

웹로그 분석을 통한 의약품 정보 검색 주제별 이용 패턴에 관한 연구

A Study on the Usage Patterns of Medicine Information Through Web Log Analysis

조경원, 우영운*
 춘해대학, 동덕대학교*

Cho Kyoung-Won, Woo Young-Woon*
 Choonhae College, Dongdui University

요약

최근에 인터넷을 통한 의약품 정보가 범람하고 있다. 그러나 지금까지 일반인들이 웹페이지를 통하여 의약품 정보를 검색하는 패턴과 획득하는 방법에 대한 연구 결과가 미비한 실정이다. 따라서 이 논문에서는 의약품 정보를 제공하는 특정 회사의 웹 로그 파일을 WiseLog 툴을 사용하여 분석하였다. 특히 사용자 유형에 따른 웹페이지 사용 형태와 웹페이지 메뉴 사용 형태, 그리고 검색 메뉴 사용 형태의 3 가지에 대한 웹 로그 파일 통계 결과 자료를 분석하였다. 그 결과 인터넷에서의 의약품 정보를 제공하는 회사들을 위한 보완 및 개선 방안들을 제시하였다.

Abstract

There are lots of medicine information on the internet recently. But there is no specific research result about search patterns or acquisition methods of medicine information on web pages for lay people until now. In this paper, I analyzed the web log files of a certain company providing medicine information using the WiseLog tool. I analyzed three kinds of statistic result of the web log files such as the status of web page usage by types of users, the status of web page menu usage, and the status of search menu usage. As results, I proposed methods of supplement and improvement for companies providing medicine information on the internet.

I. 서론

인터넷이 전 세계 지식 창고의 역할을 훌륭하게 수행함에 따라 이를 통하여 숨은 정보가 교류되고 있으며, 그 부산물인 웹 로그 역시 분석을 필요로 하는 방대한 양의 데이터로 존재하게 되었다[1]. 또한 e-business가 공공 서비스, 전자 상거래, 포털 및 콘텐츠 서비스, 기업의 홍보에 이르기까지 다양한 분야에서 web를 통하여 이루어지고 있다. 이것은 병원과 의료계에서도 예외일 수 없다. 2000년 의약분업 실시에 따라 의료 환경과 의약 정보의 수요가 급속도로

증가하는 변화가 있으며 이에 대응하여 많은 의약정보 제공기관에서 웹사이트를 통해 의약정보를 제공하고 있다. 인터넷을 이용하면 비교적 저렴한 비용으로 많은 사람들이 동시에 언제, 어디서나, 어떤 시간 이든지 정보를 의약정보 검색이 가능하기 때문이다. e-health의 구성요소로서 e-customer가 큰 부분을 차지하며 병원CRM에서의 핵도 의료의 콘텐츠가 된다. 김지희[2]가 연구한 의약정보제공기관의 질의분석을 통한 정보수용의 평가를 보면 1997년 4월부터 2000년 9월까지 제공된 618건의 정보제공에 대한 평가한 결과, 의약정보질의자가 초기에 병원약사가 주

를 이루던 [병원약사(78.6%), 약국약사(16.7%)] 것에 반해 후기로 갈수록 약국약사의 비율이 상승[병원약사(33.0%), 약국약사(37.0%)]되었으며 일반인 질문자도 증가되고 계층이 다양화 되었다. 따라서, 효과적으로 의약정보제공을 하는 사이트를 구성하기 위해서는 이들 사용자가 의약정보제공사이트에 대해 어떻게 생각하고 있는지에 대한 정기적인 수요조사가 필수적이다. 사용자들이 자주 방문하고, 자주 이용하는 과정은 무엇인지, 사용자들이 거의 방문하지 않는 과정은 무엇인지, 웹 사이트의 메뉴 중 선호되는 메뉴는 무엇인지 등 이들의 요구는 의약정보제공 사이트의 개선에 많은 도움이 될 수 있다. 그러나, 기존의 의약정보수요조사는 전적으로 설문에만 의존하고 있는바 사용자가 실제 생각하고 있는 내용이 정확히 반영이 되었는지 파악이 힘들고, 한정된 질문 안에서 한정된 정보파악만이 가능하다.

