*Journal of the Korean Data & Information Science Society*
*2009,* **20**(3), 493–504

한국데이터정보과학회지

# 웹 로그데이터를 이용한 대학입시 지원자 행태 분석†

최승배[1] · 강창완[2] · 조장식[3]

[12]동의대학교 데이터정보학과 · [3]경성대학교 정보통계학과

접수 2009년 2월 17일, 수정 2009년 5월 20일, 게재확정 2009년 5월 23일

**요 약**

홈페이지는 홈페이지를 운영하고 있는 모든 조직체들을 대변하는 얼굴이다. 웹 로그데이터는 홈페이지를 방문하는 사람들의 행적을 나타낸다. 웹 로그데이터를 분석함으로써 홈페이지 운용에 대한 유용한 정보를 얻을 수 있고, 이러한 정보를 이용하여 효율적인 홈페이지 관리 및 고객관계 관리를 수행할 수 있다. 본 연구에서는 D대학교의 홈페이지에서 얻어진 웹 로그데이터를 분석함으로써 효율적인 홈페이지 관리와 신입생 유치를 위한 홍보 전략을 세우는데 기초적인 정보를 제공한다.

주요용어: 데이터마이닝, 로그분석항목, 로그항목, 웹마이닝.

## 1. 서론

최근 컴퓨터의 발달로 어느 기업 또는 기관에서 얻어지는 데이터의 양은 헤아릴 수 없을 정도로 방대하다. 특히 웹 (web)상에서 얻어지는 데이터는 어느 타 분야에서 얻어지는 데이터보다 양적인 측면에서 비교할 수 없을 만큼 방대하다.

웹은 컴퓨터의 대중화, 하드웨어 기술의 진보, 소프트웨어 패키지의 사용법 상 (윈도우즈화)의 용이성, 웹을 통한 비즈니스의 가능성 등의 이유로 1990년대 후반기에 괄목상대한 발전이 있었다. 이와 같이 웹의 지속적인 성장과 함께 정보시스템이 보편화됨으로써 쓸모가 있든 없든 간에 웹상에서 얻어지는 데이터의 양은 실로 엄청나게 쏟아지고 있다. 이러한 대용량의 데이터 속에서 유용한 정보를 얻는다는 것은 '모래알 속에서 진주를 찾기 어렵다'라는 속담에 비유할 수 있을 정도로 어려운 경우가 많이 있다. 그러나 대용량 데이터로부터 정보획득 문제는 방대한 데이터에서 필요로 하는 정보를 찾는 기법인 데이터마이닝 (datamining)이 해결해 줄 수 있다.

웹마이닝 (webmining)은 웹상에서 얻어지는 데이터를 처리하는 데이터마이닝의 한 분야하고 할 수 있다. 즉, 웹마이닝은 웹상에서 얻어지는 방대한 데이터에서 유용한 정보를 얻어내는 기술로서 웹마이닝 도구를 사용하여 유용한 정보를 알아내고, 이러한 정보를 이용하여 다양한 분야에서 목적에 맞는 의사결정을 시도하는 것이라고 할 수 있다.

현재까지 웹 또는 웹마이닝과 관련하여 많은 연구가 활발히 진행되어 왔다. 가장 최근의 연구로서 강창완 등 (2009)은 한국자료분석학회 홈페이지에 대한 개괄적인 분석을 통하여 홈페이지의 개선점을 제시하고 있다. 강창완 등 (2001)은 구인/구직 웹 사이트에 대한 웹 로그데이터 분석을 수행하였다.최

† 이 논문은 2008학년도 동의대학교 교내연구비에 의해 연구되었음 (2008AA090).
[1] 교신저자: (614-714) 부산광역시 부산진구 가야동 산 24, 동의대학교 데이터정보학과, 부교수.
E-mail: csb4851@deu.ac.kr
[2] (614-714) 부산광역시 부산진구 가야동 산 24, 동의대학교 데이터정보학과, 교수.
[3] (608-736) 부산광역시 남구 대연 3동 314-79, 경성대학교 정보통계학과, 교수.

경호 (2008)는 통계학과 관련된 홈페이지들을 분석하여 각 홈페이지들을 비교하고 있으며, 최승배 등 (2002)은 와이즈로그 (WiseLog)와 웹트랜즈 (WebTrends) 두 제품을 고려하여 웹 로그분석기들을 비교하였다. 그리고 최승배 등 (2002)은 D학과의 학술정보 사이트를 개설하여 웹마이닝 기법을 이용하여 분석하였고, 최승배 등 (2003)은 웹 로그데이터의 정제기법을 개발한 바 있다. 또한 웹 로그데이터 분석의 유사한 연구로서 박한우 (2004a, 2004b) 등이 있다. 그리고 입시와 관련한 논문으로 김종태 등 (2009)은 2026년까지 대구광역시와 경상북도 지역의 고등학교 3학년 학생 수에 대한 예측과 대학 입학 정원수와 비교 분석을 하였고, 최승배 등 (2008)은 데이터마이닝을 이용하여 대학에 합격한 입시생들의 등록하지 않는 이탈을 최소화하는 방안을 제안하고 있다.

일반적으로 홈페이지를 분석하기 위해서는, 반드시 필요한 것은 아니지만, 분석 툴을 필요로 한다. 웹 로그데이터 (web log data) 분석을 목적으로 하는 분석 툴은 2000년 이전까지는 웹트렌즈와 와이즈로그가 대부분의 기업에서 많이 사용되었다. 그러나 이러한 툴들은 외국산이고 고가이기 때문에 2000년대에 들어서면서 웹 로그데이터 분석 및 개발 업체인 (주)넷스루와 (주)이너버스 등은 저가이면서 국산화의 필요성을 제기하였다. 이러한 배경 하에서 여러 국내 업체들이 한글판 분석기를 출시하여 상용화되기 시작하면서 기존의 분석기에 비하여 기능면에서 막강한 국산화 시대가 열렸다. 본 연구에서 웹 로그데이터 분석을 위해서 사용한 툴은 (주)넷스루에서 개발된 와이즈로그이다.

