JOISでの漢字検索サービスの開始

On the start of the online retrieval service with Kanji (Chinese character) in JOIS

SAKIHARA Hisayoshi

Commentary article

[著者抄録] JOIS では、従来、半角英数カナ記号での検索しかできなかったが、平成8年1月より、JOIS の機能拡充の一環として、漢字検索サービスを開始した。今回開始した漢字検索サービスは、付加的サービスとしての漢字タグ検索である。この漢字タグ検索について、サービスの概要、漢字タグ検索のメリット・デメリット、漢字タグ検索の実際について、解説する。

[著者付与キーワード] JOIS, 漢字検索, カナ文字検索, データベース, オンラインシステム

[Author Abstract] In addition to online retrieval with alphanumerics and one-byte Japanese Kana, Kanji has become available for online retrieval as part of the function expansion of JOIS. This new service starts in January, 1996. In this paper, an outline of the new service, merits and demerits of the new service, and the online retrieval actually with Kanji are explained.

[Keywords by Author] JOIS, online retrieval with Kanji, online retrieval with Japanese kana, database, online system

* 日本科学技術情報センター業務部 (〒102 千代田区四番町5-3) Tel. 03 (5214) 8419
JICST Chief Information Officer (5-3 Yonban-cho, Chiyoda-ku, 102)

1. はじめに

JOIS (JICST オンライン情報システム) では、従来、出力は全角文字・漢字カナまじりで、入力 は半角文字・カタカナという形で主要なデータベースのサービスを行ってきた。

その中で、入力も全角文字、すなわち漢字タグ検索 (以下漢字検索と省略) を可能にしてほしいとの要望が各方面よりあった。

そこで、平成8年1月より、JOIS の機能拡充の一環として、従来の半角英数カナ記号での検索 (以下カナ文字検索) に加えて漢字検索サービスを開始することになった。本稿では今回導入された漢字検索サービスの概要、漢字検索のメリット・デメリット、漢字検索の実際について解説する。

2. 漢字検索サービスの概要

漢字検索のサービス対象となる主要な検索項目はファイルによって異なるが、以下の通りである。

- ・全キーワード
- ・統制語 (JICST シソーラス登録語)
- ・下位語を含まない統制語
- ・標題・抄録切り出し語
- ・英文・原文標題切り出し語
- ・準ディスクリプタ
- ・著者名
- ・著者所属機関名
- ・資料名

また、漢字検索に対応する検索機能は、完全一致検索、マスク文字検索、論理式演算、近接演算、¥E, ¥LK, 二次検索 (ストリングサーチのみ) となっている。

後ほど、漢字検索のメリットのところでも詳しく触れるが、今回の漢字検索においてはキーワード検索 (全キーワードから準ディスクリプタまでを含む検索) では“故障”と“湖沼”, “宇宙船”と“宇宙線”などの同音異形異義語, 同形異音同義語など, また, 著者名検索においては, “健一”と“憲一”と“謙一”の区別が可能となるなどの利

点がある。なお、漢字検索可能ファイルは以下の9ファイルである。

- ・JICST (BJICST, 研修ファイルを含む)
- ・JICST7580 (BJICST7580を含む)
- ・JCLEARING
- ・JQUICK
- ・JCATALOG
- ・JMEDICINE (BJMEDICINE, 研修ファイルを含む)
- ・JAFIC
- ・OSAKA-UE
- ・NUCLEN

3. 漢字検索のメリット・デメリット

今回の漢字検索のメリットを従来のカナ文字検索と比較して、検索例をまじえて説明する。

なお、本稿での検索例はテスト環境での検索であるため、件数が通常環境での検索に比べてかなり少なくなっている。

3. 1 キーワード検索 (同音異形異義語)

同音異形異義語の検索では、漢字検索が有効と言える。なお、“宇宙船”は統制語では“宇宙飛行体”となっており、本来的には統制語にそって“宇宙飛行体”と検索すべきところ、ここでは参考までに“宇宙船”で検索した。また、ここでは限定句つきの統制語を例としたが、“美容院”と“病院”・“事故”と“自己”・“高温”と“高音”と“恒温”の区別など自然語検索、あるいは自然語検索との組み合わせでも漢字検索が有効と言える (図1)。

3. 2 キーワード検索 (医学用語)

