

東京大学大学院新領域創成科学研究科

人間環境学専攻

修士論文

セマンティック・ウェブ技術を用いた
歴史研究支援システムの開発

2014 年 2 月 7 日提出

指導教員 大和 裕幸 教授 印

学生証番号 47-126749

中村 覚

目次

目次.....	I
図目次.....	V
表目次.....	VIII
第1章 序論.....	1
1.1 電子資料の普及と活用.....	2
1.2 歴史研究.....	3
1.3 目的.....	5
1.4 本論文の構成.....	5
第2章 関連研究.....	7
2.1 はじめに.....	8
2.2 要素技術.....	8
2.2.1 デジタルアーカイブ.....	8
2.2.2 セマンティック・ウェブ.....	13
2.2.3 類似画像検索.....	20
2.3 既存研究.....	25
2.3.1 セマンティック・ウェブ技術を応用した研究.....	26
2.3.2 歴史研究支援を目的とした研究.....	26
2.3.3 文書自動分類システム「KASHIWADE」.....	29
2.4 本研究の位置づけ.....	32
第3章 提案手法.....	34
3.1 はじめに.....	35
3.2 歴史研究プロセスの課題と必要要件の整理.....	35
3.2.1 歴史研究プロセスの課題.....	35
3.2.2 課題解決に向けた必要要件の定義.....	37
3.3 提案する歴史研究環境.....	39

3.4 セマンティック・ウェブ技術を用いた史料管理手法	41
3.4.1 メタデータフィールドの設計	42
3.4.2 オントロジーを用いた概念の体系化	46
3.4.3 セマンティック・ウェブ技術による有用性	47
第4章 開発した歴史研究支援システム	49
4.1 はじめに	50
4.2 開発したシステムの概要	50
4.3 史料の収集	51
4.3.1 メタデータ登録機能	52
4.3.2 ファイル登録機能	53
4.3.3 メタデータ入力項目の詳細	54
4.4 史料の整理	55
4.4.1 メタデータ編集機能	55
4.4.2 オントロジーの編集機能	56
4.5 史料の分析	58
4.5.1 検索結果表示機能	58
4.5.2 検索機能	59
4.5.3 史料の並び替え機能	60
4.5.4 史料の比較機能	61
4.6 プラグインによるメタデータの分析機能	62
第5章 ケーススタディ：長門型戦艦の設計履歴分析	66
5.1 はじめに	67
5.2 ケーススタディ内容	67
5.3 分析対象史料	69
5.4 ケーススタディにおける前処理	71
5.4.1 オントロジーの追加	71
5.4.2 局所エッジヒストグラムを用いたインデックスの作成	71
5.5 ケーススタディの流れ	72
5.6 史料の収集	74
5.7 史料の整理	74

5.8 史料の分析	76
5.9 分析結果を利用した史料の再検索	77
5.10 ケーススタディ結果	78
5.11 ケーススタディのまとめ	79
第6章 実験：外務省来電の送付先決定過程分析	81
6.1 はじめに	82
6.2 実験内容	82
6.3 分析対象史料	83
6.4 実験の流れ	85
6.5 メタデータフィールドの選定	86
6.6 史料の収集	87
6.7 史料の整理	90
6.8 史料の分析	92
6.8.1 筆跡と作成日の関係に基づいた分析	94
6.8.2 筆跡と閲覧者の関係に基づいた分析	95
6.9 実験結果	97
6.10 実験のまとめ	98
第7章 考察	99
7.1 ケーススタディにおける検索精度評価	100
7.1.1 検索精度評価指標	100
7.1.2 精度評価における検索方法	101
7.1.3 オントロジーを用いた検索精度の評価	101
7.1.4 類似画像を用いた検索精度の評価	103
7.1.5 特定した年代を用いた検索精度の評価	105
7.2 既存の歴史研究プロセスの課題点に対する評価	108
7.2.1 「史料の収集」プロセスにおける評価	108
7.2.2 「史料の整理」プロセスにおける評価	109
7.2.3 「史料の分析」プロセスにおける評価	110
7.2.4 提案手法に基づく歴史研究プロセスの改善	111
7.2.5 来電における標準メタデータセットの提案	113

第 8 章 結論	116
8.1 結論	117
8.2 今後の展望	117
8.2.1 「デジタルキュレーションシステム」 への応用	117
8.2.2 他分野の文献調査プロセスへの応用	118
謝辞	119
参考文献	122

図目次

図 1-1 歴史研究者の史料（画像ファイル）と調査結果の管理方法例.....	4
図 2-1 平賀譲デジタルアーカイブ[10]	10
図 2-2 検索結果画面[10].....	11
図 2-3 史料の閲覧画面[10]	12
図 2-4 アジア歴史資料センターの検索結果画面[15]	13
図 2-5 セマンティック・ウェブにおける要素技術体系[17].....	14
図 2-6 RDF のグラフ表現例.....	16
図 2-7 RDF のシリアル形式の表現例.....	16
図 2-8 RDF のシリアル形式の表現例.....	17
図 2-9 RDF スキーマのグラフ表現例.....	18
図 2-10 オントロジーのシリアル形式の表現例	19
図 2-11 SPARQL のクエリ例	20
図 2-12 部分画像と画像ブロックの分割イメージ[28]	23
図 2-13 5つのエッジ種類とフィルタ[28]	24
図 2-14 画像ブロックとラベル例[28]	24
図 2-15 「事物つながり検索」における相関図[35]	27
図 2-16 スケッチ図からの検索[5].....	28
図 2-17 Nagasaki Archive のユーザインタフェース[36]	29
図 2-18 KASHIWADE システム概要図[37]	30
図 2-19 メタデータ管理画面[37]	31
図 2-20 本研究の位置づけ	33
図 3-1 歴史研究プロセスの課題点の整理	36
図 3-2 提案する歴史研究環境.....	39
図 3-3 セマンティック・ウェブ技術を用いた史料管理手法例	41
図 3-4 同義語及び階層概念の記述例	46
図 3-5 セマンティック・ウェブ技術で記述された史料情報の有用性.....	47
図 4-1 システム概要図	50

図 4-2	メタデータ登録 UI	52
図 4-3	ファイル登録 UI.....	53
図 4-4	メタデータ入力項目の詳細.....	54
図 4-5	メタデータ編集インタフェース.....	56
図 4-6	オントロジー編集インタフェース	57
図 4-7	検索結果インタフェース	58
図 4-8	メタデータの確認機能.....	59
図 4-9	検索インタフェース	60
図 4-10	史料の並び替え機能の例	61
図 4-11	史料の比較機能.....	62
図 4-12	プラグインの実行環境.....	63
図 4-13	プラグインの実行インタフェース	63
図 5-1	戦艦「長門」[41]（上）と「加賀」[42]（下）	68
図 5-2	平賀譲デジタルアーカイブ史料の RDF 記述例	70
図 5-3	局所エッジヒストグラムを用いたインデックスの作成	72
図 5-4	ケーススタディの流れ.....	73
図 5-5	煙突本数に関する調査結果の蓄積例.....	74
図 5-6	形状に関する類似画像の表示例.....	75
図 5-7	史料群を時系列に並び替えた検索結果画面	76
図 5-8	比較機能を用いた煙突本数の比較結果	77
図 5-9	長門の搭載汽缶数に関する報告書	78
図 5-10	加賀の搭載汽缶数に関する報告書	79
図 6-1	外務省における来電の処理過程.....	82
図 6-2	分析対象とする来電の例	84
図 6-3	実験の流れ	85
図 6-4	史料収集プロセス	88
図 6-5	メタデータ登録機能を用いた史料の登録例	89
図 6-6	アジア歴史資料センターの公開史料（左）と撮影史料（右）の比較	90
図 6-7	メタデータ編集機能を用いた史料情報の整理例	91
図 6-8	蓄積された史料情報の RDF 表現例.....	92
図 6-9	送付先の記述における筆跡例	93

図 6-10 月別の送付先指示筆跡の出現頻度	94
図 6-11 筆跡 B における閲覧者の登場割合	95
図 6-12 筆跡 A における閲覧者の登場割合	96
図 6-13 筆跡 A と閲覧者の関係分析	97
図 7-1 適合率・再現率	100
図 7-2 クエリとした利用した画像（左）と不正解画像（右）例	105
図 7-3 正解史料（上）と史料が持つメタデータ（下）	106
図 7-4 歴史研究プロセスの課題点の整理	108
図 7-5 提案手法に基づく歴史研究プロセスの改善	112
図 7-6 クラウドソーシング	113
図 7-7 来電に関する標準メタデータセット例	114
図 8-1 システムイメージ	118

表目次

表 2-1	グラディエントの計算に用いる微分オペレータ	22
表 2-2	ラプラシアン of 計算に用いる微分オペレータ	22
表 3-1	史料管理のためのメタデータフィールド	43
表 3-2	ユーザの調査結果等を記述するためのメタデータフィールド	44
表 3-3	複数人によるメタデータ管理を行うためのメタデータフィールド	45
表 4-1	システムに蓄積されたメタデータの Microsoft Excel 出力例	65
表 5-1	平賀讓デジタルアーカイブで使用されている書誌情報例	69
表 5-2	長門型戦艦の暗号名一覧[43]	71
表 6-1	分析に用いた史料一覧	84
表 6-2	選定したメタデータフィールド	87
表 7-1	正解史料群	102
表 7-2	オントロジーを用いた検索精度の評価結果	103
表 7-3	類似画像検索を用いた検索精度の評価結果	104
表 7-4	作成年月日に関する制約を加えた検索精度の評価結果	107

第1章 序論

1.1 電子資料の普及と活用	2
1.2 歴史研究	3
1.3 目的	5
1.4 本論文の構成	5

1.1 電子資料の普及と活用

近年の急速な情報技術の発達に伴い、企業や行政、大学といったさまざまな分野・組織で電子化された資料が普及している。電子資料を活用することによって計算機による知識抽出や地理的・時間的ギャップを超えた情報共有が可能になる等、その活用によって多くの効果が期待できる[1]。

企業においては、2005年に施行されたe-文書法[2]によって、財務・税務関係の帳票類や取締役会議事録など、商法や税法で保管が義務づけられている文書について、紙文書だけでなく電子化された文書ファイルでの保存が認められるようになった。これにより文書・帳票類の保管にかかるコストの削減や、企業間取引の電子化の加速といった様々な効果が期待されている。また電子文書から企業にとって有益な情報を機械的に抽出する研究も盛んに行われている。たとえば竹内ら[3]はコールセンターにおける顧客とのやりとりの記録をテキストマイニング技術によって分析するシステムを開発した。彼らは電子文書から問い合わせ内容の話題や増減傾向を抽出することでFAQを自動生成し、それを商品知識の教育や生産性の向上などの業務プロセスの改善に繋げている。

教育の分野では情報技術を用いて学習や教育を行うeラーニングが普及している。eラーニングは「教材・学習材」と「学習管理システム(Learning Management System, LMS)」から構成されるが、「教材・学習材」については紙文書が従来提供してきた文章や静止画像に加えて、動画や音声などを組み合わせたマルチメディア形態の提供が可能となっている。また「学習管理システム」によって、教師による教材の保管、学習者への教材の適切な配信、学習者の学習履歴や試験問題の成績などの統合的な管理が可能となり、授業のあり方が変化しつつある。黒瀬ら[4]は3次元CAD教育において自由曲面、自由曲線の学習支援を行うシステムを開発し、創成・発見的学習環境の実現や学生の関心度の向上といった成果を上げている。

行政においては2013年6月17日から18日に行われたG8サミットにおいて「G8オープンデータ憲章[6]」が合意された。この憲章の一部を以下に抜粋する。

原則1：原則としてのオープンデータ

- データによっては、公表出来ないという合理的な理由があることを認識しつつ、この憲章で示されているように、政府のデータすべてが、原則として公表されるという期待を醸成する。

原則3：すべての者が利用できる

- 幅広い用途のために、誰もが入手可能なオープンな形式でデータを公表する。
- 可能な限り多くのデータを公表する。

原則5：イノベーションのためのデータの公表

- オープンデータ・リテラシーを高め、オープンデータに携わる人々を育成する。
- 将来世代のデータイノベーターの能力を強化する。

このオープンデータという考えは、CSVやXMLなどの再利用可能かつ機械処理が可能な形で情報を広く一般に公開することで情報の存在を世の中に伝え、第三者の利用を自由化している点に特徴がある。現在世界各国では政府が保有するデータを公開する「オープンガバナンスポータル」が立ち上げられている。日本における端的な例としては「データシティ鯖江[7]」が挙げられる。鯖江市は市の人口や気温、施設の位置情報といったデータをホームページ上で公開している。そしてこれまでに災害時の避難所の位置を表示するアプリケーションや、市内のバスの位置情報をリアルタイムに表示するアプリケーションが第三者によって開発されるなど、新しいムーブメントの発信地として注目を集めている。このような鯖江市の取り組みも電子資料を活用した事例の一つである。

1.2 歴史研究

歴史研究の分野でも電子化された史料が普及しており、博物館、美術館、公文書館や図書館などではデジタルアーカイブ化が進められている。従来保存の観点から公開が難しかった貴重史料がデジタル化され、インターネット上で広く公開されることで、場所・時間の制約を受けることなく、複数人が同時に史料にアクセスできるようになった。電子史料の活用を目的とした研究として、小早川ら[5]は画像処理技術を用いて類似する画像史料の検索や比較といった新たな切り口からの史料研究を可能とするシステムを開発している。

一方で、複数人の歴史研究者へのヒアリングの結果、史料の収集に関してはデジタ

ルアーカイブを用いて容易に行うことが可能となったが、調査から得られた知見の蓄積や知見の再利用による分析などは各研究者が独自の方法で行っていることが分かった。図 1-1 にはある研究者の史料、および調査結果の管理方法を示している。

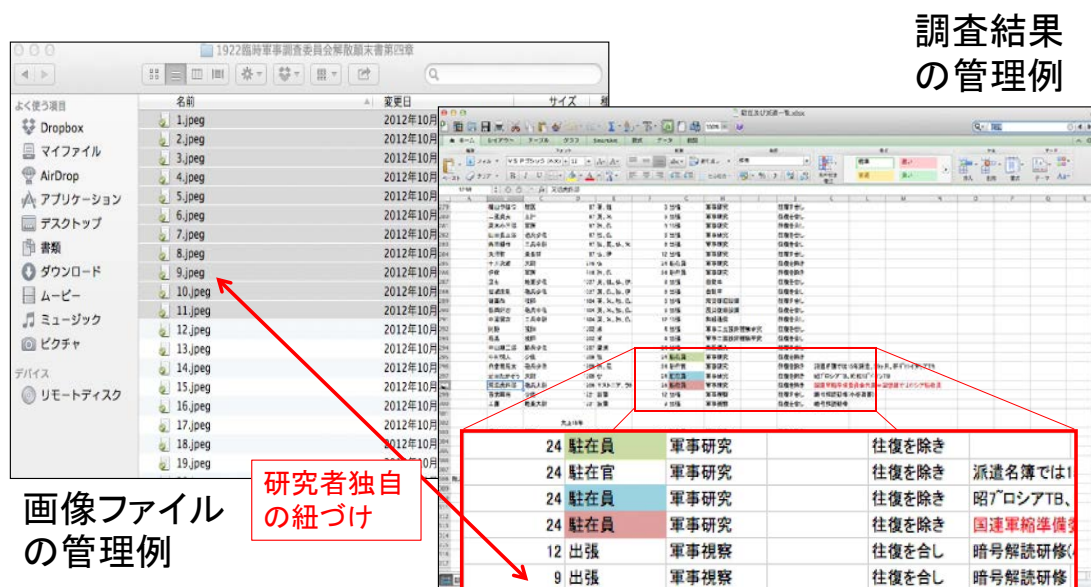


図 1-1 歴史研究者の史料（画像ファイル）と調査結果の管理方法例

（左：フォルダに格納された史料、右：資料毎の調査結果をまとめた Microsoft Excel）

左図に示すように、デジタルアーカイブ上で公開されている史料を研究者の PC といったローカル環境にダウンロードし、フォルダの階層に基づいて史料の分類を行っている。また右図に示すように、分析から得られた知見等は別途 Microsoft Excel 等のツールに保存している。ここでの問題点として、史料とそれに関する情報が別ファイルに記述されているため、再度史料を閲覧する際にはデジタルアーカイブにアクセスする、またはダウンロードした資料を PC 内で検索する必要がある、デジタルアーカイブなどインターネット上で史料が URL 等の識別子を与えられた上で公開されているメリットを生かすことができていない。また各研究者の調査結果がローカル環境に蓄積されるため、他の研究者との情報共有を行うことは困難である。例えば各研究者の調査結果や知見等は各人が独自に与えた識別子によって史料との紐づけが行われ

ており、またそれらは自由記述文で記述されている等、他の研究者が容易にそれらの情報を利用することは難しい。その結果、歴史研究は各研究者の暗黙知に頼る部分が多く、検索漏れや忘却による歴史研究の低品質化、さらにある研究者が亡くなった際、彼の研究結果や成果を後世に残すことができないといった問題を抱えている。

1.3 目的

近年の資料の電子化によってさまざまな分野でその利活用が進められているが、1.2 で述べたように歴史研究の分野ではそのメリットを十分に生かすことができていない。

本研究ではセマンティック・ウェブ技術を用いて歴史研究を支援するシステムの開発を行う。具体的には電子史料をクラウド環境で管理し、書誌情報や研究者の調査結果を史料のメタデータとして管理を行う。それらメタデータを再利用することで、史料の並び替えや比較、計算機を用いた自動分類といった分析を支援する。さらに従来の歴史研究の属人性の問題に対して、クラウド上で複数研究者が史料の収集や整理、分析を共同で行うことが可能な環境を構築し、各研究者の労力の軽減や複数研究者間の協調学習を可能とする歴史研究手法を提案する。また開発したシステム上でのケーススタディ及び実験を行うことで、本研究で提案する手法の評価を行う。

1.4 本論文の構成

第 2 章では、本研究と関連する要素技術や研究事例を説明し、本研究の位置づけについて述べる。

第 3 章では、歴史研究を支援するために必要な要件を定義し、それを実現するための手法を提案する。

第 4 章では、第 3 章で提案した手法を実装したシステムについて述べる。

第 5 章、第 6 章では、開発したシステム上で行ったケーススタディ及び実験について説明する。

第 7 章では、ケーススタディ及び実験に基づいた提案手法、開発したシステムの有効性について考察する。

第 8 章では、本論の結論と今後の展望について述べる。

第2章 関連研究

2.1 はじめに	8
2.2 要素技術	8
2.2.1 デジタルアーカイブ	8
2.2.2 セマンティック・ウェブ	13
2.2.3 類似画像検索	20
2.3 既存研究	25
2.3.1 セマンティック・ウェブ技術を応用した研究	26
2.3.2 歴史研究支援を目的とした研究	26
2.3.3 文書自動分類システム「KASHIWADE」	29
2.4 本研究の位置づけ	32

2.1 はじめに

本章では、本研究で用いる要素技術や関連する研究事例について述べ、本研究の位置づけについて述べる。

2.2 要素技術

2.2.1 デジタルアーカイブ

デジタルアーカイブとは博物館・美術館・公文書館や図書館の収蔵品を始め有形・無形の文化資源等をデジタル化して保存等を行うことである。デジタル化することによって、文化資源等の修復・公開や、ネットワーク等を通じた利用も容易となる。

デジタルアーカイブという言葉は 1990 年代中頃、東京大学名誉教授である月尾嘉男[8]によって、国内で初めて使われた。2003 年 7 月に決定された「e-Japan 戦略 II」[9]では、コンテンツ産業などの国際競争力の向上や、海外における日本文化への理解の向上を図る手段の一つとして、放送・出版などのコンテンツや、美術館・博物館・図書館などの所蔵品、Web 情報、特色のある文化などのデジタルアーカイブ化、および国内外への情報発信の推進を掲げている。同年 8 月に決定された「e-Japan 重点計画-2003」や、2004 年 2 月に決定された「e-Japan 戦略 II 加速化パッケージ」でも、教育用コンテンツの充実・普及や、国などの有するコンテンツの保存と利用機会の拡大を図るために、引き続き、各種コンテンツのデジタルアーカイブ化を推進していくとされている。

デジタルアーカイブによって期待される効果として、以下の項目が挙げられる。

- (1) 史料の破損・劣化防止
- (2) 新しい表現の実現
- (3) 時間的、地理的な制約を超えた史料提供
- (4) 様々な角度からの史料検索

(1) に関しては史料をデジタル化することにより、破損を恐れることなく貴重史料の提供が可能となる。史料を保護するために制限がある複製についても、印刷するこ

とで提供が可能となる。(2) は複数の史料の映像を部分的に切り出し、再合成することや、様々な解説・音声などを付け加えることにより、元の史料を比較して多くの情報を含んだ形での情報提供が可能となる。(3) については計算機を用い、いつでもどこからでも史料へアクセスが可能となる。また、複数の利用者が同時に史料を閲覧できる。(4) は史料のDB化することで、様々な角度から史料の検索や分類を行うことができ、史料を新たな切り口から閲覧・分析することが可能となる。

以下では本研究のケーススタディ及び実験で対象とした二つのデジタルアーカイブについて説明する。

2.2.1.1 平賀讓デジタルアーカイブ

図 2-1 に示される平賀讓デジタルアーカイブ[10]では、海軍造船中将および第 13 代東京帝国大学総長であった平賀讓が生前に遺した艦艇計画・建造関係の技術史料約 44,000 点が画像化され、東京大学附属図書館電子化コレクションで一般に公開されている。



図 2-1 平賀譲デジタルアーカイブ[10]

平賀譲デジタルアーカイブを公開するにあたり、44,000 点の史料を整理されていた封筒毎に分け、5,232 件にまとめられた。それら各史料に史料名や判明している範囲内での作成年月日などの書誌情報が付与されている。また、史料は軍艦の艦型、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ている。特に軍艦設計案や軍艦建造技術を記した史料群は、日本海軍の軍艦建造黎明期から太平洋戦争に至る当時の軍艦設計・建造の実状を明らかにする貴重な史料群であると言える。本デジタルアーカイブは徐々に知られるところとなり、八八艦隊関係の研究に利用され、論文[11][12]や書籍[13]として出版されている。

次に平賀譲デジタルアーカイブのシステムについて述べる。平賀譲デジタルアーカイブは、堀内カラー[14]製のビューワーと検索システムを有している。図 2-2 は検索画面を示す。



図 2-2 検索結果画面[10]

画面左部の検索画面で書誌情報別の検索を行い、画面の右部に検索結果が表示される。各検索結果から図 2-3 に示す各史料の閲覧画面に移動することでき、史料の詳細情報や史料の画像データを閲覧することができる。各画像データは拡大・縮小させることができ、また史料の印刷やダウンロードも可能である。



図 2-3 史料の閲覧画面[10]

2.2.1.2 アジア歴史資料センター

アジア歴史資料センター（Japan Center for Asian Historical Record）[15]は、日本が保管するアジア近隣諸国との関係に関する歴史史料を提供している。インターネットを通じて明治期から第二次世界大戦終結までの期間に関する情報提供を行っており、日本の国立公文書館によって運営されている。2011年4月時点では、約2,246万画像、162万件の目録データを提供している。収録対象の史料は、それぞれ日本の国立公文書館、外務省外交史料館、防衛省防衛研究所図書館が保管するアジア歴史資料であり、デジタル化に応じて順次公開されている。

アジア歴史資料センターでは 2.2.1.1 の平賀譲デジタルアーカイブと同様、目録などが史料に付与された基本的な書誌情報による検索機能や、史料の拡大・縮小機能、印刷・ダウンロード機能が提供されている。



図 2-4 アジア歴史資料センターの検索結果画面[15]

2.2.2 セマンティック・ウェブ

ティム・バーナーズ＝リーは「現在のウェブはセマンティックウェブ（意味を持つウェブ）と呼ぶ新しい形式のコンピュータ網に進化していく。コンピュータがコンテンツの意味を理解し推論することによって、インターネットの可能性は飛躍的に拡大するであろう[16]」と述べている。

セマンティック・ウェブの目的はウェブページの閲覧という行為に、データの交換の側面に加えて、意味の疎通を付け加えることにある。計算機に「自律的な処理」の能力を与えるためには、情報リソースに意味を持たせる必要があり、「メタデータ」と「オントロジー」が重要となる。メタデータとは、「全文検索の限界を解決するための、データに対する構造的データ」である。またオントロジーとは「ある特定分野の概念や知識、分類体系や推論ルール集」のことを指し、「語彙の定義」や「語彙と語彙の関係」を記述する。

たとえば、「アドレス」という単語は、「住所」、「メールアドレス」、「メモリアドレス」等、様々な意味に解釈される。ホームページ上に「アドレス」と書かれている場

合、人間は前後の文脈から「住所」であると判断できるが、計算機では困難である。したがって「アドレス」に何のアドレスかを説明するメタデータを付与することによって、計算機による理解が可能となる。

2.2.2.1 セマンティック・ウェブの要素技術

セマンティック・ウェブを実現するための要素技術体系を図 2-5 に示す。

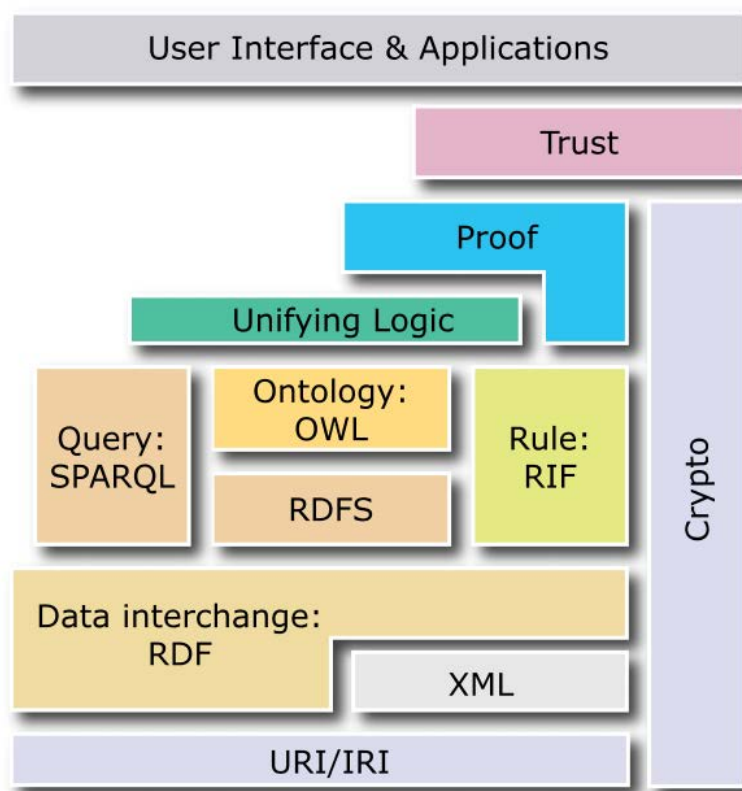


図 2-5 セマンティック・ウェブにおける要素技術体系[17]

上記の体系の内、本研究に関連する要素技術の説明を以下に加える。

- URI (Uniform Resource Identifier) : 統一資源識別子といわれ、URL の考え方を拡張したものである。
- XML (Extensible Markup Language) : 文書やデータの意味や構造を記述するためのマークアップ言語の一つである。「タグ」と呼ばれる特定の文字列で文に

情報の意味や構造、装飾などを埋め込んでいくことができる。

- Namespaces（名前空間）：名前の集合を分割することで、衝突の可能性を低減しつつ、参照を容易にする概念である。
- RDF（Resource Description Framework）：メタデータを記述するための仕組みである。
- RDF Schema（Resource Description Framework Schema、RDF スキーマ）：クラス、プロパティなどの語彙を定義し、メタデータの持つ意味を規定する。
- OWL（Web Ontology Language）：RDF スキーマより精密な意味情報管理を行う「オントロジー」を記述する言語である。
- SPARQL（SPARQL Protocol and RDF Query Language）：RDF 用の検索言語であり、RDB の「SQL」に相当する。

次に上記の要素技術から、本研究で特に重要となる技術について詳細な説明を加える。

2.2.2.2 RDF

RDF は W3C によって標準化されたデータ形式であり、1 つの情報を主語、述語、目的語の 3 つ組(トリプル)で表現する。RDF で表現する場合、目的語はリテラル（単なる文字列）でも良いが、主語と述語は URI で表現されたリソースである必要がある。

なお、URI は一定の書式によってリソース（資源）を指し示す識別子であり、http/https や ftp などのスキームで始まり、コロン（:）による区切りの後にスキームごとに定義された書式に基づいてリソースを示す。また URI によって示されるリソースは計算機が扱うデータに限らず、人や会社、書籍などを示すことも可能である。

たとえば次の日本語の文「<http://157.82.251.163/kashiwade/>というホームページの名前は KASHIWADE であり、作成者は中村覚である」を、RDF の「文」として捉える。この RDF を有向ラベル付きグラフ(directed labeled graph)を用いて表現した例を図 2-6 に示す。RDF グラフでは、リソースを楕円、述語をアーク（矢印）、リテラルを長方形で表現する。アークの矢印は主語を始点とし、目的語を終点としている。また RDF/XML によるシリアル形式の表現例を図 2-7 に示す。

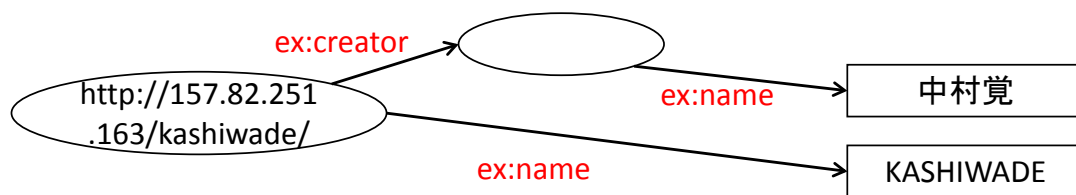


図 2-6 RDF のグラフ表現例

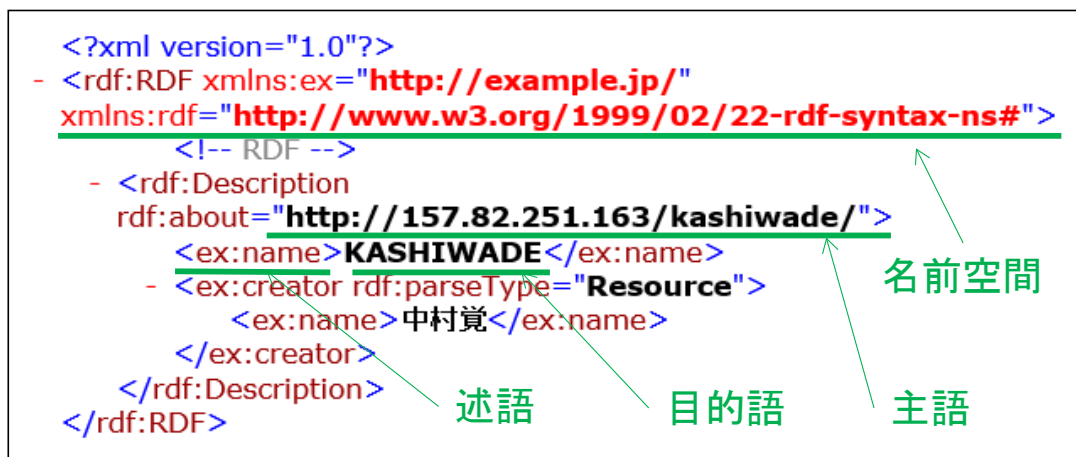


図 2-7 RDF のシリアル形式の表現例

2.2.2.3 RDF Schema

RDF はリソースの述語やリソース間の関係を記述するモデルを提供するが、そこで用いられる述語そのものについては定義しない。creator、date、channel といった述語は、RDF Schema を用いて定義し、それを URI で参照する。RDF Schema は、これらの述語や一般的なリソースのカテゴリを定義するための基本的なメカニズムを提供し、仕様で用意される基本クラス、基本的な述語などを用いて、必要なクラスを新たに定義し、そのインスタンスやサブクラスを導出する。

