

구조화된 소셜 메타데이터를 활용한 이미지 자료의 시맨틱 검색에 관한 실험적 연구*

An Experimental Study on Semantic Searches for Image Data Using Structured Social Metadata

김 현 희(Hyun-Hee Kim)**

김 용 호(Yong-Ho Kim)***

목 차

1. 서론	4.3 구조화된 폭소노미 시스템 설계
2. 이론적 배경	4.4 태그 기반 시스템
3. 통제어휘 대 폭소노미 기반 이미지 색인 기술 비교	5. 구조화된 폭소노미 시스템의 평가
4. 실험 시스템 설계	5.1 실험 설계
4.1 개요	5.2 평가 결과 분석
4.2 표본 이미지 선정과 태그 분석	6. 결론

초 록

본 연구는 이미지의 시맨틱 검색을 위해서 동등어, 동의어 및 관련어를 질의 확장시에 사용하여 태그를 통제된 구조화된 폭소노미 시스템의 모형을 제안하였다. 그런 다음 제안한 시스템의 효율성을 평가하기 위해서 실험을 통해서 이 시스템을 태그를 전혀 통제하지 않은 태그 기반 시스템과 검색 효율성(재현율과 정확률)과 만족도 측면에서 비교해 보았다. 이외에 검색 효율성이 질의 확장 방식에 따라서 어떤 차이를 보이는지 조사해 보았다. 실험 결과, 제안된 구조화된 폭소노미 시스템이 태그 기반 시스템 보다 재현율, 정확률 및 만족도에서 더 높게 나타났으며 그 차이도 통계적으로 유의미한 것으로 나타났다. 한편 질의 확장 방식에 따라서 재현율은 차이가 없었지만 정확률은 부분적으로 차이가 있는 것으로 나타났다. 본 연구 결과는 라이브러리 2.0 시대의 디지털 도서관 시스템에 적용되어 디지털 자원에 대한 접근성을 향상시킬 수 있을 것이다.

ABSTRACT

We designed a structured folksonomy system in which queries can be expanded through tag control: equivalent, synonym or related tags are bound together, in order to improve the retrieval efficiency (recall and precision) of image data. Then, we evaluated the proposed system by comparing it to a tag-based system without tag control in terms of recall, precision, and user satisfaction. Furthermore, we also investigated which query expansion method is the most efficient in terms of retrieval performance. The experimental results showed that the recall, precision, and user satisfaction rates of the proposed system are statistically higher than the rates of the tag-based system, respectively. On the other hand, there are significant differences among the precision rates of query expansion methods but there are no significant differences among their recall rates. The proposed system can be utilized as a guide on how to effectively index and retrieve the digital content of digital library systems in the Library 2.0 era.

키워드: 소셜 메타데이터, 태그, 폭소노미, 이미지, 시맨틱 검색, 질의 확장, 디지털 도서관, 통제어휘, 플리커
Recall, Precision, Structured Folksonomy System, Tag Control, Query Expansion

* 이 논문은 2009년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2009-32A-H00005).

** 명지대학교 문헌정보학과 교수(kimhh@mju.ac.kr)

*** 부경대학교 신문방송학과 교수(kimyho@pknu.ac.kr)

논문접수일자: 2010년 1월 30일 최초심사일자: 2010년 2월 3일 게재확정일자: 2010년 2월 11일
한국문헌정보학회지, 44(1): 117-135, 2010. [DOI:10.4275/KSLIS.2010.44.1.117]

1. 서론

이용자들의 요구에 따라서 디지털 도서관은 텍스트 자료 이외에 이미지, 사진, 포스터 등의 자료들을 디지털 콘텐츠로 구축하기 시작하였다. 이에 따라서 이미지 자료에 대한 색인 작업에 대한 관심도 점차 커지고 있다. 이미지 색인에 대한 접근은 크게 두 가지 즉, 개념 기반 색인과 콘텐츠 기반 색인으로 구분될 수 있다. 개념 기반 색인은 사람들에 의해서 이미지 속성과 시맨틱 내용이 언어적으로 확인되고 기술된다. 반면, 콘텐츠 기반 색인은 색, 모양 또는 텍스처와 같은 이미지 특성들이 컴퓨터 소프트웨어에 의해서 자동으로 확인되고 추출되어 생성된다(Matusiak 2006).

이상적으로는 이 두 종류의 색인은 서로 각기 다른 특성이 있기 때문에 서로 결합하여 사용하면 그 효과가 증대될 수 있음을 어렵지 않게 예측할 수 있다. 비록 성능면에서 아직 뛰어나지 않지만 이 두 방법을 결합하여 이미지를 검색하는 방식들이 연구되고 있다(Sinha & Jain 2008; Gurel 2009). 베를린의 응용과학대에서 제공하는 Pixolu가 그 중 하나이다(Pixelu 2009). 이 시스템은 개념(키워드)으로 이미지를 검색한 후 검색된 이미지 중에 하나를 선택하면 그 이미지와 비슷한 이미지를 콘텐츠 기반으로 분석하여 재검색해준다. 이러한 통합 방식은 개념 및 콘텐츠 기반의 색인 이론과 기술에 의해서 더 효율적으로 발전할 것이며 또한 대용량의 데이터베이스에 적용될 수 있을 것이다.

본 연구에서는 이러한 통합 방식의 기초 연구로 개념 기반의 이미지 색인과 검색에 초점

을 맞춰서 어떤 연구와 기술이 진행되어 오고 있는지 살펴보고, 개선 방안을 모색해 보고자 한다. 전통적으로 개념 기반 이미지 색인은 사서 또는 색인자에 의해서 수행되어 오고 있으나 최근 폭소노미의 등장으로 일반 이용자에 의해서도 색인 작업이 수행되고 있다. 색인 업무는 색인자가 이미지의 의미를 결정하고 이미지의 내용을 기술하는 색인어를 할당하는 것과 관련이 있으며, 이러한 업무는 전통적으로 분류 체계나 시소러스와 같은 통제 어휘에 의존하여 왔다. 주제 색인에 대한 이러한 접근은 이미지가 훈련된 전문가가 코딩할 수 있는 고정된 의미를 갖고 있고 그러한 이미지의 고정된 의미가 보편적으로 받아들여진다는 가정에 근거하고 있다. 그러나 현실은 이러한 가정과 많이 다르게 나타난다. 문헌의 의미는 하나의 표준화된 해석보다 그 문헌을 읽는 사람과 문헌 간의 상호작용에서 나온다고 볼 수 있다. 따라서 주제 색인자의 색인을 넘어서 일반 이용자에 의해서 주제 색인어를 할당하는 방법인 소셜 태깅을 활용하는 것도 의미가 있다고 할 수 있다(Rafferty & Hilderley 2007).

최근 이용자들은 자료를 기술할 때 '태그'라는 주제로 기술한다. 태그는 정보에 대한 이용자의 개념 모델을 반영하고 태그는 저자와 이용자의 언어를 실제적으로 나타낸다(Quintarelli 2005; Peters & Stock 2007). 태그는 특히 이미지에 유용할 것이라고 생각되는데 이미지에 포함된 객체, 사건, 활동들을 정확하게 표현하는 태그들을 할당하기 때문이다. 또한 많은 태그 어휘들은 비디오에 적용되어 추가적인 기술을 제공한다(Geisler & Burns 2007).

