

目录

- 一、 字母使用热力图..... 2
- 二、 衡量按键使用均衡性和输入效率..... 2
 - 1.均衡性..... 2
 - 2.输入效率 6
- 三、 新的编码方案 7
 - 1.编码方案描述 7
 - 2.新旧编码方案平衡性对比..... 9
 - 3.新旧编码方案输入效率对比 11
- 附录一： 拼音种类数及其频率 12
- 附录二： 新旧编码表 30

一、字母使用热力图

本篇文章中, 26 个字母使用的热力图如图 1-1 所示, 由于要考虑打字效率, 故以普通键盘的排列方式展现。

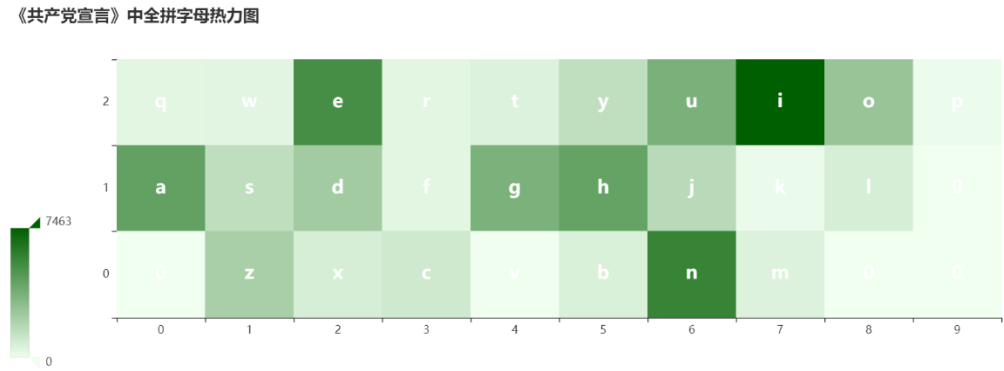


图 1-1 字母键使用热力图

二、衡量按键使用均衡性和输入效率

1. 均衡性

均衡性主要从单字母出现次数、输入文章过程中各个手指移动距离和敲击次数两个维度来衡量, 各字母出现次数越趋近, 手指移动距离越平均, 均衡性越好。

1) 单字母出现次数的均衡性

统计本篇文章中以全拼方式输入时, 各个字母出现的次数, 如图 2-1 所示, 计算每一字母出现的频率, 如表 2-1 所示。从图中可以看出, 在使用拼音输入时, 各字母使用次数差异较大, 使用次数最多的是 “l”, 为 7463 次, 使用次数最少的是 “v”, 为 27 次。经计算, 样本的平均值为 2057.0, 标准差为 1984.492, 极值为 7436, 字母使用次数标准差和极值较大, 故在 “字母出现次数的均衡性” 的标准下, 此种输入方式均衡性较差。

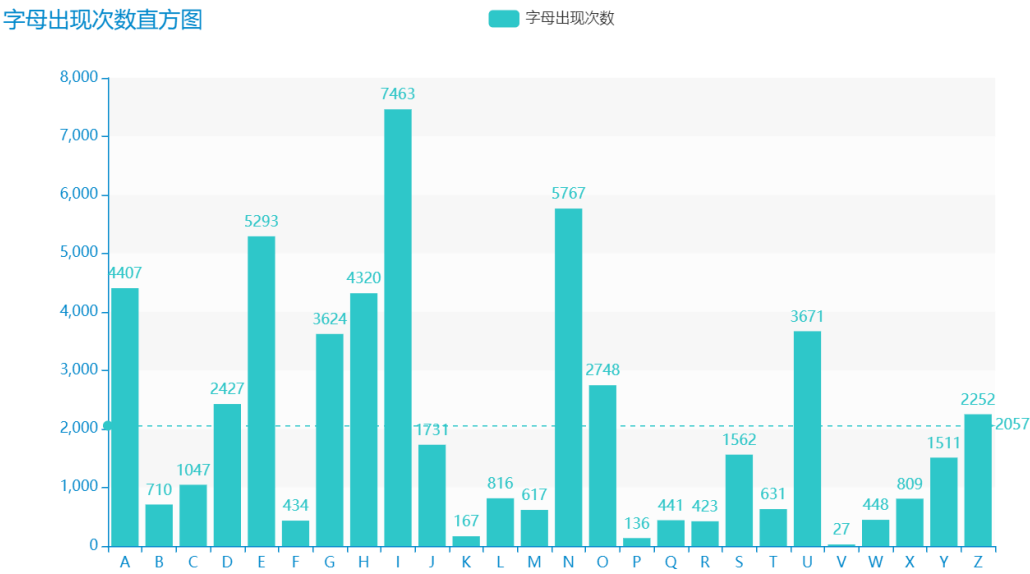


图 2-1 全拼输入时各字母出现的次数直方图

表 2-1 全拼时各字母出现的频率

A	0.08240155566358776	N	0.107830672001795
B	0.013275494558917019	O	0.051381773306907
C	0.019576680004487493	P	0.0025429116338207248
D	0.04537975393590367	Q	0.008245764930256909
E	0.09896787704274335	R	0.007909203096368871
F	0.008114879772633784	S	0.029206088029617442
G	0.06776111588945813	T	0.011798362065741745
H	0.0807748401331289	U	0.06863991623349912
I	0.13954227590591228	V	0.0005048427508320556
J	0.0323660296922329	W	0.008376650087880034
K	0.0031225459032945663	X	0.015126584645301222
L	0.015257469802924348	Y	0.02825249616693467
M	0.011536591750495494	Z	0.042107624995325534

2) 各个手指移动距离和敲击次数

依照目前公认的标准打字手法，在打字过程中，双手的放置方式如下：

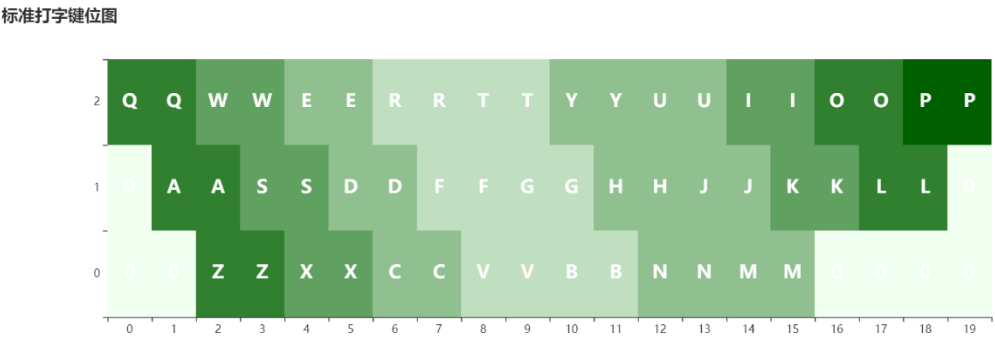


图 2-2 使用标准打字方式的键位图

其中，不同颜色的色块代表不同的手指的放置位置，从左到右依次为左手小指、左手无名指、左手中指、左手拇指、右手拇指、右手中指、右手无名指、右手小指所负责的区域，各手指在打字过程中负责的区域如表 2-2 所示：

表 2-2 标准打字方式下手指负责区域

左手				右手			
小指	无名指	中指	食指	食指	中指	无名指	小指
Q. A. Z	W. S. X	E. D. C	R. F. V. T. G. B	Y. H. N. U. J. M	I. K	O. L	P

假设打字者是明智的，能严格遵循标准打字方式，即手指只在各自负责的字母区域移动，移动过程中按最短路径转移到按键中点位置，且初始时除右手小指外的其他手指均处在中间一排按键上，可列出每个手指的移动距离矩阵：

