

デジタルアーカイブを利用した 歴史研究支援システムの構築

大和裕幸[†] 稗方和夫^{††} 畑野勇[†] 新木仁士[†]

歴史に関する史料にメタデータを付与して管理する歴史研究支援システムを提案する。このシステムは平賀譲デジタルアーカイブを基盤としている。はじめに著者の1人が行ってきた従来の歴史研究手順を整理し、史料の管理という問題を抽出した。その結果を踏まえ、web技術を用いて史料の書誌情報、所在するURLと研究者のコメントをまとめて保存・管理することで史料同士を関連付ける基本機能を持つシステムを構築する。そして歴史文献や書籍など紙媒体史料もスキニングによりデジタル化し、アップロードすることで他の史料と同等に取り扱う。また文章のみのテキスト史料だけでなく、テキストデータとして扱えない図面や写真などの様々な史料も画像ファイルとして同様に扱うことができるデータ構造とした。事例として巡洋艦コスト削減の歴史研究を行い、提案するシステムの有効性を立証した。

The Development of The Support System for Historical Research

Hiroyuki Yamato[†] Kazuo Hiekata^{††} Isamu Hatano[†] and
Hitoshi Araki[†]

A system of supporting historical research is proposed. The system which is based on Hiraga Yuzuru Digital archive attaches metadata to historical documents and manage them. Considering the traditional historical research process, the problem of historical material management is extracted. The function which makes historical materials related developed by web technology. The effectiveness of proposed system is evaluated by verifying the historical research hypothesis about the reduction of cruiser's construction cost.

1. はじめに

近年、多数のデジタルアーカイブが構築され、一般に認知されるようになった¹⁾。他 web サイトと連携する横断検索が可能な事例²⁾も見られ始め、デジタルアーカイブの利便性が高まってきている。しかし求める情報を得た後にこそ新たな知識を生み出す利用者の行動は始まる。その行動を支援するための仕組みを整えることが求められている。

本研究では平賀譲デジタルアーカイブ上の史料を基軸に、外部 web サイトの史料や紙媒体の史料と連携し、研究活動を支援するシステムを構築する。歴史研究全体の手順に即した支援システムを構築する先行事例は極めて少ないが、近似する例としては伊東ら³⁾、赤石ら⁴⁾の研究が挙げられる。テキスト文書だけでなく図面や写真などの史料を画像ファイルとして取り扱う本研究とは異なり、伊東らはテキスト文書を対象史料を限定した。そして史料に記載される文章をテキスト入力し、データ化することによって歴史史料を扱う支援システムを構築した。また赤石らは史料を時間や場所によって史料データを3次元空間に配置し、比較・分析することを主眼とする可視化システムを構築した。一方本研究では、近年発達した web 技術を用いて歴史研究手順に即した史料データの管理支援システムを構築する。

小沢ら⁵⁾、佐藤ら⁶⁾は社会科学分野においてデータベース技術を導入する意欲的な試みをデータベースシステム黎明期に行った。そして情報技術が発達するにつれて、史料を探し出すことに焦点を置き、史料の書誌情報に対して全文検索を行う朴らの研究⁷⁾、概念間の関係を事前に記述してコンテンツを分類するオントロジ検索を行う研谷らの研究⁸⁾が発表された。本研究では、史料を発見した後の研究手順も含めて支援することが歴史研究手順の改善に効果的であると考えた。そこで歴史研究者の史料の保存管理方法に着目し、デジタルアーカイブを用いた歴史研究手順に適した歴史研究統合支援システムの構築を行う。

2. 平賀譲デジタルアーカイブ

日本海軍において海軍造船中将まで登りつめ、軍艦の神様と呼ばれ、また東京帝国大学第13代総長を務めた平賀譲という人物がいた。平賀譲デジタルアーカイブ¹⁰⁾には、彼の残した文書44,000点が画像ファイルとして収められweb上で公開されている。それらのファイルには文書名や判明している範囲内での史料作成年月日など書誌情報が文書単位毎に付与されている。また史料は軍艦の船型、軍艦構造、報告書、手紙・写真など52種類のカテゴリに分類され、絞り込み検索が可能な環境が整えられている。