이러한 배경 하에 최근에 급부상하고 있는 웹 로그 파일 분석을 매우 시기 적절한 분석 방법이라 할 수 있다. 왜냐하면 일반적으로 웹 사이트를 방문한 사람들의 방문 데이터에는 누가 언제 무엇을 요청했고 무엇을 가져갔는지, 웹 서버에 얼마나 많은 사람이 왔는지, 어디에서 왔으며 무엇을 제일 좋아하는지, 무엇을 싫어하는지, 가장 오래 보는 또는 가장 많이 보는 페이지는 무엇인지 등등의 방문자 기록이 정리된 자료가 제공되기 때문이다. 특히 웹 로그 파일분석은 최근에 새로운 화두로 떠오른 데이터 마이닝 개념과 결합되어 웹 마이닝으로 불리기도 한다. 이 연구에서는 의약정보제공사이트의 웹 로그를 분석하여 웹 로그가 갖는 의미를 살펴보고 이를 바탕으로 의약정보 제공사이트의 효과적인 정보제공을 위한 전략을 제시하고자 한다.

II. 연구 배경

1. 로그파일

고객이 웹사이트에 들어와서 클릭하는 모든 데이터

는 웹 서버의 로그에 기록된다. 기록되는 내용으로는 접속한 사용자의 IP주소, 접속한 날짜와 시간, 데이터 전송을 위해 사용된 HTTP 방법과 프로토콜, 전송된 문서의 경로, 전송된 데이터의 바이트 수, 전송상태 등의 정보가 기록된다[3]. 로그파일은 웹서버가 지정하는 곳에 위치하며, 웹서버관리자가 웹서버를 설치할 때 로그파일을 생성할 수도 있는데, transfer, access, error, referrer 및 agent 로그파일 등이 있다 [4]. 로그 분석은 분석 목적에 따라 단지 로그 정보를 분석하는 것에 한정시키기도 하고 로그 정보를 기반으로 한 보다 다양한 정보를 분석하는 확장된 개념으로 확대시키기도 한다[5]. 웹로그 분석을 통하여 방문자들에 대한 접속 유희시간 측정, 상품과 서비스를 교차한 교차 판매 계획 수립, 캠페인 효과에 대한 평가, 자료 배치를 위한 웹 공간의 효율적인 논리적 구조 등의 결과를 얻을 수 있다[6].

2. 웹마이닝

데이터마이닝이란 방대한 양의 데이터 속에서 직접, 혹은 간접적인 방법으로 의미있는 패턴이나 규칙을 탐사하고 분석하는 방법이다. 이러한 데이터마이닝의 기법들은 웹에서 정보를 탐색하고 구성하는데 이용할 수 있다. 즉, 웹으로부터 유용한 정보를 발견하고 분석하는 것을 총칭하여 웹마이닝이라 한다. 웹마이닝에 관한 연구가 다수 이루어지고 있는데, Cooley(1999)등은 웹사용자들의 데이터를 사용자, 콘텐츠, 구조의 세가지로 분류하고 이들을 로짓함수를 이용하여 패턴을 추출한 결과 사용자의 패턴을 분석하여 가중치가 높은 웹페이지를 분류하였다[7][8].

웹에서 사용자의 패턴을 분석하여 이들의 연관관계를 분석하고 이를 전략적으로 활용하려는 연구도 이루어지고 있다. Mobasher(1999)등은 사용자가 자주 방문하는 사이트를 중심으로 연관관계를 분석하는 새로운 매키니즘을 제안하기도 하였다[9].

III. 연구 방법

본 연구에서 사용되는 틀은 Nethru사에서 출시한 WiseLog Enterprise를 활용하여 웹 로그를 분석하였다.

WiseLog는 웹사이트 방문 고객에 대한 분석을 통해 e-Business 전략을 수립할 수 있도록 지원하는 웹고객 분석 패키지로서, 다양한 조건의 타겟팅 대상 고객 추출 및 사용자 세그멘테이션을 하며 다양한 시스템 환경에 반영되기 때문에 비전문가도 쉽게 운영할 수 있다는 장점이 있다. 본 연구의 틀은 [그림 1]과 같다.

