

平成30年度修士論文題目一覧

統合新領域学府ライブラリーサイエンス専攻修士課程

学位記番号	申請学位	学生氏名	研究課題	論文公開
479	修士 (ライブラリーサイエンス)	ヤマグチ タカオ 山口 隆夫	ウェブによる図書館の情報発信と図書館利用の関係	公開
522	修士 (ライブラリーサイエンス)	ハク シュンハ 白 春波	九州大学所蔵宮崎文庫における賃金資料および労働組合資料 一戦後日本の賃金交渉過程解明という視点から	公開
520	修士 (ライブラリーサイエンス)	シュウイ テツ 周 緯哲	中国における汪兆銘国民政府関係文書に関する研究	非公開
521	修士 (ライブラリーサイエンス)	トウ イチヘイ 唐 一平	くずし字認識手法とそのためのデータセット構築	公開

ウェブによる図書館の情報発信と図書館利用の関係

Relationship between Library Information Dissemination on the Web and Library Usage.

2FS16206T 山口 隆夫 YAMAGUCHI Takao

公共図書館の広報活動の手段として、ウェブによる情報発信が増加している。また、人々が情報を入手するためにインターネットを利用する時間は年々増加傾向にある。図書館はウェブを利用することにより、不特定多数の利用者に対する能動的な情報発信が可能となった。しかし、ウェブによる図書館の情報発信と図書館利用との関係は不明であり、ウェブによる図書館の広報活動は、図書館利用を促進するのかという問題が提起される。

本研究の第一の目的は、ウェブによる情報発信が図書館利用に影響を与えているかを検証することである。図書館評価において、「図書館利用」はアウトプットと捉えられ、インプットは図書館の「経営資源」である。そこで、アウトプットだけでなく、インプットに着目して、図書館の経営資源がウェブでの情報発信に影響を与えているかを検証することも本研究の目的である。さらに、運営形態（直営か指定管理者か）は経営資源や図書館利用にも影響を与えているため、運営形態の違いが情報発信に影響を与えているかを検証することも目的とする。

図書館利用と経営資源の指標として、図書館の業務統計を使用し、図書館利用の指標には貸出冊数と来館者数を、図書館の経営資源の指標には職員数、蔵書冊数を用いた。具体的には以下の仮説を立てた。

仮説(1) 図書館利用：情報発信をしている図書館の来館者数、貸出冊数は多い。

仮説(2) 図書館の経営資源：職員数、蔵書冊数の多い図書館は、情報発信が盛んである。

仮説(3) 運営形態：指定管理者制の図書館は、直営の図書館に比べ情報発信を盛んに行っている。

情報発信の実態調査は、全国 791 市及び東京 23 特別区の公共図書館 805 館のウェブサイトを対象にした。情報発信の調査項目として、「図書館だより」「子どものページ」「ブックリスト」「おはなし会」を選定した。2017 年 8 月から 11 月にかけて、情報の掲載の有無を一館ずつアクセスして確認した。情報発信の現況を集計した結果、全国的におはなし会の情報発信の割合が 88.4%と高く、ブックリストは 28.6%で低いこと、また、地域別にみると関東など大都市のある地域の方が、情報発信の実施割合は高かった。業務統計のデータとして、日本図書館協会による全国統計から、調査対象館の奉仕人口、貸出数計、職員数計、来館者数、蔵書冊数計を抽出し、人口や地域差による偏りを正規化できるよう、それぞれの数値を奉仕人口で割った「密度」を算出した。さらに、同協会の全国調査から、調査対象館のうち、指定管理者制度の導入館を特定した。これら

のデータをもとに、仮説(1)と(2)について、情報発信を実施している図書館と実施していない図書館の業務データに関して、Wilcoxon の順位和検定により中央値の差の検定により分析した。また、仮説(3)について、直営と指定管理者の情報発信の実施割合を、2 群の母比率の差の検定により分析した。

