

## 순위 조작 방지를 위한 소셜 태깅 서비스 설계

정한영<sup>o</sup>, 최옥경<sup>\*</sup>, 예홍진<sup>\*</sup>

<sup>o</sup>아주대학교 지식정보보안학과

e-mail: {google, okchoi, hjyeh}@ajou.ac.kr

# A Design of Social Tagging Services for Prevention of Manipulation of Ranking

Han-Young Jung<sup>o</sup>, Okkyung Choi<sup>\*</sup>, Hongjin Yeh<sup>\*</sup>

<sup>o</sup>Dept. of Knowledge Information Security, Graduate School of Ajou University

### ● 요약 ●

최근 소셜 네트워크 서비스(Social Network Services)를 활용한 소셜 태깅 서비스에 대한 연구가 활발하게 진행 중이다. 특히 태그 기술을 이용한 협력적 태깅 시스템(collaborative tagging system)은 북마크, 문서, 사진, 동영상과 같은 웹 자원을 조직화하고 공유할 수 있는 수단으로 제공하고 있다. 그러나 광고 홍보 목적을 가진 스파머들은 콘텐츠와 관련 없는 태그를 달아 놓아 검색 키워드와 무관한 결과 값이 검색되어 웹 검색 서비스를 이용하는 사용자들에게 불편을 주고 있다. 따라서 본 연구에서는 콘텐츠와 태그의 연관성을 높이기 위해 태그에 일반 사용자가 유사도를 입력할 수 있는 추천 시스템을 적용하여 순위 조작 방지 방법을 제안하였다.

키워드: 소셜 태깅 서비스(Social Tagging Service), SNS(Social Network Services), 순위 조작(Ranking Manipulation)

## I. 서론

투자 전문가들이 공동 운영하는 블로그 ‘더빅픽처(www.ritholtz.com/blog)’에서 60초마다 인터넷에서 일어나는 일들을 한눈에 볼 수 있는 인포그래픽을 게재하였다. 페이스북은 69만개 이상의 업데이트, 블로그 포스팅 1천500개, 유튜브 동영상 600개 업로드, 구글 검색 69만 4천445개. 이것은 하루 이용량이 아니라 1분 동안에 일어나는 일들이다. 이와 같이 빠른 전파성으로 소셜 네트워크 서비스(SNS)는 빠르게 사용이 급증되면서 태그 기술을 이용한 딜리셔스(Delicious)나 플리커(Flickr)와 같은 협력적 태깅 시스템(collaborative tagging system)은 웹 사용자가 북마크, 문서, 사진, 동영상과 같은 웹 자원을 조직화하고 공유할 수 있는 수단으로 제공되고 있다.[1] 협업적 태깅 시스템을 통해 사용자들은 자원에 태그를 달 수 있다. 태깅 시스템은 특별한 사용 기술 없이도 이용할 수 있어 많은 사용자들에 의해 매우 짧은 기간 안에 광대한 양의 정보가 축적되었고 지금도 엄청난 정보들이 저장되고 있다.[2] 태깅 시스템에 사용자가 올려놓은 의미 있는 정보에 대한 태그를 달아 놓아 다른 사람들이 검색을 통해 원하는 정보를 효율적으로 공유하는 시스템으로 발전하고 있다. 그러나 광고의 목적을 가진 스파머들은 콘텐츠와 상관없는 태그를 달아 놓아 검색과 전혀 상관없는 결과가 검색되게 하는 불편함을 주고 있다. 하지만 소셜 태깅 서비스를 이용하여 정확한 랭킹(Ranking)을 지원 한다는 것은 쉬운 일이 아니다.

따라서 본 연구에서는 이러한 콘텐츠와 태그의 연관성을 높이기

위해 태그에 유사도를 적용하여 랭킹 조작을 방지할 수 있는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로서 소셜 태깅 서비스와 폭소노미의 개념, 집단지성을 활용한 추천 시스템에 대해 알아보고 3장 제안 방법에서는 시스템 흐름도와 게시 글 정보와 태그 정보에 정확성을 높이기 위한 유사도 적용 방법을 제시하고 있다. 마지막으로 4장에서는 결론 및 향후 연구에 대한 설명을 한다.

