

하이퍼텍스트 한글백과사전의 개발

전 경현*, 강 현규, 김 용일, 박 상규, 최 기선
한국과학기술원 전산학과

Development of Hypertext Encyclopedia

KyongHun Jeon, HyunKyu Kang, YoungIl Kim, SangKyu Park, KeySun Choi
Computer Science Dept. KAIST

요 약

개인용 컴퓨터가 많이 보급되고, CD-ROM과 같은 대용량의 기록장치가, 널리 퍼지면서, 전자화된 사전이 널리 사용되게 되었다. 그러나, 이러한 전자사전들은 대개 매우 단순화된 탐색기능만을 제공하므로, 전자화의 잇점을 크게 살리지 못하고 있다. 따라서, 컴퓨터의 힘을 빌린, 보통의 종이로된 사전보다 훨씬 유용한 전자사전을 생각해 볼 수 있을 것이다. 이러한 노력이 과거 80년대 말에 하이퍼텍스트로 시작되어 이제는 매우 상업적인 곳까지 사용되어지고있다. 국내에서도 전자사전의 개발은 다양하게 진행되어오고 있으나, 대부분 컴퓨터를 위한 사전, 다시말하면, 빠른 접근시간, 적은 메모리사용만을 고려한 사전에 그치고 있다.

본 논문에서는 이러한 필요성에 부합하여, 백과사전자료를 가지고 사람을 위한 전자백과사전을 개발하는 방법을 소개한다. 특히, 이러한 개발과정을 자동화할 수 있음을 보이고, 백과사전을 하이퍼텍스트로 자동변환하는 방법을 자세히 기술한다. 이 방법은 원시 자료를 하이퍼텍스트로 변환하기 좋은 정형화된 중간자료로 바꾸고, 이 중간자료와 기초적인 명사추출 알고리즘을 이용하여, 각 노드들 사이에 링크를 만드는 것이다. 또한, 이 방법을 이용하여 개발한 HE (Hypertext Encyclopedia) 시스템을 소개한다.

I. 서 론

종이로된 두꺼운 책들이 한두장의 컴퓨터 디스크에 담기는 세상이 되었다. 개인용 컴퓨터의 보급과 CD-ROM기술의 발달에 힘입어 이러한 전자책들은 다루기 쉽고, 편리한 도구가 되었다. 그러나, 단지 책의 매개물을 종이에서 디스크로 바꾸는 수준에서 벗어나, 새로운 개념의 책을 설계하고 구현하고자 하는 노력이 1980년대 후반들어 하이퍼텍스트를 통하여

광범위하게 일어났고, 앞으로도 계속되는 추세에 있다. 특히, 사전이나, 기술문서등을 하이퍼텍스트화 하였을 때 얻는 큰 장점인 지루한 탐색시간의 대폭 축소는 이들 문서들의 하이퍼텍스트화를 더욱 촉진 시키고 있다[1,4].

흔히, 백과사전을 비롯한 사전류의 문헌을 읽을 때에는, 표제어의 설명문중에서 새로운 표제어를 발견하고 이에 궁금증을 느껴 다시 찾는 지루하고 단조로운 일이 허다하게 발생한다. 어느때는 표제어라고 추측했던 단어가 사전의 표제어로 등록되어있지 않아서, 다른 사전을 뒤져야 하는 경우도 생긴다. 누구나 이를 한번쯤은 경험해 보았을 것이고, 이러한 탐색작업을 극도로 단순화 시킬 수 있는 방법을 찾고자 할 것이다. 잘못된 탐색을 방지하기 위하여, 설명문중에 나타나는 다른 표제어들은 미리 다른 글씨체로 표시해 놓을 수 있다. 그러면, 적어도 잘못된 표제어를 탐색하느라 시간낭비하는 일은 없어질 것이다. 또한, 표시되어있는 표제어가 궁금할 경우에는 인덱스를 뒤져서, 그 표제어의 설명문이 기록되어있는 페이지를 찾아낼 수 있다. 이 경우에 조금 더 개선한다면, 그 표제어를 보는 것과 동시에 매우 간단한 의사표시로 그 설명문의 내용을 바로 찾을 수 있다면, 더이상 좋을 것이 없다고 하겠다. 이러한 욕구를 모두 충족시켜주는 것이 바로 하이퍼텍스트화된 사전이고, 백과사전을 하이퍼텍스트화 하는 작업은 본 논문에서 앞으로 기술하는 바와 같이 매우 체계적으로 자동화 될 수 있다[1,3].

