

음성 인식용 웹페이지를 위한 웹서비스 제공자의 구현

오지영*, 김윤중**
한밭대학교 컴퓨터공학과

Implementation of the Web Service Provider for the Speech Recognition Web Page

Jee-Young, Oh*, Yoon-Joong, Kim**
Dept of Computer Engineering, Hanbat National University

요약

본 논문은 일반 웹페이지를 음성인식이 가능한 웹페이지로 전환하고, 이 페이지가 사용 될 수 있는 웹서비스를 구현하였다. 본 연구에서 구현한 시스템은 웹서비스 소비자와 웹서비스 제공자로 구성되어 있다. 웹서비스 소비자는 다음에 설명하는 두개의 웹서비스 제공자를 호출하는 기능과 재구성된 웹페이지와 xml 문서를 저장하는 기능, xml 문서로부터 사용자의 음성과 매핑되는 URL을 검색하는 기능을 포함하고 있다. 웹서비스 제공자는 웹페이지를 변환하는 웹서비스 제공자와 음성인식 웹서비스 제공자이다. 웹페이지 변환 웹서비스 제공자는 일반 웹페이지를 분석하여 필요한 태그를 변환하는 기능과 하이퍼링크 값인 URL을 추출하는 기능으로 구성되어 있다. 사용자의 음성을 분석하고 인식하는 음성인식기는 기존의 연구에서 구현된 음성인식 웹서비스 제공자를 이용하였다.

1. 서론

음성인식기술이 발전함에 따라 인터넷에도 음성인식기술을 적용하고자 하는 욕구가 증대되고 있다. 이와 같이 음성 인터페이스의 기대가 증가함에 따라 이미 관련된 많은 연구가 진행되었으며 현재 웹과 관련된 음성인식시스템에 대한 연구가 활발하게 진행되고 있다[1~4].

이들 연구의 음성인식기능이 웹사이트에 적용되기 위해서는 많은 비용과 시간을 투자하여 각각의 웹사이트에 음성 데이터베이스를 구축하고 인식과정을 수행해야 하는 단점을 지니고 있다.

이러한 단점을 보완하기 위하여 '웹서비스를 위한 음성인식시스템'[5]연구가 제안된 바 있다. 이 연구에서는 음성인식 웹서비스 제공자가 구현되어 있다.

이와 같은 음성인식 웹서비스 제공자를 사용하기 위하여 '웹환경에 기반한 음성인식시스템'[6]과 'XML 웹서비스와 인식기술을 이용한 웹브라우저'[7]를 연구한 바 있다. 그러나 이 연구에서는 웹 개발자가 음성

인식이 가능한 웹페이지를 직접 개발해야 한다.

본 연구에서는 음성인식에 대한 지식이 없는 웹 개발자도 일반 웹페이지를 음성인식이 가능한 웹페이지로 재구성할 수 있는 웹페이지변환 웹서비스 제공자를 구현하였다.

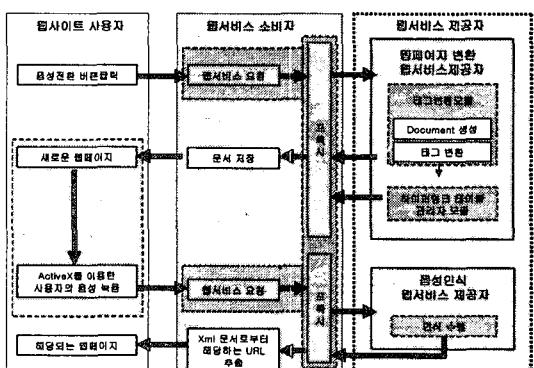


그림 1. 음성인식 사이트를 위한 웹서비스 제공자

본 연구에서 구현한 시스템은 그림 1과 같이 구성되어 있다. 첫 번째 웹사이트를 이용하고 음성을 녹음하는 웹사이트 사용자와 두 번째 음성인식 기능을 필요로 하는 웹사이트 서버인 웹서비스 소비자, 세 번째 웹페이지를 변환하는 웹서비스 제공자가 마지막 웹서비스로 제공되는 인식기로 구성되어 있다. 본 연구에서는 이와 같은 구성에서 웹서비스 소비자와 웹페이지 변환 웹서비스 제공자를 구현하였다.

본 논문의 구성은 1장에서는 서론, 2장에서는 음성인식용 웹사이트 사용자, 3장에서는 음성인식용 웹서비스 소비자에 대하여 설명하고 4장에서는 웹페이지 변환 웹서비스를 소개한다. 5장에서는 실험 및 결과와 마지막 6장에서는 결론을 기술한다.

