

사용자 검색 의도 추출을 위한 검색로그 분석¹⁾

지혜성*, 류기곤*, 임희석*

*고려대학교 컴퓨터교육과

e-mail: hyesung84@korea.ac.kr

gon0121@korea.ac.kr

limhseok@korea.ac.kr

Search Log Analysis for Extract User's Search Intention

Hye-Sung Ji*, Ki-Gon Lyu*, Heui-Seok Lim*

*Dept of Computer Science Education, Korea University

요 약

본 연구에서는 사용자 검색로그를 분석하여 사용자의 검색 목적에 따라 분류하고 그 안에 내제되어 있는 사용자의 검색 의도를 찾고자 하였다. 분석은 질의어 110개에 대한 검색로그를 기반으로 검색 목적에 따라 Navigational, Informational, Transactional로 분류하였다. 또한, 질의어를 카테고리별로 분류하였으며 각 결과를 가지고 사용자 검색 의도가 내제되었는지에 대하여 분석하였다. 분석 결과 각 질의어에 따른 검색 목적에 따라서 분포는 다르지만 검색 목적에 따른 검색 의도가 3가지 모두 내제되어 있음을 알 수 있었다. 또한, Informational의 경우에는 질의어에 대한 서로 다른 정보가 나타났으며, 질의어 안에서 사용자의 검색 의도가 나타남을 확인할 수 있었다.

1. 서론

인터넷이 발달하기 시작한 1990년대 이후로 많은 사용자들이 인터넷을 이용하여 정보를 획득하였으며, 웹 검색을 위한 정보검색 분야의 연구도 이와 함께 수행되어 왔다. 최근 컴퓨터의 하드웨어적 발전과 더불어서 검색 알고리즘 역시 많은 발전을 거듭하였으며, Page Rank와 같은 좋은 검색 성능을 가지는 검색 알고리즘이 개발되어 현재 서비스되고 있다. 그러나 웹 2.0으로 진화됨과 동시에 웹에서의 정보량은 폭발적으로 증가하였으며, 이러한 현상은 사용자들로 하여금 원하는 정보를 찾는 데 어려움을 주고 있다. 기업용 검색엔진 회사 Convera의 조사에 따르면 검색엔진 사용자 중에서 21%만이 검색 결과에 대해 만족하고 있으며, 나머지 사용자들은 검색 결과에서 원하는 정보를 얻지 못하거나 의도와는 다른 결과로 인하여 검색 결과를 신뢰하지 못하는 것으로 나타났다. 이는 알고리즘의 성능과는 다르게 사용자들이 원하는 정보를 찾는 부분에서는 크게 발전하지 못했음을 의미한다.

따라서 사용자가 원하는 정보를 찾아서 보다 편리하고 효율적으로 제공하는 연구는 검색의 질을 향상시키는데 중요한 연구가 된다. 검색의 질을 향상시키기 위한 다양한 연구 중에서 검색 트랜잭션 로그를 분석하는 연구는 효과적인 콘텐츠 구축과 효율적인 검색엔진 개발에 있어서 중

요한 자료로 활용될 수 있다[1]. 검색 트랜잭션 로그(이하 검색로그)란, 사용자와 검색 시스템 사이의 모든 상호 작용을 기록한 정보로서 사용자가 입력한 질의어와 조회한 문서를 기록한 클릭로그로 구성되어 있으며, 사용자들의 검색 행위에 대한 실질적인 정보를 반영하고 있다. 따라서 질의어에 대한 검색로그 분석을 통하여 질의어에 내포되어 있는 사용자들의 검색 목적이나 의도를 찾을 수 있다.

그러나 기존 검색로그를 이용한 연구에서는 로그를 분석하여 시간의 흐름에 따른 검색의 변화 혹은 검색 행태의 변화에 대한 분석이 주를 이루고 있으며[2], 검색로그 안에 내제되어 있는 사용자의 검색 의도에 대한 연구는 아직까지 미비한 실정이다.

