

YDK : 한국어 통합정보사전 개발 시스템의 설계 및 구현

최용준*, 황도삼¹, 최기선*

¹영남대학교 컴퓨터공학과, *한국과학기술원 전산학과

YDK : A Thesaurus Development System for Korean Language

Yongjun Choi¹, Dosam Hwang*, Keysun Choi*

¹Yeungnam Univ., *KAIST

요 약

본 논문에서는 다국어 통합정보사전 구축을 위한 한국어 용어의 통합정보사전을 구축하기 위해 정립된 사전개발 시스템의 개발 방법론을 바탕으로 한 통합정보사전개발 시스템(YDK: Yongjun Dosam Keysun)을 설계하고 구현한다. 개발한 YDK는 기존의 각종 사전의 다양한 사전정보를 통합할 수 있을 뿐 아니라 여러 자연언어처리 시스템들과의 접속을 통해 언어처리를 위한 사전정보를 손쉽게 통합할 수 있어 고품질의 전자사전을 개발할 수 있다.

I. 서론

자연언어처리 프로그램들은 대개 많은 양의 문법정보, 의미정보, 용례 등을 필요로 한다. 이러한 정보는 전자사전을 통해 제공되며, 제공되는 정보의 양과 질에 따라 프로그램의 성능도 영향을 많이 받는다.

전자사전을 구축하는 일은 매우 노동집약적이면서도 전문지식을 필요로 하며, 이러한 사전은 일단 완성이 되었다 하더라도 지속적인 단어의 추가가 필요하며, 필드의 추가 및 변경의 요구에 따라 새로운 형태의 사전을 만들어야 하는 경우가 많다. TDMS와 같은 전자사전관리 시스템은 여러 형태의 전자사전으로부터 정보를 추출하기 때문에 작업의 어려움은 개선되었으나, 여러 자연언어처리 시스템의 처리결과를 직접 참고할 수는 없으므로, 보

다 많은 양의 정보를 필요로 하는 통합정보사전에는 적합하지 않다.

본 논문에서는 사전개발시스템의 개발 방법론을 정립하고, 이를 바탕으로 한 통합정보사전 개발 시스템(YDK: Yongjun Dosam Keysun)을 설계하고 구현한다. 개발한 YDK는 기존 사전의 정보를 통합할 수 있을 뿐 아니라 여러 자연언어처리 시스템들과의 접속을 통해 다양한 정보를 손쉽게 통합할 수 있어 고품질의 전자사전을 개발할 수 있다.

II. YDK의 설계

2.1 사전구축 방법론에 따른 사전개발 도구

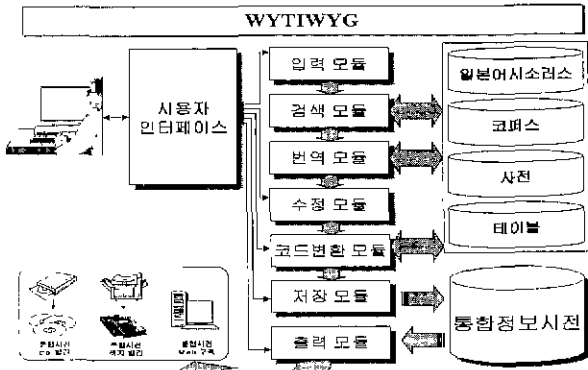
통합정보사전을 개발하기 위한 사전구축 방법으로 수작업 방식과 반자동 방식이 있다. 수작업은 사람의 손을

© 본 연구는 1997.12.4부터 1998.10.3까지 과기부의 STEPXXXX과제인 "대용량 국어정보 심층처리 및 품질관리기술 개발"(KAIST)의 위탁과제로 수행하였다.

