友有聲

中文與中文互聯網全球化之方略

魏凌雲（重慶交大1946級）

廿十世紀的最後十年，互聯網(Internet)的興趣與發展，逐漸形成人類智識寶庫，無所不包，無遠弗屆，不僅使個人隨心所欲，予取予求，而且使商業行為龍騰虎躍，互聯網對於人類文明的影響，實難估計。

近年來網路使用人口不斷成長，根據統計,1997年9月全美互聯網用户有四千萬,至1998年九月增加至五千四百萬，成長率為37%，同期間內，網頁從二億七千萬增長至七億兩千萬，成長率為166%，互聯網蔓延速度相當驚人。

在中國大陸，1998年底，網民只有210萬，到1999年底，已達890萬，增長率為323%，估計到2000年底將突破兩千萬，三年之間增加十倍，在台灣情形相似，過去三年。網民由40萬到現在超過400萬，網路使用人口比例近18%。

互聯網的迅速發展與擴充，使我們看到中西網路的貫通，將促成人類兩大文化的融合，最使我們欣慰的是，藉此機會，中文與中華文化可以傳播全球，使21世紀成為中國人的世紀。

但是要使西方人士，出入中文網路，悠哉游哉，我們得為他們考慮兩個問題：(1) 假令他們懂中文，但國語拼音，羅馬拼音，拆字根（如倉頡碼、王碼)，可能不高明，甚至根本不會寫中國字，不知從何筆下手。因此現在所有中文輸入法，對他們都不能用，(2) 如他們不懂中文，想上「中文網」，有何速成方法？

事實上，不僅西方人士，中國大陸三分之二的人口，會有同樣情形。我們說的「全球化」，當然包括「全國化」在內，因此將來「中文互聯網全球化」希望上網的人口，可能超過十億以上，他們中大半會遭遇上述的兩個問題之一，如果這兩個問題不能有效地解決，則所謂「全球化」，不知進到何年何月，這與這幾年的迅速成長率無關，因為那只限於智識圈而已。

本文的主旨在提供解決上述兩大問題的方案。

中文通碼可使人人上網

為使人人（包括文盲及洋人）上網，我們得設計一種「中文通碼」(Chinese Common Code,CCC)，其目的在使其「通行全球」，因此它必須符合下列嚴格要求：

(1) 人人稱便，處處可用
(2) 見字識碼
(3) 不需猜度，不需記憶，不需訓練
(4) 使用者不需懂中文基本知識：拼音部首，字根，筆順

依照上述要求，我們只有從中文筆劃著眼，根據分析，中文筆劃，有四種基本型態：口、橫、直、雜，它們的確實定義如下：

友聲有聲

口：有四邊四角，每邊每角，各自獨立，不與其他筆劃共用或合併。例如：田，只算一口，不能算四口，因後者中每一口與鄰口共邊共角。

橫：水平或近於水平筆劃，全筆粗細一致，一端或兩端可帶小點或小勾。

直：垂直或近於垂直筆劃，全筆粗細一致，下端可帶小勾，向左或向右。

雜：所有非橫非直筆劃，如點、挑、撇、捺、彎。

任何及所有中文字都由這四種基本型態的筆劃組成無一例外，這些簡單圖形：口、橫、直、雜，任何人一眼便看得出來，不需猜度，不需記憶，更不需訓練。

中文通碼（簡稱魏碼）便是利用口、橫、直、雜數目的適當組成，先設：

A＝口數　B＝橫數
C＝直數　D＝雜數

（一）字碼規則ＡＢＣＤ四位碼字中的（口、橫、直、雜）數

例：中1010　國2104
杏1112　丗0122
叶1110

很明顯地，魏碼尤其適用於簡體字，對於大陸民眾，海外華人及外國人，都很方便。

（二）詞碼規則

(1) 雙音詞：$(ABC)_1$，$(BC)_2$，第一字的（口、橫、直）數加第二字的（橫、直）數。

例：中國10110
先生03231

（2）三音詞：$(ABC)_1$，$(BC)_2B_3$
例：三不知030112
急口令041001

（3）多音詞：$(ABC)_1(BC)_2B_3N$
最後N為詞中的字數。
例：莫名其妙1321044
一手難掩天下人耳目0102189

詞碼不論詞的長短，最多七位，不計雜數，較字碼更為容易，中國文字，詞多於字，上述設計經濟實用，便人不少。

為減少魏碼使用困難，加訂下列規則：

規則一：凡Ａ、Ｂ、Ｃ、Ｄ中有大於9者，均做9計算
例：龍0953

規則二：凡遇到下列牽連字根，統做雜劃，其Ｄ值預先規定如下：
字根：辶　厶　幺　糸　彡　糸
D值：3　3　5　5　5　8
例：　繼0219

由本規則，上字中的雜劃，一看便大於9，故不必細算直接當作9

規矩三：若橫或直的附勾不短於主劃的二分之一者，或與下一筆劃相連者，應作雜劃計算
例：了0111　衣0115

規則四：如有同碼字，則在魏碼之後加一副碼Ｓ或ＳＰ。Ｓ為音碼，代表單字或詞中最後一字羅馬拼音的第一字母，Ｐ為位碼，自1開始，最常用者佔第一位
例：人0002R　八0002B　入0002Y

