VXML 편집기와 음성 브라우저의 설계 및 구현

김경란°, 홍기형
성 신 여 자 대 학 교 컴 퓨 터 정 보 학 부
{krkim, khhong}@cs.sungshin.ac.kr

Design and Implementation of Voice Browser and VXML editor

Kyoung-Ran Kim° and Ki-Hyung Hong
School of Computer Science and Engineering, Sungshin Women's University

요 약

VXML은 VoiceXML 포럼에서 제안한 대화형 음성 인터페이스 표준으로 음성 관련 소프트웨어 개발을 용이하게 하는 장점이 있다. 또한 음성을 통해 인터넷의 다양한 정보를 검색할 수 있어 급속히 성장하고 있고 웹의 사용과 편리한 인터페이스의 사용자 요구를 충족시킬 수 있다. 본 논문에서는 VXML 문서를 작성할 수 있는 편집기와 VXML 문서를 처리하여 사용자와 상호 작용할 수 있는 대화형 음성 브라우저의 설계와 구현에 대해 소개한다.

1. 서론

음성은 인간이 가지고 있는 기본적인 능력으로 자연스럽고 효과적인 정보교환 수단이다. 음성은 인간과 인간사이의 의사 소통 수단으로서 뿐만 아니라 인간과 기계와의 통신을 위한 수단으로 사용된다.

VXML(Voice eXtensible Markup Language)[1] 은 음성 명령 표준어로 음성 관련 소프트웨어 개발을 용이하게 하기 위해 VoiceXML 포럼에서 1999년 8월에 제안하여 현재 W3C(World Wide Web Consortium) 에서 표준화 진행 중인 마크 업 언어이다. VXML의 주요 이점은 전화와 음성 관련 소프트웨어를 통해 인터넷의 다양한 정보를 검색하고 음성 정보 서비스의 개발을 용이하게 한다. 아직까지 VXML이 특히 음성인식[5]을 이용해 전화로 웹 정보에 접속하기에 충분한 상태는 아니다.

이에 본 논문에서는 VXML 문서를 처리하여 음성을 통한 사용자와의 대화형 브라우저를 구현하여 웹 기반의 서비스 질을 향상하고자 한다.

2. VXML 의 사양 및 음성 브라우저의 구조

2.1 VXML 의 사양

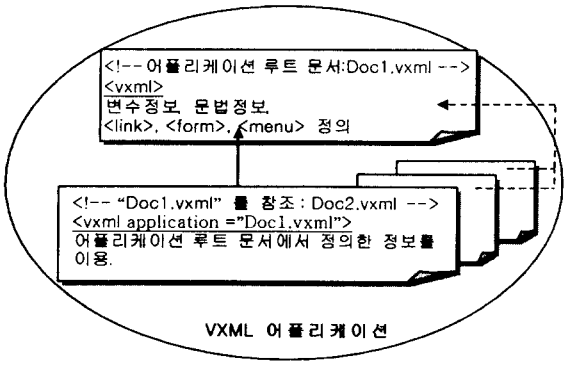
VoiceXML 포럼에서 정의한 VXML스펙(0.9)[1]은 XML[3]에 기초하며 총 41개의 엘리먼트들로 구성되어 있다. 이들은 크게 9가지의 범주로 분류할 수 있는데,<표1> 과 같다.

루트	vxml, meta
다이얼로그	form, menu, choice
프롬프트	prompt, enumerate, reprompt
필드	field, var, initial, block, assign, clear, value
이벤트	catch, error, help, link, noinput, nomatch, throw
오디오 출력	audio, break, div, emp, pros, sayas[4]
오디오 입력	dtmf, grammar, record, transcribe
콜(call)제어	disconnect, transfer
제어 흐름	if, elseif, else, exit, filled, goto
객체	object

<표1> VXML 엘리먼트의 범주

2.2 VXML 문서와 VXML 어플리케이션과의 관계

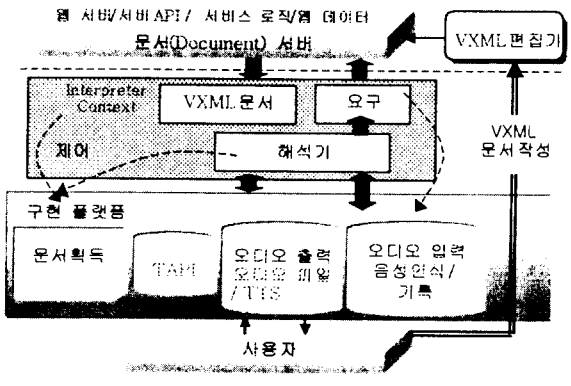
VXML 어플리케이션은 하나의 루트(root)VXML 문서와 이 루트 문서를 참조하는 다수의 비 루트 VXML 문서로 구성된다. 사용자가 어플리케이션에 속한 비 루트 문서를 실행하도록 요구하면 해석기는 이 문서의 어플리케이션 루트 문서를 로드 하여 어플리케이션 루트 문서에서 정의한 변수 및 문법, <link>, <form>, <menu> 등에 대한 정보를 이용 할 수 있도록 유지 한다. 이러한 관계를 이용해 동일한 어플리케이션에 속한 비 루트 문서들은 정보를 공유 할 수 있다. [그림1]은 이러한 관계를 보여준다.



