

はじめに

楽しい情報検索の世界へようこそ！

本書は Apache Solr (アパッチ ソーラー；以下 Solr と記します) を体系的に解説した、日本語による入門書です。Solr は Apache Lucene (アパッチ ルシーン；以下 Lucene と記します) を使って構築されたオープンソースの検索エンジン・サーバです。Lucene はすでに世界中のシステムで使われている実績を持っていますが、Java で書かれた検索エンジン・ライブラリであり、利用するには Java のプログラムをたくさん書かなければいけません。それに対して Solr は、Lucene をラップして HTTP/XML という RESTful な今どきのインタフェースでアクセスできるようになっています。もちろんインタフェースを簡単にしただけでなく、機能や性能を向上させるいろいろな仕掛けもプラスされています。自動車为例えるなら、Lucene は高信頼&高性能のエンジンであり、Solr はそのエンジンを使って組み立てられカーナビまで備えた完成車といえるでしょう。そのため Solr は、2006 年に公開されるやいなや急速に市場に受け入れられ、利用者が急拡大しました。

前著『Apache Lucene 入門』(技術評論社) 出版から約 4 年が経ちました。その間、インターネットやイントラネットで生産される情報は増え続け、情報活用のための検索機能は当時よりも一層必要とされている実感があります。今やソフトウェア技術者にとって、検索エンジンの知識はミドルウェアやデータベースに次いで必要とされる技術要素になっているといえるかもしれません。

私はこれまでさまざまな情報システムへの検索機能導入に携わる経験をし、今度はそこで出会った仲間と本書を著す機会を得ました。執筆者一同は日常的に Solr に触れ、この素晴らしい検索エンジンをもっと多くの人に知ってもらいたいとうずうずしていましたが、今ようやくそのときが訪れました(なかなか出せなかったのは、Solr 1.4 のリリースを待っていたからであります!)。本書によって一人でも多くの方が Solr を理解し活用して、検索機能を付加したりアプリケーションをより魅力的に変身させられたら執筆者一同のこの上ない喜びです。

それではまた、本文でお会いしましょう！

2010年1月
関口 宏司

●免責

本書に記載された内容は、情報の提供だけを目的としています。したがって、本書を用いた運用は、必ずお客様自身の責任と判断によって行ってください。これらの情報の運用の結果について、技術評論社および著者はいかなる責任も負いません。

本書記載の情報は、2010年2月現在のものを掲載していますので、ご利用時には、変更されている場合もあります。

また、ソフトウェアに関する記述は、特に断りのないかぎり、2010年2月現在のバージョンをもとにしています。ソフトウェアはバージョンアップされる場合があり、本書での説明とは機能内容や画面図などが異なってしまうこともあり得ます。本書ご購入の前に、必ずバージョン番号をご確認ください。

以上の注意事項をご承諾いただいた上で、本書をご利用願います。これらの注意事項をお読みいただくに、お問い合わせいただいても、技術評論社および著者は対処しかねます。あらかじめ、ご承知おきください。

●商標、登録商標について

・本書に登場する製品名などは、一般に各社の登録商標または商標です。なお、本文中に™、®などのマークは特に記載していません。

謝辞

まず Solr の産みの親である Yonik Seeley さんに感謝いたします。Yonik さんはその類い希なるプログラミングの才を存分に発揮して、高性能検索エンジン Lucene を、Web 時代にマッチした軽快なインタフェースを通じて誰でも簡単に利用できるように仕立て上げました。そして CNET 社にも感謝いたします。Solr は CNET 社で開発されましたが、CNET 社が Apache にソースコードを寄贈してコードが公開されたことでユーザが一気に広がり、今日の Solr コミュニティの基礎を築きました。また関口個人としては、担当編集者の池本公平さんと Lucene/Solr コミュニティに感謝いたします。特に池本さんは本書が出版できるよう、技術評論社を説得してくれました。そして Lucene/Solr コミュニティは、私が(株)ロンウイットを創業するきっかけとなり、そこそこの仕事の自由と世界中のエンジニアとコラボレートしながらソフトウェアを構築していくというエキサイティングな日常をプレゼントしてくれました。慢性的な寝不足というおまけを添えて。みんな、ありがとう！

対象読者

本書は、情報検索に興味を持つ、あらゆる人を対象としています。学生から社会人、週末プログラマから職業プログラマ、SE、プログラミングをしない情報システムを使うだけの人、検索エンジンを比較検討しようとしている人……制限はありません。それぞれの立場で Solr をお楽しみください。しかしながら本書は Solr の技術解説書でもあり、Solr をインストールしたりサンプルコードを設定したりする場面もあります。そのときは技術的なバックグラウンドのある方は若干有利でしょう。自分の経験に感謝しつつ、コマンドを入力したり、検索して結果を確認したり、より深くお楽しみください。

本書の構成と読み方

本書は全 10 章から構成されますが、すべてに目を通す必要はありません。第 1 章の前半では検索エンジンの基本を紹介しています。すでに検索エンジンについての知識があれば、読まなくても構いません。次に Solr のインストール方法を説明しています。本書は Solr を使いながら説明しているところが多いので、ここでぜひ手元の PC に Solr を準備しておくことをお勧めします。最後にアーキテクチャを解説していますので、ここもぜひ目を通しておくといいでしょう。

第 2 章から第 4 章は Solr の基本知識である「スキーマ定義」「インデックス作成」および「検索」について体系的に説明しています。ぜひ一通り読んでおくことをお勧めします。

第 5 章はプログラマのための章です。Solr に HTTP で検索リクエストを送ると、検索結果は XML で返ってきます。そのため、XML を HTML などに変換するフロントエンドが必要ですが第 5 章はさまざまなプログラミング言語でフロントエンドをプログラミングする方法を紹介しています。

第 6 章以降はさらに Solr を活用したい人たち向けのパートです。第 6 章ではリレーショナルデータベースなどの情報リソースから Solr のインデックスにデータを取り込むためのしくみ「データインポートハンドラ」の使い方を紹介しています。第 7 章では Solr コアを複数持つためのしくみ「マルチコア」の使い方を紹介しています。第 8 章は巨大なインデックスを分割して検索するしくみである「分散検索」の設定方法や使い方を紹介しています。第 9 章は Solr の検索機能を応用してレコメンデーションやスペルチェックなどの楽しい仕掛けをアプリケーションに付加できる「サーチコンポーネント」を多数紹介しています。そして第 10 章はユーザ企業の立場から経験上得られた貴重な知見や Solr の癖などの情報を紹介しています。

サンプルコードのダウンロードと本書のサポート

本書で紹介しているサンプルプログラムや設定ファイルは、技術評論社のホームページをはじめ、執筆陣が勤務する下記の会社のホームページからダウンロードできます。

株式会社シーマーク	http://www.seamark.co.jp/
株式会社ロンウイット	http://www.rondhuit.com/

本書の執筆には万全を期しましたが、ソフトウェア同様、残念ながら不具合が紛れ込む可能性があります。万一誤りを発見したり、手順どおり試したのに動かないことなどがありましたら、上記ホームページの問い合わせフォームなどからお気軽にご連絡ください。また、下記ブログのコメント欄でも受け付けます。

関口宏司の Lucene ブログ	http://lucene.jugem.jp/
------------------	---

いただいた問い合わせには、個別に回答したり、ブログの記事上で回答したり、正誤表を作成したりなど、何らかの対応をしたいと思います。なお対応には時間がかかる場合がありますので、あらかじめご了承ください。