

데이터베이스 모델링 기법을 이용한 국가표준식물목록 전산화 연구

이 유 미¹ · 김 휘^{2,*}

¹산림청 국립수목원 산림생물조사과, ²목포대학교 생명과학부 생약자원전공

산림청 국립수목원과 한국식물분류학회가 공동으로 국내 최초로 전산화된 식물목록 데이터베이스를 구축하였다. 그 결과로 재배종을 포함한 약 7,000 분류군을 대상으로 항상 최신의 자료를 갱신할 수 있는 국가표준식물목록을 온라인상에 구현하였다. 국가표준식물목록은 37명의 국내 전문분류학자가 참여한 우리나라 식물목록과 관련된 최대 규모의 프로젝트이다. 본 연구는 국가표준식물목록 생성과정을 분석하고 이를 통해 관계형 데이터베이스 모델을 수립하였다. 개발된 광릉 데이터베이스 모델은 식물목록과 관련된 문헌과 다양한 분류학 자료를 제공할 수 있도록 설계되었다. 이 모델은 다양한 분류학적 수준의 학명과 잡종명, 재배명, 이명, 기본명, 국명 및 명명법과 관련된 기타 자료를 함께 제공할 수 있다. 따라서, 국가표준식물목록 데이터베이스는 앞으로 개발될 생물정보시스템들의 개발에 있어 식물의 학명과 관련된 정확하고 일관성이 있는 표준 데이터베이스 기반을 제공하게 되었다.

주요어: 관계형 모델, 국가식물표준목록, 데이터베이스, 생물다양성정보학

서 론

생물다양성 정보학의 정의는 개체수준에서 생물다양성 정보를 다루는 모든 종류의 IT도구를 이용하는 것이다(Berendsohn *et al.*, 2002). 생물다양성정보학은 생물정보학의 한 분야이지만 환경정보학과 분자수준에서의 생물정보학사이에 위치한 분야라고 볼 수 있으며 특히 개체수준이상의 정보를 다룬다(Berendsohn *et al.*, 2003). 생물다양성정보학 분야 중 분류학 정보와 관련된 데이터베이스(이하 DB)의 개발은 단순한 식물목록정보 제공에서 시작하여 기준표본과 명명법상의 다양한 분류학 정보를 DB에 반영하는 과정에서 큰 발전을 이룩하였다(Berendsohn *et al.*, 2002). 특히 식물목록관련 DB의 경우 초기 International Organization of

*교신저자: 전화 061-450-2665, 전송 061-450-6445, huikim@mokpo.ac.kr

접수: 2007년 6월 15일/완료: 2007년 9월 17일

Plant Information (IOPI) 모델 개발(Berendsohn, 1994)부터 최근의 베를린모델(Berendsohn *et al.*, 2002)의 개발에 까지 다양한 DB 모델이 개발되었다. 과거에도 식물관련 DB의 표준, 혹은 최소한 서로 다른 시스템간의 데이터 교환을 가능하게 하려는 노력은 몇 차례 시도되어 왔다. 특히, International Transfer Format (ITF; Anonymous, 1987)과 Herbarium information standards and protocols for interchange of data (HISPID; Croft, 1989; Conn, 1996)는 이러한 초기 시도의 예이다. 한 개의 분류군 - 한 개의 정명 위주의 단순한 생물목록 DB만으로는 식물분류학에 있어서 복잡한 학명 및 명명법, 기준표본 및 표본관련 정보를 통합하는 것은 불가능하다. 초기의 이러한 단순 모델을 벗어나기 위하여 Taxonomic Database Working Group (TDWG)의 활동에 의하여 IOPI에 제공되는 전 세계 식물목록에 대한 DB 모델이 최초로 제안되었다. 그러나, 초기 IOPI 모델의 경우에도 DB 구조의 미비로 인하여 많은 데이터의 손실로 이어지게 되었고(Berendsohn, 1997), 이후 개선된 베를린모델의 경우 생물다양성정보학 분야에서 분류학정보 DB의 전범으로 활용되고 있다(Berendsohn *et al.*, 2002). 전세계적으로 생물다양성의 위기가 강조되고 생물자원의 효율적인 관리를 위하여 생물다양성정보학에 대한 수요가 급증하고 있다. 그러나, 국내에서 개발된 관련 DB들의 경우 한 개의 분류군 - 한 개의 정명 위주의 단순모델로 국내에서 국지적인 사용은 가능하나 필요성 증대되는 국제적인 표준화 작업과는 거리가 멀다(Berendsohn, 1995; 백 등, 2005).

