

홍 기 형 교수지도
석사학위청구논문

음성 DB를 위한 메타데이터
구축 및 관리 도구

2006

성신여자대학교 교육대학원
교육학과 전자계산교육전공
주 영 희

음성 DB를 위한 메타데이터
구축 및 관리 도구

홍 기 형 교수지도

이 논문을 석사학위논문으로 제출함

2006년 5월

성신여자대학교 교육대학원
교육학과 전자계산교육전공
주 영 희

인 준 서

주 영 희의 석사학위 논문으로 인준함.

심사위원 _____인

심사위원 _____인

심사위원 _____인

성신여자대학교 교육대학원

논문 개요

본 논문에서는 산업용 음성 DB를 위한 메타데이터 표준안을 제안하고, 산업용 음성 DB 메타데이터 관리도구를 설계 및 구현하였다. 메타데이터는 음성 DB의 재사용과 이식성을 증대시키기 위한 것이다. 제안하는 메타데이터는 XML 기반으로 엘리먼트들의 그룹핑과 계층구조를 도입하여, 메타데이터의 구조를 체계화하여 객체지향 방법으로 설계하였다.

음성 DB 메타데이터를 위한 관리도구는 대상 음성 DB 메타데이터 검색기능과 음성 DB 메타데이터 생성 및 수정기능을 제공한다. 검색기능은 기본정보뿐만 아니라 화자, 채널, 발화목록에 따라 사용자가 원하는 화자정보, 음성파일 및 전사정보의 검색을 지원한다. 또한 새로운 음성 DB 구축 시에 사용자가 쉽고 효율적으로 음성 DB 구성 요소를 각 정보 별로 저장, 수정할 수 있도록 음성 DB 메타데이터 생성 및 수정 기능을 제공하도록 구현하였다.

목차

I. 서론	1
II. 관련 연구	3
III. 산업용 음성 DB 메타데이터	5
1. 산업용 음성 DB 모델	5
2. 산업용 음성 DB 메타데이터 모델	7
3. XML 기반 음성 DB 메타데이터	9
3.1. 기본정보 (General 엘리먼트)	10
3.2. 수집채널 (Channel 엘리먼트)	12
3.3. 화자 정보 (Speaker 엘리먼트)	13
3.4. 발화목록 정보 (UtteranceList 엘리먼트)	15
3.5. 음성파일 정보 (File 엘리먼트)	16
3.6. 전사 정보 (Transcription 엘리먼트)	18
3.7. 구현 예	19
IV. 산업용 음성 DB 메타데이터 관리도구 설계 및 구현	24
1. 검색기능 설계 및 구현	24
1.1. 검색기능 설계	24
1.2. 검색기능 구현	26
1.2.1. 구현 환경	27
1.2.2. 운영 방식	28
1.3. 검색기능 구현 결과	28
2. 구축, 수정기능 설계 및 구현	37
2.1. 구축 및 수정기능 설계	37

2.2.	구축 및 수정기능 구현.....	39
2.2.1.	구현 환경	41
2.3.	구축 및 수정기능 구현 결과	41
V.	결론 및 향후 과제.....	48

참고문헌

ABSTRACT

그림 목차

[그림 III-1] 산업용 음성 DB 메타데이터 구성	6
[그림 III-2] 음성DB 메타데이터 구성요소 및 관계도 (ER 모델).....	8
[그림 III-3] 음성 파일 하나와 다른 정보와의 관계도.....	8
[그림 III-4] 산업용 음성 DB 메타데이터 스키마 구조	9
[그림 III-5] General 엘리먼트 구조.....	10
[그림 III-6] Channel 엘리먼트 구조	12
[그림 III-7] Speaker 엘리먼트 구조.....	14
[그림 III-8] UtteranceList 엘리먼트 구조.....	15
[그림 III-9] File 엘리먼트 구조	16
[그림 IV-1] 검색기능 흐름.....	26
[그림 IV-2] Speech DB Search 대화창.....	28
[그림 IV-3] File Open 대화창	29
[그림 IV-4] 산업용 음성 DB 검색 대화창	30
[그림 IV-5] 기본 정보 문서.....	31
[그림 IV-6] 발화목록 검색 대화창.....	32
[그림 IV-7] 발화문장 선택 대화창.....	33
[그림 IV-8] 음성 DB 검색 대화창.....	34
[그림 IV-9] 음성 DB 검색 대화창.....	35
[그림 IV-10] 검색 결과 화면	36
[그림 IV-11] 검색 결과 화면	36
[그림 IV-12] 검색 결과 화면	37
[그림 IV-13] 구축 및 수정기능 흐름도	39
[그림 IV-14] 관리도구 구축 및 수정 전체 시스템 흐름도	40
[그림 IV-15] 구축 및 수정 메인 화면	41

[그림 IV-16] 파일 메뉴 대화창	43
[그림 IV-17] 음성 DB 열기 대화창	43
[그림 IV-18] 수집채널 정보 입력 대화창	44
[그림 IV-19] 수집채널 정보 입력 대화창	44
[그림 IV-20] 발화자 정보 입력 대화창	45
[그림 IV-21] 발화목록 정보 입력 대화창	45
[그림 IV-22] 발화목록 별 발화문장 입력 대화창	46

표 목차

[표 III-1] 기본정보 엘리먼트 및 속성.....	11
[표 III-2] 수집채널 엘리먼트 및 속성.....	13
[표 III-3] 화자 엘리먼트 및 속성	14
[표 III-4] 화자 엘리먼트 및 속성	15
[표 III-5] 발화목록 엘리먼트 및 속성.....	16
[표 III-6] 파일정보 엘리먼트 및 속성.....	17
[표 III-7] 전사정보 엘리먼트 및 속성.....	18
[표 III-8] XML 기반 음성 DB 메타데이터 예제	19
[표 III-9] XML 기반 음성 DB 메타데이터 예제	21
[표 III-10] XML 기반 음성 DB 메타데이터 예제	22
[표 III-11] XML 기반 음성 DB 메타데이터 예제.....	22
[표 III-12] XML 기반 음성 DB 메타데이터 예제.....	23
[표 IV-1] 구현 환경.....	27

I. 서론

산업용 음성 데이터베이스(이하 산업용 음성 DB)는 자동차, 로봇, 가전제품 등 다양한 산업 제품을 위한 음성 사용자 인터페이스(음성인식/합성) 기술의 개발에 있어서 필수적이며 가장 첫 단계에서 구축되어야 한다.

다양한 산업 제품의 사용 환경을 반영한 산업용 음성 DB의 효과적인 구축 및 활용은 음성 사용자 인터페이스 기술의 효율적인 개발과 성능향상에 있어 매우 중요하며, 다양한 산업 제품을 사용하기 위한 음성언어정보가 충실히 표현된 DB, 개발자가 기술개발에 용이하게 활용할 수 있는 구조화된 DB 구성이 요구된다. 그러나 지금까지 국내에서는 산업용 음성 DB 표준화에 대한 필요성은 충분히 인지하고 있었으나 구체적인 노력이 미미하였다. 최근 대량의 산업용 음성 DB 구축이 진행되고 있으나, 각기 다른 기관에서 구축한 DB의 호환성 문제가 발생하고 있다. 이에 따라 산업용 음성 DB의 호환성을 높이고, 활용성을 증대시키기 위해 메타데이터 표기방법 및 구조의 표준화가 절실히 필요한 시점이다.

각 업체, 연구소, 대학에서 구축한 산업용 음성 DB의 표기방법의 통일은 DB 유통 활성화를 꾀할 수 있어 국가적으로 자원의 효율적 활용이 가능하다. 구체적인 필요성은 다음과 같다.

첫째, 산업용 제품의 음성인식/합성 엔진 개발자들이 용이하게 다양한 음성 DB를 쉽게 활용할 수 있으므로, DB 사용의 편의성이 증대되어 DB 전처리에 소요되는 시간과 인력을 줄일 수 있다.

둘째, 각 업체, 연구소, 대학에서 개별적으로 구축한 산업용 음성 DB의 경우, 현재까지는 개별 기관별로 독자적인 정보 표현 양식을 사용함으로써, 유사한 DB의 중복 구축이 많으며, 상호 호환되지 않음으로써

타기관의 DB를 활용하기 위해서는 별도의 변환 프로그램을 개발하여야 하는 등의 문제가 있으나, 본 통일된 메타데이터의 사용을 통하여 이러한 문제를 해결할 수 있다.

셋째, 음성 DB의 경우에 그 크기가 클수록 음성처리 기술의 성능을 향상시키는 데 도움이 되므로, 표준 메타데이터를 통하여, 기 구축한 유사 DB를 통합하여 대량의 DB 구축이 가능하다면, 음성처리 기술의 성능 향상에 크게 기여할 것이다.

본 논문의 목적은 산업용 음성 DB 메타데이터 표준을 향후 확장성을 보장하기 위하여 세계적 공개 표준인 XML(eXtended Markup Language) 기반으로 제시하고, 음성 DB 메타데이터 관리도구의 설계 및 구현이다.

본 논문의 구성은 다음과 같다. II장에서는 음성 DB 표준화와 관련한 기존연구를 기술한다. III장에서는 제안하는 산업용 음성 DB 메타데이터를, IV장에서는 산업용 음성 DB 메타데이터 관리도구 설계 및 구현 내용에 대해서 각각 기술하였다. 마지막으로 V장에서 결론 및 향후 연구 과제를 제시하였다.

II. 관련 연구

국외의 경우 프랑스의 ELRA(The European Language Resources Association) [1]와 미국의 LDC(The Linguistic Data Consortium) [2]를 주축으로 음성 DB 구축 및 표준화를 꾸준히 수행해 오고 있다. ELRA는 표준화 대상으로 수집시스템, 수집환경, 음성 DB, 전사방법, 검증방법 등을 다양한 분야에서 수행하고 있다. LDC는 1970년대에 DARPA(Defense Advanced Research Project Agency)를 통하여 수집한 DB를 기반으로 음성 DB 표준화 작업을 수행하였다. 대학, 기업, 정부연구소의 컨소시엄으로 음성 DB, 텍스트 DB, 어휘목록 및 기타 연구개발을 위한 자원을 제작, 수집 및 보급한다. 구축된 데이터베이스의 보급뿐만 아니라 새로운 데이터를 적극적으로 제작하며 다른 나라 언어로 구성되는 음성 DB 제작 및 배포도 지원한다. SALT(Speech Application Language Tags) 포럼 [3]은 멀티모달 환경에서의 음성 인터페이스에 대한 표준화 작업을 수행 중이며, W3C 컨소시엄 [4]은 VoiceXML, Speech Recognition Grammar(SRGS), Speech Synthesis(SSML), Call Control(CCXML) 등의 W3C 음성 브라우저의 표준화가 진행 중이다. 그러나 국제 표준화 활동은 영어권의 음성 특성을 반영하고 있어, 우리말 음성의 고유한 특성을 반영해야 할 국내 표준화에 그대로 채택하여 사용하기에 적합하지 않은 부분이 있다 [5].

국내의 경우 음성 DB 메타데이터 표준안은 2003년도에 ETRI(한국전자 통신연구원) 음성/언어정보 연구센터에서 개발한 공통 음성 DB 표준안 [6] [7]이 있으나, 정보통신망 기반 음성 기술 개발을 위한 음성 DB 메타데이터 규격으로 개발되었다. 공통 음성 DB 표준안 [7]은 음성 DB 정보를 크게 기본정보, 음성정보, 전사정보,

화자정보, 환경정보, 파일정보, 기타정보로 나누었다. 공통 음성 DB 표준안은 국내에서 처음으로 시도한 음성 DB 메타데이터 규격이라는 점에서 의미가 있으나, 다음과 같은 단점이 있다.

첫째, 음성 DB 메타데이터를 단순히 2단계 구조로 정의하였다. 이러한 평면적인 구조는 음성 DB를 구성하는 객체 사이의 집합적 포함관계나 연관관계를 정확하게 표현하기에 부족하다. 공통 음성 DB 표준안에서 기술한 모델은 하나의 음성 파일에 대한 것인지, 음성 DB, 즉 다수의 음성 파일에 대한 것인지가 불명확하다.

둘째, 화자, 음성녹음 채널, 음성 파일 등 DB를 구성하는 객체의 식별자와 이를 참조하는 객체 참조 개념이 결여되어 있다. 그러므로 동일한 정보의 중복 및 특정 음성 파일의 화자 식별이 어렵다. 예를 들어, 발화자가 100명이고 발화 문장이 100개가 존재할 경우 메타데이터에 총 10,000개의 음성 파일에 대한 정보가 존재하게 된다. 이 경우 발화자 한 명당 발화 문장 100개씩을 메타데이터 문서에 중복하여 입력해야 한다.

본 논문에서 제안하는 산업용 음성 DB 메타데이터는 국제기관인 LDC, ELRA의 용어 및 분류정보를 참조하고 [7]의 모델을 보다 객체 지향 구조로 보완하였다.

