

목 차

제 I 장. 사업 개요	1
제1절. 사업의 목적	1
제2절. 사업 목표	1
제3절. 사업 내용 및 추진 경과	1
1. 21세기 세종계획 성과물 관리 지원	2
2. 21세기 세종계획 성과물 배포	2
3. 한국어 연구 정보 제공 및 민원 상담	3
4. 사업 추진 경과	4
제 II 장. 21세기 세종계획 성과물 관리 지원	5
제1절. 성과물 배포 대상 선정	5
1. 오프라인 배포 대상 성과물	5
2. 오프라인 배포 대상 성과물 안내서	8
제2절. 성과물 헤더 변환	14
1. TEI-P3 말뭉치 형판	14
2. TEI-P5 말뭉치 형판	21
3. 헤더 변환 방법	28
4. 도구 사용법	31
5. 헤더 변환 결과	35
제 III 장. 21세기 세종계획 성과물 배포	40
제1절. 21세기 세종계획 성과물 관리 DB	40
제2절. 21세기 세종계획 온라인 배포 시스템 개발	41
1. 배포 시스템 메뉴 구성	41
2. 자료 유형별 상세 정보	42
3. 배포 시스템 사용자 인터페이스	44
제3절. 21세기 세종계획 오프라인 배포	48
제 IV 장. 국어 연구 정보 제공 및 민원 상담	50
제1절. 21세기 세종계획 홈페이지 운영	50

제 V 장. 사업 결과물	54
제1절. 21세기 세종계획 성과물 관리 지원	54
제2절. 21세기 세종계획 성과물 배포	54
제3절. 국어 연구 정보 제공 및 민원 상담	54
제4절. 보고서 작성	55
제 VI 장. 전산 파일(CD-ROM) 설명	56
제1절. “0. 보고서” 폴더	56
제2절. “1. 21세기 세종계획 성과물 관리” 폴더	56
제3절. “2. 21세기 세종계획 성과물 배포” 폴더	56
제4절. “3. 국어 연구 정보 제공 및 민원 상담” 폴더	57

표 목 차

[표 1] 2007년도 사업 추진 일정	4
[표 2] 전자 사전 구축 성과	10
[표 3] “0. 보고서” 폴더의 내용 설명	56
[표 4] “1. 21세기 세종계획 성과물 관리” 폴더의 내용 설명	56
[표 5] “2. 21세기 세종계획 성과물 배포” 폴더의 내용 설명	57
[표 6] “3. 국어 연구 정보 제공 및 민원 상담” 폴더의 내용 설명	57

그림 목 차

[그림 1] 21세기 세종계획 사업 구성도	8
[그림 2] 작업 선택	32
[그림 3] 파일 인코딩	32
[그림 4] 제목 추출 및 파일명 변경	33
[그림 5] TEI-P5 파일 생성	34
[그림 6] 파일명 변경과 TEI-P5 변환	35
[그림 7] 구문에 대한 결과물	36
[그림 8] 구문 결과물 (UTF8)	37
[그림 9] 한영병렬 원시 결과물	38
[그림 10] 한영병렬 원시(UTF8)	39
[그림 11] 자료실 관리를 위한 ER 다이어그램	40
[그림 12] 성과물 배포 시스템 ER 다이어그램	41
[그림 13] 성과물 배포 시스템 초기 화면	45
[그림 14] 말뭉치 목록 화면	46
[그림 15] 말뭉치 상세 보기 화면	46
[그림 16] 자료 사용 동의 화면	47
[그림 17] 성과물 검색 결과 화면	47
[그림 18] 통계 정보 보기	48
[그림 19] 배포 시디 표지	49
[그림 20] 시디 구성	49
[그림 21] 홈페이지 사업 내용 추가	51
[그림 22] 게시판 수정 및 게시물 이전	51
[그림 23] 공지 사항	52
[그림 24] 묻고 답하기 게시판	53

제 I 장. 사업 개요

제1절. 사업의 목적

산·학·연 연구·개발자들 및 관련 기관들을 대상으로, “21세기 세종계획” 성과물의 활용성 제고 및 확산 유도를 위한 성과물의 관리 및 배포 지원을 목표로 하며, 이의 달성을 위하여 다음과 같은 세부 수행 목표를 설정하였다.

- 21세기 세종계획 성과물 관리 지원
- 21세기 세종계획 성과물 배포
- 국어 연구 정보 제공 및 민원 상담

제2절. 사업 목표

본 사업은 21세기 세종계획 성과물의 활용성 증대 및 확산 유도를 위한 성과물의 관리 및 배포 지원을 목표로 하며, 이의 달성을 위하여 요구되는 다음과 같은 단위 업무들을 수행하였다.

- 21세기 세종계획 성과물 관리 지원
- 21세기 세종계획 성과물 배포
- 국어 연구 정보 제공 및 민원 상담

제3절. 사업 내용 및 추진 경과

2007년도 당해년도의 본 사업은 아래와 같은 세부 단위 업무들을 중심으로 수

행됨.

- 21세기 세종계획 성과물 관리 지원
- 21세기 세종계획 성과물 배포
- 국어 연구 정보 제공 및 민원 상담

1. 21세기 세종계획 성과물 관리 지원

- 21세기 세종계획 성과물 관리 지원은 온라인 배포 대상 성과물 선정, 오프라인 배포 성과물 대상 선정을 통해 배포 성과물을 선정하는 작업을 포함함
 - 온라인 배포 성과물 대상 선정
 - 온라인 배포 성과물 구조 작성
 - 오프라인 배포 성과물 대상 선정
 - 오프라인 배포 성과물 구조 작성
- 말뭉치 헤더 변환
 - 말뭉치 헤더 변환을 위한 DTD 추출
 - 성과물 형식 변환 프로그램 개발
- 영문 사이트를 위한 번역

2. 21세기 세종계획 성과물 배포

- 성과물 관리 DB 개발
 - 성과물의 관리 및 배포를 위한 성과물 DB 개발
- 온라인 성과물 배포 시스템 개발
 - 배포 성과물의 메뉴 구조 설계
 - 배포 대상 성과물의 일괄 다운로드 기능 제공

- 배포 대상 성과물의 내려받기 통계 제공
- 온라인 성과물 배포 시스템 홍보
 - 2006년 사업에서 성과물을 제공 받은 사용자 명단을 이용한 이메일 홍보
- 오프라인 성과물 배포용 시디 제작
 - 개별 분과들로부터 제공받은 오프라인 배포용 최종 성과물을 시디로 제작

3. 한국어 연구 정보 제공 및 민원 상담

- 21세기 세종계획 홈페이지 운영
 - 기존에 개발된 세종계획 홈페이지 운영 및 내용 갱신
- 관련 분야 학회 및 대학을 중심으로 연구 정보 제공 관련 홍보 수행.
 - 구체적 이용 사례의 정리 및 홍보를 통한 방문자 유도
 - 국어 정보화 아카데미와 연계하여 이용자 방문 유도

4. 사업 추진 경과

2007년도 사업은 [표 1]와 같이 추진되었다.

[표 1] 2007년도 사업 추진 일정

사 업 내 용		추진 일정												비 고
		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	
성과물 관리	온라인/오프라인 배포 대상 성과물 선정													100%
	성과물 형식 검증 프로그램 개발													100%
	성과물 관리 DB 개발													100%
	성과물 배포 시스템 설계													
성과물 배포 시스템	성과물 배포 시스템 개발													
	성과물 배포 시스템 홍보													
	오프라인 배포용 CD 제작													
국어 연구 정보 제공 및 민원 상담														100%
산출물 정리 및 보고서 작성														100%

제 II 장. 21세기 세종계획 성과물 관리 지원

제1절. 성과물 배포 대상 선정

1. 오프라인 배포 대상 성과물

가. 말뭉치

- 사업 개요
- 말뭉치 파일

▷ 현대

✓ 현대 문어

- ★ 원시 말뭉치
- ★ 형태소 분석 말뭉치
- ★ 형태 의미 분석 말뭉치
- ★ 구문 분석 말뭉치
- ★ 현대 문어 말뭉치 구축 지침
- ★ 현대 문어 말뭉치 목록 ('말뭉치목록_현대문어.xls'에 링크해주시면 됩니다.)

✓ 현대 구어

- ★ 원시 말뭉치
- ★ 형태소 분석 말뭉치
- ★ 현대 구어 말뭉치 구축 지침('현대구어말뭉치_구축지침.hwp' 링크)
- ★ 현대 구어 원시 말뭉치 목록('말뭉치목록_현대구어_원시말뭉치.xls' 링크)
- ★ 현대 구어 형태소 분석 말뭉치 목록('말뭉치목록_현대구어_형태소분석말뭉치.xls' 링크)

▷ 병렬

✓ 한영 병렬

- ★ 원시 말뭉치

- ★ 형태소 분석 말뭉치
- ★ 한영 병렬 말뭉치 구축 지침(‘한영병렬말뭉치_구축지침.hwp’ 링크)
- ★ 한영 병렬 말뭉치 목록(‘말뭉치목록_한영병렬_배포용.xls’ 링크)
- ✓ 한일 병렬
 - ★ 원시 말뭉치
 - ★ 형태소 분석 말뭉치
 - ★ 한일 병렬 말뭉치 구축 지침(‘한일병렬말뭉치_구축지침.hwp’ 링크)
 - ★ 한일 병렬 말뭉치 목록(‘말뭉치목록_한일병렬_배포용.xls’ 링크)
- 활용 프로그램:
 - ▷ 현대 국어 기초 말뭉치 활용 시스템
 - ✓ 사용자 설명서
 - ✓ HanMaru_setup.exe (가능하면 ‘HanMaru_setup.exe’ 자동 구동)
 - ▷ 한영 병렬 말뭉치 용례 검색기
 - ✓ 한영 병렬 말뭉치 용례 검색기
 - ✓ 원시 말뭉치 마법사
 - ★ hep wiz.exe (가능하면 ‘hep wiz.exe’ 자동 구동)
 - ✓ 주석 말뭉치 마법사
 - ★ hep wiz mp.exe (가능하면 ‘hep wiz mp.exe’ 자동 구동)
 - ✓ 사용자 설명서
 - ▷ 지능형 형태소 분석기
 - ✓ 사용자 설명서
 - ✓ 지능형 형태소 분석기 (가능하면 ‘지능형형태소분석기.exe’ 자동 구동)

나. 전자 사전

- 사업 개요 및 사전 설명서
- 기술 목록
 - ▷ 기초 전자 사전
 - ▷ 상세 전자 사전

- 전자 사전 검색기

- ▷ 세종 기초 사전 단순 검색기 ('세종기초사전단순검색기.exe' 자동 구동)
- ▷ 세종 상세 사전 단순 검색기 ('세종상세사전단순검색기.exe' 자동 구동)
- ▷ 세종 상세 사전 상세 검색기 ('세종상세사전상세검색기.exe' 자동 구동)

다. 한민족 언어 정보화

- 사업 개요
- 한민족 언어 정보화 검색 프로그램 ('setup.exe' 자동 구동)
- 남북한 언어 비교 사전 및 방언
- 어문 규정 데이터베이스
- 어휘 역사

라. 전문 용어/문자 코드/글꼴 개발/정보화 인력 양성

- 전문 용어
 - ▷ 사업 개요
 - ▷ 전문 용어 검색기 ('UllAppSJ.exe' 자동 구동)
 - ▷ 전문 용어 대응 목록
- 문자 코드
 - ▷ 사업 개요
 - ▷ 유니코드 한자 검색기
 - ▷ 문자 코드 데이터베이스
- 글꼴 개발
 - ▷ 사업 개요
 - ▷ 옛 문헌 한글 글꼴 발굴 복원 연구
 - ✓ 언해납약증치방
 - ✓ 동국정운체
 - ✓ 옛 문헌 한글 글꼴 개발 실용화 방안 연구 일러두기 ('옛 문헌 한글 글꼴 개발 실용화 방안 연구_일러두기_2004.txt' 링크)
 - ▷ 한글 글꼴 용어 사전
 - ▷ 관련 논문
- 정보화 인력 양성

- ▷ 사업 개요
- ▷ 국어정보화 아카데미 강의록

2. 오프라인 배포 대상 성과물 안내서

가. 21세기 세종계획 사업 개요

21세기 세종계획의 최종 목표는 우리말과 우리글을 바탕으로 하는 정보 사회 건설이다.