“胃”や“歯”の前方一致検索は、ここでの例のように漢字検索とカナ文字検索とでは大きな差がでる。“歯”については、“シ”と読む場合もあり、さらに前方一致検索がカナ文字検索では困難となる。なお、このことは、医学用語以外の用語、例えば、“光”という語からはじまる用語についても、“ヒカリ”と読む場合、“コウ”と読む場合が

あり、同様のことが言える（図2）。

また、“胃”とか“歯”とかの用語については、シソーラスでその語からはじまる統制語を探すのは他の“イ”や“ハ”からはじまる“胃”や“歯”以外の言葉が間にはいり大変である。そういう時、¥Eコマンドを利用したい所だが、その場合本検索例のように、漢字検索は非常に有効であると言える（図3）。

3. 3 著者名検索

著者名検索では、読みが同姓・同名であっても、漢字検索でより精度の高い検索が可能となる。また、いくつかの読みが可能な漢字についても、漢字検索は有効であると言える。例えば、“河野”という姓を検索したい場合、“カワノ”あるいは“コウノ”という読みがあり、カナ文字検索では論理和をとる必要があるが、漢字検索では問題なく検索できる。

ただし、難しい漢字の場合、例えば、“山崎”の“崎”の字などは、記事上では“崎”や“さき”の字で収録されていることがある。検索には論理和をとったりする必要もあり、カナ文字検索の方が容易にできる。

しかし、概ね、漢字検索により、著者名と著者所属機関名のかけあわせをして検索する必要はな

くなったと言える（図4）。

3. 4 著者名検索（JICST7580 ファイル）

JICST7580ファイルでは、従来は読みが姓にしか付与されていなかったため、著者名検索は姓のみからの検索となり、著者所属機関名とのかけあわせが必要な場合が生じたり、あるいは、かなり精度の低い、ノイズの多い検索となっていた。しかし、漢字検索の導入により、JICST ファイルと同様な検索が可能となり、精度の高い検索ができることとなる（図5）。

3. 5 資料名検索

JICST ファイルでの資料名検索の例であるが、従来のローマ字切り出し語のかけあわせによる検索と比べて、和文誌の資料名の検索については、かなり容易になる（図6）。

以上、漢字検索のメリットについて説明してきたが、漢字検索には、ファイル中の表記の不統一（同音異形同義語）、同音異音異義語、JICST 使用文字種の制限（JIS 第二水準の未使用）、検索語の文字数の制限（全角14文字）等の問題もあり、現状では漢字検索よりカナ文字検索の方が網羅的検索がうまくいく場合もある。

しかし、今回の漢字検索サービスは付加的な

	U: ウチュウセン		
[1]	S: 35	ウチ1ウセン	
	U: 宇宙線		
[2]	S: 12	宇宙線	
	U: 宇宙船		
[3]	S: 23	宇宙船	
	U: KW:セイレソ		
	NO	件数	検索語
	1	3	KW:"セイレソ(メソシアケ)"
	2	47	KW:"セイレソ(REFINING)"
	3	92	KW:"セイレソ(SMELTING)"
	U: KW:精練		
[4]	S: 3	KW:精練	
	U: KW:精錬		
[5]	S: 47	KW:精錬	
	U: KW:製錬		
[6]	S: 92	KW:製錬	

図1 同音異形異義語・キーワード検索例

```

      U: 胃&
[ 7] S:      228  胃&
      U: 歯&
[ 8] S:      104  歯&
      U: い&
      処理中      4,663  AL:イイキ      →異性的
      処理中      9,205  AL:イイヨ      →胃腸
      処理中     11,682  AL:イハラキ      →茨城
      処理中     13,501  AL:インデックス      →インデックス
[ 9] S:     13,947  い&
      U: ハ
[10] S:      168  ハ

