
Ajax 를 이용한 사용자의 웹 페이지 이용 행태 분석

Analysis User Action in Web Pages using Ajax technique

이동훈, Dong-Hoon* Lee 윤태복, Tae-Bok Yoon**,
김건수, Kun-Su Kim**, 이지형, Jee-Hyong Lee ***

요약 ~ 웹 페이지의 증가에 따라 문서의 질을 파악하는 것은 매우 중요한 문제가 되었다. 방문횟수와 같은 빈도 측정에 기반한 평가는 문서에 대한 사용자의 반응을 반영하기에 부족하다. 사용자는 문서가 사용자의 의도와 관련이 부족하거나 필요가 없는 경우 이용 시간이 매우 짧아진다는 특성을 보인다. 또한 사용자는 웹 페이지 이용 시 다양한 행위를 통해 페이지를 이용하게 된다. 마우스 포인터의 이동이나 클릭, 페이지 스크롤 등 웹 브라우저에서만 이루어지는 다양한 행위가 존재한다. JavaScript 는 브라우저에서 발생하는 이벤트 정보를 추적할 수 있고, 이렇게 수집된 정보를 사용자가 이용하는 도중에 서버로 전송이 가능하다. 본 논문에서는 사용자의 이용 정보를 수집하여 웹 페이지 이용을 분석하여 페이지 평가에 대한 기준을 제시한다.

Abstract ~ Web page evaluation is important issue in the Internet, Web pages are increasing extremely fast. The web page evaluation based on frequency, like the count of the page view (PV), is not sufficient way even it is used variously. Because users never use the unnecessary or irrelevant web pages for a long time. We concentrated on user's visiting duration time for the evaluation web pages. And we can collect user actions. Users do some action when users using the web page in the web browser. The movements of mouse pointer, mouse button click, page scrolling and so on are produced in the web browser. JavaScript can collect user action and Ajax can send collected data to server when user using the web browser without no user notification.

핵심어: *Web page evaluation, user action, Ajax*

본 연구는 21 세기 프론티어 연구개발 사업의 일환으로 추진되고 있는 유비쿼터스컴퓨팅 및 네트워크원천기반기술개발사업의 지원을 받았습니다.

*주저자 : 성균관대학교 전자전기컴퓨터공학부 석사과정 e-mail: idoun@skku.edu

**공동저자 : 성균관대학교 전자전기컴퓨터공학부 박사과정 e-mail: tbyoon@skku.edu

**공동저자 : 성균관대학교 전자전기컴퓨터공학부 석사과정 e-mail: kkundi@skku.edu

1. 서론

현대 인터넷의 웹 페이지 증가속도는 대단히 빠르다. 따라서 인터넷의 수많은 문서들의 질을 파악하여 원하는 것을 찾아내는 것은 매우 중요한 이슈가 되고 있다[1]. 검색 엔진들은 더 좋은 결과를 사용자들에게 제공하기 위하여 경쟁하고 있고, 일반적인 웹 게시물이나 블로그들도 콘텐츠를 평가하여 추천해 주는 기능을 제공하고 있다. 방문 횟수와 같은 '빈도 측정'에 기반한 웹 페이지 평가는 대중적인 평가 기준의 하나로서 꾸준히 이용되고 있다. 그러나 이 방법 만으로는 문서에 대한 사용자의 반응을 반영하기에 부족하다. 게시판에서 주로 사용되는 이 방법은, 사용자가 웹 페이지 조회 시 참고로 하는 정보가 제목에 노출된 내용 혹은 본문의 일부이다. 따라서 웹 페이지의 질을 반영하였다고 판단하기에 부족하다. 특히, 사용자의 의도 또는 목적과 관련이 부족하거나 정보가 거리가 먼 경우 혹은 잘못된 페이지일 경우 이용 시간이 매우 짧아진다는 특성을 보이므로, 모든 방문 횟수가 동일한 가치를 지닌다고 말할 수 없다. 사용자의 웹 페이지 이용 시간과 페이지의 정보의 질에는 상관관계가 있다고 판단하게 되었으며 이를 콘텐츠의 품질 평가방법에 적용하는 방법을 연구하였다.

