

《老子》五种版本网络检索版及相关问题¹

洪华清*
南洋理工大学

陶红印**
洛杉矶加州大学

内容提要：古籍检索的电子化目前还面临很多亟待解决的问题，如软件工具如何与传统索引结合、资源如何共建共享、相关领域的标准制定以及技术更新的影响、技术迁移等等。本文介绍了一个网络版老子文本检索系统以及我们在探索处理五种老子版本的电子化、数据库的建设以及网络检索的设计和应用等方面的一些尝试。希望能对古籍的整理出版以及相关学术研究带来一定的促进作用。

关键词：老子、语料库、逐字索引、人文电子化

一、老子研究与老子版本的选择

老子及其哲学观点是中国思想文化史的核心部分，对其文本和哲学思想的研究也已有了两千年以上的历史。老子研究和其他经典研究一样，都有不同学派和手段之争。清代乾嘉学派注重文字释读，跟西方所谓的语文学（philology）方法有相通之处。从现代学术研究的角度来看，重视文本自身的奥义，研究同一篇章的版本的流通及演变，研究不同篇章之间的互文关系（intertextuality）等，既是理解文本和篇章的必要步骤，也是学术研究的哲学理念。这也是我们参与的“老子文本两千年之演变及相关简帛资料研究”这项研究课题的一个基本研究取向。²

在基于文本的老子研究中，刘笑敢的《老子古今：五种对勘与析评引论

*新加坡南洋理工大学国立教育学院专职研究员。（电邮：huaqing.hong@nie.edu.sg）

**洛杉矶加州大学亚洲语言与文化系及应用语言学系教授。（电邮：tao@humnet.ucla.edu）

¹ 本研究计划为香港中文大学刘笑敢教授所主持的香港政府大学教育资助委员会资助的项目「《老子》文本两千年之演变及相关简帛资料研究」之一部分，并受到洛杉矶加州大学「数字人文计划」对陶红印的资助。在数据加工和数据整理方面我们得到了郑健珍（新加坡南洋理工大学）和林尔嵘（洛杉矶加州大学）的研究助理。作者感谢刘笑敢教授提供的《老子》电子文本的原始材料以及对这个项目的提出提出的多个建议。文中任何错误皆为我们个人的责任。

² 参看 Liu, Xiaogan: “From bamboo slips to received versions: Common features in the transformation of the Laozi,” *Harvard Journal of Asiatic Studies* 53.2 (2003): 337-382。刘笑敢：《老子古今：五种对勘与析评引论（上、下卷）》（北京：中国社会科学出版社，2006）等。

（上、下卷）》是具有开拓性质的专著。该书对勘了《老子》的五种有代表性的版本，即 1）1993 年发现于湖北荆门的郭店竹简本；2）1973 年发现于湖南长沙的帛书；3）正统道藏的（唐代）傅奕本；4）河上公注本；5）王弼注本。并对诸本文本的异同和流变作了逐字逐句的分析比较。更为可贵的是，该书还把这五个版本的每个文句的不同版本排列在一起，编成索引，为研究者提供了极大的便利。

传统的出版技术当然有其无可替代的作用，不过随着电脑技术的日益发展，许多电子版本的优势也逐渐为我们认识。³这里报告的《老子》文本电子化的项目就是试图在这些方面做一些进一步的尝试。

本计划的主要目标有三个。第一，文本电子化；第二，在文本电子化的基础上建立关系化的数据库；第三，为研究者提供一个最大程度上便利的检索平台；第四，为研究者提供一些只有电子化手段才能提供的研究材料。我们认为这四个目标及其实现使得这项计划能够超出一般文献电子化项目的目标和服务范围。

从经典文献电子化的角度来看，《老子》电子化计划也有其特殊的意义。一般的文献电子化计划常常涉及一个版本，内容比较单一，这项计划涉及多个版本，虽然《老子》的多项版本在整体结构方面有统一的地方，但其间细节差异有时相当大。⁴这就为电子化的努力带来了新的挑战。本文报告的方案应该对其他类似的项目有一定的启示作用。

