

CA1827 XXXXXXXXXX
**ウェブスケールディスカバリと
 日本語コンテンツをめぐる諸課題
 ー海外における日本研究の支援を踏まえて**

いのかつのり
 飯野勝則*

1. はじめに

日本におけるウェブスケールディスカバリ (Web Scale Discovery、以下 WSD) は、学術情報を統合的に検索するツールとして、大学図書館において着実に普及しつつある (CA1772 参照)。しかし、検索対象となる日本語コンテンツの収録状況は、英語コンテンツに比して、未だ十分なものとは言いがたい。従って、WSD における日本語コンテンツを充実させることは、利用者ニーズに直面する国内の WSD 導入館にとって喫緊の課題であることは間違いない。一方で、WSD のセントラルインデックスを有するという特性 (CA1772 参照) を勘案するに、日本語コンテンツの充実がもたらす効果は、国外にも広く波及することが予想される。本稿ではこの状況を巡る諸課題について、関係するベンダーの方々へのヒアリングや国内外の WSD 導入館との意見交換で得られた知見をもとに、改めて考えてみたい。

2. BIG4 における日本語コンテンツの現状

2.1 日本語コンテンツの収録状況

WSD には、OCLC の WorldCat Local⁽¹⁾、ProQuest 社の Summon⁽²⁾、EBSCO 社の EBSCO Discovery Service⁽³⁾、Ex Libris 社の Primo Central⁽⁴⁾ という、通称 BIG4 とよばれる、海外ベンダーによる代表的製品が存在する⁽⁵⁾。2014 年 6 月末現在、WSD の国内導入館は総計で 80 館ほどと推定される。また日本語コンテンツに着目すると、CiNII Articles (NII-ELS)、JAIRO、国立国会図書館雑誌記事索引、J-STAGE といった公的な性格が強く、利用者の需要も旺盛なデータベースに関しては、各 WSD に既に収録を済ませているか、収録に向けた動きを見せる状況となっている。商用データベースでは、JapanKnowledge Lib、magazineplus、医中誌 Web などが同様の状況にあるが、どちらかと言えば例外的である。たとえば利用者のニーズが大きい朝日、読売、毎日といった新聞データベースは、全く収録されていない。

2.2 WSD に収録されたメタデータの言語別割合

WSD に収録されたメタデータに関し、例として佛教大学図書館の Summon を用いて、言語別の割合を

調査した。その結果、上位 5 言語のメタデータ件数と割合は以下ようになった。すなわち、(1) 英語 7 億 7,060 万件 (78%)、(2) ドイツ語 8,530 万件 (8.5%)、(3) 中国語 4,470 万件 (4.5%)、(4) 日本語 3,500 万件 (3.5%)、(5) スペイン語 1,300 万件 (1.3%)、である (2014 年 6 月 20 日現在)。英語のメタデータ件数が群を抜いている一方、他の言語のメタデータ件数は相対的に少ない。日本語は第 4 番目に位置するものの、その割合は全体の 3.5% に過ぎない状況である。

2.3 WSD への収録が進まない要因

筆者が複数のコンテンツベンダーと各 WSD ベンダーにヒアリングしたところによると、コンテンツベンダーの視点からは、WSD へのコンテンツ提供について、懸念や困惑を抱かせるような各種の阻害要因が存在する。これらの要因を類型化すると以下のようになる。

(1) 心理的要因

海外に拠点を置く WSD ベンダーとのやり取りに拒否感を持ったり、WSD に検索をさせることで、その下風に立つことになりはしないかといった不安感が該当する。

(2) ライセンス的要因

海外に拠点を置く WSD ベンダーと契約書を交わす段階での管轄裁判所の設定や、契約を解除した場合のメタデータの処分方法など、法理面での懸念が該当する。

(3) 技術的要因

運営するデータベースが、WSD との間で、データを効率的に連携するための仕組みを持っていなかったり、コンテンツの一次情報の URL がパーマネントでないといった点で、連携が技術的に困難である場合がある。

(4) ビジネス的要因

WSD に二次情報データベースのコンテンツを提供した場合、運営するデータベース本体へのアクセスが減少することが予想されるが、それを理由に本体のデータベースが解約されるのではないかと、この懸念を持つ場合がある。また、データベースの価格体系を同時アクセス数によって決定している場合、WSD に提供したコンテンツへのアクセス状況を制御することができないことから、ビジネスモデルの構築に行き詰まる場合がある。

これらの要因はコンテンツベンダーが WSD が何たるかを認知して、初めて生じる問題でもある。しかし、今回のヒアリング時には、WSD ベンダーから「コンテンツベンダーにおける WSD の認知度が低く、折衝しにくい場合がある」との声も寄せられており、現実

* 佛教大学図書館

には、必ずしも全てのコンテンツベンダーがWSDを認知しているわけではない。

3. 図書館がとるべきアクション

3.1 コンテンツベンダーへのアプローチ

前章の状況を踏まえると、WSDにおける日本語コンテンツの収録が進展するためには、(A) コンテンツベンダーにおいてWSDの認知が深まる必要があるであり、その上で、(B) WSDへのコンテンツ提供に関する阻害要因の克服が必要となると考えられる。

