

# 클릭스트림 분석을 이용한 즐거찾기 메뉴 자동 생성 시스템

권준아, 손철수, 김원중  
순천대학교 컴퓨터학과  
e-mail:kwj@sunchon.ac.kr

## Automated Favorite Menu Creating System On Clickstream Analyzing

Jun-a Kwon, Cheol-su Son, Won-Jung Kim  
Dept. of Computer Science, Sunchon National University

### 요 약

인터넷을 이용한 정보의 공유, 활용 및 전자상거래가 활성화되면서 많은 양의 콘텐츠가 웹 사이트에서 서비스되고 있다. 사용자는 신속하게 정보를 획득하기 위하여 웹 브라우저의 즐겨찾기 기능을 이용한다. 기존의 즐겨찾기 기능은 사용자가 해당 URL을 즐겨찾기에 등록할 것인지를 판단하고 수작업으로 등록하여 관리해야하는 문제점을 가지고 있다. 본 논문에서는 즐겨찾기 목록 관리의 문제점을 해결하기 위하여 사용자가 웹 브라우저를 이용하여 사이트 방문시 발생하는 클릭 스트림을 사용자 컴퓨터에 저장하고 그 자료를 분석하여 즐겨찾기 목록에 해당 URL을 자동으로 등록하고, 또한 즐겨찾기 목록이 동적으로 관리될 수 있는 즐겨찾기 메뉴 자동생성 시스템을 구현 하였다.

### 1. 서 론

인터넷을 이용한 정보의 공유, 활용 및 전자상거래가 활성화되면서 많은 양의 콘텐츠가 웹 사이트에서 서비스되고 있으며 지금도 계속 증가하고 있다. 사용자에게는 많은 양의 웹 콘텐츠 중에서 자신이 필요한 정보를 빠른 시간 안에 정확하게 찾는 것이 매우 중요하다. 인터넷 사용자의 신속한 정보 획득을 위하여 정보 검색 엔진뿐만 아니라 사용자별 맞춤형 동적 화면 제공 시스템 그리고 사용자가 즐겨찾는 URL 관리 등이 사용되고 있다.

정보 검색 엔진의 사용은 대량의 데이터베이스에서 정보를 검색할 수 있다는 장점은 있으나 재현율과 정확율이 낮고 다음에 다시 정보를 신속하게 사용하기 위하여 특정한 방법으로 저장하거나 즐겨찾기에 등록하여야 한다.

사용자가 접속한 웹 사이트에서 사용자의 클릭스

트림을 분석하여 사용자별 관심사항이나 선호도에 따른 맞춤형 동적 메뉴 또는 화면을 제공하는 개인화 방법도 있으나, 이러한 기능을 구현하기 위해서는 사이트마다 클릭스트림 분석과 사용자 맞춤형 화면 시스템이 구현되어야 하는 문제점이 있다[1].

사용자의 브라우저에서 관리되고 있는 즐겨찾기 기능은 사용자가 방문하고자 하는 사이트의 URL을 기억하고 있지 않더라도 차후에 접근이 쉽도록 한다. 그러나 현재의 즐겨찾기 목록의 관리 방식은 다음 2가지의 단점을 가지고 있다. 첫째, 사용자가 사이트를 즐겨찾기 목록에 관리할 것인지 여부를 직접 판단하여야 하는 불편한 점이 있다. 둘째, 즐겨찾기 목록에 등록하는 것도 사용자가 수동적으로 하여야 한다는 것이다.

본 논문에서는 위와 같이 정보 검색 엔진, 사용자 맞춤형 동적 화면 제공 시스템 그리고 즐겨찾기 목록 관리의 문제점을 해결하기 위하여 사용자가 웹

브라우저를 이용하여 사이트 방문시 발생하는 클릭 스트림을 사용자 컴퓨터에 저장하고 그 자료를 분석하여 즐겨찾기 목록에 해당 URL을 자동으로 등록하고 또한 즐겨찾기 목록이 동적으로 관리 될 수 있는 즐겨찾기 메뉴 자동생성 시스템을 구현 하였다.

## 2. 관련연구

### 2.1 웹 로그

웹 서버는 사용자의 웹 서비스에 대한 요청과 제공하는 정보를 로그 파일에 저장하게 된다. CLF(Common Logfile Format)는 NCSA (National Center for Supercomputing Applications) 계열의 웹 서버에서 사용하는 표준 로그 파일 형식으로 Access Logfile 또는 Transfer Logfile이라고 불리며 Host, AuthUser, Time, Request, Status, Volume과 같은 필드들로 구성되어 기록된다[2].

