

# 《全衡》网上中文输入系统--功能与特性

张小衡

香港理工大学中文及双语学系

ctxzhang@polyu.edu.hk

**摘要:** 在广泛使用普通话的中国内地和台湾,前者单用简体字,后者单用繁体字,两地的汉字输入问题已经得到较好的解决,因此很多人不再热衷于这方面的研究。然而对于香港和国际上的中文电脑用户来说,这个问题还没有得到令人满意的解决。于是我们着手研制《全衡》——第一个较全面考虑香港和国际需求的网上中文输入系统。《全衡》的主要特性包括:采用 Unicode 编码在互联网上工作;提供粤语拼音和汉语拼音等四种输入法;每种输入法都支持直接输入繁体字和简体字;允许横向查询字词信息,同时注意语言规范,支持中文学习。本文详细介绍和讨论《全衡》的功能与特性。

**关键词:** 网上中文输入、粤语拼音输入法、简体字和繁体字、信息检索

## AllBalanced Web-Based Chinese Character Input System: Functions and Features

Zhang Xiaoheng

Department of Chinese & Bilingual Studies, Hong Kong Polytechnic University

ctxzhang@polyu.edu.hk

**ABSTRACT:** A pilot Web-based Chinese character input system, called AllBalanced (全衡), has been developed in pursuit of an optimal balanced solution to the multi-dimensional need of Chinese computer users in Hong Kong and in the international society. At its present stage of development, AllBalanced is supported by a dictionary of over 50,000 word entries encoded in Unicode. Other important features of the system include:

(a) Runs on the Web and eliminates the need to install input software on your hard disc; (b) Supports word-based Chinese character input using Jyutping(粤拼) and Sucheng(速成), in addition to the important input methods of Hanyu Pinyin (汉语拼音) and Changjie(仓颉); (c) Allows each of the four input methods to directly input traditional Chinese characters, simplified Chinese characters, as well as a number of Hong Kong local characters; (d) Enables retrieval of important information on Chinese characters and Chinese words, including their traditional characters, simplified characters, Hanyu Pinyin, Jyutpin, Changjie code and Sucheng code.

**Keywords:** Web-based Chinese character input ; Jyutping, Simplified and traditional Chinese characters, Information retrieval.

## 1 引言

很多学者都认为，拼音输入法应该是一般汉字输入用户的首选方法[1, 2]，实际上这在中国内地和台湾地区已经成为事实。然而，香港的情况却不同，据我们抽样调查[3, p32]，本地大、中学的学生和教师电脑用户中有百分之八十以上仍以形码和手写笔为最常用的中文输入途径。

众所周知，目前手写笔输入汉字的效率是很低的，即使在将来的某一天汉字识别的技术水平能达到 100% 的正确率，手写笔汉字输入还是赶不上传统的纸笔书写速度。而这种速度用拼音输入法却轻而易举就能达到。因此手写笔输入法只能作为一种辅助方法来使用。

一般来讲，与音码相比形码的最重要优点是汉字输入速度高。然而香港的大多数形码输入法用户却没有享受到这个好处。形码用户中有三分之二的人在用“速成输入法”[3, p32]。速成输入法也叫简易输入法和首尾输入法，是仓颉输入法的简化版，每字只取其仓颉码的首尾两个字母。市面上的速成输入法软件仅能支持逐字输入。这样，学习速成输入法虽然比仓颉输入法“速成”了许多，但重码现象却相当严重，输入速度一般达不到 20 字/分[3, p36]，远逊于拼音输入法。

速成输入法为什么会在香港受到如此青睐呢？其实选用速成输入法对于大多数的香港人来说是一件不得已的事。原因是：仓颉难学易忘，普通话和汉语拼音又不够水平。香港的中文输入软件一般都来自台湾和内地，较少考虑到香港的特殊需要，除了仓颉、速成、汉语拼音之外，已经没有什么可以考虑的选择空间了。