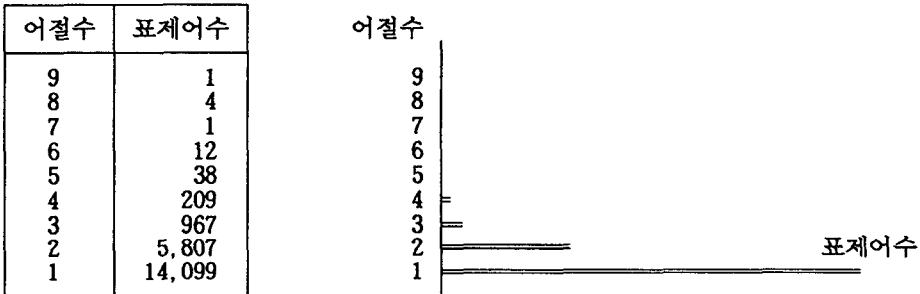
컨클린은 그의 논문에서 하이퍼텍스트를 “문헌내에서 또는 문헌간에 컴퓨터에의하여 지원되는 링크를 가진 구조”라고 정의하였다[2]. 링크는 매우 광범위한 개념이나, 간단히 정보의 연결된 상태를 나타낸다고 할 수 있다. 위에서와 같이 백과사전 표제어의 설명문에 나타나는 또다른 표제어를 이용하여, 다시 백과사전을 뒤척거리면, 이러한 표제어가 바로 링크의 역할을 한다고 말할 수있다. 일반적으로 하이퍼텍스트는 문서내에 존재하는 여러가지 형태로 나타난 링크를 선택함으로써 새로운 내용을 독자에게 보여주는 형식을 취하고 있다. 이러한 구조는 백과사전에 적용하였을 때, 무척 자연스러운 것이다. 즉, 표제어의 설명문중에 나타나는 다른 표제어를 마우스나 키보드, 또는 터치스크린인 경우에는 손을 이용하여, 선택할 수 있고, 선택 즉시 새로운 표제어의 내용이 보여지는 것이다. 마치 생각이 자연스럽게 전이하듯이, 디스플레이상에 표시되는 내용이 전이된다. 이처럼 하이퍼텍스트는 백과사전을 위한 매우 훌륭한 구조이다.

본 논문에서의 위에서 바라본 하이퍼텍스트 한글백과사전의 필요성과 유용성을 바탕으로, 하이퍼텍스트 한글백과사전(HE : Hypertext Encyclopedia) 개발 경험을 기술하였다. 2장은 하이퍼텍스트화를 위한 데이터인 한글 백과사전의 규모와 분석정보를 다루었고, 3장은 최종적으로 생성된 하이퍼텍스트 한글백과사전의 형태와 기본적인 탐색방법을 기술하였으며, 4장은 2장에서 기술된 한글 백과사전이 어떻게 3장의 하이퍼텍스트 한글백과사전으로 변환되어지는가의 방법론을 기술하였다. 그리고, 5장은 하이퍼텍스트화 경험을 바탕으로 취

약점과 개선방향및 응용방향을 기술하였다.

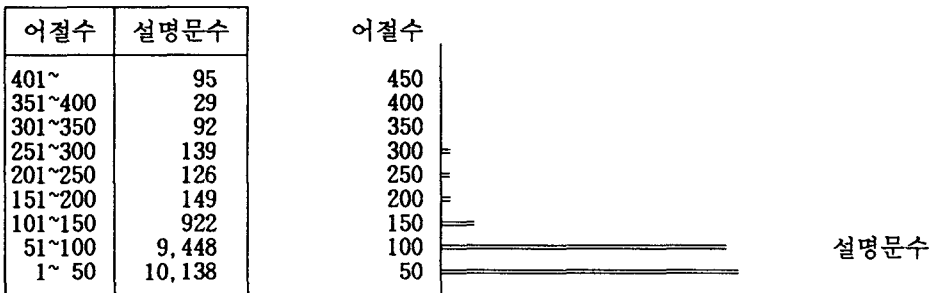
II. 한글백과사전 자료

HE의 개발에 사용된 한글백과사전은 계몽사 학생백과사전 총 6권이다[7]. 이는 학생들을 대상으로 22,000여개의 표제어를 골라 그의 설명을 실은 것이다. 원시 데이터는 약 9메가바이트크기이다. 백과사전은 크게 표제어와 표제어의 설명문으로 이루어져 있으며, 표제어는 백과사전에서 기술하고자 하는 개념을 지칭하는 단어또는 어절이다. 원 표제어의 총 갯수는 21,530개이나, 필요에의하여 동음이의어인 표제어들을 한데 묶었으므로, 구현에 사용된 표제어의 총 갯수는 21,138개이었다. 가장 긴 표제어의 어절수는 9이고, 평균적인 어절수는 1.41이다. 그림 1은 표제어의 어절수 분포이다.



