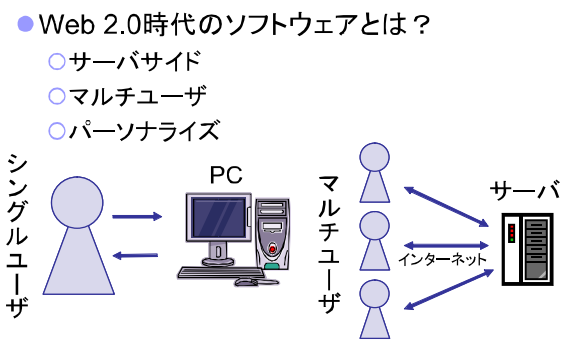
Social IME: サーバサイド日本語入力とログ活用サービス
～みんなで育てる日本語入力～

奥野 陽（慶應義塾大学理工学研究科）

1. 背景

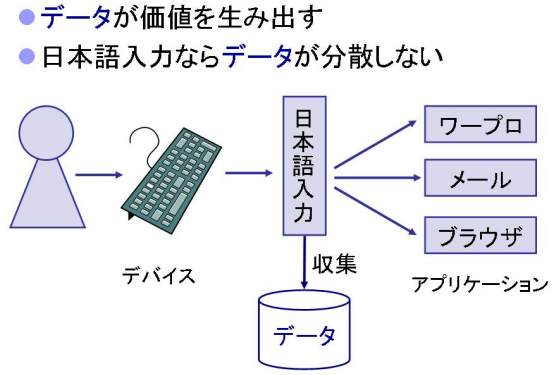
Web 2.0 の時代と言われ、さまざまなソフトウェアがサーバサイドに移行しています。今まででは考えられなかったような個人的な情報、例えばブックマークを大量に集積することで"**集合知**"を作り出せることが分かってきています。しかし、必要なデータを十分に集めることができないケースも少なくありません。



2. 目的

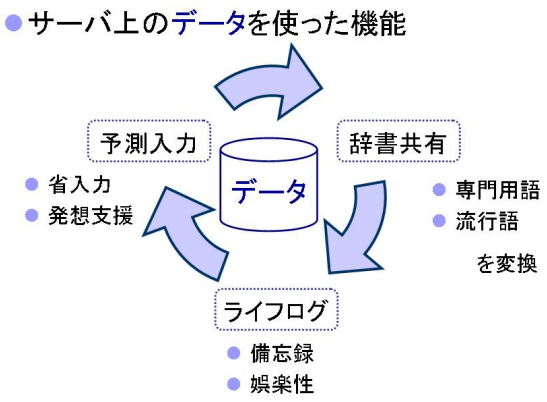
私は、この時代に集合知を集めるための方法として、**日本語入力**に着目しました。なぜなら IME による漢字かな交じり文の入力は、PC でのデータ入力の多くの割合を占めるからです。

従来の日本語入力はスタンドアロンを前提に設計されていました。しかし日本語入力をサーバサイドで実現することで、膨大な**変換のログ**や**単語の辞書**を集めることができます。そして、その膨大なデータよる**集合知**を活用したサービスを提供できます。このようなシステムは今までに存在しません。まさに未踏の領域です。



3. 開発の内容

“Social IME”の概要は、右の図のようになります。専用のクライアントソフトと、かな漢字変換を行うサーバを合わせたシステムです。クライアントは、Windows のデフォルトの IME の代わりとして違和感なく使えるように実装されています。また、外部の Web サイトに自動投稿を行い、ライフログ的なサービスを提供します。



Social IME ならではの機能として、「**予測入力**」、「**辞書共有**」、「**ライフログ**」の 3 つがあります。これらの機能はサーバ上のデータを使ったものであり、集合知の具体的な応用例として、ユーザにアピールできるポイントとなっ

ております

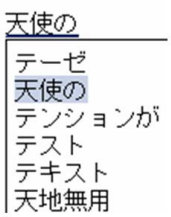
予測入力は TAB キーによって行います。右の図の例のように、先頭の 2～3 文字を入力してから予測を行うと全てのユーザの過去の履歴から新しい順に候補を表示します。また、1 文字で予測を行うと思いもかけない候補が表示されることがあり、発想支援などにも利用することができます。

- 最初の数文字＋TABキーで予測入力
- 省入力や発想支援に利用

例

おね→お願いします
あぷり→アプリケーション
にほ→日本語
みと→未踏
そふ→ソフトウェア

「て」+TAB



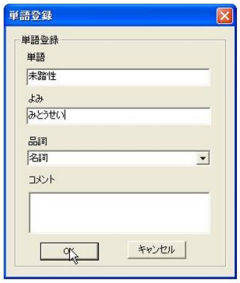
単語登録画面では、変換できない単語を登録すると、次から変換できるようになります。このように登録された専門用語や流行語の辞書は全ユーザで共有されるため、デフォルトでも右の図のような単語を変換することができます。また、「はてなダイアリーキーワード」に登録された 20 万語以上の単語を変換に用いることもできます。

