

[2023/2024]九州大学大学院統合新領域学府ライブラリーサイエンス専攻年報 : 2023/2024

<https://doi.org/10.15017/7170855>

出版情報 : 九州大学大学院統合新領域学府ライブラリーサイエンス専攻年報. 2023/2024, pp.1-, 2024-03. Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

バージョン :

権利関係 : Creative Commons Attribution 4.0 International





Annual Report

Department of Library Science **2023/2024**
Graduate School of Integrated Frontier Sciences
Kyushu University

九州大学 大学院統合新領域学府
ライブラリーサイエンス専攻 年報 2023/2024



Annual Report

Department of Library Science **2023/2024**
Graduate School of Integrated Frontier Sciences
Kyushu University

Contents

専攻長挨拶	2
「大学における研究データ管理の意義と支援人材育成」国際シンポジウム開催報告	3
「東京大学史料編纂所「正保琉球国絵図デジタルアーカイブ」の公開」	
デジタル公開の現場から	5
国際会議(ICADL2023)参加記	7
インターンシップ体験記	9
博士論文の紹介	11
修士論文の紹介	13
新任教員紹介	17
研究データ管理(RDM)支援人材育成プログラムの概要	20
専攻紹介	21
活動日誌	22



専攻長挨拶

コロナ後の新しい取り組み

—「研究データ管理 (RDM) 支援人材育成プログラム」の開始—

ライブラリーサイエンス専攻長

藤岡 健太郎

九州大学大学院統合新領域学府ライブラリーサイエンス専攻は、ユーザーの視点に立って、情報の管理・提供を担う人材の育成と、関連分野の研究を行うことを目標として、これまで教育・研究に取り組んでまいりました。

3年あまり続いたコロナ禍はようやく終息し、大学はほぼ以前の姿を取り戻しました。とはいえ、コロナ禍以前と様変わりしたところも少なくありません。特にオンライン会議システムがコロナ禍の中で普及した結果、会議や打合せ、研究発表等は、コロナ後も多くオンラインで行われています。博士課程に社会人学生の多い当専攻の場合、遠方であってもオンラインで研究発表や研究指導ができるところはたいへん重宝しています。一方で、対面でのコミュニケーションの機会が減ったことで、今後長期的には弊害も出てくるのではないかという危惧は、多くの方が感じているところではないでしょうか。

コロナ禍と直接関係するわけではありませんが、その渦中では、以前からのオープンデータ・オープンサイエンスの流れがより強まったと感じられます。この流れの中で、研究データの適切な管理もより強く求められるようになりました。ライブラリーサイエンス専攻では研究者がより適切に研究データを管理できるよう支援する人材を育成するため、2023年10月より、「研究データ管理 (RDM) 支援人材育成プログラム」を開始しました。これは「履修証明プログラム」と呼ばれるもので、主として社会人向けの専門的な授業を用意し、所定の単位を取得すれば専門的な知識・技能を習得したことを証明する履修証明書を発行するというものです。初年度は、各大学の図書館員と本学の大学院生が受講しています。

本専攻ではこうした新しい取り組みをこれからも進めてまいります。関係者のみなさまにおかれましては、今後ともライブラリーサイエンス専攻をよろしくお願いいたします。

symposium

国際シンポジウム開催報告



「大学における研究データ管理の意義と支援人材育成」

2023年6月19日に、米国イリノイ大学アーバナ・シャンペーン校のPeter Darch先生とJ. Stephen Downie先生をお招きし、国際シンポジウム「大学における研究データ管理の意義と支援人材育成」(同時通訳付)を開催しました。ライブラリーサイエンス専攻からも、冨浦教授と石田教授が登壇しました。対面とオンラインのハイブリッドで開催され、全国から166名の参加登録があり、各講演やパネルディスカッションにおいて多くの質問が寄せられ、関心の高さがうかがえました。

日 時：2023年6月19日(月) 13:00～17:05

会 場：九州大学伊都キャンパス理系図書館3階大会議室(オンライン配信とのハイブリッド開催)

主 催：九州大学データ駆動イノベーション推進本部(研究データ管理支援部門)

共 催：九州大学附属図書館・大学院統合新領域学府ライブラリーサイエンス専攻

後 援：九州地区大学図書館協議会、日本学術振興会科学研究費「デジタルヒューマニティーズを促進するオープンデータ環境およびシステム基盤の構築」(代表・石田栄美)

プログラム

13:00-13:05 開会挨拶

九州大学データ駆動イノベーション推進本部・副本部長 / 附属図書館長 谷口倫一郎

13:05-13:10 趣旨説明

九州大学データ駆動イノベーション推進本部研究データ管理支援部門・教授 石田栄美

13:10-14:10 講演1「図書館における研究データ基盤の構築：文化と価値観の影響」

イリノイ大学アーバナ・シャンペーン校情報学部・助教 Peter Darch

14:10-15:00 講演2「人文社会科学におけるビックデータの活用：HathiTrust Research Centerの事例」

イリノイ大学アーバナ・シャンペーン校情報学部・教授 J. Stephen Downie

15:10-15:40 講演3「研究データ管理支援部門の活動」

九州大学データ駆動イノベーション推進本部研究データ管理支援部門・部門長 冨浦洋一

15:40-16:10 講演4「ライブラリーサイエンス専攻の研究データ管理支援人材履修証明プログラム」

九州大学データ駆動イノベーション推進本部研究データ管理支援部門・教授 石田栄美

16:15-17:00 パネルディスカッション

17:00-17:05 閉会挨拶

九州大学統合新領域学府ライブラリーサイエンス専攻・専攻長 藤岡健太郎



近年、研究者は研究において生成・収集したデータを管理・公開することが求められています。大学においては、研究者が適切に研究データを管理・公開するための環境を整備することが期待されており、現在、日本の様々な大学で、研究データ管理・公開のための環境をどのように整備していくかが模索されています。九州大学では、2022年度からデータ駆動イノベーション推進本部に研究データ管理支援部門を設置し、研究データを管理・保存するためのデータ管理用ストレージの整備や、研究の各ステップにおいて必要となる研究データ管理に関する支援を開始したところです。研究データ管理は、研究者が従来行っていた事柄ですが、研究助成機関がデータ管理計画(DMP)の作成を求めるようになったり、研究において収集・作成したデータをできるだけ公開することが求められるようになりました。これらはこれまで以上に負担が増える部分もあり、研究者の適切な管理・公開への意識は高いわけではありません。適切な研究データ管理を促進するためには、研究者や研究者を支援する人材が、研究データ管理の意義を理解したうえで、どのような管理や支援をしていくべきかを検討する必要があります。

本シンポジウムでは、上記のような問題意識のもと、米国イリノイ大学アーバナ・シャンペーン校情報学部(iSchool)の研究者を招き、研究データ管理の意義や必要性について考える機会としました。Peter Darch先生からは、図書館における研究データ基盤の構築について、文化や価値観がどのように影響しているかについて講演してもらいました。研究事例として、大規模なデータセットの公開の場合、大学図書館の特徴や文化によってデータ公開の方法が異なることなどの紹介がありました。J. Stephen Downie先生からは、HathiTrust Research Centerにおける大規模なデータセットを利用した研究事例の紹介がありました。HathiTrustデジタルライブラリーでは、様々なジャンルの書籍を電子化しており、提供しています。一部はテキスト化したデータも計量的な分析ができる形で提供されています。これらのデータは、デジタルヒューマニティーズをはじめとする様々な研究に貢献しています。HathiTrust Research Centerでは、これらのデータを用いた計算的分析が可能な環境を提供しています。講演

では、HathiTrust Research Centerの概要、およびに研究プロジェクト、研究事例の紹介がありました。講演3では、富浦先生から、九州大学の研究データ管理支援部門の活動の紹介がありました。研究データ管理・公開ポリシーおよび同解説の策定、研究データ管理用ストレージシステムの導入、研究データ管理リテラシー教材、データ管理計画の作成支援などについて、それぞれどのような活動を行っているかが紹介されました。特に、2023年3月に公開された「九州大学研究データ管理・公開ポリシー」に関しては、ポイント等も含めて紹介がありました。講演4では、2023年度秋学期からライブラリーサイエンス専攻に設置された研究データ管理(RDM)支援人材育成プログラムの概要について紹介がありました。このプログラムは、1年間の履修証明プログラムで(5科目、5単位)、本専攻に入学できなくても履修することが可能であり、履修が完了すると履修証明書を得ることができます。大学において研究データ管理を支援する人材の必要性から、特に大学などで研究データ管理支援に携わる人のためのプログラムとなっていること、また、1科目は九州大学に来て履修する必要がありますが、それ以外はオンラインで履修が可能なことなどが紹介されました。さらに、パネルディスカッションでは、各講演者に対する質問に回答した後、研究者として研究データ管理支援サービスに期待することやサブジェクトライブラリアンがいない日本の図書館ではどのような支援が可能か等についての議論を行いました。

研究データ管理のための環境をどのように整備していくか、また、実際にどのような支援が必要なのかということは実務的な面からも検討していく必要がありますが、情報の管理と提供のありかたを探索するライブラリーサイエンス専攻にとっても、非常に重要な研究テーマです。今後も、このような機会を持ち、検討を重ねていく予定です。

当日の講演やパネルディスカッションの動画(オリジナル言語版、日本語版、英語版)と講演資料を、以下で公開しています。誰でも見るができますので、ぜひご確認ください。

<https://hdl.handle.net/2324/6790816>



東京大学史料編纂所 「正保琉球国絵図デジタルアーカイブ」の公開

黒嶋 敏 (東京大学史料編纂所 准教授)

