

DVD 版内村鑑三全集と学術資料電子出版

當山 日出夫[†]

DVD 版『内村鑑三全集』全集が刊行となった。これは、書籍版の全集を忠実にワープロで再現した PDF を利用したものである。読書用の表示と、検索用の機能をあわせもっている。また、PDF のデータを独自に、利用者が利用することも可能である。これは、既存の書籍のデジタル化、電子出版において、ひとつの方向性をしめすものである。本発表では、この DVD 版全集の概要について報告するとともに、その電子出版としての意義について考えることとする。

DVD “ UCHIMURA KANZO ZENSHU” and Academic Digital books

TOUYAMA Hideo[†]

DVD version 'UCHIMURA KANZO' complete volume was published. This is the one using PDF. The display for reading can be able to be done, and to retrieve the word. Moreover, it is also possible that the user uses the data of PDF. In this research announcement, it reports on this DVD version complete volume. The meaning as the e-publishing is described.

1. はじめに

内村鑑三は、近代の日本を代表する思想家（キリスト教信徒）である。その全著作は、最新の全集としては、岩波書店版の全 40 巻本にまとめられている。この『全集』を、すべてデジタル化した企画が完成した。DVD 版『内村鑑三全集』である。

これは、内村鑑三全集 DVD 版出版会の手になるものである。一部に、文字・検索において若干の問題点があるとはいうものの、全 40 巻の生涯にわたる全著作について、デジタル化し、それに検索機能を付加して、一般の市販として提供した意義はきわめて高いものがある。

本稿では、この DVD 版『内村鑑三全集』の概要について紹介するとともに、その電子出版として、現代における意義について考えてみたい。

2. 内村鑑三について

2.1

- ・1861年（万延2年） 江戸に生まれる。
- ・1877年（明治10年） 札幌農学校に入学。在学中に、「イエスを信ずる者の契約」に署名する。
- ・1878年（明治11年） 宣教師M・C・ハリスから受洗。
- ・1884年（明治17年） 農商務省に勤務。渡米、アマスト大学に入学。
- ・1888年（明治21年） 帰国。
- ・1891年（明治24年） 「不敬事件」によって、第一高等中学校を退職。
- ・1893年（明治26年） 『基督信徒の慰め』
- ・1894年（明治27年） 『日本及び日本人』（『代表的日本人』）英文
- ・1895年（明治28年） 『余はいかにしてキリスト信徒となりしか』（英文）
- ・1897年（明治30年） 『後世への最大遺物』 『万朝報』入社
- ・1900年（明治33年） 『聖書之研究』創刊 無教会主義を表明 聖書研究会
- ・1903年（明治36年） 日露戦争を前にして「非戦論」を展開、『万朝報』退社。以後『聖書之研究』と聖書研究会を中心に活動をつづける。
- ・1930年（昭和5年） 70年にわたる生涯を終える。

[†] 立命館大学グローバルCOE (DH-JAC) 客員研究員

内村は、無教会主義をとるなど、特異な傾向があるとはいうものの、日本近代の思想史、特にキリスト教思想史において、屹立した存在であり、現在でも、なお多大の影響がある。その代表的著作『余はいかにしてキリスト信徒となりしか』『代表的日本人』など、いまでも文庫本で読めるものとしてある。また、研究誌『内村鑑三研究』も刊行されている。2010年の時点で、第43号まで刊行（年1回）。

2.2

内村は、およそ70年にわたる生涯において、膨大な著作を残している。その著作は、現在までに幾度かまとめられてきている。現時点での通行の全集としては、1980年に刊行の、『内村鑑三全集』（全40巻）である。そして、その後、2001年に、増補版として、「第二刷」が刊行されている。DVD版『内村鑑三全集』は、この「第二刷」の版をもとにしている。

3.DVD版『内村鑑三全集』の経緯

DVD版『内村鑑三全集』（以下DVD版全集と略す）は、前述、第二刷『内村鑑三全集』（全40巻）を底本としている（以下、これは書籍版全集と略すことにする）。それは、次のような方法で作成されている。

