

# 인터넷 내용 선별을 위한 플랫폼



인터넷이 새로운 매체로 등장하면서 유통되는 정보의 질적 평가가 중요한 이슈가 되고 있다. 특히 폭력, 음란, 선정성을 다룬 내용물이 인터넷을 통해 여과없이 유통되면서 어린이, 청소년들에게 악영향을 끼치고 있다. 따라서 적합하지 못한 정보를 차단하고 선별할 수 있는 체계와 기술 개발은 이러한 인터넷의 부작용을 해소할 수 있는 방안이 될 수 있다. 본 호에서는 인터넷 내용 선별을 위한 체계로서 관련 소프트웨어의 사실상 표준으로 자리잡고 있는 PICS를 소개하고 새로운 메타데이터 시스템으로서의 발전 방향을 살펴본다.

이재진/ 한국데이터베이스진흥센터 정책연구과

## 연재 순서

1. 메타데이터의 개요
2. DC(Dublin Core)
3. GILS(Government Information Locator Service)
4. IAFA Templates
5. MARC
6. PICS(Platform for Internet Content Selection)..... 이번호
7. RFC 1807
8. SOIF
9. TEI header
10. URC(Uniform Resource Characteristics)
11. Warwick Framework RDF(Resource Description Framework)
12. 메타데이터 향후 방향

## 1. 개요

인터넷 이용의 폭발적인 증가는 다양한 이용자층의 확대와 인터넷 내용의 양적 팽창과 다양화를 의미한다. 그러나 모든 인터넷 내용이 모든 이용자에게 적절한 것이 될 수는 없으며 특히 인터넷에는 어린이와 청소년의 정서 개발과 교육상 부적절한 내용이 다수 포함되어 있다.

이러한 문제점을 직시한 W3C의 인터넷 전문가들은 1995년 8월부터 어른과 어린이, 모두를 위해 이용자가 ① 적절한 내용을 쉽게 찾을 수 있고 ② 부적절하거나 원치 않는 내용은 거부할 수 있도록 하는 기술 규격의 개발에 착수하였다. 이렇게 하여 1996년초에 인터넷 내용 선별을 위한 체계로서 PICS(Platform for Internet Content Selection)가 탄생하게 되었다.

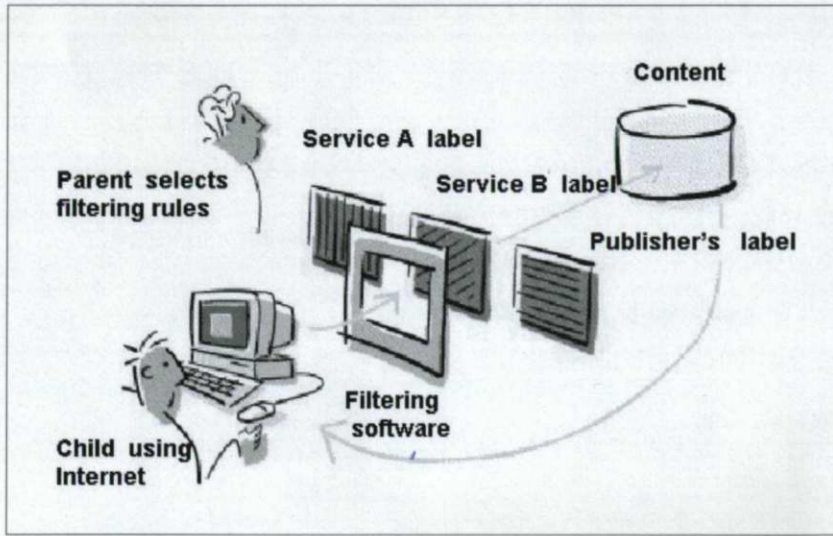
PICS는 부적절한 내용에 대한 접근 통제를 위해 만들어졌지만 레이블(label)을 이용하여 인터넷 자원을 기술하고 있기 때문에 일반적으로 '메타데이터' 시스템으로 간주된다. 즉, 레이블에서 사용하는 어휘는 어린이나 청소년에게 적절한지의 여부를 나타내기보다는 문서 내용의 질적 수준을 기술한다. 이러한 PICS 레이블은 특정 자원을 효과적으로 탐색·검색하기 위해서도 사용될 수 있다.

