社会産業理工学研究交流会2025

自然観の偏在と共生システムにおける基盤構造の分析

瓜生真也（デザイン型AI教育研究センター/理工学部・助教）
uryu.shinya@tokushima-u.ac.jp

【はじめに】地球規模での“知識の空白”

生物多様性の全体像を理解しようとするとき、まだ解明されていない部分、「**知識の空白**」が大きく存在。

【問題】空白の正体：『自然観』の理解不足

人間社会が自然をどう捉えるかという『自然観』の定量的な理解が不足。
自然観は文化や言語、地域で大きく異なる（文化ギャップが存在する）。

【目的】自然観の偏りを定量的に解明し、その背後にある知識の構造を解析

①知識の偏りを見える化

文化や言語の違いで生じる
知識の偏りや非対称性を可視化

②関係性のネットワーク化

種や分類群の関係性を
ネットワークとして整理

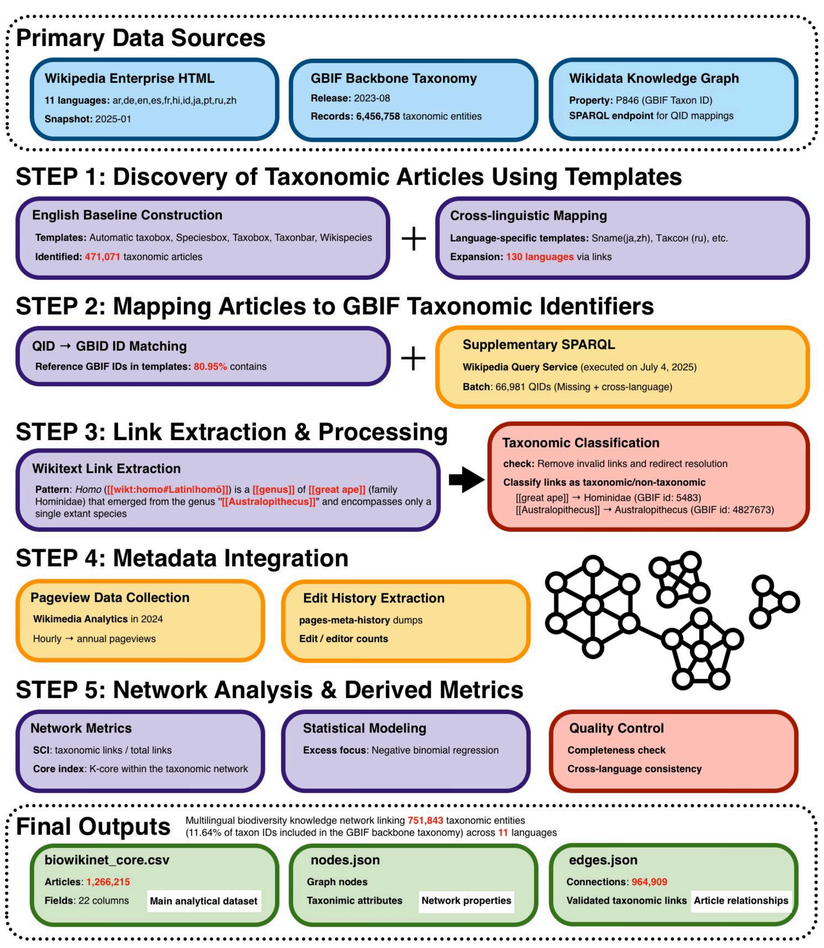
③隠れた知識の発見

AIの活用により、
未解明の関係性や知識を抽出

Biowikinet: 生物多様性の知識ギャップを捉える多言語データセット

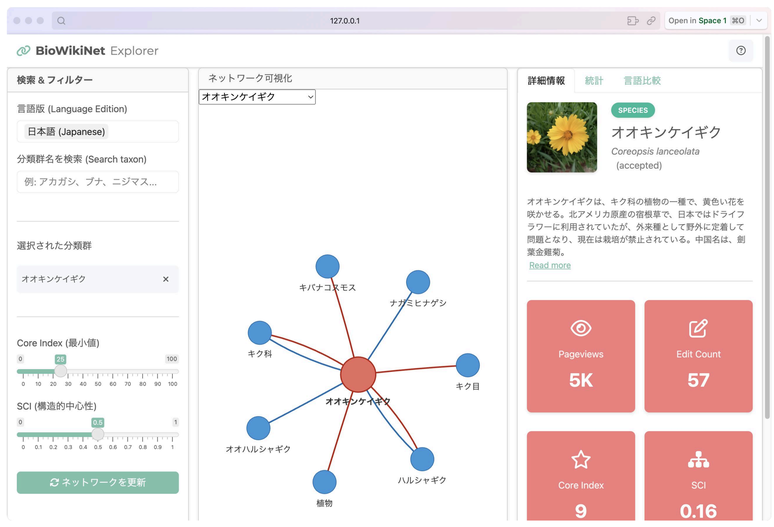
**11言語版のWikipediaから
126万の記事を抽出、
75.1万の生物種・分類群の関係をネットワーク化**

Shinya Uryu (2025).
BioWikiNet: A Multilingual Network Dataset Revealing Biodiversity Knowledge Gaps.
Figshare. <https://doi.org/10.6084/m9.figshare.29431577>

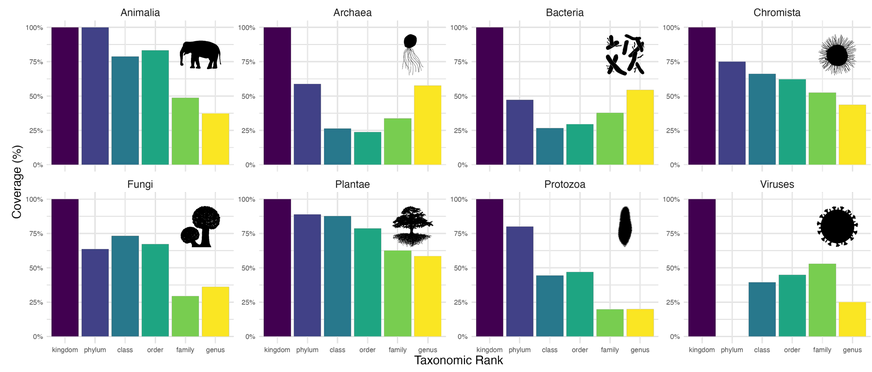


