

rstWeb

User Guide – Version 1.0.0

title: rstWeb User Guide

rstWeb version: 1.0.0

guide version: 1.0.0

date: 28 June 2015

author: Amir Zeldes

e-mail: amir.zeldes@georgetown.edu

homepage: <http://corpling.uis.georgetown.edu/rstweb/info/>

Contents

Introduction.....	2
Getting Started	2
Server Installation	2
Logging in	3
Annotation.....	3
Opening a document	3
Segmenting units.....	4
Editing structure.....	5
Administration	6
Projects.....	6
Documents	7
Import.....	7
Users	8
Database.....	8
Configuration file.....	9

Introduction

rstWeb is an open source, browser based annotation tool for discourse analyses in Rhetorical Structure Theory. It is meant to support collaborative, online annotation projects using just a Web browser, without the need to install software for annotators, though there will also be a standalone local version (**coming soon!**) for offline use if you do not have access to a server.

Getting Started

Server Installation

1. Make sure Python 2.X is installed (preferably 2.6 or newer)
2. Unpack all of the files to the directory they will be served from
3. Configure your Web server to have read, write and execute privileges in the relevant folder
4. You may want to forbid users from interacting with files other than the top level python scripts in the main rstWeb directory (in particular, no one should have access to the configuration files)
5. If you're using Apache, here is a possible configuration file:

```
Alias "/rstweb" "/var/www/html/rstweb"
<Directory "/var/www/html/rstweb/">

RewriteEngine On
RewriteBase /
DirectoryIndex open.py

<IfModule mime_module>
AddType application/x-httpd-py .py
</IfModule>

RedirectMatch 404 ".+\. (py|c|o|db|txt|rs3|ini)$"
RedirectMatch 404 ".*/(modules|export|import|templates|users).*$"

AddType text/html *
Options Indexes FollowSymLinks MultiViews
Options +FollowSymLinks
Options +ExecCGI
AddHandler cgi-script .py
AllowOverride None
Order allow,deny
allow from all
</Directory>
```

6. Use the administration interface to change the passwords/user names to secure your system.
7. You're all set!

Troubleshooting

If you're having trouble, it's possible some permissions are set wrong, or that your server needs to be configured to execute the Python scripts. Otherwise, the entry point for the program is the script *open.py*. If you're using the Apache configuration above, this acts as the directory index, so you can simply direct users to `http://.../<rstwebsdirectory>/`.

Logging in



Local version – coming soon!

If you're using the local version, you can use the default login and password: admin <> admin.
(Not yet available)

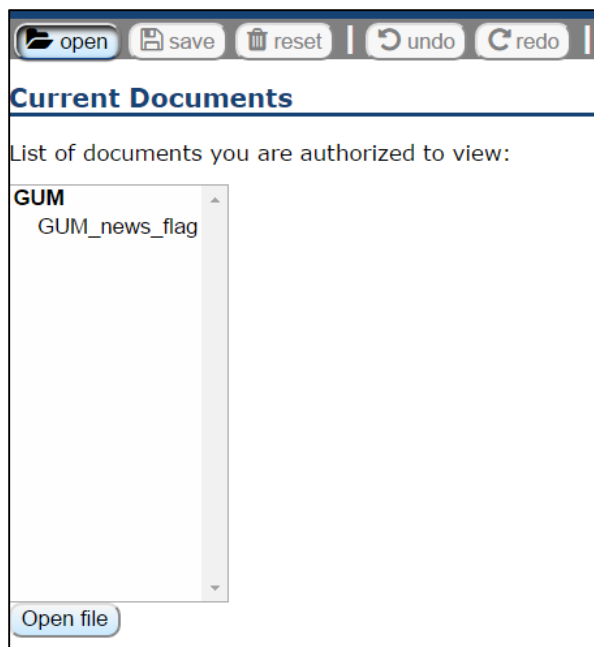
Server version

If you are working on a server over the internet, a login will be provided for you by an administrator (see User Management below). To log in, go to the server URL, e.g. <http://corpling.uis.georgetown.edu/rstweb/> - if you are not already logged in, you will be prompted to enter your credentials.

Annotation

Opening a document

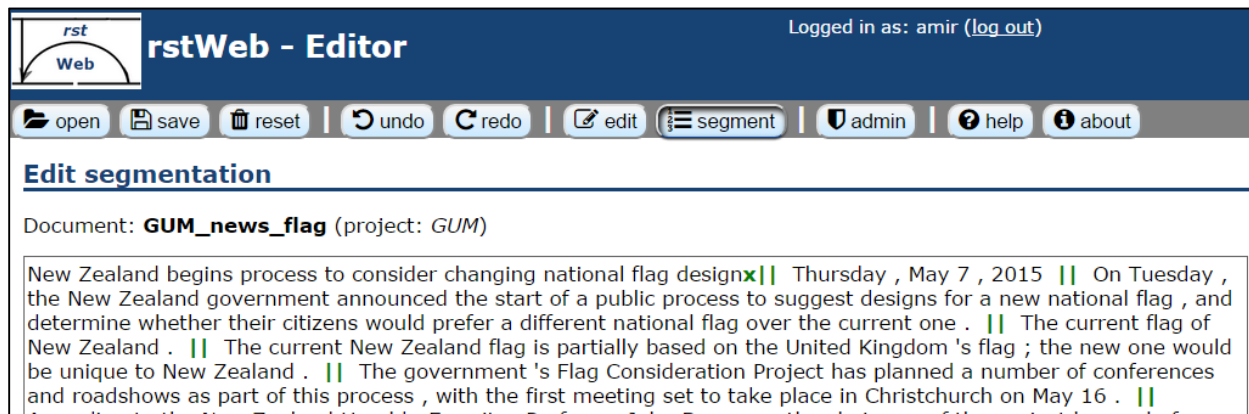
Once you are logged in, you will see a list of the documents you are allowed to edit, grouped by annotation project. You can return to this screen using the *open* button.



