

CD-ROM辞書を用いた パーソナル辞書システムの実現

宮丸郁胤 早川栄一 並木美太郎 高橋延匡

東京農工大学工学部電子情報工学科

本報告では、個人用の辞書をCD-ROM辞書の情報を用いながら、作成できるシステムの実現について述べる。計算機を使う機会が増えるにつれて、紙面などで管理していたような個人的な情報を、計算機を用いて管理したいという要求が増えてきている。そこで、単語の情報に注目し、CD-ROM辞書の情報を利用しながら、ユーザ個人の辞書を管理することができるシステムを開発した。このシステムにより、辞書に単語の情報を追加できたり、内容を編集したりすることが可能になった。さらにテキスト編集と辞書の検索システムとの連携を取ることで、辞書の情報を利用できるだけでなく、辞書にテキストの情報を取り込むことも可能になった。

Implementation of a Personal Dictionary System using CD-ROM Dictionaries

Ikutane Miyamaru, Eiichi Hayakawa, Mitarou Namiki and
Nobumasa Takahashi

Dept. of Computer Science,
Tokyo University of Agriculture and Technology

This report describes the implementation of a system that can make a personal dictionary using information from CD-ROM dictionaries. Computerized dictionaries can be used, but usually they are not able to make a personalized dictionary or add notes to the dictionary. This system is based on a virtual word card model, it can make and edit a personal dictionary using word cards. A virtual word card can be made from a CD-ROM dictionary's words, and information can be added to the card. Information from dictionaries can be used or captured in order to link editing a text and looking up a dictionary.

1. はじめに

近年、計算機が普及するにつれて、個人的な用途に計算機を利用することが増えてきた。そこで、個人的な情報を管理することが重要となってきた。

個人的な情報には、メモなどの断片的な情報やまとまった情報など、さまざまな情報があり、それを一つの管理方法で管理することは好ましくない。

我々は、個人的な情報の中で、単語に関する情報に注目し、これを管理するシステムの研究をおこなっている。

個人的な単語の情報を考えた場合に、次の2通りがあると考えられる。

(1) 辞書にない単語の情報

個人で扱う単語などには、辞書に載っていない単語も多い。例えば、計算機に関する専門用語などについては、ほとんど辞書に載っていないことが多い。また、固有名詞などの情報も一般的なものだけしか載っていない。

論文を書くときなどには、個人的に論文で定義して使用している用語などの定義部分を、論文の執筆中に参照して、内容の確認などを行いたい場合がある。そこで、そこれらの用語の定義などを個人的な辞書に入れておき参照したい。

(2) 辞書の単語に対するメモなどの情報

辞書に載っている単語に対して、個人的に例文や意味の補足情報を書き込みたい場合がある。その単語の意味の中でよく使う部分に、マーカで線を引いておいたり、英語の論文を読んでいるときに、自分で英語で書くときの参考になるような言い回しを例文として英和辞典に書き込んでおいたりしたい。

辞書に対して、例文やメモを追加したり、単語を追加したりして、個人用の辞書を作成したいという要求がある。既存のシステムでも、辞書にメモ書きができるものもある [1] が、自分で単語を追加できないなどの問題もある。

本システムでは、(1)、(2) のような単語に関する

個人的な情報を扱うことを目的としている。

2. パーソナル辞書のモデル

2.1 パーソナル辞書のモデル

個人用に用いる辞書を仮想化するに当たって、一般的に個人的に単語を管理するのに用いられる単語カードのモデルを用いて、辞書の仮想化を行う。この単語カードのモデルを用いることで、パーソナルな辞書としてユーザにわかりやすく、一つの単語の情報が一枚のカードに書き込まれるため、単語の情報もわかりやすく管理することができる。そこで、本システムでは、パーソナル辞書を単語カードを元にした「仮想単語カード」というモデルを用いて表現する。