태그는 경직되고, 진부하고, 이용자 언어와

는 거리가 먼 것으로 평가되는 통제어휘에 비해서 이용자의 실제 언어를 반영하며 미래에 정보를 검색할 때 사용될 것으로 예측되는 용어로 정의된다. 태그는 빠르고 간단하게 만들 수 있고, 전문적 색인과 비교해서 훈련이 필요 없고 비용이나 전문 지식이 요구되지 않는 장점이 있다(Hammond et al. 2005; Rosenfeld 2005). 그러나, 태그는 종종 모호하고, 매우 개인적이고 부정확하거나, 오타, 다국어, 복합 단어로 이루어져 있어서 검색의 어려움을 갖고 있다(Guy & Tonkin 2006). 또한 태그는 동의어, 동음이의어 등에 대한 통제가 이루어지지 않고 태그간 계층적 구조가 없는 문제점이 있다.

본 연구는 폭소노미의 대중성과 편리성을 통제 어휘의 시맨틱 구조와 결합하는 방안을 모색하기 위한 기초 작업으로 온라인 기반의 디지털 콘텐츠 환경에서의 태그의 통제 효과를 조사해 보고자 한다. 구체적으로 태그의 동등어, 동의어 및 관련어를 질의 확장 시에 사용하여 태그를 통제한 후 시맨틱 검색을 한 경우와 태그를 전혀 통제하지 않은 경우를 비교하여 검색의 재현율, 정확률 및 만족도 측면에서 어떤 차이를 보이는지 실험을 통해서 증명해 보고자 한다.

2. 이론적 배경

2.1 통제어휘와 폭소노미

전통적인 검색 도구는 분류 시스템, 문헌에 대해 통합된 접근을 제공하는 통제어휘인 시소러스, 주제명표목표 등으로 이루어져 있다. 이

러한 검색 도구들은 다양한 자원에 접근점을 제공하는데 한계에 이르고 있다. 통제어휘의 또 다른 문제점으로는 정보 전문가와 이용자간의 언어의 차이를 들 수 있고, 더 나아가 통제어휘를 갱신하여 관리해야 하는 어려움이 있다는 점이다(Steele 2009).

최근에는 이용자가 직접 문헌을 주석하거나 태그로 색인하는 방법인 폭소노미를 사용하고 있다. 소셜 메타데이터(social metadata), 이용자 생성 메타데이터(user-created metadata)로 불리는 폭소노미는 이용자로 하여금 정보 객체에 그 정보를 잘 표현해 준다고 판단한 태그를 입력하게 함으로써 색인어 작성과 정보 분류에 참여하도록 한다(Matusiak 2006; 이정미 2007).

이러한 폭소노미는 비용 대 효과면에서 유리하며 특히 수집물 규모가 큰 경우에 유용하게 활용할 수 있다. 연구자들은 폭소노미를 통제어휘가 갖는 문제점을 해결할 수 있는 대안으로 제안하고 있지만 이에 앞서서 해결해야 할 문제점도 많다. 예를 들어서, 폭소노미는 어휘 통제를 하지 않기 때문에 정보 검색에서 재현율과 정확률이 저하될 가능성이 높다. 따라서 이러한 폭소노미와 통제어휘의 장점을 살려서 이 둘을 결합할 방법을 생각해 볼 수 있다. 다시 말해서, 폭소노미의 대중성, 편리성 및 유연성을 통제 어휘의 시맨틱 및 고품질 구조와 결합하는 방안을 연구할 필요가 있다. <표 1>은 이 두 방식을 개념, 입력방식, 작성자, 구조 등의 측면에서 비교한 것이다(김민경 2009). 이상적으로는 이 두 접근방식들을 지속적인 피드백 고리내에서 서로 영향을 받으면서 발전시키는 것이다.

〈표 1〉 통제어휘와 폭소노미 비교

	통제어휘(controlled vocabulary)	폭소노미(folksonomy)
개념정의	검색을 위해서 지식을 조직하는 방법을 제공하는 것	이용자 생성 텍사노미를 기술하기 위하여 'folk'와 'taxonomy'를 결합하여 만든 표현(Vander 2007)
입력방식	특정 시스템(예, 도서관자동화시스템)을 통해 통제어로 입력	주로 웹을 통해 자연어로 입력
작성자	정보 전문가	이용자
구조	계층 구조	수평적 네트워크 구조
상호작용	사 서 → 이용자	이용자 ↔ 이용자, 이용자 ↔ 사 서
실례	주제명 표목표, 시소리스	태그, 주석

2.2 선행 연구

2.2.1 폭소노미와 통제어휘와의 비교

Matusiak(2006)는 이용자 생성 메타데이터(폭소노미)의 장단점을 조사하고, 폭소노미를 메타데이터 스키마와 통제 어휘 도구와 비교하였다. 비교 결과, 폭소노미가 비록 이미지 색인의 문제점을 해결하지 못했지만 이를 활용하면 이용자 언어를 반영할 수 있고, 이용자 중심의 색인으로 갈 수 있다고 제안하였다.

Lin et al. (2006)은 Connotea의 45개의 의학 분야 학술지 논문에 사용된 태그와 PubMed에서 색인어에 사용된 MeSH의 용어를 비교했다. 비교 결과, 단지 11%의 태그만이 MeSH와 일치한 것으로 나타났다. 유사하게 Bruce(2008)도 북마크 웹사이트인 CiteULike의 태그와 ERIC에서 기술하는 통제어휘와 어느 정도 일치하는지 비교·분석해 보았다. 비교 결과, 7.6%가 일치하였는데 이는 정확한 일치만 분석한 것으로, 품사와 의미적 분석을 포함한 후속 연구를 제안하였다.

Daly와 Ballantyne(2009)은 소셜 작업의 교육과 연구를 위한 리포지토리인 Learning Exchange의 색인어(통제 어휘)를 확장하기

위해서 이용자의 언어(폭소노미)를 수용할 수 있는 가능성을 측정하는 연구를 수행하였다. 이 연구는 사람들에게 이미지에 태그를 할당하게 한 다음, 이러한 태그들은 통제어휘와 비교해 보았는데, 상당 부분 이 두 언어가 일치함을 보여주었다. 이러한 결과는 통제 어휘와 함께 태그가 색인어로 사용될 수 있는 가능성을 열어 놓은 것이다. 또한 탐색시에 태그 그룹과 폭소노미를 사용할 수 있게 하는 기능은 SCIE(Social Care Institute for Excellence's) 어휘에 익숙하지 않은 사람들에게 특히 유용하게 사용될 수 있다고 보고하였다.

Yi와 Chan(2009)은 폭소노미 태그를 사용한 딜리셔스와 통제어휘인 LCSH를 단어 매칭에 의거하여 유사도를 분석하였다. 실험 결과, LCSH와 태그간의 일치도는 대략 66%로 나타났다. 이들은 이 둘 간의 매칭에 장애가 되는 요인들 즉, 태그의 비일관성, 다양한 복합 단어 태그(예, social-software)의 사용 등에 대해서 기술하고 태그의 최적화를 위해서 같은 개념을 나타내는 다른 표현의 태그들 그리고 복합 단어 태그 패턴을 정규화시켜주는 작업이 필요하다고 기술하였다. 이들은 이러한 문제들을 해결하기 위한 태그 어간 사용 또는 여러 복합 단

어 패턴을 하나의 일관성 있는 형식으로의 전환시키기와 같은 구체적인 방안들을 제안하였다. Yoon(2009)은 플리커의 태그들을 미국의 회도서관의 시소러스(Thesaurus For Graphic Materials)와 비교하여 태그들이 시소러스를 구성하는데 유용하게 사용될 수 있는지에 대해서 조사하였다. 비교 결과, 태그들이 시소러스에 포함될 개념을 선택하고 선택된 개념사이의 의미적 관계를 명확하게 하는데 유용하게 사용될 수 있다고 기술하고 있다. 이외에 이 연구는 태그의 특성과 시소러스의 특성간의 차이에 대해서도 기술하고 있다.

