

웹 검색질의어 분석을 통한 사회·문화적 특성에 관한 연구*

A Study on the Social and Cultural Characteristics of Web Queries

김 성 희**
Seong-Hee Kim

차 례

- | | |
|---------|---------|
| 1. 서 론 | 4. 결과분석 |
| 2. 선행연구 | 5. 결 론 |
| 3. 연구방법 | · 참고문헌 |

초 록

본 연구에서는 인터넷 정보 검색엔진으로부터 2007년부터 2009년까지 3년 동안 인기 검색어를 주제별로 분석함으로써 이용자의 검색어들이 시간이 흐름에 따라 주제별 변화양상을 살펴보았다. 또한 2009년도 네이버에서 제공하고 있는 상위 500개의 인기 검색어를 이용자의 검색의도에 따라 정보획득형(informational), 탐색형(navigational), 트랜잭션(transactional)형태로 구분하여 특성을 분석하였다. 그 결과 시간의 흐름에 따라 이용자의 관심 주제가 다르게 나타남을 알 수 있었다. 또한 이용자의 검색의도에 따라 인기 검색어를 분류한 결과 정보획득형 82%, 탐색형 10.8%, 트랜잭션형 7.2%로 나타났다. 이러한 연구는 미래의 검색엔진 또는 포털시스템에서 주제별 콘텐츠를 구축하여 제공하는 정책수립에도 도움이 될 것으로 기대한다. 더 나아가서 이용자들의 사회적, 문화적 관심사를 분석함으로써 인기키워드의 주제 분석을 통해 나타난 사회·문화적 특성을 파악하는 데 도움이 될 수 있을 것이다.

키 워 드

이용자, 검색의도, 검색엔진, 정보획득, 탐색, 트랜잭션, 검색질의어

* 이 논문은 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.

** 중앙대학교 문헌정보학과 교수

(Professor, Dept. Library and Information Science, Chung-Ang University, seonghee@cau.ac.kr)

• 논문접수일자: 2011년 8월 13일

• 최종심사(수정)일자: 2011년 8월 18일

• 게재확정일자: 2011년 9월 11일

ABSTRACT

This study aims to focus on classifying the search engine queries according to web query topic and the different user intents behind web queries. First, we classified 10,000 web query data set by topic. The results showed that there was significant differences in interesting topics across time. Also, we categorized 500 popular queries in web search engine as informational, navigational, or transactional. As a result, 82 percent of web queries are informational in nature, with about 10.8 percent for navigational and 7.2 percent for transactional. This results will help establish the policy to provide internet contents based on user's intent and also find out the social and cultural characteristics.

KEYWORDS

User Intent, Informational Goal, Navigational Goal, Transactional Goal, Web Queries

1. 서 론

현대사회는 정보화 시대로 이용자들은 수많은 인터넷 정보 속에서 원하는 정보를 언제 어디서나 손쉽게 제공받을 수 있게 되었으며 인터넷과 인터넷을 통한 정보검색은 사회전반은 물론 일상생활에 있어 매우 중요한 부분을 차지하고 있다. 방송통신위원회와 한국인터넷진흥원이 발표한 “2010년 인터넷 이용실태조사” 보고서에 따르면, 우리나라의 만 3세 이상 인구의 인터넷이용률은 77.8%로 인터넷 이용자 수는 37,010천 명으로 조사되었다. 2010년 5월 기준 우리나라 가구의 컴퓨터 보유율은 81.8%로 전년대비 0.4%p 증가했으며, 가구의 인터넷 보급률 또한 81.6%로 전년대비 0.4%p 증가한 것으로 나타나 이를 입증해주고 있다.

검색엔진은 인터넷상에서 정보를 쉽게 찾을 수 있게 도와주는 도구를 말한다. 이러한 검색엔진을 통해서 이용자들은 검색어를 직접 입력하거나 검색엔진에서 제시한 몇 가지 항목들 중에서 관련 항목을 선택함으로써 원하는 정보를 제공받게 된다(Hsieh-Yee 2001). 그동안 검색엔진에 관한 연구는 검색질의어와 문서 검색결과간의 정확성이나 관련성을 높이기 위한 알고리즘 개발과 같은 검색엔진의 성능을 향상시키기 위한 연구들이 많이 진행되어 왔다. 그러나 이러한 연구만으로는 이용자를 만족시킬 수 없었다. 이용자들은 무한한 정보의 홍수 속에서 이용자가 원하는 정보뿐만 아니라 불필요한 정보까지도 함께 제공되어 원하는 정보를 검색하는데 어려움이 있다. 더구나 일부 포털 사이트의 경우 검색엔진을 통해 입력되는 질의

어의 수가 하루 수천만 개를 넘고 있다. 하루에 검색엔진을 통해 입력되는 검색질의어의 크기도 방대하지만 그 내용도 워낙 다양하기 때문에 이용자들이 어떠한 목적을 가지고 정보를 검색하며 어떤 의도를 갖고 검색엔진을 이용하는지를 파악할 필요가 있다(Liu, Lieberman, and Selker 2002). 최근 들어서는 대부분의 주요 검색엔진 포털사이트에서는 매일 혹은 실시간으로 검색엔진을 통해 입력되는 질의어의 빈도와 순위 등을 제공하면서 빈도가 높은 질의어 위주로 내용이나 특징을 분석하고 있는 상황이다. 그 동안 여러 연구들을 통해 검색질의어가 이용자의 검색의도 또는 검색목적 표현하는 핵심요소라는 사실이 입증되어 왔다(Belkin, et al. 1993; Efthimiadis 2000). 또한 Pirolli(2007)에 의하면 검색질의어는 이용자의 정보요구가 외적으로 노출되는 표현이라고 주장하였다.

인터넷 검색엔진에 들어오는 엄청나게 많은 검색어들을 이용자 입장에서 다양한 분류기준에 의하여 분류를 시도해 왔다. 즉, 이용자 입장에서 선호되는 핵심 키워드나 콘텐츠를 찾아내고 다양한 기준에 의거해서 인터넷 콘텐츠를 새롭게 분류를 함으로써 새로운 웹 콘텐츠 분류체계 및 검색 행태 등을 찾아볼 수 있다. 다양한 웹 콘텐츠의 분류기준 가운데 웹 검색질의어를 이용자의 검색의도나 검색목적 기준으로써 분류하는 방법은 기본적으로 3가지 유형이 있다(Border 2002; Rose and Levinson 2004; Jansen, Booth, and Spink 2008). 이들

3가지 유형은 자료획득형(Informational goal), 탐색형(Navigational goal), 트랜잭션형(Transaction goal)이다. 자료획득형은 원하는 정보를 바로 찾고자 하는 유형의 검색어이며 탐색형은 특정 웹사이트로 이동하려는 의도를 갖는 검색질의어 유형이며 트랜잭션형은 파일을 다운로드 받거나 전자상거래를 하려는 의도를 갖는 검색질의어 유형이다. 이런 분류기준은 다음 절에 설명이 되어 있다.

