

令和元年6月14日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00456

研究課題名(和文) アノテーション付与型画像データベースプラットフォームの応用に関する方法論構築

研究課題名(英文) Methodology for Application of Database Platform System for Annotated Images

研究代表者

和氣 愛仁 (WAKI, Toshihito)

筑波大学・人文社会系・准教授

研究者番号：70361293

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、研究代表者がこれまで構築してきた「アノテーション付与型画像データベースシステム(資料画像にテキスト情報を付与し、画像から文字・単語等の言語データを検索できるシステム)」の機能拡張を行った。主要な内容は以下の2点である。(1)画像へのアクセスの標準化と総合運用性の確保を目的として普及が進んでいるIIIFの枠組みを用いて、画像の著作権の保護に配慮した上で、外部システムとデータ連係を行う仕組みを実装した。(2)画像資料内の特定ページの本文テキストをgoogle等の検索エンジンに提供することで、検索エンジンから当該資料の特定ページに直接アクセスすることを可能にする仕組みを実装した。

研究成果の学術的意義や社会的意義

最近では、世界中で様々な人文学資料のデジタル化が急速に進行している。しかしながら、資料の共有という観点から見ると、資料の著作権等の理由により、個人的利用、組織内利用、条件付き公開など、様々なレベルのものがあり得る。そのような状況下でも、本研究に示した方法を用いれば、資料を適正な保護のもとに必要な範囲内で、かつ、標準的な手段によって公開することができる。もし公開の範囲に変更が生じれば、非常に簡便な方法でその範囲を変更できる。このことにより、人文学資料のより効果的な共有と発見性の向上に寄与できることが本研究の重要な意義である。

研究成果の概要(英文)：With the present research we have extended some of the functionalities of the Database Platform System for Annotated Images we have created so far. The main functions extended are as follows. (1) With the functions of International Image Interoperability Framework (IIIF), the system enables sharing images in consideration of image copyright; (2) by providing the text of a particular page in a multi-image document to a search engine such as Google, the system allows the search engine to directly access a particular page of the document images

研究分野：日本語学(日本語文法論)、人文情報学

キーワード：人文情報学 画像資料 データベース

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

研究代表者はこれまで、多様な人文学資料を少ない手間で共通的に扱うことを目指して、アノテーション付与型画像データベースの汎用プラットフォーム(以下「本システム」)を構築してきた。本システムは、様々な時代・地域の言語資料画像を、共通のユーザインターフェイスと標準的なデータ構造により取り扱えるようにするもので、資料画像とテキストデータを結びつけ、画像から直接文字あるいは単語等の言語学的データを検索できるようなシステムである。研究開始段階で、このプラットフォームを利用して、古代エジプト語神官文字パピルス資料、明治期日本語文典資料、古代シュメール・アッカド語楔形文字年度版資料のデータベースを構築済みであった。本システムは、多様な言語資料を共通の手法で扱うことができるという意味で汎用性を持つものではあったが、画像へのアクセスの標準化と総合運用性の確保を目的として近年世界的に普及が進んでいる IIIF (International Image Interoperability Framework) には対応しておらず、より標準的な方法で画像を扱えるようにする必要が生じていた。IIIF に対応することで、本システム上の言語資料も、より効果的に共有が進み、発見性の向上が期待できる。そのようなことから、本システムをより応用的に利用するための方法論構築を目的として本研究を開始した。

2. 研究の目的

本研究は、アノテーション付与型画像データベースの汎用プラットフォームの完成度をより向上させるとともに、このプラットフォームを通じて出力される資料データを様々な外部システムと連携させ、蓄積したデータを広範かつ有効に活用するために必要になる技術的・理論的な方法論を確立することを目的とする。このような共通プラットフォームを通じて出力されるデータが、既存のものを含む各種のシステムと有機的に連携することが可能になれば、個別に閉じた世界の中で扱われてきた人文学資料がより標準的な尺度の元で共通的に扱えることになり、人文学全体に対して極めて大きなインパクトを与えることができる。

ただし、単純に資料画像を公開すればよいというものではなく、資料の公開範囲には、資料の著作権等の理由により、個人的利用、組織内利用、条件付き公開など、様々なレベルのものがあり得る。そのような状況下でも、資料を適正な保護のもとに必要な範囲内で公開できなければならない。また、その手段は、可能な限り標準的なものに従うべきである。そのようなことから、資料画像を IIIF によって公開しつつ、必要な範囲に対して画像へのアクセス権限を適切にコントロールできる仕組みを構築することを目的として本システムの機能強化を図った。

