

令和元年6月26日現在

機関番号：72690

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00412

研究課題名（和文）学際領域の発展を推進し支援する化合物表現法の開発

研究課題名（英文）Development of a method for expressing compounds to integrate interdisciplinary areas

研究代表者

山田 一作（小山内一作）（Yamada, Issaku）

公益財団法人野口研究所・研究部・研究員

研究者番号：50370185

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：各学問分野においては様々なデータ表現法がある。これらの分野毎に特色があり、異分野とのデータ統合は困難な場合が多い。本研究では、多くの研究分野において用いられている化学物質の化学構造に着目し、その部分構造を鍵として異なる分野のデータを統合化することを実現するための化合物の表現法を開発した。そして化学構造からその他の情報へもつながることができた。

研究成果の学術的意義や社会的意義

学問領域間では同じようなデータであっても表現方法が異なることにより情報の共有化や利活用がこれまで困難であった。本研究では化学物質の化学構造を鍵とした表現方法を開発することで、複数の領域における情報の共有化を実現するための障壁を低減させる方法を開発した。

研究成果の概要（英文）：There are various data representation methods in each academic field, and each field is distinctive, and it is often difficult to integrate data with different fields. In this research, we focused on the chemical structures of chemical substances used in many research fields, and developed a method of expressing compounds to realize integrating data of different fields by using the partial structure as a key.

研究分野：糖鎖情報科学

キーワード：化学構造 セマンティックウェブ オントロジー 複合糖質 糖鎖 蛋白質 脂質

1. 研究開始当初の背景

コンピュータやインターネットがない時代は、必要な情報を調査する際、雑誌や本を購入し目次や索引などから目的の情報に関連するデータを調べ、さらに引用文献等を調査するという大変な作業を行っていた。現在、同様のことを行うのは非常に簡単な作業になっている。このように、現在のウェブ検索では必要とする情報へたどり着くことが可能となってきたが、各情報に正確なメタデータが付加されていないことによる不要データの混在などの問題があり、得られた情報の確認の検証作業が負担となっている。メタデータが付与されオントロジーで整備されたデータが RDF で記述されたセマンティック・ウェブの実現された世界では、各データに適切なメタデータが付加されていることにより、コンピュータが正しくデータを認識して処理を行うことが可能となる。研究においてこのような正確性やデータの入手の容易さは重要であり、積極的にセマンティック・ウェブへの対応が進められていた。

このようなことから様々な学問領域においてセマンティック・ウェブを活用し、データの利活用促進を行う環境が急速に構築されてきていた。これは世界的な潮流となりデータ利用の新しい基盤構築へむけた大きな変革期であると考えられる。このながれに沿うようにライフサイエンス分野においては、科学技術振興機構(JST)・バイオサイエンスデータベースセンター(NBDC)による統合化推進プログラムにおいて、データベースのセマンティック・ウェブへの対応が加速され、様々な領域において研究開発が実施されていた。しかし、既存のデータのみでは学問領域を横断しデータを繋いでいくことは容易ではなく、学問分野毎の規則に従ってデータが作成されることで、学問分野内でのデータの連携・統合化は急速に進んでいた。一方、学問領域間の連携は進んでいるもののいくつかの問題点を抱えていた。その中で化学物質は様々な学問領域において重要な位置を占めており、化学構造式をハブとしたデータの統合化は NBDC, EBI, NCBI をはじめ様々なところで実施されていた。

糖鎖関連データにおけるセマンティック・ウェブへの対応は、当該領域の研究者が協力して国際的仕様を策定し、連携に向けた取り組みを行い、セマンティック・ウェブ技術の実装によりグライコミクスデータをリンクすることが可能であることを明らかにした。さらに、グライコミクスにおけるオントロジー GlycoRDF を整備した。このように、糖鎖関連データのセマンティック・ウェブへの対応について、日本およびアメリカ、ドイツ、オーストラリア、ロシア、スイス等の研究者が協力して実施しているが、糖鎖以外の領域との連携については、未整備であり基盤となる表記法の開発が望まれていた。

