Collaborative Electronic Records Project

EMAIL PRESERVATION PARSER

Installation Guide

December 2008

Preface

The Email Preservation Parser was developed as part of the Collaborative Electronic Records Project (CERP). The Rockefeller Archive Center and the Smithsonian Institution Archives partnered in this three-year project to research and implement a system and tools for the preservation of digital records with an emphasis on the special challenge of preserving email. The project was funded in large part by the Rockefeller Foundation.

Contents

Introduction

The Email Preservation Parser is designed for use on a computer workstation by individuals familiar with the normal operation of desktop computers.

The purpose of the Parser is to migrate groups of email records into an XML file that captures the email records in situ, complete with their attachments, i.e., in the organizational context in which they were kept by the email account owner. This is referred to as an email account since this is most frequently how email is transferred to the CERP partner archives.

The Parser is designed to be used with groups of email records that have been separated from their original email system and transferred into the custody of an archival organization.

The Parser

The Components

The Parser consists of an application and a web-based interface running in a virtual machine environment. The preservation parser application runs on a Smalltalk virtual machine environment called Squeak. The web-based user interface (UI) runs as a web service within the Squeak environment, almost eliminating the need to work directly in the standard Squeak interface.

The Squeak virtual machine environment works with a wide variety of operating systems including Microsoft Windows XP Pro, Linux, and Apple Macintosh OSX. The web-based UI is designed to operate within the Smalltalk environment and does not require access to the Internet.

Parser testing has focused on the Microsoft Windows XP Pro environment, with Mozilla Firefox version 2 and Microsoft Internet Explorer (IE) versions 6 and 7 web browsers. Some testing of the base parser, without the Web-based user interface, was done in a Linux environment. However, the full parser tool, including its Web UI, has been tested in Windows XP Pro only.

These installation instructions have been written to be used with either IE or Firefox. To install the Parser, you will need a ZIP tool such as WinZip or StuffIt. You will also need rights to install applications on your computer workstation.

Installing the Squeak Virtual Machine Environment

Download the Squeak environment.

Using your web browser, go to http://squeak.org. This site provides lots of information about this virtual machine environment. However, the Email Preservation Parser uses an earlier version of Squeak, so do not download Squeak from this URL.

When you are ready to download Squeak, go to the Squeak archive at:
http://ftp.squeak.org

Scroll down to **3.9/** and click on it. You will see:
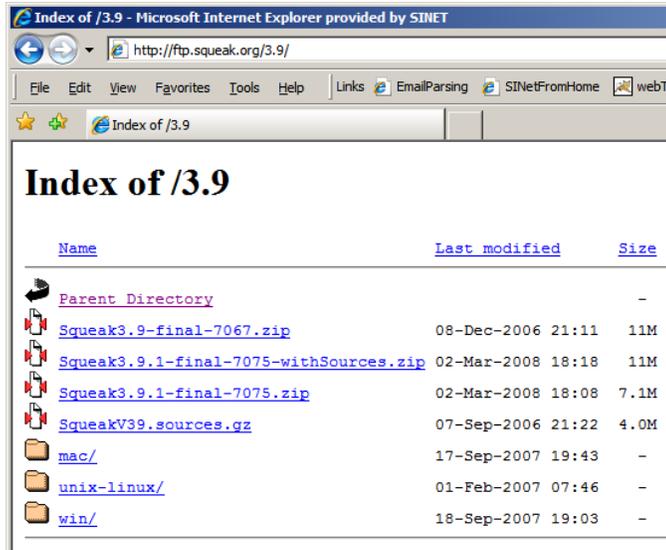
Click on **win/**. You will see:
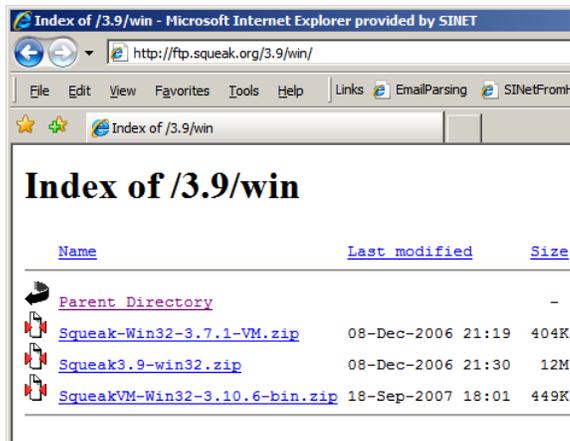
Click on **Squeak3.9-win32.zip**. Choose Save and select the location, such as the C:\ drive, where you will temporarily store this zip file. Close your web browser.

If your ZIP tool can create new folders when unzipping a file, specify **EmailParser** as the new folder into which it will unzip the components.

If this feature is not available in your ZIP tool, create a folder labeled **EmailParser**.

> Open **My Computer** and navigate to the **C:** drive.
> From the **File** menu, choose **New**, then **Folder**.
> Rename the new folder **EmailParser.**

Open the Squeak3.9-win32.zip file and extract its contents in to the EmailParser folder you just created.

> **Note**.: Depending on your ZIP extraction tool, the contents you extracted might be located in a subfolder inside the EmailParser folder. If this is the case, move those files from that subfolder into the EmailParser folder.

Once this is complete, you should have seven files in the EmailParser folder. (The image below displays all filename extensions such as .exe and .dll.)

Name	Size
↰ .. (up a level)	
.DS_Store	7 KB
Squeak.exe	1,078 KB
Squeak3.9-final-7067.changes	107 KB
Squeak3.9-final-7067.image	16,406 KB
SqueakFFIPrims.dll	31 KB
SqueakV39.sources	17,585 KB
WelcomeSqueak39	4 KB

Next, download the remaining Email Preservation Parser files. To do this, go to the CERP website's Parser download page http://siarchives.si.edu/cerp/parserdownload.htm. Click on **parsercomponents.zip** and save this file to your desktop.

Extract the contents of this ZIP file into the **EmailParser** folder you created earlier. The contents of the zip file are:

➢ CryptoAdditions.2.cs
➢ index.html
➢ SHA1Plugin.dll

- SHA1-Plugin.st
- EmailParser-Misc.5.cs
- EmailParser-Model.st
- EmailParser-WebUI.st
- images (folder containing MailAccountBkgrd.gif)
- Email_Accounts (empty)

Note: If a subfolder was created during extraction of this second ZIP file, move this second set of files and folders as you did in the previous step.

Note: Some ZIP tools may not extract the Email_Accounts folder because it is an empty folder (you will use it later.) If this is the case, simply create a new subfolder in the EmailParser folder. Label this subfolder Email_Accounts.

