

特集「図書館情報学とAIの新展開」にあたって

清田 陽司
(株) LIFULL)

大向 一輝
(東京大学)

1. はじめに

本特集は、図書館情報学およびその関連分野（文献情報学、学術情報流通、知識工学など）においてAI研究と関連の深い取組みにフォーカスし、AI研究から図書館情報学への影響、図書館情報学からAI研究への影響の双方を俯瞰することによって、今後の両者のコラボレーションを通じた研究の発展に資することを主な目的として企画し、4本の記事で構成した。

国立国会図書館次世代システム開発研究室の青池 亨氏・川島隆徳氏には、書架分類の機械学習による自動付与、画像処理技術の資料画像への適用による各種サービス（挿絵の類似画像検索機能、挿絵への自動タグ付与、紙面の白色化、見開き位置検出など）の実装、各種データセットの公開など、国立国会図書館における最先端の取組みについてご執筆いただいた。

国立歴史民族博物館の橋本雄太氏には、歴史資料の翻刻（文字を解読して活字化し、全文検索などの再利用を可能にすること）にクラウドソーシングの手法を適用し、大量の歴史資料の解読を一気に推し進めようとしている「みんなで翻刻」プロジェクトについてご寄稿いただいた。特に、本誌35巻3号の北本朝展氏による記事[北本20]でも取り上げられたAIくずし字認識の応用がもたらした効果について、紙幅を割いてご解説いただいている。

大向によるWikidataに関する記事は、意味ネットワーク、セマンティックWeb、DBpediaなど、AI研究が対象としてきた知識ベースの発展の歴史を踏まえて、2012年に開設されたWikidataの経緯、データモデルなどの解説を行うとともに、AI研究者によるWikidataへの貢献の可能性についても論じている。

Code4Lib JAPANの江草由佳氏らによる記事は、図書館業界における実務者のICTスキルの課題を踏まえてCode4Lib JAPANコミュニティが形成された経緯を述べるとともに、研究者と実務者が密接に連携し、図書館の現場に応用可能な知識やスキルを、コミュニティオブプラクティスを通じて形成する「場」としてのカンファレンス開催の取組みを紹介している。

以下では、図書館情報学とAI研究が、互いにどのように影響を与えてきたかの歴史を俯瞰したうえで、上記の各記事の内容を踏まえて、これからの両分野のコラボレーションの可能性について論じる。

2. 図書館情報学とAI研究の歴史

2020年6月にリリースされたAIマップβ2.0のマップE「AI研究の現在」(図1)には、横軸（基礎⇔応用）、縦軸（サイバー空間⇔実世界）の2軸で表現される平面上に、キーワード群が配置されている。11個の大分類の中でも、AI研究において一つの大きな潮流を形成しているのが、知識およびその表現形態としての言語やメタデータを扱う「知識の利用と共有」、「言語メディア処理」、「Webインテリジェンス」の諸分野である。第二次AIブームといわれる1980年代に盛んに研究が行われたエキスパートシステムでは、コンピュータに与える膨大な知識をどのように記述するかが大きな課題として認識され、自然言語処理分野における大量のコーパスから知識を自動的に獲得するアプローチや、WWW(World Wide Web)やセマンティックWebに代表される多数の人々が協働して知識を整備するアプローチの発達につながっている。「機械学習」を中心に活発な研究が行われている現在でも、「AI応用」においては、知識や言語、メタデータの扱いは依然として大きな課題であり、多くの研究が進められている。

図書館情報学は、上に述べた「知識の利用と共有」、「言語メディア処理」、「Webインテリジェンス」の諸分野を研究対象に含む学問分野であるという点で、AI研究と深い関連がある。図書館情報学は、図書館学と情報科学を融合させた学問分野であり、図書というメディアに限定されず、情報そのものの生成、蓄積、利用を研究の対象としている。その源流は第二次世界大戦期の米国を発祥とするドキュメンテーション研究にあるとされている[佐藤15]。