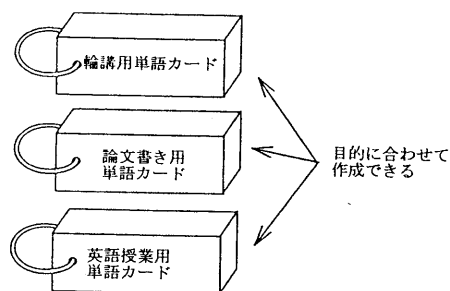


図1 仮想単語カード

2.2 仮想単語カード

単語カードは、単語の情報が書かれるカードと、カードを束ねるリングより構成されている。そこで、仮想単語カードでも、カードとリングを設けて辞書の管理を行う。

(1) リング

リングは、カードを束ねるものである。

単語の情報をユーザが蓄積する際に、単語の使う目的別に管理したいと言う場合がある。そこで、この仮想単語カードのリングには、目的を表現する名前を付けることができ、目的に合わせて作成することができる。

(2) カード

一枚のカードには、一つの単語に関する情報が

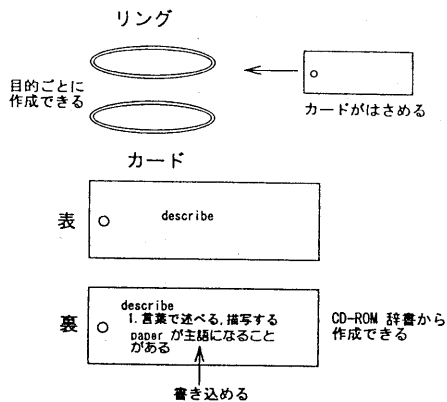


図2 リングとカード

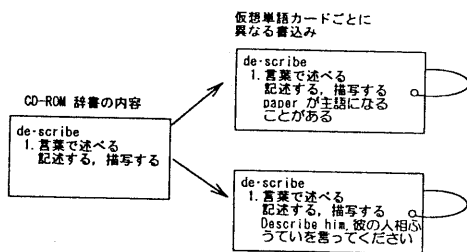


図3 CD-ROM 辞書から作成したカード

書き込まれる。カードには表と裏の情報があり、表にはキーに相当する単語のつづり書き込まれ、カードの検索はこのつづりを用いて検索を行う。カードの裏にはその単語に関する情報が書き込まれる。

カードは、ユーザが作成するだけでなく、CD-ROM 辞書から自動的に作成することが可能である。どちらの場合も同じように、後から情報を追加することが可能である。

異なるリングにより束ねられたカードの中で、同じ CD-ROM 辞書の単語からカードがあり、このカードにユーザが情報を書き込んだ場合、情報の元は同じでもカードとしては異なるので、それぞれの書込みは独立して存在する。つまり、一方のカードへの変更はもう一方のカードには、反映されない。これは、仮想単語カードが違えば、そのカードの目的も異なると考えられるので、カード

に対する書込みも当然異ってくるからである。

3. 設計

3.1 設計方針

次の設計方針で、仮想単語カードモデルを用いたパーソナル辞書システムの設計を行った。

(1) 辞書に対して情報を追加できること

辞書をユーザが作成できるだけでなく、後から自由に情報を追加することができなければ、個人的な辞書の作成を行うことが困難である。そこで、辞書に対して後から情報を追加できることが必要である。

(2) CD-ROM 辞書を用いること

辞書に入れるべき単語をユーザがすべて入力することは、困難である。また、普通の辞書に載っている単語までもユーザが入力しなければならないと、ユーザの手間が増し、辞書の作成が容易ではなくなる。そこで、普通の辞書に載っているような単語は、その辞書の情報を利用することで、ユーザは個人的な情報を入れるだけでよい。ここで用いる辞書として、情報量や種類が多く、個人で入手しやすい辞書である CD-ROM 辞書を用いる。

(3) CD-ROM 辞書の情報をコピーしない

CD-ROM 辞書の単語から、カードを作成する際に、情報のコピーを行いそのコピーを改変するという方法がある。しかし、コピーを行うと、一つの単語の情報だけでも、かなりの情報量を持つため、ディスク容量を多く消費してしまう。また、CD-ROM 辞書の情報のコピーを行うことは、著作権の問題でも好ましくない。そこで、CD-ROM 辞書の情報のコピーを行わずに、CD-ROM 辞書の単語からカードを作成する。