You are finished downloading files! You should have the following contents in your EmailParser directory:

- CryptoAdditions.2.cs
- EmailParser-Misc.5.cs
- EmailParser-Model.st
- EmailParser-WebUI.st
- index.html
- SHA1Plugin.dll
- SHA1-Plugin.st
- Squeak3.9-final-7067.changes
- Squeak3.9-final-7067.image
- Squeak.exe
- SqueakFFIPrims.dll
- SqueakV39.sources
- WelcomeSqueak39
- images folder (folder containing MailAccountBkgrd.gif)
- Email_Accounts folder (empty)

Now you will teach Windows to open the Squeak **images** file with Squeak.exe automatically.

Right click on **Squeak3.9-final-7067.image** and choose **Open**. You will be prompted to select a program to use. Choose "Select program from a list."
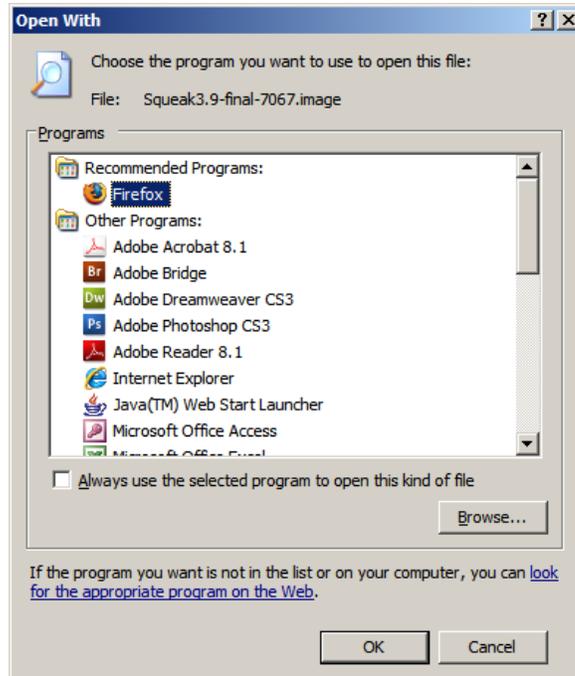
Click on Browse and navigate to **Squeak.exe** in the EmailParser folder. Select it and click on the Open button.

The Squeak icon will now appear in the list of available programs. Make sure to check "Always use the selected program to open this type of file" and choose OK.
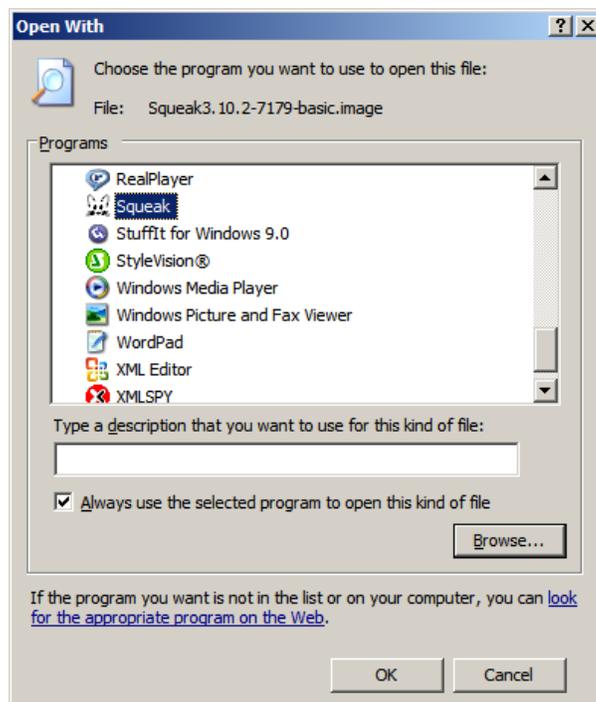
Launch the default Squeak environment

Double-click on the **Squeak3.9-final-7067.image** file. The Squeak basic image desktop will open. You will see the screen below. This is confirmation that the basic Squeak virtual machine environment has been successfully installed.
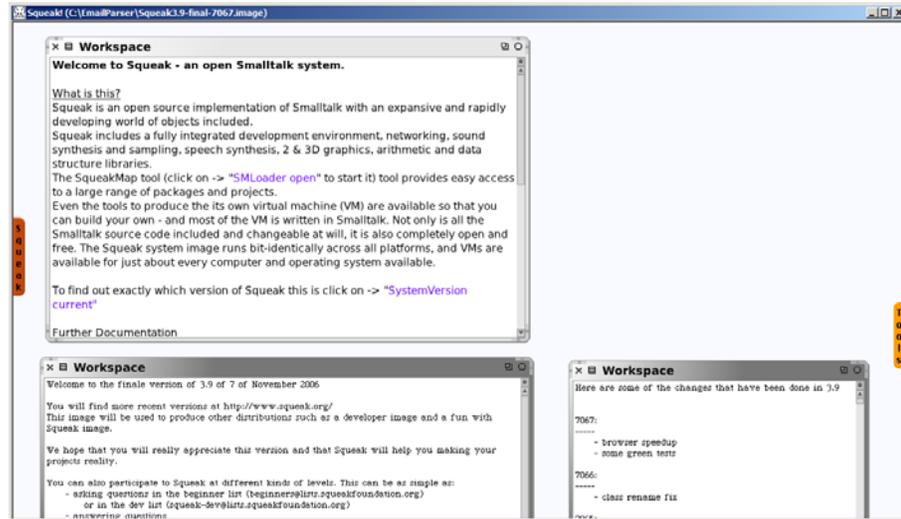
Now that you have verified the installation works, it is time to back it up. Close the Squeak window by clicking the "X" box in the far upper right corner of the Squeak window. When prompted, choose to exit without saving. Backup your Squeak installation by making a copy of the entire **EmailParser** folder and its contents.

