

차세대 웹 기반의 콘텐츠 관리 기법

김동수*, 김현구**, 양진석**, 정태명**,
*성균관대학교 컴퓨터공학과
**성균관대학교 컴퓨터공학과
e-mail : dskim96@imtl.skku.ac.kr

The Next Generation Web based Content Management technique

Dong-Soo Kim*, Hyun-Koo Kim**, Jin-Seok Yang**, Tai-Myung Chung**
*School of Information and Communication Engineering, Sungkyunkwan University
** School of Information and Communication Engineering, Sungkyunkwan University

요 약

한 조직에서 생성되어 조직 내부에서 사용되거나 제 3 자가 사용 가능하도록 전산 기반으로 작성되어 저장된 정보를 콘텐츠라 한다. 이러한 정보들은 항상 신속하고 정확하게 재사용됨으로써 조직의 생산성을 높일 수 있다. 그러한 이유로 콘텐츠를 직렬하게 관리하는 것이 중요한 이슈로 부각되고 있으며, 이미 특정 분야의 콘텐츠를 관리하는 기술은 조직에서 보편적으로 사용될 만큼 발전해 왔다. 본 논문에서는 특정 분야의 콘텐츠 관리에 치우치지 않고, 다양한 종류의 콘텐츠를 차세대 웹 기반 환경에서 통합 관리 할수 있는 콘텐츠 관리 기술에 대해 제안한다.

1. 서론

한 조직에서 생성되어 조직의 내부에서 사용되거나 제 3 자가 사용 가능하도록 전산 기반으로 작성되어 저장된 정보를 콘텐츠라 하며, 조직의 운영에 있어서 매우 중요한 기반 정보가 된다.

전산 기반으로 저장된 정보들은 지속적으로 사용되게 되며, 항상 필요한 정보를 정확하고 신속하게 제공할 수 있어야 한다. 찾아진 정보를 활용하는 경우, 해당 단일 정보만을 사용하는 것이 아니라 정보를 찾는 사람의 목적에 따라 여러 가지 정보와 함께 사용된다.

그러기 위해 콘텐츠는 크게 생성, 변경, 승인, 출판, 보관, 분배, 재사용, 공유, 소멸 등의 과정을 겪게 되며, 이러한 과정들의 관리를 일반적으로 “콘텐츠 관리”라 한다. 적절한 콘텐츠 관리를 통하여 기반 콘텐츠를 활용한 새로운 콘텐츠의 생성등이 용이하게 되며, 이는 곧 조직의 생산성과 직결된다.

본 논문에서는 현재 보편적으로 사용되는 콘텐츠 관리 기법의 문제점을 지적하고, 차세대 네트워크 인프라에 적용 가능한 새로운 콘텐츠 관리 기법을 제시한다.

2. 관련연구

콘텐츠 관리에 있어서 관리대상의 단위인 “콘텐츠”란, 콘텐츠 사용 대상자에게 서비스되는 모든 것들을 의미한다. 이러한 콘텐츠를 관리하는 “콘텐츠 관리 시스템”이란, 실제로 서비스하고 관리하기 위한 디자인 및 프로그램 등을 관리 단위로 설정하지 않은 기존의 정보관리 시스템과 달리 효율적인 정보의 재가공을 가능하게 할 수 있도록, 실제 내용인 콘텐츠와 그 내용을 가공하여 보여주는 프리젠테이션을 분리하는 방법을 기본적으로 사용한다. 실제로 가공하지 않은 데이터를 사용자 뷰와 별도로 분리 시킴으로써 가공하지 않은 데이터에 여러 가지 뷰를 더하여 다양한 형태의 콘텐츠를 가공할 수 있으며, 여러 데이터 소스의 일부만을 추출하여 새로운 콘텐츠를 만들어 낼 수 있다.

위와 같은 생성,출판 등의 과정을 거친 콘텐츠는 여러 콘텐츠 사용 분야로 분배 되기도 하며, 때에 따라 재사용 될 필요성이 있으며, 그 용도를 다하게 되면 소멸하게 된다. 이와 같이 콘텐츠의 생성부터 소멸

까지의 과정을 관리하는 것을 “컨텐츠 관리”라 한다.

