

VoiceXML을 사용한 음성 인식 항공 정보 시스템에서의 JAVA Beans Component 구조에 관한 연구

장준식*, 윤재석*, 김국보*
*대진대학교 컴퓨터공학과

A Study on the JAVA Beans Component Architecture in Speech Recognition Flight Information System Using VoiceXML

JoonSik Jang, Jaeseog Yoon, Gukboh Kim
Dept. of Computer Engineering, Daejin Univ.

E-mail : nicejjs@voice-xml.net, jsyoon@road.daejin.ac.kr, kgb@road.daejin.ac.kr

요 약

최근까지 웹은 컴퓨터 상에서의 디스플레이, 키보드, 포인팅 장치들과 같은 비주얼 인터페이스를 통해서 정보 전달 및 서비스를 해오고 있다. 또한 이들은 일부의 모바일용 서비스를 제외하고 대부분이 익스플로어나 네스케이프 등의 웹브라우저를 지원하는 서비스를 해오고 있다. 이와 같은 시스템은 시간과 공간에 제약이 있으며 지원하는 브라우저가 있어야 하는 단점이 있다. 전화의 보급률은 컴퓨터나 기타 장치들에 비해 높고, 음성은 사람에게 쉽게 다가갈 수 있고 편하게 사용할 수 있는 인터페이스이다. 본 논문에서는 지금까지의 보는 것 중심의 웹 서비스를 듣고 말하는 웹 서비스로 음성 인식 항공 정보 시스템으로 설계·구현하였다.

1. 서론

사람들 사이에서 기본적으로 많이 사용되고 있는 통신 수단은 음성이다. 컴퓨터와의 대화를 하는 수단을 보다 인간적이고 가깝게 하기 위해서 음성인식기술이 필요하게 되었고 이를 응용한 많은 분야들이 연구되고 있다. 오늘날의 셀룰러 폰, PDA, 랩탑과 같은 컴퓨터들은 컴퓨터 기술의 발달로 점점 소형화되어 편리하게 휴대를 할 수 있게 되었다. 이들

은 기존의 PC에서만 가능했던 웹을 서핑하거나 웹의 콘텐츠를 읽는 등의 기능들을 휴대하고 다니면서 사용할 수 있게 하였다. 하지만 차를 운전 중에 전자메일을 확인하는 것과 같은 서비스는 사용할 수 없다는 단점을 안고 있다. 이러한 단점을 해결해주는 것이 오랫동안 사람에게 익숙한 음성을 사용하는 것이다. 그리고 이러한 음성을 현재 널리 퍼져 있는 웹으로의 접근할 수 있게 하는 방법의 필요성이 대두되었다.

기존의 GUI(Graphic User Interface)를 VUI(Voice

User Interface)로 전환하고 웹이라는 거대한 네트워크 망도 이용할 수 있게 하는 방안으로 VoiceXML(Voice eXtensible Markup Language)이 개발되었다. VoiceXML은 VoiceXML Forum을 통해 연구되어 버전 1.0 이 발표되면서 W3C의 Voice 그룹에 표준안으로 등록되었다. VoiceXML은 인터넷을 음성으로 접근하게 하는데 있어 가장 효과적인 언어이며, 컴퓨터와 사용자 간의 대화형 표준 마크업 언어라는 점에서 기존의 웹 서비스를 재사용할 수 있게 하고 개발의 편리함과 다른 시스템과의 호환성을 가지게 하였다. 음성 인식 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 진행중인 것도 있다[1-6]. 특히 기존에 ASP(Active Server Page)로써 구현한 윈도우 서버 기반에서 운용되는 시스템에 관한 연구가 진행되었다[4]. 이 연구에서는 ASP의 웹과의 제한성으로 인해 시스템의 속도면, 이식성 등에서 제약을 가져왔다.

