

웹기반 교육에서 학습자별 학습현황 분석에 관한 연구

신지연[†] · 정옥란^{††} · 조동섭^{†††}

요 약

웹기반 교육에서 학습과정을 평가한다는 것은 개별 학습자들의 학습 활동을 평가하는 것을 의미하기 때문에 학습자의 특정 수업내용에 대한 학습 시간, 학습 패턴, 학습 참여도(의견 교환, 질문), 학습 환경 등의 정보가 요구된다. 본 연구의 목적은 웹 기반 교육에서 쟁점이 되고 있는 학습과정 평가문제를 해결하기 위해 최적의 웹 로그 마이닝을 이용하여 학습자 개인별 학습현황에 관한 정보를 얻어 이를 수행 평가에 반영하고자 함이다. 연구 내용 및 결과로는 먼저, 학습현황 분석을 위한 항목을 선정하고 웹 로그 마이닝을 위한 로그 데이터 전처리 과정을 실행하였다. 다음으로는, 위의 웹 로그 데이터를 기초로 학습자별 데이터베이스를 구축하고 질의어를 사용하여 학습현황을 분석하였다.

The Analysis of Individual Learning Status on Web-Based Instruction

Ji-Yeun Shin[†] · Ok-Ran Jeong^{††} · Dong-Sub Cho^{†††}

ABSTRACT

In Web Based Instruction, as evaluation of learning process means individual student's learning activity, it demands data on learning time, pattern, participation, environment in a specific learning contents. The purpose of this paper is to reflect analysis results of individual student's learning status in achievement evaluation using the most suitable web log mining to settle evaluation problem of learning process, an issue in web based instruction. The contents and results of this study are as following. First, conformity item for learning status analysis is determined and web log data preprocessing is executed. Second, on the basis of web log data, I construct student's database and analyze learning status using data mining techniques.

1. 서 론

웹 기반 교육은 전통적 교실 수업에 비해 수요자 중심의 학습 방법으로 교육 현장에서 인정되고 있다.

고 있지만, 교육 과정 면에서 다양한 사용자의 요구를 수렴하기 어렵고, 학습자의 모든 피드백 상황을 만들어 주지 못하고 있다. 또한, 학습자 수업통제 및 학습상황, 학습 참여도 파악이 어렵다. 이러한 문제를 해결하기 위해서는 학습자의 요구를 정확히 파악하여 개인별 학습을 가능하게 하고, 실시간으로 학습자의 학습현황을 파악하여 적절한 피드백을 제공해야 한다. 평가의 윤용 방안으로 학습자의 학습경로나 학습과정을 확인하는 방법 등을 동원한 질적인 평가가 되어야 한다 [12]. 이를 위해서는 학습에 대한 철저한 분석이 선행되어야 한다. 웹서버는 웹서비스에 대한 요구

[†] 정회원: 이화여자대학교 교육대학원 석사

^{††} 정회원: 이화여자대학교 컴퓨터학과 박사과정

^{†††} 종신회원: 이화여자대학교 컴퓨터학과 교수

논문접수: 2002년 9월 6일 심사완료: 2003년 4월 14일

*본 논문은 2002년도 두뇌한국21사업에 의하여 지원되었음.

청과 제공에 대한 기록을 모두 로그파일이라는 파일에 저장한다. 웹 엑세스 로그는 방문자가 웹 서버에 접속을 할 때 생성되는 파일로서 로그 파일을 통해 파악할 수 있는 사실은 방문자의 방문 경로, 접속 도메인(Host), 시간대, 요일, 전송프로토콜, 접속상태 및 데이터 이동상태 현황, 전송된 데이터량 등의 추이도 파악할 수 있다[9]. 웹 로그 마이닝은 웹 데이터로부터 방문자의 엑세스 패턴(access pattern)을 유도하기 위해 데이터 마이닝 기술을 적용한 웹 마이닝의 한 분야(Web Usage Mining)이다[9]. 이는 여러 가지 측면에서 상당히 중요한 의미를 갖는다. 교육 웹사이트 운영자에게는 학습자의 관심도, 학습량 혹은 학습 패턴, 학습 참여도 등을 측정할 수 있는 주요 지표로써 활용될 수 있기 때문이다. 그리고 이러한 분석 결과는 웹 기반 학습의 운영에서 가장 중요한 평가에 반영된다[15].

웹을 활용한 웹 기반 교육에 있어서 쟁점이 되고 있는 문제 중 하나는, 이 같은 교육의 결과를 어떻게 평가할 것인가 하는 문제이다. 학습과정을 평가한다는 것은 개별 학습자들의 학습 활동을 평가하는 것을 의미한다. 웹 기반 교육에서는 학습자의 특정 수업내용에 대한 학습 시간, 학습 패턴, 학습 참여도(의견 교환, 질문), 학습 환경 등의 정보를 수집하고 이를 평가하는 것이 요구된다. 웹 로그 마이닝을 통해 이러한 유용한 정보를 얻을 수 있다. 웹 로그 파일로부터 학습자별 학습현황을 분석한 다음, 이를 수행평가에 반영하면 된다[12]. 웹 로그 데이터는 그 자체만으로는 한계가 있다. 보통의 로그 분석 툴은 IP(Internet Protocol)로 유일 방문자를 카운트하게 되어 있는데, 대부분의 ISP들은 유동 IP 할당 방법을 사용하고 있기 때문에, 한 방문자가 여러 번 접속했을 경우, 서로 다른 방문자로 카운트될 우려가 있기 때문에 정확한 데이터라고 볼 수 없다. 또한, 한 사람이 여러 대의 컴퓨터를 사용해서 접근하거나 혹은 한 컴퓨터를 여러 사람이 쓰는 경우에도 정확하지 않다고 볼 수 있다[11]. 이에 본 논문은 웹 로그 정보, 학습자 정보 그리고 세션 및 쿠키 정보를 연계하여 ID 매칭의 최적의 웹 로그 마이닝을 이용하여 최대한 정확한 정보를 이끌어내는데 중점을 두었다. 연구 내용 및 결과는 다음과 같다. 첫째, 이론적 배경이 되는 웹 기반 원격 교육 및 웹 마이닝의 개요 조사가

이루어졌다. 둘째, 학습현황 분석을 위한 항목을 선정하고, 최적의 웹 로그 마이닝을 위한 로그 데이터 전처리 과정을 실행했다. 셋째, 위의 웹 로그 데이터를 기초로 학습자별 데이터베이스를 구축하고 질의어를 사용하여 학습현황을 분석하였다. 분석결과, 학습자 개인별 학습 진도(단원별·영역별 학습 시간, 총히트수, 연·월·주별 히트수, 마지막 일주일 히트수, 최다 참조 사이트), 학습 패턴, 학습 참여도(게시판, 웹채팅, 이메일, 공지사항, 자료실 이용시간), 학습환경(onsite·offsite 히트수, 운영체제별 히트수, 브라우저별 히트수) 정보를 얻을 수 있었다.