본 연구에서는 연구내용을 크게 두 관점에서 수행하고자 한다. 첫째, D대학교 전체 홈페이지를 분석하여 전반적인 이용실태에 대한 정보를 도출하는 것이다. 둘째, D대학교 홈페이지 분석을 통하여 신입생들이 입시와 관련하여 D대학교 홈페이지에서 어떤 행태를 보이고 있는 지에 대해서 알아보는 것이다. 본 연구의 분석결과에서 도출된 정보를 통하여 향후 어떤 방향으로 홈페이지를 운영할 지에 대한 효율적인 운영방안과 입시와 관련하여 신입생을 유치하기 위한 홍보 및 보완해야 할 사항 등에 대해서 제안을 하고자 한다.

본 연구에서는 2절에서 웹 로그데이터 분석에 대해서 간략히 소개하고, 3절에서는 웹 로그 분석에 대한 결과를 제시한다. 마지막으로 4절에서는 결론으로서 본 연구에 대한 분석결과를 제시하고, 분석결과에 근거하여 D대학교의 홈페이지를 운영하는데 있어서 좀 더 나은 운영을 위한 제안을 한다.

## 2. 웹 로그데이터 분석

### 2.1. 분석절차

웹 로그데이터 분석은 '로그분석항목' 설정, '로그항목' 설정, 선별 및 정제, 변환, 로그 분석, 웹 사이트에의 결과 반영과 같은 일련의 과정으로 진행된다. 그림 2.1은 웹마이닝의 수행과정을 흐름도로 보여주고 있다. 웹 로그데이터 분석을 위한 보다 자세한 내용은 최승배 등 (2008)을 참조해 주기 바란다.

#### 2.1.1. 로그분석항목 설정

'로그분석항목 (log analysis item)'은 여러 '로그분석항목' 중에서 분석에 필요한 '로그분석항목'만을 분석대상으로 하기 위해서 설정하는 단계이다. 웹 로그데이터의 '분석항목'으로서 '페이지 뷰 (page view) 수 추이', '많이 방문한 페이지', '오래 머문 페이지', '많이 시작한 페이지', '많이 이탈한 페이지', '히트수 추이', '많이 히트된 파일', '방문수 추이', '페이지 뷰 수 높은 방문자', '히트수 높은 방문자', '방문수 높은 방문자' 등 많은 항목들이 존재한다. 여기서 모든 분석항목들이 기업, 공공기관 등에서 유용한 정보가 될 수는 없다. 따라서 효율적으로 웹 사이트를 운영하기 위해서 웹 로그데이터 분석을 하기 전에 어떠한 분석항목들이 필요한가를 미리 정해 놓아야 한다.

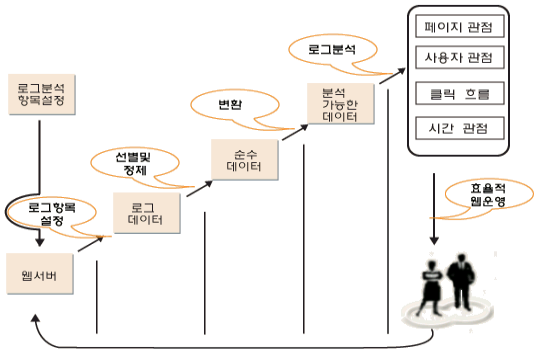



**그림 2.1** 웹마이닝의 분석 흐름도 (출처 : http://itec.incheon.ac.kr/)

**표 2.1** 로그 분석항목과 로그항목

| | 로그항목 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 로그분석항목 | date | time | c-ip | cs-user name | s-ip | s-port | cs-method | cs-uri -stem | cs-uri -query | sc-status | cs (User-Agent) |
| 페이지 뷰 추이 | | O | | | | | | O | | | |
| 많이 방문한 페이지 | O | O | | | | | | O | | | |
| 오래 머문 페이지 | O | O | O | | | | | O | | | |
| 많이 시작한 페이지 | O | O | O | | | | | O | | | |
| 많이 이탈한 페이지 | O | O | O | | | | | O | | | |
| 히트수 추이 | O | O | | | | | | O | | | |
| 많이 히트된 파일 | O | O | | | | | | O | | | |
| 방문수 추이 | O | O | O | | | | | | | | |
| 페이지뷰 수 높은 방문자 | O | O | | O | | | | O | | | |
| 히트수 높은 방문자 | O | O | O | O | | | | | | | |
| 방문수 높은 방문자 | O | O | O | | | | | | | | |
| 에러 경험 많은 방문자 | O | O | O | | | | | | | O | |
| 오래 머문 방문자 | O | O | O | | | | | O | | | |
| 브라우저 | O | O | | | | | | | | | O |
| 운영체제 | O | O | | | | | | | | | O |

### 2.1.2. 로그항목 설정

'로그항목 (log item)'은 웹 로그데이터에 있는 각 필드를 의미한다. '로그항목' 설정 단계에서는 분석에 필요한 '로그항목'들을 설정하는 단계이다. '로그항목' 설정단계에서는 분석할 '로그항목'들을 설정한 후, 분석에 필요한 '로그항목'들만 웹 로그데이터에 남겨지도록 하는 단계이다. 아무런 설정 없이 웹 로그데이터를 생성시키게 되면 로그 데이터의 크기는 매우 커지게 되기 때문에 생성된 데이터의 크기만으로도 분석을 할 수 없는 상황을 초래할 수 있다.

