# 心理学の文献をWebサイトで公開するための方法について
## ― 日本心理学会大会の抄録集のデジタル化索引を例とした一考察 ―

岩　橋　俊　哉[注1)]

# A method which publishes the psychological literature on the Web site
## —One consideration which designates the digitalization index of the preliminary reports of the Japanese Psychological Association conference as one example —

Iwahashi Toshiya

### 序

日本心理学会の年次大会は，今年度で開催が68回目であり，その抄録集は，１大会分がＢ５版サイズで約1000ページからなる膨大な資料である。しかし，その抄録集は，審査制ではないということもあり，多くの場合管理がきちんとなされていないことが多いように思われる。しかし，論文の形式が整い，時期も同時，そしてその内容からも心理学史の見地から見てと重要な記録であると考えられる。そこで心理学史研究の見地から，日本心理学会大会の抄録集の索引をデジタル化してインターネット上に公開することで，資料の検索性を高め，心理学史研究を促進する一助とすることを考えた。

学術雑誌の論文については，NACSIS-ELS（http://els.nii.ac.jp/nacsis-els-j.php3?top）などに代表されるように，心理学を含めた多くの分野で全テキストのデジタル化が進められている。まだ紙の雑誌に全面的に置き換えられるほどではないが，コンピュータの画面上での閲覧がかなり容易になってきた。また海外の例では米国心理学会（American Psychological Association : http://www.apa.org/）で，一部の雑誌がWeb上で公開されているなど，論文の電子化とインターネット上での公開が様々なところで進められている。

デジタル化は，現時点ではまだかなり手間がかかるものであるが，その方法は技術的には確立されている。ただし，著作権の問題があり，全文をデジタル化する際には，これをきちんと法律的に処理しておかないと問題が生ずるので注意が必要である。その上，全文のデジタル化は，必ずしも検索作業を効率化するわけではない。通常の書籍のように閲覧することが現在のコンピュータの画面上ではまだ困難である。この問題は，主として画面情報の処理速度に起因する問題であ

るので今後，改善が進めば書籍より有利になる可能性もある。辞書のような書籍では，結果として閲覧する箇所が少ないので電子化して閲覧することは有利であるのだが，現時点では通常の閲覧は，書籍のほうが有利である。

### 索引のみのデジタル化の試み

そこで，索引のみをデジタル化するという方法を考えた。これはパーソナル・コンピュータのファイルシステムで検索によく用いられている手法と同様な方法である（例として，ビレッジセンター社製のサーチクロス（http://www.villagecenter.co.jp/soft/searchx/），Apple 社（http://www.apple.com/）の Mac OS での旧バージョンの Sherlock など。あるいは電子化された図書館の検索システムのようなものと比較できる。この方法では，抄録集本体の存在は検索に必須であるため，書籍の保管などは従来の方法で行うことになるが，その検索性は飛躍的に高まるのではないかと考えられる。

### 索引データのデザイン

まず，方針として，索引およびその元データは独立したファイルとして作成すること，索引の元データとしてのファィルを最初に作成する，そのファイルを順次インターネット上に公開してゆくことにした。最終的には索引化したファイルを作成することが理想ではあるが，用語表記の統一が成されていないという問題があるので，それを踏まえて，最初は用いられている用語をそのまま記録することとした。

書式として，1）ファイルフォーマットは，扱いやすいテキストファイルとする，2）本文から，固有名詞のみ抽出し，単語はすべてカンマで区切る（CSV: Comma Separated Variables 形式），3）削除する単語は名詞以外の接続詞（例として，しかし，そこで，あるいは，において，である，次に，と，など），抄録集で頻出する見出し語，言い回し（例として，目的，結果，検討，分析，有為，考察，要因，考えられる，問題，必要がある，方法，考察，妥当，結果，反映，示している，最近，特徴）である。

この方針で作成すると索引データのは，以下の例のようになる。

**元の文**

...社会的弱者の被害が増加している可能性を検討するために被害者に占める各年齢層の比率を検討した。その結果，犯罪全体及び殺人において，1994，1995年頃を境に60歳以上の高齢者の比率が急増するという傾向が見られた。また，暴行においても高齢者の比率が漸増する傾向がみられた。しかし，幼児や児童の比率には顕著な傾向は確認できなかった。…

