

JAPAN/MARC 書名の KWIC 索引システム

福 島 又 一^{*1} 上 田 修 一^{*2}
中 山 和 彦^{*3}

和図書の検索に用いる索引誌を作成するための日本語 KWIC 索引システムを開発した。このシステムは、漢字かな混り文より辞書や簡単な構文規則を用いて見出し語を抽出し、その見出し語を文字種、画数、部首、五十音等のデータを使って検索しやすい順に配列するものである。処理データは、JAPAN/MARC の書名である。JAPAN/MARC には、書名の読みが分かち書きされて収録されているが、分かち書きが複合語でなく、単語単位に分けられている。このシステムでは、汎用性を持たせたシステムとするため、及び分かち書きが複合語でないために漢字かな混り文を直接処理している。

1. はじめに

図書の検索には、図書カードを用いた著者名、書名、分類による検索方法が古くから行われている。また近年には、計算機を使い文献情報をデータベース化したオンライン検索の方法が開発され、進められている。

索引方法のひとつとして、計算機により一括処理を行って作成する KWIC 索引誌がある。KWIC 索引誌は、オンライン検索に必要な計算機や端末機を用いずに手軽に利用できる利点がある。この KWIC (Key Word In Context) 索引は、英文の索引として広く実用化されている方法であり、文脈より見出し語を抽出し、その見出し語をアルファベット順にソートして、見出し語とともに見出し語の前後の文脈も表示する方法である。

- *1 Mataichi Fukushima 埼玉大学工学部
〒338 浦和市下大久保255
Tel. (0488)52-2111
- *2 Shuichi Ueda 慶応義塾大学文学部
〒108 東京都港区三田2-15-45
Tel. (03)453-3920
- *3 Kazuhiko Nakayama 筑波大学電子情報工学系
〒305 茨城県新治郡桜村
Tel. (0298)53-2369 (原稿受領 1982.11.11)

本報告は、日本語の漢字かな混り文である JAPAN/MARC 書名の KWIC 索引システム及びそのシステムを作成するうえでの日本語処理の問題点について述べたものである。

2. 日本語処理の問題点

KWIC 索引システムを開発していくうえで問題となる日本語について英語と比較しながら述べる。

2.1 文字種

英文での文字種は、アルファベット26文字の大文字、小文字及び記号を含めても数十種しかない。これらの文字には、情報処理面からみても、ISO (International Organization for Standardization, 国際標準化機構) コード、ASC II (American National Standard Code for Information Interchange, 米国国家規格協会) コード、EBCDIC (Extended Binary Coded Decimal Interchange Code) 等のほぼ統一された情報交換用コードがある。

日本語の字種には、平仮名、片仮名、漢字がある。特に漢字の文字数については、正確に調べることは難しいといわれており、現在最も多数の漢字を収録している辞典のひとつである諸橋徹次「大漢和辞典」の収録語数は、49,964字である。ま

た、人名や地名に使われている国字には同辞典にないものもあるので一般に5万字を超えているといわれている。しかし、これらの5万字にもおよぶ文字も利用分野ごとの使用頻度からみると図1に示すように一般に3千種類程度である¹⁾。この図は、新聞²⁾、雑誌³⁾、JAPAN/MARC 書名(1979年分)に使用されている漢字の使用度数累積百分率を示したもので、使用頻度の大きい順に配列した漢字を一定の字数で区切って累積したものである。

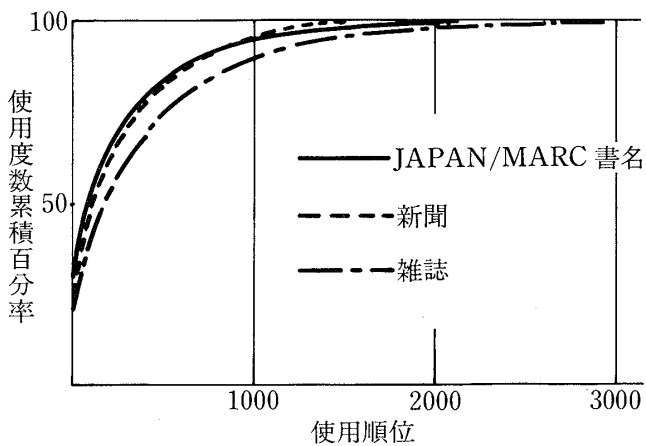


図1 漢字の使用順位と使用度数

これらの漢字を情報処理面では、基本的な漢字2,965字を第1水準漢字集合、また3,384字を第2水準集合としてJIS (Japan Industrial Standards, 日本工業規格) 漢字符号系 JIS C6226 としてコード化されている。しかし、それ以外の文字は、統一されたコード化がなされていない。

