

검색엔진을 이용한 VoiceXML 검색 인터프리터 개발에 관한 설계

이동민, 윤영선, 은성배
한남대학교 정보통신공학과

Design of VoiceXML interpreter for search system

Lee Dong-Min, Yun Young-Sun, Eun Sung-Bae

Department of Information and Communication Engineering, Hannam University

E-mail : cop2007@hannam.ac.kr

Abstract

본 연구에서는 휴대 통신 단말 기기나 기존의 전화망에서의 음성 인터페이스를 이용해 인터넷 웹 검색을 가능하게 하는 VoiceXML 인터프리터를 개발하기 위한 설계를 논의한다. 기존의 VoiceXML 인터프리터 시스템은 미리 정해진 시나리오로 구성된 정적 페이지만을 대상으로 하나 제안하는 방법은 인터넷 웹 사이트에서 음성 인터페이스만을 이용해 원하는 정보에 대한 검색결과를 HTML 로 해석하여 VXML 인터페이스를 통하여 사용자에게 전달한다.

에 따라 기업용 솔루션에 대한 요구가 늘어나고 있다. 다양한 기업용 솔루션에 맞는 음성 서비스를 제공하기 위해 VoiceXML 이 제안되었다. 웹의 특정 응용에서 사용하기 위한 마크 업 언어인 XML 의 한 응용분야인 VoiceXML 은 최근 들어 음성 인터페이스를 위한 중요한 언어가 되고 있다.

본 논문에서는 VoiceXML 에 대하여 간략하게 살펴보고, 기존 VoiceXML 서비스 시스템의 문제점과 이들 문제점을 해결하기 위한 VoiceXML 인터프리터를 제안한다.

I. 서론

인터넷 기반의 데이터 통신이 급격히 증가하면서 기존 전화망인 회선망을 흡수 또는 대체하기 위한 시도가 활발히 진행되고 있다. 이는 인터넷 인프라가 고속 대용량으로 점차 발전되고 전화망과 같이 거의 모든 지역에 확산됨으로써 실시간 전송이 특징인 전화망을 수용할 수 있게 된 것이다.

유, 무선 전화 사용자들이 급격히 증가하고, 이에 따른, 무선 인터넷 접속 서비스가 활성화되면서, 시각적 접근 방법에 의존해 왔던 웹 서비스가 청각적 접근 수단을 제공하고, 이에 맞는 음성 서비스가 등장함

II. VoiceXML

웹을 이용하여 정보를 제공할 때 가장 널리 사용되는 문서의 표준 형식은 HTML (HyperText Markup Language) 또는 이를 확장한 XML (eXtensible Markup Language)이며, 음성 인터페이스를 사용하는 사용자에게 웹 서버의 정보를 제공할 때 사용하도록 정의한 형식이 VoiceXML 이다. 즉 VoiceXML 은 음성 인터페이스를 지원하는 웹 문서를 작성하기 위한 마크 업 언어이며, XML 을 기반으로 하여 만들어졌다.

XML 은 기존의 웹 문서 표준 형식인 HTML 의 확장 어려움, 비체계적, 비 객체지향적이라는 한계를 극복하기 위해 만들어졌다. XML 은 W3C(WWW

Consortium)가 제정한 마크 업 언어이다. XML은 HTML과 달리 개발자가 용도에 따라 정의한 독자적인 태그를 사용할 수 있으며, 1 대 다수의 확장 링크 기능을 제공한다. 일반적으로 XML로 설계된 마크 업 언어를 XML 응용이라고도 부르는데 VoiceXML도 XML 응용의 일종으로 간주할 수 있다. VoiceXML은 일반 웹 브라우저에 음성 입출력을 추가하기 위한 것이 아니라 전화와 같은 음성 인터페이스를 사용하는 통신 기기를 이용해 웹 서버에 접속하는 환경을 가정하고 설계된 음성 웹을 위한 마크 업 언어이다.



