*“Optimizing the performance of computer systems has always been an art relegated to a few individuals who happen to have the ‘right skills’.”*

**Amir H. Majidimehr**  
Optimizing Unix for Performance, 1994



**Flow**

Протокол взаимодействия

Версия 1.0

# Оглавление

[Оглавление 2](#_Toc473827373)

[История документа 3](#_Toc473827374)

[Введение 4](#_Toc473827375)

[Описание проекта 4](#_Toc473827376)

[Обзор архитектуры 4](#_Toc473827377)

[Общий алгоритм взаимодействия 4](#_Toc473827378)

[Протокол сетевого взаимодействия 4](#_Toc473827379)

[Протокол 5](#_Toc473827380)

[Ядро → Граббер 5](#_Toc473827381)

[Граббер → Ядро 5](#_Toc473827382)

[JSON-объекты 6](#_Toc473827383)

[PingResponse 6](#_Toc473827384)

[ErrorInfo 6](#_Toc473827385)

[ExceptionInfo 7](#_Toc473827386)

[GrabRequest 7](#_Toc473827387)

[Period 8](#_Toc473827388)

[GrabResponse 9](#_Toc473827389)

[GrabResultsRequest 9](#_Toc473827390)

[GrabberInfo 10](#_Toc473827391)

[AuthorInfo 11](#_Toc473827392)

[SourceInfo 11](#_Toc473827393)

[GrabResults 12](#_Toc473827394)

[PostInfo 12](#_Toc473827395)

[ProcessingInfo 14](#_Toc473827396)

[GrabResultsResponse 14](#_Toc473827397)

[Приложение А. Пример JSON GrabRequest 17](#_Toc473827398)

[Приложение Б. Пример JSON GrabResults 18](#_Toc473827399)

# История документа

|  |  |  |  |
| --- | --- | --- | --- |
| **Автор** | **Дата** | **Версия** | **Комментарий** |
| Бахарев С.А. | 2.02.2017 | 0.1 alpha | Документ создан. Описаны базовые вещи, черновой вариант протокола. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Введение

## Описание проекта

Flow – это внутренний проект, предназначенный для сбора новостей с различных веб-ресурсов и рассылки этих новостей по подписке. Поиск новостей осуществляется по ключевым словам, и логически делится на «каналы подписки». Система будет предоставлять пользователям возможность выбирать свои каналы подписки.

В будущем подразумевается использовать Flow для создания тематических подписок, чтобы сотрудники компании были в курсе свежих тенденций в интересных им областях.

## Обзор архитектуры

Архитектурно Flow делится на две основные части: ядро и грабберы.

Грабберы – это сборщики информации, каждый из которых работает с одним конкретным веб-ресурсом. Примеры таких ресурсов – Twitter, Habrahabr, Facebook, VK, различные блоги и т.п. Грабберы реализуются как миниатюрные веб-сервисы на самых разных языках программирования.

Ядро является одновременно веб-приложением для менеджмента подписок, агрегатором информации, получаемой от грабберов и рассылает письма со сводкой новостей подписчикам.

## Общий алгоритм взаимодействия

Грабберы являются веб-сервисами, и пока они не получают команды от ядра, ничего не делают.

Периодически, согласно своему расписанию, ядро отправляет им запрос на осуществление поиска новостей по некоторым ключевым словам, указывая интервал дат, в который должны попадать найденные новости. Граббер отвечает серверу условно «запрос принят в работу».

Когда граббер завершает поиск новостей, он формирует сводку из полученной информации и отправляет её уже на веб-сервис ядра. При этом исходное сообщение (от ядра ко грабберу) и итоговое (от граббера к ядру) связаны через уникальный идентификатор, который ядро присваивает каждому своему запросу к каждому грабберу.

## Протокол сетевого взаимодействия

Всё общение между ядром и грабберами осуществляется по протоколу HTTP, все сообщения – в формате JSON.

