韓國經營科學會誌
第23卷 第2號
1998年 6月

지식 추상화 계층의 구축과 관리

허순영* · 문개현*

Management of Knowledge Abstraction Hierarchy

Soon-Young Huh* · Kae-Hyun Moon*

Abstract

Cooperative query answering is a research effort to develop a fault-tolerant and intelligent database system using the semantic knowledge base constructed from the underlying database. Such knowledge base has two aspects of usage. One is supporting the cooperative query answering process for providing both an exact answer and neighborhood information relevant to a query. The other is supporting ongoing maintenance of the knowledge base for accommodating the changes in the knowledge content and database usage purpose. Existing studies have mostly focused on the cooperative query answering process but paid little attention on the dynamic knowledge base maintenance.

This paper proposes a multi-level knowledge representation framework called Knowledge Abstraction Hierarchy (KAH) that can not only support cooperative query answering but also permit dynamic knowledge maintenance. The KAH consists of two types of knowledge abstraction hierarchies. The value abstraction hierarchy is constructed by abstract values that are hierarchically derived from specific data values in the underlying database on the basis of generalization and specialization relationships. The domain abstraction hierarchy is built on the various domains of the data values and incorporates the classification relationship between super-domains and sub-domains. On the basis of the KAH, a knowledge abstraction database is constructed on the relational data model and accommodates diverse knowledge maintenance needs and flexibly facilitates cooperative query answering. In terms of the knowledge maintenance, database operations are discussed for the cases where either the internal contents for a given KAH change or the structures of the KAH itself change. In terms of cooperative query answering,

* 한국과학기술원 테크노경영대학원

database operations are discussed for both the generalization and specialization processes, and the conceptual query handling. A prototype system has been implemented at KAIST that demonstrates the usefulness of KAH in ordinary database application systems.

1. 서 론

기존의 데이터베이스 시스템은 스키마의 관점에서 정확하게 표현되지 못한 질의나 요구사항이 모호하게 표현된 질의는 처리할 수 없으며, 또한 질의 조건을 정확하게 만족하는 답만을 산출해 주기 때문에 정확한 답이 존재하지 않는 경우 아무런 정보도 제공해 주지 못한다. 따라서 사용자들은 자신이 원하는 답을 얻기 위해서는 데이터베이스 스키마와 문제 영역에 관한 지식을 갖고 있어야만 한다. 그러한 문제점을 해결하기 위한 노력으로 협력적 질의응답(cooperative query answering)에 관한 연구가 수행되어져 왔다[6, 8, 9, 16, 18, 19, 24]. 협력적 질의응답은 질의의 의도를 분석하고 데이터베이스에 내포된 여러가지 의미론적 지식을 활용하여, 정확한 답은 물론 근사적인 답이나 질의와 관련된 유용한 정보를 제공하는 지능적인 시스템을 개발하는데 초점을 맞추고 있다.

협력적 질의응답을 통해 이러한 추가적인 정보를 얻기위한 질의처리의 핵심 과정은 질의 변환이라고 할 수 있는데, 이는 질의의 조건을 완화함으로써 이루어 진다. 질의 완화란 하위 개념 수준의 원시 데이터와 상위 개념 수준의 추상적 개념들 사이의 의미론적 관계를 이용하여 질의의 탐색 영역을 확장하는 것을 말한다. 따라서, 협력적 질의응답에서는 데이터의 의미론을 끌어낼 수 있게 해주는 지식 표현 프레임웍의 확립이 매우 중요하다고 할 수 있다. 지금까지 제시되어온 다양한 지식 표현 프레임웍들은 의미론적 거리, 규칙(rule), 개념적 분류에 기초한 기법들로 구분되어질 수 있다.

의미론적 거리 기법은 한 쌍의 데이터 개체간의 유사성 정도를 수치화된 거리 개념으로 나타낸다[14, 15, 17]. 이 기법의 장점은 일단 데이터 개체들간의 거리가 결정되면, 질의 완화 알고리즘을 쉽고 효율적으로 개발할 수 있다는 것이다. 하지만 양적/질적 데이터 모두의 유사성을 수량화된 하나의 일률적인 척도로써 나타내기 힘들고, 다양한 데이터 개체간의 의미론적 유사성을 평가하는 객관적인 기준이 존재하지 않는다는 한계점을 갖는다.

규칙 기반 기법[3, 9, 11]은 일차형 논리를 형식상의 프레임웍으로 사용하여 데이터의 의미론적 정보들과 무결성 제약을 표현한다. 이 기법에서는 기반 데이터베이스가 기본형 술어의 집합으로 이해되고, 데이터베이스에 대한 질의문은 찾고자 하는 정보들이 자유 변수로서 명세되는 술어 규칙으로 표현된다. 질의 처리는 해결(resolution)과 추론 메카니즘을 통해서 이루어 지는데, 이때 필요한 질의 완화는 무결성 제약들을 조정함으로써 이루어 진다. 규칙 기반 기법의 한계점은 질의 완화 과정을 유도해 나갈 체계적인 시스템이 부족하고 질의 응답 과정의 직관성이 떨어진다는 것이다.

개념적 분류 기법은 서로 유사한 원시 데이터 개체들을 하나의 추상적인 클래스로 묶고,

각 클래스들이 다른 클래스들과 부모/자식 관계를 갖도록 한다[1, 2, 6, 8, 21, 24]. 따라서 동일한 클래스에 속한 데이터 개체들은 서로 유사한 개체들에 해당하고, 부모 클래스에 속한 개체들은 자식 클래스에 속한 개체들보다 더욱 일반화된 개체들로 여겨진다. 개념적 분류 기법은 특히 질적이고 분류 가능한 데이터 개체들에 적용할 경우에 장점을 갖는다. 또한 이 기법은 인간의 사고 체계와 유사하고 많은 조직에서 데이터의 의미론을 관리하기 위한 방법으로 이미 사용하고 있기 때문에, 데이터 개체들간의 유사성 관리를 위하여 쉽게 구현 되어질 수 있다.

이러한 기존의 연구들은 각각 데이터의 의미론적 지식을 표현하기 위한 프레임웍과 그를 바탕으로 한 협력적 질의처리 과정을 제시하고 있지만, 지속적인 지식의 변화를 효과적으로 처리해 줄 수 있는 지식 관리 측면을 고려하지 못하고 있다. 결과적으로 지식 표현 프레임웍에 의해 표현된 지식의 변화를 동적으로 처리해 주는데 많은 문제점을 갖게 되고, 지식을 실제로 관리하게 되는 데이터베이스를 확장 시키기도 힘들게 된다. 이 논문은 이러한 문제점을 해결하기 위하여, 지식 추상화 계층(KAH : Knowledge Abstraction Hierarchy)이라는 지식 표현 프레임웍을 제시한다. KAH는 데이터베이스에 대한 의미론적 지식과 실제 데이터 값을 통합하여 협력적 질의응답을 수행할 수 있게 해줄 뿐 아니라, 동적인 지식 관리를 용이하게 해 주는 효과적이고 직관적인 지식 표현 프레임웍이라고 할 수 있다.

이 논문의 구성은 다음과 같다. 2장에서는 지식 추상화 계층 KAH를 값의 추상화와 도메인의 추상화라는 두개의 추상화 관점에서 소개한다. 3장에서는 KAH에 포함된 지식을 실제적으로 관리하게 되는 추상화 정보 데이터베이스를 구축한다. 4장은 지식 관리와 협력적 질의 응답을 수행하는데 필요한 추상화 정보 데이터베이스 연산들을 정의한다. 마지막으로 5장에 결론을 제시한다.

2. 지식 추상화 계층

협력적 질의 처리는 질의의 의도를 분석하고 질의 조건을 완화함으로써 좀 더 넓은 탐색 영역을 갖는 새로운 질의로 변환하는 과정을 포함한다. 질의 변환을 수행하기 위해서는 질의문에 나타나는 여러가지 데이터 값들을 그와 관련된 다른 값들로 치환하는 것이 필요하게 되고, 이때 일반화(generalization)/세분화(specialization)와 같은 데이터 값들 사이의 의미론적 관계를 이용하게 된다. 데이터 추상화(abstraction)는 데이터 값들 사이의 이러한 의미론적 관계를 파악하는 효과적인 방법으로 인식되어 왔다[23]. 이 논문은 데이터 추상화를 사용하여 데이터 값들 사이의 의미론적 관계를 표현하는 프레임웍으로서 KAH를 제시한다.

2.1 지식 추상화 계층의 구성

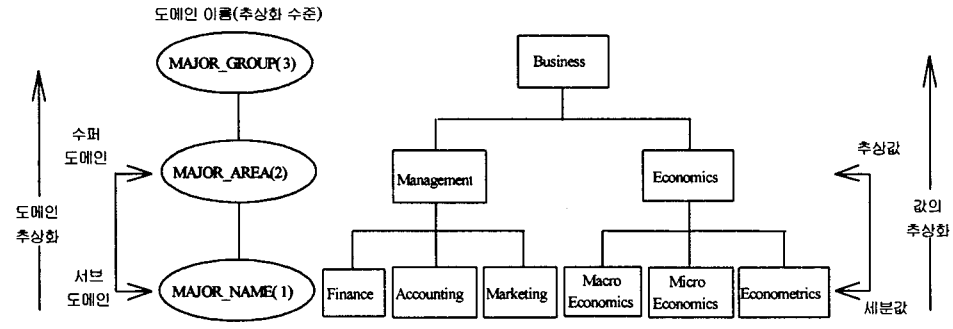
의미론적 데이터 모델링 기법에서 데이터 추상화는 특정 데이터 개체의 세부 사항 집합을 개념적인 수준에서 기술함으로써, 전체적으로 인식되고 명명될 수 있게 해주는 기본 개념으로 사용되어 진다. 그러한 데이터 추상화는 보통 구체적인 세부 사항들을 은폐함으로써 추상화된 개체와 그리고 그들 사이의 관계에 초점을 맞추게 된다. 따라서 데이터 추상화는 질의 변환을 수행하기 위하여 데이터 값들을 서로 연관 시키

는데 있어서 매우 유용하게 사용될 수 있다.

지식 추상화 계층은 하나의 지식 표현 프레임웍으로서, 데이터 추상화를 기반으로 하여 데이터 값과 데이터베이스에 대한 의미론적 지식을 다단계의 개념적 수준에서 서로 다른 형태로 표현할 수 있게 해준다. 예를 들어 그림 1의 대학 전공에 관한 지식 추상화 계층을 살펴보자. 그림 1은 대학 전공들을 전공 이름(MAJOR_NAME), 전공 분야(MAJOR_REA), 전공 그룹(MAJOR_GROUP)등의 다단계 추상화 수준으로 분류한다. 그러한 전공 이름, 영역, 그룹에 포함되는 값들은 기반 데이터베이스에 실제로 존재하는 값일 수도 있고, 존재하지는 않지만 전공에 관한 지식을 관리하기 위하여 인위적으로 추가된 값일 수도 있다. 이러한 상태에서 일반화와 분류(classification)를 통하여 추상화 계층이 구축되어 진다.

