

平成29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
468	修士 (ライブラリーサイエンス)	チン セイオク 陳 聖億	中国の統計データの信憑性を巡る議論について - 失業率とGDPを中心に -	公開
469	修士 (ライブラリーサイエンス)	ハギワラ ヤスコ 萩原 泰子	研究者による論文選択の実態に基づく文献検索講習会の提案	公開
470	修士 (ライブラリーサイエンス)	フクシマ カナ 福嶋 香奈	大学における授業外ライティングサポートの質保証	公開
471	修士 (ライブラリーサイエンス)	ホンダ ユウヤ 本田 優也	電子漫画読者コメント分析	公開
472	修士 (ライブラリーサイエンス)	ミズタニ エミコ 水谷 江美子	蘚苔類標本データベースのためのデータセットの分析	公開

---

# 中国の統計データの信憑性を巡る議論について—失業率と GDP データを中心として—

## A Study on Credibility of Chinese Statistical Data - Focusing on Unemployment Rate and GDP Data

---

2FS16205K 陳 聖億 CHEN Shengyi

---

1978 年以前、中国政府は三年大飢饉時期と大躍進運動を含め、政治的原因で統計データの虚偽報告や水増しが常態化していた。1978 年以降、改革開放政策の実施とともに、中国政府は GDP 統計や失業率などの統計データの集計と公表を始めてきた。しかし、三年大飢饉時期と大躍進運動の時代から残された慣行および公表したデータと現実の乖離により、世界中において、中国の統計データへの不信感が高まってきた。

2007 年、中国の現首相、李克強は地方政府の統計データへの批判を契機に、中国の統計データへの不信感が一層高くなってきた。

こうした背景のもと、本研究では、中国政府により公表されたデータへの不信感が高まっている環境の下で、中国国家统计局のデータに光を当て、これまでの中国の統計データに関する先行研究を分析することで、中国の統計データはなぜ世界中から疑われているか、データの信憑性は高いかどうか、現実のデータとどのように乖離しているのかを考察したい。そのため、本研究では、中国国家统计局のデータ GDP と失業率という 2 つの最も重要な統計データを取り上げ、この 2 つのデータの集計および計算方法について詳細な分析を行った。

研究の手順として、まず、文献調査と公的機関のデータに基づき、GDP と失業率に関する詳細な分析を行った。また、中国の GDP と失業率のデータに関する文献考察の現状を調べることによって、中国統計データの信憑性を研究する重要性と本研究の位置づけを示した。

中国の GDP 統計データについては、本研究では、中国の GDP 統計データの信憑性に関する先行研究と実際のデータを詳しく分析することによって、中国国家统计局により公表された全体的なデータの信憑性がある程度あるものの、中国地方政府から開示されたデータには信憑性が低く、誤差も大きいことを示した。誤差が生じた主要な原因は、国民経済計算方式の急激な転換、地方政府の集計・統計能力の限界、業績を求めるためデータへの改ざんと過大評価が挙げられる。

そのため、高橋洋一 (2016) のような中国の統計データは現実との乖離が激しく、少なくとも 3 倍くらいの乖離があると主張している人が現れた。それに対して、加之・三竝 (2016) を含む研究者は先行研究に基づき独自推計を行った結果を用いて、中国の GDP 統計の信頼性を考察してみた。その結果、統計データには誤差が存在するが、中国の GDP 統計データは国際基準に合致した統計手法や概念の整備が進んでいるため、この程度の誤差が許される範囲に

留まることも示した。

また、中国の失業率に関する統計データについては、中国の失業者と失業率に関する定義を考察するために、労働経済学により定められた失業率の定義と日米の失業者と失業率の定義を用いて、中国の失業率はどのように集計および計算されるのか、計算の方法と集計の基準は労働経済学や国際基準から離れているかどうかを考察した。

その結果、中国の国際労働期間が定めた就業と失業の定義に関する国際的基準に合致しないことが多くの研究者に指摘されている。また、中国の失業率の集計と計算は、国際基準から大きく離れていることも明らかになった。そのなか、失業率の統計は都市戸籍を持つ者のみが対象で、出稼ぎ労働者、農民工、失業したが失業登記の対象にならない労働者などが対象外となることと、失業保険の金額の低さによる登記人数と現実の失業者の乖離に関する問題が最も深刻である。

