

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。今後も国際コンソーシアムによる全遺伝子に対するノックアウトマウス作成プロジェクト等に伴い、重要性和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、いくつかの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ッ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必要がある為、我々は日本マウスクリニッ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変換し、メタ分析などのデータマイニングを行う予定である。

2. 今後の計画・展望

現在、国際的なマウス表現型解析のコミュニティにおいて、表現型解析データの国際共有の為にポータルサイト構築について話し合われており、今後データの増加が予想される。その為、複数の研究施設からのマウス表現型解析データを用いたメタ解析などのデータマイニングを行っていく為にいくつかの解析方法を RICC を用いて試行し、解析の為にシステム構築に役立てていきたい。

3. 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現在準備中であり、またデータが届いた後でも計算の前処理に多くの試行と作業を必要とする。今後届く予定のデータに対し、オントロジーを用いた処理を行う為の書式変換、他研究施設とデータを比較可能とする為のオントロジーデータの作成、オントロジーを用いた、相互比較可能なデータへの変換、表現型解析データや実験プロトコルに対して、数値計算

をする為の数値データへの変換作業を行い、作成したデータを用いてメタ分析等のいくつかの解析手法を RICC で試行し、解析システム構築に役立てる。

#### 4. 利用研究成果が無かった場合の理由

計算対象のデータがまだ準備中の段階であり、また、本研究においては、数値計算以前のデータの前処理の重要性が高く、実際に計算を行う段階までに多くの作業がある為。