

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月 9日現在

機関番号：13201

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500085

研究課題名（和文） 国際標準 Ajax 版日本現存朝鮮古書原文画像データベース・システムの研究

研究課題名（英文） Research on original text images database of old Korean books in Japan by international standard Ajax techniques

研究代表者

高井 正三（TAKAI SHOSO）

富山大学・総合情報基盤センター・教授

研究者番号：80206705

研究成果の概要（和文）：

国際標準の Ajax 技法を使用した日本現存朝鮮古書原文画像データベース・システムは、Microsoft 2008 Server R2 OS のサーバー上に、Apache HTTP Server 2.2, Apache Tomcat 7.0, Postgres SQL 9.0, Enterprise Apache PHP をインストールし、JSP, Java Framework JSF, JavaScript, HTML を使用して、画像の閲覧システムを構築し、画面遷移を伴わず、画像の拡大・縮小表示操作、画像の上下左右移動をスムーズに実行できるようになった。しかしながら、原文画像の著作権取得の遅れ、大学情報基盤センターの大規模な情報システムの更新に伴う、10年以上使用してきたデータ処理用 Unix OS サーバーとソート/マージ・プログラムがシステムの全面的に入れ替えにより、使用できなくなるという環境の変化などの要因が重なり、原文画像のメタデータ・データベースとの整合および、現行の日本現存朝鮮古書データベース・システム DOKB との結合を実現するには至らなかった。

研究成果の概要（英文）：

The original textual images database system of old Korean books in Japan with international standard Ajax techniques was implemented on MS Windows 2008 Server R2. We installed Open Source software named Apache HTTP Server 2.2, Apache Tomcat 7.0, Postgres SQL 9.0 and Enterprise Apache PHP on this server, and we developed the information system in order to display original textual images with JSP, Java framework JSF, JavaScript and HTML, and with Ajax techniques. That system can display expansion and reduction image of original textual image size, and can move image up and down, left and right by operation of computer mouse without screen transition. But we could not complete to combine these image database with the metadata database of these original textual images, and with the current database DOKB that stands for Database of Old Korean Books in Japan to create a new Database system because of delay of the copyright of original textual images acquisition and because of large scale changes of data processing environment of our university. Cause of biggest things is no longer available the data processing systems and sort/merge program that have been used more than ten years by large scale information system renewals in Information Technology Center.

交付決定額

(金額単位：円)

|        | 直接経費      | 間接経費    | 合計        |
|--------|-----------|---------|-----------|
| 2010年度 | 1,600,000 | 480,000 | 2,080,000 |
| 2011年度 | 800,000   | 240,000 | 1,040,000 |
| 2012年度 | 800,000   | 240,000 | 1,040,000 |
| 総計     | 3,200,000 | 960,000 | 4,160,000 |

研究分野：複合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：データベース，朝鮮古書書誌 DB，古書原文画像 DB，Ajax 技法，Unicode，Java システム

### 1. 研究開始当初の背景

元富山大学人文学部教授藤本幸夫氏が 36 年以上にわたって、日本現存朝鮮古書を实地調査・収集した書誌情報は約 16,000 件で、現在までその書誌情報データを電子化し、順次データベース化を進めているが、この調査の傍ら撮影した、原文画像 26,543 枚のフィルムが未整理の状態であった。本研究は、この原文フィルム画像を高解像度でデジタル化し、原著文献の内容のみならず、刻字名の研究、古書の角質や蔵書印の研究が可能となるように、高解像度画像データを提供し、Google Maps のように、Web ブラウザで画面遷移なしに前後左右の移動、縮小拡大、回転表示など、軽快な閲覧ができるよう、Ajax 版の画像 DB アクセス・ツールを研究開発・整備しようと考えたことに始まる。

### 2. 研究の目的

本研究の目的は、元富山大学人文学部教授藤本幸夫氏が 36 年以上にわたって实地調査・収集した時撮影した、日本現存朝鮮古書の原文画像 26,543 枚の総てをデータベース (DB) 化し、国際標準の Ajax 版の画像閲覧 DB システムとして完成し、活用できるようにすることである。そして、この DB の名称を DOKB (Data Base of Old Korean Books existed in Japan) として、書誌情報とともに世界に公表し、全世界の朝鮮及び朝鮮本の研究者に提供することである。