현존하는 컨텐츠 관리 시스템들은 조직의 문서관리, 웹 사이트의 컨텐츠 관리 등 각 분야별로 특화된 관리에 초점이 맞춰진 관리 시스템이 존재할 뿐 범용적인 컨텐츠 관리를 위한 관리 기법을 적용한 시스템은 아직 널리 사용되고 있지 않다. 이러한 경우 다른 조직의 컨텐츠를 사용하거나 혹은 다른 시스템을 사용하는 경우 컨텐츠의 재사용을 하는데 있어서 호환성 및 문서의 구조 표준의 부재로 효과적이고 정확한 컨텐츠 사용 및 관리가 이루어지기 힘들다.

시스템 및 어플리케이션의 통합을 위하여 가장 널리 사용되는 컨텐츠 관리 시스템 개발 플랫폼은 웹이며, 가장 보편화된 인터넷 플랫폼으로 자리 잡은 웹은 앞으로는 Semantic Web 기반의 개발 및 표준화로 보다 많은 분야의 프레임웍으로 사용 될 것이다. 이미 많은 엔터프라이즈 시스템들이 웹 기반으로 개발되고 있으나, 각 시스템들은 웹 혹은 엔터프라이즈 문서관리 중 한쪽에 치우친 관리만을 할 수 있는 한계를 지니고 있다.

시스템 및 어플리케이션 통합등의 다양한 분야에 활용되는 웹 기반의 장점을 충분히 활용하여 기존 문서관리 컨텐츠와 웹 컨텐츠를 통합 관리 할수 있는 새로운 관리 기법이 필요한 실정이다.

3. 차세대 웹 기반의 컨텐츠의 관리 구조

현대는 웹이 점차적으로 가장 큰 영향력을 끼칠 수 있는 미디어가 되어가고 있는 추세이다. 이와 더불어 한 조직에서 사용되어지는 어플리케이션 역시 웹을 기반으로 하는 시스템이 점차 보편화 되어 가고 있다. 이는 어떠한 전산 기반 정보이든지 Web 을 통해 접근할 수도 있다는 것을 말한다. 즉, 단순히 웹 컨텐츠 관리를 위해서 뿐만이 아니라, 일반적인 컨텐츠를 관리하기 위한 수단으로 웹을 사용할 수 있다.

3.1 Semantic Web 기술의 수용

차세대 웹, 즉 Semantic Web 은 기존 World Wide Web 의 한계를 극복하기 위한 새로운 개념의 웹으로, 웹의 창시자인 Berners-Lee 와 W3C 가 표준화를 위해 노력하고 있다.

Semantic Web 의 등장 배경은 인터넷 상의 정보가 기하급수적으로 증가하고, 다량의 정보 분석을 위한 하부 구조가 존재 하지 않아 시스템 간의 커뮤니케이션에 어려움이 많아졌기 때문이다. 이는 곧 Web 표현의 핵심인 HTML 로 작성된 웹 컨텐츠의 구조 및 내용을 정확하게 파악할 수 없기 때문이며, 이는 구조적이지 못한 HTML 의 한계점이기도 하다.

Semantic Web 은 현재 Web 의 확장이며, 자원의 의미를 체계적으로 정의하여 컴퓨터와 사람의 협력적인 운용을 유도하는 방법이다. 이는 Web 정보의 명확한 정의와 연결에 기반하여 효율적인 검색, 자동화, 통합, 재사용을 시도하는 새로운 접근 방법이다. 여기서 주목할 점은 Semantic Web 이 Intelligent Agents 라는 개념을 제안하고 있다는데 있다. 이 Intelligent Agent 는

사용자의 언어를 정확히 이해하고, 사용자를 대신하여 정확한 정보를 실시간으로 제공하는 것을 지향한다.

차세대 Web 이 기반 네트워크로 자리 잡게 될 경우 이를 기반으로 설계된 컨텐츠 관리 시스템은 보다 더 넓은 범위의, 능동적이며 지능적인 정확한 컨텐츠 관리가 가능할 것이다. 이러한 능동적이고 지능적이며 정확한 컨텐츠 관리를 가능하게 해줄 수 있는 Semantic Web 의 기반 기술은 Uniform Resource Identifier(URI), XML(eXtensible Markup Language) Namespace, XML(eXtensible Markup Language) schema, Resource Description Framework(RDF), RDF schema, DARPA Agent Markup Language(DAML), Ontology Inference Layer(OIL), Web Ontology Language(OWL) 등과 Intelligent Agents 기술이다.