实际上，香港对汉语输入工具的功能的要求比内地和台湾要多得多，可以说是国际上的需要加上香港的特殊需要。众所周知，香港人习惯于用繁体字，讲广州话；但自回归祖国后写简体字，讲普通话的需要与日俱增。因此理想的汉字输入工具应该既支持汉语拼音输入法又支持粤语拼音输入法，同时提供本地常用的形码输入法，而且每种输入法都能处理繁体字和简体字。此外，香港是一个国际都市，与世界各地的华人都有语言沟通的需要或可能，香港人在本地和外地使用的电脑可以说是五花八门，他们要求自己常用的输入法软件在这些电脑上能方便地用到，哪怕不是自己的电脑。还有一个重要要求，就是理想的汉字输入工具应该符合语言规范，有利于语言学习。

《全衡》(AllBalanced)是由香港理工大学中文及双语学系网上中文输入研究小组研制的一个面向香港与国际需要的网上中文输入系统，其设计基本原则是以香港的各种需要为重点，同时照顾其它各地的利益，全面考虑，以达最佳平衡。《全衡》在互联网上工作，提供汉语拼音和粤语拼音等四种输入法，支持输入简体字、繁体字和带调汉语拼音，同时

允许横向查询字词信息，重视语文规范，支持中文学习。下面分节介绍讨论这些主要功能与特性及其使用方法。

## 2 一个 Unicode 网站

作为一个中文输入系统，《全衡》的特点首先在于它是一个采用 Unicode 的 WWW 网站。整个网站仅占 5MB 的存贮空间，可以上载到 Internet 或 Intranet 上，一切支持 Unicode 国际标准的机器平台都可以直接访问这个网站连线使用，不受地域的限制，不需要花费时间和资源在客户电脑的硬盘上安装中文输入软件。《全衡》也可以放在 CD，zip 等可移动存贮媒体上随身携带，需要中文写作时直接将盘片插入电脑中就可按离线 WWW 的方式使用，同样不需安装。图一就是《全衡》直接从 CD 启动后显示出来的网站主页。



