

---

# 음성인식 시스템에서의 Voice Web Browsing에 관한 연구

윤재석\*

A Study on Voice Web Browsing in Automatic Speech Recognition Application System

Jae-Seog Yoon\*

---

이 논문은 2002년도 대전대학교 학술연구비지원에 의한 것임

---

## 요 약

본 연구에서는 지금까지의 GUI 중심의 웹 어플리케이션을 VUI 중심의 웹 어플리케이션으로 구현하기 위한 음성 인식 항공 정보 시스템을 설계 구현하였다. 이러한 ASP(Active Server Page)로써 구현한 윈도우 서버 기반에서 운용되는 시스템에 관한 Web 관련 ASR(Automatic Speech Recognition)연구가 최근 상당한 연구가 이루어지고 있지만 ASP의 웹과의 제한성으로 인해 시스템의 속도면, 이식성 등에서 제약을 가져왔다. 이와 같은 제약성을 해결하기 위해 본 연구에서는 음성 정보 및 동적 VoiceXML을 구현하는 자바 빈즈(JAVA Beans) 컴포넌트 구조에 대해서 연구하였다. 또한 본 연구에서는 Remote AWT(Abstract Windows Toolkit)기술을 이용하여 GUI 및 VUI 에서의 음성 및 그래픽 정보를 동시에 전달 가능하게 하는 Voice 웹 브라우저의 가능성을 확인하였다.

## ABSTRACT

In this study, Automatic Speech Recognition Application System is designed and implemented to realize transformation from present GUI-centered web services to VUI-centered web service. Due to ASP's restriction with web in reusability and portability, in this study, Automatic Speech Recognition Application System with Javabeans Component Architecture is devised and studied. Also the voice web browsing which is able to transfer voice and graphic information simultaneously is studied using Remote AWT(Abstract Windows Toolkit).

## 키워드

VoiceXML, Browsing, JavaBeans Component Architecture, RemoteAWT

## 1. 서론

음성인식 기술이 실용화 수준으로 발전함에 따라 웹 기반 음성 포털을 비롯한 음성인식 응용 분야가 새로운 이슈로 떠오르고 있다. 이러한 Web 관련 ASR(Automatic Speech Recognition)연구가

최근 상당한 연구가 이루어졌으며 어떤 것들은 진행 중인 것도 있다. 인터넷 기술의 발전으로 음성 및 컴퓨터 소프트웨어 연구자들은 인터넷 어플리케이션에 음성 기능을 통합하기 위해서 많은 노력을 기울이고 있다[1]. 웹을 기반으로 한 개발과 음

---

\*대전대학교 컴퓨터공학과

접수일자 : 2003. 5. 28

성 입·출력 어플리케이션의 상호 작용에 필요한 콘텐츠 전송을 위해서 XML을 기반으로 한 VoiceXML이 표준으로 떠오르고 있다[2].

사용자와 시스템간의 다이얼로그를 정의하는 VoiceXML을 동적으로 생성하기 위한 방법으로 ASP(Active Server Page)로써 구현한 윈도우 서버 기반에서 운용되는 시스템에 관한 Web 관련 ASR(Automatic Speech Recognition)연구가 최근 상당한 연구가 이루어 있지만 시스템의 제한성으로 인해 시스템의 이식성, 재사용성 등에서 제약을 가져왔다[3]. 이와 같은 제약성을 해결하기 위해서 본 연구에서는 JavaBeans Component Architecture로써 시스템의 각 부분을 재사용 가능한 컴포넌트로 작성함으로써 이와 같은 시스템의 제약을 해결하기 위한 방법을 연구해 보았다. 또한 VUI(Voice User Interface) 시스템이 가지는 음성 정보의 제약을 보완하기 위한 GUI(Graphic User Interface) 어플리케이션을 작성하고, 이것을 RemoteAWT(Abstract Windows Toolkit)[4]를 이용하여 실시간으로 사용자의 컴퓨터로 전송하게 함으로써 좀 더 효율적인 Voice 웹 브라우징을 위한 시스템에 대해서 연구하여 보았다.

