

能楽番組データの Web 検索システムについて

梶山女学園大学現代マネジメント学部

三木邦弘

平成 18 年 3 月 31 日

1 はじめに

東海能楽研究会は平成 6 年頃から東海地区で行われた能楽の番組データベースを作成している [1, 2, 3]。これまでは、データを整理したものを冊子の形にして、主として能楽研究者のために提供してきた [5]。しかし、データ量の増大とともに冊子の形態では保管や検索などに不便を生じてきている。そこで元のデータを電子的なまま研究者に提供する事が考えられるので、既にいくつかの方式を検討した [6]。ところが、データを提供されてもそれを処理できなければならず、能楽研究者にとっては簡単な検索で十分な場合は、むしろ検索ができるようにする方が良い。そこでインターネットを通じてオンラインで検索ができるようにすることを試みた。能楽研究者が研究のために必要な情報は一般的な検索などでは得られないことが想定されるが、ある程度はそのような要求に答えられたと思われる。なお検索システムの過半は、平成 15 年度の生活科学部生活社会科学科の卒業生である浅田真希の卒業研究によるものである [7]。

2 検索システムに求められるもの

能楽番組検索システムの作成にあたってはいくつかの要求があった。

- できるだけ縦書きの表示とすること。
- 人名に使用されている特殊な漢字に対応すること。
- 指定された語を含む番組が検索できること。
- 人名、曲名、会名などに指定された語を含む番組が検索できること。

まず「縦書き表示」に関しては、冊子を作成した段階で、能楽研究者には横書きの表示は極めて不評であることが判明していたからである。一方欧州の研究所で誕生した WWW (World Wide Web) は、欧文と同様に文字の並びは左から右への横書きであり、各行は上から下へと並んでいるのが基本である。近年 HTML の規格の拡張により和文やアラビア語のような異なる文字の並びにも対応できるようになったが、検索システム作成段階では様々な問題が生じた。

能楽関係者の人名(芸名)には特殊な漢字が用いられることがあるので、その対応には冊子の場合も苦勞した。印刷に関しては、JIS に存在しない漢字の部分は PostScript で記述した外字と置き換えた。今回は画面上で読めれば良いと言うレベルでの対応とし、外字用のイメージファイルを用意した。

簡易な検索機能として、指定した語を含む番組を検索できるようにしている。データ自体にはタグが付加されているので、タグと絡めて精度の良い検索が本来可能であるが、そのためには利用者がタグについて十分な知識を持っていることが前提となる。そのような前提知識なしに検索することも必要であろう。例えば人名に関しては、約2万7千のうち10回以上登場する名前は4000程度であり、大抵の場合全て出してもあまり問題にならない量になる [5]。

3 縦書き表示に関する問題

縦書きの和文は、文字が上から下へと並んで1行となり、行は右から左へと並んでいる。ただ和文と言っても現在では、新聞の記事や文庫本などを除くと大抵横書きで用いられている。日本語の文字はほぼ正方形なので、縦書きでも横書きでも同じものが使用できるが、一部の記号で問題を生じることがある。

まず文字の並び方に関しては、WWWではCSS (Cascade Style Sheet) の writing-mode 属性を利用して指定する。ただしこの属性に対応しているブラウザは限られており、Internet Explorer のバージョン5.5以降から対応しているが、Netscape や FireFox¹ などは対応していない。Writing-mode 属性には、「tb-rl」と「lr-tb」の二つの値がある。「tb-rl」は「上から下、右から左」を意味し、この属性が適用される範囲では「上から下、右から左」の順に表示される。同様に「lr-tb」は「左から右、上から下」を表す。これは一般的な英文などが表示される時の文字の並び方である。よって縦書きの表示の際には「tb-rl」を指定することになる。

この writing-mode 属性は継承されないという特徴があり、HTMLの個々のタグに指定しなければならぬ。さらに個々のタグが表示する要素の並び方が、上下方向より左から右方向が優先されると不都合が生じる。つまりウィンドウの幅よりも小さな要素は左から右に並べて表示される。せっかく中身が縦書きで行が右から左へ並んでいる要素が、左から右に並んだのではとても読めない画面になる。そこで表示されるもの全体を示すHTMLのBodyタグに、direction属性として「rtl」を指定する。これは右から左と言う意味であり、これによって各要素が右から左へ並ぶだけでなく、横に長い番組でもその最初の部分(右端)が最初に表示されるようになる。