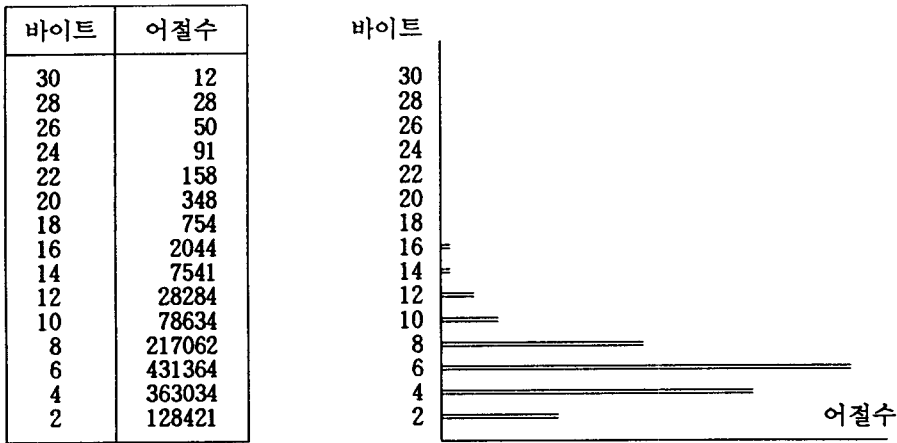
[그림 1] 표제어의 길이 분포

그림 1의 결과를 주목하면, 60%이상이 단일어절 표제어로 이루어져있다는 것을 알 수 있으며, 10어절 이상으로 이루어진 표제어는 없다는 것을 알 수 있다. 설명문은 각각의 표제어가 뜻하는 의미를 다른 표제어로 표현할 수 있는 개념이나, 아니면 더 기초적인 개념을 사용하여 기술한 영역이다. 각각의 표제어마다 붙어있는 설명문의 어절수는 최대 1,000여개 부터 최소 20여개까지 다양하며, 평균적인 어절수는 약 50이다. 다음은 설명문의 어절수 분포이다.



[그림 2] 설명문의 어절수 분포

백과사전 자료가 포함하고 있는 총 어절 수는 1,257,852개이다. 다음은 백과사전의 어절길이 분포이다.



[그림 3] 어절길이 분포

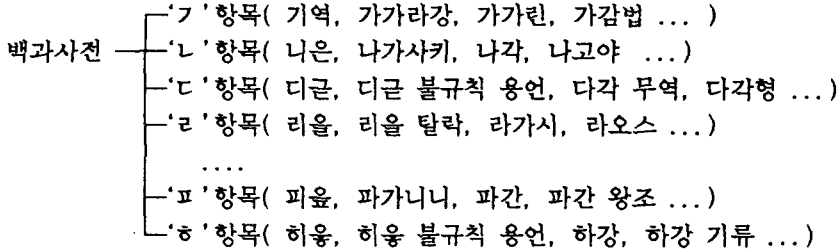
계몽사 학생백과사전은 단순한 문자들의 나열이상으로, 그림과 표 그리고, 특별한 부가 설명을 포함하고 있다. HE의 변환작업은 표제어와 설명문만을 대상으로 하였으며, 기타 자료는 포함하지 않았다.

Ⅲ. 하이퍼텍스트 한글백과사전

하이퍼텍스트 한글백과사전(HE : Hypertext Encyclopedia)은 IBM-PC상에서 C언어를 사용하여 구현되었으며, 브라우저로 “한글 MS-WINDOWS 3.1 한글 도움말” 프로그램을 사용한다. 그 결과로, 매우 편리한 윈도우 환경을 사용할 수 있으므로 초보자가 시스템을 사용하기 위하여 들이는 시간을 크게 단축할 수 있으며, 한글 MS-WINDOWS 3.1에 익숙한 사람은 새로운 시스템을 공부해야하는 번거로움을 피할 수 있다는 장점이 있다. 한글 윈도우 3.1도움말 시스템은 그림링크나, 프로그램수행링크, 팝업윈도우링크 및 기타 여러가지 형태의 링크를 지원할 수 있지만, HE는 링크버튼을 눌렀을 경우, 윈도우의 내용이 링크가 지시하는 새로운 노드의 내용으로 채워지는 링크형태(Link Topic)를 사용하였다.

