

2-계층 클러스터링을 사용한 웹 사용자 그룹의 행동규칙추출방법에 관한 연구

황 준 원[†] · 송 두 헌^{††} · 이 창 훈^{†††}

요 약

유용한 웹 사용자 그룹을 파악하고 이들의 행동패턴을 찾는 것은 eCRM에서 매우 중요하다. 그러나 온라인 사용자 데이터에는 불확실한 정보가 많이 포함되어 있어 이를 바탕으로 유사한 성향을 가진 사용자 그룹을 생성하는 경우 신뢰성이 떨어지는 문제점이 있다. 본 논문에서는 불확실성이 포함된 사용자와 페이지의 서로 다른 두 데이터 계층의 상호작용을 통해 좀 더 신뢰성 있는 사용자 그룹을 생성하고 데이터에 내재된 이들의 행동패턴을 추출하는 방법을 제시하였다. 그리고 C4.5를 사용하여 생성된 행동규칙과의 비교를 통해 본 논문에서 제시하는 방법과의 비교분석을 실시하였다.

키워드 : 웹 마이닝, 클러스터링, eCRM, C4.5

A Study on Behavior Rule Induction Method of Web User Group using 2-tier Clustering

JunWon Hwang[†] · Doo Heon Song^{††} · ChangHoon Lee^{†††}

ABSTRACT

It is very important to identify useful web user group and induce their behavior pattern in eCRM domain. Inducing user group with a similar inclination, a reliability of user group decreases because there is an uncertainty in online user data. In this paper, we have applied the 2-tier clustering, which uses the outcome of interaction with data from other tiers. Also we propose a method which induces user behavior pattern from a cluster and compare C4.5 with our method.

Key Words : Web Mining, Clustering, eCRM, C4.5

1. 서 론

온라인에서 수집한 사용자 데이터를 분석하여 유사한 성질을 가진 사용자 그룹을 파악하고 그들의 행동특성을 분석하는 일은 eCRM 뿐만 아니라 온라인 사이트의 운영을 위해서도 중요한 일이다[1]. 일반적으로 사용자가 직접 입력한 온라인 등록 정보 및 기타 다른 경로를 통해 확보한 데이터에 대해 데이터마이닝 기법을 사용하여 사용자 그룹을 생성한다[2,3]. 그리고 웹 페이지를 대상으로 유사도가 높은 웹 페이지 그룹을 생성하고 웹 로그 분석을 통해 생성된 그룹이 온라인 상에서 어떤 행동특성을 갖는지 파악한다[4,5]. 그러나 사용자가 입력한 인구통계학적 정보들은 부정확하거나 입력되지 않은 경우가 많아 사용자 데이터에는 많은 불확실성이 내포되어 있다. 또한 웹 로그에 기록된 사용자의 페이지 방

문정보, 구매정보 등과 같은 행동정보량이 부족하여 이를 바탕으로 유사한 행동패턴을 갖는 사용자 그룹을 파악하는 것은 쉽지 않은 일이다. 이러한 문제점을 해결하기 위해 페이지 클러스터를 생성하고 웹 로그 분석을 통해 사용자와 페이지 클러스터의 연관정보를 구축한 후 사용자 데이터를 대상으로 클러스터링 할 때 연관정보를 함께 이용함으로써 사용자 그룹의 정확성을 높여주는 방식이 제시되었다[6].

행동규칙은 데이터에 내재된 사용자의 행동패턴을 규칙의 형태로 표현한 것으로 본 논문에서는 생성된 사용자 그룹에서 행동규칙을 추출하는 방법을 제시하였다. 또한 규칙 생성을 위해 많이 사용되는 C4.5[7]를 사용자 데이터에 적용하여 행동규칙을 생성하고 2-계층 클러스터링 방법과의 비교 분석을 실시하여 제안하는 방법의 타당성을 검증하였다. 그리고 서로 다른 데이터 계층에 여러 클러스터링 알고리즘을 적용하고 사용자 데이터의 불확실성의 정도와 페이지 계층에 대한 연관정보의 양에 대한 다양한 셋팅을 구성하였다. 이를 통해 여러 조건에서 2-계층 클러스터링과 단일계층 클

† 준 회 원 : 건국대학교 컴퓨터공학과 박사과정수료
†† 정 회 원 : 용인송담대학 컴퓨터게임정보과 교수
††† 종신회원 : 건국대학교 컴퓨터공학과 교수
논문접수 : 2007년 8월 31일, 심사완료 : 2007년 10월 29일

러스터링과의 품질분석을 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 2-계층 클러스터링의 개념과 이를 웹에 적용하는 방법을 설명하였고, 3장에서는 생성된 클러스터에서 행동규칙을 추출하고 평가하는 방법을 설명하였고, 4장에서는 단일계층 및 2-계층 클러스터링에서 생성된 클러스터의 품질분석을 실시하고 C4.5와 행동규칙을 비교하였으며 5장에서는 결론을 기술하였다.

2. 관련연구

2.1 클러스터링 알고리즘

클러스터링 알고리즘은 크게 분할적 클러스터링(Partitional Clustering)과 계층적 클러스터링(Hierarchical Clustering) 알고리즘으로 분류할 수 있다. 분할적 클러스터링은 클러스터의 계층을 고려하지 않고 평면적으로 클러스터링 하는 방법으로 일반적으로 미리 몇 개의 클러스터로 나뉘어 질 것 인지를 예상하고 클러스터의 개수를 정한다. K-means는 분할적 클러스터링 방법으로 전체 데이터 집합에서 임의로 k 개의 seed point를 선정하고 이를 클러스터의 중점으로 시작한다. 이후 데이터 개체에 대한 소속 클러스터의 재할당 과정과 목적함수의 평가를 반복적으로 수행하여 목적함수를 최적화 한다.

계층적 클러스터링은 가장 유사한 두 개체를 선택하여 병합해 가는 병합적 계층 클러스터링(Agglomerative Hierarchical Clustering) 방법과 가장 먼 개체를 선택하여 나누어 나가는 분할적 계층 클러스터링(Divisive Hierarchical Clustering) 방법이 있다. 두 클러스터의 유사도를 측정하는 방법에는 최단연결법, 최장연결법, 중심연결법, 평균연결법이 있다[8].

2.2 2-계층 클러스터링 개념

2-계층 클러스터링 기법은 한 계층의 클러스터링 결과를 다른 계층의 클러스터링에 이용하는 방법이다. 이는 클러스터링을 할 때 자신의 콘텐츠 정보 뿐만 아니라 다른 계층의 데이터에 대한 링크 정보를 함께 이용함으로써 콘텐츠의 부정확성에서 오는 오차를 줄일 수 있는 장점이 있다[9]. (그림 1)은 2-계층 클러스터링의 개념을 나타낸 것으로 다른 계층에 대한 링크정보를 이용하여 클러스터링을 수행하는 과정을 보여주고 있다.

상위계층과 하위 계층에 6개의 노드로 이루어진 2개의 데이터 집합이 있고 두 데이터 사이에는 링크 정보가 존재한다. 상위 계층의 노드 1의 하위 계층에 대한 링크정보는 [1, 0, 0, 0, 0, 0]이고 노드 5의 링크정보는 [0, 0, 1, 0, 1, 1] 이다(단계1). 두 계층 가운데 한 계층에서 콘텐츠 정보에 의해 클러스터링이 실시되는데 (그림 1)에서는 하위계층에서 먼저 실시되었다(단계 2). 하위 계층의 노드들이 클러스터화 되면서 상위 계층의 노드에 대한 링크정보는 클러스터에 대한 링크정보로 갱신된다(단계 3). 상위 계층의 노드 1의 링크정보는 [1, 0]로 노드 5의 링크정보는 [1, 1]로 갱신된다. 노드 5의 경우 1번 클러스터에 대한 링크정보값이 2번 클러스터에 대한 것보다 더 크다. 따라서 상위 계층의 클러스터링을 할 때 다른 계층에 대한 링크 정보를 이용한다면 노드 5는 1번 클러스터에 속하게 된다.

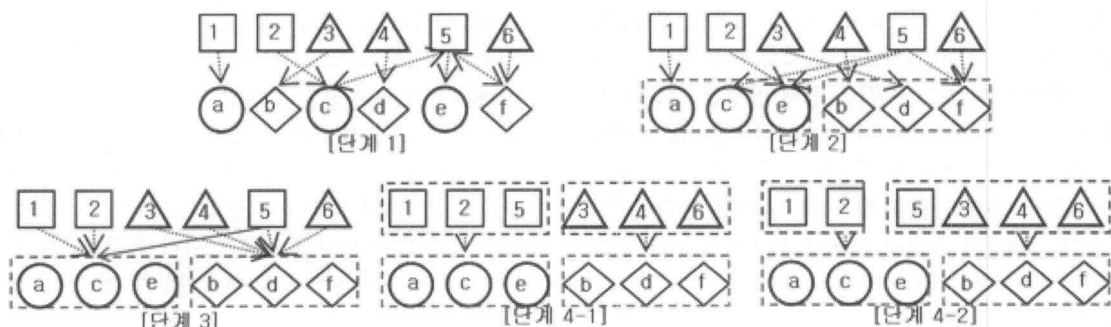
2-계층 클러스터링을 위해서는 두 계층을 서로 연관시켜 줄 수 있는 방법이 있어야 하는데 식 (1)은 2-계층 클러스터링의 거리공식이다.

$$D = \alpha D_c + (1-\alpha)D_l, 0 \leq \alpha \leq 1 \quad (1)$$

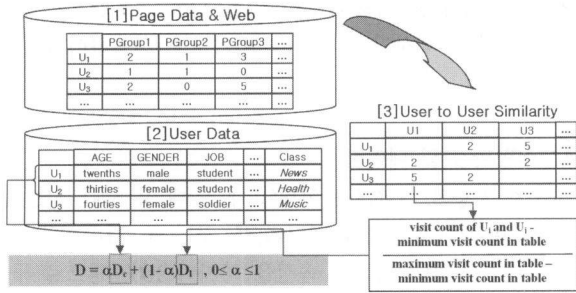
D_c 는 콘텐츠 클러스터링에서의 거리값이고 D_l 는 다른 계층과의 링크정보값이다. α 는 상수값으로 어느 계층에 더 가중치를 부여할지를 설정해준다. (단계2)에서의 하위 계층의 클러스터링의 경우 α 는 1이므로 콘텐츠 정보만을 사용하였다. (단계 4-1)에서의 클러스터링에서 α 는 0으로 링크정보만을 사용하였다. 만일 콘텐츠 정보가 링크정보보다 더 높은 비중으로 사용되었다면 노드 5가 1번 클러스터에 할당되지 않았을 수도 있고 이는 두 계층의 정보를 모두 이용했기 때문인데 어느 계층에 더욱 가중치를 두었느냐에 따라 달라질 수 있다(단계 4-2). 따라서 콘텐츠 정보와 링크정보를 모두 사용하는 $0 < \alpha < 1$ 인 경우만이 2-계층 클러스터링이고 $\alpha=0$ 혹은 $\alpha=1$ 인 경우는 단일 계층 클러스터링을 나타낸다.