Lawson(2009)은 OCLC WorldCat 분석에 의해서 정의된 31개의 주제 영역에 있는 155개의 단행본에 할당된 LCSH(통제어)를 아마존과 라이브러리썬에 있는 자료 중 동일한 표제에 할당된 태그들과 비교·분석해 보았다. 비교 결과, 이용자에 의해서 제공된 소셜 태깅은 LC 주제 표목과 일치하는 많은 객관적인 용어를 포함하고 있었고, 이러한 소셜 태깅은 개선된 정보 접근을 위해서 서지 레코드에 부여될 수 있다고 기술하고 있다. Thomas et al.(2009)은 폭소노미가 LCSH를 대신할 수 있는지 조사하였고, 폭소노미는 자원에 대한 추가적인 접근점을 제공함으로써 LCSH를 확장할 수 있다고 기술하고 있다.

국내 연구를 살펴보면, 이성숙(2008)은 국내 대학도서관의 폭소노미 태그의 형태적 특성을 분석하여, 통제어휘 개발 가이드라인과 비교해본 후 어떠한 특징이 있는지 조사하였다. 조사 결과, 시소러스 가이드라인의 여러 형태적 분석 기준에 90% 부합하는 것을 알 수 있었다. 다른 한편, 시소러스의 가이드라인에 부합되지

않는 부분도 발견되었는데 이는 태그의 단점인 모호성과 다의성과 관련이 있는 것으로 나타났다. 따라서, 폭소노미가 대학 도서관 시스템에 효과적으로 활용되기 위해서는 '태그 부여와 관련된 가이드라인' 제시와 같은 태그의 최적화 방법이 필요하다고 제안하였다. 한승희(2009)는 집단지성의 결집체인 위키피디아(Wikipedia)를 이용하여 ASIS&T 시소러스를 갱신하고, 그 결과를 평가해 봄으로써 시소러스 갱신에 위키피디아의 활용 가능성을 조사하였다. 이 연구는 실험을 통해서 위키 시소러스가 용어 포괄성의 측면에서 ASIS&T 시소러스 보다 더 우수하며, 위키피디아가 시소러스 갱신에 적합한 자원으로 활용될 수 있음을 증명하였다.

2.2.2 폭소노미의 효율성과 개선 방안

Kolbitsch(2007)는 플리커의 언어 문제를 해결하기 위해서 플리커 질의를 WordNet(워드넷)을 활용하여 확장한 WordFlickr 시스템을 제안하였다. 제안된 WordFlickr 시스템과 기존 Flickr 시스템과의 비교를 위해서 비공식적인 실험을 수행하였는데 검색 효율성에 대한 구체적인 결과는 얻지 못했다. 다만, 실험을 통해서 WordFlickr 시스템의 탐색 결과에서 가장 빈도가 높은 태그들이 의미적으로 Flickr 시스템의 탐색 결과 보다 이용자의 초기 질문과 더 유사한 것으로 나타났다고 보고하였다. Hayman(2007)은 통제어휘인 시소러스에 의해 관리되는 폭소노미 모형을 소개하였다. 이 연구는 태그의 입력창에 이용자가 태그를 입력하면 등록되어 있는 시소러스를 제공하여 입력하려는 태그와 일치하거나 관련된 것을 선택할 수 있도록 하였다. 이러한 시소러스 사용은 선

택 사항이긴 하지만 폭소노미 고유의 특성을 잃어버릴 우려가 있다고 본다. 따라서, 폭소노미의 원래 목적을 훼손시키지 않으면서 태그를 최적화시키는 방안이 요망된다.

Bar-Ilan et al.(2008)은 12개 이미지에 대해서 자유롭게 태그를 입력하는 방식과 일반 주제, 사건의 장소, 객체 유형과 같이 미리 정해진 필드에 따라서 정보를 입력하는 방식을 비교, 분석해 보았다. 비교 결과, 후자인 구조화된 정보 입력 방식은 이용자에게 가이드를 제공하는 한편 요소의 의미에 대한 다른 해석을 유발하여 태그 질을 개선시키기 보다는 오히려 저하시키는 결과를 낳았다. 이외에 자유롭게 태그를 입력하는 방식은 태그의 더 다양한 범위를 허용하였다.

Morrison(2008)은 폭소노미의 검색 효율성을 실험을 통해서 조사하였다. 폭소노미를 사용하고 있는 사회적 북마크 웹사이트들과 인터넷 디렉토리 및 검색 엔진과 비교하여 재현율과 정확률 측면에서 비교해 보았다. 비교 결과, 검색 엔진이 재현율과 정확률에서 모두 우수한 것으로 나타났고, 디렉토리가 폭소노미 보다 더 정확한 것으로 나타났다. Kern et al.(2008)은 폭소노미는 멀티미디어 정보를 조직하는데 도움을 준다고 보았다. 이들은 폭소노미를 추가적 메타데이터(예, 표제)와 어떻게 확장할 것인지, 그리고 태그 추천을 위한 확장된 폭소노미의 품질을 논의하였다.

국내 연구를 살펴보면, 최윤희(2009)는 플리커의 이미지 데이터들과 이들의 메타데이터인 태그들로부터 위치정보 태그와 공간 관계를 추출한 후 시맨틱 웹의 핵심 개념인 OWL 형태의 온톨로지로 자동 구축하였다. 구축된 시스템은

폭소노미 데이터를 온톨로지의 의미적이고 기술적인 정보로 연결시켜 줌으로써 개념적인 관계를 보완하고 보다 정확한 탐색 결과를 제공할 수 있도록 하였다. 한편 이선숙(2008)은 온톨로지 접근 방법을 활용하여 태그의 의미를 명확히 하고 태그간의 관계를 설정하여 이를 태깅에 접목시키는 시스템을 구현하였다.

이시화 등(2008)은 폭소노미가 가지고 있는 문제점인 부정확한 태그로 인한 낮은 검색 결과의 문제점을 해결하기 위해서 실험 시스템을 제안하였다. 구체적으로 이들은 연관 태그 간의 매핑을 통해서 태그 가중치 매트릭스를 생성하고 이를 기반으로 하여 연관성 높은 태그로 구성된 클러스터를 생성하기 위한 알고리즘을 제안하였다. 끝으로 이강표 등(2008)은 딜리셔스 태그들을 대상으로 하여, 태그들이 위키피디아 텍스트에서 출현하는 빈도수를 기반으로 태그간 상하위 관계를 산출하는 통계적인 모형을 제안하였다.

선행 연구들을 통해서 폭소노미의 대중성을 통제 어휘의 시맨틱과 결합하는 방안을 연구할 필요가 있음을 확인하였다. 최근 미의회도서관은 검색 효율성을 높이기 위해서 도서관이 보유한 사진 자료에 일반 이용자들이 태그를 부여하도록 하는 프로젝트를 수행하였다. 그 결과, 10개월 동안 2,518명의 플리커 이용자들에 의해서 67,176개의 태그가 부여되었고 이러한 결과가 매우 긍정적이고 유익했다고 보고하고 있다(Springer et al. 2008).

