음성출력/학습기능을 지원하는 컴퓨터용어 약어 전자사전 설계

*김홍섭,. 이현걸, 김철호, 이금석
동국대학교 교육대학원 전산교육과

A Design of Electronic Dictionary for Computer English Abbreviation Supporting Voice and Study Functions

Kim. Hong-Seop, Lee. Hyun-Geol, Kim. Cheol-Ho, Lee. Keum-Seok
Computer science education graduate course of Dong-Kuk University

컴퓨터용어에 대한 약어들을 기존 사전관리 방법에서 벗어나 컴퓨터에 수록하여, 음성을 포함한 한글 및 영문풀이 조회, 추가등록, 삭제, 수정을 가능케 하여 어휘변화 및 의미파악에 신속히 대처하고, 영문, 한글 단어를 이용한 약어검색 기능과 시스템에서 임의로 문제를 출제, 학습자 수준을 측정해볼 수 있는 컴퓨터 보조학습(CAI)과 knowledge base 교체시 타분야에서 활용이 가능하도록 DB화한 약어 전문가체제로, 업무 활용자(학습자)와 컴퓨터 상호작용에 의한 개인차를 극복할 수 있도록 컴퓨터용어 약어 전자사전을 설계하였다.

*김홍섭, 직장 : 특전사령부 전산실장: TEL) 02-402-0345
409-0541~9 (교 5530)
학교 : 동국대학교 교육대학원 전산교육과 266-5498.
자택 : 서울시 송파구 거여동 사자 APT 101동 104호 404-9233
이력 : '82 동국대학교 전산학과 졸, '83 정보사령부 전산장교
'87 정보학교 전산교관, '88~ 현 특전사령부 전산실장
'90 숭실대학교 정보과학대학원 컴퓨터공학 전공
'92 현재 동국대학교 교육대학원 전산교육과 3학기 재학중

음성출력/학습기능을 지원하는 컴퓨터용어 약어 전자사전 설계

*김홍섭, 이현걸, 김철호, 이금석
동국대학교 교육대학원 전산교육과

A Design of Electronic Dictionary for Computer English Abbreviation Supporting Voice and Study Functions

Kim. Hong-Seop, Lee. Hyun-Geol, Kim. Cheol-Ho, Lee. Keum-Seok
Computer science education graduate course of Dong-Kuk University

요 약

컴퓨터용어에 대한 약어들을 기존 사전관리 방법에서 벗어나 컴퓨터에 수록하여, 음성을 포함한 한글 및 영문풀이 조회, 추가등록, 삭제, 수정을 가능케 하여 어휘변화 및 의미파악에 신속히 대처하고, 영문, 한글 단어를 이용한 약어검색 기능과 시스템에서 임의로 문제를 출제, 학습자 수준을 측정해볼 수 있는 컴퓨터 보조학습(CAI)과 knowledge base 교체시 타분야에서 활용이 가능하도록 DB화한 약어 전문가체제로, 업무 활용자(학습자)와 컴퓨터 상호작용에 의한 개인차를 극복할 수 있도록 컴퓨터용어 약어 전자사전을 설계하였다.

I. 서 론

정보화사회의 발전추세에 따라 컴퓨터에 관련된 용어들이 폭발적으로 증가하고 있으며, 현대인의 편리성과 단순성 추구로 용어표기에 따른 어려움을 해소하고 상호의사소통을 보다 빠르게 전달하기 위하여 축약된 약어들의 사용이 필수적인 요소로 자리잡아 가고 있다. 예를들면, CPU(Central Processing Unit)는 그 의미가 '중앙처리장치'라는 것을 이젠 자연스럽게 인식되고 있으며, WYSIWYG(What you see is what you get) 와 같은 약어까지 등장하고 있어 컴퓨터 용어 약어 상호정의 및 효과적인 관리 필요성이 증대되고 있다.

일반적으로 정치, 경제, 사회등 여러 분야에서 사용하는 용어들을 살펴보면 모두들 나름대로의 고유용어들이 존재한다는 것을 알 수 있다. 뿐만아니라 이들 약어의 사용은 어려운 용어와 많은 단어로 표현된 개체를 가장 간단한 철자로써 대략 3~8자로 함축되어 있다. 이로인해 동일한 개체에 대해 정의하는 사람마다 2~3개의 약어가 할당되는 경우도 발생하여 일반적인 언어들이 추구하는 객관성이 결여되기도

* 특전사령부 전산실

한다. 또한 다수의 개체가 동일한 형태의 약어로 함축되어지는 현상도 약어가 지니고 있으며, 이 외에 새로운 개체가 생성되고 변화하며 낡은 것은 쇠퇴하여 소멸함에 따라 약어 역시 함께 생성, 변화, 소멸되는 현상도 약어의 중요한 특성이라 할 수 있다.