二、电子版设计原则及方法

我们这里采用的电子化版本的底本完全以刘笑敢《老子古今》所列为准，包括所有的文句、字型释读以及语句排列的顺序等。

如前所说，本计划的主要目标有四个：文本电子化、数据索引建库、便利的检索平台以及使用电子化以后的新研究材料。下面依次说明。

2.1 文本的电子化

文本的电子化在电脑极为普及的今天已经不是什么创新之举。不过本项目需要应对的特殊问题是早期文本中的偏僻字或未定形及未再用的汉字。这个问题在竹简本和帛书本中尤为严重。目前流行的解决办法有自造字体、用户安装。这个办法无疑会给用户带来不便，因为并不是每个用户都了解自造字体或者熟悉字体安装所需要的程序。由于用户操作系统的不同，自造字体的安装也常给用户带来很多困惑。对开发者来说，自造字体也是一件非常费时费力的工作。而且，确定在用户的机器上能准确显示难检字符只是解决了问题的一个方面，用户是否能够输出这些自造的字体以便检索就是另一方面的问题了。考虑到这些因素，在文本电子化的过程中我

³ Tao, Hongyin, Bao Zhiming, Liu Xiaogan, and Zhang Min. "Building an Indexical Classical Chinese Philosophical Dictionary Based on Electronic Corpora." Proceedings of the International Conference on Chinese Computing '96. Institute of Systems Science, National University of Singapore. 1996: 165-171.

⁴ 参看刘笑敢：《老子古今：五种对勘与析评引论（上、下卷）》。

们采取了安装大字符集和难检字图片化这两种方法相结合的解决办法。

首先，我们把从事文史工作尤其是古籍整理的学者常用的中日韩汉字超大字符集（SuperCJK）通用字体支持包推荐给用户安装。该大字符集是通用于 Windows 98/NT/2000/ME/XP/2003/Vista 等系统的 Unicode 字体支持包，能较为理想地实现国际标准超大字符集内全部七万多中日韩汉字的同台显示。拥有了这套中文处理利器，即可解决一般电脑中有缺字或打不出生僻字的难题。

其次，在大字符集的基础上，仍然有很多难检字无法在普通电脑里正确显示。这里所说的难检字指的是那些即使用户安装了超大字符集仍然没法正确显示的特殊字（如下图中的例子），也就是我们所说的偏僻字、未定形字以及未再用的汉字。这也是所有古籍电子化所面临的一个问题。经过讨论，我们认为把这样的特殊字符变成图片是一个简易的办法。考虑到用户没有办法直接输入图片来检索，而且图片在数据库索引和检索方面（参下）也会有困难。我们的解决办法是结合图片和内部编码索引，使检索图片变得可能。这样的结果使得用户不必安装自造字体也可以阅读并检索难检字。如下图所示。为了方便用户检索这些特殊字符，我们也把这些难检字索引列表显示，用户只需点击相应的图片即可检索。



2.2. 数据索引和建库

要建立便利的检索平台，以数据库为基础的动态检索方法是目前一个较为流行的办法。该方法的优点是，用户不必受限于语料库制作者所设定的词串类型而可以根据自己需要作各类检索。

数据库的语言有很多种，本着简单易用的方针，我们在这里采用了 MySQL。简单说来，所建的数据库中主要有两个表。一个是 ts_texts 表，用来存储文本信息。这个表中的内容主要是根据输入的文本文件来得到，其中的 c_setid 是从用户提供的 set.xls(x) 集合文件中得到。它的结构如下：