例えば RDF で記述した 2.2.2.2 について、RDF Schema を用いて Homepage クラ

スの定義と「Homepage の作成者」という述語の定義を行う。Homepage クラスの定義については、Homepage クラスを `rdfs:Class` タイプのリソースとして定義することでクラス派生を実現できる。「Homepage の作成者」という述語の定義については、たとえば `rdf:Property` タイプの派生とし、`rdfs:domain` と `rdfs:range` によってそれぞれ主語が Homepage クラスのインスタンスであり、目的語が Person クラスのインスタンスであると定義する。以上の定義を RDF/XML によるシリアル形式で表現した例を図 2-8 に示す。また、この RDF スキーマを 2.2.2.2 の RDF に適用したグラフ表現例を図 2-9 に示す。

```
<?xml version="1.0"?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    <!-- Homepageクラスの定義 -->
    - <rdf:Description rdf:ID="Homepage">
      <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
        schema#Class"/>
    </rdf:Description>
    <!-- creatorという述語の定義 -->
    - <rdf:Description rdf:about="http://example.jp/creator">
      <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-
        syntax-ns#Property"/>
      <rdfs:domain rdf:resource="http://example.jp/Homepage"/>
      <rdfs:range rdf:resource="http://example.jp/Person"/>
    </rdf:Description>
  </rdf:RDF>
```

図 2-8 RDF のシリアル形式の表現例

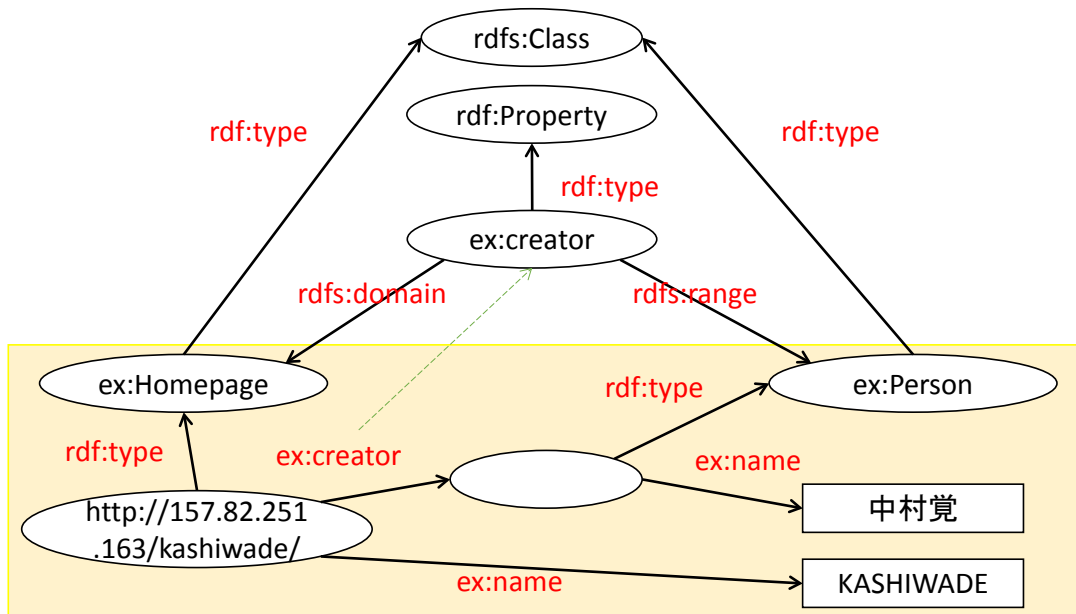


図 2-9 RDF スキーマのグラフ表現例

2.2.2.4 オントロジー

オントロジーはあるドメイン内の概念とそれらの概念間の関係のセットとしての知識の形式的な表現である。オントロジーで表現された知識を利用して、エージェントが高度な検索などを行う研究は数多く行われている[18][19][20][21]。

オントロジーは以下に示す要素から成り立っている。

- (1) オントロジー全体の情報を記述するヘッダ。
- (2) クラスの定義を記述するクラス公理。
- (3) プロパティのグローバルな定義を記述するプロパティ公理。
- (4) 固体（インスタンス）を記述する事実。

またオントロジーのクラス公理では以下のように規定されている。

- owl:intesectionOf（論理積に相当する。）
- owl:unionOf（論理和に相当する。）
- owl:complementOf（論理否定に相当する。）

この規定を使用して、たとえば、「好きなスポーツというクラスを野球とサッカー以外」というクラスを定義した例は図 2-10 のように記述する。

```
<owl:Class rdf:ID="FavoriteSports" >
  <owl:complementOf >
    <owl:Class >
      <owl:unionOf rdf:parseType="Collection" >
        <owl:Class rdf:about="#Baseball"/>
        <owl:Class rdf:about="#Soccer"/>
      </owl:unionOf >
    </owl:Class>
  </owl: complementOf >
</owl:Class>
```

図 2-10 オントロジーのシリアル形式の表現例

2.2.2.5 SPARQL

SPARQL とは RDF クエリ言語の一種であり、SPARQL クエリには基本グラフ・パターンと呼ばれる 1 組のトリプルが含まれている。サブグラフからの RDF 用語を変数に代替可能で、結果がサブグラフに同等な RDF グラフである場合、基本グラフ・パターンはその RDF データのサブグラフにマッチする。

簡単な例として、2.2.2.2 の RDF からサイト名を発見するための SPARQL クエリを図 2-11 に示す。

```
PREFIX ex: http://example.jp/

SELECT ?name

WHERE {
  <http://157.82.251.163/kashiwade/> ex:name ?name
}
```

図 2-11 SPARQL のクエリ例

クエリは **SELECT** 句と **WHERE** 句で構成される。**SELECT** 句はクエリの結果に現れる変数を識別し、**WHERE** 句はデータ・グラフにマッチする基本グラフ・パターンを提供する。この例の基本グラフ・パターンは、目的語の位置に 1 つの変数 (?name) を持つ、1 つのトリプル・パターンから成る。

2.2.3 類似画像検索

類似画像検索とは指定した画像と類似した画像を検索することであり、CBIR (Content-Based Image Retrieval) とも呼ばれる。画像間の類似の比較には色[22]、形状[23]、テクスチャ[24]、特徴点[25][26]等の様々な特徴量を用いて行われ、類似画像を用いたプロセス改善に関する研究が多く行われている[27]。

ここでは「形状」に関する特徴量の抽出、つまりエッジ検出手法について述べる。本研究では平賀譲デジタルアーカイブで公開されている史料群に対して類似画像検索を行うが、これら公開されている資料は白黒画像であり、また「図面」といった画像の形状に基づいた情報が重要となるためである。

2.2.3.1 エッジ

エッジとは物体の外縁を表す線、または画像を特徴づける線要素である。画像処理ではエッジを抽出することを「エッジ抽出」または「輪郭抽出」と呼ぶ。画像中の物

体と物体、あるいは物体と背景の境界がエッジであるため、「画像の濃度や色の変化が急である場所」がエッジであり、エッジ抽出では濃度情報に注目する。

2.2.3.2 エッジ抽出手法

エッジは濃度値が急激に変化する部分であるため、関数の変化量を算出する微分演算がエッジ抽出に利用できる。微分は一次微分（グラディエント）と二次微分（ラプラシアン）がある。またデジタル画像では離散的な値を取るため、隣接画素同士の差分を取ることで微分と代用される。微分演算を行うための隣接画素同士の演算を表現する係数の組を「微分オペレータ」と呼び、さまざまな種類が存在する。

2.2.3.3 一次微分

座標 (x, y) における濃度の勾配を表す一次微分の値は大きさとベクトル量 $G(x, y) = (fx, fy)$ として表現できる。ここで fx は x 方向の微分、 fy は y 方向の微分を示す。 fx , fy はデジタル画像では式(2.1)で計算できる。

$$\begin{aligned} fx &= f(x+1, y) - f(x, y) \\ fy &= f(x, y+1) - f(x, y) \end{aligned} \tag{2.1}$$

これら微分値が算出されると、エッジの強さは $\sqrt{fx^2 + fy^2}$ 、方向は (fx, fy) の向きとして計算できる。エッジの方向はエッジの濃度の暗部から明部の方向となる。

またグラディエントの計算に用いる微分オペレータの例を表 2-1 示す。

表 2-1 グラディエントの計算に用いる微分オペレータ

オペレータ名	通常の差分	Robertsオペレータ	Sobelオペレータ
fx を求める オペレータ	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$
fy を求める オペレータ	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$

2.2.3.4 二次微分

二次微分はエッジの強さのみを検出するのに用いられ、デジタル画像では式(2.2)で表される。またラプラシアン計算に用いる微分オペレータの例を表 2-2 示す。

$$L(x, y) = 4 \times f(x, y) - \{f(x, y - 1) + f(x, y + 1) + f(x - 1, y) + f(x + 1, y)\} \quad (2.2)$$

表 2-2 ラプラシアンの計算に用いる微分オペレータ

オペレータ名	ラプラシアン1	ラプラシアン2	ラプラシアン3
	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$

2.2.3.5 局所エッジヒストグラム

ここでは本研究で用いた局所エッジヒストグラム[28]について説明する。本手法は MPEG-7 標準化規格[29]において推奨されているエッジ分布に関する記述子である。この記述子は各画像から 80 個のヒストグラムを構成するため、画像の特徴を小さいデータ量で保持することができる点が特徴である。なお、MPEG-7 は、ISO/IEC JTC 1 の Moving Picture Experts Group(MPEG)によってつくられたマルチメディア用メタデータ表記方法の国際標準化規格である。MPEG-1,2,4 などとは異なり、動画データのエンコードが目的ではなく、XML をベースとしたメタデータ記述によるマルチメディアデータの高速な内容検索を目的としている。

本手法ではまず画像を $4 \times 4 = 16$ の部分画像に分割する。各部分画像においてその部分画像のエッジ分布を示すヒストグラムを生成する。分布の種類を抽出するため、各部分画像はさらに「画像ブロック」と呼ばれる小領域に分割する。この分割イメージを図 2-12 に示す。

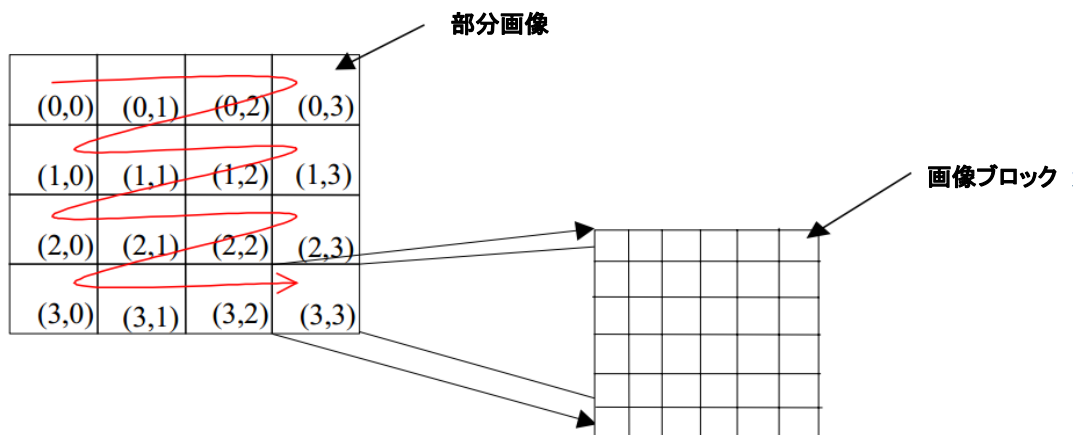


図 2-12 部分画像と画像ブロックの分割イメージ[28]

次に画像ブロックから図 2-13 に示す 5 つの種類のエッジを抽出する。画像ブロックからエッジを検出後、各部分画像における各エッジの合計値からヒストグラムを生成する。

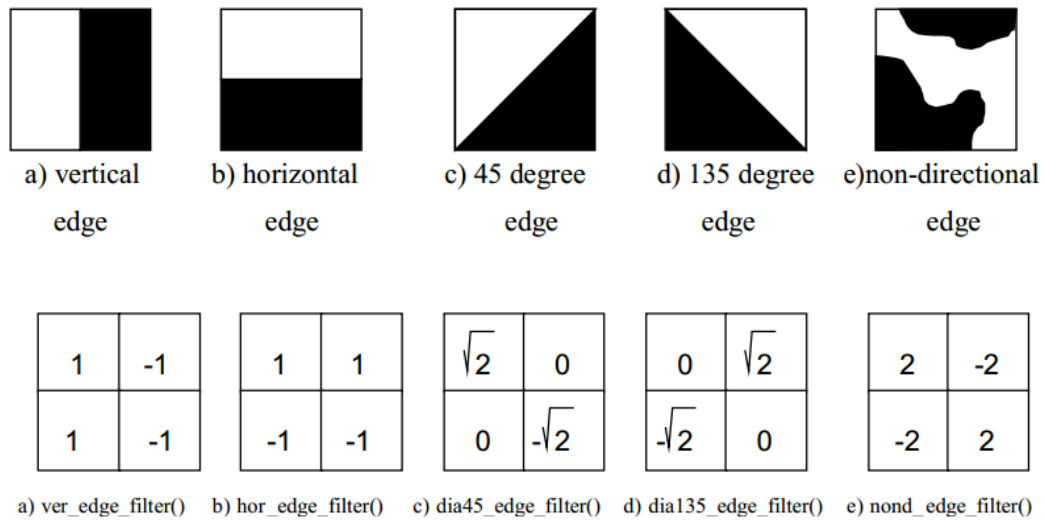


図 2-13 5つのエッジ種類とフィルタ[28]

エッジは図 2-14 に示すように各画像ブロックから検出する。ここでは各画像ブロックをさらに4つの部分ブロックに分割する。そして4つの部分ブロックから輝度の平均値を算出し、それらは図 2-13 に示すオペレータで畳み込みを行うことで、エッジの大きさを抽出する。

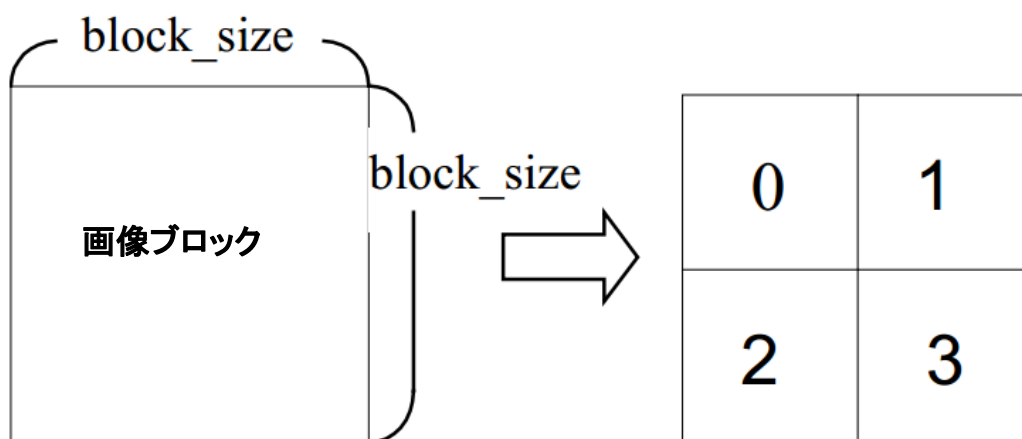


図 2-14 画像ブロックとラベル例[28]

ここでは各画像ブロックからエッジの強度を算出する方法をより具体的に説明する。今、各画像ブロックに図 2-14 に示すように 0 から 3 までのラベル付けを行ったとする。そして (i, j) 番目の画像ブロックの $k(k = 0, 1, 2, 3)$ 番目の部分ブロックについて、平均濃度 $A_k(i, j)$ とすると、各画像ブロックの各エッジの強度は式(2.3)から式(2.7)を用いることで算出される。

$$ver_edge_stg(i, j) = \sum_{k=0}^3 |A_k(i, j) \times ver_edge_filter(k)| \quad (2.3)$$

$$hor_edge_stg(i, j) = \sum_{k=0}^3 |A_k(i, j) \times hor_edge_filter(k)| \quad (2.4)$$

$$dia45_edge_stg(i, j) = \sum_{k=0}^3 |A_k(i, j) \times dia45_edge_filter(k)| \quad (2.5)$$

$$dia135_edge_stg(i, j) = \sum_{k=0}^3 |A_k(i, j) \times dia135_edge_filter(k)| \quad (2.6)$$

$$nond_edge_stg(i, j) = \sum_{k=0}^3 |A_k(i, j) \times nond_edge_filter(k)| \quad (2.7)$$

そして算出された 5 つの種類のエッジの強度について、それらの最大値が閾値を超える場合には、その画像ブロックがそのエッジを持つと判断する。この検出したエッジを用いて各画像に対してヒストグラムを作成する。異なる画像間の類似度を計算する場合には、そのヒストグラムの差分から類似度を定量化する。

2.3 既存研究

本研究では史料情報をセマンティック・ウェブ技術によって管理し、歴史研究を支援するシステムを開発することが目的である。本節ではセマンティック・ウェブ技術を文献調査に応用した研究、また歴史研究の支援を目的とした研究について述べる。さらに開発するシステムの基盤として利用した、メタデータの自動付与機能を提供する文書自動分類システム「KASHIWADE」について述べる。

2.3.1 セマンティック・ウェブ技術を応用した研究

J. Gonzalez ら[30]はセマンティック・ウェブ技術を用いて絵画や写真といった画像に半自動的なタグ付けを行うシステムを開発した。彼らは人手によるフリーテキストの方法で付与されるタグには誤字脱字や必要情報の欠落が発生し、正確に画像情報を伝えておらず、結果としてユーザが目的とする画像を得ることができていない点を課題としている。この課題に対して、タグとして用いられる語の関係をオントロジーによって定義し、それを辞書として利用して既存タグの標準化を半自動的に行った。さらにそのオントロジーを検索時における辞書としても利用することによって、計算機による効率的な検索機能を構築した

P. Castells ら[31]は新聞記事を公開するデジタルアーカイブを対象とし、ウェブ上にアップロードされる際に付与されるメタデータに表記揺れが生じ、それがアーカイブにおける検索効率の向上を妨げている点を課題としている。彼らは記事の属性（医療や政治）、記事の種類（見出しやまとめ）、およびキーワードによる3つのクラスをオントロジーによって定義し、それらを既存のアーカイブに応用し、上記の問題の解決を図っている。本研究においてメタデータを付与する際に生じる表記揺れへの対応の点で参考とした。

また C. Costilla ら[32]は複数のアーカイブの統合をし、横断的に資料の検索・閲覧を可能にする手法を提案している。彼らは個々のアーカイブの情報を Wrapper によって XML 形式に変換し、それをさらにオントロジーによって実装される Mediator によって管理することで、異なるアーカイブを統合して検索対象とする枠組みを実現した。

2.3.2 歴史研究支援を目的とした研究

伊東ら[33]はテキスト形式の史料を対象とした史料研究を支援する研究を行った。彼らはまず歴史研究に利用される歴史史料の特徴から支援を必要とするポイントを整理した。具体的には（1）史料が膨大であり、かつ形式が一定でないため、必要とする情報の抽出に膨大な時間を要する。さらに（2）抽出された情報から知見を引き出す際に、一定の基準で整理され、比較する必要があるという点である。よって彼らは（1）に対して、表現パターンテーブルを辞書として用いることで、間接的な表現で

記述された史料の抽出を可能にした。(2)については彼らが対象とする史料の特徴上、時間と場所からなる二次元空間に史料を可視化し、全体を俯瞰できるシステムを開発している。本研究では画像史料を対象とする点で異なるが、彼らが整理した史料や歴史研究の特徴はシステムを構築する上で大いに参考となった。

赤石ら[34]は伊東らが開発したシステムを基盤とし、史料を三次元空間に配置し、歴史研究における仮説作成や検定を支援する可視化システムを構築した。彼らは二次元空間地図上に、時間軸を中心とした多次元データを配置した。プロットされるデータは史料データの他に、当時の気象チャートや米価に代表される経済の動きといったデータをユーザの選択が可能で、ユーザの様々な切り口に基づいたデータ間の相関に基づいた比較や、時間的・空間的変化の多角的な解析を可能としている。

研谷ら[35]は様々な形態・内容の歴史史料を統合的に扱うことを目的とし、歴史上の事物の関係を記述するためのオントロジーを設計する手法を提案している。またそのオントロジーを用いて史料の関連付けを行い、図 2-15 に示す史料間の関係を可視化した検索システムを開発している。本研究においても、歴史上の事物の関係を記述するためのメタデータ設計の点で参考とした。

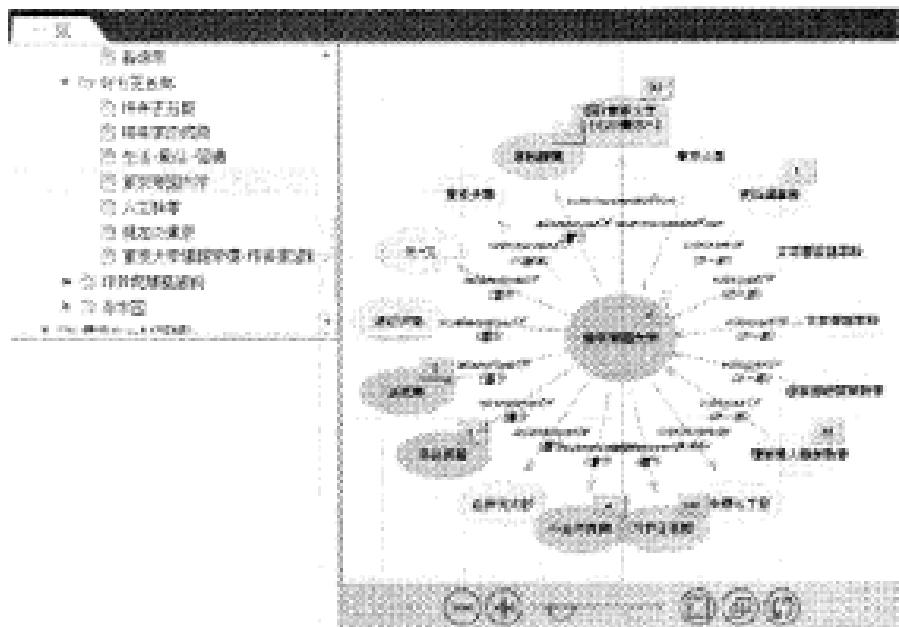


図 2-15 「事物つながり検索」における相関図[35]

アーカイブ内の画像検索を効率化することを目的とした研究として、小早川ら[5]の研究がある。彼らは従来の画像に付与されたメタデータに対するテキスト検索の限界を指摘し、ウェーブレット変換と高次局所自己相関特徴量を用いた画像データ索引を自動生成する手法を提案した。この手法を国立歴史民俗博物館の民俗学資料カードの写真 3315 枚に対して適用し、図 2-16 に示すスケッチ画像や輪郭画像、例示画像などをクエリとした検索を繰り返して画像を検索できる対話的類似画像検索システムを開発した。本研究でも画像情報に基づいた類似画像検索の実装において参考とした。

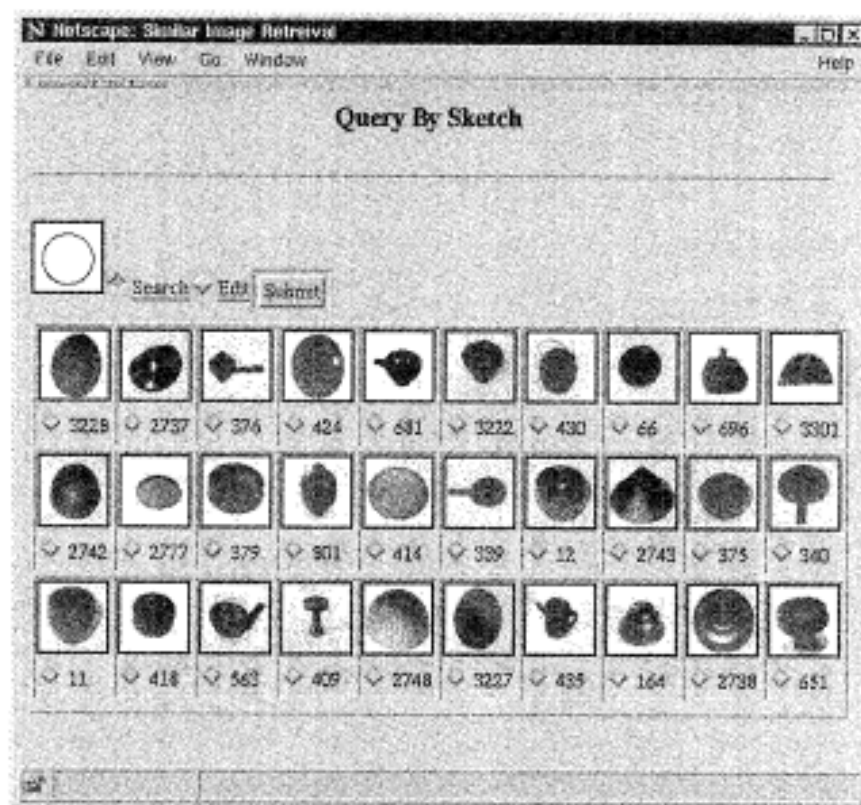


図 2-16 スケッチ図からの検索[5]

渡辺ら[36]は従来のデジタルアーカイブは単独の展示施設等の史料のデジタル化と保管、その単独ユーザによる個別利用を前提にデザインされており、扱われている事

象をより多面的・総合的に理解するきっかけが生まれにくい点を課題としている。この課題に対して図 2-17 に示すように、長崎原爆資料館の資料を Google Earth 上にマッピングし、それらの史料に対してユーザがコメントを付与していくシステムを開発した。その結果当時の被害者によって付与されたコメントによる第三者の理解度の向上や、異なるユーザの証言の差異などが発見され、事象に対する多面的・総合的な理解を促す多次元デジタルアーカイブズを構築した。本研究においても複数のユーザがコメントを共有することで創発的な学びの場を生み出す環境構築の点で参考とした。



図 2-17 Nagasaki Archive のユーザインタフェース[36]

2.3.3 文書自動分類システム「KASHIWADE」

稗方ら[37]は分類対象に依存したテキスト処理以外の共通部分をプラットフォームとして提供する「KASHIWADE」(以下、KASHIWADE)を開発した。その概要を図 2-18 に示す。

2.3.3.2 メタデータ管理機能

特定のリソースについてのメタデータフィールドとメタデータバリューは、図 2-19 に示すメタデータ管理画面に表示され、メタデータバリューの参照や更新を行うことができる。

メタデータフィールド	メタデータバリュー	メタデータ更新	メタデータクリア
http://kashiwade.org/2012/09/kd#cause	[原因要約] CORBAクライアントでメモリ破壊が発生しました。	更新	削除
http://kashiwade.org/2012/09/kd#closeCode	BF	更新	削除
http://kashiwade.org/2012/09/kd#component	OD	更新	削除
http://kashiwade.org/2012/09/kd#etc	類似げし	更新	削除
http://kashiwade.org/2012/09/kd#eventQa		更新	削除
http://kashiwade.org/2012/09/kd#group	ProductAテストデータ	更新	削除

図 2-19 メタデータ管理画面[37]

2.3.3.3 メタデータ抽出プラグイン管理機能

本システムはプラグインを読み込み、実行する機能を備えている。プラグインはシステムに保存されたバイナリデータと、リソースに付与されたメタデータ（フィールドとバリューのセット）を読み込み、既存メタデータフィールドのバリューの更新や新規メタデータの追加を行う。各プラグインはシステム内のプラグイン管理画面に一覧表示され、選択したメタデータフィールドのバリュー別、リソースの拡張子別にプラグインを実行できる。

2.4 本研究の位置づけ

2.2.1 ではデジタルアーカイブの普及に伴い、従来保存の観点から公開が難しかった史料へのアクセシビリティが向上し、学術的な発展が期待されている点について述べた。一方で既存のデジタルアーカイブの多くは「史料の公開」と「目録に基づいた基本的な分類」機能の提供に留まっており、閲覧した結果等から得られた知見の蓄積や再利用、ユーザの目的に基づいた史料の分析を行う環境は整えられていない。

またデジタルアーカイブを対象とした既存研究においては、オントロジーを用いた史料の関連付けや画像処理による類似画像検索など、従来の「目録に基づいた分類」と比べて自由度の高い検索環境の構築を行っている。しかし依然としてそれらの情報は提供者側の視点での分類であり、各ユーザの知見の蓄積や再利用を行う環境を要していない。

よって本研究では、ユーザが得た知見を蓄積し、それらの情報を再利用することでユーザのニーズにあった分類や比較を可能とする分析環境を構築することで、歴史研究における既存プロセスを支援することを目的とする。

この課題に対して、中村[38]はセマンティック・ウェブ技術を活用し、デジタルアーカイブで公開されている史料やユーザが保有する史料の書誌情報や知見等をメタデータによって管理し、知見の蓄積やその再利用を可能とする手法を提案した。またその手法に基づいた歴史研究を支援するシステムを開発し、ケーススタディを通じて提案手法の有用性を検証している。しかしこの研究では以下の点が課題として残されている。

- 研究者個人の史料管理を主眼としており、複数研究者が共同で研究を行うための環境は整えられていない。
- 人手による史料への書誌情報や知見の蓄積を主眼としており、それらの作業を行う上での各研究者の労力については考慮されていない。
- 提案手法の有用性を評価する上で、筆者自身によるケーススタディにとどまっており、歴史研究者の利用による評価は行われていない。

したがって本研究では上記課題の解決に向けて中村によって提案された手法を拡張し、開発したシステム上で歴史研究者による利用を通じた提案手法の評価を行う。ここで述べた各研究の成果と課題点、および本研究の目的を図 2-20 にまとめる。

