Jieba库概述

Jieba库是优秀的中文分词第三方库

* 中文文本需要通过分词获得单个的词语
* Jieba是优秀的中文分词第三方库，需要额外安装
* Jieba库提供三种分词模式，最简单只需要掌握一个函数

Jieba分词依靠中文词库

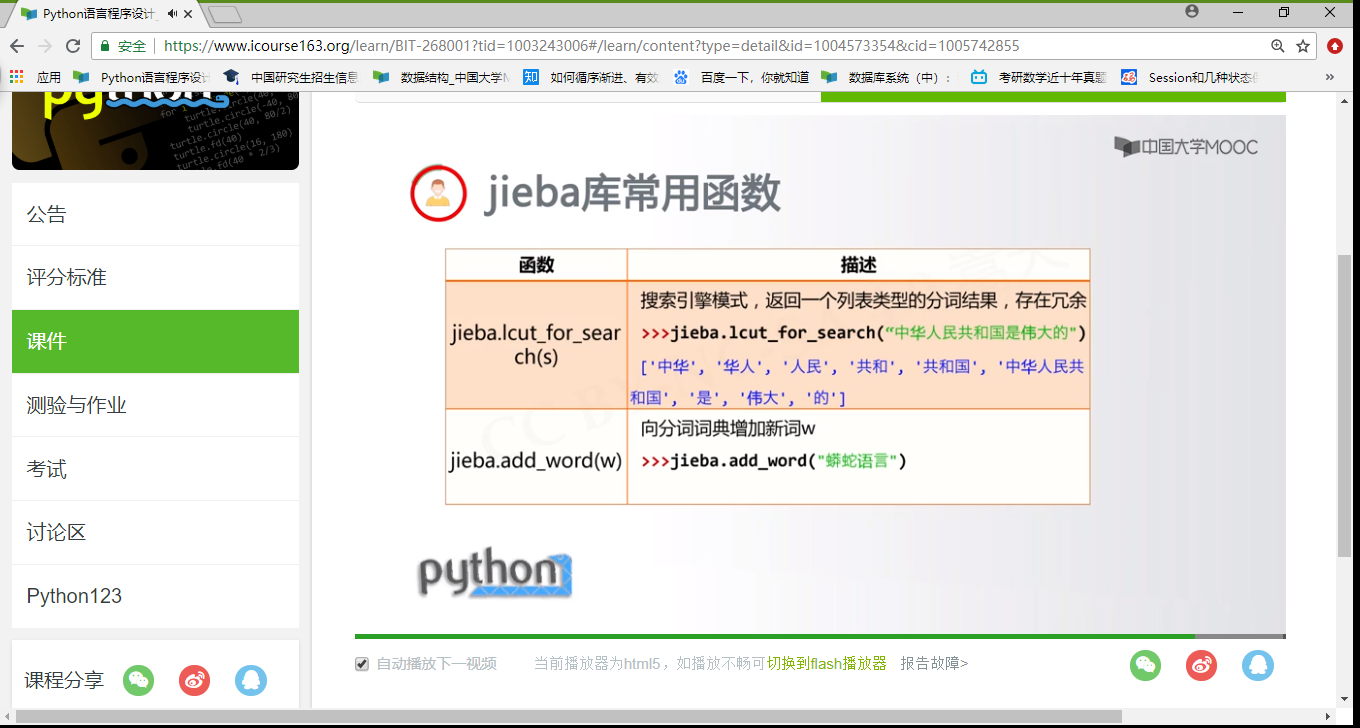
* 利用一个中文词库，确定汉字之间的关联概率
* 汉字间概率大的组成词语，形成分词结果
* 除了分词，用户还可以添加自定义的词语

Jieba库分词的三种模式

精确模式，全模式，搜索引擎模式

* 精确模式：把中文精确的切分开，不存在冗余单词
* 全模式：把文本中所有可能的词语都扫描出来，有冗余
* 搜索引擎模式：在精确模式的基础上，对长词再次切分





分词要点：

Jieba.lcut(s)