

Daily News Analysis using GeoWorlds 3.0

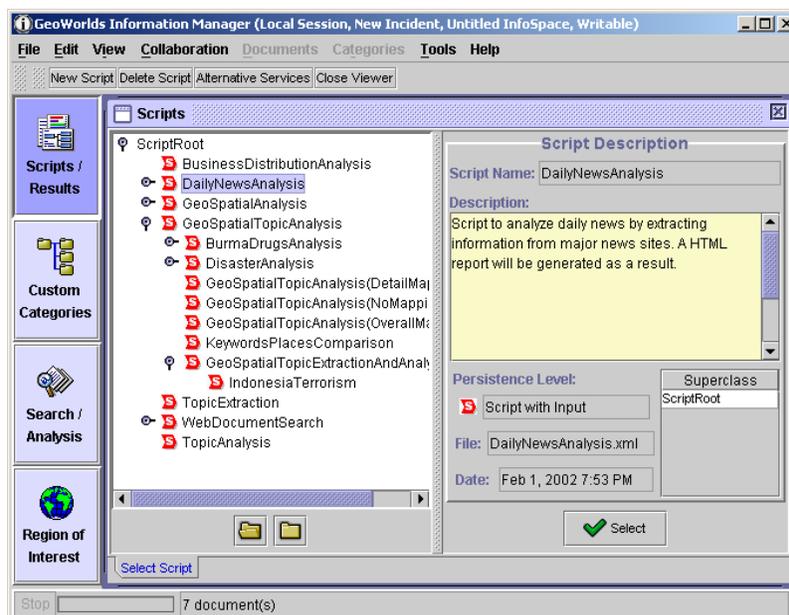
The *daily news analysis* script in GeoWorlds is an information management script that performs a geo-spatial topic analysis on news articles. It can be personalized to monitor and analyze news articles for topics and places of interest. The results of the analysis can be view through an interactive Java interface or through a Web browser. This script is a simplified version of a set of scripts that is used to generate the GeoTopics Web site (<http://www.isi.edu/geoworlds/geotopics/>).

Specifically, this script collects headline news articles from the news sites specified by the user (default input includes 10 major English-language news sites), classifies the articles based on user specified topics and place names, ranks the classifications based on frequency, cross products the two classifications (topics by places, and places by topics), and extracts new topic names (noun phrases) from the articles. As a result of executing the script, GeoWorlds displays tree viewers that show the document classifications and cross-products, histograms that show the ranked classifications and newly discovered topic names, 3D bar charts that visualize the relationships between the two categorizations, a map viewer that shows the geographical distribution of the major topics, and a HTML report that summarizes all the results.

The followings are the steps to run the script:

1. Select 'DailyNewsAnalysis' script from the script selection window

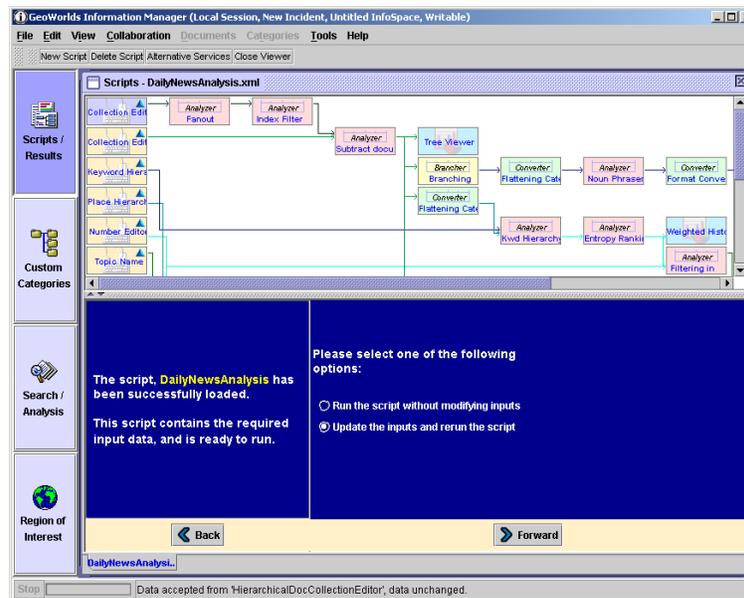
- Click on the script name, 'DailyNewsAnalysis' on the left side
- Press 'Select'  button at the bottom



2. Enter input data

- a. After the script is loaded, GeoWorlds displays a flow diagram of the script in the upper part and operation instructions in the lower part of the window. Before running the script, all the required input data have to be entered. To enter input data, select the radio button, 'Update the input and rerun the scrip' and press 'Forward' button.

