

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 2 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330395

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システムのための汎用プラットフォーム構築

研究課題名(英文) Construction of General-purpose Platform for Database System of 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) 3,700,000円

研究成果の概要(和文)：本研究では、資料画像にテキスト情報を付与し、画像から文字あるいは単語等のテキストデータを検索できるようなシステム(「アノテーション付与型画像データベースシステム」)について、さまざまな人文科学資料を共通の枠組で扱うことを可能にするために、共通のユーザインターフェイスと標準的なデータ構造モデルを作成し、汎用データベースプラットフォームシステムを構築した。本研究ではこのプラットフォームを利用し、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した。

研究成果の概要(英文)：With the present research we have created a standardized model of user interface and data structure for annotated images of various materials in humanities; ancient Egyptian hieratic script, modern Japanese grammar textbooks, Sumerian-Akkadian cuneiform, etc. The model will be a general-purpose platform which will ensure easier accessibility of materials of broader range and scope.

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

1. 研究開始当初の背景

人文学資料のデジタル化とメタデータ構築は本研究以前にも多く行われてきており、デジタルアーカイブやデジタルミュージアの形でその成果が公表されてきていた。しかし、そこで利用されている閲覧・検索システムは、単純に画像を拡大縮小できるだけの簡単なユーティリティ類を除けば、対象となる資料専用のものが用いられていることが多く、それらのシステムは相互に互換性がないことが多かった。そのため、ユーザインターフェイス・データ構造双方の再利用可能性という点で、システムとしての汎用性は乏しいと言わざるを得なかった。このような状況を改善し、さまざまな人文学資料研究に情報学的方法を用いる際の手法的な枠組みを統一することを企図して本研究を開始した。

2. 研究の目的

本研究では、資料画像にテキスト情報（アノテーション）を付与し、画像から文字あるいは単語等のテキストデータを検索できるようなデータベースシステム（これを「アノテーション付与型画像データベースシステム」と呼ぶ）について、共通のユーザインターフェイスと標準的なデータ構造モデルを作成し、より広範な資料をより少ない手間で共通的に扱えるような汎用データベースプラットフォームシステムを構築することを目的とした。このような共通のプラットフォームを利用することにより、考古学・歴史学・言語学等の研究分野を超えて研究資源の共有と研究用ツールの共通化が図られれば、暗黙知に基づいた「閉じた人文学」から、互いの研究手法を意識した「つながる人文学」への転換が図られ、その結果、人文学全体に対して極めて大きな研究上のインパクトを与えることができる。

3. 研究の方法

(1) まず始めに、研究代表者が本研究開始以前から構築を開始していた古代エジプト神官文字データベースシステムをもとにして、明治期日本語文典データベースの構築を行った。その際に、様々な時代の様々な言語を統一的にデータベース上で扱うことが可能になるような汎用的なデータ構造を設計し、そのデータ構造に基づいて、古代エジプト語神官文字パピルス資料、近代日本語文典資料、さらには古代シュメール・アッカド語楔形文字粘土板資料のデータベースを作成した。本研究のメンバーには言語学を専門とする研究者が多いため、言語学的な検討にも十分耐えうるようなデータ構造を構築することを意図した。とりわけ、古代エジプト語のような一度死滅した言語をも扱うために、固定的な文字・単語等のデータだけでなく、研究者ごとの解釈の違いも含めてデータとして保存できるようなデータ構造を定義した。

(2) ユーザインターフェイスについては、Drupal（代表的な CMS のひとつ）等のオープンソースソフトウェアや、JavaScript、HTML5 等の標準的なウェブ技術を用いて、ユーザビリティに優れた共通インターフェイスを構築することを目指した。また、様々なユーザ環境（パソコン、タブレット、スマートフォン等）に対応可能なデザインを目標にした。

(3) 最終的に、そのような汎用的なデータ構造およびユーザインターフェイスをもつシステムを、画像データベースシステムの汎用プラットフォームとしてパッケージ化することを企図した。

4. 研究成果

(1) 研究開始当初は、目的とするデータ構造として、(a)文字研究を目的としたデータ構造、(b)文法研究を目的としたデータ構造、(c)絵画・図版等の不定形資料の分析を目的としたデータ構造のようなものをそれぞれ別々に設計することを目指していたが、最終的には、文字・語・文・文章・挿絵等の図版を共通して扱うことのできる、ひとつの大きなデータ構造を設計できた。可能な限り標準的で汎用的なデータ構造の設計を目指した本研究としては、これはむしろ望ましい結果だと考える。実際に扱う言語についても、当初の計画にあった古代エジプト語と近代日本語に加えて、古代シュメール・アッカド語も同じシステム上で扱うことが可能になった。

