

GRM 기반의 생물자원정보 네트워크 구축 방안

이계준⁰ 양진호, 박형선, 안부영, 윤희준
한국과학기술정보연구원
{kjlee, spearjin, seonpark, ahnyoung, k2}@kisti.re.kr

Contracture Strategy for Network of Biological Resources Information of based GRM

KyeJun Lee⁰ Jin-Ho Yang, Hyung-Seon Park, Bu-young Ahn, Hee-Jun Yun
Dept. of Office, Korea Institute of Science and Technology Information

요 약

21세기는 생물자원을 이용한 산업이 많은 분야의 발전을 이끌어 나간다고 해도 과언이 아닐 정도로 현재 많은 분야에서 생물자원의 보전과 활용에 대한 활발한 연구가 이루어지고 있다. 이러한 것은 생물자원부국이 앞으로의 선진 국가를 의미하는 것으로 자원이 부족한 국내 실정으로는 더욱더 많은 관심을 가지고 이를 해결해야만 하는 과제를 가지고 있다.

생물분야의 연구는 타 분야와는 다르게 지금까지 외면되어 왔기 때문에 정보화에 대한 마인드가 부족하여 대부분의 정보가 디지털화되어 운용되어지지 못하고 있으며, 운용되어지는 정보는 서로 표준화되지 못해서 정보의 재사용이나 공동활용이 부족한 실정이다. 정보로서의 가치를 가지기 위해서는 표준화되어 사용자가 원하는 정보로의 가공을 거쳐 Real-Time 접근이 가능해야만 한다.

본 논문에서는 생물자원이 가지고 있는 정보의 운용을 효율적으로 구축하기 위하여 국내 생물자원 네트워크 구성, 생물자원정보 통합 방법, 표준화를 위한 DTD작성과 데이터베이스 구축을 위한 컴포넌트를 이용한 입력시스템 구축, 미디어이터 기법을 이용한 분산통합, GRM(Global Road Map) 작성을 통한 효율적인 정보의 접근이 가능한 시스템 구축을 목적으로 한다.

1. 서 론

지금까지 대부분의 정보화에서 기초연구분야는 등한 시되어 왔으며 생물자원에 대한 정보도 마찬가지였다. 근래에 들어 생물자원정보의 소중함과 필요성을 인지하게 됨으로서 생물자원의 가치, 감소 위기 극복을 위한 보존, 유용화에 대한 인식이 범지구적으로 확대되고 있으며, 국가적 차원의 보전 정책 수립에 따른 국내 생물종자원에 대한 실제 파악과 체계적인 조사·분석·정리에 대한 정보화 작업의 중요성이 큰 관심사로 대두되었다. 또한, 생물자원 정보에 대한 경제적, 환경적, 사회적 이익을 산출하기 위해서는 범세계적으로 사용 가능하고, 사용자가 인정할 수 있는 것을 기반으로 해야 한다. 정보 기술 도구들을 사용하여 생물자원 정보 데이터베이스의 상호연용이 가능한 네트워크 체계를 구축해야 한다.[1]

이와 같은 것을 가능하기 위해 생물자원 정보의 재사용과 공동활용이 가능하도록 표준화된 데이터베이스 구축을 위하여 W3C(World-Wide Web Consortium)의 인터넷 전자문서 표준인 XML(eXtensible Markup Language)을 기반으로 하는 것을 제안한다. 그리고 효율적으로 생물자원 정보의 활용을 위해 조사·분석·수집·디지털화·정보화·지식화 등을 고려할 때 생물자원이 가지는 무한한 확장성을 포함할 수 있는 분산 통합 시스템을 구축해야 한다.[2]

따라서, 본 논문에서는 각 로컬 지역의 정보에 대해 GRM(Global Road Map) 작성을 기반으로 모든 정보의 접근을 가능하도록 하여 Real-Time으로 정보운용이 가능하도록 하는 시스템 설계를 목적으로 한다.

2. 국내 현황

현재 국내의 생물자원 정보들은 국내의 고유한 종에 대한 정보에서부터 국내에서 자생하는 종에 이르기까지의 정보를 대상으로 각 기관에 종속적으로 독립적인 구축, 관리, 서비스가 이루어지는 것이 대부분이다. 그러므로 각 DB들은 여러 기관에 중복되어 있으며, 국내 데이터베이스들은 대부분 적은 규모로 서비스를 하고 있고 정보의 공유나 연계가 거의 이루어지고 있지 못하기 때문에 하나의 완전한 정보를 얻기 위해서는 여러 곳의 DB를 검색하여 사용자가 조합하는 과정을 거쳐야만 원하는 정보 전체를 얻을 수 있다. 또한, 데이터를 보유하고는 있지만 DB화되지 못하여 서비스되지 못하고 있는 정보들이 많은 실정이다.