이를 위해서

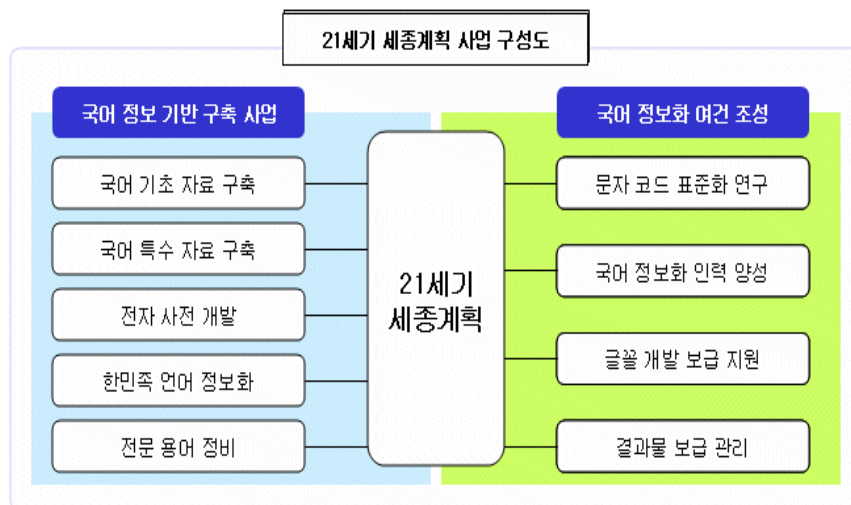
- (1) 세계 수준의 국어 기초 자료 구축을 통한 우리말 연구 기반 마련,
- (2) 표준화된 전자 사전 구축을 통한 우리말 정보화,
- (3) 한민족 언어 정보화를 통한 우리말 세계화,
- (4) 언어 정보화 관련 규격과 도구의 체계적 정리를 통한 우리말 표준화 등을 실현한다.

◆ 단계별 발전 계획

1단계 (1998~2000) : 국어 정보화 발전 단계

2단계 (2001~2003) : 국어 정보화 성숙 단계

3단계 (2004~2007) : 국어 정보화 활성화 단계



[그림 1] 21세기 세종계획 사업 구성도

<기초 자료 구축 분과>

1. 사업 목적

: 현대 한국어의 총체적 현실을 반영하는 대규모의 국가 말뭉치의 구축

2. 사업 내용

- 말뭉치 구축(원시 말뭉치, 형태 분석 말뭉치, 형태 의미 분석 말뭉치, 구문 분석 말뭉치)
- 어절 분석 표지의 표준화 연구, 구문 분석 방법론 등 말뭉치 구축 방법론 및 표준화 연구
- 현대국어 말뭉치 활용 시스템 등 기초 말뭉치 활용 도구 및 말뭉치 구축 도구 개발

3. 주요 성과

- 원시 말뭉치: 63,632,472어절
- 형태 분석 말뭉치: 15,293,877어절
- 형태 의미 분석 말뭉치: 12,691,945어절
- 구문 분석 말뭉치: 825,127어절

<특수 자료 구축 분과>

1. 사업 목적

: 시대와 지역을 망라한 총체적 한국어 연구 기반을 확보하기 위한 대규모의 국가 말뭉치의 구축

2. 사업 내용

- 말뭉치 구축(현대 국어 구어 전사 말뭉치, 한영 병렬 말뭉치, 한일 병렬 말뭉치, 북한 및 해외 한국어 말뭉치, 역사 자료 말뭉치)
- 다양한 구어적 변이형의 조사, 구어 전사 말뭉치 형태 주석 방법론 등의 기초 연구
- 다국어 병렬 말뭉치 구축을 위한 기초 연구

3. 주요 성과

- 구어 전사 말뭉치: 4,204,082어절(원시) / 1,008,669어절(형태 분석)
- 한영 병렬 말뭉치: 6,006,253어절(원시) / 1,309,330어절(형태 분석)
- 한일 병렬 말뭉치: 1,158,248어절(원시) / 297,953어절(형태 분석)
- 북한 및 해외 한국어 말뭉치: 7,708,792어절(원시)/ 1,584,976어절(형태 분석)
- 역사 말뭉치: 5,689,922어절(원시)/ 632,099어절(형태 분석)

<전자 사전 개발>

1. 사업 목적

: 한국어 기계 처리 지원을 위한 대규모 범용 전자 사전 개발

2. 사업 내용

- 전자 사전 구축 (17개 문법범주/12개 하위사전)
 - 상세전자 사전 / 기초전자 사전 / 통합전자 사전
- 전자 사전의 통합 관리 환경 완성
- 전자 사전 구축을 위한 기반 연구
- 전자 사전 구축 및 검증, 평가, 활용을 위한 각종 도구 개발

3. 주요 성과

- (1) 전자 사전 구축 : 총 60만 항목

[표 2] 전자 사전 구축 성과

	상세 전자 사전	기초 전자 사전
체언 사전	35,100	136,000
용언 사전	27,700	50,000
어근/접사 사전	1,700	1,700
의존명사/대명사/수사 사전	1,200	1,200
고유명사 사전	22,000	158,600
부사 사전	4,500	11,000
조사/어미/관형사/감탄사 사전	2,650	2,650
특수어 사전	35,000	60,000
언어 사전	9,000	9,000
관용표현 사전	5,000	5,000
복합명사구 사전	5,800	15,200
통합 전자 사전	149,650	450,350
세종 전자 사전	600,000	

- (2) 전자 사전 구축 및 검증, 평가, 활용을 위한 각종 도구 개발

- 하위사전별 입력기, 통합 입력기
- 단순 검색기, 상세 검색기(체언, 용언, 부사, 특수어, 관용표현)
- 웹 기반 관리·검색 시스템
- 세종 전자 사전 기반 형태소 및 구문 분석 시스템

<전문 용어의 정비>

1. 사업 목적

: 전문 용어의 정비를 통한 과학 기술 발전의 기반 조성

2. 사업 내용

- 전문 용어 목록 데이터베이스 구축
- 전문 용어의 국어학적 검토 및 표준안 제시
- 전문 용어 데이터베이스 서비스

3. 주요 성과

- 물리학: 15,000항목/ 화학: 15,000항목/ 생물학: 15,000항목/ 의학: 15,000항목
- 수학: 10,000항목/ 전산학: 30,000항목/ 전자전기공학: 15,000항목
- 기계공학: 15,000항목/ 산업공학: 5,000항목/ 화학공학: 15,000항목
- 재료공학: 5,000항목/ 환경공학: 5,000항목/ 건축공학: 5,000항목/ 토목공학: 5,000항목

<한민족 언어 정보화>

1. 사업 목적

: 한국어 공동체의 언어 이질감을 해소하고 국민 언어생활에 도움이 되는 자료 구축

2. 사업 내용

- 어문 규정 검색 시스템 개발
- 남북한 언어 비교 사전 구축
- 방언 검색 시스템 개발
- 어휘 역사 검색 시스템 개발

3. 주요 성과

- 어문 규정 검색 자료: 54,245항목
- 남북한 언어 비교 사전: 13,000항목
- 한국 방언 자료: 38,011항목
- 어휘 역사 자료: 5,236항목
- 한국 전통 문화 어휘 검색: 322항목
- 문학작품에 사용된 방언 검색: 2,186항목

<문자 코드 표준화 연구>

1. 사업 목적

: 비표준 문자의 표준화 및 국제적 표준 코드의 활용에 필요한 자원 개발

2. 사업 내용

- 비표준문자 수집·정리 및 국제 표준화 지원
- 문자코드 표준화 연구
 - 한자 및 옛한글 정보 데이터베이스 구축
 - 비표준문자 처리에 필요한 각종 자형 및 응용 프로그램 개발 지원
 - 비표준문자의 국제 표준화 방안 연구
- 유니코드 한자의 효율적 활용을 위한 연구
- 옛한글의 효율적 활용을 위한 연구
- 문자코드 정보 제공 및 지원체계 확립

3. 주요 성과

- 비표준 문자 수집
 - 옛한글 : 7,112,856자
 - 신출한자 : 21,659자
- <옛한글 정보 데이터베이스> 구축
- <Super CJK 한자 정보데이터베이스> 구축
- <유니코드 한자 검색기> 개발
- CJK 통합 한자 Ext.C에 한국측 제안 한자 404자 등록
- <옛 한글 자모 확장 목록> 117자 유니코드 BMP 영역 등록
- CJK 통합 한자 <Ext.D 한국측 제안 한자 목록> 2,000여 자 제출
- 심포지엄 4회, 워크숍 3회 개최
- 소식지 발간 : 연 2회 20호 발간

<국어 정보화 인력 양성>

1. 사업 목적

: 인력 양성 프로그램 운영을 통한 국어 정보화 저변 확대

2. 사업 내용

- 국어 정보화 관련 교재 개발
- 국어 정보화 아카데미 행사 개최(7회)

3. 주요 성과

- 국어 정보화 아카데미 7회 개최
- “국어정보학 개설” 편찬
- “정보학 용어 길라잡이” 편찬
- “전자 말뭉치를 이용한 사전편찬론 (Theories of Lexicography)”

<글꼴 개발 보급>

1. 사업 목적

: 다양한 한글 글꼴 개발의 촉진 및 보급의 활성화

2. 사업 내용

- 글꼴 공모전을 통한 글꼴 개발의 촉진
 - 글꼴의 조사, 연구
 - 글꼴의 유통, 보급
 - 글꼴의 교육, 홍보
- 『글꼴』 발간을 통한 글꼴 보급의 활성화
 - 한글 글꼴 개발 현황 보고
 - 글꼴 연구 논문 게재
 - 글꼴 관련 자료 소개

3. 주요 성과

- 옛 문헌 글꼴 발굴 복원 연구 - 동국정운체, 언해납약증치방체 개발
- 매년 ‘한글 글꼴 공모전’ 개최
- 『글꼴』 9회(1998~2006) 발간
- 『한글글꼴용어사전』(2000) 발간
- 디지털 서체 개발 보급 - 현대 한국 대표 서예가 서체 등

나. 21세기 세종계획 최종 성과물 자료 내용

<말뭉치>

사업 개요

말뭉치 파일: 현대 문어/현대 구어, 한영 병렬/한일 병렬, 역사

활용 프로그램: 현대국어 기초 말뭉치 활용 시스템(한마루), 지능형
형태소 분석기, 한영 병렬 말뭉치 용례 검색기(hep
man)

<전자 사전>

사업 개요 및 사전 설명서

기술 목록

전자 사전 검색기

<한민족 언어 정보화>

사업 개요

한민족 언어 정보화 검색 프로그램
남북한 언어 비교 사전 및 방언
어문 규정 데이터베이스
어휘 역사

<전문 용어/문자 코드/글꼴 개발/정보화 인력 양성>

전문 용어: 사업 개요

전문 용어 검색기

전문 용어 대응 목록

문자 코드: 사업 개요

유니코드 한자 검색기

문자 코드 데이터베이스

글꼴 개발: 사업 개요

옛 문헌 한글 글꼴 발굴 복원 연구(동국정운체, 언해납약

중치방체)

한글 글꼴 용어 사전

관련 논문

정보화 인력 양성: 사업 개요

국어정보화 아카데미 강의록

제2절. 성과물 헤더 변환

1. TEI-P3 말뭉치 형판

현재 성과물들은 TEI-P3의 헤더 형식의 변형된 형판을 가지고 있다. 따라서 헤더 변환에서는 구어와 문어, 한영(일)병렬의 3가지 형판을 우선 살펴본다.