2.2 見出し語の抽出

英文の見出し語の抽出は、単語がブランクや特殊記号で区切られる性質を利用して単語を取り出し、その単語を冠詞や前置詞等の見出し語にならない不要語辞書と参照して抽出している。

日本語文の見出し語の抽出は、単語の分割が機械的に行えないため一般に次の方法で行われている。

- 1) 単語辞書や各種単語テーブルを用いる。
- 2) 漢字、片仮名、平仮名等の字種の異なりを性質を用いる。
- 3) 語結合法則や構文規則により解析する。

2.3 見出し語の配列

英語は、一般にアルファベット順に配列される。

日本語の配列方法には、読みの五十音順、部首コード順、画数順等があるが、漢字に同音異字語、同字異音語が多くあるため複雑となり、統一された順序はない。現在一般に使われている順序には、次のようなものがある。

1) 漢和辞典 (例: 角川漢和中辞典)

- a. 親字は、「康熙字典」の部首に従う。

各部首内では、画数順に配列する。同画数内では、その漢字の現代かなづかいによる字音を標準として五十音順に配列してある。

- b. 熟語の見出し順は、字数の少ない熟語から2文字目の字の画数順に従う。同画数のものは、部首順に配列してある。ただし、字数の多いものでも、関連語は基本となる熟語の次に掲げてある。

2) 国語辞典 (例: 新明解国語辞典)

- a. 読みの五十音順による。

同一のかなの中では、清音、濁音、半濁音、また促音、拗音、直音の順に従う。

- b. 同音語の順序は、次のようである。

- 語の性質、構成により記号、造語成分、接辞、単純語、複合語の順。
- 表記により片仮名、平仮名、漢字の順。
- 語の種類により外来語、字音語、和語の順。
- 品詞により助詞、助動詞、感動詞、接続詞、副詞、連体詞、用言、名詞の順。

3) 百科辞典 (例: 平凡社世界大百科辞典の索引編)

- a. 読みの五十音順による。

読み音の促音、拗音も音順に含め、清音、濁音、半濁音の順による。

- b. 音びき(一)は、音順に含めないが、同音になる場合は音びきのあるものを、音びきのないものの後に配列する。

- c. 同音のものは、片仮名、平仮名、漢字の順に配列する。漢字表記のものは、第1字目の画数の少順、第1字目が同音のものは、順次2字目以降の画数の少順に従う。

ドクメンテーション研究 1983.02

4) JIS 漢字コード

- a. 第1水準漢字集合は、各漢字ごとに代表読み1個を定め、その五十音順に配列する。同音のものは、音読みから訓読みの順に、またその内は画数の少順に配列する。
- b. 第2水準漢字集合は、各漢字を康熙字典の214の部首に分類し、その部首内のものを画数順に配列する。また同画数のものは、代表読みの五十音順に配列する。

3. JAPAN/MARC

JAPAN/MARCは、世界的な書誌調整の一環として、日本で発行された図書の全国書誌情報を世界中で相互に交換利用するための機械可読型目録データベースである。データベースは、国立国会図書館に納本された図書に基づき作成されたもので、昭和56年4月より販布が開始されている。データの内容として、書名、著者名、出版事項などの書誌情報を含んでおり、図2にその出力例を示す。

```

001      81001611
020 A    JP
020 B    81001611
100 A    19810106 1980      OJPN 1312
251 A    宮城の自然をたずねて
251 B    野山の植物
251 F    宮城植物の会 編著
270 A    東京
270 B    第一法規出版
270 D    1980.10
275 A    233p
275 B    18×18cm
360 B    190円
551 A    ミヤギノ シゼン オ タズネテ
551 X    Miyagi no sizen o tazunete
551 B    251
658 A    ショクブツ ニホン ミヤギケン
658 X    Syokubutu Nihon Miyagiken
658 B    植物 日本 宮城県
677 A    472.123
685 A    RA255
751 A    ミヤギノ ショクブツ ノ カイ
751 X    Miyagi syokubutu no kai
751 B    宮城植物の会
905 A    RA255-160
    
```

図2 JAPAN/MARC データ例

JAPAN/MARC 書名 KWIC 索引は、このデータベースから、索引項目として書名部、表示項目として著者部、叢書名、分類番号、書誌番号を使用する。これらの項目について、KWIC 索引システムの設計に必要な基礎資料として次の分析を行った。また、ここで使用したデータは、1981年分の約4万9千件である。