실제로 컴퓨터를 활용함에 있어 컴퓨터 용어 약어는 이미 상식이 되어버렸음에도 불구하고 날로 변화해가는 약어들에 대한 학습 및 관리가 매우 어렵다는 결점이 있다. 따라서 최신 어휘의 변화와 방대한 어휘사전 관리에 대한 대책이 요망되었다.

이에 현재 책으로 편찬된 사전에 의하지 않고 개인용 컴퓨터를 이용하여 약어의 관리를 효과적으로 수행할 수 있도록 음성출력 및 학습기능을 제공하는 시스템을 설계해 보고자 한다.

Ⅱ. 시스템 구성

1. H/W

『컴퓨터용어 약어 전자사전』은 IBM PC/AT를 사용하여 개발하였으며, 호환기종인 XT, 386에서도 무난히 수행이 된다. 주메모리는 640KB정도면 충분하고, Video Card는 CGA, MDA, EGA, VGA를 지원하며, Sound Blaster Card가 장착되어 있을 경우 음악과 음성 기능을 지원한다. 프린터는 한글(완성형, 조합형 Code) 을 지원하는 제품이면 사용 가능하며, 음성기능 사용시 보조 기억장치(hard disk)의 용량은 약 30MB 정도 필요하고 Text 기능만 사용하는 경우 2D disk 1장으로 수행 가능하다.

2. S/W

개발에 사용한 S/W 로는 우선 OS로써 MS-DOS Ver.5.0이 있으며 개발 환경으로 Borland사의 Turbo-C Ver.2.0을 이용하였다. 음성 처리에는 Sound Blaster Card에서 지원해주는 Library가 이용 되었으며, Ad-Lib Composer를 이용하여 음악을 입력하였다. 또한 Graphics 화면에 그림을 나타내기 위하여 Image256과 Image Scanner가 사용 되었다.

