

국내 포털 이용자들의 멀티미디어 검색 행태 분석*

The Multimedia Searching Behavior of Korean Portal Users

박 소 연(Soyeon Park)**

목 차

- | | |
|-----------|-----------------|
| 1. 서론 | 4. 연구 결과 |
| 2. 선행 연구 | 4.1 멀티미디어 질의 분석 |
| 3. 연구 방법 | 4.2 클릭 로그 분석 |
| 3.1 자료 수집 | 5. 결론 |
| 3.2 분석 방법 | |

초 록

멀티미디어 검색은 전통적인 검색과 웹 검색을 차별화시키는 중요한 요소 중 하나이다. 본 연구에서는 국내 주요 검색 포털인 네이버 이용자들의 멀티미디어 검색 행태를 분석, 평가하였다. 이를 위하여 이용자들이 입력한 통합 검색 질의들로 구성된 질의 로그와 질의에 대한 검색 결과에서 이용자들이 조회한 문서를 기록한 클릭 로그를 분석하였다. 연구 결과, 이용자가 입력하는 멀티미디어 질의 중 음악 질의가 차지하는 비중이 가장 높았으며, 이미지 질의와 동영상 질의의 비중은 비슷한 것으로 나타났다. 한편 멀티미디어 관련 컬렉션들 중 이용자가 가장 많이 조회하는 컬렉션은 동영상 컬렉션이었고, 이어서 이미지 컬렉션, 음악 컬렉션 순으로 나타났다. 이용자가 입력하는 멀티미디어 질의나 멀티미디어 컬렉션에서 클릭하는 문서 모두 엔터테인먼트와 관련된 주제가 가장 많았다. 본 연구의 결과는 향후 포털의 멀티미디어 콘텐츠 구축 및 검색 알고리즘 개발에 활용될 수 있을 것으로 기대된다.

ABSTRACT

The main difference between web searching and traditional searching is that the web provides and supports multimedia searching. This study aims to investigate the multimedia searching behavior of users of NAVER, a major Korean search portal. In conducting this study, the query logs and click logs of a unified search service were analyzed. The results of this study show that among the multimedia queries submitted by users, audio searches are the dominant media type, followed similarly by video and image searches. On the other hand, among the multimedia documents clicked on, video is the most popular collection type followed by image and audio collections. Entertainment is the most popular topic in both multimedia queries and clicks. The results of this study can be implemented for the portal's development of multimedia content and searching algorithms.

키워드: 멀티미디어 검색 행태, 검색 포털, 로그 분석

Multimedia Searching Behavior, Search Portals, Log Analysis

* 본 연구는 NHN(주)의 지원을 받았음.

** 덕성여자대학교 문헌정보학과 부교수(sypark@duksung.ac.kr)

논문접수일자: 2010년 1월 30일 최초심사일자: 2010년 2월 3일 게재확정일자: 2010년 2월 10일

한국문헌정보학회지, 44(1): 101-115, 2010. [DOI:10.4275/KSLIS.2010.44.1.101]

1. 서론

멀티미디어 검색은 전통적인 검색과 웹 검색을 차별화시키는 중요한 요소 중 하나이다. 많은 웹 이용자들이 멀티미디어 자료를 찾기 위해 검색 포털을 이용하고, 검색 포털들은 멀티미디어 관련 콘텐츠를 확충하고, 서비스를 다양화하고 있는 현실이다. 한편 웹 검색 분야에서 다양한 연구가 수행되어 왔지만, 국내 선행 연구들 중 웹 이용자들의 멀티미디어 검색 행태를 조사한 연구는 드문 실정이다.

이에 본 연구에서는 국내 주요 검색 포털인 네이버 이용자들의 멀티미디어 검색행태를 분석, 평가하고자 한다. 이를 위하여 이용자들이 입력한 질의들로 구성된 질의 로그와 이용자들이 조회한 문서들로 구성된 대규모 클릭 로그를 활용하고자 한다. 대부분의 국내 검색 포털들은 지식 검색, 블로그, 카페, 뉴스, 디렉토리, 쇼핑, 전문 자료, 이미지, 음악, 동영상 등 다양한 컬렉션에서 정보를 제공하고 있다. 이용자들은 통합 검색을 통해 다양한 컬렉션을 동시에 검색할 수 있고, 특정한 컬렉션을 선택하여 별도의 검색을 수행할 수도 있다. 대부분의 이용자들은 컬렉션 별로 별도의 검색을 수행하기 보다는 통합 검색을 수행하는 경우가 많다. 즉 포털의 검색 서비스 중 가장 이용도가 높고 대중적인 서비스는 통합 검색 서비스라고 할 수 있다(이준호, 박소연, 권혁성 2003). 따라서 본 연구에서는 포털 이용자들의 멀티미디어 검색 행태를 분석하기 위해서 이용자들이 가장 많이 사용하는 통합 검색의 로그를 활용하고자 한다.

좀 더 구체적으로 본 연구에서는 첫째, 멀티미디어 검색어 기준을 설정하고, 이에 따라 멀

티미디어성 질의를 파악한 후 이들의 특징 및 분포를 조사하고자 한다. 둘째, 멀티미디어 자료를 포함하고 있는 이미지, 음악, 동영상 컬렉션에서 발생한 클릭 행태의 분포와 특징을 분석하고 한다. 본 연구의 결과는 웹 이용자들의 멀티미디어 이용 행태 및 정보 요구에 대한 이해를 심화시킬 것으로 기대된다. 또한 본 연구의 결과는 향후 포털의 검색 서비스의 개선에 활용될 수 있을 것으로 기대된다. 즉 본 연구의 결과는 포털 업체들의 효과적인 멀티미디어 콘텐츠 구축 및 검색 알고리즘 개발에 기여할 것으로 기대된다.