먼저, 일반화 관계는 하나의 계층 내에서 서로 유사한 데이터 값들을 상위 개념 수준의 추상값과 연계시키기 위하여 사용된다. 예를 들어, Finance, Accounting, Marketing과 같은 전공 이름들은 그들의 추상값이면서 전공 영역에 포함되는 값인 Management와 관계를 갖게 된다. 마찬가지로 Macro Economics, Micro Economics, Econometrics 역시 전공 영역에 존재하는 그들의 추상값인 Economics와 관계를 갖게 된다. 추상값인 Management와 Economics는 다시 더욱 추상화 되어 한 단계 높은 수준에 존재하는 Business와도 관계를 갖는다. 이처럼 한 계층내에서의 상위 수준은 하위 수준보다 좀 더 추상화된 데이터 표현을 가능하게 함으로써, 상위 수준에 존재하는 값은 하위 수준에 존재하는 값의 추상값으로 이해되고, 하위 수준에 존재하는 값은 상위 수준에 존재하는 값의 세분값으로 이해된다. 그리고 세분값과 그들의 추상값 사이에는 IS-A 관계가 존재하게 된다. 예를 들어 Finance는 Management이고, Management는 Business이다. 이러한 일반화 관계는 단지 데이터 값만을 고려할 뿐 그들이 속해있는 도메인은 고려하지 않기 때문에, 우리는 이와 같은 데이터 추상화를 값의 추상화라고 부르기로 한다.

다음으로, 데이터 값들과 그들이 속해있는 도메인과의 관계를 정의하기 위하여 분류(classification) 관계가 데이터 추상화의 또 다른 형태로서 사용되어 진다. 그 결과, 데이터베이스의 특정 데이터 인스턴스 값과 자신의 도메인 사이에는 INSTANCE-OF 관계가 존재하게 된다. 예를 들어, Finance, Accounting, Marketing, Macro Economics와 같이 1단계에 존재하는 값들은 MAJOR_NAME이라는 도메



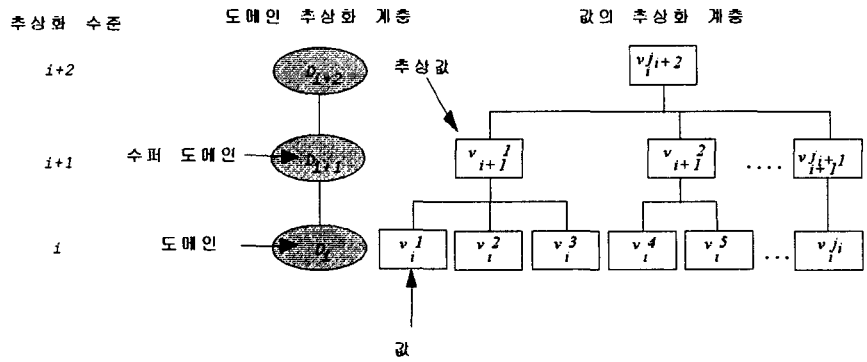
[그림 1] 대학 전공에 관한 지식 추상화 계층

인의 인스턴스들에 해당한다. 마찬가지로, Management, Economics와 같이 2단계에 존재하는 추상값들은 도메인 MAJOR_AREA의 인스턴스들이 된다. 값의 추상화 계층에서 두 개의 인접한 추상화 수준에 존재하는 값들 사이에 일반화 관계가 존재하는 것처럼, 두 개의 인접한 도메인 사이에는 수퍼 도메인/서브 도메인의 관계가 존재한다. 값의 추상화와 대응하여 우리는 그러한 관계를 도메인 추상화라고 부르기로 한다. 도메인 추상화 관계에서 수퍼 도메인은 서브 도메인보다 더욱 추상화된 도메인에 해당한다. 예를 들어, MAJOR_AREA는 MAJOR_NAME의 수퍼 도메인이므로 MAJOR_NAME보다 더욱 추상화된 도메인이다. 이것은 값의 추상화 관점에서 볼 때, MAJOR_NAME의 인스턴스 값의 추상값 모두는 MAJOR_AREA의 인스턴스 값이 됨을 의미한다. 즉, MAJOR_NAME에 포함된 Finance의 추상값은 MAJOR_AREA 도메인의 인스턴스 값중의 하나인 Management가 된다.

지식 추상화 계층은 이러한 값의 추상화 계층과 도메인 추상화 계층의 두 가지 종류의 추상화 계층으로 구성되어 진다. 값의 추상화 계층에서 한 값 v_2가 다른 값 v_1 보다 좀더 일반적인 값일 경우, v_2는 계층 내에서 $v1$보다 더 높은 수준에 위치하게 된다. 따라서 최상위 수준의 추상값은 계층 내에서 가장 일반적인 값인 반면, 최하위 수준의 값들은 가장 세분화된 값들이다. 전체 추상화 수준의 깊이는 사용자가 문제를 적절하게 묘사하기 위해서 어느 정도의 추상화가 필요한지를 판단하여 자유롭게 결정할 수 있다. 한편, 값의 추상화 계층에서 추상값과 그것의 세분값들 사이의 사상 원소수(mapping cardinality)는 일대다(one-to-many)인 것으로 가정한다. 즉, 하나의 추상값은 한 단계 낮은 추상화 수준에 존재하는 다수의 세분값들을 갖게 된다. 반대로, 한 세분값은 한 단계 높은 추상화 수준에 단지 하나의 추상값만을 갖게 된다. 하지만 값의 추상화의 반복적 특성으로 인하여, 한 세분값은 자신으로부터 최상위 수준까지, 각각 서로 다른 추상화 수준에 존재하는 다수의 추상값을 갖게 된다.

도메인 추상화 계층에서 하나의 도메인은 자신의 고유 이름과 계층 내에서의 절대적인 위치를 갖는다. 또한 그 도메인 이름에 따라서 그에 대응하여 값의 추상화 계층에 존재하는 값들의 추상화 수준이 유일하게 결정된다. 따라서, 값의 추상화 계층에서 한 값의 추상화 수준은 다른 값의 추상화 수준과의 관계에 의해서 결정되는 상대적인 개념이지만, 도메인 추상화 계층의 도메인과 연계되어져 비로소 절대적인 수준을 갖게 된다. 도메인 계층에서 한 도메인 D_2가 수퍼 도메인이고 D_1가 서브 도메인일 경우, D_2는 계층 내에서 D_1보다 더 상위 수준에 위치하게 된다. 그리고 수퍼 도메인과 서브 도메인의 사상 원소수 관계는 일대일(one-to-one)인 것으로 가정한다. 따라서 한 서브 도메인의 바로 윗 수준 수퍼 도메인은 반드시 하나만 존재하게 되고 그 역도 마찬가지이다.

결국 KAH는 값의 추상화와 도메인의 추상화를 바탕으로, 기반 데이터베이스에 존재하는 데이터 값들과 그들의 도메인들 사이의 의미론적 관계를 다단계 계층의 형태로 표현할 수 있게 해주는 지식 표현 프레임웍이라고 할 수 있다. KAH는 개별 응용 문제와 이용 관점에 따라 그 구체적인 내용이 달라지게 되므로, 하나의 데이터베이스에 대하여 다수의 KAH 인스턴스들이 존재할 수 있다.



[그림 2] 지식 추상화 계층의 형식화

2.2 지식 추상화 계층의 형식화

값의 추상화 계층과 도메인 추상화 계층에 대한 형식론은 그림 2와 같다. 그림에서 도메인과 값은 각각 타원과 사각형으로 나타내여 진다. 첨자 i는 추상화 수준을 나타내고 j_i는 추상화 수준 i에 존재하는 값들의 무작위 순번을 나타낸다. 따라서 D_i는 추상화 수준 i에 존재하는 도메인을 나타내는 반면, $v_i^{j_i}$는 도메인 D_i에 존재하는 j_i번째 값을 나타낸다.

도메인 D_i와 그것의 수퍼 도메인 D_{i+1}사이의 도메인 추상화 관계는 $D_i \Rightarrow D_{i+1}$으로 표현하고, 한 값 $v_i^{j_i}$와 그것의 추상값 $v_{i+1}^{j_{i+1}}$사이의 값의 추상화 관계는 $v_i^{j_i} \in^* v_{i+1}^{j_{i+1}}$처럼 표현한다. 예를 들어 그림 1의 도메인 추상화 계층은 MAJOR-NAME ⇒ MAJOR_AREA ⇒ MAJOR_GROUP과 같이 표현되고, Finance, Management, Business사이의 값의 추상화 관계는 Finance $\in^*$ Management $\in^*$ Business처럼 표현된다.

Finance $\in^*$ Management $\in^*$ Business rhksrP에서 나타난 것처럼, Finance와 같은 하나의 세분값은 서로 다른 추상화 수준에 존재하는 Management, Business등의 다수의 추상값을 갖을 수 있다. 만약, $v_{i+n}^{j_{i+n}}$이 $v_i^{j_i}$의 한 추상값이라면 그들 사이의 추상화 수준의 차이는 n이 된다. 이 때, 추상값 $v_{i+n}^{j_{i+n}}$을 $v_i^{j_i}$의 n단계 추상값으로 부르고 그들의 추상화 수준의 차이를 추상화의 정도라 부르기로 한다. 한 값 $v_i^{j_i}$와 그의 n단계 추상값 $v_{i+n}^{j_{i+n}}$ 사이의 값의 추상화 관계는 다음의 [식 1]과 같이 표현된다.

$$v_i^{j_i} \in^{n*} v_{i+n}^{j_{i+n}} \text{ iff } \exists v_{i+1}^{j_{i+1}}, \ldots, \exists v_{i+n-1}^{j_{i+n-1}}$$
$$s.t\, v_i^{j_i} \in^* v_{i+1}^{j_{i+1}} \in^* v_{i+2}^{j_{i+2}} \in^* \cdots v_{i+n-1}^{j_{i+n-1}} \in^* v_{i+n}^{j_{i+n}} \text{ ---[식 1]}$$

[식 1]에서 $\in^*$ 기호는 1단계 값의 추상화 관계 즉, $\overset{1}{\in^*}$를 나타낸다. 따라서, 그림 1에서의 Finance와 Business사이의 관계는 Finance $\overset{2}{\in^*}$ Business처럼 나타낼 수 있고, 이는 Business가 Finance의 2단계 추상값이라는 사실을 의미한다. 이때, 추상값들 사이에는 대수적 가법성이 존재하게 된다. $v_{i+n}^{j_{i+n}}$이 $v_i^{j_i}$의 n단계 추상값이고 $v_{i+n+m}^{j_{i+n+m}}$이 $v_{i+n}^{j_{i+n}}$의 m단계 추상값인 경우, $v_{i+n+m}^{j_{i+n+m}}$

은 $v_i^{j_i}$ 의 $n+m$단계 추상값이 된다. 다시 말해서, $v_i^{j_i} \overset{n}{\in^*} v_{i+n}^{j_{i+n}}$ 이고 $v_{i+n}^{j_{i+n}} \overset{m}{\in^*} v_{i+n+m}^{j_{i+n+m}}$ 이면, $v_i^{j_i} \overset{n+m}{\in^*} v_{i+n+m}^{j_{i+n+m}}$으로 표현할 수 있다.