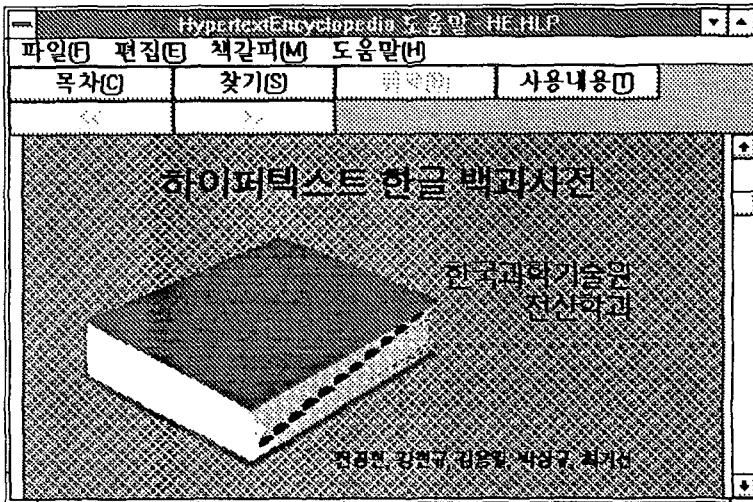
모든 표제어는 그림 4에서 볼 수 있듯이 사전순서를 담은 계층 구조를 가지고 있다. 여기서 계층구조는 의미적인 구조가 아니고, 첫번째 자음의 종류에 따라서 나뉘어진 표제어 집단들의 모임이라고 생각할 수 있을 것이다. 집단들간에는 순서가 있고, 각각의 집단내에

서도 순서가 있다. 이러한 계층구조는 표제어들을 자연스럽게 배치할 수 있도록 하며, 찾게 하는 표제어를 보다 체계적으로 찾을 수 있도록 돕는 역할을 한다.



[그림 4] HE 표제어의 계층구조

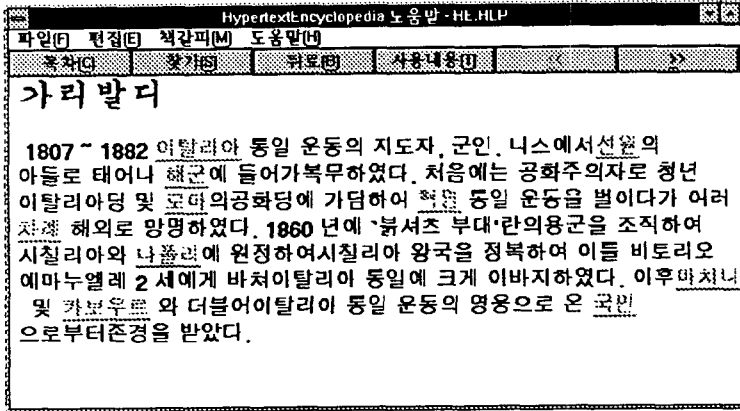
설명문에 나타나는 표제어 링크들은 한 단어이거나, 한 어절 또는 연속된 다수의 어절에 걸쳐서 나타날 수도 있다. 이러한 표제어들은 독자가 쉽게 구별할 수 있도록, 본문과는 다른 속성으로 표시된다. 그리고, 이러한 링크를 마우스나 키보드를 이용하여 활성화 시키면, 링크에의하여 연결되어있는 노드의 내용이 표시되어진다.



[그림 5] HE의 초기화면

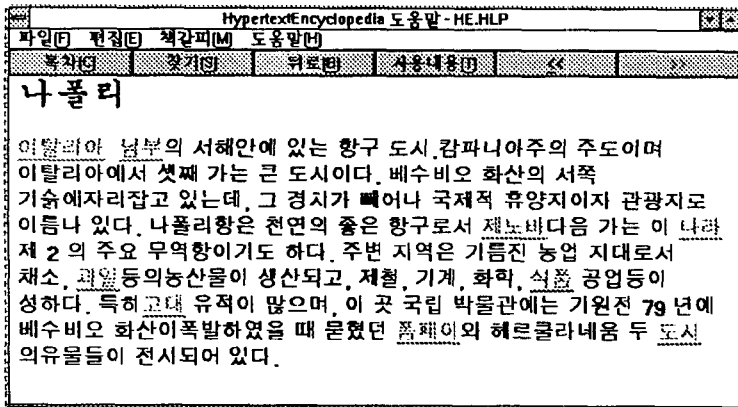
HE를 처음 한글 도움말 프로그램에 불러들이면, 그림 5와 같은 초기화면이 나타난다[그림 5][5]. 이 상태에서 원하는 단어를 찾기 위해서는 “찾기”버튼을 이용한다. 찾기 버튼은 각각의 노드의 키워드를 이용하여, 노드에 접근할 수 있도록 해준다. 찾기 버튼을 이용하여 노드 “가리발디”를 찾았다고 하자[그림 6]. 노드 “가리발디”는 ‘이탈리아’, ‘선원’, ‘해

군', '로마', '혁명', '나폴리', ... 등등의 링크를 포함하고 있으며, 이드은 밑줄이 그어진 상태로 표시된다. 그림 6상태에서 마우스로 링크 '나폴리'를 선택하였다면, 윈도우의 내용이 바뀌어지면서, 노트 "나폴리"의 내용이 나타난다[그림 7].



