

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 30日現在

機関番号：14301

研究種目：基盤研究(B)

研究期間：2010～2012

課題番号：22300083

研究課題名（和文）人文学研究を促進する協働のための情報共有基盤に関する研究

研究課題名（英文）Information Platform for Collaborative Humanity Research

研究代表者

林晋（SUSUMU HAYASHI）

京大学大学院文学研究科・教授

研究者番号：40156443

研究成果の概要（和文）：人文学、特に史料ベースの人文学のための、ネット上分散協働作業用プラットフォーム HCP サーバを構築し、歴史研究用ツール SMART-GS を、そのクライアントに改造した。また、同クライアントによるオンサイト・リアルタイム協働の概念を導入し、そのための種々の機能を開発した。

研究成果の概要（英文）：HCP, a server-client model for humanity research collaboration, and its server was built. SMART-GS, a tool for history study developed by our team, was remodeled to HCP-client. A concept onsite real-time collaboration for humanity research was introduced and many features for it were developed for SMART-GS.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	6100000	1830000	7930000
2011年度	4700000	1410000	6110000
2012年度	2900000	870000	3770000
年度			
年度			
総計	13700000	4110000	17810000

研究分野：総合領域

科研費の分科・細目：情報学・「図書館情報学・人文社会情報学」

キーワード：情報メディア、歴史情報学

1. 研究開始当初の背景

研究代表者林は、もともと情報工学者でありソフトウェア工学などを研究していたが、現在は数学基礎論史・情報史・思想史などを専門とする。林はドイツ人歴史学者にとっても分析が困難なドイツ語史料を分析するために、手書き文献資料研究用のツール SMART-GS を開発し、これを利用して歴史学・文献学上の成果を得ていた。これが他の歴史学者の注目するところとなり、日本近代史の重要史料ながら、難読性の故に文献学的研究は困難とされてきた倉富勇三郎日記（以下、倉富日記）の全翻刻プロジェクトにも

SMART-GS が採用された（「翻刻」は手書き文書等をテキスト化することを意味する人文系用語）。

しかし、SMART-GS は個人研究用に開発されたため、倉富日記全翻刻プロジェクトのような遠隔地に住む複数の歴史学者が協働して行う研究では様々な問題がおきていた。たとえば、複倉富プロジェクトでは、翻刻結果の共有を Web ファイル共有サービスや WebDAV などの既存技術を使い行っていたが、バージョン管理や編集衝突などの問題へのサポートが十分でなかった。

2. 研究の目的

以上の様な背景のもと、当初の研究の目的は、倉富プロジェクトの場合の様な遠隔地における多人数の歴史研究者による共同作業のサポートの問題を解決することであり、それを実現するために、HCP Humanities CyperPlatform というサーバ・クライアント・モデルによる情報システムを構築することであった。

HCP モデルは、基本的には、Apache SVN のようなバージョン管理・ファイル共有システムとして構想された。すでに、そういう技術がありながら、それとは別なモデル/システムを構築する理由は、SVN などの既存のバージョン管理システムが想定する、ソフトウェア・ソースや、単純なファイルシステムと、HCP が実現すべき、代表者が提唱する「画像化主義の人文情報学」におけるファイル共有のモデルが、大きく異なるからである。

画像化主義の人文情報学では、文書はデジタル画像であるとされる。そして、それは史料のデジタル化であるために、基本的には一度作られたら変更されない。それはオリジナルの代わりなので、変更は極力さけるべきだからである。

そして、変更の代わりに、オリジナル画像の上に、丁度、透明シートを置いて、その上に文字や図を書いたりするような形で、マークアップなどの追加情報を「書き込む」。つまり、透明テキスト付 PDF に似たものなのだが、画像としての文書と、「透明シート上のテキスト」がほぼ同じ形である透明テキスト付 PDF と異なり、HCP の場合は、このふたつは異なるものである。HCP では透明シート上には、オリジナル画像を補足するデータが書かれ、そのテキスト化は、別のテキストとして管理されるのである。オリジナルテキストのほとんどが手書き文書である HCP の場合には、翻刻されたテキストと図形的に重ねあわせることが元々困難であるから、この様な設計の方が自然なのである。

