

TagPlus: 폭소노미에서 동의어 태그를 이용한 검색 시스템

이선숙*, 용환승**

요약

태깅은 사용자들이 공유된 콘텐츠에 키워드의 형태로 메타 데이터를 추가하는 과정이다. 최근 이러한 태깅은 웹 상에서 더 많은 사용자들에게 사용되어지고 있는 추세인데, 이런 태깅 사이트는 사용자가 북마크, 사진, 비디오 등의 콘텐츠에 태그를 추가할 수 있도록 한다. 본 논문에서는 사용자의 참여를 바탕으로 하는 태깅 시스템의 구조와 배경 지식 또 이런 시스템이 가지는 다양한 의미와 한계들을 분석한다. 또한 WordNet 데이터베이스의 동의어 집합을 태그의 검색에 적용한 TagPlus 시스템을 제안하고 Flickr 이미지 공유 시스템으로부터 동의어 태그 검색을 가능하도록 구현하였다.

TagPlus: A Retrieval System using Synonym Tag in Folksonomy

Sun-Sook Lee*, Hwan-Seung Yong**

Abstract

Collaborative tagging describes the process by which many users add metadata in the form of keywords to shared content. Recently, collaborative tagging has grown in popularity on the web, on sites that allow users to tag bookmarks, photographs, videos and other content. In this paper, we analyze the structure and basic knowledge of collaborative tagging systems as well as their dynamical aspects. We also present a retrieval system, TagPlus, using synonym tag that is derived from WordNet database. Specifically, TagPlus, a synonym tag based system has users retrieve images from Flickr system. The proposed system show the images tagged by not only the tag that users input but also the synonyms that are synonyms with the tag.

Keywords : Collaborative tagging, Synonym tag, Folksonomy, WordNet

1. 서론

Folksonomy[1]는 인터넷 사용자가 웹페이지, 온라인 사진, 동영상 등의 콘텐츠를 특별한 제한 없이 협력하여 분류하도록 하는 분류시스템이다. 사용자들에 의해 자유롭게 선택된 콘텐츠의 식별 키워드, 카테고리 이름, 또는 메타 데이터를 태그[4]라 하고 태그를 이용하여 분류하는 작업을 태깅[4]이라 한다. 사용자는 자신들만의 유일

한 태그를 사용하여 웹페이지나 이미지에 태깅한다. 이러한 웹페이지나 이미지는 자신을 식별하도록 하는 복수개의 태그를 가질 수 있다. 동일한 태그를 가진 이미지나 웹페이지는 함께 링크되며 사용자는 유사한 웹페이지나 이미지 검색을 위하여 그 태그를 이용할 수 있다. 최근 이러한 태깅을 지원하는 새롭고 다양한 서비스가 창출되고 있다. 이러한 태깅 시스템은 사이트 제작자가 아닌 사용자들에 의해서 만들어 지고 많은 온라인 서비스들을 훨씬 더 접근이 용이하고 유용하게 만든다. 태깅 시스템의 사용이 급격히 증가함에 따라 태그의 양도 방대해지고 있다. 하지만 태깅 시스템이 지원하지 못하는 태그가 가지는 한계들도 엄연히 존재한다. 이러한 한계들로 인하여 태그 검색의 효율성에도 불구하고 때로는 만족스럽지 못하는 검색 결과를 야기한다. 태그의 동의어와 다의어 처리의 부재가 이들 한

※ 제일저자(First Author) : 이선숙
접수일자:2007년06월05일, 심사완료:2007년06월29일
* 이화여자대학교 컴퓨터공학과 석사과정
caffuchino@ewhain.net
** 이화여자대학교 공과대학 컴퓨터공학과 교수
▣ 본 연구는 한국과학재단 특정기초연구지원으로 수행되었음(R01-2006-000-10609-0).

계 중 일부이다.