[그림 6] 노트 "가리발디"

그림 6에서 볼 수 있는 바와 같이, 표제어는 궁서체로 크게 나타내 보이도록 했으며, 설명문은 윈도우 기본 폰트 크기 보다 약간 크게 나타나도록 하였다. 이는, 디스플레이의 미발달로 인하여, 아직 책을 보는 것이 같은 크기의 디스플레이상의 글자를 보는 것보다, 도움이 된다는 데서 착안한 것이며, 큰 글자가 그러한 부효과를 감소시킬 수 있을 것이다.

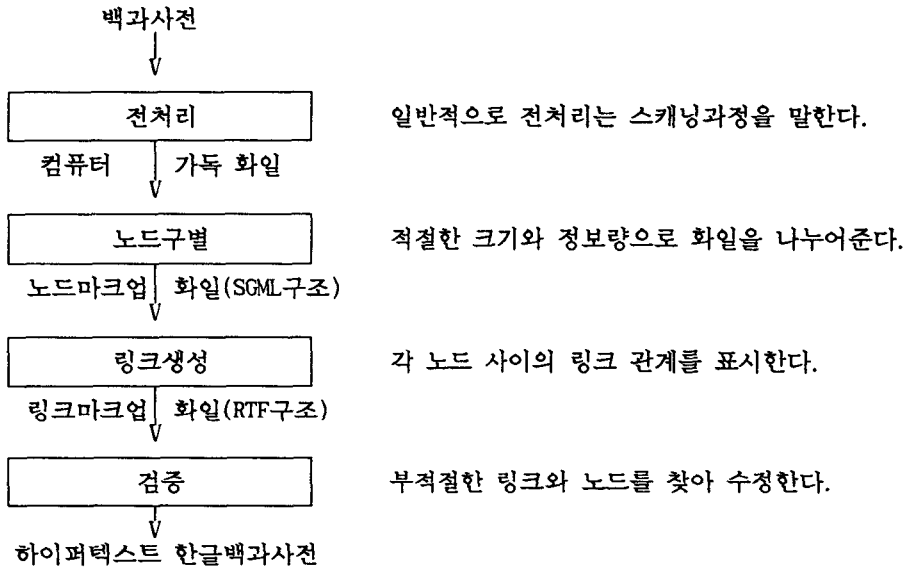


[그림 7] 그림 6에서 '나폴리' 링크를 선택한 후의 화면

윈도우 도움말은 이외에도, 지금까지 탐색한 문서들의 위치를 기억하는 '히스토리' 기능, 그리고, 주어진 노트의 순서에 의하여 이전, 이후 노트를 방문할 수 있는 순차기능등의 많은 기능을 제공하므로, 하이퍼텍스트 한글백과사전의 사용자는 이러한 기능을 모두 이용할 수 있다.

IV. 변환과정

문서를 하이퍼텍스트로 변환하는 보편적인 방법이 존재하는 것은 아니지만, 큰 범주로 나누어 보자면, 기계 가독형 화일을 생성하는 전처리 단계, 각각의 노드의 크기와 위치를 결정하는 노드구별단계, 계층구조 링크나 상호참조 링크를 생성하는 링크 생성단계, 그리고, 변환의 정확성을 위하여 링크의 유효성을 검증하는 검증단계가 있다. 그림 8은 이러한 변환단계를 도식적으로 나타내었다.



[그림 8] 하이퍼텍스트화 과정

4.1 전처리

전처리 과정은 초기자료를 하이퍼텍스트화 하기 좋은 정형화된 자료로 바꾸어주는 작업이다. 일반적으로는 종이문서를 스캐닝등을 거쳐서 전자화하고, 오류를 수정하는 작업이 포함되지만, HE의 개발에 사용한 “계몽사 학생 대백과 사전”은 이미 기계가독형으로 되어있는 자료를 구하였기 때문에, HE 개발시의 전처리과정은 코드변환작업과 자료의 오류수정작업이 포함되었다.

