



데이터 요소의 표준화(Ⅲ)

백두권

고려대 전산학과 교수

차례

I. 총론

II 데이터 요소의 생성과 표준화를 위한 프레임워크

▶ III. 도메인 식별을 위한 개념의 분류

IV. 기본 데이터 요소 속성

V. 데이터 정의의 형식화를 위한 규칙과 지침

VI. 데이터 요소를 위한 명명 및 식별 원칙

VII. 데이터 요소의 등록

III. 도메인 식별을 위한 개념의 분류

1. 배경 및 범위

현재 SC14에서 가장 활발하게 연구되고 있는 ISO 11179 표준은 국제사회에서 전역 전자정보 교환을 통한 데이터의 개방된 교환과 데이터를 용이하게 얻기 위해 표준화된 방법으로 데이터를 형식화하는 데 사용되는 이산 데이터 요소 서술과 의미론적 내용(메타 데이터)의 형식화와 유지보수의 구체적인 지침을 제시하는 것을 목적으로 한다. 이 표준은 6개의 부분 즉, 데이터 요소의 생성과 표준화를 위한 프레임워크(ISO 11179-1), 도메인 식별을 위한 개념의 분류(ISO 11179-2), 기본 데이터 요소의 속성(ISO 11179-3), 데이터 정의의 형식화를 위한 규칙과 지침(ISO 11179-4), 데이터 요소를 위한 명명

및 식별원칙(ISO 11179-5) 그리고 데이터 요소의 등록(ISO 11179-6)으로 구성되어 있다.

ISO 11179 표준중에서 제 2부 도메인 식별을 위한 개념의 분류는 데이터 요소가 의미하는 개념을 분류하기 위한 지침을 규정하고 있다. 즉 데이터 요소의 개발에 기초가 되는 개념 분류를 위한 방법론을 규정하고 있다. 이 표준을 연구하게 된 동기는 데이터 요소의 개념 분류를 위한 표준화 절차가 요구되고 있고 개방형 전자 데이터 교환의 지원하에 호환성있고 구조화된 데이터 요소의 도출을 보장하는 메카니즘을 지원해야 하는 필요성이 제기되었기 때문이다.

이 표준은 데이터 요소가 그 분류를 상속할때의 기반이 되는 개념을 분류하기 위한 방법을 명세한다. 예비과정으로서 개념 분류화를 도입하는 것은 데이터 요소 개념의 속성이 존재한다는 것을 가정한다. 이 데이터 요소 개념의 속성은 속성

과 그 값을 선택하는데 이용가능한 선택사항을 제한함으로써 상속성으로부터 직접적으로 데이터 요소의 구조에 영향을 미칠 수 있다. 이 표준에서 제공하는 분류 절차를 준수함으로써 데이터 요소의 의미를 명료하게 할 수 있고 데이터 요소의 무결성을 유지하며 설계할 수 있다.

2. 데이터 요소 분류 관점

데이터 요소의 분류는 다음과 같이 여러가지 관점에서 볼 수 있다.

1) 관점 1

데이터 요소를 단순 데이터 요소, 복합 데이터 요소 또는 합성 데이터 요소중의 하나로 분류하는 관점이다. 예를들어, 단순 데이터 요소의 예로는 연과 달에 따라 01 부터 28, 29, 30, 31의 값을 갖는 한달의 날짜를 들 수 있다. 이 관점은 또한 복합 데이터 요소도 포함한다. 즉, 010101 부터 311299의 값을 가지고 왼쪽으로부터 두자리씩 날짜, 월, 연도를 나타내게 한다. 이 예에서 01부터 12까지의 값을 가지는 달은 합성 데이터 요소가 되기 위해 고려되어야 할 날짜를 결정하는 데 사용된다. 이 관점의 다른 예는 시간의 상세화이다. 여기서 시간은 시, 분, 초, 그리고 더 정확한 표현을 위해 10분의1초, 100분의 1초, 1000분의 1초로 구성된다. 만일 그 값이 다른 복합 데이터 요소의 값의 일부분으로 사용된다면 단순 데이터 요소든 복합 데이터 요소든 합성 데이터 요소로 볼 수가 있다.

