

SIP와 VoiceXML을 이용한 다이얼로그 서버의 구현

김재철* 이명근** 이상정*
*순천향대학교 정보기술공학부, **(주)다이알로직코리아
(jck3426, sjlee)sch.ac.kr, lmk77@n-top.com

Implementation of Dialog Server using SIP and VoiceXML

Jae-Cheol Kim* Myung-Kun Lee** Sang-Jeong Lee*
Dept. of Information Technology Engineering
Dialogic Korea Co., Ltd

요 약

본 논문에서는 PSTN망을 사용하는 일반전화기 사용자와 인터넷 전화 등이 접근 할 수 있는 다이얼로그서버를 구현한다. 개발되는 다이얼로그서버는 VoiceXML로 작성된 서비스 콘텐츠를 사용하고, 이를 HTTP를 통하여 가져오며, DTMF 등의 사용자 입력을 받아서 미리 저장된 음성 데이터나 TTS 컴포넌트 등을 통한 음성 정보 서비스를 제공한다.

1. 서 론

최근 인터넷 사용자가 급증함에 따라 인터넷에 접속하기 위한 다양한 매체들이 대두되고 있다. PC를 이용한 접속 뿐만 아니라 최근에는 이동 전화를 통한 인터넷 사용도 폭발적으로 증가하게 됨에 따라 다양한 음성 서비스에 대한 요구도 늘어나게 되었다. 그러나 기존의 음성 서비스를 제공하기 위한 IVR(Interactive Voice Response)과 같은 플랫폼들은 일반적으로 수정과 유지 보수가 힘들며, 중앙 집중적이고 폐쇄적인 구조로 사용자들의 다양한 서비스 요구와 다양한 서비스 접근 매체를 제한하는 성격을 지니고 있다. 이에 따라 PSTN 전화와 인터넷 기반의 상호 접속 및 운용되는 음성 서비스들을 제공하기 위한 음성포탈 서비스 등이 등장하고 있는 실정이다.

본 논문에서는 PSTN을 사용하는 일반 전화기 사용자와 인터넷 전화 등의 사용자에게 음성 다이얼로그 서비스를 제공하는 SIP[1]와 VoiceXML[2]을 이용한 다이얼로그서버를 구현 한다.

2. SIP와 VoiceXML

SIP는 네트워크 상에서 호 설정을 위한 텍스트 기반의 시그널 링 프로토콜로서, 다이얼로그서버에서는 사용자 측과 세션을 설정하기 위해 사용되었다. 본 논문에서는 PSTN을 사용하는 일반 전화기 사용자가 다이얼로그서버에 접근하는 방법으로서 VoIP 게이트웨이[3]를 사용하였다. 게이트웨이는 PSTN 콜을 받으면 SIP 콜을 생성하여 다이얼로그서버에 콜을 전송하는 기능을 하게 되며

그밖에 전화기 사용자의 DTMF(Dual-Tone Multifrequency)를 인식하여 다이얼로그서버에 전달하는 기능을 한다.

다이얼로그서버가 처리하는 서비스 다이얼로그들은 VoiceXML로 작성되었다. VoiceXML은 웹 기반에서 음성 인터페이스를 지원하고 개발할 수 있도록 구성된 마크업 언어로서 VoiceXML 포럼에서 관리하고 있으며, 현재 VoiceXML 2.0 스펙까지 공개되어 있다. VoiceXML 문서는 사용자와 구현 플랫폼간의 음성 서비스 다이얼로그의 순서를 VoiceXML 인터프리터가 사용할 수 있는 텍스트로 정의하며, 텍스트 데이터를 음성으로 변환하여 출력하는 음성합성출력(Text-To-Speech), 미리 저장되어 있는 음성데이터를 출력하는 음성파일출력, 사용자의 입력 음성을 인식하는 음성인식, DTMF 입력인식, 음성 저장등의 기능을 기술할 수 있다.

3. 다이얼로그서버 동작

본 논문의 SIP기반의 다이얼로그서버는 두 가지 형태의 접근 매체를 통해서 사용할 수가 있다. 첫째, SIP 콜이 가능한 응용을 사용하여 다이얼로그서버에 직접 SIP INVITE 콜을 전송하고 세션을 설정하는 방식이다. 이 방식은 SIP 콜을 전송 할 수 있는 모든 SIP-UA 기반의 인스턴트 메신저, 웹 폰 및 기타 응용 구현에 음성 서비스를 제공할 수 있는 장점이 있다. 둘째로는 PSTN 망을 사용

하는 일반 전화기 사용자가 게이트웨이를 통하여 다이얼로그 서버와 세션을 설정하는 방식이다. 이 방식에서 게이트웨이는 사용자의 PSTN 콜을 받으면, 즉시 해당하는 다이얼로그서버에게 SIP INVITE 콜을 전송한다.

