

F-6 アジアオセアニア地域における生物多様性の減少解決のための世界分類学イニシアティブに関する研究

(2) GTI地域プログラム実施における生物多様性情報共有化と利用に関する研究

④ 分類学データベース利用手法の開発に関する研究

株式会社 ランス

技術開発部

小野 哲

<研究協力者>

独立行政法人 国立環境研究所

志村 純子

平成14～16年度合計予算額 5,483千円

(うち、平成16年度予算額 ※ 0千円)

「上記の予算額には、間接経費 0千円を含む」

[要旨] 生物情報データベースの整備における属性の異なるデータベース間の整合性をとるために学名とその学名の概念を記述する仕組みとして、Nomencuratorスキーマを用いたリレーショナルデータベース実装とその公開用インターフェース実装における基礎的な手順をあきらかとし、その検証をおこなった。

1. はじめに

近年の生物多様性研究の進展にともない、様々な生物情報データベースの整備が重要となっている。生物情報データベースには標本データベース、学名データベース、Geographicデータベース等があるが、特にこれらの中でも学名情報データベースは、生物多様性解析において、最も基本的な情報となるため、その重要性が高い。しかし、学名の記述や学名の分類には各生物分類学者毎に見解が異なり、同じ1つの種であっても様々な命名や分類が行われている。そのため、学名情報を統一的に扱うことが難しい。

こうした異なる学名記述間を表現する方法としてNomencuratorスキーマが提案されている。本研究では、このNomencuratorスキーマを用いた、RDBでの実装と、学名情報データベースの公開用インターフェースとして、XML・SOAPを用いた手法をTaxoNote Java Applicationという実装として研究を進めた。

2. 研究目的

本研究では、アジアオセアニア地域及び日本国内に散在する生物多様性のデータベースを統括的に管理する学名のデータベース検索を可能にすることが目的である。

3. 研究方法

モデルケースとして、筑波大学と国立環境研究所が共同で開発したNomencuratorシステムを再利用し、その仕組みをもとに各種のデータベースとの連携を行う。そのインターフェースとして

SOAPを使用し、相互のデータ変換がどのように行われることが最善であるかの技術的検証を行った。

4. 結果・考察

分野や学説によっては、同じ1つの分類体系を現すのに、様々分類体系が用いられており、従来の学名情報データベースでは、各生物分類学者毎に見解が異なる様々な分類体系をそのまま、登録することができなかった。そのため、本来生物分類学上で最も重要となる学名情報に関する統一的なデータベースを構築することが難しく、現在は、各分野毎に別個のデータベースを構築して運用していることがほとんどである。

こうした生物分類学情報を扱うスキーマとして、本年度はNomencuratorスキーマを使用した。このNomencuratorスキーマは異なる学名記述間の違いを許容することができるスキーマであり、1つの種に対して、様々な分類体系が存在した場合でも、それらを問題無く扱うことのできる。そして、Nomencuratorスキーマの概念上でのスキーマである。そこで、このNomencuratorスキーマをRDB (Fig.1) として実装する方法について考察した。

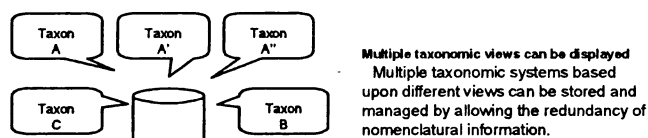


Fig.1 Nomencurator data structure implemented on RDBMS

Nomencuratorスキーマ自体は概念上の論理データ構造のため、これを物理構造であるRDBに適用させる必要があった。まず、Nomencuratorスキーマで定義されているオブジェクト毎にRDBのテーブルを作成した。物理構造にするにあたっては、一般的なデータベースでの運用が可能となるよう極力特殊な構造はとらず、シンプルな構成とした。そこで、論理構造上で表現されていたオブジェクト間のリンクはRDB向けにFig.2のように正規化した。

