

기자직의 웹 정보탐색행위 패턴 분석*

Analyzing Patterns in News Reporters' Information Seeking Behavior on the Web

권혜진(Hye-Jin Kwon)**

정동열(Dong-Youl Jeong)***

초 록

본 연구의 목적은 기자직의 웹 행동 분석을 통해 웹 정보탐색행위 패턴의 특성을 규명하는 것이다. 이를 위해 본 연구는 기자직 23명을 대상으로 트랜잭션로그를 수집하여 분석하였다. 데이터 수집을 위하여 분석 대상의 컴퓨터에 웹 트래킹 소프트웨어를 설치하였으며, 2주 동안 총 39,860개의 웹 로그가 수집되었다. 세션의 시작과 종료 패턴, 단계별 이행 패턴, 순차규칙 모형이 분석되었으며, 기자직과 일반인의 웹 이용 패턴을 비교하였다. 이를 토대로 기자직의 웹 정보탐색행위 패턴을 사실확인형 탐색, 사실확인형 검색, 조사형 탐색, 조사형 검색 등 4가지 형태로 유형화하였다.

ABSTRACT

The purpose of this study is to identify the patterns in the news reporters' information seeking behaviors by observing their web activities. For this purpose, transaction logs collected from 23 news reporters were analyzed. Web tracking software was installed to collect the data from their PCs, and a total of 39,860 web logs were collected in two weeks. Start and end pattern of sessions, transitional pattern by step, sequence rule model was analyzed and the pattern of Internet use was compared with the general public. The analysis of pattern derived a web information seeking behavior modes that consists of four types of behaviors: fact-checking browsing, fact-checking search, investigative browsing and investigative search.

키워드: 기자직, 웹, 정보탐색행위, 트랜잭션로그 분석, 이용자연구
news reporters, web, information seeking behavior, transaction logs analysis, user studies

* 이 논문은 이화여자대학교 대학원 박사학위 논문의 일부를 축약한 것임.

** 동아일보 편집국 통합뉴스센터 인터넷뉴스팀 기자(hjkwon@donga.com) (제1저자)

*** 이화여자대학교 문헌정보학과 교수(dyjeong@ewha.ac.kr) (공동저자)

■ 논문접수일자: 2010년 11월 14일 ■ 최초심사일자: 2010년 11월 20일 ■ 게재확정일자: 2010년 11월 24일
■ 정보관리학회지, 27(4): 109-130, 2010. [DOI:10.3743/KOSIM.2010.27.4.109]

1. 서론

1.1 연구의 목적

정보와 관련된 인간 행위 연구의 역사는 그 주체의 중요성이 거의 한 세기에 걸쳐 지속되어 왔으며, 직업, 사회적 역할, 인구학적 집단 등으로 더욱 세분화되고 전문화되고 있다. 또한, 1990년대 중반 이후 급속도로 이루어진 인터넷의 보급은 이용자들의 정보탐색행위를 전면적으로 변화시키고 있으며, 이러한 환경 변화 속에서 웹 정보탐색행위에 관한 연구도 증가하였다.

외국의 경우 직업별 정보탐색행위에 대한 연구는 학자를 비롯하여 엔지니어, 저널리스트, 법률가, 관리자 등 다양한 직업군을 대상으로 활발하게 진행되었다. 반면, 국내에서는 교수, 연구자, 학생 등 대부분 학술 집단을 대상으로 한 연구에 머무르고 있어 비학술 전문 집단으로 확대될 필요가 있다.

이러한 연구의 필요성에 따라 본 연구는 비학술 집단에 해당하는 기자직을 대상으로 웹 정보탐색행위를 연구하고자 하였다. 기자직은 정보의 적극적 소비자인 동시에 생산자이며, 정보탐색과 수집이 가장 중요한 업무 역할이라는 점에서 다른 직업과 구분되는 고유한 특성이 있다. 이 같은 기자의 직업적 특성으로 인해 외국에서는 오랫동안 이용자연구의 중요한 연구 대상이 되어왔으나 국내에서는 관련 연구가 미미한 실정이다.

본 연구의 목적은 기자직의 웹 정보탐색행위를 실증적으로 연구함으로써 기자직이 갖는 웹 정보탐색행위의 패턴을 규명하는 것이다. 본

연구의 세부 목적을 요약하면 다음과 같다. 첫째, 기자직의 트랜잭션로그 분석을 통해 웹 정보탐색과정에서 웹 이용 활동이 실제 어떻게 나타나는지 관찰하고 그 패턴을 분석한다. 둘째, 트랜잭션로그 분석 결과를 토대로 기자직에 적합한 웹 정보탐색행위 패턴 유형을 도출한다.

1.2 연구의 방법과 범위

본 연구는 기자직의 웹 정보탐색과정에서 웹 이용 활동이 실제 어떻게 나타나는지를 실증적으로 분석하기 위해 23명의 기자 컴퓨터에 웹 트래킹 소프트웨어를 설치하여 트랜잭션로그 데이터를 수집하여 분석하였다. 로그 수집 기간은 2009년 9월 7일에서 9월 20일까지 2주 동안으로 한정하였다.

트랜잭션로그 분석은 관찰법의 한 변형으로, 컴퓨터의 자동 기록 장치를 이용하여 사용자가 입력한 내용과 시스템의 반응을 분석하는 방법을 말한다. 이러한 데이터 수집 방식은 그 결과가 이용자 ID, URL, 접속 시간, 사용한 웹 브라우저 기능 등을 포함하기 때문에 웹 정보탐색행위에 대한 상세한 명세를 확인할 수 있는 장점이 있다. 또한 정보탐색과정이 자동적으로 기록되기 때문에 인터뷰나 설문조사 등에서 흔히 나타나는 커뮤니케이션 에러가 제거되는 것도 장점이라 할 수 있다.

본 연구의 분석 대상인 기자직의 범위는 동아일보 및 계열사에 소속된 기자직 중에서 선정하였으며, 직접 기사 작성에 참여하는 취재 기자직을 중심으로 하였다. 소속된 매체의 범위는 일간지, 주간지, 월간지, 인터넷 신문 등

가급적 다양한 매체를 포괄하였다.

1.3 선행 연구

웹 정보탐색행위 모형에 대한 연구는 탐색행위를 기초로 한 Choo 등(1998, 1999, 2003)의 연구, Chen과 Cooper(2002)의 연구, Hölscher와 Strube(2000)의 연구와 웹과 이용자 간의 상호작용을 중심으로 한 Wang 등(2000)의 연구, Hsieh-Yee(2001), Spink(2002)의 연구 등이 있다.

Choo 등(1998)의 연구는 7개 일반 기업체의 회사원 34명을 대상으로 웹 정보탐색행위 패턴을 조사하였다. 이 연구는 웹에서 일어나는 정보탐색행위를 조사하여 기존의 정보탐색행위 이론과 비교해 본 첫 번째 연구라 할 수 있으며, 기존의 정보탐색행위가 웹에서는 어떻게 변화되어 나타났는지를 보여주었다.

Chen과 Cooper(2002)의 연구는 캘리포니아 MELVYL 도서관 목록시스템의 트랜잭션 로그에서 25만 건이 넘는 검색 세션을 추출하여 웹 이용 패턴을 6개의 범주로 구분하였다. 이 연구는 이용자에 대한 배경 정보 없이 이용자의 정보탐색 목적에 따른 웹 이용 패턴을 도출해냄으로써 이용자의 웹 탐색 전략을 모형화하였다. 한편 Hölscher와 Strube(2000)는 가장 이상적으로 웹 탐색을 할 것이라 예상되는 웹 전문가 12명을 대상으로 탐색행위를 연구하여 정보탐색행위 모형을 구축하고자 하였다. 이 연구는 과거의 모형들에 비해 비선형적인 특성이 강하게 반영되어 있다.

국내에서 개별 직업 또는 학문 분야의 정보탐색행위를 다룬 연구로는 사회과학 연구자

들의 정보이용행위와 자연과학과 사회과학 연구자들의 정보이용특성을 분석한 최은주(1996, 1997)의 연구, 전자공학전공 대학원생들의 학술정보 이용행태를 분석한 박일종과 백수연(2001)의 연구 등이 있으며, 주로 학술 집단을 대상으로 연구가 이루어졌다.

인터넷이 확산된 이후에는 전문 분야별 웹 정보탐색행위에 관한 연구도 진행되었는데 이혜영과 이수영(2001)은 다양한 형태와 성격의 정보가 혼재돼 있는 인터넷 환경에서 이공계 연구자들이 어떻게 정보를 탐색하고 평가하며 이용하는지를 분석하였다.