그러나 아직까지는 통제어휘와 폭소노미를 모두 색인과 검색에 사용한 경우와 그렇지 않은 경우 또는 폭소노미에 통제어휘 개념을 적용하여 태그 통제를 한 경우와 그렇지 않은 경

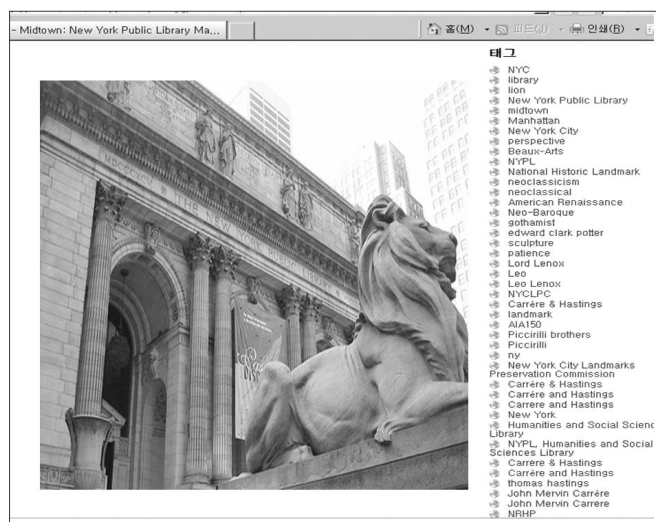
우를 비교하여 검색 효율성 측면에서 어떤 차이를 보이는지 실증적인 근거를 제시한 연구는 거의 없는 편이다. 따라서, 본 연구는 태그의 동등어, 동의어 및 관련어를 질의 확장시에 사용한 경우와 사용하지 않은 경우를 비교하여 검색 효율성 측면에서 어떤 차이를 보이는지 실험을 통해서 증명해 보고자 한다.

3. 통제어휘 대 폭소노미 기반 이미지 색인 기술 비교

본 장에서는 통제어휘 기반으로 기술되어 있는 위스콘신대학 도서관의 디지털 장서와 태그로 조직되어 있는 플리커의 자료를 비교해 보고 그 차이점을 분석해 보고자 한다. 위스콘신대학 도서관의 디지털 장서는 미국의 지리학회 도서관, 아카이브, 특수 장서 및 커리큘럼 도서

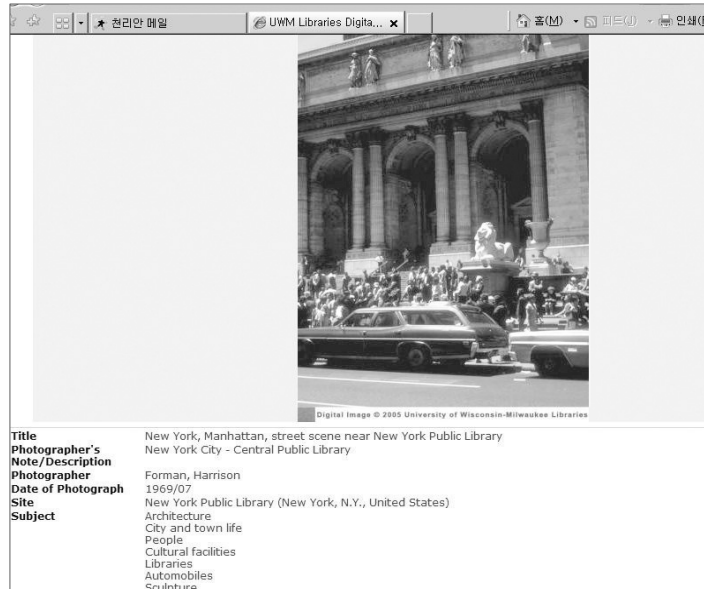
관 장서에서 41,000개 이상의 사진 이미지, 지도 및 책을 포함하고 있다. 이 디지털 장서는 CONTENTdm이라는 디지털 미디어 관리 시스템으로 구축되었고, 더블린 코아와 통제어휘 도구¹⁾를 사용한다. 한편 플리커는 2004년부터 서비스하고 있는 온라인 사진 관리 및 공유 사이트로 태그를 이용하여 사진을 관리하는 웹 2.0 서비스의 대표적인 사이트이다. 플리커는 사진 공유뿐만 아니라 사진 리포지토리로도 널리 사용되며, 한국어, 영어 등 8개 언어로 서비스되고 있으며, 현재 20억개 이상의 이미지가 데이터베이스로 구축되어 있다.

두 시스템에서 'New York Public Library'를 검색식으로 하여 검색한 결과 위스콘신대학 도서관의 디지털 장서에서는 총 23건이 검색되었고 그 중 두 건만이 적합한 자료로 보였다. <그림 1>은 그 중 한 결과물인데, 색인어가 7개가 할당되었고 다른 자료들을 살펴본 결과 색



<그림 1> 플리커 이미지

1) 그래픽 주제를 위한 LC 시소러스와 지리적 위치에 대한 지리적 명칭의 Getty 시소러스



〈그림 2〉 위스콘신 대학 이미지

언어가 모두 10개 미만이었다.

한편, 플리커의 검색 결과는 20,773건 이었다. 플리커의 태그수는 이미지에 따라서 많은 차이를 보이고 있었다. 예를 들어서 어떤 플리커 이미지는 하나의 태그만을 갖고 있는가 하면 예시한 〈그림 2〉처럼 43개의 태그를 갖고 있기도 하였다. 더 나아가 일부 이용자는 하나의 개념을 나타내는 여러 언어로 된 태그들(예, art, 예술, 藝術)을 할당하기도 하였다. 비록 이 두 그림은 비슷해 보이지만 기술 수준은 많은 차이를 보이고 있다. 위스콘신대학 도서관의 디지털 장서는 계층적 구조로 일관성 있고 자세한 기술(색인어, 자료 유형, 출처 등)을 제공하고 있다. 이에 반해 플리커 자료의 기술은 상대적으로 비계층적 구조로 일관성은 없지만 이용자의 언어를 반영하고 있다.

4. 실험 시스템 설계

4.1 개요

실험 시스템 설계에 앞서서 플리커, 유튜브, 딜리셔스 등의 폭소노미를 이용한 사이트들의 인터페이스들을 분석하였다(Shiri 2009). 분석 결과, 이러한 태그 기반 사이트들은 일반적으로 태그 그룹, 태그 목록 또는 시스템이 제안하는 태그(들)을 통해서 태그를 브라우징하는 기능을 제공하고 있었다. 특히 플리커는 태그간의 동시 출현 빈도를 이용하여 생성한 관련 태그(들)의 브라우징 기능, 태그 노트 기능 등을 갖고 있었다. 또한 딜리셔스는 이용자들이 개인 태그들을 탐색하고 필터링하는 기능도 갖고 있었다. 이러한 기능들을 참고로 하여 구조화된 폭소노미 및 태그 기반 시스템의 인터페이

스들을 설계하였다.

4.2 표본 이미지 선정과 태그 분석

표본 이미지로는 플리커 사이트의 예술과 건축 분야의 이미지들을 선정하였다. 예술과 건축 분야를 선정한 이유는 도서관에서 활용할 수 있는 학술적이면서 다양한 해석이 가능한 주제로 생각되었기 때문이다. 예술 분야 자료는 '예술' 또는 'art'를 검색어로 하여 검색된 결과 중에서 태그가 10개 이상 부여된 150개의 이미지들을 일차적으로 선정하였다. 그리고 건축 분야 자료는 '건축' 또는 'architecture'를 검색어로 하여 동일한 기준으로 150개 이미지들을 선정하였다.

선정된 300개 이미지에 부여된 총 7,890개 태그들 중 태그가 너무 짧아서 의미가 없거나 제3외국어(예, Мистецтво)로 해석이 불가능한 643개를 제외한 7,247개를 데이터베이스로 구축하였다. 7,247개의 태그 유형을 분석한 결과 이미지와 관련된 제작자, 소장 장소 그리고 해당 이미지가 속해 있는 그룹명과 같은 메타데이터(34.13%)가 차지하는 비율이 가장 높았다. 그 다음으로 장소(14.88%), 객체(12.42%), 주제(9.65%) 순으로 나타나 이미지속의 장소 및 객체 인식을 폭소노미로 보완할 수 있음을 확인할 수 있었다. 태그의 특성을 살펴본 결과, 선행 연구들이 언급한 것 처럼 태그간의 계층적 구조를 보여주지 못하고, 단복수의 구분이 없으며, 동의어는 물론 동등어 통제가 되어 있지 않다는 문제점이 있는 것으로 나타났다.