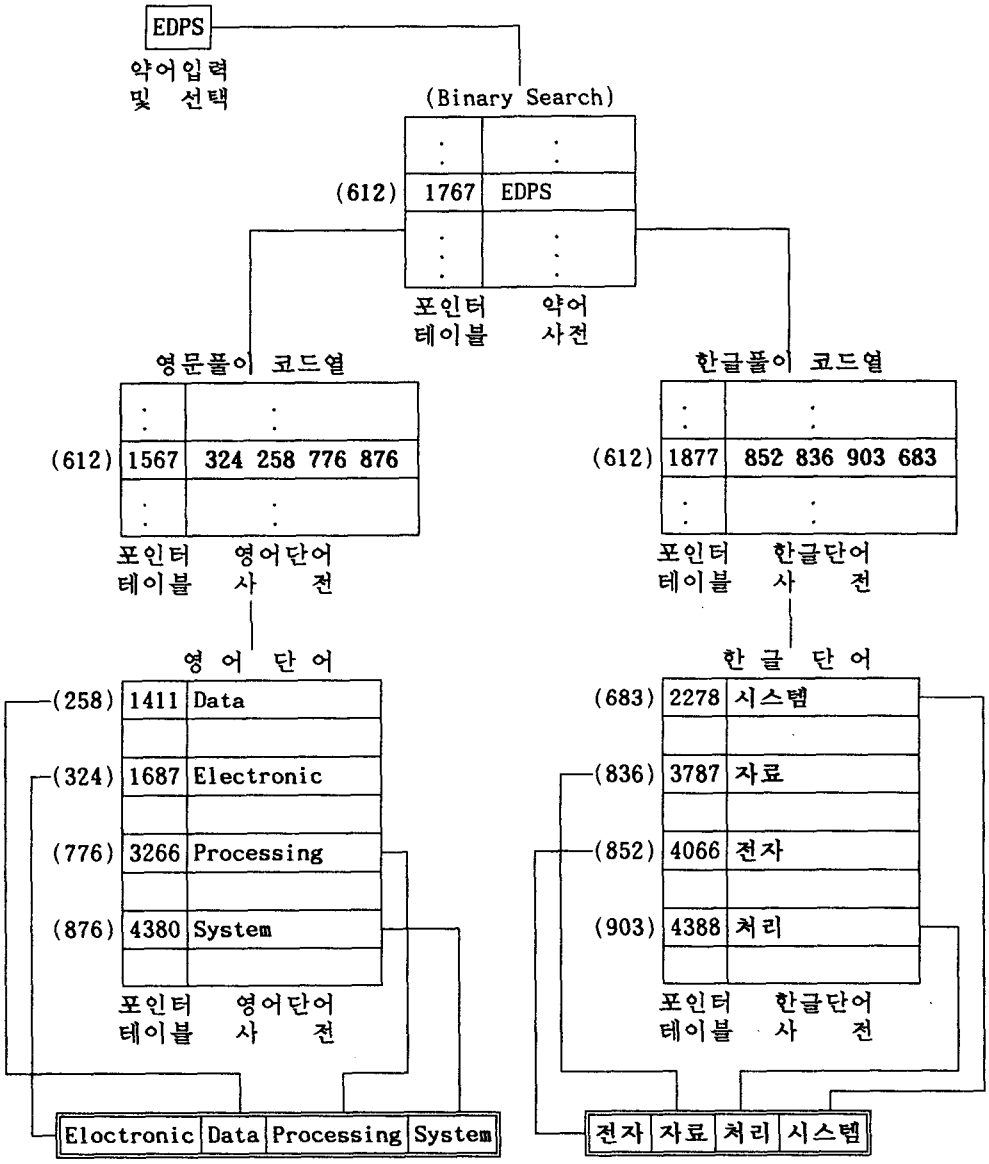
Ⅲ. 시스템 설계

1. 약어통합사전

일반적으로 약어는 전문단어들의 조합으로 이루어져 있으며, 한정된 전문단어들의 조합으로 훨씬 많은 약어들을 만들 수 있다. 따라서 본 시스템에서는 <그림 1>과 같이 다섯 개의 사전과 음성연결 테이블을 구성함으로써 전문 단어는 단어사전에 두고 모든 단어들을 코드화 하여 불필요한 메모리의 낭비를 방지하고 효과적으로 약어를 관리할 수 있도록 설계 하였다.

『약어통합사전』은 크게 영어단어사전, 한글단어사전, 약어사전, 영문코드사전, 한글코드사전의 다섯 개의 사전으로 구분되어져 있으며, 전자의 3개 사전은 ASCII 코드 문자열과 한글 완성형(KS5601) 코드 문자열로써 구성되어져 있고 후자의 2개 코드사전은 2byte 정수열로써 구성되어져 있다.

약어 조회시 약어는 약어사전에서 2진 검색을 통하여 그 위치를 얻게 되며, 이것은 다시 각각 영문코드사전과 한글코드사전을 통해서 2byte 정수열을 얻는다. 이 정수열의 정수는 각각 영어단어와 한글단어의 코드를 나타내며 각각의 코드에 해당되는 단어들을 나열함으로써 영문풀이와 한글풀이 문자열을 얻는다. <그림 2>는 약어 'EDPS'의 조회경로를 보여주고 있다.



<그림 2> 약어 'EDPS'의 조회 경로

또한 음성연결 테이블을 두어 각각을 음성과 연결하고 있다. 음성연결 테이블에서 -1은 음성이 등록되어 있지 않음을 나타낸다.

0	Abbreviation
1	Abnormal
2	Absolute
3	Acceptable
	. . .

영어단어사전
1140건, 8723 byte

0	가감
1	가변
2	가산기
3	감산기
	. . .

한글단어사전
1050건, 6185 byte

0	ABBR
1	ACR
2	ADAPT
3	EDPS
	. . .

약어사전
1501건, 6841 byte

0	132
1	68
2	162 1349
3	105
	. . .

영문코드사전
1501건, 9005 byte

0	1023
1	16 511
2	756
3	325 1498
	. . .

한글코드사전
1501건, 8731 byte

0	0
1	2
2	1
3	-1
	. . .

음성연결 테이블
1501건, 6000 byte

<그림 1> 약어통합사전 구조

약어통합사전은 <그림 1>에서 보는 바와 같이 각 단어 또는 코드열의 길이가 가변적인 이유로 사전을 읽어올 때 각각 포인터(pointer) 테이블을 생성하는데 이 포인터 테이블은 메모리에서 각 단어 또는 코드열의 위치를 가리키며 단어 또는 코드열을 얻을 때 참조하게 된다.

위와 같은 방식으로 통합사전을 구성하였을 경우 실제로 본 시스템에서 약어의 갯수는 약 1500개인데 비해 영어단어는 약 1100개, 한글단어는 약 1000개로써 사전의 크기는 54KB 정도 되며, 단어를 코드화 하지 않고 모두 text file로 풀이하였을 경우 file의 크기는 78KB가 된다. 따라서 (1)에 의하여 본 시스템은 약 31%의 메모리절감효율을 갖게 되며 <표 1>에서 보는 바와 같이 단어를 조합하여 약어를 풀이하는 방식이 메모리 절감 면에서 효과적인 것을 알 수 있다.

$$\text{메모리 절감효율} = \frac{(\text{Text File의 크기}) - (\text{약어통합사전의 크기})}{(\text{Text File의 크기})} \times 100$$

$$= \frac{78 - 54}{78} \times 100 \fallingdotseq 30.8\ (\%) \qquad (1)$$

약어 갯수	통합 사전 크기	Text File 크기	메모리 절감 효율
300	10, 317	12, 143	15. 04 %
600	22, 109	27, 972	20. 96 %
900	32, 600	41, 839	22. 08 %
1200	43, 252	57, 550	24. 84 %
1501	54, 451	78, 768	30. 87 %

<표 1> 약어의 갯수에 따른 메모리 절감 효율 (단위 : byte)

2. 음성출력

현재 인공지능분야의 연구가 활발히 이루어지면서 음성합성 및 인식에 대한 연구도 많은 성과를 거두고 있다. 그러나 현재까지의 기술로도 사람의 말이나 억양, 강약 등을 구현하기에는 앞으로도 더욱 많은 연구와 발전의 여지가 있다. 따라서 본 시스템에서는 사람의 음성을 완벽하게 재생할 수 있는 파형부호화[1] 방식을 채택하였다 [5].

