음성지원 웹 브라우저의 설계

° 신동배, 염세훈, 유재우
숭실대학교 컴퓨터학과

Design of Voice Interfacing Web Browser

° Dong-bae Shin, Se-hoon Yeam, Chae-Woo Yoo
Department of Computing, Soongsil University

요 약

인터넷 붐을 타고 현재 WWW은 수많은 자료를 인터넷 상에 축적했다. 단순히 WWW의 페이지만을 따져도 수 많은 페이지를 가지고 있고, 이제는 Gopher, Ftp, News group또한 WWW인터페이스를 통해 엑세스가 가능하다. 이러한 수많은 자료가 웹 상에서 링크로 연결되어 있다. 자료가 늘어날수록 이러한 링크 또한 늘어난다. 이런 환경에서 전통적인 입력 장치인 마우스와 키보드와 더불어 Voice를 통한 Nevigating을 제공함으로써 좀더 용이한 WWW의 사용을 제공하고자 음성지원 Web Browser를 설계하고자 한다.

1.서론

인터넷과 World Wide Web의 발달로 수 많은 정보를 온라인 상에서 이용 가능하게 되었다. 웹상에서만 해도 수 많은 홈페이지가 수 많은 서버에 연결되어 있고, 그 수는 계속 늘어나고 있다. 그러나 이러한 이용 가능한 정보는 기하급수적으로 늘어나고 있는 반면, 이러한 정보를 접근하는 방법은 상대적으로 기본적인 방법을 벗어나지 못하고 있다. 마우스나 키보드를 포인트하고 클릭하는 것이 웹 정보를 접근하는 사용자들의 기본적인 인터페이스 방식이다. 검색엔진의 출현으로 정보를 검색하는데 많은 도움을 주고 있으나, 검색엔진의 이용은 단지 정보의 추출과 출력의 기능만을 지원하는 것이다. 사용자들은 검색엔진을 이용하더라도 여전히 수많은 링크들을 검색하여야 하며 정보의 위치를 찾기 위해 많은 시간을 보내야 한다. 이러한 상황을 반영해 볼 때 웹 상에서 기존의 키보드나 마우스를 이용한 방법이 아닌 음성을 이용한 정보의 접근이 매우 이상적이다. 특히 초보자들에게는 자연스럽고 융통성 있고, 친숙한 정보전달 방법인 음성을 지원함으로써 좀더 쉽게 정보를 검색하고 추출할 수 있게 해준다.

이러한 음성지원 브라우져는 크게 시각 환경(Visual)에서의 브라우져와 비 시각 환경(Non-visual)에서의 브라우져로 구분할 수 있다. 시각 환경에서의 브라우져는 기존에 사용 가능한 전통적인 컴퓨터에서 사용하는 브라우져를 이용한 방법이며, 비 시각 환경에서의 브라우져는 사용자가 전통적인 컴퓨터가 아닌 휴대폰이나 전화를 이용하여 정보에 접근하는 방법이다. 음성을 이용한 비 시각환경에서의 정보접근은 음성을 인식하는 인식기(Recognizer)와 음성인식과 나타난 정보를 음성으로 바꿔주는 변환기인 음성 출력기(Synthesizer)의 기능이 매우 중요시 된다[4]. 반면 음성을 이용한 시각환경에서의 정보접근은 기존에 사용되는 웹 언어의 음성 지원을 위한 확장과 이를 지원하기 위한 플러그 인 또는 브라우져가 필요하다.

본 논문에서는 이 두 가지 방법의 장점을 취하여 음성 지원을 위한 웹 언어의 확장과 음성 지원 모듈의 통합으로 처리 속도 향상시킨 음성 지원 브라우져를 설계하고자 한다.