原子爆弾の開発を目的としたマンハッタン計画を端緒とするプロジェクト型科学の発達の過程で、増え続ける膨大な論文情報を扱うための手法が必要とされた。マンハッタン計画にも深く関わったMIT副学長のVannevar Bushが1945年に著した論文[Bush 45]で提案された「memex」というデバイスの構想には、蓄積された情報が相互にリンクし合い、必要な情報を容易に見つけるようにするという、現代のWWWやセマンティックWebにもつながる考え方が示されている。コンピュータを用いた論文などの全文検索システムは、Luhnによる先駆的な研究[Luhn 58]、SaltonらのグループによるSMARTシステム開発のプロジェクト[Salton 91]な

どを経て実用化され、自然言語処理の発展とも関係が深い。図書館情報学に属する計量書誌学で生み出された Garfield による引用分析手法 [Garfield 95] などは、Google のランキングアルゴリズムのキーアイデアとなった PageRank [Page 98] などにも大きな影響を与えている。図書館における利用者と図書館司書のインタラクションに着目し、情報ニーズ明確化の研究の必要性を主張した Taylor による論文 [Taylor 68] などは、質問応答システム、対話システムの研究に多くの示唆を与えてきた。

上記に述べたように、図書館情報学および周辺分野における知識、言語、メタデータの活用に関する研究の発展が、AI 研究に大きな影響を与えてきた一方で、図書館情報学も、AI や情報科学分野の発展に影響を受けてきた。特に、1990 年代に端を発する WWW の普及が、図書館情報学サービスやデータベースサービスの利用者層を、研究者や高度専門職などの限られた層から、一般市民にまで大きく広げたことが、図書館情報学の研究課題にも大きな変化をもたらした。さらに、青池氏らによる記事で示されているように、機械学習に代表される近年の AI 研究の急速な発展は、メタデータ付与の自動化、古文書の画像解析などを通じて、図書館情報サービスの高度化にも寄与しつつある。

3. 未来の図書館の実現に向けた課題

あらゆる業界において AI の浸透が専門職の役割に大きな影響を与えている流れは、図書館業界なども例外で

はない。青池氏らによる記事が取り上げている事例は、従来から司書の中心的業務とみなされてきた目録作成などの一部を AI によって代替する試みであると捉えることができる。また、橋本氏による記事が紹介している「みんなで翻刻」は、歴史学や文学分野の熟練した研究者や学芸員にしかできないと考えられてきたくずし字の翻刻を、クラウドソーシングと AI の組合せで実現しようとする取組みであると見ることができる。

一方で、こうした流れが必ずしも専門職の地位を危機にさらすものではないことも指摘されている。橋本氏は、「歴史資料の内容を適切に把握するには、くずし字を読む能力だけでは不十分であり、現代とは異なる語彙や生活習慣、時代背景についての専門的な知識と分析能力が欠かせない (中略)。こうした知識やスキルを、現在の AI により代替することは難しい」と述べている。「特定分野の訓練を積んだ少数の専門家と、インターネットを通じて連携する多数のアマチュア、そして人間の判断を部分的に肩代わりする AI という (中略) 三つのアクタを効果的に連携させる」といった新たな枠組みの構築が、図書館情報学と AI の新たな関係を築き、図書館の未来像を実現するうえで重要な鍵となるかもしれない。

国立国会図書館館長を務めた長尾 真氏による著書「未来の図書館を作るとは」[長尾 14] は、図書館が知識インフラとしての役割を今後も果たしていくためには、「情報を集め、これを知識化し活用することによって新しい情報・知識を創出し、知識インフラに加えるという形で循環的にこのシステムを強化・拡大していく」という概

