

国内所蔵生物標本を対象とした統合検索の可能性

栢野貴之 (筑波大学 生物学類 4年) 指導教官: 伊藤希 (筑波大学 生物科学系) 責任教官: 齋藤建彦 (筑波大学 生物科学系)

はじめに

コンピュータネットワークの普及にとともに、生物学的データベースが広く利用される様になってきた。遺伝子配列データベースとそれに基づく系統樹の構築はその代表例であるが、一方、分類学的データや標本データなどについてはそのデータベース化が遅れている。この不均衡を解消するため、近年、分類学的情報の整備を行なおうとする機運が高まっている。その代表格である地球規模生物多様性情報機構 (GBIF) では、必要に応じて新規データベースを構築する一方で、既存の分類学的データベースを横断的に検索するシステムを構築して、迅速に情報提供することを目指している。GBIF に参加する国及び団体は何等かの形で分類学情報データベースの公開を要求されており、わが国においても既存のデータベースについて十分な調査検討を行ない、これらを対象とした統合検索を提供する必要がある。その際、国内標本や和名等の情報はわが国からのみ提供し得るものである。そこで、本研究では全国科学博物館協議会の調査研究報告書を元に、全国の博物館が公開する生物標本情報のデータベース化の現状を明確にし、広範かつ体系的な利用のための統合検索の可能性について検討した。

方法

全国科学博物館協議会の Web サイト (<http://jcs.m.kahaku.go.jp/>) から得られた加盟館園が公開している 260 件の Web サイト全てにアクセスし、データベースを公開している URL、データベーススキーマの公開の有無、検索可能なデータ項目、検索結果として得られるデータ項目を調査した。但し、構造化された検索が可能なものだけを対象とし、静的な HTML コンテンツの集合体ないしそれに対する全文検索によって実現されているものは対象から除外した。また GBIF の国内フォーカルポイント Web サイトにある「国内の生物系データベースの現状に関する調査報告書 (http://bio.tokyo.jst.go.jp/GBIF/gbif/japanese/gbif_tyousa.htm)」(科学技術振興事業団, 2001) に挙げられている Web サイトのすべてについても同様の調査を行った。

こうして得られた結果について、複数の標本データベースのに対する統合検索スキーマである DarwinCore (http://tsadev.speciesanalyst.net/darwincore/darwin_core.asp) に定義されたデータ項目とのマッピングを行なった。

結果

全国科学博物館協議会加盟館園の Web サイト 260 件のうち標本データベースを公開している Web サイトは 8 件あったが、データベーススキーマを公開しているところはなかった。また科学技術事業団の調査に協力した Web サイト 132 件のいずれも標本データベースを公開してはなかった。公開されていたデータベース 8 件すべてで和名 (資料名) による検索が可能であったほか、学名、採集地、採集年 (月日) および分類群等の項目で検索可能なものがおよそ半数であった (表 1)。検索結果としては、すべてのデータベースで和名 (資料名)、学名、科名、採集地が得られ、ほとんどで採集年 (月日) も得られたが、一方で属名、採集者名が検索結果項目に含まれているデータベースは半数以下であった (表 2)。

複数の標本データベースに対する統合検索を目的とした DarwinCore は、必須項目として ScientificName、Kingdom、InstitutionCode、CollectionCode、CatalogNumber、Collector、Year、Month、Day、Country を要求している。これら 10 項目の全てを満たすものは一つもなかった。このうち、Kingdom などの自明な項目を除外すれば 1 件、Collector も除外する様要求を緩和すれば 5 件のデータベースが条件を満たした。従って、公

開されているデータベースの 6 割が DarwinCore の要求を概ね満たしており、これらが GBIF の目指す統合検索に迅速に対応し得ることになる。調査した 260 サイトのうち生物を対象としていたのは 80 サイトであったので、生物関連の Web リソースに対してこれらのデータベースが占める割合は 7% であった。研究者ないし博物館関係者への限定提供を念頭に、全国科学博物館協議会による標本資料データベースの標準化に関するアンケート調査で得られた未公開資料に基づき、一般公開されていない館内データベースについても同様の調査を行った。この結果の公表については現在協議中である。

考察

GBIF に対して日本からの情報提供の核となるのナショナルノードを用意する最も速い方法は、既存の生物データベースに対する統合検索ポータルサイトを実装する事である。現時点でこの対象となり得るのは上記 5 箇所であり、GBIF の中間評価が行なわれる 2005 年末までに比較的容易に国内ポータルを実現できると思われる。DarwinCore の必須項目を満たさなかった公開データベース、また一般公開されていないデータベースについても、協議の上可能であれば同様の実装を用意することが望ましい。

一方、データベースと称して情報公開はしているものの、実際には静的な HTML コンテンツの集合体であって、データベースを利用したものではないという Web サイトが多く見られた。これらに対して DarwinCore が想定している様な構造化された検索機能を提供するポータルの構築は、不可能ではないにしても困難と思われる。むしろ、Google の画像検索に見られる様な二次的検索機能をナショナルノードで提供の方が現実的と思われる。

表 1 公開されている標本データベースの検索可能項目

博物館名	和名 (資料名)	学名	採集地	採集年 (月日)	分類群	保存状態	登録番号	その他
茨城自然博物館	○	○	○	○	○	○		
群馬県立自然史博物館	○	○	○	○				英名
入間市博物館	○							
国立科学博物館 (淡水魚)	○	○	○	○	○		○	
岐阜県立博物館	○							
兵庫県立人と自然の博物館	○							
徳島県立博物館	○	○	○	○	○			
琉球大学資料館	○	○	○	○	○			方言名

表 2 公開されている標本データベースにおいて検索結果として得られた項目

博物館名	和名 (資料名)	学名	採集地	採集年 (月日)	採集者	属名	科名	目名	保存状態	登録番号	備考
茨城自然博物館	○	○	○	○		○	○	○	○	○	
群馬県立自然史博物館	○	○	○	○			○			○	検索結果は静的 HTML ファイル
入間市博物館	○	○	○				○		○	○	
国立科学博物館 (淡水魚)	○	○	○	○			○	○		○	
岐阜県立博物館	○	○	○	○	○	○	○	○	○	○	
兵庫県立人と自然の博物館	○	○	○	○			○	○			
徳島県立博物館	○	○	○	○			○				
琉球大学資料館	○	○	○	○			○	○	○	○	
Darwin Core	Scientific Name (必須)	Country (必須)	StateProvince	Year (必須)	Month (必須)	Day (必須)	Collector (必須)	Genus Family Order		Catalog Number (必須)	