

인터넷에서 메타데이터 기반구조 PICS를 이용한 협력적 필터링에 관한 연구

A Study on the Collaborative Filtering Using PICS as a Metadata Infrastructure on the Internet

최운선, 연세대학교 대학원 문헌정보학과
문성빈, 연세대학교 문헌정보학과

Choi Yun-Seon, Moon Sung-Been
Dept. of Library and Information Science, Yonsei, Univ.

본 연구에서는 인터넷 환경의 정보검색 문제를 해결하기 위한 방안으로 메타데이터 기반구조인 PICS를 분석하여 인터넷 자원의 검색 및 필터링에 적용할 수 있는 요인들을 제시하였다. 또한 시스템의 실제 구현을 도모하기 위해 의학 정보를 위한 'OMNI-PICS'라는 등급체계의 프로토타입을 설정하고, 기계가독형 RAT파일을 설계하였다. 따라서 웹에서 특정 주제 영역을 다루는 주제게이트웨이들이 자원을 협력적으로 필터링하기 위한 표준 포맷으로 PICS 등급체계를 적용하여 운용할 수 있는 구체적 지침을 제공하였다.

1 서론

인터넷에서 정보자원을 검색할 때 고려해야 할 점은 자동탐색엔진의 문제점인 정보과부하 뿐만 아니라, 정보의 품질문제까지도 생각해야 한다. 따라서 자원은 체계적으로 조직되어야 하며, 그 품질에 대한 평가도 정확해야 한다. 또한 그러한 과정은 표준화된 등급체계를 기반으로 하여 저자와 전문가, 사서 및 이용자간에 협력적으로 이루어져야 할 것이다.

본 연구의 목적은 메타데이터 기반구조인 PICS를 분석하여 웹 정보자원의 검색 및 필터링에 대한 적용가능성을 검토하고 그 요인들을 구체적으로 제시함으로써, PICS 확장구조를 바탕으로 하는 분산된 협력적 필터링 시스템을 제안하는 것이다. 또한, 제안한 시스템 구조에서 동일한 주제 분야를 다루는 주제게이트웨이들의 협력적 필터링을 가정하여, PICS 등급체계에 대한 프로토타입을 구축하고 기계가독형 RAT파일을 설계하고자 한다.

2 이론적 배경

메타데이터는 '도서관 목록'의 개념에서 '전자자원에 대한 데이터'라는 대용물의 의미로 전환된 개념이다. 이러한 메타데이터는 데이터로서 저장될 수 있고, 자원에 삽입되거나, 또 다른 문헌에 대한 정보를 포함할 수도 있다. 즉, META 태그를 이용하여 HTML 문헌에 삽입될 수 있으며, 표준 프로토콜 Z39.50 등을 통해 자원과 분리되어 접근되고, 제삼자의 등급(ratings)이 결합될 수도 있다. 그러므로, 자원의 기술사항을 목록하는 메타데이터에 자원에 대한 평가요소가 결합된다면, 검색되는 정보자원의 품질과 이용자 만족도는 더욱 향상될 것이다.

3 메타데이터 기반구조 PICS

3.1 PICS(Platform for Internet Content Selection)

월드 와이드 웹 컨소시엄(World Wide Web

Consortium : W3C)에서 개발된 PICS는 인터넷 자원과 레이블(label), 즉 메타데이터를 결합하기 위한 기반구조이며, 이러한 레이블은 자원의 기술사항(description) 정보를 제공한다. 초기 PICS1.1 규격은 인터넷의 부적절한 정보에 대한 접근을 제어하기 위한 것이었다. 이는 단순히 내용필터링만을 목적으로 개발된 것으로 수치적 평가값을 기반으로 하였지만, PICS2.0의 확장구조는 비수치값으로 문자열인 메타데이터를 수용하고, 내용(content)을 필터링하는 한정된 용도의 범위를 확장함으로써, 디지털 도서관에서 양질의 정보자원(resources)을 선별하여 필터링할 수 있도록 발전되고 있다.

3.2 PICS 기술 규격

PICS 규격은 다음의 세 가지를 정의한다.