3.2 特徴

本システムは、次の特徴を持つ。

- (1) 目的別の辞書を作成できる
- (2) CD-ROM 辞書の内容も参照できる

3.3 カードの設計

カードの表には、次の情報が格納される。

(1) キー

検索のキーとなる文字列が格納される。辞書の検索では、通常は英和辞典などでは、大文字・小文字の区別がない、アルファベットでのつづり文字列であり、国語辞典や和英辞典などでは、仮名文字の文字列などの情報が格納される。

(2) 表記文字列

検索のキーでは、大文字・小文字の区別や漢字の表記などの情報はない。同じキーが付けられているカードでも漢字などの表記が異なる場合がある。

キーを入力してカードの検索を行った場合、まず検索されたカードの持つ表記文字列が表示される。検索されたカードが複数あった場合は、この表記文字列を元にユーザは見たいカードを指定することができる。

(3) 単語の情報

実際の単語カードで、カードの裏に書かれる情報である。基本的に、一枚のカードにはキーにより指定された、一つの単語の情報が書き込まれる。この情報は、基本的にテキストの情報である。

この単語の情報は、ユーザが始めから作成することができるだけでなく、CD-ROM 辞書の単語から自動的に単語の情報を作成することもできる。

既に作成されている単語の情報に対して、後から情報を追加することができる。

後から追加できる情報は次の通りである。

(a) 文字列

単語の情報のテキストの任意の位置に文字列を挿入することができる。挿入された文字列は、CD-ROM 辞書の情報との区別して管理する。区別することにより、ユーザが書き込んだ文字列であるかどうか分かり、情報が CD-ROM 辞書の確かな情報であるかどうかの区別がつき、ユーザが不確かな情報を確かな情報として参照することを防ぐ。

(b) 文字に対しての属性

既に書かれている文字列に対して、後から属性の情報を追加することができる。属性の情報には、次の情報がある。

・アンダーライン

アンダーラインを引きその部分を強調することができる。情報の中でも重要な部分に、この情報を書き込むことで、その部分が重要であるということが、後で参照したときにすぐにわかる。

・文字列の削除

CD-ROM 辞書の単語から作成されたものであった場合、ユーザにとっては必要のない情報が含まれていることがある。また、ユーザが書き込んだ情報でも、その情報が間違った情報であった場合は、その情報は混乱の元となる危険性があるため、消してしまいたい。この属性は、その部分を表示しないようにすることができる。

この文字列の削除では、実際に情報が削除されてしまうのではなく、文字列に対して削除したという属性を持たせることにより、文字列を表示しないようにしている。このため、後で消した情報を参照したくなった場合には、その情報を見ることが可能である。

3.4 仮想単語カードの管理

ユーザがある作業を開始する場合、その作業用のリングを作成し、そのリングを用いて作業しカードを作成していくことで、その作業用の仮想単語カードが作成できる。

(1) リングの作成

そのリングを持つ仮想単語カードの名前と、その作業で用いる CD-ROM 辞書を指定することでリングを作成することができる。作業を行う際には、作成したリングの中のどのリングを使用するかを選択する。

(2) カードの検索

カードの検索を行う際に、入力されたキーに該当するカードと同時に、使用する CD-ROM 辞書と

して設定されている辞書の単語も検索し、該当するものの表記文字列をユーザに示す。ユーザは、その表記文字列から参照したいカードを選択する。ここで、CD-ROM 辞書の単語が選択された場合は、その単語の情報を自動的にカードとして作成する。次の検索からは、このカードが参照され、その元となった CD-ROM 辞書の単語は検索対象外となる。このような手順で CD-ROM 辞書からカードを作成しているため、参照した単語は必ずカードとして作成される。カードとして作成された情報に対して、ユーザは情報を付加することができる。結果として、ユーザは、CD-ROM 辞書の情報に情報を付加することができる。

3.5 CD-ROM 辞書への書込みの仮想化

CD-ROM 辞書からカードを作成する場合に、コピーを行わない方針であった。そこで、CD-ROM 辞書から作成された、カードへの書込みを仮想化する必要がある。この仮想化は、次のように行っている。

