

# 협력 태깅 환경에서의 다양한 태그 추천 방법론 비교 연구

김현우  
서울대학교  
[hwkim@idb.snu.ac.kr](mailto:hwkim@idb.snu.ac.kr)

이강표  
서울대학교  
[kplee@idb.snu.ac.kr](mailto:kplee@idb.snu.ac.kr)

김형주  
서울대학교  
[hjk@snu.ac.kr](mailto:hjk@snu.ac.kr)

## A Comparative Research of Tag Suggestion Method in Collaborative Tagging System

Hyunwoo Kim      Kangpyo Lee      Hyoung-Joo Kim  
Seoul National Univ.      Seoul National Univ.      Seoul National Univ.

### 요 약

인터넷 상에서 사용되는 태그(tag)란, 사진, 동영상과 같은 콘텐츠(contents)를 설명하는 키워드(keyword)를 의미한다. 태그는 해당 콘텐츠를 설명하는 단어이기 때문에 검색에 유용하게 사용될 수 있다. 사용자들은 이러한 태그의 유용성을 잘 알고 있지만, 실제로 사진이나 동영상에 태그를 다는 사용자들은 많지 않다. 이는 사용자들이 태그를 다는 작업 즉, 태깅(tagging)이 어렵기 때문일 수도 있고, 어떤 태그를 달아야 좋은 검색 성능이 나오는지 알기 어렵기 때문이다. 이러한 이유 때문에 사용자들을 도와주기 위한 태그 추천 시스템이 필요하게 된 것이다. 이 논문에서 우리는 현재 연구되고 있는 다양한 태그 추천 방법론들을 비교, 분석하였다.

### I. 서론

최근 YouTube<sup>1</sup>, Flickr<sup>2</sup>, del.icio.us<sup>3</sup> 등의 웹 사이트의 활성화로 인해 인터넷 상의 멀티미디어 콘텐츠의 양이 증가하고 있다. 수많은 콘텐츠 중에서 사용자들이 원하는 멀티미디어 콘텐츠를 얻기 위해서는 검색을 해야 한다. 하지만 이러한 사진, 동영상과 같은 멀티미디어 콘텐츠는 본문의 텍스트 정보를 이용할 수 있는 신문기사, 블로그 글에 비해 검색에 사용할 수 있는 정보가 많지 않다. 유일하게 사용할 수 있는 정보는 멀티미디어 콘텐츠의 제목뿐이다. 이와 같은 상황에서는 태그가 좋은 검색 결과를 나타낼 수 있다. 태그란, 블로그 글, 사진, 동영상 등 해당 콘텐츠를 설명하는 단어를 의미한다.



<그림 1> 다양한 콘텐츠에 대한 태그 사용

예를 들어, 푸른 바다와 모래사장이 펼쳐진 사진에는 '푸른 바다', '하와이', '백사장' 과 같은 태그가 붙어 있을 수 있다. 단순히 사진과 제목만 있을 때에는 그 바다가 어느 곳의 바다인지 알 수 없다. 하지만 '하와이' 라는 태그가 있다면 그 사진에 대한 정보가 훨씬 더 풍부해 지는 것이다. 동영상이나 사진에 제목 이외에도 해당 멀티미디어를 설명하는 단어인

태그가 있다면 그렇지 않은 상황보다 더 풍부한 검색 결과를 얻어낼 수 있다.

태그는 이와 같이 콘텐츠에 대한 메타데이터를 풍부하게 하고 그에 따라 검색 결과의 품질을 높일 수 있다. 태그를 콘텐츠에 추가하는 작업을 태깅이라고 한다. 일반적으로 태그는 사용자 자신이 작성한 블로그 글이나 사진, 동영상 등에 다는 것이 일반적이다. 하지만 다른 사용자가 게시한 콘텐츠에도 태그를 달 수 있는 환경을 협력 태깅 (collaborative tagging) 환경이라고 한다. 사용자의 PC 에 존재하는 웹 브라우저의 즐겨찾기를 웹 상으로 옮겨놓은 del.icio.us 의 경우에는, 사용자들이 자신의 URL 을 웹 상에 저장하고 다른 사용자들과 그 URL 을 공유할 수 있을 뿐 아니라 동일한 URL 에 대해서 여러 사용자들이 태그를 추가할 수 있다. 이러한 방식으로 del.icio.us 는 협력 태깅 환경을 제공하고 있다.