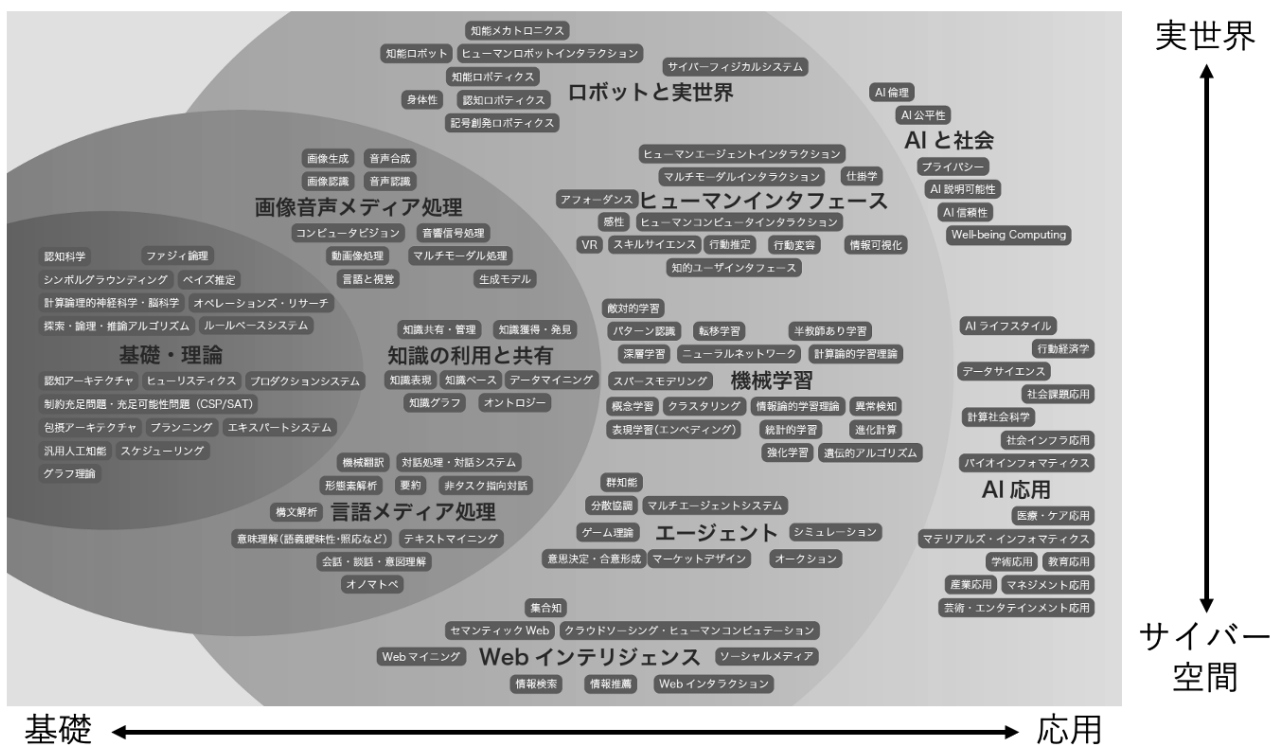


図1 AI マップβ 2.0 マップ E 「AI 研究の現在」 (©2020 人工知能学会 AI マップタスクフォース, Licensed under CC-BY 4.0, 筆者らにより追記)

念が大切であるとしている。また、図書館が扱うべき対象が、従来の本や雑誌中心から、WWW上にある音声や映像情報、さらには世界中のあらゆる情報にまで拡大している状況を踏まえ、情報を体系的(図書館学的)に扱う「情報図書館学」の必要性を主張し、以下のような内容が研究されるべきであるとしている。

- (1) マルチメディア情報
- (2) 情報の記憶
- (3) マルチメディア情報の検索
- (4) 類似性の検出と分類
- (5) メディア変換

長尾氏があげた「情報図書館学」の内容は、いずれも現在のAI・情報科学分野において盛んに研究されているものであり、青池氏らによる記事で紹介されている国立国会図書館の取組みなどは、「情報図書館学」を実現しようという試みの一つと捉えてもよいだろう。しかしながら、現時点では図書館情報学分野とAI・情報科学分野の連携はまだ弱く、手つかずの研究課題が数多く残されている。図1に照らしていえば、「画像音声メディア処理」、「機械学習」、「ヒューマンインタフェース」などの分野でも、図書館情報学との連携が期待される。今後、両者の連携がより密接に行われることが、新たなイノベーションの創出につながる可能性は大きいのではないだろうか。