3.1

書籍版全集を、全文、ワープロ（一太郎）で、入力する。これは、内村鑑三全集DVD版出版会（旧称：内村鑑三全集CD-ROM版出版会）の方たちの、手作業による入力である。これは、OCRなどはさして有効につかえない。なぜなら、内村の著作としては、総ルビ（ないしは、それに近い）ものが多い。また、圏点などもきわめて多くつかっている。基本的にすべて手作業による、本文の再入力である。

なお、書籍版の全集は、当時の印刷として、まだ活版印刷の時代である。つまり、現在のように印刷用のデータ（PDF）などがあるわけではない。ゼロから、書籍版全集を入力したものである。

また、その入力も単なる本文データの入力ではない。書籍版全集の「再現」といってよいものである。まったく書籍版全集のページ組版どおりにすべての文字を、忠実に再現（ワープロ）している。

著作によっては、行の右側にルビ（ふりがな）、左側に圏点、というような箇所もあるが、これも忠実に再現している。これは、通常のワープロの編集機能ではできない。圏点専用の行を設定して、行間調整をおこなって、なかば強引に書籍版全集の版面を再現

しているものである。

3.2

このワープロ（一太郎）で再現した全集本文をPDFとして利用している。これは、現在ならば、組版用ソフト（たとえば、InDesignなど）をつかうところである。しかし、この企画がはじまったのが、10年以上もさかのぼること、書籍版全集の完成後には準備的にスタートしていることなどを考えると、この判断は妥当なものであったといえよう。

また、企画が早くよりスタートしたせいでもあろうが、画像データの利用にはなっていない。現在であるならば、通常のパソコンで、JPEG形式などの画像データをあつかうことは容易である。また、そのためのスキャナも、低価格・高品質なものになってきている。

これも、現在であるならば、まず画像データで、書籍版をスキャンして、まず画像で提供と考えるところであるが、むしろ、そうならなかったのは幸いとすべきである。そのために、すべての書籍版の文字データを入力して、得ることができたのである。また、これは当初から、書籍版のデジタル化を、全文検索に目的を設定していたことによる。このことも見逃せない点である。

ここまで、DVD版全集出版会の、いわば「素人」（デジタル関係については）の仕事である。

3.3

そのつぎには、ワープロ（一太郎）のデータを印刷会社（精興社）にわたして、以下のような作業をすることになる。ここからは、デジタルの「プロ」の仕事になる。

ワープロで再現した書籍版全集の版面データから、テキスト（本行）を抜き出し。このテキストデータに対して、インデックスを作成し、文字列検索のデータとする。ただ、このとき、PDF化した書籍版全集のページが単位となっているため、ページ単位のテキストデータをあつかうことになった。したがって、結果的に、ページにまたがった文字列の検索ができないという不都合を生じることになっている。

また、検索用テキストのエンコーディングは、Unicode3.0のUTF-16（ただし、サロゲートペアを除く）となっている。

3.4

検索システムの組み込みとそれに連動する閲覧システムを導入する。閲覧には、LaefThrough Contents Viewer（大日本スクリーン）をつかい、HiBase（ホロン社製データベース管理システム）を使用している。これは、PDFのデータを、検索すると同時

に、書籍を閲覧するがごとくに、ページをめくったり拡大したりする機能をそなえている。

また、このLeafThroughによる閲覧からは、PDF表示に移行することも可能になっている。ここで、注目すべきは、PDF（書籍版を忠実に再現）を独立した形で、残してあることである。これだけを、AdobeReader（または、Acrobatなど、PDF閲覧用ソフト）で、閲覧したり、検索したりも可能になっている。