(2) 多様な資料を同じ操作で閲覧できるような、ユーザビリティに優れた共通インターフェイスを構築するという目的についても、CMS をベースとして利用したパッケージングを行うことで、かなりの程度達成できたものとする。ただし、現時点では簡便な方法でインストールが可能なかたちにはなっていないので、より広く一般に利用してもらうことを目指すためには、さらに簡便な方法でデータ投入とシステムの設置ができるようにする必要がある。

(3) また、研究が進展するにつれて、既存の外部システムとのデータ連携をより積極的に進めることが非常に重要になってきたので、当初計画していた TEI 形式 XML ファイルの出力機能のほかに、RDF 形式 XML ファイルの出力機能や、全文テキスト検索・出力機能の試験的な実装も行った。これについては、Web API の実装あるいは各種検索エンジンとの連携など、外部システムとのより高度なデータ連携を検討するため、平成 28 年度の科研費基盤研究(C)「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」(課題番号 16K00456)を取得して、さらに研究を深める予定である。

5. 主な発表論文等

〔雑誌論文〕(計 14 件)

[1] Yona Takahashi “A Small Dictionary of Hamar: with Some Lexical Notes on Banna and Karo” *Studies in Ethiopian Languages*, Vol.4, pp.35-75, Japan Association of Ethiopian Languages, 2015, 査読あり

[2] 田中僚・松村敦・宇陀則彦「原資料の構造を反映したデジタルアーカイブの構築」『情報処理学会研究報告(人文科学とコンピュータ研究会)』2015(9)、pp.1-2、2015、査読なし

[3] 矢澤真人「明治以降日本語の文法用語の成立と消長」『19世紀末の日韓中の言語接触と交流・伝播 近代の語彙と文法用語の再検討』国際シンポジウム資料集、2015-08、pp.91-114、2015、査読なし

[4] 矢澤真人「日本語文法教育史における品詞分類表の位置づけ」『2015 北京外国語大学中日韓語言文化比較研究国際学会議会議手冊』、pp.67-68、2015、査読なし

[5] 矢澤真人「明治以降日本語の文法用語の成立と品詞分類表」, *Language Facts and Perspectives*, 36, pp.25-48, 2015, 査読なし

[6] 高橋洋成・永井正勝・和氣愛仁「画像, TEI, LOD を用いた文字研究・言語研究のためのプラットフォームの構築」人文科学とコンピュータ研究会報告、2014-CH-105(5)、pp.1-5、2014、査読なし

[7] 永井正勝「トリノ・エジプト博物館所蔵 Papyrus Turin Cat. 1885 の神官文字に関する覚え書き」一般言語学論叢 17、pp.153-169、2014、査読あり

[8] 矢澤真人「抽象名詞主題文研究の意義」漢日語言対比研究論叢第5輯、pp.1-14、2014、査読あり

[9] 田中僚・芦川大樹・松村敦・宇陀則彦「原資料の分析に基づく図書館情報学アーカイブズの構築」人文科学とコンピュータ研究会報告、2014-CH-103(3)、pp.1-4、2014、査読なし

[10] 和氣愛仁、永井正勝「RDB と CMS を用いたアノテーション付与型画像データベースシステムの構築 データ構造とインターフェイスの標準化を目指して」情報処理学会研究報告(人文科学とコンピュータ)、2013-CH-99(7)、pp.1-8、2013、査読なし

[11] 永井正勝「モスクワ・パピルス No.120」

における接尾代名詞=w の表記とその環境」一般言語学論叢 15、pp.41-71、2013、査読あり

[12] 常川真央、松村敦、宇陀則彦「日本十進分類法を用いた類似読者発見手法」情報メディア研究 12、pp.42-51、2013、査読あり

[13] 高橋洋成「言語の多面性を織り込んだ言語資料のデジタルネットワーク」情報処理学会『人文科学とコンピュータシンポジウム論文集 2013』pp.39-44、2013、査読あり

[14] 高橋洋成「アマルナ文書の電子化: 文字研究・言語研究を目指して」情報処理学会研究報告(人文科学とコンピュータ) 2013-CH-99(6)、pp.1-7、2013、査読なし