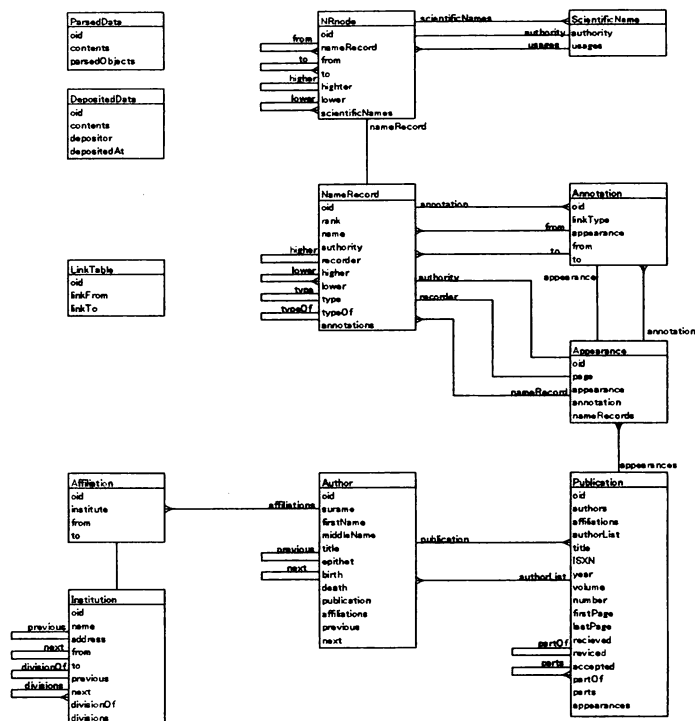


Fig.2 Nomencurator ER Diagram

XML,SOAP,JDBC Interface

Nomencuratorのスキーマをアプリケーションプログラムとして実装されたソフトウェアである Taxonoteは Java Applicationとして動作する。データベースへのアクセス用のインターフェイスとしてはXMLやJDBCのインターフェイスを用いているFig.3。これは、将来的に、Nomencuratorスキーマを用いたデータベースへのアクセスを広く公開できるようにとの考えからである。現時点でのアクセスではhttpを用いてアクセスし、サーバ側のServletを呼び出すことによりデータを取得するが、データの通信に利用されているデータはXMLのため、SOAPへの対応は容易にできるものと考えた。

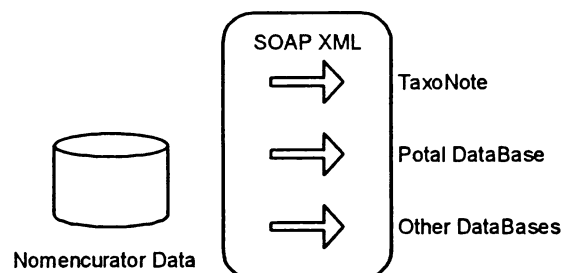


Fig.3 Nomencurator ER Diagram

実際の学名情報へのアクセスにはNomencuratorスキーマを用いたデータベースサーバに対し、次のAPIをキックすることにより行われる。この2つのAPIにより学名情報のオブジェクトをXML形式で取得することができる。

```
getNamedObject  
saveNamedObject  
searchNameUsage
```

なお、APIを複雑に構成してしまうと、あらゆるデータアクセスのパターン毎にAPIを用意しなければならなくなってしまうため、最も単純な形式として3つのAPIを定義した。

GetNamedObject APIはオブジェクト取得用のAPIであり、Nomencuratorスキーマで定義されているオブジェクトを、OIDを元に呼び出すことができる。

SearchNamedObject APIはオブジェクト保存用のAPIであり、Nomencuratorスキーマで定義されているオブジェクトを、XML形式で送信することにより、データベースに保存することができる。

各学名情報はFig. 4のような形でXMLにより定義され、APIを用いて取得できる。実際にTaxoNote Java Application側ではサーバより取得したXMLデータをパースすることにより各オブジェクトの内容を出力する。

SearchNameUsage APIは学名情報を検索するAPIである。TaxoNote Java Application では分類体系を検索するキーとして、学名、出版年、author を定義しており、これら情報をキーとして、分類体系のviewを抜き出すことができる。通常TaxoNote Java Applicationで情報を表示する際には、このひとまとまりとなった分類体系のviewを用いることとなる。各APIの呼び出しシーケンスはFig. 5の通りである。

```
-<NameUsage>  
- <oid>NameUsage::B1000000151</oid>  
- <rank>species</rank>  
- <name>Burkholderia andropogonis</name>  
- <authority />  
- <appearance>Appearance::B1000000047</appearance>  
- <higher />  
- </NameUsage>
```