### 3. 研究の方法

(1) 既にデジタル化を終了した原文画像約 6,500 枚の画像の切り出し、形式変換編集

これまでに撮影されたモノクローム 35mm フィルムを、4,000dpi (dots per inch) でデジタル化し、tiff 画像として、ナンバリングしながら保存する。次に、tiff 画像から原文を切り出し画像形式を png に変換する。更に、各 png 画像を Ajax 検索に使用できるよう 256dots×256dots の画像に分解し、ファイル名をフォルダー名にして保存する。

(2) 原文画像のメタデータの作成と現行 DOKB との結合調査

原文画像のメタデータを作成し、現行の日本現存朝鮮古書書誌情報データベースとの整合を調査する。

(3) Ajax (Asynchronous JavaScript + XML) 技法の継続調査研究

Ajax 技法を調査研究し、画面遷移のない画像表示と拡大・縮小表示、上下左右への移動

表示が可能な技法とプログラミング方法を習得する。

(4) 原文画像検索・Ajax 表示ツールの改良，データベース・システムへの組み込み・結合テスト

Ajax 版日本現存朝鮮古書原文画像 DB システムを試作、表示ツールのレビュー後、改良を重ね、既存の DOKB データベース・システムへの組み込み・結合テストを実施する。

(5) 日本現存朝鮮古書データベース DOKB への画像データ追加による DB の再編成

日本現存朝鮮古書データベース DOKB への画像データを追加するため、DOKB データベース・システムを再編成し、再構築を行う。

(6) 検索語推測システム用検索語の抽出と XML データベース化

検索語推測システムは、既存の DOKB データベース・システムから、証明や選者名などの項目を限定して、既存のデータから検索用キーワードを抽出して、検索語推測用に XML データベースを作成する。

(7) 画像データベースと検索ツール、検索語類推ツールの組み込み、統合テスト

1) 画像データベース・システムを完成する。

2) 画像検索システムと Ajax 表示ツールを完成する。

3) Ajax による検索語類推ツールを完成する。

4) 上記 3 つのシステム・ツールを統合し、テストを実施する。

(8) 原文画像を含む DOKB データベース・システムのインターネット上における公開テスト

1) インターネット上に画像データベース・システムを移植し、テストする。

2) ユーザ ID，パスワードによる画像データベース・システムを公開テストにかける。

3) 公開テストで見つかった検索・表示ツールの不具合を修正し、システムに反映する。

(9) 国内及び海外の研究者・利用者による画像データベース検索システムのレビュー

1) 国内の研究者に本 DOKB 画像データベース・システムを発表し、レビューを受ける。

2) 海外の研究者に検索語推測システムと合わせて発表し、レビューを受ける。

3) レビューを受けた結果、修正すべきところを修正し、システムに反映する。

(10) 学会，研究会にて研究成果を発表，正式公開，報告書のホームページ作成

1) ID・パスワードが必要な研究者には通知し、原文画像 DOKB データベースを公開する。

2) 研究・開発の成果を情報処理学会データベース研究会, 漢字情報処理研究会, 学術情報処理研究集会などの研究会や学会, 国際会議の“国際化 Unicode 会議”で発表する。

3) 研究成果をホームページで公開するとともに, 研究成果公開促進経費を取得して分担者の残りの課題である

(11)「日本現存朝鮮本研究(経部・史部・子部)」本(冊子体)の発行を支援する。

#### 4. 研究成果

(1) 残る原文フィルム画像約 6,500 枚のデジタル化と画像の切り出し, 形式変換

1) 既設の Nikon SUPER COOLSCAN 5000 ED スキャナを使用し, フィルム画像を 4,000dpi の解像度で読み込み, TIFF(Tagged Image File Format)画像として保存する作業は終了した。

