

# 패킷 분석을 이용한 웹 자원의 조직\*

## A Study on Organizing the Web Using Facet Analysis

유 영 준(Yeong-Jun Yoo)\*\*

### 초 록

웹 자원을 색인 및 조직하는 데에는 키워드 추출에 기반한 자동색인에 의한 방법과 기존의 분류 체계나 주제별 디렉토리를 이용하는 방법이 있다. 그러나 두 방법은 각각 표준적인 방식이 없거나 분류 구조상의 불합리성으로 인하여 이용자의 정보 요구에 제대로 대처하지 못하고 있다. 따라서 이 논문에서는 기존 분류 체계의 구조상의 한계와 특정성과 망라성과 같은 웹 자원의 특성과 관련된 문제를 지적함과 더불어 패킷 분석에 의한 웹 자원 조직의 타당성과 그 장단점을 기술하였다. 그리고 패킷 분석을 이용한 구체적인 방법론 세 가지를 제시하였다. 첫째, 패킷 분석을 기반으로 한 색인 시스템, 둘째, 열거형 분류 체계를 패킷 분류 체계로 구조를 수정한 대안, 셋째, 국내 검색 엔진의 주제별 디렉토리의 패킷 모형 등이다. 이러한 방법론들을 검토한 결과, 패킷 분석에 의한 통제어휘 구축이 웹 자원을 조직하는 수단으로 활용될 수 있는 가능성을 제시하였다.

### ABSTRACT

In indexing and organizing Web resources, there have been two basic methods: automatic indexing by extracting key words and library classification schemes or subject directories of search engines. But, both methods have failed to satisfy the user's information needs, due to the lack of standard criteria and the irrationality of its structural system. In this paper I have examined the limits of library classification scheme's structures and the problems related to the nature of Web resources such as specificity and exhaustivity. I have also attempted to explain the logicity of Web resources organization by facet analysis and its strengths and limitations. In so doing, I have proposed three specific methods in using facet analysis: firstly, indexing system by facet analysis; secondly, the alternative transformation of the enumerative classification scheme into facet classification scheme; and finally, the facet model of subject directory of domestic search engine. After examining the three methods, my study concludes that a controlled vocabulary by facet analysis can be employed as a useful method in organizing Web resources.

키워드: 패킷 분석, 웹 자원, 분류 체계, 주제별 디렉토리, 통제어휘

Facet Analysis, Web Resources, Classification Scheme, Subject Directory, Controlled Vocabulary

\* 본 논문은 2004년도 한국비블리아학회 춘계학술발표회에서 발표한 내용을 수정·보완한 것임.

\*\* 연세대학교 문헌정보학과 강사(yp6370@hanmail.net)

논문접수일자 2004년 5월 28일 논문심사일자 2004년 6월 7일 게재확정일자 2004년 6월 23일

## 1. 서론

현재 웹상에는 약 2억명 이상의 이용자가 있으며, 대략 30조 가량의 문서들이 존재하고 있다. W3C(<http://www.w3c.org>)에서는 향후 이용자 수가 10억 정도에 이르게 되고 이용 가능한 문서의 수는 더욱 많아질 것이라고 전망하고 있다(Fensel et al. 2003). 이 같은 내용을 반영이라도 하듯 최근 웹상에서 문서의 수는 천문학적인 증가 추세를 보이고 있다. 그러나 이러한 폭발적인 성장은 이용자 계층의 다양성과 함께, 정보의 검색, 접근, 제공, 유지 등을 더욱 어렵게 만들고 있다.

펜셀(Fensel et al. 2003) 등은 이용자가 특정 정보를 웹에서 검색할 경우, 상당히 많은 부적합한 정보들 때문에 많은 시간을 낭비하고, 오히려 적합한 정보를 검색할 시간과 기회를 놓치고 있으며, 웹 자원 역시 일관성 있게 유지하는 것이 매우 어려운 상황이라고 지적하고 있다. 이는 펜셀 등이 시맨틱 웹(semantic web)의 관점에서 기술한 것이지만, 이 연구에서, 패킷 분석을 이용하여 웹 자원을 효율적으로 조직함으로써 해결하려고 하는 내용과 동일한 것이라고 할 수 있다. 그리고 디바다손(Devadason et al. 2002) 역시 웹에서 검색되는 자원의 질을 신뢰할 수 없고, 인터넷 자원을 색인하고 조직하기 위한 표준적인 방법이 없음을 언급하고 있다.

웹상의 정보를 검색하기 위해 웹 자원을 색인 하는 방식은 크게 두 가지로 나뉘는데, 웹 자원으로부터 키워드를 추출하는 방식과 분류체계를 이용하는 방식이다. 분류체계를 이용하는 방식은 다시 두 가지로 구분할 수 있는데,

문헌정보학의 전통적인 정보 조직 수단인 분류체계를 이용하는 경우와 검색엔진이 자체적으로 제공하는 주제별 디렉토리 서비스다.

먼저, 기존의 분류체계를 이용하여 웹 자원을 색인하는 경우에는, 주로 열거형 분류체계를 지칭한다. 열거형 분류체계는 십진법과 같은 단일 기준에 의해 지식을 나열하기 때문에 다차원적인 성격을 갖는 지식의 본질을 제대로 반영하기 어렵다. 그리고 인쇄형태의 자료와 구별되는 웹 자원의 다양한 디지털 특성들을 수용하는데 한계가 있고, 너무 포괄적인 수준에서 정보를 색인하기 때문에 학제적인 주제의 경우 주제를 구성하는 정보의 일부가 검색되지 않거나, 접근점에서 제외되어 특정 정보를 검색하기 원하는 이용자들의 정보 요구를 만족시키기에 부적합한 색인 방식이라고 할 수 있다. 따라서 열거형 분류체계에 의한 웹 자원 조직 및 색인 방법은 이용자가 표현하는 의미를 색인시스템이 제대로 이해하지 못하는 결과를 낳게 되고 양자간의 커뮤니케이션이 원활하게 이루어지기 어렵다.

다음은 분류 체계와 유사한 구조인 주제별 디렉토리를 사용하는 경우인데, 야후(Yahoo)와 구글(Google) 등 일부 검색 엔진에서 자신의 고유한 조직 체계를 기반으로 주제 카테고리에 의해 웹사이트를 조직하는 사례이다. 이 방법은 각 검색엔진마다 상이한 주제별 디렉토리로 자원을 조직하여 표준적이고 보편적인 방식이 없다는 점과 다양한 분류 기준이 혼합된 디렉토리들로 구성됨으로써 분류 구조상에서 일관성을 상실하고 있다는 것이다. 이용자는 동일한 자원을 검색할 때 검색엔진마다 상이한 디렉토리를 검색해야 하는 불편을 감수

해야 한다.

그리고 현재 웹상에서 사용되는 검색 엔진들은 인터넷 자원을 조직하기 위하여 복잡한 색인 기법을 이용하고 있다. 대표적인 예로서 구글(Google)과 알타비스타(Altavista)와 같은 검색 엔진은 개량된 불리언 검색을 채용하고 있지만, 이들 엔진은 대부분의 이용자가 원하는 특정 정보에 대한 요구를 수용하지 못하고 있다(Patel 2002). 이처럼 이들 검색엔진에서 발생하는 문제는 웹상의 자원을 색인 하는데 서로 다른 방식들을 사용하고 있고, 그들 자신만의 순위 알고리즘(ranking algorithm)에 따라 순위를 매기는 방식을 사용하는 데서 발생하는 것이다.

