
대화형 코퍼스의 설계 및 구조적 문서화에 관한 연구

A Study in Design and Construction of Structured Documents for Dialogue Corpus

양옥렬*, 강창규**, 남명우***

혜전대학 컴퓨터멀티미디어계열*, 원광대학교 컴퓨터공학과**, 혜전대학 디지털전자디자인과***

Ok-Yul Yang(cache@hyejeon.ac.kr)*, Chang-qui Kang(ppibbak@wonkwang.ac.kr)**,
Myung-woo Nam(mwnam@hyejeon.ac.kr)***

요약

음성인식의 연구 대상은 낭독음성에서 대화음성으로 발전해가고 있다. 이를 위해서는 대량의 대화코퍼스가 필요하다. 그러나 아직 충분한 양의 대화코퍼스가 구축되어 있지 못하며 코퍼스의 주석 정보 또한 복잡하고 다양하게 표현하고 있어 효율적인 활용이 어렵다. 따라서 본 논문에서는 TEI를 기반으로 하여 대화 영역을 텔레뱅킹으로 설정하고 대화코퍼스를 구축하여 구축된 대화코퍼스의 주석 정보를 XML(eXtensible Markup Language)로 표준화할 수 있도록 DTD(Document Type Definition) 정의하고 저장 시스템을 설계하였다.

■ 중심어 : | 대화코퍼스 | 코퍼스 | 전사 | DTD | 구조적문서 |

Abstract

Dialogue speech corpora that contain sufficient dialogue speech features are needed for performance assessment of a spoken language dialogue system. And labeling information of dialogue speech corpora plays an important role for improvement of recognition rate in acoustic and language models. In this paper, we examine the methods by which labeling information of dialogue speech corpora can be structured. More specifically, we examined how to represent features of dialogue speech in a structured document based XML and how to design the repository system of the information.

■ keyword : | Dialogue | Corpus | Transcription | DTD | XML |

1. 서론

대화음성의 인식이나 합성 등으로 대표되는 대화정보 처리 기술 개발에 가장 기본적인 연구 자원은 대화코퍼스(Dialogue Corpus)이다. 미국 DARPA의 ATIS(Air Travel Information System Corpus), 일본전자공학

회, 일본음향학회, 그리고 국내에서는 ETRI, KT 등을 중심으로 대화코퍼스의 연구가 활발히 진행되고 있다. 대화코퍼스는 음성을 기록하여 보존하고 다양한 색인 정보도 가지고 있다. 이 정보에는 지정한 단어 또는 문장을 바로 찾을 수 있고 대화의 주석 정보를 포함한 자료들을 검색해 볼 수도 있다.

본 논문에서는 대화코퍼스의 주식 정보를 효과적으로 사용하기 위해 XML로 표준화하여 대화코퍼스를 구조적 문서화하는 방법을 검토하고자 한다. 이를 위해 주식 정보의 추가 및 삭제 등의 확장 가능한 형태의 DTD를 설계하여 주식정보를 표준화하고 XML을 통한 구조적 문서의 웹 서비스가 가능하도록 설계한다.

따라서 본 논문의 목적은 다음과 같다. 첫째, 텔레뱅킹 영역에서 자연스러운 대화음성을 수집한다. 둘째, 전사 규칙을 정의해서 대화음성의 전사 기준으로 적용한다. 셋째, 전사 시스템을 이용하여 대화음성을 전사한다. 이와 같은 절차로 수집되어진 대화음성 정보 및 언어 정보를 포함하는 대화코퍼스를 구축하고, 이상의 결과를 사용자가 손쉽게 활용하고 정보 검색에 이용할 수 있도록 XML DTD를 기반으로 하여 문서를 구조화한다.

본 논문의 구성은, 2장에서 대화코퍼스의 설계를 위한 검토 사항에 대해 살펴보고, 3장에서 대화음성의 세그먼트 규칙과 전사 규칙을 정의하고, 4장에서는 수집된 대화음성 자료를 전사 규칙을 기준으로 DTD와 저장 시스템을 설계하였다. 마지막으로 5장에서 결론을 맺는다.

II. 대화코퍼스의 설계

1. 대화코퍼스의 개요

최근에는 음성 연구의 방향이 낭독음성에서 대화음성으로 활성화되고 있다. 대화코퍼스의 연구에서도 HMM이나 bigram/trigram 등의 언어모델로 대표되는 통계적인 방법이 발달되면서 이를 적극적으로 활용하고 있다[1]. 따라서 대량의 대화음성 자료가 시스템의 훈련이나 학습이 요구가 증가하게 되었고 대화코퍼스의 필요성도 점점 커지고 있다. 그러나 한국어를 대상으로 하는 대화코퍼스 구축은 아직 초보 단계여서 질과 양에서 통계적인 처리 방법의 연구 재료가 되는 대화코퍼스는 충분하지 못하다[2]. 대화코퍼스의 설계 및 구축 절차는 [그림 1]과 같다.

