

1

カーリル

—図書館のオープンデータ化を促す仕組み—



応
専

吉本龍司((株)カーリル)

カーリルとは

カーリル^{☆1}は、6,500以上の公立図書館や大学図書館の所蔵情報を横断的に検索できる無料のWebサービスである。一見同様のサービスは存在するが、この後詳述するように、同種のサービスは世界的に見ても唯一である。カーリルは、図書館がWeb-OPAC（Web上に公開された蔵書目録）で提供する書籍の所蔵情報をリアルタイムに統合することで利便性の高い検索サービスを実現した。2010年3月にサービスを開始し、2014年で5年目を迎えた。公立図書館の利用者を中心に月間40万人以上に利用されている。

利用者は、近くの図書館や通っている大学など複数の図書館の蔵書をまとめて検索することができる。また蔵書がなければオンライン書店の在庫情報などを案内する。さまざまなWebサービスと相互連携しており、読書管理向けのWebサービスと組み合わせ利用することもできる。

■カーリルのコンセプト

すべての図書館に対応

サービス開始にあたっては、全国の公立図書館を網羅することを目標とした。サービスの立ち上げ時、図書館のWeb-OPAC導入率はすでに80%を超えており、これらのすべてに対応することを目標とした。普段使っている図書館や近くの図書館がカーリルに対応していないという状況では利用は広がらない。Webサービスとしては、全国規模ですべての図書館に対応しているといえるだけの規模が必要である。

☆1 <http://calil.jp/>



図-1 カーリルのトップページ

公立図書館だけでも、居住地、通勤場所、通学場所や都道県立図書館など、いくつかの図書館を組み合わせる場合も多い。

そのため、利用できる図書館を選択し、まとめて検索できる必要があると考えた。

楽しく分かりやすいデザイン

図書館へ行った経験が少ない利用者にも先入観なくサービスを利用してもらえるよう、あえて「図書館らしさ」を感じさせない、明るく楽しいデザインを目指した(図-1)。システムの制約上、検索に時間がかかる場合があるため、アニメーションを活用することで利用者を飽きさせない工夫をした。

また、各図書館がWeb-OPAC上で提供する所蔵状況の表現は統一されていないため、利用者にとっては分かりにくい。たとえば本が図書館の書棚にある状態を表す表現は、貸出可、在架、配架、排架、在庫など多様である。これらの表現を統一し、色分けなどにより分かりやすい情報提供ができる仕組みを目指した。

図書館にない本も検索

従来のWeb-OPACは、その図書館に所蔵のある本しか検索できない。利用者が図書館にない本を検

索した場合、キーワードが正しいかどうかも分からず、何度もキーワードを入力しなおさなければならず、目的の資料にたどり着くこともできない。

また、現場で働く図書館職員にヒアリングを実施したところ、多くの職員が利用者から自館にない書籍に関する問合せを受けたときに、Amazonなどのオンライン書店で検索していた。

これらの問題は、図書館にない書籍も同時に検索できれば解消できる。利用者は発見した本が図書館に所蔵されていなくても、図書館にリクエストしたり、書店で購入したりすることができる。カーリルではAmazonがAPIにより提供している書誌情報と図書館の所蔵情報を組み合わせることでこの問題を解消できると考えた。

Webと図書館をつなぐ

図書館の所蔵情報はWeb-OPACによりWeb上に公開されていたが、パーマリンクが設定されていないことが多かった。結果としてGoogleなどの検索エンジンからは検索することができない。

カーリルでは、書籍の各個別ページにパーマリンクを設定することで、検索エンジンから全国の図書館につながるハブとして機能することを目指した。

■ 開発の背景

筆者は、ある地方自治体のシステムコンサルティングの一貫として図書館システムの検討にかかわった経験がある。その自治体で導入されていた蔵書検索や予約などのWebサービスは、多額の費用をかけているにもかかわらず、ほとんど利用されていなかった。

システムの使い勝手が悪いというだけでなく、図書館職員のWebサービスに対する知識が少なく、積極的に利用者にアピールされていないなど、運用上の課題も多かった。このような経験を交えながら、開発メンバとブレインストーミングによって話し合う中で「楽しい図書館蔵書検索」というアイデアが生まれた。

