

한국어 음성 웹브라우저 설계 및 구현

(Design and Implementation of Korean Voice Web Browser)

장 영 건 [†] 조 경 환 ^{**}
 (Young Gun Jang) (Kyoung Hwan Jo)

요 약 본 논문은 HTML을 분석하여 추출된 내용을 음성으로 변환시키고, 웹 브라우저 제어 및 웹 문서의 내용 지정을 위하여 음성인식을 사용하는 음성 웹 브라우저의 설계와 구현에 관한 것이다.

이 시스템의 특징은 보편적 설계기법을 사용하여 정상인과 시각장애인을 모두 고려하여 설계되었으며, 특히 시각장애인 인터페이스를 위하여 웹 문서의 구조를 쉽게 인식할 수 있는 트리 구조를 도입하였고, 웹 브라우저 상에 태그로 표현된 모든 요소를 추출하고, 각 요소에 따라 출력음성의 속성을 다르게 하여 별도의 요소 안내음성 출력이 없이 요소 특성을 인지할 수 있게 하였다. 이 방식은 음성 특성을 세부적으로 표현할 수 있는 청각 스타일 시트의 사용을 배제한 것으로써 웹 문서 작성자가 특별한 추가적 노력 없이, 사용자가 문서내의 구조와 요소들의 특징을 쉽게 인식할 수 있는 장점이 있다.

Abstract This paper is addressed to a design and implementation of Korean voice web browser using voice technologies for controlling web browser and selecting contents in the web document, and converting them to voice after HTML analysis.

Main feature of this web browser is universal design which considers both of normal person and visual disabled, allows multi-modal interface. As voice interface for visual disabled, it supports tree structure which allows to recognize web document structure easily by only voice guidance regardless of frame usage, can handle all elements described as tag in the web document, identify them as predefined different voice property according to element property. This method gets rid of additional guidance voice for element property without audio style sheet or additional programming effort.

1. 서 론

정보의 바다인 인터넷이 활성화되면서, 방대한 정보를 효율적이고, 쾌적하게 검색할 수 있는 시스템이 요구되고 있다. 이에 따라, 웹 정보에 접근할 수 있는 웹 브라우저는 좀 더 많은 기능과 편리한 사용자 인터페이스가 요구되고 있다. 최근 음성합성 및 인식기술의 발전으로 음성기술을 이용한 웹 브라우저 개발 연구가 가속화되고 있으며, W3C[1]에서는 그 중요성을 인식하여 1998년에 음성 브라우저에 대한 Note가 발표되었고, 지속적으로 갱신되고 있다.

국내에서는 1994년 자바 애플릿을 이용한 음성인식 멀티미디어 정보시스템[2], 1998년 전자출력기와 음성합

성기를 이용한 연구[3], 웹 브라우저 상에서 음성인식을 이용한 검색시스템에 대한 연구[4]가 있었으나 기초적인 기능구현에 관한 연구이다. 본격적인 음성 웹 브라우저는 1999년에 발표한 EVANS[5]와 2000년 D&M Technology에서 발표한 보이스 익스플로러[8]이며, 가변어휘 단어 인식기를 사용하여 웹 브라우저의 메뉴와 링크정보를 통해 웹 검색을 할 수 있으나, 음성인식과 합성에 초점을 맞추어 시각인터페이스에서 제공되는 시각적 표현에 의하여 제공되는 정보가 유실되는 문제점이 있다. 음성합성기능을 이용한 시각장애이용 웹 브라우저 웹 아이[6-7]는 마우스와 단축키를 이용하며, 단축키의 사용빈도가 높아 맹아학교 등 특수교육을 받은 사용자가 편리하게 사용할 수 있다.

외국의 경우에는 1997년 10월에 발표된 IBM사의 홈페이지 리더[9]가 있으며, 시각장애인을 대상으로 개발되어 시각 장애인들이 10개의 숫자 패드를 사용하여 인터넷을 이용할 수 있다. 2000년 3월에는 일본에서 GUI

[†] 경 회 원 : 청주대학교 컴퓨터정보공학과 교수
 ygjang@chongju.ac.kr

^{**} 비 회 원 : 삼양데이터시스템(주) R&D팀 연구원
 loonie@hanmail.net

논문접수 : 2001년 1월 10일

심사완료 : 2001년 8월 9일

에 의존하지 않고 키보드와 음성 출력만을 이용한 키보드 연습 환경을 이용하여, 시각장애자들도 인터넷이나 컴퓨터를 사용할 수 있는 시스템이 개발되었다[10]. IBM사의 ViaVoice는 음성인식 및 합성 통합 패키지로, 받아쓰기 지원의 음성인식, 합성뿐 아니라 명령어 입력 레벨의 윈도우즈 제어 및 음성 웹 브라우저기능이 있으나 사용 전에 너무 많은 사용자 음성 정보 입력이 요구되어 시각장애인이 사용하기는 어렵다.

전술한 거의 모든 경우가 특정한 인터페이스만을 고려하며 일반인과 시각장애인, 지체장애인 모두를 고려하는 보편적 설계 개념을 도입하고 있지 않다. 특히 시각장애인의 경우에는 음성출력인터페이스만으로 시각인터페이스에서 제공하는 정보를 제공받을 수 있어야 한다. 웹 문서는 여러 정보(링크주소, ALT 등)가 하나의 GUI 요소로 표현되는 시각인터페이스를 구간으로 작성되었기 때문에 음성 인터페이스를 하기에는 불필요한 내용과 많은 정보를 갖고 있는 경우가 많다. 따라서 웹 문서에서 표현된 태그로, 요소의 효율적 추출과 출력정보의 특징을 보다 빠르게 사용자가 인지할 수 있는 방안과 시각인터페이스에서 제공하는 문서구조 인식의 용이성을 음성 인터페이스에서도 제공하는 방안이 연구되어야 한다.

