

# 웹 색인작성에 관한 연구

## A Study on Web Indexing

윤 구 호(Koo-ho Yoon)\*

### < 목 차 >

1. 서 론	3.3 메타 태그 작성
2. 삽입색인	3.4 메타 태그 작성과 분석을 위한 보조도구
2.1 삽입색인의 필요성	4. 웹 색인
2.2 삽입색인 작성과정	4.1 웹 색인의 필요성
2.3 삽입색인 작성의 문제점	4.2 웹 색인의 정의
2.4 삽입색인 작성을 위한 사전 준비사항	4.3 웹 색인의 기입요소
3. 메타 태그	4.4 웹 색인작성
3.1 메타 태그와 탐색엔진	4.5 웹 색인작성의 학습
3.2 주요 메타 태그	5. 결 론

### 초 록

최초의 웹 페이지가 인터넷 상에 올려진 1991년 이래, 인터넷 상의 무수한 웹 사이트 정보검색은 전통적 색인작성 방법과는 상이한 새로운 방법을 개발하고 있다. 본고는 새롭게 발전하는 웹 색인작성에 관한 기초연구로서, 그의 기초가 되는 삽입색인을 살펴보고, 웹 사이트의 정보제공을 위해 필수적인 메타 태그와 웹 색인을 위한 그들의 작성방법을 검토하고, 웹 색인작성에 관한 주요사항을 비교적 상세하게 연구검토하였다.

주제어 : 웹 색인작성, 인터넷 색인작성, 삽입색인작성, 메타 태그.

### Abstract

Since 1991 when the first Web pages were placed on the internet, information access for numerous Web sites has developed new indexing methods which are different from traditional methods. This paper, as a basic research, deals with Web indexing(Website indexing). Embedded indexing providing basics of Web indexing is examined, and essential META tags used in Web indexing are reviewed in brief. Finally, all the important issues of Web indexing are investigated in detail.

Key Words : web Indexing, internet indexing, embedded indexing, META Tag.

\* 계명대학교 문헌정보학과 교수 (khy173@keimyung.ac.kr)

· 접수일 : 2002. 5. 17 · 최초심사일 : 2002. 6. 4 · 최종심사일 : 2002. 6. 10

## 1. 서론

정보기술(IT)의 괄목한 만한 발전과 함께 정보의 배포와 이용방법도 변화하고 있다. 인터넷은 가장 빠르게 성장하는 커뮤니케이션과 출판의 매개체이다. 인터넷을 통한 정보의 배포와 이용의 특징은 이들이 신속하고 용이하게 이루어 질 수 있고, 비용이 저렴하며, 많은 사람이 정보를 공유할 수 있으며, 배포되는 자료를 저자가 완전히 통제할 수 있다는 점이다.

인간은 최소의 노력으로 최적의 정보에 즉시 접근하기를 원한다. 정보검색의 효율성은 그의 기반이 되는 색인작성에 의존한다. 색인작성은 웹 상에서 이용할 수 있는 복잡한 정보를 통제하고 접근하는데 용이하고도 필요한 과정이다.

색인작성은 텍스트에만 국한되는 것이 아니라 이제는 이미지, 사운드, 동영상 및 소프트웨어에도 적용된다. 소프트웨어는 웹에서 충분히 이용될 수 있다 많은 메타 데이터가 웹 사이트 색인작성에 사용되어 정보검색을 돕고 있다. 그러나 아직도 많은 규모가 큰 다운로드 아카이브들(archives)이 훌륭한 색인작성시스템을 제공하지 못하고 있다. 또한 웹 출판사들도 점점증하고 있으며, 규모가 큰 웹 출판사들은 자신의 웹 페이지를 작성하여 독자들을 웹에 제한하기 위해 매우 열심히 노력하고 있다.

1991년 World Wide Web의 창시자인 Tim Berners Lee가 4개의 파일을 웹에 올려준 이래 인터넷에는 수천만 내지 수억개에 달하는 문서나 파일이 존재하고 있으며, 이들은 날마다 점점증하고, 수시로 변경되고, 사라지고, 이동하게 됨으로써 내용의 통제가 어렵게 되고, 일관성이 결여되고 있다.

본고는 새롭게 발전하는 분야의 하나인 웹 색인작성에 관한 기초연구다. 먼저 웹 색인작성이라고 종종 불리워질 만큼 웹 색인작성의 기초가 되는 삽입색인에 대해 살펴보았다. 삽입색인 작성기법은 웹 색인작성의 토대를 마련해 주기 때문이다. 다음으로, 웹 사이트에 관한 정보를 제공하기 위해 많이 사용되는 메타 태그를 검토하고, 주요 메타 태그의 작성기법을 기술하였다. 끝으로, 본고의 주 목적인 웹 색인작성에 관련된 제반사항을 비교적 면밀하게 연구검토하였다.

따라서, 본고는 아직도 유년기 상태에 있는(특히, 우리나라의 경우) 웹 색인작성에 관심을 갖는 색인자와 또한 웹 색인을 사용하고 있는 도서관 사서와 일반 이용자에게 유익하리라 사료된다. 특히 전통적인 색인작성기술과 정보분석의 능력을 가진 색인자와 사서는 온라인 정보접근도구를 창안할 수 있는 제일의 후보자이므로 웹 색인작성의 다양성과 기법을 탐구하여 훌륭한 색인작성 기술을 개발할 수 있을 것이다. 보다 희망적인 웹의 미래는 이러한 기술을 언제나 보상해 줄 것이다.

## 2. 삽입색인

삽입색인법(embedded indexing)은 색인작성과정에 새로운 수준의 복잡성을 초래한다. 이는 새로운 소프트웨어 기술과 특수한 편집/색인편집 기술을 색인작성과정에서 언제나 수반된 전통적 사고와 분석에 통합할 것을 요구한다.

삽입색인작성 소프트웨어는 일반적으로 WordPerfect나 Ventura Publisher와 같은 워드 프로세싱이나 페이지 디자인 소프트웨어에서 발견된 특성이다. 삽입색인작성 소프트웨어는 색인자에게 색인기입(또는 색인기입을 위한 태그)을 문헌텍스트 파일에 직접 삽입하도록 한다. 즉, CINDEX, Macrex, 또는 SkyIndex에 색인을 기록하는 대신에 도서를 만들기 위하여 사용되는 파일에 색인기입을 직접 삽입한다. 많은 회사들이 매뉴얼과 도서의 편집, 레이아웃 및 인쇄를 프로그램 파일로부터 직접 작성하기 위하여 Microsoft Word, PageMaker, FrameMaker 또는 Quark 등을 사용한다. 페이지를 수작업으로 붙이고 교정쇄를 만드는 일은 전자출판의 일부가 아니다. 모든 것이 기록될 때부터 출판사의 인쇄 플레이트에 들어갈 때까지 파일안에 생존한다.

색인작성 코드를 파일에 삽입할 때, 출판사는 완전히 마지막까지 어떤 페이지 번호가 그 색인에 들어갈 것인가를 염려하지 않는다. 또한 필요하면, 색인작성이 완료된 후라도 파일은 레이아웃, 콘텐츠 및 원고-정리(copy-edit)의 변경을 실현할 수 있다. 이러한 접근방식은 색인작성과 텍스트가 차기 판(edition)에서 재사용 되거나 또는 매우 활동적인(hyperactive) 링크로 전환되는 것을 가능케 한다(Mauer, 1998).

### 2.1 삽입색인의 필요성

PageMaker나 Quark, Word나 WordPerfect를 사용한다고 해도 이들에는 삽입색인작성의 약간의 보편적 특성이 있다. 마치 레이아웃 기술자가 그림과 다른 디자인 요소들을 정렬하는 것 처럼 텍스트는 마치 강(江)속에 있는 페이지 위를 “표류하면서(floats)” 주위를 돌아다닌다. 이것은 색인되는 주제의 위치가 또한 주위를 표류할 수 있음을 의미한다(Wright, 1999).

예를 들어 “고무밴드”를 논의하고 있는 부분을 색인작성하려고 하는데 현재 그것은 4페이지에 있다고 하자. 그런데 만일 고무밴드의 그림을 고무밴드가 처음 논의된 문단위에 삽입하려고 한다면 이 그림이 삽입됨으로써 4페이지에 있는 “고무밴드”의 논의 부분이 다음 페이지(5페이지)로 밀려나게 될 것이다. 전자출판에서는 이러한 유형의 텍스트 이동이 매우 일반적

이다. 결과적으로 페이지 번호들은 일반적으로 출판과정의 마지막 단계까지는 결정되지 않는다. 이것은 “페이지 파괴”(“setting the page breaks” 또는 “breaking the pages”)라고 불리우며, 출판과정이 완료될 때까지 텍스트는 주위를 이동하거나 또는 표류한다. 전통적인 색인작성의 관점에서 보면 이것은 악몽이다(Mauer, 1998).

