```
isChinese: function(_str)
{
    return /^[\u4E00-\u9FA5]{0,25}$/.test(_str);
},
```

救救汉字数字化的新难民

西东
Keep it simple, stupid.

对不起我只认GBK为汉字

如果你有试过去一些网站注册，姓名错的问题或已见识过。
试举两例，看图看代码。
GBK-1995就是U+4E00~U+9FA5, 对齐Unicode 1.01(1993),名字中的[奘]不是汉字。

网站注册姓名报错信息和代码

中国人事考试网的认字范围，与上述网站一样，只认U+4E00~9FA5为汉字：

```
isChinese: function(_str)
{
    return /^[\u4E00-\u9FA5]{0,25}$/.test(_str);
},
```


汉字输入框的检查代码

汉字究竟有多少?

其实GBK只有21000字，虽然比GB2312-80的6763字多很多，只占最新Unicode含扩展G的汉字93000字的22%；与《**康熙字典**》的47000字[1]相比，也只是一半不到。

U+4E00~9FA5(或应更新为9FFC)：GBK ≈小学
+ U+3400~4DBF：GB18030-2000国家强制标准，扩展A，≈公民义务教育，中学
+ U+20000~2A6D6：GB18030-2005，扩展B部分为非强制标准，≈大学
+ U+2A700~2EBE0：GB18030-202x (待发) ，扩展C~F，≈研究生
+ U+E000~F8FF： PUA RKXX用字，≈野鸡大学
通用规范汉字表：无码，≈毕升活字印刷技术，中华民族的骄傲

粪字其实是……

GBK-1995里编码为GB+FE9F，因为当时Unicode还未收录，在PUA区里编码为U+E863,到了GB18030-2000发布时，已正式收录为U+4DAE。

知乎有篇文章提到**PUA的正确姿势**：“補充字區的(PUA)訂定只是權宜之計，讓尚無 Unicode 官方標準之前可以臨時使用。若日後被官方正式收錄，那該從補充字应被剔除，移入正式字區。空下的碼位被回收利用，再定義為其他新的補充字。因此每次新Unicode官方標準發布後，都應該立即檢查一次補充字，**將已正式編碼的字進行遷碼處理**。”

严格说来，二代身份证发行在2005年之后，不应该会有U+E863，但在2020年的现实是：U+4DAE(大粪)和U+E863(小粪)几乎平分天下，大部分系统不认识大粪，这也是造成粪字人和服务从业人员，倍受其困、不知所措的根本原因。

身份证PUA用了多少字?

公安人口信息字库RKXX PUA(编码U+E000~U+F8FF)约4700多字，至2020年3月已有3600字纳入Unicode扩展A~G正式编码，剩余字大多为类推简化字和错别字。对于GBK系统来讲，其中只有1500字能映射到双字节，其它字只能用[?]对应了,这是为什么那些GBK系统实名验证不通过的原因。其中这个PUA字有部分字用于地名，可参见《汉字编码字符集*第八辅助*集》(八辅)相关文章。

已有正式码的字一般通用环境都可以显示和打印，建议尽快去更新换证了。

至于类推简化、错别字，离数字化时代太远了，还是趁早规范用字、改名，即便是用《康熙字典》里的繁体字，也好过PUA字，因为《康熙字典》所有字2015年已全被Unicode收录。[1]

通用汉字规范表是什么?

因为Unicode本身及RKXX中的用字不规范，促成了教育部2013年《通用汉字规范表》，多方博弈的结果，最终只收录了8105字，比GBK 21000字少得多，比GB18030-2000强制标准27000字更少。但是有276字在基本集外，更有199字收录在扩展B~E及基本集的急用加字区(URO+)，因为当时Unicode扩展E等还未发布，所以2013年发布的《通用汉字规范表》原文只有图片版。[2]

身份证名字没法输入吗?

因为扩展B~及PUA的汉字并非强制标准，尤其是PUA编码的身份证字，并不是公开的信息，所以这些生僻字人在电脑、手机时代，成为数字化新难民，一点也不奇怪。

PUA字在通用环境中字库里无字形时，已有不少IT圈内人，用熟知的方法---输入内码解决。其实退回一步讲，身份证发行机构可以学学香港身份证、一些中国护照申请签证时的做法---把名字的电报码放在证件上，我们把Unicode印在证件上，服务人员可以在没有机读证件条件时，用内码输入汉字，生僻字一了百了。

样证：樂永晴，电报码：2867 3057 2532

最后，以两段碎语结束本篇，希望读者有所醒悟。

A:为人民服务无需强制。
B:让人民跑腿必须强制。

三驾码车拉一字，三叉路口来相见，
各不相让争先行，困在其中只等闲。
注：
交给教育部没有码---通规表
交给公安部随便码---RKXX PUA
交给工信部我不码---只有GB18030
回到老百姓地上爬---靠腿靠嘴讨生活

[1]哪本字典收字最多？《康熙字典》可重登宝座，2015
[2] 国务院关于公布《通用规范汉字表》的通知，2013

编辑于 2020-09-09 16:48

Unicode（统一码） 生僻字 实名制

写下你的评论...

7 条评论 默认 最新

西东 作者
有奘字人在国务院App注册成功，有需要的不妨试试，PC端目前仍是限\u4E00-9FA5.
2020-10-16 回复 2

西东 作者 ▸ **一家来写两三本书**
如果你能据理力争去改名，身份证用龑，平时写奘，应该没问题。另一方法换证用U+4DAE的标准奘，哪儿有问题投诉哪儿，道路可能比较漫长，需全国八万奘字人共同出力。
2020-10-24 回复 2

堂吉诃德
原来有八万这么多啊🤣🤣🤣
2021-05-15 回复 1

查看全部 6 条回复 >

推荐阅读

1978年起为安置陆续来中国的20多万印支难民，中国出台…

2000年在中国外交部和联合国难民署驻华代表处，联合举行纪念联合国难民署成立50周年的招待会上（联合国难民署成立于1950年12 月14日），联合国难民署驻华代表麦希伟称赞中国政府在安置29万…

天天天蓝

中国不配接受取经的难民

最近白罗斯跟波兰的冲突愈演愈烈，多达一万五的难民在白罗斯一边排队等待着通过波兰来进入欧洲。一开始波兰也是自作孽不可活，来者不拒的放难民进来，然后转手就给身后的德国送过去，你…

灾难主教

与“难民站在一国到底应不应该

18183游戏网