2. 선행 연구

국외 웹 이용자들의 멀티미디어 검색 행태를 조사한 선행 연구들로는 Jansen과 Spink 등의 연구를 들 수 있다. Jansen과 Spink 등은 1990년대 후반부터 트랜잭션 로그 분석을 통하여 웹 이용자들의 검색 행태를 조사하는 일련의 연구를 수행하여 왔다. 이들은 이러한 연구의 일환으로 웹 검색 엔진 이용자들의 멀티미디어 이용 행태를 분석해 왔다. 즉 Jansen, Goodrum, Spink(2000)는 1997년 9월 16일 하루 동안 익사이트 검색 엔진에서 생성된 211,058개의 세션과 1,025,908개의 질의를 대상으로 멀티미디어 질의를 추출하기 위한 검색어 기준을 설정하고, 이에 근거하여 이미지, 음악, 동영상 질의를 추출한 후, 이러한 질의들의 분포 및 특징을 분석하였다. Ozmutlu, Spink, Ozmutlu(2003)는 2001년도 익사이트 로그를 대상으로 유사한 분석을 수행하고, 1997년 데이터와 2001년도

데이터의 특징을 비교하고 추이를 분석하였다. 또한 Spink와 Jansen(2006)은 2005년 5월 6일 메타검색엔진인 도그파일에서 생성된 트랜잭션 로그의 일부를 사용하여 일반 웹 질의뿐만 아니라 이미지, 음악, 동영상 질의의 특징을 분석하였다. Tjondronegoro, Spink, Jansen(2009)은 2006년 5월 15일 하루 동안 도그파일에서 생성된 트랜잭션 로그 중 일부를 사용하여 도그파일 이용자들의 멀티미디어 검색 행태의 특징을 분석하였다. 또한 이러한 결과를 1997년과 2001년 익사이트 데이터로부터 도출된 결과와 비교하였다. 한편 멀티미디어의 세부 유형들 중에는 이미지 검색과 관련된 연구가 상대적으로 더 많이 수행되어 왔다(Jansen 2008; Chung and Yoon 2009).

국내 웹 검색에 관한 연구는 전산학, 경영학, 문헌정보학, 신문방송학, 심리학 등 다양한 분야에서 수행되어 왔다. 이들 중 로그 분석을 통해 웹 이용자의 검색 행태를 분석한 연구들의 경우(박소연, 이준호 2002; 이준호, 박소연, 권혁성 2003), 멀티미디어 검색보다는 일반적인 웹 검색에 초점을 맞추었다. 한편 국내 멀티미디어 검색에 관한 최근 연구는 일반적인 웹 환경보다는 주로 특정한 검색 시스템이나 검색 환경을 대상으로 수행되어 왔다. 즉 멀티미디어 검색 기법이나 검색 시스템의 개발 및 구현, 실험을 통한 검색 기법의 성능 평가 등에 집중되어 왔다(김성희 2004; 김용, 문성빈 2007; 김용, 소민호 2009; 노승민, 황인준 2003; 박창섭 2007; 유소영, 문성빈 2004; 백우진 외 2008).

따라서 국내 선행 연구들 중에서 로그 분석을 통해 일반적인 웹 이용자들의 멀티미디어 검색 행태를 분석한 연구는 찾아보기 어려운

실정이다. 또한 국내외 선행 연구들 중에서 실제 이용자가 입력한 질의뿐만 아니라 클릭한 문서들까지 분석한 연구는 드문 실정이다. 이에 본 연구에서는 국내 주요 검색 포털인 네이버를 대상으로 질의 로그와 클릭 로그에 근거하여 이용자들의 멀티미디어 검색 행태를 분석하고자 한다.

3. 연구 방법

3.1 자료 수집

본 연구에서는 국내 주요 검색 포털인 네이버 이용자들의 멀티미디어 검색 행태를 조사, 분석하였다. 네이버를 선택한 이유는 국내외 검색 포털 분야에서의 네이버의 위상과 대중성 때문이다. 네이버는 2000년대 초반 이후 국내 검색 포털들 중 시장 점유율 조사, 방문자 수 조사 등에 있어서 지속적으로 1위를 차지하고 있다. 즉, 네이버는 웹 사이트 평가 및 트래픽 분석업체인 인터넷 메트릭스(<http://www.metrixcorp.com>)와 Korean Click(<http://www.koreanclick.com>) 등의 방문자 수 조사에서 1위를 차지하여 왔다. 좀 더 구체적으로 Korean Click의 조사에 따르면, 네이버의 국내 검색 시장 점유율이 2009년도 1년 동안 64%에서 70%를 유지하고 있는 것으로 나타났다(한민옥 2010). 또한 미국 시장조사업체 Comscore(<http://www.comscore.com>)의 2009년 7월 세계 검색 엔진 순위에 따르면, 네이버는 구글, 야후, 바이두, MSN에 이어 3년 연속 5위를 차지하는 등(최광 2009), 국제적으로도 인지도가 높은 상

황이다. 따라서 네이버에 입력된 질의들이 국내 웹 이용자들이 입력한 질의들에 대한 대표성을 지니고 있다고 판단되었다.