추가된 텍스트나 재배열된 문단이 텍스트의 위치를 변경할 수 있는 것처럼 추가된 그림은 텍스트의 위치를 변경할 수 있다. 텍스트의 크기를 늘리거나 줄이거나 또는 표목형식(heading format)을 변경하는 것은 텍스트 배치를 바꿀 수 있다. 이러한 모든 변화를 통해서 코드화된 색인기입은 텍스트의 문단이나 줄(lines)에 부착되어 표류한다. 색인기입이 몇 페이지에 있는가는 문제가 되지 않는다. 색인이 편집될 때(즉, 삽입색인작성 소프트웨어에 의해 생성될 때) 색인기입에 부착된 페이지번호는 텍스트가 끝나는 위치에 부합될 것이다. 색인기입들은 이러한 모든 변화를 생존케 한다. 이는 완전히 종료될 때까지 올바른 페이지 번호에 대한 염려를 할 필요없이 출판사가 도서의 재배열이나 갱신을 할 수 있는 자유를 허락한다.

출판사들이 도서파일의 일부로서 색인기입들의 삽입에 관심을 갖는 이유는 다음과 같다(Wright, 1999).

- 색인작성은 페이지 교정쇄(page proofs)가 마무리되기 전에 시작될 수 있다. 이는 보다 빠른 “출판시간”을 의미한다. 어떤 출판사들은 이것이 삽입을 위한 첫 번째 이유라고 여긴다. 장(chapters)들은 그들이 도서에 나타난 순서에 관계없이, 각각의 장이 완료될 때 색인작성을 위해 보내질 수 있다.
- 출판에 앞선 교정이나 또는 레이아웃과 비교적 중요하지 않은 텍스트 조정에서 변경된 신판(new editions)들은 색인의 재작성(re-indexing)을 요구하지 않을 것이다. 기존의 색인작성에 대한 비교적 빠른 검토로서 충분할 것이다. 색인이 페이지번호에 매이지 않기 때문에 페이지 파괴(page breaks)의 위치 변경은 문제가 되지 않는다. 단순히 색인을 재편집하여 이러한 변화를 반영하면 된다.
- 파일안에 생존하는 색인작성은 다른 형식에서도 사용될 수 있다. 예를 들면 이들은 독자들이 올바른 페이지로 이동할 수 있게 하는 온라인 인터랙티브 색인으로 전환될 수 있다.
- 번역기들은 페이지 번호의 갱신을 염려하지 않고 단지 색인기입을 번역하면 된다.

## 2.2 삽입색인 작성과정

삽입색인 작성과정은 전통적인 색인작성과는 다르다. 삽입색인 작성과정을 요약하면 다음과 같다(Mulvany, 1994).

- 파일복사 : 파일들을 받자마자 모든 파일의 백업파일을 만든다. 이것은 매우 중요한 단계이다. 당신은 항상 원본파일에 대한 깨끗한 복사를 보유하여야 한다.
  - 기입(엔트리)의 고안 : 이 부분은 모든 단계에서 발생할 수 있다. 물론 이상적으로는 색인작성은 도서의 처음에서 시작하여 마지막에 끝나야 하지만, 전자출판에서는 항상 그런 것은 아니다. 당신은 한번에 단지 하나의 파일 상에서 작업할 수 있기 때문에 이 단계 동안 필요한 대로 당신은 고객과 파일들을 교환할 수 있다. 각각의 DTP 소프트웨어 패키지는 코드화된 기입들을 다르게 만드는 과정을 취급하므로 당신은 상호참조와 같은 항목들과 특정한 분류(sorting)를 정확하게 취급하는지 확인하기 위한 약간의 테스트를 해야 한다.
  - 색인의 편찬 : 색인을 편집하기 위하여 당신은 색인을 포맷된 버전에 편찬할 필요가 있다. 대부분의 DTP 패키지들은, 전용 색인작성 소프트웨어와는 달리, 당신이 색인을 구축하는 동안에는 색인을 디스플레이하지 않는다. 당신은 이 단계를 위하여 모든 파일을 가질 필요가 있을 것이다. 색인을 보고, 편집하고, 작업하기 위하여 편찬된 색인을 인쇄하여야 한다.
  - 편집 : 색인은 일관성과 유효범위를 위해 편집된다. 이 일은 도서의 권말색인작성(stand-alone indexing)에서는 매우 간단하지만, 삽입색인작성에서는 매우 어렵다. 왜냐하면, 대부분의 삽입색인들은 단지 속사(速寫 : snap-shots)에 불과하기 때문이다. 당신은 편찬된 색인자체에는 변경을 만들 수 없고, 원본파일로 되돌아 가서 페이지와 기입을 찾아 코드자체에 변경을 만들어야 한다.
- 인쇄출력상의 최초 편집은 싫증나는 작업이다. 인쇄출력을 통해 무엇을 변경할 것인가에 대한 주기를 만드는 작업을 하라. 그리고 그 파일들로 되돌아가서 원래의 코드들을 변경하라.

## 2.3 삽입색인 작성의 문제점

만일 당신이 삽입색인작성을 시도하려면 각 소프트웨어 패키지의 독특한 삽입기술에 정통할 필요성이 있다. 삽입된 코드들을 사용하여 색인을 작성하는 일은 색인작성 전문 소프트웨어에서의 작성보다 더 쉬우면서도 더 어렵다. 자동 페이지 번호 매김은 당신이 모든 페이지 번호를 정확히 획득하였는가에 대하여 걱정할 필요를 없애주며, 또한 레이아웃 후의 색인 변경을 매우 용이하게 갱신할 수 있음을 의미한다.

대부분의 DTP 소프트웨어 패키지들은 밀폐된 시스템(closed system)이다. 즉, 서로다른 소프트웨어 패키지를 사용하여 개발된 파일들을 교환하기가 어렵다. 그리고 전자출판 과정도 근본적으로 밀폐된 과정이다. 색인자는 일반적으로 제작자가 최종 편집과 인쇄를 위해 조정

할 필요가 있는 파일에서 작업해야 한다. 이것은 일반적으로 권말색인작성에서는 일어나지 않는 시간적 긴장을 유발한다.

당신의 컴퓨터가 고객이 사용하는 기계의 컴퓨터구성과 대등한지 확인할 필요가 있다. 소프트웨어 버전이 동일해야 하며; 레이아웃에 영향을 미치는 특수한 모듬명령(macros), 보기판(templates), 또는 추가기기(add-ons)가 비치되어야 하며; 가장 결정적으로 프린터 드라이버와 폰트 기구가 대등하여야 한다. 그렇지 않으면 폰트 교환이 색인작성을 망가뜨리기에 충분한 페이지 변경을 만들 수 있다(Malik, 1994).

페이지 범위의 설정은 삽입색인작성의 주요한 취약점의 하나다. 대부분의 삽입 소프트웨어 패키지들은 색인 모듈을 디스플레이할 때 스크린을 동굴게하여(hog) 당신이 페이지의 어느 부분에 있는지 그리고 페이지 범위가 어디에서 끝나는지의 짐작을 어렵게 한다(Mauer, 1998). 전통적 색인작성에서는 만일 텍스트가 한 페이지 이상 계속되면 단지 페이지 범위를 포함시킨다. 삽입색인작성에서는 만일 텍스트가 다른 페이지로 표류하면, 당신은 하나의 페이지 번호나 또는 페이지 범위를 사용할 것인가를 결정해야 한다. 페이지 범위의 설정은 복잡한 과정이며, 때때로 문단의 계수나 또는 강조, 다중코드화된 기입들을 요구한다(Moncrief, 1999).

삽입색인은 권말색인보다 작성에 시간이 더 오래 걸린다. 각 소프트웨어 패키지는 자신의 요구사항을 당신의 컴퓨터 메모리에 저장하여 기계를 신속하게 사용하도록 한다. 당신의 일일 평균 색인할 페이지의 목표는 특별히 당신이 복잡한 부분을 색인할 때에는 아마도 달성하기 어렵게 될 것이다. 편집은 여전히 전체시간의 3분의 1을 요구할 것이며, 만일 당신이 원본 텍스트로 되돌아가서 다중코드화된 기입들을 편집해야 한다면 더 오래 걸릴 수 있다. 따라서 충분한 시간을 허용하여야 한다.

또한 당신이 문헌전체에서 코드를 정정할 필요가 있다면 당신은 오랜 시간에 걸쳐 모든 파일을 접근하여야 한다. 만일 당신이 파일들을 기다린다면 시간을 낭비하게 된다. 편집과정 동안에는 거의 모든 기입이 약간씩 변경된다. 당신의 고객은 이를 이해하지 못하고 당신에게 모든 파일을 한번에 주려고 하지 않을 수 있으므로 이 부분의 과정에 대해 그들을 교육하여야 한다(Bonura, 1994).

만일 당신이 색인작성을 완료한 후에 누구든지 그 파일과 작업한다면 그들이 색인태그를 삭제하지 않도록 하여야 한다. 이미 코드된 논의의 중간에 첨가된 텍스트는 페이지 범위를 못쓰게 할 수 있으므로 색인에 대한 그의 영향에 대해 주의깊게 체크하여야 한다. 당신만이 색인에 발생할 수 있는 것을 이해하는 사람이므로 당신은 그것을 검토할 기회를 가져야 한다.

대부분의 삽입색인작성 프로그램들은 당신의 기입에 필요할지도 모를 소형 대문자들(small caps)이나 또는 이태릭체의 텍스트를 취급하지 못한다. 문제있는 기입의 끝에 일련의 별표를 추가하는 것은 색인작성과정의 최종단계에서 만들어지는 편집된 색인텍스트 파일안에서 특별한 요구를 가지고 있는 기입들을 탐색하는 것을 당신에게 허락할 것이다.