2. 대화의 유형

대화 유형은 대화에서 나타나는 특성(행위, 동기)과 대화자의 인사, 요청, 시사, 수락, 의지, 거절, 감사, 피드백의 표현 등이 나타난다. 대화의 유형을 다음 표 1과 같이 정리한다.

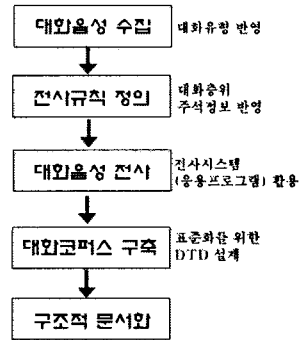


그림 1. 대화코퍼스의 구축 절차

표 1. 대화의 유형

대화 유형	구분	기준
대화 참가자	2명의 대화	2명
	2명 이상의 대화	2명 이상
작업 지향성	작업 지향의 대화	성취해야 할 뚜렷한 목적이 있음
	비 작업 지향의 대화	성취해야 할 뚜렷한 목적이 없음
응용 지향성	응용지향의 대화	산업적, 상업적인 응용
	비 응용 지향의 대화	순수한 연구
영역 제한성	제한된 영역	주제가 제한
	비 제한된 영역	주제가 제한되지 않음
대화 행위	대화 참여자의 역할	무역협상, 정보추출, 상담
인간과 기계	인간과 기계	인간과 시스템의 대화
	인간과 인간	인간과 인간의 대화

본 논문에서는 텔레뱅킹이라는 엄격한 영역 제한성 아래에서 코퍼스를 수집한다[3].

3. 대화의 층위

대화의 층위란 대화에서 나타나는 다양하고 복잡한 주식 정보를 의미한다. 그 종류로는 철자 층위, 형태통사 층위, 통사 층위, 운율 층위, 화용 층위로 구분한다. 각 층위에서 나타나는 주식 정보는 모든 층위가 동일

한 주석 정보의 표기 방식을 따르는 것이 자료의 일관성을 유지하고 자료 교환에 용이하다. 대화 층위의 구성은 그림 2와 같다.

철자 층위(Orthographic Level)의 정보는 대화음성에서 가장 기본이 되는 층위로 텍스트와 순서 정보로 구성된다[4].

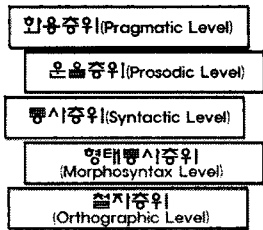


그림 2. 대화 층위의 구성

형태통사 층위(Morphosyntax Level) 정보는 대화 음성 전사에서의 주요 품사 정보, 단어 부류와 관련된 표기를 나타낸다. 주요 품사로는 명사, 동사, 형용사, 대명사, 부사, 간투사이다. 이 밖의 범주로는 부정사, 부정어와 외래어가 들어간다.

통사 층위(Syntactic Level) 정보는 대화 문장의 어절이나, 구에서 나타나는 현상에 대한 정보이다.

운율 층위(Prosodic Level) 정보는 대화 발생자의 억양, 강조, 악센트의 현상을 나타낸다. 영어에서는 어느 단어에 악센트가 나타나고 어느 유형의 악센트가 실현되는지도 알 수 있다. 그러나 한국어의 대화음성의 경우, 억양이 불분명하게 나타나기 때문에 운율 층위의 주석 정보를 적용하여 표기하기가 어렵다.

화용 층위(Pragmatic Level) 정보는 대화에서 행위의 의도(수락, 요구, 인사, 도움)를 나타낸다. 여기서는 화자의 의도가 불분명하게 나타나면 화용적 효과를 주석 정보로 표기한다.