Create the EmailParser Project

1. Launch Squeak again.
2. Right click to open the "world" menu. Select "Project," then "open."
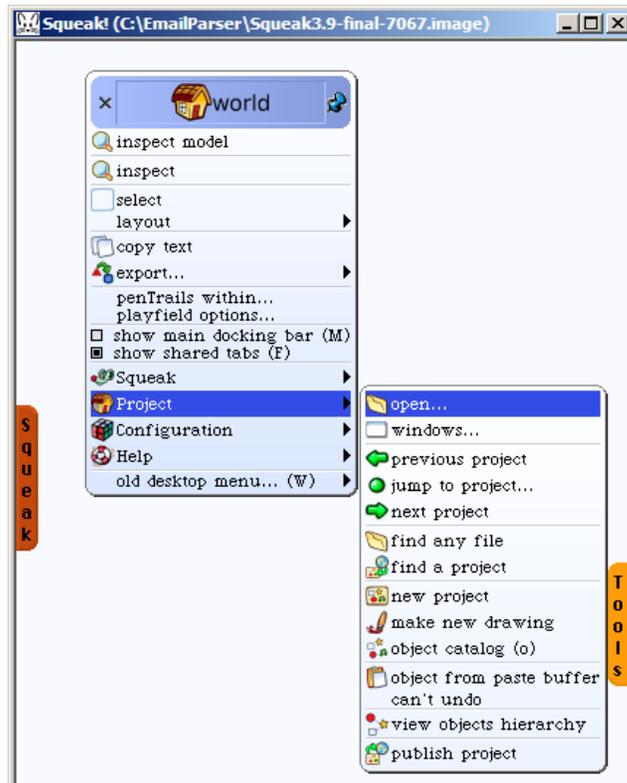
3. This brings up a new set of menu options. Choose morphic project from the bottom of this window.
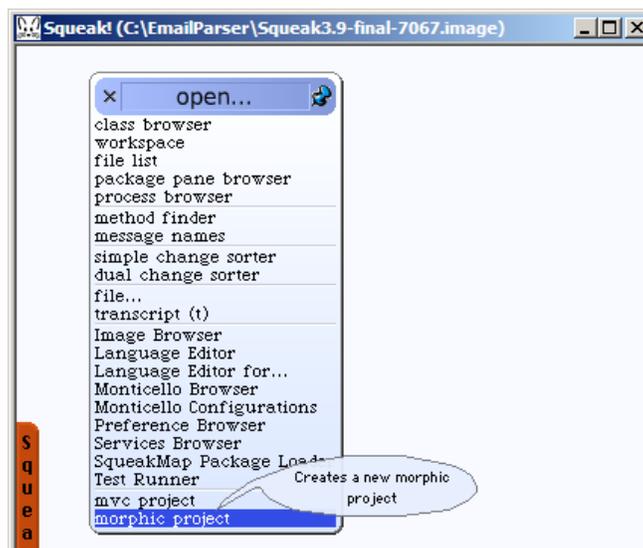
4. Now click on the window menu icon in the upper left corner to name your new project.
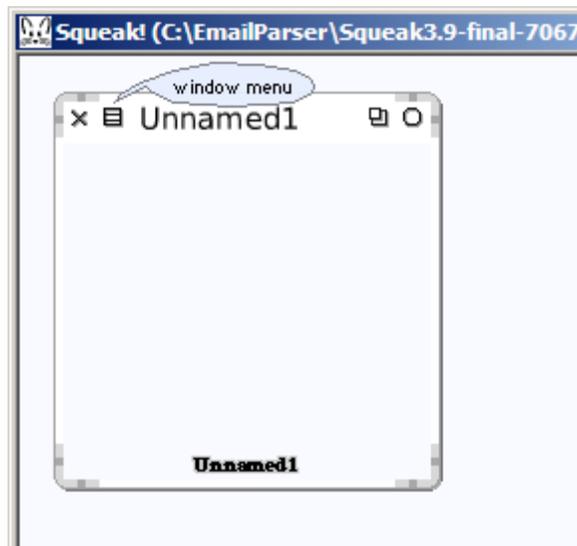
5. Choose "change title" from the drop down menu and type in "EmailParser" then click on "Accept."

You should have the following screen at this point.
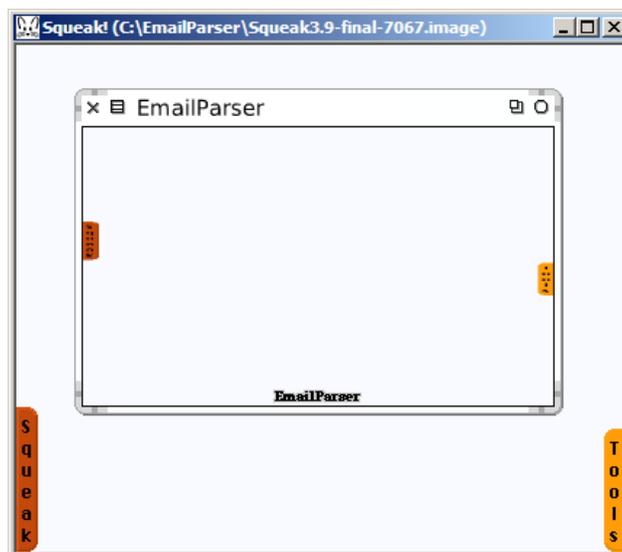
6. Click in the middle of the EmailParser window to enter the project. The EmailParser window will expand to fill the Squeak window.
7. Click on the Tools button at the right of the window. A toolbar of screen options will appear.
8. Drag the following screens into the middle of the window:

Transcript, Workspace, and File List. You may need click on the Tools button when you are ready to drag the next screen onto Squeak window.
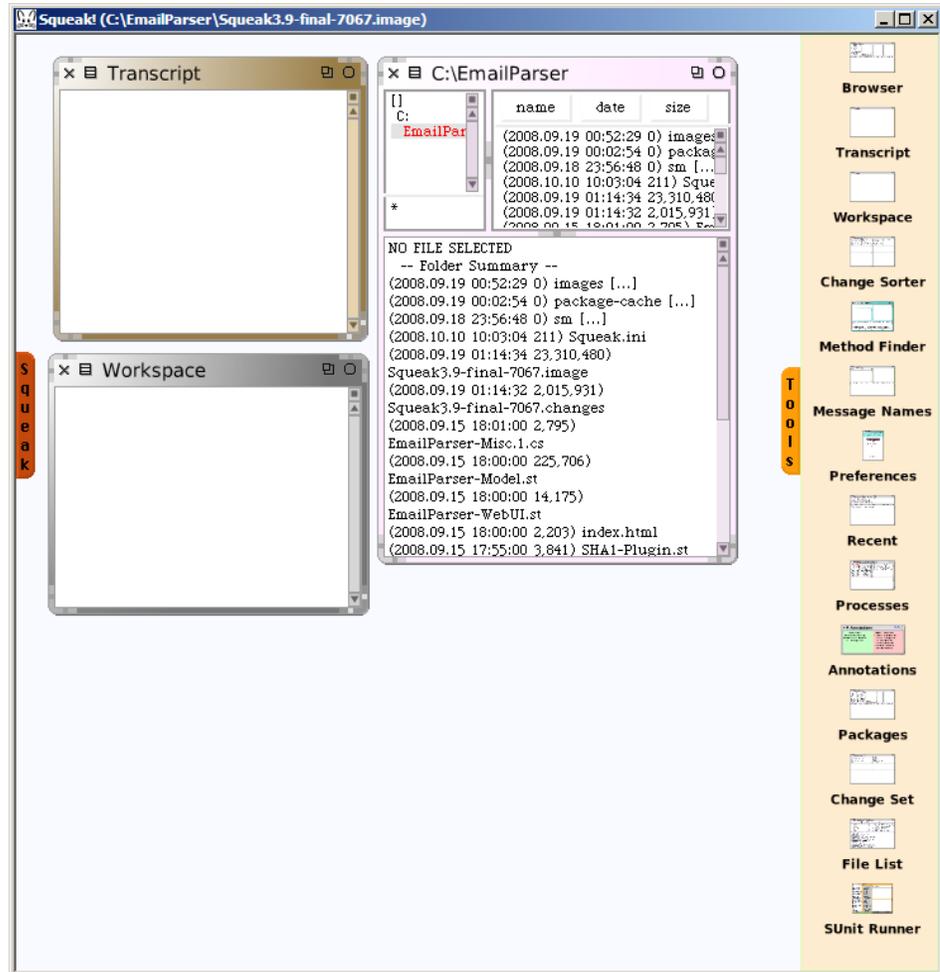
NOTE: The File List window will reflect the full path from your home drive to the EmailParser folder you created when installing Squeak. In the example, we created the folder directly on the C:\ drive.