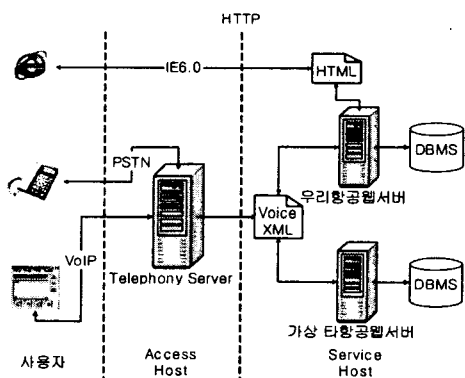
이와 같은 제약성을 해결하기 위해 본 연구에서는 가치 우리항공음성정보시스템을 구현하기 위해 VoiceXML과 웹 기반 음성 인식 시스템을 기본 실험 환경으로 하고 음성 정보 및 동적 VoiceXML을 구현하는 자바빈즈 컴포넌트 구조에 대해서 연구해 보았다. 윈도우 플랫폼에서만 가용 가능한 ASP를 자바 기반의 JSP(Java Server Page)와 Java Beans의 통합 환경과 비교해 볼 때 자바는 거의 모든 플랫폼에 이식 가능하고 시스템의 가용성, 성능면에서 ASP에 비해 우위를 나타내는 것을 알 수 있었다. 각 컴포넌트들은 이미 바이트코드로 컴파일되어 있는 상태이기 때문에 재 컴파일 시에 소요되는 시간을 줄일 수 있고 각 컴포넌트의 재사용면에서 뛰어난 면을 보였다.

또한 본 연구에서는 음성항공정보 시스템의 일부를 XML로 작성하고 VoiceXML 및 XHTML로 변환하여 VoiceXML인터프리터 및 웹 브라우저로 유효성을 테스트하여 웹사이트에서의 Voice 웹 브라우징의 가능성을 확인하여 보았다.

2. 항공 음성 정보 시스템의 설계 및 구현

2.1 실험환경 및 시스템 구성

본 연구에서 사용된 시스템의 전체 구성도는 그림 1과 같다. 시스템은 telephony 호스트와 서비스 호스트, 사용자 등으로 크게 세가지로 나눌 수 있다. telephony 호스트는 사용자의 접근과 상호 작용에 대한 처리를 한다. 본 연구에서 사용한 하드웨어는 AMD 1.2GHz와 768MB의 메모리를 탑재한 시스템을 사용하였다. 사용자가 전화로 접근이 가능하도록 하기 위해 다이얼로직 보드를 설치하였고, 일반 마이크를 사용하여 주로 음성 인식 테스트를 하였다. 또한 MS 윈도우 운영체제에 기본으로 설치되는 NetMeeting 3.0버전으로 VoIP(Voice Over IP)를 통한 테스트를 하였다. telephony 호스트에는 운영체제로 윈도우 NT4.0 Server(SP 5.0), 음성 인식 소프트웨어인 Nuance 음성인식시스템(Nuance Speech Recognition System) 7.0.4버전을 사용하였으며, VoiceXML 문서 인터프리터로 Nuance 음성 웹 서버(Nuance Voice Web Server)1.3을 사용하였다. 음성 인식 언어는 한국어 모델을 사용하여 한국어로 시스템과 대화를 할 수 있도록 하였다[7].



[그림 1] 실험에서 구축한 시스템 구성도

서비스 호스트는 웹 서버와 데이터베이스 서버로 구성 된다. 웹 서버는 아파치(Apache) 웹서버와 JSP/Servlet 컨테이너로 톰캣(Tomcat)1.4버전을 설치하였다. 데이터베이스 서버는 오라클8.1.6을 사용하

였고, 이곳에 비행정보 및 고객 정보, 예약 정보를 저장하도록 하였다[8.9].

그림1에서 웹 서버는 사용자와 시스템간의 대화를 정의하는 VoiceXML 문서와 기존의 웹 서비스도 동시에 할 수 있도록 하기 위한 HTML파일들을 저장하고 있고, Grammar와 오디오 파일들을 가지고 있다. 사용자가 원하는 정보는 아주 다양하므로 이것을 가지고 오고, 처리하는 로직은 JSP와 자바 빈즈 기술을 이용하였다.

이와 같은 시스템은 먼저 기존의 웹 서비스 구조를 재사용할 수 있었다. 개발 방법 또한 기존의 웹 콘텐츠를 만들어 내는 구조와 비슷하므로 개발 시간 및 비용을 감소 시킬 수 있었다. 두 번째로 각 시스템의 독립으로 인해 한 시스템을 교환 하거나 변경하더라도 다른 시스템이 크게 영향을 받지 않으므로 유지, 변경, 보수 면에서 효율적이었다. 셋째로 각 서비스를 구현하는 기본이 표준인 VoiceXML로 구현되어 있기 때문에 다른 시스템으로의 접근과 리소스를 사용하는 것이 가능해지게 되었다.