본 연구에서는 사용자의 검색 의도를 추출하기 위하여 실제 검색로그로부터 사용자들의 검색 목적 및 의도가 내포되어 있는지를 분석하고, 검색로그 분석에 대한 방법론을 제시함으로써 검색의 질을 향상시키고 검색의 효율을 높이고자 한다.

2. 관련연구

검색로그를 이용한 분석 및 연구는 예전부터 다양한 분야에서 수행되어 왔다. 이 장에서는 검색 로그를 분석하고 검색 목적 및 의도를 찾고자 한 연구에 초점을 맞추었다. 국내에서는 검색에 대한 질의어 분석이 주로 행해졌으며, 박소연(2011)의 연구에서는 주로 검색로그로부터 검색 행태에 대한 분석이 주를 이루고 있다. 국내에서 검색로그를 이용한 사용자의 검색 목적이나 의도를 찾는 연구는 미비

1) “이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0018014).”

한 실정이다.

국외에서 검색로그를 분석하여 사용자의 검색 목적을 찾고자 한 연구들로는 Broder 등의 연구를 들 수 있다. Broder(2002)는 질의어 안에서 사용자들의 검색 목적을 찾고자 하였으며, 질의로그를 분석하여 질의어에 따라서 특정 사이트 방문을 목적으로 가지는 Navigational 질의어, 질의어에 대한 정보를 목적으로 가지는 Informational 질의어, 질의어에 대한 상품과 같은 자원을 얻고자 하는 Transactional 질의어로 분류하였다[3]. Daniel E.와 Danny L.(2004)은 [3]의 연구를 기반으로 AltaVista의 질의로그를 이용하여 분석하였으며, 그 안에 내포되어 있는 검색 목적을 분류하였다[4]. Ricardo et al.(2006)은 질의로그를 이용하여 질의어의 목적에 따라 카테고리별로 분류하였으며, 각 카테고리에 속한 질의어들을 Informational, Not-informational, Ambiguous 3가지로 분류하였다[5].

위 연구들은 모두 질의로그를 이용하여 사용자들의 검색 목적을 찾고자 하였으며, 질의어에 따른 검색 목적을 분류하였다. 그러나 사용자들에 따라서 같은 검색어라도 검색의 목적이나 의도가 다를 수 있다는 점을 반영하지 못하였다는 점에서 한계가 있다고 볼 수 있다. 또한, Informational의 경우에는 같은 질의어 안에서도 다양한 정보가 나타날 수 있다. 예를 들어서 “애플”이라는 질의어에는 IT업계에서의 애플사를 나타낼 수도 있지만, 수순한 의미에서의 사과를 나타내는 경우도 함께 존재할 것이다.

사용자의 검색로그를 분석한 국내·외 연구들은 대체적으로 질의어의 주제 분석이나 사용자가 입력한 질의어에 대한 목적이 주된 연구 대상이었으며, 사용자의 검색 의도를 찾기 위한 연구는 드문 실정이다. 이에 본 연구에서는 검색로그를 분석하여 그 안에 내포되어 있는 사용자들의 검색 의도를 찾는 방법론에 대하여 논하고자 한다.

3. 연구 방법

본 연구에서는 국내 검색 포털 중 하나인 네이버에서 입력된 검색로그를 이용하여 사용자의 검색 의도가 로그 안에 내포되어 있는지에 대하여 분석하였다. 검색로그는 각 포털에서도 중요하게 관리되고 있는 데이터이기 때문에 연구를 위한 검색로그를 다량으로 확보한다면 보다 정확한 연구 및 분석에 용이할 것이다.

본 연구에서는 2009년 5월1일부터 30일까지의 검색된 질의어 중에서 가장 많이 검색된 상위 110개의 질의어에 대한 검색로그를 가지고 분석하였다. 분석은 크게 2가지의 목적을 가지고 실시하였다. 첫 번째로 질의어 안에서 내포되고 있는 검색 목적이나 의도가 검색로그 안에서 실질적으로 존재하는지에 대한 분석을 실시하였으며, 두 번째로 검색 목적에 따라서 질의어의 검색 의도가 어떻게 달라지는가에 대한 분석을 실시하였다.