이 때, 사용자의 이용 시간만을 이용할 경우 몇 가지 문제점이 발생하게 된다. 사용자가 단순히 브라우저에 페이지를 열어 놓고 이용하지 않았을 경우 이용이 없었음에도 페이지가 높게 평가 될 수 있다. 이용 시간을 보충해 주는 추가적인 기준이 필요하다. 우리는 사용자가 웹 페이지 조회 시 다양한 행위를 발생시키는 것을 파악하였다. 사용자는 웹 페이지 이용 시 콘텐츠 조회를 위해 마우스를 움직이거나 버튼을 클릭하는 행위를 한다. 이를 수집하여 적절한 기준을 찾아 이용 시간을 평가에 반영하는 방법을 연구하였다. 사용자 행위를 수집하기 위해 기준에는 실험용 브라우저를 만들거나 운영체제의 이벤트 정보를 가로채는 에이전트를 만들어 모니터링 하였다. 그러나 본 논문에서는 일반적인 환경에서도 정보 수집을 가능하게 하기 위한 방법을 고려하였다. 브라우저의 JavaScript 는 브라우저가 발생시키는 다양한 이벤트 정보를 수집하고 축적할 수 있다. 최근 웹 2.0 의 시류와 함께 각광받고 있는 Ajax 기술은, 수집된 이벤트 정보를 서버로 전송할 수 있게 한다. 이 기술의 특징은 사용자가 브라우저 내부에서 실행되는 것을 눈치채지 않고서도 서버와의 통신이 가능하게 한다. 일반적으로 서버와의 통신은 링크의 클릭을 통한 페이지 전환에 있으나 Ajax 기술을 통해 서버의 다양한 데이터를 전송 받을 수 있고 서버에 전송 가능하게 된 것이다. 본 논문의 목적은

실험을 통해 사용자의 이용 정보를 수집하고 이를 분석하여 페이지 평가에 대한 기준을 제시하는데 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 사용자의 이용 정보를 수집하기 위해 구현한 시스템에 대하여 설명한다. 3 장에서는 진행된 실험의 구성과 설계에 대하여 기술하였고 4 장에서는 실험의 결과를 분석하였다. 끝으로 5 장에서는 본 논문의 결론을 맺고 향후 연구 방향에 대하여 논의하였다.

2. 로그 수집 시스템

본 장에서는 사용자 정보 수집에 사용된 시스템에 대하여 소개한다. Ajax 기술을 이용한 로그 수집을 위하여 특별한 시스템을 설계하였다.

2.1 시스템 서버 구성

실험을 위한 서버의 구성은 다음과 같다.

- CPU: Intel Pentium 4 3.0 GHz
- HDD: 160 Mb
- OS: OpenSUSE Linux 10.2
- Web Server: Apache HTTP Server 2.0.26
- Web Application Server: Apache Tomcat 6.0.10
- DBMS: MySQL 5
- Java VM: Sun microsystems Java 6

실험에서는 효율적인 DBMS 이용과 페이지 구현을 위하여 Apache Tomcat 을 Web Application Server 로 이용하였다. Tomcat 을 이용함에 따라 필터와 같은 기능을 쉽게 이용할 수 있어 기존에 제작된 웹 페이지들에도 사용자 이용 정보 수집용 코드를 삽입할 수 있게 되었다.

2.2 세부 구현 사항

아래 그림은 전체적인 프로세스를 설명하고 있다. 사용자가 목록 페이지에서 특정 페이지 링크를 클릭하여 서버에 요청하면 자바스크립트 필터는 웹 페이지에 로그 모듈 코드를 삽입한다. 사용자의 웹 브라우저로 전송된 특정페이지는 로딩됨과 동시에 사용자의 행위 정보를 수집하여 로그 서버로 전송하게 된다.