さらに、中国社会科学院を含め、多くの研究機関や公的部門により公表された失業率は、中国国家统计局から公表された失業率との乖離が大きく、失業率のデータに関する一貫性が極めて低いことが明らかになった。失業率に関するデータの一貫性の低さと計算方法の問題で、中国の失業率に関するデータの信憑性が低く、高橋 (2016) に指摘されたように 20% も超えていることも懸念されている。

そのため、本研究で取り組んでいる GDP と失業率のデータの中、中国の GDP 統計データと失業率の信憑性が依然として低いままであることが明らかになった。

また、これまでの調査から得られた知見として、中国の統計データは世界中から疑われている原因は主に「データの精度」、「データの一貫性」、「統計方法の正確さ」、「データの透明感」および「データ捏造の動機」という 5 つに挙げられる。GDP と失業率の両方ともに不透明感が強いいため、世界中から疑われている。GDP の場合に、地方統計と全国統計との大幅な乖離があり、統計方法の正確さが欠けているため信憑性に大きな影響を及ぼしている。それに対して、一方、失業率の方は一貫性、精度、統計方法の正確さ、透明感のいずれも極めて低く、データ捏造の動機も十分であるため、失業率の信憑性が極めて低いことが言わざるを得ない。

最後に、本研究では、これまでの調査によって、GDP と失業率の信憑性が依然として信頼できない状態に留まることを示している。しかし、量的分析を用いてこの結果を検証していなかった。この点については今後の課題とした。

# 研究者による論文選択の実態に基づく文献検索講習会の提案

Designing Library Workshop Based on Researchers' Document Selection Behavior

2FS15060R 萩原 泰子 HAGIWARA Yasuko

現在、大学図書館には情報リテラシー教育に主体となって取り組むことが求められている。元来、大学図書館では情報リテラシー教育の一つとして文献検索講習会が行われてきた。しかし、近年では、サーチエンジンも扱うことや情報を利用するプロセスを視野に入れることも必要とされている。現在の文献検索講習会の傾向を把握するため、8学部以上を持つ19国立大学図書館のホームページで講習会の内容を調査したところ、272件中244件がデータベースの使い方を中心に説明する講習会であった。一方、研究者の情報行動に関する既往研究からは、検索の目的に応じ、データベースの他にサーチエンジンを併用し、様々な手段で文献を入手していることが明らかになっている。これらのことから、これから研究を始める利用者は、研究プロセスの中で論文が必要になる状況を認識し、その状況下でどのように論文を選択するのか、研究者が実際に行っている方法を知り、自身の研究に即した効率的な検索方法を習得することが重要であるといえる。

本研究では、研究者の予備軍として大学院修士課程の学生を対象に、研究者の論文選択の方法を参考にした文献検索講習会を提案することを目的とした。このために2つの調査を行った。まず、研究者の論文選択のプロセスを明らかにするために、研究者による文献検索を観察する調査を行った。次に、この観察調査で確認した行動の一般性を確認するため、さらに論文入手における研究者と修士課程の学生の行動を比較するための質問紙調査を行った。

観察調査は、博士課程の学生を含む10名の研究者を対象に、事前インタビュー、実験、事後インタビューの3段階で行った。実験では参加者に自身の研究課題に基づいた文献検索を行ってもらい、検索結果から全文を読もうとする論文を選んでもらった。この際、検索結果に表示されるどの要素(タイトル、抄録、著者名など)を確認したか、また、抄録中のどの文を確認したか発言してもらった。さらに読む、または読まないと判断した理由についても尋ねた。事前インタビューでは研究課題と実験で検索する内容を中心に尋ね、通常使っているデータベース等も確認した。

調査結果を分析したところ、論文選択のために確認する要素の順番については大きく3つに類型化できた。1つ目は要素をタイトル→抄録の順に確認するパターンで、6名が行っていた。2つ目はタイトル→本文と確認するパターンで、2名が行っていた。本文中では図表や参考文献が論文選択の判断に利用されていた。3つ目は、1名のみが行っていたパターンで、Googleの画像検索で実験の装置名を検索し、検索結果の画像をクリックし、さらに本文を確

