웹검색 행태 연구*

- 사용자가 스스로 쿼리를 뭉치는 방법으로 -

Web Search Behavior Analysis Based on the Self-bundling Query Method

이 중 식(Joong-Seek Lee)**

목 차

초 록

검색이 편재화 되고 있다. 사용자들은 PC를 너머 스마트폰과 스마트TV에서도 검색을 일상적으로 사용하고 있다. 따라서 사용자의 검색행태도 진화 중이다. 하지만 검색행태 연구는 서버의 트랜잭션 로그(transaction log)를 기반으로 하거나 사용자 로그(user log)를 관찰하는 경우에도 개별 쿼리(query instance)를 분석단위로 삼기에 여러 매체와 여러 시간을 가로지르는 검색 행태를 분석하기에 부족하다. 본 연구에서는 사용자가 직접 덩어리 지운 쿼리 뭉치(bundled query)를 살펴보아 시간과 매체를 가로지르며 궁금증을 해결해 나가는 사용자의 검색행동을 분석해 보았다. 연구를 위해 사용자 PC에 웹로그 캐처를 설치하고, 취합된 웹검색 기록을 사용자들이 직접 덩어리 지워 같은 궁금증을 가진 뭉치를 만들도록 하였다. 또한 각 뭉치에 대한 설문을 통해 검색의 동기, 계기, 만족도 및 검색 후 활동을 조사하였다. 사용자에 의해 만들어진 뭉치는 전화 인터뷰를 통해 검증하였고 맥락을 확인하였다. 뭉치를 통한 인터뷰는 검색 당시의 기억을 떠올리는 힌트로 작용하여 사용자의 검색 회상을 생생하게 하였다. 분석 결과 사용자들은 하루에 평균 4.75개의 검색 뭉치를 발생시키고, 각각의 검색 뭉치는 평균 2.75개의 쿼리로 구성되어 있음을 확인할 수 있었다. 또한 뭉치 내 쿼리의 발전을 '쿼리의 정교화'와 '주제의 정교화'라는 상위 범주 아래 9개의 패턴으로 확인하였다.

ABSTRACT

Web search behavior has evolved. People now search using many diverse information devices in various situations. To monitor these scattered and shifting search patterns, an improved way of learning and analysis are needed. Traditional web search studies relied on the server transaction logs and single query instance analysis. Since people use multiple smart devices and their searching occurs intermittently through a day, a bundled query research could look at the whole context as well as penetrating search needs. To observe and analyze bundled queries, we developed a proprietary research software set including a log catcher, query bundling tool, and bundle monitoring tool. In this system, users' daily search logs are sent to our analytic server, every night the users need to log on our bundling tool to package his/her queries, a built in web survey collects additional data, and our researcher performs deep interviews on a weekly basis. Out of 90 participants in the study, it was found that a normal user generates on average 4.75 query bundles a day, and each bundle contains 2.75 queries. Query bundles were categorized by; Query refinement vs. Topic refinement and 9 different sub-categories.

키워드: 검색, 뭉치, 세션, RTA, 정교화
Web Search, Bundled Query, Session, RTA, Categorization

---

* 본 연구는 서울대학교 융합대학원과 다음커뮤니케이션 UXT가 2010년에 진행한 산학프로젝트를 기반으로 한 연구임. 연구에 도움을 주신 다음커뮤니케이션에 감사드립니다.

** 서울대학교 융합과학기술대학원 차세대융합기술연구원(joonlee8@snu.ac.kr)

논문접수일자: 2011년 4월 15일 최초심사일자: 2011년 4월 19일 게재확정일자: 2011년 5월 13일
한국문헌정보학회지, 45(2): 209-228, 2011. [DOI:10.4275/KSLIS.2011.45.2.209]

1. 연구배경

검색이 편재되고 있다. 그 동안 웹의 전유물로 여겨졌던 검색(search)이 스마트폰에서 IPTV까지 모든 디바이스와 매체에 탑재되고 있다. 다양한 상황에서 검색이 필요한 이유는 매체마다 찾아야 할 데이터가 기하급수적으로 증가한 데서 기인한다. 작은 규모의 데이터에서는 분류(classification)에 의존한 탐색(browse)이 적절한 방법이었지만, 규모조차 파악이 힘든 방대한 라이브러리를 접하게 되면 검색에 의존할 수밖에 없다. 국내 IPTV 3사는 2만~10만개의 콘텐츠 라이브러리[1]가 구축되어 있으며, 애플의 앱스토어에는 현재 30만개의 앱[2]이 등록되어 있고 하루에 1천 개씩 증가하고 있다. 따라서 방대한 라이브러리를 관리하기 위해선 콘텐츠와 메타태그(meta-tag), 그리고 검색의 효율성이 중요해진다. 하지만 다양한 매체에 탑재되는 검색기능은 기술 수준에서 부족(mediocre)하고 중복(redundant)된다. 스마트 기기에는 검색이 기본 기능으로 실려 있지만 스마트 기기로 다운로드 받게 되는 수십 개의 앱(App)에도 자체적인 검색기능이 있다. 예를 들면 전화걸기 앱에도 전화번호 전용의 검색이, 메일앱에도 메일 전용 검색이, 채팅도구에도 채팅용 검색이, 지도앱에도 장소 검색이 별도로 존재한다. 이렇게 검색이 다양한 층위에서 중복 사용되는 현상은 스마트기기에서의 검색이 아직 재매개[3]화(remediation)되지 않았기 때문이다.

사용자의 검색행태도 진화한다. 검색전문가들은 사용자들이 검색도구에 대해 갖는 인식이 지난 10년간 크게 바뀌었다고 얘기한다. 과거에는 검색이 '사이트 파인더(site finder)'의 이미지였다면 지금은 '앤써링 머신(answering machine)'으로 변했다(Nielson 2004). 즉 과거의 검색이 특정 자료를 정확히 인출하기 위한 도구였다면 지금은 궁금증을 풀어나가는 대화창으로 인식된다는 것이다. 검색을 하는 우리의 마음가짐도 훨씬 긍정적이다. 전통적인 검색이 인지적으로 부담되는 작업이었다면, 최근의 검색은 훨씬 더 가망적(potent)인 활동으로 바뀌었다. 즉 머릿속에 떠오른 궁금증을 큰 탈 없이 찾을 수 있었다는 경험이 검색을 긍정적이고 시행착오를 두려워하지 않는 정보 활동으로 바꾼 것이다.

하지만 사용자 중심의 검색 연구가 부족하다. 사용자들의 검색 행태가 전에 없이 전향적으로 바뀌어가는데도 불구하고 검색에 대한 연구는 랭킹시스템의 효율화, 연관 검색어 알고리즘의 개선, 또는 음성 검색 효율화 등 주로 시스템 관련 이슈에 집중되어 있다. 따라서 스마트기기 시대의 사용자들의 검색행동은 어떤지, 또 만족도는 어떠한가를 짚어 볼 필요가 있다. 특히나 모바일 기기가 갖는 이동성과 공간(스크린 사이즈)의 제한에도 인터넷 웹검색 체제를 모바일 기기에 거의 그대로 적용하는 점은 재고해야 할 부분이다. 스마트폰 이용자는 현장에서 맥락이 요구하는 궁금증을 즉시 해결

1) 최중엽 외, 'IPTV 서비스 동향과 발전방향'에서 재인용.
2) 2010년 10월 모바일 전문조사기관 mobclix 자료. [cited 2010.10.16]. 〈Venturebeat.com〉.
3) 재매개는 Jay Bolter와 Richard Grusin이 1999년에 소개한 이론으로 새로운 매체가 등장한 뒤 그 매체만의 고유한 매체미학이 생성되는 과정을 말한다.

하고자 하는데, 이는 책상에 앉아 여러 작업을 병행하며 넓은 키보드와 큰 모니터를 통해 검색하는 경험과는 크게 다를 것이기 때문이다.