仮説(1)では、図書館だより、おはなし会の情報発信が、来館者数に関して有意な差が見られ、これらの情報発信が来館者数増に影響している可能性がある。図書館だよりは、幅広い世代への広報となるため、多くの利用者が来館する可能性を高めることや、定期的な図書館に関する情報提供が継続的な来館につながることで、子ども向けのおはなし会に関する広報は、親子連れなどの複数名での来館につながることを考えられる。また、子どものページ、おはなし会の情報発信が、貸出冊数に関して有意差が認められ、貸出冊数増に影響している可能性がある。子どものページのイベント紹介を見て来館した親子連れなどが本を借りたり、おはなし会で紹介された本を参加者が借りたりする可能性が考えられる。仮説(2)では、職員数は、ブックリストの情報発信との有意差がみられ、ブックリストの情報発信増に影響している可能性がある。職員数が少なく規模が小さい図書館の方が、蔵書に基づくブックリスト作成と情報発信が比較的容易なことが推測される。また、蔵書冊数が、図書館だより、子どものページ、ブックリストの情報発信に影響している可能性がある。蔵書冊数が少なく規模が小さい図書館の方が、自館の広報誌、児童サービスのポータルとしての子どものページ、蔵書に基づくブックリストの作成と情報発信が比較的容易な可能性を示唆している。仮説(3)の運営形態では、直営の方が指定管理者の館より、ブックリストによる情報発信を行っている館の割合が有意に高い。ブックリスト作成には、長年の経験に基づく選書能力が要求されるため、短期雇用が多い傾向がある指定管理者よりも直営の館の方が盛んな可能性がある。また、指定管理者の館の方が直営の館より、図書館だよりによる情報発信をしている館の割合が有意に高い。掲載内容やレイアウトが定型化された図書館だよりの情報発信は、情報を更新する単純作業に近いので、比較的経験が少なくても対応可能であることや、指定管理者の方が、集客目的に図書館だよりを活用する傾向が強いことが考えられる。

分析の結果、貸出冊数増に関しては、子どものページが、来館者数増に関しては、図書館だよりが影響している可能性があるという点で異なった。今回の調査からはその理由は把握できないため、今後の課題としたい。

九州大学所蔵宮崎文庫における賃金資料および労働組合資料—戦後日本の賃金交渉過程解明という視点から—

Wage Materials and Labor Union Materials in the Miyazaki Collection at Kyushu University: From the Viewpoint of Clarification of the Process of Wage Negotiations in Postwar Japan

2FS17201K 白 春波 BAI Chunbo

記録史料は公文書だけではなく、むしろ、私文書を含めた民間資料の方が多い。私文書として、たとえば、公害裁判資料と労働資料が挙げられる。現在、九州大学記録資料館が所蔵する宮崎文庫は、個人が集めた私文書である。

宮崎文庫は、1946年から1973年までの石炭企業の定期大会記録、賃金関係資料、九州炭鉱労働関係資料、生産関係資料、鉱業所案内、定期総会資料、機関誌などを含み、多くの分野をカバーしている。資料の形態は、組合旗、腕章、看板、パンフレットなど多様である。特に、賃金に関する資料が充実している。さらに、この年代は、戦後のインフレ期と戦後復興期、高度成長期という、戦後日本の賃金に最も大きな変化が起こった時期と重なる。それらの資料は歴史上の賃金決定の研究にとっても、現在あるいは将来の賃金に関する研究にとっても、重要な材料になると考える。また、宮崎文庫は戦後における日本の石炭鉱業の歴史、および企業と労働組合の交渉による賃金決定過程を明らかにするうえで、重要な一次資料に位置付けられる。

こうした背景のもと、本研究では、九州大学が所蔵する宮崎文庫の由来と特徴、全体像に焦点を当て、その実態を明らかにする。さらに、宮崎文庫に含まれる賃金資料の内容および労働組合資料をもとに、戦後日本における石炭産業の賃金が企業と労働組合の交渉によってどのように決定されたかということに関する研究に有益である資料であることを示す。最後に、その価値を示したアーカイブズとしてのこれらの資料が、今後どのように活用され得るのかについて私見を述べたい。

