

博士論文発表会(公聴会)のお知らせ

下記の通り学位論文発表会が開催されます。教員・院生・学部学生を問わず、多数ご参加ください。

日時：2025年1月30日(木) 16:00 ~ 17:00

場所：オンライン (Teams)

演題：大規模言語モデルを用いた生命科学データベースからの知識抽出手法の開発 (Development of knowledge extraction methods from life science databases using large language models)

演者：池田 秀也氏 (大学院統合生命科学研究科 数理生命科学プログラム)

要旨：

解析技術の発展とともに、生命科学分野では生命現象に関する膨大な知識が蓄積されており、特に大規模なオミクス解析の普及はその傾向を加速させている。産み出された実験データや、実験が明らかにした知識は、様々な公共データベースに蓄積され公開されている。過去の研究の積み上げてきた成果をさらなる研究へと発展させるためには、データベースを活用することは必須となっている。一方で、収載されているデータの量が膨大になっていることや、取り扱うデータの範囲の異なる多様なデータベースが多く開発者によって作られていることなどに起因して、データベースの利便性には様々な課題が生まれてきている。本研究は、実験サンプルデータベース BioSample に登録された実験メタデータの利便性を大規模言語モデル (LLM) を用いて向上させる取り組みと、様々な公共データベースのエントリー間の関係を簡便に取得できるウェブサービス TogoID (<https://togoid.dbcls.jp/>)の開発という、2つの研究で構成される。

BioSample を検索して得られた公共実験データは、様々なツールを用いて再解析することができる。また、再解析の結果、特定条件を満たす遺伝子の ID のリストが得られることが多い。こうした再解析のための入力の ID をツールの仕様に合わせて変換したり、得られた ID リストの遺伝子に関する情報を様々なデータベースから収集したりするために、TogoID が活用できる。このように、生命科学分野のデータベースの統合的利活用を促進するという目標に向けて、BioSample のメタデータの整理と TogoID での ID 変換はいずれも貢献できるものとする。

参考論文：

1. <https://doi.org/10.1093/bioinformatics/btac491>
2. <https://doi.org/10.1186/s13326-024-00322-1>

本学位論文発表会は、Teams によるオンラインミーティングです。

右の QR コード、もしくは以下のリンクからご参加ください。

<http://bit.ly/40huT7h>

または、

https://teams.microsoft.com//meetup-join/19%3ameeting_MGZhNTJhYjQtYmEIMS00MzA4LTk4OGEtNjk2MzhiYjk0MzU0%40thread.v2/0?context=%7b%22Tid%22%3a%22c40454dd-b263-4926-868d-8e12640d3750%22%2c%22Oid%22%3a%22bcad6aa4-ee2e-4b92-837c-c29c915cf139%22%7d

本セミナーは統合生命科学研究科セミナーとしてプログラム共同セミナーの対象です。

連絡先：数理生命科学プログラム 坊農 秀雅 <bonohu@hiroshima-u.ac.jp>

