
VisTrails Documentation

Release 1.7.0

University of Utah

December 05, 2011

CONTENTS

I	Preliminary Pages	1
1	Preface	3
1.1	Acknowledgements	3
II	An Introduction to VisTrails	5
2	What Is VisTrails?	7
2.1	Important Features	7
2.2	Obtaining the software	8
3	Getting Started	9
3.1	Installation	9
3.2	Quick Start	11
3.3	Manipulating VisTrails Files	11
3.4	VisTrails Basics	14
3.5	VisTrails Interaction	14
III	Learning VisTrails By Example	17
4	Creating and Modifying Workflows	19
4.1	Working with Modules	19
4.2	Adding and Deleting Modules	19
4.3	Connecting Modules	21
4.4	Changing Module Parameters	21
4.5	Changing Module Labels	23
4.6	Configuring Module Ports	23
4.7	Grouping Modules	24
4.8	Basic Modules	24
5	Interacting with the Version Tree	29
5.1	Version Tree View	29
5.2	Adding and Deleting Tags	33
5.3	Adding Version Annotations	33
5.4	Navigating Versions	33
5.5	Comparing Versions	34
6	Merging Two Version Trees	35
6.1	Example	35

7	Spreadsheet	39
7.1	The Spreadsheet Layout	39
7.2	Sending Output to the Spreadsheet	40
7.3	Interacting with the Spreadsheet	41
7.4	Launching a Web Browser	43
7.5	Saving a Spreadsheet	44
8	Querying the Version Tree	45
8.1	Query By Example	45
8.2	Textual Queries	47
8.3	Query Results	49
9	Using Analogies to Update Workflows	51
9.1	Visualization by Analogy	51
9.2	Using the Visual Diff Window	53
9.3	Using the Spreadsheet	54
9.4	Notes	57
10	Parameter Exploration	59
10.1	Creating a Parameter Exploration	59
10.2	Alternatives to Linear Interpolation	62
10.3	“Virtual Cell” Layout	66
IV	Intermediate Concepts and VisTrails Packages	67
11	Control Flow in VisTrails	69
11.1	The Map operator	69
11.2	Filtering results	73
11.3	Conditional module	74
12	The Control Flow Assistant	81
12.1	Using the Control Flow Assistant (CFA)	81
13	Connecting to a Database	85
13.1	Setup	85
13.2	Opening from a database	86
13.3	Saving to a database	87
13.4	Known Issues	87
14	Example: Web Services	89
14.1	Enabling the SUDSWebServices Package	90
14.2	Adding Web Service Packages	90
14.3	Creating a new vistrail	90
14.4	Executing the workflow	92
15	Persistence in VisTrails	95
15.1	Getting Started With Persistence	95
15.2	Using the Output of One Workflow as Input for Another	96
15.3	Managing Files in the Store	96
15.4	Examples	96
16	Embedding VisTrails Files Via Latex	103
16.1	Latex Setup and Commands	103
16.2	Including VisTrails Results in Latex	104

V	Developer's Guide	109
17	Writing VisTrails Packages	111
17.1	Introduction	111
17.2	Who Should Read This Chapter?	111
17.3	A Simple Example	111
17.4	Creating Reloadable Packages	115
17.5	Wrapping Command-line tools	116
17.6	Interfacing with the VisTrails Menu	119
17.7	Interpackage Dependencies	120
17.8	Package Requirements	121
17.9	Interaction with Caching	122
17.10	Customizing Modules and Ports	122
17.11	Generating Modules Dynamically	128
17.12	For System Administrators	130
18	Command-line Arguments	131
18.1	Starting VisTrails via the Command Line	131
18.2	Specifying a User Configuration Directory	132
18.3	Passing Database Parameters on the Command Line	133
18.4	Running VisTrails in Batch Mode	133
18.5	Executing Workflows in Parallel	136
18.6	Finding Methods Via the Command Line	137
19	Accessing the Execution Log	139
20	Example: ITK	141
20.1	Introduction to ITK	141
20.2	Preparing ITK	141
20.3	ITK and VisTrails	143
21	Control Flow in VisTrails	147
21.1	The Map operator	147
21.2	Filtering results	151
21.3	Conditional module	152
21.4	Building your own loop structure	156
VI	Indices and tables	161
	Bibliography	165
	Index	167

Part I

Preliminary Pages

PREFACE

Welcome to the *VisTrails User's Guide*. This book has been updated for version 1.6.2 of the VisTrails software.

VisTrails is a new scientific workflow management system developed at the University of Utah that provides support for data exploration and visualization. For an engineer or scientist, generating and evaluating hypotheses is an interactive process. With each change, a different, albeit related, workflow is created. VisTrails was designed to manage these rapidly-evolving workflows. By automatically managing the data, metadata, and the data exploration process, VisTrails allows you to focus on the task at hand and relieves you from tedious and time-consuming tasks involved in organizing vast volumes of data. VisTrails provides infrastructure that can be combined with and enhance existing visualization and workflow systems.

VisTrails is an open-source software system. You can contribute to VisTrails by sharing bug reports, bug fixes, and suggestions with the VisTrails community. The easiest way to get started is to sign up for the VisTrails Users mailing list. Instructions for doing this can be found on the VisTrails web site: www.vistrails.org.

This book is divided into four parts. The first part, *Introduction to VisTrails*, provides instructions on how to download and install the VisTrails software, and introduces you to its user interface. The second and longest part, “Learning VisTrails by Example,” consists of a number of tutorial chapters that guide you, step by step, through the features of VisTrails. We encourage you to try out these examples for yourself as you read this book. The third part provides information on additional features and packages. The fourth and final part is the “Developer’s Guide” and is intended for programmers who wish to add new features, packages, and modules to VisTrails.

We hope that you will find VisTrails to be a useful tool towards automating and streamlining your workflows, leading to faster discoveries and deeper insight.

For your convenience, the html version of this manual is also available at <http://www.vistrails.org/usersguide>.

About the figures: VisTrails works across multiple platforms, and the screenshots shown in this manual reflect this. Hence, some of the images in this book may vary slightly from what you see on your system, depending on the look and feel of your platform.

1.1 Acknowledgements

VisTrails research and development has been funded the Department of Energy SciDAC (VACET and SDM centers), the National Science Foundation (grants IIS-0746500, CNS-0751152, IIS-0713637, OCE-0424602, IIS-0534628, CNS-0514485, IIS-0513692, CNS-0524096, CCF-0401498, OISE-0405402, CCF-0528201, CNS-0551724), and IBM Faculty Awards (2005, 2006, 2007, and 2008).

Part II

An Introduction to VisTrails

WHAT IS VISTRAILS?

VisTrails is a new system that provides data and process management support for exploratory computational tasks. It combines features of both workflow and visualization systems. Similar to workflow systems, it allows the combination of loosely-coupled resources, specialized libraries, and grid and Web services. Similar to some visualization systems, it provides a mechanism for parameter exploration and comparison of different results. But unlike these other systems, VisTrails was designed to manage *exploratory processes* in which computational tasks evolve over time as a user iteratively formulates and tests hypotheses. A key distinguishing feature of VisTrails is its comprehensive provenance infrastructure that maintains detailed history information about the steps followed in the course of an exploratory task. VisTrails leverages this information to provide novel operations and user interfaces that streamline this process.

2.1 Important Features

One of our main uses for VisTrails has been exploratory visualization, but the system is much more general and provides many other features, such as:

- *Flexible Provenance Architecture.* VisTrails transparently tracks changes made to workflows, including all the steps followed in the exploration. The system can optionally track run-time information about the execution of workflows (*e.g.*, who executed a module, on which machine, elapsed time *etc.*). VisTrails also provides a flexible annotation framework whereby you can specify application-specific provenance information.
- *Querying and Re-using History.* The provenance information is stored in a structured way. You have a choice of using a relational database (such as MySQL or IBM DB2) or XML files in the file system. The system provides flexible and intuitive query interfaces through which you can explore and reuse provenance information. You can formulate simple keyword-based and selection queries (*e.g.*, find a visualization created by a given user) as well as structured queries (*e.g.*, find visualizations that apply simplification before an isosurface computation for irregular grid data sets).
- *Support for collaborative exploration.* The system can be configured with a database backend that can be used as a shared repository. It also provides a synchronization facility that allows multiple users to collaborate asynchronously and in a disconnected fashion—you can check in and check out changes, akin to a version control system (*e.g.*, SVN: <http://subversion.tigris.org>).
- *Extensibility.* VisTrails provides a very simple plugin functionality that can be used to dynamically add packages and libraries. Neither changes to the user interface nor re-compilation of the system are necessary. Because VisTrails is written in Python, the integration of Python-wrapped libraries is straightforward. For example, a single line in the VisTrails start-up file is needed to import all of VTK's classes.
- *Scalable Derivation of Data Products and Parameter Exploration.* VisTrails supports a series of operations for the simultaneous generation of multiple data products, including an interface that allows you to specify sets of values for different parameters in a workflow. The results of a parameter exploration can be displayed side by side in the VisTrails Spreadsheet for easy comparison.

- *Task Creation by Analogy*. Analogies are supported as first-class operations to guide semi-automated changes to multiple workflows, without requiring you to directly manipulate or edit the workflow specifications.

2.2 Obtaining the software

Visit <http://www.vistrails.org> to access the VisTrails community website. Here you will find information including instructions for obtaining the software, online documentation, video tutorials, and pointers to papers and presentations.

VisTrails is available as open source; it is released under the GPL 2.0 license. The pre-compiled versions for Windows and Mac OS X come with an installer and include a number of packages, including VTK, matplotlib, and Image Magick. Additional packages, including packages written by users, are also available (*e.g.*, ITK, Matlab, Metro). Developers can easily add new packages using the VisTrails plugin infrastructure.

GETTING STARTED

The VisTrails system is distributed both as source code and pre-built binaries, and instructions for obtaining either can be found at our website: <http://www.vistrails.org>. Because the system is written in Python using a Qt interface, it can be run on most architectures that support these two components, even if a pre-built binary is not available for your system. Section *Installation* provides instructions to guide you through installation procedures, and Section *Quick Start* gives a quick orientation and serves as a springboard for exploring the different features of VisTrails.

3.1 Installation

There are two types of VisTrails installations. The first is a binary installation that lets you use VisTrails by running the precompiled executable. The second is a full source code installation that requires you to install and compile VisTrails and all of its dependencies. Of the two types of installations, the binary version is much easier, and we encourage first-time users to use this option whenever possible. Precompiled binaries are currently available for Microsoft Windows (XP and Vista) and Mac OS X (10.4.x or higher). To obtain either a binary or source copy of VisTrails, please see our website: <http://www.vistrails.org>.

3.1.1 Installing VisTrails on Windows XP/Vista

To install VisTrails on Windows, download the installation bundle for Windows from the VisTrails website: <http://www.vistrails.org>. Unzip the file using the decompression program of your choice, then double-click the executable to begin installation (Figure *Installation wizard for Microsoft Windows XP/Vista*). Follow the prompts in the installation wizard to complete the installation process.

3.1.2 Installing VisTrails on Mac OS X

To install VisTrails on Mac OS X, download the installation bundle for Mac from the VisTrails website: <http://www.vistrails.org>. The precompiled binary currently only supports Mac OS X 10.5.x or higher. The disk image should be mounted automatically (Figure *Installing VisTrails on Mac OS X*). Once the disk image is mounted, drag the VisTrails desktop icon to the Applications folder to install the software.

3.1.3 Installing VisTrails on Ubuntu Linux

Although not a binary installation *per se*, installing VisTrails on Ubuntu Linux is nonetheless quite straightforward. VisTrails now interfaces with “apt” directly via a Python API. This allows dynamic installation of necessary packages. As a result, you do not need to manually install any of the dependent packages. Just download the VisTrails source code and execute it with:

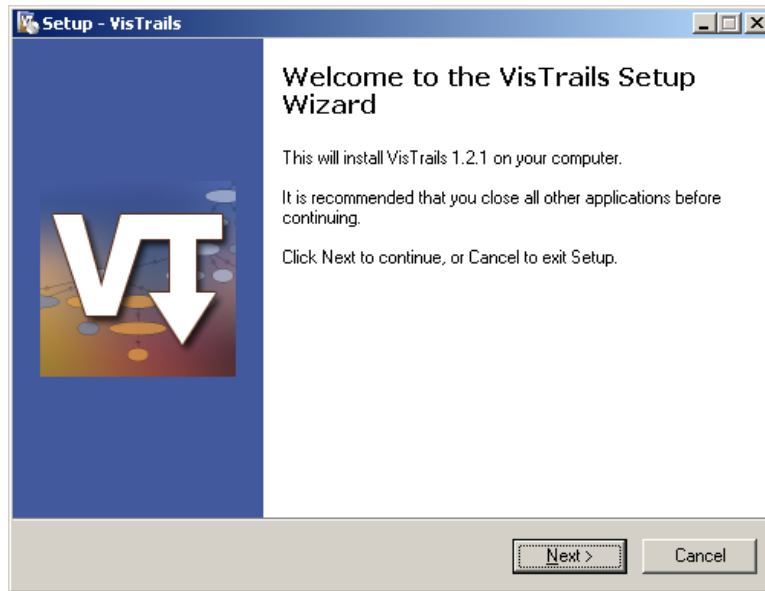


Figure 3.1: Installation wizard for Microsoft Windows XP/Vista.

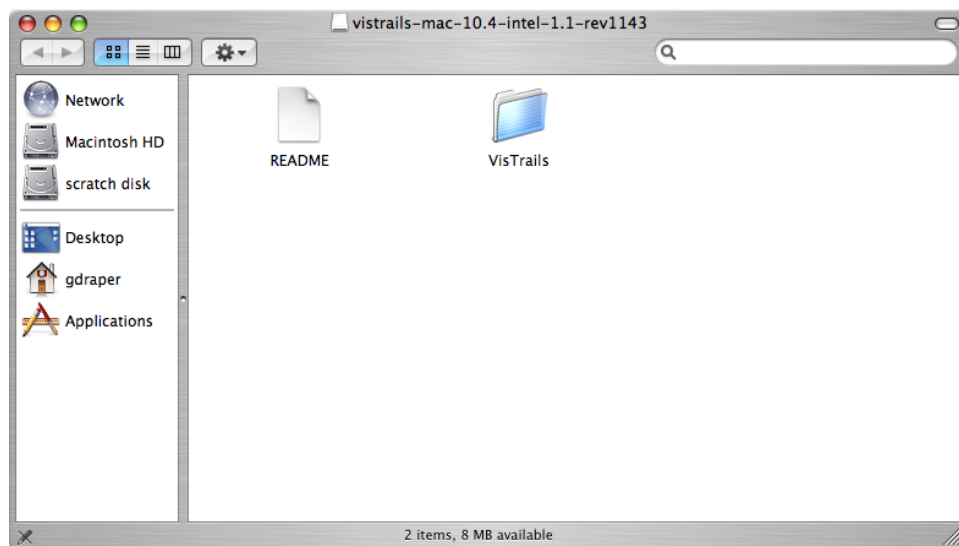


Figure 3.2: Installing VisTrails on Mac OS X.


```
python vistrails.py
```

and VisTrails should detect all necessary software and, if necessary, ask for your permission to install it.

3.1.4 Installing VisTrails from source

Installing VisTrails from source code is a non-trivial task. Rather than listing full compilation instructions in this manual, we instead provide a list of software packages upon which VisTrails is dependent, and refer you to the VisTrails website for additional details.

- Python 2.6 or higher
- Qt 4.4 or higher
- PyQt4
- SciPy
- VTK (needed to run the examples in this book)

There may also be additional dependencies, depending on which optional features of VisTrails you plan to use.

Please refer to http://www.vistrails.org/index.php/Mac_Intel_Instructions for more details.

3.2 Quick Start

On Windows and Mac OS X, you can launch VisTrails by double-clicking on the VisTrails application icon. In general, however, it is possible to start VisTrails on any system by navigating to the directory where the file `vistrails.py` is located (usually the root directory of your installation) and executing the command:

```
python vistrails.py
```

Depending on a number of factors, it can take a few seconds for the system to start up. As VisTrails loads, you may see some messages that detail the packages being loaded and initialized. This is normal operation, but if the system fails to load, these messages will provide information that may help you understand why. After everything has loaded, you will see the VisTrails Builder window as shown in Figure *VisTrails Builder Window*. If you have enabled the VisTrails Spreadsheet, you will also see a second window like that in Figure *VisTrails Spreadsheet Window*. (Note that the spreadsheet is enabled by default.) The VisTrails toolbar serves to help you navigate the various modes and functions VisTrails provides. As illustrated by Figure (*see Section Manipulating VisTrails Files*), the left side of the toolbar contains standard file manipulation buttons, and the next section provides buttons for execution and undo/redo functionality. The four buttons in the third section serve to switch between different modes to manipulate, query, and explore workflows. The right-most buttons allow you to toggle between different ways of navigating around the current canvas.

3.3 Manipulating VisTrails Files

To open a VisTrails file, or *vistrail*, you can either click the `Open` button in the toolbar or select `Open` from the `File` menu. This brings up a standard file dialog where you can select a vistrail to open. Vistrails are identified by the `.vt` file extension. Opening a vistrail adds a tab to the builder window, where each tab represents a different vistrail. Clicking a tab switches the current vistrail. Vistrails can also be stored in a database, enabling a central repository for workflows. See Chapter *Connecting to a Database* for more details about this feature. To close a vistrail, you can either choose the `Close` option from the `File` menu or click the red 'X' button on the upper right side of the builder window. If the vistrail has not been saved, you will be asked if you wish to save your work. To save a vistrail, there is

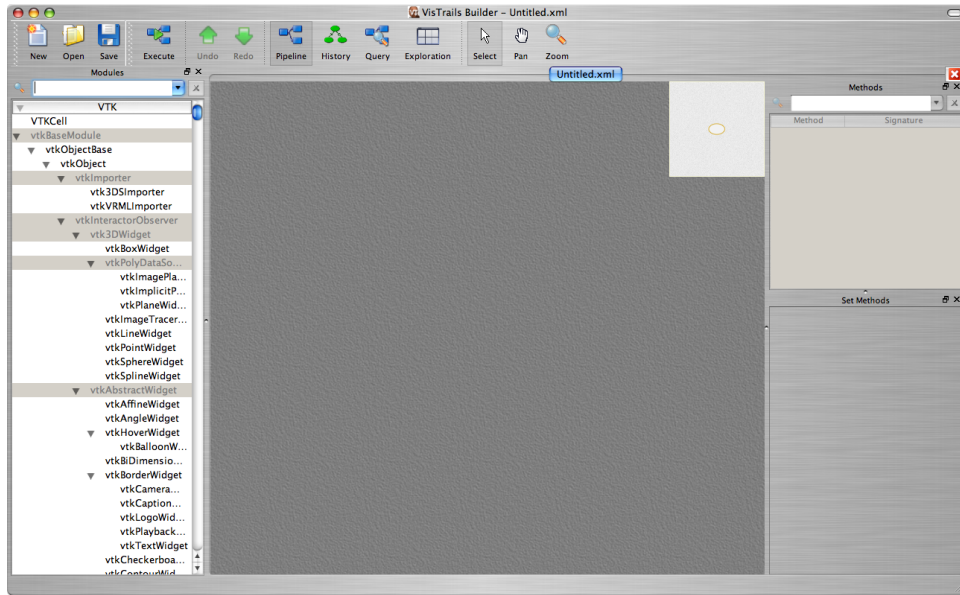


Figure 3.3: VisTrails Builder Window

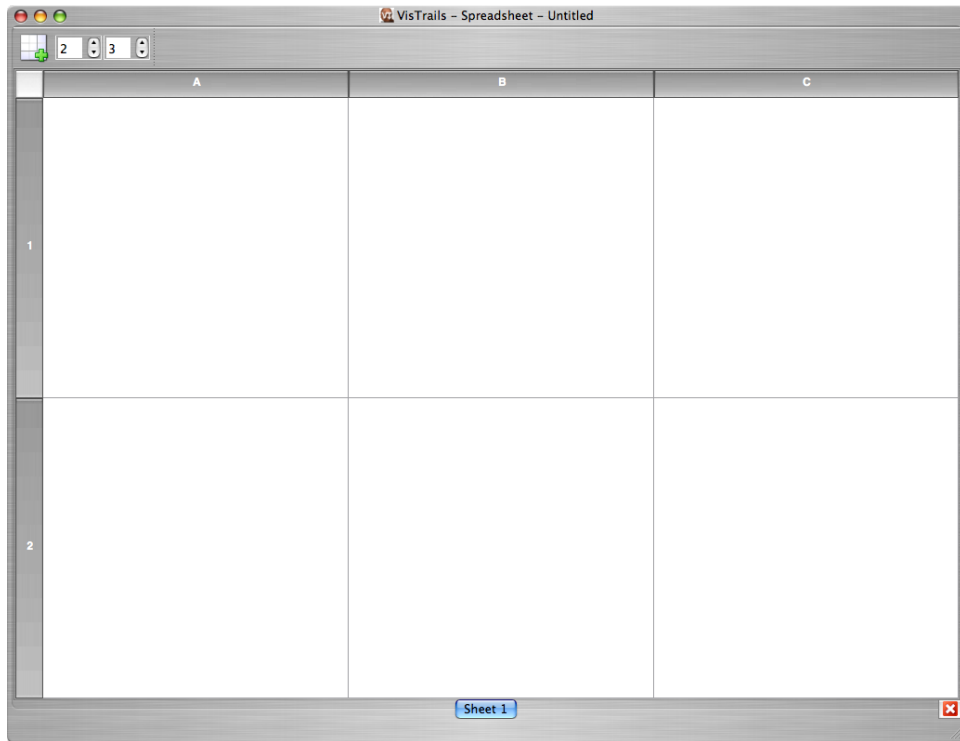


Figure 3.4: VisTrails Spreadsheet Window

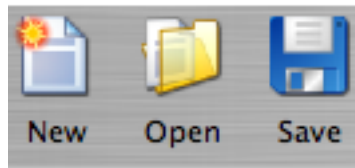


Figure 3.5: (see Section *Manipulating VisTrails Files*)

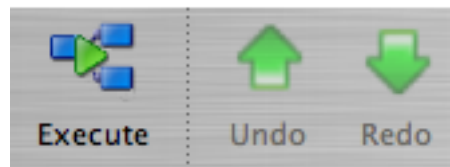


Figure 3.6: (see Section *VisTrails Interaction*)

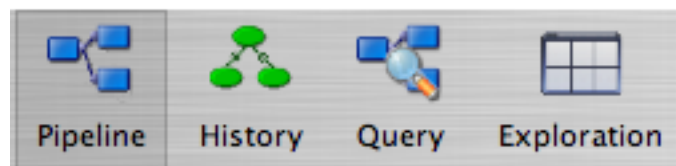


Figure 3.7: (see Chapters *Creating and Modifying Workflows*, *Interacting with the Version Tree*, *Querying the Version Tree*, *Parameter Exploration*)

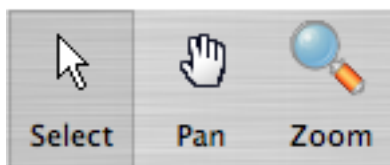


Figure 3.8: (see Section *VisTrails Interaction*)

both a button and a menu item in the `File` menu. If you would like to save the vistrail with a different name or in a different location, you can use the `Save As` option.

3.4 VisTrails Basics

In general, a *workflow* is a way to structure a complex computational process that may involve a variety of different resources and services. Instead of trying to keep track of multiple programs, scripts, and their dependencies, workflows abstract the details of computations and dependencies into a graph consisting of computational *modules* and *connections* between these modules.

The `Pipeline` button on the VisTrails toolbar accesses VisTrail’s interface for building workflows. Similar to many existing workflow systems, it allows you to interactively create workflows using an extensible library of modules and a connection protocol that helps you determine how to connect modules. To add a module to a workflow, simply drag the module’s name from the list of available modules to the workflow canvas. Each module has a set of input and output ports, and outputs from one module can be connected to inputs of another module, provided that the types match. For more information on building workflows in VisTrails, see Chapter *Creating and Modifying Workflows*.

In addition to VisTrail’s *Pipeline* interface for manipulating individual workflows, the *History* interface (accessed through the `History` button on the toolbar) contains a number of features that function on a collection of workflows. A *vistrail* is a collection of related workflows. As you explore different computational approaches or visualization techniques, a workflow may evolve in a lot of directions. VisTrails captures all of these changes automatically and transparently. Thus, you can revisit a previous version of a workflow and modify it without worrying about saving intermediate versions. This history is displayed by the VisTrails Version Tree, and different ways of interacting with this tree are discussed in Chapter *Interacting with the Version Tree*.

With a collection of workflows, one of the necessary tasks is to search for specific workflows. VisTrail’s search functionality is accessed by clicking the `Query` button on the toolbar. The criteria for these searches may vary from finding workflows modified within a specific time frame to finding workflows that contain a specific module. Because of the version history that VisTrails captures, these tasks are natural to implement and query. VisTrails has two methods for querying workflows, a simple text-based query language and a query-by-example canvas that lets you build exactly the workflow structure you are looking for. Both of these techniques are described in Chapter *Querying the Version Tree*.

The `Exploration` button allows you to explore workflows by running the same workflow with different parameters. Parameter Exploration provides an intuitive interface for computing workflows with parameters that vary in multiple dimensions. When coupled with the VisTrails Spreadsheet, parameter exploration allows you to quickly compare results and discover optimal parameter settings. See Chapter *Parameter Exploration* for specific information on using Parameter Exploration.

3.5 VisTrails Interaction

3.5.1 Workflow Execution

The `Execute` button on the toolbar serves as the “play” button for each of the modes described above. In both the `Builder` and `Version Tree` modes, it executes the current workflow. In `Query` mode, it executes the query, and in `Parameter Exploration` mode, it executes the workflow for each of the possible parameter settings.

When a workflow is executed, the module color is determined as follows:

- lilac: module was not executed
- yellow: module is currently being executed
- green: module was successfully executed

- orange: module was cached
- red: module execution failed

Note

VisTrails caches by default, so after a workflow is executed, if none of its parameters change, it won't be executed again.

If a workflow reads a file using the basic module File, VisTrails does check whether the file was modified since the last run. It does so by keeping a signature that is based on the modification time of the file. And if the file was modified, the File module and all downstream modules (the ones which depend on File) will be executed.

If you do not want VisTrails to cache executions, you can turn off caching: go to Menu Edit -> Preferences and in the General Configuration tab, change Cache execution results to Never.

If you would like your input and output data to be versioned, you can use the Persistence package.

3.5.2 Additional Interactions

The Undo and Redo buttons function in the standard way, but note that these actions are implicitly switching between different versions of a workflow. Thus, you will notice that as you undo or redo a change to a workflow, the selected version in the version tree changes. For all modes except Parameter Exploration, the center pane of VisTrails is a canvas where you can manipulate the current workflow, version tree, or query. The buttons on the right side of the toolbar allow you to change the default behavior of the primary mouse button (the left button for most multiple button mice) within this canvas. You can choose the behavior to select items in the scene, pan around the scene, or zoom in and out of the scene by selecting the given button. In addition, if you are using a 3-button mouse, the right button will zoom, and the middle button will pan. To use the zoom functionality, click and drag up to zoom out and drag down to zoom in.

Note

Pressing Ctrl-R will recenter the window. You may also zoom and pan in the window in the upper right corner of the main window (the picture-in-picture window). To recenter this picture-in-picture (PIP) window, first click on it to bring it to focus, then press Ctrl-R (or Command-R on a Mac) to recenter.

Part III

Learning VisTrails By Example

CREATING AND MODIFYING WORKFLOWS

4.1 Working with Modules

In VisTrails, modules are represented by a rectangle in the `Pipeline` view of the Builder. The name of the module is shown in bold letters in the middle of the rectangle. The input and output ports for the module are denoted by small squares on the top and bottom of the module, respectively. Modules are connected together to define the dataflow using curved black lines that go from output to input ports between modules. Each module may have also have adjustable parameters that can be viewed when a module is selected. Modules can be connected, disconnected, added, and deleted from a workflow.

As a running example in this chapter, we will make some changes to the “`vtk_book_3rd_p189.vt`” vistrail, included in the “examples” folder of the VisTrails installation.

Try it now!

Open the “`vtk_book_3rd_p189.vt`” vistrail, either by selecting `File` → `Open` from the menu, or by clicking the `Open` button on the toolbar. After opening this vistrail, click on the `Pipeline` toolbar button to enter workflow editing mode.

4.2 Adding and Deleting Modules

A list of available modules is displayed hierarchically in the `Modules` container on the left side of the VisTrails Builder (Figure *The main VisTrails Pipeline user interface. The major components are labeled.*). A core set of basic modules is always distributed with the VisTrails system. Other packages, such as VTK, are also distributed, but are not necessary for VisTrails and thus can be disabled on startup (see Chapter *Writing VisTrails Packages*). Note, however, that the VTK module *is* required for most of the examples in this book. Depending on the number of packages imported on startup, the number of modules to select from can be difficult to navigate. Thus, a simple search box is provided at the top of the container to narrow the displayed results. To add a module to the workflow, simply drag the text from the `Module` container to the workflow canvas. Modules and connections may be selected in multiple ways and are denoted by a yellow highlight. Besides directly left clicking on the object, a box selection is available by left clicking and dragging over the modules and connections in the canvas. Multiple selection can be performed with the box selection as well as by right clicking on multiple objects with the ‘Shift’ key pressed. There are several ways to manipulate selected modules in the workflow canvas. Moving them is performed by dragging a selected module using the left mouse button. Deleting selected modules is performed by pressing the ‘Delete’ key. The modules and connections can also be copied and pasted using the `Edit` menu, or with ‘Ctrl-C’ and ‘Ctrl-V’, respectively.

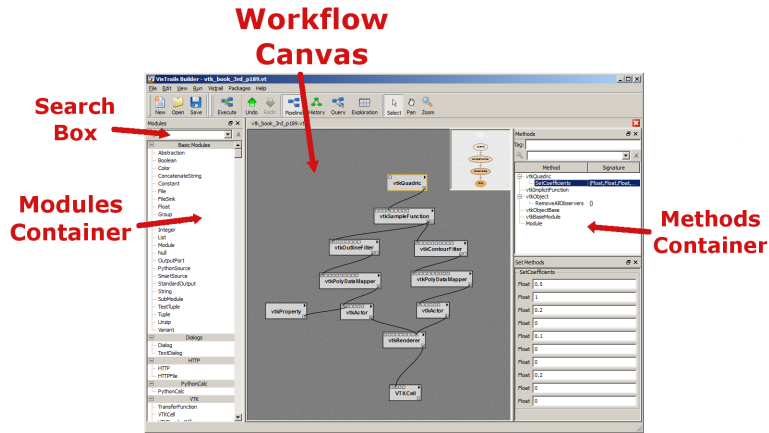


Figure 4.1: The main VisTrails Pipeline user interface. The major components are labeled.

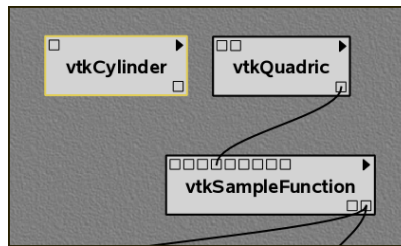


Figure 4.2: The vtkCylinder module is added to the canvas.

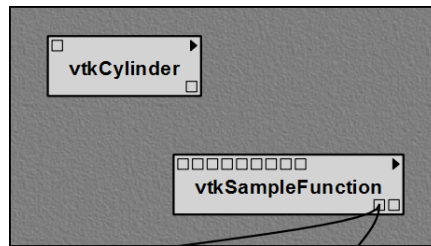


Figure 4.3: The vtkQuadric module is deleted.

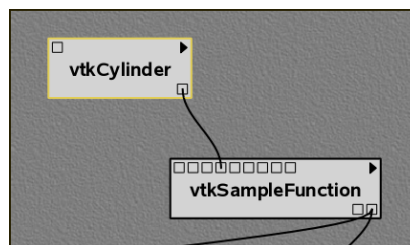


Figure 4.4: The connection replaced.

Try it now!

Let's replace the `vtkQuadric` module in our example with a `vtkCylinder` module instead. To do this, first type "vtkCylinder" into the search box of the `Module` container. As the letters are typed, the list filters the available modules to match the query. Select this module and drag the text onto an empty space in the canvas. (See Figure *The vtkCylinder module is added to the canvas.*. Then, select the `vtkQuadric` module in the canvas and press the 'Delete' key. This removes the module along with any connections it has (see Figure *The vtkQuadric module is deleted.*).

4.3 Connecting Modules

Modules are connected in VisTrails through the input and output ports at the top and bottom of the module, respectively. By hovering the mouse over the box that defines a port, the name and data type are shown in a small tooltip. To connect two ports from different modules, start by left clicking inside one port, then dragging the mouse to the other. The connection line will automatically snap to the ports in a module that have a matching datatype. Since multiple ports may match, hovering the mouse over the port to confirm the desired match may be necessary. Once a suitable match is found, releasing the left mouse button will create the connection. Note, a connection will only be made if the input and output port's data types match. To disconnect a connection between modules, the line between the modules can be selected and deleted with the 'Delete' key.

Try it now!

To connect the `vtkCylinder` module to the `vtkSampleFunction` module, place the cursor over the only output port on the `vtkCylinder` module, located on the bottom right. A tooltip should appear that reads "Output port self (vtkCylinder)." Left click on the port and drag the mouse over the `vtkSampleFunction` module. The connection should snap to the fourth input port from the left. Hovering the mouse over this port shows a tooltip that reads "Input port SetImplicitFunction (vtkImplicitFunction)." Release the mouse button to complete the connection between these two modules (see Figure *The connection replaced.*. To check for a valid dataflow, execute the workflow by pressing the `Execute` button on the toolbar, and see if the results appear in the spreadsheet.

4.4 Changing Module Parameters

The parameters for a module can be accessed in the `Methods` container located on the right side of the `Builder` window. When a module is selected from the canvas, the corresponding methods are displayed. As with the `Modules` container, a search box is provided to quickly find a desired method. By default, the `Builder` only manages methods with "set parameters." To check the set parameters, a `Set Methods` container is available below the `Methods` container. Changing a parameter can be performed directly in the `Set Methods` container. To set a parameter for the first time, click on the corresponding method and drag it into the `Set Methods` container, then enter the parameters directly into the text boxes. To remove a method, simply select the method in the `Set Methods` container and press the 'Delete' key.

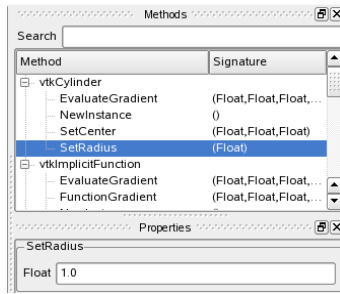


Figure 4.5: The module methods interface is shown with a change of the `SetRadius` parameter to 1.0.

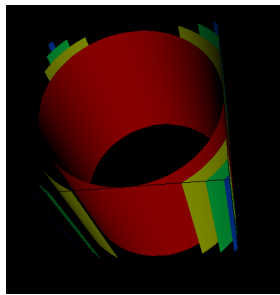


Figure 4.6: The results of the changes are displayed on execution.

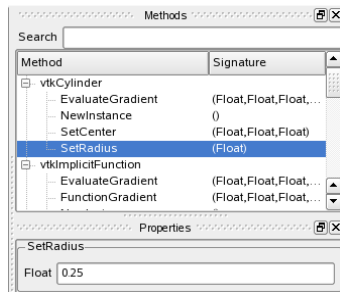


Figure 4.7: The module methods interface is shown with a change of the `SetRadius` parameter to 0.25.

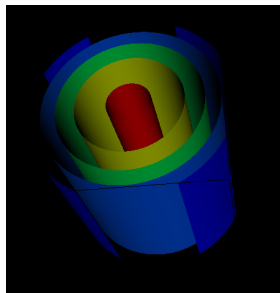


Figure 4.8: The results of the changes are displayed on execution.

Try it now!

To perform a parameter change, select the `vtkCylinder` module in the canvas. The methods are shown hierarchically in the `Methods` container. Find the `SetRadius` method and select it, then drag the highlighted text from the `Methods` container into the `Set Methods` container below. The result is a `SetRadius` box with a `Float` text input. Enter 0.25 into the text box and press the 'Enter' key. By executing the workflow, the modified visualization appears in the spreadsheet. Figure *The module methods interface is shown with a change of the `SetRadius` parameter to 1.0.* shows the interface and results of the parameter explorations.

4.5 Changing Module Labels

A new label can be assigned to a module by selecting the triangle in its top right corner to open a popup menu and selecting the `Set Module Label...` menu item. You will then be prompted to enter the new label. The new label will be displayed in the prominent position and the original module name will be displayed below it in parenthesis.

4.6 Configuring Module Ports

For convenience, all the inputs and outputs of a module are not always shown in the canvas as ports. The ports that are shown by default are defined using an option when defining the method signatures of a package. To access the full list of ports, the module configuration window is used. This is opened by selecting the triangle at the top right of a module to open a popup menu and selecting the `Edit Configuration` menu item, or alternatively by pressing 'Ctrl-E' when a module is selected. The window shows a list of input and output ports and allows you to toggle any additional ports to enable. When the configuration is complete, the new ports will appear on a module with a circle icon instead of the normal square. These new ports can then be used for connections in the same way as the others.

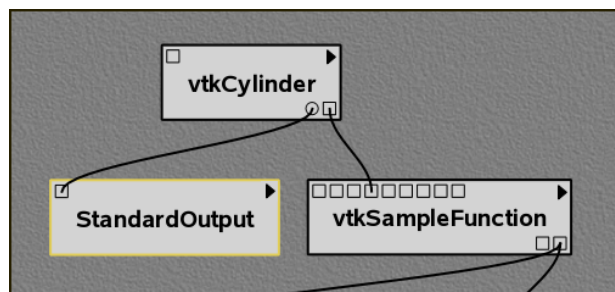


Figure 4.9: The `vtkCylinder` module is configured to show an additional `GetRadius` port, which is then connected to a `StandardOutput` module.

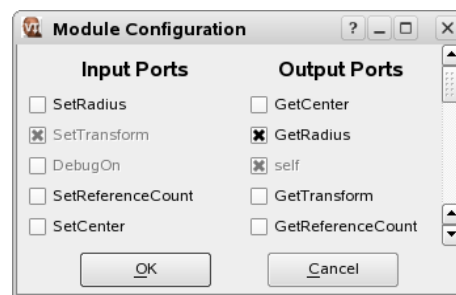


Figure 4.10: The module configuration window allows the hidden ports to be displayed.

Try it now!

As an example of configuring a module port, select the `vtkCylinder` module in the canvas and press ‘Ctrl-E’. In the newly opened configuration window, check the box for the `GetRadius` port, then click OK to close the window. A new circle port should appear on the module. Next, add a new `StandardOutput` module from the basic modules and connect the output port for `GetRadius` to the input port of `StandardOutput`. Upon execution, the value 0.25 is now output to the console. Figure *The vtkCylinder module is configured to show an additional GetRadius port, which is then connected to a StandardOutput module.* shows the new workflow together with the module configuration window.

4.7 Grouping Modules

As the number of modules in a pipeline increases, the pipeline can grow quite large and cumbersome. This also makes the pipeline more difficult to understand and maintain. With any large system, it can be helpful to cluster related pieces together and represent them as a single unit. This idea, called *encapsulation*, is commonly used in computer programming as a way of controlling complexity. VisTrails likewise supports the grouping of multiple modules together so that they can be treated as a single module. This “group module” can be thought of as a monolithic entity that performs all the same functions as its individual parts, but shields its inner details from everyone else. As such, a group module inherits all the input and output ports of the modules inside it, but only displays those ports that have connections to another module outside of the group. To borrow another term from programming languages, these visible ports might be considered the *public interface* of the group module.

Multiple modules are grouped together by first selecting them, and then choosing the `Group` option from the `Edit` menu. Alternatively, you can use the keyboard shortcut ‘Ctrl-G’.

An example may clarify how this works.

Try it now!

Select the `vtkOutlineFilter`, `vtkPolyDataMapper`, `vtkProperty`, and `vtkActor` modules on the left side of the pipeline, as shown in Figure *Box selection of four modules.*. Type ‘Ctrl-G’ to group these modules. Notice how the pipeline changes, as shown in Figure *The modules represented as a single group module.*. Since the label “Group” isn’t very descriptive, you can change this by clicking on the triangle in the top right of the module, and selecting the `Set Module Label` menu option. Type a more descriptive name, such as “BoundingBox,” into the text field and click OK. The new label is reflected in the pipeline (Figure *Renaming the group.*).

Just as any number of modules may be clustered into a group, any number of groups may be combined with other groups or modules to form still larger groups. This is done in the same way as described above.

In addition, any group may be also un-grouped; that is, restored to its individual modules. This is done by selecting the group module in the pipeline, and then choosing the `Edit` → `Ungroup` menu option. Alternatively, you can use the keyboard shortcut ‘Ctrl-Shift-G’.

4.8 Basic Modules

In addition to the modules provided by external libraries, VisTrails provides a few basic modules for convenience and to facilitate the coupling of multiple packages in one workflow. These modules mostly consist of basic data types in Python and some manipulators for them. In addition, file manipulation modules are provided to read files from disk and write files to disk. Because not every Python operation can be represented as a module, the `PythonSource`



Figure 4.11: Box selection of four modules.

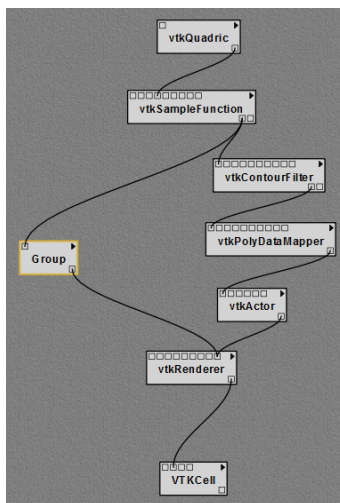


Figure 4.12: The modules represented as a single group module.

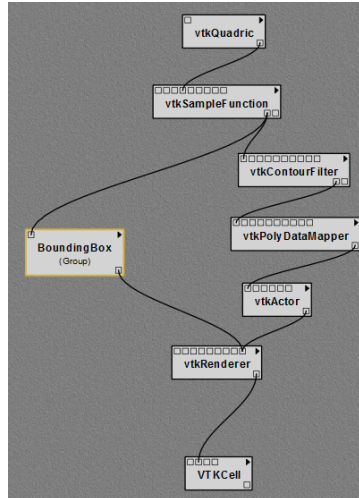


Figure 4.13: Renaming the group.

module is provided to allow you to write Python statements to be executed as part of a workflow. By pressing ‘Ctrl-E’ when a `PythonSource` module is selected in the canvas, a configuration window is opened. This window allows you to specify custom input and output ports as well as directly enter Python source to be executed in the workflow.

Try it now!

To demonstrate a `PythonSource` module, we will output the center of the cylinder using Python instead of the `StandardOutput` module. First, add a `PythonSource` module to the canvas and remove the `StandardOutput` module. Select the `PythonSource` module and press ‘Ctrl-E’ to edit the configuration. In the newly opened configuration window, create a new input port named “radius” of type `Float`. Next, in the source window enter:

```
print radius
```

then select OK to close the window. Finally, connect the `GetRadius` output of the `vtkCylinder` module to the new input port of `PythonSource`. Upon execution, the radius of the cylinder is printed to the console as before. Figure *A PythonSource module can be used to directly insert scripts into the workflow.* shows the new workflow together with the `PythonSource` configuration window.

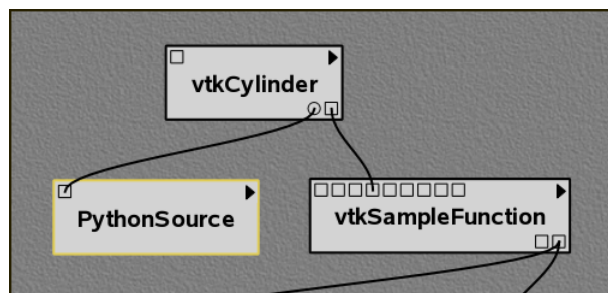


Figure 4.14: A `PythonSource` module can be used to directly insert scripts into the workflow.