2. 음성인식용 웹사이트 사용자

본 연구에서 웹사이트 사용자는 웹페이지를 이용하는 사용자이다. 웹사이트 사용자는 일반 웹페이지에 부가된 음성전환 버튼을 클릭하여 음성인식용 웹페이지로 전환한다. 이러한 음성인식용 웹페이지에서 ActiveX 컨트롤을 이용하여 음성명령을 녹음하고 해당하는 웹페이지를 로딩시킨다.

ActiveX 컨트롤은 음성의 녹음과 음성데이터의 전송, 스크립트 호출등의 기능으로 구성되어 있다. 음성 녹음 기능은 Windows의 멀티미디어 자원을 이용하기 위하여 MFC ActiveX로 구현하였다[8,9]. ActiveX 컨트롤은 녹음을 마친 후 음성인식 웹서비스 제공자를 호출하기 위하여 음성데이터를 웹서비스 소비자로 전송한다. 본 연구에서는 전송을 마친 후 ActiveX 컨트롤은 ActiveX 컨트롤의 모든 작업이 종료되었음을 웹페이지에 알리기 위하여 종료 이벤트를 발생시킨다.

3. 음성인식용 웹서비스 소비자

본 연구에서는 웹사이트 서버가 웹서비스 소비자이며, 웹페이지 변환 웹서비스 제공자와 음성인식 웹서비스 제공자를 호출한다. 웹서비스 소비자는 그림 2와 같이 웹서비스 호출기능과 문서 저장기능, xml 문서로부터 URL 검색기능으로 구성되어 있다.

본 연구에서는 .NET framework class library에서 제공되는 객체들을 이용하였다.

3.1 웹서비스 호출

웹서비스 제공자를 사용하기 위해서 웹서비스 소비자는 프록시 클래스를 생성해야 한다. 본 연구에서는 프록시 클래스를 자동으로 만들어주는 Visual Studio

.NET Tool의 웹참조 기능을 이용하였다.

생성된 프록시 클래스는 ParsingService라는 웹서비스와 Parsing 웹메서드, MakeXml 웹메서드를 정의하며 웹페이지 변환 웹서비스 제공자와 통신한다.

두 번째 프록시 클래스는 RecWebService라는 웹서비스와 Rec이라는 웹메서드를 정의하고 SOAP 메시지에 DIME(Direct Internet message Encapsulation) 타입으로 사용자의 음성을 첨부하여 음성인식 웹서비스 제공자에게 전송한다.

3.2 문서 저장 기능

웹서비스 소비자는 웹페이지 변환 웹서비스 제공자로부터 재구성된 웹페이지와 하이퍼링크 테이블을 반환받는다. 재구성된 웹페이지에는 하이퍼텍스트 앞부분에 사용자가 발음해야 할 일련번호가 삽입되어 있고 ActiveX 컨트롤이 활성화 되어있는 음성인식용 웹페이지이다.

하이퍼링크 테이블은 사용자가 발음하는 일련번호와 매핑되는 웹사이트 주소를 포함하고 있다. 본 연구에서는 하이퍼텍스트 앞에 삽입된 일련번호와 하이퍼링크를 xml 문서로 저장하였다. 이 xml 문서는 차후에 인식기에서 인식 과정을 수행한 후 반환되는 인식 결과와 비교하여 해당하는 URL을 찾을 수 있는 일종의 데이터베이스 역할을 한다.

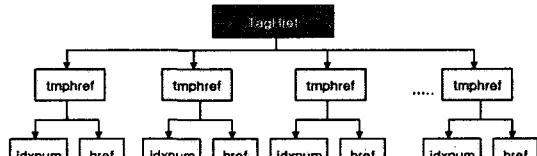


그림 5. 하이퍼링크를 저장하는 xml문서 트리 구조

3.3 xml 문서에서 URL 검색 기능

웹서비스 소비자의 마지막 기능은 음성인식 웹서비스 제공자로부터 인식결과를 수신받아 해당하는 URL로 전환하여 웹사이트 사용자에게 전송하는 것이다. xml 문서에서 인식결과와 일련번호를 비교하고 하이퍼링크를 검색한다.

4. 웹페이지 변환 웹서비스 제공자

웹페이지 변환 웹서비스 제공자는 태그 변환 모듈과 하이퍼링크 테이블 관리자 모듈로 구성되었다. 구현한 웹페이지 변환 웹서비스 제공자의 상세 내용은 그림 3과 같다.