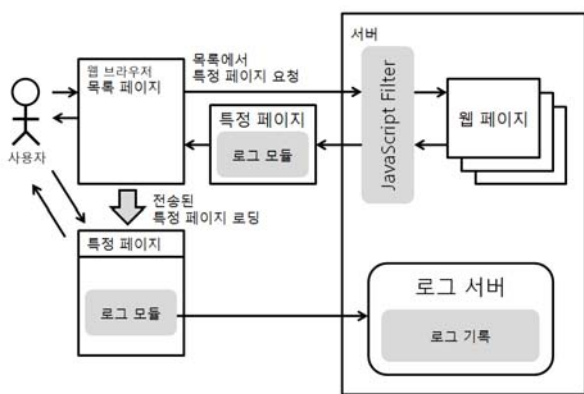


그림 1. 로그 수집을 위한 전체 프로세스

세부 구현 사항은 다음과 같다.

2.2.1 로그 모듈

로그 모듈은 웹 페이지 내부에 첨가되는 코드이다. 사용자 행위 정보 수집과 Ajax 를 통해 수집된 데이터를 서버에 전송하는 역할을 한다. 사용자 행위 정보 수집과 Ajax 코드는 모두 JavaScript 로 구성되어 있어 로그 모듈을 외부 파일로 분리할 수 있었으며 세 줄의 코드를 페이지에 추가함으로써 로그모듈을 동작하게 할 수 있다.

사용자의 행위 정보는 JavaScript 가 제공하는 이벤트 정보를 가로채 수집함으로써 가능해 진다. 웹 브라우저는 사용자가 하는 행위에 따라 반응하기 위하여 이벤트를 발생시킨다. 이 때 JavaScript 에서 이 이벤트를 검출해 낼 수 있다. JavaScript 를 이용해 어떠한 이벤트가 발생하는지 검출하여 그 이벤트 별로 데이터를 축적하며, 주기적 혹은 페이지에서 벗어날 때 서버에 수집된 정보를 전송 가능하다. 아래 그림 2 는 Ajax 가 어떻게 동작하는지 기존의 방식과 비교를 통해 설명하고 있다.

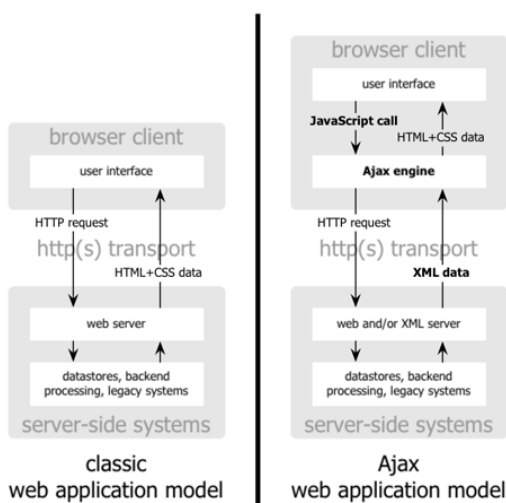


그림 2. 기존의 웹 서버 통신과 Ajax 통신의 차이[2]

Ajax 가 알려지기 이전에는 이와 같은 사용자 정보 수집을 위하여 실험용 브라우저를 구현하거나 운영체제에

발생하는 이벤트 정보를 수집하는 에이전트를 만들어 사용해 왔다. 그러나 이런 방법들은 일반적인 웹 사이트에 적용할 수 없으며 추가적인 프로그램 설치의 사용자들의 반감을 불러일으키고 나아가 보안적인 문제를 발생시킬 수 있다[3]. 따라서 순수한 JavaScript 만을 이용한 본 논문의 방법은 기존의 방법보다 사용자 친화적인 우월성을 지니게 된다.

2.2.2 자바스크립트 필터(Javascript Filter)

자바스크립트 필터는 정보 수집을 위한 로그 모듈의 코드를 웹 페이지에 자동으로 삽입해주는 역할을 한다. 필터를 통해 기존의 페이지를 손대지 않고 로그 모듈 추가가 가능하다. 따라서 대형 포털과 같은 시스템에서 본 논문에서 이용한 평가 모델을 도입할 때 페이지 수정에 대한 이슈를 해결할 수 있다.