4.3 구조화된 폭소노미 시스템 설계

4.3.1 개요

폭소노미를 구조화하는 방법은 크게 두 가지로 구분할 수 있다. 첫째는 이용자가 태그 입력시 통제어휘의 동의어 또는 상하위어 개념을 적용하여 입력한 태그와 관련된 태그를 제시하는 방법을 사용할 수 있다. 예를 들어서, 통제어휘집인 시소러스나 주제명 표목표를 사용하여 검색어로 입력된 태그 'art'의 동의어인 'artwork'을 제시하거나 또는 하위어인 'painting'을 제시할 수 있다. 다음은 이용자는 위에서 언급한 통제어휘의 구조화 정보를 활용하여 질의를 확장할 수 있다. 즉, 정보 검색시, 입력된 검색어의 동의어, 관련어 등을 검색식에 포함하여 질의를 확장시키는 방법이다.

본 연구에서는 이 두 가지 방법을 활용하여 구조화된 폭소노미 시스템을 구현하였다. 첫째는 태그 입력시 알파벳순 태그 목록 파일 뿐만 아니라 15개의 카테고리별(예, 사람/신분/그룹, 객체, 사건 활동 등)로 분류한 태그 목록 파일을 참조하여 검색어를 입력할 수 있도록 하였다. 둘째는 검색어를 입력하여 검색어의 동의어와 관련어를 검색식에 첨가하여 질의를 확장할 수 있는데 이는 선택 사항으로 하였고 등등어 처리는 자동으로 처리하였다. 이 두 가지 방법에 대한 자세한 설명은 다음장에서 기술된다.

4.3.2 태그 색인 파일 구성

태그 입력시 알파벳순 태그 목록 파일 뿐만 아니라 15개의 카테고리별로 분류한 태그 목록 파일을 참조하여 검색어를 입력할 수 있도록

오름차순 태그는 페이지 아래쪽 버튼 클릭	
1.1색상 (13개)	/ Colourful / Orange / Purple / blue / color / colorful / colour / green / pink / red / sepia / white / yellow /
1.2질감/재료(8개)	/ Digital / ceramic / concrete / metal / paint / stainless steel / steel / texture /
1.3모양/구성(7개)	/ circle / composition / dome / shapes / spiral / square / structure /
1.4숫자(2개)	10 / 365 /
2.1사람/신분/그룹 (28개)	/ Antoni Gaudi / Gaudi / Lady / Louis / Man / Women / all / angel / architect / artist / baby / child / couple / dancer / family / friend / girl / group / islamic / kid / limbert / people / persian / photographer / prayer / roman / sean / woman /
2.2 사건/활동(32개)	/ Birthday / Christmas / Storm / Sunset / UP / Vivid / action / bath / capture / dance / decay / editing / festival / flying / holiday / look / olympic / party / play / see / shoot / show / smoke / sunrise / travel / traveling / trip / view / war / wave / wedding / ww2 /
2.3객체(115개)	/ Bank / Bridge / Camera / Capitol / Car / Casa / Chair / Chapeau / Earth / Empire State Building / Eye / Great Court / Heart / Hotel / House / Ice / Lamp / Lines / MASK / One Rincon / Page / Palace / Palm / Public library / Room / Rose / Shot / Stars / Table / Tate Modern / Tower / Train / Tripod / Window / animal / aqua / big ben / bird / blue sky / boat / building / cake / campus / castle / cat / cathedral / character / church / citadel / clock / columns / couch / cross / cup / doll / dress / drop / duck / escalator / europa / film / filter / fish / floor / fountain / frames / glass / gold / golden / heritage / home / icons / kingdom / kite / lace / landmark / leaves / library / lion / louvre / mausoleum / milwaukee art museum / miniature / mirror / monument / mosque / museum / national gallery / petronas / plane / plants / postcard / rainbow / restaurant / robot / rock / roof / sand / sculpture / ship / skyscraper / sofa / staircase / stairs / stairway / statue / steps / stone / sun / tea / temple / tile / tree / umbrella / wall / wallpaper /
	/ AUS / America / Amsterdam / Argentina / Australia / Barcelona / Belgium /

〈그림 3〉 태그의 카테고리별 목록

하였다. 태그를 카테고리별로 분류하기 위해서 김현희와 김민경(2009)에 의해서 제안된 이미지 태그 카테고리 모형 (5개의 카테고리 와 17개의 하위 카테고리)에서 15개 하위 카테고리를 사용하였다. 이와 같은 태그 색인 파일을 구성하기 위해서 먼저 총 7,247개 태그를 동일한 태그로 통합한 결과 3,706종의 태그로 줄어들었다. 다시 이 중에서 출현빈도가 1개인 태그를 제외한 총 820개의 태그가 15개의 카테고리별로 분류되어 카테고리별 태그 목록 파일을 〈그림 3〉과 같이 구성하였다. 〈그림 3〉에서 아래쪽 버튼을 클릭하면 알파벳순 (오름차순) 목록으로 전환될 수 있도록 구성하였다.

4.3.3 질의 확장 절차

질의 확장은 크게 선택에 의한 확장과 자동 확장으로 구분된다. 먼저 선택에 의한 확장은 입력된 태그의 최대 3개 까지 동의어나 관련어

를 포함시켜 확장하는 방법이다. 동의어는 워드넷을 이용하여 의미상의 동의어들로 확장시켜 주는 반면 관련어들은 두 단어의 동시출현 빈도에 기초하여 구성된 플리커의 데이터베이스에서 추출하였다. 그리고 단복수, 약어, 다른 언어 표현(예, art vs. 예술) 등의 동등어 처리는 자동 확장 기능으로 포함시켰다. 이 두 용어의 차이는 동의어는 대상이 되는 태그와 의미적으로 관련되어 있지만 관련어는 동시출현빈도만을 고려하여 만든 것이기 때문에 의미적으로 꼭 연관된 태그들만 포함되는 것이 아니라는 점이다.

4.3.4 인터페이스

〈그림 4〉는 구조화된 폭소노미 시스템을 구현한 인터페이스이다. 확장 검색을 원할 경우에는 태그를 입력하고 동의어와 관련어를 열람해 본 후 동의어와 관련어를 모두 선택하거

나 또는 이 중 하나만 선택하여 질의에 포함시킬 수 있다. 예를 들어서, art를 입력하고 동의어와 관련어를 선택하면 <그림 4>에서처럼 art의 동의어와 관련어 목록을 볼 수 있다. 이와 같이 동의어와 관련어를 모두 사용하여 확장 검색을 수행할 경우 art를 태그로 가진 이미지 검색 결과 외에 동등어(예술, arte), 동의어(artwork, fine art) 그리고 관련어(painting, black, color) 태그를 가진 이미지들도 함께 검색된다. 검색어는 최대 두 개 까지 사용할 수 있도록 하고, 두 개의 검색어는 불리언 연산자 'and' 또는 'or'를 사용하여 결합할 수 있도록 하였다.