이처럼 각 검색엔진마다 웹 자원의 조직 및 색인방식이 각각 달라서 표준적인 방법이 없고, 일관성이 결여되어 있으며, 현재 사용하고 있는 방법은 이용자의 실제 정보요구와는 괴리되어 있어서, 이용자와 정보제공시스템 또는 검색엔진간의 의사소통, 즉 의미전달이 원활하게 이루어지지 않아서 검색의 정확률이 떨어질 수밖에 없는 문제가 야기된다는 점이다.

이제까지는 전통적으로 문헌정보학에서 정보 조직 수단으로 사용해 온 분류체계를 활용하여 웹 자원을 조직하는 사례들이 제시되어 왔다. DDC, UDC 등으로 대표되는 열거형 분류체계를 적용한 사례가 대부분이었다. 이 논문에서는 기존에 웹 자원을 조직하는데 적용되었던 열거형 분류체계의 단점과 그 한계를 지적하고, 그 대안으로 패킷 분석에 근거한 색인 기법을 이용하여 웹 자원을 조직하기 위한 가능성을 평가하고 그의 방법론을 제시하고자 한다.

## 2. 기존 분류 체계의 한계

여기에서는 분류구조상의 한계를 먼저 논의한 후에 웹 자원의 특성과 관련된 기존 분류 체계의 한계에 대해 살펴 보고자 한다.

### 2.1 분류 구조상의 한계

#### 2.1.1 열거형 분류 체계를 이용한 경우

웹 자원을 조직하는데 이용되는 분류 체계는 대부분 DDC나 LCC와 같은 열거형 분류 체계이다. 국내에서 웹 자원을 조직하는 수단으로 분류 체계를 이용하는데 관한 연구들도 열거형 분류 체계의 범주에 국한되어 있으며, 열거형 분류 체계의 구조 내에서 웹 자원을 조직하는 방안을 제시하였고, 열거형 분류 체계의 구조적 한계를 지적하고 그 한계를 극복할 수 있는 대안을 검토한 연구는 이루어지지 않고 있다. 또한 대부분의 연구들이 특정한 주제 분야를 대상으로 한 연구들이어서 모든 주제 분야에 적용할 수 있는 방법은 아닌 것으로 보인다(김성희 2002, 김정현, 문지현 2002, 이란주, 성기주, 양정하 2001, 이창수 2000, 최재황 1998, 최희운 1998).

외국의 경우에 이러한 흐름은 코흐(Koch et al. 1997), 맥키어난(McKiernan 1997, 1998, 1999, 2000), 사에드와 초드리(Saeed and Chaudry 2001), 우드워드(Woodward 1996) 등의 연구에도 잘 나타나 있다. 그러나 열거형 분류 체계는 근본적으로 구조적 결함을 지니고 있다. 그 이유는 인간이 사물을 인식하고 분류하는 사고 작용과 열거형 분류 체계가 지식을 분류하고 조직하는 방식이 다르기 때문에 이용

자와 분류 체계간의 의사소통에 문제가 생겨서, 열거형 분류 체계를 적용할 경우, 이용자가 원하는 의미 있고 특정한 정보를 정확하게 제공할 수 없게 된다는 것이다. 이밖에 이 열거형 분류 체계는 다음과 같은 단점을 지니고 있다.

첫째, 열거형 분류 체계는 전체 지식의 구조를 완전하게 표현하려고 시도하여 규모가 커지게 되지만, 지식의 역동적인 속성 때문에 지식의 세계에 존재하는 ‘모든’ 주제를 열거하는 것은 불가능하게 된다. 둘째, 엄격한 계층 구조로 인하여, 이 체계들은 학제적인 주제뿐만 아니라 새로운 주제를 받아들이기에 적합하지 않다. 셋째, 인터넷과 같이 특정 정보를 검색하는데 이용될 경우, 열거형 분류 체계는 매우 특정한 주제를 분류하는데 적합하지 못하다. 넷째, 열거형 분류 체계를 사용할 경우, 분류자가 정보 패키지에 맞는 체계를 선택하도록 강요당해서, 분류표에 전개되지 않은 해당 항목의 여러 중요한 측면들이 생략될 수 있게 된다. 다섯째, 열거형 분류 체계의 계층적 구조 때문에, 오류가 발견되거나, 보다 나은 구조가 개발되었을 때, 주제를 재분류하기 어렵다. 여섯째, 이들 분류 체계는 주관성과 문화적 편견에 쉽게 영향을 받게 되는데, 예를 들어, DDC의 경우 세계의 다른 종교들에 비해 기독교에 대해서는 심도 있는 분류를 하였으며, LCC의 경우 육군 및 해양군사과학에 우선순위를 준다.

이상에서 살펴 본 바와 같이 열거형 분류 체계는 논리적인 구조에 결함이 있다고 할 수 있다. 열거형 분류 체계는 기본적으로 나무 구조에 의해 조직되어 보다 특정하고 구체적인 하위 개념을 전개하는데 초점을 맞추고 있어서, 상위 개념과 하위 개념의 포섭관계만이 유일한

관계 유형으로 제시하는 경향이 있다. 이 같은 구조상의 문제로 인하여 분류 체계가 상대적으로 지나치게 광범위하게 되고, 부류(classes) 간의 결합을 위한 장치들이 제한될 수 있으며, 복잡한 의미적 내용을 표현하기에 적합하지 못하게 된다. 더욱이 학제적 환경에서의 검색이나 복잡한 속성을 갖는 대상을 검색하는 경우에 더욱 어려움을 겪을 수밖에 없는 것이다 (Broughton 2002).

### 2. 1. 2 자체 분류 체계를 이용한 경우

포털 사이트에서 제공하는 디렉토리 서비스는 분류 체계와 유사한 구조를 가지고 있다. 그러나 각 사이트마다 디렉토리의 구성이 다양하게 나타나는 등 분류 구조 측면에서 일관성을 상실하고 있다. 이처럼 웹 분류체계의 비밀관성에 대해 논의한 학자들이 있는데, 구체적으로 후돈이 논급한 사례를 살펴보면 다음과 같다.

후돈(Hudon 2000)은 웹 분류체계의 비밀관성을 주장한 대표적인 학자이다. 그가 제시한 문제는 동일한 웹페이지에서 비밀관성이 나타난다는 것이다. 이는 상이한 분류 기준을 반영하는 범주가 혼합되어 있음을 지적한 것인데, 예를 들어 ‘네이버’의 디렉토리에서 명백하게 다른 주제 영역인 ‘가정’과 ‘여성’, ‘교육’과 ‘참고자료’가 하나의 범주로 구분되어 있고, ‘엠파스’의 경우 ‘교육’과 ‘학교’, ‘정부기관’과 ‘사회’ 등이 하나의 범주로 구분된 경우 등이 바로 그것이다. ‘구글’도 ‘교육’과 ‘참고자료’가 하나의 범주에 구분되어 있고, ‘야후 코리아’에서도 유사한 문제를 지니고 있었다. 이러한 비밀관성은 웹 디렉토리 개발자들이 다양한 분류 모형을 차별화하지 못한 결과에서 나타난 것이

다(Zins 2002). 따라서 이 같은 오류를 극복하고, 보다 보편적으로 인정받을 수 있는 표준적인 조직 방법이 필요한 것으로 보인다.

이러한 비밀관성을 해결하기 위하여 진스(Zins 2002)는 주요한 상업적 포털, 사서들이 추천한 사이트 그리고 도서관 분류 체계를 웹에서 적용한 예 등의 인터넷 게이트웨이의 분류 체계를 분류 구조의 관점에서 분석하고 표준적인 분류 모형을 제시하였다. 이 연구에서는 웹에 존재하는 현상을 분석하여 유형을 제시하였다.

