

웹 콘텐츠 변환을 위한 다중 사용자 주석시스템

김재홍⁰ 장민수 손주찬

한국전자통신연구원 인터넷컴퓨팅연구부 비즈니스지식처리연구팀

(jhkim504⁰, minsu, jcsohn)@etri.re.kr

Multiparty Annotation System for Web Content Transcoding

Jaehong Kim⁰ Minsu Jang Joo-Chan Sohn

Business Knowledge Research Team, Internet Computing Department, ETRI

요 약

인터넷 접속이 가능한 장치의 종류가 계속 증가하고 있으나, 인터넷에 있는 많은 콘텐츠들은 제한된 성능을 가지는 장치들에서는 제대로 표현되지 않는다. 콘텐츠 제작비용 절감을 위해서는 기존의 콘텐츠를 재활용하는 것이 바람직하며, 본 논문에서는 이러한 목적으로 주석 편집기, 저작시점의 주석서버, 수행시점의 주석서버로 구성되는 다중 사용자 주석시스템을 설계하고 구현하였다. 구현된 시스템을 활용하면 콘텐츠 저작자는 콘텐츠에 의미 정보와 중요성을 부여하고, 콘텐츠 제공자는 웹 콘텐츠 변환을 위한 기본 주석을 제공하고 일반 사용자는 자신만의 개인화된 콘텐츠를 위한 주석을 별도로 작성할 수 있어 웹 콘텐츠 저작과 구축 그리고 사용에 관계된 모든 사용자들의 요구 사항을 충족시킬 수 있는 웹 문서 변환이 가능하다.

1. 서론

인터넷 접속이 가능한 다양한 클라이언트 장치 사용가능해짐에 따라, PDA나 CDMA 전화와 같은 장치 상에서 인터넷 접속을 하는 사용자들이 꾸준히 증가하고 있다. 그러나, 이러한 장치들은 데스크탑 컴퓨터와 동일한 처리 능력을 가지고 있지 않기 때문에 기존 웹 콘텐츠가 이런 장치상의 브라우저에서 표현되기 위해서는 변환의 과정이 필요하다. 예를 들어, 데스크탑용 큰 사이즈의 트루 컬러 이미지는 PDA와 같은 장치에서 작은 그레이스케일 이미지로 대체되어 표시될 수 있고, CDMA 전화상에서는 텍스트로 대체되거나 삭제되어 표시될 수도 있다. 이러한 웹 콘텐츠 변환은 다양한 클라이언트 장치 사용자나 서비스 제공자 모두에게 빠른 서비스 시간과 적은 비용으로 보다 적절한 콘텐츠 표시 형태를 제공하기 위해 꼭 필요한 것이다.

초기의 웹 콘텐츠 변환시스템은 대부분 구문정보에 기반한 것이었다. 그러나, 구문 정보에만 기반한 변환에서는 콘텐츠의 특정부분중 어떤 부분을 소형 화면을 가진 단말장치에서 보여주어야 할지를 결정하기가 어려워 의미적인 측면에서의 품질이 좋지 않다. 이러한 문제에 대한 대안으로 Hori 등은 주석기반(Annotation-based) 변환시스템을 제안하였다[1]. 주석기반 변환시스템을 사용하면 사용자는 WYSIWYG 주석편집기를 사용하여 작은 화면을 가지는 장치에서 보여져야 할 부분을 직접 선택할 수 있으며, 이 정보는 주석형태로 접근가능한 저장소에 저장되어 있다가 변환시스템에 의해 활용된다.

기존 연구에서는 주석을 다는 주체의 다양성에 대해서는 고려하지 않았으나, 주석을 달기 원하는 사용자 층은 다양하게 존재할 수 있다. 예를 들어, 콘텐츠 저작자, 제공자(사이트 구축자) 및 사용자는 각각 자신들의 역할에 따라 조금씩 다른 형태로 주석을 만들기를 원할 것이다. 본

논문에서는 콘텐츠 저작자, 제공자, 사용자의 변환에 대한 요구사항을 모두 반영하는 다중 사용자 주석시스템을 설계하고 구현하였다.