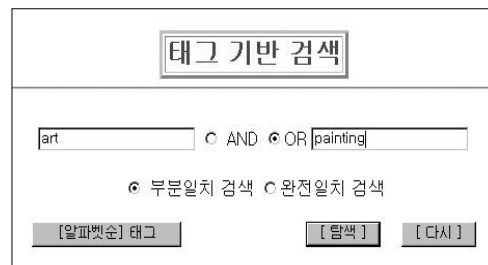


<그림 4> 구조화된 폭소노미 시스템

4.4 태그 기반 시스템

태그 기반 시스템은 태그를 색인어로 사용하지만 구조화된 폭소노미 시스템과는 달리 검색시에 자동으로 동등어 처리를 하지 않고, 또한 동의어 또는 관련어를 사용하여 질의를 확장하는 기능도 갖고 있지 않다. 이 시스템은 태그 전체를 파악할 수 있는 도구로 태그 구름 대신에 알파벳순 태그 목록을 제시하고 있다. 또한 이 시스템은 검색어의 부분 및 완전 매칭 검색 기능을 갖고 있어서 이용자들이 이 두 가지 중

한 가지 또는 두 가지 방법을 모두 사용하여 이미지를 검색할 수 있도록 하였다(그림 5 참조). 이외에 태그 기반 시스템도 구조화된 폭소노미 시스템과 검색 조건을 같게 하기 위해서 검색어는 최대 두 개 까지 사용할 수 있도록 하고, 두 개의 검색어는 불리언 연산자 'and' 또는 'or'를 사용하여 결합할 수 있도록 하였다.



<그림 5> 태그 기반 시스템

5. 구조화된 폭소노미 시스템의 평가

5.1 실험 설계

5.1.1 연구 가설과 문제

본 연구에서는 태그에 언어 통제를 한 경우와 하지 않는 경우와는 검색의 재현율과 정확률을 그리고 사용자 만족도 측면에서 어떤 차이가 있는지 알아보기 위해서 다음과 같은 세 개의 연구 가설을 이끌어 내었다.

연구 가설 1: 구조화된 폭소노미 기반 시스템과 태그 기반 시스템간의 검색 재현율에 차이가 있을 것이다.

연구 가설 2: 구조화된 폭소노미 기반 시스템과 태그 기반 시스템간의 검색 정확률에 차이가 있을 것이다.

연구 가설 3: 구조화된 폭소노미 기반 시스템과 태그 기반 시스템간의 이용자 만족도에 차이가 있을 것이다.

이외에 구조화된 폭소노미 기반 시스템의 검색시 질의 확장 방법(동어 사용, 관련어 사용, 동의어/관련어 사용 및 동의어/관련어 모두 사용 안함)이 검색 효율성(재현율과 정확률)에 어떤 영향을 미치는지에 대한 연구 문제를 조사해 보고자 한다.

5.1.2 피조사자와 설문지

제안된 시스템의 효율성을 평가하기 위해서 50명의 피조사자를 선정하여 실험을 수행하였

다. 피조사자의 동질성을 유지하기 위해서 정보 검색 과목을 수강하고 있는 50명을 남녀, 학년을 기준으로 하여 25명씩 두 개의 그룹(그룹 1과 2)으로 구분하였다. 실험에 사용된 공통 질문과 각 그룹에 사용된 기타 질문은 <표 2>와 같다.

5.1.3 평가 절차

실험 절차는 다음과 같이 네 단계로 구분하였다. 첫째, 그룹 1과 2를 두 개의 컴퓨터 실습실에 각각 배정하였다. 둘째, 각 실습실에서 실험 내용에 대한 소개를 10분 정도 하였다. 셋째, 아홉개의 질문이 수록된 질문지 유형 1을 각 실습실에 배포한 후 이용자로 하여금 시스템을

<표 2> 설문지 내용

번호	공통 질문
1	천연색이 아닌 흑백 이미지(black and white)들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.
2	오스트레일리아(Australia)의 풍경에 관한 이미지들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.
3	이미지속에 동물(animal)이 표현된 것들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.(캐릭터, 그림 등으로 동물을 표현한 이미지들도 모두 포함시키십시오)
4	미국 뉴욕시(New York)의 건축물(architecture)이 나오는 이미지들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오. 여기서 건축물은 빌딩(building), 탑(tower), 다리(bridge) 등을 모두 포함합니다.
5	커플(couple)이 손을 잡고 바라보고 있는 이미지들을 검색한 후 가장 적합하다고 생각하는 이미지를 하나 선택한 후 그 이미지 번호를 기술하십시오.
6	아이(kid) 또는 아이들(kids) 나오는 이미지들을 모두 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.
7	창문을 통해 비추이는 햇살이 아름다운 무슬림사원(muslim) 내부의 이미지들을 검색한 후 가장 적합하다고 생각하는 이미지를 하나 선택한 후 그 이미지 번호를 기술하십시오.
8	궁전(palace)이 나오는 이미지들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.
9	이미지속에 계단 또는 계단들(stair)이 나오는 이미지들을 검색한 후 적합하다고 생각하는 이미지(들)들을 선택한 후 이미지 번호(들)을 기술하십시오.
기타 질문	
그룹 1: (1)알파벳순 태그 목록과 카테고리별 태그 목록의 선호도 (2)실험 시스템의 만족도 (3)실험 시스템의 개선할 점(들)이 있다면 이를 기술해 보시오.	
그룹 2: (1)완전 일치 검색과 부분 일치 검색의 사용 빈도 (2)실험 시스템의 만족도 (3)실험 시스템의 불편함 점(들) 또는 개선할 점(들)이 있다면 이를 기술해 보시오.	

이용하여 적합한 이미지(들)을 찾은 후 이미지 번호(들)을 기술하도록 하였다. 각 질문에 대해서 7분 정도 시간을 주어서 검색 하는데 65분을 부여하였다. 넷째, 각 실습실에 기타 질문이 수록된 설문지 유형 2를 배포하고 답변하도록 하였는데 답변 시간은 15분을 부여하였다. 따라서 전체 실험 시간이 총 90분 정도가 소요되었다.

5.2 평가 결과 분석

5.2.1 데이터 분석과 연구 가설 검증

1) 구조화된 폭소노미 기반 시스템과 태그 기반 시스템간의 검색 재현율과 정확률 차이 분석

구조화된 폭소노미 기반 시스템과 태그 기반 시스템간의 검색 재현율과 정확률 차이를 검증하기 위해 t-검증을 하였다. 분석 결과, 구조화된 폭소노미 시스템이 태그 기반 시스템 보다 재현율(0.62 vs. 0.48, 등분산이 가정되지 않음)과 정확률(0.86 vs. 0.73, 등분산이 가정됨)에서 모두 높게 나타났고 유의확률수준 0.01에서

통계적으로 유의미한 차이를 보여서, 가설 1과 2가 모두 검증되었다(표 3 참조).

2) 구조화된 폭소노미 기반 시스템과 태그 기반 시스템의 만족도 차이와 개선점 분석
만족도 분석을 5점 척도(매우 만족: 5, 매우 만족하지 않음: 1)로 수행한 결과, 구조화된 폭소노미 시스템과 태그 기반 시스템의 만족도는 3.79, 3.12로 각각 나타났다. 이 두 만족도의 차이를 검증하기 위해 t-검증을 한 결과, 유의확률수준 0.01에서 통계적으로 유의미한 차이를 보여, 가설 3이 검증되었다($t_{값}=3.25, p=0.00 < 0.01$).

