

審査の結果の要旨

氏名 片山 俊明

本論文は、4章及び付録からなり、生命科学データベースの標準化と相互運用性の向上について論じている。本論文の構成では、第1章は、データベースの利活用における課題とそのためのソフトウェア開発や国際連携について論じている。第2章では、ウェブサービスを用いた主要なデータベースへのアクセスに関する標準化と相互運用性を論じており、第3章では、セマンティック・ウェブ技術を用いたデータベースのコンテンツに関する標準化と相互運用性を論じている。第4章ではこれらの技術開発研究にもとづいた改善点と将来の方向性について主に論じている。

第1章では、学位申請者が **Kyoto Encyclopedia of Genes and Genomes (KEGG)** の開発に従事していたことから、ゲノムやパスウェイのデータをコンピュータ上で容易に扱うために **BioRuby** ライブラリを開発するに至ったこと、さらに他のデータベースと連携してデータ解析を進めるためにデータベース間での標準化や相互運用性の向上の必要性について論じており、これを実現するために国際会議 **BioHackathon** を開催して国内外の主要データベース開発者間での技術協力を進めてきた経緯を、本研究の背景として論じている。

第2章では、ウェブサービスを利用して主要な生命科学データベースへのアクセス方法を標準化し、データベースエントリの内容を **BioRuby** などの機能を利用してパースおよび変換することで相互運用性を高める技術開発について述べている。これにより、基本的なデータベース利用に求められる機能は、プログラミング言語に依存せずウェブサービスを利用するだけで利用可能となっている。

一方で、上記の研究から、データベースの中身を統合的に利活用するためには、そのコンテンツ自体を標準化する必要性が明らかになった。このため、第3章では、セマンティック・ウェブ技術を用いた **TogoGenome** ゲノムデータベースの開発を行ったことについて述べている。**TogoGenome** では、ゲノムデータベースを構成する、生物種、ゲノム、遺伝子、タンパク質などの多様な情報を、セマンティック・ウェブ技術の **Resource Description Framework (RDF)** を用いて統合している。ゲノム情報に関しては **RDF** 化のための技術開発とオントロジー開発を独自に行い、国際塩基配列データベース **INSDC** や **RefSeq** の **RDF** 化を実現しており、すでに海外で **RDF** 化の行われていた **UniProt** のタンパク質機能アノテーションとともに統合している。この際に、塩基配列やアミノ酸配列の位置情報を記述するための **FALDO** オントロジーを国際連携によって標準化している。**TogoGenome** ではさらに、表現型や生育環境などの付加情報の統合を、

オントロジーを整備することによって実現している。その結果、これまでのデータベースでは困難だった、データベースの中身をオントロジーの概念階層や統制語彙を利用して絞り込み検索を行う、オントロジーに基づくファセット検索が可能となった。また、このようにして構築したデータベースの相互運用性を高めるため、**TogoGenome** では構成要素を再利用可能なモジュールとしている。そのための汎用的なフレームワークとして **TogoStanza** を開発しており、実際に国内の他のゲノムデータベースと相互運用を行い、その有効性を実証している。

審査委員会では、データベース統合におけるセマンティック・ウェブのより広い範囲のデータに対する適用性と、**TogoGenome** および **TogoStanza** の技術開発によって克服された技術的な工夫点についての質疑を中心に議論が行われた。本研究は、生命科学データベースの再利用性を向上させるための標準化と相互運用性を実現するための方向性を提示するとともに、ここで統合されたデータはデータサイエンスにおける基盤リソースとして広く利用されることが期待できる。

なお、本論文の、第2章は中尾光輝、高木利久との、第3章は川島秀一、岡本忍、守屋勇樹、千葉啓和、内藤雄樹、藤沢貴智、森宙史、高木利久との共同研究であるが、論文提出者が主体となって設計・実装と考察を行ったもので、論文提出者の寄与が十分であると判断する。

したがって、博士（科学）の学位を授与できると認める。

以上 1,788 字