이를 개선하기 위해 기존의 태깅 시스템의 태그에 WordNet 사전의 데이터를 이용하여 태그의 동의어와 동의어의 검색 또한 가능하게 하여 좀 더 효율적인 태그 검색이 이루어지도록 하는 방안을 모색하였고 이를 가능하게 하는 검색 시스템으로 TagPlus 시스템을 구현하였다. 또한 TagPlus 시스템과 Flickr의 태그 검색 결과를 비교한다.

2. 관련연구

사용자는 같은 태그를 다른 의미로 사용할 수도 있고 다양한 태그를 같은 개념에 적용할 수도 있다. 태그는 사용자의 기호나 습관에 따라 다양하게 표현되어지므로 태그에 의한 검색 시 동의어나 다의어 제어가 이루어진다면 사용자에 게 좀 더 만족스러운 결과를 줄 수 있을 것이다. 본 논문에서는 이에 관련된 최근 연구 동향에 대해서 기술한다.

2.1 시멘틱 위키피디아를 이용한 동의어의 검색 시스템: Synarcher

Synarcher는 위키피디아에서 동의어와 관련 용어 검색을 가능하게 하는 시스템이다[8]. 검색 결과는 그래프의 형식으로 보여준다. 사용자는 결과 그래프를 통해서 동의어나 관련 용어들을 컴퓨터와 대화 형식으로 검색하게 된다. 위키피디아에 글을 올리는 사용자는 그 글의 성격에 맞는 카테고리를 선택해야 한다. 또한 위키피디아의 페이지는 관련 페이지를 하이퍼링크를 통하여 소개한다. 이들 카테고리 와 하이퍼링크는 유사한 문자 검색에 이용된다.

논문에서는 웹 페이지의 부가적인 특징을 위하여 HITS 알고리즘을 확장한 Adapted HITS 알고리즘을 제안 하였다. 각 페이지는 키워드의 집합으로 구성되므로 유사한 페이지의 키워드는 원래 페이지의 키워드를 위한 동의어로 간주한다. 위키피디아의 경우 키워드는 페이지의 제목이다. 따라서 동의어 검색 문제를 유사 페이지 검색으로 접근하였다. 하이퍼링크로 연결된 페이지들 간의 유사한 정도를 허브와 권한개념에 기초하여 공식화하였다. 다른 토픽에 해당되는 그

룹으로 웹페이지를 군집화하기 위하여 계층 군집화 알고리즘을 적용하였으며 제안한 Adapted HITS 알고리즘을 기반으로 Synarcher 시스템을 구현하였다.

사용자가 키워드에 대하여 시스템은 동의어를 자동으로 검색한다. 사용자는 결과 그래프에 나타나는 동의어중 원하는 동의어에 대한 새로운 검색을 할 수 있다. 하지만 그래프에 나타나는 데이터들이 전부 동의어는 아니고 상당 부분이 관련 단어이기 때문에 동의어 검색이 완벽하게 자동화되었다고 할 수는 없다. Synarcher는 동의어가 될 수 있는 후보들을 보여주며 이들 중 실제 동의어만을 찾는 부가적인 검색을 제공한다. WordNet 2.0과 Moby Thesaurus List가 제공하는 동의어 리스트를 이용하여 전문가에 의해 제공된 관련 단어 리스트 중 동의어만 매뉴얼하게 찾아내어 Synarcher의 검색결과와 비교 실험을 하였다. Synarcher를 통해 위의 사전에 정의된 동의어가 아닌 관련 단어가 검색 결과로 나타나지는 경우가 많다.

Synarcher는 동의어뿐만 아니라 관련 용어 검색을 성공적으로 수행하였으며 동의어 사전에 등록되어 있지 않은 새로운 동의어 또한 검색했음을 보여준다. 이는 위키피디아라는 백과사전의 특징으로 인하여 나타나는 현상이라 설명되어 질수 있겠다.

2.2 태그 공간에서 검색을 향상시키는 자동화된 태그 군집화

태그 클러스터링을 통하여 관련 태그를 찾아 사용자가 원하는 정보와 좀 더 근접한 검색 결과를 제공하고자 하였다.