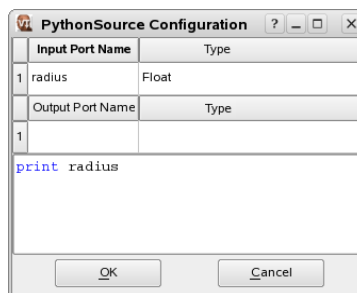


Figure 4.15: The configuration window for `PythonSource` allows multiple input and output ports to be specified along with the Python code that is to be executed.

INTERACTING WITH THE VERSION TREE

5.1 Version Tree View

The **History** button on the VisTrails toolbar lets users interact with a workflow history. It consists of a tree view in the center and the **Properties** tool container on the right for querying and managing version properties (Figure *In History mode, you can examine and annotate different versions of a workflow.*). Versions are displayed as ellipses in the tree view where the root of the tree is displayed at the top of the view. The nodes of the tree correspond to a version of a workflow while an edge between two nodes indicates that one was derived from the other. The nodes are

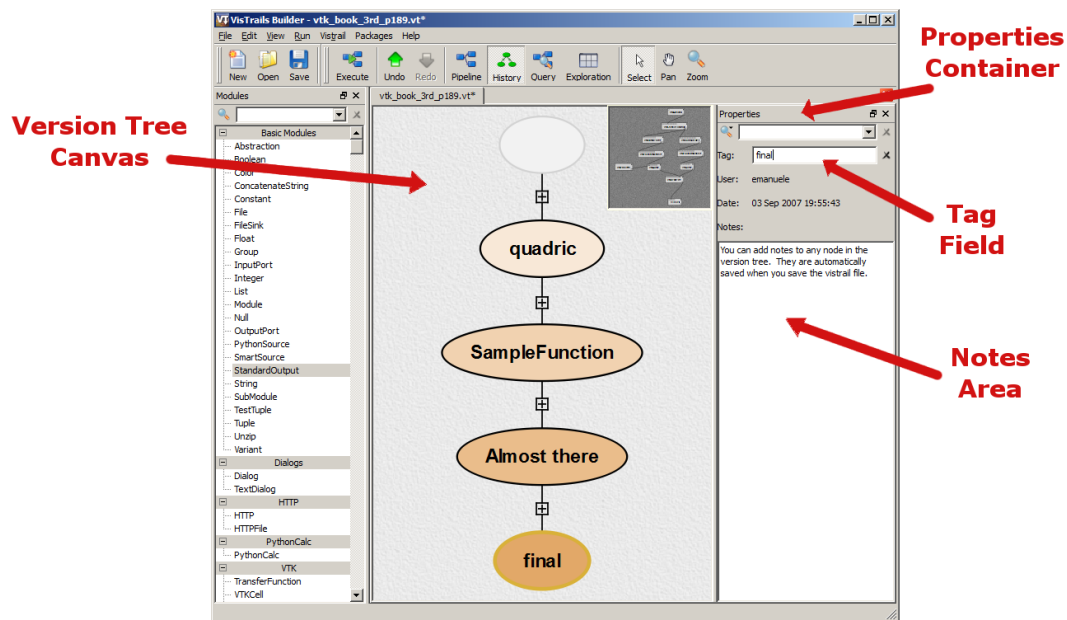


Figure 5.1: In *History* mode, you can examine and annotate different versions of a workflow.

displayed as colored ellipses, and are either blue or orange. A blue color denotes that the corresponding version was created by the current user while orange nodes were created by other users. The brightness of each node indicates how recently a version was created; brighter nodes were created more recently than dimmer ones. Each node may also have a *tag* that describes the version, and this tag is displayed as a text label in the center of the ellipse of the corresponding version.

The nodes are connected by a solid line if the child node is a direct descendant of the parent node; that is, if you have

made only a single change from the older version to the newer version. By default, only nodes that are leaves, have more than one child node, are specially tagged (see Section *Adding and Deleting Tags*), and the current version, will be displayed. To save space, other nodes will be “collapsed,” or hidden from view. Collapsed nodes are marked by the appearance of a small expansion button along an edge connecting two nodes (Figure *To conserve space, linear sequences of non-tagged nodes may be hidden from view.*). Clicking this button expands the version tree to reveal the hidden versions (Figure *They can be restored by clicking on the expansion button, which resembles a plus sign (+) inside a small box.*). The tree expansion is smoothly animated to help you maintain context from one view to the next. Clicking the button a second time collapses the nodes once again. Because most non-trivial changes to a workflow take more than action, most edges in a the version tree will be shown with these expansion buttons.



Figure 5.2: To conserve space, linear sequences of non-tagged nodes may be hidden from view.

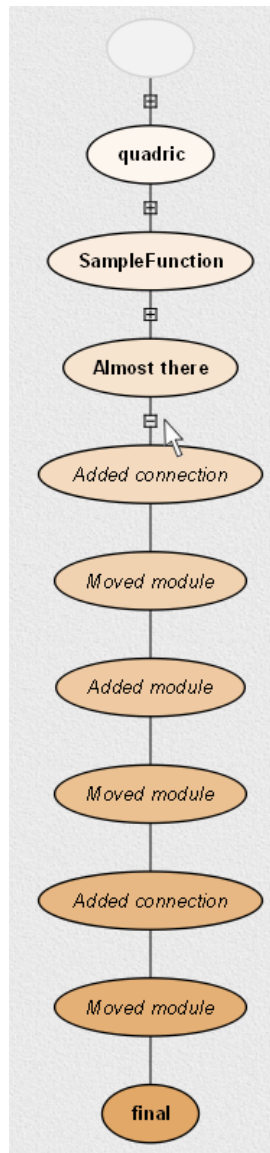


Figure 5.3: They can be restored by clicking on the expansion button, which resembles a plus sign (+) inside a small box.

Try it now!

To see an example of a version tree, load the example vistrail “vtk_book_3rd_p189.vt.” All versions will be shown in orange unless your username happens to be “emanuele.” Recall that this tree displays the structure of changes to a workflow, so let’s make some changes to see their effect on the version history. In the History view, select the node tagged *Almost there*, and then click on the Pipeline button to switch to a view of the workflow. Select a connection and delete it. Now, switch back to the History view, and notice that there is a new child node connected to *Almost there*. In addition, the line connecting the new node to its parent is solid, indicating that only a single change has been made. If we delete more connections, an expansion button would appear (Figure *All versions created by other users are shown in orange.*).

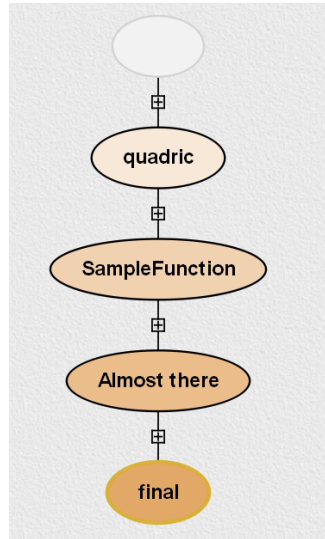


Figure 5.4: All versions created by other users are shown in orange.

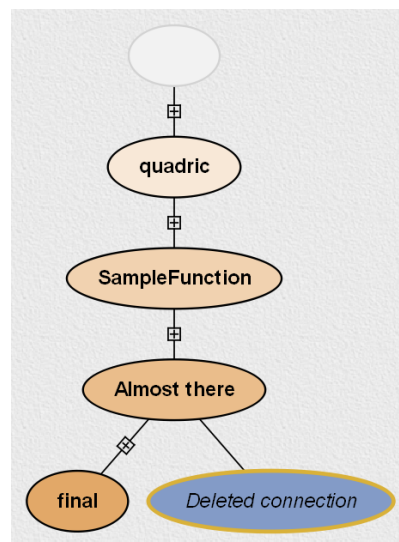


Figure 5.5: Deleting a connection results in a blue version connected by a solid line.

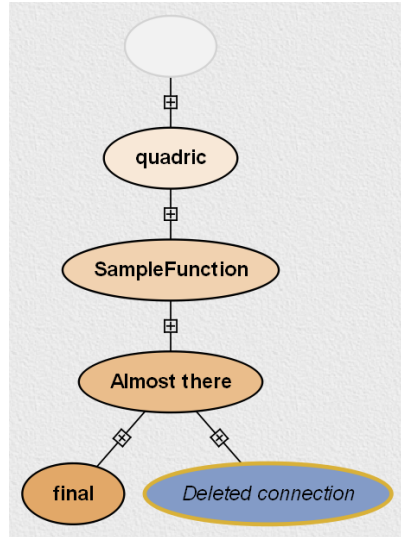


Figure 5.6: More interactions on this version cause additional versions to be created. To save space, these intermediate nodes are hidden, and an expansion button appears on the edge between the current node and the last tagged node.

5.2 Adding and Deleting Tags

As noted above, only certain nodes, including specially tagged ones, are shown by default in the version tree. To tag a version, simply click inside a selected node and type some meaningful text. The tag is created when you either click outside the node or press ‘Enter’. If you would like to change the tag to a different text, click inside the node again and modify the text as before. (Alternately, you can also create and modify tags using the `Tag` text field in the `Properties` container.)

Note that deleting all of the text in the tag field effectively deletes the tag. A second way to delete a tag is to click the ‘X’ button to the right of the text box. Removing a tag from a node may cause it to not be displayed in the default version tree view if it doesn’t satisfy any of the other criteria for display.

5.3 Adding Version Annotations

In addition to the tag field, the `Properties` container also displays information about the user who created the selected version and when that version was created. At the top of the container is a field for querying a vistrail; this functionality is described in detail in Chapter *Querying the Version Tree*. The final piece of the panel is the `Notes` field which allows users to store notes or annotations related to a version. As with tags, adding notes to a version is as easy as selecting the desired version and modifying the text field. Notes are automatically saved when you save the vistrail file.

5.4 Navigating Versions

Clicking on nodes in the version tree is not the only way to navigate different versions of a workflow; you can also use the `Undo` and `Redo` buttons. Because the version tree captures all changes to a workflow, undo and redo not only revert or reinstate changes to a workflow, but also change the currently selected version in the version tree. More precisely, undoing a change in a workflow is exactly the same as selecting the parent of the current node in the version tree. Note that because the current version is always shown in the version tree, undo and redo provide an effective way to navigate between two nodes whose intermediate versions might be currently hidden from view.

5.5 Comparing Versions

While selecting versions in the History view and using the Pipeline view to examine each version is extremely useful, this approach can be cumbersome when trying to compare two different versions. To help with such a comparison, VisTrails provides a Version Difference mechanism for quickly comparing two versions. There are two ways to compute this difference. The first is to select two versions in the tree and choose `Execute Version Difference` from the Run menu. The easier method is to drag one version onto the other. After either method, a Visual Diff window will open (see Figure A *Visual Diff showing the difference between version z-space and version textureMapper*). The difference is displayed in a manner that is very similar to the pipeline view, but modules and connections are colored based on similarity. Dark gray indicates those modules and connections that are shared between the two versions; orange and blue show modules and connections that exist in one workflow and not the other; and light gray modules are those where parameters between the two versions differ. Clicking the Legend button will bring up a window to remind you what each color corresponds to. For a module that is colored light gray, clicking on the Parameter Changes button will bring up a window that shows the difference in parameters for that module.

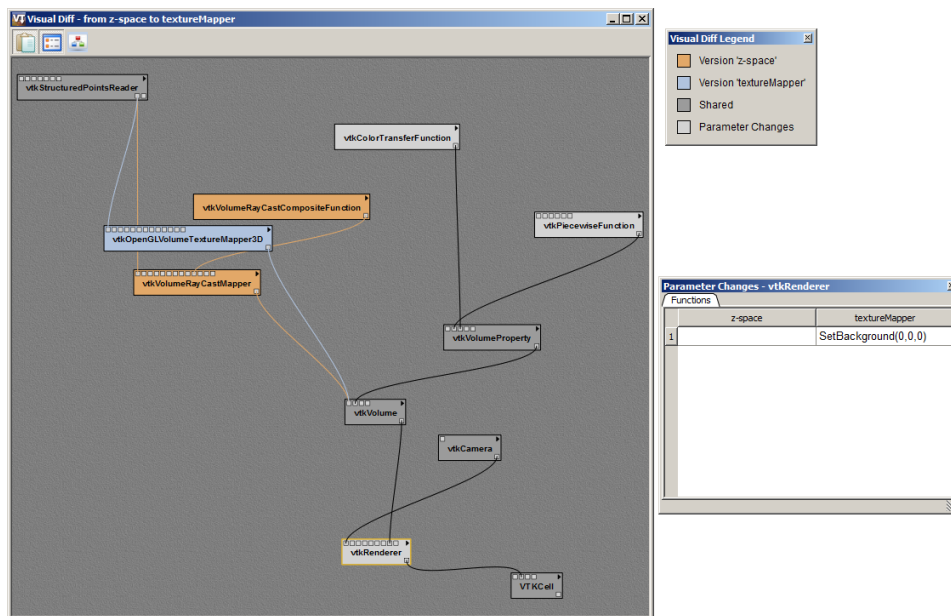


Figure 5.7: A Visual Diff showing the difference between version z-space and version textureMapper.

Try it now!

To try out this feature, open the “lung.vt” example vistrail, and click the History button. Within the version tree, click and drag the z-space version to the textureMapper version. After the diff appears, click on the Show Parameter Inspector Window button, and then click on the vtkRenderer module to see the parameter differences. We can see that one of the changes from z-space to textureMapper was to add a black background. Figure A *Visual Diff showing the difference between version z-space and version textureMapper*. shows the result of this comparison.

MERGING TWO VERSION TREES

One of the benefits of having a version tree is that branching allows users to work on multiple workflows within the same file. This is especially useful when workflows are similar or when one workflow provides output for another. However, if a user creates two different workflows in two different files and decides he/she wants them to be part of the same file/history, VisTrails allows file merging.

To merge two files:

- Open both files
- Select one of the files you would like to be merged.
- Place you mouse over the `Merge With` arrow from the Edit menu. A list of open files should appear.
- Select the file that you would like to join with the current (previously selected) file.

The history trees of both files should now be joined and placed in a new file.

6.1 Example

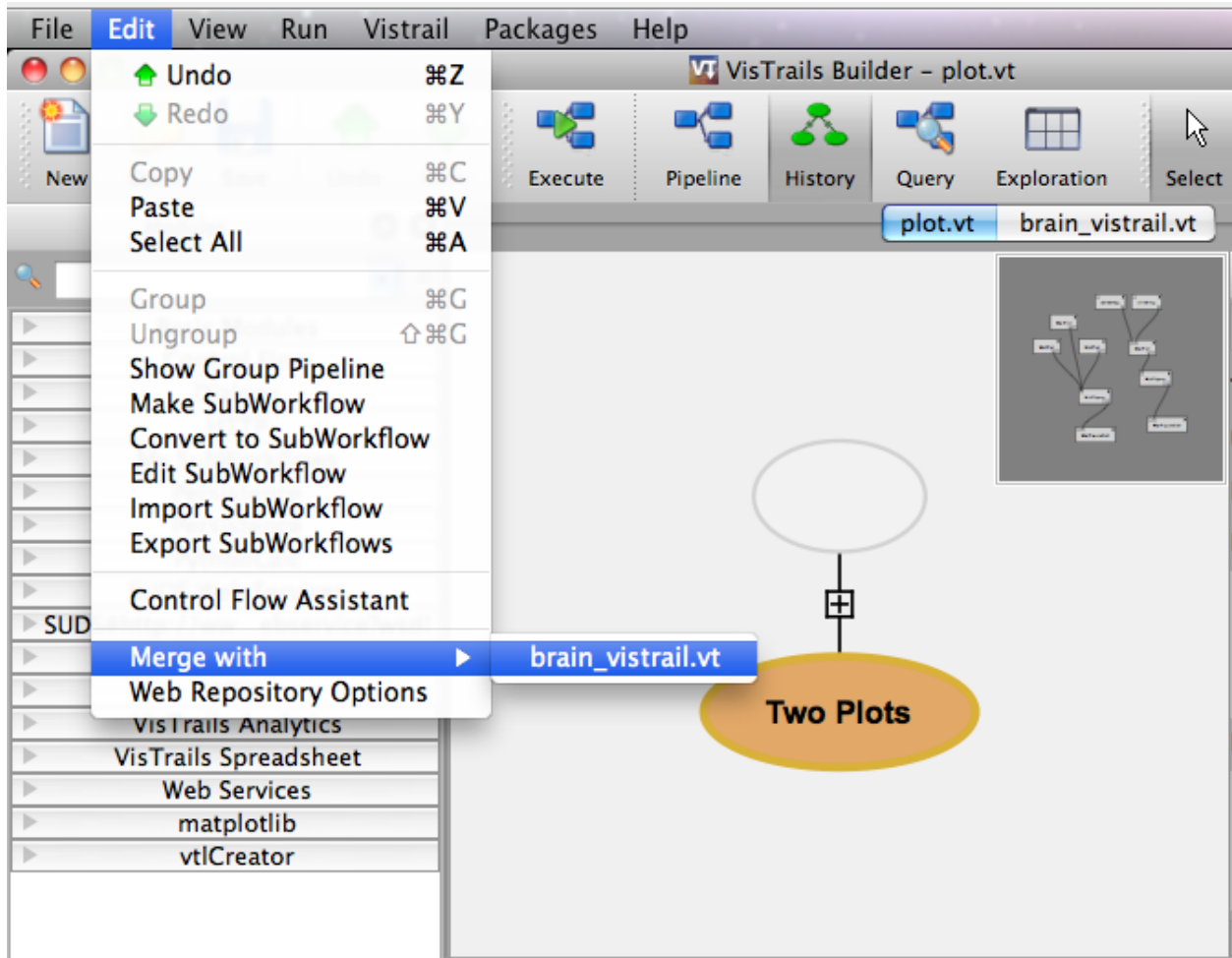


Figure 6.1: Merging two vistrails.

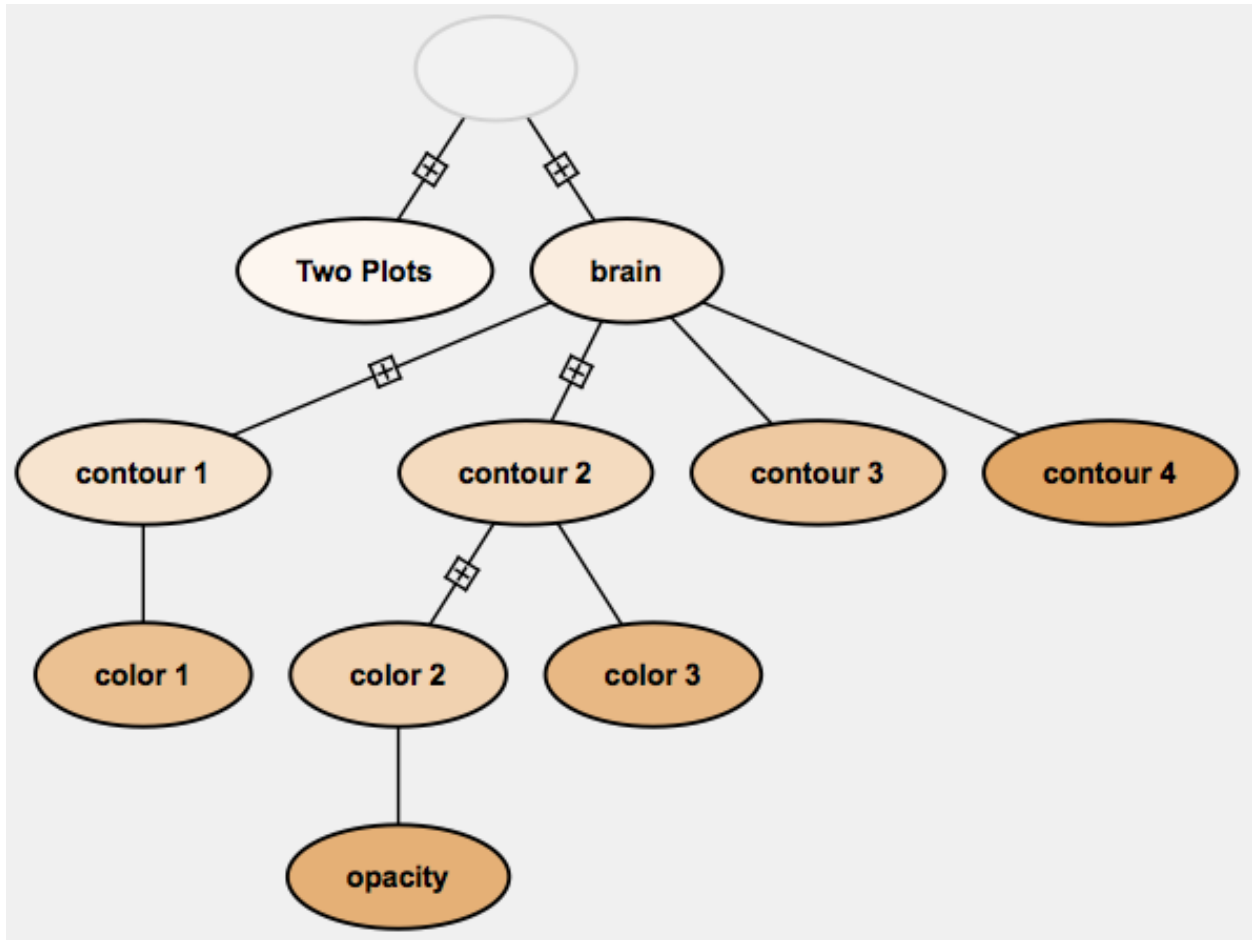


Figure 6.2: The resultant history tree.

SPREADSHEET

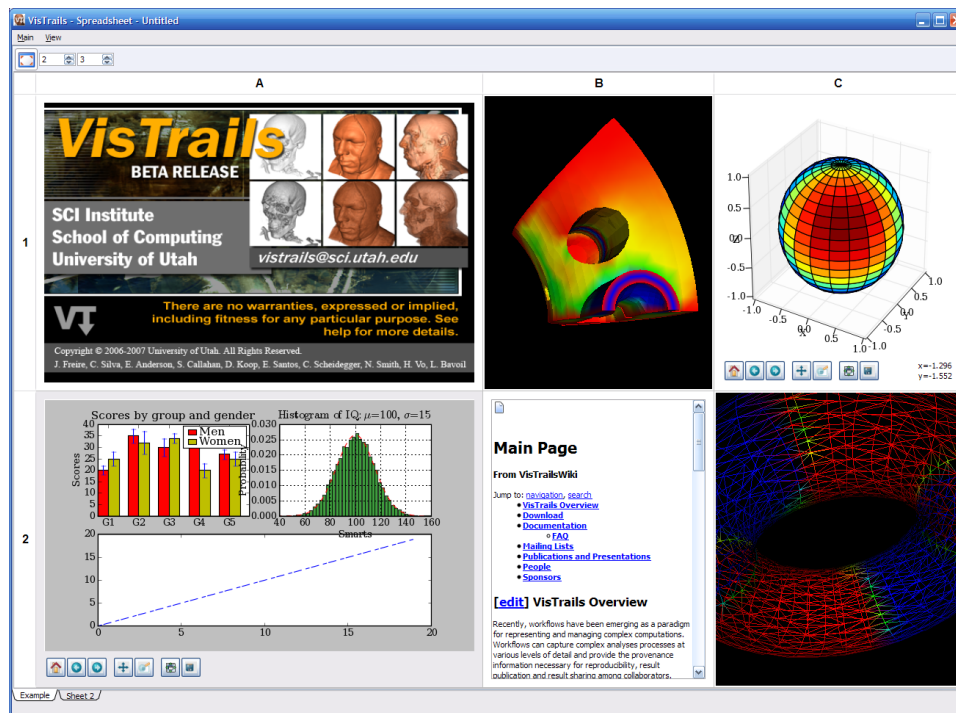


Figure 7.1: The VisTrails Spreadsheet.

As described in Section *Comparing Versions*, VisTrails has a powerful built-in mechanism to compare workflows. However, this comparison shows changes in the *design* of the workflows, and we are often also interested in differences in the *results* of workflows. The VisTrails Spreadsheet provides a simple, flexible, and extensible interface to display and compare results from workflows. Coupled with the version differences, users can explore the evolution of their workflows.

The Spreadsheet package is installed with VisTrails by default, and it can display a variety of data ranging from VTK renderings to webpages without additional configuration.

7.1 The Spreadsheet Layout

As should be expected, the VisTrails Spreadsheet consists of one or more sheets, each with a customizable number of rows and columns.

7.1.1 Custom Layout Options

To modify the layout for the active sheet, you can change both the number of rows and columns and resize individual cells. The number of rows is controlled by the left spinner in the toolbar and the number of columns by the right spinner. To resize a given row or column, click and drag on one edge of the row or column header. In addition, you can resize an individual cell by moving the mouse to lower-right corner of the cell until the cursor changes and clicking and dragging to the desired size (see Figure *Different states of a spreadsheet cell (d)*). Note that this will affect the entire layout, compressing or expanding rows and columns to generate or fill space for the resized cell. To adjust the default number of rows and columns in the spreadsheet: select `preferences` from the `VisTrails` menu, select `spreadsheet` from the `Enabled Packages` portion of the `Module Packages` tab, press the `Configure` button, and adjust the values as for `rowCount` and `colCount` as desired.

7.1.2 Multiple Spreadsheets

VisTrails supports the use of multiple spreadsheets which can be added, docked, ordered and deleted. Sheets are added either by clicking the `New Sheet` button in the `Spreadsheet` toolbar or choosing the menu item with the same name from the `Main` menu. Each of these sheets can optionally be displayed as a dock widget separated from the main spreadsheet window by dragging its tab name out of the tab bar at the bottom of the spreadsheet, allowing multiple spreadsheets to be visible at the same time. To dock a sheet back to the main window, drag it back to the tab bar or double-click on its title bar. Similarly, sheets are ordered by dragging sheet names to desired locations within the tab bar. Finally, a sheet can be deleted by clicking the red 'X' button in the lower-right corner or choosing the `Delete Sheet` menu item.

7.2 Sending Output to the Spreadsheet

Users may send results to the spreadsheet by using a spreadsheet cell. Upon inspecting the `VisTrails Spreadsheet` package (in the list of packages, to the left of the pipeline builder), one can see there are built-in cells for different kinds of data, e.g., `RichTextCell` to display HTML and plain text.

By default, an unoccupied cell on the active sheet will be chosen to display the result. However, you can specify in the pipeline exactly where a spreadsheet cell will be placed by using `CellLocation` and `SheetReference`. `CellLocation` specifies the location (row and column) of a cell when connecting to a spreadsheet cell (`VTKCell`, `ImageViewerCell`, ...). Similarly, a `SheetReference` module (when connecting to a `CellLocation`) will specify which sheet the cell will be put on given its name, minimum row size and minimum column size. There is an example of this in `examples/vtk.xml` (select the version below `Double Renderer`).

7.2.1 Advanced Cell Options

The user can define new cell types to display application-specific data. For example, we have developed `VtkCell`, `MplFigureCell`, and `OpenGLCell`. It is possible to display pretty much anything on the `Spreadsheet`!

Examples of writing cell modules can be found in: `RichTextCell`: `packages/spreadsheet/widgets/richtext/richtext.py`
`VTK`: `packages/vtk/vtkcell.py`

Here is the summary of some requirements on a cell widget:

1. It must be a Qt widget. It should inherit from `spreadsheet_cell.QCellWidget` in the `spreadsheet` package. Although any Qt Widget would work, certain features such as animation will not be available (without rewriting it).
2. It must re-implement the `updateContents()` function to take a set of inputs (usually coming from input ports of a wrapper Module) and display on the cells. `VisTrails` uses this function to update/reuse cells on the spreadsheet when new data comes in.

- It needs a wrapper VisTrails Module (inherited from `basic_widgets.SpreadsheetCell` of the spreadsheet package). Inside the `compute()` method of this module, it may call `self.display(CellWidgetType, (inputs))` to trigger the display event on the spreadsheet.

7.3 Interacting with the Spreadsheet

Currently, there are two operating modes in the Spreadsheet: Interactive Mode and Editing Mode. Interactive Mode allows users to view and interact with the spreadsheet cells, while Editing Mode provides operations for manipulating cells. The modes can be toggled via the `View` menu or their corresponding keyboard shortcuts (`'Ctrl-Shift-I'`) and `'Ctrl-Shift-E'`).

7.3.1 Interactive Mode

In Interactive Mode, users can interact directly with the viewer for an individual cell, interact with multiple cells at once, or change the layout of the sheet. Because cells can differ in their contents, interacting with a cell changes based on the type of data displayed. For example, in a cell displaying VTK data (a `VTKCell`), a user can rotate, pan, and zoom in or out using the mouse. In a sheet, a cell can be both *active* and *selected*. There can only be one active cell,

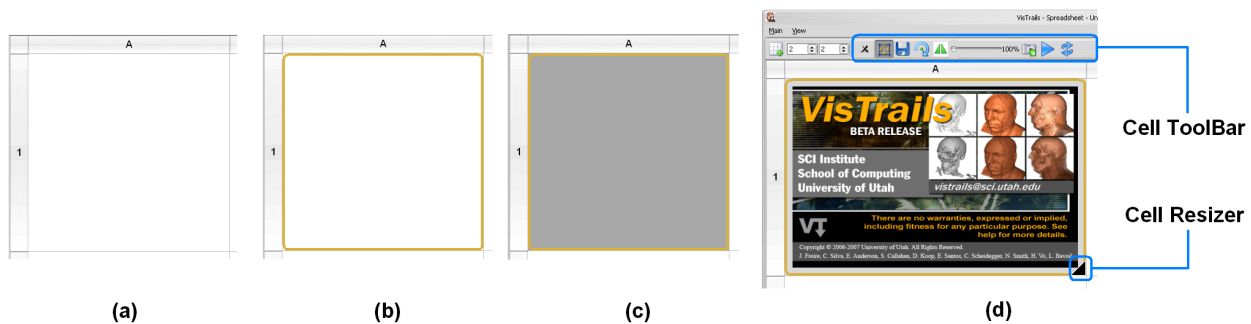


Figure 7.2: Different states of a spreadsheet cell. (a) inactive and unselected, (b) active and unselected, (c) active and selected, (d) an active cell with its toolbar and resizer.

and that cell is highlighted by a yellow or grey border. Clicking on any cell will make it active. This active cell will respond to keyboard shortcuts as well as mouse input. In contrast to the active cell, one or more cells can be selected, and the active cell need not be selected. To select multiple cells, either click on a row or column heading to toggle selection or `'Ctrl'`-click to add or remove a cell from the group of selected cells. The backgrounds of selected cells are highlighted using a platform-dependent selection color. See Figure *Different states of a spreadsheet cell* for examples of the different cell states.

Depending on the cell type, additional controls may appear in the toolbar when a cell is activated. These controls affect only the active cell, and change for different cell types. As shown by Figure *Different states of a spreadsheet cell* (d), a cell optimized for rendering 2D images (a `ImageViewerCell`) adds controls for resizing, flipping, and rotating the image in the active cell.

7.3.2 Arranging Cells

As described in Section *Custom Layout Options*, cells can be resized by either resizing rows, columns, or an individual cell. In addition to resizing, a row or column can be moved by clicking on its header and dragging it along the header bar to the desired position. See Section *Editing Mode* for instructions on moving a specific cell to a different location.

7.3.3 Synchronizing Cells

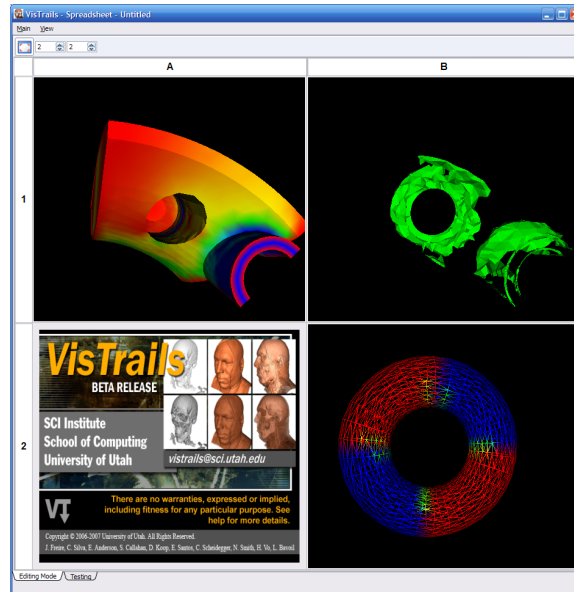


Figure 7.3: When selecting all cells, interacting with one VTK cell (A1) causes the other two VTK cells (B1 and B2) to change their camera to the same position.

Often, when a group of cells all display results from similar workflows, it is useful to interact with all of these cells at the same time. For example, for a group of `VTKCells`, it is instructive to rotate or zoom in on multiple cells at once and compare the results. For this reason, if a group of cells is selected, mouse and keyboard events for a single cell of the selection are propagated to each of the other selected cells. Currently, this feature only works for `VTKCells`, but we plan to add this to other cell types as well. An example of this functionality is shown in Figure *When selecting all cells*.

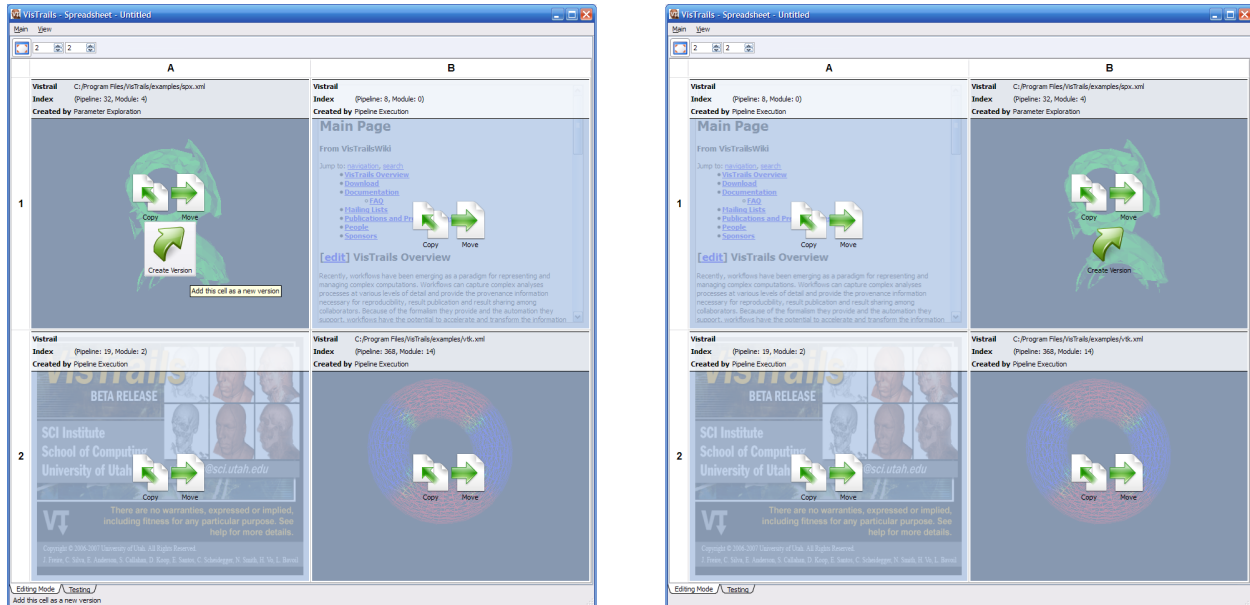
7.3.4 Editing Mode

Recall that Editing Mode can be entered either by accessing the `View` menu or by keying ‘Ctrl-Shift-E’. Editing Mode provides more operations to layout and organize spreadsheet cells. In this mode, the view for each cell is frozen and overlaid with additional information and controls (see Figure *The spreadsheet in editing mode*). The top of the overlay displays information about which vistrail, version, and type of execution were used to generate the cell. The bottom piece of the overlay contains a variety of controls to manipulate the cell depending on its state.

Cells can be moved or copied to different locations on the spreadsheet by clicking and dragging the appropriate icons (`Move` or `Copy`) for a given cell to its desired location. To move a cell to a location on a different sheet, drag the icon over the target sheet tab to bring that sheet into focus first and then drop it at the desired location. If you move a cell to an already-occupied cell, the contents of the two cells will be swapped. See Figure *The spreadsheet in editing mode* for an example of swapping two cells.

Clicking the `Locate Version` icon will highlight the node in the version tree (in the `History` view) from which the visualization in that cell was generated. The next two icons, `Create Analogy` and `Apply Analogy`, help with creating visualizations by analogy. Please refer to Chapter *Using Analogies to Update Workflows* for information about this feature.

If a cell was generated via parameter exploration (see Chapter *Parameter Exploration*), the `Create Version` button will be available to save the workflow that generated the result back to the vistrail. Clicking this button modifies the vistrail from which the cell was generated by adding a new version with the designated parameter settings. Thus, if



(a)

(b)

Figure 7.4: The spreadsheet in Editing Mode. (a) All cell widgets are replaced with an information widget (b) Two cells are swapped after drag and drop the ‘Move’ icon from A1 to B1.

you go back to the History mode of the VisTrails Builder for that vistrail, you will find that a new version has been added to the version tree.

7.4 Launching a Web Browser

It is sometimes difficult to view web pages within a spreadsheet cell due to limited space. It may therefore be desirable to launch a web browser from within the spreadsheet cell. While this functionality is not provided by VisTrails, here are some possible solutions:

1. You can use parameter exploration to generate multiple sheets so you might have an exploration that opens each page in a new sheet. Use the third column/dimension in the exploration interface to have a parameter span sheets.
2. The spreadsheet is extensible so you can write a custom spreadsheet cell widget that has a button or label with the desired link (a QLabel with openExternalLinks set to True, for example).
3. You can tweak the existing RichTextCell by adding the line “self.browser.setOpenExternalLinks(True)” at line 63 of the source file “vistrails/packages/spreadsheet/widgets/richtext/richtext.py”. Then, if your workflow creates a file with html markup text like “VisTrails” connected to a RichTextCell, clicking on the rendered link in the cell will open it in a web browser. You need to add the aforementioned line to the source to let Qt know that you want the link opened externally; by default, it will just issue an event that isn’t processed.

7.5 Saving a Spreadsheet

Warning: This is currently an experimental feature and as such is not robust. If you rename or move the vistrails used by the saved spreadsheet, the spreadsheet will not load correctly.

Because spreadsheets can include several workflow executions or parameter explorations, it is helpful to be able to save the layout of the current spreadsheet. To save a spreadsheet, simply choose the `Save` menu item from the `Main` menu, and complete the dialog. After saving a spreadsheet, you can reopen it using the `Open` menu item. A whole sheet can also be saved by selecting `Export` (either from the menu or from the toolbar).

7.5.1 Saving a Spreadsheet Image

To save an image from the spreadsheet, click on the image's cell to make it active. Then select the camera on the toolbar to take a snapshot. The system will prompt you for the location and file name where it should be saved. The other icons can be used for saving multiple images that can be used for generating an animation on demand.

QUERYING THE VERSION TREE

VisTrails is designed for manipulating collections of workflows, and an integral part of this design is the ability to quickly search through these collections. VisTrails provides two methods for querying vistrails and workflows. The first is a Query by Example interface which allows you to build query workflows and search for those with similar structures and parameters. The second is a textual interface with a straightforward syntax. For each interface, the results are *visual*: each matching version is highlighted in the `History` view, and if the query involves specific workflow characteristics, any matching entities are also highlighted in the `Pipeline` view for the current version.

8.1 Query By Example

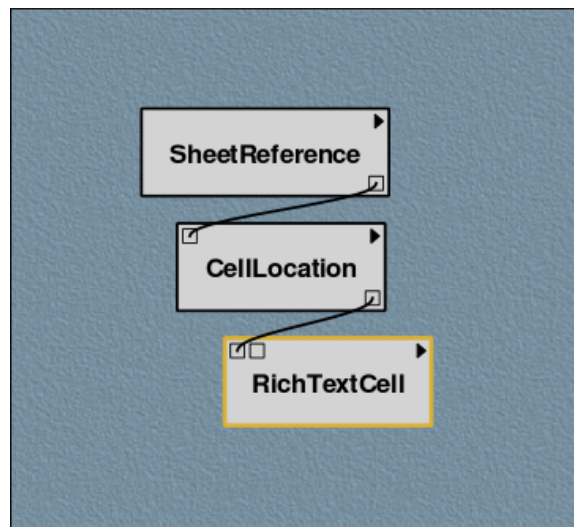


Figure 8.1: Example pipeline in `Query` mode.

One of the problems faced when trying to query a collection of workflows is the fact that structure is important. Suppose that you want to find only workflows where two modules are used in sequence. Instead of trying to translate this into a text-based syntax, it is easier to construct this relationship visually. VisTrails provides such an interface which mirrors the `Pipeline` view, allowing you to construct a (partial) workflow to serve as the search criteria.

To use the Query by Example interface, click on the `Query` button on the toolbar. This view is extremely similar to the `Pipeline` view and pipelines can be built in a similar manner. Just like the `Pipeline` view, modules are added by dragging them from the list on the left side of the window, connections are added by clicking and dragging from a port on one module to a corresponding port on another module. Figure *Example pipeline in Query mode*. shows an example pipeline that has been built in the query builder.

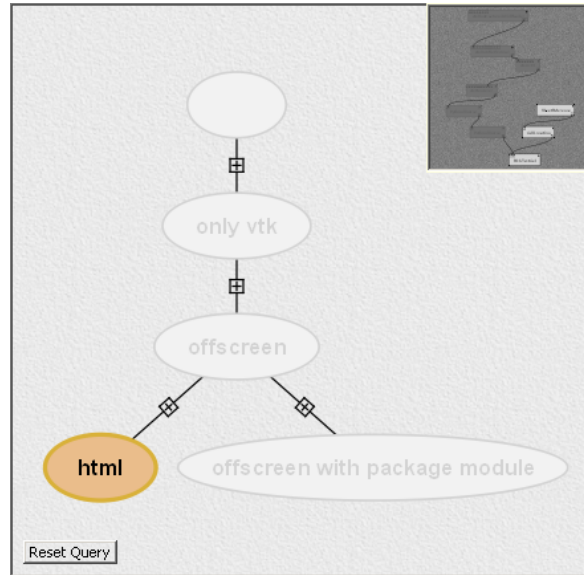


Figure 8.2: Query results in the History view.

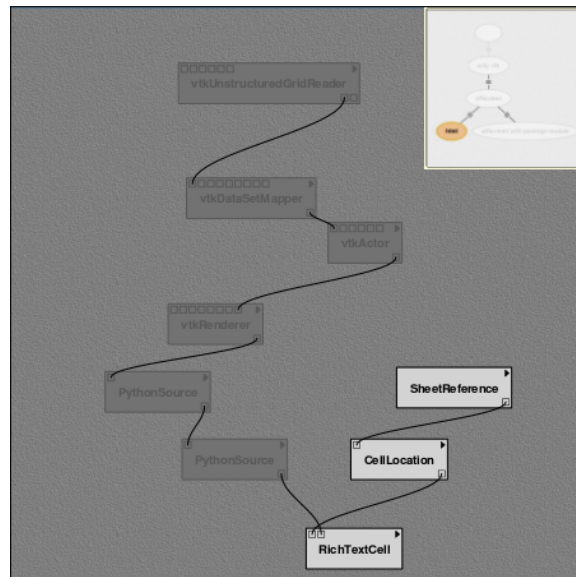


Figure 8.3: The results in the Pipeline view.

After constructing a pipeline, click the `Execute` button to begin the query. This button will be available as long as the query window is not empty. Executing the query will bring you back to the `History` view where the matching versions are displayed. Section [Query Results](#) provides information on interacting with query results.

Try it now!

Let's practice making a simple query. Open the "offscreen.vt" example vistrail. Click on the `Query` button to enter `Query` mode. Create a query like the one shown in Figure [Example pipeline in Query mode](#). by dragging the modules `SheetReference`, `CellLocation`, and `RichTextCell` onto the `Query` canvas. (These modules can be found under the "VisTrails Spreadsheet" header in the `Modules` panel.) Connect the input and output ports of the modules as shown, then click the `Execute` button to perform the query. VisTrails will automatically switch to the `History` view, with all matching versions highlighted (Figure [Query results in the History view](#)).