분석을 위해서 제공받은 110개의 질의어에 대한 검색 목적을 분류하였다. 질의어에 대한 검색 목적을 분석하기 위하여 [3]에서 제시한 3가지 검색 목적에 따른 분류를 통하

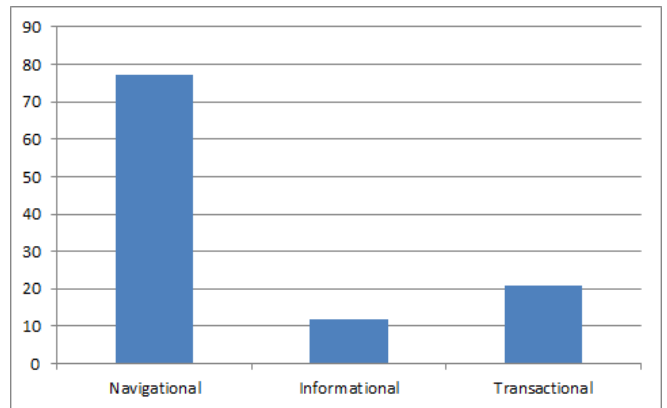
여 질의어를 수동으로 분류하였으며, [5]에서 제시한 카테고리별 분류를 통하여 질의어가 속하는 정보가 어디 부분인지에 대하여 분류하였다.

두 번째로 검색로그에서 하나의 질의어에 대한 다양한 검색 의도의 분석을 위해서 각 질의어에 대한 클릭로그를 분석하였다. 질의어에 대한 클릭로그 분포와 내용 분석은 사용자가 입력한 질의어에 대한 실질적으로 원하는 정보가 무엇이었는지에 대하여 분석할 수 있다. 본 연구에서는 개별 질의어에 대해 사용자들이 조회한 URL, 조회한 문서의 클릭빈도 정보를 이용하여 실질적으로 사용자들이 질의어에 대하여 얻고자 한 정보가 무엇인지에 대하여 분석하였다.

4. 연구 결과

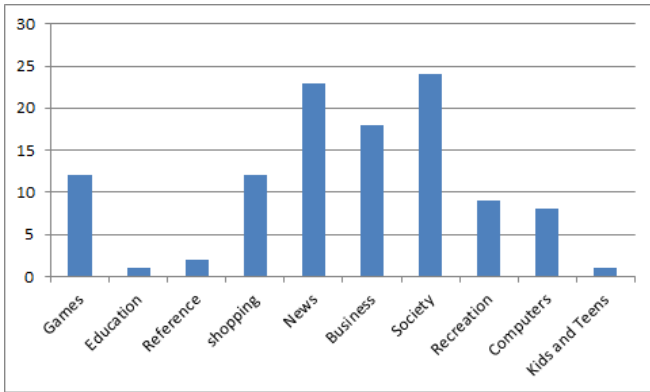
<표 1>은 분석된 검색로그에서 질의어에 대한 검색 목적 분석 결과를 나타낸다.

<표 1> 질의어 검색 목적 분석 결과



질의어 110개 중에서 약 70%의 질의어의 검색 목적이 ‘네이버’, ‘다음’, ‘싸이월드’ 등과 같이 직접적으로 사이트 방문을 목적으로 하는 Navigational 검색 목적으로 나타났으며, ‘아내의 유혹’ 등과 같은 드라마 제목이나 ‘김연아’와 같은 인물과 같이 질의어에 대한 정보를 목적으로 하는 Informational 검색 목적이 약 10% 정도로 나타났다. 약 20%가 ‘네이트온’과 같이 특정한 프로그램 이름 및 제품 이름을 질의어로 하여 이를 얻는 것을 목적으로 하는 Transactional 검색 목적을 지니는 것으로 나타났다. 또한, 분석하는 과정에서 Navigational의 검색 목적을 가지는 것처럼 보이지만, 실제의 사용자들의 목적이 Transactional인 질의어인 경우에는 Transactional 검색 목적으로 분류하였다. 예를 들어서 ‘육선’, ‘이베이’와 같은 질의어는 각 질의어에 대한 사이트를 찾아가는 것을 목적으로 하는 질의어로 보이지만 실제 사용자의 목적은 상품 검색 및 구입과 같은 Transactional 목적을 지니고 있음을 알 수 있다.

