

VoiceXML을 이용한 VUI 개발에 관한 연구

장민석 양운모⁰
국립군산대학교 컴퓨터정보과학과
msjang@kunsan.ac.kr, mica@orgio.net⁰

A Study on Development of VUI(Voice User Interface) using VoiceXML

Min-Seok Jang Woon-Mo Yang⁰
Dept. of Computer Information Science, Kunsan National University

요 약

한국현재의 컴퓨팅환경은 Text위주의 Command Line상에서의 입출력에서 GUI(Graphic User Interface)환경으로 전환되었다. 이는 사용자에게 좀더 친근한 방법으로의 컴퓨팅환경을 제공하고 있는 것이다. 하지만 아직까지 그러한 환경에 익숙해지기 위해서는 많은 습득시간이 필요하며 또한, 응용프로그램 간의 인터페이스 기능 등을 익히기 위해서는 추가적인 학습을 통해서 원활한 작업을 수행할 수 있다. 이를 해결하고자 본 연구는 음성인식/ 합성과, 현재 음성마크업 언어인 VoiceXML 등을 통해서 모색해보고자 한다.

1. 서론

음성인식/ 합성 관련연구는 상당히 오랜 역사를 가지고 있다. 외국의 경우에는 20~30년 정도의 연구기간을 갖고 있으며 우리나라 또한 타 분야에 비해 상당히 오랜 기간동안 연구되어졌으며 현재 외국과 거의 비슷한 수준으로까지 발전했다. 현재 이를 이용한 각종 음성인식 프로그램들과 가전제품 등에 응용되어 상품화 되어있으며, 일부 ARS시스템에서 음성인식을 이용한 시스템을 적용하고 있다. 본 연구는 특히 컴퓨터 응용프로그램에서의 VUI에 관한 연구로 현재까지의 VUI의 문제점들을 살피고 이를 해결하는 방법으로 VoiceXML을 통해 해결방법을 모색해보고자 한다.

VoiceXML은 XML의 서브셋으로 AT&T, 루슨트 테크놀로지, 모터롤러 등으로 구성된 VoiceXML포럼에서 제안하고 이를 W3C에서 공인한 대화형 마크업 언어의 표준이다. 이는 음성인식/ 합성 등과 같은 음성 처리 기술들을 이용하여 웹에서의 HTML을 통한 정보의 접근방법을 유무선 전화기를 이용하여 웹을 접근토록 하는 토대를 마련한다 [1]. 현재까지의 음성정보 서비스는 많은 비용을 들여 시스템을 구축하고 또한 서비스의 유지/ 보수, 갱신 등을 위해서 많은 인적/ 물적 자원을 소비되었다. 그러나 VoiceXML은 일종의 시나리오 문서로써 정보제공자가 기타 기술적인 부분을 배제하고 시나리오에 몰두함으로써 좀더 원활한 대화형 음성 응용프로그램들을 개발하고, 이를 유무선 전화기 등을 통해 이용할 수 있게 된다. 하지만 현재까지 VoiceXML관련 연구들은 유무선 전화기 등을 통해 이용하는 것에만 집중되어졌었다. 본 연구는 이러한 VoiceXML을 이용하여 VUI의 일관된 인

터페이스의 부재와 인식프로그램들이 응용프로그램에 하드코딩 되어있어 각각의 프로그램마다 사용법을 따로 익혀야하는 문제점들을 해결하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 음성인식 관련 국내외 연구들과 이이를 이용한 각종 응용프로그램과 서비스들에 대해 알아보고, 3장에서 VoiceXML의 문서 구조, 다이얼로그 모델, 문법 표현 등에 관해 기술하며, 4장에서 VUI를 위해 사용되어질 VoiceXML 엘리먼트들을 요약하고 요약된 VoiceXML을 이용하여 VUI를 제시한다. 그리고 5장에서는 본 논문의 결론을 기술한다.