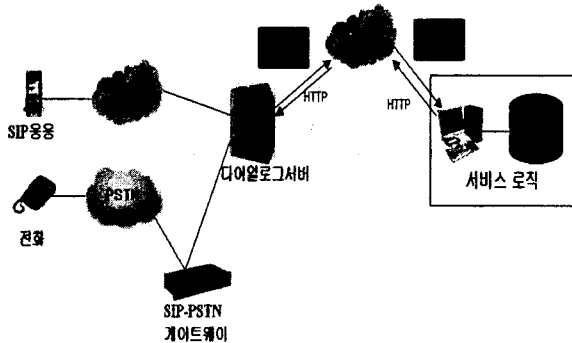


그림 1. 다이얼로그 서버의 네트워크 구성도

그림 1은 SIP를 사용하여 접속하여 여러 시스템과 상호 작용 하는 다이얼로그서버의 네트워크 구성도 이다. 일반적으로 다이얼로그서버는 웹서버의 서비스 콘텐츠, 또는 미디어 파일 등을 가져와서 사용한다. VoiceXML로 기술된 서비스 콘텐츠들은 웹 서버에 정적으로 저장되거나, 자바 서블릿이나 PHP와 같은 서버측 프로그래밍을 통해 동적으로 생성될 수 있다. 또한 다이얼로그서버는 사용자의 접근 노드로서의 기능과 서비스 처리 기능만을 처리하고 직접적인 서비스 로직과는 분리 구성된다. 이는 다이얼로그 서버가 특정 서비스 콘텐츠에 종속적이지 않고, 범용적인 애플리케이션 서버로서의 다이얼로그 서비스기능을 제공해준다. [4]

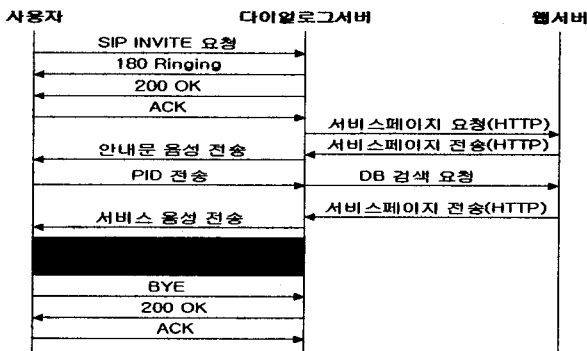


그림 2. SIP 응용 사용자의 콜 흐름도

그림 2는 SIP 콜이 가능한 응용이 다이얼로그서버와 메시지 송수신을 보여주는 그림이다. 사용자는 다이얼로그 서버에 SIP INVITE를 보내는 클라이언트로서의 UA이고 다이얼로그 서버는 이 요청을 받아들이는 서버로서의 UA이다. 그림 3은 PSTN 망을 사용하는 전화기 사용자는 직접 SIP INVITE를 다이얼로그서버에 보낼 수 없기 때문에 게이트웨이를 통한 메시지 송수신을 보여주는 그림이다. 세션이 처음 연결되면, 우선 사용자 인증을 위

한 서비스 페이지를 가져와서 처리하며, 사용자로부터 인증 번호를 입력 받게 된다. 인증이 완료되면, 사용자와 상호 작용 하는 음성 서비스를 제공하게 된다.

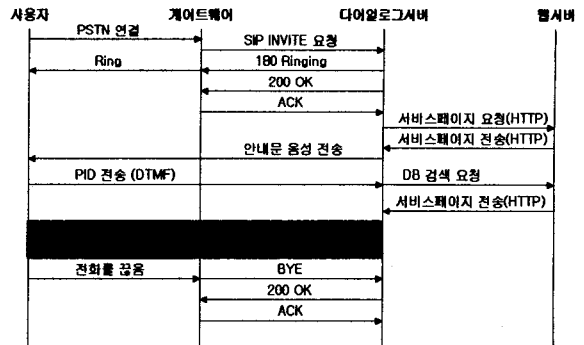


그림 3. PSTN망 사용자의 콜 흐름도

4. 다이얼로그 서버 구현 및 테스트

그림 4는 구현된 다이얼로그 서버의 구성도 이다. 외부와 세션을 설정하는 SIP스택, 음성합성용 TTS컴포넌트, XML컴포넌트, HTTP Fetcher, DB컴포넌트, 오디오컴포넌트와 이들 컴포넌트를 필요할 때 호출하고 관리하는 다이얼로그 서버 메인으로 이루어져 있다.

