

SUNのPDSによる日本語環境

電子技術総合研究所 半田 剣一、小方 一郎

Kenichi Handa、Ichirou Ogata

1. はじめに

米国では以前からネットワーク上でパブリック・ドメイン・ソフトウェア (PDS) が公開され、無料のソフトウェアが流通するようになった。ネットワークが主に UNIX の uucp ネットであったため、PDS は UNIX の上で各種の小さなツールであることが多かった。バグ情報や改良などもネットワーク上で精力的に交換され、なかには非常に有用なツールと成長したものもある。

最近になり、大きく、また非常に優秀なソフトウェアが PDS に放出され始めた。なかでも次の3つは規模の大きさと優秀さとで際立っている。

(1) フリー・ソフトウェア・ファウンデーション (FSF) による

Gnu Emacs スクリーン・エディタ

(2) マサチューセッツ工科大学による

X-Window ウィンドウ・システム

(3) Knuth, アメリカ数学会, ワシントン大学による

Unix TeX 清書システム

今回話題となるのはこれら3つのソフトウェアの日本語化である。

また日本でも優秀な無償ソフトウェア (厳密には PDS ではない) が開発された。これが

(4) 京都大学, アステック (株) による

Wnn (うんぬ) 日本語仮名漢変換システム

である。

これらを組み合わせた快適な日本語処理環境が整いつつある。これらのソフトウェアは現在急速に発展 (バージョンアップ) しつつあるが、電総研では比較的最新のバージョンが稼働していると思われるので、今回の発表では電総研でのこれらのソフトウェアの現状について紹介したい。

なお、今回例としているのは SUN ワークステーションであるが、もちろん News やその他の Unix ワークステーションへの移植も進みつつある。

PDS についてはいろいろ難しい問題もあるが、ここでは、ソース付きで研究目的で自由な利用が許されるものとして、話を進めたい。

2. 日本語エディタ

・Gnu プロジェクト

Gnu プロジェクトは、Twenex 上に作られたあのオリジナル Emacs の作者、Richard Stallman によって始められた。このプロジェクトでは、全ての人々が自由に使える優れたソフトウェアを開発しており、最終的にはオペレーティングシステムを含む、全てのソフトウェア開発環境を提供することを目標としている。

Gnu Emacs (以下 Emacs) はその最初の成果であり、Unix 及び DEC VMS 上で動く新しい Emacs である。Emacs は Stallman の Free Software の主張 (Gnu Manifest) に従って、自由に複製、改良を施すことができる。GnuManifest については [1] を参照して欲しい。

・ n e m a c s

Nemacs は、筆者がこの Gnu Manifest の精神のもとに Emacs を日本語化したものであり、だれでも自由に使うことができる。販売することも自由であるが、この場合、ソースを付け、またユーザが他のユーザにコピーを配布することを禁止してはならない。つまり、Gnu Manifest の精神である。

(1) 入力法 (仮名変換, 漢字変換とのインターフェース)

Nemacs は仮名変換, 漢字変換を直接にはサポートしていない。そこで外部の変換器を使う。外部変換器としては後節で述べる Wnn を用いることもできる。また、パソコン上の仮名漢字変換フロントエンドプロセッサと組み合わせて使うこともできる。

さらに Emacs Lisp プログラミングによって、仮名漢変換を Emacs により直接サポートするものを使っている人もいる。これらのツールもいずれパブリック・ドメインに放出されることが期待される。

(2) 出力表示装置

これであとは (なんらかの) 漢字端末があれば Nemacs が使えるわけである。漢字端末としては NEC のパソコン + 漢字端末 Emulator が電総研ではよく用いられている。

また、次節で述べる漢字 X-Window と一緒に使うこともできる。図 1 の表示例は Sun のビットマップ・ディスプレイ上の漢字 X-Window で Wnn + Nemacs を実行しているものである。