表一、文本信息

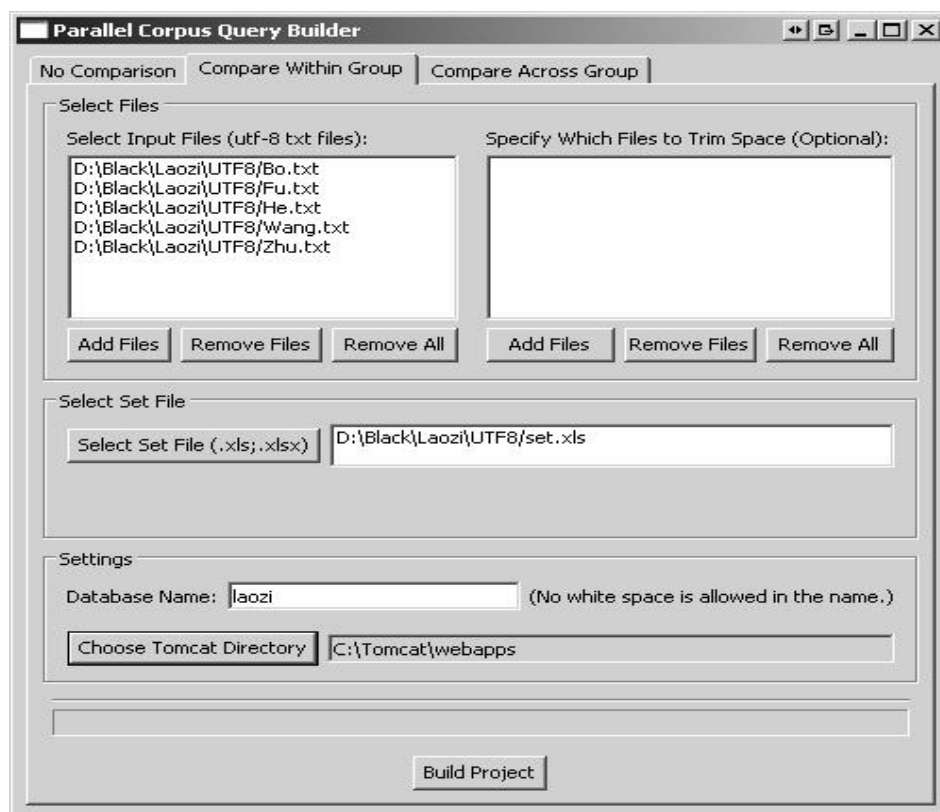
列名	数据类型	是否主键	注解
c_id	bigint(20)	是	唯一区分每条记录
c_turnno	varchar(200)	否	Turn number, 用来区分每个turn
c_sentence	blob	否	每个turn的内容
c_filename	varchar(200)	否	当前文件名
c_setid	varchar(200)	否	区分属于哪个set

另一个表是 ts_set, 用来存储 set 的信息。这个表中的内容从用户提供的 set.xls(x) 文件中得到。它的结构如下:

表二、集合文件信息

列名	数据类型	是否主键	注解
c_id	bigint(20)	是	唯一区分每条记录
c_setid	varchar(200)	否	每个 set 对应一个 id
c_member_turnno	varchar(200)	否	记录属于同一个 set 的文件的 turnno

为了方便自动处理同类文本, 我们综合考虑了语料文本的不同和潜在用户的需求的不同, 开发了一个多功能平行语料库数据库索引建站通用工具包 (Parallel Corpus Query Builder)。该程序包给用户提供了方便的图形界面 (见下图), 比较简单易用。用户只需把准备好的文本和 set 信息文件等选择好, 然后输入欲建立的数据库名称。这时程序就会自动启动 MySQL, 根据用户输入的数据库名称自动创建数据库, 然后根据输入文件的内容往数据表中插入数据完成自动索引建表, 同时生成相应的网页界面、字体设定、框架结构、javascript 等文件。程序运行完成后, 用户只需开启 Apache Tomcat 服务器和网页浏览器即可开始检索。在用户检索时, 程序会根据用户的检索内容自动生成 sql 语句, 发送到数据库中进行检索, 然后将检索结果返回到网页结果页面给用户浏览。我们将在下一节介绍这样的网页检索的几个优点。



2.3. 便利的检索平台

从便利用户的角度来说，目前最为便捷的用户接口界面应首推互联网网页浏览器。浏览器目前已经是互联网上最为流行的一个工具，绝大多数用户都应该不会对此陌生。为达到通用性，我们尽可能兼顾用户使用不同网页浏览器的情况使得检索可以在不同的浏览器（如 IE、Firefox 等）下正确运行。因此，检索平台和不同浏览器的兼容性是我们优先提供给用户的一个特色。

其次，字体编码格式的兼容也是我们着力解决的一个问题。众所周知，目前海外用户大多使用英文视窗，中国大陆和新加坡用户大多使用简体中文系统，而港澳台用户大多使用繁体中文系统。为了尽量减少不同系统的用户在浏览和检索语料时的种种不便，我们在处理语料和网页时采用了 UTF-8 作为默认编码。这样做也能够最大可能地使前文提到的中日韩汉字超大字符集（SuperCJK）通用字体包发挥作用，正确显示汉字。另外，网页的界面语言也采用 UTF-8 繁体中文和英文说明，方便用户使用。