기술	설명
URI	자원들의 고유 식별자 역할
XML namespace, schema	URI 에 의해 식별되는 이름들의 집합체이며 다양한 data 유형을 지원
RDF, RDF schema	자원의 기술을 위해 사용하는 언어이며, metadata 의 인코딩, 교환, 재사용 등을 위해 사용하고 자원간의 관계 설정
DAML, OIL	RDF 와 RDF schema 에는 부족한 모델링 요소들을 확장 하는데 사용
OWL	논리적 추론이 필요한 질의를 위해 사용 하는데, 보다 더 인간의 일상 언어에 가까운 질의 가능

[표 1] Semantic Web 기반 기술요소

OWL 의 Ontology 는 사전적 의미로 “존재론, 본체론” 등을 뜻하는 것으로, Semantic Web 에서는 논리적 추론이 필요한 질의를 위한 개념이다. Semantic Web 에서 제시하는 Intelligent Agents 의 개념은, 인터넷을 통해 분산된 근본적으로 이질의 정보 자원을 능동적으로 수집하여 이를 이해하고 통합하는 에이전트를 말한다.

이와 같은 Semantic Web 의 특성을 컨텐츠 관리 기법에 도입하게 될 경우 현재 우리가 사용하는 많은 이질의 컨텐츠에 대하여 통일된 관리 기법을 적용할 수 있다. 통일된 관리 기법을 통하여 기존의 제한된 범위의 컨텐츠를 위주로 관리하는 방식보다 효율적이고 통합된 관리 시스템의 개발이 가능하며, 이를 통해 조직 또는 개인의 컨텐츠 생산성이 향상될 것이다.

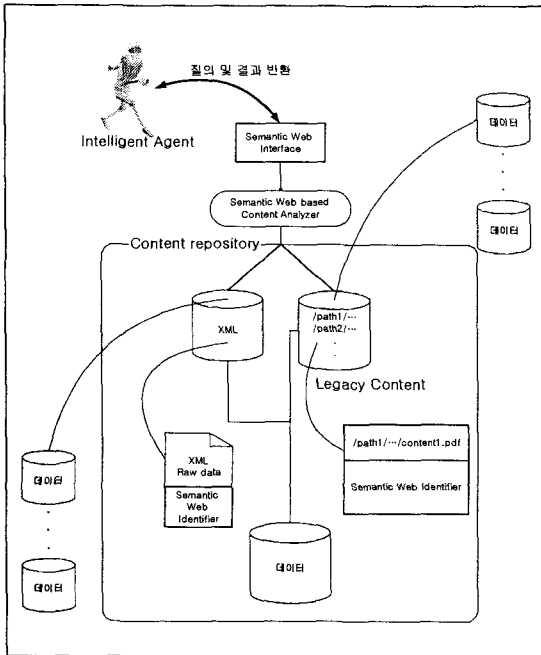
3.2 Semantic Web 기반의 content management technique

[그림 1] 에서는 Semantic Web 의 개념을 도입한 컨

텐츠 관리 시스템의 콘텐츠 라벨링 기법을 제시한다.

아래에 제시한 구조에서는 크게 두가지 저장 형식으로 구분된다. Semantic Web 에서 요구하는 자원의 서술 내용을 추가한 XML 형식의 가공되지 않은 데이터를 저장한 저장소와, 기존의 콘텐츠(legacy content), 즉 기존에 생성되어 사용되어지는 범용 파일 형식의 데이터가 저장되는 기존의 콘텐츠 저장소(legacy content repository)로 나뉜다.

XML raw data (가공되지 않은 XML 데이터)를 저장하고 있는 저장소는 객체 관계 데이터베이스(ORDBMS)로 구성한다. ORDBMS 구성을 통해 XML raw data 의 작은 조각(data fragment) 의 재사용을 최대한 효율적으로 이루어 낼 수 있다. 이와 더불어 아래의 단일 콘텐츠 저장소 외의 다른 물리적 공간의 XML raw data 를 활용할 수 있도록 해당 XML raw data 자원의 위치를 지정해 사용 할 수 있다.



[그림 1] Semantic Web 기반의 Content Management System architecture

기존의 콘텐츠(legacy content)의 관리를 위한 기법들은 계속 이슈화 되어 왔다. 대표적인 관리 기법 중 하나는 콘텐츠 파일에 파일이 손상되지 않는 범위 내의 한 부분에 콘텐츠에 대한 식별 표시(identification labeling)를 하는 것이고, 다른 하나는 파일을 특정 저장소에 저장하고, 특정 MIB(Management Information Base)에 해당 파일의 메타 정보를 저장하는 방식이다.