(3) コードについて

いま日本で使われている日本語コード体系は、大きく分けても3種類のものがある。すなわち7ビットコードの JIS コード、8ビットコードのシフトJISコード、および同じく8ビットの EUC コードである。

Nemacs はなるべく多くの機械で動くようにするために、この3種のコードをすべてサポートしている。コードは、キーボードからの入力、ディスプレイ出力、ファイルとの入出力でそれぞれ異なる場合を想定して、それぞれ独立に設定可能である。

図1の一番下のモードラインの FID フィールドが現在使用中のコードを表現している。F はファイルのコード、I は入力のコード、D はディスプレイのコードであり、それぞれに対応した J が JIS、S がシフトJIS、E が EUC コードである。

もちろん将来的にはEUC コードを用いることが望ましいわけだが、現在 TTY や 通信のソフトウェアが8ビットコードをサポートするとは限らないところが問題点である。

・ 配布

Nemacs の配布はすべてボランティア・ベースで行われている。どんな人でも自由にコピーをし、使用することができる。

Nemacs は主に JUNET 上のニュースシステムによって Emacs の Patch として配布されている。いくつかの JUNET サイトではボランティアにより過去の記事の保管が行われているので、そこに頼めばその Patch ファイルが手にはいる。

これを用いて Emacs を Nemacs にするには、やはりパブリック・ドメインのその名も patch というツールを使うのがもっとも簡単である。

Emacs 自体の手にいれ方は近くの人からボランティアでMTかカートリッジテープにコピーしてもらうのが最善である。10メガバイト以上有るので、E-mail でやりとりするのはちょっと難しい。

最後の手段として Stallman の所属する Free Software Foundation から手数料(\$150)のみで送ってもらうことができる。注文書に小切手を同封して送ればよい。Stallman は寄付も受け付けているので、できればいくらか余分に送金できればなお良い。

航空便で送ってもらうにはテープ1本、マニュアル1冊ごとに別料金(\$15)がいる。

・ 課題

Nemacs は87年の春に生まれたばかりであり、まだまだ完成途上である。夏に一度バージョンアップされた。

基本的な機能として文字列サーチが不完全である。また、Emacs の強力な機能の一つである、インクリメンタルサーチも日本語については働かない。

また、例えば、消書機能として日本語 fill-region (禁則処理を考慮しながら一定の幅に揃える)などが望まれる(日本語 auto-fill-mode はすでに組み込まれている)。

入手法ももっと使いやすくしたい。一番考えられるターゲットは Wnn との結合により、仮名漢変換モードを作ることであろう。

3. ウィンドウ・システム

漢字 X ウィンドウは MIT で開発された X-window (version 10) をベースにして、日本語テキストを表示するためのパッチを当てたものである。

X-window の日本語化についても何ヶ所かで同時並行的に進められ、既に幾つかのバージョンが出回っている。PDS としては、東京工業大学の篠田氏のもの(篠田 X)、同じく東工大の佐野・高田氏らによるもの(佐野 X)がある。特に前者は JUNE T 上のニュースに投稿されているので、稼働しているサイトも多い。

両者とも X プロトコル(XText)を改造して、日本語を表示するようにしている。つまり、サーバーとクライアントの双方の改造が必要である。クライアントとしては、シェルである(日本語化された) xterm が使える。

ただし改造の方針が両者で異なり、篠田 X が日本語については 8 ビット目が立った EUC コードを送ることによって日本語と ASCII 文字を区別しているのに対し、佐野 X はフォントを切り替えることによって両者の表示を区別している。

その他にも、サーバーは一切改造せずに、日本語フォントを 256 文字ずつの onx 形式ファイルに分けて格納し、クライアント側で必要に応じてフォントを切り替えるという方式で、X-window を改造している所もある。

ちなみに佐野 X のサーバーは Sony NEWS の X-window システムとコンパチブルであり、NEWS の jterm が SUN の上で使用可能である(逆も可)。電総研では SUN と NEWS がネットワークで繋がっていることもあって、佐野 X をインストールしてある。