로그(log) 기반의 연구가 중요하다. 새로운 검색 행동을 관찰하기 위해서는 기존의 설문이나 직접관찰을 대신해 기기에 쌓이는 로그를 분석하는 시도가 적당한 방법이다. 로그를 통한 분석은 웹 데이터마이닝이 가지는 세 가지 차원의 해석을 가능하게 한다(Liu 2009). 먼저 로그는 사용(usage)에 대한 정보를 알려주어 언제 어떤 기능을 사용자가 소모하는지를 파악할 수 있게 한다. 한편 로그는 내용(content)에 대한 분석을 가능케 하여 어떤 대화가 진행되는지 혹은 활동의 내용이 무엇인지를 범주화할 수 있게 한다. 마지막으로 로그분석을 통해 구조(structure)에 대한 해석을 가능하게 하여 참여자들의 관계망 및 내용간의 상관성을 그려낼 수 있다. 이러한 로그에의 접근은 과거에 비해 훨씬 관대해 졌는데, 트위터와 같은 서비스에서는 일련의 로그를 공개적으로 노출해 이를 바탕으로 다양한 분석 연구가 이루어지고 있다. 예를 들면 트위터의 팔로워 숫자는 영향력의 결정기준으로 무의미하다는 연구(Cha et al. 2010)나 소셜 네트워크상에서 뉴스의 전파과정을 관찰한 연구(Kwak et al. 2011)는 공개적인 로그를 기술적으로 채집하여 분석한 결과들이다. 반면 특정 행동을 관찰하기 위해 기기 내부에 숨겨진 있는 숨겨진 로그(hidden log)를 빼내는 연구와 자체 개발한 로그발생기(log catcher)를 탑재하는 연구들도 최근 증가하고 있다.

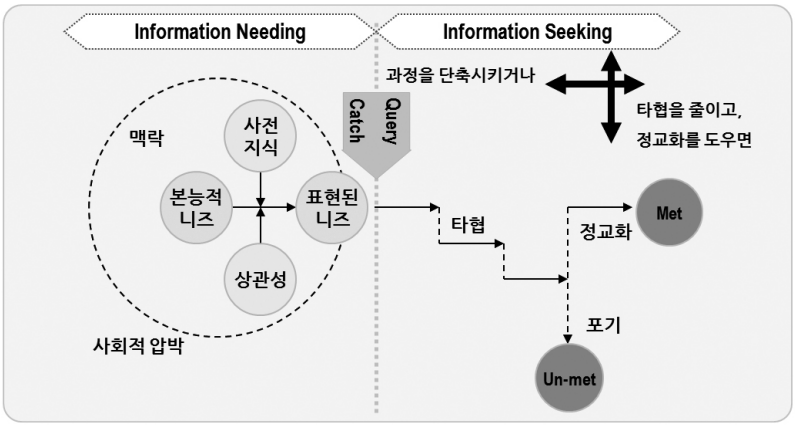
로그는 뭉치기반으로 살펴보아야 한다. 우리가 살펴보려 하는 웹검색은 일회성 활동이 아니라 연속적이며 적극적인 학습과정(active learning process)이다. 즉 하나의 검색은 독립된 것이 아니라 일련의 쿼리들을 통해 문제를 해결해 나가는 과정인 것이다. 보다 발전된 검색 연구를 위해서는, 개개인의 '개별 행동(single instance)'을 분석 단위로 삼기 보단 일련의 행동을 집합화한 '뭉치(bundled instances)'를 분석하는 것이 적당하다고 판단된다. 특히 모바일 환경에서의 사용자들은 하나의 궁금증을 여러 시간대에 나누어 여러 매체를 갈아타며 완성해 나가는 수행적 연속성(performative continuity)을 띄기에 더욱 그러하다. 이러한 현상은 개인의 웹기록을 살펴보면 쉽게 알 수 있는데 〈그림 1〉에서와 같이 컴퓨터에 쌓이는 검색기록을 보면 우리의 검색활동은 띄엄띄엄, 조금씩 정교화되는 덩어리의 형상을 갖고 있음을 알 수 있다. 따라서 검색 뭉치를 채집하여 분석할 수 있다면 정보검색 사용자의 행동에 대해 보다 풍부한 해석이 가능할 것이다.

〈그림 1〉 검색기록의 예

2. 선행연구

2.1 정보검색행동이란 무엇인가?

정보검색행동은 검색엔진을 중심으로 사용자 측면과 시스템적 측면을 나누어 보면 이 연구는 사용자 측면에 초점이 있다. 이러한 연구를 정보추구행동(information seeking behavior) 연구라 하는데 우리가 의식적으로 정보를 찾아 개인이 느낀 정보 요구(information need)를 채워나가는 활동에 대한 연구를 말한다. 부족한 정보를 채우기 위해선 전통적으로 비계획적인 브라우징(unplanned browsing)의 방법이나 계획적인 검색(planned search)의 방법이 사용되어 왔다. 하지만 정보시스템이 등장함에 따라 기계적 검색(mechanical search)이 정보추구활동의 주 활동이 되었다(Chang & Rice 1993). 전통적인 검색이든 기계적 검색이든 인간은 검색을 위해 '쿼리'를 발생시켜야 하는데 여기에는 몇 단계가 필요하다(Taylor 1962). 검색은 일종의 불편함에서 출발한다. 정보 요구의 '본능적 불편함(visceral needs)'은 지식의 갭, 또는 새로운 상황의 출현 등 여러 이유에서 시작된다. 이런 불편이 '의식(conscious)적 요구'로 자각되면 우리는 이를 쿼리로 '형식화(formalize)'하게 된다. 형식화의 단계는 개인의 교육수준에 따라 편차가 많기도 하지만 하나의 정보요구를 언어화한다는 측면에서도 인지적으로 꽤나 어려운 단계이다. 이렇게 도출된 쿼리는 검색시스템에 입력되는데 이 과정에서 검색시스템의 통제어들과 사용자가 형성한 쿼리가 '타협(compromise)'하는 과정을 거치게 된다. 이러한 과정을 〈그림 2〉와 같이 재개념화 하여 정리할 수 있는데, 실제로 많은 검색에서 사용자는 이 타협과정을 통해 원래의 요구에 채 못 미치는 결과(unmet-needs)를 얻게 된다. 따라서 검색과정을 만족스럽게 만들기 위해서는 1) 검색 과정을 단축시키거나, 2) 타협을 효율화함으로써 개선할 수 있다. 검색포털에서 발견되는 자동완성 기능은 쿼리의 도출을 쉽게 하며, 연관 검색어는 타협을 효율화시키는 검색보조기술들이다.



〈그림 2〉 Taylor의 검색과정의 도해 재개념화

2.2 과정으로서의 검색

80년대의 검색연구는 개별 검색의 효율성에 초점이 맞추어졌다. 베이츠(Bates 1993)는 검색이 하나의 질문에 하나의 답을 얻는 방식이 아니라 하나의 목표(goal)를 찾아 꼬리의 꼬리를 물고 진행되는 산딸기 찾기(berry picking)와 같다는 점을 증명했다. 이 모델은 기존의 연구와 네 가지 측면에서 차별성을 갖는데 1) 현실에서의 검색은 검색과정 중에 쿼리가 변화되고 진화되는 유동적인 행위라는 점이다. 2) 검색은 검색의 징검다리마다 사용자가 쓸모 있다고 생각되는 조각을 주어 모은 여러 정보들을 합쳐서 만족된다는 점이다. 3) 검색 과정에서 사용자는 다양한 탐색 테크닉을 활용하게 되는데, 어떨 때는 각주의 꼬리를 찾고, 어떨 때는 저자 중심으로 관점을 전환하고, 어떨 때는 결과 비교를 사용하기도 한다. 4) 검색과정에서 우리가 생각하는 것보다 훨씬 더 다양한 정보원(domain)을 넘나들며 정보를 찾는다는 점이다. 이 연구를 통해 정보 검색은 적극적 정보학습과정(active learning process)이라는 관점도 도출되었다. 현재의 복합화된 미디어 상황에서 사용자는 분명히 하나의 목표를 위해 다양한 미디어를 갈아타며 사용하기에 베리피킹 모델은 정보연구를 떠나 일반 미디어 연구로 확장이 가능한 이론이라고 볼 수 있다. 예를 들어, 인기 드라마 최종회의 결말이 궁금해진 사용자들은 신문기사, 작가 인터뷰, 네티즌들의 추측, 그리고 방송을 통해 그 내용을 탐색하며 관련정보들을 입체적으로 찾아가고 있다.