DB모델이란 자료의 전산화를 위하여 미리 정의하고 문서화한 규칙에 따라 특정한 개체의 구조를 다소 추상적인 방법을 기록하는 매체이다(Berendsohn, 1997). 이러한 모델링기법의 목적은 기존 개체의 구조를 설명하고 문서화 하거나 새로운 개체의 구조를 만들어내는 것이다. 데이터 모델링이란 현실 세계의 업무를 추상화하는 반면 데이터 설계라는 것은 모델링 과정을 통해 추출된 정보를 컴퓨터 내부에서 관리할 수 있는 형태로 구체화하며 구조와 관계를 표현하는 것이다. 이러한 모델링 결과는 DB설계를 위한 중요한 준비 작업이다. 목표 시스템이 어떤 정보를 포함할지 사전 검증할 수 있는 근거가 된다. 전산과학 분야에서 사용되는 모델링의 기법은 크게 다음과 같이 세 가지로 나눌수 있다. 첫 번째, 기능적 모델링(Function modelling; Coad and Yourdon, 1991)은 특정 정보 시스템을 구축하는데 있어 상대적으로 빠른 방법이다. 그러나, 기능적 모델링 기법은 나중에 시스템에 기능이 추가되는 것이 매우 어렵다. 두 번째, 정보 모델링(Information modelling)은 시스템과 기반 환경에 대한 깊은 이해와 지식을 필요로 한다. 이에 반해 세 번째 객체지향형 모델링(Object-oriented modelling)은 이러한 두 가지 모델링을 융합한 형태로 사용자 인터페이스의 중요성을 고려한다. 예컨대 객체지향형 DB 모델링은 복잡한 현실을 단순화시켜 컴퓨터 시스템 내부로 이식할 수 있도록 꼭 필요한 요소만 도출하는 복잡한 현실 혹은 업무에서 '관리 및 기록이 필요한 정보'를 추출하는 작업이다(Coad and Yourdon, 1991; 이, 2002).

본 연구에서 제시하는 DB모델은 다음 목표들을 수행하기 위하여 만들어 졌다. 1) 국가표준 목록심의위원회(이하 위원회)의 심의과정을 반영한 데이터를 정의하고 이를 국가표준식물목록DB(이하 목록DB)에 반영한다. 2) 위원회가 제공하는 정명 및 이명의 학명구조를 반영하기 위한 데이터 구조를 제시한다. 3) 목록DB모델은 실제 객체지향형 DB설계시 정규화과정을 통

하여 DB모델의 수정보완을 실시한다.

재 료 및 방 법

DB모델링을 위하여 목록DB 전체 시스템 요구사항을 분석하고 국가표준식물목록의 관리에 대한 업무 흐름을 정리 문서화하는 작업을 실시하였다. 위원회의 자생관속식물종과 재배종 표준명 선정과 부가적으로 제공되는 정보가 바로 DB모델링을 위한 재료가 된다. 여기서 발생하는 다양한 정보는 개체가 되고 각 개체를 묘사하는 내용은 속성이 되며, 각각의 개체가 어떻게 연결되는지의 내용은 개체간의 관계가 된다(이, 2002). 여기서 개체들의 집합은 테이블로 표현되며 시스템의 관리 대상이다. 목록DB에서는 학명, 명명자, 문헌과 혹은 이들 사이의 관계에서 발생하는 새로운 관련 정보가 하나의 테이블이 된다. DB는 자료를 저장하기 위한 시스템이지만 정보를 어떻게 효율적으로 정확히 조회할 수 있는지가 가장 중요하다. 따라서, DB모델링을 통해 다수의 테이블 어떻게 연결되며, 각 개체들은 어떤 관계를 가지고 있는가를 정의하여 DB 스키마(schema; 이, 2002)를 작성하였다.