4. AI研究資源整備への図書館情報学の貢献

大向の記事が言及しているWikidataは、2012年の開設以来、精力的に整備が進められ、現在では多くのAIシステムによって活用される重要なデータ資源となっている。特に、最近急速に普及したスマートスピーカなどの音声アシスタントの実用化に、Wikidataは非常に重要な役割を果たしている。Wikidataのデータモデルの設計には、図書館情報学における情報組織化の方法論(識別子、セマンティックWebなど)が生かされていると見ることもできる。また、Wikidataやその源流であるWikipediaの発展には、図書館情報学をバックグラウンドとする多くの人々が貢献してきた。

世界各国の図書館が整備し、公開した各種のデータセットも、AI研究における重要な研究資源になりつつある。日本国内でも、青池氏らによる記事で触れられているように、各種のデータセット(文字画像データセット、国立国会図書館デジタルコレクションのOCRテキストデータセット、資料レイアウトデータセットなど)が国立国会図書館によって公開され、機械学習コンペティションなどで活用が進められている。また、橋本氏による記事でも、「みんなで翻刻」によって作成された大量の翻刻文が、古典籍のレイアウト認識モデルを用いることで、くずし字認識AIの精度を向上させるための教師データとして利用できる可能性が示されている。

AI研究者が今後重視すべきなのは、大向の記事の末

尾で言及されているように、単にデータセット利用者として研究を行うだけでなく、データセット提供者側との双方向の関係づくり、コミュニティへの関与を通じて、研究資源の整備に積極的に関わることであろう。時には研究の成果をデータセット整備に還元するなどの貢献を行うことも、データセット提供者側との長期的な信頼醸成に寄与し、AI研究の将来的な発展の加速につながるだろう。

5. おわりに: AI研究コミュニティと社会の新たな関係構築に向けて

江草氏らによるCode4Lib JAPANコミュニティに関する記事は、研究者と実務者、さらには実務者の先にいる市民との連携のケーススタディとして、多くの示唆を含んでいる。AIやICTの発達が市民に多大な利便性をもたらす一方で、実務者に求められるスキルが急速に変化し、市民の期待に応えるサービスの提供が難しくなっているという課題は、図書館業界に限らず、社会のあちこちに見られる。新型コロナウイルス感染症(COVID-19)の感染拡大は、そうした課題を顕在化させている。保健所において患者発生状況把握が十分でなかったり、教育現場においてeラーニングの効果的な活用ができていなかったりしたのも、実務者のスキル向上の機会が十分に確保されていないことが背景にあるだろう。研究者と実務者が密接に連携し、コミュニティオププラクティスの場を創出することは、こうした課題の解決に寄与するとともに、研究コミュニティとしての新たな社会貢献のチャンネルをつくり、研究の成果を生かす機会を増やすことにつながるだろう。

また、江草氏らによる記事は、実務者と積極的に関わることは、本質的な研究課題を見いだすためのインスピレーションを得る上でも有益であるとしている。日本の研究コミュニティの弱点の一つとして、シーズ指向への大きな偏りが指摘されている[伊藤 15]が、実務者との積極的な関わりを通じてニーズ指向の研究課題を発見するという営みは、その弱点を補強するうえでも大いに役立つのではないだろうか。

図書館情報学と社会をめぐる最近の注目すべき動きの一つとしては、自然災害などで被災した文化施設に関する被災情報や救援情報をMediaWikiで集約し、効果的な復旧・復興支援につなげることを目指すsaveMLAKの取組みがある[saveMLAK 20a]。2011年に発生した東日本大震災で被害を受けた多数のMLAK=博物館・美術館(M)、図書館(L)、文書館(A)、公民館(K)の復旧・復興を支援することを当初の目的として発足したsaveMLAKには、被災地外を含む各地に在住する多数のボランティアが情報の集約と共有に関わり、MediaWikiに集約された情報は多くの支援者に活用された[岡本 12]。saveMLAKの活動は、2016年熊本地震や、続発した豪雨災害においても継続され、2020年にはCOVID-19

