デジタル・ヒューマニティーズとは？人文学研究におけるデジタル変革（DX）

ROIS-DS人文学オープンデータ共同利用センター：「データ駆動型人文学」と「人文学ビッグデータ」の展開

情報・システム研究機構 データサイエンス共同利用基盤施設（ROIS-DS）
人文学オープンデータ共同利用センター（CODH）
北本 朝展、小川 潤、加藤 幹治

どんな研究？

- **データ駆動型人文学**：情報学・統計学の最新技術を用いて人文学資料（史料）を分析
- **人文学ビッグデータ**：人文学研究の成果に基づき構築したデータセットを超学際的に活用
- **人文学のデジタル変革**：オープンサイエンスなど新しい潮流を取りこんだ人文学研究へ

何がわかる？

- **データ駆動型研究を進めるための、機械可読データセットを構築・公開**
- **オープンソースソフトウェアを公開し、各種のサービスを外部からも活用**
- **共同研究を通して知識や資源を提供**

背景・目的

データサイエンス共同利用基盤施設（DS）は、情報・システム研究機構（ROIS）内に設置された研究組織。生命科学・地球科学から人文学・社会科学まで、データサイエンスを幅広い分野で推進する7つのセンターがある。

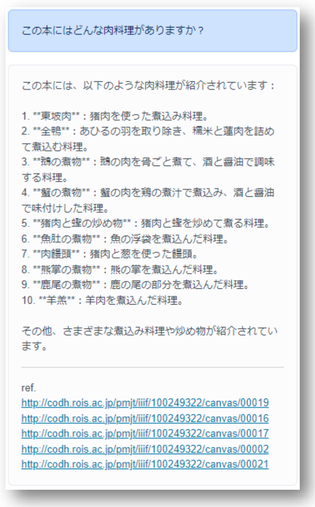
人文学者や情報学者などが分野横断的に協働し、人文学的な問いを情報学的な手法で解く、人文学資料から作る過去のビッグデータを分析する、などの研究に取り組む。人文学的な視点は、AIなどのテクノロジーを社会に取り入れる際の指針としても期待が高まっている。

研究内容

みを：AIくずし字認識アプリ

- **物体検出アルゴリズムを用いて、画像からくずし字を認識し、現代日本語文字に変換**
- **iOS・Androidアプリを無料で公開、ダウンロードは約18万回**
- **これまでの認識画像枚数は約230万枚、教育や調査に幅広い利用実績**

つくし：歴史的日本語と基盤モデル



- **AIくずし字認識を用いた、歴史的日本語の大規模データセット構築**
- **歴史的日本語データセットによる大規模言語モデルLLMの学習**
- **大規模言語モデルを活用するアプリの開発とUXの探求**
- **日本文化の新しい研究基盤を創出**

そあん：くずし字生成アプリ

- **日本の出版史上もっとも美しい書物の一つと言われる「嵯峨本」の古活字を自動切り出し**
- **デジタル組版ソフトウェアを開発し、任意の現代日本語テキストをくずし字に変換可能**
- **ウェブサービスやLINEボットを公開し、くずし字をコミュニケーションに活用する「日常化」を推進**

歴史ビッグデータ

- **過去の資料から構造化データを抽出し、現代のビッグデータ技術を活用して過去を復元**
- **地名識別子（GeoLOD）を拡大し、歴史地名と地図を結合**
- **データ構造化ツール、データ公開サイトなどを複数開発し、歴史研究のデジタル化を推進**
- **古地震や古気候などの分野で、実世界データを検証**

連絡先：ROIS-DS人文学オープンデータ共同利用センター／国立情報学研究所　北本 朝展
URL：http://codh.rois.ac.jp/　Email：kitamoto@nii.ac.jp