출판사들은 “보라”와 “도보라” 참조의 포맷팅에 관해서 크게 다르기 때문에 당신은 고객들이 상호참조를 어떻게 포맷하기를 원하는가를 알아야 한다. 당신은 참조들이 부표목 리스트의 하부에 적절하게 나타나는지 확인하기 위해 당신의 코드화된 기입에 “see also” 대신에 “zsee also”를 삽입하는 것과 같이 어떤 기교를 적용할 필요가 있을지 모른다. “z”는 색인텍스트에서 쉽게 제거될 수 있다(Wright, 1999).

## 2.4 삽입색인작성을 위한 사전 준비사항

삽입색인작성을 시작하기 위해서는 당신은 어떤 소프트웨어 패키지를 배워야 하고 또 더욱 중요하기는 어떤 소프트웨어 패키지를 구입해야 할 것인지가 매우 중요하다. 당신은 이미 기계에 당신의 최초 패키지를 갖고 있을 수 있다. 당신이 색인작성 모듈을 가지고 있는 Microsoft Word나 혹은 다른 워드 프로세서를 갖고 있는지 확인하라. 워드프로세서들은 단지 기록을 위해서가 아니라 단편(부분)들을 인쇄에 보내기 위하여 전문서적들의 출판에 많이 이용된다. 온라인 도움말과 색인작성 모듈과의 경험으로부터 색인작성 모듈을 어떻게 사용하는가를 배울 수 있다. 당신은 유사한 항목들의 기입을 촉진하기 위한 복사/붙이기(copy/paste) 특성을 사용하거나 또는 종종 사용된 용어들이나 복잡한 두문자어들을 위한 자동-텍스트 “모드 명령들”을 고안함으로써 약간의 스피드 향상 기교(tricks)를 또한 개발할 수 있다(Maislin, 1999).

당신이 보다 많은 고객에게 서비스를 제공할 수 있도록 DTP패키지를 구입해야 하는가는 패키지들이 매우 비싸기 때문에 어려운 질문이다. PageMaker나 FrameMaker의 가격은 500달러 이상이므로 대부분의 고객은 당신을 위해 이러한 패키지들을 구입하지 않을 것이다. 각각의 갱신도 또한 비용이 들며, 상이한 고객들은 종종 상이한 또는 보다 오래된 버전들을 사용한다.

이러한 패키지들은 또한 대형 컴퓨터를 요구한다. 당신은 상당한 처리능력을 요구하는 대규모 파일들과 작업하게 된다. 당신의 색인작성 과정은 편집을 위해 많은 처리능력을 요구하면서 통례적으로 모든 파일을 가동한다. 당신은 동일한 폰트와 동일한 프린터 드라이버를 사용하여 당신의 고객과 동일한 운영시스템을 가동하여야 한다. 이것은 모든 요소들이 정확하게 함께 작업하고 있는가를 확인하기 위한 당신의 시간적 투자를 요구한다. 인터넷상에서 파일들을 보내거나 돌려받는 즐거운 부분, 즉 문제를 일으킬 수 있는 과정을 망각하지 마라.

보다 값비싼 패키지들과의 경험을 위한 한가지 방법은 지역사회 대학의 컴퓨터 실험실에서 강좌를 택하는 것이다. PageMaker와 FrameMaker 수업은 전국적으로 행해지고 있으며, 실습시간은 어떻게 색인작성을 하며, 당신이 도구를 배울것인지 또는 구입할 것인지를 알아

보기 위해 사용될 수 있다. 이것은 비용이 들지 않는 맛보기 방법이며, 부수적 이점으로 당신이 고객의 작업을 이해할 수 있도록 도와주는 레이아웃을 배울 수 있다 (Wright, 1999).

가장 값이 싼 패키지 - 당신의 워드프로세싱 패키지 - 를 처음 배워라. 만일 어떤 프로젝트가 보다 값비싼 패키지를 요구하게 되면 당신이 보상받을 수 있을 것이라고 확인한 후 과감히 진행하라.

### 3. 메타 태그

메타 태그들은 웹사이트에서 메타-정보나 또는 정보에 대한 정보를 만들기 위하여 사용된다. 많은 유형의 메타 태그들이 있으나 색인작성에 가장 적합한 것들은 디스크립션 태그와 키워드 태그이다. 디스크립션 태그는 탐색엔진이 탐색결과를 리스트 할 때 종종 디스플레이 되는 사이트 콘텐츠의 간략한 요약물을 제공한다. 키워드 태그는 탐색엔진을 사용하는 사람이 적합한 사이트를 찾기 위하여 사용할 수 있는 단어나 단어구를 규정하기 위하여 사용된다.

메타 태그는 두 가지 이유 때문에 색인자에게 관심사가 된다 (Rowland, 1999). 그들은 색인작성 서비스를 위해 웹을 탐색하는 사람에게 당신의 웹사이트 색인작성 업무를 더욱 잘 보이게 만드는 수단이며, 색인자에게 업무의 잠재적인 새로운 소스(웹사이트 개발자와 웹사이트를 갖고 있는 회사를 위해 키워드 태그와 디스크립션 태그를 기록하는 일)를 제공한다. 메타 태그 기록은 적합한 키워드를 선택하기 위해 사서의 능력을 훌륭히 활용하도록 하며, 밀접하게 관련된 초록작성의 숨씨(문헌의 핵심을 한 두 문장으로 전달하는 기술)를 마련해 준다.

#### 3.1 메타 태그와 탐색엔진

탐색엔진은 웹 페이지에 대한 색인된 참조를 작성함으로써 작업할 수 있다. 이 참조들은 페이지의 제목, URL, 요약 및 키워드 리스트를 포함한다. 상이한 탐색엔진들은 요약정보와 키워드들을 다르게 선택하고, 일반적으로 페이지의 메타 태그로부터의 정보를 페이지로부터의 다른 정보와 결합한다. 당신은 메타 태그를 남용하거나 부정확하게 사용할 수 있어 탐색엔진이 당신의 사이트를 전부 무시하는 결과를 초래할 수도 있다. 어떤 웹사이트 설계자는 사람들을 사이트로 유인하기 위하여 관련이 없지만 통속적인 용어들의 키워드를 사용하여



탐색 엔진을 치장하려고 한다(경쟁자의 브랜드 명칭과 섹스-관련 용어들이 이 범주에 속한다). 이러한 오용의 잠재력 때문에 어떤 탐색엔진은 메타 태그에 한정된 키워드들에 어떠한 가중치도 주지 않고, 대신에 페이지의 본문에서 텍스트를 색인한다(Dan's Web Tips).

그럼에도 메타 태그의 사용은 약간의 이점이 있으며, 일반적으로 이러한 태그들을 만들기 위한 적은 노력은 가치가 있다. 탐색결과에서 사이트의 순위를 높이기 위하여 메타 태그를 사용하는 탐색엔진에는 HotBot와 Infoseek가 있다. Alta Vista, Excite 및 WebCrawler와 같은 다른 탐색엔진들은 메타 태그를 지원하지만, 그러나 반드시 당신의 사이트 순위를 보다 높게 하지는 않는다.

Drott(2002)는 Fortune지가 선정한 지구촌 500 개 회사로부터 60 개 웹사이트를 선정하여 그들이 자동색인작성을 위한 지원도구를 제공하는지를 2000년과 2001년에 조사하였다. 특히 "키워드"와 "디스크립션"을 위한 robot.txt와 메타 태그의 사용을 조사하였는데, 대략 절반에 못 미치는 사이트들이 이들 두 가지 지원 도구중 한가지 또는 두 가지 모두를 사용하였다. 거의 80%의 사이트들이 Java를 사용하였으며, 약 3분의 1 정도의 사이트들이 cookie들을 사용하였다. 색인작성 지원 도구, 특히 메타 태그, 사용의 증가는 웹 로봇들이 사이트들을 보다 빠르게 색인할 수 있어 웹의 전체적 색인 범위를 향상시키는 하나의 방법임을 나타낸다.

Browne과 Jerney(2001)는 웹사이트를 위한 도서 권말색인 유형의 웹 색인작성을 웹 상의 정보접근을 위한 다른 방법들과 비교 검토하였는데, 도서 권말색인 유형이 목차 테이블, 사이트 지도, 포털(portals)과 탐색엔진과 같은 전통적 방법들보다 많은 이 점을 제공한다고 하였다.

### 3.2 주요 메타 태그

대부분의 HTML 태그와는 달리, 메타 태그는 페이지에 있는 정보를 포맷하기 위한 것이 아니라 웹사이트에 관한 정보를 제공하기 위해 사용된다. 그들은 다양한 목적을 위해 브라우저와 탐색엔진에 의해 사용되는, 디스플레이 되지 않거나 또는 숨겨진, 태그들이다. 논의하려고 하는 키워드 태그와 디스크립션 태그는 많은 태그들 중의 일부이다.

