

한국어 음성합성과 인식을 이용한 웹 브라우저 설계 및 구현

조경환

*최훈일

*조철환

*장영건

*청주대학교 전산정보공학과

{loonie, choih, garfield, ygjang}@chongju.ac.kr

Design and Implementation of Web browser Using Voice synthesis & Recognition for Korean language

Kyoung-hwan, Cho, *Hoon-il, Choi,* *Chul-hwan, Cho,* *Young-gun, Jang
*Dept. of Computer Information Eng., Chongju University

요 약

인터넷의 중요성이 증가함에 따라, 웹 브라우저에 음성 인터페이스를 추가하는 연구와 개발이 이루어지고 있다. 그러나, 아직까지, 기존의 모든 웹 문서가 HTML로 작성되어 있어, 효과적인 음성 인터페이스를 하기에는 많은 어려움이 있으며, 음성이 느린 출력 매체이므로 사용자가 빠르게 인지할 수 있는 방안이 연구되어야 한다.

본 논문에서는 사용자의 웹 액세스를 높이기 위하여, 웹 브라우저에 연결되는 웹 문서에서, 각각의 객체를 추출한 후, 사용자가 그 객체에 바로 액세스를 하거나 한국어 음성으로 그 정보를 알 수 있는 방법을 사용하여, 음성으로 제어할 수 있는 한국어 음성 웹 브라우저를 설계하고 구현하였다.

음성합성과 인식을 사용하여 브라우저를 제어하기 때문에, 노약자나 어린이 또는 시각장애인들이 쉽게 웹 서핑을 할 수 있도록 도와줄 수 있고, 또한 현재 사용되고 있는 웹 문서에서의 객체추출을 사용하기 때문에 특별히 문서의 변환이 필요 없는 장점이 있다.

1. 서론

인터넷이 보급되면서, 방대한 정보로부터 구한 정보를 효율적이면서 쾌적하게 검색할 수 있는 시스템이 요구되고 있다. WWW의 등장으로 인터넷이 제공하던 단순한 텍스트위주의 서비스에서 멀티미디어 서비스로의 전환점을 가져왔으며, 전세계의 다종다양한 정보에 액세스하는 것이 가능해 졌다. 이렇게 인터넷의 중요성이 증가함에 따라, 웹 정보에 접근할 수 있는 웹 브라우저도 많은 기능이 추가되고 좀 더 사용자에게 편리하게 인터페이스 되어 가고 있다.

현재, 음성합성과 음성인식의 기술 발전으로 인해 웹 브라우저에서도 이를 이용하여, 이전에 사용하던 키보드와 마우스 인터페이스를 대신하고 있는 상태이다. 보다 편리한 인터페이스를 요구하는 어린이나 노약자 층, 시각장애인들의 경우는 기존의 웹 브라우저를 쉽게 이용하기가 어려운 실정이다.

본 연구에서는 기존의 HTML을 분석하여 추출된 내용을 음성으로 변환시키고, 웹 브라우저의 제어를 위하여 음성인식을 사용하는 음성 웹 브라우저를 설계와 구현에 관한 것이다.

음성인식의 경우, 음성 웹 브라우저^[1-2]에 사용되는 기능을 단어레벨로 정의하고, 입력음성과 입력된 단어의 매칭여부를 이용하여 웹 브라우저를 제어하며, 웹 문서에서의 정보추출을 이용하여, 폼에 대한 구조 설명, 링크 주소 리스트, 그림과 같은 멀티미디어에 대한 설명 등을 출력하는, 음성 웹 브라우저의 인터페이스에 대한 방안을 제시한다.

휴먼 인터페이스로서 음성을 이용한 경우, 음성이 느린 출력매체이기 때문에, 객체에 대한 정보를 출력한 후 직접 그 객체에 액세스하는 방식으로 사용하였다. 이러한 방식에서는 음

성을 출력하는 타이밍, 출력한 내용의 적절성, 출력된 음성의 자연성이 중요하며, 시각장애인을 대상으로 하는 IBM사의 '홈페이지 리더'가 이 방식을 사용한다.

외국의 경우 화면 읽기 프로그램을 이용하여 텔넷, 전자메일, FTP를 불편 없이 사용할 수 있으며, 전용 웹 브라우저의 개발로 WWW까지도 이용할 수 있으나, 우리나라의 경우에는 이러한 부분에서 그 연구가 미미한 실정이다. 음성 웹 브라우저는 음성합성과 인식을 사용하여, 웹 브라우저를 제어하기 때문에, 노약자나 어린이 또는 시각장애인들이 웹 서핑을 쉽게 할 수 있도록 도와줄 수 있을 것이며, 또한 현재 사용되고 있는 웹 문서에서의 객체추출을 사용하기 때문에 특별히 문서의 변환이 필요 없는 장점이 있다.