[†] 東京大学大学院新領域創成科学研究科人間環境学専攻

^{††} 東京大学大学院工学系研究科環境海洋工学専攻

3. 歴史研究手順

本研究で取り上げる歴史研究とは、歴史に関わる公文書や日記、覚え書きの形で残されたテキスト史料だけでなく、図面や表、グラフ、写真など画像史料を対象に比較分析し研究を行うこととする。具体的な事例として、巡洋艦夕張の排水量増加原因の検証を取り扱う。図1に従来の歴史研究手順と提案する歴史研究手順を示す。

また本研究においては、従来の歴史研究で取り扱われてきた歴史文献や書籍などの史料を紙媒体史料と名付ける。そしてデジタルアーカイブ上に存在する史料をweb上の史料と呼び、紙媒体史料をスキャンしデジタル化したものをデジタル化史料と呼ぶ。史料データという用語を使用した際は、web上の史料とデジタル化史料の双方を指す。さらに提案システム内で、史料データに付与される書誌情報と研究者のコメントを合わせたものをメタデータと定義する。

3.1 従来の歴史研究手順

著者の一人である歴史研究者が行ってきた従来の歴史研究手順を整理し、図1の左部に示した。従来の手順において“史料の探索”と“史料同士の考察”に多くの時間が割かれていた。特に史料同士の考察時において、考察を行っている時間よりも、求める史料がファイルのどこにあるのか探し出すことに時間が費やされていた。さらに史料をアナログで管理しているために一度ヒューマンエラーが起こると史料管理に多大な支障をきたす恐れをもつ。

この手順の中で、近年情報技術が用いられ始めた。史料が存在する分野・場所の特定、考察史料や論文の作成部分が該当部分にあたる。しかし依然として史料は紙媒体で保存され、史料管理に多くの時間と労力が割かれている。

3.2 デジタルアーカイブを用いた歴史研究手順

デジタルアーカイブが導入されたことにより、史料をweb上でいつでも閲覧することが可能となり、史料の探索時における課題点は解決された。一方、史料同士の考察時における課題を改善するには、考察手順のふたつ前に位置する“付箋を付けて史料をファイルに保存”という手順を根本的に変更する必要がある。そこでデジタルアーカイブを用いた歴史研究手順においては、“書誌情報と共に史料をデータとして保存”し、史料を管理する。その結果、史料をアナログで管理する必要がなくなり、考察時にユーザが今まで蓄積した史料を検索することが可能となる。また検索機能を効果的に用いることによって史料管理シートを作成する手順も省くことができる。これらの手順を踏むことで従来の研究手順の改善を図る。