## 2. 내용 선별 체계

PICS를 이용한 인터넷 내용 선별 과정은 <그림>로 설명할 수 있다.

필터링 소프트웨어는 어린이(혹은 인터넷 이용자)와 이용가능한 내용(content) 사이에 존재하여 어떤 내용에는 접근할 수 있도록 하고 어떤 내용에는





(그림) PICS의 내용 선별 과정

접근을 차단한다. 이 소프트웨어는 주로 두가지 정보원을 통해 접근과 차단 여부를 결정한다.

하나는 내용물을 기술(記述)한 레이블 집합으로서 이 레이블은 정보 생산자가 생성할 수도 있고 생산·이용자가 아닌 제3자가 생성할 수도 있다. 따라서 하나의 문서가 관련된 여러 가지 레이블을 가질 수 있다. 또다른 방법은 필터링 규칙을 사용하는 것이다. 이것은 어떤 종류의 레이블에 주의를 해야 하는지를 알려주는 것으로서 레이블의 특정 값에 따라 접근과 접근 불가의 판정을 내리게 된다.

PICS는 레이블에 대한 표준 포맷으로서 PICS와 호환이 가능한 모든 선별 소프트웨어는 마찬가지로 PICS와 호환이 가

능한 모든 레이블을 처리할 수 있다. 결국 <그림 1>에서와 같이 부모(소비자)는 필터링 규칙(등급 서비스)과 여러 레이블을 처리하는 필터링 소프트웨어를 선택하여 이용하게 된다.

### 3. PICS 기술 규격

PICS 기술 규격은 ① 등급 서비스의 어휘와 척도를 기술하기 위한 포맷인 '등급 서비스와 등급 체계(Rating Services and Rating Systems)', ② 레이블 구문과 레이블의 배포 방법의 포맷인 '레이블 구문과 통신 프로토콜(Label Syntax and Communication Protocols)', 그리고 ③ PICS 레이블이 기술하는 URL에 접근을 허용하거나 차단하는 필터링 규칙인 프로

파일 작성 언어를 정의한 PICS규칙(PICSRules)으로 구성되며 추가적으로 해당 문서에 레이블을 이용한 작성자의 디지털 서명을 삽입하기 위한 규격(PICS Signed Labels(DSig))있다.

#### (1) PICS 레이블

레이블은 문서 내용에 관한 정보를 기술하고 등급 정보, 검색 정보, 지적소유권 등을 표시할 수 있다. PICS의 레이블은 ① 서비스 식별기호(service identifier), ② 레이블 옵션(label options), 그리고 ③ 등급(rating)으로 구성된다. 서비스 식별 기호는 등급 서비스에 의해 선택된 URL을 가리키며 레이블 옵션은 등급 자체와 등급이 매겨진 문서에 대한 추가 정보를 제공한다. 등급은 문서를 기술한 속성(attribute)-값(value)의 집합이다. PICS 레이블의 예는 다음과 같다.

<예 1> PICS 레이블

```
(PICS-1.1 "http://www.gcf.org/v2.5"
→ 서비스 식별기호
by "John Doe"
labels on "1994.11.05T08:15-0500"
→ 레이블 옵션
until "1995.12.31T23:59-0000"
for "http://w3.org/PICS/Overview.html"
rating (suds 0.5 density 0 color/hue 1)
for "http://w3.org/PICS/Underview.html"
by "Jane Doe"
rating (subject 2 density 1 color/hue 1)
→ 등급(수치값)
```

레이블 옵션은 문서에 관한 정보, 레이블 자체에 관한 정보, 그리고 기타 정보의 세가지로 나눌 수 있다. 이 옵션을 유형별로 구별하여 설명해 보면 <표 1>과 같다.

