

## 1 汉字在互联网时代中的尴尬处境

- 1978年12月 全国汉字编码学术交流会在青岛召开这次青岛会议 聚集了来自中科院和各地高校的众多学生 他们讨论的问题在如今看来难以想象 如何把汉字输入电脑
- 在如今各种输入法横行天下 年轻人打字速度快得飞起的年代 很难想象在七八十年代的中国 中文输入成为一项卡脖子的技术 原因也很简单 汉字实在是太多了 1980年发布的国标GB2312 自己包括了6763个汉字而2005年发布的国标 GB18030-2005 收录了高达7万244个汉字 而且和西文字母不同汉字并没有一个类似ABCD之类的排序方法 如何把而无序的汉字和键盘上的按键一一对应 在当时并没有一个现成的解法

### e.g.1

汉字在信息处理尤其是信息输入上远远落后于西方国家 1984年洛杉矶奥运会 当法新社记者看到中国记者还在手写新闻发稿 直接阴阳怪气了一番7000名记者 只有中国人还在用手写

### e.g.2

当时的学界不乏汉字落后论的声音 有些极端学者甚至认为在计算机时代 拼音文字将取代方块字成为中文的出路 汉字拼音化 拼音代替汉字甚至一语双文的论调层出不穷 有些地方的学校甚至出现了 语文全拼音化的教学

## 2 如何摆脱此困境？

在计算机技术起飞的前夜 汉字输入问题必须得到解决 会上众多专家教授都提出了不同的意见 主流的看法是 开发一款专门为汉字输入而设计的键盘当时的汉字专用键盘有三种方案

### • 整字大键盘

也就是把几千个汉字 按照部首分区放在一块像桌子一样大的键盘上哪里要用电脑里 如果一块键盘不够 那还可以放第二款 第三款这个方案优点很突出 没有学习成本很直接 但缺点也很突出 除了简单直接以外都是缺点 不能盲打 速度极慢 每分钟能输入十个汉字就是极限 而且一块键盘造价几千块 设备大 成本高不实用

### • 主副键盘

是仿照日本人设计主副键盘，这个方案由一块主键盘和一块副键盘构成 副键盘负责定分区主键盘 按对应的按键就能输入分区中特定的汉字 可以看出这个方案是对上面那一个的一个改进 减小了设备体积可以双手打字 速度也快了很多 但造成了一个巨大的bug 打字员需要记住每个汉字处在哪个分区 排序在哪个位置 哪怕只有几千个汉字 都是巨大的学习成本 这个方案基本上没有普及的可能性

### • 256键中键盘

汉字激光排版之父王选院士提出的256键中键盘方案 这个方案把所有汉字 拆成了1000多个笔画和部首把它们放在256个案件上 只要按照结构组合 这些零部件就能像拼乐高一样拼出想要的汉字 中键盘第一次用到了拆字的方法也就是汉字编码类型中的形码 后来的五笔真码 以及此前朱邦复在1976年所创制的仓颉 输入法使用的都是形码 相比前两套系统中 键盘方案显然是先进很多 但还远远称不上完美毕竟英文的QWERTY键盘 只有26个字母键 键位少 够轻便而且已经在国际上通用了 如果可以 谁不想直接用现成的东西 奈何汉字确实太特殊了

## 3 五笔输入法的出现

### • 王永明发明汉字专用键盘

就在青岛会议召开的同一年 河南南阳科委的王永明 也加入了汉字输入法开发的工作 和当时的主流想法类似他一开始也希望制造一款专属于汉字的键盘 为了实现这一目标 王永明把现代汉语词典中1万2000多个字 逐一分解进行了分类统计归纳出了600多个组字单位 王永明将他们命名为字根 并且根据出现频度 选择了其中最常用的125种在他的努力下 他把中文专用键盘从256个 压到了188个 又

压到了62个到1982年 王永明开发出了36键方案 而且这一切都是在没有计算机 纯人工统计的情况下开发出来的 这时摆在他面前的有两条路：

- 一是推广36件方案 这套方案在1982年已经上机成功 作为当时最先进的方案其实已经有普及的可能性
- 另一条路则是回头继续研究压缩键位 压到完全适配26键的QWERTY键盘 要知道键位压缩的越少重码也就是一组按键对应的不同文字就会越多 而选择会极大影响效率 到了这一步 再往下压缩已经很难了 为了降低重码 经过将近一年的研究 王永民发明了莫比自行识别法 将每个字最后一个笔画编码 定位将重码率降低了一个数量级

1983年2月28日 王永民发明的五笔字型宣告诞生

### 五笔的厉害之处：

- 首先就在于快 因为引入了横竖撇捺折五种笔画的分区 使得用户最多只需要按此按键 就能锁定唯一的汉字没有除码 意味着用户不再需要挑选想要的字词 经过甚至可以实现双盲输入 也就是不看键盘也不看屏幕打字
- 而且相比大键盘和中键盘 五笔的键位少 学习成本更低许多八零后可能还会被五笔口诀 什么王旁青头兼五一土 是二千十寸女 而熟练的无比用户甚至不需要这些依靠肌肉记忆就能输入
- 其次就是五笔完美兼容QWERTY键盘 不需要专门再开发新的硬件 在五笔发布之初因为不需要专用键盘 甚至被学界认为是崇洋媚外 食古不化 但市场和用户是用脚投票的

1984年 国家科委和国防科工委先后发布红头文件 向全国全军推广五笔字型

e.g. 3

1986年也在同年以数10万美元的价格 将使用权出售给美国 DEC公司 成为中国第一个出口美国的电脑专利技术 此后IBM微软 苹果陆续购买专利使用权 而东南亚国家如马来西亚和新加坡 更是至今还在广泛使用五笔

