# Tagging

## 方法1 concur

输入：请求打标签的文本 text

输出：n个标签词及其权重

缩略词：

* k = keyword; kw = keyword weight; 由jieba找出的关键词和权重
* c = concur word; cw = concur word weight; 由ItemFreq找出的共现词及其权重
* t = tag word; tw = tag weight 最终得到的 tag word 和计算出来的对应权重，

给定要打标签的文本text和标签集合TagSet

* 用jieba找出前n个keyword和他们的权值，
* 将kw归一化
* 对每一个在共现矩阵中找到其对应记录：
* for every
  + if
    - if then
    - else
* 得到矩阵TagMat：
* 在TagMat中依据相同词合并同类项 ；其中n是同类项的数目，得到：
* (用yakmo)找出输入文本text的cluster id，记作： （需要先用article2vector将text转换为向量）
* For every (上一步找到的tag集合)
  + Let tagword在text所属cluster的概率。
  + Let THRESHOLD是事先定义的阈值如0.5

## 方法2：knn

替换 cwij 为 （2-knn\_weight）/ 2