# Протокол

## Ядро → Граббер

### Проверка работоспособности: /ping

Запрос: GET, JSON-объект отсутствует

Ответ: JSON-объект PingResponse

### Старт сбора новостей: /grab

Запрос: POST, JSON-объект GrabRequest

Ответ: JSON-объект GrabResponse

## Граббер → Ядро

### Результаты поиска: /results

Запрос: POST, JSON-объект GrabResultsRequest

Ответ: JSON-объект GrabResultsResponse

# JSON-объекты

## PingResponse

|  |  |
| --- | --- |
| result | |
| Тип | Boolean |
| Обязательное | Да |
| Пояснение | Если граббер готов к работе, должно иметь значение true. |
| error | |
| Тип | ErrorInfo |
| Обязательное | Нет |
| Пояснение | Если граббер не готов к работе (значение result = false), описывает проблему. |

## ErrorInfo

|  |  |
| --- | --- |
| text | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Русскоязычное понятное осмысленное описание ошибки. |
| time | |
| Тип | String, дата и время в стандартном формате JS |
| Обязательное | Нет |
| Пояснение | Дата и время регистрации (отлова) ошибки. |
| exception | |
| Тип | ExceptionInfo |
| Обязательное | Нет |
| Пояснение | Информация об исключении, вызвавшем эту ошибку. |

## ExceptionInfo

|  |  |
| --- | --- |
| className | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Класс исключения. |
| message | |
| Тип | String |
| Обязательное | Нет |
| Пояснение | Сообщение исключения. |
| stackTrace | |
| Тип | Массив String |
| Обязательное | Нет |
| Пояснение | Стек вызовов исключения. |
| inner | |
| Тип | ExceptionInfo |
| Обязательное | Нет |
| Пояснение | Аналогичная информация о внутреннем исключении. |

## GrabRequest

|  |  |
| --- | --- |
| requestId | |
| Тип | Number |
| Обязательное | Да |
| Пояснение | Уникальный идентификатор запроса, который используется как связь между запросом от ядра ко грабберу и итоговым запросом от граббера к ядру с результатами поиска новостей. |
| version | |
| Тип | Number |
| Обязательное | Да |
| Пояснение | Версия протокола. |
| keywords | |
| Тип | Массив String |
| Обязательное | Да |
| Пояснение | Массив поисковых запросов на целевой ресурс, по каждому из которых граббер должен отработать отдельно. |
| period | |
| Тип | Period |
| Обязательное | Да |
| Пояснение | Версия протокола. |

## Period

|  |  |
| --- | --- |
| fromDate | |
| Тип | String, дата в стандартном формате JS |
| Обязательное | Да |
| Пояснение | Дата начала периода. |
| toDate | |
| Тип | String, дата в стандартном формате JS |
| Обязательное | Да |
| Пояснение | Дата окончания периода. |

## GrabResponse

|  |  |
| --- | --- |
| result | |
| Тип | Boolean |
| Обязательное | Да |
| Пояснение | Если граббер взялся за работу, должно иметь значение true. |
| error | |
| Тип | ErrorInfo |
| Обязательное | Нет |
| Пояснение | Если граббер не смог начать работу (значение result = false), описывает проблему. |

## GrabResultsRequest

|  |  |
| --- | --- |
| requestId | |
| Тип | Number |
| Обязательное | Да |
| Пояснение | Идентификатор requestId, изначально присланный ядром грабберу. |
| version | |
| Тип | Number |
| Обязательное | Да |
| Пояснение | Версия протокола, в которой граббер составлял сообщение. |
| grabber | |
| Тип | GrabberInfo |
| Обязательное | Да |
| Пояснение | Описание граббера. |
| results | |
| Тип | Массив GrabResults |
| Обязательное | Да |
| Пояснение | Результаты поиска по указанным ядром ключевым словам. |

## GrabberInfo

|  |  |
| --- | --- |
| name | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Человеческое название граббера. |
| version | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Версия граббера в формате, принятом в рамках проекта. |
| authors | |
| Тип | Массив AuthorInfo |
| Обязательное | Да |
| Пояснение | Перечисление авторов граббера. |
| source | |
| Тип | SourceInfo |
| Обязательное | Да |
| Пояснение | Описание источника новостей, с которым работает граббер. |

## AuthorInfo

|  |  |
| --- | --- |
| name | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Фамилия и имя автора на русском языке. |
| email | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Адрес электронной почты автора в домене @pflb.ru. |

## SourceInfo

|  |  |
| --- | --- |
| title | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Краткое человеческое наименование ресурса. Если есть возможность, то русскоязычная версия. |
| url | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Адрес главной страницы веб-сайта источника новостей. |
| logo | |
| Тип | String, бинарное значение, закодированное в Base64 |
| Обязательное | Нет |
| Пояснение | Логотип источника новостей в формате PNG или JPG, закодированный в base64, с исходным размером (до base64) не более 100 Кб. |