カードを作成する際には、CD-ROM 辞書でのその単語の情報が格納されている、位置の情報をカードに記録しておく。そして、カードの情報を参照した場合には、その位置の情報を元に CD-ROM 辞書からその単語の情報を取り出す。

CD-ROM から作成したカードに対しての情報の追加は、書込みの方法が文字列の挿入や文字列に対するアンダーラインの付加の情報である。そこで、書込みを行った後の状態を保存するのではなく、書込みの手順を保存しておき、図4のように、カードの情報を取り出す際に、手順を元に情報を復元するという方法を取る。これにより、CD-ROM 辞書から作成したカードの内容に少し書き込んだ場合でも、すべてを保存しなくてよいので、ディスク容量の節約になる。

上記の方法により、CD-ROM 辞書から作成されたカードでも、ユーザが作成したカードでもどちらでも、後から情報を追加することができ、さらに CD-ROM 辞書のコピーを行わずに済む。

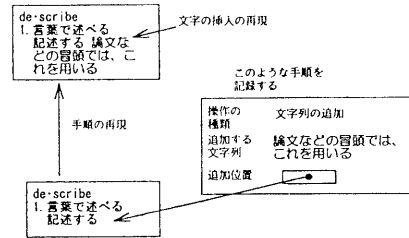


図4 書込みの仮想化

4. 実現

4.1 実現内容

まず、仮想単語カードを管理する部分とそのユーザインタフェース部分の作成を行った。仮想単語カードの部分では、上記の仮想単語カードを管理している。また、そのユーザインタフェースは次のようになっている。

検索したいカードのキーを入力するとそのキーを持つカードの表記文字列の一覧が表示されるので、ユーザはその一覧の中から見たいカードを選ぶ。そこで、カードの内容が表示される。もし、ユーザがそのカードの情報に文字列を書き込みたい場合は、カードの内容が表示されている状態で、文字列書込みの機能を選択し、文字列の挿入位置を指定した後で、書き込みたい文字列を入力することにより文字列の書込みができる。また、アンダーラインや文字列の削除も同様に機能を選んで、カーソルで開始位置と終了位置を指定することでできる。

上記の仮想単語カードの参照・情報の追加を行うものを作成した後、実際に使用した結果、テキストエディタなどでの作業中にカードの参照ができないことから、非常に使いづらさを感じた。そこで、このツールにテキスト編集機能を追加することにした。そのテキスト編集機能で可能なことは、次の通りである。

(1) 文書編集集中のカードの検索

文書中のカードの検索ができることで、文書作成中にカードの情報を参照することができ、作業の効率があがる。カードの情報を参照したい状況

というのは、文書作成中に特に多いと考えられる。そのため、文書作成中にカードの情報を検索できなければ、カードを用いる機会がかなり減るものと考えられる。

(2) カードの内容の辞書への取込み

ユーザがよく用いる情報をカードにしておき、カードの情報から文書に取込むようにすることにより、ユーザの入力の手間を省くだけでなく、そのカードの情報を複数の箇所で見取込む際に、情報の一貫性も取ることができる。

(3) 編集中の文書からカードへの情報の取込み

上記のカードの内容を文書に取込むのとは逆に、文書の内容の一部をカードに書き込む機能である。これにより、文書に書かれている情報をカードに書き込みたいときに、再入力が必要ない。また、編集した文書の一部を例文としてカードに書き込んでおき、後で文書に取込んで使うなどの使い方もできる。

個人でデータベースなどに情報を入力するのは非常に手間がかかるため、個人ではデータベースをあまり使わない場合が多い。そこで、このように既にある情報から、カードに情報を取込むことができれば、個人での情報収集において非常に効果的であると考えられる。

4.2 実現

本システムは、PC-9801 の MS-DOS 上で、C を用いて作成した。規模はソースリストで約 7000 行、うちテキスト編集機能のモジュールは約 300 0 行である。

実現したシステムで、研究社の「英和・和英中辞典電子ブック版」を用いて、辞書を参照しながら、計算機に関する英語の文章を読んでいるときの画面を図5～7に示す。

図5では文書編集中に、その文書中の単語をカードから検索した後の結果である。編集中の文書にある単語を検索する場合には、検索したい単語の先頭の部分にカーソルを移動して、その文字が