사람이 들을 수 있는 주파수 영역은 대략 20Hz~20KHz로 알려져 있으며, 보통 사람들이 말하는 음성은 4KHz를 넘지 않는다. 따라서 중첩현상이[2] 발생하지 않도록 하기 위해서는 적어도 8KHz 이상의 주파수가 필요하지만 본 시스템에서는 보조기억장치의 용량을 고려하여 6KHz를 사용하였으며, 6KHz로 양자화 하여도 음질이 저하되지 않고 충분히 복원되는 것을 확인하였다.

양자화 하여 얻어진 음성 Data는 양이 많기 때문에 보조기억장치 (hard disk)를 이용하여 관리하여야 하는데 File의 이름을 설정하는데 있어서 시스템 분석에서 밝힌 바와 같이 형태는 같으나 서로 다른 의미를 지닌 약어가 존재할 수 있으므로 약어 사전에서의 약어의 위치를 이용하여 영문풀이 음성의 File 이름은 'VOC0873.ENG', 한글풀이 음성의 File 이름은 'VOC0873.HGL' 등과 같이 나타내었다. 이와 같은 방법을 사용하였을 때 새로운 약어가 추가되거나 기존의 약어가 삭제되는 경우에 따라 약어의 위치가 변하게 되는 경우가 발생하는데 이 경우 data file의 이름을 모두 변경한다는 것은 hard disk의 access 속도를 고려할 때 비효율적이다. 따라서 <그림 1>에 보인 바와 같이 음성연결 테이블을 두어 참조하게 함으로써 약어의 추가, 삭제시 음성연결 테이블의 내용만을 바꾸어 줄 수 있도록 설계하였다.

음성합성 방식에 비하여 파형부호화 방식은 사람의 음성을 완벽하게 재생할 수 있다는 잇점이 있지만, 반면에 모든 data를 직접 마이크를 이용하여 녹음하여야 하고, 따라서 음성 data의 양도 방대해 진다는 단점이 있다.

또한 같은 단어라 하더라도 문장에서의 위치와 쓰임에 따라 억양과 강약이 다르기 때문에 약어통합사전과 같이 코드화 하여 관리할 수도 없다. 그러므로 모든 문장을 직접 입력해야 하는데 이러한 경우 막대한 양의 음성 data 관리에 문제점이 발생한다. 실제로 본 시스템에서 입력한 약 1500 개의 약어에 대한 음성을 모두 입력한 결과 약 30MB의 data와 3000 여개의 file을 관리해야 하는 문제점이 발생하였다.

본 시스템이 운영체제로서 사용하고 있는 MS-DOS를 살펴볼 때 hard disk에서 한 개의 subdirectory에 포함할 수 있는 file이나 subdirectory의 수는 64개 미만이 적당하며 64개 초과시마다 access 속도가 조금씩 느려진다. 따라서 3000여개의 File을 한 개의 subdirectory에서 관리한다면 access 속도는 무척 느려질 것이고 실제로 약 1500 개의 File이 한 개의 subdirectory에 존재할 경우 한 개의 file을 access하는데 평균적으로 약 3초의 시간이 소요된다는 것이 확인되었다.

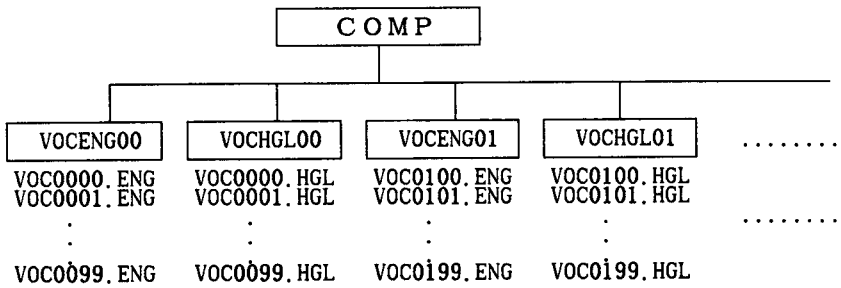
---

1) 파형부호화 : analog 파형을 digital로 변환하여 양자화한 data를 기록해두고 합성시는 digital에서 analog로 다시 변환하여 재생하는 방식으로 자연음에 가까운 음성을 재생할 수 있으나 주파수가 높아 자료량이 매우 많다는 단점이 있다. 컴퓨터 교육연구 제1권, 교육용 S/W에서의 음성지원 방법, 1992. 4. 123 - 129 page 참고.

2) 중첩현상 : analog 파형을 양자화할 경우 analog 파형 최대 주파수의 2배 이상으로 양자화하지 않으면 복원시 파형이 서로 중첩되어 잡음이 발생한다.

