

지능정보단말용 멀티모달 브라우저 설계

이동우⁰ 조수선 한동원
한국전자통신연구원 정보기전연구부
(hermes⁰, sscho, dwhan)@etri.re.kr

Design of Multi-modal Browser for Post-PC

Dong-Woo Lee⁰ Su-Sun Cho Dong-Won Han
Dept. of Information Appliance Technology,
Electronics and Telecommunications Research Institute

요 약

정보통신 분야의 기술 발달이 가속화 됨에 따라 종합 정보기기 개념의 PC와는 달리 정보이용 환경과 사용목적에 따라 용도가 단순하면서도 제공 기능별로 특화되고 사용자 편의성을 극대화시키는 새로운 컴퓨팅 기술과 정보기기의 요구가 증대되고 있다. 그러나 Post-PC의 특성상 기존 PC와는 달리 입출력의 제한 및 시스템 자원의 제약사항이 많은 게 사실이다. 본 논문에서는 이러한 환경에 맞는 웹 브라우저를 설계하고 개발하는 것이 목적이다. Post-PC의 입출력의 제약 사항을 극복하기 위한 방법으로 입출력의 다양성을 지원하는 다중인식 지원 선택형 브라우저를 설계한다.

1. 서론

정보통신 분야의 기술 발달이 가속화 됨에 따라 종합 정보기기 개념의 PC와는 달리 정보이용 환경과 사용목적에 따라 용도가 단순하면서도 제공 기능별로 특화되고 사용자 편의성을 극대화시키는 새로운 컴퓨팅 기술과 정보기기의 요구가 증대되고 있다. 기존 PC는 높은 가격에 휴대성과 편리성이 결여되는 반면, Post-PC는 non-PC 개념의 정보기기로 이러한 요구를 만족시키고 있다. 유선망 중심의 기존 PC는 정보창출과 생산자 역할을 하게 되고, 무선인터넷의 급속한 확산으로 정보소비자 역할을 담당할 신 개념의 정보기기인 Post-PC가 필연적으로 요구되고 있는 실정이다. 또한 인터넷 접속환경의 다양화로 사용자가 자신에게 가장 익숙하고 편리한 방법으로 언제, 어디서든지 인터넷을 사용할 수 있는 Post-PC 통신 환경으로 변화되면서 정보 서비스 욕구의 다양화로 Post-PC를 통한 지식정보 서비스, 멀티미디어 서비스와 아울러 지능형 서비스의 요구가 증대되고 있다. 이러한 인터넷 서비스의 바탕은 웹 브라우저에 있다고 해도 과언은 아니다. 하지만 Post-PC의 특성상 기존 PC와는 달리 입출력의 제한 및 시스템 자원의 제약사항이 많은 게 사실이다. 본

논문에서는 이러한 환경에 맞는 웹 브라우저를 설계하고 개발하는 것이 목적이다.

2. 멀티 모달

앞서 말한 Post-PC의 입출력의 제약 사항을 극복하기 위한 방법으로 입출력의 다양성을 지원하는 다중인식 지원 선택형 브라우저를 설계한다. 다중인식 지원 시각/청각 선택형 웹브라우저(이후, 선택형 브라우저로 표기)는 기존의 시각형 웹 콘텐츠뿐만 아니라 청각형 웹 콘텐츠를 지원하고 음성 및 글러브 입력을 지원하는 멀티 모달 브라우저이다. “시각/청각 선택형”은 출력의 다양성을 나타내고 “다중인식”은 입력의 다양성을 표현하는 단어이다.

본 논문에서 설계하는 선택형 브라우저의 출력은 지능정보단말의 LCD 디스플레이와 HMD를 이용한 출력을 지원한다. 가로방향 보기를 기본으로 하며 좌우 스크롤은 없고 상하 스크롤만 있는 것을 기본으로 한다. 선택형 브라우저의 입력은 스타일러스에 의한 화면 상의 선택은 물론 음성인식 및 글러브 동작에 의한 입력을 추가로 지원한다.

2.1 사용 모드

지능정보단말에서 선택형 브라우저는 기본적인 시각형 브라우저 기능을 모두 제공하고 이에 덧붙여 청각 및 음성 인식 기능을 제공하는 것이므로 주요 용도는 “ 손 동작이 불가능한 상황에서 음성인식으로 웹 브라우저할 때(입력은 음성, 출력은 시각 또는 청각형)” 이다. 이와 같은 선택형 브라우저의 주요 용도를 가정하고, 그 외에도 실제로 구현하였을 때 의미있는 사용모드가 무엇인지 찾아내면 다음 표와 같다.

