Import Documents

Import text documents from folders.

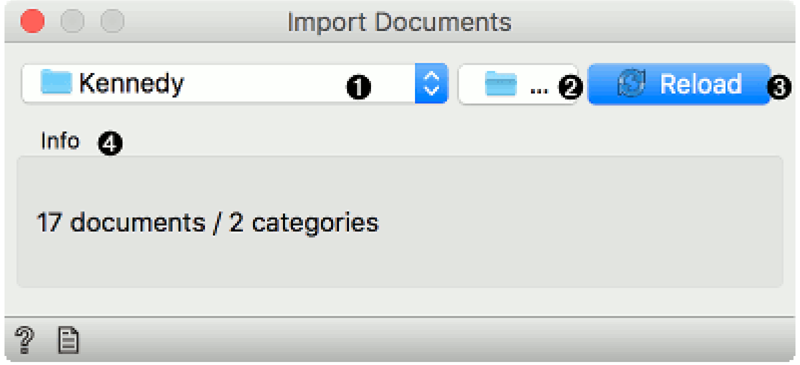
Inputs

- None

Outputs

- Corpus: A collection of documents from the local machine.

Import Documents widget retrieves text files from folders and creates a corpus. The widget reads .txt, .docx, .odt, .pdf and .xml files. If a folder contains subfolders, they will be used as class labels.



