JGAAP 5 User Guide

Written by: Jean Zambri

Files may be uploaded in one or more of the following formats: Word documents (.doc or .docx), PDF files (.pdf), HTML (.htm or .html) or Text files (.txt). JGAAP will convert documents in various formats to plain text as they are loaded.
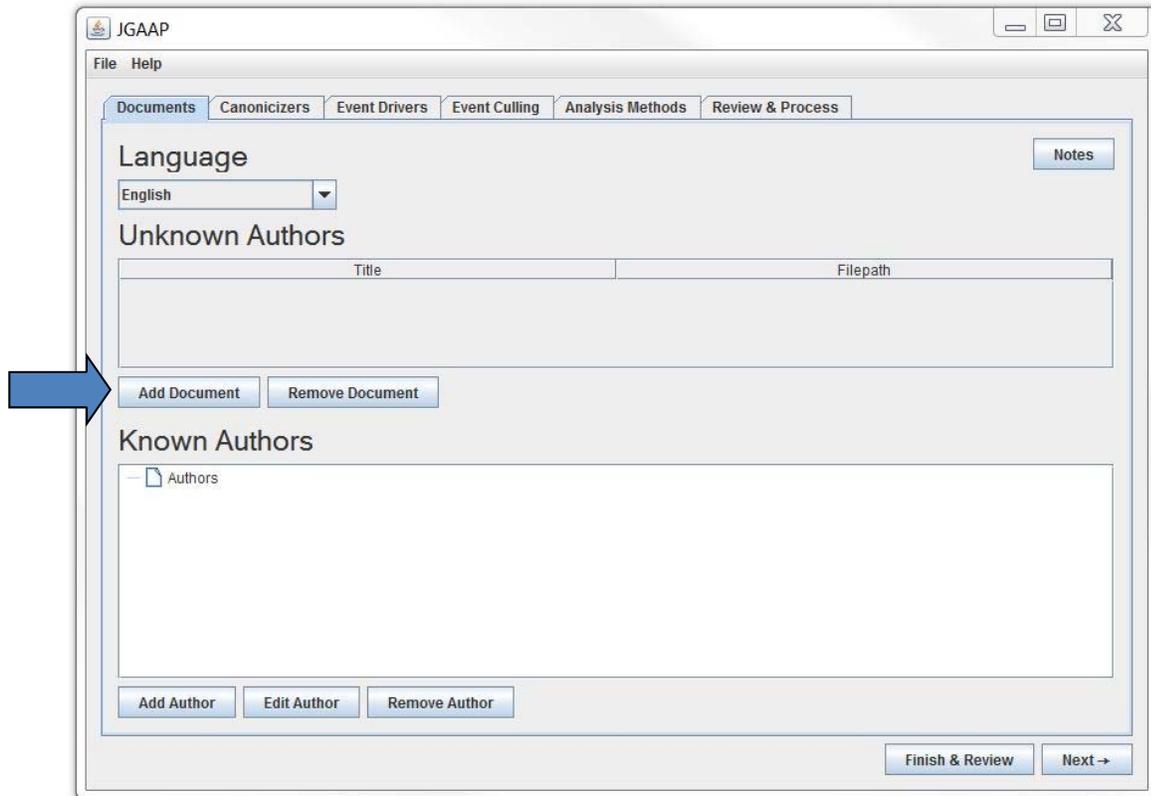
To add a document to the **Unknown Authors** box, click on "Add Document" (blue arrow). A window will open (see picture below) where you may select your desired documents. After making your selections, click on "Open" (purple arrow) to add the documents to the **Unknown Authors** box.
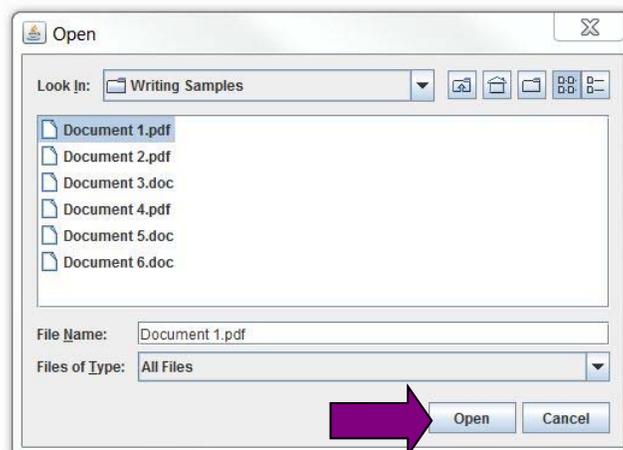
The document now appears in the **Unknown Authors** box (blue arrow).
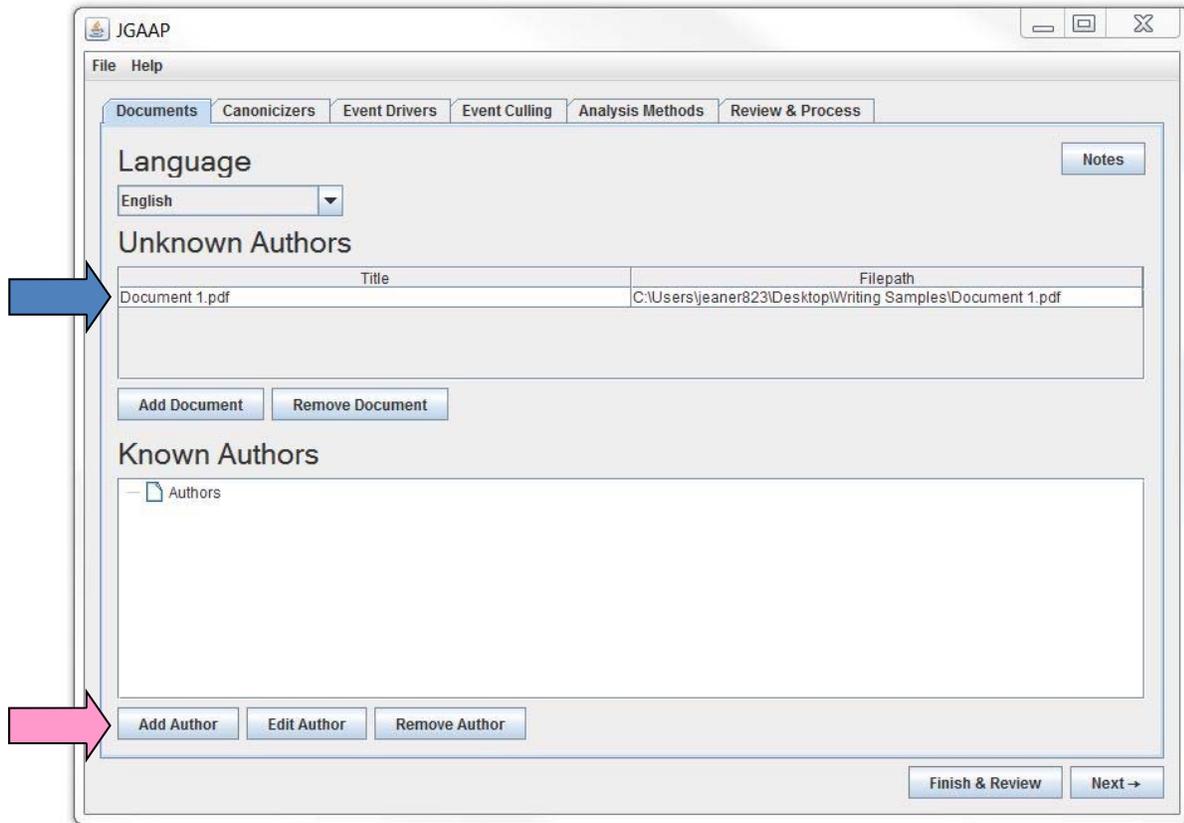
JGAAP will always pick one of the known authors as the most likely author of any document. This means that you need at least two known authors for a meaningful analysis.