2. 웹 마이닝과 웹 기반 원격교육

2.1 웹 마이닝 정의

웹의 확산과 함께 이로부터 얻을 수 있는 정보 역시 방대하게 되었다. 이는 관련 정보 템색의 어려움, 다량의 정보로부터 지식의 창출, 고객 또는 사용자에 대한 학습, 정보의 개인화 등의 문제를 대두시켰다. 이러한 문제들을 해소하기 위해 웹으로부터 정보를 얻을 수 있는 자동화된 툴이 필요하게 되었고, 얻은 정보를 이용하여 웹 사용자들의 패턴을 식별할 수 있는 방법 또한 필요하게 되었다. 웹 데이터 마이닝이란 웹으로부터 얻어지는 정보를 찾아내어 분석하는 것을 말한다. 즉, 웹에서 발생하거나 웹사이트에 저장한 데이터를 대상으로 유용한 패턴을 찾아내는 것이다[12]. 트래픽, 등록과 거래 정보, 사용 패턴 등 인터넷에서 벌어지는 모든 웹 데이터를 전통적인 데이터 마이닝 기법에 접목시킨 기술로, 실시간으로 활용할 수 있다. 그러므로 웹 데이터 마이닝은 실시간으로 웹 데이터를 분석하여 진정한 의미의 개인화 서비스를 가능하게 하며 CRM, SCM등에 적용될 수 있다.

2.2 웹 마이닝 방법

웹 마이닝 단계는 전처리 과정, 패턴 템색, 패턴분석으로 구분된다. 전처리 과정 단계에서는 획득된 데이터를 분석에 적합한 형태로 정제하며

이는 원천 데이터에 비해 용량이 감소하게 된다. 다음은 패턴탐색 과정으로 기술적 통계분석 방법, 연관 분석, 클러스터링, 분류, 순차 분석 등의 기법이 있다. 마지막, 패턴분석 과정에서는 웹 데이터 마이닝을 이용하여 분석된 정보를 사용자들이 보다 더 쉽게 이해할 수 있도록 통계(statistics), 가시화(visualization), 유용성 분석(usability analysis), 데이터베이스 질의(database querying) 등을 이용한다. 사용자의 웹사이트 이용 패턴을 분석하기 위해서는 사용자가 웹을 사용할 때마다 사용자가 웹서버로 보낸 요구를 기록해야 한다. 이 데이터를 웹 로그라고 한다. 웹 로그에서 패턴 발견을 위해 데이터를 추출, 변환, 정제하는 일련의 과정을 전처리 과정이라 할 수 있다. 정확한 결과를 얻기 위해서는 전처리 과정이 매우 중요하다. 대부분의 웹사이트에는 웹서버가 자동으로 생성하는 표준 웹 로그 형식(common log format)으로 저장된 웹 로그 파일만이 있다. 그러나 여기에 저장된 정보의 한계 때문에 실제 사용자의 웹 방문 패턴을 100% 정확하게 얻을 수 없다. 표준 웹 로그 형식의 웹 로그만으로는 사용자를 완전히 구분할 수 없기 때문이다. 따라서 정확한 결과를 얻기 위해 사용자와 세션을 구분할 수 있는 모듈을 웹서버에서 제공하거나, 각각의 페이지에 적당한 실행 코드를 삽입해야 한다. 사용자 구분 처리와 세션 처리 등의 전처리 과정으로 기존 웹서버를 유지한 채 데이터 마이닝을 수행할 수도 있다. 또 사용자 정보, 세션 정보 등 마이닝에 필요한 모든 정보를 웹 로그에 저장하고 있다 해도 웹 로그 내의 필요 없는 정보를 제거하는 등의 이유로 전처리 과정은 필요하다.

웹 사용 마이닝(web usage mining)은 웹 사용자의 사용 패턴을 분석하는 것이다. 이를 통해 웹 사용자의 행동을 접속 통계 정보 이상으로 이해할 수 있고, 또한 웹 페이지의 이용 패턴을 알 수 있게 된다. 결국 이 정보는 사용자에게 더욱 친숙하게 페이지를 재구성하거나, 웹서버 로드 벨런스, 사용자별 맞춤형 웹 페이지 구성 등에 이용된다. 웹 사용 마이닝은 응용 분야도 많을 뿐 아니라, 웹 마이닝에서 주요하게 다루어지는

부분이기 때문에, 현재는 협의의 웹 마이닝을 뜻하기도 한다. 웹 사용 마이닝은 초기 데이터를 마이닝 알고리즘의 입력 형태로 적절하게 바꾸어 주는 전처리(preprocessing)과정, 전처리 과정에서 얻은 데이터에서 유용한 정보를 얻기 위한 패턴발견(pattern discovery)과정, 마지막으로 생성된 규칙과 패턴을 분석(pattern analysis)하는 과정으로 나눌 수 있다[16].

특히, 본 논문의 핵심인 웹 사용 마이닝에 쓰이는 데이터는 객관적 데이터와 주관적 데이터로 구분할 수 있다. 객관적 자료에는 내용(content), 구조(structure), 사용자 프로파일(user profile), 사용(usage) 등이 있으며 주관적 데이터에는 조사, 평가 등이 있다. 이러한 데이터는 서버(해당 사이트의 서버 내에 축적되는 정보), 클라이언트(쿠키, java, CGI등의 프로그램을 이용하여 사용자의 쿠도 추적), 프록시(back, forward 등에 관한 데이터 확보)로부터 얻어질 수 있다[14].

2.3 웹 기반 원격교육

웹 기반 교육은 월드와이드웹에 기반한 교육을 말하는 것으로 월드와이드웹의 특성을 교육의 효과성과 효율성 증진에 활용하는 교육을 의미한다. 웹 기반 교육이 가지는 잠재된 특징은 웹이 가지고 있는 여러 가지 예상하기 어려운 교육적, 비교육적, 교육 중립적인 특성들이 교육에 도입될 수 있다는 점이다. 또한 교육 내용과 방법이 공개되어 교사와 학습자의 모든 상호작용이 객관적으로 관찰될 수 있고 기록되며 교육의 노하우가 축적되어 궁극적으로 교육이 보다 과학적으로 이루어지게 하는 수단이 될 수 있다는 점이다. 또 다른 특징을 살펴보면 웹 기반 교육은 수요자 중심 교육의 개념을 강화한 것이다. 즉, 웹 기반 교육의 시간적·공간적·방법적 융통성은 다양한 부류의 학습자들에게 각기 다른 내용과 방법으로 수요자 중심 교육을 실현할 최적의 대안이 될 수 있다.

시간과 공간의 제약이 있는 집합교육의 한계를 벗어나 학습자의 수요와 요구에 맞는 열린 학습을 가능하게 한다는 장점과 상호작용을 할 수 있다는 특성에 대해 웹 기반 교육은 그 교육적 잠

재력이 무한한 것으로 인정받고 있다. 그러나 그 꿈을 실현하기 위해서 우수한 웹 기반 교수-학습 체제 개발과 웹 기반 교육 프로그램을 개발하는 일이 무엇보다도 우선이라고 할 수 있다. 첨단 정보통신 공학의 발달로 인해 새로운 교육체제로 등장하고 있는 웹 기반 교육 프로그램은 학습내용과 관련된 다양한 견해와 최신의 정보를 수시로 참조할 수 있다는 점, 텍스트를 비롯하여 이미지, 그래픽, 오디오와 비디오 등 풍부한 멀티미디어 데이터베이스에 기초한 학습이 가능하다는 점, 동시적·비동시적 상호작용이 가능하다는 점, 그리고 원격지 학습자들과의 활발한 협력학습을 통한 문제해결 활동을 가능하게 한다는 점에서 기존의 교육 환경을 보다 역동적이고 학습자 중심적인 환경으로 변모시킬 가능성성이 매우 높은 것으로 인식되고 있다.