Note that `Query by Example` provides the capability to iteratively refine searches by adding more criteria. For example, if you were interested in workflows that contain a certain module, you may find that such a query returns too many results. You could then refine the query to find only those workflows where the given module has a parameter setting that falls within a given range. This is done by specifying parameter values in the `Methods` panel on the right side of the window. One major difference between the `Pipeline` view and the `Query` view is that you can use comparison operations, such as '`<`' and '`>`', in parameter values. The following example illustrates this.

Try it now!

Open the "terminator.vt" example file, and enter `Query` mode. Drag the `vtkActor` module from the `Modules` panel onto the `Query` canvas. `Execute` the query, and see which versions of the workflow contain a `vtkActor` modules. Return to the `Query` view, select the `vtkActor` icon, then drag the `RotateZ` method from the `Methods` panel to the `Set Methods` panel. In the `RotateZ` text field, type '`> 90`'. When you `Execute` the query this time, you will notice that the results are different. This is because we are searching for versions that not only contain a `vtkActor` module, but that also use a value greater than 90 in this module's `RotateZ` method. Your results should resemble those in Figure [Query result showing all workflows in the "terminator.vt" example that contain the module vtkActor](#).

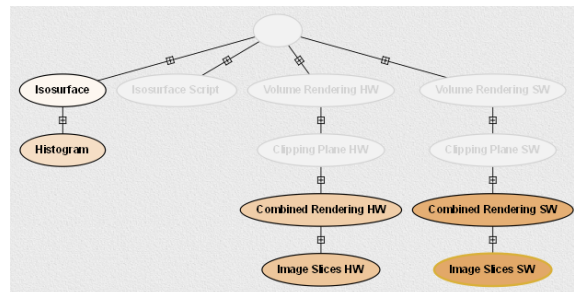


Figure 8.4: Query result showing all workflows in the "terminator.vt" example that contain the module `vtkActor`.

8.2 Textual Queries

There are many ways to search for versions in the version tree using textual queries, but they all rely on a simple text box for input. Begin a search by activating the `History` view. The search box is in the `Properties` panel, and can be identified by the magnifying glass icon next to it. If you enter query text,

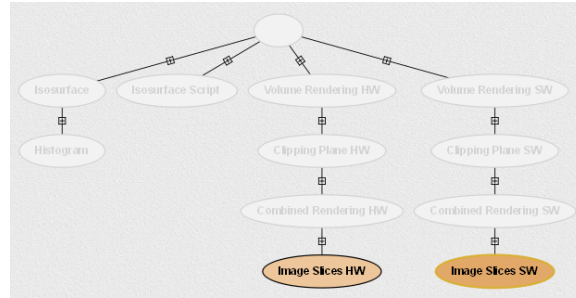


Figure 8.5: The same query refined to show only those workflows whose RotateZ value is greater than 90.

VisTrails will attempt to match logical categories, but if your query is more specific, VisTrails has special syntax to markup the query. To execute a query, simply press the ‘Enter’ key after typing your query.

Table 8.1: Syntax for querying specific information using textual queries.

Search Type	Syntax
User name	user: <i>user name</i>
Annotation	notes: <i>phrase</i>
Tag	name: <i>version tag</i>
Date	before: <i>date</i> <i>relative time</i> after: <i>date</i> <i>relative time</i>

Note

Since we allow regular expressions in our search box, question marks are treated as meta-characters. Thus, searching for “?” returns everything and “abc?” will return everything containing “abc”. You need to use “\?” instead to search for “?”. So the search for “??” would be “\?\?”.

Table *Syntax for querying specific information using textual queries*. lists the different ways to markup a query. Note that you can search by user name to see which changes a particular user has made, and also by date to see which changes were made in a specific time frame. When searching by date, you can search for all changes before or after a given date or an amount of time relative to the present. If searching for changes before or after a specific date, the date can be entered in a variety of formats. The simplest is ‘*day month year*,’ but if the year is omitted, the current year is used. The month may be specified by either its full name or an abbreviation. For example, ‘before: 18 November 2004’ and ‘after: 20 Dec’ are both valid queries. If searching by relative time, you can prepend the amount of time relative to the present including the units to ‘ago’. An example of this type of query is ‘after: 30 minutes ago’. The available units are seconds, minutes, hours, days, months, or years.

You can concatenate simple search statements to create a compound query to search across different criteria or for a specific range. For example, to search for workflows whose tag includes ‘brain’ and were created by the user ‘johnsmith’, the query would be ‘name: brain user: johnsmith’. To search for all workflows created between April 1 and June 1, the query would be ‘after: April 1 before: June 1’.

Try it now!

Open the “terminator.vt” example file, and enter History mode. Let’s look for all workflows that were created after July 1, 2007. In the search box in the Properties panel, type ‘after: 1 july 2007’ and press ‘Enter’. The expected result is shown in Figure *Results of a query to find any changes made after July 1, 2007..*

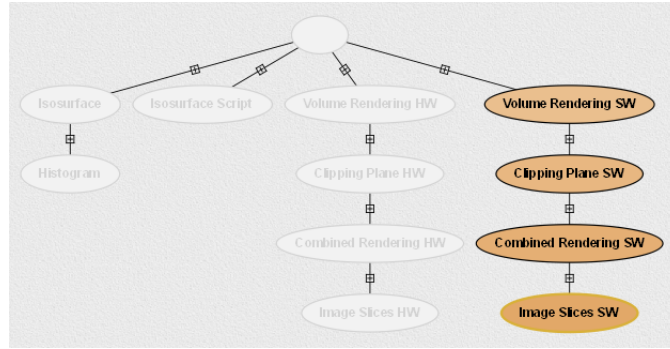


Figure 8.6: Results of a query to find any changes made after July 1, 2007.

8.3 Query Results

After executing either a query by example or a textual query, the matching versions are highlighted in the version tree. In addition, there is a button named `Reset Query` in the lower-left of the version tree that allows you to reset the query, returning the view to normal. For queries by example, if you click on a specific matching version and change to the `Pipeline` view, the matching structure will also be highlighted. Figure *Query results in the History view* shows the results of the query by example in Figure *Example pipeline in Query mode*, in both the `History` and `Pipeline` views. While in the `History` view, you can select two different ways of viewing search results. The magnifying glass icon to the left of the textual search box contains an dropdown menu with two options: “Search” and “Refine” (Figure *Clicking the button to the left of the query text box accesses a dropdown menu*). The first displays results by simply highlighting the matching nodes while the second condenses the tree to show only the versions that match. For large `vistrails`, this second method can help you determine relationships between the matching versions more easily.

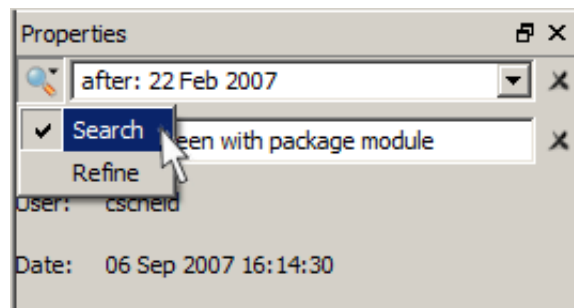


Figure 8.7: Clicking the button to the left of the query text box accesses a dropdown menu.

In addition, `VisTrails` keeps track of the most recent textual queries, and repeating these queries can be accomplished by selecting the recent query from the dropdown menu attached to the search box. You can also clear recent searches using this menu. Finally, the ‘X’ button next to the search box will reset the query and restore the normal view of the version tree.

USING ANALOGIES TO UPDATE WORKFLOWS

In Chapter *Interacting with the Version Tree*, we saw how the provenance data maintained by VisTrails allows you to compare different versions of a workflow. In Chapter *Querying the Version Tree*, you learned how this same provenance information forms the basis of an elegant Query by Example mechanism, letting you find all versions of a workflow that match a given arrangement of modules. In this chapter, we will see yet another benefit of the VisTrails provenance architecture. Through a process we call *visualization by analogy*, you can reuse pipeline information to create new visualizations semi-automatically without directly editing the workflow.

9.1 Visualization by Analogy

The main idea behind visualization by analogy is as follows. Given two versions of a workflow (called the “source” and “target” versions, respectively), VisTrails can automatically find the differences between them and apply those differences to another (potentially unrelated) workflow. This powerful feature lets you create a new visualization without having to add or remove modules to/from the pipeline. VisTrails takes care of these details for you behind the scenes.

There are two distinct user interfaces for constructing visualizations by analogy. In the first, you set up the analogy in the `Visual Diff` window. In the second, you interact with the `Spreadsheet`. Both ways are logically equivalent, and which method you use will be largely a matter of personal preference.

Before explaining either approach, however, let’s first set up the vistrail that we’ll be using as a running example in this chapter.

Try it now!

Open the “vtk_http.vt” vistrail, located in the examples directory of the VisTrails distribution. Go to the History view, and select the version labeled tetra mesh contour. This will be our “source” workflow. Execute this workflow, and take a look at the resultant visualization in the Spreadsheet.

We will now create our “target” workflow. Switch to the Pipeline view and add a new module, vtkSmoothPolyDataFilter between the vtkContourFilter and vtkDataSetMapper modules. Your modified pipeline should resemble the one shown in Figure *Modified pipeline for use in our example.*

Next, let’s adjust some of the parameters for the new module. Select the vtkSmoothPolyDataFilter module. In the Methods panel, drag the SetNumberOfIterations method and the SetRelaxationFactor method to the Set Methods panel. In the text box for SetNumberOfIterations, type 20 and in the text box for SetRelaxationFactor, type 0.5. Now, execute this workflow, and compare the two results in the Spreadsheet.

Return to the History view, and give your new version a meaningful tag such as smoothed (Figure *Corresponding version tree.*). Finally, select the Fran Cut version and execute it too. Your spreadsheet should now resemble the one shown in Figure (*top left*) *Original tetrahedral mesh model.* (*top right*) *Smoothed tetrahedral mesh.* (*bottom left*) *“Fran Cut” model.*

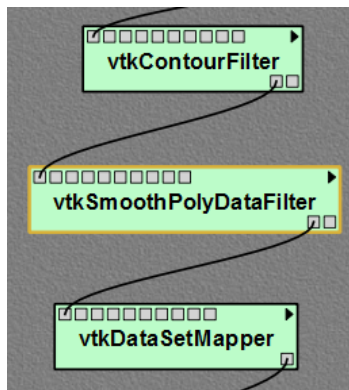


Figure 9.1: Modified pipeline for use in our example.

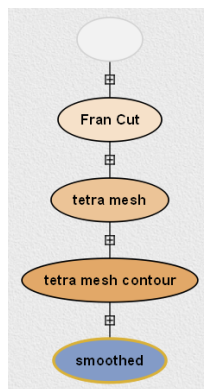


Figure 9.2: Corresponding version tree.

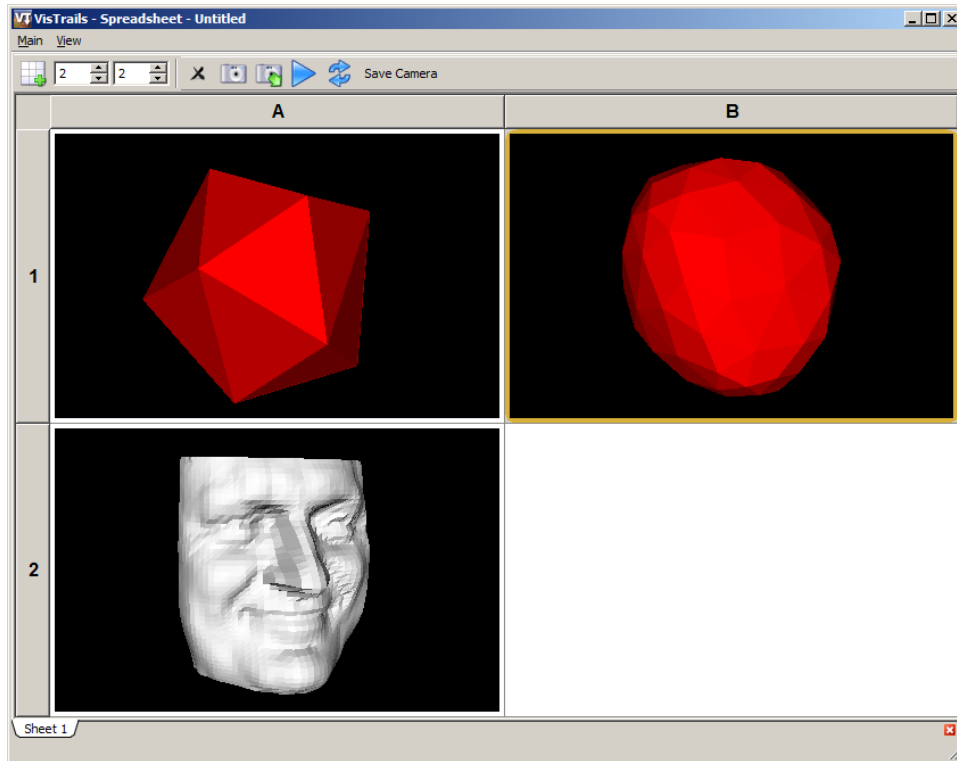


Figure 9.3: (top left) Original tetrahedral mesh model. (top right) Smoothed tetrahedral mesh. (bottom left) “Fran Cut” model.

9.2 Using the Visual Diff Window

By creating an analogy, you’re telling VisTrails to analyze and store the differences between the “source” and “target” versions of a workflow. Then when you apply that analogy to another (perhaps completely different) version of the workflow, VisTrails attempts to make similar types of changes to this other workflow.

One way to create an analogy is to run a `Visual Diff` between the “source” and “target” workflows. Recall from Chapter *Interacting with the Version Tree* that to perform a `Visual Diff` between two versions of a workflow, you need to drag the icon for one version atop the icon for the other. However, in the case of analogies, the sequence is important. In order for the analogy to work correctly, the icon for the *source* version of the workflow must be dragged atop the icon for the *target* version (not vice versa).

In the toolbar of the `Visual Diff` window, there is a button whose tooltip is labeled `Create Analogy` (Figure *Click the Create Analogy button in the Visual Diff window to create a named analogy.*). Clicking the `Create Analogy` will open up a dialog that lets you give this analogy a descriptive name. Once the analogy has a name, you can then apply it to any version of the workflow. This is done by returning to the `History` view, and selecting then right-clicking the version you want to apply the analogy to. A menu will appear, showing you a list of available analogies. Choose the one you want, and VisTrails will attempt to apply the selected analogy to this version of the workflow.



Figure 9.4: Click the Create Analogy button in the Visual Diff window to create a named analogy.

Try it now!

In the History view, drag the tetra mesh contour icon (the “source” version) atop the smoothed icon (the “target” version). A Visual Diff window will open. Click the Create Analogy button in the toolbar and then choose a name for this analogy, for example “SmoothFilter.” Close the Visual Diff window. Select the Fran Cut icon in the History view so that it is highlighted, then right-click to access the Perform analogy menu. Choose the name of the analogy you just made (Figure *Access the Perform analogy menu by right-clicking on a selected icon in the version tree.*). A new version icon will appear as a child of the current icon. Select the new icon, and click Execute to run this new version of the workflow. The resulting visualization will appear in the Spreadsheet (Figure *Result of applying a smoothing analogy to a different workflow.*).

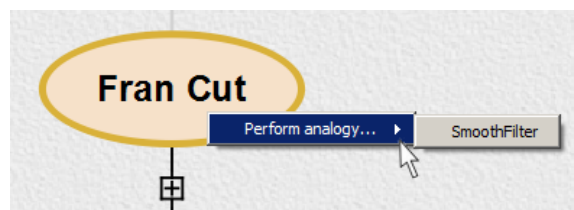


Figure 9.5: Access the Perform analogy menu by right-clicking on a selected icon in the version tree.

9.3 Using the Spreadsheet

You can also create and apply analogies directly in the Spreadsheet, without the use of the Visual Diff window. The Spreadsheet uses a simple “drag and drop” interface to manipulate analogies, and many users find it simpler to work with than the method described above. The drawback is that the Spreadsheet does not allow you to name your analogies like the Visual Diff window does.

The Spreadsheet must be in Editing Mode in order to use analogies. Please refer to Chapter *Spreadsheet* for more

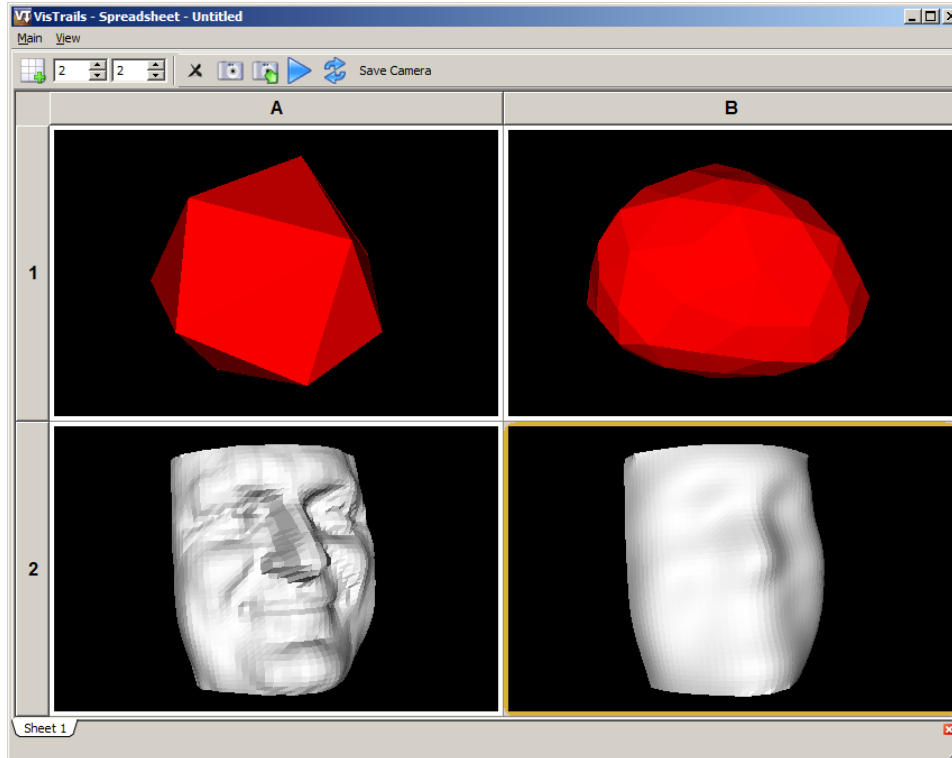


Figure 9.6: Result of applying a smoothing analogy to a different workflow.

information about the Spreadsheet’s modes. Identify the two cells representing the “source” and “target” visualizations for which you wish to create an analogy. Drag the `Create Analogy` icon *from* the “source” *to* the “target.” This creates an analogy that you can use to apply changes to other workflows. To apply an analogy to another version, drag the `Apply Analogy` icon *from* the cell containing a visualization to which you want to apply the analogy, *to* an empty cell. A new version of the workflow will be created, and rendered in the designated cell.

The following example illustrates how to use analogies within the Spreadsheet. If you completed the previous “Try it now!” exercise, first clear the cell containing the smoothed version of the `Fran Cut` model, so that it won’t interfere with the present example. The Spreadsheet should again resemble Figure (top left) *Original tetrahedral mesh model.* (top right) *Smoothed tetrahedral mesh.* (bottom left) *“Fran Cut” model.*

Try it now!

Switch to the Spreadsheet’s Editing Mode by hitting ‘Ctrl-Shift-E.’ Create the analogy by dragging the `Create Analogy` icon from the top-left cell over to the top-right cell (Figure *Drag the Create Analogy icon from the “source” cell to the “target” cell to create an analogy.*). Next, apply this analogy to the `Fran Cut` model by dragging the `Apply Analogy` icon from the bottom-left cell over to the bottom-right cell (Figure *Drag the Apply Analogy icon from the cell you wish to modify to an empty cell.*). Hit ‘Ctrl-Shift-I’ to return to Interactive Mode, and see the result of your analogy. It should resemble the output of the first example, as shown in Figure *Result of applying a smoothing analogy to a different workflow.*

Regardless of whether you use the `Visual Diff` interface or the Spreadsheet interface to create your analogy, the end result is the same. To verify this, you can inspect the `Pipeline` view for the newly created version of the workflow. All the module(s) necessary to implement the analogy’s behavior are automatically inserted by VisTrails at the correct locations in the pipeline.

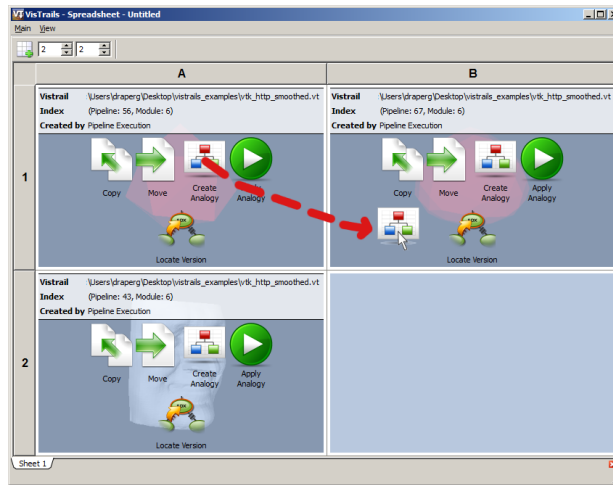


Figure 9.7: Drag the `Create Analogy` icon *from* the “source” cell *to* the “target” cell to create an analogy.

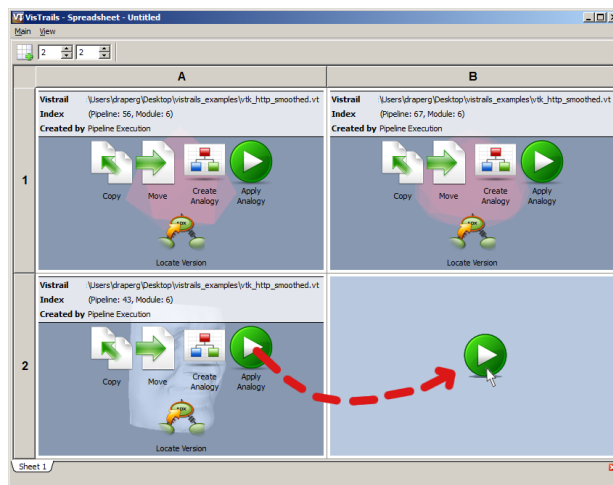


Figure 9.8: Drag the `Apply Analogy` icon *from* the cell you wish to modify *to* an empty cell.

9.4 Notes

Creating visualizations by analogy is a relatively new feature of VisTrails, and as such is not yet fully mature. Although it works well for the examples presented herein, our algorithm may fail to create intuitive visualizations on other pipelines. Furthermore, it is not yet possible to save an analogy, nor apply an analogy to a vistrail other than the one in which it was created. Please contact the VisTrails development team with any bug reports and/or suggestions.

PARAMETER EXPLORATION

While exploring workflows, one critical task is tweaking parameter values to improve simulations or visualizations. VisTrails contains an integrated parameter exploration interface that lets you thoroughly explore the parameter space and quickly identify the desired settings. By binding parameters to a range of values, you can generate a collection of results without having to tediously edit the workflow. VisTrails Parameter Exploration is Spreadsheet-aware, so you can map the intermediate results from explorations into cells of the Spreadsheet. Because the Spreadsheet provides a multi-view interface that makes efficient use of screen space, you can quickly compare the results of different parameter settings. The changes in parameters can be displayed across rows, columns, and sheets. In addition, parameters can be explored across timesteps, and then displayed in the Spreadsheet as animations. This could be used, for example, to show how pathological tissues and tumors are affected by radiation treatment in a series of scans.

10.1 Creating a Parameter Exploration

To access VisTrails Parameter Exploration for the currently-active workflow, click on the `Exploration` button in the VisTrails toolbar.

The `Parameter Exploration` view starts out with a blank central canvas wherein exploration parameters can be set up. On the right side of the window, there are a variety of panels that control aspects of the exploration.

The `Set Methods` panel contains the list of parameters that can be explored, the `Annotated Pipeline` panel displays the workflow to be explored and helps resolve ambiguities for parameter settings, and the `Spreadsheet Virtual Cell` aids users in laying out exploration results in the spreadsheet. To add parameters to an exploration, simply drag the corresponding method from the `Set Methods` panel to the center canvas. To reduce clutter, this panel only shows the methods for which parameters were assigned values in the `Pipeline` view. See [Chapter *Creating and Modifying Workflows*](#) for instructions on adding methods and parameters to a module. After dragging

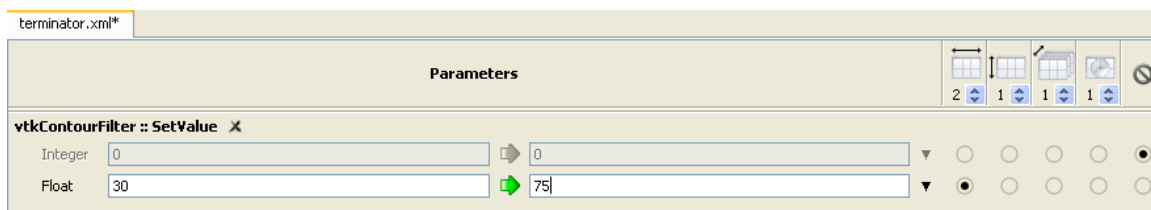
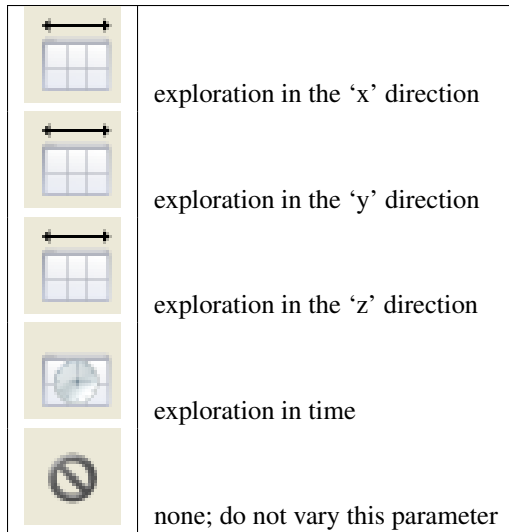


Figure 10.1: Setting values for parameter exploration.

a method to the exploration canvas, you can, for each parameter, set the collection of values to be explored and the direction in which to explore (Figure [Setting values for parameter exploration](#)). By default, intermediate parameter values are set via linear interpolation. We will discuss other options later. The five column headings in the upper-right corner of the main canvas control how the results of the parameter exploration will be displayed in the Spreadsheet. From left to right, the five controls determine:



The spinner beneath each of these icons lets you control the number of parameter values to be explored in that direction. For each parameter, you must select one of the radio buttons corresponding to a direction of exploration ('x', 'y', 'z', time, or none). Note that choosing the final column disables exploration for that parameter.

To run a parameter exploration, click the `Execute` button in the VisTrails toolbar or select `Execute Parameter Exploration` from `Run` menu.

We now reinforce the above discussion with three examples, motivated by the problem of finding isosurfaces for medical imaging. In the examples that follow, we'll look at determining the interfaces between different types of tissue captured by CT scans.

Try it now!

To begin, load the "terminator.vt" vistrail, select the "Isosurface" node in the version tree, and switch to parameter exploration. From the `Set Methods` panel, click and drag the `SetValue` method of the `vtkContourFilter` module to the center canvas.

We'd like to compare different values for the isosurfaces so change the start and end values of the "Float" parameter to "30" and "75". Since side-by-side visualization will look better on most monitors, select the radio button below the 'x' dimension control, and increase the value of the control to 2 (see Figure [Setting values for parameter exploration](#)). Execute the exploration and switch to the Spreadsheet to view the results. They should match Figure [Parameter Exploration of two isovalues as displayed in the Spreadsheet](#).

While these two isovalues show interesting features, we may wish to examine other intermediate isosurfaces. To do so, switch back to the main VisTrails window and increase the number of results to generate in the 'x' direction to four. VisTrails will calculate the intermediate values via linear interpolation, and your execution of this new exploration should match Figure [Parameter Exploration of four isovalues as displayed in the Spreadsheet](#).

In our next example, we demonstrate how multiple parameter values can be explored simultaneously. We will use both X and Y exploration directions to change the values of two parameters at the same time in the same spreadsheet.

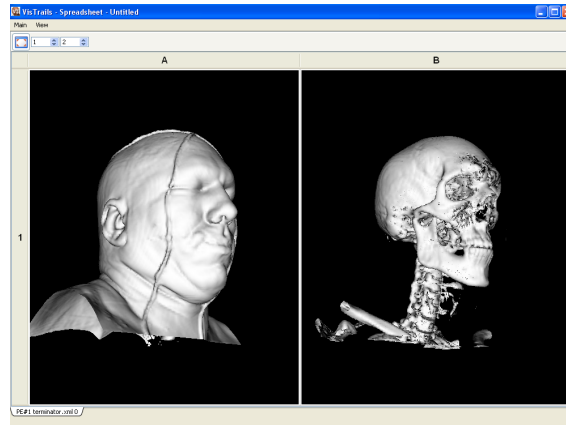


Figure 10.2: Parameter Exploration of two isovalues as displayed in the Spreadsheet.

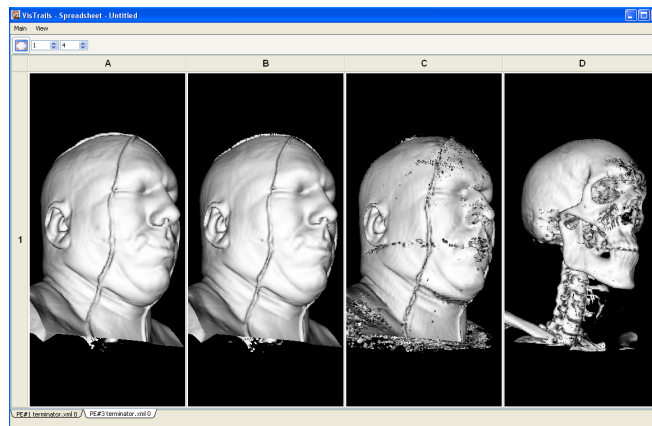


Figure 10.3: Parameter Exploration of four isovalues as displayed in the Spreadsheet.

Try it now!

In the “terminator.vt” example vistrail, make sure you’re working with the “Isosurface” version of the workflow, then go to the Pipeline view. Add the module `vtkImageResample` to the pipeline, and insert it between `vtkStructuredPointsReader` and `vtkContourFilter`, connecting the output of the reader to input of the resampler and the output of the resampler to the input of the contour filter as shown in Figure [Inserting a `vtkImageResample` module into the “terminator.vt” example pipeline.](#) Finally, select the `vtkImageResample` module and add the `SetAxisMagnificationFactor` method with parameter values 0 and 0.2. See Chapter [Creating and Modifying Workflows](#) for reminders on how to accomplish these tasks. After modifying the workflow, switch back to the Exploration view. Inside the Set Methods panel, select the `SetValue` method from the `vtkContourFilter` module, and drag it to the center canvas. Also select the `SetAxisMagnificationFactor` method from the `vtkImageResample` module and drag it to the canvas as well. Set the values as in the previous example, but set the range of the “Float” parameter of “`SetAxisMagnificationFactor`” to start at 0.2 and end at 1.0. Also, set the magnification factor to vary over the ‘y’ direction. Finally, set the exploration to generate 16 results, four in the ‘x’ direction, and four in the ‘y’ direction. Your exploration setup should match Figure [Setting up parameter exploration.](#), and after executing, you should see a result that resembles Figure [Resulting spreadsheet.](#) Notice that the isosurface changes from left to right while the images have less artifacts as the magnification factor approaches 1.0 from top to bottom.

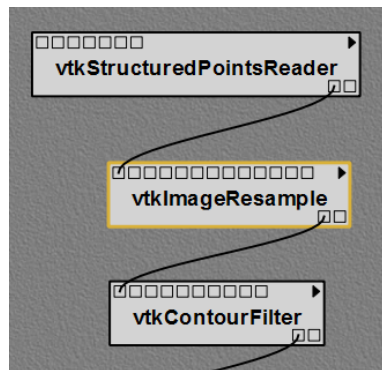


Figure 10.4: Inserting a `vtkImageResample` module into the “terminator.vt” example pipeline.

Our third example shows how to create an animation by exploring parameter values in *time*, rather than in ‘X’ or ‘Y’.

Try it now!

To create an animation, we’ll use the same “terminator.vt” example (make sure that you have the “Isosurface” version selected). Follow the same steps as in the first example, but this time, use the range from 30 to 80 and select “time” as the dimension to explore, setting the number of results to generate to 7. See Figure [Setting up parameter exploration](#) to check your settings. After executing, the Spreadsheet will show a *single* cell, but if you select that cell, you will be able to click the Play button in the toolbar. You should see an animation where each frame is the result of choosing a different isovalue. A sample frame is displayed in Figure [One frame from the resulting animation.](#)

10.2 Alternatives to Linear Interpolation

In each of the examples above, we used linear interpolation to vary the parameter values in ‘X’ and ‘Y’ and time. However, linear interpolation is only one of three methods for exploring a range of parameter values. The other two are to iterate through a simple list of values, or use a user-defined function. You can choose the desired method from

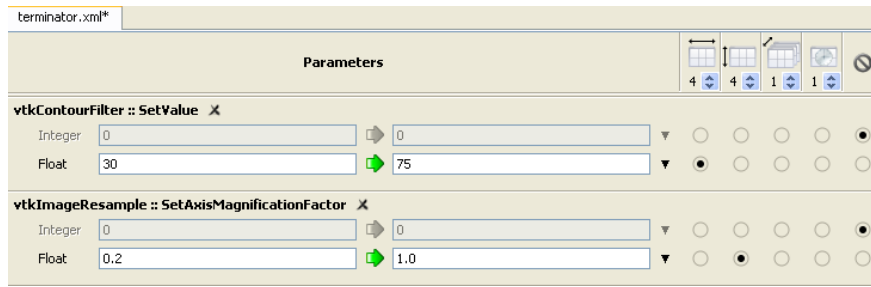


Figure 10.5: Setting up parameter exploration.

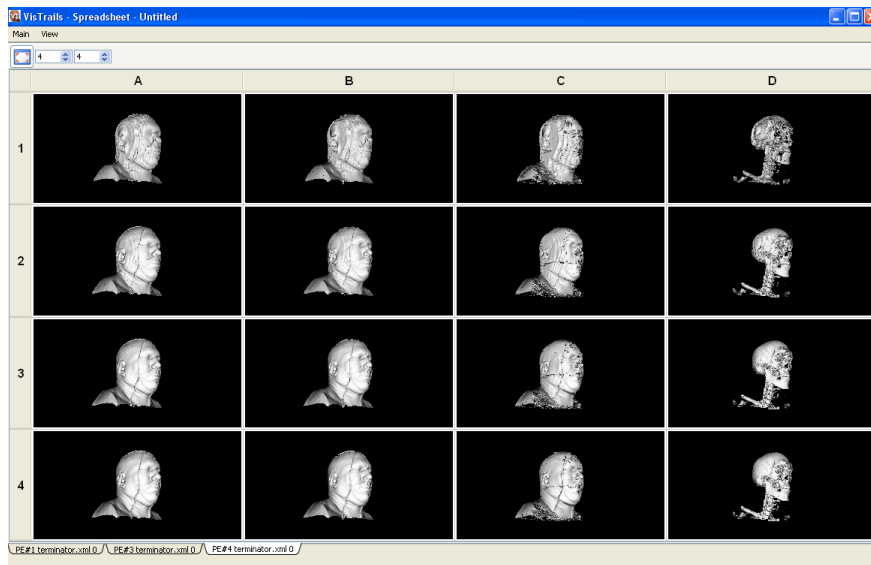


Figure 10.6: Resulting spreadsheet.
Using parameter exploration with two parameters.

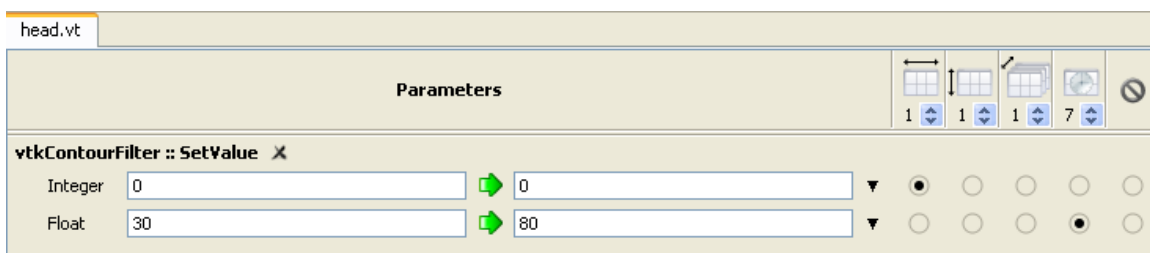


Figure 10.7: Setting up parameter exploration

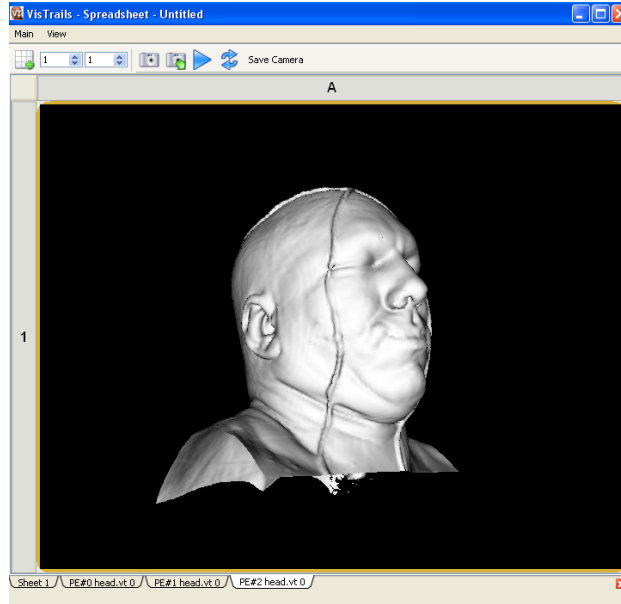


Figure 10.8: One frame from the resulting animation
Animations with parameter exploration

the drop-down menu on the right side of the parameter input field (Figure *Choose from linear interpolation, list, or user-defined function.*). For linear interpolation, the starting and ending values must be specified; for a list, the entire comma-separated list must be specified, and for a user-defined function, a Python function must be specified. For the list and user-defined functions, you can access an editor via the ‘...’ button. (See Figure *Editors for lists of values.* for examples of the list editor and Python editor widgets.) As an alternative to the list editor, you can manually enter a list using Python notation; for example, `[30, 36, 45, 75]`. As before, to set the direction in which to explore a given parameter, simply select the radio button in the column for the specified direction.

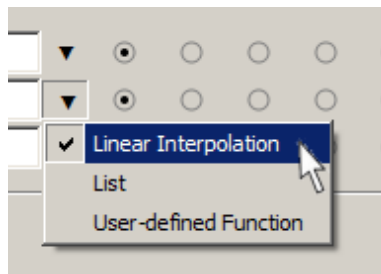


Figure 10.9: Choose from linear interpolation, list, or user-defined function.

In both the `Set Methods` and `Annotated Pipeline` panels, you may see numbered red circles. See Figure *The right panels of the Parameter Exploration window. The numbered red circles distinguish duplicate modules, and the cells in (c) determine the layout for spreadsheet results.* for an example of what this looks like. These circles appear when there is more than module of a given type in a workflow. For each type satisfying this criteria, the instances are numbered and displayed so that you can identify which part of the pipeline a module in the `Set Methods` panel corresponds to.

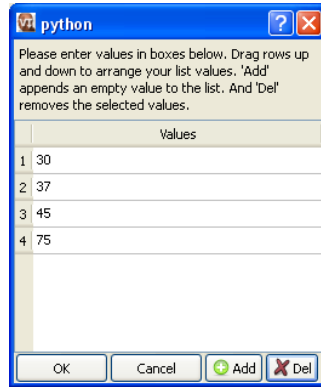


Figure 10.10: Editors for lists of values.

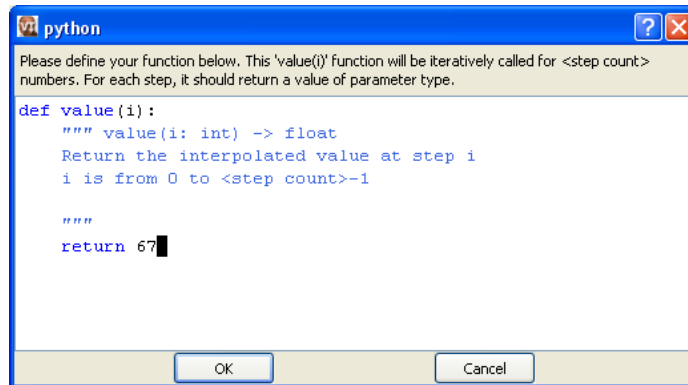
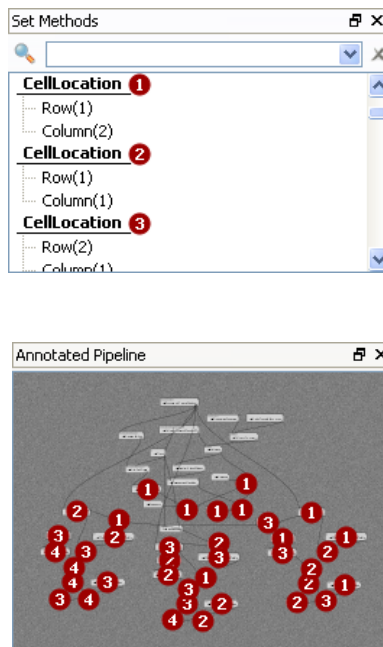


Figure 10.11: Editors for user-defined functions.



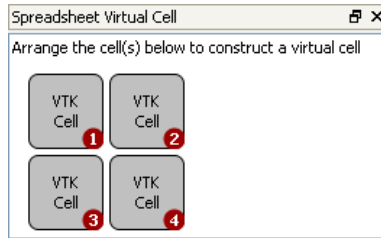


Figure 10.12: The right panels of the Parameter Exploration window. The numbered red circles distinguish duplicate modules, and the cells in (c) determine the layout for spreadsheet results.

10.3 “Virtual Cell” Layout

As stated earlier, the Spreadsheet provides integrated support for parameter explorations. Each of the directions of exploration corresponds to a visual dimension in the spreadsheet: the ‘x’ direction corresponds to columns; the ‘y’ direction to rows; the ‘z’ direction to sheets; and time to animations. However, when a workflow already outputs to more than one cell, you can layout the group of cells as it will be replicated during the exploration. For example, given a workflow with two output cells and an exploration for three parameter values in the ‘x’ direction, the resulting spreadsheet could be 1×6 or 2×3 . The Spreadsheet Virtual Cell panel controls the layout of the pattern. Drag and drop cells to position them. See Figures *The right panels of the Parameter Exploration window*, *The numbered red circles distinguish duplicate modules, and the cells in (c) determine the layout for spreadsheet results*, and *Results of the Virtual Cell arrangement*, for an example.

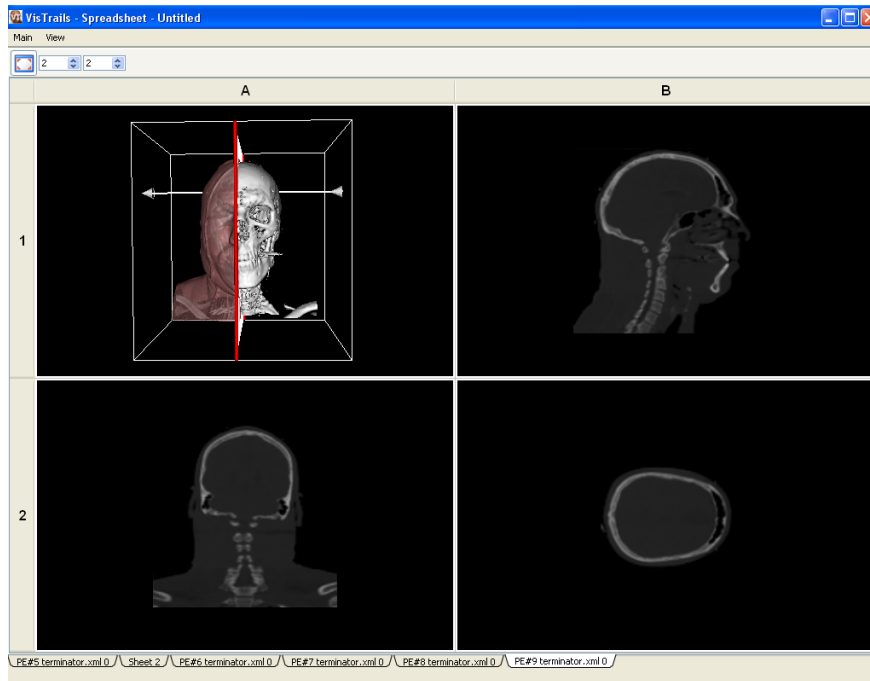


Figure 10.13: Results of the Virtual Cell arrangement.