특정 태그와 밀접하게 관련된 태그를 찾기 위한 알고리즘은 어떠한 태그들이 같은 페이지에 얼마나 자주 사용되었나를 나타내는 동시 발생 수를 기본으로 한다.

태그의 집합을 의미상 관련된 태그들의 비-홍미 그룹으로 분할하여 각 태그로 태깅된 리소스에 웨이트를 두어 태그간의 거리를 계산하고 같은 클러스터에 속하는 태그의 쌍의 유사도 수를 증가시켜 가장 유사한 N개의 유사 태그를 선택한다[9].

조정 함수(modularity function)를 이용하여 노드들의 특정 클러스터링의 질적 효율을 측정

하고 크기가 큰 클러스터는 그래프 클러스터링 알고리즘을 기반으로 조정 함수의 값을 비교하여 작은 클러스터로 분할한다. 클러스터링은 무질서한 태그 스페이스에서 자동으로 관련 태그를 찾기에 필요한 기술이며 이는 태깅 서비스의 사용자에게 좀 더 효율적인 검색을 제공한다.

2.3 이미지 검색을 위한 태깅과 시맨틱의 통합

태그만으로 이미지 검색 시 야기되는 동의어, 다른 언어의 태그, 틀린 철자, 다의어 등의 문제를 해결해보고자 태그 이외에 비주얼 특징들을 검색에 이용한다[10].

이미지의 태그는 이미지로부터 추출해 낼 수 없는 정보를 가지는 반면 개인적인 태그는 이미지들의 공통의 형태를 표현하지는 않는다. 모든 사용자들이 특정 이미지에 같은 카테고리를 공유하지도 않기 때문에 태그를 어느 한 카테고리에 대응되도록 할 수도 없다.

외관상 비슷한 이미지는 태그를 통해서 언어 지고 사용자가 이 이미지들 중에서 어떤 이미지로 원하는 검색을 할 것인지를 결정한다. 색과 색조의 선형 거리의 조합으로 계산된 거리로 이미지간의 유사 정도를 측정한다. 이런 유사 검색의 결과 이미지들을 사용자는 최종검색을 위해 사용여부를 선택하며 이런 선택은 이미지 검색에서 결과의 적절성에 대한 피드백의 역할을 한다.

일차적으로 태그에 의해 검색된 이미지로부터 비주얼 특징을 이용하여 재검색을 하여 나온 결과는 처음 사용자가 선택한 이미지와 유사한 이미지를 가져오며 그의 태그를 살펴보면 처음 태그의 동의어이거나 다른 언어로 표현되어진 태그를 실험 결과에서 볼 수 있다. 이 논문에서는 태그와 비주얼 특징들을 가지고 이미지를 검색하였을 경우 태그의 동의어를 발견할 수 있음을 증명하였다.

2.4 Flickr 클러스터

Flickr는 유사한 태그들을 그룹화 하여 관련 태그로 묶는 Flickr 클러스터를 제공한다. 예를 들면, "turkey"는 WordNet에 따르면 조류, 음식, 나라, 사람, 실존작 등의 의미를 가진다. (그림 1)의 "turkey" 클러스터를 살펴보면 나라, 음식,

터키의 경치와 관련된 4개의 클러스터를 볼 수 있다.

[istanbul](#), [mosque](#), [bluemosque](#), [blue](#), [bosphorus](#), [travel](#), [bridge](#), [sultanahmet](#), [topkapi](#), [church](#)

➔ [See more in this cluster...](#)

[thanksgiving](#), [food](#), [dinner](#), [family](#), [stuffing](#), [christmas](#), [bird](#), [holiday](#), [cooking](#), [november](#)

➔ [See more in this cluster...](#)

[türkiye](#), [türkei](#), [turquie](#), [törökország](#), [türkiye](#), [turkve](#), [turkish](#)

➔ [See more in this cluster...](#)

[sea](#), [boat](#), [water](#), [sunset](#), [sky](#), [sun](#)

➔ [See more in this cluster...](#)

(그림 1) "turkey" Flickr 클러스터

Flickr는 태그를 군집화하기 위해 태그의 동시 발생 수를 사용한다. "turkey"와 "stuffing"이 함께 태깅된 이미지는 음식 관련 클러스터에 나타나며, "turkey"와 "mosque"는 나라와 관련된 클러스터에 나타난다.