이와 같이, 대화 층위의 주석 정보는 다양하고 복잡하게 나타나기 때문에 일반적으로 표준화하여 표기하기가 어렵다. 그러나 이러한 주석 정보는 전사 규칙의 정의에서 매우 중요한 정보이다. 따라서 각 층위의 주석 정보 중에서 대화음성에 자주 발생하는 것을 기반으로 전사 규칙에 적용한다. 전사 규칙에 적용되는 주석 정보는

표 2. 주석 정보의 전사규칙 적용

대화 층위	주석 정보	전사 규칙 적용
철자층위	중첩	시간 정렬 방법
	잡음	사람에 의한 잡음과 환경적인 요소에 의한 잡음으로 구분
	불완전한 단어, 구두점	기호를 사용 마침표, 물음표, 쉼표
	알 수 없는 음성	쌍괄호 사용
형태통사 층위	간투어	철자 세트를 사용

위의 표 2와 같다.

4. 대화 자료 수집

4.1 시나리오 설계

본 논문에서는 텔레뱅킹의 시나리오를 설계하여 제작하였다. 제작된 시나리오를 기반으로 자연스러운 음성을 수집하기 위해 2-Channel 방식의 수집 시스템을 구성하였다. 위 과정에서 시나리오의 설계 및 제작 단계에서는 대화 유형의 특성을 반영하여 적용하였다. 시나리오는 텔레뱅킹에서 사용되는 다양한 단어들과 언어 구성물을 포함하도록 해야 한다. 따라서 텔레뱅킹 영역에서 필요한 정보 아이템과 이들의 순서가 상황에 맞게 설계되어야 한다[5]. 수집된 아이템을 중심으로 4개의 서브 도메인을 아래와 같이 분류하고 서브 도메인별로 세부 정보 아이템을 정의하였다.

표 3. 서브 도메인 아이템

서브 도메인	세부 아이템
공과금 납부	전화 요금 납부, 전화 요금 납부 내역 조회, 아파트 관리비 납부, 아파트 관리비 납부 내역 조회
신용카드조회	결제 대금 조회, 사용 내역, 사용 한도 조회, 현금 서비스 이체, 사용 명세서 주소지 변경
분실신고	현금 카드 분실, 직물 카드 분실, 수표 분실, 통장 분실, 인감 분실
서비스	우편 서비스, 팩스 서비스

4.2 시나리오 제작

시나리오는 실제 텔레뱅킹의 흐름에 맞게 작성해서 제작하였다. 시나리오의 서브 도메인은 공과금 납부, 신용카드 조회, 분실 신고, 서비스로 분류하여 구성하였다.

시나리오에서의 대화는 상호 주도의 방법을 사용하여 한 발화에 적당한 정보 아이템의 수와 순서를 고려하여 발화하도록 하였다.

제작된 시나리오의 예는 그림 3과 같다.

시나리오
텔레뱅킹 서비스입니다. 이름과 주민등록번호를 말씀하세요. 텔레뱅킹 사용자 번호와 비밀번호를 말씀하세요. 원하시는 서비스를 말씀하세요. 출금하실 통장의 은행명, 계좌번호, 비밀번호를 말씀하세요. 이름과 전화번호를 말씀하세요. 처리되었습니다. 원하시는 서비스를 말씀하세요. 신고하실 카드의 통장 계좌번호와 통장 비밀번호를 말씀하세요. 이름과 전화번호를 말씀하세요. 분실신고를 접수하였습니다. 원하시는 서비스를 말씀하세요. 조회하실 카드 번호와 카드 비밀번호를 말씀하세요. 향후 결제할 대금은 만원입니다. 원하시는 서비스를 말씀하세요.

그림 3. 시나리오 예

4.3 대화 음성 수집 환경

대화음성의 자료 수집 환경은 잡음이 없는 방음 시설을 갖춘 두 개의 부스를 이용하였다. 화자의 지나치게 크거나 작은 발성, 구강 노이즈, 어색한 발성 등은 관리실에서 시스템 관리자가 통제를 했다. 화자가 긴장하지 않고 자연스러운 발성을 유도하기 위해 실제 텔레뱅킹 환경을 유지 하였다[6].

대화음성의 수집 장비는 표 4와 같다.

표 4. 대화음성 수집 장비

장비 명	기능	용도
Soundcraft M-12	Mixer	녹음 채널 구성
Sennheiser HMD 25-1	Headset	발성자 착용 헤드 셋
SoundBlaster Audigy 2	Sound Card	마이크 채널 샘플링
Pentium III PC	Computer	수집 컴퓨터

4.4 수집 시스템

대화음성의 수집 작업을 수행하기 위한 시스템의 구성이 필수적이다. 대화음성 수집 시스템은 2-Channel 방식을 사용했고 수집 시스템은 다음 그림 4와 같다.