表 2-3 左手小指移动矩阵

	Q	A	Z
Q	0	1. 118	2. 236
A	1. 118	0	1. 118
Z	2. 236	1. 118	0

表 2-4 左手无名指移动矩阵

	W	S	X
W	0	1. 118	2. 236
S	1. 118	0	1. 118
X	2. 236	1. 118	0

表 2-5 左手中指移动矩阵

	E	D	C
E	0	1. 118	2. 236
D	1. 118	0	1. 118
C	2. 236	1. 118	0

表 2-6 左手食指移动矩阵

	R	F	V	T	G	B
R	0	1. 118	2. 236	1	1. 803	2. 828
F	1. 118	0	1. 118	1. 118	1	1. 803
V	2. 236	1. 118	0	2	1. 118	1
T	1	1. 118	2	0	1. 118	2. 236
G	1. 803	1	1. 118	1. 118	0	1. 118
B	2. 828	1. 803	1	2. 236	1. 118	0

表 2-7 右手中指移动矩阵

	I	K
I	0	1. 118
K	1. 118	0

表 2-8 右手食指移动矩阵

	Y	H	N	U	J	M
Y	0	1.118	2.236	1	1.803	2.828
H	1.118	0	1.118	1.118	1	1.803
N	2.236	1.118	0	2	1.118	1
U	1	1.118	2	0	1.118	2.236
J	1.803	1	1.118	1.118	0	1.118
M	2.828	1.803	1	2.236	1.118	0

表 2-9 右手无名指移动矩阵

	0	L
0	0	1.118
L	1.118	0

将文章中出现的字母按每个手指负责的区域分成八类,计算出每一类中手指移动的距离,计算方式为:

$$dis[finger] = \sum_{i=1}^m \sum_{j=1}^n dis[i][j] * times[i][j], \quad i \neq j$$

其中, $dis[i][j]$ 为手指从键 i 转移到键 j 的距离, $times[i][j]$ 为键 i 转移到键 j 的次数,当 $i=j$ 时,距离为 0。由此计算出输入示例文章时,各个手指的移动距离:

表 2-10 输入示例文章时各个手指移动距离

	左手				右手			
	小指	无名指	中指	食指	食指	中指	无名指	小指
移动距离	4398.776	1999.492	7176.048	4625.765	19677.470	353.288	1416.506	0
敲击次数	7100	2819	8767	5849	17617	7630	3564	136

由表和图可见,移动距离最大的为右手食指,其次是左手中指,在输入示例文章时,左右手移动距离比约为 0.85,左手手指移动距离远小于右手手指移动距离;对于每个手指来说,右手食指和右手中指、小指移动距离差距极大,故在“手指移动距离和敲击次数”这一衡量指标下,拼音输入的平衡性较差。

手指移动距离直方图

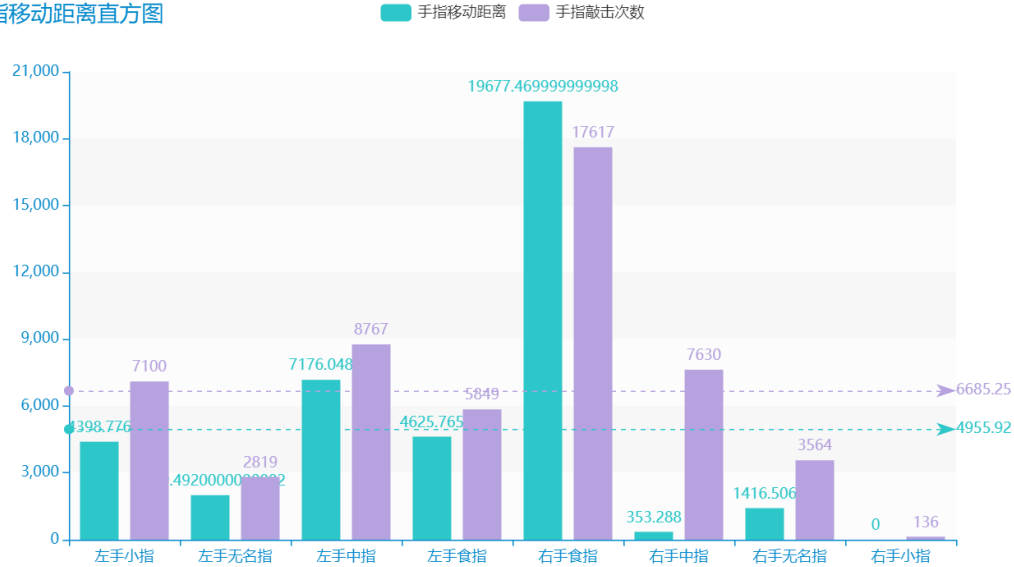


图 2-3 手指移动距离和敲击次数直方图

2. 输入效率

输入效率从三个方面进行评估：平均码长、信息熵、输入有效时间占比。

1) 平均码长

码长即全拼输入一个汉字时对应的拼音个数，将整篇文章按字翻译成拼音，统计字母总个数和字母数，得到示例文章的平均码长为 2.9519 字母/字。

2) 信息熵

信息熵常用来衡量不确定时间中的信息量多少，信息熵越大，其中包含的信息量越大，这一事件发生的不确定性越大。在一组随机事件中，事件概率分布越平均，其信息熵越大，信息熵的计算公式为：

$$H(x) = - \sum_{x \in X} p(x) \log_2 p(x)$$

其中， $p(x)$ 为事件 x 发生的概率，在示例文章中，首先统计拼音的种类数，计算出其所占频率，部分表格如下（剩余部分见附录）：

表 2-11 拼音种类数及所占频率

拼音	个数	出现概率	信息熵
a	1	5.519372999227288e-05	0.0007807228219605782
ai	6	0.00033116237995363726	0.0038282945979337454
an	8	0.00044154983993818304	0.004921133055870077

由此，可计算出本篇文章中，以拼音为一个单位字符时的信息熵为 6.955 比特，而英文的平均信息熵为 4.03 比特，中文平均信息熵为 9.65 比特，使用拼音输入这篇文章的信息熵大于英文小于中文，能传递较大信息量。

3) 输入有效时间占比

假设按键按下和抬起和手指移动一个按键距离均为一个单位时间，按键之间的空隙忽略不计，定义按键按下和抬起时间为有效时间，则输入有效时间占比为：

$$\text{有效时间比} = \frac{\text{按键按下和抬起时间}}{\text{按键按下抬起时间} + \text{手指移动时间}} \times 100\%$$

理想情况下，按键按下和抬起次数为一篇文章中全拼字母个数，手指移动距离如表 2-10 所示。由此得到：

$$\begin{aligned} \text{有效时间比} &= \frac{2.9519 \times 18118}{2.9519 \times 18118 + 4398.776 + 1999.492 + 4625.765 \dots} \times 100\% \\ &= 57.43\% \end{aligned}$$

有效时间比仅为 54.73%，效率较低。

三、新的编码方案

1. 编码方案描述

1) 拼音编码方案

由于文章确定，即信息熵确定，那么平均码长越小，输入效率越高，提出一种基于霍夫曼编码的新编码方案，其操作步骤如下：

- 统计每个拼音组合出现的次数，并算出其概率
- 使用霍夫曼编码方式，对原始字符集进行编码
- 将编码得到的二进制序列转化成 8421BCD 码（因为要在不增加字母的基础上进行编码，又 $2^4=16$ ，故选用 BCD 码）
- 将 BCD 码按 ASCII 码映射标准映射为 16 个字母的组合，用新的组合代替原拼音组合，即为新的编码方式