<표 1> 선택형 브라우저에서 의미있는 사용모드

	HT ML	Voice XML	TTS (이어폰)	ASR (마이크)	단말 Display	HMD	컨텐츠 변환
VO-HI	√				√		√
VO-VI	√			√		√	√
VAO-HI	√	√	√		√		√
VAO-VI	√	√	√	√		√	√
VAO-HVI	√	√	√	√	√		√

범례: VO. Visual Output (시각형 출력), AO. Auditory Output (청각형 출력), HI. Hand Input (손 입력), VI. Voice Input (음성 입력)

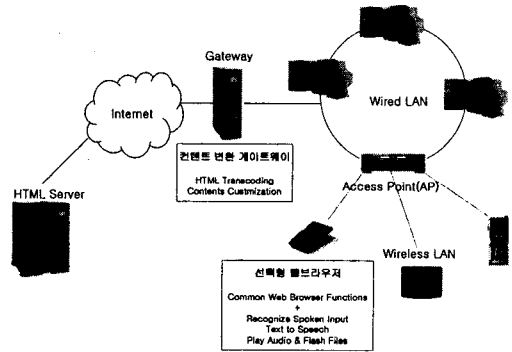
<표 1>에서와 같이 총 5 가지의 사용모드가 구현했을 때 의미가 있으며 이 중 마지막 사용모드인 VAO-HVI 형은 나머지 4 개의 사용 모드를 모두 포함하는 것이므로 가장 포괄적인 형태가 된다. 단 손 입력이 필요하므로 HMD는 사용하지 않는다. 또 각 사용모드에서 음성명령의 입력을 처리하기 위해서는 굳이 VoiceXML 이 필요없고 청각형 출력을 지원하기 위해서는 콘텐츠 일부를 VoiceXML 로 하는 것이 필요하다고 가정한다.

2.2 시스템 구성

지능정보단말의 선택형 브라우저는 멀티모달 브라우저라는 특징과 함께 콘텐츠변환 게이트웨이를 통한 웹 서비스를 지원한다는 특징을 가진다. 콘텐츠 변환은 낮은 서비스 대역폭의 무선 네트워크 및 지능정보단말의 제한적인 하드웨어 제약 때문에 필요하기도 하지만 일반적인 시각형 웹 콘텐츠를 시청각 복합형 콘텐츠로 바꾸어 주기위해서도 필요하다. 따라서 전체적인 웹 콘텐츠의 다운사이징 및 커스트마이징과 함께 적절한 시청각 출력을 위한 HTML 의 변환도 필요하다. 이를 위해서 선택형 브라우저에서는 클라이언트의 정보(화면 해상도, 색상 수, 사용 모드, 그림표시 유무 등)를

콘텐츠변환 게이트웨이에 전달한다.

물론 선택형 브라우저는 이와 같이 변환된 웹 콘텐츠를 적절히 출력할 수 있어야 한다. 콘텐츠변환 게이트웨이에 관해 자세한 것은 본 논문에서는 다루지 않는다. 전체 서비스 시스템 구성을 다음 <그림 1>과 같이 나타낸다.

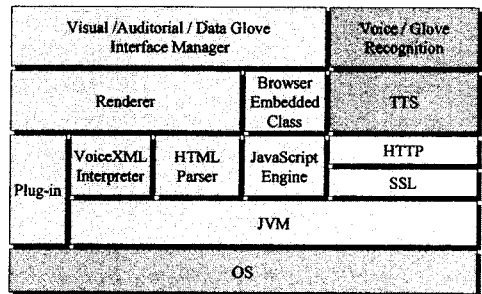


<그림 1> 선택형 브라우저의 전체 서비스 시스템 구성도

그림 1 에서 나타났듯이 지능정보단말의 선택형 브라우저는 인터넷 상의 HTML 서버로부터 웹 콘텐츠를 다운받을 때 게이트웨이와 유선 LAN 상의 액세스 포인트 장치들 거치게 된다.

콘텐츠 변환의 예를 들면, TTS 를 사용하여 웹 페이지 내용의 일부를 청각형으로 출력하고자 할 때 콘텐츠변환 게이트웨이에서는 TTS 로 보낸 HTML 데이터의 일부를 VoiceXML 로 변환할 수 있다. 선택형 브라우저는 VoiceXML prompt 태그의 텍스트를 TTS 로 보냄으로써 청각형 출력을 지원한다.

선택형 브라우저의 내부 모듈 구성도는 <그림 2>와 같다.



<그림 2> 선택형 브라우저의 모듈 구성도

선택형 브라우저는 기본적으로 자바가상머신(JVM) 상에서 구현되고 실행된다. 따라서 운영체제에 독립적이며 적당한 JVM 만 지원되면 하드웨어 및 운영체제에 상관없이 포팅이 가능하다. 단 일부 모듈은 자바가 아닌 일반 C 또는 C++ 언어로 구현되어 JNI 를 통해 자바로 구현된 브라우저에 연결이 가능하다.