3.1 書名部

書名部は、索引項目として書名、副書名、多卷

物の各巻の書名とし、表示項目として巻次、叢書名、叢書番号とした。書名部の文字長分布は、図3に示すように書名の90%が15文字以下であり、また書名部全体を組合せた場合でも90%が30文字以下であった。書名部の組合せ出現割合を表1に示す。表1より、叢書名のあるシリーズもの、書名だけの単行本、また巻次などのある出版物の割合が高いことがわかる。

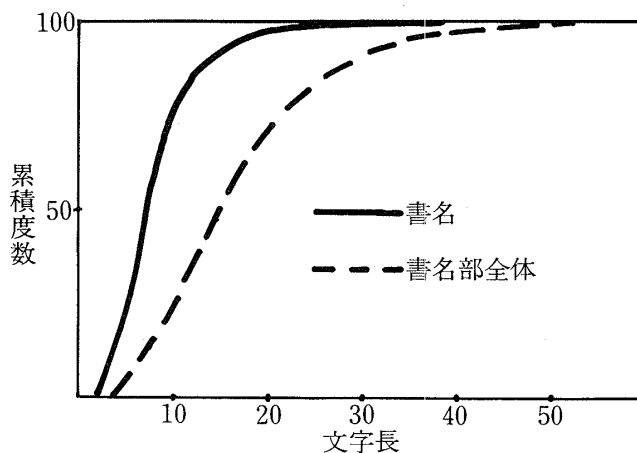


図3 書名部文字長度数分布

表1 書名項目組合せ割合

項目	組合せ						
	○	○	○	○	○	○	○
書名	○	○	○	○	○	○	○
副書名		○		○			
巻次等			○	○			
多巻物の各巻の書名						○	○
叢書名					○		○
組合せ出現割合	25%	16%	20%	3%	26%	9%	1%

表2 出現頻度の多い叢書名

順位	叢書名	順位	叢書名
1	角川文庫	7	新潮文庫
2	講談社文庫	8	ハヤカワ文庫
3	集英社文庫	9	中公文庫
4	ふるさと文庫	10	少年チャンピオン・コミックス
5	徳間文庫	11	少年サンデー・コミックス
6	文春文庫	12	農林水産統計報告

叢書名は、全体の27% (約1万3千件) である。叢書名の種類は、3千6百種に及び、その出現件数の多い叢書名を表2に示す。

3.2 著者部

著者部は、著者、多巻ものの各巻の著者、翻訳書の原著者を集めている。各著者項目には、著、共著、作、文、画、撮影、作曲、編、訳等の役割表示が付けられている。この著者部の文字長分布を図4に、また組合せ割合を表3に示す。著者部は、収録データの84%に出現し、そのうち88%が書名に対する著者である。

3.3 全国書誌番号

番号は、日本全国書誌週刊版への収録順に全データに付与する8桁の一連番号である。番号の冒頭2桁は西暦年の末尾2桁を表わしている。

3.4 NDC

数値は、日本十進分類の分類標数を示す。NDCは、収録データの82%に付けられており、その文字長を図5に示す。

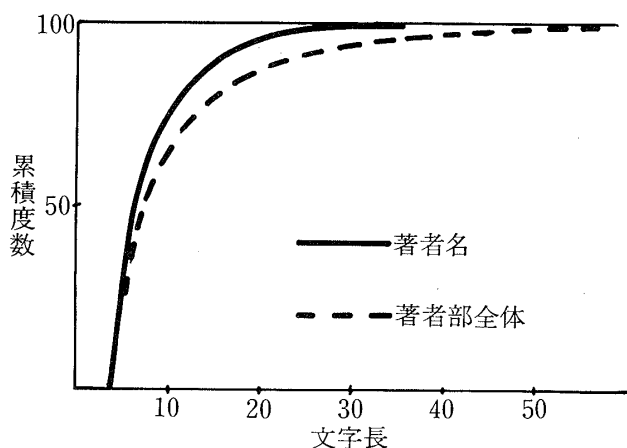


図4 著者部文字長度数分布

表3 著者項目組合せ割合

項目	組合せ		
著者	○	○	
多巻物の各巻の著者			○
翻訳書の原著者		○	
組合せ出現割合	88%	9%	1%

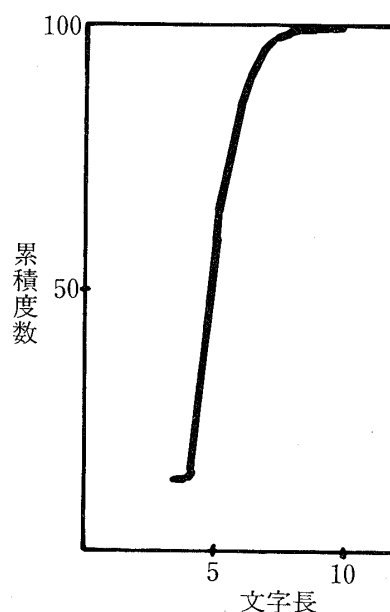


図5 NDC文字長度数分布

4. KWIC 索引システム

KWIC 索引システムの流れ図を図6に示す。このシステムは、見出し語の抽出に各種辞書を使用していること、見出し語の配列に読み、画数及び部首コードを使用していること、また出力媒体として冊子体、COMが選択できる点に特色がある。