個々の文字に関しては、大半を占める幹事やひらがなに関しては問題がないが、半角の数字やアルファベットは横向きになってしまう。よってこれらの部分は writing-mode 属性を使用して普通の並び方に戻して表示されるようにする。括弧などの引用符号やその他の記号は、縦書きと横書きで変わるものと変わらないものがある。引用符号のほとんどは、図1(右)のように、その向きが変わり適切にその役割を表現する。しかし中には図1(左)のように向きが変わらず、不自然な表示になるものもあり、個々に対応が必要となる。

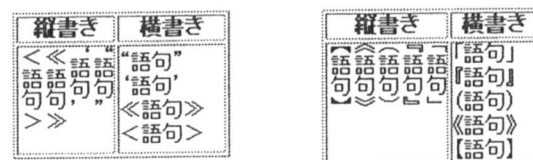


図 1: 向きが変化しない引用符号(左)と向きが変化する引用符号(右)

¹<http://www.mozilla-japan.org/>

4 特殊な漢字の扱い

能楽番組の検索システムでは、JISコードに含まれない漢字の表示に、「今昔文字鏡²」を利用している。「今昔文字鏡」は文字鏡研究会が運営しているサイトで、「今昔文字鏡」フォントや、簡易入力ツールの無償配布を行っている。ここには諸橋轍次著『大漢和辞典』（大修館書店刊）に収録される約5万の見出し漢字を全て収録しており、Unicodeとして知られる日本・中国・韓国の文字を一括で扱うための規格ISO/IEC 10646のCJK漢字も収録されている。更に、甲骨文字、西夏文字、現代中国で使われている簡体字、変体仮名まで、多種多様な文字も収録されており、合計約12万字が収録されている。

一方能楽番組データでは特殊な文字は芸名などの決まった形で出現し、入力の際には全て「●」として入力されている。前後の文字から判断して、「●」の部分「今昔文字鏡」フォントの画像に置き換えて表示する。部分的に異なる字体の異なる大きさの文字が混在するのでどうしても不釣合いな感じになるが、漢字の判読に支障は無い。

検索する語に特殊な文字を含む場合は、それをどのように入力するかが問題となる。現在の検索プログラムは指定された語を含む番組を探すために、通常の文字の部分だけ入力して検索することによっても大抵の場合問題はない。しかしそのやり方では、多数の関係ない番組が出力されることもある。幸いなことに特殊な文字を含む例はそれほど多くないため、現在では特殊な文字を含む語の一覧を表示し、その中から検索したい語をクリックするだけで検索されるようにしている。

5 検索システムの利用方法

能楽番組の検索システムへアクセスするためには、次のURLをブラウザで指定する。ブラウザとしてInternet Explorerのver.5.5以降を使用すると縦書きに表示されるが、それ以外のブラウザでは横書きで表示される。表示の際の見た目の違いだけで、その他の検索の機能などには違いは生じない。

<http://zeami.ci.sugiyama-u.ac.jp/~miki/search.html>

最初の画面にはタイトルの他、検索の画面と説明の画面へのリンクなどがある。ここで検索の画面へのリンクをクリックすると次のような画面(図2)になる。

ここで検索の目的にあわせて次のように使用する。

簡単検索

検索したい語を入力した上で「検索開始」ボタンをクリックする。検索結果は、検索語を含む番組の一覧の形で表示される。100件以上の番組が該当した場合、最初の100件がまず表示される。番組の内容を見たい場合は、番組名をクリックする。100件以降の検索結果を見たい場合は、「次のページへ」をクリックする。