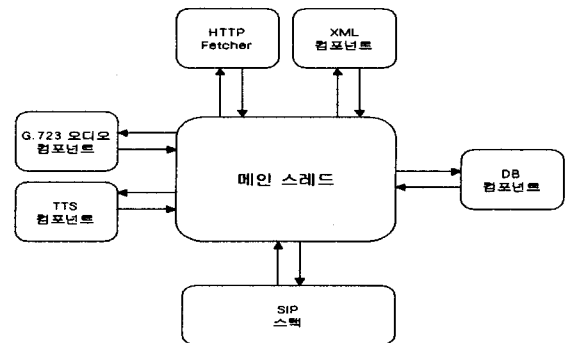


그림 4. 다이얼로그 서버의 구성도

SIP 콜을 통해 세션이 설정되면, 다이얼로그서버는, 사용자가 원하는 서비스 콘텐츠를 해당 VoiceXML 서비스 콘텐츠가 있는 웹서버로부터 HTTP Fetcher를 사용하여 받아온다. 이때 VoiceXML 서비스 콘텐츠를 가지고 있는 웹 서버는 다이얼로그서버 자체가 될 수도 있고, 원격지의 VoiceXML 서비스 제공자가 될 수도 있다. 다이얼로그 서버는 기본적인 서비스를 위한 VoiceXML 서비스 콘텐츠만을 가지고 있으며 원격지의 VoiceXML 서비스 제공자는 각 제공자마다 특화된 서비스 콘텐츠를 가지고 있을 수 있다.

다이얼로그서버가 처리할 VoiceXML 서비스 콘텐츠를 가져오면, XML 파서를 호출하여 해당 VoiceXML 서비스 콘텐츠의 파싱 처리를 요청한다. XML파서는 각 노드를 탐색하여 다이얼로그서버가 처리하기 쉽도록 DOM(Docu

ument Object Model) 트리 노드 데이터를 반환한다. XML 파서는 마이크로소프트사의 MSXML 4.0 버전을 사용하였다.

다이얼로그서버는 처리된 XML 파서의 결과물을 사용자와 대화하는 순서와 방법을 기술해 놓은 다이얼로그 데이터로 취급한다. 다이얼로그서버는 이 다이얼로그 데이터를 VoiceXML 포럼에서 제정한 스펙에 정의되어 있는 FIA(Form Interpretation Algorithm)[2]에 따라 해석한다. VoiceXML로 작성된 서비스 콘텐츠는 일반적으로 여러개의 실행 가능한 폼(Form)과 메뉴(Menu) 다이얼로그로 정의되어 있는데, 이들의 실행 순서를 결정하는 것이 FIA이며, FIA는 반복적으로 루프를 돌며 다음 실행에 해당하는 폼을 처리하게 된다. 이때 다이얼로그의 처리에 영향을 미치는 사용자의 입력은 단순한 캐릭터를 입력, PSTN망 사용자의 경우는 게이트웨이를 통해 DTMF 입력을 받으며, 사용자와 상호작용 하는 서비스를 수행하게 된다. DB 컴포넌트는 인증을 위한 사용자 정보 뿐만 아니라, 서비스 콘텐츠 기술을 제외한 가변 데이터의 저장과 요청을 제공한다. 사용자에게 제공할 가변 데이터에는 제공되는 서비스에 따라 다르다. 예를 들면 뉴스, 날씨, 증권, 스포츠, 지리정보 등 실시간으로 데이터가 변하는 서비스들이 있다. TTS 컴포넌트는 텍스트로 저장되어 있는 콘텐츠들을 PCM 음성 데이터로 변환하는 기능을 한다. 저장된 메일, 또는 미리 저장된 일정관리 등의 텍스트 데이터를 TTS 컴포넌트를 통하여 PCM 음성 데이터로 변환한다. TTS 컴포넌트는 다이얼로그 코리어나서 제공한 상용 툴을 사용하였다. G.723 오디오 컴포넌트는 TTS 컴포넌트가 제공하는 음성 데이터나, 미리 저장된 음성 데이터를 G.723 코덱으로 압축하고 사용자에게 전송하는 기능을 하게 된다. SIP 스택은 콜롬비아 대학의 오픈 소스를 이용하여 구현하였다.