분석 시 필요한 각 '로그분석항목'들에 대해서 요구되는 '로그항목'들을 정리한 결과가 표 2.1에 주어져 있다. 예를 들면, '로그분석항목' 중에서 '페이지 뷰 추이'를 분석하기 위해서는 '시간 (time)'과 '웹상에서 클릭한 페이지 (cs-uri-stem)'를 필요로 하게 된다. 웹 로그데이터분석을 위해서 세부적인 분석을 수행할 때 반드시 표 2.1에 주어진 것과 같이 설정해야 한다는 것은 아니다. 이것은 하나의 기준일 뿐이고, 웹 로그데이터분석을 수행할 때 분석자의 분석의도에 맞게 '로그항목'을 적절히 설정하면 된다.

### 2.1.3. 선별, 정제 및 변환

2.1.1절과 2.1.2절에서 각 항목들의 설정에 의해서 얻어진 웹 로그데이터는 분석 가능한 형태로 데이터를 변환해야 한다. 모든 분석에 있어서 가장 중요한 것은 양질의 데이터를 확보해야 한다는 것이다. 이러한 정제 작업은 데이터의 질을 높여 주는 역할을 한다 (최승배 등, 2003; 최승배 등, 2008). 웹 로그데이터 분석에서도 웹 로그데이터의 정제를 통하여 분석자가 원하는 결과를 얻을 수 있도록 하여야 한다. 이 과정은 웹 로그데이터 분석에 가장 시간이 많이 소요되는 과정이기는 하지만 의미 있는 분석을 수행하기 위해서는 생략되어서는 안 되는 중요한 과정이다.

### 2.1.4. 로그분석

로그분석은 어떤 분석도구를 이용하여 웹 로그데이터 분석을 실제로 수행하는 과정이다. 웹 로그데이터분석을 위한 일련의 과정을 거쳐 분석한 결과는 분석대상인 홈페이지를 보완 수정하여 보다 효율적인 홈페이지 운영을 위한 기초자료로 활용된다. 상기한 2.1.1절, 2.1.2절, 2.1.3절에서 얻어진 내용들이 만족스럽지 못할 경우는 앞 단계로 다시 되돌아가서 만족스러울 때까지 과정들을 순환시킬 수 있다.

## 2.2. 분석데이터

2008년 12월 7일부터 2009년 2월7일까지 3개월간에 D대학교의 홈페이지를 방문한 데이터 (웹 로그데이터)로서 전체 용량은 305MB이다. 이에 대한 데이터의 일부가 그림 2.2에 주어져 있다.

일반적으로 웹 로그데이터는 컴퓨터에서 '시작' ⇒ '설정' ⇒ '제어판' ⇒ '인터넷 서비스 관리자' ⇒ '기본 웹 사이트' ⇒ '속성'부분에서 (1) 어떤 형식으로 웹 로그데이터를 얻을 것이며, (2) '생성주기'를 어떻게 ('시간별', '일별', '주별', '월별') 할 것이며, (3) 어디에 '저장'할 것인가를 지정하고, (4) 웹 로그데이터 형식으로 생성시킬 '로그항목'들을 취사선택하여 그림 2.2와 같은 웹 로그데이터를 얻을 수 있다.

## 2.3. 분석내용

본 연구에서는 D대학교의 홈페이지를 다음과 같은 내용으로 분석하고자 한다. (1) 일자별로 홈페이지에 대한 방문의 추이를 종합적으로 알아본다. (2) 어떤 방문자들이 주로 방문했는지를 알아보기 위하여 방문자 관점에서 분석결과를 제공한다. 그리고 (3) 방문자들이 어떤 페이지를 많이 접속하였는지에 대한 정보를 알아보기 위하여 페이지 관점에서 분석을 수행한다.

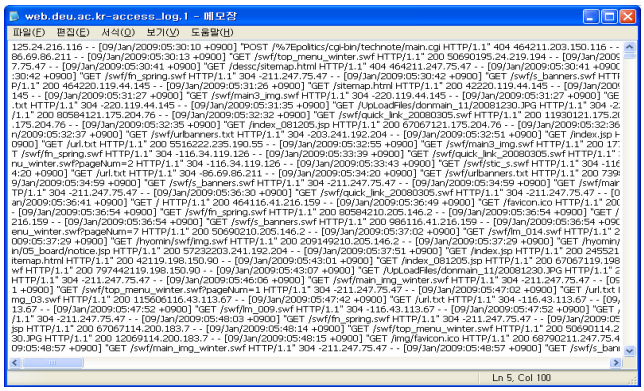



그림 2.2 분석 웹 로그데이터

## 3. 분석결과

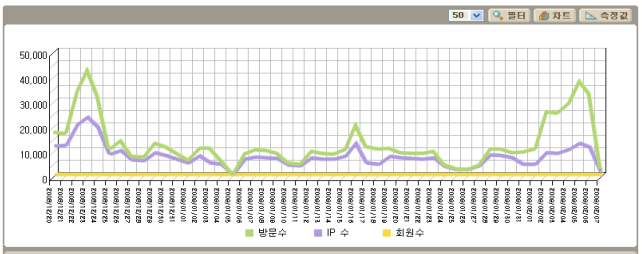
### 3.1. 종합적인 추이

종합적인 홈페이지의 추이를 일자별로 알아보기 위한 분석결과가 그림 2.3에 주어져 있다. 그림 3.1에서 위의 곡선은 방문 (접속) 수 (웹사이트의 방문자는 일련의 페이지를 보게 되는데, 이러한 과정은 하나의 방문)이고, 아래곡선은 IP 수 (해당 일에 접속한 IP에서 중복을 제거한 수)를 의미한다.