認していた。その他、インタビューでは、論文選択の理由として研究段階に言及する参加者が多く、これらが利用するデータベース等や確認要素に影響する可能性があるといえる。なお、通常利用しているデータベース等は、7名がGoogle Scholarなど学術データベース以外を挙げた。

質問紙調査は観察調査の結果を反映し、研究段階別の利用データベースと確認要素、論文選択の際の要素の確認順を尋ねる質問を作成した。また、論文の入手先、全文を読む論文が入手できている程度、学術データベースの機能の利用についても尋ねた。対象は電気電子情報工学分野の教員と大学院生とし、2大学の専攻に協力を得て実施した。

88名(研究者48名、修士課程40名)の回答結果から、修士学生は全文を読もうとする論文を入手できていると感じる程度が研究者に比べ低かった。一因として、修士学生が論文の入手方法に精通していないことが他の質問から示唆された。例えば論文の入手先については、修士学生は直接教えてもらうという回答が82.5%と他の入手先に比べて高く、研究者が利用すると回答した学会の情報、特定の雑誌、著者のホームページの利用率は低かった。また、回答率が最も高い利用データベース等は、研究者は学会デジタルライブラリであったが、修士学生はGoogle Scholarであった。検索の目的別の確認要素では、研究者は、詳しくない分野の調査では被引用数や雑誌名、研究手法調査ではキーワードなど、目的別の確認要素に特徴があった。一方、修士学生は研究手法調査以外の文献検索を行うことが少ないことが分かった。これらの結果から、研究者の論文選択の方法を講習会で説明することで修士学生が全文を読もうとする論文を入手するための支援になると言えた。

提案する文献検索講習会では、2つの調査結果を反映し、研究者の事例を挙げて論文選択の方法を説明することとした。対象は質問紙調査を受け、電気電子工学分野の修士課程とした。観察調査で類型化した3パターンを元に、画像検索、本文の図表を確認、本文のIntroductionを確認、抄録を確認の4事例を紹介し、各事例の中で、検索の目的に合せて確認する要素や本文中の箇所を説明した。最後に講習会で扱わなかった方法があれば共有するため、所属研究室で行われている論文選択の方法をグループで紹介し合う時間も設けた。

本研究では、文献検索講習会において状況に応じた多様な手順を示すため、研究者を対象とした調査に基づいて論文選択の方法を実証した。情報リテラシー教育を向上させる一案として実践を求めたい。また、事例を積み重ねていけば、今後の図書館サービスの発展にも寄与できるだろう。

---

# 大学における授業外ライティングサポートの質保証

Quality Assurance of Writing Supports Outside Courses in Universities

---

2FS16201G 福嶋 香奈 FUKUSHIMA Kana

---

2008年ごろから、日本の大学の教育方針は、授業による「知識の伝達」から「知識の創出・自主的学習」へと大きく転換してきた。そのため近年、学生の自主的な学習を支援する手段として、学習支援の重要性が高まっている。様々な学習支援の取り組みの中に、授業時間外でレポートや論文などのライティングに関する相談に対応する、授業外ライティングサポートがある。学習支援の中でライティング教育の一端を担う授業外ライティングサポートを実施する大学は増加しており、その質の保証にも目を向ける必要がある。

日本の大学における授業外ライティングサポートは、その運営組織や指導内容などがサポートによって異なっており、多様性が生じている。その理由として、国内では授業外ライティングサポートにおけるガイドラインが確立されておらず、既に実施されているサポートの中でも、そのあり方を模索している段階にあるということが指摘されている。

この状況を踏まえ、本研究では、日本におけるライティングサポートの全体像を把握し、運営体制面の課題を明らかにすることを目的とする。その上で、どのような体制でライティングサポートを運営するのが良いのかを日本の大学の現状に合わせて検討し、授業外ライティングサポートの質保証に何が必要であるかを示す。これは、今後、新たにライティングサポートを開設する大学にとって現実的な参考情報となり得る。また、既存のサポートの改善にも役立てることができる。

授業外ライティングサポートの全体像把握のため、二つの調査を行った。一つはウェブサイト調査で、国立大学全86大学のウェブサイトから授業外ライティングサポートの有無とその運営体制を調査した。その結果、全86大学中28大学において36件のライティングサポートが確認された。アンケート調査では、36件のサポートの運営組織に対し、指導者の研修に関するアンケート調査を実施した。