当時、メンバの多くは図書館を頻繁に利用しておらず、どうしたら自分が図書館を活用できるかとい

う視点でシステムの設計やデザインを進めた。

■ サービス開始と反響

サービス開始から数日で想定を超える反響があった。新しいWebサービスを常に注目して追っているプログラマーやデザイナーを中心とした、いわゆるアーリーアダプタ層に図書館の利用者が多かったことも考えられるが、実際には図書館を普段使っていないという利用者からの反響が多かった。なかには、カーリルで検索してみて、図書館に新しい本があることを初めて知ったという声もあった。

Webサービスが図書館利用のきっかけとなっていたのである。Web上のサービスとして完結するのではなく、利用者が図書館とWebを循環することを実感した。

また、当初対応していなかった大学の図書館についても要望が多く、現在ではこれらの図書館にも広く対応している。

■ 運営体制とビジネスモデル

カーリルを運営する(株)カーリルは、岐阜県に本社を置く従業員6名のベンチャー企業である。利用者がカーリルを通じて書籍を購入した際に得られるオンライン書店からのアフィリエイトや広告収入により運営されている。オープンデータを活用したビジネスモデルとしても注目されている。

カーリルの技術

■ カーリルの仕組み

カーリルは、Amazonや国立国会図書館などが提供する書誌データと、全国の図書館が分散して管理する所蔵データを組み合わせることでサービスを実現している。書誌データとは、資料のタイトルや著者などの情報であり、所蔵データとは資料がどこにどのような状態であるかという情報である。書誌データと所蔵データはISBNによって紐付けている。

従来、公立図書館も含めた図書館の所蔵データは統合的なデータ流通の仕組みが存在していなかつ

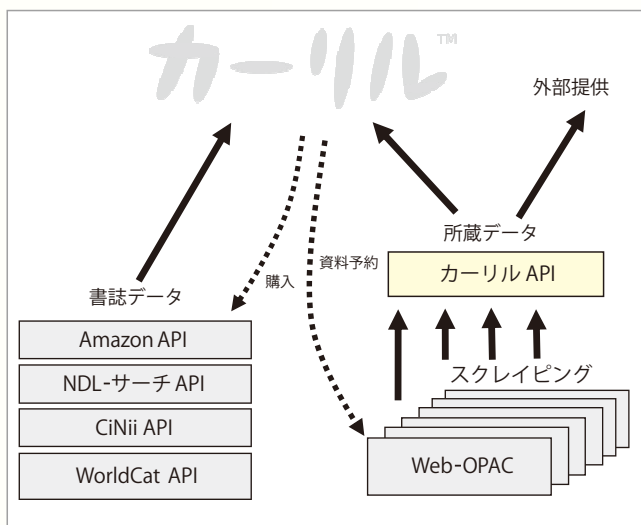


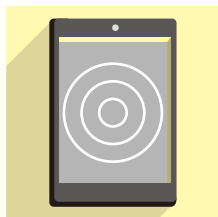
図-2 データの流れ

た。そのため、カーリルでは Web-OPAC の出力する HTML を図書館ごとに解析し、統一的なデータ形式に変換する、一般にスクレイピングと言われる手法を用いている (図-2)。

■ スクレイピング

カーリルのスクレイピングエンジンはプログラミング言語の Python によって実装されている。標準的なライブラリのみで構成することにより、サーバ環境に依存せずさまざまな環境に展開できるよう設計した。スクレイピング処理は Web-OPAC それぞれにあわせて実装しなければならないが、図書館システムは多数存在し、独自構築している図書館も多い。そのため多様なスクレイピング処理を効率的に実装することが求められた。

カーリルでは、Web-OPAC のスクレイピングに



最適化したフレームワークを開発することで、少ないコードで柔軟に開発できるようになった。このフレームワークでは、図書館システムを Python のクラスとして抽象化する。図書館のパッケージシステムを、バージョンごとに中間クラスとして定義し、共通の処理はそれを継承することで、プログラムの記述はカスタマイズ部分のみに整理されるようになっている。また、一般に行われるカスタマイズを自動的に検出し、追従するようにすることで、システム更新に伴う対応コストを最小化した。

■ 負荷制御と分散処理

カーリルでは、検索タスクを複数のサーバに分散して処理することにより高速化とスケールアウトを実現した。一方で Web-OPAC の中には同時アクセスに対して著しく性能が低いものや、同時接続数が増えるとデータベースエラーが発生するものも多い。このため、システムごとに負荷を適切に制御することによりトラブルを防いでいる。負荷制御は分散メモリ「Memcached」を利用して、複数のサーバで分散して処理している状況であっても特定の Web-OPAC に対して過負荷とならないように制御するものである。