본 연구에서는 웹 기반 음성 인식 시스템의 이점을 고려하여 각 노드들이 독립적으로 존재할 수 있도록 하였다. Access Host는 사용자의 접근을 처리하고 서비스 호스트로부터 전송 받은 VoiceXML 문서를 인터프리트하고 그 결과를 사용자에게 다시 응답해주는 기능을 하도록 하였다. 그리고 Service Host는 실제 예약이나 비행 정보를 생성, 회원 인증 등을 처리하도록 하였다.

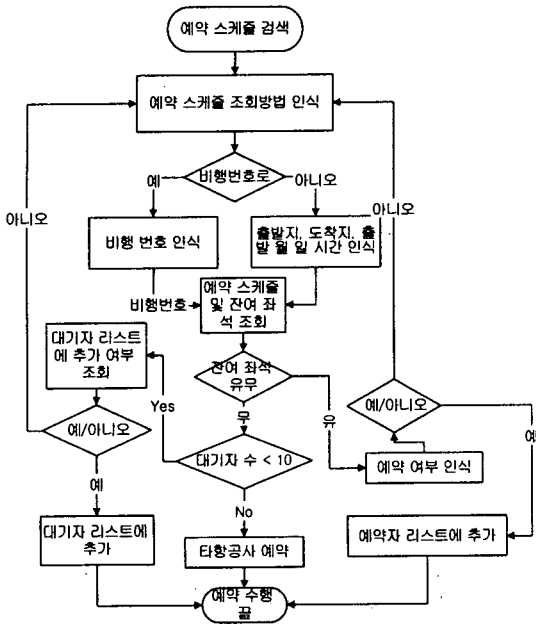
2.2 Call Flow Design

Call Flow는 사용자가 서비스에 접속하여 진행되는 과정을 도식화한 것이다. Call Flow 설계 시에 중점을 둔 부분은 기존의 웹으로 해오던 예약이나 회원 인증과 같은 부분을 음성으로 입력을 받고 처리할 시에 사용자에게 거부감이 가지 않도록 하는 것이었다. 본 시스템에 설치한 음성 인식 엔진이 자연어 처리를 지원하지만 어느 정도는 Call Flow를 통해 범위를 좁혀가는 방법을 사용했을 때 음성 인식을 높일 수 있고 시스템의 신뢰성을 보장 받을 수 있었

다. 본 연구에서 구현한 웹 기반 음성인식항공정보 시스템은 인식 언어로 한국어를 사용하였으며, 다음과 같은 기능들을 가지도록 설계되었다. 먼저 회원 번호와 비밀번호를 이용한 회원 인증 기능을 설계하였다. 회원 번호와 비밀번호는 각각 5자리와 4자리의 단위가 없는 숫자이다. 사용자 입력의 경우 DTMF를 통한 keypad로 입력 받는 방법과 사용자의 음성 입력에 의한 방법 중에서 후자로 테스트 하였다.

음성 시스템의 경우에 사용자가 말할 수 있는 스타일이 아주 다양하기 때문에 사용자의 음성 입력에 의한 처리를 위해서는 Grammar가 필요하다. Grammar는 인간과 컴퓨터 사이에 상호 작용하는 스타일과 그 구조로써, 이를 말하는 형태나 어휘 등의 패턴을 분석하여 정의한다. 이 어휘를 시스템에 무한정 정의한다는 것은 아주 어려운 일이므로 프롬프트를 적절하게 제시하여 사용자의 음성입력 패턴의 범위를 최대한으로 줄여서 인식을 높일 수 있도록 하였다. 시스템이 사용자에게 어떤 프롬프트(prompt)를 주느냐에 따라 그 정의 내용이 많이 달라지게 된다. 이 프롬프트를 효과적으로 주어야 사용자가 쉽고 정확하게 시스템을 이용할 수 있는 것이다.

그림 2는 본 연구에서 구현한 모듈 중에서 사용자가 예약을 하는 과정을 Call Flow로 나타낸 것이다. 그림 2의 각각의 처리부분을 Java Beans로서 모듈화함으로써 재사용성을 높일 수 있도록 구현하였다.