Flickr는 Flickr 클러스터를 통해 태그의 불명확한 의미를 해소하려는 방법을 제공한다. Flickr 클러스터는 사용자들이 더 나은 검색을 하도록 하지만 한계를 가지고 있다. 동시 발생 수는 어떤 이미지에 한개 이상의 태그가 존재해야 함을 의미한다. "turkey"만으로 태깅된 이미지의 양도 무시할 수 없으며 이런 이미지는 어느 클러스터에도 속할 수 없다. 게다가 Flickr 클러스터는 여전히 명확하지 않다. 몇몇 클러스터는 (그림 1)의 네 번째 클러스터의 "sea"와 "Turkey"처럼 보는 즉시 무엇을 의미하는지 알 수 없다. 이는 Flickr 클러스터가 언어적 의미를 알지 못함을 지적한다. 또한 동의어를 같은 클러스터에 속하도록 하지 않는다. 사용자가 특정 이미지를 어느 클러스터에서 찾을 수 있는지를 예측하기란 쉬운 일이 아니다.

Flickr 클러스터는 클러스터를 4개로 제한하였기 때문에 더 많은 그룹으로 군집화 될 수 있는 경우에는 자주 사용되지 않은 태그들은 도태된

다. Flickr 클러스터는 태그들 간의 의미적인 유사성보다는 어느 태그와 얼마나 자주 동시에 쓰이냐를 기준으로 하기 때문에 이들 클러스터만으로 태그들 간의 관계를 통하여 의미의 유사성을 찾기는 아직은 어렵다.

2.5 폭소노미 마이닝 기반의 웹페이지 추천 시스템

기존의 웹 페이지 추천 시스템은 필터링을 사용한다. 이는 사용자가 구매한 상품으로 사용자의 기호를 파악한다. 하지만 상품의 양이 방대해지면 필터링은 좋은 결과를 주지 못한다. 이를 해결하기 위하여 사용자와 그들의 태그간의 친화도(affinity level)를 계산하여 사용자들의 웹페이지 선호도를 표현한다. 제안된 이러한 접근은 사용자의 기호를 파악하고 유사한 기호의 사용자를 찾는데 용이하다[11]. 친화도는 관계의 정도를 보여주며 TF-IDF공식을 기본으로 추출한 값이다. 태그들 간의 유사도를 기본으로 밀접하게 관련된 태그들을 군집화 한다. 각 태그의 부모 태그들을 계산하고 이 태그들이 하나의 태그로 군집화 되면 이를 “클러스터 리더”라 한다. 한 클러스터 리더의 부모가 다른 하나에 속하면 이 두 클러스터 리더는 병합된다. 사용자와 태그 클러스터간의 친화도를 계산하여 사용자의 기호를 토픽별로 표현한다. 각 토픽에 해당하는 태그 클러스터의 추천 페이지를 계산하고 각 사용자에게 추천할 만한 추천 페이지를 계산한다.

논문은 실험을 통하여 클러스터의 크기를 최적화하여 동의어 태그를 발견하였다. 이는 동의어로 인한 태그 중복 문제를 어느 정도 해소했음을 보여준다.

3. 폭소노미에서 동의어 태그를 이용한 TagPlus 시스템

태그의 다의성과 동의 관계 문제를 개선하기 위해 태그에 WordNet 데이터의 동의어 집합을 적용하여 태그의 동의어와 다의어를 고려해서 검색하게 하는 TagPlus 시스템을 제안한다.