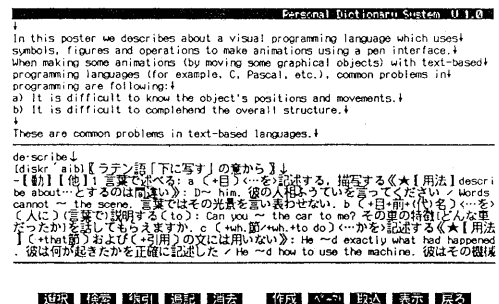


図5 実行画面1

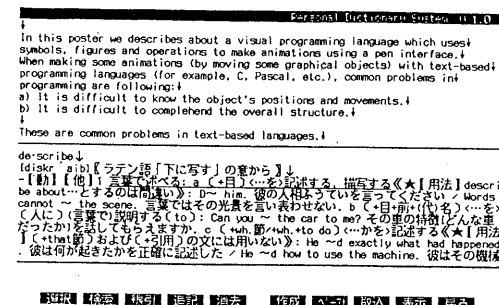


図6 実行画面2

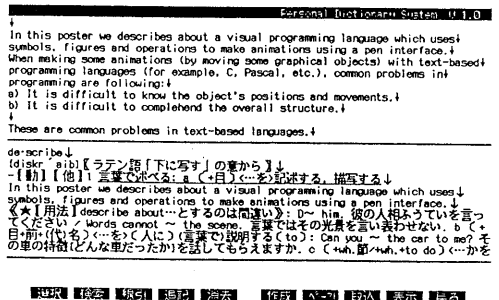


図7 実行画面3

ら次の空白までを単語とみなして切り出すことができる。この切出された文字列または、その文字列に修正を加えて、それをキーとしてカードを検索することができる。この図で検索されたカードは、CD-ROM 辞書にある単語であり、内容を表示した段階では、自動的にカードが作成された後である。

図6では、そのカードに対して、参照した位置にアンダーラインの情報を書き込んだものである。このようにして、書き込まれた情報は消すことが

できないようになっている。これは、個人での情報収集においては、情報の蓄積は重要であるので、間違っで消してしまうことを防ぐためである。その情報を見せたくない場合には、文字列の削除を用いて、通常は表示しないようにすることで可能である。

図7では、編集中の文書の中の英文を例文として、カードに情報を追加した例である。この例のように、編集中の文書に後で有効となる情報が含まれている場合が多い。そのような情報をこのようにテキスト編集と連携を取ることで蓄積することができる。

5. 評価

5.1 評価

本大学の大学院1年次のカリキュラムで輪講があり、学生は自分の研究に関係のある英語の論文を読み、その内容を OHP を用いて、プレゼンテーションを行う。

本システムを、大学院1年次の学生1名に、その輪講の準備のために、普通の英和辞典を引く代りに、使用してもらった。このとき読んだ論文は、十数ページの計算機に関する論文である。輪講の準備にかかった期間は、およそ1か月であった。また、作業の内容は、紙面に印刷された英語の文献を見ながら、本システムを使用して、その文献の日本語訳を本システムのテキスト編集機能で作成するという手順で行っていた。この作業で、作成された辞書のデータを次に示す。

表1 カードの枚数とデータサイズ

作成されたカード数	835 枚
ファイルサイズ	106,496 byte
一枚当りの平均容量	127 byte

評価者に、使用したときに感じた本システムへの要求を、いくつか挙げてもらった。要求の中で、ユーザインタフェースの細かい部分などに関する、

実現よりの要求が多かったが、設計にもかかわってくる要求として次のようなものがあった。

- ・ CD-ROM 辞書の内容が見にくい
- ・ カードに対してリンクをはりたい
- ・ 過去形などを入力して、現在形などの単語を検索したい
- ・ 編集中の文書で、どの位置でカードを検索したかの情報を知りたい

