

ROIS-DS人文学オープンデータ共同 利用センター（CODH）

<http://codh.rois.ac.jp/>

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

研究背景



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6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ためのガイドとしても重要になりつつある。

研究内容

江戸地名データ
「江戸切絵図」



商業ビッグデータ
「江戸買物案内」



観光ビッグデータ
「江戸観光案内」



文書空間 ↔ 実体空間

文書を読み解きながらデータを構造化したい
実世界の構造に合わせてデータを構造化したい



古地図のジオレファレンス



データポータル「edomi」

地名識別子

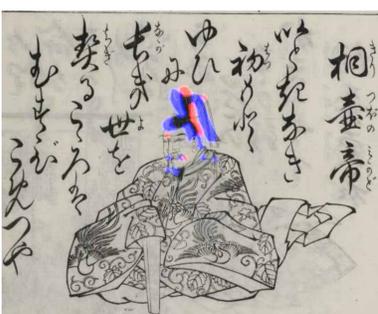


文書空間と実体空間を、識別子を介して双方向に結合する

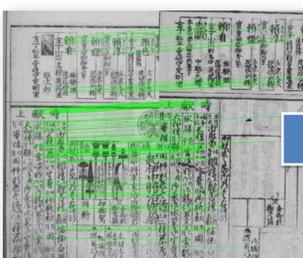


江戸ビッグデータ

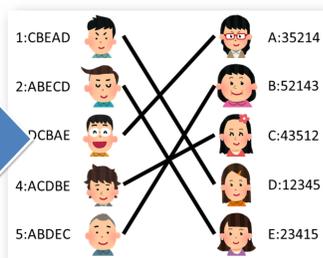
武鑑全集



同一板木だが一部だけ修正されているかを確認



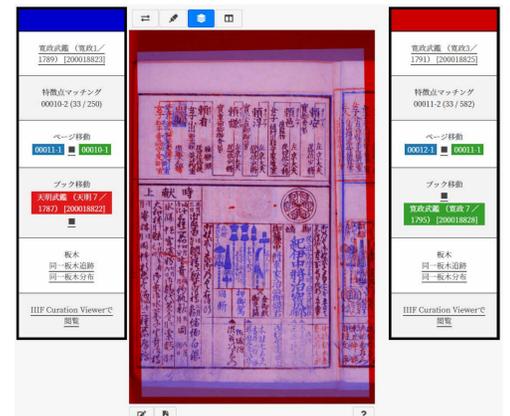
1. ページ照合：特徴点ベースの画像マッチングアルゴリズム



2. ブック照合：画像マッチング結果を利用した安定結婚アルゴリズム



3. 板木追跡：ブック照合結果に基づき、同一板木を書籍横断的に接続



vdiff.js：画像比較ツールにより同一板木の更新を強調



連絡先：北本朝展 / ROIS-DS人文学オープンデータ共同利用センター センター長

TEL : 03-4212-2584 FAX : 03-4212-2612 Email : kitamoto@nii.ac.jp