TagPlus 시스템은 Flickr의 이미지 데이터를 사용한다. Flickr는 50만이 넘는 회원들이 포스트한 이미지를 사용자들에게 보여주는 인기 있

는 사진 공유 서비스이다. TagPlus 시스템은 Flickr 시스템에서 이미지를 검색할 때 사용자의 태그뿐만 아니라 그의 동의어 또한 태그로서 검색에 이용하게 한다.

본장에서는 먼저 동의어의 처리를 가능하게 하기 위해 어떠한 처리 단계가 이루어지는 것을 설명한다. 어떤 아이템의 태그들을 보고 인간은 다른 표현 형태이지만 이들이 같은 의미를 가지는 동의어임을 알 수 있듯이 태깅 시스템에서도 이를 인지한다면 동의어가 가진 문제점은 해결될 수 있다. 이를 위하여 WordNet의 동의어 집합을 채택한다.

온톨로지의 일종으로 간주되고 있는 WordNet은 인간의 어휘 지식에 대한 심리언어학 연구의 성과를 기반으로 프린스턴 대학 인지과학 연구실이 구축해온 영어어휘 데이터베이스이다[7]. 영어 어휘의 정의뿐만 아니라 어휘를 synsets라고 하는 동의어 집합으로 나누고 이들 동의어 집합간의 의미적 관계를 기록한 데이터베이스이다.

한 단어가 여러 가지의 의미를 가질 때 각각의 의미들은 다른 동의어 집합에 속한다. 대부분의 동의어 집합은 많은 의미적 관계를 통해 다른 동의어 집합과 연결 된다. 한 동의어 집합에 속하는 모든 단어는 공통된 의미를 지니므로 이들 단어는 동의어 관계를 가진다.

WordNet은 또한 다의어 수도 제공한다. 이는 그 단어를 포함하는 동의어 집합의 수를 의미한다. 즉 한 단어가 여러 개의 동의어 집합으로 나누어지면 각각의 동의어 집합이 그 단어의 다의어가 된다. 이는 이 동의어 집합의 수만큼의 다의어가 존재함을 의미한다.

WordNet은 인간의 어휘지식을 모방한 만큼 다의성과 동의 관계를 이용하여 의미를 최대한 정확히 표현하고 있다. WordNet의 주요 특징은 단어 형이 아닌 단어의 의미를 구성요소로 하였다는 점이다. WordNet은 자연언어 처리와 정보 검색의 여러 분야에서 널리 이용되고 있다. 따라서 본 논문에서 구현하는 TagPlus 시스템은 동의어 태그를 위하여 WordNet을 이용한다.

(그림 2)와 (그림 3)은 “movie”와 “film”에 대한 WordNet에서 제공하는 사전 내용이다. (그림 2)의 “movie”와 (그림 3) “film”의 첫 번째는 영화의 의미를 가지며 서로 동의 관계이다. (그림

2)와 (그림 3)에서 알 수 있듯이 “film”과 “movie”의 첫 번째 의미는 같은 동의어 집합을 가진다.

The noun movie has 1 sense (first 1 from tagged texts)

1. (77) movie, film, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick -- (a form of entertainment that enacts a story by a sequence of images giving the illusion of continuous movement, "they went to a movie every Saturday night", "the film was shot on location")

(그림 2) “movie”의 WordNet 사전 내용

The noun film has 5 senses (first 5 from tagged texts)

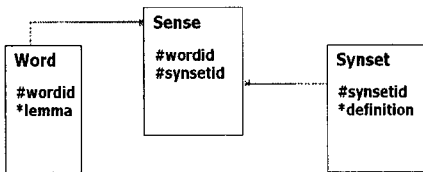
1. (401) movie, film, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick -- (a form of entertainment that enacts a story by sequence of images giving the illusion of continuous movement, "they went to a movie every Saturday night", "the film was shot on location")
2. (94) film, cinema, celluloid -- (a medium that disseminates moving pictures, "theater pieces transferred to celluloid", "this story would be good cinema", "film coverage of sporting events")
3. (42) film -- (a thin coating or layer; "the table was covered with a film of dust")
4. (28) film, plastic film -- (a thin sheet of (usually plastic and usually transparent) material used to wrap or cover things)
5. (16) film, photographic film -- (photographic material consisting of a base of celluloid covered with a photographic emulsion, used to make negatives or transparencies)