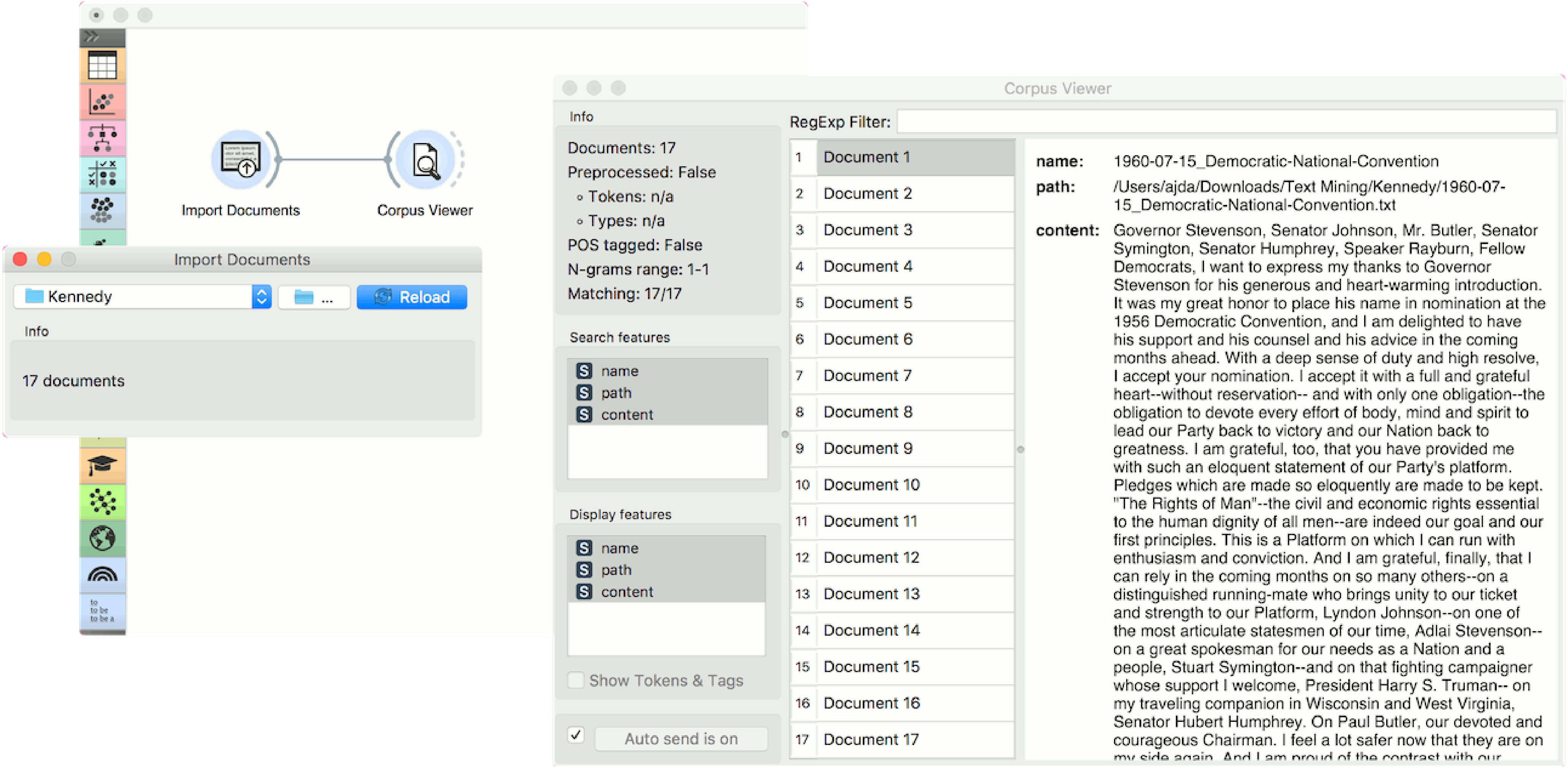
17 documents / 2 categories

1. Folder being loaded.
2. Load folder from a local machine.
3. Reload the data.
4. Number of documents retrieved.

If the widget cannot read the file for some reason, the file will be skipped. Files that were successfully retrieved will still be on the output.

Example

To retrieve the data, select the folder icon on the right side of the widget. Select the folder you wish to turn into corpus. Once the loading is finished, you will see how many documents the widget retrieved. To inspect them, connect the widget to Corpus Viewer. We've used a set of Kennedy's speeches in a plain text format.



Import Documents

Kennedy ... Reload

Info

17 documents

Corpus Viewer

Info

Documents: 17
Preprocessed: False
∘ Tokens: n/a
∘ Types: n/a
POS tagged: False
N-grams range: 1-1
Matching: 17/17

Search features

S name
S path
S content

Display features

S name
S path
S content

□ Show Tokens & Tags

☑ Auto send is on

RegExp Filter:

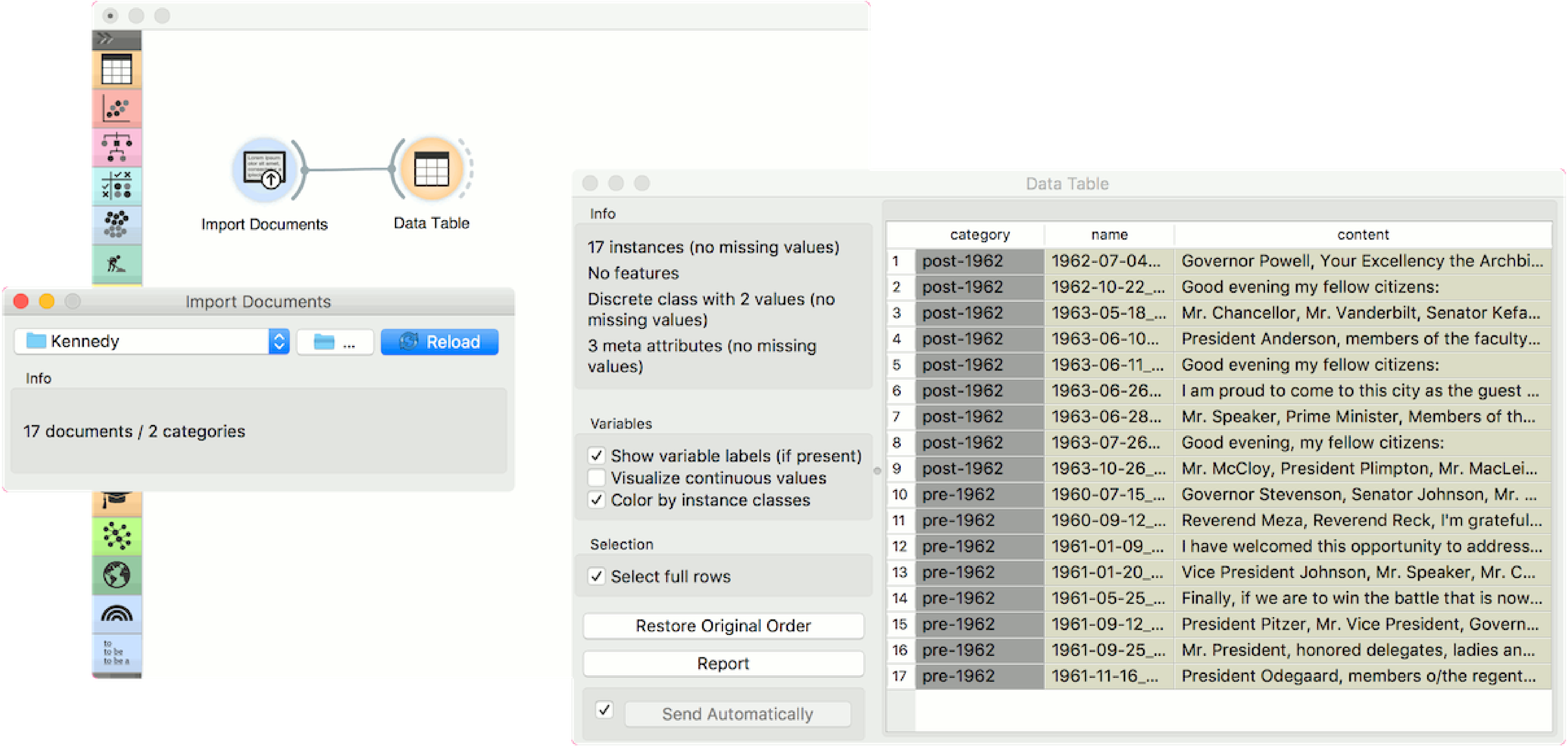
1 Document 1
2 Document 2
3 Document 3
4 Document 4
5 Document 5
6 Document 6
7 Document 7
8 Document 8
9 Document 9
10 Document 10
11 Document 11
12 Document 12
13 Document 13
14 Document 14
15 Document 15
16 Document 16
17 Document 17