위의 선행연구들이 개별 직업 또는 학문 분야에서 이루어지는 정보탐색행위 자체에 대한 분석에 초점을 두고 있다면 웹 이용자의 정보탐색행위 패턴을 분석한 김성진(2006)과 영화창작자의 정보활동 모형을 설계한 이정연(2008)의 연구는 이용자의 정보탐색행위 연구를 바탕으로 정보탐색행위 모형을 제시했다.

기자직을 대상으로 한 국내외 연구를 살펴보면, 해외에서 문헌정보학 분야에서 기자직 정보탐색행위를 포괄적으로 고찰한 최초의 연구는 Herron(1986)의 연구이다. Herron의 연구는 저널리스트들 사이에서 가장 선호하는 커뮤니케이션 형태가 '면대면 접촉이나 전화를 사용한 비공식적 구두 채널에 의한 개인적 커뮤니케이션'인 것으로 제시하였는데 그 후 이루어진 다른 연구들(Abdulla 2006; Anwar et al. 2004; Campbell 1997; Vreekamp 1995; Edem 1993; Joseph 1993)에서도 이와 같은 저널리스트의 정보원 선호도가 확인되었다.

이후의 연구로는 6년간에 걸쳐 영국 가디언지 기자들의 온라인 검색 행위에 대해 포괄적인

사례연구를 수행한 Nicholas와 Martin(1993)의 연구, 저널리스트의 정보 행위를 살펴본 Chinn(2001)의 연구, 전자적 환경과 관련하여 저널리스트의 정보 행위를 다룬 Harman(1986)의 연구 등이 있다. 특히, Fabritius(1999)는 신문 기자들의 기사 작성 과정에서 전자정보원의 역할과 정보기술의 활용, 디지털 시스템이 기자직의 업무를 어떻게 변화시켰는가에 대해 연구하였다.

기자직의 정보탐색행위를 바탕으로 모형을 구축한 연구는 그다지 많지 않다. 기자직의 정보탐색행위에 영향을 미치는 요인을 모형화한 Fabritius(1999)의 연구, 기자직의 업무 맥락에서 기사작성 과정을 분석하여 모형화한 Attfield와 Dowell(2003)의 연구 정도를 찾아볼 수 있으며 국내에서는 관련 연구가 이루어지지 않았다.

국내에서 수행된 기자직의 정보탐색행위에 관한 연구는 해외에 비해 매우 미미한 수준에 불과하다. 관련 연구로는 기자직의 웹 탐색엔진 이용에 대해 연구한 임석중(1998)의 연구, 컴퓨터를 활용한 취재보도(Computer-Assisted Reporting)에 초점을 맞추어 기자직의 인터넷 활용을 분석한 권혜진(1999)의 연구, 취재환경 변화에 따른 기자의 정보탐색행위를 심층면담 방법으로 연구한 장선화(2005)의 연구가 있다. 특히, 웹 정보탐색행위에 기반한 패턴 분석이나 모형 구축에 관한 연구는 거의 이루어지지 않아 이에 대한 연구가 필요하다.

2. 연구의 설계

2.1 데이터 수집 방법과 대상

본 연구는 기자직의 웹 정보탐색행위 패턴을 분석하기 위하여 트랜잭션로그 분석을 실시하였으며, 코리안클릭사(www.koreanclick.com)의 웹 트래킹 소프트웨어인 iTrack을 사용하여 웹 로그 데이터를 수집하였다. iTrack 소프트웨어는 코리안클릭에서 자체 개발하여 패널에게 배포하는 소프트웨어로, 패널의 동의하에 패널의 인터넷 이용 정보를 실시간으로 코리안클릭 데이터 수집 서버에 전송하는 로그전송 프로그램이다. 본 연구는 코리안클릭사 웹 서버에 본 연구를 위한 별도의 저장 공간을 마련하여 일반인의 데이터와 구분하였으며, 기자들 컴퓨터에서 일어나는 웹 이용 활동을 실시간으로 수집하였다.

트랜잭션로그 분석의 대상자는 동아일보 및 계열사에 소속된 취재기자직 중에서 매체, 취재분야, 직책, 경력을 기준으로 선정하였으며, 일간지 19명, 주간지 2명, 월간지 1명, 인터넷신문 1명 등 총 23명을 대상으로 하였다. 일간지의 경우 논설위원 1명, 정치 3명, 경제 3명, 사회 3명, 국제 3명, 문화 3명, 스포츠 3명 등 취재분야가 골고루 분포되도록 대상자를 선정하였다. 취재분야 내에서의 분포는 경력과 직책에 따라 데스크 1명, 5년 이상 평기자 1명, 5년 미만 평기자 1명 등 각 3명씩을 선정하였다.

이러한 선정 기준에 따라 8월 24일부터 28일까지 대상자들을 개별 접촉하여 연구의 의의와 내용을 설명하고 소프트웨어 설치에 대한 동의를 구하였다. 이 결과 총 23명의 취재기자들이 소프트웨어 설치를 동의하였으며, 설치에 동의한 기

〈표 1〉 트랜잭션로그 분석 대상자의 분포

패널ID	구분	경력	취재분야	인원
1	논설위원	15년 이상	경제	종합일간지 19명
2	데스크	15년 이상	정치	
3	데스크	15년 이상	경제	
4	데스크	15년 이상	사회	
5	데스크	15년 이상	국제	
6	데스크	15년 이상	문화	
7	데스크	15년 이상	스포츠	
8	평가자	5-15년	정치	
9	평가자	5-15년	경제	
10	평가자	5-15년	사회	
11	평가자	5-15년	국제	
12	평가자	5-15년	문화	
13	평가자	5-15년	스포츠	
14	평가자	5년 미만	정치	
15	평가자	5년 미만	경제	
16	평가자	5년 미만	사회	
17	평가자	5년 미만	국제	
18	평가자	5년 미만	문화	
19	평가자	5년 미만	스포츠	
20	평가자	5-15년	정치	주간지 2명
21	평가자	5-15년	문화	
22	평가자	5-15년	정치/경제	월간지 1명
23	평가자	5-15년	인터넷뉴스	인터넷신문 1명

자들에게 9월 4일 각각 e-메일을 발송하여 컴퓨터에 직접 프로그램을 설치하도록 하였다. 트랜잭션로그 분석 대상자의 분포는 〈표 1〉과 같다.

프로그램이 설치된 각각의 PC는 e-메일을 통해 부여한 패널ID로 구별하였으며, 관찰 기간은 2009년 9월 7일에서 9월 20일까지 2주 동안으로 한정하였다. 웹 로그 수집 기간과 일치하는 2주간의 일반인 이용 행태 통계도 코리안 클릭사에서 동시에 입수하여 취재기자직과 일반인의 정보탐색행위 패턴을 비교하는 데 활용하였다.

2.2 분석 방법

데이터 수집을 위해 2009년 9월 7일에서 9월 20일까지 2주 동안 기자들의 컴퓨터에서 코리안 클릭사의 웹서버에 실시간으로 전송된 로그를 코리안 클릭사의 데이터 전처리과정을 거쳐 엑셀 파일로 입수하였다. 트랜잭션에서 발생한 자료 중에서 이미지 파일(jpg, gif 등) 등은 분석 내용과 상관이 없기 때문에 전처리 과정에서 삭제되었다. 전처리 과정을 거쳐 엑셀 파일로 입수한 로그 건수는 총 39,860건이었으며,

한 로그의 필드는 SERVER_DATE, ACCESS_DAY, PANEL_ID, BROWSER_ID, SITE_ID, SECTIONS, DOMAIN_URL, QUERY_DECODE, REF_SITE_ID, REF_URL, REF_PAGE 등 27개로 구성되어 있다.

39,860개의 로그 자료는 본 연구의 분석 대상에 맞게 데이터 축소 및 분류의 후처리 과정을 거쳤다. 데이터 축소에서는 불필요한 사이트의 로그, 예를 들면 사내 인트라넷과 은행 등의 사이트를 삭제하였는데 이 과정에서 분석대상의 로그 건수가 27,810건으로 감소하였다. 분석을 위한 후처리 과정에서는 동일 세션을 표시하는 세션ID, 전 탐색행위와의 시간차, 키워드 개수, 업무성격, 접속시간, 오전오후, 시간대, 탐색형태 등의 필드가 추가되었다. 로그데이터의 후처리 과정에서는 Microsoft Excel 2007을 사용하였으며, 후처리 과정을 거친 로그데이터는 SPSS 12.0 for Windows와 웹 데이터마이닝 소프트웨어인 SPSS PASW Modeler 13.0 (구 클레멘타인)을 사용하여 분석하였다.

