

インターネットを利用した外国人・小学生のための漢字学習 支援システムの開発

龍岡亮二 吉村ミツ 前田和昭

中部大学経営情報学科

〒487 愛知県春日井市松本町1200番地

tatsuoka@isc.chubu.ac.jp

yosimura@solan.chubu.ac.jp

kaz@solan.chubu.ac.jp

あらまし 日本語を読み書きする能力は、小学校低学年での漢字習得の度合、とくに、筆順を反復練習する方法での漢字習得の度合に影響されると言われている。ところが、限られた小学校の授業時間の中では、個人差が大きい反復練習の時間を十分に取る事ができないのが現状である。外国人(留学生)の場合も、この点では同じ状況にある。

そこで我々は、留学生と小学生を対象にして、個人単位で漢字の筆順と形を反復学習できるシステムを開発している。このシステムでは、グラフィカルユーザインタフェースによる使いやすさ、インターネットからのアクセス、プラットフォーム(PCやワークステーション)非依存の3点を考慮して設計を行なった。本稿では、この漢字学習支援システムの現状について報告する。

Development of a Kanji Learning System for Elementary School Children and Foreign Students through Use of Internet

Ryoji Tatsuoka, Mitsu Yoshimura and Kazuaki Maeda

College of Business Administration and Information Science, Chubu University

Matsumoto-cho 1200, Kasugai, Aichi, 487, Japan

tatsuoka@isc.chubu.ac.jp

yosimura@solan.chubu.ac.jp

kaz@solan.chubu.ac.jp

Abstract The ability to read and write in Japanese depends on the knowledge level of Kanji characters, especially on the practical knowledge of their stroke orders. By means of repeated hard practice to draw a Kanji character, students learn to follow its correct stroke order. To reduce their hard work, we have developed a Kanji learning system for foreign students and for Japanese elementary school children at an initial step of Kanji education. Our design policies are an easy-to-use graphical user interface, use of Internet and platform (PC and Workstation) independence. This paper discusses the current status of our project.

1 はじめに

欧米外国人や小学生を対象にした漢字教育システムの研究は、様々な形で報告されており^{1, 2, 3)}、反復による個人練習を支援するソフトウェアの開発も行なわれている。

一方、国語教育の現場では、小学生における漢字の習得度が読解にも大きな影響があるといわれているにも拘らず、漢字学習には個人差が大きく習得に時間がかかるという困難を抱えている。

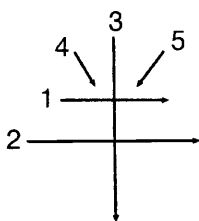


図 1: 留学生による「半」の字の筆順例
図中の数字は書き順を意味し、矢印はストロークの向きを表す。

また、欧米の外国人留学生による漢字の習得も同様な問題を抱えている。例えば、図1のように「半」の字で筆順が全くことなるものや、「出」という字の場合、先に「山」を学習したため、「山」を二つ重ねた書き方をする。すなわち、すでに学習したパターンによって誤りを生じさせるケースである。この場合、二つの「山」が完全に接触した形でなければ、読み手には違和感を与えよう。

外国人留学生と小学生では、幾分違いはあると思われるが、筆順を含めた漢字の形を正しく理解することは、漢字学習上重要な点であり、著しく反復が要求されるものである。

我々は、ゲーム感覚で、興味を持って学習できる形のグラフィカルユーザーインターフェースを持った支援システムの開発を目指している。

このような目的を持ったシステムモデルを検討するため、X-window上で作動するプロトタイプ—XKanji—が作成され³⁾、漢字教育研究の関係者からシステム構成について意見ならびに評価を既に得ている。

ここでは、XKanjiを拡張し、実際の教育現場に適用できるシステム—JKanji—について検討を行なう。

2 システム構成

今回、システムを拡張する際にとった主な指針は次の点である。

1. プラットフォーム非依存性
ワークステーションおよびパソコン (Windows95, MacOS) のいずれの機械にも依存しないバイナリ互換であること。
2. ネットワークの利用
データへのアクセスおよび Netscape 等の WWW ブラウザでの利用を考慮する。