## II. 본 론

### 1. 웹 기반 음성 인식 항공 정보 시스템

본 연구에서는 음성으로 웹을 브라우징하기 위한 시스템을 그림 1과 같이 구성하였다. 이 시스템은 시스템과 사용자간의 다이얼로그를 정의하는 VoiceXML로 작성된 콘텐츠를 브라우징하는 것을 전제로 한다.

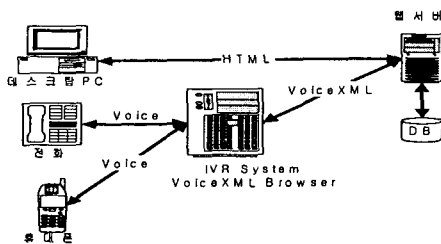


그림 1. 시스템 구성도  
Fig. 1 System Structure

그림1에서 웹서버는 VoiceXML 문서와 오디오 파일, Grammar 파일을 제공한다. 사용자는 일반적인 방법인 IE(Internet Explorer)나 전화, VoIP를 사용하는 MS NetMeeting 등으로 음성 브라우징 시스템에 접속할 수 있다. 사용자와의 음성 인터페이스를 제공하는 IVR(Interactive Voice Response)시스템에는 음성인식 엔진으로 Nuance Speech Recognition System을 사용하였고 VoiceXML로 구현된 웹사이트를 브라우징하기 위한 VoiceXML 브라우저로 Nuance Voice Web Server를 사용하였다[6].

VoiceXML 브라우저에는 VoiceXML 인터프리터가 내장되어 있어 HTTP를 통하여 웹서버로부터 VoiceXML문서를 요청한 후에, 이것을 인터프리터 하면서 IVR 시스템에 접속한 사용자에게 오디오 정보를 제공하거나 사용자의 음성 입력을 받아 들인다. 그림2는 본 연구에서 구현한 비행 정보 시스템의 Call Flow 다이어그램을 나타내는 것이다. 이와 같은 시스템 구성으로 음성 인식에 의한 항공 스케줄 정보를 조회하고 예약 및 예약·확인을 할 수 있는 시스템을 구현하였다.

사용자에게 다양한 Call Flow를 제공하고 좀 더 인터랙티브한 시스템 구현을 위하여 동적으로 VoiceXML 문서가 생성될 필요가 있다. 이를 위해서 본 연구에서의 웹기반 음성 인식 항공 정보 시스템은 각 사용자가 비행 스케줄을 검색하고 이것을 VoiceXML로 변환하는 처리를 하기 위하여 자바 기반의 Servlet을 사용하였으며 이것의 실행 환경으로 Apache Tomcat을 사용하였다[8,9]. 서블릿은 자바 언어를 사용해야하는 제약이 있기는 하나 JavaBeans Component Architecture의 이점을 활용할 수 있고 플랫폼에 의존하지 않는 장점을 가지고 있다. 본 연구에서는 JavaBeans Component Architecture를 따르도록 시스템을 설계함으로써 각 시스템을 구성하는 컴포넌트들의 재사용율이 증가하도록 하였다.

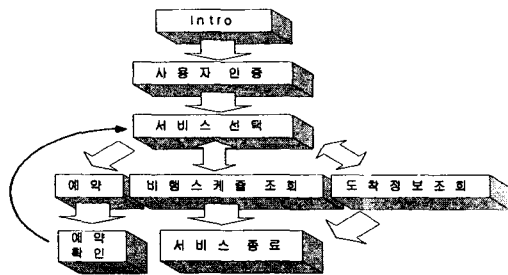


그림 2. 비행 정보 시스템을 위한 기본 Call Flow Diagram  
Fig. 2 basic Call Flow Diagram

## 2. JavaBeans Component Architecture

사용자와 시스템간의 다이얼로그를 정의하는 VoiceXML을 동적으로 생성하기 위한 방법으로 ASP(Active Server Page)로써 구현한 윈도우 서버 기반에서 운용되는 시스템에 관한 Web 관련 ASR(Automatic Speech Recognition)연구가 최근 상당한 연구가 이루어 있지만 웹과의 제한성으로 인해 시스템의 속도면, 이식성 등에서 제약을 가져왔다[3]. 이와 같은 제약성을 해결하기 위해서 본 연구에서는 JavaBeans Component Architecture로써 시스템의 각 부분을 재사용 가능한 컴포넌트로 작성함으로써 ASP 사용시스템의 제약을 해결하기 위한 방법을 연구해 보았다. 또한 ASP 사용시스템에서 라이브러리 기반의 재사용은 다양한 Component들을 조합하기 힘들고 코드 재사용에 한계를 가지기 때문에 큰 효과를 보기 힘들다[6].