① 레이블 구문과 전송(Label Syntax and Communications) : 레이블의 형식과 전송 방법을 정의하며, 레이블은 서비스 식별기호(service identifier), 레이블 옵션(label options), 등급(rating)으로 구성된다.

```
(PICS-1.1 "http://old.rsac.org/v1.0" labels ⇒ 서비스
식별기호
on "1994.11.05T08:15-0500" ⇒ 레이블 옵션
until "1995.12.31.T23:59-0000"
for "http://www.gcf.org/stuif.html"
by "John Doe"
ratings (( 3 s 2 v 0)) ⇒ 등급(수치값)
language=3, sex=2, violence=0
```

<그림 1> PICS 레이블

② 등급서비스와 등급체계(Rating Services and Systems) : 등급서비스는 인터넷 자원에 대한 레이블을 제공하는 개인이나 기관을 의미한다. 한편, 등급체계는 레이블링에 사용되는 범주를 규명하며, 각 범주에 대한 척도 및 값을 부여하는데 사용되는 기준을 기술한다.

③ PICSRules : PICS 레이블이 기술하는 URL에 대한 접근을 허용하고 차단하는 필터링 규칙을 표현하기 위한 언어이다.

3.3 PICS 확장구조

W3C는 새로운 작업그룹인 PICS-NG 작업그룹(PICS Next Generation Working Group)을 형성하였다.

```
(PICS-2.0 (service "http://metadata.net/DC/V1.0") ⇒ 서비스
(label ⇒ 식별자
(for "http://www.marilyn.net/book/" ⇒ 레이블 옵션
by "Renato Iannella <renato@dstc.edu.au>."
on "1997.01.01T08:15-1000"
until "2000.12.31T23:59-0000")
(label (name "title")
type/name "Photographs of ⇒ 등급정보
Marilyn Monroe and Others"(문자열 데이터)
language/en)
(label (name "creator")
type/name "Paul Flora" ⇒ 등급정보
type/affiliation "Acme Organization"
type/phone "+61 555 555 555"
language/en )))
.....생략
```

<그림 2> PICS2.0 레이블

4 PICS와 인터넷 자원의 협력적 필터링

4.1 협력적 필터링

협력적 필터링(collaborative filtering)은 다른 사람의 '등급'을 이용하여 자원을 필터링하는 기법이다. 이러한 등급의 목적은 자원의 품질을 증가시키거나, 특정 분야에서 부적절하게 여겨지는 특정 자원을 피할 수 있게 하기 위한 것이다.

따라서, 협력적 필터링 기법은 인터넷의 정보과 부하 문제를 해결하며, 인터넷 자원을 선별하는데 유용하게 사용될 수 있을 것이다.

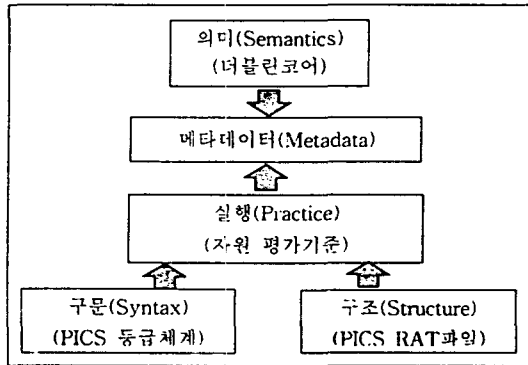
4.2 PICS를 이용한 정보자원 필터링 기법

인터넷 자원을 필터링하는데 PICS 확장구조를 이용할 수 있는 요인들은 다음과 같다.

첫째, 웹 환경의 다양한 자원 유형을 고려한 필터링의 중요성에 주안점을 두었다. 즉, 웹 자원을 색인하고 기술하는 메타데이터를 다룰 수 있어야 한다. PICS는 <그림 3>과 같이 그러한 메타데이터

를 전송하는 기반구조로서, 메타데이터의 상호운용성(interoperability)을 제공할 수 있을 것이다.

지 않는다는 이점이 있으며, 레이블을 전자적으로 발행할 수 있으므로 실제적인 운용을 활성화시킬 수 있을 것이다.



<그림 3> PICS의 상호운용적 구조

5 PICS 기반 협력적 필터링 시스템

5.1 시스템 구조

본 연구에서 제안하는 시스템은 PICS 확장구조의 표준화된 등급체계를 기반으로 하며, 여러 등급 서비스들로 구성되는 협력적 필터링 시스템이다.