본 논문은 HTML을 분석하여 추출된 내용을 음성으로 변환시키고, 웹 브라우저 제어 및 웹 문서의 내용 지정을 위하여 음성인식을 사용하는 음성 웹 브라우저의 설계와 구현에 관한 것이다.

주요한 문제로서, 사용자 인터페이스로 음성, 키보드, 마우스의 조작을 허용하는 다중모드를 지원하는 문제와 MS사의 웹 브라우저 Explorer의 기능을 확장하여 한국어 음성인식 및 합성을 하는 문제, 그리고 GUI의 웹 문서의 구조와 정보의 특징을 음성을 이용하여 사용자가 쉽게 인식할 수 있는 인터페이스의 편의성에 대한 문제를 다룬 것이다.

구현할 음성 웹 브라우저는 한국어를 대상으로 하고, 단어레벨의 가변어휘 음성인식기를 사용하였다. 본 논문에서는 태그를 이용하여 문서 구조를 트리 구조로 변환하여 음성으로 안내하는 방식과 인식된 문서 요소에 따라 출력음성의 속성(남/여, 피치, 떨림 정도 등)을 변화시켜 웹 문서상의 요소들의 특징을 표현하는 새로운 방식을 제시하였다.

2. 음성 브라우저와 장애인 인터페이스

음성합성 및 인식 기술의 발전으로 웹에서 음성 인터

페이스가 이루어지고 있으며, 시각장애자를 주 대상으로 음성 명령으로 웹 서핑을 하고, 선택된 웹 문서를 읽어 주는 기능에 이용하고 있다. W3C의 음성브라우저 작업 그룹에서, 2000년 8월 8일에 음성 상호작용을 이용하여 웹에 접근하도록 하는 규격을 개발하는 것을 목적으로 하는 음성합성 제어를 위한 XML 마크업 언어 규격인 Working Draft로 발표되었다[13].

웹 브라우저 측면에서, 음성기술은 음성합성기반과 음성기반의 브라우저로 구분할 수 있으며, 음성합성기반 브라우저는 키보드나 마우스, 그리고 음성합성을 이용하고, 음성 브라우저는 음성인식과 음성합성을 모두 이용한다.

2.1 음성 브라우저의 모델 구조

음성 브라우저는 전화와 같이 화면이 없는 장치를 통하여 웹 액세스를 하는 수단으로써 제안되었다[11]. W3C의 음성 브라우저 작업 그룹에서는 전형적인 음성 브라우저 응용을 위하여 [그림 1]과 같은 모델구조를 개발하였다. 이 모델구조는 전형적인 시스템의 주요 요소의 가능한 배열을 보여주며, 동기화된 다중 모드 입력과 다중 모드 출력을 지원하도록 설계되었다.

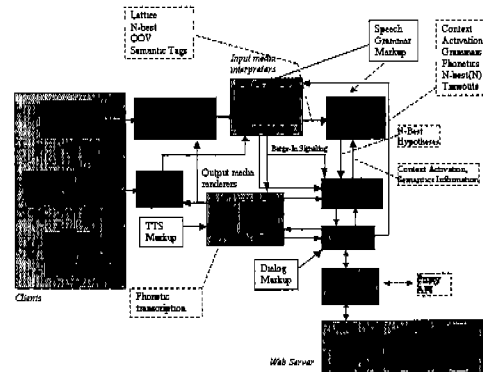


그림 1 W3C의 음성 브라우저 모델 구조

2.2 음성 브라우저의 기능

W3C에서 1998년에 음성 브라우저가 가져야 할 기능에 대하여 기술하였으며, 브라우저는 음성입력과 합성, 그리고 미리 녹음된 음성을 이용하여 음성 출력을 하며, 키보드와 음성 인식의 조합을 입력 수단으로 한다[12]. 이 기술은 전화기로부터의 웹 브라우저를 실현하는데 쓰일 수 있다., W3C는 2000년 9월에 개정된 안을 발표한 바 있으며, 1) 다른 미디어, 2) 이동, 3) 오퍼레이터와 모호성, 4) 청각 스타일 시트, 5) 액세스 키, 6)

TTS(Text to Speech), 7) 음성파일, 8) 음성인식문법 등에 대하여 상세히 규격을 제시하고 있다.

2.3 시각장애인 인터페이스

웹 정보를 기술하는 HTML코드는 기본적으로 시각적 인터페이스를 전제로 만들어졌다. 오늘날 웹 저자는 다양한 시각적 효과를 사용하여 가능한 한 많은 정보를 한 페이지에 표현하고자 한다. 이 정보는 시각적으로 그룹화되어 조각나 있다. 이러한 형식은 비 시각적 웹 액세스를 더욱 어렵게 한다. 시각장애인 사용자는 음성 브라우저를 사용하여 웹 내용을 태그 순서대로 읽지만, 웹 페이지의 내용이 어떠한 구조적 특징을 갖고, 내용적 특징이 무엇인지를 이해하기 위해서는 접속하고 있는 웹 페이지의 전체 내용을 순서적으로 들어보고, 추상화하여야 한다. 이러한 방식은 순차 파일을 대상으로 필요한 정보를 찾는 방식과 유사하며, 매우 비효율적 정보 검색 방식이다. 이러한 문제를 해결하고자 최근 IBM에서는 정보의 중계자로서 프록시 서버를 사용하여 탐색하는 웹 페이지에 주석을 붙여서 사용자에게 되돌려주는 시스템을 제안하고 있다[16,17].