[그림 2] 실험에서 설계한 예약 Call Flow

예약을 하는 방법은 비행 번호와 일반 정보를 통하여 원하는 비행 스케줄을 검색하게 하였고 각 비행 스케줄의 상황에 따라서 다른 처리를 하도록 하였다. 또한 사용자가 예약 서비스에서 예약확인 및 취소할 수 있도록 구현하였다.

VoiceXML을 통하여 사용자에게 비행정보나 예약하는 방법을 알려주게 하였는데, 그 상황 및 비행 스케줄의 내용이 다양하므로 이와 같은 서비스 로직들은 JSP와 Java Beans를 이용하여 내부적으로 처리하도록 하였다. 사용자가 의도하는 바에 따라서 VoiceXML을 생성하고 인터프리터함으로 훨씬 다양한 서비스 로직을 가질 수 있게 되었다. 각각의 클래스별 기능들을 일반 웹 브라우저를 이용하였을 때도 같은 처리 결과를 볼 수 있게 하였다. 이것은 전화나 넷미팅을 통해서 예약을 하더라도 일반 웹 브라우저를 통해서 다시 확인할 수 있고 그 반대로 가능하게 하였다는 것을 의미한다.

3. JAVA Beans를 이용한 서비스 로직 컴포넌트 설계

본 연구에서 동적 VoiceXML 태그 생성 및 각 내부 로직을 처리하기 위해 이용한 기술인 JSP(Java Server Page)와 Java Beans는 자바에 근간을 두고 있다. XML과 유사한 태그를 바로 쓸 수 있고 페이지를 생성하는 로직을 캡슐화하는 스크립트릿을 사용한다. 또한 응용 로직은 이들 태그와 스크립트릿으로 페이지에 접근하는 서버기반 자원(자바빈즈 컴포넌트 아키텍처)를 포함시킬 수 있다. JSP와 자바빈즈 컴포넌트는 디자인과 디스플레이, 지원되는 재사용 가능한 컴포넌트 기반의 설계로부터 페이지 로직을 분리함으로써 다른 웹기반 응용에 비해 더 빠르고 쉽게 만들어 낼 수 있다[8,9].

이전에 ASP로 구현한 것은 지원되는 플랫폼이 아주 한정적이었다. 만약 웹 서버를 윈도우 기반에서 ASP로 웹 서비스를 운영하는 IIS(Internet Information Service)에서 다른 웹 서버나 OS로 변경을 하게 되면 지금까지 사용한 코드들을 모두 그 플랫폼에 맞는 코드로 변환을 해주어야 한다. 하지만 자바빈즈 컴포넌트 구조는 자바 응용 환경을 위한 플랫폼 독립적 구조이고, 컴포넌트의 재사용에 많은 이점을 가지고 있다. 각각의 자바 빈즈들은 패키징화 되어 있기 때문에 재사용이나 재배포면에서 강력한 기능을 가진다.

ASP에서 구현한 모듈들은 각 기능들을 프로시저로 모듈화 하였으나 컴파일된 상태가 아니기 때문에 사용자가 요청할 때마다 서버에서 컴파일을 했으나 자바 빈즈는 바이트코드로 변환된 상태이기 때문에 재컴파일 시간을 줄일 수 있었다. 페이지에 동시 접근 횟수가 많아지게 되었을 때 ASP로 구현한 웹서버에 비해 훨씬 안정되고 빠른 응답시간을 보여주었다. 자바와 같은 거의 완벽에 가까운 정도로 객체지향을 지원하는 언어에서 자바 빈즈로 구현할 시에 각 패키지들은 클래스로 구현되고 상위에서 하위의 상속의 구조를 가지게 된다. 각 모듈들은 상위에서 구현한 클래스의 변수나 메소드들을 상속 받을 수 있다. 이것으로 상위에서 구현한 메소드를 하위에서 재사용할 수도 있고 다시 재정의도 할 수 있게 된다.