메타 키워드 태그는 당신의 웹페이지를 서술하는 키워드들을 포함한다. 메타 디스크립션 태그는 당신의 사이트나 서비스의 서술을 포함한다. 당신의 색인작성 슝씨는 이들 안에 들어간다. 당신은 키워드들을 선택하고 간결한 서술을 기록한다. 당신이 집착해야 하는 어휘들의 통제 리스트는 없다. 키워드를 기록하는 과정은 모든 단어가 특정한 순서가 없이, 그리고 한정된 부기입이 없이 키워드 태그에 연결되는 것을 제외하고는 통상적인 색인작성과정과 유사하다.

전통적인 색인작성에서와 같이 당신은 항상 이용자를 염두에 두어야 한다. 즉, 이용자가 당신의 사이트에 있는 정보를 찾기 위해 어떠한 키워드들이나 또는 키워드들의 조합을 선택할 것인가? 숙련된 웹 열성자로서 당신은 어떠한 용어들을 당신의 마음에 드는 검색엔진에 타이프 할 것인가? 웹이 싫은(Web-phobic) 이웃집 사람이, 만일 정보를 찾을 필요가 있다면 어떠한 용어들을 사용할 것인가? 등을 고려하여야 한다(Metatags).

당신은 키워드들의 이형(異形), 영국식 철자법, 틀린 철자들을 포함하고자 할 것이다. 예를 들면, 키워드가 "archaeology"라면 당신은 또한 "archeology"를, "millennium"이라면 또한 "millenium"과 "milennium"을 포함시킬 수 있다

단지 하나의 지침은 키워드들은 800-1000자 이상을 사용하지 않아야 하며, 몇 번이고 반복되어서는 안 된다는 것이다. 어떤 검색엔진은 당신의 사이트를 전혀 리스트하지 않음으로써 반복되는 키워드들에 대해 당신을 벌할 것이다. 하나의 실용적 방법은 키워드를 그 자체로서 또는 다른 단어와의 조합으로 일곱 번 이상 사용해서는 안 된다는 것이다. 예를 들면, 색인자는 "indexing, indexer, medical indexing, embedded indexing, scientific indexing, book indexing, back-of-the-book indexing"을 사용할 수 있지만 더 많은 "indexing"의 사용은 검색엔진으로 하여금 그 사이트를 무시하게 할 수 있다.

메타 디스크립션 태그는 당신에게 사이트의 특징을 명확하고 간결하게 서술하도록 허용한다. 당신의 초록작성과 요약작성 슝씨는 이곳에서 사용될 것이다. 이 태그는 당신의 페이지에 대한 정확한 서술을 검색엔진에게 제공하기 때문에 유용하다. 검색엔진은 검색결과 페이지 위에 당신의 사이트를 가져올 때 당신의 서술을 사용 할 수 있다. 디스크립션 태그가 없어도 검색엔진은 자신의 서술을 기록할 수 있거나 또는 당신의 웹 페이지로부터 최초의 하나 내지 두 개의 문장을 단순히 사용할 수 있다. 당신의 디스크립션 태그는 200자 이하의 문자로 되어야 한다(META Tag Resources).

### 3.3 메타 태그 작성

만일 당신이 기본적인 HTML을 잘 알면 메타 태그를 위한 코딩은 비교적 간단하다. 메타 태그들은 페이지의 <HEAD>부분 안에 배치된다. 그들은 NAME과 CONTENT라고 하는 두 가지 속성을 갖는다. NAME 속성은 당신이 제공하는 메타정보의 유형(당신이 사용하는 메타 태그의 유형)을 기술하기 위해 사용된다. CONTENT 태그는 실제 정보(이 경우에는 키워드들과 디스크립션)를 제공하기 위해 사용된다. 메타 태그들을 만들기 위하여 당신은 단지 아래와 같은 코드를 빈자리에 채워 넣으면 된다(<TITLE>태그는 메타 태그가 아니지만 그들이 또한 페이지의 <HEAD>부분에 포함되는 태그의 하나이기 때문에 여기에 포함되며

<TITLE>태그의 콘텐츠도 또한 탐색엔진에 의해 사용된다.

```
<HEAD>
<TITLE>"_____ "<TITLE>
<META NAME="keywords" CONTENT="_____ ">
<META NAME="description" CONTENT="_____ ">
</HEAD>
```

사이트의 제목, 사이트를 기술하는 키워드들의 리스트와 사이트의 디스크립션을 <HEAD>에 삽입하라. 다음은 하나의 예시이다.

```
<HEAD>
<TITLE>"웹 색인작성에 관한 연구(A Study on Web Indexing)"<TITLE>
<META NAME="Keywords" CONTENT="Web Indexing, Internet Indexing, Embedded Indexing, META Tag, 웹 색인작성, 인터넷 색인작성, 삽입색인작성, 메타 태그">
<META NAME="description" CONTENT=" 웹 색인작성에 관한 연구는 웹 색인작성과 관련된 삽입색인 및 메타태그에 관한 연구와 웹 색인의 필요성, 정의, 기입요소, 색인작성 및 웹 색인작성의 학습 등을 다루고 있다.">
</HEAD>
```

키워드들은 또한 <TITLE> 태그에 포함될 수 있다. <TITLE>태그에 당신의 회사명칭을 넣는 대신 당신의 사업 선전문구나 또는 소수의 키워드를 포함시켜라.

당신의 홈페이지와 최상위 수준 페이지들을 위해 메타 태그를 만들어라. 만일 당신의 사이트가 상당히 작거나 또는 페이지들이 독특한 콘텐츠를 포함하면 모든 페이지를 포함시켜라. 모든 페이지를 위해 동일한 키워드들과 디스크립션을 사용하는 것보다는 오히려 각각의 페이지에 대한 당신의 키워드들과 디스크립션을 만들어라.

### 3.4 메타 태그 작성과 분석을 위한 보조도구

메타 태그를 작성하고 분석하는 것을 허용하는 많은 온라인과 다운로드 할 수 있는 도구들이 있다. 메타 태그들은 작성하기가 매우 용이하기 때문에 메타 태그 생성프로그램에의 투자는 실제로 필요하지 않지만, 그러나 이 프로그램들은 신속하고, 사용하기에 재미있고, 그리고 종종 무료다. 이들은 당신의 유료 웹 왕래를 증대시키기 위해 다른 방법들을 제공하는 회사의 선전용 방책으로서 종종 주어진다. 당신은 자신의 키워드와 디스크립션을 삽입하여야 하지만, 그러나 도구들이 당신을 위해 모든 귀찮은 HTML 태그들을 기록한다. 그들은 일반

적으로 그들의 제품에 대한 선전용 메시지를 포함하지만, 그러나 당신은 당신의 페이지에 코드를 삽입하기 전에 이들 광고를 삭제할 수 있다(Meta Tag Builder).

당신은 수많은 온라인 태그 생성 프로그램을 발견할 수 있을 것인데, 이들은 당신의 결과를 웹사이트 상에 즉시 보여주거나 또는 e메일로 보내줄 것이다. 당신이 해야할 모든 일은 광고메시지를 당신이 원하는데로 제거하거나 또는 제거하지 않으면서 당신의 웹 페이지에 HTML코드를 강조하고, 복사하고, 붙이는 것이다.

또한 당신이 기록한 키워드들과 디스크립션들을 분석하여 당신의 메타 태그를 어떻게 개선할 것인가에 관한 힌트를 주는 프로그램들이 있다. 또 다른 프로그램들은 당신을 위해 당신의 키워드들과 디스크립션들을 실제로 기록해 줄 것이다. 이러한 프로그램들은, 약간은 무료 또는 유료 선전용 제품으로서, 웹으로부터 쉽게 다운로드 할 수 있다(Meta Medic).

## 4. 웹 색인

### 4.1 웹 색인의 필요성

인간은 정보경제에서 매우 중요한 역할을 수행한다. 보다 빠른 pc, 보다 큰 스크린, 혹은 보다 신속한 모델들은 스스로는 생산성을 높이지 못한다. 정보를 찾기 위한 탐색기술이나 또는 탐색전문가의 고용은 과거 어느때보다 더욱 중요하다. 미디어가 변화되고, 사람들이 지구촌 어느 곳에 살고 있든지 집에서 매우 많은 정보를 접근할 수 있기 때문에 상이한 국가에서 상이한 언어를 사용하는 사람들 간에 정보를 공유할 필요성이 있게 된다.

만일 당신이 웹 색인의 구축방법을 안다면 당신은 세계 도처에 있는 고객에 접근할 수 있고, 또한 고객에게 봉사 할 수 있다. 대량의 정보가 다양한 형태로 정보의 배포를 위해 매일 “출판”된다. 저자들은 웹에 책을 적재하고(loading); 연주자들은 오디오와 비디오 클립(clips)을 적재하고; 10대들은 게임을 적재하고; 데이터베이스 컨설턴트들은 지식베이스를 적재하고; 신문과 잡지 편집자들은 기사와 인용(발췌)을 적재하고 있으며, 그리고 인터넷 제공의 e-zines, e메일 뉴스레터 및 웹 뉴스 사이트들은 끊임없이 증가하고 있다.

