



데이터 요소의 표준화(VI)

정인기

고려대 전산학과

차례

- I. 총론
- II 데이터 요소의 생성과 표준화를 위한 프레임워크
- III. 도메인 식별을 위한 개념의 분류
- IV. 기본 데이터 요소 속성
- V. 데이터 정의의 명식화를 위한 규칙과 지침
- ▶ VI. 데이터 요소를 위한 명명 및 식별 원칙
- VII. 데이터 요소의 등록

I. 데이터 요소를 위한 명명 및 식별 원칙

1. 적용 범위 및 분야

이 표준은 데이터 요소의 명명 및 식별을 위한 원칙, 규칙 및 지침을 포함하며 데이터 요소 식별의 원소와 구조를 서술한다. 식별은 협의로는 등록 안에서 데이터 요소의 유일한 식별을 설정하기 위한 수단만으로 정의된다.

이 표준은 데이터 요소를 속성화하고, 분류하고, 정의하고, 등록하기 위한 규칙과 절차를 설정하는 것들을 가진 결합에서 사용될 수 있는 수단이 될 수 있다.

2. 데이터 구조의 식별을 위한 원칙

2.1 속성의 식별

네 가지의 관련된 속성들의 집합이 데이터명

으로서 제공되고 등록할 때 데이터 요소를 차별화하여 각 데이터 요소를 식별한다. 속성들은 다음과 같다.

- 이름
- 문맥
- 등록 식별자
- 등록 기관

이 속성들의 기본 원칙들은 다음과 같다.

2.2 이름과 문맥

각 데이터 요소는 하나 혹은 그 이상의 이름을 가진다. 이름은 특정한 문맥 내에서 특수한 활용을 가진다. 엄격하게 구조화된 이름은 데이터 관리 조직을 위해 생성될 수 있고, 선정된 이름은 사용자에게 의해 명세될 수 있으며, 단축된 이름은 특정한 프로그래밍 언어나 데이터베이스 관리 시스템과 같은 특정한 소프트웨어 시스템을

위해 생성될 수 있다.

많은 데이터 요소를 위한 이름들은 각 문맥 안에서 개발될 수 있다. 명명법(보통 규칙의 집합)은 이름들이 그 문맥 안에서 어떻게 형식화되는가를 명세하기 위하여 각 문맥을 위해 설정된다. 명명법은 그 문맥의 모든 관계있는 측면을 망라한다. 또한 다음과 같은 것들에 적용될 수 있다.

- a. 문맥의 목적
- b. 이름을 설정하는 기관
- c. 이름에서 사용되는 단어의 출처와 내용을 감독하는 의미론적 규칙, 예를 들어 데이터 모델에서 파생된 단어, 규율에서 공통적으로 사용되는 단어 등
- d. 요구되는 단어 순서를 망라하는 구문론적 규칙
- e. 제어 단어 목록, 이름의 길이, 문자 집합, 언어 등을 망라하는 어휘론적 규칙
- f. 이러한 문맥 안에서 이름들이 유일해야 하는가를 설정하는 규칙

명명법에 대한 이러한 규칙은 4장에서 설명하기로 한다.

2.3 등록 식별자와 등록 기관

하나의 데이터 요소에 대해서 적어도 하나의 유일한 등록 식별자가 요구된다. 등록 식별자는 데이터 요소의 도메인과 정의가 변할 때까지 변하지 않는다. 각 등록 식별자는 등록 기관에 의해 할당되며 등록 기관 안에서 유일하다.

데이터 요소가 복수의 등록 기관으로부터 등록 식별자를 할당받는다면 등록 식별자와 등록 기관은 데이터 요소의 식별을 위해 모두 필요하다. 데이터 요소의 특정 측면이 변하면 데이터 요소는 새로운 버전이 등록된다.

등록 식별자는 데이터 요소를 식별하는 보편적인 수단이며 정보 시스템, 조직, 혹은 특정 데

이터를 공유하기를 원하지만 동일한 이름이나 문맥을 이용할 수 없는 다른 부류들 간에 데이터를 교환하기 위한 기반을 제공할 수 있다. 유일한 등록 식별자는 또한 등록 식별자가 하나 이상의 자연어에 대해 설정된 문맥과 관련될 때 언어 번역을 위해서도 유용하다.

등록 기관을 위한 요구 사항은 이 국제 표준의 ISO 11179-6에 명세되어 있다.

3. 데이터의 등록 식별을 위한 규칙

데이터의 등록 식별을 위한 규칙은 다음과 같다.

① 각 데이터 요소는 등록 기관 안에서 유일한 등록 식별자를 가져야 한다.

② 등록 식별자와 등록 기관의 조합은 데이터 요소에 대하여 유일한 식별자가 되어야 한다.

③ 등록 식별자를 할당하기 위해서 데이터 요소는 다음과 같은 규칙을 따라야 한다.