2. 음성인식 국내외 연구

2.1 국외 현황

미국의 경우 1970년부터 미 국방성의 주도에 의한 ARPA 프로젝트의 일환으로 연속음성인식에 대한 본격적인 연구가 시작되어 진행되던 중 1984년부터 DARPA 프로젝트로 바뀌어 대용량 어휘 음성인식 및 구어체 언어 이해 연구가 진행되었다. 이 연구의 장기적인 목표는 100,000 단어의 어휘를 대상으로 한 연속 음성인식이며 단기적으로는 10,000 단어의 어휘를 사용하여 연속음성을 인식하되 95% 이상의 단어 인식률을 얻는 것을 목표로 하고 있다.

일본의 경우 1986년이래 15년간의 장기 계획으로 자동통역전화개발을 추진해 오고 있으며, 1987년에는 국가주도에 의한 인간과 기계화의 구어체대화를 목표로하는 "Advanced Man-Machine Interface Through spoken Language" 계획이 시작되어 대화체 언어이해 및 소음환경에서의 음성인식에 관한 연구가 진행되어 많은 결과를

도출하였다.

유럽의 경우 1983년부터 ESPRIT 프로젝트를 중심으로 하여 현재까지 약 40여 개의 프로젝트를 수행하고 있으며, 최근에는 대화체 음성인식과 사외적 요구에 중심을 둔 새로운 프로젝트들이 진행 중에 있다.

독일에서는 Grrman VERBOBILE 프로젝트(영역 : 약속 스케줄)에서 구문론적, 운율적 경계에 대한 레이블링을 도입하여 122개의 문장에 대하여 96%이상의 단어인식률을 달성하였다.

2.2 국내현황

국내에서도 1980년도에 돌면서부터 본격적인 음성인식에 관한 연구가 이루어져 오고 있다. 개발된 시스템으로는 한국전자통신 연구소의 자동통역시스템, 한국통신의 증권정보 안내시스템, 삼성전자의 음성구동 퍼스널 컴퓨터, 음성구동 셀룰러폰(삼성, LG), 음성메모장치(공성통신)등이 있으며 현재 성능개선 또는 상용화 중에 있다. 또 음성에 의한 로봇 제어에 관한 연구, 음성에 의한 자동항법 장치 등에 관한 연구도 활발히 진행되고 있다 [2].

3. VoiceXML

Voice XML은 AT&T, IBM, 루슨트 테크놀로지, 모토라 등 정보통신 분야의 4개 거대 기업체에 의해 설립된 Voice XML 포럼에서 제안한 마크업 언어로서, 음성 애플리케이션 개발을 위해 고안된 XML문서 형식의 일종이다. Voice XML포럼은 1999년 8월 Voice XML 0.9 버전을 발표후 작년 3월 이를 크게 보완한 버전 1.0을 정식 제안했다. 세계 인터넷 환경을 주도하고 있는 W3C 컨소시엄에서는 Voice XML 포럼의 제안을 받아들여 작년 5월 22일 Voice XML을 웹의 대화형 마크업 언어 표준으로 공인했다. 기존에도 WTE(The Web Telephony Engine), WAP(Wireless Application Protocol)등이 Voice XML 이전에 개발되어져 있었다.

VoiceXML은 음성 응용프로그램으로 인터넷 정보를 전달하고 음성 응용 프로그램 제작자를 하위레벨 프로그래밍이나 자원관리로부터 벗어나도록 도와준다. 이는 개발자가 음성 시나리오에 집중함으로써 개발시 많은 이점을 얻을 수 있다.

VoiceXML은 다음과 같은 특징을 갖는다.

- VoiceXML은 서비스 로직에서 사용자 상호작용 코드를 분리할 수 있다.
- 음성 서비스와 DB 서비스를 통합하여 클라이언트-서버환경으로 제공 할 수 있다.