## GrabResults

|  |  |
| --- | --- |
| query | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Поисковый запрос, соответствующий указанным результатам. |
| posts | |
| Тип | Массив PostInfo |
| Обязательное | Да |
| Пояснение | Описание источника новостей, с которым работает граббер. |
| errors | |
| Тип | Массив ErrorInfo |
| Обязательное | Нет |
| Пояснение | Описания ошибок, произошедших в процессе обработки этого поискового запроса. |
| processing | |
| Тип | ProcessingInfo |
| Обязательное | Да |
| Пояснение | Метрики и другие характеристики процесса поиска. |

## PostInfo

|  |  |
| --- | --- |
| url | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Адрес поста на ресурсе-источнике. |
| original | |
| Тип | String |
| Обязательное | Нет |
| Пояснение | Если пост на ресурсе является отсылкой к другой странице (возможно, даже не другом сайте), то тут указывается конечная ссылка на источник. |
| language | |
| Тип | String |
| Обязательное | Да |
| Пояснение | Язык поста: “ru” или “en”. |
| title | |
| Тип | String |
| Обязательное | Да, если нет аннотации (annotation) |
| Пояснение | Заголовок поста, должен быть как можно короче. |
| annotation | |
| Тип | String |
| Обязательное | Да, если нет заголовка (title) |
| Пояснение | Аннотация к посту. Например, первый его абзац. Желательно не более 1000 символов, всё равно на ядре при составлении письма придётся обрезать ради экономии места. |
| popularity | |
| Тип | Number |
| Обязательное | Нет |
| Пояснение | Условное значение популярности поста. В первую очередь важна популярность относительно остальных постов с этого же ресурса, поэтому формула, по которой считается популярность, лежит на граббере. |
| image | |
| Тип | String |
| Обязательное | Нет |
| Пояснение | Картинка, прикреплённая к посту, в формате PNG или JPG, закодированная в base64, с исходным размером (до base64) не более 200 Кб. |

## ProcessingInfo

|  |  |
| --- | --- |
| startTime | |
| Тип | String, дата и время в стандартном формате JS |
| Обязательное | Да |
| Пояснение | Время начала обработки конкретного поискового запроса. |
| finishTime | |
| Тип | String, дата и время в стандартном формате JS |
| Обязательное | Да |
| Пояснение | Время завершения обработки конкретного поискового запроса. |

## GrabResultsResponse

|  |  |
| --- | --- |
| result | |
| Тип | Boolean |
| Обязательное | Да |
| Пояснение | Если результаты успешно прошли валидацию, должно иметь значение true. |
| error | |
| Тип | ErrorInfo |
| Обязательное | Нет |
| Пояснение | Если результаты не прошли валидацию, описание обнаруженных проблем. |

# Приложение А. Пример JSON GrabRequest

{

requestId: 13,

version: 1,

queries: [

"selenium",

"#testautomation"

],

period: {

fromDate: "2017-01-01",

toDate: "2017-01-31"

}

}

# Приложение Б. Пример JSON GrabResults

{

requestId: 13,

version: 1,

grabber: {

name: "PUPER-граббер для Twitter",

version: "0.1.1",

authors: [

{

name: "Эрнст Константин",

email: "k.ernst@pflb.ru"

}

],

source: {

title: "Твиттер",

url: "https://twitter.com",

logo: "..."

}

},

results: [

{

query: "selenium",

posts: [

{

url: "https://twitter.com/testrepository/status/827134899127717892",

original: "https://testingrepository.com/capybara-get-attribute/",

language: "en",

title: "How to get element's attribute with #ruby #capybara",

popularity: 2,

image: "..."

},

{

url: "https://twitter.com/testrepository/status/827134899127717892",

original: "https://testingrepository.com/capybara-get-attribute/",

language: "en",

title: "Selenium 3: Обсуждаем внедрение, Делимся опытом",

popularity: 7

}

],

errors: [

{

text: "Не удалось прикрепить изображение к посту https://twitter.com/testrepository/status/827134899127717892",

time: "2017-01-30T03:24:00",

exception: {

className: "ru.pflb.flow.exceptions.InvalidImageFormatException",

message: "недопустимый формат изображения: bmp",

stackTrace: [

"ImageConverter.java: Line 58",

"ImageAttach.java: Line 201",

"..."

]

}

}

],

processing: {

start: "2017-01-30T03:22:09",

finish: "2017-01-30T03:29:11"

}

},

{

query: "#testautomation",

posts: [],

errors: [],

processing: {

start: "2017-01-30T03:22:12",

finish: "2017-01-30T03:22:19"

}

}

]

}