新编码方式如下（完整表见附录）：

表 3-1 新旧编码表

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为 BCD 码（倒序）	转化为拼音（正序）
dao	7	0000000	103	0 0	aa
lian	9	000000100	26	4 0 0	aae
kan	9	000000101	26	5 0 0	aaf
man	10	0000001100	13	12 0 0	aam
gou	10	0000001101	13	13 0 0	aan
ya	10	0000001110	13	14 0 0	aao
ma	10	0000001111	13	15 0 0	aap
di	7	0000010	104	2 0	ac
ju	8	00000110	52	6 0	ag
chang	8	00000111	53	7 0	ah
neng	8	00001000	53	8 0	ai
ba	8	00001001	53	9 0	aj

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
guan	7	0000101	106	5 0	af
hui	6	000011	213	3 0	ad
ge	6	000100	213	4 0	ae

由此便得到了与原编码一一对应的新编码表，相对于原来的 26 个字母，新编码表仅使用了 16 个字母。

2) 键盘排列方案

考虑的各个手指使用的平衡性、手指的总移动距离和手指的灵活性差异，给出键盘排列方案：

- 在新编码的基础上统计 16 个字母的使用次数，并从大到小排列
- 按照使用大频率放在第一排（初始时手指按到的地方）、大频率优先分配给拇指/右手（灵活性好）、最大频率对应最小频率（移动距离小）的原则分配键盘布局

新键盘布局 and 此布局下的热力图如下：

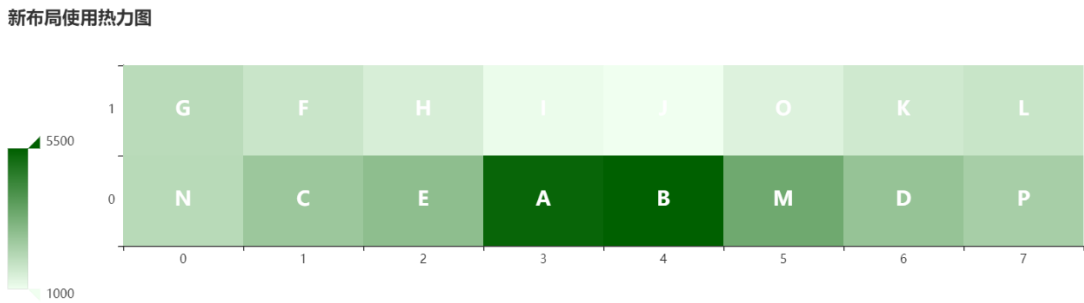


图 3-1 新键盘布局热力图

2. 新旧编码方案平衡性对比

1) 单字母出现次数均衡性

在新的编码方案下，各个字母出现次数如图 3-2 所示，从图中可以看出，在使用新编码方式输入时，各字母使用次数差异变小，使用次数最多的是“B”，为 5500 次，使用次数最少的是“J”，为 1006 次。经计算，样本的平均值为 2429.19，标准差为 1297.50，极值为 4494，字母使用次数标准差和极值相对旧编码都有较大改善，故在“字母出现次数的均衡性”的标准下，新的编码方式较旧的编码方式效果好。

新旧编码方式在数据上的对比如表 3-2 所示。

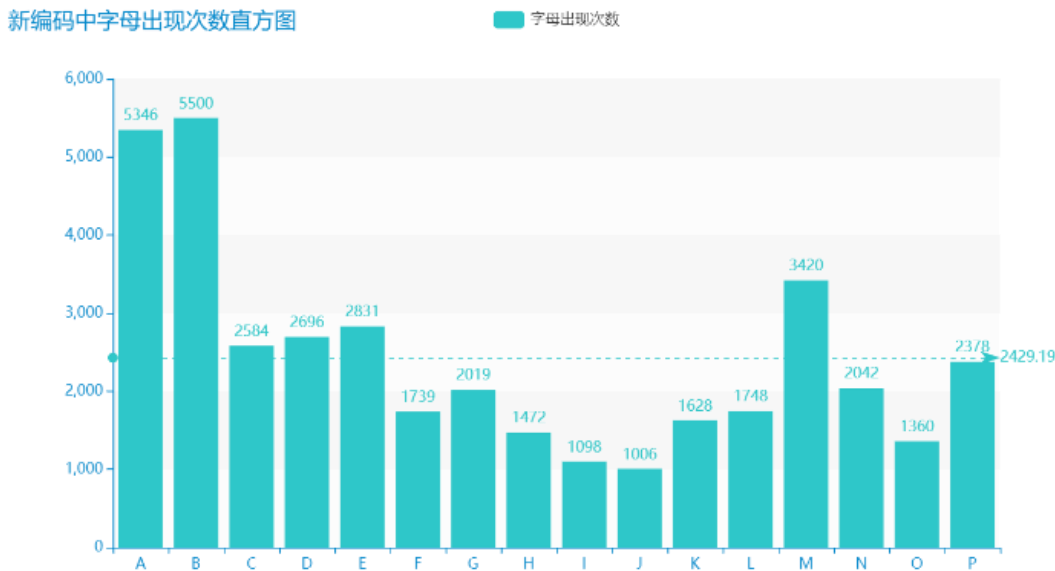


图 3-2 新编码中各个字母出现次数直方图

表 3-2 新旧编码对比

	最大值	最小值	极值	平均值	标准差
旧编码	7463	27	7436	2057.0	1984.492
新编码	5500	1006	4494	2429.19	1297.5

2) 各个手指移动距离和敲击次数

在新编码方案下，移动任一手指的距离都为 1，根据 2.1.2 节中的计算方式，可得新旧编码下，各个手指的移动距离为：

表 3-3 新旧编码下手指移动距离

	左手				右手			
	小指	无名指	中指	食指	食指	中指	无名指	小指
旧	4398.776	1999.492	7176.048	4625.765	19677.470	353.288	1416.506	0

新	7992	2110	2949	980	7762	6364	6264	4204
---	------	------	------	-----	------	------	------	------

表 3-4 新旧编码下手指敲击次数

	左手				右手			
	小指	无名指	中指	食指	食指	中指	无名指	小指
旧	7100	2819	8767	5849	17617	7630	3564	136
新	4061	4323	4303	6352	6506	4780	4324	4126

由表可见，在新编码方案下，各个手指的移动距离和敲击次数更加均衡，也有所降低。

3. 新旧编码方案输入效率对比

1) 平均码长

新编码方式下，平均码长为 2.1452，较原来的平均码长 2.9519 缩短了 27.33% 左右，输入字符总数减小，输入效率提高。

2) 有效输入时间比

按照 2.2.3 节中的算法，新编码方案下的有效输入时间比为：

$$\text{有效时间比} = \frac{2.1452 \times 18118}{2.1452 \times 18118 + 7992 + 2110 + 2949 \dots} \times 100\% = 50.16\%$$

旧编码的有效时间比为 54.73%，虽然新编码在有效时间比方面并未提高，但其输入的总时间从 93129.87 个时间单位缩短到了 77491.73 个时间单位。

附录一：拼音种类数及其频率

拼音	个数	出现概率	信息熵
a	1	5.519372999227288e-05	0.0007807228219605782
ai	6	0.00033116237995363726	0.0038282945979337454
an	8	0.00044154983993818304	0.004921133055870077
ao	1	5.519372999227288e-05	0.0007807228219605782
ba	53	0.0029252676895904625	0.024622608929632604
bai	9	0.0004967435699304559	0.0054518655360169446
ban	33	0.001821393089745005	0.016576028533749256
bang	4	0.00022077491996909152	0.00268134144790413
bao	30	0.0016558118997681863	0.015296796819748158
bei	64	0.0035323987195054643	0.028771868288444222
ben	59	0.0032564300695441	0.026906230781337623
beng	1	5.519372999227288e-05	0.0007807228219605782
bi	41	0.002262942929683188	0.019885801271123223
bian	50	0.0027596864996136438	0.023460867365346704
biao	21	0.0011590683298377303	0.0113041832418348
bie	15	0.0008279059498840931	0.008476304359758172
bing	33	0.001821393089745005	0.016576028533749256