2.3 트랜잭션 로그(transaction log) 분석

사용자의 웹행동을 관찰하기 위해서 크게 두 가지 방법이 가능하다. 첫째 검색서버에서 쿼리의 기록을 분석하는 것이다. 검색엔진에서 수집한 데이터의 분석을 트랜잭션 로그 분석(transaction log analysis)이라 하는데 1999년 알타비스타(Alta-Vista) 이용자들이 6주간 남긴 2억 8천 5백만 개의 이용자 세션, 9억 9천만개의 쿼리를 분석한 연구(Silverstein 1999)가 트랜젝션 로그 분석의 시초라 할 수 있다. 이는 공급자단에 유용한 연구 방식으로 쿼리의 발생 빈도, 평균적인 쿼리의 길이, 전형적인 쿼리 세션의 발견 및 쿼리간의 연관성 분석이 가능하다. 이러한 데이터는 웹 검색 전략을 형성하는데 사용되기도 하고 검색쿼리들이 어떻게 사용자의 검색행동으로 바뀌었는지를 확인하는데도 사용되었다. 예를 들면 한 세션 내에서 검색어의 숫자나 쿼리를 변경한 숫자가 검색어 재구성에 어떤 영향을 미쳤는지 등의 분석이 가능했다. 같은 서버의 통시간적 비교도 가능한데 2002년에 알타비스타에서 생성된 100만개의 쿼리 주제를 분류한 실버스타인의 1999년 연구(Silverstein 1999)에 비해 2006년에는 쿼리의 주제가 다양해지고 엔터테인먼트성 쿼리가 증가함을 다른 연구에서 발견하기도 했다(Jansen 2006). 하지만 모든 사용자들이 검색 전에 로그인을 하지 않기에 쿠기를 기반으로 한 서버사이드(server-side) 연구들은 특정 개인을 추적하기 어려운 한계가 있다. 한편 검색 엔진으로부터 쿼리 세션을 구분하고 그 내부의 룰을 찾아내는 연구도 진행되었는데 이러한 연구는 연관 검색어를 도출하는 규칙을 만드는데 사용되었다(Shi & Yang 2007).

연관 검색어 연구는 검색엔진이 가진 주제 분류 간의 관계성을 도출하는 목표를 가지고 있다. 즉 사용자들이 입력한 키워드의 연관보다는 제공되는 검색결과의 연관성을 가시화하는데 목표가 있다. 또 다른 연구에선 검색 엔진에서 사용자의 세션을 분리해내고 검색어를 통해 검색 동기를 추론해 보는 연구도 진행되었다(Rose & Levinson 2004).

2.4 사용자 로그 중심의 연구

검색회사가 아닌 이상 서버사이드의 데이터에 접근하기는 상당히 힘들다. 따라서 사용자의 검색로그를 바탕으로 검색행동을 분석하는 연구가 최근 많이 이루어지고 있다. 피실험자의 컴퓨터에 WebTracker라는 툴을 설치하여 사용자의 URL 요청과 응답을 수집하는 연구(Choo et al. 1999)는 정량적 데이터와 심층 인터뷰를 병행하여 검색 동기와 만족도를 파악했다. 반면 자연스런 환경에서 검색 행동과 맥락을 알아보기 위해 웹로그(web-log)가 수집되는 노트북을 제공하고 학생들에게 검색활동을 하도록 한 뒤 지속적으로 사용자의 웹브라우저 사용방식에 대한 로그를 수집한 연구는 그 결과로 학생들의 정보검색 동기가 사실확인(fact finding), 정보 수집(information gathering), 둘러보기(browsing), 자료 처리(transacting), 기타 등으로 나뉘어진다는 결론을 제시하기도 했다(Keller et al. 2007).

2.5 검색 뭉치 분석

컴퓨터공학에서 사용되는 세션(session) 개념은 정보교환의 효율을 위해 두 컴퓨터가 1회 이상의 통신을 지속할 때 서로의 연결상태를 준영구적(semi-permanent) 단계로 끌어올림을 말한다. 이렇게 되면 세션 내의 통신은 서로를 인지하는 단계를 건너 뛰어 무척 빠르게 진행된다. 이러한 세션은 시간적 개념을 동반하며 시작점과 종료점을 갖게 된다. 따라서 '세션 분석(session analysis)'이라 함은 일련의 목적을 가지고 연속적으로 진행한 활동의 분석을 말한다. 검색 연구에서도 트랜잭션 로그를 통한 세션 분석의 시도가 많이 있다. 트랜잭션 로그로 세션을 정의하는 방법은 서버가 할당한 이용자 식별자(user identifier)를 쿠키에 넣어 검색을 요청한 브라우져에 보내 구성된다. 쿠키를 통해 세션이 얼만큼 지속되는가를 측정하고 쿼리를 모으는 방법(Spink et al. 2001)은 특정 컴퓨터에 서로 다른 사용자 또는 다른 목적의 검색이 발생하면 분석이 어려운 단점이 있다. 이를 보완하기 위해 검색을 요청한 컴퓨터에 5분 후 자동 소멸하는 쿠키를 심는 방법이 고안되었는데(Silverstein et al. 1999), 5분 안에 검색이 계속 발생하면 쿠키의 소멸을 5분씩 연장하여 시간적 설정이 갖는 우려를 보완하기도 했다. 국내에서는 네이버의 로그 분석을 통해 한 명의 이용자가 단일한 정보 요구를 지니고 처음 디렉토리에 접근했을 때부터 디렉토리 접근이 종료될 때까지를 한 세션으로 분석한 사례가 있다(박소연 et al. 2004).

트랜잭션 로그를 이용해 규칙기반으로 쿼리의 덩어리를 형성하는 '세션'과 달리 본 연구에서는 사용자가 직접 쿼리의 덩어리를 만들게 하였다. 이는 세션과 개념적으로 대별되기에 '뭉치(bundle)'라 부르기로 하였다. 사용자에 의해

덩어리지어지는 뭉치는 규칙기반의 세션에 비해 다음과 같은 장점을 갖는다. 먼저 뭉치는 주인이 누구인지 명확하다. 일반적으로 세션은 트랜잭션 로그를 기반으로 하기에 개인을 식별하기보단 정보기기를 식별한다. 쿼리 정보와 사용자의 정보를 모두 알게 되면 추가적으로 분석할 수 있는 내용이 많아진다. 또한 세션은 규칙기반으로 측정되기에 아주 긴 시간 또는 간헐적으로 지속되는 관심을 측정하기 어렵다. 반면 뭉치는 사용자가 관여하여 의미적으로 묶어나가기에 시간적으로 인접하지 않는 궁금증을 관찰할 수 있다. 또한 향후에 실험을 확대하여 여러대의 PC와 모바일 장치의 쿼리를 뭉치게 되면 기기간을 오가며 검색하는 관심도 분석할 수 있는 여지가 있다.

3. 연구문제

이 연구는 정보장치의 고도화, 검색의 보편화, 그리고 검색행동의 진화 속에 최근 사용자들의 검색활동은 어떤 모습을 갖는지를 살펴보는데 목표를 두고 있다. 이를 위해 기존 연구와는 다른 접근을 시도하려 하는데, 첫째는 사용자 측면에서 발생하는 검색 로그를 적극적으로 채집하여 활용하고, 둘째로는 관찰의 단위를 개별 행동(single instance)이 아닌 집합적 뭉치 활동(bundled instances)으로 확대하여 개별행동이 놓치기 쉬운 맥락성을 확보해 보려 한다. 이런 전제 조건에서 다음과 같은 연구 문제를 제시할 수 있다.