그림 3.1의 결과를 보면 12월 21일부터 12월 25일 사이와 2월 2일부터 2월 7일 사이의 기간에서 페이뷰 (page views : 웹사이트내의 페이지 파일에 대한 접근의 총합으로서 웹 사이트의 이용도을 보여주는 수치) 수가 가장 많이 나타나고 있음을 볼 수 있다. 이 시기들이 D대학교 홈페이지를 가장 많이 이용한 시기라고 할 수 있다. 12월 21일부터 12월 25일 사이의 시기는 정시의 원서접수기간이고, 2월 2일부터 2월 7일 사이의 시기는 합격자 등록기간이다. 이러한 결과로 볼 때, D대학교의 홈페이지에 가장 많이 방문한 시기가 12월 21일부터 12월 25일 사이와 2월 2일부터 2월 7일 사이의 기간에 방문한 방문객은 2009학년도 신입생임을 예상할 수 있다. 다른 기간들의 D대학교 홈페이지에 대한 종합적인 방문추이에 대해서는 그림 3.1을 참조하면 된다.

### 3.2. 방문자 관점

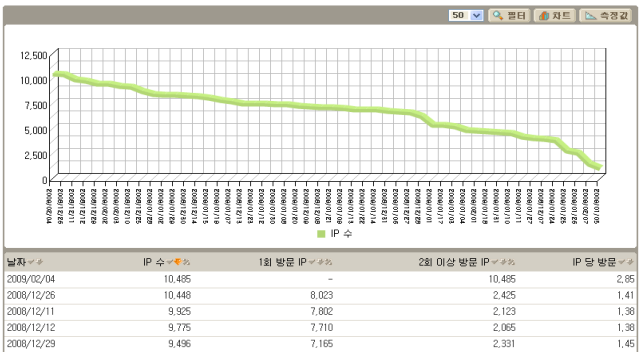
D대학교 홈페이지에 방문하는 방문자들에 대한 행태를 방문자 관점에서 분석한 결과를 제시한다. 이를 위하여 'IP 방문자 추이', '지역별 방문순위'에 대한 결과를 중점적으로 다룬다. 'IP 방문자 추이'는 해당 일에 접속한 IP에서 중복을 제거한 수인 IP 수로 홈페이지에 많이 방문한 결과를 보여준다. 이에 대한 결과는 그림 3.2에 주어져 있는데 이 결과는 IP 수로 내림 정렬한 결과이다. 그림 3.2를 보면 IP 방문자수로 많이 방문한 일자는 2009년 2월 4일에 제일 많이 방문했고 (10,485회), 이 날에는 모든 IP 방문자가 2회 이상 방문한 것으로 나타났다. 그 다음이 2008년 12월 26일에 10,448회로 1회 방문이 8,023건, 2회 이상 방문한 IP 방문자가 2,425회 방문한 것으로 나타났다. 2009년 2월 4일과 2008년 12월 26일 두 날을 보면, 입시와 관련된 날로서 2월 4일은 입학생들이 등록금 납부 마지막 날이고, 2008년 12월 26일은 추가합격자 발표 날과 관련이 있는 날로 나타났다.




| 날자 | 방문수 | IP 수 | 회원수 | 페이지 뷰 | 방문당 페이지뷰 | 방문당 머문 시간 min | 회원 방문률 |
|---|---|---|---|---|---|---|---|
| 2008/12/20 | 18,104 | 12,140 | - | 135,023 | 7.46 | 3.02 | - |
| 2008/12/21 | 17,654 | 12,360 | - | 121,727 | 6.90 | 3.06 | - |
| 2008/12/22 | 34,547 | 20,703 | - | 257,281 | 7.45 | 4.24 | - |
| 2008/12/23 | 44,101 | 24,167 | - | 356,928 | 8.09 | 4.79 | - |
| 2008/12/24 | 32,570 | 19,709 | - | 271,130 | 8.32 | 4.53 | - |

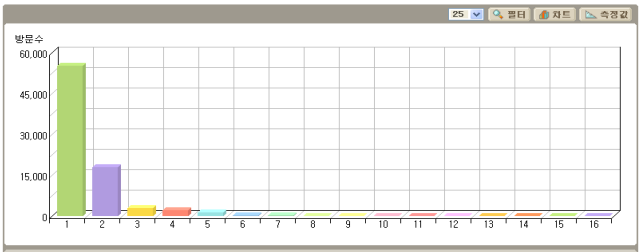
**그림 3.1** 종합적인 추이

다음으로 어느 지역에서 많이 방문했는지에 대해서 알아보기 위해서 지역별 방문자 순위에 대해서 분석하였고, 그 결과는 그림 3.3에 주어져 있다.




| 날자 | IP 수 | 1회 방문 IP | 2회 이상 방문 IP | IP 당 방문 |
|---|---|---|---|---|
| 2009/02/04 | 10,485 | - | 10,485 | 2.85 |
| 2008/12/26 | 10,448 | 8,023 | 2,425 | 1.41 |
| 2008/12/11 | 9,925 | 7,802 | 2,123 | 1.38 |
| 2008/12/12 | 9,775 | 7,710 | 2,065 | 1.38 |
| 2008/12/29 | 9,496 | 7,165 | 2,331 | 1.45 |