2.3 2-계층 클러스터링의 웹 적용

2-계층 클러스터링은 서로 다른 두 데이터 계층이 있고 그 사이에는 두 계층을 연관시켜 주는 링크정보가 존재해야 적용이 가능하다. 웹 데이터는 크게 사용자 데이터와 페이지 데이터로 구분 할 수 있고 이 둘간의 링크정보는 웹 로그의 사용자의 페이지 방문정보를 분석함으로써 얻을 수 있



(그림 1) 2-계층 클러스터링의 개념



(그림 2) 웹에서의 2-계층 클러스터링

다[10]. (그림 2)는 2-계층 클러스터링 공식을 사용하여 사용자 1과 사용자 2간의 거리를 구하는 것을 보여주고 있다.

먼저 D_c 를 구하기 위해 사용자 1과 사용자 2간의 콘텐츠 정보에 대해 범주형(nominal) 데이터의 경우는 VDM[11] 기법을 사용한 후 유클리디안 거리를 적용한다. 그리고 페이지 데이터의 콘텐츠 정보를 이용하여 페이지 클러스터를 구하고 웹 로그를 분석하여 사용자와 페이지그룹간의 방문정보 테이블을 생성한다. 그리고 이것으로부터 사용자 간의 유사도 테이블을 생성하고 D_1 을 구한 후 D_c 와 더한다. 여기서 D_1 , D_c 은 똑같은 범위의 거리값을 갖는데 $0 \leq D_1, D_{cl} \leq 1$ 이다.

사용자 클러스터를 생성 하기 위한 2-계층 클러스터링 알고리즘은 다음과 같다.

- 단계 1 : 페이지 계층에서 페이지 콘텐츠를 이용하여 페이지 클러스터를 생성한다. 웹 로그에 있는 사용자의 페이지 방문정보와 생성된 페이지 클러스터를 이용해 <사용자 - 페이지 클러스터> 간의 유사도 테이블을 작성한다.
- 단계 2 : 사용자 계층에서 사용자 콘텐츠와 전단계에서 만들어진 <사용자 - 페이지 클러스터> 간의 유사도 테이블의 링크 정보를 이용하여 사용자 클러스터를 생성한다. 웹 로그에 있는 사용자의 페이지 방문정보와 생성된 사용자 클러스터를 이용하여 <사용자 클러스터 - 페이지> 간의 유사도 테이블을 작성 또는 업데이트 한다.
- 단계 3 : 페이지 계층의 콘텐츠와 <사용자 클러스터 - 페이지> 간의 유사도 테이블의 링크 정보 값을 이용하여 클러스터링 한다. 생성된 클러스터를 이용해 <사용자 - 페이지 클러스터> 간의 유사도 테이블을 업데이트 한다.
- 단계 4 : 생성된 전체 클러스터들의 밀도의 합이 전 단계 보다 낮으면 단계 2로 가고 높으면 종료한다.

3. 행동 규칙 추출

본 논문에서의 실험은 실제 데이터와 유사하게 가상데이터를 생성하여 실시되었다. 이를 위해 필요한 데이터 생성 방법과 생성된 클러스터로부터 행동규칙을 추출하고 평가하는 방법을 제시하였다.

3.1 행동규칙 설정방법

일반적으로 사용자들은 고유한 행동패턴을 갖는다. 이를 위해 사용자들의 특정한 행동패턴을 데이터 속에 내재시키기 위하여 데이터 생성 규칙을 적용하였는데 이것을 통해 각 사용자 데이터와 웹 로그의 내용이 결정된다. 데이터 생성 규칙은 내재규칙과 분포결정규칙으로 구성된다. 내재규칙은 데이터 내의 행동패턴을 규칙으로 정의한 것이고 분포결정규칙은 각 내재규칙이 어느 정도의 경향을 가지고 그 행동패턴들을 결정하는지를 정의한 것으로 생성규칙은 다음과 같은 내용을 갖는다.

$$(\text{속성, 속성 값, 분포도})^+ \rightarrow (\text{방문자 비율, 페이지 클래스, 방문률}) \quad (2)$$

예를 들어 “나이가 10대이고 학문이 취미인 학생들은 과학 관련 페이지를 자주 방문한다” 라는 행동패턴에 대한 생성규칙은 “(AGE 10대 20) \wedge (JOB 학생 50) \wedge (HOBBY 과학 50) \rightarrow (60 과학 40)” 로 표현 할 수 있다. 이 생성규칙의 내용은 전체 데이터 중에서 10대가 차지하는 비율을 20%로 하고 그 중에서 직업이 학생인 비율이 50%, 취미가 과학인 비율이 50%로 사용자데이터를 생성한다. 그리고 이러한 데이터에서 60%의 사용자에게 대해 사이트 내의 “과학” 관련 페이지 방문률을 40%로 하고 나머지 60%는 임의로 다른 페이지를 방문하도록 생성함을 의미한다. 그리고 생성규칙에서 지정되지 않는 내용은 모두 임의로 설정되었다.

3.2 행동 규칙 추출 및 평가방법

생성된 클러스터의 품질(Quality)을 평가하기 위해 정보 검색 분야에서 널리 사용되고 있는 재현률(Recall)과 정확도(Precision)를 사용하였다[12]. 전체 클러스터의 크기가 $|C|$ 이고 i 번째 클러스터의 개체 크기가 $|C_i|$ 이며 클러스터 내에서 이 행동패턴을 따르는 개체 수를 $|C_i|$, 전체에서의 수를 $|P_j|$ 라고 하면 재현률과 정확도는 다음과 같다.

$$Recall = \left(\sum_{i=1}^{|C|} |C_i| \right) / |P_j| \quad (3)$$

$$Precision = \sum_{i=1}^{|C|} \frac{|C_i|}{|C_i|} \quad (4)$$

사용자 계층의 데이터를 클러스터링하여 생성된 클러스터의 대표값은 그 사용자그룹의 대표값으로 이는 행동규칙에서 조건부의 내용을 의미한다. 클러스터 내의 각 속성의 최대 빈발값을 대표값으로 선정하면 개체가 n 개의 속성으로 이루어진 경우에 클러스터는 n 개의 값으로 표현된다. 이는 항상 n 개의 조건값을 갖는 규칙을 생성하기 때문에 필요 없는 조건이 제거된 최적화된 규칙을 표현할 수 없다. 따라서 최대 빈발값이 그 속성값들을 대표할 수 있는지의 여부를 판별하여 그 비율이 유의수준을 넘을 경우 조건으로 채택하고 그렇지 못한 경우는 기각함으로써 클러스터가 갖는 규칙의 유연성을 부여하였다. 본 논문에서는 추출된 최대빈발값이 95%의 신뢰성을 갖도록 유의수준의 설정을 5%로 하

었다. 그리고 생성된 각 클러스터가 갖는 개체의 크기 $|C_i|$ 에 따라 $|C_i| \leq 8$ 개인 경우는 이항분포, $8 < |C_i| < 30$ 인 경우는 카이제곱분포, $30 \leq |C_i|$ 인 경우는 표준정규분포를 각각 적용하였다.

추출된 행동규칙이 데이터에 내재규칙을 어느 정도 포함하고 있는지 분석하여 추출된 규칙의 유용성을 측정하였다. 속성 집합 $A = \{a_1, a_2, \dots, a_m\}$ 이고, 각 속성이 갖는 범주형 데이터 값의 크기집합 $S = \{s_1, s_2, \dots, s_m \mid 1 \leq m \leq |A|\}$ 일 때, i 번째 속성 값의 집합 $V_i = \{v_{i1}, v_{i2}, \dots, v_{ij} \mid 1 \leq i \leq |A|, 1 \leq j \leq s_j\}$ 이다. $v_i \in V$ 에 대해 데이터에 내재시킨 규칙 $r = "v_1 \wedge v_2 \wedge \dots \wedge v_m \rightarrow page\ class"$ 이고 찾은 규칙은 $r' = "v'_1 \wedge v'_2 \wedge \dots \wedge v'_m \rightarrow page\ class"$ 이다. $v_i \neq v'_i$ 인 속성의 수가 k 개 라면 추출규칙 r' 를 평가하는 공식은 다음과 같다.

- if $k = 0$: 추출된 모든 규칙 값이 내재규칙에 존재하는 경우

$$Eval(r') = n/m$$

- if $k \geq 1$: 추출된 규칙 값이 내재규칙과 다른 것이 존재하는 경우

$$Eval(r') = 1 - \frac{k}{|A|}$$

생성 가능한 모든 페이지 클래스 수가 N 이고 발견된 페이지 클래스 수는 N' 이다. 행동규칙의 목적패턴에서 같은 페이지 클래스를 갖는 클러스터의 집합을 $W = \{w_1, w_2, \dots, w_i \mid 1 \leq i \leq N\}$ 라 하면

$$Inst(w) = \sum_{i=1}^{|w_i|} |C_i|, \quad Inst(w') = \sum_{i=1}^{|w'_i|} |C'_i|$$

페이지 클래스 단위별 규칙 평가도 $Eval(R_i)$ 와 추출된 행동 패턴의 다양성을 반영하는 전체 규칙 평가도 $Eval(R)$ 은 다음과 같다.

$$Eval(R_i) = \sum_{i=1}^{i=N'} \frac{Inst(w'_i)}{Inst(w_i)} \times Eval(r'_i) \quad (5)$$

$$Eval(R) = Eval(R_i) \times \frac{N'}{N} \quad (6)$$

4. 실험

성질이 다른 두 계층의 데이터에 여러 클러스터링 알고리즘을 사용하는 메타모델을 구성하여 클러스터링 알고리즘의 조합이 어떤 성능을 내는지를 다양한 데이터 셋팅을 통해 살펴보았다. 그리고 생성된 클러스터의 품질을 분석하기 위해

정확도와 재현률을 적용하였고 클러스터에서 추출한 행동규칙의 품질을 분석하기 위해 규칙평가도를 사용하여 C4.5와 비교하였다.