3.5

以上のながれは、全集刊行時から年代を追って再整理すると次のようになる。DVD版全集所収「出版会の発足と刊行までの経緯」から、適宜整理すると、

- ・1980年 『内村鑑三全集』刊行開始 岩波書店
- ・1998年 OCRによる読み取りをこころみはじめる
- ・2001年 『内村鑑三全集』（第二刷）刊行
- ・2001年 岩波書店と交渉（全集の編集権について）委員会発足 精興社との協議でPDFの採用に決定
- ・2003年 データ入力を委員会において行うことに決断 内村鑑三全集CD-ROM版出版会を結成
- ・2006年 テスト版として、第26巻（1921年）を作成
- ・2007年 全巻のデータ入力・校正
- ・2008年 検索と閲覧の別立てにすることに方針を決める DVD版とする
- ・2009年 7月、作業完了 出版会の名称を「内村鑑三全集DVD版出版会」と改める
8月『DVD版内村鑑三全集』刊行

4. DVD版全集の概要

以下、DVD版全集について、その利用にあたっての概要を述べる

4.1

基本となる動作環境は、Microsoft Windows XP（SP2） Windows Vista、となっている。また、32ビット版のみ（パンフレット）。なお、筆者の使用環境である、Windows 7（32ビット）においても、正常に作動することを確認しておきたい。

スペックとしては、必要メモリは、XP：512MB（推奨：1GB以上）、Vista：1GB（推奨：2GB以上）。1GB以上のドライブ空き容量、（HDにコピーして使用の場合は6GB以上が必要）。また、片面二層DVDのドライブが必要である。

これらのスペックは、現在のパソコン環境としては、さほど特殊なものではなく、ごく通常のマシンで正常に作動する。また、HDに全データをコピーしても利用できる。

4.2

インストールはきわめて簡単である。DVD版全集をドライブにいれて、LeafThroughCVをHDにインストールするだけである。また、AdobeReader（Ver.9）を必要とする。

データを、DVDのままでも使用することもできるし、また、それをHDにコピーして使用することも可能である。プロテクトなどはかけていない。これは、学術的な電子出版として、良心的な処理であると同時に、また、その利用の質をたかめるためにも重要なことであると判断される。

4.3

システムの起動も、簡単になっている。スタートの画面から「スタート」をクリックするだけでよい。起動の直後には、「読書用表示」と「検索」の二つのタブから、いずれかを選択するようになっている。そして、「読書用表示」と「検索」は、随時きりかえが可能である。

5. 読書用表示

まず、読書用表示の機能から説明する。

5.1

DVD版全集の一つの特徴は、読書用の画面と検索機能を分離しているところにある。読書用表示画面では、上下二段にわかれて、上段に巻数（全40巻のうちの第何巻であるかがしめされ）、下段にはその該当する巻の目次がしめされる。もちろん、これは、書籍版全集に忠実に、である。

そして、読みたい箇所（目次）をクリックすると、その文章の巻頭が表示される。これは、最初に述べたごとく、書籍版全集にきわめて忠実に作ってある、というより、それを再現したものとして画面に表示される。

読む場合には、1ページずつクリックしてページをめくっていく感覚で、次のページに移動する。1ページずつのページめくりでは間に合わない場合に、10ページごとにめくこともできる。もちろん、ページをめくる方向を逆にする、左からめくる／右からめくる、も自在である。また、ページをめくる速度調整機能もついている。

画面の表示サイズは、任意に決められる。

ウィンドウのサイズにあわせて、その範囲内で全画面（見開き）表示の大きさを変えることが可能。

部分的に、表示しているページの部分を拡大して見ることもできる。これらの操作は、マウス右ボタンのメニューからの「ズームナビゲーション」によって可能である。また、この「ズームナビゲーション」それ自体を見えなくすることもできるようになっている。以上がLeafThroughによる基本の機能である。

5.2

マウス右ボタンメニューにある「アプリケーションで開く」を使用すると、LeafThroughではなく、PDFをそのまま見ることができ、PDFに遷移する。このとき、PDFの該当するページが表示されるようになっている。

PDF（フォント埋め込み）であるから、ここから引用（コピー）が自由にできる。これは、検索結果画面からも同様のことが可能であり、閲覧と検索、そして、それを使つての知的生産（簡単にいえば、これをつかって論文を書く）が、うまく結合することになる。この点は、このDVD版全集で評価されるべき特徴のひとつであろう。