	目的	課題
既存のデジタルアーカイブ	<ul style="list-style-type: none"> ・歴史的資料へのアクセスの容易化 ・資料の破損、劣化の防止 	<ul style="list-style-type: none"> ・「資料の閲覧」と「目録に基づいた分類」の提供に留まり、史料の分析や再利用を目的とした研究環境は整えられていない。
関連研究	<ul style="list-style-type: none"> ・オントロジーを利用した関連資料の検索 ・画像情報に基づいた類似画像検索 ・地図へのマッピングによる可視化 	<ul style="list-style-type: none"> ・「目録に基づいた分類」に比べ、高度な検索を可能としているが、知見の蓄積や再利用を目的とした環境は整えられていない。
中村による研究	<ul style="list-style-type: none"> ・セマンティック・ウェブ技術を用いた資料情報の管理手法の提案 ・各ユーザによる知見の蓄積や再利用を可能とするプロトタイプシステムの開発 	<ul style="list-style-type: none"> ・実証実験によるシステムの評価 ・複数人による情報共有のための枠組み ・人手によるメタデータ付与における負荷
本研究	<ul style="list-style-type: none"> ・実証実験によるシステム検証 ・計算機によるメタデータ自動付与環境の構築 ・複数研究者による情報共有のためのメタデータ設計および環境の構築 	

図 2-20 本研究の位置づけ

第3章 提案手法

3.1 はじめに	35
3.2 歴史研究プロセスの課題と必要要件の整理	35
3.2.1 歴史研究プロセスの課題	35
3.2.2 課題解決に向けた必要要件の定義	37
3.3 提案する歴史研究環境	39
3.4 セマンティック・ウェブ技術を用いた史料管理手法	41
3.4.1 メタデータフィールドの設計	42
3.4.2 オントロジーを用いた概念の体系化	46
3.4.3 セマンティック・ウェブ技術による有用性	47

3.1 はじめに

本研究では歴史研究のプロセス改善を目的とし、電子化された史料の有効活用を可能とする研究環境を構築する。したがって本章では、まず歴史研究プロセスとその課題について整理し、必要な機能や環境などの必要要件の洗い出しを行う。次にこれらの要件に基づいた新しい歴史研究環境を提案し、最後にその実現に向けたセマンティック・ウェブ技術による史料情報の記述手法について述べる。

3.2 歴史研究プロセスの課題と必要要件の整理

3.2.1 歴史研究プロセスの課題

三浦[39]は計算機によって歴史研究を支援するシステムの開発を行った。彼は歴史学者へのインタビューなどを通じて、歴史研究におけるプロセスを整理し、計算機によって支援可能な問題の洗い出しを行っている。そこでは、以下に示す研究プロセスが一つの典型例だとしている。

「歴史学では現存する不完全なデータ（史料）を歴史学者の主観を基に分類し、その分類されたデータを時間の流れといった関係から整理し、独自の仮説を構築し、さらにそれを裏付けるデータを探して検証を行っていく」

また彼は史料の特徴と、その特徴によって歴史研究を困難としている点を整理している。具体的には図 3-1 に示すように、「史料収集の困難さ」、「収集した史料から必要な情報のみを抽出することの困難さ」、「抽出された情報を俯瞰し有用な知見を得ることの困難さ」の三点を挙げている。

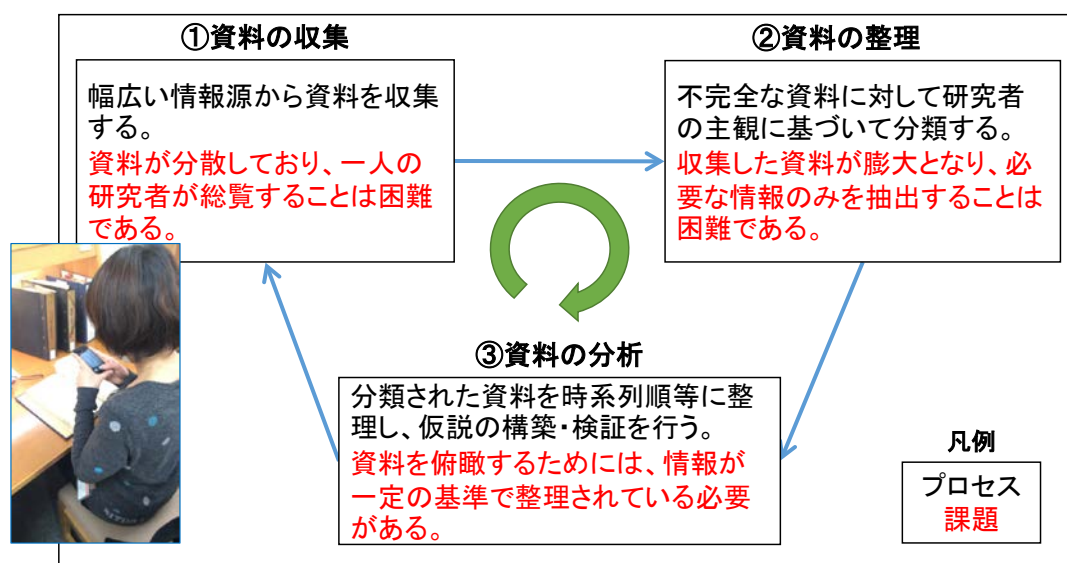


図 3-1 歴史研究プロセスの課題点の整理

ここでは三浦がまとめている史料の収集・整理・分析プロセスにおける各課題点について述べ、さらに筆者が行った複数の歴史研究者へのヒアリング結果に基づいて、その具体例を示す。

「史料収集の困難さ」は史料が各地に分散されていて、一つの事象に関する史料を一人の研究者が総覧することは容易でないと述べられている。具体的な史料の収集プロセスとしては、史料を保有している文書館や史料館に赴き、必要な史料をデジタルカメラ等で撮影し、各研究者の PC などの媒体に保存する。またこれらのプロセスは基本的に歴史研究者が一人で行っていることが確認され、この史料収集プロセスに多大な時間を要していることが分かった。

「収集した必要な情報のみを抽出することの困難さ」は、史料が一か所に収集された場合においても、その総数は膨大であり、また編集形式が一定でないため、それら史料群から各研究者の目的に応じて必要な部分のみを抽出するには史料を通読しつつ内容に基づいた検索を行わなければならないことに起因すると述べている。具体的には収集した史料を一つ一つ閲覧し、研究目的に応じた情報を Microsoft Excel 等のツールで管理していることがヒアリングから確認された。

「抽出された情報を俯瞰し有用な知見を得ることの困難さ」は、多大な労力をかけて関心ある情報を抽出した場合においても、それを俯瞰することができなければ一般

的知見を引き出すことは困難であることを意味する。抽出した情報を俯瞰するためには、それらの情報が一定の基準のもとで整理されなければならないと述べている。実際、各研究者の調査結果は記述形式が定まっておらず、他の研究者が容易に理解できる形式で管理されていないことがヒアリングから確認できた。また従来情報の整理に利用されてきたノートやカード等の紙のメディアでは限られたタイプの情報しか載せられず、また史料とは独立に管理されるため、整理された史料数が膨大になると、全体を俯瞰することが難しくなる点を課題として挙げている。

3.2.2 課題解決に向けた必要要件の定義

3.2.1 の各プロセスにおける課題解決に向けた必要要件を定義する。

3.2.2.1 史料の収集

史料の収集プロセスに関して、デジタルアーカイブで公開されている史料の活用を考える。デジタルアーカイブはインターネットを利用することで、時間的・空間的制約を受けずに史料にアクセスすることが可能である。しかし 2.4 で述べたように各デジタルアーカイブは史料の閲覧機能のみを提供しており、それら史料を研究者の目的に応じて分類する等の環境は整えられていない。よってクラウド上に各研究者が史料や調査結果を蓄積することが可能な環境を構築し、デジタルアーカイブで公開されている史料の URL を参照することで、各研究者のリポジトリに史料を収集可能とする。これにより、デジタルアーカイブで公開されている幅広い史料を分析対象とすることができる。一方、研究者が必要とする史料がすべてデジタルアーカイブ上で公開されているとは限らない。したがってこれまで歴史研究者が行ってきたように、紙媒体の史料も収集する必要がある。この点については複数の研究者がクラウド上に構築されたリポジトリを共有し、複数の研究者が共同で史料を収集し、撮影された画像史料等を登録する。これにより各研究者は他の研究者が収集した史料等を分析対象とすることが可能となり、膨大な史料を収集する際の各研究者の労力の軽減や幅広い対象からの史料収集が期待できる。

3.2.2.2 史料の整理

史料の整理プロセスについても複数研究者が共同で史料の整理を行う環境が必要である。この結果、複数研究者で史料内容を確認するプロセスを分担することができる。

このプロセスの分担では、予め複数研究者で抽出する情報を共有しておく必要がある。また歴史研究者へのヒアリングの結果、歴史研究においては各研究者が得た知見は研究の独自性につながるという観点から共有する情報の取捨、および公開範囲を制限することが必要であることを確認した。さらに他の研究者が蓄積した情報はその信憑性等から誰による情報なのかといった情報の出所を明示化する必要があることも確認できた。よって複数研究者による情報共有を行う環境の構築には、これらの点を考慮した史料管理手法が必要となる。加えて人手による史料の整理には限界があるため、電子化された史料の利点を活用し、画像処理やテキストマイニングといった計算機による情報の抽出を可能とする環境も必要である。これにより人手による整理の労力軽減や、人手では気づけなかった新たな発見を支援することが期待される。

3.2.2.3 史料の分析

史料の分析プロセスでは史料を俯瞰するために、情報が一定の基準で整理され、それらが再利用可能な形であることが必要である。事実、歴史研究者へのヒアリングから、多くの歴史研究者が史料と調査結果といった情報を別の場所で管理していることが確認されている。よって史料とその調査結果等が紐づけられ、相互に利用可能な管理手法が必要となる。さらに蓄積された史料が膨大となった際には、人手による俯瞰的な分析には限界がある。したがって蓄積された調査結果に対する計算機を利用した分析を可能とする環境も必要となる。この結果研究者に新たな視点を与え、研究における仮説の立案や検証の支援を行うことが期待される。

3.2.2.4 歴史研究のプロセス改善に向けた必要要件の定義

これらの考察から既存の課題を解決し、歴史研究プロセスを改善するため、本研究では以下に示す環境や機能を必要要件として定義する。

- 複数研究者による史料の収集や整理、分析を可能とする環境
- 史料整理や分析時に計算機による処理が可能な環境
- 史料と調査結果が紐づけられた形での史料情報の記述手法
- 複数研究者によって蓄積された情報の公開範囲の設定や出所の明示化を可能とする史料管理手法

3.3 提案する歴史研究環境

3.2 で挙げた必要要件の実現に向けて、本研究では図 3-2 に示す歴史研究環境を提案する。提案する研究環境はクラウド上に史料やその調査結果を蓄積する DB を構築する。そしてこの DB を歴史研究者やアーキビスト、情報技術者といった複数研究者で共有することで、複数研究者による情報の収集、整理、分析が可能となり、協調的な研究環境を構築する。なおアーキビストとは新たに入手した収集物に対して、史料の年代の特定や説明書きの追加を行うことで、史料に書誌情報を付与する業務を担う人物である。デジタルアーカイブ等で公開される史料は彼らが付与する情報に基づいて検索されるため、彼らの業務を支援することも歴史研究プロセスの改善に必要である。

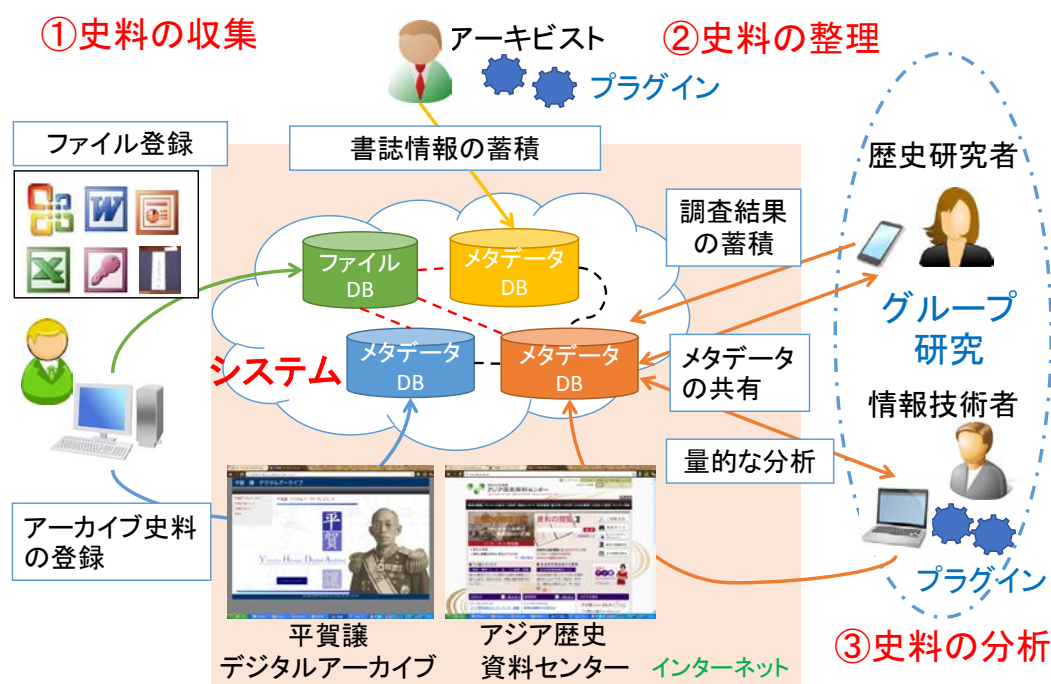


図 3-2 提案する歴史研究環境

収集プロセスでは、研究者が PC 等で管理している史料やデジタルアーカイブで公開されている史料をクラウド上に構築された DB に登録する形で行う。整理プロセス

では、収集した史料から得た調査結果を研究者が蓄積する他、画像処理技術等を用いて機械的に情報を抽出する。分析プロセスでは、DB に登録された書誌情報や調査結果を利用することで、研究目的に応じた史料の分類や比較を行う。

DB に蓄積される史料はセマンティック・ウェブ技術によって管理することで、調査結果をメタデータとして史料に付与し、史料と史料情報が紐づけられた形での管理が可能となる。これにより調査結果はメタデータフィールドとバリューのペアとして記述されるため、情報の記述形式の統一化が実現できる。さらに蓄積された調査結果等の公開範囲や登録者に関する情報についてもメタデータとして記述することで、調査結果を複数研究者で共有する際の情報の明示化が可能となる。これらの実現手法の詳細については次節 3.4 で述べる。

またプラグイン形式で史料情報を機械的に抽出可能な環境を構築することで、メタデータとして蓄積された調査結果等の定量的な分析を可能とし、蓄積された史料の傾向や特徴の抽出といった俯瞰的な分析を支援する。

3.4 セマンティック・ウェブ技術を用いた史料管理手法

セマンティック・ウェブとは 2.2.2 で説明したように、Web ページに記述された内容について、それが何を意味するかを表すメタデータを一定の規則に従って付与し、情報を構造化する構想である。この技術を応用し、歴史研究者が史料分析から得た知見を史料のメタデータとして付与することで、それらの情報が構造化され、統一した記述形式による調査結果の管理や利用が可能となる。さらにそれらの情報の公開範囲や付与した人物の情報をメタデータとして明示化することによって、複数研究者間での情報共有に必要な要件を実現する。さらにオントロジーを用いて各研究者の暗黙知を体系化することによって、他の研究者との知識共有や計算機による自律的な理解が可能とする。セマンティック・ウェブ技術を用いた史料管理手法の例を図 3-3 に示す。

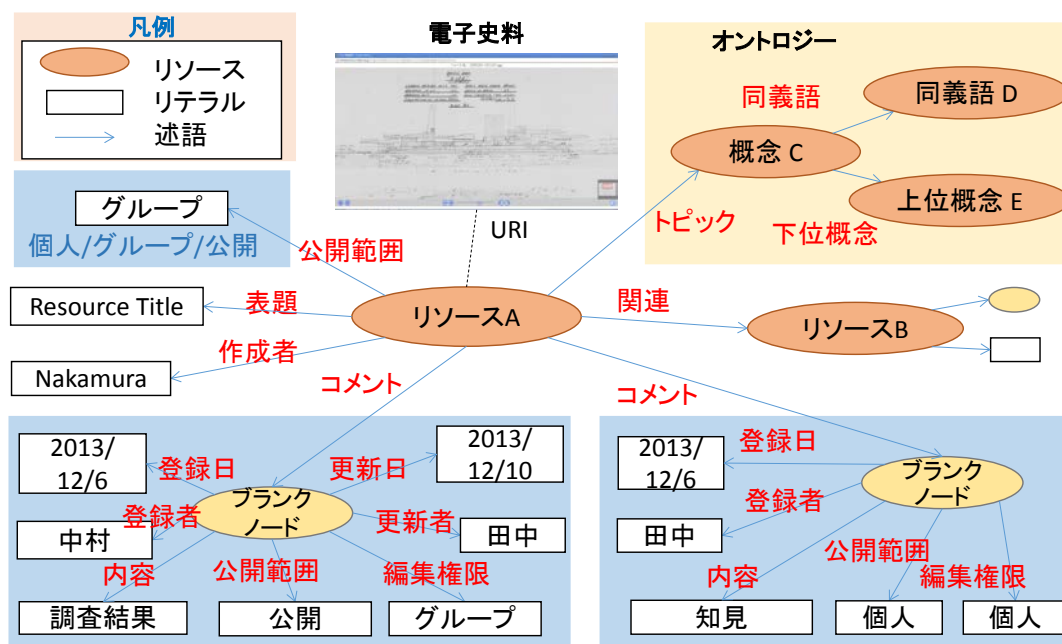


図 3-3 セマンティック・ウェブ技術を用いた史料管理手法例

各史料は URI によって識別される RDF モデルによって記述され、そのメタデータとして書誌情報や調査結果を管理する。さらに各メタデータの公開範囲や編集権限、登録者や更新者といった情報を記述するために、ブランクノードを導入した記述手法を提案する。例

例えばメタデータの公開範囲を個人、グループ、全体から選択することで、史料に付与された各メタデータについて他の研究者による閲覧範囲を制限できる。またメタデータの更新日や更新者情報を付与することで、それがいつ誰によって更新されたのか等の情報を明示化する。これを実現するためのメタデータフィールドの設計については 3.4.1 で述べる。

またリソースに関する概念の定義を図中右上に示すようにオントロジーとして定義することによって、研究者の暗黙知の体系化や計算機による史料の自律的な関連付け等が可能となる。これについては 3.4.2 で詳細を説明する。

3.4.1 メタデータフィールドの設計

メタデータを記述するためのメタデータフィールドの設計について述べる。本研究では史料を管理するためのメタデータと、調査結果等のユーザの知見を記述するためのメタデータの二種類のメタデータフィールドを定義した。

3.4.1.1 史料管理のためのメタデータフィールド

史料を管理するために定義したメタデータフィールドの一覧を表 3-1 に示す。

表 3-1 史料管理のためのメタデータフィールド

ID	日本語名	Property
e1	表題	dc:title
e2	ファイル名	dc:source
e3	登録者	dc:publisher
e4	更新者	kd:modifier
e5	URL	dc:identifier
e6	拡張子	dcterms:medium
e7	参照権限	dcterms:accessRights
e8	更新日	dcterms:modified
e9	登録日	dcterms:created
e10	リソースID	kd:resourceId
e11	平賀DA-ID	kd:hiragaId
e12	サムネイルURI	kd:thumbnailUri

これらのメタデータフィールドの定義について、ウェブや文書の作者、タイトル、作成日といった書誌的な情報をメタデータとして記述するためのボキャブラリを定めている「Dublin Core[40]」を利用した。なお、表中のフィールド ID「e5: URL」は、ウェブ上で公開されている史料について、RDF 表現において各史料の識別子として与えられる URI と区別するために使用する。

加えて、「Dublin Core」では提供されていない語彙を定義するために「kd (<http://kashiwade.org/2012/09/kd#>)」という名前空間を独自に定義した。具体的には「更新者 (kd:modifier)」「リソース ID (kd:resourceId)」「平賀 DA-ID (kd:hiragaId)」および「サムネイル URI (kd:thumbnailUri)」である。更新者は対象とする史料のいずれかのメタデータを最後に修正したユーザ名を記述する。リソース ID はファイル DB に登録したファイルとメタデータ DB の URI を紐づけるためのメタデータである。また平賀 DA-ID は平賀譲デジタルアーカイブで各史料に付与されている ID を示し、これを参照することでアーカイブ内の史料との紐づけを行う。サムネイル URI は Microsoft Office や画像ファイルを登録する際、前者では PDF、後者では低解像度の画像ファイルがそれぞれサムネイル画像としてファイル DB に追加登録され、それら

のサムネイル画像と登録した史料との紐づけを行うためのメタデータである。これらのメタデータは各史料に一つずつ付与されることによって史料を管理する。

3.4.1.2 ユーザの調査結果等を記述するためのメタデータフィールド

ユーザの調査結果から得られた知見等を記述するために定義したメタデータフィールドの一覧を表 3-2 に示す。

表 3-2 ユーザの調査結果等を記述するためのメタデータフィールド

ID	日本語名	Property
o1	作成日	dc:date
o2	キーワード	dc:subject
o3	作成者	dc:creator
o4	文書種類	kd:type
o5	カード目録	kd:card
o6	カテゴリ	kd:category
o7	ノート	kd:note
o8	グループ	kd:group

このメタデータフィールドを定義するにあたり、「Dublin Core[40]」をベースとして使用した。これらのメタデータフィールドとしては、「作成日 (dc:date)」「キーワード (dc:subject)」および「作成者 (dc:creator)」が該当する。またその他のフィールドとしては「平賀譲デジタルアーカイブ[10]」で使用されている項目を利用した。具体的には「図面」や「メモ」等が記述される「文書種類」、アーカイブ史料の目録を記述する「カード目録」、および戦艦名や論文、報告書といった平賀譲デジタルアーカイブで用いられている分類を記述する「カテゴリ」が該当する。さらに表中のフィールド ID「o7：ノート」は各ユーザが調査から得た知見等をメモするための役割を持ち、登録するメタデータが具体的なフィールドに該当しない場合に一時的に用いるフィールドである。

なお本項で定義したユーザの調査結果を記述するためのメタデータフィールドはあくまで暫定的なものであり、これらのフィールド以外にも各研究者が自由にフィールドを追加できるようにした。これは各研究者の研究目的に応じて分類基準が異なり、調査結果を記述するための項目は一意に決めることができないためである。一方でメタデータフィールドを自由に追加できることによってフィールド数が膨大かつ煩雑となる可能性がある。この問題に対しては今後研究者による利用ログなどから共通して利用されるフィールドを抽出し、歴史研究に必要な一般語彙として定義していくことを想定している。

さらにこれらのメタデータについては、複数研究者によるメタデータ管理を行うために表 3-3 に示すメタデータフィールドを定義した。

表 3-3 複数人によるメタデータ管理を行うためのメタデータフィールド

ID	日本語名	Property
a1	更新者	kd:modifier
a2	更新日	dcterms:modified
a3	内容	dc:description
a4	参照権限	dcterms:accessRights
a5	編集権限	dcterms:accrualPolicy
a6	登録者	dc:publisher

これらは表 3-2 で示したメタデータフィールドの値を記述するためのメタデータ（メタデータのメタデータ）である。書誌情報や調査結果に登録者や更新者、更新日等のメタデータをさらに付与することによって、メタデータの編集履歴等を異なるユーザ間で確認することができる。表中のフィールド ID「a4：参照権限」については、メタデータの公開範囲を「個人」「グループ」「全体」の三種類から選択する。た

例えば史料は他の研究者に公開するが、該当するメタデータは公開したくない場合には公開範囲を「個人」として設定する。これは特に歴史研究の場合において、得られた知見を研究の新規性・独自性の確保という目的で他のユーザに公開したくないケースに対応するためである。またフィールド ID「a5：編集権限」については、他のユーザによるメタデータの編集権限を「個人」「グループ」「全体」から選択する。たとえばこの項目を「個人」に設定することによって他のユーザによる上書きを禁止することができる。逆に信頼度の高いユーザとのグループにおける研究を行っている場合には、本項目を「グループ」に設定することで、同一グループのユーザによる更新を許可し、協調的なメタデータの蓄積が可能となる。

3.4.2 オントロジーを用いた概念の体系化

ここでは 2.2.2 で説明した RDF Schema および OWL を利用して、研究対象分野の概念の体系化を行う。本研究では対象とする概念の同義語および階層概念を RDF によって記述する。図 3-4 に RDF/XML 形式による記述例を示す。

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">

  <rdf:Description rdf:about="http://kashiwade.org/2012/09/kd/concept#長門">
    <rdfs:subClassOf rdf:resource="http://kashiwade.org/2012/09/kd/concept#戦艦"/>
    <owl:sameAs rdf:resource="http://kashiwade.org/2012/09/kd/concept#A123"/>
    <owl:sameAs rdf:resource="http://kashiwade.org/2012/09/kd/concept#A124"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://kashiwade.org/2012/09/kd/concept#戦艦">
    <rdfs:subClassOf rdf:resource="http://kashiwade.org/2012/09/kd/concept#軍艦"/>
  </rdf:Description>