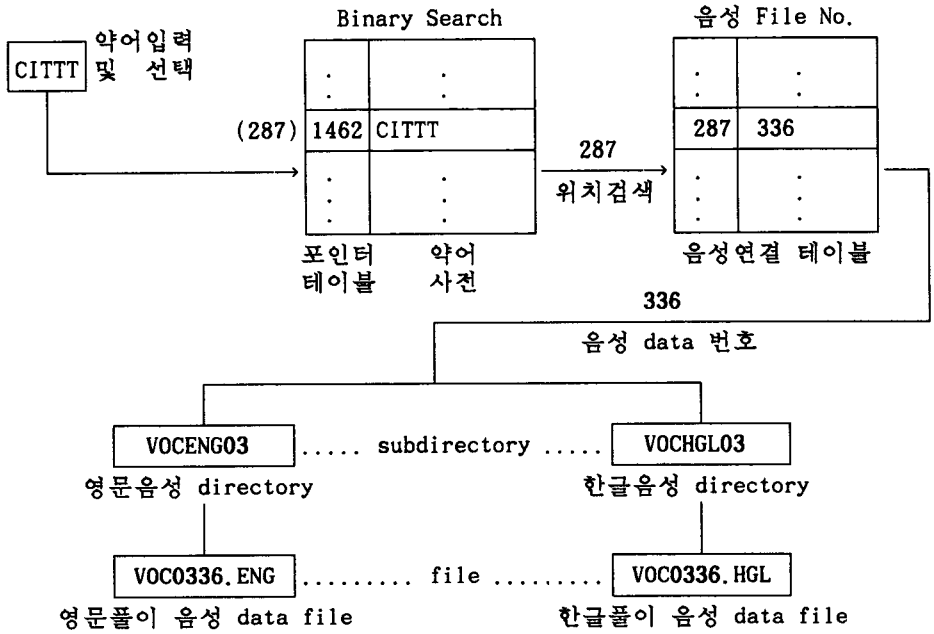
영어 혹은 한글 풀이의 음성을 듣기 위하여 3초의 시간을 기다려야 한다는 것은 매우 답답하고 비효율적인 일이 될 것이다.

위의 문제를 해결하기 위한 방법으로 여러 개의 subdirectory를 두어 data file을 분산하는 방법을 고려하였다. 그 결과 <그림 3>과 같이 영어 16개, 한글 16개의 총 32개의 subdirectory를 두고 각각의 subdirectory에는 100개의 data file을 두어 관리함으로써 file access 시간의 문제를 해결하였으며 실제로 이와 같은 방법으로 data를 관리한 결과 file access 시간은 거의 무시할 정도에 지나지 않는다는 것이 실증되었다.



<그림 3> 음성 Data File의 Directory 구조

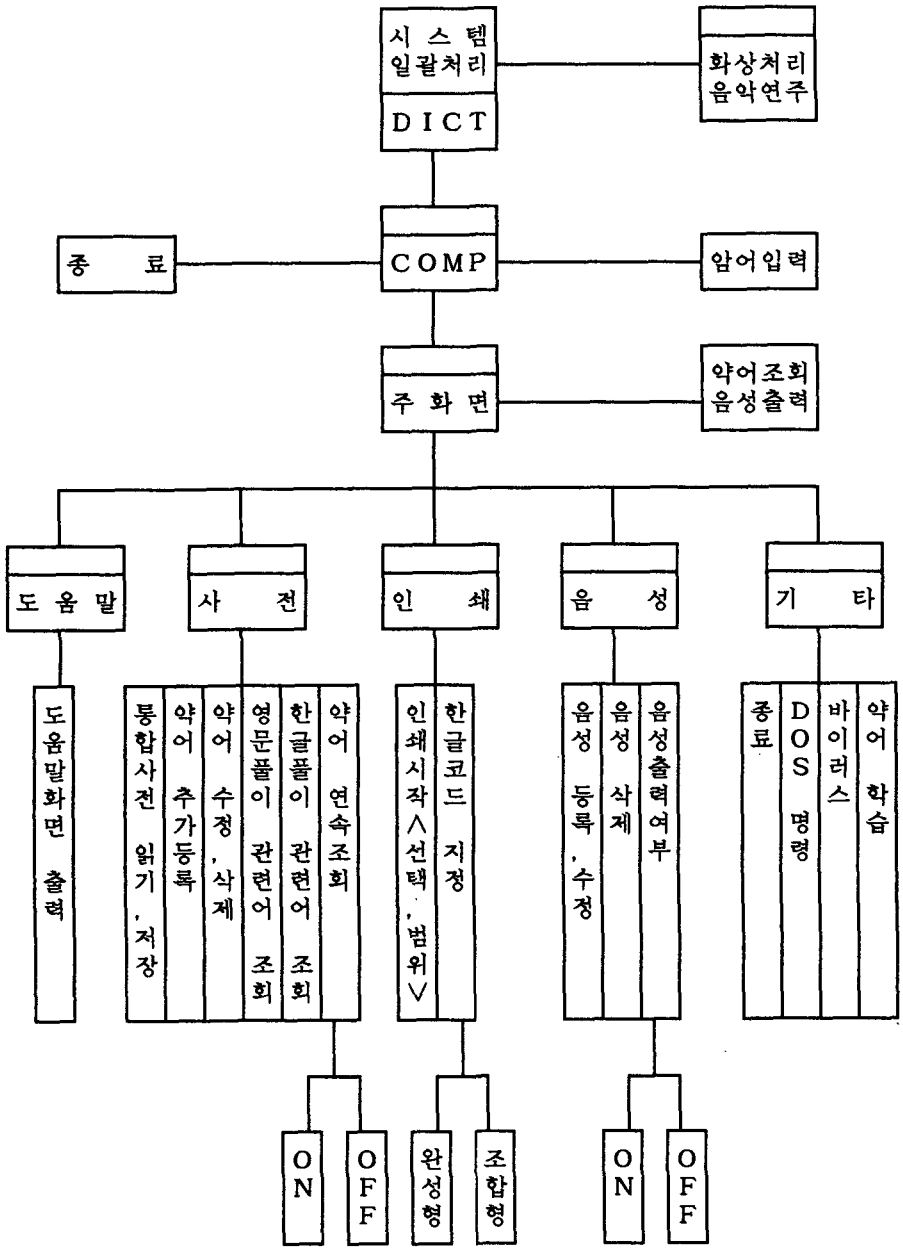
악어의 음성 출력시 악어사전을 통하여 2진 검색에 의해서 악어의 위치를 얻게 되며 다시 음성연결 테이블을 참조하여 음성 data 번호를 얻게 되고 이 번호를 가지고서 음성 file의 이름과 directory이름을 얻게 된다. <그림 4>에 악어 'CCITT'의 음성 data file 검색 경로를 나타내었다.



