

# 중포된 객체에 대한 질의처리를 위한 다차원 중포 속성 색인기법

이종학, 김규평  
 jhlee11@cuth.cataegu.ac.kr, g9521005@cuth.cataegu.ac.kr  
 대구호성가톨릭대학교 전자정보공학부

## A Multidimensional Nested-Attribute Indexing for Queries on Nested Objects

Jong-Hak Lee, Kyu-Pyong Kim  
 School of Electronics and Information Engineering, Catholic Univ. of Taegu-Hyosung

### 요약

본 논문에서는 객체지향 데이터베이스의 중포된 객체에 대한 질의처리를 효율적으로 지원하기 위한 다차원 중포 속성 색인기법을 제안한다. 중포된 객체에 대한 기존의 색인기법들은 일차원 색인구조를 이용함으로써 중포된 객체의 속성과 클래스 계층이 포함된 다양한 형태의 질의들에 대한 처리를 효율적으로 지원하지 못하는 문제점을 가지고 있다. 다차원 중포 속성 색인기법에서는 다차원 화일구조를 이용하여 중포 속성의 키값 도메인과 함께 중포 속성을 표현하는 경로상의 모든 속성에 대해 각 속성이 정의된 클래스 계층마다 클래스 식별자 도메인을 할당함으로써, 다차원 도메인 공간상에서 색인 엔트리들의 클러스터링을 다룬다. 따라서, 다차원 중포 속성 색인기법에서는 기존의 색인기법에서 지원하기 어려운 질의의 대상 범위가 클래스 계층상의 임의의 클래스들로 제한되거나, 질의에 포함된 복합 속성들의 도메인이 클래스 계층상의 임의의 클래스들로 제한되는 경우에도 효율적으로 지원할 수 있다.

### 1. 서론

객체지향 데이터베이스는 클래스 상속(inheritance) 개념에 의해 클래스들 사이에 하나의 계층구조를 이룬다. 즉, 하나의 클래스는 여러개의 서브클래스를 가지며, 각 서브클래스는 또 다른 여러 서브클래스들을 가진다. 따라서, 객체지향 데이터베이스에서는 질의의 대상 범위를 두가지 경우로 나타낼 수 있다. 한 경우는 질의의 대상 범위를 질의에 나타나는 클래스만으로 한정하는 것이고, 또 다른 경우는 질의의 대상 범위를 질의에 나타나는 클래스와 그의 모든 서브클래스들을 포함하는 것이다[3,6,7].

또한, 객체지향 데이터베이스는 클래스 집산화(aggregation) 개념에 의하여 한 클래스가 가지는 속성의 도메인이 또 다른 클래스가 되게 함으로써(이러한 속성을 복합 속성이라 함) 클래스 집산화 계층을 이룬다. 따라서, 클래스 집산화 계층상의 모든 클래스에서 정의된 어떠한 속성도 논리적으로는 루트 클래스의 속성이라고 볼 수 있다. 우리는 이런 속성을 루트 클래스의 중포 속성(nested attribute)[2,3,6]이라 한다. 이로 인하여 객체지향 질의어에서는 기존의 관계형 데이터베이스에서의 질의어와는 달리 중포 속성에 조건이 주어지는 중포 술어(nested predicate)[2,6]를 가진다는 특징이 있다. 중포 속성을 표현하기 위해서는 경로식(path expression)[3,5]이 사용되며, 경로식은 루트 클래스로부터 클래스 집산화 계층을 따라 나타나는 속성들의 나열로 표현된다.

중포 속성을 표현하는 경로식에는 속성값(객체 참조자 · Obj)들의 참조관계에 의한 객체와 객체 사이에 암시적 조인(implicit join)[1]의 의미를 가지고 있다. 이러한 암시적 조인은 데이터베이스 스키마에 의해 미리 예상이 가능하다. 따라서, 질의에 자주 나타나는 중포 속성에 대한 암시적 조인을 미리 계산하여 그 결과를 색인으로 구축하여 놓음으로써, 질의처리시 이를 이용하여 성능 향상을 꾀할 수 있으며 이를 중포 속성에 대한 색인기법[2,3,5,11]이라 한다. 그러나, 중포 속성에 대한 기존의 색인기법들은 B-tree와 같은 일차원 색인구조를 이용함으로써, 클래스 상속에 의한 특징으로 질의의 대상 범위가 클래스 계층상의 임의의 클래스들로 제한되거나, 경로식에 나타나는 속성의 도메인이 클래스 계층상의 임의의 클래스들로 제한이 되는 질의들을 지원하기 어려운 문제점을 가지고 있다.