9. If the Tools bar does not retract, click on the "Tools" tab to hide the bar.

Install the Web-UI Service

1. Click on the main screen of the Squeak window to display the World menu options.
2. Click on "open" to display the "open" menu options.
3. From that list, select SqueakMap Package Loader.
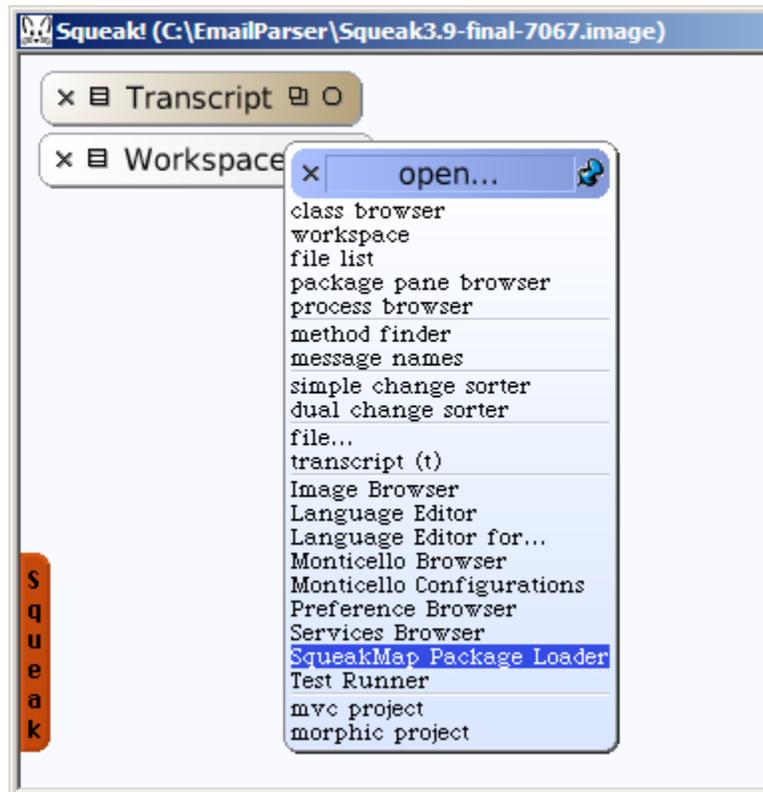
4. In the upper left hand pane of the window that appears, scroll down and select "UpdateStreams for SqueakMap."
5. Right click and choose "update map from the net."
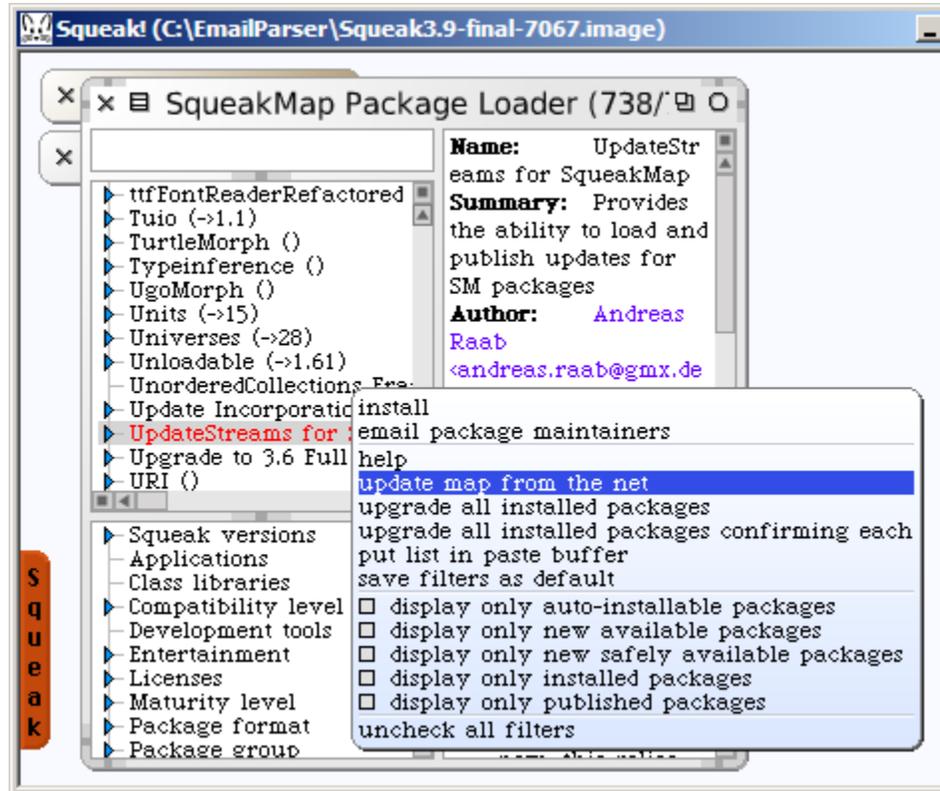
6. When that is complete, scroll up in the upper left pane to "DynamicBindings (1.21)".
7. Click on the triangle symbol to the left to expand the DynamicBindings list.
8. Find the DynamicBindings 1.21 entry and select it.
9. Right click and choose "install."
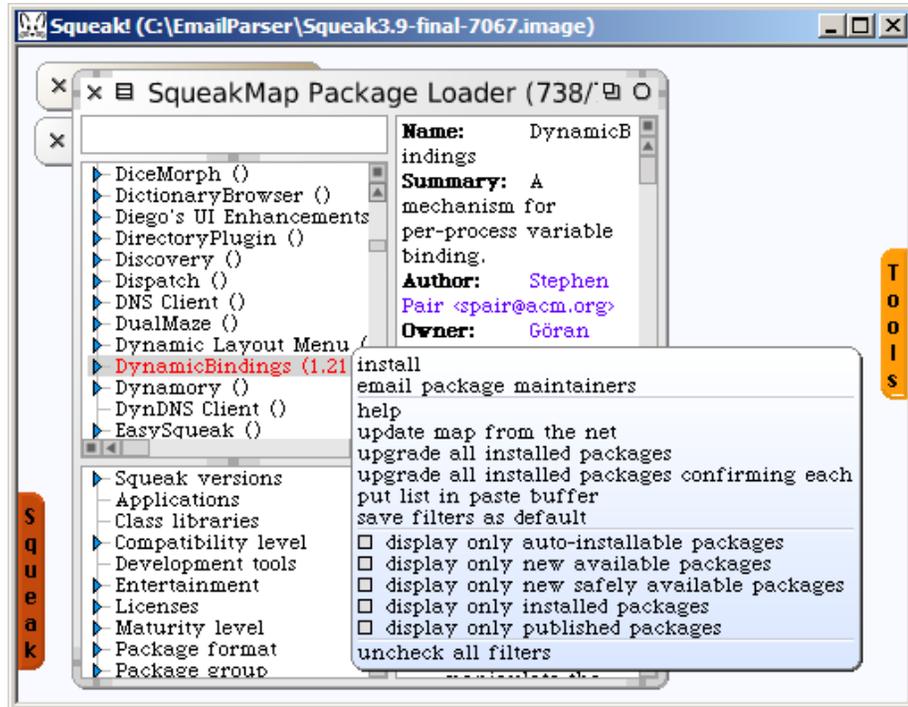
10. Repeat the process for the following files. Again order and version are very important. Click on the triangle symbol to expand a list and select the correct version.