4.1 見出し語の抽出

見出し語の抽出は、各種辞書の参照、語結合法則の応用、字種の異なりの性質等により、次の方法で処理する。

1) 削除語辞書による処理

文字列を削除語辞書と比較して等しければ削除する。この削除語辞書は、索引語として不用な語(不用語)及びJAPAN/MARC書名に出現頻度の高い語を収録している。辞書は、追加、変更が容易なように設計した、漢字、平仮名、アルファベットの3種類より成っている。

現在の収録語は、漢字では「概要」、「調査」、「研究」、「報告」等の41語、平仮名として「あゆみ」、「おける」、「ために」等の16語、またアルファベットとしてa, for, in, of等の13語を収録している。

例

学校基本調査結果速報
進路決定のために
The shinkansen

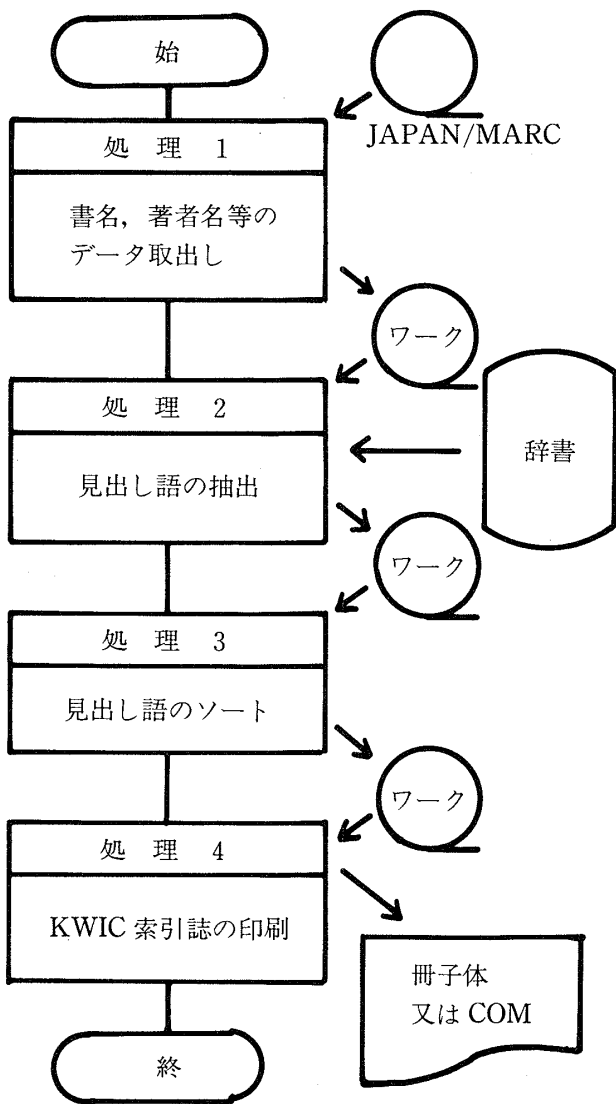


図6 KWIC 索引システム流れ図

~~~~部が削除語を示す。

2) 個有語辞書による処理

文字列を個有語辞書と比較して等しければ見出し語とする。辞書には、書名データより抽出した人名や地名などの個有名詞、例えば札幌、筑波、空海、杜甫等の622語を収録している。また、この辞書は、単語の追加、変更が容易なように設計されている。

例

京都府歳入歳出決算書