아래 HTML 코드는 자바스크립트 필터에 의해 웹 페이지에 자동으로 기록되는 코드이다. 이 코드가 기록됨에 따라 수정된 웹 페이지가 사용자의 브라우저에 로딩됨과 동시에 사용자의 정보를 수집하기 시작한다.

```
<script type='text/javascript'>
  var thisUrl = 'http://...';
  var pageSig = '34234234';
</script>
<script src='/js/ahCollector.js'></script>
```

thisUrl 변수는 현재 페이지의 고정 URL 을 의미하며 수집된 사용자 행위 정보가 서버로 전송될 때 같이 전송되어 정보가 어떤 페이지에서 수집되었는지 구분할 수 있게 한다. pageSig 는 페이지 접속 시 매번 새로 임의로 부여되는 번호이다. 같은 페이지더라도 새로 로딩되어 중복으로 방문한 경우에 이전에 방문한 기록과 구분할 수 있도록 추가하였다. 이 변수 역시 서버로 정보가 전송될 때 함께 전송된다.

아래 그림 3 은 자바스크립트 필터의 내부에서 어떤 일을 수행하는지 보여준다. 자바스크립트 필터는 필터 적용 대상을 지정할 수 있어 모든 페이지에 적용하지 않고 임의 혹은 지정된 페이지의 정보만 수집하도록 할 수 있다. 따라서 예상되는 서버에의 부하를 경감시킬 수 있다.

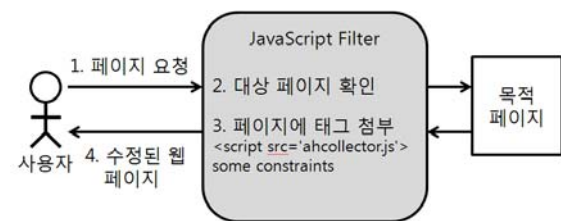


그림 3. 자바스크립트 필터

2.2.3 로그 서버

로그 모듈을 통해 수집된 정보는 로그 서버로 전송된다. 로그서버는 이 정보를 처리하여 저장한다. 이 때 데이터베이스를 이용해 수집된 데이터를 정리하게 된다. 데이터베이스에 저장된 데이터는 쿼리를 이용하여 다양한 방법으로 조회가 가능하다. 여러 가지 측면으로 통계를 구해 분석할 수 있다. 또한 이 로그 서버에 페이지 평가 모듈을 추가한다면 수집된 정보를 바탕으로 웹 페이지의 가치를 구할 수 있게 된다.

3. 실험

실험은 '추천 영화' 키워드로 검색된 웹 페이지에서 선별된 20 개를 이용해 진행되었다. 참여인원은 46 명이다. 실험 대상은 간단한 신상정보를 입력한 뒤 시스템에 로그인 하게 되고, 그 뒤 검색 결과 리스트를 보게 된다.

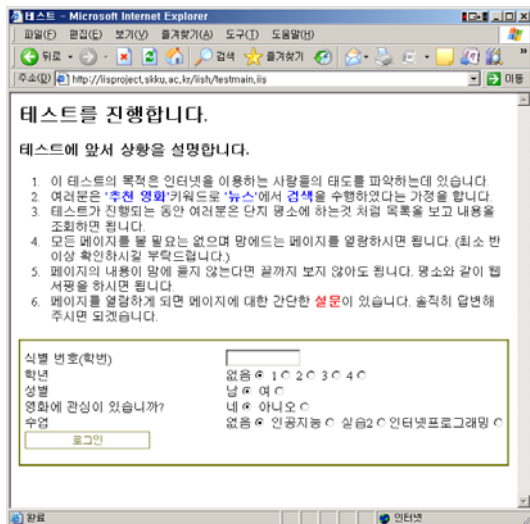


그림 4. 로그인 페이지와 개인 신상 설문

아래 그림 5 는 뉴스 카테고리에서 수집된 페이지 20 개를 나열한 것이다. 접속 시마다 순서를 임의로 배열하여 상위에 나열된 것에 사용자가 집중되는 것을 방지하였다.