이와 같은 제약성을 해결하기 위해서 본 연구에서는 동적으로 음성 정보 및 Call Flow를 VoiceXML 생성하는 시스템의 구현을 위하여 JavaBeans Component Architecture로서 시스템을 설계하였다.

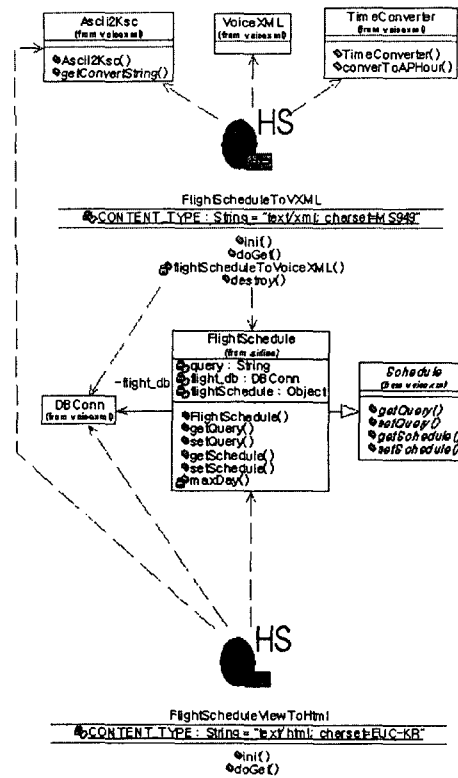


그림 3. UML로 표현한 비행 스케줄JavaBeans component  
Fig. 3 Flight Schedule JAVABeans Component expressed by UML

JavaBeans는 자바 언어로 구현한 재사용 가능한 컴포넌트 모델로 정의할 수 있다. JDBC(Java Database Connectivity) API와 같은 클래스 라이브러리들을 이용하여 사용자가 JavaBeans 내에 데이터베이스에 접근하거나 질의를 하는 부분을 메소드로 만들어 두어 다시 재사용 할 수 있게 하는 것이다. 또한 JavaBeans Component Architecture를 따르지 않는 객체의 상태를 알 수 없는 단순 자바 객체와 구별되고, Java Virture Machine(JVM)에서 작동한다[4].

JavaBeans의 데이터 필드는 private 접근자를 가져야 하는데 데이터 필드들은 JavaBeans의 property가 되고 accessor method나 mutator method를 통해서만 접근이 가능하다[5]. 이렇게 함으로써 컴포넌트의 소스 코드를 보지 않더라도

각 컴포넌트를 재사용할 수 있도록 하였다.

그림3은 본 연구에서 구현한 음성 인식 항공 정보 시스템에서 비행 스케줄에 관한 property를 가지는 FlightSchedule JavaBean 컴포넌트가 본 시스템에서 어떻게 쓰이고 있는지를 UML의 클래스 다이어그램으로 보이고 있다. FlightSchedule 컴포넌트는 accessor method로 getQuery()와 getFlightSchedule()을 가지고 있다. 또한 mutator method로는 setQuery()와 getSchedule()을 가지고 있다. 이 method들은 각각 데이터베이스에 질의할 내용의 설정과 비행 스케줄 반환의 기능을 가지고 있다. 웹서버상의 서블릿이 JDBC 드라이버를 통해서 데이터베이스로부터 비행 스케줄 정보를 받아서 VoiceXML이나 HTML로 변환하도록 하였다. 그림3 상에 있는 HS(Http Servlet)는 각 beans들을 조합해서 원하는 VoiceXML이나 HTML로 변환되어 VoiceXML 인터프리터나 IE(Internet Explorer)로 반환하는 역할을 하는 서블릿을 나타낸다. 이 FlightSchedule은 VoiceXML을 생성하는 FlightScheduleToVXML 서블릿 뿐만 아니라 HTML을 생성하는 FlightScheduleViewToHtml에서도 사용이 가능하다. 그림3의 VoiceXML 클래스는 VoiceXML 스펙 상의 각 태그 생성에 관련된 컴포넌트로 동적으로 VoiceXML을 생성하고자 하는 다른 서블릿에서도 재사용이 가능하도록 하였다.