4.1 실험환경 및 데이터

사용자 계층의 데이터는 연령(10대, 20대 등 5개), 성별(남, 여), 결혼여부(기혼, 미혼), 직업(학생, 군인 등 20개), 취미(컴퓨터, 건강 등 10개)의 5개 속성을 가지며 1000 명으로 구성되었다. 웹 문서 계층은 Open Directory Project[13]에 있는 카테고리로부터 10개를 선정하여 각각 100개의 페이지를 임의의 선택함으로써 10개 클래스와 1000개의 페이지로 구성되었다. 10개 페이지 클래스에 대해 10개의 행동규칙을 설정하였고 모든 사용자는 각자 20개의 페이지를 방문하도록 웹 로그 데이터가 생성되었다. 사용자 데이터의 경우 3.1 절에서 언급한 식 (2)에서 방문자 비율을 30%와 90%로 방문율을 40%와 80%의 2가지로 설정하였는데 이 수치값은 단순히 비율의 높고 낮은 정도를 설정한 것이다. 이러한 설정은 2-계층 클러스터링의 거리공식에서 D_c 와 D_i 에 영향을 주는 것으로 D_c 와 D_i 의 성능이 전체에 미치는 영향을 분석하기 위한 것이다. 클러스터링 알고리즘은 크게 분할적 클러스터링 알고리즘과 계층적 클러스터링 알고리즘으로 분류할 수 있는데 본 논문에서는 각 계열에서 많이 이용되는 K-means와 병합적 계층 알고리즘을 사용하여 각 계층에 적용하였다. 이를 통해 서로 다른 특성을 갖는 클러스터링 알고리즘의 조합이 전체 성능에 어떤 영향을 미치는지 살펴보았다.

4.2 클러스터의 품질 분석

생성된 클러스터의 품질을 분석하기 위해 일반적으로 널리 사용되는 정확도와 재현률을 측정하였는데 실험결과에서 K-means 알고리즘은 K로 병합적 계층 클러스터링 알고리즘은 H로 표기하였다. 모든 표들은 식 (1)에서 가중치 α 의 값을 0부터 1까지 증가시켜 10회 실험하여 구한 평균값을 나타낸 것이다. 2-계층 클러스터링 구간은 $0 < \alpha < 1$ 으로 표의 내용은 이 구간의 평균값이다.

정확도 분석에서 사용자 데이터 계층만을 이용한 단일계층 클러스터링($\alpha=1$)은 30.1%이지만 2-계층 클러스터링의 경우는 64.3%로 매우 좋은 결과를 보여주고 있다. 그리고 링크 정보만을 이용한 단일계층 클러스터링($\alpha=0$)의 경우도 58.7%의 높은 정확도를 보여주고 있다. 링크정보는 페이지 데이터 계층과 사용자의 웹 로그 분석을 통해 생성되는데 방문자비율과 페이지 방문률이 높을수록 좋은 결과를 보인다. 이는 같은 행동유형을 갖는 사용자 비율이 높을수록 2-계층 클러스터링에서 콘텐츠 클러스터링의 정확도가 높아지기 때문이다. 또한 페이지 방문률이 높을수록 페이지 방문기록에 대한 정보량이 증가함으로써 사용자 콘텐츠의 부정확성에서 오는 오차를 링크정보를 통해 줄일 수 있는 확률이 높아지기 때문이다. 재현률의 경우도 정확도와 유사한 패턴을 가져 $\alpha=0$ 과 $\alpha=1$ 일 때 46.4%와 71.1%이나 2-계층 클러스터링의 경우 75.5%의 좋은 결과를 보여주고 있다.

〈표 1〉 데이터 구성에 따른 정확도

알고리즘		방문자 비율 - 방문률	α=0	α=1	2-tier
Page	User				
H	K	30-40	0.34	0.14	0.34
H	K	30-80	0.56	0.14	0.62
H	K	90-40	0.48	0.48	0.55
H	K	90-80	0.75	0.45	0.80
H	H	30-40	0.34	0.18	0.35
H	H	30-80	0.77	0.15	0.85
H	H	90-40	0.41	0.44	0.54
H	H	90-80	0.82	0.46	0.89
K	K	30-40	0.34	0.13	0.38
K	K	30-80	0.62	0.13	0.65
K	K	90-40	0.54	0.44	0.59
K	K	90-80	0.80	0.45	0.85
K	H	30-40	0.37	0.18	0.44
K	H	30-80	0.85	0.15	0.9
K	H	90-40	0.52	0.44	0.6
K	H	90-80	0.88	0.46	0.93
평균			0.587	0.301	0.643

〈표 2〉 데이터 구성에 따른 재현률

알고리즘		방문자 비율 - 방문률	α=0	α=1	2-tier
Page	User				
H	K	30-40	0.41	0.42	0.46
H	K	30-80	0.74	0.35	0.78
H	K	90-40	0.61	0.64	0.67
H	K	90-80	0.89	0.63	0.92
H	H	30-40	0.43	0.27	0.47
H	H	30-80	0.87	0.25	0.92
H	H	90-40	0.51	0.61	0.65
H	H	90-80	0.90	0.62	0.94
K	K	30-40	0.53	0.37	0.53
K	K	30-80	0.80	0.31	0.81
K	K	90-40	0.72	0.59	0.75
K	K	90-80	0.90	0.61	0.94
K	H	30-40	0.54	0.27	0.57
K	H	30-80	0.93	0.25	0.96
K	H	90-40	0.66	0.61	0.73
K	H	90-80	0.94	0.62	0.98
평균			0.711	0.464	0.755

서로 다른 성질을 갖는 사용자 데이터와 페이지 데이터 계층에 K-means와 병합적 계층 클러스터링 기법을 적용하여 어느 경우에 좋은 성능을 보이는 지를 살펴보았다. 알고리즘에 관계없이 정확도와 재현률 모두 항상 2-계층 클러스터링이 좋은 결과를 보여주었다. 정확도의 경우 2-계층 클러스터링이 단일계층 클러스터링 α=0, α=1일 때보다 5.6% 35.3%가 향상되었고 재현률의 경우는 4.4%, 29.1%의 향상을 보여주었다. 특히 페이지 계층에 관계없이 사용자 계층에 병합적 계층 클러스터링 기법을 사용하였을 때 더 좋은 성능을 보여주었다.

〈표 3〉 알고리즘에 따른 정확도

알고리즘		α=0	α=1	2-tier
Page	User			
H	K	0.53	0.30	0.58
H	H	0.59	0.26	0.66
K	K	0.58	0.29	0.62
K	H	0.66	0.31	0.72
평균		0.587	0.290	0.643

4.3 클러스터의 규칙평가도 분석

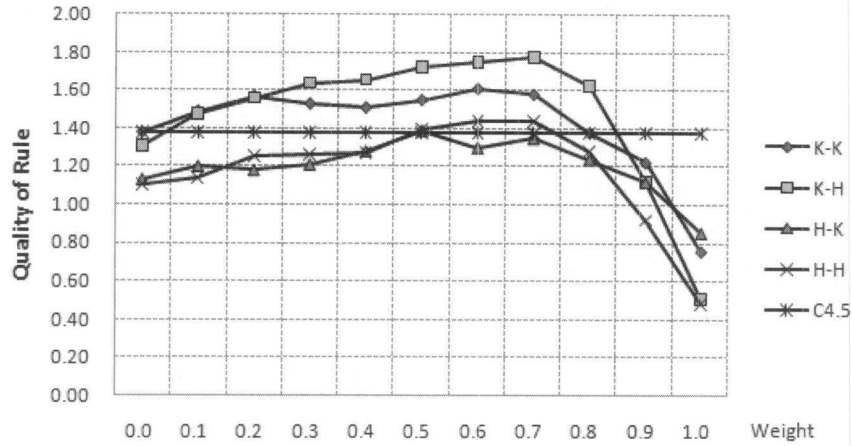
클러스터에서 추출된 규칙과 C4.5의 규칙에 대해 식 (6)의 규칙평가도를 적용하여 데이터에 내재된 행동패턴을 얼마나 잘 나타내었는지 분석하였는데 일반적으로 클러스터의 정확도와 재현률이 높은 2-계층 구간에서 규칙평가도의 값도 비례하여 높게 나타남을 알 수 있었다.

먼저 C4.5는 α(가중치)의 영향을 받지 않으므로 동일한 결과를 나타내지만 2-계층 클러스터링의 경우 2-계층 클러스터링 구간(0<α<1)에서 우수한 결과를 보이고 있다. 특히 페이지 계층의 알고리즘에는 관계없이 사용자 계층에 병합적 계층 클러스터링 알고리즘을 사용했을 때 우수한 결과를 나타내고 있다. 방문자 비율이 30%인 경우에는 비교적 2-계층 클러스터링이 C4.5 보다 우수하고 90%인 경우에는

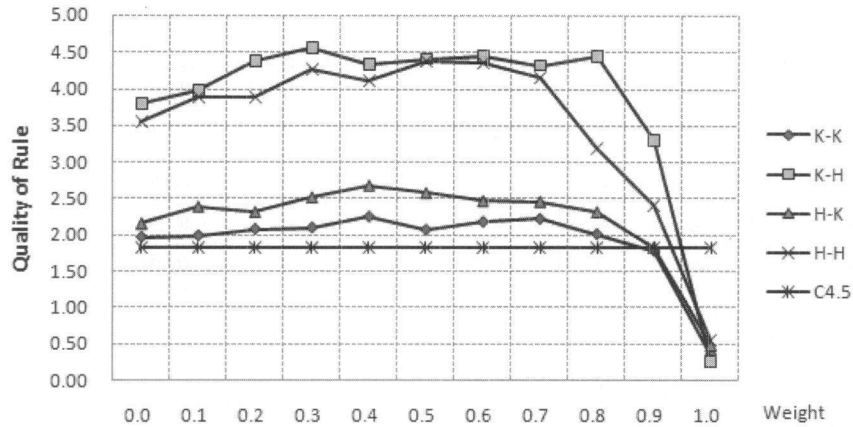
〈표 4〉 알고리즘에 따른 재현률

알고리즘		α=0	α=1	2-tier
Page	User			
H	K	0.66	0.51	0.71
H	H	0.68	0.44	0.75
K	K	0.74	0.47	0.76
K	H	0.77	0.44	0.81
평균		0.711	0.464	0.755

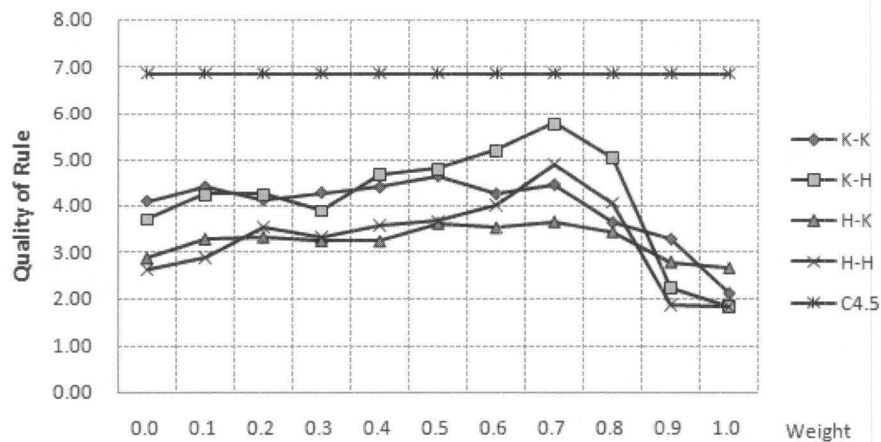
C4.5가 2-계층 클러스터링보다 좋은 결과를 나타내고 있다. 방문자 비율이 높다는 것은 전체 사용자 데이터에서 같은 행동패턴을 갖는 사용자들이 많다는 것을 나타내며 이들은