Part IV

Intermediate Concepts and VisTrails Packages

CONTROL FLOW IN VISTRAILS

Scientific workflows usually follow a dataflow model, but, in some cases, control structures, including loops and conditionals, are necessary to accomplish certain tasks. VisTrails provides the `Control Flow` package to support these and other structures.

11.1 The Map operator

In functional programming, `map` is a high-order function that applies a given function to a list (each element of the list is processed using this function) and returns a sequence of results. The `Map` module provides this functionality for workflows in VisTrails. Note that this module provides simple looping as it can be used to iterate through a list of inputs.

In order to use the `Map` module, a `ListOfElements` type representing a list of data structure is also provided. Two additional modules help users combine elements of lists: `Dot` combines the elements of two lists like the dot product (the first element of the first list is combined with the first of the second one, the second element of the first list is combined with the second of the second one, and so on), and `Cross` combines the elements as the cross product (all the possible combinations between the elements of the lists); both modules return a list of tuples. In addition to two lists of elements as input ports, they take an optional boolean input named “CombineTuple”. This input is useful when one or both input lists have tuples as elements; if this port is selected, and its value is `False`, the elements of the list will not be combined in just one tuple, e.g., $(1, 2) + 3 \rightarrow ((1, 2), 3)$; otherwise, the elements will be combined, e.g., $(1, 2) + 3 \rightarrow (1, 2, 3)$.

The `Map` module has four input ports:

- “FunctionPort”: this port receives the module (via the “self” output port) that represents the function to be applied for each element of the input list; if the function uses more than one module, you must use a `Group` (see Chapter *Creating and Modifying Workflows*) or a `SubWorkflow` and connect that composite module to this port;
- “InputPort”: this port receives a list of the names of the input ports that represent the individual arguments of the function;
- “OutputPort”: this port receives the name of the output port that represents the individual result of the function;
- “InputList”: this port receives the input list for the loop; it must be a list of tuples if more than one function input port was chosen.

The output port “Result” produces a list of results, one for each element in the input list.

To better show how to use the `Map` module, let’s use a workflow as an example. Inside the “examples” directory of the VisTrails distribution, open the “triangle_area.vt” vistrail. Now, select the “Surface Area” version. This version basically calculates the area of a given isosurface. We are going to modify this version, in order to calculate the areas of the isosurface given by contour values in a list. Then, we will create a 2D plot to show all the areas.

Begin by deleting the `StandardOutput` module, and the connection between the `vtkDataSetReader` and the `vtkContourFilter` modules. Then, drag the following modules to the canvas:

- `Map`
- `ListOfElements`
- `Cross`
- `MplPlot` (under “matplotlib”)
- `MplFigure` (under “matplotlib”)
- `MplFigureCell` (under “matplotlib”)
- `InputPort` (under “Basic Modules”) - you will need two of them
- `OutputPort` (under “Basic Modules”)
- `PythonSource` (under “Basic Modules”)

Notice that when you drag `Map` to the pipeline canvas it will be drawn in a different shape from the other modules. This is a visual cue to help distinguish control modules from other modules. All control modules have the same shape.

Select the `vtkContourFilter` module and delete its method “`SetValue`” in the `Set Methods` container. Then, open its configuration dialog (you can use the “`Ctrl-E`” or “`Command-E`” keyboard shortcut) and enable this method (the input port “`SetValue`”) by clicking on it, and pressing `OK`.

Then, connect the modules as shown in Figure *Connecting a subset of the modules to be grouped as a SubWorkflow*.

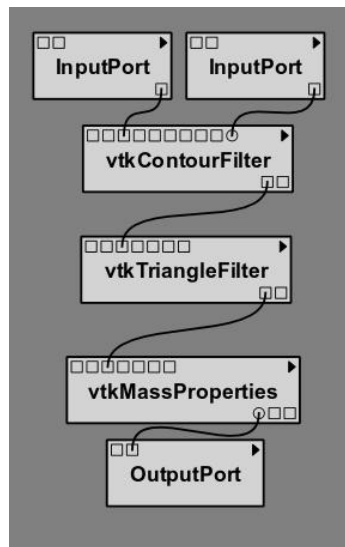


Figure 11.1: Connecting a subset of the modules to be grouped as a `SubWorkflow`

These modules represent the function we wish to map: each element of the input list will be processed using them. Because we have more than one module, we need to create a `Group` or a `SubWorkflow` to identify the entire function. The `InputPort` and the `OutputPort` modules are necessary to expose these ports in the `Group/SubWorkflow` structure.

In this example, we will use a `SubWorkflow` structure. Select all the modules shown in Figure *Connecting a subset of the modules to be grouped as a SubWorkflow*, go to the `Edit` menu, and then click on `Make SubWorkflow`. You can name it `CalculateArea`. Select this `SubWorkflow` and open its configuration. When the configuration dialog opens, enable the output port “`self`” and press `OK`. You will need this port to connect to the `Map` module.

Note

When using `Map`, the module (or subworkflow) used as function port in the map module **MUST** be a function, i.e., it can only define 1 output port.

Now, select the `MplPlot` module and open its configuration dialog. Inside it, add two input ports of type `ListOfElements`: “InputList” and “X_Values”. Also, copy the code listed below, in order to create the necessary information for the 2D plot, into the source text area and save your changes using the OK button.

```
subplot (212)

dashes = [1, 3]
list1 = self.getInputFromPort("InputList")
list2 = self.getInputFromPort("X_values")
list3 = []

for i in xrange(len(list1)):
    list3.append(list2[i][1])

l, = plot(list3, list1, marker="o", markerfacecolor="red",
          markersize=7, label="IsoSurface Areas", linewidth=1.5)

l.set_dashes(dashes)
```

Next, edit the `PythonSource` module by adding an output port “List” of type `ListOfElements`, copying the following code to the source text area, and saving these changes. The code will create a range of contour values that we will use as our input list.

```
result = []

for i in xrange(4, 256, 4):
    result.append(i)

self.setResult("List", result)
```

It may be helpful to identify this `PythonSource` module by labeling it as `RangeList`. Connect all the modules as shown in Figure *All the modules connected in the canvas*.

You will set some parameters now:

- `HTTPFile`: set the parameter “url” to <http://www.sci.utah.edu/~cscheid/stuff/head.120.vtk>
- `ListOfElements`: set the parameter “value” to `[0]`
- `Map`: set the parameter “InputPort” to `["SetValue"]` and the parameter “OutputPort” to `GetSurfaceArea`

The workflow is now ready to be executed. When you execute the workflow, you will see the `SubWorkflow CalculateArea` executing several times (one time for each value of the input list). It’s important to notice that, although only the module connected to `Map` (in this example, the `SubWorkflow Calculate Area`) will be in the loop, the modules above it will be also used in each loop iteration; the difference is that they are going to be executed only for the first iteration; in all other iterations, the results will be taken from the cache.

When the workflow finishes its execution, the VisTrails Spreadsheet will contain a 2D plot (Figure *The result in the VisTrails Spreadsheet*).

This example can be found in the version “Surface Area with Map”, inside the “triangle_area.vt” vistrail.

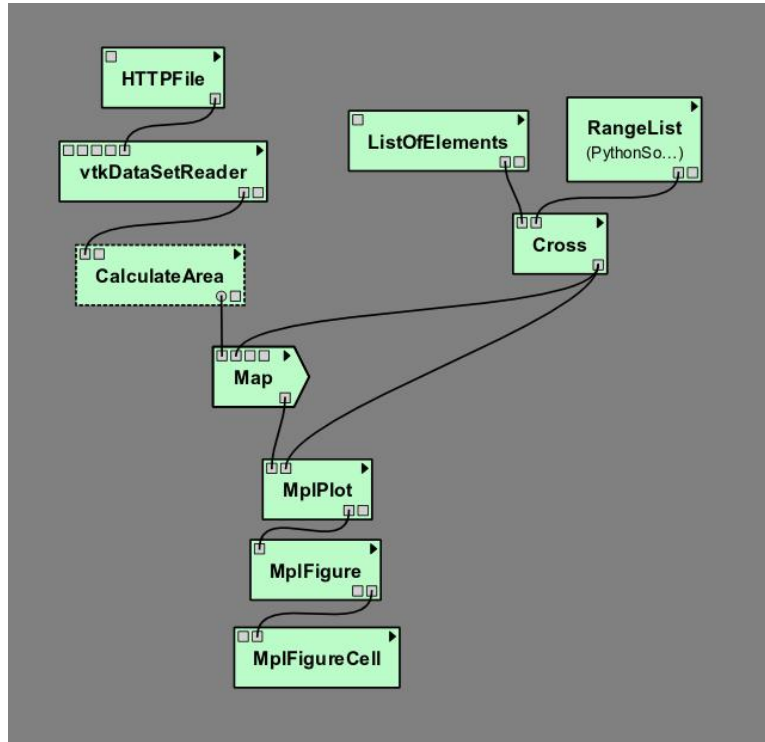


Figure 11.2: All the modules connected in the canvas

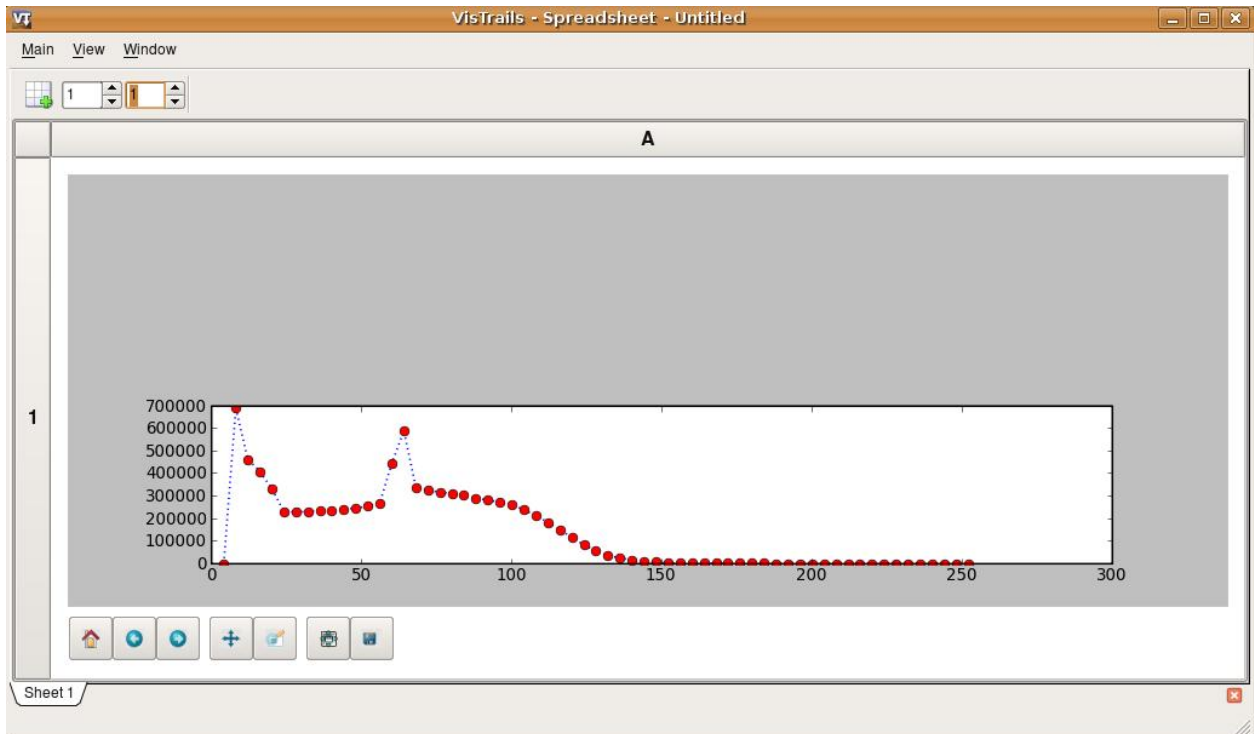


Figure 11.3: The result in the VisTrails Spreadsheet

11.2 Filtering results

When computing large lists of results, it can be useful to selectively reduce the list during execution to avoid unnecessary computation.

The `Filter` module was developed to address this issue. It receives an input list and, based on a specified boolean condition, returns only elements of the list that satisfy the condition. Its ports are the same as those in the `Map` module. The difference between these modules is related to the structure: in `Filter`, the output port “FunctionPort” cannot receive any function, but only a condition; in other words, the selected port in “OutputPort” must return a boolean value or `Filter` will not work. Thus, `Filter` will not return a list with boolean values as `Map` would do, but rather the elements of the input list for which the condition evaluated to `True`.

To better understand how `Filter` works, let’s modify our earlier example to filter out areas less than 200,000. With the previous `vistrial` open (you can use the “Surface Area with `Map`” version), add the following modules to the canvas:

- `Filter`
- `PythonSource` (under “Basic Modules”)

Edit the configuration of `PythonSource` by adding an input port of type `Float` named “Area”, and an output port of type `Boolean` named “Condition”, and writing the following code inside the source text area:

```
area = self.getInputFromPort("Area")

if area>200000.00:
    self.setResult("Condition", True)
else:
    self.setResult("Condition", False)
```

Press the OK button. You can label this `PythonSource` as `FilterCondition`. Now, reorganize the modules in the canvas as shown in Figure *The new organization of the modules in the canvas*.

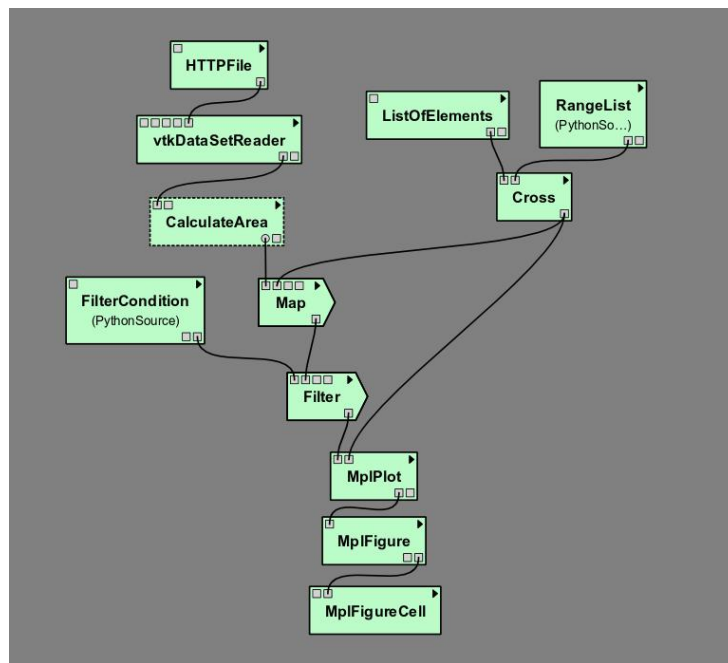


Figure 11.4: The new organization of the modules in the canvas

Select the `Filter` module and set the values of its parameters to the following:

- “InputPort”: [”Area”]
- “OutputPort”: *Condition*

When you execute this workflow, it will generate another plot that is similar to the one from the Map example, but only areas above 200,000 are considered (Figure *The result in the VisTrails spreadsheet*).

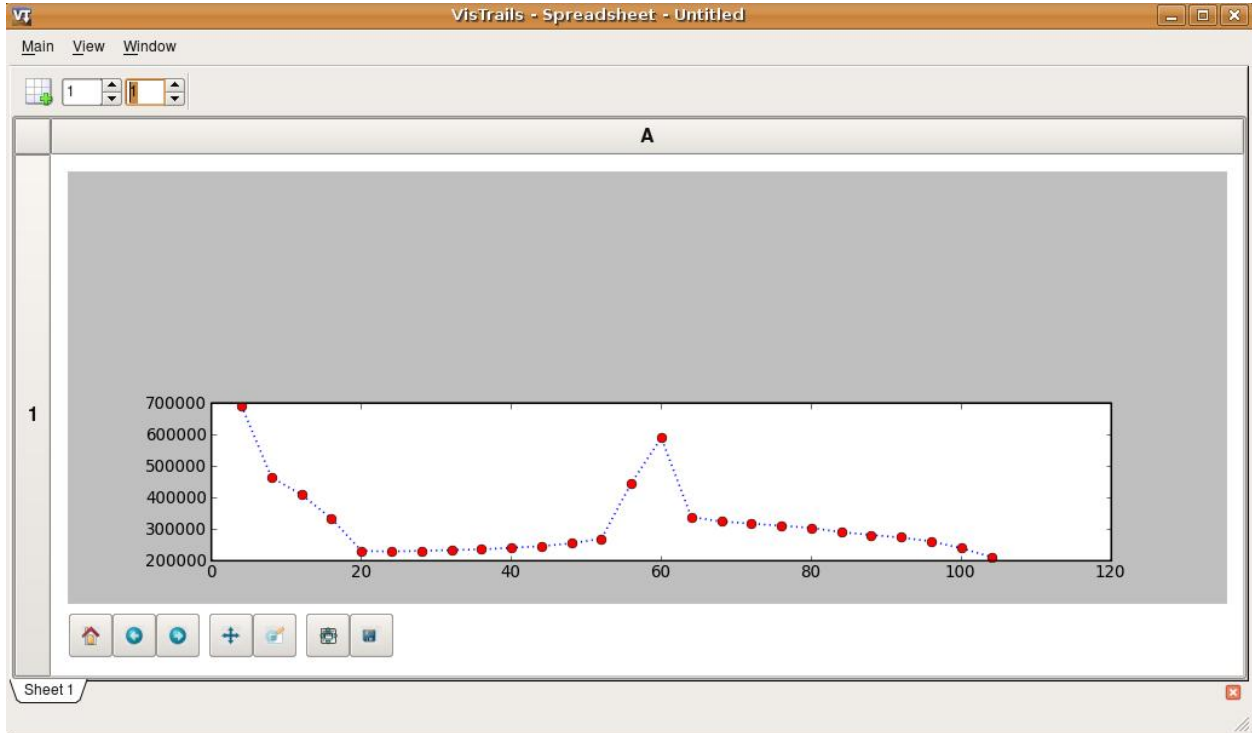


Figure 11.5: The result in the VisTrails spreadsheet

This example is already inside the “triangle_area.vt” vistrail, in the “Surface Area with Map and Filter” version.

Later in this chapter, you will see how to combine Map and Filter in one single module, by creating your own control structure.

11.3 Conditional module

Conditional statements are a very important control flow structure frequently used in programming languages, and the `if` structure is probably the most common of these structures. In scientific workflows, for example, an `if` structure can be used to select the part of the pipeline to be executed based on a boolean condition.

For this reason, the `Control Flow` package also includes an `If` module. Its input ports are:

- “Condition”: this port receives a boolean value that will specify the direction of computation;
- “TruePort”: this port receives the part of the workflow that will be executed if the condition value is `True`; you don’t need to group or make a `SubWorkflow` in this case: just connect the output port “self” of the last module in this port;
- “FalsePort”: this port receives the part of the workflow that will be executed if the condition value is `False`; as with the
- “TruePort” port, you don’t need to group or make a `SubWorkflow`;

- “TrueOutputPorts”: this port receives a list that contains the names of the output ports of the module connected to “TruePort” that you want the result of; this port is optional;
- “FalseOutputPorts”: this port receives a list that contains the names of the output ports of the module connected to “FalsePort” that you want the result of; this port is optional.

The `If` module has an output port named “Result” that returns a list with the results of the specified output ports of “TrueOutputPorts” or “FalseOutputPorts”, depending on the condition value. If only one output port is chosen, the result of this port will not be returned in a list. If “TrueOutputPorts” or “FalseOutputPorts” are not enabled, “Result” returns `None`.

Let’s do now a simple example to show how exactly this module works. This example is from the bioinformatics domain, and takes a string as the input; if it’s a structure identifier, a web service from the European Bioinformatics Institute, or simply EBI (<http://www.ebi.ac.uk/>), a centre of researches in bioinformatics, is used to get the structure in the PDB format, a standard representation for macromolecular structure, and then, the `VTK` package is used to show the protein in the VisTrails Spreadsheet; otherwise, the input is assumed to be invalid, and a message is generated in the Spreadsheet too.

First, the EBI’s web service must be enabled. For this, you need to add the following url to the `wSDLList` configuration:

```
http://www.ebi.ac.uk/Tools/webservices/wSDL/WSDbfetch.wSDL
```

Don’t forget to ensure that the `webServices` package is enabled in the Preferences dialog. For more information about web services in VisTrails, see Chapter *Example: Web Services*.

Now, you’re going to drag the following modules to the canvas:

- `If`
- `fetchData` (under “Methods” for the current web service)
- `vtkPDBReader` (under “VTK”)
- `vtkDataSetMapper` (under “VTK”)
- `vtkActor` (under “VTK”)
- `vtkRenderer` (under “VTK”)
- `VTKCell` (under “VTK”)
- `PythonSource` (under “Basic Modules”) - you will need three of them
- `String` (under “Basic Modules”)
- `RichTextCell` (under “VisTrails Spreadsheet”)

Select one of the `PythonSource` modules, and open its configuration dialog. Inside it, add one input port of type `String`, named “PDB_format”, and one output port of type `File`, named “File”. Then, write the following code:

```
1 PDB_format = self.getInputFromPort('PDB_format')
2
3 output = self.interpreter.filePool.create_file()
4 file_ = open(str(output.name), 'w')
5 file_.write(PDB_format)
6
7 self.setResult('File', file_)
8
9 file_.close()
```

You can name this module as `CreateFile`. Now, set some parameters of `fetchData`:

- “format”: `pdb`

- “style”: *raw*

Next, connect some modules as shown in Figure *Some modules of the workflow connected*.

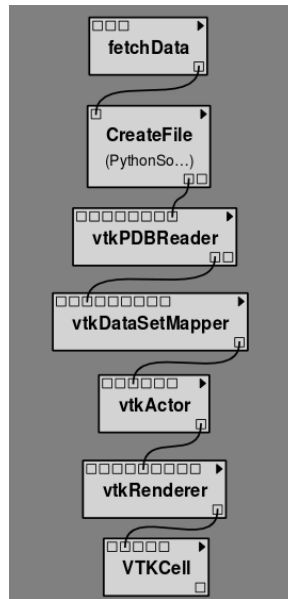


Figure 11.6: Some modules of the workflow connected

The aim of this group of modules is to get the PDB format of the structure ID, through the web service, and then make the visualization with the VTK package. This is the part of the workflow that will be executed if the input is a structure identifier.

Next, select another `PythonSource` module and open its configuration dialog too. One input port named “Structure”, of type `String`, and one output port named “Is_ID”, of type `Boolean`, must be added, as well as the code below:

```

1  if "\n" in structure:
2      lineLen = structure.index("\n")
3  else:
4      lineLen = -1
5  if lineLen < 1:
6      lineLen = len(structure)
7
8  if ":" in structure:
9      index = structure.index(":")
10 else:
11     index = -1
12
13 if (structure[0] != "ID ") and (index > 0) and (index < lineLen):
14     is_ID = True
15 else:
16     is_ID = False
17
18 self.setResult("Is_ID", is_ID)

```

Name this module as `Is_ID`. This module will be the condition for the `If` structure.

Now, select the last `PythonSource` module, and, inside its configuration, add one input port of type `String`, named “Input”, and one output port of type `File`, named “html”. Then, copy the code below:

```

1 input = self.getInputFromPort("Input")
2
3 output = self.interpreter.filePool.create_file()
4 f = open(str(output.name), 'w')
5 text = '<HTML><TITLE>Protein Visualization</TITLE><BODY BGCOLOR="#FFFFFF">'
6 f.write(text)
7 text = '<H2>Protein Visualization Workflow</H2>'
8 f.write(text)
9 text = '<H3>The following input is not an ID from a protein:</H3>'
10 text += '<H4>' + str(input) + '</H4>'
11 text += '<H3>The visualization cannot be done.</H3>'
12 f.write(text)
13
14 text = '</BODY></HTML>'
15 f.write(text)
16
17 self.setResult('html', f)
18
19 f.close()

```

Name this PythonSource as Not_ID. This module will print a message in the VisTrails Spreadsheet when the input is not a structure identifier.

Finally, the String module can be named as Workflow_Input, to make it clear that it takes the input of the workflow. Also, open the configuration dialog of RichTextCell to enable the output port “self”, so it can be connected to the If module. Then, connect all the modules as shown in Figure *All the modules connected*.

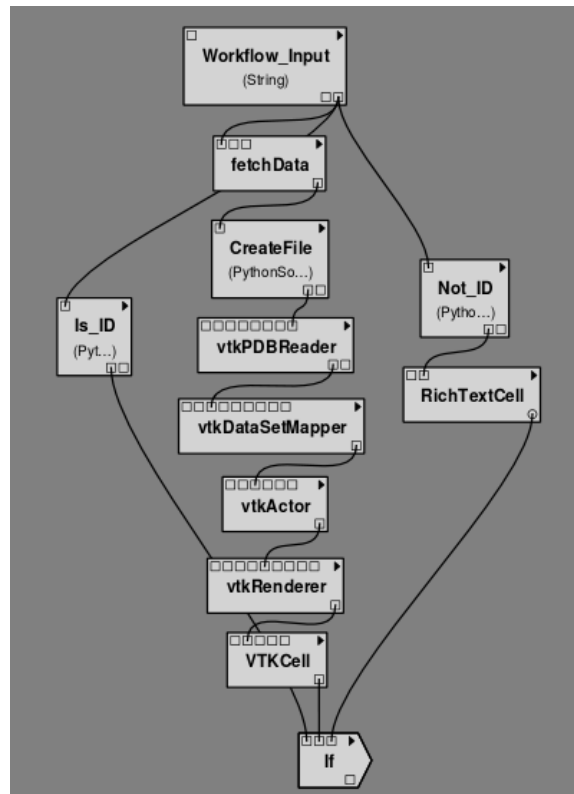


Figure 11.7: All the modules connected

In order to better organize the disposal of the modules, group all the modules shown in Figure *Some modules*

of the workflow connected by selecting them, going to the Edit menu and clicking on Group. Name it as `Generate_Visualization`. Your workflow must correspond to the one shown in Figure *The final workflow, using the Group structure*.

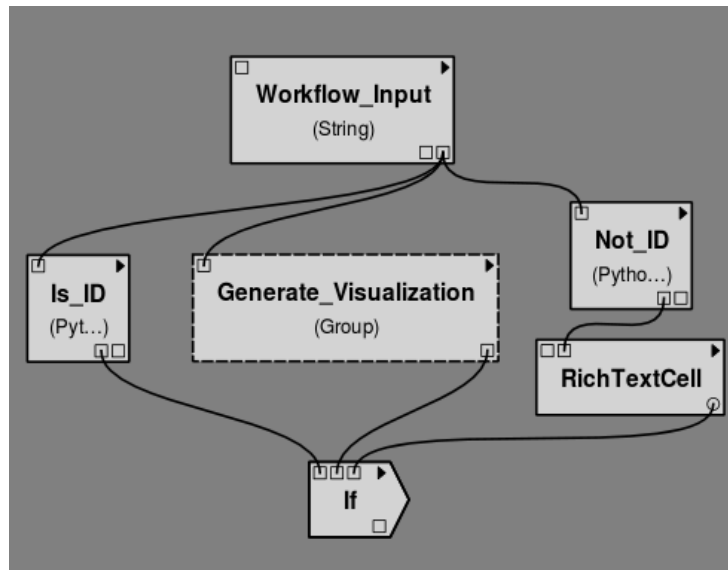


Figure 11.8: The final workflow, using the Group structure

Note that this implementation follows exactly the initial specification of the workflow. If the input is a structure identifier (`Is_ID` returns `True`), `Generate_Visualization` will be executed; otherwise (`Is_ID` returns `False`), `Not_ID` and `RichTextCell` will create an error message in the VisTrails Spreadsheet.

For the workflow execution, set the parameter “value” of the `Workflow_Input` module to `PDB:3BG0`. This entry is an ID from a protein; so, the condition will be `True`, and the `Generate_Visualization` group will be executed, generating the visualization show in Figure *The visualization of the protein in the VisTrails Spreadsheet*.

If you change the value from the input port “value” to `protein`, for example, the condition will be `False`, and the message shown in Figure *The message in the Spreadsheet, generated when the input is not a structure ID* will be generated in the Spreadsheet.

This example can be found inside the “examples” directory, in the “protein_visualization.vt” vistrail. It was partially based on the workflow “Structure_or_ID”, which can be found at <http://www.myexperiment.org/workflows/225>.

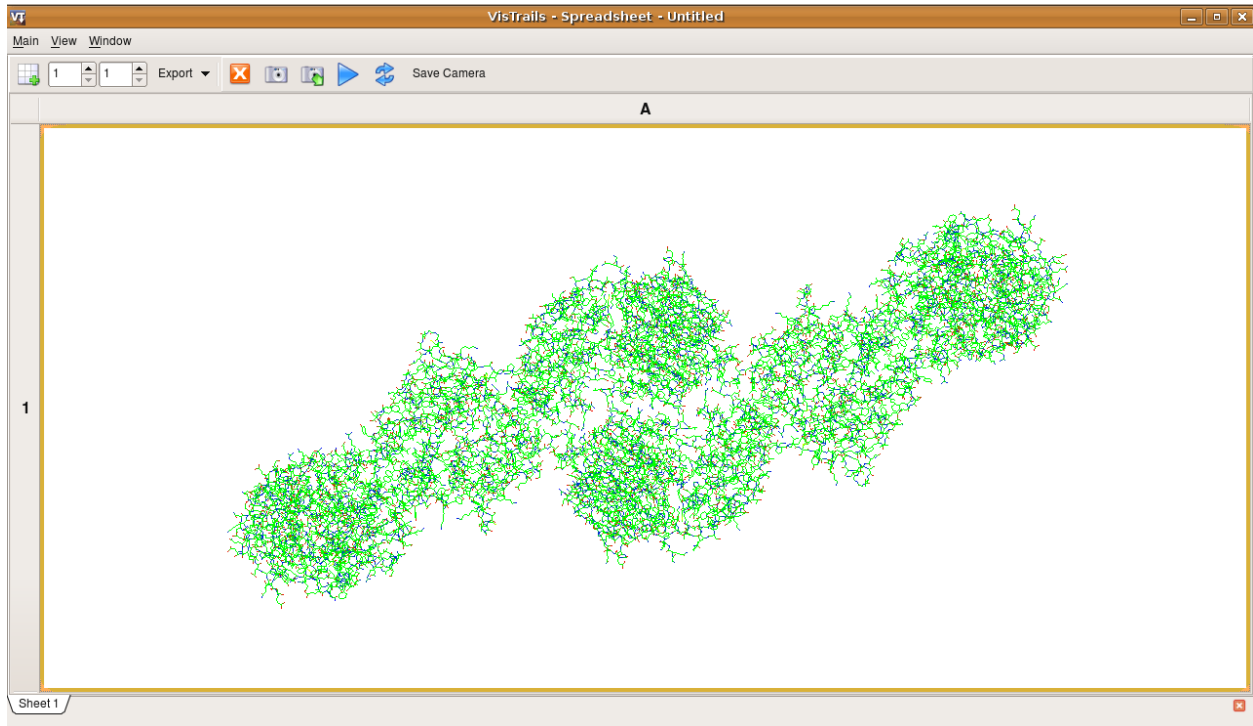


Figure 11.9: The visualization of the protein in the VisTrails Spreadsheet

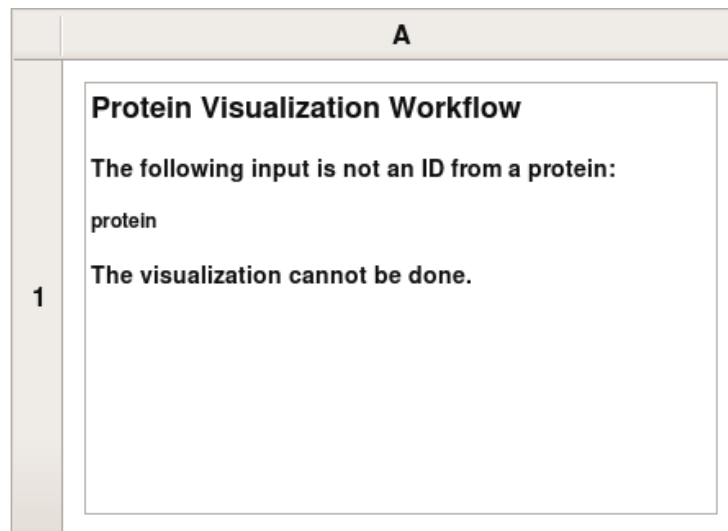


Figure 11.10: The message in the Spreadsheet, generated when the input is not a structure ID

THE CONTROL FLOW ASSISTANT

12.1 Using the Control Flow Assistant (CFA)

In order to simplify the process of creating a control flow loop that uses the Map module, VisTrails has a Control Flow Assistant (CFA). To use the CFA, you must:

1. Enable all ports (in the Module Configuration) that you wish to use as inputs or outputs.
2. Select the modules in the workflow that will form the basis of your mapped input-output loop.
3. Go to 'Edit->Control Flow Assistant' to launch the CFA using the selected modules.
4. Select the input ports that you wish to loop over using List modules as input.
5. Select the output port that you wish to use for the values in the output List.
6. Click OK, and the CFA will generate the control flow structure as a Group module.
7. Connect a List input to each of the inputs on the control flow Group.
8. Connect the Group's output List (output port 'Result') to a suitable module/port, or create a PythonSource module to handle the List output.

Note: All existing connections to input and output ports selected in steps 4 and 5 will be removed.

12.1.1 List Input

By default, the List inputs will be used sequentially, one from each List, which requires that all List inputs be the same length. As another option, the Group created by the CFA has a boolean 'UseCartesianProduct' parameter. If this parameter is set to 'true' then the cartesian product of all of the input Lists will be used as the input for the Map. Use caution when using this parameter, as the number of inputs can grow very rapidly with just a few List inputs.

12.1.2 List Output

There are several ways to handle the output List. One option would be to send the output List to a StandardOutput module to display its contents. Another option is to simply ignore the output List, in the case where you just want part of the workflow to execute multiple times using different inputs. For example, if the mapped portion of the workflow contains a VTKCell, and you just want to generate a new VTKCell for each input, you should select the 'self' port of the VTKCell module when choosing the output port in the CFA, and then ignore the output List. For more specialized handling of the output List, you may wish to create a PythonSource module.

12.1.3 Custom List

For advanced users, the default behavior, or cartesian product behavior may not be sufficient for your needs. If this is the case, the 'UserDefinedInputList' parameter allows you to manually specify the input list. If this parameter is defined, it will override any input lists already defined or connected. The format for this user-provided input list must be a list of lists of tuples. Each inner list represents a single loop execution, and contains tuples (or single values for functions taking only one argument) representing the arguments for each input function to be used in that loop execution. The order of the argument tuples in the inner lists should match the order in which the functions appear on the module generated by the CFA.

For example, if the loop has two input functions defined, in order, as SetXY(x, y) and SetZ(z), and we want two executions of the loop, the input list would be: [[(x1, y1), z1], [(x2, y2), z2]]

12.1.4 Parameter Exploration

One useful purpose for the CFA is to provide a version-based approach to parameter exploration. To create a parameter exploration for a workflow, you could simply select all modules in the workflow, making sure the ports for the desired parameters are enabled, then launch the CFA and select the ports of the parameters you wish to explore. By providing a list for each parameter, you can create a parameter exploration that directly uses the version tree.

Try it Now!

Processing a List of values with PythonCalc:

1. Go to 'Edit->Preferences', select the 'Module Packages' tab, and enable the 'pythonCalc' package if it is not already enabled.
2. Click on File->New to start a new VisTrail.
3. Add the following modules from the module registry to the VisTrail: a) One 'PythonCalc' module from the 'pythonCalc' package b) One 'List' module from the 'Basic Modules' package c) One 'StandardOutput' module from the 'Basic Modules' package
4. Set the List 'value' parameter to: [1.0, 2.0, 3.0, 4.0, 5.0]
5. Set the PythonCalc 'op' parameter to: '*'
6. Set the PythonCalc 'value2' parameter to: 2.0
7. With the PythonCalc module selected, go to 'Edit->Control Flow Assistant': a) Click on the input port 'value1' and ensure it is highlighted b) Click on the output port 'value' and ensure it is highlighted c) Click 'OK' to close the window and build the loop structure as a Group module
8. Connect the 'List' module's output port 'value' to the 'Group' module's input port 'value1'.
9. Connect the 'Group' module's output port 'Result' to the 'StandardOutput' module's input port 'value'
10. Click on 'Run->Execute Current Workflow'
11. In your Standard Output console, you should see a List containing the computation for each element in the input list: [2.0, 4.0, 6.0, 8.0, 10.0]

Try it Now!

Performing a Parameter Exploration:

1. Go to 'File->Open', explore to the VisTrails examples folder, and open 'spx.vt'
2. Open the History view and select the version tagged as 'decimate'.
3. Open the Pipeline view.
4. Open the module configuration for the 'vtkContourFilter' module and enable the 'SetValue' input port.
5. Click on 'Edit->Select All'.
6. With all modules selected, go to 'Edit->Control Flow Assistant': a) Click on the 'vtkContourFilter' module's input port 'SetValue' and ensure it is highlighted b) Click on the 'VTKCell' module's output port 'self' and ensure it is highlighted c) Click 'OK' to close the window and build the loop structure as a Group module
7. Select the newly created 'Group' module, and set the 'SetValue' parameter to: [(0, 0.5), (0, 0.75), (0, 1.0)]
8. Click on 'Run->Execute Current Workflow'
9. In your VisTrails Spreadsheet, you should see three visualizations, one for each set of input parameters to the 'SetValue' port of 'vtkContourFilter'.

CONNECTING TO A DATABASE

As an environment for collaborative scientific exploration, VisTrails supports both stand-alone, file-based storage and relational storage of vistrails. With a relational database supporting VisTrails, multiple users can easily collaborate on projects without copying files back and forth. At the same time, if you choose to work without being connected to a database, you can save your work locally to files. Finally, VisTrails can import and export to both types of storage so you can import a vistrail from the database, save it locally as a file on your computer, make and save changes, and export those changes back to the database. Remember that because the complete workflow evolution is always saved, other users will not overwrite your changes, and vice versa. This prevents users from “stepping on each other’s feet.”

By default, VisTrails works with local files stored in the “.vt” format (essentially compressed XML). You can save a vistrail as uncompressed XML by saving the file with a “.xml” extension. When saving a vistrail, the system displays a standard save dialog. These files have a version associated with them so when the schema for these files may change, VisTrails will be able to import older versions. The current version of the XML schema can be found in the distribution at:

```
vistrails/db/versions/v0_9_1/schemas/xml/vistrail.xsd
```

where `v0_9_1` is the current version.

13.1 Setup

As described earlier, VisTrails supports relational database storage as well as file-based storage. Currently, VisTrails has been tested with MySQL, but in the future, we plan to support most standard relational database systems.

13.1.1 Setting up the database

Before using VisTrails with a relational database, you must have a database installed and have access to create, access, and modify that database. If you are planning to deploy for institution-wide access, you should consult your system administrator to determine the correct configuration. The database schema for VisTrails can be found in the distribution at:

```
vistrails/db/versions/v0_9_1/schemas/sql/vistrails.sql
```

where `v0_9_1` is the current version. This schema contains a sequence of SQL commands that define the tables needed for storing vistrails.

After you or someone else has created the database for the vistrails, you will need the following information:

1. *hostname*: the name or IP address of the machine that stores the database (`localhost` if it is your own machine)
2. *port*: the port number that you connect to the database on

3. *user*: the username that should be used to access and modify the vistrails database
4. *password*: the password for the account corresponding to the given user
5. *database name*: the name of the database where the vistrails are to be stored.

13.1.2 Setting up VisTrails

If you are planning to use the database for most of your work, you can configure VisTrails to open vistrails from the database by default. To do so, select the `Preferences` option from the `Edit` menu. (On Mac OS X, the `Preferences` option is found under the `VisTrails` menu.) When the `Preferences` window opens, check the “Read/Write to database by default” box in the `General Configuration` tab. You can switch the default back to a file-based interaction by unchecking this box.

13.2 Opening from a database

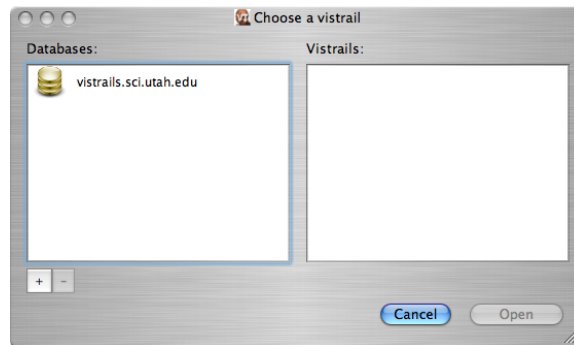


Figure 13.1: Opening a vistrail from the database.

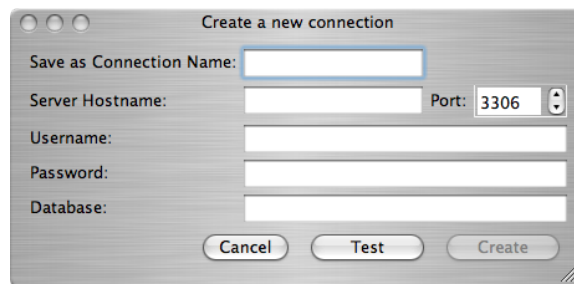


Figure 13.2: Creating a new database connection.

To open a vistrail from a relational database, choose the `Import` option from the `File` menu. You should see a dialog like the one pictured in Figure *Opening a vistrail from the database*. (Alternatively, if you have set VisTrails to use a relational database by default (see Section *Setting up VisTrails*), then you should select `File` → `Open` from the menu or the `Open` button on the toolbar.)

If you have previously connected to databases using VisTrails, you should see a list of these databases in the left column of the dialog. If not, you will need to add one. To do so, click the `+` icon in the lower-left corner. This will bring up a dialog like that shown in Figure *Creating a new database connection*. To set up a connection, you will need the database connection information outlined in Section *Setting up the database*. After filling in that information, you can test the connection by clicking the `Test` button. If the test succeeds, click the `Create` button to add the database to the available sources for vistrails.

The database you wish to use should now be listed in the left column. Clicking on that row will query the database for a list of vistrails available from the database and display them in the right column. To open a vistrail, select the desired vistrail and click the `Open` button or simply double-click the vistrail. When the vistrail has loaded, you will be able to interact with it in exactly the same way as a vistrail loaded from a file.

13.3 Saving to a database

If you opened a vistrail from the database, the default save action will be to save that vistrail back to the database. There will be no dialogs displayed—the database the vistrail was loaded from will be automatically updated.

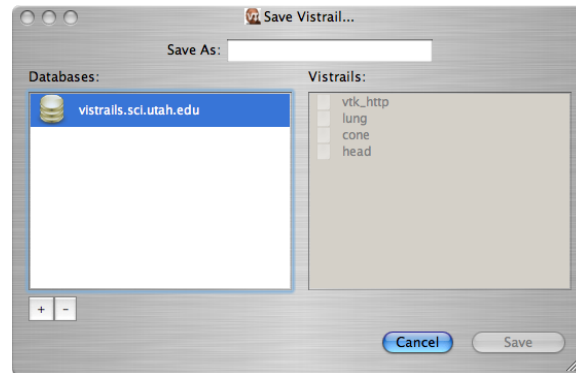


Figure 13.3: Saving a vistrail to the database.

If you opened the vistrail from a file, you will need to select either `Save As` or `Export` from the `File` menu, depending on whether VisTrails uses the database by default (see Section *Setting up VisTrails*). You will be shown a dialog similar to the one in Figure *Saving a vistrail to the database*. As discussed in Section *Opening from a database*, you can create a new connection to the database or use an existing one. Note that the name of the vistrail must differ from those already stored on the database, and clicking the `Save` button will persist the changes to the database.