To add documents to the **Known Authors** box, click on "Add Author" (pink arrow). A window will open (see picture below). Click on "Add Document" (green arrow).
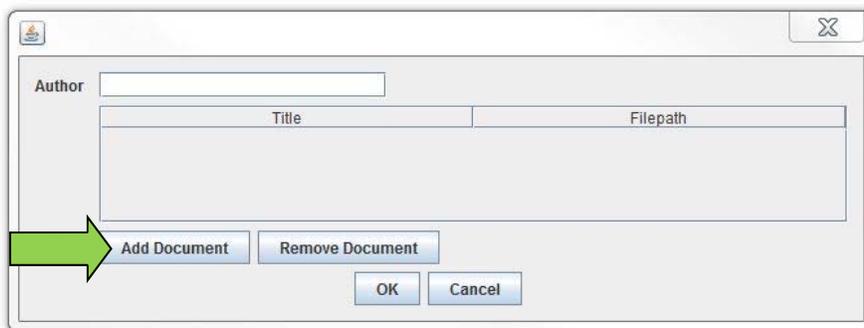
Another window will open where you may select your desired documents (see example below). After making your selections, click on "Open" (green arrow).
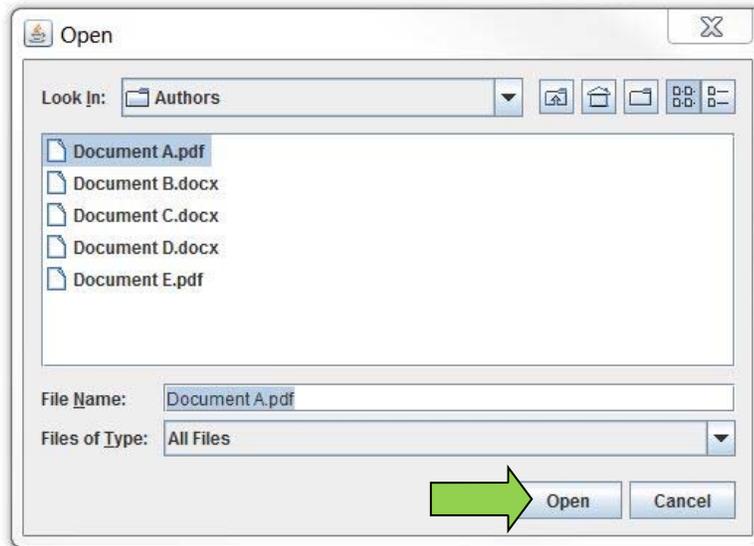
Next you will type in the name of the author for the selected document(s) (pink arrow). When you are finished, click on "OK" (purple arrow).
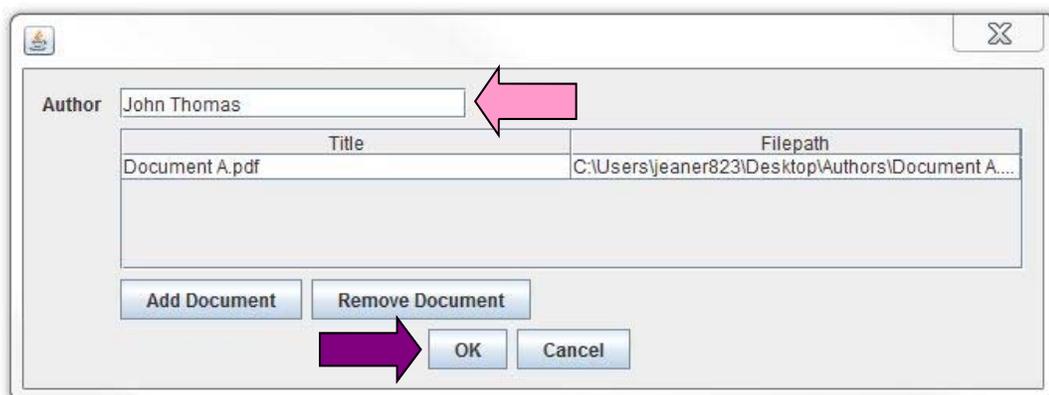
The document now appears in the **Known Authors** box (blue arrow).