ウェブサイト調査の結果、58.3%の授業外ライティングサポートが大学図書館によって運営されており、最も多かった。また、約60%のサポートで指導者に大学院生を採用していることが分かった。また、アンケート調査の結果、67.9%のサポートで指導者への研修を実施していた。研修内容は、指導者同士のミーティングや学内外のセミナーの受講等、様々であったが、ライティングに関する科目の履修やライティングのプロの教職員とのセッション練習を組み合わせて指導者が十分な指導スキルを身につけるまでの研修プログラムを確立させているサポートは、研修を

実施している19件中2件のみであった。

これらの結果から、ライティングの専門家ではない大学院生が主な指導者であるにも関わらず、多くのライティングサポートでは指導者の研修体制が確立できていないという課題が明らかになった。その理由として、ライティング教育が専門の組織ではない大学図書館が主な運営組織であるため、指導者の研修を図書館員が行っており、専門的なライティング研修が困難であることが考えられる。

この課題を解決するために、授業外ライティングサポートにおける指導者の研修のガイドラインを確立する必要がある。本研究では、研修体制のガイドライン策定に向けた取り組みとして、ライティング指導者に最低限必要なスキルを設定することを試みた。調査結果に基づいて、サポートの運営組織は大学図書館で、大学院生が主な指導者である場合を想定して、スキルを設定する。

ライティングサポートの指導者に必要なスキルは、知識を中心としたライティングスキルと、指導中の態度を含むチュータリングのスキルに分けられる。本研究では、ライティングの専門家ではない大学図書館の担当者や大学院生でも研修が可能なものとして、基本的なライティングスキルに焦点をあて、指導者に必要なスキルを提示した。スキルを設定するため、日本とアメリカの大学の授業外ライティングサポートにおいて実際に利用者向けに必要なスキルとして提示されているものを、文献やウェブサイトから合わせて16事例収集した。

収集したスキルを、ライティングプロセスに応じて、ライティングの基本的な考え方、書く前の準備段階、草稿・執筆の段階、校正・体裁の段階、特殊な場面の5つのカテゴリに分類し、最もライティング指導者に必要だと考えられるスキルを、指導者に求められるコアスキルとして設定した。各サポートが、このコアスキルセットに沿って研修体制を整えていけば、指導者は指導に必要なライティングスキルを身につけることができ、ライティングサポートの質の保証を実現することができる。

サポートの内容に応じて、指導者に求められるスキルのレベルも異なる。本研究では基本的なライティングスキルのみを設定したが、例えば授業の課題レポートだけでなく、学術論文や英語論文も指導する場合など、より専門的なライティングサポートを行う場合には、サポート内容に応じたスキルをコアスキルに加えることで、より発展的なスキルセットを作ることも可能である。本研究ではライティングスキルのみ焦点を当てたが、今後は、チュータリングスキルについても同様に基準を確立させていく必要がある。

---

# 電子漫画読者コメント分析

Analysis of online manga user comment

---

2FS16204E 本田 優也 HONDA Yuya

---

## 1. はじめに

スマートフォンの普及に伴い、電子漫画市場が拡大している。紙媒体の漫画と異なり、電子漫画は読者がコメントを投稿できる。読者コメントから作者や編集者(以下、作者側と表記)は今後の展開や次作品の参考になる情報を得ることができると期待される。作者側が、膨大な読者コメントを確認するのは大変な作業になるため、統計的な分析で上記のような有用な情報を提供できることが望ましい。

本研究では、毎回の「話」について読者が書いたコメントを分析する「話」に着目した分析と、コメントに出現する「登場人物」に着目した分析を行い、読者が作品のどこに関心を持っているかの推定に有用な情報の抽出を行った。

本研究の結果は、作者側の意図と読者の関心が一致するかを検討するヒントになると考えられる。一致すれば意図が伝わったと分り、そうでなければ一致しない理由を考えることで、その後の漫画作りの参考になると考えられる。具体的な対象データとして、多くの利用者がいる電子漫画アプリ comico の最人気作品 ReLIFE の読者コメントを分析した。

