外部数据说明

### host\_feature.csv:

数据来源：

通过向网站发出请求，同时对于网址进行whois解析，得到对应与主机相关的特征。

数据维度：

|  |  |  |
| --- | --- | --- |
| 数据项 | 数据名称 | 备注 |
| Ip | ip地址 |  |
| Title | 标题 |  |
| Domain\_name | 域名 |  |
| Registrar | 登记 |  |
| Whois\_server | whois服务 |  |
| Updated\_date | 更新时间 |  |
| Creation\_date | 创建时间 |  |
| Expiration\_date | 到期时间 |  |
| Status | Status |  |
| Emails | 邮箱 |  |
| Dnssec | DNS解析 |  |
| Org | 组织 |  |
| State | 地点 |  |
| Country | 国家 |  |
| City | 城市 |  |
| zipcode | 邮编 |  |

### content\_feature.csv:

数据来源：

通过请求网站响应后，获取对应页面的HTML信息，在其中提取标题内容、关键文本等信息。

数据维度：

|  |  |  |
| --- | --- | --- |
| 数据项 | 数据名称 | 备注 |
| url | 网址 |  |
| Word | 关键文本 |  |

### shoot\_feature.txt:

数据来源：

获取网站的一些宏观信息，包括网站是否响应、网站世界排名、网站流量等信息。

数据维度：

|  |  |  |
| --- | --- | --- |
| 数据项 | 数据名称 | 备注 |
| Connection | 响应 |  |
| Rank | ALEXA排名 |  |
| Visit | 访问流量 |  |

### snapshot.zip:

数据来源：

获取网站的图片快照信息，通过python中的截图工具。

数据维度：

|  |  |  |
| --- | --- | --- |
| 数据项 | 数据名称 | 备注 |
| Picture | 图片 |  |