

# SOUPİ — ПРОЕКТ ВЕБ-ИНСТРУМЕНТА ДЛЯ ПОТОКОВОЙ ОБРАБОТКИ АСТРОНОМИЧЕСКИХ ИЗОБРАЖЕНИЙ

В.Н. Черненко

## **Аннотация.**

Проект Soupi является новой ветвью в развитии практически "замороженного" проекта (Youri) "You Pipe" -- персональный конвейер обработки наблюдений (Monnerville M. at all, 2010).

Это гибкий, ориентированный на астронома веб-интерфейс приложений, обеспечивающий высокий уровень функциональности выполнения обработки научных данных из исходных FITS-изображений. Он построен на основе открытых программных инструментариях, в том числе разработанных сообществом TERAPIX Парижских Института Астрофизики и Университета Марии и Пьера Кюри с целью организации процесса обработки на компьютерах кластера. Интерфейс позволяет управлять заданиями обработки в режиме реального времени и облегчает процесс взаимодействия между браузером пользователя и большим набором данных. На стороне сервера работают скрипты на языке программирования Python, вызываемые веб-фреймворком Django. На стороне клиента использована Ajax - технология на основе библиотек Javascript Prototype и script.aculo.us.

Основными отличительными элементами нового проекта является переработка под использование полноценной СУБД PostgreSQL с расширением PostGIS, что позволит повысить масштабируемость процесса обработки данных.

Встраиваемые модули и описания наблюдательных инструментов включают новые настройки, в том числе на отечественный прибор Scorpio-2.

## **1. Введение.**

Soupi -- это современный универсальный "frontend" для интернет-грида, который подходит и для обработки больших архивных "залежей" наблюдательного материала, таких как банк данных CAO РАН, так и небольшого количества персональных наблюдений с разрозненным набором FITS-изображений.

Необходимость создания такого инструментария вызвано в первую очередь накоплением большого количества неупорядоченного наблюдательного материала в CAO, требующего астрометрических калибровок и отождествлений объектов в автоматизированном режиме.

Soupi предназначен для запуска и контроля выполнения заданиями обработки данных, отслеживая все этапы и параметры в процессе жизненного цикла обработки.

Различные варианты реализации графического интерфейса для создания персонального конвейера в грид можно найти например на сайте проекта Taverna, развиваемого Манчестерским университетом в Англии.

Существенное отличие проекта в том, что он не ориентирован только на грид-сервисы.

Астрономам часто даже весьма нежелательно, чтобы их наблюдательный материал куда-то передавался по Интернет с использованием сторонних вычислительных ресурсов, а подвергался обработке не покидая пределы организации.

Процесс взаимодействия основных компонентов Soupi изображен на рисунке 1.