로그 분석을 위해선 세션의 구분이 필요한데,

세션은 트랜잭션로그 분석의 기본 단위로써 일반적으로 한 명의 이용자가 단일한 검색 목적을 지니고 처음 검색을 시작하여 검색을 종료하기까지의 일련의 과정으로 정의된다(박소연, 이준호 2002). 본 연구는 Choo 등의 연구(1998)와 같이 30분의 시간차를 기준으로 세션을 구분하였다. 트랜잭션로그 데이터의 분석 범위는 프라이버시 문제를 피하기 위하여 세부적인 검색 내용은 분석에서 제외하였고 웹 행동(Web move)만을 분석하는 것으로 범위를 한정하였다.

각각의 로그는 웹 행동 유형에 따라, 시작은 S, 연결은 C, 브라우징은 B, 차별화는 D, 모니터링은 M, 추출은 E를 코드로 부여하여 필드를 추가하였다. 차별화의 경우 Choo 등(1998)의 연구에서는 북마크나 인쇄, 복사와 붙여넣기 등을 이용하여 특정 사이트로 이동하는 경우인 것으로 정의했으나, 본 연구에 사용된 트래킹 소프트웨어는 북마크, 인쇄, 복사 등의 활동을 추적하는 기능이 없는 문제로 실제 웹 탐색행위 분석 대상에서는 제외하였다. 본 연구에서 채택한 웹 활동의 유형과 정의는 <표 2>와 같다.

<표 2> 웹 행동의 유형과 정의

웹 행동 유형	코드	Choo 등의 연구(1998)	본 연구
시작	S	• 전 로그와의 시간차가 30분을 초과한 경우	• 전 로그와의 시간차 30분을 초과한 경우
연결	C	• 시작 페이지에서 다른 콘텐츠 관련 사이트로 링크를 타고 이동한 경우	• 유입 사이트와 현 접속 사이트가 다른 경우. 즉 다른 사이트에서 이동한 경우
브라우징	B	• 최상위 페이지나 목록, 제목, 사이트맵 등을 훑어보는 경우	• 동일 사이트 내 클릭으로 이동한 경우
차별화	D	• 북마크, 인쇄, 복사, 붙여넣기 등을 이용하여 사이트를 이동하는 경우	-
모니터링	M	• 푸시 등의 기능을 이용하여 업데이트된 정보를 받거나 특정 사이트를 재방문하는 경우	• 직접 주소를 입력하고 들어온 경우. 즉 유입 경로와 검색 질의어가 없는 경우
추출	E	• 정보를 추출하기 위해 검색	• 검색 질의어가 존재하는 경우
종료		• 세션 종료	• 세션 종료

3. 연구 결과 및 분석

3.1 세션별 패턴 분석

27,810건의 전체 로그를 대상으로 탐색행위별 분포를 살펴보면, 가장 많이 나타난 탐색행위는 브라우징(B)으로 47.5%인 13,219건이었으며, 그 다음이 추출(E)이 6,528건으로 23.5%, 모니터링(M)이 5,674건으로 20.4%, 연결(C)이 2,389건으로 8.6%의 순으로 나타났다(표 3 참조).

추출(E) 행위가 나타난 6,528건을 대상으로 검색 질의어 수를 분석해 보면 1단어를 사용하여 검색한 것이 3,486건으로 53.4%를 차지해 과반수를 넘는 것으로 나타났다. 2단어 검색은 1,772건으로 27.1%를 차지했으며, 3단어 검색

이 832건으로 12.7%, 5단어 이상이 160건으로 2.5%였다. 평균 질의어 수는 1.78단어였으며 최대 질의어 수는 19단어였다(표 4 참조).

한 세션 당 로그 건수, 즉 한 세션이 몇 번의 탐색행위로 이어졌는지를 살펴보면, 1건으로 세션이 종료된 것은 103건으로 11.8%를 차지했으며, 2~10건이 326건으로 37.3%, 11~50건이 298건으로 34.1%, 50건 이상인 146건으로 16.7%였다. 한 세션은 최소 1건부터 최대 1,132건까지 연속하여 이어졌으며, 한 세션 당 평균 탐색행위는 31.8건인 것으로 나타났다(표 5 참조). 한 세션이 1건으로 끝난 경우를 구체적으로 살펴보면, 모니터링(M)이 69건으로 67.0%, 추출(E)이 21건(20.4%)이었고, 브라우징(B)이 7건으로 6.8%, 연결(C)이 6건으로 5.8%를 차지하였다.

〈표 3〉 전체 로그의 탐색행위별 분포

구분	모니터링(M)	추출(E)	브라우징(B)	연결(C)	합계
빈도수	5,674	6,528	13,219	2,389	27,810
백분율	20.4%	23.5%	47.5%	8.6%	100.0%

〈표 4〉 전체 추출(E) 행위의 질의어 수 분포

구분	1개	2개	3개	4개	5개 이상	합계	평균	최대값
빈도수	3,486	1,772	832	278	160	6,528	1.78	19
백분율	53.4%	27.1%	12.7%	4.3%	2.5%	100.0%		

〈표 5〉 한 세션 당 로그 건수 분포

한 세션 당 로그 건수	빈도수	백분율	평균 로그 건수	최대 로그 건수
1건	103	11.8%	31.8	1,132
2~10건	326	37.3%		
11~50건	298	34.1%		
50건 이상	146	16.7%		
합계	873	100.0%		

3.2 시작과 종료 패턴 분석

3.2.1 세션의 시작 패턴

873개 세션의 시작 패턴을 살펴보면 모니터링(M)으로 시작된 것이 510건으로 58.4%를 차지하여 가장 많았으며, 추출(E)이 165건으로 18.9%, 브라우징(B)이 103건으로 11.8%, 연결(C)이 95건으로 10.9% 순으로 나타났다(표 6 참조).

3.2.2 웹 정보탐색 시간대와 직책별 시작 패턴

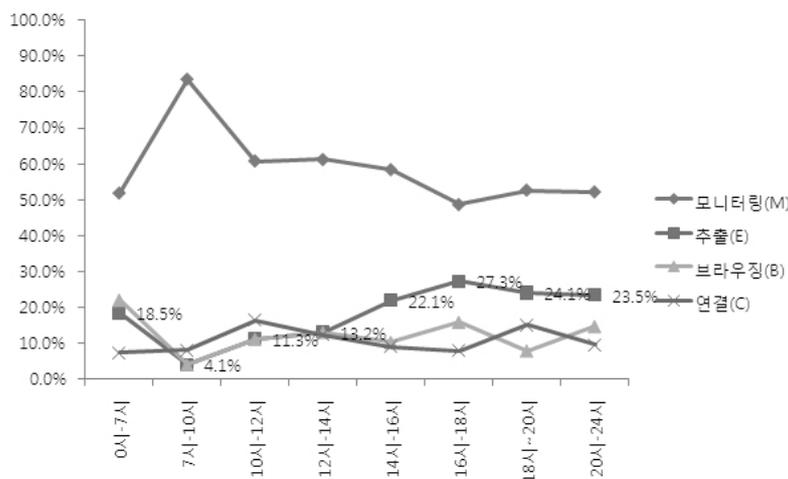
다음은 기자직이 웹 정보탐색을 시작할 때 시간대와 직책에 따라 웹 행동 유형의 차이가 있는지 살펴보았다. 시간대를 고려하는 이유는 기사 아이디어 수집 단계와 집필 단계의 시작 패턴

차이를 보기위한 것이다. 세션의 시작을 시간대로 세분화하여 검색의 비중 변화를 살펴보면, 기자직의 웹 정보탐색 시작 패턴이 시간대와 매우 밀접한 관계가 있음을 알 수 있다(그림 1 참조).

