

3G-1

文書処理ワークステーション上の
英和電子辞書システム(1) - 実現 -

宮内 忠信、岡 満美子、松永 義文
富士ゼロックス(株) システム技術研究所

1. はじめに

既存の冊子体の辞書/事典などをより活用するため、電子化して利用することが一般化している。大量の言語情報を高速に扱うこと自体は、CD-ROM検索システム[1]などにより容易になってきている。ところが、対象が大きくなるほど利用者が本来求める情報を獲得するのが困難なのが現状であり、より柔軟な検索が求められている。特に、ワークステーション上での電子辞書は、その文書処理環境などと統合化することにより種々の電子化の効用が期待できるはずである。

このような背景から、我々はオフィスワークステーションJStar上に英和電子辞書システムを実現した。本稿では、このシステムに関し上記のような問題を解決する機能とその実現について述べる。

2. 基本方針

本電子辞書の実現にあたっては、対象ユーザとしてワークステーション上で作業をする通常のユーザを想定している。今回、原辞書として英和辞書(見出し数約4万)を選んだのは、実際にもっとも利用される辞書の一つであるだけでなく、英文理解や作成といった一般ユーザの作業を支援する実験の素材として適当だからであり、対象とする辞書を限定するものではない。ただし、扱う情報はひとまず文字列のみとする。

利用者やデータを限定せず日常的に使える辞書システムとするためには、次のような機能が必要であると考えた。

- ① ワークステーション上に統合された簡便なユーザインタフェース
- ② 多様な検索キーからの柔軟な検索
- ③ 検索の高速性とこれを実現するデータ構造

3. 機能と実現

3.1 ユーザインタフェース

実現した英和電子辞書システムのイメージを図1に示す。

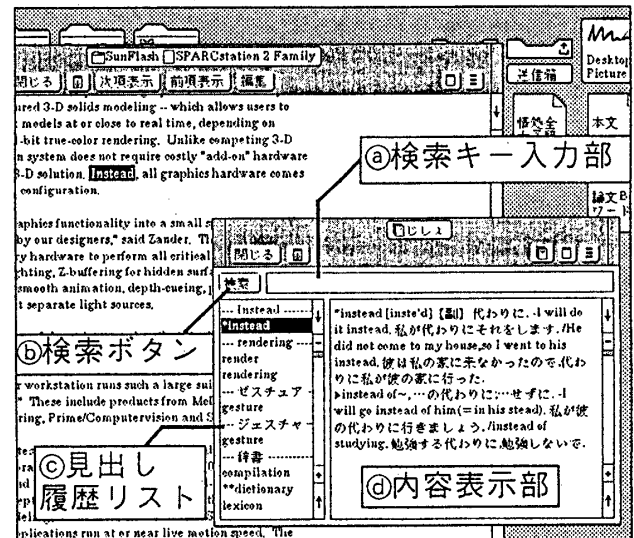


図1. 英和電子辞書システムのイメージ

辞書システムは一つのウィンドウとして実現され、次の4つのエリアを持つ。

- ① 検索キー入力部
- ② 検索ボタン
- ③ 見出し履歴リスト
- ④ 内容表示部

本ユーザインタフェースは次のような利用者の操作から見た検索の流れに沿って設計した。

(1) 検索キー指定

検索キーの指定には、画面上の任意の文字列をマウスでセレクトする方法と、タイプなどにより①に直接検索キーを入力する方法があり、それぞれ読解/文作成などの用途により使い分ける。

(2) 検索指示

検索キーを指定した後に②をマウスでクリックすることにより検索を開始する。特に英文の電子メールを読むような場合は、マウス操作のみで済む。なお、①に直接入力するときは、キーボード上の特定のキーでも検索指示ができるため、この場合キーボードから手を離す煩わしさが無い。

(3)候補選択

検索に成功すると、検索結果の候補の見出しを指定した検索キーとともに◎に表示する。また、同じ単語を何度も検索する手間を省くため、検索履歴をスタック状に保持する。参照したい見出しを選択することで内容が表示されるが、検索時に候補が一つしかない場合は自動的に選択される。