사용자들이 태깅의 유용성과 가치에 대해서 인식 하고 있음에도 불구하고, 태그를 사용하는 사용자가 많지 않은 것은 콘텐츠에 태그를 다는 작업이 귀찮다고 생각하거나, 어떤 태그를 달아야 더 많은 사용자들에게 노출되고, 좋은 검색 결과를 나타낼 수 있는지 모르기 때문일 수 있다. 이러한 이유들 때문에 사용자들을 도와줄 수 있는 태그 추천 시스템이 필요하게 된 것이다. 다양한 콘텐츠에 태그를 추가할 수 있는 협력 태깅 환경에서, 기존의 다른 사용자들이 달아놓은 태그 데이터를 이용해서 태그 추천을 사용하면 태깅 과정이 간단해지고, 좋은 검색 결과를 얻어낼 수 있다.

### II. 태그 추천 방법론

현재 많은 사람들이 다양한 태그 추천 방법에 대해서 연구를 진행하고 있다. 대부분의 연구들은 기존의 사용자들이 달아놓은 태그와 해당 콘텐츠를 저장하고 콘텐츠와 태그를 분석하여 새로운 입력에

<sup>1</sup> <http://www.youtube.com>  
<sup>2</sup> <http://www.flickr.com>  
<sup>3</sup> <http://www.delicious.com>

대해서 태그를 추천하는 방식을 취하고 있다.

태그를 추가하려고 하는 콘텐츠가 블로그 글이나 신문기사와 같이 텍스트 정보를 이용할 수 있는 경우에는, 텍스트 정보를 이용할 수 없는 사진이나 동영상과 비교하여 상대적으로 쉽게 더 좋은 결과를 얻어낼 수 있다 [1]. 사용자가 블로그 글을 입력하면 기존에 저장된 블로그 글 중에서 비슷한 글을 찾아내서 그 글에 달려있는 태그를 추가하는 방식으로 태그를 추천해 줄 수 있다. 사용자가 입력한 글과 비슷한 블로그 글들을 선정한 후에 그 글에 달려있는 태그들을 추출하여 모은다(aggregation). 모여진 태그들 중에서 관련된 태그를 선별(filtering)하고 랭킹(ranking)을 매겨서 사용자에게 추천해 주는 방식으로 진행된다. 비슷한 방식으로 콘텐츠의 제목을 기반으로 태그를 추출한 뒤에 글 안에서 동시 출현(co-occurrence)한 태그들을 이용해서 추천하고자 하는 태그의 범위를 확장시키는 방법도 있다 [2]. 앞의 사용자가 입력한 글과 비슷한 글을 찾아내는 방법과 다른 점은, 앞의 방법은 블로그의 텍스트 전체를 사용하지만 이 방법은 텍스트의 일부분 만을 사용한다는 것이다.

위의 방법들과는 다르게 콘텐츠의 본문 텍스트는 하나도 사용하지 않고 태그를 추천하는 방법이 있다 [3]. 사람들이 기존에 Flickr 라는 사진 공유 사이트에 달아 놓은 태그 데이터를 저장하여 동시 출현 빈도(co-occurrence count)를 계산해 두고, 사용자가 새로운 태그를 입력하면 그 태그와 동시 출현 빈도가 가장 높은 태그들을 추천 후보로 정한다. 그 후보들 중에서 관련도가 높은 태그를 선별하고 랭킹을 정해서 관련도가 높은 순서대로 상위 태그들을 추천해 주는 방식이다. 본문 텍스트 정보를 이용하지 않는 또 다른 방법에는, 사용자가 이전에 입력한 태그를 바탕으로 태그 추천을 해주는 방식이 있다 [4]. 사용자가 이전에 어떤 태그를 더 많이 사용하고 어떤 태그를 적게 사용하는지에 따라서 추천하는 태그가 달라진다. 또한 사용자가 입력한 태그가 어떤 태그와 더 많이 동시 출현했는지에 따라서 추천하는 태그가 달라진다.

이전의 방법들은 본문 텍스트를 이용하기 때문에 사진이나 동영상과 같이 본문 텍스트가 없는 멀티미디어 콘텐츠에는 적용하기 어렵다. 하지만 방금 설명한 방법들은 본문 텍스트를 전혀 이용하지 않기 때문에 텍스트가 있는 블로그 글, 신문기사 등에 적용 가능할 뿐만 아니라 제목 이외의 텍스트가 전혀 없는 콘텐츠에도 적용이 가능하다는 것이 장점이다.

