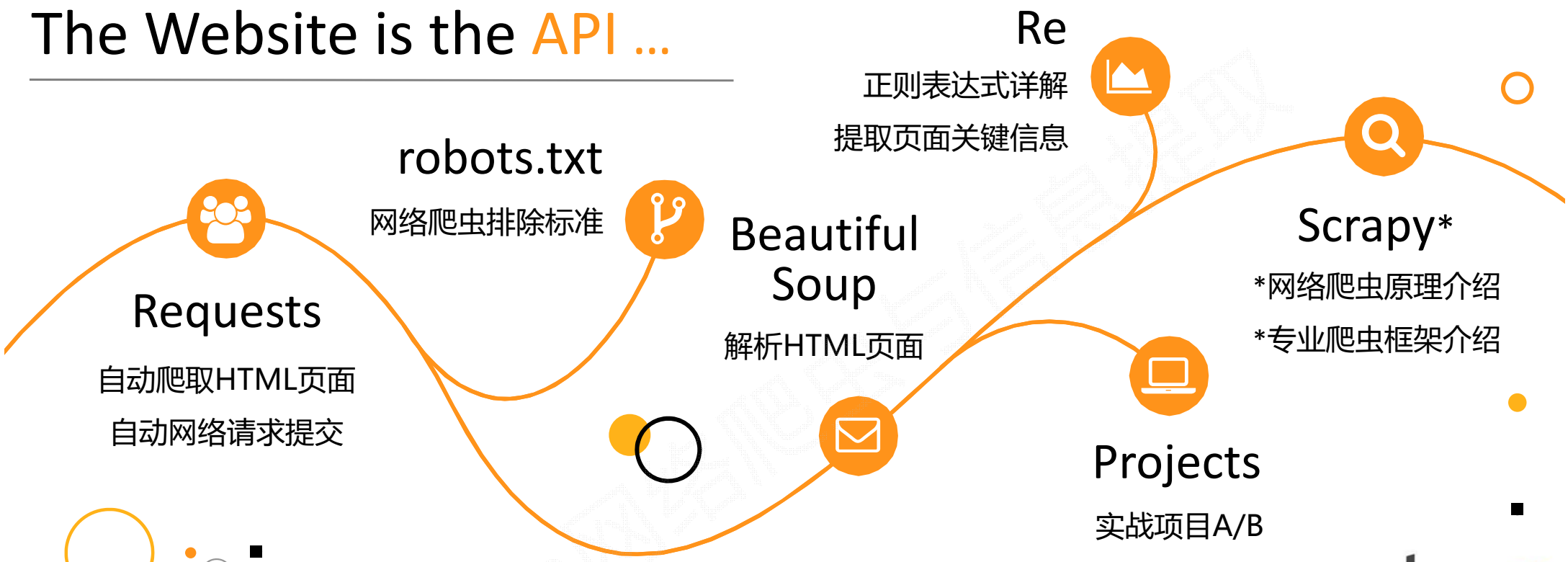
“网络爬虫”课程回顾和总结

WS00

嵩天

www.python123.org



掌握定向网络数据爬取和网页解析的基本能力

python
弹指之间 · 享受创新

Python网络爬虫与信息提取

O4X -Tian

技术路线

requests-bs4-re

scrapy（5+2结构）

本课程实例

- 京东商品页面的爬取
- 亚马逊商品页面的爬取
- 百度/360搜索关键字提交
- 网络图片的爬取和存储
- IP地址归属地的自动查询
- 中国大学排名定向爬虫
- 淘宝商品比价定向爬虫
- 股票数据定向爬虫
- 股票数据Scrapy爬虫

技术路线展望

requests-bs4-re

scrapy

+ PhantomJS

scrapy-*

表单提交、爬取周期、入库存储

https://pypi.python.org

致谢

- 北京理工大学
- 爱课程中国大学MOOC平台
- 拍摄和后期团队
- 助教：袁炜佳、李天龙

“君子曰：学不可以已。积土成山，风雨兴焉。”

——荀子《劝学》