

음성 웹브라우저에서의 문서구조안내 및 발화속도제어

조철환⁰ 최훈일 연제용 장영건
청주대학교 전산정보공학과
(garfield⁰, choihi, cheyong, ygjang)⁰@chongju.ac.kr

Guidance of Web Document Structure and Voice Firing Rate Control in the Voice Web Browser

Chul-hwan Cho⁰, Hoon-il Choi, Che-yong Yon, Young-gun Jang
Dept. of Computer Information Engineering, Chongju University

요 약

본 논문은 HTML을 분석하여 추출된 내용을 트리로 표현하여 문서구조안내에 이용하고, 웹 문서의 내용의 속도 필요성에 따라 실시간으로 음성 발화속도를 제어할 수 있는 음성 웹 브라우저의 설계와 구현에 관한 것이다.

이 시스템의 특징은 웹 브라우저 상에 태그로 표현된 모든 요소를 추출하고, 이러한 정보를 트리로 표현하고 음성인식으로 정보를 선택하도록 하고, 선택한 정보의 난이도와 필요성에 따라 사용자가 실시간 발화속도제어를 통하여 정보를 쉽게 알 수 있도록 했다. 이 방식은 문서의 내용에 따른 구조를 쉽게 인식하여 사용자가 빠른 시간 내에 필요한 정보를 수집할 수 있고, 문서가 발음되는 것을 청취하여 문서의 필요성을 인식하고, 속도 필요성에 따라 실시간으로 낭독 속도를 제어할 수 있는 장점이 있다.

1. 서 론

인터넷으로 정보의 배포와 수집의 보편적 수단으로 정착되고, 일상생활수단으로서 쓰이고 있다. 그러나 장애인, 특히 시각장애인의 경우에는 인터넷 접근수단인 브라우저가 시각적 인터페이스만을 제공하기 있기 때문에 이러한 보편적 정보 수단의 거의 사용할 수 없으며, 따라서 정상인과 시각장애인 사이의 정보 불평등이 심화되고 있다.

이에 따라, 웹 정보에 접근할 수 있는 웹브라우저는 좀 더 많은 기능과 편리한 사용자 인터페이스가 요구되고 있다. 최근 음성합성 및 인식기술의 발전으로 음성기술을 이용한 다양한 분야에서 사용되고 있으며, 개발 연구가 가속화되고 있다.

웹 문서는 여러 정보(링크주소, ALT 등)가 하나의 GUI 요소로 표현되는 시각인터페이스를 단간으로 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 방안, 시각인터페이스에서 제공하는 문서구조 인식의 용이성과 사용자의 필요성에 따라 문서내용을 한 화면에 유지시키는 기능을 음성 인터페이스에서도 제공하는 방안이 연구되어야 한다.

구현할 음성 웹브라우저는 한국어를 대상으로 하고, 단어레벨의 가변어휘 음성인식기를 사용하였다. 본 논문에서는 HTML을 분석하여 추출된 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 방식과 실시간 발화속도제어를 통해 사용자의 편의성을 고려하였다.

2. 음성 웹브라우저와 장애인 인터페이스

음성합성 및 인식 기술의 발전으로 웹에서 음성 인터페이스가 이루어지고 있으며, 시각장애자를 주 대상으로 음성 명령으

로 웹 서핑을 하고, 선택된 웹 문서를 읽어주는 기능에 이용하고 있다. W3C의 음성 웹브라우저 작업 그룹에서, 2000년 8월 8일에 음성 상호작용을 이용하여 웹에 접근하도록 하는 규격을 개발하는 것을 목적으로 하는 음성합성 제어를 위한 XML 마크업 언어 규격이 Working Draft로 발표되었다^[1].

웹브라우저 측면에서, 음성기술은 음성합성기반과 음성기반의 웹브라우저로 구분할 수 있으며, 음성합성기반 웹브라우저는 키보드나 마우스, 그리고 음성합성을 이용하고, 음성 웹브라우저는 음성인식과 음성합성을 모두 이용한다.

2.1 음성 웹브라우저의 기능

W3C에서 1998년에 음성 웹브라우저가 가져야 할 기능에 대하여 기술하였으며, 웹브라우저는 음성입력과 합성, 그리고 미리 녹음된 음성을 이용하여 음성 출력을 하며, 키보드와 음성 인식의 조합을 입력 수단으로 한다^[2].