그림4는 FlightScheduleToVXML 클래스의 코드로써 FlightSchedule과 VoiceXML 등의 컴포넌트들이 사용되며, 각 Property를 getXXX, setXXX 메소드를 통해 반환 및 설정하는 것을 볼 수 있다. JavaBeans는 각 beans의 내부 property를 변경하는 것으로 각각의 용도에 맞도록 재사용할 수 있도록 하였다. 이처럼 JavaBeans Component Architecture를 가지도록 시스템의 구성하면 각 component의 재사용율이 증가하게 된다.

### 3. RemoteAWT를 이용한 Voice Browsing

본 연구에서는 음성 인식 항공 정보 시스템에서 음성으로만 정보를 전달할 시에 발생하는

```

import javax.servlet.*;
import javax.servlet.http.*;
//import 생략
public class FlightScheduleToVXML extends
HttpServlet {
//Initialize global variables
public void init() throws ServletException { }
//Process the HTTP Get request
public void doGet(HttpServletRequest request,
HttpServletResponse response) throws
ServletException, IOException {
//field 선언 생략
PrintWriter out = response.getWriter();
DBConn dbconn = new DBConn();
VoiceXML vxml = new VoiceXML(out);
FlightSchedule schedule =
new FlightSchedule(dbconn);
TimeConverter tconverter =
new TimeConverter();
schedule.setQuery(flight_number,origin,destination,month,
day,hour,period);
//RemoteAWT Start
String cmd = "rjava.bat GUIMain "+ "\"?origin="+
+origin+"&destination="+destination+"&month="+month+
h+"&day="+day+"&hour="+hour+"";
Process p = Runtime.getRuntime().exec(cmd);
String[][] flightSchedule =
(String[][])schedule.getSchedule();
vxml.Header();
vxml.VxmlStart(null,"2.0","ko-KR",null);
vxml.FormStart("viewschedule");
//vxml generate
try {
if (flightSchedule.length > 0) //배행 스케줄을
VoiceXML형식으로 생성
flightScheduleToVoiceXML(vxml,
flightSchedule, tconverter);
//생략
}catch (Exception e) { e.printStackTrace(out); }
//생략
}
//데이터베이스에서 검색된 비행정보를 VoiceXML로
변환
private void flightScheduleToVoiceXML(VoiceXML
vxml, String[][] rset, TimeConverter tconverter)
throws Exception { // method 내용 생략 }
//Clean up resources
public void destroy() { }
}
    
```

그림 4. FlightScheduleToVXML 클래스 코드  
Fig. 4 FlightScheduleToVXML Class Code

제약성을 해결하도록 사용자에게 GUI 어플리케이션을 실시간으로 보낼 수 있도록 하였다. 이때 Remote AWT[7](Abstract Windows Toolkit)기술을 이용하여 GUI 및 VUI에서의 음성 및 그래픽 정보를 동시에 전달 가능하게 하는 Voice 웹브라우저의 가능성을 확인하여 보았다.