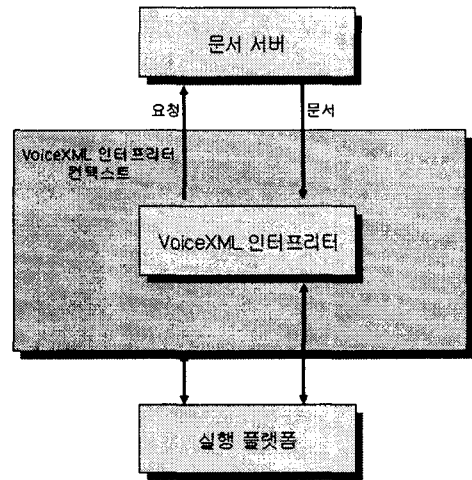
<그림 1> HTML 문서와 웹 브라우저



<그림 2> VoiceXML 문서와 VoiceXML 게이트웨이

위의 그림 1과 그림 2는 HTML 문서와 VoiceXML 문서의 처리 과정을 간략화 한 것이다. 그림 1은 웹 브라우저를 가지고 인터넷 망을 통해 웹 서버에 접근하여 HTML 파일을 처리하는 과정을 나타낸 것이며, 그림 2는 전화를 통해 VoiceXML에 접근하는 과정을 나타낸 것이다. 전화기는 웹 브라우저와는 달리 HTTP (HyperText Transport Protocol)를 통해서 웹 서버에 연결하거나 VoiceXML 문서를 분석하는 능력이 없기 때문에 전화기와 웹 서버 사이에 전화 접속 기능과 HTTP를 통해 웹 서버에 접속하는 기능을 갖춘 'VoiceXML Gateway' 장치가 필요하다.

VoiceXML Gateway는 전화 접속 기능은 물론이고 HTML의 웹 브라우저처럼 웹 서버에 URL을 전송하고 웹 서버가 보낸 VXML 문서를 분석하고 렌더링하는 기능도 수행하며, 음성인식과 음성합성 기능을 포함하고 있다. 웹 브라우저가 HTML 파일을 렌더링한 결과는 화면에 글자나 그림으로 나타나지만 VoiceXML Gateway가 VoiceXML을 렌더링한 결과는 음성이나 오디오 신호가 되어 전화를 통해 사용자에게 전달된다. VoiceXML Gateway의 기능 가운데 VoiceXML 파일을 분석하고 렌더링하는 기능을 'VoiceXML 인터프리터'라고 한다.



<그림 3> VoiceXML 구조 모델

그림 3에서 문서 서버는 웹 서버를 말하며 HTTP 클라이언트 응용 프로그램에 해당하는 VoiceXML 인터프리터가 VoiceXML 인터프리터 컨텍스트를 통해 요청하는 것을 처리한다. 문서 서버는 VoiceXML 인터프리터의 요청에 따라 적절한 처리 과정을 거쳐서 VoiceXML 문서를 만들고, 그 문서를 HTTP 요청에 의한 응답으로 VoiceXML 인터프리터에 전달한다. VoiceXML 인터프리터는 서버가 보낸 VoiceXML 문서를 해석하고 사용자와 상호작용(interaction)을 수행하기 위해 수행 플랫폼을 제어하는 컴퓨터 프로그램을 말한다. VoiceXML 인터프리터 컨텍스트는 웹 서버와 HTTP를 통해 데이터를 주고 받는 HTTP 통신을 담당하며, VoiceXML 인터프리터가 VoiceXML 문서를 해석하게 하고 VoiceXML 인터프리터와는 독립적으로 실행 플랫폼과 상호작용을 한다. 실행 플랫폼은 하드웨어와 소프트웨어를 모두 포함하며, 전화 수신 기능, 전화 호 전환(call transfer) 기능, 음성인식 기능, 음성 합성 기능, 음성과 오디오 재생 기능, 음성과 오디오 녹음 기능 등을 수행한다. 그림 3에서 보면 VoiceXML 인터프리터와 VoiceXML 인터프리터 컨텍스트가 모두 실행 플랫폼을 제어할 수 있는 것을 알 수 있다.

VoiceXML 기술은 다음과 같은 장점을 가지고 있다. 먼저 웹 서버와 같이 이미 잘 구축되어 있는 인터넷 기반의 장비나 콘텐츠를 쉽게 사용할 수 있다. 웹을 기반으로 하는 많은 서비스가 이미 내부의 비즈니스 로직과 사용자 인터페이스를 분리하는 추세이다. 내부 비즈니스 로직은 대개 C, Javascript, perl, python, Java Servlet,

PHP 와 같은 고급 프로그래밍 언어를 사용하여 구현하고, 사용자 인터페이스에 필요한 화면 표시 부분은 HTML 로 구현하는 것이 일반적이다. 여기서 HTML 을 생성하는 부분을 VoiceXML 이 생성되도록 수정하기만 하면 기존의 웹 서버가 가지고 있거나 연동되는 정보를 그대로 음성 정보 서비스 시스템의 콘텐츠로 활용 할 수 있다.