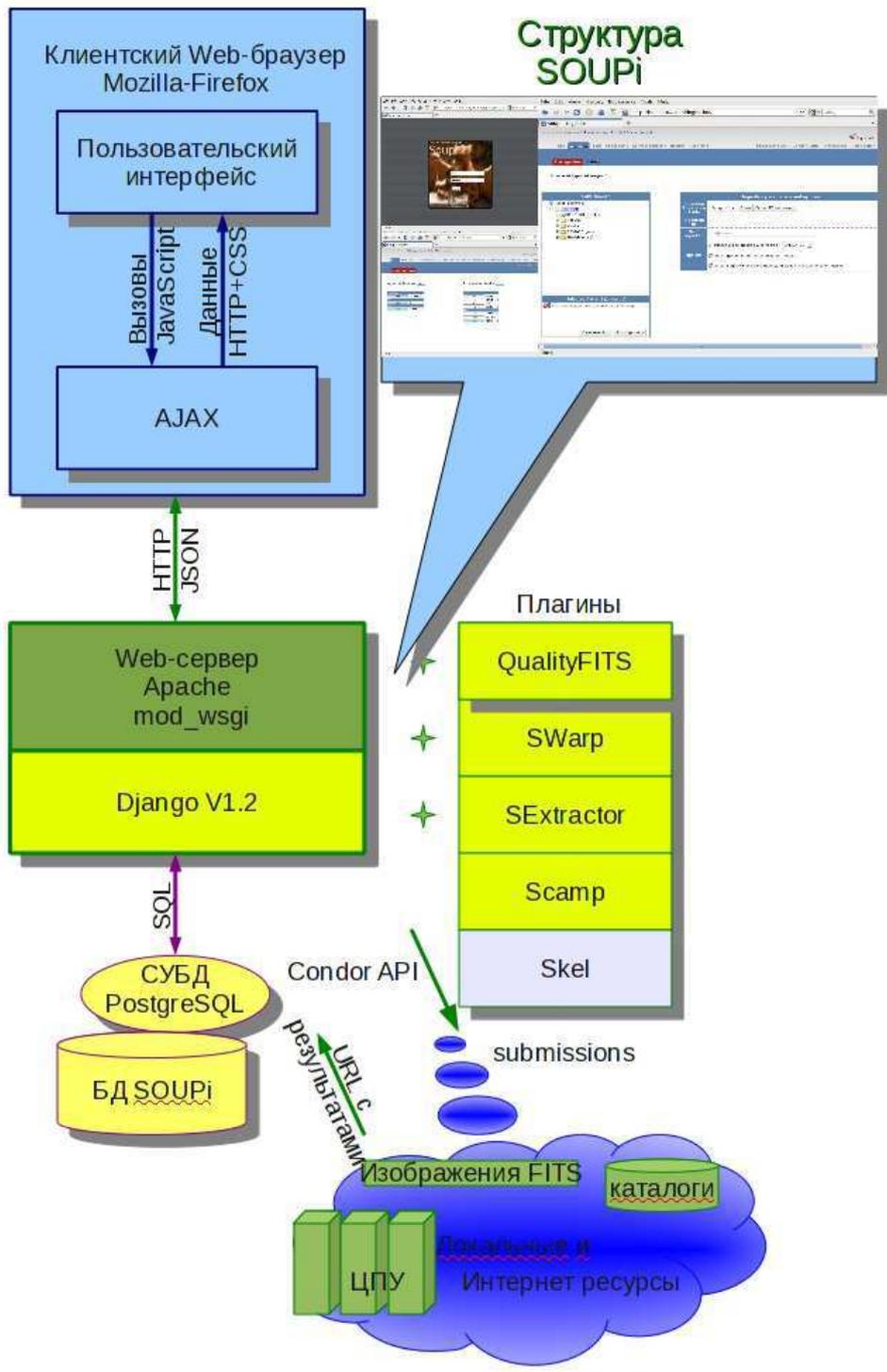


Рис. 1

Цикл начинается с загрузки (ingestion) FITS-файлов с информацией об изображениях в базу данных и управлением обработкой вплоть до конечных результатов (откалиброванных изображений, каталогов).

## **2. Элемент загрузка данных -- "Ingestion".**

Загрузка изображения -- Ingestion, является основной точкой входа для пользователя Soupi.

Благодаря виджетам с функцией select пользователь выбирает один или несколько каналов (фильтров) данных FITS-изображений, доступных как локально, так и из сетевых ресурсов, таких как NFS, FTP. При этом Soupi считывает текстовые заголовки и заносит информацию в базу данных.

Сообщает при этом пользователю, в том числе и по E-mail об окончании загрузки.

Загруженные данные могут быть доступны пользователю для обработки в любое время посредством удобных инструментов выборки изображений (см. далее).

## **3. Элемент обработка данных -- "Processing".**

Soupi является инструментом, который действует как высокоуровневая оболочка для свободных низкоуровневых программных пакетов обработки астрономических изображений.

К моменту публикации запланирована поддержка программных пакетов: QualityFITS -- оценка качества изображения, SCAMP -- астрометрическая и фотометрическая калибровка инструмента (Bertin 2006), SWarp -- инструмент для интерполяции и сложения изображений, а также SExtractor -- инструмент для извлечения точечных источников и каталогизации (Bertin et al. 1996).