본 연구의 웹 기반 음성인식 비행 정보 시스템의 경우에 비행 스케줄 정보를 음성으로 전달을 받게 되면 텍스트를 통해서 스케줄을 검색하는 것에 비해 정보 전달에 많은 시간이 걸린다. 특정 VoiceXML에서 제공하는 음성 데이터를 텍스트 기반의 HTML로 재가공하고, 이 HTML을 특정 VoiceXML 페이지가 요청될 때 이와 상관된 HTML을 동시에 사용자에게 보여줄 수 있도록 GUI 어플리케이션을 작성하였다. 원격지 상의 시스템에 본 시스템에서 작성한 GUI 어플리케이션을 보내기 위하여 RemoteAWT기술을 사용하였다. RemoteAWT는 자바용 Abstract Windowing Toolkit을 구현한 것이며, Sun의 JDK 클래스 라이브러라인 AWT API를 사용해서 자바 어플리케이션을 디스플레이하고 GUI 인터페이스로 상호간에 작용할 수 있도록 한 것이다. 이와 같은 시스템에서는 클라이언트가 GUI어플리케이션을 사용자 시스템에 설치하는 것이 아니므로 어플리케이션이 수정되더라도 별도의 업데이트를 필요로 하지 않으며, GUI 어플리케이션을 제공하는 서버가 GUI를 디스플레이 할 수 없더라도 목적 클라이언트에는 이것을 보낼 수 있다. 또한 목적 클라이언트가 MS Windows를 사용하거나 X-Windows를 사용하더라도 이상 없이 실행된다. 서버와 클라이언트 사이의 원격 인터페이스를 통해서 클라이언트의 컴퓨터에 GUI 어플리케이션이 전송되는 구조를 가지고 있다. 이 시스템은 원격지 상에 AWT API를 사용한 어플리케이션을 보내는 RemoteAWT Host와 보낸 GUI 어플리케이션을 받는 RemoteAWT Client로 나뉜다. RemoteAWT Interface를 수행하기 위한 Host 프로그램과 Client 프로그램이 설치된다.

그림5는 GUI 어플리케이션을 사용자에게 전송하기 위한 구조를 나타내는 것이다. 웹서버는 VoiceXML 인터프리터의 요청을 받은 서블릿을 컴파일하면서, 동시에 RemoteAWT가 작동하여 원격 사용자의 컴퓨터에 GUI 어플리케이션을 전송한다.

각 VoiceXML 요청이 들어오면 서블릿이 서버에서 실행되면서 이 파라미터와 상응하는 어플리케이션을 사용자에게 전달한다. 그림4와 같은 코

드로 작성된 서블릿이 웹서버에서 실행되면 원격지의 목적 클라이언트에 GUI 어플리케이션을 전송하기 위하여 RemoteAWT Host 프로그램이 실행되도록 하였다.

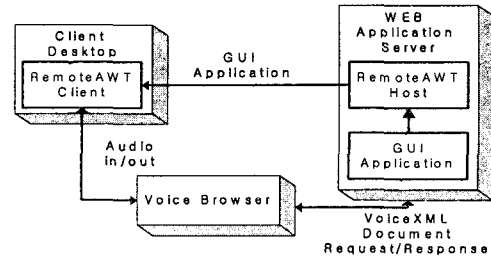


그림 5. RemoteAWT GUI 어플리케이션 전송 구조

Fig. 5 RemoteAWT GUI Application Transmission Structure

이 때 웹 서버내의 각 VoiceXML을 생성하는 서블릿 내에는 그에 상응하는 GUI 어플리케이션과 HTML의 URL 파라미터들을 포함하고 있다. 이 때 서블릿에 들어가는 사용자의 컴퓨터 IP 파라미터의 값에 따라서 GUI 어플리케이션이 각각 다른 사용자의 컴퓨터로도 전송된다. 본 연구에서는 각 사용자를 미리 알고 있다고 가정하고 테스트하였다. 그림6은 사용자가 VoiceBrowser로 비행 스케줄을 검색하기 위해 출발지와 도착지, 출발일자를 말한 후에 VoiceBrowser가 요청한 서블릿 코드인 그림4의 코드가 실행되었을 때 클라이언트의 컴퓨터에 전송된 GUI 어플리케이션을 나타낸다. 그림6의 GUI 브라우저는 웹상의 HTML 문서를 브라우징할 수 있도록 구현된 것으로 서블릿이 요청될 시에 전달된 URL 파라미터의 값에 해당하는 웹페이지를 브라우저 창에 표시한다.

음성 정보와 동시에 그에 상응하는 비주얼 정보를 사용자의 컴퓨터 화면에 전송하게 함으로써 검색한 비행 스케줄 정보를 오디오 출력만으로 듣는 것이 아니라 그림6과 같은 화면을 보고 비행번호로써 원하는 비행 스케줄을 선택할 수 있게 하였다.