在国内 五笔也成为一代中国人最早接触和学习的汉字 输入法 在电脑还不普及的90年代使用电脑被认为是一种职业技能 国内有大量的培训班 而五笔就是这些培训班最重要的授课内容 当年的许多国产电脑键盘上不仅有英文字母也标注了五笔字根 方便学习

## 4 拼音输入法的崛起

### 4.1 拼音输入法的诞生以及初期的问题

就是这样一个当代活字印刷术 为何却沦为拼音输入法的手下败将成为了小众的代名词和时代的眼泪呢

#### • ABC输入法诞生

1995年 智能ABC输入法诞生 这是一款经典的拼音输入法 和现在的拼音输入法比起来 当年的智能ABC和智能两个字没有半毛钱关系 它的存在就是提供了一个庞大的字词库 让用户查找拼音对应的文字和词语而他的词库其实也缺乏实用性 因为那个年代的词库主要是从官媒和各类报刊 杂志的报告中统计高频次得来的 因此这样的词库其实非常不符合普通人的语音需要 也没有联网更新词库的方式 现在拼音输入法必备的联想输入 模糊音 常用词汇记忆等功能在当时更是天方夜谭 在效率上和五笔是远远不能相提并论

#### • 为何输入效率低

拼音输入 为什么效率低一个点就是 拼音的重码率非常高 不考虑音调 只有400多种组合而这些组合要对应的可是成千上万个汉字 这意味着拼音不止要打字 还要选字 往往是输入容易找字难 尤其是某些读音对应着大量的读音字 比如之之师 如果你的名字里有个一字恰巧还是生僻字 而你只会用拼音输入法 那么恭喜你 国标GB18030自己里光是读一音的字就有407个 打字一时爽 找字火葬场

#### • 为什么不如五笔输入法

而王永明提出的形码设计三原理第一原理就是香浓型原理 也就是重码率要低 确保编码的唯一性 最好达到2%以内 由于拼音里面大量多音字 同音字效率显然是不及格的 为了改善效率 后来又有了双拼输入 不过只要是基于拼音重码就是解决不了的问题

此外五笔字形在按键盘布上是经过设计 从这张图就可以看出来 同样是基于QWERTY键盘五笔打字在两只手的负荷率上基本是平衡的 按拼音新输入就是右手忙死 左手放假 而且如果你用标准的打字法 会发现左手最强壮的手指 好像没有什么做 但是力量最小的小指反而负荷很大 简单来说拼音打字在人体工程学上也不是很科学 所以说从效率和科学性上来讲 五笔是远超拼音输入法的存在

## 4.2 转机

真正改变了拼音输入法尴尬局面的 一是靠内存二是靠互联网：

- 内存

作为一种极其依赖词库的输入方式 拼音输入法需要大量的存储空间来存储词库 在80年代电脑内存按KI计算存储空间 按照B来计算的年代 根本存不下足够大的词库 直到90年代主流电脑的硬盘容量已经达到了GB级 内存也达到了数十个兆B的水平 这才有了越来越庞大的拼音字库 字词库更大了能够匹配的词汇也就多了 效率相对就提高了

- 互联网

而互联网带来的则是更大的变化 2006年搜狗输入法横空出世 它的基本原理上和智能ABC 紫光拼音 并没有太大的区别 唯一的不同是他通过搜狗的搜索引擎极大地扩展了词库 用户输入的拼音不再是与本地词库对应 而是与整个中文互联网的内容匹配 除此之外搜狗还包括了整句输入 联想输入等功能 要知道整句打字的效率要比一个字 一个字打可高太多了而且很多时候你不用打完整的拼音系统 就会帮你补完 此外它还会记录你常用的高频词 随着你的使用它会越来越顺手 这些功能的加入 让拼音输入法的效率直接起飞

搜狗输入法一年内抢下了90%的市场份额 即使后来腾讯 百度 谷歌也纷纷加入战局搜狗的市场份额也没调到70%以下 但客观一点来说 这些输入法在本质上并没有让拼音成为更精确 重码率更低的输入方式它只是用大数据 大词库和联网功能强行提高了匹配效率

e.g. 4

这就像我听过的一个关于程序员的笑话 产品性能不够怎么办一个程序员会有两种选择 一是猛干六个月 写一套很复杂的程序 提高系统另一个是花半年休假在家睡大觉 等你一觉醒来 更牛逼的硬件已经降价了 直接换硬件就能解决问题

## 4.3 优越之处

- 门槛更低，符合人体特征

而拼音真正相比无比的优越之处 在于它是一种门槛更低的输入方式 汉字是字形自生自义 三位一体的文字这三者中我们往往是先理解了自身和记忆 也就是先学听说再学读写 而且我们的思维也往往是以语音方式进行的 当我们想事情的时候脑袋里本来也会有一个声音

- 契合社会背景与教育体系

相比五笔这样的形码 拼音作为真码 门槛更低也更符合我们的本能 更可怕的是 拼音的背后是一整个9年制义务教育体系 拼音输入法并不是没有学习门槛只是很少有人需要专门去学习拼音 从小学一年级第一堂语文课开始 我们就在学习啊 我已无语了一直到高考 拼音就没有离开过语音学习 以中国目前90%以上的9年制 义务教育巩固率来看拼音对这一代中国年轻人来说 几乎就是出厂设置

有意思的是 五笔的用户群中反而有不少六零后 他们没有接受过完整的拼音教育 本能读会写 这批中老年人用起无比反而更加顺畅 只是在中国 越来越完备的教育体系 和庞大的拼音使用人口面前五笔毫无胜算 为什么 互联网公司愿意花大力气开发拼音输入法 也是因为用户基数大商业上具备想象空间 巨头们才愿意在输入法这样的边缘市场 投入人力物力 到了移动互联网时代拼音已经真正意义上坐稳了江山