분석 결과, 기사 계획인 발제를 준비하는 오전 7시~10시에는 추출(E) 즉, 검색의 비중이 4.1%에 불과했으나, 오후 발제를 준비하는 10시~12시엔 11.3%, 12시~14시엔 13.2%로 검색 비중이 상승하였다. 오후 발제가 끝난 후 기사 작성 단계인 2시~4시에는 검색 비중이 22.1%, 오후 4시~6시에는 27.3%로 발제 단계보다 검색 비중이 2배 이상 높아짐을 확인할 수 있다. 기사 송고 후 초판 기사를 수정하는 오후 6시~8시 사이에도 검색의 비중이 24.1%로 오전에

〈표 6〉 웹 정보탐색 시 시작 행동 패턴

구분	모니터링(M)	추출(E)	브라우징(B)	연결(C)	합계
빈도수	510	165	103	95	873
백분율	58.4%	18.9%	11.8%	10.9%	100.0%



〈그림 1〉 시간대에 따른 시작 패턴의 변화

비해 2배 이상 높았다.

간부직인 데스크와 평기자의 직책에 따른 차이를 살펴보면, 데스크는 모니터링(M) 행위가 45.6%로 평기자의 62.4%보다 훨씬 낮았다. 추출(E) 행위의 경우 데스크가 24.3%인 반면, 취재 기자는 17.2%로 데스크의 추출(E) 행위 비중이 더 높게 나타났다. 브라우징(B) 행위는 데스크가 20.4%인 반면, 취재기자는 9.1%여서 역시 데스크의 브라우징(B) 행위 비중이 더 높았다. 연결(C) 행위에서는 취재부서 데스크가 9.7%인 반면, 취재기자는 11.2%로 나타나 취재기자의 연결(C) 행위 비중이 다소 높은 것으로 나타났다(그림 2 참조).

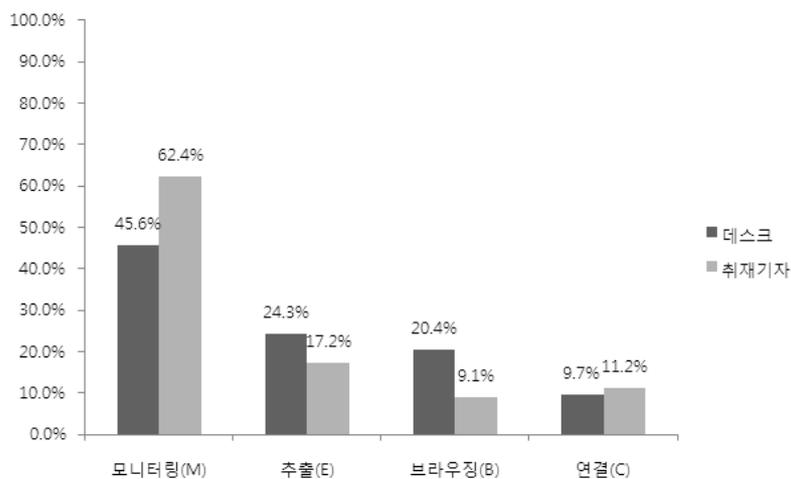
다시 말하면, 취재기자는 데스크에 비해 모니터링(M)과 연결(C) 행위가 더 활발했으며, 데스크는 취재기자에 비해 추출(E)과 브라우징(B)이 더 활발하다고 할 수 있다. 이는 취재기자의 경우 기사 발제를 위한 아이디어 수집과 기사 작성을 위해 일상적인 모니터링이 필요하고, 타 사이트로 이동하여 정보를 찾는 조

사 요구가 더 높은 것으로 해석할 수 있다. 세션 시작 시 시간대와 직책별 패턴을 자세히 살펴보면 <표 7>과 같다.

3.2.3 세션 시작 시 웹 행동별 출발 사이트

다음은 세션을 시작할 때 시작하는 웹 행동에 따라 출발 사이트의 차이가 있는지를 <표 8>과 같이 분석하였다.

세션 시작 시 전체적인 출발 사이트를 살펴보면, 네이버가 59.1%를 차지하여 구글 7.8%, 야후 7.6%, 다음 2.6%, 네이트 2.3%에 비해 압도적으로 높은 의존도를 나타냈다. 웹 행동 별로 살펴보면 브라우징(B)으로 시작할 때의 네이버 의존도가 71.8%로 특히 높다는 것을 알 수 있으며, 연결(C)로 시작할 때 야후의 이용도 제법 활발했다. 네이버, 구글, 야후, 다음, 네이트의 5대 사이트가 각각의 웹 행동에서 차지하는 비중을 살펴보면, 추출(E), 즉 검색으로 시작할 때는 100.0% 5대 사이트로 출발하여 검색의 포털 사이트 의존도가 가장 높았다. 반



<그림 2> 직책에 따른 시작 패턴의 차이

〈표 7〉 시간대와 직책별 시작 패턴

단위: 건(백분율)

구분	시간대	모니터링(M)	추출(E)	브라우징(B)	연결(C)	합계
시간대	0시~7시	14 (51.9%)	5 (18.5%)	6 (22.2%)	2 (7.4%)	27 (100.0%)
	7시~10시	81 (83.5%)	4 (4.1%)	4 (4.1%)	8 (8.2%)	97 (100.0%)
	10시~12시	59 (60.8%)	11 (11.3%)	11 (11.3%)	16 (16.5%)	97 (100.0%)
	12시~14시	74 (61.2%)	16 (13.2%)	16 (13.2%)	15 (12.4%)	121 (100.0%)
	14시~16시	90 (58.4%)	34 (22.1%)	16 (10.4%)	14 (9.1%)	154 (100.0%)
	16시~18시	73 (48.7%)	41 (27.3%)	24 (16.0%)	12 (8.0%)	150 (100.0%)
	18시~20시	59 (52.7%)	27 (24.1%)	9 (8.0%)	17 (15.2%)	112 (100.0%)
	20시~24시	60 (52.2%)	27 (23.5%)	17 (14.8%)	11 (9.6%)	115 (100.0%)
	합계	510 (58.4%)	165 (18.9%)	103 (11.8%)	95 (10.9%)	873 (100.0%)
직무	데스크	94 (45.6%)	50 (24.3%)	42 (20.4%)	20 (9.7%)	206 (100.0%)
	취재기자	416 (62.4%)	115 (17.2%)	61 (9.1%)	75 (11.2%)	667 (100.0%)
	합계	510 (58.4)	165 (18.9%)	103 (11.8%)	95 (10.9%)	873 (100.0%)

〈표 8〉 세션 시작 시 웹 행동별 출발 사이트 분포

사이트	모니터링(M)	추출(E)	브라우징(B)	연결(C)	합계
네이버	302 (59.2%)	108 (65.5%)	74 (71.8%)	32 (33.7%)	516 (59.1%)
구글	38 (7.5%)	29 (17.6%)	0 (0.0%)	1 (1.1%)	68 (7.8%)
야후	23 (4.5%)	26 (15.8%)	0 (0.0%)	17 (17.9%)	66 (7.6%)
다음	19 (3.7%)	1 (0.6%)	2 (1.9%)	1 (1.1%)	23 (2.6%)
네이트	18 (3.5%)	1 (0.6%)	0 (0.0%)	1 (1.1%)	20 (2.3%)
5대 사이트 합계	400 (78.4%)	165 (100.0%)	76 (73.8%)	52 (54.7%)	693 (79.4%)

면, 연결(C) 행위의 경우 5대 사이트에서 출발하는 비중이 54.7%에 불과해 세션 시작 시 웹 행동 중 포털 사이트 의존도가 가장 낮았다.

3.2.4 시작과 종료 행동의 관계

웹 정보탐색 시 세션의 시작과 종료 행동의 관계를 분석하기 위해 시작형태 내 백분율과 종료형태 내 백분율을 함께 살펴보았다(표 9 참조).

웹 정보탐색 시 세션의 종료가 어떤 행위로 끝났는지 살펴보면, 전체 873개의 세션 중 브라우징(B)으로 종료한 것이 323건으로 37.0%를 차지하여 가장 많았다. 세션 시작과 종료 행위의 관계를 살펴보면, 모니터링(M)에서 시작하여 모니터링(M)으로 끝난 경우가 32.4%, 브라우징(B)으로 끝난 경우가 33.7%로 M-M와 M-B 형태의 패턴이 비슷하게 많았다. 추출(E)에서 시작했을 때는 추출(E)로 끝난 경우가 48.5%를 차지하여 E-E 형태의 패턴이 가장 많

았으며, 브라우징(B)으로 시작한 경우 브라우징(B)으로 끝난 것이 67.0%에 달해 B-B 형태의 패턴이 가장 많았다. 연결(C)로 시작했을 때는 브라우징(B)으로 끝난 경우가 44.2%로 C-B 형태의 패턴이 가장 많았다.