그림 3은 본 연구에서 구현한 자바빈즈 컴포넌트

트 패키지들의 상속 관계를 나타낸 것이다. 각각의 클래스는 상속 구조를 가짐으로 상위에서 구현한 클래스를 하위 클래스에서 상속 받을 수 있다.

```

o class java.lang.Object
o class net.voicexml.dbconn
o class net.voicexml.ascii2ksc
o class net.voicexml.ModVXML
o class net.voicexml.VoiceXML
o class net.voicexml.airline.CheckID
o class net.voicexml.airline.getFlyingInfo
o class net.voicexml.airline.IsMember
o class net.voicexml.airline.reservationCheck
o class net.voicexml.airline.reservationProcess
    
```

[그림 3] 실험에서 구현한 자바빈즈 컴포넌트 상속

모듈에서 가장 기본적인 기능을 하는 VoiceXML 클래스는 VoiceXML 1.0 스펙에 있는 거의 모든 태그를 생성할 수 있다. 이 VoiceXML 클래스를 기반으로 다른 패키지들이 각자의 기능을 수행하도록 구현하였다. `ascii2ksc` 클래스는 아스키코드를 한글 인코딩으로 변환하는 것이다. `dbconn` 클래스는 JDBC 인터페이스를 통한 데이터베이스와의 접근하는 부분을 클래스로 구현한 것이다. `ModVXML` 클래스는 년, 달, 월, 시간, 분 등과 기타 한국어의 특징상 달라지는 부분을 처리하는 것이다. 시간에 따라 오전 또는 오후, 숫자 발음에 있어 하나 또는 일, 아홉 또는 구 등의 발음 상의 차이를 구별해서 음성으로 전달하기 위한 것이다. `IsMemeber` 클래스는 회원 인증을 수행하는 클래스이며, `reservationCheck` 클래스는 예약 확인을 위한 클래스이다. 사용자가 조회한 비행 스케줄의 비행기를 예약하는 `reservationProcess` 클래스와 비행 정보를 생성하는 `getFlyingInfo` 클래스가 있다. 각각의 클래스들은 상위 클래스에서 구현된 기능들을 상속 받아 사용할 수 있도록 하였다.

```

<%@ page
contentType="text/xml;charset=euc-kr"
session="true" %><% //생략
net.voicexml.VoiceXML
    vxml = new net.voicexml.VoiceXML(out);
net.voicexml.airline.getFlyingInfo
    gFI = new net.voicexml.airline.getFlyingInfo(out);
net.voicexml.ascii2ksc a2k = new net.voicexml.ascii2ksc();
    
```

```

//파라미터로 각각의 값들을 입력받는다.
//생략
vxml.Header(); vxml.VxmlStart(null,null,"ko-KR");
vxml.FormStart("PlayInfo"); gFI.connectDB();
gFI.getFlyingInfo("reservation_other",
    0, a2k.ascii2ksc(origin), a2k.ascii2ksc(destination),
    d_month, d_day, d_hour, 1);
gFI.disconnectDB();
if (gFI.state==0) //입력 받은 비행번호가 존재하지 않을때
{ //다른 예약 스케줄의 조회여부를 물어본다.
//생략
}
vxml.FormEnd();
vxml.VxmlEnd(); %>
    
```

[그림 4] `getFlyingInfo` 클래스를 JSP에서 JavaBeans로 사용

위 그림 4의 JSP소스는 입력 받은 출발지와 도착지, 비행시간으로 비행 스케줄을 검색하고 예약의 처리 여부를 입력하도록 요청하도록 한 것이다. 데이터베이스에 접근하여 비행 스케줄을 검색하는 것과 VoiceXML 태그를 생성하는 메서드, 비행 스케줄을 VoiceXML로 변환하는 메서드들은 각각의 VoiceXML, `getFlyingInfo` 클래스로 캡슐화하였다

4. 음성 웹 브라우저

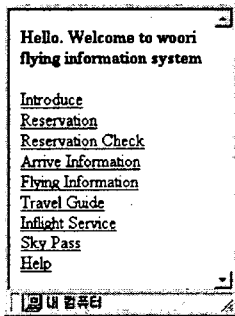
본 연구에서는 XML문서 포맷 변환 기술인 XSLT(XML Stylesheet Language Transformation)을 응용하여 웹 서버 시스템 내에 VoiceXML과 HTML 등의 스크립트 자원이 분리 되는 단점을 해결해 보고, 이들을 VoiceXML로 변환하여 음성 웹 브라우저의 가능성을 연구해 보았다.