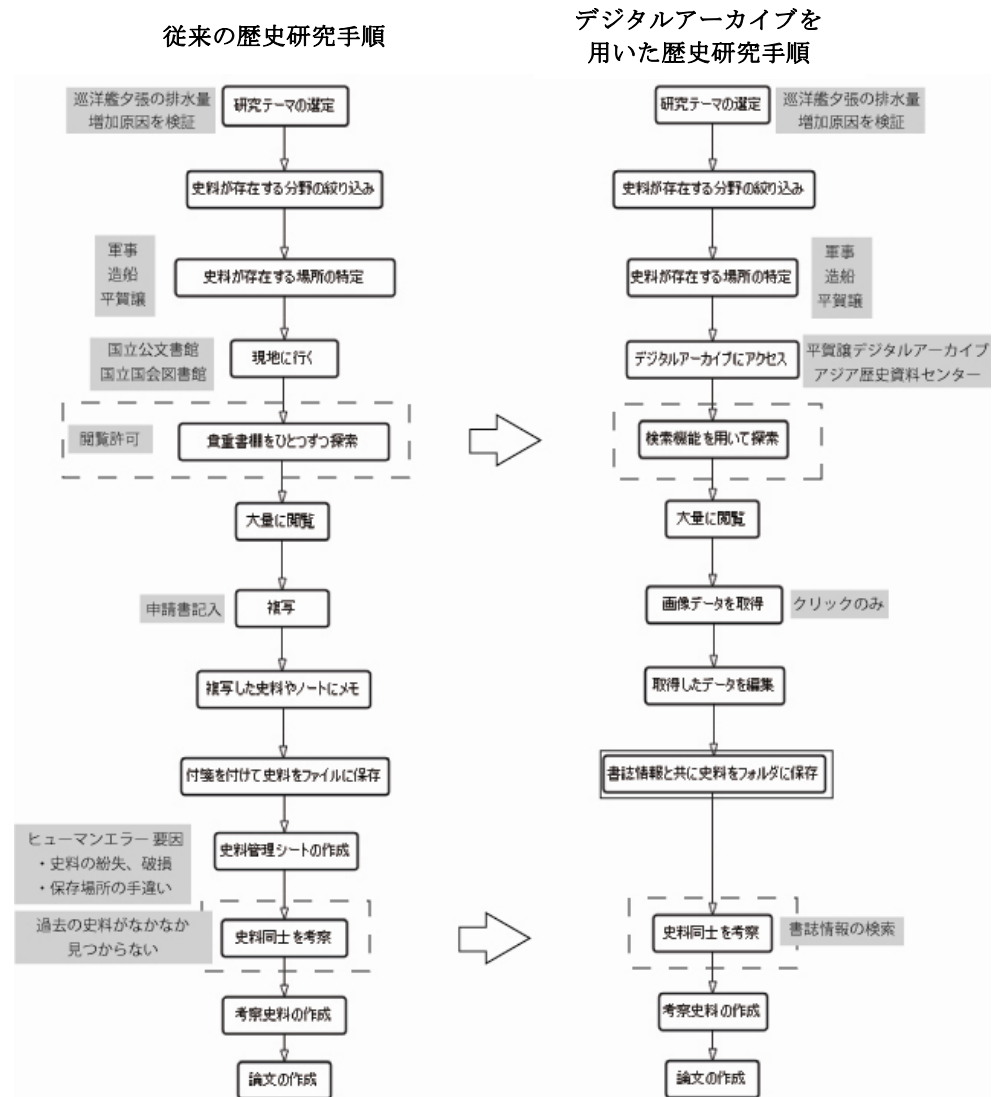


図1 歴史研究手順の比較

4. 提案する歴史研究支援システム

提案するシステム全体は HTML ブラウザをインタフェースとして持ち、Java 言語を用いて記述されている。提案する歴史研究支援システムの概念図を図 2 に示す。

システムは大きく分けて史料データの保存、メタデータの付与、蓄積した史料への検索という 3 つの部分で構成される。本システムはローカル環境に仮想サーバを導入し、その上でクライアント・サーバ型システムを web 上で動作することを想定して構築した。またセマンティック web の基幹技術のひとつであるメタデータ技術をデータ構造として使用した。簡略化のために XML 形式で記述し、スキーマは独自のものを用いた。

4.1 システムの基本機能

史料データにメタデータを付与して保存・管理し、検索機能によって目的の史料を得るといった基本機能を持つ。そのためフォルダ分けによって史料を分類しユーザの手によって整理するといった作業には最低限の機能しか設けない。

検索が実行されるたびに検索の対象となるデータの性質を読み取り、検索結果に含めるかどうかを判定するのは著しく非効率である。よって項目ごとにタグ付けを行い、予め検索をし易くする。そのために検索時の処理の効率化を図るメタデータを作成する。提案するシステムの概念図を図 2 に示し、順を追って説明する。

4.1.1 史料データの保存

史料データを保存する際、web 上の史料の場合は URL をファイルへ書き込む。対して、紙媒体史料をスキャニングによりデジタル化したデジタル史料、ユーザが作成した word や PowerPoint などの資料データの場合はサーバへデータをアップロードする。この時、アップロードした個人ディレクトリ内の URL をファイルへ書き込む。

4.1.2 メタデータの付与

史料データを保存する際、同時にメタデータを史料へと付与し、ファイルに書き込む。図 3 はメタデータの入力画面である。中央部に前項で述べた史料データの URL が書き込まれている。複数の史料データの URL をひとつのファイルに書き込み、文書名や書誌情報に加えてユーザ自身のコメントを入力し、保存・管理することで史料同士を関係付ける。そしてアップロードボタンを押すと、自動的に項目毎にタグ付けを行い、史料へのリンクアドレスと書誌情報をまとめてメタデータとして XML 形式で保存する。このとき平賀譲デジタルアーカイブ内の史料を保存する場合は、書誌情報を予め読み込み、ユーザ自身が直接入力する手間を省いた。デジタル史料の場合はユーザが書誌情報を入力する。また史料の原文や史料関与年月日などの項目を追加入力可能である。この時の XML データ構造を図 4 に示す。