표 5. 대화음성 수집 방식

구 분	WOZ 방식	2-Channel 방식
대화 참여자	2명	2명
대화 참여 대상	인간-기계	인간-인간
시나리오	필요	필요
대화	제한적	자연스러움

기존의 대화음성 수집 시스템의 방식은 인간과 기계(Wizard of Oz) 시뮬레이션을 이용한 방법은 대화가 제약되어 자연스러운 대화음성이 수집되지 못했다. 따라서 본 논문에서는 2-Channel 방식의 수집 시스템을 구성하여 자연스러운 대화음성을 수집하였다. 위의 표 5는 두 수집 방식을 비교한 예이다.

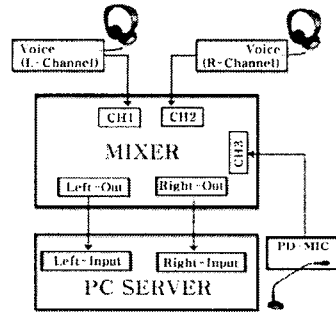


그림 4. 대화음성 수집 시스템

입력된 두 채널의 발화 음성은 녹음되어 16KHz로 샘플링하고 16bit로 양자화된 후, 저장하였다.

대화음성 자료는 시스템 발화와 사용자 발화가 자연스럽게 이루어지도록 생성하였다. 생성된 대화음성 자료는 시스템 발화 음성과 사용자 발화 음성, 각 89개의 텔레뱅킹 문장을 발화하였다. 한 대화에서 7개의 시나리오로 구분하여 발성해 7개의 웨이브 파일을 생성하였다. 총 10명의 화자가 대화에 참가하여 5번의 대화가 이루어졌다. 수집된 대화음성 자료의 양은 다음 표 6과 같다.

표 6. 수집된 음성 자료

발화자	발성 문장	웨이브 파일
시스템 발화	89	7
사용자 발화	89	
총	890	35

III. 대화코퍼스 전사

대화음성의 세그먼트 규칙과 전사 규칙은 명확하게 정의되지 못하고 있다. 각 연구소, 업체마다 서로 다른 세그먼트 규칙과 전사 규칙을 적용하고 외국어를 기준으로 한 규칙도 한국어의 어휘 문맥과 틀려 적용하기 어렵다. 따라서 본 논문에서 대화음성에 적용하기 위해 세그먼트 규칙과 전사 규칙을 정의하고 이를 구조적 문서인 XML DTD를 이용하여 표준화하고자 한다. 다음 [그림 5]와 같은 절차로 정의했다.

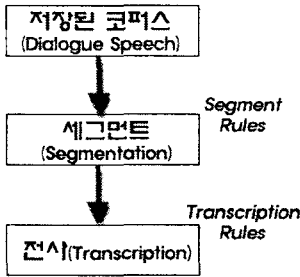


그림 5. 대화음성의 전사 과정

위 과정에서 전사 규칙은 대화의 층위에서 나타나는 주석 정보를 기반으로 철자 전사와 발음 전사의 규칙을 정의했다.

1. 세그먼트 규칙

대화음성의 세그먼트 규칙은 다음과 같다.

표 7. 세그먼트 규칙

항목	내용
차례	- 화자가 바뀌는 순서의 경계를 나타냄(text) - 대화에서 화자를 식별할 수 있는 방법
구간	- 발화에서 500m/sec 이상의 포우즈가 존재 - 포우즈 전후에서 세그먼트하여 표기
발성	- 동일 화자의 발성 내에서 문장 단위로 차례를 세분화
위치	- 휴지구간이나 호흡에 의한 음성의 자연스런 경계점 - 문장의 끝에 위치시킴

세그먼트에서 주의할 점은 발화의 중간에 위치할 수

없고 대화음성에서 음성이 중복된 구간이 없으면 동일 화자의 이전 발성 구간에 접칠 수 없다. 또한 파열음 ('교', '트', '크', '스')으로 시작되거나 끝나는 단어의 앞과 뒤는 소리가 잘리지 않게 주의해야 한다.

2. 전사 규칙

본 논문에서는 대화음성 데이터의 내용을 기술하기 위해 2개의 전사 규칙(철자전사, 발음전사)으로 구분하여 정의하였다.

2.1 철자 전사 규칙

대화의 층위의 주석 정보를 기반으로 철자 전사 규칙을 다음과 같이 정의한다.

가. 철자(Orthography)전사

대화음성을 글자 그대로를 텍스트로 표현한 것이며 전사에서 가장 기본이 된다. 화자의 발성된 음성은 철자 전사에서 한글 표준 맞춤법에 따른다. 대화음성의 철자 전사는 차례로 화자의 발성을 구분한다. 텍스트와 차례의 특성은 다음과 같다.