이외에 그룹 1의 이용자들을 대상으로 알파벳순 태그 목록과 카테고리별 태그 목록 중 어느 목록을 더 선호 하느냐는 질문에 20명(80%)의 이용자가 알파벳순 태그 목록을 더 선호한다고 답변하였다. 선호 이유는 사용의 편리함과 익숙함을 가장 많이 꼽았고, 그 다음으로 카테고리의 어려움 및 부족을 언급하였다. 3명(12%)의 이용자는 카테고리별 태그 목록을 선호하였고 선호 이유는 어떤 태그를 선택해야 할지 모를때 특히 유용하고, 관련 검색어를 찾

〈표 3〉 t-검증 (구조화된 폭소노미 vs. 태그)

기술 통계값(항목별 평균)					
재현율(구조화된 폭소노미)	재현율(태그)	정확률(구조화된 폭소노미)	정확률(태그)		
0.62	0.48	0.86	0.73		
t-검증 결과					
대응	t값	유의확률(p값)	대응	t값	유의확률(p값)
재현율(구조화된 폭소노미) vs. 재현율(태그)	4.62	0.00**	정확률(구조화된 폭소노미) vs. 정확률(태그)	4.60	0.00**
등분산 검정(재현율) ($F=0.01, p(=0.93) > 0.05$) (등분산이 가정되지 않음)					
등분산 검정(정확률) ($F=41.24, p(=0.00) < 0.05$) (등분산이 가정됨)					

는데 용이하다고 답변하였다. 2명(8%)의 이용자는 상황에 따라 두 가지를 함께 병행하여 사용하였다고 답변하였다. 또한 그룹 1의 이용자들은 시스템의 개선점으로 태그수 부족(5명), 이미지 크기 작음(4명), 데이터베이스양 부족(3명), 관련어가 의미적으로 연관되어 있지 않음(3명) 등을 언급하였다.

한편 그룹 2의 완전 일치 검색과 부분 일치 검색의 사용 빈도의 분석 결과, 부분 일치 검색(58%)이 완전 일치 검색(42%) 보다 더 많이 사용되었다. 검색시 불편한 점 또는 개선점은 검색어(접근점)판단 또는 동의어나 연관어가 잘 생각나지 않는 경우 검색의 어려움을 가장 많이 언급하였다(15명). 그 다음으로 관련 태그들이 너무 많아서 검색결과 부적합한 그림에 대한 판단 기준의 부재(13명), 한국어지원이 없다/영어라서 접근점의 어려움(6명), 고급검색의 필요/추가적인 연산자 필요(5명) 등이 많

이 언급되었다.

5.2.2 연구 문제 분석

각 질문에서 네 개의 질의 확장 방법(동의어 사용, 관련어 사용, 동의어/관련어 사용 및 동의어/관련어 모두 사용 안함)에 따라서 검색의 효율성(재현율과 정확률)이 차이가 있는지 알아보기 위해서 일원배치 분산분석을 하였다. 집단이 하나 이하의 케이스로 사후 검정이 불가능한 질문번호 3, 6, 7을 제외한 6개 질문의 분석 결과는 <표 4>와 같다. 분석 결과, 질의 확장 방법이 재현율에 미치는 영향은 없었으나, 정확률(질문 1, 5, 및 8)에 영향을 미치는 것으로 나타났다. 사후 검증 결과, 질문 1과 5인 경우에 동의어/관련어를 사용하지 않은 경우에 정확률이 가장 높게 나타났으며, 질문 8에서는 관련어만을 사용했을 때 정확률이 가장 높은 것으로 나타났다.

<표 4> 일원배치 분산분석 결과(재현율과 정확률)

질문번호	평가 기준	F값	유의확률 (p값)
1	재현율	2.14	0.12
	정확률	3.11	0.04*
2	재현율	1.19	0.33
	정확률	0.94	0.44
4	재현율	0.34	0.80
	정확률	1.51	0.24
5	재현율	2.23	0.11
	정확률	4.43	0.01**
8	재현율	2.65	0.09
	정확률	10.74	0.00**
9	재현율	0.87	0.47
	정확률	0.39	0.76

6. 결론

자료 유형이 디지털화됨에 따라서 점점 더 많은 이용자들이 OPAC을 통해서 소장하고 있는 자료 이외에 관심 있는 사진이나 포스터 등을 검색하기를 원하고 있다(Choi and Hsieh-Yee, 2008). 이와 같은 추세에 맞춰서 본 연구에서는 폭소노미를 활용하여 시맨틱 검색이 가능한 이미지 검색 시스템을 구현하여 그 효율성을 평가해 보았고, 그 결과는 다음과 같다.

첫째, 태그를 전혀 통제하지 않은 태그 기반 시스템의 검색 결과, 재현율과 정확률이 0.48과 0.73으로 각각 나타나 일단 색인어로서의 기능은 있는 것으로 나타났다. 그러나 재현율이 상대적으로 낮게 나타나 재현율을 좀 더 향상시킬 필요가 있어 보였다. 이를 위해서 단복수, 약어, 다른 언어 표현 등의 동등어 통제와 함께 동의어 및 관련어를 그룹핑하여 질의를 확장할 수 있는 구조화된 폭소노미 시스템을 구현하여 재현율은 0.48에서 0.62로 향상되었고, 정확률도 0.73에서 0.86으로 향상되었다. 이러한 결과로 인하여 가설 1과 2가 검증되었다.

둘째, 질의 확장시 태그의 동시출현빈도에 의해서만 추출한 관련어는 의미적으로 연관이 없는 태그들도 다수 포함되어 있었다. 다시 말해서, 이와 같은 관련어 사용은 플리커와 같은 대용량 데이터베이스에서는 효과적일 수 있으나 소규모의 데이터베이스에서는 신중히 사용해야 할 것으로 보인다. 또 다른 방법으로 관련어 선정 기준을 태그들의 동시출현빈도에 기초 하면서 의미적으로도 연관되어 있는 태그로 제

한하는 것도 고려해 볼 수 있다.

셋째, 구조화된 폭소노미 기반 시스템과 태그 기반 시스템에 대한 이용자 만족도는 3.79, 3.12로 각각 나타났고, 이 두 만족도의 차이는 통계적으로 유의미한 것으로 나타나 가설 3은 검증되었다. 제안된 시스템을 사용한 그룹 1에서 언급한 시스템 개선점은 태그수 부족(5명), 이미지 크기 작음(4명), 데이터베이스양 부족(3명), 관련어가 의미적으로 연관되어 있지 않음(3명) 등을 지적하였다. 한편 태그 기반 시스템을 사용한 그룹 2에서 기술한 시스템 사용시 어려운 점은 검색어(접근점) 판단 또는 동의어나 관련어가 잘 생각나지 않는 경우 검색의 어려움(15명), 관련 태그들이 너무 많아서 검색 결과 부적합한 그림에 대한 판단 기준의 부족(13명)을 가장 많이 언급하였다.

넷째, 질의 확장 방법이 재현율에 어떤 영향을 미칠 것인지에 대한 연구 문제는 명확하게 증명되지 못했지만 정확률 측면에서는 부분적으로 증명되었다. 즉, 두 개의 질문은 동의어/관련어를 사용하지 않은 경우에 정확률이 가장 높게 나타났고, 한 개의 질문에서는 관련어만을 사용했을 때 정확률이 가장 높은 것으로 나타나 일관성 있는 패턴을 보이지는 못했다.

끝으로, 본 연구 결과는 서론에서 언급한 개념 기반 색인과 콘텐츠 기반 색인을 결합한 통합 이미지 색인 기법 연구에 대한 기초 자료로 활용될 수 있을 것이다. 또한 이 연구 결과는 라이브러리 2.0 시대의 디지털 도서관 시스템에 적용되어 디지털 자원에 대한 접근성을 향상시킬 수 있을 것이다.

