

審査の結果の要旨

氏名 川島 秀一

本論文は、3章及び付録からなり、セマンティック・ウェブ技術を用いた複数のデータベースの統合に関して論じている。本論文の構成では、第1章は前書き、第2章は20種類のアミノ酸に関する物理化学的な指標のデータベース AAindex について、第3章は本論文の本題であるバイオサイエンスデータベースセンター (NBDC) の RDF ポータルについて、主に論じている。

AAindex は、タンパク質アミノ酸配列の物理化学的な数値的性質と配列アラインメントに用いられる置換行列のデータベースである。1990年代に開発されたフラットファイル形式の古いデータベースであるが、近年まで更新が続けられている。

複数のデータベースを統合して利用する問題に関して、本論文ではセマンティック・ウェブ技術を利用した。セマンティック・ウェブで用いられる情報は、XMLによる情報記述方式である Resource Description Framework (RDF)を用いて、主語、述語、目的語の3つのオブジェクトの組で表現され、個々のオブジェクトは Uniform Resource Identifier (URI)によって識別される。さらに、語彙レベルでのデータベース間での共通化のためには、Web Ontology Language (OWL)で記述することが有効である。しかしながら、(1) 同一オブジェクトが複数の URI を持つ、(2) 同じ概念に対して異なるオントロジーのクラスやプロパティが存在する、(3) 同一情報が異なる RDF で表現される、などの問題がデータベースの統合を困難にしている。学位申請者は、(1) ~ (3) を解決するための実用的なガイドラインを提案し、NBDC の協力を得て NDBC RDF portal という RDF データセットのリポジトリを開発した (<https://integbio.jp/rdf/>)。このリポジトリには、学位申請者自身がレビューを行って登録した 21 の RDF データセットが登録されている。その結果、RDF データセット間の高い相互運用性を実現した

審査委員会では、ガイドラインの策定と NDBC RDF portal に関する質疑を中心に議論が行われた。レポジトリに登録されている RDF トリプルは、455 億件を超えており、今後の生命科学データサイエンスにおける基盤リソースとして期待できる。

なお、本論文の、第2章は緒方博之、金久實との、第3章が片山俊明、畠中恵菜、榎田達矢、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1101字