- 変換できない単語を登録できる
- みんなで専門用語や流行語の辞書を共有

例

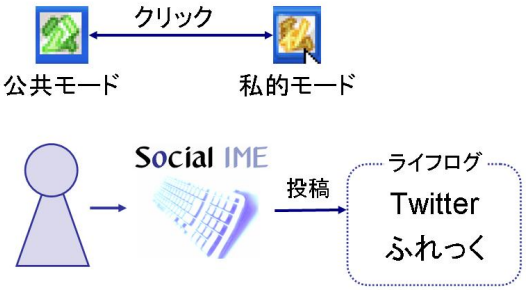
未踏性（みとうせい）
集合知（しゅうごうち）
創発（そうはつ）

亀田 和毅（ともき）
初音ミク（はつねみく）
東方 永夜抄（えいやしょう）



ライフログの開発は**公共モード時の外部サイトへの自動投稿**として実装を行いました。言語バーで「公」モードと「私」モードを切り替えることで、自動投稿を行うかどうかを選択することができます。IME のログは一種のライフログとしてとらえられ、現在は Twitter とふれっくに対応しています。

- 公共モード時、外部サイトに自動投稿



Social IME をインストールすると、**言語バー**のアイコンが右の図のように変更されます。左から順に、「入力切替」ボタン、「半角・全角」ボタン、「公共モード」ボタン、「単語登録」ボタンです。公共モードについては後述します。単語登録ボタンを押すと、単語登録ダイアログを表示します。

4. 従来の技術との相違

Social IME は**未踏性**の高いソフトウェアです。 サーバサイドで変換を行うというアイデアは、古くは UNIX の Canna、最近では Google Suggest などが発想としては近い。しかし最も普及している Windows でサーバサイドの変換を行った例も、インターネット全体で辞書や変換結果の共有を行った例も、今までに存在しません。**集合知**の

実現にはインターネット全体でデータを収集する必要があると考えられており、集合知の実現を目指して開発された日本語入力は、現在のところ Social IME の他に存在しません。

5. 期待される効果

・かな漢字変換分野へのインパクト

今までのかな漢字変換のアルゴリズムは、「正しい日本語」の文法に関する知識と開発者の感覚から職人芸的に作られていました。サーバサイドかな漢字変換では、ユーザが**実際にどのように利用しているか**をリアルタイムに反映したチューニングができます。またアルゴリズムを変更してもアップデートが不要なので、**ユーザのフィードバック**を受けやすいという利点があります。これらによって、柔軟な開発体制を敷くことができるようになりました。このような観点から、かな漢字変換分野の開発者にとって Social IME の登場が驚きを持って受け止められました。

・Windows IME 分野へのインパクト

従来、Windows 上で利用される IME は Microsft IME と ATOK の寡占状態にありました。MS IME は OS デフォルトの IME として、かなりの変換精度を持っています。しかし、クライアント PC にインストールされるタイプである以上、登録されている辞書などのデータ量に限界があります。また、ATOK は有料のソフトウェアであり、導入に当たっての敷居が高いという問題点があります。

無料で使える Social IME の登場は、このように Windows で利用できる IME 自体が少なく、**十分な開発競争が行われていない**現状に一石を投じることとなりました。

・エンドユーザへのインパクト

Social IME は現在のところユーザからは主に特殊な単語を変換する際に便利な IME として受け止められているようです。今後の開発により、より一般的なケースにおいても有用な機能を増やすことで、多くのユーザに利用される新しい入力システムに成長する可能性を秘めています。

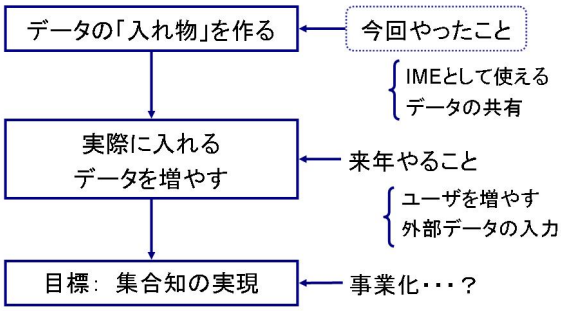
6. 普及の見通し

以下の URL で Social IME のベータ版の公開を行なっています。

http://www.social-ime.com/

2008 年 1 月 18 日現在、1,000 件以上のダウンロードがされていますが、十分とはいえません。今後の広報活動に加えて、サーバの能力の向上、新機能の開発などを通して普及を計っていく予定です。

長期的に見ると、今回のプロジェクトは右の図のように**「データの入れ物を作る」**フェーズだったと言うことができます。今後はまず来年の開発によって「実際に入れるデータを増やす」フェーズに入ります。最終的には集合知に実現を目指して、事業化も視野に入れています。



7. 開発者名（所属）

奥野 陽（慶應義塾大学理工学研究科）

E-mail:nokuno@nokuno.jp
URL: http://d.hatena.ne.jp/nokuno/