```

図2 医学用語・キーワード検索例(1)

```

U: ¥E KW:胃
NO      件数      検索語
@E21      15      KW:維持管理
@E22      15      KW:緯度
@E23      33 *    KW:胃
@E24        8      KW:胃液
@E25      16      KW:胃液分泌
@E26      17      KW:胃液分泌抑制薬
@E27        3      KW:胃炎
@E28      32      KW:胃疾患
@E29      15      KW:胃腫よう
@E30        5      KW:胃切除
S: 継続するときは ¥Eを入力してください
U: ¥E
NO      件数      検索語
@E31     109      KW:胃腸管
@E32      95      KW:胃腸疾患
@E33        3      KW:胃腸出血
@E34      12      KW:胃潰よう
@E35        5      KW:胃内投与
@E36      16      KW:胃粘膜
S: 出力終わりました
U: ¥E KW:い
NO      件数      検索語
@E37        1      KW:アノモニウムイリト
@E38       57      KW:アノモニウムカゴウアツ
@E39      33 *    KW:い      →胃
@E40        3      KW:イータソウ      →イータ相
@E41      13      KW:インカイイト      →委員会制度
@E42        1      KW:イエカゾク      →イエカ属
@E43        8      KW:イエキ      →胃液
@E44      16      KW:イエキフンビョ      →胃液分泌
@E45      17      KW:イエキフンビョクセイヤク      →胃液分泌抑制薬
@E46        1      KW:イェバエ      →イエバエ
S: 継続するときは ¥Eを入力してください

```

図3 医学用語・キーワード検索例(2)

サービスであり、漢字検索がデメリットとなるようなキーワード検索については、カナ文字検索で従来通りの検索をすることで、回避できる。

一般的には、自然語、特に同音異形異義語や短い言葉の前方一致検索、著者名、とりわけ JICST 7580 ファイルでの著者名、資料名の検索に漢字検索は有効であり、統制語からの検索には、向き不向きがあると言える。

4. 漢字検索の実際

実際に漢字検索をするためには、パソコン側の機能によるカナ漢字変換 (FEP) を立ち上げる必要がある。FEP の立ち上げ方はパソコン・ソフトにより異なるので、個々に対応する必要がある。検索コード等については、従来通り半角で入力し、必要に応じて全角文字による入力を行い、漢字検索を行う。

	U:	AU:	佐藤謙一	
[11]	S:	1	AU:	佐藤謙一
	U:	AU:	佐藤健一	
[12]	S:	1	AU:	佐藤健一
	U:	AU:	佐藤研一	
[13]	S:	1	AU:	佐藤研一
	U:	AU:	サトウ ケンイチ	
[14]	S:	3	AU:	サトウ ケンイチ

図4 著者名検索例

以下に実際の検索テーマと漢字検索例を示す。
テーマ：航空機事故 (図7)

“航空機”，“事故”ともに統制語であるが、カナ文字検索では，“ジコ”の中に“自己”が含まれて検索されることがわかる。

テーマ：胃の診断について (図8)

カナ文字検索では、本来 KW: で統制語に限定して検索すべき内容である。しかし、KW: で検索すると、絞りすぎの検索になり、網羅的検索とは言えなくなる。そこで、このようなテーマでは、漢字検索が有効となり、場合によって、胃の前方一致での検索も駆使して、より網羅的検索が可能となる。

5. お わ り に

漢字検索のメリット・デメリットのところでも触れたが、メリットの多い漢字検索にも、やはり

	U:	AU:	鈴木孝	
[1]	S:	1	AU:	鈴木孝
	U:	AU:	ススキ	
[2]	S:	130	AU:	ススキ
	U:	AU:	ススキ タカシ	
[3]	S:	0	AU:	ススキ タカシ

図5 著者名検索例 (7580 ファイル)

	U:	JT:	高分子論文集	
[1]	S:	1	JT:	高分子論文集
	U:	JT:	KOBUNSHI(W)RONBUNSHU	
[2]	S:	1	JT:	KOBUNSHI[W]JT:RONBUNSHU
	U:	JT:	コウブンシロフンシユウ	
[3]	S:	0	JT:	コウブンシロフンシユウ
	U:	JN:	G0122A	
[4]	S:	1	JN:	G0122A

図6 資料名検索例

デメリットも共存しており、現状ではその部分については、カナ文字検索で補完して検索することになる。今後、異表記等の対応辞書の作成等による改善が行われれば、漢字検索はより有用なものとなると思われる。

なお、平成8年1月のJOISの機能拡充は、本稿で取り上げた漢字検索サービスの開始の他に、記

事の重複削除、検索結果の統計的解析を容易とするための新規コマンド¥MAP, ¥SELのサービスがあり、これら新規コマンドは漢字の検索項目でも使用可能である。さらに、平成8年4月から¥P, ¥OFFコマンドの機能改善等も予定しており、より利便性の高いJOISを目標に、今後ともJOISの機能拡充を推進していく。

