一．Task（用于爬虫和检查端通信）：

message Task{

required string hash = 1;

required string rel\_path = 2;

required bytes task\_descriptor = 3;

required bytes target\_addition = 4;

}

|  |  |
| --- | --- |
| hash | 目标的hash值 |
| rel\_path | 相对于默认保存路径的相对路径 |
| **task\_descriptor** | 展示端下达的任务，序列化前是字典型 |
| **target\_addition** | 具体被检查项信息，序列化前是字典型 |

1. task\_descriptor(由展示端产生，用于下达命令)

**task\_descriptor:** {“type”: ”,

“kwds”: ””,

”siteurl”: ””,

“sitename”: ””,

“dirname”: “”,

“taskid”:””

“rank”:””

}

|  |  |
| --- | --- |
| type | 任务类型 (sitecheck、emailcheck、weibocheck、weixincheck) |
| kwds | 关键字 (以英文“,”分隔，如“机密,秘密”) |
| siteurl | 任务入口 (如网站主页，邮件所在的文件夹路径) |
| sitename | 任务名称 (如“百度”) |
| dirname | 任务名称的拼音 (如“baidu”)，用于创建该任务的临时文件夹 |
| taskid | 任务的唯一标识符 |
| rank | 设置爬取得页面层数 |

三．**target\_addition:** 分多种情况，分类信息由**task\_descriptor**的type提供

1. type == “sitecheck”(网页检查) 或 “weibocheck”(微博检查) 或 “weixincheck”(微信检查)

task\_content: {“rel\_path”: ””,

“hash”: ””

“title”:””,

“parent\_id”: ””,

“web\_url”: ””,

“file\_url”: ””,

}

|  |  |
| --- | --- |
| hash | 该被检查项的hash值 |
| rel\_path | 相对于默认保存路径的相对路径 |
| title | 网页标题 |
| parent\_id | 父网页的hash值 |
| web\_url | 所在网页的url |
| file\_url | 该被检查项的url |

2. type== “emailcheck”(邮件检查)

task\_content:{“hash”: ””,

“rel\_path”: ””,

“mail\_header”: ””}

|  |  |
| --- | --- |
| hash | 该被检查项的hash值 |
| rel\_path | 相对于默认保存路径的相对路径 |
| mail\_header | 邮件头部信息，字典型 |

**Command:（用于展示端发送结束命令）**

**命令形式：**{

“taskid”:”tasked\_xxx”,

“type”: “task\_type”

“command”:”command\_line”

}

**例如：**{“taskid”: ”123”, ”type”:”sitecheck”, “command”: “kill”}