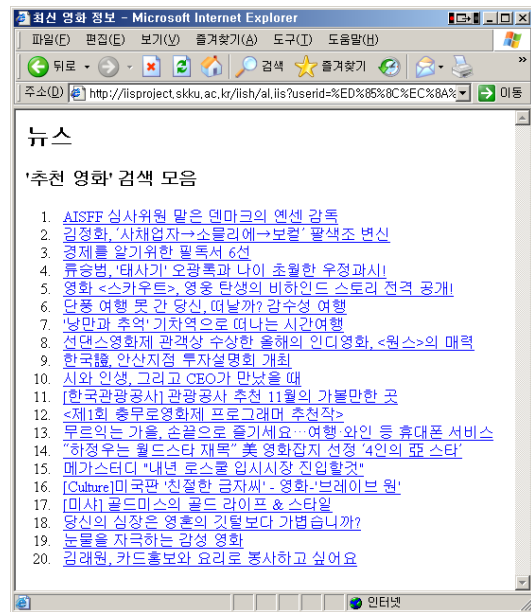


그림 5. 로그인 페이지와 개인 신상 설문

실험대상자에게는 사전에 실험 중 평소에 웹 서핑을 하듯이 자연스럽게 볼 것을 당부하였다. 리스트에 나열된 제목을 클릭하여 각 페이지에 접속하게 되면 로그 모듈은 자동으로 정보를 수집하여 서버에 전송한다. 이 때 JavaScript 만을 이용하므로 사용자의 컴퓨터에는 아무것도 설치되지 않는다. 각 페이지 우측에는 페이지가 검색 의도인 '추천 영화' 에 맞는지 묻는 설문이 있다. 이 설문은 1 에서 5 사이의 라디오 버튼으로 구성되어 있으며 점수가 높을수록 긍정을 의미한다.



그림 6. 사용자에게 노출되는 기사 페이지와 설문조사

피실험자의 컴퓨터의 사양에는 제한을 두지 않았으나 웹 브라우저의 경우 '한글 Internet Explorer 6' 로 제한하였다. 브라우저에 따라 이벤트 수집에 대한 구현이 조금씩 달라 수집되는 정보에 있어 작지만 차이가 발생하였고 이를 방지하기 위하여 브라우저에 대한 제한을 두고 실험하였다.

3. 결과

수행된 실험은 다각도의 분석이 필요하다. 382 개의 수집된 페이지 조회수(Page view, PV)를 분석하였다. 아래 표 1 은 기본적인 통계이다. 설문지에 응답한 평가 점수 별 사용자 행위 정보의 평균을 나타낸다. 일반적으로 설문 점수가 높을수록 사용자의 이용 시간과 행위가 높게 수집되는 모습을 보이나, 일부의 경우 기대한 것과 일치하지 않는 값이 수집됨을 볼 수 있다.

표 1. 점수 별 사용자 행위 데이터의 평균

설문 점수	이용 시간	마우스 휠	마우스 클릭	블록 선택	마우스 움직임
1	14.12	10.43	0.75	7.57	161.24
2	18.30	13.72	1.04	3.75	266.30
3	24.33	17.00	1.84	10.51	323.33
4	23.97	15.55	1.15	6.22	310.07
5	30.79	19.53	1.21	1.98	318.98

그림 7 은 좀 더 쉬운 비교를 위하여 표 1 을 그래프로 표현한 것이다. 마우스 움직임을 뜻하는 mousemove 항목의 경우 숫자의 단위가 다른 항목들에 비해 커 10 으로 나눈 숫자를 사용하여 그래프를 표현하였다.