就检索本身来说，网页首页界面的设计把几种可能的检索方法都列在上面，用户可以自由选择。当一个检索界面打开后，主界面的选择仍然在页面上右方置顶保留，为用户提供了方便的退回功能，而不必把各个次界面逐一关闭。如用户可以根据自己需要的检索项输入检索，也可以打开网页上提供的特殊字列表检索，还可以打开预制好的两个词语列表（一个按出现的频率排列，另一个是按词语的长短排列）选择词条后敲后面的 query 即可进行检索（具体操作步骤见第 3 节。）

另一个便利措施是，为方便不同用户系统语言和输入语言的不同，我们还设置了一个繁简字自动匹配转换的机制。当用户输入简体汉字符发送检索后，该机制将在背后运行，自动识别用户输入的字符并转换成对应的繁体字符进行检索，然后返回结果给用户。同样，繁体汉字的检索也能自动转换成对应的简体汉字来输出结果。

2.4. 电子化后提供的新的研究材料

电子化以后除了可以在数据库的基础上进行动态检索外，还可以提供什么新的研究材料？至少一项很重要的材料是词串及不同词串的频率。在检索界面中，用户可以看到两个词表，一个是单字词表，另一个是词串词表。词串词表是使用汉语自动分词工具生成的，无人工干预。这样做的好处是，机器可以不以人的先入之见而纯粹根据频率客观地把常见的共现语词找出。下面是几个常见的四字组和三字组词串及其频率。

慎終如始	4
慎終若始	2
民不畏死	2
深根固柢	3
深根固蒂	1
玄之又玄	4
絕仁棄義	4
絕聖棄智	3
美言不信	4
被褐懷玉	2
金玉滿堂	2
長生久視	5
不可不	3
不可以	8
不得已	7
不知足	5
之所以	7
從事於	5
以至於	4
大丈夫	4

从上面的列表我们可以注意到，除了老子里边的名言如“玄之又玄”、“根深固蒂”之外，还有一些虚词性的成分如“不得已”“之所以”“以至于”等，也一览无余。有可能为我们比较不同版本的语言特征起到一定的作用。而这一类的语言现象一般的检索文本是不会提供的，即使研究者一般也不会去注意（尤其是那些含虚词语类的成分组合及其频率）。

字词检索更是提供了多种印刷文本上不容易看到的信息，例如只在某一个版本

上出现的字词以及音序、频率排列检索方法等（详情见下节）。

五種版本篇章對比的程序使得版本之间的对比一目了然。版本之间的用词（甚至后来插入的标点符号）的差异可以通过不同颜色显示出来（详见下节最后部分）。

检索平台上提供了顺序显示（Concordance）和对应显示（Parallel/Set View）两种浏览检索结果的方法。顺序显示采用了语料库检索结果展示的通行的 KWIC（Key Words in Context）方法，把关键词（检索项）的所有结果列印在语境的中间，从而使研究人员能充分利用语境来审视检索项的使用情况和搭配规律。与此互补，对应显示极大地方便了研究人员逐一审视不同版本之间的异同，从而把繁琐的手工对应工作变成了简单有效的对比研究，数据信手可得。如从下面的检索结果我们可以看到，“道”字在五種版本的老子里共有 71 组对应的检索结果。就第 7 组结果来说，虽然帛书本、河上本、王弼本和傅奕本各有一句，但是这四个版本中的语句有多种变异。另外，这一句在竹简本中也没有记录。

7	# ¹ Null	竹簡_Z_77_01
	# ¹ 天之道，猶張弓也，	帛書_B_77_01
	# ¹ 天之道，其猶張弓乎？	河上_H_77_01
	# ¹ 天之道，其猶張弓與！	王弼_W_77_01
	# ¹ 天之道，其猶張弓者歟，	傅奕_F_77_01

由上可知，利用网络检索程序，可以对电子文本文件进行多种参数分析，统计字词的出现次数和出现频率，分析字词的使用情况，还可以分析语句的结构特征。人力难以胜任的工作在电子手段下变得非常容易。文本的语言文字分析可以显示一篇文字在行文措词方面的特征，对研究文本内容很有帮助，而版本之间的对比既可以做到宏观、全面，也可以达到细致入微。