로 직접 모든 사전자료를 수집하고 입력하여 구축하는 방법으로 많은 노력과 시간 및 비용이 들어가는 단점이 있다. 이에 비해 반자동 방식은 사전구축 시스템을 이용하여 구축하는 방법으로 많은 양의 데이터를 빠른 시간에 입력할 수 있을 뿐 아니라 실지적인 사전원시정보를 가지고 있는 코퍼스나 전자사전과 같은 정보원으로 부터 사전구축에 필요한 정보들을 자동으로 추출할 수 있으므로 고품질의 사전을 개발할 수 있다 본 연구에서는 반자동 방식으로 통합정보사전을 개발하기 위해, 통합정보사전 개발 시스템인 YDK(YongjunDosamKeysun) 시스템을 개발하였다.

2.2 YDK

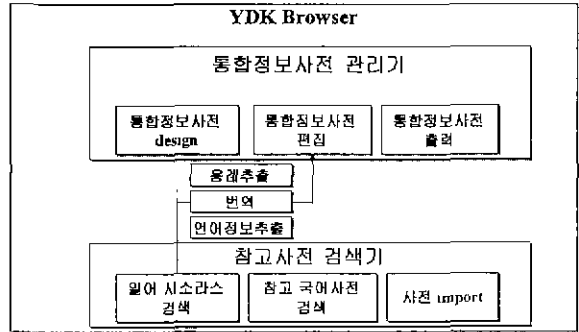
YDK는 사전개발환경의 개발 개념으로 WYTIWYG(What You Think Is What You Get)을 가지고 있다. YDK는 생각할 수 있는 모든 자연언어 자원들을 이용할 수 있도록 전자사전과 자연언어 처리 도구들을 하나의 인터페이스 내에 통합한다. 분산되어 있는 자연언어처리 자원들이 각각의 자원에 적합한 모듈을 개발하고, 각 모듈들의 인터페이스를 표준화시켜 통합하는 체계를 따른다 이를 [그림 1]에 보인다.



[그림 1] YDK

2.3 시스템 구성

YDK는 통합정보사전 관리기와 참고사전 검색기 및 자연언어처리 시스템과의 접속처리부로 구성된다, 이를 [그림 2]에 나타낸다.



[그림 2] YDK 시스템 구성도

(1) 통합정보사전 관리기

통합정보사전을 설계하고 자료를 입력하는 기능을 가진다. 사전 항목을 정의하거나 변경할 수 있으며, 항목 자료를 입력할 수 있다.

(2) 참고사전 검색기

현존하고 있는 여러 전자사전들을 참고할 수 있는 기능이다.

(3) 자연언어처리 시스템과의 접속처리부

자연언어처리 시스템들로부터의 정보 획득을 위한 처리부로 서버 시스템에 접속하여 Query를 넘겨주고 처리결과를 돌려받는 역할을 한다. YDK는 서버 시스템에 설치되어 있는 YDK지원 시스템과의 교섭을 통해 자연언어처리 시스템과 연결된다.

III. YDK의 구현

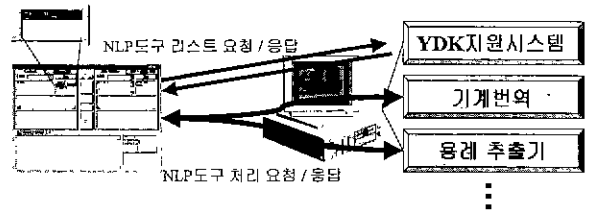
3.1 YDK의 구현 환경

YDK는 Intel Pentium PC, MS-Windows95환경에서 MS-Visual C++ 5.0과 MS-Visual BASIC 5.0을 이용하여 개발하였으며, 서버에 설치되는 YDK지원 시스템은 SUN ULTRASparc, Solaris 2.5환경에서 CERN httpd 3.0 a와 gcc 2.7.2.3과 gdbm 1.7.3을 이용하여 개발하였다.