2.4 웹 기반 교육에서의 학습현황 분석의 필요성

현재 개발되어 운영중인 웹 기반 교육 시스템을 살펴보면 모든 학습자에게 동일한 학습 내용을 제공하고 토론이나 질문을 할 수 있는 게시판 등을 제공하고 있다. 평가에서도 단지 객관식 문제은행식의 평가가 대부분이며 출석률과 진도율을 확인하여 점수화하고 있기는 하지만 정상적이지 못한 방법에 대해 제재할 수 있는 방법이 사실상 없다.

학습현황 분석을 하는 목적으로는 각각의 학습자에게 각기 다른 내용의 학습 내용을 제공할 수 있을 뿐만 아니라 학습자의 수준에 맞춰 진도나 평가의 내용을 달리하여 학습효과를 높이는 것이며, 또한 학습자가 원하는 내용을 보다 빨리 선택하여 열람할 수 있도록 도와주어 학습시에 불필요한 시간을 줄이는 것이다.

웹 기반 교육의 중심은 학습자이다. 원격교육 시스템의 여러 장점 중 하나가 학습자들의 능동적인 참여를 유도할 수 있다는 것이다. 이는 전통적인 면대면 교육 장면에서 학습자가 겪을 수 있는 교수자에 대한 심리적인 부담이 없기 때문이다. 따라서, 교수자에 대한 심리적인 부담으로 인하여 면대면 교육에서 소극적이고 수동적인 학

습태도를 가지는 학습자도 웹 기반 교육에서는 능동적으로 학습에 참여하고, 자기 의견을 제시하며 다른 학습자들과의 활발한 토론을 할 수 있게 된다.

그러나 이러한 웹 기반 교육의 장점은 교수자 혹은 교재 개발자 측면에서 보면 단점이 될 수 있다. 왜냐하면, 면대면 교육에서 가질 수 있는 교수자 측면에서의 장점 즉, 학습자의 반응을 살피고, 이에 대한 즉각적인 재반응을 하거나 혹은 학습자들의 취약점을 파악하고 이를 보완할 수 있다는 장점을 잊게 된다. 하지만, 웹 기반 교육 시스템의 학습 현황 분석 결과를 이용하게 된다면 학습자에 대한 추적 및 제어 기능을 제공함으로써 상당부분 해소할 수 있을 것이다.

이처럼 웹 기반 교육의 많은 장점을 살리기 위해서 꼭 필요한 전처리 단계인 학습현황의 분석은 필수적 일 것이다. 본 논문에서는 교수자나 학습자에게 도움을 주는 기반이 될 수 있도록, 웹 로그 마이닝을 단계별로 적용하여 학습현황을 효과적으로 분석하였다.

3. 학습현황 분석을 위한 웹 로그 마이닝

본 장에서는 학습자 개인별 학습현황을 분석하기 위한 항목들을 설정하고, 이를 분석하기 위해 웹 로그 마이닝을 이용하는 방법에 대해 알아본다.

3.1 학습현황 분석을 위한 항목 설정

일정기간 동안 학습자의 학습현황을 분석하기 위해서는 먼저 분석하고자 하는 항목들을 설정해야 한다. 본 절에서는 학습현황을 말해주는 4가지 항목을 알아본다. 이 항목은 기존의 “학습자의 가상수업 참여에 영향을 미치는 요인 연구”를 참고하여 설정한 것이다[2].

(1) 학습 진도 확인

학습자가 웹브라우저를 사용하여 시스템에 접속한 회수와 각 접속이 일어난 시간을 로그 데이

터로부터 얻어내어 학습이 진행된 순서와 각 접속 시각간에 소요된 시간 등 학습자 관련 데이터를 유도할 수 있다. 학습자가 웹 기반 교육 시스템에 접속한 후에 학습자가 한 행동 예를 들면, 네비게이션 경로와 같은 학습자 관련 기록이 로그 파일에 남게 된다. 학습자가 웹 기반 원격교육 사이트를 이동하면서 남긴 로그를 분석하여 의미있는 정보를 얻어낼 수 있다. 학습자로부터 얻어낼 수 있는 정보는 개별 학습자에 해당하는 부분과 전체 학습자 집단에 해당되는 부분으로 구분할 수 있다. 본 논문에서는 개별 학습자에 관한 학습현황 분석이 목적이므로 이에 초점을 맞춘다. 따라서, 학습자가 일정 기간 학습한 내용을 단원별, 영역별로 분석하여 화면에 보여준다. 일정 기간 총히트수, 해당 연/월/주별 히트수, 마지막 일주일 동안 히트수, 단원별, 영역별 학습시간을 분석한다.

(2) 학습 패턴 확인

웹 기반 원격교육의 경우 CD-ROM이나 다른 컴퓨터 기반 교육과 다른 점은 학습 내용이 순차적인 구조에 제한을 받지 않는다는 점이다. 물론, 기존의 CD-ROM을 통한 학습이나 기타 다른 방법을 통한 컴퓨터 기반 학습에서도 교재를 구성하는 방법에 따라 순차성이라는 제한에서 벗어날 수 있지만, 웹을 통한 교육의 경우 웹의 특성상 이러한 비순차적 구조가 자연스럽게 내재되어 있다. 웹 기반 교육에서 학습자가 학습을 진행하는 단위는 한 페이지의 HTML 문서가 된다. 따라서 페이지 분석을 통해 학습의 패턴을 알 수 있다[6].

- 직선형

직선형은 한 권의 책을 처음부터 읽어나가듯이 학습하는 것을 말한다. 1단원부터 순차적으로 학습하는 것이다. 즉, 일정한 순서에 의해 학습하는 경우를 말한다.

- 위계형

위계형 구조는 특정한 위계에 따라 학습 내용을 상위와 하위 단위의 관계로 학습해 나가는 것을 말한다. 이런 경우 학습자는 쉽게 전체 구조를 파악할 수 있다는 장점이 있지만, 한편으로는 필요

요한 정보만을 학습하는데 어려움을 겪을 수도 있다.

- 네트워크형(웹형)

네트워크형은 학습한 내용이 거미줄 형태로 구성되어 있는 것을 말한다. 학습자들은 학습을 진행하면서 학습 내용을 자유롭고 폭넓게 학습할 수 있다는 장점이 있다. 그러나 학습자가 학습 과정 전체적인 구조를 이해할 수 없게 되어 결과적으로 학습의 효율이 떨어질 수도 있다.

(3) 학습 참여도 확인

웹 기반 교육에 있어서 학습자는 장시간 혼자 학습해야 하는 부담이 있는 만큼 교수자가 적절한 자극을 꾸준히 주어 학습자로 하여금 성취에 대한 만족감을 고조시켜 혼자하는 학습에 대한 불안함과 실망감을 제거시켜 주어야 한다.

상호작용은 학습자-학습자, 교수자-학습자, 학습자-컨텐츠 간의 상호작용으로 구분할 수 있다. 학습자-학습자 상호작용은 동료들간의 토론, 그룹활동, 협동학습 등을 말하고, 교수자-학습자 상호작용은 질문 및 피드백을, 학습자-컨텐츠 상호작용은 조사, 자료수집, 과제의 선택과 학습속도의 조절이 있다. 이러한 상호작용을 통해 학습자의 동기가 유발되고 흥미가 유지될 수 있는 것이다. 더군다나 웹 기반 교육에서의 상호작용은 시간과 공간의 관념을 초월하여 가능하기 때문에 이러한 장점을 충분히 살린다면 학습자의 흥미와 동기를 충분히 살릴 수 있다.

