

第三课 网络数据的获取与表示

第三节 网页下载模块

主 讲: Robin

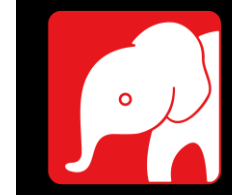
网页下载模块

- 将URL对应的网页下载到本地或读入内存(字符串)
- `requests`模块
- `requests.get(url)`返回`response`对象
- `response`对象
 - `status_codes`: 状态码
 - `headers`: 网页header信息
 - `encoding`: 网页编码
 - `text`: 请求返回的文本信息
 - `content`: 以字节形式返回的非文本信息



常见的HTTP状态码

- 200: 请求成功
- 404: 请求的资源不存在
- 500: 内部服务器错误



只为遇见明天更优秀的你！