국내 웹 이용자들의 멀티미디어 검색 및 이용 행태를 조사하기 위하여 본 연구에서는 네이버 통합 검색 질의 로그와 클릭 로그를 분석하였다. 일반적으로 웹 검색 서비스의 로그는 이용자들이 입력한 질의를 기록한 질의로그와 질의에 대한 검색 결과에서 이용자가 조회한 문서를 기록한 클릭 로그로 구성된다. 개별 질의에 대한 클릭 로그는 이용자들이 조회한 문서로 구성되며, 질의별로 클릭 횟수가 100만 회에 달하는 경우도 있으므로, 대다수 이용자들의 정보 요구가 집대성된 것으로 간주될 수 있다. 본 연구에서는 2007년 5월 17일과 2007년 8월 10일 이용자들이 네이버 통합검색 창에 입력한 전체 질의 중에서 각각 700개씩 총 1,400개의 질의를 무작위로 선정하여 분석하였다. 질의 선정 이전에 로그 정제 작업을 통하여 로봇이나 프로그램이 자동으로 입력하는 질의, 편법을 이용하여 클릭 횟수를 조작하는 클릭 어뷰즈(click abuse), 클린 검색에 의해 접속이 차단된 성인성 질의 등을 제외하였다. 한편 하루에 입력되는 통합 검색 질의들 중 인기 질의들은 여러 번 반복하여 출현한다는 특징이 있다. 즉 “네이버”, “다음”과 같은 질의나 인기 연예인과 관련된 질의들의 경우 수십만 회에서 수백만 회까지 입력되기도 한다. 이러한 모집단의 특징을 부분적으로 반영하기 위하여 질의 추출 시 질의가 중복되어 포함되는 것을 허용하였다.

2007년 5월 17일과 8월 10일의 날짜를 선택한 이유는 이 날짜들이 주중의 평일을 대표할

수 있는 다른 요일의 날짜이기 때문이다. 또한 이용자들의 구체적인 검색 행태가 기록된 가장 최신 로그가 포털들의 대외비로 간주되어 확보되기 어려운 현실에서, 연구를 수행하던 시점에서 비교적 최신의 질의들을 구할 수 있는 날이기 때문이다. 실제로 검색 포털의 로그를 분석한 국외 선행 연구들의 경우 현실적인 제약 때문에 7년 이상 경과된 로그를 분석하는 경우도 존재한다(Jansen 2008). 약 3개월의 간격을 두고 질의를 추출한 것은 이 기간 동안 이용자들의 멀티미디어 이용 행태의 변화나 추이를 분석하기 위해서이다. 한 날짜에서 700개의 질의를 선택한 이유는 하루에 네이버에 입력되는 통합 검색 질의의 수를 고려할 때, 표본 오차 95% 신뢰수준 $\pm 4\%$ 와 $\pm 5\%$ 를 허용할 경우 필요한 표본의 크기가 각각 600개와 384개로 통계학 문헌에서 제시되고 있기 때문이다(Arkin and Colton 1963).

본 연구에서는 이렇게 선택된 질의들에 대하여 이용자가 조회한 문서들로 구성된 클릭 로그도 분석하였다. 일반적으로 네이버를 비롯한 검색 포털들의 클릭 로그에는 개별 질의에 대해 이용자가 조회한 문서의 URL, 조회한 문서의 클릭 빈도, 문서가 소속된 컬렉션에 대한 정보 등이 기록된다. 좀 더 구체적으로 본 연구에서 활용된 클릭 로그를 살펴보면, 2007년 5월 17일에는 700개의 질의에 대해 총 23,297개의 웹 문서가 약 700만회 조회되었으며, 2007년 8월 10일에는 700개의 질의에 대해 총 31,487개의 웹 문서가 약 832만회 조회되었다. 본 연구에서 활용한 클릭 로그에는 이용자들의 이러한 클릭 행태와 관련된 세부 정보가 저장되어 있다.

3.2 분석 방법

이용자들의 멀티미디어 정보 요구를 조사하기 위하여, 이용자들이 입력한 질의들 중 멀티미디어성 질의를 파악하는 작업이 필요하다. 본 연구에서는 멀티미디어 질의를 파악하기 위하여 국외 선행 연구에서 사용된 기준을 (Jansen, Goodrum, Spink 2000; Ozmutlu, Spink, Ozmutlu 2003) 참고하고, 본 연구에서 사용된 네이버 질의로그와 클릭로그를 상세히 분석하였다. Ozmutlu, Spink, Ozmutlu(2003)는 특정 멀티미디어 검색어 포함 여부를 멀티미디어 질의 기준으로 활용하였다. 즉 오디오 질의는 “audio,” “cd,” “concerts”와 같은 27개의 오디오 관련 검색어, 이미지 질의는 “image,” “picture” 등의 30개의 이미지 관련 검색어, 비디오 질의는 “video,” “animated”와 같은 13개

의 비디오 관련 검색어를 포함하는 질의로 정의하였다. 한편 이들의 멀티미디어 검색어 기준은 90년대 후반과 2000년대 초반 익사이트로그를 대상으로 하였으므로, 본 연구에서는 이들의 기준을 업데이트하고, 한국적 상황을 반영하여 <표 1>과 같이 멀티미디어 검색어 기준을 개발하였다. 즉 본 연구에서는 이미지, 음악, 동영상 질의를 <표 1>에서 제시된 검색어를 포함하는 질의로 정의하였다. 멀티미디어 검색어 선정 시 모호한 경우에는 클릭 로그를 참고하여, 멀티미디어 자료에 대한 정보 요구가 존재하는지를 확인하였다.

클릭 로그 분석을 위하여, 특정 질의에 대하여 이용자가 조회한 문서들을 컬렉션별로 수합하였다. 즉 질의별로, 멀티미디어 컬렉션인, 이미지, 음악, 동영상 컬렉션에서 조회된 문서들을 수합하여 그 분포 및 특징을 분석하였다.