가. 구어 원시(형태소 분석) 말뭉치 형판

```
<!DOCTYPE tei.2 SYSTEM "c:\sgml\dtd\tei2.dtd" [  
  <!ENTITY % TEI.corpus "INCLUDE">  
  <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
```

```

    <!ENTITY % TEL.extensions.dtd SYSTEM "sejong1.dtd">
]>

<tei.2>

<teiHeader>
<fileDesc> (전자 서지 사항)
  <titleStmt>
    <title> 【자료제목】 </title>
    <author> 【저자】 </author>
    <sponsor> 【스폰서】 </sponsor>
    → <sponsor>대한민국 문화관광부</sponsor>로 고정, ‘국립국어
연구원, 국립국어원’ 등으로 되어 있는 것은 오류이므로 무조건 변환
    <respStmt><resp> 【자료구축내용】 </resp>
      <name> 【기관명】 </name>
    </respStmt>
  </titleStmt>
  <extent> 【어절수】 </extent>
  <publicationStmt>
    <distributor> 【배포 책임 기관】 </distributor>
    → <distributor>국립국어원</distributor>로 고정, ‘국립국어연구
원’ 등으로 되어 있는 것은 오류이므로 무조건 변환
    <idno> 【파일명】 </idno>
    <availability><p> 【사용·배포 제약】 </p></availability>
  </publicationStmt>
  <sourceDesc>
    <p> 【원전정보】 </p>
  </sourceDesc>
</fileDesc>
<encodingDesc> (전자화 관련 방법)
  <projectDesc><p> 【구축 목적】 </p></projectDesc>
  <samplingDecl><p> 【구축방법】 </p></samplingDecl>
  <editorialDecl><p> 【편집 원리와 실행 원칙】 </p></editorialDecl>
</encodingDesc>
<profileDesc> (텍스트 유형 정보)

```

```

    <creation><date> 【텍스트생성일】 </date></creation>
    <langUsage>
        <language id=KO usage=99> 【텍스트에 나타나는 언어】
    </language>
    </langUsage>
    <particDesc>
        +<person id= 【발화자번호】 sex= 【성별(M/F)】 age= 【연
령】 ><p> 【대화참여자】 </p></person>
        <person id= 【발화자번호】 sex= 【성별(M/F)】 age= 【연
령】 ><p> 【대화참여자】 </p></person>
        <person id= 【발화자번호】 sex= 【성별(M/F)】 age= 【연
령】 ><p> 【대화참여자】 </p></person>
    </particDesc>
    <settingDesc>
        <p> 【대화장면정보】 </p>
    </settingDesc>
    <textClass>
        <catRef scheme='SJ21' target=' 【분류코드】 '> 【텍스트종류】
    </catRef>
    </textClass>
    </profileDesc>
    <revisionDesc> (입력·수정 과정 정보)
        <change><date> 【전자텍스트구축및수정일자】 </date>
            <respStmt><resp> 【지적책임성격】 </resp><name> 【구축및
수정자】 </name></respStmt>
            <item> 【구축및수정작업내용】 </item>
        </change>
    </revisionDesc>
    </teiHeader>

    <text>(본문)
        <u who= 【발화자번호】 ><s n= 【발화일련번호】 >… </s></u>
    </text>

    </tei.2>

```


나. 문어 원시(형태소 분석) 말뭉치 형판

헤더 형식 변환은 기초 자료 구축 분과 말뭉치 전체(원시, 형태 분석, 형태 의미 분석, 구문 분석), 특수 자료 구축 분과의 역사 자료 말뭉치, 북한 말뭉치 전체(원시, 형태 분석)에 적용된다.

```
<!DOCTYPE tei.2 SYSTEM "c:\sgml\dtd\tei2.dtd" [  
    <!ENTITY % TEL.corpus "INCLUDE">  
    <!ENTITY % TEL.extensions.ent SYSTEM "sejong1.ent">  
    <!ENTITY % TEL.extensions.dtd SYSTEM "sejong1.dtd">  
>
```

```
<tei.2>
```

```
<teiHeader>  
<fileDesc> (전자 서지 사항)  
    <titleStmt>  
        <title> 【자료제목】 </title>  
        <author> 【저자】 </author>  
        <sponsor> 【스폰서】 </sponsor>  
        <respStmt><resp> 【자료구축내용】 </resp>  
            <name> 【기관명】 </name>  
        </respStmt>  
    </titleStmt>  
    <extent> 【어절수】 </extent>  
    <publicationStmt>  
        <distributor> 【배포 책임 기관】 </distributor>  
        <idno> 【파일명】 </idno>  
        <availability><p> 【사용·배포 제약】 </p></availability>  
    </publicationStmt>  
    *<notesStmt>  
        <note><p> 【기타주석】 </p></note>  
    </notesStmt>  
<sourceDesc>
```

```

        <bibl><author> 【원전저자】 </author>
            <title> 【원전제목】 </title>
            <pubPlace> 【출판지】 </pubPlace>
            <publisher> 【출판사】 </publisher>
            <date> 【출판일】 </date>
        </bibl>
    </sourceDesc>
</fileDesc>
<encodingDesc> (전자화 관련 방법)
    <projectDesc><p> 【구축 목적】 </p>
    </projectDesc>
    <samplingDecl><p> 【구축방법】 </p>
    </samplingDecl>
    <editorialDecl><p> 【편집 원리와 실행 원칙】 </p>
    </editorialDecl>
</encodingDesc>
<profileDesc> (텍스트 유형 정보)
    <creation><date> 【텍스트생성일】 </date></creation>
    <langUsage>
        <language id=KO usage=99> 【텍스트에 나타나는 언어】
    </language>
    </langUsage>
    <textClass>
        <catRef scheme='SJ21' target=' 【분류코드】 ' > 【텍스트종류】
    </catRef>
    </textClass>
</profileDesc>
<revisionDesc> (입력·수정 과정 정보)
    *<respStmt><resp> 【지적책임성격】 </resp><name> 【책임자】
</name></respStmt>
    +<change><date> 【전자텍스트구축및수정일자】 </date>
        <respStmt><resp> 【지적책임성격】 </resp><name> 【구축
및수정자】 </name></respStmt>
        <item> 【구축및수정작업내용】 </item>
    </change>

```

```

</revisionDesc>
</teiHeader>

<text> (본문)
<body>
  <p> ... </p>
</body>
</text>

</tei.2>

```

다. 한영(일)병렬 원시(형태소 분석) 말뭉치 형판

```

<cesDoc version=3.24>
<cesHeader version="2.3">
  <fileDesc>
    <titleStmt>
      <h.title> 【자료제목】 </h.title>
      <respStmt>
        <respType> 【자료구축내용】 </respType>
        <respName> 【기관명】 </respName>
      </respStmt>
    </titleStmt>
    <extent>
      <wordcount> 【어절수】 </wordcount>
    </extent>
    <publicationStmt>
      <distributor> 【배포 책임 기관】 </distributor>
      <pubAddress> 【주소지】 </pubAddress>
      <availability status=restricted> 【사용·배포 제약】 </availability>
      <PubDate> 【발간일】 </PubDate>
    </publicationStmt>
    <sourceDesc>
      <biblStruct>
        <monogr>

```

```

        <h.title> 【원전제목】 </h.title>
        <h.author> 【원전저자】 </h.author>
        <imprint>
            <pubPlace> 【출판지】 </pubPlace>
            <publisher> 【출판사】 </publisher>
            <pubDate> 【출판일】 </pubDate>
        </imprint>
    </monogr>
</biblStruct>
</sourceDesc>
</fileDesc>
<encodingDesc>
    <projectDesc> 【구축 목적】 </projectDesc>
    <editorialDecl> 【편집 원리와 실행 원칙】 </editorialDecl>
    <samplingDecl> 【구축방법】 </samplingDecl>
    <refsDecl> </refsDecl>
</encodingDesc>
<profileDesc>
    <creation><date> 【텍스트생성일】 </date></creation>
    <langUsage>
        <language id="en" iso639="en"> 【텍스트에 나타나는 언어】
    </language>
    </langUsage>
    <wsdUsage>
        <writingSystem id="iso8859-1"></writingSystem>
    </wsdUsage>
    <textClass>
        <catRef scheme="SJ21" target=" 【분류코드】 "> 【텍스트종류】
    </catRef>
    </textClass>
    <translations>
        <translation trans.loc="rb0001ek.hwp" lang=kr
wsd="iso2022-kr">
        <translator></translator>
    </translations>

```

```

</profileDesc>
<revisionDesc>
  <change>
    <changeDate> 【전자텍스트구축및수정일자】 <changeDate>
    <respName> 【책임자】 </respName>
    <h.item> 【구축및수정작업내용】 </h.item>
  </change>
</revisionDesc>
</cesHeader>
<text>
</text>
</cesDoc>

```

2. TEI-P5 말뭉치 형판

현재 성과물들은 TEI-P3의 헤더 형식의 변형된 형판을 가지고 있다. 따라서 헤더 변환에서는 구어와 문어, 한영(일)병렬의 3가지 형판을 우선 살펴본다.

