

|                                                     |
|-----------------------------------------------------|
| Universal Multiple-Octet Coded Character Set<br>UCS |
|-----------------------------------------------------|

ISO/IEC JTC1/SC2/WG2    IRG N 989

Date: 2003-11-14

|               |                                                    |
|---------------|----------------------------------------------------|
| Source:       | Beijing UniHan Digital Technology Co., Ltd         |
| Title:        | CJK Ideographs in Ancient Chinese Classics (古籍用汉字) |
| Status :      | FYI                                                |
| Distribution: | IRG Members and Ideographic Experts                |
| Medium :      | Electronic                                         |

This paper reveals partial outcome from the Project, Hanzi Usage Frequency Statistics in Ancient Classics, which is one of the key projects sponsored by The National Language Committee of China (国家语委).

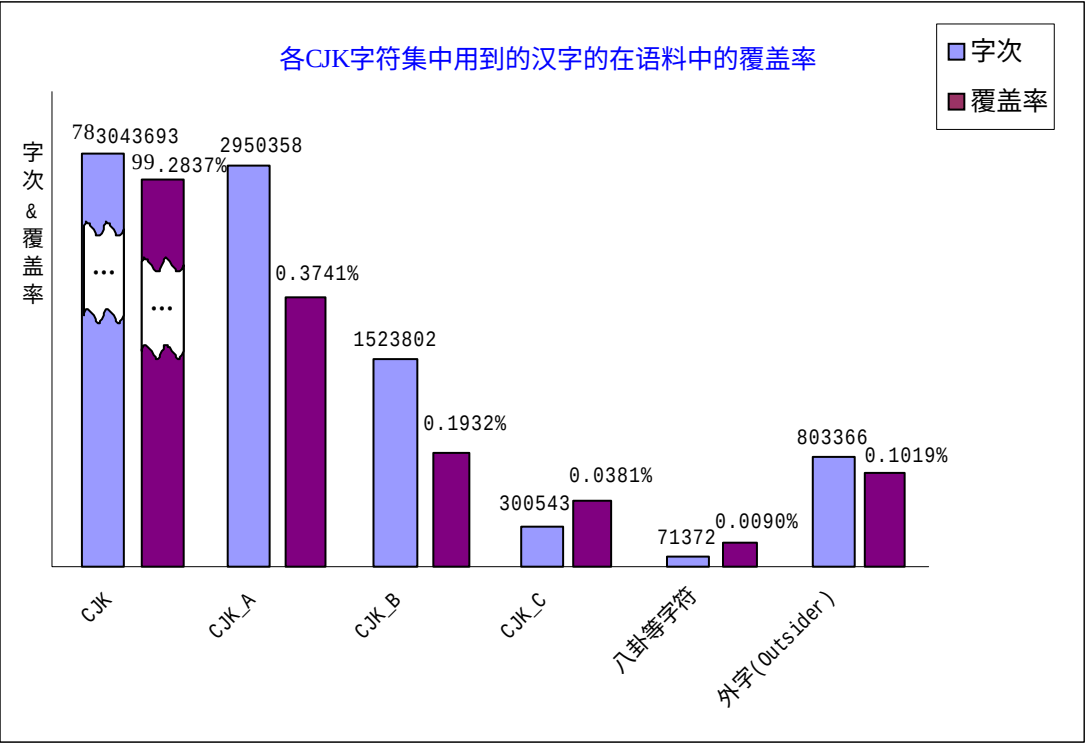
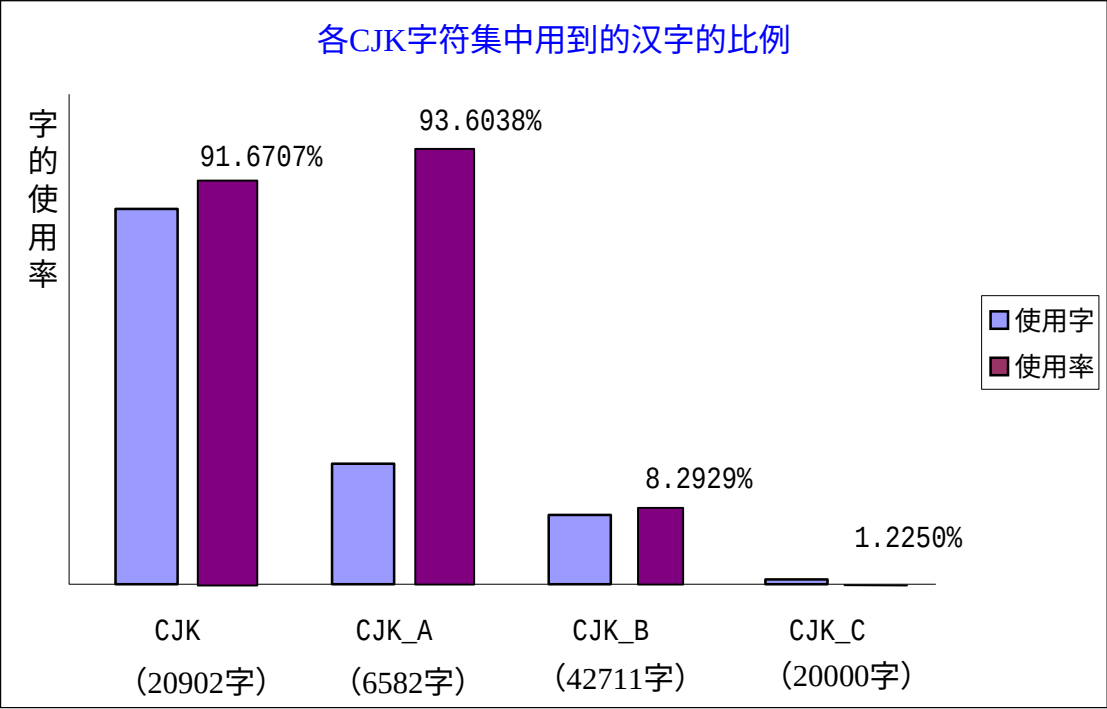
The project conducted statistics based on 800 million Hanzi data from large scale classics 《文渊阁四库全书》 and 《四部丛刊》 .

