V07b

JVOの研究開発（ワークフロー機能の向上）

田中昌宏、白崎裕治、大石雅寿、川野元聡、本田敏志、水本好彦、大江将史(国立天文台)、
安田直樹(東大宇宙線研)、増永良文(お茶の水女子大)、
石原康秀、堤純平(富士通)、中本啓之、小林佑介、坂本道人(セック)

JVO 口頭発表
3月29日（木）10:36-11:12 [F会場]

V04a	全体進捗　大石雅寿
V05a	すばる望遠鏡データ解析機能の導入　白崎裕治
V06b	分光データの取り込み　川野元聡
V07b	ワークフロー機能の向上　田中昌宏

jvo　検索

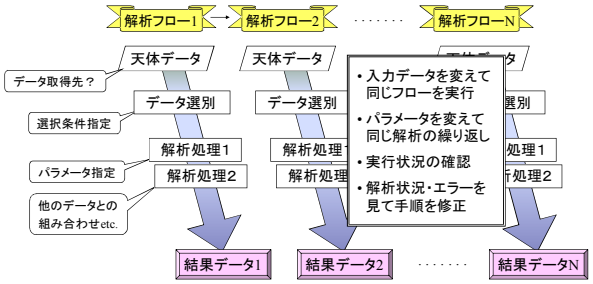
→ http://jvo.nao.ac.jp/

バーチャル天文台 (VO) では、世界中に分散配置された膨大な天文データアーカイブを高度に使いこなすことを目的とし、データアーカイブへのアクセス・連携方法の標準化、およびその標準に基づくデータベース/ソフトウェアシステムの開発を行っている。

我々が開発を進めてきたJVOでは、ネットワークを通じて提供されるサービスを利用し、それらを連携したデータ解析を可能にするため、ワークフロー機能の開発を行っている。昨年度は、基本的な制御構造および並列処理実行の記述が可能なワークフロー言語を設計し、そのワークフローに基づく処理システムの開発を行った。本年度の開発では、ワークフローの機能を向上させるため、ビルトイン関数を記述・追加する仕組みを開発した。これによって処理機能を充実させることで、ワークフローの利便性を向上させる。また、ワークフロー言語を拡張してテキストデータ1行ごとの処理機能を追加することにより、柔軟なデータ処理の記述を可能にした。さらに実行状況ステータスの導入により、ワークフロー処理の実行状況の把握を容易にする。ワークフローを記述する言語は、現在は計算機での扱いが容易なXMLで開発しているが、将来的にはユーザが容易にGUIを用いて記述できるようにする計画である。

ワークフロー言語の必要性

解析処理の流れの例:



このような手順を手作業でおこなうのは非効率。

ワークフロー言語により、処理手順を記述し、

自動実行ができれば効率的な処理が可能。

ワークフロー言語 開発の目標

- 複数のサービスをつなぐ一連の処理（ワークフロー）の記述
- Webサービスなどの外部サービスの利用
- ワークフローの自動実行
- 並列実行
- 実行状況をリアルタイムで確認
- 実行結果をログから確認

実装のポイント

- WF言語として、XMLを採用
  - 字句・構文解析の開発が不要
  - XMLスキーマによる仕様設計
- 実行エンジンとして、Groovyを利用
  - 実行エンジンの開発が不要
  - XMLからGroovyへの変換ルールを設計 (XSLTを利用)

ワークフローシステムの開発

2006年度の開発

- ワークフロー言語仕様の見直し、及び機能追加
  - タグを整理して記述量を低減
  - 配列定義の導入
  - awk構文 (テキストファイル1行毎のループ) の導入
  - そのほか制限事項の撤廃
- ビルトイン関数設定機能
  - 利用可能なコマンドやその利用方法を、設定ファイルに書くことにより、ビルトイン関数を登録できる。
- ステータス・ログ機能の設計

ワークフロー ログ機能

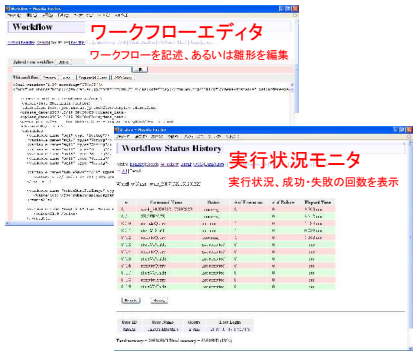
- アクティビティステータス
  - ワークフローの各実行要素(アクティビティ)に実行状況のステータスを付加
  - counter=実行回数、failed=失敗回数
- ワークフローログ
  - アクティビティの実行状況を実行順に出力
  - ※ ワークフロー単位で作成