분석 대상은 각 사이트의 홈페이지에 제시된 분류 체계의 최상위 범주만을 대상으로 하였고, 해당 범주가 갖는 텍스트의 의미를 분석하였다. 그리고 해당 범주가 어느 모형에 속하는지를 분석하기 위해 세 요소를 기준으로 삼았는데 그 요소들은 다음과 같다. 첫째는 개념 구조로서 용어의 의미와 다른 용어와의 논리적 관계이다. 둘째는 문맥으로서 체계 내에서 용어의 위치이다. 셋째는 연결된 웹 자원의 속성 등이었다. 이 과정은 결과적으로 범주의 어휘가 지니는 의미를 분석한 것이라고 볼 수 있다. 이 어휘 분석을 통해 분류 구조적으로 문제가 있는 범주를 분류 모형에 맞게 조정하는 대안을 제시하였는데, 분류 모형이란 구분 기준으로서의 모형과 분류 체계로서의 모형의 의미를 모두 갖는 개념이다.

결론적으로 제시된 분류 모형은 8개이었는데, 1) 예술, 비즈니스, 교육 등과 같은 주제(subjects), 2) 조직체나 사람(people) 등과 같은 대상(objects), 3) 채팅, 전자메일, 쇼핑 등과 같은 응용분야(applications), 4) 어린이, 성인, 여성 등과 같은 이용자(users), 5) 캐나다, 영국, 미국과 같은 지역(locations), 6) 알

마나, 사전, 지도 등과 같은 참고자료(reference), 7) 그래픽, 그림, 라디오 등과 같은 매체(media)인데 매체는 테크놀로지와 매체를 모두 포함하는 모형, 8) 독일어, 스페인어, 불어 등과 같은 언어(languages) 등의 모형이었다.

진스는 8개의 분류 모형을 제시했을 뿐만 아니라 개별 모형을 통합하여 자원을 조직하는 방법을 제시하였다. 이는 각 모형을 하나의 패킷으로 인식하고 패킷을 통합하여 자원을 조직하는 방법을 의미한다. 주제별 디렉토리 분석을 통해 패킷을 범주화하고 패킷과 패킷을 합성하는 패킷 분석의 방법론을 적용한 것으로 파악할 수 있다.

## 2.2 웹 자원의 특성과 관련된 기존 분류 체계의 한계

정보자원의 의미를 식별하는 문제는 그 자원의 본질적 특성이라고 할 수 있다. 웹 자원을 색인 하는 과정에 영향을 주는 특성이 있는데, 자료의 지적 내용이 매우 복잡하고, 특정성(specificity)과 망라성(exhaustivity)을 모두 만족시킬 수 있는 수준 높은 색인이 필요하다는 것이다(Broughton 2002).

따라서 디지털 환경에서 '상호참조(cross-referencing)'를 가능하게 하는 도구인 하이퍼텍스트를 활용하는 방안을 고려할 만 하다(Broughton 2002). 하이퍼텍스트는 최소한 두 가지 방식으로 검색을 개선시킬 수 있다. 첫째, 검색 결과를 시각적으로 디스플레이 할 때 이용자에게 한 번에 많은 계층을 제시하지 않는 동시에 위계 구조내의 세부 내용을 그대

로 유지하면서 연속적인 전개과정을 제시하는데 이용될 수 있다는 것이다. 둘째, 하이퍼텍스트를 이용함으로써 네트워크 자원에 대한 다중 접근점 개념을 지원하고 시스템 구조의 규칙성을 강화할 수 있다는 점이다.

기존의 열거형 분류 체계가 전자문서를 폭넓게 조직하고 브라우징 하는데 도움이 되지만, 자원을 정밀하게 조직하고 효과적인 검색 도구로서의 기능은 수행하기 힘들다(Devadason et al. 2002). 특히 전자문서의 경우 신속하게 갱신되고, 갱신에 될 때마다 그들의 주제가 변경되므로, 열거형 분류체계에 의존하여 웹 자원을 조직할 경우 그 목적을 완전하게 이룰 수 없을 것이다. 또한 색인어를 사용하여 개념을 기술하는데, 사용 가능한 계층 수준이 하나 또는 둘 정도밖에 안 될 경우 특정성이 떨어지는 결과가 나오기 때문에 열거형 분류 체계에 기반을 두고 있는 색인 시스템이 제공하는 포괄적인 수준의 주제기술은 주제 지향적 데이터베이스에 계속적으로 문제를 야기 시킬 것이다(Broughton 2002). 하나의 예를 들어 보면, 'Visual Arts Data Service' 데이터베이스에 저장된 '19세기 정치 만화'를 다룬 자원은 '19세기 정치사'에 관심을 가지고 있는 이용자는 검색할 수 없다는 것이다. 이 같은 이유는 이용자가 원하는 항목이 '비주얼 아트 - 만화 - 19세기'와 같은 색인어 용어열로 색인될 것이기 때문이다. 결국, 이러한 현상은 학문이 발전하고 전문화됨에 따라 학제적 연구가 보편화되고 있는 연구 경향(Iyer 1995)과 웹에 존재하는 자원의 유형이 멀티미디어 형태로 변화하는 흐름을 감안할 경우 이들 변화에 대처하기 어렵게 될 것이다.

### 3. 패킷 분석에 의한 웹 자원의 조직

이상에서 살펴본 열거형 분류체계를 이용한 웹 자원 조직의 한계를 극복할 수 있는 대안으로 패킷 분석을 이용한 웹 자원의 조직을 고려할 수 있다. 패킷 분석은 랑가나단(Ranganathan)이 창안한 콜론분류 즉 분석합성형 분류 체계에서 개념을 분석하는 방법을 말한다. 이 패킷 분석을 이용하여 웹 자원을 조직할 경우, 그 주제 영역으로부터 추출한 개념을 사용하여 구조화하기 때문에 다른 주제 영역과 구별되면서 특정한 주제 영역의 고유한 개념 구조를 구축할 수 있게 된다.

그리고 웹사이트 설계자, 색인자, 이용자에게 그 페이지의 내용을 기술하고 식별할 수 있는 공통 언어를 제공한다는 점에서 유용할 것이다. 이처럼 패킷 분석을 이용하는 것은 웹 자원의 디지털 특성과 같은 새로운 개념과 학제적인 주제에 대해서 매우 수용적이기 때문에, 빠르게 성장하는 웹 환경에 적합한 정보 조직 수단이 될 수 있을 것이다.

#### 3.1 웹 자원 조직 수단으로서 패킷 분석의 타당성

열거형 분류와 함께 지식을 구조화하는 대표적 유형은 분석합성형(analytico-synthetic) 분류 체계이다. 이 분류 체계는 개념간의 계층구조를 일일이 제시하는 대신, 문헌의 주제를 소수의 관점(특성이나 측면)으로 분석하여 파악하고 이들 관점마다 분류기호를 부여한 다음 이를 일정한 원칙에 따라 합성하여 주제를 표현하는 방식이다. 분석합성형 분류체계에

서는 개념을 일정한 특성에 따라 구성요소로 분석하는데 이를 패시 분석이라고 한다(김태수 2000). 여기에서 패시(facet)이란 문헌의 주제를 구성하는 소수의 관점이나 특성 또는 측면을 말하는 것으로서, 패시는 분명하게 정의되고, 상호 배타적이고, 그리고 집합적으로 완전한 부류(class)나 특정 주제의 측면, 속성 또는 특성을 말하는 것이다(Maple 1995).