본 논문에서는 이 문제점을 해소하는 방식으로써 접속하고 있는 웹 페이지의 HTML코드의 태그를 해석하여, 각 요소들을 부모 관계로서 표현하는 트리 구조로 변환하여, 트리 구조를 따라 웹 내용을 액세스하는 방식을 제안하고, 이를 구현하였다. 이 방식은 웹 문서의 트리 구조가 내용적 특징의 구조를 표현한다는 전제에서 출발하며, 대부분의 웹 저자들이 이 전제에 따라 문서를 작성하고 있다. 이러한 이동 방식은 웹 문서를 대상으로 하는 음성 브라우저에서 구현된 바가 없고, 특히 시각장애인이 웹 문서의 내용을 액세스할 때 정보 검색을 보다 효율적으로 지원하는 특징이다.

시각적 인터페이스에서 제공되는 특징 중의 하나는 색상과 형태적 특징을 사용한다는 것이다. 보통 링크의 경우 푸른색 색상을 사용하며, 링크를 마우스로 지정하면, 포인터의 형태가 변하여, 지정한 것이 링크임을 시각적으로 표현하여 정보를 전달한다. 링크 이외에도 표, 그림, 객체 등은 그 형태로써 해당 요소의 특징을 인식할 수 있다. 그러나 일반적인 음성 브라우저로써 해당 요소를 접근할 때는 단순히 그 요소의 내용을 음성으로 들려주거나, 요소의 특징을 사용자에게 인식시키기 위해서는 그 요소의 특징을 별도의 음성으로 들려주는 방식을 사용한다. 이 방식은 음성 전달이 문자데이터에 비하여 수 천배의 데이터의 전달이 필요하다는 것을 감안하면, 링크나 표를 많이 갖고 있는 페이지에서는 전체적인 정보 검색시간이 크게 늘어나는 결과를 가져온다. 본 논

문에서는 이러한 요소의 특징을 각 요소에 지정된 음색을 사용하고, 사용자가 이 음색효과를 구별하여 인식하는 방법을 제안하고, 구현하였다.

2.4 보편적 설계와 다중모드

보편적 설계는 가능한 모든 다양한 사용자의 요구를 고려하여 모든 상품, 환경, 통신 등이 설계되어야 한다는 개념에 근거한 전세계적인 운동이다. 음성 브라우저의 사용자는 시각 장애인이 될 수도 있고, 팔이나 손의 기능이 없거나 불편한 지체장애인이 될 수도 있고, 자동차를 운전중인 정상인이 될 수도 있다. 즉 음성 브라우저는 매우 다양한 사용자의 요구를 갖고 있으며, 장애인 사용자의 경우도 장애의 정도와 종류가 매우 다양하다. 이러한 다양한 계층의 사용자가 특별한 별도의 도구에 의존하지 않고, 임의의 목적을 달성할 수 있도록 하자는 것이 보편적 설계의 목적이며, 정보의 취득성에서 중요한 원칙이다[18].

본 논문에서 구현한 음성 웹브라우저는 웹 문서의 구조를 트리로 표현하고, 이 트리 상에서 이동하도록 되어 있다. 웹 문서의 구조와 내용을 파악하기 위한 이 기능을 사용할 때 가장 편리한 사용자 인터페이스는 키패드에서 상하를 표시하는 화살표 키이다. 물론 전혀 키보드를 사용할 수 없는 상태의 중증 지체 장애자의 경우는 음성 명령을 사용해야 할 것이지만, 대부분의 교육을 받은 시각장애인들은 키보드 사용에 익숙하다. 웹 페이지를 검색하다 보면 텍스트박스 인터페이스가 제공되는 경우가 종종 있다. 지체 장애의 경우에는 마우스나 키보드 사용이 불편하지만 전혀 불가능하지 않은 사용자는 이 텍스트박스를 마우스로 클릭하고, 키보드를 사용할 수 있어야 한다. 즉 사용자의 편의성 및 요구가 접속하고 있는 웹 문서의 종류 및 상태, 또는 사용자가 처한 환경에 따라 바뀔 수 있으며, 이에 따라 적절한 인터페이스가 제공되어야 한다. 본 논문에서는 이러한 이유로 구현한 음성 웹 브라우저에 다중 모드 인터페이스 기능을 지원하였다.

3. 한국어 음성 웹 브라우저의 설계

현재 진행되고 있는 W3C의 음성 웹 브라우저의 표준화는 매우 광범위한 영역을 다루고 있다. 그러나 일반적으로 발표된 음성 웹 브라우저는 W3C의 음성 웹 브라우저 영역의 일부분만을 다루고 있고, 지속적으로 발전해 가고 있다.