この結果、HCP では、SVN などは扱わず、むしろ、Flickr や Google Picasa などの WEB サービスが扱うような、原則的にはバージョンが変わらず、一方で、非常に大きなオリジナル画像史料ファイル群と、SVN などが管理するような、軽量である一方、頻繁かつ大規模に変更される、画像アノテーションとしての幾何学的数値情報や、リンク情報などからなる「透明シート上のアノテーション」の情報と「別途管理される翻刻などのテキスト」の二つに大別されるのである。

基本的には、後者は XML テキストなのであるが、対となる前者の画像情報についての幾何学的データとリンク情報を含む点が、通常の SVN とは大きく異なるのである。

例えば、画像データには行の概念があり、

それが翻刻テキストの行と結び付けられており、この部分は透明テキスト付 PDF に似た構造を持っているのである（しかし、続け字などのため、文字単位の対応は現実的にはほぼ不可能である）。

当初計画では、これらの HCP 特有の構造を反映するバージョン管理システムの基礎として、多角形としての画像の行情報と、その部分に対応する画像の断片からなる抽象行の概念を導入し、それを元に HCP を作成する計画であった。これにより、重い画像データの管理のパフォーマンスを改善するとともに、史料の論理的な構造を反映したデータ管理、特に、gs-file と呼ばれるアノテーションなどの、ユーザ作成のテキスト・データ、リンク・データなどの作成・管理・転送などを一般化・自由化することを意図していたのである。

ところが、次の「研究の方法」で述べる様に、HCP デザイン構築のためのエスノグラフィとなった、代表者林と京都大学田辺元史料研究会の実際の協同史料研究では、抽象行の概念は、ほとんど意味を持たず、また、HCP の最初の構想では、画像検索の前処理用として導入されていた画像の行の概念が、むしろ、共同翻刻の際に、参加者間のコミュニケーションのための道具としての役割が重要であり、画像検索を利用しない場合でさえ、画像行が協同作業の要の役割を果たすことが確認された。

これは画像の共有がプロジェクトで投影された画像で行われるために、画像の特定の行を指し示すことが難しく、また、コミュニケーションが専ら音声で行われるために、行の指定が、「何行目」という風に行番号で呼ぶことにより行われることが自然でかつ簡便だからである。また、書籍による演習、さらには対面による指導の際にも 1対1の場合を除けば、同じようにして行番号によるコミュニケーションが行われていることを思い出せば、これは多人数による文書の解析の際の基本的コミュニケーション手段だったのである。

そして、これは将来、同じ画像が遠隔地の研究者によりシェアされた場合にも、同じ状況が生じるものと考えられた。

この様に画像中の行を、画像における行の位置指示者として理解する場合、画像から行を抽象化してしまうという抽象行のアイデアは、むしろ不自然であった。

このため抽象行のアイデアを捨て、「コミュニケーションの道具」「文書画像における行・位置の指示者としての画像行」という新たな視点を導入し、これに基づいて HCP を再考し、モデルと実装を行うこととした。このエスノグラフィによる知見は幸いにも、プロジェクトの初期に得ることができた

め、研究目的の方針転換を早期に行うことができた。

また、申請時には、HCP 完成後に、HCP のバージョン管理機能を利用して、丁度、Wikipedia 風の協同翻刻機能、Folk-transcription を実現する予定であったが、研究開始後、画像検索などを除けば、計画していたものと、ほぼ同じ機能を持つ、Transcribe Bentham というシステムが英国ロンドン大学にて、哲学者ベンサムの手稿翻刻プロジェクトのために構築され、既に実用に供して、大きな成果をあげつつあることが判明したため、この計画は行わないことにした。