본 논문에서는 동료 학습자와의 상호작용을 가능하게 하는 수단으로 의견 교환 게시판과 웹 채팅, 이메일을, 교수자-학습자 상호작용에서는 질문 게시판을, 학습자-컨텐츠 상호작용에서는 공지사항, 자료실을 이용하도록 한 다음, 그 접속시간을 분석한다. <게시판>은 과목에 대한 것은 물론 다른 분야에 대해서도 자유롭게 의견을 말할 수 있게 마련한 공간이다. 이 공간을 적절히 활용하면 동료 학습자간, 교수자-학습자간에 대화가 가능하며 면대면 수업에서와는 다른 친밀감을 갖게 된다. <웹 채팅>은 수업과 관계된 주제를 정해놓고 학습자들과 교수자가 가상공간에서 자유롭게 토론할 수 있도록 마련한 공간이다. 서

로 사전에 시간을 정한 뒤, 정해진 시간에 맞추어 공간의 제약없이 접속하여 활용한다. <공지사항>은 일종의 수업 게시판이다. 이 곳은 교수자만이 사용할 수 있으며 수업에 필요한 각종 공지사항을 이 곳에 올려놓는다. 학습자들은 항상 이 곳을 방문하여 수업에 필요한 자료 및 과제 등이 게시되어 있는지를 확인해야 한다. 학습자들은 게시물을 확인하고 교수자가 지시하는 바에 따라 자료실에서 자료를 다운로드 받거나 올리도록 한다[5].

(4) 학습 환경 확인

학습자가 학습 진행을 위해 사용하는 시스템 환경에 대한 분석이다. Onsite/Offsite 접속회수, 접속시 사용한 운영체제 및 브라우저 정보를 얻어낸다. Onsite/Offsite 접속회수는 먼저 로컬 IP 주소를 정의해두고 이를 기준으로 On/Off로 구분 한다. 또한 학습자가 학습시 이용한 시스템의 운영체제(Windows, Macintosh, Linux, Unix, Other) 및 브라우저(MS Explorer, Netscape, Lynx, Spiders, Other)가 무엇이었으며 이에 따른 히트수를 분석한다.

3.2 웹 로그 마이닝 단계

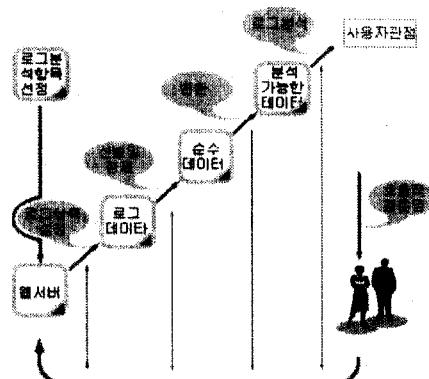
학습현황 분석을 위해 웹 로그 마이닝을 이용하는데, 웹 로그 데이터의 구성 및 전처리 과정을 알아본다.

(1) 웹 로그 데이터의 이해

웹서버는 웹서비스에 대한 요청과 제공에 대한 기록을 모두 로그파일이라는 파일에 저장한다. 웹서버에는 액세스(Access), 에러(Error), 참조(Referer), 에이전트(Agent) 등 4개의 로그파일이 웹서버의 일거수일투족을 기록하고 있다. 본 논문에서는 학습자의 학습 현황에 관심이 있으므로 일반적인 사항을 모두 기록하는 액세스 로그파일에 초점을 둔다. 로그파일의 형식은 NCSA 계열의 웹서버에서 사용하는 파일형식인 CLF(Common Logfile Format)를 따른다.

(2) 웹 로그 데이터의 전처리과정

- 분석을 위한 로그 항목 선정 및 웹서버 설정



(그림 1) 로그마이닝 단계

아파치는 로그파일을 기본적으로 CLF(Common Log Format) 형식으로 저장한다. 필요에 따라 LogFormat directive를 이용하여 다른 형식으로 저장할 수 있다. CLF형식은 스페이스로 구분되며 다음과 같은 형식으로 기록된다.

만약 값은 가지지 않는 항목들은 '-'로 기록된다.

본 논문에서는 학습자별 학습 현황을 파악하기 위해 사용자 IP Address, 인증된 사용자 아이디, 접속시간, 요청한 문서 및 프로토콜, 참조자, 사용자 브라우저 정보가 필요하다. 분석할 항목들을 정했다면 분석에 필요한 항목들이 로그에 남겨 지도록 해야 한다. 아파치는 다양한 로그 파일의 포맷을 지원하는데, 본 논문은 위의 로그 항목들을 모두 포함하는 combined 포맷으로 서버를 설정한다.

```
LogFormat "%h %u %a %W%q %s %b %{Referer}i %{User-Agent}i" combined  
LogFormat "%h %u %a %W%q %s %b" common  
LogFormat "%{Referer}i - %U" referer  
LogFormat "%{User-Agent}i" agent
```

(그림 2) 아파치 서버의 로그 포맷 형식

CustomLog logs/access.log combined

(그림 3) 설정된 로그포맷

- 로그 데이터의 정제

웹 로그 마이닝에서 가장 염두에 두어야 할 것은 “Garbage In Garbage Out”(쓰레기가 들어가면 쓰레기가 나온다)이다. 데이터 정제는 바로 로그 데이터에 무수히 널려 있는 Garbage(쓰레기 정보)들을 걸러 내는 과정이다. 로그 데이터에는 과잉 또는 분석 작업에 불필요한 항목을 포함하고 있다. 예를 들어, 대부분의 이미지 파일 항목은 관련이 없거나 필요 이상이다. 이미지 파일도 웹 페이지처럼 독립적으로 기록된다. 이는 웹 파일이나 동영상 파일도 마찬가지다. 불필요한 항목이라는 것은 사이트의 성격과 활용도에 따라 달라질 수 있으므로 신중하게 설정해야 한다. 로그 분석에 가장 시간이 많이 걸리고 또 중요한 과정이 바로 이 과정이다.

본 논문에서는 학습 현황 파악이 목적이므로 각 학습자가 요청한 이미지 파일에 관한 로그 엔트리(Log Entry)는 제거할 필요가 있다. 또한, loop-back interface를 통한 요청과 robots.txt 파일에 대한 요청 역시 의미가 없다. 학습 현황 파악을 위해 각 학생들은 회원 가입을 한 후 인증을 통해 원하는 학습을 시작할 수 있도록 전제하고 있기 때문에 Authuser 항목이 '-'인 로그 데이터도 분석에 필요하지 않다. 그렇다면 위의 로그 엔트리들을 모두 제거하도록 다음 그림4와 같이 서버를 설정해야 한다. 사용자 요청의 특성에 기초해 로그 파일로부터 특정 로그 엔트리를 배제하고자 할 때 Conditional Logging을 이용한다. 이 경우, 환경변수의 도움을 받는다. 먼저, 사용자 요청이 특정 조건을 만족시키도록 환경변수(SetEnvIf)를 set시킨다. 그 다음, 환경 변수가 set된 요청을 포함하거나 배제하도록 CustomLog directive의 env= 절을 사용한다.

