

1. 有如下词典，每个词的费用 $C = -\log P$, P 为该词在语料库中的使用频率。对 “为人民工作” 进行分词。
- (1) 分别给出正向和逆向最大匹配法的分词结果；
 - (2) 画出词图，用最短路径法分词；
 - (3) 用最大概率法分词，给出详细计算过程。

词	费用
为人	4.2
人民	2.8
民工	3.2
工作	2.5
为	3.6
人	3.4
民	4.5
工	4.0
作	4.8

2. 一个分词系统分词得到如下结果：
- 据/ 中央/ 气象台/ 预报/, 未来/三天/, 四川/盆/地/等/地/将有/一次/降/雨天气/过程/。
- 假设正确分词结果如下：
- 据/ 中央气象台/ 预报/, 未来/三/天/, 四川盆地/等/地/将有/一/次/降雨/天气/过程/。
- 计算此次分词的准确率、召回率及 F-Measure 值（标点符号不计入词数）。

3. 用 HMM 和 Viterbi 算法进行词性标注

人民 收入 和 生活 水平 进一步 提高

n n c/p/v n/v n/a n/d n/v

词性转移表

Tag	n	c	p	v	a	d
n	80000	10000	12000	80000	5000	10000
c	30000	1000	2000	20000	10000	5000
p	40000	1000	500	5000	5000	10000
v	50000	5000	5000	20000	4000	10000
a	30000	8000	4000	7000	1000	20000
d	20000	10000	6000	30000	5000	9000

词语频度表

词语	词性	频次		词语	词性	频次
人民	n	5000		水平	n	4000
收入	n	4000		水平	a	1000

和	c	2000		进一步	n	1000
和	p	1000		进一步	d	2000
和	v	200		提高	n	1000
生活	n	5000		提高	v	4000
生活	v	2000				

词性频度表

词性	频次
n	200000
c	100000
p	100000
v	200000
a	100000
d	100000