5.3

つまり、このDVD版全集では、読書用表示として、LeafThroughによるものと、PDFによるものと、利用者が自由に選択可能になっている。それぞれのソフトの特徴を活用した設計となっている。

6. 検索の基本機能

テキスト（内村鑑三全集）をデジタル化する意義は、ただ読むだけではない。そもそも意図として、検索がある。そのためのデジタル化であった。以下、DVD版全集における検索機能についてのべる。

DVD版全集を起動の後、「検索」のタブをクリックして選択すれば、検索用の画面に切りかわる。

6.1

検索語のウィンドウに、検索したい文字列を入力して、「検索」ボタンをクリックするだけである。その結果は、各用例について一行ずつ一覧表示される。表示されるのは、巻数、ページ数、そして本文の一部である。

マウスで選択した行については、画面下部に、別ウィンドウとして文章が表示される。このとき、検索文字列については、ハイライト表示するようになっている。

さらに、その箇所の本文を見たいとおもえば、クリックすると、先の読書用表示の画面に遷移する。そしてこの場合でも、検索文字列については、黄色でハイライト表示になっている。ここから先の操作は、読書用表示と同じである。ディスプレイでそのまま該当のページを読むこともできるし、また、部分的に拡大して文字を大きくすることもできる。さらには、マウス右ボタンメニューから、PDF表示に移行して、そこから、テキストを引用（コピー）することが可能である。

6.2

検索結果をCSV出力することもできる。これは、検索結果が一覧表示されている状態のときに、「CSV出力」のボタンをクリックするだけでよい。デフォルトでは、「内村鑑三全集検索結果.csv」のファイル名となっている。このファイル名は、保存時に変更可能。（ただし、出力は700件までと制限がある。）

6.3

「全件表示」とあるボタンをつかうと、40件までと制限はあるが、検索結果の読書用表示のサムネール一覧が見られる。これにも、巻数・ページ数の所在表記がある。

7. 検索を便利にする機能

内村鑑三の全集という特殊性に配慮して、DVD版全集の検索機能には、次のような補助的な機能が付加してある。

7.1

文字の同一視機能。

これは、以下の文字を同じと見なして検索する機能である。

- ・異体字：時代により変化した字形を同一文字とみなす（例：汽／瀛、灯／燈等）
- ・英字の大文字と小文字（Zとz等）
- ・拗促音等に使用される小文字のひらがな・カタカナ（あいうえおやゆよアイウエオヤ ュョ）を通常のひらがな・カタカナと見なす。
- ・濁音および半濁音と清音（ハ、バ、パ等）
- ・旧かな（ゐとい、キとイ、ゑとえ、エとエ）
- ・全角カタカナと半角カタカナ
- ・数字、英字、記号の全角と半角

7.2

まさしくこれこそ内村鑑三全集ならではの機能であるが、聖書名同一視の機能がある。

内村の書いた文章には、当然ながら多くの聖書名が出てくる。それも、古い明治時代に日本で使われた聖書である。現在とは表記が異なっている。それらを、まとめて同一して検索可能とする。

たとえば、『ルカによる福音書』（新共同訳名）についてみた場合、『路加伝』『路加』『ルカ伝』『路可伝』『路可』などが、使用されている。この例であれば、「ルカ伝」を検索文字列として入力しても、その他の「路加伝・路加・路可伝・路可」もヒットするのである。