<그림 1> VXML 문서 VXML 어플리케이션의 관계

2.3 음성 브라우저 구조

음성 브라우저(Voice Browser)[2]는 마크 업 언어를 해석하고 음성 입력 해석 및 음성 출력을 생성 할 수 있는 소프트웨어 장치이다. 음성 브라우저를 구현하기 위해서는 음성 인식 엔진과 사용자에게 음성 정보를 제공해 줄 수 있는 음성 합성 엔진[6], 전화를 통해 엑세스 할 수 있도록 하기 위한 전화 음성 인식 엔진이 필요하다.
본 논문에서는 음성 브라우저 해석기와 VXML 문서 생성을 위한 편집기를 구현하였고 [그림2]은 이들의 관계를 나타낸다.

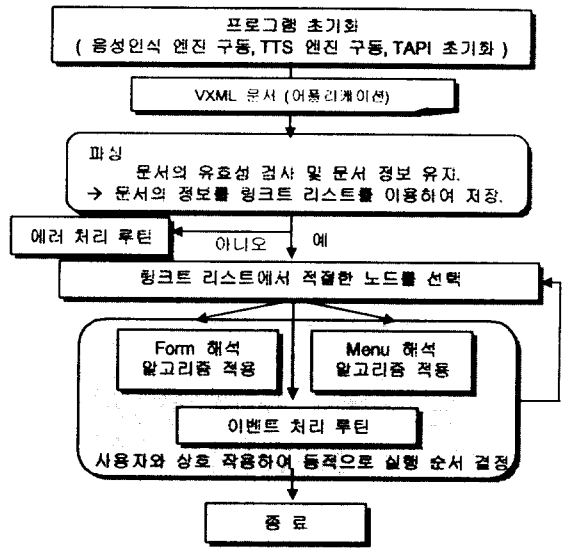


<그림 2> 음성 브라우저의 구조

3. 음성 브라우저 시스템 설계 및 구현

3.1 시스템 수행 과정

시스템은 [그림3]의 과정으로 진행된다.



<그림 3> 프로그램 수행 과정

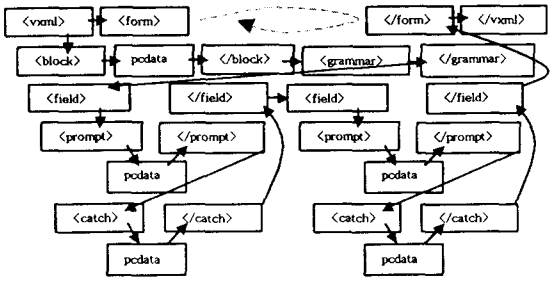
프로그램이 시작될 때 음성인식 엔진과 TTS(Text -To-Speech) 엔진 그리고 전화 음성 인식 인터페이스를 위한 TAPI (Telephony API)엔진의 초기화 과정을 거친다.
사용자가 입력으로 VXML 문서(어플리케이션)를 제공 하면, VXML 문서에 대한 DTD 정보를 이용하여 문서의 유효성 검사 및 문서 정보를 [그림4]의 링크트 리스트로 유지한다. 특히, VXML 파서는 <element name/> 형태 즉, empty 형의 엘리먼트가 존재하면 <element name></element name> 의 시작 태그와 끝 태그의 쌍으로 전환하여 제공하고 끝 태그에 대한 정보도 링크트 리스트의 한 노드에 할당하여 유지한다. 이것은 시작태그와 끝 태그 사이에 존재하는 엘리먼트들의 반복적인 수행의 위치를 식별 하기 위함이다.
다음 [그림4]는 VXML 문서의 파싱 과정 후 생성된 링크드 리스트이다. 링크트 리스트에 연결된 노드의 순서는 문서 자체의 표면적 실행 순서이며 사용자와 상호 작용하여 동적으로 수행 순서를 결정한다.

3.2 수행순서에 영향을 미치는 요소

수행순서에 영향을 미치는 요소로는 크게 엘리먼트 자체의 특성에 의존하는 것과 가드(guard)변수의 상태 변경에 의한 것, 이벤트 처리에 의한 것 세가지로 볼 수 있다.
첫째, 엘리먼트 특성에 의한 것은 <goto>, <link> 엘리먼트등을 통한 다른 문서, 새로운 다이얼로그 또는 현재의 폼 내의 다른 필드로 전이 하는 것이다. 전이의 위치는 이 엘리먼트 속성 값에 의존한다.

```
<?xml version="1.0"?>
<vxml>
<form id = "날씨_정보">
  <block> 날씨 안내 서비스입니다 </block>
  <grammar src="날씨.gram#지명" scope="document"/>
  <field name="나라">
    <prompt> 어떤 나라의 날씨 정보를 원하십니까? </prompt>
    <catch event="help">
      알고 싶은 나라이름을 말씀해 주십시오.
    </catch>
  </field>
  <field name="도시">
    <prompt> 어떤 도시의 날씨 정보를 원하십니까? </prompt>
    <catch event="help">
      알고 싶은 도시이름을 말씀해 주십시오.
    </catch>
  </field>
</form>
</vxml>
```

```
C : 날씨 안내 서비스입니다. 어떤 나라의 날씨 정보를 원하십니까?
H : help
C : 알고 싶은 나라이름을 말씀해 주십시오.
H : 한국
C : 어떤 도시의 날씨 정보를 원하십니까?
H : 서 장시만
C : 무슨 말인지 모르겠습니다. 어떤 도시의 날씨 정보를 원하십니까?
H : 서울
```