본 연구에서는 이러한 작업을 고도로 축약하여 보여주는 개체-관계 모형도(Entity-Relationship Diagram; ERD, Fig. 1)를 제시하여 개체와 관계를 제시하였다. 모델링에서 개체를 추출한 후에는 개체 간 의존 관계를 파악하게 되고 개체간 관계를 분석하게 되면 새로운 개체를 생성하게 되고 각 개체의 관계를 파악할 수 있다. 여기서 사용되는 개체는 'entity type(개체 타입)'의 테이블의 형식으로 제공된다. 이 테이블은 열(column)의 윗부분에 속성(attribute)이 제공되면 각 행(row)에는 개체(entity)가 자리 잡게 된다(이, 2002). DB시스템에서는 둘 이상의 개체를 조합하여 정보를 추출하는 경우가 많다. 또한 각 개체에 포함된 정보를 정확하게 검색할 수 있도록 개체에 포함된 데이터를 유일하게 구분 할 수 있는 정보(Primary Key, 이하 Pk)를 설정한다(이, 2002). 그리고 각 속성 크기와 자료 유형을 결정하며, 속성에 null 값은 지정되지 않는 등의 제약 사항을 지정한다. 두개의 'entity type'은 내부의 특정한 속성들(식별자, key로 표현)사이의 특정한 관계로 연결시켰다. 본 연구에서 제공하는 ERD에서는 두개의 'entity type'간의 관계는 기수성(cardinality)이라는 조건이 부여된다. 기수성은 개체가 참여할 수 있는 관계의 수를 나타내고 '1'은 반드시 한개, 'C'는 0 혹은 1개, 'N'은 1부터 다수, 그리고 'CN'은 0부터 다수를 나타낸다. 'C'의 경우는 조건적 관계로 아무 개체가 연결되지 않아도 되는 경우이다(Berendsohn, 1997). 이러한 기수성의 정의는 외래 식별자(foreign key; 이하 Fk)는 반드시 primary key와 항상 일치해야한다는 DB의 참조무결성 법칙(reference integrity rule)을 기준으로 하고 있다(이, 2002). Fig. 1에서와 같이 식물의 학명은 반드시 하나의 rank를 수용해야하는 경우 '1'로 표기되고 그에 반해 학명의 rank는 0에서 여러 개의 학명과 연결될 수 있을 경우 'CN'으로 기수성이 정의된다(Berendsohn, 1995). 모델 수립후 하향식 분석을 통해 구축 대상 개체, 관계, 속성 등을 파악한 후 DB설계



Fig. 1. Cardinality, entity relationship diagram: '1' means that a plant name has exactly 1 corresponding taxonomic rank. 'CN' means that every taxonomic rank is assigned to 0 to many plant names (Berendsohn, 1997)

시 누락된 항목은 없는지 다시금 모델을 검토하는 작업을 수행하였다(Berendsohn, 1997). ERD를 이용한 DB의 모델의 장성은 ER-win이라는 프로그램을 이용하였다(Anonymous, 2007).

결 과 및 고 찰

DB모델: 2007년 4월 현재 우리나라에 분포하는 관속식물 자생종 204과 1,143속 4,881 분류군과 재배품종 134과 424속 2,266 분류군에 대한 국가표준식물목록DB (<http://www.koreaplants.go.kr:9101>)를 구축하였다. 국가표준식물목록은 2003년부터 2005년까지 37명의 분류군별 전문학자에 의해 제출된 학명과 관련 이명 및 국명과 문헌을 바탕으로 한국식물분류학회와 산림청 국립수목원이 공동으로 설립한 국가표준식물목록위원회(이하 위원회)의 심의에 의해 완성되었다(조 등, 2007). 완성된 목록이 담고 있는 정보인 학명과 이명, 문헌 그리고 추천국명 등의 데이터를 전산화하고, 사업기간 이후에는 식물분류학회지를 중심으로 새로운 분류학적 연구에서 제시된 학명상의 변동에 대하여 위원회의 심의의 절차를 반영하도록 설계하였다(Fig. 2). 목록DB의 관계형 DB 모델을 개발하기 위하여 학명, 문헌, 국명, 명명자 등의 주요 테이블과 그 속성을 정하고 이들 테이블간의 관계간의 관계를 설정하였다. 이는 최종적으로 Korea Plants Name Index (KPNI) Gwangreung model이라는 관계형 DB 모델로 실제 web 상에서 구현되는 DB 프로그램(Oracle 8.0)을 이용하여 목록DB를 실시간으로 제공하고 있다.

본 연구에서 제시한 DB모델(Gwangreung model ver. 1.00)은 물리적 혹은 논리적으로 개체의 기능을 프로그램상에서 확인하는데 아주 유용하다. 이러한 확인작업 혹은 정규화작업은 DB모델을 더 정교하게 만들 목적으로 실행된다. 또한, 정규화 작업을 통해 기존의 개체를 새

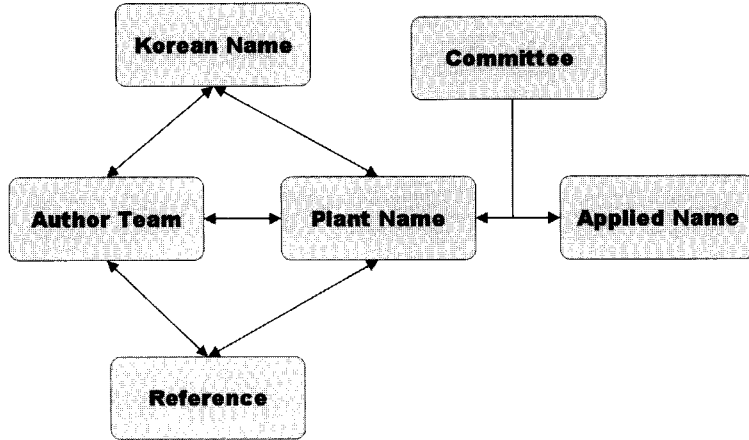


Fig. 2. Basic data structure of KPNI botanical database

롭게 응용하거나 새로운 개체를 추가할 경우 같은 방식으로 수행된다. 따라서, 본 연구에서와 같이 관계형 DB개발을 위해서 필히 개념적, 논리적, 물리적 모델링 단계를 거치게 된다. 개념적 단계의 개체, 속성, 관계 등 이 세 가지의 핵심개념은 용어의 표현만 달리하면서 실제 구현단계인 물리적 모델링까지 그대로 유지된다.