<표 1> 멀티미디어 검색어 기준¹⁾

이미지	음악	동영상
“그림이름”	“곡제목”	“특정프로그램의 ~특집”
그림	“특정가수의 곡제목”	“특정프로그램의 ~회”
동양화	~곡(예: 즉흥환상곡)	“특정프로그램의 ~회”
만화	~송(예: 당근송)	~신 또는 썸(예: 전투썸)
메인사진	~요(예: 노동요)	다시보기
메인화면	~집(예: 보아1집)	다운
명화	가요	다운로드
바탕화면	노래	동영상
배경화면	다운	라이브
비트맵	다운로드	뮤비
사진	동요	뮤직비디오
서양화	듀엣곡	미리보기
수채화	라이브	비디오
아트	뮤직	비디오클럽

1) 이 기준에서 “ ”은 특정 그림 이름이나 곡제목과 같은 고유 명사를 나타냄.

이미지	음악	동영상
엠블럼	민요	애니메이션
웹툰/카툰	배경음악	영상
유화	신곡	예고편
이미지	앨범	티저
자화상	엔딩	ucc
정물화	엠펙쓰리	
지도	연주	
초상화	오프닝	
캐리커처	음악	
캐릭터	전곡듣기	
포스터	찬양	
포토	최신곡	
풍경	캐롤	
photo	컬러링	
picture	클래식	
image	트랙	
	팝송	
	music	
	ost	
	song	
파일형식(bitmap, bmp, dcm, eps, gif, jpeg, jpg, pbm, pct, pcx, pdd, pgm, pict, png, psb, psd, pxr, raw, tga, tif, tiff)	파일형식(cda, fla, mid, MIDI, mp3, mr, ogg, sds, wav, wma)	파일형식(asf, avi, flv, k3g, mjpeg, mkv, mov, mov8, mjpeg, mkv, mov, mp4, mpeg, mpg, skm, svi, vod, wmv)

4. 연구 결과

4.1 멀티미디어 질의 분석

〈표 1〉에서 제시된 멀티미디어 검색어 기준에 따라 네이버의 멀티미디어 질의를 분석한 결과는 〈표 2〉와 같다. 〈표 2〉에 따르면, 2007년 5월 17일과 8월 10일 전체 질의 중 멀티미디어 질의가 차지하는 비중은 각각 6.6%와 6%로 나타났다. 또한 양일 모두 멀티미디어 질의 중 음악 질의가 차지하는 비중이 가장 높았다. 2007년 5월 17일 질의의 경우, 전체 질의 중 음악 질의는 26개, 이미지 질의는 12개, 동영상

질의는 8개이었으며, 8월 10일의 경우 음악 질의는 25개, 동영상 질의는 11개, 이미지 질의는 6개이었다. 3개월 동안 전반적인 멀티미디어 질의의 분포나 세부 유형의 분포에 있어서 큰 변화 없이 유사한 것으로 나타났다. 한편 전체 700개 질의들 중 중복되는 질의를 제거한 후, 멀티미디어 질의의 기술통계를 분석한 결과는 〈표 3〉과 같다. 멀티미디어 질의의 비중이 약간 증가한 것을 제외하면, 〈표 3〉의 결과는 〈표 2〉의 결과와 유사하다고 할 수 있다.

한편 이러한 결과는 국외 선행 연구의 결과와 상당한 차이를 보고 있다. Ozmutlu, Spink, Ozmutlu(2003)의 연구에 따르면, 익사이트 멀

〈표 2〉 멀티미디어 질의의 기술통계(중복 질의 포함 시)

	이미지 질의	음악 질의	동영상 질의	총계	전체 질의수
2007년 5월 17일	12(1.7%)	26(3.7%)	8(1.1%)	46(6.6%)	700
2007년 8월 10일	6(0.9%)	25(3.6%)	11(1.6%)	42(6%)	700

〈표 3〉 멀티미디어 질의의 기술통계(중복 질의 제거 시)

	이미지 질의	음악 질의	동영상 질의	총계	전체 질의수
2007년 5월 17일	12(1.9%)	26(4.0%)	8(1.2%)	46(7.2%)	642
2007년 8월 10일	6(0.9%)	25(3.8%)	11(1.7%)	42(6.3%)	662

티미디어 질의의 비중은 1997년, 1999년, 2001년의 로그 자료에서 각각 전체 질의의 3.7%, 3.7%, 1.7%로 나타났다. 2006년 5월 15일의 도그파일의 로그를 분석한 Tjondronegoro, Spink, Jansen(2009)의 연구 결과에 따르면, 네이버와는 달리 전체 멀티미디어 질의들 중 이미지 질의의 비중이 가장 높았고, 이어서 음악, 동영상 질의의 순으로 나타났다. 또한 멀티미디어 질의의 주제에 있어서 엔터테인먼트의 비중이 높았지만, 의학이나 스포츠, 기술 등의 주제도 이전보다 증가하는 추세라고 보고되었다.

〈표 4〉와 〈표 5〉는 〈표 1〉의 기준에 따라 분류된 멀티미디어 질의 리스트를 날짜별로 보여 준다. 한편 박소연, 이준호, 김지승(2005)은 웹 질의의 주제를 분석하기 위하여 건강, 게임, 과학, 교육/학문, 금융/경제, 기관, 기업, 뉴스/미디어, 라이프스타일, 문화/예술, 사회, 성인, 쇼핑, 엔터테인먼트, 지역/여행, 컴퓨터/인터넷 등 16개의 주제 범주로 구성된 분류 체계를 제안하였다. 엔터테인먼트의 하위 범주에는 드라마, 방송, 애니메이션, 연예인, 영화, 음악, 운세 등이 포함된다. 이들이 제시한 분류 체계에 따라 멀티미디어 질의의 주제를 살펴보면, 8월 10

일의 경우 전체 멀티미디어 질의 42개 중 88%인 37개가 엔터테인먼트에 관한 질의이고, 5월 17일의 경우 전체 멀티미디어 질의 46개 중 87%인 40개가 엔터테인먼트에 관한 질의로, 엔터테인먼트의 비중이 압도적임을 알 수 있다. 음악 질의의 경우 “콜라 첫사랑,” “현영의 누나의 꿈” 처럼 특정한 노래 제목이나 특정 가수의 특정한 노래 제목을 직접적으로 찾는 경우가 많았으며, 동영상 질의의 경우에도 특정 TV 드라마나 프로그램의 특정 회차를 찾는 경우가 많았다. 이미지 질의의 경우, “블리치275화,” “나루토366화”와 같은 구체적인 질의와 “메인사진,” “이쁜 메인,” “이쁜 사진”과 같은 일반적인 질의의 비중이 비슷하였다.