**索引データ**

...社会的，弱者，被害，被害者，各年齢層，比率，犯罪全体，殺人，1994，1995，60歳，　以上，高齢者，比率，暴行，高齢者，比率，幼児，児童，比率…

さらに，各発表の要素を区別し，以下のタグで囲む[注3]。データベースとしての書式を整える。

<record></record>　1レコード（= 一ページ分）
<category></category>　分野
<title></title>　論文の題目　＊本文と同様にキーワード化する
<author></author>　著者　＊姓と名は分離しない
<position></position>　所属
<key></key>　キーワード
<body></body>　本文
<yomi></yomi>　著者名の読み
<caption></caption>　図表の見出し　＊本文と同様にキーワード化する
<page></page>　ページ番号

## 索引の利用方法

一番単純な方法では，エディターなどで索引データファイルを開き，その中からキーワードや本文の単語を検索し，該当するページ（<page></page>）タグ内を参照する，ということになる。

## 索引の作成手順（試験データの作成）

まず，スキャナーで印刷されたデータを読み取って画像ファイルを作成する[注2]。スキャナーは，読み取り速度を重視してEpson製ES-9000Hを選択した。その速度は，600dpiで32枚/1分程度である。ちなみにファイルサイズは，1ページあたり3 MBとなる（抄録一冊は約1000ページなので，一冊あたり3 GBとなる）。OCRアプリケーションには，メディアドライブ社製（http://mediadrive.jp/）のWinReaderProを使用した。このアプリケーションは，プロ用途であり，個人向けよりは変換精度が高い製品であるが，それでも変換精度にはやや難があった。変換速度はコンピュータのハードウェアに依存するので一概には言えないが，今回は特に問題は感じられなかった。ちなみに手書き文字の場合には目視ではかなり可能かと思われる文字でも変換は数パーセントという結果で，ほぼ不可能といってよい状態であった。この場合にはデータをキーボードから入力せざるを得ない。作成されたテキストファイルは，1ページあたり約6 KBである。校正は人手を用いて行うことになる。心理学についてある程度の素養がある学生などに依頼することになったが，所要時間は1ページあたり約10分程度であった。印刷された文をキーボードから入力

する場合には１ページあたり１時間程度かかると考えられる。

## データの保管について

画像ファイルとテキストファイルを保存するメディアは，１年分を入れるにはDVD（4.7GB）一枚分が必要である。テキストファイルは１年分で６MB程度なので，ほとんどが画像ファイルで占められることになる。

## 索引データの公開方法

画像ファイルおよび，全テキストは現時点では公開できないが，作成された索引データを，作業としては途中であるが，サイトで公開することにする(注4)。このデータを元に索引を作成することができる。最初はテスト公開とし，テキストファイル形式の索引データを順に公開する。最終的には１年分の抄録一冊が１ファイルとなるようにする。サイズは数MB程度であろう。

## 今後の展開

今後の展開としては，ふたつの方向がある。ひとつは索引の対象を増やしてゆくことがひとつ。もうひとつはより洗練された検索ができる索引データベースを設計することである。どちらも重要であるが，当面はこの対象データの充実につとめたいと考えている。

### 参照URL

Apple Computer Inc.　http://www.apple.com/

EPSON　http://www.i-love-epson.co.jp/

メディアドライブ　http://mediadrive.jp/

NACSIS-ELS　http://els.nii.ac.jp/nacsis-els-j.php3?top

サーチクロス　http://www.villagecenter.co.jp/soft/searchx/

### 謝辞

デジタル化した抄録原稿の校正に際して，千葉大学教育学部の大葦治氏に助言等をいただきました。ここに謝意を表します。

### 脚注

注１．本研究は，平成15年度の大東文化大学特別研究費の助成を受けたものである。

注２．スキャナーを利用する際，資料の紙が薄く裏の文字が透けて移ってしまう場合には，裏に黒い紙をはさむことでそれを防ぐことができる。

注３．タグの名称は，任意に設定したもので変更不可能ではない。

注４．テスト公開のURLは，http://www.daito.ac.jp/~iwahashi/Psyc/

(2004年９月14日受理)