2.Voice Browser의 필요성

웹 상에서 정보를 찾기 위해서는 많은 홈페이지를 검색해야 한다. 이러한 홈페이지들은 검색 할 때에 마우스나 키보드를 이용하는 방법은 부족함이 있다. 키보드는 링크를 따라가기에 너무 불편하고, 마우스는 Form 형태의 입력을 할 수 없으니, Voice Interface를 추가하면, 좀더 쉽게 WWW을 사용할 수 있다. 다시 말하면, WWW을 사용하는 데에 음성을 인터페이스로 사용하면, 더욱 용이하게 Web를 사용 가능하게 한다. Form형태의 입력 시나, 단순한 Nevigation에서도 Voice Interface는 편리하다. 예를 들어 현재의 브라우져에서 이전 페이지로 되돌아가는 Forward와 다음 페이지로 가는 back를 생각하여 보자. 마우스로 해당 아이콘을

클릭하는 것 보다 단순히 음성으로 'Forward' 나 'Back' 을 말하는 방법이 편리하다. 마우스와 이러한 Voice Nevigation을 사용한다면 좀 더 빠르게 Link 사이를 이동할 수 있을 것이다. 구현을 위해서는 브라우져의 설치시에 이러한 음성을 미리 저장한다고 가정하면 인식률도 높을 수 있다. 최근에는 PC에서도 멀티 유저 시스템이기 때문에 각자의 메일함이나 폴더(디렉토리)를 따로 가지고 있다. 이를 이용해서 유저마다 각자의 음성정보를 따로 가지고 있게 하면, 여러 사람이 PC를 사용하더라도 문제가 발생하지 않게 할 수 있다

3.Voice Browser 구현 예

웹 상에서 Voice Interface와 Speech Navigation을 지원하기 위해서는 여러가지 구현 방법이 있을 수 있다. 거기에 따라서 좀더 좋은 인식율과 듣기 좋은 Voice를 얻을 수 있다.

여기서는 각 구현에 따른 장단점과 구현방법에 대해서 살펴본다. Texas Instruments SAM은 Web 브라우져의 Voice 인터페이스로써 HTML 페이지에서 하이퍼링크를 뽑아내어 자동으로 문맥자유문법을 만들고 억양그래프를 생성해서 유저가 키보드와 마우스 대신 음성으로 Web Navigation을 할 수 있도록 해준다. OGI SLAM 시스템도 비슷하다. BBS SPIN은 유저의 음성을 캡쳐해서 VQ코드로 바꾸어서 대용량의 Vocabulary 를 가진 음성 인식 서버로 보내서 텍스트로 바꾼다. 그리고 그 텍스트가 미리 지정된 페이지들 중의 하나이면, 해당 페이지가 유저의 PC로 보내지고, 다른 텍스트이면, 서치엔진으로 보내진다.[3] Massachusetts의 WEB-GALAXY는 음성명령으로 마우스를 완전히 대치해서 사용하는 것이 아니라, 마우스로 다른 일을 하거나, 마우스를 사용할 수 없는 경우에 마우스에 부가적인 입력장치로 Voice Interface를 사용한다.[1]

본 논문에서는 두 가지의 구현방법을 고려하였다. 첫 번째로 HTML 태그를 확장하는 방법을 생각해 볼 수 있다. Speech에 관련되어 필요한 억양에 관한 정보, 높낮이, 음성, 악센트에 관한 정보를 음성출력기를 위해서 넣는 방법이다.

Pitch
Value:<hertz> | x-low | low |medium | high | x-high
Initial:medium
Applies to : all elements
Inherited: yes
Percentage values : NA -- [2]

<표 1> 엑센트에 관한 스타일시트정보

<표 1>은 W3C에서 논의중인 음성관련정보를 스타일 시트에 넣어 정의해 놓은 것인데 음성의 피치(높낮이)에 대한 정보이다. 음성 인식을 위해서는 긴 링크 태그를 위해서 중요한 단어 하나 만를 인식하면, 곧장 그 링크로 갈수 있도록 스타일 시트정보에 집어 넣을 수 있다. 그러나 이러한 방법들의 단점은 현재의 수많은 페이지에 대해서는 Voice기능을 제공하지 못한다.