2) 관점 2

이 관점에서는 데이터 요소를 주제 영역에 따라 분류한다. 의학 데이터 요소, 은행 데이터 요소, 무역 데이터 요소, 군사 데이터 요소, 환경 데이터 요소의 집단을 그 예로 들 수 있다. 이러

한 접근 방법을 사용할 경우에는 서로 다른 영역에 있어서의 빈번한 겹침이 발생한다. 따라서 개방된 전자적 데이터 교환을 위해서는 서로 다른 영역에 있어 데이터 요소 구조에 대한 조정이 필요하다. 주제가 특정 조직에 관련된 것이라면, 또한 그 요구를 충족시킬 수 있는 데이터 요소의 수집을 고려해야 한다.

3) 관점 3

이 관점에서는 정보를 제공하는데 있어 수행되는 기본적 기능에 따라서 데이터 요소로 발전할 수 있는 모든 개념을 분류할 수 있다. 모든 주제 영역에 속해 있는 개념들은 모두 이 관점에 속하며 응용과는 무관한 문맥에서 생각된다. 적합한 데이터 요소가 개방된 전자적 데이터 교환을 지원할 수 있도록 유도되어야 한다는 데에서부터 일반적인 기반이 발전되어야 한다는 것인 이와 같은 인식이 필요하다. 그러나 모든 개념이 데이터 요소로 부터 발전한 기반으로 역할해야 하는 것은 아니다.

검증된 개념을 선택하기 위해 평가 공정이 필요하며 전체 지식내에 있는 모든 개념이 전부 잠정적인 후보이다. 전체 지식은 세계에 존재하는 모든 지식의 덩어리로 볼 수 있다. 알려지지 않은 지식은 전체 지식의 일부분이며 아직 인간에게 알려지지 않은 지식이다. 인간의 지식은 전체 지식중에 알려진 부분이며 객체, 객체의 속성들, 그리고 객체와 객체사이의 관계에 부여된 값으로 표현되어 현재 보여지는 정보로도 볼 수 있다.

객체는 공간이나 시간적으로 물리적인 것이든, 마음에 의해 형성되는 개념적인 것이든지 사람의 두뇌 활동에 의해서 인식되어지는 것이다. 객체는 현상이나 실체로 여겨지기도 한다. 현상은 외부적인 공간이나 시간 상에서 감각에 의해 인식되는 객체이다. 실체는 이성에 의해 인식되고 그 결과로 사고되어지지만 느낌만으로는 알 수 없는



객체이다. 현상이나 실체는 이름이나 의미 등 그것과 연관된 특성에 따라 특징 지워지며, 이들 특성은 속성이라고도 한다.

속성화 된 객체는 그 객체를 서술하는 정성적이거나 정량적인 속성과 그 값이 연결되도록 되어진 객체를 말한다. 속성화된 객체는 아이디어를 생성하게 되는데 아이디어는 어떤 것에 대한 실제적으로 감각을 통해 인지되는 한 순간 또는 다른 순간의 이미지이거나 인지되지 않았더라도 약간의 정보만으로 시각화 될 수 있는 이미지이다.