三、网络电子版的使用

用户可以用任何一个网页浏览器浏览《老子》五種版本檢索网址：

<http://score.crpp.nie.edu.sg/laozi/index.htm>

用户目前可以在主页上看到系统提供的四大检索方式。1)《老子》五種版本平行檢索 (Laozi Parallel Corpus Online Query); 2)《老子》五種版本字詞使用頻率索引 (Index of All Characters); 3)《老子》五種版本全文索引 (Index of Full Texts); 4)《老子》五種版本篇章對比 Text Comparison in Detail

第一种检索方式是“五種版本平行檢索”(Laozi Parallel Corpus Online Query)。这个检索方法的目的在于便于用户比较同组字句在各个版本中的出现情况。初始状态下系统自动选择五种版本同时进行比较，但是用户可以通过打勾选择任何几个版本进行比较。

SEARCH（檢索項）給用戶提供了輸入的窗口。但是用戶也可以利用其他方式輸入，包括：Special Character (特殊字符)、Word List A (詞語列表一)和 Word List B (詞語列表二)。用戶打開任何一個連接都可以進一步通過鼠標做出選擇。

上一節已經提到，檢索平臺上提供了順序顯示（Concordance）和對應顯示（Parallel/Set View）兩種瀏覽檢索結果的方法。這兩種瀏覽檢索方式只需要用戶用鼠標選擇即可，不需要另外輸入檢索詞條。下面兩圖顯示它們的位置：

順序顯示：



The screenshot shows a web interface for searching. At the top, there is a section titled "SELECT TEXTS (選擇文本)" with five checkboxes, all of which are checked: "竹簡本", "帛書本", "傳奕本", "河上本", and "王弼本". Below this is the "SEARCH (檢索項):" section, which contains a text input field with the character "道". Underneath the search input are three links: "Special Character (特殊字符)", "Word List A (詞語列表一)", and "Word List B (詞語列表二)". Further down is the "File Info Window Size:" section, which includes a label "(定義文本信息範圍):" followed by a dropdown menu showing the number "2". Below that is the "Results per page:" section, which includes a label "(指定結果頁面行數):" followed by a text input field containing the number "80". At the bottom of the interface are two buttons: "Concordance (關鍵詞順序顯示)" and "Sentence Parallel (句子對應顯示)". An arrow points from the left towards the "Concordance" button.

例如：“上”的順序顯示結果：

query item (共找到82例当前检索项).

Left Context 左边语境	Query Item 检索项	Right Context 右边语境	File Info 文本信息
	大	上	下知有之，
	銛為	上	，弗美也。
	故吉事	上	左，喪事上右。
故吉事上左，喪事	上	右。	竹簡_Z_31_09
	上	將軍居右，	竹簡_Z_31_11
	上	士聞道，勤能行於其中。	竹簡_Z_41_01
	上	德如谷，大白如辱，	竹簡_Z_41_07
其在民	上	也，以言下之。	竹簡_Z_66_06
其在民	上	也，民弗厚也；	竹簡_Z_66_07
不	上	賢，使民不爭。	帛書_B_03_01
	上	善如水。	帛書_B_08_01

Result Pages: 1 2 3 4 5 6 7 8 9 Next

对应显示：

SELECT TEXTS (選擇文本)

☒ 竹簡本

☒ 帛書本

☒ 傳奕本

☒ 河上本

☒ 王弼本

SEARCH (檢索項) :

道

Special Character (特殊字符)

Word List A (詞語列表一)

Word List B (詞語列表二)

File Info Window Size:

(定義文本信息範圍): 2

Results per page:

(指定結果頁面行數): 80

Concordance

(關鍵詞順序顯示)

Sentence Parallel

(句子對應顯示)



参看上一节“道”的对应显示结果。

如果一个版本缺乏某些句段，系统会显示 Null（缺值）来对此加以提示。

在选择难检字(特殊字符)进行检索的时候，用户只需点击与图片对应的 Query 选项，再点击顺序显示（Concordance）或句子对应显示（Sentence Parallel），检索结果就出现在结果窗口里（有关不同窗口的选择问题，请参看本节最后一段的说明）。