<그림 4> 악어 'CITTT'의 음성 Data File 검색 경로

3. 설계도

컴퓨터용어 약어 전자사전의 주요 설계 내용을 도표로 나타내 보면 <그림 5>와 같다. 화상처리 및 음악연주는 특정 소속기관의 이미지를 부각시키기 위한 logo화면으로 생략시 주화면의 실행화일 COMP 명령으로 실행이 가능하다.

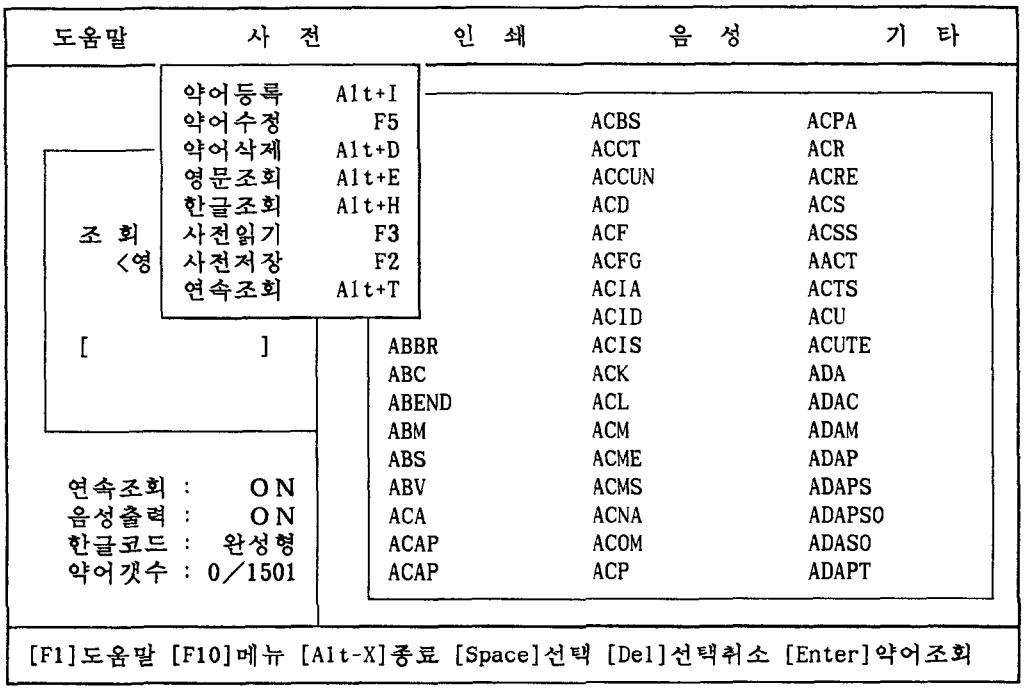


<그림 5> 컴퓨터용어약어 자동번역 시스템 설계 내용

Ⅳ. 시스템 구현

1. 주메뉴

부메뉴 및 세부기능 설명은 생략하였으며 주메뉴는 <그림 6> 과 같다.