このような処理の対象となっている聖書名はつぎのごとくである。新共同訳の名称と、内村が使用している用例の一つだけを事例としてしめす。

創世記＝創世記 出エジプト記＝出埃及記 レビ記＝利未記 民数記＝民数紀略
申命記＝申命記 ヨシュア記＝約書亜記 士師記＝士師記 ルツ記＝路得記
サムエル記上＝撒母耳前書 サムエル記下＝撒母耳後書 列王記上＝列王紀略上
列王記下＝列王紀略下 歴代誌上＝歴代誌略上 歴代誌下＝歴代誌略下
エズラ記＝以士喇書 ネヘミヤ記＝尼希米亜記 エステル記＝以士帖書
ヨブ記＝約伯記 詩篇＝詩篇 箴言＝箴言 コヘレトの言葉＝伝道之書 雅歌＝雅歌
イザヤ書＝以賽亜書 エレミヤ書＝耶利米亜記 哀歌＝耶利米亜哀歌
エゼキエル書＝以西結書 ダニエル書＝但以理書 ホセア書＝何西阿書
ヨセル書＝約耳書 アモス書＝亜麼士書 オバデヤ書＝阿巴底亜書 ヨナ書＝約拿書
ミカ書＝米迦書 ナホム書＝拿翁書 ハバクク書＝哈巴谷書 ゼファニヤ書＝西番雅書
ハガイ書＝哈基書 ゼカリヤ書＝撒加利亞書 マラキ書＝馬拉基書
マタイによる福音書＝馬太伝 マルコによる福音書＝馬可伝
ルカによる福音書＝路加伝 ヨハネによる福音書＝約翰伝 使徒言行録＝使徒行伝
ローマの信徒への手紙＝羅馬書 コリントの信徒への手紙一＝哥林多前書
コリントの信徒への手紙二＝哥林多後書 ガラテヤの信徒への手紙＝加拉太書
エフォソの信徒への手紙＝以弗所書 フィリピの信徒への手紙＝腓立比書
コロサイの信徒への手紙＝哥羅西書 テサロニケ信徒への手紙一＝帖撒羅尼迦前書
テサロニケ信徒への手紙二＝帖撒羅尼迦後書 テモテへの手紙一＝提摩太前書
テモテへの手紙二＝提摩太後書 テトスへの手紙＝提多書
フィレモンへの手紙＝腓利門書 ヘブライ人への手紙＝希伯來書
ヤコブの手紙＝雅各書 ペトロの手紙一＝彼得前書 ペトロの手紙二＝彼得後書

ヨハネの手紙一＝約翰第一書 ヨハネの手紙二＝約翰第二書
ヨハネの手紙三＝約翰第三書 ユダの手紙＝猶太書 ヨハネの黙示録＝約翰黙示録

8. 検索機能の問題点

以上のように、すぐれた文字列検索機能をもつDVD版全集であるが、若干の問題がないではない。その問題点のいくつかを以下にしめす。

まず、検索の対象となる本文（テキスト）は、全集の本行のみである。先にのべたごとく、まずワープロで作成した書籍版全集を再現したデータがあり、それをPDFに変換する。そのPDFから、さらに検索のためのテキストを抽出し、インデックスを作成してある。

8.1

検索の対象からはルビ（ふりがな）がはずれてしまう。特に、内村鑑三の文章の場合、総ルビに近いものが多い。また、明治期の文章の特徴でもあるが、本行の本文そのものよりも、ルビの方がむしろ「本文」であるというべき性格もっている。

たとえば、「あはれみ」という、おそらく内村鑑三の思想研究にとって重要な意味をもつであろう語についてみても、本行で使用される表記（漢字）としては、「憐憫・矜恤・恤矜・恤・慈悲・憐恤」などがある。仮名表記でも「あはれみ」ともある。

ちなみに、「あはれみ」を検索すると、7件のヒットがある。しかし、「憐憫」で検索をかけると、132件のヒットになる。しかし、この漢字表記の文字列に、どのようなルビが付されているかは、検索結果の画面からは直接には読み取れない。検索の結果を読書用表示で表示して、さらにPDF表示に遷移しないと、ルビつきの、つまり、オリジナルの書籍版全集の本文の状態を見ることができないのである。

8.2

漢字については、次のように処理されている。マニュアルから引用すると、「検索機能については使用できる文字種・字体の範囲を、サロゲートペアを利用する拡張領域を対象としないUnicode3.0のUTF-16としています。一方PDF表示画面では、その字形を可能な限り底本に合わせて、外字を使用しています。従って、上記範囲内の文字を用いて検索ができない一部の漢字については、表示されているものとは別の文字で検索することになり、検索時には、便宜上検索可能なUTF-16範囲内の文字を指定していただくこととなります」とある。