또 다른 방법으로는 단순히 브라우져에 음성인식 기능과 Voice Synthesizer 기능을 집어넣을 수도 있다. 이 경우의 장점으로는 현재의 페이지들에 대해서도 Voice기능을 제공 받을 수 있고, 비 시각 상태에서 Web Browsing이 가능하다. 다시 말해서 이동중인 자동차에서 음성만으로 정보를 추출하는 등의 비 시각환경에서 무선 전화기를 사용해 간단한 정보를 전달하고, 정보를 추출할 수 있다. 이러한 분야에서 많은 응용이 가능한데, 스크린으로 볼 수 없는 상태에서 WEB이용을 할 수 있기 때문이다. 그러나 이 경우 미리 WAV의 형식등으로 저장된 대용량의 Voice DB가 지역마다 필요하다. 좀더 좋은 인식률을 위해서는 HTTP Server와 Client Brower사이에 Middle tier로써 음성인식 서버를 두어서 Client 브라우져가 Voice를 특정한 정보로 바꾸어 그 정보를 음성인식서버에 보내고, 이 서버에서는 대용량의 vocabulary DB를 사용하여 Text로 바꾸어서 HTTP Server로 보내주는 방식으로 움직일 수도 있다. 이러한 Middle Server가 있으면, 현재의 전화 ARS(Auto Response System : 전화 예약 시스템)과 Web Site를 가지고 있는 경우에는 전화 예약 시스템은 Internet Web Site의 예약 페이지로 통합되고 없어질 수도 있을 것이다. 이러한 상황을 미루어 보아 본 연구에서는 위의 두가지 방법의 장점을 취하고자 한다.

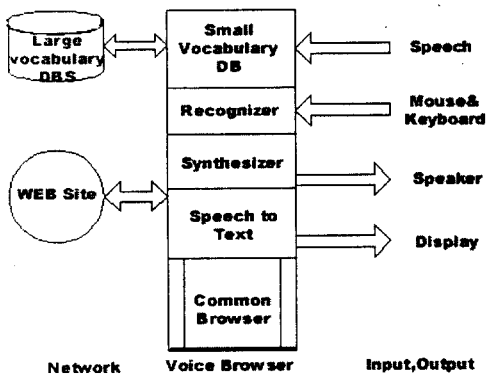
4.Voice Browser의 구성요소

Voice Browser의 전체적인 구성은 <그림 1>과 같다. 사용자가 Mouse나 Keyboard로 명령을 내리면, Voice Browser내부에 있는 Common Browser(이전의 보통 브라우져)부분에서 처리하고, Speech명령을 입력하면 Reconizer(인식기)가 명령을 접수한다. 그리고, Speech을 Text로 바꾸기 위해 1차적으로 내부의 Small Vocabulary DB를 검색한다.자신의 E-mail주소라든지 전화번호 주소 등은 내부에 저장하고 있어야 한다. 예를 들어 이러한 기능이 제공되면, 유저가 새로 산 소프트웨어를 등록하는 데에 필요한 Form 입력을 아주 쉽게 그리고 전혀 틀리지 않게 입력하여 전송할 수 있다.

내부의 DB에서 입력된 Speech에 해당하는 정보를 찾을 수 없으면, 2차적으로 Network상에 있는 Large

Vocabulary DBS에 요청을 해서 Text로 바꿀 수 있는지 알아본다. 그리고, 마지막으로 Large Vocabulary DBS에서도 처리하지 못하면, Speech to Text모듈에서 Text로 바꾸게 된다. 이 처리하는 순서는 음성명령 처리시 요구되는 정확성, 속도등의 요구사항에 따라 결정된다. 음성명령을 Text로 바꾼 후에는 Common Browser가 Web Site와 통신을 하여 정보를 검색하여 결과를 가져온다. 이 결과에 대해서 Synthesizer(음성출력기)가 적절한 작업을 수행하여 Speaker로 출력해 주고, 나머지 부분에 대해서는 Common Browser부분에서 디스플레이 해준다.

그리고, Voice정보 자체가 Network을 통해서 전달하게 되면, 많은 전송량이 필요할 수 있기 때문에 Client는 입력된 Voice를 분석해서 적절한 중간코드로 바꾸어서 통신하여야 한다.