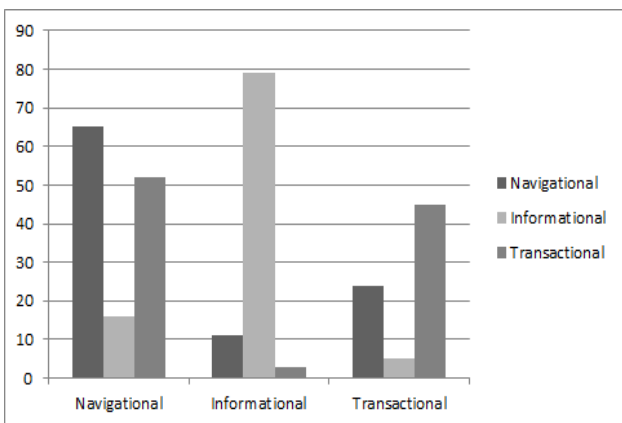
<표 2> 질의어 카테고리별 분류 결과



<표 2>는 질의어에 대한 카테고리별 분류 결과를 나타낸 것이다. 분류는 ODP²⁾에서 선정한 카테고리를 기반으로(Arts, Games, Kids and Teens, Reference, Shopping, World, Business, Health, News, Society, Computers, Home, Recreation, Science, Sports) 수동으로 분류 하였다. 총 질의어 110개 중에서 나타나지 않은 카테고리를 제외한 나머지 10개의 카테고리 중에서 Society가 가장 많은 비중을 차지하였으며, News와 Business등이 그 다음으로 많이 나타남을 볼 수 있다. Society 분야가 가장 많이 나타나는 이유는 분류과정에서 포털 사이트에 관련된 질의어가 많았으며, 이를 Society 분야로 분류하였기 때문이다. 따라서 사용자들이 포털 사이트를 통하여 제공받는 정보를 많이 이용하거나, 블로그, 카페 등을 가입하여 그 안에서 다양한 활동하기 때문으로 분석할 수 있을 것이다. 본 연구에서는 질의어에 따른 검색 목적과 카테고리 분류를 통하여 나타난 결과를 토대로 검색의 목적이 아니라 질의어 안에 내포되어 있는 실제 검색 의도가 검색로그 안에 내포되어 있는지 그리고 어떻게 검색 의도를 추출해야 하는지에 대하여 분석하였다.

<표 3>은 <표 1>에서 나타난 검색 목적별로 분류된 질의어의 클릭 로그를 분석한 결과이다.

<표 3> 질의어 카테고리별 분류 결과



2) Open Directory Project <http://dmoz.org>

위의 결과에서 나타난 것과 같이 3가지 검색 목적 안에서도 분포에 따라 다르지만 모두 동일하게 3가지 검색 목적이 내포되어 있음을 알 수 있다.

Navigational 검색 목적의 경우에는 질의어에 따라서 많은 차이가 나타났다. Navigational 검색 의도를 그대로 지니고 있는 질의어도 존재하지만, 질의어 ‘아이온’의 경우에는 ‘아이온’이라는 온라인 게임을 하기 위한 게임 사이트 방문을 의도로 하는 것(Navigational)과 게임 공략과 같은 정보를 얻기 위한 의도(Informational), 게임 아이템 및 머니를 구입하기 위한 매매 의도(Transactional)가 모두 나타났다.

Informational 검색 목적의 경우에는 대부분의 질의어가 정보를 얻기 위한 의도이기 때문에 URL 분석 결과에서도 모두 Informational 검색 의도가 나타났다. 그러나 질의어에 대한 정보가 다양하게 나타나는 특징을 보여주었다. 질의어 ‘김연아’의 경우에는 피겨스케이팅과 관련된 스포츠 정보와 CF 및 TV 출연과 같은 연예 정보 등으로 나타났다. 이는 사용자에게 따라 같은 질의어 안에서 얻고자 하는 정보가 틀림을 의미하고 있으며, 각각의 얻고자 한 정보들은 모두 사용자 검색 의도로 볼 수 있다.