PICS 레이블을 전송하는 방식은 3가지가 있다. 첫번째 방식은 HTML 문서 안에 삽입하는 방식이다. 이 경우, HTML 문서의 HEAD 요소안에서 META 태그를 사

<표> 레이블 옵션

레이블 옵션	유형	내용
문서 정보	at quoted-ISO-date	등급이 매겨진 항목의 최종 개정일
레이블 자체 정보	by quotedname	레이블 생성을 책임지는 등급 서비스의 사람이나 기관
	for quotedURL	등급이 매겨진 항목의 URL
	on quoted-ISO-date	등급이 매겨진 날짜
기타 정보	until quoted-ISO-date	등급이 파기된 날짜
	comment quotedname	레이블을 보는 사람을 위한 정보
	comment-label quotedURL	생략되지 않은 완전한 레이블을 연결하는 URL
	extention	확장 메커니즘



용하며 HTTP 헤더와 동등한 메카니즘을 따른다. 사용 예는 다음과 같다.

#### (예 2) HTML의 META 태그를 이용한 레이블의 전송

```
(head)
(META http-equiv="PICS-Label" content="
(PICS-1.1 "http://www.gcf.org/v2.5"
labels on "1994.11.05T08:15-0500"
until "1995.12.31T23:59-0000"
for "http://w3.org/PICS/Overview.html"
rating (suds 0.5 density 0 color/hue 1)
)
</head>
```

두번째는 RFC822 헤더를 사용하는 프로토콜을 통해 전송되는 문서에 포함되어 전송하는 방식이다. 이때 HTTP 서버는 클라이언트(웹 브라우저)가 문서를 요청할 때 PICS 레이블 헤더를 포함하도록 한다.

세번째 방식은 문서와 레이블을 분리하여 보내는 것이다. 클라이언트는 HTTP 프로토콜을 사용하는 레이블국(label bureau)에 레이블을 요청할 수 있다. 레이블국은 HTTP 서버로서 특정 질의 구문을 처리하여 HTTP가 아닌 다른 프로토콜을 통해 이용할 수 있는 다른 서버에 있는 문서에 대한 레이블을 제공한다.

## (2) PICS 등급 서비스와 등급 체계

등급 서비스란 인터넷상의 정보에 대한 내용 레이블을 제공하는 개인, 기관 혹은 회사를 지칭한다. 등급 체계는 레이블을 작성할 때 사용되는 속성과 속성에 따른 값의 범위, 그리고 값의 할당시 사용되는 범주를 설명하는 것으로 등급 어휘(vocabulary)라고도 한다. 미국내에서 상영되는 영화는 관람 허용 척도를 여러 가지 등급으로 나누고 있는데 이 때 사용되는 허용 척도인 G, PG, PG-13, R 그리고 NC-17이 이러한 등급 체계의 대표적인

예이다.

등급 서비스가 제공하는 레이블은 등급 체계에 기초한다. 각 등급 체계는 유효한 URL에 의해 식별되며 이는 여러 가지 등급 서비스가 동일한 등급 체계를 사용할 수 있도록 하기 위함이다. 새로운 MIME 유형인 application/pics-service로 정의한 등급 서비스의 예는 다음과 같다.

#### (예 3) 등급 서비스

```
PICS-version 1.1)
(rating-system "http://www.gcf.org/ratings") →
등급 체계의 식별기호(URL)
(rating-service "http://www.gcf.org/v1.0/") →
등급 서비스의 식별기호(URL)
(icon "icons/gcf.gif" → 등급 서비스 관련 아이콘
(name "The Good Clean Fun Rating System")
(description "Everything you ever wanted to
know about soap,
cleaners, and related products. For
demonstration purposes only.")

(category → 범주
(transmit-as "suds") → 레이블에서 사용되는 간략
전송명
(name "Soapsuds Index") → 전송명
(min 0.0) → 척도
(max 1.0))

(category
(transmit-as "density")
(name "suds density")
(label (name "none") (value 0) (icon
"icons/none.gif"))
(label (name "lots") (value 1) (icon
"icons/lots.gif"))
```

## (3) PICSRules

PICSRules은 URL을 기술하는 PICS 레이블에 기반하여 URL에 대한 접근 허용과 차단을 위한 필터링 규칙인 프로파일 작성하기 위한 언어이다. 일반적인 프로파일 작성을 위한 규칙을 사용하는 목적은 프로파일의 공유와 설치, 에이전트나 탐색 엔진, 프록시, 기타 서버와의 통신, 그리고 여러 가지 필터링 제품들을

사용할 수 있게 하기 위함이다.