· VoiceXML 웹서버는 여러 가지 서비스 로직을 수행하여 VoiceXML문서를 동적으로 제작할 수 있다.

· VoiceXML은 기존 웹기반의 네트워크를 이용함으로써 쉽게 네트워크 기반의 어플리케이션을 제작할 수 있다 [1][3].

4. VUI를 위한 VoiceXML을 이용한 음성인터페이스에 적합한 엘리먼트 요약/ VUI제시

VoiceXML은 XML의 서브셋이므로 현재의 XML파서와 음성인식/합성 모듈을 이용하여 구현할 수 있을 것이다. 본 논문에서 제시하는 VoiceXML을 이용한 음성 인터페이스는 다음과 같은 구조를 갖고 동작을 한다.

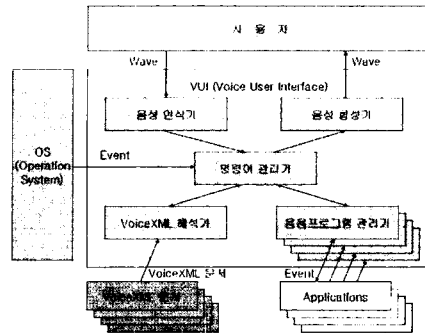


그림4-1 본 논문이 제시하는 VUI 구조

그림 4-1 에서 VoiceXML은 사용자에게 응용프로그램의 기능을 설명 해주거나 사용자가 명령을 내렸을 경우에 해당하는 동작을 응용프로그램에게 전달하여 수행토록 한다. VoiceXML은 XML의 서브셋인 만큼 아주 적은 학습으로도 사용법을 익힐 수 있는 장점이 있다.

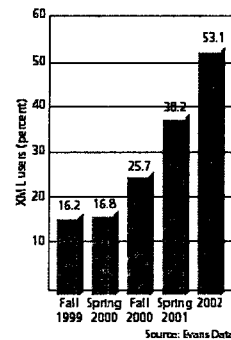


그림 4-2 XML 사용자 증가 비율

그림 4-2 에서 보는바와 같이 증가하고 있는 상황임

로 이는 사용자가 충분히 VoiceXML을 직접 작성할 수 있음을 이야기하며, 개발자 또한 음성 입출력부분을 좀 더 적은 노력으로 개발 할 수 있음을 이야기한다. 이러한 VoiceXML은 인터넷을 통해 사용자들 스스로가 작성하여 배포할 수도 있을 것이다 [4].

VoiceXML 엘리먼트들은 전화기를 이용하여 VoiceXML 브라우저에 접속해 수행할 경우 야기되어지는 것들을 고려해서 만들어진 엘리먼트들이 존재하므로 이를 전부 수용하여 음성인터페이스를 구성하는 것은 비효율적이다. 그러므로 본 논문은 다음과 같은 시나리오를 바탕으로 표 4-1 과 같이 실제 음성인터페이스 구성시 필요한 42개의 엘리먼트들을 추출하였다. 로컬한 응용프로그램 내에서 실행이 되어지므로 전화기의 특성을 갖는 엘리먼트들은 제외하였으며, 본 논문에서 제시하는 VUI 환경에 필요한 엘리먼트들을 세가지 시나리오를 바탕으로 추출하였다.

[표4-1]. 본 논문에서 제시하는 엘리먼트 추출

form, initial, field, option, menu, choice, subdialog, enumerate grammar, goto, return, exit, return, catch, error, help, noinput, nomatch, throw, filled, if, else, elseif, option, audio, block, break, div, emp, param, prompt, pros, record, reutrn, sayas, script, link, value, clear, var, assign, meta

그리고, 추출한 엘리먼트들을 바탕으로 현재 사용되어지는 응용프로그램의 GUI환경의 메뉴구성을 어떻게 대체 할 수 있는지 알아보기로 하자. 일반적인 응용프로그램들의 실행 및 메뉴구성, 작동방법 등은 크게 세가지 방법으로 작업을 수행을 한다.