4.2 노드 구별

노드구별과정 적당한 정보량으로 분리하는 작업이므로, 의미적인 분리가 가능하여야한다. 따라서 본질적으로 어려운 작업이라 할 수 있으나, HE의 개발에서는 백과사전이라는 특

수성에 의하여, 각각의 표제어와 설명문이 쉽게 분리할 수 있었으며, 이러한 노드 구별의 결과로 각각의 표제어집단별로 마크업된 화일을 만들어 주었다. 여기서 사용된 마크업은 SGML형식을 빌린 것으로, 백과사전을 하나이상의 레코드의 나열이라 정의하고, 각각의 레코드는 레코드번호와 표제어, 그리고 설명문으로 이루어진다[8]. SGML형식으로 기술한 문서형태정의(DTD : Document Type Definitions)는 다음과 같다[그림 9].

<!--	ELEMENTS	MIN	CONTENT -->
<!ELEMENT	encyclopedia	--	(record+)>
<!ELEMENT	record	-0	(id, title, description)>
<!ELEMENT	id	00	(#PCDATA)>
<!ELEMENT	title	00	(#PCDATA)>
<!ELEMENT	description	00	(#PCDATA)>

[그림 9] 노드 화일의 문서화일구조(DTD)

노드는 원시백과사전 데이터화일에서 분리된 표제어와 설명문을 그대로 따와서 만들지 않았고, 몇가지 수정을 가하였다. 그 중에서 가장 중요한 것은 동음이의어의 처리인데, 백과사전의 표제어 “가사”는 1) 가정 생활을 꾸려 가는데 필요한 여러가지 일. 2) 걸으로 보기에 숨도 안쉬고 심장도 멈춰 죽은 것같이 보이나 뒤에 다시 살아날 수 있는 상태. 3) 중이 입는 법의. 4) 우리 나라의 전통 성악곡의 한 갈래. 5) 조선 초기에 나타나 조선 말기 까지 짓고 불렀던 시가의 한 형식. 무려 5가지의 중의성을 가지며, 각각은 원시백과사전에서 독립된 표제어로 나타난다. HE의 개발에서는 이러한 중의성을 피해가기 위하여 이러한 중의적인 노드들을 모두 한데 묶어 하나의 노드로 하고, 각각의 노드들을 하위항목으로 만들었다. 이는 중의성의 해결을 백과사전을 읽는 독자에게 자연스럽게 맡긴 것이라 할 수 있다.

```

<record>
<id>
ga:00079
</id>
<title>
가리발디
</title>
<description>
1807 ~ 1882 이탈리아 통일 운동의 지도자, 군인.
니스에서 선원의 아들로 태어나 해군에 들어가 복무하였다.
처음에는 공화주의자로 청년 이탈리아당 및 로마의 공화당에 가담하여 혁명
통일 운동을 벌이다가 여러 차례 해외로 망명하였다.
1860년에 '부셔츠 부대'란 의용군을 조직하여 시칠리아와 나폴리에
원정하여 시칠리아 왕국을 정복하여 이를 비토리오 에마누엘레 2 세에게 바쳐
이탈리아 통일에 크게 이바지하였다.
이후 마치니 및 카보우르와 더불어 이탈리아 통일
운동의 영웅으로 온 국민으로부터 존경을 받았다.
</description>
</record>

```

[그림 10] 노드마크업의 예(노드 : “가리발디”)

그림 10에 나타난 레코드번호(ID)는 각각의 표제어그룹에서 몇번째 순서인가를 나타내고 있다. 즉, 표제어 “가리발디”는 ‘가’ 그룹의 79번째 표제어이다. 모든 노드는 위와 같이 SGML형식으로 노드 마크업된다.

4.3 링크 생성

링크의 생성또한 노드 구별문제 못지 않은 의미적인 상관관계의 문제가 발생한다. 의미적인 상관관계의 강도차이라든가, 실제로 이러한 관계가 있다고 할 것인가 부정할 것인가 하는 등등의 제 문제가 있다. 따라서, 링크의 생성에는 의미분석단계가 필요로 되지만, 단순한 구문관계만을 이용해서도 매우 만족할 만한 결과를 얻을 수가 있었다. 다음은 HE의 개발에서 사용된 링크 생성 기법을 기술한다.

링크의 생성을 위하여, 기본적으로 필요로되는 것은 설명문중에 있는 표제어를 추출해내는 것이다. 이러한 표제어 추출을 위하여 미리, 각각의 표제어와 그 ID를 표에 등록하여 놓고[그림 11], 설명문을 조사하면서 표제어 표에 등록된 부분이 있는가를 조사한다.