日本語を表示するときの問題となるのはフォントである。現在 JIS フォントを使い 16x16, 24x24 のフォントが使用可能である。これで普通の用途には十分である。特に 24x24 は SUN の High Resolution Mono ディスプレイ(1600x1280)でみると結構美しい。JIS フォントはごく僅かの手数料で手にはいるので誰にでも使ってもらえると思う。またパブリック・ドメインに 14x14 のフォントが出回っている。

X-window に関しては、最近やっと version 11 が出回り始めた。このバージョンの日本語化にあたっては、前バージョンの様に各所で勝手にやるのではなく、なんらかの標準を設けようという動きがある。各メーカーにも是非この標準化に協力していただきたいというのが、ユーザからの切実な希望である。

配布に関しては、Nemaes と同じで、篠田 X がオリジナルの X-window に対するパッチの形でニュース上に投稿されているので、JUNE T が利用できるサ

イトは用意に手に入るであろう。

この原稿を書いている今、JUNET上のニュースに東工大の佐野氏作成のkterm (version 11 に対応した漢字版 xterm) が投稿された。サーバーの改造は不要になっている。発表の時点では、より詳しい報告ができるものと思う。

4. 日本語の入力

Wnn (うんぬ) は、京都大学の数理解析研究所が、(株)アステックと(株)立石電機の協力を得て開発した、主に UNIX の上で動く仮名漢字変換システムである。

ソースは C 言語で書かれているため、移植は比較的容易と言われている。このため、色々な計算機の上で、同じ日本語環境を提供できる。

Wnn はいわゆる、フロント・エンド・プロセッサである。具体的には UNIX の仮想端末機能 (PTY) を用いて、ユーザのキー入力 (例えば、ローマ字入力) を仮名漢字混じりの出力としてアプリケーションに送り込む。

ユーザからみた Wnn は大変使い勝手の良いシステムである。これは、キーの定義などユーザに開放できる部分はできるだけ開放した結果ともいえる。これによって、ユーザごとに例えば常用するキーボードの種類に応じて好みのキー定義で使うことができる。

普通のキー定義は Emacs ライクなコントロール・キーによるコマンドとローマ字仮名変換、仮名漢字変換を組み合わせたものである。この場合キー定義はコマンドのみならず、ローマ字・仮名の対応 (例えば 'Xi' で 'い') なども定義できる。

また、JIS 仮名入力と仮名漢変換を組み合わせたか、さらに、句点入力、T-code 入力などの直接入力も勿論可能である。これはキー定義の表のプログラムだけで可能な訳である。もっともそのような表を作ること自体は結構大変だが。

最近、半角仮名入力や、Emacs コマンド入力に便利な「桜川表」(京大の桜川氏による) がネットワークで発表された。筆者らはともにこの表を若干修飾したものを喜んで用いている。

Wnn は、複数のプロセスが協調して仕事をする、サーバ・クライアント方式を採用した実現法をとる。辞書情報を扱うプロセスは jserver と呼ばれる。jserver はクライアントである、wnn (ユーザのプロセス) からのリモート・プロシージャ・コールの形で要求を受け付け、仮名漢変換サービスを行う。プロセス間通信にはソケットを用いる。ネットワーク間のソケット結合を通じて、jserver をネットワークで共有する、inet 化 Wnn も開発されつつある。

jserver は辞書をメモリー中に読み込むため 2Mbyte 以上のメモリを消費する上、辞書探索の特性上局所性が少ないため仮想記憶もうまく働かず、計算機の多少の負担となるのである。そういう意味でも inet 化は待たれる所である。

仮名漢字変換法は、後ろ向きの n 文節評価値最大法というもので、単語頻度、

文節長さ、自立語長さ、などのパラメータに重みを付けて足し算したものが単文節の評価値となり、この単文節の評価値の和を評価値とするものである。これは、普通のワープロなどの変換方式である、2文節最長一致（2文節に関して、文節長さだけをパラメータとするもの）を含んだ、より一般性のあるものとなっている。重みに関しては、これをクライアントごとにダイナミックに変更可能なので、さまざまな評価法を試みることができる。参考文献として、[2]を挙げておく。

