

## 웹 2.0을 위한 효율적인 태그 관리 시스템의 설계 및 구축

강필구<sup>0</sup>, 김남중, 이예슬, 채진석

인천대학교 컴퓨터공학과

{kpg2976, water09z, yadmny, jschae}@incheon.ac.kr

### A Design and Implementation of Effective Tag Management System for WEB 2.0

Pilgu Kang<sup>0</sup>, NamJoong Kim, Yeseul Lee, Jinseok Chae

Dept. of Computer Science and Engineering, University of Incheon

#### 요약

웹 2.0이 등장하게 되면서 태그를 사용하는 사람들이 점차 늘어나고 있다. 이렇게 태그를 사용하는 사람이 점차 늘어나게 되면 태그를 효율적으로 관리하는 시스템이 필요하게 된다. 본 논문에서는 태그의 장단점을 알아보고 이를 통해서 태그의 장점을 극대화 하고 태그의 단점을 보완할 수 있는 모델링 방법을 제시하였다. 또한, 대표 태그를 설정하는 방법, 대표 태그를 이용하여 연관 태그를 검색하는 방법에 대한 구현 방안을 제안하였다.

#### 1. 서론

인터넷의 발달과 사용자의 적극적인 참여에 힘입어 웹 서비스 환경은 다양하게 변화하고 있다. 기존의 웹 서비스가 정적이고 수동적인데 반해 최근의 웹 서비스는 점차 동적이고 능동적으로 변화하고 있다. 이러한 웹 서비스 변화의 흐름을 잘 반영하는 것이 웹 2.0이다[1].

웹 2.0의 주요한 특징은 사용자가 능동적으로 참여한다는 것이다[2]. 기존의 웹이나 시맨틱 웹은 리스트 구조의 로그를 연상시킨다. 게시판처럼 리스트를 보여주고 검색을 통해 결과를 추출해내며, 컴퓨터에 의해 자동으로 정보를 생산하거나 수집한다. 하지만 웹 2.0에서 대부분의 정보는 사용자의 의해 생산되고, 사용자가 붙인 태그에 의해 분류되고, 연관성 있는 태그에 의해 서로 관계를 맺는다.

사용자는 게시판이나 블로그를 통하여 만든 정보를 태그를 이용하여 분류한다. 이는 기존의 단일 분류 구조를 벗어나 다중 분류 구조의 확장성을 제공한다. 하지만 다중 분류 구조의 확장성은 의미론적인 단점을 가지고 있다. 책을 소개하는 글에 있어 사용자는 분류 구조에 책이라고 태그를 붙이지만 다양한 생각과 사고를 하는 사용자들은 도서, book 같이 다양한 태그를 붙일 수 있다. 이는 의미론적으로는 같을 수 있지만, 엄밀히 말하면 서로 다른 태그가 되는 것이다.

웹 2.0이 등장하게 되면서 엄청나게 쏟아져 나오는 정보들을 분류하고 관계를 맺어줘야 할 필요성이 있다. 가공된 검색의 로우 데이터로서가 아닌 진정한 분류 및 관계를 형성해줘야 하고, 생산된 로우 데이터를 효율적으로 설계하고 관리하기 위한 방안을 제시할 필요가 있다[3].

본 연구의 구성은 2장에서 웹 2.0과 웹 2.0에서 사용되는 태그에 대해서 살펴보고, 3장에서는 블로그를 이용하여 RSS에서 추출한 태그 정보 효율적으로 저장하는 시스템을 설계하고 4장에서 구현 방안을 제시하였다. 마지막으로 5

장에서 결론을 제시하였다.