2. 다중 사용자 주석시스템

2.1 주석기반 변환의 기본 개념

주석기반 웹 콘텐츠 변환의 기본 개념은 [1]에서 이미 소개된 바 있다. 본 논문에서는 이 기본 개념을 공유하기 때문에 이에 대해 간단히 설명한다. 대부분의 존재하는 웹 콘텐츠는 데스크탑용으로 작성되었으며, 이러한 콘텐츠는 주석이라고 부르는 메타정보를 활용하여 다른 장치에서도 효과적으로 표현될 수 있다. 예를 들어, 그림 1과 같은 샘플 콘텐츠가 있다고 하면, 주석이 작성되지 않은 일반적인 브라우저에서의 표현은 그림 2와 같다.

```
<html>
<head>
<title>Annotation-based Transcoding System</title>
</head>
<body>
<p>Annotation-based Transcoding System</p>
<p>To be kept</p>
<p>To be replaced</p>
<p>To be removed</p>
</body>
</html>
```

그림 1 샘플 HTML 파일 소스

Annotation-based Transcoding System
To be kept
To be replaced
To be removed

그림 2 주석이 적용되지 않은 브라우저 표시

그러나, 그림 3과 같이 HTML 파일의 경로상에 주석이 작성되어 주석에 기술된 바와 같이 keep, replace 또는 remove 가 적용된 후의 브라우저 표시는 그림 4와 같이 된다.

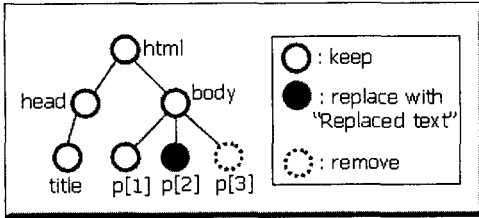


그림 3 주석의 예

**Annotation-based Transcoding System
To be kept
Replaced text**

그림 4 주석이 적용된 후의 브라우저 표시

주석기반 변환시스템은 이와 같은 개념에 기반한다. 주석의 어휘로 사용될 수 있는 것은 위의 예에서와 같이 keep, replace 그리고 remove가 기본이 될 수 있으나 제한은 없으며 자신의 용도에 맞게 다양하게 사용될 수 있다.

2.2 다중 사용자 주석의 필요성

상기 설명한 기존 연구에서는 주석을 이용한 웹 문서변환의 기본 개념을 소개하였으나, 주석을 작성하는 주체의 층을 다양하게 세분화하지는 않고, 하나의 주석작성 주체만 존재하는 모델에 기반한 연구를 하였다. 그러나, 다음과 같은 이유에서 주석을 작성하는 주체가 보다 세분화될 필요가 있다.

- 콘텐츠를 작성하는 저작자의 입장에서 자신이 중요하다고 생각하는 콘텐츠 부분이 변환시스템에 의해 임의로 삭제될 수 있다. 따라서, 이러한 경우 콘텐츠가 어떠한 형태로든 유지되어야 할 필요가 있다.
- 콘텐츠 제공자가 서비스 제공자 측면에서 작성한 주석의 결과 페이지가 사용자들이 원하는 것과 다를 수 있다. 이러한 경우, 콘텐츠 사용자가 직접 자신들의 기호에 맞는 콘텐츠를 구성할 수 있도록 주석 작성이 가능하도록 해야 한다.

이러한 이유로 본 논문에서는 주석 작성의 주체를 표1에서와 같이 3부류로 구분하고 각각의 역할을 규정하였다.

표 1 주석 작성 주체별 역할

작성주체	역할
콘텐츠 저작자	콘텐츠에 의미 및 중요도 부여
콘텐츠 제공자	변환을 위한 기본 주석 제공
콘텐츠 사용자	사용자 선호에 맞는 주석 작성

2.3 주석언어

그림 5는 위와 같은 목적으로 사용하기 위해 개발된 주석언어(Annotation Language)의 DTD이다.

주석언어는 최상위 어휘 salt(Simple Annotation Language for Transcoding)로 시작하며 이것은 주석을 달 웹 콘텐츠의

위치(targetURL)와 관련 정보(version, title, author) 및 다수의 주석(annotation+)으로 구성된다. 주석(annotation)의 속성 role과 priority를 사용하여 저작자는 콘텐츠의 특정 부분이 의미하는 역할(role: header, footer, advertisement 등)과 그 부분의 중요도(priority: high, medium, low)를 지정할 수 있다. 특별히, 중요도가 medium일 때 대체가능한 자원들(altResource+)을 지정할 수 있다.