가. 구어 원시(형태소 분석) 말뭉치 형판

```

<TEI>
  <teiHeader>
    <fileDesc> (전자 서지 사항)
      <titleStmt>
        <title> 【자료제목】 </title>
        <author> 【저자】 </author>
        <sponsor> 【스폰서】 </sponsor>
        → <sponsor>대한민국 문화관광부</sponsor>로 고정, ‘국립국어
        연구원, 국립국어원’ 등으로 되어 있는 것은 오류이므로 무조건 변환
        <respStmt><resp> 【자료구축내용】 </resp>
          <name> 【기관명】 </name>
        </respStmt>
      </titleStmt>
    <extent> 【어절수】 </extent>

```

```

<publicationStmnt>
  <distributor> 【배포 책임 기관】 </distributor>
  → <distributor>국립국어원</distributor>로 고정, ‘국립국어연구
원’ 등으로 되어 있는 것은 오류이므로 무조건 변환
  <idno type="SJ file"> 【파일명】 </idno>
  <availability><p> 【사용·배포 제약】 </p></availability>
</publicationStmnt>
<sourceDesc>
  <p> 【원전정보】 </p>
</sourceDesc>
</fileDesc>
<encodingDesc> (전자화 관련 방법)
  <projectDesc><p> 【구축 목적】 </p></projectDesc>
  <samplingDecl><p> 【구축방법】 </p></samplingDecl>
  <editorialDecl><p> 【편집 원리와 실행 원칙】 </p></editorialDecl>
</encodingDesc>
<profileDesc> (텍스트 유형 정보)
  <creation><date> 【텍스트생성일】 </date></creation>
  <langUsage>
    <language id="KO" usage="99"> 【텍스트에 나타나는 언어】
  </language>
  </langUsage>
<particDesc>
  <listPerson>
    +<person xml:id=" 【발화자번호】 " sex=" 【성별(M/F)】 " age=" 【연
령】 "><p> 【대화참여자】 </p>
    </person>
    <person xml:id=" 【발화자번호】 " sex=" 【성별(M/F)】 " age=" 【연
령】 "><p> 【대화참여자】 </p>
    </person>
    <person xml:id=" 【발화자번호】 " sex=" 【성별(M/F)】 " age=" 【연
령】 "><p> 【대화참여자】 </p>
    </person>
  </listPerson>
</particDesc>

```

```

<settingDesc>
    <p> 【대화장면정보】 </p>
</settingDesc>
<textClass>
    <catRef scheme="SJ21" target=" 【분류코드】 "> 【텍스트종
류】 </catRef>
</textClass>
</profileDesc>
<revisionDesc> (입력·수정 과정 정보)
    <list>
        +<item>
            <date when=" 【전자텍스트구축및수정일자】 "> 【전자텍스트구축및
수정일자】 </date> 【구축및수정작업내용】 - 【지적책임성격】 - 【구축
및수정자】 </item>
        </list>
    </revisionDesc>
</teiHeader>

<contents> (본문)
<!-- 본문은 주석처리
    <text>
        <u who=" 【발화자번호】 "><s n=" 【발화일련번호】 ">... </s></u>
    </text>
-->
</contents>
</TEI>

```

나. 문어 원시 말뭉치 형판

```

<TEI>
    <teiHeader>
        <fileDesc> (전자 서지 사항)
            <titleStmt>
                <title> 【자료제목】 </title>
                <author> 【저자】 </author>
            </titleStmt>
        </fileDesc>
    </teiHeader>

```

```

    <sponsor> 【스폰서】 </sponsor>
    <respStmt><resp> 【자료구축내용】 </resp>
        <name> 【기관명】 </name>
    </respStmt>
</titleStmt>
<extent> 【어절수】 </extent>
<publicationStmt>
    <distributor> 【배포 책임 기관】 </distributor>
    <idno type="SJ file"> 【파일명】 </idno>
    <availability><p> 【사용·배포 제약】 </p></availability>
</publicationStmt>
*<notesStmt>
    <note><p> 【기타주석】 </p></note>
</notesStmt>
<sourceDesc>
    <bibl><author> 【원전저자】 </author>
        <title> 【원전제목】 </title>
        <pubPlace> 【출판지】 </pubPlace>
        <publisher> 【출판사】 </publisher>
        <date when=" 【출판일】 "> 【출판일】 </date>
    </bibl>
</sourceDesc>
</fileDesc>
<encodingDesc> (전자화 관련 방법)
    <projectDesc><p> 【구축 목적】 </p>
    </projectDesc>
    <samplingDecl><p> 【구축방법】 </p>
    </samplingDecl>
    <editorialDecl><p> 【편집 원리와 실행 원칙】 </p>
    </editorialDecl>
</encodingDesc>
<profileDesc> (텍스트 유형 정보)
    <creation><date> 【텍스트생성일】 </date></creation>
    <langUsage>
        <language id="KO" usage="99"> 【텍스트에 나타나는 언어】

```



```

</language>
  </langUsage>
  <textClass>
    <catRef scheme="SJ21" target="【분류코드】">【텍스트종
류】</catRef>
  </textClass>
</profileDesc>
<revisionDesc> (입력·수정 과정 정보)
  *<respStmt><resp>【지적책임성격】</resp><name>【책임자】
</name></respStmt>
<list>
  +<item>
    <date when="【전자텍스트구축및수정일자】">【전자텍스트구축및
수정일자】</date> 【구축및수정작업내용】 - 【지적책임성격】 - 【구축
및수정자】</item>
</list>
</revisionDesc>
</teiHeader>

<contents> (본문)
<!-- 본문은 주석처리
  <text>
    <u who="【발화자번호】"><s n="【발화일련번호】">... </s></u>
  </text>
-->
</contents>

</TEI>

```

다. 병렬 말뭉치 형판

```

<TEI>
  <teiHeader>
    <fileDesc>(전자 서지 사항)
      <titleStmt>

```

```

        <title>[자료 제목]</title> //<--<h.title>[자료 제목]</h.title>
    <extent> 【어절수】 </extent>
    </titleStmt>
    <respStmt><resp> 【자료구축내용】 </resp>
        <name> 【기관명】 </name>
    </respStmt>
    <publicationStmt>
        <distributor>국립국어원</distributor>
    <idno> 【파일명】 </idno>
        <pubPlace> 【출판지】 </pubPlace>
    //<--<pubAddress>[출판지]</pubAddress>
        <availability>[상용·배포 제약]</availability> //<-- <availability
status=restricted>배포 불가</availability>
        <date>[출판일]</date> // <--
    <PubDate>[출판일]</PubDate>
    </publicationStmt>
    <notesStmt>
        <note><p> 【기타주석】 </p></note>
    </notesStmt> // 기존 버전에서는 닫는 태그가 없었네요.
    <sourceDesc>
        <bibl>//<--<biblStruct>
            //<monogr> --> 제거
            <author>[원전 저자]</author> // <--<h.title>[원전
제목]</h.title>
            <title>[원전 제목]</title> // <--<h.author>[원전 저자]
</h.author>
            //<imprint> --> 제거
            <pubPlace>[출판지]</pubPlace>
            <publisher>[출판사]</publisher>
            <date when=" 【출판일】">[출판일]</date> //
    <--<pubDate>[출판일]</pubDate>
            // </imprint> --> 제거
            // </monogr> --> 제거
        </bibl> //<--</biblStruct>
    </sourceDesc>
    </fileDesc>
    <encodingDesc> (전자화 관련 방법)

```

```

    <projectDesc><p> 【구축 목적】 </p> </projectDesc>
    <samplingDecl><p> 【구축방법】 </p> </samplingDecl>
    <editorialDecl><p> 【편집 원리와 실행 원칙】 </p> </editorialDecl>
    <refsDecl> 【설명】 </refsDecl> 제거
</encodingDesc>
    <profileDesc> (텍스트 유형 정보)
        <creation><date> 【텍스트생성일】 </date></creation>
        <langUsage>
            <language id="KO" usage="99"> 【텍스트에 나타나는 언어】
</language>
        </langUsage>
        <!-- <wsdUsage> --> 제거
            <writingSystem id="iso8859-1"></writingSystem>
        </wsdUsage> --> 제거 -->
        <textClass>
            <catRef scheme="SJ21" target=" 【분류코드】 "> 【텍스트종류】
</catRef>
        </textClass>
        <translations>
            <translation trans.loc="파일명" lang="kr" wsd="iso2022-kr"/>
            <translator></translator>
        </translations>
    </profileDesc>
    <revisionDesc> (입력·수정 과정 정보)
        <!-- <change>
            <changeDate>[수정일]</changeDate>
            <respName>[책임자] </respName>
            <respName>[책임자] </respName>
            <hitem>본문 마크업, POS 태깅</hitem>
        </change>
==> 아래의<change when="[수정일]"> </change>와 같이 수정
-- >
        <change when="[수정일]">
            <list type="ordered">
<respStmt>

```

```

<resp> 【지적책임성격】 </resp>
<Name> 【책임자】 </Name>
</respStmnt>
// 나열일 때
    <item> 【구축및수정작업내용】 -    【지적책임성격】    -    【구축및수정자】
</item>
</list>

    </change>
</revisionDesc>
</teiHeader> //<--</cesHeader>

```

3. 헤더 변환 방법

본 변환은 아래의 4가지 단계를 거친다.

가. 파일 변환

현재 성과물은 한글 파일 형태를 가진다. 하지만 변환 작업을 수행하기 위해서는 텍스트 파일로의 변환이 필요하다. HwpConv라는 프로그램을 통해 유니코드 형태의 텍스트 파일로 변환한 경우에는 파일 변환 작업을 거치지 않아도 되나, 유니코드 형태가 아닌 파일의 경우에는 유니코드로 인코딩을 변환한다.

나. 전처리 작업

이 작업은 입력으로 들어오는 TEI-P3 파일을 Well-formed XML형태로 변환하는 전처리 과정이다. 로컬 패스를 가지는 DTD 선언 부분을 <?xml version="1.0" encoding="UTF-8"?>와 같은 XML 선언으로 바꾼다. 그리고 어트리뷰트의 경우는 ‘나 “를 이용하여 나타내야하므로 attribute의 값이 선언된 부분은 ”“ 기호를 붙여주었다. 그리고 현재 변환의 대상이 되는 태그는 헤더 부분이므로 실제 body에 해당되는 부분은 <contents>태그를 이용하여 구분하였으며 body내의 태그는 변환에 처리되지 않게 하기 위해서 <!-- 주석처리 -->를 하였다.

다. 파일명 추출 및 파일 변환

파일명 추출을 위해서는 전처리를 거친 TEI-P3 XML파일을 파싱하여 데이터를 추출한다. 현재 파일명 형식은 아래와 같다.

주석단계_매체_자료명_출판연도.확장자

* 주석단계 : 말뭉치별 폴더 이름 앞부분으로 분류된다.

원시

한일병렬 - 원시 - 형태소 분석

한일병렬과 같이 서브 폴더를 갖는 경우에는 “한일병렬.원시”와 같은 형태를 가진다.

* 매체 : teiHeader안의 textClass 속성 내의 분류 코드 중 세 번째 자리 코드로부터 추출한다.

아래와 같이 teiHeader안의 textClass 엘리먼트의 자식 노드인 catRef의 어트리뷰트 값의 3번째 코드값을 가져온다.

```
<teiHeader>
  <textClass>
    <catRef scheme='SJ21' target='M2103'></catRef>
  </textClass>
</teiHeader>
```

* 자료명 : teiHeader 내의 <title> 정보를 이용한다.

```
<teiHeader>
<fileDesc>
  <titleStmt>
    <title> 조선일보 90년 좌담기사</title>
```

* 출판연도 : <SourceDesc><date>영역 내의 연도만 추출하여 이용한다.

```
<source>
  <date>1990/01/01</date>
  <page>07</page>
</source>
```

위와 같이 XML에서 해당되는 데이터를 추출한 후, 현재 파일을 rename한다.