본 논문에서 제시하는 기존의 콘텐츠에 대한 관리 기법으로는 위의 대표적인 관리 기법들 중 파일이 손상되지 않는 범위 내에서 파일의 특정 부분에 Semantic Web 에서 정의하는 자원 설명 내용을 라벨링 하는 방법이다. 이는 향후 Semantic Web 이 기반 인프

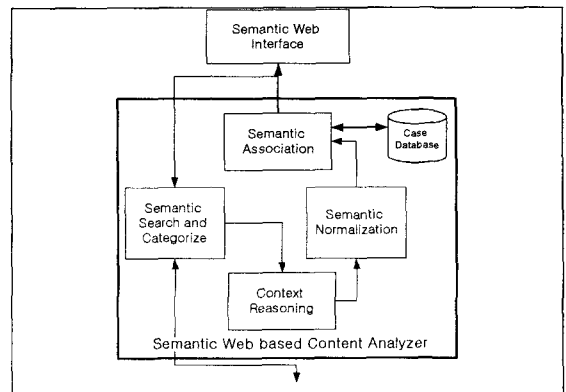
라로 널리 사용 될 때를 위해 앞으로 생성되는 모든 범용 파일 형식에 본 논문에서 제시하는 labeling 기법을 권고하기 위함이다. 이는 기존 콘텐츠 관리 시스템의 비효율성을 한편으로 답습한 모습을 띄긴 하지만, 라벨링 하는 내용 및 구조는 차세대 웹 기반에서 더 큰 발전성을 내포하고 있는 방법이다. 이러한 방식을 지원하기 위하여 본 논문이 제시하는 시스템에서는 기존의 콘텐츠에 대한 라벨링 툴을 제공해야한다.

콘텐츠 저장소를 크게 두 가지로 나눈 본 시스템 구조에서는 서로 상이한 두 콘텐츠 저장소 (XML raw data repository, legacy content repository) 를 Intelligent Agent 가 검색 할 때 서로 상이한 기준을 갖고 접근하는 것을 방지하기 위해 Semantic Web 기반 콘텐츠 분석기 (Semantic Web based Content Analyzer) 와 Semantic Web 인터페이스 (Semantic Web Interface)를 통해 검색 질의 및 콘텐츠에 대한 관리를 하게 된다.

3.3 Semantic Web based content analyzer 와 interface

본 논문에서 제시한 시스템에서는 기존의 콘텐츠를 콘텐츠 저장소에 저장하되, Semantic Web 기반 식별자를 라벨링 하는 기법을 도입함으로써 콘텐츠를 semantic web 기반으로 관리할 수 있게 한다. 이러한 경우 Semantic Web 기반에서 요구하는 형식과 완전하게 호환 되지 않을 수 있다.

이러한 기존의 콘텐츠의 Semantic Web 기반의 인프라에서의 이질성을 해결하기 위하여 콘텐츠 저장소와 Semantic Web 인프라와의 사이에 콘텐츠 분석기 (Content Analyzer) 를 두고, 콘텐츠 저장소에서 추출되어져 나오는 모든 콘텐츠에 대한 구조 및 내용 분석을 수행하여 Semantic Web 에서 요구하는 형식으로 정형화 해주는 역할을 하는 부분이 Semantic Web 기반 콘텐츠 분석기(Semantic Web based content analyzer)이며, 콘텐츠 저장소와 Intelligent Agent 혹은 사용자 연결시켜주는 것이 인터페이스(interface)이다. 본 논문에서는 인터페이스 부분에 대한 구현에 대해서는 아직 언급 하지 않는다.



[그림 2] Semantic Web based Content Analyzer

[그림 1] 은 Semantic Web 기반 콘텐츠 분석기 (Semantic Web based Content Analyzer)의 구성을 나타내고 있다.

컨텐츠 분석기의 각 모듈의 역할은 다음과 같다.

- **Semantic 검색, 분류 모듈 (Semantic Search and Categorize module)**

: Intelligent Agent 를 통해 요청 받은 검색을 정확히 수행하기 위하여 콘텐츠에 대한 분류를 수행한다.