(그림 3) 규칙평가도 (30%, 40%)



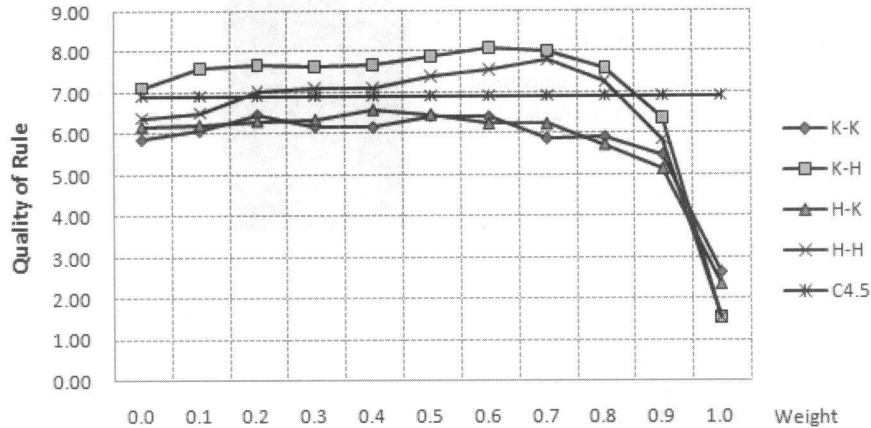
(그림 4) 규칙평가도 (30%, 80%)



(그림 5) 규칙 평가도 (90%, 40%)

서로 유사한 콘텐츠 정보를 갖는다. C4.5는 같은 행동패턴을 갖는 사용자들을 분류할 때 가장 영향을 미치는 속성 값을 선정하고 이것들은 곧 사용자 그룹이 갖는 행동규칙의 조건부의 내용을 구성하게 된다. 방문자 비율이 높으면 속성값을 선택할 때 오차를 줄이게 되어 정확도를 높일 수 있기 때문에 방문자 비율이 높을수록 월등히 좋은 성능을 보인

다. 그러나 방문률이 40%에서 80%로 높아지는 경우에는 별 차이를 보이지 않는다. 방문률은 사용자의 콘텐츠 정보와는 관계없이 페이지 계층에 대한 연관성을 높이는 효과를 주는 것으로 C4.5에서는 이 정보를 이용하지 않기 때문이다. 그러나 2-계층 클러스터링의 경우는 콘텐츠 정보와 링크 정보를 모두 사용하기 때문에 방문자 비율과 방문률이 증가함에 따



(그림 6) 규칙평가도 (90%, 80%)

라 규칙평가도 값이 커지는 것을 알 수 있다. 특히 방문자 비율 보다 방문률이 더욱 규칙평가도에 영향을 주고 있다. 이는 웹 로그를 분석하여 얻는 사용자의 페이지 방문정보가 행동규칙을 설정하는데 사용자의 콘텐츠 정보보다 더 유익하다는 것을 보여준다.

5. 결 론

일반적으로 웹 사용자들의 정보량은 부족하고 불확실성을 갖고 있어 사용자 데이터만을 갖고 사용자 그룹을 형성하고 웹 로그 분석을 통해 행동규칙을 추출하는 방식은 정확도가 떨어질 수 있다. 2-계층 클러스터링은 서로 다른 계층의 상호작용에 의한 잇점을 피하기 때문에 웹 사이트에서 유사 사용자의 그룹형성과 행동패턴을 추출하는데 효과적이다.

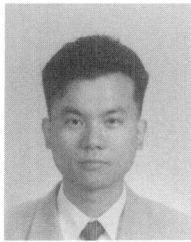
본 논문에서는 웹 사용자 그룹을 형성하고 이들의 행동패턴을 효율적으로 추출하기 위해 사용자 데이터 계층과 페이지 데이터 계층 간의 연관관계를 고려한 2-계층 클러스터링을 실시하였다. 그리고 각 계층에 서로 다른 클러스터링 알고리즘을 적용하고 분석함으로써 특정 알고리즘에 관계 없이 2-계층 클러스터링이 단일계층 클러스터링 보다 우수하다는 것을 알 수 있었다. 또한 성질이 다른 데이터 계층에 서로 다른 알고리즘을 적용하는 경우 좀더 좋은 성능을 가질 수 있음을 보여주었다.

그리고 본 논문에서는 생성된 사용자 클러스터로부터 사용자의 행동규칙을 추출하는 방법을 제시하였고 추출된 행동규칙이 내재된 행동패턴을 얼마나 잘 나타내는지를 C4.5와의 비교를 통해 살펴보았다. C4.5는 같은 행동패턴을 가지는 사용자들의 콘텐츠 정보가 유사할 때 좋은 결과를 보여주었지만 콘텐츠 정보가 어느 정도 다른 경우에는 나쁜 결과를 나타내었다. 2-계층 클러스터링은 사용자의 콘텐츠 정보 뿐만 아니라 페이지 데이터와의 연관성 정보를 함께 이용하기 때문에 콘텐츠 정보가 어느 정도 다르더라도 연관 정보를 같이 사용하므로 좋은 결과를 보여주었다. 그리고 클러스터의 정확도가 높을수록 행동규칙의 유용성이 높음을 실험을 통해 알 수 있었다.

참 고 문 헌

- [1] Vincent S. Tseng, Jeng-Chuan Chang, and Kawuu W. Lin, "Electronic commerce technologies(ECT): Mining and Prediction of Temporal Navigation patterns for personalized services in e-commerce," Proceedings of the 2006 ACM symposium on Applied computing, pp.867-871, 2006.
- [2] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Web Data Mining: Effective personalization based on association rule discovery from web usage data," Proceedings of the 3rd international workshop on Web information and data management, pp.9-15, 2001.
- [3] R. Kosala, and H. Blockeel, "Web Mining Research: A Survey," ACM SIGKDD Explorations, Vol.2 No.1, pp.1-15, 2000.
- [4] Y. Xie, and V. V. Phoha, "Web user clustering from access log using belief function," Proceedings of the 1st international conference on Knowledge capture, pp.202-208, 2001.
- [5] M. Gery, and H. Haddad, "Web clustering and usage mining: Evaluation of Web Usage Mining Approaches for User's Next Request Prediction," Proceedings of the 5th ACM international workshop on Web information and data management, pp.74-81, 2003.
- [6] J. W. Hwang, D. H. Song, and C. H. Lee, "Performance Analysis of 2-tier Clustering," 2006 International Conference Hybrid Information Technology, vol.2, pp.542-547, 2006
- [7] J. R. Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.
- [8] Brian S Everitt, "Hierarchical Clustering Techniques", Cluster Analysis, pp.55-90, 1993
- [9] H.J. Zeng, Z. Checn, and W. Y. Ma, "A Unified Framework for Clustering Heterogeneous Web Objects," Proceedings of the 3rd International Conference on Web Information Systems Engineering, pp.161-172, 2002.
- [10] J. Srivastava, R. Cooley, M. Deshpande, and P. N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," ACM SIGKDD Explorations, Vol.1 No.2, pp.12-23, 2000.

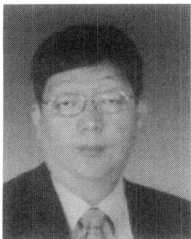
- [11] C. Stanfill, and D. Waltz, "Toward memory-based reasoning," Communications of the ACM, Vol.29, No.12, 1986.
- [12] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of document clustering techniques," KDD Workshop on Text Mining, 2000.
- [13] Open Directory Project, <http://dmoz.org/>, 2007.



황 준 원

e-mail : jwhwang@konkuk.ac.kr
 1998년 건국대학교 컴퓨터공학과(학사)
 2000년 건국대학교 컴퓨터공학과
 (공학석사)
 2002년~현 재 건국대학교 컴퓨터공학과
 박사과정수료

관심분야: eCRM, 웹 마이닝, 데이터마이닝 등



송 두 현

e-mail : dsong@ysec.ac.kr
 1981년 서울대학교 계산통계학과(학사)
 1983년 한국과학기술원 전산학과(석사)
 1994년 캘리포니아대학교 전산학과
 박사과정수료
 1983년~1986년 KIST 연구원

1997년~현 재 용인송담대학 컴퓨터게임정보과 교수
 관심분야: 기계학습, 데이터마이닝, 데이터베이스, 보안 등



이 창 훈

e-mail : chlee@konkuk.ac.kr

1980년 연세대학교 수학과(학사)
 1977년 한국과학기술원 전산학과(석사)
 1993년 한국과학기술원 전산학과(박사)
 1996년~2000년 건국대학교 서울캠퍼스
 정보통신원 원장

2000년~2002년 건국대학교 정보통신대학원 원장
 2001년~2002년 건국대학교 정보통신대학 학장
 1980년~현 재 건국대학교 컴퓨터공학과 교수
 관심분야: 지능시스템, 운영체제, 보안, 전자상거래 등

웹 기반의 동적 웹 서버 페이지 생성기 설계 및 구현

신 용 민[†] · 김 병 기^{††}

요 약

인터넷이 대중화 되면서 내부 데이터베이스로 관리되고 있는 정보를 동적 웹 서버 페이지를 작성하여 웹으로 제공하는 다양한 웹 어플리케이션 개발이 수행되고 있으나, 대부분의 경우 체계적인 개발 방법론의 채택 없이 직접 프로그램을 작성하거나, 규모에 맞지 않는 대규모 개발 방법론을 적용하여 개발의 효율성을 떨어뜨리고 있다. 웹 어플리케이션이 체계적인 개발 방법론을 따르지 않고 스크립트 언어를 사용하여 임의 개발 뒤에 따라 프로그램 개발의 생산성, 유지 보수성, 그리고 재 사용성을 저하시키게 된다. 본 논문에서는 빠르고 효과적인 스크립트 기반의 웹 어플리케이션 개발을 위하여 데이터베이스를 이용한 동적 웹 서버 페이지 자동 작성 도구를 설계하고 이를 구현하였다. 데이터베이스를 이용한 동적 웹 서버 페이지 패턴을 분석하여 정형화된 스크립트 모델을 제시하고 데이터 바운드 컨트롤 태그 생성기를 표준 스크립트로 작성하여 웹 어플리케이션 개발과 유지보수에 활용되어 생산성 향상에 기여할 수 있다.