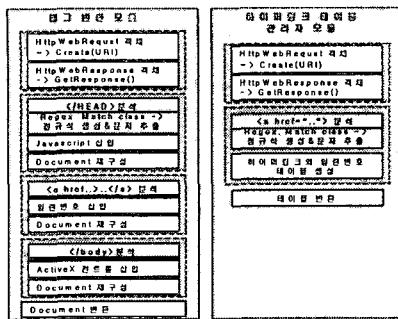


그림 6. 웹페이지 변환 웹서비스 제공자

4.1 태그 변환 모듈

본 연구에서는 웹페이지에서 태그를 검색하기 위하여 정규식을 작성하였다. 정규식을 작성하기 위하여 .NET Framework class library에서 제공하는 Regex 객체와 Match 객체를 사용하였다.

태그 변환 모듈은 웹페이지에서 필요한 태그들을 분석하여 추출하고 필요한 자바스크립트, 일련번호, ActiveX 컨트롤의 구성들을 삽입하여 웹페이지를 재구성한다.

첫 번째 </HEAD> 태그 변환은 웹페이지의 <HEAD>...</HEAD> 사이에 자바 스크립트를 삽입한다. 자바스크립트는 2장에서 기술한 ActiveX 컨트롤의 종료 이벤트를 수신하는 기능과 음성인식 웹서비스 제공자를 호출하고 음성데이터를 전송하는 웹페이지를 호출하는 기능을 포함한다.

두 번째 기능은 >. 태그를 이용하여 태그 앞에 있는 하이퍼 텍스트의 위치를 분석한다. 하이퍼 텍스트의 위치가 파악되면 하이퍼 텍스트 앞부분에 일련번호를 삽입한다.

마지막 기능은 </body> 태그를 분석하여 </body> 태그 앞에 <object..> 태그를 삽입하는 기능이다. ActiveX 컨트롤은 웹페이지 내에서 <Object Classid="" data="" codebase=""> 태그로 삽입되어 있다. 본 연구에서는 ActiveX 컨트롤을 활성화 시키는 방안으로 </body>앞에 <object> 태그를 추가하였다.

4.2 하이퍼링크 테이블 관리자 모듈

본 연구에서는 하이퍼 텍스트를 클릭하는 대신 음성으로 처리되므로 사용자가 발음하는 음성과 매핑되는 하이퍼링크를 추출하여 웹서비스 소비자에게 반환한다. 본 연구에서 추출된 하이퍼링크 문자열은 Match 클래스에 저장되며 추출된 문자열의 위치를 기억한다.

Match 클래스에 저장된 하이퍼링크와 사용자가 발음해야 하는 일련번호를 2차원 배열 형태의 테이블로 생성한다.

5. 실험 및 결과

5.1 실험

본 연구에서 구현된 음성인식 사이트를 위한 웹서비스 제공자의 성능을 분석하기 위하여 다음과 같은 실험을 하였다. 실험 환경은 Windows 2000 server의 OS와 .NET framework 1.0, Visual Studio .NET tool을 이용하였다. 실험 과정은 첫 번째로 일반 웹페이지에 음성 전환 버튼을 부가한 후 음성인식용 웹페이지로 변환하는 단계와 두 번째 음성인식용 웹페이지에서 음성을 녹음한 후 해당하는 웹페이지로 전환하는 단계로 실험하였다.

5.1.1 웹페이지 변환 실험

첫 번째 단계에서 일반 웹페이지에 음성 전환 버튼을 부가하기 위해 표 5와 같은 코드를 작성한다.

```
<INPUT type="button" value="음성전환버튼"
onclick="location.href='SendDime1.aspx?url=http://203.2
30.96.36/testhttp/wavetest.aspx'">
```

표 5. 음성전환버튼의 html 코드

일반 html 버튼인 음성 전환 버튼은 onclick 메서드를 이용하여 SendDime1.aspx 페이지를 호출하도록 작성한다. SendDime1.aspx 페이지는 3장에서 설명한 웹서비스 소비자의 기능을 자동으로 처리해 주는 웹페이지이다. 이때 필요한 파라미터는 변환해야 하는 일반 웹페이지의 url이다.

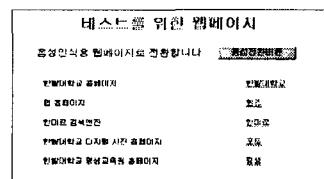


그림 10. 실험을 위한 웹페이지

음성 전환 버튼이 부가되면 음성 전환 버튼을 클릭하여 음성인식용 웹페이지로 변환한다. 그림 11은 웹페이지 변환 웹서비스 제공자에서 변환 과정을 수행한 후 재구성된 음성인식용 웹페이지이다. ActiveX 컨트롤과 하이퍼 텍스트 앞부분에 일련번호가 삽입된 것을 알 수 있다.