본 논문에서는 이와 같은 문제점들을 해결하기 위하여, 다차원 화일구조를 이용한 중포된 객체에 대한 다차원 중포 속성 색인기법을

제안한다. 중포 속성에 대한 다차원 색인구조에서는 중포 속성의 키값 도메인과 함께 중포 속성을 표현하는 경로식에 나타나는 모든 속성들에 대해 각 속성이 정의된 클래스 계층마다 한 축의 클래스 식별자 도메인이 할당된 다차원 도메인 공간에 주어진 색인 엔트리(객체 식별자인 Obj로 대치함)들의 클러스터링 문제불 더룬다. 이와 같은 다차원 중포 속성 색인기법에서는 기존의 일차원 색인구조를 이용한 색인기법들에서 문제가 되는 질의의 대상 범위가 클래스 계층상의 임의의 클래스들로 제한되거나, 경로식에 나타나는 속성의 도메인이 클래스 계층상의 임의의 클래스들로 제한이 되는 질의들의 처리를 한번의 색인 탐색으로 가능하게 한다.

### 2. 객체지향 질의어

본 논문에서는 Atkinson et al [1]과 Kim [7] 등이 제안한 보편적인 객체지향 데이터 모델을 사용한다. 그림 1은 클래스 Persons에 대한 스키마 그래프이다. 그림에서 클래스 Persons은 서브 클래스 Employees와 Students 그리고 서브 클래스 Employees와 Students의 서브 클래스들을 포함하는 클래스 상속 계층구조와 Vehicles과 Companies를 포함하는 클래스 집산화 계층구조의 루트이다.

클래스 상속 계층에서 클래스 C의 그의 모든 서브 클래스들을 원소로 하는 집합을 C'로 표기한다. 예를들어 그림 1에서 클래스 Persons'는 집합 {Persons, Employees, Students, Engineers, Computer-eng., Graduates, Under-graduates}이고, Students'는 {Students, Graduates, Under-graduates}이다.

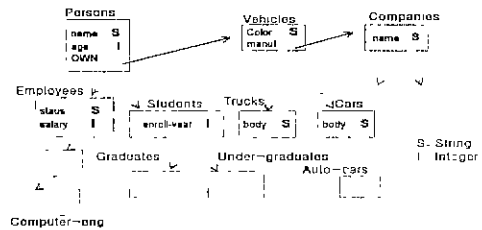


그림 1: 클래스 Persons에 대한 스키마 그래프

객체지향 질의어[6]는 Select, From, Where 절로 구성되며 각 절에서 객체지향 개념을 지원하도록 확장되었다. From 절에는 질의의 대상이 되는 클래스를 기술하며, Where 절에는 단순 속성에 대한 조건인 단순 술어(simple predicate)와 더불어 중포 속성에 대한 조건인 중포 술어(nested predicate)[2]를 사용할 수 있다

예제 1 다음은 그림 1에서 클래스 Persons을 루트로 하는 클래스 계층구조 상의 모든 클래스로부터 P.own.color = "Blue"인 객체의 name을 찾는 질의이다.

```
Select      P.name
From        Persons* P
Where       P.own.color = "Blue".
```

경로식(path expression)[3,5,6]은 객체지향 질의에서 중포 속성을 표현하는 도구이다. 경로식은 클래스 집단체 계층구조상에서 클래스 이름과 속성의 교차적인 나열(sequence)로서 다음과 같은 형태를 가진다 단, A<sub>i</sub> 뒤의 중괄호( )는 선택적임을 나타내는 표시이다.

$$P = C_1.A_1([C_2]).A_2([C_3]) \dots A_n([C_{n+1}]) \quad (1)$$

경로 P에서 클래스 C<sub>i</sub>를 타겟 클래스, 속성 A<sub>i</sub>의 도메인이 되는 C<sub>i+1</sub>을 A<sub>i</sub>의 도메인 클래스라 정의한다. 타겟 클래스와 도메인 클래스는 경로에서 클래스 계층구조에 속하는 특정 클래스로 한정(limit)될 수 있으며[6,7], 이를 클래스 대체(class substitution)라 한다. 이러한 클래스 대체는 질의의 범위를 특정한 클래스로 한정할 수 있도록 하여 클래스 상속의 개념을 객체지향 질의에 표현하도록 한 것이다[6]. 다음 예제 2는 예제 1의 질의로부터 클래스 대체를 보여주고 있다