この TIFF 画像から必要な画像を抽出については, スケール付き画像が多く, フィルム・サイズの 35mm で統一した方が, 256×256dots 分解後の分割数が統一できるので, 原文部分の画像抽出を必要最小限にして, 開発した 256×256dots 分解ソフトウェア「にごろ.exe」を用いてバッチ処理し, 256×256dots の Ajax 用 png 画像に分解する作業を実施した。



図 1-1 オリジナル画像

(ファイル: 70009\_00.tif)

2) Ajax による表示ツールで配信するために, 256×256dots の画像に分解して, 位置や解像度の情報を組み合わせ, PNG 形式で保存する作業は, Hot Soup Processor で開発した

「にごろ.exe」と命名したバッチ処理プログラムにより終了した。



図 1-2 「にごろ.exe」による画像分割処理

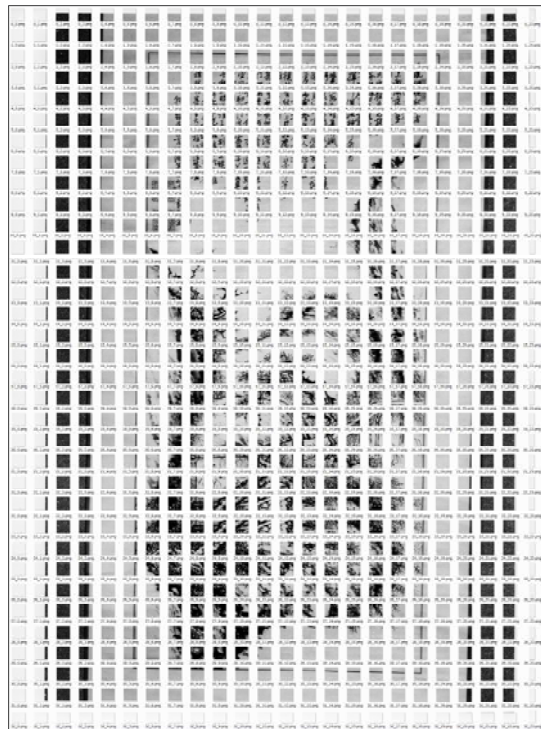


図 1-3 画像を 256×256dots の png 形式画像に分解したもの(フォルダー名: 70009\_00)

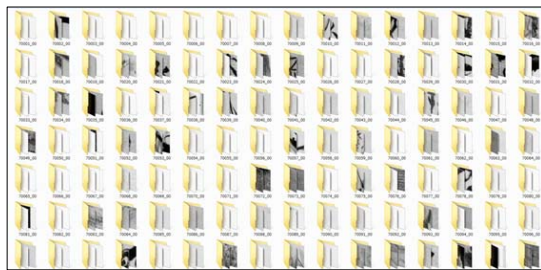


図 1-4 256×256dots に分解された png 形式画像のフォルダー群

(画像ファイル名→フォルダー名)

(2) 原文画像のメタデータの作成と現行 DOKB との結合調査



原文画像のメタデータを再確認し、既存 DOKB データとの整合を調査する作業は、原文画像データベースが完成した後に、作業を開始する予定で、進んでいないのが現状である。従って、既存 DOKB データとの結合が可能な場合は、メタデータに DOKB 参照番号を付加する作業、集部以外の原文画像については独自の参照番号を付加する作業などが、依然として残っている。

### (3) Ajax(Asynchronous JavaScript and XML)技法の継続調査研究

Ajax, XML, JavaScript に関する研究を続け、ツールの改良とテストを実施し、技法を使いこなす作業、Ajax 技法を使用して、画像表示ツールを改良し、テストを繰り返す作業、各国 IME と検索語類推ツールを改良する作業などは、順次進行中である。

(4) 古書原文画像検索, Ajax 表示ツールを改良, DB システムへの組込み・結合テスト  
1) Ajax 技法を使用した, 画面遷移のない日本現存朝鮮古書原文画像検索システムは, Windows 2008 R2 Sever 上に, HTTP Server 2.2, Apache Tomcat 7.0, Postgres SQL 9.0, Enterprise Apache PHP をインストールし, HTML 版の画像閲覧システムを構築した。