**그림 3.2** IP 방문자수 추이




| | 지역 | 방문수 | 페이지 뷰 | 머문 시간 min | 전송량 KB |
|---|---|---|---|---|---|
| 1 | 부산 | 55,323 | 697,607 | 535,211 | 113,223,399 |
| 2 | 서울 | 17,407 | 123,375 | 65,640 | 40,616,631 |
| 3 | 경기도 | 2,300 | 22,031 | 12,468 | 8,528,105 |
| 4 | 경상남도 | 1,597 | 16,229 | 7,404 | 4,313,068 |
| 5 | 강원도 | 544 | 18,206 | 655 | 1,416,891 |

**그림 3.3** 지역별 방문자

그림 3.3의 결과를 보면, '부산'이 방문 수 55,323회로 가장 많이 방문한 것으로 나타났고, 그 다음이 서울 (17,407회), 경기도 (2,300회), 경상남도 (1,597회), 강원도 (544), 울산 (407회), 경상북도 (383회) 순으로 나타났다. 분석결과를 입시와 관련하여 해석하면 D대학교 홈페이지 방문자는 부산이 제일 많은 결과로 나타난 것은 당연한 결과라고 할 수 있다. 그러나 의외로 부산과 인접된 지역인 경상남도, 울산, 경북이 서울과 경기도에 비해 방문수가 적은 것으로 나타났다. 이 결과는 입시기간과 상관없이 일반적으로 서울과 경기도 수도권에서 많이 방문하고 있다고 해야 할 것이다.

## 3.3. 페이지 관점

### 3.3.1. 방문페이지

본 절에서는 방문자가 어떤 페이지를 많이 보고 있는지에 대한 정보를 제공하는 페이지 관점에서 분석한 결과를 제시한다. 이에 대한 결과는 그림 3.4에 주어져 있다. 그림 3.4를 보면 'http://www.deu.ac.kr'이 1,395,295회로 제일 많이 방문한 페이지로 나타났다. 그러나 이것은 어느 기관의 방문자가 홈페이지에 방문할 경우, 통상적으로 메인페이지를 통하여 세부적인 페이지를 보기 때문에 이 페이지에 대한 결과는 의미가 없다.

그 다음 많이 방문한 페이지를 순서대로 몇 개만 고려하면 다음과 같다.

1. 2순위 : 'http://www.deu.ac.kr//sitemap.html' (1,183,065회),
2. 3순위 : 'http://www.deu.ac.kr//top_menu_winter.swf' (1,174,823회),
3. 4순위 : 'http://www.deu.ac.kr//index_081205' (669,211회),
4. 5순위 : 'http://www.deu.ac.kr//hyomin' (403,530회)로 나타났다.

2순위인 'http://www.deu.ac.kr//sitemap.html'은 전체사이트를 맵 형태로 나타내어 주는 페이지로 이 사이트 역시 메인 페이지로 고려할 수 있어 순위에 의미가 없다고 할 수 있다. 3순위로 나타난 'http://www.deu.ac.kr//index_081205'은 메인 페이지 내 교수초빙 1차 합격자 발표 통보와 관련된 페이지이다. 'http://www.deu.ac.kr//top_menu_winter.swf' 페이지는 D대학교 소개 및 입학정보 등이 있는 메뉴 페이지이고, 'http://www.deu.ac.kr//hyomin'페이지는 기숙사에 대한 정보를 알려주는 페이지이다. 분석시기가 입시기간이고, 3순위와 4순위를 고려한다면 입학과 관련된 페이지에 신입생들이 많이 접속했음을 알 수 있다.

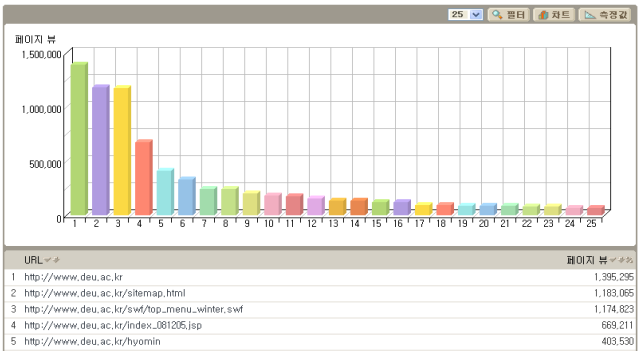



**그림 3.4** 방문한 페이지

### 3.3.2. 다운로드 페이지

본 절에서는 페이지를 다운로드한 경우, 많이 다운로드한 페이지별로 분석한 결과이다. 이에 대한 결과는 그림 3.5에 주어져 있다. 분석결과 제일 많이 다운로드한 페이지는 '발전기금약정서 (히트: 58,759)', '교류이수지원서 (히트: 49,931)', '이수학점인정지원서 (히트: 39,677)', '이수계획서 (히트: 19,583)', '학적부 기재사항 정정원 (히트: 12,452)' 순으로 나타났다. 여기서 '히트 (hits)'란 웹 사이트 내의 모든 파일에 대한 접근한 총합을 나타낸다. 위의 결과들을 보면 다운로드 페이지는 주로 학부생들이 주로 이용한 페이지이고, 이 페이지는 학부생들이 학부생활 시 유용하게 사용되는 페이지임을 알 수 있다.

### 3.3.3. 학과별 방문순위

본 연구의 연구기간이 정시기간이기 때문에 대부분의 방문자가 신입생이라는 가정 하에서 학과별 방문 순위를 알아보았다. 학과 페이지별로 많이 방문한 학과에 대해서 정리한 결과가 표 3.1에 주어져 있다.