### 2.2 웹 마이닝

웹 마이닝은 데이터 마이닝의 한 분야로서 사용자들의 클릭스트림의 집합체인 로그파일을 이용하여 사용자들의 패턴을 분석하는 기술이다. 웹 마이닝은 크게 Web Content Mining, Web Usage Mining, Web Structure Mining의 세 가지로 분류 된다[3].

### 2.3 클릭스트림

클릭스트림(Clickstream)은 웹 사이트에서 사용자의 이동 경로를 분석할 수 있는 모든 행위 패턴 정보를 말한다. 사용자가 웹 사이트에 접속한 순간부터 웹 페이지를 이동하기 위해 마우스를 이용하여 클릭 이벤트를 발생시키는 일련된 흐름의 정보들이다.

### 2.4 개인화

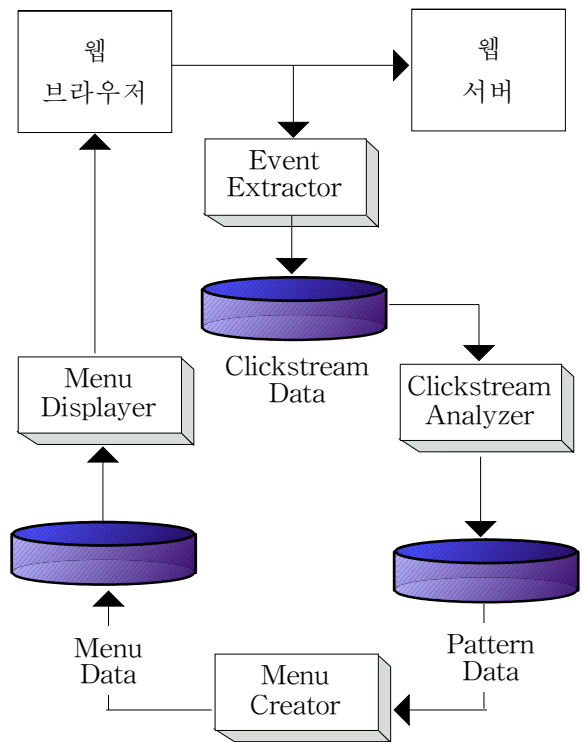
개인화(Personalization)된 웹 사이트는 사용자에게 무엇을 원하는지 직접 물어보는 것이 아니라, 사용자들의 특성을 통계학 및 인공지능 기법을 활용하는 데이터마이닝을 통해 분류 또는 세그먼트화하여 개인이 원하는 것을 예측하여 서비스하는 것이다 [3][4]. 예를 들면 전자 상거래 업체들은 사용자의 개인적인 취향에 따라 자신의 페이지를 구성하고 사용자의 구매 기록이나 접속 기록에 따라 제품을 추천 받을 수 있는 기능을 제공한다.

## 3. 즐겨찾기 메뉴 생성 시스템

본 연구에서는 사용자가 웹 브라우저를 통하여 접속하는 사이트들과 URL의 클릭스트림을 분석하여 즐겨찾기 메뉴 자동 생성 시스템을 구현하였다.

### 3.1 시스템 구조

[그림 1]은 클릭스트림 분석을 이용한 즐겨찾기 메뉴 자동 생성 시스템의 구조를 보여준다. 사용자가 웹 브라우저에서 URL 링크를 클릭하여 웹 서버의 페이지를 요청할 때 Event Extractor가 클릭 스트림을 추출하여 Clickstream Data를 축적한다. 그리고 Clickstream Analyzer는 요약된 Pattern Data를 생성한다. Menu Creator는 Pattern Data로부터 추출한 URL을 Menu Data에 추가하거나 수정한다. 웹 브라우저에 플러그인 된 Menu Displayer는 Menu Data를 이용하여 즐겨찾기 목록을 사용자에게 제공한다.