</rdf:RDF>
```

図 3-4 同義語及び階層概念の記述例

これは「軍艦長門は暗号名として A123、A124 を持つ」という情報を体系化した例である。まず RDF リソースとして「長門」を定義し、RDF Schema で定義されている「subClassOf」という対象リソースの上位概念を示すメタデータフィールドを用いて、「長門」が「戦艦」の下位概念であることを示す。さらに「戦艦」は「軍艦」の下

位概念であるといった関係を同様に記述することによって、「長門」は「軍艦」の下位概念であるという関係を計算機が理解可能となる。

さらに OWL で定義されている「sameAs」という異なるリソースが同義であることを表現するメタデータフィールドを用いて、「A123」や「A124」といった概念が「長門」と同義関係にあることを定義する。これにより「A123」は「長門」の暗号名であり、実体としては同一であることを計算機に理解させることが可能となる。

このように RDF Schema や OWL を用いて概念間の関係をオントロジーとして体系化することによって、計算機が自律的に情報の意味を理解することが可能となる。活用例としては、「長門」という文字列を検索した際、計算機がオントロジーを参照することで自動的に「A123」や「A124」といった文字列を補完し、文字列「A123」等をメタデータとして含む史料についても抽出することが可能となる。これによりある研究者がオントロジーとして概念を体系化することで、これらの知識のない他の研究者も知識を共有することが可能となり、異なる分野の研究者が共同で研究する際の障壁を低減することができる。

3.4.3 セマンティック・ウェブ技術による有用性

従来の史料情報の管理方法とセマンティック・ウェブ技術で記述された史料情報の管理方法を比較した図を図 3-5 に示す。

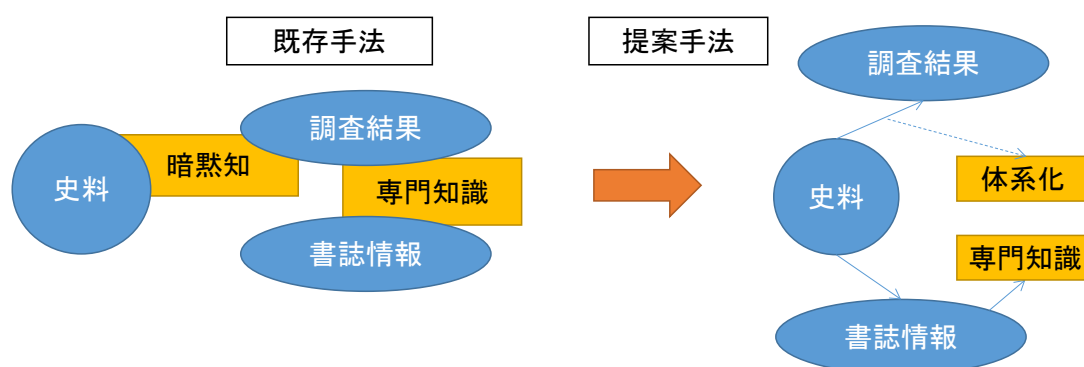


図 3-5 セマンティック・ウェブ技術で記述された史料情報の有用性

セマンティック・ウェブ技術を用いることによって、史料（ファイル）と調査結果（メ

タデータ)を紐づけた形での整理が可能となり、従来の史料と調査結果が別リポジトリで管理されていた問題を解決することができる。また従来の管理方法に起因する史料と調査結果の各研究者の暗黙知をノード間のリンクの形で体系化することにより、異なる研究者間の情報共有が可能となる。さらに各研究者の専門知識をオントロジーによって構造化することで、計算機による自律的な理解を支援することができる。

第4章 開発した歴史研究支援システム

4.1 はじめに	50
4.2 開発したシステムの概要	50
4.3 史料の収集	51
4.3.1 メタデータ登録機能.....	52
4.3.2 ファイル登録機能	53
4.3.3 メタデータ入力項目の詳細	54
4.4 史料の整理	55
4.4.1 メタデータ編集機能.....	55
4.4.2 オントロジーの編集機能.....	56
4.5 史料の分析	58
4.5.1 検索結果表示機能	58
4.5.2 検索機能	59
4.5.3 史料の並び替え機能.....	60
4.5.4 史料の比較機能.....	61
4.6 プラグインによるメタデータの分析機能.....	62

4.1 はじめに

本章では開発した歴史研究支援システムの概要について説明し、その後システムが提供する各機能について説明する

4.2 開発したシステムの概要

開発した歴史研究支援システムの概要図を図 4-1 に示す。

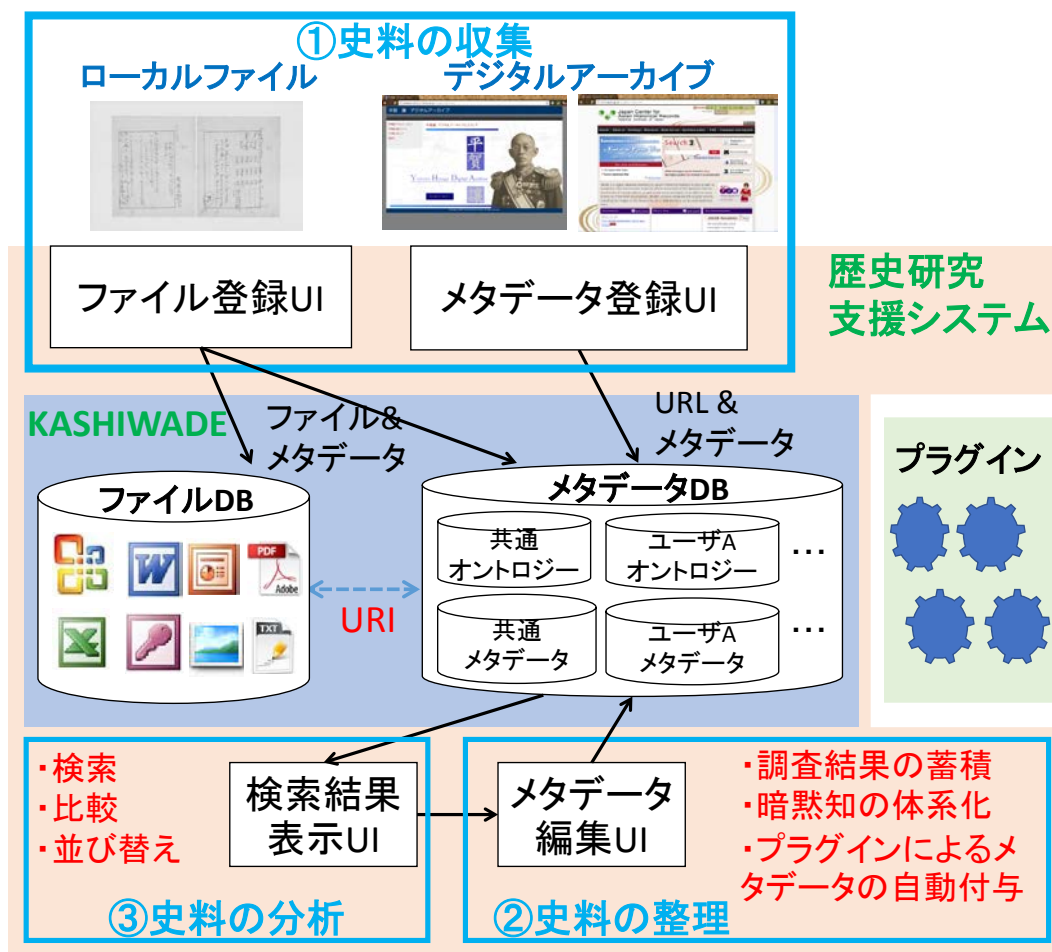


図 4-1 システム概要図

提案するシステムはウェブ上に構築されるウェブアプリケーションである。システムの実装には Java を使用し、システムは Web ブラウザ Google Chrome 上で動作確認を行った。

本システムは DB 構造やプラグインの実行環境に 2.3.3 で説明した文書自動分類システム「KASHIWADÉ」を用いている。その上に 3.2 で定義した歴史研究プロセスにおける「①史料の収集」「②史料の整理」および「③史料の分析」を支援するインタフェースを構築した。システムは登録されたユーザ毎の DB を持ち、基本的に各ユーザの DB を用いて歴史研究を行う。加えて後述する史料やメタデータに関する参照権限に基づいて、他のユーザの DB へアクセスすることで情報の共有を行う。システム上における歴史研究プロセスは、まず「①史料の収集」機能を用いて、研究者が画像史料やデジタルアーカイブで公開されている史料をシステムに登録する。その際、各史料には識別子として URI が割り当てられ、3.4 で示した RDF 構造に従ってメタデータ DB に登録される。画像史料等のファイルをシステムに登録する場合には、ファイルはファイル DB に格納され、先に割り振った URI によってメタデータ DB との紐づけがなされる。そして登録されたメタデータを利用して、「②史料の整理」「③史料の分析」を行う。「②史料の整理」機能では、人手によるメタデータの追加や更新、およびプラグインを用いた機械的なメタデータの付与を行うことで、調査結果の蓄積や史料の分類を行う。「③史料の分析」機能では整理されたメタデータを利用することで、史料の絞り込みや比較、並び替え等による分析を行う。これらのプロセスがシステム上で繰り返されることによって、歴史研究が行われる。

次にシステムが提供する各機能について、「①史料の収集」「②史料の整理」「③史料の分析」の三つに大別して説明する。

4.3 史料の収集

ここでは史料をシステムに登録するための機能について述べる。本機能はデジタルアーカイブによって公開されている史料を代表とする、安定した URL を持つ史料に登録する「メタデータ登録機能」、および PC やタブレット端末内に保存された画像史料等の電子ファイルを登録する「ファイル登録機能」の二つに分けて説明する。

4.3.1 メタデータ登録機能

図 4-2 に示すインタフェースを用いて、デジタルアーカイブで公開されている史料をシステムに登録する。



図 4-2 メタデータ登録 UI

左画面にはメタデータの登録インタフェース、右画面には史料を閲覧するための Web ブラウザを持つ。このインタフェースでは対象とする史料が公開されている URL、史料名および公開範囲が入力必須項目であり、他のメタデータを追加登録するためには「項目の追加」ボタンから追加する。URL を入力すると、右画面のブラウザにそのページ内容が自動的に表示される。図中ではアジア歴史資料センターで公開されている史料の URL を入力することによって、そのページ内容が右画面に表示されている例を示している。

メタデータの入力項目の詳細については 4.3.3 において後述する。

4.3.2 ファイル登録機能

図 4-3 に示すインタフェースを用いて、ユーザの PC やタブレット端末内にある電子史料をシステムに登録する。

The screenshot shows a web browser window with the URL `157.82.251.163/kashiwade/hiraga/uploadResources?lang=ja`. The interface is divided into two main sections.

Metadata Form (Top):

- Buttons: "項目の追加" (Add Item), "新規項目の追加" (Add New Item).
- Table Headers: 項目 (Item), 内容 (Content), 公開範囲 (Public Range), 編集権限 (Edit Permission), 更新者 (Updater), 編集 (Edit).
- Form Fields:
 - 作成年月日 (Creation Date): 2013 年 12 月 23 日
 - グループ (Group): 五国会議 (Gomoku Kaigi)
 - 公開範囲 (Public Range): [User Icon]
 - 編集権限 (Edit Permission): [User Icon]
 - 更新者 (Updater): nakamura
 - 更新 (Update) button
 - 編集権限 (Edit Permission): [User Icon]
 - 更新者 (Updater): nakamura
 - 追加 (Add) button

File Upload Section (Bottom):

- Buttons: "+ Add files...", "Start upload", "Cancel upload", "全選択" (Select All).
- File List:

Thumbnail	File Name	Size	Start	Cancel
	10010101-005_001.jpg	598.12 KB	[Start]	[Cancel]
	10010101-006_001.jpg	550.78 KB	[Start]	[Cancel]
	10010101-007_001.jpg	593.01 KB	[Start]	[Cancel]

図 4-3 ファイル登録 UI

本画面では画面上部にメタデータ登録インタフェース、画面下部には登録するファイルのリストが表示される。ファイルの登録は図中の「Add files」ボタンをクリックし、各種端末内のファイルを必要数選択する。選択後には図に示すように各史料のサムネイルが表示され、登録するファイルを確認することができる。なお本システムでは Microsoft Office や PDF、JPEG や PNG といった各画像形式のイメージファイルなど、あらゆる拡張子のファイルを登録することができる。

画面上部のメタデータ登録インタフェースで入力されたメタデータは、画面下部に表示されたファイル群に一括で付与される。したがってグループ等のある分類に基づいて選択された史料群をシステムにアップロードする場合には、その分類基準をメタデータとして複数史料に付与することができる。メタデータの項目の追加は 4.3.1 と同様に「項目の追加」ボタンから行う。またメタデータの一括登録の他、登録した各

史料の表題を修正する機能も提供する。

メタデータの入力項目の詳細については 4.3.3 において後述する。

4.3.3 メタデータ入力項目の詳細

4.3.1 のメタデータ登録機能、4.3.2 のファイル登録機能におけるメタデータの入力項目に関する説明を加える。メタデータ入力時の拡大図を図 4-4 に示す。








項目	内容	公開範囲	編集権限
URL	http://www.jacar.go.jp/DAS/meta/image_B06150179500?IS_KIND=SimpleSummary&IS_KEY_S1=%E4%BA%94%E5%9B%BD%E4%BC%9A%E8%AD%B0&IS_STYLE=default&IS_TAG_S1=InfoD&		
表題	巴里最高会議		
公開範囲			
作成年月日	2013 年 12 月 16 日		
閲覧者	<div> <input type="text"/> <ul style="list-style-type: none"> 澤田、幣原、埴原、武者小路 小川、幣原、埴原、武者小路 小川、幣原、芳澤、武者小路 澤田、幣原、埴原 澤田、幣原、埴原、武者小路、小村 澤田、幣原、埴原、武者小路、青木 澤田、幣原、芳澤、武者小路 澤田、幣原、芳澤、青木 </div>		  

図 4-4 メタデータ入力項目の詳細

図中の三行目は史料の公開範囲を指定する項目であり、「個人」「グループ」「全体」の三種類から選択する。これにより他ユーザによる史料の閲覧の可否を設定することができる。四行目以降はユーザが任意に登録するメタデータである。各メタデータは項目、内容、公開範囲、編集権限の四つの選択・入力項目を持つ。

項目はシステム内に登録済みのメタデータフィールドからプルダウン形式で選択す

る。

内容については書誌情報や得られた知見をテキスト形式で入力する。その際、図 4-4 に示すように選択した項目毎に、システム内に登録済みのメタデータの利用頻度上位 10 件が候補群として表示される。これによりメタデータを入力する労力の軽減や、語の表記揺れ等を吸収することができる。特にグループ名といった史料の分類を行うためのメタデータは複数の史料に対して同一のメタデータが繰り返し付与されることが多く、本機能がその労力軽減に寄与すると考えられる。

さらに公開範囲および編集権限については、表 3-3 で示したメタデータの「公開範囲」「編集権限」を指定する。ここではそれぞれ「個人」「グループ」「全体」の三種類からプルダウン形式で選択する。これにより史料に付与された各メタデータについても、その公開範囲や編集権限の設定が可能となる。

4.4 史料の整理

4.4.1 メタデータ編集機能

史料の書誌情報の追加や更新、調査から得られた知見の蓄積等は図 4-5 に示すメタデータ編集画面を用いて行う。

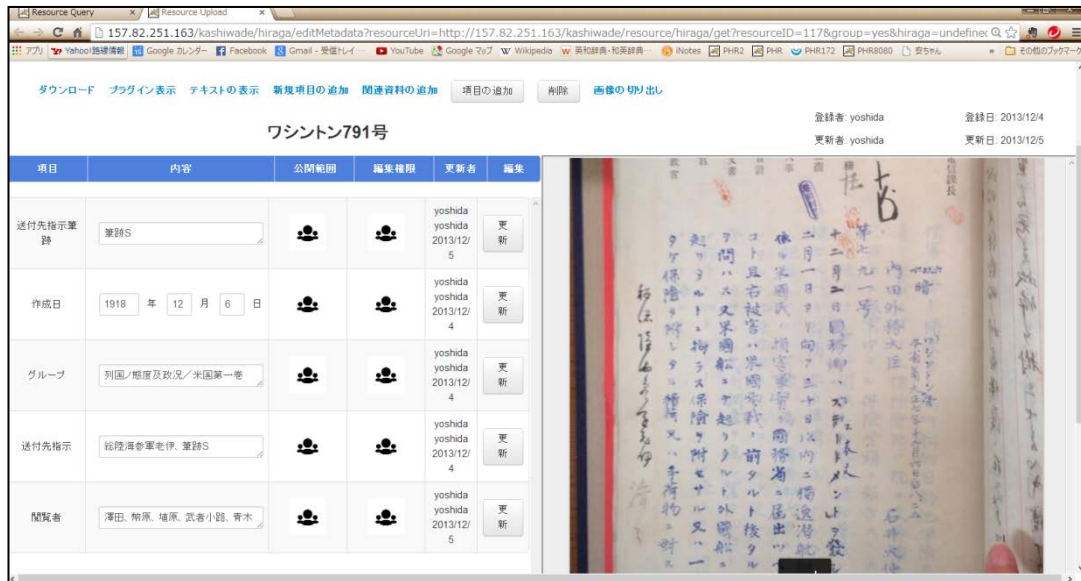


図 4-5 メタデータ編集インタフェース

この編集画面では 4.3 で説明した史料収集のためのインタフェースと同様に、史料が持つメタデータと史料が同画面に表示されるため、史料の閲覧と同時にメタデータの編集が可能である。また各メタデータは 4.3.3 で説明した項目を持つ。その際、図左部の各行の 5 行目に示される更新者という項目ではメタデータを登録したユーザ、メタデータを修正した最終更新者、およびその日付が表示される。これを参照することで、メタデータの更新履歴を各ユーザが確認することができる。

他のユーザが登録した史料を参照する場合には、メタデータの「公開範囲」「編集権限」の設定によって表示される項目が異なる。まず公開範囲が「個人」、つまり他ユーザへの公開を拒否している場合には、本画面にはそのメタデータが表示されない。次に編集権限が「個人」に設定されている場合、「更新」ボタンが非表示となり、他のユーザによるメタデータの上書きを許可しない。これらの機能によって複数研究者でメタデータを共有する際に必要となるメタデータの公開範囲や信頼性等の管理を行う。

4.4.2 オントロジーの編集機能

3.4.2 で説明したオントロジーによる各研究者の専門知識等の構造化は図 4-6 に示



本インタフェースで編集したオントロジは検索時の辞書として用いられ、検索画面で入力した検索語に関連する上位概念または同義語が検索クエリに自動で追加されることで、関連資料を含めた検索が可能となる。

4.5 史料の分析

4.4 でメタデータとして整理した史料情報に基づいて史料分析を支援する機能について、以下のそれぞれの機能について順に説明する。

- 検索結果の表示機能
- 史料の検索機能
- 検索結果の並び替え機能
- 複数史料の比較機能

4.5.1 検索結果表示機能

検索結果インターフェースを図 4-7 に示す。

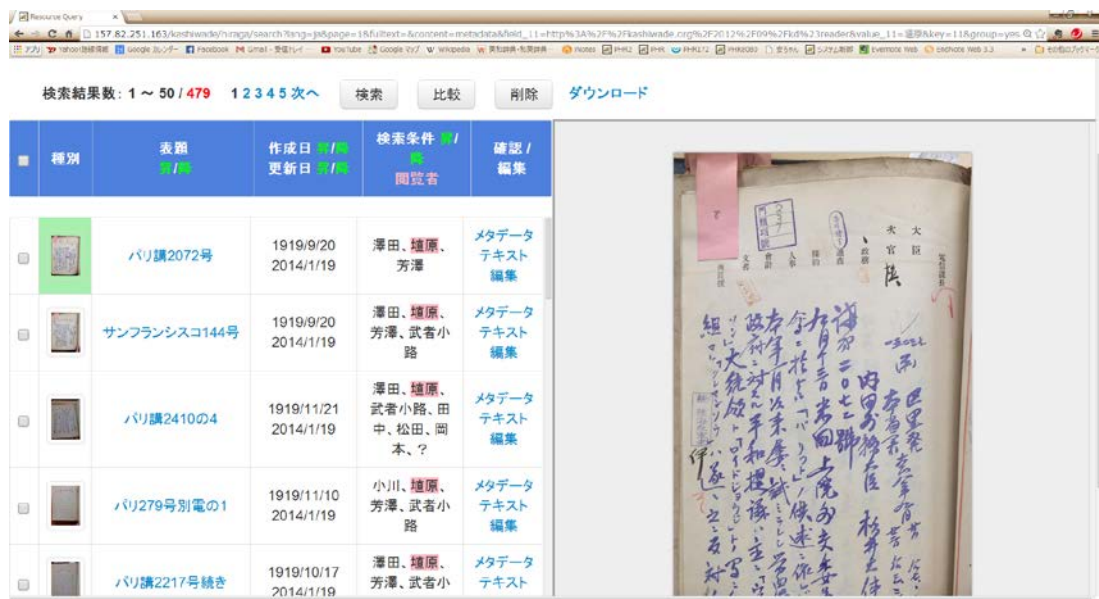