III. 산업용 음성 DB 메타데이터

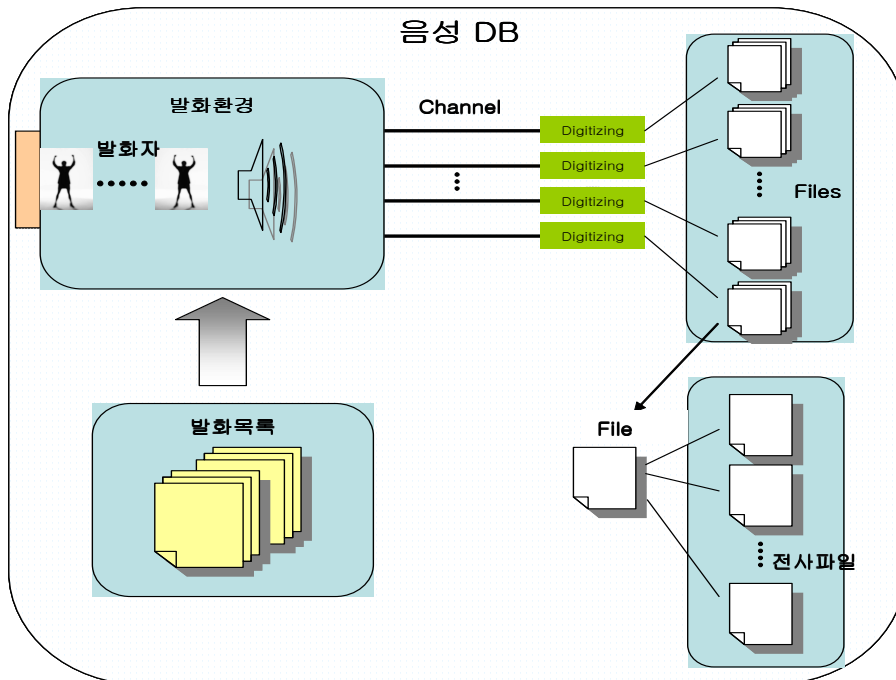
1. 산업용 음성 DB 모델

자동차, 완구, 가전 기기, 산업용 기기 등 다양한 소음 및 주변 환경에서 높은 성능의 음성 기술(인식/합성/인증/코딩/강화) 개발 및 시험을 위해서는, 해당 산업용 음성 기술이 사용되는 동일한 환경에서 대량의 음성 데이터베이스의 구축이 필수적이다. 즉, 자동차에서 사용하려고 하는 음성 인식 기술의 개발을 위해서는 자동차의 주행 환경에서 사용자가 수행할 수 있는 음성 명령을 녹음하여, 기술의 개발에 사용하려면 자동차 환경에서 성능이 높은 인식 기술을 개발할 수 있다. 또한 음성은 화자의 특성에 따라 동일한 환경이라고 하더라도 많은 차이가 있다. 사투리의 구사여부, 개인 구강 구조의 특성, 성별, 나이 등에 따라 동일한 환경, 동일한 단어를 발성하더라도 음성 신호에서 많은 차이를 보이므로, 화자 독립형의 음성 기술의 개발에서는 가능한 많은 수의 화자로부터 음성을 수집하여야 한다. 따라서 완구용 음성 기술을 위한 음성 DB, 자동차용 음성 기술을 위한 음성 DB 등과 같이 산업용 음성 기술 개발을 위해서는 해당 기기의 사용 환경에서 많은 수의 화자로부터 음성을 수집하는 것이 필수적이다.

산업용 음성 DB는 산업 응용에서 적합한 음성 기술(인식/합성/인증/코딩/강화) 개발 및 시험을 위하여, 다수의 화자로부터 녹취한 음성 파일의 집합으로 정의한다.

산업용 음성 DB 메타데이터는 산업용 음성 DB의 기본정보(구축 기관 정보 등), 음성파일의 디지털링 방법, 화자 정보 등과 같이 산업용 음성 DB에 대한 정보를 말한다.

본 논문에서 제시하는 표준안의 대상이 되는 산업용 음성 DB는 [그림 III-1]과 같은 구성을 전제로 한다.



[그림 III-1] 산업용 음성 DB 메타데이터 구성

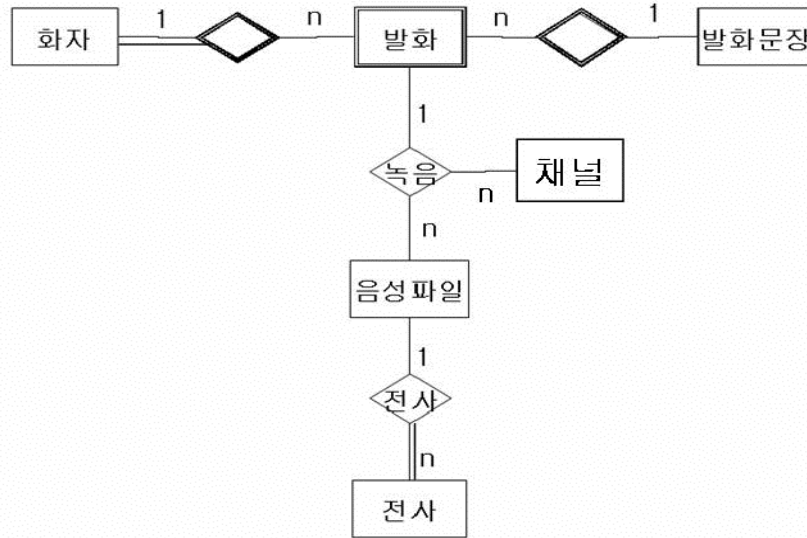
- 1) 음성 DB 내의 음성 파일은 다양한 발화환경에서 발화한 것을 수집한 것으로 가정하였다. **발화환경**은 방음실, 백화점, 자동차 안, 가정과 같은 화자의 주변 환경을 의미한다.
- 2) 하나의 음성 DB는 화자의 음성을 여러 가지 방법으로 수집하기 위해서 다수의 녹음 채널이 있을 수 있다. **채널**이란 녹음 시 사용하는 마이크, 디지털화 형식, 통신망(전화망, IP network 등), 발화환경 등을 의미한다.
- 3) 하나의 **음성파일**은, 화자가 발화목록의 한 문장 또는 단어를 발화한 음성을 특정한 채널로 녹음하고 디지털화하여 저장한 것이다.
- 4) 하나의 **발화목록**은 여러 개의 발화문장 또는 단어로 구성될 수 있다.

- 5) 하나의 음성파일은 여러 종류의 **전사정보**를 가질 수 있다.
- 6) **화자** 한 명은 다수의 발화목록을 발성할 수 있다.
- 7) 발화시점을 다르게 하여 화자는 같은 발화목록을 여러 번 반복 발성하여 녹음할 수 있다.

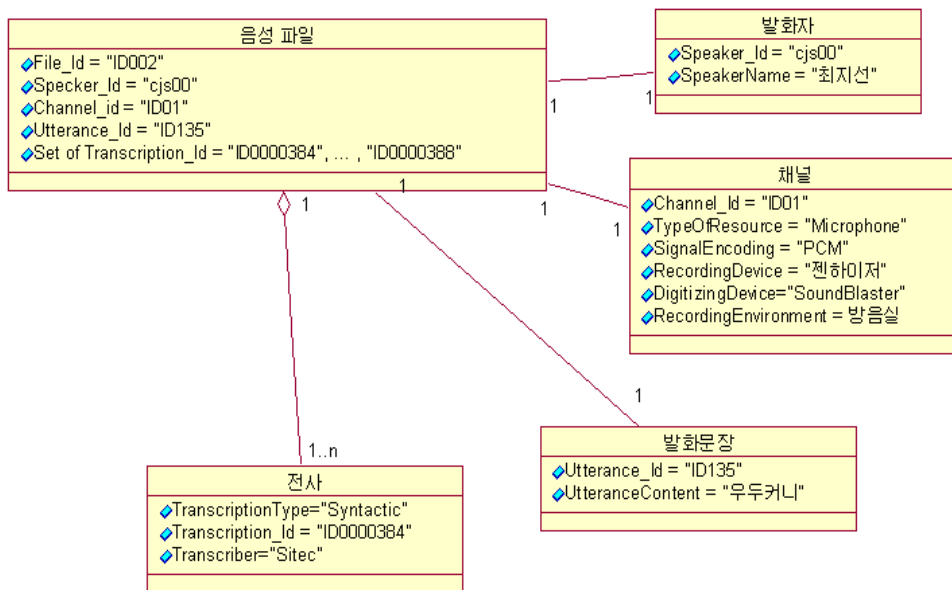
2. 산업용 음성 DB 메타데이터 모델

음성 DB의 구성요소는 화자, 발화문장, 채널, 음성파일, 전사정보이다. 음성 DB 메타데이터 구성요소들의 관계는 [그림 III-2]와 같다. 화자 한 명이 여러 개의 문장을 반복 발화할 수 있으므로 발화자와 발화는 1:n의 관계이다. 발화문장 하나는 여러 화자가 반복하여 발화하므로, 발화와 발화문장의 관계는 n:1이 된다. [그림 III-2]에서 '발화'는 화자가 발화하는 행위 자체를 하나의 개체로 바라보기 위한 약한 개체(weak entity)이다. 화자 한 명이 하나의 발화문장을 발화할 때, 다수의 채널로 녹음할 수 있으므로 채널별로 음성파일이 생성되어 발화와 음성파일은 1:n의 관계이다. 이렇게 생성된 한 개의 음성파일에는 전사정보가 전사의 종류에 따라 여러 개 존재할 수 있다. 따라서, 하나의 음성파일과 전사정보는 1:n의 관계를 가진다.

하나의 음성파일과 관련 있는 화자, 채널, 발화문장, 전사정보와의 구체적인 관계는 [그림 III-3]과 같다. 음성파일에는 식별자(*File_Id*)가 있으며, 누가 발화하였는지 나타내기 위해 발화자의 식별자 (*Speaker_Id*)를 참조한다. 어느 채널을 통해 녹음된 파일인지 나타내기 위해 채널의 식별자(*Channel_Id*)를 참조한다. 어떤 문장에 대한 발화인지를 나타내기 위해 발화문장 식별자(*Utterance_Id*)를 참조하며, 해당 음성 파일과 관련한 전사정보를 모두 참조(Set of *Transcription_Id*)하고 있다.



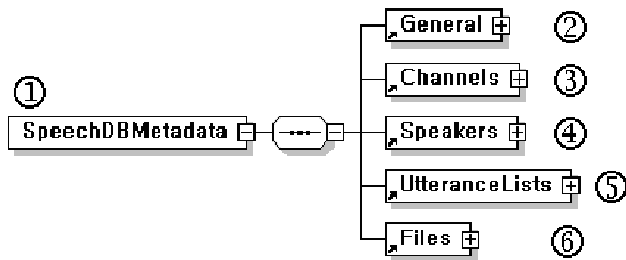
[그림 III-2] 음성 DB 메타데이터 구성요소 및 관계도 (ER 모델)



[그림 III-3] 음성 파일 하나와 다른 정보와의 관계도

3. XML 기반 음성 DB 메타데이터

앞 절에서 살펴본 바와 같이 산업용 음성 DB는 수집채널, 화자 정보, 발화목록, 음성파일, 전사 정보로 구성된다. 제안하는 메타데이터 규격은 기존의 SiTEC과 ETRI에서 추진한 표준화 작업[6][7]과 비교하여 XML[8]의 특징을 최대한 살릴 수 있도록 정의하였다. 화자, 채널, 발화목록, 파일, 전사 정보에 각각의 식별자를 두고 구성요소 사이의 관계를 식별자의 참조로 표시할 수 있도록 하였다. 또한 연관되는 엘리먼트들의 그룹핑과 계층구조를 도입하여, 메타데이터의 구조를 체계화하여, 향후 음성 DB의 검색이나 재사용이 용이하도록 설계하였다. XML 스키마로 정의한 메타데이터 구조는 [그림 III-4]와 같다. 본 메타데이터 정의에서 사용한 엘리먼트와 속성의 이름은 ELRA와 LDC에서 부분적으로 참고하였다[1][2]. [그림 III-4]의 ①번 **SpeechDBMetadata**는 메타데이터인 XML 문서의 루트 엘리먼트이다.

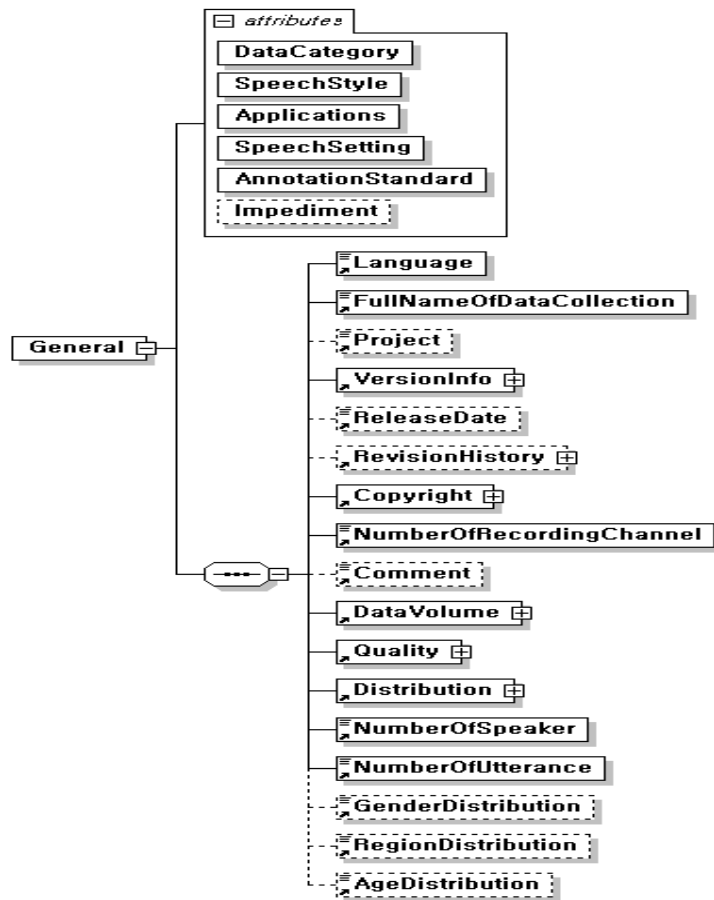


[그림 III-4] 산업용 음성 DB 메타데이터 스키마 구조

루트 엘리먼트 하위에 총 5 개의 엘리먼트가 존재하며, 5 개의 엘리먼트는 [그림 III-2]의 화자, 음성파일, 채널, 발화문장에 해당하는 **Speakers, Files, Channels, UtteranceLists** 와 음성 DB의 저작권 등과 같은 기본정보 기술을 위한 **General** 엘리먼트로 구성된다.

i. 기본정보 (General 엘리먼트)

[그림 III-4]의 ②번 General은 해당 음성 DB의 기본정보를 위한 엘리먼트이다. 기본 정보는 전체 음성 DB에 해당하는 공통 정보, 저작권에 관한 정보, 배포할 시 필요한 정보 및 발화자 정보나 발화목록 정보로부터 유도될 수 있는 정보로 나누어질 수 있다. General 엘리먼트 구조는 [그림 III-5]과 같다.