라. 태그 변환

태그 변환의 경우, 전처리를 거친 TEI-P3 XML을 이용한다. 현재 태그 변환은 JDOM 파서를 이용하여 구현되었다. 구어와 문어의 경우에는 변환 대상이 되는 태그가 중복되는 것이 많고 구어가 몇 가지 변환 대상 태그를 더 갖고 있어서 두개를 하나의 메소드로 구현하였고, 변환 대상이 되는 태그가 많이 차이가 나는 한영(일) 병렬의 경우에는 또 다른 메소드로 구현하였다. 그래서 처음에 변환 작업을 하기 전에 사용자에게 한영(일) 병렬과 나머지 변환의 구분을 따로 입력받는 과정을 거친다.

<sponsor>과 <distributor>의 경우, 태그를 갖고 있지 않으면 태그를 삽입하고 태그 값은 대한민국 문화관광부와 국립국어원으로 변환한다.

```
<tei.2> ⇨ <TEI>
<indo> ⇨ <indo type="SJ file">
```

와 같이 단순히 태그명을 바꾸거나 attribute를 추가하는 변환 유형이 있으며, 아래와 같이 새로운 태그를 추가하기도 한다.

```
<particDesc>
  +<person id= 【발화자번호】 sex= 【성별(M/F)】 age= 【연령】><p> 【대화참여자】
</particDesc>
  ↓
<particDesc>
  <listPerson>
    +<person xml:id=" 【발화자번호】 " sex=" 【성별(M/F)】 " age=" 【연령】
"><p> 【대화참여자】 </p>
```

```

</listPerson>
</particDesc>

```

그리고 아래와 같이 태그를 삭제하거나 구조를 완전히 변환하기도 한다.

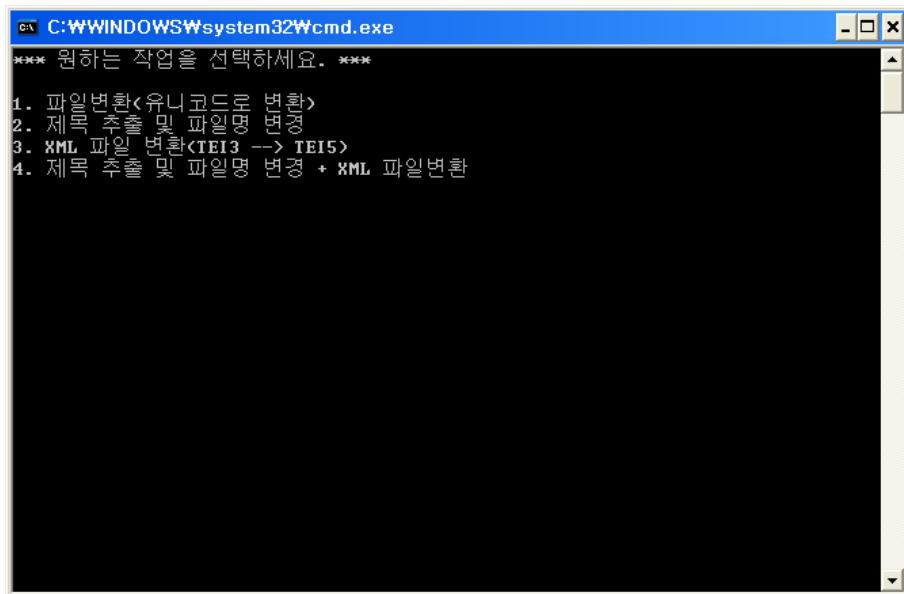
```

<revisionDesc>
  *<respStmt><resp> 【지적책임성격】 </resp><name> 【책임자】
</name></respStmt>
  +<change><date> 【전자텍스트구축및수정일자】 </date>
    <respStmt><resp> 【지적책임성격】 </resp><name> 【구축
및수정자】 </name></respStmt>
    <item> 【구축및수정작업내용】 </item>
  </change>
</revisionDesc>
↓
<revisionDesc>
  <list>
    +<item>
      <date when=" 【전자텍스트구축및수정일자】 "> 【전자텍스트구축및
수정일자】 </date> 【구축및수정작업내용】 - 【지적책임성격】 - 【구축
및수정자】 </item>
    </list>
  </revisionDesc>

```

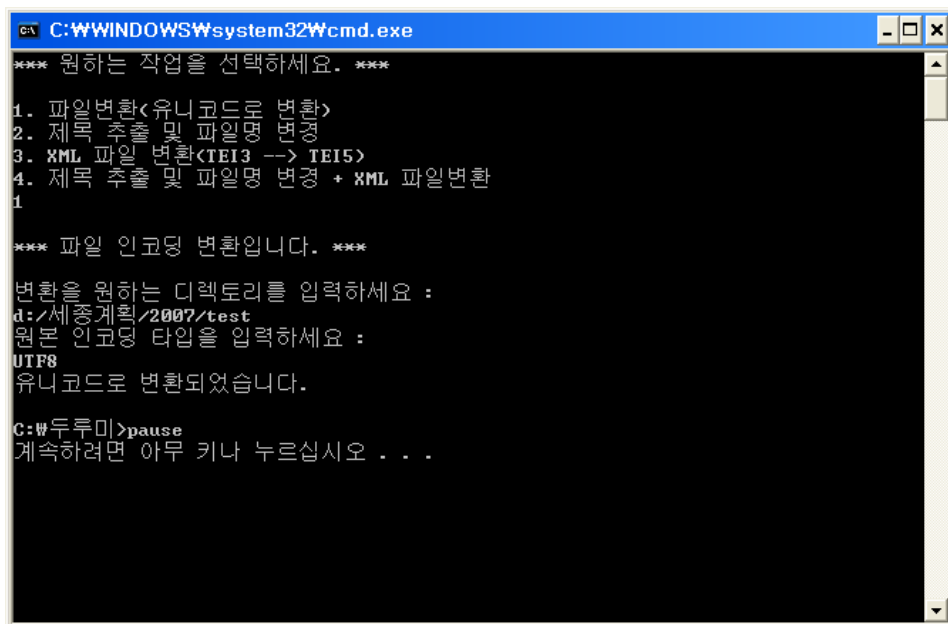
4. 도구 사용법

JRE를 설치한 후, isrl.bat를 실행하면 된다. 배치 파일을 실행하면 아래와 같
은 화면이 나타난다.



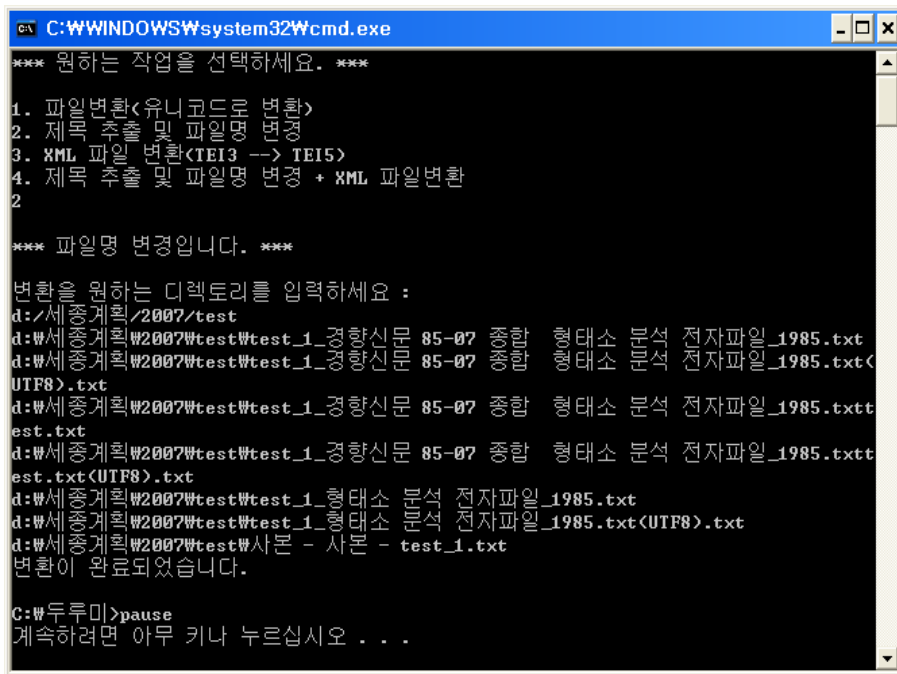
[그림 2] 작업 선택

파일 인코딩만 변환을 원할 경우에는 1번을 선택한다.



[그림 3] 파일 인코딩

변환을 원하는 디렉토리를 입력하고 원본 파일의 인코딩 타입을 입력하면 폴더내의 모든 파일의 인코딩을 유니코드로 변환한다. TEI P3->P5로 파일 변환은 하지 않고 파일명만 바꿀 경우에는 2번 제목 추출 및 파일명 변경 메뉴를 선택한다. 앞에서 설명한 파일 제목 형식에 따라 새로운 파일명으로 이름이 변환된다.



[그림 4] 제목 추출 및 파일명 변경

3번의 경우에는 파일명은 그대로 유지한 채 변환하여 새로운 TEI-P5 파일을 생성하는 것이다. 이 메뉴의 경우에는 변환을 원하는 성과물이 한영(일)병렬인지 나머지 파일인지를 체크해줘야 변환 규칙이 구분된다. 4번의 경우에는 파일명 변경과 함께 XML파일을 TEI-P3에서 TEI-P5로 변환하는 과정을 함께 거친다.

```
C:\WINDOWS\system32\cmd.exe
*** 원하는 작업을 선택하세요. ***
1. 파일변환<유니코드로 변환>
2. 제목 추출 및 파일명 변경
3. XML 파일 변환<TEI3 --> TEI5>
4. 제목 추출 및 파일명 변경 + XML 파일변환
3

*** XML 파일 변환입니다. ***

변환을 원하는 디렉토리를 입력하세요 :
d:/세종계획/2007/test

** 변환을 원하는 분과의 성과물을 선택하세요..
1. 한영<일>병렬
2. 나머지 변환
2

d:\세종계획\2007\test\구어원시1.txt
d:\세종계획\2007\test\구어원시2.txt
d:\세종계획\2007\test\구어원시3.txt
변환이 완료되었습니다.

C:\>pause
계속하려면 아무 키나 누르십시오 . . .
```