현재 많은 웹사이트에서 운영하고 있는 HTML(HyperText Markup Language)을 다른 포맷으로 변환을 할 경우에 HTML 페이지의 비구조적 특성과 순수한 데이터와 디자인을 분리해 내는 것에 어느 정도의 한계가 있음을 알 수 있었다. XML은 HTML에 비해 구조적이며 확장성을 가지고 있기 때문에 변환이 HTML에 비해 효율적이었다. 이것 때문에 XML로 구현한 웹사이트를 실험 대상으로 하였다. 본 항공음성정보시스템의 일부를 XML로 구현하고

이를 XHTML과 VoiceXML 형태로 변환하여 각각을 웹브라우저와 VoiceXML 인터프리터로 유효성을 테스트하여 보았다.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE root SYSTEM "./default.dtd">
<?xml-stylesheet type="text/xsl" href="/xml2html.xsl"?>
<root>
  <menu>
    <voice>Hello. Welcome to woori flying information
system</voice>
    <item id="1" link="/introduce.xml"
vlink="/introduce.vxml">Introduce </item>
    <!--생략 -->
    <item id="9" link="/help.xml"
vlink="/help.vxml">Help</item>
  </menu>
</root>
```

(a)



(b)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE vxml PUBLIC
"http://203.237.81.195/xslt/voicexml1-0.dtd">
<vxml version="1.0">
  <menu>
    <prompt>hello here is woo ri airport system<enumerate/>
    </prompt>
    <choice next="introduce.vxml">introduce</choice>
    <!--생략 -->
    <choice next="help.vxml">help</choice>
  </menu>
</vxml>
```

(c)

[그림 5] 원본 XML문서에서 XHTML과 VoiceXML로 의 변환. (a)원본XML (b)HTML(c)VoiceXML

위의 그림 5는 원본 XML을 XSLT Processor로 처리한 결과를 각각 보여주고 있다. XML로 웹의 컨텐

츠를 작성하고 이를 접근하는 인터프리터에 따라 다른 문서를 전송 받게 함으로써 XML로 이루어진 웹 사이트를 웹 브라우저와 음성으로 동시에 접근 가능하게 하고 콘텐츠를 통합할 수 있었다.

5. 결론

본 연구에서는 VoiceXML과 음성 인식 시스템을 실험 환경으로 구축하여 보았으며, 현재의 웹 기반 시스템과의 연동이나 통합을 하기 위해서 자바기반의 자바빈즈 컴포넌트 구조를 이용하였다. 이것을 음성인식과 음성 정보를 제공하는 항공정보 시스템으로 구현하여 보았다. 이전 연구한 ASP모듈에 비해 자바빈즈 구조는 이식성과 안정성, 모듈화에서 더 좋은 성능을 보였다. 또한 XML의 포맷 변환 기술을 이용할 경우 같은 데이터를 사용자에게 비주얼한 정보뿐만 아니라 음성 정보도 함께 제공 가능함을 알아 보고 음성에 의한 웹 브라우저의 가능성을 열어 보았다.

이제 웹은 보여지는 것 뿐만 아니라 듣고 말로서 정보를 검색할 수 있게 될 것이다. 풍부한 확장성을 제공하는 XML은 이것을 더 가속화 시킬 것이다. 또한 현재의 거대한 인터넷망과 결합하게 됨으로써 음성 웹이라는 새로운 개념의 네트워크가 형성 될 것이다.

[참고문헌]

- [1] V. Digalakis, L. Neumeyer, and M.Perakis. Quantization of cepstral parameters for speech recognition over the World Wide Web. IEEE Journal on Sel Areas in Communications, pages 82-90, Jan. 1999.
- [2] Z. Tu and P. Loizou. Speech recognition over the Internet using Java. IEEE International Confrence on Acoustics, Speech, and Signal Processing, Phoeix, AZ: pages 2367-70, Mar. 1999
- [3] C.T.Hemphill and Y.K.Muthusamy. Developing Web-based speech applications. European Conference on Speech Communication and Technology, pages 895-898, Sep. 1997.

- [4] 장준식 외, VoiceXML을 이용한 음성 인식시스템
에서의 ASP 모듈 연구, 한국해양정보통신학회,
Vol 5 No 2, 2001년10월
- [5] World Wide Web Consortium, <http://www.w3.org>
- [6] Voice extensible Markup Language VoiceXML, Version
1.0, VoiceXML Forum, Mar. 2000,
<http://www.voicexml.org>.
- [7] Nuance Developer Networks, <http://extranet.nuance.com> .
- [8] JAVA Beans, <http://java.sun.com/products/javabeans>.
- [9] JavaServer Pages(TM) Technology,
<http://java.sun.com/products/jsp/index.html>