<그림4> vxml 문서 처리 예제

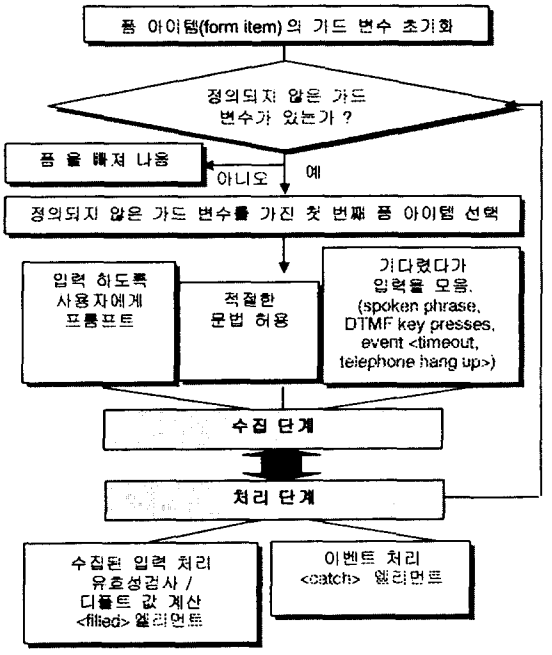
둘째, 가드 변수의 상태에 의한 전이는 <assign> 또는 <clear> 엘리먼트를 통한 가드 값 또는 변수 값의 변경이다. 이 가드 변수는 폼(form) 실행 시 가장 중요한 요소로 수행 순서를 바꾸는데 결정적인 역할을 한다. 셋째, 이벤트 발생으로 인한 이벤트 처리루틴으로 전이하는 것이다. 이벤트는 어플리케이션 문서에서 정의한 사용자 정의 이벤트와 문서 해석기에 의해 미리 정의된 이벤트로 나눌 수 있다.

3.3 해석 알고리즘

VXML 문서는 일종의 다이얼로그 형태로 크게 폼(form)과 메뉴(menu) 두 부분으로 나뉜다.[1] 폼은 사용자에 정보를 보여주고 입력을 모으는 역할을 하며 메뉴는 다음에 무엇을 할지 선택을 제공한다. 메뉴 아이템의 알고리즘은 단일 필드로 구성된 폼의 알고리즘과 같다. [그림 5]는 폼 아이템의 기본적인 해석 알고리즘이다.

3.4 구현환경

본 논문에서 구현한 시스템은 Windows NT4.0, Microsoft Visual C++ 6.0 을 기반으로 하였다.



<그림 5> 폼의 해석 알고리즘

4. 결론 및 향후 연구 과제

VXML의 궁극적인 목표는 인터넷 접속장치 없이 전화를 통해 인터넷의 전자우편, 날씨정보, 철도 항공편 안내 및 예약, 여행정보 등을 검색하고, 웹 페이지에 다양한 음성정보를 실을 수 있게 하는 음성기반 기술이다. VXML이 상용화할 경우 음성을 활용한 웹 기반의 콘텐츠 서비스와 음성인식 어플리케이션 개발에 중요한 영향을 미치게 될 것이다. 향후 연구과제로는 사용자가 VXML 문법을 모르더라도 flow chart 형태의 그래픽 사용자 인터페이스 이용하여 다이얼로그 모델을 제시하면 자동으로 VXML 문서를 생성해 주는 편집기를 구현하고자 한다.

참고 문헌

[1] VoiceXML forum, " Voice Extensible Markup Language VoiceXML" (Version 0.9), http://www.vxmlforum.org/specs
[2] W3C, " Voice Browsers" , Working Draft 23 December 1999, http://www.w3.org/TR/voice-intro/
[3] W3C, " Extensible Markup Language(XML) 1.0" February1998, http://www.xml.com/axml/testaxml.htm
[4] Bell-Lab, " A Synthesis Markup Language" (version 1.0) http://www.bell-labs.com/project/tts/sable.html
[5] 오영환, " 음성 언어 정보 처리 연구의 동향" 정보과학회지 제16권 2호, 1998
[6] 이준우, 김세린, 이종석 " 음성합성 기술의 현재와 전망" 정보과학회지 제16권 2호, 1998