13.4 Known Issues

Currently, saving a vistrail to the database will *overwrite* the vistrail currently stored on the database. However, we plan to add synchronization soon so that all explorations are captured. Thus, be aware that if two users have the same vistrail loaded from the database at the same time, and both users save their changes, only the second user's changes will be captured.

EXAMPLE: WEB SERVICES

A *web service* grants you programmatic access to an online data source via a straightforward API. In this chapter, you will learn how to invoke web services from within VisTrails workflows. We will build a simple workflow that invokes a web service and generates an HTML table with the results. Our current example is intentionally simple; for more in-depth examples, please refer to the VisTrails website.

Where we’re going in this chapter: The European Bioinformatics Institute maintains ChEBI,¹ a database of over 15,000 chemical compounds. Each entity is referenced by a unique ID number, called its *chebiID*. To see an example of the kind of queries we will build in this example, go to <http://www.ebi.ac.uk/chebi/webServices.do> and scroll down until you find the web form labeled “getCompleteEntity.” (Figure (a) *Web browser interface for the ChEBI database.*). If you type 15357 into the text field, it will return a long string of data in XML format about this chemical. We learn, among other things, that this chemical’s name is *acetylenedicarboxylate(-2)*.

To try another query, scroll down to the area labeled “getOntologyChildren” and type 15357 into the text field. This returns an XML representation of this chemical’s ontology children. In this case, the result is a single chemical, *acetylenedicarboxylate(-1)*, whose chebiID is 30782 (Figure (b) *Results from a “getOntologyChildren” query.*).

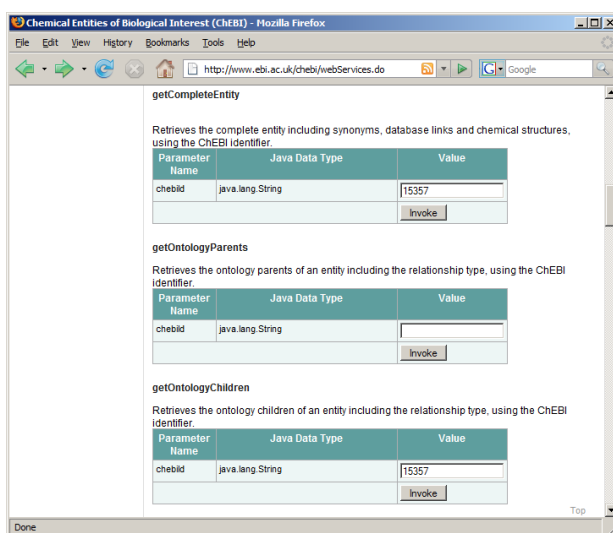


Figure 14.1: (a) Web browser interface for the ChEBI database.

In this example, we will build a workflow that accesses a web service to perform the second of these two queries. Because we’re using a web service, we don’t need a browser — we will perform this query programmatically within VisTrails.

¹ ChEBI is an acronym for Chemical Entities of Biological Interest.

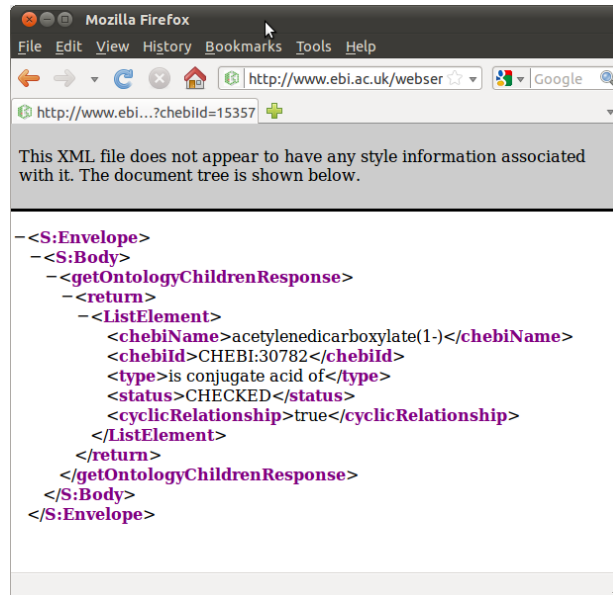


Figure 14.2: (b) Results from a “getOntologyChildren” query.

14.1 Enabling the SUDSWebServices Package

In order to use web services in VisTrails, you need to ensure that the `SUDSWebServices` package is enabled in the `Preferences` dialog. (Please refer to Chapter [Writing VisTrails Packages](#) for more information on enabling packages.)

14.2 Adding Web Service Packages

Within the `Module Packages` tab of the `Preferences` dialog, click the `Configure` button to open the configuration dialog for this package (`SUDSWebServices`). Select the `wsdlList` and click on the `Value` field. This is where you will enter the URL(s) of the web service(s) you wish to access. If there is more than one URL, place a semicolon (;) between each URL, but *not* after the final URL. In other words, the URLs must be semicolon-delimited, but not semicolon-terminated.

For our example, we need the following URL:

```
http://www.ebi.ac.uk/webservices/chebi/webservice?wsdl
```

After closing the dialog, you need to reload the `SUDSWebServices` package in order to load the changes. Then, close the `Preferences` dialog. A new package will be created for each URL provided.

Alternatively, you may add a web service package by clicking the secondary mouse button on the “SUDS Web Services” package in the module palette and entering the corresponding URL. You may remove a web service by clicking the secondary mouse button on the corresponding package in the module palette and selecting `Remove this Web Service`.

14.3 Creating a new vistrail

After configuring the `SUDSWebServices` package properly, there will be a `SUDSWebServices` entry in your `Modules` panel. The `SUDSWebServices` package will generate a module for each published method in a web

service.

Start with a new empty workflow in the Pipeline view, and drag the following modules to the canvas.

- String (under “Basic Modules”)
- getOntologyChildren (under “Methods” for the current web service)
- getOntologyChildrenResponse (under “Types” for the current web service)
- OntologyDataItemList (under “Types” for the current web service)
- PythonSource (under “Basic Modules”)
- RichTextCell (under “Spreadsheet”)

As discussed in Chapter *Creating and Modifying Workflows*, PythonSource has no input and output ports by default; we need to create some. Open the configuration dialog for PythonSource by selecting this module in the pipeline canvas and typing ‘Ctrl-E’. Add a new input port named “ontologyDataItemList” of type List, and a new output port named “outfile” of type File. (Please refer to Chapter *Creating and Modifying Workflows* for more information about configuring and using the PythonSource module.)

We will now add some Python code to this module. This code generates a simple HTML table based on the information retrieved from the web service query. Type or paste the following source code into the PythonSource configuration dialog:

```
dataitemlist = self.getInputFromPort("ontologyDataItemList")
output1 = self.interpreter.filePool.create_file()
f1 = open(str(output1.name), "w")
text = "<HTML><TITLE>Chebi WebService</TITLE><BODY BGCOLOR=#FFFFFF>"
f1.write(text)
text = "<H2>getOntologyChildren Query</H2><BR>"
f1.write(text)
text = "<CENTER><table border = 1><tr><TH>ChebiId</TH> <TH>ChebiName</TH>"
text += "<TH>Comments</TH> <TH>Type</TH> <TH>Status</TH>"
text += "<TH>CyclicRelationship</TH></tr>"
f1.write(text)
for element in dataitemlist:
    if not hasattr(element, 'Comments') or str(element.Comments) == '[]':
        comment = ""
    else:
        comment = str(element.Comments)
    line = "<tr><td>" + str(element.chebiId) + "</td><td>" + str(element.chebiName)
    line += "</td><td>" + comment + "</td><td>" + str(element.type) + "</td><td>"
    line += str(element.status) + "</td><td>" + str(element.cyclicRelationship)
    line += "</td></tr>"
    f1.write(line)
text = "</table></CENTER></BODY></HTML>"
f1.write(text)
self.setResult("outfile", output1)
f1.close()
```

Click OK to close the dialog. One of the ports we need to use is an optional port. Select the OntologyDataItemList module and type ‘Ctrl-E’. Check the box for ListElement under Output Ports and select OK to close the dialog. Now connect the modules together as shown in Figure *Our example pipeline*.

Our workflow is now complete except for one crucial element: the starting point. We need to pass a chebiID string to the workflow in order to look up information about a chemical. We do this by assigning a chebiID string to the the String module at the top of the pipeline. Highlight the String module in the canvas, then in the Methods panel at the right, drag the “value” parameter down to the Set Methods panel. Type CHEBI:15357 into the text field.

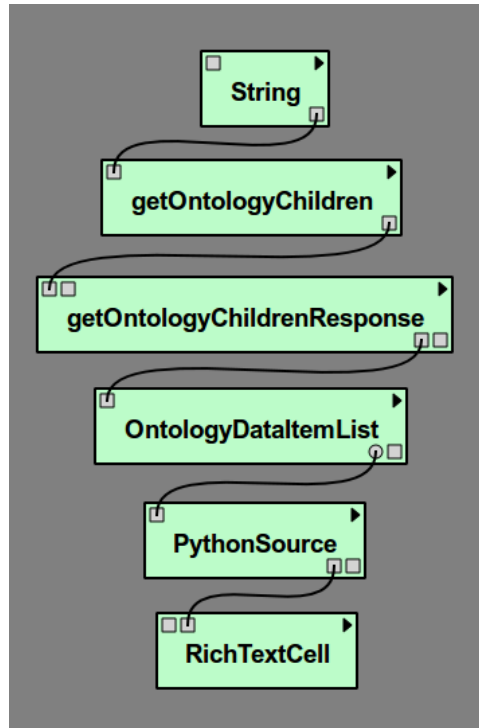
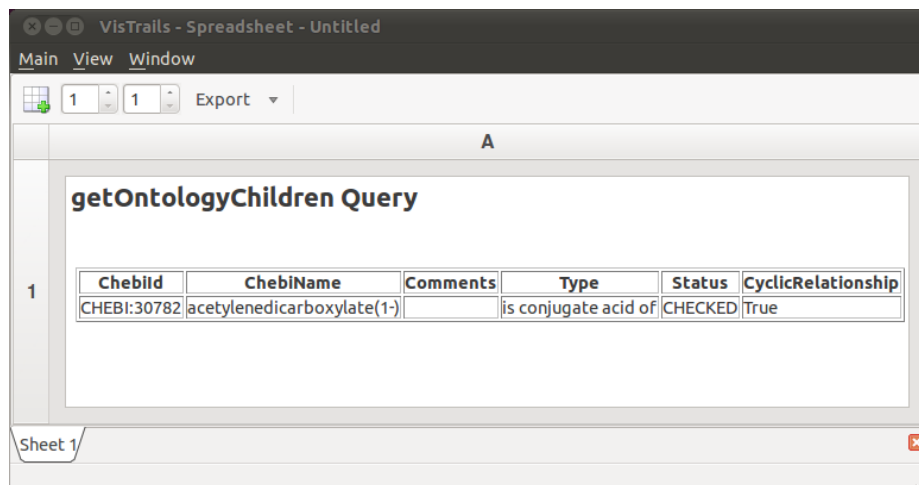


Figure 14.3: Our example pipeline.

14.4 Executing the workflow

The workflow is now ready to be visualized. Click the `Execute` button to send the current pipeline with the current parameters to a `RichTextCell` within the VisTrails Spreadsheet. Your result should resemble Figure *The HTML table generated by our workflow*. As you can see, the “ontology children” query returns the same information as before, but without the use of a web browser. In addition, we used a small Python program (via the `PythonSource` module) to transform the raw XML into a readable HTML table.



VisTrails - Spreadsheet - Untitled

Main View Window

1 1 Export

A

getOntologyChildren Query

ChebiId	ChebiName	Comments	Type	Status	CyclicRelationship
CHEBI:30782	acetylenedicarboxylate(1-)		is conjugate acid of	CHECKED	True

Sheet 1

Figure 14.4: The HTML table generated by our workflow.

PERSISTENCE IN VISTRAILS

The `Persistence` package in Vistrails helps improve reproducibility by associating versions of data files with their provenance, and minimize the need to rerun lengthy executions by keeping intermediate persistent files. Although, we will focus primarily on the use of persistent files, persistent directories are used in the same manner. The difference is that a persistent directory deals with multiple files within a directory rather than a single file.

15.1 Getting Started With Persistence

A persistent file is simply a file that is kept in a repository and identified by an id and version string and annotated with a signature and content hash. To begin, notice that there are three persistent file/directory types: input, intermediate, and output. It is helpful to understand the differences among these files as well as their distinguishing characteristics and configuration options.

15.1.1 Input Files

To use a persistent input file, after dragging it onto the canvas, it is necessary to edit the configuration (CTRL+E). For an input file, one can either create a new reference to an existing local file, or use a file that exists in the database. To create a new reference, select `Create New Reference` and either enter the path to the file, or select the folder icon on the right to browse local directories for a file. Then, give the file a name and any appropriate tags to help identify the file. Select `OK`. The file will be added to the repository. If the file is already in the repository, select `Use Existing Reference` and select the appropriate file. You may use the search box to search for particular files by name, tag, ID, or content hash.

Note

A new version of a persistent file is created each time its contents change. A persistent input file will always use the most recent version of a file if it is assigned to the root of the file tree.

The `PersistentInputFile` module can also be used to read or write data directly from or to a local file. To use this feature, you will need to set `localPath` point to a local file. This can be done in the `Set Methods Panel`, the `Configuration Dialog`, or by connecting a file to the `localPath` input port. Then, the local file will be read when `readLocal` is set to `true` and written when `writeLocal` is set to `true`. To do this using the `Configuration Dialog`, select `Keep Local Version`, enter the file location or click on the folder icon to select a file, then select `Read From Local Path`.

15.1.2 Output Files

To use a persistent output file, after dragging it onto the canvas, it is necessary to edit the configuration (CTRL+E). Notice that the Configuration Dialog for the output file is the same as that of a the input file except that an option to `Always Create New Reference` exists. Selecting this new option will cause a new file to be created and added to the persistent store each time the workflow is executed. The new file does not get a name or tag, so it can be difficult to identify the newly created file.

If you prefer to have a new version of the file created each time the workflow is changed, you should choose either of the other options. The other two options and the option to read from or write to a local path are used in the same manner as with the input file. However, when the input file is read from a local path, it is not also read from the persistent store. In contrast, when an output file is written to a local path, it is also written to the selected entry in the persistent store.

15.1.3 Intermediate Files

An intermediate file is the same as the output file except that its contents can be used in further calculations. Thus, lengthy computations upstream of the intermediate file will only be recalculated when the upstream workflow changes, but will not need to be rerun when only the downstream workflow changes. Intermediate files by default are set to always create a new reference. Since these files do not need to be manually annotated or named, configuration is optional, but is the same as the output file configuration.

15.2 Using the Output of One Workflow as Input for Another

You need to configure the persistence modules using the module's configuration dialog. After adding a `PersistentOutputFile` to the workflow, click on the triangle in the upper-right corner of the `PersistentOutputFile`, and select "Edit Configuration" from the menu that appears. In this dialog, select "Create New Reference" and give the reference a name (and any space-delimited tags). Upon running that workflow, the data will be written to the persistent store. In the second workflow where you wish to use that file, add a `PersistentInputFile` and go to its configuration dialog in the same manner as with the output file. In that dialog, select "Use Existing Reference" and select the data that you just added in the first workflow in the list of files below. Now, when you run that workflow, it will grab the data from the persistent store.

15.3 Managing Files in the Store

You may manage the files in the store by selecting `Packages -> Persistence -> Manage Store`. You are then free to save files from the store to local files, or delete files from the store. However, please be aware that files deleted from the store are not recoverable. Also, for versioned files, a specific version of a file can not be deleted without deleting all versions of that file. To delete all versions of the file, select the root version and press `Delete`.

15.4 Examples

Try it Now!**Base Workflow**

1. Drag the following modules to the canvas and connect them in the order in which they are named: HTTPFile, vtkUnstructuredGridReader, vtkDataSetMapper, vtkActor, vtkRenderer, VTKRenderOffscreen. Refer to Figure *Example 1.1* to ensure connections are correct.
2. Select the HTTPFile module and set the url to:
<http://www.vistrails.org/download/download.php?type=DATA&id=spx.vtk>
3. (Optional) Select the vtkRenderer module and drag SetBackgroundWidget to the Set Methods Panel. Select the background color of your choice.

Persistent Output

4. Switch to the History view and tag the current version as “Base Workflow”. Then, switch back to the pipeline view.
5. Enable the persistence package.
6. Drag the PersistentOutputFile module to the canvas and connect the output from VTKRenderOffscreen to its value port. See Figure *Example 1.2*.
7. With the PersistentOutputFile module selected, press CTRL-E to edit the module configuration.
8. Select Create New Reference, name it “persistence1_1”, and give it a “Persistent Output” tag. Select Ok.
9. Execute the workflow.

Persistent Input

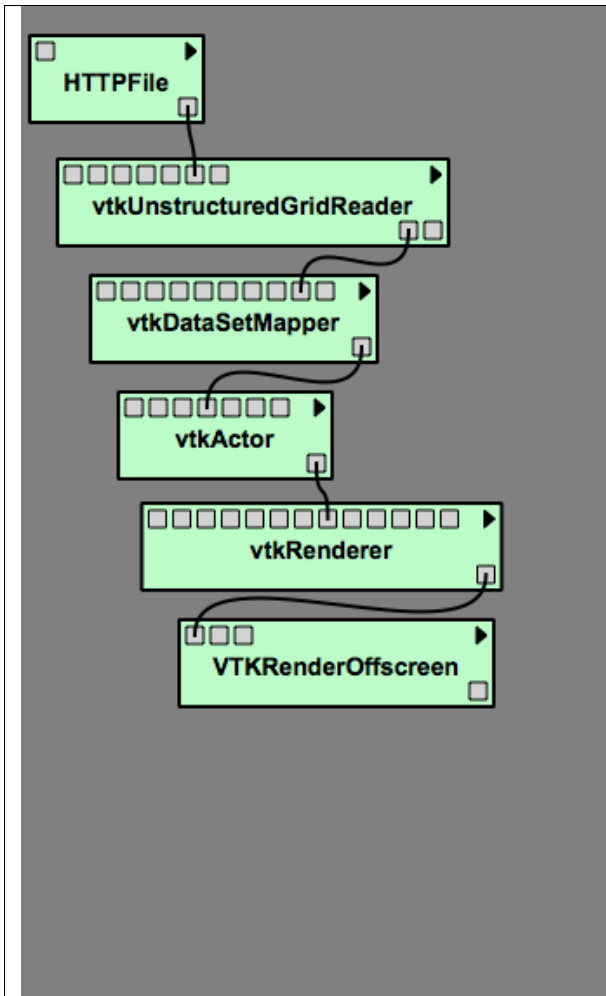
10. Switch to the History view again and tag the current version as “Persistent Output”. Then, select the root of the version tree and go back to the pipeline view.
11. Drag the PersistentInputFile and ImageViewerCell modules to the canvas and connect them. See Figure *Example 1.3*.
12. Edit the configuration of the PersistentInputFile. Select “Use Existing Reference” and select the file named “persistence1_1” with the “Persistent Output” tag. Select Ok. See Figure *Example 1.4*.
13. Execute the workflow. An image should be displayed in the VisTrails spreadsheet.

Persistent Intermediate

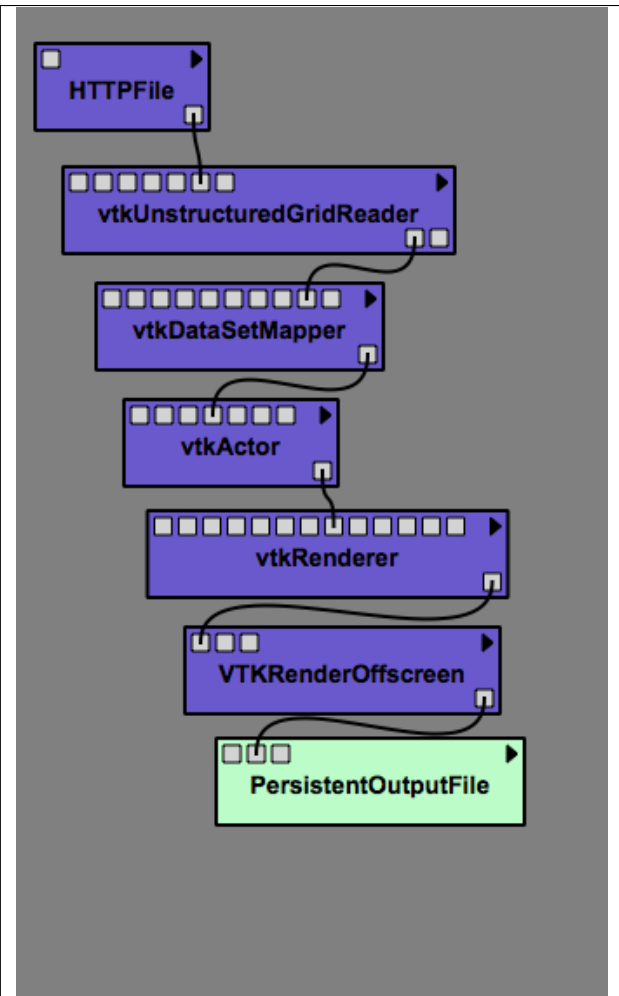
14. Switch to the History view and tag the version with “Persistent Input”, then select the “Base Workflow” version and switch to the pipeline view again.
15. Drag the PersistentIntermediateFile and ImageViewerCell modules to the canvas and connect them as shown in Figure *Example 1.5*.
16. Execute the workflow.
17. In the History view, tag the version with “Persistent Intermediate”. See Figure *Example 1.6*.

Experimentation

You can now change the background color in the “Persistent Output” version and execute the pipeline. The changes should show up when you execute the pipeline of the “Persistent Input” version. You can also change the background color in the “Persistent Intermediate” version, but this example merely demonstrates how to use the module. Performance increase will not be seen here due to the lack of lengthy computations.



Example 1.1 - Base Workflow - An example pipeline without persistence.



Example 1.2 - Persistent Output File

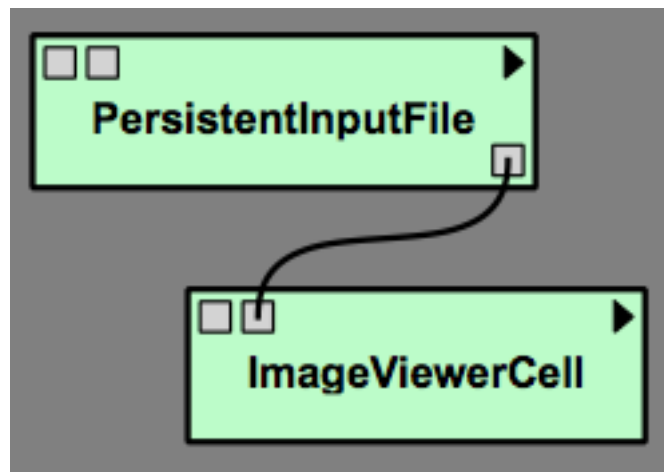


Figure 15.1: Example 1.3 - Persistent Input File

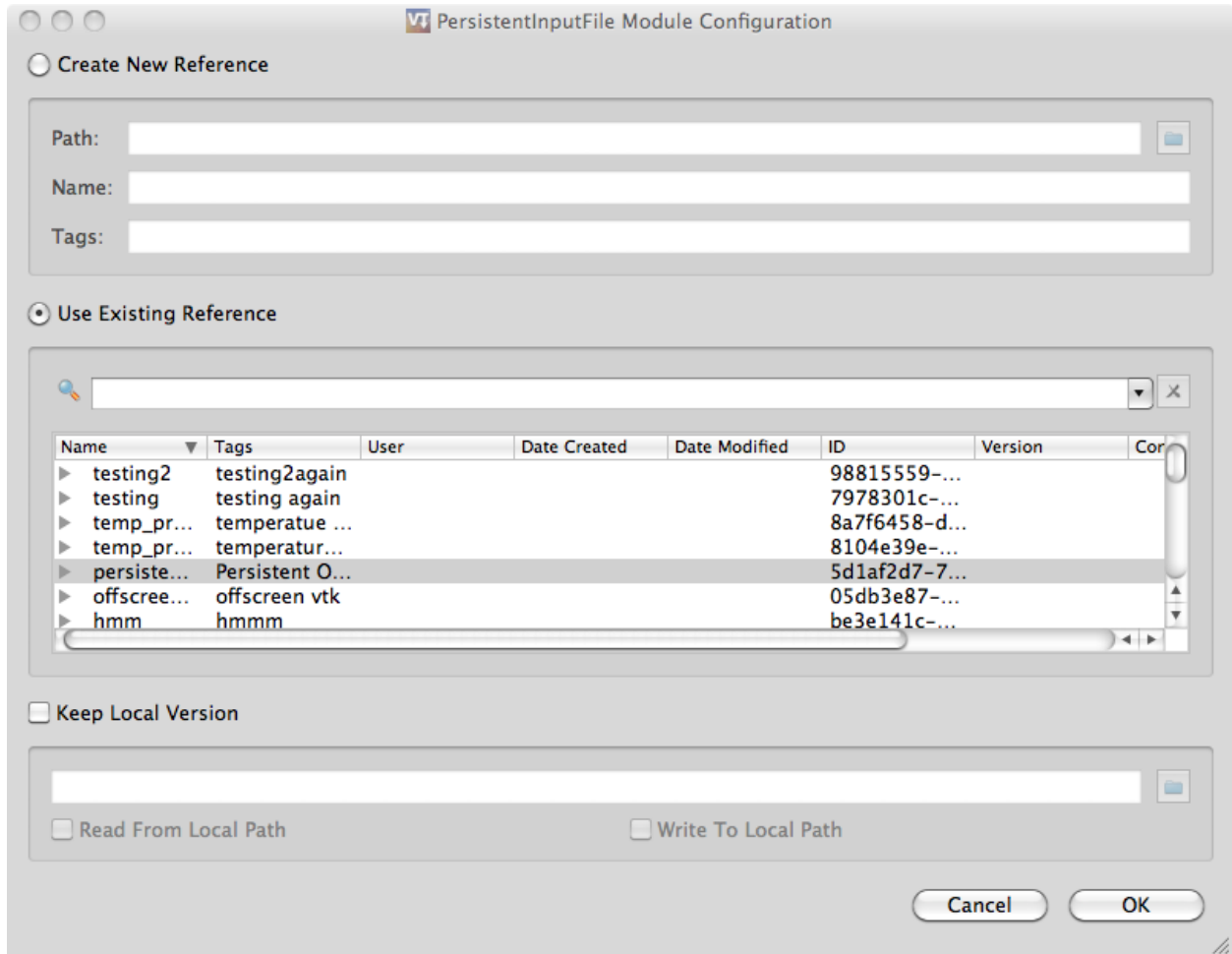


Figure 15.2: Example 1.4 - Selecting an existing reference

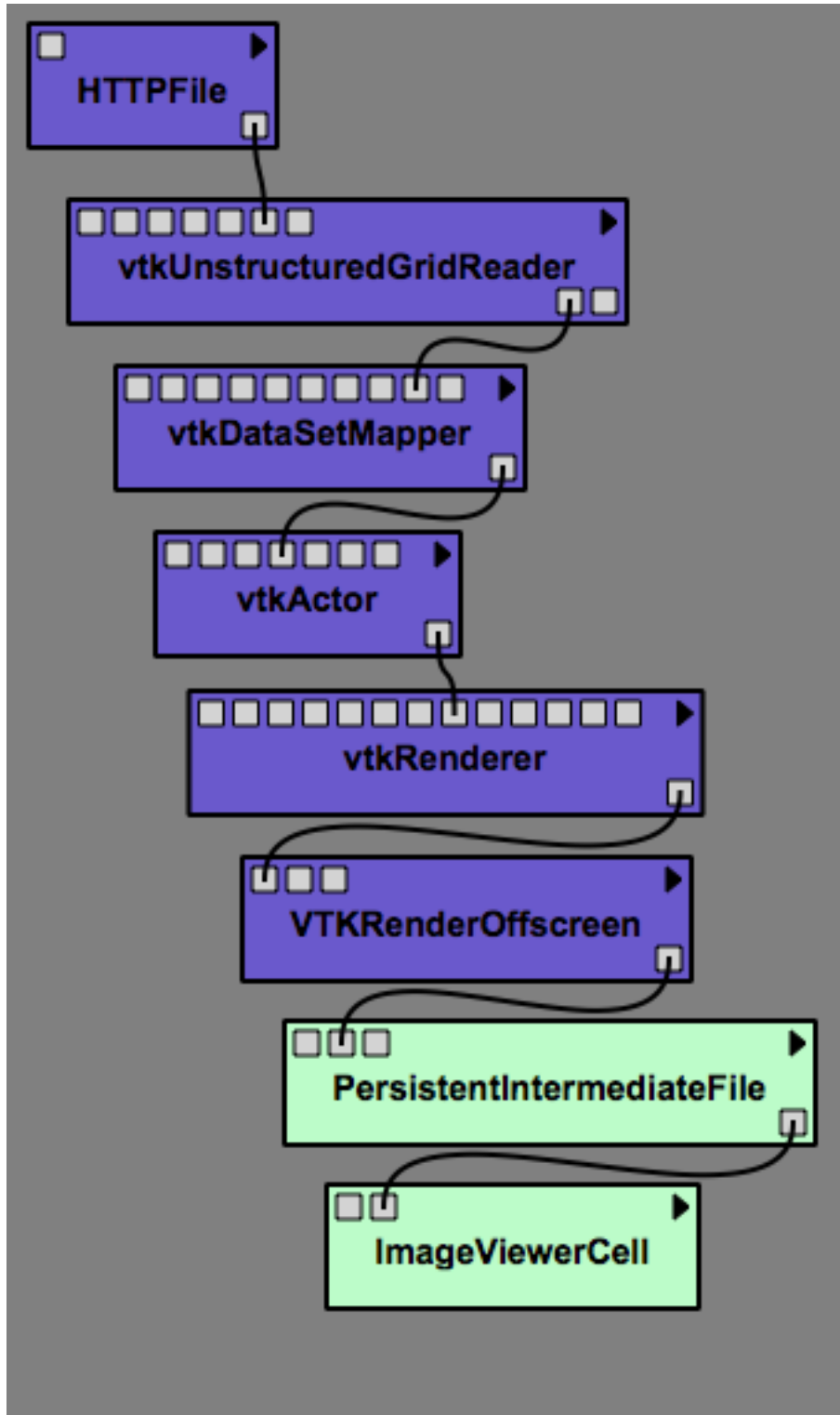


Figure 15.3: Example 1.5 - Persistent Intermediate File

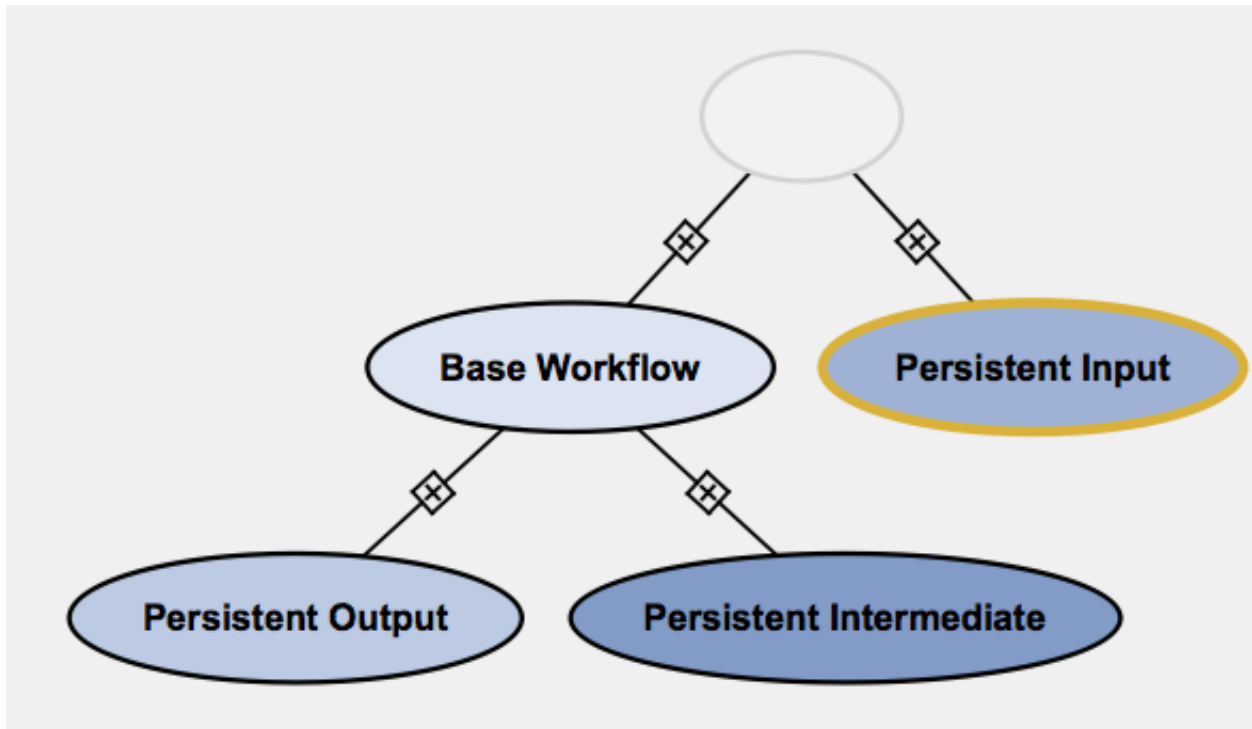


Figure 15.4: Example 1.6 - The History Tree

EMBEDDING VISTRAILS FILES VIA LATEX

The VisTrails Latex extension allows you to embed a result from a VisTrails file in a Latex document. Images to be included in the Latex document will be generated through VisTrails and can be linked to the VisTrails file and version from which it was generated. In other words, Latex call VisTrails to generate an image for a resulting PDF document. The resulting image can be set up so, when clicked, the generating VisTrails file will be opened in Vistrails.

16.1 Latex Setup and Commands

This section contains instructions for setting up Latex files to use VisTrails. For a complete example of this setup, you may also refer to `example.tex` in VisTrails' `extensions/latex` directory.

To use the Latex extension, copy `vistrails.sty` and `includevistrail.py` from the `extensions/latex` directory to the same directory as your `.tex` files. Then, add the following line to the beginning of the latex file:

```
\usepackage{vistrails}
```

By default, VisTrails will be executed at `www.vistrails.org` and the images downloaded to your hard drive. If you want to run a local copy of VisTrails instead, add the path to your python file or executable:

```
\renewcommand{\vistrailspath}{/path/to/vistrails.py}
```

By default, images are set up to link to their corresponding VisTrail. However, if you are using your own web server, additional setup is required (see *Setup For Use With Files on MediaWiki or a Web Server*). Otherwise, to setup the images without links (not clickable), add:

```
\renewcommand{\vistrailsdownload}{}
```

Finally, if you don't have python on your path, use this to set the python interpreter:

```
\renewcommand{\vistrailspythonpath}{/path/to/python/executable}
```

Note: If you set the `vistrailspythonpath` to an invalid path VisTrails will use cached files if they exist.

16.1.1 Setup For Use With Files on MediaWiki or a Web Server

Many VisTrails files and/or data is stored in a database that readers of a pdf document might not have access to. If the files are also accessible through the web, the following instructions explain setup that will allow readers to download the VisTrail or workflow through MediaWiki or a web server.

MediaWiki

To setup your MediaWiki for use with VisTrails:

- In your wiki/extensions folder, create a config.php file based on the config.php.sample file located in VisTrails' extensions/http folder.
- Copy download.php, functions.php, and vistrailsExtension.php from the extensions/mediawiki folder to your wiki/extensions folder and update these files according to your needs.
- Configure your .tex files with:

```
\renewcommand{\vistrailsdownload}{http://yourwebserver.somethingelse/download.php}
```

Web Server

To configure VisTrails to run on your web server:

- Create a config.php file based on the config.php.sample file located in VisTrails' extensions/http folder.
- Copy functions.php and downloads.php from the extensions/mediawiki folder and update them according to your needs.
- **Depending on the functionality you want to make available, copy any or all of the following files:**

- run_vistrails.php - will execute workflows and return images and pdf files
- get_db_vistrail_list.php - will return a list of VisTrails available in the database
- get_vt_xml.php - will return a VisTrail in xml format
- get_wf_pdf.php - will return a workflow graph in pdf format
- get_wf_xml.php - will return a workflow in xml format

- Configure your .tex files with:

```
\renewcommand{\vistrailspath}{http://yourwebserver.somethingelse/run_vistrails.php}
\renewcommand{\vistrailsdownload}{http://yourwebserver.somethingelse/download.php}
```

16.2 Including VisTrails Results in Latex

To include VisTrails Results in a Latex file:

- Select the `History` view.
- Ensure that a version is selected.
- Press the `Embed` button at the bottom of the Properties Panel. This will launch a dialog with embedding options (see Figure *Embedding Options*).
- Select the result that you would like to display. The choices are: workflow results, workflow graph, and history tree graph.
- Select `Latex`.
- You should then choose from a number of “Embed” and “Download” options which will be explained in the tables below.
- Select “Copy to Clipboard”
- Paste clipboard contents into you Latex document
- Run `pdflatex` with the `-shell-escape` option: `pdflatex -shell-escape example.tex`.

Note on using local VisTrails files: Relative or absolute filenames can be used in the .tex file, but absolute filenames are used in the pdf. Thus, if the absolute location of the file has changed, the pdf will need to be regenerated even if the relative location of the file has not changed. Also, the VisTrails `Embed` function assumes the .vt file is in the same directory as the .tex file. You will need to change this to an absolute filename if it is not.

Table 16.1: Configuration Options

Option	Latex Flag	Description
Workflow Results	version=<...>	Show the results of the specified version.
Workflow Graph	version=<...> showworkflow	Show the workflow instead of the results.
History Tree Graph	showtree	Show the version tree instead of the results.

Table 16.2: Embed Options

Option	Latex Flag	Description
As PDF	pdf	Download images as pdf files.
Smart Tag	tag=<...>	Allows you to include a version's tag. If a tag is provided, version can be omitted and buildalways is implicit.
Cache Images	buildalways (do not include for caching)	When caching desired, the buildalways flag should not be included. If it is included, VisTrails will be called regardless of whether or not it has been called for the same host, db, version, port and vt_id.
Include .vtl	getvttl	Causes the .vttl file to be downloaded.

Table 16.3: Download Options

Option	Latex Flag	Description
Include Workflow	embedworkflow	Download the workflow only.
Execute Workflow	execute	Will cause the workflow to be executed when it is opened.
Include Full Tree	includefulltree	Download the complete VisTrail.
Show Spreadsheet Only	showspreadsheetonly	Will initially only show the spreadsheet.

16.2.1 Example

The following is an example command for including a VisTrails image in Latex:

```
\vistrails[host=vistrails.sci.utah.edu,
db=vistrails,
version=<version_number>,
vtid=<vistrails_id>,
tag=<tag>
port=3306, %The default value for port is 3306.
buildalways,
execute,
showspreadsheetonly,
pdf,
showworkflow,
showtree,
```

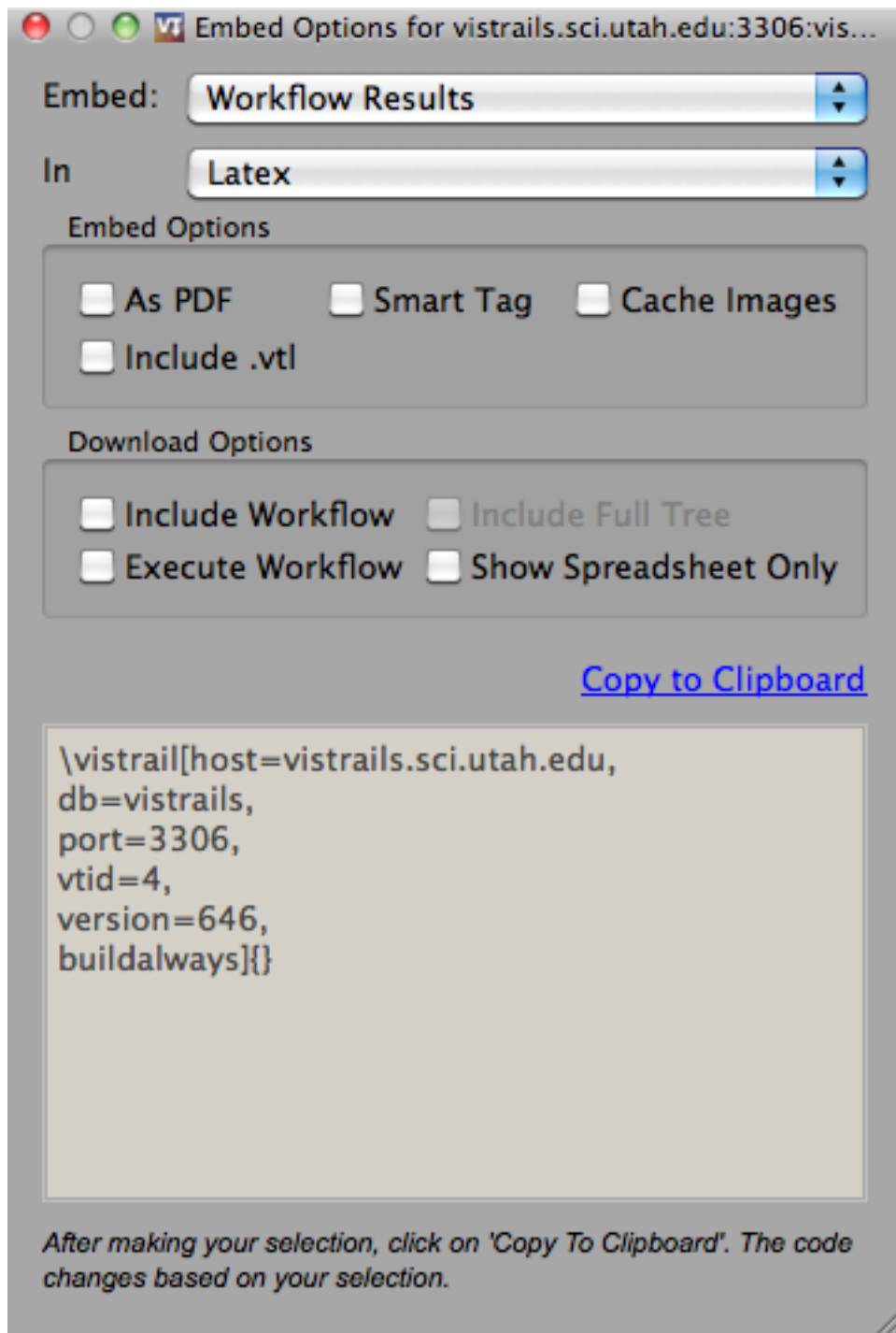


Figure 16.1: Embedding Options


```
getvtl,  
]{width=0.45\linewidth} %Options you would give to the \includegraphics{} command.
```

See example.tex in the extensions/latex directory for a complete example of usage.

16.2.2 Additional Notes

After running at least once, VisTrails will cache the images and latex instructions. The latex code will be in the “cached” folder and the images in vistrails_images.

VisTrails will create in the current directory a directory called vistrails_images/host_db_vtid_version with the png files generated by the spreadsheet.

Part V

Developer's Guide

WRITING VISTRAILS PACKAGES

17.1 Introduction

VisTrails provides a plugin infrastructure to integrate user-defined functions and libraries. Specifically, users can incorporate their own visualization and simulation code into pipelines by defining custom modules (or wrappers). These modules are bundled in what we call *packages*. A vistrails package is simply a collection of Python classes stored in one or more files, respecting some conventions that will be described shortly. Each of these classes will represent a new module. In this chapter, we will build progressively more complicated modules. Note that even though each section introduces a specific large feature of the VisTrails package mechanism, new small features are highlighted and explained as we go along. Because of this, we recommend at least skimming through the entire chapter at least once.

17.2 Who Should Read This Chapter?

This chapter is written for developers who wish to extend VisTrails with customized modules, tailored for their specific needs. It is assumed that you have experience writing code in the Python programming language. Teaching the syntax of Python is beyond the scope of this manual; for experienced programmers who would like a compact introduction to Python, we recommend the book *Python in a Nutshell* by Alex Martelli (published by O'Reilly).