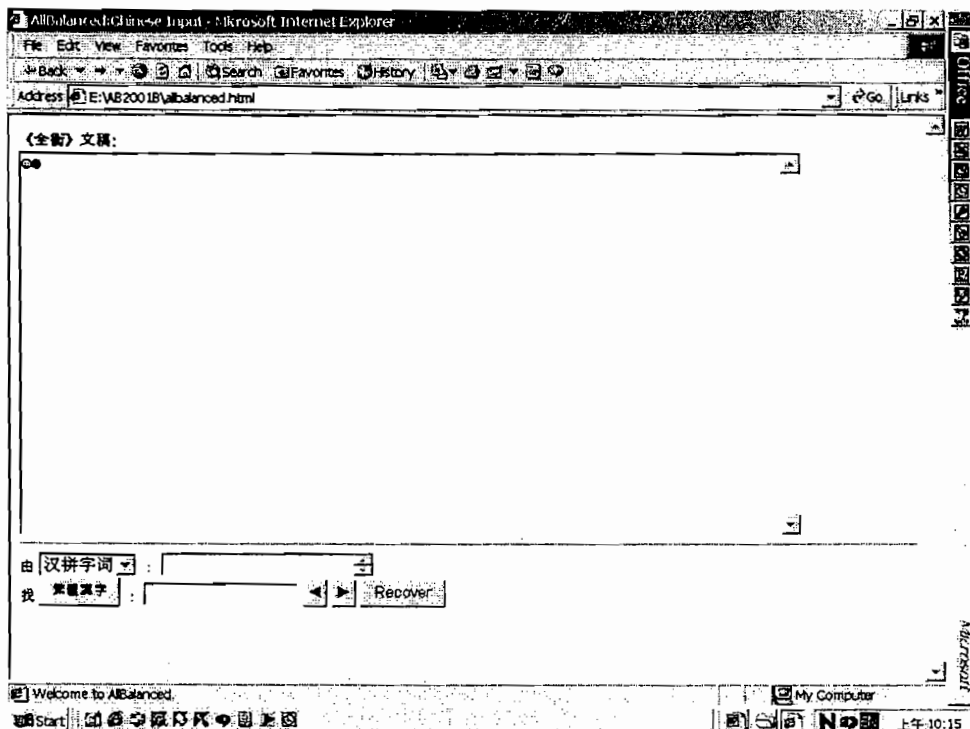
图一《全衡》直接从 CD 启动后显示出来的网站主页。

当然《全衡》也可以装到个人电脑或局域网的硬盘中来使用，安装程序极为简单，只要直接将整个网站文件夹复制到硬盘中就行了。

《全衡》直接采用 Unicode 国际标准代码，这样既可更好地与 WWW 和互联网整体技术环境配合，又可在越来越多的机器平台上使用，包括 Windows 2000, Win Me, Win 98。(中

文版)等。另一个重要好处是 Unicode 有一个相当大的字符集,其汉字部分涵括 GB 和 Big5 的所有繁体字和简体字,还有一些香港字。

《全衡》的主要功能集中在核心系统中,可通过网站主页的第一个菜单项目连接启动,核心系统就绪后屏幕上会出现其工作页面,如图二所示。



图二 核心系统用户界面 (English Win2000 + IE 5.0)

如果是在 Internet 或 Intranet 上使用《全衡》,核心系统用户界面出现后,就可以离线工作,因为《全衡》的程序采用客户 (client-side) 的模式运作。这意味着《全衡》只需依靠网络一两分钟,既安全又省网费。

### 3 提供粤语拼音等四种输入法

为了方便香港用户,尤其是哪些普通话程度较低的用户,《全衡》提供粤语拼音输入法,采用的是香港语言学学会的粤语拼音方案和《粤语拼音字表》[4]。这个拼音方案在设计时就考虑到信息技术的需要,所有的声母和韵母都用英文字母表示,声调用 1-6 六个数字表示,因此用作汉字输入码时均可直接从英文键盘打入,不需要再编码。

现有的其他粤语拼音输入法一般只允许逐字输入,《全衡》的输入法既支持字输入又

支持词<sup>1</sup>输入。当然我们也提供支持以字和词为单位的汉语拼音输入法。这样不论用户使用普通话还是粤语都能享受到“我手写我口”汉字输入的乐趣。两种拼音输入码均采用全拼的方式，汉语拼音的 u 用 v 表示。两种拼音输入法中的单字拼音既可以带声调（汉语拼音用 1、2、3、4 加在末尾表示第几声，轻声用 5 表示），也可以不带调。含有两个字或两个以上的词目前只能通过不带调的拼音输入。

除汉语拼音输入法和粤语拼音输入法之外，我们还提供香港常用的仓颉和速成两种形码输入法。其中速成输入法支持以字和词为单位的汉字输入，仓颉输入法只输入单字。《全衡》的输入法之间可随时转换，相互查询代码，方便中文写作和语言学习，尤其是方便形码输入法的用户学习和使用拼音输入法。

## 4 支持输入简体字、繁体字和带调汉语拼音

《全衡》目前提供的四种输入法每一种都可以直接输入简体字和繁体字，而且简繁体的输入模式可以随时转换，可在同一文稿中使用两种字体。字体输入模式转变时，屏幕上相对于当前输入码的候选字词的字体也随着改变。这既方便输入又有利于学习。

据我们了解，其他输入法软件一般只能支持简体字或繁体字，对于两种字体都需要的人来说会带来不便，尤其是在同一文章中繁简两体并用的情形。靠简-繁转换的手段（如 Wood2000 上设置的功能）来处理已输入的汉字的方法不大可靠，据统计[5]，Unicode 字符集的 3516 对涉及繁简变化的汉字中，一简对多繁（或异）的情形约占三分之一。此外还有少量的一繁对多简的个例。这种一对多的情形使得电脑作转换时常常会出错[5]。因此直接用繁体或简体字输入会可靠得多。