Transactional 검색 목적의 경우에는 ‘롯데닷컴’과 같은 홈쇼핑 사이트나 ‘옥션’, ‘g마켓’과 같은 인터넷 쇼핑 사이트나 ‘네이트온’과 같이 프로그램이나 파일 이름을 질의어로 사용하고 있다. URL 분석 결과, ‘네이트온’과 같이 원하는 목적이 프로그램이나 파일 명과 같은 경우에는 주로 Transactional 검색 의도가 대부분임에 반하여, 쇼핑 관련 질의어에서는 Navigational의 검색 의도도 많이 나타나고 있었다. 또한, Transactional 검색 목적의 경우에는 ‘다운’, ‘다운로드’, ‘구입’과 같이 Transactional에 맞는 특정 키워드들이 자주 등장하였다. 따라서 질의어에 따라서 사용자에게 URL 자체를 제공하는 경우도 있지만, 질의어와 함께 나타나는 Transactional 키워드를 분석하여 이에 맞는 결과를 제공하는 것도 하나의 방법이 될 것이다.

본 연구에서는 검색로그를 분석한 결과 질의어에 따라서 검색 목적을 3가지로 분류하였으며, 결과적으로는 질의어에 따라 분포는 다르지만, 3가지 검색 목적을 모두 가지고 있음을 알 수 있었다. 따라서 질의어를 분석하여 검색 목적에 맞는 검색 의도를 제공하는 것이 검색의 질을 향상시킬 수 있는 방법이 될 것이다.

4. 결론

본 연구에서는 검색의 질을 향상시키고 검색의 효율을 높이기 위하여 검색로그를 분석하고, 그 안에 내제되어 있는 사용자의 검색 의도를 찾고자 하였다. 검색 목적은 [3]에서 분류한 방식을 사용하였다. Navigational 검색 목적은 질의어에 대한 사이트 방문을 주요 목적으로 하고 있기 때문에 질의어에 대한 사이트를 제공하는 것 자체가 검색 의도로 볼 수 있으며, Informational 검색 목적은 질의어에 대한 정보가 목적을 이루고 있다. 질의어에 따라서

여러 가지 다른 정보들로 구성되어 있으며 각각의 정보들은 사용자가 원하는 검색 의도로 볼 수 있다. 따라서 클릭 로그의 URL에서 나타나는 정보들을 분석하고 분류 및 클러스터링 방법으로 각각 나누어서 보다 편리하게 사용자들에게 정보를 제공할 수 있을 것이다. 마지막으로 Transactional 검색 목적은 질의어에 따라서 다르게 나타나고 있으나, 주로 Navigational 목적과 유사하게 겹쳐서 나타나는 경우가 많이 존재한다. 또한, URL 분석을 통하여 Transactional에서 주로 나타나는 키워드를 분석하여 이를 분류할 수도 있을 것이다.

차후 연구에서는 자동으로 검색로그를 분석하고 분류하는 시스템을 개발하여 이를 실제로 사용자에게 제공하고 얼마나 효과가 있는가에 대한 검증이 이루어져야 할 것이다. 또한, 유사한 질의어들이 많이 존재하기 때문에 이를 어떻게 처리할 것인가에 대한 연구도 필요로 할 것이다.

참고문헌

- [1] 클릭 로그에 근거한 네이버 검색 질의의 형태 및 주제 분석, 한국문헌정보학회지, 제39권, 제1호, 2005.
- [2] 웹 검색 행태의 추이 및 변화 분석, 한국문헌정보학회지, 제45권, 제1호, 2011.
- [3] Broder A, A taxonomy of web search, SIGIR, Fall 2002, Vol 36, No 2.
- [4] Daniel E. Rose, Danny Levinson, Understanding User Goals in Web Search, Proceedings of the 13th international conference World Wide Web, May, 2004
- [5] Ricard Baeza-Yates, Liliana Calderon-Benavides, Cristina Gonzalez-Caro, The Intention Behind Web Queries, SPIRE 2006, pp 98-109, 2006