However, if you do not yet know Python but are familiar with another object-oriented language such as Java or C#, you should be able to get the gist of these examples from looking at the code and by reading our line-by-line commentaries.

17.3 A Simple Example

Here is a simplified example of a very simple user-defined module:

```
1 class Divide(Module):
2     def compute(self):
3         arg1 = self.getInputFromPort("arg1")
4         arg2 = self.getInputFromPort("arg2")
5         if arg2 == 0.0:
6             raise ModuleError(self, "Division by zero")
7         self.setResult("result", arg1 / arg2)
8
9     _input_ports = [('arg1', '(edu.utah.sci.vistrails.basic:Float)', \
10                    {"labels": str(["dividend"])}), \
11                   ('arg2', '(edu.utah.sci.vistrails.basic:Float)', \
12                    {"labels": str(["divisor"])})]
```

```
13     _output_ports = [('result', '(edu.utah.sci.vistrails.basic:Float)',\  
14                       {"labels": str(["quotient"])}])  
15  
16 _modules = [Divide]  
17 #old syntax  
18 #registry.addModule(Divide)  
19 #registry.addInputPort(Divide, "arg1", (basic.Float, 'dividend'))  
20 #registry.addInputPort(Divide, "arg2", (basic.Float, 'divisor'))  
21 #registry.addOutputPort(Divide, "result", (basic.Float, 'quotient'))
```

New VisTrails modules must subclass from `Module`, the base class that defines basic functionality. The only required override is the `compute()` method, which performs the actual module computation. Input and output is specified through ports, which currently have to be explicitly registered with VisTrails. However, this is straightforward, and done through method calls to the module registry. An example of a (slightly) more complicated module follows:

```
1  import core.modules.module_registry  
2  from core.modules.vistrails_module import Module, ModuleError  
3  
4  version = "0.9.0"  
5  name = "PythonCalc"  
6  identifier = "edu.utah.sci.vistrails.pythoncalc"  
7  
8  class PythonCalc(Module):  
9      """PythonCalc is a module that performs simple arithmetic operations on its inputs."""  
10  
11     def compute(self):  
12         v1 = self.getInputFromPort("value1")  
13         v2 = self.getInputFromPort("value2")  
14         result = self.op(v1, v2)  
15         self.setResult("value", result)  
16  
17     def op(self, v1, v2):  
18         op = self.getInputFromPort("op")  
19         if op == '+': return v1 + v2  
20         elif op == '-': return v1 - v2  
21         elif op == '*': return v1 * v2  
22         elif op == '/': return v1 / v2  
23         else: raise ModuleError(self, "unrecognized operation: '%s'" % op)  
24  
25     #####  
26  
27     def initialize(*args, **keywords):  
28  
29         # We'll first create a local alias for the module registry so that  
30         # we can refer to it in a shorter way.  
31         reg = core.modules.module_registry.registry  
32  
33         reg.add_module(PythonCalc)  
34         reg.add_input_port(PythonCalc, "value1",  
35                             (core.modules.basic_modules.Float, 'the first argument'))  
36         reg.add_input_port(PythonCalc, "value2",  
37                             (core.modules.basic_modules.Float, 'the second argument'))  
38         reg.add_input_port(PythonCalc, "op",  
39                             (core.modules.basic_modules.String, 'the operation'))  
40         reg.add_output_port(PythonCalc, "value",  
41                             (core.modules.basic_modules.Float, 'the result'))
```

To try this out in VisTrails, save the file above in the `.vistrails/userpackages` subdirectory of your home

directory, with the filename `pythoncalc.py`. Then, click on the `Edit` menu (or the `VisTrails` menu on Mac OS X), select the `Preferences` option and select the `Module Packages` tab. A dialog similar to what is shown in Figure *All available packages...* should appear. Select the `pythonCalc` package, then click on `Enable`. This should move the package to the `Enabled packages` list. Close the dialog. The package and module should now be visible in the VisTrails builder.

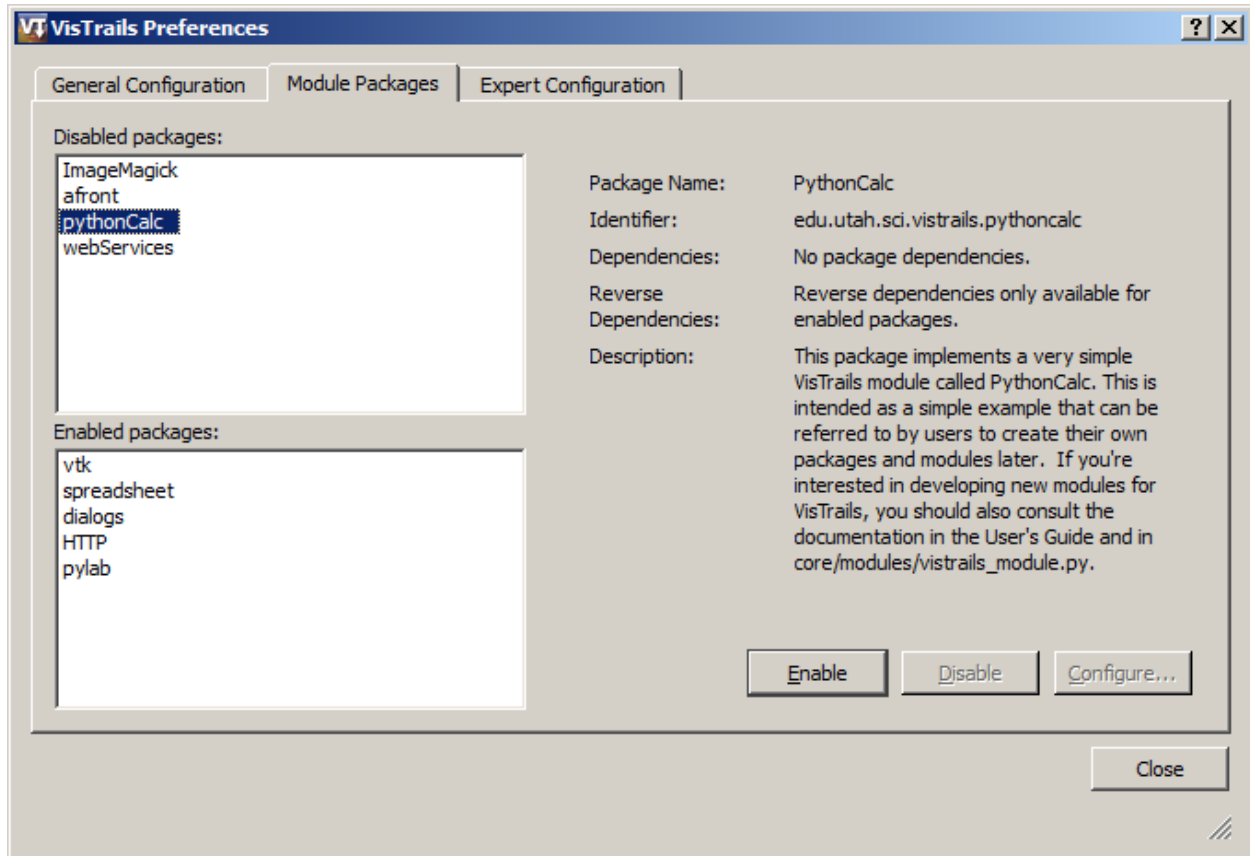


Figure 17.1: All available packages can be enabled and disabled with the VisTrails preferences dialog.

Now create a workflow similar to what is shown in Figure *A simple workflow that uses PythonCalc, a user-defined module..* When executed, this workflow will print the following on your terminal:

```
7.0
```

Let's now examine how this works. The first two lines simply import required components. The next three lines give VisTrails meta-information about the package. `Version` is simply information about the package version. This might be tied to the underlying library or not. The only recommended guideline is that compatibility is not broken across minor releases, but this is not enforced in any way. `Name` is a human-readable name for the package.

The most important piece of metadata, however, is the package *identifier*, stored in the variable called `identifier`. This is a string that must be globally unique across all packages, not only in your system, but in any possible system. We recommend using an identifier similar to Java's package identifiers. These look essentially like regular DNS names, but the word order is reversed. This makes sorting on the strings a lot more meaningful. You should generally go for `institution.project.packagename` for a package related to a certain project from some institution, and `institution.creatorname` for a personally developed package. If you are wrapping third-party functionality, *do not* use their institution's DNS, use your own. The rationale for this is that the third party itself might decide to create their own VisTrails package, and you do not want to introduce conflicts.

Line 8 is where we actually start defining a new module. Every VisTrails module corresponds to a Python class that

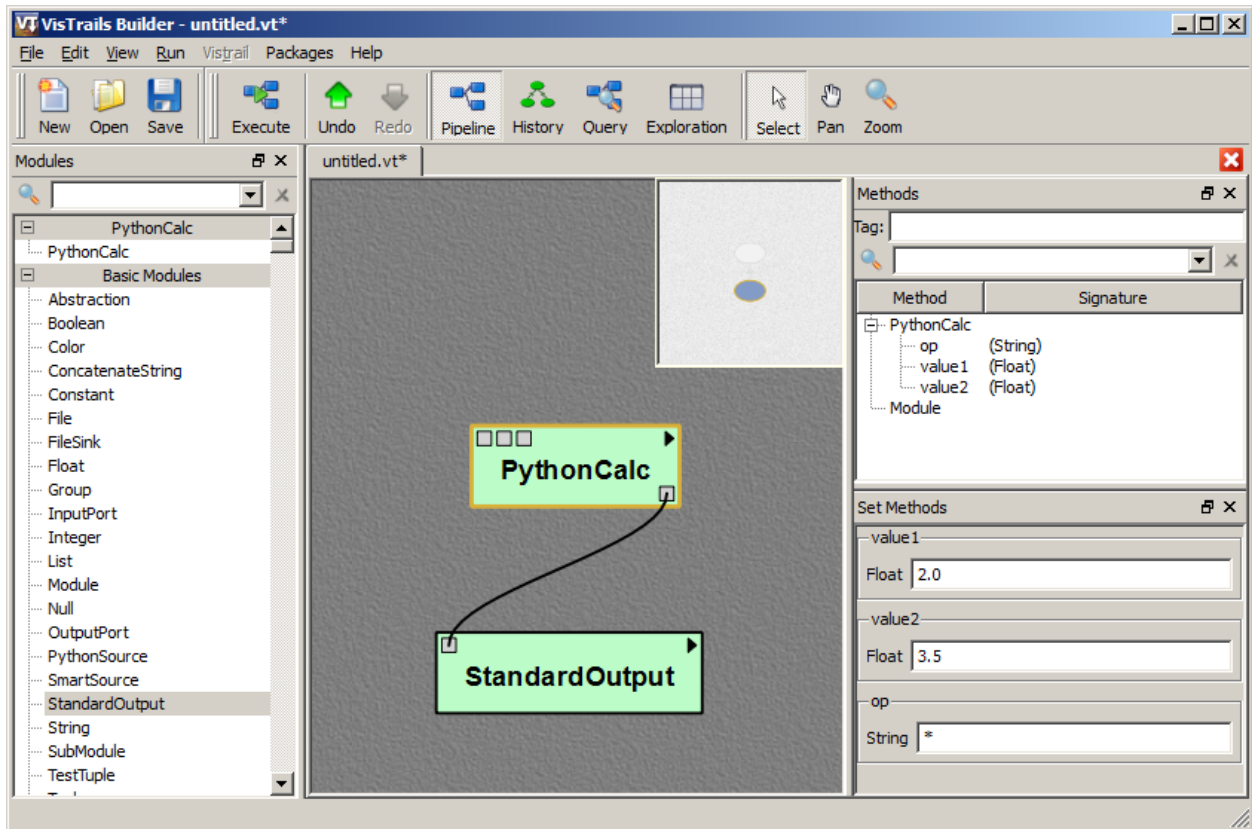


Figure 17.2: A simple workflow that uses PythonCalc, a user-defined module.

ultimately derives from the `Module` class, which is defined in `core.modules.vistrails_module`. Each module must implement a `compute` method that takes no extra parameters, such as on Line 11. This method represents the actual computation that happens in a module. This computation typically involves getting the necessary input and generating the output. We will now see how that works.

Line 12 shows how to extract input from a port. Specifically, we're getting the values passed to input ports `value1` and `value2`. We then perform some operation with these values, and need to report the output on an output port, so that it is available for downstream modules. This is done on Line 15, where the result is set to port `value`. Let us now look more carefully at the remainder of the class definition. Notice that developers are allowed to define extra helper methods, for example the `op` method on Line 17. These helper methods can naturally use the ports API. The other important feature of the `op` method is *error checking*. `PythonCalc` requires a string that represents the operation to be performed with the two numbers. If the string is invalid, it signals an error by simply raising a Python exception, `ModuleError`, that is provided in `core.modules.vistrails_module`. This exception expects two parameters: the module that generated the exception (typically `self`) and a string describing the error. In the Pipeline view, this error message is displayed in the tooltip that appears when the user "mouses over" the `PythonCalc` module icon. That is all that it takes in terms of module behavior. The rest of the code is meant to interact with VisTrails, and let the system know about the modules and ports being exposed. To do that, you must provide a function called `initialize` in the main body of the package file (the function starting on Line 27). The first thing is usually to register the module itself, such as on Line 33. Then, we need to tell VisTrails about the input and output ports we want to expose. Input ports are set with the `add_input_port` method in the registry, and output ports with `add_output_port`. These calls take three parameters. The first parameter is the module you're adding a new port to. The second one is simply the name of the port, and the third one is a description of the parameter. In most cases, this is just a pair, where the first element is a VisTrails module representing the module type being passed, and the second element is a string describing it. Notice that the types being used are VisTrails modules (Line 35), and not Python types.

That is it — you have successfully created a new package and module. From now on, we will look at more complicated examples, and more advanced features of the package mechanism.

17.4 Creating Reloadable Packages

When creating or making changes to packages, it is often desirable to reload the package without having to restart VisTrails. To create a package that is reloadable, users should create a new directory for the package in `userpackages` directory. This new directory should have the same name as the package and should contain an `__init__.py` file and an `init.py` file. The `identifier`, `name`, `version`, `configuration`, and `package_dependencies` fields/methods should be in `__init__.py`. An example of `__init__.py` from Vistrails' `pylab` package follows.

```

1 identifier = 'edu.utah.sci.vistrails.matplotlib'
2 name = 'matplotlib'
3 version = '0.9.0'
4
5 def package_dependencies():
6     import core.packagemanager
7     manager = core.packagemanager.get_package_manager()
8     if manager.has_package('edu.utah.sci.vistrails.spreadsheet'):
9         return ['edu.utah.sci.vistrails.spreadsheet']
10    else:
11        return []
12
13 def package_requirements():
14    import core.requirements
15    if not core.requirements.python_module_exists('matplotlib'):
16        raise core.requirements.MissingRequirement('matplotlib')
17    if not core.requirements.python_module_exists('pylab'):
18        raise core.requirements.MissingRequirement('pylab')
```

Imports (excluding `core.configuration`), other class definitions, and the initialize method should be in the `init.py` file. Finally, to reload a package, select the `reload` button from the Preferences dialog's `Module Packages` tab.

Note

To make the previous example *A Simple Example* reloadable, rather than having just one file `pythoncalc.py`, one would have a `pythoncalc` directory with the “version”, “name”, and “identifier” lines in `__init__.py` and all other lines in `init.py`.

17.5 Wrapping Command-line tools

Many existing programs are readily available through a command-line interface. Also, many existing workflows are first implemented through scripts, which work primarily with command-line tools. This section describes how to wrap command-line applications so they can be used with VisTrails. We will use as a running example the `afront` package, which wraps `afront`, a command-line program for generating 3D triangle meshes.¹ We will wrap the basic functionality in three different modules: `Afront`, `AfrontIso`, and `MeshQualityHistogram`.

Each of these modules will be implemented by a Python class, and they will all invoke the `afront` binary. `Afront` is the base execution module, and `AfrontIso` requires extra parameters on top of the original ones. Because of this, we will implement `AfrontIso` as a subclass of `Afront`. `MeshQualityHistogram`, however, requires entirely different parameters, and so will not be a subclass of `Afront`. A first attempt at writing this package might look something like this:

```
1 from core.modules.vistrails_module import Module
2 ... # other import statements
3
4 name = "Afront"
5 version = "0.1.0"
6 identifier = "edu.utah.sci.vistrails.afront"
7
8 class Afront(Module):
9     def compute(self):
10         ... # invokes afront
11
12 class AfrontIso(Afront):
13     def compute(self):
14         ... # invokes afront with additional parameters
15
16 class MeshQualityHistogram(Module):
17     def compute(self):
18         ... # invokes afront with completely different parameters
19
20 def initialize():
21     ...
```

17.5.1 Class Mixins

While this approach is a good start, it does require significant duplication of effort. Each module must contain code to invoke the `afront` binary and pass it some parameters. Since this functionality is required by all three modules,

¹ This package is not included in binary distributions of VisTrails, but is available in the source code distribution. The stand-alone `Afront` utility is available at <http://afront.sourceforge.net>.

we would like to put this code in a separate class called, say, `AfrontRun`, and let each of our modules inherit from it. `AfrontRun` itself is not a module, and thus does not extend the `Module` class. So our three modules will inherit from *both* `AfrontRun` *and* `Module`. Helper classes such as this are often referred to as *mixin classes*.² *hi* hello

```

1 from core.modules.vistrails_module import Module, ModuleError
2 from core.system import list2cmdline
3 import os
4
5 class AfrontRun(object):
6     _debug = False
7     def run(self, args):
8         cmd = ['afront', '-nogui'] + args
9         cmdline = list2cmdline(cmd)
10        if self._debug:
11            print cmdline
12        result = os.system(cmdline)
13        if result != 0:
14            raise ModuleError(self, "Execution failed")
15
16 class Afront(Module, AfrontRun):
17     ...
18
19 class MeshQualityHistogram(Module, AfrontRun):
20     ...

```

Now every module in the `afront` package has access to `run()`. The other new feature in this snippet is `list2cmdline`, which turns a list of strings into a command line. It does this in a careful way (protecting arguments with spaces, for example). Notice that we use a call to a shell (`os.system()`) to invoke `afront`. This is frequently the easiest way to get third-party functionality into VisTrails.

17.5.2 Package Configuration

There are two obvious shortcomings in the way `run()` is implemented. First, the code assumes `afront` is available in the system path, which might not be true in practice. Second, the debugging variable is inaccessible from the GUI, where it could be very useful. VisTrails provides a way to configure a package through the VisTrails Preferences dialog. It is very simple to provide your own configuration; just add a configuration attribute to your package, as follows:

```

1 from core.configuration import ConfigurationObject
2 from core.modules.vistrails_module import Module, ModuleError
3 from core.system import list2cmdline
4 import os
5
6 configuration = ConfigurationObject(path=(None, str),
7                                   debug=False)
8
9 class AfrontRun(object):
10
11     def run(self, args):
12         if configuration.check('path'): # there's a set directory
13             afront_cmd = configuration.path + '/afront'
14         else: # Assume afront is on path
15             afront_cmd = 'afront'
16         cmd = [afront_cmd, '-nogui'] + args

```

² Programmers who are more comfortable with single-inheritance languages like Java and C# may be unfamiliar with mixins. A mixin class is similar to an *interface* in Java or C#, except that a mixin provides an implementation as well. Python's support for multiple inheritance allows subclasses to "import" functionality as needed from a mixin class, without artificially cluttering the base class's interface.

```

17     cmdline = list2cmdline(cmd)
18     if configuration.debug:
19         print cmdline
20     ...
21     ...

```

Let us first look at how to specify configuration options. Named arguments passed to the `ConfigurationObject` constructor (Lines 6 and 7) become attributes in the object. If the attribute has a default value, simply pass it to the constructor. If the attribute should be unset by default, pass the constructor a pair whose first element is `None` and second element is the *type* of the expected parameter. Currently, the valid types are `bool`, `int`, `float` and `str`.

To use the configuration object in your code, you can simply access the attributes (as on line 18). This is fine when there is a default value set for the attribute. In the case of `path`, however, the absence of a value is encoded by a tuple (`None`, `str`), so using it directly is inconvenient. That is where the `check()` method comes in (line 12). It returns `False` if there is no set value, and returns the value otherwise.

Perhaps the biggest advantage of using a configuration object is that the values can be changed through a GUI, and they are persistent across VisTrails sessions. To configure a package, open the Preferences menu (VisTrails → Preferences on Mac OS X, or Edit → Preferences on other platforms). Then, select the package you want to configure by clicking on it (a package must be enabled to be configurable). If the `Configure` button is disabled, it means the package does not have a configuration object. When you do click `Configure`, a dialog like the one in Figure *Configuration window for a package that provides a configuration object*. will appear.

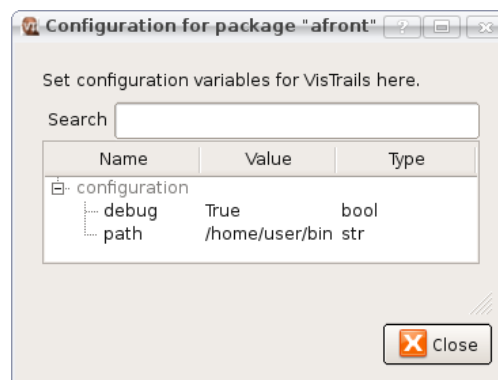


Figure 17.3: Configuration window for a package that provides a configuration object.

To edit the value for a particular field, double-click on it, and change the value. The values set in this dialog are persistent across VisTrails sessions, being saved on a per-user basis.

17.5.3 Temporary File Management

Command-line programs typically generate files as outputs. On complicated pipelines, many files get created and passed to other modules. To facilitate the use of files as communication objects, VisTrails provides basic infrastructure for temporary file management. This way, package developers do not have to worry about file ownership and lifetimes.

To use this infrastructure, it must be possible to tell the program being called which filename to use as output. VisTrails can accommodate some filename requirements (in particular, specific filename extensions might be important in Windows environments, and these can be set), but it must be possible to direct the output to a certain filename.

We will use `Afront`'s `compute()` method to illustrate the feature.

```

1     ...
2     class Afront(Module, AfrontRun):
3

```

```

4     def compute(self):
5         o = self.interpreter.filePool.create_file(suffix='.m')
6         args = []
7         if not self.hasInputFromPort("file"):
8             raise ModuleError(self, "Needs input file")
9         args.append(self.getInputFromPort("file").name)
10        if self.hasInputFromPort("rho"):
11            args.append("-rho")
12            args.append(str(self.getInputFromPort("rho")))
13        if self.hasInputFromPort("eta"):
14            args.append("-reduction")
15            args.append(str(self.getInputFromPort("eta")))
16        args.append("-outname")
17        args.append(o.name)
18        args.append("-tri")
19        self.run(args)
20        self.setResult("output", o)
21    ...

```

Line 5 shows how to create a temporary file during the execution of a pipeline. There are a few new things happening, so let us look at them one at a time. Every module holds a reference to the current *interpreter*, the object responsible for orchestrating the execution of a pipeline. This object has a `filePool`, which is what we will use to create a pipeline, through the `create_file` method. This method takes optionally a named parameter `suffix`, which forces the temporary file that will be created to have the right extension.

The file pool returns an instance of `basic_modules.File`, a module that is provided by the basic VisTrails packages. There are two important things you should know about `File`. First, it has a `name` attribute that stores the name of the file it represents. In this case, it is the name of the recently-created temporary file. This allows you to safely use this file when calling a shell, as seen on Line 17. The other important feature is that it can be passed directly to an output port, so that this file can be used by subsequent modules. This is shown on Line 20.

The above code also introduces the boolean function `hasInputFromPort` (see Lines 7, 10, and 13). This is a simple error-checking function that verifies that the port has incoming data before the program attempts to read from it. It is considered good practice to call this function before invoking `getInputFromPort` for any input port.

Accommodating badly-designed programs Even though it is considered bad design for a command-line program not to allow the specification of an output filename, there do exist programs that lack this functionality. In this case, a possible workaround is to execute the command-line tool, and move the generated file to the name given by VisTrails.

17.6 Interfacing with the VisTrails Menu

As we saw in Section *Package Configuration*, using the `ConfigurationObject` class is one way to “hook” your custom package into the VisTrails GUI. However, this is not the only way to integrate your package with the user interface. VisTrails also supports a mechanism whereby your package can add new options underneath the Packages menu (Figure *Packages can integrate their own commands into the main VisTrails menu.*).

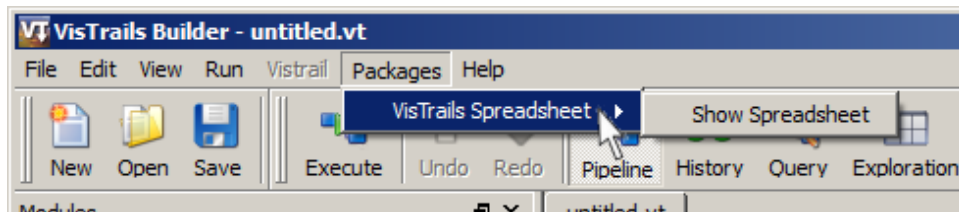


Figure 17.4: Packages can integrate their own commands into the main VisTrails menu.

This is done by adding a function named `menu_items` to your main package file. This function takes no parameters, and should return a tuple of pairs for each new menu item to be added. The first element of each pair is the label of the menu item as it should appear in the GUI. The second element of the pair is the name of the callback function to be invoked when the user selects that menu item.

As an example, we include below the implementation of `menu_items` from the VisTrails Spreadsheet package:

```
1 def menu_items():
2     """menu_items() -> tuple of (str,function)
3     It returns a list of pairs containing text for the menu and a
4     callback function that will be executed when that menu item is selected.
5     """
6
7     def show_spreadsheet():
8         spreadsheetWindow.show()
9         lst = []
10        lst.append(("Show Spreadsheet", show_spreadsheet))
11    return tuple(lst)
```

Writing your own `menu_items` function is straightforward; simply use the provided example as a basis, and substitute labels and callback functions as appropriate for your specific module. Although the Spreadsheet package currently only implements one new menu option, you are free to add as many as you see fit; just append additional pairs to the list (see Line 10 of the example code) before the function returns.

The Packages menu is organized hierarchically, as illustrated in Figure *Packages can integrate their own commands into the main VisTrails menu.* Each package that contributes a `menu_items` function will receive an entry in the Packages menu. The actual menu items for each package will appear in a submenu.

17.7 Interpackage Dependencies

When creating more sophisticated VisTrails packages, you might want to create a new module that requires a module *from another package*. For example, using modules from different packages as input ports, or even subclassing modules from other packages, require management of interpackage dependencies. VisTrails needs to know about these so that packages can be initialized in the correct order. To specify these dependencies, you should add a function named `package_dependencies` to your package. This function should return a list containing the identifier strings of the required packages.

As an example of this function's usage, let's take a look at a (simplified) code segment from the VTK package, which is included in the standard VisTrails distribution:

```
1 def package_dependencies():
2     return ['edu.utah.sci.vistrails.spreadsheet']
```

As you can see, the `package_dependencies` function is quite straightforward; it simply returns a list of the identifiers for the packages required by the VTK package. In this case, the list contains just a single string, as the VisTrails Spreadsheet is the only package dependency for the VTK package.

The simple approach taken by the above code works well for the majority of cases, but in practice you may want to add some error-checking to your `package_dependencies` function. This allows VisTrails to recover gracefully in the unlikely event that the Spreadsheet package is missing. Below is the complete `package_dependencies` function for the VTK package:

```
1 def package_dependencies():
2     import core.packagemanager
3     manager = core.packagemanager.get_package_manager()
4     if manager.has_package('edu.utah.sci.vistrails.spreadsheet'):
5         return ['edu.utah.sci.vistrails.spreadsheet']
```

```

6     else:
7         return []

```

The above code segment also demonstrates the VisTrails API function `has_package` which simply verifies that a given package exists in the system.

17.8 Package Requirements

In Section *Interpackage Dependencies*, we saw how packages can depend on other packages. However, some packages may also depend on the presence of external libraries (in the form of Python modules) or executable files in order to run correctly.

17.8.1 Required Python Modules

To check for the presence of a required Python module, you should add a function named `package_requirements` to your package. This function need not return any value; however it may raise exceptions or output error messages as necessary. Here is an example of the syntax of the `package_requirements` function, taken from the VisTrails VTK package:

```

1 def package_requirements():
2     import core.requirements
3     if not core.requirements.python_module_exists('vtk'):
4         raise core.requirements.MissingRequirement('vtk')
5     if not core.requirements.python_module_exists('PyQt4'):
6         print 'PyQt4 is not available. There will be no interaction',
7         print 'between VTK and the spreadsheet.'
8     import vtk

```

A key element of `package_requirements` is the use of the function `python_module_exists` (see Lines 3 and 5), which checks whether a given module has been installed in your local Python system.

17.8.2 Required Executables

As explained in Section *Wrapping Command-line tools*, a common motivation for writing new VisTrails modules is to wrap existing command-line tools. To this end, the VisTrails API provides a function called `executable_file_exists` which checks for the presence of specific executables on the path.

Here is an example of its usage, taken from the `initialize` function of the `ImageMagick` package. This package is included in the standard VisTrails distribution. The following code snippet checks to see if `convert`, a command-line program associated with the *ImageMagick* suite of graphics utilities, is on the path.

```

import core.requirements

...

if (not core.requirements.executable_file_exists('convert')):
    raise core.requirements.MissingRequirement("ImageMagick suite")

```

Note that this function is not strictly required in order to wrap third party executables into a module. Using a `Configuration` object (see Section *Package Configuration*) that lets the user specify the path to an executable may be a cleaner solution.

For additional information or examples of any of the functions described above, please refer to the VisTrails source code or contact the VisTrails development team.

17.9 Interaction with Caching

VisTrails provides a caching mechanism, in which portions of pipelines that are common across different executions are automatically shared. However, some modules should not be shared. Caching control is therefore up to the package developer. By default, caching is enabled. So a developer that doesn't want caching to apply must make small changes to the module. For example, look at the StandardOutput module:

```
from core.modules.vistrails_module import Module, newModule, \
    NotCacheable, ModuleError
(...)
class StandardOutput(NotCacheable, Module):
    """StandardOutput is a VisTrails Module that simply prints the
    value connected on its port to standard output. It is intended
    mostly as a debugging device."""

    def compute(self):
        v = self.getInputFromPort("value")
        print v
```

By subclassing from NotCacheable and Module (or one of its subclasses), we are telling VisTrails not to cache this module, or anything downstream from it.

VisTrails also allows a more sophisticated decision on whether or not to use caching. To do that, a user simply overrides the method `is_cacheable` to return the appropriate value (the default implementation returns `True`). For example, in the *teem* <<http://teem.sourceforge.net/>> package, there's a module that generates a scalar field with random numbers. This is non-deterministic, so shouldn't be cached. However, this module only generates non-deterministic values in special occasions, depending on its input port values. To keep efficiency when caching is possible, while still maintaining correctness, that module implements the following override:

```
class Unu1op(Unu):
(...)
    def is_cacheable(self):
        return not self.getInputFromPort('op') in ['rand', 'nrand']
(...)
```

Notice that the module explicitly uses inputs to decide whether it should be cached. This allows reasonably fine-grained control over the process.

17.10 Customizing Modules and Ports

Here we will explore the options for registry initialization of modules and ports which was introduced in Section A *Simple Example*. There is a new syntax for specifying modules in packages:

```
1 _modules = [MyModule1, (MyModule2, {'option_name' : 'value'})]
```

Observe that `_modules` is assigned a list of modules to be registered, and module options can be provided as keyword arguments by specifying a tuple (class, kwargs). Similarly, ports are defined by providing a list of tuples of the form (portName, portSignature, optional=False, sort_key=-1). For example:

```
1 class MyModule(Module):
2     def compute(self):
3         pass
4
5     _input_ports = [('firstInput', String), ('secondInput', Integer, True)]
6     _output_ports = [('firstOutput', String), ('secondOutput', String)]
```


Notice that “String” and “Integer” were used for the portSignature instead of `edu.utah.sci.vistrails.basic:String` and `edu.utah.sci.vistrails.basic:Integer`. That is because the current package, `edu.utah.sci.vistrails.basic` is used by default.

Note

The old syntax (`reg.add_module(...)`, `reg.add_input_port(...)`, and `reg.add_output_port(...)`) is still supported.

17.10.1 Configuring Modules

Hierarchy and Visibility There are a few options that assist in the organization and display of modules: `namespace`, `abstract`, and `hide_descriptor`. The `namespace` option can be used to define a hierarchy in the module palette, which hierarchies can be nested through the use of the ‘|’ character. For example:

```
_modules = [MyModule1, (MyModule2, {'namespace': 'MyNamespace'})]
or
_modules = [MyModule1, (MyModule2, {'namespace': 'ParentNamespace|
    ChildNamespace'})]
```

The other options, `abstract` and `hide_descriptor` can be used to prevent modules from appearing in the module palette. `Abstract` is for use with modules that should never be instantiated in the workflow and will not add the item to the module palette. On the other hand, `hide_descriptor` will add the item to the palette, but hides it. This will prevent users from adding the module to a pipeline, but allow code to add it programmatically. To use either of these options, `abstract` or `hide_descriptor`, set it to `True`:

```
_modules = [AnotherModule, (InvisibleModule, {'abstract': True})]
or
_modules = [AnotherModule, (InvisibleModule, {'hide_descriptor': True})]
```

Defining Module Shapes and Colors VisTrails allows users to define custom colors and shapes to modules. This must be done at module registration time, by using the `moduleColor` and `moduleFringe` options. For example:

```
_modules = [(Afront, {'moduleColor' : (1.0, 0.0, 0.0),
    'moduleFringe' : [(0.0, 0.0),
        (0.2, 0.0),
        (0.2, 0.4),
        (0.0, 0.4),
        (0.0, 1.0)]})])]
```

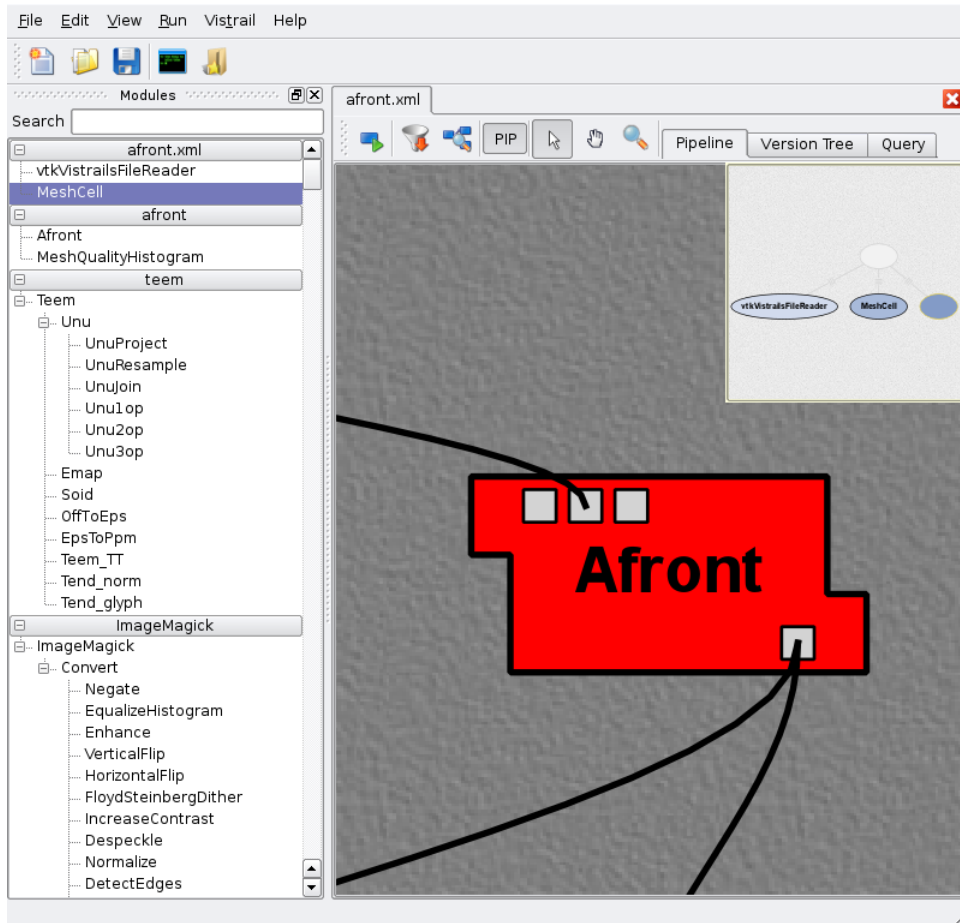
gives:

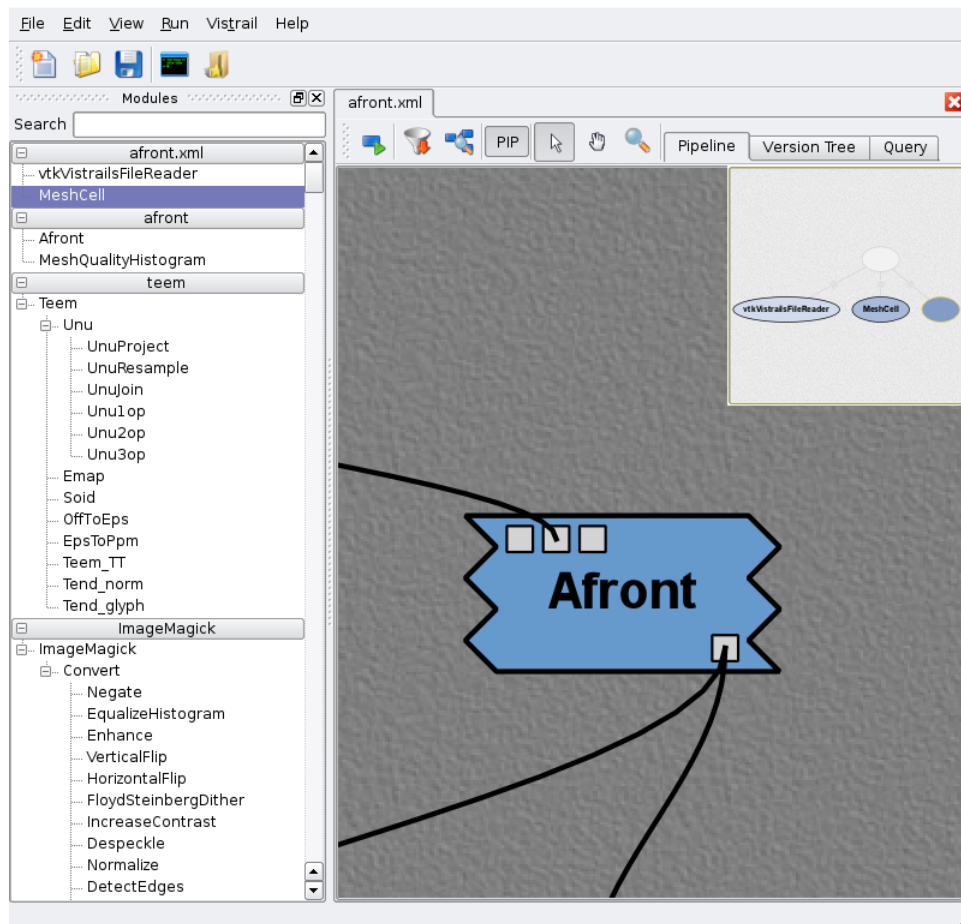
and

```
_modules = [(Afront, {'moduleColor': (0.4,0.6,0.8),
    'moduleFringe' : [(0.0, 0.0),
        (0.2, 0.0),
        (0.0, 0.2),
        (0.2, 0.4),
        (0.0, 0.6),
        (0.2, 0.8),
        (0.0, 1.0)]})])]
```

gives:

The `moduleColor` parameter must be a tuple of three floats between 0 and 1 that specify RGB colors for the module background, while `moduleFringe` is a list of pairs of floats that specify points as they go around a side of the module

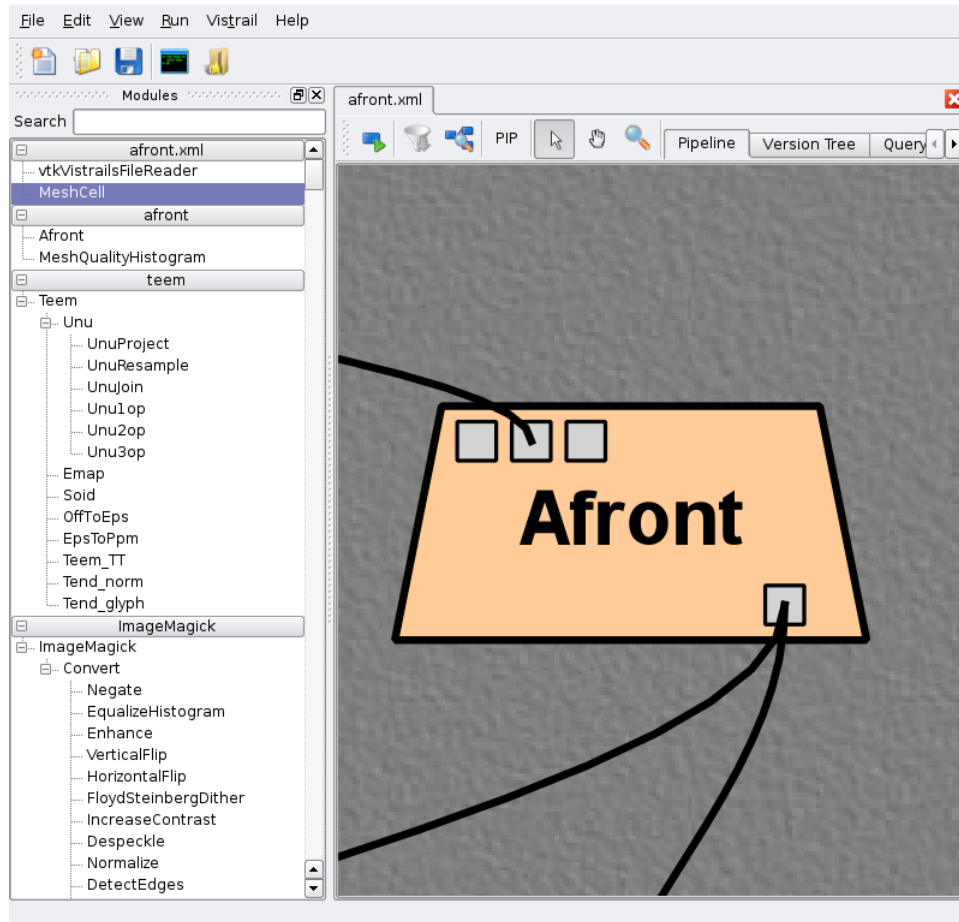




(the same one is used to go from the top-right corner to bottom-right corner, and from the bottom-left corner to the top-left one. If this is not enough, let the developers know!)

Alternatively, you can use different fringes for the left and right borders:

```
_modules = [(Afront, {'moduleColor': (1.0,0.8,0.6),
                    'moduleLeftFringe' : [(0.0, 0.0),
                                           (-0.2, 0.0),
                                           (0.0, 1.0)],
                    'moduleRightFringe' : [(0.0, 0.0),
                                           (0.2, 1.0),
                                           (0.0, 1.0)]})])]
```



17.10.2 Configuring Ports

Default Values and Labels In versions 1.4 and greater, package developers can add labels and default values for parameters. To add this functionality, you need to use the defaults and labels keyword arguments and pass the values as strings. For example,

```
1 class TestDefaults (Module):
2     _input_ports = [('f1', '(edu.utah.sci.vistrails.basic:Float, \
3                     edu.utah.sci.vistrails.basic:String)',
4                     {"defaults": str([1.23, "abc"]),
5                      "labels": str(["temp", "name"])}])
6     _modules = [TestDefaults]
```

or in the older syntax,

```

1 def initialize():
2     reg = core.modules.module_registry.get_module_registry()
3     reg.add_module(TestDefaults)
4     reg.add_input_port(TestDefaults, "f1", [Float, String],
5                          defaults=str([1.23, "abc"]),
6                          labels=str(["temp", "name"]))

```

Making a Port Optional To add a port that is optional, simply set optional to true:

```

_input_ports = [('MyPort', '(edu.utah.sci.vistrails.basic:String)',
                 {"optional": True})]

reg.add_input_port(MyModule, "MyPort",
                  (edu.utah.sci.vistrails.basic:String, 'MyPort Name'),
                  optional=True)

```

Multiple Inputs For compatibility reasons, we do need to allow multiple connections to an input port. However, most package developers should never have to use this, and so we do our best to hide it. The default behavior for getting inputs from a port, then, is to always return a single input.

