

ライフサイエンス統合データベースセンター
山口 敦子
理化学研究所 情報基盤センター
戀津 魁

ライフサイエンス統合データベースセンター
山本 泰智
理化学研究所 バイオリソースセンター
榎屋 啓志

大阪大学 産業科学研究所
古崎 晃司
理化学研究所 情報基盤センター
小林 紀郎

SPARQL Builder は、SPARQL言語の知識がなくとも、また、対象データセットの構造を知らなくとも、クラス間関係提示を用いた対話的なGUIを介してSPARQLクエリを生成することができるウェブ上のサービスである。ユーザは対象データセットやそのデータセットを提供するSPARQLエンドポイントに対して、知識を持たないことを前提にしているため、システム側で適切なデータセットやSPARQLエンドポイントを利用するよう、ユーザを誘導する必要がある。

それを実現するため、SPARQLエンドポイントの信頼性監視システム **Umaka/Yummy Data** と連携し、蓄積されたメタデータを利用して、信頼性が確保されたSPARQLエンドポイントのみを利用するSPARQL Builderシステムの構築を行った。

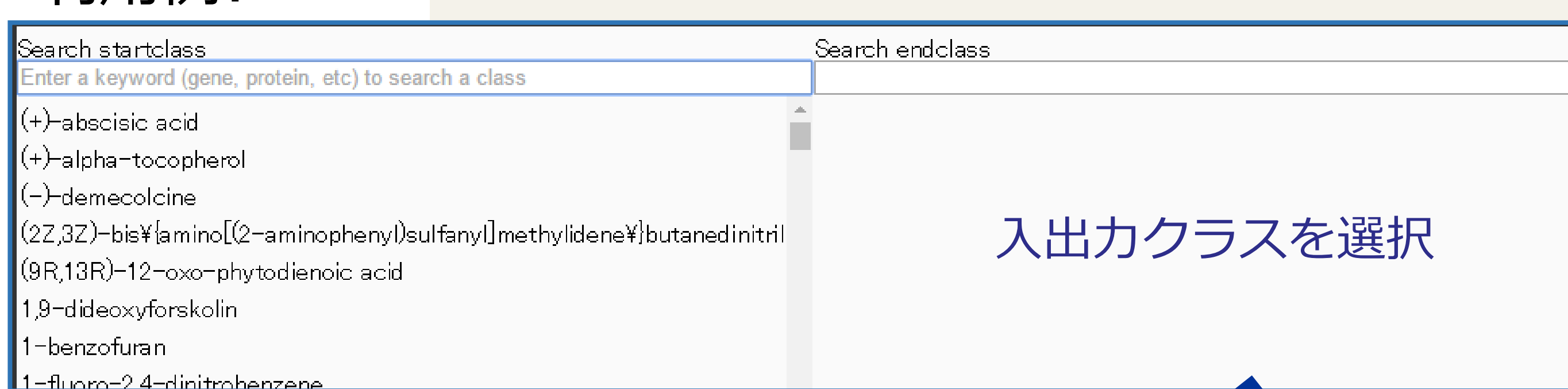
SPARQL Builder

SPARQL (RDFの標準検索言語) クエリの作成補助システム

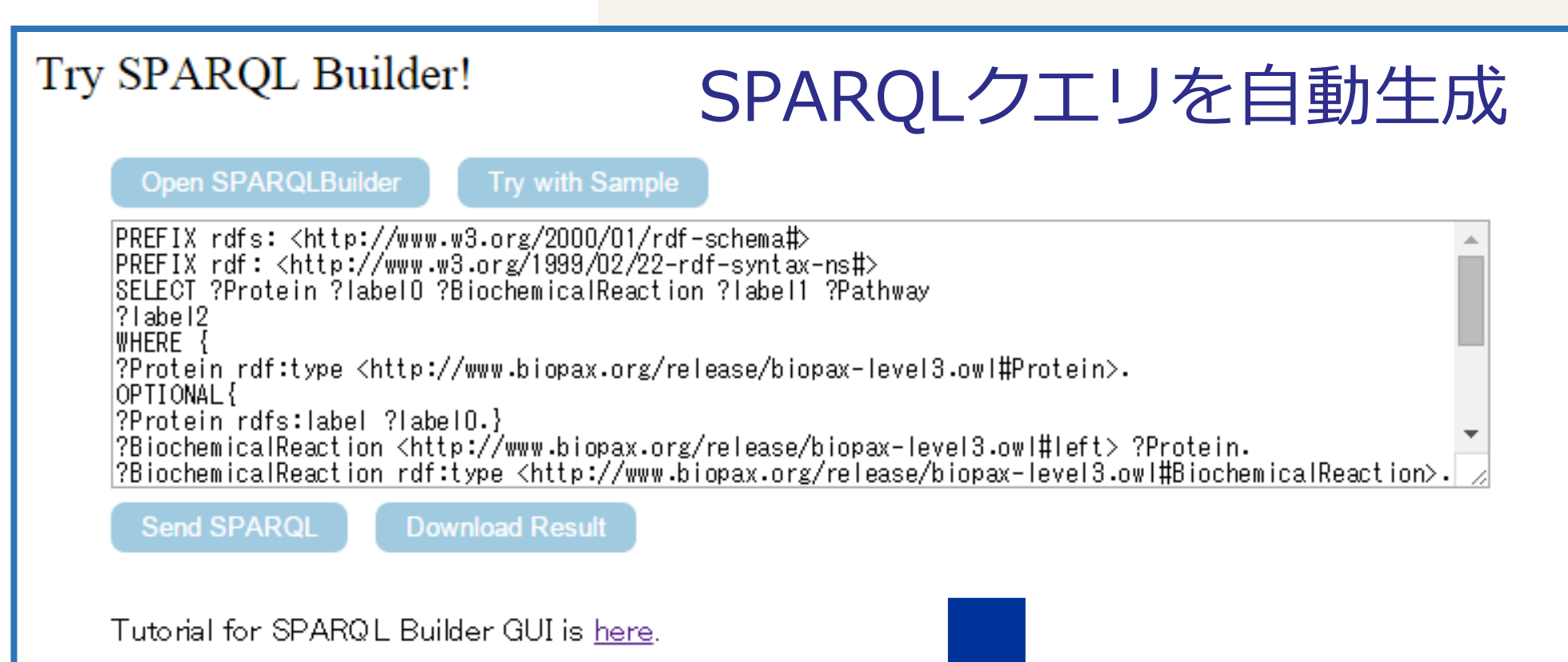
セマンティックウェブの知識がなくともDBのスキーマを知らなくともクラス(Gene, Protein, Pathway など)とクラス間関係を指定するとSPARQL クエリが書けるシステム