[그림 5] TEI-P5 파일 생성

```

C:\WINDOWS\system32\cmd.exe
*** 원하는 작업을 선택하세요. ***

1. 파일 변환<유니코드로 변환>
2. 제목 추출 및 파일명 변경
3. XML 파일 변환<TEI3 --> TEI5>
4. 제목 추출 및 파일명 변경 + XML 파일 변환
4

*** 파일명 변경 & XML 파일 변환입니다. ***
변환을 원하는 디렉토리를 입력하세요 :
d:/ISRL Documents-1/[수탁사업.2007]21세기세계계획/변환작업/한영병렬원시
변환을 원하는 문과의 결과물을 선택하세요..
1. 한영<일>병렬
2. 나머지 변환
1
D:\ISRL Documents-1\수탁사업.2007\21세기세계계획\변환작업\한영병렬원시\Wr-b0001a
l.txt
D:\ISRL Documents-1\수탁사업.2007\21세기세계계획\변환작업\한영병렬원시\Wr-b0001e
e.txt
D:\ISRL Documents-1\수탁사업.2007\21세기세계계획\변환작업\한영병렬원시\Wr-b0001e
k.txt
D:\ISRL Documents-1\수탁사업.2007\21세기세계계획\변환작업\한영병렬원시\Wr-b0113a
l.txt
D:\ISRL Documents-1\수탁사업.2007\21세기세계계획\변환작업\한영병렬원시\Wr-b0113e
e.txt
-

```

[그림 6] 파일명 변경과 TEI-P5 변환

5. 헤더 변환 결과

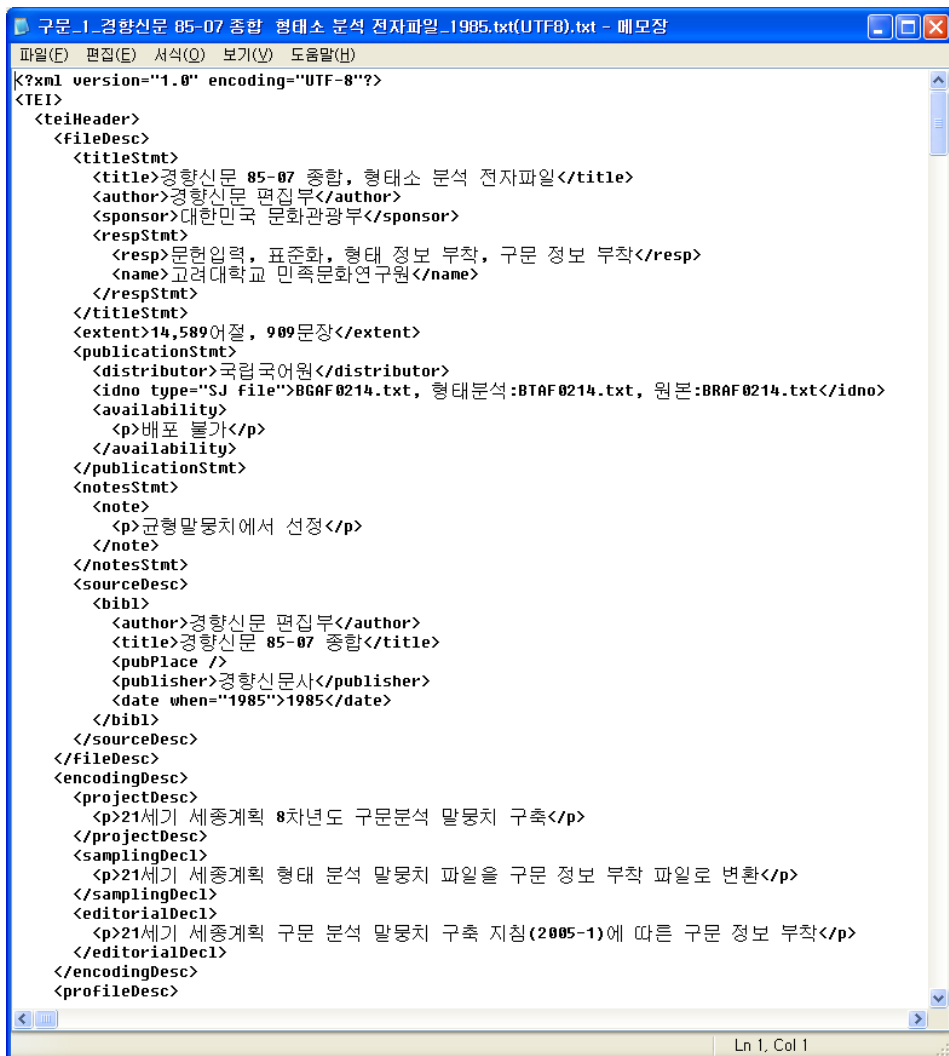
아래 그림은 구문에 대한 결과물과 한영병렬에 대한 원본과 변환 후 결과물을 보여준다.

```

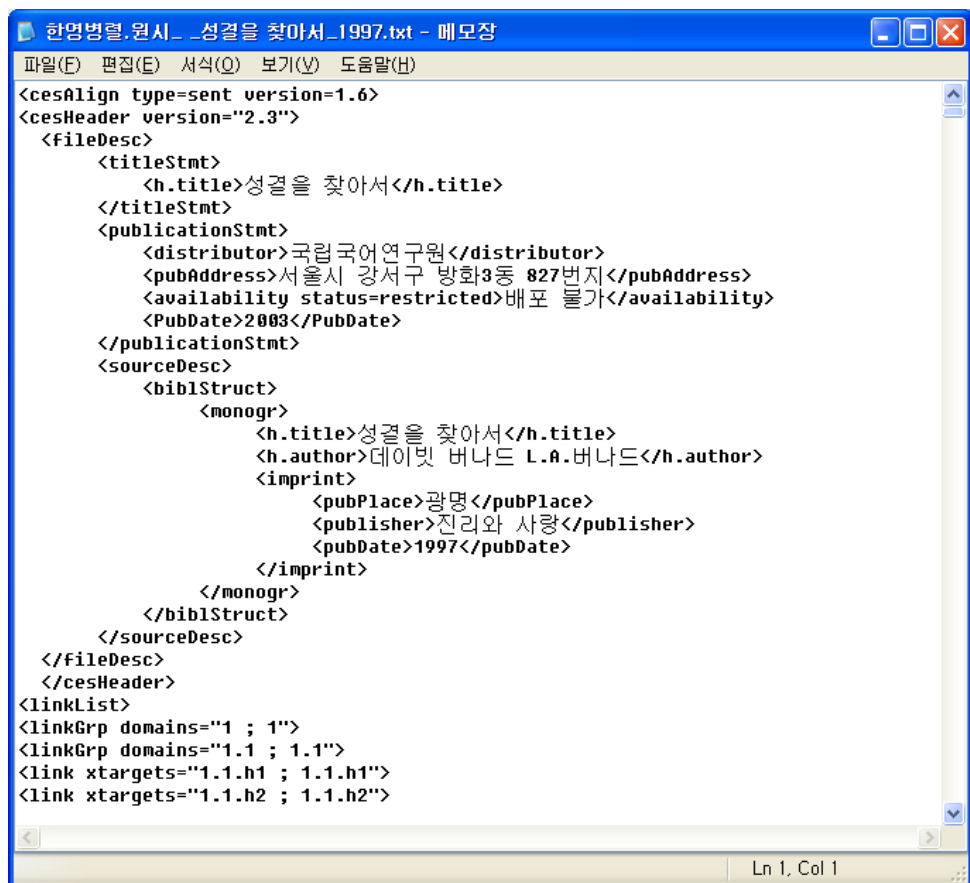
<!DOCTYPE tei.2 SYSTEM "c:\wsgml\wtdtd\tei2.dtd" [
<!ENTITY % TEI.corpus "INCLUDE">
<!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
<!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd"> ]>
<tei.2>
<teiHeader>
<fileDesc>
<titleStmt>
<title>경향신문 85-07 종합, 형태소 분석 전자파일</title>
<author>경향신문 편집부</author>
<sponsor>국립국어원</sponsor>
<respStmt>
<resp>문헌입력, 표준화, 형태 정보 부착, 구문 정보 부착</resp>
<name>고려대학교 민족문화연구원</name>
</respStmt>
</titleStmt>
<extent>14,589어절, 909문장</extent>
<publicationStmt>
<distributor>국립국어원</distributor>
<idno>BGA0052.txt, 형태분석:BTAF0052.txt, 원본:BR0052.txt</idno>
<availability>
<p>배포 불가</p>
</availability>
</publicationStmt>
<notesStmt>
<note>
<p>균형말뭉치에서 선정</p>
</note>
</notesStmt>
<sourceDesc>
<bibl>
<author>경향신문 편집부</author>
<title>경향신문 85-07 종합</title>
<pubPlace></pubPlace>
<publisher>경향신문사</publisher>
<date>1985</date>
</bibl>
</sourceDesc>
</fileDesc>
<encodingDesc>
<projectDesc>
<p>21세기 세종계획 8차년도 구문분석 말뭉치 구축</p>
</projectDesc>
<samplingDecl>
<p>21세기 세종계획 형태 분석 말뭉치 파일을 구문 정보 부착 파일로 변환</p>
</samplingDecl>
<editorialDecl>
<p>21세기 세종계획 구문 분석 말뭉치 구축 지침(2005-1)에 따른 구문 정보 부착</p>

```

[그림 7] 구문에 대한 결과물



[그림 8] 구문 결과물 (UTF8)

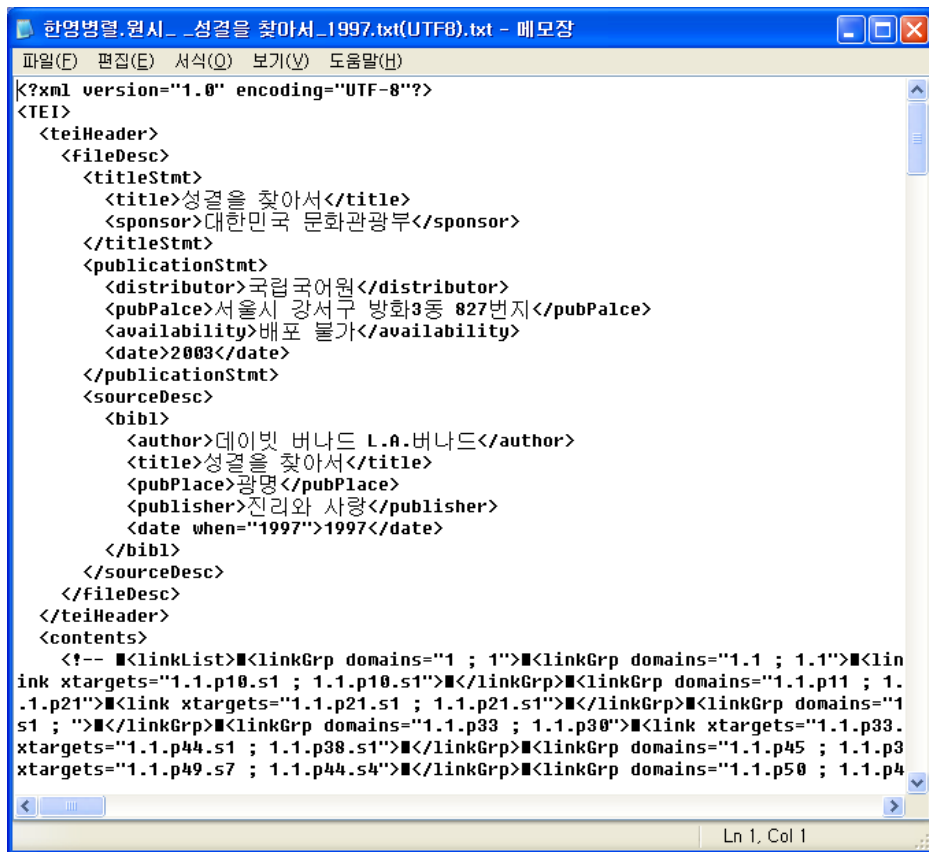


The screenshot shows a text editor window titled "한영병렬 원시...성결을 찾아서_1997.txt - 메모장". The window contains XML code for a bibliographic record. The code is as follows:

```
<cesAlign type=sent version=1.6>
<cesHeader version="2.3">
  <fileDesc>
    <titleStmt>
      <h.title>성결을 찾아서</h.title>
    </titleStmt>
    <publicationStmt>
      <distributor>국립국어연구원</distributor>
      <pubAddress>서울시 강서구 방화3동 827번지</pubAddress>
      <availability status=restricted>배포 불가</availability>
      <PubDate>2003</PubDate>
    </publicationStmt>
    <sourceDesc>
      <biblStruct>
        <monogr>
          <h.title>성결을 찾아서</h.title>
          <h.author>데이빗 버나드 L.A.버나드</h.author>
          <imprint>
            <pubPlace>광명</pubPlace>
            <publisher>진리와 사랑</publisher>
            <pubDate>1997</pubDate>
          </imprint>
        </monogr>
      </biblStruct>
    </sourceDesc>
  </fileDesc>
</cesHeader>
<linkList>
<linkGrp domains="1 ; 1">
<linkGrp domains="1.1 ; 1.1">
<link xtargets="1.1.h1 ; 1.1.h1">
<link xtargets="1.1.h2 ; 1.1.h2">
```

The status bar at the bottom right indicates "Ln 1, Col 1".