[그림 1] 시스템 구조도

### 3.2 Event Extractor

정확한 즐겨찾기 메뉴 생성과 동적인 관리를 하기 위하여 Clickstream Data에는 사용자가 클릭한 URL과 접속 일시를 관리하여야 한다. 또한 효율적인 분석과 자료 관리를 하기위하여 전처리 과정을

거쳐야 한다[5]. 전처리 과정에서는 클릭스트림 데이터 중에서 분석에 필요하지 않은 데이터를 제외시키는 데이터 정제(Data Cleaning) 과정을 거쳐야 한다. [표1]은 Clickstream Data Table의 스키마이다.

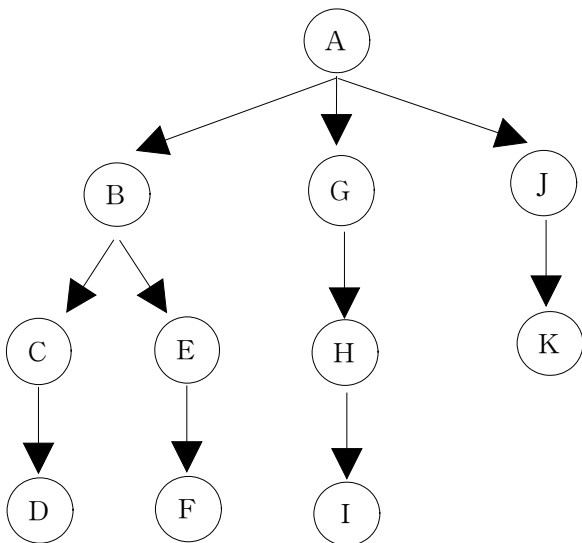
[표 1] Clickstream Data Table Schema

No.	Field Name	Type
1	Access_Time	Datetime
2	URL	VARCHAR2(128)
3	DOC_SIZE	NUMBER
4	CONN_DATE	VARCHAR2(14)

### 3.3 Clickstream Analyzer

저장된 모든 웹 페이지 URL 정보를 메뉴 생성 시스템의 Menu Creator 입력 데이터로 사용하지 않는다. Clickstream Analyzer는 사용자가 찾고자하는 콘텐츠 URL 정보만을 분석해서 일정시간 이상의 체류시간을 가진 웹 페이지 URL 정보만을 Menu Creator의 입력 데이터인 Pattern Data에 저장한다. 체류시간은 사용자가 브라우저를 통해 특정 웹 페이지에 실제로 머문 시간을 말한다. 체류시간이 매우 짧은 웹 페이지는 사용자가 이 페이지를 실수로 열었거나 실제 찾고자하는 웹 페이지까지의 이동 경로 역할을 하는 경우가 대부분이다[5].

본 논문에서는 Menu Creator를 위한 입력 데이터 정제를 위해 Clickstream Analyzer가 체류시간의 평균을 구하고 웹 페이지의 데이터 크기를 비교하여 분석하였다. [그림 2]는 Clickstream Analyzer가 분석한 사용자 패턴을 트리형식으로 나타낸 것이다.



[그림 2] 분석된 사용자 패턴 트리

[표 2]는 사용자의 패턴 트리중 하나를 추출하여 패턴 데이터의 평균 체류시간과 데이터의 크기를 나타낸 것이다. 사용자는  $A \Rightarrow B \Rightarrow C \Rightarrow D$  경로로 이동하였다.

[표 2] 웹 페이지별 체류시간과 데이터 크기

웹 페이지	체류시간	데이터 크기
A	2 초	48464 Byte
B	2 초	14913 Byte
C	8 초	71488 Byte
D	30 초	77244 Byte

A 페이지에서의 체류 시간을 계산하기 위해서는 B 페이지에 접속한 시간에서 A 페이지에 접속한 시간을 빼면 된다. 각각의 웹 페이지 체류시간 중 데이터 크기가 비슷한 C와 D페이지의 경우 D 페이지는 평균 체류시간이 27초이지만 C 페이지에서의 평균 체류시간은 7초로 많은 차이가 있다. 따라서  $A \Rightarrow B \Rightarrow C$  까지의 정보는 단지 D 페이지를 찾아오기 위한 경로로만 사용된 것을 알 수 있기 때문에 D 페이지만 입력데이터로 사용된다. 그리고 사용자가 D 페이지에서 더 이상의 클릭스트림을 발생시키지 않았을 경우 세션이 Time Out되고 체류시간을 구할 수 없기 때문에 메뉴 생성 시스템의 입력 데이터로 사용하지 않고 정제하게 된다. D 페이지가 메뉴 생성 시스템의 입력 데이터로 사용되어 사용자에게 제공 되었을 경우  $A \Rightarrow B \Rightarrow C \Rightarrow D$  경로를 통해 D 페이지에 접근하는데 소요되는 평균 시간 13초 중 B와 C 페이지를 거치지 않아도 된다. 이 사용자의 경우 사용자별 메뉴를 제공하였을 때 D 페이지에 접근하는데 평균 10초 이상의 시간이 줄었다. 또 다른 패턴 데이터에서도 사용자별 메뉴를 제공하였을 경우 경로 역할을 하는 페이지에서 소요되는 시간이 줄어드는 결과를 얻을 수 있었다. [표 3]은 Pattern Data Table의 스키마이다.