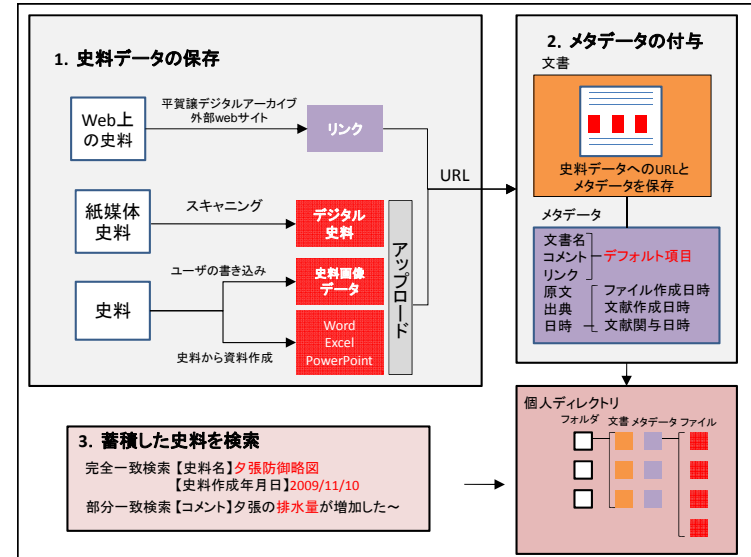


図 2 提案するシステム概念図

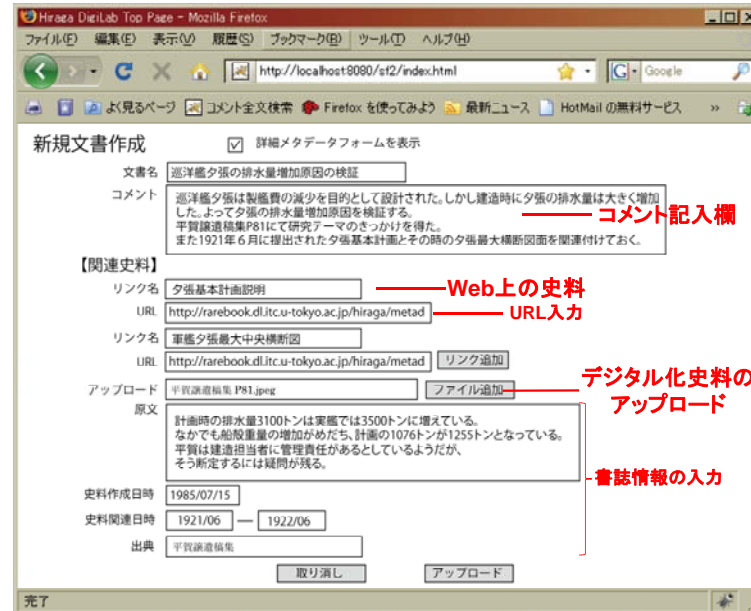


図 3 メタデータ付与画面

4.1.3 蓄積した史料への検索

ひとつずつの文書ファイルの書誌情報はメタデータとして保存されているため、検索時には検索対象のタグに対して検索キーワードと完全に一致するかどうか判断を行う完全一致検索を行う。ただし、ユーザが自然文を入力するコメントタグに関してはテストデータを一文字ずつ分割し、部分一致検索を行う。

4.2 システム使用手順

図3, 4, 5は巡洋艦夕張の排水量増加原因の検証という文書名のファイルに、平賀デジタルアーカイブ上の夕張基本計画、軍艦夕張最大中央横断図とデジタル化された平賀譲遺稿集 P81 という3つの史料を関連付けて保存した際のインタフェース画面やデータ構造である。

4.2.1 史料の探索

平賀譲デジタルアーカイブ上の史料の場合、システム内の専用検索ページを使用する。検索結果として史料の各種書誌情報が表示され、同時に画像閲覧画面へも直接リンクされる。平賀譲デジタルアーカイブ以外の Web 上の史料の場合、目的の史料を発見したらその URL を取得する。紙媒体史料の場合は従来の歴史研究手順と同様の探索手法をとる。

4.2.2 史料データの保存

平賀譲デジタルアーカイブ上の史料の場合、コメントを付けて保存することがボタンひとつで行える。このときリンクや付属する書誌情報も同時に保存される。その時書き込まれた結果が図4である。外部サイト上の史料の場合、新規文書作成画面を呼び出し、リンク入力フォームに保存する史料の URL を入力する。紙媒体史料の場合、スキャニングによりデジタル化した後、ファイルをアップロードし保存する。