实际上，直接输入简繁体字的功能和简繁体文本互相翻译转换的功能在技术上是不相矛盾的。直接输入的文本也可以再作繁简翻译。所以在输入法上的简繁体字工作模式转换功能所带来的灵活性是附加的。

据了解，除《全衡》以外只有 Win2000 上的 MSPY (微软拼音) 输入法具有直接输入简繁体字的功能。但 MSPY 不能随时改变一个候选字词的繁简字体。出现简体字后，如果要其繁体字只能重新打入输入码，反之亦然。

《全衡》的用户还可通过输入数字标调的汉语拼音来得到单字的标准标调拼音。例如：输入“xiang1 gang3 ming2 tian1 hui4 geng4 hao3 !”可得到“xiāng gǎng míng tiān huì gèng hǎo !”。

---

<sup>1</sup> 如不特别说明，本文所说的“词”包括词典中的单词和短语。

## 5 文稿输入与编辑

以下分几个方面介绍如何利用《全衡》输入和编辑文稿。

### 5.1 汉字输入

在《全衡》上，输入法与字体的选择与配合相当灵活，中文输入相当简便。核心系统启动后，初始的输入法和汉字字体分别定为“汉语拼音”和“繁体字”（如图二所示）。如需要更改，可用鼠标器点击（Click）“由”字后面的输入法菜单和“找”字后面的“繁体字-简体字-标调拼音”转换按钮。

然后将输入码打在输入法名称右边的小文字框(text box)里面（输入码会同时处现在上方文稿区中的◎●之间），并按空格键。与该输入码对应的第一个汉语字词就会出现在字体按钮右边的文字框里（同时显示在◎●中间）。如果该字或者词正好是所需要的，则按键“v”，将其输入到文稿的末尾。如果不是所需要的，则再按空格键（或使用“▶”按钮）查看下一个重码字词。如要查看前面一个重码字词，可使用“◀”按钮。这种动作可以重复多次或混合使用，直到目标字词出现或者遍历词典里与给定输入码对应的所有字词。出现候选字词后，输入码并没有消失，可以对任何部分直接修改。修改后再在末尾加空格，就可以按更正后的代码重新查找汉语字词。这种功能 Win2000 上的输入法是不具备的。

如果要将新输入的文字插入到文稿的其它地方，应先在該处标上插入号“◎●”（可使用 cut-and-paste 或者 drag 的方法移动“◎●”。如果文稿中有多个插入号则以最后一个为准。）。使用符号“◎●”的原因有：（1）不与一般书写字符混淆；（2）营造一种轻松快乐的气氛；（3）对于中国人来说黑人白人是典型的外国人。用“◎●”代表中文和英文的分界线直观合理，此范围以外的可作外文直接输入，以内的是中文输入。

《全衡》目前只能逐个输入字、词或短语，不能像 MSPY 那样通过拼音汉字智能转换允许连续编码输入。其实拼音汉字智能转换的技术似乎还不太成熟。例如在 MSPY 上为了输入“中文及双语学系”打入“zhongwenjishuangyuxuexi”时，得到的是“中文几双雨靴系”，其他的例子还有：不止气度引得子（不知其读音的字）、补选装（不需安装）、健全（极安全）。