[그림 III-5] General 엘리먼트 구조

[표 III-1]은 기본정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다. 표에서 속성은 이탤릭체로 표기하였고 엘리먼트는 정자로 표기하였다.

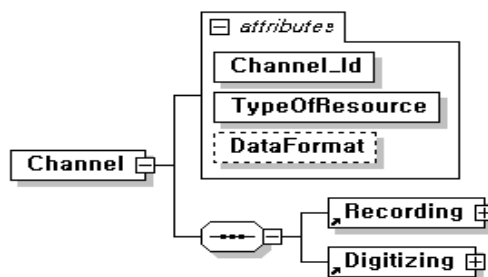
분류	엘리먼트 및 속성	내용
음성 DB 공통 정보	Language	음성DB 발성에 사용된 언어를 기록하는 것으로 ISO 3166 표기 방식
	FullNameOfDataCollection	음성 DB의 이름
	VersionInfo	버전과 음성DB 버전별로 녹음 구축기간
	NumberOfRecordingChannel	동시녹음 채널 수
	DataCategory	음성 DB의 분류 (가능한 값 : IsolatedWord, Digit, ReadSpeech, Psuedo ReadSpeech, Interaction, Text, Verification, Speech Synthesis, SpeakerCircumstance, Other, Orthography, Accent, FreeAnswer Question, Syllables, VCVSequence, IsolatedDigits, Continous Sentences, PRS, PBS, YesNoQuestions)
	SpeechStyle	발화할 내용의 스타일, 종류 (가능한 값 : Spontaneous, Read, Elicited, Prepared, Prompted, Other)
	Applications	상용서비스를 고려할 때 본 음성 DB의 응용 분야 (가능한 값 : DiscourseAnalysis, VoiceControl, LanguageIdentification, SpeakerIdentification, SpeakerVerification, SpeechRecognition, Spoken-대화창ueSystems, Other)
	SpeechSetting	단독발화(monologue), 대화체(대화창ue), 3인 이상 대화체(multilogue) 구별
	Impediment	언어장애 분류, 마비말 장애 (뇌성마비), 청각장애, 음성장애(가능한 값 : Dysarthrias, HearingImpairments, SpeechImpediment, Other)
저작권	Authors	음성DB 저작자의 이름
	Distributor	음성DB 설계, 구축을 주관한 기관명
	Performer	음성DB를 최종 배포한 기관명
배포시 필요한 정보	DataVolume	DB전체 크기 및 재생시간
	Distribution	DB압축 여부, 배포방법, 배포분류
	Quality	품질상태, SNR, ClippingRate와 같은 음성 DB 품질
유도 정보	NumberOfSpeaker	발성한 총 발화자 수
	NumberOfUtterance	총 화자가 발성한 총 발화 횟수

[표 III-1] 기본정보 엘리먼트 및 속성

ii. 수집채널 (Channel 엘리먼트)

[그림 III-4]의 ③번 Channels는 음성 DB 수집에서 사용한 다수의 채널에 대한 정보를 위한 엘리먼트이다. Channels 엘리먼트의 하위 엘리먼트로 Channel 엘리먼트를 다수 가질 수 있으며, 각 Channel 엘리먼트는 특정 수집채널에 대한 정보를 기술한다. Channel 엘리먼트 구조는 [그림 III-6]과 같다.

수집채널은 화자가 발생하는 음성을 어떤 방식으로 조작하고, 저장하였는지를 의미한다. 하나의 음성 DB는 여러 개의 채널로부터 구축될 수 있다. 따라서 다채널 설정이 가능하고, 각 채널 별로 레코딩 방식과 디지털링 방식을 다르게 할 수 있다. 각 채널은 녹음 시 발화환경을 달리하여 설정할 수 있다. 발화환경은 녹음환경이 사무실, 야외, 백화점 등과 같은 어떤 장소인지를 명시한다. *Channel_Id*는 채널을 구별할 수 있는 식별자이다. 수집채널은 크게 디지털링 방법과 레코딩 방법으로 분류하여 정의하였다.



[그림 III-6] Channel 엘리먼트 구조

[표 III-2]는 수집채널의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

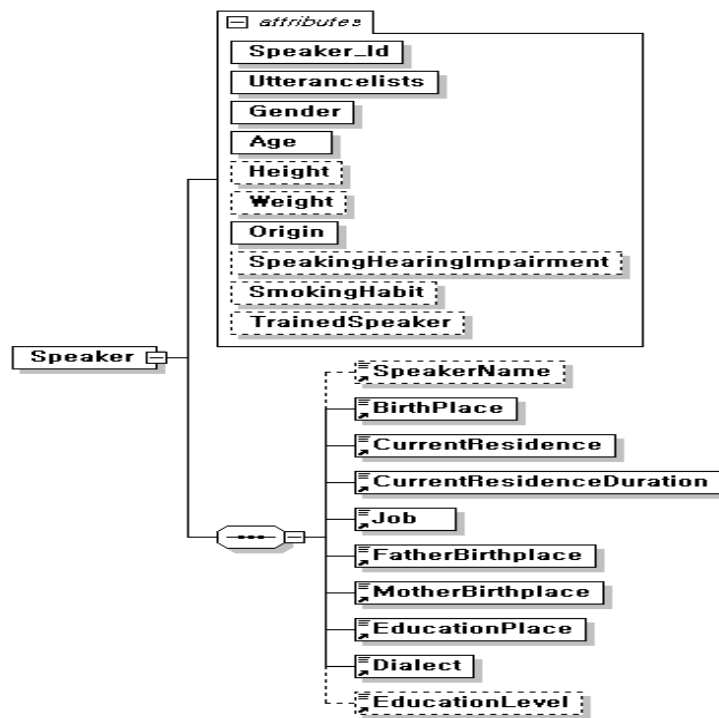
분류	엘리먼트 및 속성	내용
채널 정보	<i>Channel_Id</i>	채널 식별자
	<i>TypeOfResource</i>	수집환경 중 유선/무선전화망, VoIP망 등 통신망 종류 (가능한 값 : PC, PDA, Telephone, Cellular Phone, VoIP, Microphone, Broadcast News, Other)
레코딩	Recording Environment	녹음 시 발화환경
	RecordingDevice	녹음 시 사용된 마이크, 헤드셋 등 입력장치의 모델명과 주요규격
	TalkingDistance	원거리 음성입력의 경우, 화자와 입력장치와의 발성거리
디지털타이징	DigitizingDevice	입력장치로부터 녹취된 음성을 디지털타이징하기 위해 사용된 A/D장비의 모델명과 주요규격
	SignalEncoding	디지털 음성데이터의 인코딩 방법
	ByteOrder	상위 바이트와 하위 바이트의 순서
	BitsPerSampling	음성 샘플 당 차지하는 비트 수
	SamplingRate	음성데이터의 샘플링 주파수

[표 III-2] 수집채널 엘리먼트 및 속성

iii. 화자 정보 (Speaker 엘리먼트)

[그림 III-4]의 ④번 **Speakers**는 화자 정보를 나타내는 엘리먼트이다. **Speakers**는 다수의 **Speaker** 엘리먼트를 가질 수 있고, 하나의 **Speaker** 엘리먼트는 특정 화자의 정보에 해당한다. **Speaker** 엘리먼트 구조는 [그림 III-7]과 같다.

화자는 여러 명이 존재할 수 있고 각각의 화자는 *Speaker_Id*로 구분한다. 발화자 한 명은 여러 개 발화목록을 발화할 수 있다. 특정 화자가 발화한 목록은 *Utterancelists* 속성으로 나타낸다. *Utterancelists* 속성은 발화목록의 식별자인 *Utterancelist_Id*를 집합 참조한다.



[그림 III-7] Speaker 엘리먼트 구조

[표 III-3]과 [표 III-4]는 발화자 정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

분류	엘리먼트 및 속성	내용
화자 기본정보	<i>Speaker_Id</i>	화자의 식별자. 화자명의 영문 이니셜 3자리와 숫자 2자리 조합으로 표현. 영문 이니셜이 중복될 경우, 2자리 숫자로 구분
	SpeakerName	발성화자의 한글 이름
	<i>Gender</i>	발성화자의 성별
	<i>Age</i>	발성화자의 나이
	Birthplace	발성화자의 태어난 지역
	EducationPlace	화자가 초등학교 재학 시 거주했던 지역명
	CurrentlyResidence Duration	발성화자의 현 거주 지역에서의 거주기간

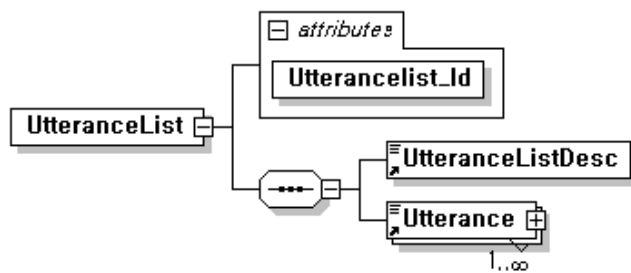
[표 III-3] 화자 엘리먼트 및 속성

분류	엘리먼트 및 속성	내용
화자 기본정보	Job	화자의 직업
	Dialect	화자가 사용하는 방언
	<i>Height</i>	화자의 신장
	<i>Weight</i>	화자의 몸무게
참조	<i>Utterancelists</i>	해당화자가 발화한 발화목록 식별자 (<i>Utterancelist_Id</i>)의 집합참조
화자 특수정보	<i>Origin</i>	외국인인지 내국인인지 표기
	<i>TrainedSpeaker</i>	훈련된 발화자 여부
	<i>SmokingHabit</i>	화자의 흡연 습관
	<i>SpeakingHearing Impairments</i>	화자가 듣거나 말하는 데 장애여부

[표 III-4] 화자 엘리먼트 및 속성

iv. **발화목록 정보 (UtteranceList 엘리먼트)**

[그림 III-4]의 ⑤번 **UtteranceLists**는 DB 수집에서 사용한 발화목록 정보이다. 하나의 DB 수집에서 다수의 발화목록이 있을 수 있으므로, 하위 엘리먼트로 다수의 **UtteranceList** 엘리먼트를 가진다. **UtteranceList** 엘리먼트의 구조는 [그림 III-8]과 같다. 하나의 발화목록은 발화문장의 집합이다. 발화목록의 식별자는 *Utterancelist_Id*이다. 발화목록에 대한 설명(**UtteranceListDesc**)과 발화목록을 구성하는 발화문장의 수만큼 각 발화문장의 정보를 나타내는 **Utterance** 엘리먼트로 구성된다. 하나의 발화문장은 *Utterance_Id*로 식별한다.



[그림 III-8] UtteranceList 엘리먼트 구조

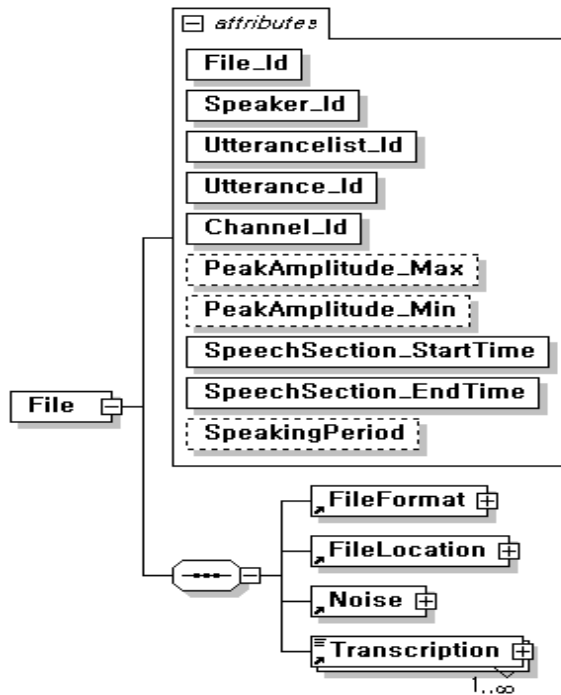
[표 III-5]는 발화목록 정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

분류	엘리먼트 및 속성	내용
발화목록	<i>Utterancelist_Id</i>	발화목록 식별자
	UtteranceListDesc	발화목록 설명
발화문장	Utterance	발화 단어, 문장의 실제내용
	<i>Utterance_Id</i>	발화문장 식별자

[표 III-5] 발화목록 엘리먼트 및 속성

v. 음성파일 정보 (File 엘리먼트)

[그림 III-4]의 ⑥번 Files는 DB를 구성하는 각 음성파일에 대한 정보를 담고 있는 File 엘리먼트를 DB에 존재하는 음성파일의 개수만큼 하위 엘리먼트로 가진다. File 엘리먼트 구조는 [그림 III-9]와 같다.