2. 研究の目的

セマンティック・ウェブへの対応が様々な分野において実施されているが、ライフサイエンスにおける遺伝子、蛋白質、糖鎖、代謝物、脂質などの化合物の記述方法は研究分野毎に異なることが現状であり、学問領域を横断した繋がりを構築することは容易ではない。そこで、化合物記述方法の共通基盤を開発し、セマンティック・ウェブ技術を活用して学際領域の発展を推進し支援するための基盤を築くために、化学物質の全化学構造を活用する手法を拡張し、化学物質を複数の部分構造とし、これらを適切に再構築することで、糖質、蛋白質、脂質について部分化学物質構造を鍵として学問分野のデータを繋ぐことを目的とする。

3. 研究の方法

本研究の目的を実現するためには、化合物を適切な部位で部分構造に分割する必要がある。そこで既に部分構造となりうる構造表記法を開発している糖鎖構造を含む複合糖質を研究対象の化合物構造とした。複合糖質の中でも糖蛋白質および糖脂質を対象とすることとした。複合糖質の核となる糖鎖構造に関しては、これまでグライコミクス分野において開発してきた糖鎖構造の表記法(WURCS)を利用することができる。そこで蛋白質及び脂質分野における表記について調査検討した結果、蛋白質の構造は、アミノ酸配列と結合部位として FALDO (Feature Annotation Location Description Ontology)を利用して表現する。脂質構造については IUPAC における Nomenclature of Lipids に示された表記法はあるが、データベース毎に構造表記法について異なるところもあり、統一した表現は困難であった。したがって、無理に統一した表現方法を利用しないこととし、結合部位については蛋白質と同様に FALDO を利用することとした。これらの部分構造をハブとする URI として、既存オントロジーを利用し、新たに複合糖質のオントロジーを開発、分野横断的な RDF スキーマの構築し、RDF データを作成し、分野横断的なデータを統合し、新しい化合物表現方法を開発する。

4. 研究成果

ライフサイエンス分野の情報についてセマンティック・ウェブ技術を活用して統合・連携し、学

際領域の発展を推進し支援するための基盤を構築するために、各分野におけるセマンティック・ウェブ技術の活用事例からその特徴と問題点の整理を行った。まず、セマンティック・ウェブを指向した糖鎖構造表記法である Web3 Unique Representation of Carbohydrate Structures (WURCS)を有効活用するために、糖鎖分野の GlyTouCan, GlycoNAVI, UniCarbKB などの各データベースおよび Glyco, GlycoRDF などのオントロジーの整理を実施した。そして糖鎖構造表記について研究者と議論した。脂質、蛋白質および代謝物分野については各分野のデータベース (LipidBank, Lipid MAPS, PDB, Uniprot, KNApSACk など) の格納情報について調査・整理した。蛋白質のデータについて UniProt における糖鎖修飾の記述方法の調査を実施した。そして、糖蛋白質や糖脂質などの複合糖質の表記法の開発を行った。糖タンパク質の Resource Description Framework (RDF)を用いた記述では、糖鎖とタンパク質の複雑な組み合わせについても表現できるように開発を実施した。化合物の部分構造のやグリコフォームなどについて SemanticScience Integrated Ontology (SIO)を利用、糖鎖情報は GlycoRDF オントロジー、結合部位情報は FALDO (Feature Annotation Location Description Ontology)を利用し、複合糖質の構造及び関連する疾患や書誌情報などを統合した複合糖質のためのオントロジーを開発し、RDF スキーマの作成、論文記載データの R D F 化を実施してこれらの情報が分野横断的に利用できることを確認した。

5 . 主な発表論文等

〔雑誌論文〕(計 0件)

〔学会発表〕(計 6件)

〔図書〕(計 0件)

〔産業財産権〕

出願状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

6 . 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号(8桁)：

(2)研究協力者
研究協力者氏名：
ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。