

爬虫需求文档

需求概述

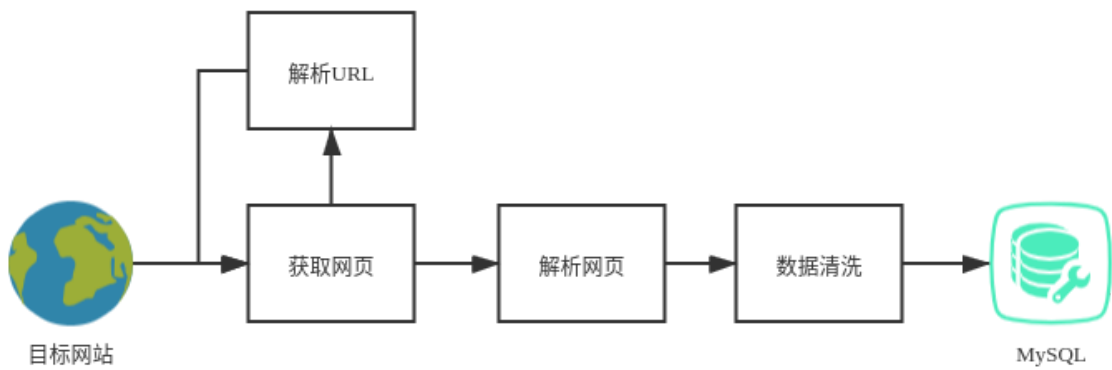
爬虫目标网站表

名称	网址	说明
N营养素需求推荐量	https://www.59baike.com/a/42787-33	这是《中国人民膳食指南》设定的标准，结合年龄、性别、体重就可以算出个人营养当日需求
F食物成分	https://fq.chinafcd.org/	食物成分详细说明每种食物包含的营养元素，比如碳水化合物、脂肪、蛋白质、维生素等
M菜谱	https://www.shipuxiu.com/zuofadaquan/861852/	菜谱有食物名称和食物用量

说明：

- 1. N是nutrition首字母，F是food首字母，M是menu首字母；后面数据库表数据源用此表示

数据流程



数据库表(所有重量单位为g)

- 云数据库账户

```
ip:122.51.38.230
port:3306
user:xiangli
passwd:123456
db_name:HealthyLife
```

- N：营养素推荐摄入量表(nutrition_energy)

age int	sex varchar(4)	PAL varchar(8)	energy float	portein float	fat_rate float

说明：

1. *各年龄组的能量的RNI(推荐摄入量)与其EAR(平均需要量)相同。
2. **为AI(适宜摄入量)，非母乳喂养应增加20%。
3. ▲PAL为体力活动水平。
4. 凡表中缺数字则表示未制定该参考值

- N：nutrition_minerals

age int	sex varchar(4)	PAL varchar(8)	Ca float	P float	K float	Na float
Mg float	Fe float	I float	Zn float	Se float	Cu float	F float
Cr float	Mn float	Mo float				

说明：

1. 正数表示AI（适宜量），负数表示RNIs（推荐摄入量）
2. PAL：正常、孕妇（早期、中期、晚期、乳母）
3. 凡表中缺数字则表示未制定该参考值

- N：nutrition_vitamin

age int	sex varchar(8)	PAL varchar(8)	A float	D float	E float	B1 float	B2 float
B6 float	B12 float	C float	pantothenic_acid float	folic_acid float	nicotinic_acid float	cholie float	biotin float

说明：

1. (1) *α-TE为α-生育酚当量。
2. (2) 凡表中缺数字则表示未制定该参考值
3. PAL：正常、孕妇（早期、中期、晚期、乳母）

- N：nutrition_microelement_ULs

age int	sex varchar(8)	PAL varchar(8)	Ca float	P float	Mg float	Fe float	I float
I float	Zn float	Se float	Cu float	F float	Cr float	Mn float	Mo float
A float	D float	B1 float	C float	folic_acid float	nicotinitc_acid float	choline float	

说明：

- 1. ▲60岁以上磷的UL(可耐受的最高摄入量)为3000mg。
- 2. 凡表中缺数字则表示未制定该参考值

- N：nutrition_EARs

age int	sex varchar(8)	PAL varchar(8)	portein float	Zn float	Se float	A float
D float	B1 float	B2 float	C float	folic_acid float		

- F：食物成分表（food_ingredients）

food_id int	name varchar(64)	Edible float	Water float	Energy float	Protein float	Fat float	Cholesterol float
Ash float	CHO float	Dietary_fiber float	Carotene float	Vitamin float	α- TEfloat	Thiamin float	Riboflavin float
Niacin float	Vitamin_C float	Ca float	P float	K float	Na float	Mg float	Fe float
Zn float	Se float	Cu float	Mn float	I float	SFA float	MUFA float	PUFA float
type							

说明：

- 1. Edible:食部，用小数标书百分比
- 2. Water：水分，用小数表示百分比
- 3. energy：能量单位为kj
- 4. 其它食物成分单位为g
- 5. type：表示成分类别，0 能量及相关成分，1 维生素，2 矿物质，3 脂肪酸

- M：菜谱表（cook_book）

menu_id int	name varchar(64)	picture varchar(256)	intrduction varchar(1024)	spectrum varchar(16)

说明：

- 1. picture：保存图片的文件路径
- 2. spectrum：菜谱谱系，比如川菜、杭帮菜、徽菜、粤菜、东北菜

- F+M：食材表(food_material)

menu_id int	food_id int	quantity float					

说明：

1. menu_id来自cook_book表的主键menu_id,food_id来自food表的主键food_id
2. 爬取菜谱后，先在cook_book表生成菜谱记录的menu_id,然后返回这个menu_id，根据解析的食材去food表查询food_id,重量来自菜谱解析

参考源码

源码地址

<https://github.com/XiangLiHealthy/XiangLiHealthyLife/tree/master/spider>

爬虫类与对象图

<https://www.processon.com/view/link/5e718b07e4b011fcce9c7683>

学习资料推荐

- 爬虫原理：<https://www.jianshu.com/p/2411490d6464>
 - requests----->html网页下载工具
 - beautifulsoup---->网页解析工具
 - pymysql----->数据保存到mysql的工具
- 简单示例：<https://github.com/voidking/baike-spider>
- 开发环境搭建：<https://www.mainblog.cn/199.html>
- 实战进阶：<https://zhuanlan.zhihu.com/p/21479334>