If on your module you need multiple inputs connected to a single port, use the 'forceGetInputListFromPort' method. It will return a list of all the data items coming through the port. The spreadsheet package uses this feature, so look there for usage examples (vistrails/packages/spreadsheet/basic_widgets.py) **Port Types** To define ports to be of types that are not imported into the package, pass an identifier string as the portSignature:

```

<module_string> := <package_identifier>[:<namespace>|]<module_name>,
<port_signature> := (<module_string>*)

```

For example,

```

registry.add_input_port(MyModule, 'myInputPort', \
                       '(edu.utah.sci.vistrails.basic:String)')

```

or

```

_input_ports = [('myInputPort', '(edu.utah.sci.vistrails.basic:String)')]

```

Varying Output According to the Input There are a few ways to tackle this - each has its own benefits and pitfalls. Firstly, module connections do respect class hierarchies as we're familiar with in object oriented languages. For instance, A module can output a Constant of which String, Float, Integer, etc are specifications. In this way, you can have a subclass of something like MyData be passed out of the module and the connections will be established regardless of the sub-type. This is a bit dangerous though. Modules downstream of such a class may not really know how to operate on certain types derived from the super-class. Extreme care must be taken both when creating the modules as well as connecting them to prevent things like this from happening.

A second method that is employed in several different packages is the idea of a container class. For instance, the NumSciPy package uses a relatively generic container "Numpy Array" to encapsulate the data. Of course, these encapsulating objects can store dictionaries that other modules can easily access and understand how to operate on. Although this method is slightly more work, the benefits of a stricter typing of ports is beneficial - particularly upon interfacing with other packages that may depend on strongly typed constants (for example). **Determining Whether or Not a Module is Attached to an Output Port** The outputPorts dictionary of the base Module stores connection information. Thus, you should be able to check

```
("myPortName" in self.outputPorts)
```

on the parent module to check if there are any downstream connections from the port "myPortName". This might be used, for example, to only set results for output ports that will be used. **Note**, however, that the caching algorithm assumes that all outputs are set so adding a new connection to a previously unconnected output port will not work as

desired if that module is cached. For this reason, I would currently recommend making such a module not cacheable. Another possibility is overriding the `update()` method to check the `outputPorts` and set the `upToDate` flag if they are not equal. In a single, limited test, this seemed to work, but be warned that it is not fully tested. Here is an example:

```
1 class TestModule(Module):
2     _output_ports = [('a1', '(edu.utah.sci.vistrails.basic:String)'),
3                     ('a2', '(edu.utah.sci.vistrails.basic:String)')]
4     def __init__(self):
5         Module.__init__(self)
6         self._cached_output_ports = set()
7
8     def update(self):
9         if len(set(self.outputPorts) - self._cached_output_ports) > 0:
10            self.upToDate = False
11            Module.update(self)
12
13    def compute(self):
14        if "a1" in self.outputPorts:
15            self.setResult("a1", "test")
16        if "a2" in self.outputPorts:
17            self.setResult("a2", "test2")
18        self._cached_output_ports = set(self.outputPorts)
```

17.11 Generating Modules Dynamically

When wrapping existing libraries or trying to generate modules in a more procedural manner, it is useful to dynamically generate modules. In our work, we have created some shortcuts to make this easier. In addition, the list of modules can also be based on the package configuration. Here is some example code:

`__init__.py`

```
1 from core.configuration import ConfigurationObject
2
3 identifier = "edu.utah.sci.dakoop.auto_example"
4 version = "0.0.1"
5 name = "AutoExample"
6
7 configuration = ConfigurationObject(use_b=True)
```

`init.py`

The `expand_ports` and `build_modules` methods are functions to help the construction of the modules easier. The key parts are the `new_module` call and setting the `_modules` variable.

```
1 from core.modules.vistrails_module import new_module, Module
2
3 identifier = "edu.utah.sci.dakoop.auto_example"
4
5 def expand_ports(port_list):
6     new_port_list = []
7     for port in port_list:
8         port_spec = port[1]
9         if type(port_spec) == str: # or unicode...
10            if port_spec.startswith('('):
11                port_spec = port_spec[1:]
12            if port_spec.endswith(')'):
13                port_spec = port_spec[:-1]
```

```

14     new_spec_list = []
15     for spec in port_spec.split(','):
16         spec = spec.strip()
17         parts = spec.split(':', 1)
18         print 'parts:', parts
19         namespace = None
20         if len(parts) > 1:
21             mod_parts = parts[1].rsplit('|', 1)
22             if len(mod_parts) > 1:
23                 namespace, module_name = mod_parts
24             else:
25                 module_name = parts[1]
26             if len(parts[0].split('.')) == 1:
27                 id_str = 'edu.utah.sci.vistrails.' + parts[0]
28             else:
29                 id_str = parts[0]
30         else:
31             mod_parts = spec.rsplit('|', 1)
32             if len(mod_parts) > 1:
33                 namespace, module_name = mod_parts
34             else:
35                 module_name = spec
36             id_str = identifier
37         if namespace:
38             new_spec_list.append(id_str + ':' + module_name + \
39                                 ':' + namespace)
40         else:
41             new_spec_list.append(id_str + ':' + module_name)
42     port_spec = '(' + ','.join(new_spec_list) + ')'
43     new_port_list.append((port[0], port_spec) + port[2:])
44     print new_port_list
45     return new_port_list
46
47 def build_modules(module_descs):
48     new_classes = {}
49     for m_name, m_dict in module_descs:
50         m_doc = m_dict.get("_doc", None)
51         m_inputs = m_dict.get("_inputs", [])
52         m_outputs = m_dict.get("_outputs", [])
53         if "_inputs" in m_dict:
54             del m_dict["_inputs"]
55         if "_outputs" in m_dict:
56             del m_dict["_outputs"]
57         if "_doc" in m_dict:
58             del m_dict["_doc"]
59         klass_dict = {}
60         if "_compute" in m_dict:
61             klass_dict["compute"] = m_dict["_compute"]
62             del m_dict["_compute"]
63         m_class = new_module(Module, m_name, klass_dict, m_doc)
64         m_class._input_ports = expand_ports(m_inputs)
65         m_class._output_ports = expand_ports(m_outputs)
66         new_classes[m_name] = (m_class, m_dict)
67     return new_classes.values()
68
69 def initialize():
70     global _modules
71     def a_compute(self):

```

```
72     a = self.getInputFromPort("a")
73     i = 0
74     if self.hasInputFromPort("i"):
75         i = self.getInputFromPort("i")
76     if a == "abc":
77         i += 100
78     self.setResult("b", i)
79
80     module_descs = [("ModuleA", {"_inputs": [("a", "basic:String")],
81                                   "_outputs": [("b", "basic:Integer")],
82                                   "_doc": "ModuleA documentation",
83                                   "_compute": a_compute,
84                                   "_namespace": "Test"}),
85                    ("ModuleB", {"_inputs": [("a", "Test|ModuleA")],
86                                   "_outputs": [("b", "Test|ModuleA")],
87                                   "_doc": "ModuleB documentation"})
88
89     ]
90
91     if configuration.use_b:
92         _modules = build_modules(module_descs)
93     else:
94         _modules = build_modules(module_descs[:1])
95
96 _modules = []
```

17.12 For System Administrators

Most users will want to put their custom packages in their

```
~/ .vistrails/userpackages
```

directory, as described in Section *A Simple Example*. This makes the package available to the current user only. However, if you are a power user or a system administrator, you may wish to make certain packages available to all users of a VisTrails installation. To do this, copy the appropriate package files and/or directories to the

```
vistrails/packages
```

directory of the VisTrails distribution. The packages will be made visible to users the next time they launch VisTrails.

COMMAND-LINE ARGUMENTS

18.1 Starting VisTrails via the Command Line

VisTrails supports a number of command-line arguments that let you modify certain attributes and behaviors of the program. When invoking VisTrails from the command line, the arguments are placed after the “vistrails.py” filename. For example,

```
python vistrails.py -n
```

suppresses the VisTrails splash screen. Table *Command line arguments supported by VisTrails* contains a complete list of the command line switches supported by vistrails. Each command line switch has both a short form and a long form. The two forms are logically equivalent, and which one you use is a matter of personal preference. The short form consists of a single minus sign “-” followed by a single letter. The longer form uses two minus signs “--” followed by a descriptive word. For example, the above command to suppress the splash screen could have been written as:

```
python vistrails.py --nosplash
```

In addition to the explicit switches listed in Table *Command line arguments supported by VisTrails*, the VisTrails command line also lets you indicate the filename of the vistrail you wish to open. For example, assuming your “examples” directory is one level above your current working directory, this is how you would tell VisTrails to load the “lung.vt” example at startup:

```
python vistrails.py ../examples/lung.vt
```

Moreover, if you want VisTrails to start on a *specific version* of the pipeline within the vistrail, you can indicate that version’s tag name on the command line. The filename and version tag should be separated by a colon. For example, to start VisTrails with the `colormap` version of the “lung.vt” vistrail, use:

```
python vistrails.py ../examples/lung.vt:colormap
```

In the event that the version you want to open contains a space in its tag name, simply surround the entire “filename:tag” pair in double quotes. For example:

```
python vistrails.py "../examples/lung.vt:Axial View"
```

You can also open up multiple vistrails at once by listing more than one vistrail file on the command line. This causes the vistrails to be opened in separate tabs, just as if you had opened them via the GUI. For example:

```
python vistrails.py ../examples/lung.vt ../examples/head.vt
```

You can specify version tags in conjunction with multiple filenames. Here is an example of an elaborate command-line invocation that opens two vistrails and sets each one to a specific version:

```
python vistrails.py "../examples/lung.vt:Axial View"  
../examples/head.vt:bone
```

Note:

As of this writing, the VisTrails development team is refactoring the implementation of many of the command-line switches presented in Table *Command line arguments supported by VisTrails*. As such, depending on your version of VisTrails, the results you achieve may not match those described. For a list of known issues with the command line switches, please refer to the VisTrails website.

Table 18.1: Command line arguments supported by VisTrails.

Short form	Long form	Description
-h	--help	Print a help message and exit.
-S <i>/path</i>	--startup= <i>/path</i>	Set user configuration directory (default is <code>~/vistrails</code>)
-?		Print a help message and exit.
-v	--version	Print version information and exit.
-V <i>num</i>	--verbose= <i>num</i>	Set verbosity level (0–2, default=0, higher means more verbose).
-b	--noninteractive	Run in non-interactive (batch) mode.
-n	--nosplash	Do not display splash screen on startup.
-c <i>num</i>	--cache= <i>num</i>	Enable/disable caching (0 to disable, nonzero to enable. Default is enabled).
-m <i>num</i>	--movies= <i>num</i>	Set automatic movie creation on spreadsheet (0 or 1, default=1). Set this to zero to work around VTK bug with offscreen renderer and OpenGL texture3D mappers.
-s	--multiheads	Display the Builder and Spreadsheet on different screens (if available).
-x	--maximized	Maximize Builder and Spreadsheet windows at startup.
-l	--nologger	Disable logging.
-d	--debugsignals	Debug Qt Signals.
-a <i>params</i>	--parameters= <i>params</i>	Set workflow parameters (non-interactive mode only).
-e <i>dir</i>	--dumpcells= <i>dir</i>	Set directory to dump spreadsheet cells before exiting (non-interactive mode only).
-t <i>host</i>	--host= <i>host</i>	Set hostname or IP address of database server.
-r <i>port</i>	--port= <i>port</i>	Set database port.
-f <i>dbName</i>	--db= <i>dbName</i>	Set database name.
-u <i>userName</i>	--user= <i>userName</i>	Set database username.

18.2 Specifying a User Configuration Directory

In addition to the default `.vistrails` directory, VisTrails allows you to create and use additional configuration directories. First, you

```
python vistrails.py -S /path_to_new_directory/new_directory_name.
```

This will both create a new directory containing default configuration files and directories, and launch VisTrails, which will use the newly created files for configuration. The user is then free to add desired configurations to the new directory. Once a configuration directory exists, subsequent calls using the directory name (`python vistrails.py -S /path_to_directory/existing_directory`) will launch VisTrails using the ‘existing_directory’ for configuration and a new directory will not be created.

Note: If you would like to copy configuration directories, you must change the references in `copy_of_directory/startup.xml` to point to the new directory instead of the original.

18.3 Passing Database Parameters on the Command Line

As discussed in Chapter *Connecting to a Database*, VisTrails can read and write vistrails stored in a relational database as well as in a filesystem. VisTrails allows you to specify the name of the database server, the database name, the port number, and the username on the command line. This potentially saves you the trouble of filling out the same information on the database connection dialog. Note that, for security reasons, VisTrails does not allow you to include a database password on the command line; you must still type your password into the database connection dialog when VisTrails opens.

The last four rows of Table *Command line arguments supported by VisTrails* show the command-line switches that pertain to database connectivity. Be advised that these switches were designed primarily for use by VTL files (see Section *Using “Vistrail Link” Files*) and as such, are not necessarily user-friendly. In particular, these switches are ignored unless you also specify the vistrail ID and version name on the command line. For example, to open the `contour` version of the “`spx`” vistrail (whose ID is 5) from the database “`vistrails`” residing on the host “`vistrails.sci.utah.edu`” with a username of “`vistrails`”:

```
python vistrails.py -t vistrails.sci.utah.edu -f vistrails -u
vistrails 5:contour
```

Once VisTrails opens, you will be prompted to enter the password. Upon successful authentication, the vistrail is loaded from the database and opened to the pipeline corresponding to the specified version.

18.3.1 Using “Vistrail Link” Files

As discussed in Chapter *Connecting to a Database*, one of the advantages of storing your vistrails on a database is that you can collaborate with others without having to pass around a `.vt` file or force all users to use a shared filesystem. A disadvantage is that you need to remember the parameters with which to connect to the database. Using a “Vistrail Link” (VTL) file reduces this inconvenience, and also eliminates the need to include the associated command-line switches to connect to the database.

A VTL is a very small text (XML) file that contains the parameters required to load a vistrail from a database. VTL files are intended for use with a VisTrails-enabled wiki. You can open a VTL either by saving the file and passing its filename to the command line, or by configuring your web browser to do this for you. Here is the syntax for using a VTL file on the command line:

```
python vistrails.py sample.vtl
```

Internally, VisTrails parses the VTL file and loads the vistrail from the database exactly as if you had included its full parameter list on the command line.

Note:

VTL is a relatively new feature of VisTrails, and as such is neither fully developed nor completely documented. Please contact the VisTrails development team with any bug reports and/or suggestions.

18.4 Running VisTrails in Batch Mode

Although VisTrails is primarily intended to be run as an interactive, graphical client application, it also supports non-interactive use. VisTrails can thus be invoked programmatically, eg as part of a shell script. You can tell VisTrails to

start in non-interactive mode by using the “-b” or “-noninteractive” command line switch when launching vistrails. ¹

Running VisTrails in non-interactive mode has little effect, however, without an additional command line argument indicating which vistrail to load. Since we are running VisTrails as part of a batch process, it only makes sense to execute vistrails whose output is something tangible, such as a file. A vistrail whose only output is an interactive rendering in a `VTKCell`, for instance, would not be well-suited for running in batch mode.

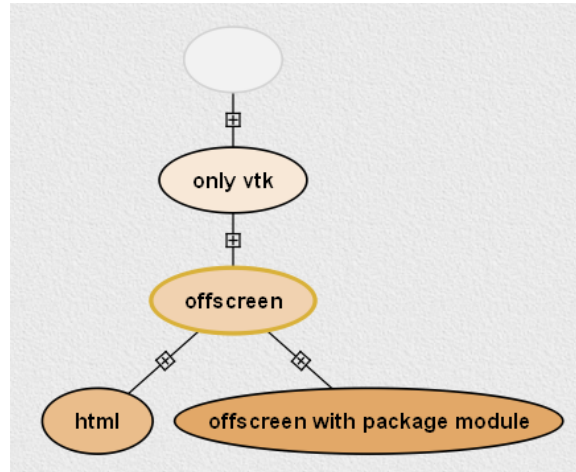


Figure 18.1: The different versions of the “offscreen.vt” vistrail offer various forms of output.

Consider the following example. The “offscreen.vt” vistrail (included in the “examples” directory) has a variety of output options, depending on which version you select in the History view (Figure *The different versions of the “offscreen.vt” vistrail offer various forms of output.*). The version tagged `only vtk` displays its output as an interactive VTK rendering. The version tagged `html` creates a simple web page in the Spreadsheet. The `offscreen` version, however, outputs an image file named “image.png”. Since its output (a file) can be easily accessed outside of VisTrails, this version is an ideal candidate for running in batch mode. To try it, invoke VisTrails as shown, specifying both the name of the vistrail file and the desired version:

```
python vistrails.py -b ../examples/offscreen.vt:offscreen
```

As you would expect, this command runs to completion without opening any windows. Instead, it silently loads the requested pipeline, executes it, and closes. Assuming it ran correctly, this pipeline should have created a file named “image.png” in the current directory. When you view this file, it should resemble the picture in Figure *Running the offscreen version of “offscreen.vt” in batch mode produces an image named “image.png”.*

18.4.1 Running a Specific Workflow in Batch Mode

To run a specific workflow in batch mode, call VisTrails with the following options:

```
python vistrails.py -b path_to_vistrails_file:pipeline
```

where pipeline can be a version **tag name** or version **id**.

Note

If you downloaded the MacOS X bundle, you can run VisTrails from the command line via the following commands in the terminal. Change the current directory to wherever VisTrails was installed (often `/Applications`), and then type: `Vistrails.app/Contents/MacOS/vistrails [<cmd_line_options>]`

¹ The parameter “-b” stands for “batch.” In this chapter, we use the terms “batch mode” and “non-interactive mode” synonymously.

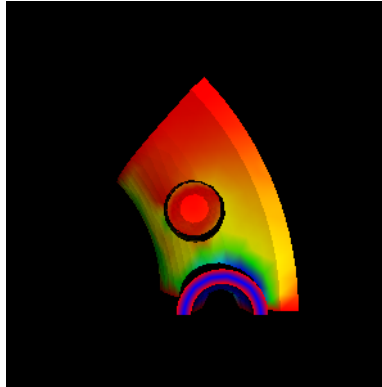


Figure 18.2: Running the `offscreen` version of “`offscreen.vt`” in batch mode produces an image named “`image.png`”.

18.4.2 Running a Workflow with Specific Parameters

An alias is a name assigned to a parameter that allows you to reference that parameter in batch mode. An alias is created by clicking on the type of an existing parameter in VisTrails, then entering a name for it.

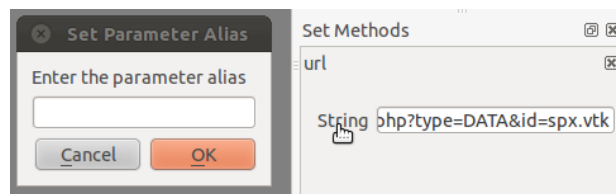


Figure 18.3: Example of creating an alias

Users can change workflow parameters that have an alias through the command line.

For example, `offscreen` pipeline in `offscreen.vt` always creates the file called `image.png`. If you want generate it with a different filename:

```
python vistrails.py -b ../examples/offscreen.vt:offscreen
-a"filename=other.png"
```

`filename` in the example above is the alias name assigned to the parameter in the value method inside the `String` module. When running a pipeline from the command line, VisTrails will try to start the spreadsheet automatically if the pipeline requires it. For example, this other execution will also start the spreadsheet (attention to how `$` characters are escaped when running on bash):

```
python vistrails.py -b ../examples/head.vt:aliases -a"isovalue=30\$\&\$diffuse_color=0.8,
0.4, 0.2"
```

You can also execute more than one pipeline on the command line:

```
python vistrails.py -b ../examples/head.vt:aliases ../examples/spx.vt:spx \
-a"isovalue=30"
```

Use the `-a` parameter only once regardless the number of pipelines.

18.4.3 Accessing a Database in Batch Mode

As discussed in Section *Passing Database Parameters on the Command Line*, you can specify most of the parameters of your database connection on the command line, but the password must be entered through the GUI. This poses a problem for running VisTrails in non-interactive mode, since no database connection dialog will be opened. If your batch process needs to access vistrails stored on a database, the current workaround is to create a special account on the database (probably one with read-only access) that does *not* require a password, and use this account for connecting to the database in batch mode.

18.4.4 Using VisTrails as a Server

Using the VisTrails server mode, it is possible to execute workflows and control VisTrails through another application. For example, the CrowdLabs Web portal (<http://www.crowdlabs.org>) accesses a VisTrails sever to execute workflows, retrieve and display vistrail trees and workflows.

The way you access the server is by doing XML-RPC calls. In the current VisTrails release, we include a set of PHP scripts that can talk to a VisTrails server instance. They are in “extensions/http” folder. The files are reasonably well documented. Also, it should be not difficult to create python scripts to access the server (just use xmlrpclib module).

Note that the VisTrails server requires the provenance and workflows to be in a database. More detailed instructions on how to setup the server and the database are available here:

http://www.crowdlabs.org/site_media/static/dev_docs/vistrails_server_setup.html

http://www.crowdlabs.org/site_media/static/dev_docs/vistrails_database_setup.html

If what you want is just to execute a series of workflows in batch mode, a simpler solution would be to use the VisTrails client in batch mode (see Section *Running VisTrails in Batch Mode*).

18.5 Executing Workflows in Parallel

The VisTrails server can only execute pipelines in parallel if there’s more than one instance of VisTrails running. The command

```
self.rpcserver=ThreadedXMLRPCServer((self.temp_xml_rpc_options.server,
self.temp_xml_rpc_options.port))
```

starts a multithreaded version of the XML-RPC server, so it will create a thread for each request received by the server. The problem is that Qt/PyQT doesn’t allow these multiple threads to create GUI objects. Only the main thread can. To overcome this limitation, the multithreaded version can instantiate other single threaded versions of VisTrails and put them in a queue, so workflow executions and other GUI-related requests, such as generating workflow graphs and history trees can be forwarded to this queue, and each instance takes turns in answering the request. If the results are in the cache, the multithreaded version answers the requests directly.

Note that this infrastructure works on Linux only. To make this work on Windows, you have to create a script similar to `start_vistrails_xvfb.sh` (located in the scripts folder) where you can send the number of other instances via command-line options to VisTrails. The command line options are:

```
python vistrails_server.py -T <ADDRESS> -R <PORT> -O<NUMBER_OF_OTHER_VISTRAILS_INSTANCES>
[-M] &
```

If you want the main vistrails instance to be multithreaded, use the `-M` at the end.

After creating this script, update function `start_other_instances` in `vistrails/gui/application_server.py` lines 1007-1023 and set the script variable to point to your script. You may also have to change the arguments sent to your script (line 1016: for example, you don’t need to set a virtual display). You will need to change the path to the `stop_vistrails_server.py` script (on line 1026) according to your installation path.

18.6 Finding Methods Via the Command Line

We have tried to make some methods more accessible in the console via an api. You can import the api via `import api` in the console and see the available methods with `dir(api)`. To open a vistrail:

```
import api
api.open_vistrail_from_file('/Applications/VisTrails/examples/terminator.vt')
```

To execute a version of a workflow, you currently have to go through the controller:

```
api.select_version('Histogram')
api.get_current_controller().execute_current_workflow()
```

Currently, only a subset of VisTrails functionality is directly available from the api. However, since VisTrails is written in python, you can dig down starting with the `VistrailsApplication` or `controller` object to expose most of our internal methods. If you have suggestions for calls to be added to the api, please let us know.

One other feature that we're working on, but is still in progress is the ability to construct workflows via the console. For example:

```
vtk = load_package('edu.utah.sci.vistrails.vtk')
vtk.vtkDataSetReader() # adds a vtkDataSetReader module to the pipeline
# click on the new module
a = selected_modules()[0] # get the one currently selected module
a.SetFile('/vistrails/examples/data/head120.vtk') # sets the SetFile\
parameter for the data set reader
b = vtk.vtkContourFilter() # adds a vtkContourFilter module to the\
pipeline and saves to var b
b.SetInputConnection0(a.GetOutputPort0()) # connects a's GetOutputPort0\
port to b's SetInputConnection0
```


ACCESSING THE EXECUTION LOG

The code responsible for storing execution information is located in the “core/log” directories, and the code that generates much of that information is in “core/interpreter/cached.py”. Modules can add execution-specific annotations to provenance via `annotate()` calls during execution, but much of the data (like timing and errors) is captured by the `LogController` and `CachedInterpreter` (the execution engine) objects. To analyze the log from a `vistrail (.vt)` file, you might have something like the following:

```
import core.log.log
import db.services.io

def run(fname):
    # open the .vt bundle specified by the filename "fname"
    bundle = db.services.io.open_vistrail_bundle_from_zip_xml(fname) [0]
    # get the log filename
    log_fname = bundle.vistrail.db_log_filename
    if log_fname is not None:
        # open the log
        log = db.services.io.open_log_from_xml(log_fname, True)
        # convert the log from a db object
        core.log.log.Log.convert(log)
        for workflow_exec in log.workflow_execs:
            print 'workflow version:', workflow_exec.parent_version
            print 'time started:', workflow_exec.ts_start
            print 'time ended:', workflow_exec.ts_end
            print 'modules executed:', [i.module_id
                                        for i in workflow_exec.item_execs]

if __name__ == '__main__':
    run("some_vistrail.vt")
```

You should be able to see what information is available by looking at the “core/log” classes.

EXAMPLE: ITK

20.1 Introduction to ITK

The *Insight Toolkit*,¹ or ITK, is an open-source software system initially designed to support the Visible Human Project [C1]. ITK is under continual development, being updated to employ cutting-edge segmentation and registration algorithms for multiple dimensions. For more information about ITK, please refer to [C2] and [C3].

In order to facilitate the implementation of processing mechanisms specific to the medical imaging community, ITK provides a robust set of general purpose image processing tools. These image processing tools are available to users through the standard ITK Filter interface. Although ITK is implemented in C++, through the use of CMake² and CableSwig³, the functionality of ITK is made available to languages such as TCL, Java, and Python. In addition, much of the functionality of ITK is also available in a VisTrails package. The ITK package is not currently included in the binary distribution of VisTrails, but it may be downloaded separately from the VisTrails website.

Note: The WrapITK library, upon which the VisTrails ITK package depends, is well-tested on Linux and Mac OS X platforms. It is, however, known to have issues under Windows.

20.2 Preparing ITK

At the time of this writing, the latest stable release of ITK is 3.6.0. In order to incorporate the functionality of ITK into the VisTrails system, it first must be built and installed. In the following sections, we will describe in detail the process of downloading, building, and installing ITK and all the required components needed to use it.

20.2.1 Downloading ITK

ITK can be downloaded in either source tarballs or via public CVS access to the ITK source repository. The following instructions take advantage of the CVS source repository; however, source tarballs can be downloaded from: <http://www.itk.org/>

These instructions can be found, in part, at the ITK website. To use CVS, you must have a CVS client installed on your system. To download the ITK library, issue the following commands:

```
cvs -d :pserver:anonymous@www.itk.org:/cvsroot/Insight login
password: insight
```

```
cvs -d :pserver:anonymous@www.itk.org:/cvsroot/Insight co Insight
```

¹ The Insight Toolkit is sometimes referred to by the longer name *Insight Segmentation and Registration Toolkit*.

² CMake cross-platform make system. <http://www.cmake.org>

³ CableSwig Interface generator. <http://www.itk.org/HTML/CableSwig.html>

Change directory into the newly created `Insight/Utilities` directory and issue the following command:

```
cvs -d :pserver:anonymous@public.kitware.com:/cvsroot/CableSwig co CableSwig
```

This checkout includes `CableSwig` in the ITK system, allowing it to be built automatically during compilation of ITK itself.

20.2.2 Building the ITK Libraries

ITK requires CMake to be installed and available on your system. As of ITK version 3.2.0, CMake version 2.4.6 or greater must be used to prevent compilation errors. In order to simplify updating ITK to later versions of the software, we perform an out-of-source build. To do this, we first create a directory outside the **Insight** directory created for us during the CVS checkout process.

```
mkdir itk
cd itk
```

We now run CMake, or the GUI-based version `ccmake`, in this directory. If you're using Windows, you may wish to run the Windows-specific `cmake-gui.exe` instead.

```
ccmake ../Insight
```

Note: The above command assumes that the **Insight** directory exists at the same level as the **itk** directory that we just created.

The following advanced CMake variables must be set to the appropriate values:

CMake Variable	Value
BUILD_SHARED_LIBS	ON
INSTALL_WRAP_ITK_COMPATIBILITY	ON
ITK_CSWIG_PYTHON	OFF
ITK_USE_REVIEW	ON
USE_WRAP_ITK	ON
WRAP_ITK_PYTHON	ON
WRAP_ITK_JAVA	OFF

Note: Some CMake variables are only available based on the state of others. If a variable is missing from the list, set what is visible and re-configure, this will often allow you to see and set additional parameters.

After generating the appropriate files and exiting `ccmake`, the standard build process can be completed. To compile (on Linux or Mac OS X), run:

```
make
sudo make install
```

On Windows, the build process is governed by the type of project or Makefile generated by CMake.

Note:

It is possible to use ITK without installing it. To do this, the environment variables `LD_LIBRARY_PATH` and `PYTHONPATH` must be set to the appropriate build directories:

```
LD_LIBRARY_PATH=/Path_To_itk/bin
PYTHONPATH=/Path_To_itk/Wrapping/WrapITK/Python
```

At this point, ITK is build and installed. To validate this, open a Python shell and issue the following commands:

```
>>> import itk
>>> itk.Image[itk.US, 2]
```

The above commands should both complete without error, and should produce the output:

```
<class 'itkImage.itkImageUS2'>
```

The WrapITK implementation used to wrap ITK for use in Python lazily instantiates required classes. This means that even if the import succeeds, the instantiation of the above `itk.Image` class may fail. This is particularly common if the environment variable `LD_LIBRARY_PATH` is incorrectly set.

20.3 ITK and VisTrails

When built and installed with the appropriate Python bindings included, ITK can be used from VisTrails through the ITK package. As mentioned previously, ITK is a third-party package and is not included in the general VisTrails distribution. However, like many third-party packages, it is available from the VisTrails website.⁴

The VisTrails ITK package is under continual development with the latest versions being announced on the VisTrails website. After downloading the package and extracting it into the `.vistrails/userpackages` directory, you can enable it through the `Module Packages` tab in the `Preferences` dialog. Please refer to Chapter [Writing VisTrails Packages](#) for instructions on how to do this.

Upon starting VisTrails, the ITK package modules will be made available to the Builder Window.

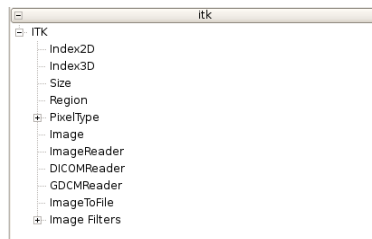


Figure 20.1: (a) The VisTrails ITK Package Structure Overview



Figure 20.2: (b) The ITK Package Supported PixelTypes

20.3.1 ITK Package Organization

The ITK VisTrails package loosely mimics the ITK functionality hierarchy. The package's top level consists of base classes, containers, and file readers as shown in Figure (a) *The VisTrails ITK Package Structure Overview*. Also

⁴ Please Note: At the time of this writing, the VisTrails ITK package is not a complete wrapping of all ITK functionality. If you would like to contact the author regarding the wrapped functionality, please do so through the e-mail address on the VisTrails website.

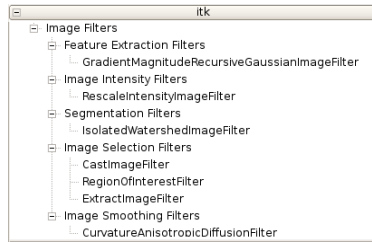


Figure 20.3: (c) The ITK Package Filter Structure

available at the top level is the `PixelType` module and the specific types used to create and execute ITK-based pipelines.

Currently, the ITK Image Filters are organized into functional groups. The five filter types, as show in Figure (c) *The ITK Package Filter Structure*, are:

- Feature Extraction Filters
- Image Intensity Filters
- Segmentation Filters
- Image Selection Filters
- Image Smoothing Filters

All filter types currently have at least one representative ITK filter wrapped and usable from within the VisTrails environment.

20.3.2 Reading DICOM Volumes

DICOM is a standard format for exchanging medical images. ITK includes DICOM support through the GDCM libraries.⁵ It is worthwhile to note that at this time these libraries are currently not a complete implementation of the DICOM standard.

DICOM volumes can be integrated into VisTrails through the use of either the `GDCMReader` or `DICOMReader` modules in the ITK package. For the rest of this example, we will use the `GDCMReader` module as its performance is slightly higher than the `DICOMReader` implementation.

Figure *VisTrails workflow utilizing ITK to extract a single slice from a DICOM volume. The slice is chosen by first forming a Region to extract. The result is viewed through the use of standard VisTrails Spreadsheet modules.* shows the use of the `GDCMReader` module. In order to properly read a DICOM volume, the `GDCMReader` must be supplied with the dimension of the volume to be read and the directory containing the series to read. By default, `WrapITK` supports two- and three-dimensional volumes. In order to include support for higher dimensions, the appropriate `WrapITK` variable must be set within CMake, *before* compiling ITK.

20.3.3 Volume Processing With ITK and VisTrails

Typically, DICOM volumes are written with no 16-bit unsigned shorts. Unfortunately, most systems allow the display of only 8-bit values. Because of the higher precision inherent in DICOM data, it is often preferable to perform any computation, segmentation, or processing on the data prior to rescaling in order to utilize as much information as possible.

⁵ Grass roots DiCoM Project. <http://www.creatis.insa-lyon.fr/Public/Gdcm/>

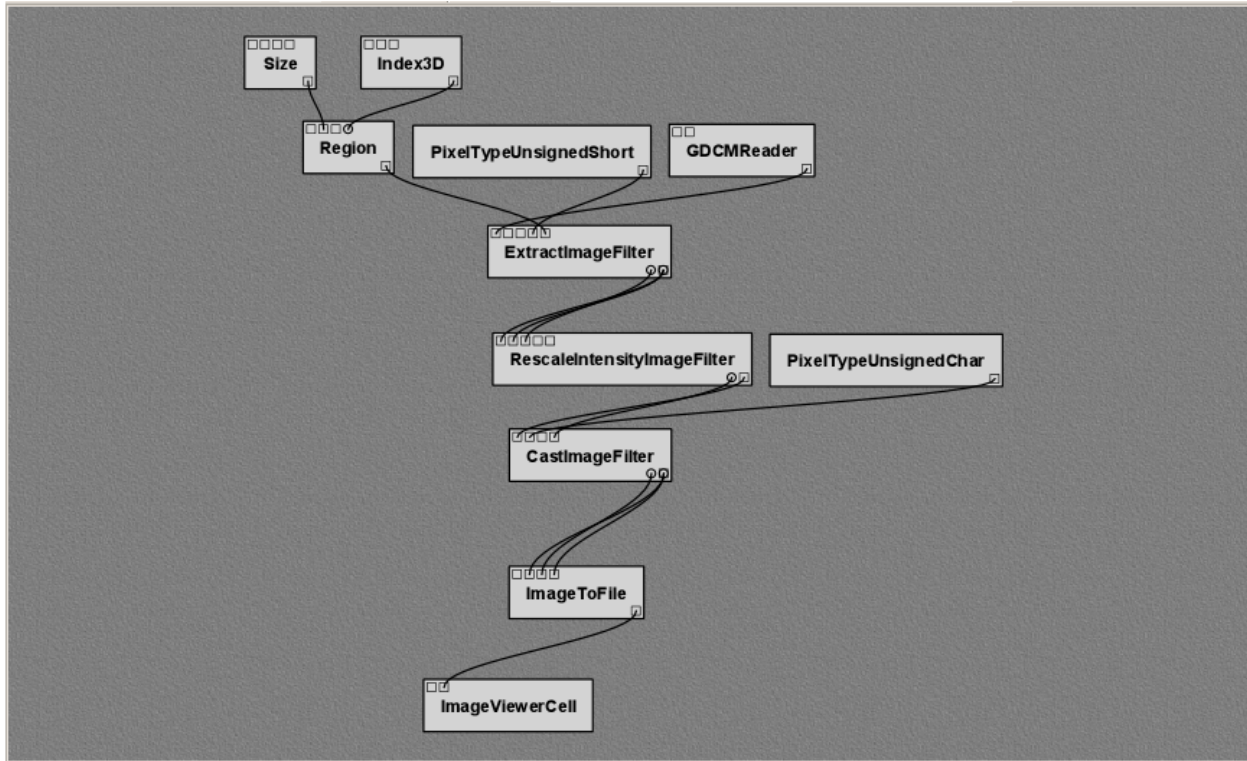


Figure 20.4: VisTrails workflow utilizing ITK to extract a single slice from a DICOM volume. The slice is chosen by first forming a Region to extract. The result is viewed through the use of standard VisTrails Spreadsheet modules.

20.3.4 Volume Processing With ITK and VisTrails

ITK image filters are typically templated based on the dimensionality of the data being processed, as well as the data type being processed. In VisTrails, these parameters are handled through the use of `PixelType` modules. Although any ITK Filter wrapped in `vistrails` can accept any of these `PixelTypes`, the underlying implementation may not be compatible with the input `PixelType`. Using `PixelTypes` incompatible with the underlying filter implementations is the most frequent cause of error when executing otherwise functional pipelines in VisTrails.

When processing volumes, it is often necessary to extract a single slice from the volume at different stages of the processing pipeline. This is possible in VisTrails through the use of the `ExtractImageFilter`. Given a volume, a `Region`, and Dimensionality information, the `ExtractImageFilter` can extract a single slice from the data volume that can be used in further processing, previewing the results, or writing to disk. An example workflow that extracts a slice from a DICOM volume can be seen in Figure *VisTrails workflow utilizing ITK to extract a single slice from a DICOM volume. The slice is chosen by first forming a Region to extract. The result is viewed through the use of standard VisTrails Spreadsheet modules.*

20.3.5 Visualizing the results

Although ITK's processing filters and the DICOM standard both support 16-bit processing and storage, many image viewers are capable of displaying in only 8-bit resolution using the unsigned char `PixelType`. Since the output of an ITK processing workflow is an image, it makes sense to view it as such. This means that we are required to both remap the data values in the image to 8 bits as well as perform a casting operation to change the data type from unsigned shorts to unsigned chars. These operations are performed through the use of the `RescaleIntensityImageFilter` and the `CastImageFilter`. Figure *VisTrails workflow utilizing ITK to extract a single slice from a DICOM volume. The slice is chosen by first forming a Region to extract. The result is viewed through the use of standard VisTrails Spread-*

sheet modules. demonstrates the use of the `RescaleIntensityImageFilter` and the `CastImageFilter` in conjunction with the `ImageToFile` and `ImageViewerCell` modules to view the resulting slice in the VisTrails Spreadsheet.

Citations

CONTROL FLOW IN VISTRAILS

Scientific workflows usually follow a dataflow model, but, in some cases, control structures, including loops and conditionals, are necessary to accomplish certain tasks. VisTrails provides the `Control Flow` package to support these and other structures.

21.1 The Map operator

In functional programming, `map` is a high-order function that applies a given function to a list (each element of the list is processed using this function) and returns a sequence of results. The `Map` module provides this functionality for workflows in VisTrails. Note that this module provides simple looping as it can be used to iterate through a list of inputs.

In order to use the `Map` module, a `ListOfElements` type representing a list of data structure is also provided. Two additional modules help users combine elements of lists: `Dot` combines the elements of two lists like the dot product (the first element of the first list is combined with the first of the second one, the second element of the first list is combined with the second of the second one, and so on), and `Cross` combines the elements as the cross product (all the possible combinations between the elements of the lists); both modules return a list of tuples. In addition to two lists of elements as input ports, they take an optional boolean input named “CombineTuple”. This input is useful when one or both input lists have tuples as elements; if this port is selected, and its value is `False`, the elements of the list will not be combined in just one tuple, e.g., $(1, 2) + 3 \rightarrow ((1, 2), 3)$; otherwise, the elements will be combined, e.g., $(1, 2) + 3 \rightarrow (1, 2, 3)$.

The `Map` module has four input ports:

- “FunctionPort”: this port receives the module (via the “self” output port) that represents the function to be applied for each element of the input list; if the function uses more than one module, you must use a `Group` (see Chapter *Creating and Modifying Workflows*) or a `SubWorkflow` and connect that composite module to this port;
- “InputPort”: this port receives a list of the names of the input ports that represent the individual arguments of the function;
- “OutputPort”: this port receives the name of the output port that represents the individual result of the function;
- “InputList”: this port receives the input list for the loop; it must be a list of tuples if more than one function input port was chosen.

The output port “Result” produces a list of results, one for each element in the input list.

To better show how to use the `Map` module, let’s use a workflow as an example. Inside the “examples” directory of the VisTrails distribution, open the “triangle_area.vt” vistrail. Now, select the “Surface Area” version. This version basically calculates the area of a given isosurface. We are going to modify this version, in order to calculate the areas of the isosurface given by contour values in a list. Then, we will create a 2D plot to show all the areas.

Begin by deleting the `StandardOutput` module, and the connection between the `vtkDataSetReader` and the `vtkContourFilter` modules. Then, drag the following modules to the canvas:

- `Map`
- `ListOfElements`
- `Cross`
- `MplPlot` (under “matplotlib”)
- `MplFigure` (under “matplotlib”)
- `MplFigureCell` (under “matplotlib”)
- `InputPort` (under “Basic Modules”) - you will need two of them
- `OutputPort` (under “Basic Modules”)
- `PythonSource` (under “Basic Modules”)

Notice that when you drag `Map` to the pipeline canvas it will be drawn in a different shape from the other modules. This is a visual cue to help distinguish control modules from other modules. All control modules have the same shape.

Select the `vtkContourFilter` module and delete its method “`SetValue`” in the `Set Methods` container. Then, open its configuration dialog (you can use the “`Ctrl-E`” or “`Command-E`” keyboard shortcut) and enable this method (the input port “`SetValue`”) by clicking on it, and pressing `OK`.

Then, connect the modules as shown in Figure *Connecting a subset of the modules to be grouped as a SubWorkflow*.

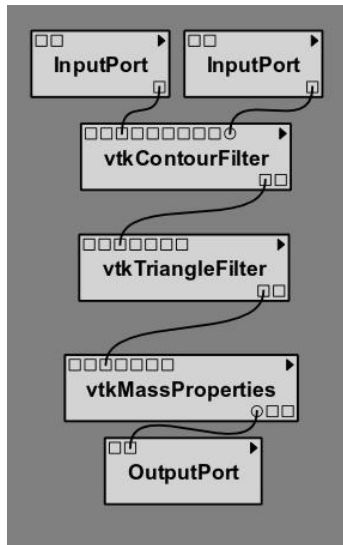


Figure 21.1: Connecting a subset of the modules to be grouped as a `SubWorkflow`

These modules represent the function we wish to map: each element of the input list will be processed using them. Because we have more than one module, we need to create a `Group` or a `SubWorkflow` to identify the entire function. The `InputPort` and the `OutputPort` modules are necessary to expose these ports in the `Group/SubWorkflow` structure.

In this example, we will use a `SubWorkflow` structure. Select all the modules shown in Figure *Connecting a subset of the modules to be grouped as a SubWorkflow*, go to the `Edit` menu, and then click on `Make SubWorkflow`. You can name it `CalculateArea`. Select this `SubWorkflow` and open its configuration. When the configuration dialog opens, enable the output port “`self`” and press `OK`. You will need this port to connect to the `Map` module.

Note

When using `Map`, the module (or subworkflow) used as function port in the map module **MUST** be a function, i.e., it can only define 1 output port.