name: 1960-07-15_Democratic-National-Convention

path: /Users/ajda/Downloads/Text Mining/Kennedy/1960-07-15_Democratic-National-Convention.txt

content: Governor Stevenson, Senator Johnson, Mr. Butler, Senator Symington, Senator Humphrey, Speaker Rayburn, Fellow Democrats, I want to express my thanks to Governor Stevenson for his generous and heart-warming introduction. It was my great honor to place his name in nomination at the 1956 Democratic Convention, and I am delighted to have his support and his counsel and his advice in the coming months ahead. With a deep sense of duty and high resolve, I accept your nomination. I accept it with a full and grateful heart--without reservation-- and with only one obligation--the obligation to devote every effort of body, mind and spirit to lead our Party back to victory and our Nation back to greatness. I am grateful, too, that you have provided me with such an eloquent statement of our Party's platform. Pledges which are made so eloquently are made to be kept. "The Rights of Man"--the civil and economic rights essential to the human dignity of all men--are indeed our goal and our first principles. This is a Platform on which I can run with enthusiasm and conviction. And I am grateful, finally, that I can rely in the coming months on so many others--on a distinguished running-mate who brings unity to our ticket and strength to our Platform, Lyndon Johnson--on one of the most articulate statesmen of our time, Adlai Stevenson--on a great spokesman for our needs as a Nation and a people, Stuart Symington--and on that fighting campaigner whose support I welcome, President Harry S. Truman-- on my traveling companion in Wisconsin and West Virginia, Senator Hubert Humphrey. On Paul Butler, our devoted and courageous Chairman. I feel a lot safer now that they are on my side again. And I am proud of the contrast with our

Now let us try it with subfolders. We have placed Kennedy's speeches in two folders - pre-1962 and post-1962. If I load the parent folder, these two subfolders will be used as class labels. Check the output of the widget in a **Data Table**.



Import Documents

Kennedy ... Reload

Info

17 documents / 2 categories

Data Table

Info

17 instances (no missing values)

No features

Discrete class with 2 values (no missing values)

3 meta attributes (no missing values)

Variables

☑ Show variable labels (if present)

□ Visualize continuous values

☑ Color by instance classes

Selection

☑ Select full rows

Restore Original Order

Report

☑ Send Automatically

	category	name	content
1	post-1962	1962-07-04...	Governor Powell, Your Excellency the Archbi...
2	post-1962	1962-10-22_...	Good evening my fellow citizens:
3	post-1962	1963-05-18_...	Mr. Chancellor, Mr. Vanderbilt, Senator Kefa...
4	post-1962	1963-06-10...	President Anderson, members of the faculty...
5	post-1962	1963-06-11_...	Good evening my fellow citizens:
6	post-1962	1963-06-26...	I am proud to come to this city as the guest ...
7	post-1962	1963-06-28...	Mr. Speaker, Prime Minister, Members of th...
8	post-1962	1963-07-26...	Good evening, my fellow citizens:
9	post-1962	1963-10-26_...	Mr. McCloy, President Plimpton, Mr. MacLei...
10	pre-1962	1960-07-15_...	Governor Stevenson, Senator Johnson, Mr. ...
11	pre-1962	1960-09-12_...	Reverend Meza, Reverend Reck, I'm grateful...
12	pre-1962	1961-01-09_...	I have welcomed this opportunity to address...
13	pre-1962	1961-01-20_...	Vice President Johnson, Mr. Speaker, Mr. C...
14	pre-1962	1961-05-25_...	Finally, if we are to win the battle that is now...
15	pre-1962	1961-09-12_...	President Pitzer, Mr. Vice President, Govern...
16	pre-1962	1961-09-25_...	Mr. President, honored delegates, ladies an...
17	pre-1962	1961-11-16_...	President Odegaard, members o/the regent...