
VoiceXML VUI Browser 설계/구현

장민석 · 예상후

국립군산대학교, 컴퓨터정보학과

Design and Implementation of VoiceXML VUI Browser

Min-seok Jang · Sang-hoo Yea

Kunsan National University, Dept. of Computer Information Science

E-mail : msjang@cs.kunsan.ac.kr, badpotat@cs.kunsan.ac.kr

요 약

현재의 웹 환경은 HTML로 구성이 되어있으며 이로 인해 하이퍼링크를 따라가기 위해 마우스 클릭을 통해 작업하는 GUI 환경이 주를 이룬다. 하지만 이러한 방법은 인간이 가장 손쉽게 사용하는 음성과 비교해 볼 때 상당히 불편한 축에 속한다. 본 논문에서는 이를 해결하기 위해 현재 무르익은 음성인식/합성 기술과 전화기를 통해 정보를 제공하고자 하는 XML의 파생언어인 VoiceXML을 이용하여, 현재의 HTML 주축인 GUI 웹 환경을 VoiceXML을 이용하여 VUI(Voice User Interface) 환경으로 전환해 주는 VoiceXML VUI Browser를 설계/구현하였다.

ABSTRACT

The present Web surroundings is composed of HTML(Hypertext Mark-up Language) and thereby users obtains web informations mainly in GUI(Graphical User Interface) environment by clicking mouse in order to keep up with hyperlinked informations.

However it is very inconvenient to work in this environment comparing with easily accessed one in which human's voice is utilized for obtaining informations.

Using VoiceXML, resulted from XML, for supplying the information through telephone on the basis of the contemporary matured technology of voice recognition/synthesis to work out the inconvenience problem, this paper presents the research results about VoiceXML Web Browser designed and implemented for realizing its technology.

키워드

XML, VoiceXML, 음성인식, 음성합성, VUI

1. 서 론

현재의 컴퓨팅환경은 GUI(Graphic User Interface)가 주류를 이루고 있다. 이는 기존의 Text 환경의 인터페이스보다 확실히 간편한 인터페이스를 가능하게 해준다. 현재의 웹서비스 또한 이러한 GUI환경을 택하고 있다. 하지만 HTML로 구성되어 있는 웹 페이지들은 정보를 단순히 시각에만 의존하고 있으며 이는 포털서비스와 같은 거대한 사이트에서는 익숙하지 않은 사용자가 사용하기에는 너무 많은 정보를 포함하고 있다. 이를 해결하기 위해 VoiceXML을 이용하여 VUI

환경을 마련한 VoiceXML VUI Browser를 설계/구현하였다. 현재 VoiceXML은 VoiceXML 포럼에서 AT&T, IBM, 루슨트 테크놀로지, 모토로라 등 정보통신 분야의 4개 거대 기업체를 중심으로 표준화가 진행 중이며 2002년 4월에 ver 2.0이 표준으로 인정되었다 [1][2]. 또한 W3C에서도 VoiceXML을 이용하여 웹에서 VUI를 제공하기 위해 XHTML+Voice라는 이름으로 논의가 진행 중이다. 본 논문에서는 VoiceXML을 기존의 XML Island 기술을 이용하여 VoiceXML 버전 1.0을 HTML에 첨가한 웹페이지를 대상으로 작동하도록 설계/구현하였다.

II. 국내외 VoiceXML 기술개발 현황

VoiceXML연구가 진행되는 동안에도 관련 업체에서는 이를 사용하기 위한 각종 응용솔루션들을 개발하고 있는 상태이다. 특히 IBM사는 자사의 WebSphere 서버의 Voice Toolkit을 이용해 VoiceXML을 서비스할 수 있도록 개발했으며, IT의 대표기업인 MS사는 VoiceXML이라 명명하지는 않지만 VoiceXML의 성격을 반영하는 XML을 이용하여 자사의 상품 등에 적용하고자 노력하고 있다. 이외에도 뉘앙스, 스피치 워크스 등 음성 서비스 업체에서 또한 자사의 플랫폼 및 서버등을 개발/업그레이드하여 VoiceXML을 수용하고자하고 있다.

VoiceXML은 XML의 서브셋이므로 XML의 영향을 가장 많이 받는다고 할 수 있다. 그러므로 VoiceXML을 위한 파서 등은 기존의 XML파서를 이용해서 개발함으로써 개발비용을 절감할 수 있다. 현재 VoiceXML을 지원하는 파서들은 대부분 외국 제품이며 또한 그나마 있는 국내의 파서들은 아직 공개가 되어있지 않은 상태이다.