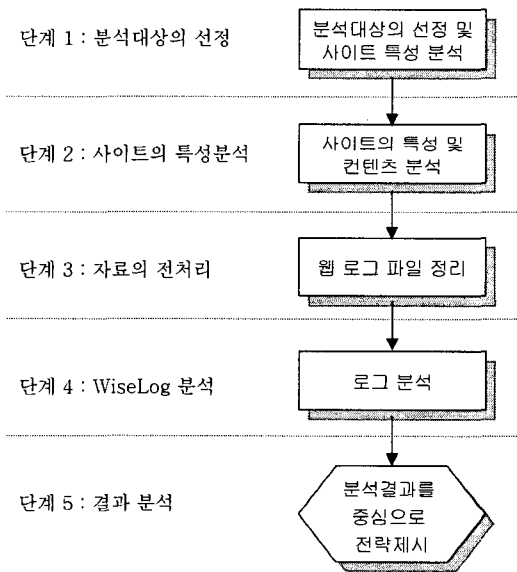


그림 1. 연구의 틀

1. 대상 선정(Log Gathering)

의약정보제공 웹사이트의 로그 파일을 중심으로 이를 분석하기 위하여 우선 분석대상을 선정해야 하는데 본 연구에서는 한 의약정보제공사이트를 대상으로 2주일동안 접속한 로그파일을 사용하였으며 이를 WiseLog를 활용하여 웹로그 분석을 실시하였다. 연구에 사용된 사이트는 검색뿐 아니라, 쇼핑, 뉴스/학

술, 건강정보, 구인/구매 핵심 인터넷서비스를 한데 모은 사이트이다. 분석자료는 2005년5월24일부터 6월6일까지의 2주일치 자료를 대상으로 하였다. 아래의 [표 1]에는 각 일자별 페이지뷰 수가 제시되어 있다.

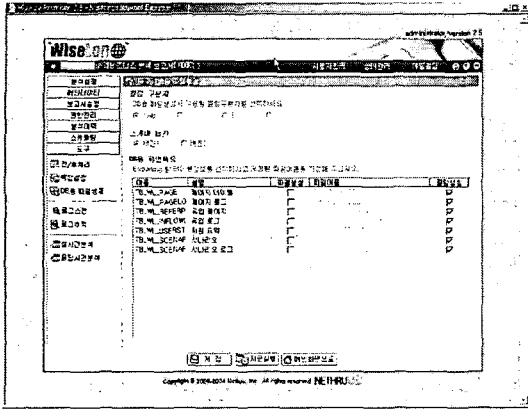
[표 1] 각 일자별 페이지뷰 수

시간	페이지 뷰	회원당 페이지 뷰
2005-05-24	263,893	36
2005-05-25	255,844	37
2005-05-26	253,515	36
2005-05-27	231,694	35
2005-05-28	143,289	33
2005-05-29	85,059	45
2005-05-30	254,627	35
2005-05-31	246,689	35
2005-06-01	255,504	36
2005-06-02	254,841	36
2005-06-03	248,409	36
2005-06-04	150,805	33
2005-06-05	74,993	44
2005-06-06	113,156	39
2005-06-07	262,282	36
합계	3,094,602	36

2. 사이트의 특성분석

효율적인 전략을 제시하기 위해서는 먼저 의약정보 제공 사이트의 특성을 면밀히 분석하는 것이 중요하다. 대상 사이트의 특징은 정보를 검색하기 위해서 반드시 회원등록이 이루어져야 하며 회원으로 등록하면 의약정보 검색을 비롯한 건강정보, 의약계 뉴스, 의약마케팅서비스 등을 동시에 이용할 수 있다. 의사, 약사 등 의료계종사자를 위한 정보가 다수이나 소비자 의료정보 활용이 많아지고 있는 현실이므로 이 사이트의 방문자 속성별 분석이 필수적이며 전략적으로 활용할 핵심적인 콘텐츠를 중심으로 이들간의 연계성을 분석하는 것이 중요하다.

3. 자료의 전처리 및 분석 설정



▶▶ 그림 2. 웹로그분석 설정화면

자료선택에 있어서, 트랜잭션에서 발생한 자료에서 이미지 파일, include파일(css 등)은 의미가 없으므로 분석에 포함하지 않았다. 프로파일 제목, 기본 도메인을 지정하고 웹서버에서 지정한 기본페이지를 그대로 사용하였다. 로그 파일 정보, 서버 이름, 해당 로그파일을 생성한 도메인을 설정하였다.