大学・博物館をはじめとする文化財の所蔵機関にとって、所蔵品の公開は重要な責務ではあるものの、頭を悩ませてしまうのが絵図ではないでしょうか。とくに大型絵図になると取り扱いが難しく、場所や人員を用意しなければ展開すらできないために、やむなく保存庫に眠らせているとの話もよく聞かれます。

筆者の勤務する東京大学史料編纂所でも、その事情は同じです。閲覧のニーズにお応えすることができず、研究資源として利用しうる複製もないまま、公開方法が長年の課題となって（つまりは公開が先延ばしされて）きた所蔵品に、国宝「島津家文書」の国絵図があります。そのなかでも「正保琉球国絵図」は、現在の沖縄本島を中心とする島々を3舗に分けて仕立てた絵図で、1649年に薩摩藩が江戸幕府に提出した絵図の精巧な写し図です。琉球王国の版図を描いた大型絵図としては最も古いものとされ、サンゴ礁などの地形描画や豊富な文字情報もあることから、歴史学や地理学など各方面から注目されてきました。

しかしこの絵図は、長辺が7mを超える大きさがネックとなり、史料編纂所の狭い建物の中では広げることもできません。簡単には原本を閲覧に供せない以上、公開方法として最適解となるのはデジタル公開です。そのためには、まず高精細の画像データを得る必要がありますが、カメラでは1舗あたり100コマ近い分割撮影になり、接合時の作業量や修正個所が大きくなる懸念が残ります。

そこで検討の末に、筆者が研究代表者となった科学研究費補助金などの外部資金を用いて、文化財専門業者によるデジタルスキャンを実施することになったのです。

デジタルスキャンは絵図の周囲にスキャナー用のレールを設置するため、さらに広い作業面積を必要とし、しかも文化財を広げることに適した環境を確保しなければなりません。幸いにも国立歴史民俗博物館（千葉県佐倉市）の展示室をお借りすることができ、2020年に絵図を搬送してスキャンを行いました（図①）。ちなみに搬送も文化財の専門業者によるものです。これに先立ち、絵図にある折れジワなどの凹凸を補正しておかなければならず、関係機関と協議しながら、史料編纂所の史料保存技術室にて平滑化の作業を行いました（図②）。スキャンは一度に広い面積の画像データを得ることができる利点がありますが、それでも1舗あたり4時間を要するため、全3舗に2日をかけました（図③）。

獲得できた画像データを基に、次は、デジタル公開の方法を探っていきます。大型絵図をweb公開した先行事例としては、国立公文書館デジタルアーカイブ^①など、デジタル画像の閲覧に主眼を置くものがありました。一方で、人文学オープンデータ共同利用センターから公開されていた江戸マップβ版^②は、江戸切絵図の閲覧だけではなく、抽出された地名の検索、絵図上の位置や現代地図との重ね合わせなどの機能を持っており、歴史地理情報を



図① デジタルスキャンに向けて、7人がかりで絵図を広げる
(2020年12月、国立歴史民俗博物館)



図② 絵図の補正作業 絵図の上に吸水性の高い厚手の紙を敷き、加湿した後に、押し板を乗せて絵図の凹凸を減らしていく。

重視した先駆的な取り組みとして大変興味深いデジタルアーカイブであるといえます。これらの事例に学ぶとともに、史料編纂所の前近代日本史情報国際センターから支援を受けながら、画像閲覧に加え各種の機能を盛り込んだデジタルアーカイブを構築していくことになりました。

こうして2021年よりweb公開を開始したものが、「正保琉球国絵図デジタルアーカイブ」⁽³⁾になります(図④)。絵図の画像データ量は解像度300dpiで1鋪あたり約1.6GBにもなりますが、史料編纂所で2015年より採用していた画像の管理・共有の国際規格であるIIIF (International Image Interoperability Framework) は、大きなサイズの画像もスムーズに操作でき、今回のような容量の大きいデータにも適性を持っていました。またIIIFはデジタルアーカイブ同士の相互運用をスムーズにする利点があり、IIIF 対応の画像ビューワMiradorを利用して、類例になる国立公文書館所蔵の絵図と並列表示することもできます(デジタルアーカイブの「比較」欄)。

さらにIIIFのキュレーション機能を使い、絵図に書き込まれた文字を全てテキスト化し、アノテーションを付してデータを構造化しました。これにより、書き込まれた文字情報のテキスト検索が可能となっただけでなく(「検索」欄、図⑤)、島や航路などの属性による分類(「カテゴリ」欄)、石高による村などの並べ替え(「石高」欄)のほか、地名に付したGIS情報から現代地図との重ね合わせ表示(「現代地図」欄)なども可能になりました。また地名からは、辞書類の総合データベースであるジャパンナレッジ⁽⁴⁾にある『日本歴史地名大系』の該当項目にもリンクでき、派生的な利用が可能になっています。

このような各種の機能を搭載することで、「正保琉球国絵図デジタルアーカイブ」は幅広い層に利用していただけるものを目指しました。たとえば、専門家以外にはハードルになりやすい崩し字が、容易にテキスト検索できるようになったことで、一般の方にも使いやすいものになったといえます。研究目的による閲覧だけでなく、学校教育の場など多様な利用方法の広がりを期待できるのではないのでしょうか。

また「正保琉球国絵図デジタルアーカイブ」では、近年のオープンデータ化の流れを受け、画像利用はCC BY (クリエイティブ・コモンズ・ライセンス、所蔵者情報を明示して利用可能) 相当のオープンデータとし、メタデータもダウンロードできるようになっています。前近代の絵図のなかには、現代の基準・感覚からすると不適切な表現や情報が盛り込まれ全面的な公開にそぐわないものもありますが、今回は文字情報のテキスト化やアノテーション付与という基礎作業を通じて図中の情報を二重三重にチェックし、

公開やオープンデータ化に支障がないことを確認しました。

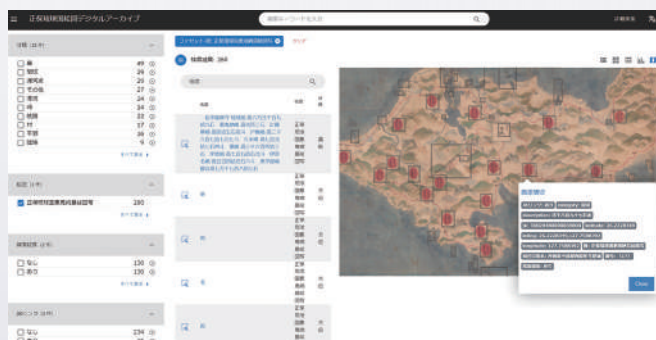
従来の画像のみを閲覧できるweb公開とは異なり、データを構造化し付加的な機能を搭載した公開方法は、史料編纂所にとっても新たな試みとなるものです。ぜひ多くの方に体験していただくとともに、今後の改善につながるご意見を頂戴できますと幸いです。



図③ スキャン中の様子 スキャナーの可動域に即して、分割してスキャンを実施した。



図④ 「正保琉球国絵図デジタルアーカイブ」トップページ



図⑤ アーカイブの検索結果一覧 絵図上の矩形の線がアノテーション設定の範囲を示す。

(1) <https://www.digital.archives.go.jp/DAS/pickup/view/category/categoryArchives/0300000000/default/00>

(2) <http://codh.rois.ac.jp/edo-maps/>

(3) <https://www.hi.u-tokyo.ac.jp/collection/digitalgallery/ryukyu/>

(4) <https://japancknowledge.com/>



Presenting Research Paper at the 25th International Conference on Asia-Pacific Digital Libraries, Taipei, Taiwan

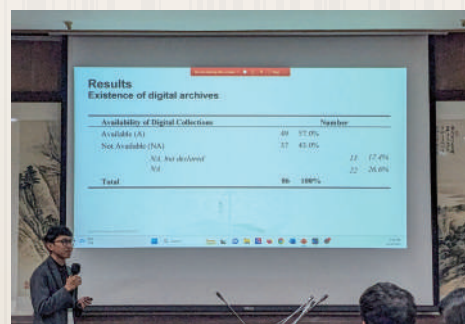
Widiatmoko Adi Putranto (Department of Library Science, D2)

The 25th International Conference on Asia-Pacific Digital Libraries was held between 4-7 December 2023 in Taipei, Taiwan. There were six delegates from Kyushu University attended the event. I came to the conference particularly to present my research entitled “Opening access to digital collections: the state of cultural materials in Indonesian higher education institutions”, which was also accepted as full paper in the conference proceeding. The proceeding is now also accessible via the Springer Portal (<https://link.springer.com/book/10.1007/978-981-99-8085-7>). This report briefly describes the overview of my participation activities in the conference.

The conference consisted of diverse programs. I attended the pre-conference workshop of “Metadata Models for Digital Archiving of Intangible and Experiential Cultural Entities” at the International Conference Hall in National Taiwan Normal University Library. The workshop was hosted by Professor Shigeo Sugimoto from University of Tsukuba, Japan. The workshop mainly discussed the ongoing issues on managing currently existing massive numbers of digital archives and metadata. It also touched on the development of diverse and previously unimaginable materials to be archived. There were several international presenters providing insights from their research on the issues. Several presenters were based in Europe but their research objects were Japanese materials, so it was interesting to look into their perspectives on managing materials from another country. The workshop itself was more about challenging the participants with the issue of archiving ‘intangible’ and ‘experiential’ cultural entities in form of natural phenomena or intangible experience such as disasters or fireworks. Professor Sugimoto provided provocative view in how objects like fireworks can be archived in three different categories: the firework as ephemeral object in the air, the firework as the tangible object in the form of its ball gunpowder, and the firework as the event (Sugimoto 2023). Additionally, the presentations are also openly available via the ICADL webpages (<https://ifdik2023.conf.tw/site/page.aspx?pid=287&sid=1521&lang=en>).