Fig.4 Description of nomenclatural information with XML

}

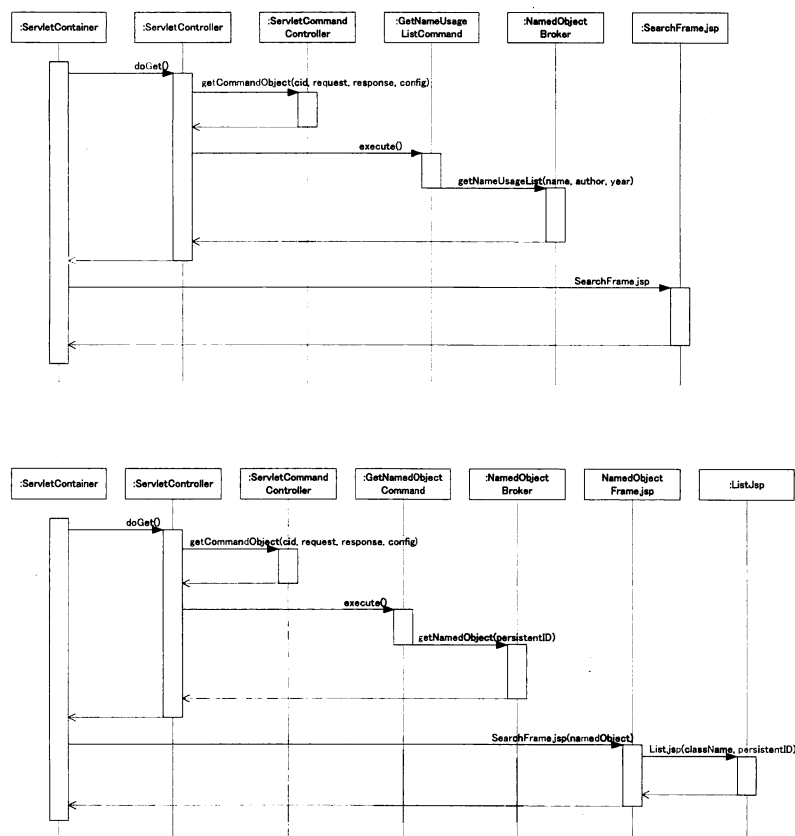


Fig.5 Data access sequence

Comparison of taxonomic view

Nomenclatorスキーマは異なる学名記述間のデータ表現することができるが、従来からある学名情報の表現方法では、こうした複数の見解に基づく学名情報を表現することが難しい。例えば、従来から分類体系の表現に用いられるGUIとしてツリー形式のGUIがある。(Fig. 6) この方式はTaxoNote Java Applicationでも、用いているが、2つの分類体系を比較・表現することは難しい。TaxoNote Java Applicationでは、ツリー形式以外の表現方法として、異なる学名記述間のデータをより効果的に表現するためには、Fig. 7のようなGUI表現を用いることもできる。

この表現方法では、2つ以上の分類体系を指定することによって、各階層毎 (family, genus, species等) の相違を明確に出力することが可能である。

ットを用いた広範囲でのデータベースの連携を可能とするものである。これは生物多様性の研究において、標本データベース、学名データベース、Geographicデータベース等の連携に応用できる。

また、将来的には本実装を改良することで、容易にWeb Serviceに対応することが可能であり、あらゆる学術データベースとのシームレスな連携が可能となると考えられる。

そこでSOAPインターフェースを使用しNomencuratorスキーマを使用してSOAPインターフェースによるデータベースの公開に関する技術的検証を行った。

5. 本研究により得られた成果

標準化されたデータ交換のためのインターフェースはSOAPによって効率よく公開でき再利用が可能であることが明確になった。しかし、その研究をすすめてゆく過程でアメリカ及びヨーロッパにおいて標準化されつつあるデータ連携のシステムであるDiGIRを知ることとなった。DiGIRはすでにいくつかの実績をもっており、分類学の世界では標準になりつつあるデータ連携システムである。このシステムを深く研究することによってSOAPインターフェースと同じかあるいはより効率のよい連携システムを実現できることがわかった。

本研究の目的の初期段階としてのデータ連携はそれによって実現が可能である。次にはデータを効率よく再利用するためのインターフェースとしてフィールドガイド出力のためのデータフォーマットのあり方、それを操作するためのインターフェースとその手法を研究する必要がある。そのためには、DIGGERの連携方式を採用し、そのデータの構造を明確にしフィールドガイドに必要なデータ項目を定義するための研究ステップが必要である。

6. 引用文献

7. 国際共同研究等の状況

なし

8. 研究成果の発表状況

(1) 誌上発表（学術誌・書籍）

＜論文（査読あり）＞

なし

＜その他誌上発表（査読なし）＞

- ① 佐藤 聡、伊藤 希、小野 哲、柁原 宏、志村純子：情報処理学会研究報告 2002-DBS-128, 255-262 (2002) 「生物種情報データベースのためのデータ交換方式」
- ② Satoshi Ono, Ryo Fujimoto, Takehisa Okada, Hiroshi Kajihara, Akira Sato, Nozomi Ytow and Junko Shimura: The implementation of database applying Nomencurator schema and the development of the user interface for it, 国立環境研究所報告 R-175-2003, 287-293 (2003)

- ③ Akira Sato, Hiroshi Kajihara, Takeshi Sagara, Satoshi Ono and Junko Shimura (2002):
Implementation Method for Data Exchange of Biodiversity Information Databases, *1st
GTI Global Taxonomy Initiative Regional Workshop in Asia-Abstract* p25, Putrajaya,
Malaysia, September 2002.

(2) 口頭発表
なし

(3) 出願特許
なし

(4) 受賞等
なし

(5) マスコミ等への公表・報道等

- ① 月刊DBマガジン 2004年10月号 076-081ページ PostgreSQL実力判定&120%活用術事例
- ② 国立環境研究所「GTIプロジェクト」

9. 成果の政策的な寄与・貢献について

本研究成果は、多様性条約に基づく世界分類学イニシアティブにおいて、データベースの統合とその利用技術のためのモデルケースとして、東南アジアを含む発展途上国のデータベース交換の仕組みについてその継続的利用に貢献する。