

GNU/Linuxなどにおける、よりよい日本語環境の実現

Implementation of Improved Japanese Input Method for GNU/Linux and Other Systems

田畑 悠介¹⁾ 鵜川 始陽²⁾
信岡 孝佳³⁾ 高井 幸輔⁴⁾
Yusuke TABATA Tomoharu UGAWA
Takayoshi NOBUOKA Kosuke TAKAI

- 1) 京都大学大学院情報学研究科 Email: yusuke@kmc.gr.jp
- 2) 京都大学大学院情報学研究科 Email: foosen@kuis.kyoto-u.ac.jp
- 3) 株式会社トランス・ニュー・テクノロジー 京都研究室 Email: taka@trans-nt.com
- 4) 京都大学大学院情報学研究科 Email: takai@vlsi.kucee.kyoto-u.ac.jp

ABSTRACT.

In these days, many kinds of software such as operating system, desktop environment, web browser, or office suites are available as an open source software, and most of them are fully useful for daily use. However Japanese input method systems are not so useful. Systems widely used these days was developed for UNIX workstations in more than 10 years ago and their developments are stopped. In this 10 years, the computing environment was radically changed. For example, security is becoming an important issue, and computers are becoming powerful. We developped new Japanese input system "Anthy" in this project. Anthy is designed for modern personal computers with Linux or other UNIX like operations systems. Anthy's conversion modules is provided as a dynamically linkable library instead of a server to keep higher security. Anthy adopts simple algorithm and does deeper search for precise conversion. Additionally, we developped some applications for Anthy to be used in many environments: Emacs, X window system and GTK.

1 背景

近年 Linux などのオープンソースソフトウェアを用いて各種サーバから組込み機器などの広い範囲のコンピューティング環境を構築することができるようになりつつある。またデスクトップシステムにおいては近代的なデスクトップ環境やウェブブラウザ, オフィススイーツなども登場しており, 常用に耐えるものとなりつつある。

しかし, 我々日本人がコンピュータを使用する際には欧米などの多くの国の人々とは違い日本語入力のためのシステムが必須であるが, 現在の Linux などで使用できる Canna や FreeWnn などのオープンソースの日本語入力システムは 1990 年前後に国産 Unix ワークステーション用として作成されたものであり, その進化は 1990 年代後半においてはほとんど停止してしまっている。

2 目的

1 章で述べた問題を解決するために, 本プロジェクトではオープンソースな日本語入力システムの構成を行う。また, ソフトウェアを開発するだけでなく, ノウハウの公開を行うことによって着実にメンテナンスされるソフトウェアを目指す。

3 内容および成果

本プロジェクトでは次のような内容の作業を行った。

- かな漢字変換エンジンの性能の向上
プロジェクト開始以前から開発を行ってきたかな漢字

変換エンジン Anthy の改良を行った。

- かな漢字変換エンジンのインターフェイスの作成
各種のデスクトップシステムやアプリケーションから日本語の入力を行うためのユーザインタフェースおよび, そのユーザインタフェースを作成するためのライブラリのインタフェースの作成を行った。

これらについて説明する。

3.1 変換エンジンの性能の向上

かな漢字変換エンジン Anthy は以前からシンプルなアルゴリズムを使用していたが, それをよりシンプルにして計算の量を増やすことによって高い変換精度を実現することができた。

また, ソフトウェアを公開してフィードバックを得ることによってパラメータのチューニングなども行うことによって性能の向上を実現できた。

3.2 インタフェースの作成

作成したかな漢字変換エンジンを使うためのアプリケーションおよびライブラリを作成した。

- 入力ライブラリ
かな漢字変換を用いて日本語を入力するためには入力途中の状態の管理や設定の管理などを行う必要がある。このためのライブラリを変換エンジンとは別に作成した。
- emacs 対応
Linux において広く使われているテキストエディタである emacs で Anthy を使用するためのソフトウェアを作成した。既存のシステムの多くはかな漢字変換エ

エンジンとの通信を暗号化されていない TCP を用いて行い、ユーザの認証を省略しているためセキュリティやプライバシーの保護に問題があるが、今回作成したソフトウェアは安全なパイプ通信を用いている。

- XIM サーバ
X Window System で広く使用されている入力プロトコルである XIM プロトコルを利用するサーバを作成した。
既存の XIM サーバよりも近代的なユーザインタフェースを実現している。また、ユーザインタフェースと XIM サーバのパートを分離することができるので、各種のツールキットに対応させることも可能である。
- gtk+-immodules
Linux で広く使われているツールキットである gtk+ の次期バージョンでは入力システムのための immodules というフレームワークを持っている。これはオブジェクト指向のシステムであり、シンプルな入力システムから日本語のような複雑な入力システムまで様々なものに対応する能力を持っている。
本プロジェクトではこのシステム上に Anthy を用いた入力のためのモジュールを作成した。
- その他
入力ライブラリのテストを兼ねて、エディタや web ブラウザ (w3m) への Anthy による入力の実装を行った。

4 評価

既存のオープンソースの日本語入力システムと比較した場合、次のような特徴を持つソフトウェアを作成することができた。

- 高性能
シンプルなアルゴリズムで大量の計算を行うという方針をとることによって、既存のシステムを越える精度でかな漢字変換を実現することができるようになった。
- 高い安全性
近年のソフトウェアは以前のものと比較して高いセキュリティが要求されている。しかし、既存の日本語入力システムの多くは設計された時代が古いためセキュリティやプライバシーを保護する機能が低い。これに対して今回作成したシステムは外部からの接続を受け付けないう設計されているため高い安全性を実現することができた。
- シンプル
できる限り単純なアルゴリズムを採用することにより、既存のオープンソースの日本語入力システムよりも理解が容易なシステムにすることができた。これによって開発への参加が容易になると思われる。

本プロジェクトで作成したソフトウェアは現時点では数十人に常用されており、これからも継続的に開発を続けることによりより広く使用されるソフトウェアとなることが期待できる。

5 参加企業及び機関

なし

6 参考文献

なし