사실 검색엔진 이용자들의 검색사용 패턴을 신속하고 정확하게 분석하고 예측하여 이용자의 검색의도를 파악하는 것은 이용자 만족에 있어서 매우 중요한 문제이다. 현재 대다수 포털사이트에서는 검색엔진을 통해 입력된 질의어들의 일간 혹은 주간 순위를 제공하고 있다. 이러한 순위 정보는 포털사이트의 콘텐츠를 구성하는데 도움을 주고 또한 사용자에게 최근 사회의 관심 동향이 어디로 흘러가는지 보여주면서 인터넷 소비자 행동 패턴을 예측할 수 있는 정보를 준다는 점에서 매우 중요하게 사용되고 있다. 또한 최근에는 이러한 순위 정보가 인터넷 광고에까지 연결되어 사용되고 있어 그 중요성이 갈수록 강조되고 있다(박상규 등 2007). 또한 질의어는 정보 대한 요구를 직접 표출하고 있는 표현이기 때문에 사회적, 문화적 흐름을 파악하는데 매우 중요하다고 할 수 있다. 예를 들어 일본의 지진에 대한 뉴스 발표가 나면 곧바로 네티즌들은 상세정보 획득을 위해 검색엔진 포털사이트 검색창에 '일본 지진', '쓰나미', '원전폭발'과 같은 키워

드들이 입력되고 검색질의어 순위에서 해당 키워드 들이 급상승하게 된다. 물론 이와 같은 사회적인 현상은 질의어 분석을 통하지 않더라도 동향 파악에 무리가 없겠지만 '얼짱'과 같은 문화적 키워드에 대한 동향은 네티즌 사회에서 자생적으로 발생해 증폭되는 문화현상 이므로 사회 문화 전반에 걸친 변화를 감지하기란 쉽지 않다(박상규 등 2007). 이런 동향 변화를 질의어 분석을 통해 빠르게 인식하거나 예측할 수 있다. 또한 이러한 인기 질의어를 광고와 결합시켜 최근 포털사이트들은 막대한 수입을 벌어들이고 있다. 따라서 최근 포털사이트들은 차세대 서비스 방향을 결정하기 위해 이런 분석에 많은 시간과 노력을 투자하고 있다.

이러한 맥락에서 본 연구에서는 국내 포털 업체 중의 하나인 네이버에서 제공하는 인터넷 정보검색엔진으로부터 2007년부터 3년 동안 검색어를 주제별로 분석함으로써 이용자의 검색어들이 시간이 흐름에 따라 주제별 변화양상을 살펴보았다. 또한 2009년도 네이버 검색 엔진의 상위 500개의 인기 검색어를 전문가 3인에 의하여 이용자의 검색의도에 따라 정보획득형(Informational), 탐색형(Navigational), 트랜잭션형(Transactional) 형태로 인터넷 콘텐츠를 분류하고 인기 검색어에 나타난 특성을 분석하였다. 이러한 연구는 미래의 정보검색엔진 개발자들의 이용자의 검색의도를 미리 파악하고 예측하여 적합한 문서를 상위에 검색될 수 있게 하는데 도움을 줄 수 있으며 더 나아가서 이용자들의 사회적, 문화적 관심사를 분석

함으로써 인기키워드의 주제 분석을 통해 나타난 사회·문화적 특성을 파악하는데 도움이 될 수 있을 것이다.

2. 선행연구

본 연구에서는 시간의 흐름에 따라 웹 검색 질의어의 주제들이 어떻게 변화하는지를 살펴보고 웹 환경하에서 이용자들이 실제로 어떠한 검색의도나 검색목적 을 가지고 있는지를 분석하였다. 즉, 검색포털사이트의 검색질의어를 분석하고 검색질의어들을 주제별로 그리고 검색의도에 따라 분류함으로써 문화적, 사회적 특성들을 파악하는데 도움이 되고자 하였다. 이와 관련된 선행연구들을 이용자의 검색의도에 따른 검색질의어 분석과 관련된 연구와 검색질의어들의 주제를 분류한 연구로 나누어서 살펴보았다.

먼저 그 동안 이용자의 정보추구행위에 관한 연구가 많이 수행되어 왔다. 대표적인 정보이용행태와 관련된 모델들로는 Bates(1989), Dervin(1999), Ellise(1989), Johnson(1997), Krikelas(1983), Kuhlthau(1991), Leckie, Pettigrew and Sylvain(1996), and Wilson(1981, 1999) 등에 의해 개발된 모델들이다. 이러한 정보이용행태에 관련된 모델 중에서 먼저 이용자의 검색목적이나 검색의도를 수행한 연구들로는 Bates(1979)의 연구로 이용자의 정보검색행태에 대해 초창기에 이루어졌으며 정보검색 시스

템에서 이용자의 인지적 상태를 설명함으로써 정보검색 시스템 디자인에 반영하려는 연구가 Belkin(1982)에 의해 수행되었다. 이 연구에서는 변칙적 지식상태(Anomalous state of Knowledge: ASK)에 대한 프레임워크를 제시하여 서로 다른 이용자들의 정보요구행태에 대한 차이점을 분석하였다. 또한 이용자들의 검색패턴이나 행동을 관찰하여, 이용자의 검색목적을 파악하는 연구가 수행되었다. Silverstein et al.(1999)은 6주 동안 알타비스타 이용자들이 9억 9천만 개 이상의 질의를 트랜잭션 로그를 이용하여 분석하였다. Spink et al.(2001)은 1997년부터 2001년까지 2년에 한 번씩 하루를 선정하고 그날 익사이트에 입력된 질의들 중 무작위로 추출된 약 2,500개의 주제를 분류하였다. 그 결과 검색질의어가 엔터테인먼트와 성관계로부터 전자 상거래 관련으로 변화하였으나 이용자들의 전반적인 검색행태는 변하지 않았음을 보고하였다. Broder(2002)는 검색질의어를 3가지 유형: 탐색형, 정보획득형, 트랜잭션형에 따라 분석하였다. 이 연구 결과에 따르면 웹 질의어의 48%가 정보획득형이며 20%가 탐색형이고 30%가 트랜잭션형 질의어로 나타났다. 나머지 2%로는 3가지 유형 어디에도 포함되지 않는 질의어 유형으로 조사되었다. Spink와 Jansen(2004)은 다양한 검색엔진을 이용해서 전자상거래 관련된 질의어들을 조사하였는데 전자상거래 관련 질의어들은 전체 검색질의어의 12%에서부터 24%에 이르기까지 검색엔진에 따라 다양한 것으로 파악하였

다. Jansen, Spink와 Pedersen(2005)은 2002년 알타비스타 검색로그로부터 상위 검색질의어를 분석한 결과 상위 15개의 질의어는 모두 탐색형 질의어인 것으로 보고하였다. Jansen, Spink와 Pederson(2005)은 1997년부터 2002년까지 익사이트 이용자들이 주로 검색한 주제를 연도별로 분석하여 비교하였다. 그 결과 엔터테인먼트, 레크리에이션 관련 검색질의어는 많이 감소한 반면, 컴퓨터, 인터넷 관련 검색어는 비슷한 수준을 유지하는 것으로 나타났다. 또한, 전자상거래, 여행, 고용, 경제 관련 질의가 증가하였으며 인물, 장소, 사물에 관한 주제가 가파르게 상승하였다고 보고하였다. 이 연구자들은 또한 2002년 알타비스타에서 생성된 질의들로부터 2,603개를 무작위로 추출하여 이를 주제별로 분류하고 Silverstein et al.(1999)의 연구결과와 비교하였다. 그 결과 질의어의 주제가 좀 다양해졌으며 성과 관련된 주제가 감소된 것으로 나타났다. Rose와 Levson(2004)은 알타비스타(altavista)의 검색로그를 대상으로 검색질의어를 이용자의 검색목적에 따라 분류하였다. 약 100-200여개의 검색질의어를 정보획득형, 탐색형, 자원형(resource)에 따라 구조적으로 분류하였다. 그 결과 질의어의 62%는 정보획득형이었고 13%는 탐색형, 24%는 리소스형인 것으로 나타났다. 또한 이 3개의 유형을 기준으로 하여 검색목적에 따라 질의어를 상세하게 분류하였다. 국내의 연구로는 박소연, 이준호(2002, 2005)는 1년 동안 네이버에 입력된 검색질의를