- **문맥 추론 모듈 (Context Reasoning module)**

: Ontology 외의 콘텐츠에 대한 분류를 제공하는 기술이다. 문맥 추론 모듈은 구성 메타 데이터 (syntactic metadata)와 의미 메타 데이터(semantic metadata)로 콘텐츠의 구성적, 의미적 전후관계를 분석한다.

Syntactic metadata 는 콘텐츠의 구성에 대한 정보를 말하며, semantic metadata 는 그 자체로 의미를 포함하고 있는 정보들을 말한다. Ontology 가 semantic metadata 에 문맥(context)를 부여한다.

- **Semantic 규격화 모듈 (Semantic Normalization)**

: 각기 다른 semantic 정보를 규격화 해주는 모듈이다. 규격화의 과정은 두 가지 단계로 나누어진다.

첫번째로 정보와 형식에 관계 없이 같은 영역 (domain) 혹은 분류(category) 내의 가능한 모든 메타 데이터를 연관 짓는다.

두번째로 메타 데이터가 동일한 객체를 지칭하는 다른 이름들을 균일화 한다.

- **Semantic 조합 모듈 (Semantic Association module)**

: Semantic 조합 모듈은 검색된 결과와 키워드 간의 적합성 여부를 결정하여 최종적으로 검색 결과를 조합한다. 키워드와 검색 결과 사이의 적합성 여부는

이와 같이 콘텐츠 저장소(Content repository)에 저장 되어 있던 조직 내의 일반적인 도큐먼트 자료와, 콘텐츠 관리 시스템에 의해 관리되는 XML 기반의 콘텐츠 모두 Semantic Web 기반 콘텐츠 분석기(Semantic Web based content analyzer)를 통하여 구조 및 내용이 분석되며, 이를 Semantic Web 기반의 포맷으로 정형화 되어 사용자의 질의에 대한 결과 값으로 보내지게 된다. 이러한 과정을 통해 사용자는 조직 내부의 일반 도큐먼트 콘텐츠 등과 Semantic Web 기반의 일반 Web 콘텐츠를 호환성 및 정확성의 문제 없이 재가공 등을 할 수 있게 되며, 결과적으로 각기 다른 기반의 콘텐츠를 호환성 문제 없이 재가공 할 수 있게 되는 효과를 기대할 수 있다.

4. 향후 과제 및 결론

본 논문에 제시한 관리 기법을 지원하기 위하여 우선적으로 Semantic Web 기반 콘텐츠 분석기(Semantic Web based content analyzer) 와 Semantic Web 인터페이스 (Semantic Web Interface) 의 구현이 가장 우선적으로 이

루어질 예정이며, Semantic Web 인터페이스에 대한 자세한 설계 및 구현이 이루어질 것이다. 그와 더불어 Intelligent Agent 의 프로토타입 모델을 구현하여 논문에 제시된 관리 기법의 시플레이션이 이루어질 것이다.

하지만 무엇보다 W3C 의 Semantic Web 표준 제정이 가장 중요한 사항이다. 더불어 Semantic Web specification 에 명시되어 있는 Intelligent Agent 의 구현 정도 및 성능에 따라 콘텐츠의 검색 및 재사용 효율이 결정될 것이다.

통일 되지 않은 콘텐츠 관리 기법을 차세대 웹 (Semantic Web) 기반의 관리 기법으로 통일 함으로써 보다 다양한 콘텐츠 형식에 대하여 정확하고 통일된 관리가 이루어질 수 있을 것이다.

참고문헌

[1] Feature Synopsis for OWL Lite and OWL, "http://www.w3.org/TR/owl-features/", W3C Working Draft.
 [2] JoAnn T.Hackos, "Content Management for Dynamic Web Delivery", WILEY Computer Publishing,2002.
 [3] RDF Model Theory, "http://www.w3.org/TR/rdf-mt/", W3C Working Draft.
 [4] Requirements for a Web Ontology Language, "http://www.w3.org/TR/webont-req/", W3C
 [5]. Resource Description Framework (RDF) Model and Syntax Specification, "http://www.w3.org/TR/1999/REC-rdf-syntax-19990222", W3C
 [6] Resource Description Framework (RDF): Concepts and Abstract Data Model, "http://www.w3.org/TR/rdf-concepts/", W3C Working Draft.
 [7] VIGNETTE, "Vignette Content Management Server", 2001.
 [8] Web Ontology Language (OWL) Guide Version 1.0, "http://www.w3.org/TR/2002/WD-owl-guide-20021104/", W3C