つまり、いくつかの漢字は、ダイレクトに検索できない。たとえば、「妍」と表示され

ている文字については、「姐」の文字を使用しなければならない。ただ、これらは使用頻度としては、そう多くない文字であるし、マニュアルに一覧表示されているので、それを見ることで、概略を把握できる。実用面ではこれは特に問題とならないと思われる。

8.3

むしろ問題であるのは、本文（書籍版全集）との文字の字体の違いである。本文のデータ入力、基本的に、JIS X 0208 によっている。したがって、いわゆる拡張新字体を多くふくむものとなっている。「祈祷」「冒流」などである。

また、JIS X 0213:2004 との間で字体がちがってくるという問題もある。つまり、フォント埋め込みPDFで見えている文字を、自分のコンピュータ環境でワープロやエディタにコピーしたら字体が違ってしまいうという例である。「葛」「昂」などが、それに該当する。

このような字体のコンピュータ環境による問題が発生するのは、現時点の日本の状況ではやむをえないことかもしれない。また、このDVD版全集の企画が、0213:04 の普及（Windows Vista）以前からスタートしていることを考えれば、（それだけ長期間にわたっての難事業であったわけであるが）、やむを得ないとすべきであろう。この問題については、コンピュータ文字についてある程度の知識があれば、あらかじめ予想できる範囲の問題でもあり、現実の運用面において大きな障害となるとは思われない。

8.4

文字の同一視は便利な機能ではあるが、言語研究の視点からはやや問題がある。たとえば、「エ」（清音）と「ヱ」（濁音）が対応するようになっている。しかし、「ルーズベルト」に対応するのは「ルーズベルト」であって、「ルーズエルト」ではない。言い換えれば、「ルーズエルト」で「ルーズベルト」を検索するかどうか、という問題である。

9. 電子出版としての意義

さて、以上、DVD版『内村全集』について、その紹介をしてきた。ここで、視点をかえて、その電子出版としての意義について考えてみたい。

9.1

本稿を執筆しつつある時期は、まさに、電子出版をめぐっておおきな転換点の時期にさしかかっているといえるであろう。たとえば、以下のようなことがらがある。

- ・グーグルブックス（これは、世界的規模で見て）
- ・国立国会図書館の大規模電子図書館構想（日本国内の事例として）

・電子ブックリーダーの登場。Kindle（Amazon）、iPad（Apple）など。

このような、電子書籍、デジタル・ライブラリについては、本稿の範囲をこえた大きな、また、緻密な議論が必要になる。ここでは、本稿であつたDVD版『内村全集』が、全体としてこのような書籍のデジタル化のながれのなかで、どのような位置づけになり、どのような意義があるのか、について、わずかばかりの論考をこころみたい。

このとき、考える観点としては、知的生産のための学術資源としてどうであるか、ということから考えてみることにする。

9.2

画像ではなく、テキストデータとして提供されているということ。しかも、それが、もとの書籍版全集に厳密に対応するかたちで提供されていること、である。

デジタル書籍の学術利用の場合、おそらく一般のデジタル書籍利用と異なる点は、「引用」「典拠」としての確実性である。『内村鑑三全集』を、ただ、読書のために読むという読者もいないではないであろう。しかし、現在の利用状況を想像すると、その利用の多くは学術利用であると推測してまちがいないであろう。その場合、必要なことは、

- ・確実に引用できること（コピーできること）
- ・典拠が明確にできること

これは、学術資料の利用についての、最低限の基本ルールである。逆に言えば、これが確実に保証されないでは、（その内容の質とはまた違ったレベルで）、学術利用のための資料としては問題があるといつてよい。

DVD版全集の場合、書籍版全集を基本的に忠実にデジタル化したものであるもので、論文を書くなどの際、何巻の何ページ、何行目であるか、確実に特定できるし、それは、まさに、書籍版全集という紙の固定されたものによって保証されている。