다음으로는 최근에 상당한 수준으로 실용화하여 진행된 음성 합성 및 음성 인식 기술을 쉽게 활용할 수 있다. 기존의 음성 정보 서비스 시스템의 입출력 기능은 녹음된 음성을 들려주고, 전화기에 있는 숫자와 기호만을 입력 받는 것으로 제한되어 있다. 반면에 VoiceXML 을 기반으로 한 음성정보 서비스 시스템은 기존의 입출력 외에도 사람의 음성을 입력으로 직접 받아들여 처리한다거나 수리로 변하는 내용을 합성한 음성으로 제공하는 등 다양한 입출력을 처리할 수 있다. 특히 이런 서비스를 개발할 때 개발자들이 음성인식과 음성 합성에 관한 전문 지식을 갖추지 않아도 무리가 없다는 것이 또 다른 장점이 될 것이다.

마지막으로 VoiceXML 기반 음성 정보 서비스 시스템은 인터넷을 이용하는 분산 시스템이므로, 앞으로 VoiceXML Gateway 의 전화 회선을 임대해 주는 음성 정보 호스팅 서비스가 일반화되면 웹 서버 하나만을 가지고도 전국적 혹은 국제적으로 음성 정보 서비스를 제공할 수 있게 된다 [1,2].

III. 기존 VoiceXML 서비스의 문제점

VoiceXML 은 위에 언급한 장점들을 기반으로 점차 확대 적용되고 있는 추세이다. 그러나 현재 대부분의 전화를 이용하는 음성 서비스들은 이미 녹음되어 있는 음성을 사용하거나 VoiceXML 을 사용하더라도 이미 정해져 있는 시나리오에 따라서 서비스를 제공하고 있다. 이런 방식은 ARS(Automatic Response System)와 같은 방식에서는 사용이 가능하나 기존의 웹 문서를 음성 인터페이스로 제공하는 할 경우에는 적합하지 않다는 단점이 있다 [2].

인터넷은 이미 우리의 생활에서 대단히 중요한 부분들을 차지하고 있다. 현 생활에 있어 대부분의 정보가 인터넷에 있다고 해도 무방할 것이다. 이에 따라 인터넷에 접근하기 위한 수단도 날이 갈수록 발전해 나가고 있

다. 유선 네트워크와 PC 를 사용해서 접근하는 방식은 이미 보편화되어 있고 PDA 나 휴대전화 등 무선 통신을 이용한 접근 또한 점차 증가해 나가고 있다.

그러나 휴대통신기기의 제약사항 중 하나가 바로 접근 인터페이스의 어려움 일 것이다. 현재 나와있는 기기들의 대부분은 디스플레이 장치를 접촉한다거나 펜 입력 또는 별도의 자판을 통한 입력이다. 이런 입력 장치들은 기계를 어려워하는 이들에게는 거부감을 일으키는 요소가 될 수 있다.

접근 인터페이스를 단순화하고 누구나 접근이 용이하도록 할 수 있는 것이 음성을 이용한 인터페이스이다. 그러나 음성 인식 장치나 음성 합성 장치를 휴대 통신 기기에 직접 설치하는 것은 금전적이나 장치 효율적인 면에서 바람직하지 못하다고 할 수 있다. 이런 모든 사항을 종합해 볼 때 VoiceXML 은 휴대통신기기를 통한 접근에 있어서 가장 효과적인 대안이 될 수 있다.

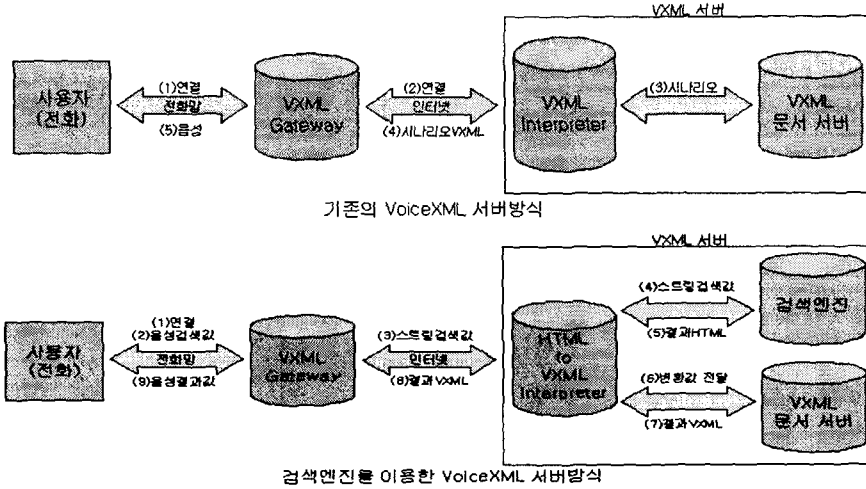
그러나 기존의 VoiceXML 서비스는 기존에 잘 구성되어 있는 인터넷 웹 서비스 망과 연동되지 못하고 있는 단점이 있다. VoiceXML 서비스가 기존의 웹 서버에 VoiceXML 서버를 구성할 수 있다는 점에 비추어 볼 때 비효율적이라고 할 수 있겠다.