- 로그 데이터의 변환

선별과정을 통해 얻은 순수 데이터를 분석 가능한 데이터로 변환시켜줄 필요가 있다. 이는 순수로그파일의 각 로그 엔트리를 항목별로 분리시켜 데이터베이스에 저장해야 함을 의미한다. 먼저 로그파일을 항목별로 분리해야 한다.

일단, 로그 파일을 열고 각 로그 엔트리를 파

```
# Mark requests for image file
SetEnvIf Request_URI \.gif\$ exceptlist
SetEnvIf Request_URI \.png\$ exceptlist
SetEnvIf Request_URI \.jpg\$ exceptlist

SetEnvIf Request_URI ^/logs/ exceptlist

# Mark requests from the loop-back interface
SetEnvIf Remote_Addr "127\.\d{1,3}.\d{1,3}.\d{1,3}.\d{1,3}" exceptlist

# Mark requests for the robots.txt file
SetEnvIf Request_URI "/robots\.\w{3}\$" exceptlist

# Mark requests from unauthenticated user
SetEnvIf Remote_User "" exceptlist

|CustomLog logs/access.log combined env=!exceptlist
```

(그림 4) 아파치 서버 설정

싱한다. 각 로그 엔트리의 처음과 끝 공백은 제거한 다음, split() 함수를 이용하여 각 로그 엔트리를 공백을 기준으로 분리하여 배열 변수 \$field에 저장한다. 각 field를 알기 쉬운 이름으로 assign한다. 이 때 time string은 다시 ':'을 기준으로 한번 더 분리한다. 미국 시간을 한국 시간으로 바꾸기 위해 다음과 같은 연산을 한다.

$\$UKhr=(\$tmp[0]+8)\%24 :$

line의 각 필드를 위한 배열을 생성하면 split() 함수의 역할은 종료된다.

다음으로, 필드별로 분리한 로그데이터를 데이터베이스에 저장한다. 데이터베이스에 connect한 후, 해당 데이터베이스를 선택한다. foreach()문을 이용하여 각 배열의 필드들을 데이터베이스에 질의어(INSERT INTO ~문)로 삽입한다. 이를 기초로 로그 마이닝이 시작된다.

3.3 세션 데이터 및 학습자 정보와의 연계

본 절에서는 세션 데이터를 통한 인증 및 ID매칭을 통한 웹 로그 마이닝에 대해 알아본다.

(1) 쿠키 인증

서버와 클라이언트가 웹에서는 계속적으로 연결되어 있는 것이 아니고 페이지에 대한 요청과 전송이라는 하나의 사이클을 단위로 연결이 끊어진다. 매 페이지마다 사용자 정보의 지속적인 유지가 필요할 경우 쿠키가 이용된다. 쿠키 인증은 HTTP 인증과 같은 절차를 가지며 다만, 지원해

주는 환경변수가 아닌 인증에 사용할 쿠키를 선택해서 사용하면 된다. 예를 들어 사용자 이름을 user라고 하고 패스워드를 passwd라는 쿠키를 사용한다면 인증을 걸고 싶은 프로그램들 앞부분에 쿠키값 user와 passwd를 검사하여 없거나 틀리다면 인증을 받을 화면으로 넘어가도록 하는 부분을 삽입하면 된다.

쿠키를 이용해서 인증을 하는 장점으로는 별도의 인증창이 뜨는 것이 아니라 HTML form으로 ID와 패스워드를 입력받으므로 사용자의 거부감이 적고 홈페이지의 메인 화면에 디자인해서 사용할 수 있다는 것이다. 그러나 쿠키는 인터넷에서 암호화되지 않은 데이터의 송수신에선 그 중간에서 데이터를 가로채 해석할 수 있다. 그러므로 중요한 아이디나 패스워드를 클라이언트에 저장하고 페이지를 이동할 때마다 서버와 주고 받는다면 중간에서 정보를 가로챌 확률은 증가한다.

(2) 세션 인증

세션은 특정사용자가 홈페이지에 접속한 후 연결이 끊어질 때까지 지속적으로 그 사용자를 인식할 수 있도록 추적하는 것이다. 또한 쿠키와 같은 성격과 기능을 가지는데, 쿠키와 구별되는 중요한 차이점은 쿠키가 사용자의 컴퓨터에 저장된다면 세션은 서버 쪽에 세션 데이터를 둔다는 점이다. 이는 보안적인 측면에서 상당한 안정성을 높일 수 있다. 세션은 클라이언트와 서버가 통신할 때 서버에서 세션을 생성하면 클라이언트에서는 이 세션의 ID만을 쿠키 등으로 저장한다. 이 세션 ID는 서버에서 접속한 각 클라이언트마다 유일하게 생성하여 중복되는 일이 없다. 이렇게 세션을 생성한 후 정보를 기록할 때는 클라이언트에서 서버로 한 번만 전송되고 서버에 저장된다. 그러므로 세션을 중요하기 전까지는 더 이상 서버로 아이디나 패스워드를 전송할 필요가 없다. 단지 세션 ID만 전송하면 되고, 서버에선 세션 ID에 해당되는 세션 정보를 가져오면 된다. 따라서 쿠키보다 조금 더 안전하다.

(3) IP매칭 .versus. ID매칭

로그 데이터에 저장되는 클라이언트 IP를 살펴보자. 클라이언트 IP는 사용자가 인터넷을 시작한 위치 정보에 해당된다. 클라이언트 IP정보를 통해 웹사이트에 접속하는 사용자들의 정보를 파악할 수 있다. 하지만 사용자가 유동 IP를 사용할 경우, 또는 프록시 서버(Proxy Server)를 이용할 경우 등에서는 정확한 클라이언트 IP를 파악할 수 없다. 기본적으로 많은 분석 툴은 방문자를 IP를 가지고 처리한다. 한 번 방문자는 해당 일에 IP가 한 번 나타났다는 의미이고 두 번 방문은 두 번 나타났다는 얘기다. 여기서 IP가 한 ROW 즉 한 페이지뷰만 나타났다는 얘기가 아니다. 세션 단위로 표현 할 때 한 세션에 대해서 나타났다는 의미이고 여러 번 방문했다는 의미는 여러 세션에 나타났다는 의미이다. 그래서 본 논문에서는 IP매칭 대신 ID매칭을 고려했다. 전학년의 학생들은 회원가입 절차를 거쳐 얻은 ID 및 Password를 가지고 로그인을 한 후 원하는 사이트에서 학습이 가능함을 전제로 한다. 회원가입 과정에서 학년, 반, 번호, 이름, id, password, 관심있는 컴퓨터 분야에 관한 정보를 입력하게 되고, 이 정보를 바탕으로 학습자 데이터베이스를 구축한다. 이 학습자 정보는 로그 마이닝 과정에서 웹로그 데이터 및 세션 데이터와 연계되어 원하는 결과를 얻을 수 있도록 한다. 로그인 과정을 거친 후 사용자의 행동을 기록한 로그 엔트리는 다음과 같다.

```
211.37.134.119 - saehee [12/Mar/2002:10:37:02
+0900] "GET
/cgi-bin/way-board/way-board.cgi?db=flea03
HTTP/1.1" 200 1050
"http://211.37.134.119/login.html" "Mozilla/4.0
(compatible; MSIE 5.5; Windows 98; Win 9x
4.90)"
```

(그림 5) 로그인 과정 후 로그 엔트리의 구성

로그 엔트리의 Authuser field에 사용자의 로그인 아이디가 기록되어 있음을 확인할 수 있다. 이 아이디를 기본으로 학습자를 구분하게 된다.