본 논문에서는 음성검색의 효율적 인터페이스를 제공하기 위하여, 웹 문서의 구조를 트리 구조로 표현하여 음성안내에 이용하는 방안과 텍스트를 처리하는 새로운

방안을 제시한다. 웹 문서의 구조를 트리 구조로 변환하고, 키보드 조작으로 트리 상에서 이동하며, 선택된 내용을 음성합성을 통하여 안내한다. 음성출력의 문서 의미전달을 효율적으로 하기 위하여 복잡하고 부가적인 청각스타일시트를 사용하는 대신에, 문서 객체 요소 특성을 이용하여 요소의 특징을 명시적 방식으로 제시한다. 이 방법은 청각스타일시트 사용에 비하여 다양성과 세밀성을 포기하는 대신 부가적 정보를 생성할 필요가 없는 간편성을 제공하고, 문서요소의 속성을 명시적으로 음성으로 제시하는 방법에 비하여 음성 출력 양을 줄일 수 있다.

따라서 제안하는 방식은 시각 인터페이스와 음성인터페이스를 동시에 갖는 정상인보다는 시각장애인에게 보다 효율적이며, 한국어를 대상으로 하고, 인터페이스의 음향화, 대안화를 위하여 키보드와 마우스를 사용할 수 있는 다중모드를 지원한다.

3.1 모델 구조

제안된 한국어 음성 웹 브라우저의 모델 구조를 [그림 2]에 표시하였다.

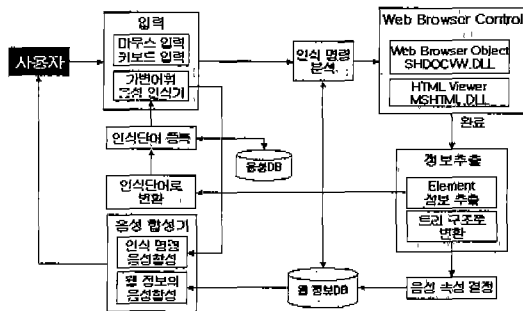


그림 2 한국어 음성 웹 브라우저의 모델 구조

한국어 음성 웹 브라우저의 구조는 크게 입력모듈, 인식명령분석 모듈, 웹 브라우저 모듈, 정보추출 모듈, 인식단어 등록 모듈 및 음성합성 모듈로 구성된다.

음성인식기는 웹 페이지의 이동에 따라 웹 문서의 내용을 음성인식 사전에 등록하여 사용할 수 있는 가변어휘 음성인식기를 사용하고, 음성합성기는 정보추출 모듈에서 생성된 음성속성 값을 이용하기 위해 음성속성(남/여, 피치, 속도 등)을 조절할 수 있는 합성기를 사용한다. 그림 2의 모델 구조에서 각 모듈에 대한 설명은 다음과 같다.

3.1.1 입력

본 연구에서는 다양한 사용자의 욕구충족과 다양성에

대한 대안으로써 다중모드를 허용한다. 사용자는 웹 브라우저 제어를 위하여 마우스, 키보드, 음성입력을 사용하여 웹 브라우저에 대한 기본적인 제어명령과 웹 페이지에서 추출된 명령을 선택할 수 있다. 음성입력의 경우는 키보드나 마우스 동작과 같이 항상 입력이 가능하고, 입력상황과 그 인식 결과를 화면과 음성으로 표시한다.

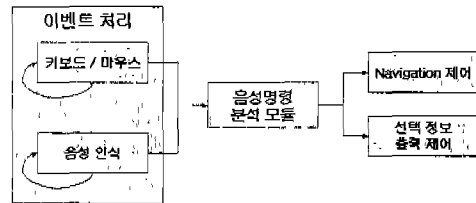


그림 3 다중모드 처리 구조

3.1.2 인식명령 분석

인식명령 분석 모듈은 사용자에 의해 입력된 음성명령, 키보드, 또는 마우스에 의해, 선택된 명령을 분석하여, 웹 브라우저 제어, 선택된 정보에 대한 제어명령을 웹 브라우저 모듈에 전송, 또는 웹 정보 DB에 저장된 정보를 음성합성 모듈로 전송한다.

3.1.3 항해제어

항해제어 모듈은 인식 명령 분석 모듈에서 전송된 제어명령을 통해 웹 브라우저의 이동(홈으로 이동, 이전/다음으로 이동, 기타), 또는 링크 요소의 URL이나 이벤트를 이용하여 웹 브라우저를 제어하게 된다.

3.1.4 정보추출

정보추출 모듈은 웹 브라우저에 웹 페이지가 표시되면, 필요한 요소의 정보를 추출하고, 트리 구조로 표현한다. 요소의 속성에 따라 음성속성을 결정한 후 웹 정보 DB에 저장하는 기능을 담당한다.

3.1.5 인식단어 등록

웹 정보 DB에 저장된 정보를 이용하여 요소의 정보를 이용하여 인식명령을 생성한 후, 음성인식기에 등록한다. 인식단어는 단어와 문자열의 길이를 이용하여 생성한다.

3.1.6 음성합성 모듈

명령 분석 모듈에 의해 수신된 DB의 음성속성 정보를 이용하여 음성합성기의 출력 음성 속성을 설정하고, 텍스트 정보를 이용하여 음성으로 출력한다.