PICSRules은 PICS 레이블에 기반하여 접근을 금지하고 허용하는 경우가 있는데 각 예를 살펴보면 다음과 같다.

#### (예 4) PICS 레이블에 기반한 접근 금지

```
1 (PicsRule-1.1
2 (
3 serviceinfo (
4 "http://www.coolness.org/ratings/V1.html"
5 shortname "Cool"
6 bureauURL "http://labelbureau.coolness.
org/Ratings"
7 UseEmbedded "N"
8 )
9 Policy (RejectIf "((Cool.Coolness (<= 3) or
(Cool.Graphics >= 3))")
10 Policy (AcceptIf "otherwise")
11 )
12 )
```

이 경우, 문서는 'Cool' 등급 서비스(<http://www.coolness.org/ratings/V1.html>)를 사용하며 레이블은 레이블국(<http://labelbureau.coolness.org/Ratings>)에서 가져온다. 제9행에서 보는 바와 같이 Cool의 정도가 충분하지 않거나(척도 3이하), 그래픽이 너무 많은 경우(척도 3이상)에는 차단된다.

#### (예 5) PICS 레이블에 기반한 접근 허용

```
1 (PicsRule-1.1
2 (
3 ServiceInfo (
4 name "http://www.coolness.org/ratings/
V1.html"
5 shortname "Cool"
6 bureauURL "http://labelbureau.coolness.org/
Ratings"
7 )
8 Policy (RejectUnless "(Cool.Coolness)")
9 Policy (AcceptIf "((Cool.Coolness > 3) and
(Cool.Graphics < 3))")
10 Policy (RejectIf "otherwise")
11 )
12 )
```

이 경우는, 접근 차단의 경우와는 반대로 Cool의 정도가 충분하거나(척도 3이



상) 그래픽이 너무 많지 않으면(척도 3이하) 접근이 허용된다.

#### 4. 관련 소프트웨어

PICS는 레이블 작성이나 등급을 매기기 위한 제품이나 서비스가 아니라 이를 수행할 수 있도록 돕는 기반구조이며 이용자는 PICS 자체가 아닌 PICS와 호환 가능한 등급 체계나 필터링 소프트웨어를 독립적으로 선택하여 이용하게 된다.

따라서 PICS와 호환 가능한 소프트웨어가 다수 개발되어 있다. 이미 1995년 초에는 Microsoft, Netscape, SurfWatch, CyberPatrol사가 PICS 호환 제품을 내놓았으며, 1996년말에 AOL, AT&T WorldNet, CompuServe, Prodigy사 등은 PICS와 호환되는 자유 차단 소프트웨어를 내놓았다.

이러한 소프트웨어에 제공될 PICS와 호환되는 등급 체계로는 RSACi와 SafeSurf가 대표적이다. RSACi(Recreational Software Advisory Council on the Internet)는 오락용(게임) 소프트웨어와 인터넷과 같은 매체의 등급을 매기기 위한 전문가 모임에서 제정한 것으로 폭력, 신체노출, 섹스, 언어의 4개 분야를 각각 5단계의 등급으로 나누고 있다. (<http://www.rsac.org/>)

SafeSurf는 부모가 아이들을 위해 인터넷상의 유해 정보를 차단할 수 있도록 하기 위한 등급 체계로서 나이, 모독행위, 이성애(異性愛), 동성애(同性愛), 신체노출, 폭력, 섹스, 불관용, 약물남용, 기타, 도박 등의 세부 분야에 대해 각각 9가지 척도를 제공한다. (<http://www.safesurf.com/index.html>)

현재 국내에는 PICS 호환 선별 소프트

웨어로서 한국전산원의 NCAPatrol 1.5가 개발되어 있다.