检索系统给用户提供了检索结果显示页面的灵活性。为了方便页面快速登载，检索的结果页面采取了分页显示。考虑到不同用户的使用习惯和对结果页面的再处理要求，左边的选项区也很人性化地提供了定义每页显示多少结果行的功能。用户可以输入数字规定在一个页面上显示多少笔检索结果。假如想要把某个检索项的结果页面整体保存下来，可以把这个数字改大，这样就可以在一个页面里存下该检索项的所有记录了。例如：



SELECT TEXTS (選擇文本)

- ☒ 竹簡本
- ☒ 帛書本
- ☒ 傳奕本
- ☒ 河上本
- ☒ 王弼本

SEARCH (檢索項):

上

Special Character (特殊字符)
Word List A (詞語列表一)
Word List B (詞語列表二)

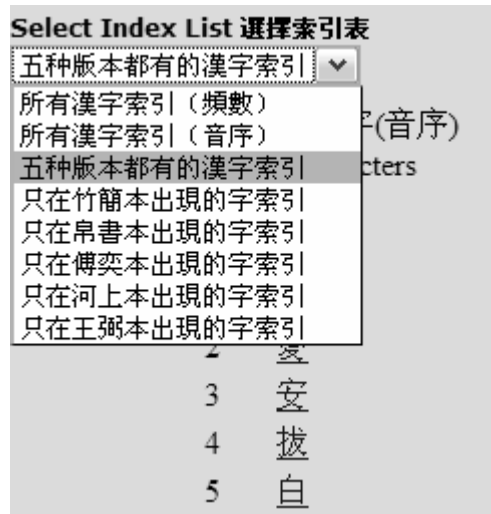
File Info Window Size:
(定義文本信息範圍): 2

Results per page:
(指定結果頁面行數): 50

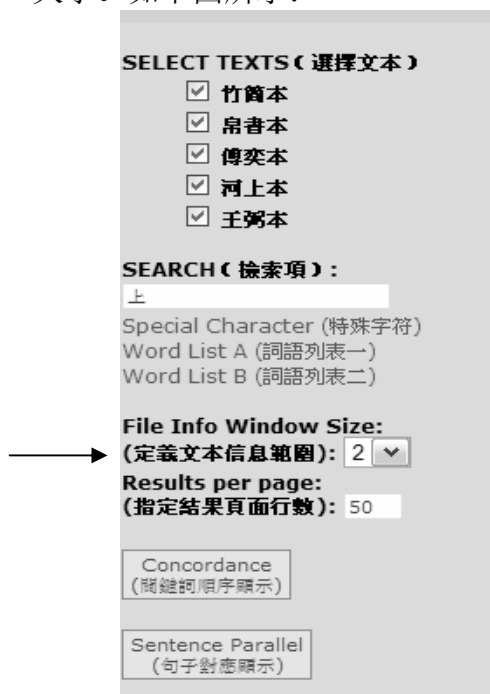
Concordance
(關鍵詞順序顯示)

Sentence Parallel
(句子對應顯示)

第二，五種版本字詞索引（Index of All Characters）。这个索引是以字词的频率和音序排列的。用户可以有多种选择，包括：所有版本中的汉字频率索引、所有版本中的汉字音序索引、只在五个版本中的任何一个版本中出现的汉字索引等。把不同版本中的字词单列索引，可以方便研究者从字词使用的变化上考察版本流变的情况。



如果要察看上下文，用戶也可以點擊“文本信息 (File Info)”由此得到原始文档的上下文信息。文本信息默认只给出检索项所在句子前后分各两句（即默认值为 2）。用户可以点选左边的定义范围的数字来更改要显示的语境的大小。如下图所示：



第三．五種版本全文索引 (Full Text Index)。这里把五种版本的全文排列出来，用户可以根据不同版本来选择显示原文文本。

Select Texts for Concordances 選擇文本
☒ 竹簡 ☒ 帛書 ☒ 傳奕 ☒ 河上 ☒ 王弼
Results per page 檢索結果每頁的行數: 30
File Info Window Size 文本信息範圍: 2