랑가나단은 패시 분류 체계를 구축하는 과정을 세 단계로 나누었다. 즉 패시 분석과 합성 과정은 다음과 같다. 첫째는 주제 영역을 구성 요소로 분석하는 과정인 개념 단계(Idea Plane)이고, 둘째는 구성 요소를 표현하기에 적합한 전문용어를 선정하는 표현 단계(Verbal Plane)이며, 셋째는 기호체계를 사용해서 구성 요소를 표현하는 기호단계(Notational Plane)이다. 엘리스와 바스콘셀로스(Ellis & Vasconcelos 1999)는 비록 사용한 용어는 다르지만 이들 내용에 대해 다음과 같이 설명하고 있다. 먼저 특정 주제 영역을 간단한 개념 용어로 분석하고, 이들 용어들을 패시로 군집화하고, 패시의 순서를 배열하기 위해 기호를 부여한 후, 여러 패시으로부터 용어를 조합하여 문헌의 내용을 합성하여 구조화하는데, 이들 분석 과정의 결과, 용어로 구성된 통제어휘(controlled vocabulary)가 생성되게 된다. 이 통제어휘는 시소러스나 주제명표목과 유사한 구조를 가지며, 개념 용어들은 어떤 관념적 개념화에 기반한 것이 아니라 해당 주제 영역에서 사용되는 실제 용어들로부터 추출되고 그 주제 영역을 반영한 결과이다. 다시 설명하면, 패시 분석 모형에 근거한 분류 체계는 열거형 분류 체계와는 다른 원칙을 적용하며 지식 전

체를 열거하려는 관점이 아니라 주제를 구성하는 개념들로부터 출발한다는 것이다. 따라서 패시 분석에 의한 색인 수단을 개발하는데 있어서 첫 번째 단계는 사전이나 백과사전, 표준이 될 만한 텍스트, 논문과 같은 자료들로부터 소규모의 용어 데이터베이스를 구축하는 것이다. 이러한 방식은 시소러스를 구축하는 방식이 동일하다고 볼 수 있다.

기존 분류 체계와 달리 패시 분석은 주제의 특정한 한 측면 그 이상의 표현이 가능하다. SLAIS(School of Library, Archive & Information Studies at University College London)는 정부가 재정을 지원하는 인문학 관련 주제 게이트웨이인 AHDS(The Arts and Humanities Data Service, <http://www.ahds.ac.uk>)와 Humbul(<http://www.humbul.ac.uk>)을 통합하려는 연구계획에서, 패시 분석을 엄격한 용어학적 분석이라고 정의하였다. 그리고 해당 주제 영역의 어휘를 패시와 용어열로 조직하고 그 결과로 의미론적 관계와 구문 분석적 관계 모두를 명확하게 표현할 수 있는 복잡한 지식 구조가 구축되는 것을 패시 분석의 최종 목표로 제시하였다(Broughton 2002). 이러한 관점에서 패시 분석을 이용하여 웹 자원을 조직하는 목적은 특정한 주제 영역의 개념들을 철저하게 분석하고 주제 영역을 구성하는 용어들을 합성함으로써, 이용자의 정보요구가 무엇인지를 정확하게 분석하여 이용자가 원하는 자원만을 특정하게 제공하는 것을 목표로 한다. 이는 웹 자원의 특정성과 망라성 모두를 향상시킬 수 있는 보다 높은 수준의 색인 기법으로 패시 분석 이론을 적용하려는 것이다.

패시 분석을 이용한 가장 대표적인 분류 체

계는 랑가나단이 창안한 콜론분류(Colon Classification)이다. 랑가나단은 의학, 수학 혹은 문헌정보학과 같은 기본 주제를 전제하고 순차적으로 개체 개념을 합성하는 방식으로 지식을 조직하였다. 이에 관해 구체적인 주제를 예로 들어 설명하면 다음과 같다. “Research in the cure of Tuberculosis of lungs by X-ray conducted in India in 1950”이라는 제목의 주제 또는 자원의 제목을 가정해 보자. 콜론분류와 같은 패킷 분석을 이용한 주제의 패킷 분석은 주제의 상이한 패킷을 다음과 같이 표현한다:

Medicine:Lungs:Tuberculosis:Treatment:X-ray:Research.India'1950

위의 결과는 개념을 표현하는 용어와 용어 간의 관계를 패킷 기호를 사용하여 합성한 것이다. 즉 분류의 대상이 갖는 주제의 의미 있는 측면을 모두 표현할 수 있다는 것을 보여주고 있다. 앞 절에서 언급한 특정성과 망라성을 모두 만족시킬 수 있는 수준 높은 색인 방법으로 웹 자원을 조직하는데 활용될 수 있는 가능성이 이러한 측면에 있는 것이다. 그리고 패킷 분석은 복잡한 주제 내용에 적합하고 ‘역동적인(dynamic)’ 색인 도구라 할 수 있다.

### 3. 2 패킷 분석 이용의 장점과 단점

먼저 패킷 분석을 웹 자원의 조직 수단으로 이용할 경우 장점은 아래와 같다.

첫째, 학제적인 주제에 대해서 유연성을 제

공 한다. 패킷 분석에 의한 색인 기법에서는 기본 개념만이 전개되어 있기 때문에, 문서가 학제적 관점을 가진다고 할지라도, 색인자는 특정한 항목을 위해 작성한 색인어를 생성하기 위해서 어떤 개념도 다른 개념과 조합할 수 있고, 패킷 간의 조합은 규칙적이고 예측 가능하다.

둘째, 새로운 개념을 편입하기가 용이하다: 새로운 주제가 전개될 때, 열거형 체계를 사용하는 색인자는 그 체계가 해당 주제를 위한 용어를 제공할 때까지 기다려야 하지만, 패킷 분석을 사용할 경우, 기본 개념으로 시작하여 그 주제를 완성해 나가기 때문에, 새로운 주제를 합성하는데, 이미 존재하는 용어들을 조합하는 것은 훨씬 용이하며, 사전에 명세 되지 않은 복잡한 개념도 구조화 할 수 있게 된다.

셋째, 주제의 특정성을 표현할 수 있다: 패킷 분석을 사용함으로써 표현할 수 있는 주제의 내용이 기존 분류 체계를 활용한 색인 기법보다 더 세세하고 특정하다.

넷째, 결합 순서(citation order)가 반드시 필요한 것은 아니다: 이용자가 패킷의 어떤 조합으로도 문헌을 검색할 수 있기 때문에 분류 번호의 결합 순서는 필요하지 않으며, 이러한 점이 오히려 주제들이 광범위하게 분산되어 있는 인터넷상에서 특히 유용하게 된다.

다섯째, 다른 패킷을 추가할 수 있다: 목표 집단, 언어, 관점과 같은 문서의 범위를 나타내는 부가적인 패킷을 용이하게 추가할 수 있다 (Patel 2002). 예를 들면 웹 자원의 멀티미디어와 관련된 특성을 독립된 패킷으로 범주화할 수 있게 된다.

이처럼 패킷 분석을 웹 자원의 조직 수단으

로 이용할 경우, 특정한 정보의 검색 및 학제적 성격을 띤 정보의 검색 요구와 관련된 문제에 대처할 수 있으며, 패킷 분석의 고유한 유연성과 역동성이 웹 환경에 보다 적합할 것으로 보인다. 그리고 기존의 열거형 분류 체계가 전통적인 도서관 자료 분류로서 갖는 폐쇄성이 웹 자원을 조직하는 수단으로 활용되는데 문제가 되지 않음을 알 수 있다.

패킷 분석의 단점은 아래와 같다:

첫째, DDC와 LCC와는 달리 이용자들 대다수에게 매우 낯설다.

둘째, 매우 복잡하고 경험이 없이는 사용하기가 어렵다.

셋째, 분류표의 계층 구조를 브라우징할 수 있도록 통합하기가 어렵다.