세션유지시간을 30분으로 한정하여 그 시간동안 히트 요청이 없으면 방문을 마친 것으로 간주하도록 하였으며 본 사이트가 회원제사이트이므로 회원ID는 쿠키에서 추출하는 방식으로 설정하였다.

IV. 실험 결과

방문자 유형에 따른 페이지 이용 현황을 분석한 결과는 [표 2]에 제시되었다. [표 3]의 결과는 의약품에 종사하는 비의료인의 페이지 뷰와 의료인으로 진출하기 위한 관련 학생들의 경우를 제외한 결과이다. 이 결과에서 알 수 있듯이 3가지 유형에 대한 일반인의 페이지 뷰 비율이 약 29% 정도로 상당히 높음을 알 수 있었다.

페이지별 방문수는 사이트의 첫 화면이 가장 많지만 페이지뷰 수로 보면, 의약품 검색이 가장 많은 접속을 나타냈고, 사이트 홈, 뉴스/학술, 건강정보, 구인

[표 2] 방문자 유형에 따른 페이지 이용 현황

방문자 유형	페이지 뷰
의사	784,061
약사	287,760
일반인	435,747

/구매, 의약품마케팅 등의 순으로 나타났다. 이는 의약 정보 제공 사이트의 특성상 의약품검색을 주목적으로 방문한다는 것이다. [표 3]의 메뉴 이용 현황을 보면 알 수 있듯이, 대부분의 회원 수가 중복되지만 전체 페이지 뷰 중에 의약품검색 페이지가 전체의 약 65%정도를 차지하고 있다.

[표 3] 메뉴 이용 현황

메뉴분석	페이지 뷰
의약품 검색 페이지	729,112
검색 외 페이지	397,146

의약 정보 검색 메뉴별로 이용자들의 이용 현황을 분석한 결과는 [표 4]에 제시되었다. [표 5]에서 알 수 있듯이, 이용자들은 의약품에 대한 정보를 얻기 위하여 웹 페이지를 가장 많이 이용하는 것으로 나타났으며, 전체 페이지뷰의 약 64%를 제품 정보 검색에 할애하고 있음을 알 수 있었다. 특히 제품의 상세 정보보다 기본 정보만을 획득하고 있는 것으로 나타났으며 전체 페이지뷰의 약 54%를 기본 정보 획득에 할애하고 있음을 알 수 있었다.

[표 4] 의약품정보 검색 메뉴 이용 현황

순서	메뉴폴더	페이지 뷰
1	제품	466,557
	기본정보	395,152
	상세정보 외	71,405
2	성분	91,389
3	종합검색	143,072
4	의약품 식별	22,141
5	복약지도 검색	1,972
6	상호작용 검색	1,149
7	신장애약용량 검색	713
8	신제품 검색	1,473
9	수정요청	220
10	제품정보 올리기	426

V. 결과 고찰

첫 번째 제시한 방문자 유형에 따른 페이지 이용 현황 분석 결과에 따르면 의사나 약사들 이외에 의약품업과 전혀 관계가 없는 일반인들의 의약품 정보 사이트의 페이지 뷰가 약사들보다도 높은 것으로 나타났다. 이는 의약품 분업 이후에 일반인들이 처방전을 쉽게 접하게 되고 그에 따라 자신이나 가족이 복용하는 의약품에 대한 정보 욕구가 강해졌기 때문이라고 판단된다. 따라서 일반인들의 정보 획득 욕구에 알맞은 사용 편의성을 갖춘 웹 페이지 운영이 향후 필요할 것으로 예상된다.

두 번째 제시한 메뉴 이용 현황 분석 결과에 따르면 이용자들이 의약품정보 제공 사이트를 이용할 경우 메인 페이지나 기타 정보 제공 페이지보다 의약품 검색 페이지를 약 2배정도 많이 액세스를 하고 있다는 것을 알 수 있었다.