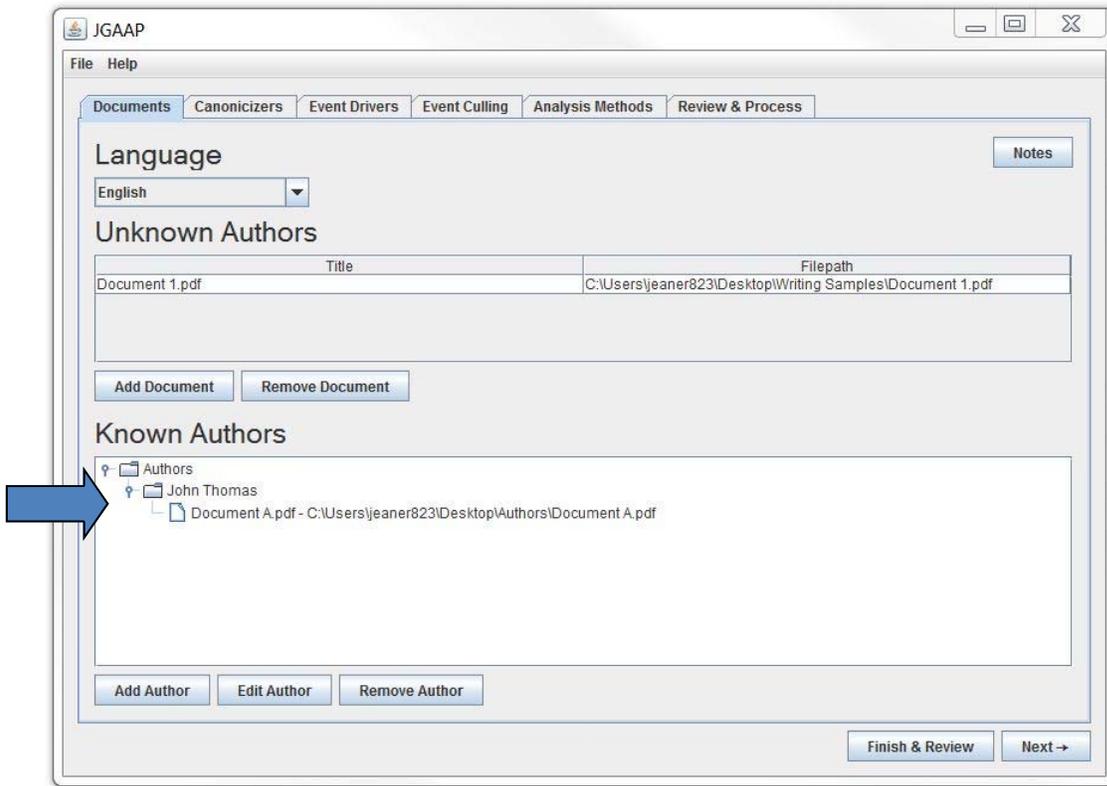
See documents added in example below. When you have finished loading all of your documents, click on "Next" (purple arrow) to move on to the **Canonicizers** tab.
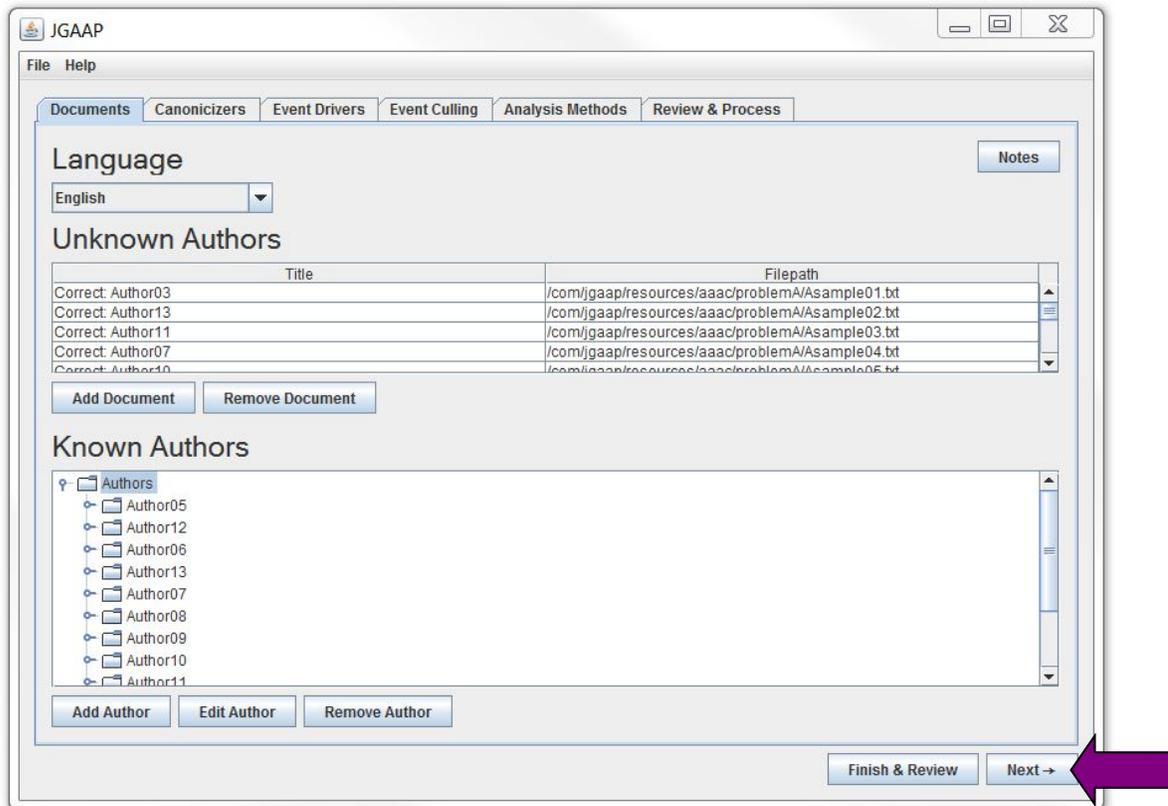
On the **Canonicizers** tab (pictured below), you may select your desired canonicization techniques . These techniques will be applied to each document in order to standardize the documents before features are extracted from them. *Please Note: Canonicizers are optional so you do not have to make a selection unless desired.*
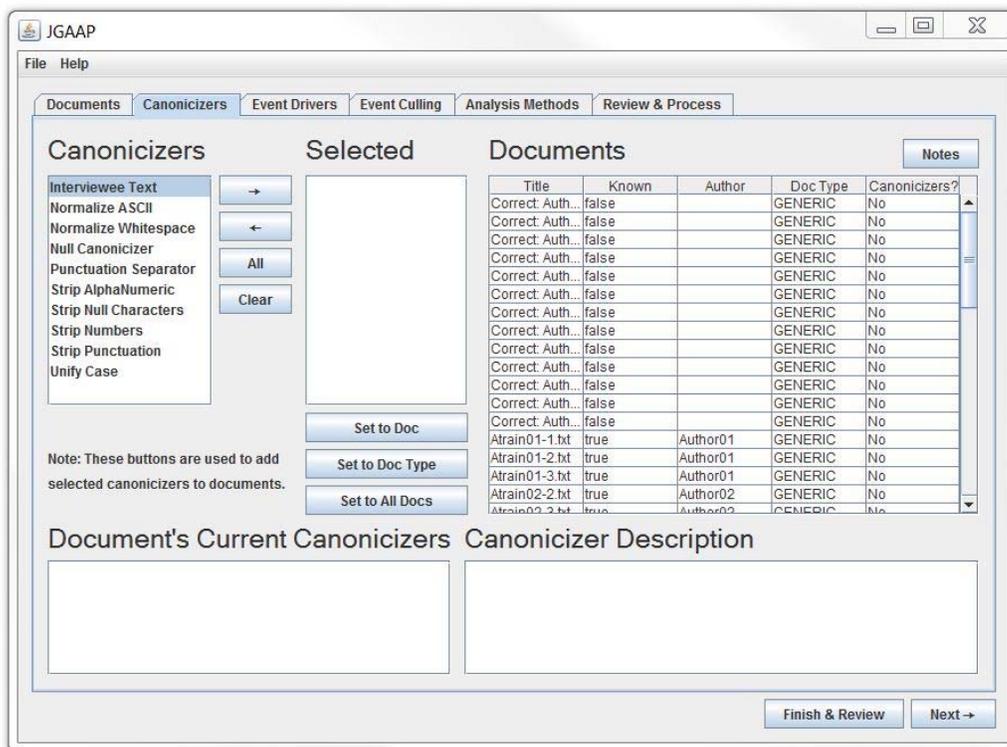