키워드 : 웹 어플리케이션, 동적 서버 페이지, 사이트 빌더

Design and Implementation of Dynamic Web Server Page Builder on Web

Shin Yong Min[†] · Kim Byung Ki^{††}

ABSTRACT

Along with the trend of internet use, various web application developments have been performed to provide information that was managed in the internal database on the web by making a web server page. However, in most cases, a direct program was made without a systematic developmental methodology or with the application of a huge developmental methodology that is inappropriate and decreased the efficiency of the development. A web application that fails to follow a systematic developmental methodology and uses a script language can decrease the productivity of the program development, maintenance, and reuse. In this thesis, the auto writing tool for a dynamic web server page was designed and established by using a database for web application development based on a fast and effective script. It suggests a regularized script model and makes a standardized script for the data bound control tag creator by analyzing a dynamic web server page pattern with the database in order to contribute to productivity by being used in the web application development and maintenance.

Key Words : Web Application, Dynamic Server Page, Site Builder

1. 서 론

동적 웹 서버 페이지는 웹 브라우저에서 요청한 페이지를 웹 프로그램이 데이터베이스로부터 데이터를 추출하여 콘텐츠로 구성하여 보여주는 것으로서, Oracle, Microsoft, IBM과 같은 데이터베이스 벤더가 데이터베이스 관리용 도구를 제공하고 있지만, 웹 콘텐츠를 위한 페이지구성의 기능은 제공하지 않고 있다[1]. 엔터프라이즈 환경에서 웹 기반 어플리케이션은 SUN의 EJB (Enterprise Java-Beans), 마이크로소프트의 DCOM 또는 OMG의 CCM(CORBA Component Model)과 같은 분산 컴포넌트로 개발하여 처리하는 사례도 있지만

대부분의 웹 어플리케이션 환경에서는 동적 웹 페이지 작성에 유리한 ASP(Active Server Page), JSP(Java Server Page), PHP(Professional HTML Preprocessor)와 같은 스크립트 언어를 이용하여 개발되고 있다[2, 3, 4].

스크립트 지원 언어는 어플리케이션 개발 언어가 가지고 있는 강건한(robust) 특성을 완화하여 비교적 자유롭게 프로그램 할 수 있도록 되어 있어서 프로그램 코드를 패치(patch) 적인 형태로 구현하는 경우가 많기 때문에 대부분의 프로그램 소스가 개발자의 성향에 따라 다양한 형태로 개발됨으로써 정형화 되지 못하고 있다. 또 프로그램의 재사용 측면에서도 스크립트 소스를 컴포넌트로 만들어 사용하지 않고 유사한 프로그램을 복사하여 코드 레벨에서 수정하여 사용하는 정도에 그치고 있다[5].

본 논문에서는 데이터베이스를 이용한 동적 웹 서버 페이

† 정 회 원 : (주)가민정보시스템 대표이사
 †† 중신회원 : 전남대학교 전자컴퓨터정보통신공학부 교수
 논문접수 : 2007년 2월 22일, 심사완료 : 2007년 11월 21일

지 패턴을 분석하여 정형화된 스크립트 모델을 제시하고, 데이터 바운드 컨트롤 태그 생성기를 스크립터로 구현하였다. 본 논문의 결과물은 스크립트 기반의 웹 어플리케이션 개발을 위한 소프트웨어 프레임워크를 제공하고 대부분의 데이터베이스 기반 웹 콘텐츠의 경우 프로그램을 작성하지 않고도 웹 페이지를 작성할 수 있도록 기여할 수 있다.

논문의 2장에서는 관련 연구를 기술하고, 3장에서는 데이터베이스를 이용한 동적 웹 서버 페이지의 패턴을 분석하고 웹 서버 페이지 생성기를 설계한다. 4장에서는 3장에서 설계한 내용을 토대로 한 구현을 설명하고 마지막으로 5장에서는 결론과 향후 연구 방향을 제시한다.

2. 관련연구

2.1 상용화된 웹 어플리케이션 개발 도구

웹 어플리케이션에 대한 개발방안 및 모델링 방안들이 제기되고 있으나 상용화된 방법론은 거의 없으면, 규모가 큰 조직의 경우 나름대로의 개발 방법론을 가지고 있지만, 중소규모의 경우에는 제대로 된 개발 방법론 없이 직접 프로그램을 작성하거나, 규모에 맞지 않는 대규모 개발 방법론을 적용하여 효율성을 떨어뜨리고 있다.

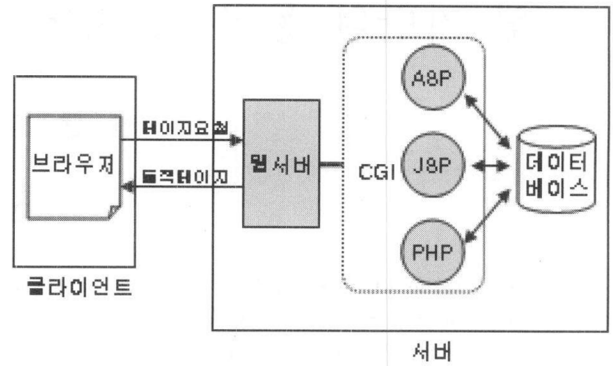
현재 웹 어플리케이션에서 일반적으로 사용하고 있는 방법론들은 기존에 존재하는 범용적인 개발 방법론을 바탕으로 약간의 특성화를 통하여 수행하는 것이다. RUP[6]와 같은 대규모 방법론에서 웹 어플리케이션을 위한 확장 방법들이 소개되고 있지만 다양한 웹 페이지를 수용하기에 부적절한 상황이다.

웹 사이트 구축을 지원해주는 콘텐츠 관리 시스템들의 경우 사이트 관리 및 웹 프로그램 관리에는 우수한 기능을 제공하지만 웹 페이지를 자동으로 작성해주는 기능은 지원하지 않는다[7, 8]. 또, 홈페이지 빌더 같은 제품들도 사이트 생성 및 관리를 위한 기능은 제공하지만 데이터베이스와 연동되는 웹 페이지를 자동으로 만들어 주는 기능은 지원하지 않는다[9, 10].

2.2 동적 웹 페이지 처리

웹에서 처리되는 어플리케이션은 서버와 클라이언트간의 정보 전달을 위해 HTTP 프로토콜을 사용한다. HTTP 프로토콜은 상태를 관리 하지 않기(Stateless) 때문에 하나의 페이지에 대해 하나의 트랜잭션이 처리되는 페이지 특성을 가지고 있다[11].

(그림 1)에 스크립트 기반의 동적 웹 페이지 생성 환경을 구성하는 클라이언트, 웹 서버, 스크립트 엔진과 데이터베이스의 관계가 표시되어 있다. 클라이언트가 브라우저를 통해 웹 서버에 있는 동적 웹 페이지 스크립트를 URI(Uniform Resource Identifier) 형태로 호출하면, 웹 서버의 스크립트는 이 요청을 해석하고 스크립트를 번역한 다음 번역된 문장을 실행시킨다. 데이터베이스 정보를 필요로 할 경우는 데이터베이스로부터 정보를 가져와 표현에 적합한 태그를



(그림 1) 동적 웹 서버페이지 처리 환경

불인 다음 HTML형태로 클라이언트에 동적 페이지의 결과를 전송한다.

2.3 스크립트 언어의 응용

스크립트 언어는 성능이나 개발 생산성에 대한 의구심으로 어플리케이션 개발 프로젝트에서 외면하는 경향이 있다. 하지만 동적 웹 서버 페이지 작성에서 스크립트 언어는 성공적으로 선호되고 있다. PERL 스크립트 언어는 기존에 C로 작성되던 CGI 개발 과정을 훨씬 수월하게 대체 할 수 있게 해주었다. 이러한 영향은 PHP나 ASP 그리고 JSP와 같은 웹 전용 스크립트 언어를 탄생하게 했으며 오늘날 대부분의 웹 애플리케이션 시스템은 특정 모듈 또는 컴포넌트를 제외하고 대부분 웹 전용 스크립트 언어로 작성되고 있다[12].

2.4 HTML

HTML은 SGML(Standard Generalized Markup Language)을 모체로 하여 국제 표준 기구인 ISO에 의해 1986년에 채택되어 웹에서 사용되는 표준 언어이다. 여기서 마크업 언어란 문서 내용에 글꼴, 문단 모양과 같은 각종 정보를 표시해서 문서의 모양과 구조를 결정하는 언어를 의미한다. 그리고 하이퍼텍스트는 하이퍼링크를 이용하여 웹 정보를 연결해 놓은 문서를 의미한다[13].

HTML은 1999년 12월에 HTML 4.01이 최종 권고안으로 나오고 종료되었으며, 현재는HTML을 XML로 확장시킨 XHTML이 발표되었다[14].

2.5 스타일 시트

스타일 시트란 문서의 외형인 글자 크기, 글자 모양, 줄 간격, 배경 색상, 테두리 장식 등을 다양하고 효과적으로 지정할 수 있도록 지원하는 개념이다. 스타일 시트에는 CSS(Cascading Style Sheet)와 JSSS(JavaScript Style Sheet)의 두 가지 문법이 사용되고 있다. CSS는 속성과 속성값을 이용하여 스타일을 지정하지만, JSSS는 자바 스크립트를 이용하여 스타일을 지정한다. JSSS는 넷스케이프만 지원하기 때문에 보편적인 스타일 시트를 지정하려면 CSS를 주로 사용해야 한다[7].

1996년 12월에 CSS 레벨1 권고안이 나온 이래로 1998년

5월에 CSS 레벨1에 특정 미디어 지원이 추가된 CSS 레벨2가 제출되었고 현재는 CSS 레벨2 Revision 1이 확정되어 있으며 CSS 레벨 3이 개발 중에 있다[15].

2.6 ASP

ASP(Active Server Pages)는 1995년 말에 등장한 IIS (Internet Information Server)의 세 번째 버전으로 마이크로소프트의 웹 서버인 IIS 3.0 이상에서만 동작하는 특별한 페이지이다. ASP 페이지는 Web을 프로그래밍 할 수 있도록 해 주는 서버에서 동작하는 페이지로서 기존의 HTML 페이지는 문서 내용이 미리 작성되어 있는 정적인 페이지 인데 반해서 클라이언트가 요청하는 시점에 문서 내용을 만들어서 응답하는 동적인 페이지 구성을 가질 수 있게 해 준다.

현재 최종 버전인 IIS 6.0은 안정성, 성능, 관리, 보안 및 응용 프로그램 서비스가 강화되어 윈도우2003서버에 포함되어 설치할 수 있다[16].

2.7 JScript

JScript는 Microsoft가 ECMA 262 언어 규정(ECMAScript 버전 3)을 구현한 것으로 객체 기반의 스크립팅 언어이다. 비록 JScript가 C++ 같은 객체 지향 언어보다 기능은 적지만, 웹 페이지를 다루기에는 충분한 기능을 가지고 있다.