感染拡大に伴う MLAK の休館・開館状況に関する情報の集約が行われている [saveMLAK 20b]。休館・開館情報の把握には、同じくボランティアによる図書館や自治体 Web サイト、SNS などの巡回が行われている。このような各種情報源の巡回やチェックにも、機械学習や自然言語処理などの技術の活用を通じて、AI 研究が貢献できる部分は大きいであろう。

より多くの AI 研究者の方々が、社会との新たな関係構築へのヒントを見いだすうえで、本特集の内容が少しでも参考になれば幸いである。

◇ 参 考 文 献 ◇

- [Bush 45] Bush, V.: As we may think, *The Atlantic*, No. 176, pp. 101-108 (1945), <https://www.ias.ac.in/article/fulltext/reso/005/11/0094-0103> (accessed 2020-10-12)
- [Garfield 95] Garfield, E.: New international professional society signals the maturing of scientometrics and informetrics, *The Scientist*, Vol. 9, No. 16 (1995), [http://www.garfield.library.upenn.edu/commentaries/tsv09\(16\)p11y19950821.pdf](http://www.garfield.library.upenn.edu/commentaries/tsv09(16)p11y19950821.pdf) (accessed 2020-10-12)
- [伊藤 15] 伊藤貴之: イノベーションのための産学連携と基礎教育に関する一考察, *人工知能*, Vol. 30, No. 3, pp. 337-343 (2015), https://doi.org/10.11517/jjsai.30.3_337 (accessed 2020-10-12)
- [北本 20] 北本朝展: Kaggle くずし字認識—世界規模の人文系コンペ開催への挑戦—, *人工知能*, Vol. 35, No. 3, pp. 366-376 (2020), https://doi.org/10.11517/jjsai.35.3_366 (accessed 2020-10-12)
- [Luhn 58] Luhn, H. P.: The automatic creation of literature abstracts, *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159-165 (1958), <https://doi.org/10.1147/rd.22.0159> (accessed 2020-10-12)
- [長尾 14] 長尾 真 著, LRG 編: 未来の図書館を作るとは, 達人出版会 (2014), <https://tatsu-zine.com/books/miraino-toshokan> (accessed 2020-10-12)
- [岡本 12] 岡本 真: saveMLAK の活動と課題, そして図書館への支援を巡って, *情報管理*, Vol. 54, No. 12, pp. 808-818 (2012), <https://doi.org/10.1241/johokanri.54.808> (accessed 2020-10-12)
- [Page 98] Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank citation ranking: Bringing order to the web, Technical Report, Stanford InfoLab (1998), <http://ilpubs.stanford.edu:8090/422/> (accessed 2020-10-12)
- [Salton 91] Salton, G.: The smart document retrieval project, *Proc. 14th Annual International ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 356-358 (1991), <https://doi.org/10.1145/122860.122897> (accessed 2020-10-12)
- [佐藤 15] 佐藤 翔: マンハッタン計画と「電子図書館の神話」: 学術情報流通の近現代史, *Musa: 博物館学芸員課程年報*, Vol. 29, pp. 7-18 (2015), https://www.i-repository.net/il/meta_pub/G00001450TEMON_504150305 (accessed 2020-10-12)
- [saveMLAK 20a] saveMLAK: 博物館・美術館, 図書館, 文書館, 公民館の被災・救援情報 (2020), <https://savemlak.jp/> (accessed 2020-10-12)
- [saveMLAK 20b] saveMLAK プロジェクト: saveMLAK ニュースレター, 第 65 号 (2020-09-15), https://savemlak.jp/savemlak/images/c/c5/saveMLAK_Newsletter_20200915.pdf (accessed 2020-10-12)
- [Taylor 68] Taylor, R. S.: Question-negotiation and information seeking in libraries, *College and Research Libraries*, Vol. 29, No. 3, pp. 178-194 (1968), https://www.ideals.illinois.edu/bitstream/handle/2142/38236/crl_29_03_178_opt.pdf (accessed 2020-10-12)