<http://sparqlbuilder.org/>

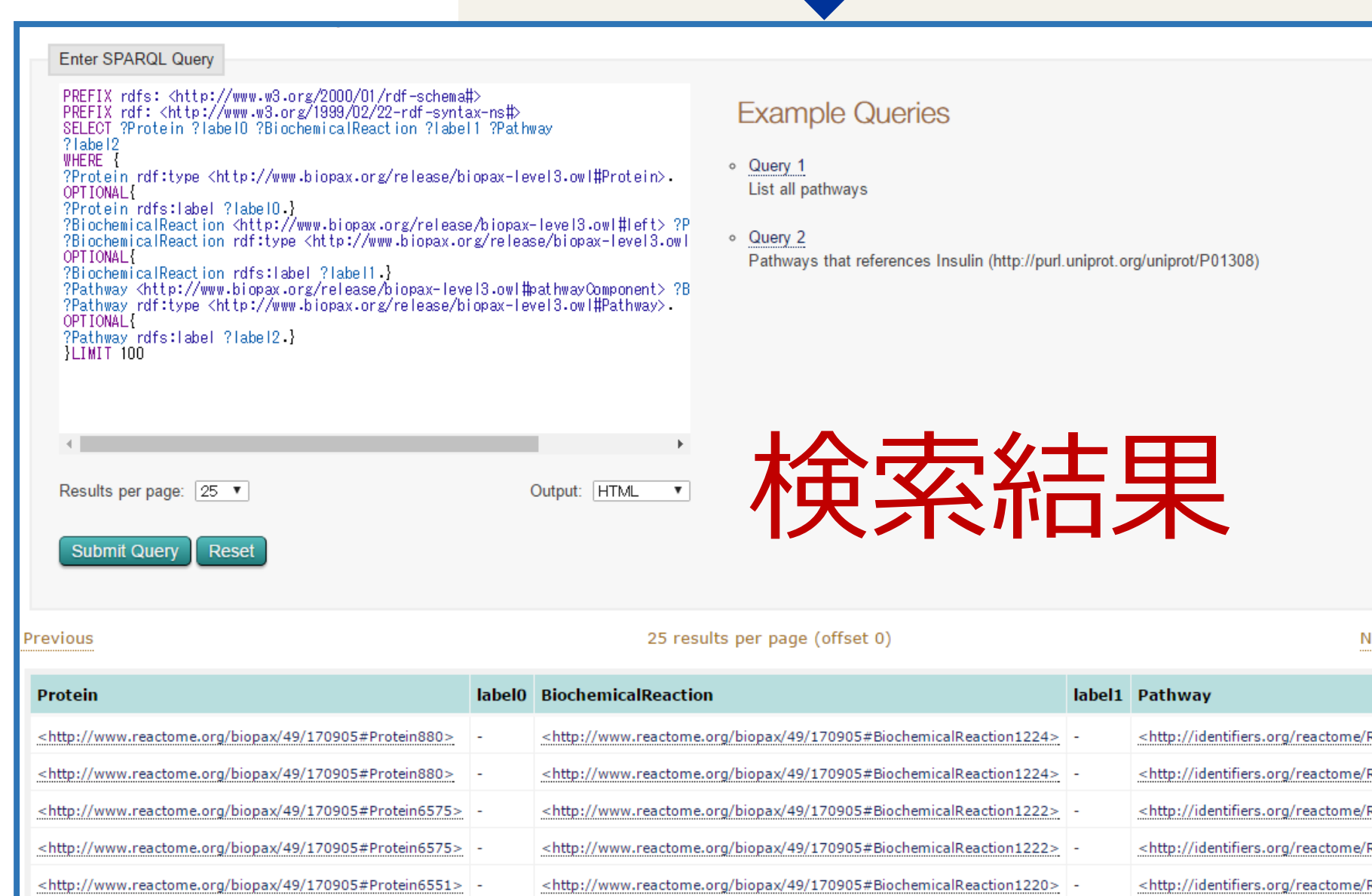
利用例:



入出カクラスを選択

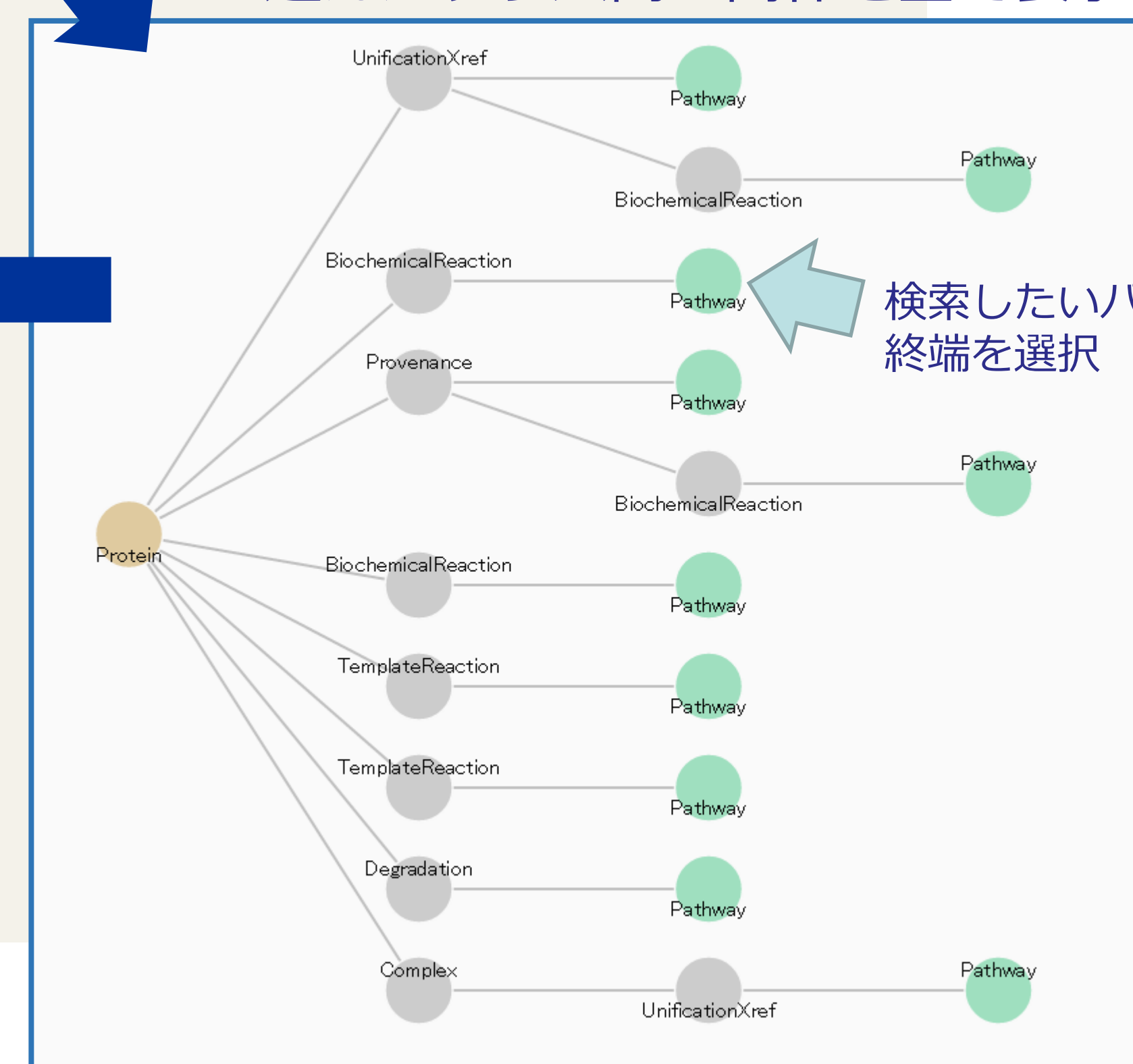


SPARQLエンドポイントへクエリを投げる



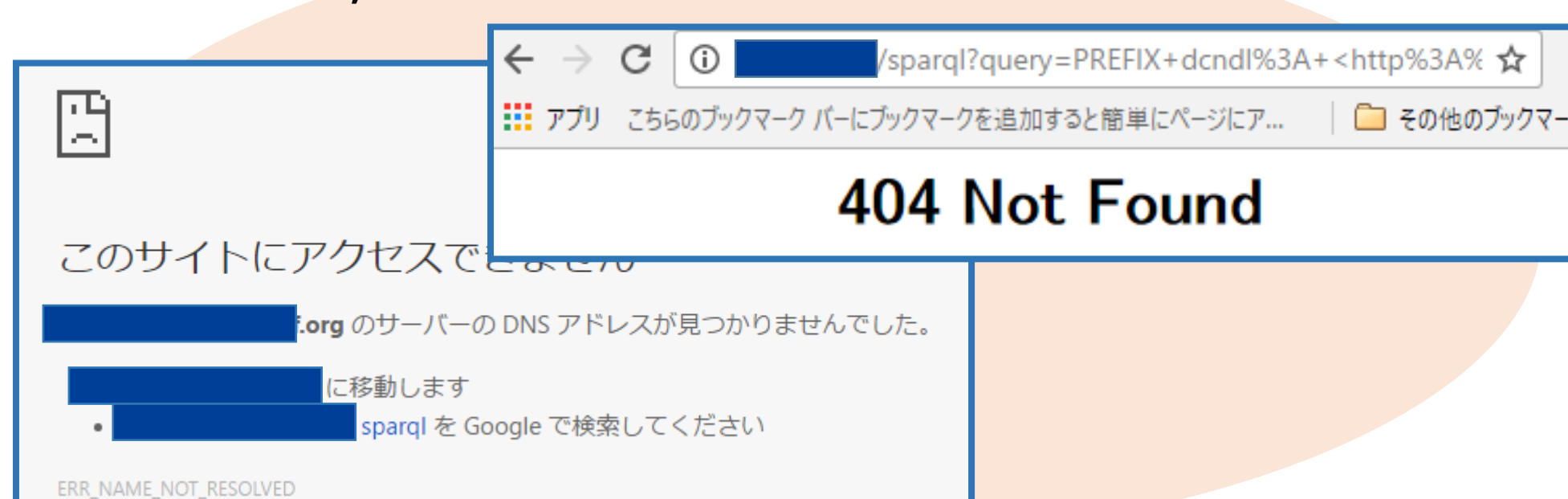
