

貴重書デジタルアーカイブのための画像処理ワークフローの構築

Image Processing Workflow for the Digital Archive of Rare Books

北本 朝展 (Asanobu KITAMOTO) 国立情報学研究所 <http://dsr.nii.ac.jp/>

どんな研究？

貴重書をデジタルアーカイブして活用するためには、種々の画像処理技法を効果的に組み合わせる必要があります。そこで一連の画像処理手法を提案し、画像取得から画像検索までを一貫して自動処理するシステムを構築しました。

何が嬉しい？

デジタルカメラで撮影した画像を、デジタル漂白処理で読みやすくしたり、自動文字認識で全文検索できるようにしたり、画像自動抽出で画像検索ができるようにしたりなど、デジタルアーカイブの利用方法が大きく広がります。

画像処理ワークフロー

画像取得

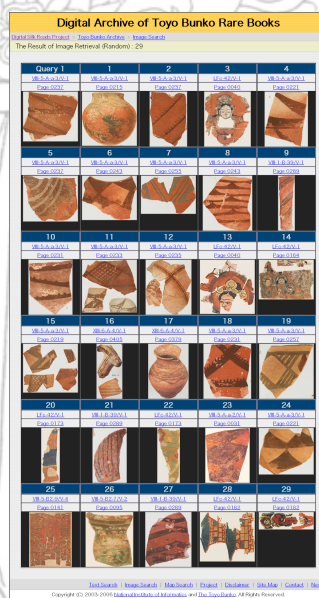
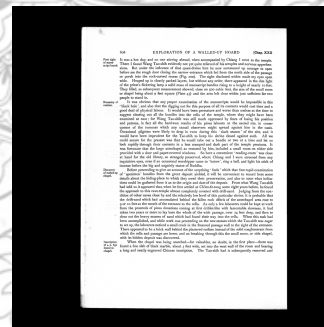
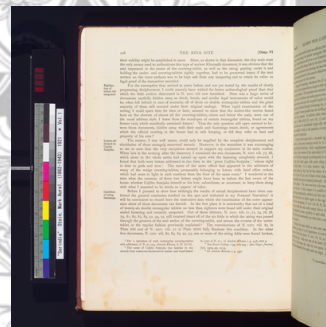
ページ領域抽出

デジタル漂白

コンテンツ抽出

コンテンツ索引付け

ウェブサイト



ページ領域抽出およびデジタル漂白処理については、約14,000ページに対して完全に自動化を達成。コンテンツ抽出にはまだ誤りが多く、研究の余地があります。その成果はウェブサイトでご覧下さい。

ネットワークを利用するデジタルアーカイブ共同構築システム

Network-based Collaborative System for Digital Archives

ASPICO

小野 欽司

北本 朝展

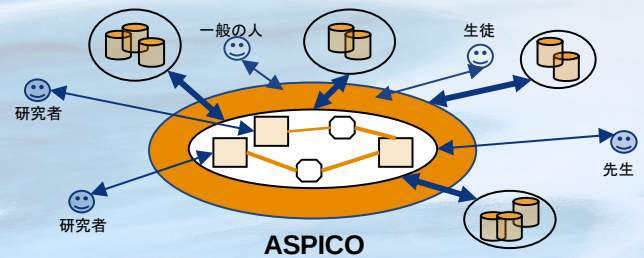
Dominique DEUFF

池崎 友博

Nigar ABBASZADE

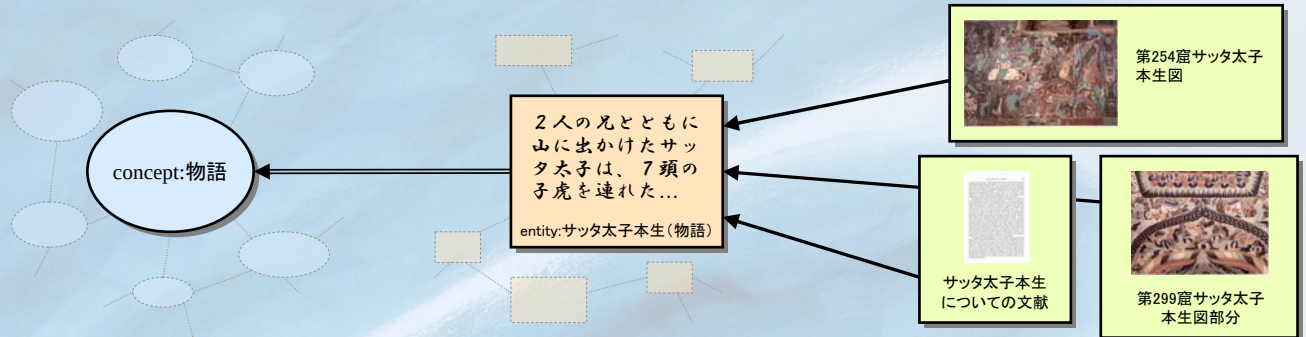
研究の目的

本研究では「デジタル・シルクロード」プロジェクトの一環として、文化遺産関連のさまざまな分野の研究者が、データの入力から公開までを遠隔／共同で行うことができるシステムの構築を目標としています。さらに、将来、蓄積されたデータを教育や新たな研究などに利用することを目指しています。