5. 消書システム

日本語 TeX (JTeX) についても二つの PDS バージョンが知られている。一つは Ascii の大野氏と倉沢氏のもので、もう一つは NTT の齊藤氏によるものである。これらは近い将来に少なくともその仕様だけは統一されるはずである。

ここではソフトウェア科学会で配布予定の NTT の齊藤氏バージョンについてのべる。これは日本 TeX ユーザグループでの発表及び Tug Boat (米国 TeX Users Group の会報) に投稿された齊藤氏の論文に基づくものである。

JTeX は TeX の全ての機能、例えばあの美しい数式表示、を維持しながら日本語で論文消書ができるようにしたものである。もちろん幅ぞろえ、日本語禁則処理は自動的に処理される。勿論、TeX マクロである LaTeX も使える。

また、X-window 上でのプレビューア (画面上で刷り上がり確かめるもの) である xdvi にも齊藤氏による改造が加えられ、稼働している。

フォントは大日本印刷によってサイト毎に 95,000 円で提供される。これは 50 種ものサイズと明朝、ゴシックの 2 つの書体を含んでいる、大きなものである。

課題としては、プリンター・サポートがいまのところ不足している。確認されているのは Apple Laser Writer および Imagen 8/300, Imagen 3320 である。これを はやいうちに Canon Laser Shot (AJ2) や Sony News LBP などの安価なプリンタで動くようにしたい。

Canon Laser Shot に関しては (英文用の) デバイス・ドライバが日本 TeX ユーザ会からパブリック・ドメインに既に提供されている。これに簡単な改造を施せばよいだろう。

BiBTeX (LaTeX 用文献整理システム) など、TeX Ware と呼ばれるツールの日本語化も課題である。また、フォントは、いまのところパブリック・ドメインのものは使えない。

実はこの原稿の時点ではまだ電総研では JTeX が稼働していないが、発表時点では動いている予定である。

6. おわりに

日本語環境を整える上での難関はまだ多い。完全 PDS 化にあたって問題

となることが幾つかある。ここで問題点を列記し皆さんのご協力を仰ぎたい。

(1) 日本語フォント

日本語フォントはある意味で PDS 日本語環境最大のネックである。いま PDS であるのはアステックの橘氏による 14x14 ドットの PDS フォントだけである。

その他に自由に使用可能なフォントとして電総研で使っているのが、日本規格協会によるいわゆる 16x16 ドット、24x24 ドット J I S フォントである。このフォントは自由に使うことができ、また日本規格協会から若干の手数料で手に入るのだが、これを自由に複製して良いかどうかについては明確な回答が得られていない。

このフォントは「規格」であり、本来 PDS の性格を持つ筈であるので、大学、研究機関での日本語処理研究の土台を提供するという意味でも、関係者のご配慮を是非お願いしたい。

また清書システムで用いるために、書体、大きさともに PDS フォントのいっそうの充実が強く望まれる。

(2) 日本語辞書

現在 Wnn で用いられている日本語辞書は、日本電子協会(???)の配布するものであり、ロイヤリティーが必要である。

これに代わる辞書として現在 JUNET ネットワーク上で公共辞書プロジェクト (public プロジェクト) が進行しつつある。これは、各人が人力(登録)した辞書をネットワーク上で収集し、一つの辞書を作り上げようというものである。基本となる比較的小さな辞書も始めに提供される筈である。このためネットワーク上のみなさんの協力次第では比較的簡単に完成することがみこめる。勿論、電総研有志でもできるかぎり協力するつもりである。

(3) PDS に関する法律上の問題

ここで紹介したソフトウェアは厳密に考えれば PDS と呼べないものも含まれている。例えば、Wnn はいまのところ自由な配布が認められていない。また、Gnu については、GnuManifest に従った強い使用制限がある。もっともこれは前にも述べたように Gnu ソフトウェアをより多くの人に使って貰うためであるが。

その他 PDS が引き起こすであろう問題として、例えば

- ・ 配布・複製は自由にできるのか?
- ・ 改良などに関する権利はどうなるのか?
- ・ 他の(営利目的の)ソフトに組み込めるのか?
- ・ 使用権が打ち切られることはないのか?

