

文科系研究支援のためのコンコーダンスを用いた

文書研究システムの設計と実現

坂口基彦 早川栄一 並木美太郎 高橋延匡  
東京農工大学 工学部

5 B - 5

1 はじめに

近年 Internet を通し古文書の画像などが公開され、それを利用して文献の研究を行なえるようになった。そのため計算機上での文献研究を支援する環境が必要とされている。本稿では、コンコーダンスを用いて文献研究を支援するコンコーダンスシステムの第2版の設計と実現について述べる。

2 コンコーダンスシステム第2版の背景

従来、計算機上でコンコーダンスを扱う場合には、作成に重点がおかれた。しかし実際の文献研究では、コンコーダンスを使って文献を参照する、また明確には定義されていない役割だが、コンコーダンスへの書き込みで研究の成果を残すといったことが非常に重要であると考えた。そこで文字列検索を用いコンコーダンスを自動的に作成し、作成されたコンコーダンスを用い文献を参照する、研究の成果を残すなど一連の過程を支援するコンコーダンスシステムを考案し、初版を MS-DOS 上で実現した。また初版を用いて夏目漱石の『坊っちゃん』、琉球王国外交文書集『歴代宝案』に対し数時間でコンコーダンスを作成した [1]。

一方、文献研究では実際の古文書から得られる筆跡などの情報も重要である。計算機上で画像を用いる場合、テキストで検索し、その結果に対応する画像が表示されれば効率よく筆跡を参照できる。そこで動作環境をウィンドウシステムに移行しテキストに対応する画像が参照できるシステム第2版を設計した。

3 コンコーダンスシステム第2版の概要

3.1 第2版のコンコーダンスの項目の構成

本システムでは、文献中の情報をコンコーダンスで統一して表現するため、従来のコンコーダンスを拡張した。図1に第2版の項目の構成を示す。

3.2 コンコーダンスシステム第2版の画面構成

テキストを表示するテキストビューア、属性、書き込みなどの項目の情報が表示される項目ウィンドウ、KWIC が表示されるコンコーダンスウィンドウ

からなる。それぞれ独立したウィンドウであり、作業内容に応じた配置が可能である。コンコーダンスウィンドウは、大きさを変えることで、KWICの表示長を動的に変えることができる。

(1)項目名 赤シャツ

(2)ふりがな あかしやつ

(3)検索条件 第一候補 赤シャツ と [ ] と [ ]  
第二候補 教頭 と [ ] と [ ]  
第三候補 [ ] と [ ] と [ ]

(4)項目作成型 ●自動作成 ○半自動作成 ○手動作成

(5)属性 人物

(6)関連項目

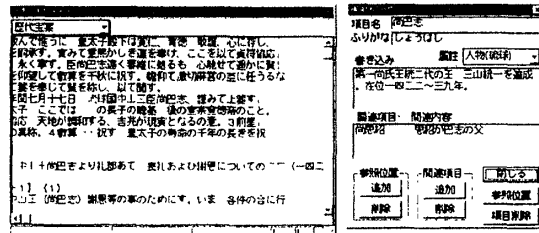
関連項目名	関連内容
文学士	赤シャツは文学士である
オホホホ	赤シャツの笑ひ方

(7)書き込み 坊っちゃんが担任した中学校の教頭。いつも赤シャツを着ているので、坊っちゃんが赤シャツと言うあだ名をつけた

(8)参照位置(KWIC)

文書名	検索構造	行	文脈
坊っちゃん	二	307	だ。しかもそれが赤シャツだから人をばかに
坊っちゃん	二	308	この男は年が年中 赤シャツを着るんだそうだ
坊っちゃん	二	355	校長は理、教頭は赤シャツ。英語の教師はう
坊っちゃん	三	452	でいたから、理も赤シャツも、ちっとも恐ろ

図1 項目の構成



文書名	検索構造	行	文脈
歴代宝案	九	209	中山王 高松 御指図本
歴代宝案	九	213	く 藤、藤 内 巴 吉 遠く 豊 城 に出
歴代宝案	二	222	中山王 高松 より 礼部 あり
歴代宝案	二	225	中山王 (高松) 謝恩等の事
歴代宝案	二	280	中山王 高松 より 礼部 あり

図2 第2版の画面構成

3.3 第2版の作業例

本システムでは、文献を読み進めながらコンコーダンスの項目を逐次作成していく。

作業例を図3に示す。まず(1)テキストビューア上で、項目として作成したい語句を指定する。(2)項目作成ウィンドウが開き、属性、検索条件などを入力する。(3)自動的に参照位置(KWIC)が検索されコンコーダンスウィンドウに表示される。参照したい個所のKWICを選択すると、その個所がテキストビューアに表示される。テキストを参照し、わかったことなどを(4)項目ウィンドウに書き込んで残すことができる。項目が作成された語句は属性の色で表示され、クリックすれば書き込んだ研究成果を参照できる。