2.2 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 그룹화 되어 조각나 있다. 이러한 형식은 비 시각적 웹 역세를 더욱 어렵게 한다. 시각장애인 사용자는 음성 웹브라우저를 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 들여보고, 추상화하여야 한다. 이러한 방식은 순차 파일을 대상으로 필요한 정보를 찾는 방식과 유사하며, 매우 비효율적 정보 검색 방식이다. 이러

한 문제를 해결하고자 최근 IBM에서는 정보의 중계자로서 프록시 서버를 사용하여 탐색하는 웹 페이지에 주석을 붙여서 사용자에게 되돌려주는 시스템을 제안하고 있다^[3-4].

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.

시각적 인터페이스에서 제공되는 특징 중의 하나는 화면에 나타나는 문서의 필요성과 사용자의 인식정도에 따라 화면 유지시간을 마우스로 조절할 수 있다는 것이다. 그러나 음성인터페이스의 경우에는 발음되는 시간을 제외하고는 그 정보를 유지시키는 것이 불가능하며, 문서내용은 사용자의 기억에 의존한다. 사용자의 이해를 향상시키기 위해서는 문서내용의 난이도와 사용자의 필요성에 따라 문서내용을 읽어주는 속도를 사용자가 선택하게 하는 인터페이스가 요구된다.

본 논문에서는 사용자의 선택에 따라 웹 문서 발화속도를 실시간으로 변화시키는 기능을 설계하고, 구현하였다.

3. 음성 웹브라우저의 설계

현재 진행되고 있는 W3C의 음성 웹브라우저의 표준화는 매우 광범위한 영역을 다루고 있다. 그러나 일반적으로 발표된 음성 웹브라우저는 W3C의 음성 웹브라우저 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 안내하며, 실시간 발화속도제어를 통해 음성출력의 문서 의미전달을 정확하게 한다.

3.1 모델 구조

제안된 음성 웹브라우저의 모델 구조를 그림 1에 표시하였다.

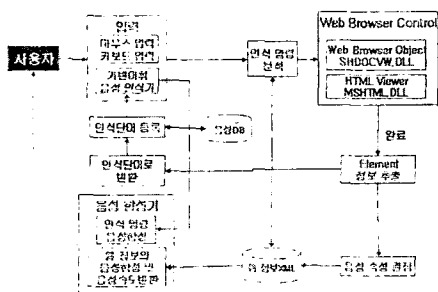


그림 1. 음성 웹브라우저의 모델 구조

음성 웹브라우저의 구조는 크게 입력모듈, 인식명령분석 모

듈, 웹브라우저 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.

4. 구현

4.1 시스템 구성

본 논문에서 구현한 시스템은 음성 합성기로 LGTTS와 단어 레벨 음성인식기로 VoiceLink를 사용하였고, 윈도우즈2000 운영체제 환경 내에서 VC++를 사용하여 그림 2처럼 구현하였다.

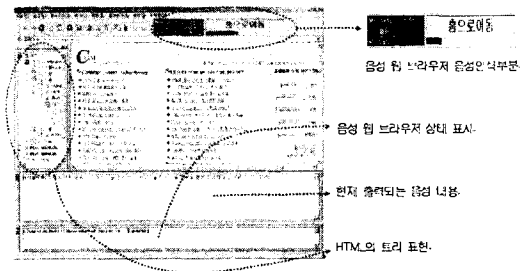


그림 2. 구현된 웹 보이스

가변어휘 음성인식기를 사용하여 웹브라우저 제어에 필요한 수십 가지의 기본 제어를 할 수 있는 명령어와 웹 페이지의 변화에 대하여 동적으로 변화하는 수십 개의 동적 명령어를 음성인식 대상으로 하였다. 웹브라우저는 마이크로소프트사에서 기본적으로 제공하는 웹브라우저 컨트롤을 사용하여 제작하였으며, 웹브라우저 제어는 URL 및 링크요소에 대하여 Navigate2 Method와 Document Object의 Element Click 이벤트를 사용하여 처리하였다.

4.2 음성명령

음성인식을 위한 명령어는 기본적인 웹브라우저 제어명령(뒤로이동, 앞으로이동, 홈으로이동, 새로고침, 멈춤, 프로그램종료, 전체읽기, 주소입력, 다음노드읽기, 이전노드읽기, 다음읽기, 이전읽기, 확인, 즐겨찾기추가)과 웹 페이지의 변화에 따라 새롭게 만들어지는 추가음성명령이 있다. 추가음성명령은 텍스트 정보나 ALT정보의 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.