3. 研究の方法

本研究の特徴として、実際に役立つシステムを作り出すことを最優先にするという方針があった。そのため代表者が長年研究を続けていたソフトウェア工学の手法を使い開発したが、もっとも重要な役割を果たしたのが、エスノグラフィー、つまり、現実の利用の場に溶け込んで行う観察によるユーザの要求研究である。

すでに目的の欄に書いたように、これは、現在は人文学、特に京都学派の思想史の研究を行っている代表者林が自らの研究に、協同翻刻のツールとして、SMART-GS を適用することにより行われた。

HCP 開発体制は、(i) HCP モデル作成：相原を中心として相原と林が協同で担当、(ii) HCP サーバ作成：相原担当、設計は相原、実装は外注、(iii) HCP クライアント作成

(SMART-GS の HCP クライアント化)：林担当、実装は林と雇用した 2 名の研究員、および、林の学生、sourceforge.jp におけるオープンソース開発によるボランティア、という体制で行われた。また、マニュアルの作成はコードの作成者が担当し、英語化はアルバイトを使い行った。

以上の方法の中で、特に本研究の特徴として強調すべきは、林、および、研究員の内の一名である哲学者久木田が、田辺元史料研究会による田辺元史料研究に常に参加し(林は主催)、その経験を開発に反映し続けたことだろう。それにより、HCP の画像行の位置づけについての基本方針の変更だけでなく、マーカー、バスケット、画像行と翻刻行の対応などの、実用上極めて重要となった諸機能の導入が可能となったのである。

4. 研究成果

本研究の研究成果の最大のもの、当然ながら、HCP モデルの構築と、その実装である。

それは、具体的には、(1) HCP モデルの構築、(2) HCP サーバの構築、(3) HCP クライアントの構築、(4) 完成したシステムの

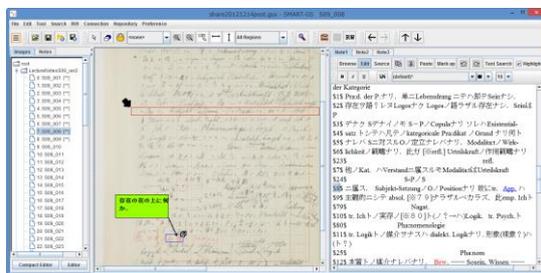
歴史研究への応用である。

これらの研究成果は、すべて GPL2 ライセンスにより、広く社会に還元する予定であるが、すでに (3) のクライアントの内、HCP サーバに関連しない部分、特にマーカー、バスケットなどの、オンサイト・リアルタイム協同のための機能を搭載した SMART-GS0.9 は、<http://sourceforge.jp/projects/smart-gs/> において公開している。近い将来、サーバも含めて、HCP のフル機能を GPL2 で公開することを計画している。

すでに公開されている SMART-GS0.9 における本研究の成果として、特筆すべきは、すべての資源に独自の URL が付与されたことがあげられる。これは表面上は特にめだつことがない成果であるが、将来の開発を考えると、非常に大きな成果である。これにより、今後、新しく開発される機能が有機的に結合できるようになるからである。

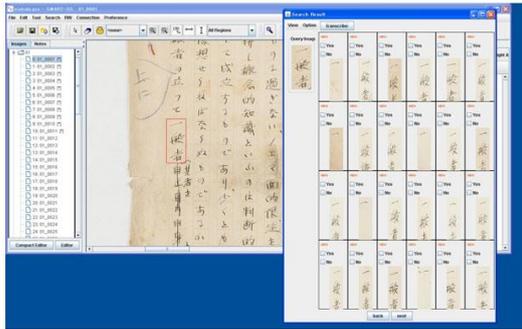
つまり、今回の研究で最も特筆されるべき成果は、公開されているツール SMART-GS0.9 を見ただけでは理解できないものの、将来的な開発の整合性を保証する基礎づけなのである。