[그림 III-9] File 엘리먼트 구조

File은 음성 DB에 존재하는 각 음성파일에 대한 정보를 기술하는 엘리먼트이다. 해당 음성파일이 누가(*Speaker_Id*) 어떤 문장(*Utterancelist_Id* 및 *Utterance_Id*)을 발화한 것이며, 어떤 채널(*Channel_Id*)로 디지털이징한 것인지를 기술한다. 음성파일 자체정보로는 식별자(*File_Id*)와 저장위치(*FileLocation*) 등의 물리적 정보와, 음성구간을 나타내는 *SpeechSection_StartTime*과 *SpeechSection_EndTime* 애트리뷰트, 그리고 음성파일에 존재할 수 있는 다수의 잡음구간을 표시하기 위하여 **Noise** 엘리먼트를 가진다. 그리고 해당 음성파일에 대한 전사정보를 위한 다수의 **Transcription** 엘리먼트를 가질 수 있다. [표 III-6]는 파일정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

분류	엘리먼트 및 속성		내용
파일 기본정보	<i>File_id</i>		음성파일 식별자
	SpeakingPeriod		녹음 주기와 관련하여 파일별로 발화시점 정보
물리적 정보	FileLocation		음성파일명과 위치
	FileFormat		파일의 포맷, 헤더크기와 파일크기
	<i>SpeechSection_StartTime</i>		음성이 시작하는 시간
	<i>SpeechSection_EndTime</i>		음성이 끝나는 시간
잡음정보	Noise	NoiseSection	잡음이 시작하는 시간과 끝나는 시간
		<i>NumberOfNoise</i>	1개의 음성파일 내 다양한 잡음이 포함되어 있을 경우 잡음의 개수, 잡음개수의 수만큼 잡음구간 (Noise Section) 존재
전사	Transcription		음성파일에 대한 전사정보
참조	<i>Speaker_id</i>		해당 음성 파일의 발화자의 식별자 참조
	<i>Channel_id</i>		수집채널의 식별자 참조
	<i>Utterancelist_id</i>		발화목록 식별자 참조
	<i>Utterance_id</i>		발화문장 식별자 참조

[표 III-6] 파일정보 엘리먼트 및 속성

vi. 전사 정보 (Transcription 엘리먼트)

[그림 III-9]의 Transcription 엘리먼트는 상위 엘리먼트인 File이 기술하고 있는 음성파일에 존재하는 전사정보를 위한 엘리먼트이다. 하나의 음성파일은 하나 이상의 전사정보가 있을 수 있으므로, 하나의 File 엘리먼트는 다수의 Transcription 엘리먼트를 가질 수 있다. [그림 III-3]에서 Set of *Transcription_Id*는 XML 스키마로 구현하면서 File의 하위 엘리먼트인 Transcription을 참조하는 구조로 표현하였다.

전사정보는 녹음된 파일 하나에 대하여 전사의 종류에 따라 여러 개가 존재할 수 있다. 하나의 음성파일에 해당하는 여러 개의 전사는 *Transcription_Id*로 구분한다. *TranscriptionType*은 Orthographic(철자 전사 표기), Morphological(형태소전사 표기), Phonetic(발음전사 표기), Syntactic(구문전사 표기), Semantic(의미전사 표기), Prosodic(운율전사 표기)와 같은 전사 종류를 의미한다. 각 종류별로 해당되는 전사내용을 Transcription 엘리먼트의 내용으로 기술한다. *Transcriber*는 전사자 정보이다.

[표 III-7]은 전사정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

분류	엘리먼트 및 속성	내용
전사정보	<i>Transcription_Id</i>	전사정보 식별자
	<i>TranscriptionType</i>	전사 방법의 종류
	<i>Transcriber</i>	전사정보의 전사기관 또는 전사자의 정보

[표 III-7] 전사정보 엘리먼트 및 속성

vii. 구현 예

[표 III-8]은 제안하는 산업용 음성 DB 메타데이터 규격을 특정 음성 DB에 적용한 기본정보의 예이다.

```
<SpeechDBMetadata>
  <General SpeechSetting="Monologue" SpeechStyle="Elicited" Applications="VoiceControl"
    AnnotationStandard="NIST-LDC" DataCategory="VCVSequence" >
    <Language>KOR</Language>
    <FullNameOfDataCollection>멀티모달 음성인터페이스를 위한 DB</FullNameOfDataCollection>
    <Project>MVD1</Project>
    <VersionInfo>
      <Version>1.0</Version>
      <RecordingPeriod>2003.05.02~2004.11.08</RecordingPeriod>
    </VersionInfo>
    <ReleaseDate>2005-01-13</ReleaseDate>
    <RevisionHistory>
      <RevisionDate>2004-09-13</RevisionDate>
      <RevisionDetail>발화자 mkj00의 avi파일 오류 수정</RevisionDetail>
    </RevisionHistory>
    <Copyright>
      <Authors>박지환</Authors>
      <Distributor>Sitec</Distributor>
      <Performer>MIPS</Performer>
    </Copyright>
    <NumberOfRecordingChannel>3</NumberOfRecordingChannel>
    <Comment/>
    <DataVolume>
      <DataSize>5.0GB</DataSize>
      <DataDuration>14:20:0Z</DataDuration>
    </DataVolume>
    <Quality OverAllQuality="VeryGood">
      <SoundQualityMeasures>
        <SNR>8.9dB</SNR>
        <CrossTalk>4.5dB</CrossTalk>
        <ClippingRate>1.6%</ClippingRate>
        <BackgroundNoise>0.03dB</BackgroundNoise>
        <ErrorRate>0.4%</ErrorRate>
      </SoundQualityMeasures>
    </Quality>
    <Distribution DistributionCategory="Train" DistributionMedia="DVD " Compression="Zip"/>
    ....
  </General>
```

[표 III-8] XML 기반 음성 DB 메타데이터 예제

[표 III-8]의 예를 보면 음성 DB 이름이 ‘멀티모달 음성 인터페이스를 위한 DB’로 음성 DB를 구축한 상위 프로젝트가 ‘MVD1’이며, 발화시 사용 언어는 ‘KOR’(한국어)를 사용하였다는 것을 알 수 있다. 음성 DB의 분류는 ‘VCVSequence’, 발화할 내용의 스타일은 ‘Elicited’, 상용서비스를 고려할 때 본 음성DB의 응용분야는 ‘VoiceControl’, 발화형태는 ‘Monologue’(단독발화)로 녹음하였고, 전사정보 표기 방법은 ‘NIST/LDC’을 사용하였음을 의미한다. 버전은 ‘1.0’, 구축기간은 ‘2003년 5월 2일부터 2004년 11월 8일’까지이며, 배포날짜는 ‘2005년 1월 13일’인 것을 의미한다. 저작권 정보로서 저작자 ‘박지환’, 배포기관은 ‘SiTEC’, 수행기관은 ‘성신여대 MIPS’란 것을 알 수 있다. 해당 음성 DB의 전체 물리적 크기는 ‘5.0GB’, 음성 DB 런타임시간은 ‘14:20:0Z’임을 알 수 있다. 음성 DB 품질은 ‘VeryGood’으로 신호대잡음비는 ‘8.9dB’, 누화현상은 ‘4.5dB’, 클리핑 비율은 ‘1.6%’, 잡음은 ‘0.03dB’, DB오류율은 ‘0.4%’임을 의미한다. 배포분류는 ‘Train’(학습용)이고, 배포 시 ‘DVD’로 배포되며, 압축형식은 ‘Zip’이라는 것을 알 수 있다.

[표 III-9]은 채널정보를 메타데이터로 구현한 예다. 예를 보면, 채널 1번은 수집환경의 통신망 종류가 ‘Microphone’이고 음성신호의 저장 단위 방식은 ‘FloatingPoint’를 사용한 것을 의미한다. 녹음 시 발화환경은 ‘방음실’이고 제조사가 ‘젠하이저’, 모델명이 ‘M30’이고 ‘dynamic’ 규격인 마이크를 녹음 시 사용하였다는 것을 의미한다. 또한 마이크와 발화자의 발성거리는 ‘35cm’이고, 구축된 음성 DB에서 사용한 A/D 장비가 ‘SoundBlaster’에서 만들어진 ‘SB129’ 모델이며, 규격은 ‘digital’인 것을 의미한다. 디지털 음성데이터의 인코딩 방법은 ‘PCM’, 바이트 순서는 ‘LittleEndian’, 음성샘플 당 차지하는 비트수는 ‘8Bit’,

샘플링 주파수는 ‘150KHz’임을 의미한다.

```

<Channels>
  <Channel TypeOfResource="Microphone" Channel_Id="ID01" DataFormat="FloatingPoint">
    <Recording>
      <RecordingEnvironment>방음실</RecordingEnvironment>
      <RecordingDevice>
        <Manufacturer>젠하이저</Manufacturer>
        <Model>M30</Model>
        <Spec>dynamic</Spec>
      </RecordingDevice>
      <TalkingDistance>35cm</TalkingDistance>
    </Recording>
    <Digitizing>
      <DigitizingDevice>
        <Manufacturer>SoundBlaster</Manufacturer>
        <Model>SB129</Model>
        <Spec>digital</Spec>
      </DigitizingDevice>
      <SignalEncoding type="PCM"/>
      <ByteOrder order="LittleEndian"/>
      <BitsPerSample>8Bit</BitsPerSample>
      <SamplingRate>150KHz</SamplingRate>
    </Digitizing>
  </Channel>
  <Channel/>
  ...
</Channels>

```

[표 III-9] XML 기반 음성 DB 메타데이터 예제

[표 III-10]은 화자정보를 메타데이터로 구현한 예다. 화자 식별자 ‘csy00’은 ‘Native’(내국인), ‘비흡연자’이고 훈련된 발화자가 아니며 나이는 ‘20세’, ‘Female’(여자), ‘언어장애가 없고’, 발화한 목록은 ‘ID03’과 ‘ID08’임을 알 수 있다. 화자 이름은 ‘천신영’, 출생지는 ‘서울’, 현재 거주지역은 ‘서울’, 서울에서 거주한 기간은 ‘20년’임을 의미한다.

```

<Speakers>
  <Speaker Speaker_Id="csy00" Origin="Native" SmokingHabit="no" TrainedSpeaker="no" Age="20"
    Utterancelists="ID03 ID08" SpeakingHearingImpairment="no" Gender="Female">
    <SpeakerName>천신영</SpeakerName>
    <BirthPlace>서울</BirthPlace>
    <CurrentlyResidence>서울</CurrentlyResidence>
    <CurrentlyResidenceDuration>20</CurrentlyResidenceDuration>
    ...
  </Speaker>
</Speaker>
...
</Speakers>

```

[표 III-10] XML 기반 음성 DB 메타데이터 예제

[표 III-11]은 발화목록을 메타데이터로 구현한 예다. 발화목록 식별자 ‘ID03’은 ‘1번부터 105번까지 발화문장’을 가지고 있다는 것을 알 수 있다. 발화문장 ‘ID000001’번은 ‘영’, 발화문장 ‘ID000002’번은 ‘공’이란 단어를 의미한다.

```

<UtteranceLists>
  <UtteranceList Utterancelist_Id="ID03">
    <UtteranceListDesc>1번부터 105번까지 발화목록</UtteranceListDesc>
    <Utterance Utterance_Id="ID000001">영</Utterance>
    <Utterance Utterance_Id="ID000002">공</Utterance>
    ...
  </UtteranceList>
</UtteranceList>
...
</UtteranceLists>

```

[표 III-11] XML 기반 음성 DB 메타데이터 예제

[표 III-12]은 파일정보와 전사정보를 메타데이터로 구현한 예다. 예를 보면, 파일 ‘ID001’은 발화목록 ‘ID03’에 있는 발화문장 ‘ID000001’을 채널 ‘ID01’을 통해 화자 ‘cjs00’이 발화하여 녹음한 음성파일인 것을 알 수 있다. 파일 형식은 ‘Wav’이고 파일헤더는 ‘0Byte’, 파일크기는 ‘1024Byte’란 것을 의미한다. 파일이름은 ‘dig01.wav’이며 파일위치는 ‘D:\data \digit\female\f2cjs0’이다. 파일 안에 포함된 잡음은 ‘한 개’ 존재하고 잡음 종류는 음악, 잡음 시작시간은 ‘00:01:34’, 끝나는 시간은 ‘00:01:50’임을 알 수 있다. 전사종류는 ‘Syntactic’, 전사 식별자는 ‘ID0001’, 전사기관은 ‘SiTEC’, 전사내용은 ‘영’을 의미한다.

```

<Files>
  <File File_Id="ID001" Utterancelist_Id="ID03" Channel_Id="ID01" Utterance_Id="ID000001"
    Speaker_Id="cjs00">
    <FileFormat format="Wav">
      <HeaderSize>0Byte</HeaderSize>
      <FileSize>1024Byte</FileSize>
    </FileFormat>
    <FileLocation>
      <SpeechFileName>dig01.wav</SpeechFileName>
      <DirectoryPath>D:\data\digit\female\f2cjs0</DirectoryPath>
    </FileLocation>
    <Noise NumberOfNoise="1">
      <NoiseSection>
        <NoiseType>음악</NoiseType>
        <NoiseStartTime>00:01:34</NoiseStartTime>
        <NoiseEndTime>00:01:50</NoiseEndTime>
      </NoiseSection>
    </Noise>
    <Transcription TranscriptionType="Syntactic" Transcription_Id="ID0001" Transcriber="SiTEC">
      영</Transcription>
    </File>
    <File/> ...
  </Files>
</SpeechDBMetadata>

```

[표 III-12] XML 기반 음성 DB 메타데이터 예제

IV. 산업용 음성 DB 메타데이터 관리도구 설계 및 구현

1. 검색기능 설계 및 구현

1.1. 검색기능 설계

산업용 음성 DB 메타데이터 관리도구는 메타데이터 검색기능, 생성 및 수정기능을 제공하도록 설계하였다. 검색 기능의 요구사항은 다음과 같다.

첫째, 검색 시 하나의 음성 DB 메타데이터를 대상으로 한다.

검색기능은 III장에서 정의한 음성 DB 메타데이터 스키마에 적합한 문서가 완성되어 존재한다고 가정한 후 검색 조건에 따라 필요한 정보를 추출하여 검색결과를 제공한다.