예제 2 예제 1의 Where 절에 있는 중포 술어로부터 타겟 클래스를 Students\*, Students로 대체하거나, 각각에 대하여 복합 속성 own의 도메인 클래스를 Vehicles\*에 속하는 Cars\*과 Cars로 대체하여 만들어진 중포 술어들이다.

```
Pd1. Students*.own color = "Blue"
Pd2. Students.own color = "Blue"
Pd4. Students.own[Cars*] color = "Blue"
Pd5. Students.own[Cars] color = "Blue"
```

경로식 P에서 경로 인스턴스(path instantiation)는 다음 조건을 만족하는 객체들의 리스트 (O<sub>1</sub>, O<sub>2</sub>, ..., O<sub>n+1</sub>)로 정의한다[2,6] (1) 객체 O<sub>i</sub>은 클래스 C<sub>i</sub>의 객체이다 (2) 객체 O<sub>i</sub> (1 < i <= n+1)는 클래스 C<sub>i</sub>의 객체로서 객체 O<sub>i-1</sub>의 속성 A<sub>i-1</sub>의 값이다. 예를 들어, 경로식 Persons.own[Trucks] manuf.name에서 객체 s ∈ Persons가 속성 own의 값으로 객체 t ∈ Trucks의 Oid를 가지고, 객체 t에서 속성 manuf의 값으로 객체 c ∈ Companies의 Oid를 가지며, 객체 c에서 속성 name의 값으로 hyundai를 가진다면, 객체들의 나열인 (s, t, c, hyundai)는 경로식 Persons.own[Trucks] manuf.name에 대한 경로 인스턴스가 된다.

### 3. 중포된 객체에 대한 질의처리를 위한 기존의 색인기법

지금까지 중포 속성에 대한 색인기법을 위해 제안된 색인구조들로 중포 색인(nested index)[2], 경로 색인(path index)[2], ASR(Access Support Relation)[5], 그리고 JIH(Join Index Hierarchy)[11] 등이 있다. 그러나, 이러한 색인구조들은 클래스 상속에 의한 객체지향 데이터 모델의 특징을 반영하지 못하는 것들로서 타겟 클래스의 대체 또는 도메인 클래스의 대체가 있는 경로식으로 표현된 질의는 지원하지 못한다 즉, 질의의 내상 범위가 클래스 상속 계층상의 임의의 클래스들로 제한되거나, 경로식에 나타나는 어떠한 속성의 도메인이 클래스 상속 계층상의 임의의 클래스들로 제한이 되는 질의들을 지원하지 못한다

Bertino[4]는 질의의 대상 범위가 임의의 클래스들로 제한되는 경로식으로 표현된 중포 속성에 대한 질의를 지원할 수 있는 NIX(Nested-Inherited Index)라고 하는 색인구조를 제안하였다. 그러나, NIX는 색인된 중포 속성으로부터 각 타겟 클래스까지의 경로 정보를 유지하지 않기 때문에 도메인 클래스 대체가 있는 경로식을 가지는 질의어를 지원하지는 못한다 즉, 질의의 대상 범위가 클래스 계층상의 임의의 클래스들로 제한되는 질의들은 지원하지 못하나, 색인된 중포 속성을 표현하는 경로식에 나타나는 어떠한 속성의 도메인이 클래스 계층상의 임의의 클래스들로 제한이 되는 질의들은

지원하지 못하는 색인구조이다

### 4. 다차원 중포 속성 색인기법

그림 2는 객체지향 데이터베이스의 스키마 그래프에서 세 개의 클래스 계층으로 이루어진 하나의 경로를 나타내는 경로 스키마이다. 그림 2의 경로 스키마에서, 클래스 계층 A에서 정의된 복합 속성 a의 도메인은 클래스 계층 B이며, 클래스 계층 B에서 정의된 단순 속성 b는 클래스 계층 A의 중포 속성이다 그리고, A'와 B'는 각각 클래스 계층 A와 B의 서브클래스 계층을 나타낸다

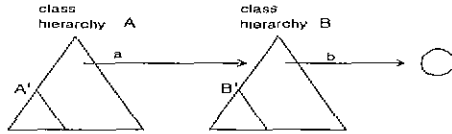


그림 2. 객체지향 데이터베이스의 경로 스키마.