(4)内容参照

選択した見出しに対応する内容は、見出しとともに◎に表示される。表示内容は参照だけでなく自由にセレクトできるため、孫引きや文書への転記が容易である。

上記以外の、例えば選択中の見出しに関する辞書順の前/後単語表示といった指示はウィンドウ上部のプルダウンメニューにまとめ、繁雑さを避けるようにした。また、以上の操作はJStar上のアプリケーションとして一貫性を持つ。

今回実現したのは英和辞書のみであるが、本来辞書は多種多様であるため、各種辞書に統一的に適用可能なユーザインタフェースが要求される。本システムでは、キーワードを指定して候補から望みの検索結果を得るといった検索の流れに沿っており、単純ながら電子辞書の本質と思われる部分のみを抽出したものである。

3.2検索キー

本システムはキーワード検索を基本としており、次に示す多様な検索を実現することにより、利用者の検索意図に応じた柔軟な検索を提供する。

(1)英単語からの検索

見出し語だけでなく、活用形/変化形や大文字を含む表記からでももとの単語が直接検索される。また、完全に一致するものがない場合、冊子体の辞書のように辞書順で前後の近い単語を表示する。

(2)語義からの検索

英和辞書の対訳表記から自動的に語義を抽出し、キーとして登録することで、日本語の単語からの和英辞書的な検索ができる。

(3)カタカナ語からの検索

日本語文においてカタカナ語の占める割合は多い。こうしたカタカナ語からもとの英単語を得る試みはいくつか報告されているが[2]、共通の問題として表記のゆれが存在する。本システムでは全単語に対し表記のゆれを含む発音のカタカナ表記から直接検索することが可能である。

上記(1)~(3)のキーは、内部的には転置キーとして個々に設定されるが、検索キー指定時に利用者はキーの種類による検索モードなどの区別を意識する必要はない。例えば"gesture","身振り","ゼスチュア","ジェスチャー"のいずれからでも単に検索開始を指示するだけで見出し語"gesture"に対応する内容が得られる。なお、検索キーはすべて自動的に抽出したうえで付加しており、総数約20万に及ぶ。

3.3検索速度とデータ構造

対象となるデータ量が巨大になるにもかかわらず、インタラクティブに利用することが必要であるため、検索には高速性が要求される。この点で、CD-ROMベースの電子辞書は検索に平均2~3秒程度を要し、若干問題がある。本システムは、ストレスなく辞書を利用するため、検索開始から内容表示まで1秒程度を目標とした。このため、柔軟な検索に関しては前述のキーワード検索を基本とするものにとどめ、前方/後方一致やワイルドカードなどは実現していない。

このような大量の静的なキーからの高速な検索に適した手法の一つとして、ハッシングが挙げられる。そこで、データ構造としてチェーン付きのハッシュファイルをベースとし、重複を許した転置キーを設定できるものを用意した。このとき、実体データを辞書順に配置することにより前後検索を実現している。

検索速度については、8087-JStar II上で検索指示から候補表示までに目標どおり平均1秒程度のパフォーマンスを得た。なお、データサイズはインデックス情報を含めた合計で約6.5MBである。

4. おわりに

本英和電子辞書システムの実現により、ワークステーション上の統合化環境で電子辞書を利用する基礎ができた。評価については一般の利用者による実際の利用が不可欠であり、[3]で詳細に述べる。

参考文献

- [1] 鈴木,長谷部,喜多山:「広辞苑CD-ROM検索ツールのユーザインタフェース」,情報処理学会第42回全国大会 1L-5 (1991)
- [2] 野美山:「カタカナ外来語の表記の揺れの解消」,情報処理学会第41回全国大会 6S-3 (1990)
- [3] 岡,宮内,松永:「文書処理ワークステーション上の英和電子辞書システム(2) -評価-」,情報処理学会第44回全国大会 3G-2 (1992)