Selecting a canonicizer (brown arrow) will display its corresponding description (green arrow). When choosing canonicizers, you have three options: you may set a canonicizer to one or more specific documents (pink arrow), or to a specific document type (i.e. .pdf, .doc, etc.) (gray arrow), or you may choose to set it to all of the documents in the list (red arrow). Once you have applied a canonicizer to a document, you will notice that the Canonicizer column changes to "Yes" for that document (blue arrow).
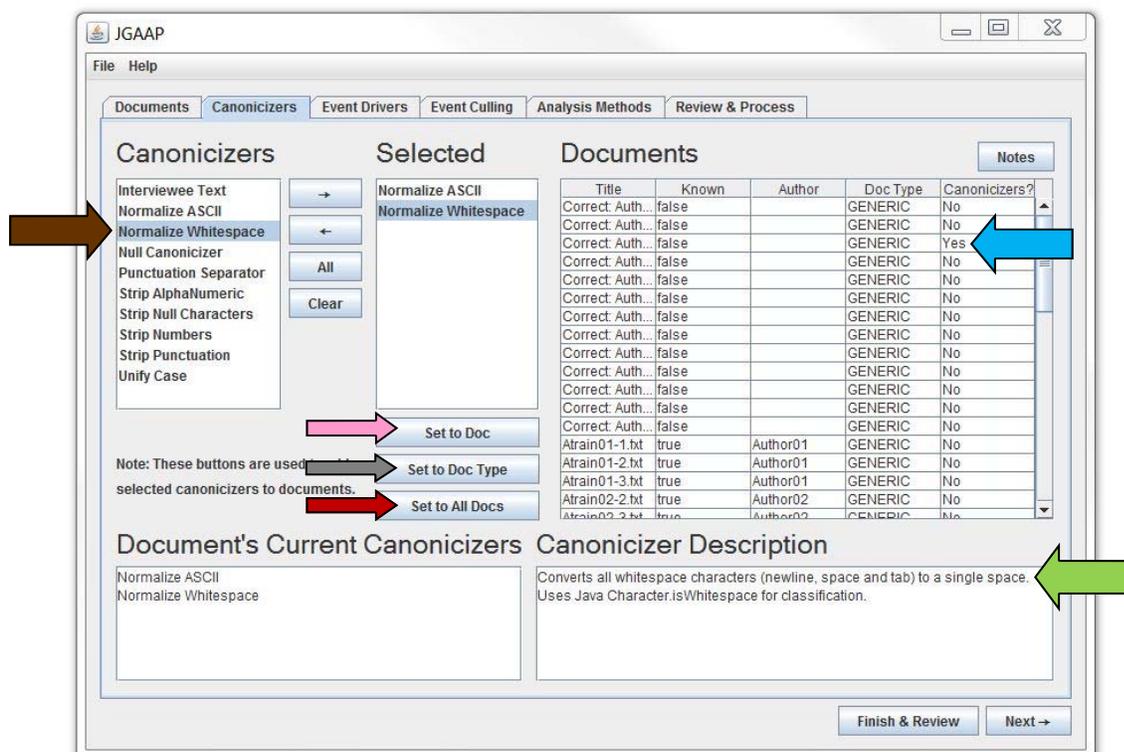
To remove a canonicizer that you have already selected, click on the back arrow (see blue arrow in example below).   You may also click on "Clear" to remove all of your selections (pink arrow). Remember to click on "Set to All Docs" to apply your changes.