### 5.2 偏旁部首和标点符号的输入

《全衡》的每一种输入法都支持汉字偏旁部首和标点符号的输入。在单词“偏旁”的“同码词”中可以查到不成字的偏旁部首。例如在汉拼输入法中打入“pianpang”或在粤

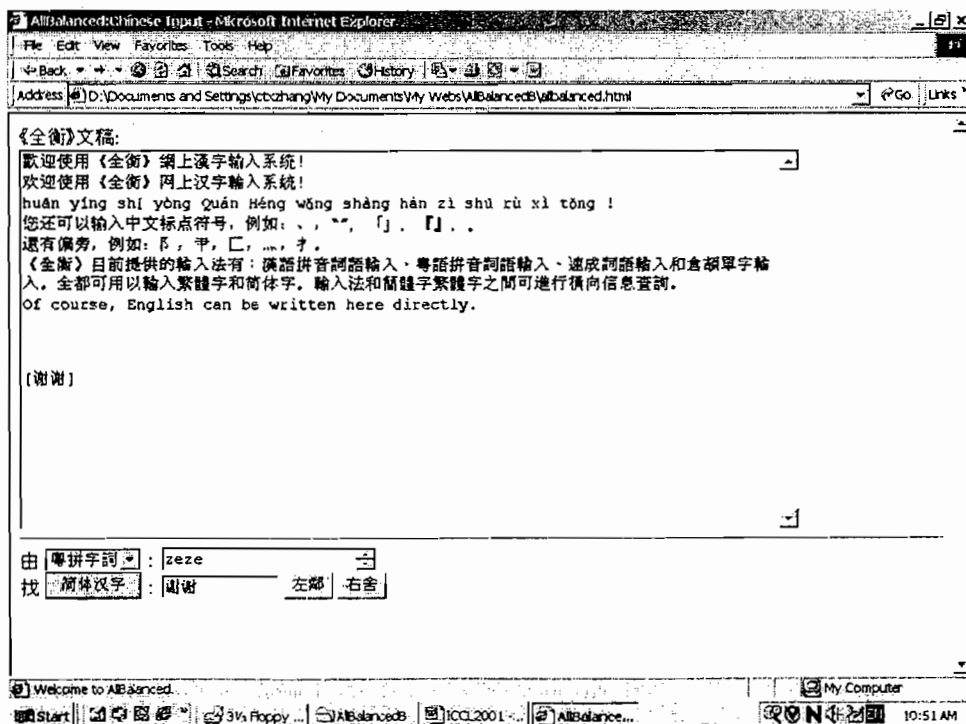
拼输入法中打入“pinpong”，除“偏旁”这个词之外，还可以得到“𠃉”、“𠃊”、“𠃋”、“𠃌”和“𠃍”等。

至于标点符号，与英文同形的可以直接在文稿中输入，汉语特有的标点符号以类似的英文字符为“输入码”，方法是：“. →。”、“, →,”、“< →《”、“> →》”、“: → :”。其中“A→B”表示从输入码A可以得到中文符号B。单词“顿号”和“引号”的“同码词”则包括这两种中文标点符号的各种繁简体表示。

### 5.3 英文输入

英文词语可以直接打入文稿框内。用户也可通过《全衡》外面的中文输入法直接在文稿中打汉字。在☉范围之外输入的中外文字《全衡》都不做改变。因此《全衡》与电脑上原有的各种输入法完全兼容。

图三是《全衡》中文输入的一个实例，没有用到其它中文输入法软件。当前的中文输入模式已改为“通过粤语拼音输入简体汉字”。



图三：《全衡》的各种字符输入

### 5.4 文字编辑和文件处理

《全衡》的主要功能在于汉字输入，同时支持一些简单的文本编辑操作。用户可以用浏览器的文字编辑功能，如 Copy, Cut, Paste, Drag and drop 等等直接在文稿上作修改。Right-click(用鼠标器右键单击)文稿框内还可以看到 Undo 等功能。点击“Recover”(恢复)按钮则可将文稿内容恢复到上次用 v 命令后的状态。因此，Undo 和 Recover 具有安全保护作用。

文稿中的内容可以通过 Copy-and-Paste 转移到 MS Word 等其它文书处理系统或者文本编辑软件上做进一步的编辑处理和格式排版，以及文件存盘。

## 6 信息查询

一般的输入法依靠的是一个简单的“输入码-汉字”对照表，语言知识含量较低。《全衡》使用的是一个拥有五万多词条的词典，每一词条讲述一个词，信息包括该词的繁体字、简体字、汉语拼音、粤语拼音、仓颉码、速成码等项。用户可以从任何一项入手，查找其它各项信息。