- 연구문제 1: 뭉치 기반의 웹검색 활동은 어떠한 모습을 갖는가?
  쿼리 뭉치들의 통계적 특성은 무엇인가? 뭉치는 몇 개의 쿼리들이 있으며, 궁금증의 평균 지속시간은 얼마인가? 개인은 하루에 몇 개의 뭉치를 발생하는가?
- 연구문제 2: 검색 뭉치의 외부적 특성은 무엇인가?
  검색 뭉치가 발생하게 된 동기와 계기는 무엇인가? 검색과정에서 다른 도구들은 어떻게 사용되는가? 어떠한 검색보조기술이 사용되는가?
- 연구문제 3: 검색 뭉치의 내부적 특성은 무엇인가?
  뭉치 내의 쿼리들은 어떤 발전 과정을 거치는가? 정교화에는 어떤 패터닝이 가능한가? 정교화 패턴의 분포와 상호적 관계는 어떠한가?

이 연구는 연구문제에 대한 실증적 해답을 얻는 목적 이외에도 새로운 관찰 방법의 도입을 통해 복합화된 점보검색 활동을 조사하는 뭉치 기반의 연구방법이 갖는 방법적 문제점을 검토하는데도 의미를 두고 진행하였다.

4. 연구방법

4.1 파일럿, 표집, 그리고 조작적 정의

먼저 20대~40대의 8명을 대상으로 6일간 파일럿 실험을 진행하였다. 학생:직장인, 남:녀 비율을 각각 50:50으로 표집한 파일럿 집단을 통해 자체적으로 개발한 실험툴과 설문을 검증

하였다. 이를 바탕으로 개선된 툴을 사용해 진행된 본 실험에서는 인구통계학적 기준에 의거해 나이, 성별, 직업군을 고려한 20~30대 남녀 100명을 모집하였다. 연구진이 개발한 분석도구[4]의 특성상 표집은 모두 마이크로소프트 윈도우 환경에서 인터넷익스플로러를 주 브라우저로 사용하는 집단으로 한정하였다. 표집은 전문 리쿠르팅 회사의 도움을 받았다. 중도 탈락한 8명을 제외한 92명이 실험을 완료하였고 이들의 구성비는 〈표 1〉과 같다. 동영상으로 제작된 매뉴얼을 통해 '검색 쿼리 뭉치기'의 방법을 설명했고 이틀 간의 도구 숙지 기간을 주었다. 2010년 7월 22일부터 본 실험 데이터가 수집되어 14일간 데이터를 모았고, 1주일에 한 번씩 전화 인터뷰를 통해 정량적인 데이터를 보완하였다.

〈표 1〉 표집 구성

구분	구성비	구분	구성비
남성	45명	20대	50명
여성	47명	30대	42명

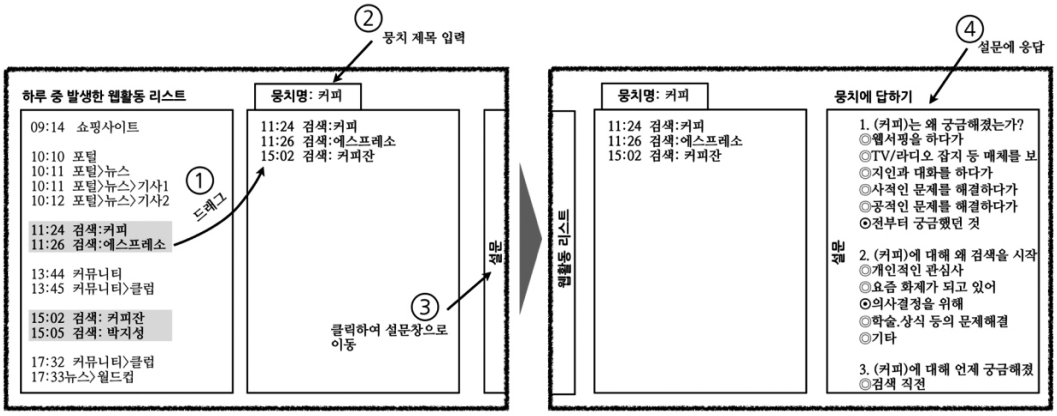
피실험자의 검색어 뭉치기 과제는 일일 단위로 진행되었는데, '하루 동안의 웹 사용량'을 구분하기 위해 '하루'의 단위를 정의할 필요가 생겼다. 파일럿 실험을 통해 대부분의 피실험자가 밤 12시가 무렵에도 활발한 인터넷 사용을 보였기에 활동량이 감소하는 새벽 5시부터 다음날 오전 5시까지를 하루로 정의하였다.

한편, 수집된 데이터는 총 14일치였지만, 피실험자에 따라 검색활동이 아주 없는 날들도 있었다. 뭉치가 발생한 날만이 의미를 가진다고 결정하였다. 한편 사용자가 직접 참여하는 검색 뭉치기는 초기에 집중도가 높으나 일주일이 지나감에 따라 떨어지는 현상이 나타났다. 따라서 피실험자간의 뭉치발생일 편차(7~14일)를 고려해 분석에서는 도구 숙지기간이 끝난 뒤 실험참여의 집중도가 높은 날 중 뭉치가 다수 발생한 7일간의 데이터만 사용하기로 했다.

4.2 데이터 수집

데이터의 수집을 위해 사용자의 동의 하에 피실험자의 PC에 '로그 캐쳐(Log Catcher)'라는 자체 개발한 소프트웨어를 설치하였다. '로그 캐쳐'는 백그라운드로 작동하기에 사용자의 자연스런 컴퓨팅 활동을 방해하지 않는다. 이 도구는 사용자의 모든 웹활동을 일단위로 모아 매일 저녁 분석서버에 모아진 데이터는 연구자뿐 아니라 사용자에게도 보여진다. '쿼리 뭉치기'를 위해 '뭉치기 툴(Query Bundler)'을 제작했다. 이 도구는 크게 세 개의 컬럼으로 구성되어 있으나 사용자가 처음 로그인하면 〈그림 3〉의 좌측처럼 두 개의 컬럼만 먼저 보이게 된다. 좌측에는 하루 종일 발생한 웹활동이 나열되고 그 중 검색활동은 짙은 색으로 구분된다. 사용자는 '관련 있는 검색어'를 모아달라는 요청에 따라 의미있는 관계를 가진 쿼리를 오른편으로 드래그할 수 있는데, 뭉치기가 끝나면 상단에 제목을 입력할 수 있다. 이 과정에서 실수를 하게 되면 쿼리를 원래대로 돌려 놓을 수도 있어 사용자가 걱정 없이 도구를 사용할 수 있었다. 실험 기획 단계에서 쿼리 뭉침을 텍스트 마이닝 기술

4) 로그 캐쳐는 마이크로소프트의 닷넷프레임워크에서 자바로 개발되었으며 쿼리 모니터는 아도비 플래쉬 기반으로 제작되었다.



〈그림 3〉 뭉치기툴(Query Cluster) 다이어그램

을 이용하여 규칙기반(rule-base)으로 구현할까 고민했으나, 인간의 궁금증이 가지는 복잡성과 복합성을 고려해서 사용자 참여 방식을 선택했다. 실험과정에서 사용자에 의한 뭉치기의 완성도는 초기에 60% 수준, 1차 인터뷰 후 피실험자의 학습 및 실험자의 교정에 의해 80%에 도달했다. 향후에는 텍스트마이닝으로 1차 뭉치기를 형성하고 이를 사용자에게 제시하여 교정하는 방식도 생각해 볼 수 있다.