기반으로 이용자의 검색 행태 추이를 분석하였다. 이 연구에서는 이용자가 조회한 문서를 기록한 클릭로그에 근거하여 1년 동안 네이버 이용자들이 입력한 질의어의 형태 및 주제의 전반적인 특징을 조사하였다. 그 결과 네이버 이용자가 가장 많이 검색한 주제는 컴퓨터/인터넷, 엔터테인먼트, 쇼핑, 게임, 교육, 기업, 라이프스타일, 금융/경제 순으로 나타났다. 박소연, 이준호(2005)의 연구에 따르면 국내 웹 이용자들이 국외 웹 이용자들보다 컴퓨터/인터넷과 엔터테인먼트, 특히 인터넷포털 사이트, 커뮤니티 사이트, 연예인에 대한 관심도가 매우 높다고 할 수 있으며, 이는 양국 간의 문화적 차이를 반영한다고 할 수 있다. 또한 성인사이트 관련주제가 차지하는 비중이 외국에 비해 많이 낮다고 하였다. 박상규(2007) 등은 검색엔진을 통해 인터넷 검색 패턴을 분석하는데 활용될 수 있는 질의어 분포의 정상성을 Pareto 분포와 Zipf 모형을 이용해 검토하였다. 이들 연구에서는 검색엔진을 통해 들어오는 수많은 검색 질의어가 어떤 패턴(pattern)을 이루고 있는지를 분석하기 위해서 빈도가 높은 질의어를 1년을 기준으로 표본조사 해 본 결과 검색 입력어가 일정한 패턴을 띠고 있었으며 검색질의어가 일정한 패턴을 갖는다는 것은 사람들의 관심사에 어떤 흐름이 존재한다는 것으로 해석하였다. 따라서 이러한 패턴을 파악하게 되면 사람들의 사회적 관심사에 대한 전체적인 흐름을 분석해 낼 수 있다는 점을 제시하였다. 실제 2주일간의 일간 검색어 분포를 빈도 - 순서 그림

으로 그린 후 이를 Pareto 모형으로 적합시킨 결과, 그 기울기가 1.33에서 1.34 정도로 일관되게 측정되고 있어 Pareto 분포를 이용해 일간 검색질의어 분포의 정상성을 평가할 수 있음을 제시하였다. 이들 연구에 따르면 정상성(regularity)을 유지하고 있는 날들과 그렇지 않은 날들이 존재할 경우 비정상 패턴을 보다 심도 있게 분석할 수 있다고 하였다. 만약 어떤 날의 패턴이 이 정상성으로부터 벗어나 있다면 그 날은 뭔가 특별한 사회적 이슈가 등장한 날로 간주하고 분석할 수 있게 되어, 사회적 변동에 대한 대처와 예측이 가능해 진다고 보았다. 따라서 검색질의어의 정상성을 통해 검색질의어의 변동을 분석할 수 있는 기준 척도로 사용 가능할 것임을 시사하였다. 이상에서와 선행연구를 보면 그 동안의 연구결과는 검색어를 주제별로 적절하게 분류되어 주제의 특성 및 분포를 설명하였다. 또한 검색질의어를 이용자의 검색사용 의도에 따라 정보획득형, 탐색형, 트랜잭션으로 구분하여 연구들이 진행되어 왔음을 알 수 있었다. 하지만 인터넷 포털사이트의 인기 검색어들의 연도별로 관심주제를 비교하거나 사회·문화적 특성을 분석하지는 않았다. 따라서 본 연구에서는 네이버에서 제공하고 있는 검색엔진의 인기검색질의어들의 주제를 연도별로 분석해서 주제의 흐름과 특징들을 분석하였으며 이용자의 검색의도에 따라 검색질의어를 정보획득형, 탐색형, 트랜잭션형으로 구분하여 분석하였다. 이런 연구결과는 포털 사용자들이 효율적으로 구축된 콘텐츠를 잘 사용

할 수 있는 검색엔진 개발 환경을 제공할 수 있다. 또한 검색질의어들의 전반적인 주제를 이용자의 검색의도에 따른 분석을 의 특성을 분석하거나 특정 패턴에 관한 연구를 함으로써 지속적으로 변하는 사회·문화적 현상이나 관심을 반영할 수 있는지를 살펴보고자 하였다.

3. 연구방법

정보검색 포털 업체인 네이버는 2007년부터 매년 '네이버 트렌드 연감'은 1위부터 1만 위까지의 통합 검색어를 바탕으로 인물, 생활, 경제, 엔터테인먼트, 쇼핑, 건강 등 총 15개 분야를 대상으로 분류하여 관심분야 연구자들의 연구를 수행하는데 도움을 주고자 제공하여 오고 있다. 본 연구에서는 이러한 네이버 트렌드 연감에서 제공하고 있는 통합검색어들의 주제들을 2007년부터 3년간의 걸쳐 관심주제의 변화양상을 분석하였다. 또한 문헌정보학 관련 배경지식을 갖고 있는 전문가 3인이 네이버에서 제공하고 있는 2009년도의 인기 검색어 500개를 이용자의 검색의도에 따라 정보획득형, 탐색형, 트랜잭션형으로 분석하였다. 구체적인 내용은 다음과 같다.

3.1 주제 분류

인터넷 콘텐츠란 인터넷 등의 통신망을 통해 제공되는 각종 디지털정보로서 디지털로

처리된 다양한 형태의 정보를 의미한다. 인터넷 콘텐츠는 다양한 플랫폼으로 공급/이용되고 있고 상업적, 비상업적 콘텐츠가 혼재되어 있으며 콘텐츠의 형태가 다양하여 일관성 있는 분류체계를 마련하기 어렵다. 한국 인터넷진흥원(2007)에 따르면 일반적으로 인터넷 콘텐츠 분류는 매우 다양한 방법이 적용되고 있는데 일반적으로 산업별 분류, 비즈니스별 분류, 주제별 분류, 전달매체별 분류, 기타 등으로 구분할 수 있다. 이들 가운데 주제별 분류는 콘텐츠가 담고 있는 내용을 기준으로 분류하는 방식이다. 인터넷 이용자의 콘텐츠 이용행태를 이해하는데 적합하며 비상업적 콘텐츠가 포함된다는 장점이 있다. 포털사이트의 콘텐츠 분류체계인 디렉터리 개념과 유사하며, 순위사이트에서 사이트를 분류할 때에도 이 분류기준을 사용한다(한국인터넷진흥원 2006). 본 연구에서는 네이버에서 제공하는 주제 분류로 주제 카테고리는 경제, 환경, 스포츠, 사회와 정치, 문화와 예술, 건강, 교육과 학문, 컴퓨터와 인터넷, 세계와 여행, 게임, 뉴스와 미디어, 쇼핑, 생활, 엔터테인먼트, 인물과 사람으로 구분되었다.

3.2 이용자 검색의도에 따른 검색질의어 분류기준

본 연구에서는 2009년도 네이버 검색포털 사이트에서 상위 500개의 인기검색질의어¹⁾를 기존의 선행연구들(Jansen, Spink, and Pedersen

2005: Jansen, Booth, and Spink 2008: Rose and Levinson 2004)을 바탕으로 해서 이용자 검색의도에 따라 3개의 유형: 1) 정보획득형(Informational), 2) 탐색형(Navigational), 3) 트랜잭션형(Transactional)으로 분류하였다. 문헌정보관련 지식을 갖고 있는 3명이 평가자가 500개의 질의어를 3개의 유형에 따라 분류하였으며 분석 전에 검색의도에 따른 분류기준에 대해 충분한 토론을 거쳤다. 질의유형및 세부질의어의 분류기준 설명하면 다음과 같다.