（両者を関連づけて表示する機能を実装予定）

ログ機能の設計
- ログ管理クラス→ログファイルの作成/書き込み (Log4J)
- ログ検索クラス→ログファイルの検索

ログに出力される項目
- 時刻
- ログレベル　… INFO, DEBUG, ERROR
- アクティビティID
- 種別　… 実行前 / 実行中 / 実行後

JVOポータルからワークフローを実行



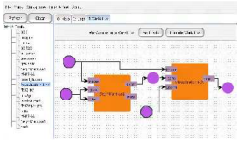
今後の方針

- IVOAにおける世界標準の策定に貢献する
- ワークフローの書き易さの向上

XMLの手書きは困難。そのため、
- テンプレートの充実
- 簡単な文法の言語との変換
- GUIワークフロービルダ

などの対応を予定。

GUIワークフロービルダの例：**JFLOW**
ストラブールデータセンターで開発されたワークフロービルダ



XMLワークフロー言語の仕様

基本的なプログラム言語と同じ制御構文に加え、並列実行、ファイル行ループ、外部サービス実行などの機能を備える。

変数宣言

```
<variables>
    <variable name="tableName" type="String">
        <value>ivo://jvo/subaru/spcam</value>
    </variable>
    <variable name="magLimit" type="double">
        <value>21.0</value>
    </variable>
    <variable name="id" type="int">
        <value>1</value>
    </variable>
    <variable name="url" type="String[]"/>
    <variable name="vot" type="VOTABLE[]"/>
    <variable name="params" type="Object[]"/>
    <variable name="map" type="Map"/>
    <variable name="list" type="List"/>
</variables>
```

variables タグ内の variable タグで変数を定義し、value タグで初期値を設定。

連続実行

```
<sequence>
    . . .
</sequence>
```

実行要素の列を逐次的に実行。

並列実行

```
<flow>
    . . .
</flow>
```

実行要素の列を並列実行。

条件ループ

```
<while condition="condition">
    . . .
</while>
```

条件を満たす間、繰り返し実行。

ループ・並列ループ

```
<for init="0" max="n" variable="i" step="1">
    . . .
</for>
```

```
<parfor items="a" variable="i[_count]">
    . . .
</parfor>
```

リスト/配列の要素毎、あるいは指定回数だけ繰り返し実行。

属性：

items	リスト、配列、VOTable
init	格納変数の初期値
max	繰り返し回数
step	格納変数の増分
variable	格納する変数

ビルトイン関数実行

```
<command xsi:type="builtin" name="executeQuery">
    <input>
        <varRef>sql</varRef>
    </input>
    <output>
        <varRef>vot</varRef>
    </output>
</command>
```

input要素を引数としてビルトインコマンドを実行し、戻り値をoutputの変数に格納。引数として、varRefで変数、literalで値を記述。

変数代入

```
<set name="where" literal="where region=Circle(ra,dec),@SIZE@)"/>
<set name="sql" ref="select+from+where"/>
```

literal属性に値を記述して代入。または、ref属性に変数または式を指定し、値を代入。

条件分岐

```
<if condition="qsoName==1">
    <then>
        . . .
    </then>
    <else>
        . . .
    </else>
</if>
```

条件が真ならばthenを、偽ならばelseを実行。

複数条件分岐

```
<switch>
    <case condition="param==1">
        . . .
    </case>
    . . .
    <otherwise>
    . . .
    </otherwise>
</switch>
```

複数のcaseから条件に合うものを実行し、どれにも合わなければotherwiseを実行。

ファイル行ループ

```
<awk file="tmp.dat" variable="l">
    . . .
</awk>
```

ファイルから行を読み込み、行毎に繰り返し実行。

属性：

file	入力ファイル
separator	区切り文字(正規表現)
variable	格納する変数
from	開始行
to	終了行

ビルトイン関数一覧（追加可能）

copyText	指定されたファイルを移動する
executeHyperZ	Photometric redshift を計算するサービスを呼び出す
executeQuery	JVOQLに基づいて検索実行する
getCurrentWorkDir	ワーキングディレクトリを取得
invoke	Webサービスを呼び出す
loadVOTable	VOTableを読み込んで、JAXBオブジェクトを返す。
storeVOTable	VOTableをファイルに保存
loadAsString	ファイルの内容を文字列として返す。
loadAsDataHandler	データハンドラオブジェクトを返す。
getFile	擬似パス名で指定されたファイルを表すFileオブジェクトを返す。
replace	文字列中の正規表現に一致する部分文字列全てを置き換える。
split	文字列をスペース区切りで分割する。
getColumnByName	VOTableから指定したカラムを取り出す
rm	ファイルを削除
sort	ファイルをソート
xmatch	VOTableを座標によるクロスマッチ結合を行う。
filenameMatch	ファイル名のパターンマッチング