For more information, visit [www.unihan.com.cn](http://www.unihan.com.cn) later of Dec. 2003.

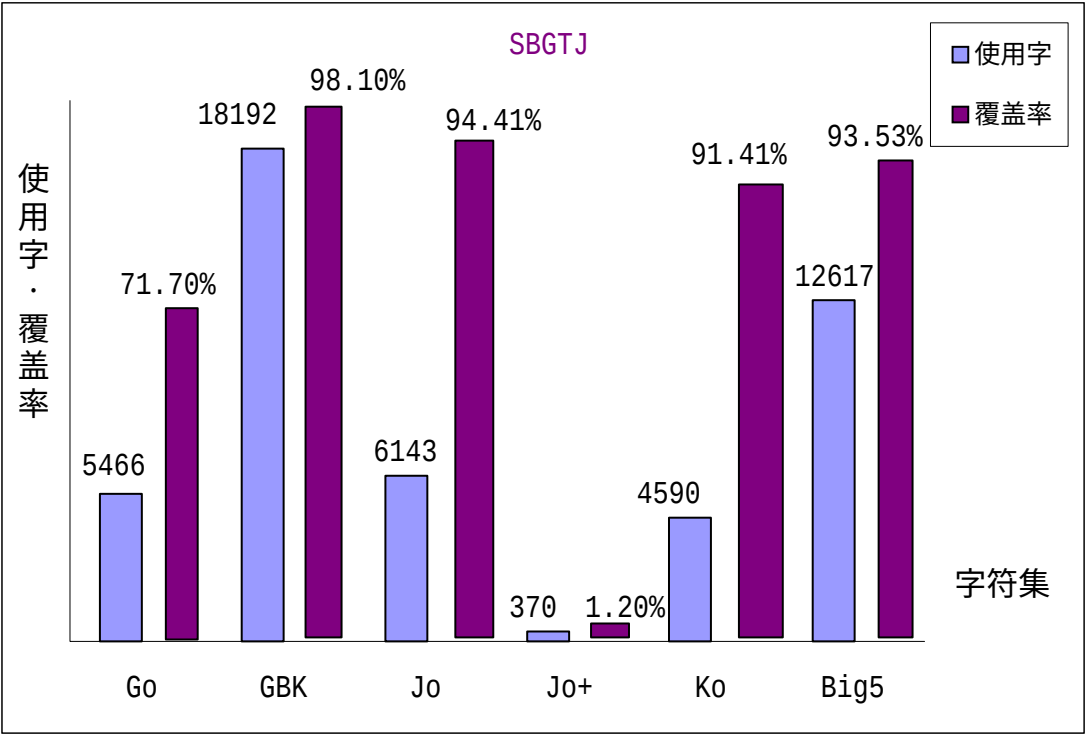
**1. The Number of Hanzi Used in the Classics (excluding the dictionary category 小学类) :**