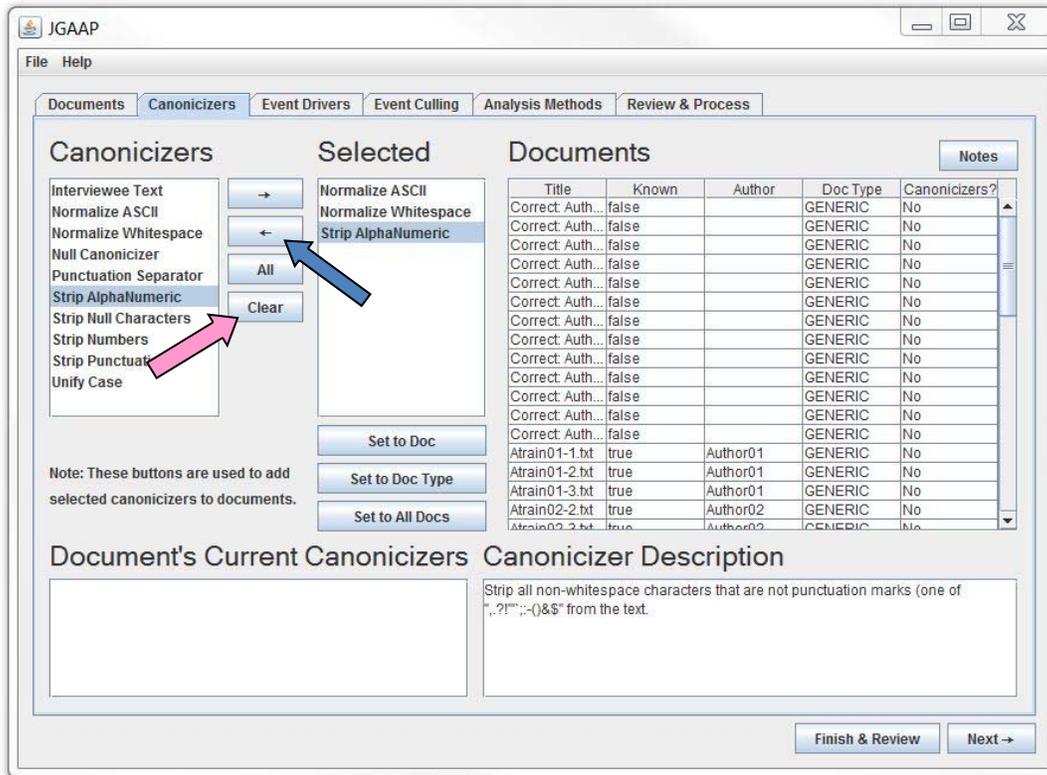
When you have finished applying canonicizers to your list of documents, click on "Next" (purple arrow) to continue to the **Event Drivers** tab.
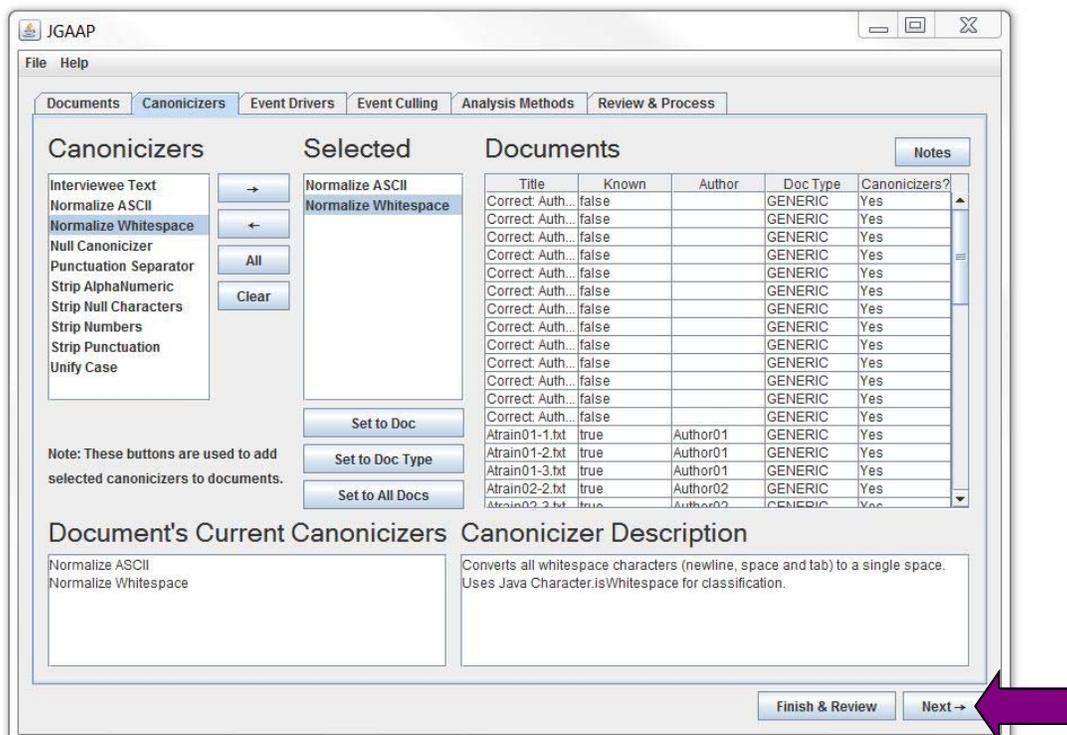
On the **Event Drivers** tab (pictured below), you may select your desired event drivers. The features extracted from each document will depend on the event drivers chosen. Selecting a event driver (brown arrow) will display its corresponding description (green arrow).