The verb film has 2 senses (first 1 from tagged texts)

1. (3) film, shoot, take -- (make a film or photograph of something, "take a scene", "shoot a movie")
2. film -- (record in film; "The coronation was filmed")

(그림 3) “film”의 WordNet 사전 내용

TagPlus 시스템에서는 태그가 하나의 단어가 아닌 <태그, 동의어 집합>과 같은 확장된 태그로 검색을 수행한다. 특정 태그의 동의어 집합을 알기 위해서는 (그림 4)의 데이터 구조가 필요하다.



(그림 4) 데이터 구조

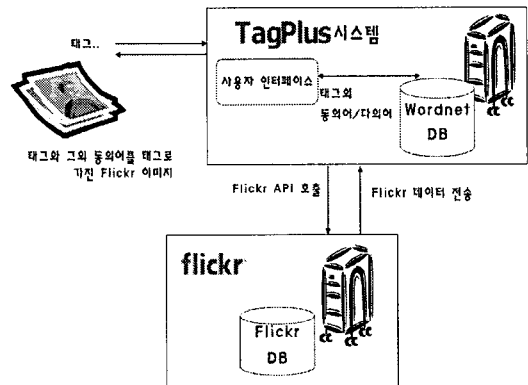
Word 테이블은 단어의 고유한 식별자인 wordid와 정의인 lemma를 갖는 단어의 기본 테이블이다. Synset 테이블은 동의어 집합을 정의한다. 각 동의어 집합에 대한 식별자인 synsetid를 가지며 의미는 definition 속성에 정의되어 있다. Sense 테이블은 각 단어와 그의 동의어 집합 사이의 관계 테이블이다. 한 단어는 여러 개의 동의어 집합을 가질 수 있으며 이때의 각각의 단어와 동의어 집합의 세트는 다의성

을 지님을 알 수 있다. 같은 동의어 집합을 갖는 서로 다른 단어들은 서로 동의어 관계에 있다.

예를 들어, 사용자가 (그림 3)의 첫 번째 의미를 갖는 “film”을 태그로 사용한다고 가정하면 “film”과 동의어 관계에 있는 “movie”, “picture”, “moving picture”, “motion picture”, “picture show”, “pic”, “flick”등의 단어 또한 태그로 사용하여 검색에 이용한다. 이는 위의 단어들이 같은 동의어 집합 아이디 ‘106525881’을 가지므로 가능하다. 따라서 TagPlus 시스템의 검색은 다음과 같은 형식을 취하게 된다.

Searching(리소스,48351,106525881)

괄호안의 48351은 “film”의 wordid이며 106525881은 “film”이 (그림 3)의 첫 번째 의미일 경우의 동의어 집합 아이디이다. “film”이 아닌 “movie” 나 “motion picture”등의 단어를 태그로 이용하여 검색 하더라도 사용자가 검색하고자 하는 의미가 영화라면 동의어에 속한 어떤 단어를 이용하여 검색 하더라도 같은 결과를 볼 수 있도록 하였다.



(그림 5) TagPlus 시스템의 프로세스

TagPlus 시스템의 프로세스는 (그림 5)와 같다. 이 시스템은 사용자가 특정 태그로 검색을 할 경우 그 태그의 동의어와 다의어 관계를 WordNet 데이터베이스에서 찾아 이들 동의어도 태그에 포함하여 Flickr 시스템에서 이 모든 태그로 이미지를 검색하여 사용자에게 결과를 보여준다. 본 시스템의 구현을 위해서는 Apache

어를 중간 결과로 보여줌으로써 사용자에게 좀 더 많은 관련 이미지를 볼 수 있는 기회를 제공한다.