웹 색인작업은 여러가지 일을 의미하게 되었다. 이들은 처음에는 웹 사이트들에 대한 주제-관련 링크들의 가상도서관이었다. 나중에는 사람들이 웹을 통해 접근할 수 있는 온라인 도서와 서지들을 올려주기 시작하였다. 이는 곧 데이터베이스-구동(database-driven) 웹 색인이 되었다. 그리고 개별 웹 사이트들이 보다 크게 성장함으로써 사람들은 사이트의 내적 콘텐츠에 대한 웹 사이트 색인을 만들 필요가 있게 되었다(Walker, 1999).

## 4.2 웹 색인의 정의

웹 사이트에 대한 색인들은 인쇄된 색인들과 공유하는 특성이 있다. 모든 색인의 목적은 Wheatley(1902)가 말한것처럼 “도서에 분리되어 있는 유사한 주제에 대한 모든 항목(items)을 모으는 것이다.” 색인은 본질적으로 리스트이다. 그러나 우리 모두가 아는 것처럼 여기에는 많은 유형의 리스트가 있다. 예를 들어, Merriam Webster Dictionary(1993)에 의하면 서지는 “특정 주제, 기간 또는 저자에 관련된 저술들의 설명적이거나 비평적인 주석을 종종 갖는 리스트이다.” 목록은 상세한 기술사항을 갖는 체계적으로 정렬된 항목들의 리스트이다. 전통적으로 색인은 특정한 문헌의 내용을 분석하여 생기는 고유명사, 이름 및 토픽들을 자모순으로 리스트한 것이다.

웹에서의 용어 “색인”은 보다 광범위하고 보다 일반적인 의미로 발전하였다. 종종 “웹 색인”은 근본적으로 이용자를 다른 웹사이트로 안내하기 위해 설계된 URL의 집합으로 간주된다. 이러한 웹 사이트들은 수작업으로 또는 탐색엔진이나 스파이더(spiders)에 의해 수집될 수 있다. 여기에는 일반적으로 웹사이트 자신들에 대한 내용분석은 거의 없다. 이러한 유형의 리스트는 전통적인 색인이라기 보다는 목록이라고 볼 수 있다. 그러나 주의깊은 선정과 주석의 면에서 보면 더욱 많은 것을 갖고 있는 목록이라고 할 수 있다 (Walker, 1996)

비록 색인 내에 정보를 조직하는 수 많은 방법이 있지만, 웹 색인은 통상 자모순으로 배열된다. 다른 방법(예를 들어, 시대나 또는 범주에 의한 방법)으로 정보를 조직하는 것은 정보를 분류하는 것이다. 그러나 Wheatley(1902)의 진술처럼 “우리는 자모순 색인에는 가능한 한 분류가 거의 없기를 원한다.”

## 4.3 웹 색인의 기입요소

인쇄색인과 같이 웹색인에서의 기입은 최소한 두 가지의 요소, 즉 표제어(headword 또는 head entry)와 소재표시자(locator)를 항상 포함한다. 인쇄색인에서의 locator는 일반적으로 페이지번호인 반면에, 웹색인에서의 그것은 파일명칭 과/또는 앵커명칭이다. 앵커는 <A>와 </A> HTML태그를 사용하여 배치된다. 앵커는 최초의 <A>코드에 뒤따라오는 두 가지 유형의 속성(attributes)을 사용하는 내부 및 외부 문헌들에 대한 링크이다. 첫번째의 HREF 속성(헤더참조)은 이용자를 색인내의 기입으로부터 이용자가 찾는 정보의 위치로 이용자를 지시하는 두번째의 NAME속성(locator참조)으로 안내한다. 색인자는 언제나 이 두가지의 속성을 만들어야 하며 그들을 식별하는 코드를 삽입하여야 한다. 색인을 포함하는 파일은 HREF 속성으로 구성될 것이다. NAME 속성은 색인되는 다양한 문헌을 통해 색인자에 의해 배치될

것이다(Broccoli, 1999).

웹 색인기입을 위한 코딩의 예를 보면 다음과 같다.

윤구호 <A HREF="yoon.html#yoonA">1</A>

위의 웹 색인기입에서 HREF속성 "yoon.html"는 기입으로 참조시키는 웹 페이지 명칭이다. "#yoonA"는 앵커가 삽입된 웹 페이지 안에 있는 포인트를 식별한다. 숫자 "1"은 locator이다. 웹 페이지 상에서 locator는 윤구호에 관한 정보에 대한 링크이다

아래의 코드는 yoon.html 페이지 상의 윤구호 정보가 앵커를 만들기 위한 NAME속성으로 어떻게 식별되는가를 나타낸다.

<P><B><A NAME="yoonA">윤구호</A></B><I>A Study on Web Indexing  
</I>서울 : 한국도서관정보학회, 2002</P>

이 예시에서 나타난 다른 태그들은 패러그래프를 드러내기 위한 <P></P> 태그들, 굵은 활자 텍스트(bold text)를 만들기 위한 <B></B> 태그들, 그리고 이탤릭체의 텍스트를 만들기 위한 <I></I>태그들이다.

가장 단순한 웹 색인은 자모순으로 리스트 되고 HTML로 코드화된 이러한 기입들의 집합이다. HTML은 프로그래밍 언어가 아닌 포매팅 언어다. 이는 초기의 워드 프로세싱 프로그램과 유사하다. 웹 색인을 위한 HTML 코딩의 예시는 다음과 같다

웹색인, <A HREF = "dictionary.html#webindex"></A><BR>  
윤구호, <A HREF = "yoon.html#yoonA"></A><BR>  
한국도서관정보학회, <A HREF = "kliss.html#lisinfo"></A><BR>  
이 예시에서 <BR>태그는 행(줄) 바꾸기를 위해 사용된다

#### 4.4 웹 색인작성

웹은 도서의 권말색인처럼 정보에 대한 신속하고 용이한 접근을 허용한다. 초기의 웹에는 소량의 진보된 색인과 소수의 간단한 색인이 있었으나 시간이 경과하면서 많은 복잡한 웹 데이터베이스와 주석달린 색인이 증가하였다. 일반적으로 기입들은 링크들의 주석달린 리스트였다. 오늘날에는 많은 색인자가 정보를 분류하기 위해 데이터베이스와 메타데이터를 사용한다. 한편, 고객은 무엇이 리스트되어야 하며, 웹사이트가 어떻게 디자인되어야 하는가를 구술한다.

웹 사이트 색인작성은 소심한 사람을 위한 직무(job)가 아니다(Rowland, 1999). 웹 색인작성을 시도하려는 사람은 이미 숙달된 색인자가 되어야 하며, 웹 페이지 작성의 철저한 경험을 가져야 한다. HTML에 대한 지식은 필수적이다. 사실, 웹 색인작성 업무의 특성에 따라, 당신은 웹 색인작성이 색인작성 기술보다 더 많은 HTML 지식을 요구한다는 것을 알 수 있

을 것이다. 삽입색인작성처럼 웹사이트 색인작성은 전통적인 색인작성보다 훨씬 많은 일을 요구한다 : 용어의 선택이 강요되고, 그리고 색인작성과정은 보다 느리고, 보다 도전적이다. 편집과정이 지루하고 시간이 걸릴 수 있기 때문에 기입들은 조심스럽게 선택되어야 한다. 하이퍼텍스트 링크들과 앵커들이 추가되어야 하며, 또한 링크들이 의도한대로 일하는지 확인하기 위하여 색인자체가 테스트 되어야 한다.

### 1) 웹 사이트 색인작성의 도전

웹 사이트 색인작성은 약간의 도전을 제시한다(켄들딩, 1997).

첫째는 색인될 정보의 적시성(timeliness)이다. 모든 정보는 저장수명(shelf life)을 갖고 있다. 어떤 정보는 다른 정보보다 긴 저장수명을 갖는다. 만일 웹사이트 상의 정보가 종종 갱신된다면 색인을 최신성있게 유지하는 것은 어려울 것이다.

둘째는 다운로드 시간이다. 인쇄색인에서는 이용가능한 제한된 페이지 수 때문에 길이가 종종 고려된다. 웹 색인에서는 길이는 이용자가 색인을 다운로드하여 그것을 사용하기 위해 요구되는 시간으로 측정된다. 만일 색인이 다운로드하기에 너무 오래 걸리면 사용되지 않을 것이다. 온라인 이용자들은 참을성이 없기로 평판이 나 있다.

셋째는 문헌자체의 특징이다. 도서들은 선형(linear)인 반면에 전자문헌들은 비선형이다. 예를 들면, 전자문헌들은 계층적일 수 있다. 정보는 사이트의 도처에 산재될 수 있다. 이러한 경우 색인은 모든 정보를 함께 모으기 위하여 중요하다. 색인은 사이트에 대한 게이트웨이가 된다. 웹 색인에는 페이지 범위가 없으므로 어떤 주제의 광범위한 취급과 조잡한 취급을 구별하기가 어렵다.