The next program I attended was the ICADL conference. The conference theme for this year is “Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration” with three keynote speakers, Dr Mark Liao, Dr Gillian Oliver, and Professor Shigeo Sugimoto. Their presentation materials are also openly available on the ICADL webpage. In my opinion, providing open access for the conference materials is a progressive step towards disseminating knowledge for the benefit of the public. I presented my research paper on the third day of the conference. The paper was co-written with my supervisor, Professor Ishita Emi, and my colleague Regina Dwi Shalsa Mayzana from Universitas Gadjah Mada in Indonesia. My research mainly discusses about the problems of opening access to digitized cultural heritage materials by examining the current state of their digital collections at the Indonesian higher education (HE) institutions in terms of their accessibility to the users. The accessibility is based on the online presence of the university libraries and the ability of the websites to provide sufficient information about the cultural heritage collections from the perspectives of the users. Eight Indonesian HE institutions with cultural heritage collections were identified. However, they have not been able to provide sufficient or explicit information on the cultural heritage collections online, including information on what collections are available, the basic descriptions of the items, and what users can do with the material.

Therefore, raising awareness among Indonesian HE institutions, particularly the libraries, on the importance of opening access to cultural heritage collections is required. To conclude, the ICADL 2023 was one of the most mind-provoking conferences I ever attended. I hope many academic conferences in the future can make their research materials and discussions openly accessible to the public.



IFDIK(ICADL)2023報告

山崎 基数 (ライブラリーサイエンス専攻 修士2年)

はじめに

私は2023年12月4日～12月7日の間に行われたInternational Conference on Asia-Pacific Digital Libraries (ICADL 2023)にてポスター発表を行った。会場は台湾、台北に位置する国家図書館にて行われた。

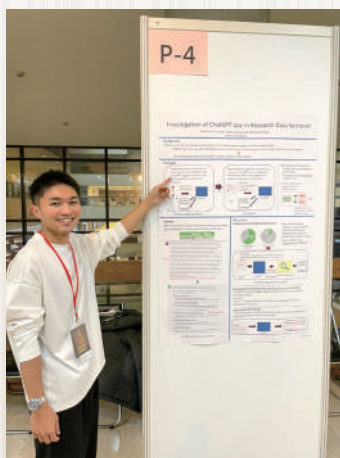
研究概要

近年、膨大な量の研究データが生成されるようになり研究データを活用するためにはそれらを効率的に検索することが重要となっている。従来の検索エンジンやキーワードベースの検索手法では、ユーザが適切なキーワードやフレーズを入力する必要がある、目的のデータに関する詳細な情報を持っていない場合は十分な検索結果を得ることが難しい。

本研究では、研究データ検索の際にLLM (大規模言語モデル) を利用することでユーザが持つ情報と目的の研究データを結びつけることが出来ないかその可能性を模索した。LLMの例の1つとしてChatGPTを利用した。ユーザが研究データを検索する際に、ChatGPTに自身の研究の概要を伝えることで、その研究に役立つ研究データを提案させることを考えた。実際には、ChatGPTに対してある研究データを引用した既存の論文の概要を入力し、その論文の研究を行うのに必要なデータを提案させ、提案された研究データの中に引用された研究データが含まれているかを調査した。実験を行った結果、約6割の研究データと論文のペアにおいて目的の研究データがChatGPTにおいて提案された。この結果により研究データ検索にLLMの利用が有効な可能性があることが分かった。

発表について

私は今回の発表が初めての学会発表であったため、とても緊張した。学会発表が初めてということだけではなく、国際会議でもあったため不安や緊張はより大きなものであった。今回、私はポスター発表を行ったが、発表時間が1時間半と知った時は長時間だろう、最後までやり通せるだろうという気持ちでいっぱいだった。発表は2日目の最後、16時からであったが1日目や2日目の午前は自分のポスターを見直したり、発表の流れを確認したり、どのような質問が来るのだろうかとかあれやこれやと考えていた。私は英語が堪能ではないためきちんと質問を聞きとれるだろうか、相手にわかるように説明できるだろうかと不安を抱えていた。ただ、嬉しいことに発表の前にコーヒーブレイクがあり、そこではおいしいお茶菓子や軽食などが用意



されていて、後述する他大の学生と会話をしながらそれらを口に出すことで緊張をほぐせたため発表が開始する直前にはリラックスすることが出来ていた。

いざ発表の時間になるとすぐに何名か私のポスターの前来てくださり話をするようになった。拙い英語ではあったが一通りの説明を終え、そこから質疑応答・議論を行うことになった。始めは緊張していたこともあり上手く話せずにあたふたしていたところ、隣にいてくれた指導教員の清水先生にご助力をいただき、なんとか最初の発表を終えることができた。その後もあまり途切れることもなく次から次へと私のもとへ話を聞きに来てくれる人がいて発表の時間はあっという間に過ぎ去っていった。発表の前は英語が聞き取れるだろうかという不安で満ち溢れていたが、聴衆の方もゆっくり質問してくれたり平易な英語で聞き直してくれたりとしてくれたおかげで何とか質問に答えることができた。時には上手く答えられなかった時もあったと思うが、上手くこちらの意図が伝わった時にはとても達成感があった。正直、学会の直前にはいきたくない、発表したくないという気持ちもあったがいざ終えてみるといってよかったと心から思う。海外で英語を用いて自分の研究について発表するという貴重な経験を積ませてもらったことにとても感謝している。

他大の学生との交流

指導教員の清水先生のお知り合いということで他大学の先生や学生と交流する機会があった。学会1日目の夜には食事をともにしたり、2日目の発表直前にもそちらの学生と話す機会があり、今回の発表をするにあたって不安や悩みを共有することでリラックスして発表することができた。発表の後には一緒にバーに行って台湾のウィスキーを飲んだりもした。遠い台湾の地で新たな友人を作ることができたのも今回の学会の収穫の1つであった。



謝辞

最後に、今回の学会発表を行うにあたり、ご助力いただいた指導教員の清水先生、冨浦先生をはじめライブラリーサイエンス専攻の先生方や関係者の方にお礼申し上げます。

インターンシップに参加して

橘 万里絵 (ライブラリーサイエンス専攻 修士1年)

私が今回インターンで訪れたのは、日立ドキュメントソリューションズのEPCプロジェクト本部(以下Eプ本)である。EPCとはEngineering(設計)、Procurement(調達)、Construction(建設)の頭文字をとった言葉で、発電所建設などの大規模なプロジェクトのことを指す。Eプ本では主に、原子力発電所の建設に関わるマネジメントを行なっている。前半は主に各部署の説明、後半は業務体験を行なった。

Eプ本では、プロジェクトマネジメント支援業務を「企画」「開発」「運用」に体系化している。Eプ本の中でも、OM(おしごとマネジメント)室、PM(プロジェクトマネジメント)室、PI(プロジェクトIT)室などに別れており、その中にさらに図書管理(文書管理)、工程管理、製品納期管理、工認など、担当ごとに細かい区分がある。インターンでは、各部署に関する説明を受けることができた。

インターン前半は説明だけでなく、オリジンパークや臨海工場の視察も行った。オリジンパークでは、日立の創設者である小平浪平氏の生い立ちに始まり、日立の歴史を学ぶことができた。臨海工場では、原子力発電所の部品が製作されている。知識としては知っていても実物を目にすると思像以上に大きく驚いた。また、「プロトレZ」という日立ドキュメントソリューションが開発したプロジェクトマネジメントの体験ゲームを行う機会もあった。参加者全員で一つのチームとしてプロジェクト成功させるゲームで、用語や対応のヒントはファシリテーターが説明してくれ、プロジェクトマネジメント支援の難しさを学ぶことができた。

インターン後半は原発関連の工程・図書管理とOM室の業務体験を行なった。工程管理とは、細かいスケジュール管理のことで、学園祭を例に自分で工程管理表を作成したり、実際の工程管理システムの操作を行ったりした。図書管理とは文書管理のことで、業務体験としては原発建設の関連書類の紐付けを行なった。原子力発電所の建設のような大規模なプロジェクトでは、スケジュール管理を厳密に行う必要があり、扱う文書も膨大な量のため、それぞれ担当を細かく分けて管理することの重要性が感じられた。また、OM室は、他の部署と少し異なり、原発以外の業務のマネジメントを多く行っており、業務体験としてPEST分析を行った。「今後の働き方はどう変わっていくか」というテーマで、それぞれがニュース記事等を調べ、それらをPolitics(政策)、Economy(経済)、Society(社会)、Technology(技術)に分類して複数人でアイデアを出し議論することで、さまざまな視点の気づきを得ることができ、発展した内容まで考えることができた。

今回のインターンシップでは、自分が今まで学習していたことに深く関わるものもあれば、全く新しい分野も多くあった。今回のインターンで、理系出身でも技術職でなく、理系で培った知識を使いながらさまざまな分野で活躍している方々とお会いすることができ、将来の選択肢に関して視野が広がった。

