

課題名 (タイトル) :

マウス表現型情報の国際的統合及びデータマイニング

利用者氏名 : 脇 和規

所属 : 筑波研究所 バイオリソースセンター マウス表現型知識化研究開発ユニット

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

マウスはヒト疾患研究等において、遺伝子操作が可能な実験動物として中心的に用いられている重要なバイオリソースである。特に近年、国際マウス表現型解析コンソーシアムによる約 2 万遺伝子に対するノックアウトマウスの網羅的表現型解析のプロジェクト等に伴い、重要性とリソースの数がさらに増加していくと考えられる。さらに、各国で理研バイオリソースセンターにおける日本マウスクリニクのような網羅的なマウス表現型解析の為に組織が作られており、標準系統や多くのミュータントマウス等の解析が行われている為、今後マウス表現型解析実験データの増加も期待される。

従って、ますます増加していくマウスリソースが、相互に比較可能な解析データと共に入手出来るようにする為に、実験データとプロトコルの国際的な共有と統合が重要となる。しかし、異なる研究施設間のデータでは、それぞれ検査項目、結果を示す語彙や測定値の単位等に相違があり、データを単純に比較することが出来ない。さらに多くのデータから必要なものを探すのも困難になる。

そのような問題を解決する為に、オントロジーを取り入れて語彙の意味情報を扱うことが可能なセマンティックウェブが、次世代標準の技術の一つとして期待されている。理化学研究所においてもデータベースを統合する為の共通基盤として、セマンティックウェブを応用した理研サイネスが開発・運用されており、さらに、今年 7 月に行われたマウス表現型解析データについての国際会議「International Phenome Integration Meeting RIKEN/InterPhenome/CASIMIR」においても、セマンティックウェブの技術を用いた情報統合を国際的に行っていくことが話し合われた。

それらの背景から、今後多量の表現型解析データをセマンティックウェブの為に形式にしていけることが必要になる為、我々は日本マウスクリニクの表現型解析データをなるべく簡単に自動的にセマンティックウェブ用の形式にするシステムを開発している。さらに、異なる研究機関の表現型解析データをオントロジーを用いて比較解析可能なデータにし、メタ分析などのデータマイニングを行うことにより、各解析結果からより信頼性の高い結論を得ることや、結果の差異と各実験・解析内容とサンプルの違いから、解析結果の違いを生ずる原因を探ることに役立てられると考える。

2. 具体的な利用内容、計算方法

本課題は、次の 4 つの項目からなる。

- 表現型解析データの収集。
- オントロジーの作製。
- 収集したデータを解析用に変換。
- オントロジーを用いた、データの関連性の推論。
- データマイニング。

これらのうち、現在“オントロジーを用いた、データの関連性の推論”について、RICC を利用している。

(1) オントロジーを用いた、データの関連性の推論。

独自に作成したオントロジーを用い、理化学研究所バイオリソースセンターにあるマウス表現型解析プラットフォームである日本マウスクリニクのマウス表現型解析データと、欧州連合のマウス表現型解析コンソーシアムである Eumodic のマウス表現型解析データについて統合的解析を行う為、オントロジーとセマンティックウェブのデータを用いて推論処理を行う為のプログラムの開発を行っている。その過程において、開発を行っているプログラムの試用の為に、既存の公開されているオントロジー及びセマンティックウェブのデータである、ほ乳類表現型のオントロジーデ

ータについて、セマンティックウェブにおける標準的な書式である RDF における subclass of の関係の推移的推論処理を行った。

3. 結果

ほ乳類表現型のオントロジーデータについて、セマンティックウェブにおける標準的な書式である RDF における subclass of の関係の推移的推論処理を行った結果、推論前のオントロジーの最上位のクラスと subclass of の関係にあるクラスの数 が 33 であるのに対し、推論後の数は 7085 であった。

4. まとめ

記載した方法以外にも、より複雑な推論処理を試行したが、メモリ不足や実行時間が長い等の問題があった為、一層の工夫が必要である。プログラム開発の過程において今後より多くの試行が必要である為、今後も RICC を活用し、必要な機能や方法を検討して開発を行っていく。

5. 今後の計画・展望

来年度より、国際マウス表現型解析コンソーシアム (IMPC) における活動が始まる予定の為、表現型解析データの国際共有や活用についての動きが活性化していくことが予想される。その為、今後もマウス表現型解析情報の国際統合とそれらを用いたメタ解析などのデータマイニングについて、いくつかの方法により RICC を用いて試行し、システム構築に役立て、国際的な表現型解析の基盤構築へ寄与するとともに、解析結果を成果として役立てていきたい。

6. RICC の継続利用を希望の場合は、これまで利用した状況 (どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか) や、継続して利用する際に行う具体的な内容

計算対象となる日本マウスクリニックのマウス表現型解析データについては、最近いくつか出てきた為、現在それらのデータに合わせて処理プログラムとオントロジーデータを作成中である。そ

れらのデータを使って、推論処理を行う為のプログラムを既存の公開されているオントロジーデータで試用しながら開発を行っており、実際の解析データの準備が整い次第、これまで解析を行ったデータに加え、表現型解析実験データについても同様の解析を行っていく。また、さらに他のアルゴリズムでの解析も検討していく。

7. 利用研究成果が無かった場合の理由

最近、処理対象の実験データをいくつか受け取ったので、実データに合わせて処理プログラムを開発中。