これをさらに言い換えるならば、デジタル書籍について、その「引用」「典拠」についてどこまで確実に特定可能であるのか、という問いかけにもなる。この観点は、意外と重要な点であるように筆者には思える。検索などの面で、デジタル書籍が便利だからといってつかっても、その引用の典拠を明示するために、また改めて書籍版と対応関係をさがすという、あまり意味のないことに労力をついやさなければならないことにもなりかねない。あるいは、書籍版の無い（Born Digital）のものについては、どのような典拠明示の方法が、現実的にありうるのか、また、その方法が共有化できるのか、問題となる。

書籍の学術利用という側面からみたとき、DVD版内村全集の方法は、非常に価値のあ

る方法ということができよう、

9.3

自由にコピーができること。これも上述の、引用・典拠の確定とならんで、非常に重要なことである。

DVD版内村全集は、もし、いまから企画をたてるなら、画像データとして書籍版全集をスキャンする方法をえらぶ、その選択肢もある。(たとえば、国立国会図書館の近代デジタルライブラリーなどの電子図書館が、この方式であるように。)

これには、一長一短がある。もとの書籍の版面レイアウトや、どのような組版がされているのかを見ようと思えば、画像データ方式の方がすぐれている。一方で、「引用」ということを考えた場合、画像データではどうにもならない。DVD版内村全集のように、フォント埋め込みPDFでも同時に本文のデータが提供されているならば、そこから、任意に、パソコン上で、必要箇所をコピーすることができる。それも、正確・忠実に、である。

これは、知的生産のツールとしてみた場合、コピー可能(さらにいえば、再編集可能)という、きわめて合理的な方向に道をひらくものである。知的生産のために貢献するものである。

いうまでもないことであるが、人文学系の研究、それも文献資料に依拠した研究は、検索と引用、これをいかにスムーズにできるか、ストレス無くパソコン上で実現するか、に、研究支援ツールとしての価値がある。この意味では、DVD版内村鑑三全集は、この文字列検索とその結果を本文データから正確に引用(コピー)できるという意味で、まさに、研究者の視点にたった利用を考えてあるといえよう。

9.4

検索と読書が両立していることも見逃せない。これは、上述の観点とも深く関連するが、人文学研究におけるテキストデータ利用の最大の目的は検索にある、といってもいいだろう。だが、ここで、検索のみを重視すると、検索結果の文字列が(たとえば、KWICのように)出力するだけでおわってしまっは意味がない。検索する語は、あくまでも原文のなかでどのようにつかわれているか、そこのなかにもどして改めて吟味するという作業が必要である。特に、内村鑑三など思想史にかかわる研究では、語の検索結果をリストにただけでは意味がないともいえる。原テキストの文脈のなかにもどして、再度、原文を読む必要がある。これは研究の方法論の基本である。

検索結果の一覧表示から、読書用画面へ、さらにPDF表示。そして、必要なら、そこ

からコピーして、自分のワープロなりエディタなりにて移行して、論文が書ける。そして、必要ならば、もとのオリジナルの書籍版全集の該当箇所を読むこともすぐにできる。いや、このDVD版全集を本当につかいこなそうと思うのであれば、書籍版全集は必須であるといえよう。

デジタル書籍は、旧来の紙の書籍と対立したりする、それを駆逐するものとしては、ここでは存在していない。むしろ、書籍版全集を、よりよく読むための、支援ツールとして、デジタルのDVD版全集がある、そのような位置づけになっている。

9.5

これは、スペックとは関係ないが非常に重要な点として、ISBNのついた形式で、書店の流通ルートにのって販売されるものである、ということも指摘しておかなければならない。現実には、まだ、書籍版全集(第二刷)があるなかで、その編集にかかわる権利をクリアしたうえで、妥当と思われる価格での市販は、ある意味で重要な意義がある。そして、特別に、コピーのプロテクトなどかけてはいない。個人利用であれば、自分のパソコンのHDにコピーして、自由に利用できる。