~~~~部が個有語を示す。

3) 単語辞書による処理

文字列を単語辞書と比較して等しければ見出し語とする。辞書は、新明解国語辞典の漢字部から

2文字以上の単語を7文字までの文字長で抽出編集した約5万5千語を収録している。また新明解国語辞典の漢字部と読み部の文字長を図7に示す。

例

動物の不思議な感覚

~~~~部が単語を示す。

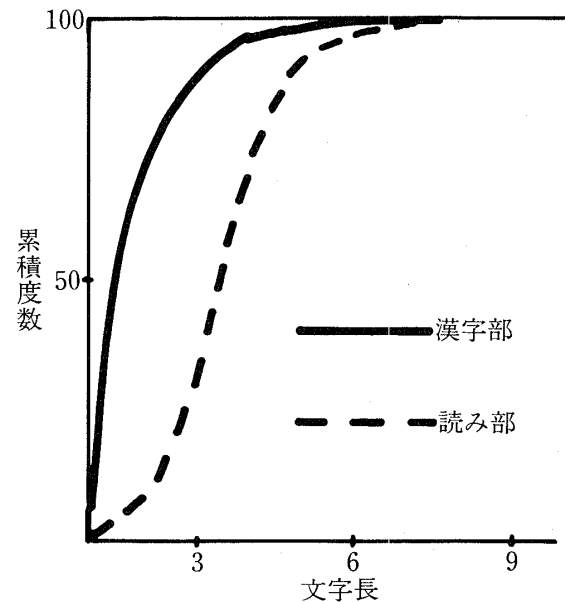


図7 新明解国語辞典読み部、漢字部文字長分布

4) 片仮名文字列の処理

片仮名文字列は、書名の先頭から始まる文字列、及び書名中で2文字以上の文字列の場合を見出し語にする。

例

誰でもできるムギづくり

アイソトープの工業利用

~~~~部が片仮名見出し語を示す。

5) 平仮名文字列の処理

平仮名文字列は、書名の先頭から始まる2文字以上の文字列、及び平仮名接続詞(の、と、は、を、も)に続く2文字以上の文字列を見出し語とする。

例

幸福のためいき

ペロニカはにんきもの

~~~~部が平仮名見出し語を示す。

6) アルファベット文字列の処理

アルファベット文字列は、先に述べた削除語辞書、及び1文字長の語を除き、他の文字列を見出し語にする。

例

The Wired Society

マイコンのための FDOS

~~~~部がアルファベット見出し語を示す。

7) 数字及び記号文字列の処理

数字及び記号文字列は、全て削除語とする。

例

買いたい心理 110 カ条

西暦2000年の地球

~~~~部が削除語を示す。

#### 8) 接辞語による処理

ここでは、上記段階で処理できない漢字文字列を不明語とよぶことにする。また以下で使用する接頭接尾語は、新聞の語彙調査で他の語と結合する1字漢語の使用頻度の多い語<sup>2),5)</sup>を参考に収録した。

a. 不明語が1文字で、その次の文字列が見出し語または削除語であれば、その不明語を接頭語テーブルと比較して処理する。現在テーブルには、同、大、新、無等の12語を収録している。

例

新民事執行法の解説

~~~~部が接頭語を示す。

b. 不明語の前が見出し語または削除語の場合には、接尾語テーブルと比較して処理する。このテーブルには、的、者、学、法等の12語を収録している。

例

わかりやすい労働法常識

~~~~部が接尾語を示す。

c. 不明語が2文字以上からなり、その次の文字が地名用接尾語であれば、不明語を見出し語とする。このテーブルには、県、市、郡、町、村の5語を収録している。

例

大和市文化財調査報告書