3.2 요소단위 정보추출과 음성속성결정

음성인식 및 합성을 이용한 음성 웹 브라우저의 경우, 효율적인 음성 인식 단어 추출과 사용자의 빠른 인지를

위해 출력할 정보의 양을 줄이는 것이 중요하다[14]. 웹 페이지내의 요소 속성을 이용하여 음성 속성을 결정함으로써, 사용자에게 들려주어야 할 정보의 양을 줄이는 방법을 사용하고, 간소화된 요소정보를 이용하여 음성인식 명령을 추출함으로써, 사용자가 빠르고 편리하고 웹 브라우저를 제어하거나 웹 정보를 인지할 수 있다.

4. 구현

4.1 시스템 구성

본 연구에서 구현한 시스템은 음성 합성기로 LGTTS와 단어레벨 음성인식기로 VoiceLink를 사용하였고, 윈도우즈2000 운영체제 환경 내에서 VC++를 사용하여 구현하였다.

가변어휘 음성인식기를 사용하여 웹 브라우저 제어에 필요한 수십 가지의 기본 제어를 할 수 있는 명령어와 웹 페이지의 변화에 대하여 동적으로 변화하는 수십 개의 동적 명령어를 음성인식 대상으로 하였다. 음성인식과, 합성된 음성의 출력을 동시에 할 수 있게 하기 위하여 Duplex를 지원하는 사운드카드를 사용하고, 웹 브라우저는 마이크로소프트사에서 기본적으로 제공하는 웹 브라우저 컨트롤을 사용하여 제작하였으며, 웹 브라우저 제어는 URL 및 링크요소에 대하여 Navigate2 Method와 Document Object의 Element Click 이벤트를 사용하여 처리하였다.

4.2 데이터베이스 구조

음성 웹 브라우저의 전체 정보를 저장하기 위한 데이터베이스 구조를 표 1에 나타내었으며, 총 7개의 필드로 구성된다.

표 1 웹 정보를 저장하기 위한 DB구조

필드이름	종 류
음성명령	문자열
프레임	0, 1, 2, ...
속성	A, S, IMG, INPUT, OPTION, ...
고유인덱스	숫자 문자열
속성값	링크주소, 요소 값에 대한 문자열 또는 값
음성속성	W+, W0, W-, M-, M0, M-
세부내용	음성합성은 위한 세부 내용

4.2.1 음성명령

음성인식을 위한 명령어는 기본적인 웹 브라우저 제어 명령과 웹 페이지의 변화에 따라 새롭게 만들어지는 추가음성명령이 있다. 추가음성명령은 텍스트 정보나 ALT

정보의 처음 두 단어를 사용하여 음성명령을 만들기 때문에 사용자가 쉽게 암기하거나 사용할 수 있다. 이러한 정보가 없는 경우는 요소의 속성과 고유번호를 이용하여 생성하며, 그림의 경우 ALT정보가 없으면 추가되지 않는다.

표 2 추가 음성 명령 일부

음성명령	기능
링크35	링크35 URL로 선택
[링크타이틀]로 이동	선택요소의 URL로 이동
입력[요소]	입력상자[요소]에 포커스
이미지[요소]	이미지[요소]에 포커스
내용[요소]	선택요소 내용 선택
테이블30	테이블30 요소로 포커스
~	~
[즐거찾기]로 이동	선택된 즐겨찾기 포커스

추가적으로 생성되는 음성 명령의 일부를 표 2에 나타내었다. 링크와 즐겨찾기를 입력하게 되면, 해당 음성아이콘에 초점이 이동하게 되고, "이동", 또는 "확인"과 같은 기본 제어명령에 의해 해당 URL로 이동하게 된다.

4.2.2 프레임 번호

프레임 번호는 현재 웹 브라우저 창에 표시되는 웹 페이지가 프레임으로 구성되어 있을 경우, 프레임에 대한 인덱스 번호를 저장한다. 만약 프레임이 없는 경우는 기본 값으로 '0' 값이 된다.

4.2.3 속성

웹 페이지 정보는 크게 링크주소에 의한 이동정보, 단순히 문자로 표현된 정보, 어떤 정보를 입력하거나 선택할 수 있는 입력정보로 구분할 수 있다. A는 이동할 수 있는 정보를 뜻하며, S는 문자열로 표시되는 정보, INPUT은 입력을 받을 수 있는 텍스트박스, OPTION은 선택할 수 있는 옵션박스 등을 나타낸다. 만약 이미지에 링크가 설정되어 있다면 ALT정보와 링크정보를 사용하며 속성은 A, 세부내용은 ALT정보 또는 '이미지 링크'등의 요소의 특징을 나타내는 이름으로 설정한다.

4.2.4 고유번호

아이디는 웹 페이지 내에서 각 요소를 구분하기 위한 유일한 숫자로 표시한 것이며, 현재 선택된 정보가 웹 페이지의 무엇인지를 나타내는 역할을 하도록 고유번호를 부여한다.

4.2.5 속성값

속성값은 링크의 경우는 링크주소, 입력박스나 옵션박스 등과 같은 경우에는 그 요소의 값을 표시한다.

4.2.6 음성속성

음성속성은 크게 M, W와 -, 0, +의 조합으로 나누어진다. M은 남자 목소리를 나타내며, W는 여자 목소리를 나타내며, -, 0, +은 음성피치를 나타낸다. 일반적으로 입력받을 수 있는 정보에 대해서는 'M', 읽기 정보만 제공하는 경우는 'W'로 설정된다. 예를 들어 Text 정보로 된 문자열의 경우는 'W0'이며, 링크와 같은 이동할 수 있는 정보는 'M0'으로 설정된다.