拼音	个数	出现概率	信息熵
bo	25	0.0013798432498068219	0.013110276932480175
bu	272	0.015012694557898222	0.09094199818923421
cai	62	0.0034220112595209186	0.028029488146014057
can	3	0.00016558118997681863	0.0020797284889436914
cang	2	0.00011038745998454576	0.0014510581839366108
cao	3	0.00016558118997681863	0.0020797284889436914
ceng	15	0.0008279059498840931	0.008476304359758172
cha	7	0.00038635610994591015	0.004380421026800368
chan	490	0.02704492769621371	0.14086345585311852
chang	53	0.0029252676895904625	0.024622608929632604
chao	6	0.00033116237995363726	0.0038282945979337454
che	5	0.0002759686499613644	0.003262834748149463
chen	3	0.00016558118997681863	0.0020797284889436914
cheng	112	0.006181697759134562	0.04535994539226764
chi	15	0.0008279059498840931	0.008476304359758172
chong	15	0.0008279059498840931	0.008476304359758172
chou	3	0.00016558118997681863	0.0020797284889436914
chu	73	0.00402914228943592	0.03205308210874969
chuan	9	0.0004967435699304559	0.0054518655360169446

拼音	个数	出现概率	信息熵
chuang	12	0.0006623247599072745	0.006994264435960216
chui	1	5.519372999227288e-05	0.0007807228219605782
chun	11	0.0006071310299150017	0.006487622760022578
chuo	1	5.519372999227288e-05	0.0007807228219605782
ci	48	0.002649299039629098	0.022678459664582668
cong	35	0.0019317805497295506	0.017416649602427843
cu	6	0.00033116237995363726	0.0038282945979337454
cui	9	0.0004967435699304559	0.0054518655360169446
cun	43	0.0023733303896677337	0.02069276229958744
cuo	5	0.0002759686499613644	0.003262834748149463
da	67	0.003697979909482283	0.029876153115918257
dai	90	0.004967435699304559	0.03801719115110318
dan	60	0.0033116237995363725	0.027281969839959945
dang	83	0.004581079589358649	0.03559543120242803
dao	103	0.005684954189204106	0.04240200147881386
de	1392	0.07682967214924384	0.28443824310998866
deng	38	0.0020973617397063694	0.01866066485505301
di	104	0.005740147919196379	0.04273365837811177
dian	12	0.0006623247599072745	0.006994264435960216

拼音	个数	出现概率	信息熵
diao	17	0.0009382934098686389	0.009437048526301693
ding	15	0.0008279059498840931	0.008476304359758172
dong	121	0.0066784413290650185	0.04826023926326673
dou	45	0.0024837178496522794	0.021492313425203868
du	97	0.005353791809250469	0.04039555441390766
duan	37	0.0020421680097140963	0.01824816540115652
dui	109	0.006016116569157744	0.04438060173429884
dun	4	0.00022077491996909152	0.00268134144790413
duo	33	0.001821393089745005	0.016576028533749256
e	6	0.00033116237995363726	0.0038282945979337454
er	106	0.005850535379180925	0.04339468248008428
fa	135	0.007451153548956839	0.052667141313399654
fan	71	0.003918754829451374	0.0313319691347374
fang	63	0.0034772049895131912	0.02840131020980549
fei	27	0.0014902307097913678	0.013993636815608444
fen	51	0.002814880229605917	0.023849665970958348
feng	32	0.0017661993597527321	0.016152133503974844
fo	4	0.00022077491996909152	0.00268134144790413
fou	1	5.519372999227288e-05	0.0007807228219605782

拼音	个数	出现概率	信息熵
fu	50	0.0027596864996136438	0.023460867365346704
gai	30	0.0016558118997681863	0.015296796819748158
gan	11	0.0006071310299150017	0.006487622760022578
gang	4	0.00022077491996909152	0.00268134144790413
gao	9	0.0004967435699304559	0.0054518655360169446
ge	213	0.011756264488354123	0.07536266904161114
gei	9	0.0004967435699304559	0.0054518655360169446
gen	6	0.00033116237995363726	0.0038282945979337454
geng	15	0.0008279059498840931	0.008476304359758172
gong	271	0.01495750082790595	0.09068713377400689
gou	13	0.0007175184898995474	0.007494262766962614
gu	35	0.0019317805497295506	0.017416649602427843
guai	2	0.00011038745998454576	0.0014510581839366108
guan	106	0.005850535379180925	0.04339468248008428
guang	5	0.0002759686499613644	0.003262834748149463
gui	23	0.0012694557898222762	0.012214162989356944
guo	195	0.010762777348493211	0.07036494789898269
ha	2	0.00011038745998454576	0.0014510581839366108
hai	5	0.0002759686499613644	0.003262834748149463

拼音	个数	出现概率	信息熵
han	2	0.00011038745998454576	0.0014510581839366108
hang	5	0.0002759686499613644	0.003262834748149463
hao	14	0.0007727122198918203	0.007988129833708916
he	228	0.012584170438238217	0.07943438044478858
hen	9	0.0004967435699304559	0.0054518655360169446
heng	6	0.00033116237995363726	0.0038282945979337454
hong	1	5.519372999227288e-05	0.0007807228219605782
hou	40	0.002207749199690915	0.01947943038612296
hu	37	0.0020421680097140963	0.01824816540115652
hua	38	0.0020973617397063694	0.01866066485505301
huai	6	0.00033116237995363726	0.0038282945979337454
huan	47	0.002594105309636825	0.022284784199160697
huang	8	0.00044154983993818304	0.004921133055870077
hui	213	0.011756264488354123	0.07536266904161114
hun	3	0.00016558118997681863	0.0020797284889436914
huo	81	0.0044706921293741035	0.034895031069498164
ji	606	0.03344740037531736	0.16395788914417353
jia	88	0.0048570482393200136	0.03732983736222058
jian	142	0.007837509658902748	0.054826428610572055

拼音	个数	出现概率	信息熵
jiang	15	0.0008279059498840931	0.008476304359758172
jiao	57	0.0031460426095595542	0.026150680330316253
jie	438	0.02417485373661552	0.12982740228292833
jin	87	0.00480185450932774	0.036984808231685255
jing	88	0.0048570482393200136	0.03732983736222058
jiu	118	0.0065128601390882	0.047299601423587044
ju	52	0.0028700739595981894	0.024236903148654076
juan	3	0.00016558118997681863	0.0020797284889436914
jue	22	0.0012142620598300034	0.011760983460215152
jun	15	0.0008279059498840931	0.008476304359758172
kai	25	0.0013798432498068219	0.013110276932480175
kan	26	0.0014350369797990947	0.013553488554126133
kang	6	0.00033116237995363726	0.0038282945979337454
kao	4	0.00022077491996909152	0.00268134144790413
ke	50	0.0027596864996136438	0.023460867365346704
ken	2	0.00011038745998454576	0.0014510581839366108
kong	13	0.0007175184898995474	0.007494262766962614
kou	8	0.00044154983993818304	0.004921133055870077
ku	3	0.00016558118997681863	0.0020797284889436914