VoiceXML은 현재 2002년 4월에 버전 2.0을 발표하고 이후로도 계속 표준화를 진행중이다. 이러한 표준화 진행에 우리나라에서는 보이즈웨어와 삼성전자가 VoiceXML 포럼에 참가하여 함께 일을 하고 있다. 또한, 각 대학의 연구실에서 이와 관련된 각종 기반 기술을 개발하고 있다. 하지만 이러한 연구들은 아직까지 인터프리터 내지는 파서개발에 그치고 있어 좀더 응용할 수 있는 기반기술을 개발해야 하는 실정이다. 또한 W3C에서도 본 연구와 같이 VoiceXML을 웹 환경에서의 인터페이스로 사용하기 위해 2001년 12월부터 논의중이다 [4].

국외에서는 2002년 3월에 열린 VoiceXML Conference & Expo를 살펴보면 VoiceXML 기술관련 연구로는 VoiceXML을 좀더 효과적으로 이용하기 위한 VoiceXML의 Dialog & Grammar에 관한 연구와 각종 플랫폼에서의 효과적인 VoiceXML구축 방법 등이 축을 이루었다 [3]. 이를 보면 국내에서는 아직까지 기반기술에 중점을 맞추고, 국외에서는 이미 실용화 단계의 작업을 하고 있는 상황이다. 본 연구는 이러한 점을 인지하고, 국내에서 각 VoiceXML관련 기술을 이용하여 현재 바로 사용할 수 있는 기술을 연구하였다.

III. VoiceXML VUI

VoiceXML VUI를 제공하기 위한 방법으로 Browser는 웹 서버로부터 스크립트와 DB에 의해 구성된다.

HTML문서를 전달받아 HTML내부에 XML Island로 첨가된 VoiceXML 문서를 추출하여 이를 음성인식기/합성기를 통해 사용자와 적절히 반응하여 VoiceXML 문서 처리를 수행하게 하였다. 또한 일반적인 HTML 브라우저에서 사용되어지는 앵커태그는 기존의 방법과 동일하게 마우스를 통해 브라우징을 수행하며, 필요에 의해서 VoiceXML로 작성된 Dialog를 통해 음성 인터페이스를 제공하게 하였다. 이렇게 함으로써 현재의 환경에 최대한 수정 없이 사용되어질 수 있어 사용자의 접근을 용이하게 하며 또한 개발자들에게도 간단한 VoiceXML을 익힘으로써 다양한 방법으로 웹페이지를 개발할 수 있는 토대를 마련할 수 있으리라 본다. 이런 VoiceXML문서는 다음의 [그림 3.1]에서와 같은 모습으로 HTML문서상에 포함되어 VoiceXML VUI Browser에서 처리가 이루어진다.

```
<HTML>
<TITLE>VoiceXML 을 이용한 웹 VUI환경제공</TITLE>
<HEAD>음성을 이용할 수 있습니다.</HEAD>
<BODY>
  <XML ID="vxml">
    <vxml version="1.0">
      <form>
        .....
      </form>
    </vxml>
  </XML>
</BODY>
</HTML>
```

[그림 3.1] VoiceXML을 HTML에 삽입하여 사용하는 모습

XML Island의 특징은 HTML내에 의미를 갖는 데이터를 삽입하여 이를 HTML상의 각종 스크립트로 제어하여 HTML을 좀더 풍부한 응용을 할 수 있도록 도와주는 기술이다. 이것은 실제 HTML을 브라우징 할 경우에는 보여지지 않는 내용이다.

IV. VoiceXML VUI Browser 설계/구현

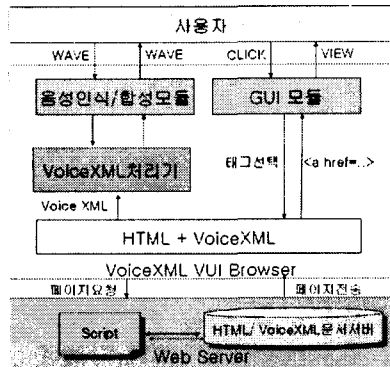
본 논문에서 구현한 VoiceXML VUI Browser는 다음과 같은 환경에서 구현되었다.