위의 결과를 종합해 보면, 추출(E)로 시작한 경우를 제외하면 모니터링(M), 브라우징(B), 연결(C) 모두 브라우징(B)으로 끝난 비율이 높으며, 추출(E)로 시작한 경우에만 추출(E)로 끝난 비율이 높다는 것을 알 수 있다.

3.2.5 전체적인 시작 종료 패턴 분포

전체적으로 시작과 종료 패턴을 살펴보면, 모니터링(M)에서 시작하여 브라우징(B)로 끝난 것이 전체 873건의 세션 중 172건(19.7%)로 가장 많았으며, 브라우징(B)으로 시작하여 연결(C)로 끝난 경우는 8건(0.9%)에 불과해 가장 낮은 비율을 나타냈다(표 10 참조).

〈표 9〉 시작과 종료 행동의 관계

시작 행동	구분	종료				
		모니터링(M)	추출(E)	브라우징(B)	연결(C)	합계
모니터링(M)	세션 수	165	114	172	59	510
	시작형태 내 백분율	32.4%	22.4%	33.7%	11.6%	100.0%
	종료형태 내 백분율	74.3%	51.6%	53.3%	55.1%	58.4%
추출(E)	세션 수	22	80	40	23	165
	시작형태 내 백분율	13.3%	48.5%	24.2%	13.9%	100.0%
	종료형태 내 백분율	9.9%	36.2%	12.4%	21.5%	18.9%
브라우징(B)	세션 수	14	12	69	8	103
	시작형태 내 백분율	13.6%	11.7%	67.0%	7.8%	100.0%
	종료형태 내 백분율	6.3%	5.4%	21.4%	7.5%	11.8%
연결(C)	세션 수	21	15	42	17	95
	시작형태 내 백분율	22.1%	15.8%	44.2%	17.9%	100.0%
	종료형태 내 백분율	9.5%	6.8%	13.0%	15.9%	10.9%
전체 세션 수		222	221	323	107	873
전체 시작형태 내 백분율		25.4%	25.3%	37.0%	12.3%	100.0%

〈표 10〉 시작 종료 패턴의 분포

시작-종료 행위	세션 수	백분율
M-B	172	19.7%
M-M	165	18.9%
M-E	114	13.1%
E-E	80	9.2%
B-B	69	7.9%
M-C	59	6.8%
C-B	42	4.8%
E-B	40	4.6%
E-C	23	2.6%
E-M	22	2.5%
C-M	21	2.4%
C-C	17	1.9%
C-E	15	1.7%
B-M	14	1.6%
B-E	12	1.4%
B-C	8	0.9%
총합계	873	100.0%

3.3 단계별 이행 패턴 분석

기자직의 웹 탐색행위가 세션 시작의 1단계에서 2단계로 이행할 때 어떤 행위로 이어졌는지 단계별 이행 패턴을 살펴보았다(표 11 참조).

시작단계인 1단계에서 모니터링(M)으로 시작한 세션을 살펴보면, 총 510건 중 141건(27.6%)이 다시 모니터링으로 이행하여 가장 높은 비율을 차지하였다. 추출(E)로 시작된 세션은 총 165건 중 79건(47.9%)이 다시 추출(E)로 이행한 경우가 가장 많았다. 브라우징(B)으로 시작된 세션은 총 103건 중 57건(55.3%)이 다시 브라우징(B)으로 이행하여 가장 높은 비율을 차지했다. 연결(C)로 시작된 세션에선 총 95건 중 브라우징(B)으로 이행한 것이 25건(26.3%)

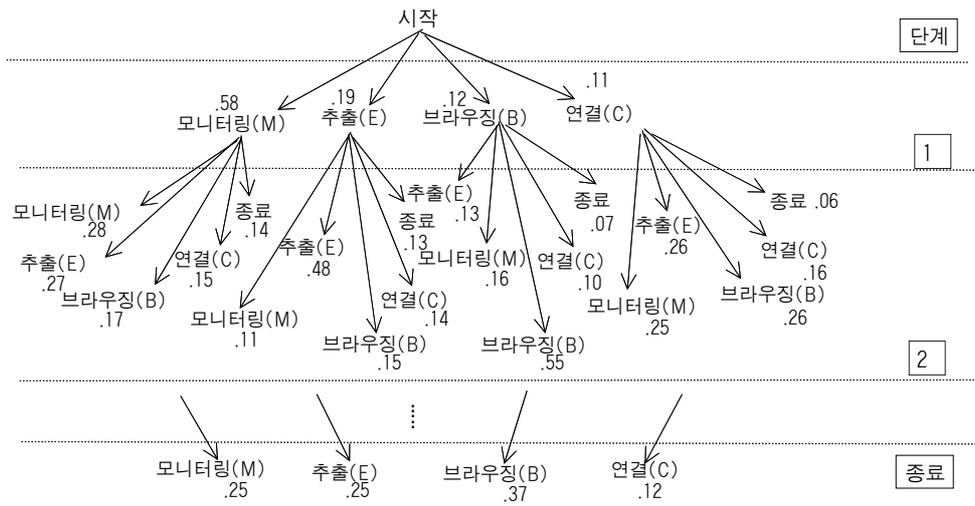
으로 가장 많았다.

위를 결과를 종합해 보면, 전체 873개 세션에서 1-2단계 이행 패턴 중 가장 높은 비율을 차지한 것은 M-M, 즉 모니터링에서 모니터링으로 이행한 패턴(16.2%)이며, M-E, 즉 모니터링에서 추출로 이행한 패턴(15.7%)이 그 다음으로 많았다.

1단계 탐색 행위 안에서 세분화하여 살펴보면, 브라우징(B)에서 시작했을 때 브라우징(B)으로 이행한 패턴이 55.3%로 가장 비율이 높았고, 추출(E)로 시작했을 때 다시 추출(E)로 이행한 패턴이 47.9%로 그 다음으로 비율이 높았다. 본 연구의 결과를 토대로 시작에서 2단계 및 종료까지의 이행 패턴을 확률로 표시하여 모형화하면 〈그림 3〉과 같다.

<표 11> 단계별 이행 패턴 분석

1단계	2단계	세션 수	소계 내 백분율	합계 내 백분율
모니터링(M)	모니터링(M)	141	27.6%	16.2%
	추출(E)	137	26.9%	15.7%
	브라우저(B)	86	16.9%	9.9%
	연결(C)	77	15.1%	8.8%
	종료	69	13.5%	7.9%
	소계	510	100.0%	58.4%
추출(E)	추출(E)	79	47.9%	9.0%
	브라우저(B)	24	14.5%	2.7%
	연결(C)	23	13.9%	2.6%
	종료	21	12.7%	2.4%
	모니터링(M)	18	10.9%	2.1%
	소계	165	100.0%	18.9%
브라우저(B)	브라우저(B)	57	55.3%	6.5%
	모니터링(M)	16	15.5%	1.8%
	추출(E)	13	12.6%	1.5%
	연결(C)	10	9.7%	1.1%
	종료	7	6.8%	0.8%
	소계	103	100.0%	11.8%
연결(C)	브라우저(B)	25	26.3%	2.9%
	추출(E)	25	26.3%	2.9%
	모니터링(M)	24	25.3%	2.7%
	연결(C)	15	15.8%	1.7%
	종료	6	6.3%	0.7%
	소계	95	100.0%	10.9%
합 계		873	-	100.0%



<그림 3> 기자직의 웹 탐색 확률 모형

3.4 순차규칙 모형 분석

본 연구는 앞서 세션의 시작과 종료, 시작 1 단계에서 2단계로의 이행 패턴에 대해 분석 하였으나, 한 세션은 최대 1,132건의 탐색 행위로 연결되어 있어 중간 단계의 탐색 행위 이행 패턴에 대해서는 추가 분석이 필요하다. 이를 위해 본 연구는 웹 데이터마이닝 방법에서 제공하는 연관성 규칙 모형 중 하나인 순차규칙(Sequence) 모형 분석을 시도하였다. 연관성 규칙 모형이란 데이터에 존재하는 항목간의 관계를 살펴보고 이로부터 유용한 관계를 찾아내고자 할 때 이용되는 기법이다. 순차규칙 모형에서는 인스턴스와 규칙지지도, 신뢰도 지수가 사용된다. 순차규칙 모형에서 사용되는 지수와 계산식을 정리하면 다음과 같다(표 12 참조).

본 연구는 순차규칙 모형 분석을 하기 위해 SPSS 파일로 변환한 로그 파일을 PASW Modeler 13.0(구 클레멘타인) 프로그램으로 불러온 후, Session ID, 시간대, 탐색행위의 3개 필드를 사용하여 연관성 규칙 분석을 시행하였다.