선택형 브라우저는 <그림 2>에서 보이듯이 일반적인 웹 브라우저의 필수 모듈인 HTTP/SSL, HTML Parser 및 Renderer 를 포함한다. 또한 자바 스크립트를 지원하기 위한 모듈인 자바 스크립트 엔진 및 브라우저 내장 객체 모듈을 포함한다.

2.3 멀티모달 지원 모듈

시각/청각 인터페이스를 지원하기 위한 인터페이스 관리자 및 입출력 모듈이 필요한데 음성 및 데이터 글러브를 통한 인터페이스 관리자, 음성 인식 및 합성 모듈과 데이터 글러브 동작 인식 모듈, 음성 마크업랭귀지를 지원하기 위해 표준 VoiceXMLInterpreter 등을 포함한다. 또한 브라우저와 독립된 모듈로서 TTS(Text to Speech) 및 음성/데이터 글러브 인식 모듈이 필요하다.

2.3.1 음성 및 데이터 글러브 인식 인터페이스

사용자의 음성을 브라우저의 입력신호로 사용하기 위해서 음성인식 인터페이스 모듈에서는 HTML 문서 상에서 음성인식이 가능한 목록을 만들고 내장된 음성인식 하드웨어 칩의 출력을 비교하여 적절한 동작을 수행하도록 한다.

데이터글러브의 경우는 브라우저 외부의 글러브 인식 모듈에서 글러브의 신호를 시스템의 마우스 신호와 동일한 신호로 바꾸어 주므로 브라우저에서는 일반 마우스와 같이 사용을 한다.

2.3.2 VoiceXML 지원 모듈 및 TTS (Text To Speech)

전화 단말과 연결된 VoiceXML 인터프리터 및 보이스 포털 서비스를 위한 일반적인 VoiceXML 문서는 대개 다음과 같은 내용을 포함한다. 1) spoken prompts (synthetic speech) 2) output of audio files and streams 3) recognition of spoken words and phrases 4) recognition of touch tone (DTMF) key presses 5) recording of spoken input 6) control of dialog flow 7)

telephony control (call transfer and hangup).

그러나 선택형 브라우저는 단말내에 VoiceXML 인터프리터와 음성 인식 및 합성 모듈을 포함하고 있으므로 선택형 브라우저가 대상으로 하는 VoiceXML 문서는 위의 7 가지 항목 중 1), 2), 3), 6) 항목의 내용으로 구성되어 있다고 가정한다. 즉 선택형 브라우저는 음성 및 오디오의 출력과 음성 인식, 그리고 다이얼로그 컨트롤은 포함하고 음성 녹음, DTMF 인식 및 텔레포니 컨트롤은 포함하지 않는 VoiceXML 문서를 취급한다. 따라서 선택형 브라우저의 VoiceXML 지원 모듈은 이와 같이 제한된 내용의 VoiceXML 을 처리한다.

선택형 브라우저의 VoiceXML 지원 모듈은 해당하는 VoiceXML 파일을 파싱하고 VoiceXML 의 출력 부분을 TTS 에 넘겨주는 일들을 수행한다. VoiceXML Parser 는 콘텐츠변환 게이트웨이를 통해 전달된 VoiceXML 파일을 파싱하여 ASR API 및 Prompt API 를 호출하여 음성 인식 모듈 및 TTS 모듈과 인터페이스 한다.

선택형 브라우저를 위한 TTS 는 30Mbyte 이내의 음성 DB 및 실행엔진을 갖는 소형 합성기를 사용한다.

3. 결론

본 논문에서는 Post-PC 의 입출력의 제약 사항을 극복하기 위한 방법으로 입출력의 다양성을 지원하는 다중인식 지원 선택형 브라우저를 설계하고 멀티모달 지원을 위해서 필요한 모듈들을 정의하고 설계하였다. 이러한 멀티모달 브라우저는 손으로부터 자유로워야 하는 산업현장의 작업장 등에서 활용이 가능하다.

앞으로도 지능정보단말의 취약점인 입출력문제의 개선을 위해 보다 편리한 입출력 방법에 대한 고찰이 필요하겠다.

4. 참고 문헌

- [1] S. Maes, " Multimodal browser architecture, Overview on the support of multi-modal browsers in 3GPP," IBM, Aug. 2001, <http://www.w3.org/2002/mmi/2002/MM-Arch-Maes-20010820.pdf>.
- [2] VoiceXML Forum, Voice eXtensible Markup Language (VoiceXML™) version 1.0, May, 2000.