

金庸小说全集（16本）信息熵表

#	分词模型	语料字数	分词个数	平均词长	信息熵（比特/词）
1	unigram	7282998	4293450	1.696304	12.172905
2	bigram	7282998	4293449	1.78891	6.962872
3	trigram	7282998	4293448	3.12219	2.302969

金庸小说全集（16本）信息熵表

#	小说名	语料字数	一元分词个数	一元平均词长	一元模型信息熵（比特/词）	二元分词个数	二元平均词长	二元模型信息熵（比特/词）	三元分词个数	三元平均词长	三元模型信息熵（比特/词）
1	三十三剑客图	53399	31290	1.706584	11.681066	31289	3.116431	2.94602	31288	3.651336	0.272666
2	书剑恩仇录	435906	253493	1.719598	11.719617	253492	2.535216	5.02564	253491	3.505679	1.038883
3	侠客行	309887	183372	1.689936	11.188972	183371	2.392521	4.955795	183370	3.423526	1.12706
4	倚天屠龙记	818422	474982	1.723059	11.760183	474981	2.334135	5.518218	474980	3.423599	1.328639
5	天龙八部	1021603	604558	1.689835	11.736275	604557	2.212131	5.677541	604556	3.336879	1.480923
6	射雕英雄传	767899	451854	1.699441	11.834495	451853	2.371189	5.469081	451852	3.438073	1.266258
7	白马啸西风	57437	35157	1.633729	10.242509	35156	2.542155	3.991294	35155	3.458854	0.741034
8	碧血剑	416533	242879	1.714982	11.747485	242878	2.566857	4.996851	242877	3.519263	0.993505
9	神雕侠侣	827823	498750	1.659795	11.561237	498749	2.223978	5.671443	498748	3.3594	1.434044
10	笑傲江湖	824948	482596	1.709397	11.398787	482595	2.164898	5.618827	482594	3.327497	1.527482
11	越女剑	13663	8063	1.694531	10.072456	8062	2.99876	2.529249	8061	3.617293	0.328286
12	连城诀	194542	117258	1.659094	11.043223	117257	2.46787	4.691516	117256	3.457981	0.955575
13	雪山飞狐	113428	67310	1.685158	11.110252	67309	2.680013	4.110558	67308	3.501634	0.693631
14	飞狐外传	375225	220938	1.698327	11.528414	220937	2.48216	4.994856	220936	3.473576	1.053167
15	鸳鸯刀	29290	17173	1.705584	10.478126	17172	2.925751	3.085121	17171	3.599324	0.432056
16	鹿鼎记	1022993	603777	1.694323	11.447505	603776	2.098619	5.752031	603775	3.281991	1.619806