以上の指針を満足させるため、Java 言語でシステムを記述した。このシステムを JKanji と呼ぶことにする。Applet の形式で用いることも、Application で構成することも可能である。

2.1 本体構成

JKanji 本体の外観は、図2に示すように、「評価結果表示部 (EvalView)」「入力部 (Writer)」「漢字パターン表示部 (ShowWindow)」「漢字選択部 (Selector)」および「環境設定部 (Controller)」の5部分から構成される。

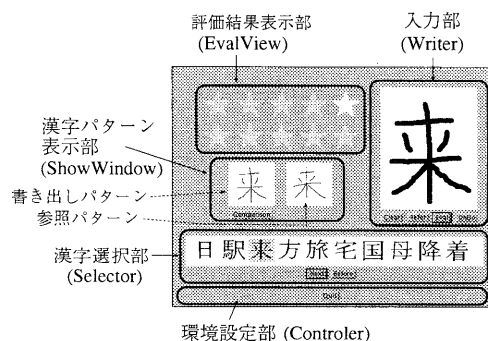


図 2: JKanji 本体外観

学習者が「来」という文字をマウスで入力し、評価を行なわせた状態を示している。

学習者は、次の手順で漢字を学習する。

1. 漢字選択部より、学習しようとする漢字をマウスを使って選ぶ。
この時、漢字パターン表示部の右側ウィンドウに“参照パターン”が示される。

2. 参照パターンを参考に入力部で漢字を書く。

漢字入力には現在のところマウスで行なう。書き終わった時点で入力部の該当ボタンをクリックすれば、学習者が書いた“書き出しパターン”と参照パターン間の筆順と形に関する評価が行なわれる。この時、書き出しパターンは漢字パターン表示部の左側ウィンドウに表示され、参照パターンとの比較が行なえる。

また、必要に応じて入力部の背景に参照パターンを表示させることができ、それをなぞる形式の学習も可能である。もちろん、ウィンドウのクリア、直前のストロークの無効化(Undo)も可能である。

3. 成績の確認。

評価結果の善し悪しに応じて5段階評価で得点化し、評価結果表示部の星印を点灯させる。星印の並びが二段になっているが、一つは筆順を考慮した「形」の類似度で、他は筆順の正誤に関する情報で得点化したものを示している。評価の方法に関しては後に述べる。

加えて、学習者が筆順を間違えたストローク部分は漢字パターン表示部の参照パターンに赤色で表示される。

4. 筆順の比較。

漢字パターン表示部のウィンドウをクリックすれば、参照パターンまたは書き出しパターンが低速で表示される。また、この部分にあるボタンを押せば、両パターンの低速表示を同期を取った形で行なえる。学習者はこれらの機能を使って、筆順の違いを確認する。

なお、表示速度の変更は環境設定部で行なわせるが、現在この部分をどのような形態にするかは未定である。

このように、対象とする外国人留学生および

小学校低学年の生徒がゲーム感覚で反復練習できるように、結果表示をグラフィックス化し、直観的に到達度が把握できる形に外観を組み立ててある。

また、各部分は基本的に独立しており、参照パターンと入力部で書き出されたパターンで決まる状態に応じて各部の動作が引き起こされる。従って、上記の手順を必ず踏む必要はなく、書き出しパターンの途中結果に対して筆順比較を行なうことも可能であるし、途中で別の漢字に変更することも可能である。