2. 음성 웹 브라우저의 현황

음성 웹 브라우저는 아직까지 그 연구가 미미한 실정이었으며, 현재 개발 연구가 진행되고 있는 상태이다.

국내의 경우에는 1994년 자바애플릿을 이용하여 음성인식을 한 후, 키보드나 마우스 클릭 없이 멀티미디어 정보를 제공하고, 전송하는 시스템 연구^[3]가 진행되었고, 1998년에는 접자출력기와 음성합성기를 이용한 연구^[4]와 웹브라우저 상에서 음성인식을 이용한 검색시스템^[5] 등이 연구되었다.

최근에는 시각장애인을 위한 웹 브라우저 '웹 아이'^[6-7]가 개발되었으며, 마우스와 단축키를 이용하여 웹 브라우저를 제어한다.

외국의 경우 대표적으로 IBM사의 '홈페이지 리더'^[8]가 있으

며, 시각장애인을 대상으로 하여 개발되어 시각 장애인들이 인터넷을 자유롭게 이용할 수 있게 되어 있으며, 10개의 숫자 패드를 사용하여 편의성을 높였으며 신속이동, 북마크, 통합전자우편 등의 기능도 제공한다.

현재, 한국어를 대상으로 하는 음성인식, 합성 인터페이스 웹 브라우저의 연구는 미미한 실정이다.

3. 음성 인식 및 합성

최근 "음성인식 및 내추럴 자연스러운 대화 기술"로 인해 컴퓨터와의 모든 대화가 마치 옆에 있는 친구와 대화하는 것처럼 자연스럽게 이루어지는 기술이 급속한 속도로 성숙해가고 있다. 외국 음성인식은 화자적용 방식을 통해 95%가 넘는 인식률을 자랑하고 있다. 화자적용은 화자독립의 경우 사용자의 목소리를 녹음해 훈련시키는 과정이 없다는 장점과 화자중속의 경우 사용자의 발음하는 습관과 억양 등을 데이터베이스화해 감기가 걸려 목소리가 변하지 않는 이상 거의 완벽한 인식률을 보인다. 장점만을 수용한 기술이다. 때문에 화자적용을 채택한 받아쓰기 프로그램에서는 어떤 사용자든지 데이터베이스 안에 저장한 단어를 말하기만 하면 인식하지만, 사용자의 음성 데이터베이스를 만드는 훈련 과정을 거치도록 해서 인식률을 높일 수 있고, 아울러 받아쓰기를 많이 할수록 인식률이 높아지는 특징을 갖는다. 그러나 현재의 음성 인식 기술은 아직까지 다음과 같은 한계⁹⁾를 가지고 있다. 본 연구에서는 화자독립의 단어레벨의 음성인식기를 사용하였고, 음성인식기는 음소를 대상으로 DB를 구성하기 때문에 적은 용량의 DB를 사용한다.

- 동시에 인식 불가능
- 방대한 음성 DB 필요
- 자연언어 처리 기술의 한계
- 환경(잡음, 마이크 성능)에 제약
- 비슷한 발음 인식 문제

4. 웹 문서

웹 문서는 HTML(Hyper Text Markup Language)로서 웹 브라우저에서 하이퍼텍스트 기능을 가지고 있다. HTML문서의 모양과 행동양식을 정해주는 하나의 명령어로 태그를 사용하며, 브라우저에게 화면에 나타나는 방식을 설명해 주게 된다. 태그는 보통 "< >"로 둘러 쌓여 다른 문장들과 구별되며, 객체로 표현할 수 있다. 보통 HTML 태그는 <태그>와 </태그>으로 짝을 이뤄 사용되고, 이 둘 사이에 문장이 들어가게 된다. 시작 태그와 마지막 태그로 객체를 추출할 수 있으며 본 연구에서는 각 객체를 추출하여 사용하는 방법을 사용하였다. 표 1은 HTML 문서에서 사용되는 3가지 형식의 태그이다.

표 1. 태그구조

<ul style="list-style-type: none"> • <태그>문장</태그> • <태그 속성=인자> 문장 </태그> • <태그>
--

5. 시스템 설계

본 연구에서 구현된 시스템은 문장 합성기로 LGTTS와 음성인식기로 단어레벨의 VoiceLink를 사용하였고 윈도우즈 2000 환경 내에서 Visual Basic과 VC++를 사용하여 구현하였다. 단어레벨의 음성인식기를 사용하여 자연어는 인식할 수 없기 때

문에 웹 브라우저 제어에 필요한 수십 가지의 단어를 대상으로 하였다. 웹 브라우저는 마이크로소프트사에서 기본적으로 제공하는 웹 브라우저 컨트롤을 사용하여 제작하고 메일모듈은 MAPI를 사용하였다. 그림 1은 시스템 전체 구성도를 나타낸 것이다.