넷째는 다양한 참조이다. 인쇄문헌에서는 이들의 제시가 용이하지만, 온라인에서는 훨씬 어렵다. 몇 가지의 선택이 있다. 표목이나 부표목들은 그들이 핫 링크(hot link)들임을 가리키는 밑줄긋기로 반복될 수 있다. 다른 방법은 표목이나 부표목 뒤에 빈도수를 보이는 숫자나 문자를 갖는 것이다. 그러나, 만일 스크린이 줄들로 꽉 채워지면, 색인은 시각적으로 매력적이지 못할 것이며 성급한 이용자는 스크린을 통해 정사(精査) 할 시간을 취하지 않을 것이다. 비록 스크린 기술이 빠르게 향상되고 있지만 많은 세목을 포함하는 웹 색인들은 브라우즈하고 이용하기에 매우 어렵다. 이것이 패러그래프 양식의 색인이나 다수의 부기입들을 갖는 색인의 발견이 희박한 하나의 이유이다.

물론, 이러한 도전들을 극복하는 동안에 성공적인 웹색인은 이용자가 얼마나 빠르게 색인 페이지를 왕래하여 원하는 정보를 찾을 수 있는가로 측정된다는 사실을 당신은 기억해야 한다. 보다 훌륭한 색인은 이용자가 색인에서 정보를 찾아 문헌에 접근하는 것을 보다 빠르게 해준다.

## 2) 웹 사이트 색인작성의 다양성

다양한 웹사이트와 고객이 있는 것처럼 다양한 형태의 사이트-기반 웹 색인작성이 있다. 그 유형을 요약해 보면 다음과 같다(Rowland, 1999).

- 당신이 색인작성하려는 사이트의 유형 : 예를 들면, 온라인 출판물, 광범위한 토픽들에 대한 수 많은 연구보고서의 결과를 저장하는 정보-팩 사이트; 특수 토픽에 대한 규모가 적은 일반적 정보사이트; 또는 콘텐츠가 빠르게 변경되는 역동적 사이트 등.
- 당신의 업무 : 당신이 웹사이트를 작성해야 하는지; 혹은 당신이 웹사이트 운영에 포함되는지; 또는 당신이 특별히 웹 색인작성 업무를 위한 웹마스터 팀의 일원으로 고용되는지 등.
- 프로젝트의 성격 : 사이트의 보다 적은 수정을 요구하는 일시적 색인작성 프로젝트인지; 본질적으로 정규적 수정을 요구하는 사이트인지(예를 들면, 온라인 뉴스레터); 또는 날마다 변경되는 사이트인지 등.
- 색인의 성격 : 색인이 단순한 항해 보조물로서 시도되는지; 혹은 상세한 연구보조물로서 시도되는지; 또는 사이트에 제공되는 메뉴가 탐색도구와 같은 다른 도구인지 등.

## 3) 웹 사이트 색인작성 과정

웹사이트 색인작성은 색인대상물의 유형에 따라 다르지만, 일반적으로 공통적인 작성과정을 뉴스레터를 예로 들어 기술해 본다(Rowland, 1999).

- ① 색인용어의 선정 : 웹 색인작성 과정의 첫 단계로서 색인에 포함될 용어를 선정한다. 이는 연속간행물 색인작성과 유사하다. 용어가 이미 색인에 있는지와 그것이 어떻게 표현되었는지, 어떤 철자법 모순이 있는지, 용어가 부기입들을 갖고 있는지 없는지, 그리고 그 개념이 현재의 색인에서 “보라”참조를 가지고 있는지와 그 개념이 다른 용어아래 색인되어야 하는지를 결정하기 위하여 현존하는 색인을 참고하여야 한다.
- ② 색인용어들에 대한 고객의 검토 : 용어리스트를 고객에게 일반적으로 팩스나 e-mail로 보내어, 고객이 회사의 표현양식, 적합한 전문용어, 정확성 및 현존색인과의 일관성을 검토하도록 한다. 검토된 용어리스트는, 제안과 함께, 보통 2~3일 내로 반송되도록 한다. 고객의 변경과 제안을 논의하고, 새로운 전문용어에 순응하기 위하여 또는 회사의 명칭 변경을 반영하기 위하여 현존 색인기입의 변경여부를 결정한다. 여러 면에서 이와 같은 고객의 초기 검토는 도움이 된다.
- ③ 색인용어들의 통합 : 고객이 요구하였거나 또는 당신이 동의한 모든 변경을 만들어 혼합 색인(composite index)을 작성하는 색인작성 프로그램(예컨대, CININDEX, 윈도우 95버전)



에 입력한다. 보통 이 시점에서 용어들이 각 신문기사아래 독립적으로 리스트되었던 초기단계에서는 분명하지 않았던 약간의 명백한 불일치를 알게 된다. 가끔 고객에게 문의하면서 이러한 착오들을 조정한다. “Case-in-Point” 색인에서는 하이퍼텍스트 링크들이 용어가 나타나는 기사명에 작성된다. 기사명은 다음과 같은 코드로 시작된다 : 기사가 사례연구인 경우는 CS, 리포트인 경우는 R, 뉴스레터 기사인 경우는 N으로 시작된다.

- ④ 임시 웹페이지의 작성 : Netscape Composer안에 있는 HTML 문서를 열어 필요한 만큼 임시 웹페이지를 편집한다. 링크를 만드는 실제 횟수를 최소화하기 위한 복사와 붙이기 방식(copy-and-paste approach)을 사용하면서 모든 기사를 위한 하이퍼텍스트 링크들을 추가한다.

색인을 포맷하기 위하여 HTML 정의 리스트를 사용한다. 이는 두 개의 태그를 갖고 있는데; <DT> 태그는 용어를 위해 사용되고, <DD> 태그는 용어정의를 위해 사용된다. 웹 색인에서는, <DT> 태그는 주기입을 위해 사용될 수 있으며, <DD> 태그는 부기입을 위해 사용될 수 있다.

- ⑤ 불일치에 대한 추적 : 용어를 입력할 때 새로운 용어와 기존의 용어와의 처음 비교에서 또는 고객이 프로세스를 검토하는 동안에 드러나지 않았던 어법( wording)의 문제와 불일치가 일반적으로 발견된다. 이 때에 별로 중요하지 않은 변경은 수행하고, 그리고 해결할 수 없는 문제(예를 들어, 유사하지만 동일하지 않는 회사 명칭 들)에 대해서는 고객에게 문의한다.

불일치를 추적하는 일은 약간 지루하고 귀찮은 일이 되기 쉽다. 종종, 두 개의 파일을 비교하여 모든 중요한 변경이 옮겨졌는지 확인하여 가능한 한 CINDEX 파일을 정확하게 만든다.

- ⑥ 앵커의 추가 : 앵커들의 추가는 HTML에서는 비교적 용이하지만, 시간이 걸리며, 앵커의 정확한 배치를 요구하며, 또한 광범위한 항해(extended passages : 이와 동등한 온라인 페이지 범위)나 또는 동일한 기사안에서 분리된 참조들을 어떻게 취급할 것인지에 관한 판정을 위해 고객과 여러 번 통화할 것을 요구한다.

색인 페이지에 앵커를 삽입할 때 각각의 주 기입과 약간의 부기입에 앵커를 표시한다. 이는 “보라” 및 “도보라” 참조를 용이하게 만들 수 있게 하며, 또한 광범위한 키워드 색인(도치색인이라고 할 수 있음)시스템을 지원한다. 이러한 “도치” 색인들은 독자에게 관련된 기사를 찾으도록 도와준다. 각 기사의 끝에는 기사를 위한 주제와 회사명칭 색인기입을 나타내는 “키워드”리스트에 대한 하나의 링크가 있다. 하나의 키워드를 클릭함으로써, 이용자는 동일한 키워드로 색인된 다른 기사들을 찾을 수 있는 색인으로 되돌아가게 된다. 그러므로, 각 색인기입용어에 이어진 앵커들은 링크들이 키워드 페이지에서 일할 수 있도록 해야한다.

- ⑦ 색인의 테스트와 수정 및 포맷의 재 체크 : 온라인 상의 색인모습을 보기 위하여 브라우저에서 색인을 주의깊게 검토함으로써 색인을 테스트하고, 필요한 변경을 만든다. 그리고 나서 색인이 원하는 방식으로 보이는지 확인하기 위하여 포맷팅을 다시 체크한다.
- ⑧ 색인대상물을 위한 키워드 색인페이지의 작성 : 주요 색인 페이지를 발송한 후 “도치”색인을 위한 키워드 페이지를 작성한다. 모든 키워드 페이지는 이용자가 방금 읽은 기사로 돌아가거나, 새로운 기사로 가거나, 또는 관련 기사를 보기위해 색인으로 보내지는 것을 허용하는 링크들을 갖고 있다. 이 과정은 시간이 걸리지만 어렵지는 않다.
- ⑨ 색인의 갱신 : 예를 들어 “Case-in-Point” 색인은 실제로 6개 색인으로 되어있다. 24페이지의 기본색인(main index) 이외에, 또한 회사, 사람, 제품, 서비스 및 주요 주제를 리스트하고 있는 특수화된 길면서, 한 페이지로 된 5개 범주의 색인이 있다. 따라서 최초의 색인기입 리스트로 돌아가서(색인작성 과정에서 수정된 대로)그들을 이 5가지 범주로 나눈다.
- ⑩ 색인파일의 송부와 재송부 : 작업이 끝나면, 특수화된 색인파일을 e-mail로 웹마스터에게 보낸다.
- ⑪ “CINDEX” 파일의 갱신 : 이제 색인작성과정에서 취한 다양한 주석에 근거하여 최신호를 위한 “CINDEX” 파일을 갱신한다. 그리고 나서 최신색인파일을 누가 색인파일에 복사한다. 간단히 색인을 체크하고, 색인파일과 텍스트파일로서 이 파일을 보관한다. “Word”안에 있는 텍스트 파일을 열어 색인을 인쇄한다. 최종적으로 색인의 디스크파일을 고객에게 보낸다.