3.2.1 정보획득형

정보획득형은 어떤 주제와 관련된 정보의 내용을 검색하고자 하는 경우로 주로 방법, 종류, 유형 등에 관한 검색질의어가 여기에 해당된다. 주로 리스트형태의 질의 유형이나 복수개의 결과들이 나타날 수 있는 질의어가 여기에 해당된다. 이 부분은 검색결과 창에서 원하는 결과를 읽을 수 있으며 다른 사이트로 이동이 필요 없는 검색질의어의 유형을 의미한다고 볼 수 있다. 또는 탐색형이나 트랜잭션형이 아닌 형태의 경우 여기에 해당된다고 할 수 있다. 질의어 수는 다른 유형에 비해 상대적으로 길다. 이러한 유형의 질의어로는 “어린이대공원 입장료”와 같이 특정한 사실을 알고 싶어 하는 질의어나 아니면 특정주제대한 일반적인 정보로 “어린이대공원”과 같이 어떤 주제에 관련된 일반적인 정보라든지 “카레와 잘 어울리

는 음식” 또는 특정 서비스를 받을 수 있는 곳을 질의하는 검색어 등이 여기에 포함된다.

3.2.2 탐색형

특정 URL 혹은 출처 즉 홈페이지의 주소로 이동하기 위한 검색으로 회사, 기업체나 조직의 사이트주소를 검색하는 질의어 등이 여기에 해당된다. 일반적으로 질의어 길이 또는 키워드 수는 다른 유형에 비해 짧으며 찾고자 하는 결과는 여러 개보다는 하나일 가능성이 많다. 일반적으로 이런 유형의 검색질의어는 이용자가 특정사이트를 마음속에 두고 검색을 하는 경우가 많다. 다음과 같은 질의어가 이런 유형에 속한다고 볼 수 있다.

- 그랜드 하얏트 호텔(예상되는 사이트:
<http://www.grandhyattseoul.co.kr>)
- 삼성(예상되는 사이트:
<http://www.samsung.com>)
- 대한항공(예상되는 사이트:
<http://kr.koreanair.com/>)

이런 유형의 질의어는 기존의 정보검색 분야에서는 “known item”을 의미하는 것이라고 볼 수 있다. 탐색형 질의어는 두 가지 형태로 구분할 수 있다. 첫째 탐색형에서 트랜잭션 형태로 이동하는 질의어의 형태로 이용자가 의도하는 URL이 트랜잭션 사이트일 경우이

1) 네이버에서는 2007년부터 상위 인기 검색어를 네이버 연감을 통해서 제공하고 있다.

다. 두 번째로는 탐색형에서 정보획득형으로 이동하는 질의어의 형태로 이용자가 의도하는 최종 URL 역시 정보획득 사이트가 될 수 있는 경우로 “네이버”, “야후” 등의 검색어들이 여기에 해당된다.

3.2.3 트랜잭션형

음원이나 프로그램 다운로드 혹은 뭔가 액션(action)을 원하는 검색질의어가 여기에 해당된다. 소프트웨어, 비디오, 오디오, 게임, 구매, 이미지, 노래, 영화와 같은 질의어 등이 여기에 해당된다. 특히, 오락이나 전자상거래와 관련된 질의어, 팟캐스트(Podcasts), 전자책 등도 여기에 해당된다. 예를 들어 상품을 온라인에서 판매 또는 구매하려는 의도가 있다면, 이용자 질의어는 트랜잭션 질의어로 분류된다. 물론, 짧은 이용자의 질의어 속에는 꼭 한 가지 의도만 가지고 있다고는 할 수 없다. 예를 들어, “WindowXP”라는 질의어에는 이것에 대한 특징이나 정보를 알고 싶어서, 혹은 해당 사이트의 URL로 이동하기 위해서 혹은 다운로드 하고 싶어서 이용하는 이용자들도 있다. 구체적인 예는 다음과 같다.

- 획득(obtain): 구체적인 자원이나 아이템을 얻는 질의어로 “음악악보” 등이 여기에 해당된다.
- 다운로드(download): 다운로드하기 위해 파일을 찾는 질의어로 “mp3 다운로드”가 여기에 해당된다.

- 인터랙트(interact): 다른 사이트와 프로그램이나 자원을 서로 상호 작용하는 것과 관련된 것으로 이러한 질의어의 예로는 “전자책구매” 등과 같이 상품 구매 등을 들 수 있다.

이상의 분류기준에 의거해서 2009년도 네이버 상위 500개의 검색질의어를 3명의 평가자가 분류하였다. 먼저 3명의 분류평가자는 문헌정보학 배경을 갖고 있는 전문가로 분류기준에 대해 서로 충분한 토의과정을 거치고 20개의 질의어를 먼저 분류를 한 다음 문제점을 살펴보고 서로의 의견을 교환한 후 각 질의어를 분류하였다. 일부 질의어 중에 특정 분야에 분류하기 애매한 경우에는 3명이 충분한 토의 과정 및 직접 검색어를 검색사이트에 입력해서 얻은 경험을 토대로 분류하도록 노력하였다.

4. 결과분석

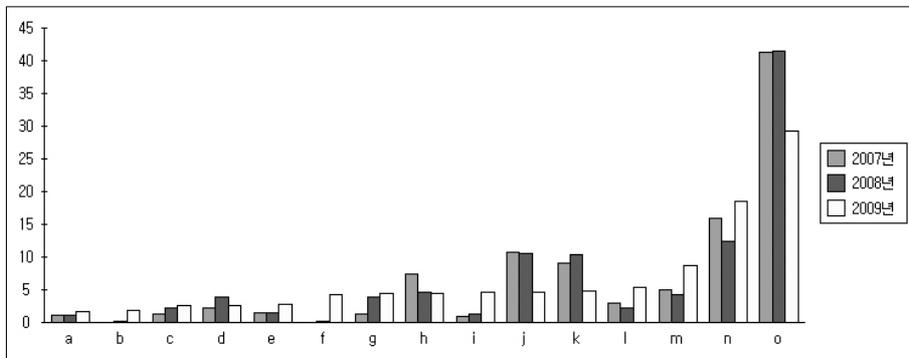
4.1 웹 검색질의어의 주제별 분석

네이버가 해마다 15개의 주제 분야에 따라 제공하고 있는 주제별 검색어를 연도별로 정리하여 관심주제 변화 양상을 분석하였다. <표 1>은 통합검색어 1천위 주제별 검색어 점유율이다.

〈표 1〉 주제별 검색어 점유율

(퍼센트: %)

순위	주제	주제 표시	2007년	2008년	2009년	평균
1	인물과 사람	o	41.3%	41.5%	29.31%	37.37%
2	엔터테인먼트	n	15.8%	12.4%	18.5%	15.57%
3	게임	j	10.7%	10.6%	4.64%	8.65%
4	뉴스와 미디어	k	9.1%	10.4%	4.82%	8.11%
5	생활	m	5%	4.3%	8.67%	5.99%
6	컴퓨터와 인터넷	h	7.3%	4.5%	4.43%	5.41%
7	쇼핑	l	3%	2.2%	5.32%	3.51%
8	교육과 학문	g	1.3%	3.9%	4.42%	3.21%
9	사회와 정치	d	2.2%	3.8%	2.57%	2.86%
10	세계와 여행	i	0.9%	1.3%	4.57%	2.26%
11	스포츠	c	1.2%	2.2%	2.5%	1.97%
12	문화와 예술	e	1.4%	1.5%	2.8%	1.9%
13	건강	f	0%	0.2%	4.25%	1.48%
14	경제	a	1%	1.1%	1.68%	1.26%
15	환경	b		0.1%	1.87%	0.98%
	합계		100%	100%	100%	100%