유형 노드 설정에서는 각 필드의 데이터 유형을 연속형과 범주형으로 구분하여 설정하였

으며, 각 필드의 방향은 입력과 출력으로 구분하여 설정하였다. 순차규칙의 필드 설정에서는 동일 세션 안에서만 순차 규칙을 찾기 위하여 세션 ID 필드로 동일 세션을 구분하였고, 시간대 필드를 사용하여 시간 순서대로 순차규칙을 설정하였다. 순차규칙을 설정할 때 최대 순차규칙 크기를 설정해야 하는데 최대 탐색행위 수를 10개로 설정하여 분석한 결과, 총 1,912개의 연관성이 추출되었다. 지나치게 많은 규칙이 추출되는 것을 피하기 위해 본 연구에서는 최대 순차규칙 크기를 5개로 제한하였으며, 최소 규칙 지지도와 최소 규칙 신뢰도를 20% 이상으로 설정하였다. 위와 같은 단계를 거쳐 순차규칙 모형 분석을 실시한 결과, 총 915개의 연관성이 추출되었다.

915개의 연관성 규칙은 인스턴스, 규칙지지도, 신뢰도 등 3가지 지수를 기준으로 살펴볼 수 있으며, 순차규칙 분석에서는 특히 신뢰도를 중요시한다. 신뢰도는 순차적으로 발생하는 패턴의 예측에 주안점을 두고 있으며, 예측의 관점에서는 신뢰도가 높을수록 좋은 순차규칙이라고 할 수 있다. 이러한 관점에 따라, 915개의 순차규칙을 신뢰도가 높은 순서로 정렬하였으며, 그 중 신뢰도가 높은 상위 15개의 순차규칙을 도출하면 <표 13>과 같다.

<표 12> 순차규칙 모형에서 사용되는 지수와 계산식

지수	계산식
인스턴스(Instances)	$(A \cap C)$ 가 참인 경우의 거래수
규칙 지지도(Rule Support)	$Pr(A \cap C) = \frac{(A \cap C) \text{ 거래수}}{\text{전체 거래수}}$
신뢰도(Confidence)	$Pr(A C) = \frac{(A \cap C) \text{ 거래수}}{A \text{ 거래수}}$

A: 전항값(Antecenant), C: 후항값(Consequent), 전체 거래수: 적어도 하나의 참값을 가지는 케이스.

〈표 13〉 순차규칙 모형 분석에 의한 탐색행위 패턴

순차규칙	인스턴스	규칙지지도(%)	신뢰도(%)
C-M-B-B-B	265	30.3	93.6
C-M-C-B-B	202	23.1	92.7
B-M-C-B-B	219	25.1	92.4
C-E-B-B-B	206	23.6	92.4
C-E-E-E-E	214	24.5	91.9
C-C-B-B-B	226	25.9	91.1
M-C-C-B-B	214	24.5	91.1
C-E-M-B-B	192	22.0	91.0
C-B-B-B-B	293	33.5	91.0
B-E-E-E-E	250	28.6	90.9
C-C-E-E-E	189	21.6	90.9
M-C-M-M-M	240	27.5	90.6
B-B-B-B-B	373	42.7	90.5
C-E-B-E-E	181	20.7	90.5
B-B-C-B-B	238	27.2	90.5

위에서와 같이 최대 순차규칙 크기를 5개로 제한하고 최소 규칙 지지도와 신뢰도를 20% 이상으로 설정했을 때, 신뢰도가 가장 높은 웹 탐색행위 패턴은 연결(C)-모니터링(M)-브라우징(B)-브라우징(B)-브라우징(B) 패턴으로 나타났다. 이러한 C-M-B-B-B 패턴의 인스턴스, 규칙지지도, 신뢰도 값을 해석해 보면, 전향인 연결(C)-모니터링(M)-브라우징(B)-브라우징(B)과 후향인 브라우징(B)이 동시에 나타나는 거래 수는 265건이며, 이는 전체 거래수 중에서 30.3%를 차지한다. 또한, 전향이 연결(C)-모니터링(M)-브라우징(B)-브라우징(B)일 때 후향이 브라우징(B)으로 이어질 확률은 93.6%인 것으로 해석할 수 있다.

3.5 기자직과 일반인의 웹 이용 비교

기자직과 일반인의 웹 이용 현황을 비교하기 위해 본 연구는 트랜잭션로그 수집 기간과 동

일한 기간의 일반인 이용 현황 통계를 코리안 클릭사에서 함께 입수하였다. 기자직과 일반인의 웹 이용 비교에는 동일한 기준을 적용하기 위해 후처리를 거치지 않은 39,860건의 데이터를 사용하여 비교하였다(표 14 참조).

기자직과 일반인의 주 평균 방문 사이트 수를 비교해 보면, 기자직은 조사 기간 중 평균 39.8개 사이트를 방문하였으며, 일반인의 평균 방문 사이트 수인 32.3개에 비해 7.5개 많은 사이트를 방문하였다. 평균 인터넷 이용일수를 비교해 보면, 기자직은 주 평균 5.49일을 이용했으나 일반인은 5.33일을 이용하여 기자직이 일반인보다 조금 높은 이용일수를 나타내었다. 평균 페이지뷰(PV)는 기자직이 주 평균 1074.6건을 기록한 반면, 일반인은 1145.6건을 기록하여 일반인이 기자직보다 다소 높았다. 이 때, 페이지뷰(PV)란 측정기간 중 해당사이트에 방문한 순방문자들의 총 페이지 조회 수를 말하며, 순방문자(UV: Unique Visitors)는 측정 기간 중 1회 이상 해당

〈표 14〉 기자직과 일반인의 주 평균 웹 이용현황 비교

구분		기자직	일반인	기자직과의 차이
주 평균 이용 통계	방문 사이트 수	39.8	32.3	+7.5
	인터넷 이용일수	5.49	5.33	+0.16
	페이지뷰(PV)	1074.6	1145.6	-71
	체류시간(분)	718.6	616.4	+102.2
	검색 평균 질의어 수	171	53	+118
서비스별 이용시간(분)과 점유율	검색	264,592 (16.6%)	1,813,599 (5.1%)	+11.5%
	커뮤니티	119,058 (7.5%)	4,859,253 (13.6%)	-6.2%
	뉴스	120,360 (7.5%)	2,244,594 (6.3%)	+1.2%
포털사이트 검색 점유율	naver.com	84.2%	66.5%	+17.7%
	yahoo.co.kr	9.0%	3.4%	+5.6%
	google.co.kr	5.5%	2.0%	+3.5%
	daum.net	1.2%	21.9%	-20.7%
	nate.com	0.3%	5.5%	-5.2%

사이트에 방문한 중복되지 않은 방문자를 의미한다.

평균 체류시간(분)은 기자직의 체류시간이 주 평균 718.6분인 반면, 일반인은 616.4분을 나타내 기자직의 인터넷 이용시간이 일반인보다 102.2분 더 많았다.

검색 평균 질의어 수를 비교해 보면, 기자직의 검색 평균 질의어 수가 주 평균 171건인 반면, 일반인은 53건에 불과해 기자직이 일반인에 비해 검색할 때의 평균 질의어 수가 3배 이상 높았다. 서비스별 이용시간(분)과 점유율을 살펴보면, 포털사이트 내 섹션을 기준으로 할 때 총 이용시간 중 검색 서비스의 점유율은 기자직이 16.6%인 반면, 일반인은 5.1%에 불과해 기자직이 일반인에 비해 3배 이상 검색 이용 비중이 높았다. 커뮤니티 서비스의 경우 일반인은 13.6%의 점유율을 보인 반면, 기자직은 7.5%에 불과해 일반인이 기자직보다 커뮤니티

서비스를 훨씬 많이 이용하였다. 뉴스 서비스는 기자직이 7.5%, 일반인이 6.3%로 기자직이 일반인에 비해 뉴스 이용 비중 다소 높은 것으로 나타났다.