4.2 클릭 로그 분석

2007년 5월 17일의 경우 이미지 컬렉션에 소속된 이미지 문서들에 대한 총 조회 수는 112,315회, 동영상 컬렉션에 소속된 동영상에 대한 총 조회 수는 102,166회, 음악 컬렉션에 대한 소속된 음악에 대한 총 조회 수는 15,330회로 나타났다. 또한 2007년 8월 10일의 경우 동영상 컬

〈표 4〉 멀티미디어 질의 리스트(2007년 5월 17일)

이미지	음악	동영상
9850사진저장	19곡은혜의복음	로맨스헌터15회
겨울날씨그림	나이트음악	무한도전53화
네이버웹툰	렉시하늘위로다운	밤비노5화
반고흐에서피카소까지그림	로꾸꺼	야동영상
블리치275화	마플피wma	야에니
우에키의법칙플러스39화	비오는날듣기좋은노래	웃음충전소타짱2다시보기
이쁜메인	사랑으로빛진날들	유승민탁구동영상
이쁜사진	사랑을위하여셋소폰연주	클레이모어7화
주변에풍경이보이는사진이나그림	아리랑다운	
중국지도	아빠함내세요	
포스터디자인	양과5집	
편이나간사진	음악무료다운로드	
	이기찬사랑그게뭐라고	
	잘해주지말걸 그랬어	
	편의전쟁ost	
	최신음악	
	코요테어글리ost	
	콜라츠사랑	
	클레지콰이스weety	
	특수공작원아이언맨ost	
	환상즉흥곡	
	believe-a.i	
	life is cool	
	mp3추천	
	sg워너비_ordinary_people.mp3	
	you give me something	

〈표 5〉 멀티미디어 질의 리스트(2007년 8월 10일)

이미지	음악	동영상
19금 만화	개와늑대의시간ost	남규리키스신
나루토366화	나에게로떠나는여행처럼신나는노래	놀러와공포특집
메인사진	노래방에서부르기좋은노래	무릎팍도사심형래다시보기
십우도	머피의법칙djdoc	밤비노마지막회
엠블럼	박효신눈의꽃	선택남녀다시보기
지도	빠빠노래	성인영화
	설레임배경음악	지피지기4회
	세븐mr	칼잡이오수정1회
	스트레칭음악	커피프린스1호점13화
	신화퍼펙트맨	커피프린스1호점2회

이미지	음악	동영상
	愛以外の何でもない	flv
	예지원상송	
	오종혁죽을만큼	
	오카리나로연주할만한클래식	
	올림픽포스가디언오프닝	
	요구르팅노래다운	
	우울할때듣는노래	
	자전거탄풍경-보물.wma	
	최신곡	
	컬러링다운로드	
	현영누나의꿈	
	ft아일랜드천둥	
	have you ever seen the rain	
	sex bomb	
	you raise me up	

렉션에서 발생된 총 조회 수는 122,829회, 이미지 컬렉션에서 발생된 총 조회 수는 103,986회, 음악 컬렉션에서 발생된 총 조회 수는 9,710회로 나타났다. 두 날짜의 클릭 로그를 종합하여 분석한 결과, 멀티미디어 관련 컬렉션들 중 가장 많은 클릭이 발생한 컬렉션은 동영상 컬렉션이었고, 근소한 차이로 이미지 컬렉션이 뒤를 이었고, 음악 컬렉션에서 가장 적은 수의 클릭이 발생하였다.

멀티미디어 질의 분석 시 음악 관련 질의가 가장 큰 비중을 차지하였으나, 클릭 로그 분석 시에는 음악 컬렉션의 클릭 수가 가장 적었다. 이는 이용자들이 음악 관련 질의를 입력 후 노출된 결과에서 음악 컬렉션보다는 블로그 컬렉션이나 카페 컬렉션 등을 선택하여 블로그나 카페에서 제공되는 음악을 감상하였기 때문인 것으로 추정된다. 또한 조회된 동영상 자료들의 상당수가 판도라TV나 고평TV, Diideo 등의 사이트에 이용자들이 올려놓은 국내 TV 프로그램, 드라마 등이었다. 한편 강화된 저작권법

에서는 이처럼 이용자가 저작권법에 위반되는 음악, 동영상 자료를 올리거나 포털이 이러한 자료를 제공하는 것이 모두 허용되지 않기 때문에, 자료 수집 당시와 현재 사이에 이용자들의 음악, 동영상 조회 행태가 다소 변화되었을 것으로 예상된다.