한편, 도메인 추상화 계층에서는, n단계 수퍼 도메인을 생각할 수 있다. D_{i+n}이 D_i의 n단계 수퍼 도메인인 경우, 그들의 추상화 수준 사이의 차이는 n이 되고 다음의 [식 2]와 같이 정의할 수 있다.

$$D_i \overset{n}{\Rightarrow} D_{i+n} s.t v_i^{j_i} \overset{n}{\in^*} v_{i+n}^{j_{i+n}}, v_i^{j_i} \in D_i, v_{i+1}^{j_{i+1}} \in D_{i+1}, \ldots,$$

$$v_{i+n}^{j_{i+n}} \in D_{i+n} \text{ for all } j_i \quad \text{[식 2]}$$

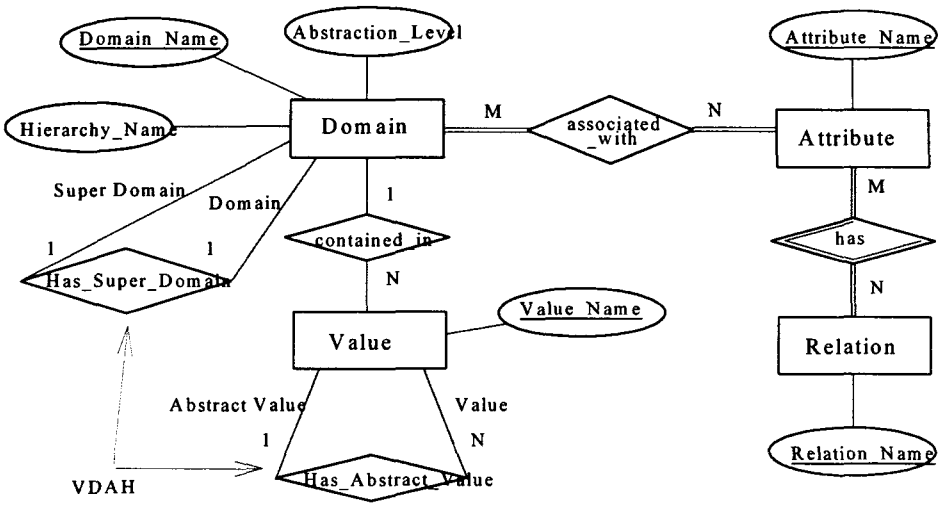
[식 2]에서 는 1단계 도메인 추상화 즉 $\overset{n}{\Rightarrow}$의 특별한 경우로서 n=1일 경우를 나타낸다. 그림 1의 MAJOR_NAME과 MAJOR_GROUP사이의 관계는 MAJOR_NAME $\overset{2}{\Rightarrow}$ MAJOR_ GROUP처럼 표현할 수 있다. 값의 추상화 계층과 마찬가지로, 도메인 추상화 계층의 도메인들 사이에도 추상화의 가법성이 존재한다. 다시 말해서 $D_i \overset{n}{\Rightarrow} D_{i+n}$이고 $D_{i+n} \overset{n}{\Rightarrow} D_{i+n+n}$이면 $D_i \overset{n+m}{\Rightarrow} D_{i+n+m}$과 같이 표현될 수 있다.

3. 추상화 정보 데이터베이스

KAH가 표현하는 값과 도메인에 관한 의미론적 지식은 협력적 질의처리를 수행하는데 활용될 수 있으며, 이를 위하여 이 논문에서는 관계형 데이터 모델을 바탕으로 추상화 정보 데이터베이스를 구축한다. 추상화 정보 데이터베이스란 KAH에 포함된 모든 의미론적 지식을 관리하여 협력적 질의 처리 과정을 용이하게 수행할 수 있게 해주는 지식베이스라고 할 수 있다. 그러한 추상화 정보 데이터베이스의 설계시 다음의 세 가지 이슈들이 고려되어야 한다. 첫째, 추상화 정보 데이터베이스는 기반 데이터베이스의 변화로 인한 추상화 지식의 변화를 동적으로 처리 가능함으로써 효과적으로 유지, 관리될 수 있어야 한다. 다시 말해서 기반 데이터베이스와 관련된 지식의 추가, 삭제, 갱신 등을 용이하게 수행할 수 있어야 한다. 둘째, 관계형 데이터 모델을 이용하여 추상화 정보 데이터베이스를 구축함으로써, 기반 데이터베이스와 그의 추상화 정보 데이터베이스 모두가 관계형 질의어등과 같은 단일 형식론으로 다루어질 수 있어야 한다. 세째, 추상화 정보 데이터베이스는 협력적 질의응답을 수행하기 위하여 데이터 값이나 도메인에 대해 일반화/세분화등과 같은 핵심적인 협력적 질의 처리 연산을 수행할 수 있어야 한다. 이러한 점들을 고려하여, 추상화 정보 데이터베이스에 대한 개체-관계도를 제시하고 관계형 스키마를 설계한다.

3.1 개념적 모델링

추상화 정보 데이터베이스에 대한 개념적 모델링을 수행하는데 있어, 먼저 단일 KAH에 대한 데이터 모델을 구축하고, 그 다음에 다수의 KAH를 반영하도록 데이터 모델을 확장하기로 한다. 단일 KAH에 대한 추상화 정보 데이터베이스 모델은 그림 3에 나타난 것과 같이 VALUE, DOMAIN, ATTRIBUTE, RELATION 등의 4개의 개체(entity)로 구성된다.



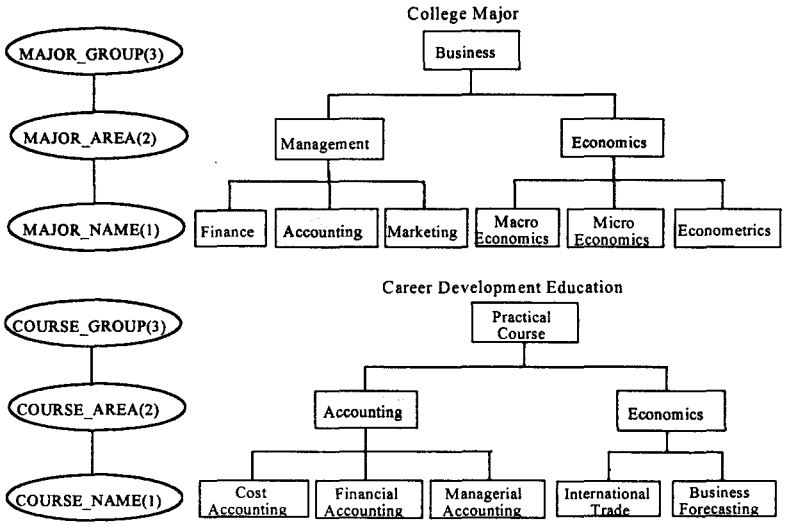
[그림 3] 단일 KAH에 대한 개체-관계도.

VALUE, DOMAIN 개체는 각각 값의 추상화 계층과 도메인 추상화 계층을 구성하는 값과 도메인을 나타낸다. 값의 추상화 계층에서 세분 값과 추상값 사이에는 반복적 일반화 관계가 존재하므로, VALUE 개체는 Has_Abstract_Value라는 자기 순환적 관계를 가지게 된다. 또한 사상 원소수는 일대다(one-to-many)가 된다.

도메인 추상화 계층과 관련하여, VALUE 개체는 DOMAIN 개체와 contained_in 관계를 갖는다. 하나의 도메인은 보통 다수의 값을 포함하고 있으므로 DOMAIN과 VALUE 개체의 사상 원소수는 일대다(one-to-many)가 된다. DOMAIN 개체는 계층에서의 자신의 위치를 결정해 주는 유일한 이름과 추상화 수준을 가지고 있으며, 수퍼 도메인과 서브 도메인 사이에 Has_Super_Domain 이라는 자기 순환적 일반화 관계를 가지고 있다. 이때, 하나의 수퍼 도메인은 하나의 서브 도메인을 가지기 때문에 Has_Super_Domain 관계의 사상 원소수는 일대일(one-to-one)이 된다. DOMAIN 개체는 기반 데이터베이스에 존재하는 릴레이션들의 개별 애트리뷰트들을 나타내는 ATTRIBUTE 개체와 관계를 갖는다. 다수의 애트리뷰트들이 하나의 동일한 도메인을 사용할 수 있으므로, 그들 사이의 사상 원소수는 일대다(one-to-many)가 된다.

다음은, 하나의 기반 데이터베이스에 대하여 다수의 KAH가 필요한 경우로 데이터 모델을 확장하기로 한다. 그림 4는 동일한 데이터베이스로부터 유도되는 두개의 KAH를 보여준다. 위의 KAH는 대학의 전공에 관한 College Major 계층을 나타내고, 반면에 아래의 KAH는 직원 교육 과목에 관한 Career Development Education 계층을 나타낸다. 필요할 경우, 동일한 데이터베이스로부터 또 다른 세 번째 추상화 계층이 추가될 수 있다.

데이터 모델이 다수의 KAH를 반영하도록 확장하는 것은, 그림 5에 나타난 것처럼 KAH 개체를 추가하여 DOMAIN, VALUE 개체와

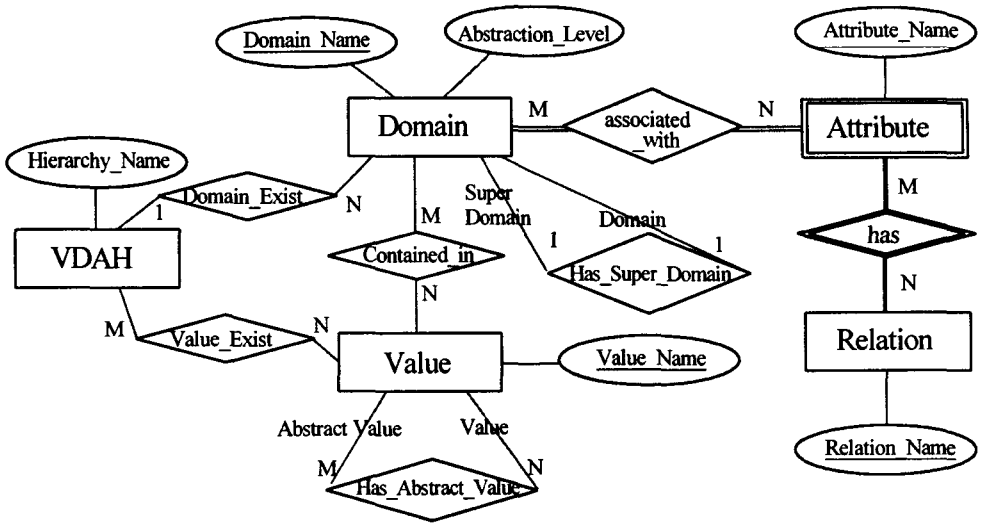


[그림 4] 단일 기반 데이터베이스로부터 유도된 다수의 KAH.

관계를 갖게 함으로써 이루어진다. 실제 여러 응용 문제들의 특성을 살펴볼 때, 다수의 KAH가 존재하는 경우 하나의 도메인은 자신이 속한 계층내의 도메인들과만 의미론적 관계를 갖지만, 하나의 값은 서로 다른 계층에 존재하는 다른 값들과도 의미론적 관계를 갖는 경우가 많다. 따라서 본 논문에서는 다수의 KAH가 존재하는 경우 하나의 도메인은 반드시 하나의 KAH에만 존재하는 반면, 하나의 값은 다수의 KAH에 존재할 수 있다고 가정한다. 결국 하나의 값을 고려할 때, 그 값이 어느 KAH에 존재하는가에 따라 서로 다른 도메인에 속할 수 있게 된다. 결과적으로, 한 값은 다수의 도메인에 속할 수 있기 때문에 그림 3에 나타난 DOMAIN과 VALUE 개체 사이의 일대다(one-to-many) 관계는 그림 5에서 다대다(many-to-many) 관계로 변하게 된다.

예를 들어 그림 4에 나타난 한 값 Economics는 서로 다른 계층에 존재하는 서로 다른 두개의 도메인에 속해 있다. College Major 계층에서는, MAJOR_AREA에 속해 있는 반면, Career Development Education 계층에서는 COURSE_AREA 도메인에 속해 있다. 이러한 경우, Economics는 자신이 속한 도메인에 따라 서로 다른 추상값과 세분값 집합을 갖게 된다. 따라서, Economics의 1단계 추상값을 구하고자 할 때, MAJOR_AREA가 도메인인 경우에는 Business가 1단계 추상값이 되고, COURSE_AREA가 도메인인 경우에는 Practical Course가 1단계 추상값이 된다. 같은 이유로, 한 값은 어느 도메인에 속해 있는가에 따라 서로 다른 1단계 세분값 집합을 갖게 된다. 본 논문에서는 이러한 특성을 추상값과 세분값의 도메인 종속성이라 부르기로 한다.