그림 4와 같은 경로 스키마에 대하여 중포 속성 b에 조건이 주어진 중포 술어에는 중포 속성을 표현하는 경로식의 특성에 따라 다음과 같은 세가지 형태로 분류할 수 있다

- 중포 술어1 : A.a b θ Val
- 중포 술어2 : A'.a b θ Val
- 중포 술어3 : A'.a[B']b θ Val

이들은 모두 중포 속성 b에 조건이 주어진 술어들이지만, 각 중포 술어를 만족하는 결과는 서로 다르게 된다 즉, 중포 술어1은 클래스 계층 A의 모든 클래스를 질의의 대상으로 중포 속성 b에 조건이 주어진 술어이고, 중포 술어2는 중포 술어1에서 질의의 대상을 클래스 계층 A에서 A'에 속하는 클래스들만으로 한정된 것이다 그리고, 중포 술어3은 중포 술어에서 복합 속성 a의 도메인으로 클래스 계층 B에서 B'에 속하는 클래스들로 한정된 것이다 중포 술어2에서 사용된 경로식에는 타겟 클래스 대체가 있는 경우이며, 중포 술어3에서 사용된 경로식에는 타겟 클래스 대체와 도메인 클래스 대체가 있는 경우이다

다차원 중포 속성 색인기법에서는 (경로길이+1) 차원의 색인구조를 이용하여, 중포 속성을 표현하는 경로상의 각 클래스 계층마다 클래스 식별자들로 구성된 한 차원씩의 클래스 식별자 도메인을 할당한다 예를 들어, 그림 2와 같은 경로 스키마에서 중포 속성 b에 대한 색인구조로서 삼차원 색인구조를 이용하여 다음과 같은 삼차원 도메인 공간을 구성할 수 있다. 그림 3은 이와 같이 구성된 삼차원 도메인 공간상에서 중포 술어3에 대한 질의 영역을 표현한 것이다

- 축1 : 중포 속성 b의 키값 도메인
- 축2 : 클래스 계층 A의 클래스 식별자 도메인
- 축3 : 클래스 계층 B의 클래스 식별자 도메인

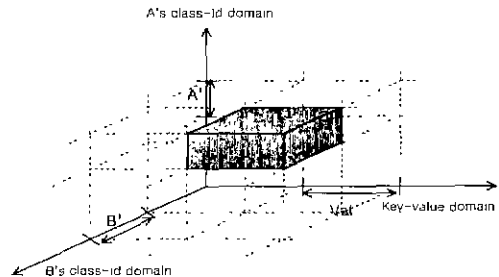


그림 3' 중포 술어3에 대한 질의 영역

본 논문에서는 중포 속성에 대한 색인구조로 다차원 화인구조의 하나인 계층 그리드 화일[10]을 이용하여, 색인 엔트리를 구성하는 방법에 따라 확장된 중포 색인구조와 확장된 경로 색인구조로 분류한다. 확장된 중포 색인구조(extended nested index)는 색인 엔트리를 색인된 중포 속성의 타겟 클래스 계층에 속하는 객체에 대한 객체 식별자(즉, Oid)들로 구성하며, 확장된 경로 색인구조(extended path index)는 색인 엔트리를 색인된 중포 속성에 대한 경로 인스턴

스(즉, Old 리스트)들로 구성한다. 그림 4는 확장된 중포 색인구조의 색인 페이지 구조를 나타내며, 그림 5는 확장된 경로 색인구조의 색인 페이지 구조를 나타낸다.