한편 멀티미디어 컬렉션에서 이용자들의 클릭이 가장 많이 발생한 질의들은 <표 6>과 같다. 5월 17일의 경우, 이미지 컬렉션과 동영상 컬렉션에서 모두 가장 많은 클릭을 발생시킨 질의는 “아이비굴욕”이었으며, 음악 컬렉션에서는 “양파 5집”이 가장 많은 클릭을 발생시켰다. 8월 10일의 경우 동영상 컬렉션에서 이용자들이 가장 많이 조회한 질의는 “진중권”이었는데, 조회된 동영상 대부분은 100분토론 TV 프로그램에서 진중권이 영화 디위에 대해 발언한 내용을 녹화한 동영상이었다. 즉 양일 모두 멀티미디어 컬렉션에서 최다 클릭이 발생한 질의는 “조로증”을 제외하면, 모두 연예 관련 질의들이었으므로 나타났다. 또한 최다 클릭 질의들

〈표 6〉 컬렉션별 최다 클릭 질의

	2007년 5월 17일		2007년 8월 10일	
	질의	클릭 횟수	질의	클릭 횟수
이미지 컬렉션	아이비 굴욕	36,077	조로중	10,351
음악 컬렉션	양파 5집	6,268	커피프린스 1호점	2,572
동영상 컬렉션	아이비 굴욕	31,091	진중권	27,052

의 경우, 질의 자체에 멀티미디어 검색어 대신 인명을 포함한다는 특징을 보이고 있다.

〈표 7〉과 〈표 8〉은 이미지, 음악, 동영상 컬렉션에서 각각 클릭 횟수가 1,000회 이상인 질의에 관한 기술통계를 보여 준다. 5월 17일과 8월 10일의 질의별 총 클릭 횟수의 중간값이 각각 239.5와 122이고, 이미지, 음악, 동영상 컬렉션에서의 평균 클릭 횟수가 모두 200이하임을 고려할 때, 클릭 횟수가 1,000회 이상인 질의는 멀티미디어에 관한 정보요구가 강한 인기 질의라고 볼 수 있다. 클릭 횟수가 1,000회 이상인 질의들은 음악 컬렉션보다 이미지 컬렉션과 동영상 컬렉션에서 더 많이 나타났다. 이러한 질의들의 분포는 중복 질의 포함 시에는 5월 17일이 약간 높았고, 중복 질의 제거 시에는 두 날짜

간에 큰 차이가 없는 것으로 나타났다.

〈표 9〉와 〈표 10〉은 이미지, 음악, 동영상 컬렉션에서 클릭 횟수가 1,000회 이상인 질의들에 대한 세부 정보를 보여 준다. 이 두 표로부터, 클릭 횟수가 1,000회 이상인 질의들 중에 이용자들이 단일 멀티미디어 컬렉션의 문서를 집중적으로 조회하는 경우도 있지만, 복수의 멀티미디어 컬렉션의 문서들을 동시에 많이 조회하는 경우도 있음을 알 수 있다. 특히 이미지를 조회하는 이용자들은 동영상도 함께 조회하는 경향이 있음을 알 수 있다. 8월 10일 “커피프린스 1호점” 질의의 경우 이미지, 동영상, 음악 컬렉션에서 모두 많은 클릭이 발생하였다. 또한 멀티미디어 컬렉션에서 이용자들이 많이 조회한 질의들의 주제도 대부분 엔터테인먼트와 관련된

〈표 7〉 컬렉션별 클릭 수 1,000 회 이상 질의(중복 질의 포함 시)²⁾

	이미지 컬렉션	음악 컬렉션	동영상 컬렉션	총계	전체질의 수
2007년 5월 17일	49(6.6%)	3(0.4%)	42(5.7%)	94(11.7%)	738
2007년 8월 10일	31(4.3%)	3(0.4%)	24(3.3%)	58(8%)	720

〈표 8〉 컬렉션별 클릭 수 1,000 회 이상 질의(중복 질의 제거 시)

	이미지 컬렉션	음악 컬렉션	동영상 컬렉션	총계	전체질의 수
2007년 5월 17일	15(2.3%)	3(0.5%)	14(2.1%)	32(4.9%)	653
2007년 8월 10일	17(2.5%)	3(0.4%)	17(2.5%)	37(5.4%)	675

2) 한 질의가 동시에 여러 컬렉션에서 조회되는 경우, 표에 중복되어 포함되기 때문에 〈표 7〉과 〈표 8〉의 질의 총계는 분석된 질의 수보다 약간 많다.

〈표 9〉 컬렉션별 클릭횟수 1,000회 이상 질의(2007년 5월 17일 질의)

단일 컬렉션	2개 컬렉션
이미지	이미지&동영상
고은아	메이비
다음	바다괴물
블리치275화	스타크레프트2
이승신	아이비
편의전쟁시청률	아이비굴욕
음악	안녕아빠
양파5집	원목초등학교
편의전쟁ost	정재연
동영상	편의전쟁
김미려방송사고	트위기컷
던파	이미지&음악
커밍아웃	윤하

〈표 10〉 컬렉션별 클릭횟수 1,000회 이상 질의(2007년 8월 10일 질의)

단일 컬렉션	2개 컬렉션	3개 컬렉션
이미지	이미지&동영상	이미지&음악&동영상
구글폰	진중권	커피프린스1호점
나루토366화	웹카운트다운인재팬	
다음	100분토론	
웹실화	조로증	
에밀리브라우닝	토사카준이치	
은혁미니홈피	소녀시대	
동영상	디워관객수	
남규리키스신	한영	
노다메칸타빌레	이흥기	
던파	이미지&음악	
미로밴드	남규리	
미로밴드너바나	음악&동영상	
조연진	ft아일랜드천둥	

어 있음을 알 수 있다. 즉 5월 17일 질의의 경우에는 21개 질의들 중 약 71%에 해당하는 15개 질의가 8월 10일의 24개 질의들 중 75%인 18개 질의가 엔터테인먼트에 대한 질의들이었다. 한편 2006년 5월 15일의 도그파일의 로그를 분석

한 Tjondronegoro, Spink, Jansen(2009)의 연구 결과에 따르면, 가장 많이 조회된 질의들의 경우, 성(性)과 관련된 질의가 압도적으로 많아서 네이버와 상당한 차이를 보이고 있었다.