표 8. 철자 주석의 표기

구분	단위	표기	예
철자	text	없음	자연스럽게 발화된 하나의 대화 표본
	turn	없음	A: 통장 비밀번호를 말씀해주세요. B: 삼사오말이야.
	표준철자/발음에서 벗어나는 발음	+	텔레뱅킹을 +신청했거든
	숫자 발음	한글	비밀 번호는 삼사오칠이야

나. 주석 정보 표기

문장의 각 부분 사이에 표시하여 논리적인 관계와 문장의 정확한 의미를 전달하기 위한 수단으로 구두점을 사용한다.

언어가 가지고 있는 모호성을 피하기 위해 기호를 사용한다. 예를 들어, "NASA"는 원래 단어는 "National Aeronautics and Space"의 여러 단어들로 구성된 표현이다. 여러 단어로 구성되어 있지만 "NASA"는 하나의 단어로 발성한다. 여러 단어를 보통 한 단어로 발성

하지만 개별 문자로 발성하는 두 음자(Acronyms)들의 경우에는 끊어서 발성한다. 예를 들어, "F·B·I" 단어의 원래 단어는 "Federal Bureau of Investigation"의 여러 단어들로 구성된 표현이다.

표 9. 기호 주석의 표기

구분	단위	표기	예
기호	여러 단어를 하나의 단어로 발성	@	이번 비용은 @NASA 차례입니다.
	두 음자의 개별문자	~	이 계좌는 ~FBI 계좌입니다.
	고유명사	^	이 텔레뱅킹은 ^KOREA 은행입니다.
	불완전한 단어	-	그- 그것도 잔액 조회야.
	잘못 발성된 단어	+	텔레뱅킹으로 자동이체 +하란다.
특이한 형태의 단어	*	*오케버리 텔레뱅킹을 할게.	

또한, 말하기에 앞서 "어", "그러니까", "거기" 등의 앞세우는 간투어, 사람 이외에 현상에서 발생하는 소리(잡음) 등도 구분 대상이다. 이밖에도 대화 상호간에 참가자의 중복(중첩)된 대화의 경우 아래와 같이 표기한다.

표 10. 잡음/중첩 주석의 표기

구분	단위	표기	예	
잡음	사람 잡음	웃음	(laugh)	전화 예약 서비스를 (laugh)를 신청할래.
		기침	(cough)	번호(cough)는 일사오팔이야.
		재채기	(sneeze)	텔레뱅킹 은행은 농협(sneeze)이야.
		숨소리	(breath)	{breath}텔레뱅킹 비밀번호는 삼사오팔이야
	입술 소리	(lip smack)	아파트는 동신아파트이고 (lip smack)동은 육 동 백 칠호야.	
사람이 아닌 잡음	단발성 잡음	[noise]	전화번호[noise]는 육칠삼에 사사구이야	
	구간 잡음	[noise] [/noise]	전화번호[noise]는 육칠삼에 사사구이고 날짜는 칠월 3일이야.[/noise]	
중첩	중첩의 시작과 끝	<o> /<o>	A:텔레뱅킹 사용자 <o>번호를 알고 싶어.</o> B: <o>일삼 이 팔 칠 오 구 야.</o>	

그밖에도 대화에서는 완전하지 못한 발성을 위한 불완전단어, 잘 못 발성된 단어 불완전한 단어, 알 수 없는

음성, 알 수도 없고 추측도 어려운 음성, 전사하기 어려운 부분을 포함한 음성, 재시작, 다른 언어 사용 음성 등이 존재한다.

2.2 발음 전사 규칙

발음 전사는 대화음성 데이터의 음향적 내용을 기술하며 화자의 발성 내용을 소리 나는 형태로 정의한다. 발음 전사는 들리는 음성을 가장 정확하게 소리 나는 대로 적는 것이 목적이다. 여기에는 받침소리, 중성+초성의 형태 등의 다양한 발음상에서 나타나는 정보의 전사를 의미한다.

표 11. 전사가 어려운 대화음성 주석/기타 표기

구분	단위	표기	예
전사가 어려운 음성	알 수 없는 음성	((텍스트))	자동이체 ((어)) 언제 ((근데)) 했냐.
	알 수도 없고 추측도 할 수 없는 음성	(())	수표번호(())
	전사하기 어려운 음성	[[skip]]	언제 은행에 [[skip]] 갈래.
기타	재 시작	—	은행 — 은행이 아니고
	다른 언어를 사용	<language 텍스트> (언어를 알고 전사 할 수 있는 경우) <language (())> (언어는 알지만 전사 할 수 없는 경우)	<English telebanking you know?> <English (())>

IV. 전사 정보의 구조적 문서화

1. 대화음성 전사 시스템

LDC의 Transcriber 전사 시스템을 사용하여 수집된 대화음성을 전사하였다. 전사 과정은 그림 6과 같다.