検索結果

選んだクラス間の関係を全て表示



検索したいパスの終端を選択

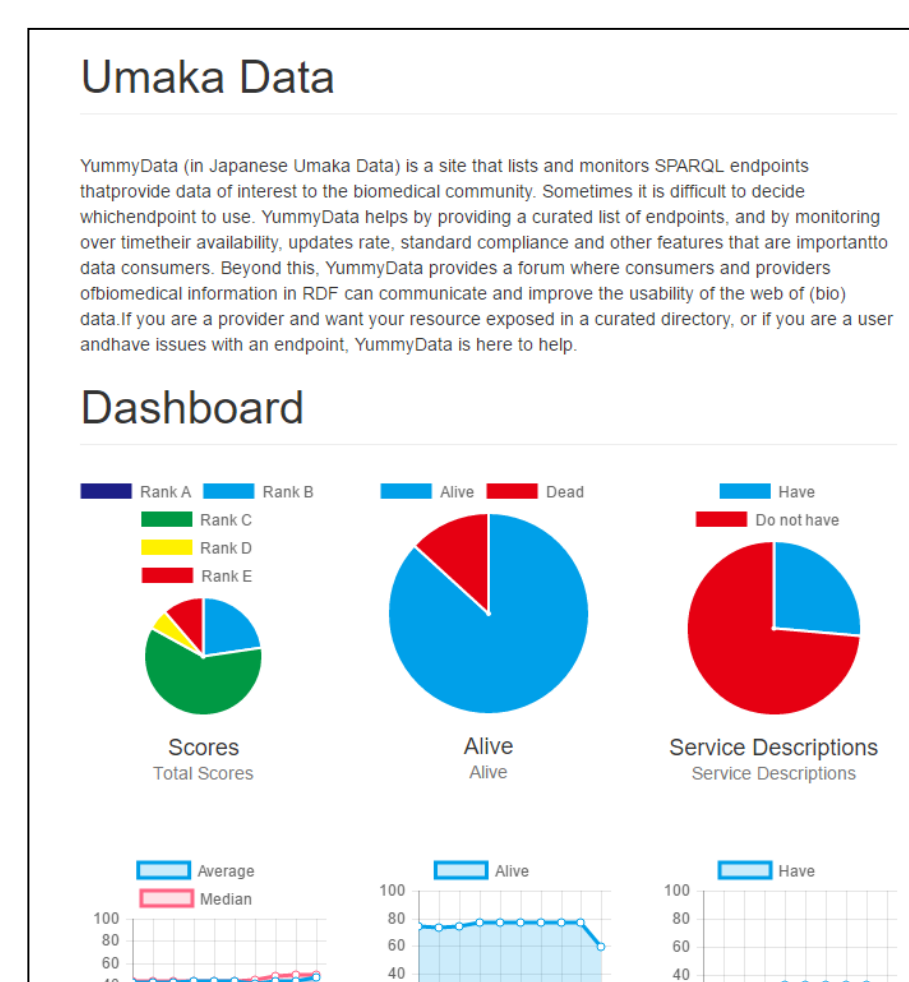
ところが、エンドポイントによっては...