기자직과 일반인이 검색 시 어떤 포털 사이트를 주로 가는지 포털사이트 검색점유율을 통해 비교해 보았다. 포털사이트 검색 점유율이란 이용자가 질의어를 입력하여 검색을 실행할 때 이용한 포털 사이트의 비율을 의미한다. 포털사이트의 검색 점유율을 비교해 보면, 기자직의 경우 검색을 할 때 네이버에서 검색하는 비율이 84.2%인 반면, 일반인은 66.5%여서 기자직의 네이버 이용도가 일반인에 비해 훨씬 높았다. 야후의 경우 기자직이 9.0%, 일반인이 3.4%로 기자직이 5.6% 더 높았으며, 구글 역시 기자직이 5.5%, 일반인이 2.0%로 기자직이 3.5% 높았다. 이에 비해 다음 사이트는 기자직이 검색할 때 1.2% 밖에 이용하지 않는 반면,

일반인은 21.9% 이용률을 나타내 기자직의 이용률이 일반인의 이용률보다 20.7%나 낮아 큰 차이를 보였다. 네이트의 경우엔 기자직과 일반인의 검색 이용도가 모두 낮은 수준이었으며, 기자직이 0.3%, 일반인이 5.5%로 기자직의 이용률이 일반인의 이용률보다 5.2% 낮았다.

포털사이트의 검색 점유율에서 기자직과 일반인의 차이가 가장 큰 것은 다음 사이트인데, 그 이유를 분석해보면 다음 뉴스에선 주요 일간지 기사가 제공되지 않기 때문에 뉴스 검색을 자주 하는 기자직의 경우 검색 시 다음 사이트의 선호도가 매우 낮은 것으로 해석된다.

4. 기자직의 웹 정보탐색행위 패턴

4.1 Choo 등(1998) 연구의 적용

Choo 등(1998)은 그들이 수행한 일련의 연구(1998, 1999, 2000, 2003)에서 웹 정보탐색행위의 유형을 비의도 탐색(undirected viewing), 의도 탐색(conditioned viewing), 비공식 검색(informal search), 공식 검색(formal search)의 4가지 유형으로 구분하고 이를 웹 정보탐색행위에 적용하였다.

Choo 등(1998)에 따르면, 비의도 탐색은 특정 목표나 정보요구 없이 웹을 탐색하다가 뜻밖의 정보를 접하는 탐색이라 정의하였으며, 탐색행위는 주로 시작과 연결 행위가 나타난다고 보았다. 의도 탐색은 이용자가 선택한 주제에 대한 정보를 찾거나 특정 유형의 정보를 보기 위해 웹을 사용하는 의도적 탐색이라 정의

하였고, 탐색행위는 주로 브라우징, 차별화, 모니터링이 나타나는 것으로 보았다. 비공식 검색의 경우 특정 주제에 대한 지식과 이해를 깊게 하기 위해 정보를 검색하는 경우로 정의되었으며 탐색행위는 차별화, 모니터링, 추출이 나타난다고 제시하였다. 공식 검색은 특정 주제와 특정 형태의 정보를 얻기 위해 구체적인 계획을 세워 비교적 상세한 정보를 찾는 경우로 정의되며 탐색행위는 모니터링, 추출이 주로 나타난다고 분석하였다. Choo 등(1998)이 제시한 유형 구분을 본 연구에 적용함에 있어서 유형 구分的 쟁점은 탐색과 검색, 의도와 비의도를 어떤 기준으로 구분하느냐의 문제이다.

본 연구는 탐색과 검색의 구분을 한 세션 안에서 추출(E) 행위의 유무로 구분하였으며, 탐색 안에서의 비의도와 의도 구분은 한 세션 안에서 모니터링(M)의 유무로 구분하였다. 여기에서 추출(E) 행위란 검색 질의어가 존재하는 경우를 말하여, 모니터링(M) 행위는 직접 주소를 입력하고 들어온 경우, 즉 유입경로와 검색 질의어가 없는 로그를 뜻한다. 공식검색과 비공식 검색의 구분은 브라우징(B) 행위의 유무로 설정하였다.

이와 같은 기준으로 Choo 등(1998)의 유형 구분을 본 연구에 적용한 결과, 기자직의 웹 정보탐색행위에서 가장 많이 나타난 유형은 공식 검색으로 총 873건의 세션 중에서 493건(56.5%)을 차지하였고, 탐색 행위는 브라우징, 연결, 모니터링, 추출이 모두 나타났다. 그 다음은 의도 탐색으로 227건(26.0%)을 차지하였으며, 탐색행위는 브라우징, 연결, 모니터링이 나타났다. 비공식 검색의 경우엔 106건(12.1%)으로 연결, 모니터링, 추출 행위가 발생하였다. 이에 반해

비의도 탐색은 47건으로 5.4%에 불과했으며, 브라우징과 연결 행위가 나타났다.

4.2 기자직의 웹 정보탐색행위 유형

앞에서와 같이 Choo 등(1998)의 연구를 본 연구에 적용한 결과, 기자직의 공식 검색에 대한 높은 선호도는 확인되었으나, 웹 정보탐색 시 한 세션이 얼마나 짧게, 혹은 길게 지속되었는지에 대한 고려가 부족하다는 문제점이 발견되었다. 본 연구는 트랜잭션로그 분석에 기초한 웹 정보탐색행위 패턴 연구에서 세션의 길이가 매우 유용한 구분 기준이 된다고 판단하고 새로운 유형 구분을 시도하였다.

유형 구분의 기준은 탐색과 검색, 세션 길이의 장단이며 이러한 기준에 따라 사실확인형 탐색, 조사형 탐색, 사실확인형 검색, 조사형 검색 등 4가지의 유형을 도출하였다. 탐색과 검색 구분의 기준은 한 세션 안에서 추출(E) 행위의 유무로 구분하였고, 사실확인형과 조사형의 구

분은 한 세션 당 탐색 행위의 건수를 기준으로 구분하였다. 이와 같은 기준을 적용하여 4가지 유형을 정의해 보면 다음 <표 15>와 같다.

구분의 기준이 되는 탐색 행위 건수는 10건으로 정했는데, 그 이유는 탐색 행위가 10건 이내로 짧게 끝난 세션이 전체 873개 중에서 429건으로 중간 값에 가깝기 때문이다. 이와 같은 유형 구분 방법을 바탕으로 본 연구의 결과를 적용해 보면 <표 16>과 같다.

본 연구의 기자직 웹 정보탐색행위 패턴에서 가장 많이 나타난 정보탐색 유형은 사실확인형 검색으로 367건(42.0%)를 차지하였고, 그 다음은 조사형 검색 232건(26.6%), 사실확인형 탐색 197건(22.6%), 조사형 탐색 77건(8.8%) 순으로 나타났다. 즉, 간단한 정보를 검색하기 위해 10회 이하의 탐색행위로 세션을 끝내는 패턴이 가장 많았으며, 그 다음으로는 특정 정보를 찾기 위해 브라우징, 연결, 모니터링, 추출 등 다양한 행위로 이루어진 조사형 검색 패턴이 많았다. 검색 행위가 수반되지 않으면서 10회

<표 15> 기자직의 웹 정보탐색행위 유형 정의

유형	정의
사실확인형 탐색	새로운 정보를 발견하거나 특정 정보를 얻기 위해 브라우징 중심으로 비교적 짧게 정보를 찾는 유형
조사형 탐색	새로운 정보를 발견하거나 특정 정보를 얻기 위해 브라우징 중심으로 비교적 길게 정보를 찾는 유형
사실확인형 검색	새로운 정보를 발견하거나 특정 정보를 얻기 위해 질의어 검색 중심으로 비교적 짧게 정보를 찾는 유형
조사형 검색	새로운 정보를 발견하거나 특정 정보를 얻기 위해 브라우징과 질의어 검색 등을 이용하여 비교적 길게 정보를 찾는 유형

<표 16> 기자직의 웹 정보탐색행위 유형

유형	세션 수	브라우징	연결	모니터링	추출
사실확인형 탐색	197(22.6%)	✓	✓	✓	
조사형 탐색	77(8.8%)	✓	✓	✓	
사실확인형 검색	367(42.0%)	✓	✓	✓	✓
조사형 검색	232(26.6%)	✓	✓	✓	✓

이하로 간단한 탐색을 하는 사실확인형 탐색 패턴은 조사형 검색 패턴보다 적게 나타났으며, 가장 건수가 적은 패턴은 검색 행위 없이 10회가 넘는 탐색행위로 이루어진 조사형 탐색이었다.

5. 결론

본 연구는 기자직을 대상으로 웹 정보탐색행위의 특성을 규명하고자 하였으며 이를 위해 23명의 취재기자를 대상으로 2주 동안 39,860개의 웹 로그를 수집하여 트랜잭션로그 분석을 실시하였다.