現状、WSDへの日本語データベースの収録は、初めに図書館からWSDベンダーに要望が出され、その要望をうけたWSDベンダーがコンテンツベンダーと折衝するという流れが一般的である。しかしコンテンツベンダーにとって、WSDベンダーは、直接の顧客ではなく、影響力には限界がある。それゆえ、図書館が問題解決に向けて、積極的に関与していただくことが望ましいと考える。たとえば(A)については、図書館がコンテンツベンダーに対し、直接WSDが何たるかを伝え、積極的に認知度を上げていくといった努力が求められるだろう。一方、(B)については、コンテンツベンダーの懸念を払拭できるような提案やアドバイスを行っていくことが必要となる。とくに阻害要因のうち、コンテンツベンダーの収益確保に絡む、「ビジネス的要因」に類型化される事象については、顧客の立場から、WSDに適したビジネスモデルを提案するなど積極的な関与を行うことが有効であろう。

3.2 ビジネスモデルの提案

実のところ、ビジネスモデルの提案といっても、それほど難しく考える必要はない。国内外を問わず、WSDにおけるコンテンツベンダーのビジネスモデルの事例は増え続けている。図書館は顧客として、コンテンツベンダーに対し、適切な事例紹介を行うことで、十分にその役割を果たすことができる。

たとえば2.3(4)で言及した「データベースの解約の恐れ」に対しては、「本体解約後には、WSDでのコンテンツ利用は一切できなくなる」ような契約を締結すれば、本体のアクセスが減っても、契約解除にはつながらないとアドバイスできる。そのほか、同時アクセス数による価格体系で運用するデータベースについては、「WSDでの使用料」を見込んだ価格体系を作ってはどうかという提案を行うこともできる。

また、WSDへのコンテンツ提供が、ビジネスの拡大につながる要素を持っていることもアピールする必要があるだろう。とくに一次情報データベースであれば、WSDを通じたアクセスの増加により、図書館側で、同時アクセス数を増やそうという動きが出てくる可能

性もある。更に、WSDのセントラルインデックスを通して、日本のみならず、諸外国の図書館でコンテンツが検索され、目にさらされるようになることで、データベースの知名度が高まることも期待できる。これは海外からの新規の問い合わせや、契約に結び付く要素となるだろう。

今後、図書館はWSDにおけるビジネスモデル情報のハブとして、コンテンツベンダー、WSDベンダー、各種代理店の橋渡しを行い、WSDにおける日本語コンテンツの収録促進に寄与していくことが望ましいと考える。

4. WSDにみえる海外日本研究の危機

WSDにおける日本語コンテンツの貧弱さは、海外、とくにWSDが普及した米国において危機的状況を招いている。例として、以下に米国ミシガン大学図書館のSummonで2014年5月30日に、『枕草子』を検索した場合の上位7レコードの内容を示す。なお、この時点でのヒットレコードは144件であった。

表示順	言語	コンテンツタイプ	作者	レコードタイトル	図書/雑誌/叢書名等
1	中国語	雑誌論文	清少納言 [日]	枕草子	視野
2	中国語	雑誌論文	清少納言 [日]	枕草子	文苑
3	中国語	雑誌論文	宋茜茜	《枕草子》与《浮生六记》中“趣”之比较	华北水利水电学院学报：社会科学版
4	日本語	電子ブック		枕草子	新編日本古典文学全集
5	日本語	参考文献		枕草子	日本大百科全書
6	中国語	雑誌論文	涂云帆	又宁静又美好——《枕草子》读书小记	小溪流：成长校园
7	中国語	雑誌論文	姚继中 林茜茜	日本文学理念(五)「枕草子」之和雅——「をかし」	日语知识

図1 『枕草子』の上位7レコード (協力：ミシガン大学図書館 横田カーター啓子氏、2014-05-30)

WSDの収録コンテンツは膨大であることから、上位に適切な検索結果が表示されることは重要である。ところが、ここでは日本文学を代表する『枕草子』の検索結果であるにも関わらず、日本語レコードは2件しか含まれていない。また上位3レコードはすべて中国語のコンテンツとなっている。このような状況が生じた理由は以下のようなものである。

(1) 中国語データベースは有償のため、図書館において「購読」の意識があり、WSDの検索対象に設定する作業が行われていたが、国立国会図書館雑誌記事索引に代表される、日本語のオープンアクセス形式のデータベースには、図書館側に「購読」の意識がなく、WSDの検索に設定する作業が全く行われていなかった。

(2) 日本における WSD においては、おおむね日本語コンテンツが優先的に表示される設定になっているが、海外においては、純粋にメタデータの内容のみで「関連度」が判断されるため、貧弱なメタデータは下位に埋もれがちになる。中国語コンテンツの場合、日本語コンテンツに比べ、抄録や本文テキストデータを含む、エンリッチメントされたメタデータになっていることが多く、「関連度」の分析により、上位に表示される傾向が強い⁽⁶⁾。

