DRAE GUI User Manual

### Daniel J. Powell Leidos

### Draft: February 13, 2015

Contents

[Introduction: The DRAE System 1](#_Toc411933484)

[Installation 1](#_Toc411933485)

[Use 3](#_Toc411933486)

[GUI Anatomy 3](#_Toc411933487)

[Opening the GUI and Loading PDF Documents 4](#_Toc411933488)

[Annotating a table using the Quick Tag bar 4](#_Toc411933489)

[To create a table Working Set 4](#_Toc411933490)

[Annotating Columns 5](#_Toc411933491)

[Annotate Rows 6](#_Toc411933492)

[Tagging sections to be ignored 7](#_Toc411933493)

# Introduction: The DRAE System

DRAE is a system – currently still under development -- for analyzing table and diagrammatic content in PDF documents.

Current functionality is limited to marking up table and diagram content (which can be stored in a JSON overlay file) and generating JSON representations of table content.

The document explains how to install DRAE, start the GUI, markup PDF documents, and save the results as JSON files.

# Installation

DRAE is distributed as a Java 1.7 jar file. It has been tested on Windows and Mac systems, but should also run without modification on most Unix or Linux systems.

DRAE uses the pdf2xml utility to do backend processing of PDF files. So, before running DRAE, you will need to install pdftoxml and tell DRAE where to find it. The pdftoxml utility can be found at <http://sourceforge.net/projects/pdf2xml/>. For the Mac there are downloadable binaries at the SourceForge site.

DRAE needs to be given the location of the pdftoxml executable in the drae-settings.end config file.

Installations steps:

* 1. Install pdftoxml for your system.
  2. Unzip draegui.zip into a folder [ex: draegui\_install/]
  3. Open the file drae-settings.edn
  4. Edit the line :pdftoxml-executable so that the path points to the pdftoxml executable that you just installed

# Use

## GUI Anatomy

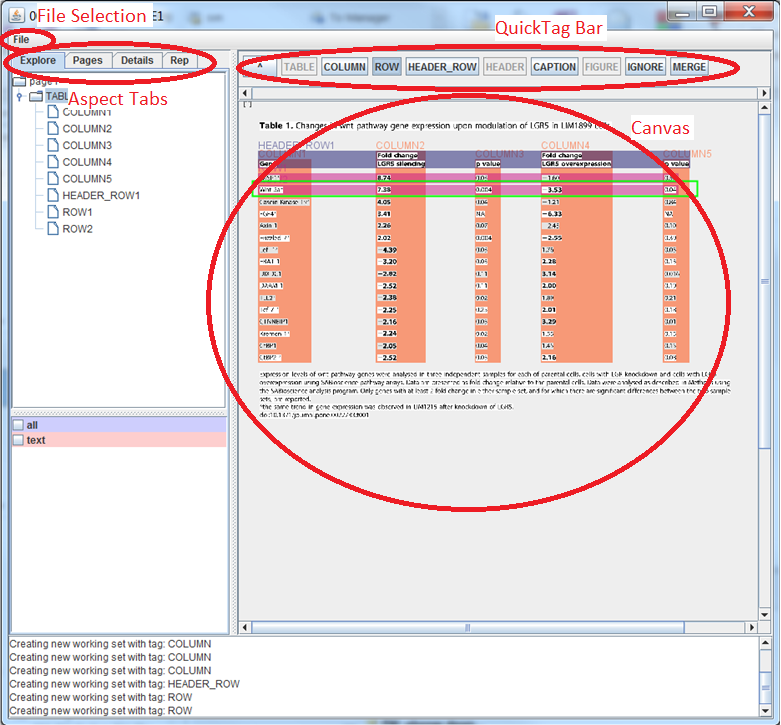


Figure . The GUI has many parts.

1. File Selection Menu: This menu allows you to open PDFs and save and load working set overlays. Options are:
   1. Browse files to open a PDF
   2. Save overlay information to a json file
   3. Load json file for the current pdf
2. Aspect Tabs: These tabs show four different views of the current working set. The tabs are:
   1. *Explore*: Navigate the working set hierarchy from the page level on down, and select the current working set.
   2. *Pages*: Navigate between pages using thumbnails
   3. *Details*: A complete view of all the PDF elements in the current working set.
   4. *Rep*: View and navigate the representation levels in the current working set. Representation levels may include text processing results, table analysis results, and other hierarchical representations.
3. Quick Tag Bar: This command bar is designed for rapid tagging of tables.
   1. Toggle on a button to enable Quick Tagging. If Quick Tagging is enabled, when a rectangle is drawn on the canvas, it will automatically create a Working Set from the selected area and tag it with the selected type
   2. Click the [^] button to navigate up one level
4. Canvas: A view of the current working set. Click to select a working set or element, double click a working set to drill down to that working set. Click and drag a rectangle to group select or quick tag. Right click to access the context sensitive menu.