The corpus contains 788,693,134 Hanzi-appearances (字次) from 39000 volumes of 4000 books. The exact number of Hanzi (字种) used is **30,135**.

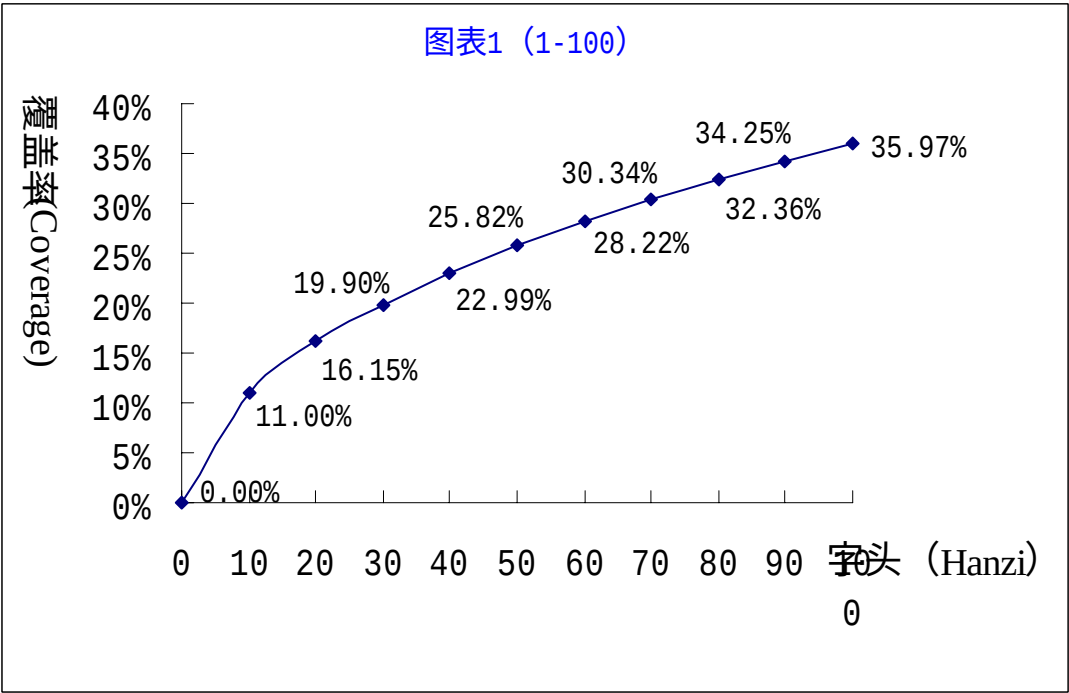
2. The Used Hanzi Allocation in CJK Blocks and Their Contribution.



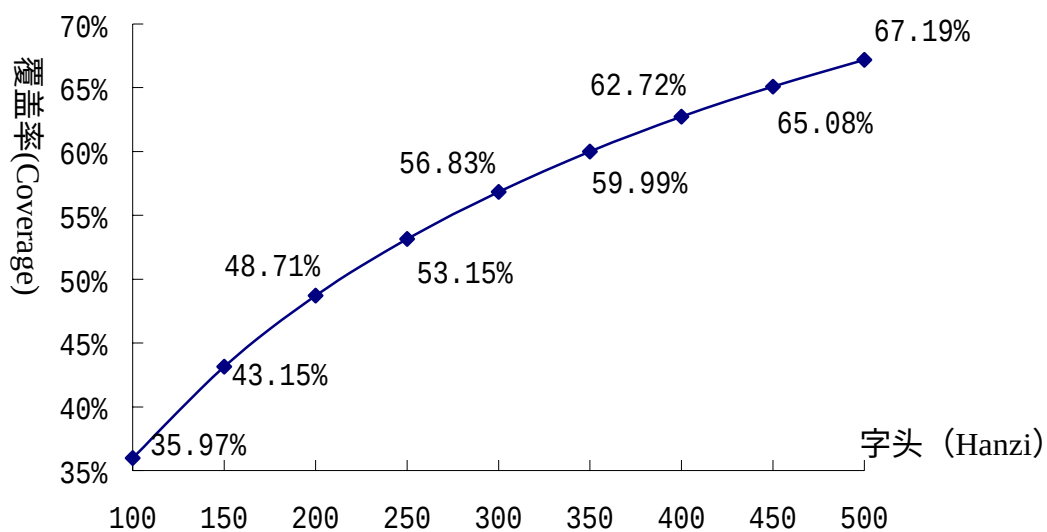
3. The Coverage of the Ideographs in C-J-K Code Pages to the Classics