둘째, 검색 조건은 대상 음성 DB 메타데이터의 내용에서 추출하여 대상 음성 DB 메타데이터가 바뀔 때마다 동적인 화면으로 제공해야 한다. 음성 DB는 서울, 경기 지역의 출신자를 대상으로만 하거나 60세 이상의 노인만을 대상으로 하거나 혹은 30대의 남자, 여자만을 대상으로 하는 것과 같이 여러 가지 조건에서 구축될 수 있다. 이런 음성 DB의 특성상 모든 음성 DB 메타데이터의 검색 조건을 한 가지 패턴으로 고정하여 제공하는 것보다 해당 음성 DB 메타데이터의 정보를 추출하여 가능한 검색 조건을 제공하는 것이 사용자에게 매력적이며 효과적이다.

검색 조건은 크게 화자, 채널, 발화목록, 3가지 영역으로 구분한다. 화자는 지역, 연령, 성별로 구분하고 채널은 녹음 시 사용한 통신망환경, 인코딩 방법으로 구분한다. 화자, 채널, 발화목록 각각의 검색 조건은 모두선택, 대상 음성 DB 메타데이터에서 읽어낸 값들을 조건으로 선택, 직접 검색으로

나누어 검색한다.

화자는 발화자의 출생지역을 의미하는 지역, 연령, 성별로 각각 검색할 수 있도록 세분화한다. 예를 들면, 서울 출신 화자 중 20대이고 여자인 화자가 발화한 음성파일, 화자정보, 전사정보를 검색할 수 있어야 한다.

채널은 음성 DB 수집 시 사용한 통신망환경과 인코딩으로 세분화 하여 검색할 수 있도록 한다. 음성 DB는 서로 다른 방법의 채널 통신망환경을 사용하여 수집 가능하므로 사용자가 PDA 또는 마이크를 사용한 채널에서 Linear 인코딩 방법을 이용하여 녹음한 음성파일 및 그 음성파일을 발화한 화자를 검색할 수 있어야 한다.

발화목록은 전체 발화문장을 선택하거나 발화목록 별로 발화문장을 개별적으로 선택할 수 있도록 한다.

사용자가 검색 조건 중 아무것도 선택을 안 하는 경우에는 검색 조건 모두를 선택한 것으로 검색하도록 한다. 검색 조건인 지역, 연령, 성별, 통신망환경, 인코딩, 발화문장 조건은 서로 교집합으로 연산되어 결과 조건에 맞는 정보를 검색한다.

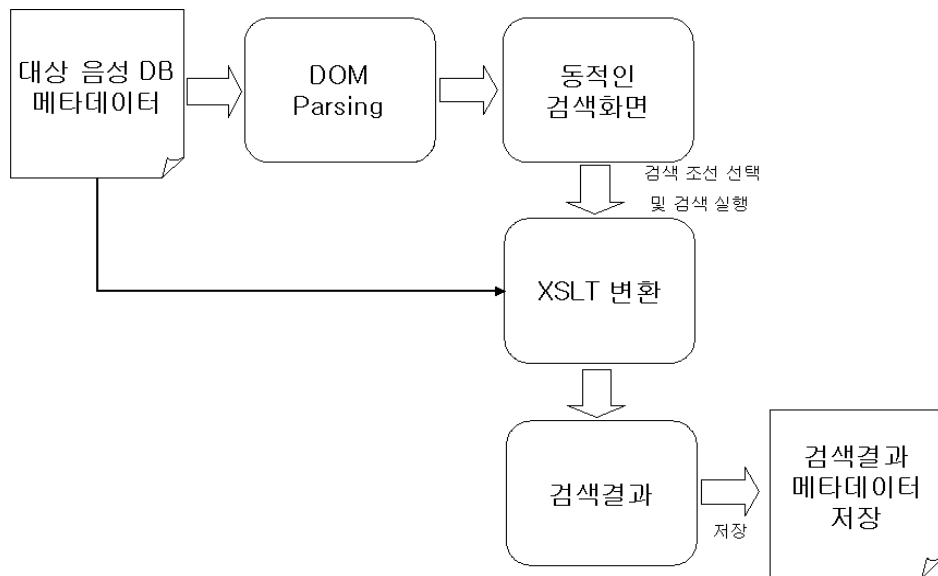
셋째, 결과 조건은 화자정보, 음성파일, 전사정보로 구분한다. 결과 조건은 화자정보, 음성파일, 전사정보로 검색 조건에 맞는 결과를 합집합 연산하여 결과로 출력해야 한다.

화자정보를 선택하면 검색 조건에 맞는 화자정보를 출력하고, 화자정보와 음성파일을 선택하면 검색 조건에 맞는 음성파일과 화자정보를 검색 결과로 출력한다. 전사정보는 결과 조건으로 음성파일을 선택하고 전사정보를 선택했을 때 해당 메타데이터에 전사정보가 존재한다면 결과에 출력한다.

넷째, 출력 결과는 III장에서 제안한 메타데이터 스키마 문법에 맞는 형태로 저장될 수 있어야 하며, 음성파일을 재생하거나 메타데이터와 음성파일을 사용자가 지정하는 곳으로 저장할 수 있어야 한다. 이것은 기존의 메타데이터로부터 새로운 메타데이터를 생성하여 기존에 구축된 여러 음성 DB의 일부 혹은 전체의 재사용을 가능하게 한다.

1.2. 검색기능 구현

검색기능 흐름은 [그림 IV-1]과 같다. 사용자가 선택한 메타데이터에 포함되어 있는 정보를 DOM[9]으로 파싱한다. DOM으로 먼저 메타데이터를 파싱하는 것은 대상 음성 DB 메타데이터의 내용에 따라 검색화면이 달라지는 동적인 검색화면을 제공하기 위해서이다.



[그림 IV-1] 검색기능 흐름

해당 음성 DB 메타데이터 정보에 따라 동적인 검색 화면이 제공되고, 사용자로부터 입력 받는 검색 조건은 검색기능을 위해 마련한 여러 개의 XSLT[10]~[12] 문서에 반영된다. 검색 조건이 반영된 XSLT를 이용하여 원본 메타데이터를 변환하여 검색 결과를 구한다. 검색 결과는 사용자가 선택한 조건에 따라 추출된 정보(음성파일)를 재생할 수 있고, 기존 대상 음성 DB 메타데이터로부터 검색 결과 정보만을 저장한 새로운 내용의 메타데이터를 저장할 수 있다.

검색은 다음의 세 가지가 가능하도록 구현하였다.

① 기본정보보기 : 내부적으로 저장되어 있는 XSLT 파일을 통하여 변환된 HTML[13] 파일 형태로 기본정보를 제공한다.

② 검색 : 사용자가 선택한 조건이 내부적으로 저장되어 있는 여러 개의 XSLT 파일에 여러 개의 파라미터로 전달되고, XSLT 파일에 의해서 대상 음성 DB 메타데이터가 변환되어 최종 검색 결과를 제공한다.

③ 메타데이터 저장 : 사용자가 검색한 결과를 지정하는 곳에 새로운 메타데이터로 저장한다.

1.2.1. 구현 환경

음성 DB 메타데이터 관리도구 구현 환경은 다음과 같다.

하드웨어	
CPU	Intel Pentium4 /3.0 GHz
RAM	1.0 GB
HDD	80 GB
소프트웨어	
운영체제	Window XP
Programming Language	JDK 1.5
XML Parser	Xalan 2.7
Programming Tool	Eclipse 3.0

[표 IV-1] 구현 환경

1.2.2. 운영 방식

음성 DB 메타데이터 관리도구는 JDK 1.5를 토대로 설계 되었으므로 JDK가 설치되어 있는 곳이라면 어느 곳에서라도 실행 시킬 수 있다. 따라서 본 도구를 실행 하기 위해서는 JDK 1.5를 인스톨 한 뒤 실행한다.

본 논문에서 설계한 음성 DB 메타데이터 관리도구는 수정, 추가, 삭제 작업이 용이하도록 DOM 파서를 사용하고 XSLT를 이용하여 메타데이터를 변환, 생성한다.

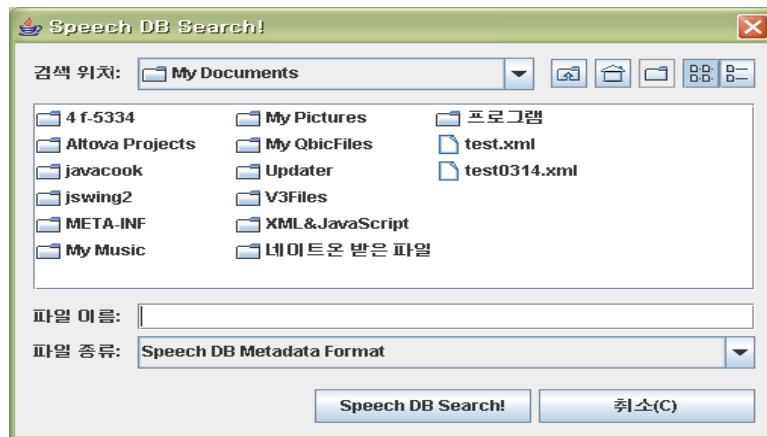
1.3. 검색기능 구현 결과

사용자는 Speech DB Search 대화창을 통해서 검색할 대상 음성 DB 메타데이터를 선택한다. Speech DB Search 대화창은 [그림 IV-2]에서 보는 바와 같이 대상 메타데이터를 선택하는 버튼으로 구성된다.



[그림 IV-2] Speech DB Search 대화창

사용자가 [그림 IV-2]의 버튼을 선택하면 [그림 IV-3]과 같이 검색 대상이 될 음성 DB 메타데이터를 선택하는 File Open 대화창이 나타난다. 파일 필터링 작업을 통해 XML 파일을 우선적으로 보여준다.



[그림 IV-3] File Open 대화창

[그림 IV-3]에서 이미 작성되어 있는 음성 DB 메타데이터를 선택하면 [그림 IV-4]과 같이 산업용 음성 DB 검색 대화창이 나타난다.

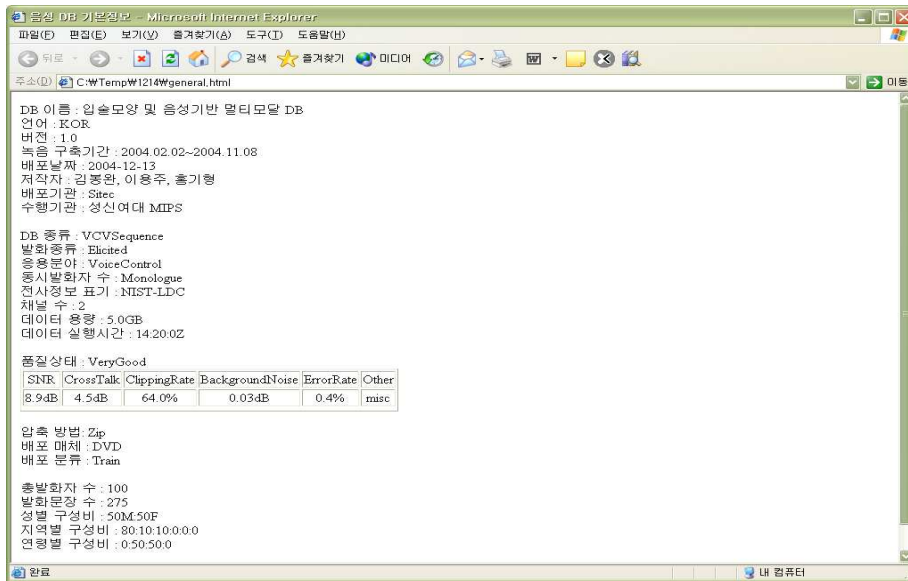
① 음성 DB 기본정보보기 버튼

음성 DB 기본정보보기 버튼을 선택하면 [그림 IV-5]와 같이 음성 DB 메타데이터의 기본 정보만을 추출한 문서를 결과로 생성하여 보여준다. [그림 IV-4]의 음성 DB 기본정보보기 버튼 왼쪽에 있는 텍스트들은 대상 음성 DB 메타데이터 기본정보의 요약 정보를 보이고 있다.



[그림 IV-4] 산업용 음성 DB 검색 대화창

[그림 IV-4]의 검색 조건은 대상 음성 DB 메타데이터에 존재하는 정보로 구성한다. 예를 들어 화자의 출생지에 제주도가 존재하면 제주도를 화면에서 선택할 수 있도록 추가되고, 서울이 존재하지 않으면 서울 지역은 화면에 보이지 않는다. 즉, 검색 도구 메인 화면은 대상 음성 DB 메타데이터 정보에 의하여 동적으로 변화하게 된다.



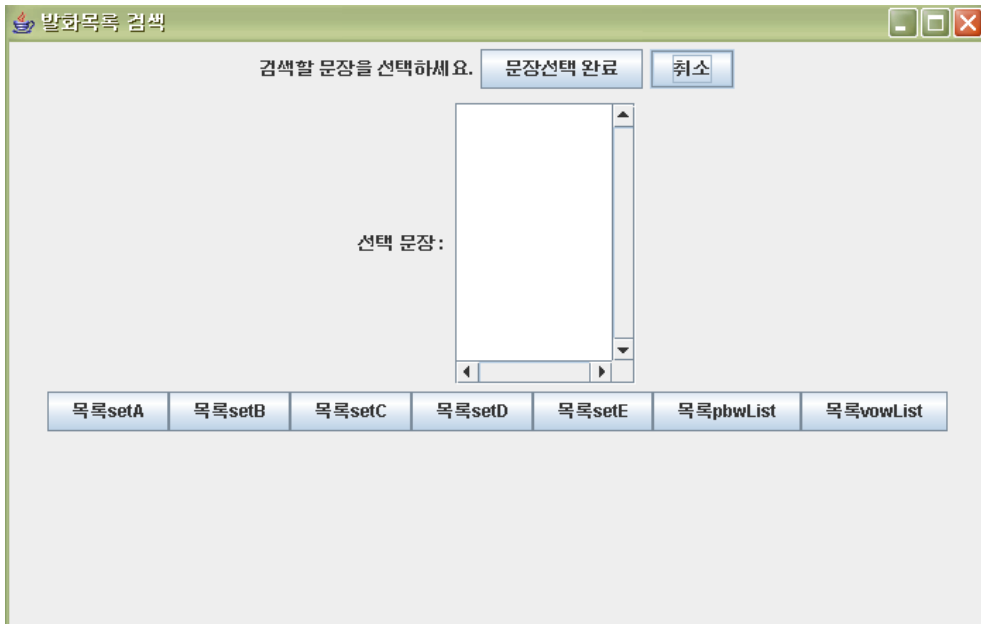
[그림 IV-5] 기본 정보 문서

② 발화목록검색 버튼

발화목록검색 버튼을 선택하면 [그림 IV-6]와 같이 발화목록 별로 검색할 발화문장을 선택할 수 있도록 발화목록 검색 대화창이 나타난다.