5. 결론

본 연구에서는 국내 주요 검색 포털인 네이버 이용자의 멀티미디어 검색 행태를 조사, 분석하였다. 이를 위하여 2007년 5월 17일과 8월 10일의 네이버 통합 검색 질의 로그와 클릭 로그를 활용하였다. 연구 결과, 이용자가 입력하는 멀티미디어 질의 중 음악 질의가 차지하는 비중이 가장 높았으며, 이미지 질의와 동영상 질의의 비중은 비슷한 것으로 나타났다. 한편 멀티미디어 관련 컬렉션들 중 이용자가 가장 많이 조회하는 컬렉션은 동영상 컬렉션이었고, 이어서 이미지 컬렉션, 음악 컬렉션 순으로 나타났다. 이용자가 입력하는 멀티미디어 질의나 멀티미디어 컬렉션에서 클릭하는 문서 모두 엔터테인먼트와 관련된 주제가 가장 많은 것으로 나타났다. 특히 최신 드라마나 음악, 연예 관련 뉴스 등에 대한 정보 요구가 강한 것으로 나타났다. 또한 3개월 동안 멀티미디어 질의 검색 행태나 클릭 행태는 큰 변화 없이 일정한 것으로 나타났다.

본 연구의 결과는 향후 포털들의 보다 효과적인 멀티미디어 콘텐츠 구축 및 효율적인 검색 알고리즘 개발에 기여할 것으로 기대된다. 검색 포털들이 멀티미디어 관련 콘텐츠 구축 시, 이용자들이 검색하는 질의나 조회하는 문서에서 엔터테인먼트의 비중이 매우 높고, 시류에 민감한 주제가 큰 비중을 차지하며, 연예인과 같은 인물에 대한 정보 요구가 강하다는 사실을 반영할 수 있을 것이다. 특히 멀티미디어 컬렉션에서 이용자들의 클릭이 가장 많이 발생한 질의들의 경우, 질의에 멀티미디어 검색어 대신 단순히 인명이나 드라마명 등을 포

함한다는 점도 주목할 만하다. 또한 시한성 있는 질의들이 큰 비중을 차지한다는 사실은 멀티미디어 콘텐츠의 지속적인 업데이트가 필요함을 시사한다. 일반적으로, 검색 포털이나 시스템들은 검색 시스템과 검색 기술의 평가를 위하여 테스트 컬렉션을 구축할 수 있는데, 검색 포털들이 멀티미디어 테스트 컬렉션을 구축할 경우에도 이러한 특징을 반영할 수 있을 것이다. 이용자들이 멀티미디어 자료 조회 시 동시에 여러 매체를 조회한다는 특징은 현재의 통합 검색의 결과 노출 방식이 이용자 요구에 부응한다는 것을 암시한다. 한편, 스마트폰 등의 등장으로 이용자들의 멀티미디어 자료에 대한 요구와 이용도가 더욱 증가할 것으로 예상되므로, 향후 이용자들의 멀티미디어에 대한 정보 요구 및 이용 행태에 관한 지속적인 연구가 요청된다.

본 연구의 수행 결과 향후 연구가 요구되는 사항들은 다음과 같다. 첫째, 본 연구에서는 검색 포털 이용자들의 멀티미디어 검색 행태를 질의 로그와 클릭 로그에 근거하여 분석하였다. 따라서 이용자들이 특정한 방식으로 행동하는 이유, 이용자들의 멀티미디어 검색 서비스 및 결과에 대한 만족도 등을 심층적으로 조사하기 위해서는 인터뷰, 관찰 등을 통한 별도의 보완 작업이 요구된다. 둘째 향후 연구에서는 세분화되고 전문적인 기준들에 따라 포털의 멀티미디어 검색 서비스를 분석, 평가하는 것이 필요할 것으로 보인다. 예를 들어 이미지의 해상도, 음악의 음질, 동영상의 화질, 재생 속도 및 다운로드 속도 등의 측면에서 평가할 수 있으며, 포털들에서 제공하는 다양한 부가 서비스에 대해 평가할 수 있을 것이다. 셋째, 웹과 관련된 정보

환경이 급변한다는 점을 고려하여 보다 최신 로그 자료에 근거한 분석 작업이 바람직하다. 위에서도 언급되었듯이 전 세계적으로 강화되고 있는 저작권법이 이용자들의 음악 및 동영상 자료 이용 행태에 영향을 미칠 것으로 추정된다. 넷째, 본 연구에서는 이들 동안 입력된 질의와 클릭 문서들을 대상으로 멀티미디어 검색

행태를 분석하였다. 향후 연구에서는 보다 장기간에 걸쳐 수집된 질의를 통한 분석 작업 및 멀티미디어 검색 행태에 대한 추이 분석이 요구된다. 마지막으로 국외 웹 이용자와 국내 웹 이용자의 멀티미디어 검색 행태에 대한 심층적인 비교 분석이 요청된다.