협력적 질의 응답의 관점에서 볼 때, 그림 5는 추상화 정보 데이터베이스가 질의를 어떻게 분석하게 되는지를 보여준다. 기존의 전통적인



[그림 5] 추상화 정보 데이터베이스에 대한 개체-관계도.

질의 처리에서는, 질의 조건에 명시해서 찾고자 하는 목표값이 데이터베이스내의 애트리뷰트 값들과 직접적으로 비교 되어진다. 예를 들어, Finance를 전공한 직원을 찾고자 하는 경우, 질의문에는 목표값 Finance와 전공을 나타내는 애트리뷰트, 이를 테면 Major가 명시된다. 이때, Major 애트리뷰트가 취하는 여러 값들과 Finance를 직접 비교하여, 정확히 일치하는 경우 질의의 결과로서 제시한다. 하지만, 협력적 질의 응답에서는 DOMAIN 개체가 목표값과 명시된 애트리뷰트 사이에 매개체로서 작용하여, 질의 조건을 정확히 만족하는 값들이 존재하지 않거나 유사한 값들이 요구되는 경우에, 그러한 유사한 값들이 애트리뷰트 값들과 비교되어 선택될 수 있게 해준다. 다시 말해서, DOMAIN 개체는 기존의 전통적인 질의 처리 과정에서는 가능하지 못했던 다음과 같은 일반화와 세분화 과정을 통하여 다양한 질의 완화 경로를 생성해 주는 것이다.

먼저 한 값에 대한 일반화 과정은 값의 추상화 계층을 따라 올라 가면서 그 값의 추상값들을 찾아내는 것을 말한다. 예를 들어 그림 4의 Finance에 대한 일반화 과정은 그의 1단계 추상값 Management를 찾는것에 해당한다. 그리고 그림 5에 제시된 개체-관계도에서 일반화 과정의 의미는 한 값이 Has_Abstract_Value 관계를 통하여 자신의 추상값과 일반화 관계를 갖는 것을 말한다. 보다 일반적인 임의의 n단계 추상값을 생성해 내는 좀 더 강력한 일반화 과정은 도메인 추상화 계층을 함께 사용함으로써 이루어진다. 그 이유는 질의문에 한 값이 주어졌을 때, 그 값의 해당 도메인과 수퍼 도메인은 도메인 추상화 계층을 통해 즉각적으로 찾아낼 수 있으며, 다음 수퍼 도메인 역시 현재 수퍼 도메인을 바탕으로 반복적으로 찾아낼 수 있어, 결과적으로 1단계 추상값보다 더 넓은 범위의 추상값들을 생성해 낼 수 있기 때문이다. 이것은 그림 4에서 Finance의 추상값 Management와, 그와 유사한 값으로 Economics까지 찾아내는 것에 해당한다. 반대로, 한 값에 대한 세

분과 과정은 값의 추상화 계층을 따라 내려가면서 그 값의 세분값들을 찾아낸다. 협력적 질의응답은 이러한 일반화와 세분화 과정을 이용하여 질의문에 명시된 목표값과 애트리뷰트 값을 직접적으로 비교하는 대신, 그들의 추상값이나 세분값들과 비교함으로써 좀더 넓은 범위의 값들을 선택할 수 있게 되는 것이다. 협력적 질의 응답에서의 일반화와 세분화 과정에 대한 좀 더 세부적인 사항은 4장에서 다루어 진다.

3.2 관계형 스키마 설계

KAH에 포함된 지식을 관리하기 위하여 추상화 정보 데이터베이스를 구축하는데 있어 관계형 데이터 모델을 이용하여 스키마를 설계함으로써, 하나의 단일화된 형식론을 통해 기반 데이터베이스와 추상화 정보 데이터베이스 모두를 다룰 수 있게 된다. 설계된 관계형 스키마는 기반 데이터베이스나 KAH에 발생하게 될 진화적 변화에 대하여 유연성을 가지고 적절히 대처할 수 있어야 하며, 협력적 질의 응답을 위해 일반화와 세분화 과정과 같은 질의처리 연산을 용이하게 수행할 수 있어야 한다. 그림 6은 그림 5의 개체-관계도로부터 추상화 정보 데이터베이스의 관계형 스키마를 설계하고, 그를 이용하여 그림 4에 존재하는 KAH에 대하여 추상화 정보 데이터베이스를 구축한 결과를 보여주고 있다.

그림 5에 나타난 값의 추상화 계층과 도메인 추상화 계층은 각각 그림 6에서 VALUE_ABSTRACTION 릴레이션과 DOMAIN_ABSTRACTION 릴레이션으로 구체화되어 나타난다. 기반 데이터베이스에 존재하는 릴레이션들과 그들의 애트리뷰트에 관한 정보는 ATTRIBUTE_MAPPING 릴레이션으로 나타난다.

구축된 추상화 정보 데이터베이스를 이용하여 한 값의 추상값을 구하고자 할 때, 앞에서 설명한 추상값의 도메인 종속성으로 인하여 그 값의 이름뿐 아니라 그 값의 도메인까지 알아야 한다. 따라서 VALUE_ABSTRACTION 릴레이션에서, Value와 Domain 두 애트리뷰트가 Abstract_Value 애트리뷰트 값을 결정하는 복합키가 된다. 예를 들어, 그림 6의 VALUE_ABSTRACTION 릴레이션에서 MAJOR_NAME 도메인에 속하는 Accounting의 추상값은, Practical Course가 아니라 Management가 된다. 마찬가지로, Accounting의 형제(sibling)값들은 동일한 도메인 MAJOR_NAME에 속하고 동일한 추상값 Management를 갖는 Finance, Marketing등이 된다.

한편, 도메인의 추상화 관계를 이용하여 한 도메인의 수퍼 도메인을 찾아내기 위해서는 그 도메인의 이름만 알면 되므로, DOMAIN_ABSTRACTION 릴레이션의 키 애트리뷰트는 Domain이 된다. 또한 DOMAIN_ABSTRACTION은 각 도메인이 속해 있는 KAH의 이름과 도메인의 추상화 수준도 관리한다. VALUE_ABSTRACTION 릴레이션도 도메인과 관련된 정보를 관리 하고 있는데, 이를테면 각 값이 속해 있는 도메인의 이름이 그것이다.

이러한 DOMAIN_ABSTRACTION 릴레이션은 특히 한 값의 n단계 추상값을 구하기 위해서 반드시 필요한 릴레이션이다. 예를 들어, 그림 6의 VALUE_ABSTRACTION 릴레이션을 이용하여 Econometrics의 2단계 추상값을 구하고자 할 때, 단순히 1단계 추

VALUE_ABSTRACTION((Value, Domain, Abstract_Value))
DOMAIN_ABSTRACTION((Domain, Super_Domain, Hierarchy, Abstraction_Level))
ATTRIBUTE_MAPPING((Relation, Attribute, Domain))

VALUE_ABSTRACTION

Value	Domain	Abstract_Value
Finance	MAJOR_NAME	Management
Accounting	*MAJOR_NAME*	*Management*
Marketing	MAJOR_NAME	Management
Macro Economics	MAJOR_NAME	Economics
Micro Economics	MAJOR_NAME	Economics
Econometrics	MAJOR_NAME	Economics
Management	MAJOR_AREA	Business
Economics	*MAJOR_AREA*	*Business*
Business	MAJOR_GROUP	
Cost Accounting	COURSE_NAME	Accounting
Financial Accounting	COURSE_NAME	Accounting
Managerial Accounting	COURSE_NAME	Accounting
International Trade	COURSE_NAME	Economics
Business Forecasting	COURSE_NAME	Economics
Accounting	*COURSE_AREA*	*Practice Course*
Economics	*COURSE_AREA*	*Practice Course*
Practice Course	COURSE_GROUP	
. . .	. . .	. . .

DOMAIN_ABSTRACTION

Domain	Super_Domain	Hierarchy	Abstraction_Level
MAJOR_NAME	MAJOR_AREA	College Major	1
MAJOR_AREA	MAJOR_GROUP	College Major	2
MAJOR_GROUP		College Major	3
COURSE_NAME	COURSE_AREA	Career Development Education	1
COURSE_AREA	COURSE_GROUP	Career Development Education	2
COURSE_GROUP		Career Development Education	3
. . .	. . .	. . .	. . .

ATTRIBUTE_MAPPING

Relation	Attribute	Domain
EMPLOYEE_RECORD	EMP_ID	ID
EMPLOYEE_RECORD	MAJOR_NAME	MAJOR_NAME
JOB_CAREER	EMP_ID	ID
JOB_CAREER	PERFORMED_JOB	MAJOR_AREA
DEPT_REQUIRE_MAJOR	DEPT_NAME	DEPT
DEPT_REQUIRE_MAJOR	REQUIRED_MAJOR	MAJOR_GROUP
....	....	....

[그림 6] 지식 추상화 정보 데이터베이스의 구성 릴레이션.

상값을 연속해서 두번 구함으로써 얻을 수는 없다. 왜냐하면, Econometrics의 1단계 추상값인 Economics에 대하여 다시 1단계 추상값을 구하고자 할 때, MAJOR_AREA 도메인에 속하는 Economics와 COURSE_ AREA에 속하는 Economics등, 두개의 Economics가 존재하기 때문이다. 둘 중에서 Econometrics의 1단계 추상값에 해당하는 것이 어느 것인지를 판단하기 위해서는, Econometrics가 속한 도메인 즉, MAJOR_NAME의 수퍼 도메인이 MAJOR_AREA와 COURSE_AREA 중에서 어느것 인지를 알아내야 한다. 이를 위하여 DOMAIN_ABSTRACTION 릴레이션을 이용하여 MAJOR_AREA가 찾고자 하는 수퍼 도메인임을 알아낼 수 있고, 그 다음에 VALUE_ABSTRACTION 릴레이션에서 Economics와 MAJOR_AREA를 복합키로 사용하여 Business가 Econometrics의 2단계 추상값임을 알아낸다.

이처럼, 임의의 n단계 추상값을 구하기 위해서 VALUE_ABSTRACTION 릴레이션 외에도 DOMAIN_ABSTRACTION 릴레이션까지 필요로 하는 것은 VALUE_ABSTRACTION의 스키마 구조와도 관련이 있다. VALUE_ ABSTRACTION 릴레이션과 DOMAIN_ ABSTRACTION 릴레이션의 각 투플은 한 값과 그 추상값으로 구성된 쌍, 그리고 한 도메인과 그 수퍼 도메인으로 구성된 쌍의 형태로서, 단지 1단계 추상화 관계만을 나타내고 있다. 일단 추상화 지식이 추상화 정보 데이터베이스에 그러한 1단계 추상화의 형태로 저장되면, 임의의 추상화 수준에 존재하는 추상값은 추상화의 가법성에 기초하여 1단계 추상값을 반복적으로 찾아냄으로써 구할 수 있게 된다.

마지막으로 ATTRIBUTE_MAPPING 릴레이션은 기반 데이터베이스에 존재하는 릴레이션들의 애트리뷰트가 갖는 도메인을 관리함으로써, 질의문의 의도를 분석하고 VALUE_ ABSTRACTION, DOMAIN_ABSTRACTION 등과 결합하여 다양한 질의 완화 경로를 생성해 내기 위해 사용된다. 기반 데이터베이스에서 동일한 애트리뷰트 이름이 여러 릴레이션에서 사용될 수 있기 때문에 Relation과 Attribute 모두가 ATTRIBUTE_MAPPING 릴레이션의 복합키가 된다. 협력적 질의응답과 관련하여 ATTRIBUTE_MAPPING 릴레이션이 이용되는 방법은 4장에서 좀더 자세하게 논의될 것이다.