JScript는 Java와 간접적으로 관련되어 있기는 하지만 Java 언어의 축소판이나 간소화한 버전이 아니다. JScript는 프로그램 구현이 제한적이다. 예를 들어 독립적인 응용 프로그램을 쓸 수 없고 파일을 읽거나 쓸 수 있도록 언어 자체가 내장된 기능을 지원하지 않는다. 게다가 JScript의 스크립트는 해석기 또는 ASP(Active Server Pages), Internet Explorer 또는 Windows Script Host 같은 “스크립팅 호스트”가 있어야만 실행될 수 있다[17].

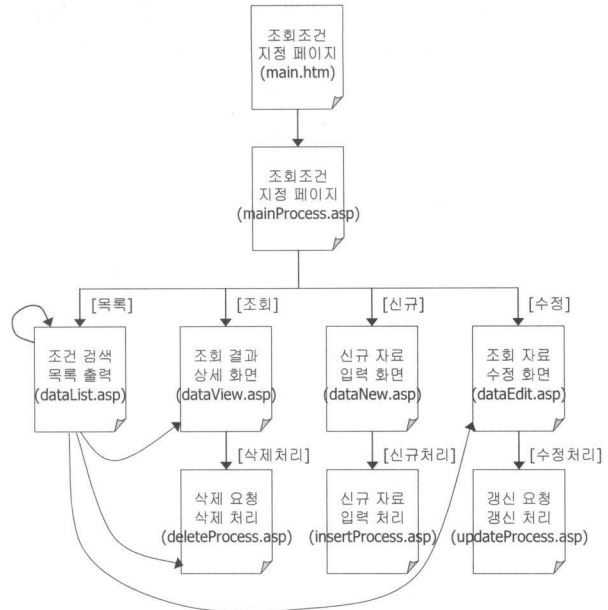
3. 동적 웹 서버 페이지 생성기 분석 · 설계

3.1 웹 서버 페이지 작성 패턴 분석

데이터베이스를 다루는 웹 어플리케이션은 (그림 2)와 같은 네비게이션 체계를 가지고 있다. 이러한 구조는 하나의 데이터베이스 뷰에 대한 어플리케이션을 구현하기 위해 9개 정도의 웹 페이지 스크립트가 있어야 한다는 것을 보여주고 있으며 만약 데이터베이스 뷰가 변경되면 웹 페이지 스크립트도 변경해야 하는 문제를 가지고 있다. 또한 조회, 입력, 갱신, 삭제와 같은 일련의 처리를 이해하기 위해 여러 개의 스크립트 파일을 찾아봐야 하는 불편함이 있다.

(그림 3)은 대부분의 동적 웹 서버 페이지 생성 프로그램에 있는 유사 스크립트 코드 블록이며 각 스크립트 행에 대한 설명은 다음과 같다.

- 01행은 이 프로그램이 트랜잭션 처리가 있는 페이지이고 사용되는 스크립트 언어가 JScript임을 선언하고 있다.
- 02행은 웹 어플리케이션 모두가 참조하기 위한 공용



(그림 2) 데이터베이스 관련 웹 어플리케이션의 네비게이션 체계

```

01: <%@ Transaction=Supported Language=JavaScript%>
02: <!--#include file="inc/globalConstant.asp" -->
03: <%
04: var ls_code = Request.QueryString("code");
05: var ls_form = Request.Form("data");
06:
07: // action에 따른 데이터베이스 객체 생성 및 처리
08:
09: %>
10: <HTML>
11: <HEAD><TITLE>My Page</TITLE>
12: <LINK REL="stylesheet" TYPE="text/css"
13: HREF="<%=gs_mainDirectory%>inc/default.css">
14: <SCRIPT LANGUAGE=javascript SRC="com/clientValidate.js"></SCRIPT>
15: <FORM name=fm_form method=post action="coUser.asp?<%=ls_queryString%>">
16: <TABLE align="<%=gi_tableAlign%>" border=0 cellpadding=0 cellspacing=0>
17: <%
18: gf_setButton("If_confirm", ls_button);
19: %>
    
```

(그림 3) 웹 서버 페이지의 스크립트 패턴

변수의 SSI(Server Side Include)를 지정하고 있다.

- 03행부터 09행까지는 동적 페이지를 생성하기 전에 데이터베이스 조회 등 서버 측에서 수행되어야 할 처리 로직이 들어가는 부분이다.
- 12행은 HTML 문서의 글자 크기, 글자 모양, 배경 색상 등의 지정을 웹 어플리케이션 전체에 걸쳐 일괄 적용하기 위한 CSS 외부 파일 이름을 지정하고 있다.
- 13행은 클라이언트 폼 검증을 위한 여러 가지 자바 스크립트가 코딩 되어있는 스크립트 파일을 연결하고 있다.
- 18행은 명령 버튼을 생성해주는 공통 함수를 호출하는 스크립트다.

3.2 동적 웹 서버 페이지 생성기 설계

웹 서버 페이지를 생성하는 동일 패턴의 웹 페이지 템플

리스트를 클래스라고 하기로 한다. 하나의 클래스는 하나 또는 그 이상의 데이터베이스 테이블을 조인한 뷰 또는 동적 SQL 조회에 의한 레코드 셋을 보여주거나, 해당 테이블에 대한 입력, 수정, 삭제와 같은 일련의 SQL 명령에 의한 데이터베이스 트랜잭션을 처리하게 된다.

(그림 4)는 웹 서버 페이지 생성기의 클래스 다이어그램이다. nv_class테이블은 전체 사이트에서 관리되는 모든 템플릿 페이지의 정보를 가지고 있다. nv_classProfile은 웹 서버 페이지를 동적으로 작성해서 내보낼 classTemplate 에서 사용할 페이지 작성 정보를 가지고 있다. nv_classTable 은 테이블을 동적으로 생성하고 관리하기 위한 칼럼정보를 가지고 있으며 nv_classField는 웹 서버 페이지의 페이지 작성을 위한 입력 혹은 출력 필드의 속성을 가지고 있다.

입력 폼에 있는 각종 필드의 유효성 검사를 위한 클라이언트 스크립트는 cli-entValidate.js에 자바 스크립트로 구현되어 있다. nv_classField에 있는 페이지 구성 필드의 속성에 따른 HTML TAG 작성은 globalTag의 메소드를 호출하여 작성하고 타이틀, 리스트, 상세 보기, 버튼 표시, 페이지 표시, 링크 연결, URL 리다이렉션 등 전체 페이지 구성 및 출력은 페이지의 유형에 따라서 Master.asp 혹은 Detail.asp

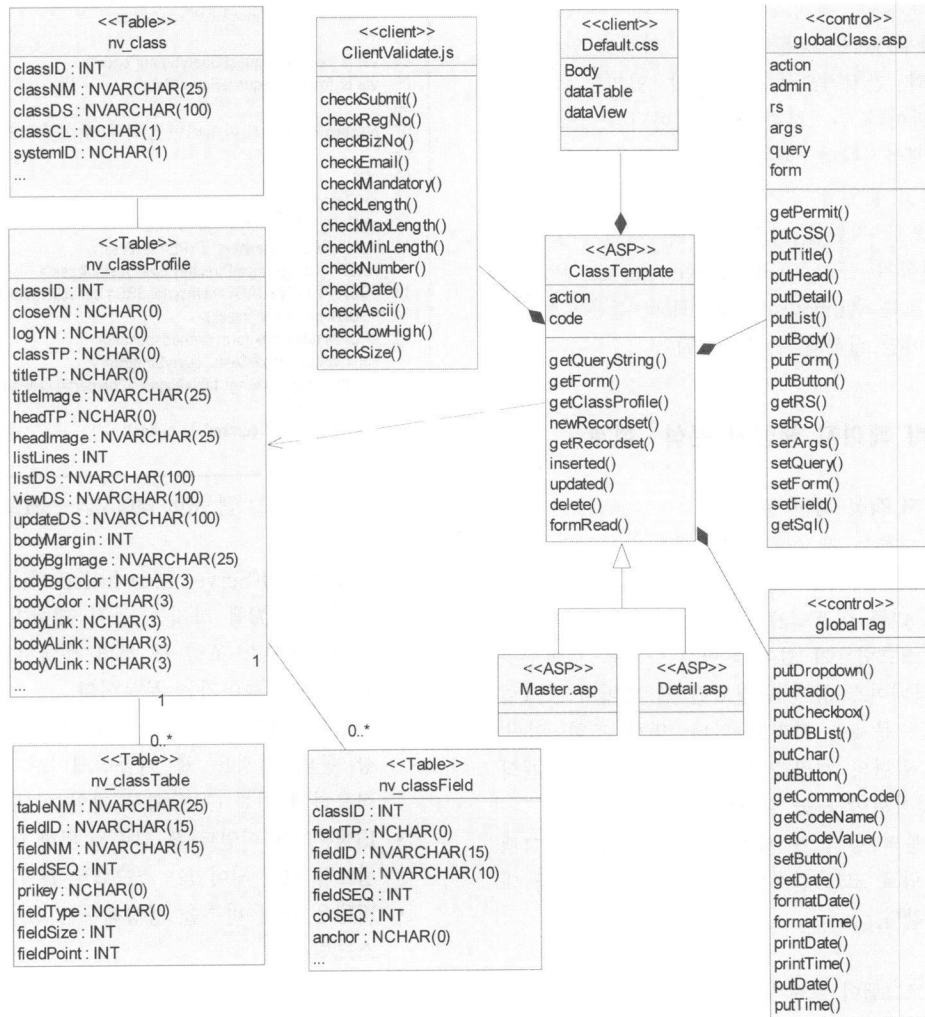
가 처리하게 된다. Master.asp는 목록과 상세보기가 하나의 화면에 표시되는 형태이고, Detail.asp는 목록과 상세보기가 별도의 화면으로 표시되는 형태이다.

4. 동적 웹 서버 페이지 생성기 구현

앞 절의 분석·설계를 바탕으로 데이터베이스 관련 웹 서버 페이지 생성 표준 스크립트는 IIS 기반의 ASP를 사용하여 구현하였다. ASP의 기본 언어인 VBScript 는 언어 호환성이 약하고 객체 지향 언어가 아니기 때문에 다른 언어로의 확장을 고려하여 자바 스크립트와 호환성이 있는 JScript 언어를 사용하였다. 데이터베이스는 Access를 사용하여 구현하였으나 범용의 RDBMS로도 구현이 가능하도록 확장성을 부여하였다. 생성기의 이름은 Dynamic Web Server Page Builder로 하며 간단히 DWSP Builder로 표현한다.

4.1 웹 테이블 관리

(그림 5)는 웹 테이블 관리 화면이다. 웹 페이지에서 사용할 테이블을 새로 만들고 그 테이블을 구성하기 위해 칼럼을 정의할 수 있으며, 등록된 테이블을 웹에서 직접 Create,



(그림 4) 웹 서버 페이지 생성 클래스 다이어그램

Alter, Delete 할 수 있다. (그림 5)의 (a)는 이미 등록되어있는 테이블의 목록이 나타나고 여기에서 테이블을 선택하여 수정하거나 삭제할 수 있고 새로운 테이블을 신규 작성 하게 된다. (그림 5)의 (b)는 테이블 신규 작성에서 필요한 컬럼을 등록하는 화면이다.

