

# 株式会社セブンネットショッピング

オープンソース検索エンジン Solr/Lucene の採用で、検索精度とパフォーマンスの大幅な向上を実現。

2011 年 11 月に検索サービスをリニューアルし、さらにユーザー・フレンドリーな EC サイトとして新しいスタートを切ったセブンネットショッピング。国内の大手 EC サイトとしては初となるオープンソースの検索エンジン Apache Solr/Apache Lucene（以下 Solr/Lucene）と、形態素解析器 lucene-gosen を採用しています。開発に携わったセブンインターネットラボの北林洋太さん、都築和仁さん、本間文康さん、田中晴菜さんから、プロジェクトの全体像や今後の目標についてお話をうかがいます。



## プロフィール

株式会社セブンインターネットラボ



ソーシャルソリューション本部  
システム 2 部  
シニアリーダー  
北林洋太さん



ソーシャルソリューション本部  
システム 1 部  
リーダー  
都築和仁さん



ソーシャルソリューション本部

システム 2 部

システム 3 チーム チーフ

本間文康さん



ソーシャルソリューション本部

システム 2 部

システム 3 チーム

田中晴菜さん

## 検索精度とパフォーマンスの向上を目的としたリニューアルを実施。

■ 今回のプロジェクトの概要について教えてください。

北林／セブンネットショッピングはセブン＆アイ・ホールディングスの総合ネット通販として 2009 年 12 月にサービス・イン。アイテム数約 500 万、会員数 900 万人という国内有数の規模を誇る EC サイトです。検索エンジンとして、当初は国産の商用ソフトを採用していましたが、検索精度やパフォーマンスの改善を課題としていました。

具体的には文章解析に N-gram 方式のものが用いられていたため、お客様がサザンオールスターズの CD を購入しようとして「サザン」と入力しているのに「サザンカ」に関連する商品が上位に表示されてしまっていたのです。

また画面生成のために多数のクエリーが走る構造だったため、検索結果の表示に時間を要することがありました。

こうした問題を解決するため、新しい検索エンジンの検討に入ったのが 2010 年 11 月のことです。最終的な候補として残ったのは、商用ソフトと、オープンソース（OSS）の Solr/Lucene のふたつ。検索精度、実装の容易性、保守性、コストなどさまざまな観点から比較検証しました。検証環境で実験してみると、Solr/Lucene はサーバーなどのスペックが低くても、パフォーマンスが低下しにくいことが分かりました。

こうしたことから Solr/Lucene の採用が決まり、2011 年 5 月から開発がスタート。11 月 18 日に新しい検索エンジンへの切り替えが完了しました。

## OSS ならではの開かれた開発体制に、商用ソフト以上の安心感があった。

■ OSS を採用することに不安はありませんでしたか。

北林／商用ソフトは、メーカーという閉じた環境で開発やバージョンアップが行われています。機能を知るためにはメーカーが用意した説明書を読むしかありません。

以前の検索エンジンを使った開発では、不明点をメーカーに尋ねてもタイムリーな回答を得られなかったことがありました。

それに対して、OSS はオープンな環境。特に Solr/Lucene はコミュニティが非常に活発に活動しています。何か疑問点が生じて、メーリングリストに質問を投げれば、誰かが答えてくれます。またバージョンアップもメーカー主導で行われるものではないため、更新の頻度が落ちるということはありません。

EC サイトという性格上、お客様のニーズを仮説検証し、常に辞書登録を更新し、設定をチューニングしていく必要があります。こうした点を総合的に考えると、むしろ OSS である Solr/Lucene を選択するほうが安心感がありました。

都築／もちろん、ロンウィットという会社が存在していたのも安心材料のひとつでした。日本人唯一のコミッターとして Solr/Lucene の開発に参加している関口さんから、直接、トレーニングやサポートを受けることができるわけですから、OSS を不安視する必要はまったくないだろうという判断になったのです。

## 検索時の負荷を分散するため、複数のサーバーで機能を分担。

■ 実際の開発においてはどのような苦労がありましたか。

北林／Solr/Lucene を採用したことによる苦労は特にありません。今回、私たちが重視していたのは、とにかくパフォーマンスを向上させること。99%の確率で、1 秒以内に検索結果を返す、というのが目標でした。

そのため、Solr のキャッシュに加えてバックエンドにインデックス生成用の別サーバーを立てるなどの工夫をしています。

今回のプロジェクトでは実装に入る前に 1 カ月半ほどかけてプロトタイプを使って性能検証を行い、その性能検証結果を

受けて、ハードウェア構成を決定するという手順を取っています。

都築／プロトタイプの開発時には、特定のキーワードで検索結果がヒットしない、という問題も起こりました。ロンウイットに質問を投げたところ、非常に短時間で回答をもらうことができ、設定を変更することで解決しました。また、Solr のマルチコア、マージを活用した検索インデックス切り替え方式の妥当性検証などにもロンウイットに支援していただきました。