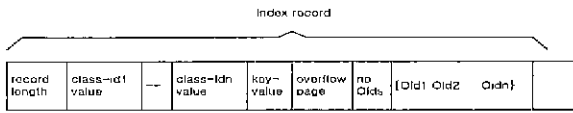


그림 4 확장된 중포 색인구조의 색인 페이지 구조

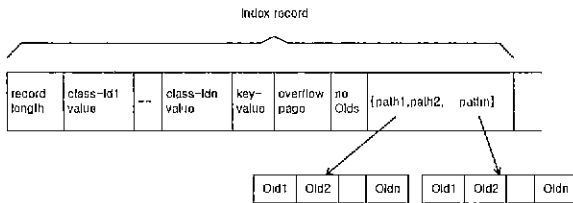


그림 5 확장된 경로 색인구조의 색인 페이지 구조

중포 속성에 대한 색인기법은 경로상의 참조관계에 의한 암시적 조인의 연산을 미리 계산하여 두는 기법으로 중포 술어를 만족하는 객체들의 탐색에는 매우 유용하지만, 상대적으로 색인구조의 유지 비용을 많이 필요로 한다[2]. 이러한 색인구조의 유지 비용은 색인을 유지하는 경로의 길이에 비례하기 때문에, 경로의 길이가 매우 길 때는 경로를 여러개의 서브경로로 나누어서, 서브경로별로 여러개의 색인구조를 유지함으로써 유지 비용을 줄일 수 있다. 특히, 길이가 1인 서브경로에 대한 색인구조는 참고 문헌[9]에서 제안한 이차원 클래스 계층 색인구조가 된다. 즉, 경로상의 클래스 계층마다 이차원 클래스 색인구조를 단계적으로 구축함으로써 중포 술어를 만족하는 다중 클래스 계층의 객체들을 여러번의 단계적인 색인 검색으로 구할 수 있다. 우리는 이와같이 경로상의 클래스 계층마다 구축하는 색인구조를 **다중 이차원 클래스 색인구조(multiple 2-dimensional class index)**라 정의한다.

### 5. 성능 평가

먼저, 본 논문에서 제안한 각 색인구조의 검색 성능에 대하여 살펴보면, 색인 엔트리를 색인된 중포 속성의 경로 인스턴스로 구성하는 확장된 중포 색인구조가 색인 엔트리를 타겟 클래스 계층에 속하는 객체 식별자만으로 구성하는 확장된 경로 색인구조에 비하여 저장 공간의 오버헤드로 인하여 검색 성능이 떨어지게 되고, 경로상의 클래스 계층마다 이차원 클래스 색인구조를 단계적으로 구축함으로써 중포 속성을 표현하는 경로길이 만큼의 단계적 색인검색이 필요한 다중 이차원 클래스 색인구조가 가장 나쁘게 된다. 즉, 확장된 중포 색인구조가 가장 좋은 검색 성능을 가지게 되고, 확장된 경로 색인구조가 그 다음으로 좋은 검색 성능을 가지게 된다.

그러나, 각 색인구조의 운영에 따른 유지 비용에 대한 오버헤드는 색인된 중포 속성의 경로 길이에 따라 많은 차이를 보이게 된다. 다음은 각 색인구조의 유지 비용에 대한 분석 결과로서, 색인구조를 구축된 경로의 길이에 따라 가장 적합한 색인구조를 선택하는 기준이 된다.

첫째, 색인구조를 구축할 경로의 길이가 2인 경우에는, 확장된 중포 색인구조가 가장 적합한 색인구조이다. 이것은 경로의 길이가 2인 경우, 유지 비용은 세가지 색인구조 모두 비슷한 반면에 확장된 중포 색인구조에서 검색 성능이 가장 좋기 때문이다.

둘째, 색인구조를 구축할 경로의 길이가 3인 경우에는, 일반적으로 확장된 경로 색인구조가 가장 적합한 색인구조이다. 이것은 확장된 경로 색인구조의 검색 성능은 확장된 중포 색인구조에 필적하는 반면에 데이터베이스 변경에 의한 유지 비용이 적게 되기 때문이다. 그러나, 데이터베이스의 변경이 첫번째 클래스 계층과 두번째 클래스 계층에서 주로 발생하고, 각 클래스 계층의 객체 참조 공유도의 값이 크게 되면 확장된 중포 색인구조가 가장 적합한 색인구조이다. 이것은 데이터베이스의 변경이 첫번째 클래스 계층과 두번째 클래스 계층에서 발생할 경우에는 확장된 중포 색인구조에서도 갱신을 위한 역방향 운영이 필요없게 되므로 색인구조의 유지 비용이 확장된 경

로 색인구조에서와 같게 되기 때문이다.

셋째, 색인구조를 구축할 경로의 길이가 4이상일 경우에는, 경로로 길이가 1, 2 또는 3이 되는 서브경로들로 분할한 다음에, 각 서브경로에 따라 확장된 중포 색인구조 또는 확장된 경로 색인구조를 할당하는 정책은 사용하는 것이 좋다. 이것은 객체 참조 공유도가 매우 낮은 경우(거의 1일때)가 아니면 확장된 중포 색인구조와 확장된 경로 색인구조 모두 사용하기가 어렵기 때문이다.