また、今回、特に工程管理や図書管理について学ぶことが多かったが、これは研究データや公文書の管理ともよく似ていると思った。今後研究をする上でも、それ以外においても、今回のインターンを通して学んだ工程管理(スケジュール管理)やデータ管理について役立てたい。

レコードマネジャー研修生として — 麻生レコードマネジメント株式会社でのインターンシップ体験談

丁瑞楷 (ライブラリーサイエンス専攻 修士1年)

レコードマネジメント会社は、情報の重要性が高まる現代社会で非常に重要な役割を果たし、組織がデータと文書を効果的に管理し、法的コンプライアンスを維持できるようサポートしている。業務内容は顧客や市場に応じて異なるが、情報の安全性、アクセス性、合法性の確保が常に中心に置かれている。レコードマネジャー業務は、組織内の情報やデータを効果的に管理し、保管するための重要な役割を果たす。

ライブラリーサイエンス専攻の学生として、ユーザーのニーズや知の創造と継承プロセスを把握する能力を身につけるとともにレコードマネジャーの仕事の実態を現場で認識するために、2023年8月に筆者は麻生レコードマネジメント株式会社で研修生として働くことを体験した。その度の実習を通じて、いろいろな経験した。

まず、筆者はデータベースの基本原則やデータ整理の方法についてのトレーニングを受けた。データの作成、編集、削除、バックアップ、復元などの基本的な操作を学んだ。機密性の高いデータの取り扱い方法やセキュリティポリシーに関する研修を受け、データへの不正アクセスや漏洩を防ぐための対策を理解し、特に公文書の対処について、業界や地域の法的規制やコンプライアンス要件とそれに適合する方法を勉強した。



スキャニングを体験する筆者

また、書類の電子化について以下のような研修を受けた。最初のステップは、紙の公文書をスキャナーやデジタルカメラを使用してデジタル形式に変換することであった。このプロセスには高品質のスキャン機器と光学文字認識(OCR)ソフトウェアを使った。電子化された公文書は適切なバックアップとセキュリティ対策を施す必要もあり、データの損失を防ぎ、機密情報を守るために適切な保護策を講じることを教わった。そして、デジタル公文書を整理し、適切なカテゴリに分類し、これにより、文書の見つけやすさとアクセスが向上することがわかった。

専門的な作業の他にも、レコードマネジメント会社の従業員としての経験をいろいろとすることができた。テクノロジーの進化や業界の変化に追従して最新のベストプラクティスやツールについて学び続けることやチーム内での協力とコミュニケーションスキルを練習して他のチームと連携してデータ管理のプロセスを効果的に実行することに加え、先輩たちとの接し方も大事な研修経験になった。特に、会社の先輩たちはみんなとても親切で、いろいろ教えてくれた。

その一週間の実習を通じて、レコードマネジメントの精神もある程度理解できた。レコードマネジメントは、顧客の情報を戦略的に活用し、効果的な情報管理を提供することで生産性向上や働き方改革を支援し、情報に関連するリスクから顧客の情報を保護する使命を担っていると思う。LSS専攻の学生として、このようにレコードマネジャーの仕事実態を身近に体験するチャンスを得たことは、まことに幸いだと思う。なによりも、親切で丁寧な麻生レコードマネジメント株式会社の先輩たちのお陰で、本当に短い時間であるが、今度のインターンシップの経験が一生忘れられない大切な思い出になった。



異なる種類やサイズの資料に対応する高性能のスキャン機器

令和5年度 国立公文書館実習に参加して

劉俊延 (ライブラリーサイエンス専攻 修士1年)

概要

2023年8月21日(月)から9月1日(金)まで、2週間にわたり国立公文書館実習に参加した。1週目は「アーカイブズ研修I」を受講し、2週目は本館とつくば分館で実務研修に参加した。

1週目「アーカイブズ研修I」

8月21日(月)からの5日間、様々なテーマの講座とグループディスカッションに参加した。まず、アーカイブズ(公文書館)の役割と歴史に関する内容が取り上げられた。具体的には、国際標準やアーカイブズの理論的な考察、日本型アーカイブズに焦点を当てた歴史の概観が説明された。また、「公文書等の管理に関する法律」や「特定歴史公文書等の保存、利用及び廃棄に関するガイドライン」についての法令に関する説明も受けた。その後の講義では、評価選別業務の基準や国内の評価・選別実施状況についての情報が提供され、電子公文書の保存管理やデジタルアーカイブ、目録作成に関する知識も深めることができた。

グループディスカッションでは、「公文書の評価選別」のグループに参加した。自己紹介と問題意識の共有を通じて、グループメンバーとの交流を深めた。私は実務経験がないため、職員の皆様の問題意識を理解する貴重な機会となった。次に、2回目のグループディスカッションでは、前回に残った課題を中心にグループ討論が継続され、各グループが担当テーマに関する業務上の課題と解決策について報告した。また、現場で直面されている課題について学び、他のグループの報告からも得られた現場の貴重な情報を活かして今後の研究に役立てていきたい。

2週目「実務研修」

8月28日(月)からの5日間、実務研修が東京本館とつくば分館において行われた。保存、修復、利用、展示、評価選別、目録作成、利用審査など国立公文書館の業務を実際に体験させていただいた。

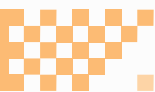
保存と展示に関する実務については、書庫清掃・排架整理、資料の埃取り・カバーかけ、金属除去、スキャニング、デジタル化作業を学んだ。また、資料の展示計画のプレゼンテーションと実際の展示方法を実践した。

利用と修復に関する実務については、利用業務では資料の出納や誤返却防止の方法について学び、修復業務では虫損のドライクリーニングや和装本の裏打ち作業を体験した。

評価選別、目録作成、利用審査に関する実務については、モデルを使った評価選別の実習や特定歴史公文書等の目録作成に取り組んだ。また、審査基準に基づいた資料の公開可否の判断方法を学んだ。

最終日には、二週間の実習で得た知見や経験の振り返りと今後の技能・能力、キャリアについての専門官と意見を交換した。実際の業務の多様性や専門性を理解し、アーキビストの仕事が理論だけでなく倫理観も求められることを理解した。

最後に、統合新領域学府がこのような貴重なチャンスを頂いたことを感謝申し上げる。今後、国立公文書から学んできた知識をより一層自分の研究に活用していきたい。



小柏 香穂理 (お茶の水女子大学 教学IR・教育開発・学修支援センター)

2014年に山口大学大学評価室での自己点検評価システムの運用に携わってから、大学評価やIR (Institutional Research) の分野における実践研究を続けて約10年になります。これまでずっと同じ目的のもと、研究を進めてきたつもりでしたが、今回、博士論文をまとめるにあたって一番苦労した点は、博士論文の根幹(串)を示すことであつたと思います。指導教員の三輪宗弘先生よりご助言をいただきながら、何度も繰り返し、議論をいたしました。三輪先生からのご助言によって、自分の中でぼんやりしていたものが、少しずつクリアになってきたように思います。これまでの研究活動のデータを一度まっさらの状態にして、そこからもう一度再構築していくことで本当に必要なもの(串)ができたと思っています。

データに基づく大学評価に関する実証研究 ―実質的な中期計画文書の特徴と意思決定過程―

【学位論文の要旨】

自己点検評価と継続的な改善はいかなる組織でも必須である。我が国の大学でも、2004(平成16)年に認証評価と法人評価が義務化され、データに基づく説明が社会的にも求められている。国立大学では2004年4月に法人評価の開始とともに中期計画を策定することが義務化された。さらに、2020年4月に私立大学においても中期計画の策定が義務化され、中期計画への関心は高まっている。しかし、大学が策定した目標や計画には、記述が抽象的で目標が明確ではない点があり、データに基づく評価を適切に行うことが難しいと厳しく指摘されている。また中期計画を策定する際には大学の意思決定が重要な要素になるが、意思決定の過程を示す文書や記録に関する組織的な対応(管理)がなされていないことも指摘されている。これらの課題を解決するために、大学評価が義務化された背景や、国立大学と私立大学の法令や組織構造の違いによる課題を取り上げる。具体的には、国立大学においては法人化以降の体制整備、私立大学においては財政破綻を起こさない財政管理が大きな課題となっている。また私立大学では、国立大学とは異なり、義務化されても中期計画を策定しているかどうか、実際に文部科学省(国の機関)においてチェックすることはしていないため、内容的にどの程度の中期計画を策定しているか、各大学によってかなりのばらつきがある可能性が高い。本研究ではこれらの課題解決に向けた方策を提案し、様々な事例研究により中期計画策定の本質的な成功要因の一端を明らかにした。

第1章では、国立大学の中期計画文書についての分析を行った。中期計画文書の問題は、目標や計画の書き方が明確でない場合と、大学の特徴が明示されていない場合に大別できる。明確な記述として、文部科学省からは、「目的、投入(インプット)、活動(アクティビティ)、アウトプット、アウトカム」の5つの要素からなるロジックモデルに基づく記述が推奨されている。上述の5つの要素のラベ

ルが付けられた文の集合を学習データとして、機械学習を適用することで自動識別の性能と識別のための特徴語を分析した。「活動」については高い識別性能で、解釈可能な少数の特徴語が得られた。一方、「アウトカム」については低い識別性能であり、この要素が明確な記述を困難にする要因であることを定量的に示した。次に、自大学の特徴記述の課題については、記述の良し悪しを他大学との識別難易度としてとらえ分析し、24大学の中で機械学習による自動識別性能の高い大学の特徴を明らかにした。また、キーワードを検索語として入力して共起関係により抽出された各大学の文書の特徴語を、2次元上に可視化するシステムを実現した。