4.2.3 史料同士の考察

過去に蓄積した史料を基に探索する。図5は蓄積された史料を一覧表示する個人ディレクトリ画面である。個人ディレクトリ内では同じ史料に関する文書はまとめて作成日時順に表示され、研究履歴を一覧で把握することができる。同時に保存された3つの史料を一目で把握することができ、クリックすると web 上の史料ならば史料の存在する web ページへリンクし、デジタル化史料ならばダウンロード出来る。

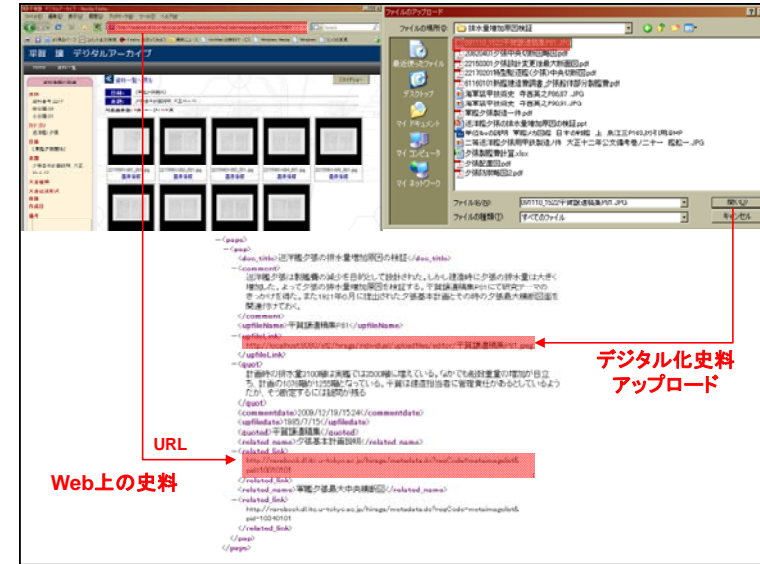


図4 史料間の関係を記述するデータ構造

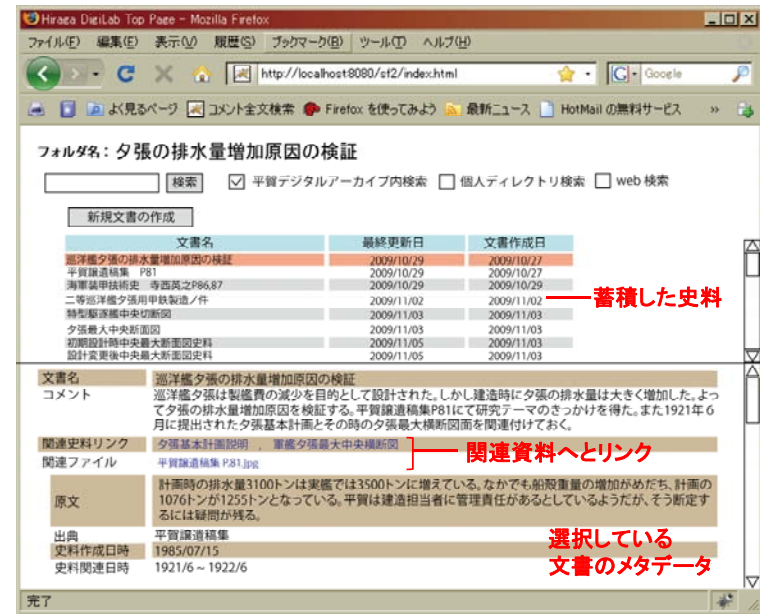


図5 個人ディレクトリ画面

5.2 夕張の建造後排水量増加原因の検証

巡洋艦夕張の設計は 1921 年 6 月に基本計画と図面が海軍艦政本部へと提出され、着工へ向けて準備されていたが、1922 年 3 月、「呉製鋼部維持ヲ兼ネ夕張ノ防御力増加ノ意味ヲ以テ全船ノ高張力鋼材ヲ甲鉄ニ変更」という設計変更訓令が下った。この訓令によって使用部材が高張力鋼材から NVNC 甲鉄へと変更された。そこで排水量の増加原因はこの設計変更にあると仮定した。1921 年 6 月に提出された図面と、設計変更後の図面において変更されている使用部材や材厚を比較し、その重量差を求め仮定を検証する。