Administrator users may see all documents. Note that every user has their own copy of the documents they are working on, so that multiple annotated versions of each document can exist simultaneously.

There are two modes while annotating: **segmentation mode** and **edit structure mode**.

Segmenting units



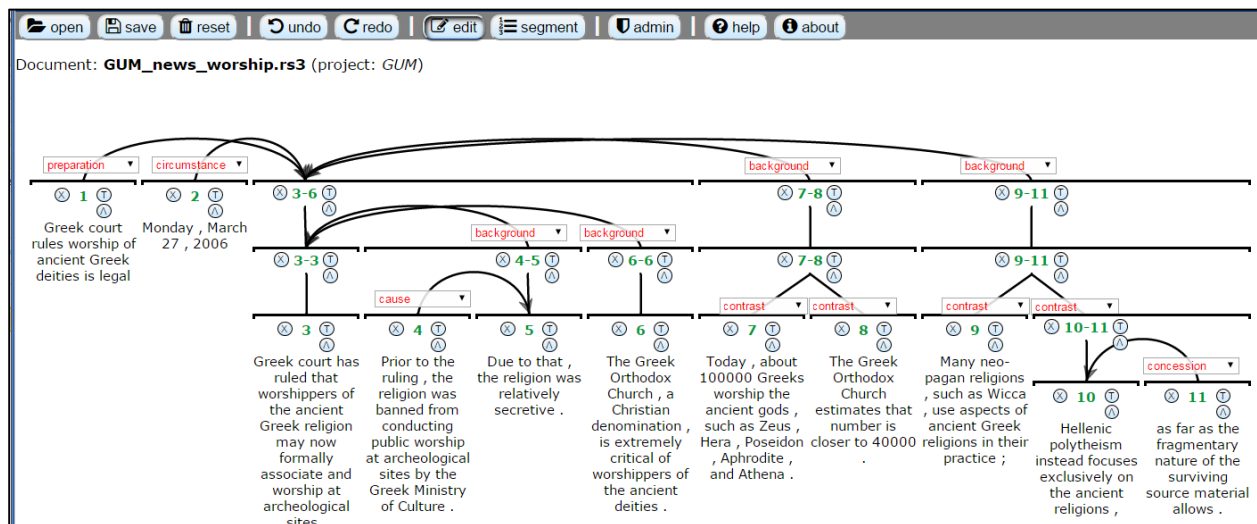
In segmentation mode, you can split the text into Elementary Discourse Units (EDUs). Initial EDU borders can be set by either importing an already segmented .rs3 file (e.g. from RSTTool), or by importing a plain text file with one EDU per line (see Import under Administration). You can alter segmentations at any time by pressing the *segment* button.

To add segments, click on a space between words. To remove segments, hover over the green dividers between segments and click on the x that appears to remove them.

It is possible to add and remove segmentations of already structured texts. In this case, adding a segment will cause the first part of that segment to remain linked to the annotation graph, and the second part will be inserted as a new unattached EDU for annotation. Deleting a segmentation border will merge the second unit into the first, so that the relations involving the first part of the new merged segment are retained, but those for the second part are deleted.

Editing structure

To annotate RST relations use the *edit* button. The interface below shows an annotated document with relations already in place.




In edit mode, each elementary discourse unit (EDU) has its text under a line with a number giving its position in the text. Similarly, groups of units have a range of numbers. You can connect nodes, unlink them, and group them together under spans and multinuclear nodes. You can undo and redo actions at any point using the appropriate buttons.


Don't forget to save your work! If you make a mistake and want to go back to the original form of the document when it was imported, you can also use the *reset* button (**warning:** this will delete all of your annotations for this document!)