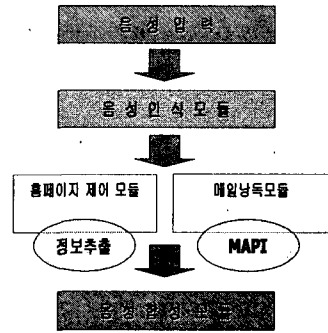


그림 1. 시스템 구성도

5-1. 내부 시스템 수행 과정

그림 2는 프로그램의 수행절차를 나타낸 것이며, 사용자의 입력을 음성인식으로 한다. 음성인식의 경우, 복잡한 음성은 인식률의 저하를 가져오기 때문에, 수십 가지의 입력된 데이터를 가지고 사용한다. 예를 들면, "앞으로 이동", "뒤로 이동", "홈페이지로 이동", "메일읽기", "출거찾기", "이동", "추가", "삭제" 등과 같은 필수적인 기능과 링크정보와 같은 수십 가지의 추가 정보를 대상으로 음성인식을 수행한다.

음성인식을 통하여, 그 기능을 수행하고 해당 홈페이지로 접속을 하게 되면, 그 홈페이지의 HTML문의 내용에서 각각의 객체를 추출하고 그 객체의 정보를 DB에 저장한다. 음성명령으로 링크주소나, 홈페이지의 내용, 타이틀, 테이블, 그림 등의 정보를 음성으로 출력하게 된다.

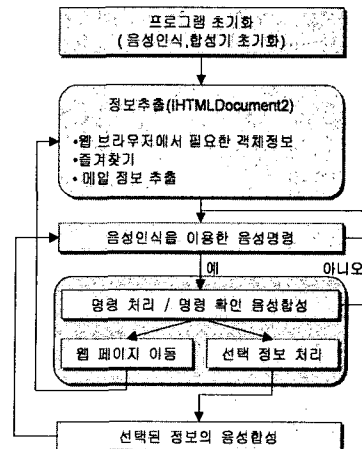


그림 2. 프로그램 수행 절차

5-2. 음성 인터페이스

현재의 음성 인식 기술이 한계를 지니고 있기 때문에 인식률을 높이기 위해 수십 가지의 기본 명령을 사용하고, 추가적으로 제공되는 정보를 입력할 수 있게 하였다. 또, 입력단어에 약

간의 융통성을 주어 일반적으로 같은 성격의 명령에 대해서 같은 실행을 하게 한다. 모든 인식 단어에서는 일관성 있는 단어를 이용하여 사용하기 편리하게 하고, 객체 정보는 트리 구조와 같은 개념으로 저장하여 상황에 따라 직접 이동할 수 있게 하였다.⁹⁻¹⁰⁾

브라우저는 홈페이지를 이동하게 되면, 홈페이지 타이틀을 읽어주고, 홈페이지의 HTML문서를 해석하여 그 객체에 대한 정보를 사용자에게 들려주게 되며, 사용자는 각 객체를 선택하고 객체에 대한 정보를 읽거나 전체 내용을 읽을 수 있다.

즐거찾기 경우는, 익스플로러에 등록되어 있는 즐겨찾기 정보를 사용자에게 들려주고, 사용자는 바로 해당 사이트로 이동할 수 있게 한다.

메일의 경우는 아웃룩 익스프레스에 등록된 계정의 POP3를 이용하여 메일의 수, 제목, 송신자, 보낸 일자, 및 내용을 음성으로 읽어주고, 삭제할 수 있다.

5-3. DB 구조

표 2는 객체정보와 메일, 즐겨찾기 등의 정보를 저장하는 데이터 구조를 나타낸 것이다. ID는 각 정보에서의 인식번호를 나타낸 것이며, sObj는 객체에 대한 속성을 나타낸 것이다. 즐겨찾기와 링크정보는 'A', 그림은 'IMG'와 같이 나타내었다. sText는 각 정보에 대한 보충설명으로 사용자에게 들려주기 위한 데이터이다.

표 2. 사용된 DB구조

객체 정보		메일 정보		즐거찾기 정보	
ID	번호	ID	번호	ID	번호
sObj	속성	sObj	속성	sObj	속성
sName	이름	sSub	제목	sSub	타이틀
sType	타입	sDate	날짜	sLink	링크주소
sLink	링크주소	sSender	보낸사람	sText	보충설명
sValue	상태값	sText	내용		
sText	내용				
sDes	보충설명				

5-4. 구현

그림 3은 구현된 화면을 나타낸 것이다. 객체정보와 메일정보, 즐겨찾기 정보는 트리 구조로 나타내었으며, 명령어 입력창과 음성어 입력되는 상황을 화면으로 볼 수 있다. 음성명령에 대한 응답으로 각 명령의 출력 정보를 음성으로 출력, 화면으로 표시하여 사용자가 편하게 사용할 수 있게 하였다.

