

【予稿集】

AIで作れるでしょと言われてしまう日本の図書館目録について

木村麻衣子*

*日本女子大学 kimuram@fc.jwu.ac.jp

1. 目録データ自動作成の夢

生成 AI 時代¹の到来を待つまでもなく、面倒な目録作成作業は、いつでも図書館業務の「合理化」の対象であった^{2,3}。自動索引や自動分類には60年以上の研究の積み重ねがあり⁴、一部では実用化もされている⁵。目録の記述部分の自動化についても、1969年にタイトル・ページからの情報の抽出というアイデアが見られ⁶、1984年には目録プロセスの一部を自動化する試みが見られた⁷。

Brzustowicz は、2023年2月に ChatGPT (2023年2月13日更新版) を用いて、WorldCat に含まれない資料1件を含む6件の WorldCat 用 MARC レコードの作成を試み、手作業で作成する MARC レコードとほぼ同じ完成度であったと述べた⁸。この論文に掲載された MARC レコードの内容が、実際には多くの誤りを含むものであることが後に指摘されており、“ChatGPT は人間のように MARC や RDA のルールを理解することはできない”⁹との反論がなされた。しかし、その後も ChatGPT の機能は進化しており¹⁰、何らかの形で情報源を与えれば、ChatGPT がそれらしき目録データを作ってくれることは事実であり、目録作業に生成 AI を活用できる可能性は、確かにある。

かといって、今後は目録作業の全てを生成 AI に任せることができ、目録業務やカタログガーは不要となる、と考えるのは早計である。本発表では、目録の3つの側面を整理した上で、日本の図書館目録の現状の問題点を4つ挙げ、これらの問題が生成 AI によって解決する見込みは薄く、むしろ人間の手に残る作業はこれまでよりも高度化することを確認したい。

本稿で述べることは発表者の執筆時点での考えであり、今後の研究や技術の発展の動向により、

考えが変化しうることをお断りしておく。

2. 目録の3つの側面

発表者は、図書館目録に限らず、古典籍の目録なども含めた「目録」には少なくとも3つの側面があると考えている。

1 点目は、知識体系の表現手段としての側面である。清代の目録学者である章学誠が“辨章学术，考鏡源流”と説明しているように、古典籍の「目録」とは、それぞれの書物の形態、来歴やテキストの系統を他本と見比べながら分析し、学問の源流の中に該本を位置付ける研究の成果物である¹¹。そこでは、目録全体が、一つの知識体系を表現している¹²。ポール・オトレの「世界、知識、学術、書物」の構想図からも、類似の思想を読み取ることができる¹³。

現代の図書館目録においては、分類表そのものが知識体系の表現であることを認めたとしても、分類記号を付与された図書館目録自体が、知識体系の表現であるとは考えにくい。図書館分類は、現実には資料の排架場所を決める意味合いが強く、書誌分類であっても、多くは図書館側の様々な都合が加味された上で付与されているからである。

2 点目は、コレクションのリストとしての側面である。図書館目録は、所蔵者が独自に定める蔵書構築基準等に基づき収集した所蔵資料の一覧であり、所蔵されたものとされなかったものを同時に示しているという意味で、所蔵者の思想を反映している。

ところが、近年の図書館目録では、利用可能な電子資源の書誌データが、OPAC には収録されないケースが見られる¹⁴。さらに、所蔵していないものが図書館目録に含まれるケースも出てきた。慶應義塾大学メディアセンターでは、2016年4月

より DDA (Demand Driven Acquisition) を導入しており、OPAC 上に図書館が未購入の電子書籍の書誌データが表示される¹⁵。これらの図書館目録からは、コレクションのリストという側面が失われつつある。

3 点目は、ディスカバリー・ツールとしての側面である。FRBR より前に言われていた図書館目録の 2 つの機能 (識別機能と集中機能) も、FRBR の 4 つの利用者タスク (発見・識別・選択・入手) も、この側面の域を出るものではない。ディスカバリー・ツールとしての目録は、利用者が目的とする情報資源にたどり着くことができることを保証する。現代の図書館目録では、この側面が他の側面よりも圧倒的に重視されている。