図 4-1 画像閲覧システムの初期画面

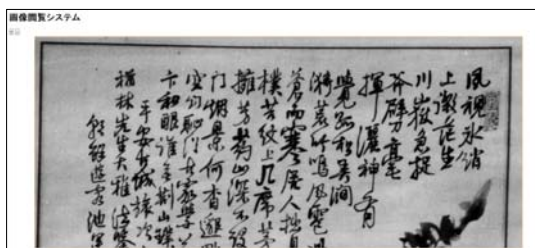


図 4-2 画像閲覧システムによる原文画像の拡大画面

(マウスのホイール操作で拡大縮小)

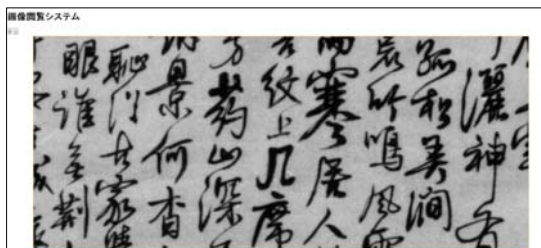


図 4-3 画像閲覧システムによる原文画像の

最大拡大画面(「寒」を中心に上下左右動作を実行)

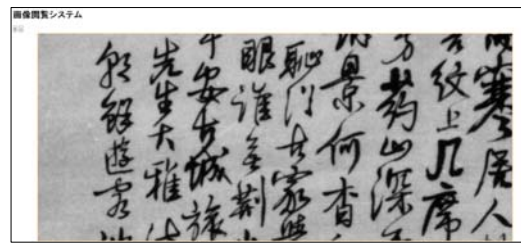


図 4-4 画像閲覧システムによる「寒」を中心に右へドラッグ移動した画面

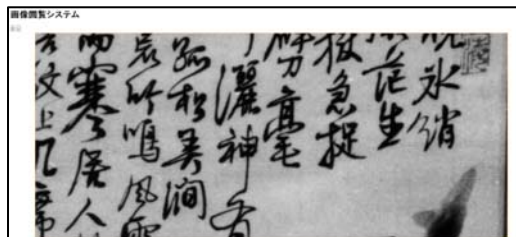


図 4-5 画像閲覧システムによる「寒」を中心に左へドラッグ移動した画面

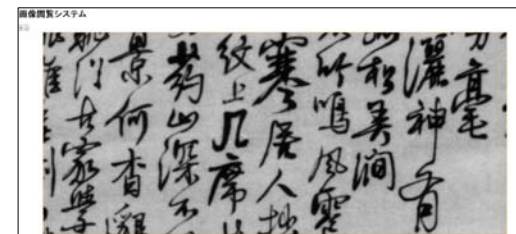


図 4-6 画像閲覧システムによる「寒」を中心に上へドラッグ移動した画面

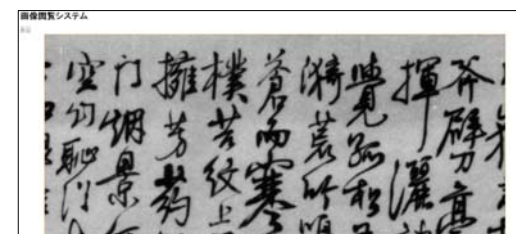


図 4-7 画像閲覧システムによる「寒」を中心に下へドラッグ移動した画面

2) 現状では、画像のメタデータ DB から原文画像を検索し、画像ファイルを表示して、Ajax 機能による表示ツールを使って、画面の遷移なしに、画像の拡大・縮小をマウスのホイール操作で、画像の上下左右の移動をマウスのドラッグ操作で行える画像閲覧システムを開発して、テスト段階に来ており、DOKB システムへの組み込み・結合テストを行う作業は未着工である。

(5) 日本現存朝鮮古書データベース DOKB への画像データ追加による DB の再編成  
既存の DOKB を再編成し、画像データ、メタデータを追加するためのテーブルを登