[표 3] Pattern Data Table Schema

No.	Field Name	Type
1	URL_ID	NUMBER
2	URL	VARCHAR2(128)
3	ORDER_NUM	NUMBER
4	GUBUN_CODE	CHAR(1)
5	DWELL_TIME	NUMBER
6	VIEW_COUNT	NUMBER
7	SUPPORT_RATE	NUMBER

### 3.4 Menu Creator

즐거찾기 메뉴를 자동으로 생성하는 Menu Creator는 Pattern Data에 저장된 자료에서 새로운 URL을 추출해서 기존의 Menu Data에 추가하는 역할을 한다.

### 3.5 Menu Displayer

즐거찾기 메뉴는 사용자가 방문한 각각의 웹 페이지별 체류시간이 길고 클릭스트림 데이터중에서 방문 회수가 많은 순서로 출력하였다. 또한 새로운 웹 페이지가 추가되었을 때 사용자의 방문 회수가 적어 즐거찾기에서 제외되는 경우를 해결하기 위해 사용자가 해당 URL을 고정형으로 설정할 수 있는 기능을 제공한다. [표 4]는 Menu Data Table의 스키마이다.

[표 4] Menu Data Table Schema

No.	Field Name	Type
1	MENU_ID	VARCHAR2(20)
2	URL_ID	VARCHAR2(20)
3	TITLE	VARCHAR2(50)
4	ORDER_NO	NUMBER
5	ACTIVE_YN	CHAR(1)
6	REG_DATE	NUMBER

## 4. 결 론

본 논문에서는 사용자가 보다 신속하고 정확하게 인터넷에서 자료를 획득할 수 있도록 기존의 웹 브라우저에서 제공하는 즐거찾기 목록 관리의 수동성 문제를 제기하였다. 또한 이를 해결하기 위하여 사용자가 인터넷의 웹 콘텐츠를 획득하는 과정에서 웹 브라우저에서 발생하는 클릭스트림을 추출하여 저장하고, 저장된 클릭스트림 정보를 기반으로 평균 체류시간을 계산, 정제 하고 분석하여 즐거찾기 메뉴를 자동으로 생성하는 시스템을 구현하였다.

사용자는 이러한 즐거찾기 메뉴 자동 생성 시스템을 설치하여 인터넷 웹사이트의 콘텐츠를 사용함으로써 즐거찾기를 추가하기 위한 심적인 부담이 줄어들고 통계에 의한 동적인 즐거찾기 메뉴가 생성됨으로써 활용성을 높일 수 있다.

즐거찾기 목록에 저장된 URL의 존재 유무를 점검하여 유효성을 유지하고 URL의 웹 페이지 내용이 변경될 경우 사용자가 인식할 수 있도록 하는 방안을 추가적으로 연구할 예정이다.

## 참고 문헌

- [ 1 ] 이은경, 이기현, 조근식, “사용자 군집을 이용한 개인화 된 웹 페이지 추천”, 한국정보과학회 2002년 춘계학술대회, pp.0241-0243, 2002
- [ 2 ] <http://www.w3.org/Daemon/User/Config/Logging.html>
- [ 3 ] Mobasher, B., Cooley, R., Srivastava, J., “Web Mining: Information and Pattern Discovery on the World Wide Web,” In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence(ICTAI'97), November, 1997.
- [ 4 ] Personalization Consortium  
<http://www.personalization.org>
- [ 5 ] Srivastava, J. Cooley, R., Deshpande, M. & Tan P.N. Web Usage Mining: Discovery and Application of Usage Patterns from Web Data. SIGKDD Explanations, 1. 2000.