

スペルミスに対し正しい単語を示唆するシステムを開発、販売

2004年2月16日

株式会社ランゲージ・クラフト（本社：東京都目黒区、代表取締役：関根聡、<http://languagecraft.jp>）は、キーワード検索においてミススペルやかな漢変換ミスにより検索結果が見つからないという状況に対応できる、スペルコレクションのシステムを開発しました。

本システムを利用すれば、「もののけ姫」と誤った綴りで検索がされても「探しているのは“もののけ姫”ではありませんか？」と表示することができます。

例えば、WEBの検索システムに「ウィンドーズ」と入力したら、普通、期待した検索結果は出ません。「もののけ姫」「風邪の歌を聞け」「外付ディスク」でも出ることはありません。これはユーザーのミススペルによるものですが、このような間違えは、実はよく起こっています。弊社の調査では「ウィンドーズ」では33%、「風の歌...」「外付け...」共に14%、「もののけ姫」でも5%ものミススペルがWEBページに見つかりました。平均5%のミススペル率だとしても、検索頻度や売上を掛け合わせた損失を考えれば非常に大きな問題になっています。

英語の検索をgoogle, yahoo, MSNなどでやってみると分かりますが、英語ではミススペルを示唆してくれる機能があります。しかし、よく知られているアルゴリズムは26文字の英語では有効ですが、日本語では文字の数の多さから同じ機能を実現することは不可能です。

ランゲージ・クラフト研究所は、新しいアルゴリズムを開発し、日本語のミススペルを示唆できる機能を実現しました。頻出する表記の揺れ辞書を用意しておくという方法とも異なり、どのような入力にも対応できます。30万語の辞書を利用した実験では、候補の1位が正しい単語だった割合が93%、3位まで含めると100%という精度で、速度も平均0.01秒と瞬時に計算いたします。（ただし、精度の保証は致しません）

今回、ランゲージ・クラフト研究所はこの日本語スペルコレクションを2つの方式で販売いたします。

- 専用サーバーマシンをお客様の所に置く方式
契約金：400万円（2台目以降300万円）
月間ライセンス料：40万円
- 弊社のマシンを共有する方式
契約金：20万円
月間ライセンス料：8万円から（アクセス数によって変動します）

共に、上記以外に辞書のインデックス作成費用（20万円より。単語数によって変動します）が必要です。単語リスト作成、読みの付与、インストールなどのオプションも用意しております。契約金の支払いは1ヶ月システムを利用した後という形になっています。

システムの詳細およびデモはランゲージ・クラフト研究所のホームページ<http://languagecraft.jp>にてご覧いただけます。3月15日(月)午後2時より、東京・千駄ヶ谷の津田ホール会議室において製品説明会を行う予定にしております。

株式会社ランゲージ・クラフト研究所は、言葉に関する研究開発を通して、使いやすいコンピューター環境を実現していこうと考えております。本研究所は15年以上自然言語処理の研究に携わってきたニューヨーク大学の関根助教授を中心とするソフトウェア開発、技術コンサルティングの会社です。このたびは小さな委託研究の成果からこのような製品が生まれ、広く世に問うことに致しました。なにとぞよろしく願いいたします。