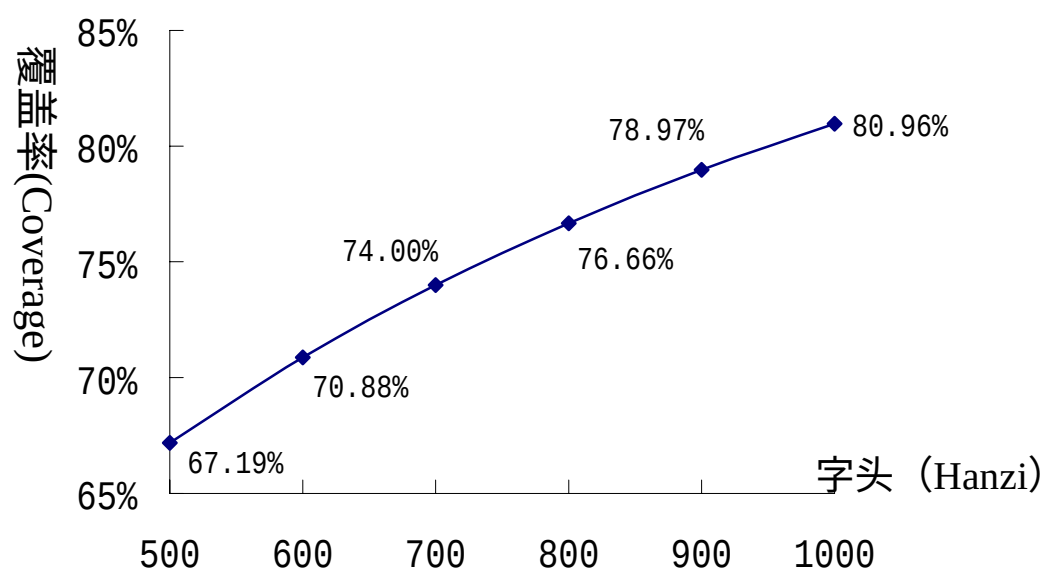
4. The Frequency Curves



图表2 (100-500)



图表3 (500-1000)



图表4 (1000-10000)