拼0324P 1 披0324P 2坡0324P 3
一代01011D1　一丁01011D2

副碼(S或SP)，只存於字典，電腦或資料庫，常人不必記。

現在我們來談中文通碼如何使人人上網。假定有人想在互聯網上查看香港股票行情，譬如股票名稱為：中信、長江、大光明，照上述詞碼規則，他可容易打入：
10131（中信）04221（長江）010224（大光明）

如果發現同碼詞，他只需用滑鼠移動箭頭向所要的名字按一下就行，各地股票名稱不會超過一千個，因此同碼股票名很難得見，不會引起困擾。

再如一位洋學生，學中國歷史，在圖書館查不到「八王之亂」的內容，他可利用互聯網接到北京圖書館，在「歷史」類，打入：

000	31	1	4
八	王	之	亂

立刻便可得到所需資料，這是何等方便。

中文通碼可使人人上網，在於上網者多半是查詢資料（如上兩例），而查詢資料僅要輸入一個中文詞足夠，依前述詞碼規則，完全不計雜劃，而詞碼最多七位，簡單明瞭，任何人都會用，較之窮拼音、拆字根，難易之間，不可以道理計。

「中文通卡」能使文盲識字

「中文通卡」的設計是：卡內儲存常用中文三千字，繁體、簡體俱備，查一字，將「魏碼」順次按入，螢屏上即顯出所有同碼字，為縮小體積，卡上不加26個英文字母，副碼全用數字，依使用頻率排列（1為最高），每字之下顯出副碼，看要查的字下面副碼，按入後，首先聽到該字的標準國語發音，同時螢屏上顯出該字的中英文意義，為便於文盲領會中文意義可用語音表達，而且一筆一筆地依序寫出該字的全部筆劃，於是字的音、形、義以及寫法，可立即聽到和看到。

「中文通卡」有如手掌大小，可完全放在身上口袋內，隨時隨地可用，假定一個文盲每天學十個中文字，則一年便可認識三千字，文盲可從此文「明」。

利用「中文通卡」，可徹底解除中國五千年來的文盲問題，他不僅可使現有大陸約兩億文盲可在三年內掃清，而且使文盲數永遠保持到近於零（除非此卡供應不足）。

更大的意義「中文通卡」可使全世界人易於及樂於學習中文，它與互聯網聯手，一方面促長「中文互聯網」全球化，一方面藉後者之助，使中文全球化。

遠的不說，在中國進入世貿組織以後，國內市場大開，外人將如潮水湧入，爭往各地，尋求最佳機會，大發利市，但是他們不懂中文，一出門就會遭遇困難，如果有「中文通卡」在手，則許多問題可迎刃而解（譬如說，不會將吃飯唸成吃糞）。

海外中文教育，一直是滯礙難行，我曾義務擔任中文學校校長五年，深知其中辛酸。首先，許多中國小孩不願學中文，認為中文無用，而且難唸難記難寫；；第二，中文學校每週只上課兩三小時，時間不夠教學需要；第三，小孩回家後無人督促輔導，聽其荒廢，所學大半忘記，幾年

下來，小孩能說幾句簡單中文，已經是很不錯了；寫字作文根本談不上。如果有「中文通卡」，則上述情形可大為改善，只要老師管教嚴格，要學生照卡練習發音、寫字、作文，假以相當時日，成績必然可觀。

由於美國大學入學試驗SATT，將中文列入可考科目之一，高中增加中文選課，因此美國高中生選讀中文人數大增。而大學中文課更為加強。 一九九六年八月北京舉行國際漢語會議，有廿八個國家代表參加，據統計全球有六十個國家，一千多學校有中文課，隨著中國加入世貿組織後，西洋學中文熱將更會熱烈，然而洋人學中文的困難，尤甚於華人子弟，因為他們求助無門，唯一的希望只有「中文通」卡。

前面說過，「中文通卡」當與「中文互聯網」聯手，互輔互助，可迅速使中文及中文互聯網全球化。現在中文互聯網在大陸及台灣均飛快成長，估計到2003年，中國網民可超過一億，那時中國農民及西方人士將被吸引，大量投入中文網，因之「中文通」卡的需要將激劇提升，依我們審慎的估計，「中文通卡」的年產量要在五千萬片以上，才能配合中文網成長的速度。開發「中文通」卡，成本甚低，而其市場包括中國大陸及全世界，十年之內少需五億片，其效用可根絕中國文盲及使中文全球化，一舉數得，雄心企業家何樂不為？