#### 5. PICS의 활용

PICS는 내용 선별을 위한 기술 규격으로 개발되었지만 등급 체계 적용외에도 디지털 저작물에 대한 기술을 통해 메타데이터로서 인터넷상의 수많은 자원의 탐색·검색을 위해서도 활용될 수 있다. 기존의 탐색엔진은 인터넷 문서의 내용을 색인한 것을 기반으로 하는 검색방식을 취하고 있다.


이때 레이블을 사용하는 PICS 체계를 활용하면 레이블 정보만을 이용하여 검색을 할 수 있으며 등급에 따른 정보의 품질 제한도 가능하여 보다 효과적인 정보 검색이 가능하다. 또한 레이블을 이용하여 컴퓨터 바이러스 보호나 디지털 서명과 저작물의 소유권 및 사용권 표시 등의 분야에도 활용할 수 있다.

#### 6. 향후 방향

다양한 문화가 공존하는 인터넷 세계에서 개별 이용자가 자신에게 적절한 정보

만을 취하고 위험하거나 부적절한 정보는 차단하기란 쉬운 일이 아니다. PICS는 이를 기술적으로 가능토록 하는 규격으로서 매우 유용한 도구이다. 그러나 아직까지 전세계 인터넷 사이트의 1%미만에만 레이블이 부여되어 있는 것으로 파악되고 있어 그 실효성에 의문을 제기하지 않을 수 없다.

따라서 인터넷 자원의 제작자나 이용자의 적극적인 참여와 PICS와 호환될 수 있는 관련 소프트웨어의 개발이 활발하게 이루어져야 한다. 또한 국내에서 개발된 PICS 호환 선별 소프트웨어가 많지 않은 상황에서 우선적으로 국내 환경에 맞는 적절한 등급 체계의 개발을 통해 다양한 소프트웨어의 개발을 촉진해야 한다.

PICS는 앞으로 내용 선별 분야에서 뿐만 아니라 메타데이터 시스템으로서 웹의 중요한 부분을 차지하게 될 것이며, 더욱 복잡한 상거래, 커뮤니케이션, 색인, 및 탐색 서비스를 가능하도록 하기 때문에 차세대 레이블 포맷을 개발하는 기초 작업이 될 것으로 전망된다. 

#### 참고문헌

이병만. 1998. "인터넷 내용선별 체계(PICS)". 제1회 정보·통신 표준화 심포지움. 1998.6.2-3. 490-499. (서울: 한국과학기술회관).

최문선. 1999. (인터넷 메타데이터 기반구조 PICS를 이용한 협력적 필터링에 관한 연구). 석사학위논문, 연세대학교 대학원, 문헌정보학과.

W3C. "Platform for Internet Content Selection (PICS)". [online]. [cited 1999.1.8]. (<http://www.w3.org/PICS/>).

Resnick, P. 1997. "Filtering Information on the Internet". Scientific American, March: 106-108. [online]. [cited 1999.1.8]. (<http://www.sciam.com/0397issue/0397resnick.html>).

Resnick, P. and Miller, J. 1996. "PICS: Internet Access Controls Without Censorship". Communications of the ACM. 39(10): 87-93. [online]. [cited 1999.1.9]. (<http://www.w3.org/PICS/iacwcv2.htm>).

Miller, J. Resnick, P. and Singer, D. 1996. "Rating Services and Rating Systems(and Their Machine Readable Descriptions)". Ver.1.1. [online]. [cited 1999.1.8]. (<http://www.w3.org/TR/REC-PICS-services>).

Krauskopf, T. et al. 1996. "PICS Label Distribution Label Syntax and Communication Protocols". Ver.1.1. [online]. [cited 1999.1.8]. (<http://www.w3.org/TR/REC-PICS-labels>).

Evans, C. et al. 1997. "PICS Rules 1.1". [online]. [cited 1999.1.8]. (<http://www.w3.org/TR/REC-PICSRules>).

Chu, Y. et al. 1998. "PICS Signed Labels(DSig) 1.0 Specification". [online]. [cited 1999.1.8]. (<http://www.w3.org/TR/REC-DSig-label/>)