図 4-7 検索結果インターフェース

画面左部には検索結果一覧が表形式で表示される。各行には左から、史料の比較を行うためのチェックボックス、史料のサムネイル、表題や作成年月日といった主要なメタデータ、及び検索クエリとして使用したフィールドの内容が表示される。この一

覧画面を用いることによって、大まかな史料情報を確認することができる。より詳細な確認を行う場合には、各サムネイル画像をクリックすることによって画面右部に史料が表示され、拡大・縮小機能によって史料の確認作業を支援する。さらに史料が持つすべてのメタデータを確認する場合には、各行の右端にある「メタデータ」というリンクから、以下図 4-8 に示すインターフェースを用いて確認する。



図 4-8 メタデータの確認機能

4.5.2 検索機能

検索インターフェースを図 4-9 に示す。

キーワード検索	<input type="text"/>		
	<input checked="" type="radio"/> Metadata <input type="radio"/> Content <input type="radio"/> Both		
表題 ▼	<input type="text" value="ワシントン"/>	<input checked="" type="radio"/>	
検索対象範囲	<input type="checkbox"/> グループ  <input checked="" type="checkbox"/> 全体 		
表示結果数	<input type="radio"/> 20 <input type="radio"/> 50 <input checked="" type="radio"/> 100	表示順序	<input checked="" type="radio"/> 表題 <input type="radio"/> 作成年月日 <input type="radio"/> 更新年月日
検索方式	<input checked="" type="radio"/> AND <input type="radio"/> OR	正規表現	<input checked="" type="radio"/> 含める <input type="radio"/> 含めない
<input type="button" value="項目の追加"/>		<input type="button" value="検索"/>	

図 4-9 検索インタフェース

システム内の史料検索は主にメタデータ検索によって行う。メタデータ検索では史料の持つメタデータフィールド毎に検索語を入力して検索を行う。ユーザは検索の対象となるフィールドを一つまたは複数指定し、そこに検索語を入力する。図では「表題」フィールドに「ワシントン」という文字列を含む史料を検索している例である。フィールドを複数指定する場合には、図中左下の「項目の追加」ボタンを押下することで、入力項目を追加することができる。

その他に史料が持つメタデータに対して全文検索を行う「キーワード検索」や、検索対象とする DB の範囲の指定を行うことができる。検索対象範囲は公開範囲が「全体」である史料、「グループ」である史料、「個人」である史料の三通りから選択できる。

4.5.3 史料の並び替え機能

検索結果画面に表示される各メタデータフィールドには並び替えのためのリンクが

あり、各リンクをクリックすることによって表題や作成年月日などのフィールド別に昇順・降順に並び替えることができる。この並び替え機能によって、史料の分類や様々な角度からの分析を支援する。特に史料を作成年月日で降順、または昇順に並び替えることによって、史料の時系列に沿った分析を行うことができる。

種別	表題	作成日	登録者	確認 / 編集
	スズメ	1915/03 2013/12/5	yoshida	リサーチ リスト 編集
	バノ1254号03	1918/12/16 2013/12/18	yoshida	リサーチ リスト 編集
	バノ1254号05	1918/12/16 2013/12/18	yoshida	リサーチ リスト 編集
	バノ1482号	1915/03 2013/12/17	yoshida	リサーチ リスト 編集
	バノ158号	1918/01 2013/12/18	yoshida	リサーチ リスト 編集
	バノ1254号03	1918/12/16 2013/12/17	yoshida	リサーチ リスト 編集
	バノ1659号04	1918/12/16 2013/12/17	yoshida	リサーチ リスト 編集

種別	表題	作成日	登録者	確認 / 編集
	ワシントン179号	1918/12/6 2013/12/5	yoshida	リサーチ リスト 編集
	スズメ	1918/03 2013/12/5	yoshida	リサーチ リスト 編集
	ワシントン188号	1918/04 2013/12/4	yoshida	リサーチ リスト 編集
	バノ1254号	1918/03 2013/12/4	yoshida	リサーチ リスト 編集
	ワシントン299号	1918/11 2013/12/5	yoshida	リサーチ リスト 編集
	ワシントン207号	1918/03 2013/12/4	yoshida	リサーチ リスト 編集
	ワシントン209号	1918/03 2013/12/4	yoshida	リサーチ リスト 編集

図 4-10 史料の並び替え機能の例

4.5.4 史料の比較機能

さらに一覧画面の各検索結果にはチェックボックスが用意されており、比較機能を用いることでチェックされた史料を図 4-11 に示す同一画面で史料の比較を行うことができる。この機能を並び替え機能と組み合わせることによって、史料における時系列の変化の比較等を容易に行うことが可能となる。

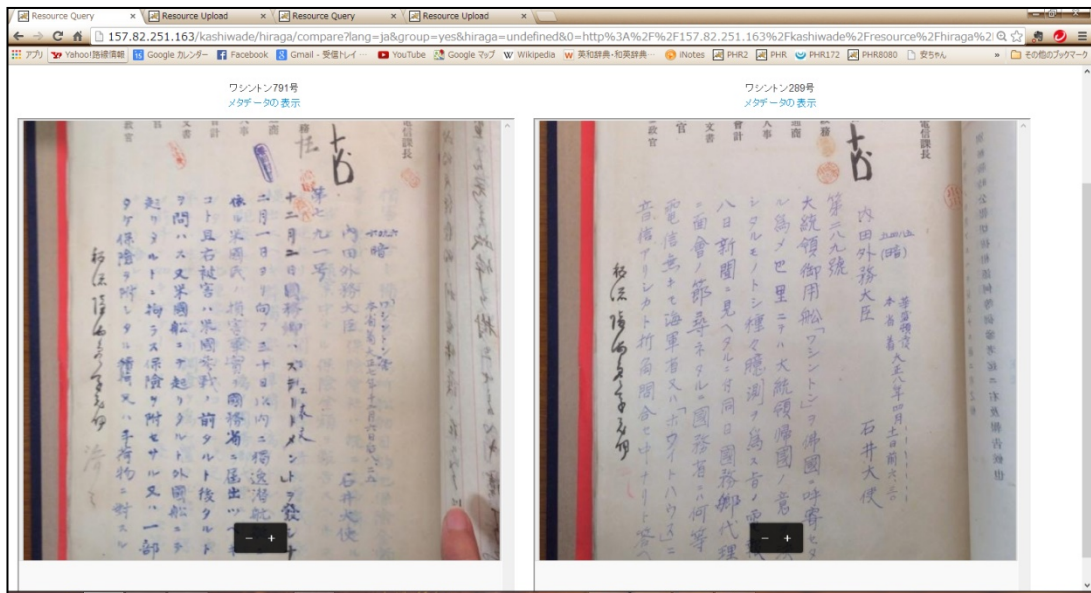


図 4-11 史料の比較機能

4.6 プラグインによるメタデータの分析機能

本システムは KASHIWADE 上に構築されているため、KASHIWADE のプラグイン機能を用いて、登録されたリソースやメタデータに対して機械的にメタデータを付与することができる。プラグインの実行環境の概要を図 4-12 に示す。

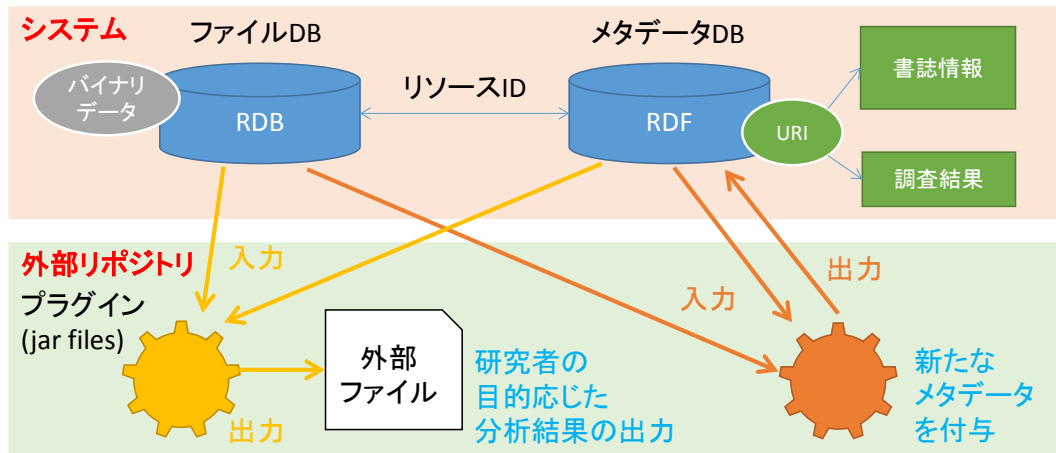


図 4-12 プラグインの実行環境

プラグインはサーバ上のシステムとは異なるリポジトリで管理され、システムは指定したプラグインを呼び出す形で実行する。各プラグインはシステムに登録されたリソース（バイナリファイル）やメタデータ（文字列）を引数とし、新たに抽出したメタデータの付与や外部ファイルへの出力を行う。プラグインはシステムから独立しているため、システムを改変することなく、目的に応じて自由に追加や修正することが可能である。システムからプラグインを呼び出すインターフェースを図 4-13 に示す。

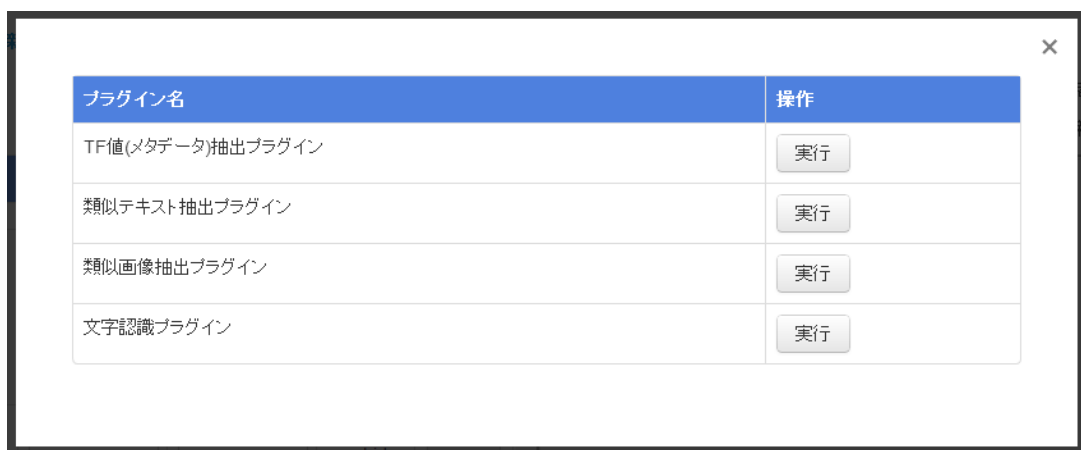


図 4-13 プラグインの実行インターフェース

本インタフェースは 4.4.1 で説明したメタデータ編集インタフェースから呼び出すことができ、対象とする史料情報を引数として実行する。図中の各プラグインを選択して実行することによって、プラグイン内に記述されたアルゴリズムに基づいて新たなメタデータの自動付与等を行う。

新たにメタデータを付与する例として、画像から文字列を抽出する文字認識プラグインを開発・実行する際には、システムに登録された画像資料のバイナリデータを引数として、プラグイン内のアルゴリズムにしたがって文字列を抽出し、その文字列を新たにその画像のメタデータとしてシステムに登録できる。またリソースだけではなくメタデータを引数としてプラグインを実行できる。たとえば先の例で各画像から抽出した文字列を格納したメタデータを引数とし、それらに形態素解析を行うプラグインを実行することで、画像中の形態素の出現頻度情報を新たにメタデータとしてシステム内に蓄積できる。さらにこの処理を複数画像に行い、機械的に付与した形態素の頻度情報を複数画像間で比較することによって、画像のテキスト情報に基づいた類似画像情報等を新たに抽出することができる。このようにシステム内に蓄積されたメタデータを再利用するプラグイン機能を用いることによって機械的にメタデータを付与することが可能となる。機械的に付与されたメタデータの正当性については最終的にユーザによる確認作業が必要ではあるが、人手で新たにメタデータを付与する作業と比較した場合、その労力の軽減が期待できる。

また外部ファイルへの出力としては、システム内の蓄積されたメタデータに対してプラグインを実行し、量的な分析結果を出力する例が考えられる。表 4-1 にシステム内の蓄積されたメタデータを Microsoft Excel 形式で出力した例を示す。

表 4-1 システムに蓄積されたメタデータの Microsoft Excel 出力例

リソース ID	送付先指示	登録者	更新日時	グループ	閲覧者	送信者	更新者	表題	ノート	登録日	作成日	送付先指示筆跡
126	総陸海参軍老伊、筆跡S	yoshida	20131204	列国ノ態度及政況ノ米国第一巻	小川、幣原、埴原、武者小路		yoshida	パリ講856号	隅に謎文字	20131204	19190505	
127	総陸海参軍老伊	yoshida	20131205	列国ノ態度及政況ノ米国第一巻	澤田、幣原、埴原、武者小路	松井大使	yoshida	パリ講1015号		20131204	19190516	筆跡S
128	総陸海参軍老伊	yoshida	20131205	列国ノ態度及政況ノ米国第一巻	澤田、幣原、埴原	石井大使	yoshida	ワシントン370号		20131205	19190521	筆跡S
129	総陸海参軍老伊、筆跡S	yoshida	20131205	列国ノ態度及政況ノ米国第一巻	澤田、幣原、埴原、武者小路	石井大使	yoshida	ワシントン386号		20131205	19190525	

図中の一行目はメタデータフィールドであり、第一列目はシステム内の各資料に割り当てられたIDである。このように各史料に対してメタデータフィールドとメタデータバリューのペアによって史料情報や研究者による調査結果が記述されるため、一定の記述形式が一定の基準を満たしながらシステム内に蓄積される。このため、メタデータフィールド別のメタデータバリューの出現頻度やある基準（文字列に形態素解析を適用後の特徴ベクトルなど）における史料間の類似度等を算出するプラグイン等を個別に開発することによって、研究者の視点に立った分析を機械的に行うことが可能となる。本機能については第6章の実験において具体的に述べる。

第5章 ケーススタディ：長門型戦艦の設計履歴 分析

5.1 はじめに	67
5.2 ケーススタディ内容	67
5.3 分析対象史料	69
5.4 ケーススタディにおける前処理	71
5.4.1 オントロジーの追加	71
5.4.2 局所エッジヒストグラムを用いたインデックスの作成	71
5.5 ケーススタディの流れ	72
5.6 史料の収集	74
5.7 史料の整理	74
5.8 史料の分析	76
5.9 分析結果を利用した史料の再検索	77
5.10 ケーススタディ結果	78
5.11 ケーススタディのまとめ	79

5.1 はじめに

本ケーススタディでは既存の歴史研究事例を開発したシステム上で行い、提案手法および開発したシステムの検証を行う。既存の歴史研究事例として、中村[38]が対象とした長門型戦艦の設計履歴分析におけるケーススタディを対象とする。特にセマンティック・ウェブ技術を用いた史料情報の記述手法、メタデータに基づいた史料の分類や比較機能、プラグインを用いた計算機によるメタデータの自動付与機能の有用性について検証する。

なお本ケーススタディの実施者は筆者一名である。

5.2 ケーススタディ内容

本ケーススタディは、書籍「八八艦隊計画」[13]で紹介されている研究を参考とした。八八艦隊計画の戦艦を担った軍艦である、長門型戦艦「長門」（以下、長門）と加賀型戦艦「加賀」（以下、加賀）について、両者間の煙突本数の減少に関する原因追究を行っている。両戦艦は同系統かつ近い時期に作成されたにも関わらず、図 5-1 に見るように煙突本数が前者は二本であるのに対して、後者は一本であり、この原因が長い間明確でなかった。書籍ではこの原因を搭載汽缶数の減少としており、当時期に搭載されていた汽缶である「ロ号艦本式缶」の技術革新が起こり、一缶当りの容量が増加したことがその原因であると記述されている。実際、この研究においては平賀譲デジタルアーカイブで公開されている史料が多く用いられている。よって本ケーススタディでは、この原因追究に関するシナリオを本システム上で再現し、煙突本数の減少に関する報告書を見つけることを目的とする。

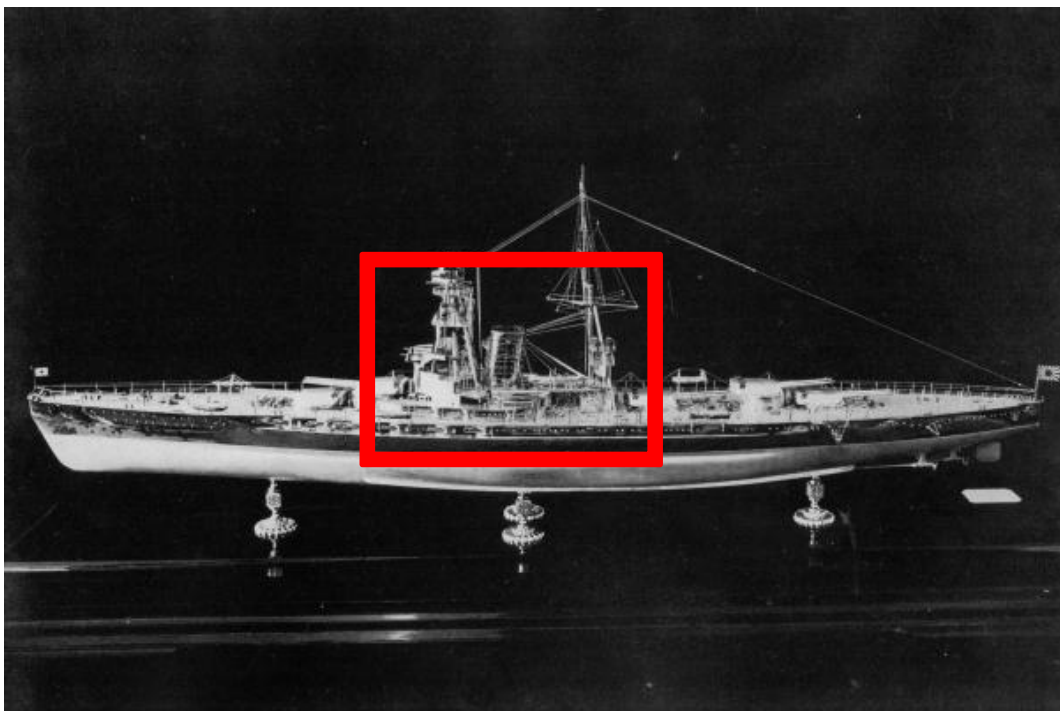
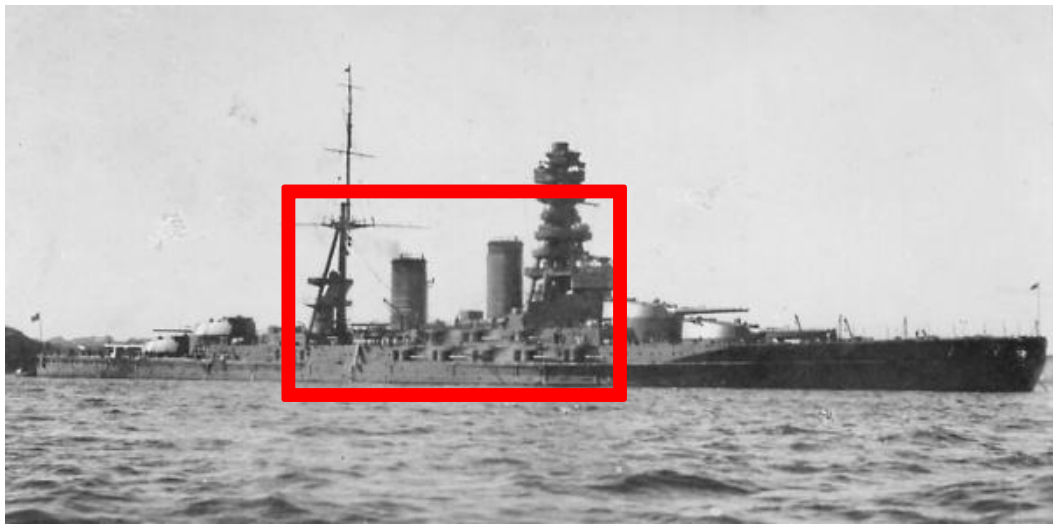


図 5-1 戦艦「長門」[41] (上) と「加賀」[42] (下)

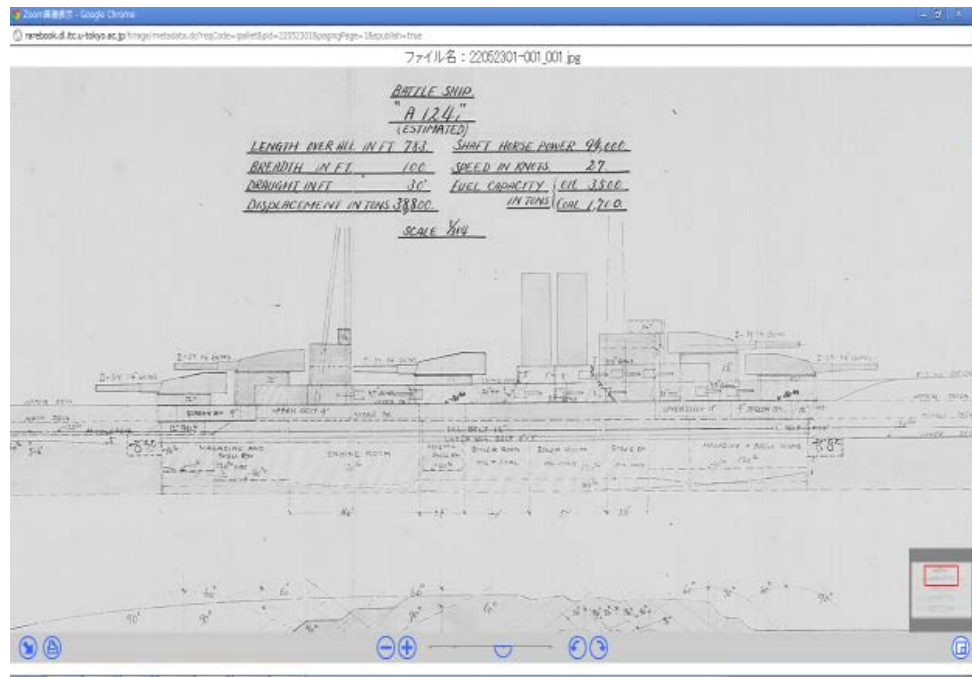
5.3 分析対象史料

本ケーススタディで対象とする史料は平賀譲デジタルアーカイブで公開されている史料 30,058 件を対象とする。これらの各画像史料に平賀譲デジタルアーカイブで付与されている書誌情報を付与する。この書誌情報を表 5-1 に示す。

表 5-1 平賀譲デジタルアーカイブで使用されている書誌情報例

標題	文書種類	文書記述形式	枚	年月日(新)	備考	ID	カテゴリ
製艦費ノ減少ニ就テ	意見書	海軍野紙タイプ印刷	29	1929/2/28		10010101	E0100
なし	人名一覧	万国工業会議用箋タイプ	1			10020101	H0101
昭和四年十月廿三日 万国工業会議会議委員長 稲田三之助 平賀譲殿	平賀宛書簡	万国工業会議用箋ペン書き	2	1929/10/23		10020201	H0101
EXTRA EDITION OF OFFICIAL BULLETIN THE WORLD ENGINEERING CONGRESS TOKYO 1929	万国工業会告示	英文印刷物	1	1929/10/31		10020301	H0101
昭和四年十月廿七日 万国工業会議会議委員長 稲田三之助 平賀譲殿	平賀宛書簡	万国工業会議用箋タイプ	2			10020401	H0101
部会理事(日本人)	理事一覧表	タイプ(ペン書き記入有り)	1			10020501	H0101
昭和三年八月 日 倭万国工業会議論文委員長殿 論文委員 近藤基樹・末廣恭二・山内不二雄・平賀譲	平賀(等)発書簡	タイプ	1	1928/8/1	1928/8	10020601	H0101

平賀譲デジタルアーカイブでは表題、文書種類、文書記述形式、作成年月日といった項目別に、その値について研究者がわかる範囲で整理されている。この書誌情報を用いた各史料に関する RDF の記述例を図 5-2 に示す。