전사과정은 수집된 대화음성에 세그먼트 규칙을 적용하여 발성 구간을 세그먼트로 구분하고 전사 규칙을 적용해서 전사하였다. 위와 같은 방법으로 텔레뱅킹 영역에서 대화코퍼스를 구축하고 대화음성 및 언어 자료를 생성하였다.

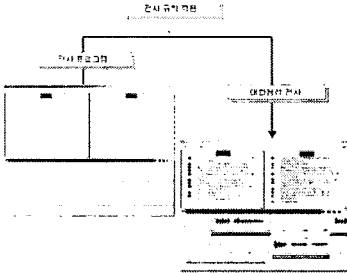


그림 6. 전사과정 및 응용 프로그램 형태

사용자 발화의 전사도 시스템 발화의 전사와 동일하다. 그림 7은 Transcriber 시스템의 전사 편집 에디터에서 전사된 자료로 대화의 주석 정보에서 잡음의 입술 소리와 단발성 잡음, 구간 잡음을 표기하여 전사하였다.

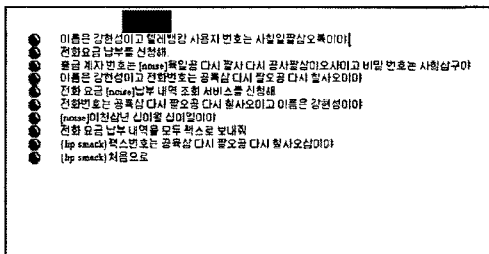


그림 7. 사용자의 발화 전사

수집된 대화음성의 전사 자료는 시스템 발화 음성과 사용자 발화 음성을 각각 전사하였다. 구축된 전사 자료는 시스템 발화 음성과 사용자 발화 음성을 전사해서 한 대화에서 7회 전사하였으며 이를 총 5회 반복하였다. 따라서 35개의 전사파일을 생성하였다.

2. XML 문서 구조화

대화코퍼스의 다양한 주석 정보를 의미태그로 선정하여 표기 방안을 표준화 하였다. 또한 이를 표현을 위해 XML을 사용하였다.

표 12. 화자/텍스트 정보

Tag Name	Attribute Name	Meaning
speaker	/name/	이름 정보
	/sex/	성별 정보
	/region/	지역 정보
	/date/	발성날짜 정보
	/language/	언어 정보
Tag Name	Attribute Name	Meaning
text	none	실질적인대화 정보
utterance	/who/	발화자 정보
turn	none	한 화자의 발화 정보

2.1 태그 정의

주석 정보를 XML로 표기하기 위해서는 먼저 의미태그를 선정하고 DTD를 정의해야 한다. 대화코퍼스의 화자 정보와 주석 정보의 의미 태그는 TEI(Text Encoding Initiative)와 LDC(Linguistic Data Consortium)의 기준을 기반으로 작성하였다[7]. 현재 구성된 DTD는 화자, 텍스트, 전사 등 3개의 정보만을 의미태그와 속성으로 분류하였다.

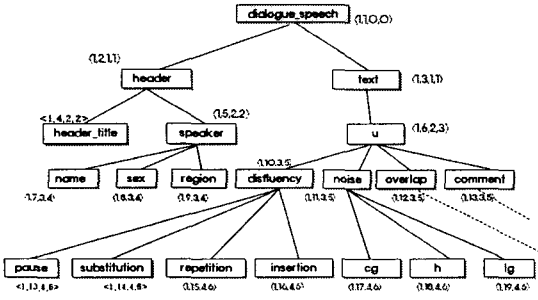
표 13. 전사 정보

Tag Name	Attribute Name	Meaning
disfluency	/pause/	간투사표현
	/substitution/	대용 표현
	/repetition/	반복 표현
	/insertion/	삽입 표현
	/deletion/	삭제 표현
noise	/cg/	기침소리 표현
	/h/	숨쉬는 소리 표현
	/lg/	웃음소리 표현
	/ls/	입술소리 표현
	/sp/	숨소리 표현
	/other/	기타 잡음 표현
overlap	/overlap_length/	중복 표현
comment	/P/	마침표 표현
	/C/	쉼표 표현
	/Q/	물음표 표현
	/E/	느낌표 표현
foreign	/lang/	외국어 표현
verbalism	/origin/	축약되기 전 단어 표현
dialect	/standard/	방언 표현

2.2 DTD 선정

대화코퍼스 DTD 전체적인 구조는 header DTD와 text DTD로 분류하여 구조화하였다. header DTD에

는 화자 정보 및 대화 코퍼스 제목 정보를 나타내며 text DTD는 실제 전사 정보를 나타낸다.