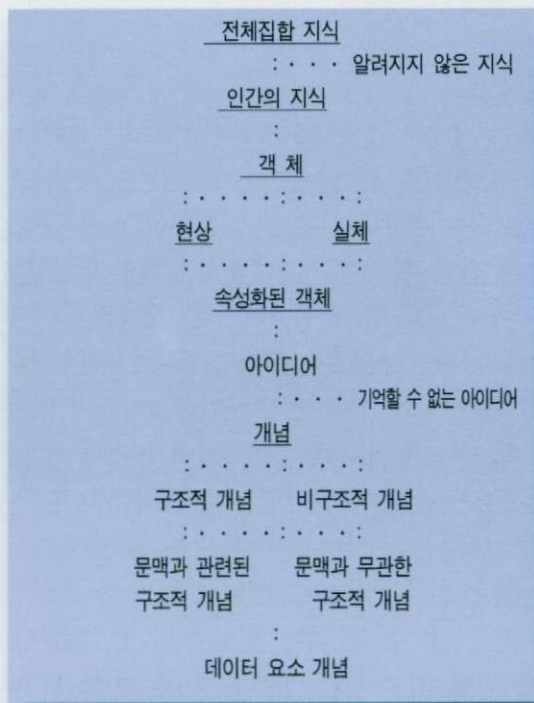
기억할 수 없는 아이디어는 회상할 수 없거나 따라서 인식하지 못하는 아이디어이다. 객체가 그 속성에 의해 커지고, 빈번히 만나게 되어 아이디어로 인식되어 질때, 이 객체는 개념으로 여겨진다. 개념은 전형적인 것이나 일반적인 것을 제외한 세부적인 것이 부족한 영역에 대해 많은 예들을 기억 할 수 있게 된 다음, 마음이 인식하게 되는 것에 대한 아이디어이다. 개념은 언어에 의해 표현될 수 있으며, 독자성을 주는 의미와 이름을 제공하는 정의와 같은 속성의 값을 이루기 위한 분산된 문자로 표현되기도 한다.

개념은 구조적이거나 비구조적일 수 있다. 비구조적인 개념은 연결된 인스턴스의 집합을 가지고 있지 않다. 구조적 개념은 연결된 인스턴스의 집합을 가지고 있다. 예를 들어 “색”이라는 개념이 있다고 하면 이것은 색들의 집합과 연결될 수 있다. 구조적 개념은 문맥과 관련하여 또는 관계 없이 발생한다. 개념 “색”은 그 질에 대한 부가적인 단어가 아무것도 주어지지 않았다면 문맥과 무관한 것이 된다. 하지만 그 개념이 “사람눈에 대한 색”이나 “사람의 머리카락 색”이 된다면 문맥은 각 개별적인 집합에 대하여 생각할 수 있도록 해준다. 문맥과 무관한 구조적 개념은 데이터 요소의 형태로 표현될 수 없다. 하지만 문맥과 관련있는 구조적 개념은 데이터 요소의 형태로 표

현 가능하며 이것이 바로 데이터 요소 개념이다.

자동화된 데이터 처리 장치가 사회에 도입 됨으로써 많은 개념을 그 속성의 값들이 컴퓨터가 처리 할 수 있도록 구조화해야할 필요가 증대되게 되었다. 개념을 구성하는 모든 속성과 그에 연결된 값들이 모여져 이러한 형태의 구조를 이룬다면 이것이 데이터 요소이다. <그림 1>은 데이터 요소 개념에 대한 진화를 그림으로 나타낸 것이다.

<그림 1> 데이터 요소 개념의 진화



3. 데이터 요소 개념 속성

데이터 요소 개념은 그들의 속성에 의해 유일하게 특성화된다. 이것은 데이터 요소 개념들을 올바르게 분류하게 하는 토대가 된다. 데이터 요소 개념에 배정된 분류는 그 분류로부터 유도된 포괄적 데이터 요소에 의해 상속되며 또한 포괄

적 데이터 요소로부터 유도된 응용 데이터 요소에 의해 상속된다.

데이터 요소 개념은 데이터 요소 개념 속성이라 불리는 알려진 특성을 갖는다. 이들 속성에 배정된 값은 그것들을 정량화 또는 정성화한다. 특정 데이터 요소 개념과 연관된 속성들의 모임을 속성집합이라 한다. 각 데이터 요소 개념은 각각의 속성 집합에 대해 동일한 값을 갖지 않기 때문에 유일하다. 데이터 요소 개념 속성의 세심한 선택과 설계 그리고 데이터 요소 개념 설계자에 의해 할당된 값의 비교 및 검증은 결과 데이터 요소 개념이 유일하게 구분되고, 엄밀히 정의되며, 정확하고 완전히 명명되며, 올바르게 위치하며 분류 시스템 내에서 관련된 다른 데이터 요소 개념과 제대로 연관될 수 있게 해준다.