```

<rdf:Description rdf:about="http://172.16.53.65:8080/kashiwade/resource/hiraga/get?resourceID=da22052301">
  <j.0:modified>20131210</j.0:modified>
  <dc:title>BATTLE SHIP "A124" (ESTIMATED) </dc:title>
  <dc:identifier>
    http://rarebook.dl.itc.u-tokyo.ac.jp/hiraga/metadata.do?reqCode=ipallet&pid=22052301&pagingPage=1&spub
  </dc:identifier>
  <j.0:created>20111027</j.0:created>
  <dc:date rdf:nodeID="A5787"/>
  <kd:thumbnailUri rdf:resource="http://172.16.53.65:8080/kashiwade/resources/images/thumbnail/da.jpg"/>
  <kd:docType rdf:nodeID="A6176"/>
  <kd:note rdf:nodeID="A13378"/>
  <dc:source>22052301.tmp</dc:source>
  <kd:catalog rdf:nodeID="A6164"/>
  <kd:hiragaId>22052301</kd:hiragaId>
  <kd:resourceId>da22052301</kd:resourceId>
  <kd:category rdf:nodeID="A5862"/>
  <j.0:accessRights>public</j.0:accessRights>
  <kd:modifier>admin</kd:modifier>
  <j.0:medium>TMP</j.0:medium>
  <dc:publisher>admin</dc:publisher>
  <kd:group rdf:nodeID="A5810"/>
</rdf:Description>

```

図 5-2 平賀譲デジタルアーカイブ史料の RDF 記述例

5.4 ケーススタディにおける前処理

5.4.1 オントロジーの追加

本ケーススタディを行うにあたり、4.4.2 で説明したオントロジー編集インタフェースを用いて、長門に関する同義語等をオントロジーに追加する。長門型戦艦は表 5-2 に示すように「A112」や「A115」といった暗号名を持ち、平賀譲デジタルアーカイブでもこれらの表記が長門に関する史料において多く用いられている。したがってこれらの暗号名に加えて英語表記である「NAGATO」を「長門」の同義語としてオントロジーに記述する。

表 5-2 長門型戦艦の暗号名一覧[43]

Items	常備 排水量 T	速力 kt	出力 shp	垂間長ft × 水線幅ft	缶室区画 数 Mix/Oil	主砲 口径× 門数	バー ベット in	水線 装甲 in	甲板 防御 in	水中 縦隔壁 in
A112 長門型	33,800	2026/1/2	80,000	660-1/2 × 95	M×1+O × 3	41cm×8	12	12	3	3
A115	34,700	30	113,000	735×97		同上	同上	9(傾斜)	同上	同上
A116	36,700	同上	120,000	同上		同上	同上(煙路防御強化)			
A117	37,100	32	145,000	780×97		同上	同上			
A118	39,400	同上	150,000	800×97		同上	同上(煙路防御強化)			
A119	38,400	30	120,000	790×97		41cm×10	同上			
A120	40,400	同上	124,000	800×97		同上	同上(煙路防御強化)			
A121	40,600	32	150,000	830×97		同上	同上			
A122	43,500	同上	157,000	875×97		同上	同上(煙路防御強化)			
A123	36,600	27	90,000	700×97		同上	同上(煙路防御強化)			
A124	37,200	同上	91,000	700×100		同上	12	12	3	3
A124'	38,800	同上	94,000	730×100		同上	同上(煙路防御強化)			
A125 陸奥変体	33,800	2026/1/2	80,000	660-1/2 × 102	M×1+O × 3	同上	12	12(傾斜)	3	3
A126	39,300	同上	88,000	700×100	同上	同上	同上	10(傾斜)	2002/3/4	同上

5.4.2 局所エッジヒストグラムを用いたインデックスの作成

本ケーススタディでは 2.2.3.5 で説明した局所エッジヒストグラムを用いて、対象とする画像に類似する画像を機械的に抽出し、それを史料の整理プロセスにおいて利用する。したがって本ケーススタディでは各画像から局所エッジヒストグラムを作成し、それをインデックスとして格納するプラグインを実装した。この実装には画像処

理ライブラリの一つである Lire[44]を利用した。このプラグインを利用し、システム内に登録された 30,058 件の史料から局所エッジヒストグラムに関するインデックスを予め構築する。この流れを図 5-3 に示す。

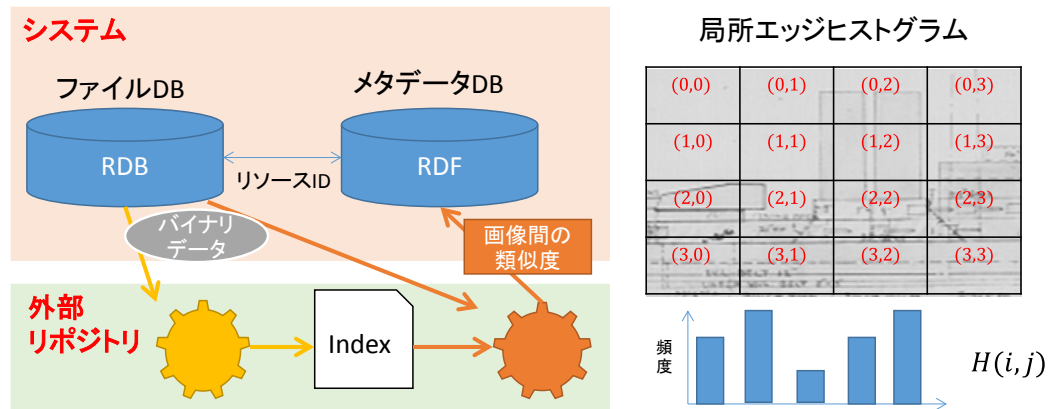


図 5-3 局所エッジヒストグラムを用いたインデックスの作成

図中の黄色で示されたプラグインがインデックスを作成するプラグインに該当する。この前処理で作成されたインデックスに対して、各画像情報の局所エッジヒストグラムに基づいて画像間の類似度を算出するプラグイン(図中橙色)を実行することによって、類似画像に関するメタデータを各画像に自動的に付与する。

5.5 ケーススタディの流れ

ケーススタディの流れを図 6-3 に示す。本ケーススタディは 3.2 で示した歴史研究プロセスに準ずる形で行う。

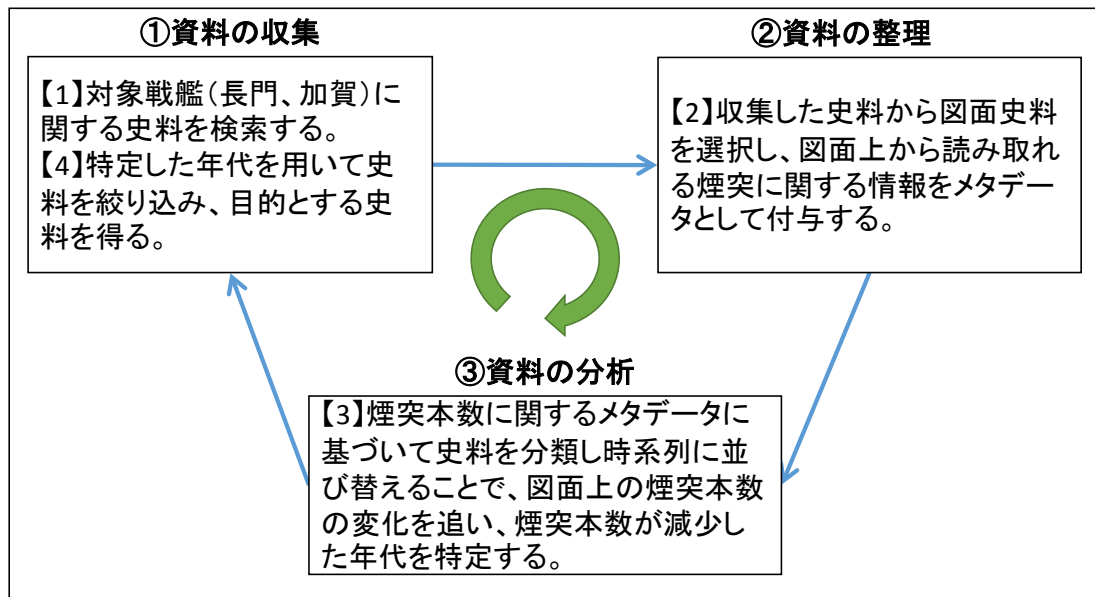


図 5-4 ケーススタディの流れ

- 【1】 まず長門および加賀に関する史料を検索する。なお注意点として、これまで「史料の収集」は史料をシステム内に登録するプロセスとしていたが、本ケーススタディでは平賀譲デジタルアーカイブで公開されている史料がシステムに登録済みであるため、それら史料群から分析対象とする史料を検索する行為とする。
- 【2】 次にそれら史料に煙突本数に関するメタデータを付与し、史料の整理を行う。具体的には煙突が読み取れる図面史料を選択し、それらに煙突に関する情報をメタデータとして付与する。
- 【3】 前プロセスで付与したメタデータに基づいて分類した史料を時系列に並び替える。そして図面上の煙突本数の変移を追うことで、煙突本数の減少が生じた年代を特定する。
- 【4】 その特定した年代前後に煙突本数の変化の原因について記述した報告書等が作成されているという仮説のもと、特定した年代を検索クエリに追加し、再度検索を行う。このように対象とする史料を絞り込んだ上で、煙突本数減少の原因である搭載汽缶数について記述した報告書を得ることが本ケーススタディの目的となる。

5.6 史料の収集

長門の同義語を追加したオントロジーを用いて、分析対象とする史料群を網羅的に取得する。具体的には全文検索を用いて、いずれかのメタデータフィールドに文字列「長門」、および長門の同義語を含む史料を取得する。結果として 707 件の検索結果を得た。

5.7 史料の整理

得られた史料群から各史料を閲覧し、史料内容を確認する。その際、煙突本数を読み取れることができる図面史料である場合は、図 5-5 に示すようにメタデータ編集インタフェースを用いて、「ノート」フィールドに「煙突本数〇〇本」等の形で煙突本数に関する調査結果をメタデータとして蓄積する。

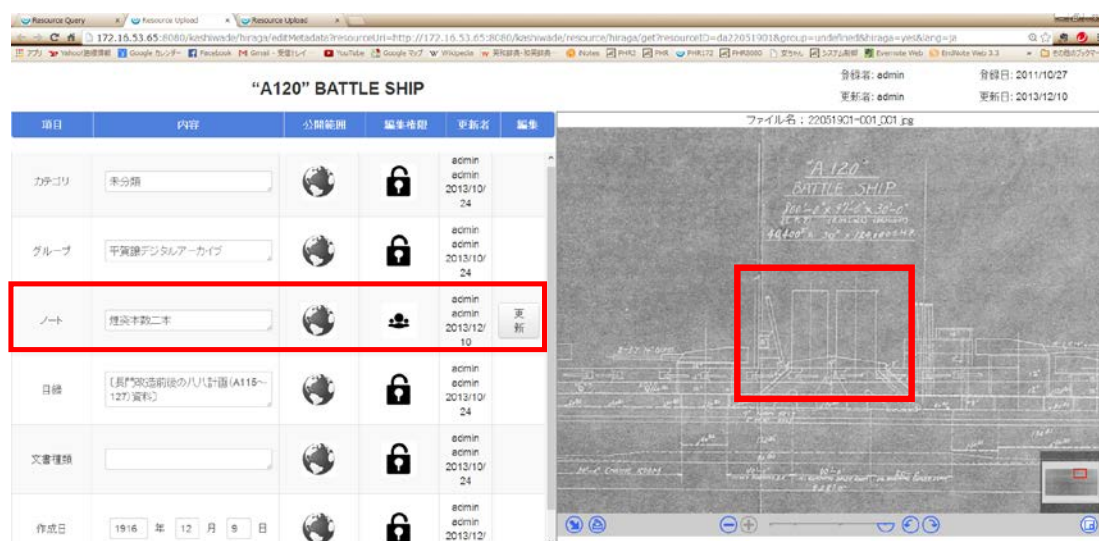


図 5-5 煙突本数に関する調査結果の蓄積例

中村[38]が開発したシステムでのケーススタディでは他の図面史料を収集する場合、一度検索結果画面に戻り、他の各史料についてその内容をそれぞれ確認する必要がある。一方、本論文ではプラグイン機能を用いて、画像情報を機械的に処理するこ

とで図面史料の収集プロセスを省力化する。具体的にはある図面史料を発見した際、その図面に対して局所エッジヒストグラムに基づいた類似画像を抽出するプラグインを実行する。このプラグインは対象とする画像から局所エッジヒストグラムを抽出し、5.4.2 で作成したインデックスに対して類似度を算出し、その上位 10 件の史料 ID をメタデータとして付与する。このメタデータを利用することで、図 5-6 に示すように対象史料に関する類似画像がメタデータ編集画面に表示される。ここに表示された史料群から優先的に検索していくことで、目的とする長門の図面史料の収集を行う。この類似画像検索を用いることで、対象史料群中のすべての図面史料を収集することはできないが、目的とする図面史料には共通して、「カード目録」フィールドに「[長門改造前後の八八計画 (A115~127) 資料]」というメタデータが付与されていることを発見し、これらのメタデータをクエリに追加することで、効率的に図面史料を収集することができた。

結果として、14 件の図面史料に対して煙突本数に関する情報をメタデータとして付与した。

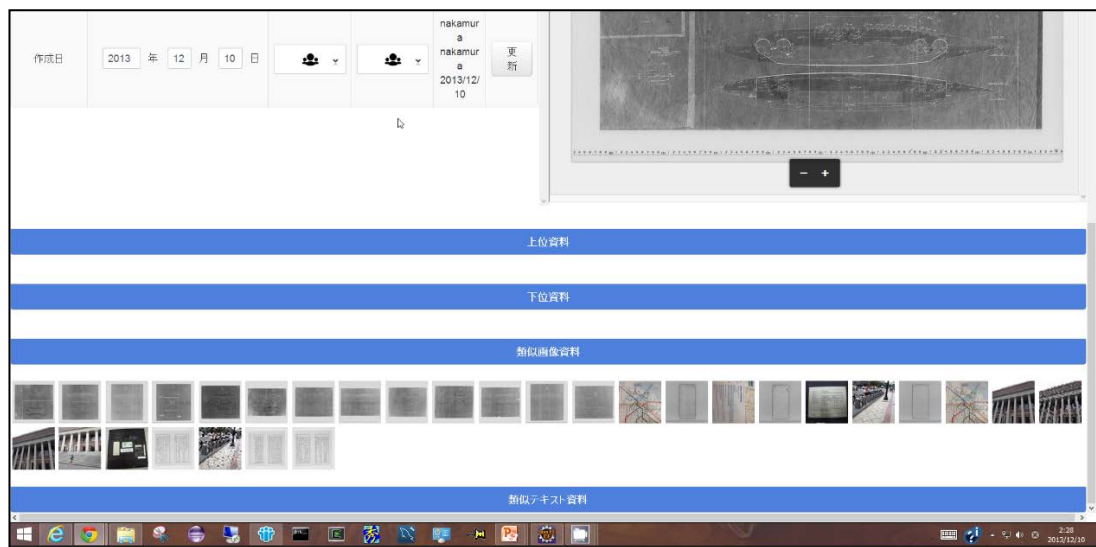


図 5-6 形状に関する類似画像の表示例

5.8 史料の分析

ここでは図面上の煙突本数の変化を確認するために、まずノートフィールドに文字列「煙突」を含む史料をメタデータ検索することによって、前プロセスで整理した 14 件の図面史料を取得する。次に煙突本数の時系列に基づいた変化を確認するために、得られた史料群を「作成年月日」フィールドに関して昇順に並び替える。並び替えた後の検索結果画面を図 5-7 に示す。



種別	表題	作成日 昇/降 更新日 昇/降	検索条件 ノート	確認 / 編集
図面史料	"A119" BATTLE SHIP	1916/12/9 2013/12/10	煙突本数二本	メタデータ テキスト 編集
図面史料	"A123" BATTLE SHIP	1917/1/11 2013/12/10	煙突本数二本	メタデータ テキスト 編集
図面史料	BATTLE SHIP"A124"(ESTIMATED)	1917/1/11 2013/12/10	煙突本数二本	メタデータ テキスト 編集
図面史料	"A124" BATTLE SHIP	1917/1/11 2013/12/10	煙突本数二本	メタデータ テキスト 編集
図面史料	"A127" BATTLE SHIP	1918/2/26 2013/12/10	煙突本数一本	メタデータ テキスト 編集
図面史料	"A126C" BATTLE SHIP	1918/2/26 2013/12/11	煙突本数一本	メタデータ テキスト 編集

図 5-7 史料群を時系列に並び替えた検索結果画面

並び替えられた史料群について、付与された煙突本数に関する記述を各史料について確認した結果、1917 年 1 月 11 日に作成された「"A124" BATTLE SHIP」という史料より以前に作成された図面史料では二本の煙突が記述されているのに対し、1918 年 2 月 26 日に作成された「"A126" BATTLE SHIP」という史料より以降に作成された史料では一本の煙突が記述されていた。これらの史料を比較機能によって同一画面に表示した結果を図 5-8 に示す。



図 5-8 比較機能を用いた煙突本数の比較結果

この分析結果から 1917 年に図面上で煙突本数が二本から一本に減少したことが確認された。この発見に基づいて、特定した年代付近に煙突本数の減少に関する理由を記述した報告書の存在を仮定する。

5.9 分析結果を利用した史料の再検索

ここでは特定した年代を検索クエリに追加することで、分析対象とする史料の絞り込みを行う。具体的には特定した年代「1917 年」について、その前後 5 年を含めた「1912 年から 1922 年」に作成された長門に関する史料を検索する。具体的には、メタデータフィールドに文字列「長門」を含む全文検索の条件に、「作成年月日」フィールドの値が「19120101」以上、かつ「19221231」以下であるという条件を加えた検索を行う。

この結果 105 件の史料に絞り込むことができ、これらの史料内容をそれぞれ確認することで、1920 年 9 月 25 日に作成された報告書「軍艦長門（8/10 予行）運転成績摘要」に、搭載汽缶数に関して専焼缶 15 缶、および混焼缶 6 缶の計 21 缶であると記述されていることが確認できた。この報告書を図 5-9 に示す。

艦長門(長門)運轉成績摘要														九年九月廿	
主軸		副軸		馬力		使用缶数		噴口使用数		機集		中			
左舷	右舷	左舷	右舷	左舷	右舷	左舷	右舷	左舷	右舷	左舷	右舷	左舷	右舷		
207	208	204		14,186	12,910	6	13,232	15	6	27	27	27	27	2495	
96	204	194	故障	11,448	12,426	2	11,991	15	6	27	27	27	27	2550	
障															
2028															

図 5-9 長門の搭載汽缶数に関する報告書

5.10 ケーススタディ結果

長門に関して行った分析プロセスを加賀についても同様に適用する。その結果、「戦艦加賀・土佐 要目一覧」という報告書から、専焼缶 8 缶、および混焼缶 4 缶の計 12 缶の汽缶が搭載されていたことが確認できた。この報告書を図 5-10 に示す。

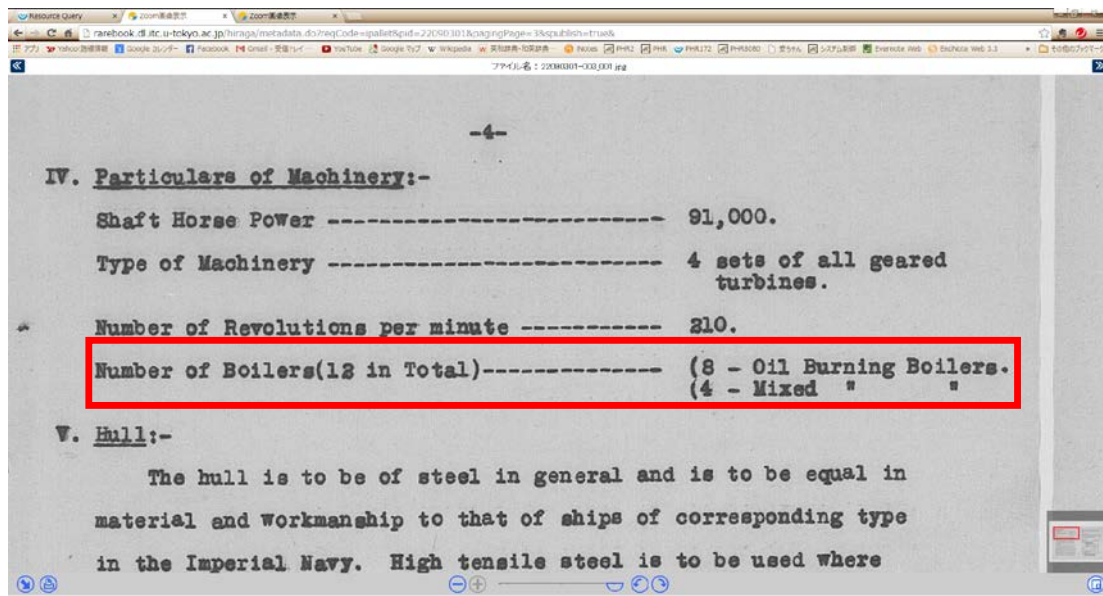


図 5-10 加賀の搭載汽缶数に関する報告書

これらの分析結果から、ケーススタディの目的であった長門・加賀の両戦艦の間で搭載汽缶数が 21 缶から 12 缶に減少していることを、平賀譲デジタルアーカイブで公開されている史料から確認することができた。

5.11 ケーススタディのまとめ

本ケーススタディでは既存の歴史研究事例を開発したシステム上で行い、提案手法および開発したシステムの検証を行った。史料の収集プロセスでは研究者が持つ暗黙知をオントロジーとして定義することで、計算機による網羅的な史料の検索が可能となった。また煙突本数に関する調査結果を史料のメタデータとして付与することによって、研究者の目的に応じた分類や比較といった分析が可能となった。具体的には煙突本数に関するメタデータに基づいた検索を行うことで、煙突本数が記述された図面史料に分類し、それらを並び替え機能や比較機能を用いて分析した結果、煙突本数が減少した年代を特定した。この特定した年代に基づいて史料の絞り込みを行うことで、対象とする戦艦の搭載汽缶数に関する報告書を容易に得ることができ、それらを対象戦艦間で比較した結果、搭載汽缶数が減少しているという既存の研究事例と同じ

発見をすることができた。

さらに図面史料を収集する際に計算機によるメタデータの自動付与を行うことによって、画像（エッジ）情報に基づいた図面史料を自動的に抽出することが可能となった。

第6章 実験：外務省来電の送付先決定過程分析

6.1 はじめに	82
6.2 実験内容	82
6.3 分析対象史料	83
6.4 実験の流れ	85
6.5 メタデータフィールドの選定	86
6.6 史料の収集	87
6.7 史料の整理	90
6.8 史料の分析	92
6.8.1 筆跡と作成日の関係に基づいた分析	94
6.8.2 筆跡と閲覧者の関係に基づいた分析	95
6.9 実験結果	97
6.10 実験のまとめ	98

6.1 はじめに

第5章では筆者個人が既存の歴史研究をシステム上で再現することによって、提案手法の評価を行った。本実験では開発したシステム上で歴史研究者が研究を行うことで、歴史研究者による実利用に基づいた提案手法および開発したシステムの評価を目的とする。さらに史料の収集・整理・分析の各プロセスを複数研究者で分担・共有することで、本研究で提案する複数研究者による歴史研究手法の評価を行う。

なお、本実験は東京大学人文社会系研究科日本文化研究専攻の修士課程の学生一名（以下、研究者 X）、著者（技術者 Y）、および史料収集を補助する研究室学生二名の計四名で行った。

6.2 実験内容

本実験では第一次世界大戦後の在外公館からの来電に関する送付決定過程の分析を行う。外務省における来電の処理過程を図 6-1 に示す。

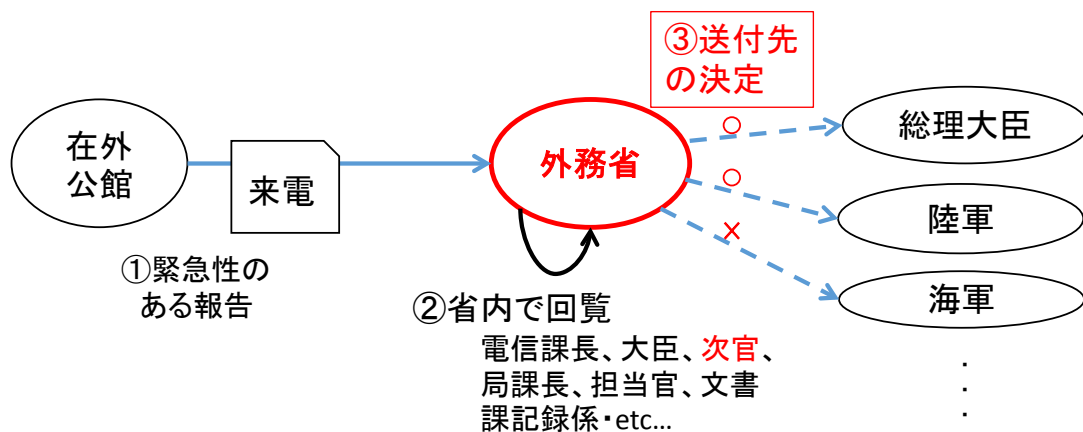


図 6-1 外務省における来電の処理過程

来電は在外公館が緊急性を要する報告を行う際に送付される外交文書である。外務省は受領した来電をまず省内で回覧する。回覧時の基本的なプロセスは「電信課長→（大臣→）次官→主管局課長→電信課→文書課記録係」となる。省内での回覧後に総

理大臣や陸軍、海軍等の他機関へ送付する。この他機関への送付先は外務省に決定権があり、情報を誰に公開するかを決定できる点で、他機関に対して優位性を持っていた。しかし省内でこれを決定する権限を有していた人物、つまり来電の送付先を決定した人物は明確になっていない。当時の外務省と他機関の関係性を分析する際には、来電の送付先の傾向等は重要な要素となり、特に来電の送付先を決定していた人物情報は特に重要である。

この送付先決定者の特定を目的として、研究者 X は送付先を指示している筆跡に基づいた分析を行った。その結果、埴原正直(以下、埴原)が外務次官に就任する 1919 年 9 月以前では、雑纂中の来電には電信課長・外務次官・主管局課長の私印、花押が確認でき、送付先指示の筆跡が次官の交代と共に変化していることが確認された。これにより、首相や軍部への送付判断は外務次官によって行われていたという仮説を立案している。一方、1919 年 9 月以降の埴原次官時代では来電の送付先が複数の筆跡によって記述されていることを確認した。当該期間は外務次官が埴原から変化していないことを考慮すると、外務次官以外の人物が送付先を指示していることを意味する。また同時期に送付先に訂正線が記述されている来電や異なる筆跡で新たな送付先が追記されている来電の存在も確認された。これらの発見から埴原次官時代では従来の外務次官が送付先を決定する処理過程に変化が生じていることが示唆された。

しかしこの先行研究では 50 件弱の来電の分析に留まり、誰が送付先を指示または訂正したのかといった具体的な人物特定には至らず、これ以上の議論が進められていない。よって本実験では来電に対して閲覧者や筆跡に関する情報をメタデータとして整理し、それらを用いて史料の分類・比較といった量的な分析を複数研究者で行うことで、これらの人物を特定することを目的とする。さらに特定した人物に基づき、当時の来電の送付先の決定がどのような過程で行われていたかを明らかにすることを目的とする。

6.3 分析対象史料

分析対象とする史料は先行研究によって来電送付の処理過程が変化したことが示唆された埴原次官時代、1919 年（大正 8）年 9 月から 1922 年（大正 11 年）前後の来電とした。具体的な来電の例を図 6-2 に示す。



表 6-1 分析に用いた史料一覧

84

これら史料は 1918 年以降の巴里平和諸条約に関する来電を含む史料である。巴里平和諸条約とは、1919 年 1 月 18 日から開会され、第一次世界大戦における連合国が中央同盟国の講和条件等について討議した会議において制定された諸条約である。本会議では世界各国の首脳が集まり、講和問題だけではなく、国際連盟を含めた新たな国際体制構築についても討議された。なお、これらの史料は外務省外交史料館が「外務省記録[45]」として所蔵している史料の一部であり、またアジア歴史資料センター[15]において公開されている。「外務省記録」とは日本の外務省により公開された外交に関する記録文書である。

6.4 実験の流れ

実験の流れを図 6-3 に示す。本実験は 3.2 で示した歴史研究プロセスに準ずる形で行う。

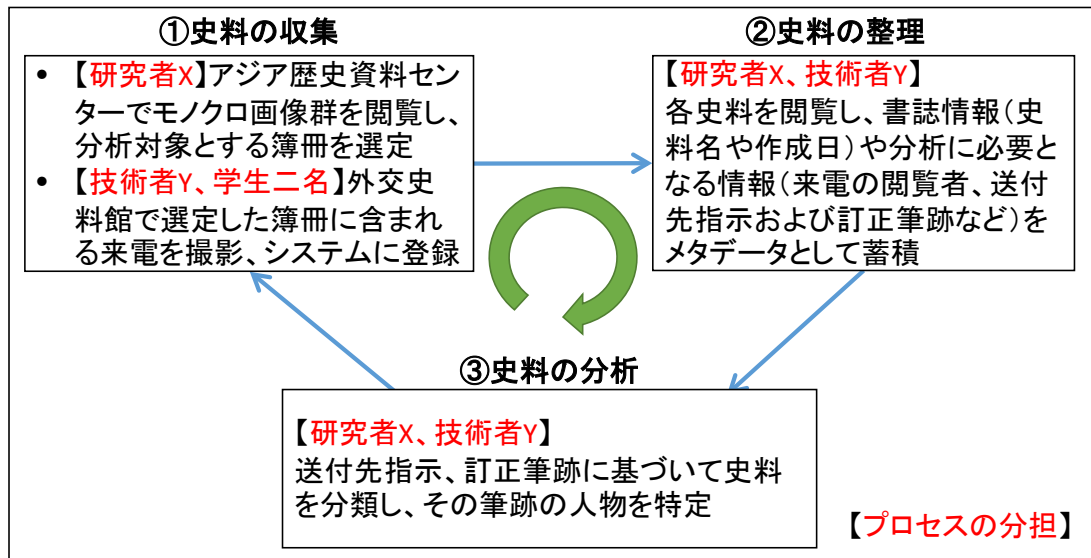


図 6-3 実験の流れ

史料の収集では、分析対象期間の史料や情勢について深い見識を持つ研究者 X がアジア歴史資料センターで公開されている史料を閲覧し、分析対象とする簿冊を選定す

る。これは次の外交史料館での撮影プロセスにおいて、開館時間等の関係から現地で
の簿冊の選定プロセスを省略するためである。次に技術者 Y と学生二名が外交史料館
に赴き、研究者 X が選定した簿冊から図 6-2 に示した形状の来電を撮影し、システム
に登録する。

史料の整理では登録された来電に対して、予め選定したメタデータフィールドを用
いて筆跡や閲覧者情報を付与する。本プロセスは研究者 X が中心となり、技術者 Y が
補足する形で行った。これは研究者 X が古文書の読解に長けており、また史料の整理
過程で得られた気づきが仮説の立案につながるためである。

史料の分析では整理された情報に基づいて史料を分類し、筆跡の人物の特定を行う。
本作業についても研究者 X と技術者 Y が共同で行う。ここでの役割としては、まず研
究者 X が史料の整理の過程でいくつかの仮説を立案する。次に立案された各仮説の検
証を目的として、技術者 Y が蓄積された情報に対してプラグインを用いた定量的な分
析を行う。このプロセスを繰り返すことで、各筆跡の人物特定を行う。

6.5 メタデータフィールドの選定

本実験では複数の研究者が共同で史料収集プロセスにおける史料の登録、史料整理
プロセスにおけるメタデータの蓄積を行うため、予め各史料に付与するメタデータ
フィールドの選定を行う必要がある。本選定については研究者 X と技術者 Y が史料の
分析方法まで想定して、以下に示すメタデータフィールドを選定し、これらをシステ
ムに予め登録した。本実験において研究者 A が各史料に付与したメタデータの例を表
6-2 に示す。

表 6-2 選定したメタデータフィールド

メタデータフィールド	メタデータバリュー	目的	付与手段
拡張子	JPG	史料管理	自動
登録日	20131215		
更新日時	20140104		
リソースID	542		
登録者	yoshida		
更新者	nakamura		
表題	バリ講2038号		
作成日	19190915	調査結果の保存	手動
閲覧者	澤田、芳澤		
送付先指示筆跡	A、筆		
送付先指示	総陸海参軍老伊		
送付先訂正筆跡	B、筆		
送付先訂正	総		
グループ	五国会議／第3巻		
カテゴリ	五国会議動向		
ノート	埴原「何ノ件ナリヤ」、主管指示が次官でない？		

6.6 史料の収集

まず史料を収集し、システムに登録するプロセスについて説明する。史料収集プロセスのイメージを図 6-4 に示す。

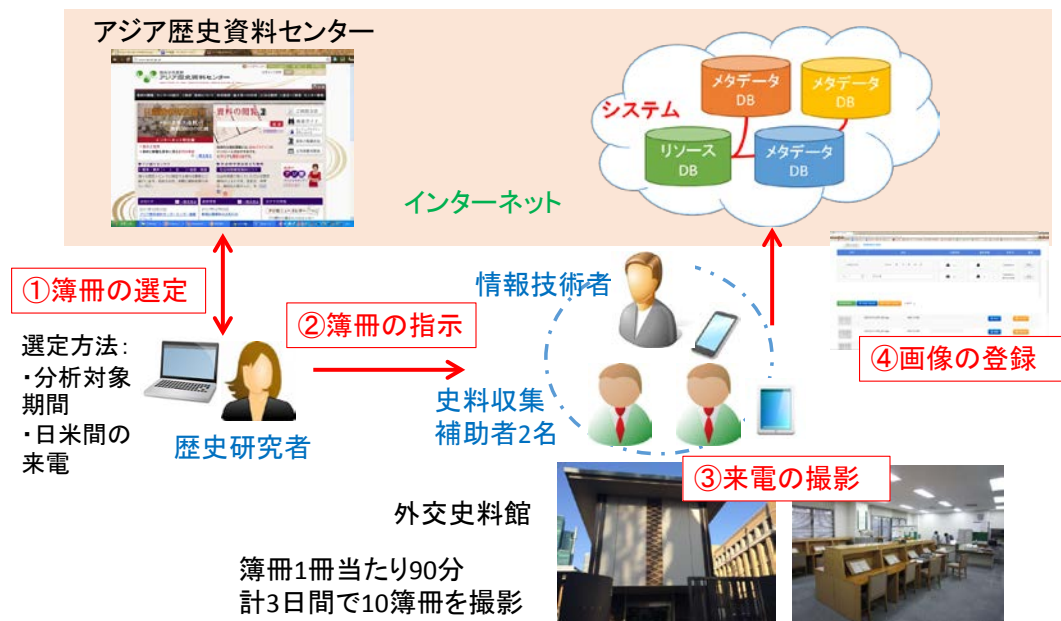


図 6-4 史料収集プロセス

図 6-4 の①では、まず研究者 X がアジア歴史資料センターで公開されている史料を閲覧し、分析対象とする簿冊の選定を行う。この際、必要に応じて図 6-5 に示すように 4.3.1 で説明したメタデータ登録機能によってシステムに史料を登録する。本プロセスは先行研究等から分析対象期間の史料や情勢について深い見識を持つ研究者 X が行う。



図 6-5 メタデータ登録機能を用いた史料の登録例

図 6-5 ではアジア歴史資料センターで公開されている「五国会議第三巻：講第 1970 号[46]」という史料を登録している例である。しかしアジア歴史資料センターで登録されている史料はモノクロ画像として公開されており、分析対象とする「閲覧者の筆跡」などを確認するには適していない。これは筆跡の文字色として赤や青が存在し、色使いも人物を特定するために有益な情報であるためである。

図 6-4 の②から④では、研究者 X が選定した簿冊を基に、史料収集者である技術者 Y および史料収集補助者二名が実際に外交史料館に赴き、三日間を要して計 471 件の来電を撮影し、その画像を 4.3.2 で説明したファイル登録機能を用いてシステムに登録する。この作業については図 6-2 に示した分析対象とする来電の形式に類似した史料を簿冊から抽出し、デジタルカメラにて撮影する単純な作業である。登録の際には簿冊を単位として、「グループ」フィールドに簿冊名を入力して複数画像を一括登録した。参考として 4.5.4 で説明した史料の比較機能を用いて、アジア歴史資料センターで公開されている史料と、実際に撮影した史料の比較したものを図 6-6 に示す。公開史料では黒字で記述されている部分について、撮影史料では赤や青の文字色が使われていることが確認できる。

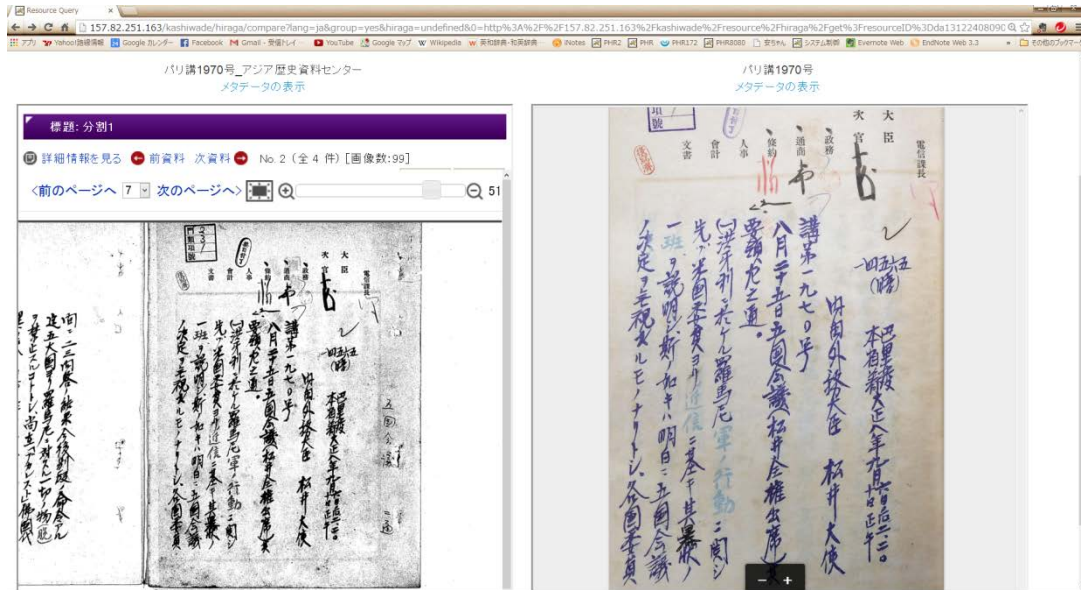


図 6-6 アジア歴史資料センターの公開史料（左）と撮影史料（右）の比較

6.7 史料の整理

次にシステムに登録した史料群に対して、6.5 で定義したメタデータフィールドを用いて筆跡や閲覧者等の情報を付与する流れについて説明する。本プロセスは研究者 X が中心となり、技術者 Y が補足する形で行った。これは研究者 X が古文書の読解に長けており、また史料の整理過程で得られた気づきが仮説の立案につながるためである。また技術者 Y は研究者 X が付与したメタデータの表記揺れの修正や、後述する分析プロセスにおいて新たに必要となったメタデータの蓄積を行った。これはすでに研究者 X が付与したメタデータを確認することで、古文書の読解に長けていない技術者 Y でも史料内容を把握することが可能となったためである。

研究者 X は史料名や作成日といった書誌情報に加えて、閲覧者のサインや筆跡に基づいた来電の閲覧者の抽出や、複数の筆跡によって記述された送付先の抽出を行う。史料情報を整理している例を図 6-7 に示す。

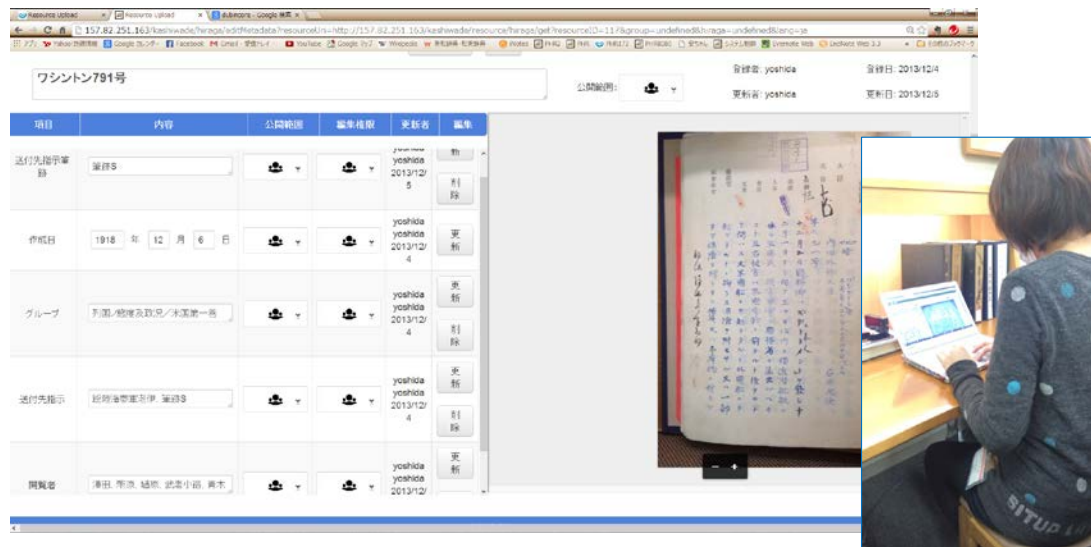


図 6-7 メタデータ編集機能を用いた史料情報の整理例

登録した各史料について、メタデータ編集インターフェースを用いて史料情報をメタデータとして付与する。その際、史料の拡大・縮小機能を用いることで、閲覧者の筆跡等の細部の確認を行う。また「項目の追加」機能を用いて、必要となるメタデータフィールドを追加し、その値を入力する。蓄積された史料情報の RDF の例を図 6-8 に示す。