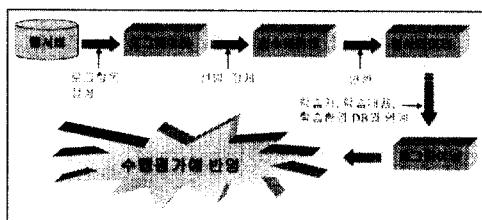
4. 학습현황 분석 시스템 설계 및 구현

본 장에서는 학습현황 분석을 위한 웹 로그 마이닝을 구현하고 그 결과를 제시한다. 본 논문의 프로그램 구현 환경<표 1> 및 전체적 시스템 구성<표 2>은 다음과 같다.

< 표1 > 시스템 구현 환경

H/W	CPU	PentiumIII
	HDD	20GB
	RAM	128MB
S/W	운영체제	Windows xx
	웹서버	Apache Web Server
	개발도구 및 언어	Photoshop6.0, Dreamweaver, MySQL6.0, PHP4

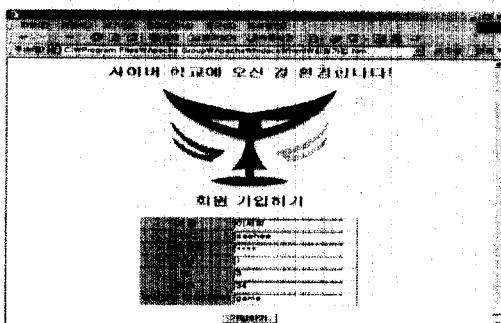
< 표2 > 시스템 구성도



4.1 학습자 인증 과정

(1) 학습자 로그인

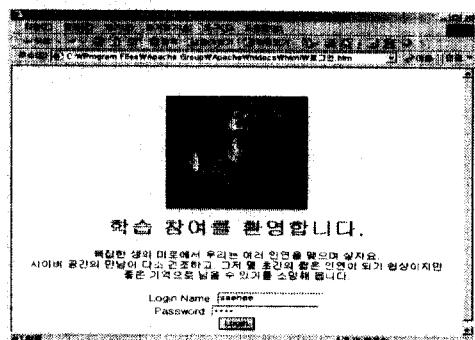
학습자는 우선 회원가입을 거친다. 회원 가입 시 필요한 정보를 입력하면 학습자 데이터베이스(student)에 필드별로 저장이 된다. 그림 6과 그림7은 이 과정을 보여준다.



(그림 6) 학습자 회원 가입 화면

(그림 7) 학습자 데이터베이스에 회원 정보 저장

그런 다음, 그림 8과 같이 로그인을 하게 된다.



(그림 8) 학습자 로그인 화면

(2) 학습자 인증

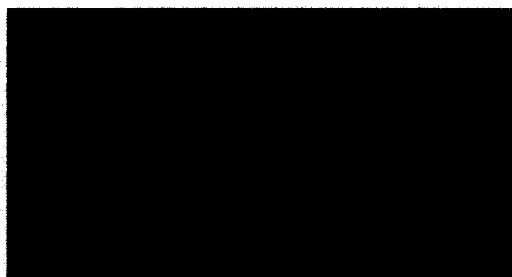
학습자가 로그인에 성공하면 인증 과정을 거친다. 로그인 폼으로부터 온 정보가 set되어 있으면 학습자 데이터베이스의 userid, password와 로그인 정보를 비교한다. 우선, 학습자 데이터베이스에 연결하고 로그인 정보와 일치하는 엔트리가 있는지 확인한다. 일치하는 엔트리가 있으면 세션을 시작하고, 세션 변수(\$mysession)에 저장한 후 변수를 등록한다. 인증에 성공한 것이다. 만일 일치하는 엔트리가 없다면 인증에 실패한 것으로 접근이 불가능하다.

로그인 정보가 없다면 세션 변수를 통해서 인증이 이루어진다. 위와 마찬가지로 학습자 데이터베이스에 연결하여 세션변수 정보와 일치하는 엔트리가 있는지 확인하여 인증 성공/실패 여부를 결정한다. 세션을 통한 인증은 세션 데이터가 서버에 저장되어 있으므로(세션 데이터베이스) 클라이언트와 서버가 통신할 때 서버에서 세션을 생성하면 클라이언트에서는 이 세션의 ID만을 쿠키로 저장한다. 이렇게 세션을 생성한 후 정보를

기록할 때는 클라이언트에서 서버로 한 번만 전송되고 서버에 저장된다. 그러므로 세션을 종료하기 전까지는 더 이상 서버로 아이디나 패스워드를 전송할 필요가 없다. 단지 세션 ID만 전송하면 되고, 서버에선 세션 ID에 해당되는 세션 정보를 가져오면 된다.

4.2 데이터베이스 설계 및 구축

위의 과정에서 순수한 로그 엔트리들을 분석이 가능하도록 변환하기 위해서는 데이터베이스를 이용해야 한다. 각 로그엔트리들을 field별로 분리한 후, 해당 데이터베이스에 저장하여 원하는 분석 작업을 시작한다. 그러기 위해서는 데이터베이스 설계가 선행되어야 한다. 먼저, 데이터베이스 구축 후에 테이블(TABLE)을 생성한다. 학습자 DB는 회원가입 정보인 학습자 아이디, 패스워드, 학년, 반, 번호, 이름, 관심있는 컴퓨터 분야로 구성된다. 학습내용 DB는 학습단원 및 영역의 과목명으로, 학습환경 DB는 아이디, 날짜, agent, referer 필드로 구성된다. 로그 DB는 사용자 아이디, 학습단원, 영역, 날짜, 학습시간 필드로 구성된다. 앞의 DB 중 실제 로그 DB 모습은 그림 9와 같다.



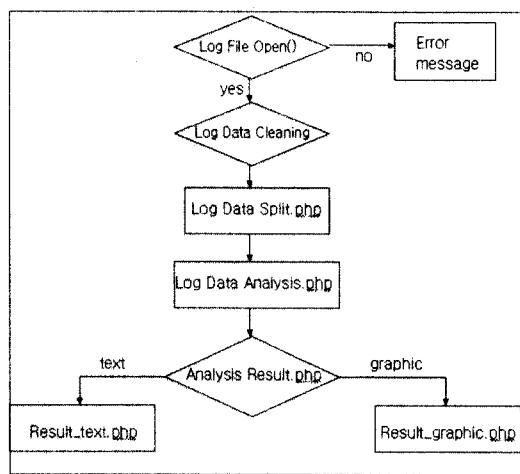
(그림 9) 로그 데이터베이스의 구성

4.3 웹 로그 마이닝 과정

4.3.1 로그 데이터베이스 생성과정

학습자가 회원 인증을 거쳐 원하는 학습을하게 되면, 서버는 이 과정을 로그 파일에 기록한다. 이 로그 데이터를 기초로 마이닝을 하기 위

해서는 로그 데이터를 필드별로 분리해야 한다. 엑세스 로그 파일을 split()함수를 이용하여 필드별로 분리한 다음 로그 데이터베이스에 저장하는 데이터 변환 작업을 수행한다.