研究の手順として、まず、先行研究や文献調査に基づき、日本の賃金に関する詳細な分析を行った。次に、宮崎文庫の由来、特徴、全体像、九州大学に移管された経緯と理由を明らかにした。そのうえで、宮崎文庫における賃金関係資料と労働組合資料を研究する重要性を示した。また、宮崎文庫の賃金関係資料および労働組合資料から得られた知見を先行研究と比較し、分析を行った。

事例研究としては、宮崎文庫のなかの大企業、労働組合、中小企業の資料の代表例である日炭高松資料、全炭鉱（全国石炭鉱業労働組合）資料、志岐炭鉱資料に分析対象を絞った。賃金関係資料および労働組合資料を取り出し、石炭鉱業と炭労および各炭鉱の労組との交渉による賃金決定過程について詳細な分析を行った。

日炭高松資料に関して、まず、選択した理由をアーカイブズ学の視点から詳しく説明した。そして、日炭高松資料のなかのアイテムから、賃金に関する資料の年代は戦後日本の賃金に最も大きな変化が起こった時期と重なること

が分かった。さらに、団体交渉のうちの対角線交渉に絞ることができ、賃金闘争の実態を明確にした。

戦後の電産型賃金体系の導入の影響で、日本の各産業の賃金体系は大きく変わった。石炭産業の賃金体系も、1946年10月に電産型賃金体系に基づいて、決定されていた。また、日炭高松資料から、戦後復興期に入った後、石炭産業の賃金体系は電産型賃金体系と同様に、生活給賃金から能率給賃金へ移行したことが明らかになった。

宮崎文庫のなかの全炭鉱資料を分析した結果、全炭鉱が、石炭産業における重要な労働組合として、炭鉱の賃金決定、労働条件の改善などに不可欠な役割を果たしたことは間違いない。全炭鉱資料のなかでは、昭和30年代前半の資料が最も多い。そして、記録されている賃金決定基準も多様である。経済的環境はその基準の一つである。全炭鉱資料のなかに残された、昭和30年代の日本国内の主要エネルギー供給の推移に関するデータから、経済的環境が石炭産業における賃金決定に大きな影響を与えたことが分かった。

志岐炭鉱資料から大企業と中小企業間の賃金格差に関する問題を明らかにした。しかし、大企業と異なり、志岐炭鉱の鉱員賃金に関して、基準内賃金の本人給は、請負給、職務給、固定給、奨励給、資格給という5つの下位項目を含んでいる。賃金決定に関しては生活給の思想から職務、能率への転換が進み、賃金決定の要素が増えたことが明らかになった。中小企業の労働組合は、賃金の固定給が低い反面、奨励給や資格給を活用して大企業との賃金の格差を縮めようとしていたと考える。

収集アーカイブズの最も重要な目的は、研究に有益なデータを提供し、証拠としての過去のことやそれから得られる知見を把握すること、そして、その目的を認識したうえで公開して利用に供することである。本論で展開してきた、宮崎文庫における賃金に関わる資料および労働組合資料の分析は、収集アーカイブズの目的を明確化する行為である。このような行為は、アーカイブズ資料の利活用のためには不可欠である。

本論で扱った宮崎文庫は、労働組合の資料であり、賃金に関する記録資料がたくさん残っている。それらの資料を分析することで、当時の労働運動やその状況下で決定された賃金の実態を明らかにできる。宮崎文庫の一次資料で戦前、戦中、戦後の日本の賃金制度の変遷を跡づけることができる。

くずし字認識手法とそのためのデータセット構築

Recognition Method of Kuzushi Characters and Construction of Characters Segmentation Dataset

2FS17203M 唐 一平 TANG Yiping

近年、情報技術は、情報系分野だけでなく、様々な分野に応用されてきた。例えば、膨大な情報を持つ人文学の分野において、手書き文字の認識や古文書に記載された文字の認識などに応用されており、文字認識技術はますます情報のデジタル化の重要な要素技術になっている。

現時点では、手書きアラビア語や手書き英語の認識は既に実用レベルになっているが、手書き漢字の認識はまだ実用レベルに達していない。さらに、くずし字という日本の古い文字の認識は手書き漢字より難しいといえる。そのため、現在のくずし字の認識はまだ人手で行われているのが現状である。