각각의 주석(annotation)은 다수의 명령(command+)으로 구성되며, 각 명령(command)은 주석을 달기 위한 장치의 조건(condition)과 조건이 만족되었을 때 수행하는 행위(action: keep, replace, remove), 그리고 특별히 action이 replace일 경우 대체가능한 자원들 자체 또는 자원들의 주소(altResource+)로 구성된다.

```
<?xml version="1.0" encoding="UTF-8"?>
<ELEMENT salt (information, annotation+)>
<ELEMENT information (version, title, author)>
<ATTLIST information
  targetURL CDATA #REQUIRED
  >
<ELEMENT annotation (command+)>
<ATTLIST annotation
  target CDATA #REQUIRED
  type (author | cp | user) #REQUIRED
  role (header | footer | sidemenu | decoration |
  advertisement | maincontent | subcontent) #IMPLIED
  priority (high | medium | low) #IMPLIED
  >
<ELEMENT command (condition, action, altResource+)>
<ATTLIST command
  type (author | cp | user) #REQUIRED
  >
<ELEMENT version (#PCDATA)>
<ELEMENT title (#PCDATA)>
<ELEMENT author (#PCDATA)>
<ELEMENT condition (#PCDATA)>
<ELEMENT action (#PCDATA)>
<ELEMENT altResource (#PCDATA)>
```

그림 5 주석언어의 DTD

본 논문에서 사용하는 주석언어는 저작자에게는 콘텐츠의 역할(role), 중요도(priority) 및 대체 가능한 자원(altResource)을 지정할 수 있게 하고, 콘텐츠 제공자 또는 사용자에게는 장치가 어떠한 조건(condition)을 만족할 경우 콘텐츠의 특정 노드(target)를 keep 또는 다른 대체자원(altResource)으로 replace하라는 힌트를 주는 목적으로 개발되었으며, 최소한의 기능만을 가진다.

2.4 다중 사용자 주석의 경험

본 논문에서 개발한 다중 사용자 주석시스템은 콘텐츠 저작자, 제공자 그리고 사용자 각각이 웹 콘텐츠에 대해 주석 작성이 가능하다. 따라서, 다음과 같이 주석의 사용이 상충되는 경우가 발생할 수 있다. 이와 같은 경우 상충된 주석을 정합하여 변환을 위한 최종 결과를 생성해주는 메커니즘이 필요하며, 본 논문에서는 각 경우에 따라 표 3과 같은 메커니즘을 사용한다. (표 3에서 AA는 저작자 주석, CPA는 제공자 주석, UA는 사용자 주석을 나타낸다.)

표 2 주석이 상충되는 경우의 예

주석의 대상	그림 3의 html/body/p[2]
컨텐츠 저작자	priority = "medium"
컨텐츠 제공자	condition = "deviceType = 'pda'" action = "remove"
컨텐츠 사용자	condition = "deviceType = 'pda'" action = "replace" altResource = "Replaced text"

표 3 주석정합 매커니즘

AA만 존재	priority가 high/medium/low일 때 각각 keep, replace, remove로 처리
CPA만 존재	CPA의 action에 기술된 처리수행
UA만 존재	UA의 action에 기술된 처리수행
AA와 CPA가 존재	AA의 priority가 high일 때 keep으로 처리, 그 외의 경우 CPA의 action에 기술된 처리수행
AA와 UA가 존재	AA의 priority가 high일 때 keep으로 처리, 그 외의 경우 UA의 action에 기술된 처리수행
AA, CPA 와 UA가 존재	AA와 UA가 존재하는 경우와 동일하게 처리
CPA와 UA가 존재	UA만 존재하는 경우와 동일하게 처리

표 3은 본 논문의 주석정합 매커니즘에서 각 주석 주체별 주석의 우선순위가 AA, UA, CPA 순임을 보여준다.