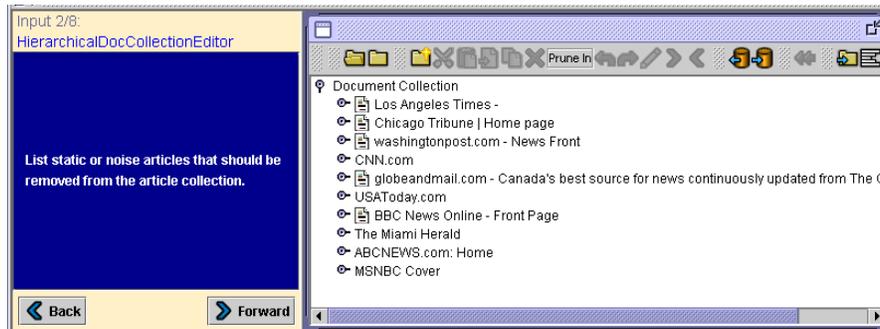
(Choosing the 'Run the script without modifying inputs' option directs GeoWorlds to run the script by using the default input values. This is not recommended for this script, because it overwrites results from previous runs and it does not allow for personalization of the script. If you select this option, skip to step 3).



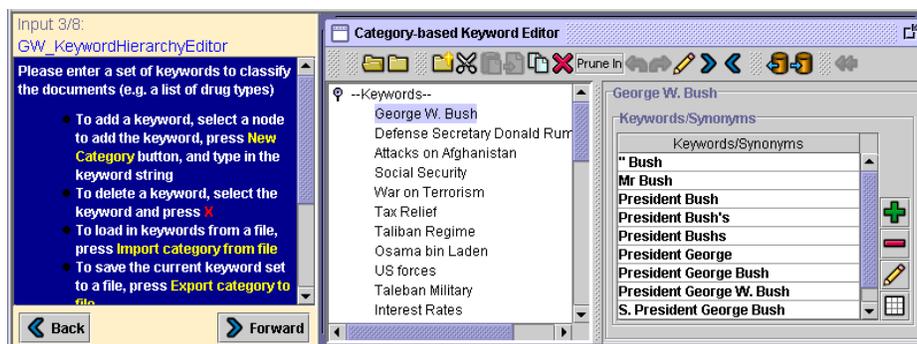
- b. Organize the list of news sites to collect articles. New news site can be added by connecting the home page of the news site by using your default Web browser, selecting the root, 'Document Collections' of the list, and pressing 'Import URL' button (The button may not be visible if the size of the GeoWorlds window is too small. In such case, please enlarge the GeoWorlds window). Press 'Forward' button when you finish editing the list.



- c. Edit the list of noise pages that are filtered out from the document collection. The noise pages are static pages such as the pages that include news company profile and contact information, static-URL pages that show daily weather and stock information, and commercial pages. The script already includes a set of noise pages for each pre-entered news site. To add more noise pages, select the corresponding news site node (create a new news-site node if it is needed) and import the noise page by using the same method explained at the previous step. Press 'Forward' button when you finish editing the list.



- d. Edit the topic list. The words and phrases in the topic list are used to classify the news articles. *Sample topics are included in the script for testing. You'll need to remove these default topics and organize a task-specific topic list for your interest or information management task.* Synonyms of a topic name can be defined. To add a synonym, select the topic on the left-side tree, press 'Add'  button on the right side, type in the synonym, and press Enter key. To remove a synonym, select the synonym to remove and press 'Remove'  button. An organized topic list can be saved to a file so that it can be reused for future analyses. To save a topic list, press 'Export'  button, enter a file name to the file chooser displayed, and click on 'Save' button. To insert in an existing topic file, select the location in the topic list to insert (such as the '--Keywords--' node), press 'Import'  button, select a topic file to import at the file chooser displayed, and press 'Read' button. Press 'Forward' button when you finish editing the list.