特殊な漢字を含む語の場合は、「JISに無い漢字を含む語の検索はこちら」をクリックすると一覧が表示されるので、その中で検索したい語をクリックすると検索される。

²<http://www.mojikyo.org/>

簡単検索

キーワードを入力して下さい。(E)に無い漢字を含む語の検索は「E」あり

検索開始

リセット

人名指定の検索

人名を入力して下さい。複数の検索をした場合は、下にも人名を入力して下さい。

演能団体 もしくは主催者についても追跡したい場合は入力して下さい。

複数の人名を指定した場合、同じ演目にも出演していなければなりません。●はい ○いいえ

検索開始

リセット

曲名指定の検索

曲名を入力して下さい。

検索開始

リセット

日付指定の検索

日付を選択して下さい。

年

月

検索開始

リセット

分類項目からの検索

次の分類項目をクリックして下さい。その下分類されたものの一覧が表示されます。なお、○の中の数字はその項目に含まれるものの数を示しています。

- (2) 下(991) 小(17676) シン(7812) 口(7157) 口(1)
- (3053) 演目(1006) 舞(499) 会(5321) 解説(58) 問(1474)
- 問(191) 問(88) 問(41) 問(17) 問(10) 問(7) 問(2) 共催
- (2) 共催者(44) 協賛者(2) 区(1) 形態(32) 後(145)
- 後(1104) 後見(1104) 後小鼓(1) 三番三(18)
- 三番(73) 指導者(2) 次(1097) 主催(8) 主催者(13)
- 小鼓(3122) 場所(1155) 神楽(6) 千歳(537) 前(1) 前(1)
- 前(1) 前小鼓(1) 太鼓(307) 大鼓(736) 題(32) 題目(1)
- 担当(6) 地謡(1158) 注記(540) 笛(1025) 典拠(6) 独調(1)
- 日時(6015) 入場料(157) 備考(494) 補助(6) 面箱(32) 謡(10649)

図 2: 検索の画面

人名指定の検索

複数の人名を指定すると、同じ番組に共に含まれる場合のみ検索することができる。さらに条件を同じ演目で共演したかどうかまで狭くすることも可能である。能楽のプロは招かれて様々な素人の会にも出演する機会が多い。そのような会での出演、またはそうでない会での出演を調べやすいように、団体や主催者の条件を付けられるようにしている。

曲名指定の検索

能楽で演じられる曲にも流行があり、そのような調査に利用することができる。ただ流派により曲名が異なるものもあり、それらを一つにまとめるようなことはしていないデータなので、十分注意する必要がある。

日付指定の検索

冊子体での本編の代わりとなるものである。指定された年月の番組が日付の順番に出てくる。

分類項目からの検索

番組データに付けられたタグごとに分類したものである。例えば「演目」をクリックすると、演目のタグが付けられたものが図2の場合、1006種類あり、その一覧が表示される(図3)。そこでさらに見たい演目をクリックすると、それを含む番組の一覧が表示される。数千件を越えるものになると五十音順に並んでいても探し出すのに苦勞するが、前述の曲名の表記のゆれなどは見つけやすい。例えば図3においても、「阿古木」と「阿漕」のような例が見られる。

「演目」項目計 1006 個	
●(22)	●犬盗人(1) 口之段(1) あまりの徒然
(1)	あやめ酒(2) いたいけしたる物(1) いろは
(2)	うさぎ(2) おかしき天狗(7) おひやし(1)
(1)	お調へ(1) お用の尼(2) お冷し(4) かぶ
(1)	れ太郎(1) くるす桜(4) こがとり(4) さんざ
(2)	しびり(14) じゃじゃ馬馴し(1) じゃじゃ馬
(1)	馴らし(1) ともしび(1) とらまだら(1) とり
(1)	かえばや(1) はりこ丸(1) ひょうたん(1) み
(1)	んなで遊ぼうポンポコリン(1) よしの葉(16) ク
(1)	ソロギ(1) シェイクスピア新作狂言(1) スカパ
(1)	の悪だくみ(1) モリエール新作狂言(1) 唾の一
(1)	声(1) 阿古屋松(10) 阿古木(23) 阿漕(79) 葵
(1143)	悪女(1) 悪太郎(68) 悪坊(17) 芦刈
(1042709)	綾鼓(32) 鮎之段(42) 栗田口(17) 安宅
(288)	安達原(288) 庵梅(3) 鞍馬(1) 鞍馬参
(17)	鞍馬参り(3) 鞍馬天狗(172) 鞍馬智(4)
(1)	以呂波(2) 伊勢詣(2) 伊文字(25) 伊呂波
(61)	夷昆沙門(1) 夷大黒(35) 夷昆沙門(11)
(1)	井戸茶碗(1) 井杭(2) 井筒(1129) 越之音取
(1)	一角仙人(17) 一管(1) 一字題(7) 一声
(1)	一天四海(1) 稲荷山(1) 因幡堂(69) 引
(42)	引数算(1) 隠狸(74) 右近(100) 右近左
(10)	右流左止(1) 宇治の晒(39) 烏追舟(1)
(98)	烏頭(98) 烏帽子折(62) 羽衣(237) 雨月(228) 雨
(86)	鶯(86) 鶯(1) 鶯の鳥(4) 鶯ノ殿(1) 鶯羽
(2)	鶯飼(509) 鶯之段(686) 姥捨(1) 浦下部
(1)	浦島(2) 瓜盗人(75) 雲山(1) 雲雀山