참 고 문 헌

- [1] 김성희. 2004. 내용기반 이미지 및 비디오 검색 시스템 성능분석에 관한 연구. 『한국비블리아학회지』, 15(2): 97-115.
- [2] 김용, 문성빈. 2007. A Study on Recommendation System Using Data Mining Techniques for Large-sized Music Contents. 『정보관리학회지』, 24(2): 89-104.
- [3] 김용, 소민호. 2009. XML 기반의 동영상콘텐츠 검색 시스템 설계 및 구현. 『정보관리학회지』, 26(4): 113-128.
- [4] 노승민, 황인준. 2003. 멀티미디어 검색 시스템의 설계 및 구현. 『정보과학회논문지: 데이터베이스』, 30(5): 494-506.
- [5] 박소연, 이준호. 2002. 로그 분석을 통한 이용자의 웹 검색 이용 행태에 관한 연구. 『정보관리학회지』, 19(3): 111-122.
- [6] 박소연, 이준호, 김지승. 2005. 클릭 로그에 근거한 네이버 검색 질의의 형태 및 주제 분석. 『한국문헌정보학회지』, 39(1): 266-278.
- [7] 박창섭. 2007. 의미적 연관성을 이용한 멀티미디어 정보 검색. 『인터넷정보학회논문지』, 8(5): 67-79.
- [8] 유소영, 문성빈. 2004. 심미적 인상을 이용한 이미지 검색에 관한 실험적 연구. 『정보관리학회지』, 21(4): 187-208.
- [9] 백우진, 정선은, 김기영, 안의근, 신문선. 2008. Content-based image retrieval using data fusion strategy. 『정보관리학회지』, 25(2): 49-68.
- [10] 이준호, 박소연, 권혁성. 2003. 질의 로그 분석을 통한 네이버 이용자의 검색 행태 연구. 『정보관리학회지』, 20(2): 27-40.
- [11] 최광. 2009. 네이버, 세계 검색엔진순위 3년 연속 5위. 『매일경제』, 2009년 9월 2일. [online].

- [cited 2010.01.25]. <<http://news.mk.co.kr/outside/view.php?year=2009&no=462755>>.
- [12] 한민옥. 2010. NHN · 다음 · SK컴즈, 차세대 검색경쟁 '3사 3색'. 『디지털타임스』, 2010년 1월 4일. [online]. [cited 2010.01.29].
<http://www.dt.co.kr/contents.html?article_no=2010010502010431699001>.
- [13] Arkin, H., & Colton, R. 1963. *Tables for Statisticians*. New York: Barnes & Noble Inc.
- [14] Chung, E., & Yoon, J. 2009. "Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of *Flickr* tags and Web image search queries." *Information Research*, 14(3). paper 408. [online]. [cited 2009.12.26].
<<http://InformationR.net/ir/14-3/paper408.html>>.
- [15] Jansen, B. J. 2008. "Searching for digital images on the Web." *Journal of Documentation*, 64(1): 81-101.
- [16] Jansen, B. J., Goodrum, A., & Spink, A. 2000. "Searching for multimedia: analysis of audio, video, and image Web queries." *World Wide Web*, 3: 249-254.
- [17] Ozmutlu, C., Spink, A., & Ozmutlu, S. 2003. "Multimedia Web searching trends, 1997-2001." *Information Processing and Management*, 39(4): 611-621.
- [18] Spink, A., & Jansen, B. J. 2006. "Searching multimedia federated content Web collections." *Online Information Review*, 30(5): 485-495.
- [19] Tjondronegoro, D., Spink, A., & Jansen, B. J. 2009. "A study and comparison of multimedia Web searching: 1997-2006." *Journal of the American Society for Information Science and Technology*, 60(9): 1756-1768.

• 국문 참고자료의 영어 표기

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

- [1] Seong-Hee Kim. 2004. "A Study on the Performance Analysis of Content-based Image & Video Retrieval Systems." *Journal of the Korean Biblia Society for Library and Information Science*, 15(2): 97-115.
- [2] Yong Kim, & Sung-Been Moon. 2007. "A Study on Recommendation System Using Data Mining Techniques for Large-sized Music Contents.." *Journal of the Korean Society for Information Management*, 24(2): 89-104.
- [3] Yong Kim, & Min-Ho So. 2009. "A Study on Implementation of XML-Based Information Retrieval System for Video Contents." *Journal of the Korean Society for Information Management*, 26(4): 113-128.

- [4] Seung-Min Rho, & Een-Jun Hwang. 2003. "Design and Implementation of Multimedia Retrieval a System." *Journal of KIISE: : Database*, 30(5): 494-506.
- [5] Soyeon Park, & Joon-ho Lee. 2002. "Investigation Web Search behavior via Query Analysis." *Journal of the Korean Society for Information Management*, 19(3): 111-122.
- [6] Soyeon Park, Joon-Ho Lee, & Ji Seoung Kim. 2005. "An Analysis of Query Types and Topics Submitted to Naver." *Journal of the Korean Society for Library and Information Science*, 39(1): 266-278.
- [7] Park Chang-Sup. 2007. "Multimedia Information Retrieval Using Semantic Relevancy." *Review of Korean Society for Internet Information*, 8(5): 67-79.
- [8] So-Young Yu, & Sung-Been Moon. 2004. "An Exploratory Study of Image Retrieval Using Aesthetic Impressions." *Journal of the Korean Society for Information Management*, 21(4): 187-208.
- [9] Woojin Paik, Sun-Eun Jung, Gi-Young Kim, Euigun Ahn, & Moon-Sun Shin. 2008. Content-based image retrieval using data fusion strategy. *Journal of the Korean Society for Information Management*, 25(2): 49-68.
- [10] Joon-Ho Kee, Soyeon Park, & Hyuk-Sung Kwon. 2003. "Information Seeking Behavior of the NAVER Users via Query Log Analysis." *Journal of the Korean Society for Information Management*, 20(2): 27-40.
- [11] Kwang Choi. 2009. "Naver, SeGye Geomsaekengine Sunwi 3nyun yeonsok 5wi." *Maeil Business Newspaper*, 2 September 2009. [online]. [cited 2010.01.25].
 <<http://news.mk.co.kr/outside/view.php?year=2009&no=462755>>.
- [12] Min-Ok Han. 2010. "NHN · DAUM · SK communications Chasedae Geomsaekg-Gyeongjaeng '3Sa 3 Saek'." *DigitalTimes*, 4 January 2010. [online]. [cited 2010.01.29].
 <http://www.dt.co.kr/contents.html?article_no=2010010502010431699001>.