표 3.1에 주어진 결과는 지면 상 상위를 차지하고 있는 단대별 3개 학과 (3개 학과로 구성되어 있지 않은 단대가 있음) 이하들만을 고려한 결과이며, 지원율은 모집단위 (가군, 나군, 다군)들이 구분되어 있

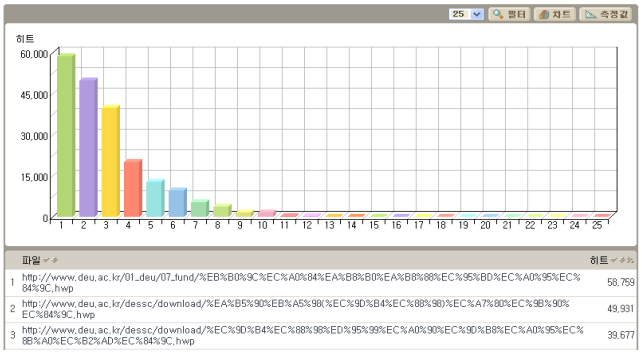



**그림 3.5** 다운로드 페이지

**표 3.1** 학과별 방문순위 (단대별 상위 3개 학과)

| 단대 | 학과 | 페이지 뷰 수 | 09년도 지원율(정시) | 지원자 수 |
|---|---|---|---|---|
| 인문대학 | 일어일문학과 | 928 | 3.89 | 256.74 |
| | 평생교육학과 | 897 | 3.16 | 126.40 |
| | 영어영문학과 | 871 | 4.00 | 240.00 |
| 법정대학 | 사회복지학과 | 1,425 | 3.19 | 191.40 |
| | 법·경찰행정학부 | 1,320 | 3.12 | 405.60 |
| | 행정학과 | 727 | 3.23 | 161.50 |
| 상경대학 | 호텔컨벤션경영학과 | 1,450 | 4.63 | 208.35 |
| | 외식산업경영학과 | 1,442 | 3.63 | 181.50 |
| | 경영학과 | 1,364 | 4.27 | 555.10 |
| 자연과학대학 | 간호학과 | 1,276 | 5.78 | 289.00 |
| | 임상병리학과 | 849 | 3.70 | 185.00 |
| | 치위생학과 | 822 | 4.67 | 140.10 |
| 한의과대학 | 한의학과 | 1,035 | 7.27 | 363.50 |
| 생활과학대학 | 식품영양학과 | 767 | 3.49 | 191.95 |
| | 패션디자인학과 | 704 | 4.12 | 226.60 |
| | 보육가정상담학과 | 657 | 2.37 | 118.50 |
| 공과대학 | 기계공학과 | 988 | 4.30 | 344.00 |
| | 컴퓨터공학과 | 814 | 4.13 | 330.40 |
| | 조선해양공학과 | 786 | 3.45 | 207.00 |
| 영상정보대학 | 영상정보공학과 | 743 | 4.30 | 215.00 |
| | 멀티미디어공학과 | 666 | 3.80 | 228.00 |
| | 디지털문화콘텐츠공학과 | 665 | 2.92 | 116.80 |
| 예술디자인대학 | 산업디자인학과 | 1,557 | 3.03 | 212.10 |
| | 음악학과 | 1,384 | 5.00 | 350.00 |
| | 미술학과 | 985 | 1.48 | 44.40 |
| 체육과학대학 | 체육학과 | 799 | 5.32 | 319.20 |
| | 특수체육학과 | 563 | 1.78 | 71.20 |
| | 레저스포츠학과 | 552 | 5.24 | 314.40 |

는 경우 전체 모집인원 수에서 지원자 수로 나눈 결과이다. 그리고 학과의 홈 페이지 방문 (페이지 뷰 수)과 2009년도 정시의 지원율과 관련성이 있는지에 대해서 상관분석을 통하여 알아보았다. 그러나 정시의 지원율은 각 학과마다 모집인원이 상이하기 때문에 지원율 대신으로 각 학과의 지원율에 모집정원을 곱한 총지원자 수를 이용하였다. 전체 학과들의 측면에서의 상관분석 결과 p값이 0.008로 유의수준 0.05하에서 매우 유의하게 나타났다.

### 3.4. 시스템 운영측면

#### 3.4.1. 에러

D대학교의 홈페이지를 3개월간 운영함에 있어서 에러적인 측면에서 분석한 결과가 그림 3.6에 주어져 있다. 그림 3.6을 보면 에러코드 404가 240,932건으로 가장 많이 발생했음을 볼 수 있고, 나머지는 에러코드 404에 비하여 아주 미미하다. 에러코드 404는 요청된 URL에 대한 페이지가 존재하지 않는 경우에 발생된다. 이러한 에러는 페이지 요청 시 방문자의 PC의 문제로 발생될 수 있고, 실제로 홈페이지에서 방문자가 원할 때 서버의 문제로 발생될 수 있는 문제이다.