価格的にも妥当であると筆者は判断する。現時点で、書籍版全集の古書価格とほぼ同額で、自由に購入可能というのは、ある意味で、非常にオープンな姿勢であると判断するものである。

10. デジタル書籍の一つのあり方

以上のように考えてみた場合、DVD版内村鑑三全集は、書籍版全集があつてこそ、その価値があるということがわかる。また、同時に、書籍版全集も、DVD版全集があるからこそ、これまででできなかった読み方(研究)が可能になる。双方が、それぞれに価値を高めあうものであるという点で、非常にすぐれた企画であるというべきであろう。

しかし、これも、まず、内村鑑三という人の存在、編年編集を基本とした書籍版内村鑑三全集の存在、を前提にしなければならない。このモデルが、他の、書籍のデジタル版にあてはまるということは無いであろう。あくまでも、内村において特殊なこととして、まずとらえておくべきであろう。

そのうえで、このデジタル出版のモデルが、適応可能なものは、内村鑑三以外にもまだあるのではないか。ここで、その具体名を出すのは避けることにするが、日本近代の文学者・思想家の全集においては、可能性のあるものがいくつかあるように思える。このDVD版内村鑑三全集をモデルとして、次なるころみながなされることに期待するもの

である。

11. 今後の課題

やっと完成したDVD版全集であるが、これで完成というわけではない。今後の課題についていくつか考えてみたい。

11.1

このDVD版全集をつかって、何ができるのかの研究。

実は、この点が一番重要であるにもかかわらず、あまり進んでいない。実際に使ってみて、種々の問題点を考えることが次の段階にむけて重要である。

11.2

文字列の検索から、語の検索へ。

現時点でのDVD版全集では、検索しているのは、文字列である。コンコーダンスとして、語の検索機能は持っていない。また、文字列を検索した場合でも、検索対象文字列がページの切れ目にまたがってしまっている場合は、検索対象から落ちてしまうという欠点をもっている。少なくともこれはどうにかしなければならない。

さいわいなことに、PDFからは、プレーンな文字列データを取り出すことは可能である。また、ルビ（ふりがな）もPDFのなかには埋め込んである。今後の課題としては、これをつかって、再加工して、自然言語処理の技術をつかって、本格的な内村鑑三全集コンコーダンスの作成である。

この意味では、DVD版全集の完成は、その第一歩を踏み出した段階、しかし、非常に堅固な第一歩であるということができよう。

11.3

編集とは何かという視点。

書籍版全集は完全な編年編集である。したがって、内村が生涯をかけて刊行をつづけた『聖書之研究』が、まとまって収録されていない。すくなくとも個人的に利用の範囲内であるならば、PDFを再編集して、『聖書之研究』だけをとりだすことも簡単である。また、そのプリントアウトもできる。

この意味では、個人の全集を編纂するとき、デジタル技術が加わることによってどのようなことが可能になるのか、非常に大きな問題があるいえよう。ただ、このことについては、編集文献学の視点を参照しつつ、今後の課題としておきたい。

謝辞

本稿でとりあげたDVD版全集にたずさわった、内村鑑三全集DVD版出版会の方々に感謝いたします。

参考文献

- 1) 當山日出夫, 2009, 「『内村鑑三全集』 デジタル版の文字処理について」, 『東洋学へのコンピュータ利用 第20回研究セミナー』(京都大学人文科学研究所), pp.5-18
- 2) 當山日出夫, 2009, 「アーカイブズにおける文字と文字コード」, 日本アーカイブズ学会, 2009年度大会自由論題研究発表会(2009年4月26日, 学習院大学)
- 3) 當山日出夫, 2009, 「デジタル文字の共有と継承について」, 『情報処理学会研究報告 2009-CH-81』(東京大学)
- 4) 當山日出夫, 2010, 「DVD版『内村鑑三全集』について」, 『東洋学へのコンピュータ利用 第21回研究セミナー』(京都大学人文科学研究所), pp.9-22