設計変更前と変更後の最大中央切断図を示す図 7、図 8 において 2 点を比較した結果、防御甲板では一平方フィート当たりの重量が 35lbs から 40lbs へと増加し、一部の船殻構造材についても 40lbs から 50lbs へと増加していた。またいずれも使用部材は高張力鋼材から NVNC 甲鉄へと変更された。ここで中央切断図と防御略図(図 9)から機関室部分の部材寸法を測定し部材面積を求め、それぞれの使用材料値を乗じて機関室部分の部材重量を計算する。その結果、設計変更前は 307.5 噸、設計変更後は 324.5 噸であったため、機関室部分における設計変更による重量増加は 17.0 噸という結果が得られた。

新艦建造費に記載されている夕張の機関室部分防御材重量は 343.5 噸であった。これは、図面から読み取った値と比べると設計変更後より更に増加している。軍艦設計において排水量は設計開始時に定められ、その値を超えることは許されない。そのため使用材料が変更されたとしても既定排水量に収まるよう設計を修正する。しかし、記録によると設計変更後の図面から得た重量値よりも建造後の重量値のほうが大幅に増加している。機関室部分のみでこれだけの差が出たということは艦全体においてはより差が大きくなることは自明である。よって、夕張の排水量増加原因の一因として使用材料の変更が挙げられることが立証できた。また、設計時と建造時の実状に差があることも確認が出来、巡洋艦夕張においても通説が正しいことが立証された。

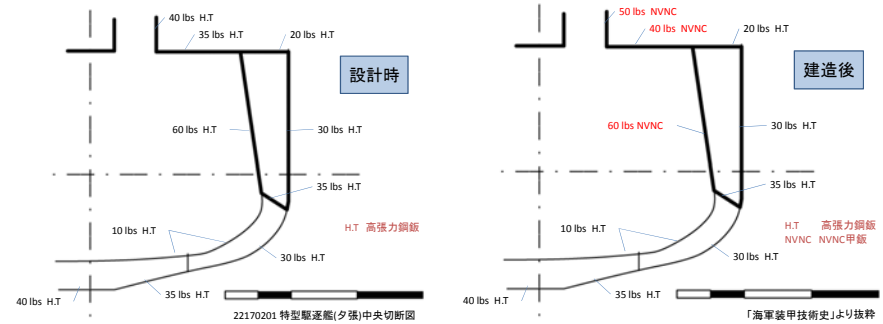
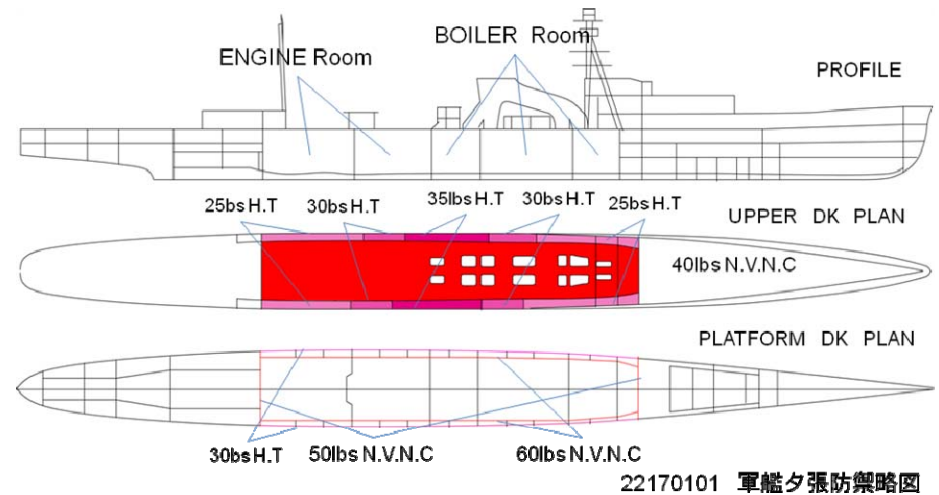


図 7 設計変更前夕張最大横断面図

図 8 設計変更後夕張最大横断面図



22170101 軍艦夕張防禦略図

図 9 設計変更後夕張防禦略図

6. 考察

提案システムの史料保存機能によって、史料を挟んだファイルを人手で管理する必要がなくなり、史料管理に割かれていた多くの時間が低減された。また史料を画像ファイルとして扱い、書誌情報と共に保存することにより史料の検索を行うことが出来た。その結果、過去の史料を探し出す手間が軽減し、更に史料管理シートを作成する手順を省くことができた。