[그림 9] 한영병렬 원시 결과물



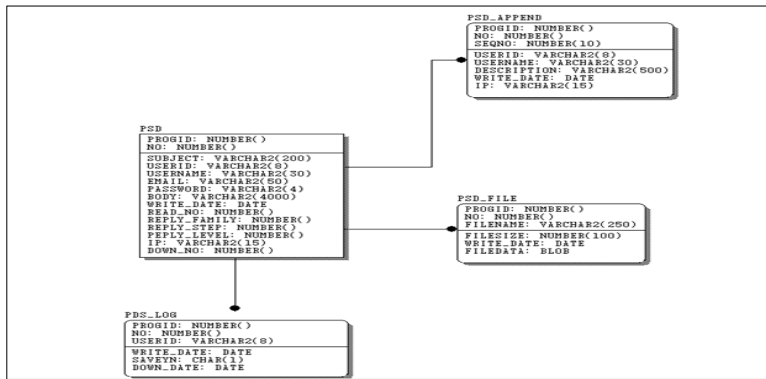
[그림 10] 한영병렬 원시(UTF8)

제 III 장. 21세기 세종계획 성과물 배포

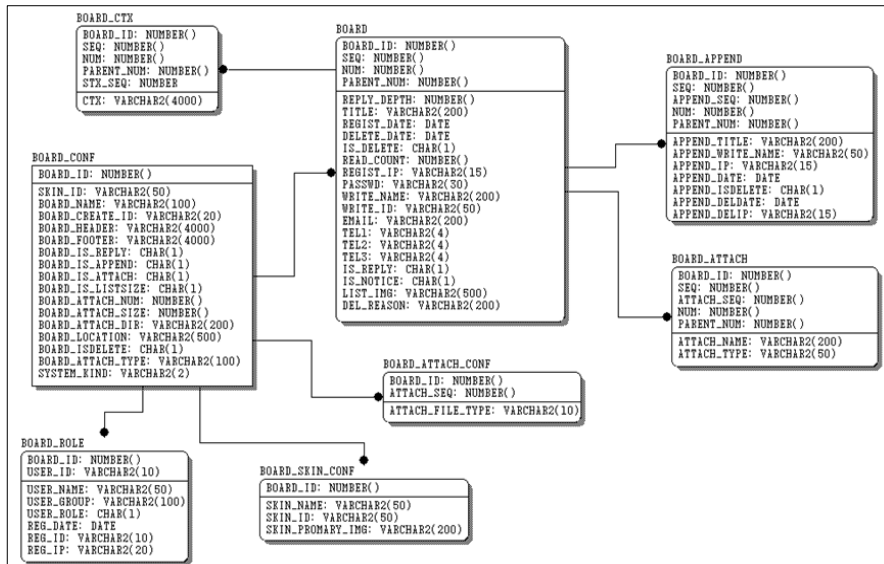
21세기 세종계획 홈페이지의 하위 메뉴로 들어가는 배포 시스템은 성과물 관리 및 다운로드 기능을 전담하며, 배포 결과의 통계 정보를 제공해 준다. 21세기 세종계획의 성과물이 실질적인 연구·개발의 기초 자료로 널리 활용될 수 있기 위해서는 이를 필요로 하는 연구자나 전문가들에게 보급 및 배포를 통하여 이를 활용할 수 있도록 하는 것이 중요하다. 21세기 성과물 배포 시스템은 성과물 배포 및 관리 기능을 웹을 통해 제공함으로써, 21세기 세종계획 성과물의 확산 및 효율적인 관리를 지원할 수 있다. 21세기 세종계획 홈페이지의 하위 메뉴로 들어가는 배포 시스템은 성과물 관리 및 다운로드 기능을 전담하며, 배포 결과의 통계 정보를 제공해 준다.

제1절. 21세기 세종계획 성과물 관리 DB

21세기 세종계획 성과물의 관리 정보의 등록 및 편집, 배포 시스템의 기능 제공, 배포 현황 관리를 위한 DB 스키마를 설계한다.



[그림 11] 자료실 관리를 위한 ER 다이어그램



[그림 12] 성과물 배포 시스템 ER 다이어그램

제2절. 21세기 세종계획 온라인 배포 시스템 개발

1. 배포 시스템 메뉴 구성

배포 시스템의 메뉴는 다음과 같이 구성된다.

○ 말뭉치

- 말뭉치:

- 현대-문어
- 현대-구어
- 병렬-한영
- 병렬-한일
- 역사-역사
- 역사-구비

- 활용 도구

○ 전자 사전

- 전자 사전

- 전자 사전 검색기
- 기타 결과물
 - 국어 기초 자료 구축
 - 국어 특수 자료 구축
 - 전자 사전 개발
 - 한민족 언어 정보화
 - 전문 용어 정비
 - 문자 코드 표준화
 - 국어 정보화 인력 양성
 - 글꼴 개발 보급 지원
- 프로그램
- 보고서

2. 자료 유형별 상세 정보

자료 유형별 상세 정보 항목은 다음과 같다.

말뭉치

제목
 자료명
 저자
 출판사
 출판 연도
 말뭉치 종류
 주석 단계
 언어
 매체
 등록자
 등록일
 조회/내려 받기
 첨부 파일
 파일 이름
 파일 형식
 파일 크기

전자 사전

제목

사전 종류: 상세, 기초 중 하나 선택

사전 범주

항목 수

등록자

등록일

조회/내려 받기

첨부 파일

파일 이름

파일 형식

파일 크기

설명 : 설명 파일의 다운로드 및 열기 가능

파일 이름

파일 형식

파일 크기

설명 내용

기타 결과물

제목

분과

사업 연도

등록자

등록일

조회/내려 받기

첨부 파일

파일 이름

파일 형식

파일 크기

설명 : 설명 파일의 다운로드 및 열기 가능

파일 이름

파일 형식

파일 크기

설명 내용

프로그램

제목

분과

사업 연도

등록자

등록일

조회/내려 받기

첨부 파일

파일 이름

파일 형식

파일 크기

설명 : 설명 파일의 다운로드 및 열기 가능

파일 이름

파일 형식

파일 크기

설명 내용

보고서

제목

분과

사업 연도

등록자

등록일

조회/내려 받기

첨부 파일

파일 이름

파일 형식

파일 크기

3. 배포 시스템 사용자 인터페이스

21세기 세종계획 배포 시스템은 21세기 세종계획 홈페이지의 하위 메뉴로 연계되며, 성과물 배포와 관련된 기능만을 제공하기 위해 개발되었다. 일반 사용자 관리와 기타 기능은 21세기 세종계획 홈페이지가 전담하며, 성과물의 다운로드, 자료의 검색, Q/A 게시판 기능만 제공한다.

○ 성과물 배포 시스템 초기 화면

- 21세기 세종계획 홈페이지의 성과물 배포 메뉴와 연계
- 성과물 검색 입력 부분, 분류별 목록 페이지, Q/A 게시판, 관리 페이지로 구성
- 관리 페이지는 관리자 모드로 로그인 경우에만 나타남
- 21세기 세종계획 홈페이지의 사용자 목록을 공동으로 사용
- 사용자 정보는 쿠키를 통해 전달 받음



[그림 13] 성과물 배포 시스템 초기 화면

○ 분류별 목록 페이지

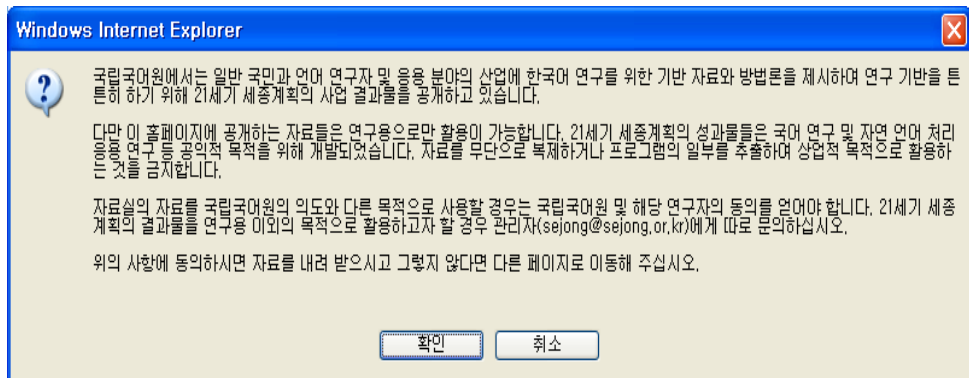
- 선택된 자료 분류에 적합한 자료 목록을 제공
- 자료 목록에 자료실로 자료를 담기 위한 선택 박스를 제공
- 목록 내 검색 기능 제공
- 목록 항목별 정렬 기능 제공