拼音	个数	出现概率	信息熵
kua	2	0.00011038745998454576	0.0014510581839366108
kuai	5	0.0002759686499613644	0.003262834748149463
kuang	9	0.0004967435699304559	0.0054518655360169446
kui	1	5.519372999227288e-05	0.0007807228219605782
kun	3	0.00016558118997681863	0.0020797284889436914
kuo	10	0.0005519372999227288	0.005973732196376197
la	5	0.0002759686499613644	0.003262834748149463
lai	125	0.00689921624903411	0.04953190062106516
lan	4	0.00022077491996909152	0.00268134144790413
lao	46	0.0025389115796445524	0.021889414399069336
le	176	0.009714096478640027	0.06494557824580113
lei	7	0.00038635610994591015	0.004380421026800368
leng	4	0.00022077491996909152	0.00268134144790413
li	267	0.014736725907936859	0.08966472689097318
lian	26	0.0014350369797990947	0.013553488554126133
liang	29	0.0016006181697759134	0.014865189187629505
liao	8	0.00044154983993818304	0.004921133055870077
lie	12	0.0006623247599072745	0.006994264435960216
ling	27	0.0014902307097913678	0.013993636815608444

拼音	个数	出现概率	信息熵
liu	8	0.00044154983993818304	0.004921133055870077
long	3	0.00016558118997681863	0.0020797284889436914
lou	2	0.00011038745998454576	0.0014510581839366108
lu	15	0.0008279059498840931	0.008476304359758172
luan	2	0.00011038745998454576	0.0014510581839366108
lun	20	0.0011038745998454575	0.010843589792906937
luo	11	0.0006071310299150017	0.006487622760022578
lv	17	0.0009382934098686389	0.009437048526301693
lve	2	0.00011038745998454576	0.0014510581839366108
ma	13	0.0007175184898995474	0.007494262766962614
mai	23	0.0012694557898222762	0.012214162989356944
man	13	0.0007175184898995474	0.007494262766962614
mang	3	0.00016558118997681863	0.0020797284889436914
mao	11	0.0006071310299150017	0.006487622760022578
mei	48	0.002649299039629098	0.022678459664582668
men	214	0.011811458218346395	0.0756366700409786
meng	6	0.00033116237995363726	0.0038282945979337454
mi	9	0.0004967435699304559	0.0054518655360169446
mian	49	0.002704492769621371	0.02307047609913682

拼音	个数	出现概率	信息熵
miao	1	5.519372999227288e-05	0.0007807228219605782
mie	57	0.0031460426095595542	0.026150680330316253
min	64	0.0035323987195054643	0.028771868288444222
ming	73	0.00402914228943592	0.03205308210874969
mo	7	0.00038635610994591015	0.004380421026800368
mou	7	0.00038635610994591015	0.004380421026800368
mu	19	0.0010486808698531847	0.010379013297379691
na	43	0.0023733303896677337	0.02069276229958744
nan	6	0.00033116237995363726	0.0038282945979337454
nang	1	5.519372999227288e-05	0.0007807228219605782
nei	21	0.0011590683298377303	0.0113041832418348
neng	53	0.0029252676895904625	0.024622608929632604
ni	39	0.0021525554696986424	0.019071068600693276
nian	20	0.0011038745998454575	0.010843589792906937
nie	1	5.519372999227288e-05	0.0007807228219605782
ning	1	5.519372999227288e-05	0.0007807228219605782
nong	25	0.0013798432498068219	0.013110276932480175
nu	16	0.0008830996798763661	0.008959166431863787
nuo	1	5.519372999227288e-05	0.0007807228219605782

拼音	个数	出现概率	信息熵
nv	8	0.00044154983993818304	0.004921133055870077
ou	7	0.00038635610994591015	0.004380421026800368
pai	17	0.0009382934098686389	0.009437048526301693
pan	10	0.0005519372999227288	0.005973732196376197
pang	2	0.00011038745998454576	0.0014510581839366108
pao	4	0.00022077491996909152	0.00268134144790413
pei	7	0.00038635610994591015	0.004380421026800368
peng	2	0.00011038745998454576	0.0014510581839366108
pi	17	0.0009382934098686389	0.009437048526301693
pian	5	0.0002759686499613644	0.003262834748149463
pie	2	0.00011038745998454576	0.0014510581839366108
pin	25	0.0013798432498068219	0.013110276932480175
ping	12	0.0006623247599072745	0.006994264435960216
po	21	0.0011590683298377303	0.0113041832418348
pu	12	0.0006623247599072745	0.006994264435960216
qi	135	0.007451153548956839	0.052667141313399654
qia	1	5.519372999227288e-05	0.0007807228219605782
qian	25	0.0013798432498068219	0.013110276932480175
qiang	12	0.0006623247599072745	0.006994264435960216

拼音	个数	出现概率	信息熵
qiao	2	0.00011038745998454576	0.0014510581839366108
qie	87	0.00480185450932774	0.036984808231685255
qin	2	0.00011038745998454576	0.0014510581839366108
qing	10	0.0005519372999227288	0.005973732196376197
qiu	18	0.0009934871398609118	0.009910243932172977
qu	79	0.004360304669389557	0.034190698303801066
quan	47	0.002594105309636825	0.022284784199160697
que	18	0.0009934871398609118	0.009910243932172977
qun	5	0.0002759686499613644	0.003262834748149463
ran	31	0.0017110056297604593	0.015725749702767487
rang	4	0.00022077491996909152	0.00268134144790413
rao	1	5.519372999227288e-05	0.0007807228219605782
re	2	0.00011038745998454576	0.0014510581839366108
ren	231	0.012749751628215034	0.08023912174776335
ri	14	0.0007727122198918203	0.007988129833708916
rong	9	0.0004967435699304559	0.0054518655360169446
ru	22	0.0012142620598300034	0.011760983460215152
rui	2	0.00011038745998454576	0.0014510581839366108
ruo	1	5.519372999227288e-05	0.0007807228219605782

拼音	个数	出现概率	信息熵
sa	2	0.00011038745998454576	0.0014510581839366108
sai	2	0.00011038745998454576	0.0014510581839366108
san	7	0.00038635610994591015	0.004380421026800368
se	7	0.00038635610994591015	0.004380421026800368
seng	5	0.0002759686499613644	0.003262834748149463
sha	2	0.00011038745998454576	0.0014510581839366108
shan	6	0.00033116237995363726	0.0038282945979337454
shang	68	0.0037531736394745556	0.030241846826257662
shao	16	0.0008830996798763661	0.008959166431863787
she	164	0.009051771718732752	0.06143966164702739
shen	51	0.002814880229605917	0.023849665970958348
sheng	161	0.008886190528755934	0.06055245020625397
shi	731	0.040346616624351475	0.1868616628707767
shou	67	0.003697979909482283	0.029876153115918257
shu	42	0.0023181366596754607	0.02029022982399414
shuang	1	5.519372999227288e-05	0.0007807228219605782
shui	10	0.0005519372999227288	0.005973732196376197
shuo	49	0.002704492769621371	0.02307047609913682
si	44	0.0024285241196600068	0.021093442800770298