음성으로 웹을 브라우징 시에 GUI 정보를 함께 제공할 수 있게 하여 음성으로 VoiceXML로 구현

된 사이트를 브라우징하면서 동시에 HTML로 구현된 웹 페이지를 함께 브라우징할 수 있는 가능성을 확인 해주었다.

는 Multi-modal 웹 브라우징 시스템에 관한 연구가 필요할 것이다.

비행번호	출발지	도착지	출발시간	도착시간	비행시간
2008	서울	제주	4월 1일 11시10분	4월 1일 12시10분	39
2009	서울	제주	4월 2일 11시10분	4월 2일 12시10분	40
2010	서울	제주	4월 3일 11시10분	4월 3일 12시10분	40
8282	서울	제주	4월 30일 9시10분	4월 30일 10시30분	90
8283	서울	제주	4월 30일 15시10분	4월 30일 16시30분	90

원하는 비행번호를 클릭하십시오

그림 6. GUI Client 프로그램이 FlightScheduleViewToHTML 서버릿에 의해 생성된 비행 스케줄 정보를 나타내는 화면  
Fig. 6 Picture presented Schedule information

### III. 결론

본 연구에서는 음성 기능이 결합된 웹 브라우징 시스템을 구성하고, 지금까지의 GUI 중심의 웹 어플리케이션을 VUI 중심의 웹 어플리케이션으로 구현하기 위한 음성 인식 항공 정보 시스템을 설계 구현하였다. 이와 같은 시스템을 구현 시에 이식성과 재사용성이 뛰어난 JavaBeans component architecture를 적용하여 보았다. 또한 VoiceXML로 구현된 시스템에서 음성으로 전달이 용이하지 않은 정보를 전달 시에 원격지 상에 GUI 어플리케이션을 보내어 그 정보를 제공할 수 있게 하였다. 비행 정보를 전달 시에 음성 정보만 아닌 GUI 정보도 함께 제공하는 시스템을 구현함으로써 음성으로 기존의 GUI 웹을 브라우징할 수 있는 가능성을 열어 보았다.

본 연구에서 구현한 RemoteAWT를 이용한 시스템은 GUI 어플리케이션을 원격지의 컴퓨터에 전송할 시에 약간의 지연이 있었다. 이것은 웹서버에서 RemoteAWT로 GUI 어플리케이션을 보내기 위한 프로세스를 생성하는데 걸리는 시간과 TCP/IP를 통하여 GUI 어플리케이션을 전송되는데 걸리는 시간이다.

이와 같은 병목 현상을 줄이고 좀 더 효과적으로 웹을 브라우징할 수 있게 하기 위하여 GUI와 VUI를 동기화하고 다양한 입력 출력기능을 가지

### 참고 문헌

- [1] Z. Tu and P. Loizou. "Speech recognition over the Internet using Java", IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ: pages 2367-70, Mar. 1999
- [2] VoiceXML, <http://www.voicexmlforum.org>
- [3] 윤재석 외, "VoiceXML을 이용한 음성 인식 시스템에서의 ASP 모듈 연구", 한국해양정보통신학회, Vol 5 No 2, 2001년10월
- [4] RemoteAWT for java, "<http://www.alphaworks.ibm.com/tech/remotewtforjava>
- [5] Don Doherty, Rick Leinecker, "JavaBeans Unleashed" SAMS, 49-66, 2000
- [6] Nuance Developer Network, <http://extranet.nuance.com>
- [7] Gu-Beom Jeong, Guk-Boh Kim, "A Study on Component Customization Concept for Component-Based Reuse Environment", Journal of Electronics & Computer Science, Volume 4 No.2, 13-20, 2002
- [8] Java servlet technology, "<http://java.sun.com/products/servlet/>"
- [9] Jakarta Tomcat, "<http://jakarta.apache.org/>"

### 저자 소개



#### 윤재석(Jae Seog, Yoon)

1982. 경북대학교 전자공학과 공학사

1985. 연세대학교 대학원 전자공학과 석사

1991. Washington Univ. 전기 및 컴퓨터공학과 M.S.

1997. 연세대학교 대학원 전자공학과 공학박사

1998-현재 대전대학교 컴퓨터공학과 조교수 재직중

※ 관심분야 : 음성인식 멀티미디어 응용 시스템, Voice XML, e-비즈니스, 디지털 시스템