〔学会発表〕(計 14 件)

[1] 永井正勝「中エジプト語における文法研究の試み 動詞述語文を事例として」, 日本言語学会第 151 回大会、「ワークショップ 古代文字文献を資料とした死言語の文法研究 中エジプト語・契丹語・シュメール語・西夏語の事例から」, 2015 年 11 月 29 日、名古屋大学(愛知県名古屋市)

[2] 矢澤真人「日本語文法教育史における品詞分類表の位置づけ」2015 北京外国語大学中日韓語言文化比較研究国際学会議、2015 年 10 月 17 日~2015 年 10 月 18 日、北京外国語大学(中国・北京)

[3] 永井正勝・和氣愛仁・高橋洋成「古代エジプト語神官文字資料のコーパス作成とその利用: “外字” で書かれた文字言語のドキュメンテーションの試み」日本言語学会第 150 回大会、2015 年 6 月 21 日、大東文化大学(東京都板橋区)

[4] Toshihito Waki, Masakatsu Nagai, Yona Takahashi, “On the Development of a Web-based Platform for Egyptian Hieratic Papyri, Akkadian Cuneiform Tablets, and Japanese Schoolbooks: Toward a New Grammatological and Linguistic Studies”, *Digital Humanities Worldwide*, 「スーパーグローバル大学創成支援」採択事業キックオフ・ワークショップ、2015 年 02 月 15 日、筑波大学(茨城県つくば市)

[5] 高橋洋成・永井正勝・和氣愛仁「画像, TEI, LOD を用いた文字研究・言語研究のためのプラットフォームの構築」第 105 回人文科学とコンピュータ研究会、2015 年 01 月 31 日、大阪国際大学守口キャンパス(大阪府守口市)

[6] 和氣愛仁「明治期日本語文典の画像データベースの構築と他言語文化資料への展開」第 27 回日本語教育連絡会議、2014 年 08 月 24 日(ハンガリー・バラトンサールスー)

[7] 高橋洋成「言語の多面性を織り込んだ言語資料のデジタルネットワーク」人文科学とコンピュータシンポジウム「じんもんこん2013」, 2013年12月12日、京都大学(京都府京都市)

[8] 和氣愛仁、永井正勝、高橋洋成「古代エジプト語の高精細画像を用いた言語データベースシステム」ワークショップ「デジタル・ヒューマニティーズと人文科学の再接続」, 2013年11月09日、筑波大学(茨城県つくば市)

[9] 矢澤真人「文法研究と国語教育の文法、日本語教育の文法」2013年度日本語教育及び日本文化研究国際シンポジウム(招待講演), 2013年11月09日、上海外国語大学(中国・上海)

[10] 永井正勝「古代エジプト神官文字文書のアノテーション付与型データベース Hieratic Database Project (HDB) の取り組み」日本オリエント学会第55回大会, 2013年10月27日、京都外国語大学(京都府京都市)

[11] 矢澤真人「日本語の抽象名詞と文型」漢日対比言語学研究会 2013年度大会, 2013年08月21日、福建師範大学(中国・福州市)

[12] 和氣愛仁、永井正勝「RDBとCMSを用いたアノテーション付与型画像データベースシステムの構築 データ構造とインターフェイスの標準化を目指して」第99回情報処理学会人文科学とコンピュータ研究会, 2013年08月03日、筑波大学(東京都文京区)

[13] 和氣愛仁、宇陀則彦、永崎研宣、松村敦「閉じる研究と開く研究の接点を目指して～筑波人文情報学研究会の挑戦～」第99回情報処理学会人文科学とコンピュータ研究会, 2013年08月03日、筑波大学(東京都文京区)

[14] 高橋洋成「アマルナ文書の電子化：文字研究・言語研究を目指して」第99回情報処理学会人文科学とコンピュータ研究会, 2013年08月03日、筑波大学(東京都文京区)

〔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

(2) 研究分担者

矢澤 真人 (YAZAWA, Makoto)

筑波大学・人文社会系・教授

研究者番号：30182314

宇陀 則彦 (UDA, Norihiko)

筑波大学・図書館情報メディア系・准教授

研究者番号：50261813

永井 正勝 (NAGAI, Masakatsu)

筑波大学・人文社会系・助教

研究者番号：70578369

高橋 洋成 (TAKAHASHI, Yona)

筑波大学・人文社会系・非常勤研究員

研究者番号：90647702