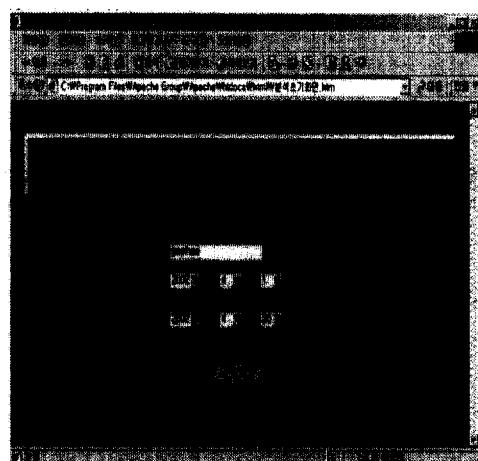


(그림 10) 로그 마이닝 흐름도

4.3.2 학습 현황 분석을 위한 로그 마이닝

(1) 분석 초기화면

학습 현황을 알고 싶은 학습자의 아이디와 학습 기간을 입력하면 분석이 시작된다. 분석 결과는 텍스트 및 그래픽 화면 중 선택할 수 있다.



(그림 11) 학습현황 파악을 위한 분석 초기화면

(2) 로그 마이닝 과정

학습 현황 분석을 위해 입력한 학습자 아이디 및 학습 기간을 만족하는 엔트리만을 선택하여 다시 학습자 개인의 데이블을 생성한다. 이 테이블에 적절한 질의어를 이용하여 학습 현황 분석 결과를 얻는다.

- 학습 진도 분석

먼저, 학습 기간 동안의 Total Distinct Hits(\$total)를 얻기 위해 "SELECT DISTINCT(ip) FROM shlee ORDER BY ip" 문을 이용한다. 또한 학습 기간의 차(\$avgday)를 구해 이 값으로 위의 \$total을 나누면 하루 평균 히트수를 알 수 있고, 이를 다시 24로 나누면 시간당 평균 히트수를 얻을 수 있다. Total Overall Hits(\$bigtotal) 역시 마찬가지로 계산할 수 있다.

분석 기간에 해당하는 년, 월, 주별 히트수를 얻기 위해 연도별, 월별, 주별로 그룹지어 (GROUP BY, ORDER BY 이용) 질의한다. 예를 들어 연도별 히트수를 원한다면 "SELECT COUNT(*) AS cnt, YEAR(received) AS year FROM shlee GROUP BY year ORDER BY year" 문을 이용하면 된다.

다음, 단원별·영역별 학습 시간을 분석하기 위해서 각 단원 및 영역에 해당되는 페이지에서의 학습자가 머문 시간을 이끌어낸다. 예를 들면, 1단원에 해당되는 페이지가 www.cyberedu.net/first_chap이고, 각 영역들은 1 단원 시작 페이지의 하위에 www.cyberedu.net/first_chap/com_history.html와 같은 형태가 된다. 즉, 로그 데이터베이스에서 요청한 파일 경로가 www.cyberedu.net/first_chap%인 엔트리들의 접속시간을 계산하면 된다.

- 학습 패턴 분석

학습자의 학습 패턴을 알기 위해서는 각 페이지간 이동 경로를 분석해야 한다. 학습자가 1단원에 해당되는 페이지들을 순차적으로 접속한 후 2단원으로 이동했다면 이 경우 직선형 패턴이라 볼 수 있다. 그렇지 않고 일정한 순서없이 학습자가 원하는 페이지로 임의로 이동했다면 네트워크형으로 학습한 것이다. 따라서, 학습 기간동안

로그 데이터베이스의 페이지 이동 경로를 분석한다.

- 학습 참여도 분석

학습자가 동료 학습자 및 교수자와의 상호작용 정도를 알기 위해서는 게시판, 채팅, 자료실 등을 이용한 시간을 알아야 한다. 이 역시 학습 진도 분석시 이용했던 방법을 적용한다. 해당 페이지 상에서 접속시간을 계산한다. 먼저 동료 학습자 간 상호작용 정도를 알기 위해 게시판, 채팅 페이지를 살펴본다. 로그 데이터베이스의 엔트리 중 게시판, 채팅 페이지에 접속한 엔트리들만 선택한 다음, 접속시간을 계산한다.

- 학습 환경 분석

Onsite vs Offsite Hits를 알기 위해서는 미리 로컬한 ip 주소를 정의해야 한다. 본 논문에서는 203.139.%를 로컬이라 정의한다. 따라서 ip 주소가 위와 다른 경우는 로컬 네트워크의 외부 주소로 간주한다.

운영체제별/브라우저별 히트수를 얻기 위해서는 원하는 운영체제 및 브라우저와 일치하는 엔트리를 구하는 질의어를 이용한다. 예를 들어 윈도우즈를 이용한 히트수를 구하는 경우 "SELECT count(browser) as cnt FROM shlee WHERE browser LIKE '%win%'" 문을 이용한다.

- 분석 결과

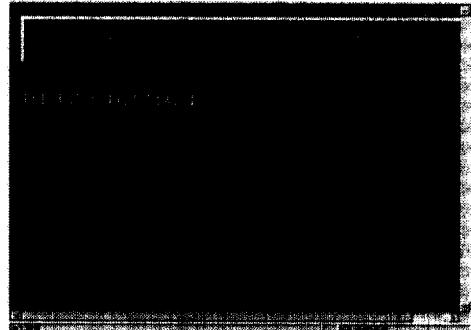
이 분석 결과는 개인별로 학습현황을 테스트하는 실험을 보여 주는 것이다. 여기서 보여지는 그림은 실험대상 중 임의의 한명의 결과를 상세하게 보여주고 있다. 분석 결과는 텍스트 및 그래픽 화면으로 확인할 수 있다. 이 분석결과를 통해 텍스트를 선택한 경우 (그림 12)~(그림 15)과 같은 화면을 볼 수 있으며, 시작적인 효과를 경험하고 싶다면 그래픽 화면을 선택할 수 있다. (그림 16)~(그림 19)은 분석 결과를 그래픽으로 나타낸 것이다. 이는 분석 과정을 통한 결과를 PHP4의 GD Library를 이용하여 그라프로 표현한 것이다. 결과로 이 학습자는 2단원에 대한 학습시간이 가장 길었으며, 직선형 학습 패턴이며, 웹 채팅을 이용한 학습 참여가 가장 많았으며, 브라우저는 Explorer를 대부분 이용하였다는 것

을 알 수 있었다.

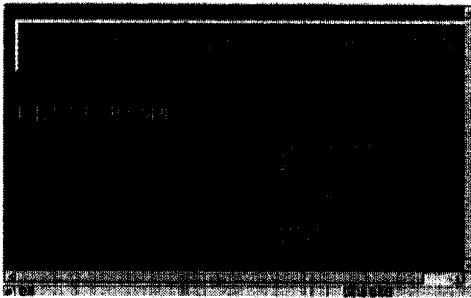
[텍스트 결과]



