

日本語文章推敲支援ツール『推敲』における応答時間

6B-2

倉田昌典 菅沼 明 牛島和夫

(九州大学 工学部)

1. はじめに

日本語文章推敲支援ツール『推敲』は、日本語文章ファイルを解析して、推敲に役立つ情報を提供するツールである。文章作成のような仕事は個人的な営みであるから、『推敲』の最新の版をパソコン PC-9801 上に実現した¹⁾。このツールの開発にあたっては、実用的な長さの文章（約 1 万字、論文誌刷り上がり 7~8 ページ程度）を、待ち遠しくない時間（数秒以内）で処理できることを目標とした。しかし、パソコン上で『推敲』を実現するのは処理速度の点で不利である。そこで、先の目標が達成されているかどうか、パソコン版『推敲』の処理速度（コマンド毎の応答時間）を測定した。その結果、1 万字程度の入力文章を、一部のコマンドを除いて 1 秒程で処理できることを確認したので報告する。

2. 「推敲」の概要

『推敲』は、文章中の問題点を見つける作業を対話的に支援する。たとえば、受身の使用について検討したいなら、メニュー(図1)から、[A.受身]のコマンドを発する。その結果、文章中で受身の候補が使用されているところが反転表示される(図2)ので、指摘された受身の候補を一つ一つ、画面をスクロールさせながら文章末まで検討してゆけばよい。図3は、コマンド[N.文頭の特徴]の結果である。文章中の文が文頭の文字からソートされているので、文頭の特徴がわかる。図4はコマンド[K.カタカナ表記のゆれ]の結果である。文章中に出現するカタカナ列をソートして、ユニークに表示(上段のウインドウ)、K W I Cで表示(中段)したものである。これは、コマンド[Q.字種毎の文字列表示]でカタカナ列を指定した場合の結果と全く同じものである。カタカナ表記にゆれがある場合は、それらが近くに寄り集まるので、容易に発見できる。そのほか、『推敲』を用いて、図1のマクロコマンドメニューにあるような問題点について文章を見直すことができる。

図1のマクロコマンドは、プリミティブな基本コマンド（主に、検索、ソート、集合演算、表示の4種類がある）を複数組み合わせたものになっている。基本コマンドモードにおいて、それらをうまく組み合わせて使用することによって、マクロコマンドでは扱えないような問題点についても検討することができる。

パソコン版『推敲』は、PC-9801 (RAM 512K 以上) の MS-DOS 上で動作する。処理の対象となる日本語文章ファイルは、shift JIS コードを用いた通常の MS-DOS 上の日本語テキストファイルであり、最大 10 万字のものまで処理できる。『推敲』の詳細については、文献 1, 2, 3) を参照されたい。

3 応答時間の測定

『推敲』では対話的に処理を行うので、キーボードからコマンドを発してから、その結果が画面に表示されるまでの応答時間を測定することにした。この応答時間の測定をマクロコマンドに対して行った（基本コマンドの応答時間はマクロコマンドのそれよりも短い）。測定の対象とした入力文章には、それぞれ文字数の

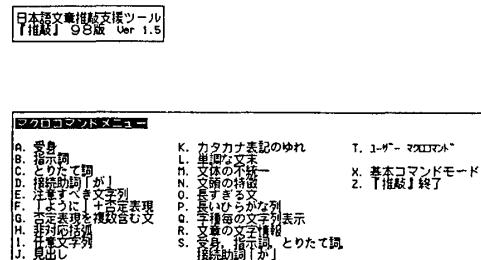


図1 マクロコマンドメニュー

圖2. (A) 爰鳥科的特徵

| No. | 文 | 句 | 文 | 末 | 文No. | 文卡 |
|------|---|---|---|---|------|-----|
| S44: | それを以下示す。 | | | | 513 | 398 |
| S45: | たゞ、TS-WINDOWS上で後に「接続」を実現できのように配慮すれば、必ずしも「接続」の手順は、 | | | | 514 | 398 |
| S46: | たゞ、少し、その手順が、接続する相手のエラーコードによって、多少の違いがある。 | | | | 442 | 398 |
| S47: | たゞ、少し、その手順が、接続する相手のエラーコードによって、多少の違いがある。 | | | | 848 | 398 |
| S48: | たゞ、少し、その手順が、接続する相手のエラーコードによって、多少の違いがある。 | | | | 425 | 398 |
| S49: | たゞ、少し、「コード」の「一一」のように、記号であつて、力加法に含められ、 | | | | 870 | 398 |
| S50: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 592 | 398 |
| S51: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 526 | 398 |
| S52: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 888 | 398 |
| S53: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 112 | 398 |
| S54: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 783 | 398 |
| S55: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 748 | 398 |
| S56: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 113 | 398 |
| S57: | たゞ、少し、その場合、プログラム作成において、ハードウェアに依存する、長さの手順図である文字列で、 | | | | 168 | 398 |
| S58: | つまり、「接続」によって、エラーコードが、文書中の箇所を読みいかに「接続」 | | | | 232 | 398 |
| S59: | つまり、「接続」によって、エラーコードが、文書中の箇所を読みいかに「接続」 | | | | 232 | 398 |