サービス開始当初は、一部の図書館システムにおいて、図書館システムの停止などのトラブルがあった。これらの問題はカーリル側の負荷制御と開発会社の不具合修正により解決している。

同時に、カーリルでの利用者数が多く、また常識的な Web システムの応答速度と比べて、著しく性能の低いシステムについては該当の図書館や開発元に性能の向上を要請するなど、利用者の満足度向上に向けた取り組みを実施している。これらの性能の低いシステムは、データベースのインデックス設定の不足など、導入時の作業漏れなどによるものが多かった。

■ 運用管理の効率化

一般的に図書館システムは 5 年程度で更新される上、随時のカスタマイズや新しい図書館のオープン

などに伴い仕様が変更となる。これらに追随するための維持コストが莫大なものとなるため、従来、スクレイピングは大規模なシステムの構築には向いていないとされてきた。

カーリルではパターン化されたカスタマイズを自動検出する技術に加え、HTTP 層のエラーやスクレイピングの解析エラーを統計的に分析することにより、システムの変更を早期に検出することができるようになった。これら品質管理システム (図-3) の整備により、Web-OPAC の仕様が変更された場合も、短期間で対応できるようになった。

また、国立国会図書館の ISIL (図書館および関連組織のための国際標準識別子) や国立情報学研究所の NACSIS-CAT/ILL 参加組織情報などのオープンデータと図書館の情報を紐付けて、変化を監視することで図書館の名称変更や移転などにも効率的に対応している。

■ API の提供

カーリルは、自らの Web サービスを運営するために運用している API を開発者にも広く公開している。API によって、開発者は、スクレイピングエンジンを独自に開発しなくても図書館の情報を活用した新しいサービスやアプリケーションが開発できるようになった。現在、次に挙げる 2 つの API を提供しており、JSON または XML で情報を取得できる。

- (1) 図書館基本データの取得…図書館の基本的な情報と図書館の ID を返す。
- (2) 所蔵データの取得…ISBN と図書館 ID を指定することにより蔵書の有無と貸出状態を返す。

カーリルの図書館 API を活用したアプリケーションは 500 種類を超えており、スマートフォン向けのアプリケーションやブラウザ向けの拡張機能、英語の多読の支援サービスなど多岐に渡っている。

Web サービスとしてのカーリルも、API のデータを活用した実装の 1 つといえる。

個人の開発者であっても、API を活用することで新しい図書館のサービスを開発できるようになった。また、ブログパーツなどにより開発の知識を持たな

カーリル library-base				
最近発見された所蔵 (10 条目)				
ステータス	システムID	LIBID	正式名称	LIBKEY
巡回館(地域)	Hokkaido_Bibw	113201	美瑛市立図書館	巡回館
会図書館・公民館	Wakayama_Mhatew	113200	みなべ町立図書館津川分室	津川分室
会図書館・公民館	Wakayama_Mhatew	113199	みなべ町立図書館高城分室	高城分室
会図書館・公民館	Aomori_Hiroaki	113198	私前市立私前図書館駅前分室「こども読本の森」	こども読本
巡回館・BM	Hyogo_Nagawa	113196	播磨川町移動図書館「イナくるっと号」	移動図書館

最近更新されたシステム			
更新日	システムID	システム名	パッケージ
2014年2月3日	Fuku_Pref	福岡県立図書館	iLibera
2014年2月3日	Osaka_Kaisei	総務省海津市	LICS-Rte
2014年2月3日	Fukuoka_Pref	福岡県立図書館	iLibera
2014年2月3日	Choshi_Nihama	茨城県新治市	NAJIS V3
2014年2月3日	Hyogo_Nagawa	兵庫県播磨川町	iLibera21/We V2

図-3 カーリルの管理画面

い場合でも簡単に API を利用できる仕組みの提供も開始した。

国立国会図書館が運営する「国立国会図書館サーチ」や、(株) 図書館流通センターが運営するオンライン書店「TRC ブックポータル」など公共機関や企業による活用も広がっている。

現在 API の提供するデータの約 30% 程度が外部のアプリケーション向けに提供されているが、API の提供を通して図書館の所蔵データの流通性を高めることで図書館の新しいニーズが創出されると期待している。