B. 목록버튼

[그림 IV-6]에서 목록setA, 목록setB와 같은 버튼은 대상 음성 DB 메타데이터에 존재하는 발화목록 개수만큼 생성된다. 목록버튼 이름은 대상 음성 DB 메타데이터에 있는 발화목록 식별자에 의해 결정된다. 사용자가 목록 setE 버튼을 선택하면 해당되는 발화문장이 리스트로 [그림 IV-7]과 같이 나타난다. 사용자가 검색을 원하는 발화문장을 선택하면 선택문장은 TextArea에 보여진다.



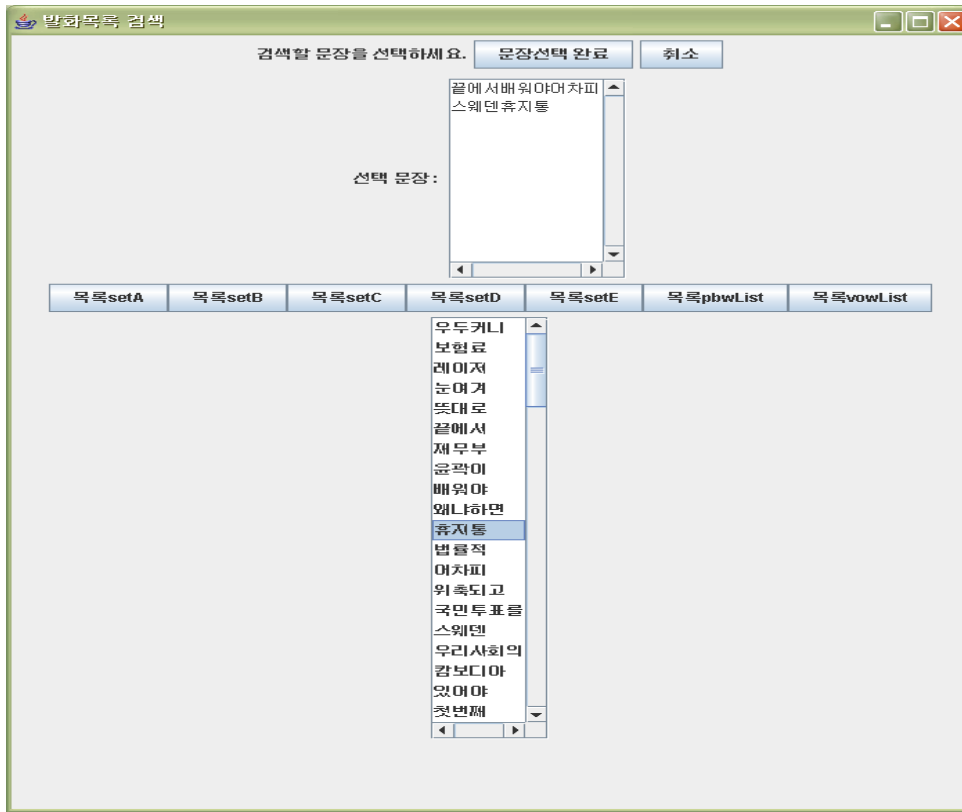
[그림 IV-6] 발화목록 검색 대화창

C. 문장선택 완료 버튼

[그림 IV-7]에서 사용자가 검색할 발화문장을 모두 선택한 후에 문장선택 완료 버튼을 선택하면 [그림 IV-4]으로 돌아가서 발화문장 TextField에 선택한 발화문장 항목을 보여준다.

D. 취소 버튼

[그림 IV-7]에서 사용자가 취소 버튼을 선택하면 선택했던 문장들이 모두 초기화된다. 취소 버튼을 선택하면 [그림 IV-6] 상태로 돌아가게 된다.



[그림 IV-7] 발화문장 선택 대화창

③ 검색 버튼

사용자가 [그림 IV-8]과 같이 검색 조건을 선택하고, 결과 조건은 화자정보를 선택한 후에 검색 버튼을 누르면 결과화면은 [그림 IV-10]과 같다. 검색 버튼은 모든 조건을 선택한 후에 최종적으로 검색 결과를 보기 원할 때 선택한다. 검색버튼을 선택하면 [그림 IV-9] 음성 DB 검색 대화창에서 보듯이 메타데이터 저장과 새로운 음성 DB 메타데이터 생성이라는 버튼이 생성된다.



[그림 IV-8] 음성 DB 검색 대화창

사용자가 [그림 IV-8]과 같이 검색 조건을 선택하고 결과 조건으로 화자정보와 음성파일을 선택하고 검색 버튼을 누르면 결과화면은 [그림 IV-11]와 같다. 음성파일을 선택하고 검색하게 되면 해당 음성파일을 재생할 수 있도록 구현하였다. 결과 조건으로 화자정보, 음성파일, 전사정보를 선택하고 검색 조건은 [그림 IV-8]과 같다면 결과화면은 [그림 IV-12]과 같다. 전사정보는 음성파일 선택 시 대상 메타데이터에 전사정보가 존재할 경우에만 출력된다.

④ Clear 버튼

[그림 IV-8]에서 Clear 버튼을 선택하면 선택했던 모든 조건들은 초기화되고 [그림 IV-4]으로 돌아가게 된다. 사용자는 원하는 조건으로 대상 음성 DB 메타데이터를 다시 검색할 수 있다.



[그림 IV-9] 음성 DB 검색 대화창

⑤ 메타데이터 저장 버튼

[그림 IV-9]의 메타데이터 저장 버튼은 검색 화면에서 검색 버튼을 누를 때 활성화된다. 사용자가 선택한 검색 조건으로 검색한 결과를 사용자가 지정한 디렉토리에 새로운 메타데이터로 저장할 수 있다.

⑥ 새 음성 DB 생성 버튼

[그림 IV-9]의 새 음성 DB 생성 버튼은 메타데이터 저장 버튼과 함께 검색 버튼을 누를 때 활성화된다. 사용자가 원하는 검색 조건으로 검색한 결과를 사용자가 지정한 디렉토리에 새 메타데이터와 함께 음성 파일을 저장하여 음성 DB를 새로 구성할 수 있다.

번호	출생지	발화자 식별자	발화 목적	성별	나이	현재거주지역	아버지 고향	어머니 고향	12세 이전 성장지역	발언
1	경기	cjs00	setE	Female	20	경기	서울	경북	경기	없음
2	서울	csj00	setE	Female	20	서울	전남	서울	서울	없음
3	경남	cbs00	setE	Male	25	경기	부산	서울	서울	없음

[그림 IV-10] 검색 결과 화면

번호	출생지	발화자 식별자	발화 목적	성별	나이	현재거주지역	아버지 고향	어머니 고향	12세 이전 성장지역	발언
1	경기	cjs00	setE	Female	20	경기	서울	경북	경기	없음
2	서울	csj00	setE	Female	20	서울	전남	서울	서울	없음
3	경남	cbs00	setE	Male	25	경기	부산	서울	서울	없음

번호	파일식별자	발화자식별자	발화목적	발화문장	채널번호	발화시작시간	발화마침시간	파일이름	파일위치	파일크기
1	ID0006	cjs00	setE	ID367	ID01	00:00:07	00:0:21	pbw367.wav	WdataWpbwWFemaleWf2cjs0	35KB
2	ID000188	csj00	setE	ID367	ID01	00:00:07	00:0:21	pbw367.wav	WdataWpbwWFemaleWf2cjs0	39KB
3	ID00098	cbs00	setE	ID367	ID01	00:00:07	00:0:21	pbw367.wav	WdataWpbwWmaleWm2cbs0	38KB
4	ID0009	cjs00	setE	ID370	ID01	00:00:07	00:0:21	pbw370.wav	WdataWpbwWFemaleWf2cjs0	33KB
5	ID000191	csj00	setE	ID370	ID01	00:00:07	00:0:21	pbw370.wav	WdataWpbwWFemaleWf2cjs0	40KB
6	ID000101	cbs00	setE	ID370	ID01	00:00:07	00:0:21	pbw370.wav	WdataWpbwWmaleWm2cbs0	35KB
7	ID00011	cjs00	setE	ID372	ID01	00:00:07	00:0:21	pbw372.wav	WdataWpbwWFemaleWf2cjs0	34KB
8	ID000193	csj00	setE	ID372	ID01	00:00:07	00:0:21	pbw372.wav	WdataWpbwWFemaleWf2cjs0	40KB
9	ID000103	cbs00	setE	ID372	ID01	00:00:07	00:0:21	pbw372.wav	WdataWpbwWmaleWm2cbs0	39KB
10	ID00013	cjs00	setE	ID374	ID01	00:00:07	00:0:21	pbw374.wav	WdataWpbwWFemaleWf2cjs0	35KB
11	ID000195	csj00	setE	ID374	ID01	00:00:07	00:0:21	pbw374.wav	WdataWpbwWFemaleWf2cjs0	39KB
12	ID000105	cbs00	setE	ID374	ID01	00:00:07	00:0:21	pbw374.wav	WdataWpbwWmaleWm2cbs0	39KB
13	ID00016	cjs00	setE	ID377	ID01	00:00:07	00:0:21	pbw377.wav	WdataWpbwWFemaleWf2cjs0	33KB
14	ID000198	csj00	setE	ID377	ID01	00:00:07	00:0:21	pbw377.wav	WdataWpbwWFemaleWf2cjs0	40KB
15	ID000108	cbs00	setE	ID377	ID01	00:00:07	00:0:21	pbw377.wav	WdataWpbwWmaleWm2cbs0	37KB

[그림 IV-11] 검색 결과 화면

번호	출생지	발화자 식별자	발화 목록	성별	나이	현재거주지역	아버지 고향	어머니 고향	12세이전 성장지역	방언
1	경기	cjs00	setE	Female	20	경기	서울	경북	경기	없음
2	서울	csj00	setE	Female	20	서울	전남	서울	서울	없음
3	경남	cbs00	setE	Male	25	경기	부산	서울	서울	없음

번호	파일식별자	발화자식별자	발화목록	발화문장	채널번호	발화시작시간	발화마침시간	파일이름	파일위치	파일크기	전사
1	ID0005	cjs00	setE	ID366	ID01	00:00:07	00:0:21	pbw366.wav	WdataWpbwWFemaleWF2cjs0	33KB	뜻대로
2	ID000187	csj00	setE	ID366	ID01	00:00:07	00:0:21	pbw366.wav	WdataWpbwWFemaleWF2csj0	35KB	뜻대로
3	ID00097	cbs00	setE	ID366	ID01	00:00:07	00:0:21	pbw366.wav	WdataWpbwWmaleWm2cbs0	35KB	뜻대로
4	ID00010	cjs00	setE	ID371	ID01	00:00:07	00:0:21	pbw371.wav	WdataWpbwWFemaleWF2cjs0	38KB	왜냐하면
5	ID000192	csj00	setE	ID371	ID01	00:00:07	00:0:21	pbw371.wav	WdataWpbwWFemaleWF2csj0	43KB	왜냐하면
6	ID000102	cbs00	setE	ID371	ID01	00:00:07	00:0:21	pbw371.wav	WdataWpbwWmaleWm2cbs0	40KB	왜냐하면
7	ID00014	cjs00	setE	ID375	ID01	00:00:07	00:0:21	pbw375.wav	WdataWpbwWFemaleWF2cjs0	40KB	위축되고
8	ID000196	csj00	setE	ID375	ID01	00:00:07	00:0:21	pbw375.wav	WdataWpbwWFemaleWF2csj0	44KB	위축되고
9	ID000106	cbs00	setE	ID375	ID01	00:00:07	00:0:21	pbw375.wav	WdataWpbwWmaleWm2cbs0	42KB	위축되고

[그림 IV-12] 검색 결과 화면

2. 구축, 수정기능 설계 및 구현

2.1. 구축 및 수정기능 설계

산업용 음성 DB 메타데이터 관리도구의 구축 및 수정기능의 요구사항은 다음과 같다.

첫째, 산업용 음성 DB 구축 시에 대용량으로 데이터를 저장, 관리하는 경우가 많기 때문에 산업용 음성 DB를 구축하는 사용자가 음성 DB 기본 정보, 수집채널 정보, 화자 정보, 발화목록 정보, 파일 및 전사정보를 모두 한 번에 입력하기는 힘든 일이다. 따라서 5가지

정보 별로 입력하여 임시 저장할 수 있어야 하며, 추후에 최종적으로 음성 DB 메타데이터를 생성, 저장할 수 있어야 한다.