〈그림 1〉 주제별 검색어 점유율

(a: 경제, b: 환경, c: 스포츠, d: 사회와 정치, e: 문화와 예술, f: 건강, g: 교육과 학문, h: 컴퓨터와 인터넷, i: 세계와 여행, j: 게임, k: 뉴스와 미디어, l: 쇼핑, m: 생활, n: 엔터테인먼트, o: 인물과 사람)

〈표 1〉에 나타난 바와 같이 주제는 네이버에서 분류한 것으로 모두 15개로 구분되었다. 주제 카테고리는 경제, 환경, 스포츠, 사회와

정치, 문화와 예술, 건강, 교육과 학문, 컴퓨터와 인터넷, 세계와 여행, 게임, 뉴스와 미디어, 쇼핑, 생활, 엔터테인먼트, 인물과 사람으로 구

분되었다. 그 결과를 살펴보면 <표 1>에서 보는 것과 같이 15개의 주제 분야 중에서 인물과 사람분야가 전체 중에 약 37%를 차지하였다. 2007년과 2008년에는 각각 41.3%, 41.5%를 차지하였으며 2009년에는 29.31%로 나타나 약 10% 이상 점유율이 하락한 것으로 분석되었다. 다음으로 높은 점유율을 차지하는 주제 분야로는 엔터테인먼트인데 2007년에 15.8%, 2008년에는 12.4%로 약 3.4% 관심도가 떨어진 것으로 나타났으나 2009년에 다시 18.5%로 증가하였다. 이어서 게임분야가 2007년과 2008년에 각각 10.7%, 10.6%로 나타났으나 2009년에 4.64%로 절반가량이 관심도가 낮아진 것으로 나타났다. 뉴스와 미디어도 게임분야와 마찬가지로 2007년과 2008년에 9.1%, 10.4%로 비슷하게 나타났으나 2009년에는 4.82%로 전년도 대비 절반가량의 점유율을 보였다. 컴퓨터와 인터넷 분야는 2007년도에는 7.3%로 높은 점유율을 나타냈으나 2008년과 2009년에는 4.5%, 4.43%로 나타나서 관심도가 떨어졌다. 생활관련 주제 분야에서는 2007년과 2008년에는 5.0%, 4.3%로 나타났으며 2009년에는 8.67%로 나타났다. 이는 지난해와 비교해서 거의 두 배 가까이 상승한 것으로 나타났다. 쇼핑관련 주제 분야에서는 2007년도와 2008년도 각각 3.0%, 2.2%로 나타났으며 2009년도에는 5.32%로 나타나 전년도 대비 두 배 이상 높았다.

스포츠 분야에서는 2007년에는 1.2%, 2008년에는 2.2%, 2009년에는 2.50%로 꾸준히 상

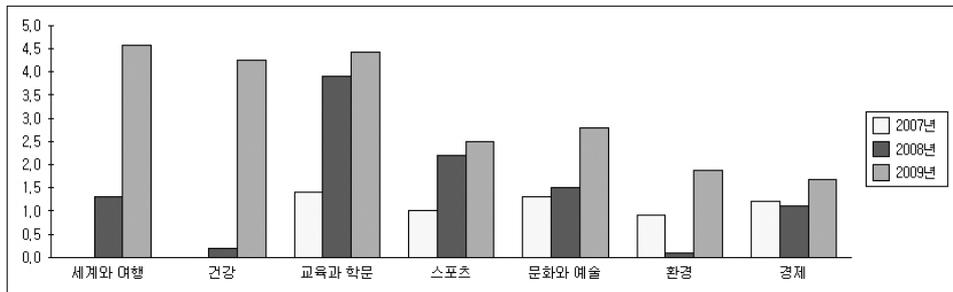
승하는 것으로 나타났다. 교육과 학문분야에서는 2007년도 1.3%, 2008년도 3.9%, 2009년도에는 4.42%로 계속해서 증가하고 있는 것으로 나타났다. 사회와 정치 분야에는 2007년 2.2%, 2008년 3.8%, 2009년 2.57%로 나타나 2008년도에 상대적으로 높은 비율을 차지하였다. 문화와 예술분야에서는 2007년도 1.4%, 2008년 1.5%, 2009년 2.8%로 과거 2년에 비해 거의 두 배 가까이 상승하였다. 세계와 여행 분야에서는 2007년도 0.9%, 2008년도 1.3%, 2009년도 4.57%로 과거 2년에 비교해서 3배 이상 증가한 것으로 나타났다. 그 이외에도 경제 분야에서는 2007년 1.0%, 2008년 1.1%, 2009년 1.68%로 꾸준히 증가하였으며 환경 분야도 계속해서 증가하고 있는 것으로 나타났다.

이상에서 15개 주제 분야별로 웹 검색질의어의 점유율을 년도 별로 살펴보았다. 지난 3년 동안 계속해서 이용자의 관심사가 증가한 분야는 <표 2>와 같다. <표 2>에서 보는바와 같이 이용자의 검색질의어가 지난 3년간 계속 증가한 주제 분야로는 15개 분야 중에서 경제, 환경, 스포츠, 문화와 예술, 건강, 교육과 학문, 세계와 여행 분야 등 7개 분야였으며 이 중에서도 상대적으로 가장 많이 증가한 주제 분야로는 세계와 여행 분야였으며 이어서 건강관련 분야와 교육과 학문 분야도 많은 증가율을 나타내었다. <그림 2>는 지난 3년 동안 계속 증가한 주제 분야를 그래프화 한 것이다.

〈표 2〉 3년간 계속 검색질의어가 증가한 주제 분야

(단위: %)

순위	주제	2007년	2008년	2009년	증감평균
1	세계와 여행	0.9	1.3	4.57	2.29
2	건강	0	0.2	4.25	2.13
3	교육과 학문	1.4	3.9	4.42	1.51
4	스포츠	1	2.2	2.5	0.75
5	문화와 예술	1.3	1.5	2.8	0.75
6	환경	0.9	0.1	1.87	0.49
7	경제	1.2	1.1	1.68	0.24



〈그림 2〉 3년간 검색질의어가 계속 증가한 주제 분야

이상에서 지난 3년간의 검색질의어를 분석한 결과는 포털 이용자들의 관심사 선호경향을 보여 주는 것으로 정치나 사회문제와 같은 다소 딱딱한 뉴스보다는 여행이나 문화와 예술 같은 좀 더 부드러운 주제 분야에 많은 관심을 나타내고 있음을 알 수 있었다. 본 분석에는 실제 2007년부터 2009년까지 상위 10개의 인기 검색어를 분리해서 분석한 결과 게임 또는 연예나 스포츠와 같은 감각적이고 흥미 있는 주제를 선호하고 있음을 알 수 있었다. 또한 구글에서 제공하는 2007년부터 2009년까지의 급상승 인기검색질의어를 분석한 결과 역시 해

마다 이용자들의 관심사가 다르게 나타남을 알 수 있었다. 급상승형 검색질의어는 갑자기 검색어 빈도가 급상승 하는 유형으로 다양한 사회적 뉴스 등 어느 정도 예측 가능한 유형일 수도 있고 문화적 양상이나 기타 감지하지 못하는 급상승 패턴이 있을 수 있다. 본 연구에서 급상승형 검색질의어는 갑자기 검색어 빈도가 급상승 하는 유형으로 다양한 사회적 뉴스 등 어느 정도 예측 가능한 유형일 수도 있으나 본 연구 결과에 의하면 문화적 양상이나 기타 감지하지 못하는 급상승 패턴이 있는 것으로 나타났다. 예를 들면 2007년 급상승 검색질의어

로 이용자 관심사를 살펴보면 게임서비스가 가장 관심 있는 것으로 나타났으며 학력위조와 같은 사회적 이슈에도 커다란 관심을 보인 것으로 나타났다. 주제영역별로 분석한 결과를 보면 세계와 여행, 건강, 교육과 학문분야에서 계속해서 관심을 보이는 것으로 나타났다. 2008년 검색량이 가장 많이 증가한 급상승어에는 '2008 베이징'이 선정되어 이용자들이 올림픽에 많은 관심을 보인 것으로 나타났다. 2009년 한국에서는 급상승 검색어(fastest rising queries)에는 신종플루부터 다양한 사회적 이슈와 관심사가 반영되었다. 따라서 이러한 인기검색질의어 분석은 사회적·문화적 관심사를 파악하는데 사용될 수 있음을 알 수 있다.