## Opening the GUI and Loading PDF Documents

### Starting the GUI

1. Run the GUI by double-clicking on the jar file (if applicable for your OS)
2. Or by run from the shell via the command “java -jar draegui.jar”.
   1. Open a command prompt (“cmd” on windows, or a shell window on Mac or Unix)
   2. Navigate to the installation folder
   3. Enter the command “java –jar dragui.jar”

### Loading a PDF document

1. Wait for the GUI to appear
2. Click file from the menu bar and browse to select the PDF.
3. The file will start to load in the background

## Annotating a table using the Quick Tag bar

Although working sets can be annotated using the normal working set and tagging facility, the quick bar permits this to be done via a set of specialized buttons. Once a quick tag button is active (i.e., pressed), all selections in the document will create a working set of the given type. This is especially useful for tagging sets of columns and rows.

### To create a table Working Set

1. Load the PDF file, as described in the previous section.
2. Navigate to the page that contains the table
3. Select [TABLE] from the quick tag bar
4. Click and drag a rectangle around the table
5. When you release the mouse button, the selected region will be tagged as a table.

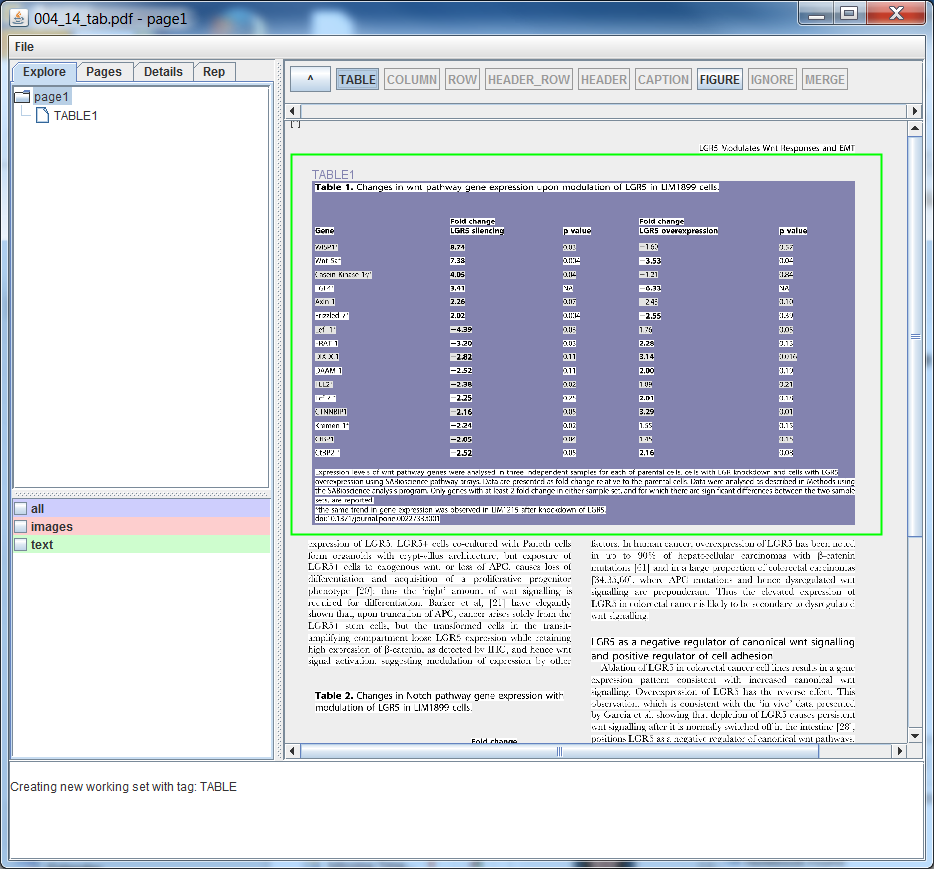


Figure Selecting the table region.

1. This will create a new Working Set with the “table” tag. Double click on the newly created table Working Set to drill down into that Working Set.