> ➢ KomServices v1.1.3.1
> ➢ KomHttpServer v7.0.5
> ➢ Seaside v2.8.2

When you "install" Seaside, you will be asked if you would like to install the Kom server. Click NO, you have already installed it. You will then be prompted to enter a username and password. You are setting this administrator level information so you can modify Seaside configuration later on, although you will not need to do so for the EmailParser. For the purpose of completing the installation, enter **admin** as your username and **seaside** as the password. We recommend you change the username and password *after* the installation is complete, using the script in the Appendix.

The Seaside install will also ask you if you want to install Scriptaculous, Comet, and RSS Support. Answer yes to each question.

11. Launch the web service, internal to Squeak, on port 9090. To do this, type or cut and paste the following into the bottom of the Transcript window exactly as seen below.

```
| oldService ma |
oldService := TcpService serviceOnPort: 9090 ifAbsent: [nil].
oldService notNil ifTrue:
        [oldService stop.
         TcpService removeService: oldService].
ma := ModuleAssembly core.
ma serverRoot: FileDirectory default fullName.
ma documentRoot: FileDirectory default fullName.
ma directoryIndex: 'index.html index.htm'.
ma serveFiles.
(HttpService startOn: 9090 named: 'httpd') plug: ma rootModule.

WAKom stop.
WAKom startOn: 9091
```

12. In the Transcript window, select the text you have entered, right click, and choose **do it**.

13. Test that the service is running by opening your browser to this URL http://localhost:9090. You should see this screen:
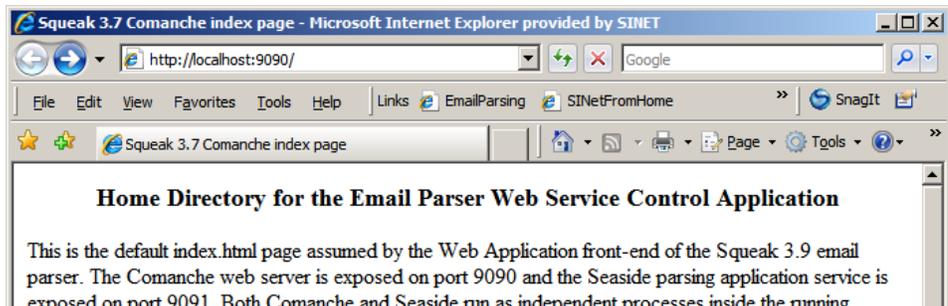
14. Now test that the service is full functional. Go to http://localhost:9091/seaside/config. You will be prompted to enter a username and password. Enter the **admin** and **seaside** defined during the Seaside installation.

The first screen you will see is

seaside ⭐

Seaside2.8a1-lr.563.mcz 8 June 2008 5:07:44 pm

/seaside/

browse	Dispatcher Viewer	Configure	Copy	Remove
comet	Directory	Configure	Copy	Remove
config	Dispatcher Editor	Configure	Copy	Remove
examples	Directory	Configure	Copy	Remove
files	Files	Configure	Copy	Remove
tests	Directory	Configure	Copy	Remove
tools	Directory	Configure	Copy	Remove