<표 1> "film"의 동의어와 동의어별 Flickr 이미지 검색 건수

Sense	Synonym	Count
Form of entertainment that enacts a story by a sequence of images giving the illusion of continuous movement	film	339984
	flick	3030
	motion picture	147
	motion-picture show	0
	movie	67901
	moving picture	28
	moving-picture show	0
	pic	19996
	picture	41953
a medium that disseminates moving picture	film	339984
	celluloid	221
	cinema	24621
a thin coating or layer	film	339984
a thin sheet of material used to wrap or cover thing	film	339984
	plastic film	5
photographic material consisting of a base of celluloid covered with a photographic emulsion; used to make negatives or transparencie	film	339984
	photographic film	0

"film" 관련 이미지를 보고자 한 사용자는 처음엔 의도하지 않았지만 본인의 선택에 의해서 "film"과 동의어관계에 있는 "movie"로 태그된 이미지 또한 검색을 할 수 있게 되며 이 검색 결과를 통해 사용자가 원하던 이미지를 찾을 가능성은 커진다. 이처럼 사용자의 다양한 표현방식이나 태그 스타일에 따라 다른 단어를 사용하여 태그되었지만 같은 의미를 담은 태그를 가진 이미지의 검색이 가능하다.

TagPlus 시스템과의 이미지 검색 건수 비교를 위하여 Flickr 시스템에서 태그로 검색시, Flickr는 태그 내 빈 칸을 허용하지 않기 때문에 "plastic film"과 같은 단어는 "plasticfilm"처럼

마치 한 단어 인 것처럼 하여 사용하였다.

또한 Flickr에서 태그 "film"은 사용자가 어떤 의미로 이미지에 그 태그를 이용했는지 알 수 없으므로 TagPlus 시스템과 비교 시, <표 1>의 몇 번 의미로 "film"이라는 태그가 이용되었는지에 대해 명확하게 정의할 수 없는 부분은 아쉬운 점이다.

4. 결론 및 향후 과제

태그 사용자들은 지역적으로 문화적으로 다양하다. 폭소노미식 접근의 강점은 개방성에 있다. 사용자는 원하는 대로 스스로의 기준에서 리소스를 묘사할 수 있다. 과연 태깅 시스템에서 다양한 사용자들 간에 태그 사용에 함의를 이끌어내는 것이 바람직한 것인지, 합의에는 도달할 수는 있는지 등에 대한 답은 쉽지만은 않다. 태그를 정리하거나 제약을 가함으로써 폭소노미의 본질과 매력을 잃을 수도 있다. 또한 개인적인 메타 데이터로부터 얻어지는 풍부한 데이터양의 축소를 가져올 수도 있지만 메타데이터의 노이즈를 감소시킴으로써 검색의 정확도와 효율성을 향상시킬 수 있을 것이다.

본 논문에서 제안한 동의어 관계에 있는 태그를 검색에 이용한 TagPlus 시스템은 단순히 무질서한 태그를 정리하거나 태그를 제한하려는 것이 목표가 아니다. 이 시스템은 현재의 폭소노미 환경의 무질서한 태그에 효율적인 검색을 위하여 WordNet 동의어 집합을 적용시켰으며 태그의 동의어 집합에 의한 동의어들의 상호 관계를 태그로 사용하여 검색하게 하였다. 또한 Flickr 시스템과 이미지 검색 방법을 비교하였다.

폭소노미의 장점 중 하나는 사용자들이 관심 있는 태그들의 링크를 쫓아 리소스에 접근하면서 자신에게 의미 있는 정보들을 발견할 수 있는 우연한 기회가 많아졌다는 것이다. 전문가가 아닌 자신과 비슷한 일반 사용자들로부터 추가되는 다양한 태그들이 비록 무질서하기도 하지만 무의미 한 것은 아니다. 다양한 사용자들은 다양한 방식으로 그때의 기분에 따라서 다르게 단어를 선택하기도 하므로 같은 오브젝트를 표현 할 수 있는 여러 가지 단어들을 전부는 아니지만 동의어를 통해서 상당 부분 유추할 수 있

고 이는 곧 검색 결과에 영향을 미친다.