< DTID, NID, Depth_ID, ETID >

-DTID: Document Type ID / -NID: Node ID / -Depth_ID: Depth ID

그림 8. DTD 구조 다이어그램

구조화된 대화코퍼스 정보 초기에 header와 text, 그리고 text의 서브 노드 front body, back의 부분으로 구분하였으나, 부분적 DOM 트리 생성이 가능하기 위해서 모든 트리의 노드를 재 색인화 하였다. 이 색인에는 엘리먼트 이름, 엘리먼트 속성 이름, 속성 인스턴스 등의 정보로 구성하였다. DTID는 구조화된 단일 문서의 파일명을 표시하기 위한 색인 정보이다. NID는 각 노드가 가지는 고유한 노드 ID를 의미하고, Depth_ID는 각 노드가 트리 구조에서 나타나는 레벨(깊이)을 표시하기 위한 색인이다. 이 깊이 색인은 특정한 범위 내에서 특정 상위 노드(parent)에 대한 하위 노드(child)의 검색이 가능하도록 하기위해 필요한 색인이다. ETID는 엘리먼트가 갖는 속성 정보를 표시하기 위한 색인으로 Depth_ID와 함께 특정 범위 내에서 존재하는 음성 정보의 특징을 검색용 색인이다.

이렇게 구성된 구조화 된 문서는 기능별로 저장 및 색인이 가능하고 문서 정보, 엘리먼트 정보, 깊이 정보와 구조적 관계에 대한 표현이 가능하며 이를 중심으로 대화 음성 정보의 구조화된 문서 저장 시스템을 설계한다.

그림 8은 DTD의 구성의 일부를 나타내며 그림 9는 의미 태그를 DTD로 정의하여 XML 편집기로 표현한 일부이다. 이러한 DTD 설계와 함께 저장시스템을 구축

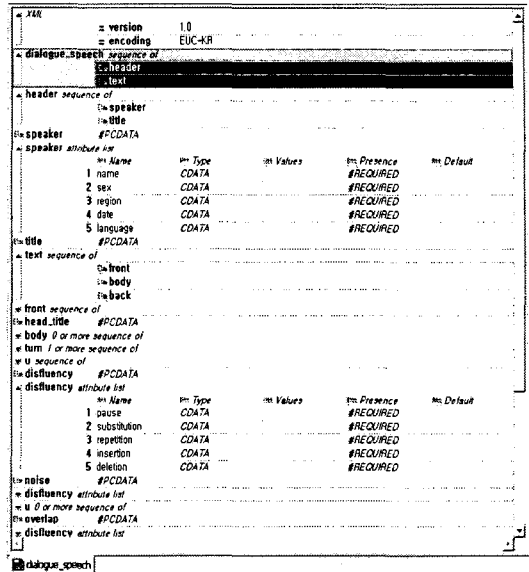


그림 9. DTD 설계

하여 구조적 문서의 웹 서비스가 가능하도록 XML 문서로 작성하고자한다[8].

2. 저장 시스템 설계

구조화 된 대화 음성 정보 문서를 바탕으로 다음과 같은 저장 시스템을 설계한다. 저장시스템의 구성도 그림 10과 같다.

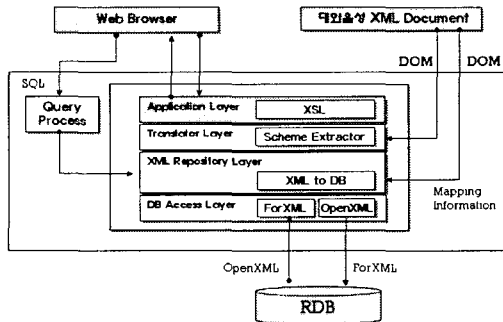


그림 10. 대화코퍼스 저장 시스템 구성도

세부 모듈로 구성된 구조적 문서는 DOM을 통해 트리 형태로 구조화되어 DB에 저장한다. 저장된 구조적

문서는 Schema Extractor에서 질의를 통해 입력된 파라미터 및 테이블 정보를 분석한다. 이를 통해 얻어진 노드별 색인 정보는 XML Repository에 전달되고 여기서 해당정보를 DB에 저장하게 된다.