Connections


- To connect nodes, drag from the numbers under one unit to the numbers under a target unit

- To change the relation between two units, use the drop down list on the connection between them
- To unlink a node from the graph, click its  button - all of the nodes connected above it will be unlinked

Spans

- Use the  button to add a span above a node. The span will group together all of the nodes connected to that node

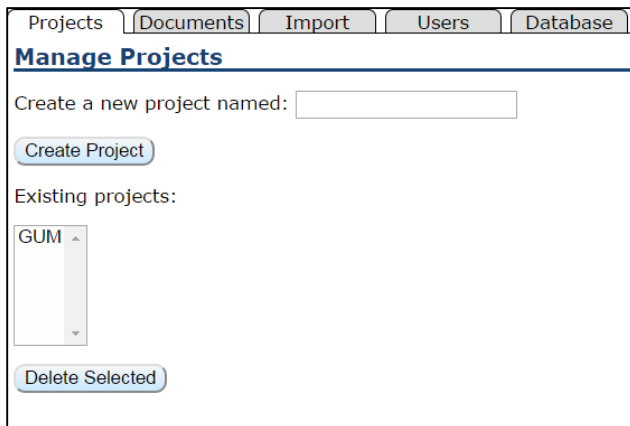
Multinuclear relations

- Use the  button to create a multinuclear node (multinuc) above a node. The multinuc can have multiple child nodes with the same relation (e.g. a set of items in contrast to each other)
- To change the relation of a multinuc to its children, change the relation box of any of its children
- When connecting a new node to a multinuc node, it will be added as a multinuc child by default. You can change the relation to a satellite relation by choosing "change to satellite" in the relation box

Administration

The administration interface is only available to administrator users (marked with level 3 in the user file in the users/ directory). It contains five tabs:

Projects



Projects | Documents | Import | Users | Database

Manage Projects

Create a new project named:

Existing projects:

GUM

This tab lets you create new projects and delete existing ones. Projects are used to group documents together. Deleting a project will remove all documents within it permanently!

Documents

The screenshot shows the 'Current Documents' tab with a list of documents under the project 'GUM'. The list includes 'GUM_news_flag' and 'GUM_news_worship.rs3'. Below the list are buttons for 'Export' and 'Delete'.

Projects	Documents	Import	Users	Database
Current Documents				
List of documents in the database:				
GUM				
GUM_news_flag				
GUM_news_worship.rs3				
Export selected document(s) to export folder as .rs3 file(s):				
<input type="button" value="Export"/>				
Delete selected document(s):				
<input type="button" value="Delete"/>				

This tab shows you a list of all documents and allows you to delete them or export existing document annotations. If multiple annotators have worked on a document, one rs3 file will be created for each annotator's version in the export/ folder.

Import

The screenshot shows the 'Import a Document' tab with three steps: 1. Upload .rs3 or plain text file: Choose File No file chosen; 2. Choose file type: rs3; 3. Import document into this project: GUM. An 'Upload' button is at the bottom.

Projects	Documents	Import	Users	Database
Import a Document				
1. Upload .rs3 or plain text file: <input type="button" value="Choose File"/> No file chosen				
2. Choose file type: <input type="text" value="rs3"/>				
3. Import document into this project: <input type="text" value="GUM"/>				
<input type="button" value="Upload"/>				

Here you can import new documents. Choose between .rs3 format (from RSTTool) and plain text (one EDU per line). When importing plain text files, the default relations that are available for annotation are determined by the file *default_rels.tab* in the *users/* configuration directory (alternatively you can choose a different location in *users/config.ini*).

The new documents will be imported into a new project and a version of the document will be created for your user. You can assign the document to other users in the Users tab. A copy of the imported document will be saved in the import/ folder and you can always revert to the imported state of the file while editing by using the *reset* button after you've opened the document.

Users

The Users tab lets you create and delete user files. These are stored as .ini files in the users/ directory.

New users can be defined as either normal users or administrators using the checkbox in the new user form at the bottom. Administrators can see all documents and have access to the administration interface.

New user names cannot be: 'config', 'default', 'temp', 'emails', '_orig' or 'pending'.

For normal users, you need to assign a user name to a document name in order to make that document visible to the annotator. You can assign multiple annotators to multiple documents simultaneously by selecting multiple lines in the Users and Documents to assign boxes. Once you assign a document, a copy of it will be created in the system which can then be annotated and exported without affecting other copies of that document.

The screenshot shows the 'User Management' window with tabs for Projects, Documents, Import, Users, and Database. The 'Users' list contains 'amir', 'guest', and 'temp'. The 'Documents to assign to:' list contains 'GUM', 'GUM_news_flag', and 'GUM_news_worship.rs3'. Below the lists are buttons for 'Delete user(s)', 'Assign', and 'Delete assignment'. At the bottom is a 'Create new user' form with fields for 'User name:', 'Real name:', 'Email address:', 'Password:', and an 'Administrator:' checkbox. A 'Create user' button is next to the checkbox.

Deleting an assignment will delete that user's version of the document, but will not affect other users' annotations and will not remove the document from the system.

Database

The screenshot shows the 'Initialize the Database' window with tabs for Projects, Documents, Import, Users, and Database. The main text reads 'Wipe and restore database structure.' followed by a red 'Warning:' and the text 'this will delete all imported documents and all edits from the data'. Below this is an 'Init DB' button.

This tab allows you to wipe the database clean and restore its original schema. This is mainly useful if the database file becomes corrupt somehow, or if you want to quickly delete all data in the database (delete all documents and projects).

Configuration file

Some aspects of the interface, especially file and directory paths for templates, default relations etc. are configured in *users/config.ini*. You can edit this file to use different templates, place import and export files in different locations, and in the future also to use automatic mail notifications to new users (not yet supported).