사용자의 쿼리 뭉치기가 끝나면, 우측의 '설문'탭을 눌러 설문창이 있는 세 번째 컬럼을 열 수 있다. 여기서 뭉치에 대한 심화된 설문을 진행하게 된다. 설문의 내용은 〈표 2〉와 같이 구조화되었는데, 크게 검색의 동기와 계기, 그리고

〈표 2〉 설문 항목의 구조화

질문 의도	질문	선택지
검색 동기 분석	( )에 대해 검색하신 동기는 무엇입니까?	1) 업무・과제를 해결하기 위해 2) 최근 화제가 되고 있는 사건, 이슈를 확인하기 위해 3) 잘 모르거나 불확실한 지식에 대한 정보를 얻기 위해 4) 의사결정을 내리는데 필요한 정보를 얻기 위해 5) 단순 호기심을 충족시키기 위해
검색 계기 분석	( )에 대한 검색을 하게 된 계기는 무엇입니까?	1) 웹서핑을 하다가 ① 지인과 대화를 하다가(온・오프라인, 전화, 대면) ② 사적인 문제를 해결하다가(쇼핑, 취미, 구직) ③ 공적인 문제를 해결하다가(업무・과제) ④ TV나 라디오, 잡지 등의 매체를 접하다가 ⑤ 전부터 알고자 했던 정보가 갑자기 떠올라서 ⑥ 기타

질문 의도	질문	선택지
검색 동기 발생 시점	( )에 대한 궁금증은 언제 생기셨습니까?	1) 바로 ① 1~3일 전 ② 일주일 이내 ③ 일주일 이상 ④ 기타
검색 과정 중 병행 여부	( )를 알아가는 과정에서 웹검색 외에 어떤 방법을 사용하셨습니까?	1) 지인 ① 웹 ② 기타 매체 ③ 스마트폰
보조기술 활용 여부	( )를 검색하면서 어떤 검색 보조기능을 사용하셨습니까?	1) 오타수정 기능 ① 자동완성 기능 ② 실시간 검색어 기능 ③ 연관 검색어 기능 ④ 기타
과정의 만족도	( )에 대한 검색 결과는 만족스러웠습니까?	매우 불만족에서 매우 만점 (5점 척도)
추가 검색 진행 여부	( )의 검색이 끝난 후에 인터넷 외에 어떤 수단을 활용하셨습니까?	1) 달성 후 종료 2) 달성 그러나 추가진행 3) 미달성 후 종료 4) 미달성 그러나 추가 진행 5) 기타
결과의 만족도	( )에 대한 검색을 통해, 이루고자 했던 목적에 얼마나 도달하셨습니까?	매우 불만족에서 매우 만점 (5점 척도)

검색 중 병행 및 후행하는 행동, 그리고 과정과 결과의 만족도로 구성되었다. 검색의 동기는 정보학에서 흔히 사용되는 문제의 해결(problem solving), 의사결정(decision making), 사회적 의미짓기(sense-making), 호기심 및 불확실성의 감소(reducing uncertainty)로 구성하였다. 한편 검색의 계기는 내적(internal)요인과 외적(external)으로 구분할 수 있는데, 내적 요인으로는 평소의 궁금증, 사적인 문제, 공적인 문제를, 외적 요인으로는 지인과의 대화, 웹서핑, 타 매체의 자극으로 구성했다. 설문은 파일럿 실험에서 사용된 주관식 문항의 결과를 바탕으로 재조정되었다.

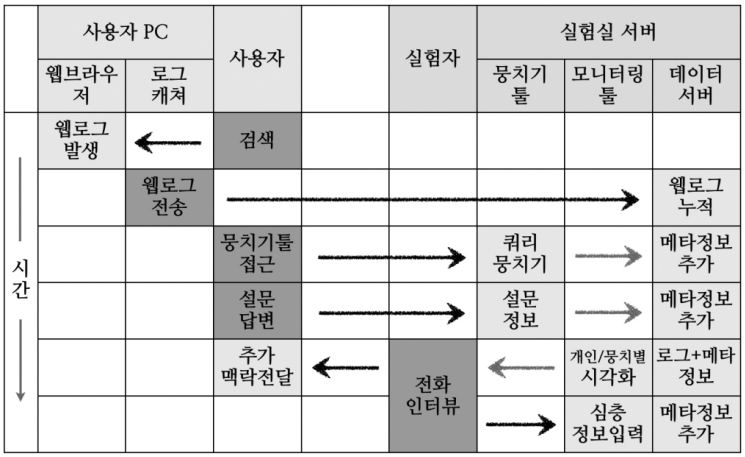
사용자가 매일 쿼리를 뭉치고 설문을 답하게 되면 연구원들은 〈그림 4〉의 모니터링 툴(Monitoring Tool)를 통해 피실험자의 진행상황을 파악할 수 있다. 연구자는 1주에 한 번씩 피실험자들에게 직접 전화를 걸어 1) 뭉치기에 어려움이 없는지, 2) 뭉치 제목에 대한 검증, 3) 설문 중 기타 항목 및 자유답변 사항을 확인하는 심층 인터뷰를 진행하였다. 여기서 뭉치로 대변되는 궁금증이 어떻게 생기고 어떻게 해결되었는지를 질문했다. 이 과정은 평균 30분 정도 진행된다. 〈그림 4〉의 모니터링 툴을 살펴보면 실험의 순서를 살펴볼 수 있다. 먼저 연구자는 1) 실험일과 2) 피실험자를 선택하여 특정 피실험자가 특정일에 형성한 3) 뭉치리스트를 볼 수 있다. 이 중 한 뭉치를 선택하면 4) 뭉치 내 쿼리들이

〈그림 4〉 모니터링툴

나열된다. 연구자는 전체적인 뭉치의 얼개와 5) 웹설문 내용을 살펴본 뒤 전화 인터뷰를 통해 궁금증이 어떻게 생겼고 어떤 과정을 통해 해결되었는지를 6) RTA창에 정리하여 입력한다. 여기서 사용된 회상적 인터뷰 기법(RTA: Retrospective Interview Technique)은 피실험자에게 과거를 기억할 수 있는 단서를 제공함으로써 피실험자가 과거 경험에 대한 회상을 쉽게 할 수 있도록 한다(Baxter 1994). 이 연구에서는 피실험자에게 회상의 단서가 되는 뭉치의 이름과 검색 기록을 제시함으로 그들로 하여금 기억의 인출이 자유 회상(free recall)에서 힌트 회상(hinted recall)수준으로 높여져 보다 생생한 경험을 말할 수 있다.

전체 실험의 흐름을 〈표 3〉으로 정리해 볼 수 있다. 먼저 사용자의 검색이 일어나면 웹로그가 로그캐처에 쌓인다. 이 로그는 지정된 시간에 데이터 서버로 전송되고, 사용자는 매일 뭉치기툴에 접근하여 쿼리를 뭉치고 설문에 답하게 된다. 일주일에 한 번 실험자는 사용자와 전화인터뷰를 진행하고 이 과정에서 모니터링 툴에 시각화된 개인별/뭉치별 정보를 활용하게 된다. 심층인터뷰를 통해 취합된 정보는 데이터 서버에 누적된다.

〈표 3〉 시스템 구성과 정보의 데이터의 흐름



5. 분석 및 결과

5.1 기술통계 분석

먼저, 검색 뭉치들의 기술적 통계를 살펴 보았다. 92명의 피실험자들이 일주일간 발생시킨 총 쿼리는 45,853개였고 총 2,968개의 뭉치가 도출되었다. 이 중 쿼리 1개인 뭉치는 1,306개로 전체 뭉치의 44%를 차지한다. 2001년 Excite의 트랜잭션 로그를 분석한 연구(Jansen et al. 2002)에서 쿼리 1개의 뭉치 비율이 30.8%인 점과 비교하면 본 실험에서는 짧은 뭉치가 더 많이 발견된 셈이다. 피실험자들은 일인당 하루 평균 4.7개의 뭉치를 만들어 냈으며 쿼리 1개로 이루어진 뭉치를 제외하면 한 개의 뭉치는 평균 4.48개의 쿼리로 구성되었다. Excite의 연구에서 뭉치 내 쿼리가 2.84개인 것과 비교하면 우리의 경우 훨씬 뭉치 내부가 촘촘하다고 볼 수 있다.