데이터 요소 개념의 속성 집합은 데이터 요소 설계자가 포괄적 데이터 요소의 속성집합을 설계하고자 할 때 정보를 도출하기 위해 처음으로 탐색해야 하는 가장 중요한 원시 참조자료이다. 구조 및 내용 정보 모두 데이터 요소 개념 속성으로부터 그들의 대응되는 포괄적 데이터 요소 속성에 의해 상속된다.

1) 데이터 요소 개념 속성의 유형

데이터 요소 개념 속성은 그것이 언급하는 데이터 요소 개념의 특성에 따라 관리, 관계 또는 표현 데이터 요소 개념 속성으로 구분된다.

① 관리 데이터 요소 개념 속성

관리 데이터 요소 개념 속성은 데이터 요소 개념의 통제, 개발, 유지보수, 운영, 구성 또는 사용과 관련된 특성을 나타내는 속성이다.

② 관계 데이터 요소 개념 속성

관계 데이터 요소 개념 속성은 데이터 요소 개념과 하나 또는 그이상의 다른 데이터 요소 개념 속성과의 관련성에 대한 특성을 나타내는 속성이

다.

③ 표현 데이터 요소 개념 속성

표현 데이터 요소 개념 속성은 데이터 요소의 구조를 나타내는 속성이다.

2) 데이터 요소 개념 속성의 성질

데이터 요소 개념 속성은 데이터 요소 개념 등록 권한에 따라 필수 또는 선택 데이터 요소 개념 속성으로 구분된다.

① 필수 데이터 요소 개념 속성

필수 데이터 요소 개념 속성은 데이터 요소 개념의 속성집합에서 필수적인 데이터 요소 개념의 특성을 나타낸다. 모든 필수 데이터 요소 개념 속성은 분류 시스템을 포함하기 위한 등록 권한에서 고려될 수 있도록 후보 데이터 요소 개념의 구조에 필수적으로 나타나야 한다.

② 선택 데이터 요소 개념 속성

선택 데이터 요소 개념 속성은 데이터 요소 개념의 속성집합에 필수적이지는 않지만 데이터 요소 개념 설계자의 의함에 따라 포함될 수도 있는 특성을 나타낸다. 선택 데이터 요소 개념 속성은 분류 시스템을 포함하기 위한 등록 권한에서 고려되기 위해 후보 데이터 요소 개념의 구조에 필수적으로 나타나야 하는 것은 아니다.

4. 데이터 요소 개념 분류

데이터 요소 개념 분류법은 포괄적 데이터 요소의 개발을 지원할 수 있는 특정 데이터 요소 개념의 연관, 분류, 정의, 구분, 명명 및 순서화를 위한 보조기능을 하는 체계적인 구조로 정의된다.

1) 데이터 요소 개념 분류 계층

데이터 요소 개념 분류법은 데이터 요소 개념을 삽입 및 조회할 수 있는 계층적 구조를 제공한

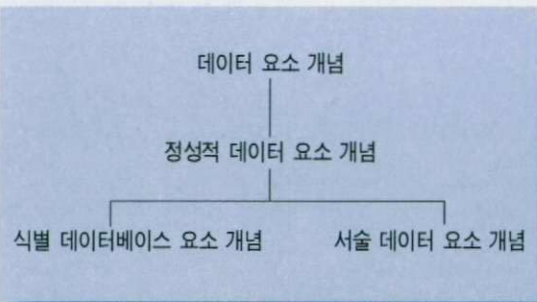


다. 분류과정 동안 각 데이터 요소 개념은 개념이 나타내는 주제의 정제도에 따라 개념적 매크로 클래스, 하나 혹은 그이상의 개념적 하부 매크로 클래스, 개념적 클래스 그리고 개념적 하부 클래스의 하나 혹은 그이상의 수준과 연관된다.