新しい図書館システム

■ つながる図書館システム

カーリルは図書館の外から図書館を便利にするサービスであるが、図書館もまた積極的に外部の Web サービスと連携することで、利用者の利便性向上を図っている。

本や雑誌・論文の情報を Web サービス間でリンクする方法として OpenURL という標準プロトコルが提唱されている。カーリルは OpenURL に準拠しているため、Web-OPAC が OpenURL に対応していれば、容易に連携することができる。この機能を活用して、Web-OPAC からカーリルへリンクすることによって連携する図書館も多い。図書館は、Web-OPAC を起点として別の図書館の所蔵を容易に案内できるようになった。

また、図書館員自身が JavaScript により後付けで



図-4 野田市立図書館の Web-OPAC

Web-OPAC の機能を拡張する「OPAC+ (オパックプラス)」という手法も試行されている。千葉県野田市立図書館 (図-4) ではこの手法により、カーリルはもとより、さまざまな Web サービスとの連携を実現した。

■ 図書館のオープンデータ

カーリルは、図書館の所蔵情報をオープンデータと捉え、APIとしてこれらの情報の共通形式を定義することで、情報の流通性を高める取り組みを推進してきた。図書館においては、国立国会図書館の統合検索サービス「国立国会図書館サーチ」や国立情報学研究所の論文・書籍検索サービス「CiNii」においても API が提供されるなど、APIによるオープンデータの提供が活発化している。最近では、大阪府立図書館がデジタルアーカイブに関する API を提供するなど、その動きは公立図書館にも広がっている。

■ 新しいサービスの実現

オープンデータを組み合わせることで、新しい図書館サービスを創出することができる。カーリルが取り組むカーリルタッチ (図-5) もその1つである。カーリルタッチでは、図書館の書棚のラベルに内蔵された IC タグの情報を利用者のスマートフォンで読み取ることにより、Web と本棚を行き来しながら情報探索をすることができる。テーマにあわせてカーリルの API による図書館の所蔵情報はもとより、



図-5 カーリルタッチの動作イメージ

関連する論文やデジタルアーカイブ、図書館に寄せられたレファレンス事例などさまざまな情報源を横断的に表示することができる。

このような新しいサービスは、APIを活用し、従来の図書館システムと疎結合できるようになったことで実現しやすくなった。

■ リンクお断り

電子書籍やデジタル化資料の普及が始まっており、図書館における Web サービスの重要性は増している。新しい技術や利用者のニーズを敏感に捉え、自ら変化する図書館が増えている一方で、「リンクお断り」に代表される昔ながらの Web 運営を続けている図書館も多い。

こうした図書館でも導入しているシステムは耐用年数が経てば更新されるため大差はない。しかし、結果として Web サービスが積極的に運用されないため、運用レベルでの格差が大幅に増大している。図書館が公共サービスとして提供される限り、多くの利用者は住んでいる場所や通っている学校に制約され図書館を選ぶことができない。決して諦めることなくこうした図書館との対話を続け、ボトムアップを図っていききたい。

■ 図書館のビッグデータを活かす

これまで図書館のデータは、その図書館の資料を管理するためにのみ使われていた。カーリルの開発したスクレイピング技術により、これらの情報を集約できるようになり、ビッグデータと言える規模と

1. カーリル図書館のオープンデータ化を促す仕組み



図-6 図書館データの解析

なった。ここ数年でビッグデータを対象としたデータ解析技術は急激に一般化した。

このビッグデータは、たとえば出版社にとってはマーケティングのための重要なヒントとなる。品切れのため書店に並ばなくなった本であっても、全国の図書館の貸出率からニーズを推定したり、同じテーマの本の売上予測に活かしたりできる。このように図書館のデータを出版に活かす実験を出版業界とも連携しながら進めている（図-6）。

また、このビッグデータは図書館の運営に活か

せる可能性がある。全国の図書館の所蔵率や貸出率、資料の場所（展示コーナーや書庫など）を統計的に分析することで、本の購入や書庫への移動を最適化することができるだろう。全国の図書館の情報を集約することで、利用者の個人情報を除いたオープンデータだけでもできることはたくさんありそうだ。

図書館の持つオープンデータ、ビッグデータを利用者、開発者、図書館、出版などさまざまな立場の人々に流通しやすくする仕組みをつくることで、「図書館をもっと楽しく」というミッションを実現していきたい。

(2014年2月4日受付)

■ 吉本龍司 ryuuji@calil.jp

企業や行政のシステム構築や、Webサービスの立ち上げにかかわる。2010年、カーリルの立ち上げでスクレイピングエンジンの開発を担当。2012年、(株)カーリル代表取締役就任。