3. 시스템 구현 및 활용 시나리오

다중 사용자 주석시스템은 그림 6과 같이 주석편집기(Annotation Editor), 저작시점(Authoring-time) 주석서버(Annotation Server), 수행시점(Runtime) 주석서버로 구성되며, 활용시나리오는 다음과 같다.

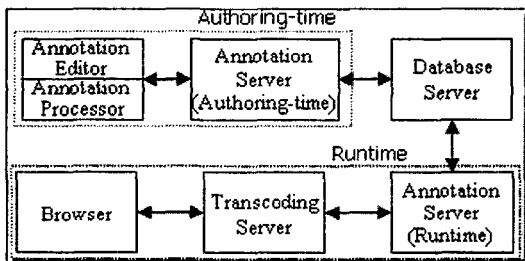


그림 6 시스템 구성 및 응용 형태

(저작시점) 컨텐츠 저작자, 제공자 또는 사용자가 주석편집기를 이용하여 기존 웹 컨텐츠에 주석을 달면 주석처리기(Annotation Processor)가 2장에서 설명과 같이 주석정합을 하고 저작시점 주석서버가 정합된 결과를 데이터베이스에 저장한다. 본 구현에서 주석문서는 XML 형태이고, 다수의 주석이 정합된 결과는 XSL 형태이다.

- (수행시점 단계 1) 브라우저에서 웹 컨텐츠에 대한 요청이 있으면 변환서버(Transcoding Server)는 수행시점 주석서버에 요청된 URL에 해당하는 주석(XSL 파일)을 요청하고, 주석서버는 결과를 변환서버에 반환한다.
- (수행시점 단계 2) 변환서버는 반환된 주석을 이용하여 keep, replace, remove 등의 처리를 한다(원본 웹 컨텐츠와 반환된 XSL을 입력으로 하여 XSLT 수행). 처리가 완료되면 기존 컨텐츠의 양이 요청한 장치의 능력에 적합한 양으로 감소되어 있다.
- (수행시점 단계 3) 변환서버는 태그변환(예, HTML→WML) 등 다른 부가적인 처리가 필요할 경우 이를 수행하고, 최종 결과를 브라우저에 반환한다.

본 논문의 다중 사용자 주석시스템은 표 4와 같이 구현되었다.

표 4 시스템 구현도구

주석편집기	VC++, MS IE 컴포넌트
주석처리기	JAVA
저작시점 주석서버	COM+ 컴포넌트, 주석편집기와는 SOAP 프로토콜로 통신
데이터베이스	MS SQL Server
수행시점 저작서버	IIS, ASP

구현된 다중 사용자 주석시스템을 사용하여 [2,3]의 변환시스템과 연동하여 PDA, CDMA 전화 등에 대해 테스트해 본 결과 간단한 작업만으로 기존 컨텐츠를 재활용하여 작은 화면을 가지는 장치들에 대한 컨텐츠의 작성이 가능하다는 것을 확인하였다.

4. 결론

기존 웹 컨텐츠를 제한된 성능을 가지는 다양한 클라이언트 장치에서 재활용하여 서비스 가능하도록 하기 위해 본 논문에서는 컨텐츠 저작자, 제공자 및 사용자의 변환에 대한 요구사항을 모두 반영하는 다중 사용자 주석시스템을 설계하고 구현하였다. 구현된 다중 사용자 주석시스템은 컨텐츠 변환시스템과 연계하여 변환해야 할 컨텐츠의 양을 줄이는 목적 및 무선 컨텐츠 개인화 목적 등에 맞게 효과적으로 사용될 수 있다.

참고문헌

[1] Masahiro Hori, Goh Kondoh, Kohichi Ono, Shin-ichi Hirose, Sandeep Singhal, "Annotation-Based Web Content Transcoding", Proc. of the 9th International World Wide Web Conference, May. 2000.
 [2] Jaehong Kim et al., "Internet Content transcoding framework for heterogeneous client devices", CALS/EC KOREA 2001 International Conference, pp.379-391, 2001.
 [3] 이미경 외, "유무선 문서 변환 시스템의 태그변환기 설계 및 구현", 한국정보과학회 가을 학술발표논문집 제 29권, 제 2호(III), pp 43-45, 2002.