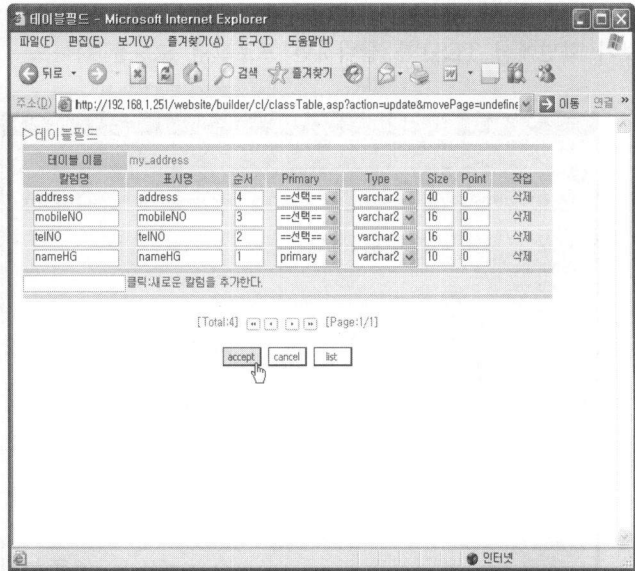
4.2 클래스 관리

(그림 6)은 클래스 관리 화면이다. (그림 6)의 (a)에 있는 클래스 목록보기에서 프로파일 열에 나온 '로그'는 해당 클래스 프로그램 사용정보를 보여주고 '작성'은 프로파일이 아직 등록 안된 클래스의 프로파일을 새로 작성하며 '수정'은 이미 등록된 클래스의 프로파일 수정하기, 그리고 '삭제'는 이미 등록된 클래스의 프로파일 삭제하기이다. 프로파일 열에 나온 '수정'은 클래스 정보를 수정하고 '삭제'는 클래스

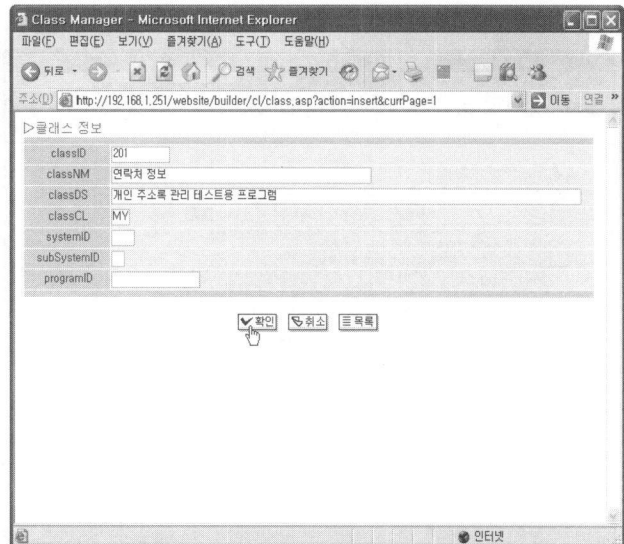
정보를 삭제한다. (그림 6)의 (b)는 클래스의 상세 정보이다.

4.3 클래스 프로파일 정보 관리

(그림 7)은 클래스 프로파일 정보 관리 화면이다. 클래스의 봉쇄 여부와 사용 로그, 목록과 편집 보기 화면을 위한 SQL 문장, 그리고 접근 권한이나 필드 속성, 표현 양식 등을 지정하는 클래스 프로파일을 관리한다. (그림 7)의 (a)에 있는 SQL 입력항목의 '필드설정' 버튼은 각각 목록 화면, 상세보기 화면, 편집 화면에서 보여줄 화면 구성 요소(컬럼 혹은 필드)를 설정한다. (그림 7)의 (b)에 있는 '자동생성' 버튼은 insert, update, delete를 위한 SQL 문장을 목록뷰 SQL에 지정된 SQL쿼리 문장과 편집뷰SQL에 지정된 SQL 쿼리 문장을 참조하여 자동으로 생성시켜 주는 프로그램을 호출한다.



(그림 5) 웹 테이블 관리 화면



(그림 6) 클래스 프로파일 정보 관리 화면

4.4 필드 관리

(그림 8)은 클래스와 관련된 테이블 뷰의 각 칼럼의 화면 표시 속성을 정의하는 필드관리 화면이다. 클래스 프로파일 처음 작성되어 아직 필드 속성이 정의되어 있는 않았을 때는 (그림 8)의 (a)에서 '신규' 버튼을 눌러서 뷰에 등록된 기본 필드를 (그림 8)의 (b)와 같이 자동 생성하여 준다.

4.5 템플릿 디버그

(그림 9)는 템플릿 디버그 실행 화면이다. (그림 9)의 (a)는 클래스 템플릿에 설정된 속성에 의해 ASP 프로그램 코드로 어떻게 구현되는지를 확인하고 (그림 9)의 (b)는 ASP 프로그램에 의해 HTML 태그로 구현된 것을 확인한다.

4.6 웹 페이지 테스트

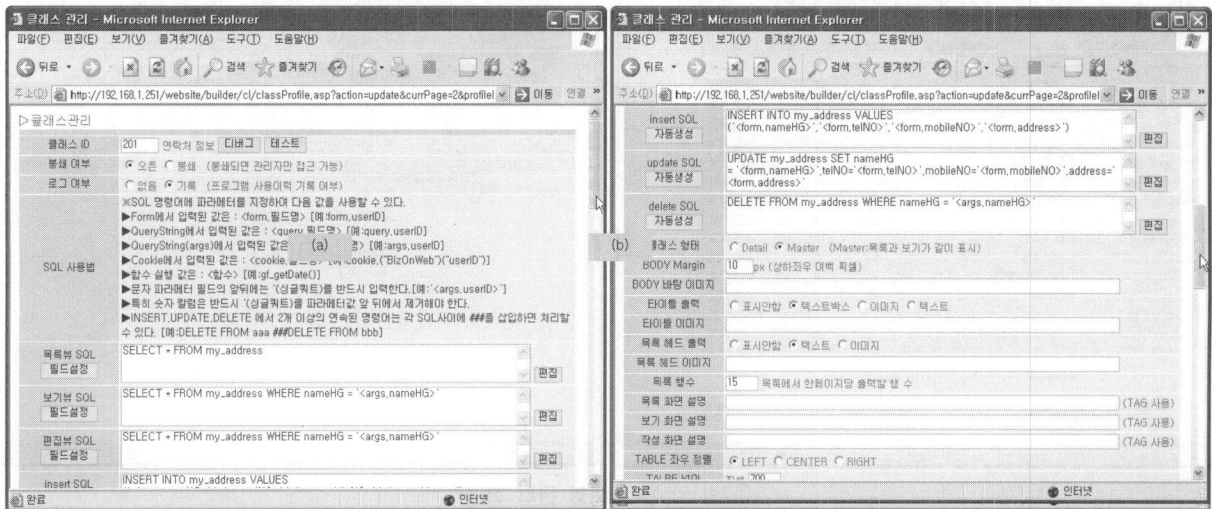
(그림 10)은 템플릿 디버그 기능을 통해 웹 페이지 작동 테스트를 한 결과이다. (그림 10)의 (a)에 있는 HTML 보기 화면에서 TEST를 클릭하면 프로그램 코딩을 하나도

하지 않은 상태에서 (그림 10)의 (a) 및 (그림 10)의 (b)와 같이 목록보기와 신규 자료 입력, 갱신, 삭제 등이 원활히 이루어 짐을 확인할 수 있다.

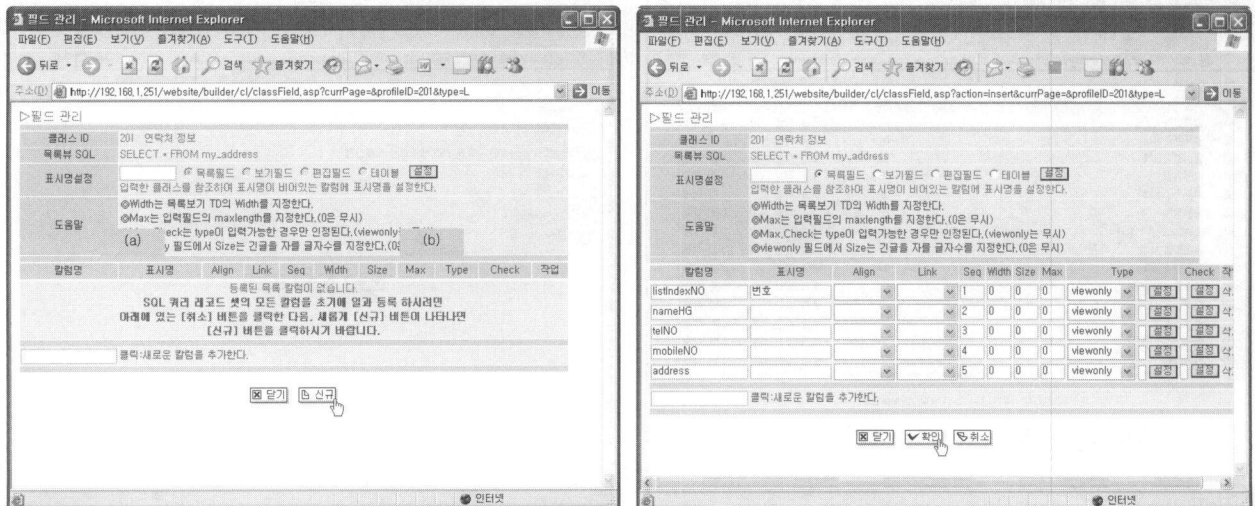
5. 상용화된 웹 어플리케이션 개발 도구와의 비교

본 장에서는 본 논문에서 구현한 스크립트 기반의 웹 어플리케이션 소프트웨어 프레임워크와 데이터베이스 기반 동적 웹 서버 페이지 생성기를 이용한 웹 사이트 구축 방법과 상용화된 웹 어플리케이션 개발 도구를 이용한 웹 사이트 구축 방법의 특징을 비교하고 그 결과를 <표 1>로 제시한다.