録し、データベースの入力準備を行う作業は、設計段階止まりである。

#### (6) 検索語推測システム用検索語の抽出とXMLデータベース化

DOKBのマスター・データから検索語を切り出しは、一部はできているが、これまで使用してきたAIXサーバーのソート・プログラムが、総合情報基盤センターのシステム更新に伴って撤去され、XMLデータベース化作業は中断したままである。

#### (7) 画像データベースと検索ツール、検索語類推ツールの組み込み、統合テスト

1) 画像データベースを作成するためのデータの整備はできたが、256×256dotsの画像に分解した画像フォルダーを蓄積し、呼び出すデータベース・システムは、開発途上で完成に至っていない。

2) Ajax表示ツールを使用した画像を閲覧するためのシステムは完成したが、メタデータから検索して画像を呼び出すシステムは構築中である。

3) Ajax版の試作品はできているが、検索語のソートを実施し、XMLデータベース化を待って、完成を目指している。

4) 3つのシステム・ツールを統合とテストは実現に至らなかった。

#### (8) 原文画像を含むDOKBデータベース・システムのインターネット上における公開テストについて

1) インターネット上に画像データベース・システム移植し、テスト。2) ユーザID、パスワードによる画像データベース・システムを公開テスト。3) 公開テストで見つかった検索・表示ツールの不具合を修正し、システムに反映。の3つは、共同研究者の元教授藤本幸夫氏から、著作権処理をしていないので、当分の間はインターネット上へは公開しないでほしい旨の要請を受け、公開待ちである。

#### (9) 国内及び海外の研究者・利用者による画像データベース検索システムのレビュー

毎年開催される国内の漢字文献情報処理研究会総会や国際化Unicode会議に出席し、研究者・利用者による画像データベース検索システムについて、意見を交換し、討論を重ねてきているが、朝鮮古書関係のデータには多少の興味のある人はいるが、画像の研究者は殆ど皆無の状況であった。

#### (10) 学会、研究会にて研究成果を発表、正式公開、報告書のホームページ作成

1) ID・パスワードが必要な研究者に通知し、原文画像DOKBデータベースを公開と、2) 研究・開発の成果を情報処理学会データベ

ス研究会、漢字文献情報処理研究会、学術情報処理研究集会などの研究会や学会、国際会議の“国際化Unicode会議”での討論は順次実施してきたが、公開は準備段階である。

3) 研究成果のホームページでの公開と、研究成果公開促進経費の取得は未だ実現せず、分担者の残りの課題である。

(11) 「日本現存朝鮮本研究(経部・史部・子部)」本(冊子体)の発行を支援したい。史部については、2度冊子体の仮製本を出版したが、研究成果公開促進経費の科研費は、2回とも付かなかった。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

AjaxとJavaフレームワークJSFによる古書書誌データベース・システムの構築, 高井正三, 第49回IBMユーザー・シンポジウム論文, pp.1-25, 2011. (奨励賞受賞)

[学会発表](計0件)

なし

[図書](計3件)

・高井正三, 上木佐季子, 新里泰孝(共著), "2011年版大学生の情報リテラシー 大学生のICT活用標準テキスト(第5版)", 富山大学出版会, ISBN978-4-340-53017-5, 2011.4.1.

・高井正三, 上木佐季子, 新里泰孝(共著), "2012年版大学生の情報リテラシー 大学生のICT活用標準テキスト(第6版)", 富山大学出版会, ISBN978-4-340-53018-2, 2012.4.1.

・高井正三, 上木佐季子, 新里泰孝(共著), "2013年版大学生の情報リテラシー 大学生のICT活用標準テキスト(第7版)", 富山大学出版会, ISBN978-4-340-53021-2, 2013.4.1.

[その他]

ホームページ等

<http://stl30.itc.u-toyama.ac.jp/dokb/>

<http://stl50.itc.u-toyama.ac.jp/dokb/> (予定)

#### 6. 研究組織

##### (1) 研究代表者

高井 正三 (TAKAI SHOSO)

富山大学・総合情報基盤センター・教授  
研究者番号: 80206705

##### (2) 研究分担者

藤本 幸夫 (FUJIMOTO YUKIO)

麗澤大学・言語研究センター・客員教授  
研究者番号: 70093458