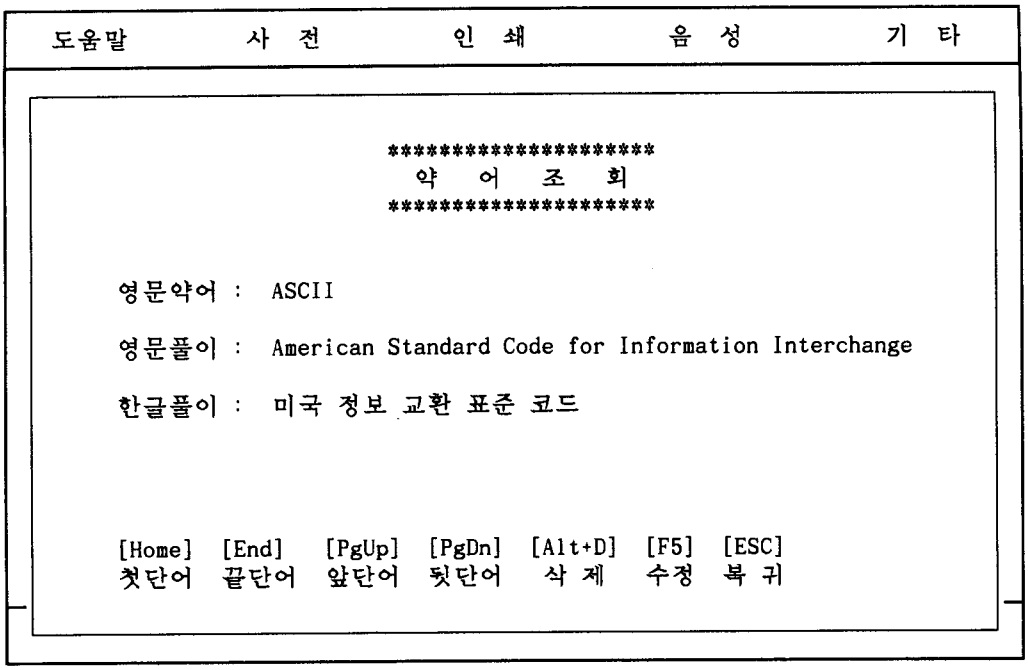
<그림 6> 자동번역 시스템의 주메뉴

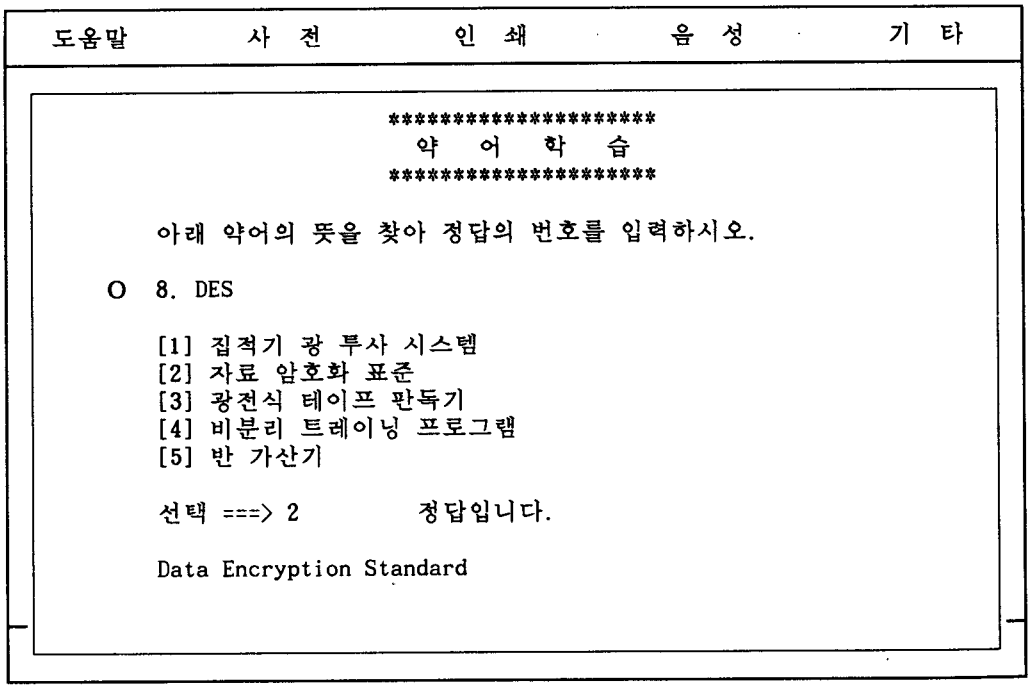
2. 특징

(가) VGA에서 지원해주는 색상을 최대로 이용한 full-down menu 방식 이용
(나) 음성처리기능 지원으로 시청각 효과 실현
(다) 코드화된 단어구조를 통하여 신속한 약어 검색 및 메모리 절약
(라) 단축키(hot-key) 이용, menu를 거치지 않고 직접 원하는 작업 가능
(마) 영어 혹은 한글 단어를 통한 관련 약어 검색
(바) 연속조회와 선택조회 기능 지원
(사) 약 1500건에 달하는 풍부한 약어관리와 최신약어의 추가적인 등록 가능
(아) 출력장치(printer)를 이용한 사전 인쇄 (한글 완성, 조합형 코드 가능)
(자) 약어 학습 기능을 통한 약어 암기 (CAI, CBE)
(차) 과학, 행정, 예술 제분야 DATA 입력시 활용가
(Knowledge Base 교체로 해당분야 약어번역 전문가 체제로 발전)

3. 약어조회 및 학습 화면

제분야의 업무활용 및 학습자 개인차에 의한 어휘 등록, 수정, 삭제가 가능한 약어조회 화면을 나타내 보았으며, 시스템에서 출제하는 임의의 문제를 test해볼 수 있는 학습화면을 아래에 예시하였다.





V. 결론

이상에서 음성출력 및 학습기능을 지원한 컴퓨터용어약어사전을 설계하여 보았다. 본 시스템은 인간의 인지구조와 유사하게 비순차적(nonsequential), 즉 direct 검색 방법과 첨단 multimedia를 최대한 활용, 사용자 편리성을 강조한 hypermedia 기법을 이용하여 컴퓨터 용어 약어의 조회, 등록, 삭제, 수정을 보다 빠르고 용이하게 하여, 최신 약어의 어휘변화에 대처하며, 영문 또는 한글단어에 대한 관련 약어를 찾을 수 있도록 설계해 보았다. 또한 그래픽 및 음성출력으로 인한 시청각 효과를 얻음으로써 약어에 대한 정보검색과, 임의의 문제를 출제, test하여 약어를 손쉽게 이해할 수 있도록 하는 약어학습 및 작업을 처리하는 동안 계속적인 배경음악의 동시 연주와 MS-DOS Shell기능을 제공하고, 필요에 따라 도움말을 얻을 수 있도록 배려하였다. 본 시스템은 PC-286 (AT) 상에서 Turbo-C 2.0을 이용하여 구현, 컴퓨터 활용 방안을 검토 시험 운영한 바 XT, 386 호환성 및 학습자 개인차에 따른 학습환경을 제공할 수 있는 컴퓨터 보조학습 도구로서 약어관리에 있어 질적 변혁이 가능하게 하는 전문가 시스템으로 만들 수 있었다.