이러한 단점을 해결하기 위한 방안으로는 단계적으로 이용자의 정보 요구를 구체화시키는 과정을 적용하는 것이다. 이용자에게 먼저 패킷을 설명하고 포괄적인 주제에서 구체적인 주제로 주제의 범위를 조정하면서, 한 번에 하나의 주제를 입력하도록 이용자에게 제시함으로써 상호 작용이 가능한 이용자 인터페이스를 제공한다면 이러한 문제점들을 해결할 수 있다. 웹에서 정보를 검색할 때 패킷 분석은 이용자가 자신의 요구를 보다 정확하게 표현하도록 인터페이스를 제공하고, 패킷의 어떤 조합을 사용해서도 문서를 검색할 수 있는 자유를 이용자에게 줄 것이다.

## 4. 패킷 분석을 이용한 구체적인 방법론

### 4.1 패킷 기반 색인 시스템

웹 자원을 조직하는 방법은 색인에 의한 방식과 분류 체계에 의한 방식으로 나눌 수 있다. 전자는 키워드 검색이며 후자는 디렉토리 서비스로 검색 엔진이 이용자에게 제공한다. 그리고 색인을 이용한 조직은 단어 기반 기법(word based methods)과 개념 기반 기법(concept based methods)으로 나눌 수 있다. 최근에는 단어 동시 출현 기법을 사용하여 부분적으로 의미와 관련된 문제를 해결하려는 검색 엔진이 있지만, 단어 기반 색인 기법은 문서에 기록된 단어를 자동 색인 기법을 사용해서 색인을 작성하는 것으로써, 원 문서로부터 추출한 용어의 정확한 의미는 고려하지 않는다. 그러나 개념 기반 색인 기법은 문서에 사용된 용어를 그대로 추출하는 것이 아니라 문서에 사용된 용어가 나타내는 개념 확인, 즉 용어의 의미에 초점을 맞추는 것이다. 그리고 이 개념 기반 색인 기법은 상이한 개념간의 관계를 나타내기 위해 상호 참조나 관련 있는 주제들을 군집함으로써 구조화한다.

주제 색인에서 개념의 생성은 인간의 마음 속에서 개념을 형성하는 것과 유사하다. 패킷 분석을 이용한 색인 기법은 인간이 개념을 인지하고 분류하는 자연범주화 모형을 기반으로 함으로써 개념 색인 기법의 방법론으로 활용될 수 있다. 여기에서는 개념 기반 기법에 의거한 패킷 분석을 이용한 분류 및 색인 기법을 적용한 연구들을 살펴보고 웹상에서의 적용 가능성

을 평가하고자 한다.

#### 4. 1. 1 색인 기법

디바다손은, 패킷 분석 개념에 초점을 맞춘 랑가나단의 연쇄 색인 시스템(Chain Indexing System)에서 시작하여, 바타카리아(Bhattacharyya)의 POPSI(Postulate based Permuted Subject Indexing)를 토대로 한 DSIS(Deep Structure Indexing System)를 이용하여 웹 자원 조직에 관한 연구를 하였다(Devadason et al. 2002). 이 색인 시스템은 주제명표목을 구성하는 개념을 학문영역(Discipline; Base), 개체(Entity; Core), 속성(Property), 행동(Action)의 기초 패킷으로 범주화할 수 있다는 전제로 출발한다. 즉 자원과 관련된 모든 내용과 형식을 패킷으로 범주화하고 패킷 간의 합성에 의해서 자원을 색인하고 조직하는 방법을 적용하였는데, 이 색인 기법을 구체적으로 설명하면 다음과 같다.

이 색인 기법에서는 위에 제시한 기초 패킷과는 독립적으로 존재하는 특수한 구성 요소로 한정어(Modifier)로 자원을 색인 한다. 이 한정어는 개념의 유형을 지시하는데 사용되며,

모양(Form), 시간(Time), 장소(Place)와 환경(Environment) 등의 공통 한정어와 기초 패킷으로부터 도출된 특수 한정어(Special Modifier)로 구분되며, 이 특수 한정어는 독립 한정어(Independent Modifiers)와 종속 한정어(Dependent Modifiers)로 나뉘어 진다. 또한 개념이 어느 기초 패킷과 하위 패킷에 속하는지를 표시하는 숫자 기호를 배정하는데, 기호의 예는 아래의 표 1과 같다.

하위구분(subdivisions)은 기초 패킷이나 한정어 모두에 적용할 수 있다. 구체적인 예는 아래의 주제명표목 생성 과정에 제시하였다.

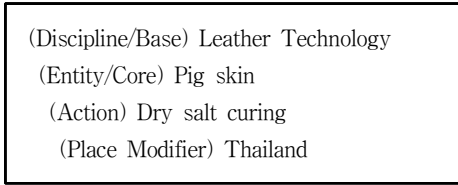
자원에서 다루어지는 구체적인 주제들이 주제명표목을 작성하는 근거가 되며 자원의 제목을 예로 들어서 주제명표목을 생성하는 기본 규칙을 제시하면 다음과 같다.

“In leather technology, dry salt curing of pig skin in Thailand”라는 제목을 분석하여 이들 구성 요소가 속하는 기초 패킷과 하위 패킷을 식별한다. 주제명표목을 구성하는 용어들은 아래의 문장 규칙에 따라 정규화된 문장은 그림 1로 표현된다.

이들 주제명표목을 구성하는 용어들은 상위 용어를 찾기 위해 분석된다. 해당 요소가 전체

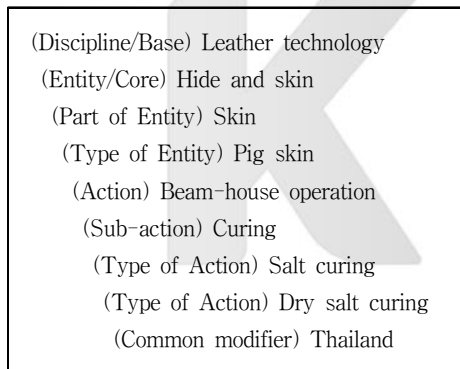
(표 1) 기초 패킷, 하위 패킷 및 한정어의 기호

기호	기초 패킷 및 한정어	기호	하위 패킷 및 한정어
0	Form Modifier	.3	Constituent
2	Time Modifier	.4	Part
3	Environment Modifier	.5	Modifier of Kind 1 including Phase Relation Modifier
4	Place Modifier	.6	Species/Type, including those created by Modifiers of Kind 2
9	Discipline/Base		
8	Entity/Core Object		
.1	Action		
.2	Property		



(그림 1)

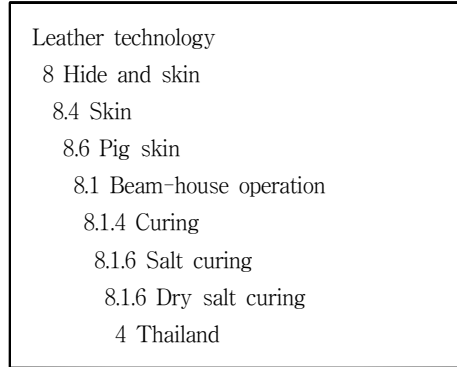
주제의 맥락에서 종/유형(species/type)인지 부분(part)인지 구성요소(constituent)인지를 밝혀내는 것이다. 상위 용어를 확인하는 과정은 기초 패킷의 개념에 이를 때까지 계속되는데, 이러한 목적을 달성하기 위해서 시소러스, 용어 집, 사전과 같은 용어학적 근거들이 이용된다. 확인된 상위 용어 각각은 해당 용어보다 앞서 결정되고, 이어서 아래의 그림 2와 같이 조정된 주제명표목을 생성하게 된다.