트랜잭션로그 분석 결과, 가장 많이 나타난 탐색행위는 브라우징(47.5%), 추출(23.5%), 모니터링(20.4%), 연결(8.6%) 순으로 나타났다. 30분을 기준으로 세션을 구분할 때 873개의 세션이 도출되었고, 세션의 시작 패턴은 모니터링(58.4%), 추출(18.9%), 브라우징(11.8%), 연결(10.9%) 순으로 나타났으며, 시작 패턴은 시간대, 업무에 따라 차이가 있었다. 세션의 시작과 종료 패턴을 살펴보면, 모니터링(M)에서 시작하여 브라우징(B)로 끝난 것이 19.7%로 가장 많았다. 단계별 이행 패턴을 살펴보면, 1-2단계 이행 패턴 중 모니터링(M)에서 모니터링(M)으로 이행한 패턴이 16.2%를 차지하여 가장 높은 비율을 차지하였다. 전체적인 탐색행위 이행 패턴을 분석하기 위해 본 연구는 순차규칙 모형 분석을 실시하였다. 최대 순차규칙 크기를 5개로 제한하고, 최소 규칙 지지도와 신뢰도를 20% 이상으로 설정하였을 때 총 915개의 연관성이 도출되었다. 이 중에서 신뢰도가 가장 높은 패턴은 C-M-B-B-B 패턴으로, 연결(C), 모

니터링(M), 브라우징(B), 브라우징(B), 브라우징(B)이 순서대로 나올 확률이 93.6%였다. 기자직과 일반인의 웹 이용 패턴을 비교해 보면, 평균 방문 사이트 수, 평균 인터넷 이용일수, 평균 체류시간(분)은 기자직이 일반인보다 높았으며, 평균 페이지뷰(PV)는 일반인이 기자직보다 다소 높았다. 검색 평균 질의어 수를 비교해 보면, 기자직이 일반인에 비해 평균 질의어 수가 3배 이상 높았다. 포털사이트 내 섹션을 기준으로 할 때 총 이용시간 중 검색 서비스의 점유율은 기자직이 일반인에 비해 3배 이상 높았다.

위와 같은 트랜잭션로그 분석 결과를 토대로 기자직의 웹 탐색행위 패턴을 검색 질의어의 유무, 세션 길이를 기준으로 사실확인형 탐색, 조사형 탐색, 사실확인형 검색, 조사형 검색 등 4가지 유형으로 구분하였다. 4가지 유형에서 가장 많이 나타난 유형은 사실확인형 검색(42.0%)이었으며, 그 다음으로 조사형 검색(26.6%), 사실확인형 탐색(22.6%), 조사형 탐색(8.8%) 순으로 나타났다.

본 연구는 국내 이용자연구에 있어서 연구가 미미했던 기자직을 대상으로 웹 정보탐색행위 패턴을 트랜잭션로그 분석을 통해 연구함으로써, 비학술 전문 집단의 웹 정보탐색행위에 관한 실증적 데이터를 제공하였다는 점에서 의의를 갖는다.

본 연구를 통해 드러난 기자직의 웹 정보탐색행위 특성은 적극적 정보소비자이자 정보생산자의 특성을 공유하는 콘텐츠 생산자들, 예를 들면 PD, 드라마 구성 작가, 블로거 등 다양한 직군에 적용될 수 있을 것이다. 향후 더욱 다양한 전문 집단을 대상으로 후속 연구가 이어지길 기대한다.

참 고 문 헌

- 권혜진. 1999. 『인터넷을 활용한 취재보도에 관한 연구』. 석사학위논문, 이화여자대학교 정보과학대학원.
- 김성진. 2006. 웹 이용자의 정보탐색행위 패턴 분석. 『정보관리학회지』, 23(4): 197-214.
- 박소연, 이준호. 2002. 로그 분석을 통한 이용자의 웹 문서 검색 행태에 관한 연구. 『정보관리학회지』, 19(3): 111-122.
- 박일중, 백수연. 2001. 전자공학전공 대학원생들의 학술정보 이용행태 분석. 『한국도서관·정보학회지』, 32(3): 31-58.
- 이정연. 2008. 영화창작자의 정보활동모형 설계에 관한 질적 연구. 『한국문헌정보학회지』, 42(4): 419-439.
- 이해영, 이수영. 2001. 인터넷 정보의 탐색, 평가 및 활용: 대학 이공계 연구자의 사례를 중심으로. 『정보관리학회지』, 18(4): 163-181.
- 임석중. 1998. 『기자들의 WWW 탐색엔진 이용에 관한 연구』. 석사학위논문, 중앙대학교 대학원.
- 장선화. 2005. 취재환경 변화에 따른 기자의 정보추구행태 연구. 『한국문헌정보학회지』, 39(3): 165-182.
- 정동열, 김성진. 2003. 『이론정보학』. 서울: 한국도서관협회.
- 최은주. 1996. 사회과학 연구자의 정보이용행태에 관한 연구. 『한국문헌정보학회지』, 30(4): 13-38.
- _____. 1997. 자연과학 및 사회과학 연구자들의 정보이용특성 분석. 『정보관리학회지』, 14(1): 27-45.
- Abdulla, Abdulnasir Taleb Mansoor. 2006. *Technology and Internet-related Information Behaviors of Print Journalists in Kuwait*. Ph.D. diss., University of Pittsburgh.
- Anwar, M. A., Husain, Al-Ansari, and A. Abdullah. 2004. "Information seeking behaviour of kuwaiti journalists." *Libri*, 54(1): 228-236.
- Attfield, Simon and John, Dowell. 2003. "Information seeking and use by newspaper journalists." *Journal of Documentation*, 59(2): 187-204.
- Campbell, Foina. 1997. "Journalistic construction of news: information gathering." *New Library World*, 98(1133): 60-63.
- Chen, Hui-Min and M. D. Cooper. 2002. "Stochastic modeling of usage patterns in a web-based information system." *Journal of the American Society for Information science and Technology*, 53(7): 536-548.
- Chinn, L. A. 2001. *The Information Seeking Behavior and Needs of Journalists in Context*. MSLS diss., Univ of NC at Chapel Hill.
- Choo, Chun Wei and Christine, Marton. 2003. "Information seeking on the web by

- women in it professions." *Internet Research*, 13(4): 267-280.
- Choo, Chun Wei, Brian. Detlor, and Don. Turnbull. 1998. "A behavioral model of information seeking on the web: Preliminary results of a study of how managers and IT specialists use the web." *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, Pittsburgh, PA.
- Choo, Chun Wei, Brian. Detlor, and Don. Turnbull. 1999. "Information seeking on the web: An integrated model of browsing and searching." *Proceedings of the 62nd Annual Meeting of the American Society for Information Science*, Washington, DC.
- Edem, U. S. 1993. "Information needs and information seeking behavior patterns of journalists in selected Nigerian towns." *Journal of Library and Information Science*, 19(2): 1-14.
- Fabritius, H. 1999. *Triangulation as a Multi-Perspective Strategy in a Qualitative Study of Information Seeking Behaviour of Journalists: Exploring the Contexts of Information Behaviour*, Taylor Graham Publishing London, UK.
- Harman, Jennifer. 1986. "Reuters: A survey of End-LTser searching." *Aslib Proceedings*, 38(1): 35-42.
- Herron, Nancy Lee. 1986. *Information Seeking Behavior and the Perceptions of Information Channels by Journalists of two Daily Metropolitan Newspapers*, Ph.D. diss., University of Pittsburgh, School of Library and Information Science.
- Hölscher, C. and G. Strube. 2000. "Web search behavior of internet experts and newbies." *Computer Networks*, 33(1): 337-346.
- Hsieh-Yee, I. 2001. "Research on web search behavior." *Library and Information Science Research*, 23: 167-185.
- Joseph, R. 1993. "How Indian journalists use libraries?" *Information Development*, 9(1): 70-75.
- Nicholas, D. and H. Martin. 1993. "End-user coming of age? Six years of end-user searching at the Guardian." *Online Review*, 17(2): 83-89.
- Spink, Amanda. 2002. "A user-centered approach to evaluating human interaction with web search engines." *Information Processing and Management*, 38(3): 401-426.
- Vreekamp, H. 1995. "The information seeking attitudes of non-metropolitan journalists: A qualitative study of two communities and their primary and secondary sources." *The Electronic Library*, 13(1): 43-51.
- Wang, Peiling, B. H. William, and Carol. Tenopir. 2000. "Users' interaction with world wide web resources: an explo-

ratory study using a holistic approach.”
Information Processing and Manage-

ment, 36(2): 229-251.