이러한 세가지 릴레이션들의 정규화 정도를 살펴보면, 모든 결정자가 후보키가 되기 때문에, 제 3정규형보다 더욱 강력한 보이스/코드(Boyce Codd) 정규화에 해당한다. 따라서 기존의 연구들에서 제시된 릴레이션들이 정규화 되지 못함으로써 발생 하였던 지식 베이스 관리상의 여러 문제점들이 본 논문에서는 발생하지 않게 된다.

4. 추상화 정보 데이터베이스 연산

구축된 추상화 정보 데이터베이스는 KAH에 포함된 지식의 변화를 동적으로 처리하고 다양한 유형의 협력적 질의를 처리할 수 있어야 한다. 이 장에서는 추상화 정보 데이터베이스가 지식 관리와 협력적 질의처리를 수행하기 위해 필요한 연산들을 논의한다. 첫 번째, 지식관리의 관점에서 볼 때 KAH에 포함된 의미론적 지식은 시간이 지남에 따라 계속해서 변화해 나가

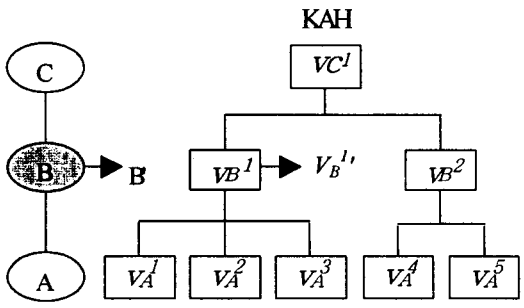
고, 이러한 지식의 변화는 추상화 정보 데이터베이스에 적절하게 반영 되어야 한다. 따라서 KAH에 일어날 수 있는 가능한 변화들을 검토하고, 그러한 변화들이 추상화 정보 데이터베이스에서 어떻게 처리되는지를 설명한다. 두 번째 협력적 질의응답의 관점에서, 추상화 정보 데이터베이스가 질의 변환을 수행하기 위해 필요한 연산들 즉, 한 값이나 도메인에 대한 일반화, 세분화 과정을 어떻게 수행하는지 설명한다. 마지막으로 기존의 다른 연구들과 비교할 때, 본 논문에서 제시하는 추상화 정보 데이터베이스가 갖는 장점을 논의한다.

4.1 지식 관리 연산

기반 데이터베이스에 대한 추상화의 의미론적 지식뿐 아니라 KAH의 이용관점이 변해 나감에 따라, KAH에 포함된 지식이 변화해 나간다. KAH에 포함된 지식의 변화는 KAH의 내용상의 변화와 구조적 변화 두 가지로 구분할 수 있다. 내용상의 변화는 KAH 형태상의 구조는 변화하지 않은 채, 도메인이나 값 인스턴스들이 진화적으로 변경되어 나가는 것이다. 반면 구조적인 변화는 한 노드의 차수 즉, 세분값의 갯수가 변한다든지, KAH의 추상화 수준의 깊이가 변화하는 등, 트리 (tree) 구조가 변화하는 것을 의미한다. 추상화 정보 데이터베이스가 그러한 변화들을 처리하기 위하여 사용하는 연산들은 다음과 같다.

1) 내용값의 변화

KAH의 내용값의 변화는 값의 추상화 계층과 도메인 추상화 계층 모두에서 일어날 수 있다. 먼저, 값의 추상화 계층의 경우에 대한 예로서 그림 1의 Management가 다른 값 Administration으로 변경되는 경우를 생각해 보자. 한 값이 변화하게 되면, 그 값은 물론 변화하는 값이 관련된 추상화 관계 역시 변화하게 된다. 다시 말해서, KAH의 내용값의 변화는 자신의 변화뿐 아니라, 1단계 추상화 관계의 데이터베이스 무결성을 유지하기 위해 자신과 관련된 다른 값들의 후속적인 변경을 필요로 한다. 이러한 내용값의 변경은 그림 7에서 $v_B{}^1$이 $v_B{}^{1'}$ 로 변하는 경우로 설명될 수 있다.



[그림 7] KAH의 내용값 변화.

그림 7에서 값 $v_B{}^1$은 그의 자식값 $v_A{}^1$, $v_A{}^2$, $v_A{}^3$등과 1단계 추상화 관계를 형성하고 있기 때문에, $v_B{}^1$이 $v_B{}^{1'}$로 변화하는 것은 VALUE_ABSTRACTION 릴레이션에서 $v_B{}^1$이 $v_B{}^{1'}$로 갱신되는 것뿐 아니라, $v_A{}^1$, $v_A{}^2$, $v_A{}^3$와의 추상화 관계도 갱신되어야 한다. 따라서 Value 애트리뷰트가 값 $v_B{}^1$을 취하고 있는 투플은 그 값을 $v_B{}^{1'}$으로 갱신 하여야 한다. 또한 1단계 추상화 관계의 데이터베이스 무결성을 만족시키기 위해서, Abstract_Value 애트리뷰트가 값 $v_B{}^1$을 취하고 있는 투플들 역시 그 값을 $v_B{}^{1'}$로 갱신 하여야 한다. 이러한 갱신 작업은 Value 애트리뷰

VALUE_ABSTRACTION

Value	Domain	Abstract_Value
Finance	MAJOR_NAME	***Management***
Accounting	MAJOR_NAME	***Management***
Marketing	MAJOR_NAME	***Management***
Macro Economics	MAJOR_NAME	Economics
Micro Economics	MAJOR_NAME	Economics
Econometrics	MAJOR_NAME	Economics
Management	MAJOR_AREA	Business
Economics	MAJOR_AREA	Business
Business	MAJOR_GROUP	
. . .	. . .	. . .

(a) 변화 전

Value	Domain	Abstract_Value
Finance	MAJOR_NAME	***Administration***
Accounting	MAJOR_NAME	***Administration***
Marketing	MAJOR_NAME	***Administration***
Macro Economics	MAJOR_NAME	Economics
Micro Economics	MAJOR_NAME	Economics
Econometrics	MAJOR_NAME	Economics
Administration	MAJOR_AREA	Business
Economics	MAJOR_AREA	Business
Business	MAJOR_GROUP	
. . .	. . .	. . .

(b) 변화 후

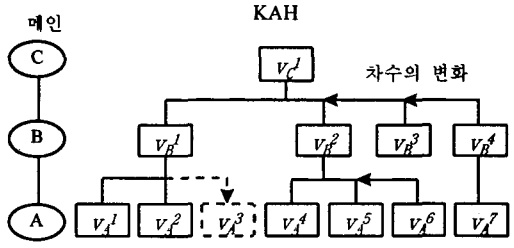
[그림 8] KAH의 내용값 변화에 대한 지식 추상화 정보 데이터베이스의 처리 결과.

트가 v_A^1, v_A^2, v_A^3를 취하고 있는 투플들을 대상으로 하게 된다. 그림 8은 Management가 Administration으로 변함에 따라 VALUE_ABSTRACTION이 갱신된 결과를 보여주고 있다.

다음은 도메인 추상화 계층에서 하나의 도메인이 변경되는 경우를 고려해 보자. 도메인의 변화에 대한 처리는, VALUE_ABSTRACTION, DOMAIN_ABSTRACTION, ATTRIBUTE_MAPPING등 세가지 릴레이션 모두에 대한 갱신 작업을 필요로 한다는 점을 제외하고는 값의 변경시에 대한 처리와 유사하다. 그림 7에서 도메인 B가 B'로 변하는 경우를 생각해 보자. 먼저 도메인 B와 관련된 모든 기반 데이터베이스 애트리뷰트들에 대해서 B의 변화를 처리해 주어야 하며, 이것은 ATTRIBUTE_MAPPING 릴레이션에서 이루어 진다. 따라서 ATTRIBUTE_MAPPING 릴레이션에서 Domain 애트리뷰트 값이 B인 모든 투플들에 대하여 그 값을 B'로 갱신하여야 한다. 두 번째, 변화된 도메인 B'와 관련된 추상화 관계 역시 갱신되어야 하며, 그것은 DOMAIN_ABSTRACTION 릴레이션에서 이루어진다. 한 도메인과 그의 수퍼 도메인 사이의 대응 관계가 일대 일(1:1)이므로 두개의 투플 즉, Domain 애트리뷰트 값이 B인 투플, 그리고 Super_Domain 애트리뷰트 값이 B인 투플이 B'로 갱신되어 진다. 세 번째, 도메인의 변화는 변화된 그 도메인에 포함 되었던 값에도 영향을 미치게 되고, 그 영향이 VALUE_ABSTRACTION 릴레이션에 나타난다. 다시 말해서, Domain 애트리뷰트 값 이 B인 모든 투플들에 대하여 그 값을 B'로 갱신하게 된다.

2) 값의 차수 변화

앞에서 설명한 KAH의 두 가지 종류의 구조적 변화 중에서, 먼저 한 노드값의 차수 변화를 고려해 보자. 한 도메인은 언제나 하나의 서브 도메인만을 갖게 되므로, 노드의 차수 변화는 값의 추상화 계층에서만 의미를 갖게 된다. 그림 9-a)는 v_A^6, v_A^7, v_B^3, v_B^4 가 추가되고, v_A^3가



(a) 값의 차수 변화.

VALUE_ABSTRACTION

Value	Domain	Abstract_ Value
v_A^1	A	v_B^1
v_A^2	A	v_B^1
v_A^3	A	v_B^1
v_A^4	A	v_B^2
v_A^5	A	v_B^2
v_B^1	B	v_C^1
v_B^2	B	v_C^1
v_C^1	C	...
v_A^6	A	v_B^2
v_A^7	A	v_B^4
v_B^3	B	v_C^1
v_B^4	B	v_C^1

(b) 추상화 정보 데이터베이스에서의 처리

[그림 9] 노드값의 차수 변화.

제거되는 경우를 나타내고 있다. 이처럼 노드를 추가하고 제거함에 따라, v_B^2과 v_C^1의 차수는 증가하고 반면에 v_B^1의 차수는 감소하게 된다. 그림 9-b)는 이러한 변화가 VALUE_ABSTRACTION 릴레이션에 반영된 결과를 나타내고 있다. 직선의 회색은 삽입되는 투플을 나타내고 점선의 회색은 삭제되는 투플을 나타낸다.

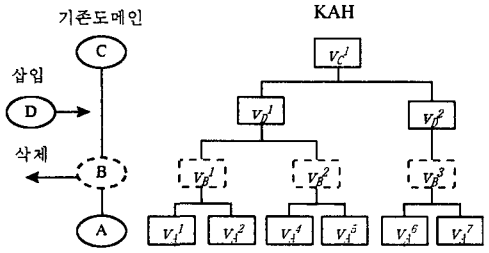
그림 9-a)에 나타난 것처럼, 한 도메인으로 새로운 값을 추가하는 것은 그 값의 추상값의 차수를 증가 시킨다. 이때 v_B^3처럼 최하위 수준에 존재하지 않으면서 세분값을 갖지 않는 경우라 하더라도 마찬가지이다. 이러한 차수의 변화는 VALUE_ABSTRACTION 릴레이션에서 처리 되는데, 추가되는 새로운 값과 그 값이 속한 도메인, 그리고 그 값의 추상값으로 구성된 새로운 투플을 삽입해야 한다. 반면에 세분값들이 제거됨으로써 한 값의 차수가 감소하는 경우에는, 제거되는 값을 키로 갖는 투플을 삭제하는 것으로 처리가 가능하다.