第2章では、私立大学を対象に、財務指標が向上している大学の中期計画文書の記述を分析した結果、具体的な記述と明確な目標が書かれていることを実証した。日本私立大学協会では、2009年以降から私大マネジメント改革に関する研究が行われており、各大学の成功事例に共通する改革内容をあげている。本章では、私立大学の文書管理の問題に触れながら、私立大学での成功事例(模範的な大学)を対象として、その中期計画文書中の特徴となる文と関連するデータを明らかにすることを目的とする。財務に関する政策・ビジョンの実効性に着目し、模範的な3大学の特徴となる文を抽出した結果、各大学の経営戦略の違いを明らかにすることができた。模範的な大学の事例では、具体的な記述と明確な目標が書かれている文が特徴となる文として抽出され、関連する数値データ(教務、財務)の所在や管理方法も組織化されていることが明らかになった。さらに中期計画文書とデータを共有(リンク)し、効果的に活用するためのメタデータを考案した。

第3章では、重要文書の公開が義務付けされている国立大学を対象に大学の事例研究を行い、入手した資料の中で、意思決定過程の記述をどの程度読み取ることができるかを分析した。前章に

において、具体的な記述(戦略)と根拠となる関連データの所在や管理方法が組織化されていることが、成功事例の大学に共通する特徴であることを明らかにしたが、どのような意思決定過程を経て中期計画文書が策定されているのかは、これらの資料だけでは不十分である。国立大学の法人化による組織改変や、私立大学においては大学と学校法人という複雑な組織体系から、意思決定過程の複雑さが指摘されており、日本のすべての大学において、意思決定過程の最適化は重要な課題である。そこで国立大学を対象に、今回、認証評価受審のための自己点検評価書の記述内容に基づいて分析した結果、大学の意思決定過程の流れの概要を把握することができた。さらに大学文書館に所蔵されている資料の調査分析に基づき、意思決定過程の記録の具体事例を示した。

第4章では、私立大学における中期計画の策定プロセスを探るために、各大学の自己点検評価書の記述内容から、各々の大学の意思決定過程の流れの概要を把握し、意思決定過程の記録の具体事例を示した。前章で論じた国立大学とは組織構造が異なるこ

とから、私立大学においては、意思決定の過程はさらに複雑であることが明白であり、最適な意思決定過程を明らかにすることは重要な課題として残っているが、意思決定機関である「理事会」と、諮問機関である「評議員会」の両者の運営実態についても、公開データにより分析した。さらに学長インタビューの分析結果をもとに、私立大学における意思決定過程の実態の解明を目指すとともに、その記録の具体事例を示した。各大学の文書管理において、自己点検評価書を重要文書として位置付け、長期的に保存することが望ましいことを示した。

中期計画文書の記述が「抽象的」で「データに基づいていない」という課題を起点として、それに起因する課題解決に向けた大学のすすめるべき方策を提案した。国立大学や私立大学の事例研究により中期計画策定の本質的な成功要因を実証することで、有益な中期計画文書とはどのような内容が盛り込まれ、記述されているのかを跡付けた。

[謝辞]

本論文の主査である九州大学附属図書館 三輪宗弘先生には、本論文を執筆するにあたり、一字一句、心を込めて文章を書くことの大切さをご指導いただきました。研究者としての生き方、考え方などを肌で感じながら学ぶことができたこと、心より御礼申し上げます。九州大学名誉教授 廣川佐千男先生には、私が本専攻に2016年10月に入学してから今日までの8年間、ご指導いただきました。研究指導だけでなく、就職活動などについても親身になって相談に乗ってくださいました。心より御礼申し上げます。副査の東京工業大学企画本部 森雅生先生には、私が山口大学大学評価室在籍時より今日までの10年間、ご指導いただきました。当時の九州大学大学院の授業を受講したことがきっかけで、本専攻をご紹介くださり、研究をはじめ種々のご助言をいただきました。心より御礼申し上げます。副査である九州大学大学文書館 藤岡健太郎先生、別府大学文学部史学・文化財学科 針谷武志先生には、本論文をまとめるにあたり、多大なるご支援とご指導を賜りましたこと、厚く御礼申し上げます。九州大学大学院システム情報科学研究院 富浦洋一先生には、不足している点などをご助言くださり、本論文の執筆にあたり、お忙しい中、丁寧にご指導くださり、厚く御礼申し上げます。

本当にこれまでの長い期間に多くの皆様からのご助言やご支援をいただきました。また本研究は多くの科学研究費の助成を受けております。この場を借りて御礼申し上げます。



IIAI-AAI 2023
郡山で開催された国際会議



熊本県庁での
公文書選別作業の様子



読むべき最新論文の判別方法

A Method for Identifying Recent Papers to Read

佐々木 陸 SASAKI Riku

新たな研究着手の際、その研究テーマに関して、扱っている課題、手法とその性能、理論などの周辺トピックを調査する。その際、レビュー論文に引用されている論文を調査することは一つの有用な手段である。しかし、実際ある研究テーマで用いられている手法やその性能などに関して、最新の動向を論じている論文は、まだレビュー論文によって引用されていない場合がほとんどである。引用データベースを用いて、レビュー論文で紹介されていた特定の論文を引用している比較的新しい論文を見つけることは可能だが、引用の理由はさまざまであり、引用元論文すべてが同じトピックを扱っているとは限らない。また論文の質を問わず、多数の論文が見つかるうえに、被引用数といった客観的な評価にもまだ差がないため、読むべきものを絞るのが難しいという問題点もある。そのため、最新の論文のサーベイは一般的な論文のものと比較して、より手間と時間を要する作業となっている。

そこで本研究では被引用論文の抄録、引用元論文の抄録、導入、結論、引用文に基づいて、引用関係にある2つ論文において、引用元論文が被引用論文と同じトピックを扱っており(条件①)、かつそのトピックを代表するよい(条件②)論文を判定する方法を提案する。ここでいうトピックとは研究対象やそれに対する主要なアプローチを指す。この方法を用いて、与えられた研究トピックの最新の論文を見つけることができる。まず、レビュー論文等を用いてその研究トピックにおける代表的な論文を1つ見つけ、これをキー論文とする。引用データベースを用いて、キー論文を引用している論文を見つけ、提案法に基づいてキー論文を引用している論文がキー論文と同じトピックを扱っており、かつそのトピックを代表するよい論文かどうかを判定する。これによってキー論文を引用している論文の中で読むべき論文を厳選することができる。これによって自身の研究トピックを扱うレビュー論文で引用されている論文や先行研究と引用関係にある論文を調べる時間を大幅に削減することにつながる。また、それらの中でまたキー論文を設定し、その引用論文を判別することで上記の2つの条件を満たすより新しい論文を見つけることができ、最終的には客観的な評価がまだない

最新の論文においても先述した2条件を満たすものを判別することができる。

本研究では引用関係にある論文ペアのうち、先述した2つの条件をともに満たすものの判別を自作したデータセットで学習した判別モデルを用いて機械的に行い、最終的には判定結果をRecall、Precisionで評価する。情報検索の一種で、見逃しが少ないことを評価するためにRecallを重視する。

判別モデルでは被引用論文の抄録、引用元論文の抄録、導入、結論、引用文を入力とし、自然言語モデルであるBERTを用いて、入力テキストの分散表現を求め、テキストを数値として扱えるようにする。最終的な出力は入力が先述した2つの条件を満たす正例である確率を表し、閾値0.5で正例・負例に判別する。

データセットは、データ作成者(annotator)の主観の影響がでないように、機械的な方法でデータを作成する。レビュー論文はある特定の研究トピックを中心に論文をサーベイしたものであるから、同じレビュー論文で引用(紹介)されている論文は、同一のトピックを扱っていて、そのトピックにおいて代表的な論文であると考えられる。レビュー論文で引用(紹介)されている論文XとXを引用している論文Yのペアでデータセットを作成した。Yがレビュー論文で引用されているものを正例、引用されていないものを負例とした。

本実験はGoogle Colaboratory上でPytorchを用いて行った。学習率やモデルの層の出力サイズなど、様々なパラメータを調節して実験を行い、評価には5-Foldクロスバリデーションを用いた。データセットを5つに分け、モデルの学習とテストをそれぞれ異なる部分で行う。そして、5回の結果の平均を評価した。その結果、最終的にPrecisionは約83%、Recallは約87%を達成した。

また、データセットの引用元論文の被引用数に正例と負例の間で差があったことから、追加検証において判別モデルの判定結果と引用元論文の被引用数の関係を調べた。その結果、この判定モデルは基本的には論文の質だけで判別を行っているわけではないが、ある一定以上の被引用数を持つ論文に

ついては、その質のよさのから正例と判定されやすいことがわかった。

本研究では引用関係にある2つの論文間の判定を行い、最終的にPrecisionは約83%、Recallは約87%という高い分類精度を達成したが、今回は比較実験を行わなかったパラメータも複数存在し、それらを含めてより細かく調整を行うことで、同モデルでもさらなる精度向上が見込まれる。この判定方法を用

いて、1つのキー論文から条件①、②をともに満たす最新の論文を検索できるシステムができると、より使いやすいものになる考ことが期待される。また本研究で作成したデータセットの論文は本研究トピックである引用分類に関するものを中心に集めたが、他分野の論文においても同様の精度を達成できれば、より多くの研究者の役に立てるものになると考えている。