- e. Edit the place name list. The news articles will be classified based on these place names. *The default places included in the script are the countries and major cities in the world.* Alternative place name spellings can be defined by following the steps explained Step 1d. Place names that have been exported from the ArcView map can be imported by pressing 'Load Place Names'  button. A place name list can be saved or loaded to/from a file by following the steps explained in Step 1d. Press 'Forward' button when you finish editing the list.



- f. Enter the maximum number of topics and place names to be use for the cross-product and displayed in the HTML report. Recommended value is 20, which means that the script uses the top 20 topics and places for generating the cross-products and the HTML report.



- g. Enter the title of the HTML report. This title is displayed in the report.



- h. Enter the name of the target directory to save the HTML report data. To avoid overwriting of previous data, please use a unique directory name. e.g., news_analysis_20020201. The directory will be created under GeoWorlds' information space directory ('C:\GeoWorlds\Client\InfoSpaces' if you installed GeoWorlds under 'C:\GeoWorlds'). If you enter an absolute path name such as 'C:\MyInfoSpaces\news_analysis_20020201', GeoWorlds will create the target directory that you specified.

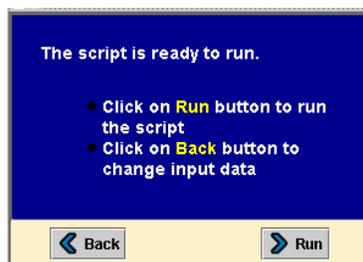


- i. Enter the name of the region for this analysis. This name is used to create files for storing the HTML report data. If the place names entered at the earlier step are in a specific region, please enter the name of that region, e.g., Indonesia.

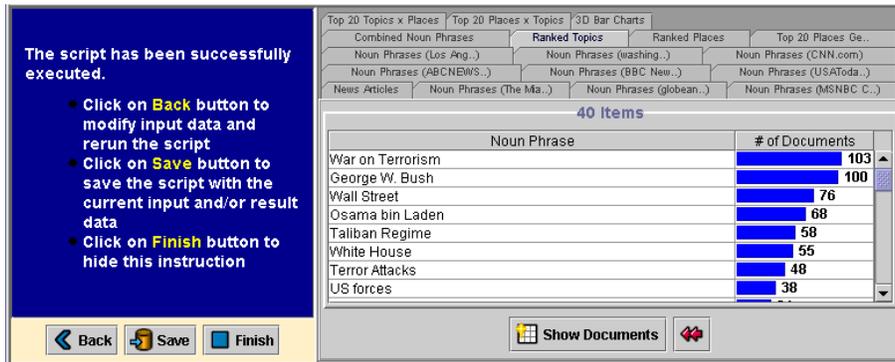


3. Run the script

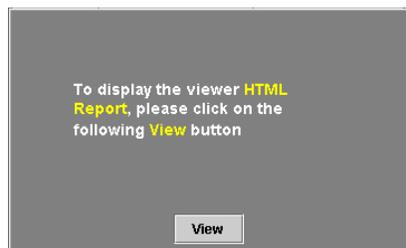
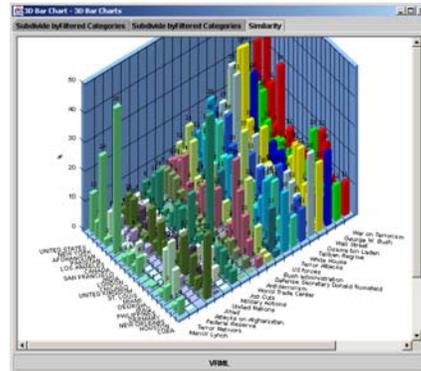
- Now, the script is ready to run. To run the script, press 'Run' button. If you want to modify input data, press 'Back' button to go back to the input item to be modified. During run-time, GeoWorlds shows the status messages and progress at the lower-left panel.



- The results of the script execution are displayed in the lower-right panel (in separate tabs for each result). Press the 'Finish' button at the bottom-left part to expand the result panel.