### Annotating Columns

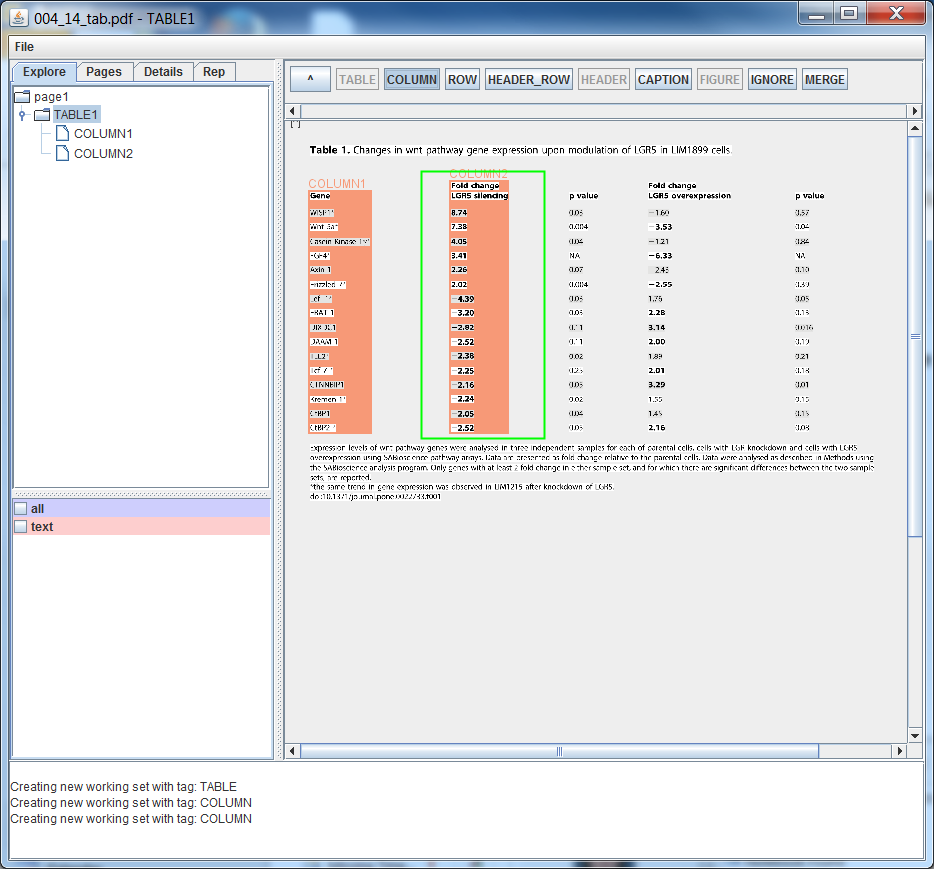
1. Navigate to the table Working Set by double clicking the tagged table region
2. Select [COLUMN] in the quick tag bar
3. Drag to select a column (including the heading for the column)

Figure Selecting a column region.

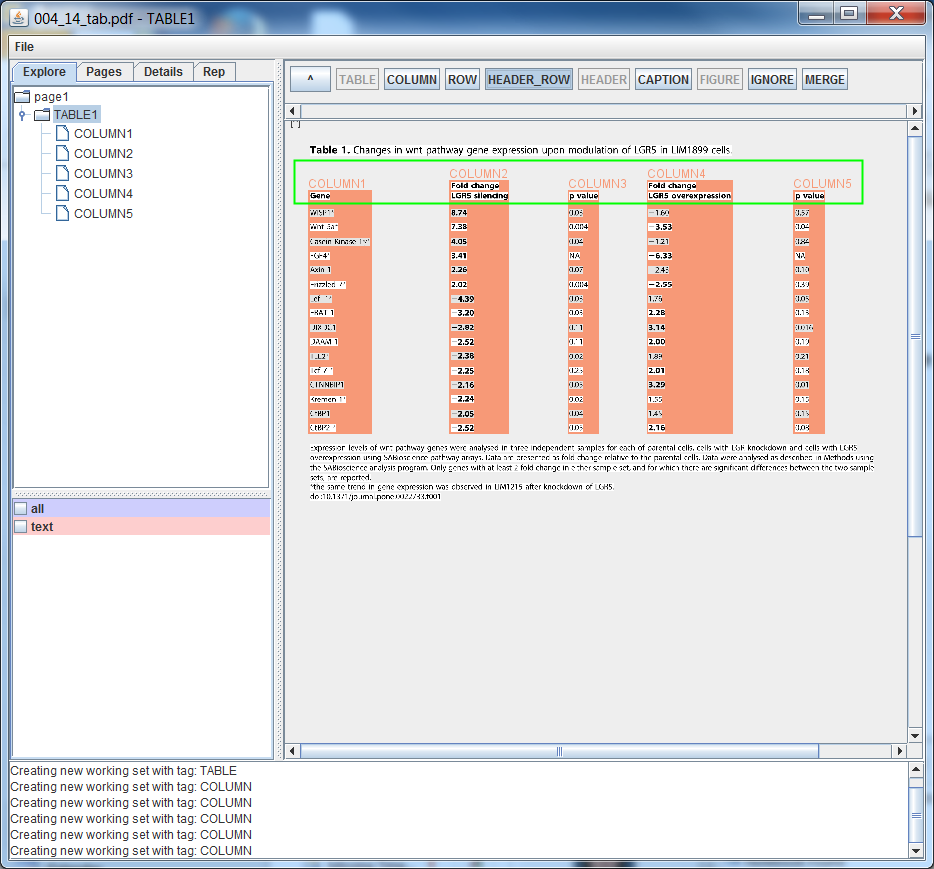
1. Drag a rectangle around each of the columns that you want to be part of the table
   1. NOTE: you don’t need to re-select the [COLUMN] quick tag button
2. After all of the columns have been tagged, click the [HEADER\_ROW] quick tag button
3. Drag a rectangle around the header rows for all columns 