~~~~部が地名用接尾語を示す。

d. 不明語が2文字以上からなり、その次に見出

し語または削除語がある場合には、その不明語を見出し語とする。

例

犬塚古墳群発掘調査報告書

~~~~部が見出し語を示す。

上記c, dで抽出した見出し語の読みは、次章で述べる漢字コード辞書により代表読みを決めた。

#### 4.2 見出し語の配列

この索引誌では、検索しやすい配列方法として、見出し語の読み順、部首コード順、画数順、漢字コード順、またそれからの併用等を試行及び検討の結果、次の配列方法を用いた。また配列処理に使用した部首コード等のデータは、JIS C6226の第1, 第2水準の6,349字について部首コード、画数、音読み、訓読みを収録した漢字コード辞書を作成した。

##### 1) 見出し語の1文字目による分類

- 字種別として、アルファベット文字を索引誌の先頭に集中させアルファベット順にし、そのあとに日本語を配列した。
- 日本語の読み別として、読みの先頭1文字目の五十音順とした。
- 同音内の順序は、平仮名、片仮名、漢字の順とする。また同音の漢字については、画数の昇順。同画数内では、部首コード順。同部首内では、JISC 6226のコード順とした。

##### 2) 見出し語の2文字目以降による分類

見出し語の2文字目以後から6文字目、書名の先頭文字の順で分類する。各桁での順序は、記号(JIS漢字コードの昇順)、アルファベット順、平仮名の五十音順、片仮名の五十音順、漢字の場合はJIS漢字コードの昇順とする。

##### 3) 全国書誌番号による分類

8桁の個有な数字である書誌番号の昇順とする。

#### 4.3 索引誌の出力形式

索引誌の印刷基本形式は、JAPAN/MARCデータの特性から、書名、著者名、NDC、全国書誌番号を次のように表示した。

ドクメンテーション研究 1983.02

## 1) 書名 (52文字分)

— ◎ — \* — ☆ — § — \* —

書名 副書名 巻次 多巻物書名 叢書名 巻次

## 2) 著者名 (26文字分)

—— ☆ —— ☆ ——

著者項目 著者項目 著者項目

## 3) 全国書誌番号 (8文字分)

## 4) NDC (6文字分)

索引誌は、A4判の大きさに1頁53行の見出し語を表示するように設計した。また大量出力用としてCOM(Computer Output on Microfilm)での出力ができる。COMは、1枚に索引誌を325(13行×25列)頁分収容するので有効である。KWIC索引誌の出力例を図8に示す。

## 5. JAPAN/MARC 書名の処理結果

使用データは、JAPAN/MARCの1979年分約3万4千件、1980年分約4万5千件、1981年分約4万9千件である。見出し語を抽出する書名部は、1冊の本に対し書名、副書名、多巻物の各巻の書名を対象としたので総計17万8千件のデータ量になり、このデータを年度単位で処理した。

処理結果は、3年間とも次に示すような傾向の比率であった。見出し語の抽出率は、書名当り約2.5件、また一冊当り約3.5件であった。見出し語の抽出方法別では、単語辞書を利用したもの約69%、個有語辞書を利用したもの約7%、語結合の関係から抽出したもの約11%、片仮名文字列のもの約9%、平仮名文字列のもの約3%、アルファベット文字列のもの0.6~2%であった。また削除語については、書名当り約0.3件発生した。この数値から、個有語辞書の役割は、先に述べたように現在の収録語数622語で7%の抽出になり、その役割の大きさがわかる。また、KWIC索引誌の頁数では、1年分で約3,000頁になり、COMに出力すると約10枚になる。

このシステムの精度を測定するため500冊分のデータをランダムに抽出して処理した結果を分析した。見出し語の抽出については、1,808件の抽出に対し、次のような誤りがあった。

○単語切り出しの誤り(7件、約0.4%)

例

兵庫県立近代美術館館蔵品目録

抽出は、兵庫、県立、近代、美術館、品目、とされた。「品目」は、「館蔵品」、「目録」の抽出誤りである。

○個有語辞書に追加したい単語(15件、約0.8%)

例

現代っ子不器用の証明

抽出は、不器用、証明、とされた。「現代っ子」の抽出ができない。「現代」は、削除語になっている。

○1文字単語が未抽出(25件、約1.4%)

例

土地と家のトラブル解決集

抽出は、土地、トラブル、解決集、となる。現在のシステムは、書名の中に現われる1文字の漢字を抽出していない。

見出し語の配列については、読みの誤りによる配列誤りが発生している。

○地名や人名(20件、約1.1%)

例

女川原子力発電所原子炉設置変更許可申請書

辞書に無い見出し語の読みは、基本として、2文字以上の漢字文字列を音読み、1文字漢字を訓読みのように優先する方法を取っている。そのため「女川」をジョガワと誤り、配列が「女性」等と同じ場所に表示される。

○辞書に無収録語(9件、0.5%)

例

油絵の新技法

上記と同様に「油絵」をユエと誤って読んでいる。

上記の誤りの多くは、個有語辞書の追加により、ある程度まで精度を上げることができるが、根本的には今後の課題となる。

## 6. おわりに

このKWIC索引システムは、JAPAN/MARCデータの処理を第一として開発したシステムであるが、汎用性を持たせるように留意している。そのため、他の図書館や資料センター等で保有する文献データベース等にも使用できると思われる。

| 著者名    | 書名        | ( 著者名 )                | ( 書誌番号 ) (NDC) | 頁         |
|--------|-----------|------------------------|----------------|-----------|
| 農林登録候補 | あらしのクリクリ谷 | 桐原三好著                  | 81007590       | 16. 3     |
| 農林登録候補 | あらしのクリクリ谷 | 佐藤テル, 大河原由紀子著          | 81009257       | 29. 2, 58 |
| 農林登録候補 | あらしのクリクリ谷 | 沖井千代子作                 | 81007964       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 沖井千代子作                 | 81003847       | 47        |
| 農林登録候補 | あらしのクリクリ谷 | ウィルバート・オードリ            | 81011944       | 491. 65   |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81011978       | 14        |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000579       | 375. 33   |
| 農林登録候補 | あらしのクリクリ谷 | 東京都中学校文芸教育研究会編         | 81001565       | 375. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 兼田正治著                  | 81007420       | 375. 19   |
| 農林登録候補 | あらしのクリクリ谷 | 兼田正治著                  | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 竹田晃著                   | 81000341       | 583. 7    |
| 農林登録候補 | あらしのクリクリ谷 | 竹田晃著                   | 81002996       | 6         |
| 農林登録候補 | あらしのクリクリ谷 | 森一著                    | 81010754       | 923       |
| 農林登録候補 | あらしのクリクリ谷 | 講談社編                   | 81001633       | 498. 3    |
| 農林登録候補 | あらしのクリクリ谷 | 三浦福雄著                  | 81000749       | 596. 22   |
| 農林登録候補 | あらしのクリクリ谷 | 三浦福雄著                  | 81012148       | 289. 1    |
| 農林登録候補 | あらしのクリクリ谷 | 福岡県中小企業総合指導所編          | 81001293       | 335. 35   |
| 農林登録候補 | あらしのクリクリ谷 | 福岡県中小企業総合指導所編          | 81011622       | 335. 35   |