(그림 2)

주제명표목의 구성 용어들은 표준 용어(standard terms; 통제어휘)로 대체되고 용어들은 등호(=)와 같은 적절한 기호와 함께 표준 용어에 추가된다. 이러한 목적을 이루기 위해서 시소러스와 분류표와 같은 어휘 통제 도구를 사용한다. 기초 패킷에 대한 적절한 기호, 하위패킷 그리고 여러 종류의 일반 한정어가 적절한

위치에 삽입된다. 그 결과로 생성된 주제명표목은 그림 3과 같다.



(그림 3)

이와 같이 해당 주제를 구성하는 개념들과 관련된 개념들을 함께 명세함으로써 자원의 특성을 높일 수 있을 것이며, 이용자가 어떤 용어로 검색하더라도 해당 자원을 이용할 수 있는 접근점을 다변화하는 효과를 발휘할 수 있을 것이다. 이상과 같은 색인 기법을 웹 자원 조직에 활용하기 위해서는 전통적인 용어열 색인을 보다 정밀하고 정확하게 적용하기 위한 연구와 함께 시소러스나 주제명 표목 등에 관한 연구가 병행되어야 한다.

#### 4. 1. 2 웹 자원을 위한 검색 시스템의 기능 및 구성 요소

POPSI/DSIS을 이용한 웹 자원을 위한 원형 시스템(prototype system)은 다음과 같은 기능을 가지도록 구축되었다.

- 시스템 기능(System functions)
- 시스템 직원 - 즉, 관리자(administrator), 매니저, 목록자, 색인자 - 접근 및 검색하는 시스템

이용자들은 아래와 같은 기능을 수행할 수 있다.

- 직원(Staff)
  - 이용자가 추천하는 사이트로부터 자원을 선정;
  - 새로운 자원을 데이터베이스에 추가;
  - 자원 기술(resources descriptions)을 요약;
  - 자원 요약이나 초록을 수정;
  - 자원을 색인;
  - 주제명 표목을 유지;
  - 표준 메타데이터에 의해 자원 기술을 수정 및 저장, 도치 색인 파일 생성;
  - 모니터 및 검토, 링크 체크 및 시스템 유지.
- 이용자(Users)
  - 계층적으로 제시된 주제명표목을 브라우징;
  - 탐색어를 사용해서 데이터베이스를 검색;
  - 주제 계층적 순서 내에서 검색된 레코드를 디스플레이;
  - 새로운 사이트를 추천;
  - 검색된 레코드에서 제공된 링크를 이용해서 원 자원에 접근.
- 시스템의 주요 구성요소(System components)
  - 직원 관리 모듈;
  - 텍스트 자동 요약 및 메타데이터 모듈;
  - 목록 및 색인 모듈
  - 데이터베이스 및 색인(DSIS) 생성 및 유지 모듈
  - 검색 및 브라우징 모듈;
  - 시스템 유지 모듈.

이와 같이 시스템을 운영하는 운영자, 시스템을 이용하는 이용자, 시스템 자체의 주요 구성 요소에 의거하여 실제 상황에서 자원을 색인하고, 검색하는 구체적인 과정 중에서 주요한 내용을 설명하면 아래와 같다.

먼저 색인 대상으로 선정된 자원은 시스템에 추가되고, 레코드는 자동 텍스트 요약과 함께 더블린 코어(Dublin Core) 데이터요소와 패킷 주제명표목으로 제시된다. 텍스트 자동 요약 모듈에 입력되는 세 값은 다음과 같다:

- 1) 의미 있는 것으로 간주된 단어의 최소 빈도;
- 2) 클러스터를 형성하는 의미 있는 단어의 최소 수;
- 3) 문장을 추출해서 요약을 생성하기 위한 클러스터의 두 의미 있는 단어간의 최대 거리.

자동 요약 모듈의 과정은 요약자가 텍스트 처리의 시작 단계에서 공통 단어를 제거하기 위해서 불용어 리스트를 사용함으로써 시작한다. 요약의 크기는 입력된 값에 따라 달라지는데, 애매한 문장을 해결하기 위해 요약을 편집할 필요가 생기기도 한다. URL을 구체적으로 명시함으로써 새로운 사이트를 직접 추가하는 것도 가능하다. 시스템은 그 자원에 자동적으로 접근해서, 메타데이터를 추출하고, 템플릿을 채우고, 요약을 위한 입력으로 취하고 요약을 추가한다. 그리고 주제명표목의 생성 과정에서 패킷 주제명표목은 입력 템플릿을 이용해서 수작업으로 입력한다. 레코드가 입력되고 시스템에 추가되면, 시스템은 레코드를 데이터베이스에 저장하고 검색을 위한 패킷 주제명표목으로부터 요구되는 색인을 준비한다.

실제로 검색어가 입력되면 시스템은 해당 용

어를 가지고 있는 패킷 주제명표목을 검색하고 그 용어들을 배열 순서에 따라 보여준다. 주제명표목이 한 페이지를 넘으면, 검색자는 이용자의 질문을 가장 잘 표현한 주제명표목을 선정하기 위해 여러 페이지를 브라우징 할 수 있거나 아니면 조합된 다른 용어를 입력할 수도 있다. 또한 브라우징을 용이하게 하기 위해 다양한 수준의 조직된 색인을 생성할 수 있도록 디스플레이를 변경하는 것이 가능하다. 그 순서는 이용자에 따라 달라질 수 있다. 이용자가 학자라면 주제명 용어열은 학문영역을 기준으로 제시될 수 있다.

그리고 자원이 색인과 검색 파일을 작성하기 위해 수집된 것이라면, 동의어를 제어하고 사용된 용어를 표준화하는 것도 가능하다. 검색자가 사용한 비표준 용어는 표준 용어로 변환할 수 있고, 검색자가 사용한 용어가 여러 계층을 동시에 포함하는 다계층적인(poly-hierarchical) 용어일 때는 용어가 출현하는 주제명 용어열의 최상위 용어가 제시됨으로써 이용자는 자신의 질의와 일치하는 적절한 용어열을 선택할 수 있다.

또한 의미 분해(semantic factoring)로 인해 동의어를 통합함으로써 시스템이 패킷 주제명표목에 나타나는 분해된 용어의 조합을 검색하는 것이 가능하다. 따라서 해당 표목을 정렬할 수 있고 용이한 브라우징을 위해 상위 수준의 색인을 제시할 수 있다. 전체 주제명표목을 제시하고, 이용자로 하여금 동시에 해당 용어를 사용해서 검색자가 의도한 의미와 일치하는 주제명표목을 선택하도록 함으로써 동음이의어의 문제를 해결하는 것이 가능하다.

따라서 전체로서의 자원을 위해서 그리고 의미 있는 하위 단위로서의 개별 자원을 위해서 이러한 주제명 용어열을 구축하는 일은 어휘 제

어 도구처럼 많은 시간을 필요로 할 것이고 다양한 유형의 용어학적 근거들이 참조되고 이용되어야만 할 것이다. 그리고 기계가 이해할 수 있는 정보의 의미론을 구현하려는 시맨틱 웹에서 의미론의 표현을 위해 제안한 온톨로지와 방법적인 측면에서는 상이한 면이 있고, 색인을 이용해서 웹 자원을 조직함에 있어서 인간적 요소를 강조하는 관점이라고 할 수 있다. 그러나 두 방법은 해결하려는 문제의 대상과 지향하는 목표나 구축하려는 웹 자원 조직 시스템의 기본적인 구조는 동일하다. 디바다손의 패킷 색인을 이용한 웹 자원 조직에서 사용한 방법론은 문헌 정보학이 이미 오래 전부터 이용해 온 전통적인 수작업과 관련되어 있지만 시맨틱 웹의 온톨로지가 지향하는 기계적 요소와 함께 접목시키려는 하이브리드(hybrid) 방식의 연구도 이루어져야 할 것이다.