Все модули обработки доступны изнутри в качестве плагинов. Некоторые из них могут быть отключены для какой-либо конкретной или фрагментированной последовательности обработки по желанию пользователя. Модульная архитектура является гибкой и универсальной.

При необходимости можно добавлять собственные плагины и расширить возможности обработки, для чего имеется скелетный шаблон с описанием API.

Soupi управляет доступом данных и защитой прав путем реализации Unix-тегов, аналогичных атрибутам файловой системы. Каждый пользователь имеет персональный, защищенный паролем профиль, где наборы данных и результаты могут иметь пользовательские разрешения для владельца, группы и прочих. Разрешения могут быть изменены, чтобы делиться изображениями, результатами обработки и даже файлами конфигурации конвейера.

## **4. Управление данными.**

Пользовательский интерфейс Soupi имеет набор виджетов -- инструментариев, предназначенных для упрощения выбора и организации данных, а также управлением и редактированием конфигурационных файлов.

### **а) Data Path Selector.**

Селектирование данных путей к файлам может быть вызвано из любой настройки подключаемого плагина и сохранено в режиме реального времени (например позволяет предварительно указать путь для внешних файлов калибровки набора изображений для плагина SCAMP). Виджет выбора нужных директорий (path) может просматривать как локальные, так и удаленные носители сетевой файловой системы и сохранить их для дальнейшего использования.

### **б) Image Selector.**

Загруженные заранее списки изображений выбираются пользователем пометкой галочками.

Некоторые критерии поиска могут быть объединены вместе и учитывать идентификатор фильтров/каналов, типы инструментов.

Когда изображение выбрано, селекцию можно отредактировать и при необходимости сохранить текущее выделение. Сохраненные выборки могут быть удалены или объединены с другими списками. В целях повышения гибкости настроек селектора добавлена функция загрузки одного текстового файла описания настроек (имена файлов изображений и контрольные MD5-суммы этих файлов) или целого пакета таких файлов, находящихся в каталоге.

Это позволяет организовать полуавтоматическую пакетную обработку без необходимости множества ручных операций. Что представляет значительные удобства при обработке наблюдений с очень большим количеством изображений, поскольку всю необходимую выборку можно подготовить заранее.

### **в) Маркировка данных тегами.**

Тег не описывается какой-либо строгой иерархией и является просто неким ключевым словом и может быть применен к части сохраняемой информации. Soupi позволяет отметить тегом загруженные изображения как некую единицу описания метаданных. Тегированные изображения могут быть в дальнейшем найдены и выбраны селектором по названию тега. Для создания нового тега или использования существующего нужно

только выбрать подходящие изображения и курсором перетащить в выделенное поле рисунок тега. В этом поле есть кнопки для маркировки или ее снятия для текущего списка отселектированных изображений. Количество тегов, которые могут быть применены к изображениям практически неограниченно.

#### г) Процессинговая корзина.

Перед запуском любой обработки в Soupi все входные данные, параметры конфигурации для плагинов и набор заданий сначала совместно связываются в элемент процессинговой корзины. Процессинговая корзина по сути аналогична корзине интернет-магазина, в которую помещают покупки. Это удобный инструмент для обмена между пользователями элементами, описывающими наборы данных и этапы их обработки. Каждый элемент в корзине может быть сохранен для последующего использования или загружен обратно в процессинговую корзину в любое время.

Механизм процессинговой корзины позволяет также менять набор реал-тайм опций перед запуском заданий в кластере.

### 5. Кластерные вычисления посредством Condor.

Soupi для организации кластерных вычислений использует программный пакет Condor.

Condor управляет распределением заданий на локальном компьютере, узлах кластера и Интернет-грид. Загрузка заданий на выполнение в среде Condor осуществляется через процессинговую корзину. Soupi собирает все необходимые входные параметры и передает их в Condor через сгенерированный стандартный submission файл – текстовый скрипт, специфицирующий параметры задания, переменные окружения, командные строки -- готовые для исполнения в операционной системе.