Now, select the `MplPlot` module and open its configuration dialog. Inside it, add two input ports of type `ListOfElements`: “InputList” and “X_Values”. Also, copy the code listed below, in order to create the necessary information for the 2D plot, into the source text area and save your changes using the OK button.

```
subplot (212)

dashes = [1, 3]
list1 = self.getInputFromPort("InputList")
list2 = self.getInputFromPort("X_values")
list3 = []

for i in xrange(len(list1)):
    list3.append(list2[i][1])

l, = plot(list3, list1, marker="o", markerfacecolor="red",
          markersize=7, label="IsoSurface Areas", linewidth=1.5)

l.set_dashes(dashes)
```

Next, edit the `PythonSource` module by adding an output port “List” of type `ListOfElements`, copying the following code to the source text area, and saving these changes. The code will create a range of contour values that we will use as our input list.

```
result = []

for i in xrange(4, 256, 4):
    result.append(i)

self.setResult("List", result)
```

It may be helpful to identify this `PythonSource` module by labeling it as `RangeList`. Connect all the modules as shown in Figure *All the modules connected in the canvas*.

You will set some parameters now:

- `HTTPFile`: set the parameter “url” to <http://www.sci.utah.edu/~cscheid/stuff/head.120.vtk>
- `ListOfElements`: set the parameter “value” to `[0]`
- `Map`: set the parameter “InputPort” to `["SetValue"]` and the parameter “OutputPort” to `GetSurfaceArea`

The workflow is now ready to be executed. When you execute the workflow, you will see the `SubWorkflow CalculateArea` executing several times (one time for each value of the input list). It’s important to notice that, although only the module connected to `Map` (in this example, the `SubWorkflow Calculate Area`) will be in the loop, the modules above it will be also used in each loop iteration; the difference is that they are going to be executed only for the first iteration; in all other iterations, the results will be taken from the cache.

When the workflow finishes its execution, the VisTrails Spreadsheet will contain a 2D plot (Figure *The result in the VisTrails Spreadsheet*).

This example can be found in the version “Surface Area with Map”, inside the “triangle_area.vt” vistrail.

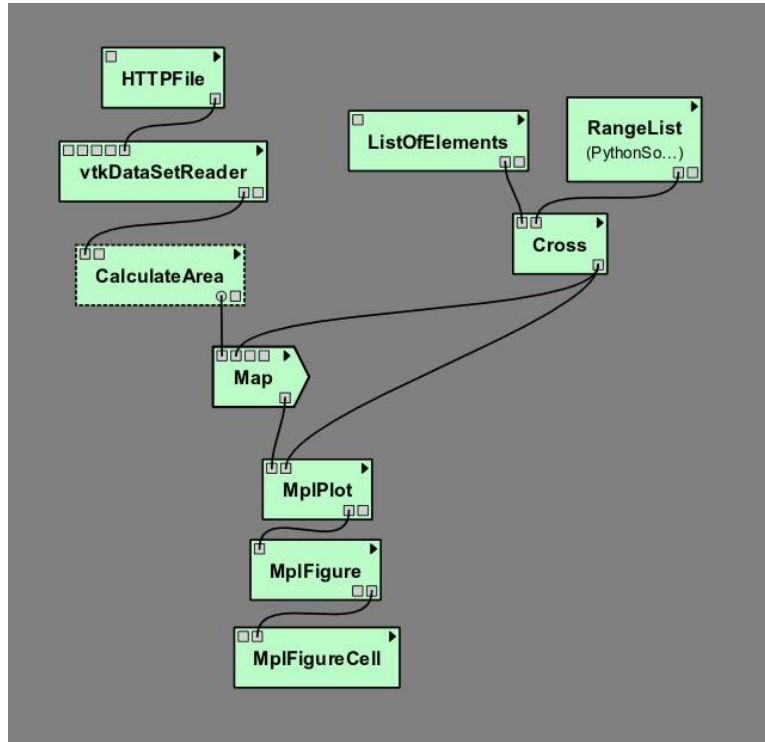


Figure 21.2: All the modules connected in the canvas

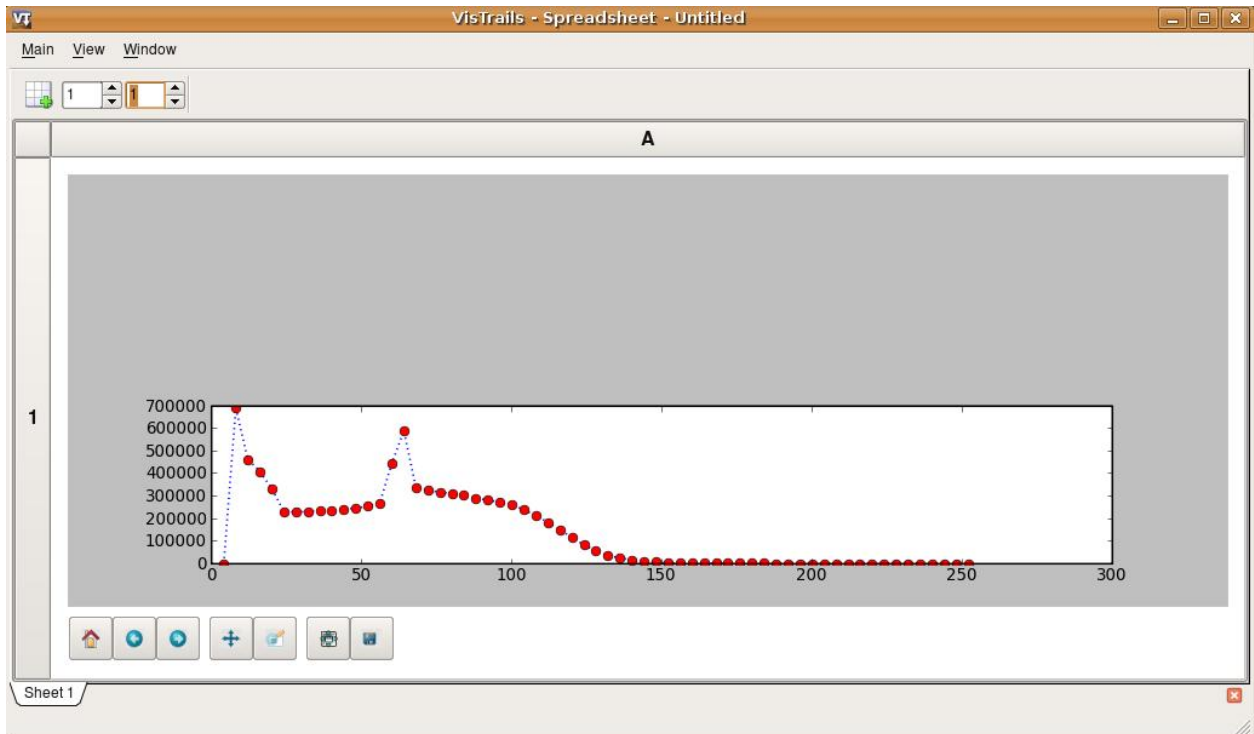


Figure 21.3: The result in the VisTrails Spreadsheet

21.2 Filtering results

When computing large lists of results, it can be useful to selectively reduce the list during execution to avoid unnecessary computation.

The `Filter` module was developed to address this issue. It receives an input list and, based on a specified boolean condition, returns only elements of the list that satisfy the condition. Its ports are the same as those in the `Map` module. The difference between these modules is related to the structure: in `Filter`, the output port “FunctionPort” cannot receive any function, but only a condition; in other words, the selected port in “OutputPort” must return a boolean value or `Filter` will not work. Thus, `Filter` will not return a list with boolean values as `Map` would do, but rather the elements of the input list for which the condition evaluated to `True`.

To better understand how `Filter` works, let’s modify our earlier example to filter out areas less than 200,000. With the previous `vistrial` open (you can use the “Surface Area with `Map`” version), add the following modules to the canvas:

- `Filter`
- `PythonSource` (under “Basic Modules”)

Edit the configuration of `PythonSource` by adding an input port of type `Float` named “Area”, and an output port of type `Boolean` named “Condition”, and writing the following code inside the source text area:

```
area = self.getInputFromPort("Area")

if area>200000.00:
    self.setResult("Condition", True)
else:
    self.setResult("Condition", False)
```

Press the OK button. You can label this `PythonSource` as `FilterCondition`. Now, reorganize the modules in the canvas as shown in Figure *The new organization of the modules in the canvas*.

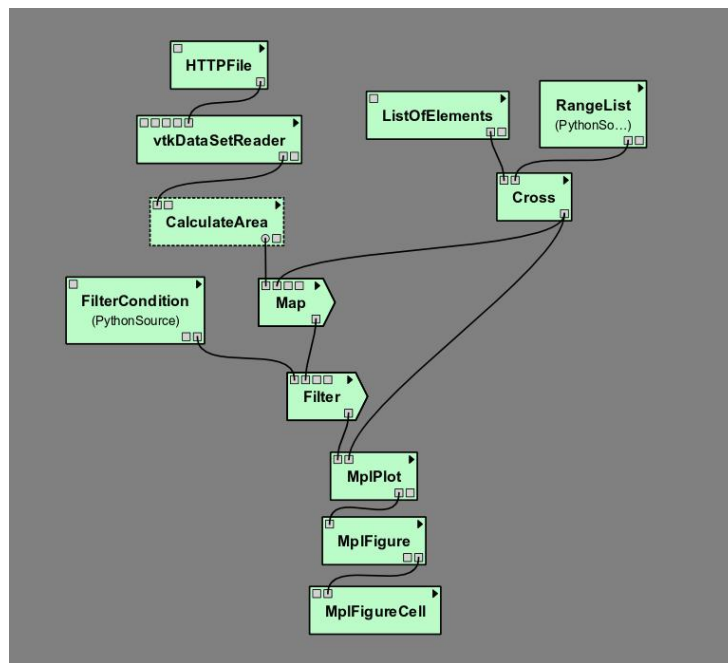


Figure 21.4: The new organization of the modules in the canvas

Select the `Filter` module and set the values of its parameters to the following:

- “InputPort”: [*Area*]
- “OutputPort”: *Condition*

When you execute this workflow, it will generate another plot that is similar to the one from the `Map` example, but only areas above 200,000 are considered (Figure *The result in the VisTrails spreadsheet*).

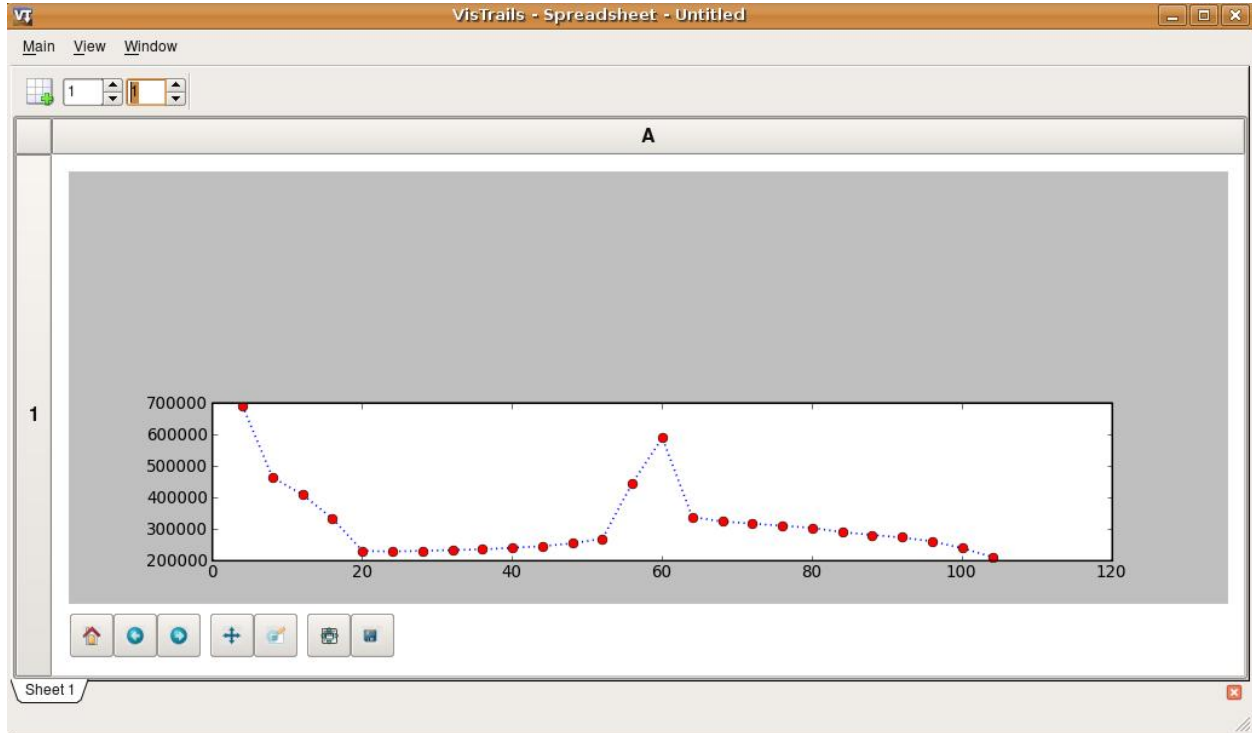


Figure 21.5: The result in the VisTrails spreadsheet

This example is already inside the “`triangle_area.vt`” vistrail, in the “Surface Area with Map and Filter” version.

Later in this chapter, you will see how to combine `Map` and `Filter` in one single module, by creating your own control structure.

21.3 Conditional module

Conditional statements are a very important control flow structure frequently used in programming languages, and the `if` structure is probably the most common of these structures. In scientific workflows, for example, an `if` structure can be used to select the part of the pipeline to be executed based on a boolean condition.

For this reason, the `Control Flow` package also includes an `If` module. Its input ports are:

- “Condition”: this port receives a boolean value that will specify the direction of computation;
- “TruePort”: this port receives the part of the workflow that will be executed if the condition value is `True`; you don’t need to group or make a `SubWorkflow` in this case: just connect the output port “self” of the last module in this port;
- “FalsePort”: this port receives the part of the workflow that will be executed if the condition value is `False`; as with the
- “TruePort” port, you don’t need to group or make a `SubWorkflow`;

- “TrueOutputPorts”: this port receives a list that contains the names of the output ports of the module connected to “TruePort” that you want the result of; this port is optional;
- “FalseOutputPorts”: this port receives a list that contains the names of the output ports of the module connected to “FalsePort” that you want the result of; this port is optional.

The `If` module has an output port named “Result” that returns a list with the results of the specified output ports of “TrueOutputPorts” or “FalseOutputPorts”, depending on the condition value. If only one output port is chosen, the result of this port will not be returned in a list. If “TrueOutputPorts” or “FalseOutputPorts” are not enabled, “Result” returns `None`.

Let’s do now a simple example to show how exactly this module works. This example is from the bioinformatics domain, and takes a string as the input; if it’s a structure identifier, a web service from the European Bioinformatics Institute, or simply EBI (<http://www.ebi.ac.uk/>), a centre of research in bioinformatics, is used to get the structure in the PDB format, a standard representation for macromolecular structure, and then, the `VTK` package is used to show the protein in the VisTrails Spreadsheet; otherwise, the input is assumed to be invalid, and a message is generated in the Spreadsheet too.

First, the EBI’s web service must be enabled. For this, you need to add the following url to the `wSDLList` configuration:

```
http://www.ebi.ac.uk/Tools/webservices/wsdL/WSDbfetch.wsdl
```

Don’t forget to ensure that the `webServices` package is enabled in the Preferences dialog. For more information about web services in VisTrails, see Chapter *Example: Web Services*.

Now, you’re going to drag the following modules to the canvas:

- `If`
- `fetchData` (under “Methods” for the current web service)
- `vtkPDBReader` (under “VTK”)
- `vtkDataSetMapper` (under “VTK”)
- `vtkActor` (under “VTK”)
- `vtkRenderer` (under “VTK”)
- `VTKCell` (under “VTK”)
- `PythonSource` (under “Basic Modules”) - you will need three of them
- `String` (under “Basic Modules”)
- `RichTextCell` (under “VisTrails Spreadsheet”)

Select one of the `PythonSource` modules, and open its configuration dialog. Inside it, add one input port of type `String`, named “PDB_format”, and one output port of type `File`, named “File”. Then, write the following code:

```
1 PDB_format = self.getInputFromPort('PDB_format')
2
3 output = self.interpreter.filePool.create_file()
4 file_ = open(str(output.name), 'w')
5 file_.write(PDB_format)
6
7 self.setResult('File', file_)
8
9 file_.close()
```

You can name this module as `CreateFile`. Now, set some parameters of `fetchData`:

- “format”: `pdb`

- “style”: *raw*

Next, connect some modules as shown in Figure *Some modules of the workflow connected*.

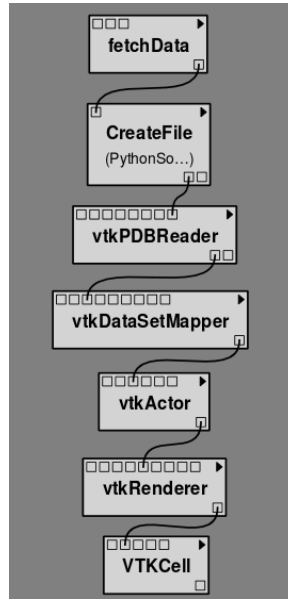


Figure 21.6: Some modules of the workflow connected

The aim of this group of modules is to get the PDB format of the structure ID, through the web service, and then make the visualization with the VTK package. This is the part of the workflow that will be executed if the input is a structure identifier.

Next, select another PythonSource module and open its configuration dialog too. One input port named “Structure”, of type `String`, and one output port named “Is_ID”, of type `Boolean`, must be added, as well as the code below:

```

1  if "\n" in structure:
2      lineLen = structure.index("\n")
3  else:
4      lineLen = -1
5  if lineLen < 1:
6      lineLen = len(structure)
7
8  if ":" in structure:
9      index = structure.index(":")
10 else:
11     index = -1
12
13 if (structure[0] != "ID ") and (index > 0) and (index < lineLen):
14     is_ID = True
15 else:
16     is_ID = False
17
18 self.setResult("Is_ID", is_ID)

```

Name this module as `Is_ID`. This module will be the condition for the `If` structure.

Now, select the last PythonSource module, and, inside its configuration, add one input port of type `String`, named “Input”, and one output port of type `File`, named “html”. Then, copy the code below:


```

1 input = self.getInputFromPort("Input")
2
3 output = self.interpreter.filePool.create_file()
4 f = open(str(output.name), 'w')
5 text = '<HTML><TITLE>Protein Visualization</TITLE><BODY BGCOLOR="#FFFFFF">'
6 f.write(text)
7 text = '<H2>Protein Visualization Workflow</H2>'
8 f.write(text)
9 text = '<H3>The following input is not an ID from a protein:</H3>'
10 text += '<H4>' + str(input) + '</H4>'
11 text += '<H3>The visualization cannot be done.</H3>'
12 f.write(text)
13
14 text = '</BODY></HTML>'
15 f.write(text)
16
17 self.setResult('html', f)
18
19 f.close()

```

Name this PythonSource as Not_ID. This module will print a message in the VisTrails Spreadsheet when the input is not a structure identifier.

Finally, the String module can be named as Workflow_Input, to make it clear that it takes the input of the workflow. Also, open the configuration dialog of RichTextCell to enable the output port “self”, so it can be connected to the If module. Then, connect all the modules as shown in Figure *All the modules connected*.

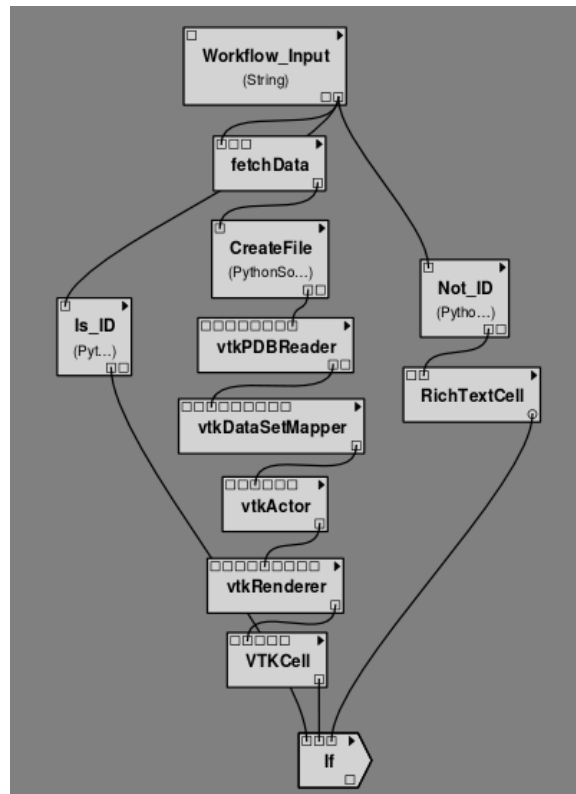


Figure 21.7: All the modules connected

In order to better organize the disposal of the modules, group all the modules shown in Figure *Some modules*

of the workflow connected by selecting them, going to the Edit menu and clicking on Group. Name it as `Generate_Visualization`. Your workflow must correspond to the one shown in Figure *The final workflow, using the Group structure*.

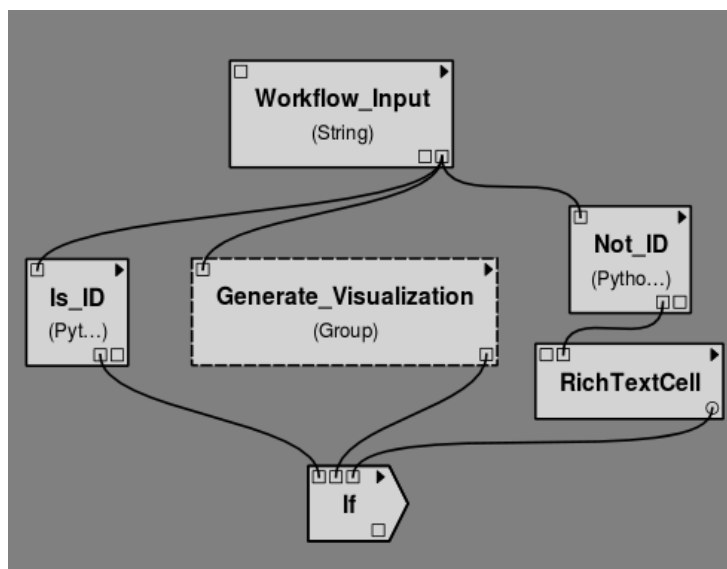


Figure 21.8: The final workflow, using the Group structure

Note that this implementation follows exactly the initial specification of the workflow. If the input is a structure identifier (`Is_ID` returns `True`), `Generate_Visualization` will be executed; otherwise (`Is_ID` returns `False`), `Not_ID` and `RichTextCell` will create an error message in the VisTrails Spreadsheet.

For the workflow execution, set the parameter “value” of the `Workflow_Input` module to `PDB:3BG0`. This entry is an ID from a protein; so, the condition will be `True`, and the `Generate_Visualization` group will be executed, generating the visualization show in Figure *The visualization of the protein in the VisTrails Spreadsheet*.

If you change the value from the input port “value” to `protein`, for example, the condition will be `False`, and the message shown in Figure *The message in the Spreadsheet, generated when the input is not a structure ID* will be generated in the Spreadsheet.

This example can be found inside the “examples” directory, in the “protein_visualization.vt” vistrail. It was partially based on the workflow “Structure_or_ID”, which can be found at <http://www.myexperiment.org/workflows/225>.

21.4 Building your own loop structure

In functional programming, `fold` is a high-order function used to encapsulate a pattern of recursion for processing lists. A simple example of a `fold` is summing the elements inside a list. If you `fold` the list `[1, 2, 3, 4]` with the sum operator, the result will be $((1+2)+3)+4 = 10$. It’s common to start with an initial value too. In the sum example, the initial value would be 0, and the result would be $((0+1)+2)+3)+4 = 10$.

With this function, a programmer can do any type of recursion. In fact, the `map` and `filter` functions, shown previously, can be implemented with `fold`. The `Control Flow` package provides a `Fold` module to enable this functionality, and the `Map` and the `Filter` modules inherit from the `Fold` class.

In fact, any control module that has this kind of recursion uses the `Fold` class. To use this functionality for your own control modules, instead of defining the `compute()` method, you need to define two other methods:

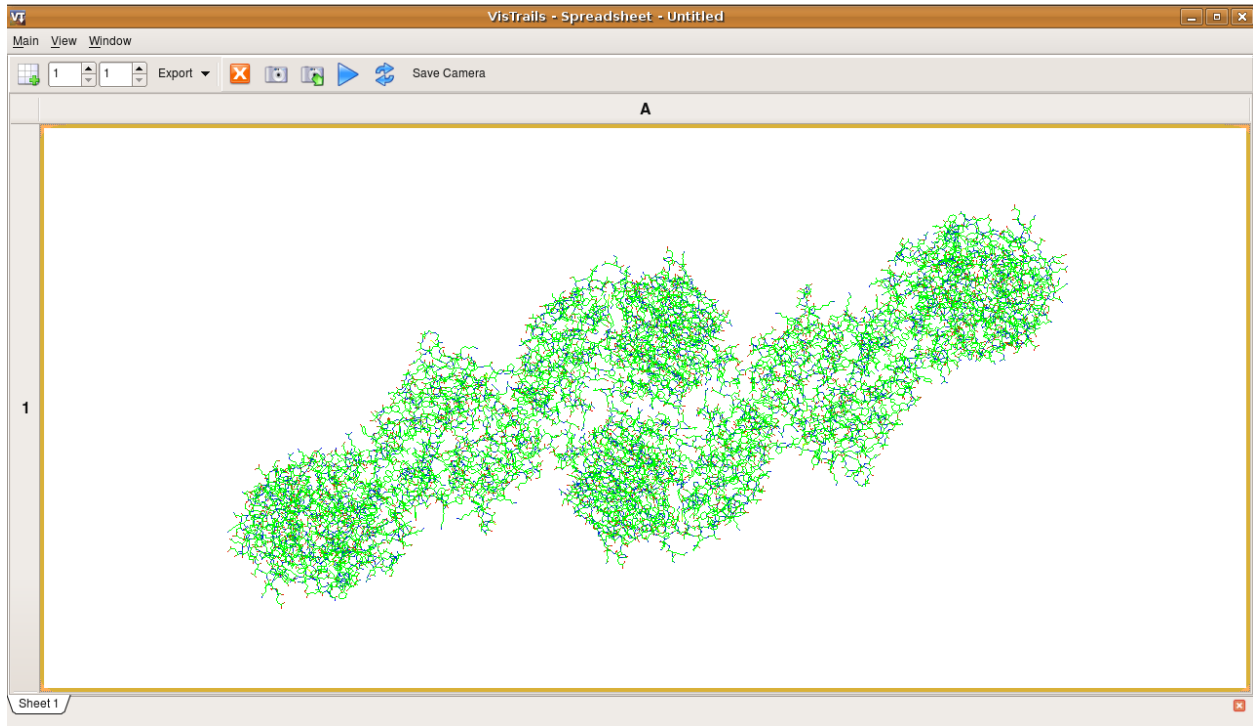


Figure 21.9: The visualization of the protein in the VisTrails Spreadsheet

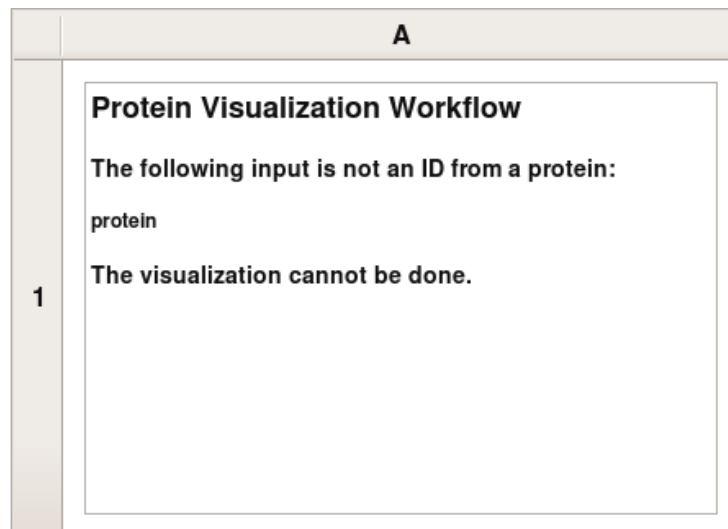


Figure 21.10: The message in the Spreadsheet, generated when the input is not a structure ID

- `setInitialValue()`: in this method, you will set the initial value of the fold operator through the `self.initialValue` attribute;
- `operation()`: in this method, you must implement the function to be applied recursively to the elements of the input list (e.g., the sum function). More specifically, you need to define the relationship between the previous iteration's result (`self.partialResult` attribute) and the current element of the list (`self.element` attribute); this method must be defined after the `setInitialValue()` one.

It's important to notice that all modules inheriting from `Fold` will have the same ports, as `Map` and `Filter`, but you can add any other ports that will be necessary for your control structure. Also, you do not need to use the input ports "FunctionPort", "InputPort" and "OutputPort". You will only use them when you create an operator like `Map` and `Filter`, which need a function to be applied for each element of the input list.

As an example, we will create a simple `Sum` module to better understand the idea. Create a new package, and the code inside it would be as follows:

```
1 from controlflow import Fold, registerControl
2
3 version = "0.1"
4 name = "My Control Modules"
5 identifier = "edu.utah.sci.my_control_modules"
6
7 def package_dependencies():
8     return ["edu.utah.sci.vistrails.control_flow"]
9
10 class Sum(Fold):
11     def setInitialValue(self):
12         self.initialValue = 0
13
14     def operation(self):
15         self.partialResult += self.element
16
17 def initialize(*args, **keywords):
18     registerControl(Sum)
```

We begin by importing the `Fold` class and the `registerControl` function from the `Control Flow` package (Line 1). The `registerControl` function is used to register the control modules, so the shape of them can be set automatically.

Also, define the variables `version`, `name` and `identifier`, as it's done for all packages. The interpackage dependency (include reference of the package chapter) is used too, as `My Control Modules` requires a module and a function from `Control Flow` (Lines 7 and 8); in this way, `VisTrails` can initialize the packages in the correct order. Then, create the class `Sum`, which inherits from `Fold`. Inside it, set the initial value to 0 inside the `setInitialValue()` method (Lines 11 and 12), and define the sum operator inside `operation()`, as shown clearly by the relation between `self.partialResult` and `self.element` (Lines 14 and 15).

The last thing we must do is define the `initialize()` method, so the package can be loaded in `VisTrails`. However, instead of calling the registry, if you do not need any other ports, you just have to call the `registerControl()` function (Line 18).

Save this package and enable it inside `VisTrails`. Create a similar workflow as shown in Figure [A workflow using the Sum module](#).

Upon executing this workflow, the sum $((((0+1)+2)+3)+4)$, should be printed on your terminal as following:

```
10
```

Note that the input ports "FunctionPort", "InputPort" and "OutputPort" were not necessary for this module. Now, let's see another example that does use them. Open the workflow we used to calculate the area of isosurfaces

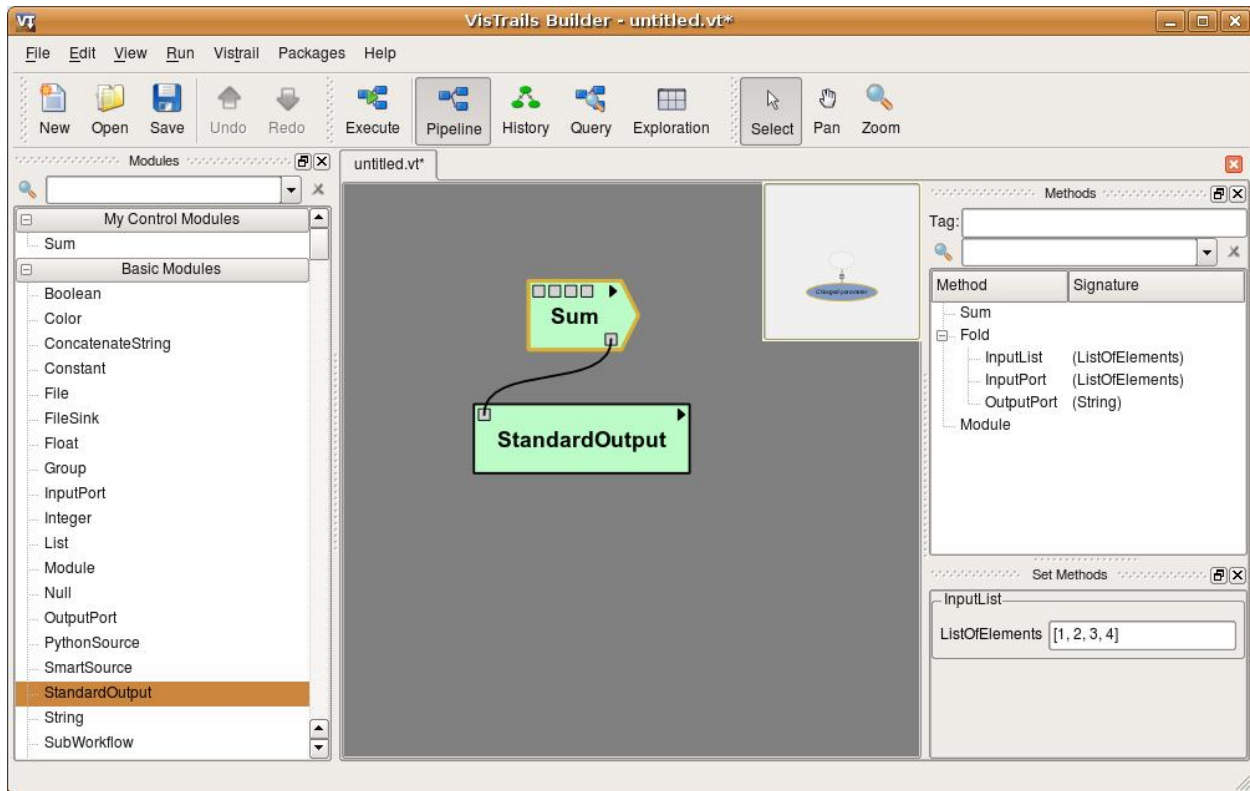


Figure 21.11: A workflow using the Sum module

(in "triangle_area.vt", "Surface Area with Map and Filter" version), and delete the Map, the Filter, and the FilterCondition (PythonSource) modules.

Now, create a single module that maps the list and filters the results, named as AreaFilter. Inside your package, add the following class:

```

1 class AreaFilter(Fold):
2     def setInitialValue(self):
3         self.initialValue = []
4
5     def operation(self):
6         area = self.elementResult
7
8         if area > 200000:
9             self.partialResult.append(area)

```

The initial value is an empty list, so the result of each element can be appended to it (Line 3). In the operation() method, the self.elementResult attribute is used (Line 6); it represents the result of the port chosen in "OutputPort"; so, it means that "FunctionPort", "InputPort" and "OutputPort" will have connections. In this workflow, self.elementResult is the area for each contour value inside the input list, and, if the area is above 200,000, it will be appended to the final result (Lines 8 and 9). We can easily see that this module does exactly the same as Map and Filter combined.

Don't forget to register this module in the initialize() function. After doing this, save the package and load it again inside VisTrails. Then, just connect AreaFilter as in Figure *The same workflow, but now with AreaFilter*.

Now, you must set some values in the following parameters of AreaFilter:

- "InputPort": ["SetValue"]

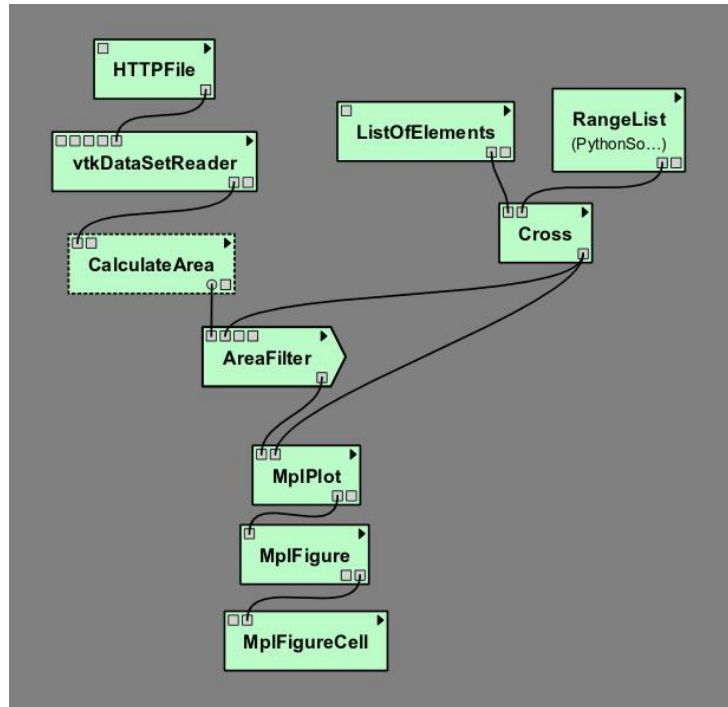


Figure 21.12: The same workflow, but now with `AreaFilter`

- “OutputPort”: `GetSurfaceArea`

When you execute this workflow, the result in the VisTrails Spreadsheet will be the same as shown previously (Figure *The result in the VisTrails spreadsheet*). It shows the flexibility of doing a recursion function by inheriting from `Fold`.

Part VI

Indices and tables

- *genindex*
- *search*

BIBLIOGRAPHY

- [C1] 18. (a) Banvard, "The visible human project image data set from inception to completion and beyond," Proceedings of CODATA, 2002.
- [C2] 12. Ibanez, W. Schroeder, L. Ng, and J. Cates, The ITK Software Guide, 2nd ed., Kitware, Inc. ISBN 1-930934-15-7, <http://www.itk.org/ItkSoftwareGuide.pdf>, 2005.
- [C3] 20. (a) Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxes, and R. Whitaker, "Engineering and algorithm design for an image processing API: A technical report on ITK - The Insight Toolkit," Proceedings of Medicine Meets Virtual Reality, pp. 586-592, 2002.

INDEX

Symbols

- : parameters
 - differences, 34
- “-output_ports“
 - Module registry, 122
- “ConfigurationObject“, 118
- “ModuleError“
 - modules, 112, 115
- “_input_ports“
 - Module registry, 122
- “_modules“
 - Module registry, 122
- “addInputPort“
 - Module registry, 112
- “addModule“
 - Module registry, 112
- “addOutputPort“
 - Module registry, 112
- “add_Module“
 - Module registry, 115
- “add_input_port“
 - Module registry, 115
- “add_output_port“
 - Module registry, 115
- “filePool“
 - packages, 118
- “initialize“
 - packages, 112, 115

A

- adding
 - connections, 21
 - modules, 19, 91
 - ports, 23
 - spreadsheet sheets, 40
 - tags, 33
- adding parameters
 - parameter exploration, 59
- analogy, 51
- animation, 59, 62
- annotations

- versions, 33

B

- basic
 - modules, 24
- batch mode, 133
- builder, 19, 26
- by example
 - queries, 45

C

- caching
 - modules, 122
- cells
 - creating, spreadsheet, 40
 - positioning, spreadsheet, 40
 - spreadsheet, 40, 41
 - web browser, spreadsheet, 43
- center, 15
- changing
 - parameters, 21
- close
 - vistrail, 11
- color
 - modules, 123
- columns
 - spreadsheet, 40
- command line arguments, 131
- comparing
 - versions, 34
- configuration
 - packages, 117
- configuration directory, 132
- connecting
 - modules, 21
- connections
 - adding, 21
 - definition, 14
 - selecting, 19
- control flow
 - parameter exploration, 82
- control flow assistant, 81

creating
 spreadsheet cells, 40

D

database, 85, 87
 issues, 87
 opening from, 86
 saving to, 87
 setup, 85

default values
 ports, 126
 spreadsheet, 40

definition
 connections, 14
 modules, 14
 vistrail, 14

deleting
 modules, 19
 ports, 23
 spreadsheet sheets, 40
 tags, 33

dependencies
 packages, 120

diff
 see versions, comparing, 34

differences
 : parameters, 34

directions
 parameter exploration, 59

docking
 spreadsheet sheets, 40

dynamic
 modules, 128

E

editing
 spreadsheet modes, 42

execute, 14

exploring
 parameters, 59, 82

F

from a database
 open, 11

G

grouping
 modules, 24

H

history, 29

I

image

 spreadsheet saving, 44

input
 persistent files, 95

input dependency
 ports, 127

interactive
 spreadsheet modes, 41

intermediate
 persistent files, 96

issues
 database, 87

L

labels
 modules, 23
 ports, 126

layout
 spreadsheet, 39

legend, 34

log, 139

M

merging, 35

methods, 21

modes
 editing, spreadsheet, 42
 interactive, spreadsheet, 41
 spreadsheet, 41

module connectivity
 ports, 127

Module registry
 “-output_ports“, 122
 “_input_ports“, 122
 “_modules“, 122
 “addInputPort“, 112
 “addModule“, 112
 “addOutputPort“, 112
 “add_Module“, 115
 “add_input_port“, 115
 “add_output_port“, 115

modules, 19
 “ModuleError“, 112, 115
 adding, 19, 91
 basic, 24
 caching, 122
 color, 123
 connecting, 21
 definition, 14
 deleting, 19
 dynamic, 128
 grouping, 24
 labels, 23
 namespaces, 123
 packages, 123

- parameters, 21
- ports, 23
- selecting, 19
- shape, 123
- ungrouping, 24
- visibility, 123
- writing new, 111

multiple inputs

- ports, 127

N

namespaces

- modules, 123

navigating

- versions, 33

non-interactive mode, 133

notes, 33

O

open

- from a database, 11
- vistrail, 11

opening from

- database, 86

optional

- ports, 127

ordering

- spreadsheet sheets, 40

output

- persistent files, 96

P

packages, 111

- “filePool“, 118
- “initialize“, 112, 115
- configuration, 117
- dependencies, 120
- modules, 123
- ports, 126
- temporary files, 118
- wrapping command-line tools, 116

pan, 15

parameter exploration, 59, 66

- adding parameters, 59
- control flow, 82
- directions, 59
- running, 59
- setting values, 59
- spreadsheet, 59, 66

parameters

- changing, 21
- exploring, 59, 82
- modules, 21

persistent files

- input, 95
- intermediate, 96
- output, 96

port

- shortcut, 123

port types

- ports, 127

ports, 21

- adding, 23
- default values, 126
- deleting, 23
- input dependency, 127
- labels, 126
- module connectivity, 127
- modules, 23
- multiple inputs, 127
- optional, 127
- packages, 126
- port types, 127

positioning

- spreadsheet cells, 40

PythonSource, 24

Q

queries, 45, 49

- by example, 45
- textual, 47
- viewing results, 49

R

redo, 15, 33

RichTextCell

- spreadsheet, 92

rows

- spreadsheet, 40

running

- parameter exploration, 59

S

save

- vistrail, 11

saving

- image, spreadsheet, 44
- spreadsheet, 44

saving to

- database, 87

search

- refine, 49

select, 15

selecting

- connections, 19
- modules, 19

server, 136

- setting values
 - parameter exploration, 59
- setup
 - database, 85
- shape
 - modules, 123
- sheets
 - adding, spreadsheet, 40
 - deleting, spreadsheet, 40
 - docking, spreadsheet, 40
 - ordering, spreadsheet, 40
- shortcut
 - port, 123
- spreadsheet, 39, 44
 - cells, 40, 41
 - cells creating, 40
 - cells positioning, 40
 - cells web browser, 43
 - columns, 40
 - default values, 40
 - layout, 39
 - modes, 41
 - modes editing, 42
 - modes interactive, 41
 - parameter exploration, 59, 66
 - RichTextCell, 92
 - rows, 40
 - saving, 44
 - saving image, 44
 - sheets adding, 40
 - sheets deleting, 40
 - sheets docking, 40
 - sheets ordering, 40
 - virtual cell, 66

T

- tab, 11
- tags, 29
 - adding, 33
 - deleting, 33
- temporary files
 - packages, 118
- textual
 - queries, 47
- toolbar, 11

U

- undo, 15, 33
- ungrouping
 - modules, 24

V

- versions, 29, 34
 - annotations, 33

- comparing, 34
- navigating, 33
- viewing, 29
- viewing
 - versions, 29
- viewing results
 - queries, 49
- virtual cell
 - spreadsheet, 66
- visibility
 - modules, 123
- vistrail
 - close, 11
 - definition, 14
 - open, 11
 - save, 11
- visual diff
 - see versions, comparing, 34

W

- web browser
 - spreadsheet cells, 43
- workflow, 14
- wrapping command-line tools
 - packages, 116
- writing new
 - modules, 111

Z

- zoom, 15