이 외에도 과학, 경제, 예술, 교육 제분야 영어 약어에 대해서도 데이타 입력 (Knowledge Base 교체) 시 자동번역의 가능성을 확인하였으며, 업무활용 및 컴퓨터 보조학습 성과를 향상시킬 수 있는 약어자동번역시스템, 약어전자사전, 대형 시스템에서 통합 S/W 일부분으로 할당, 다수인 동시 사용과 전문가체제와 접목시키는 연구를 계속 하여 발전시켜 나아간다면 컴퓨터 활용뿐만 아니라 인간의 의사 소통과 지성을 향상시키는 학습용 전문가 시스템으로서 사용되리라 제안해 본다.

<참고 문헌>

[1]. Nielsen, Jakob, Hypertext & Hypermedia, Academic Press Inc. 1990.

[2]. Shneiderman, Ben, Designing the User Interface, Addison Wesley Publishing Company, 1992.

[3]. Shneiderman, Ben & Greg Kearsley, Hypertext Hands-on, Addison Wesley Publishing Company, 1991.

[4]. 이석호, 공학 논문 작성법, 정익사, 1991.

[5]. 강명희, 교육용 S/W에서의 음성 지원 방법, 전국 학교 컴퓨터 연구회, 1992. 4

[6]. 신정구, 교육공학매체를 통한 영어교육활용화 방안책, 성균관교육대학원, 1990.

[7]. 안성근, 컴퓨터용어대사전 한국정보과학회편, 정익사, 1984.

[8]. 안성근, 표준컴퓨터 용어사전, 정익사, 1990.

[9]. 컴퓨터 용어사전 편찬위원회편, 컴퓨터용어략어사전, 크라운 출판사 1992.

[10]. 양단희, 한국어 전자사전 원형의 설계및 구현, 연세대학원 석사학위론문, 1991.

[11]. 원영주, 멀티미디어 기술과 군사적 응용, 육군사관학교 화랑대 연구소, 1992.7.

[12]. 한성국, IBM PC/XT/AT/PS/2 기술사전, 집문당, 1989.

[13]. 정택희 · 손병길 편저, CAI 개발과 활용, 교학사, 1989.

[14]. 영한 변환에 관한 기초연구 및 기계사전의 개발, 시스템공학센터, 1989.

[15]. 제 3회 학술발표회 논문집 "인간과 기계언어", 한국 정보과학회, 1991.