3) KAH 깊이의 변화

구조적 변화의 두 번째 경우는, 새로운 도메인이 추가되거나, 기존의 도메인이 삭제됨으로써 KAH의 깊이가 변화하는 경우이다. 도메인의 추가와 삭제는 항상 그 도메인에 속한 값들의 추가와 삭제를 동반한다. 그림 10은 도메인을 추가, 삭제함으로써 KAH의 깊이가 변화하는 예를 보여주고 있다.

그림의 예는 새로운 도메인 D가 도메인 B와 C사이에 삽입되고, 그 다음에 B가 제거되는 경우이다. D가 v_D^1, v_D^2와 같은 값들을 포함하고 있으므로, 도메인 D를 삽입하는 것은 값의 추상화 계층에 현재 존재하고 있는 값들 v_B^1, v_B^2 v_B^2, v_C^1에 v_D^1, v_D^2을 연결하는 것을 필요로 한다. 기존의 도메인이 삭제되는 경우에는 데이터베이스 무결성을 유지하기 위하여, 그 도메인에 속한 값들도 삭제 되어야만 한다. 결국, KAH에 한 도메인을 추가하고 삭제하는 것은 추상화 정보 데이터베이스에 대한 2단계 지식 갱신 과정을 필요로 한다. 첫번째 단계는 도메인 추상화 계층에서의 도메인의 추가나 삭제를 처리하기

위하여 DOMAIN_ABSTRACTION 릴레이션을 갱신하는 것이다. 두 번째 단계는 도메인의 추가나 삭제에 의해 후속적으로 발생하는 값의 추가, 삭제를 처리하기 위하여 VALUE_ABSTRACTION 릴레이션을 갱신하는 것이다.



[그림 10] KAH의 깊이 변화.

먼저, 도메인 D가 추가되는 경우를 고려해 보자. 도메인의 추가는 우선 DOMAIN_ABSTRACTION 릴레이션에서 2개의 투플에 영향을 미친다. Domain, Super-Domain 애트리뷰트 값이 각각 D, C인 새로운 투플이 삽입되고, C를 Super-Domain 애트리뷰트 값으로 갖는 기존의 투플은 새로운 도메인 D를 그 애트리뷰트 값으로 갖도록 갱신된다. 그림 11-a)는 이러한 D의 추가가 DOMAIN_ABSTRACTION 릴레이션에서 처리되고, 나중에 설명되어질 B의 제거까지 처리된 최종결과를 보여준다. 다음으로 도메인의 추가는 VALUE_ABSTRACTION 릴레이션에 대한 지식 갱신 작업도 요구하게 된다. 예를 들어, 도메인 D의 추가로 인하여 새로운 값 v_D^1이 기존의 값들 v_B^1과 v_C^1사이에 삽입될 때, Value와 Abstract_Value 애트리뷰트 값이 각각 v_D^1, v_C^1인 새로운 투플이 삽입되고, v_C^1을 Abstract_Value 애트리뷰트 값으로 갖는 기존의 모든 투플들은 새로운 도메인 v_D^1을 그 애트리뷰트 값으로 갖도록 갱신된다. 그림 11-b)는 이러한 과정이 VALUE_ABSTRACTION 릴레이션에서 처리된 결과를 보여준다. 그림 11에서, 추가된 도메인과 값들은 해당 릴레이션의 키 애트리뷰트에 해당하므로, 두 릴레이션 모두 키 애트리뷰트 값이 널(NULL) 이거나, 가공의 값을 갖어야만 하는 삽입 이상이 발생하지 않게 된다.

DOMAIN_ABSTRACTION

Domain	Super_Domain	Hierarchy	Level
A	D	H1	1
B	D	H1	2
C		H1	3
D	C	H1	2
...	...	H2	...

(a) DOMAIN_ABSTRACTION 처리결과

VALUE_ABSTRACTION

Value	Domain	Abstract_Value
v_A^1	A	v_D^1
v_A^2	A	v_D^1
v_A^4	A	v_D^1
v_A^5	A	v_D^1
v_A^6	A	v_D^2
v_A^7	A	v_D^2
v_B^1	B	v_D^1
v_B^2	B	v_D^1
v_B^3	B	v_D^2
v_C^1	C	
v_D^1	D	v_C^1
v_D^2	D	v_C^1
...	...	...

(b) VALUE_ABSTRACTION 처리결과

[그림 11] 깊이 변화에 대한 추상화 정보 데이터베이스의 처리 결과.

다음으로, 도메인이 제거되는 경우도 도메인의 추가되는 경우와 유사하게 처리된다. 먼저, DOMAIN_ABSTRACTION 릴레이션이 갱신된다. 그림 11-a)에 나타난 것처럼, 도메인 B가 제거되는 경우에 B의 서브 도메인인 A를 Domain 애트리뷰트 값으로 갖는 모든 투플은 자신의 Super_Domain 애트리뷰트 값으로, 제거되는 도메인 B의 수퍼 도메인 D를 갖도록 갱신 되어야만 한다. 그리고 나서, B를 Domain 애트리뷰트 값으로 갖는 투플을 삭제 하여도 삭제 이상이 발생하지 않게 된다. 두 번째, VALUE_ABSTRACTION 릴레이션에 대한 지식 갱신 작업도 이루어 진다. 도메인 A의 수퍼 도메인이 D로 바뀜에 따라, v_A^1, v_A^2, ..., v_A^6, v_A^7의 추상값도 v_D^1, v_D^2로 갱신되어야 한다. 그 다음에, 도메인 B에 속하는 값들 v_B^1, v_B^2, v_B^3를 삭제하게 된다.

4.2 협력적 질의응답 연산

이 절에서는 협력적 질의응답을 수행하기 위해 필요한 추상화 정보 데이터베이스 연산들을 설명한다. 일반적인 질의처리에서는, 질의 조건을 정확히 만족하는 답이 존재하지 않을 경우 아무런 답을 제시해 주지 못한다. 하지만 협력적 질의 응답에서는 질의 처리시 탐색 조건을 완화시켜, 추가적인 답을 포함하도록 질의 조건의 범위를 넓힐 수 있으므로, 유사하거나 일반적인 정보들을 대신 제공해 줄 수 있다. 이를 위하여, KAH를 바탕으로 질의 조건에 나타난 값들을 그들의 추상값이나 세분값들로 치환함으로써, 질의의 범위를 확대하거나 축소하는 질의 변환 과정을 수행하게 된다.

그림 12-a)에 나타난 인사 정보 데이터베이스를 기반 데이터베이스로 해서 구축된 KAH가 그림 4에 나타난 college major라고 가정하자. 만약 인사 관리 담당자가 Finance나 그와 관련된 분야를 전공한 사람을 찾고자 하여, 그림 12-b)와 같은 질의문을 작성한 경우, 직원 번호 005를 갖는 한 명의 직원만이 검색된다. 하지만 만약 KAH를 이용하여 질의 영역을 Finance의 1단계 추상값인 Management로 확대시킨 다음, 다시 질의 조건을 Management의 세분값들 Finance, Accounting, Marketing등으로 축소 시킨다면, 원래의 Finance외에 Accounting이나 Marketing등을 전공한 직원들까지 포함된 더 넓은 범위의 정보들을 추가적으로 얻을 수 있게 된다. 만약 질의 범위를 더욱 확장 시킨다면, Management 분야뿐 아니라, Macro Economics, Micro Economics, Econometrics등의 Economic 분야를 전공한 사람들을 포함하는 더 넓은 범위의 근사값들이 얻어질 수 있을 것이다.

협력적 질의응답은 보통 이러한 근사해를 얻기 위해서, 질의조건을 완화하고 축소하는 과정을 통해 질의 변형을 수행하게 된다. 질의 조건 Q가 주어졌을 때, 질의 변형의 다음과 같은 단계를 거친다.

1. 질의 조건 Q를 정확히 만족하는 해를 찾는다. 만약 실패하면,
2. 추상화된 질의를 생성하기 위하여 KAH의 상위 수준으로 이동해 가면서 질의 조건을 일반화 시킨다 (질의 일반화 과정) 즉, Q -> Q'. 다음에,
3. 추상화된 질의를 세분화 시키기 위하여 KAH의 하위 수준으로 이동해 가면서 세분 질의를 생성한다 (질의 세분화 과정), 즉 Q' -> Q' '.

EMPLOYEE_RECORD

Emp_No	Major
001	Marketing
002	Accounting
003	Econometrics
004	Accounting
005	Finance
006	Macro Economics
007	Econometrics
008	Micro Economics
009	Marketing
010	Accounting
...	...

(a) 기반 인사정보 데이터베이스

```
select      *
from        EMPLOYEE_RECORD E
where       E.Major =Finance
domain :    MAJOR_NAME  MAJOR_NAME
```

(b) Finance 전공자 검

```
select      *
from        EMPLOYEE_RECORD E
where       E.Major =Management
domain :    MAJOR_NAME  MAJOR_NAME
```

(c) Management 전공자 검색

[그림 12] 상이한 추상화 수준의 데이터 개체를 포함한 질의문.

협력적 질의응답이 위의 순서대로 질의를 변형해 나감에 따라, 그림 12-b)의 질의 조건은 major = Finance와 major in {Finance, Accounting, Marketing}으로 변하게 된다.

한편, 인사 관리 담당자가 일반적인 경영학 분야를 전공한 사람들을 찾고자 하지만, 대학 전공 분야에 관한 지식이 부족하여 "Accounting, Finance, Marketing"과 같은 값들에 익숙치 못한 경우에, 그림 12-c)에 나타난 것처럼 major = Management과 같은 개념적 수준의 질의 조건을 작성할 수 있다. 기존의 일반적인 질의 응답은 이러한 질의문을 잘못 작성된 질의문으로 판단하게 되지만, 협력적 질의응답은 이러한 질의조건에 대하여 세 번째 질의 변형 단계, 즉 세분화 과정을 수행함으로써 개념적 질의조건을 처리할 수 있게 된다. 다시 말해서, 질의 조건을 major=Management에서 major in {Finance, Accounting, Marketing}과 같이 세분화된 질의 조건으로 변형함으로써 질의 처리가 가능하게 된다. 그림 4에 나타난 KAH에서 일반화와 세분화 과정을 수행하는 것은 상당히 직관적이지만, 그림 6에 나타난 추상화 정보 데이터베이스를 이용하여 이러한 과정을 수행하기 위해서는 좀더 세부적인 논의가 필요하다.

1) 일반화 연산

추상화 정보 데이터베이스를 이용하여, 질의 조건에 명세된 값에 대해 일반화 과정을 수행하는 것은, KAH의 상위 수준으로 이동해 가면서 추상값을 구해 그 추상값으로 대체해 나가는 것을 말한다. 예를 들어, 그림 4에 나타난 College Major KAH에 존재하는 Finance를 고려해 보자. 1단계 일반화는 Finance외에 Accounting, Marketing까지 포괄하는 1단계 추상값 Management를 제공한다. 2단계 일반화는 한 수준 더 추상화된 값인 2단계 추상값 Business를 반환한다. 2단계 추상값 Business는 Economics를 포함하여 더 넓은 범위의 근사값들을 포함하는 값이다.