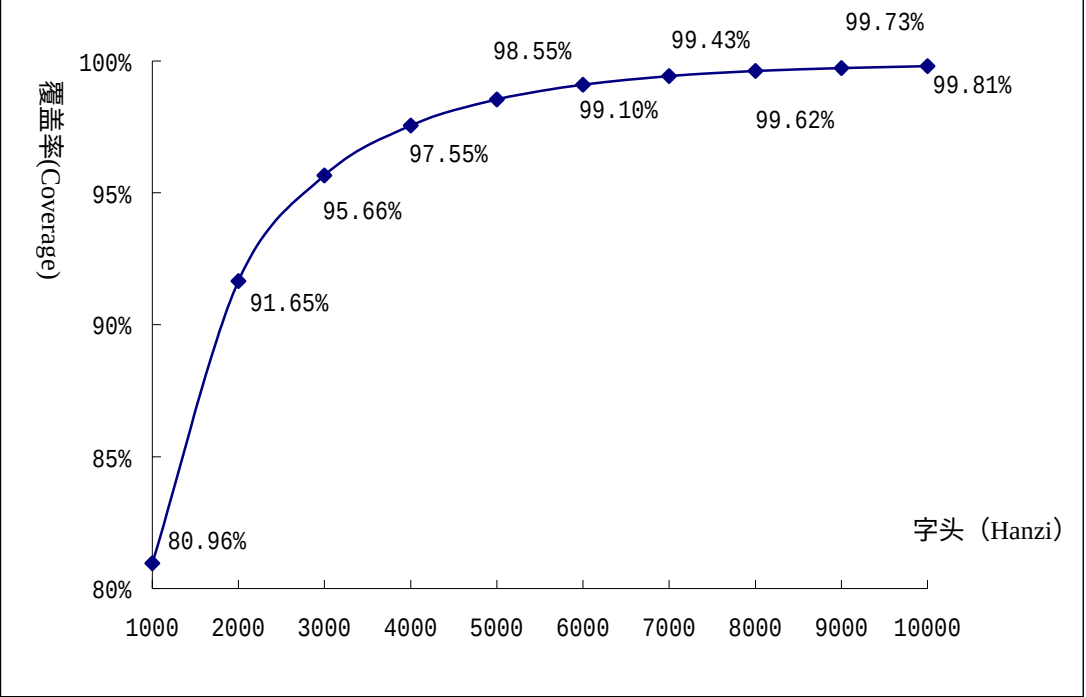
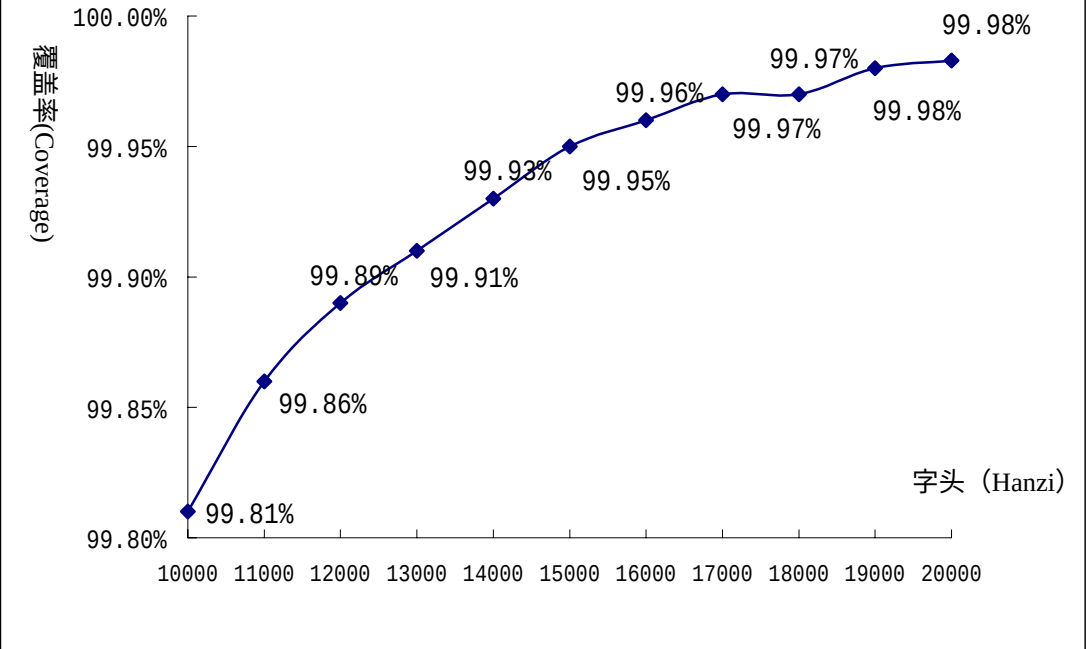
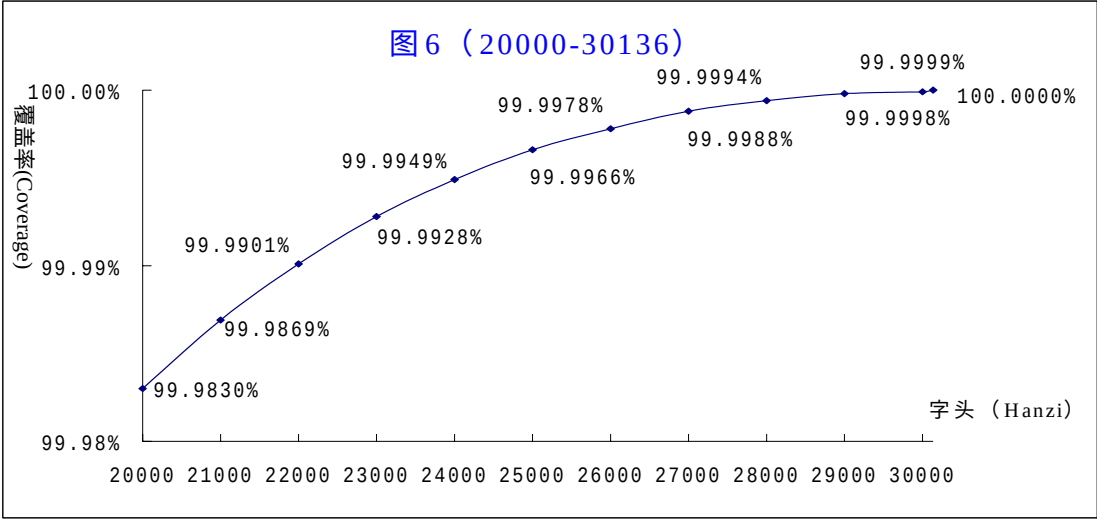