4.2 이용자의 검색의도에 따른 검색질의어 분석

본 연구는 인기검색질의어를 주제별로 분석하고 정보획득형, 트랜잭션형, 탐색형으로 인기 검색어를 분석을 함으로써 이용자의 다양한 검색목적이나 검색의도를 파악하고자 하였다. 이를 위해 사용된 데이터는 2009년에 국내 네이버 검색엔진에서 수집된 검색질의어 중에서 인기 검색어 상위 500개를 네이버 연감으로부터 추출하여 분석하였다. 여기서 정보획득형은 이용자가 필요로 하는 정보나 콘텐츠를 검색하기 위한 질의어이며 탐색형은 특정사이트를 찾아가기 위해 URL을 검색하기 위한 질의어이며, 트랜잭션형은 텍스트 형태의 정보가

아닌 파일형태의 자료를 다운로드 받거나 어떤 프로그램을 실행하기 위한 검색질의어 형태이다.

4.2.1 인기검색질의어의 주제 분석

〈표 3〉은 2009년도 네이버 연감에 나타난 인기 검색어 중에서 상위 500개의 검색질의어를 따로 추출하여 15개의 주제별로 분석하여 정리한 결과이다.

〈표 3〉에 나타난 바와 같이 인물관련 검색어가 274(54.8%)개로 타 분야에 비해 압도적으로 가장 많은 관심을 보였다. 이어서 게임(11.8%), 뉴스(10.6%), 엔터테인먼트(8.4%) 순으로 나타났다. 이 외에 생활, 컴퓨터, 사회관련 인기 검색어는 각각 3.6%, 2.8%, 2.6%로 나타났다. 스포츠, 교육, 쇼핑 관련 인기 검색어로는 1.4%, 1.2%, 1.2% 순이었으며 경제 와 문화는 각각 0.8%, 0.6%였다. 건강관련 인기 검색어로는 단 한 개가 500위 안에 있었으며 검색어는 "신종플루"였다. "신종플루"는 2009년 한국에서는 급상승 검색어(fastest rising queries)에 해당되는 것으로 급상승어는 다양한 사회적 이슈와 관심사가 반영될 수 있음을 알 수 있다. 이런 결과는 앞에서 과거 3년 동안 검색어를 주제별로 구분한 결과에서도 나타났듯이 이용자들은 주로 딱딱한 주제보다 유연하고 즐거움을 제공할 수 있는 주제에 관심이 많음을 알 수 있다.

〈표 4〉는 500개의 검색어들을 이용자의 검색의도에 따라 1) 정보획득형, 2) 탐색형, 3) 트랜잭션형으로 문헌정보학 관련 전문가 3인

〈표 3〉 인기 검색어의 주제별 분포

대분류	소분류	검색어 수	비율(%)
인물과 사람	연예인, 스포츠, 방송인, 정치·사회, 기업인	274	54.8
게임	게임	59	11.8
뉴스	국내TV라디오	53	10.6
엔터테인먼트	국·내외 영화, 국내대중음악, 방송연예화제, TV만화 애니메이션	42	8.4
생활	생활정보, 교통정보 등	18	3.6
컴퓨터와 인터넷	온라인문화	14	2.8
사회와 정치	기관단체, 국제 및 국내사회	13	2.6
스포츠	스포츠경기화제, 스포츠 레포즈, 스포츠팀	7	1.4
교육과 학문	교육기관, 교육자료, 시험자격증	6	1.2
쇼핑	가전전자용품, 생활용품, 중고용품	6	1.2
경제	금융재테크	4	0.8
문화와 예술	문화정보, 문화공간	3	0.6
건강		1	0.2
합계		500	100
비율(%)		100	100%

〈표 4〉 이용자의 검색의도에 따른 인기 검색어 분류

구분	정보획득형		탐색형		트랜잭션형	
	검색어수	퍼센트(%)	검색어수	퍼센트(%)	검색어수	퍼센트(%)
인물	248	49.6	26	5.2	0	0
게임	40	8	0	0	19	3.8
뉴스	53	10.6	0	0	0	0
엔터테인먼트	35	7	0	0	7	1.4
생활	12	2.4	5	1	1	0.2
컴퓨터	9	1.8	1	0.2	4	0.8
사회	2	0.4	11	2.2	0	0
스포츠	5	1	2	0.4	0	0
교육	1	0.2	4	0.8	1	0.2
쇼핑	1	0.2	1	0.2	4	0.8
경제	2	0.4	2	0.4	0	0
문화	1	0.2	2	0.4	0	0
건강	1	0.2	0	0	0	0
합계	410	82	54	10.8	36	7.2
평균	31.54		4.15		2.77	
표준편차	67.37		7.23		5.36	
최고값	248		26		19	
최저값	1		0		0	
비율(%)	82		10.8		7.2	

에 의하여 분류되었다. 그 결과를 살펴보면 <표 4>에 나타난 바와 같이 정보획득형이 82%로 가장 많았으며 이어서 탐색형 10.8%, 트랜잭션형 7.2% 순으로 나타났다. 이런 연구결과는 Kathuria et al.(2010)의 연구결과와 유사한 것으로 나타났다.

주제별로 구체적으로 살펴보면 인물에서는 정보획득형이 49.6%로 가장 많았고 탐색형 5.2%, 트랜잭션형은 하나도 없는 것으로 나타났다. 인물 관련 검색어의 경우에는 연예인이 압도적으로 많았으며 다음으로 사회 정치인, 스포츠 관련 인물 등으로 나타났다. 주로 사회적 이슈관련 인물, 스포츠 인물 등 다양하게 나타났다. 인물에 대한 일반적인 다양한 내용들을 검색하고자 하는 검색어들이 대부분인 것으로 나타나서 사회 이슈를 분석하는데 도움이 되는 것으로 나타났다. 게임의 경우 전체 분야 중에서 11.8%를 차지하였다. 이 들 중에서 게임의 경우 8.4%가 정보획득형이며 트랜잭션형은 3.8%로 나타나 다른 주제 분야에 비해 트랜잭션형이 상대적으로 높은 것으로 나타났다. 게임관련 검색어는 크게 두 가지 형태로 분류 되었다. 첫째는 게임 사이트에서 클라이언트 프로그램을 다운로드 받거나 CD로 설치해서 즐기는 게임관련 검색어로 이러한 검색어들은 대체적으로 게임의 스케일이 크고 복잡하며 게임공략법 등의 매뉴얼 필요 할 경우 많으며 아이템거래, 사이버머니 거래 관련 사이트가 다수 존재하며, 게임 관련 정보 사이트도 다수 존재하는 것으로 분석되었다. 두 번