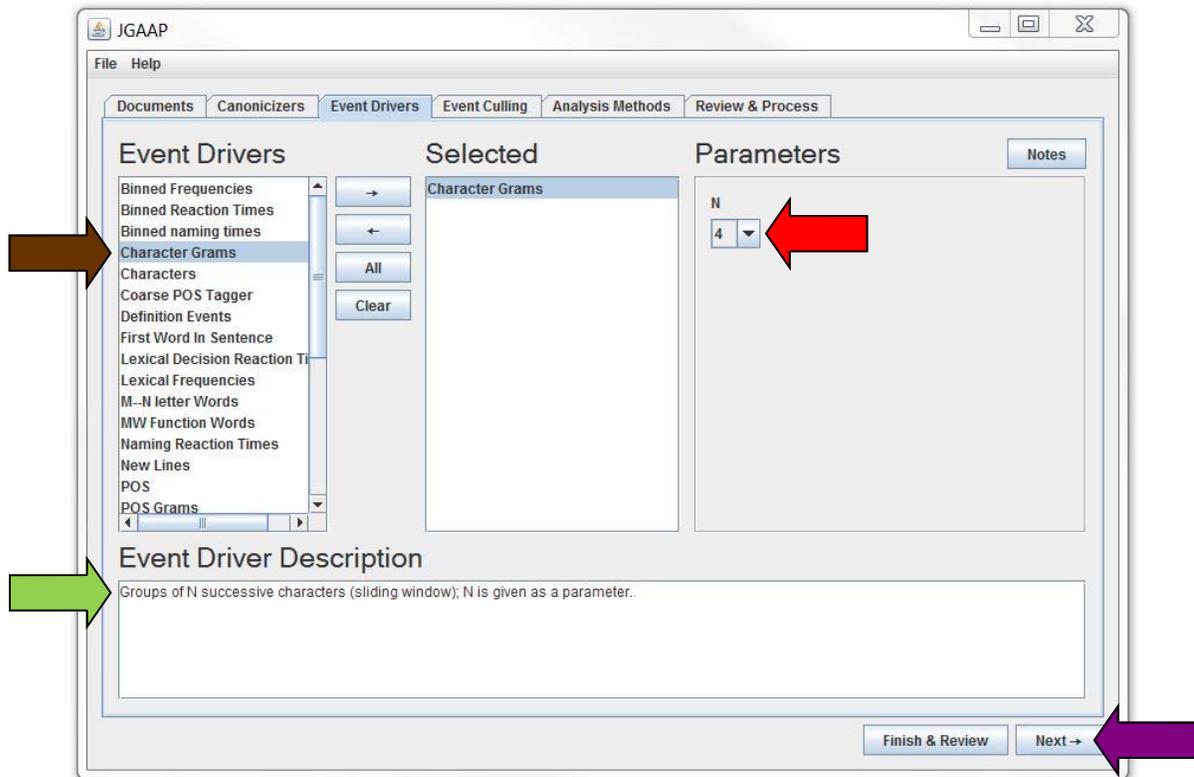
Some event drivers may require specific parameters, such as "Character Grams" in which you must choose a number for "N" from a drop-down menu. To set the Parameters, click on the drop-down box (red arrow) to change "N" to the desired parameter. When you have finished selecting event drivers (and parameters if applicable), click on "Next" (purple arrow) to continue to the **Event Culling** tab.

On the **Event Culling** tab (pictured below), you may choose whether you want to analyze only the most frequent events (green arrow), only the least common events (blue arrow), or only those events which appear in all of the documents (red arrow). If you select "Most Common Events," then any uncommon events will be eliminated from the documents prior to analysis.

For "Least Common Events" and "Most Common Events" you may set the parameters ("N") to any number of your choice. To set the Parameters, click on the drop-down box (pink arrow) to change "N" to the desired parameter by either typing the number or choosing a value from the drop-down menu. You do not have to set Parameters for the "X-treme Culler."

Please Note: Event culling order matters if you choose more than one event culler.
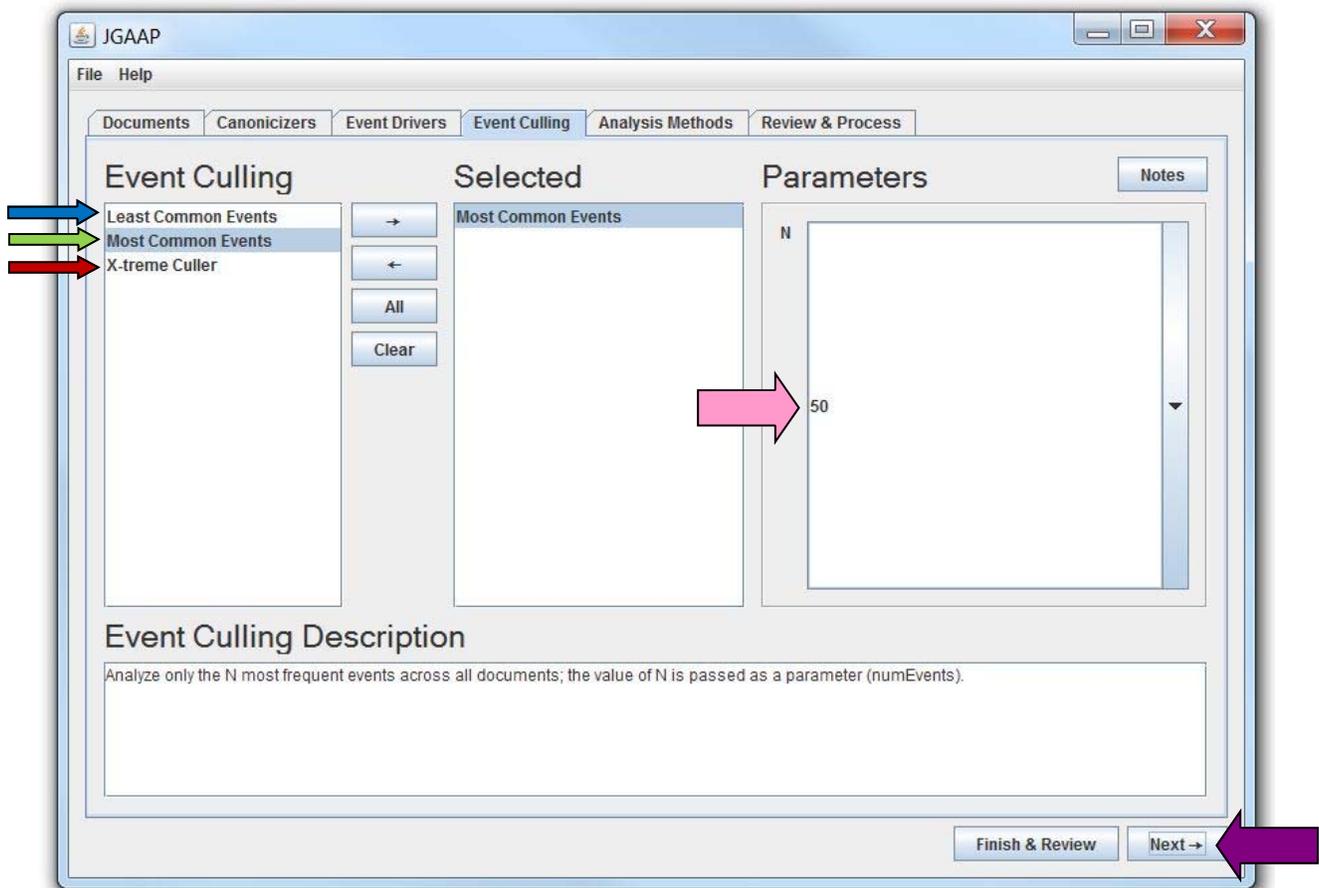
Please Note: Event cullers are optional so you do not have to make a selection unless desired.