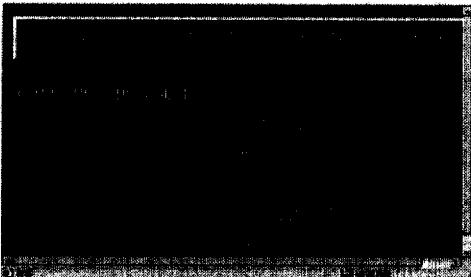
(그림 12) 텍스트 결과: 학습 진도 확인



(그림 13) 텍스트 결과: 학습 패턴 확인

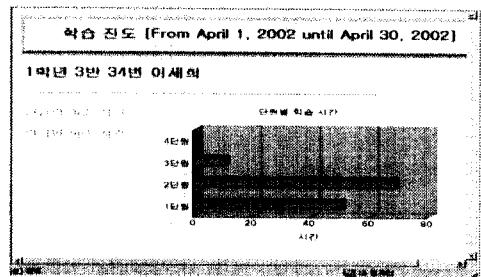


(그림14) 텍스트 결과: 학습 참여도 확인

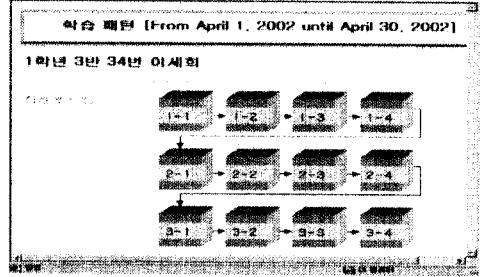


(그림 15) 텍스트 결과: 학습 환경 확인

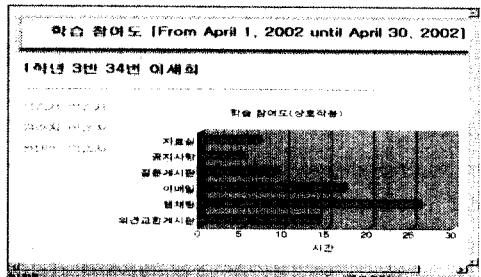
[그래픽 결과]



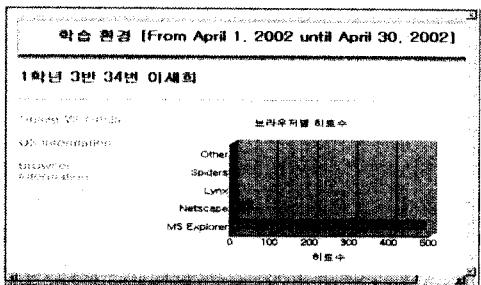
(그림 16) 그래픽 결과: 학습 진도



(그림 17) 그래픽 결과: 학습 패턴



(그림 18) 그래픽 결과: 학습 참여도



(그림 19) 그래픽 결과: 학습 환경

5. 결론 및 향후 과제

본 논문에서는 사이트를 방문한 학습자의 정보가 기록된 웹 로그 데이터를 기초로 개인별 학습현황 분석을 하였다. 기존의 로그 분석 툴이 지향한 IP매칭의 문제점을 지적하고 그 대안으로 ID매칭을 제안하였다. 학습자의 아이디를 토대로 개인별 통계 정보를 이끌어 내었다. 로그 분석 결과는 텍스트 및 그래픽 화면으로 제시했고, 이 결과에 따라 학습자의 일정 기간 학습량을 파악할 수 있다.

본 논문에서 구현한 시스템은 분석 초기에 제시한 학습기간 동안 학습자의 학습 진도, 학습 패턴, 학습 참여도, 학습 환경을 분석했다. 결과적으로 단원별·영역별 학습 시간, 학습 패턴구조, 상호작용 정도, 학습 시 이용한 시스템 정보 등을 얻을 수 있다.

본 연구의 의의는 차별화된 웹로그마이닝 기법을 이용하였다는 점이며, 부가적인 효과로 개인별 학습현황 분석을 통해 더 나은 웹 기반 교육으로 발전시킬 수 있다는 것이다. 구체적으로는 첫째, 로그에 ID가 기록되기 때문에 독립된 학습자의 구분이 명확해진다.

둘째, 수시로 학습현황 분석이 가능하므로 학습과정의 평가가 수월해진다. 학습자별 학습현황 정보는 수행 평가 데이터로 이용될 수 있다.

셋째, 면대면 평가가 아니기 때문에 교수자는 평가시 시간, 비용을 절감할 수 있고, 학습자는 평가에 대한 심적 부담을 최소화할 수 있다. 마지막으로 웹 로그 마이닝을 이용한 학습자의 학습현황 분석으로 학습자의 행동을 접속 통계 정보 이상으로 이해할 수 있다[4].

마지막으로 웹 로그 마이닝을 이용한 학습현황 분석 결과를 이용하여 학습자에게 더욱 친숙하게 페이지를 재구성하거나, 웹 서버 로드 밸런스, 학습자별 맞춤형 웹 페이지를 구성할 것과 평가에 반영하기 위해 더 세분화된 학습현황 분석 항목 선정을 향후 과제로 제시한다.

참 고 문 헌

- [1] 김영수, 강명희, 정재삼(1997). 21세기를 향한 교육공학의 이론과 실제. 교육과학사.
- [2] 김은옥(1998). 학습자의 가상수업 참여에 영향을 미치는 요인 연구. 서울대학교 대학원 석사학위논문.
- [3] 정인성(1998). 웹기반 가상수업의 교수 전략과 평가. 한국방송대학교 원격교육 심포지움 발표 자료집.
- [4] 김정숙, 조승근(1996). 방송대 원격영상강의 시스템의 도입 및 효율적인 운영방안에 관한 연구. 교육매체개발논총 제1집, pp. 31-65. 한국방송대학교 교육 매체 개발 연구소.
- [5] 정갑주, 박종선(1998). 효과적인 교수-학습을 위한 가상학습 지원시스템 분석. 정보과학회지 제 16권 10호.
- [6] 황상연(1999). 웹을 기반으로 한 학습자 진단 및 조언 시스템 구현. 한국 정보교육학회 제 4권 1호.
- [7] khan B.H.(1997). Web-based instruction.
- [8] 2000 Pacific-Asia Conf. on Knowledge Discovery and Data Mining (2000). *Mining Access Pattern efficiently from Web logs*, Proc. Kyoto Japan.
- [9] M. Carl Drott(1998). Using Web Server Logs to Improve Site Design.
- [10] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan(2000). *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. ACM SIGKDD.
- [11] <http://webmania.co.kr/reviews/benchmark>.
- [12] <http://www.itbiz.co.kr/news/enter/2000120103.html>.
- [13] <http://cs.sungshin.ac.kr/~jpark/HOME/SEARCH/sigdb2000.html>.
- [14] http://www.ecminer.com/m4_webDataMining.html.
- [15] <http://webmania.co.kr/reviews/benchmark>.
- [16] <http://maso.zdnet.co.kr/20010509/specplan/article.html>.



신지연

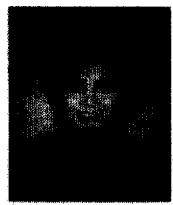
2000 고려대학교 전자계산학과
(이학사)
2002 이화여자대학교
교육대학원(교육학석사)

2003 ~ 현재 덕계중학교 교사

관심분야: 컴퓨터교육

E-Mail: shin1009@hanmail.net

정옥란



1993 전북대학교 전자계산학과
(이학사)
1998 전북대학교 대학원
정보과학과(이학석사)

1999~현재 이화여자대학교 컴퓨터학과

박사과정

관심분야: 웹 마이닝, WBI

E-Mail: orchung@ewha.ac.kr



조동석

1979 서울대학교 전기공학과
(공학사)
1981 서울대학교 대학원
전기공학과(공학석사)

1986 서울대학교 대학원
컴퓨터공학과(공학박사)

1985~현재 이화여자대학교
컴퓨터학과 교수

1996년 University of California, Irvine
Visiting Scholor

관심분야: 컴퓨터구조 및 인터넷공학,
컴퓨터비전, 컴퓨터그래픽스, 컴퓨터교육

E-Mail: dscho@ewha.ac.kr