째의 게임관련 검색어로는 게임 사이트에 접속해서 다이렉트로 게임 플레이하는 경우로서 첫 번째 유형에 비해 규모가 작고 단순한 게임인 경우가 여기에 해당된다고 할 수 있다. 이런 유형의 게임은 직관적이기 때문에 공략법이나 육성법과 같은 매뉴얼이 거의 필요 없는 것이며 주로 슈팅게임이나, 레이싱 게임 등이 해당된다. 이런 유형은 주로 트랜잭션형으로 분류되었다. 뉴스 관련 인기 검색어의 경우 정보획득형이 10.6%로 나타났으며 탐색형이나 트랜잭션형은 없었다. 주로 뉴스의 특성상 주로 간단하게 알 수 있는 검색질의어가 많은 것으로 조사되었다. 특히 국내 TV나 라디오 프로그램 관련 검색어가 대부분인 것으로 분석되었다. 엔터테인먼트 관련 검색어의 경우 정보획득형 7%이며 트랜잭션형이 1.4%로 나타났다. 트랜잭션의 경우에는 주로 프로그램 다운로드 관련 검색어가 대부분인 것으로 나타났다. 엔터테인먼트도 즐거움이나 놀이 등의 특성이 강하기 때문에 트랜잭션형이 상대적으로 많은 것으로 보인다. 생활관련 검색어로는 정보획득형이 2.4%, 탐색형 1%, 트랜잭션형이 0.2%로 나타났다. 컴퓨터 관련 검색어의 경우 정보획득형 1.8%, 탐색형 0.2%, 트랜잭션형 0.8%로 나타나 게임과 같이 트랜잭션 형태의 질의어가 상대적으로 다소 높게 나타났다. 그 이외에 사회와 스포츠의 경우 트랜잭션이 없는 것으로 나타났으며 사회의 경우에는 정보획득형이 0.4%이고 탐색형이 2.2%로 탐색형의 검색질의어가 더 많은 것으로 조사되었다. 이는 사회관련

검색어들이 주로 기관 또는 단체관련 검색어로 나타났다. 예를 들면 병무청, 코레일, 대법원 등이 여기에 해당되므로 이들 기관이나 단체 등의 사이트로 이동하기 위한 경우가 대부분이므로 탐색형이 많은 것으로 분석된다. 스포츠의 경우 정보획득형 1%, 탐색형이 0.4%으로 나타났다. 교육의 경우 정보획득형이 0.2%, 탐색형이 0.8%, 트랜잭션형이 0.2%로 나타나 탐색형이 가장 많고 정보획득형과 트랜잭션형이 동일하게 나타났다. 이는 교육관련 검색어 중에 교육기관과 관련된 검색어가 많아서 이들 교육기관에 연결하기 위한 탐색형이 많은 것으로 파악된다. 쇼핑의 경우에는 트랜잭션형 0.8%로 다른 유형의 질의어보다 가장 많은 것으로 나타났는데 이는 쇼핑과 관련해서 구매형 검색어들이 많은 것으로 판단되었다. 정보획득형과 탐색형은 각각 0.2%로 나타났다. 나머지 경제, 문화, 건강 관련 검색어에는 트랜잭션형태는 없었다. 경제의 경우 정보획득형 및 탐색형이 동일하게 0.4%로 나타났는데 주로 금융재테크 관련 검색어들이었다. 문화의 경우에는 정보획득형이 0.2%, 탐색형이 0.4%로 나타나 상대적으로 탐색형이 더 많은 것으로 분석되었다. 마지막으로 건강관련 인기 검색어로는 500개 검색어 중에 한 개만 있었으며 검색어는 “신종플루”였다. 이는 정보획득형으로 분류되었다. 이와 관련된 연구로 Kathuria et al. (2010)의 연구에서 보면 130,00개의 검색엔진의 검색질의어를 분석한 결과 정보획득형은 75%, 탐색형과 트랜잭션형이 각각 12%로 분석되었

다. 이는 본 연구결과와 유사하다고 볼 수 있다. 반면에 Jansen와 Booth(2010)는 20,000개의 검색질의어를 정보획득형, 탐색형, 트랜잭션형으로 나누어서 자동분류를 하였는데 정보획득형은 51.3%, 탐색형은 33.5%, 트랜잭션형은 15.3%로 다른 연구결과들에 비해 정보획득형이 다소 감소하였으나 여전히 가장 높은 퍼센트를 차지하고 있었다. 이 연구에 의하면 탐색형 및 트랜잭션형의 검색어들을 좀 더 구체적으로 분석하였는데 그 결과를 보면 검색의도에 따른 유형별로 살펴보면 탐색형의 경우 비즈니스 관련 질의어는 51.9%, 홀리데이(holidays) 관련 검색어는 50.8%, 조직과 관련된 검색어는 72.1%가 탐색형 질의어들이었다. 또한 트랜잭션 질의어유형과 관련해서는 아직까지 성관련 질의어가 62.3%로 가장 많았고 이어서 쇼핑(35.0%), 컴퓨팅(27.7%), 게임(24.8%) 순이었다. 이상의 결과들을 종합해 보면 정보획득형 검색어가 가장 많고 이어서 탐색형, 트랜잭션형의 순으로 나타남을 알 수 있었다.

참고로 2009년 상위 10개의 검색어만을 분리시켜 분석해본 결과 ‘사회적 이슈검색어’가 최상위 10개 검색어 중 3개를 차지하였다. 남녀 별로 구분해 살펴보면 남성보다 여성이 더 많은 것으로 나타났다. 2009년도는 연예 및 오락과 관련된 검색질의어가 상위에 랭크되었으며 사회 전반의 특징이었던 ‘여성 파워’를 반영하고 있는 것으로 나타났다. 2009년도에는 2008년이나 2007년에 비해 게임관련 검색질의어는 상위 10위에 포함되지 않았다.

이상에서 지난 3년간의 상위 10개 검색질의 어를 분석한 결과 주제에 따라 검색의도가 다르게 나타난 것임을 알 수 있었다. 예컨대, 대부분은 정보획득형의 검색어가 많았지만 게임이나 쇼핑의 경우 트랜잭션형이 많았고 사회 관련 분야 및 교육 분야에서는 탐색형이 많은 것으로 나타났다.

5. 결론

본 연구에서는 국내 포털업체 중의 하나인 네이버 인터넷 정보 검색엔진으로부터 2007년부터 3년 동안 검색어를 네이버 연감으로부터 추출하여 분석함으로써 이용자의 검색어들이 시간이 흐름에 따라 주제별 변화양상을 살펴 보았다. 또한 2009년도 네이버 검색엔진의 상위 500개의 인기 검색어를 이용자의 검색의도에 따라 정보획득형(informational), 탐색형(navigational), 트랜잭션(transactional) 형태로 인기 검색어를 분류하고 인기 검색어에 나타난 특성을 분석하였다. 그 결과를 요약하면 다음과 같다.

첫째 네이버가 제공하는 웹 검색질의어를 15개의 주제 분야에 따라 2007년부터 2009년도 까지 연도 별로 분석한 결과 15개의 주제 분야 중에서 인물과 사람분야가 전체 중에 약 40%를 차지하였다. 다음으로 높은 점유율을 차지하는 주제 분야로는 엔터테인먼트인, 게임, 뉴

스와 생활 등의 순으로 나타났다.

둘째, 3년 동안 계속해서 이용자의 관심사가 증가한 분야는 15개 분야 중에서 경제, 환경, 스포츠, 문화와 예술, 건강, 교육과 학문, 세계와 여행 분야 등 7개 분야였으며 이 중에서도 상대적으로 가장 많이 증가한 주제 분야로는 세계와 여행 분야였으며 이어서 건강관련 분야와 교육과 학문 분야도 많은 증가율을 나타내었다. 이 결과는 포털 이용자들의 관심사 선호경향을 보여 주는 것으로 정치나 사회문제와 같은 다소 딱딱한 뉴스보다는 연예나 스포츠와 같은 감각적이고 흥미 있는 주제를 선호하고 있음을 알 수 있었다. 또한 급상승 검색어를 분석한 결과 역시 해마다 이용자들의 관심사가 다르게 나타남을 알 수 있었다.