한 값과 그 값이 속한 도메인이 주어졌을 때, 1단계 일반화 과정을 통하여 그 값의 1단계 추상값을 구하는 것은, 단지 VALUE_ABSTRACTION 릴레이션만을 필요로 한다. 왜냐하면, VALUE_ABSTRACTION 릴레이션이 Abstract_Value 애트리뷰트에 1단계 추상

값을 이미 관리하고 있기 때문이다. 이때 주목할 것은, 한 값의 추상값을 구하기 위해서는 그 값이 존재하는 도메인도 함께 주어져야 한다는 것이다. 예를 들어, 그림 6의 VALUE_ABSTRACTION 릴레이션을 이용하여 College Major KAH에 존재하는 Accounting의 1단계 추상값을 구하면, Practice Course가 아니라 Management가 된다. 그 이유는 고려하는 Accounting의 도메인이 Career Development Education KAH에 존재하는 COURSE_AREA가 아니라, College Major KAH에 존재하는 MAJOR_NAME이기 때문이다.

반면에, n단계 일반화 과정은 VALUE_ABSTRACTION 릴레이션과 DOMAIN_ABSTRACTION 릴레이션 모두를 필요로 한다. 원칙적으로, 임의의 n단계 추상값은 1단계 일반화 과정을 n번 반복함으로써 얻어질 수 있지만, 추상값의 도메인 종속성으로 인하여 수퍼 도메인 정보를 필요로 하는데, 이것은 DOMAIN_ABSTRACTION 릴레이션을 이용해서 얻을 수 있기 때문이다. 예를 들어, Career Development Education KAH에 존재하여 COURSE_NAME 도메인에 속하는 Cost Accounting의 2단계 추상값을 구해보자. 먼저 1단계 일반화 과정은 Cost Accounting의 도메인이 COURSE_NAME임을 이용하여, VALUE_ABSTRCTION 릴레이션으로부터 1단계 추상값 Accounting을 찾아낸다. 하지만, 두 번째로 Accounting에 대한 1단계 일반화 과정을 수행하기 위해서는, Accounting의 도메인을 알아야만 한다. 그것은 물론 COURSE_NAME의 수퍼 도메인에 해당한다. 이러한 수퍼 도메인 정보는 DOMAIN_ABSTRACTION 릴레이션을 이용하여 얻어질 수 있고, 그 결과 COURSE_AREA가 수퍼 도메인이 된다. 3단계 추상값 역시, 2단계 추상값과 2단계 추상값이 존재하는 도메인 정보를 가지고 세 번째 1단계 일반화 과정을 수행함으로써 찾아질 수 있다. 이러한 과정을 계속 반복함으로써, 임의의 n단계 추상값이 구해질 수 있게 된다.

KAH에서 상위의 추상화 수준에 존재하는 하나의 추상값은 하위 수준에 존재하는 다수의 세분값에 해당한다. 따라서, 협력적 질의응답에서 한 추상값을 사용해 질의를 수행하는 것은 다수의 세분값들을 사용해 질의하는 것과 동일한 의미를 갖게 된다.

2) 세분화 연산

하나의 추상값은 그보다 하위 수준에 존재하는 다수의 세분값들에 대응한다. 주어진 값에 대한 세분화 과정은 KAH의 하위 수준으로 이동해 가면서 임의의 n단계 세분값을 찾아내는 것을 말한다. 예를 들어 그림 4의 College Major 계층에 존재하는 Business를 고려해 보자. Business에 대한 1단계 세분화 과정은 1단계 세분값으로 Management와 Economics를 반환한다. 2단계 세분화는 Finance에서 Econometrics에 이르는 6개의 값들 즉, 2단계 세분값들을 반환하며, 그것들은 Management와 Economics의 1단계 세분값들에 해당한다.

일반화 과정과 유사하게, 세분화 과정 역시 임의의 n단계 세분값을 찾아내기 위해서는 VALUE_ABSTRACTION 릴레이션과 DOMAIN_ABSTRACTION 릴레이션 모두를 필요로 한다. 먼저, 주어진 값에 대한 1단계 세

분화 과정을 살펴 보기로 하자. VALUE_ABSTRACTION 릴레이션을 이용하여 주어진 값의 1단계 세분값들을 찾아내기 위해서는 추상값과 세분값의 도메인 종속성으로 인하여 세분값들이 존재하게 될 도메인을 먼저 알아내야 한다. 이것은 DOMAIN_ABSTRACTION 릴레이션을 이용하여 주어진 값이 존재하는 도메인의 서브 도메인을 찾아냄으로써 가능하게 된다. 세분값들이 존재하게 될 도메인이 결정되면, VALUE_ABSTRACTION 릴레이션을 이용하여 주어진 값의 세분값들을 찾아낼 수 있다.

예를 들어, 그림 6의 추상화 정보 데이터베이스를 이용하여 그림 4에서 College Major KAH에 존재하는 Economics의 세분값을 구해 보기로 하자. Economics의 1단계 세분값을 구하기 위해서는, 세분값의 도메인 종속성으로 인하여 Economics의 세분값들이 존재하게 될 도메인을 먼저 알아야 한다. 이는 Economics가 속하는 도메인 MAJOR_AREA의 서브 도메인에 해당하며, DOMAIN_ABSTRACTION 릴레이션을 이용하여 MAJOR_NAME가 됨을 알 수 있다. 그 다음에, VALUE_ABSTRACTION 릴레이션에서 Domain과 Abstract_Value 애트리뷰트 값으로 각각 MAJOR_NAME과 Economics를 갖는 투플들을 선택함으로써 Economics의 세분값 Macro Economics, Micro Economics, Econometrics등을 찾아낸다. n단계 세분화 과정은 1단계 세분화 과정을 동일하게 n번 반복 함으로써 이루어질 수 있다.

결론적으로, 협력적 질의응답을 수행하기 위한 일반화와 세분화 과정은 값의 추상화 지식과 도메인 추상화 지식 모두를 필요로 한다. 다시 말해서, 추상화 정보 데이터베이스에 저장된 값의 1단계 추상화 지식을 이용하여 임의의 n단계 추상값 혹은 세분값을 구하는 과정은, 추상값과 세분값의 도메인 종속성으로 인하여 값의 추상화 지식 외에 도메인 추상화 지식을 추가로 필요로 한다. 따라서 KAH에 존재하는 도메인 추상화 지식은 동적인 지식관리를 가능하게 해줄 뿐 아니라, 값의 추상화 지식을 1단계 형태로 관리하고도 임의의 n단계 일반화와 세분화 과정을 수행해 내는 것을 가능하게 해 주는 것이다.

3) 추상화 정보 데이터베이스를 이용한 개념적 질의

추상화 정보 데이터베이스는 추상화 수준이 다른 데이터 개체들이 비교 되어지는 개념적 질의를 처리할 수 있다. 예를 들어, 그림 12-c)의 질의조건 Major=Management에서 비교되어지는 두 개체는 서로 상이한 추상화 수준을 가지고 있다. 다시 말해서, 선택 연산 애트리뷰트 Major는 College Major KAH에 존재하는 MAJOR_NAME을 도메인으로 가지고 있고, 반면에 비교 대상값 Management는 MAJOR_AREA를 도메인으로 갖는다. 이러한 질의 조건은 사용자의 실수로 작성된 것일 수도 있지만, 더 넓은 범위에 속한 다양한 전공들이 검색되어져 "Major in Management area" 혹은 "Major in (Finance, Marketing, Accounting)" 등으로 해석 되어지기를 바라면서 의도적으로 작성된 질의문일 수도 있다. 어떤 경우든, 기존의 질의처리는 Major 애트리뷰트가 Management를 애트리뷰트 값으로 갖지 않을 뿐 아니라, Management와 Major 애트리뷰트의 도메인이 애초부터 서로 다르기 때문에 질의를 만족하는 답이 존재하지 않게 되고, 결국 어떠한 값도 제

시해 주지 못한다. 하지만 동일한 질의에 대해서, 본 논문에서 제시하는 추상화 정보 데이터베이스는 목표값이 선택 연산 애트리뷰트의 도메인 MAJOR_NAME에 속한 값이 아니라는 점을 파악한 뒤 질의의 의도를 묻는다. 다시 말해서, Management가 MAJOR_NAME보다 높은 추상화 수준을 갖는 MAJOR_AREA 도메인에 존재하는 값이라는 사실을 주목 하면서, 사용자가 좀 더 높은 추상화 수준인 개념적인 수준에서 질의를 처리하고 싶은지 묻는다.

이러한 개념적 질의를 처리하는데 있어 추상화 정보 데이터베이스는 질의에 명시된 애트리뷰트 Major와 목표값 Management의 도메인의 추상화 수준을 비교하고 질의 변환 방법을 먼저 정한다. 그림 12-c)에서 Major 애트리뷰트의 추상화 수준은 Management 값의 추상화 수준보다 1단계 낮다. 이처럼 추상화 수준에 차이가 존재할 때, 두 가지 방법의 질의변환이 가능하다. 하나는 추상화 수준이 낮은 쪽을 일반화하는 것이다. 이것은 그림 12-c)에서 EMPLOYEE_RECORD 릴레이션의 모든 Major 애트리뷰트 값들을 그들의 1단계 추상값으로, 이를테면 Marketing을 Management로 일반화 시키고, 일반화된 값을 Management와 비교하는 것에 해당한다. 다른 하나는 추상화 수준이 더 높은 쪽을 세분화하는 것이다. 그림 12-c)에서, 추상값 Management를 Finance, Accounting, Marketing 등의 다수의 세분값들로 세분화하여 그들을 Major 애트리뷰트 값과 비교하는 것이다.

결과적으로 본 논문에서 제시하는 추상화 정보 데이터베이스는 보다 많은 유형의 협력적 질의를 처리할 수 있게 해 주고, 사용자와의 상호작용도 향상시켜 준다. 결국, KAH는 다른 연구들에서 제시한 지식 표현 프레임웍보다 더욱 풍부한 의미론적 지식을 표현함으로써, 동적인 지식관리뿐 아니라 더욱 효과적인 협력적 질의응답을 가능하게 한다고 할 수 있다.

4) 다른 연구와의 비교

본 논문에서 제시하는 추상화 정보 데이터베이스를 기존의 객체지향 데이터베이스 기법, 규칙 기반 기법, 개념적 분류 기법등과 비교하면 다음과 같다. 첫째, 값의 추상화와 도메인의 추상화 개념 모두를 포함하는 것은 기존의 어떤 기법에서도 발견되지 않는다. 객체지향 데이터베이스 기법에는 클래스와 인스턴스라는 두개의 층(layer)이 존재한다. 클래스 층에서는 동일한 특성을 갖는 개체들을 하나의 클래스로 그룹화하기 위하여 분류(classification)의 개념이 사용된다. 클래스들은 그들 사이의 일반화와 세분화 관계를 바탕으로 수퍼 클래스/서브 클래스의 관계를 갖는다. 분류 관점은 개체 인스턴스 값 자체보다는 그들의 전체적인 타입(type)에 관심을 갖는다. 이러한 클래스-지향적 추상화는 KAH에서의 도메인 추상화 계층과 거의 유사하다. 하지만, 인스턴스 층에서는 하나의 클래스에 속하는 여러 인스턴스들이 자신보다 더욱 추상화된 여러 인스턴스 그룹들로 분류 되어지는 값의 추상화와 같은 개념이 존재하지 않는다. 따라서 객체지향 데이터베이스에서는 한 서브 클래스 인스턴스들의 특정 집합이 수퍼 클래스에 존재하는 추상화된 개별 인스턴스 하나와 직접적으로 관련을 갖지는 않는다.