이에 본 연구에서는 휴대 통신기기에서 가장 효율적인 인터페이스라고 할 수 있는 음성 인터페이스를 통한 인터넷 웹 망과의 연동에 있어서 기초가 될 수 있는 VoiceXML 서버를 통한 웹 페이지 검색이 가능하도록 검색엔진을 이용한 VoiceXML 검색 인터프리터에 대한 설계 방안을 제시하고자 한다.

IV. 제안하는 VoiceXML 인터프리터

본 연구는 추후 인터넷상의 모든 문서를 일반 PC 상에서 인터넷 브라우저를 통해 검색엔진으로 검색하듯이 음성 인터페이스를 통해 전화로 검색할 수 있는 방안을 제시한다.

그러나 현재의 많은 인터넷 웹 페이지들은 수많은 프로그래밍 언어들로 구현되어 있고 사용자 인터페이스를 위한 그림과 동영상 등으로 꾸며져 있기 때문에 효과적으로 음성 인터페이스를 이용하여 구현하기 위한 태그의 구성에는 어려움이 많다. 이에 인터넷 웹 페이지의 음성 구현을 위한 시작으로 그림과 동영상 등이 배제된 텍스트 기반의 검색엔진을 통한 검색 결과값을



<그림 4> 기존 방식과 검색엔진 이용 방식 개요

VoiceXML로 구현하는 방법을 제시한다.

제안된 방식은 그림 4에서 보이는 것과 같이 기존의 VoiceXML 서버 방식과 다른 방식을 취하고 있다. 사용자가 무선이나 유선 전화망인 PSTN(Public Switched Telephone Network)을 이용해 VoiceXML Gateway에 접속하고 Gateway는 사용자가 입력하는 음성을 문자열로 처리해 VoiceXML 서버의 검색엔진에 전달한다. 검색엔진은 전달된 문자열을 검색 값으로 입력하여 기존의 PC용 인터넷 브라우저에서 실행되는 바와 같은 웹 검색을 실시하고 그 결과를 HTML 문서로 제공한다. 결과 HTML 문서는 HTMLtoVXML Interpreter에 전달되어 음성 출력이 가능한 VoiceXML 문서로 변환되게 된다. 이때 결과 HTML 문서는 이미 정해진 문서 형식으로 제공되어 있는 경우에만 VoiceXML 문서로 변환이 가능하다. VoiceXML로 변환된 문서는 VoiceXML 문서 서버에 저장되고 VoiceXML Gateway로 전달되고 문서를 음성으로 변환하여 전화망을 통해 사용자에게 들려주게 되는 것이다.

이는 기존의 VoiceXML 인터프리터에 HTML 문서를 VoiceXML 문서로 변환시켜주는 HTMLtoVXML 기능을 구현함으로써 가능하다. 본 연구에서는 기존 연구를 통하여 구현된 HTMLtoVXML 인터프리터[2]를 수정하여 검색 엔진과의 인터페이스를 추가하여 검색 문서를 해석하도록 인터프리터를 설계하였다.

설계된 HTMLtoVXML 기능은 HTML 문서에 포함된 문자열 태그들을 VoiceXML 태그들로 대체하고 HTML 태그 사이에 있는 텍스트 문자열들을 VoiceXML 태그들 사이에 삽입함으로써 이루어진다. 삽입된 문자

열들은 음성 인터페이스를 가진 통신기기를 통해 사용자에게 음성 또는 음향으로 전달되는 것이다.

V. 결론

HTMLtoVXML 인터프리터는 비단 검색엔진을 이용한 사이트 검색뿐만 아니라 추후에 이루어질 모든 인터넷 웹 페이지의 음성 인터페이스를 통한 액세스를 가능하게 할 기반이 될 수 있을 것이다. 이는 휴대 무선 통신 기기의 인터페이스 간편화 및 기존 전화 통신망 재사용이라는 측면뿐만 아니라 시각 장애인들의 인터넷 정보 활용이라는 측면에서도 중요한 일이 될 수 있을 것이다.

참고문헌

- [1] 박섭형, "음성 웹 애플리케이션 구축을 위한 VoiceXML", 한빛 미디어
- [2], "VXML을 이용한 기업용 VoIP 통합 시스템 개발", 정보통신부, 정보통신산업기술개발사업 최종연구개발결과보고서, 2003-02-03