Select a Version to View 選擇要顯示的版本
☐ 竹簡 ☐ 帛書 ☒ 傳奕 ☐ 河上 ☐ 王弼

F_01_01 道可道，非常道。
 F_01_02 名可名，非常名。
 F_01_03 無名，天地之始。
 F_01_04 有名，萬物之母。
 F_01_05 故常無欲，以觀其妙：
 F_01_06 常有欲，以觀其微。
 F_01_07 此兩者同出而異名，同謂之玄。
 F_01_08 玄之又玄，衆妙之門。
 F_02_01 天下皆知美之為美，斯惡已。
 F_02_02 皆知善之為善，斯不善已。
 F_02_03 故有無之相生，難易之相成，
 F_02_04 長短之相形，高下之相傾，
 F_02_05 音聲之相和，前後之相隨，
 F_02_06 是以聖人處無為之事，

更有用的是，面对网页左边的全文文本，用户可以点击文中的任何一字而得到五种版本的检索结果。例如，点击左边的任何一个“可”就可得到下图右边的结果。而且，用户也可以点击结果中的任何一个字词，从而得到一个新的逐字索引。这也是全文索引的一个优点。

<傳01.01> 道可道，非常道。
 <傳01.02> 名可名，非常名。
 <傳01.03> 無名，天地之始。
 <傳01.04> 有名，萬物之母。
 <傳01.05> 故常無欲，以觀其妙：
 <傳01.06> 常有欲，以觀其微。
 <傳01.07> 此兩者同出而異名，同謂之玄。
 <傳01.08> 玄之又玄，衆妙之門。
 <傳02.01> 天下皆知美之為美，斯惡已。
 <傳02.02> 皆知善之為善，斯不善已。
 <傳02.03> 故有無之相生，難易之相成，
 <傳02.04> 長短之相形，高下之相傾，
 <傳02.05> 音聲之相和，前後之相隨，
 <傳02.06> 是以聖人處無為之事，
 <傳02.07> 行不言之教。

可 160			
No. 序号	Left Context 左边语境	Headword 索引词	Right (右边)
1		道，可	道也，非恆道也。
2		名，可	名也，非恆名也。
3		不見，可	欲，使民不亂。
4		揣而允之，不可	長葆也。
5		若，可	以託天下矣；
6		女，可	以寄天下矣。
7		三者不可	致詰，
8		尋尋呵不可	名也，
9		微妙玄達，深不可	識。
10		夫唯不可	識，
11		上可所田 亦不可	以不田上