[표 4.1]에서 나타낸 바와 같이 운영 체제로는 마이크로소프트사의 윈도우즈 2000 서버를 사용하였으며, 개발 언어로는 Java와 같이 플랫폼에 독립적인 개발 환경을 제공하는 닷넷 프레임워크 상에서의 개발언어인 C#을 이용하였다[7]. 이는 추후 각종 OS상에서도

[표 4.1] VoiceXML Web Browser 개발 환경

구 분	내 용
운영 체제	Windows 2000 Server Service Pack 2
개발 언어	C# (Dot NET Frame Work)
XML 파서	MSXML
음성 인식/합성 엔진	MSAGENT 2.0 (한국어 TTS: L&H TTS 3000)

작동을 할 수 있을 것이다. 파서로는 MS사에서 무료로 배포하는 MSXML을 이용하였다. 또한 음성 인식/합성 엔진으로는 MS사의 Speech 기술로 구현된 MSAGENT 2.0을 이용하였다. 이는 Speech 5.0의 영어/중국어 인식을 할 수 있는 인터페이스를 마련해 주며, 합성엔진으로는 다른 업체의 모듈을 이용할 수 있도록 되어 있다[8]. 이는 현재까지 공개되어있는 인식/합성엔진이 없는 관계로 부족하지만 이를 이용하였다. [그림 4.1]은 VoiceXML VUI Browser의 시스템 구성도이다.

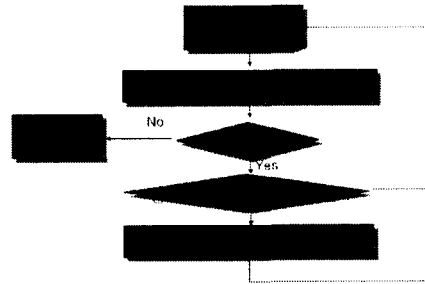


[그림 4.1] VoiceXML VUI Browser의 시스템 구성도

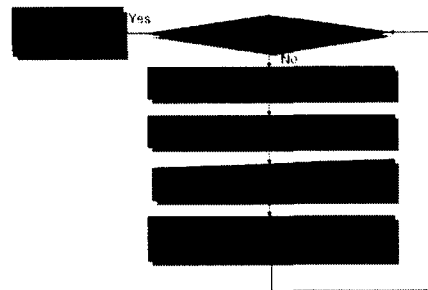
[그림 4.1]에서 VoiceXML VUI Browser는 앞서 설명한 바와 같이 HTML내에 XML Island로 함께 전달된 VoiceXML문서를 추출하여 이를 VoiceXML 버전 1.0을 기준으로 파싱을 하고, VoiceXML문서내의 Dialog를 적절한 처리를 하여 사용자에게 VUI를 제공하게 된다. [그림 4.2]은 처리기에서 VoiceXML을 처리하는 과정을 플로차트로 나타낸 그림이다.

[그림 4.2]에서 VoiceXML에서 처리해야할 품들의 아이템을 선택하여 이를 품해석 알고리즘을 적용하여 사용자에게 인터랙티브한 VUI환경을 제공하게 된다. [그림 4.3]은 품해석 알고리즘을 플로차트로 나타낸 것이다[6].

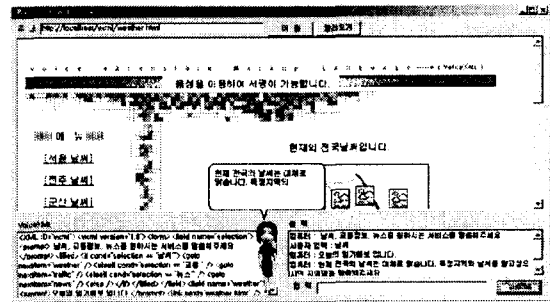
품해석알고리즘에서는 [그림 4.3]에서 나타낸 바와 같이 실제 사용자에게 나타내야 할 정보를 음성으로



[그림 4.2] VoiceXML 처리과정



[그림 4.3] 품해석 알고리즘



[그림 4.4] VoiceXML Web Browser

출력하고, 또한 사용자로부터 입력을 받아 다 적절한 반응을 나타낸다. [그림 4.4]는 본 논문에서 구현한 VoiceXML Web Browser의 모습이다.