提案するシステム上で行った夕張の建造後排水量増加原因の検証では、テキストだけでなく図面なども含む史料データを読み解くことで通説の検証を行った。検証の中で中央切断図を蓄積した史料から探し出す際、検索結果として示される史料をファイル作成日時でソートすることにより、関連史料を閲覧した時の軌跡が明確に表示された。更に同時に示される関連資料を追うことにより、今まで研究者の中で関連付けられていなかった史料を新たに紐付けるきっかけが生まれた。このことからデジタルアーカイブ上の史料を用いる歴史研究において、史料データにメタデータを付与し管理する本支援システムは有効であった。

7. 結論

従来の歴史研究手順とデジタルアーカイブを用いた研究手順を比較し、その問題点について分析を行った。その結果に基づき、計算機によって歴史研究を支援するシステムを試験的に構築した。また試験システム上で巡洋艦夕張の排水量増加原因を検証し、従来の手順ならば数十ページに及ぶ史料ファイルを作成する必要がなくなり、かつ史料管理に費やす時間を削減した。このことにより従来の歴史研究手順の改善に効果的であることを立証した。今後は史料を探し出す手順において、より精度の高い検索結果を得るために検索機能を強化し、支援システムの構成を検討していく必要がある。

謝辞

本研究は独立行政法人日本学術振興会科学研究費補助金による挑戦的萌芽研究21650230の補助を受けて実施したものである。

参考文献

- 1) A. P. Bishop, "Document structure and digital libraries: how researchers mobilize information in journal articles," *Information Processing & Management*, vol. 35, pp. 255-279, (1999)
- 2) 国立公文書館/横断検索, <http://zgw.digital.archives.go.jp/GlobalFinder/htdocs/index.html>
- 3) 伊東幸宏, 小西達裕, 三浦崇, 赤塚大輔, 田村貞雄, 阿部圭一, 赤石美奈, 中谷広正, テキ

- スト史料の抜粋・分類機能と分類結果の俯瞰機能による歴史学研究支援, 情報処理学会論文誌, vol.40(3), pp.821-830, (1999)
- 4) 赤石美奈, 岡田義広, 中谷広正, 伊東幸宏, 田村貞雄, 史料の管理・検索・可視化機能を持つ歴史学研究支援統合環境の構築, 情報処理学会論文誌, vol.40(3), pp. 831-839, (1999)
 - 5) 朴明哲, 森本雅史, 立花純児, 村川猛彦, 宇都宮啓吾, 中川優, 人文研究を支援するデータベースシステム: 聖教検索および系図表示, 情報知識学会誌, vol. 17, pp. 105-110, (2007)
 - 6) 小沢一雅, 考古学研究支援型データベースシステムの構成, 情報処理学会論文誌, vol.26(5), pp.936-945, (1985)
 - 7) 佐藤真知子, 橋原秀晴, 井岡幹博, 黒川雅人, 洪政国, 杉田繁治, 久保正敏, 山本泰則, 民族学研究支援のための標本画像検索システム, 情報処理学会論文誌 vol.29(12), pp.1108-1118, (1988)
 - 8) 研谷紀夫, 馬場章, 文化資源オントロジーの構築とその活用, 情報知識学会誌, vol.17(2), pp.129-134, (2007)
 - 9) 溝口理一郎, オントロジー研究の基礎と応用, 人工知能学会誌, vol. 14(6), pp.977-988, (1999)
 - 10) 東京大学附属図書館 平賀譲デジタルアーカイブ, <http://rarebook.dl.itc.u-tokyo.ac.jp/hiraga/>
 - 11) 平賀譲: 平賀譲遺稿集, 出版協同社(1985)
 - 12) 寺西英之: 海軍装甲技術史, 慶友社(2007)
 - 13) 堀川一男: 海軍製鋼技術物語, アグネ技術センター(2000)
 - 14) 山下正喜: 三菱造船所の原価計算, 創成社(1995)
 - 15) 泉江三: 軍艦メカニズム図鑑 日本の戦艦 上, グランプリ出版(2001)
 - 16) 海軍艦政本部第4部: 公文備考 軍艦夕張製造一件, 国立公文書館 アジア歴史資料センター, http://www.jacar.go.jp/DAS/meta/listPhoto?IS_STYLE=default&ID=M2008061910511536944