4.2.7 세부내용

세부내용은 음성합성기를 이용하여 읽어주는 대상이 되며, DB 리스트가 선택이 되면 음성합성기는 선택된 요소의 세부내용을 읽어주게 된다.

4.3 음성처리 흐름도

그림 4는 음성인식 및 합성처리구조를 나타낸 것이며, 웹 페이지 이동에 의해 임의의 홈페이지로 이동하면, 그 홈페이지의 HTML문의 내용에서 각각의 요소를 추출하고 그 요소의 정보와 요소에 대한 음성의 속성을 DB에 저장한다.

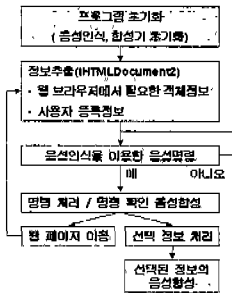


그림 4 음성처리 흐름도

음성인식기는 항상 동작을 하고 있어, 음성이 출력되는 동안에도 음성인식을 할 수 있다. 사용자는 음성명령을 사용하여 링크주소, 홈페이지의 내용, 이미지 및 기타 웹 페이지에서 표시된 요소를 선택하여 음성으로 듣거나 링크요소를 이용하여 이동을 할 수 있다.

4.4 요소단위 정보추출과 음성속성결정

그림 5, 6은 각각 정보추출과 음성속성 결정에 대한 것이다. 웹 브라우저에 표현된 웹 문서에서 각각의 요소 속성을 추출하고 속성에 따라 출력음성의 속성을 결정한다. 출력음성 속성이 결정되면 요소에 대하여 고유번호 및 링크정보나 텍스트정보, 입력 텍스트박스 등의 값을 추출하여 DB에 저장을 한다.

요소의 대표적인 속성으로는 링크 속성을 갖거나 그렇지 않은 경우가 있다. 따라서, 링크 속성을 갖는 요소는 남자 목소리, 그렇지 않은 경우는 여자 목소리의 속성을

갖게 한다. 일반 텍스트 정보로 표시되는 테이블이나 문장, 링크 속성을 갖지 않은 그림의 경우는 여자 목소리이며, 테이블은 피치가 높게, 그림의 경우는 낮게 설정하였다.

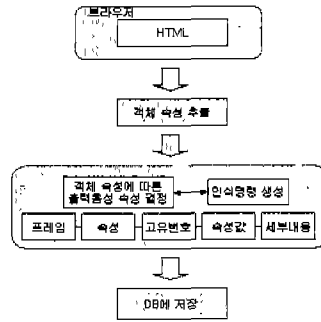


그림 5 정보 추출

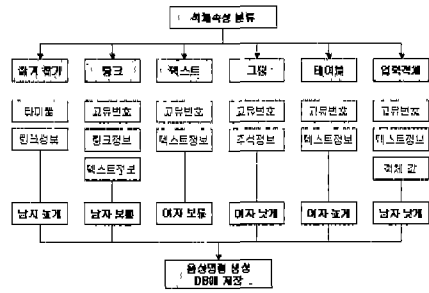


그림 6 음성속성 결정

4.5 트리구조의 표현

윈도우즈 운용체제에서 트리구조를 표현하는 데에 사용되는 상속구조와 메소드인 CTreeCtrl과 CTreeView를 이용하여 HTML문서의 요소와 상속구조의 분석으로부터 얻어진 표현할 요소를 트리 구조로 재구성하였다.

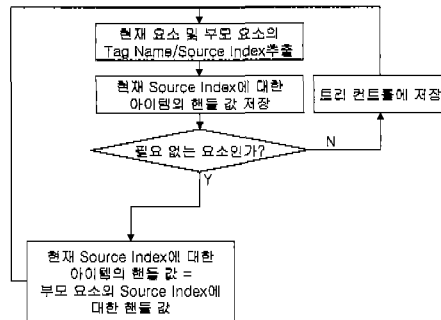


그림 7 트리 추가 알고리즘

트리 생성 알고리즘은 그림 7과 같다. 얻어진 트리는 그림 8의 좌측에 나타나는 바와 같이 시각적으로도 표현되고, 키보드 조작에 의하여 선택되며, 음성으로 안내되어 웹 문서의 구조를 파악하는데 도움을 준다.

5. 시험 및 토의

시험조건 설정과 방법은 시험 목적에 따라 정성적 비교와 정량적 비교로 나눌 수 있으며, 피실험자의 자질에 따라라도 실험결과가 달라질 수 있다. 본 실험에서는 방법론 측면에서 Christion [15]의 방법을 참고하였다.

5.1 출력 음성 비교 시험

웹 페이지의 전체 내용을 음성으로 출력하는 경우, 단일음성으로 출력할 경우는 {요소정보, 내용}과 같이 선택된 요소가 링크인지 이미지인지 등의 부가적인 정보를 출력해 주어야 한다. 그러나 각 요소의 타입에 따라 출력음성의 속성 변화시켜 출력하게 되면, {요소정보}와 같이 부가적인 정보를 출력하지 않아도 되기 때문에 출력시간을 절약할 수 있다. 특히 많은 검색사이트와 같이 링크가 많거나 이미지, 또는 컨트롤이 많은 경우는 많은 시간을 절약할 수 있게 된다. 표 3은 2 종류의 국내·외의 브라우저와 본 논문에서 구현한 웹 브라우저의 음성 출력 측면에서 비교 실험한 내용을 요약하였다. 아후코리아의 경우 메인 페이지는 링크 속성을 가진 요소가 총 212개이다. 따라서 단일음성으로 출력했을 경우 212개×2자("링크")를 내용정보 외에 요소정보를 출력해야 한다.