この側面において、近年では「探索」という利用者タスクも重視されている。FRBR 以降の IFLA の標準類は、図書館の目録データを図書館外のメタデータと連携させることの重要性を強調しつつ、一方で、あえて図書館目録の高度化を図ってきた。それが「関連」の追加であり、「関連」なしに「探索」機能は果たせない。「関連」は、利用者に、ある情報資源に關係する別の情報資源、あるいは、ある著者に關係する別の個人や団体を示すものであり、利用者の当初の目的を超えた情報資源を提供する。

版本学者である姚伯岳は、FRBR 以降の西洋の目録の動きは、実用主義を捨て、「関連」を重視することで“辨章學術，考鏡源流”に似てきていると述べた¹⁶。ある情報資源を、他の情報資源との關係性の中に置いてみせるという行為は、確かに 1 つ目の側面を思わせる。尤も、FRBR や後継の IFLA LRM には目録全体が知識体系の表現である、という意識までは見られない。

FRBR 以降推進されてきた、図書館の目録データを LOD (Linked Open Data) として公開し、図書館外の世界に使ってもらおうとする動きは、データの外部提供という 4 つ目の側面をもたらしたと言えるかもしれないが、ここでは詳述しない。

3. 日本の図書館目録の現状

現在の日本の図書館目録がディスカバリー・ツールとしての役割を十分に果たしているかといえば、残念ながらそうではない。発表者は、日本の図書館目録には、3 つ目の側面に限っても、少なくとも以下 4 点の問題があると考えている。

1 点目に、日本の図書館目録は、件名標目からの網羅的検索が困難である。件名は、情報資源を主題から漏れなく「発見」するために必要であり、タイトルにも目次にも、時には全文の中にも現れない、主題を表す名辞が、主題分析によって付与される。複雑な主題も、複数の件名を一定の規則に従った順序で結合することによって表現され (事前結合)、分類記号よりも主題の特定性が高い主題目録法ということになっている (特定記入の原則)¹⁷。特定性の高い件名は、類似の主題を持つ情報資源同士の「識別」にも役立つはずである。

ところが、複数の細目を事前結合した件名標目は付与が難しい上、OPAC でうまく検索されないという難点がある。さらに、CiNii Booksなどで付与されている件名を見ると、分類が決まれば件名が決まるような特定のでない件名が数多く付与されていると言わざるを得ない。

2 点目に、日本の図書館目録の多くは、件名以外の標目のコントロールも不完全であり、このことが、「発見」のみならず「探索」機能の実現をも阻害している。少なくとも、FRBR の実体「著作」を「探索」する機能を図書館目録に追加するためには、「著作」の典拠データを新たに構築しなければならない。

3 点目に、日本の図書館目録の一部においては、一部の情報資源に対する「識別」機能の実現が危うくなっている。

FRBR の「識別」機能に求められるのは、情報資源の性質が理解できるだけのデータ (多くの場合それは標目ではなく記述部分に現れる) を利用者を与えることと、類似の情報資源との区別ができるようにすることである。

NACSIS-CAT は 2020 年より、外部機関作成書誌データの活用を念頭に、NACSIS-CAT で採用するものとは異なる目録規則で作成された書誌デー

タの存在を許容し、さらに同一の情報資源に対する複数の書誌データ（並立書誌）の存在を許容した¹⁸。同一の情報資源に対して複数の書誌データが存在するならば、それらが同じ情報資源なのか、別の情報資源なのかを識別するのは利用者ということになる。本来、ISBN等の識別子で一意に識別できない情報資源に対しては、それらを識別できるだけの詳細な記述を提供すべきであるが、今回の決定がそのような前提に立っているとは思われない。