개념적 분류 기법에서는 다단계 지식 표현 프레임웍에 기초하여 값의 추상화 개념을 제시한 첫번째 연구로서 타입 추상화 계층(type abstraction hierarchy)[6, 8]을 들 수 있다. 타

입 추상화 계층은 수퍼 타입과 서브 타입간에 다수의 인스턴스 층이 존재하고, 따라서 서브 타입의 인스턴스들은 수퍼 타입의 서로 다른 인스턴스들과 의미론적 관계를 갖을 수 있다. 하지만 타입 추상화 계층이 도메인 추상화 계층을 포함하고 있지 않기 때문에, 일반화와 세분화 관계는 단지 인스턴스 층에만 적용이 되고 타입 층에 적용되지는 않는다. 이것은 타입 계층을 따라 다수의 타입들이 존재한다고 가정할 뿐, 그들 타입간의 관계를 명시적으로 정의하지는 않음을 의미한다. 하지만, KAH에서는 일반화와 세분화 관계가 인스턴스 층과 타입 층 모두에 적용된다. 따라서 서브 타입의 한 인스턴스는 자신의 수퍼 타입에 존재하는 특정 추상화 인스턴스와 관계를 갖고, 추상화 인스턴스의 도메인은 서브 타입의 인스턴스의 수퍼 도메인으로 간주된다.

둘째, 추상화 계층의 변화를 관리하는데 있어, KAH 추상화 정보 데이터베이스는 다른 기법들보다 두 가지 면에서 더욱 효과적이다. 먼저, 값의 추상화 지식을 관리하는 VALUE_ABSTRACTION 릴레이션은 정규화된 릴레이션으로서 다른 기법들의 비정규화된 릴레이션들에 비해 여러 가지 장점을 갖는다. 정규화를 가능케 하는 근본적인 원인은 하나의 투플에 값의 1단계 추상화 관계만을 나타내기 때문이다. 다단계 추상화 정보는 여러 개의 1단계 추상화 투플들을 순환적으로 연결함으로써 얻어질 수 있다. 이러한 정규화의 결과로, 추상화 계층에 존재하는 값의 변화는 변화되는 값을 포함하는 투플들을 갱신함으로써 간단하게 처리될 수 있다. 이와는 대조적으로, 타입 추상화 계층을 포함한 다른 기법들은 하나의 투플에 다단계의 값의 추상화 지식을 표현한다. 그로인해 정규화되지 못한 릴레이션 스키마를 갖게 되고, 추상화 계층에 존재하는 값들이 변화하게 될 때 삽입, 갱신, 삭제 이상을 불러 일으킨다.

다음으로, KAH-기반 기법은 지식 추상화 계층의 구조와는 독립적인 추상화 정보 데이터베이스를 제공한다. 다시 말해서, 추상화 계층에 도메인을 추가하고 삭제함으로써 발생하는 지식의 변화는 VALUE_ABSTRACTION 릴레이션과 DOMAIN_ABSTRACTION 릴레이션에서 평범한 데이터베이스 트랜잭션을 수행함으로써 처리된다. 하지만 객체지향 데이터베이스 기법과 다른 개념적 분류 기법에서는, 추상화 계층 구조가 지식 베이스의 스키마에 파묻히게 되고, 계층의 구조적인 변화는 데이터베이스 스키마 자체의 변화를 요구하게 되어, 전체적인 추상화 정보 데이터베이스 재구성의 결과를 낳게 된다.

셋째, 논리적 질의 완화 규칙에 바탕을 둔 규칙 기반 기법은 질의 완화 과정의 직관성이 떨어지고, 사용자와의 상호 작용을 통하여 질의 완화 과정을 제어하기 어렵다[3, 9, 11]. 하지만, KAH 기반 추상화 정보 데이터베이스는 다른 기법들보다 좀 더 유연한 질의 변환을 수행할 수 있게 해 준다. 이것은 추상화 정보 데이터베이스가 단지 질의완화에 직접적으로 사용되는 값의 추상화 지식만을 관리하는 것이 아니고, 값의 도메인, 도메인의 수퍼 도메인, 도메인의 추상화 수준과 같은 도메인 관련 지식을 추가적으로 관리하기 때문이다. 추상화 수준에 관한 정보는 질의 완화 과정에서 사용자가 원하는 정도의 질의 완화를 점진적으로 혹은 직접적으로 수행할 수 있게 해준다. 결과적으로 보다 유연한 질의 완화 제어를 사용자에게 제공함으로써, 보다 효과적이고 효율적인 질의 처리가 이루어진다.

5. 결 론

본 논문에서는 다양한 유형의 협력적 질의 응답 뿐 아니라 보다 효과적인 지식 관리를 지원하기 위하여, 지식 추상화 계층(KAH)이라는 지식 표현 프레임웍을 제시하였다. KAH는 값의 추상화와 도메인의 추상화라는 두 가지 종류의 데이터 추상화를 기본 개념으로 갖는 다단계 지식 추상화 계층이다. 값의 추상화는 세분값과 추상값 사이의 일반화/세분화 관계를 표현한다. 이러한 값의 추상화를 통하여 최하위 수준에 존재하는 값이 가장 세분화된 값이 되고, 최상위 수준에 존재하는 값이 가장 추상화된 값이 되도록 값의 추상화 계층을 점진적으로 구축해 나간다. 도메인 추상화는 수퍼 도메인과 서브 도메인 사이의 추상화 관계를 표현하고, 값의 추상화 계층과 병렬적으로 도메인 추상화 계층을 구축해 나간다.

KAH에 포함된 의미론적 지식을 관리하기 위하여, 관계형 데이터 모델을 바탕으로 추상화 정보 데이터베이스를 설계하였다. 구축된 추상화 정보 데이터베이스는 KAH에 포함된 지식의 변화를 동적으로 관리해 줄 수 있음은 물론, 보다 다양한 형태의 질의에 대한 협력적 질의 응답을 효과적으로 수행할 수 있게 해준다. 지식 추상화 계층을 구성하는 두 가지 추상화 계층 중에서, 협력적 질의응답의 핵심 과정인 질의 완화와 좀 더 밀접한 관계를 갖는 것은 값의 추상화 계층이라고 할 수 있다. 질의 완화는 질의의 결과가 좀 더 일반적이고 유사한 값까지 포함할 수 있도록 질의 조건을 확장함으로써 이루어진다. 한편, 도메인 추상화 계층은 KAH의 내용과 구조상의 동적인 변화를 처리해 주는 KAH의 관리 측면과 관련이 있다. 다시 말해서, KAH의 구조와는 독립적으로 추상화 정보 데이터베이스가 구축되고, 두가지 추상화 계층 모두가 정규화된 릴레이션으로 구현되기 때문에, KAH의 구조상의 변화는 추상화 정보 데이터베이스 스키마를 변화시킬 필요 없이 단지 도메인의 삽입, 삭제와 같은 일반적인 데이터베이스 연산을 통해서 처리된다.

앞으로의 연구에서는, KAH와 질의 완화 과정에 대한 형식론의 확립과 확장된 SQL의 개발이라는 두 가지 분야에 초점을 맞추고자 한다. KAH의 형식론은 본 논문에서 검토된 일반화, 세분화와 같은 질의 완화 과정이 완전하고 정확한 것임을 증명하기 위하여 관계 해석에 기초를 두고 개발되고 있는 중이다. 협력적 질의 응답의 측면에서는 근사 조인 연산에 관련된 질의 완화 과정이 추가적으로 연구될 것이다. 또한, KAH 및 추상화 정보 데이터베이스의 유연성과 사용자 친숙성을 향상 시키기 위하여 확장 SQL과 보다 범용적인 연산자가 개발될 것이다.

REFERENCE

[1] Y. Cai, N. Cercone, and J. Han, "Attribute-Oriented Induction in Relational Databases", in Knowledge Discovery in Databases, AAAI Press/The MIT Press, 1993.

[2] Q. Chen, W. Chu and R. Lee, "Providing Cooperative Answers via Knowledge-Based Type Abstraction and Refinement", in Proc. of the 5th International Symposium on Methodologies for Intelligent Systems, Knox-

ville, TE, 1990.

[3] L. Cholvy and R. Demolombe, "Queryng a Rule Base," in Proc. 1st Int. Conf. Expert Database Syst., 1986, pp.365-371

[4] W. Chu, H. Yang, K. Chiang, M. Minock, G. Chow and C. Larson, "CoBase: A Scalable and Extensible Cooperative Information System", International Journal of Intelligence Information Systems. Vol.6, 1996.

[6] W. Chu and Q. Chen, "A Structured Approach for Cooperative Query Answering", IEEE Transactions on Knowledge and Data Engineering, Vol.5, No.5, October 1994.

[7] W. Chu, R. Lee, and Q. Chenm, "Using Type Inference and Induced Rules to Provide Intensional Answers", in Proc. of the 7th International Conference on Data Engineering, Kobe, Japan, April 1991.

[8] W. Chu, Q. Chen, and R. Lee, "Cooperative Query Answering via Type Abstraction Hierachy", in Proc. of the 1990 International Conference on Cooperative Knowledge Base System, Springer-Verlag, 1991.

[9] F. Cuppens and R. Demolombe, "Cooperative Answering: A Methodologies to Provide Intelligent Access to Databases", in Proc. of 2nd International Conference on Expert Database Systems, Huntsville, Alabama, October 1990.

[10] P. Godfrey, J. Minker, and L. Novik, "An Architecture for a Cooperative Database System", Proceedings of the 1994 International Conference on Applications of Databases, June, 1994.

[11] A. Hemerly, M. Casanova, and A. Furtado, "Exploiting User Models to Avoid Misconstruals," Nonstandard Queries and Nonstandard Answers, Oxford Science Publications, pp.73-98, 1994.

[12] S. Y. Huh and K. H. Moon, "A Knowledge Abstraction Hierarchy for Cooperative Query Answering", Submitted for Publication.

[13] S. Y. Huh and J. W. Lee, "Providing Approximate Answers Using a Knowledge Abstraction Hierarchy", Submitted for Publication.

[14] Ichikawa, T. ARES: a relational database with the capability of performing flexible interpretation of queries. IEEE Transactions on Software Engineering, SE -12, 5(May 1986).

[15] A. K. Jain and R. C. Dubes. Algorithms for Cluster Analysis. Prentice Hall, Englewood Cliffs, NJ, 1998.

[16] M. J. Minock and W. Chu, "Explanation for Cooperative Information Systems", Proceedings of Ninth International Symposium on Methodologies for Intelligent Systems. June 1996.

[17] A. Motro, "VAGUE: A User Interface to Relational Database that Permits Vague Queries", ACM Transactions on Office Information Sustems, Vol.6, No.3, July 1988.

[18] A. Motro, "FLEX: A Tolerent and Coo-

perative User Interface to Databases", IEEE Transactions on Knowledge and Data Engineering, Vol.2, No.2, June 1990.

[19] A. Motro. "Intensional Answers to Database Queries", IEEE Trabnsactions on Knowledge and Data Engineering, Vol.6, No.3, June 1994.

[20] A. Scheer, Enterprise-Wide Data Modeling: Information Systems in Industry, Springer -Verlag, 1989.

[21] C. D. Shum and Muntz, "An Information -Theoretic Study on Aggregate Responses", in Proc. 14th International Conference on Very Large Databases, Los Altos, CA: Morgan Kaughmann, 1988.

[22] E. Turban, Decision Support and Expert Systems: Management Support Systems, Macmillan, 1988.

[23] J. D. Ullman, Database and Knowledge-Base Systems, Vol.1, Computer Science Press, 1987.

[24] S. V. Vrbsky and W. S. Liu, "APPROXIMATE-A Query Processor that Produces Monotonically Improving Approximate Answers", IEEE Transactions on Knowledge and Data Engineering, Vol.5, No.6, December 1993.