

Rosette®文字・言語コード判別システム

NTT レゾナント株式会社様

会社概要：

NTT レゾナントは、インターネットポータルサイト「goo」の運営のほか、Eラーニングサービス、映像コミュニケーションサービスなど、インターネットユーザの幅広いニーズに応える利便性の高いサービスを提供しています。

事例概要：

NTT レゾナントが運営するインターネットポータルサイト goo は、2005 年 12 月に、インターネット上の画像、動画、音楽データを任意のキーワードで検索できる「画像・動画・音楽検索」サービスを開始しました。本サービスでは、ウェブ上に公開された画像、動画、音楽をキーワードで検索できます。検索DBは常時、追加・更新され、いつも最新の情報を検索できます。

Basis Technology 製品導入背景：

「画像、動画、音楽検索」は画像などにつけられた説明文をキーワードで検索します。適切な検索結果を提供するためには、画像などが掲載された Web ページの言語に合った処理を行って索引を生成しなければなりません。それには、対象ページの言語、文字コードの正確な判別が求められます。この課題を解決するために、ほぼ 100%という高い精度で判別を行う Basis Technology の Rosette® 言語・文字コード判別システム(RLI : Rosette® Language and encoding Identifier) が採用されました。さらに、ユーザに常に最新の情報を提供するために、「画像・動画・音楽検索」の索引も、日々、更新していく必要があります。こういった用途には何よりも安定性と処理速度が求められますが、Rosette 言語・文字コード判別システムはこの点でも高く評価されています。

導入効果：

Rosette 言語・文字コード判別システムの利用により、対象ページの言語や文字コードを正しく認識することが可能となり、最適な自然言語処理を行なうことができ、効率的で精度の高い検索が可能になりました。Rosette 言語・文字コード判別システムは処理速度が速く非常に安定しているので、サービス開始後の継続的な検索DBの追加・更新も効率的かつ順調に進められており、ユーザに常に最新の情報を提供しています。

NTT レゾナント、ポータル事業本部技術マーケティング部担当部長の竹野浩氏のコメント：

「Basis Technology の言語・文字コード判別システムの導入により、インターネット上のテキストデータの記述言語を明確に判別することにより、最適な自然言語処理をおこなうこと

が可能になり、検索の効率と精度の向上に大変役立っています。また、膨大な量のテキストデータも瞬時に処理することができる処理速度、高い言語判別の精度など、この製品のパフォーマンスにとっても満足しています。」

構成図：