## 2. 「話」に着目した分析

連載漫画は毎週公開される「話」の系列といえる。コメント数の多い「話」に読者間の盛り上がりがあるといえる。どの「話」でコメント数が多いのか、また、その「話」で読者が何についてコメントしているのかを分析した。

第118話はコメント総数14986件で最多の「話」である。読者は第118話で最も盛り上がっていると考えられる。一方、第30話から第37話のコメント数は1000件ほどであり、この期間で読者はあまり盛り上がっていないと推定される。

読者が漫画作品の何に関心を持ったか分かれば、漫画作りに役立つ。ある「話」でコメントに頻出する単語は重要であると考え、各単語の出現頻度を見た。コメント総数上位の5話(118、25、144、27、113話)について出現頻度上位の単語を抽出した。その結果、各話から抽出された単語には「いい」や「くん」などの単語が上位に出てきた。出現頻度では読者に使われやすい単語が上位に現れ、ある「話」での読者の関心は分からない。

そこで、第*i*話のコメントとそれ以外のコメントを判別する機械学習で第*i*話のコメントについての特徴語を見るため、SVMの属性選択を適用した。具体的には、第*i*話へのコメント集合を正例、その他の話へのコメント集合を負例として、線形カーネルのSVMを適用し、得られたモデルでの単語の重みをその単語の重要度として分析した。この重要度が高い単語は、読者のその話への関心を示す可能性がある。

118話の特徴語には「ストレート」「被験者」などの内容に関わる単語や、「衝撃」といった内容への読者の印象を表す単語を抽出できた。コメント数が上位5話それぞれで内容に関わる特徴語を抽出できた。また、内容に関係ない運営主催のイベント、アニメ化などに関する単語も確認された。抽出された単語集合から、内容に関係ない単語を除くと、読者のその話についての関心を示すだろう。作者側でイベント、アニメ化などに関する単語を除外することは可能であると考えられる。

## 3. 「登場人物」に着目した分析

登場人物は漫画作品において重要な要素である。読者が登場人物のどこに関心があるのか分かると、作者側が意図する人物像と比較することができ、その後の人物作りに役立つ。なお、人物名を収集する際、前もって人手で略称や愛称を統合した。

読者の登場人物への関心を、コメント内で各登場人物と共起する単語の出現頻度から見た。結果はコメント総数で見た時と同様、読者の関心の分かる単語は抽出できなかった。

そこで、各登場人物の名前が現れるコメントの集合を正例としてSVMと属性選択を適用し、抽出された特徴語で分析した。抽出された単語は、読者の「登場人物」への関心を示す可能性がある。登場人物「狩生」への特徴語から「ツンデレ」という単語が抽出できた。読者は「狩生」を「ツンデレ」な人物と捉えていると推定する。登場人物「日代」には「被験者」という単語が抽出できた。「被験者」は、「話」に基づく分析の第118話の特徴語でも抽出された。登場人物「小野屋」から「怪しい」という単語が抽出できた。読者は「小野屋」の人物像に怪しさを感じていると推定でき、実際色々な人物像を推測するコメントが多数確認できた。

# 蘚苔類標本データベースのためのデータセットの分析

Analysis of Datasets for Bryophyte Specimen Database

2FS15061N 水谷 江美子 MIZUTANI Emiko

蘚苔類分野の研究者にとって、コケ植物を乾燥した標本は研究を遂行するために必要不可欠なものであり、研究者自らも蘚苔類の植物を採集し、乾燥標本を作成する。標本には、学名、採集地、管理番号などのメタ情報が所蔵機関ごとに付与されている。蘚苔類を対象にした研究は、しばしば標本を用いておこなわれるが、標本データへのアクセスを容易にするには、公開されていない標本データをデータベースとして公開し、公開されている既存のデータベースの間でも相互にデータを交換可能にする必要がある。更に、このためには、蘚苔類標本データを記述するための標準的なデータ項目を設定する必要がある。この動きの一つとして、地球規模生物多様性情報 (GBIF) の活動がある。

本研究では、蘚苔類研究者がデータ登録をおこなうことを想定し、蘚苔類に必要なデータセットを提案した。まず、実際の蘚苔類データベースで使用されているデータ項目を明らかにするため、既存の蘚苔類データベースで用いられているデータ項目と乾燥標本ラベルに記載されている情報を調査した。これらの調査結果を GBIF のメタデータである Darwin Core と対照することによって、蘚苔類標本データのためのデータセットを提案した。