Once you have made your event culler selection(s) and are ready to move on, click on "Next" (purple arrow) to continue to the **Analysis Methods** tab.

On the **Analysis Methods** tab (pictured below), you will select an analysis method to analyze your documents.

*Please Note: If you choose the **Frequency Centroid Driver** or **Nearest Neighbor Driver**, you will have to select a **Distance Function**. You do not have to choose a distance function for the other analysis methods.*

In the example below, "Nearest Neighbor Driver" (blue arrow) is selected with "Cosine Distance" as the Distance Function (red arrow).
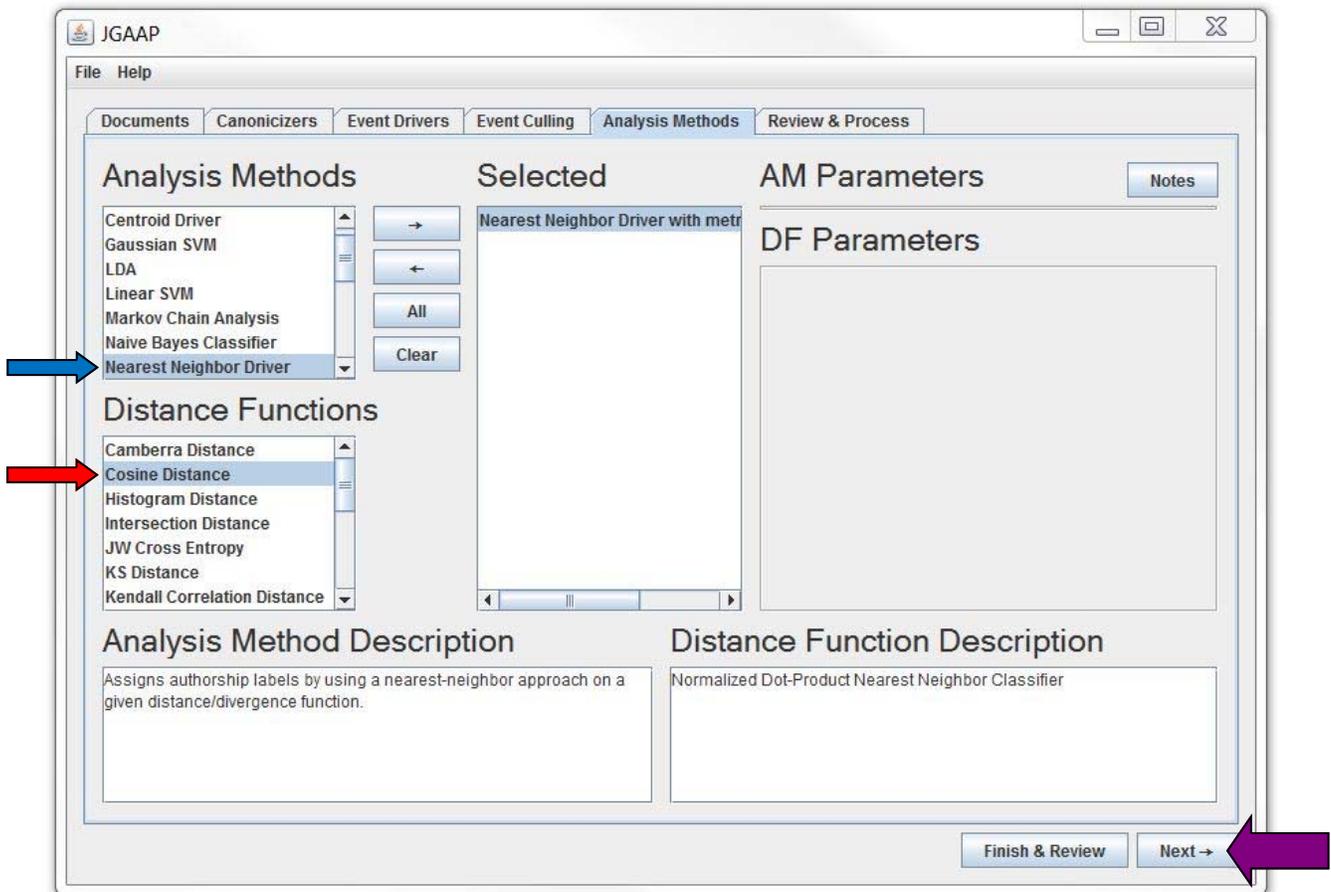
After choosing an analysis method, click on "Next" (purple arrow) to continue to the **Review & Process** tab.

On the **Review & Process** tab (pictured below), you may review all of the selections you've made from the previous tabs. If you'd like to make any changes, you may click on the tab you wish to make changes to (blue arrow), or you may click on the green Headings (brown arrow) which will take you back to the corresponding tab.
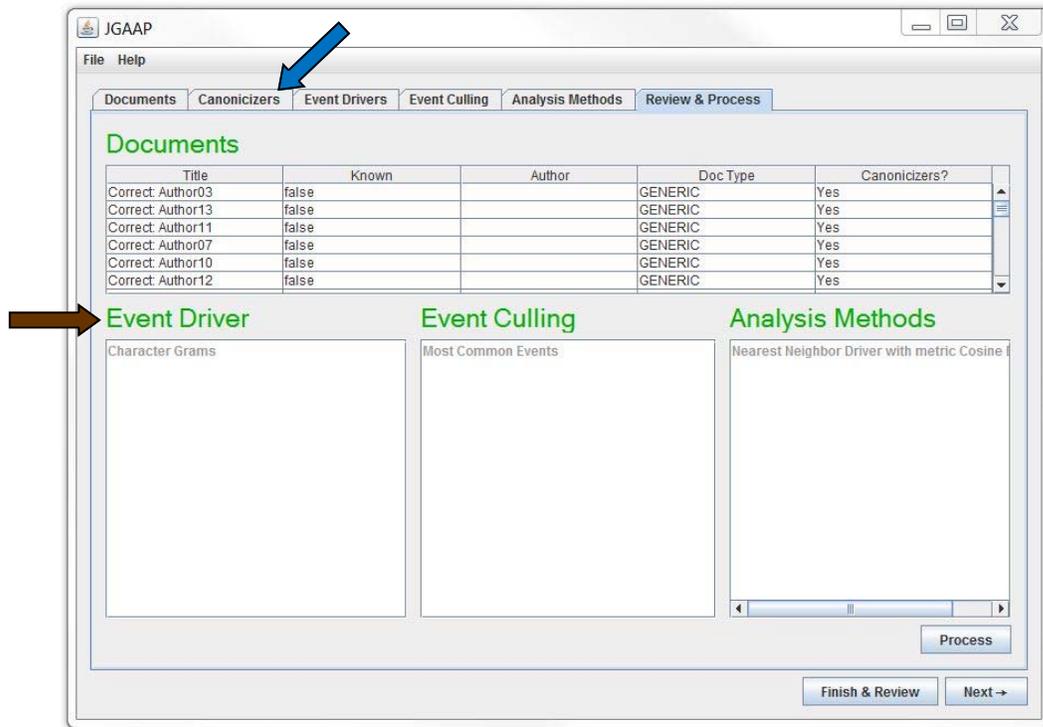
If a mistake has been made on any of the tabs, the heading will be red instead of green to draw your attention to the error(s). In the example below, the **Event Driver** heading is red because no Event Drivers were selected (red arrow).