2.2 データ

上述の事情を把握するために、データと各部分の関連が図3に示してある。

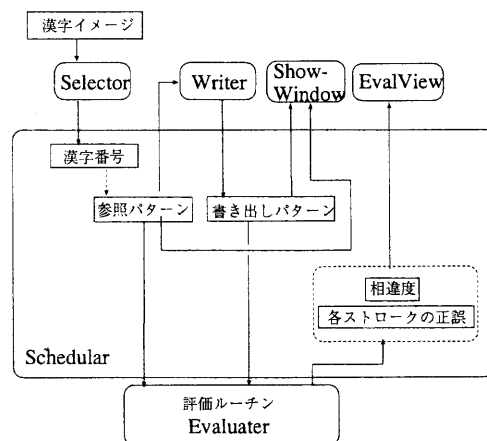


図3: データから見た各部の関連

角の丸い四角で囲んだものは図2に示した構成部品を、単なる四角で囲んだものはデータを意味する。

このシステムで使われるパターンデータはストロークを構成するために必要な時系列の座標列であり、特に図3中の参照パターンは、漢字選択部で選ばれた漢字に割り振られた番号—漢字番号—に対応した座標列データを漢字データベースから読み込んでくる。このデータベースについては3節に詳しく述べる。

さらに、表示させる漢字、即ち学習させたい漢字は、学習者の程度に合わせて、教師が予め

選択して与えられるようになっており、これは表示漢字の漢字番号一覧ファイルをパラメータとしてシステムに与えることで果たせる。また、Applet の場合は<PARAM>タグを用いて漢字番号のリストを与えれば良い。

また、漢字選択部の漢字は GIF 形式ファイルに格納してあるイメージデータを表示している。この理由は外国から本システムにアクセスする場合、日本語のフォントをユーザの計算機が実装しておく必要性を排除するためである*。

JKanji 本体では、例えば漢字選択部で漢字番号が決定した時点での各部へのデータ転送といった、本体各部に跨るデータのやりとりなどの作業工程は Scheduler クラスのオブジェクトで管理しており、各部から送られてくるメッセージデータを解析することで次の工程を実行している。メッセージのやりとりは、Watcher と名付けた Thread によって監視している。

3 データ管理

XKanji では漢字の参照パターンデータをファイルに直接保存していた。システムを拡張する際、XKanji のデータ管理をそのまま踏襲することはできない。その理由は次の 2 点である。

1. Applet だけで JKanji を作ろうとすると、参照パターン、漢字イメージ、ユーザの履歴などのデータの扱いに問題が起きる。それは、Netscape 上で実行される Applet が、局所ファイルにアクセスできないという制約を持つからである。そこで、必要なデータを外部記憶装置上に保持することができなくなり、システムの構築が難しくなる。
2. 将来、漢字学習システムの機能を拡充していくことを考えると、マルチメディア(動画、画像、音声)データの扱いに対処する必要がある。そのような状況ではデータの種類と数が増えることが予想されるため、ファイルを直接操作する方式ではデータ管理が複雑になる。

*その為、JKanji では SJIS などの日本語文字コードセットを直接用いることはしない。

第 1 点目の問題に関しては次の 3.1 節に、第 2 点に関しては 3.2 節に対処法を述べる。

3.1 クライアントとサーバの分離

ユーザインタフェースと漢字の評価を管理するクライアント側(本体)と、データを管理するサーバ側—漢字サーバ—の 2 つに分離した。Netscape 上で実行される Applet は、Applet を獲得したサーバとの間でだけネットワーク入出力を許可されている。従って、関係しているデータをサーバ側に置き、必要な時にネットワーク入出力を使ってデータを受信/送信するようにすれば、上記の問題は解決する。

3.2 データベース

本格的なデータベース管理システムを使って漢字情報(漢字のデータとそれに関連するデータ)を管理する。JKanji では、無料配布されているデータベース管理システム POSTGRES95⁵⁾を採用し、利用している[†]。