사용자들의 협력 필터링 (collaborative filtering) 방법을 이용해서 태그 추천을 하는 방식도 있다 [5]. 사용자, 태그, 콘텐츠 사이의 점수를 매겨서 순위가 높은 태그를 추천하는 방식이다. 태그들을 군집화 (clustering)해서 저장해 두고, 사용자가 태그를 입력하면 그 태그와 같은 군집(cluster) 내에 있는 태그를 추천할 수도 있다 [6]. 이 방법은 태그 추천에 직접 사용된 것은 아니고, 태그 구름 (tag cloud)을 효과적으로 만들기 위한 방법으로 사용되었다.

태그 구름이란, 그림 2 와 같이 태그가 사용된 빈도를 중심으로 많이 사용된 태그의 글자 크기를 크게 나타내고 출현 빈도가 작은 태그의 글자 크기를 작게 나타내는 방식으로 태그의 출현 분포를 나타내는 시각화 (visualization) 방법이다. 이 방법에서는 단순히 출현 빈도만을 생각하는 것이 아니라, 태그들을 군집화 하여 같은 군집끼리 태그 구름을 만들어 시각화 하였다. 이러한 방법을 태그 구름에만 사용할 것이 아니라 태그 추천에도 응용해 볼 수 있다. 미리 태그들을 군집화하여 데이터를 저장해 두고 사용자가 태그를 입력하면 그 태그와 같은 군집에 있는 태그를 추천해 주는 것이다.

### III. 결론

협력 태깅 환경에서 다양한 태그 추천 방법에 대해서 살펴보았다. 대부분의 태그 추천 방법이 태그 사이의 동시 출현 빈도를 기반으로 태그들 사이의 관계를 추출해 낸다. 그리고 그 데이터를 이용해서 태그 추천을 한다. 태그 추천 방법에는 본문 텍스트 정보를 이용하는 방법과 텍스트 정보를 이용하지 않는 방법이 있다. 텍스트 정보를 이용하는 방법은 멀티미디어 콘텐츠에는 사용할 수 없는 대신에 더욱 정확한 태그 추천이 가능하다. 텍스트 정보를 이용하는 방법은 사진과 동영상과 같이 본문이 없는 콘텐츠에 적용이 가능하지만 텍스트 정보를 이용한 방법에 비해 결과가 나쁠 수 있다. 또한 사용자들이 입력한 정보를 이용할 때에, 전체 데이터를 보는 방법은 많은 사람들이 관심 있는 분야에 대해서 좋은 결과를 나타낼 수 있다. 각 사용자의 과거 데이터만을 보는 방법은 그 사용자에게 딱 맞는 개인화된 결과를 나타낼 수 있지만 그 사용자가 이전에 입력하지 않은 태그는 추천이 불가능하다는 단점이 있다.

앞서 언급한 바와 같이 대부분의 방법론이 동시 출현 기법을 사용하고 있다. 다양한 태그 추천 방법론의 장점들을 동시 출현 빈도뿐 아니라 다른 방법을 함께 이용해서 더욱 발전시키면 좋은 결과를 나타내는 태그 추천 방법을 연구할 수 있을 것이다.

### IV. 감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 육성 지원사업(IITA-2008-C1090-0810-0031), BK-21 정보기술사업단, 국토해양부 첨단도시개발 사업의 연구비지원(07 첨단도시 A01)의 연구 결과로 수행 되었습니다.

### References

- [1] G. Mishne, "AutoTag: a collaborative approach to automated tag assignment for weblog posts," WWW, 2006.
- [2] M. Lipczak, "Tag Recommendation for Folksonomies Oriented towards Individual Users," ECML PKDD Discovery Challenge, 2008.
- [3] B. Sigurbjornsson, and R. Zwol, "Flickr Tag Recommendation based on Collective Knowledge," WWW, 2008.
- [4] N. Garg, and I. Weber, "Personalized tag suggestion for flickr," WWW, 2008.
- [5] K. Tso-Sutter, L. Marinho, and L. Schmidt-Thieme, "Tag-aware recommender systems by fusion of collaborative filtering algorithms," SAC, 2008.
- [6] Y. Hassan-Montero, and V. Herrero-Solana, "Improving Tag-Clouds as Visual Information Retrieval Interfaces.," InSciT, 2006



<그림 2> 태그 구름 (tag cloud)