流通を考慮した研究データのメタデータ付与支援

Support for adding metadata to research data in consideration of distribution

篠原 桐 SHINOHARA Kiri

研究データとは、研究のために収集・作成・観測したデータで、研究の成果である論文の根拠となるデータ、及び研究成果そのものであるデータの両方を含むものである。近年、オープンサイエンスの高まりにより、研究データの公開が促進されており、その利活用によるイノベーション創出が期待されている。Scopusに掲載されているデータ論文の掲載数も、2014年では2件だったのに対し、2023年では7341件とその数を大きく伸ばしている。また、Web of ScienceのData Citation Indexによると、毎年100万件近くものデータセットが登録されている。公開されている研究データを利用するためには、利用者が研究データを検索し、それを発見できる必要がある。

研究データを検索する際は、一般的に、研究データに付与されているメタデータを通して検索が行われる。研究データが広く利用されるためには、研究データが作成された分野とは専門の異なる利用者からも検索される必要があるが、専門としていない分野の検索クエリを思いつくことは困難である。そのため、研究データの流通性を高めるためには、メタデータを付与する際に、専門としていない利用者が検索しやすいメタデータを付与する必要がある。

一般的に、研究データに付与するメタデータの登録は、その研究データの作成者が行う。しかし、作成者によるメタデータだけでは、記述が不十分である場合も多く、有効な検索が行えない可能性がある。

そこで、本研究では生成AIを利用することで研究データの

作成者が研究データにメタデータを付与する際に、研究データが検索される可能性の高い、流通することを考慮したメタデータの付与の支援ができるのではないかと考え、その可能性について調査を行った。

本研究では、生成AIのモデルとしてChatGPTを利用した。Scopusからデータ論文を収集し、各データ論文に対し、ChatGPTに研究データのタイトルと、データ論文に書かれた概要を入力し、このデータを利用する可能性のある研究を考慮したキーワードを、与えた情報を含まないよう10個提案させた。提案されたキーワードを調査するために、主観評価と客観評価の2種類の評価を行った。主観評価では、48個のデータ論文に対し、提案されたキーワードを5段階で評価した。評価は数字が大きいほど高いものとなる。客観評価では、50個のデータ論文に対し、ChatGPTに提案されたキーワードを検索クエリとし、Semantic Scholarを利用して、そのデータ論文を引用した論文が検索可能かを調査することで、間接的に評価を行った。

主観評価に用いた48個のデータ論文に対し提案された合計480個のキーワードの中で、与えた情報に含まれていない新たなキーワードは実際には401個だった。データ論文あたり平均8.35個の新たなキーワードが提案された。

提案されたキーワードの主観評価では、自身で評価を行うと共に、その評価が信頼できるかを確認するために、第三者に同様に主観評価をしてもらい、その結果との κ 係数を調べた。

第三者による評価では、48個の研究データのうち、10個は評価基準の学習に利用し、残りの38個の研究データに対して評価を行った。その結果、自身の評価の平均スコアは3.45、第三者による評価の平均は3.24となった。 κ 係数は、0.50となり、十分な一致を示すことはできなかった。評価3以上のキーワードは、研究データにメタデータとして付与できると考えられるため、評価を3以上と2以下の2値で分類し、その2値に対して κ 係数を調べたところ、0.66となり、十分な一致を示すことができた。一致度が高く、評価3以上のものが多いことから、提案されたキーワードは、メタデータとして付与される可能性のあるものが多く提案されていると考えられる。

間接的な客観評価では、50個のデータ論文に対し、31個のデータ論文において、そのデータ論文を引用している論文を検索することができた。また、検索に成功した31個のデータ論文に対し、元々持っていたキーワードを検索クエリとして、同様にそのデータ論文を引用している論文を検索したところ、18個のデータ論文では、元々持っていたキーワードでは検索を行うこ

とができなかった。このことから、この18件において、提案されたキーワードをメタデータとして付与することで検索される可能性が高まると考えられる。

新たなキーワードの提案数、主観評価によるキーワードの有用性、客観評価によるキーワードの妥当性から、メタデータの付与の支援に有効に利用できる可能性が高いと考えられる。

本研究では、生成AIを利用することで、研究データのメタデータとして、利用用途を考慮したキーワードの提案を支援できると考え、その可能性について議論した。その結果、新たなキーワードの候補の推薦に有効である可能性が示された。しかし、検索結果の向上については、利用法を考慮したキーワードを付与できる可能性を示すことはできたが、直接的な客観評価をすることができないことなどから、可能性を示唆するに留まった。今後は、主観評価を行う人数を増やす、評価を行うデータ数を増やす、データの偏りをなくすなどを行うことで、提案されたキーワードの有効性についてより議論していく必要がある。

研究内容からの研究データ検索

Research data retrieval by using research summary

近年、インターネットをはじめとする情報技術の発達により急速なペースで蓄積される研究データは、科学研究や技術の進展において不可欠なものとなっている。新しい発見や知見を得るためには、これらのデータを適切に利用して研究を行うことが重要である。研究データの検索はデータに付与されたメタデータに対してのキーワード検索によって行われる。しかし、従来のキーワードベースの検索手法ではユーザが適切なキーワードやフレーズを入力する必要があり、目的の研究データに関する詳細な情報を持っていない場合は十分な検索結果を得ることが難しい。そこで、本研究では研究データ検索におけるLLM(大規模言語モデル)の有効性を検討した。LLMに基づいたChatGPTを活用し、より研究者が欲しい研究データにたどりつきやすくするための手法を提案する。ChatGPTはLLMに基づいているため、ユーザが入力した文章に対して人間のように自然な回答を行うことができる。ユーザがChatGPTに自身の研究の概要を伝えることにより関連する研究データや役に立つ研究データについて情報を得られる可能性があると考えた。キーワード検索と異なり、LLMを利用することで入力

山崎 基数 YAMASAKI Motokazu

した文章から関連のある研究データを推測し、ユーザが特定の研究データに関する詳細な情報を持たない場合でも目的の研究データを得やすくなる可能性があると考えた。

本研究では研究者が計画している研究の内容を文章で表現し、その文章から自身の研究を行うのに役に立つ研究データを入手するのにChatGPTが有用であるか検証した。しかし、実際に研究者に研究内容を入力してもらい実験を行うことは多くの研究者の協力を得る必要があることから困難である。そのため、我々はChatGPTに既存の論文のAbstractを入力し、その論文に書かれている研究の役に立つ研究データを提案させることを考えた。

本研究は実験Iを行った後に、実験Iで得た知見をもとに実験IIを行った。基本的な実験の手順は同じである。実験Iと実験IIの相違点としては主に使用した論文や研究データの変更である。まず、実験Iと実験IIで共通する手順について説明する。ChatGPTはある時点までの情報を学習しており、それに基づいて回答を生成している。ChatGPTが学習していない論文のAbstractを利用することで研究者がこれから行う研究の

概要を入力することを再現した。具体的には既存の論文のAbstractと研究データをChatGPTに提案させるための文章を入力した。研究データ、正確には研究データに付随するデータ論文を50個集めた。また、研究データ1つにつき研究データを引用している論文を2本集めた。つまり研究データ(に付随する論文)50個とそれを引用している論文100本に対して実験を行った。ChatGPTに提案させた研究データの中に目的の研究データ(研究データを引用している論文とペアとなっている研究データ)が含まれているかを調査し、含まれている場合には「発見」、含まれていない場合には「未発見」としてその数をカウントした。

実験Iでは100本の論文のうち65本で研究データの発見をすることができた。しかし、実験Iで使用した論文にはAbstractに研究データ名が直接記載されているものや研究データを引用しているが研究の本筋と研究データが結びついていないものが含まれていた。Abstractに欲しいデータの名前が記載されている論文は研究者が目的の研究データに詳しくない状況を再現するのに不適切であると考えた。また、研究データが論文内で活用されていないものも研究者が自身の行いたい研究の役に立つデータを探す状況の再現を行う上で不適切な論文であると考えた。そのため、Abstractに研究データ名が記載されている論文、研究データが活用されていない論文を差し替えて実験IIを行った。実験IIでは、100本の論文のうち67本

の論文で研究データを発見できた。更に、キーワード検索との比較を行うために実験で使用した論文のAbstractから検索ワードを抽出し、それを用いてGoogle Dataset SearchとGoogle Scholarのそれぞれで検索を行った。キーワードの選定に公平性を持たせるためにChatGPTを用いてAbstractからキーワードを抽出した。Google Dataset Search、Google Scholarでの検索結果上位30件における研究データ発見数はそれぞれ17個と23個でChatGPTに提案をさせた場合と比べて少なかった。

本研究では研究データ検索におけるLLMの利用可能性を調査するためにChatGPTを用いて研究内容からの研究データ検索が行えるか実験した。実験の結果、実験で使用した100本の論文のうち、67本の論文において目的の研究データを発見することができた。これによりLLMを利用することが研究データの検索を行うのに役立つ可能性を確認した。今後の課題としては、目的のデータによりたどり着きやすくするための質問文の工夫を行うことや回答内容について更なる分析を行うことが挙げられる。また、見つかった研究データやChatGPTが頻繁に提案した研究データのメタデータを分析することで研究データを見つけやすくするためのメタデータについて知見を得られる可能性がある。このことは研究データを従来のキーワード検索で見つけやすくすることにも寄与する可能性がある。