Settings

Default entry point

browse Submit Query

Add entry point

Name:
Type: Application Add

New Session Configure Toggle Halos Profiler Memory Terminate XHTML 38/0 ms

15. Click on the **examples** link, then the **counter** link on the next page that displays. After having done this, you will see:
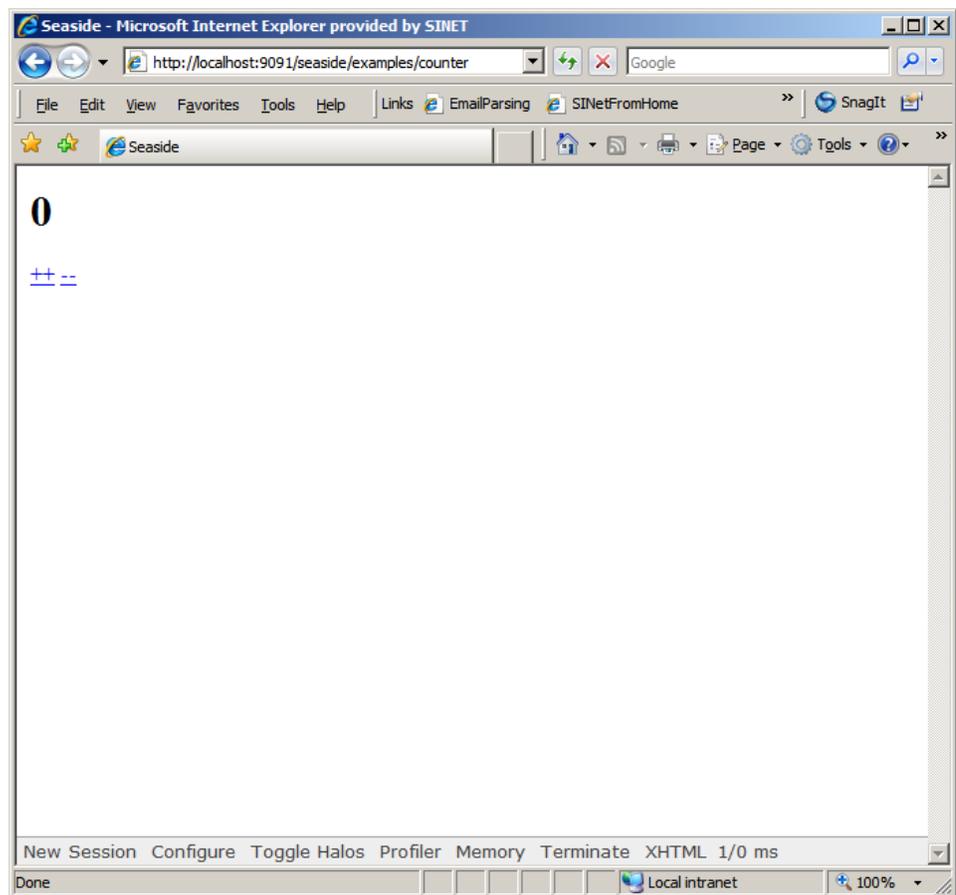
16. Clicking on either the ++ or -- should increment or decrement the counter respectively. Having achieved these results, the Seaside web service has been confirmed to be functioning properly.
17. Close your web browser.

"File In" the Parser Components

You will now begin to "file in" a number of files to the EmailParser project. To do this you will use Squeak's File List screen.

Navigating in the File List window.

The two upper panes will be where you work in this window. The upper left pane shows where in the computer's file directory tree you are and can be used to navigate *up* the tree only.

The upper right pane shows the contents of the folder you are in. To select a file or folder in this pane, click once. To open a folder in this pane, click twice.

"File in" the EmailParser files

1. Make sure the folder "EmailParser" is selected in the upper left pane. If it is not, then navigate to it.
2. In the upper right pane, find the file "CryptoAdditions.2.cs".
3. Click once to select it.
4. Now right click and choose "install into new change set" from the menu options that display.
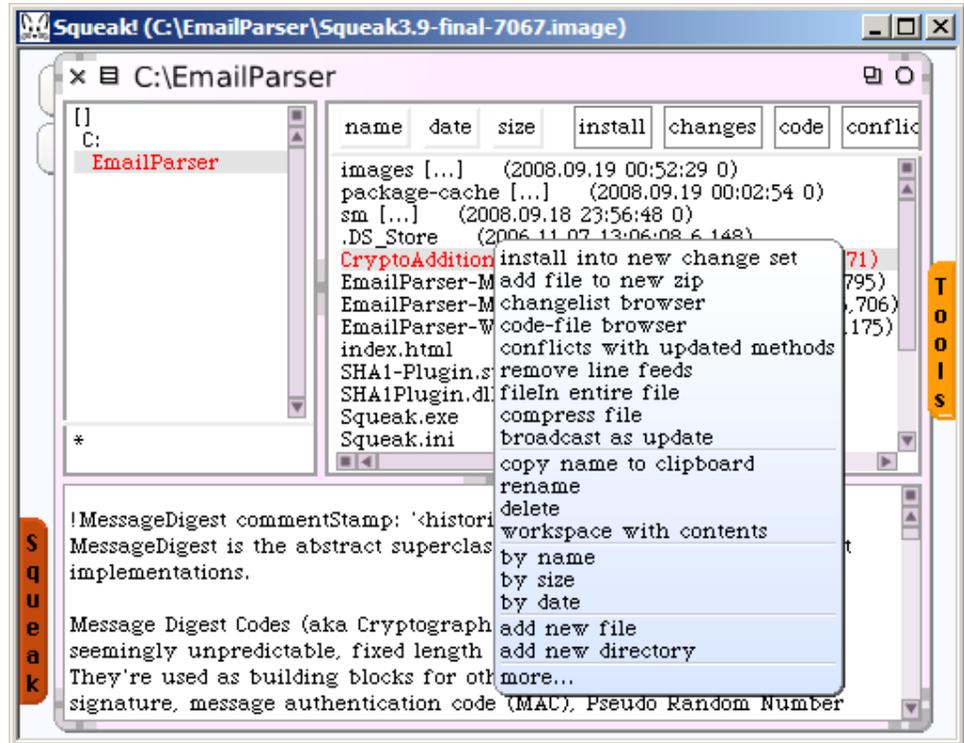
5. Repeat the process for the following other files, *in order*:
   - ➢ EmailParser-Misc.5.cs
   - ➢ EmailParser-Model.st
   - ➢ EmailParser-WebUI.st
   - ➢ SHA1-Plugin.st

6. Close the File List window by clicking on the "x" in the upper left.

Configure the Web User Interface

Configure the Seaside web service to recognize the Parser

1. Open your web browser to **http://localhost:9091/seaside/config/**. You may be prompted to log in again.

2. Under **Add entry point**, enter the name **EmailParsing** and click **Add**.

Adding EmailParsing as an application will open the configuration page for the EmailParsing service. Towards the bottom of the screen, you will see the field **Root Component**.