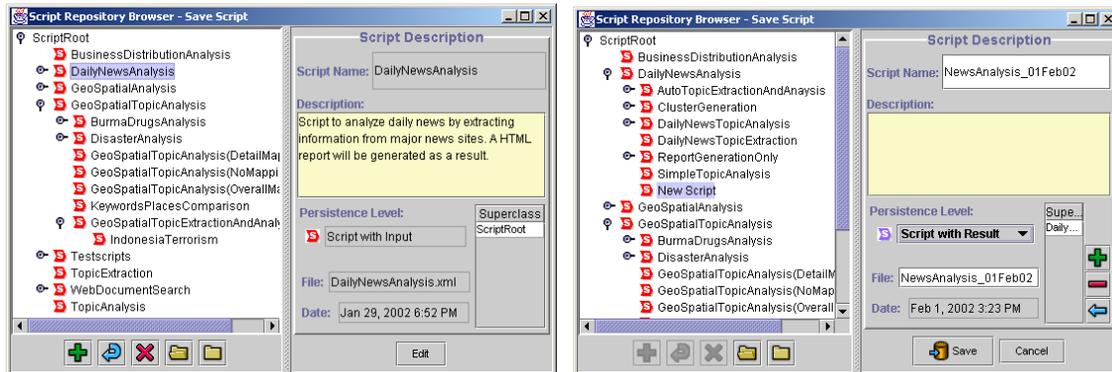
- The results that require special viewer software are shown separately in external viewer frames. In this script, the 3D bar charts, document mappings on the ArcView map and the HTML report are results that require separate viewers. For these results, the corresponding tabs show only the description about the result and a 'View' button. Press this button to redisplay the result in the external viewer.



- To modify any of the input data and rerun the script, press 'Back' button. You can come back to the result mode at any time by double-clicking on any of the viewer nodes in the flow-diagram. Viewer nodes can be identified by a background image that shows a bar chart and a pie chart as such as:

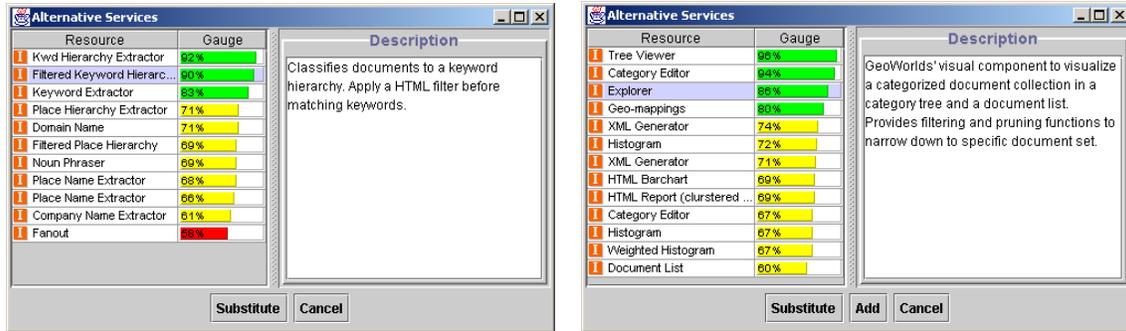


- To save the script to the script repository, press 'Save' button. Then the script repository browser is displayed. To save the script along with the result, select a category to place the scrip, press 'Add'  button, enter the name of the script and a text description, and select 'Script with results' item in the 'Persistence Level' combo box, and press 'Save' button (please hit Enter key whenever you finish editing the script name. Then, the corresponding file names are automatically updated).



4. Reconfigure the script

- The script can be reconfigured to display results in alternative viewers, and to use different analysis services. To get a list of alternative services, click on a service node (then, a star  is displayed on the node), and press 'Alternative Services' button at the tool bar. If input services are currently connected to the node, GeoWorlds asks whether you want to use the current input services or add a new set of input services. To use the current input service, press 'Yes' button. Then, a window that shows the list of alternative services is displayed. For example, the followings are the lists of alternative services for the 'Keyword Hierarchy Extraction' and 'Tree Viewer' for one of the cross-product viewers in this script:



- To replace the current service with an alternative one, select one of the alternative services, and press 'Substitute' button. For a viewer, you can also add a new viewer service to show the result in an alternative viewer. To add a new viewer, press 'Add' button at the service list.

If you replace a viewer or add a new viewer at the result mode, the new viewer displays the result immediately. When an analysis service is replaced with an alternative one, the script should be rerun to generate a new set of results.