当然，如果用户不打算检索或比较所有五种文本，同样可以在左上角的文本选择框里做出相应选择。

最后一种检索方式是五種版本篇章對比 (Text Comparison in Detail)。这个功能可以把任何文本之间的差异用彩色显示出来。

逐章對比
Chapter-by-Chapter

第01章

第02章

第03章

第04章

第05章

第06章

第07章

第08章

第09章

第10章

第11章

第12章

第13章

第14章

第15章

第16章

第17章

第18章

第19章

第20章

第21章

第22章

第23章

第24章

第25章

第26章

第27章

第28章

第29章

第30章

第31章

第32章

第33章

第34章

第35章

第36章

第37章

第38章

第39章

第40章

第41章

第42章

第43章

第44章

第45章

第46章

第47章

第48章

第49章

第50章

第51章

第52章

第53章

第54章

第55章

第56章

第57章

第58章

第59章

第60章

第61章

第62章

第63章

第64章

第65章

第66章

第67章

第68章

第69章

第70章

第71章

第72章

第73章

第74章

第75章

第76章

第77章

第78章

第79章

第80章

第81章

全文對比
Text-by-Text

傳奕與王弼

傳奕與河上

傳奕與帛書

傳奕與竹簡

回到主頁 Back to Home

傳奕 vs. 王弼

1 01_01

道可道，非常道。

2 01_02

名可名，非常名。

3 01_03

無名，天地之始。

4 01_04

有名，萬物之母。

5 01_05

故常無慾，以觀其妙。

6 01_06

常有欲，以觀其微。

7 01_07

此兩者同出而異名，同謂之玄。

8 01_08

玄之又玄，眾妙之門。

傳奕 vs. 河上

1 01_01

道可道，非常道。

2 01_02

名可名，非常名。

3 01_03

無名，天地之始。

4 01_04

有名，萬物之母。

5 01_05

故常無慾，以觀其妙。

6 01_06

常有欲，以觀其微。

7 01_07

此兩者同出而異名，同謂之玄。

8 01_08

玄之又玄，眾妙之門。

傳奕 vs. 帛書

1 01_01

道，可道也，非恆道也。

2 01_02

名，可名也，非恆名也。

3 01_03

無名，萬物之始也。

4 01_04

有名，萬物之母也。

5 01_05

故恆無欲也，以觀其妙。

6 01_06

恆有欲也，以觀其所微。

用户可以选择任何两个版本的任何一个章节进行比较。

四、光盘版的使用

为方便在没有网络连接的情况下也可以检索，用户也可以考虑使用光盘版的五種版本老子网络检索程序包。对光盘版感兴趣的读者可以在主网页下载光盘映像或直接跟我们联系索取。用户可以使用通用的光盘烧录程序如 Nero Burning Rom、Sony Record Now、Smart DVD/CD Burner、Roxio Easy Media Creator 等把下载好的光盘映像（ISO 文件）烧录成普通光盘即可使用。我们也打算在网页上随着网络版的更新而及时更新光盘映像。

光盘版和网络版在数据库内容、检索界面等方面没有任何差别，主要差别就是网络版是在开发者的服务器上运行，用户需要用互联网来联机使用；而光盘版是在用户自己的计算机上运行，用户不需要连接互联网。使用时用户也不需要安装任何程序，只要插入光盘，程序就会自动运行，检索界面即可弹出，检索办法和前面介绍过的网络版检索办法完全一样。当然，对于高级用户来说，光盘版还有一个优点，就是用户可以根据情况对检索界面进行定制和修改。不过，对于数据库本身的修改普通用户是没有权限的。

五、结束语

古籍资源电子化，是古籍整理工作发展的必然趋势，前景十分远大。电子化的

13

古籍在利用方面较传统古籍有许多优点，除了简单方便、省时省力、准确性高、传输快捷等因素之外，还有不易漏检、能穷尽、能统计等传统古籍研究很难或者根本不可能有的优点。⁵尽管如此，古籍索引编制的电子化还面临很多亟待解决的问题，如软件工具与传统索引的结合、资源的共建共享、相关领域的标准制定以及技术更新的影响、技术迁移等等。⁶

在这篇短文里我们介绍了自己的一点实践和经验。在处理五种老子版本电子化、数据库的建设以及网络检索的设计和应用方面我们也总结出了一些比较行之有效的办法。例如，在文本电子化时要既要考虑到文本本身的情况，也要照顾到不同用户的不同需求，从而选取有效的办法加工和处理数据。具体来说，在字体的选择、工具的开发、界面的设置、功能的选取、使用的便利以及研究的重点等方面都需要做全盘的考虑。

老子五种版本的网络检索版可以看作是一个特殊语料库，通过这个样本我们看到古籍电子化为研究者可能带来的诸多便利。希望我们在这次实践中所碰到的问题以及探索出来的解决问题的一般办法以及开发的通用工具将会对古籍整理出版以及学术研究带来一定的有利作用。

应该承认，古籍电子化虽然已经受到广泛关注，但目前古籍电子化的工作还是刚刚起步，大规模的数据库也是近年才陆续面世，方兴未艾的势头展现了多方位的广阔前景。⁷然而，经典文献的长期利用与技术更新却不会简单地随着数据库的建立而得到推广。这还需要文献专业研究人员和计算机技术人员等多方面的互动和努力，从而改善数据库的设计，完善检索的方法，最终达到充分利用电子语料库促进学术研究的目的。

⁵ 张尚英：《试论古籍电子化的理想模式》，《中国典籍与文化》2001年4期。

⁶ 参看毛建军：《古籍索引的电子化实践》，《中国索引》，2006年4期。宋继华、胡佳佳、孟蓬生、王宁：《古今汉语平行语料库的语料构建》，《现代教育技术》，2008年第1期。蔡奇林：《巴利数字文献资源现况评述与未来展望——兼谈「初期佛教圣典多语本平行语料库」的几点构想》，《中华佛学学报》2006年19卷，页107-176。

⁷ 参看王桂平：《我国古籍数字化的现状及展望》，《图书情报知识》2000年4期。朱小健：《古籍整理通用系统及其中字典的编纂》，《语言文字应用》2000年3期。