## II. 관련 연구

### 1. 관련연구

#### 1.1 소셜 태깅 서비스

소셜 태깅(Social Tagging)은 사진이나 음악, 공동 주체의 텍스트 등에 키워드를 다는 것이다. 이는 컴퓨터에 의해 생산되는 정형화된 텍소노미와 다르게 사용자들 의해 제안하여 만들어지는 즉, 협업적 생산을 통해 만들어지는 결과물이다. 이러한 소셜 태깅은 폭소노미라고 부르기도 한다.

소셜 태깅은 세 개의 주요한 객체인 사용자(user), 태그(tag), 리소스(resource)가 있으며 기존의 정보와 분류라는 일대일 관계를 깨고 일대다로 묶어 주기 때문에 예상하지 못했던 콘텐츠가 태그를 통해 사용자와 리소스와의 관계를 서로 연결시켜준다. 우연히

만난 태그는 새로운 리소스를 접하게 되면서 끊임없는 항해를 계속 할 수 있으며, 자연스럽게 사용자와 태그, 리소스가 유기적으로 연결 된다.[3]

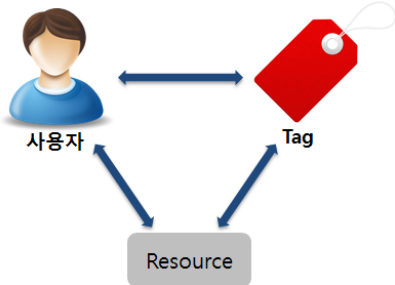


그림 1. 소셜 태깅 서비스의 관계도  
Fig. 1. Relationship for Tagging Social Services

### 1.2 집단지성과 추천 시스템

웹2.0의 원동력은 소수의 전문가가 아닌 일반 사용자가 만들어 가는 인터넷 환경이다. 즉, 일반 사용자가 참여하고 만든 자원을 공유하여 협업과 집단지성을 통해 만들어가는 인터넷 환경이라고 할 수 있다. 유용한 정보들과 사진, 동영상과 같은 일반 사용자들이 제공하는 데이터들이 폭발적으로 늘어나면서 보다 효과적으로 사용자가 원하는 정보를 제공하기 위해 추천시스템에 대한 필요가 대두되고 있다. 현재의 추천 시스템은 주로 집단지성을 이용한다 [4]. 사용자가 제공하는 데이터에 태그(tag), 사용자의 평점(User's Rating) 정보들이 포함되면 이러한 데이터는 추천 시스템에 있어서 유용한 정보를 제공하게 된다.[5] 이에 본 논문에서는 상업 혹은 악의적인 목적으로 콘텐츠와 태그를 이용한 랭킹 조작 방법을 방지할 수 있도록 일반 사용자가 태그에 유사도 점수를 직접 입력하는 추천 시스템을 적용하는 기법을 제안한다.

## III. 제안 방법

### 1. 제안 방법

#### 1.1 시스템 흐름도

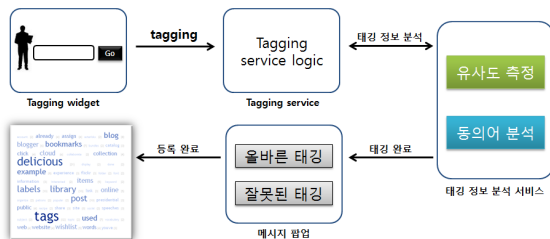


그림 2. 시스템 흐름도  
Fig. 2. System Flow Chart

[그림 2]은 태그 등록 단계로 단계별 프로세스를 살펴보면 다음

과 같다.

- ① 사용자는 블로그나 소셜 북마크와 같은 소셜 네트워크 서비스 콘텐츠에 관련 태그(tag)를 입력한다.
- ② 입력된 태그는 태깅 서비스 단계를 거쳐 태깅 정보 분석 서비스에서 동의어 분석과 유사도 측정 과정을 거친다.
- ③ 입력된 태깅(tagging)들은 태깅 정보 분석 서비스를 통해 올바른 태깅 혹은 잘못된 태깅을 판별하여 메시지 팝업창을 통해 그 결과를 반환 시킨다.
- ④ 만약 올바른 태깅이면 [등록 완료]라는 메시지와 함께 소셜 태깅 서비스에 등록되고 잘못된 태그의 경우에는 사용자에게 재입력을 요구한다.