現在、データベース管理システムで保持しているデータは、漢字コード表と参照パターンである。

漢字コード表は

漢字番号	漢字コード	漢字イメージ ファイル名(GIF)
------	-------	----------------------

の 3 種類のデータからなる[‡]。

また、参照パターンは、

漢字番号	順序番号	x 座標値	y 座標値
------	------	-------	-------

の 4 種類のデータからなる。順序番号は、参照パターンを構成する座標の時間的順序を示している。これを使えば、特定の漢字の座標列を作り出すことができる。

[†]POSTGRES95 は、Stonebraker らによって開発された拡張関係データベースシステム POSTGRES⁴⁾を土台にして開発されたシステムである。従来の関係データベースの機能を保持しつつ、それを越えるものとしてクラス・継承・型・関数の機能を提供している。

[‡]イメージファイルは現在のところ完備には至っていない。

3.3 JKanji におけるデータアクセス

「クライアントとサーバの分離」、「本格的なデータ管理」の節で述べた事項を考慮した上で、JKanji の構成を図 4 のようにした。

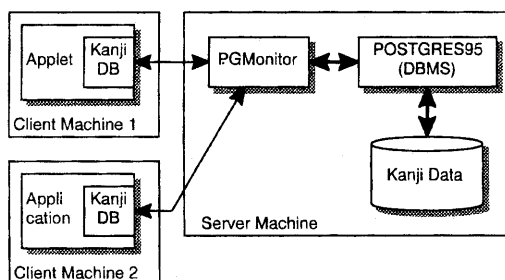


図 4: JKanji の構成

図 4 中の要素は、以下のような機能を提供する。

Applet, Application : 漢字学習システム本体
ユーザインターフェースと漢字評価機能からなる。

KanjiDB : 漢字情報にアクセスするための Java クラス
漢字サーバの存在を意識しなくても、漢字情報を獲得できるようにするための Java のクラス。

PGMonitor : POSTGRES95 にアクセスするためのプログラム
POSTGRES95 が稼働している計算機上で動き、POSTGRES95 に特化した機能を提供する C のプログラム。漢字サーバへのアクセス状況を監視できるようになっている。KanjiDB と PGMonitor の間は、C と Java の socket 機能を使って情報を交換している。

POSTGRES95(DBMS) : 漢字情報を管理するデータベース管理システム
PGMonitor と POSTGRES95 の間は、POSTGRES95 が提供するライブラリ libpq.a を使って情報を交換している。
これらの間の処理の流れは以下のである。

1. Applet または Application から要求がある

と、KanjiDB オブジェクトが PGMonitor へ問い合わせを出す。

2. PGMonitor は、その問い合わせを監視しながら POSTGRES95 へ送り、結果を受け取る。
3. PGMonitor は、KanjiDB オブジェクトへ結果を送る。KanjiDB オブジェクトは、必要に応じてその結果を加工し、Applet または Application へ結果を渡す。

JKanji の機能を充実させるには、漢字サーバで管理する漢字情報を増やす必要がある。今後、マルチメディア (動画, 画像, 音声) の扱いを考慮しながら、漢字サーバが管理する情報および機能を充実していく予定である。

4 評価ルーチン

ここでは、学習者による書き出しパターンが、どの程度正しいかを評価する方法について述べる。なお、評価ルーチンは Thread として実行される。

4.1 評価法

JKanji では、書き出しパターンの筆順や形が参照パターンにどの程度似ているかを、オンライン文字認識の分野で用いられている手法で評価する。その評価の手順は次のようになる。

1. 書出しパターンの位置および大きさの正規化⁷⁾
2. 文字単位の形の評価
 - 2.1) 正規化後の座標系列を $(N-1)$ 分割し、 N 個の座標系列 (N -標本点) に標本化⁷⁾する。
 - 2.2) 正規化、標本化 (N -標本点) された参照パターンと DP マッチング⁶⁾を行い相違度 $dist1$ を求める。
3. 筆順及びストローク単位の形の評価
この評価は、書き出しパターンと参照パターンのストロークの数が等しいときに行う。