표 3 음성 웹 브라우저의 비교

비교 대상	웹 아이 1.0 (국내)	홈페이지 리더2.5(국외)	구현시스템
입·출력	단축키/음성	키패드/음성	음성,키보드, 마우스/음성
브라우저 엔진	익스플로러	네스케이프	익스플로러
음성출력	{요소정보, 내용}	{요소정보, 내용}	{내용}
링크	{링크, 텍스트}	{Link, 텍스트} 음성변환지원	{텍스트}
이미지	부분지원	{요소정보, 내용}	{주석}
이미지링크	{링크그림, 주석}	{주석}	{주석}
입력상자	{입력상자}	{Text Input}	{입력상자}

5.2 웹 서핑시험

음성에 의한 웹 브라우저는 작업완료시간에 있어서 전통적인 마우스에 의한 것보다 분명한 시간 지연이 있기

때문에 작업을 하기 위해서는 웹 항해뿐만 아니라 문서의 형식과 내용을 파악하여야 하며, 그러기 위해서는 문서 내용을 음성으로 표현한 안내 음성을 순차적으로 듣고 판단하여야한다. 음성 브라우저는 시각적 효과를 배제하기 위하여 모니터를 끄고 수행하였다. 실제로 음성 브라우저가 시각인터페이스와 더불어 제공되지만, 보조적 수단이 아닌 주 수단으로 사용됐을 때의 유효성을 시험하기 위해서는 강제적으로 시각 인터페이스를 제거하였다.

5.2.1 시험절차

본 실험에서는 동일한 시스템 규격(펜티엄 III 333, 128M RAM, Window2000 OS)과 환경(다수의 컴퓨터가 있는 연구실 환경)에서 8명의 피실험자가 실험에 참여하였다. 이 중 6명은 남성이고, 2명은 여성이다. 피실험자는 모두 컴퓨터정보공학과 학부생과 대학원생으로써 일반적 웹 항해에 친숙하며, 모든 사람이 한국어에 능숙하다.

표 4 검색문제 유형

대상 홈페이지	http://www.ksrd.or.kr
문제	공두리 홈페이지에서 최근 공지사항 중 프로그래머 모집에 관한 사항을 찾을 것

표 5 문제에서 사용된 웹 페이지의 요소 종류와 수

요소	페이지 메인페이지	공지사항	답이 있는 페이지
전체 음성 명령 단어수	8	40	21
링크 단어수	2	32	14

본 실험은 특정사이트를 대상으로 표 4와 같은 검색에 대하여, 음성 웹 브라우저 만으로 사용자가 웹 서핑을 할 수 있는지와 속도도에 의한 작업시간 단축정도를 테스트하였다.

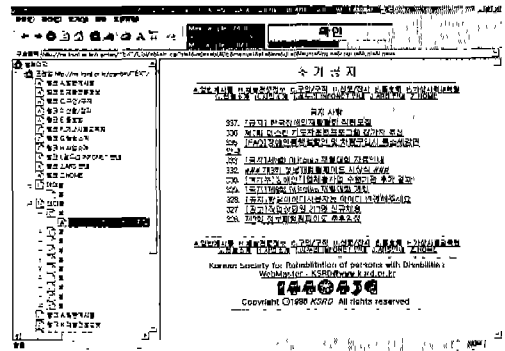


그림 8 공지사항 웹 페이지

5.2.2 결과

3회의 실험을 통하여 측정된 작업완료시간은 표 6과 같다. 실험결과에 의하면 참가자의 브라우저 사용경험과 문서구조의 인식정도에 따라 완료시간 차이가 현저하게 나타남을 알 수 있으며, 초보 사용자의 경우 최초의 작업보다 3번째 사용시에는 평균 18.9%의 작업시간이 단축되었다. 또한 음성만으로 웹 서핑이 모두 가능하였다.

표 6 실험 결과

참가자	음성 브라우저 사용경험	인터넷 사용 경험	제1회	제2회	제3회
화자1	×	○	5:20	4:50	3:30
화자2	2 회	○	3:30	3:40	3:20
화자3	×	○	4:50	4:25	4:05
화자4	×	○	4:40	3:50	3:40
화자5	×	○	4:30	3:57	3:30
화자6	×	○	5:05	5:10	4:32
화자7	×	○	4:32	4:10	4:12
화자8	다수	○	3:10	3:20	2:50

6. 결 론

본 연구에서는 보편적 설계 개념을 도입한 한국어 음성 웹 브라우저를 설계하고, 구현하였다. 정상인과 시각장애인, 지체장애인 모두에게 음성명령에 의한 브라우저 제어, 내용지정과 지정된 문서를 음성으로 읽어주는 기능을 제공하며, 키보드, 마우스를 이용할 수 있다.

시각장애인을 위한 인터페이스로써 음성만으로 웹 문서의 구조 특성을 인식할 수 있는 트리구조 음성안내 및 검색 기능과 추출된 요소의 속성을 이용하여 날짜, 여자, 떨림, 피치를 조절하여 출력함으로써 복잡한 구조의 HTML문서를 다양한 음성으로 출력하게 하였다.