などが考えられる。

また、PDS を配布する側の問題として

- ・ 損害を賠償する責任はあるのか？

などが考えられる。筆者は法律問題の専門家ではないので、ここではこれ以上詳しく論じることはできない。読み易い文献として[3]をあげたい。また、Gnu Manifest は「ソフトウェア共産主義宣言」であり、考えようによっては大変過激な読みものである。大変面白いので是非一読をお勧めしたい。

このように、PDS はさまざまな問題をはらんではいるが、Gnu Emacs, X-Window, TeX などは既に研究になくてはならないものとなっている。問題が難しいからといって諦めず、すこしでも研究用ソフトウェアの自由な流通にご協力をいただきたい。

今回話題となったソフトウェアについてはソースも公開されているから改良も自由である。日本語処理の発展のためにも自由に使える日本語資源の一層の充実に皆さんの協力をお願いする次第である。

参考文献

- [1] 野崎昭弘 「GNU宣言」
Computer Today 1988年1月号, サイエンス社
- [2] 桜川貴司 「開かれた日本語入力システム Wnn」
ビット 1987年9月号, 共立出版
- [3] 佐野稔 「フリーソフトウェア」
Computer Today 1988年1月号, サイエンス社

本 PDF ファイルは 1988 年発行の「第 29 回プログラミング・シンポジウム報告集」をスキャンし、項目ごとに整理して、情報処理学会電子図書館「情報学広場」に掲載するものです。

この出版物は情報処理学会への著作権譲渡がなされていませんが、情報処理学会公式 Web サイトに、下記「過去のプログラミング・シンポジウム報告集の利用許諾について」を掲載し、権利者の検索をおこないました。そのうえで同意をいただいたもの、お申し出のなかったものを掲載しています。

https://www.ipsj.or.jp/topics/Past_reports.html

過去のプログラミング・シンポジウム報告集の利用許諾について

情報処理学会発行の出版物著作権は平成 12 年から情報処理学会著作権規程に従い、学会に帰属することになっています。

プログラミング・シンポジウムの報告集は、情報処理学会と設立の事情が異なるため、この改訂がシンポジウム内部で徹底しておらず、情報処理学会の他の出版物が情報学広場 (=情報処理学会電子図書館) で公開されているにも拘らず、古い報告集には公開されていないものが少からずありました。

プログラミング・シンポジウムは昭和 59 年に情報処理学会の一部門になりましたが、それ以前の報告集も含め、この度学会の他の出版物と同様の扱いにしたいと考えます。過去のすべての報告集の論文について、著作権者（論文を執筆された故人の相続人）を探し出して利用許諾に関する同意を頂くことは困難ですので、一定期間の権利者搜索の努力をしたうえで、著作権者が見つからない場合も論文を情報学広場に掲載させていただきたいと思います。その後、著作権者が発見され、情報学広場への掲載の継続に同意が得られなかった場合には、当該論文については、掲載を停止致します。

この措置にご意見のある方は、プログラミング・シンポジウムの辻尚史運営委員長 (tsuji@math.s.chiba-u.ac.jp) までお申し出ください。

加えて、著作権者について情報をお持ちの方は事務局まで情報をお寄せくださいますようお願い申し上げます。

期間： 2020 年 12 月 18 日 ~ 2021 年 3 月 19 日

掲載日： 2020 年 12 月 18 日

プログラミング・シンポジウム委員会

情報処理学会著作権規程

<https://www.ipsj.or.jp/copyright/ronbun/copyright.html>