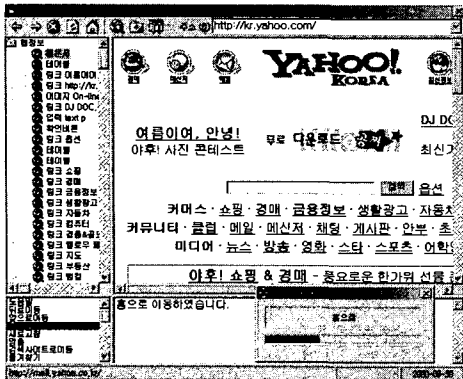


그림 3. 음성 웹 브라우저 화면

표 3. 음성 브라우저 비교

기능	웹아이	홈페이지 리더	구현된 브라우저
입력	마우스/단축키	마우스/키패드	키보드/음성
출력	GUI/음성	문자/음성	GUI/문자/음성
출력정보	(영/한국어) 단어/줄단위 화면읽기	(영어) 객체단위 읽기	(영/한국어) 객체단위 읽기
라이브러리 환경	익스플로러	넷스케이프	익스플로러

표 3은 대표적인 음성 브라우저를 대상으로 입출력 인터페이스 면에서 평가한 표이다.

6. 결론 및 향후 방향

본 연구에서는 웹 브라우저에 전달된 웹 문서에서 각각의 객체를 추출하여 음성합성과 인식을 이용하여 한국어 음성 웹 브라우저를 구현하였다. 구현된 브라우저는 음성으로 웹 브라우저를 제어할 수 있으며, 즐겨찾기, 웹 서핑, 웹 페이지 및 객체 정보 읽기, 메일읽기 등을 할 수 있다.

객체를 이용한 방법은 느린 출력메체인 음성에서 좀 더 빠르게 사용자가 인식하고, 원하는 정보를 바로 액세스할 수 있다는 장점이 있으며, 웹 페이지에 있는 입력박스에 값을 넣어줄 수 있다. 그러나 현재 페이지에 대한 정보를 사용자가 알아야 하기 때문에 전체 페이지 정보를 선택적으로 들어야 하며, 이미지나 도표, 테이블 등의 정보는 주석 설명을 이용하기 때문에 이러한 정보가 웹 문서에 없다면 그 정보를 사용자가 음성으로 알 수 없는 단점이 있다.

음성 웹 브라우저는 어린이나 노약자 또는 시각장애자들이 좀 더 편리하게 사용할 수 있을 것이며, 향후에는 키보드 숫자 키패드와 음성인식을 동시에 사용하고, 각 객체에 대한 정보를 자연스럽고 빠르게 사용자가 인지할 수 있도록 하는 방안이 연구되어야 할 것이다.

7. 참고 문헌

- [1] Michael Robin, Jim Larson, "Voice Browsers - An introduction and glossary for the requirements drafts", <http://www.w3.org/TR/1999/WD-voice-intro-19991223/>, 1999. 12
- [2] Michael Robin, Jim Larson, "Model Architecture for Voice Browser Systems", <http://www.w3.org/TR/1999/WD-voice-architecture-19991223/>, 1999. 12
- [3] 김기원, 정지원, 송정길, "음성인식 웹브라우저에 관한 연구", 한국정보처리학회 '96춘계 논문집, 1996. 4
- [4] 박찬용, 장병태, "시각장애자를 위한 인터넷 웹 브라우저 개발", 대한전자공학회지, 제11권 2호, pp829-832, 1998
- [5] 이항성, 김희권, "웹 브라우저 상에서 한국어 음성인식을 이용한 정보검색시스템", 음성통신 및 신호처리 워크샵 논문집 제15회, pp35-38, 1998
- [6] 김석일, (주)e-Trek Infodigm, "시각장애인용 인터넷 통합 솔루션 패키지 개발", 충북대학교 보도자료, 2000. 4. 18
- [7] 이승수, 민경석, 주용덕, 김석일, "시각장애인을 위한 인터넷 제27회 솔루션의 구현", 한국정보과학회 제27회 춘계 논문집, 2000. 4
- [8] <http://www.tonggu-gch.ed.seoul.kr/infonews/i4.htm>
- [9] 강준구, "음성 인식 시스템에서 사용자 인터페이스 연구", 한국과학기술원 학위논문, 1994. 12
- [10] 김재갑, "인간-컴퓨터 인터페이스", 한국심리학회 98 동계 연구 세미나 자료, 1998, <http://galaxy.channeli.net/jkap/cogeng/hci98.htm>