5.2 考察

実際に使用してもらった作業は、本学大学院1年次のカリキュラムに輪講があり、そこで自分の研究に関する英語の論文を読んで、内容について発表しなければならない。そのような状況であったので、ユーザによるカードの作成や、情報の書込みの機能を使用してもらう機会があまりなかった。しかし、CD-ROM 辞書の内容を参照する機会が多いことが、判明した。ユーザ辞書のファイルサイズは、一枚当たり 127 バイトであったが、CD-ROM 辞書の内容のコピーを行っていたとすると、仮に一つの単語の情報を 2K バイトと仮定すると、約 1.7 メガバイトもの容量になってしまう。このことから、CD-ROM 辞書の内容のコピーを行わないことが、正しい選択であったことがわかった。

また、本システムに対する要求に関する考察を次に述べる。

- ・ CD-ROM 辞書の内容が見にくい

これは、使用した CD-ROM 辞書の情報が見にくい構成になっていたことが問題である。しかし、個人的な辞書として考えた場合、このような見にくい情報もユーザが自分でみやすいように変えることができる必要がある。

しかし、ユーザが初めて見る単語まで、いちいち見やすいように変更するのは、余計な手間がかかる。そこで、ある程度自動でインデントするなどの処理を行い、さらにユーザがそのインデントでは不満がある場合には、ユーザが変更できるようにする必要がある。

・編集中の文書で、どの位置でカードを検索したかの情報を知りたい

この問題は、辞書だけではなく、作成した文書との関係である。この要求により、辞書だけの情報だけでなく、他の文書などのデータとの関係についても考える必要があることが判明した。

この評価からわかった、次版での追加すべき機能を次に示す。

(1) 辞書の情報の整理

CD-ROM 辞書に格納されている情報が見にくい場合は、辞書の構造を利用して、自動でインデントなどの処理を行う機能を追加する。さらに、ユーザが自分で情報を見やすいように変更できる機能を追加する。

(2) カードに対してリンクを張れるようにする

カードの情報に別のカードに対するリンクの情報を追加して、そのリンクの張られたカードは、入力して検索を行わずに済むようにする。このリンクの情報は、文字の書込みと同じく、ユーザが自由に情報を追加できるようにする。

(3) カードの検索方法を増やす

現在のバージョンでは、カードの検索は、カードのキーと検索文字列が一致するかどうかだけによる検索である。これに加えて、要求にもあった英単語の原形以外を入力して原形を検索するような検索を追加する。

また、カードの内容の情報に対して文字列検索を行うような検索方法を追加する予定である。

(4) グラフィックなど、文字以外の情報も扱えるようにする

文字列の情報だけでは、説明できない情報もある。そのような情報を無理に文字で表現するのは、わかりにくく、情報の入力の際のユーザの負担を大きくする。そこで、グラフィックなどで表現することでわかりやすい情報は、グラフィックで記述できるようにしたい。そこで次版では、グラフ

ィック情報も取込むことができるようにする。

(5) テキスト編集機能からの情報もカードに登録する

テキスト編集機能があることにより、テキスト編集中の情報をカードに対して取込むことが可能である。現在は、文書を取込むなどしかできないが、どこでカードを参照したかという情報が必要な場面が実際に評価のときにあった。それらの情報をカードの情報として自動的に記録する機構を設ける。

6. おわりに

本報告では、仮想単語カードのモデルによるパーソナル辞書システムの実現について述べた。本システムの実現により次のことができるようになった。

- (1) CD-ROM 辞書を用いて、パーソナルな辞書を作成できるようになった。
- (2) CD-ROM 辞書の内容に対して、情報が追加できた。
- (3) テキスト編集との組み合わせにより、情報の蓄積や利用が可能になった。

今後の課題は次の通りである。

- (1) 本システムでの評価でわかったことをもとに、第二版の設計を行う
- (2) 設計をもとに第二版の実現を行う
- (3) ユーザ辞書を実際に作成する例で実現したシステムの評価を行う

参考文献

- [1] 一太郎 Ver.6.3 付属 JS 辞書検索ツールキットマニュアル, 1995.
- [2] 宮丸郁胤, 他: 文書作成用パーソナル辞書システム, 情報処理学会第 48 回全国大会 2V-6, 1994
- [3] 宮丸郁胤, 他: CD-ROM 辞書を用いたパーソナル辞書システム, 情報処理学会第 50 回全国大会 3N-8, 1995