Atitit 汉字编码大总结attilax总结 艾提拉总结

Atitit 汉字编码 gbk Unicode big5 sjis Shift\_JIS EUC-JP, JIS, Shift-JIS转码方法[转载]

Gbk比utf8可以减少50%占用，压缩后依然这么影响存在。。Gbk用了俩个字节16bit

规划企业内部汉字编码标准 uke，针对3500常用汉字，原则上需要12bit就够

垂直领域汉字编码。医疗行业汉字编码，医辽行业垂直领域使用汉字应该在一千左右，通过编码10bit可以满足，对比gbk的16bit，可以减少到gbk的60%

参考资料

Atitit 3500个常用汉字 attilax整理 v2 qaa.docx

1000个高频汉字.doc