따라서 시각장애자가 정상인이 시각인터페이스를 통하여 인지할 수 있는 웹 페이지 구조, 내용과 요소의 정보를 음성인터페이스를 통하여 동시에 인지할 수 있고, 요소 특성을 별도의 음성으로 안내하지 않아도 되며, 프레임을 지원하기 때문에 화면 전체에서 원하는 정보에 바로 액세스할 수 있다는 장점이 있다.

실험결과, 시각장애인이 쉽게 웹 문서의 구조를 인식할 수 있었고, 출력음색의 차이로써 요소 특성을 인지할 수 있고, 훈련에 의하여 작업시간이 크게 단축될 수 있다는 것을 확인할 수 있었다. 그러나 음성만을 이용한 웹 브라우징은 웹 서핑은 가능하지만 현재 페이지에 대

한 정보와 인식 가능한 단어를 사용자가 알아야 하기 때문에 전체 페이지 정보를 듣거나 선택적으로 들어야 하며, 이미지나 도표, 테이블 등의 정보는 타이틀이나 ALT정보 없이는 그 정보가 어떠한 것인지 알 수 없고, 단어 레벨의 음성인식기를 사용하였기 때문에 메일을 보내거나, 웹 문서내의 텍스트 입력박스에 임의의 값은 넣어줄 수 없다는 제한점을 가지고 있다.

아직까지 웹 페이지에서 구조적으로 표현된 시각인터페이스를 음성으로 인터페이스하기에는 중요 키워드 추출이나 정보의 요약과 같은 어려움이 있기 때문에 보다 많은 연구가 진행되어야 할 것이다. 음성 인터페이스를 더욱 효율적으로 제공하기 위해서는 음성출력의 내용에 따라 출력 중에 사용자가 출력속도를 임의로 조절할 수 있는 기능이 요구된다.

참 고 문 헌

[1] "Voice Browser Activity," <http://www.w3.org/Voice/>

[2] 김기원, 정지원, 송정길, "음성인식 웹 브라우저에 관한 연구", 한국정보처리학회 '96춘계 춘계학술발표 논문집, 1996. 4

[3] 박찬용, 장병태, "시각장애자를 위한 인터넷 웹 브라우저 개발", 대한전자공학회 추계학술발표 논문집, 제21권 2호, pp.829-832, 1998

[4] 이항섭 등, "웹 브라우저 상에서 한국어 음성인식을 이용한 정보검색시스템", 음성통신 및 신호처리 워크샵 논문집, 제15회, pp.35-38, 1998

[5] 이항섭, "가변어휘 단어 인식기를 사용한 음성 명령 웹 브라우저", 한국음향학회지 제18권 2호, pp.48-52, 1999

[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 솔루션의 구현", 한국정보과학회 제27회 춘계 학술대회 논문집, 2000. 4

[8] "보이스 익스플로러", <http://www.dnmtech.co.kr>

[9] "홈페이지 리더", <http://www.austin.ibm.com/sns/hpr.html>

[10] http://japanweb.lycos.co.kr/cgi-bin/unisoft_tg.cgi?URLNAME=www-vox.dj.kit.ac.jp/nishi/work/2000-03-09-wit.html

[11] M. K. Brown, D. A. Dahl, "Model Architecture for Voice Browser Systems," W3C Working Draft, Dec, 1999, <http://www.w3.org/TR/1999/WD-voice-architecture-19991223>

[12] Dave Raggett, Ben-Natan, "Voice Browser," W3C Workshop, Jan, 1998, <http://www.w3.org/TR/1998/NOTE-voice-0128>

- [13] Mark R. Walker, Andrew Hunt, "음성 인터페이스 체계를 위한 음성 합성 마크업 언어 규격", <http://www.w3.org/TR/2000/WD-speech-synthesis-20000808>
- [14] Nicole Yankelovich, Gina-Anne Levow, Matt Marx, "Issues in Speech User Interfaces," CHI '95 Proceedings papers, 1995
- [15] Christian K. et al, "A comparison of voice controlled and mouse controlled web browsing," Proc. ACM ASSETS conference, Nov, 2000.
- [16] Chieko Asakawa et al, "Annotation Based Transcoding for Nonvisual Web Access," Proc. ASSET'00, pp.172-179, Nov., 2000
- [17] Hironobu Takagi et al, "Transcoding Proxy for Nonvisual Access," Proc. ASSET'00, pp.164-171, Nov. 2000
- [18] Eric Bergman et al, "Towards Accessible Human-Computer Interaction," <http://www.sun.com/access/developers/updt.HCI.advance.html>



장 영 건

1980년 인하대학교 전자공학과 학사.
 1979년 ~ 1983년 국방과학연구소 연구원.
 1983년 ~ 1994년 대우중공업 중앙연구소 책임연구원. 1991년 인하대학교 전자공학과 석사(정보공학). 1995년 인하대학교 전자공학과 박사(정보공학). 1995년 ~ 1996년 고등기술연구원 책임연구원. 1996년 ~ 현재 청주대학교 컴퓨터정보공학과 조교수. 관심분야는 HCI, CTI, 음성정보처리물 이용한 웹 프로그래밍, 재활공학



조 경 환

1999년 청주대학교 반도체공학과 학사.
 2001년 청주대학교 전산정보공학과 석사.
 2001년 ~ 현재 삼양데이터시스템(주) R&D팀 연구원.