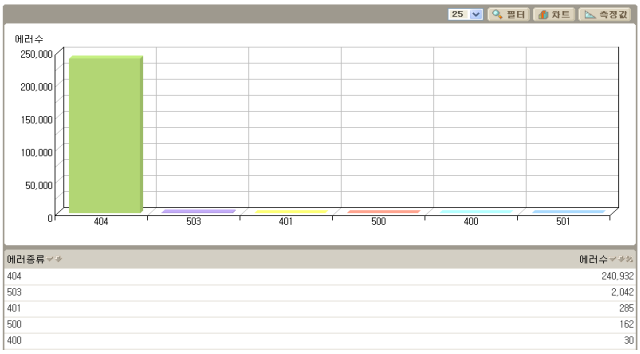



**그림 3.6** 에러의 종류

#### 3.4.2. 전송량

전송량에 관한 분석결과는 그림 3.7에 주어져 있다. 그림 3.7에서 '최대전송량'은 웹 사이트에서 전송되어 나가는 데이터의 총합을 시간 간격으로 측정하여 단위 시간당 계산한 값이다. 이것은 웹 사이트의 순간적인 이용도를 보여줌과 동시에 네트워크의 순간적인 이용도를 보여주기도 한다. 그림 3.7의 결과를 보면 2008년 12월 23일에 4,992,048KB로 가장 많이 방문자에게 전송되었고, 2008년 12월 24일에 3,233,275KB로 두 번째로 전송량이 많은 것으로 나타났다. 이러한 결과는 12월 19일부터 12월24일까지 원서접수 마감일임을 감안할 때, 원서접수를 위한 모집요강 등을 D대학교의 홈페이지에서 전송한 것으로 추측된다.

## 4. 결 론

본 연구에서는 2009학년도 정시기간에 웹 로그데이터를 수집하여 2009학년도 신입생들의 D대학교 홈페이지에서 움직임들을 파악하여 차후년도의 입시전략을 기획하는데 기초자료로 활용하는 것에 목표를 두었다. 그러나 사정에 의해서 수시 2학기의 데이터에 대해서는 분석하지 못하고 정시기간에 대해서만 분석하였다.

물론 D대학교의 홈페이지를 분석하는 것을 신입생들로 한정하는 것은 무리가 있다. 그러나 분석기간이 입시기간임을 감안하여 홈페이지에 방문하는 방문자들의 대다수가 신입생임을 가정하고 분석한 결과이다. 따라서 종합적으로 홈페이지를 분석하되 신입생들의 행태에 중점을 두고 분석결과를 해석한다. 분석결과는 다음과 같다.

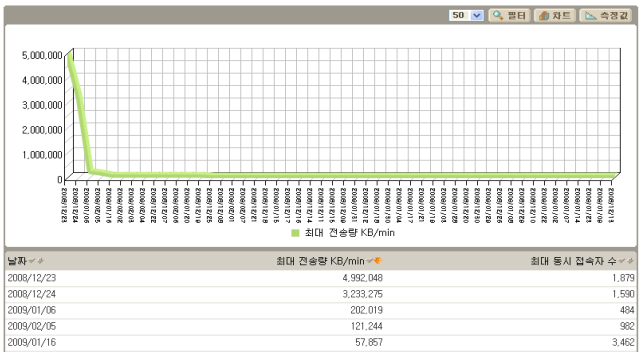



**그림 3.7** 전송량

첫째, 종합적인 추이로 보면 D대학교의 홈페이지에 가장 많이 방문한 시기는 12월 21일부터 12월 25일 사이와 2월 2일부터 2월 7일 사이로 나타났다. 12월 21일부터 12월 25일 사이의 시기는 정시의 원서접수기간이고, 2월 2일부터 2월 7일 사이의 시기는 합격자 등록기간이기 때문에 이 시기에 가장 많이 방문한 것으로 나타났다고 할 수 있다. 따라서 D대학교 당국은 신입생과 관련된 페이지를 보완 또는 획기적인 페이지를 신설함으로써 신입생의 유치를 위해서 노력해야 할 것으로 판단된다.

둘째, 방문자 관점에서 'IP 방문자 추이', '지역별 방문순위' 두 가지로 홈페이지를 분석한 결과 'IP 방문자 수'로 많이 방문한 일자는 2009년 2월 4일에 제일 많이 방문한 것으로 나타났고 (10,485회), 2008년 12월 26일에 10,448회로 그 다음으로 많이 방문한 것으로 나타났다. 2009년 2월 4일과 2008년 12월 26일 두 날을 보면, 입시와 관련된 날로서 2월 4일은 입학생들이 등록금 납부 마지막 날이고, 2008년 12월 26일은 추가합격자 발표 날과 관련이 있는 날이다. '지역별 방문순위'의 분석결과는 '부산'이 방문 수 55,323회로 가장 많이 방문한 것으로 나타났고, 그 다음이 서울 (17,407회), 경기도 (2,300회), 경상남도 (1,597회), 강원도 (544회), 울산 (407회), 경상북도 (383회) 순으로 나타났다. 이러한 지역별 분석결과를 입시와 관련하여 해석한다면 부산과 부산인접지역뿐만 아니라 수도권으로의 새로운 입시전략을 세울 필요가 있다고 할 수 있다.

셋째, 페이지 관점에서 '방문페이지'와 '다운로드 페이지' 두 가지 점에서 분석하였다. '방문페이지'에 대한 분석결과 D대학교 소개 및 입학정보 등이 있는 메뉴 페이지인 'http://www.deu.ac.kr//top_menu_winter.swf'과 기숙사에 대한 정보를 알려주는 페이지인 'http://www.deu.ac.kr//hyomin'의 페이지에 많이 방문한 것으로 나타났다. 페이지 관점에서도 볼 수 있듯이 입학과 관련되어 있고, 특히 부산 이외의 신입생들에게 관심이 많은 기숙사 페이지에 신입생들이 많이 접속했음을 알 수 있다. 따라서 D대학교 당국에서는 입학생 유치를 위한 기숙사 관련 페이지와 신입생들과 직접적인 관련성이 있는 보다 다양하고 새로운 페이지를 개발할 필요가 있다고 생각된다. 그리고 학과의 홈 페이지 방문 (페이지 뷰 수)과 2009년도 정시의 지원율과 관련성이 있는지에 대해서 알아보았다. 홈페이지 방문과 학과 지원율의 상관분석 결과 유의한 결과 (p값: 0.008)를 보여 학과 홈페이지를 많이 방문할수록 지원율이 높음을 알 수 있다. 따라서 홈페이지 관리는 신입생 유치 전략의 일환이 될 수 있기 때문에 대학 당국과 각 학과에서는 홈페이지의 관리에 소홀함이 없어야 할 것으로 판단된다.

