● 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，2013 年 10 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，2013 年 10 月現在で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かずに，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おりません。

はじめに

楽しい検索の世界へようこそ！

今日，検索を利用しない生活は想像ができないくらいに，身の回りに検索技術がありふれています。普段，私たちが何気なく検索しているところで，実は本書で紹介する Apache Solr（アパッチソーラー：以下 Solr と記します）の技術が利用されているかもしれません。

本書はそんな Solr を体系的に解説した書籍となります。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 は非常に優れたライブラリですが，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。Solr はその手間を解消するために HTTP でアクセスできる REST ライクなインタフェースを用意しています。また，インタフェースだけでなく，機能や使い勝手を向上するしくみがプラスされています。

前回の『Apache Solr 入門』の出版から約 3 年が経ちました。Lucene/Solr もバージョンが 4.5.1（2013 年 10 月現在）となり，前のバージョン（Solr 1.4）に比べると大量データを扱うためのしくみや位置情報を活用した検索など，機能が大きく追加変更されています。これは，同様に Lucene をベースとした検索エンジン ElasticSearch が好敵手として登場しコミュニティが良い形で活性化されたこと，また，利用の拡大によりさまざまなニーズに対応してきたことの結果でしょう。

また，最近では Solr を核としたプロダクトや OSS（Cloudera Search や Yokozuna）も登場するなど，Solr 理解の重要性がより高まっている状況といえます。このように活発なコミュニティに支えられ，さまざまなシステムを支えるミドルウェアとなりつつある Solr に関して，このたび，ユーザの中でも特に経験豊かな執筆陣により，改訂版を出版できる運びとなりました。

本書が皆さまの理解に対する一助となることを願っています。

では，また本文でお会いしましょう！

2013 年 10 月

大谷　純

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。さらに，Christian Moenさんにも感謝いたします。Christianさんにより Kuromojiという日本語形態素解析ライブラリが寄贈されたことによりSolrでより簡単に日本語での検索が行えるようになりました。また，大谷個人としては，担当編集者の池本さんとリクルートテクノロジーズの中野さん，ロンウイットの関口さん，Lucene/Solrコミュニティに感謝いたします。特に池本さん，中野さんは本書の改訂版を出版できるように尽力してくれました。関口さんには前回の『Apache Solr入門』を書くきっかけと，いろいろと勉強させていただく機会を与えてもらっています。そして，Lucene/Solrコミュニティは，私に日々開発者MLからのエキサイティングな英語のメールをプレゼントしてくれ，英語の勉強をする機会を与えてくれました。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9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。Chapter 1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かまいません。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良いでしょう。

Chapter 2からChapter 4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。ぜひ一とおり読んでおくことをお勧めします。

Chapter 5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/JSONで返ってきます。そのため，XML/JSONをHTMLなどに変換するフロントエンドが必要ですがChapter 5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。

Chapter 6以降はさらにSolrを活用したい人たち向けのパートです。Chapter 6ではリレーショナルデータベース，ファイルサーバなどのさまざまな情報リソースから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」や「ManifoldCF」の使い方を紹介しています。

Chapter 7では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」の紹介や，全文検索以外の応用的な検索機能である位置情報検索やJoin検索を紹介します。

Chapter 8はスケーラビリティを確保するためのしくみである，巨大なインデックスを分割して検索する「分散検索」の設定方法や使い方，レプリケーションの方法，また，Solr 4から新たに導入されたSolrCloudと呼ばれるクラスタ構成の設定方法や使い方についても紹介します。

そしてChapter 9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。

最後に，付録として各種API（スキーマ操作，マルチコア操作，コレクション操作）の簡単な説明を記載してあります。

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
株式会社ロンウィット	http://www.rondhuit.co.jp

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。

@johtaniの日記 2nd　http://blog.johtani.info

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。

Contents

［目次］

付 録