図 3: 演目の一覧(一部)

6 検索の利用状況

この検索システムの改良の参考にするために、利用者の検索状況の記録を採集している。記録をまとめたものが表1である。「アクセス件数」はその月の全ての検索の総計である。「項目一覧」は、分類項目のどれかをクリックして一覧を表示した回数、「項目一覧」はさらにその一覧の中でどれかをクリックして番組の一覧を表示させた回数である。この「項目一覧」の回数が非常に大きな数になっているが、恐らく Google などの Web 検索システムのロボットが巡回して来たものと思われる。

表 1: 検索システムの利用状況

年	月	アクセス件数	簡易検索	人名検索	曲名検索	日付検索	項目一覧	項目表示
2004	8	540	41	32	13	2	265	187
	9	7,395	138	119	70	7	1,664	5,397
	10	4,587	62	15	3	9	695	3,803
	11	68,823	11	15	29	0	154	68,614
	12	5,159	112	36	25	1	447	4,538
2005	1	5,007	200	36	34	1	74	4,662
	2	6,433	350	77	34	7	925	5,040
	3	84,788	297	28	7	0	2,022	82,434
	4	5,641	161	54	13	1	1,091	4,321
	5	4,880	199	33	3	0	2,234	2,411
	6	46,872	228	85	16	4	1,954	44,585
	7	9,736	252	40	16	3	1,923	7,502
	8	64,942	214	90	11	11	1,469	63,147
	9	48,668	237	171	40	2	933	47,285
	10	121,270	203	35	17	3	1,038	119,974
	11	74,609	170	19	11	1	2,565	71,843
	12	127,671	203	40	28	2	2,873	124,525
2006	1	39,603	262	41	3	3	3,934	35,360
	2	23,319	185	16	0	7	3,844	19,267
	3	75,505	168	40	22	3	3,929	71,343
平均		41,272	185	51	20	3	1,702	39,312

7 おわりに

能楽番組の検索システムを作成する際に生じた問題点やその解決方法について述べた。特に検索に関しては何も工夫はしていないが、それほどデータ量も多くなく、検索時間などの問題は無い。今後はデータ量も増加するが、恐らくコンピュータの処理能力の向上の方がはるかに進むであろう。現在のところ考えられる問題点は、

- 検索結果を印刷すると綺麗に印字されない。
- 検索結果をワープロなどに移すと縦書きなどの書式が乱れる。
- データの入力・校正作業などが追いつかない。

などがある。最初の二つは検索結果を保存したり、引用して論文などに使用する際に問題となる。

参考文献

- [1] 東海能楽研究会：「東海能楽年鑑 平成5年版」、平成6年10月、中部日本教育文化会
- [2] 深谷哲監修、笥鉦一、飯塚恵理人編集：「東海地域能楽番組一覧 明治元年～昭和26年」、平成9年7月、椋山女学園大学
- [3] 深谷哲監修、笥鉦一、飯塚恵理人編集：「近代名古屋の能楽を支えた人々」全三巻、平成13年7月、東海能楽研究会発行
- [4] 飯塚恵理人、三木邦弘、深谷哲：「東海地域能楽番組データベースの使用プログラム(1) - 能楽番組データ検査プログラム "CHECK1" を中心に -」、平成12年3月、『椋山女学園大学研究論集』第31号人文科学篇 p.55～68、椋山女学園大学
- [5] 三木邦弘：「能楽番組データベースからプログラムで読み取れるもの」、平成14年1月、椋山女学園総合クリエイティブセンター「創」第4号 p.23～32、椋山女学園総合クリエイティブセンター
- [6] 三木邦弘：「能楽番組データの記述形式についての考察 - 独自形式、XML、RDBの比較検討 -」、平成15年3月、平成12年度～平成14年度科学研究費補助金・基盤研究(C)(2)研究成果報告書「近世・近代能楽資料の収集・整理とデータベース化 - 東海地域を中心に -」課題番号 12610457 p.5～14
- [7] 浅田真希：「能楽検索システムの試作」、平成16年3月、「卒論ノート」vol.10 p.1～15、椋山女学園大学生生活科学部生活社会科学科