通过《全衡》某种输入法找到一个字词后，如果要了解该字词的有关信息，可通过改变输入法来实现，输入法选项中包括词典词条的各类信息的名称。简体字和繁体字也可以作为特殊的“输入法”来用，从而可以从词典中了解到外来字词的有关信息。具体的做法是：先将输入法转为繁体字或简体字，然后通过某外部输入法或 Copy-and-Paste 将一个繁体或简体字词放入输入码小框里，接着按空格键，如果查找成功，下面的汉字输出框里就会有相应的汉字出现。这时如果改变输入法，就能看到该汉字或单词的有关信息。这种灵活的词典信息查询功能对于语言学习与运用是很有益的。

## 7 结束语

《全衡》是第一个面向香港和国际实际需要的功能较为齐全的网络中文输入系统。它提供粤语拼音和汉语拼音等四种输入法，每种输入法都支持直接输入繁体字和简体字。通过《权衡》的输入法查找到的每一个字词都可以横向查询到其繁体字、简体字、汉语拼音、粤语拼音、仓颉码、速成码等有用信息，支持中文学习。经初步在香港理工大学中文及双语学系的学生中内部试用，效果是令人鼓舞的。

然而现在的《全衡》还只是一个发展中的软件，很不成熟，许多方面的功能都有待补



充或改善。首先需要进一步完善的是作为《全衡》的核心知识库的词典。该词典是以《粤语拼音字表》，香港理工大学中文三地语料库词频表和 Win98（简体中文版）的全拼输入法词表为基础建立起来的，信息量相当丰富，但编辑工作也很艰巨，还有不少地方需要修改更正。此外还要添加数量可观的香港常用字和常用词。

另一个问题是，现在用拼音输入多字词时不能用声调。为了提高灵活性和工作效率，多字词输入应该同单字输入的情况一样，允许拼音输入码既可以带声调，又可以不带声调或带部分声调。

我们还打算增加郑码输入法，因为郑码同时照顾简体字和繁体字，而且郑码输入法是 Windows2000 中来自中国内地的唯一形码输入法。

尽管从理论上讲，作为网站的《全衡》应该在任何支持 Unicode 的 WWW 客户机器上都可以使用，但实际上还存在一些技术性问题。《全衡》是在“Windows2000 + IE(Internet Explorer)”的平台上研制的，在“WindowsMe +IE”和“Windows98（台湾中文版） + IE”的环境中也能正常工作。与其它工作平台的兼容问题也是有待解决的。

此外在语言规范性，用户界面的亲和性和系统的智能性等方面都应该进一步加强。我们的目标是将《全衡》发展成面向香港和国际社会的，既可用于中文输入又有利于中文学习的，语言素质较高智能性较好的方便实用网上中文系统。

## 参考文献

- [1] 张普. 汉语拼音输入法应该是我国的首选中文输入法. 见: 张普. 汉字编码键盘输入文集. 北京: 中国标准出版社, 1997.
- [2] 张小衡. 香港也应该以拼音输入法为首选汉字输入法. 见: 黄昌宁和董振东. 计算语言学文集. 北京: 清华大学出版社, 1999, pp424-431.
- [3] 邝施雯. 香港中、大学生与教师学习及使用中文输入法情况研究. 香港理工大学中文及双语学系学士学位论文, 1999. (导师: 张小衡)
- [4] 香港语言学学会粤语拼音字表编写小组. 《粤语拼音字表》. 香港: 香港语言学学会, 1997.
- [5] 郭小武. 电子文本的简繁转换. 语言文字应用. 2000 年第 4 期.

鸣谢: 本课题先后得到香港理工大学的两次资助(账号为 GS968 和 1-9827), 香港语言学学会提供《粤语拼音字表》, 北京中易电子公司提供汉语拼音码表, 香港理工大学中文教学中心提供三地语料库词频表, 本校副校长梁天培教授亲自担任我们的顾问, 苏詠昌博士在课题管理上给予大力支持, 特此鸣谢. 参加本课题研究的人员还有张群显博士、邓思颖博士、谭世宝博士和沈枫琪小姐。