3. 研究の方法

(1) 上記の目的を達成するため、外部システムとの連携のための方法論構築に関して、まず、本システム内部での画像の取り扱いを、IIIF 画像 API に準拠した形に改めた。本システムは、以前は資料画像のソースを Zoomify ネイティブのタイル画像(ローカルファイル)に依っていたが、今回、このソース画像取得部分を、IIIF に対応した画像サーバ Cantaloupe を導入して独立させ、システム内部の画像利用、および、外部公開用の両方のためにこれを利用することとした。

(2) 著作権に配慮した画像の公開のために、Cantaloupe のデリゲートスクリプトを利用して、必要な範囲に対して画像へのアクセス権限を適切にコントロールできる仕組みを実装した。また、画像情報 info.json に対して、画像の権利およびライセンスに関するプロパティ(attribution、license、logo の3つ)を追加した。

(3) 画像資料に対する IIIF マニフェストおよび IIIF アノテーションリストを、ソース画像およびデータベースから自動的に生成するようにした。この IIIF マニフェストを Mirador 等の外部ビューアに読み込ませると、本システム上に保存されている言語データが、外部ビューアのアノテーションポップアップ内に表示される。アノテーションポップアップ内に表示される文字列には本システムの検索ページへのリンクタグが記述されており、このリンクをクリックすると、外部ビューア上から本システムの検索ページへ飛んで、当該の文字・単語・品詞を直接検索することができる。

(4) 資料がオープンアクセスを許可している場合は、資料画像ページに本文テキストを配置し、これを google 等の検索エンジンに提供することによって、検索エンジンから資料内の特定ページにアクセスできるようにした。本システムでは、資料自体が何ページで構成されていても、Drupal 上では1画面のみ作成すればよいようにした。その一方、資料画像中の各ページに対しては仮想的に固有の URL が与えられるので、google 等でテキスト検索を行った場合、検索結果から資料画像中の特定のページに直接アクセスすることができる。

4. 研究成果

(1) 以上に述べた機能強化により、画像の著作権に一定の配慮をしつつ、よりインタラクティブな形で資料画像や言語データの共有を図ることが可能になった。資料・データの公開におけ

る複数の許可レベルをひとつの枠組みで扱えるようにすることは非常に重要な意義を持っている。本システムが実装した画像へのアクセス認可判断基準はあくまでひとつの例ではあるが、基本的な考え方と具体的な実装において一定程度の成果は示し得たと考える。また、外部システムとの連携についても、例えば Mirador のアノテーション内へのリンク記述などはひとつの例であるが、どのような仕組みを整えることがより大きなデータ交流・知識交流に結びついていくかということについては、今後さらに検討を深めたい。

(2) 今回の本システムの機能強化は、もともと IIIF の採用により画像資料の共有を有効に行うことを企図したものであったが、結果的には IIIF の利用は内部的なシステム構築作業においても非常に有効であることがあらためて確認された。例えば特定の座標でトリミングした画像を取得する場合、Zoomify 独自のタイル画像を使用していた時は、座標を計算して画像を切り抜く処理を自前でプログラムしていたが、この処理を IIIF 画像 API に任せることで、大幅に処理の簡略化をはかることができた。IIIF は画像の共有という文脈で語られることが比較的多い印象であるが、必ずしも画像の共有を前提としなくとも、画像処理に IIIF を採用することには十分な意義があると認められる。

5. 主な発表論文等

〔雑誌論文〕(計 11 件)

- 和氣愛仁, 永井正勝, 高橋洋成. 「アノテーション付与型画像データベースプラットフォームの IIIF 対応」『研究報告人文科学とコンピュータ』2019-CH-119(15), pp.1-6, 2019, 査読なし
- 永井正勝, 和氣愛仁, 高橋洋成. 「文字資料を対象とするデータベース構築に適した言語学的記述のあり方について」『研究報告人文科学とコンピュータ』2019-CH-119(14), pp.1-7, 2019, 査読なし
- Norihiko Uda, Chieko Mizoue, Saori Donkai, Saki Ishimura. "Information seeking behavior of older adults in a public library in Japan," Library and Information Science Research e-journal(LIBRES), vol.28-1, pp.1-12, 2018, 査読あり
- 中村覚, 成田健太郎, 永井正勝. 「LinkedData 化した典拠データと IIIF を用いた法帖の異版比較支援システムの開発」『人文科学とコンピュータシンポジウム論文集』2018, pp.297-302, 2018, 査読あり
- 上原究一, 永井正勝, 中村覚, 中尾道子, 近藤隼人, 荒木達雄, 蓑輪顕量. 「図書館における木版本のデジタル化と利活用の可能性-IIIF と TEI を用いた『水滸伝』諸版本のデジタル化を通じて」『人文科学とコンピュータシンポジウム論文集』2018, pp.381-388, 2018, 査読あり
- 渡部航太郎, 松村敦, 宇陀則彦. 「論文標題図書館情報学アーカイブズの利用を目的とした DBpedia の活用」『研究報告人文科学とコンピュータ』, 2018-CH-116(4), pp.1-4, 2018, 査読なし
- 野田香蓮, 松村敦, 宇陀則彦. 「絵図の理解を目指した読図プロセスの可視化」『研究報告人文科学とコンピュータ』, 2018-CH-116(14), pp.1-7, 2018, 査読なし
- 高橋洋成. 「旧約聖書における親族・部族用語の意味論：歴代誌史書と民数記との比較から」『New 聖書翻訳』No.3, pp.31-44, 2017, 査読なし
- 永井正勝. 「「パピルス・アボット」における完了 sDm=f 形の接尾表記について-ボトムアップ型の言語記述を目指して」『実験音声学・言語学研究』pp.75-90, 2017, 査読あり
- Norihiko Uda. "Development of a new information literacy education: A bibliographic search game," Proceedings of the 7th International Conference on Asia-Pacific Library and Information Education and Practice, A-LIEP 2016, pp.434-446, 2016, 査読あり
- 永井正勝. 「エジプト語における文連鎖：エジプト語を支える「らしさ」の継承」『古代オリエント研究の地平：小川英雄先生傘寿記念献呈論文集』pp.127-144, 2016, 査読なし