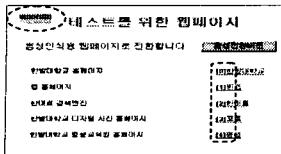


그림 11. NET으로 구현한 웹사이트 웹페이지

이 단계에서는 재구성된 웹페이지뿐 아니라 하이퍼링크를 저장한 xml 문서가 생성된다. 그림 12는 3.2절에서 설명한 xml 문서의 일부분이다.



그림 12. 웹페이지의 하이퍼링크 값을 저장한 xml

5.1.2 음성녹음 실험

두 번째 단계로 그림 11의 재구성된 웹페이지에서 ActiveX 컨트롤로 음성을 녹음하여 전송하고 음성인식 웹서비스 제공자를 호출 한 후 인식과정을 수행시킨다. 실험대상은 화자종속 남성 2인, 화자종속 여성 2인, 화자독립 남성 2인, 화자독립 여성 3인의 음성으로 3회씩 녹음하여 실험하였다. 화자종속적인 음성은 85% 이상, 화자 독립적인 음성은 70% 정도가 해당하는 URL을 찾는 것을 알 수 있다.

5.2 실험 검토 사항

이와 같은 실험으로 일반 웹페이지의 분석과 변환, 그리고 하이퍼링크 추출이 정확하게 이루어지는 것을 알 수 있었다. 또한 본 연구에서 사용한 음성인식시스템에서 발생하는 오류를 제외하고는 정확하게 해당하는 웹페이지를 검색하는 것도 알 수 있었다. 정확한 인식과정을 수행하는 음성인식시스템을 사용할 경우 사용자 음성에 매핑되는 URL을 검색하는 과정이 더욱 향상 될 것이라 판단되어진다.

6. 결론

본 연구에서는 일반 웹페이지를 음성인식용 웹페이지로 변환해 주는 웹페이지 변환 웹서비스를 구현하였다. 웹서비스 소비자는 웹페이지변환 웹서비스 제공

자와 음성인식 웹서비스 제공자를 요청하는 기능과 재구성된 웹페이지와 xml 문서를 저장하는 기능, xml 문서에서 음성에 해당하는 url을 검색하는 기능을 포함한다. 웹페이지 변환 웹서비스 제공자는 태그 변환 모듈과 하이퍼링크 관리자 모듈로 구성되어 있다.

본 연구에서 구현된 시스템의 성능을 분석하기 위하여 작성한 테스트용 웹페이지와 기존에 제작되어 사용 중인 웹사이트의 웹페이지로 실험하였다. 실험결과는 화자 독립적인 음성은 약 85% 이상, 화자 종속적인 음성은 약 70% 정도로 해당하는 URL을 찾는 것을 확인할 수 있었다.

향후 이 연구가 계속되어 현재 하이퍼링크 텍스트 외에 이미지등에 적용되어 있는 하이퍼링크까지 확장할 예정이며 사용자가 발음해야 할 일련번호도 확장시킬 예정이다. 또한 실험 대상을 .NET 웹사이트 개발자 뿐만 아니라 모든 웹사이트 개발 언어에서도 쉽게 사용할 수 있는 방안을 연구 중이다.

참고문헌

- [1] 이항섭외, “웹 브라우저 상에서 한국어 음성인식을 이용한 정보검색 시스템”, 음성통신 및 신호처리워크샵, 15권1호, 1998, pp35-38
- [2] 이재희외, “웹기반의 화자확인시스템 설계에 관한 연구”, 한국음향학회지, 제19권4호, 2000, pp23-30
- [3] 장영건외, “한국어 음성 웹브라우저 설계 및 구현”, 정보과학회논문지, 제7권5호, 2001.10, pp458-466
- [4] 박은주외, “<a href> 태그 추출을 이용한 웹 문서구조화”, 한국멀티미디어학회, 2000 추계학술발표 논문집, pp121-124
- [5] 오신영, “웹서비스를 위한 음성인식시스템의 구현”, 한밭대학교, 정보통신전문대학원 논문집, 제1권1호, 2003.2, pp87-97
- [6] 오지영외2인, “웹환경에 기반한 음성인식시스템의 구현” 한국기술응용학회, 2002 춘계학술발표논문집, 2002.6, pp70-76
- [7] 오지영, “XML 웹서비스와 인식기술을 이용한 웹브라우저”, 한밭대학교, 정보통신전문대학원 논문집, 제1권1호, 2003.2, pp99-114
- [8] 이상엽, 「Visual C++ Programming Bible」, 영진출판사, 1998
- [9] 김용성, 「Visual C++ Programming Bible」, 영진출판사, 2002