



RDFストア間データ連結フレームワーク の開発およびオーソログ解析への適用

千葉啓和
基礎生物学研究所



©2014 千葉啓和(基礎生物学研究所) licensed under CC表示2.1日本

背景と目的

- **バイオデータベースのRDF化が進んでいる**
 - インターネット上の分散データベースが実現しつつある
- **SPARQLクライアントが未成熟なため、問い合わせ際の作業効率が低い**
 - 特に、RDFストア間でのデータ連結はやりにくい



- RDFストアに効率よく問い合わせを行うための枠組みを開発
 - **複数のエンドポイントへのアクセスにも対応**

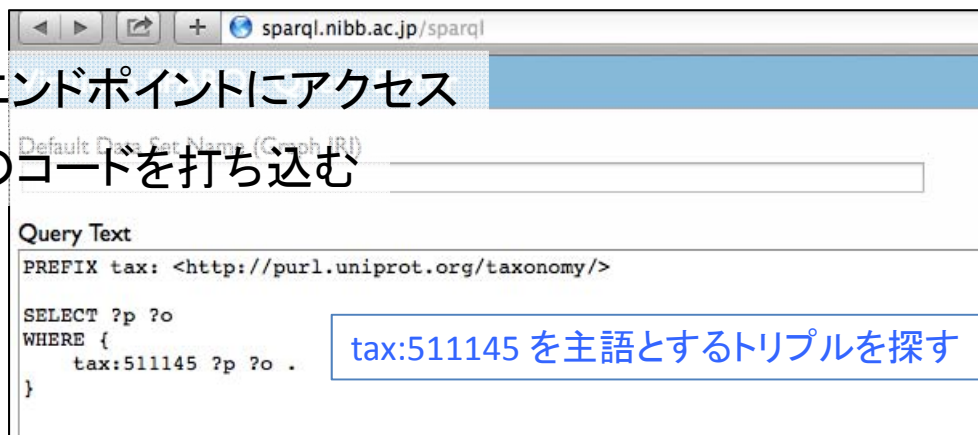


- **オーソログ情報を利用した遺伝子情報の統合的解析の試み**

通常のSPARQL実行方法

SPARQLエンドポイントにアクセス

SPARQLのコードを打ち込む



Query Text

```

PREFIX tax: <http://purl.uniprot.org/taxonomy/>

SELECT ?p ?o
WHERE {
  tax:511145 ?p ?o .
}

```

tax:511145 を主語とするトリプルを探す



p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://purl.uniprot.org/core/Taxon
http://www.w3.org/2000/01/rdf-schema#subClassOf	http://purl.uniprot.org/taxonomy/83333
http://purl.uniprot.org/core/otherName	"Escherichia coli MG1655"
http://purl.uniprot.org/core/otherName	"Escherichia coli str. K12 substr. MG1655"
http://purl.uniprot.org/core/otherName	"Escherichia coli str. MG1655"
http://purl.uniprot.org/core/otherName	"Escherichia coli strain MG1655"
http://purl.uniprot.org/core/partOfLineage	0
http://purl.uniprot.org/core/reviewed	0
http://purl.uniprot.org/core/scientificName	"Escherichia coli str. K-12 substr. MG1655"
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/1208340
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/1385755
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694514
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694515
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694516
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694517
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694518
http://www.w3.org/2004/02/skos/core#narrowerTransitive	http://purl.uniprot.org/taxonomy/694519

✓ 単純な処理にも煩雑な操作が必要

✓ 検索結果を利用した解析がしにくい



SPANG: a command-line client supporting query generation for distributed SPARQL endpoints

- コマンドによってSPARQLを動的に生成
 - 典型的コードパターンの生成
 - prefix宣言の生成
 - 基本的な問い合わせ作業を効率化
- クエリ毎にプロセスを生成
 - プロセス間通信により他の処理と組み合わせる
 - 複数のクエリを組み合わせることも可能
 - SPARQLコードのモジュール化による運用性の向上