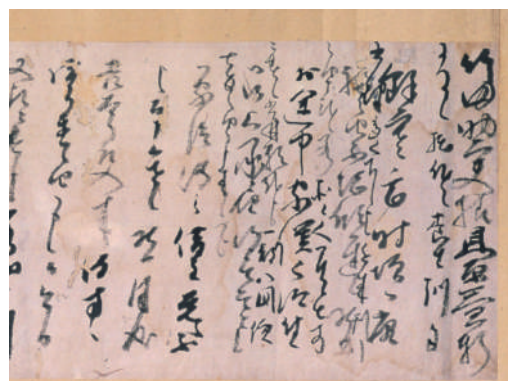
木土 博成 (ライブラリーサイエンス専攻 准教授)

2023年4月に比較社会文化研究院に着任し、本専攻を担当しています。わたしの専門は歴史学、なかでも、日本の近世期(16~19c)を対象に、主に九州・琉球の歴史を研究しています。

歴史学に限らず、人文・社会科学などの分野では、近年、大きな変革がおきています。それは、デジタルアーカイブズの波です。たとえば、古文書や古記録・絵図といった歴史資料(史料)が、各所蔵機関のHPなどで続々とデジタル公開されており、本学でも附属図書館HPの「貴重資料デジタルアーカイブ」で見ることができます。また、著名な史料集や、自治体が編纂したその地方の歴史書である自治体史が、テキストデータやPDF形式で公開されてきているのも、近年の特徴です。



九州大学附属図書館「貴重資料デジタルアーカイブ」より



書状(貝原益軒)部分

かつて、歴史は足で書くもの、つまり史料館や史料が伝わっている現場を歩き回って史料を見出して書くもの、と教わりましたが、いまや歴史学の論文は自宅で、パソコンを前に「目で書くもの」になりつつあるのです。

ただ、不思議なことに、史料や論文へのアクセス環境が良くなったという声はよく耳にしても、それにともなって研究の質が上がったという話はあまり聞かれません。どちらかというと、近年の歴史学をはじめとする人文科学研究は細分化がすすみ、かえって時代像や本質が見えにくくなってきているという嘆きの声が多いのです。それは要するに、検索して得られたデータを組み合わせるだけでは、そこに人間の思索が介在せず、薄っぺらいものになってしまうという懸念です。

こうした現状に危機感を持ちつつ、それでもなお、わたしは人間の可能性を信じています。人はデジタル技術を使いながら、深い思考水準を保つ努力をすることができます。それはつまるところ、他人ではなく自分の人生を日々、ああでもないこうでもないと思ひながら、胸を張って生きていこうとする努力のことです。こうした努力を重ねることで、よりよい情報の受け手になれるのです。

一方、ライブラリアンといった情報の発信側もまた、たとえばデジタル画像をたんに公開して安堵するのではなく、利用者をより深いところに誘う仕掛けを作ることができます。仕掛けのアイデアもまた、AIや他人ではなく、自分自身の生活や仕事を日々、ああでもないこうでもないと思ひながら前に進めていくなかで、芽生えてくるのです。

情報は情報ですが、それを発信し、受信するのは人であり、その人には、表には出ない試行錯誤の過程がある。この当たり前のことを大切に、これまでの人文科学が培ってきた良質な伝統と、急速に進むデジタル技術との融合を意識し、教育・研究にあたる所存です。



谷口 雄太 (ライブラリーサイエンス専攻 准教授)

私は2023年4月に情報基盤研究開発センターの教育情報基盤研究部門に准教授として着任し、同年5月よりライブラリーサイエンス専攻の専任教員となりました。専門はデータマイニングで、特に近年は教育におけるデータサイエンスともいえるLearning Analytics分野で研究をしています。

近年、教育や学習の分野においてもデジタル技術の導入が進んでいます。大学では履修登録、出欠管理、資料配布、成績登録などが既にデジタル化され、大学以外でもコロナ禍でのオンライン授業の普及やGIGAスクール構想の実施により学習環境のデジタル化が加速しました。デジタル教科書の導入も進むことで、デジタル技術の活用は今後さらに広がると考えられます。

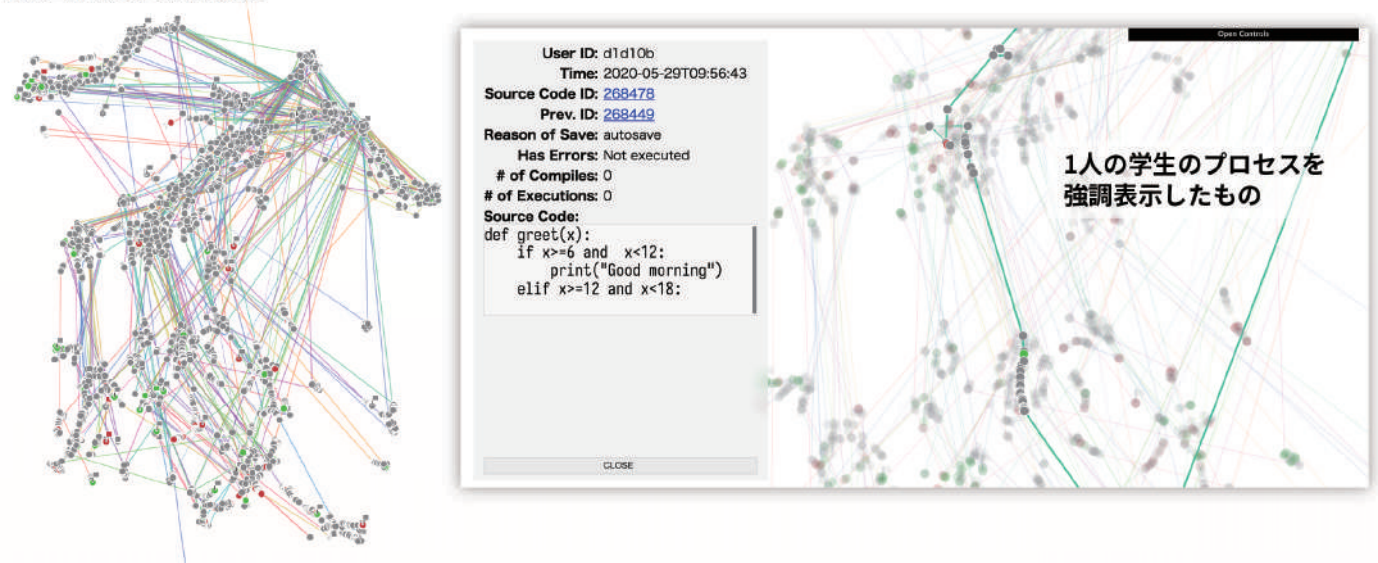
こうしたデジタル化にともない、提出物や成績などの情報、教師や学習者の活動が電子的に記録されるようになっていきます。蓄積された多様な教育学習データを分析し、その結果得られた知見を用いてよりよい教育や学習を実現しようとするのがLearning Analyticsという研究分野です。私は中でもプログラミング学習を対象として研究に取り組んでいます。

プログラミングは近年重要なスキルの1つとして認識され、その学習に注目が集まっています。しかし、プログラミングの学習には多数の困難があり、学習者への適切な支援が重要となりますが、教師による支援だけでは限界があります。私の研究では学習者のプログラミングの過程を分析することで、教師や学習者へ有益なフィードバックを行うための手法を開発しています。例えば以下の図は、我々が開発したプログラミングの過程を可視化する手法を使い、実際の授業におけるクラス全員のプログラミングの様子を視覚的に把握できるようにしたものです。このような図を用いれば、教師にとっては学習の進捗や問題の解き方を把握するのに役立ちますし、学習者にとっては自身を他者と比較することでモチベーションを維持したり様々な解答を知るのに役立つと考えられます。

蓄積されたデータからいかにして有益な情報を取り出すか、取り出した情報をどのようにして必要とする人々に届けるかはこの情報化社会においてとても重要な課題です。ライブラリーサイエンス専攻の他の先生方とともに、このような課題に取り組み、またライブラリーサイエンスを学ぶ学生の方々の目標実現に貢献して参ります。

クラス全体のプログラミング演習の過程

※ 1色が1人の学生に対応





中川 奈津子 (ライブラリーサイエンス専攻 准教授)

2024年2月に着任しました。専門は言語学と人文情報学です。

ことばの仕組み、特にヒトの認知的な傾向や談話における実際の使用がことばにどのように反映し、ことばを制約しているかについて研究しています。実際に産出されたデータ（コーパス）を使って質的・量的に調べたり、フィールドに行って母語話者の方に聞き取り調査をしたり、談話を録音・録画してコーパスを作ったりしています。このようにして得た研究データにどのような情報を付与して公開できるかについても取り組んでいます。また、ことばを学ぶことの社会的意義も重視し、地域コミュニティの人々やデザイナーらとともに方言を調べたり、絵本を作って展示したり、ことばの仕組みを考えさせるようなゲーム作りも行ってきました。さらにこれからは、地域の言語や文化を有機的に表現できるような、デジタルの世界と現実世界を結ぶネットワークづくりのために働いていきたいと思っています。このように、ことばを研究することと、デジタルデータを作り出したりそれに様々な情報を付与してことばについて探求することは私の研究テーマにおいては不可分であり、したがって言語学と人文情報学も密接に結びついています。



撮影協力: 村松佳優 (ムシミル)

地域コミュニティで調査を行っているとき、言語だけを取り出して調査したり、成果を公開したりするのが困難になります。貴重な歴史的証言や文化的知識についての語りも蓄積されるので、このような証言や語りと言語の知識を結びつけて、有機的な知のネットワークとして表現して公開していきたいと考えています。

