

READ ME

为了确保所有脚本能正常运行，请确保一下模组已被安装：

1. *SnowNLP*
2. *baidu - aip* (通过 `pip install baidu-aip` 安装)
3. *python - OPenCV*
4. *pyluence*
5. *numpy*

代码一共分为六个部分，分别被存储在一下六个文件夹中：

1. *lshash* 用来对网页上获得的图片进行局部哈希特征提取以及哈希降维
2. *resnet* 用来训练logo搜索的分类器
3. *resnet*训练集数据增强 用来对现有获得的logo图片进行旋转变色等等操作获得一个更大更全的数据集
4. 爬虫数据整合与建立索引 用来整合不同网站不同格式的数据并建索引
5. 爬虫与数据处理 用来存储各种爬虫和进行评论情感分析
6. 索引与网页 存储这建立完全的索引和网页源码，启动搜索网页需要在 *Python2* 环境下运行该文件夹下的 *code.py*

同时部分上传的脚本为数据处理整合所使用的脚本，由于从网页上获得的原数据未一起上传且很多中间处理结果没有保存，所以这些脚本无法正常运行，请不要尝试运行他们，包括存储在如下文件夹中的脚本：

1. *lshash*
2. *resnet*
3. *resnet*训练集数据增强
4. 爬虫数据整合与建立索引
5. 爬虫与数据处理

网页源码使用的语言为python(web.py)+html+javascript+css；

打开网页时请使用主流（Chrome）浏览器，否则部分功能不兼容；

关于以图搜图和logo搜索的部分请不要尝试在没有上传文件的情况下点击搜索，否则否则会带来错误结果和程序异常。