본 연구에서 제시한 객체지향형의 Gwangreung model은 심의의 과정을 도입하여 새로운 학명의 변동에 대한 위원회의 판단을 수용할 수 있는 구조를 제시하였다. 또한, 일단 제안된 학명은 심의의 결과로 실제 하나의 학명으로 인정되고(Fig. 2) 학명은 합법명만으로 구성되며 비합법명의 경우 따로 database에 기록하도록 하였다. 또한, 합법명(이명과 정명을 포함한 모든 분류군의 학명)은 분류학적 지위에 관련 없이 하나의 학명 테이블로 구성하였다.

분류군 학명: 목록DB에서 다루는 분류군명은 속명과 같은 일명법으로 이루어진 학명 및 종명과 같은 이명법으로 이루어진 학명, 종하분류군을 이루는 삼명법 혹은 2개 이상의 종하위명을 기록하는 다명법으로 구성된다. 잡종명과 제배종명도 독립된 학명 형식으로 정하였다(Table 1). 또한, 모든 학명에는 고유의 분류학적 계급을 부여한다. 종이하 분류군은 종이하 분류군의 계급을 표시하고 속과 종사이의 분류군명도 분류학적 계급이 반드시 부여된다.

학명과 관련된 개체에는 분류학적 혹은 명명법상의 판단(이명, 정명)을 하나의 테이블에 표기하지 않고 학명과 학명의 관계 테이블을 독립된 개체로 추출하여 학명을 연결하였다(Fig. 3). 학명간의 연결은 학명관계(Relation 테이블)에 의해 두개의 학명이 연결되며 이 연결은 '1'과 'CN'의 기수성으로 연결된다. 이러한 학명과 학명의 관계는 관계내용 (Relation Qualifier)

Table 1. Allowed name structures for name data table (Berendsohn, 1997).

Name structure	Example
Monomial	<i>Pinus</i>
Binomial	<i>Pinus densiflora</i>
Trinomial	<i>Pinus densiflora</i> for. <i>erecta</i>
Polynomial	<i>Carex breviculmis</i> subsp. <i>royleana</i> for. <i>wardiana</i>
Hybrid	<i>Quercus</i> × <i>dentatomongolica</i>
Cultivar	<i>Juniperus rigida</i> 'Aureo'

Table 2. Entities and data type for entity type 'Plant Name'

Entities	Description	Data Elements
NameId	Primary key for table Name	CHAR(PK)
DelFlag	Indication of data deletion	CHAR
FullName	Complete latin name string including author s	VARCHAR2
WildCultivarFlag	Indicator of Wild plants and Cultivar	CHAR
HybridFlag	Indication of hybrids	CHAR
MultiRankName	Indication of Multirank Name	CHAR
GenusSubdivisionEpi	Genus subdivision epithet	VARCHAR2
GenusID	Pointer to Genus table	INTEGER(FK)
InfraSpeciesEpi1	epithet for subspecies	INTEGER(FK)
InfraSpeciesEpi2	epithet for variety	VARCHAR2
InfraSpeciesEpi3	epithet for forma	CHAR
MultiRankName	epithet string for multinomial name	CHAR
CultivarGroupName	Cultivar group	VARCHAR2
RankId	Pointer to table Rank	SMALLINT
SpeciesEpi	Species epithet	VARCHAR2
CultivarHybridName	Cultivar or Hybrid epithet	CHAR
BasAuthorTeamId	Pointer to basionym authors in the Author Team table	INTEGER(FK)
AuthorTeamId	Pointer to authors in the AuthorTeam table	INTEGER(FK)
ExBasAuthorTeamId	Pointer to ex basionym authors in the Author Team table	INTEGER(FK)
ExAuthorTeamId	Pointer to ex authors in the AuthorTeam table	INTEGER(FK)
EngName	English name	VARCHAR2
JpnName	Japanese name	VARCHAR2
PrkName	DPRK name	VARCHAR2
Image	Image data	VARCHAR2
CreatedWho	Person who created the record	VARCHAR2(FK)
CreatedWhen	Date and time when record was created	DATE
UpdatedWho	Person who updated the record	VARCHAR2(FK)
UpdatedWhen	Date and time when record was updated	DATE