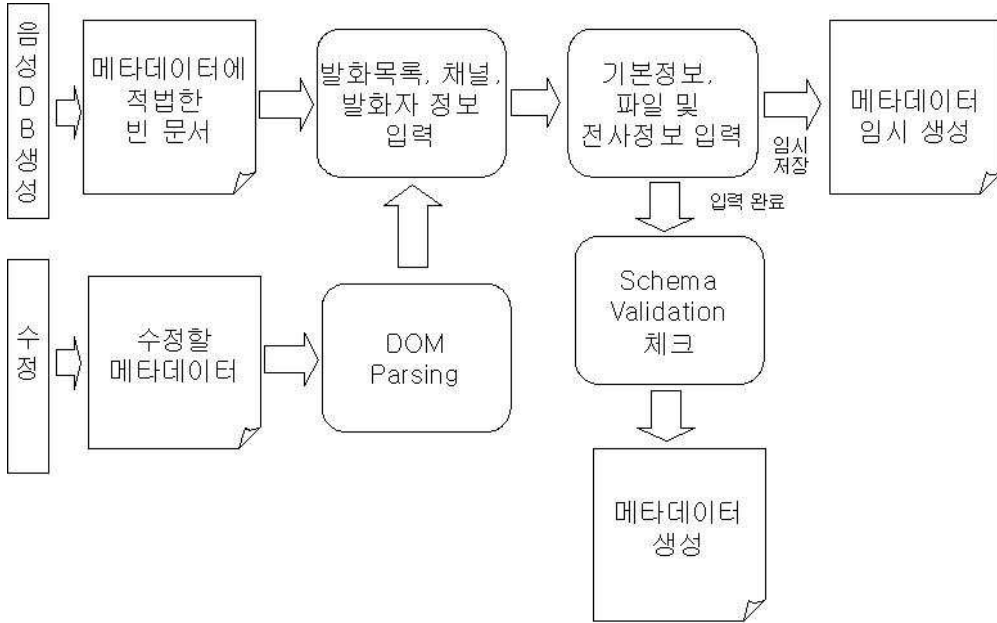
둘째, 수집채널, 화자정보, 발화목록 정보 입력을 먼저 수행 한 후에 파일정보 입력이 가능하다. III장에서 설명하였듯이 본 음성 DB 메타데이터 표준안은 객체 참조 개념을 도입하여 관계 있는 객체간에 상호참조 하도록 정의하였다. 음성파일 정보 한 개를 저장할 때 참조하는 속성은 발화자 식별자, 수집채널 식별자, 발화목록 식별자, 발화문장 식별자이다. 이와 같은 식별자는 속성값이 존재 해야지만 참조 가능하므로 수집채널, 화자정보, 발화목록을 먼저 입력 한 후 파일정보를 입력해야 한다. 따라서, 각 정보에서 서로 참조하는 속성값들은 사용자로부터 먼저 입력된 속성값을 읽어서 자동적으로 볼 수 있도록 하여 사용자에게 편리성을 제공해야 한다.

셋째, 수집채널, 화자, 음성파일, 발화목록 정보 등은 데이터가 여러 개인 경우가 대부분이므로 같은 형식의 다른 데이터를 여러 번 저장할 수 있는 방법을 제공해야 한다.

넷째, 음성 DB 구축이 완료되는 시점에서는 메타데이터에 적합한지 스키마 유효성 검사가 필요하다. III장에서 제안한 메타데이터 스키마 문법에 적합해야만 새로운 음성 DB 메타데이터를 생성한다.

2.2. 구축 및 수정기능 구현

산업용 음성 DB 메타데이터 구축 및 수정기능 흐름은[그림 IV-13]과 같다.



[그림 IV-13] 구축 및 수정기능 흐름도

첫째, 새로운 음성 DB 메타데이터를 생성하려면 기본적으로 현재 디렉토리에 내용은 비어있지만 스키마에 적법한 메타데이터 문서를 한 개 생성한다.

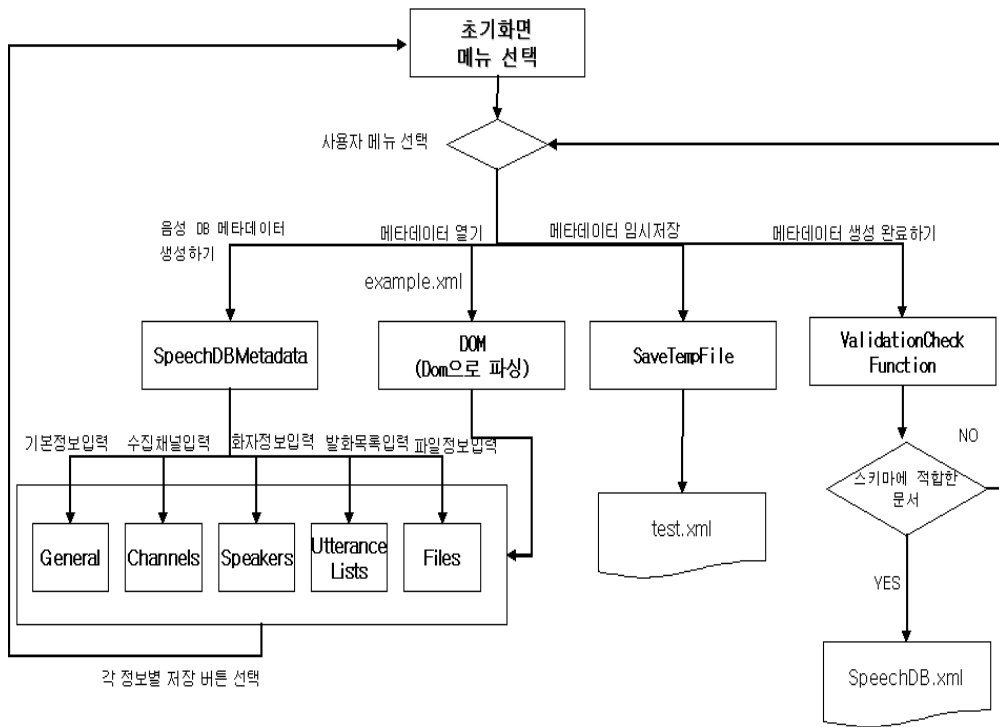
둘째, 이미 구축되었거나 전에 작업한 음성 DB 메타데이터 파일이 존재하는 경우 사용자가 파일을 선택하면 해당 메타데이터 문서를 DOM으로 파싱한다. DOM을 이용하여 파싱하는 것은 사용자가 데이터를 수정 및 삭제할 경우 메타데이터에 존재하는 정보가

있다면 화면에 제공하고 입력을 받을 수 있도록 하기 위해서이다.

셋째, 사용자로부터 입력 받은 모든 데이터는 임시저장 할 수 있다. 사용자로부터 입력 받은 정보는 음성 DB 메타데이터가 완전히 구축되기 전에는 임시로 저장하여, 구축의 편의를 도모한다.

사용자가 구축한 음성 DB에 관련 되는 모든 정보를 입력한 후에 최종적으로 스키마에 적합한 메타데이터인지 아닌지를 확인한 후 적법하다면 메타데이터를 최종적으로 생성한다.

구축 및 수정기능 시스템 흐름도는 [그림 IV-14]과 같다.



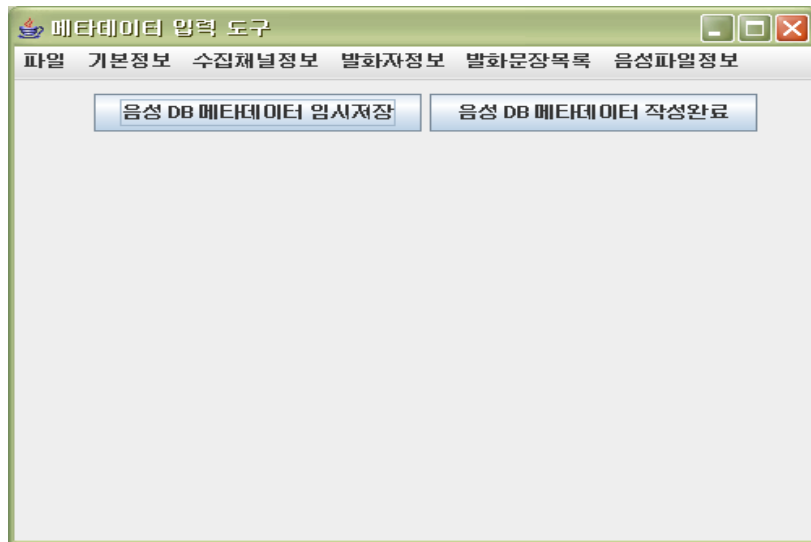
[그림 IV-14] 관리도구 구축 및 수정 전체 시스템 흐름도

2.2.1. 구현 환경

음성 DB 메타데이터 구축 및 수정기능 구현 환경 및 운영방식은 검색기능 구현([표 IV-1])과 동일하다.

2.3. 구축 및 수정기능 구현 결과

관리도구의 구축 및 수정 메인 화면은 [그림 IV-15]와 같다. 사용자가 파일 메뉴 중 음성 DB 만들거나 열기를 선택하면 음성 DB 메타데이터를 생성 및 수정할 수 있다. 음성 DB 만들기는 새로운 음성 DB 메타데이터를 생성하는 것이고 열기는 수정할 메타데이터를 선택한 후 정보 별로 사용자의 입력에 따라 메타데이터 내용을 수정한다.

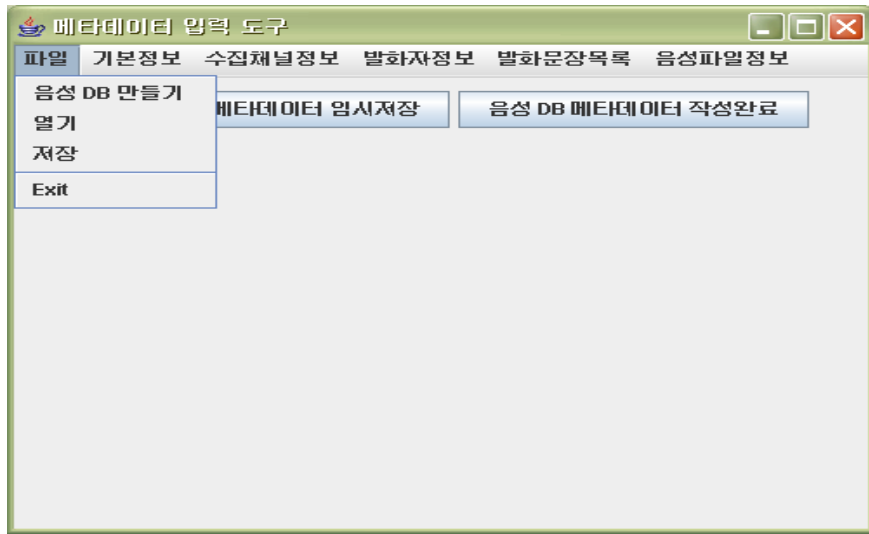


[그림 IV-15] 구축 및 수정 메인 화면

[그림 IV-15]에서 보듯이 산업용 음성 DB 메타데이터 관리도구의 구축 및 수정기능은 메뉴기반으로 구현하였다. 관리도구는 크게 다음의 여섯 가지 메뉴를 제공한다.

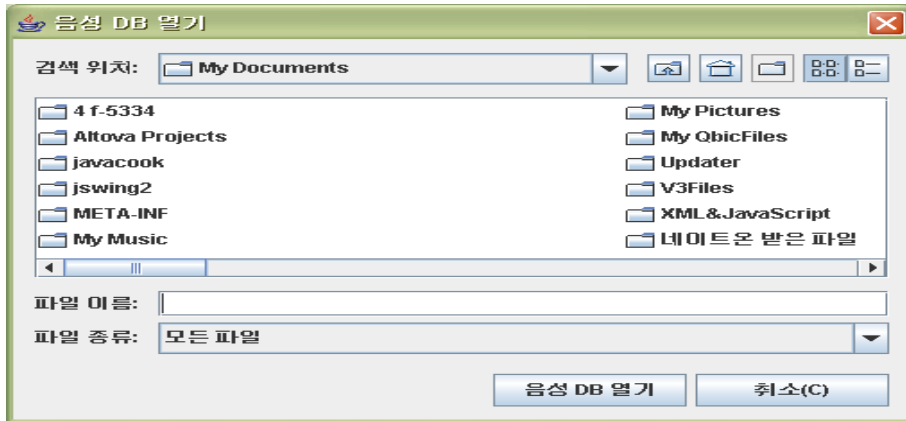
- ① 파일 : 음성 DB 만들기, 열기, 저장, 종료 기능을 제공한다. 음성 DB 만들기는 현재 디렉토리에 기본 메타데이터를 생성하고, 열기는 미리 만들어진 메타데이터를 수정하는 기능을 제공한다.
- ② 기본정보 입력 : 음성 DB 메타데이터의 기본정보 입력 및 수정 기능을 제공한다.
- ③ 수집채널 입력 : 여러 개의 수집채널 정보 입력, 수정, 삭제 기능을 제공한다.
- ④ 화자정보 입력 : 여러 명의 화자정보 입력, 수정, 삭제 기능을 제공한다.
- ⑤ 발화목록 입력 : 다수의 발화목록에 다수의 발화문장을 입력, 수정, 삭제하는 기능을 제공한다.
- ⑥ 파일정보 입력 : 녹음 작업을 통해 얻은 물리적 음성파일 정보를 저장하는 파일정보를 입력, 수정, 삭제하는 기능을 제공한다. 전사정보 존재 시 입력하는 기능을 함께 제공한다.

사용자가 파일 메뉴를 선택한 후 [그림 IV-16]와 같이 음성 DB 만들기를 선택하면 메타데이터 스카마에 적법하고, 내용은 비어있는 빈 문서가 한 개 생성된다.