#### 4) 웹 사이트 색인작성을 위한 제안

규모가 큰 웹 색인의 설계에 참여하거나 또는 관심이 있는 사람을 위하여 약간의 제안을 하고자 한다(Rowland, 1999).

- “Acxiom” 색인에서처럼, 공백의 확인 불가능한 링크들을 제공하거나 또는 기본기입(main entry)에 대한 하나의 링크를 제공하는 것보다는 오히려 부수적 정보를 제공하기 위한 부기입과 부-부기입(sub-subentries)을 사용하라.
- “도치색인”은 이용자를 관련된 웹 페이지로 안내하는데 매우 유용하다. 또 다른 방법은 일반적으로 상호작용적 백과사전에서 하는 것처럼, 각 기사의 끝에 있는 “보라” 참조를 통해서 안내하는 것이다. 각 기사의 끝에 있는 링크들이나 또는 웹 페이지의 최신 리스트(즉, 갱신된 리스트)를 유지하는 일은, 색인작성과정이 어떤 방식으로든 자동화되지 않았다면, 색인을 준비하는데 상당한 작업을 추가할 것이다
- 다수의 색인시스템(예를 들면, 주제, 이름, 토픽, 제품, 회사 등)은 방대한 색인을 보다

적고, 보다 읽기 쉬운 단편들로 나눌 수 있게 도와준다.

- 다양한 웹 저작 프로그램들과 웹 색인작성 프로그램을 조사하라. 그러나 이들 프로그램이 당신의 웹 설계 문제를 해결해 주리라고는 기대하지 마라. HTML에 대한 훌륭한 기본지식은 필수적이다.
- 고객이 웹사이트 주위에서 정보 접근 방법을 찾을 수 있도록 도와주기 위하여 오로지 색인에만 의존하지 마라. 웹사이트는 또한 메뉴, 버튼, 이미지 맵과 또는 사이트 맵으로 잘 설계된 항해시스템을 필요로 한다. 현장 탐색엔진(on-site search engine)은, 고객이 정보접근을 위한 방법들의 선택을 필요로 하기 때문에, 또한 도움이 될 수 있다
- 서비스의 값을 공정하게 계산하라. 당신은 색인작성과 웹사이트 설계 서비스를 제공한다. 웹 색인작성에서는 페이지 당 요금이 의미가 없기 때문에 일반적으로 시간당 요금이 필연적이다.
- 웹 색인작성 방법과 전략 및 관련 웹 기술에 관해 최신성을 유지하라. 새로운 기술을 사용하라. 그러나 당신의 “보통 이용자(average user)”의 컴퓨터 능력을 비난하지 마라. 당신의 아이디어를 공유하기 위해 다른 사람과 네트워크를 형성하고(연락을 취하고), 웹 이용자를 위한 효과적인 색인을 작성하는 새로운 방법들을 배워라.

## 4.5 웹 색인작성의 학습

어떤 주제에 근거한 색인을 작성하거나 또는 특수한 관점에서 웹 사이트의 색인을 작성하는 일은 색인자에게 종이 대신에 HTML을 사용한 색인작성의 실제 경험을 쌓도록 하며, 또한 색인자는 문헌과 상호작용하기 위한 색인을 설계하는 것을 배우게 된다.

사람들이 흔히 범하는 실수는 웹 사이트가 어떤 페이지에 부여한 페이지 명칭만을 사용하는 것이다. 따라서 색인자는 그 명칭이 색인된 주제를 반영할 수 있도록 링크의 구축과 링크의 개명에 관해 알아야 한다. 효율적인 웹 색인작성을 위해서는 시소러스에 대한 충분한 이해와, HTML의 사용방법 및 앵커(anchor)에 대해서도 어느 정도 알아야 하며, 또한 탐색엔진과 피드백에 관해 학습하는 것도 매우 중요하다.

### 1) 링크(Link)

웹은 항상 변화한다. 일주일동안 작업한 링크가 다음 주에는 사라질 수 있다. 그러므로 당신의 색인을 정정하기 위하여 HTML에서 작업하는 경우에는 링크의 구축에 관해 어느 정도

알아야 한다. 링크를 유지하는 것은 매우 시간이 걸리는 일이다. 하이퍼텍스트 세계는 페이지 번호를 포함하지 않는다. 정보의 일부에 접근하기 위해서는 그의 URL을 알아야 하며, 그 URL에 대한 링크를 만들어야 한다.

인터넷 상에 정보 위치를 구축하기 위해서는 프로토콜(<http://>)과 도메인([www.kmu.ac.kr](http://www.kmu.ac.kr))을 상세히 기입하면 파일명칭([links.htm](http://www.kmu.ac.kr/~khyoon/web_indexing/links.htm))과 빈칸을 포함한 페이지(#subject-oriented lists)내의 앵커가 뒤따라 오는 URL([http://www.kmu.ac.kr/~khyoon/web\\_indexing/links.htm#subject-oriented\\_list](http://www.kmu.ac.kr/~khyoon/web_indexing/links.htm#subject-oriented_list))을 준다. 이것은, 예를 들면, 우리에게 Kmu Web Server 상의 페이지에 있는 페러그래프 — 웹 색인의 훌륭한 서지사항 — 을 지시할 것이다. 이에 대한 접근을 제공하기 위해서는 이 URL을 당신의 색인에 삽입하여야 한다.

웹 사이트에서 색인작성을 할 때는 적절한 링크들을 사용한다. 만일 사이트를 위한 파일들이 하나의 디렉토리 안에 있다면 색인내의 모든 링크는 단지 파일 명칭이 될 수 있다. 예를 들면, 완전한 URL(또는 완전무결한 링크):<http://www.abc.com/productx.html> 대신에 [productx.html](http://www.abc.com/productx.html) 이 될 수 있다. 이것은 링크들의 기능을 변경하지 않을 것이다. 그러나 이것은 당신이 오프라인으로 보다 쉽게 일하도록 해주며, 또한 색인을 단지 코드화할 필요없이 장차 당신의 웹 사이트와 다른 디렉토리나 도메인에 대한 웹 사이트의 연관된 색인을 이동할 수 있도록 해주는 보다 분명한 HTML을 제공할 것이다.

사람들은 이해할 수 있는 방법으로 링크들을 명명하는 중요성을 인식하기 시작하였다. 그러나 아직도 기괴하거나 불가해한(bizarre or arcane) 명칭이 많이 있는데(예를 들어, 방대한 숫자로 된 URL들), 그 의미를 분명하게 하기 위해 색인에 이러한 링크들을 개명하는 일은 색인자의 업무이다. 도서의 색인작성에서처럼 색인자는 이중-배치(double-post)를 할 수 있고, 여러 곳에 링크를 설치할 수 있다.

## 2) 시소러스

시소러스는 관계를 보여주는 미로에서 상위, 하위 및 관련 용어들을 갖는 거대한 정보지도(map)와 유사한 특수한 정보분야를 위한 통제 어휘집이다. 훌륭한 웹색인작성을 위해서 시소러스는 조리있게 색인에 삽입될 페이지를 인코드(encode)하거나, 색인용어를 삽입하기 위해 사용될 수 있다.

브라우저는 적합정보를 찾기 위해 웹-기반 시소러스를 통과할 수 있다(funnel down). 시소러스는 이용자가 웹사이트에 있는 정보를 탐색할 때 이용자의 여러 가지 견해를 조화하도록 도와줄 수 있다.

### 3) HTML

HTML은 SGML의 단순화된 버전이다. 이는 문서를 표현하기 위해 콘텐츠를 제공하고 정보를 레이아웃하는 일련의 태그들이다. 예를 들면, `<H1>Main Heading </H1>`은 Main Heading의 단어들을 갖는 수준1의 heading이다. 기본적인 링크 HTML코드는 `<A HREF="http://www.abc.com/">ABC</A>`이다.

이는 어떤 웹페이지 상에 있는 단어 ABC로부터 HTTP 프로토콜을 사용하여 www.abc.com 이라고 일컫는 다른 웹사이트로 브라우저를 이동하게 할 것이다. 당신은 앵커(anchor: 단지 페이지의 맨 위가 아니라 페이지내의 어느 포인트로 이동하게 해주는 HTML문헌에 있는 태그)들을 삽입하기 위해 외부의 웹 문헌을 편집할 수 없다.

