



Title	漢籍電子テキスト所蔵情報集約の提言
Author(s)	井上, 了
Citation	中国研究集刊. 2000, 26, p. 1-3
Version Type	VoR
URL	<a href="https://doi.org/10.18910/61005">https://doi.org/10.18910/61005</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# 漢籍電子テキスト所蔵情報集約の提言

井 上 了

いわゆる電子テキストとは、文献を電算機に入力し作成された文書ファイルである。漢籍についても多数が作成されており、インターネット上などには膨大な量の「電子漢籍」が公開されている。東洋学の分野において電子テキストを利用した研究は一般的ではないが、今後は検索などの用途のために「電子漢籍」の需要は拡大するであろう。

何らかの文献について電子テキストを利用するためには、まずその文献の電子テキストの有無や所在を調査せねばならない。また新たに電子テキストを作成・公開しようとするならば、二重入力 of の愚を避けるためにも、まず先行する電子テキストについて調査すべきである。しかしこの分野には『東洋学文献類目』のような便利な目録が存在せず、雑誌類の彙報欄などにも電子テキスト関係の情報は掲載されないのが通例である。今後は「電子漢籍」についても「テキストを公開した者は原則としてそこに登録する/される」といった「権威ある目録」が、然るべき組織によりネット上に運営されるべきであろう。漢字文献情報処理研究会(一)が準備を進めているという「漢字文献電子テキスト総覧」(二)はこれに近い企画かと推測されるが、いまだ公開には至っていないようである。現状では、「電子漢籍」の所蔵情報については検索エンジンおよびリンク集を頼りとして不確実な調査を行わざるを得ないと筆者は認識している。

漢籍電子テキスト関係の検索エンジンは十種以上存在する。もちろん、目的とする文献名を入力すれば直ちに検索結果を返してくれる検索エンジンは便利である。しかし、一覧表示を行えず、検索結果のみを返してくる検索エンジンでは、文字コード非統合問題などに起因する「取りこぼし」の可能性を完全には排除できない(三)。またこれらの検索エンジンの多くは登録型であり(robotによる電子テキスト情報の自動蒐集は構造上困難であろう)、多くの電子テキストを掲載しているサイトについては、所載の電子テキストすべてをエンジンに登録していないことのほうが多い。

漢籍電子テキスト関係のリンク集は、検索エンジンよりもはるかに多く存在する。しかしその多くは、電子テキストを掲載しているサイトを並べ、各サイトに何のテキストが存在するかを列挙したものである(次頁図左側)。このよう

なリンク集の形式は、目的とする文献が定まっている場合には不便なものと感じられよう。利用者の大半が必要としているのは、「A大学には『論語』と『孟子』の電子テキストがある、B研究所には『論語』と『孝経』と『大学』の電子テキストがある」という情報の列挙ではなく、『孟子』の電子テキストはA大学とCサイトとにある」という情報なのではないか。そのような情報を提供するためには、文献別に所蔵サイトを分類し直した目録＝リンク集が便利であることはいうまでもない(四)。

実際に電子テキストの情報を収集する際には、極めて詳細な情報までを必要とするわけではない。『論語』のようにネット上に数十種類もの電子テキストが氾濫しているものについてはともかく、数種類程度しか存在しない文献については、必要な文献がそもそも電子テキスト化されているのか否か、されているのならその電子テキストはどこにあるのか、ということが判明すればよい。あとは所在サイトに飛び、当該テキストを実見すればよいのである。かような所在情報が提供されるだけでも、節約される労力（検索の労力および二重入力の労力）は膨大なものとなろう。

たとえば筆者は、経・史・子部（釈家を除く）について下図右側のような簡単な目録＝リンク集を作成しており、所属研究室のサーバ(五)を利用してとりあえず公開している。

#### 現存する多くのリンク集の形式

##### A大学

『論語』(Big5)

『孟子』(GB)

##### B研究所

『論語』(JIS)

『孝経』(Big5)

『大学』(UTF-8)

##### Cサイト

『中庸』(S-JIS)

『孟子』(GB)

:

#### 筆者が提案するリンク集の形式

##### 『論語』

A大学 (Big5)

B研究所 (JIS)

##### 『孟子』

A大学 (GB)

Cサイト (GB)

##### 『大学』

B研究所 (UTF-8)

##### 『中庸』

Cサイト (S-JIS)

:

この筆者の目録については、「縦のものを横に並び替えただけのもの」だとの評価も頂いている。だが「並べ替え」とは、検索のために最初に必要な作業でもあろう。

筆者の目録は、調査した機関も少なく、チェックも甘い代物で、とうてい実用的なもの（これ一つにて足る、という意味での）ではない。この目録は筆者の提案を示す見本にすぎないからである。しかし、このような拙い目録でも、とりあえず公開してしまうことにより一定の効果を発揮する。現にこの目録に対しては、各機関や個人からの情報（誤りの指摘やお叱り）が集まってきており、フィードバックによる充実が加わりつつある。

訂正が困難な冊子体目録ではないのだから、とりあえず公開して広く協力を求め、お叱りを受けた部分について順次訂正していくほうがよいのではないか。このような姿勢は研究者として誠実ではないかもしれないが、遺漏なきを目指してチェックを重ね、いつまでも公開しないよりは、まだしも後学を益することになろう。

もちろん、所属研究室の全面的な支援を頂いているとはいえ、個人作業に近い形でこのようなコンテンツを維持するのは不自然である。ましてや実験的な目録から「権威ある目録」への脱皮など想像すらできない。筆者としては、一日も早く「権威ある電子テキスト目録」が立ち上げられ、このような私的な目録が不要となることを願うものである。そのような目録が構築されるに際して以上に述べたような方法が考慮されれば、筆者にとり望外の幸いである。

(一) <http://jaet.gr.jp/>。(2000年3月現在)

(二) 漢字文献情報処理研究会編『電腦中国学』(好文出版、1998年) 208ページ。

(三) たとえば、「内経」で検索を実行しても「内経」がヒットしないという現象ならば比較的理解されやすく、「内経」で検索をやり直すことに思い至る利用者もいるであろう。しかし「内経」で検索して「内経」がヒットしないという事態(UTF-8では、口部の「内」と入部の「内」とは別字とされる)は理解されにくく、結果として「内経」の電子テキストに到達し得ない利用者が生じる虞れがある。

(四) これら二者の相違については、『東洋史研究』の「近刊叢欄」欄と『日本中国学会報』の「学界展望」欄との体裁の相違を想起されれば理解されやすいであろう。

(五) <http://bun165.let.osaka-u.ac.jp/>。(2000年3月現在)