〔学会発表〕(計 9 件)

- 和氣愛仁. 「言語資料を対象とした画像 = テキスト連携データベースプラットフォームの開発と今後の展望」『東京外国語大学アジア・アフリカ言語文化研究所情報資源利用研究センター (IRC)DH ワークショップ「30 年後も使えるデータ」を目指す』, 2018 (招待講演)
- 和氣愛仁, 永井正勝, 高橋洋成. 「アノテーション付与型画像データベースプラットフォームの IIIF 対応」情報処理学会第 119 回人文科学とコンピュータ研究会発表会, 2018
- 永井正勝, 和氣愛仁, 高橋洋成. 「文字資料を対象とするデータベース構築に適した言語学的記述のあり方について」情報処理学会第 119 回人文科学とコンピュータ研究会発表会, 2018
- Yona Takahashi, Masakatsu Nagai, Toshihito Waki. "Application of the Concept of "Linkbase" for Digitalization of Linguistic Resources and Analysis," Japan Association for Digital Humanities Conference 2017, 2017 (国際学会)
- 高橋洋成, 永井正勝. 「古代エジプト語とアッカド語における Tense-Aspect-Mood の対照

研究：対訳資料に出現する動詞形の記述」日本オリエント学会第 59 回大会，2017
永井正勝，高橋洋成。「古代エジプト語聖刻文字資料とアッカド語楔形文字資料の対訳データベースの構想：「対応関係の可視化」のための研究プラットフォームの形成」，日本オリエント学会，2016 年 11 月 13 日，慶應義塾大学三田キャンパス（東京都港区）
高橋洋成。「楔形文字粘土板における「空白」のマークアップ」，デジタル・ヒューマニティーズ講演会&ワークショップ，2016 年 12 月 16 日，筑波大学筑波キャンパス（茨城県つくば市）
宇陀則彦。「図書館情報学アーカイブズの概要」，デジタル・ヒューマニティーズ講演会&ワークショップ，2016 年 12 月 16 日，筑波大学筑波キャンパス（茨城県つくば市）
和氣愛仁。「アノテーション付与型画像データベースプラットフォーム WDB」，デジタル・ヒューマニティーズ講演会&ワークショップ，2016 年 12 月 16 日，筑波大学筑波キャンパス（茨城県つくば市）

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

[1] WDB: General-purpose Platform for Database System
<https://wdb.jinsha.tsukuba.ac.jp/>

[2] Hieratic Database Project
<https://wdb.jinsha.tsukuba.ac.jp/hdb/>

[3] 近代日本語文典集成
<https://wdb.jinsha.tsukuba.ac.jp/jgt/>

[4] Sumero-Akkadian Cuneiform Tablets
<https://wdb.jinsha.tsukuba.ac.jp/xsux/>

6 . 研究組織

(1)研究分担者

研究分担者氏名：宇陀 則彦

ローマ字氏名：(UDA, Norihiko)

所属研究機関名：筑波大学

部局名：図書館情報メディア系

職名：准教授

研究者番号(8桁)：5 0 2 6 1 8 1 3

研究分担者氏名：永井 正勝

ローマ字氏名：(NAGAI, Masakatsu)

所属研究機関名：東京大学

部局名：附属図書館

職名：特任准教授

研究者番号(8桁)：7 0 5 7 8 3 6 9

研究分担者氏名：高橋 洋成

ローマ字氏名：(TAKAHASHI, Yona)

所属研究機関名：筑波大学

部局名：人文社会系

職名：非常勤研究員

研究者番号(8桁)：9 0 6 4 7 7 0 2

[削除：2018年3月23日]

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。