#### 2. 관련 연구

##### 2.1 태그의 정의

태그는 학생들의 이름표, 수하물의 딱지, 제품의 상표를 뜻한다. 웹에서도 태그 기술은 어떤 글이나 자료에 붙여놓은 추가정보를 뜻한다. 태그는 하나의 정보에 부가적인 설명을 기록할 수 있다는 점에서 키워드와 유사하며 글을 분류 할 수 있다는 점에서 카테고리과 유사하다[4]. 하지만 태그는 여러 단계를 거칠 필요가 없고 특별한 규칙이 없기 때문에 사용자의 편의성이 높아진다. 또한 태그는 사용자가 마음대로 지정한 태그가 중요한 정보가 되고 태그와 태그 사이에서 입체적으로 다양한 정보들과의 관계를 형성하며 서로 사용할 수 있다는 점에서 키워드와 카테고리과 다른 장점을 가지고 있다[5].

##### 2.2 태그의 활용

태그가 웹 2.0에서 가장 주목 받는 이유는 글과 글 사이의 관계를 맺을 수 있다는 것이다. 예를 들어 사용자가 다음과 같은 태그를 붙었다고 가정해보자.

- 글 1. 설기현 맹 활약 [태그:설기현, EPL, 축구]
- 글 2. 수원 삼성 우승 [태그:K리그, 축구, 우승]

전혀 다른 글 1과 글 2는 축구라는 태그를 통해서 연결 될 수 있다. 단순히 글 1을 보고 쓰는 사람도 글 2와도 연결이 될 수 있다. 여기에서의 두 가지 개념, 연관 태그와 연관 글이 있는데, 연관 태그란 하나의 중심 태그, 즉 축구라는 태그와 연관도가 높은 태그의 집합을 의미한다. 연관 글은 각 정보의 태그들을 통해서 글이 연결 될 수 있고 중복된 태그의 개수가 높아질수록 정보의 연관도가 높다고 평가할 수 있다. 그리고 연관 태그

\* 이 논문은 산업자원부 지정 인천대학교 동북아전자물류 연구센터의 지원에 의한 것이다.

의 모음에서 가장 높은 연관성을 가지고 있는 것이 대표 태그가 될 수 있다[6].

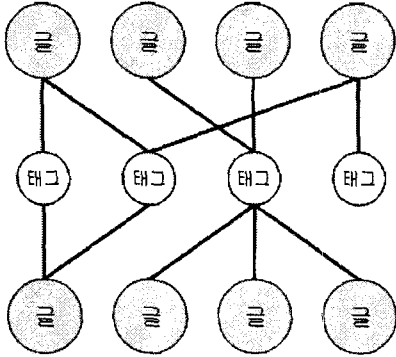


그림 1. 태그를 통한 글 사이의 관계

또한 태그는 정보의 트렌드를 분석 할 수 있다. 최근에 생산되는 정보들에서 집중적으로 사용되는 태그를 찾아내고, 그러한 태그들을 통해서 관련 정보를 습득하는 시스템, 정보에 태그가 붙어진다면 트렌드 분석도 가능할 것이다.

### 2.3 태그 구름과 태그 맵

태그 구름(Tag Cloud)은 태그들을 아무런 분류 없이 나열한 것을 가리키는데, 그 모양이 구름 같다고 하여 태그 구름이라고 부른다. 태그 맵(Tag Map)은 서로 연관성 있는 태그들을 태그 구름 형식으로 모아 놓은 것을 가리킨다. 관련된 연관 태그를 링크로 묶고, 태그의 규모를 적용시켜보면 하나의 큰 태그 맵을 만들 수 있다. 태그 구름은 한눈에 모든 태그들의 트렌드를 살펴볼 수 있지만 전체적인 태그의 연관성에 대한 표현은 부족하다. 태그 맵은 단순히 트렌드의 측면을 증가하여 태그와 태그 사이의 연관 관계까지 표현하고, 나아가서는 태그를 탐색해 나가는 중요한 방식 중 하나가 될 수 있을 것이다.

## 3. 시스템 설계

### 3.1 데이터베이스 모델링

태그 관리 시스템은 표 1과 같이 크게 4개의 테이블로 구성된다.