```

<?xml version="1.0"?>
<rdf:RDF xmlns:kd="http://kashiwade.org/2012/09/kd#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/">
  - <rdf:Description rdf:about="http://157.82.251.163/kashiwade/resource/hiraga/get?resourceID=120">
    <kd:thumbnailUri rdf:resource="http://157.82.251.163/kashiwade/resource/hiraga/get?resourceID=120"/>
    <kd:modifier>yoshida</kd:modifier>
    <dcterms:medium>JPG</dcterms:medium>
    - <dc:date rdf:parseType="Resource">
      <dcterms:accrualPolicy>group</dcterms:accrualPolicy>
      <dc:publisher>yoshida</dc:publisher>
      <kd:modifier>yoshida</kd:modifier>
      <dcterms:accessRights>group</dcterms:accessRights>
      <dc:description>19190304</dc:description>
      <dcterms:modified>20131204</dcterms:modified>
    </dc:date>
    <dc:title>ワシントン188号</dc:title>
    <kd:resourceId>120</kd:resourceId>
    <dc:identifier>http://157.82.251.163/kashiwade/hiraga/zoom?resourceUri=http://157.82.251.163/kashiwade/resource/hiraga/get?resourceID=120</dc:identifier>
    <dc:publisher>yoshida</dc:publisher>
    - <kd:director rdf:parseType="Resource">
      <dcterms:accrualPolicy>group</dcterms:accrualPolicy>
      <dc:publisher>yoshida</dc:publisher>
      <kd:modifier>yoshida</kd:modifier>
      <dcterms:accessRights>group</dcterms:accessRights>
      <dc:description>総陸海参軍老伊、筆跡S</dc:description>
      <dcterms:modified>20131204</dcterms:modified>
    </kd:director>
    <dc:source>IMG_9136.jpg</dc:source>
    <dcterms:accessRights>group</dcterms:accessRights>
    <dcterms:created>20131204</dcterms:created>
    <dcterms:modified>20131204</dcterms:modified>
    - <kd:reader rdf:parseType="Resource">
      <dcterms:accrualPolicy>group</dcterms:accrualPolicy>
      <dc:publisher>yoshida</dc:publisher>
      <kd:modifier>yoshida</kd:modifier>
      <dcterms:accessRights>group</dcterms:accessRights>
      <dc:description>澤田、幣原、芳澤、青木</dc:description>
      <dcterms:modified>20131204</dcterms:modified>
    </kd:reader>
  </rdf:Description>
</RDF>

```

図 6-8 蓄積された史料情報の RDF 表現例

「表題 (dc:title)」や「登録日 (dcterms:created)」といった史料管理を目的としたメタデータは各史料に与えられた URI に対して直接紐づけられている。それに対し、「閲覧者 (kd:reader)」や「送付先指示筆跡 (kd:director)」といったメタデータフィールドのバリューについては、その「内容 (dc:description)」の他に、「更新者 (kd:modifier)」や「公開範囲 (dcterms:accessRights)」等の複数研究者で共有するためのメタデータが登録される。

6.8 史料の分析

ここでは 5.7 で整理した筆跡情報や閲覧者等の史料情報を利用して、史料の分類や比較といった分析を行い、各筆跡の人物特定を行う。本プロセスについては研究者 X と技術者 Y によって行い、研究者 X が史料の整理プロセスで得た知見や関連する知識に基づいて仮説を立案し、技術者 Y がその検証に必要な情報を蓄積されたメタデータ

から機械的に抽出し、仮説の量的な検証を行うことで分析を行う。

史料の整理の結果、対象とする 471 件の来電には四種類の筆跡が確認された。本実験ではこれらの各筆跡に対して「A」「B」「S」「印」というラベルを付与して説明する。図 6-9 に送付先の記述における各筆跡の例を示す。各筆跡の一文字目はすべて総理大臣への送付先を示す「総」の筆跡である。

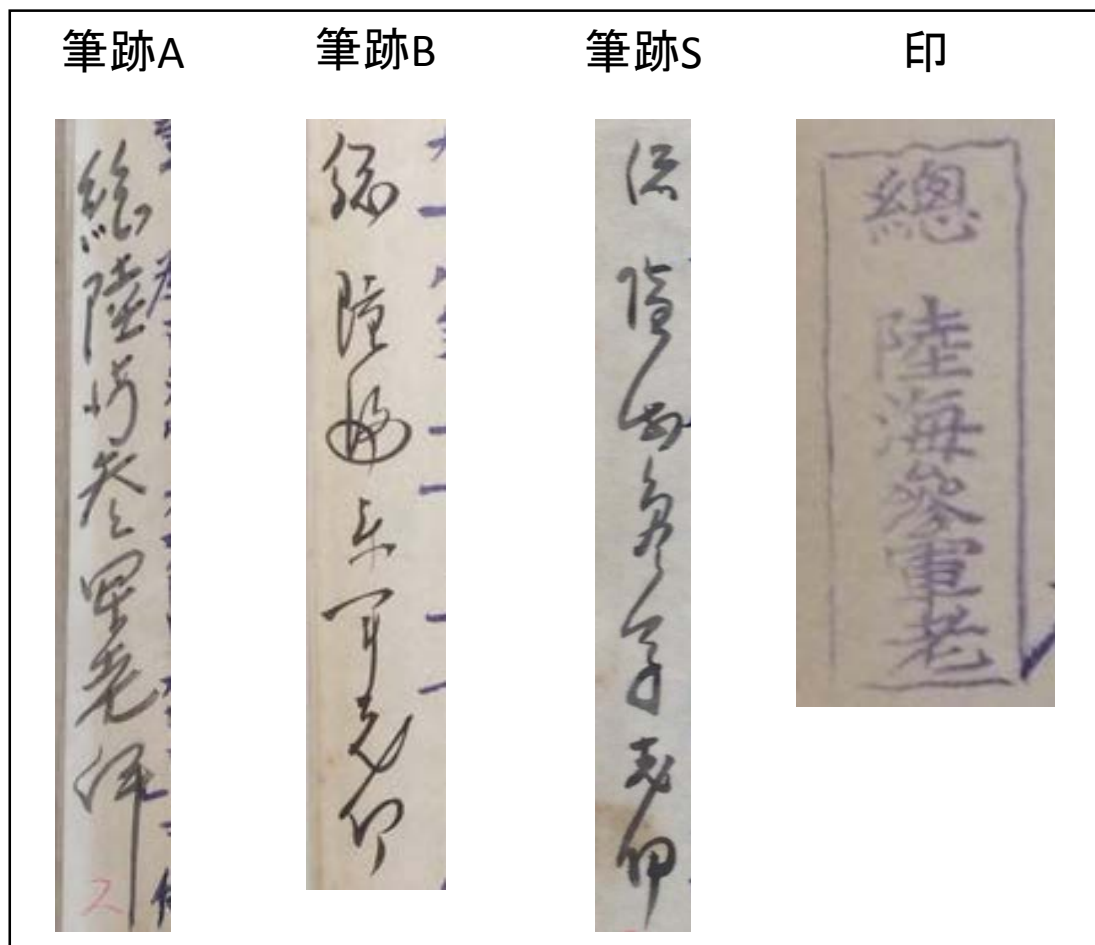


図 6-9 送付先の記述における筆跡例

これらの各筆跡の人物特定に向けて、本実験では筆跡と作成日の関係、および筆跡と閲覧者の関係の二点の観点から分析を行った。以下ではそれぞれの分析の詳細について述べる。

6.8.1 筆跡と作成日の関係に基づいた分析

まず来電を時系列に並び替えることで、筆跡の出現時期に基づいた分析を行った。この分析結果を図 6-10 に示すが、これは技術者 Y がプラグインを開発し、471 件の来電から筆跡と作成日に関するメタデータを引数とし、月別の送付先指示筆跡の出現頻度を出力したものである。

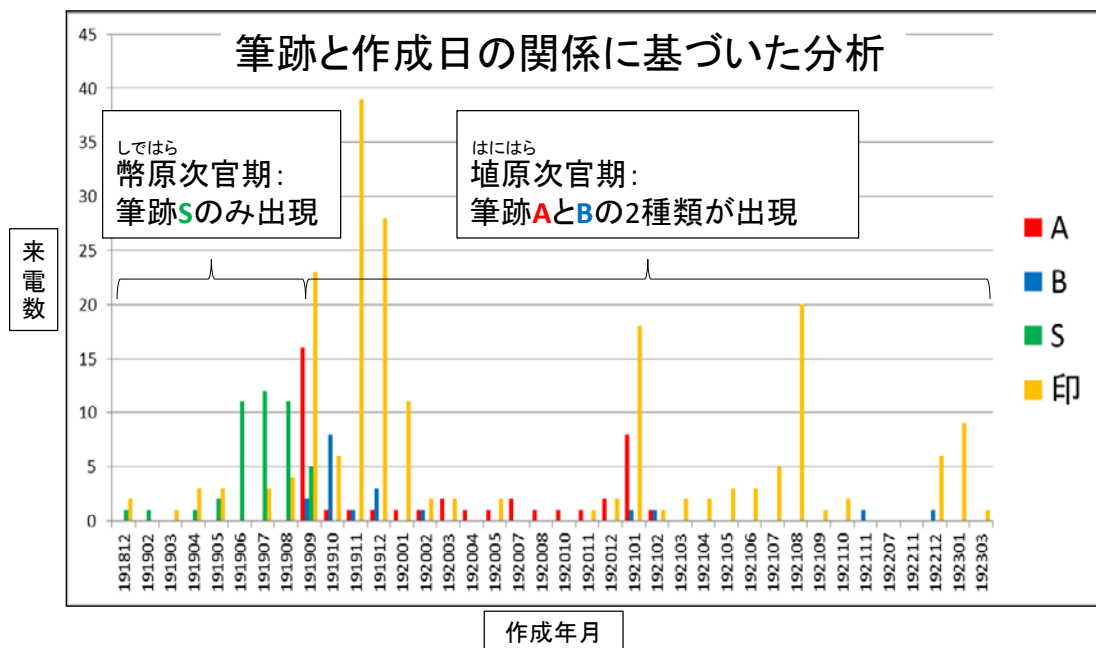


図 6-10 月別の送付先指示筆跡の出現頻度

この結果から筆跡 S は 1919 年 9 月以前にのみ登場していることが確認できる。これは幣原喜重郎（以下、幣原）が外務次官を務めていた時期である。これにより、研究者 X が先行研究で示唆したように、1919 年 9 月以前では外務次官が送付先を決定しているという仮説を指示した結果となっている。また埴原が外務次官に就任した 9 月以降では少なくとも筆跡 A、B の二種類の筆跡によって送付先が指示されていることが視覚的に確認できるため、外務次官以外の人物が来電の送付先を指示していることがわかる。また当該期間では「印」による送付先指示が増大していることが確認できるが、この原因については明らかにすることができなかった。これについては他の

文献を参照することで、今後明らかにする必要がある。

6.8.2 筆跡と閲覧者の関係に基づいた分析

次に筆跡別の来電における閲覧者の登場割合に基づいた分析を行う。これは研究者 X が史料の整理プロセスにおいて、埴原が来電閲覧時に記入するサインと筆跡 B が類似するという気づきを得たためである。この分析結果を図 6-11 に示すが、これについても 6.8.1 と同様に技術者 Y が開発したプラグインによって出力した結果である。

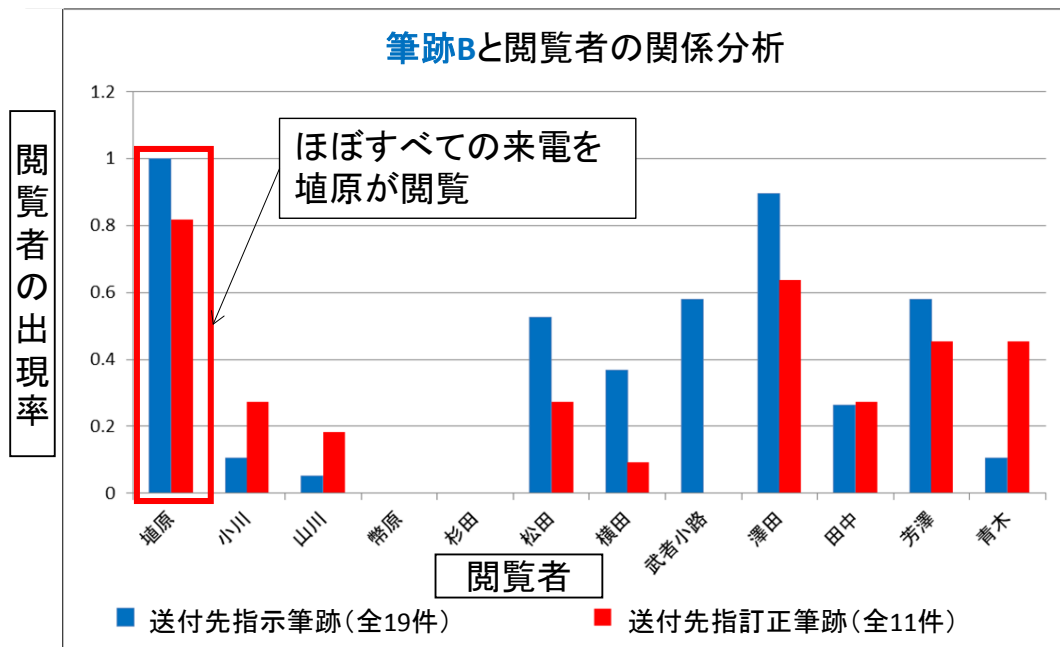


図 6-11 筆跡 B における閲覧者の登場割合

この結果から筆跡 B によって送付先が指示(19 件)または訂正(11 件)されたほぼすべての来電において、埴原が閲覧していることが分かる。これは研究者 X が立てた筆跡 B が埴原であるという仮説を支持する一つの量的な結果といえる。

さらにこの発見に基づき、高い割合で閲覧している人物は送付先を指示した人物である可能性が高いという仮説を立て、技術者 Y が筆跡 A によって送付先が指示されている来電に対して同様の分析を行った。その結果を図 6-12 に示す。

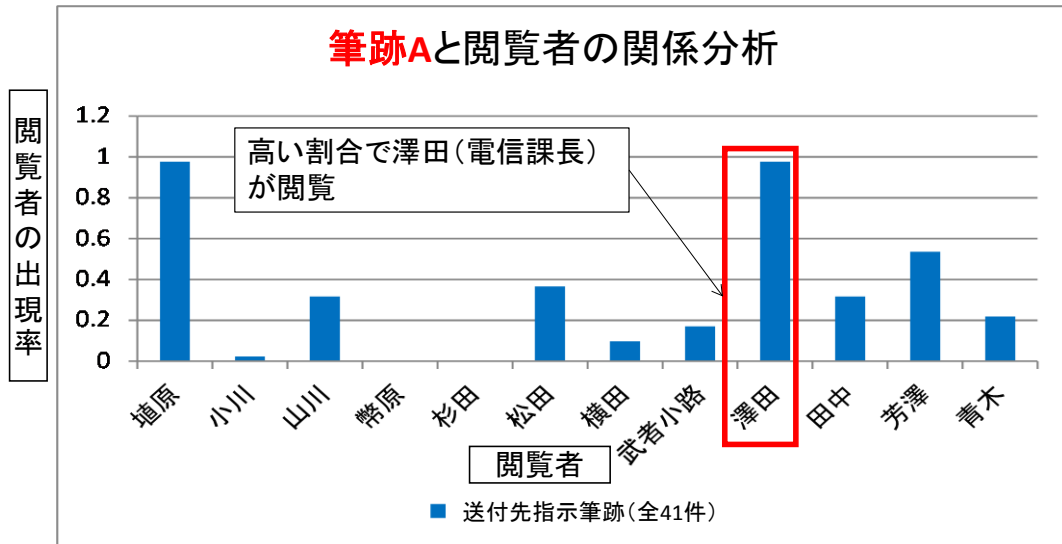


図 6-12 筆跡 A における閲覧者の登場割合

この結果、電信課長であった澤田が高い割合で閲覧していることが分かった。また研究者 X が他文献[47]を調査した結果、1921 年 3 月 3 日から 9 月 3 日まで澤田は皇太子の外遊に同行していることが判明し、図 6-10 に示した月別の筆跡の出現頻度に関する出力結果と照合した結果、当該期間には筆跡 A による送付先指示は行われていないことが分かった。これについて、筆跡と作成日の関係に基づいた分析結果を図 6-13 に再掲する。

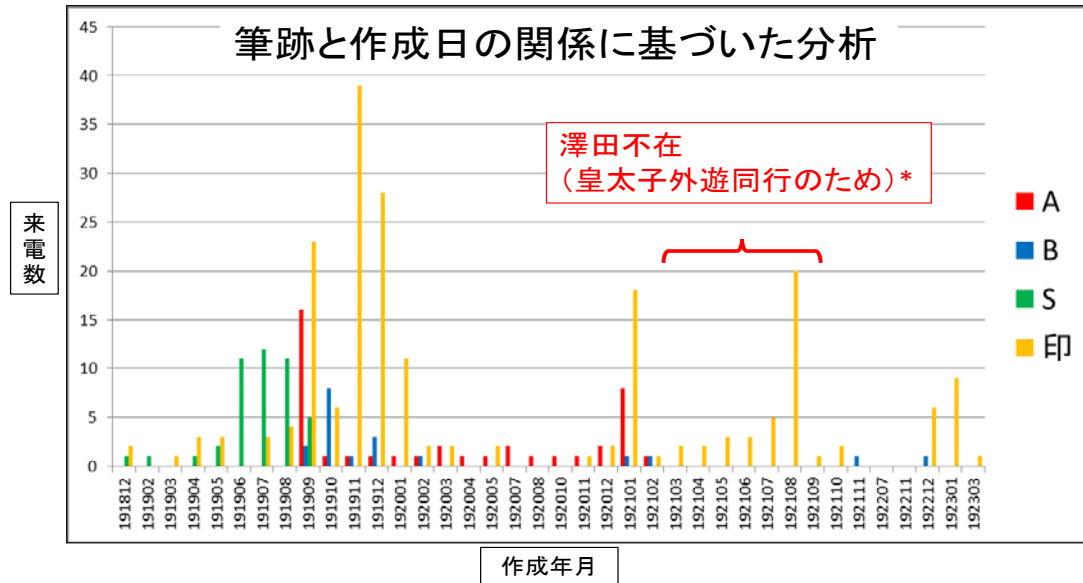


図 6-13 筆跡 A と閲覧者の関係分析

これらは筆跡 A が澤田であることを支持する結果である。さらに筆跡 A と澤田が来電閲覧時に記入するサインと照合した結果、全 41 件中、明らかに筆記具が同一と思われるものが 16 件存在することが確認できた。また不一致であった残りの 25 件中 17 件において、澤田の閲了サインに使われている筆跡は赤鉛筆や青鉛筆であり、事務規定上、送付先指示の際は筆ないしペンに持ち替えた可能性が示唆された。これは研究者 X の先行研究から、伝達事項の書き込みに色鉛筆を禁じられていたという知見に基づく。この検証に基づき、本実験では筆跡 A は澤田によるものであると結論づけた。

6.9 実験結果

本実験では分析対象とした 471 件の来電において、送付先の指示を行っている各筆跡の人物に関する仮説を立案することができた。具体的には筆跡 A は澤田電信課長、筆跡 B は埴原次官、筆跡 S は幣原次官であると結論づけた。またこれらの発見から、幣原次官期には外務次官が送付先を単独決定していたのに対し、埴原次官期には電信課長である澤田が送付先を仮決定し、次官である埴原が訂正することで最終的な送付先を決定していたという仮説を立案することができた。またこの仮説の検証を目的として、6.8.1 で行った月別

の筆跡の出現頻度を分析した結果、送付先訂正筆跡として埴原の筆跡 **B** のみが確認され、澤田の筆跡 **A** による訂正は 0 件であることが確認でき、立案した仮説を支持する結果を得ることができた。

今後はさらに多くの史料、また異なる形式の外交文書を分析対象として、本実験で立案した仮説の検証を行う。さらに仮説として立案した処理プロセスが行われた原因等についても引き続き分析を進める。

6.10 実験のまとめ

本実験では開発したシステム上で歴史研究者が研究を行うことで、提案手法および開発したシステムの歴史研究者による実利用評価を行った。特に史料収集プロセスのコスト(労力・時間)の削減が可能となった。具体的には本プロセスを三名で分担することにより、収集史料数については先行研究で対象とした 50 件弱から 471 件まで増加させることができた。また収集時間に関しては一冊当たり 90 分要する簿冊を 3 日間で 10 冊撮影した。さらにこれらの各研究者が登録した史料を複数研究者でクラウド上の DB を共有することによって、量的な分析による来電の送付先決定者を特定することが可能となった。

さらに本実験では歴史研究者が立案した仮説を情報技術者の量的な分析によって検証し、また情報技術者が蓄積された調査結果を可視化することによって歴史研究者への新たな仮説の提案を行った。このように異なる分野の研究者が共同で歴史研究に参画することによって、協調的な史料分析が可能となり、本研究で提案する複数研究者による歴史研究環境の有用性を確認することができた。

第7章 考察

7.1 ケーススタディにおける検索精度評価	100
7.1.1 検索精度評価指標	100
7.1.2 精度評価における検索方法	101
7.1.3 オントロジーを用いた検索精度の評価	101
7.1.3.1 正解史料群	101
7.1.3.2 検索方法	102
7.1.3.3 検索精度の評価結果	102
7.1.4 類似画像を用いた検索精度の評価	103
7.1.4.1 正解史料群	103
7.1.4.2 検索方法	104
7.1.4.3 検索精度の評価結果	104
7.1.5 特定した年代を用いた検索精度の評価	105
7.1.5.1 正解史料	105
7.1.5.2 検索方法	107
7.1.5.3 検索精度の評価結果	107
7.2 既存の歴史研究プロセスの課題点に対する評価	108
7.2.1 「史料の収集」プロセスにおける評価	108
7.2.2 「史料の整理」プロセスにおける評価	109
7.2.3 「史料の分析」プロセスにおける評価	110
7.2.4 提案手法に基づく歴史研究プロセスの改善	111
7.2.5 来電における標準メタデータセットの提案	113

7.1 ケーススタディにおける検索精度評価

第5章では長門型戦艦の設計書の変更履歴を分析にすることによって、煙突本数が減少した原因追究を開発したシステム上で行った。その過程において、5.6 ではオントロジーを利用した史料の検索、5.7 では類似画像検索を利用した図面史料の収集、および5.9 では分析から特定した年代を検索クエリに追加した報告書の抽出を行った。ここではこれらの検索毎の目的に基づいて、検索精度の評価を行うことで、ケーススタディにおける定量的な評価を行う。

7.1.1 検索精度評価指標

本項では検索精度の評価を図7-1に示す適合率・再現率（F値）の指標を用いて行う。

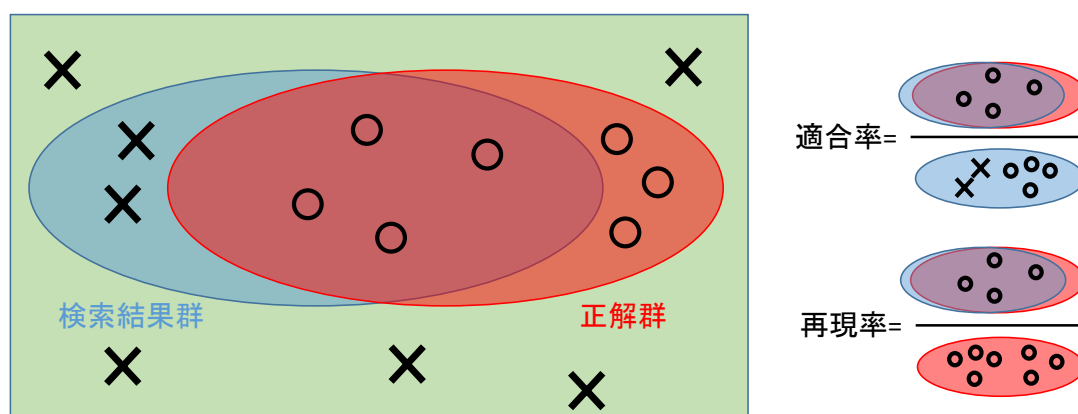


図 7-1 適合率・再現率

適合率は検索結果群に対して検索結果群中の正解の割合を示し、再現率は正解群に対しての検索結果群中の正解の割合を示す。一般にこれらの二つの指標は相反性にある。したがって異なるシステムの検索精度を比較する際には、式(7.8)に示す再現率と適合率の調査平均を取ったF値が用いられる。

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (7.8)$$

7.1.2 精度評価における検索方法

本項では精度評価を行う際の検索方法について説明する。ここでは以下に示す検索方法を対象とする。

- 全文検索
- フィールド「表題」に対するメタデータ検索
- フィールド「カテゴリ」に対するメタデータ検索

これらの検索方法の選定理由としては、検索対象とする史料群に付与されているメタデータを予め知らないユーザを想定し、該当ユーザの使用する可能性が高い検索方法とした。

全文検索では史料に付与されているメタデータに関して、いずれかのフィールドに検索クエリとする文字列を含む史料を検索する手法である。再現率は高くなるが、適合率は低くなる傾向がある。

フィールド「表題」を用いたメタデータ検索は、史料名に検索クエリとする文字列を含む史料を検索する。表題は全ての史料に付与されており、また史料情報を端的に示すという点で、一般的なユーザが検索対象とし易いフィールドであると考えたためである。

フィールド「カテゴリ」を用いたメタデータ検索は、上記の検索語を入力する検索とは異なり、予め各史料に付与されている有限のカテゴリから選択して検索する方法である。平賀譲デジタルアーカイブでは「主力艦：長門型」や「軍艦構造：溶接」といった 52 のカテゴリが付与されているため、登録済みのカテゴリから選択して検索できる点で、登録されている史料に関する知識が少ないユーザでも利用しやすいメタデータである判断したためである。

7.1.3 オントロジーを用いた検索精度の評価

本項ではオントロジーを用いて、長門に関する史料の検索精度の評価を行う。

7.1.3.1 正解史料群

正解とする史料群は 5.6 で分析対象とした煙突本数が確認できる長門の図面史料 14 件とする。これら史料の一覧を表 7-1 に示す。

表 7-1 正解史料群

番号	表題	カテゴリ	カード目録
1	軍艦長門改正図	主力艦：長門型	〔長門防禦関係〕
2	軍艦長門改正案(C)	未分類	「軍艦長門改正図」等
3	“A116” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
4	“A118” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
5	“A119” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
6	“A120” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
7	“A122” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
8	“A123” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
9	“A124” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
10	BATTLE SHIP“ A124”(ESTIMATED)	未分類	〔長門改造前後の八八計画(A115～127)資料〕
11	“A126” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
12	BATTLE SHIP“ A126A”	未分類	〔長門改造前後の八八計画(A115～127)資料〕
13	“A126B” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕
14	“A126C” BATTLE SHIP	未分類	〔長門改造前後の八八計画(A115～127)資料〕

表中の番号 1 の史料は表題、カテゴリともに「長門」という文字列を含むが、番号 2 の史料は表題にのみ文字列「長門」を含む。さらに番号 3 以降の史料については文字列「長門」を含まないことが確認できる。ただし、カード目録には全正解史料について文字列「長門」を含む。

7.1.3.2 検索方法

本検索精度評価を行うための検索方法について説明する。使用するメタデータフィールドは 7.1.2 で説明した三種類を使用し、それぞれの検索文字列は「長門」とする。その際、比較対象としてオントロジーを利用しない検索と利用する検索を比較する。オントロジーを利用しない検索は指定したフィールドに文字列「長門」を含むか否かを判別する検索である。一方、オントロジーを利用した検索は長門の暗号名「A123」や「A124」、および英語表記である「NAGATO」等の長門の同義語として定義された語を検索文字列に追加し、これらの追加された文字列のいずれかを指定したフィールドに含むか否かを判断する検索である。なおカテゴリを用いた検索ではカテゴリ「主力艦：長門」を使用し、オントロジーを利用した検索は対象外とする。

7.1.3.3 検索精度の評価結果

7.1.3.2 のそれぞれの検索方法について、オントロジーを用いた検索と用いない検索を行った際の評価結果を表 7-2 に示す。

表 7-2 オントロジーを用いた検索精度の評価結果

検索方法	オントロジー 利用の有無	検索結果数	正解数	適合率	再現率	F値
全文検索	無	550	14	0.03	1.00	0.05
メタデータ検索(表題)	無	168	2	0.01	0.14	0.02
メタデータ検索(カテゴリ)	無	185	1	0.01	0.07	0.01
全文検索	有	707	14	0.02	1.00	0.04
メタデータ検索(表題)	有	508	14	0.03	1.00	0.05

図中の赤色の行は F 値が最も高かった検索方法を示す。この結果からオントロジーを利用しない場合の全文検索と、オントロジーを利用した場合のフィールド「表題」に対するメタデータ検索が最も高い F 値を得た。

オントロジーを用いない全文検索を行った際には、7.1.3.1 で述べたようにフィールド「カード目録」に文字列「長門」を含むため、再現率が 1 となり利用した検索方法の中では高い F 値を示している。

一方オントロジーを利用した全文検索では、再現率は 1 であるが正解でない検索結果数が増すことで適合率が下がり、結果として F 値が下がっている。

これらの F 値を用いた考察ではオントロジーを用いた検索の有用性を一概に示すことはできないが、ケーススタディにおける 5.6 の検索の目的は「長門に関する史料を網羅的に検索する」ことであった。つまり高い再現率を要求されており、その点ではオントロジーを利用した二種類の検索方式の両方で再現率 1 を示している。したがって第 5 章のケーススタディでの目的においては、オントロジーを用いた検索に優位があると結論づけられる。

7.1.4 類似画像を用いた検索精度の評価

本項ではプラグインを用いて類似画像情報を機械的に付与したメタデータを利用して図面史料を検索する際の検索精度の評価を行う。

7.1.4.1 正解史料群

正解史料群は 7.1.3.1 で説明した史料 14 件とする。

7.1.4.2 検索方法

本検索精度評価を行うための検索方法について説明する。ここでは文字列情報のみを用いた検索と、画像の形状に関する情報を加えた検索を比較する。前者については、7.1.3.2 で説明した五種類（オントロジーを利用しない全文検索、表題を利用したメタデータ検索、オントロジーを利用した全文検索、表題を利用したメタデータ検索、およびカテゴリを用いたメタデータ検索）とする。後者については、前者の文字列情報を含み、かつ画像の形状に関して類似する資料上位 10 件を検索結果とした。より具体的には各正解史料の画像情報を検索クエリとし、上位 10 件中に含まれる他の正解史料群の割合が適合率となる。なお正解史料は 14 件存在するため、適合率・再現率（および F 値）の算出には 14 件の平均値を利用している。

7.1.4.3 検索精度の評価結果

7.1.4.2 のそれぞれの検索方法について、クエリとして画像情報を追加しない検索と、追加した検索を行った際の評価結果を表 7-3 に示す。

表 7-3 類似画像検索を用いた検索精度の評価結果

検索方法	オントロジー 利用の有無	類似画像検索 の有無	検索結果数	正解数	適合率	再現率	F値
全文検索	無	無	550	14	0.03	1.00	0.05
メタデータ検索(表題)	無	無	168	2	0.01	0.14	0.02
メタデータ検索(カテゴリ)	無	無	185	1	0.01	0.07	0.01
全文検索	有	無	707	14	0.02	1.00	0.04
メタデータ検索(表題)	有	無	508	14	0.03	1.00	0.05
全文検索	無	有	10	3.79	0.38	0.27	0.32
メタデータ検索(表題)	無	有	10	0.64	0.06	0.05	0.05
メタデータ検索(カテゴリ)	無	有	10	0.36	0.04	0.03	0.03
全文検索	有	有	10	3.64	0.36	0.26	0.30
メタデータ検索(表題)	有	有	10	4.00	0.40	0.29	0.33

表上部の画像情報を利用しない検索方式の精度は 7.1.3.3 と同値である。一方、表下部に示す画像情報を利用した検索では、検索結果数が 10 件に限定されるため再現率が低くなるが、二種類の検索方式を除いて 4 件弱の正解数を含むことで高い適合率を示している。低い適合率・再現率を示したオントロジーを利用しない「表題」フィールドに対するメタデータ検索、および「カテゴリ」フィールドに対するメタデータ検索では、検索クエリとして利用した文字列情報から正解数がそれぞれ 2 件、1 件と限

定されるため、画像情報を利用の有無に関わらず低い検索精度を示す。よってこれら2種類の検索方式を除外した場合、画像情報を利用した検索方式ではF値が総じて0.3以上となり、文字列情報のみを利用した検索と比較して高い検索精度を示すことが確認できた。

なお画像情報を利用した検索において、不正解となった画像例を図 7-2 に示す。

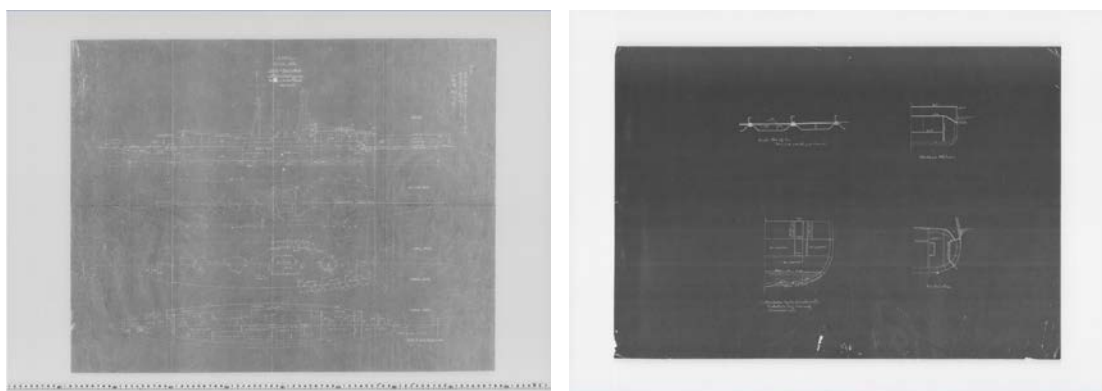


図 7-2 クエリとした利用した画像（左）と不正解画像（右）例

不正解となった画像も図面ではあるが、クエリとして利用した画像の形状とは大きく異なる点を確認できる。この課題に対して本研究では局所エッジヒストグラムを特徴量として利用しているが、D. K. Park ら [28]はこの特徴量に加えて全体および準全体エッジヒストグラムを特徴量として扱い、類似画像検索の精度向上を実現している。今後これらのアルゴリズムを利用した精度向上が課題となる。

7.1.5 特定した年代を用いた検索精度の評価

本項では特定した年代情報を利用して、長門の煙突本数の減少に関する報告書を検索する際の検索精度の評価を行う。

7.1.5.1 正解史料

正解史料は図 7-3 に示す史料 1 件とする。

0123456789

軍艦長門(予行)運轉成績摘要

九月九日施行

番号	種類	主砲				副砲				使用砲数				命中数				総砲撃							
		左舷	右舷	中央	合計	左舷	右舷	中央	合計	左舷	右舷	中央	合計	左舷	右舷	中央	合計	左舷	右舷	中央	合計				
1	普通	2588	212	107	208	204	1416	2910	1323	15	6	27	27	27	27	245	255	470	470						
2	普通	2375	198	196	204	194	1145	1242	1194	15	6	27	27	27	27	255	255	464	475						
		総平均 202.8																							
		235	205	201	206	177	1271	2465	1118	15	6	27	27	27	27	254	254	470	475						
番号	種類	主砲				副砲				使用砲数				命中数				総砲撃							
1	普通	177	173	220	183	50	40	50	52	25	25	27	27	27	27	28	28	28	28						
2	普通	141	175	180	155	40	40	50	40	27	28	27	27	27	27	28	28	28	28						
		142	174	200	170	45	40	50	46	27	28	27	27	27	27	28	28	28	28						

平均砲水量 三六五〇

軍艦長門(8/10予行)運轉成績摘要

項目	内容	公開範囲	編集権限	更新者	編集
文書種類	手書き謄写			admin admin 2013/10/24	
目録	「長門計画ノ件(大正六年九月廿八日)」等(平賀書)			admin admin 2013/10/24	
グループ	平賀譲デジタルアーカイブ			admin admin 2013/10/24	
カテゴリ	主力艦:長門型			admin admin 2013/10/24	
作成日	1920 年 9 月 25 日			admin admin 2013/10/24	

図 7-3 正解史料(上)と史料が持つメタデータ(下)

本史料は「1920 年 9 月 25 日」に作成された「軍艦長門 (8/10 予行) 運転成績摘要」という史料である。「カード目録」フィールドに文字列「長門」を含み、また「主力艦：長門」というカテゴリが割り当てられている。

7.1.5.2 検索方法

本検索精度評価を行うための検索方法について説明する。ここでは文字列「長門」を用いた検索と、作成年月日に関する情報を加えた検索を比較する。前者については、7.1.3.2 で説明した五種類（オントロジーを利用しない全文検索、表題を利用したメタデータ検索、オントロジーを利用した全文検索、表題を利用したメタデータ検索、およびカテゴリを用いたメタデータ検索）とする。後者については、前者の文字列情報を含み、かつ作成年月日に関して制約を加えた検索結果とする。より具体的にはケーススタディで特定した「1917 年」の前後 5 年を含めた「1912 年から 1922 年」に作成された史料を検索結果とする。

7.1.5.3 検索精度の評価結果

7.1.5.2 のそれぞれの検索方法について、作成年月日に関するクエリの制約を追加しない検索と、追加した検索を行った際の評価結果を表 7-4 に示す。

表 7-4 作成年月日に関する制約を加えた検索精度の評価結果

検索方法	オントロジー 利用の有無	作成年月日の 指定の有無	検索結果数	正解数	適合率	再現率	F値
全文検索	無	無	550	1	0.00	1.00	0.00
メタデータ検索(表題)	無	無	168	1	0.01	1.00	0.01
メタデータ検索(カテゴリ)	無	無	185	1	0.01	1.00	0.01
全文検索	有	無	611	1	0.00	1.00	0.00
メタデータ検索(表題)	有	無	352	1	0.00	1.00	0.01
全文検索	無	有	97	1	0.01	1.00	0.02
メタデータ検索(表題)	無	有	47	1	0.02	1.00	0.04
メタデータ検索(カテゴリ)	無	有	81	1	0.01	1.00	0.02
全文検索	有	有	105	1	0.01	1.00	0.02
メタデータ検索(表題)	有	有	75	1	0.01	1.00	0.03

上部の画像情報を利用しない検索方式の精度は 7.1.3.3 と同値である。一方、表下部に示す作成年月日の制約を加えた検索では、正解数に変化はないが検索結果数が絞

り込まれるため適合率が高くなっている。結果として F 値が向上し、本検索の目的であった分析対象史料の絞り込みが確認できる。

7.2 既存の歴史研究プロセスの課題点に対する評価

本節では 3.2 で挙げた歴史研究プロセスの課題点について、提案手法および開発したシステムの有用性の検証を行う。図 7-4 にその課題点を再掲する。

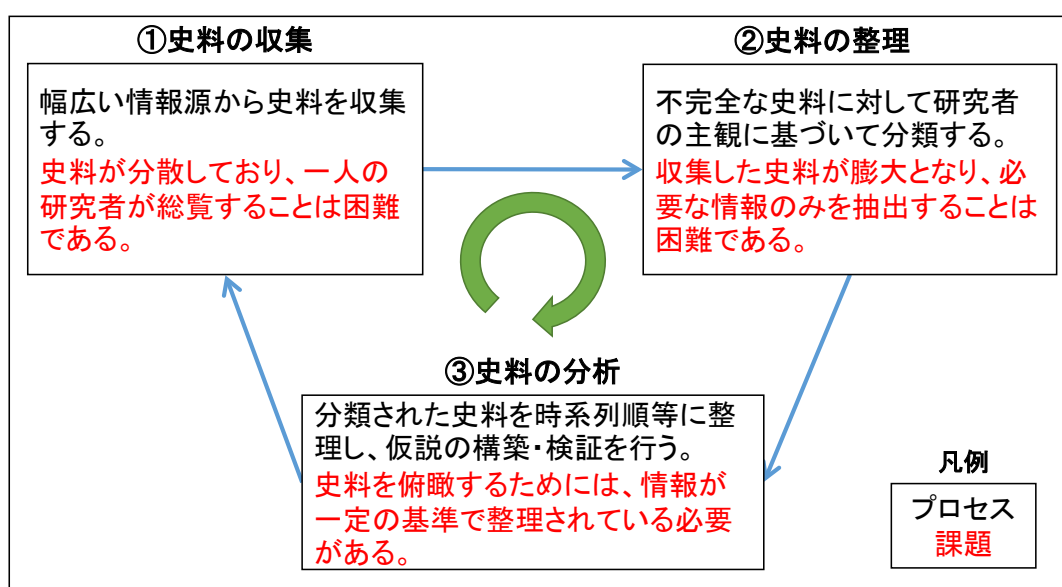


図 7-4 歴史研究プロセスの課題点の整理

7.2.1 「史料の収集」プロセスにおける評価

史料の収集プロセスでは幅広い情報源から史料を収集する必要があるが、現在の課題として史料が分散しており、一人の研究者が総覧することは困難である点が課題として挙げられている。この課題点に対して、本研究では以下に示す手法を提案した。

- デジタルアーカイブの活用
- 複数研究者による史料収集

デジタルアーカイブの活用については、システムをウェブアプリケーションとして

構築することで、他のデジタルアーカイブとの親和性を図った。具体的には 4.3.1 で説明したメタデータ登録機能においてデジタルアーカイブで公開されている史料の URL を参照することによって、容易に史料を研究者のリポジトリに蓄積できる環境を構築した。このようにデジタルアーカイブで公開されている史料群についてはその利点を最大限生かすことで効率的な史料収集が可能となった。