참 고 문 헌

- [1] 김민경. 2010. 『폭소노미를 활용한 이미지 검색의 효율성 개선 방안에 관한 연구』. 석사학위 논문, 명지대학교 대학원, 문헌정보학과.
- [2] 김현희, 김민경. 2009. 플리커 이미지 자료에 대한 이용자 태깅 행태 분석과 활용 방안. 『정보관리연구』, 40(2): 71-94.
- [3] 이강표 외. 2008. FolksoViz: Wikipedia 본문을 이용한 상하위 관계 기반 폭소노미 시각화 기법. 『정보과학회논문지: 컴퓨팅의 실제 및 레터』, 14(4): 401-411.
- [4] 이선숙. 2008. 『OntoSonomy: 온톨로지 기반의 폭소노미 확장법』. 석사학위 논문, 이화여자대학교 대학원, 컴퓨터정보통신공학과.
- [5] 이성숙. 2008. 대학도서관 폭소노미 태그의 형태적 특성에 관한 연구. 『한국문헌정보학회지』, 42(4): 463-480.
- [6] 이시화 외. 2008. Web2.0 환경에서의 효율적인 이미지 검색을 위한 태그 클러스터링 시스템의 설계 및 구현. 『멀티미디어학회논문지』, 11(8): 1169-1178.
- [7] 이정미. 2007. 폭소노미의 개념적 접근과 웹정보 서비스에의 적용. 『한국비블리아 학회지』, 18(2): 141-159.
- [8] 최윤희. 2009. 『폭소노미에서 위치태그 분석을 통한 공간관계 추출 기법』. 석사학위논문, 이화여자대학교 대학원, 컴퓨터정보통신공학과.
- [10] 한승희. 2009. 집단지성을 활용한 시소러스 갱신에 관한 연구. 『정보관리학회지』, 26(3): 25-43.
- [11] Bar-Ilan, J., et al. 2008. "Structured versus unstructured tagging: a case study." *Online Information Review*, 32(5): 635-647.
- [12] Bruce, R. 2008. "Descriptor and folksonomy concurrence in education related scholarly research." *Webology*, 5(3). [online]. <<http://www.webology.ir/>>.
- [13] Choi, Y., & Hsieh-Yee, I. 2008. "Subject access for images in an OPAC." *Proceedings of the ASIST Annual Meeting*. Columbus, Ohio.
- [14] Daly, E., & Ballantyne, N. 2009. "Ensuring the discoverability of digital images for social work education: an online "tagging" survey to test controlled vocabularies." *Webology*, 6(2). [online]. <<http://www.webology.ir/>>.
- [15] Gurel, M. 2009. *Expert assisted exploration of photographs: Supporting users in exploring visual media through subjective aesthetic attributes and crowd-sourced tags*. M.A. thesis, University of Dublin.
- [16] Guy, M., & Tonkin, E. 2006. "Tidying up tags?" *D-Lib Magazine*, 12(1). [online]. <<http://www.dlib.org/dlib/january06/guy/01guy.html>>.

- [17] Hammond, T., Hannay, T., Lund, B., & Scott, J. 2005. "Social bookmarking tools (I)." *D-Lib Magazine*, 11(4). [online]. <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>.
- [18] Hayman, S. 2007. "Folksonomies and tagging: New developments in social bookmarking." *Proceedings of the Ark Group Conference: Developing and Improving Classification Schemes*. [online]. <<http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/arkhayman.pdf>>.
- [19] Kern, R., Granitzer, M., & Pammer, V. 2008. "Extending folksonomies for image tagging." *Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, 7-9 May 2008. 126-129.
- [20] Kolbitsch, J. 2007. "WordFlickr: A solution to the vocabulary problem in social tagging systems." *Proceedings of I-MEDIA '07 and I-SEMANTICS '07*, Sept 2007. Graz, Austria.
- [21] Lawson, K. G. 2009. "Mining social tagging data for enhanced subject access for readers and researchers." *The Journal of Academic Librarianship*, 35(6): 574-582.
- [22] Lin, X., et al. 2006. "Exploring characteristics of social classification." *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop*, Austin, Texas (US).
- [23] Matusiak, Krystyna K. 2006. "Towards user-centered indexing in digital image collections." *OCLC Systems & Services: International Digital Library*, 22(4): 283-298.
- [24] Morrison, J. 2008. "Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web." *Information Processing and Management*, 44: 1562-1579.
- [25] Peters, I., & Stock, W. G. 2007. "Folksonomy and information retrieval." *Proceedings of the ASIST Annual Meeting*, October 19-24, 2007. Milwaukee, Wisconsin: America Society for Information Science Technology.
- [26] Pixolu. 2009. [online]. <<http://www.pixolu.de/>>.
- [27] Quintarelli, E. 2005. "Folksonomies: Power to the people." *ISKO Italy UniMIB meeting*, June 24, 2005. Milan: ISKO Italia.
- [28] Rafferty, P., & Hilderley, R. 2007. "Flickr and Democratic Indexing: dialogic approaches to indexing." *Aslib Proceedings: New Information Perspectives*, 59(4/5): 397-410.
- [29] Rosenfeld, L. 2005. "Folksonomies? how about metadata ecologies?" [online]. <http://www.louisrosenfeld.com/home/bloug_archive/000330.html>.
- [30] Shiri, A. 2009. "An examination of social tagging interface features and functionalities An analytical comparison." *Online Information Review*, 33(5): 901-919.
- [31] Sinha, P., & Jain, R. 2008. "Classification and annotation of digital photos using optical context

- data.” In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, 309-318, New York, NY, USA: ACM.
- [32] Springer, M., et al. 2008. *For the common good: the Library of Congress Flickr pilot project*. Washington, DC : Library of Congress.
- [33] Steele, T. 2009. “The new cooperative cataloging.” *Library Hi Tech*, 27(1): 68-77.
- [34] Thomas, M., et al. 2009. “To tag or not to tag?” *Library Hi Tech*, 27(3): 411-434.
- [35] Trant, J. 2003. “Image retrieval benchmark database service: a needs assessment and preliminary development plan.” [online].
<<http://www.clir.org/pubs/reports/trant04/tranttext.htm>>.
- [36] Vander Wal, T. 2007. “Folksonomy coinage and definition.” [online].
<<http://vanderwal.net/folksonomy.html>>.
- [37] Yi, K., & Chan, L. M. 2009. “Linking folksonomy to Library of Congress subject headings: an exploratory study.” *Journal of Documentation*, 65(6): 872-900.
- [38] Yoon, J. 2009. “Towards a user-oriented thesaurus for non-domain-specific image collections.” *Information Processing and Management*, 45: 452-468.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Min-Kyung Kim. 2010. *A study on Improving the Search Efficiency of Image Data Using Folksonomies*. Graduate Thesis, Myongji University.
- [2] Hyun-Hee Kim, & Min-Kyung Kim. 2009. “Investigating the End-User Tagging Behavior and its Implications in Flickr.” *Journal of Information Management*, 40(2): 71-94.
- [3] Kangpyo Lee, et al. 2008. “FolksoViz: A Subsumption-based Folksonomy Visualization Using the Wikipedia.” *Journal of KIISE: Computing Practices and Letters*, 14(4): 401-411.
- [4] Sun-Sook Lee. 2008. *OntoSonomy: Ontology-based Extension of Folksonomy*. Graduate Thesis, Ehwa Womans University.
- [5] Sung-Sook Lee. 2008. “A Study on Form of Folksonomy Tags in University Libraries.” *Journal of the Korean Society for Library and Information Science*, 42(4): 463-480.
- [6] Si-Hwa Lee, et al. 2008. “Design and Implementation of Tag Clustering System for Efficient Image Retrieval in Web2.0 Environment.” *Journal of Korea Multimedia Society v.*, 11(8): 1169-1178.
- [7] Jeong-Mee Lee. 2007. “A Conceptual Access to the Folksonomy and Its Application on the

Web Information Services.” *Journal of the Korean Biblia Society for Library and Information Science*, 18(2): 141-159.

- [9] Yun-Hee Choi. 2009. *Extraction method of spatial relation by analyzing location tag in folksonomy*. Graduate Thesis, Ehwa Womans University.
- [10] Seung-Hee Han. 2009. “Thesaurus Updating Using Collective Intelligence: Based on Wikipedia Encyclopedia.” *Journal of the Korean Society for Information Management*, 26(3): 25-43.