| 農林登録候補 | あらしのクリクリ谷 | 山川篤著                   | 81015718       | 18        |
| 農林登録候補 | あらしのクリクリ谷 | 山川篤著                   | 81012692       | 950. 2    |
| 農林登録候補 | あらしのクリクリ谷 | 相川正道, 翠川修共著            | 81007896       | 661. 6    |
| 農林登録候補 | あらしのクリクリ谷 | 相川正道, 翠川修共著            | 81005482       | 494. 93   |
| 農林登録候補 | あらしのクリクリ谷 | 中国四国農政局鳥取統計情報事務所吉出張所編集 | 81000205       | 612. 17   |
| 農林登録候補 | あらしのクリクリ谷 | 中国四国農政局鳥取統計情報事務所吉出張所編集 | 81011876       | 612. 17   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81010924       | 402. 91   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81014664       | 402. 91   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81003098       | 9         |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81009903       | 509. 21   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81013814       | 615       |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81015616       | 16        |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81004017       | 382. 13   |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81012862       | 2         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81010992       | 2         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000579       | 375. 33   |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81016296       | 913. 6    |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000102       | 318. 2    |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | Milton H. Donald       | 81013338       | 493. 92   |
| 農林登録候補 | あらしのクリクリ谷 | Milton H. Donald       | 81007862       | 517. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 筒井勉一著                  | 81001805       | 791       |
| 農林登録候補 | あらしのクリクリ谷 | 筒井勉一著                  | 81012488       | 753       |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81003711       | 791. 5    |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81004051       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81013746       | 640. 59   |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81005414       | 370. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81015888       | 210. 1    |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81016228       | 796       |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81016592       | 210. 1    |

図 8 KWIC 索引誌出力例

| 著者名    | 書名        | ( 著者名 )                | ( 書誌番号 ) (NDC) | 頁         |
|--------|-----------|------------------------|----------------|-----------|
| 農林登録候補 | あらしのクリクリ谷 | 桐原三好著                  | 81007590       | 16. 3     |
| 農林登録候補 | あらしのクリクリ谷 | 佐藤テル, 大河原由紀子著          | 81009257       | 29. 2, 58 |
| 農林登録候補 | あらしのクリクリ谷 | 沖井千代子作                 | 81007964       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 沖井千代子作                 | 81003847       | 47        |
| 農林登録候補 | あらしのクリクリ谷 | ウィルバート・オードリ            | 81011944       | 491. 65   |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81011978       | 14        |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000579       | 375. 33   |
| 農林登録候補 | あらしのクリクリ谷 | 東京都中学校文芸教育研究会編         | 81001565       | 375. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 兼田正治著                  | 81007420       | 375. 19   |
| 農林登録候補 | あらしのクリクリ谷 | 兼田正治著                  | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 竹田晃著                   | 81000341       | 583. 7    |
| 農林登録候補 | あらしのクリクリ谷 | 竹田晃著                   | 81002996       | 6         |
| 農林登録候補 | あらしのクリクリ谷 | 森一著                    | 81010754       | 923       |
| 農林登録候補 | あらしのクリクリ谷 | 講談社編                   | 81001633       | 498. 3    |
| 農林登録候補 | あらしのクリクリ谷 | 三浦福雄著                  | 81000749       | 596. 22   |
| 農林登録候補 | あらしのクリクリ谷 | 三浦福雄著                  | 81012148       | 289. 1    |
| 農林登録候補 | あらしのクリクリ谷 | 福岡県中小企業総合指導所編          | 81001293       | 335. 35   |
| 農林登録候補 | あらしのクリクリ谷 | 福岡県中小企業総合指導所編          | 81011622       | 335. 35   |