BioWikiNetデータセット構築のためのデータ処理ワークフロー。（1）テンプレートベースによる分類学的記事の発見、（2）GBIF分類学的識別子へのマッピング、（3）記事間リンクの抽出と分類、（4）ページビューデータと編集メタデータの統合、（5）派生指標を用いたネットワーク分析

（例えば）「**オオキンケイギク**」は...
どんな生物・分類群と繋がりをもつ？
日本と、**中国語圏**、**英語圏**で生態学的な地位が異なる？



データセットをインタラクティブに探索するためのShinyウェブアプリケーションを構築



データセットにおける8つの界にわたる分類学的網羅状況を示す。各階層レベルでGBIFバックボーン分類体系の何パーセントが表現されているかを可視化している。

【展望】AI活用による知識抽出、AIの影響評価（目的③の詳細）

①Wikipedia記事の構造（wikilink）を大規模言語モデル（LLM）で解析
→学術論文よりも包括的な**生物種間の関係を自動的に評価**、知識グラフとしてLLMの性能向上に貢献可能

②生物保全に関する固有表現認識モデルの構築
→既存の固有表現認識タスクでは評価されない「生物・分類群（TAXON）、環境（ENVIRONMENT）」を追加。
多言語に対応することで地域の生物情報抽出に寄与する

③大規模言語モデルによる生物保全に関するタスク（分類群、生息地域、絶滅リスク等）の性能評価
→ **現行のLLM、基礎的な知識は有しているものの、保全評価に必要な生態学的推論能力を欠いている**。
分類群ごとの保全努力や絶滅リスク評価数がLLMの知識にも影響。
保全に関する人間の評価を補助するためにLLMの活用が効果的（Shinya Uryu. 2025. arXiv）