이러한 현상은 정보화에 대한 지식과 인력의 부족으로 인한 부분과 기존의 보유하고 있는 정보들에 대한 공개를 기피하는 부분으로 크게 나뉘볼 수 있다. 우선, 지금까지 생물자원 정보를 분류하고 연구하는 분야의 전문가들은 전산화에 대한 마인드(Mind) 부족과 보유하고 있는 정보에 대한 서비스를 고려하지 않은 수집으로 인해 데이터베이스 구축에 많은 어려움을 수반하고 있다. 그리고 공개를 기피하는 부분은 정보로서의 가치가 높고 뛰어난 정보들에 대한 보호 정책 및 무단 복제에 대한 대책이 없다는 것이 문제로 제기된다. 이러한 것을 해결하기 위해 정보의 내용 보호차원에서는 저작권법을 적용, 이미지에 대해서는 워터마크(Water Mark)와 전용 뷰어(Viewer) 기법을 적용 등의 방법과 지속적인 지원이

가능한 사업을 통해 상호 우호적인 관계를 기초로 하여 숨겨져 있는 정보들에 대한 정보화 작업을 준비 중에 있다.

생물자원 정보는 멀티미디어 정보로서 텍스트, 이미지, 동영상, GIS 정보들로 구성되어 있으나 현재 데이터베이스 대부분은 이러한 정보를 고루 갖추고 서비스하지 못하고 있으며, 가지고 있지만 기술적인 적용이 미비한 실정이다. 또한, 국내에는 대학·연구소·협회·기업 등을 기반으로 생물자원 정보에 대한 내용 검색, 사이버 박물관, 자연사 박물관, 종자은행 등의 다양한 서비스가 이루어지고 있으나 규모나 내용 면에서 내실을 다져야 할 필요성을 가지고 있다. 국가적인 경쟁력을 가지기 위해서는 유관기관들 간의 네트워크를 통한 정보 공유와 표준화를 기반으로 하는 통합 추진으로 Macro와 Micro의 두 가지 모듈을 겸비한 데이터베이스 구축이 필요하다.

3. 통합시스템

3.1 XML 기반 정보 표준화

생물자원 정보는 분류학적인 부분에서 식물, 동물, 곤충, 미생물 등으로 나뉘지며, 각각은 내용정보, 생태정보, 이미지나 동영상 정보, 관련 논문정보나 문헌정보 등으로 구성되어진다. 또한, 신규 및 기존 정보의 분산 통합, 생물자원 표준화를 통한 데이터베이스와 네트워크 구축을 위하여 XML(eXtensible Markup Language) 기반으로 DB를 구축하기 위해 표준 DTD를 작성하였다.[2]

표준 DTD는 생물 종(Species)정보와 콘텐츠(Content) 정보로 나뉘서 구성되며, 생물 종정보는 분류학적 정보를 대상으로 횡적인 통합으로 고려하였으며, 콘텐츠 정보는 종설명정보, 멀티미디어정보, 서식지정보, 참고문헌정보, 명명자정보, 관련 정보, 파생정보 등으로 구분하였으며, 각 분야별 생물정보의 종적인 통합을 위한 것으로 그 특징과 확장성을 고려하였다.

3.2 Mediator 기반 분산 통합

현재 국내 생물자원 정보 데이터베이스 구축을 위해서는 전체 생물자원 정보 표준 내에서 각 분야별 스키마 통합이 이루어져야 하며, 표준 스키마를 기반으로 구축된 정보는 분산통합시스템에 의해 통합과 검색이 이루어진다. 구축된 정보를 편집·가공하는 시스템을 거쳐 데이터베이스가 구축되며, 이때, 멀티미디어정보와 지리 정보에 대한 기술을 부합하여 보다 완전한 정보를 생성한다.