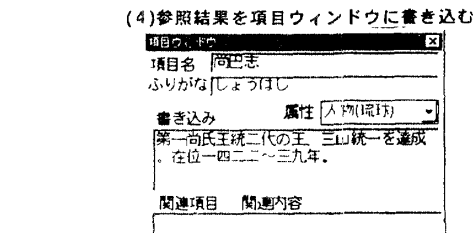
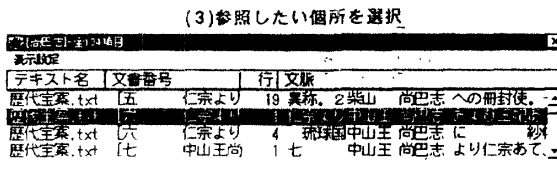
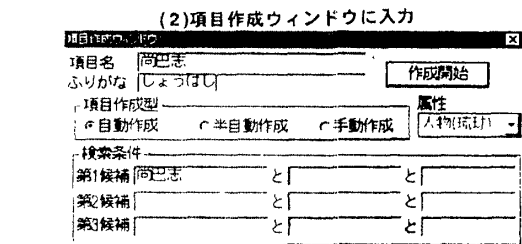
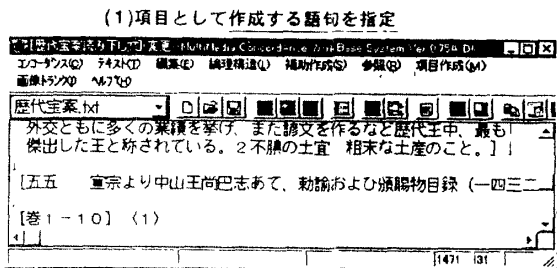


図3 第2版の作業課程の例

#### 4 画像トランクの概要

第2版において、テキストと対応させて画像を表示する画像トランクという画像ビューアを考案した。画像トランクは(1)コンコーダシステムとアプリケーション間通信で動作する、(2)古文書など画像データが大量な場合でも、あらかじめ登録し管理するので容易に参照できる、などの特徴を持つ。大量の画像をトランクにつめておくイメージで管理するので、画像トランクとした。画像トランクは、コンコーダシステムのビューア上で文字がダブルクリックされたとき、コンコーダシステムで参照位置が選択されたときに、メッセージを受け取り対応する画像を更新し、表示する。

##### 4.1 テキストと画像間のリンクの作成

本システムは、計算機の専門家でない研究者が対象なので、直感的にリンク作成を行えることが重要である。またユーザは古文書に精通しているので、一字一字正確リンクを張る必要はなく、対応する大体の個所が表示されればよい。そこで直感的でかつリンク作成の手間を減らす方法として、自動割付けを考案した。文献の画像であれば、基本的には行は

等間隔であると考えられる。そこで、図4のように対応するテキストデータの行数で、ユーザが指定した画像の範囲を均等に割り付ける。表示するときは、その割り付けられた個所(点線の部分)を画面の中心に表示する。この方法を用いれば行がおおよそ均等に並んでいる画像であれば、従来は行単位などであったリンク作成がページ単位でできる。また画像の範囲を指定するとき図4の割付けの枠が表示されるので直感的にリンクを作成できる。

##### 4.2 リンク作成の例

まずテキストビューア上で、画像と対応させるテキストの範囲を指定する。次に図4に示した割付け方法から画像にあったものを選択する。そして画像トランク上で、テキストに対応する画像の範囲を指定する。図5に例を示す。

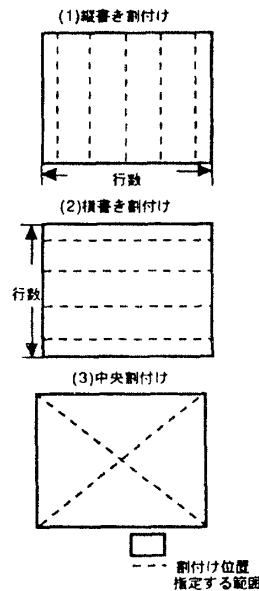


図4 自動割付け方法

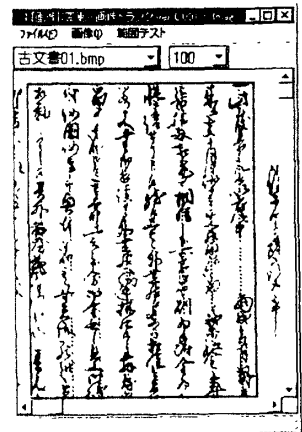


図5 画像リンク作成の例

#### 5 コンコーダシステム第2版の実現

0 コンコーダシステム第2版と画像トランクをWindows95上に実現した。言語Cで書かれ、ソースはそれぞれ約11500行、約4200行である。

#### 6 おわりに

本稿では、コンコーダシステム第2版の設計と実現について述べた。今後の課題として、実際の研究への幅広い応用が挙げられる。

謝辞 本研究は、文部省科学研究費補助金(重点領域研究「沖縄の歴史情報」課題番号06208102)により行なわれた。

##### 参考文献

[1]坂口他、コンコーダシステムを指向したテキストデータベースの研究、情報処理学会、96-CH-29-3,1996