3. Click the drop-down menu and select **EmailParserRootComponent**.
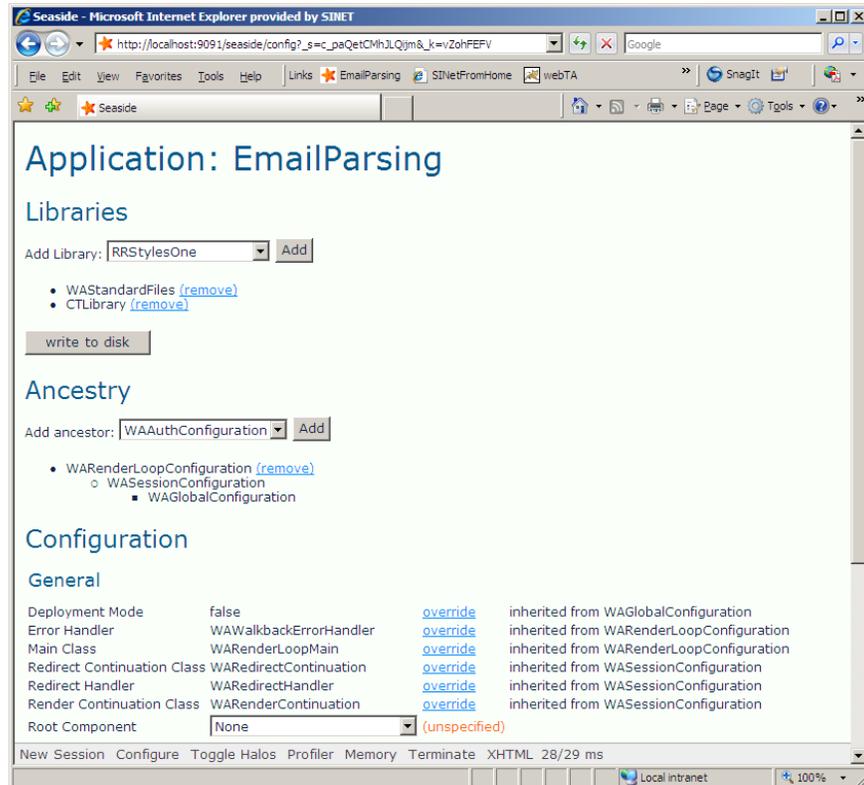
4. Go to the bottom of the page and click **Save.**
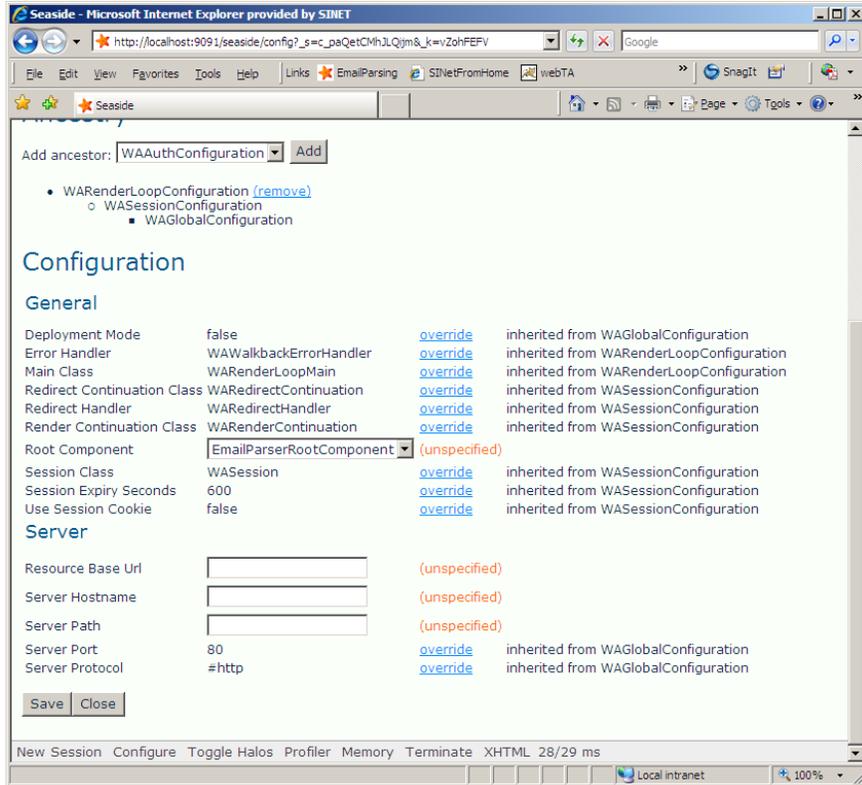
5. Finally click the **Close** button. This will return you to the original Seaside config page. You will note that **EmailParsing** is now on the list of links at the top of the page.
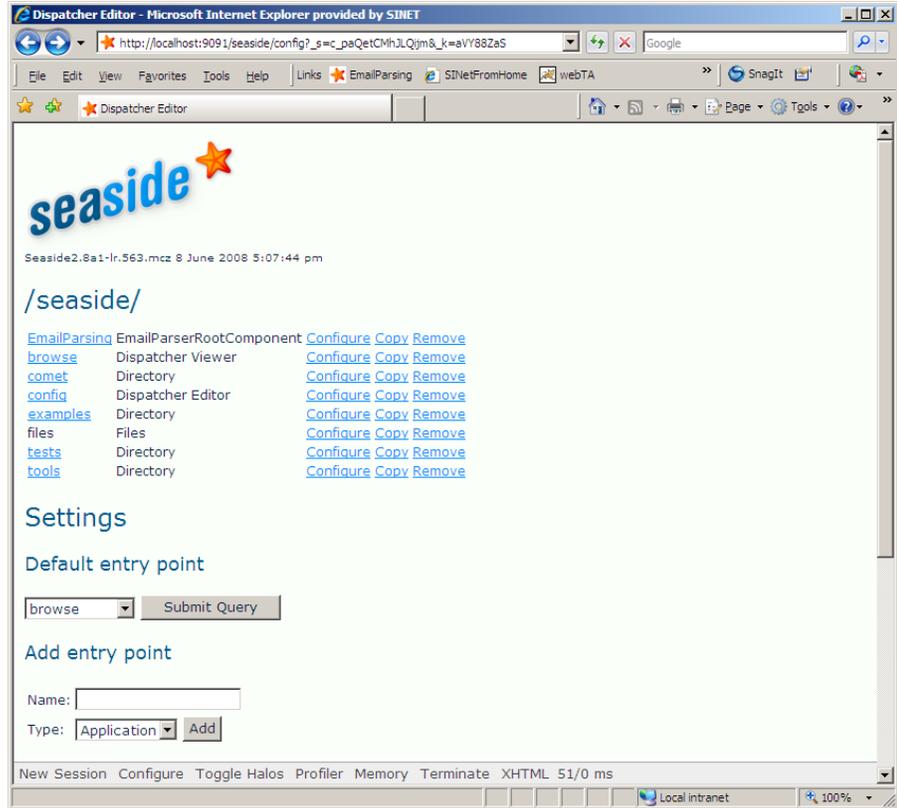
Now you are ready to launch the Parser. But first, save your work up to this point