국내 생물자원 정보에 대한 DB 구축은 각 분야의 정보들이 전문가들에 의해 입력하는 순간 데이터베이스화 되고 사용자들에게 서비스되며, 데이터베이스가 모든 사용자들의 참여로 확장될 수 있도록 정보의 수정이나 추가를 가능하게 하고 이에 대한 면밀한 평가를 부가하여 정보의 생성·평가·서비스가 하나 되는 시스템 구축이 되어야 한다.

이러한 것이 가능하며, 기구축된 정보와 신규 구축되는 정보의 논리적인 통합을 위해 사용되어지는 미디어에

터 기법을 적용하였다.[3]

4. GRM 기반 네트워크

GRM은 분산되어 있는 정보에 대한 접근을 위한 정보를 가지고 있어 사용자들의 서비스 요청이 있을 경우에 Mediator 시스템을 통해 GRM의 정보를 검색하고 결과를 가지고 로컬 정보를 검색하게 된다. 또한 로컬 정보에 대한 전체적이 Local Map이 존재하고 이것을 모두 포함하는 GRM이 만들어진다. 다음은 GRM 시스템의 전체 구성이다.

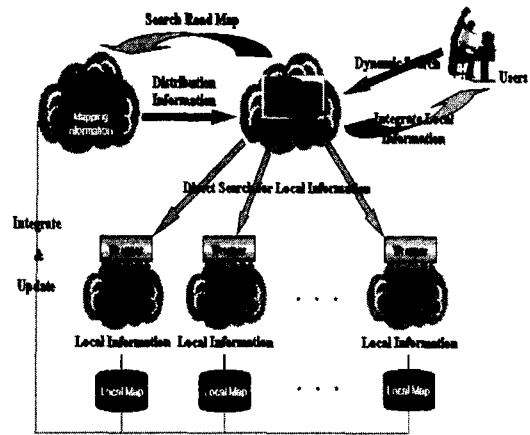


그림 1 GRM 시스템의 전체 구성

GRM 시스템은 로컬의 정보를 논리적인 관계(Association)에서 얻어지는 계층구조를 기반으로 만들어지며, 정보들 간의 관계를 정의하기 위한 기본 개념은 다음과 같다.

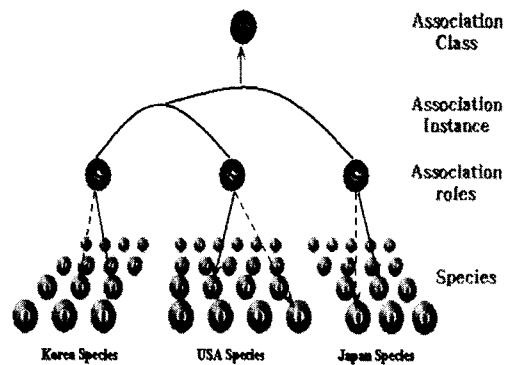


그림 2 GRM의 개념

GRM은 각각의 개별적인 개체(Object)들로부터 시작을 하며, 이러한 개체들 간에 정보로서의 관련성이 같은 것들 간에 하나의 의미 부여를 통해 계층적인 구조를 가지도록 설계하는 것이다. 이것은 정보의 확장이 보장되게

되며, 원하는 정보를 다시 재정의 하여 자신의 시스템에 사용할 수 있도록 한다. 따라서 본 개념을 기반으로 해서 GRM을 생물자원정보에 적용하여 계층적으로 나타내면 다음과 같은 단계로 구분할 수 있다.

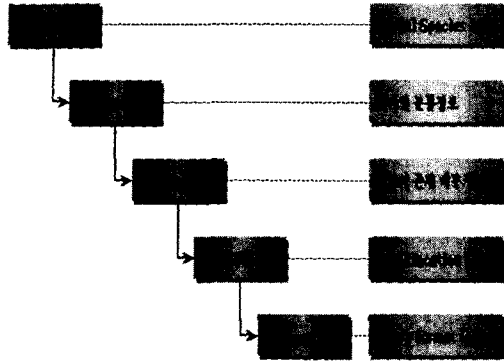


그림 3 GRM의 구조

첫째, Species Info는 종정보인 종·속·과·목·강·문·계와 일반명(Common Name : 한글명, 영문명 등)을 기준으로 구성된다.

둘째, Classification Info는 종정보 이외의 정보를 보다 세분화함으로써 검색의 효율과 분산화를 위한 분류정보이다.

셋째, Filter Info는 검색 대상이 되는 내용이 존재하는가의 정보를 가지고 있어 정보의 검색 시에 필터링 해 주는 기능을 수행한다.

