# **Catcher**

*Class in package 'Class Model'*

页面捕捉器

Catcher

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **OUTGOING STRUCTURAL RELATIONSHIPS** |
| --- |
| Realization from Catcher to I\_Catcher  [ Direction is 'Source -> Destination'. ] |

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Catcher | Target: Public (Class) Cheerio |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Catcher | Target: Public (Class) Processor |

# **Cheerio**

*Class in package 'Class Model'*

Cheerio

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Catcher | Target: Public (Class) Cheerio |

# **CollectController**

*Class in package 'Class Model'*

采集控制器

CollectController

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/8/31 星期一. Last modified 2015/9/6 星期日

| **ATTRIBUTES** |
| --- |
| request : Private  [ Is static False. Containment is Not Specified. ] |

|  |
| --- |
| mongo : Private  [ Is static False. Containment is Not Specified. ] |

|  |
| --- |
| catchers : Private  [ Is static False. Containment is Not Specified. ] |
| urls : Private  存储所有访问过的页面  [ Is static False. Containment is Not Specified. ] |
| finishCount : Private  完成页面计数  [ Is static False. Containment is Not Specified. ] |

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) Source |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public Pages (Class) Page |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) Request |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) MongoDB |

| **OPERATIONS** |
| --- |
| addPage (op : var ) : Public  添加页面  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |

|  |
| --- |
| collect () : Public  开始触发采集  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |

|  |
| --- |
| load (source\_package\_path : var ) : Public  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |
| registerCatcherForPage (catcher\_name : var , page : var ) : Private  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |
| log (callback : Function ) : Private  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |
| request (page : Page ) : Private  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |
| hasUrl (url : var ) : Private  [ Is static False. Is abstract False. Is return array False. Is query False. Is synchronized False. ] |

# **MongoDB**

*Class in package 'Class Model'*

MongoDB数据库操作端，npm中的mongo模块。

MongoDB

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) MongoDB |

# **Page**

*Class in package 'Class Model'*

页面

Page

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **ATTRIBUTES** |
| --- |
| url : var Private  页面url  [ Is static False. Containment is Not Specified. ] |

|  |
| --- |
| headers : Object Private  页面头信息  [ Is static False. Containment is Not Specified. ] |

|  |
| --- |
| body : var Private  页面内容  [ Is static False. Containment is Not Specified. ] |
| data : var Private  页面捕获数据  [ Is static False. Containment is Not Specified. ] |
| catchers : I\_Catcher Private  页面使用的内容捕捉器  [ Is static False. Containment is Not Specified. ] |
| undoneCount : int Private  未完成捕捉数  [ Is static False. Containment is Not Specified. ] |
| request.retry : int Private  页面请求信息：请求失败重试次数  [ Is static False. Containment is Not Specified. ] |
| request.status : int Private  请求状态  [ Is static False. Containment is Not Specified. ] |
| request.statusCode : var Private  请求状态码  [ Is static False. Containment is Not Specified. ] |
| request.error : var Private  页面请求错误  [ Is static False. Containment is Not Specified. ] |

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Page | Target: Public (Interface) I\_Catcher |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public Pages (Class) Page |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Source | Target: Public entryPages (Class) Page |

# **Processor**

*Class in package 'Class Model'*

Processor

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **OUTGOING STRUCTURAL RELATIONSHIPS** |
| --- |
| Realization from Processor to I\_Processor  [ Direction is 'Source -> Destination'. ] |

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Catcher | Target: Public (Class) Processor |

# **Request**

*Class in package 'Class Model'*

页面请求模块，使用npm中的request模块。

Request

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/9/17 星期四. Last modified 2015/9/17 星期四

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) Request |

# **Source**

*Class in package 'Class Model'*

Source

Version 1.0 Phase 1.0 Proposed

Administrator created on 2015/8/31 星期一. Last modified 2015/9/6 星期日

| **ATTRIBUTES** |
| --- |
| entryPages : Page Private  入口页面  [ Is static False. Containment is Not Specified. ] |

|  |
| --- |
| catchers : I\_Catcher Private  页面捕获器  [ Is static False. Containment is Not Specified. ] |

| **ASSOCIATIONS** | |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Source | Target: Public (Interface) I\_Catcher |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) Source | Target: Public entryPages (Class) Page |

|  |  |
| --- | --- |
| Association (direction: Source -> Destination) | |
| Source: Public (Class) CollectController | Target: Public (Class) Source |