Flickr는 TagPlus 시스템에서 제공하는 동의어와 동의어 태그에 대해 인지하지 못하므로 TagPlus 시스템과 같은 결과를 얻기 위해서는 사용자가 이미지에 태깅 할 때 해당 태그의 모든 동의어를 찾아 이들 모두를 태그로 추가하는 수밖에 없다. 이는 사용자가 같은 의미를 갖는 단어들을 여러 번 태깅하는 비용을 추가시킨다.

본 논문에서 제안한 WordNet 데이터베이스의 동의어 집합을 적용시킨 TagPlus 시스템을 기반으로 단어의 상 하위 개념, 관계어등도 태그를 이용한 정보검색에 이용 할 수 있는 시스템의 확장은 향후 연구 과제이다. 동의어 태그에 의한 검색 결과와 마찬가지로 특정 태그의 관계어를 이용하거나 상 하위 개념을 적용 시킨 정보검색 또한 사용자의 검색결과 만족도에 크게 기여할 것으로 기대된다.

참고문헌

[1] Tim O'reilly, "What Is Web 2.0," Sep 30, 2005.
 [2] Adam Mathes, "Folksonomies-Cooperative Classification and Communication Through Shared Metadata," Computer Mediated Communication - LIS590C MC Graduate School of Library and Information Science. Dec, 2004
 [3] 이선숙, 용환승, "TagPlus: 폭소노미에서 동의어 태그를 이용한 검색 기법," 제 26회 정보처리학회 추계학술발표대회, 충북대학교, pp. 389-392, 2006년 11월 10-11.
 [4] Clay Shirkey, "Ontology is Overrated: Categories, Links, and Tags," Clay Shirky's Writings About the Internet Economics & Culture, Media & Community Apr 2005.
 [5] Mariek Guy and Emma Tonkin, "Folksonomies : Tidying up Tags?," D-Lib Magazine. Vol 12, Jan 2006.
 [6] Flickr Open API <http://www.flickr.com/services/api>
 [7] WordNet <http://wordnet.princeton.edu/>
 [8] Andrew Krizhanovsky, "Synonym Search in Wikipedia: Synarcher" arXiv:cs.IR/0606097 v2 Jun 23 2006
 [9] Grigory Begelman and Philipp Keller and Frank Smdja, "Automated Tag Clustering: Improving search and exploration in the tag space," Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland (2006), May 22-26, 2006

[10] Melanie Aurnhammer, Peter Hanappe, Luc Steels, "Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval," Proceedings WWW 2006, Collaborative Web Tagging Workshop (2006), Edinburgh, UK, May 22-26, 2006
 [11] Satoshi Niwa and Takuo Doi and Shinichi Honiden, "Web Page Recommender System based on Folksonomy Mining for ITNG '06 Submissions," Proceedings of 3rd International Conference on Information Technology: New Generations, pp.388-393, 2006.

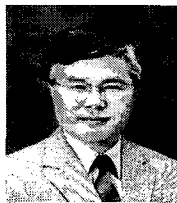
이 선 숙



1995년 :성신여자대학교 전산학과

2006년~현재 : 이화여자대학교 컴퓨터 정보통신학과 대학원 석사과정
 관심분야 : Web 2.0, 정보검색, 유비쿼터스 컴퓨팅

용 환 승



1983년 : 서울대학교 컴퓨터공학과 학사
 1985년 : 서울대학교 컴퓨터공학과 공학석사
 1994년 : 서울대학교 컴퓨터공학과 공학박사

1995년~현재 : 이화여자대학교 컴퓨터 정보통신학과 교수
 관심분야 : 객체관계데이터베이스, 데이터마이닝, 유비쿼터스 데이터베이스