しかし、現在公開されている SMART-GS0.9 に限定して言えば、その最大の成果は、オンサイト・リアルタイム協同翻刻のための諸機能であり、それには次の図で示す画像行と翻刻行の対応(右のペインの行番号を指示することにより、左のペインの画像行が一時的領域として表示されている)。



の機能や、特定の場所を記述しやすいマーカー機能、それらを簡単に保存できるバスケット機能などの多くの便利な機能の導入であると言えるだろう。実際、これらにより、田辺元史料研究は飛躍的に簡単になった。

そして、これらの機能以外にも、多くの改良が成された。たとえば、その完成までは、数百の画像に対する検索で検索ツールが能力の上限をむかえた画像検索を、並列処理などの方法を利用して、数千枚までの検索や、時間さえかければ、どの様に多い画像も逐次処理できるような画像検索機能は、その代表的例であろう。次に、その機能を用いて、五千枚近い、西田幾多郎手書き史料から、「一般者」という文言を画像検索した例を示す。



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)

- ① Kumiko Fujisawa, Kenro Aihara, Estimation of Interest from Physical Actions Captured by Familiar User Device, Whole Body Interaction, 査読有, pp.187-195, 2011.
DOI: 10.1007/978-0-85729-433-3_15

[学会発表] (計7件)

- ① 久木田水生、林晋、大浦真、SMART-GS——史料研究のためのソフトウェアツール、応用哲学会ワークショップ「デジタル・ヒューマニティーズの現在」2013年4月20日、南山大学
- ② 大浦真、林晋、久木田水生、SMART-GS: 文献研究のためのソフトウェアツール、言語処理学会ポスターセッション、2013年3月12日、名古屋大学、査読有
- ③ 久木田水生、林晋、大浦真、「デジタル・ヒューマニティーズの現在」、白眉センター&応用哲学・倫理学教育研究センター共催セミナー(招待講演)、2012年12月10日、京都大学
- ④ 林晋、久木田水生、大浦真、Smart-GSによる手稿解析——その実際(林)、SMART-GSネットワーク版による共同翻刻(久木田)、経済学史学会ヤングスカラーセミナー(招待講演とSMART-GS講習会実施)、2012年12月1日、大阪学院大学
- ⑤ Hayashi S., Aihara K., Kukida M., Ohura M., SMART-GS system: a software for historians by historians, JADH2012, 2012年9月16日、東京大学、アブストラクト集
<http://www.jadh.org/JADH2012-Abstracts-Online.pdf>, 査読有
- ⑥ 林晋、Smart-GSによる手稿解析——その実際、情報処理学会第84回デジタルドキュメント研究発表会、2012.01.20, 招待講演
- ⑦ 相原 健郎, 林 晋, 画像化主義に基づく

文献資料研究用ツール SMART-GS とその発展、情報処理学会 デジタル・ドキュメント研究会研究報告, 79(5), pp.1-5, 2011/7/22, 函館, 査読有

[図書] (計0件)

該当なし

[産業財産権]

該当なし

○出願状況 (計0件)

該当なし

○取得状況 (計0件)

該当なし

[その他]

ホームページ等

SMART-GS 開発および公開のサイト

<http://sourceforge.jp/projects/smart-gs/>

SMART-GS オンライン・マニュアル
和文

<http://smart-gs.sourceforge.jp/manual/ja/>

英文

<http://smart-gs.sourceforge.jp/manual/en/>

京都学派アーカイブ

<http://www.kyoto-gakuha.info>

京都大学文学研究科所蔵の西田幾多郎手書き文書を一括画像検索する SMART-GS を提供。サイトとデータの作成は他科研費の成果だが、これを可能にした並行検索機能は、本研究の成果。

6. 研究組織

(1) 研究代表者

林晋 (SUSUMU HAYASHI)

京都大学大学院文学研究科・教授

研究者番号: 40156443

(2) 研究分担者

相原健郎 (AIHARA KENROU)

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号: 90300706

(3) 連携研究者

なし