- ISO 11179-2로부터 파생되어야 하고,
- ISO 11179-3을 따르는 속성이어야 하고,
- ISO 11179-4에 따라 정의되어야 하고,
- ISO 11179-5에 따라 명명되어야 하며,
- ISO 11179-6에 따라 등록되어야 한다.





④ 데이터 요소는 문맥 안에서 적어도 하나 이상의 이름을 가져야 한다.

4. 구조적 명명법을 위한 지침

특정 문맥을 위한 구조적 이름을 엄격하게 생성하기 위한 명명법을 개발하는데 사용될 수 있는 지침은 다음과 같다. 명명법의 예제는 부록에서 다룬다.

규칙은 이름이 개발됨에 따라 원칙으로부터 파생되어진다. 이들 규칙에 따라 형성된 이름들은 간략하게 된 구문 때문에 원래의 것보다 다른 언어로 쉽게 번역될 수 있다. 구문론적, 의미론적 및 어휘론적 규칙은 회사 혹은 표준 설정 조직과 같은 조직에 따라 많이 변형된다. 각각은 그들의 문맥 안에서 이름에 대한 규칙을 설정할 수 있다.

데이터 요소 이름은 원소의 이름으로부터 형성되어야 하며 각각은 하나의 이름 안에서 상대적 혹은 절대적 위치(구문론적으로)와 의미(의미론적으로)를 할당받는다. 그것들은 어휘론적 규

칙에 영향을 받는다. 그들은 또한 꼭 필요한 것은 아니지만 분리 기호로 경계를 정할 수 있다.

각 원소의 값의 도메인, 집합 혹은 범위들은 하나의 기관, 예를 들어, 국제적 사업 부문의 명명 표준을 위한 회사나 인정받은 위원회 안에서의 데이터 관리기관에 의해 엄격하게 통제받아야 한다.

4.1 이름의 의미론적 내용을 관리하는 원칙

4.1.1 원소의 의미

원소는 의미가 논리적 참조에 따라 적어둔 것에 대한 이상적인 용어들로 구성된다. 원소들은 지시대상(Referent), 특성(Property) 및 수식어구(Modifier)로써 서술된다.

(1) 지시대상 (Referent)

개체-관계 도표(ERD, Entity Relationship Diagram) 혹은 객체 모델과 같은 모델 방법론의 사용은 보다 높은 수준의 모델 개체에 관하여 모든 데이터 요소를 이상적으로 위치시키는 방법이다. 개체-관계 모델 개체의 속성들은 방법론의 응용에 따라 서로 다른 것과 관련되는 데이터 요소와 같다. 객체 모델에서 데이터 요소는 객체 속성으로써 표현된다.

이들 모델 속성들의 많은 것들은 모델 개체 사이의 상수로 남아 있다. 예를 들어 빌딩이나 기구의 조각 같은 실세계 객체의 위치, 사람이나 사업체의 주소, 그리고 사원의 식별자로서 사용되는 사원 번호와 같은 회사 전체에 넓게 분포되어 있는 많은 코드 등이 바로 그것이다.

모델은 데이터 요소에 대한 하나의 분류 스킴을 제공한다. 데이터 요소는 모델 개체 이름을 지시하는 사상에 의해서 발생된 모델링 개체를 가지고 식별할 수 있다. 이러한 이름들의 예는 사원 (Employee), 가격(Cost), 훈련(Training) 그



리고 회원(Member) 등과 같다.

(2) 특성 (Property)

특성의 집합은 분류법에서 원소의 집합으로부터 개발된다. 이 집합은 이산적이고, 완전한 용어들로 구성된다. 특성은 표제어를 포함할 수 있다.

특성의 예로서는 이름(Name), 계수(Count), 코드(Code), 비율(Rate), 속도(Velocity), 길이 단위(Length Measure), 높이 단위(Height Measure), 그리고 숫자(Number) 등이 있다.

특성은 데이터 요소의 정의에서 자연스럽게 나타난다.

두 개의 구조 집합으로부터의 원소의 사용은 범주화의 보완적인 방법을 제공한다.

(3) 수식어구 (Modifier)

수식어구는 데이터 요소를 유일하게 식별할 필요가 있을 때 특성과 지시대상에 첨가될 수 있다. 이 수식어구들은 구조 집합들로부터 하나의 문맥으로 유래될 수 있다. 명명법에 대한 규칙에서 수식어구의 개수에 대한 제한이 권장된다.

수식어구는 다른 원소의 의미를 제한한다. 의미론적 제한은 같은 의미의 요소의 인식을 따라

중복을 감소시키고 데이터 재사용의 발생을 증가시키기 위하여 수식어구에 대하여 묘사된다. 용어의 시소러스(Thesaurus)와 같은 메카니즘은 이러한 노력을 용이하게 한다.

4.2 이름의 형식을 관리하는 원칙

4.2.1 구문론적 지침

명명법은 이름 안에서 원소들의 정렬을 명세한다. 이 정렬은 다음 두 개 중의 하나 혹은 조합으로 명세할 수 있다.