표 1. 태그 관리 시스템 주요 테이블

테이블이름	설명
ITEMLIST	RSS 또는 사용자가 입력한 글의 정보
TAGLIST	태그 정보를 객체화 하여 관리
TAGMAP	글과 태그를 연결하는 게이트웨이
TAGANCESTOR	태그 계층, 연관태그 설정

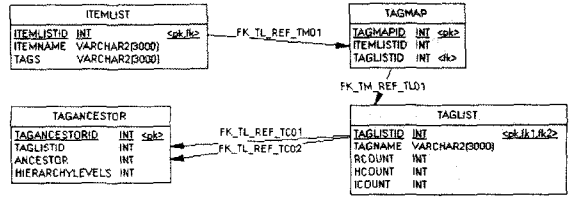


그림 2. 태그 관리 시스템 주요 테이블 스키마

태그 관리 시스템은 태그 관리를 유용하게 하는 것을 목적으로 하고 있다. 태그 관리를 유용하게 하는 방법에는 두 가지가 있다.

1. TAGMAP 테이블을 이용한 글과 태그를 연결하는 게이트웨이 방식.
2. TAGANCESTOR 테이블을 이용한 태그 하이러키 방식.

태그 게이트웨이 방식은 객체화된 태그를 글과 따로 분리함으로써 사용자가 원하는 형식으로 다양화 보여줄 수 있고, 태그 하이러키 방식은 태그와 태그 사이에 연결 관계에 따른 연관 태그, 대표 태그를 설정할 수 있다.

### 3.2 태그 정보 생성

태그 정보가 생성된다는 것은 사용자가 새로운 글을 입력하거나 RSS를 통해서 새로운 정보를 수집하는 것을 의미한다. 예를 들어 사용자가 입력한 태그가 책, 이문열, 삼국지를 입력했다고 가정한다면 테이블에는 다음과 같은 정보가 입력된다.

표 2. ITEMLIST 테이블

TAGLISTID	ITEMNAME	TAGS
1	삼국지를 읽고	책, 이문열, 삼국지

표 3. TAGLIST 테이블

TAGLISTID	TAGNAME	RCOUNT	HCOUNT	ICOUNT
1	책	0	0	1
2	이문열	0	0	1
3	삼국지	0	0	1

표 4. TAGMAP 테이블

TAGMAPID	ITEMLISTID	TAGLISTID
1	1	1
2	1	2
3	1	3

태그 정보를 입력시 ITEMLIST, TAGLIST, TAGMAP 테이블

들에 정보가 서로 관계를 맺으면서 입력된다. 또한 TAGLIST 테이블에 ICOUNT 컬럼의 값을 하나 증가 시킨다.

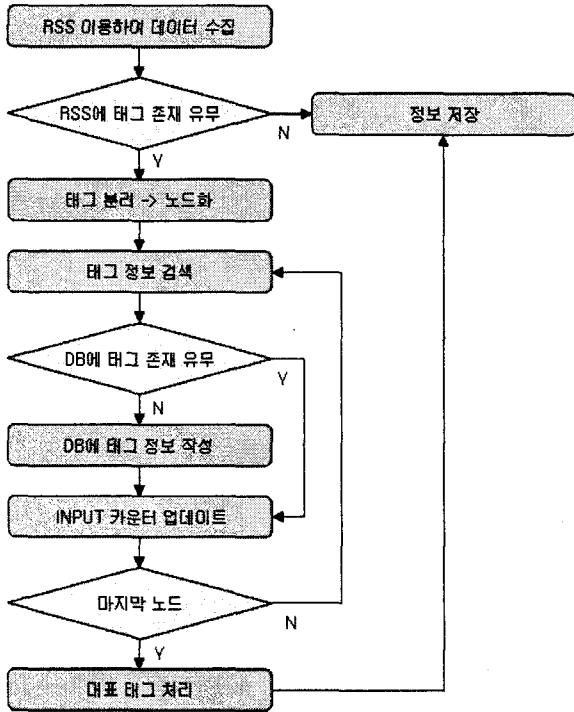


그림 3. 태그 정보 생성

### 3.3 태그 정보 조회

태그 정보가 조회되는 것은 사용자가 게시물을 조회하는 것과 태그 맵을 이용하여 태그를 검색하는 경우로 나눌 수 있다. 태그 정보를 조회 하는 경우는 DB에 새로운 정보는 입력하지 않고 조회한 결과에 따라서 특정 값을 수정한다. 사용자가 게시물을 조회하는 경우는 TAGLIST 테이블에 RCOUNT 컬럼의 값을 하나 증가 시키고 태그를 검색하는 경우는 TAGLIST 테이블에 HCOUNT 컬럼의 값을 하나 증가 시킨다.