<그림 1> Voice Browser의 구성요소

5.Voice Browser의 설계

본 논문에서는 Voice Browser의 구현을 HTML 태그의 확장과 음성인식기 , 음성출력기의 내장 이 두 가지의 구현방법을 적절히 다 사용해서 두 가지의 장점을 최대한 많이 취하는 모델을 개발하고 있다.

(1) 태그의 확장

웹 페이지를 만든 저작자는 브라우져에서 음성 출력시 그 출력을 조절하고 여러가지 효과를 주기를 원할 것이다. 이 때문에 웹 페이지에 소리의 크기, 높낮이, 음성의 종류, 엑센트정보, 빠르기, 효과음등의 정보가 삽입되어야 한다. 기본적으로 여러가지 형식들도 정의를 하여 웹 페이지 저작자가 사용할 수 있도록 해주어야 한다.

음성인식을 위한 태그확장도 필요한데, 예를들면, 너무 긴 링크를 읽어 인식하려면, 인식의 정확도가 떨어지므로 키워드만을 인식해서 처리해야 하거나, 입력된 음성을 텍스트로 변환하여 웹 서버로 전송할 것인지 어떤 필요에 의해 음성정보를 그대로 전송할 것인지를 나타내 주는 태그도 필요하다. 입력된 음성을 그대로 웹서버에 보내는 경우, 음성을 통해 본인인지를 확인하는 보안에 관련된 홈페이지의 경우가 그러하다.

(2) 음성인식기 , 음성출력기 내장

이 경우 대용량의 DB가 필요하고, 처리능력을 높이고, 여러가지 서비스를 더 제공하기 위해서는 Middle tier에 여러 기능을 하는 여러 서버를 삽입하여야 하고, Client 브라우져가 Middle Server들에게 데이터를 보내도록 설계했다.

브라우져를 사용하여 웹을 검색할 경우 특히 많이 사용하는 음성명령과 아주 가끔 음성으로 구분하여서 자주 사용되는 정보는 Local machine에 저장해야한다. 그리고 가끔 사용하는 음성들은 Middle tier의 서버에 저장해 놓아야 하며, 드물게 사용하는 음성들은 Speech To Text 모듈에 의존해야 한다. 이러한 음성인식 방법의 순위 결정은 정보의 정확한 인식이 어느 정도 필요한지 필요성 , 그리고 사용되는 음성명령의 빈도수에 따라 결정되며, 정확한 변환이 이루어 지도록 결정해야 한다.

6.결론

Web에서의 Voice Interface가 편리하다는 것은 이제 의심할 여지가 없다. 다시 말하면, 현재의 브라우져에 음성 인터페이스를 추가 할 경우 좀더 쉽고, 빠르게 정보를 검색하고 가져올 수 있다. 비 시각환경에서의 WWW이용도 가능해지고, 마우스나 키보드등의 인터페이스가 불가능한 환경에서도 WWW을 사용할 수 있다.

아직 여러 부분의 연구가 더 계속 되어야 하겠지만, WWW Browsing에서의 음성기능 추가는 아주 간단한 WWW Navigation에서부터 HTML 태그가 확장되어 좀더 중요한 기능을 하는 것까지 개발하면 할수록 더욱 WWW을 사용하기 편리하고 다양하게 해줄 것이다.

References

[1] R.Lau, G.Flammia, C.Pao, and V.Zue, "WebGALAX Y:Beyond point and click a conversational interface to a browser", in Proc. Sixth International World Wide Web Conference (M. R. Genesereth and A. Patterson, eds.), Santa Clara, CA, pp. 119-128, Apr 1997.

[2] T.Raman,"Cascaded speech style sheets",in Proc .Sixth International World Wide Web Conference (M. R. Genesereth and A. Patterson, eds.),Santa Clara, CA, pp. 109-117, Ar. 1997.

[3] D. Stallard," Demonstration of BBN SPIN (Speech over the Internet)".Presented at MIT, Cambridge, MA(no published cite),1997.

[4]"Requirements for markup language for HTTP-mediated interactive voice response services", peter Danisls en,Nils Klarlund,David Ladd,Peter Mataga, J.Christopher Ramming,Kenneth Rehor. AT&T Lab-Research, Lucent Technologies-Bell Laboratories, Motorola ICSD,1997