## 2. 구현 화면

### 2.1 구현 환경

사용된 프로그램은 windows 운영체제에서 apache2, HTML, PHP를 사용하였으며 데이터베이스는 Mysql를 사용하였다.

### 2.2 사용자 구현 화면

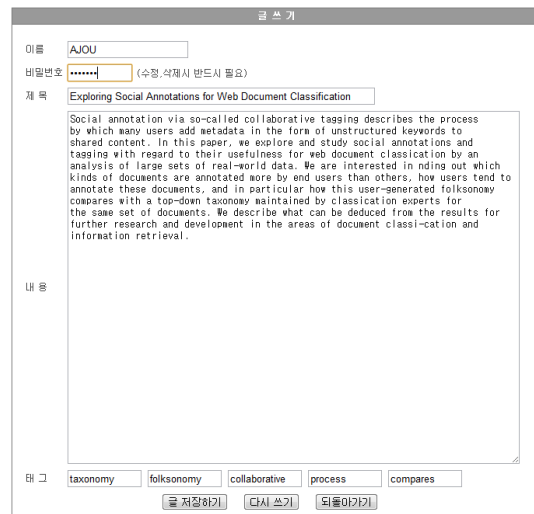


그림 3. User 글쓰기 입력 화면  
Fig. 3. User Input Screen

[그림 3]은 사용자가 게시판에 글을 저장할 수 있는 화면으로 User 구성 화면으로 이름, 비밀번호, 제목, 내용, 글쓰기 form 아래에 태그 5개를 입력할 수 있다. 일반 게시판과 같이 글 저장하기 버튼을 누르게 되면 작성한 글에 대한 정보가 Mysql DB에 저장 이 된다.

### 2.3 Search User 구현 화면

[그림 4]은 게시 글 목록 구성 화면이다. 사용자들이 작성한 글 들은 현재 웹 사이트에 저장 되어있다.

검색을 하지 않아도 게시 글의 제목을 클릭하며 내용을 볼 수 있으며 만약 특정 키워드로 검색을 원한다면 사용자들이 작성한 태그(tag)를 통해 검색이 되어 검색된 글만 목록에 보여 진다.

게시글 목록

번호	제목	글쓴이
30	Exploring Social Annotations for Web Document Classification	AJOU
27	Exploring Social Annotations	Michael
26	rtwetwet	
25	fish	holand
24	Betta From Wikipedia, the free ency	wifew
23	Male and Female Betta Fish Behavior	Korean
22	Can i keep my betta fish in a bowl?	brody
21	Betta Fish and Betta Fish Care	fishman
20	Siamese Fighting Fish	jeny

[검색] (글쓰기)

그림 4. 사용자가 작성한 글 목록  
Fig. 4. Bulletin List written by User

[그림 5]의 구성화면은 사용자가 작성한 제목, 글쓴이, 내용과 태그로 구성되어 있다. 사용자가 입력한 각 태그 아래에는 숫자 입력 상자에 0.0부터 1.0까지의 유사도를 입력하여 저장 할 수 있다. 소셜 네트워크 서비스 사용자들은 사용자가 작성한 게시글의 내용과 태그들을 확인 후 관련성이 높은 태그라고 판단이 되면 0.8이상의 유사도를 부여하고 내용과 관련성이 낮다고 판단되는 태그에 대해서는 0.8미만의 값을 부여한다. 단, 태그의 유사도 입력 횟수를 제한하기 위해 웹 사이트에 로그인한 소셜 네트워크 서비스 사용자만 태그에 유사도 값을 입력할 수 있다. 태그에 유사도를 입력한 후 Submit 버튼을 누르게 되면 0.8이상의 점수를 저장된다.

그림 5. 'folksonomy'에 의한 사용자 콘텐츠  
Fig. 5. User Contents by 'Folksonomy'

[그림 6]은 소셜 네트워크 사용자가 유사도를 입력하고 0.8이상의 점수를 입력하였을 경우 5개의 태그 중 taxonomy, folksonomy, collaborative 태그만 보여주도록 한 화면이다. 입력한 태그 유사도는 DB에 누적되어 저장된다. 평균 점수가 0.8을 계속 유지할 경우 상위 순위 랭킹에 속하게 되고 0.8미만의 점수를 가진 태그는 하위 랭킹에 속하게 된다. 그래서 소셜 네트워크 서비스 사용자가 태그를

검색하였을 때 상위 랭킹의 검색 결과가 우선적으로 보이게 된다.