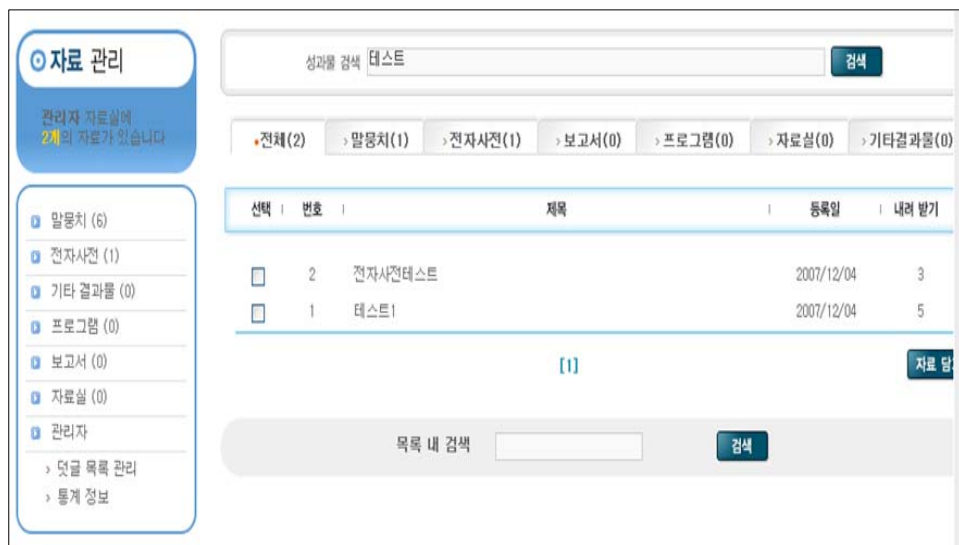
[그림 14] 말뭉치 목록 화면



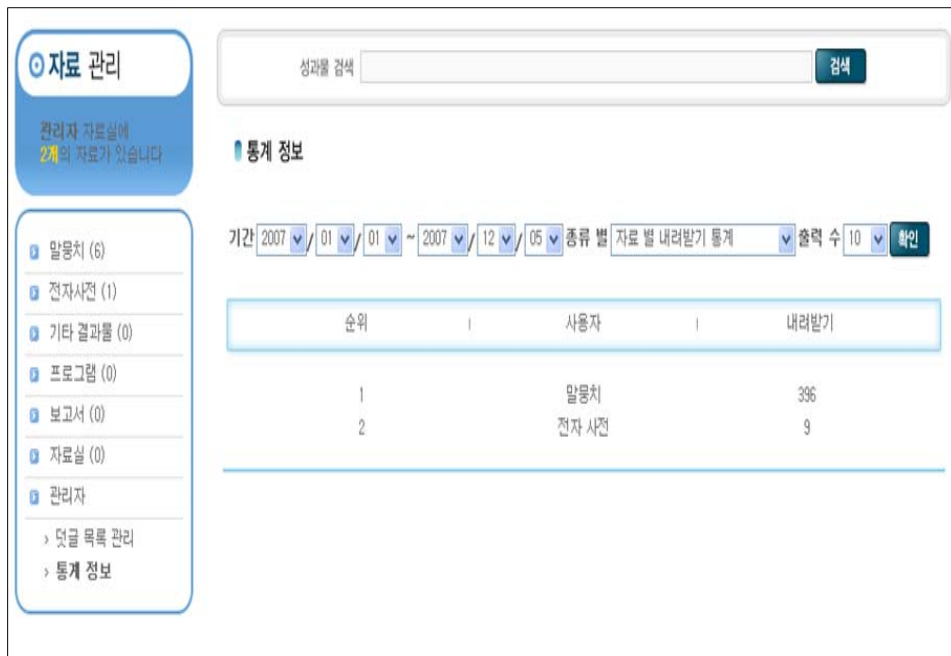
[그림 15] 말뭉치 상세 보기 화면



[그림 16] 자료 사용 동의 화면



[그림 17] 성과물 검색 결과 화면



[그림 18] 통계 정보 보기

제3절. 21세기 세종계획 오프라인 배포

21세기 세종계획 최종 성과물의 오프라인 배포를 위해 시디와 디브디가 포함된 성과물을 제작하여 배포한다.



[그림 19] 배포 시디 표지



[그림 20] 시디 구성

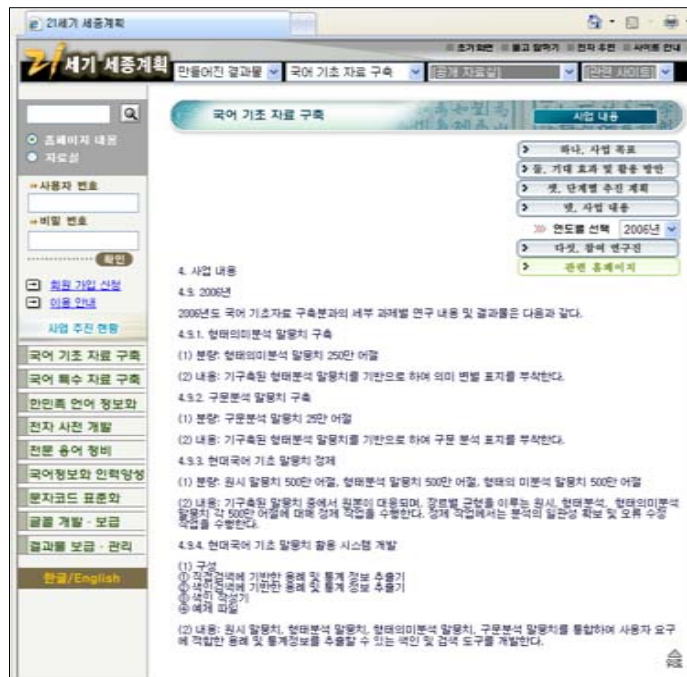
제 IV 장. 국어 연구 정보 제공 및 민원 상담

제1절. 21세기 세종계획 홈페이지 운영

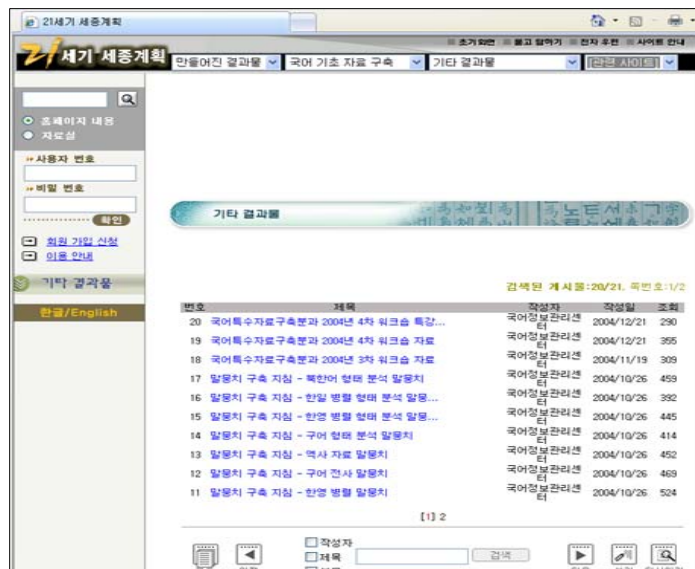
21세기 세종계획 성과물의 보급과 더불어 성과물 활용의 극대화를 위하여, 한국과학기술정보연구원은 21세기 세종계획 홈페이지를 위탁 운영하고 있다. 홈페이지는 세종계획의 전체적인 사업목적과 필요성 등에 관한 사업 전반에 걸친 사업 소개와 사업추진 현황 등을 제공하고 있다. 한국과학기술정보연구원은 21세기 성과물의 활용을 극대화하기 위한 방안으로 홈페이지를 통해 사업 진행 및 관련 분야 동향에 홍보를 적극적으로 시행하고 있으며, 이를 위해 사업과 관련된 학회의 학술대회나 학술발표대회 등에 관한 개최 정보를 주기적으로 갱신하여, 최신 현황을 유지하고 있다.

○ 홈페이지 수정

- 이용 안내 페이지 삭제
- 사업 소개 자료 수정 : 사업 내용, 만들어진 결과물
- 사업 추진 현황 : 각 분과별로 2004, 2005, 2006 년 사업 내용 및 참여 연구진
- 공개 자료실 : 일부 게시판 삭제 및 게시물 이전



[그림 21] 홈페이지 사업 내용 추가



[그림 22] 게시판 수정 및 게시물 이전

- 홈페이지 운영
 - 학술대회 학회, 세미나 공지
 - Q/A 게시판 운영



[그림 23] 공지 사항



[그림 24] 묻고 답하기 게시판

제 V 장. 사업 결과물

제1절. 21세기 세종계획 성과물 관리 지원

결과물의 내용	○ 말뭉치 헤더 변환을 위한 도구 ○ 영문 번역 자료
최종 결과물	○ CD-ROM 형태의 전산 파일 10부

제2절. 21세기 세종계획 성과물 배포

결과물의 내용	○ 온라인 성과물 배포 시스템 ○ 성과물 배포 시디 디자인 및 소책자
최종 결과물	○ CD-ROM 형태의 전산 파일 10부 ○ CD-ROM, DVD-ROM 형태의 성과물 배포시디(1000부)

제3절. 국어 연구 정보 제공 및 민원 상담

결과물의 내용	○ 21세기 세종계획 홈페이지 게시물 보완 자료
최종 결과물	○ CD-ROM 형태의 전산 파일 10부

제4절. 보고서 작성

최종 결과물	○ 최종 사업 보고서 100부 발간
--------	---------------------

제 VI 장. 전산 파일(CD-ROM) 설명

제1절. “0. 보고서” 폴더

- 본 사업의 수행 결과에 대한 최종 보고서를 포함한다.

[표 3] “0. 보고서” 폴더의 내용 설명

파 일	내 용
보고서.zip	“21세기 세종계획 성과물 관리 및 배포 지원” 사업의 최종 보고서

제2절. “1. 21세기 세종계획 성과물 관리” 폴더

- 말뭉치 헤더 변환을 위한 도구를 포함한다.

[표 4] “1. 21세기 세종계획 성과물 관리” 폴더의 내용 설명

파 일	내 용
헤더변환도구.zip	말뭉치 헤더 변환을 위한 도구 프로그램
영문번역자료.zip	영문 번역에 사용된 한글 자료와 영문 자료

제3절. “2. 21세기 세종계획 성과물 배포” 폴더

- 21세기 세종계획 성과물 배포를 위한 온라인 성과물 배포 시스템의 개발 프로그램을 포함한다.

[표 5] “2. 21세기 세종계획 성과물 배포” 폴더의 내용 설명

과 일	내 용
온라인배포시스템.zip	성과물 온라인 배포 시스템 프로그램
오프라인 성과물 배포 작업.zip	오프라인 성과물 배포를 위해 제작된 시디 디자인, 자 켓, 소책자 내용

제4절. “3. 국어 연구 정보 제공 및 민원 상담” 폴더

- 21세기 세종계획 홈페이지 관리 운영에 관한 캡처 이미지를 포함한다.

[표 6] “3. 국어 연구 정보 제공 및 민원 상담” 폴더의 내용 설명

과 일	내 용
21 세 종 계 획 홈 페 이 지.zip	21세기 세종 계획 홈페이지 보완 파일

주 의

1. 이 보고서는 국립국어원의 21세기 세종계획 ‘21세기 세종계획 성과물 관리 및 배포 지원’ 연구 용역 결과 보고서이다.
2. 이 연구 보고서의 내용을 대외적으로 발표할 때에는 반드시 국립국어원의 21세기 세종계획 연구 결과물임을 밝혀야 한다.

21세기 세종계획 성과물 관리 및 배포 지원

2007. 12. 8. 인쇄

2007. 12. 8. 발행

발행자 : 국립국어원
서울시 강서구 방화3동 827
전화 2669-9775

인쇄처 : (주)대덕사
대전시 유성구 어은동 116-11
전화 042) 861-2181~3