### 4) 앵커 (Anchor)

HTML에서 앵커는 용어, 구절, 이미지 및 기타 다른 모든 정보 객체에 대해 선언될 수 있다. 같은 HTML 문서 내의 다른 장소로 이동하기 위한 하이퍼텍스트 링크는 다음과 같이 정의하며, 이때 이러한 앵커를 특히 named anchor라고 부른다.

```
<A NAME="앵커 이름">링크를 만들 부분</A>
```

또 어떠한 HTML 파일 이름이라도 앵커나 연결 대상이 될 수 있다는 사실에 유의할 필요가 있는데, 하이퍼 링크를 마우스로 클릭했을 때 보여줄 대상 정보가 다른 문서에 존재하는 경우에는 다음과 같이 선언한다.

```
<A HREF="파일이름.html">링크를 만들부분</A>
```

만약 어떤 문서내의 특정 위치가 보여지길 원하는 앵커나 링크라면 다음과 같이 선언한다.

```
<A HREF="파일이름.html#앵커이름">링크를 만들 부분</A>
```

하이퍼텍스트의 고민은 사람들이 매우 빠르게 길을 잃게 되는 것이다. 여기 저기로의 이동은 프로그래머에게는 재미가 있을 지 몰라도 이용자에게는 싫증나고 괴로운 일일 수 있다.

### 5) 탐색엔진과 피드백

당신은 색인에 추가할 정보를 찾기 위하여, 그리고 또 당신의 사이트에 대한 적중건수를 증가시키기 위하여 탐색엔진을 효과적으로 사용하는 방법을 배울 필요가 있다. 적합정보를 찾기 위한 가장 좋은 접근은 특정성을 위해 선택된 키워드로 시작하는 것과 그리고 만일 적중건수가 없으면 거슬러 올라가서 다시 일하는 것을 포함한다. 당신이 검색한 것을 통해 스크롤(scroll : 표시화면 내용을 순차적으로 1행씩 올리거나 내리는 일) 할 때 당신의 상식을

사용하고 매력적인(흥미있는) 사이트를 주시하라. 가끔 쇼핑사이트나 온라인 게시판 사이트들이 검색결과를 넘치게 하는데 이는 그들의 헤더(header)에 코드화된 키워드들이 있기 때문이다. 소스코드를 보고 당신이 선택한 페이지가 왜 검색되었는지를 설명해 주는 페이지에 어떤 메타 태그가 나타나는지 살펴보아라. 당신은 최초에 시도했던 키워드보다 탐색을 위해 더 적합한 키워드를 발견할지도 모른다. 그러면 더 좋은 결과를 얻기 위해 당신의 탐색을 정제할 수 있다.

당신의 색인 사이트에 대한 적중건수를 향상시키기 위해서는 계수기를 추가하라. 계수기는 사람들이 당신의 사이트를 방문한 회수를 알려준다. 당신의 색인 페이지 안에 메타 태그를 설치할 것을 권유한다. 만일 당신이 당신의 색인에 대해 보다 많은 적중건수를 얻는다면 그것은 당신이 올바른 트랙에 있음을 알려주는 것이다. 당신은 또한 당신의 서버나 ISP에 의해 제공될 수 있는 통계적 추적방법을 통해서 적중건수가 어디서 오는지를 확인할 수 있다. 당신의 URL을 메일링 리스트에 의해 배포하거나 또는 탐색엔진에 공급함으로써 보다 광범위하게 알려지도록 하라. 접촉을 위한 e-mail 주소를 남겨두라. 당신은 독자층을 만족시키기 위하여 당신의 색인을 수정할 수 있을 것이다.

## 5. 결 론

본고는 웹 색인작성에 관한 기초연구이다. 먼저, 웹 색인의 기초가 되는 삽입색인을 살펴보고; 다음으로, 웹 색인작성에서 많이 사용되는 메타 태그를 검토하고, 끝으로, 웹 색인작성법과 관련사항들을 면밀히 연구 검토하였다. 그 요점은 다음과 같다.

1) 삽입색인은 전자출판(DTP)에서 많이 사용되는 새로운 기법으로 웹 색인작성의 토대를 마련해 준다.

2) 메타 태그는 웹 사이트의 정보제공을 위해 많이 사용된다. 특히, 디스크립션 태그와 키워드 태그는 웹 색인작성에 가장 적합한 태그로서, 그들의 작성기법은 색인자의 특별한 관심사가 되며 그의 능력을 발휘할 수 있는 분야이다.

3) 웹 색인의 기입요소, 색인작성의 다양성 및 작성과정, 웹 색인작성을 위한 학습사항 등을 중점적으로 논의하였다.

실제로 웹 색인작성이라는 주제에 대한 결론은 없다. 우리가 웹 색인작성에 관한 약간의 질문에 답할 수 있는 동안에 더 많은 질문이 항상 발생한다. 웹 색인작성은 새롭게 발전하는

분야다. 웹에 대한 색인이 정보검색의 효과적인 방법이 될 수 있음을 우리는 이미 알고 있다. 정보분석의 능력을 가진 색인자들은 온라인 접근도구를 창안하기 위한 첫 번째의 후보들이다. 만일 당신이 색인작성과 웹 페이지 디자인에 관심이 있다면, 웹색인작성의 다양성을 탐구하고 아울러 독자들의 유동성과 가변성을 감지할 필요가 있다. 어떠한 단 하나의 해답도 모든 사이트와 환경에 작용할 수는 없을 것이다. 당신의 상상력을 사용하여 이 새롭고 항상 변하는 미디어에서 이용자와 색인자-웹 마스터의 요구를 충족시켜 주는 새로운 색인작성 기술을 개발하여야 한다.

우리는 기술에 대한 공포를 생각하지 않아야 한다. 그렇지 않으면 미지의 세계가 우리를 방해한다. 웹이라고 하는 바다는 여전히 파도가 일고, 정보의 바다는 어둡고, 깊고, 그리고 불확실성으로 가득차 있다. 그러나 그것은 어떠한 향해가 될 것인지 모른다(Broccoli.1999). 훌륭한 웹 색인작성 기술은 언제나 보상받을 수 있을 것이다.

## 참 고 문 헌

- Bonura, Larry. 1994. *The Art of Indexing*. New York : John Wiley.
- Broccoli, Kevin and Gerry Van Ravenswaay. 1999. "Web Indexing Anchors Away!". In : *Beyond Book Indexing*, ed. by Diane Brenner and Marilyn Rowland. Medford, NJ : Information Today, Inc., 2000.
- Browne, Glenda and Jonathan Jerney. 2001. *Website Indexing : Enhancing Access to Information within Websites*. Adelaide, Australia: Auslib Press.
- Castro, Elizabeth. 1998. *HTML for the World Wide Web : a Visual QuickStart Guide*. Berkeley, CA : Peachpit Press.
- "Dan's Web Tips : Title, META tags, and Search Engine Robots".  
(<http://www.softdisk.com/comp/dan/webtips/titles.html>).
- Drott, M. Carl. 2002. "Indexing Aids at Corporate Websites : the Use of Robots.txt and META Tags", *Information Processing & Management*, Vol. 38, No.2. pp. 209-219.
- Maislin, Seth. "Tutorial on Index Tagging". (<http://www.oreilly.com/people/staff/seth/tutor.html>).
- Malik, Ed. 1994. "Indexing at Frame Technology Corporation in What Editors Need and Want : are They Getting It?", *The Changing Landscapes of Indexing, The Pro-*

- ceedings of the 26th Annual Meeting of the American Society of Indexers*, San Diego, California, May 13-14, American Society of Indexers.
- Mauer, Peg. 1998. "Embedded Indexing : What is It and How Do I Do It?", *Key Words*, Vol. 6, No.1. pp. 14-15.
- Mauer, Peg. 1998. "Embedded Indexing in FrameMaker", *Key Words*, Vol. 6, No.5. pp. 1-6.
- Merriam Webster's Collegiate Dictionary*. 1993. 10th edition. Springfield, MA : Merriam Webster.
- "Meta Medic". (<http://www.northernwebs.com/set/setsimjr.html>)
- "Meta Tag Builder". (<http://vancouver-webpages.com/META/mk-metas.html>).
- "Meta Tag Resources". ([http://webdeveloper.com/html/html\\_metatag\\_res.html](http://webdeveloper.com/html/html_metatag_res.html)).
- "Metatags". (<http://www.metatages.net>).
- Mulvany, Nancy. 1994. *Indexing Books*. Chicago: University of Chicago Press.
- Rosenfeld, Lou. 1997. "Organizing Your Site from A to Z". (<http://webreview.com/wr/pub/97/10/03/arch/index.html>).
- Rowland, Marilyn J. 1999. "<META> Tags". In: *Beyond Book Indexing*, ed. by Diane Brenner and Marilyn Rowland. Medford, NJ : Information Today, Inc., 2000.
- Walker, Dwight. 1996. "Web Indexing: an Exercise in Hypertext Navigation", *LASIE*, Vol. 27, No.3. pp. 50-58.
- Walker, Dwight. 1998. "Web Indexing : the State of Play", *Online Currents*, Vol. 13, No.2. pp. 7-24.
- Walker, Dwight. 1999. "Subject-Oriented Web Indexing". In: *Beyond Book Indexing*, ed. by Diane Brenner and Marilyn Rowland. Medford, NJ : Information Today, Inc., 2000.
- Wellisch, Hans. 1995. *Indexing from A to Z*. New York : H.W. Wilson.
- Wheatley, Henry Benjamin. 1902. *How to Make an Index*. London : Elliot Stock.
- Wright, Jan C. 1999. "The World of Embedded Indexing". In: *Beyond Book Indexing*, ed. by Diane Brenner and Marilyn Rowland. Medford, NJ : Information Today, Inc., 2000.