시스템은 <그림 5>와 같이, PICS 레이블을 생성하는 저자와 제삼자 기관인 등급서비스, 그리고 레이블을 독립적으로 저장, 관리, 배포하는 레이블국으로 구성된다.

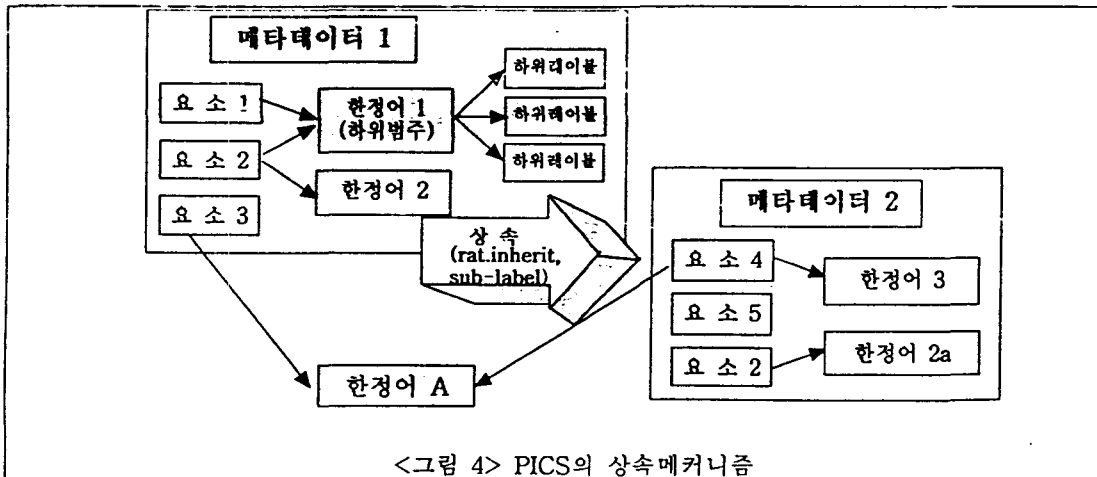
첫째, PICS 레이블은 저자가 작성하는 기술레이블과 등급서비스들에 의한 평가레이블로 구성된다. 이러한 레이블은 정보검색의 정확률을 높이고, 특정한 선별기준에 따라 탐색을 제한할 수 있으므로 재현율이 향상될 수 있다. 이러한 구조에서 전문가, 사서, 협회 및 기관 등은 분산화된 등급서비스로서 자원을 협력적으로 필터링하는 역할을 할 수 있다.

둘째, 레이블국은 등급서비스의 레이블을 독립적으로 저장, 관리하며, 이용자들은 이를 통해 여러 등급서비스들의 레이블을 참고할 수 있으므로 한 기관의 견해에 의한 독단적인 평가 메커니즘과 자원 품질의 편중성이 해결된다.

PICS의 상속 메커니즘(그림 4)은 메타데이터 스키마들간의 의미상속에 적용되어, 데이터의 중복이 방지되고 공통요소의 공유가 가능해진다. 또한, 이것은 PICS가 기존의 메타데이터 스키마를 상속하여 계속 확장될 수 있는 가능성을 보여주는 것이다.

둘째는 기존 검색시스템의 문제점을 해결할 수 있는 협력적 필터링의 구현가능 여부이다. 이는 PICS의 제삼자 레이블링으로 가능하다.

셋째는 그러한 PICS 기반 시스템의 실제 운용 가능성 측면이다. PICS는 범용적이고 가치중립적인 특성을 갖고 있어 특정 목적이나 영역에만 한정되



<그림 4> PICS의 상속메커니즘

또한 이용자 수준에서 이루어지는 필터링에 의한 피드백 정보는 등급서비스의 레이블 생성과정에 제공되어, 등급서비스 기관들과 이용자의 협력적인 등급체계 구조를 이룰 수 있다.

셋째, 자원평가는 두 단계, 즉 제삼자의 평가레이블 생성과정과 이용자 수준의 필터링에서 협력적으로 이루어지므로, 이용자는 정보자원의 품질을 신뢰할 수 있을 뿐만 아니라, 선별기준에 따라 탐색수준을 제한하여 정보자원을 선별할 수 있게 된다.