4点目に、以上の問題点に関心のある図書館員はおそらくきわめて少数である。組織化を専門とする研究者も減少の一途を辿っている。

4. 生成AIはこれらの問題を解決できない

前章で述べた4つの問題点は、生成AI等の技術を以て解決することが難しいと考えられる。

1点目の件名標目について、米国国立医学図書館(NLM)は2021年に、MEDLINEに収録される書誌データへの医学件名標目表(MeSH)による件名標目付与の自動化を発表した¹⁹。しかし、主題が特定分野に限定されない情報資源への、特定性の高い件名標目の自動付与は、実装レベルでは実現していない。

主題分析のプロセスは情報資源ごとに異なる判断が求められ、言語化しづらく、こうすれば適切な件名が付与できる、というマニュアルは存在しない。人間が言語化できない作業を生成AIに学習させるのは困難である。生成AIはこれまでの実例をもとにそれらしい件名を付与するであろうが、その付与が適切であるかどうかは(主題による探索を重視するのであれば)人間が評価せざるをえない。また、新しい件名標目や、件名標目付与のルールは人間が策定することになる。

2点目に関して、件名標目以外の典拠コントロールとは、新しい典拠データを作成する場合でも、既存の典拠データにリンクする場合でも、2つの実体と同じ実体であるか否かの判断を要する。この判断は容易な場合もあれば難しい場合もある。難しい場合は、通常、対象情報資源に手がかりが

乏しく、様々な情報資源を調査の上で結論を出すことになるが、この作業は生成AIにも難しいはずである。生成AIが判断を下す際、明確な出典を挙げるができるのかも疑問である。

さらに、生成AI等の技術によって、FRBRの「関連」を自動付与できるかどうかは不明である。手元の情報源から簡単に見つけられる程度の関連(翻訳と原著など)であれば可能と思われるが、一見してわからない「関連」の付与は、その必要性の判断も含め、難易度が非常に高い。少なくとも、それぞれのデータベースにおいて、利用者の「探索」の助けとなりうる「関連」の種類を決定するのは人間であろう。

3点目に関して、体現形にISBN等の識別子が付与されていない場合、あるいは付与されていてもそれが識別に使用できない場合に、識別に足るだけのエレメントを選択し記述することは、生成AIには困難であろう。なぜなら、とりわけ特殊コレクションにおいて、識別に足るエレメントは情報資源ごとに異なり、マニュアル化できないからである。例えば古典籍では、記述の一部が異なるだけで全く異なる資料であることを意味する場合があり、類似の資料だからといって、同じように記述すればよいというものではない。

定型的な情報源からごく一般的なエレメントを転記するだけであれば、生成AIにも可能と思われるが、「発見」や「識別」のためにどのエレメントを取り出すべきなのかを予め決めておくのは人間である。目録規則にはエレメントの膨大なリストがあるが、その目録規則を策定するのも、それぞれの図書館やデータベースの実情に合わせて、その中から自館に必要なエレメントを選択し、適用細則を策定するのも人間である。生成AIには目録規則や適用細則の提案はできるかもしれないが、利用者ニーズをふまえた最終判断は下せまい。

4点目に関して、発表者が問題点として述べた事柄に(同意してくださるかどうかは別として)関心を寄せてくださる図書館員は、目録について一定の知識と経験をお持ちのはずであるが、こうした図書館員は、目録作業への生成AIの導入で

さらに減少する可能性がある。

5. 目録超高度化時代の到来

以上に見てきたように、目録作成作業において生成 AI を活用したとしても、難度の高い作業は人間の手に残るだろう。一方、情報源から決められたエレメントを転記するだけの作業は自動化されるため、目録担当職員はさらに減るであろう。わずかに残った職員で、高度な判断を迫られる作業を一手に引き受けなければならない。そのような職員をどのように養成していくのかは、今後の大きな課題である。目録担当になる可能性が少ないからと、司書課程から組織系の科目が削減され