图5 (10000-30136)





## 5. The Samples of the Hanzi-Frequency List

国家语委十五科技攻关资助项目"古籍用字统计"

### 古籍字频统计

2003年8月28日版

©书同文数字化技术有限公司

|                                                            |                                                            |                                                            |                                                            |                                                            |                                                            |                                                            |                                                            |                                                            |                                                            |
|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------|
| No:00121 U:-05P06<br>從<br>000114504<br>14.51886% 39.2132%  | No:00122 U:-06T2C<br>本<br>000113869<br>14.45009% 38.2577%  | No:00123 U:-091CC<br>里<br>000113387<br>14.43389% 38.5020%  | No:00124 U:-06C11<br>民<br>000113544<br>14.39635% 38.6460%  | No:00125 U:-04E16<br>世<br>000113352<br>14.36679% 38.7868%  | No:00126 U:-05B06<br>家<br>0001127204<br>14.20205% 36.0325% | No:00127 U:-03F49<br>復<br>0001125786<br>14.27409% 40.0753% | No:00128 U:-0843C<br>萬<br>0001123460<br>14.24406% 40.2177% | No:00129 U:-04F96<br>來<br>0001107721<br>14.04809% 40.3582% | No:00130 U:-0324D<br>前<br>0001107002<br>14.03096% 40.4085% |
| No:00131 U:-0514B<br>八<br>0001102116<br>13.97340% 40.6283% | No:00132 U:-05206<br>分<br>0001099806<br>13.94215% 40.7777% | No:00133 U:-04E02<br>七<br>0001090712<br>13.93079% 40.9170% | No:00134 U:-05B57<br>字<br>0001085101<br>13.75035% 41.0540% | No:00135 U:-05B72<br>史<br>0001082850<br>13.72869% 41.1919% | No:00136 U:-05B78<br>學<br>0001060275<br>13.55759% 41.3275% | No:00137 U:-033B4<br>古<br>0001045829<br>13.51389% 41.4636% | No:00138 U:-0671D<br>朝<br>0001040727<br>13.44976% 41.5971% | No:00139 U:-06990<br>門<br>0001037137<br>13.40369% 41.7311% | No:00140 U:-03175<br>兵<br>0001036242<br>13.39235% 41.8631% |
| No:00141 U:-06C3F<br>江<br>0001041488<br>13.20529% 41.9971% | No:00142 U:-0591A<br>多<br>0001036721<br>13.14409% 42.1286% | No:00143 U:-033B8<br>去<br>0001034837<br>13.12094% 42.2598% | No:00144 U:-04B4E<br>乎<br>0001034450<br>13.10335% 42.3903% | No:00145 U:-05D53<br>必<br>0001027461<br>13.02790% 42.5211% | No:00146 U:-05C0F<br>小<br>0001015034<br>12.86889% 42.6488% | No:00147 U:-04E04<br>令<br>0001012880<br>12.84259% 42.7782% | No:00148 U:-07B49<br>義<br>0000986060<br>12.51429% 42.9034% | No:00149 U:-07D03<br>經<br>0000980014<br>12.48819% 43.0282% | No:00150 U:-07B93<br>石<br>0000977116<br>12.38695% 43.1521% |
| No:00151 U:-04B06<br>武<br>0000968830<br>12.28340% 43.2797% | No:00152 U:-05343<br>千<br>0000967872<br>12.27189% 43.3077% | No:00153 U:-06B32<br>欲<br>0000966404<br>12.25329% 43.5202% | No:00154 U:-06625<br>春<br>0000965580<br>12.24279% 43.6427% | No:00155 U:-05916<br>外<br>0000964738<br>12.23219% 43.7650% | No:00156 U:-06B78<br>數<br>0000963243<br>12.21319% 43.8871% | No:00157 U:-09BAC<br>馬<br>0000962209<br>12.21279% 44.0002% | No:00158 U:-0767D<br>白<br>0000961276<br>12.18829% 44.1211% | No:00159 U:-05C2B<br>師<br>0000960231<br>12.12429% 44.2524% | No:00160 U:-06B78<br>歸<br>0000951853<br>12.06629% 44.3720% |
| No:00161 U:-05B97<br>宗<br>0000950051<br>12.00000% 44.5000% | No:00162 U:-05372<br>即<br>0000948444<br>11.93333% 44.6333% | No:00163 U:-0337B<br>司<br>0000947271<br>11.86667% 44.7667% | No:00164 U:-06F22<br>漢<br>0000946108<br>11.80000% 44.9000% | No:00165 U:-05468<br>周<br>0000945042<br>11.73333% 45.0333% | No:00166 U:-07B49<br>等<br>0000944010<br>11.66667% 45.1667% | No:00167 U:-04E0B<br>主<br>0000942978<br>11.60000% 45.3000% | No:00168 U:-091D1<br>金<br>0000941946<br>11.53333% 45.4333% | No:00169 U:-0547D<br>命<br>0000940914<br>11.46667% 45.5667% | No:00170 U:-03C45<br>居<br>0000939882<br>11.40000% 45.7000% |