#### 4.2 열거형 분류 체계를 패킷 분류 체계로 구조 수정

열거형 분류 체계를 대표하는 DDC의 분류 구조를 패킷 구조로 변형하여 온라인 정보 검색에 대한 접근을 개선하려는 시도가 있었다 (Pollitt, Tinker and Braekevelt 1998). 이 방법론은 기존 분류 체계의 계층구조가 주제를 단일한 구분 기준에 의해 나열하는 방식을 채택함으로써 관련된 정보를 분산시키는 비효율적인 결과를 낳았다고 주장한다. 다시 말하면, 동일한 주제에 관한 내용들이 상이한 학문영역에 분산되어 분류되고 조직된다는 것이다(김태수 2000).

이와 같이 정보가 분산되어 조직되면 다면적

이고 역동적인 속성을 지닌 지식이 단면적이고 정적인 분류 체계에 의해 조직됨으로써 이미 결정된 통로를 따라 정보를 검색하는 결과를 나타내게 된다. 이렇게 열거형 분류 체계가 제공하는 일차원적인 검색 방식으로는 지식의 다차원적인 특성을 제대로 반영할 수 없으며, 이용자의 다양한 정보 요구에도 대처할 수 없다. 따라서 지식의 다면적인 속성을 반영함과 동시에 이용자가 요구하는 방식으로 정보를 검색하는 것이 더 효율적인 검색 방식이라고 할 수 있다. 이러한 개념은 관점 기반 검색(view-based searching)을 통해 적용되었다.

관점 기반 검색 인터페이스에서는, 예를 들어 Dewey for Windows에서 'textiles'로 검색했을 때, 'Textiles'(677), 'Cleaning, color, coating, related technologies'(677), 'Quantitative analysis - Other methods'(545.8), 'West Yorkshire'(-4281, Table 2) 등의 결과를 얻을 수 있다. 이 같은 결과는 일차원적이고 정적인 인터페이스에서는 불가능하고 이용자는 한 번에 하나의 주류(main classes)만을 선택해서 동일한 위계구조 내에서 위아래 방향으로만 이동할 수 있다. 관점 기반 검색 시스템의 이용자는 한 패킷 내에서 디스크립터를 선택함으로써 관점을 좁힐 수 있고 동시에 관점을 넓힐 수도 있다. 이러한 선택은 다른 관점의 내용에도 영향을 주고, 디스플레이되는 폴더나 항목의 수를 변화시킬 수 있다. 관점 기반 검색에서는 일차원 공간에서 위아래 방향으로 자원을 검색할 수 있을 뿐 아니라 다차원 정보 공간으로 이루어진 무한한 수의 교차하는 통로를 이용하여 정보를 검색할 수 있다.

관점 기반 검색 인터페이스는 다차원적인 속

성을 가지고 있는 주제의 다양한 관점을 패킷 분석이라는 수단을 사용하여 보다 정확하게 조직하여 구조화했고, 특정한 자원에 접근하는 통로를 다변화시켰다는 평가를 내릴 수 있다. 학제적 연구와 다양한 디지털 속성을 갖는 웹 자원 조직에도 활용할 수 있는 방법이다.

#### 4.3 국내 검색 엔진의 주제별 디렉토리의 패킷 모형

2장에서 언급한 진스의 연구는 2001년 당시의 현상을 분석하였고 그 이후에 분석 대상인 분류 체계가 긍정적인 방향으로 개선되었다. 따라서 진스가 적용한 분석 방법을 국내의 검색 엔진에 적용하여 분류 모형을 분석했는데, 포털 사이트인 '네이버'의 최상위 디렉토리를 대상으로 하였다.

네이버에서 제공하는 주제별 디렉토리는 아래 표 2와 같다.

그리고 진스가 제시한 8개 모형으로 분석한 결과는 아래 표 3과 같다.

분석 결과 주제 모형이 가장 많았고, 매체 모형, 응용분야 모형, 지리 모형이 독립된 패킷으로 범주화되었으나, 주제 모형과 다른 모형들이 혼합된 패킷이 4개나 나타났고, '레크리에이션' 범주는 세 개 이상의 모형이 혼합된 것으로 나타났다. 그리고 언어 모형은 존재하지 않는 것으로 나타났다.

분석 결과에 의하면, 주제 모형이 가장 주요한 모형으로 나타났으나, 상이한 분류 구조를 갖는 모형들이 혼합된 범주들이 확인됨으로써 이러한 범주들은 보다 분명한 기준에 의해 조정되어야 할 것으로 보인다.

(표 2) 네이버의 주제별 디렉토리

뉴스, 미디어	엔터테인먼트, 예술
비즈니스, 경제	컴퓨터, 인터넷
쇼핑	게임
가정, 여성	레크리에이션
사회, 문화	스포츠
학문, 과학	건강, 의학
교육, 참고자료	지역정보

(표 3) 진스 모형에 의한 분석 결과

모형의 유형	디렉토리 이름
주제모형	비즈니스, 경제/ 학문, 과학/ 스포츠/ 건강, 의학
응용분야 모형	쇼핑/ 게임
지리 모형	지역정보
매체 모형	뉴스, 미디어/컴퓨터, 인터넷
주제와 이용자 모형의 합성	가정, 여성
주제와 참고 모형의 합성	교육, 참고자료
주제와 응용분야의 합성	엔터테인먼트, 예술
주제와 대상 모형의 합성	사회, 문화
응용, 대상, 매체 등의 합성	레크리에이션

이상에서 살펴 본 내용을 요약하면 다음과 같다. 패킷 분석을 이용한 웹 자원 조직 방법은 시소러스, 자동분류 및 자동요약, 그리고 자동 색인과 수작업 색인을 모두 사용하여 조직하는 하이브리드 색인 기법을 포함하는 매우 다양한 영역의 공동 연구를 필요로 한다. 그리고 패킷 분석 기법을 적용하여 기존의 열거형 분류 체계를 수정하는 연구는 DDC에 익숙해져 있는 실무 환경에 적용해 볼 만한 대안이 될 수 있다. 또한 상업적 포털 뿐 만 아니라 분류 체계를 이용해서 자원을 조직하고 검색하는 서비스를 제공하는 모든 웹사이트에서 8개의 모형을 통합하여 패킷 분석을 적용한다면 보다 논리적이고 일관성 있는 색인 시스템을 구현할 수 있을 것이다.

## 5. 결론

량가나단은 용어의 의미와 문맥은 그 용어가 색인시스템 내에서 어디에 위치하는가에 달려 있다고 하였다. 즉 어느 패킷에 속하느냐에 따라 그 개념의 고유한 의미가 결정된다는 사실을 지적하고 있다. 자원을 제공하는 시스템이 이용자가 의미하는 요구에 일치하는 자원을 제공하기 위해 이용자의 다양한 관점을 반영할 수 있어야 한다는 뜻이다.

이러한 관점에서 웹을 하나의 거대한 도서관이라고 가정할 때, 량가나단의 도서관 5 법칙의 의미가 새롭게 해석되어야 할 것이다. 특히 모든 이용자에게 그 이용자가 원하는 자원을 제공해야 한다는 제 2 법칙과 웹은 성장하는 유기체

와 같은 역동적인 존재라는 제 5 법칙을 되새겨 보아야 할 것이다.

웹 환경은 매우 빠르게 변화하면서 성장하고 있다. 특히 거대한 정보량과 광범위하게 분산되어 있는 자원을 색인하고 조직할 적합한 수단을 필요로 한다. 그 수단은 다양한 이용자의 정보 요구를 만족시킬 수 있는 표준과 체계를 갖추어야 하고, 학제적 정보를 정확하게 검색할 수 있어야 하며, 웹 자원의 특징을 반영할 수 있어야 한다.