実のところ、(1)については、WSD の検索の対象に日本語コンテンツを設定するという、必要最低限の対応に過ぎない。このため、その効果は限定的である。図 2 に、この設定作業を行った後の 2014 年 6 月 26 日の検索における、上位 7 レコードを示す。なおヒットレコード件数は 1,723 件である。

表示順	言語	コンテンツタイプ	作者	レコードタイトル	図書/雑誌/叢書名等
1	中国語	図書	清少納言著 周作人译	枕草子	苦雨斋译丛
2	中国語	雑誌論文	清少納言 [日]	枕草子	視野
3	中国語	雑誌論文	清少納言 [日]	枕草子	文苑
4	中国語	雑誌論文	宋茜茜	《枕草子》与《浮生六记》中“趣”之比较	华北水利水电学院学报：社会科学版
5	日本語	図書	清少納言作、バリエイ・アートワークス漫画	枕草子	まんがで読破
6	日本語	図書	坂本和子 朗読、田辺聖子 解説	枕草子	新潮 CD
7	日本語	電子ブック		枕草子	新編日本古典文学全集

図 2 『枕草子』の上位 7 レコード (協力：ミシガン大学図書館 横田カーター啓子氏、2014-06-26)

検索結果のトップレコードに変化が見られ、また日本語のレコードが増えたものの、依然として上位の 3 レコードは中国語のままである。

米国の大学は、日本研究の拠点として長く知日派を排出する揺籃となってきたと筆者は考える。しかし、米国の WSD において、日本語での検索にも関わらず、検索結果に日本の学術情報が出現しなかったり⁽⁷⁾、他のアジア諸国のコンテンツに埋もれるという現状が続けば、日本研究の将来は相当危ういものになるだろう。とくに日本研究の中心となる、日本語の人文社会系の学術誌の電子化が、遅々として進んでいない現状を踏まえると⁽⁸⁾、その事態は非常に深刻である。

5. おわりに

WSD に日本発のコンテンツの収録を推し進めることは、日本国内のエンドユーザーに対するサービスの向上にも直結する、重要な課題であるのみならず、世界的な学術情報流通の枠組みのなかでも大きな意味を持

つ。本稿では、WSD における日本語コンテンツの収録を加速するために、図書館がとるべきアクションについて述べたが、今後は人文社会系日本語コンテンツの電子化促進やメタデータのエンリッチメントをどう行うのかという課題も検討する必要がある。また本稿では言及していないが、グローバルな視点に立てば、理工系の英語のコンテンツも、日本の技術立国としての地位を担保するために収録の促進が求められる⁽⁹⁾。日本の政府・学術関係者には、WSD が海外の研究者や学生に与える広報効果や影響力を正確に認識して、WSD に向き合う姿勢が求められるのではないだろうか。

- (1) “WorldCat Local”. OCLC.
<http://www.oclc.org/en-asiapacific/worldcat-local.html>, (accessed 2014-06-25).
- (2) “The Summon Service”. ProQuest.
<http://www.proquest.com/products-services/The-Summon-Service.html>, (accessed 2014-06-25).
- (3) “EBSCO Discovery Service”. EBSCO Publishing.
<http://www.ebscohost.com/discovery/eds-about/>, (accessed 2014-06-25).
- (4) “Primo Central Index”. Ex Libris.
<http://www.exlibrisgroup.com/category/PrimoCentral/>, (accessed 2014-06-25).
- (5) 2014 年 6 月には、米国情報標準化機構 (NISO) により、Open Discovery Initiative (ODI) ワーキンググループの最終報告に基づく、WSD の仕様に関する推奨指針も公開された。
- (6) たとえば Summon においては、ユーザーインターフェースで利用する言語のコンテンツが優先的に表示される設定となっている。すなわち日本語の検索画面を利用すれば、日本語が上位に表示されるが、英語の検索画面を利用すると、日本語も中国語も優先されない、ニュートラルな状況となる。また OPAC など、ローカルコンテンツについては、上位に表示される設定となっている。なお「関連度」の判定において、抄録はタイトルに次いで 2 番目に重要な項目とされており、全文を渡ぐとされている。
- (7) なお日本漢字と中国簡体字といった言語を超えた異体字の統合検索も、WorldCat Local を除き十分には対応できていないようである (例：「検索」と「检索」)
- (8) 佐藤藤一ほか. J-STAGE 新システムが加速する国内学術論文誌の電子化と流通. 情報管理. 2012, 55 (2), p. 106-114.
https://www.jstage.jst.go.jp/article/johokanri/55/2/55_2_106/_article/-char/ja/, (参照 2014-06-25).
- (9) 飯野勝則. 佛教大学図書館におけるウェブスケールディスカバリー Summon の導入効果と課題. 情報管理. 2014, 57 (2), p. 99-108.
https://www.jstage.jst.go.jp/article/johokanri/57/2/57_99/_article/-char/ja/, (参照 2014-06-25).

[受理：2014-08-18]

Iino Katsunori.

Problems Concerning Japanese Database Contents in Web-Scale Discovery Services - From the Viewpoint of the Necessity of Increasing and Enriching Metadata in Japanese.