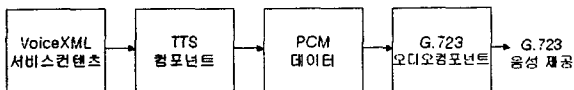


그림5. 음성 데이터 제공 과정

그림 6은 VoiceXML 서비스 콘텐츠 URL을 포함한 메시지의 요청 및 이의 응답 메시지 형식을 보여주는 그림이다. VoiceXML 서비스 콘텐츠의 URL은 SIP INVITE에서 sip:dialog.vxml.http%3a//cs.sch.ac.kr/vxml/script2.vxml@cs.sch.ac.kr 과 같이 설정 할 수 있다. SIP URL은 ‘:’ 를 사용할 수 없으므로 반드시 %3a로 대체 해서 기술해야만 한다. 만약 명시적으로 지정되지 않았다면, 다이얼로그서버의 SIP스택에 설정된 기본 서비스 URL을 가르키게 된다.[6] 다이얼로그서버를 이용한 응용으로는 뉴스, 날씨 증권, 스포츠, 지리정보와 같은 정보 서비스분야가 있다. 또한 banking서비스, 전자상거래와 같은 e-커머스 분야, 음성 메일이나 음성 게시판과 같은 음성저장 서비스 분야, 음성 인식이나, 음성 콘텐츠 등의 대행서비스 분야(VASP), 그 밖에 음성을 통한 홈 네트워킹 정보가전 서비스 분야에 적용할 수 있다.

```

INVITE sip:dialog.vxml.http%3a//cs.sch.ac.kr/vxml/script2.vxml@cs.sch.ac.kr SIP/2.0
Via: SIP/2.0/UDP arch.sch.ac.kr:5060; branch=103438946-0
From: sip:jick@arch.sch.ac.kr
To: sip:dialog.vxml.http%3a//cs.sch.ac.kr/vxml/script2.vxml@cs.sch.ac.kr
Subject: test call for voicexml
CSeq: 1 INVITE
Call-ID: 1901789152@arch.sch.ac.kr
Contact: <sip:jick@arch.sch.ac.kr>
Accept: application/sdp
Content-Type: application/sdp
Content-Length: 162

v=0
o=arch.sch.ac.kr 18487 41 IN IP4 arch.sch.ac.kr
s=SIP Library call
c=IN IP4 arch.sch.ac.kr
t=320138078 0
m=audio 10000 RTP/AVP 4
a=rtpmap:4 g723/6300

SIP/2.0 200 OK
Via: SIP/2.0/UDP arch.sch.ac.kr:5060;branch=103438946-0
From: sip:jick@arch.sch.ac.kr
To: <sip:dialog.vxml.http%3a//cs.sch.ac.kr/vxml/script2.vxml@cs.sch.ac.kr>
Call-ID: 1901789152@arch.sch.ac.kr
CSeq: 1 INVITE
Content-Type: application/sdp
Content-Length: 205

v=0
o=cs.sch.ac.kr 2004728824 2004728824 IN IP4 cs.sch.ac.kr
s=SIP Call
c=IN IP4 cs.sch.ac.kr
t=0 0
m=audio 40108 RTP/AVP 0 4
a=rtpmap:4 g723/6300
    
```

그림6. SIP-VoiceXML 요청과 응답 메시지 형식

5. 결론 및 향후 계획

본 논문에서는 PSTN망을 이용하는 일반 전화기 사용자와 SIP콜이 가능한 모든 응용에 서비스 콘텐츠에 독립적인 다이얼로그 서버를 구현하였다. 서비스 콘텐츠는 원격지의 웹 서버로부터 정적 또는 동적으로 생성된 VoiceXML을 사용하였고, 텍스트 데이터는 TTS 컴포넌트를 통해 음성으로 변환하고 G.723 오디오 컴포넌트를 통해 압축 후 전송하였다. 앞으로 작성된 G.723 컴포넌트의 미디어 전송 부분을 RTP/RTCP로 대체하고, 더 다양한 압축 코덱을 지원하며, 음성합성 뿐 아니라 음성 인식 기능 추가, VoiceXML 2.0 스펙을 지원하도록 개발하여 다양한 서비스가 가능한 다이얼로그서버를 개발할 예정이다.

6. 참고문헌

- [1] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. "SIP: session initiation protocol" IETF RFC 2543, Mar. 1999
- [2] VoiceXML Forum, "Voice extensible markup language (VoiceXML) version 1.00," VoiceXML forum specification, VoiceXML Forum, Mar. 2000.
- [3] 이명근, 권진호, 이상정, " SIP 기반 VoIP 게이트웨이 설계 및 구현," 한국정보처리학회 2002 춘계학술발표논문집(하), 제 9권 제1호, p.1361-1364, 2002년 4월 13일
- [4] J. Resenberg "Distributed Algorithms and Protocols for Scalable Internet Telephony" Columbia University, Ph.D thesis 2001
- [5] T. Ball, V. Bonnewell, P. Danielsen, P. Mataga, K. Rehor. " Speech-Enabled Services Using TelePortal Software and VoiceXML" Bell Labs Technical Journal July-September 2000
- [6] J. Rosenberg, et al. "A SIP Interface to VoiceXML Dialog Servers" IETF Internet Draft July, 2001