図3 (N.文頭の特徴) の結果

| 文 字 列 在 引 | 列の 長さ | 規則 |
|---------------------------------------|----------|------------------|
| 6 「イチフローピーディスク | 48 | (12) (1) |
| 7 「ワンドラウ | 41 | (5) (6) |
| 8 「エスケープシーケンス | 47 | (18) (1) |
| 9 「エディタ | 45 | (4) (1) |
| 10 「エディター | 45 | (5) (1) |
| 11 「オーブン | 50 | (4) (1) |
| 12 「オブジェクトコード | 51 | (9) (1) |
| 13 「オブジェクト | 52 | (1) (1) |
| 文 字 列 中 引 | 列の 長さ | 規則 |
| 45 ダーに用意されている「ワンドラウ | 93 | 接頭語の接続力が十分 |
| 46 「ワンドラウ | 93 | 接頭語のどちらは一般に |
| 47 だし、画面表示関係の「エスケープシーケンス | 93 | か用意されており、接 |
| 48 はカタカナ表記では「エディタ | 93 | エディタ」のようによく |
| 49 は「エディタ」と「エディター | 93 | 規格化文書の「オブジェクトコード |
| 50 日本語文字列を自由に「オブジェクト | 93 | の「オブジェクト」 |
| 51 「オブジェクト | 93 | がわかるところを客觀的 |
| 52 日本語文字列を自由に「オブジェクト | 93 | な表現としている。 |
| 文 字 列 入 六 方 文 字 内 容 | 列の 長さ | 規則 |
| 176: ABC(記入)によりソートして最後の又脚と共に表したものである。 | 176 | |

つかる。

異なるもの4種類を用意した。応答時間の測定には、打鍵データ収集システムを用いた⁴⁾。これは、PC-9801 MS-DOS 内に常駐して、MS-DOS 上のアプリケーション実行時におけるキー入力、画面更新、ディスクアクセス等のイベント発生時刻を 1/100 秒の精度で記録することができるシステムである。測定は PC-9801 VX21 (CPU i80286、クロック 10MHz) で行った。なお、パソコン版『推敲』は全ての処理を主記憶上で行うので、二次記憶装置の速度から影響を受けない。

処理速度としてコマンドの応答時間の他に、画面スクロールの速さも測定した。『推敲』では、ほとんどのマクロコマンドの結果がスクロールする画面に表示されるため、スクロールの速さもコマンドの応答時間と並んで重要であると考えたからである。スクロールの速さには高速と標準の2種類があるので、それぞれ測定した。

4. 測定結果

コマンド応答時間の測定結果を図5、6に示す。図から、応答時間が入力文章の大きさにはほぼ比例していることがわかる。厳密にいえば、マクロコマンド中でソート処理（クイックソート）を行っているもの（図6）は比例しない。コマンド Q1 は漢字列、Q4 はひらがな列と、ソートの対象となる文字列数が数千と極めて多いので、他のコマンドと比べて応答時間がかなり長くなっている。図から、入力文章が1万字程度であるなら、先の2つのコマンドを除いて、応答時間が1秒程度であることがわかる。また、入力文章が3万字程度の大きさでも、応答時間は数秒である。

画面スクロールの速さは、500 行あたり、6.5 秒（高速時）、および 20.2 秒（標準時）であった。1万字の入力文章は、『推敲』では、400 行程度の長さになるので、文章の始めから終わりまでスクロールさせるのに必要な時間はさらに短い。

5. 字面解析について

『推敲』がパソコン上でも非常に高速な処理を実現している理由は、日本語文章の解析に文法解析や辞書を用いずに、字面解析だけを採用しているからである^{5,6,7)}。