- ① 상대적 정렬은 다른 원소에 대한 원소들의 위치를 명세한다. 예를 들어, 수식어구는 수식되는 원소가 나타나기 전에 항상 나타날 것을 요구한다.
- ② 절대적 정렬은 원소의 고정된 발생을 명세한다. 예를 들어, 특성은 항상 이름의 마지막 원소가 될 것을 요구한다.

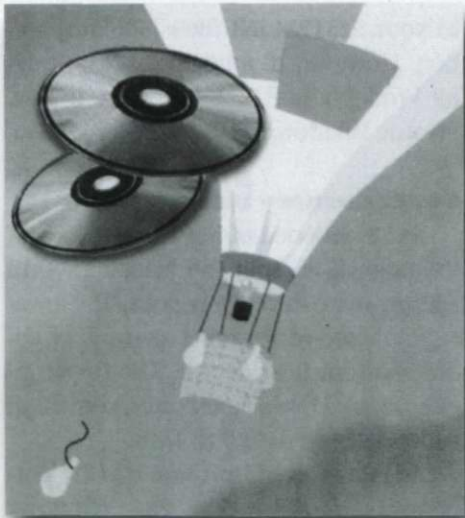
4.2.2 어휘론적 지침

이 규칙들은 원소, 동의어, 약어, 원소 길이 등에서 허용되는 단어와 허용되지 않는 단어들에 대하여 언급한다.

5. 시소러스 응용 지침

사용자가 이름 원소에 대한 여러 동의어, 유사어 및 동형이의어 등을 찾을 수 있는 시소러스는 유용한 도구이다. 이것은 선정된 이름 용어와 다른 용어 사이의 의미론적 연결을 제공할 수 있다. 동형이의어의 사용에 대한 지침 외에 시소러스는 다음을 포함하는 선택에 따라 사용자를 유도할 수 있다.

- 등치 (Equivalence) - 하나의 단어 혹은 용어가 하나의 개념 표현에 대하여 다른 것보다 우선임
- 계층 (Hierarchy) - 광의의 용어와 협의의 용어 사이의 관계가





데이터베이스 리포트①

데이터 요소의 표준화(VI)

상위 분류 혹은 하위 분류의 수준으로 표현됨

- 연합 (Association) - 두 개 혹은 그 이상의 용어들이 동일한 계층에 속하거나 혹은 다른 계층에 속하더라도 의미론적 혹은 개념적으로 연합됨

이름에서 사용되는 원소를 위한 시소러스는 등록원을 위해 관심 있는 부분에 개발되거나 분배된다. 뿐만 아니라 주제 분야에 대한 시소러스들의 개발이 권장된다. 시소러스는 그래픽 묘사에 대한 보충으로서 언어적 표현으로 구조를 서술하는데 사용될 수 있다. 선정된 용어는 원소 이름이 되며 선정된 용어들간의 관계들은 구조에서 원소들의 위치로 표현된다.

부록 명명법 예제

이 규칙들은 4장에 서술된 지침으로부터 마련되었으며 주제 분야 기관의 재량에 따라 문맥 이름의 개발에 적용될 수 있다.

1. 의미론적 규칙

① 지시대상은 예를 들어 전사적 모델 구조 집합에서 찾을 수 있는 회사에 대한 것들을 표현한다.

예 : Cost

② 하나 이상의 지시대상이 존재해야 한다.

③ 특성은 분류 시스템 구조로부터 파생될 수 있으며 데이터의 클래스 혹은 범주를 표현한다.

예 : Total Amount

④ 하나 이상의 특성이 존재해야 한다.

⑤ 수식어구는 주제 분야의 기관에 의해 결정됨으로써 파생될 수 있으며 데이터 요소를 서술하기 위한 필요로 첨가되고 명세된 문맥 안에서 유일하게 만들어진다. 하나의 수식어구는 특성의

부분이 될 수 있으며 다른 수식어구의 순서는 중요하지 않다. 수식어구는 선택적이다.

예 : Budget Period Total Amount

2. 구문론적 규칙

① 지시대상은 이름에서 첫 번째 위치에 있어야 한다.

② 수식어구는 수식되는 원소에 선행되어야 한다. 수식어구의 순서는 데이터 요소 이름을 구별하기 위해 사용되지 않는다.

③ 특성은 가장 마지막 위치에 있어야 한다.

예 : Cost : Budget Period :

[Total Amount]

④ 지시대상 : 수식어구 : [특성]

3. 어휘론적 규칙

① 명사는 단수형으로만 사용된다. 동사는 현재형으로 사용된다.

② 다중-단어 용어에서 이름 원소들과 단어들은 공백으로 분리한다. 특수 문자는 허용되지 않는다.

③ 이름에서 모든 단어들은 혼합된다.

④ 약어 및 두문자어 등은 허용된다.

예 : Cost Budget Period Total Amount

DC