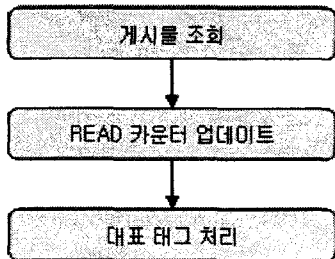


그림 4. 게시물 조회

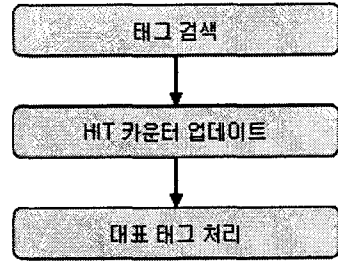


그림 5. 태그 검색

### 3.2 연관 태그, 대표 태그 설정

태그와 태그를 연결하기 위해서는 사용자가 직접 태그와 태그 사이의 관계를 정의해야 한다. 예를 책과 도서, 책과 book을 연결한다고 가정하면 책과 도서, 책과 book, 도서와 book는 서로 연관 태그로 등록된다. 그리고 책과 도서, book중에서 대표 태그는 태그 중 RCOUNT, HCOUNT, ICOUNT의 합이 가장 높은 것으로 결정된다. 따라서 표 5에 따라서 대표 태그는 책으로 결정 된다.

표 5. TAGLIST 테이블

TAGLISTID	TAGNAME	RCOUNT	HCOUNT	ICOUNT
1	책	0	0	2
4	도서	0	0	1
5	book	0	0	1

표 6. TAGANCESTOR 테이블

TAGANCESTORID	TAGLISTID	ANCESTOR	HIERACHYLEVELS
1	1	1	0
2	4	1	1
3	5	1	1

## 4. 시스템 구현

### 4.1 개발 환경

태그 관리 시스템은 윈도우 환경에서 ASP와 Visual Basic 6.0, Visual C#.Net을 이용하여 개발되었으며, DB는 오라클로 구현하였다.

### 4.2 개발 내용

블로그에서 입력한 글을 RSS 엔진을 통해서 수집하고 수집한 정보를 대상으로 하여 태그를 생성한다. 생성된 태그는 객체화 하여 관리하며 사용자가 글을 작성할 때 태그의 사용빈도, 사용자가 글을 읽고 태그를 검색할 때 태그의 조회빈도를 기준으로 하여 연관 태그 중 대표 태

그를 생성한다. 생성된 연관 태그는 그림 6과 같이 트리 형식으로 표현할 수 있고, 사용자가 태그 검색 시 대표 태그를 중심으로 그림 7과 같이 연관 태그까지 같이 검색 한다.

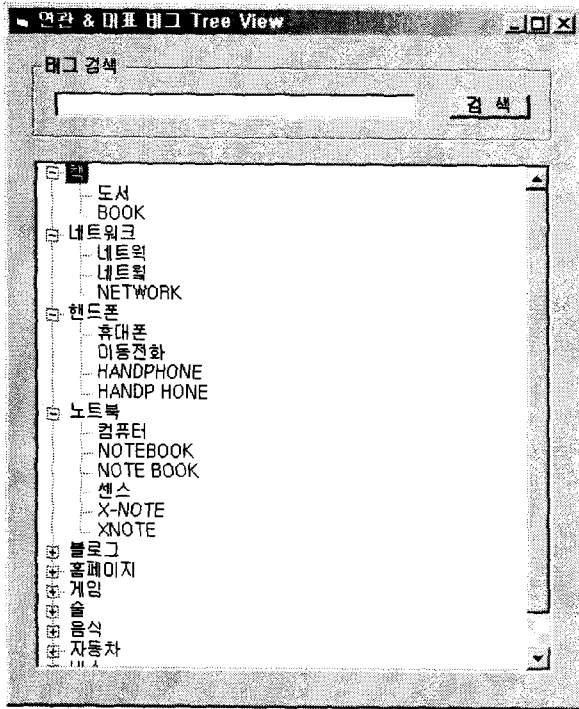


그림 6. 연관 및 대표 태그 Tree View

태그 : 책

이 태그는 현재까지 2개의 글이 연결되어 있습니다.  
이 태그는 {도서, BOOK} 태그를 대표합니다.

