 알고리즘,” 멀티미디어학회 논문지, vol.7, no.10, pp.1339~1353, 2004

[10] 김종근, “모바일 UCC폰피에 독립적인 실시간 모바일 광고 합성 기법,” 정보처리학회논문지D, vol.17-D, no.5, pp. 371~382, 2010.10

[11] MPEG MDS Group, “Information technology - Multimedia framework (MPEG-21) - Part 2: Digital Item Declaration,” ISO/IEC TR 21000-1:2005, Final Draft.

김 종 근



2007년 : 숭실대학교 대학원 미디어학과(공학석사)
 2009년 : 숭실대학교 대학원 미디어학과(박사수료)

관심분야 : 디지털콘텐츠, 멀티미디어, 모바일시스템

심 근 호



2006년 :고려대학교 전산학과
(학사)

2010년 :송실대학교 대학원 미디어
학과(석·박사 통합과정 수
료)

관심분야 : 모바일, 마크업, 멀티미디어

이 요 셉



2010년 :송실대학교 대학원 미디어
학과(공학석사)

2008년~2010년 :(주)유니웹스

2011년~현 재 :코원시스템(주)

관심분야 : 디지털콘텐츠, 인터넷미디어, 유비쿼터스

임 영 환



1979년 :한국과학원 대학원 전산학
과(공학석사)

1986년 :Northwestern University 전
산학과(공학박사)

1979년~1996년 :한국전자통신연구소 책임연구원

1996년~현 재 :송실대학교 미디어학부 교수

관심분야 : 웹마이닝, 인공지능, 멀티미디어, 모바일