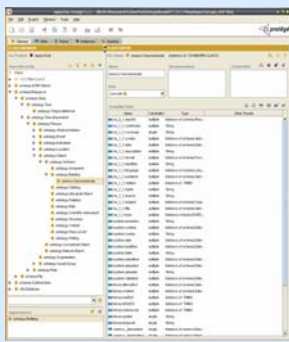


システムの特徴

ASPICOは、文化遺産情報をオントロジー・エンティティ・データという3つの階層で表現し、それらの関係をRDFによって記述するという設計に基づいており、部分画像へのアノテーションやエンティティの地図へのマッピング等の特徴的な機能を備えています。

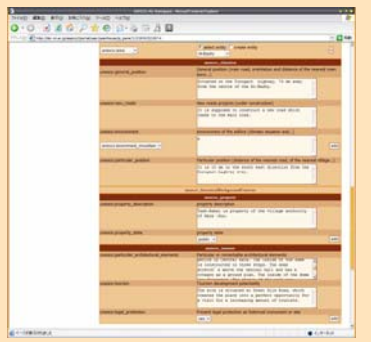


オントロジーの作成



Protégéによるオントロジーの編集

エンティティデータの作成



オントロジーから入力フォームを生成

データへのアノテーション

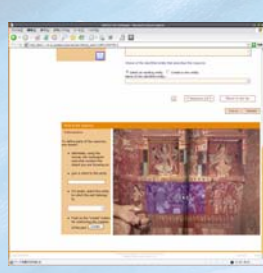


エンティティへのリンクを記述

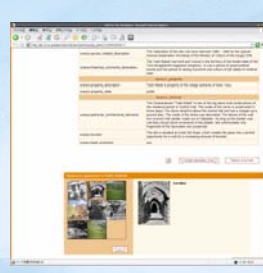
英・仏・露 インターフェイス



部分画像への アノテーション



エンティティ関連 データの閲覧



エンティティを 地図上で表示



文化の研究と教育を促進する貴重書デジタルアーカイブ

北本朝展 大西磨希子 ドミニク・デュフ マイヤー恵加 村松賢子
神田涼 佐藤園子 山本毅雄 小野欽司 池崎友博

貴重書デジタルアーカイブ

「東洋文庫所蔵」図像史料マルチメディアデータベース (<http://dsr.nii.ac.jp/toyobunko/>) では、
財団法人 東洋文庫 所蔵のシルクロードに関連する学術文
献や写真集などの貴重書を、全ページ、デジタル資料として
ウェブサイトで無料公開しています。
(2006年6月現在 計53冊 総ページ数13,973)

研究の目的

デジタルアーカイブには、①資料の劣化や破損・消滅を防ぐ、②広くアクセス性を高める、③資料の検索性が向上する、といった利点があるだけでなく、④デジタル修復による再現や原資料では困難な画像の張り合わせ作業など、デジタル資料ならではの研究・教育への活用が可能です。

本プロジェクトでは、貴重書デジタルアーカイブを新たな情報基盤として研究や教育に活かすための様々な取り組みを行っています。

活用の試み

「東洋文庫所蔵」図像史料マルチメディアデータベース



多言語
(現在・英語・仏語・独語・露語・伊語・
瑞語・漢語の8ヶ国語)

専門性が高い学術資料中心

研究 原著そのものの情報(テキスト・図)が見たい

1. 検索 ・OCRや翻刻による全文検索
・連想検索
・多言語キーワード検索(MASS)



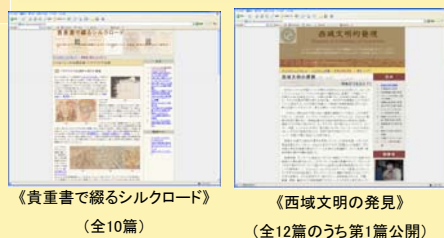
2. 表示 ・サムネイル
・カラー画像
・白黒高解像度画像
・見開き
・キャプション一覧
・画像の回転機能



一般 分かりやすい切り口でシルクロードやアーカイブ資料について知りたい

さまざまなトピックに沿って、原著とは別の文脈や別の分類によって提示
→ もととなったデジタル資料の窓口を増やす意義も

1. 文章主体(シルクロード談義)



2. 画像主体

(準備中)



子供 アーカイブ資料を教育に活用したい

- ・シルクロードに関する基本的な情報
- ・シルクロードを探索した人物や遺跡などの紹介
- ・遊びをととした学習(めりえ・クイズ)
- ・教育者用ページ