본 논문에서는 사용자가 게시글을 입력하여 저장할 수 있는 글 쓰기 form구성 화면과 저장된 글을 볼 수 있는 게시글 목록과 태그를 검색할 수 있는 기능을 가진 구성화면으로 되어 있다. 그리고 소셜 네트워크 서비스 사용자가 검색된 게시글을 확인 후 소셜 네트워크 서비스 사용자가 유사도 입력에 참여를 원한다면 로그인 후 유사도를 넣을 수 있도록 하였다.

게시글의 내용과 연관성이 높은 태그가 검색 순위에 먼저 보이도록 하기 위해서 다음과 같은 방법이 적용되었다.

그림 6. 유사도 0.80이상인 태그  
Fig. 6. similarity over 0.8 of Tag

- 1) 사용자가 입력한 내용과 태그에 대한 연관성을 0.0부터 1.0까지의 유사도 값을 적용함으로써 내용과 관련이 적은 태그가 상위랭킹에 보이지 않아 키워드를 통해 검색을 할 경우 관련성이 높은 글을 검색할 수 있는 장점이 있다.
- 2) 사용자가 입력한 내용과 태그에 소셜 네트워크 서비스 사용자가 유사도를 직접 적용함으로써 내용과 관련이 없거나 적은 태그에 0.8미만의 점수를 부여함으로써 태그를 통해 검색을 할 경우 0.8이상의 연관성이 높은 글을 우선적으로 검색할 수 있는 장점이 있다.

#### IV. 결론

본 논문에서는 기존 소셜 서비스 시스템에서 가지는 잘못된 태그입력으로 검색 내용과 다른 결과가 나오는 문제점을 해결하기 위해 효율적이고 정확한 랭킹이 가능하도록 게시글의 내용과 태그를 비교하여 태그에 유사도 값을 적용하는 방법을 제안하였다. 제안한 방법은 직접 사람이 태그와 내용을 비교 분석함으로써 정확도가 높아지지만 소셜 네트워크 서비스 사용자가 로그인하는 과정이 다소 번거로울 수 있는 문제점이 있다. 또한 사용자가 작성한 글과 태그에 동의어와 유의어를 측정하여 사용자에게 정확한 랭킹

을 제공하는 것이 제안하고 있다. 그러나 구현 화면은 소셜 네트워크 서비스 사용자가 판단하여 유사도 값을 입력하는 방법을 보여 주고 있다.

향후 연구에서는 태그의 정보 및 유사도 값을 정확하게 측정하여 검색의 정확성을 높일 수 있는 랭킹 알고리즘으로 발전시켜 기존 다양한 랭킹 알고리즘과 비교 평가를 통해 더욱 효율적인 랭킹 알고리즘으로 발전시키기 위한 연구가 필요하다.

## 참고문헌

- [1] M. G. Noll and C. Meinel, Exploring social annotations for web document classification. In Proc. of ACM Symposium on Applied Computing, pages 2315{2320, Fortaleza, Brazil, 2008.
- [2] Hyujung Park, Sangyu Rho, A Folksonomy Ranking Framework: A Semantic Graph-based Approach, Asia Pacific Journal of Information Systems, 2011
- [3] Joohee Bae, Kyungwon Lee, A Study of User Interests and Tag Classification related to resources in a Social Tagging System, HCI, 2009
- [4] Tim O'Reilly, What Is Web 2.0, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [5] KyeongJong Lee, Gi Hyun Gong, Sanggoo Lee, Tag-Based Collaborative Filtering Approach Using Analysis of the Correlation, Between User's Preference and Tags, kiise Vol. 34, No. 2(C), 2007
- [6] Sujin Park, Sihwa Lee, Daehoon Hwang, A Web Contents Ranking System using Related Tag & Similar User Weight, Journal of Korea Multimedia Society Vol. 14, No. 4, pp. 567-576, 2011