가름솔	ga:00076
가름수	ga:00077
가리개	ga:00078
가리발디	ga:00079
가리비	ga:00080
가마	ga:00081
가마니	ga:00083
...	
나폴레옹. 법전	na:00094
나폴레옹. 3. 세	na:00095
나폴리	na:00097
나폴리. 왕국	na:00098
나프타	na:00099
나프탈렌	na:00100

[그림 11] 표제어 목록(부분)

그림 11에서 ‘.’는 조사생략처리를 한후 연결되었다는 의미를 가진다. 즉, “가격의 혁명”이라는 표제어는 “가격. 혁명”으로 표제어목록에 기재된다. 명사구로 이루어진 복합명사는 앞서 II장에서 살펴보았다시피 무려 33%에 이른다. 따라서, 이들 명사구 표제어를 추출하지 못한다면, 백과사전의 독자를 무척 실망 시키게 될 것이다. 복합명사구의 처리를 위하여, 현재까지 조사생략처리한 어절들을 결합시켜서, 표제어 목록을 열람한뒤, 표제어 목록에 있는지, 없는지, 있다면, 새로운 어절이 추가되면서, 더 긴 표제어를 생성할 가능성은 없는지를 검사하고, 가능한한 가장 긴 표제어를 선택한다. 즉, “가격 탄력성”의 경우에는 “가격”만으로도 표제어목록에 있으므로, 링크를 생성할 수 있지만, 뒤에 “탄력성”, “혁명”

등의 새 어절이 추가될 수 있으므로, 링크생성을 유보하다가, 뒤따라 나오는 “탄력성”을 알고, “가격 탄력성”이라는 링크를 생성하게 된다. 이처럼 상호 참조 링크의 생성에 있어서는 최장 표제어를 링크로 삼았다. 조사 생략 처리를 위하여 사용한 조사 사전의 엔트리 갯수는 모두 596개이며, 조사 중에서 조사사기 보다는 접사로써 많이 사용되는 ‘도’, ‘이’ 등은 생략하였다.

한음절 표제어의 경우에, 특히 ‘이’와 같은 표제어는 “기생충의 일종”이라는 뜻을 가지지만, 실제 설명문중에 대명사의 의미로 많이 출현하게된다. 따라서, 이러한 한음절 표제어를 모두 링크로 만들어주면, 부적절한 링크가 다수 생겨나게 되서 불합리하므로, 한음절 표제어는 아예 표제어 목록에도 집어넣지 않았다. 결국 한음절 표제어에 접근 하는 방법으로는 “찾기”기능을 이용하든지 아니면, 순서가 가까운 표제어를 찾아서, 거기서 순차적으로 접근 하는 방법밖에는 없다.

기타, 링크 생성과정중 현재 처리중이 노드의 표제어가, 그 노드의 설명문 중에 많이 등장한다. 따라서, 이들을 링크로 만들어주면, 무용한 링크가 다수 발생하게 되므로, 이들을 없애주어야한다.

링크 생성 단계의 결과로 나타나는 것은 링크가 마크업된 RTF(Rich Text Format)화일이다[6]. SGML HyTime이 아닌 RTF화일을 링크 생성의 최종단계로 사용한 것은 RTF가 MS-WINDOWS 도움말 프로그램의 데이터를 생성할 수 있기 때문이다. 즉, 링크 생성의 결과로 만들어진 RTF화일은 MS-WINDOWS Help Compiler를 이용하여 컴파일되어, MS-WINDOWS 도움말 화일인 *.hlp 화일로 만들어진다. 다음은 최종 생성된 RTF화일 중 일부이다[그림 12].

```

#{\footnote # ga000079 }
${\footnote $ 가리발디 }
K{\footnote K 가리발디 }
+{\footnote + ga:00079 }

\pard{\plain\f1\b\fs48 가리발디}\par
\pard{\par
\pard{
1807 ~ 1882 {\uldb 이탈리아}{\v aa002949} 통일 운동의 지도자, 군인, 니스에서
{\uldb 선원}{\v sa001030}의 아들로 태어나 {\uldb 해군}{\v ha000366}에 들어가
복무하였다. 처음에는 공화주의자로 청년 이탈리아당 및 {\uldb 로마}{\v ra000229}의
공화당에 가담하여 {\uldb 혁명}{\v ha000636} 통일 운동을 벌이다가 여러
{\uldb 차례}{\v ca000016} 해외로 망명하였다. 1860년에 ‘부셔즈 부대’란 의용군을
조직하여 시칠리아와 {\uldb 나폴리}{\v na000097}에 원정하여 시칠리아 왕국을
정복하여 이를 비토리오 에마누엘레 2 세에게 바쳐 이탈리아 통일에 크게
이바지하였다. 이후 {\uldb 마치니}{\v ma000168} 및
{\uldb 카보우르}{\v ka000051}와 더불어 이탈리아 통일 운동의 영웅으로 온
{\uldb 국민}{\v ga002188} 으로부터 존경을 받았다.
}\par
}\page

```

[그림 12] 링크마크업의 예(노드 : “가리발디”)

그림 12에서 “\v” 뒤에 나타난 id들은 링크가 활성화되었을 때 보여줘야하는 노드의 ID이다.