検索結果が返らない
→エンドポイントの監視が必要

Umaka/Yummy Data

SPARQL エンドポイント監視サービス
SPARQLエンドポイントの情報を収集



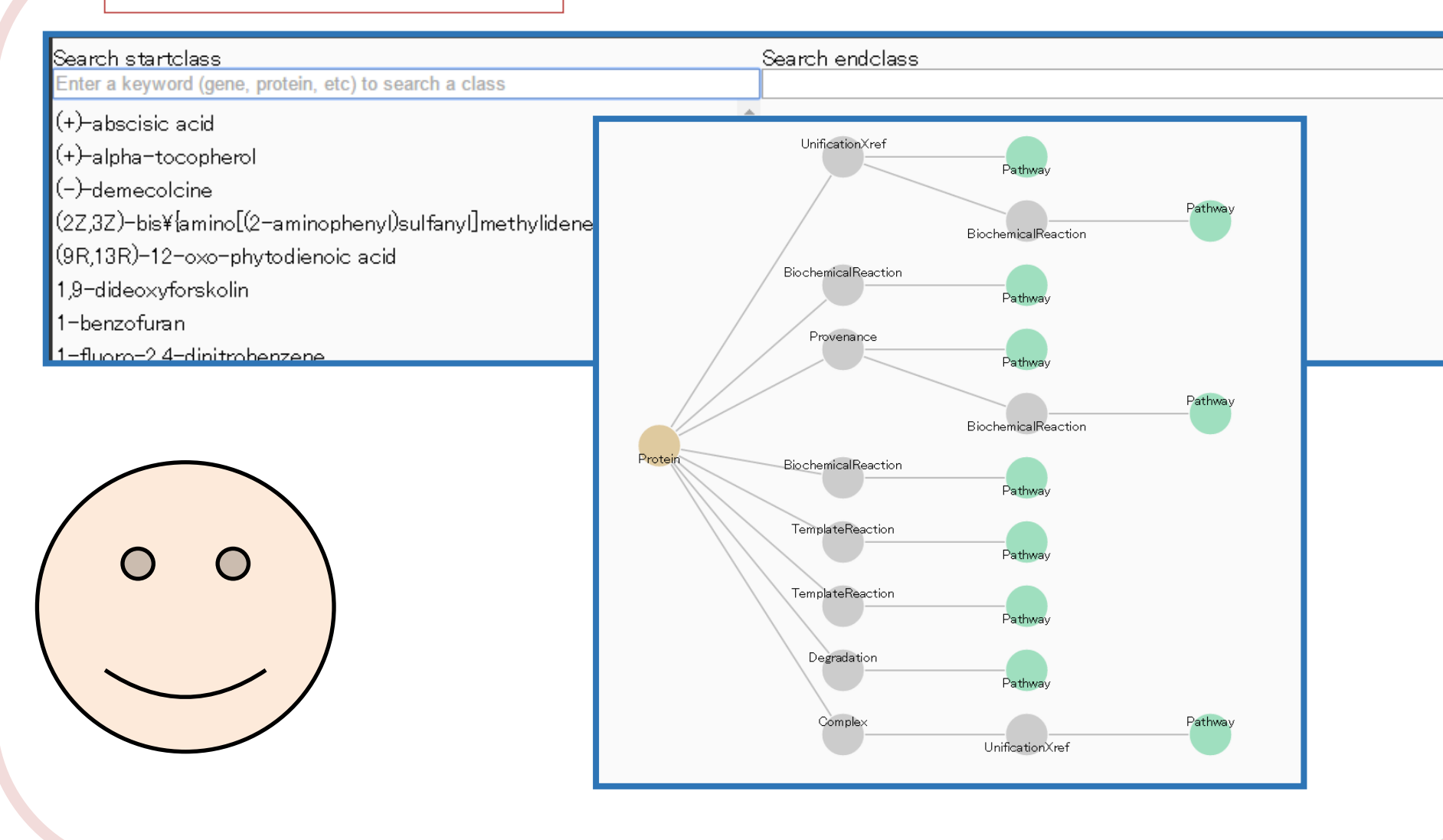
- ・ 死活情報
 - ・ 更新頻度
 - ・ クエリ応答時間
- など (詳しくはポスター12参照)

Umaka API
Alive率 ≥ 80%かつクエリ応答時間 ≤ 10秒のエンドポイントのリスト

安定したエンドポイントから信頼性のある結果の取得

SPARQL Builder

リスト内のエンドポイントに紐づいた情報のみをユーザに提示



113の生物系DBのエンドポイントが対象
ウェブページのみならずAPI(Umaka API)でも情報取得可能

今後の課題

Umaka/Yummy Data のAlive率、クエリ応答時間以外の項目の活用
例)クラスやパスのランキングに更新頻度を使うなど