- 3.1) 書き出しパターン (ストローク数 s) の各ストロークの座標系列を $(M-1)$ 分割し、 M -標本点ベクトル $(\mathbf{W}^{(\nu)} = (W_i^{(\nu)}), i = 1, \dots, M; \nu = 1, \dots, s)$ を求める。
- 3.2) 標本化 (M -標本点) された参照パターンの μ -番目のストローク $(\mathbf{R}^{(\mu)})$ と $\mathbf{W}^{(\nu)}$ に対し DP マッチングを行ない相違度 $\Delta_{\mu\nu}$ を求める。
- 3.3) 各 μ に対し、最小相違度を示すストローク ($\min(\Delta_{\mu,1}, \dots, \Delta_{\mu,s})$ となるストローク ν^*) を探し、参照パターンのストローク (μ) と同じ ($\mu = \nu^*$) かどうかを調べて結果を $\text{eval}[\mu]$ とする。さらに相違度の和 ($\sum_{\mu} \Delta_{\mu\nu^*}$) を求め、ストロークベースの相違度 dist2 とする。

dist1 , dist2 に基づき形の最終評価とし、 eval に基づき筆順の評価とする。

4.2 文字およびストロークの標本点の数 (N, M) について

字形 (標本点数 N) またはストローク (標本点数 M) は、各々の標本点で近似表現できる度合と、DP マッチングに要する計算時間のトレードオフで決めなければならない。

1. 文字単位の標本点の数 N
文字単位においては次の部分の近似表現に注意すべきである。
 - 1.1) 画数の多い文字の「ごん篇」や「さんずい」などの短いストローク
 - 1.2) 各ストロークの長さ—例えば、「木篇」の各ストローク
 - 1.3) 「しんにょう」の微妙な形の部分
以上、 $(N-1)$ 分割した各部の長さは、評価したい形の最も短い部分よりも短くなければならない。このような配慮が必要な字例を図 5(a),(c) に示す。
2. ストローク単位の標本点の数 M
次のストローク部分の近似表現に注意すべきである。

2.1) 特徴的な屈折点をもつ「聞」、「約」、「風」などの右上角の部分 (図 5(b))

2.2) 長いストローク「しんにょう」の複雑な曲部の形 (図 5(c))

以上、 $(M-1)$ 分割部分の長さは、ストロークの評価したい部分より短くなければならない。

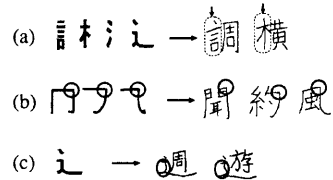


図 5: 問題となる漢字例

以上のように、評価ルーチンのパラメータ (N, M) の最適値は文字によって異なる。現在は一定値が JKanji に設定されているが、今後はこの情報も漢字データベースに記録しておく必要がある。

なお、本研究の一部は文部省科研費・重点領域研究 2-18207223 を使って行なわれた。

参考文献

- 1) Komori, S., *Ann. Symp.*, pp.96-99 (1995).
- 2) 山崎, 井口, 桜井, 電子情報通信学会論文誌 J72-D-II, 9, pp.1493-1500 (1989).
- 3) 龍岡, 吉村, 信学技報 Vol.95, No.96, ET96-36, pp.103-110 (1996).
- 4) Rowe, L. A. and Stonebraker, M., The POSTGRES Data Model, *Proc. of VLDB-87*, Vol. 5, No. 2, pp. 83 - 96 (1987).
- 5) Yu, A. and Chen, J., <http://www.ki.net/postgres95/postgres95-manual/> (1995).
- 6) Sakoe, H. and Chiba, S., *IEEE Trans. Acoustics, Speech and Signal Proc.*, ASSP-26, pp.43-49 (1978).
- 7) Yoshimura, M. et al., *IEICE Trans. Vol.E* 74, No.7 JULY pp.2083-2092 (1991).