4.4 검증

검증단계는 사람의 지능을 필요로하는 의미검증과 기계적으로 링크의 시점과 종점의 유효성을 검사하는 방법이 있는데, HE의 개발에 있어서는 기계적 검증방법만을 적용하였다. 기계적 검증방법은 잘못된 표제어목록으로 부터 발생하는 오류를 찾아낼 수 있다. 백과사전과 같은 대량의 자료를 다룰 때에는 기계적 검증방법만으로도 꽤많은 오류를 찾아 낼 수 있다.

V.결 론

이상에서 “백과 사전”을 하이퍼텍스트로 바꾸는 작업을 개략적으로나마 훑어보았다. 백과사전은 본질적으로 하이퍼텍스트화 하기에 좋은 점을 많이 가지고 있었다. 첫째로, 표제어별로 노드를 구별하면 되기에, 노드 구별에 관련된 복잡한 문제를 피할 수 있었고, 둘째로, 독자가 순차적으로 읽지 않을 것이라 예상하고 만든 책이기에 하이퍼텍스트가 가지는 비순차성과도 적합하다. 따라서, “백과 사전”은 의미 처리없이 기계적으로 하이퍼텍스트화 하기에 도 무척 좋은 응용이었다. 그러나, 앞으로 범용 하이퍼텍스트 변환기는 의미처리를 필요로하게 될 것이다.

HE는 독자적인 브라우저를 가지고 있지 않고, 한글 MS-WINDOWS 3.1의 “한글 도움말” 프로그램을 사용하기 때문에, 사용자가 손쉽게 접근할 수 있는 반면, 백과사전이라는 특성을 크게 반영하지 못한 면이 있다. 즉, 사용자는 책을 읽는 다는 느낌보다는 편리한 데이터베이스시스템을 사용하는 느낌을 많이 받을 것이다. 따라서, 실제로 책을 보는 듯한 책 메타포어(Metaphor)를 사용한 브라우저의 개발이 필요로된다. 또한, 의미처리를 하지 않았기 때문에 중의적인 의미를 갖는 표제어가 잘못 링크로 만들어지는 일이 많이 발생하였다. 단순한 신택스 처리만은 쉽게 한계에 부딪친다.

백과사전은 상식의 집합체라고 생각할 수 있을 만큼 범용이다. 이는 백과사전의 표제어가 무척 객관적으로 선택된 것이며, 이는 개념창고의 역할을 할 수 있을 것으로 보인다. 즉, 임의의 문서에 존재하는 표제어를 백과사전식으로 설명할 수 있는 것이다. 따라서, HE는 그 자체로써도 충분히 유용하지만, HE를 일반적인 “문서에서 하이퍼텍스트 자동 변환 시스템”에서, 개념창고로 사용할 예정이다. 이는 HE의 표제어를 일반적인 개념이라 생각하고, 설명문을 개념의 설명이라 생각하여, 일반적인 문서에 범용의 Point-to-Node 링크구조를 부여할 것이다.

참 고 문 헌

- [1] Glushko, R.J "Transforming text into hypertext for a compact disc encyclopedia", In CHI'89 Proceedings, ACM, New York, pp.293-298, May, 1989
- [2] Conklin, J. "Hypertext: An Introduction and survey", Computer, 20(9), Sept. 1987, 17-41
- [3] Raymond, D., Tompa, F. "Hypertext and the new Oxford English Dictionary", Communications of the ACM, 31(7), pp.871-879, July 1988
- [4] Hiroshi Naito, Shinji Yamashita, Youichi Matzuyama, Koichi Masegi, "Transforming Text into Hypertext for an On-line Dictionary", 情報處理學會論文誌(일본), 34(2), pp.320-330, Feb 1993
- [5] Microsoft, Microsoft Windows Version 3.1 User's Guide, Microsoft Corporation, 1992
- [6] Microsoft, Microsoft Windows 3.1 SDK Programmer's Guide, Microsoft Corporation, 1992
- [7] 계몽사 편집부, "학생 백과 사전", 총 6권, 계몽사, 1992
- [8] Goldfarb, C.F, The SGML Handbook, Oxford University Press, NewYork, 1990