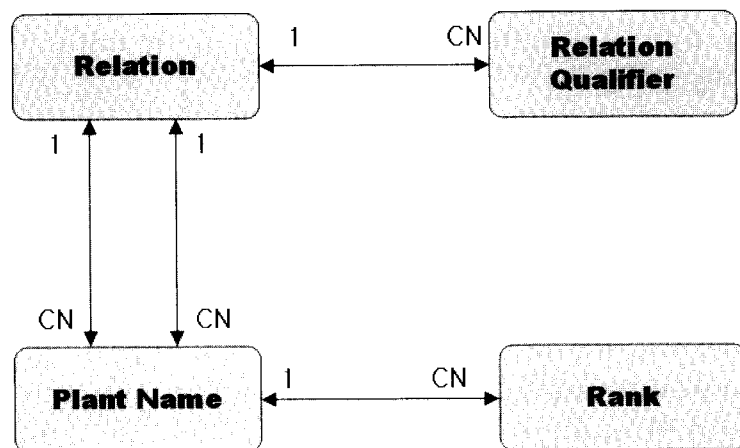


Fig. 3 Entity relationship diagram for KPNI botanical names

테이블에서 정의하는 관계 즉, 이명, 비합법 등의 관계로 정의되며 '1'과 'CN'의 기수성으로 연결된다. 이렇게 정의된 학명간의 관계에 의해 모든 식물학명은 특정한 분류학적 계급(rank)을 갖게 된다. 모든 학명의 데이터는 새로운 학명을 만들지 않는 이상 계급이 바뀌지 않는 제한 조건을 두었다. 학명과 관련된 분현, 명명자, 국명 등의 데이터를 연쇄적으로 연결해야 새로운 학명을 만들 수 있다.

각 학명 데이터는 ICBN(Greuter *et al.*, 2000)에서 규정하고 있는 다양한 분류학적 지위에 해당되는 분류학적 계급 테이블(rank 테이블)에 대한 Fk로 연결된다. 여기서 정의된 Fk에 의해 학명의 지위가 정해지며 동시에 학명의 표기 방식도 일명법을 사용하는 속명 등의 학명과 이명법, 삼명법 등으로 정해진다(Table 2). 일반종과 재배종의 구별은 WildCultivarFlag에 의해 구별된다. 따라서, 재배종으로 표기된 개체는 국제재배식물명규약인 ICNCP (Trehane *et al.*, 1995)에 따라 출판된 학명을 사용한다. 잡종명에 대한 학명의 부여는 학명 테이블에 Hybridflag을 두어 잡종명을 표시하였다. Hybridflag은 속수준, 종수준, 변종 수준에 따라 그 flag의 속성을 달리함으로써 표기 될 수 있고 두 개의 부모종이 표현되는 잡종식의 경우 이를 표기하도록 Hybridflag내의 속성으로 처리하였다. 종이하분류군명은 rank에 따라 가장 많이 사용되는 종이하 분류군인 아종, 변종, 품종을 각각 InfraSpeciesEpi1, InfraSpeciesEpi2, InfraSpeciesEpi3의 개체에 입력이 가능하도록 하였다. 또한, 삼명법 이상의 다명법 학명의 경우 MultiRankName flag을 두어 학명의 표기를 삼명법의 data를 포함하여 FullName 개체에 직접 삼명법으로 제시되는 이상의 학명에 대한 표기를 가능하도록 하였다(Table 1). 학명의

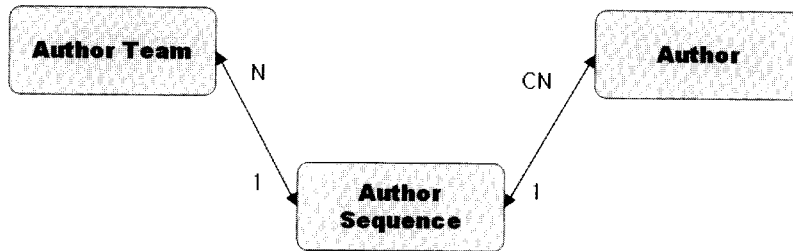


Fig. 4 Entity relation daigram for entity type 'Author Team'.