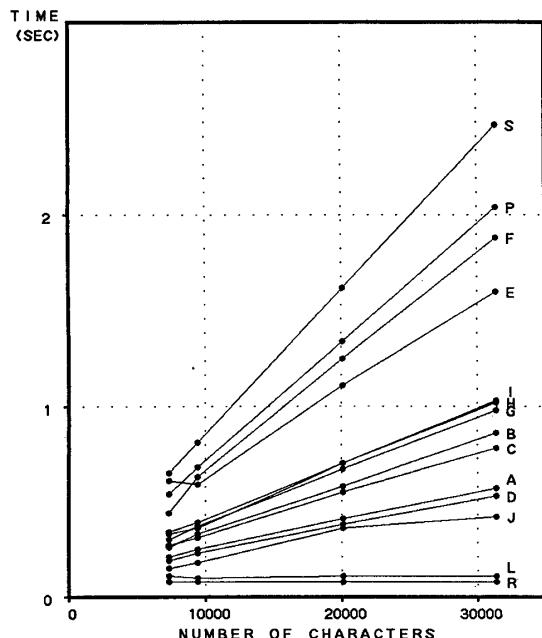


図5 応答時間の測定結果（ソート処理を含まないコマンド）

解析の精度は、文法解析などと比べると、劣るかも知れない。しかし、文法解析をしても結果にあいまいさが生ずるであろう。また、『推敲』の解析結果を利用者が必ず吟味することを前提とすれば、解析誤りのうち、「指摘すべきものを指摘しない」という誤りは許されないが、「指摘すべきでないものを指摘する」という誤りならばある程度許容できる。実際、『推敲』が後者の誤りを犯す率は、一部を除いて 5% 未満である⁸⁾。たとえ、この率が高くても、誤りを含めた指摘件数が入力文章 1 万字あたり 20 くらいまでであれば、十分実用に耐えうると考える。

6. おわりに

『推敲』では、日本語文章の解析にある程度の誤りを許容することによって、字面解析の方法を用いることが可能となり、その結果、極めて高速な処理速度を実現することができた。もちろん、字面解析だけで処理できる範囲はかなり限定されたものであろう。しかし、『推敲』のような対話的な環境においては、応答時間の短さが大事である。

謝辞

応答時間測定の際には打鍵データ収集システムを使用した。このシステムを提供して下さった製品科学研究所の森川治氏、ならびに、その使用を勧めて下さった東工大の木村泉教授に感謝します。

参考文献

- 1) 倉田昌典他：日本語文章推敲支援ツール『推敲』のパーソナルコンピュータでの実用化、情処第35回全国大会講演論文集、1987.1E-3, pp. 2459-2460
- 2) 倉田昌典他：日本語文章推敲支援ツール『推敲』のパソコン上の実現と使用、情処第29回プログラミングシンポジウム報告集、1988, pp. 45-54
- 3) 日本語文章推敲支援ツール『推敲』(Ver. 1.5) 使用手引書、九州大学工学部情報工学科計算機ソフトウェア研究室、1988
- 4) 森川治：時間情報を利用した制御を可能にする MS-DOS の機能拡張について、文書処理とヒューマンインターフェース研究会報告集、1987, No. 14-2
- 5) 牛島和夫他：日本語文章推敲支援ツールにおける受身形の抽出法、情処論文誌、1987, Vol. 28, No. 8, pp. 894-897
- 6) 菅沼明他：日本語文章推敲支援ツールにおける字面解析 I、情処第34回全国大会講演論文集、1987.6X-1, pp. 1347-1348
- 7) 菅沼明他：日本語文章推敲支援ツールにおける字面解析 II、情処第35回全国大会講演論文集、1987.1E-4, pp. 2461-2462
- 8) 菅沼明他：日本語文章推敲支援ツールにおける字面解析の手法の検証、情処第37回全国大会講演論文集、1988

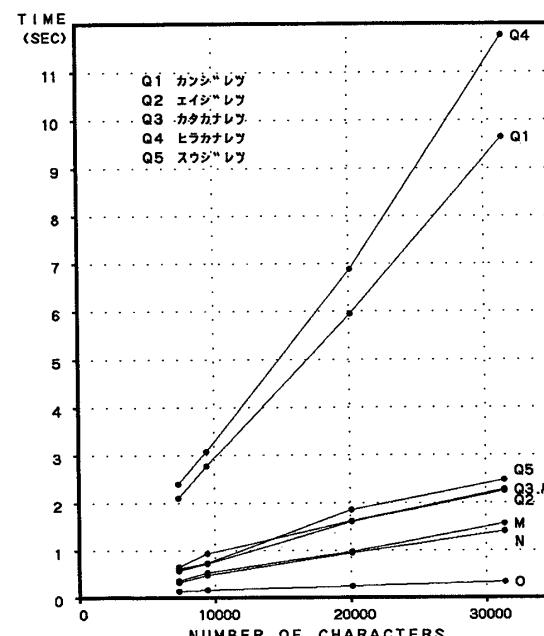


図6 応答時間の測定結果（ソート処理を含むコマンド）