넷째, Local URL Info는 분산되어 존재하는 Database의 위치 정보를 가지고 있어 Global 정보와 Local 정보와의 Mapping 기능을 수행한다.

다섯째, Query Format는 Local 정보를 검색하기 위해 각각의 Database에 맞는 질의 형태로 변환하여 검색하고 결과를 반환해 준다.

5. 결론

GRM 기반의 생물자원정보 네트워크를 구축하기 위해서 수행되어야 하는 부분으로 크게 나누어볼 수 있다.

첫째, 생물자원정보 데이터베이스 구축의 두 가지 목적은 첫째, 생물종자원(Species) 정보 구축, 둘째, 이것들을 기초로 하여 내용정보, 이미지정보, 동영상정보, GIS정보, 관련정보 등을 구축한다. 이것은 국제적인 흐름에 따르는 것이며, 국내에서 반드시 구축되어야 하는 Catalogue정보와 관련 정보를 동시에 구축할 수 있는 것이다.

둘째, Catalogue정보를 구축하기 위해서는 각 분야 전문가들과 공동으로 공동활용과 검색을 위한 표준화된 구성요소를 추출하고 각 분야의 정보를 모두 포함할 수 있는 표준 DTD 작성한다. 또한 이렇게 만들어진 정보들의 따르는 로컬 시스템을 구축하는데 여기에는 입력시스템과 Global 시스템과 연동이 가능한 모듈을 포함하는

시스템이 구축되어야 한다.

셋째, 로컬시스템을 통해 각각의 분야에서 만들어지는 정보들의 표준화가 이루어지고 이렇게 데이터베이스화된 정보를 Centralize한 개념의 하나의 데이터베이스를 구축하는 것이 아닌 Mediator기법을 이용하여 분산된 정보 Real-time검색이 가능한 시스템을 구축하여 서비스 하게 된다.

넷째, 가장 선행되어야 할 것은 생물종에 대한 데이터베이스이며, 이것을 기반으로 종에 관련정보들이 구축되어진다. 이렇듯, 정보화에 대한 기준이 세워지고 하나의 종에 대한 완벽한 관련 정보까지 구축되어진 다음에 종들끼리의 Association(연관)·Relation(관련) 정보가 완벽하게 적용할 때만이 올바른 데이터베이스 구축 및 운영이 가능해진다.

다섯째, 정보의 효율적인 접근을 통한 서비스를 위하여 GRM과 같은 정보에 대한 맵을 만들어 모든 정보의 접근을 보장하므로 정보유통을 위한 기반을 조성한다.

본 논문에서는 생물자원정보를 대상으로 국내 모든 생물자원에 대한 접근을 가능하게 하며, 표준화된 데이터베이스 구축, 분산통합을 통한 정보의 고유성 및 정보 제공으로 인한 손실을 최소화하기 위한 네트워크 체계 구축 및 필요한 시스템과 방법에 대해 전반적으로 제안을 하였다. 제안된 내용을 기반으로 정보를 구축하고 유통한다면 생물자원정보 네트워크 체계가 확립 될 것이다.

참고문헌

- [1]국립환경연구원, "생물자원주권 확보를 위한 국립생물자원보존관의 역할", 환경의날 기념 국제세미나, 2002
- [2]한림원, "국내 생물자원 정보 DB 및 네트워크 운영 체계 확립", 생물자원정보 콜로퀴움, 2002
- [3]충남대학교 소프트웨어연구센터, "XML 문서 저장/검색 및 분산 문서 시스템의 설계 및 구현", 연구보고서, 2001
- [4]이계준, 조현양, 최재황, 손강렬, "효율적인 KSCI 체계 구축을 위한 XML 기반 모델 설계", 한국과학기술정보연구원, 2001 가을 정보과학회 학술발표논문집, 2001
- [5]한국전자통신연구원, "이질 자료 모델의 충돌 해결 방안 연구", 중간보고서, 2001
- [6]Integrated Taxonomic Information System, <http://www.itis.usda.gov>
- [7]GBIF(The Global Biodiversity Information Facility) <http://www.gbif.org/index.html>
- [8]Species 2000 <http://www.sp2000.org>
- [9]Center for Computational Biology & Bioinformatics,
- [10]National Center for Biotechnology Information,
- [12]손현석, 생물정보학 (Bioinformatics)해의동향, 한국과학기술정보연구원, 지식정보인프라지