# 采集器部署文档

###### 环境搭建

1.软件版本

Node.js:8.x及以上

2.TypeScript配置

* 1. 使用npm命令安装TypeScript

npm install -g typescript

* 1. tsconfig.json

{

"compilerOptions": {

"target": "es6",

"allowJs": true,

"sourceMap": true,

"moduleResolution": "node",

"emitDecoratorMetadata": true,

"experimentalDecorators": true,

"removeComments": false,

"noImplicitAny": false

},

"lib": [

"es2015"

],

"include": [

"./\*\*/\*"

],

"exclude": [

"node\_modules"

]

}

3.npm依赖包

package.json

{

"name": "spider",

"version": "1.0.0",

"dependencies": {

"axios": "^0.19.0",

"chalk": "^2.3.2",

"puppeteer": "^1.11.0",

"request": "^2.88.0",

"socks5-http-client": "^1.0.4"

},

"devDependencies": {

"@types/fs-extra": "^5.0.1",

"@types/jest": "^22.1.1",

"@types/lru-cache": "^4.1.0",

"@types/reflect-metadata": "^0.1.0",

"jest": "^22.1.4",

"ts-jest": "^22.0.2",

"ts-loader": "^4.3.0",

"ts-node": "^7.0.1",

"tslint": "^5.9.1",

"tslint-config-standard": "^7.0.0",

"typescript": "^2.6.2"

}

}

1. 依赖包下载

1)定位到package.json所在文件夹下(此项目为spiders文件夹下)

2)执行npm install下载package.json文件下的依赖包

3)如若需要添加某个特定的依赖包执行 npm install 包名

###### 采集器启动

1. 生成采集器界面

**1)首次运行需要执行以下firstRun.bat脚本生成采集器界面快捷方式**



1. json文件配置

每个网站的采集脚本都对应一个json配置文件(scripts文件夹下),

例如 y8com对应的json文件如下

{

"category-page":[

    {"url":"https://zh.y8.com/categories/girls","name":"girls"},

    {"url":"https://zh.y8.com/categories/shooting","name":"shooting"},

    {"url":"https://zh.y8.com/categories/strategy","name":"strategy"}

],

"crawl-type":"browser"

}

url为采集脚本执行的初始网站地址,name为对应的游戏类型

可自行按需添加

1. **具体script命令行执行**
   1. 游戏json数据采集

/\*\*

\*node ts-node

\*tsPath 脚本文件路径

\*outfilenName 输出文件名

\*startPage 初始网站地址

\*headless 无头浏览器,默认true

\*/

const [node, tsPath, outfileName, startPage, headless=true, ...args] = process.argv

**在scripts文件夹下执行**

**ts-node up.zdhm.xyz.ts xyz.json http://up.zdhm.xyz**

* 1. swf文件下载

/\*\*

\*node ts-node

\*tsPath 脚本文件路径

\*outfilenName 下载文件名

\*startPage 初始网页地址

\*outDir 下载目录

\*headless 无头浏览器,默认true

\*/

const [node, tsPath, outfileName, startPage, outDir = "../data/4399.com/swf", headless = true, ...args] = process.argv;

**在scripts文件夹下执行**

**ts-node 4399download.ts 1.swf http://www.4399.com/flash/207016.htm**