## 6. The Tool to Check Frequency of Hanzi Variants

书同文古籍字频查询系统 国家语委十五科技攻关重点项目

UniHan 北京书同文数字化技术有限公司

选择数据库 四部四库

☐ 单字查询 
☐ 多字查询

☐ 关联字查询

| No | Unicode | Hanzi | 字次       | 覆盖率%%     | 累加覆...   |
|----|---------|-------|----------|-----------|----------|
| 1  | 0-4E4B  | 之     | 19891628 | 252.20998 | 2.5221   |
| 2  | 0-4E0D  | 不     | 9733996  | 123.41931 | 3.75629  |
| 3  | 0-4EE5  | 以     | 9480560  | 120.20594 | 4.95835  |
| 4  | 0-4E5F  | 也     | 7881524  | 99.93144  | 5.95767  |
| 5  | 0-800C  | 而     | 7697614  | 97.59961  | 6.93366  |
| 6  | 0-5176  | 其     | 6896797  | 87.44589  | 7.80812  |
| 7  | 0-4EBA  | 人     | 6750929  | 85.59640  | 8.66409  |
| 8  | 0-70BA  | 為     | 6615021  | 83.87319  | 9.50282  |
| 9  | 0-6709  | 有     | 6209740  | 78.73455  | 10.29016 |

☒ No
 ☐ Unicode
 ☐ 字次
 ☐ 覆盖率%%
 ☐ 累加覆盖率%
  1
  10

书同文古籍字频查询系统 国家语委十五科技攻关重点项目

UniHan 北京书同文数字化技术有限公司

选择数据库 四部四库

☐ 单字查询 
☐ 多字查询

☒ 关联字查询

| No    | Unicode | Hanzi | 字次    | 覆盖率%%   |
|-------|---------|-------|-------|---------|
| 29924 | 0-94C1  | 铁     | 1     | 0.00001 |
| 2888  | 0-9421  | 鐵     | 22351 | 0.28339 |
| 1475  | 0-9435  | 鐵     | 80315 | 1.01833 |
| 6101  | 0-9244  | 鉄     | 3098  | 0.03928 |
| 7014  | 0-9295  | 鐵     | 1965  | 0.02491 |
| 9065  | 0-496B  | 鐵     | 691   | 0.00876 |