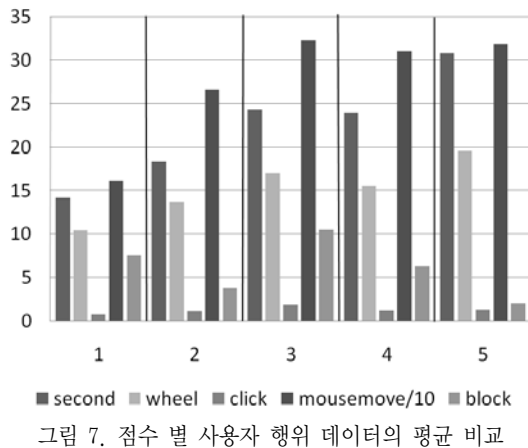


그림 7. 점수 별 사용자 행위 데이터의 평균 비교

표 1 과 그림 7 을 통해 볼 때 사용자들은 3 점을 준 페이지들에서 문제점을 보여주고 있다. 조사 결과 3 점 받은 페이지의 콘텐츠가 다른 페이지들에 비해 좀더 텍스트가 많거나 전체 내용을 살펴야 하는 경우가 많았다. 그리고, 이용시간을 뜻하는 second 와 마우스 휠을 뜻하는 wheel, 마우스 움직임을 뜻하는 mousemove 는 대체적으로 설문조사 점수가 높아짐에 따라 증가하는 모습을 보여주고 있으나 마우스 버튼 클릭을 뜻하는 click 과 블록 선택을 뜻하는 block 의 경우 본 실험의 가정과 벗어나고 있음을 볼 수 있다.

표 2 는 설문조사에서 점수 별 사용자들의 평가 횟수를 나타낸 것이다. 사용자들은 주로 2, 3, 4 점을 많이 주고 있음을 볼 수 있으며 5 점과 같은 높은 점수의 경우 가장 적은 횟수를 기록하였다.

표 2. 점수 별 사용자들의 평가 횟수

설문 점수	1	2	3	4	5
평가 수	75	82	83	89	53

수집된 정보를 이용해 웹 페이지 평가를 하기 위해서는 각 사용자 행위의 빈도가 사용자들이 느끼는 페이지에 대한 평가와 비례하여야 한다. 따라서 위에서 제시한 것에 추가적으로 행위 별로 개별 적인 분석을 수행하였다. 아래 그림 8~12 는 사용자의 각각의 행위 별 설문 점수와의 상관관계를 나타낸 그래프이다. 그래프의 왼쪽 변은 수집된 행위의 횟수이며, 아래 변은 설문 점수이다. 설문 점수 별로 그래프 막대가 위치하는데, 막대 오른쪽에 있는 숫자의 의미는 다음과 같다. 맨 위에 위치한 숫자는 설문 점수를 부여한 피실험자 중 가장 많은 반응의 횟수이다. 가운데 위치한 숫자는 설문 점수를 부여한 피 실험자들의 반응 횟수의 평균이며, 맨 아래에 위치한 숫자는 최소 반응 횟수를 뜻한다.