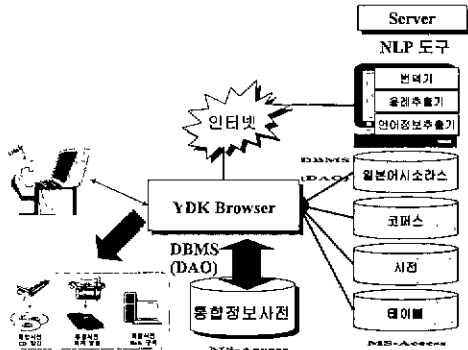
3.2 YDK Version 1.0

YDK Version 1.0은 단일 사용자를 위한 시스템으로

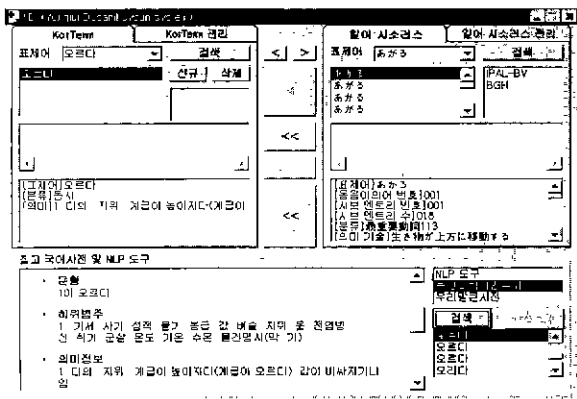
사전정보는 사용자 시스템의 디스크에 저장되며, 자연언어처리 시스템들은 서버 시스템에 설치되어 있다. 사용자는 사전정보를 검색하고자 할 때나, 자연언어처리 시스템의 결과를 필요로 할 때는 간단한 버튼 조작만으로 결과를 얻을 수 있다. YDK Version 1.0의 구성을 [그림3]에 보이고, 실제로 구현한 YDK의 실행예를 [그림 4]에 보인다.



[그림 5] YDK지원 시스템



[그림 3] YDK Version 1.0의 구성



[그림 4] YDK의 실행예

YDK는 실행시마다 YDK 서버시스템(nlp.yeungnam.ac.kr)에 접속하여 등록된 자연언어처리 도구의 리스트를 얻어서 리스트창에 표시해 준다. 사용자의 자연언어 처리 입력(버튼 클릭)을 받게되면 서버 시스템에 설치된 자연언어처리 도구에 접속하여 Query를 넘겨주고 그 처리 결과를 넘겨 받아 사용자에게 보여준다. 이를 [그림 5]에 나타낸다

3.3 결과 및 고찰

YDK는 한국어 용언의 다국어 통합정보사전인 YDK-Term개발에 활용되고 있다. 특히 YDK의 자연언어처리 시스템과의 연계부분은 보다 쉽게 자연언어처리 시스템들을 이용하게 함으로써 사전정보의 정확성을 높이는 데 결정적 역할을 하고 있다.

IV. 결론

본 논문에서는 사전개발시스템 개발 방법론을 정립하였으며, 이를 바탕으로 한 통합정보사전개발 시스템(YDK:Yongjun Dosam Keysun)을 설계하고 구현하였다. 개발한 YDK는 기존 사전의 정보를 통합할 수 있을 뿐만 아니라 여러 자연언어처리 시스템들과의 접속을 통해 다양한 정보를 손쉽게 통합할 수 있어 고품질의 전자사전을 개발할 수 있다는 장점이 있다.

향후 연구과제로는 다중 사용자를 지원하는 YDK-Multiuser Distributed 버전을 개발하여 보다 효율적인 사전 개발 환경을 구축하는 것이 있다.

- [1] 이재성 외3, “텍스트 및 전자사전 관리시스템의 설계”, 한국정보과학회&한국인지과학회, 제8회 한국어 정보처리 학술대회 논문집, pp.408-414, 1996
- [2] 최병진 외3, “표준화를 위한 일반사전의 논리 구조”, 한국정보과학회&한국인지과학회, 제8회 한국어 정보처리 학술대회 논문집, pp.415-423, 1996
- [3] 한국과학기술원, “텍스트코퍼스 및 전자사전 관리시스템(TDMS)”, 과학기술처, 통합 국어정보베이스 최종보고서, pp.17-150, 1996
- [4] 황도삼 외2, “자연언어처리”, 흥릉과학출판사, 1998