사용자가 하루 동안 특정 주제에 대한 검색 행동을 얼마나 오래 지속했는가를 측정하는 관심의 지속시간을 계산하기 위해 다음의 공식을 사용했다.

$$CD = t(p_n) - t(p_1)$$

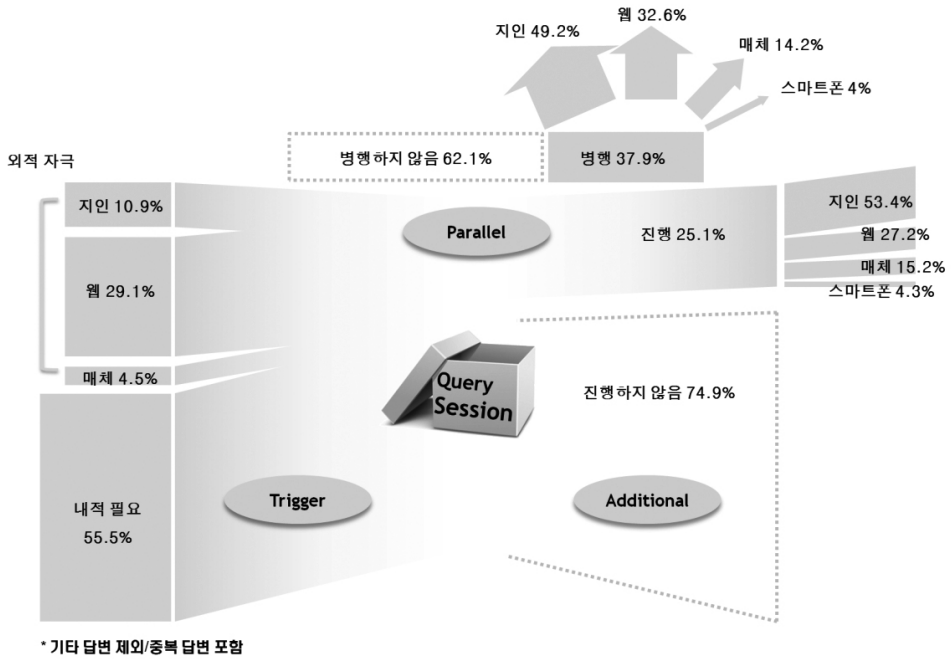
t는 시간, p_n은 한 뭉치에서 마지막 페이지, p_1은 뭉치의 첫 페이지

즉 관심의 지속시간은 뭉치의 마지막 시간에서 시작 시간을 뺀 것으로 계산된다. 이와 같이 뭉치의 마지막 시간에서 첫 시간을 뺀 결과로 수행 시간을 산정하는 방식은 얀센 등(Jansen et al. 2007)이 웹 검색 연구에서 과업 달성 시간을 측정하는 방법을 사용했다. 관심의 지속시간은 한 뭉치를 구성하는 페이지에 머문 시간의 총합이 아니기 때문에 사용자의 다른 웹 행동을 포함한다. 본 연구에서 관심의 지속시간은 최소 0초에서 최대 24시간까지의 분포를 가지며 평균값은 124분으로 나타났지만, 중앙값은 13.4분으로 전체의 70%의 뭉치에서 관심의 지속시

간이 1시간을 넘지 않는 것으로 관찰되었다. 특히 전체 뭉치의 51.3%가 15분 미만의 지속시간을 가졌는데, 이 값은 얀센이 52%의 사용자가 15분 미만의 쿼리 세션 시간을 나타낸다고 연구한 것과 비슷한 결과를 보였다(Jansen 2003).

5.2 뭉치의 외부적 특성

웹설문을 통해 뭉치 행위의 전후 관계를 살펴볼 수 있다. 즉, 검색뭉치는 왜 발생했으며, 검색 중에는 어떤 병행적인 노력을 했고 또한 검색활동은 어떻게 종료되었는지를 〈그림 5〉로 도식화해 보았다. 설문을 통해 검색은 내적 요구에 의한 동인이 55.5%를 차지해 가장 큰 비중을 차지한다. 외적 자극으로는 웹(29.1%), 지인(10.9%), 기타매체(4.5%)가 차지한다. 웹검색에서 웹으로부터 발생하는 이용 동기가 30%에 미치지 않는 이유는 포털 환경에서 생긴 궁금증을 검색이 아닌 브라우징으로 상당 부분 해결되기 때문이라 짐작된다. 이렇게 시작된 검색과정 중 약 38%의 뭉치에서 웹검색 이외의 정보활동을 병행이 나타나고 있다. 38%가 다른 행위와 병행을 한다는 점은 그만큼 검색에서 만족하지 못하거나 해결되지 못한 부분이 존재한다. 흥미로운 점은 전화 또는 대면을 통해 지인(49.2%)에게 묻거나, 검색 이외의 커뮤니티 웹사이트(32.6%) 등을 찾는 활동이 주요 병행활동으로 나타나고 있다. 친구 또는 주변 전문가에게 궁금증을 해소하는 행위는 정보추구행동에서 빈번히 나타나는 행위이며 영향력도 크다. 한편 커뮤니티나 블로그 등을 통해 검색에서 해결되지 않는 문제를 해결해나가는 점은 익



〈그림 5〉 뭉치의 계기, 병행, 후행 활동

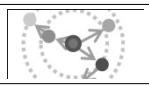
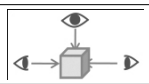
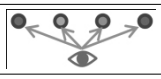
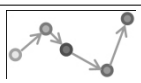
숙한 정보원에서의 브라우징이 유용한 정보 찾기의 수단임을 나타낸다. 검색뭉치의 종료과정에서도 약 25.1%가 후행활동을 진행하는데, 그 중 친구(53.4%)에게 묻는 비중이 크게 나타나는데 이는 사람을 통한 정보활동이 웹검색 활동과 병행하는 주요 정보원임을 알 수 있다.

5.3 뭉치 내부의 정교화

뭉치는 여러 개의 쿼리들로 구성되어 있다. 그러면 뭉치 내부의 쿼리들은 어떤 관계를 가지는가? 검색이 진행되면서 쿼리들은 정밀해지는가? 아니면 쿼리들은 궁금증 주변만 애두르는가? 이러한 궁금증을 풀기 위해 뭉치 내 쿼리들간의 관계를 패턴화해 보았다. 먼저 하나의 뭉치에 대해 2인 이상이 그 내용을 관찰하여 범주화 코딩을 진행 하였다. 기본 코딩으로는 구문적인(syntactic) '쿼리의 정교화'와 의미적인(semantic) '주제의 정교화'를 제시하였다. 쿼리의 정교화란 궁금증을 형식화(formalize)하는 과정으로 먼저 적당한 단어를 떠올린 뒤 이를 검색엔진이 저장하고 매칭해주는 인덱스 용어와 맞추어 가는 과정을 말한다. 예를 들면 '부천교차로 → 부천IC'와 같이 검색엔진이 사용하는 통제어(controlled keyword)에 가까워지는 것을 의미한다. 반면 주제의 정교화는 검색과정을 통해 사용자의 머릿속에 있는 궁금증을 섬세하게 발전시켜 나가는 과정을 말한다. 예를 들면 '커피 → 에스프레소 → 에스프레스 롱고'를 들 수 있다.

두 개의 상위 코딩을 300개의 뭉치에 적용하면서 9개의 패턴이 발견되었다. 9개의 패턴은 먼저 쿼리의 정교화 수준에선 교정, 단순화, 구체화, 변경의 4개의 소 패턴이, 주제의 정교화 수준에선 확장, 탐색, 전환, 수행화, 입체화의 5개의 소 패턴이 도출하였다. 이렇게 도출된 9개의 코딩기준에 따라 1,662개의 전체 뭉치를 살펴본 결과 패턴의 이름과 성격이 조정되었고 최종적으로 9개의 범주가 도출되었다. 코딩 과정에서 애매한 성격을 가진 뭉치는 약 2% 정도였다. 이 9개의 정교화 패턴이 본 연구에서는 가장 중요한 관찰이라 볼 수 있다(〈표 4〉 참조). 9개의 패턴들은 다음과 같다. 먼저, 쿼리의 맞춤법을 고쳐나가는 '교정(fix)'이 있다. 하지만 교정의 비중은 그리 높지 않은데 이는 검색엔진이 제공하는 자동완성기능이나 철자 교정 기능이 좋아졌기 때문이다. 그 다음으로 쿼리를 문장형태로 길게 입력한 다음 단어의 개수를 '줄임(shorten)' 또는 단어의 수를 더해가는 '늘림(addition)'의 패턴이 나타난다. 줄임보다는 늘림의 비중이 높은데 이는 쿼리에 쿼리를 더했을 때 'or 연산'으로 작동되는 고급검색기능을 사용자들이 부지불식간에 학습했기 때문이다. 마지막으로 쿼리를 검색데이터 내부에 인덱스된 단어와 합치시키는 '통제어화(controlled)'를 발견할 수 있다. 이 정교화는 검색영역에 대한 사전지식을 필요로 한다. 반면 주제의 정교화에선 먼저 '관심의 확장(expansion)'이 주요 패턴으로 발견된다. 일정한 관심사를 갖고 관련 단어들을 확장해 가며 궁금증을 풀어가는 과정이다. 이 경우 좁은 영역에서 넓은 영역으로 확장되기도 하지만 넓은 영역에서 좁은 영역으로 좁혀지는 패턴도 나타난다. 그 다음으로 하나의 사건이나 인물에 대해 사진, 동영상, 뉴스 등을 갈아타며 궁금증을 입체적으로 해소하는 패턴이 나타난다. 포털이 멀티미디어 검색 기능을