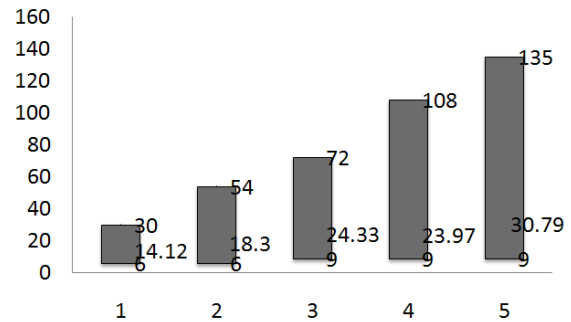


그림 8. 점수 별 사용자 이용 시간에 따른 데이터의 수치 비교

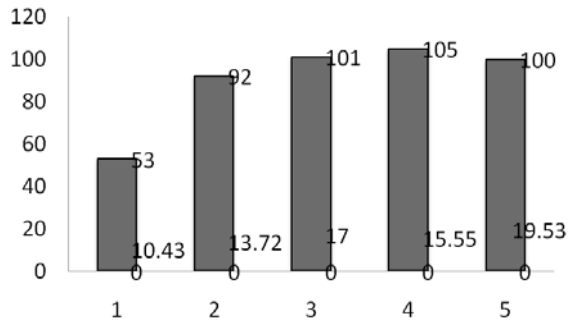


그림 9. 점수 별 사용자 마우스 휠 누적 수치에 따른 데이터의 수치 비교

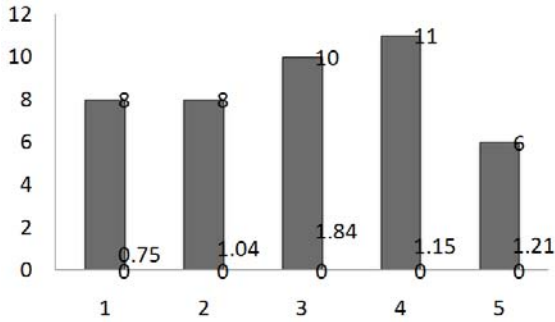


그림 10. 점수 별 사용자 마우스 버튼 클릭 누적 수치에 따른 데이터의 수치 비교

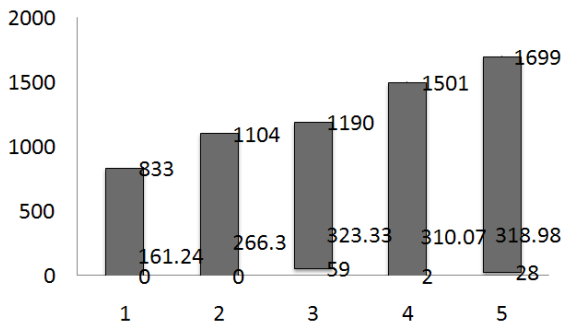


그림 11. 점수 별 사용자 마우스 움직임 누적 수치에 따른 데이터의 수치 비교

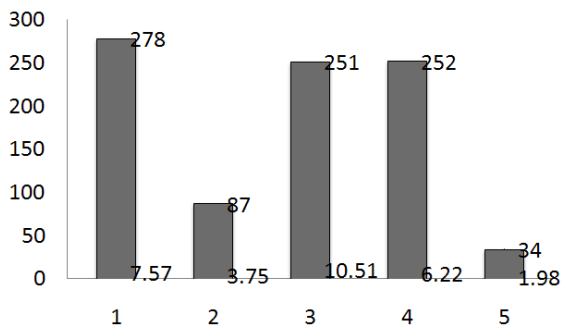


그림 12. 점수 별 사용자 블록 선택 누적 수치에 따른 데이터의 수치 비교

그림 12의 사용자 블록 선택 누적수치에 따른 데이터의 수치 비교 그래프의 경우에만 최소 누적 값이 모두 0 이어 생략하였다. 따라서 막대의 아래쪽에 위치한 수치는 평균값을 의미한다.

앞에서 언급한 것과 같이 그래프들을 통하여 볼 때 마우스 버튼 클릭과 블록 선택 행위는 웹 페이지 평가의 기준으로 알맞지 못함을 볼 수 있다. 따라서 Ajax 를 이용하여 웹 페이지 평가를 위해서 수집되어야 하는 사용자 행위는 이용시간, 마우스 휠 움직임, 마우스 움직임의 세가지로 압축된다.

4. 결론

사용자의 행위 데이터를 이용하여 웹 페이지를 평가하는 방법을 제안하기 위해 수행된 본 실험은 실험을 통해 JavaScript 의 이벤트 중 페이지 평가에 유효하게 사용될 수 있는 항목을 찾아냈으며, Ajax 로 이 행위 정보를 수집할 수 있는 구체적인 방법을 제공하였다. 추후 이 기준을 이용하여 웹 페이지를 평가하는 기계학습 방법을 통한 데이터 마이닝 적용이 필요할 것으로 보인다[4]. 또한 설문점수 3 점의 경우 4 점보다 높은 수치가 나오는 사용자 행위 항목이 발견되었다. 이를 고려하기 위해서는 콘텐츠의 양과 같은 추가적인 항목이 필요한 것으로 보인다. 추후 사용자의 성향과 웹 페이지의 콘텐츠의 관계에 대한 연구가 필요하다.

참고문헌

- [1] A. Gulli, A. Signorini, "The indexable web is more than 11.5 billion pages", the 14th international conference on World Wide Web, pp. 902-903, 2005.
- [2] J. Garrett, "Ajax: A New Approach to Web Applications", <http://www.adaptivepath.com/publications/essays/archives/000385.php>, 2005
- [3] R. Baraglia, F. Silvestri, "Dynamic personalization of web sites without user intervention", Communications of the ACM, vol. 50, issue 2, New York:ACM, pp.63-67, 2007.
- [4] E. Alpaydin, "Introduction to Machine Learning," The MIT Press, pp. 229-273, 2004.