本論文では、古文書のくずし字で書かれたひらがなを対象とし、認識精度をより高くする二つの手法を提案する。

既存手法として2017年に開催されたくずし字チャレンジというアルゴリズムコンテストの最優秀賞チームの手法があげられる。最優秀賞チームの手法は長さ可変のスライド窓を用いて可能な区切りの候補の枠を多数作成する。その後、各区切り候補枠で囲った内容を、それぞれに単一文字画像認識ネットワークに入力し、出力される各文字種の予測結果(確率)と比較することにより、確率が高い文字種とその枠を求める。各文字に対して、複数の候補枠を統合して最終的な文字枠を決定し、その文字枠に対応する文字種を最終結果とする。

しかし、この手法では、長さだけ可変のスライド窓により候補枠を作成することで、文字が候補枠のサイズと合致することを期待しているが、それは現実的ではない。

提案手法①は、CNN (Convolutional Neural Networks)の画像特徴を抽出する特性を利用し、文字種の選定を行うと同時に、当該文字の枠も学習させることを実現する。これを行うために、くずし字画像に各文字の区切り情報が必要である。そこで、本論文では、(1)文字の座標情報を持つCODH(Center for Open Data in the Humanities)データセット、(2)学習用問題データセットであるPRMU(Pattern Recognition and Media Understanding)データセットと合わせ、くずし字区切りデータセットを作成した。なお、データセット作成の結果は国際会議JADHに採択された[1]。この新しいデータセットを用いて、くずし字画像の各文字の区切り情報と文字種情報を同時に学習することで、くずし

字向けニューラルネットワークの重み付けファイルを得ることができる。さらに、この重み付けファイルを用いて、くずし字文字画像を予測して複数の候補枠を出力することにより、その複数の候補枠の統合手法を提案する。

提案手法②は区切り枠の予測のみとなる。従来の手書き文字認識によく使われる多文字画像を単一文字ずつに分割してから認識するという手順をくずし字の認識に適応するという手法である。通常の手書き文字を分割する際に、ヒストグラム法のような文字と文字の間に空白な部分が存在していることを利用し、分割を行うことが多いが、このような方法はくずし字のように繋がって、重なっている文字には対応できない。

そこで、画像の色や材料の変化のようなより豊かな特徴を利用し、繋がっている文字でも分割できる手法を提案する。選択性検索法は画像の中の色やテクスチャ特徴に基づいて、類似度が高い近接領域を結合し、目標対象の領域を認識する方法である。これを用いて、くずし字向け枠決定手法を提案する。この枠決定手法では、選択性検索法で予測した枠とヒストグラム法で予測した枠により、候補枠を作成し、見本データセット(くずし字のすべての字形画像を一枚ずつ保存したもの)とマッチングすることにより、最も文字らしい画像を囲った枠を選定する。

最後に、提案手法の実験的評価を行う。評価指標として誤認識率と枠の一致率(誤差を許す)を用いる。七万個の3文字データを用いて学習を行い、7000個のテスト用データを用いて評価する。実験結果は、提案手法①の誤認識率は0.7%であり、[2]の誤認識率12.4%と比べ、さらに良い結果を出した。また、枠の評価について、提案手法②の枠予測結果は提案手法①より良い結果を出したことを明らかにした。例えば、評価パラメータ ϵ が0.07の場合に、提案手法②の枠予測結果は86.5%であり、提案手法①の枠予測結果は72.2%である。

参考文献:

[1]Y. Tang, K. Hatano, E. Ishita, T. Nakatoh, and T. Kawahira. Construction of Japanese historical hand-written characters segmentation data from the CODH data sets. JADH2018, 2018.

[2] Nguyen, H. T., Ly, N. T., Nguyen, K. C., Nguyen, C. T., & Nakagawa, M. (2017). "Attempts to recognize anomalously deformed Kana in Japanese historical documents." In *Proceedings of the 4th International Workshop on Historical Document Imaging and Processing (HIP2017)*, 31- 36.