의사들은 처방을 내리기 위하여 의약품 정보를 획득하기 위해 웹 페이지를 많이 사용하고 있는 것으로 판단할 수 있으면 일반인들은 관심 있는 의약품에 대한 정보를 웹 페이지를 통해 많이 획득하고 있음을 알 수 있었다. 따라서 의료인이든 일반인이든 의약품 검색을 더욱 신속하고 편리하게, 그리고 정확한 정보를 획득할 수 있도록 웹 페이지에서 제공되는 정보의 질을 향상 시킬 필요가 있을 것으로 예상된다.

세 번째 제시한 의약품 정보 검색 메뉴 이용 현황 분석 결과에 따르면 이용자들이 의약품의 기본 정보만을 보는 경우가 기타 정보를 보는 것보다 많음을 알 수 있었다. 이 결과는 처방을 내리기 위한 의사들은 물론 일반인들도 웹 페이지를 통하여 의약품 정보를 얻고자 하는 경우에 기본 정보에만 의존해서 약의 효능과 부작용을 신속히 파악하고자 하기 때문인 것으로 파악된다.

따라서 의약품 정보를 제공하는 업체들은 기본 정보의 정확성에 더욱 만전을 기하여 오용과 남용의 문제가 발생하지 않도록 할 필요가 있다. 또한 기본 정보들에 언급되는 전문 용어들을 일반인들도 쉽게 알

수 있도록 용어 해설에 관련된 페이지를 함께 제공하여 일반인들의 정보 접근성을 높일 수 있는 방향으로의 개선이 필요할 것으로 예상된다.

VI. 결론

이 논문에서는 특정 의약품정보 제공 사이트의 웹 로그를 WiseLog라는 툴로써 분석한 후 획득된 여러 가지 통계 자료들 중 3가지 부분에 대한 통계 자료 결과를 제시하고 그 결과를 분석하였다. 첫 번째가 방문자 유형에 따른 페이지 이용 현황 분석이며 두 번째가 메뉴 이용 현황 분석이며 세 번째가 의약품 정보 검색 메뉴 이용 현황 분석이다.

세 가지 부분에 대하여 분석 결과를 토대로 하여 의약품 정보 제공 업체들이 향후 개선해야 할 점과 보완해야 할 점을 제시하였다.

향후 연구 과제로는 웹 로그 분석 결과를 데이터베이스화하고 사용자 데이터베이스와 연동하여 개별 이용자들이 대한 의료 정보 수요를 파악하고 그에 따른 능동적 의료 정보 제공 메커니즘을 구축하여 더욱 지능적인 의료 정보 제공 사이트를 구현할 필요가 있을 것으로 예상된다.

■ 참고 문헌 ■

- [1] 김석기, 안정용, 한경수, 한범수, "웹 로그(Web Log) 분석을 통한 정보의 활용", 한국통계학회 학술발표회논문집, pp.123-127, 2000.
- [3] 변시우, "인터넷 비즈니스를 위한 웹로그 분석", 2001 한국경영과학회 학술대회 논문집, pp.256-259.
- [4] 손용배, 이용봉, "웹마이닝을 통한 도서관 홈페이지의 사용편의성에 관한 연구", 제8회 한국정보관리학회 학술대회 논문집, pp.213-218.
- [5] Kim, J.K., Lee, K.C, Chung, N.H., Kwon, S.J., Cho, Y.H., "Analysis of web Log Using Clementine Data Minin Solution", Information Systems Review , Vol.4, No.1, pp.47-67.
- [6] Bamshad Mobasher, Namit Jain, Eui-Hong(Sam) Han, Jaideep Srivastava. (1996). Web Mining:

- Pattern Discovery from World Wide Web Transactions. Technical Report 96-050, Dept. of Computer Science, Univ. of Minnesota, Minneapolis.
- [7] Cooley, R., Tan P.N. and J.Srivastava, "Discovery of interesting usage pattern from web data ", Technical Report TR 99-022. University of Minnesota, 1999.
- [8] Jaideep Srivastavay, Robert Cooleyz, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data" ACM SIGKDD, Jan 2000. Volume 1, Issue 2-page 12
- [9] Mobasher, B., Cooley, R. and J. Srivastava, "Automatic Personalization Based on Web Usage Mining" Proc. of the ACM SIGKDD international conference on Konwledge discovery and data mining, pp.123-132, 1999.