[그림 4.4]의 VoiceXML VUI Browser는 크게 세가지 화면으로 구분된다. 우선 HTML을 나타내는 메인 화면이 있고, HTML에서 추출한 VoiceXML을 나타내는 VoiceXML출력화면이 있으며, 사용자와의 인터페이스를 시각적으로 나타내는 Input/ Output 인터페이스 부분이 있다. 현재 VoiceXML Web Browser는 음성인식/합성부분의 미약함으로 인해 음성 입력은 입력창과 병행하여 작동하며 출력은 그림에서처럼 캐릭터에 의해 음성출력과 동시에 텍스트 출력을 병행하도록 하였다. [그림 4.5]은 본 논문에서 테스트를 위해 제작한 간단한 VoiceXML문서를 포함한 HTML 페이지이다.

```

<HTML>
<TITLE>.....</TITLE>
<HEAD>
<XML ID="vxml">
<vxml version = "1.0">
<form>
<field name="selection">
<prompt>날씨, 교통정보, 뉴스중 원하시는 서비스를
말씀해주세요</prompt>
<filled>
<if cond = "selection == '날씨'">
<goto nextitem = "weather" />
<elseif cond = "selection == '교통'" />
.....
<field name = "traffic">
<prompt>현재의 교통정보를 말씀드리겠습니다
.</prompt>
<link next = "traffic.html"/>
</field>
<field name = "news">
<prompt>오늘의 주요 뉴스입니다.</prompt>
<link next = "news.html"/>
.....
    
```

[그림 4.5] VoiceXML을첨가한 HTML페이지

V. VoiceXML VUI Browser 비교

본 논문에서 구현한 VoiceXML Web Browser와 기존에 제작되어진 음성 인식 브라우저의 차이점은 [표 5-4]와 같이 구분할 수 있다.

[표 5-4] 기존 음성인식 브라우저와의 차이점

구분	VoiceXML VUI Browser	기존 음성인식 Web Browser
제품명	VoiceXML Web Browser	Sayser (voicenic), VoiceBrowser (보이스웨어)
VUI제공 방법	VoiceXML을 이용한 사용자 정의	Application Menu, 기존 HTML에서 Image link, Text link등 각종 링크정보 수집
표준화	VoiceXML 포럼	없음
기타	공개된 음성인식/ TTS엔진 이용	업체에서 개발한 음성인식/ TTS 엔진을 이용하여 개발

[표 5-4]에서 나타낸 바와 같이 본 논문에서 구현한 VoiceXML VUI Browser는 기존의 음성인식 브라우저가 브라우저의 메뉴 등을 음성으로 작동케 하는 것과는 달리 웹 상에서 VUI를 부가적으로 사용할 수 있도록 도움을 주고 있다. 또한 VoiceXML의 특성상 유연한 메뉴 구성을 할 수 있어 페이지 특성에 맞는 음성 서비스를 실시할 수 있다.

VI. 결론/ 향후 연구과제

지금까지 국내외에서 연구된 것들은 대부분 VoiceXML을 전화기에서 사용하기 위한 방법들 위주로 연구되었으며 개발되어진 브라우저 또한 PC환경에서 VoiceXML을 시뮬레이트 하는 수준에 머물렀었다. 이러한 VoiceXML관련 연구를 전화기뿐만 아니라 현재 많이 제작되어진 웹 사이트에서 VoiceXML의 장점인 사용자와 기계간의 상호 대화식 VUI환경을 제공할 수 있도록 VoiceXML VUI Browser를 구현하였다. 이는 기존의 단조로운 마우스 클릭과는 달리 사이트 방문자에게 좀더 다양한 인터페이스를 제공하여 편의를 제공할 수 있다. 하지만 본 논문에서 구현한 VoiceXML VUI Browser는 JSGF(Java Speech Grammar Format) 처리를 하기 위한 처리기가 구현되지 않아 앞으로 실용화하기 위해서는 이 부분에 대한 연구가 이루어져야 하며, 효과적인 VUI 인터페이스 구축을 위한 Dialog관련 연구가 필요하다.

참고문헌

- [1] 홍용택, XML과 VoiceXML의 효율적인 통합방안 제시, 군산대학교 교육대학원 석사, 2002. 2.
- [2] VoiceXML 포럼, <http://www.voicexml.org>
- [3] voicexmlplanet, <http://www.voicexmlplanet.com>
- [4] 월드와이드웹 컨소시엄, <http://www.w3c.org>
- [5] 김경란, VoiceXML 기반 음성 브라우저의 설계 및 구현, 성신여자대학교 대학원 석사, 2001.
- [6] 신현경, 추상구분트리 기반의 VoiceXML 해석기 설계 및 구현, 숭실대학교 대학원, 2001. 6.
- [7] 마이크로 소프트 닷넷, <http://www.microsoft.com/net/>
- [8] 마이크로 소프트 스피치, <http://www.microsoft.com/speech/>