研究データ管理(RDM)支援人材育成プログラムの概要

(1)本プログラム設置の背景

2021年3月に閣議決定された『第6期科学技術・イノベーション基本計画』に示されたとおり、大学等の研究機関は、新たな研究システムの構築のために「信頼性のある研究データの適切な管理・利活用促進のための環境整備」と強化が喫緊の課題として求められています。また、2023年1月には、文部科学省オープンサイエンス時代における大学図書館の在り方検討部会による審議のまとめにおいて、大学図書館等のこれまで大学等の研究機関において研究支援をしてきた部署が、研究データ管理支援にも積極的に関わる必要性が示されました。このように、日本の大学等の研究機関においては、研究データ管理に向けた環境整備だけではなく、研究者が適切に、また効率的に研究データを管理することができるよう支援人材が求められています。

九州大学統合新領域学府ライブラリーサイエンス専攻では、これまで、ユーザーの視点に立った新たな情報の管理と提供の場を提案することを目指して、高度な情報専門職の育成を行ってきました。この度、上記に示した支援人材の育成ニーズにこたえるため、令和5年度から研究データ管理支援に焦点を絞り、研究者が研究データ管理(RDM)をすることを支援する人材を育成する「研究データ管理支援人材育成プログラム」を開講することにいたしました。

(2)育成する人材像

研究データ管理支援人材育成プログラムでは、以下のような人材の育成を目的としています。

- 1.研究データ管理の遂行にあたり適切な支援ができる人材
- 2.各々の研究組織において研究データ管理支援のための体制構築やサービス設計ができる人材
- 3.研究データ管理支援人材として組織を超えたネットワークを構築できる人材

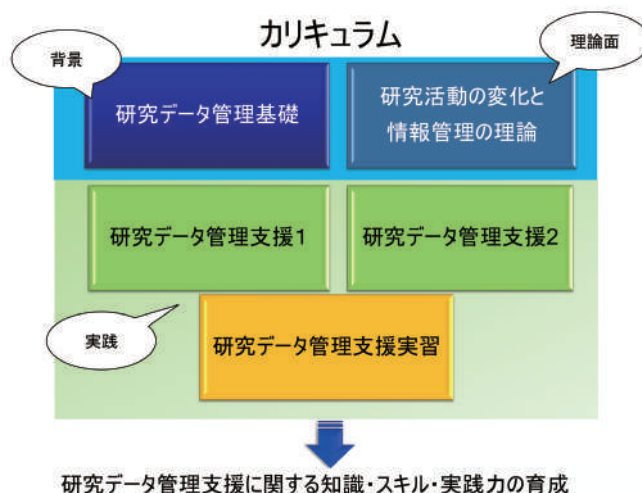
大学をはじめとする研究機関の図書館職員、URA、研究推進に関わる部署等の第一線で活躍する実務家などが積極的に履修されることを期待します。研究データ管理支援に関する最初の一步を踏み出すためのプログラムです。初歩から学ぶことができますので、前提知識や経験は必要ありません。また、九州大学のすべての学府に在籍する大学院生も受講できますので、受講を期待します。

(3)カリキュラム

カリキュラムは、5科目5単位(1科目、15時間)で構成されており、全て必修です。各科目の関係性は以下の図、科目の詳細については「科目一覧」(ライブラリーサイエンス専攻HPより確認可能)をご覧ください。

「研究データ管理基礎」は、秋学期(10月中旬頃)の土日を使って、九州大学伊都キャンパスにて対面授業で実施します。「研究活動の変化の情報管理の理論」は秋学期の土曜日にオンラインのライブ授業で実施します。その他の3科目については、次年度の春学期の土曜日に開講予定です。社会人の方々が受講しやすいようにできるだけ配慮いたします。

各年度の授業スケジュールの詳細については、「科目一覧」(ライブラリーサイエンス専攻HPに掲載)をご確認ください。



研究データ管理支援に関する知識・スキル・実践力の育成

(4)履修証明プログラムの概要

「研究データ管理支援人材育成プログラム」では、社会人を統合新領域学府の科目等履修生として受け入れ、履修により合格した科目については統合新領域学府により単位を認定します。本プログラムでは、75時間以上の課程の修了者(5科目5単位以上の修得)に対して履修証明書を交付します。

九州大学の大学院生が本プログラムの科目を履修した上で、統合新領域学府により合格科目の単位認定を受け、「修了要件」に記載の要件を満たした場合も、履修証明書を交付します。

(5)その他

プログラムの概要、科目や募集要項などの詳細は、ライブラリーサイエンス専攻HPをご覧ください。

ライブラリーサイエンス専攻HP ▶ <https://www.ifs.kyushu-u.ac.jp/lss/>

専攻紹介

ライブラリーサイエンス専攻では、新しい枠組みにもとづき、
知の創造に関わる課題を知の統合によって解決し、
ユーザーの視点に立った情報の管理・提供を担う人材を養成します。

学 府 名：大学院統合新領域学府 Graduate School of Integrated Frontier Sciences

専 攻 名：ライブラリーサイエンス専攻 Department of Library Science

学 位 名：修士（ライブラリーサイエンス） Master of Library Science

博士（ライブラリーサイエンス） Doctor of Library Science

入学定員：修士課程 10名

博士後期課程 3名

開設時期：修士課程 平成23（2011）年4月

博士後期課程 平成25（2013）年4月

教育研究上の理念／入学から修了まで

【理念】ユーザーにとって真に意義ある情報の管理・提供の実現



情報の管理・提供に興味を持つ、修士課程修了者



情報の管理・提供に携わる社会人



情報の管理・提供に興味を持つ、学部卒業生

【博士課程】2013年4月開設、入学定員3名

情報の管理・提供に関する課題を解決し、ライブラリーサイエンスを深化・発展させる人材（研究者）の養成

- ①図書館情報学、アーカイブズ学、情報科学をカバーするT字型の研究能力の養成
- ②ユーザーの視点に立って、意義ある情報の管理・提供を考察する能力を養成
- ③従来の枠組みにとらわれずに、情報の管理・提供に関するより良い枠組みを探究する能力を養成

【修士課程】2011年4月開設、入学定員10名

情報の管理・提供の場で実践的に活躍する人材の養成

- ①ユーザーのニーズと知の創造・継承プロセスを把握させる教育
- ②図書館情報学とアーカイブズ学を統合した一体化教育
- ③情報の管理・提供を実現する情報通信技術の教育
- ④情報法制と情報流通制度に関する教育
- ⑤情報の管理・提供の在り方、ライブラリーの新たな機能を探究する能力の養成

大学等の教育・研究機関
ライブラリーサイエンスを
深化・発展させ、人材養成
を行う大学等の教員

情報の管理・提供組織
情報の管理・提供の新たな
モデルを提案し、組織を
牽引するリーダー

公文書館

記録管理の専門企業

図書館

情報サービス関連企業

情報通信技術関連企業

研究機関

ライブラリーサイエンス専攻活動日誌 2023年度

2023年6月5日(月)

「ライブラリーサイエンスを知る:2023春」+入試説明会 オンライン開催

2023年8月22日(火)

「ライブラリーサイエンスを知る:2023夏」+入試説明会 オンライン開催

2023年9月28日(木)

修士論文中間発表会 九州大学伊都キャンパス イースト1号館B-324講義室

ライブラリーサイエンス専攻懇親会

九州大学伊都キャンパスItri-ito(イトリー・イト)

2023年10月14日(土)

研究データ管理(RDM)支援人材育成プログラム開講

2023年11月22日(水)

入試説明会

九州大学伊都キャンパスイースト1号館B-340講義室 オンライン併用

2024年2月12日(月)

博士論文公聴会 九州大学西新プラザ多目的室

2024年2月20日(火)

修士論文公聴会 九州大学伊都キャンパス イースト1号館B-324講義室

ライブラリーサイエンス送別会



ライブラリーサイエンス専攻送別会

教員一覧

▶ 教授

石田 栄美	データ駆動イノベーション推進本部
川平 敏文	人文科学研究院(2024年1月まで)
富浦 洋一	システム情報科学研究院
藤岡 健太郎	大学文書館
三輪 宗弘	附属図書館

▶ 准教授

赤司 友徳	大学文書館
大賀 哲	法学研究院
木土 博成	比較社会文化研究院
清水 敏之	附属図書館
谷口 雄太	情報基盤研究開発センター
中川 奈津子	人文科学研究院
渡邊 由紀子	附属図書館

博士学位授与

- ▶ 小柏香穂理 データに基づく大学評価に関する実証研究—実質的な中期計画文書の特徴と意思決定過程—

修士論文題目一覧

- ▶ 佐々木 陸 読むべき最新論文の判別方法
篠原 桐 流通を考慮した研究データのメタデータ付与支援
山崎 基数 研究内容からの研究データ検索

問い合わせ先

九州大学大学院 統合新領域学府

〒819-0395 福岡市西区元岡744番地

ライブラリーサイエンス専攻

工学部等教務課統合新領域係 TEL(092)802-6501

※詳しくは、統合新領域学府のホームページをご参照ください。

<https://www.ifs.kyushu-u.ac.jp> <https://www.ifs.kyushu-u.ac.jp/lss/>

令和6年(2024)年3月発行

編集発行 九州大学大学院統合新領域学府 ライブラリーサイエンス専攻
〒819-0395 福岡市西区元岡744番地
工学部等教務課統合新領域係(イースト事務室)
TEL:092-802-6501

印刷 城島印刷株式会社