데이터 요소 개념 분류법 내에서 발생하는 데이터 요소 개념은 처음에 정성적 데이터 요소 개념 또는 정량적 데이터 요소 개념이라 불리는 둘 중 하나의 매크로 클래스내에서 분류된다. 이 규칙의 예외는 각각의 사실적 표현이 free-text 형식으로 나타나거나 정성적 또는 정량적인 사항을 언급하거나 기정의된 특성 집합과 특성의 최대수에 의해 제한될때 free-text 데이터 요소 개념이라 한다.

2) 정성적 데이터 요소 개념

정성적 데이터 요소 개념은 데이터 요소 개념 각각의 사실상의 표현이 식별이나 서술에 의해 특정 실체나 실체의 클래스를 경계짓고 수정하거나 제한하는 데이터 요소 개념이다. 각 정성적 데이터 요소 개념은 식별 데이터 요소 개념이나 서술 데이터 요소 개념중 하나의 적당한 개념적 하부 매크로 클래스로 분류되어야 한다. 다음의 스키마는 상위 정성적 수준에서 데이터 요소 개념 분류의 관점을 보여준다.



① 식별 데이터 요소 개념

식별 데이터 요소 개념은 각각의 사실상의 표

현이 특정 실체나 실체의 클래스에 명사 혹은 코드화된 요소를 치환함으로써 일치성을 증명하는 방법으로 동일성 및 관련성을 보여주는 데이터 요소 개념이다. 하나의 실체 혹은 실체의 클래스는 코드화된 요소와 같은 명사 혹은 명사를 위한 대응물로 식별한다. 실체는 행위, 동물, 아이디어, 사건, 사람, 장소, 식물, 양, 관계, 상태, 물질 혹은 사물이 될 수 있다. 각 식별 데이터 요소 개념은 명사적 식별 데이터 요소 개념 혹은 코드화된 식별 데이터 요소 개념 중 하나의 적당한 개념적 클래스로 분류되어야 한다.

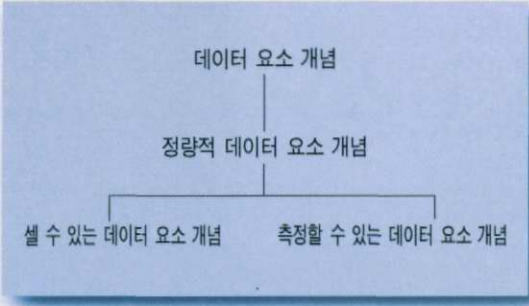
② 서술 데이터 요소 개념

서술 데이터 요소 개념은 사실 표현의 각각이 특정 실체나 실체들의 한 클래스 혹은 특정 실체들 사이나 실체들의 클래스들 사이의 내부나 외부의 서술적 형태에 근거한 양적인 관계의 내부나 외부의 서술적 형태를 언급하는 데이터 요소 개념이다. 각각의 서술적 데이터 요소 개념은 서술적으로 특징지을 수 있는 데이터 요소 개념이나 서술적으로 분류할 수 있는 데이터 요소 개념 중 어느 하나로써 적당한 개념 클래스로 세분화되어야만 한다.

3) 정량적인 데이터 요소 개념

정량적인 데이터 요소 개념은 사실표현의 각각이 특정한 실체나 실체들의 클래스를 표준 단위로 측정된 카운트나 측정된 실제적이거나 추정된 값 또는 특정한 실체들 사이나 실체들의 클래스들 사이의 양적인 관계에 대한 실제적인 혹은 추정된 값인 데이터 요소 개념이다.