次に複数研究者による史料収集については、開発したシステム上で複数研究者がリポジトリを共有可能な環境を構築した。またそのために史料の公開範囲、メタデータの公開範囲や編集権限を設定可能な史料管理手法およびシステムを開発することで、各研究者の新規性の保護や各情報の信憑性を確保した形での情報共有が可能となった。第 6 章の実験では、簿冊一冊当たり 90 分程度の時間を要する史料収集プロセスを三名で分担することで、10 冊の簿冊から 500 件弱の史料を三日間という短期間で収集することができた。これにより従来の研究者個人による研究では分析対象史料が 50 件弱であった研究成果に対する量的な分析が可能となり、また研究者の史料収集における労力の軽減が確認できた。

一方、本研究では同一の研究目的を有する四名の研究者による実験による評価であるため、史料の公開範囲やメタデータの編集権限といった機能については十分な評価を行うことができなかった。今後は異なる研究グループが混在するような大規模な研究者グループによる歴史研究に本システムを適用し、提案手法の有効性を検証する必要がある。

7.2.2 「史料の整理」プロセスにおける評価

史料の整理プロセスでは収集した不完全な史料に対して研究者の主観に基づいて分類する必要があるが、史料数が膨大になるに従い、必要な情報のみを抽出することは困難であるといった課題が存在した。この課題に対して、本研究では以下に示す手法を提案した。

- セマンティック・ウェブ技術による史料情報の記述手法
- プラグインを用いたメタデータの自動付与

従来は 1.2 で説明したように史料とその書誌情報や調査結果は独立して蓄積されていた。これにより史料の閲覧と調査結果等の蓄積は別プロセスとして行う必要があり、調査結果等に基づいた史料の再検索は容易ではなかった。これに対して、本研究では

各史料に URI を与え、RDF を用いて史料情報を管理する手法を提案することで、史料情報がメタデータとして史料に直接紐づく形での整理が可能となった。さらに 4.4.1 で説明したメタデータ編集インタフェースを用いることで、史料の閲覧と同時にメタデータを編集することが可能となり、史料整理プロセスの省力化が実現した。本件に関しては歴史研究者へのヒアリングから、従来は史料数が膨大になった際に史料の所在の忘却や史料情報との関連づけの漏れ等が多く生じ、量的な分析は困難であったが、開発したシステムを利用することにより、これらの問題が改善されたという意見が得られた。

プラグインを用いたメタデータの自動付与については、第 5 章のケーススタディにおいて画像からエッジ情報を抽出し、形状が類似する画像情報をメタデータとして付与した。このメタデータを利用し、ユーザに類似画像を表示することで、膨大な史料を逐次確認する労力を軽減し、また 7.1 の結果から検索精度の向上が確認された。今後は文字認識を利用した史料画像からの文字列抽出や、類似するメタデータを持つ史料の機械学習による自動分類などを適用し、従来人手で行っていた分類における労力や情報量の比較による有効性の評価を行う。

7.2.3 「史料の分析」プロセスにおける評価

史料の分析プロセスでは分類された史料を時系列順等に整理し、仮説の構築・検証を行う必要があるが、史料を俯瞰するためには情報が一定の基準で整理されている必要があり、この点が従来の歴史研究における課題であった。この課題に対して、本研究では以下に示す手法を提案した。

- メタデータの利用に基づいた史料の分析
- 複数研究者による協調的な分析

従来は Microsoft Word や Excel に目録や調査結果を自由記述の形で記述しており、7.2.2 でも述べたように史料との紐づけも手動で行う必要があった。提案手法では調査結果等は史料にメタデータとして蓄積され、各メタデータはフィールドと値のペアとして記述されるため、情報の記述基準の統一化を実現できた。その結果、これらのメタデータを検索や並び替え等で利用することで史料分析を支援する環境を構築した。第 5 章のケーススタディでは、ユーザが付与したメタデータを検索に利用することによって、図面史料の煙突本数に関する情報といった、ユーザの目的に合わせた史料の

絞り込みが可能となった。また史料に付与されたメタデータに基づいて史料の並び替えや比較を行うことで、史料分析を支援することができた。さらに第 6 章の実験では、蓄積されたメタデータに対してプラグインを適用することで、膨大な史料から必要に応じた情報を機械的に抽出することが可能となった。この結果、史料や調査結果の出現頻度等を可視化することで俯瞰的な分析が可能となった。

また複数研究者が史料や調査結果を共有し、史料分析を共同で行うことが可能な環境を構築した。これにより、第 6 章の実験では歴史研究者と情報技術者という異なる分野の研究者による協調的な研究環境が実現できた。具体的には歴史研究者が立案した仮説に対して、情報技術者が定量的な検証を行う研究スタイルの有効性が示唆された。また情報技術者が蓄積された史料をグラフ等で可視化することによって、歴史研究者が見落としていた情報を想起させることが可能となった。これらの評価から従来の研究者個人で行われていた分析と比較して、本提案手法が有用に機能することが示唆された。今後はより多くの研究者を対象とした評価を行い、提案手法による効果に関するより詳細な分析を行う。

7.2.4 提案手法に基づく歴史研究プロセスの改善

7.2.1 から 7.2.3 で行った各プロセスの考察に基づき、既存の歴史研究プロセスと提案する歴史研究プロセスの比較を図 7-5 に示す。

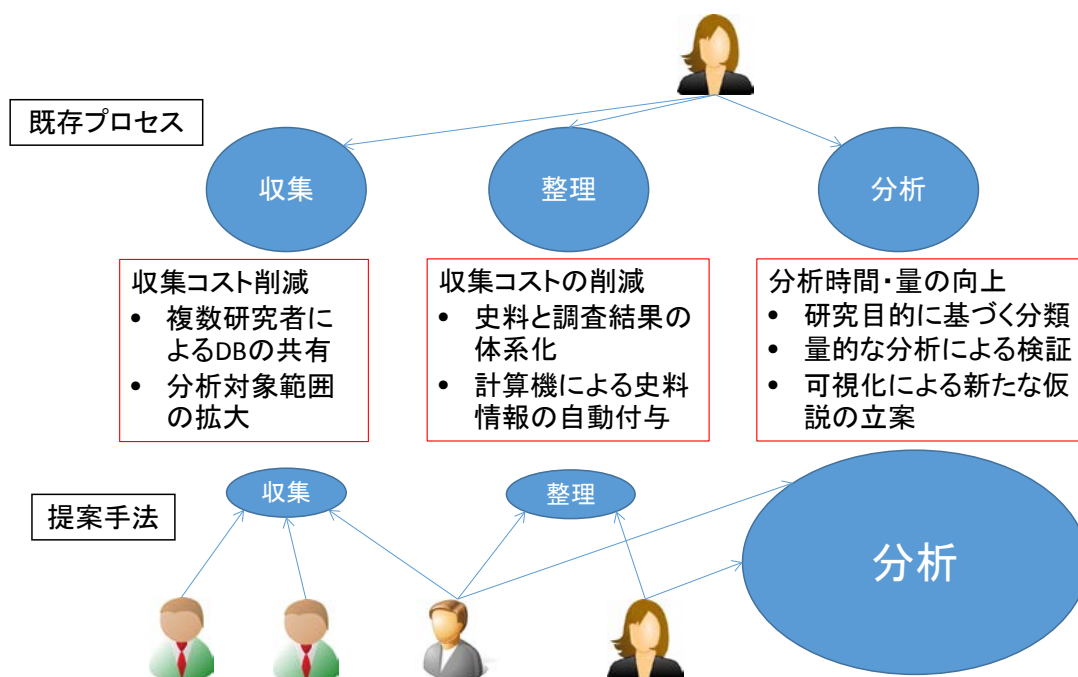


図 7-5 提案手法に基づく歴史研究プロセスの改善

本研究では複数研究者による歴史研究を支援する手法の提案及びシステムを開発することによって、歴史研究者の史料の収集・整理プロセスの労力を低減し、分析により多くの時間を割くことが可能となった。これまでの考察と重複するが、史料の収集プロセスでは複数研究者で史料収集プロセスを分担し、各研究者のDBを共有することによって、分析対象範囲の拡大が可能となった。また史料の収集に関してはセマンティック・ウェブ技術によって史料情報を記述することにより、史料と調査結果の体系化による調査結果の蓄積の効率化や、調査結果に基づいた史料の再検索時の手戻りを削減することができた。史料の分析プロセスでは、メタデータとして史料に付与された調査結果を用いることで、研究に目的に応じた史料分析が可能となった。さらに歴史研究者と情報技術者が史実に対する仮説の立案と検証を相互に行うことで、協調的な史料分析環境を実現することができた。

一方、このような歴史研究環境を実現するには、情報技術者が歴史研究に参画する必要がある、そのインセンティブを与える必要がある。この点に関しては、近年注目されているクラウドソーシングの適用が考えられる。このイメージを図 7-6 に示す。

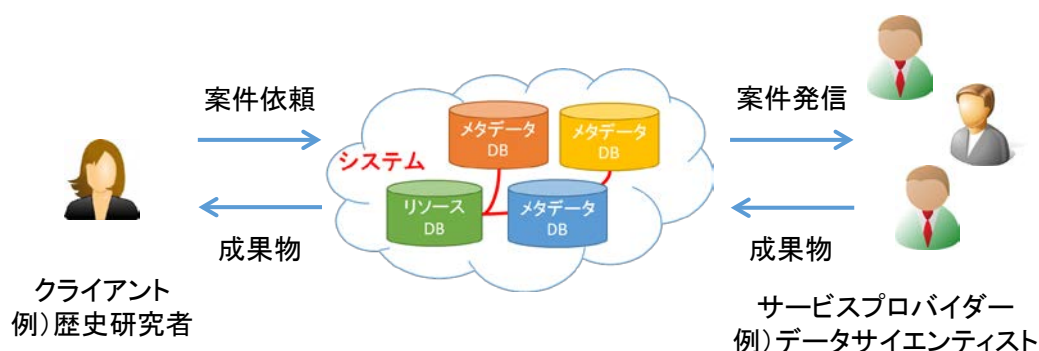


図 7-6 クラウドソーシング

クラウドソーシングとはインターネットを利用し、不特定多数の人への業務の発注や受注者の募集を行うウェブサービスである。これは本研究で提案するクラウド上で史料情報を管理する環境、およびプラグインを用いた計算機による史料分析環境との親和性が高いと考えられる。歴史研究者が調査結果をクラウド上に蓄積する。そして統計等の歴史研究者にとっては馴染みのない数理的な分析が必要となった際、情報技術者等のサービスプロバイダーに業務として発注、またはそれに代替されるインセンティブを提示する。そして受注した情報技術者、データサイエンティストが指定された DB の情報を分析するプラグインを開発することにより、異なる専門を持つ研究者または人々による協調的な研究環境の実現に寄与することが可能となる。

7.2.5 来電における標準メタデータセットの提案

第 6 章の実験では、外交史料館が保有する来電を収集し、表題や作成日等の書誌情報の他に、来電の送付先や閲覧者の情報をメタデータとして付与した。このメタデータを利用することで分類や比較等の分析が可能となり、来電の送付先決定者の人物特定を行った。一方、6.6 で述べたように、これらの来電の一部はアジア歴史資料センターで公開されている史料であるが、これらの来電には目録といった基本的な書誌情報しか付与されておらず、本実験で行ったように筆跡の照合による人物特定や閲覧者の分布に基づいた分析等は不可能である。

よって本研究では、本実験で来電分析における有用性を確認した表 6-2 に示したメタデータフィールドを、図 7-7 に示すように来電における標準メタデータセットとし

て提案する。

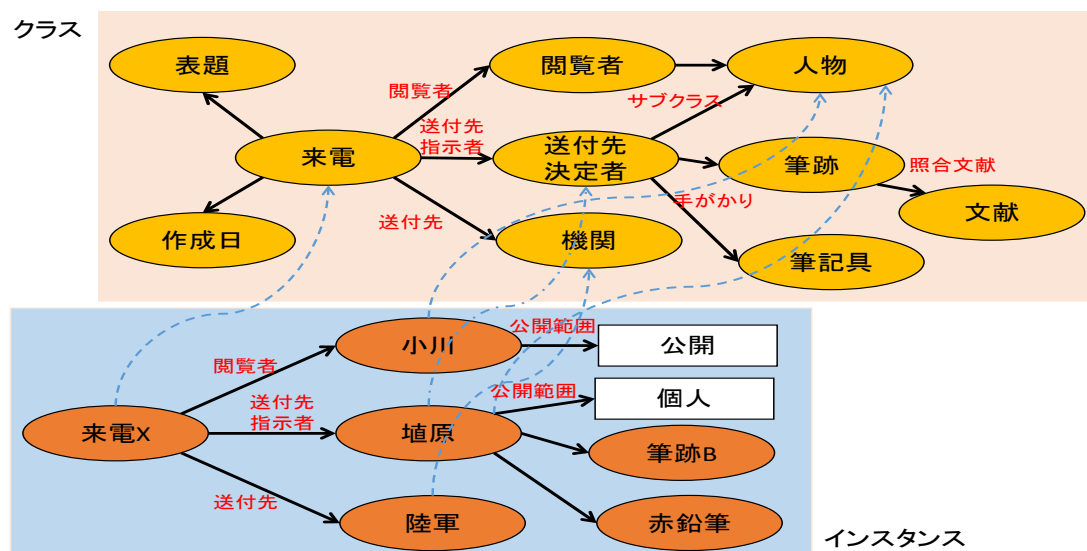


図 7-7 来電に関する標準メタデータセット例

これらが標準メタデータセットとして定義されることによって、新たに来電を収集した際に、史料内容からここで定義したメタデータフィールドの値を付与するプロセスへと単純化することができる。これにより複数の研究者やアーキビストが来電に関する書誌情報を付与する際にも、この標準メタデータセットを参照することで、研究者各々による付与方法の揺れ等を吸収することができ、史料へのメタデータの付与プロセスの効率化が期待できる。

さらに研究者が来電に付与するメタデータの意味（セマンティクス）を予め **RDF Schema** およびオントロジーを用いて定義することによって、付与された情報の意味を計算機が自律的に理解することが可能となる。例えば「来電の閲覧者を記述するプロパティ『閲覧者』のバリューはクラス『人物』のサブクラスである」と定義することにより、各研究者が入力した値を計算機が人物だと認識することができる。これにより従来の文字列に基づいた検索と比較し、セマンティクスに基づいた情報抽出が可能となる。また「送付先決定者の特定に寄与する情報は『筆跡』や『筆記具の種類』である」等の関係を定義することにより、計算機が自動的にそれらの情報を抽出し、関連概念や不足情報の提示といった自律的な分析が可能となる。今後は歴史研究者の

暗黙知を体系化し、計算機による半自動的な史料分析を可能とする環境構築を目指す。

第8章 結論

8.1 結論	117
8.2 今後の展望	117
8.2.1 「デジタルキュレーションシステム」への応用	117
8.2.2 他分野の文献調査プロセスへの応用	118

8.1 結論

本研究ではセマンティック・ウェブ技術を用いて歴史研究を支援するシステムを開発した。具体的には電子史料をクラウド環境で管理し、書誌情報や研究者の調査結果を史料のメタデータとして管理を行う手法を提案した。またそれらのメタデータを利用することで、史料の並び替えや比較といった分析を可能とする環境を構築した。さらに従来の歴史研究の属人性の問題に対して、複数研究者が史料の収集や整理、分析を共同で行うことが可能なシステムを開発した。

また開発したシステム上におけるケーススタディ及び実験によって提案手法の有効性を検証した。その結果、史料の収集や整理における各研究者の労力の軽減や効率化に寄与することが確認できた。また歴史研究者と情報技術者が共同で実験を行うことで、歴史研究者が立案した仮説に対する量的な分析や、データの可視化による新たな気づきの創出など、複数研究者による協調的な歴史研究環境を実現することができた。

8.2 今後の展望

今後の展望として、以下の二点を挙げる。

8.2.1 「デジタルキュレーションシステム」への応用

本研究では基本的なメタデータが予め付与された史料に対して、調査結果等をメタデータとして付与することで、研究者個人または複数研究者による歴史研究を支援することを目的とした。今後は本研究で提案した複数研究者による史料の収集や整理、プラグインを用いた計算機による史料情報の自動抽出技術を応用し、図 8-1 に示すようなデジタルアーカイブの構築を支援する「デジタルキュレーションシステム」を開発することを目指す。これは利用者やアーキビストが入力した史料に対して、計算機がセマンティック・ウェブ技術や画像処理、テキストマイニング等の技術によって史料情報を自律的に理解し、それらを書誌情報として自動付与するようなシステムである。これにより現在アーカイブ作業における課題である、人手による書誌情報(メタデータ)の労力を軽減し、デジタルアーカイブの普及や学術

の発展につなげていきたい。

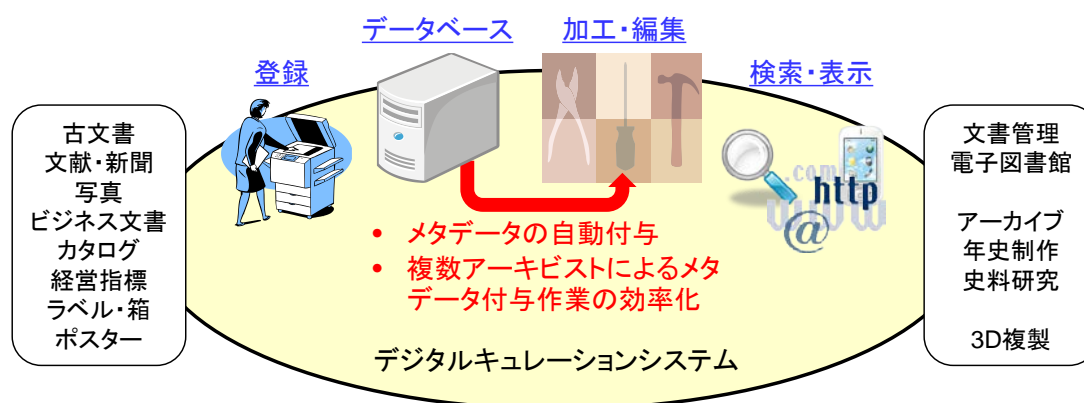


図 8-1 システムイメージ

8.2.2 他分野の文献調査プロセスへの応用

本研究では画像を主とした電子史料を活用することで、歴史研究を支援することを目的とした。今後は電子書籍を利用した読書プロセス、デジタル教材を利用した授業プロセス、企業内文書を利用した業務プロセスなどに提案手法を応用し、既存プロセスの改善や新たな価値を創出する環境の構築を目指す。

謝辞

本研究を進めるにあたって、多くの方々にご助力を頂きました。ここに感謝の意を述べさせていただきます。

東京大学大学院人文社会系研究科日本文化研究専攻 教授 鈴木淳先生、修士一年の吉田ますみ様には、提案手法や開発したシステム、歴史研究プロセス等について、貴重なご助言をいただきました。特に吉田様には本研究の実験に関して多大なご協力、ご支援をいただきました。心より深く感謝いたします。

東京大学大学院新領域創成科学研究科人間環境学専攻 教授 大和裕幸先生には大変ご多忙の中貴重な時間を割いていただき、研究の方針はもとより、一般教養、学生としての心構え等、様々なご指導をいただきました。厳しくも真に学生のことを考えてのアドバイスは、どれも深く心に残る言葉ばかりでありました。心から深く感謝いたします。

東京大学大学院新領域創成科学研究科人間環境学専攻 准教授 稗方和夫先生には、特に技術面においてもっとも多くの具体的な有意義なアドバイスをいただきました。加えて、研究に対する姿勢、プログラミング作法、論文の書き方、仕事に対する心構えなど様々な面でご指導、ご教示いただきました。心より深く感謝いたします。

電気通信大学大学院情報理工学研究科 教授 増田宏先生、東京大学大学院工学系研究科システム創成学専攻 准教授 白山晋先生には設計工学研究室大輪講などにおいて、的確で具体的なアドバイスをいただきました。大変感謝いたします。

Global Project Design.inc の Bryan R Moser 博士には、インターンシップを通じて有意義な議論させていただきました。心より深く感謝いたします。

東洋文庫の牧野元紀様には、東洋文庫における史料の管理手法や歴史研究における

必要要件の定義等、多くの有用な情報及び助言をいただきました。深く感謝いたします。

Yahoo! Japan 研究所の坪内孝太博士には、研究室の教員として、また頼れる先輩として、公私ともに大変お世話になりました。時には厳しく時には優しく私に指導していただき、多くのことを学ばせて頂きました。ありがとうございました。

東京大学大学院工学系研究科システム創成学専攻 技術専門員 榎本昌一様には、ネットワーク、サーバをはじめとした様々な情報技術に関するサポートをいただきました。深く御礼申し上げます。

水野淳子様、山本和子様、谷村育子様、上村美江子様、山田祐子様には、研究に専念できるように膨大な数の煩雑な事務手続きを行っていただきました。また、生活面においても、常に優しく接していただきました。本当にありがとうございました。

共同研究員である、富士通株式会社の岡田伊策様、齋藤稔様、笈田佳彰様とは、研究室の勉強会や研究のディスカッションなどで大変お世話になりました。ビジネス視点でのお話や情報技術に関する最先端のお話などは大変に興味深く、積極的にディスカッションさせていただきました。本当にありがとうございました。

研究室の卒業生である柳澤龍様、于佰鑫様、鹿渡俊介様、木村彰吾様、杉浦孝光様、長谷川嵩様、曹樺楠様、三浦慎也様、砂川辰徳様、舩島一樹様、研究室の先輩である、金載烈様、満行泰河様、孫晶ギョク様には研究とプライベートの両面で大変お世話になりました。様々な思い出を共有できたことは忘れることのできない思い出です。本当にありがとうございました。

研究室の後輩である、石黒慧様、河野裕様、廣井貴彬様、深田直人様、黒川達也様、角野為耶様、津村和輝様、宮崎邦洋様、安藤早紀様、齋藤智輝様、松原洸也様、和中真之介様には日頃の研究室生活において、多大なるサポートを受けました。ありがとうございました

ました。

東京大学工学院工学系研究科システム創成学専攻に所属していた設計工学研究室の皆様とは、設計研共通のイベントを通して研究に関する活発な議論をすることができました。

最後になりましたが、研究生生活を支えてくれた家族に感謝いたします。本当にありがとうございました。

参考文献

- [1]. Ur-Rahman, N., and Harding, J.A.: Textual data mining for industrial knowledge management and text classification: A business oriented approach, Expert Systems with Applications, Vol.39, No. 5, pp. 4729-4739, 2012.
- [2]. e-文書法の施行について, Available at
<<http://www.kantei.go.jp/jp/singi/it2/others/e-bunsyoun.html>>. Accessed on : Dec 20th 2013.
- [3]. Takeuchi, H., Venkata, L., Nasukawa, T., and Roy, S.: Getting insights from the voices of customers: Conversation mining at a contact center, Information Sciences, Vol. 179, No. 11, pp. 1584-1591, 2009.
- [4]. 黒瀬能幸, 矢野米雄, 富田 豊: インターネットを利用した 3 次元 CAD 学習支援システムの構築, 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, 80 巻, 4 号, pp. 933-942, 1997.
- [5]. 小早川倫広, 星守, 大森匡, 照井武彦: ウェーブレット変換を用いた対話的類似画像検索と民俗資料データベースへの適用 (<特集>人文科学とコンピュータ), 情報処理学会論文誌, 40 巻, 3 号, pp. 899-911, 1999.
- [6]. 外務省, オープンデータ憲章 (概要), Available at
<http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html>. Accessed on : Dec 20th 2013.
- [7]. データシティ鯖江, Available at
<<http://www.city.sabae.fukui.jp/pageview.html?id=11552>>. Accessed on : Dec 20th 2013.
- [8]. 月尾嘉男, デジタル・アーカイブの功罪, Available at
<<http://www.tsukio.com/denki2.html>>. Accessed on : Dec 20th 2013.
- [9]. e-Japan 戦略 II, Available at
<<http://www.kantei.go.jp/jp/singi/it2/kettei/030702ejapan.pdf>>. Accessed on : Dec 20th 2013.
- [10]. 平賀譲デジタルアーカイブ, Available at

-
- <<http://rarebook.dl.itc.u-tokyo.ac.jp/hiraga/>>. Accessed on : Dec 20th 2013.
- [11]. 川瀬晃：日本海軍における電気溶接の艦艇への採用について，海事史研究，第 67 号，pp. 37-55, 2010.
- [12]. 川瀬晃：平賀譲と第四艦隊事件，海事史研究，第 68 号，pp. 40-54, 2011.
- [13]. 学研パブリッシング：帝国海軍の礎 八八艦隊計画（歴史群像シリーズ），学習研究社，2011.
- [14]. 堀内カラー，Available at <<http://www.horiuchi-color.co.jp/>>. Accessed on : Dec 20th 2013.
- [15]. アジア歴史資料センター，Available at <<http://www.jacar.go.jp/>>. Accessed on : Dec 20th 2013.
- [16]. Berners-Lee, T., Hendler, J., and Lassila, O.: The semantic Web, Scientific American, Vol.284, No. 5, pp. 34-43, 2001.
- [17]. Semantic Web "Layer Cake", Available at <<http://www.w3.org/2007/03/layerCake.png>>. Accessed on: Dec 20th 2013.
- [18]. 松尾豊，Web 上の情報からの人間関係ネットワークの抽出，人工知能学会論文誌，20 卷，5 号，pp. 337-345, 2005.
- [19]. Breitman, K.K.: Ontology as a Requirements Engineering Product , 11th IEEE International Requirements Engineering Conference (RE'03), pp. 309-319, 2003.
- [20]. Fox, M.S.: An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behavior, Computers in Industry, Vol.29, pp. 123-134, 1996.
- [21]. Hiekata, K., Yamato, H., and Tsujimoto, S.: Ontology Based Knowledge Extraction for Shipyard Fabrication Workshop Reports, Expert Systems with Applications, Vol. 37, No.11, pp. 7380-7386, 2010.
- [22]. Hafner, J., Sawhney, H.S., Equitz, W., Flickner, M., and Niblack, W.: Efficient color histogram indexing for quadratic form distance functions, Pattern Analysis and Machine Intelligence, Vol.17, No.7, pp. 729-736, 1995.
-

-
- [23]. Chatzichristofis, S. A., and Boutalis, Y. S.: CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval, the Proceedings of the 6th international conference on Computer vision systems, Vol.5008, pp. 312-322, 2008.
- [24]. Manjunath, B.S., Ohm, J.-R., Vasudevan, V.V., and Yamada, A.: Color and texture descriptors, Circuits and Systems for Video Technology, Vol.11, No.6, pp. 703-715, 2001.
- [25]. Yang, J., Jiang, Y.-G., Hauptmann, A. G., and Ngo, C.-W. : Evaluating bag-of-visual-words representations in scene classification, the Proceedings of the international workshop on Workshop on multimedia information retrieval, pp. 197-206, 2007.
- [26]. Tirilly, P., Claveau, V., Gros, P.: Language modeling for bag-of-visual words image categorization, the Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 249-258, 2008.
- [27]. Tobin, and Kenneth, W.: Using a patient image archive to diagnose retinopathy, Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp. 5441-5444, 2008.
- [28]. Park, D. K., Jeon, Y. S., and Won, C. S.: Efficient use of local edge histogram descriptor, the Proceedings of the 2000 ACM workshops on Multimedia, pp. 51-54, 2000.
- [29]. Mpeg-7 - Standards | MPEG, Available at
<<http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>>.
Accessed on : Dec 20th 2013.
- [30]. González, J., Villena, J., Moreno, C., and Martínez-Fernández, J.: Semiautomatic Extraction of Thesauri and Semantic Search in a Digital Image Archive, Proceedings of the 10th International Conference on Electronic Publishing, pp. 279-290, 2006.
- [31]. Castells, P., Perdrix, F., Pulido, E., Rico, M., Benjamins, R., Contreras, J., Lorés, J., and Neptuno, F.: Semantic Web Technologies for a Digital
-

-
- Newspaper Archive, Springer, pp. 445-458, 2004.
- [32]. Costilla, C.: Semantic Web digital archive integration, Database and Expert Systems Applications, pp. 179-185, 2004.
- [33]. 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 (<特集>人文科学とコンピュータ), 情報処理学会論文誌, 40 巻, 3 号, pp. 821-830, 1999.
- [34]. 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 (<特集>人文科学とコンピュータ), 情報処理学会論文誌, 40 巻, 3 号, pp. 831-839, 1999.
- [35]. 研谷紀夫, 馬場章: 文化資源オントロジの構築とその活用(<特集>第 15 回(2007 年度)年次大会(研究報告会&総会)), 情報知識学会誌, 17 巻, 2 号, pp. 129-134, 2007.
- [36]. 渡邊英徳, 坂田晃一, 北原和也, 鳥巢智行, 大瀬良亮, 阿久津由美, 中丸由貴, 草野史興: "Nagasaki Archive": 事象の多面的・総合的な理解を促す多元的デジタルアーカイブズ, 日本バーチャルリアリティ学会論文誌, 16 巻, 3 号, pp. 497-505, 2011.
- [37]. 稗方和夫, 大和裕幸, 中村覚, 岡田伊策, 齋藤稔, 安藤峻: インシデントレポートの自動分類とその分析環境の構築と評価, 人工知能学会全国大会論文集(CD-ROM), 27 巻, 2013.
- [38]. 中村覚: セマンティックウェブ技術を用いたデジタルアーカイブ史料研究支援システムの構築, 卒業論文, 2011.
- [39]. 三浦崇, 伊東幸宏, 小西達裕, 田村貞雄, 赤石, 美奈, 中谷広正, 阿部圭一: 歴史学研究支援システムの構築”, 情報処理学会研究報告, 97 巻, 48 号, pp. 25-30, 1997.
- [40]. Dublin Core Metadata Initiative, Available at <<http://dublincore.org/>>. Accessed on: Dec 20th 2013.
- [41]. 長門 (戦艦) , Available at <[http://ja.wikipedia.org/wiki/%E9%95%B7%E9%96%80_\(%E6%88%A6%E8%89%A6\)](http://ja.wikipedia.org/wiki/%E9%95%B7%E9%96%80_(%E6%88%A6%E8%89%A6))>. Accessed on : Dec 20th 2013.
- [42]. 加賀 (空母) , Available at
-

- <[http://ja.wikipedia.org/wiki/%E5%8A%A0%E8%B3%80_\(%E7%A9%BA%E6%AF%8D\)>](http://ja.wikipedia.org/wiki/%E5%8A%A0%E8%B3%80_(%E7%A9%BA%E6%AF%8D)>)>. Accessed on : Dec 20th 2013.
- [43]. 夢の八八艦隊 , Available at
<http://www.geocities.jp/kigiken/88fleet_1.html>. Accessed on : Dec 20th 2013.
- [44]. Lire (Lucene Image REtrieval), Available at
<<http://www.lire-project.net/>>. Accessed on : Dec 20th 2013.
- [45]. 外務省文書 (MF) , Available at
<<http://rnavi.ndl.go.jp/kensei/entry/Gaimusho.php>>. Accessed on : Dec 20th 2013.
- [46]. JACAR(アジア歴史資料センター)Ref.B06150296900、五国会議 第三卷 (B-2-3-1-055-00-003)(外務省外交史料館)
- [47]. 外務省: 外務省年鑑: 大正 11 年, クレス出版, pp. 292, 1999.