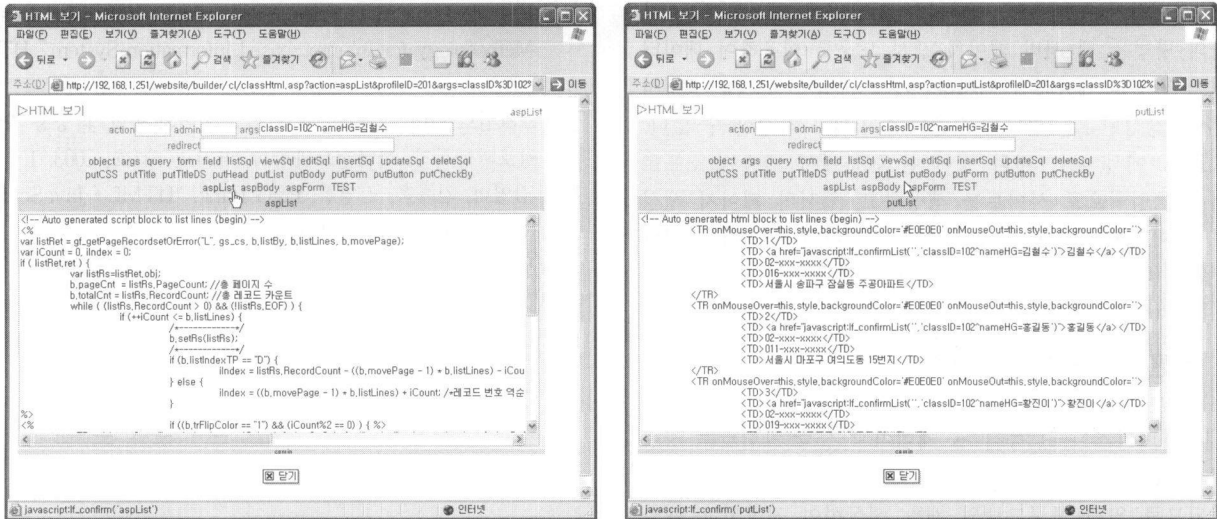
초기 사이트 생성의 경우 콘텐츠 관리 시스템과 홈 페이지 빌더의 경우는 개발 도구가 제공하는 기능에서 선택하여 페이지 위주의 사이트를 생성하는 반면, 본 논문의 방법은 스크립트 기반의 소프트웨어 프레임워크를 이용하기 때문에 개발자에게 많은 유연성을 제공한다. 페이지 디자인을 할 때 콘텐츠 관리 시스템이나 홈페이지 빌더는 제공되는 스킴



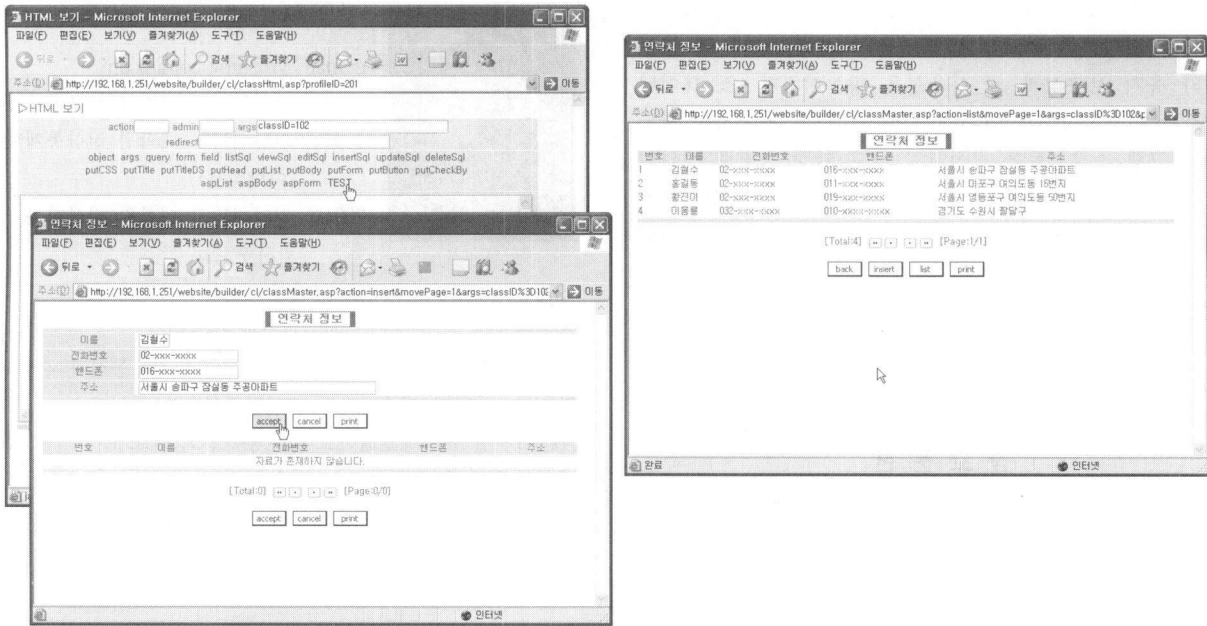
(그림 7) 필드 관리 화면



(그림 8) 필드 관리 화면



(그림 9) 템플릿 디버깅



(그림 10) 웹 페이지 테스트

<표 1> 상용화된 웹 어플리케이션 개발 도구와의 비교

비교항목	개발도구	본 논문의 방법	콘텐츠 관리 시스템		홈 페이지 빌더	
			E[7]	C[8]	D[9]	T[10]
사이트 생성 방법	프레임워크		사이트관리자	사이트관리자	위저드	메뉴
네비게이션 구성	O	O	O	O	O	O
템플릿 제공 방식	소스방식		메뉴방식	메뉴방식	스킨방식	스킨방식
페이지 디자인	생성기이용		에디터	에디터	스킨방식	스킨방식
100% 웹 기반	O	O	O	O	O	O
스크립트 소스 생성	O	X	X	X	X	X
스크립트 디버거	O	X	X	X	X	X
데이터베이스관리지원	O	X	X	O	X	X
DB바운드 필드 작성	생성기이용	X	X	확장태그	X	X

에서 선택하거나 비주얼 에디터를 사용하여 작성하지만, 본 논문의 방법은 웹 서버 페이지 생성 도구를 이용하여 데이터베이스 필드를 페이지에 배치하거나 스크립트 소스를 생성시킨 다음 생성된 스크립트를 수정하여 임의로 작성할 수 있다.

<표 1>에서 보는 바와 같이 본 논문에서 제시한 소프트웨어 프레임워크로 사이트를 생성하고, 웹서버 페이지 생성기로 데이터베이스 관련 콘텐츠를 작성하면 스크립트 기반의 웹 어플리케이션을 개발할 때 매우 효율적이다.

5. 결론 및 향후 연구 과제

데이터베이스 관련 동적 웹 페이지 작성 스크립트는 정형화된 코딩 가이드가 없고 개발자의 성향에 따라 스크립트가 작성되므로 인해 프로그램의 개발 생산성이 저하되고 개발된 프로그램의 유지보수가 어려울 뿐만 아니라 재 사용성도 매우 부족하였다.

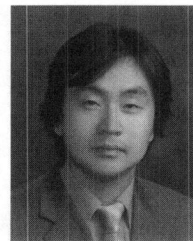
본 연구에서는 데이터베이스 관련 웹 서버 페이지 패턴을 분석하여 정형화된 스크립트 모델을 작성함으로써 웹 프로그램의 유지 보수성과 재 사용성을 향상시켰다. 또 데이터베이스를 이용하여 만들어지는 동적 웹 서버 페이지를 별도의 프로그램 스크립트를 작성하지 않고 클래스 별로 정의된 클래스 프로파일을 읽어 생성해 주는 도구를 개발 함으로써 웹을 통한 데이터베이스 조작을 가능하게 하였고 웹 어플리케이션 개발 생산성을 향상시켰다.

향후 연구과제로는 데이터베이스 관련 웹 페이지 생성 스크립트를 어떤 플랫폼에서나 사용이 가능하도록 JAVA 기반의 스크립트 및 XML로 확장할 계획이다.

참 고 문 헌

[1] EJB : Sun, Enterprise JavaBeans Specification Version 1.1, Sun Microsystems Inc, 1999
 [2] CCM : OMG, CCM Revised Submission, OMG TC Document orbos/99-07-01, 1999
 [3] DCOM : Microsoft, DCOM Technical Overview, URL : http://msdn.microsoft.com/library/backgrnd/html/msdn_dcomtec.htm, 1996
 [4] Clemens Kerer, Engin Kirda, and Roman Kurmanowystsch, "A Generic Content Management Tool for Web Databases," IEEE Internet Computing, Vol.6, No.4, pp.38-42, 2002. 7.
 [5] D. Lowe, and R. Webby, "Web Development Process Modeling and Project Scoping: Work in Process," First Internation Workshop on Web Engineering, WWW7 Conference, 1998. 4
 [6] Rational Unified Process 2002.05.00.25, Rational Software Corporation
 [7] http://www.3070.co.kr/_ezaid/project/templateshopnew.jsp
 [8] <http://www.contentwise.co.kr/cms/cms.jsp>
 [9] <http://www.dubuweb.com/web/aboutus/aboutus01.php>
 [10] <http://www.homepagebuilder.co.kr/together/>

[11] F.A. de Lima, R.T. Price, "Towards an Integrated Design Methodology for Internet-based Information System," HTF5, 1998
 [12] 스크립트 언어 : 김익순, "스크립트 언어의 동향 및 응용 방안," 정보과학회지, 한국정보과학회, 제21권 제1호, 2003. 1.
 [13] 김영미, 김륜옥, 김영실, P&G Bank, "HTML + JavaScript + Css/DOM," 헤지원, 2002.
 [14] <http://www.w3.org/MarkUp/>
 [15] <http://www.w3.org/Style/CSS>
 [16] <http://www.microsoft.com/korea/TechNet/iis/Default.asp>
 [17] <http://www.microsoft.com/korea/scripting/Default.asp>



신 용 민

e-mail : ymshin@comin.com

1992년 조선대학교 전자공학과(공학사)
 1999년 전남대학교 대학원 전산통계학과 (이학석사)
 2004년 전남대학교 대학원 전산학과 박사수료

1979년~1993년 (주)금호 전산실
 1994년~현 재 (주)가민정보시스템 대표이사
 2000년~2005년 전남 도립남도대학 컴퓨터정보통신과 겸임교수
 2002년~2004년 광주대학교 디지털경영학부 겸임교수
 2004년~현 재 목포대학교 정보공학부 겸임교수
 관심분야 : 소프트웨어공학, 객체지향시스템, 컴포넌트기반소프트웨어, 웹 어플리케이션



김 병 기

e-mail : bgkim@chonnam.ac.kr

1978년 전남대학교 수학과(이학사)
 1980년 전남대학교 대학원 수학과 (이학석사)
 2000년 전북대학교 대학원 수학과 (이학박사)

2000년~2002년 전남대학교 자연과학대학 학장
 2000년~2003년 한국정보처리학회 S/W공학연구회 운영위원장
 2007년~2007년 한국정보처리학회 회장
 1981년~현 재 전남대학교 전자컴퓨터정보통신공학부 교수
 관심분야 : 소프트웨어공학, 객체지향시스템, 컴포넌트기반소프트웨어 개발