調査対象とした既存のデータベースは、日本のものは 6 件であり、その内訳は検索形式 3 件、リスト形式 2 件、冊子体 1 件である。選定には、まず、日本分類学会連合が日本蘚苔類学会からの情報提供を受けて作成した蘚苔類標本情報国内重要コレクションリストから、標本所蔵が 1 万件以上であり、かつウェブサイトで公開されている 4 データベースを選んだ。次に、GBIF のデータ登録統計の Japan リストに含まれており、データセットに Bryophyte(or Moss)の用語を含む公開されたデータベースから、上記の 4 データベースと重複しない 1 データベースを選定した。さらに、冊子体形式で公開されている 1 件を選択した。海外のデータベースは、GBIF の国別統計リストで Bryophyte(or Moss)の用語を含むデータセットを持つ 36 機関と、Google で「bryophyte(or moss) specimen database」を検索した結果得られた中で上記と重複しない 39 データベースから、英語であり、かつ、公開されている 14 データベースを選択した。以上の結果得られた日本と海外を合わせた 20 データベースについて、それぞれ最低 5 回の検索をおこない、検索結果として表示されるデータから、データ項目を調べた。

次に、公益財団法人服部植物研究所が所蔵する蘚苔類乾燥標本の保管用袋に貼付されたラベルに記載されている情報を調査した。当研究所は蘚苔類を専門とする世界唯一

の研究機関であり、約 48 万点の乾燥標本を所蔵しているため、ラベルの記載内容から多種多様な情報を得ることができると判断し、調査対象とした。対象とする標本ラベルは、標本ラベル全体を概観し、記載内容に差異がないことから、学名決定の際に使用されたホロタイプ標本 151 件に絞った。これらに対し、ラベルの記載情報を抽出しながら、項目名を設定していったところ、22 のデータ項目が抽出された。

2つの調査から得られたデータ項目を Darwin Core の項目に照らし合わせ、蘚苔類標本情報のデータセットを選定した。データセットの項目を選定するにあたって、既存データベースや標本ラベルから抽出されたデータ項目のうち、Darwin Core のデータ項目に 1 件でも対応した項目は蘚苔類研究者が必要だと判断した項目とみなし、選定した。対応する項目がなかったものでも、必要と判断される項目は選定した。以下、選定した項目を Darwin Core のレベル毎に示す。「record-level terms」レベルは調査結果では対応する項目がなかったが、データ登録の際の必須項目であるため、21 項目すべてを選定した。「occurrence」では、22 項目中 12 項目が調査結果の項目と対応し、対応がなかった 10 項目のうち植物に関するデータが存在する 6 項目を残し、計 18 項目を選定した。「event」は、15 項目のうち 8 項目が対応し、対応がなかった 7 項目のうち標本採集の期間や場所、プロジェクトに関する 4 項目は必要と判断し、12 項目を選定した。「location」は 44 項目のうち 19 項目が対応し、対応しなかった項目のうち国コードと水中に生存する蘚苔類のために水深海拔の項目は残し、24 項目を選定した。「geological context」の 18 項目はすべて対応がなく、内容が古代動物関連項目のため、一つも選定しなかった。「identification」は 8 項目のうち 6 項目は対応しており、対応のなかった 2 項目のうち蘚苔類の同定過程で必要のない 1 項目を省き、7 項目を選定した。「taxon」は 33 項目のうち 16 項目が対応しなかったが、この中から学名に関する 6 項目を残し、23 項目を選定した。以上、Darwin Core からは 105 項目を選定した。さらに、標本ラベル調査から、ラベル表記で特徴あるものを記載する項目、他機関からの移管・委譲・寄贈を明示する項目の 2 項目を加え、最終的に提案するデータセットは 107 項目である。項目数が多いとデータ入力に負担がかかるが、提案するデータセットの項目数は Darwin Core の項目数と比較すると 65%であり、少ない項目数であった。

今後は、提案したデータセットを蘚苔類研究者が使用することにより、検証と分析を重ね、改善する。