U: ¥FILE 010			
JICST		(1994.01)	24,660 (1995.10.21 UPDATE)
JICST COPYRIGHT			
S: 質問を開始します 1995.11.30 15:27:56 質問番号 L6R27A01			
U: 航空機			
[1]	S:	250	航空機
	U:	事故	
[2]	S:	107	事故
	U:	1*2	
[3]	S:	7	1*2
	U:	コウクキ	
[4]	S:	250	コウクキ
	U:	シコ	
[5]	S:	471	シコ
	U:	4*5	
[6]	S:	8	4*5
	U:	6*3	
[7]	S:	1	6*3
	U:	¥P A	
#000001* JICST COPYRIGHT			
CN 93A0918741, M94063886			
TI 第二世代スマートアクチュエータ			
ET Second Generation Smart Actuator.			
AU DELLER R W (HR Texttron, Inc.)			
JN D0244B (0148-7191) SAE Tech Pap Ser (Soc Aut			
omot Eng)			
RP SAE-932585			
VN PAGE. 10 p 1993			
CI (C) (a1) (EN) (USA) (写真 10)			
AB 駆動制御用電子回路を組み込んだ油圧シリンダを開発した。このシリンダはコンポーネントレベルの診断機能、軽量、光通信の採用によるEMI耐性の向上等の利点を有し、開発にあたっては、航空機の厳しい使用環境条件をクリアすることに重点が置かれた。ここでは、本装置の形態や構成、システム制御回路やシリンダ駆動用サーボ回路の構成、システム監視機能や自己診断機能の概要を示した			
CC QK05020Z (629.7.062)			
KW 戦闘機; エルロン; 操縦系統; 油圧シリンダ; サーボ制御; 故障診断; 軽量化; 光通信			

図7 航空機事故に関するテーマ検索

U: ¥FILE 050	
JMEDICINE (1994.01) 21,336 (1995.11.06 UPDATE)	
JICST COPYRIGHT	
S: 質問を開始します 1995.11.30 15:37:40 質問番号 L6R37A01	
U: 胃	
[1] S:	279 胃
U: 診断	
[2] S:	4,676 診断
U: 1*2	
[3] S:	89 1*2
U: イ	
[4] S:	319 イ
U: シンダソ	
[5] S:	4,676 シンダソ
U: 4*5	
[6] S:	100 4*5
U: 6#3	
[7] S:	11 6#3
U: ¥P T/1	
#000001* JICST COPYRIGHT	
TI 超高速シネX線CT画像を用いた脳底部血管の抽出	
KT チョウコウソクシネンCTガゾウヲモチタイノウテイブケツカンノチュウシュツ	
S: 出力終わりました (SSNO[7] から 1件)	
U: 3	
[8] S:	89 3
U: ¥P A/2	
#000002* JICST COPYRIGHT	
CN 94A0001132, L94103760	
TI 早期胃癌合併大網原発Intra-abdominal fibromatosisの1例	
ET Greater Omentum in an Early Gastric Cancer Patient. Report of a Case.	
AU 松崎茂, 高木洋行, 中村雅彦, 吉沢晋一 (波田総合病院); 石井恵子 (信州大医病院)	
JN Y0221A (0286-1135) Oncologia	
VN VOL. 26, NO. 6 PAGE. 713 - 717 1993	
CI (A) (a1) (JA) (JPN) (写真3, 参14)	
AB 60歳男性で早期胃癌を合併したintra-abdominal fibromatosisと診断の1例を報告した。腹部CTの結果, 本例は腫瘍のsolidとcysticな部分より成る大網に原発する極めて稀な症例と判明。組織学的にはムコ多糖の増加, 粘液変性に加え血管の拡張を観測した。手術は胃全摘に加え横行結腸切除を施行した。本症の病理組織学的所見などに文献的考察を加えた	
CC GE03010Y (616-006-071)	
KW 計算機トモグラフィー; 生検; 線維腫; 血管造影; 大網; 胃切除; 組織診; 症例報告; 胃腫瘍; 予後	
FT [腹部内線維腫; ガードナー症候群]	

図8 胃の診断に関するテーマ検索