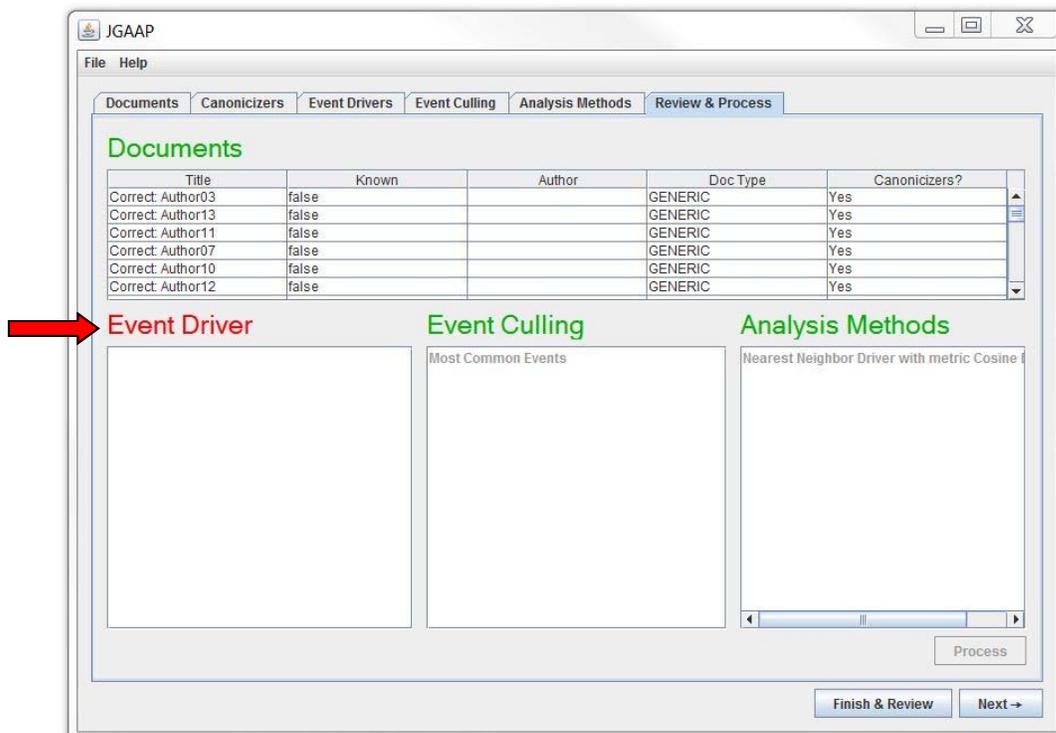
Once you have verified that all of the selections are correct and you are ready to obtain your results, click on "Process" (purple arrow).
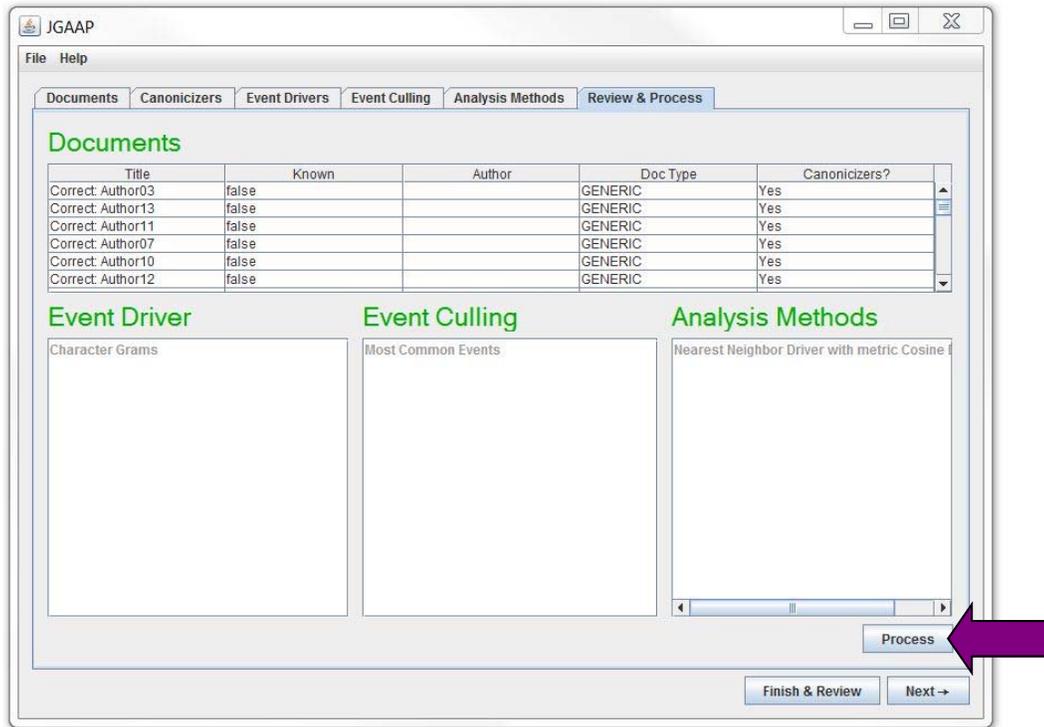
A new window will open with the results. In this case, the program determined that Author10 is the most likely author for the document in question (red arrow).