이 태그의 최신 블로그

글 제목	블로그 이름	시간
비밀번호 찾기 방법	중성비	09.03
비밀번호	중성비	09.03

그림 7. 연관 태그 검색 결과

대표 태그를 이용하여 연관 태그에 속한 태그 객체를 조회하는 질의는 다음과 같다.

```

SELECT
    TAGLISTID,
    TAGNAME
FROM TAGLIST
WHERE TAGLISTID IN
    ( SELECT
        ANCESTOR
        FROM TAGANCESTOR
        WHERE TAGLISTID =
            ( SELECT TAGLISTID
              FROM TAGLIST
              WHERE TAGNAME = :SearchText
            )
    )
    
```

### 4.3 성능 평가

태그 관리 시스템은 추가(insert)나 갱신(update) 비용이 메타정보를 하나의 테이블 구조를 이용하는 스키마 구조보다 많이 들어간다. 하지만 그 비용보다는 선택(select)의 비용이 훨씬 줄어든다.

표 7은 100만개의 데이터가 입력된 상태에서 동일한 환경에서 동일한 데이터를 반복입력하고, 동일한 내용을 검색하고 트리 형식으로 사용자에게 보여주는 것을 50회 반복하고 그에 따른 각 스키마의 추가, 갱신, 선택 비용을 보여준다.

표 7. 테이블 구조에 따른 추가, 갱신, 선택 비용

	태그 관리 시스템	다른 스키마 구조
추가 비용	0.07s	0.03s
갱신 비용	0.12s	0.07s
선택 비용	0.02s	0.89s

표 7에서 볼 수 있듯이 선택 비용에서 큰 효과를 보여준다. 이는 전체의 태그를 보여주거나 태그를 트리 형식으로 표현하거나, 사용자가 원하는 태그를 표시할 때 효과적으로 관리 할 수 있다.

### 5. 결론 및 향후 과제

본 연구는 웹 2.0이 이슈화 되면서 대두되고 있는 블로그나 게시판 등에서 생성되는 태그를 관리하기 위한 태그 관리 시스템을 구현에 관한 연구로서 태그를 유연하게 관리할 수 있고, 관련 있는 태그를 연결함으로써 사용자가 원하는 태그 결과를 좀 더 유연하게 연결시킬 수 있게 되었다. 연결 태그를 생성하는데 있어 사용자가 생성된 태그의 의미를 생각하여 태그와 태그 사이의 관계를 맺어야 하는데, 이를 자동화 하여 입력할 수 있도록 확장이 필요하다. 또한 대표 태그를 선정하는 기준을

확장하고 불법적으로 조회빈도를 높일 수 있는 현상을 방지해야 한다. 본 태그 관리 시스템은 블로그나 게시판 이외에도 게시판 구조로 이루어진 어떠한 구조에서도 적용할 수 있도록 표준화 적용 방안 연구가 필요하다.

## 6. 참고 문헌

- [1] <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [2] 최호찬, "인터넷의 새로운 문화 블로그", 경향잡지 pp107-109, 2004
- [3] <http://www.onlamp.com/pub/a/onlamp/2006/06/08/designing-tag-clouds.html>
- [4] <http://www.dbguide.net/know/know101003.jsp?IDX=1049&K=TITLE&K2=REGID&catenum=14>
- [5] <http://coelomic.wordpress.com/2006/01/02/tag-clouds/>
- [6] 박영욱, "웹2.0 구현의 핵심, 태그", 마이크로 소프트웨어 5월호, 2006