### 6 결론

본 논문에서는 객체지향 데이터베이스의 중포된 객체의 중포 속성에 대한 색인기법으로 다차원 화일구조를 이용하는 방향에 관하여 논의하였다. 중포 속성에 대한 기존의 색인기법에서는 B<sup>+</sup>-tree와 같은 인차원 색인구조를 이용함으로써, 객체지향 데이터베이스를 구성하는 여러 클래스 계층과 경로식으로 표현된 복잡한 형태의 질의들의 처리를 지원하지 못하는 문제의 한계가 있었다. 제안된 다차원 중포 속성 색인기법은 기존의 색인기법에서 지원하기 어려운 질의의 대상 범위가 클래스 계층상의 임의의 클래스들로 제한되거나, 질의에 포함된 복합 속성들의 도메인이 클래스 계층상의 임의의 클래스들로 제한되는 경우에도 효율적으로 지원될 수 있음을 보였다.

또한, 본 논문에서는 중포 속성에 대한 다차원 색인구조로 다차원 화일구조의 하나인 계층 그리드 화일을 이용하여, 색인 엔트리를 색인된 중포 속성의 타겟 클래스 계층의 객체 식별자(즉, Old)들로 구성하는 확장된 중포 색인구조와 색인 엔트리를 색인된 중포 속성에 대한 경로 인스턴스(즉, Old 리스트)들로 구성하는 확장된 경로 색인구조를 제안하고, 각 색인구조의 운영에 따른 성능을 비교 평가 하였다. 그 결과 경로의 길이가 4이상인 경우에 경로의 길이가 1, 2, 또는 3이 되는 서브경로로 나누어서 각 서브경로별로 적합한 색인구조를 할당해야 함을 알 수 있었다. 그러나, 이러한 색인구조의 할당 문제는 데이터베이스의 특성, 작업 부하(work load), 그리고 각 색인구조의 검색 비용과 유지 비용등을 고려하는 매우 복잡한 문제로서 앞으로 더 연구해야 할 연구과제이다.

### 참고 문헌

- [1] Atkinson, M et al, "The Object-Oriented Database System Manifesto," In *Proc. Intl. Conf. on Deductive and Object-Oriented Databases*, pp 40--57, Kyoto, Japan, Dec. 1989.
- [2] Bertino, E and Kim, W, "Indexing Techniques for Queries on Nested Objects," *IEEE Trans on Knowledge and Data Eng.*, Vol. 1, No 2, pp 196--214, June 1989.
- [3] Bertino, E. et al., "Object-Oriented Query Languages: The Notion and the Issues," *IEEE Trans on Knowledge and Data Engineering*, Vol. 1, No 3, pp 223--237, June 1989.
- [4] Bertino, E. and Foscoli, P, "Index Organizations for Object-Oriented Database Systems," *IEEE Trans. on Knowledge and Data Eng.*, Vol. 7, No. 2, pp. 193--209, April 1995
- [5] Kemper, A. and Moerkotte, G., "Access Support Relations: An Indexing Method for Object Bases," *Information Systems*, Vol 17, No. 2, pp 117--145, 1992
- [6] Kifer, M., Kim, W., and Sagiv, Y., "Querying Object-Oriented Databases," In *Proc. Intl. Conf. on Management of Data*, ACM SIGMOD, San Diego, Calif., pp 393--402, May 1992.
- [7] Kim, W., "Object-Oriented Database Systems Promises, Reality, and Futures," *Modern Database Systems: The Object Model, Interoperability, and Beyond*, Kim, W ed, pp. 255-280, Addison-Wesley, 1995
- [8] Lee, J H et al, "A Physical Database Design Method for Multidimensional File Organizations," *International Journal on Information Sciences*, Vol 102, No. 3, pp 32-65, 1997
- [9] 이 종학, 황 규영, "다차원 색인구조를 이용한 객체지향 데이터베이스의 클래스 계층 색인기법," *한국정보과학회 논문지(B)*, 제 26권, 제 3호, pp 365-379, 1999. 3.
- [10] Whang, K Y and Krishnamurthy, R, Multilevel Grid Files, IBM Research Report RC11516, IBM Thomas J Watson Research Center, Nov 1985
- [11] Xie, Z. and Han, J, "Join Index Hierarchies for Supporting Efficient Navigations in Object-Oriented Databases," In *Proc Intl. Conf. on Very Large Data Bases* pp. 522-533, Santiago, Chile, Sept 1994