拼音	个数	出现概率	信息熵
su	14	0.0007727122198918203	0.007988129833708916
suan	2	0.00011038745998454576	0.0014510581839366108
sui	24	0.001324649519814549	0.012663879352105884
suo	87	0.00480185450932774	0.036984808231685255
ta	239	0.013191301468153218	0.08237003852477871
tai	16	0.0008830996798763661	0.008959166431863787
tan	12	0.0006623247599072745	0.006994264435960216
tang	3	0.00016558118997681863	0.0020797284889436914
tao	7	0.00038635610994591015	0.004380421026800368
te	16	0.0008830996798763661	0.008959166431863787
ti	51	0.002814880229605917	0.023849665970958348
tian	10	0.0005519372999227288	0.005973732196376197
tiao	42	0.0023181366596754607	0.02029022982399414
tie	3	0.00016558118997681863	0.0020797284889436914
ting	16	0.0008830996798763661	0.008959166431863787
tong	166	0.009162159178717297	0.06202870322613876
tou	3	0.00016558118997681863	0.0020797284889436914
tu	28	0.0015454244397836406	0.01443083522763419
tuan	3	0.00016558118997681863	0.0020797284889436914

拼音	个数	出现概率	信息熵
tui	11	0.0006071310299150017	0.006487622760022578
tun	1	5.519372999227288e-05	0.0007807228219605782
tuo	4	0.00022077491996909152	0.00268134144790413
wa	7	0.00038635610994591015	0.004380421026800368
wai	10	0.0005519372999227288	0.005973732196376197
wan	22	0.0012142620598300034	0.011760983460215152
wang	17	0.0009382934098686389	0.009437048526301693
wei	152	0.008389446958825477	0.057863765502561094
wen	50	0.0027596864996136438	0.023460867365346704
weng	1	5.519372999227288e-05	0.0007807228219605782
wo	38	0.0020973617397063694	0.01866066485505301
wu	151	0.008334253228833204	0.057562448045725696
xi	95	0.005243404349265924	0.03972025861548332
xia	38	0.0020973617397063694	0.01866066485505301
xian	121	0.0066784413290650185	0.04826023926326673
xiang	97	0.005353791809250469	0.04039555441390766
xiao	113	0.006236891489126835	0.04568496275653617
xie	68	0.0037531736394745556	0.030241846826257662
xin	49	0.002704492769621371	0.02307047609913682

拼音	个数	出现概率	信息熵
xing	121	0.0066784413290650185	0.04826023926326673
xiong	4	0.00022077491996909152	0.00268134144790413
xiu	3	0.00016558118997681863	0.0020797284889436914
xu	40	0.002207749199690915	0.01947943038612296
xuan	10	0.0005519372999227288	0.005973732196376197
xue	45	0.0024837178496522794	0.021492313425203868
xun	5	0.0002759686499613644	0.003262834748149463
ya	13	0.0007175184898995474	0.007494262766962614
yan	31	0.0017110056297604593	0.015725749702767487
yang	51	0.002814880229605917	0.023849665970958348
yao	93	0.0051330168892813775	0.03904160982322311
ye	136	0.007506347278949111	0.052977346373566214
yi	539	0.02974942046583508	0.1508591512948596
yin	64	0.0035323987195054643	0.028771868288444222
ying	37	0.0020421680097140963	0.01824816540115652
yong	77	0.004249917209405011	0.03348233332399756
you	232	0.012804945358207307	0.0805066774264141
yu	120	0.006623247599072745	0.04794069208084714
yuan	37	0.0020421680097140963	0.01824816540115652

拼音	个数	出现概率	信息熵
yue	42	0.0023181366596754607	0.02029022982399414
yun	39	0.0021525554696986424	0.019071068600693276
zai	229	0.012639364168230489	0.07970297476224318
zan	3	0.00016558118997681863	0.0020797284889436914
zao	28	0.0015454244397836406	0.01443083522763419
ze	15	0.0008279059498840931	0.008476304359758172
zen	3	0.00016558118997681863	0.0020797284889436914
zeng	18	0.0009934871398609118	0.009910243932172977
zha	6	0.00033116237995363726	0.0038282945979337454
zhai	1	5.519372999227288e-05	0.0007807228219605782
zhan	79	0.004360304669389557	0.034190698303801066
zhang	17	0.0009382934098686389	0.009437048526301693
zhao	47	0.002594105309636825	0.022284784199160697
zhe	267	0.014736725907936859	0.08966472689097318
zhen	22	0.0012142620598300034	0.011760983460215152
zheng	193	0.010652389888508666	0.06980169206566064
zhi	352	0.019428192957280054	0.1104629635343222
zhong	221	0.012197814328292306	0.07754435422699867
zhou	14	0.0007727122198918203	0.007988129833708916

拼音	个数	出现概率	信息熵
zhu	134	0.007395959818964566	0.05235634641287194
zhua	1	5.519372999227288e-05	0.0007807228219605782
zhuan	14	0.0007727122198918203	0.007988129833708916
zhuang	18	0.0009934871398609118	0.009910243932172977
zhun	4	0.00022077491996909152	0.00268134144790413
zhuo	1	5.519372999227288e-05	0.0007807228219605782
zi	357	0.019704161607241418	0.11163108672360077
zong	29	0.0016006181697759134	0.014865189187629505
zou	4	0.00022077491996909152	0.00268134144790413
zu	77	0.004249917209405011	0.03348233332399756
zui	39	0.0021525554696986424	0.019071068600693276
zun	5	0.0002759686499613644	0.003262834748149463
zuo	54	0.0029804614195827355	0.02500681221163415

附录二：新旧编码表

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
dao	7	0000000	103	0 0	aa
lian	9	000000100	26	4 0 0	aae
kan	9	000000101	26	5 0 0	aaf
man	10	0000001100	13	12 0 0	aam
gou	10	0000001101	13	13 0 0	aan
ya	10	0000001110	13	14 0 0	aao
ma	10	0000001111	13	15 0 0	aap
di	7	0000010	104	2 0	ac
ju	8	00000110	52	6 0	ag
chang	8	00000111	53	7 0	ah
neng	8	00001000	53	8 0	ai
ba	8	00001001	53	9 0	aj
guan	7	0000101	106	5 0	af
hui	6	000011	213	3 0	ad
ge	6	000100	213	4 0	ae
er	7	0001010	106	10 0	ak

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
kong	10	0001011000	13	8 5 0	afi
mou	11	00010110010	7	2 11 0	alc
se	11	00010110011	7	3 11 0	ald
fei	9	000101101	27	13 2 0	acn
zuo	8	00010111	54	7 1	bh
men	6	000110	214	6 0	ag
dui	7	0001110	109	14 0	ao
ling	9	000111100	27	12 3 0	adm
hao	10	0001111010	14	10 7 0	ahk
su	10	0001111011	14	11 7 0	ahl
wa	11	00011111000	7	8 15 0	api
mo	11	00011111001	7	9 15 0	apj
san	11	00011111010	7	10 15 0	apk
tao	11	00011111011	7	11 15 0	apl
tu	9	000111111	28	15 3 0	adp
jie	5	00100	438	4 0	ae
zhong	6	001010	221	10 0	ak
cheng	7	0010110	112	6 1	bg

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
ri	10	0010111000	14	8 11 0	ali
zhou	10	0010111001	14	9 11 0	alj
lei	11	00101110100	7	4 7 1	bhe
ou	11	00101110101	7	5 7 1	bhf
zhuan	10	0010111011	14	11 11 0	all
zao	9	001011110	28	14 5 0	afo
zong	9	001011111	29	15 5 0	afp
xiao	7	0011000	113	8 1	bi
mie	8	00110010	57	2 3	dc
jiao	8	00110011	57	3 3	dd
he	6	001101	228	13 0	an
zai	6	001110	229	14 0	ao
ren	6	001111	231	15 0	ap
you	6	010000	232	0 1	ba
liang	9	010001000	29	8 8 0	aii
lu	10	0100010010	15	2 1 1	bbc
ze	10	0100010011	15	3 1 1	bbd
ben	8	01000101	59	5 4	ef