通碼字典利惠中西中文字典

目前流行的檢索法有兩種：第一是康熙部首（214），第二是羅馬拼音。康熙部首引起的困難，林語堂曾鑑下列的例子：

知字矢部　而和字口部
蜜字虫部　而密字宀部
粵字米部　而奧字大部
鳳字鳥部　而凰字几部
卒字十部　而率字玄部

林認為這簡直是與人開玩笑，他感嘆地說「康熙字典，豈有此理」。1935年教育部國語統一批評康熙部首，用了十六字的考語：
「徘徊古金，迷亂本末，檢尋不便，控制無方」

再說羅馬拼音對於中文齒音可用到C、Ch，G、J、Q、S、Sh、Z、Zh，普通人那能分得清楚？查字典，有時是為了查生字，生字根本不知其音，如何查法？

就是為了上述緣故，現代中國人很少翻中文字典，看書或看報，稍有不懂的文字，便馬虎混過去，不求甚解，今日一般民眾，甚至大學生中文程度日漸低落，少翻字典，是最大的原因。

至於洋人到了中國想學中文，除非進「外語學校」，可說望字興嘆，最多學會幾句簡單中國話而已，想看中文，沒有一本字典可查，只好徒呼負負，所以辦起事來，非用懂洋文的中國人不可。

總結來說，現代中國人文化低落及中國文化難以西流，與字典不易使用，大有關係，因此我們必須在康熙部首與羅馬拼音之外另覓途徑，以解決這個困難。

中文通碼是跳出「部首」與「拼音」之外的一種方法，它用的「口、橫、直、雜」四種基本型態，是人人用眼分辨得清清楚楚的，它用的四個數字，也是人人屈

指可數的，對於簡體，尤其一目了然，通碼不需猜度，不需記憶，更不需訓練，在字典上，用數字排列秩序，遠比用「部首」和「拼音」，要來得清楚自然，方便得多。

再者，在中文字典，詞遠比字多，我在中文詞源（台灣出版）上查出：

單字：14000（取整數）
雙音詞：65000
三音詞：9700
多音詞：7700

同碼字及同碼詞，在加副碼（S或SP）後可完全分開，字與詞間，以及不同字數之詞間，由於通碼位數不同，決無同碼。

在「部首字典」或「拼音字典」中，詞都排在單字之後。在上述的中文詞源，以「三」字居首的雙音詞，就有379個字之多，如果運氣不好，查其中的其一「雙音詞」，可能要花十分鐘，這是多麼不方便，通碼的詞碼，不計「雜」數，可以見詞知碼，因此在字典中，一查即得，全不費力。

當大眾發現利用一種新法查字典，輕鬆愉快，而且可以提高智識程度，增加閱讀興趣，那麼人們（中國人及洋人）必樂於勤翻字典，這種習慣的養成便無形提升全國文化水準及加速中國文化西流，有利於中文全球化。

結論

自從一百多年前，李鴻章提倡仿效西洋「船堅砲利」以來，中國本土內的科技發展，一直依循「仿效」的老路，迄今為止，本國科技產物絕少有堪稱「中華首創，通行全球」，而為世人所樂用者。倒是在海外，盛行針灸、人參、銀杏、大蒜、綠茶和太極拳，難道我們這一代中國的智慧硬是不及我們的祖先？我的答案是，我們的「觀念」與「抱負」有問題。在觀念上我們以「西洋至上」，在抱負上，我們以「賺錢第一」，這兩者十足窒息我們的智慧，難以發揮。

如果我們期望「廿一世紀是中國人的世紀」，那們我們必須改變上述的觀念和抱負，新的觀念應當是「中華至上」，而新的抱負應該是「助人第一」。

本文所提出的「中文通碼」和「中文通卡」正是出於新的觀念與抱負，如果實現，則中文與中文互聯網全球化，指日可待。

參考資料

1.L.Y.Wei, A Chinese indexing system designed for world-wide use.Proa of the 1998 International Conf on Computer Processing of Chinese and Oriental Language Torunto,Aug 29-Sept 1,1998,pp.316-320.

2.L.Y.Wei, A semantio coding system of Chinese. Language based on universal use principle, Ibid, Taipei, Taiwan,Aug 13-16,1991,pp20-23.

3.魏凌雲,中文編碼學,全華科技圖書公司,台北,1991

4.L.y.Wei,Univeral indexing system of Chinese words.Copyright Certificate of Registration,420270,Canada, Dec.18,1992

5.L.y.Wei,A solution to four problems in Chinese culture.北美華人學術討論會,芝加哥,7月25日,1993

6.L.y.Wei,A simple solution to removal of illiteracy in China.全美中國研究協會第四次年會,紐約,10月30-11月1日,1998