| 農林登録候補 | あらしのクリクリ谷 | 山川篤著                   | 81015718       | 18        |
| 農林登録候補 | あらしのクリクリ谷 | 山川篤著                   | 81012692       | 950. 2    |
| 農林登録候補 | あらしのクリクリ谷 | 相川正道, 翠川修共著            | 81007896       | 661. 6    |
| 農林登録候補 | あらしのクリクリ谷 | 相川正道, 翠川修共著            | 81005482       | 494. 93   |
| 農林登録候補 | あらしのクリクリ谷 | 中国四国農政局鳥取統計情報事務所吉出張所編集 | 81000205       | 612. 17   |
| 農林登録候補 | あらしのクリクリ谷 | 中国四国農政局鳥取統計情報事務所吉出張所編集 | 81011876       | 612. 17   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81010924       | 402. 91   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81014664       | 402. 91   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81003098       | 9         |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81009903       | 509. 21   |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81013814       | 615       |
| 農林登録候補 | あらしのクリクリ谷 | 函館統計情報事務所編集            | 81015616       | 16        |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81004017       | 382. 13   |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81012862       | 2         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81010992       | 2         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 群馬県教育委員会編              | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000579       | 375. 33   |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81016296       | 913. 6    |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81000102       | 318. 2    |
| 農林登録候補 | あらしのクリクリ谷 | 大谷猛夫著                  | 81009971       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | Milton H. Donald       | 81013338       | 493. 92   |
| 農林登録候補 | あらしのクリクリ谷 | Milton H. Donald       | 81007862       | 517. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 筒井勉一著                  | 81001805       | 791       |
| 農林登録候補 | あらしのクリクリ谷 | 筒井勉一著                  | 81012488       | 753       |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81003711       | 791. 5    |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81004051       | 1         |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81013746       | 640. 59   |
| 農林登録候補 | あらしのクリクリ谷 | 古賀健蔵著                  | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81005244       | 4         |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81005414       | 370. 8    |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81015888       | 210. 1    |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81016228       | 796       |
| 農林登録候補 | あらしのクリクリ谷 | 和歌森太郎著作集刊行委員会編         | 81016592       | 210. 1    |



ドクメンテーション研究 1983.02

しかし、この KWIC 索引システムは、現在急速に研究が進められている日本語処理の応用システムのため、十分完成されたシステムとは言えない。今後も、日本語文の自動分かち書き、日本語ソー  
ト等の研究を通じて KWIC 索引システムの精度を、より向上させるつもりである。

### 参 考 文 献

- 1) 林大. “図説日本語”. 東京, 角川書店, 1982.02, p.244.
- 2) “国立国語研究所報告56: 現代新聞の漢字”. 東京, 秀英出版, 1976.
- 3) “国立国語研究所報告22: 現代雑誌九十種の用語用字 第二分冊 漢字表”. 1963.
- 4) 国立国会図書館. “JAPAN/MARC マニュアル”. 1981. 3.
- 5) 林大. “図説日本語”. 東京, 角川書店, 1982. 2, p.108.
- 6) 西村恕彦. “文字列の配列順序についての問題”. 情報処理, **10** (1), 21 (1969).
- 7) 植村俊亮. “漢字かな混り文 KWIC 索引”. 情報処理, **10** (5), 270 (1969).
- 8) 長尾 真他. “日本語文献における重要語の自動抽出”. 情報処理, **17** (2), 110 (1976).
- 9) 長尾 真他. “国語辞書の記憶と日本語文の自動分割”. 情報処理, **19** (6), 514 (1978).
- 10) 星野雅英. “実用型重要語抽出システム(国文学研究論文タイトルに限定して)”. ドクメンテーション研究, **28** (9), 383 (1978).
- 11) 松坂忠則. “ワカチ書きの手引き”. 日本経営出版会, 1970.

**On Japanese KWIC indexing system**, Mataichi Fukushima (Saitama University), Shuichi Ueda (Keio University) & Kazuhiko Nakayama (Tsukuba University).

There are few researches on Japanese KWIC indexing system, although practical applications of English KWIC indexing have been rather extensively studied. In these method, keywords and their contexts can be alphabetically arranged even by manual. This paper describes a method of Japanese KWIC indexing system for JAPAN/MARC. Furthermore, some problems of processing Japanese for creating the system is discussed.