첫째, 응용프로그램의 실행

둘째, 풀다운(Pull Down) 메뉴의 선택

셋째, 도구모음 메뉴의 선택/ 단축키에 의한 메뉴 선택

위의 세가지 방법을 VoiceXML을 통해 수행할 경우 다음과 같은 사항을 고려하여 VoiceXML 문서를 만들어야 할 것이다.

첫번째 응용프로그램 실행은, 대부분 시작프로그램에 등록되어있으므로 이를 명령어 관리기가 검색하여 VoiceXML 문서를 만들어낸다. 이때 필요한 것은 응용프로그램을 실행시키기 위한 프로그램의 이름을 결정하는 것이며, 이때 사용자가 알기 쉽고 음성인식기가 정확하게 인식하기 위해서는 단어의 발음이 명확히 구분되는 것이 좋다. 또한 해당 응용프로그램의 메뉴들을 파악하여 이를 VoiceXML 문서상에서 지원할 수 있도록 해야 한다. 일반적인 메뉴로는 파일, 편집, 보기 등 이 있으며,

프로그램에 맞는 메뉴구성들이 있는데 이를 인식하여 VoiceXML 문서를 만들거나 사용자가 직접 입력할 수 있도록 해야한다.

두번째, 풀다운 메뉴는 계층적 구조를 갖고 있으므로 가장 일반적인 VoiceXML 문서를 만들 수 있다. 이때 주로 사용되는 것은 분기 엘리먼트를 이용하여 VoiceXML을 이용한 음성인터페이스를 제작한다.

세번째 도구모음/ 단축키의 경우 풀다운 메뉴의 첫번째 품과 겹치지 않는 범위 내에서 작성되어야 한다. 만일 같은 음성 명령으로 작동하게 할 경우 현재 로드 되어있는 VoiceXML 문서 내에서 명령처리를 수행하기 때문에 사용자가 원치 않는 명령이 수행될 수 있거나, 여러 프로그램을 동시에 수행 할 경우 사용자의 실수로 원치 않는 프로그램에게 명령이 하달 될 수 있다.

5. 결론

본 연구는 VoiceXML 문서를 기존의 VUI환경에 접목하여 보다 유연한 VUI환경을 제시하였다. 현재 음성인식 기술은 많은 발전을 이룬 것은 사실이나 이를 적극 활용할 응용기술들이 적었으며, 이때 가장문제 되는 것 중 하나가 바로 원활한 음성 사용자 인터페이스의 부재였다. 본 논문은 이러한 상황을 인지하고 이를 현재 차세대 음성서비스 언어인 VoiceXML을 통해 모색해보고자 하였다.

본 연구를 통해 VoiceXML을 이용하여 음성 인터페이스를 활용하고자 할 때 실제 음성인터페이스 환경을 구축시 필요하지 않은 VoiceXML 엘리먼트가 있음을 알았으며, 앞으로 많은 응용프로그램에 접목하여 작성할 때에 파생되는 음성 인터페이스 시나리오를 구분하여 음성 인터페이스 개발시 지침이 될 수 있으리라 생각된다. 또한 개발자뿐만 아니라 현재 증가하고 있는 XML사용 유저들 까지 끌어들여 폭넓은 사용자 층을 끌어들일 수 있으리라 생각되며 본 논문에서 연구한 VoiceXML을 이용한 VUI를 실제 응용프로그램에 접목하여 VoiceXML의 효용성을 확인해볼 수 있을 것이다.

참고자료/ 사이트

- [1] Voice XML 포럼, <http://www.voicexml.org>
- [2] 음성인식, 디지털타임즈, 2000년 7월 4일
- [3] 홍용택, "XML과 VoiceXML의 효율적인 통합방안 제시" 군산대학교 교육대학원 석사논문, 2001. 10
- [4] Lee Garber, "The Challenges that XML Faces", October 2001