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
jiu	7	0100011	118	3 2	cd
ta	6	010010	239	2 1	bc
chong	10	0100110000	15	0 3 1	bda
ceng	10	0100110001	15	1 3 1	bdb
bie	10	0100110010	15	2 3 1	bdc
ding	10	0100110011	15	3 3 1	bdd
dan	8	01001101	60	13 4	en
gai	9	010011100	30	12 9 0	ajm
bao	9	010011101	30	13 9 0	ajn
chi	10	0100111100	15	12 3 1	bdm
jun	10	0100111101	15	13 3 1	bdn
cha	11	01001111100	7	12 7 2	chm
leng	12	010011111010	4	10 15 4	epk
dun	12	010011111011	4	11 15 4	ep1
geng	10	0100111111	15	15 3 1	bdp
yu	7	0101000	120	8 2	ci
xian	7	0101001	121	9 2	cj
dong	7	0101010	121	10 2	ck

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
xing	7	0101011	121	11 2	cl
chan	5	01011	490	11 0	al
jiang	10	0110000000	15	0 8 1	bia
an	11	01100000010	8	2 0 3	dac
nv	11	01100000011	8	3 0 3	dad
ran	9	011000001	31	1 12 0	amb
cai	8	01100001	62	1 6	gb
lai	7	0110001	125	1 3	db
yan	9	011001000	31	8 12 0	ami
tai	10	0110010010	16	2 9 1	bjc
nu	10	0110010011	16	3 9 1	bjd
fang	8	01100101	63	5 6	gf
bei	8	01100110	64	6 6	gg
yin	8	01100111	64	7 6	gh
shao	10	0110100000	16	0 10 1	bka
ting	10	0110100001	16	1 10 1	bkb
kao	12	011010001000	4	8 8 6	gii
zhun	12	011010001001	4	9 8 6	gij

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
gang	12	011010001010	4	10 8 6	gik
tuo	12	011010001011	4	11 8 6	gil
pao	12	011010001100	4	12 8 6	gim
fo	12	011010001101	4	13 8 6	gin
xiong	12	011010001110	4	14 8 6	gio
zou	12	011010001111	4	15 8 6	gip
min	8	01101001	64	9 6	gj
liu	11	01101010000	8	0 5 3	dfa
huang	11	01101010001	8	1 5 3	dfb
peng	13	0110101001000	2	0 8 4 13	anei
sa	13	0110101001001	2	0 9 4 13	anej
ken	13	0110101001010	2	0 10 4 13	anek
lve	13	0110101001011	2	0 11 4 13	anel
kou	11	01101010011	8	3 5 3	dfd
feng	9	011010101	32	5 13 0	anf
zhai	14	01101011000000	1	1 0 12 10	bkma
miao	14	01101011000001	1	1 1 12 10	bkmb

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
qia	14	01101011000010	1	1 2 12 10	bkmc
kui	14	01101011000011	1	1 3 12 10	bkmd
a	14	01101011000100	1	1 4 12 10	bkme
ning	14	01101011000101	1	1 5 12 10	bkmf
hong	14	01101011000110	1	1 6 12 10	bkmg
beng	14	01101011000111	1	1 7 12 10	bkmh
nang	14	01101011001000	1	1 8 12 10	bkmi
tun	14	01101011001001	1	1 9 12 10	bkmj
zhua	14	01101011001010	1	10 1 10 12	bkmk
nie	14	01101011001011	1	10 1 11 12	bkml
rao	14	01101011001100	1	10 1 12 12	bkmm
chuo	14	01101011001101	1	10 1 13 12	bkmn
chui	14	01101011001110	1	10 1 14 12	bkmo
ruo	14	01101011001111	1	10 1 15 12	bkmp
lan	12	011010110100	4	4 11 6	gle
bang	12	011010110101	4	5 11 6	glf

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
zhuo	14	01101011011000	1	1 8 13 10	bkni
shuang	14	01101011011001	1	1 9 13 10	bknj
fou	14	01101011011010	1	10 1 10 13	bknk
nuo	14	01101011011011	1	10 1 11 13	bknl
rang	12	011010110111	4	7 11 6	glh
pang	13	0110101110000	2	0 0 7 13	anha
suan	13	0110101110001	2	0 1 7 13	anhb
sai	13	0110101110010	2	0 2 7 13	anhc
qin	13	0110101110011	2	0 3 7 13	anhd
pie	13	0110101110100	2	0 4 7 13	anhe
re	13	0110101110101	2	0 5 7 13	anhf
luan	13	0110101110110	2	0 6 7 13	anhg
qiao	13	0110101110111	2	0 7 7 13	anhh
han	13	0110101111000	2	0 8 7 13	anhi
lou	13	0110101111001	2	0 9 7 13	anhj
weng	14	01101011110100	1	1 4 15 10	bkpe

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
ao	14	01101011110101	1	1 ^{5 15 10}	bkp ^f
cang	13	0110101111011	2	0 ^{11 7 13}	an ^h l
sha	13	0110101111100	2	0 ^{12 7 13}	an ^h m
rui	13	0110101111101	2	0 ^{13 7 13}	an ^h n
kua	13	0110101111110	2	0 ^{14 7 13}	an ^h o
guai	13	0110101111111	2	0 ^{15 7 13}	an ^h p
bing	9	011011000	33	8 13 0	an ⁱ
ban	9	011011001	33	9 13 0	an ^j
te	10	0110110100	16	4 11 1	bl ^e
liao	11	01101101010	8	10 6 3	dg ^k
gao	11	01101101011	9	11 6 3	dg ^l
duo	9	011011011	33	11 13 0	an ^l
shou	8	01101110	67	14 6	g ^o
da	8	01101111	67	15 6	g ^p
li	6	011100	267	12 1	bm
zhe	6	011101	267	13 1	bn
yi	5	01111	539	15 0	ap

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
zhu	7	1000000	134	0 4	ea
qi	7	1000001	135	1 4	eb
fa	7	1000010	135	2 4	ec
wang	10	1000011000	17	8 1 2	cbi
pai	10	1000011001	17	9 1 2	cbj
pi	10	1000011010	17	10 1 2	cbk
diao	10	1000011011	17	11 1 2	cbl
shang	8	10000111	68	7 8	ih
gong	6	100010	271	2 2	cc
bu	6	100011	272	3 2	cd
ye	7	1001000	136	8 4	ei
xie	8	10010010	68	2 9	jc
zhang	10	1001001100	17	12 4 2	cem
lv	10	1001001101	17	13 4 2	cen
gu	9	100100111	35	7 2 1	bch
cong	9	100101000	35	8 2 1	bci
zhuang	10	1001010010	18	2 5 2	cfc
zeng	10	1001010011	18	3 5 2	cfđ

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
fan	8	10010101	71	5 9	jf
jian	7	1001011	142	11 4	el
qiu	10	1001100000	18	0 6 2	cga
que	10	1001100001	18	1 6 2	cgb
gei	11	10011000100	9	4 12 4	eme
rong	11	10011000101	9	5 12 4	emf
hen	11	10011000110	9	6 12 4	emg
bai	11	10011000111	9	7 12 4	emh
kuang	11	10011001000	9	8 12 4	emi
chuan	11	10011001001	9	9 12 4	emj
mi	11	10011001010	9	10 12 4	emk
cui	11	10011001011	9	11 12 4	eml
hu	9	100110011	37	3 3 1	bdd
ming	8	10011010	73	10 9	jk
chu	8	10011011	73	11 9	j1
yuan	9	100111000	37	8 3 1	bdi
duan	9	100111001	37	9 3 1	bdj
ying	9	100111010	37	10 3 1	bdk

