

Software

Apache Solrの開発者Yonik Sheely氏インタビュー 次世代サーチ・テクノロジーフォーラム東京

4月22日、国際文化会館にてベイシス・テクノロジー(株)により「次世代サーチ・テクノロジーフォーラム東京」が開催された。同社は、海外ではアマゾンやマイクロソフト、グーグル、日本ではgoo、BIGLOBE、NTTデータ、富士通などを顧客とし、自然言語処理のエキスパートとして言語処理プラットフォーム「Rosette」を提供している。

Carl Hoffman CEOは、基調講演において、Web検索だけでなくエンタープライズ検索の重要性を述べ、検索市場の発展と技術の進化をアピールした。続いてプロダクトマネージャのSteve Kearns氏は、正規化されていない非構造化データ、つまりメールや社内文書などのメタデータをどのように検索し活用すべきか問題点を論じ、そこにRosetteを適用することで企業が求めるデータを速やかに提供可能であることを発表した。

ベイシス社の言語処理技術(RLP)はオープンソース検索エンジンであるApache SolrやLuceneと組み合わせることで利用可能であることから、Solrの開発者であるYonik Sheely(ヨニック・シーリー)氏も来日。本誌でインタビューする機会を得た。

——Apache Solrを開発した動機を教えてください。

「ご存じのようにCNETでSolrを開発しました。商用ベンダーの検索エンジンの代替として、つまりバックアッププランの1つでした」

——現在はRESTfulなWeb技術が主流ですが、Solrが完全にRESTfulではない理由を教えてください。

「Webアプリでいえばデータベース層の位置づけでSolrを開発していたこと、そして完全にREST対応にするよりも、使いやすさや実用性を優先させたからです。Solr

の機能を制限したくなかったのです」
——クラウドコンピューティングへ応用はいかがでしょうか。

「今後実装する機能として[SolrCloud]を発表しました。Solrの分散化検索機能を次のレベルに引き上げるものです。処理のオーバーヘッドを低減するのが目的です。より大きなクラスターに対応できます。

Solr自体は、Web情報の検索のために作りましたが、今後はエンタープライズサーチの機能強化を図っていきたいと考えています。PDFなどのリッチドキュメントの検索や企業独自のデータなどが対象になります。また、地域・ローカル情報に紐付いた情報を検索する技術も考えています」

企業でのIT利用が進み、すぐに使える正規化・構造化されたデータだけでなく、さまざまな形でデータが莫大に蓄積されているようになった。これらのデータを価値あるものにするには、データマイニング技術などが先行していたが、自然言語処理研究の進化や検索技術の向上によりエンタープライズサーチの重要性が増している。さらなる展開を期待したい。



●ルーシッド・イメージーション社 Yonik Sheely氏
(Solr開発者・Lucene/Solrコミッター)



● Carl Hoffman
CEO



● Steve Kearns氏