〈표 4〉 뭉치 정교화의 범주화

대분류	중분류	코드	소분류	내용	예시	다이어그램
검색의 시작		N	일반적 검색 시작 Normal Search			
		T	실시간검색어 유입 Search Trend			
쿼리의 정교화	쿼리의 기본형태 유지	F	쿼리 교정 Fix	맞춤법 교정	나인&데이 → 나잇 & 데이	
		S	쿼리 길이 줄임 Shorten	형태를 유지한채 쿼리의 단어수를 줄여나감	부천교차로가는법 → 부천IC	
		A	쿼리 길이 늘림 Addition	형태를 유지한채 쿼리의 단어수가 늘어남	에어벤더→ 에어벤더개봉일	
	쿼리의 형태 변화	C	통제어화 Controlled	쿼리가 통제어에 가까워짐	거미줄섬유 → 케블러	
관심의 정교화	강한 주제의 확장	E	관심의 확장 Expansion	동일 관심의 여러 쿼리를 발생	커피→ 에스프레소 →룽고	
		H	입체화 Holistic	동일 관심의 다른 매체를 살펴봄	박지성→ 박지성 4호골 영상	
	주제의 통일성이 약함	R	반복 탐색 Repetitive Search	주가 검색처럼 대등한 검색을 수행	김밥집→ 김가네→ 김밥천국	
		J	점핑 Jumping	연관성이나 맥락이 약한 이동	소녀시대→ 신승훈 →장동건	
	절차가 내재된 검색	P	수행화 Perform	구매와 같이 수행의 단계가 심화됨	료토호텔→ 교통→ 맛집	

갖기에 사용자는 하나의 주제에 대해 여러 형식의 검색결과가 나오는 것을 학습한데서 나오는 패턴이라 생각된다. '점핑(jumping)'은 동일한 패턴으로 반복해서 비슷한 검색어를 계속 입력하는 형태인데 주가 검색이나 맛집 검색, 음악 검색 등에 많이 나타난다. 일정한 방향성이라기보다는 관심의 일정한 폭을 유지하는 행동으로 볼 수 있다. 마지막으로 '수행화(perform)'는 쇼핑 또는 여행 등 일련의 과업을 수행하기 위해 쿼리들이 바뀌어 가는 패턴을 얘기한다. 수행화는 비중이 작지만 아주 명확한 패턴을 갖는다.

1,662개의 뭉치 안에 존재하는 쿼리쌍들의 패턴을 정리해 보면 〈표 5〉와 같은 분포를 갖는다. 먼저 첫 번째 쿼리들은 유입의 성격에 의해 '정상유입'과 '실시간검색유입'으로 나누어 볼 수 있다. 다음 두 번째 이상의 쿼리들에선 쿼리의 정교화(46%)와 주제의 정교화(54%)가 대략 절반씩의 비중을 갖는다. 쿼리의 정교화 중 가장 높은 비중은 쿼리의 변경(17%)이고 그 다음이 쿼리의 구체화(11%)이다. 주제의 정교화에서는 전환(22%)과 탐색(10%)의 비중이 크다.

이러한 분포는 첫번째 쿼리의 성격, 즉 유입의 성격에 따라 크게 달라지는데, 〈표 6〉에서와

〈표 5〉 쿼리정교화의 분포

대분류	코드	정교화	비율
유입	N	일반적 검색 시작 Normal Search	21%
	T	실시간검색어 유입 Search Trend	2%
쿼리의 정교화	F	쿼리의 교정 Fix	3%
	S	쿼리의 길이를 줄임 Shorten	5%
	A	쿼리의 길이를 늘림 Addition	11%
	C	통제어화 Controlled	17%
주제의 정교화	E	관심의 확장 Expansion	7%
	H	입체화 Holistic	2%
	R	반복 탐색 Repetitive Search	10%
	J	점핑 Jumping	22%
	P	수행화 Perform	0%

〈표 6〉 검색엔진 유입에 따른 정교화 빈도

대분류	코드	정교화	실시간 검색어로 유입 (전체 뭉치의 8%)	일반 검색창에서 시작 (전체 뭉치의 92%)
쿼리의 정교화	F	쿼리의 교정 Fix	0.8%	3.6%
	S	쿼리의 길이를 줄임 Shorten	2.2%	6.4%
	A	쿼리의 길이를 늘림 Addition	4.8%	15.2%
	C	통제어화 Controlled	8.3%	24.7%
주제의 정교화	E	관심의 확장 Expansion	5.0%	9.6%
	H	입체화 Holistic	1.2%	2.9%
	R	반복 탐색 Repetitive Search	5.3%	14.2%
	J	점핑 Jumping	72.3%	22.9%
	P	수행화 Perform	0.2%	0.5%

같이 실시간 검색어로 초기 접근한 사용자는 '전환(72%)'을 높은 비율로 사용해 특정한 관심 없이 여러 검색어들을 서핑하는 경향이 있음을 알 수 있다. 반면 정상적으로 검색창에 쿼리를 입력하여 시작한 경우에는 '쿼리 변경(24.7%)'과 '전환(22.9%)'의 순서로 검색에 대한 구체적 인식이 있음을 알 수 있다.

6. 연구의 의의 및 시사점

6.1 로그기반 연구와 RTA

최근 증가하고 있는 트위터 연구들은 로그데이터의 접근성이 높아지면서 가능해진 트랜드라 할 수 있다. 앞에서 설명했지만 로그데이터는 사용량, 내용, 관계 등 상당히 차원 높은 해

석을 가능하게 한다. 따라서 이 연구에서처럼 보안체계 안에 가려져 있는 로그 또는 분석자가 원하는 로그를 발생하도록 도구를 만들어 시도되는 연구들이 점차 중요해진다. 물론 이러한 연구는 사용자의 동의를 거쳐야 하고 필요한 만큼의 데이터만 추출하는 연구 윤리의 민감함을 가지고 있다. 이러한 어려움을 극복할 수 있다면 사적인(private) 로그의 분석은 인간활동의 내적 특성을 살펴보는데 훨씬 유용하다. 왜냐하면 최근의 정보기기들은 우리의 인지구조와 아주 밀접히 연동되기 때문에 이러한 로그들은 우리의 의식적/무의식적 정신활동의 단면을 잘 보여주기 때문이다. 이 연구에서는 로그 데이터에 추가하여 기억의 증거(memory cue)로 제시하는 RTA방식의 인터뷰를 통해 피실험자에게 자신의 내적 행동에 대한 회상을, 어려운 자유 회상(free recall)에서 조금 쉬운 힌트회상(hinted recall)수준으로 낮추어 진행하였다.