V. 결론 및 향후과제

대화 음성 연구에는 다양한 언어정보 및 대화코퍼스가 필요하다. 대화 음성의 언어모델과 음향모델의 성능 평가에 레이블링된 정보는 인식 향상에 큰 영향을 끼친다. 또한 대량의 대화코퍼스 구축을 위해서는 체계적인 절차와 방법을 정립하는 것이 중요하며 구축된 대화코퍼스는 효율적으로 관리하고 사용자가 이를 보다 더 쉽게 활용할 수 있도록 하는 것이 바람직하다.

본 논문에서는 대화 영역을 텔레뱅킹으로 한정하여 코퍼스를 구축, 특정정보의 레이블링, 마지막으로 이를 표준화된 레이블링 데이터로 사용하기 위해, TEI를 기반으로 한 구조적 문서화가 가능하도록 대화코퍼스 정보 표준 DTD를 설계하였다. 또한 설계된 DTD를 기준으로 대화음성 특징 정보를 XML 기반 구조적 문서로 저장할 수 있는 저장 스키마를 설계하여 정확하고 효율적으로 관리할 수 있는 저장 시스템 구조를 설계하였다. 이러한 저장시스템을 통해 각 문서에서 가지고 있는 공통적으로 나타나는 특징 정보의 검색이 가능하며, 대화 음성 특징연구에 매우 유익한 정보 추출이 가능하다.

향후 연구 과제로는 저장시스템의 구축을 통한 검색이 가능한 웹서비스와 대화코퍼스에 나타나는 특성 및 레이블링 정보의 확대뿐만 아니라 효율적으로 이용할 수 있는 응용 프로그램에도 적용할 수 있도록 해야 할 것이다.

참고 문헌

[1] Akira Kurematsu et al, "Identification of Utterance Intention in Japanese Spontaneous Spoken Dialogue by Use of Prosody and

Keyword Information," ICSLP, pp. 98-101, 2000.

- [2] 이용주 외, "국어정보처리기술개발 한국어 음성 DB구축에 관한 연구", 한국과학기술원, 1997.
- [3] <http://www ldc. upenn. edu>
- [4] Li Aijun et al, "Spontaneous Conversation Corpus CADCC," ORIENTAL COCOSDA WORKSHOP, 2001.
- [5] 남승석, "대화 음성 시스템을 위한 효과적인 DB 구축", 서강대학교 컴퓨터학과 석사 논문, 2002.
- [6] D. Langmann et al, "The French Telephone Speech Data Collection-PRAT Of The European SPEECHDAT(M) Project," ICSLP. Vol.4, pp. 1918-1921, 1996.
- [7] 이용주 외, "해의 음성 DB 구축 동향", 한국음향학회음성통신 및 신호처리워크샵논문집 제12회, pp. 253-260, 1995.
- [8] 양옥렬, "시소러스를 이용한 구조적 문서의 탐색을 위한 깊이중심분할 인덱싱에 관한 연구", 원광대학교 박사논문, 2002.

저자 소개

양 옥 렬(Ok-Yul Yang)

정회원



- 1995년 2월 : 원광대학교 컴퓨터 공학과(공학사)
- 1997년 2월 : 원광대학교 컴퓨터 공학과(공학석사)
- 2002년 2월 : 원광대학교 컴퓨터 공학과(공학박사)

- 1998년~2000년 : BNS MediaTech. 대표
 - 2000년~2002년 : (주)휴먼미디어테크 연구소장
 - 2003년 3월~현재 : 혜전대학 컴퓨터멀티미디어계열 전임강사
- <관심분야> : 과학기술 콘텐츠, IT, 문화 콘텐츠

강 창 규(Chang-Qui Kang)

정회원



- 2001년 2월 : 원광대학교 수학과 (이학사)
 - 2004년 2월 : 원광대학교 컴퓨터 공학과 (공학석사)
- <관심분야> : 과학기술 콘텐츠, IT, 문화 콘텐츠

남 명 우(Myung-Woo Nam)

정회원



- 1992년 2월 : 서울시립대학교 제어계측공학과(공학사)
- 1994년 2월 : 서울시립대학교 전자공학과(공학석사)
- 2001년 8월 : 서울시립대학교 전자공학과(공학박사)

- 2002년~2003년 : LG 이노텍 선임연구원
 - 2003년 3월~현재 : 혜전대학 디지털전자디자인과 전임강사
- <관심분야> : 음성인식, 신호처리