셋째 2009년도 네이버에 나타난 상위 500개의 검색질의어들을 15개의 주제별로 분석한 결과 인물관련 검색어가 274(54.8%)개로 타 분야에 비해 압도적으로 가장 많은 관심을 보였다. 이어서 게임(11.8%), 뉴스(10.6%), 엔터테인먼트(8.4%) 순으로 나타났다.

넷째, 2009년도 네이버 인기 검색어 상위 500개의 검색어들을 이용자의 검색의도 또는 검색 목적에 따라 1) 정보획득형, 2) 탐색형, 3) 트랜잭션형으로 분류한 결과 정보획득형이 82%로 가장 많았으며 이어서 탐색형 10.8%, 트랜잭션형 7.2% 순으로 나타났다.

이상의 연구결과는 미래의 정보검색엔진 개발자들의 이용자의 검색의도를 미리 파악하고 예측하여 적합한 문서를 상위에 검색될 수 있게 하는데 도움을 줄 수 있다. 또한 질의어에 내포된 이용자의 검색의도가 주제에 따라 다르게 나타나므로 이는 미래의 주제별 콘텐츠를 구성하고 제공하는 정책수립에도 도움이 될 것으로 보인다. 또한 인기키워드들은 사회이슈나 그 당시의 문화적 특성을 반영하는 것으로 나타났다. 예컨대, 신종플루나 환율문제, 여행 및 문화관련 인기 키워드들은 사회·문화적 특성이나 여행 관심사를 파악하는데 도움이 될 수 있을 것으로 판단된다. 결국 이용자의 질의어에 내포된 이용자의 의도를 파악할 수 있다면 이용자에게 더 만족스러운 양질의 적합한 검색결과를 제공해 줄 수 있을 것이다. 따라서 앞으로 검색목적이 무엇인지, 검색정보 및 검색시 중요한 속성 및 특성이 무엇인지를 분석하는 연구가 활발히 진행되기를 기대한다.

참고문헌

- 박상규 외. 2007. 검색엔진에서 일간 질의어 분포의 정상성에 관한 연구. 『정보관리학회지』, 24(4): 255-265.
- 박소연, 이준호, 김지승. 2005. 클릭 로그에 근거한 네이버 검색질의어의 형태 및 주제 분석. 『한국문헌정보학회지』, 39(1): 266-278.
- 이준호, 박소연, 권혁성. 2003. 질의 로그분석을 통한 네이버 이용자의 검색행태 연구. 『정보관리학회지』, 20(2): 28-41.
- 진범석, 지용구. 2005. 사용자의 검색목적에 포함한 검색엔진 인터페이스 디자인에 관한 연구. 『한국전자거래학회지』, 13(4): 111-124.
- 한국인터넷진흥원. 2006. 『인터넷 디지털 콘텐츠 분석방법에 관한 연구』. 보고서.
- 한국방송통신위원회, 한국인터넷진흥원. 2010. 『2010년 인터넷이용실태조사 보고서』. [cited 2011. 9. 10].
 <<http://isis.kisa.or.kr/board/index.jsp?pageId=040100&bbsId=7&itemId=771&pageIndex=1>>.
- Bates, M. J. 1979. "Information Search Tactics." *Journal of the American Society for Information Science*, 30: 205-214.
- Bates, M. J. 1989. "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." *Online Review*, 13: 407-424.
- Belkin, N. J., C. Cool, W. Croft, and J. P. Callan. 1993. "The Effect of Multiple Query Representations on Information Retrieval System Performance." *In SIGIR 93, Proceedings of the Sixteenth Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*, 339-346. New York: ACM.

- Belkin, N. J., R. N. Oddy, and H. M. Brooks. 1982. "ASK for Information Retrieval: Part, I & II." *Journal of Documentation*, 38: 61-71.
- Belkin, N. J., P. Kantor, E. A. Fox, and J. A. Shaw. 1995. "Combining Evidence of Multiple Query Representations for Information Retrieval." *Information Processing & Management*, 31(3): 431-448.
- Broder, A. 2002. "A Taxonomy of Web Search." *ACM SIGIR Forum*, 36(2): 3-10.
- Dervin, B. 1999. "On Studying Information Seeking Methodologically: The Implications of Connecting Meta Theory to Method." *Information Processing & Management*, 35(6): 727-750.
- Efthimiadis, E. N. 2000. "Interactive Query Expansion: A User-based Evaluation in a Relevance Feedback Environment." *Journal of the American Society of Information Science and Technology*, 51(11): 989-003.
- Ellis, D. 1989. "A Behavioural Approach to Information Retrieval Design." *Journal of Documentation*, 45(3): 171-212.
- Johnson, J. D. 1997. *Cancer-Related Information Seeking*, Hampton Pr.
- Kathuria, A. J., Bernard Jansen, C. Hafernik, and A. Spink. 2010. "Classifying the User Intent of Web Queries Using K-means Clustering." *Internet Research*, 20(5): 563-581.
- Krikelas, J. 1983. "Information Seeking Behaviour: Patterns and Concepts." *Drexel Library Quarterly*, 19(2): 5-20.
- Kuhlthau, C. C. 1991. "Inside the Search Process: Information Seeking from the User's Perspective." *Journal of the American Society for Information Science*, 42(5): 361-371.
- Hsieh-Yee, I. 2001. "Research on Web Search Behavior." *Library & Information Science Research*, 23: 167-185.
- Jansen, B. J., A. Spink, and J. Pedersen. 2005. "A Temporal Comparison of Alta Vista Web Searching." *Journal of the American Society for Information Science and Technology*, 56(6): 559-570.
- Jansen, B. J. and D. Booth. 2010. "Classifying Web Queries by Topic and User Intent." *In Proc. of the 28th International Conference on Human Factors in Computing Systems*: 4285-4290.
- Jansen, B. J., D. Booth, and A. Spink. 2008. "Determining the Informational, Navigational, and Transactional Intent of Web Queries." *Information Processing & Management*, 44(3): 1251-1266.
- Leckie, G. J., K. Pettigrew and C. 1996.

- "Sylvain. Modelling the Information-Seeking of Professionals: a General Model Derived from Research on Engineers, Health Care Professionals and Lawyers." *Library Quarterly*, 66(2): 161-193.
- Liu, H., H. Lieberman, and T. Selker. 2002. "GOOSE: A Goal-Oriented Search Engine with Commonsense." *In Proceedings of AH*. : 253-263.
- NHN. 2008. 『네이버 트렌드 연감 2007』.
- NHN. 2009. 『네이버 트렌드 연감 2008 : 검색어로 읽어보는 대한민국 트렌드』, 324.
- NHN. 2010. 『네이버 트렌드 연감 2009』, 네이버, Seed paper, 302.
- Pirolli, P. 2007. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, Oxford.
- Rose, D. E. and D. Levinson. 2004. "Understanding User Goals in Web Search." *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA.
- Silverstein, C., H. Marais, M. Henzinger, and M. Moricz. 1999. "Analysis of a Very Large Web Search Engine Query Log." *ACM SIGIR Forum*, 33(1): 6-12.
- Spink, A. et al. 2001. "Searching the Web: The Public and Their Queries." *Journal of the American Society for Information Science and Technology*, 52(3): 226-234.
- Wilson, T. D. 1981. "On User Studies and Information Needs." *Journal of Documentation*, 37(1): 3-15.
- Wilson, T. D. 1999. "Models in Information Behaviour Research." *Journal of Documentation*, 55(3): 249-270.