[그림 IV-16] 파일 메뉴 대화창

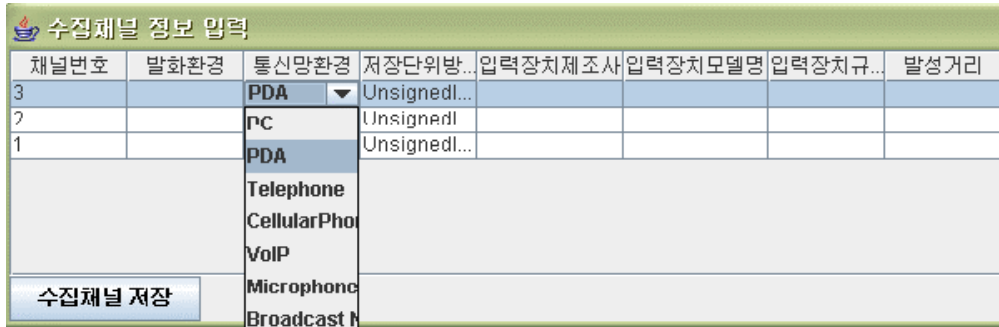
생성된 문서에 사용자는 발화자, 채널, 발화목록 정보부터 입력할 수 있다. [그림 IV-16]에서 사용자가 열기를 선택하면 [그림 IV-17]과 같이 수정할 음성 DB 메타데이터를 선택하는 대화창이 나타난다.



[그림 IV-17] 음성 DB 열기 대화창

음성 DB 열기 대화창에서 선택한 메타데이터를 DOM으로 파싱한 후 사용자가 발화자, 수집채널, 발화목록, 음성파일 정보 입력 선택 시 기존에 존재하는 정보를 보여주고, 그것에 사용자가 정보를 추가 입력 또는 수정할 수 있다.

사용자가 채널 정보를 먼저 입력하기 위해 수집채널 메뉴를 선택하면 [그림 IV-18]과 같이 수집채널 정보 입력 대화창이 나타난다. 수집채널이 여러 개일 경우가 있으므로 [그림 IV-19]와 같이 행 삽입과 행 삭제 버튼을 통해 융통성 있게 사용자가 입력하도록 구현하였다. 수집채널 뿐만 아니라 발화자, 발화목록, 음성파일, 전자정보 모두 동일하게 행 삽입과 행 삭제 버튼이 제공된다.



[그림 IV-18] 수집채널 정보 입력 대화창



[그림 IV-19] 수집채널 정보 입력 대화창

수집채널 정보를 모두 입력한 후 수집채널 저장 버튼을 선택하면 사용자로부터 입력 받은 정보를 저장한다.

[그림 IV-15]에서 발화자 정보 메뉴를 선택하면 [그림 IV-20]과 같이 발화자 정보 입력을 받는 대화창이 나타난다. 발화자 정보를 모두 입력한 후에 발화자 정보 저장 버튼을 선택하면 사용자로부터 입력 받은 정보를 저장한다.

화자번호	이름	발화목록번호	성별	나이	키	몸무게	방언	12세미전거주지	원어민구분	출생지
4			Male						Native	
3			Male						NonNative	
2			Male						Native	
1			Male						Unknown	

발화자정보 저장

[그림 IV-20] 발화자 정보 입력 대화창

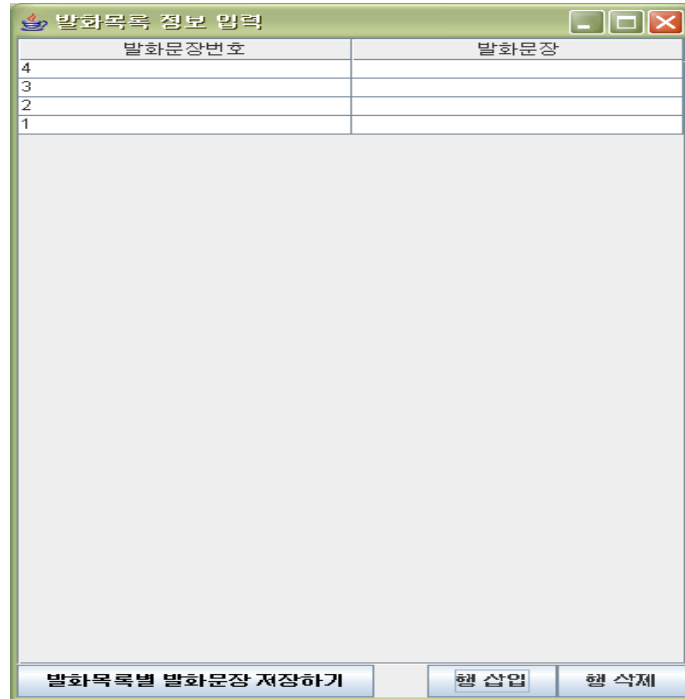
[그림 IV-15]에서 발화목록 정보 메뉴를 선택하면 [그림 IV-21]과 같이 발화목록 정보 입력을 받는 대화창이 나타난다.

발화목록번호	발화목록설명
1	

발화목록정보 저장 발화목록별 발화문장 입력하기 행 삽입 행 삭제

[그림 IV-21] 발화목록 정보 입력 대화창

여러 개의 발화목록 중 한 행을 선택하고 발화목록 별 발화문장 입력하기 버튼을 선택하면 [그림 IV-22]과 같이 발화목록 별 발화문장 입력 대화창이 나타난다.



[그림 IV-22] 발화목록 별 발화문장 입력 대화창

한 개의 발화목록에 입력할 발화문장을 모두 입력한 후 발화목록 별 발화문장 저장하기 버튼을 선택하면 사용자가 입력한 발화문장이 저장된다. 여러 개의 발화목록이 있다면 [그림 IV-22] 발화목록 별 발화문장 입력 대화창을 여러 번 열어 입력하고 발화목록 별로 저장한다.

위와 같이 발화목록, 발화자, 수집채널 정보만을 입력하거나 추후에 또 정보를 입력하고자 하는 경우, [그림 IV-15]에서 음성 DB 메타데이터

임시저장 버튼을 선택하면 스키마 유효성 검사를 하지 않고 디폴트 문서에 그대로 정보를 저장하게 된다.

모든 정보를 입력하여 음성 구축이 완료되었다면 [그림 IV-15]에서 음성 DB 메타데이터 작성완료 버튼을 선택한다. 스키마 유효성 검사를 하고 스키마에 적법하면 메타데이터로 저장한다.

V. 결론 및 향후 과제

본 논문에서는 산업용 음성 DB 메타데이터를 위한 표준안을 XML 기반으로 제시하고, 표준화된 메타데이터 관리도구를 설계 및 구현하였다. 본 논문에서 제안하는 표준안은 ETRI와 SITEC에서 기존에 제안한 규격 [7]를 객체 지향 구조로 보완하였다. 화자, 채널, 발화목록, 파일, 전사 정보에 각각의 식별자를 정의하여 구성요소 사이의 관계를 식별자의 참조로 표시할 수 있도록 하였다. 또한 연관되는 엘리먼트들의 그룹핑과 계층구조를 도입하여, 메타데이터의 구조를 체계화하여, 향후 음성 DB의 검색이나 재사용이 용이하도록 설계하였다.

관리도구에서는 다음과 같은 기능을 제공한다.

- 첫째, 새로운 음성 DB 메타데이터 생성 기능.
- 둘째, 기존에 존재하는 음성 DB 메타데이터의 삽입, 수정, 삭제 기능.
- 셋째, 음성 DB 구축 중 메타데이터 임시 저장 기능
- 넷째, 대상 음성 DB 메타데이터의 기본정보 검색 기능.
- 다섯째, 대상 음성 DB 메타데이터에 존재하는 발화자의 지역, 연령, 성별로 구분하여 발화자 정보를 검색하는 기능.
- 여섯째, 검색 조건에 맞는 화자가 발화한 음성파일을 검색, 발화한 음성 파일을 재생하는 기능.

XML 기반 산업용 음성 DB를 위한 메타데이터 표준안은 기존 구축되어 있는 음성 DB와 향후 구축될 다양한 음성 DB의 표준 메타데이터 기술 방법으로 활용 가능하다. 또한 표준 메타데이터의 빠른 보급과 확산을

위하여 설계된 산업용 음성 DB 메타데이터 관리도구의 개발로 향후 대용량 음성 DB의 메타데이터 구축이 손쉬워 질 것이다. 산업용 음성 DB 메타데이터 관리도구는 대용량의 산업용 음성 DB의 부분적 재사용 및 재구성을 활성화하는데 활용할 수 있다.

향후 계획으로는 본 관리도구는 하나의 음성 DB 메타데이터만을 대상으로 생성, 검색하도록 되어 있는데 차후 같은 정보를 여러 개의 메타데이터에 저장하거나 여러 개의 메타데이터를 한번에 같은 조건으로 검색 할 수 있는 기능을 추가하여야 할 것이다.

참고문헌

- [1] ELRA, “Description form-PROVIDER Language Resources : Speech,” http://www.elra.info/services/speech_1.4.rtf, pp.1-4, 2004.
- [2] LDC, “LDC Catalog by Type And Source,” <http://www ldc.upenn.edu/Catalog>, 2004.
- [3] Cisco Systems Inc, “Speech Application Language Tags (SALT) 1.0 Specification,” <http://www.saltforum.org/devforum/spec/SALT.1.0.a.asp>, 2003.
- [4] A. Hunt, S. McGlashan, “Speech Recognition Grammar Specification Version 1.0,” <http://www.w3.org/TR/2004/REC-speech-grammar>, 2004.
- [5] 홍기형, 이육재, “국제 음성기술 표준화 동향과 대응,” 음성통신 및 신호처리 학술대회 논문집, 20 권, 1 호, pp.185-188, 2003.
- [6] 김상훈, 이용주, “음성 DB 표준화,” 음성통신 및 신호처리 학술대회 논문집, 20 권, 1 호, pp.181-184, 2003.
- [7] 김상훈, 이영직, 한민수, “음성 DB 부가 정보 기술방안 표준화를 위한 제안,” 말소리, 제 47 호, pp.110-119, 2003.
- [8] D.C. Fallside, P. Walmsley, “XML Schema Part 0: Primer Second Edition,” <http://www.w3.org/TR/xmlschema-0/>, 2004.
- [9] L. Wood, “Document Object Model (DOM) Level 1 Specification,” <http://www.w3.org/TR/REC-DOM-Level-1/>, 1998.
- [10] J. Clark, “XSL Transformations (XSLT) Version 1.0,” <http://www.w3.org/TR/xslt>, 1999.

- [11] M.Kay, XSLT Programmer's Reference, 정보문화사, 2001.
- [12] 박미나, 김석, 윤상민, 남형주, XML & XSLT 실무 테크닉, 글로벌, 2000.
- [13] D. Raggett, A. Hors, I. Jacobs, "HTML 4.1 Specification," <http://www.w3.org/TR/html4/>, 1999.

ABSTRACT

Design and Implementation of an XML based Meta-data Management Tool for Speech Databases

Young-Hee Joo

Department of Education Computer Science

Graduated School of Education

Sungshin Women' s University

In this paper, we propose a meta-data specification for industrial speech databases and implement the management tool for the meta-data. In order to advance the reusability and portability of speech databases, the proposed meta-data is based on XML. We first define the speech databases more formally and identify objects appearing in speech databases, and then design the meta-data model for speech databases in an object-oriented way.

The management tool offers meta-data search and construction function. Users of the management tool can search not only the general information for a specific speech database, but also speaker's information, speech files and transcription information by specifying searching conditions on speakers, channels and utterance list.

By using the management tool, speech DB constructors can create and modify the meta-data for speech databases, easily and efficiently.

감사의 글

지난 2년 6개월 동안 항상 저의 곁에서 새로운 기회와 많은 경험을 하게 해주신 홍기형 교수님께 진심으로 감사를 드립니다. 매 주 세미나 때마다 좋은 말씀과 따뜻한 격려를 해주신 유원경 교수님께도 감사를 드립니다.

대학원 생활 초기부터 지금까지 연구실 생활을 함께한 사랑하는 지혜와 재영, 지영에게도 감사의 마음을 전합니다. 다른 연구실이지만 많은 시간을 함께한 문정 언니, 선숙 언니에게도 감사의 말을 전합니다. 우리가 함께 한 시간과 노력들이 좋은 결과로 이어지기를 기원합니다.

연구실 후배 경현, 수현에게도 감사의 마음을 전합니다. 짧지만 즐거운 시간을 같이 한 이경님 선생님, 이경아 선생님께도 감사를 드립니다. 비록 자주는 못 봤지만 함께 고민해 준 연구실 선배 진숙 언니, 하정 언니에게도 감사합니다.

대학교, 대학원 생활 동안 저에게 항상 활력소가 되어주고 마음을 나눈 UNSC 동기들 윤덕, 재량, 대홍, 소연, 은성, 혜경, 지훈, 정은, 종렬, 영일, 용희에게도 감사합니다. 저에게 무슨 일이 있을 때마다 함께 고민해주는 사랑하는 진희 언니, 병하 오빠, 병준, 소라, 보라, 이모, 준현, 태량, 윤재에게도 감사의 마음을 전합니다. 저에게 늘 친언니 같이 조언과 격려를 해준 시은 언니와 정민 언니에게도 감사합니다. 한결 같은 마음으로 저희 가족을 아껴주시는 큰어머니, 큰아버지께도 감사를 드립니다. 저 멀리 시골에서 항상 저에게 사랑을 주시는 외할머니께도 감사합니다.

대학원 연구실 생활을 시작할 때는 모든 일이든 열심히 하겠다던 저의 의지와는 다르게 점점 시간이 지날수록 초심을 잃은 저의 모습이 졸업을 앞 둔 지금, 많이 후회됩니다. 힘들고 지칠 때 항상 저에게 다시 도전하게끔 만드는 원동력인 저의 부모님께 진심으로 사랑과 감사의 말을 전합니다. 마냥 어린애 인줄로만 알았던 저의 사랑하는 동생, 이제는 어엿한 직장인이 된 성희에게도 감사의 마음을 전합니다.

마지막으로, 저 멀리 있는 P에게 사랑과 감사의 마음을 전합니다.

2006년 7월 연구실에서

주 영 희 올림