정량적인 데이터 요소 개념은 기본적으로 경제적, 물리적, 심리학적, 혹은 사회적인 양을 언급한다. 각 양적인 데이터 요소 개념은 헤아릴 수 있는 데이터 요소 개념이나 측정할 수 있는 데이터 요소 개념중 어느 하나로 그것에 적당한 개념적 하부 매크로 클래스로 세분화되어야만 한다.



① 셀 수 있는 데이터 요소 개념

셀 수 있는 데이터 요소 개념은 사실 표현의 각개이 특정한 실체들이나 실체들의 클래스들을 표준 단위로 계산한 총 카운트에 대한 실제적인 혹은 추정된 값인 데이터 요소이다. 각 셀 수 있는 데이터 요소 개념은 실제적인 카운트 데이터 요소 개념이나 추정된 카운트 데이터 요소 개념 중 어느 하나로 그것에 적당한 개념 클래스로 세분화되어야만 한다.

② 측정할 수 있는 데이터 요소 개념

측정할 수 있는 데이터 요소 개념은 사실 표현 각개이 특정 실체나 실체들의 클래스의 표준 단위로 측정할 수 있는 실제적이거나 추정된 값인 데이터 요소 개념이다. 각 측정할 수 있는 데이터 요소 개념은 실제 측정 데이터 요소 개념이나 추정된 측정 데이터 요소 개념 중 어느 하나로 그것에 적당한 개념 클래스로 세분화되어야만 한다.

5. 데이터 요소 개념 주격 관계

데이터 요소내의 많은 양적, 질적 데이터 요소 개념 사이에 존재하는 주격 관계는 지배적, 협조적, 그리고 종속적중 어느 하나로 관찰될 수 있다. 지배적 데이터 요소 개념은 직접적으로 하위 분류적 수준에 있는 하나 이상의 직접적으로 관련된 데이터 요소 개념들을 가진 것이다. 협조적 데이터 요소 개념은 같은 분류적 수준에 있는 하나 이상의 직접적으로 관련된 데이터 요소 개념들을 가진 것이다. 종속적 데이터 요소 개념은 직접적

으로 우위에 있는 수준에 있는 직접적으로 관련된 데이터 요소 개념을 가진 것이다. 종속적 데이터 요소 개념은 그것의 지배적인 데이터 요소 개념들의 성질을 그대로 상속한다.

일단 후보 데이터 요소 개념이 그것은 필요한 속성들을 할당 받고 등록 권한을 인정 받으면, 그것은 유사한 데이터 요소 개념들에 근접하게 데이터 요소 개념 분류법 내로 분류될 것이다. 이것들은 자주 어떤 유사하거나 동일한 속성들을 공유할 것이다.

어떤 것이 주의 깊게 이러한 속성들과 그것들의 관련되 값들을 인지했다면 그것은 자주 어떤 제한 조건들이 예상되는 데이터 요소 개념의 설계에 영향을 미칠 수 있다가 명확해질 것이다. 제안된 일련의 개념들은 종종 데이터 요소 설계자가 보충되는 이름, 정의 또는 다른 후보 데이터 요소 개념에 대한 데이터 속성을 개발하는 것을 지원 할 수도 있는 데이터 요소 분류법내에 이미 등록된 데이터 요소 개념들로부터 이용 가능하다. 그러한 데이터 요소 개념 속성 비교는 데이터 요소 개념 분류법 내에서의 진보된 개념의 호환의 출현을 확신케 하는데 도움이 된다.

속성을 가진 데이터 요소 개념이 일반적인 데이터 요소로 변형되도록 선택되는 것은 바로 데이터 요소 개념 분류의 개념 하부 클래스로부터 나온다. 어떤 주어진 데이터 요소 개념에 대한 데이터 요소 개념 속성과 그것들의 관련된 값들은 일반적인 데이터 요소 설계자에게 일반적인 데이터 요소 속성 집합속에 포함 되어져야 할 어떤 일반적인 데이터 요소 속성들과 속성 값들을 제안 할 것이다. **DC**