명명자부분에 대한 입력은 명명자 team 단위로 학명과 연결하도록 하였다. 그래서, 학명 테이블내의 명명자팀의 Fk의 경우 명명자팀 테이블과 'CN'와 'C'으로의 기수성에 의해 연결되도록 하였다. 즉, 각 학명은 0 혹은 1개의 명명자팀과 연결되며 명명자팀은 0에서 다수의 학명과 연결관계를 갖게 된다. 명명자팀의 입력항목은 명명자팀의 학명의 경우 단순히 AuthorTeamId라는 Fk로 연결시키면 된다. 그러나, 유효출판을 한 명명자가 다를 경우 이에 대한 표기를 위하여 ExAuthorTeamId라는 Fk로 연결되며 'ex'앞에 표기되도록하였다(Fig. 4). 기본명명자(Basionym author)의 경우 BasAuthorTeamId Fk로 연결되고 괄호안에 표기되며 ex-기본명 명명자팀명의 경우 ExBasAuthorTeamId Fk로 명명자팀명과 연결되고 기본명 명명자로서 괄호안에 들어가고 'ex' 앞에 표기하게 된다. 학명별로 따로 영명과 일본명 북한명을 두어 각각 EngName, JpnName, PrkName 이라는 개체를 설정하였다. 재배종(cultivar)의 경우 품종의 화상자료를 이용할 수 있도록 Image 개체를 설정하였다. DB 관리자에 의해 학명의 생성과 수정에 대한 내용을 기록할 수 있도록 CreatedWho, CreatedWhen, UpdatedWho, UpdatedWhen 등의 개체를 설정하였다.

명명자와 명명자 team: 학명에 부여되는 명명자는 최소 1명에서 그 이상의 명명자를 갖는 경우가 많다. 이러한 명명자의 데이터 속성을 학명에서 구현하기 위해서는 명명자 team이라는 개념을 도입하였다. 특히, 명명자 team의 경우 2명이상의 경우 이들의 순서를 정할 필요가 있어 명명자 순서를 정하는 독립된 테이블을 만들어 명명자와 명명자팀을 연결할 수 있도록 하였다(Berendsohn, 1994). 명명자팀(Author team)과 명명자순서(Author sequence)는 '1'과 'N'이라는 기수성에 의해 연결된다. 즉, 명명자팀은 반드시 한 개의 명명자순서를 갖고 있어야하고 명명자순서는 여러명의 명명자의 데이터를 연결하게 된다. 또한, 명명자(Author)는 명명자순서와 '1'과 'CN'의 기수성으로 연결된다. 명명자 data는 'Authors of Plant Names' (Brummit and Powell, 1992)를 국제적인 표준규격으로 사용하고 있으며 표준화된 데이터 테

Table 3. Entities and data type for entity type 'Reference'

Entities	Description	Data Elements
ReffId	Primary Key for table Reference	CHAR(PK)
AuthorTeamId	Pointer to the authorteam table	INTEGER(FK)
NomRefCache	Full nomenclatural citation	VARCHAR2
NomTitleAbbrev	Nomenclatural abbreviation of title	VARCHAR2
Reference	Reference title	VARCHAR2
Publisher	Publisher string	VARCHAR2
PageString	Pages of an article or other part in a book or journal	VARCHAR2
Edition	Edition string	VARCHAR2
Volume	Volume as a string	VARCHAR2
Series	Publication Series the reference belongs to	VARCHAR2
RefYear	Year of publication as string	DATE
CreatedWho	Person who created the record	VARCHAR2(FK)
CreatedWhen	Date and time when record was created	DATE
UpdatedWho	Person who updated the record	VARCHAR2(FK)
UpdatedWhen	Date and time when record was updated	DATE
Notes	Remarks and notes	CHAR

이블을 이용하여 Gwangreung Model에서 사용하였다.

문헌: 학명 테이블과 문헌 테이블의 기수성 관계는 'C'와 'C'의 기수성으로 연결된다. 이분헌은 또한, 저자팀 테이블과도 'C'와 'CN'의 관계로 연결된다. 문헌 필드의 데이터의 경우도 문헌과 관련된 국제적인 표준인 TL2 (Taxonomic literature; Stafleu and Cowan, 1976-1988)와 BPH (Lawrence *et al.*, 1968; Bridson and Smith, 1991) 등의 표준을 그대로 활용할 수 있도록 특히 International Plant Name Index (IPND)와 동일한 데이터를 사용할 수 있도록 설계하였으며 주요 개체에 대한 내용은 Table 3에 표기하였다.

국명: 국명테이블을 학명과 'C'와 'CN'의 관계로 연결되도록 기수성을 설정하였다. 국명문헌은 TL2나 BPH에 등록되지 않은 문헌들이 많고 이를 함께 관리할 경우의 다양한 문제점으로 인하여 독립된 국명문헌테이블을 생성하였다. 국명과 국명문헌 테이블은 'C'와 'C'의 관계를 갖도록 설정하였다. 국명의 경우 특정한 명명상의 규약이 존재하지 않는 관계로 위원의 추천 사항으로서 활용되어 추천명(조 등, 2007)과 명명상의 이명과는 다른 이명의 관계로 설정하였다. 따라서 추천명 flag를 두어 추천된 이름과 추천된 이름의 정명과 이와 관련된 국명들의 관계가 되도록 설정하도록 하였다(Table 4).