Saving the Squeak Image

1. Click on the Squeak window to display the "World" menu. At the bottom of the menu click **save.**
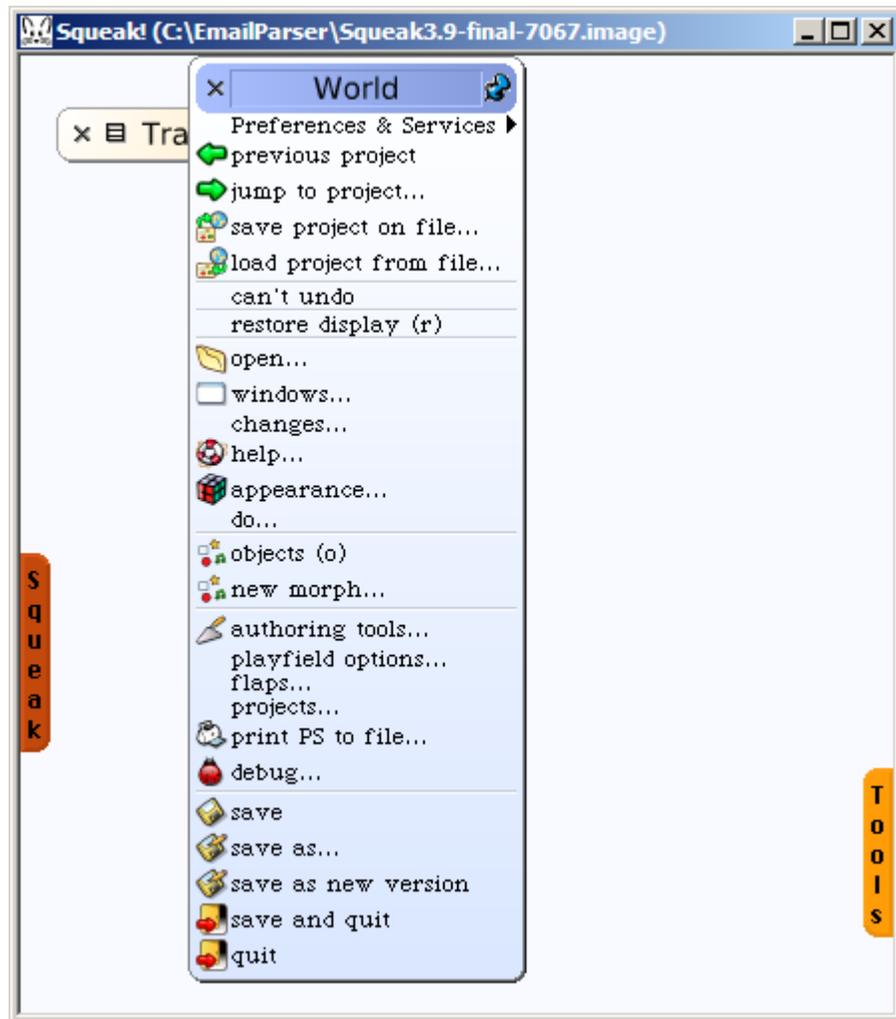
Launch the Parser

1. Open your web browser and go to
    http://localhost:9091/seaside/EmailParsing/.
    The URL is case-sensitive.
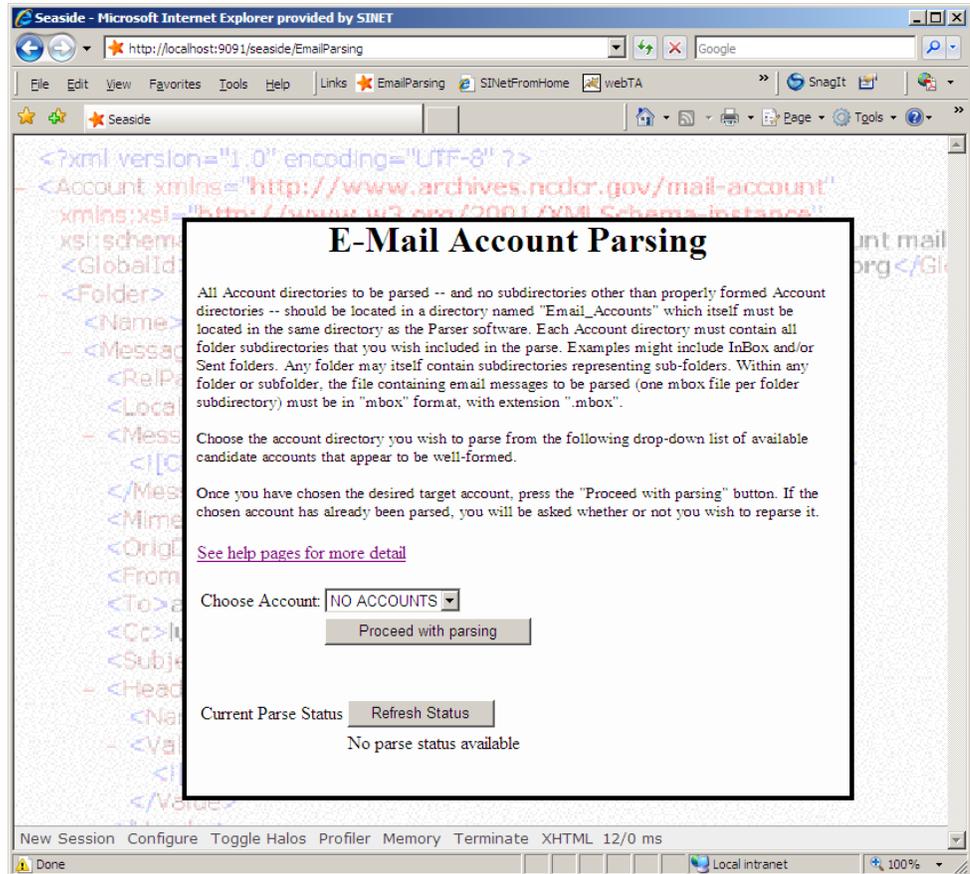
The Email Preservation Parser installation is complete.

Appendix: Changing the admin username and password

In the Transcript window, type or cut and paste the following. Be sure to replace **MyNewLogin** and **MyNewPassword** with your own choice for administrator's user name and password.

```
(WADispatcher default entryPoints at: 'config')
        preferenceAt: #login put: 'MyNewLogin'.
```

Then select and then "do it."

```
(WADispatcher default entryPoints at: 'config')
        preferenceAt: #password put: 'MyNewPassword'.
```

Then select and then "do it."

To check your login and password:

Paste these two expressions into a Workspace or the Transcript and do a "print-it" on each:

```
(WADispatcher default entryPoints at: 'config') preferenceAt: #login
```

```
(WADispatcher default entryPoints at: 'config') preferenceAt: #password
```