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
hua	9	100111011	38	11 3 1	bdl
wu	7	1001111	151	15 4	ep
ji	5	10100	606	4 1	be
wei	7	1010100	152	4 5	fe
wo	9	101010100	38	4 5 1	bfe
xia	9	101010101	38	5 5 1	bff
deng	9	101010110	38	6 5 1	bfg
mu	10	1010101110	19	14 10 2	cko
pan	11	10101011110	10	14 5 5	ffo
kuo	11	10101011111	10	15 5 5	ffp
yong	8	10101100	77	12 10	km
zu	8	10101101	77	13 10	kn
yun	9	101011100	39	12 5 1	bfm
zui	9	101011101	39	13 5 1	bfn
ni	9	101011110	39	14 5 1	bfo
pian	12	101011111000	5	8 15 10	kpi
guang	12	101011111001	5	9 15 10	kpj
che	12	101011111010	5	10 15 10	kpk

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
xun	12	101011111011	5	11 15 10	kpl
nian	10	1010111111	20	15 11 2	clp
zhan	8	10110000	79	0 11	la
qu	8	10110001	79	1 11	lb
hou	9	101100100	40	4 6 1	bge
xu	9	101100101	40	5 6 1	bgf
hang	12	101100110000	5	0 3 11	lda
qun	12	101100110001	5	1 3 11	ldb
ha	13	1011001100100	2	4 6 6 1	bgge
ku	13	1011001100101	3	5 6 6 1	bggf
hai	12	101100110011	5	3 3 11	ldd
seng	12	101100110100	5	4 3 11	lde
kuai	12	101100110101	5	5 3 11	ldf
cuo	12	101100110110	5	6 3 11	ldg
la	12	101100110111	5	7 3 11	ldh
wai	11	10110011100	10	12 9 5	fjm
xuan	11	10110011101	10	13 9 5	fjn
qing	11	10110011110	10	14 9 5	fjo

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
shui	11	10110011111	10	15 9 5	fjp
sheng	7	1011010	161	10 5	fk
huo	8	10110110	81	6 11	lg
lun	10	1011011100	20	12 13 2	cnm
tian	11	10110111010	10	10 11 5	flk
zun	12	101101110110	5	6 7 11	lhg
zan	13	1011011101110	3	14 14 6 1	bgoo
zen	13	1011011101111	3	15 14 6 1	bgop
bi	9	101101111	41	15 6 1	bgp
she	7	1011100	164	12 5	fm
tong	7	1011101	166	13 5	fn
dang	8	10111100	83	12 11	lm
shu	9	101111010	42	10 7 1	bhk
yue	9	101111011	42	11 7 1	bhl
biao	10	1011111000	21	8 15 2	cpi
po	10	1011111001	21	9 15 2	cpj
tiao	9	101111101	42	13 7 1	bhn

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
nei	10	1011111100	21	12 15 2	cpm
mao	11	10111111010	11	10 15 5	fpk
luo	11	10111111011	11	11 15 5	fpl
cun	9	101111111	43	15 7 1	bhp
de	4	1100	1392	12	m
qie	8	11010000	87	0 13	na
jin	8	11010001	87	1 13	nb
na	9	110100100	43	4 10 1	bke
jue	10	1101001010	22	10 4 3	dek
zhen	10	1101001011	22	11 4 3	del
suo	8	11010011	87	3 13	nd
gan	11	11010100000	11	0 10 6	gka
chun	11	11010100001	11	1 10 6	gkb
wan	10	1101010001	22	1 5 3	dfb
si	9	110101001	44	9 10 1	bkj
jia	8	11010101	88	5 13	nf
le	7	1101011	176	11 6	gl
zhi	6	110110	352	6 3	dg

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
zi	6	110111	357	7 3	dh
shi	5	11100	731	12 1	bm
jing	8	11101000	88	8 14	oi
ru	10	1110100100	22	4 10 3	dke
tui	11	11101001010	11	10 4 7	hek
e	12	111010010110	6	6 9 14	ojg
gen	12	111010010111	6	7 9 14	ojh
xue	9	111010011	45	3 13 1	bnd
dai	8	11101010	90	10 14	ok
dou	9	111010110	45	6 13 1	bng
gui	10	1110101110	23	14 10 3	dko
mai	10	1110101111	23	15 10 3	dkp
lao	9	111011000	46	8 13 1	bni
quan	9	111011001	47	9 13 1	bnj
yao	8	11101101	93	13 14	on
zhao	9	111011100	47	12 13 1	bnm
huan	9	111011101	47	13 13 1	bnn
xi	8	11101111	95	15 14	op

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
ci	9	111100000	48	0 14 1	boa
mei	9	111100001	48	1 14 1	bob
ping	11	11110001000	12	8 8 7	hii
tan	11	11110001001	12	9 8 7	hi j
can	13	1111000101000	3	1 8 2 14	boci
hun	13	1111000101001	3	1 9 2 14	boc j
tie	13	1111000101010	3	1 10 2 14	bock
kun	13	1111000101011	3	1 11 2 14	bocl
xiu	13	1111000101100	3	1 12 2 14	bocm
chen	13	1111000101101	3	1 13 2 14	bocn
chou	13	1111000101110	3	1 14 2 14	boco
mang	13	1111000101111	3	1 15 2 14	bocp
chuang	11	11110001100	12	12 8 7	him
lie	11	11110001101	12	13 8 7	hin
qiang	11	11110001110	12	14 8 7	hio
pu	11	11110001111	12	15 8 7	hip
zha	12	111100100000	6	0 2 15	pca

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
ai	12	111100100001	6	1 2 15	pcb
chao	12	111100100010	6	2 2 15	pcc
huai	12	111100100011	6	3 2 15	pcd
sui	10	1111001001	24	9 12 3	dmj
juan	13	1111001010000	3	0 5 14 1	bofa
tang	13	1111001010001	3	1 5 14 1	bofb
heng	12	111100101001	6	9 2 15	pcj
cao	13	1111001010100	3	4 5 14 1	bofe
tou	13	1111001010101	3	5 5 14 1	boff
long	13	1111001010110	3	6 5 14 1	bofg
tuan	13	1111001010111	3	7 5 14 1	bofh
shan	12	111100101100	6	12 2 15	pcm
nan	12	111100101101	6	13 2 15	pcn
kang	12	111100101110	6	14 2 15	pco
cu	12	111100101111	6	15 2 15	pcp
xiang	8	11110011	97	3 15	pd
zheng	7	1111010	193	10 7	hk

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
du	8	11110110	97	6 15	pg
shuo	9	111101110	49	14 14 1	boo
mian	9	111101111	49	15 14 1	bop
guo	7	1111100	195	12 7	hm
xin	9	111110100	49	4 15 1	bpe
pin	10	1111101010	25	10 14 3	dok
kai	10	1111101011	25	11 14 3	dol
ke	9	111110110	50	6 15 1	bpg
wen	9	111110111	50	7 15 1	bph
bian	9	111111000	50	8 15 1	bpi
fu	9	111111001	50	9 15 1	bpj
nong	10	1111110100	25	4 15 3	dpe
bo	10	1111110101	25	5 15 3	dpf
dian	11	11111101100	12	12 14 7	hom
meng	12	111111011010	6	15 10 13	pnk
pei	12	111111011011	7	15 11 13	pnl
qian	10	1111110111	25	7 15 3	dph

拼音	哈夫曼编码位数	哈夫曼编码	次数	转化为BCD码（倒序）	转化为拼音（正序）
fen	9	111111100	51	12 15 1	bpm
yang	9	111111101	51	13 15 1	bpn
shen	9	111111110	51	14 15 1	bpo
ti	9	111111111	51	15 15 1	bpp