또한 시맨틱 웹에서 자원의 지적 내용을 해결하는 가장 중요한 수단으로 제안한 온톨로지에 대한 이해와 패시 분석에서 구현하려고 하는 구조가 밀접한 관련을 가지고 있다. 베르너스-리(Berners-Lee's 2001)에 의하면, 분류 체계(taxnomy)는 류(classes)의 대상과 류간의 관계를 정의한다. 그리고 상위류, 하위류 그리

고 개체간의 관계는 웹 활용을 위한 가장 강력한 수단이고, 정보 교환을 가능하게 하는 기호를 가지고 있는 분류 체계의 활용이 필수적이라고 주장한다. 또한 보다 보편적인 언어를 사용함으로써 하위문화(subcultures)의 결합이 본질적인 과정이 될 수 있다고 하였다. 결과적으로 분류 체계와 온톨로지는 동일한 구조를 가지는 동시에 보편적이고 표준적인 어휘체계를 구축함으로써 정보 자원간의 상호운용성을 확보하는 수단으로 활용할 수 있을 것이다.

베르너스-리의 관점은 패시 분석을 통해 해당 주제 영역의 개념을 추출하고 그 개념을 표현하는 용어를 사용하여 웹 자원을 조직하는 통제 어휘 체계를 구축하는 것과 다르지 않다. 앞으로의 연구 과제는 패시 분석을 이용한 통제 어휘 체계를 XML을 기반으로 하는 시스템에 적용하는 것이라고 할 수 있다.

## 참 고 문 헌

- 김성희. 2002. 인터넷상의 디지털문화컨텐츠 분류 방안에 관한 연구. 『한국문헌정보학회지』, 36(3): 181-200.
- 김정현, 문지현. 2002. 농학분야 인터넷 자원의 분류 체계에 관한 연구. 『한국도서관정보학회지』, 33(3): 393-413.
- 김태수. 2000. 『분류의 이해』. 서울: 문헌정보처리연구회.
- 이란주, 성기주, 양정하. 2001. 여성학분야 인터넷 자원의 분류에 관한 연구. 『한국도서관정보학회지』, 32(3): 397-417.
- 이창수. 2000. 정보통신기술분야 인터넷 자원의 분류 체계에 관한 연구. 『한국도서관정보학회지』, 31(4): 111-138.
- 정영미. 1997. 『지식구조론』. 서울: 한국도서관협회.
- 최재항. 1998. 인터넷 학술정보 자원의 디렉토리 서비스 설계에 있어서 DDC 분류 체계의 활용에 관한 연구. 『정보관리학회지』, 15(2): 47-67.
- 최희운. 1998. 인터넷 정보서비스의 분류 체계에 대한 비교 연구: 물리학을 중심으로. 『정보관리학회지』, 15(3): 45-71.
- Berners-Lee, Tim, J. Hendler and O.

- Lassila. 2001. The Semantic Web. *Scientific American* 284(5) [cited 2004. 3. 20]  
<<http://www.sciam.com/article.cfm?chanID=sa006&colID=1&articleID=00048144-10D2-1C70-84A9809EC588EF21>>
- Broughton, Vanda. 2001. Faceted Classification as a Basis for Knowledge Organization in a Digital Environment; the Bliss Bibliographic Classification as a Model for Vocabulary Management and the Creation of Multidimensional Knowledge Structure. *The New Review of Hypermedia and Multimedia* 7: 67-102.
- Broughton, Vanda. 2002. Facet Analytical theory as a Basis for a Knowledge Organization Tool in a Subject Portal. Seventh International ISKO Conference. Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge across Boundaries. 10-13 July 2002. [Spain, Granada].
- Devadason, F. J., N. Intaraksa, P. Patamawongjariya, and K. Desai. 2002. Faceted Indexing Based System for Organizing and Accessing Internet Resources. *Knowledge Organization* 29(2): 65-77.
- Ellis, David, and Ana Vasconcelos. 1999. Ranganathan and the Net: Using Facet Analysis to Search and Organise the World Wide Web. *Aslib Proceedings* 51(1): 3-10.
- Fensel, Dieter, Handler, J., Lieberman, H., and Wahlster, W. 2003. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. Cambridge : MIT Press.
- Hudon, M. 2000. Innovation and Tradition in Knowledge Organization Schemes on the Internet, or, Finding One's Way in the Virtual Library. Proceedings of the Sixth International ISKO Conference. [Wurzberg, Germany : Ergon Verlag].
- Iyer, Hemalata. 1995. Classificatory Structures: Concepts, Relations and Representation. Textbook for Knowledge Organization, vol. 2. Indeks Verlag, Frankfurt/Main.
- Koch, Traugott. 1997. The Role of Classification Schemes in Internet Resource Description and Discovery. DESIRE Specification for Resource Description Methods Part 3. [cited 2004. 3. 20]  
<<http://www.ukoln.ac.uk/metadata/desire/classification/>>
- Maple, A. 1995. Faceted Access: a Review of the Literature. Paper Presented at the Music Library Association

- Annual Meeting 10 February. [cited 2001. 9. 11]  
<<http://www-sul.stanford.edu/depts/music/mlatest/BCC/BCC-Historical/95WGFAM2.html>> 재인용: Broughton Vanda. 2002. Facet Analytical Theory as a Basis for a Knowledge Organization Tool in a Subject Portal. Seventh International ISKO Conference. Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge across Boundaries. 10-13 July. [Spain, Granada].
- McKiernan, G. 1997. Hand-made in Iowa: Organizing the Web Along the Lincoln Highway. *D-Lib Magazine*. [cited 2004. 2. 7]  
<<http://www.dlib.org/dlib/february97/02mckiernan.html>>
- McKiernan, G. 1998. Beyond Bookmarks: a Review of Frameworks, Features, and Functionalities of Scheme for Organizing the Web. *Internet Reference Services Quarterly* 3(1): 69-77.
- McKiernan, G. 1999. Points of View: Conventional and 'Neo-conventional' Access and Navigation in Digital Collection. *Journal of Internet Cataloging* 2(1): 23-41.
- McKiernan, G. 2000. Beyond Bookmarks: Schemes for Organizing the Web. [cited 2004. 4. 21].  
<<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>>
- Patel, Dimpel. 2002. Organizing the Web: a Faceted Approach. Workshop on Information Resource Management. 13th-15th march. [DTRC, Bangalore].
- Pollitt, A. Steven, Amanda J. Tinker and Patrick A. J. Braekevelt. 1998. Improving Access to Online Information Using Dynamic Faceted Classification. *Online Information 98 Proceedings* 17-21.
- Ranganathan, S. R. 1951. Classification and Communication. Delhi University Publications, Library Science Series, 3. Delhi : University of Delhi.
- Ranganathan, S. R. 1965. A Descriptive Account of the Colon Classification. [reprinted 1990] Bangalore : Sarada Ranganathan Endowment for Library Science.
- Saeed, H. & A. S. Chaudry. 2001. Potential of Bibliographic Tools to Organize Knowledge on the Internet: the Use of Dewey Decimal Classification Scheme for Organizing Web Based Information Resources. *Knowledge Organization* 26(2): 80-96.
- Spiteri, Louise. 1998. A Simplified Model

- for Facet Analysis. *Canadian Journal of Information and Library Science* 23: 1-30.
- Svenonius, Elaine. 1983. Use of Classification in Online Retrieval. *Library Resources and Technical Services* 27(1): 76-80.
- Woodward, J. 1996. Cataloging and Classifying Information Resources on the Internet. *Annual Review of Information Science and Technology* 31: 189-220.
- Zins, Chaim. 2002. Models for Classifying Internet Resources. *Knowledge Organization* 29(1): 20-28.

K C I

к с і