5.2 OMNI-PICS 등급체계와 RAT파일 설계

5.2.1 설계 범위와 원칙

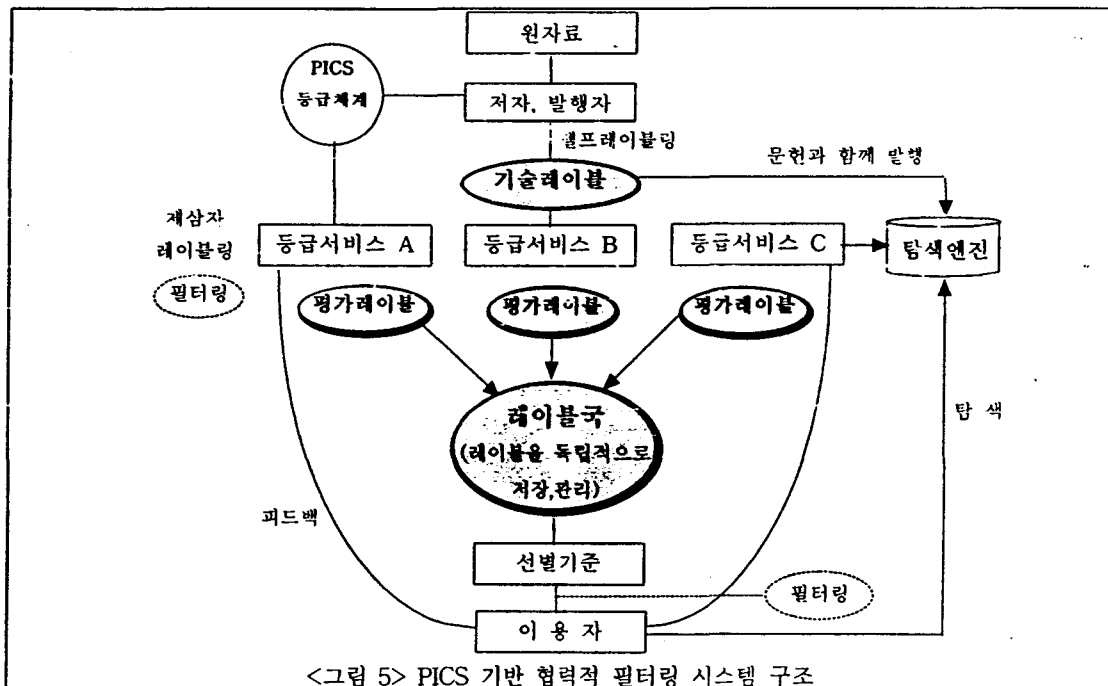
본 연구에서는 PICS 등급체계의 프로토타입을 위한 주제영역으로 의학정보 분야를 선정하였다. OMNI-PICS 등급체계는 PICS 확장구조를 기반으로 하여 두 단계 즉, 문자열 데이터인 더블린코어를 기술범주로 수용하고, 추가적으로 평가범주를 설정하여 구축하였으며, 의학정보 평가서비스인 OMNI(Organizing Medical Networked Information) 주제게이트웨이의 자원 평가기준을 참고하여 구축

하였다. OMNI-PICS 등급체계에 대한 기계가독형 RAT파일은, 먼저 기술범주로 사용된 더블린코어 RAT파일을 설계하고, PICS 확장구조를 이용하여 이를 상속하였다.

6 결론 및 제언

정보에 대한 접근과 평가는 통제적이고 하향적인(top-down) 구조가 아닌, 분산적이고 상향적인(bottom-up) 구조에서 이루어져야 할 것이다. 또한, 정보를 검열하거나 접근을 차단하는 개념보다는 이용자의 특정한 관심과 요구에 적합한 양질의 정보 탐색을 지원할 수 있는 방안에 초점이 두어져야 할 것이다.

메타데이터 기반구조로서 PICS의 사용은 디지털 도서관 사회를 촉구하게 될 메타데이터의 생성과 운용을 위한 기회를 제공하는 것으로, 이를 활용한 전자자원의 조직과 검색에 대한 지속적인 연구가 필요할 것이다.



<그림 5> PICS 기반 협력적 필터링 시스템 구조