北林／検索対象、スコアの重み付けなどは、稼動後にも頻繁に変更の必要があると考えていたので、XML ファイルでこれらを変更することで迅速に対応できるしくみにしました。また、カテゴリーごとにサーバーを分けるなど、かなりの部分で分散化を図っています。



**Solr の実力を引き出すためにも、丁寧なチューニングは不可欠。**

■テスト段階での問題点などありましたか。

北林／今回、テストには約 2 ヶ月の期間をかけています。関係者のみに公開される検証用の仮サイトを設け、100 名体制で様々なキーワードを使ったテストを実施しました。

ここでも課題になったのは、パフォーマンスの問題。例えば、セブンネットショッピングには専門店というカテゴリーがあるのですが、データが一般商品とは異なる階層にあります。そのため、クエリーが複数、走ることになり、検索結果を表示するまでに時間がかかってしまう。そこで検索実行を非同期にするなどして、パフォーマンスの向上を図っています。

都築／EC サイトとしての使い勝手の良さを実現するため、テスト段階になって仕様が変更になる部分はかなりありました。

スコアの重み付けなども、当初は Solr がデフォルトで持ってくるデータを重視していたのですが、売れ筋や予約受付中のものを重視すべきだ、といった指摘を受けて設定を変更しています。

Solr/Lucene を採用すれば、それだけで良い検索結果が得られる、というわけではなく、やはりチューニングは必要。むしろ、そうした設定の変更にも柔軟に対応できるという点が、Solr/Lucene を採用したことの大きなメリットでした。

**検索結果表示の大幅な高速化を実現。検索精度のさらなる向上を目指す。**

■リニューアルによって、どの程度の効果がありましたか。

都築／今回は形態素解析器としてやはり OSS である lucene-gosen を採用しています。採用を決定する際には性能比較を行いました。その結果、lucene-gosen の方が約 3 倍も高速であることがわかりました。

北林／検索結果の表示に 1 秒以上かかることがリニューアル後は 0.4% にまで低減し、高速化を実現しています。

ただし、検索精度を向上させるためには、手作業で辞書登録していくしかない。そこは保守運用のスタッフの尽力によるものです。

田中／辞書登録は、商品名や人名などを中心に追加しており、最終的には lucene-gosen のデフォルト約 49 万語から 70 万語まで拡張しました。その他で苦労したのは、同音異義語ですね。例えば「マツケン」で検索された時、表示すべきなのは松平健なのか、松山ケンイチなのか。お客様のニーズによってソートできるよう、チューニングを加えています。同音異義語だけでも数千個は登録しています。

本間／検索エンジンは、リリースがゴールではなく、スタート。新商品の追加や在庫状況、発売日の変更などがありますから、1 日に数回の検索インデックス更新を行っており、さらに頻度を高めるチューニング中です。

EC サイトでは、検索精度が低いと、リアル店舗で考えれば、在庫はあるのにお店に並んでいないのと同じことが起こってしまいます。チューニングに関しては、運用のフェーズに入ってから絶えず見直し続けていくことになるでしょう。

## グループ各社への導入も視野に、ノウハウを蓄積していき

い。

■今後の目標について教えてください。

北林／現在、要望として受けているものにサジェスト機能の実装があります。Solr/Lucene には検索キーワードを集積していく機能がありますから、いわゆる「もしかして機能」なども搭載していきたいですね。

将来的には、検索結果の一覧表示をパーソナライズしていく方向性になると思います。

本間／検索キーワードは、お客様の声とも言えるもの。EC サイトにとっては財産ともいえるべき重要なマーケティングデータです。それを有効に活用できるという点でも Solr/Lucene を採用したことのメリットは大きかったですね。

都築／今回、Solr/Lucene を採用することで、期待した以上の効果をあげることができました。今後は、セブン&アイ・ホールディングスの他のサイトにも Solr/Lucene を採用していくことも検討しています。その意味でもセブンネットショッピングでの検証を重ね、活用のノウハウを蓄積していきたいですね。

北林／Solr/Lucene で、こんなことができないか、など、ロンウィットには、色々と無理難題をお願いすることになるかもしれません。新しい機能が実装された時のアナウンスやパッチの更新など、セキュリティ面でのサポートにも期待しています。



株式会社セブンインターネットラボ

代表者 鈴木康弘

業務内容 IT におけるシステム開発、新規ビジネスおよび

データ活用の研究

設立 2009 年 3 月 24 日

資本金 4 億 9,995 万円

株主 株式会社セブン&アイ・ネットメディア

日本電気株式会社

株式会社セブンネットショッピング



株式会社セブンネットショッピング

代表者 鈴木康弘

業務内容 インターネットによる商品の販売、およびサービスの提供

設立 1999 年 8 月 24 日

資本金 30 億円

株主 株式会社セブン&アイ・ネットメディア

ヤフー株式会社

株式会社トーハン

鈴木康弘

※記事内のデータは、取材日現在の値です。

(取材日：2011 年 12 月)