Figure Selecting a header row region.

### Annotate Rows

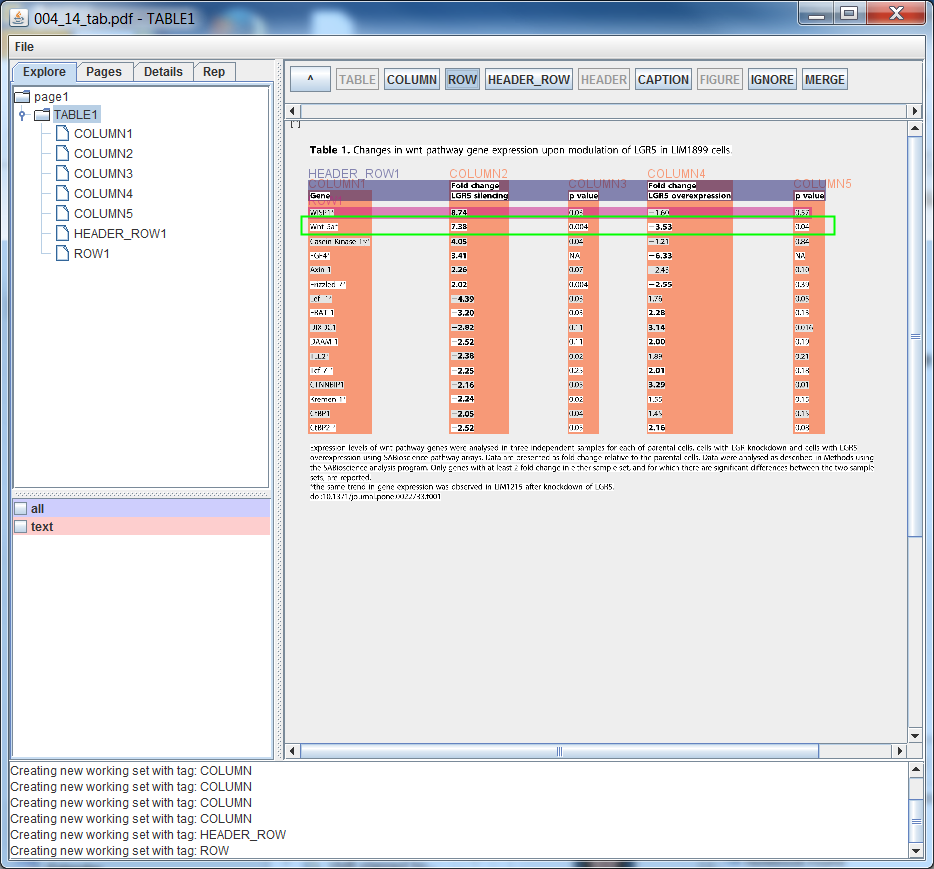
1. Navigate to the table Working Set
2. Select [ROW] in the quick tag bar
3. Drag to select a row (including the heading for the column) 

Figure Selecting a row region

1. Drag a rectangle around each of the rows that you want to be part of the table
   1. NOTE: you don’t need to re-select the [ROW] quick tag button

### Tagging sections to be ignored

If there are sections within a column that should not be taken into account by the table analysis systems they can be tagged to be ignored

1. Select the ignore Quick Tag button
2. Drag a rectangle around the area to be ignored
3. The selected area will be tagged as an ignored region
4. Unselect the [IGNORE] Quick Tag button