넷째, 시스템 운영측면에서 에러와 전송량 두 가지에 대해서 분석하였다. 에러와 관련한 분석결과 에러코드 404가 240,932건으로 가장 많이 발생했음을 알 수 있었다. 에러코드 404는 요청된 URL에 대한 페이지가 존재하지 않는 경우에 발생되는 것으로 향후 홈페이지를 관리하는데 있어서 입시와 관련하여

에러가 발생하면 신입생 유치에 문제가 발생할 수 있음을 인지하고 에러에 특별히 신경을 써야 할 것이다. 전송량에 관한 분석결과는 2008년 12월 23일과 24일에 가장 많이 전송된 것으로 나타났다. 이것은 12월 19일부터 12월24일까지 원서접수 마감일임을 감안할 때, 원서접수를 위한 모집요강 등을 D대학교의 홈페이지에서 전송한 것으로 추측된다. 이러한 결과를 보면 신입생들이 원서접수 마지막 날에 원서접수를 많이 하기 위해서 홈페이지를 방문한다고 볼 수 있기 때문에 신입생들의 견지에서 보다 빠르고 효율적으로 관련서류를 전송해 갈 수 있도록 배려를 해야 할 것으로 생각한다.

본 연구의 중요한 목적은 신입생의 행태를 분석하는 것이기 때문에 입학생 유치를 위해서는 D대학교의 홈페이지를 효율적으로 운영해야 함은 물론 각 학과의 홈페이지도 잘 관리함으로써 입학생 유치를 위한 새로운 전략을 세워야 할 것으로 판단된다. 본 연구에서 분석한 시기는 정시시기에 얻어진 웹 로그데이터만을 이용하여 분석하였는데 향후 수시기간도 함께 고려하여 분석하면 좀 더 나은 입시정보를 얻을 수 있지 않을까 하는 아쉬움이 남는다.

## 참고문헌


강창완, 김규곤, 김태완 (2009). 한국자료분석학회 웹사이트에 대한 웹로그 분석. <한국자료분석학회지>, **11**, 361-370.

강창완, 김규곤, 최승배, 정민석, 박광준, 손승한, 임승범, 정경미 (2001). 구인/구직 웹 사이트에 대한 웹 로그분석. <한국자료분석학회지>, **3**, 331-342.

김종태, 서효민, 이인락 (2009). 2026년까지 대구광역시와 경상북도 지역의 고등학교 3학년 학생수에 대한 예측과 대학 입학정원수와의 비교 분석. <한국데이터정보과학회지>, **20**, 159-169.

박한우(2004a). 정치 웹사이트 하이퍼링크 네트워크에 대한 종단연구. <한국자료분석학회지>, **6**, 901-915.

박한우(2004b). 정치 웹사이트 아웃링크 데이터를 이용한 네트워크 분석-16대 국회의원 웹사이트를 중심으로-. <한국자료분석학회지>, **6**, 1109-1121.

최경호 (2008). 통계학 관련 학회 웹사이트 현황 및 네트워크 분석. <한국자료분석학회지>, **10**, 1603-1612.

최승배, 강창완, 조장식 (2008). D대학교의 입학시험에서 합격한 신입생들에 대한 이탈에 대한 특성분석. <한국데이터정보과학회지>, **19**, 1111-1122.

최승배, 김규곤, 강만기, 문승호, 강창완, 조장식, 이정형 (2008). <Wiselog를 이용한 웹 로그데이터 분석>, 자유아카데미, 서울.

최승배, 김규곤, 강창완, 조성기, 손종관 (2002). 웹 로그 분석기들의 비교연구. <한국자료분석학회지>, **4**, 327-340.

최승배, 임승범 (2002). 웹 마이닝 기법을 이용한 학술정보 사이트 분석. <한국자료분석학회지>, **4**, 451-464.

최승배, 손종관, 김규곤, 조성기 (2003). 웹 로그데이터의 정제기법 개발 연구. <한국자료분석학회지>, **5**, 345-357.

*Journal of the Korean Data & Information Science Society*
*2009,* **20**(3), 493–504

한국데이터정보과학회지

# Behavior analysis of entrance applicants using web log data†

Seungbae Choi[1] · Changwan Kang[2] · Jangsik Cho[3]

[12]Department of Data Information Science, Dongeui University
[3]Department of Informational Statistics, Kyungsung University

Received 17 February 2009, revised 20 May 2009, accepted 23 May 2009

**Abstract**

The web log data analysis is to analysis traces which visitors remain while they drop by a web-site. Ultimately it can help to obtain a lot of useful information that can efficiently manage homepage and perform CRM(customer relationship management) using obtained information. In this paper, we provide a basic information to manage efficiently homepage of D university and to establish strategy for invitation of new pupil, as analyzing web log data for D university.

*Keywords*: Datamining, log analysis item, log item, webmining.

† This work was supported by Dong-eui University Foundation Grant (2008AA090).
[1] Corresponding Author: Associate Professor, Department of Data Information Science, Dongeui University, Busan 614-714, Korea. E-mail: csb4851@deu.ac.kr
[2] Professor, Department of Data Information Science, Dongeui University, Busan 614-714, Korea.
[3] Professor, Department of Informational Statistics, Kyungsung University, Busan 608-736, Korea.