광릉모델의 확장: 학명DB 구축의 주요 연구결과는 각 분류군의 학명과 분류군과 관련된 각종 데이터를 생성하고 이들 데이터를 서로 연결시키는 작업이다. 그러나, 속성이 서로 다른 데이터간의 공통된 속성을 추출하여 이를 연결시키는 작업은 내재적으로 복잡성을 갖고 있

Table 4. Entities and data type for entity type 'Korean Name'

Name	Description	Data Elements
KorNameId	Primary Key for table Korean Name	CHAR(PK)
NameId	Pointer to Name table	CHAR(FK)
RecomFlag	Indication of recommended Korean Name	CHAR
KorName	Korean Name	CHAR(1)
CreatedWho	Person who created the record	VARCHAR2(FK)
CreatedWhen	Date and time when record was created	DATE
UpdatedWho	Person who updated the record	VARCHAR2(FK)
UpdatedWhen	Date and time when record was updated	DATE
Notes	Remarks and notes	CHAR

다. 실제 분류군의 분류학적 범위와 이를 이용하는 분류학자와 학명의 이용자간의 분류군의 범위에 대한 이해의 차이가 존재한다. 또한, 국제식물명명규약(Greuter *et al.*, 2000)이 다루는 학명과 학명의 기준이 되는 기준표본과의 관계 및 학명과 학명간의 복잡한 명명상의 관계를 DB로 전환하기 위해서는 분류군의 분류학적 한계에 대한 미묘한 견해차의 문제점으로부터 단순한 학명의 잘못된 적용까지 다양한 내용을 포함해야 한다. 모든 명명상의 문제점은 국제식물명명규약에서 다루는 내용으로 이러한 분류군의 한계와 같은 미묘한 분류학적 문제를 해결하기 위한 방식으로 제안된 DB 모델로는 Potential taxa (Berendsohn, 1994, 1995)가 있다. Potential taxa의 개념은 미시종(microspecies)의 개념 뿐아니라 종개념의 범위 즉 넓은 의미와 좁은 의미의 개념을 포괄할 수 있으며 특히 복잡성이 증가하는 광범위한 지역 혹은 세계 수준에서의 학명 목록분야 연구로의 확장을 위해서는 앞으로 Gwangreung model내에서도 potential taxa로의 확장이 필요하다.

최근 국내에서 생물다양성 관련DB가 여러 기관에서 동시다발적으로 만들어 지고 있으나 대부분 DB기술자들에게 모든 구체적인 DB설계에 대한 업무가 일임되어 있는 경우가 많다. 특히, 참여 인력에 있어 생물학 전문지식과 DB관련 전문 기술정보의 비대칭성으로 인한 불균형은 비효율적이고 표준화 되지 못한 DB만을 양산하게 된다. 따라서, 생물학자들의 자신의 고유한 영역을 추상화된 DB로 전환하는데 적극적으로 참여하지 못할 경우 생물분야와 관련 없는 DB기술자들에 의해 DB의 구조가 왜곡되거나 비효율성이 증대된다. 컴퓨터를 이용한 정보의 공유는 정보의 방대한 양과 빠른 속도라는 장점을 가지고 있기 때문에 정보수요자에게 있어서 가치창출을 위한 훌륭한 도구로서 이용될 수 있다. 국내에 있어서도 생물종에 대한 DB구축이 활기를 띠기 시작하고 있어 이 분야 있어 큰 발전이 기대되었으나 국가적으로 통합된 자료의 발전이 이루어지지 못하고 중복적으로 투자되는 경우가 많다. 성격이 비슷한 DB간의 표준화와 이를 통한 통합 운영체계 구축, 다른 DB와 연동 등이 필요한 상황이지만

구체적인 통합의 활동은 확인되지 않고 있다. 국가적으로 식물정보의 표준화에 대한 수요가 늘어가고 있는 상황에서 국가적으로 학명과 국명의 일치된 자료를 제시할 필요가 있었다. 따라서, 데이터모델링 기법을 통하여 정명과 이명 국명과 추천명 및 분헌에 대한 검색이 가능한 데이터모델인 Gwangreung 모델을 완성하였다. 이 모델을 바탕으로 국가의 식물명과 관련된 데이터를 집대성 할 수 있고 이후 개발되는 관련 DB의 식물명 부분에 표준적 기반을 제공할 수 있는 국가표준식물목록 DB를 구축하게 되었다.

감 사 의 글

연구진반에 도움을 준 산림청 국립수목원 정승선 연구사와 참을성 있게 모든 요구사항을 DB에서 구현해준 서광종합개발(주) 박종관 과장님께 감사드립니다.