4.3 음성처리 흐름도

그림 3은 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.

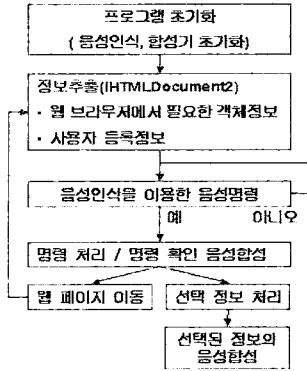


그림 3. 음성처리 흐름도

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 이용을 할 수 있다.

4.4 실시간 발화속도제어

TTS는 정보를 음성으로 출력하기 전에 -5~5의 값으로 발화속도를 설정할 수 있지만 실시간으로 발화속도를 제어할 수 없다. 발화속도제어는 표 1과 같이 다섯 가지(매우 느리게, 느리게, 보통, 빠르게, 매우 빠르게)로 속도를 제어할 수 있도록 하였다. 음성합성엔진에서 실시간 발화속도제어를 지원하지 않기 때문에 실시간으로 발화속도를 제어하기 위해서 음성합성에 사용되는 전체문자열을 일정하게 나누고, TTS의 환경을 따로 저장한 후 나누어진 문자열을 대상으로 실시간으로 발화속도제어를 하도록 구현했다.

표 1. 발화속도제어 종류

| 종류 | TTS 속도 |
|--------|--------|
| 매우 느리게 | 5 |
| 느리게 | 3 |
| 보통 | 0 |
| 빠르게 | 3 |
| 매우 빠르게 | 5 |

4.5 트리구조의 표현

윈도우즈 운영체제에서 트리구조를 표현하는 데에 사용되는 상수구조와 메소드인 CTreeCtrl과 CTreeView를 이용하여 HTML문서의 요소와 상수구조의 분석으로부터 얻어진 표현할 요소를 트리 구조로 재구성하였다. 구현된 트리는 그림 5와 같이 시각적으로도 표현되고, 키보드 조작에 의하여 선택되며, 음성으로 안내되어 웹 문서의 구조를 파악하는데 도움을 준다.

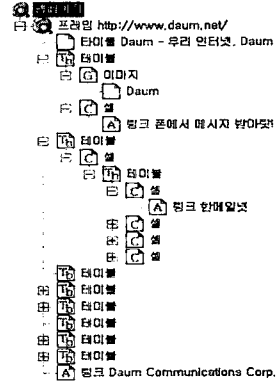


그림 5. 구현된 트리 예

5. 결론

본 연구에서는 웹 문서구조를 트리로 표현하여 음성으로 안내하고 문서내용의 이해를 높이기 위해서 실시간 발화속도제어를 도입한 음성 웹 브라우저를 설계 및 구현하였다. 정상인과 시각장애인, 지체장애인 모두에게 음성명령에 의한 웹브라우저 제어, 내용지정과 지정된 문서를 음성으로 읽어주는 기능을 제공하며, 키보드, 마우스를 이용할 수 있다.

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, 화면 전체에서 원하는 정보에 바로 액세스할 수 있다는 장점이 있다.

아직까지 웹 페이지에서 구조적으로 표현된 시각인터페이스를 음성으로 인터페이스하기에는 중요 키워드 추출이나 정보의 요약과 같은 어려움이 있기 때문에 보다 많은 연구가 진행되어야 할 것이다.

6. 참고 문헌

- [1] Mark R. Walker, Andrew Hunt, "음성 인터페이스 체제를 위한 음성 합성 마크업 언어 규격", <http://www.w3.org/TR/2000/WD-speech-synthesis-20000808>
- [2] Dave Raggett, Ben-Natan, "Voice Browser", W3C Workshop, Jan, 1998, <http://www.w3.org/TR/1998/NOTE-voice-0128>
- [3] Chieko Asakawa et al, "Annotation Based Transcoding for Nonvisual Web Access", Proc. ASSET'00, pp172-179, Nov., 2000
- [4] Hironobu Takagi et al, "Transcoding Proxy for Nonvisual Access", Proc. ASSET'00, pp164-171, Nov. 2000
- [5] Nicole Yankelovich, Gina-Anne Levow, Matt Marx, "Issues in Speech User Interfaces", CHI '95 Proceedings papers, 1995