Оперативные данные настроек плагинов, участвующих в работе конвейера, URL файлов исходных и обработанных изображений, в том числе вспомогательных сохраняются в базе данных.

Таким образом реализован механизм передачи данных от одного плагина к другому для массовой автоматизированной обработки изображений. Работа плагина может быть как полностью автономной, но как правило, требует обращений к удаленным астрофизическим базам данных и каталогам, доступа к Астрогрид, это накладывает дополнительные требования к соответствующим узлам кластера.

В зависимости от требований конкретная конфигурация кластера может включать только необходимые узлы кластера с указанными параметрами. В Soupi есть специальная страница для Condor, где специфицируется набор требований в виде настраиваемых политик и селекторов. Пользовательские политики определяют динамические правила, из которых некоторые могут быть описаны регулярными выражениями, причем будут вычислены только перед тем, как задания будут загружены в запуск. В отличие от пользовательских, некий набор доступных узлов кластера может быть выбран статически и отображается на экране.

Состояние запланированных заданий Condor можно наблюдать на соответствующей страничке мониторинга. Важная информация, такая как описание задания на удаленном хосте, время работы, владелец и текущий статус отображается и обновляется в реальном времени без перезагрузки страницы.

### 6. Заключение.

Soupi является универсальным средством организации конвейера процедур обработки данных, который может быть построен на единственном хосте, десктопе, ноутбуке - или применен на крупных вычислительных комплексах, кластерах и грид. Программа является открытой в соответствии с GPL. Первые наработки его прототипа -- Youpi были опубликованы сообществом TERAPIX в 2009 году, а последние фрагменты кода версии 0.7 доступны с весны 2011 и более не менялись.

Версия 0.7 декларирует поддержку для данных CFHT-MegaCam, CFHT-WIRCam и ESO-VISTA-VIRCAM инструментов. Последняя версия 0.8, авторская, уже под названием Soupi позволяет использование Django 1.2 и выше. Вместо MySQL СУБД-ориентированные вызовы переписаны под использование PostgreSQL с поддержкой полноценных транзакций, а описания геометрических объектов -- точек и вырезок из небесной сферы типа POLYGON в нотации расширения PostGIS. СУБД PostgreSQL теперь разрешает множественность запросов и независимых модификаций базы данных состояния процедур из Django.

Набор инструментальных таблиц расширен параметрами основного наблюдательного прибора БТА CAO РАН -- Scorpio и Scorpio-2. В ближайшей перспективе расширение набора подключаемых плагинов программами предобработки архивных снимков БТА и их координатной привязки.

Автор благодарен своим коллегам: О.П. Желенковой и В.С. Шергину за живой интерес, проявленный к работе и консультации по вопросам астрофизической обработки.

Работа по проекту поддержана грантом РФФИ 11-07-00108а.

#### ЛИТЕРАТУРА:

1. Monnerville M., Semah G., 2010, ASPC, 434,495
2. Bertin E., 2006, ASPC, 351, 112
3. Bertin E., Arnouts S., 1996, A&AS 317,393

4. Bertin et al. 2002, ASP Conference Series, 281, 228
5. Goranova, Y. Hudelot, P. Magnard, F., McCracken, H., Mellier, Y., Monnerville, M., Schultheis, M., S'emah, G., Cuillandre, J.-C., Aussel, H., 2009, The CFHTLS T0006 Release, <http://terapix.iap.fr/cplt/T0006/T0006-doc.pdf>
6. <http://youpi.terapix.fr/>, <http://www.taverna.org.uk>
7. <http://www.djangoproject.com/>
8. <http://www.prototypejs.org/>, <http://script.aculo.us/>
9. <http://www.cs.wisc.edu/condor/>
10. <http://terapix.iap.fr/>
11. <http://www.postgresql.org/>, <http://postgis.org/>
12. <http://www.sao.ru/hq/vch/rfbr11-07-00108/>