てしまえば、状況はさらに悪化する。

現在の目録が既に十全に機能を果たせていないから、あるいは難しいから、人手が足りないからといって、それらの機能を切り捨ててしまえば、やがて図書館目録は「発見」も「識別」も保証できない代物に成り下がる。質の高い目録のないところでは、質の高いパブリックサービスの提供も不可能である。目録によって利用者と情報資源を結びつけるという最低限のサービスを提供できないならば、図書館は居場所提供以外のどのような価値を利用者に提供できるというのか。

注・文献

1 本稿では、OpenAI 社が「ChatGPT (GPT-3.5)」を一般向けにリリースした 2022 年 11 月 30 日以降を「生成 AI 時代」と定義する。

2 山田米一. 図書館運営の合理化. 私立大学図書館協会会報. 1952, (2), p. 3-4.

3 記述ユニット・カード方式の導入は、目録作業の「合理化」のために行われた。森耕一. 標目と記述の分離: 目録作業の合理化のために. 図書館界. 1955, 7(6), p. 195-201.

4 図書館情報学ハンドブック. 丸善, 1988. p.471 によれば、自動分類は 1960 年代初頭から研究が進められてきた。同 p.574 によれば、自動索引の基本的な概念は 1950 年代後半に確立された。

5 Golub, Koralka. Automated subject indexing: an overview. *Cataloging & Classification Quarterly*. 59(8), 2021, p. 702-719.

6 Kilgour, Frederick G. Computerization: the advent of humanization in the college library. *Library Trends*. 18(1), 1969, p. 29-36.

7 Davies, Roy; James, Brian. Towards an expert system for cataloguing: some experiments based on AACR2. *Program*. 18(4), 1984, p. 283-297.

8 Brzustowicz, Richard. From ChatGPT to ChatGPT: the implications of artificial intelligence on library cataloging. *Information Technology and Libraries*. 2023, 42(3), <https://doi.org/10.5860/ital.v42i3.16295>.

9 Amram, Tess et al. Response to “From ChatGPT to CatGPT”. *Information Technology and Libraries*. 42(4), 2023, <https://doi.org/10.5860/ital.v42i4.16983>. 同号にほか 2 通の反論が掲載されている。

10 OpenAI が 2023 年 11 月 6 日にリリースした GPTs では、ユーザが任意の文書やデータを ChatGPT

に学習させることができる (学習の程度は不明であるにしても)。Open AI. “ChatGPT release notes”. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>, (accessed 2024-05-20).

11 木村麻衣子. “和古書・漢籍”. 図書館情報学事典. 丸善, 2023, p. 378-379.

12 漢籍の四部分類に関しては、完成された分類表が存在しないため、知識体系の表現となりえたものが目録のみである、とも言えるかもしれない。

13 根本彰. アーカイブの思想. みすず書房, 2021, p. 189-190.

14 例えば、日本女子大学図書館が提供する電子図書館 JWU LibrariE の書誌データは OPAC に掲載されていない。

15 古賀理恵子, 藤本優子. DDA を導入して: 電子資源管理部門での取り組みや変化. *MediaNet*. 2018, (25), p. 5-8, <https://www2.lib.keio.ac.jp/publication/medianet/article/pdf/02500050.pdf>.

16 姚伯岳. 我国图书馆古籍编目工作存在的问题及建议. *图书情报工作*. 2020, 64(10), p. 28-34.

17 山下栄. 特殊記入の原則と限界: 中小図書館における件名標目の特殊化方式について. 仙田正雄教授古稀記念図書館資料論集. 天理: 仙田正雄教授古稀記念会, 1970, p. 211-236.

18 これからの学術システム構築検討委員会. “NA CSIS-CAT/ILL の軽量化・合理化について (最終まとめ)”. 2018-10-19. <https://repository.nii.ac.jp/records/2000919>, (参照 2024-05-20).

19 MEDLINE 2022 Initiative: transition to a automated indexing. *NLM Technical Bulletin*. 2021, (443), https://www.nlm.nih.gov/pubs/techbull/nd21/nd21_medline_2022.html, (accessed 2024-05-20).