인 용 문 헌

- Anonymous. 1987. ITF. The international transfer format for botanic garden plant records. Pittsburgh.
- _____. 2007. Allfusion ER-win data modeler. CA, NY.
- Berendsohn, W. G. 1994. IOPI vascular plant checklist. A case model of checklist system data, version 6.0. In Global plant checklist project plan, version 7.3. Wilson, K. (ed.), International organization for plant information, Sidney. <http://www.bgbm.fu-berlin.de/iopi/iopimodel73/7301root.htm>
- _____. 1995. The concept of 'potential taxa' in databases. *Taxon* 44: 207-212.
- _____. 1997. A taxonomic information model for botanical databases: the IOPI model. *Taxon* 46: 283-309.
- _____, M. Geoffroy, K. Glück, A. Güntsch and J.-N. Li. 2002. The Berlin taxonomic information model. <http://www.bgbg.org/biodivinf/docs/bgbm-model/>
- _____, M. Döring, M. Geoffroy, K. Glück, A. Güntsch, A. Hahn, R. Jahn, W. Kusber, J.-L. Li, D. Röpert and F. Sprech. 2003. MoReTax - Handling factual information linked to taxonomic concepts in biology. BfN, Bonn.
- Bridson, G. D. R. and E. R. Smith. 1991. B-P-H/S Botanic-Periodicum-Huntianum /supplementum. Pittsburgh.
- Brummitt, R. K. and C. E. Powell. 1992. Authors of Plant Names. Royal botanic gardens,

Kew.

- Coad, E. and E. Yourdon. 1991. Object-oriented Analysis, ed. 2. Englewood Cliffs.
- Conn, B. J. 1996. HISPID3. Herbarium Information Standards and Protocols for Interchange of Data HISPID, Version 3. Royal Botanic Gardens, Sydney.
- Croft, J. R. 1989. HISPID. Herbarium Information Standards and Protocols for Interchange of Data, Version 1. Canberra.
- Greuter, W., J. McNeill, F. R. Barrie, H. M. Burdet, V. Demoulin, T. S. Filgueiras, D. H. Nicolson, P. C. Silva, J. E. Skog, P. Trehane, N. J. Turland and D. L. Hawksworth. (eds). 2000. International Code of Botanical Nomenclature (Saint Louis Code) Koeltz Scientific Books, Königstein, Germany.
- Lawrence, G. H. M., A. F. G. Buchheim, G. S. Daniels and H. Dolezal. 1968. B-P-H. Botanico-Periodicum-Huntianum. Hunt botanical library, Pittsburgh.
- Stafleu, F. A. and R. S. Cowan. 1976-1988. Taxonomic Literature, ed. 2. Regnum Veg. 94, 98, 105, 110, 112, 115, 116.
- Trehane, P., C. D. Brickell, B. R. Baum, W. L. A. Hettterscheid, A. C. Leslie, J. McNeil, S. A. Spongberg and F. Vrugtman. 1995. International Code of Nomenclature for Cultivated Plants (ICNCP or Cultivated Plant Code). Quaterjack Publishing, Wimborne.
- 백기열, 정종천, 박선주, 이종욱. 2005. 생물다양성 자료의 데이터베이스화와 온라인 관리시스템 및 분석도구 구축에 관한 연구. 한국환경과학회지. 14: 1119-1127.
- 이춘식. 2002. 데이터베이스 설계와 구축. 한빛미디어(주), 서울.
- 조동광, 이유미, 최혁재, 정승선, 남기흠, 고정은, 양선규. 2007. 국가표준식물목록. 국립수목원. 포천시.

Database Model for Korea Plant Name Index

You-Mi Lee¹ and Hui Kim^{2,*}

¹Herbarium, Korea National Arboretum, Pochun, 487-821

²Department of Medicinal Plants Resources, Mokpo National University,
Mokpo 534-729, Korea

Korea national arboretum has worked with the plant taxonomic society of Korea to make the first fully electronic floristic checklist in Korea. The result is an ever-expanding online plant name index containing scientifically authoritative, up-to-date information on the approximately 7,000 taxa including cultivars. With 37 contributing taxonomists, KPNI is the largest collaborative research projects ever assembled in Korea. A comprehensive database model for the taxonomic data from literature and other sources is presented, which was devised for the Korea National Plant Index database project (KPNI). Gwangreung database model is based on an approach using entity-relationship diagram. It encompasses taxa of all ranks, nothotaxa and hybrid formulae, cultivars, full synonymy, basionyms, Korean name, and other nomenclatural information. This paper presents an analysis of KPNI work processes and an overview how we are approaching the construction of Gwangreung database model. It can help the system engineers of other biological information systems to develop their database based on the accurate and integrative taxonomic database.

Key Words: Biodiversity informatics, checklist, database model, entity-relationship model, KPNI(Korea Plant Name Index)

*Corresponding author: Phone 82-61-450-2665, Fax 82-61-450-6445, huikim@mokpo.ac.kr

Received: 15 June 2007/Accepted: 17 September 2007