6.2 뭉치 분석

이 연구에서는 정보검색의 개별활동(instance)을 살펴 보는 대신 검색뭉치(bundle)를 관찰하는 방법을 택했다. 활동을 뭉치는 방식은 쿼리의 복잡성을 고려할 때 텍스트마이닝 방식보다 사용자가 직접 참여하는 방식을 택했다. 이를 위해 사용자가 흥미롭게 참여할 수 있는 도구와 인터페이스를 만들었는데, 실험 과정에서 피실험자들이 이 도구의 사용을 상당히 흥미로워한다는 사실을 알게 되었다. 자신의 웹 활동이 시각화되고 검색기록이 컬러코딩 되어 나타나는 점은 하루 종일 주고받은 문자기록을 되짚어 보는 반추적(reflective) 흥미가 있다. 쿼리를 드래그하여 모으는 과정은 게임이 주는 즐거움과 비슷하다는 피드백도 받았다. 또한 사용자들이 도구의 숙지과정이나 실험과정에서 탈락하거나 실수하는 경우가 거의 없어 '뭉치기 툴'이 웹환경에 친숙한 네티즌들에게 적절한 학습 수준(learning curve)을 제공함도 확인할 수 있었다. 향후에는 이런 도구를 확장하여 소셜네트웍 게임(social network game)의 형태로 발전시킬 수 있는데 그 동안 측정이 어려웠던 사용자의 의도나 동기 또는 만족도 등을 집단적으로 관찰할 수 있는 가능성이 확인되었다.

6.3 뭉치의 외부적 특성

사용자가 만든 뭉치는 규칙기반으로 분류된 세션과 달리 훨씬 다양한고 폭넓은 정보를 제공했다. 먼저 뭉치는 규칙기반의 세션에서는 불가능한 명확한 개인식별이 가능하기 때문에 정량적 데이터와 더불어 설문 등을 통한 정성적 데이터와의 연동한 분석이 가능하다. 여기에 개별 뭉치에 대한 웹설문과 RTA 인터뷰를 통해 궁금증의 동기, 계기, 병행작업과 후행작업도 살펴 볼 수 있었다.

6.4 뭉치 내부적 특성

뭉치 연구의 하이라이트는 뭉치 내부의 쿼리들이 어떻게 발전해 가는가를 살펴보는 것이다. 여기서는 쿼리의 정교화, 주제의 정교화라는 두 개의 9개의 세부 정교화 패턴을 확인하였다. 한편 본 연구의 데이터를 활용한 파생 연구에서는 뭉치 내의 검색보조기술이 검색행동에 미치는

영향을 살펴볼 수도 있었다(김문성 2011).

이 연구는 복합화된 정보검색 환경의 사용자와 그들의 행동을 이해하기 위한 새로운 연구방법 제시에 의미를 둔다. 설문 대신 로그를 사용하고, 트랜잭션 로그 대신 사용자 로그를 사용했으며, 개별 쿼리 대신 쿼리 뭉치를 살펴보았다. 이런 시도를 통해 전에 볼 수 없던 검색 행동의 미시적이고 또 거시적인 패턴을 살펴 볼 수 있었다. 이 연구는 어떤 목적(goal)을 가지지만 시간적으로 인접하지 않고 여러 매체를 아우르며 진행되는 최근의 정보검색 활동을 이해하는데 유용한 데이터 수집 방법과 분석기법을 제공한다.

이 연구에서 제시한 정보검색 뭉치의 유형화는 사용자들의 행위를 사후적(expost-facto)으로 분류해 만든 것이다. 다시 말해 이론적으로 모델을 정립한 후 데이터를 수집해 검증한 것(theory-driven)이 아니라 데이터에서 모델을 추출한 것(data-driven)이다. 하지만 공급자 측면의 트랜잭션 로그 아닌 사용자 관점의 사용자 로그를 개별 쿼리를 사용하지 않고 쿼리뭉치를 만들어 분석하는 새로운 방법의 연구로 아직 연구되지 않은 분야를 개척하는 탐색적 연구(exploratory study)로서는 가치가 있다. 앞으로 이러한 방법론을 사용한 후속연구가 계속 수행되어 많은 연구결과가 축적된다면 인간 정보행위의 보다 다양한 수준에서의 '뭉치기'가 가능해 질 것이다

참 고 문 헌

[1] 김문성. 2011.『웹 검색행태 분석을 통한 사용자 유형 연구: 검색보조기술이 사용자 검색 행동에 미치는 영향』. 석사학위논문, 서울대학교 융합과학기술대학원.

[2] 박소연, 이준호. 2001. 로그분석을 통한 인터넷 검색엔진 이용자의 웹 문서 검색 행태에 관한 연구.『한국문헌정보학회 학술논문발표집』, 14: 61-73.

[3] Bates, M. J. 1989. "The design of browsing and berrypicking techniques for the online search interface." *Online Information Review*, 13(5): 407-431.

[4] Baxter, L., Montgomery, B., & Duck, S. 1994. *Studying Interpersonal Interaction*. New York: Guilford Press.

[5] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. 2010. "Measuring user influence in twitter: The million fallacy." *Proceedings of International AAAI Conference on Weblogs and Social Media(ICWSM)*, May, 2010.

[6] Chang, S. J. 1993. "Browsing: A multidimensional framework." *Annual Review of Informational Science*, 28: 231-276.

[7] Choo, C., Detlor, B., & Turnbull, D. 1999. "Information seeking on the web: An integrated

model of browsing and searching." *Proceedings of the 62nd ASIS Annual Meeting*, 36: 3-16. Oct. 31-Nov. 4, 1999. Washington, D.C.

[8] Jansen, B., et al. 2000. "Real life, real users, and real needs: A study and analysis of user queries on the web." *Information Processing and Management*, 36(2): 207-227.

[9] Jansen, B., et al. 2002. "From e-sex to e-commerce: Web search changes, computers." *IEEE Computer*, 35(3): 107-109.

[10] Jansen, B., & Spink, A. 2003. "An analysis of web information seeking and use: Documents retrieved versus documents viewed." *Proceedings of the 4th International Conference on Internet Computing*, 65-69. June 23-26, 2003. Las Vegas, Nevada.

[11] Jansen, B., & Spink, A. 2006. "How are we searching the World Wide Web? A comparison of nine search engine transaction logs." *Information Processing and Management*, 42(1): 248-263.

[12] Jansen, B., et al. 2007. "Defining a session on web search engines." *Journal of the American Society for Information Science and Technology*, 58(6): 862-871.

[13] Kellar, M., Watters, C., & Shepherd, M. 2007. "A field study characterizing web-based information-seeking tasks." *Journal of the American Society for Information Science and Technology*, 58(7): 999-1018.

[14] Kuhlthau, 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.

[15] Kwak, H., Lee, C., Park, H., & Moon, S. 2010. "What is twitter, a social network or a news media?" *19th International World Wide Web Conference*, 591-600. Apr. 26-30, 2010. Raleigh NC.

[16] Liu, B. 1998. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Berlin: Springer.

[17] Nielson, J. 2006. *Prioritizing Web Usability*. Indianapolis: New Riders.

[18] Rice, E. 1993. *Accessing and Browsing Information and Communication*. Cambridge: MIT Press.

[19] Rose, D., & Levinson, D. 2004. *Understanding user goals in web search*. New York: ACM.

[20] Silverstein, C., et al. 1999. *Analysis of a Very Large Web Search Engine Query Log*. New York: ACM.

[21] Shi, X. 2007. "Mining related queries from web search engine query logs using an improved association rule mining model." *Journal of the American Society for Information*, 58(12): 1871-1883.

[22] Spink, A., Wolfram, D., Jansen, M. B. J., & Saracevic, T. 2001. "Searching the web: The public and their queries." *Journal of the American Society for Information Science and Technology*, 52(3): 226-234.

[23] Taylor, R. 1962. "The process of asking questions." *American Documentation*, 13(4): 391-396.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

[1] Kim, Mun Seong. 2011. *Characterizing Web Search Behavior through Patterns in Using Search Assistant Tools*. M.A. thesis, Graduate School of Convergence Science and Technology, Seoul National University.

[2] Park, So-Yeon, & Lee, Joon-Ho. 2001. "Log bunseokeul tonghan internet geomsaek enjin iyongjaui web munseo geomsaek haengtaee gwanhan yeongu." *Korean Society for Library and Information Science Haksul Nonmun Balpyojip*, 14: 61-73.