Orange Textable Prototypes v0.1.7 documentation

Release 0.1.7

2016 University of Lausanne

Contents

Welcome to the documentation of Orange Textable Prototypes.

Orange Textable Prototypes is an open-source add-on for Orange Canvas. It collects text-analytic widgets that couldn't be included in the core Orange Textable distribution for various reasons (notably because they introduce dependencies to third-party packages, or simply because they're not yet production-ready), but that may still be found useful by Orange Canvas/Textable users.

The project is brought to the community by the department of language and information sciences (SLI) at the University of Lausanne It is hosted at https://github.com/axanthos/TextablePrototypes, and the documentation can be found at http://textable-prototypes.readthedocs.org/.

Widgets

Theatre Classique

Import theater plays from the theatre-classique website (in TEI-XML).

Author

Aris Xanthos

Signals

Inputs: None

Outputs:

- `Text data`

    Segmentation covering the content of imported TEI-XML-encoded theatre plays

Description

This widget is designed to import one or more theatre plays in Orange Canvas. The plays are retrieved from http://www.theatre-classique.fr and richly encoded in TEI-XML format. The output is a segmentation containing a segment for each imported play. Each segment has 5 annotations with keys *author*, *title*, *year*, *genre*, and *url*.

The interface of **Theatre Classique** is available in two versions, according to whether or not the **Advanced Settings** checkbox is selected.

Basic interface

In its basic version (see *figure 1* below), the **Theatre Classique** widget simply lets the user select one or more plays in the catalogue of about 800 entries downloadable from the theatre-classique website. To select multiple files use either control/command-click or shift-click.
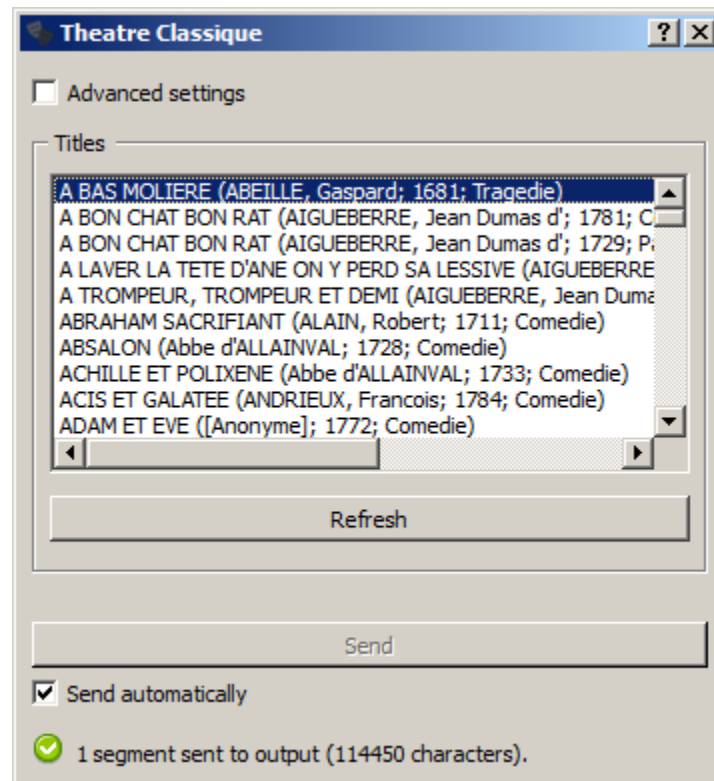

Fig. 1.1: Figure 1: **Theatre Classique** widget (basic interface).

The **Options** section allows the user to define the label of the output segmentation (**Output segmentation label**).

The **Info** section indicates the number of segments and characters in the output segmentation, or the reasons why no segmentation is emitted (no title selected, connection issues, etc.).

The **Send** button triggers the emission of a segmentation to the output connection(s). When it is selected, the **Send automatically** checkbox disables the button and the widget attempts to automatically emit a segmentation at every modification of its interface.

Advanced interface

The advanced version of **Theatre Classique** (see *figure 2* below)offers the same functionality as the basic one, and it adds the possibility of selecting only the plays of a given author/genre/title.

The **Options** and **Info** sections, as well as the **Send** button and **Send automatically**, operate in the same way as in the basic interface.
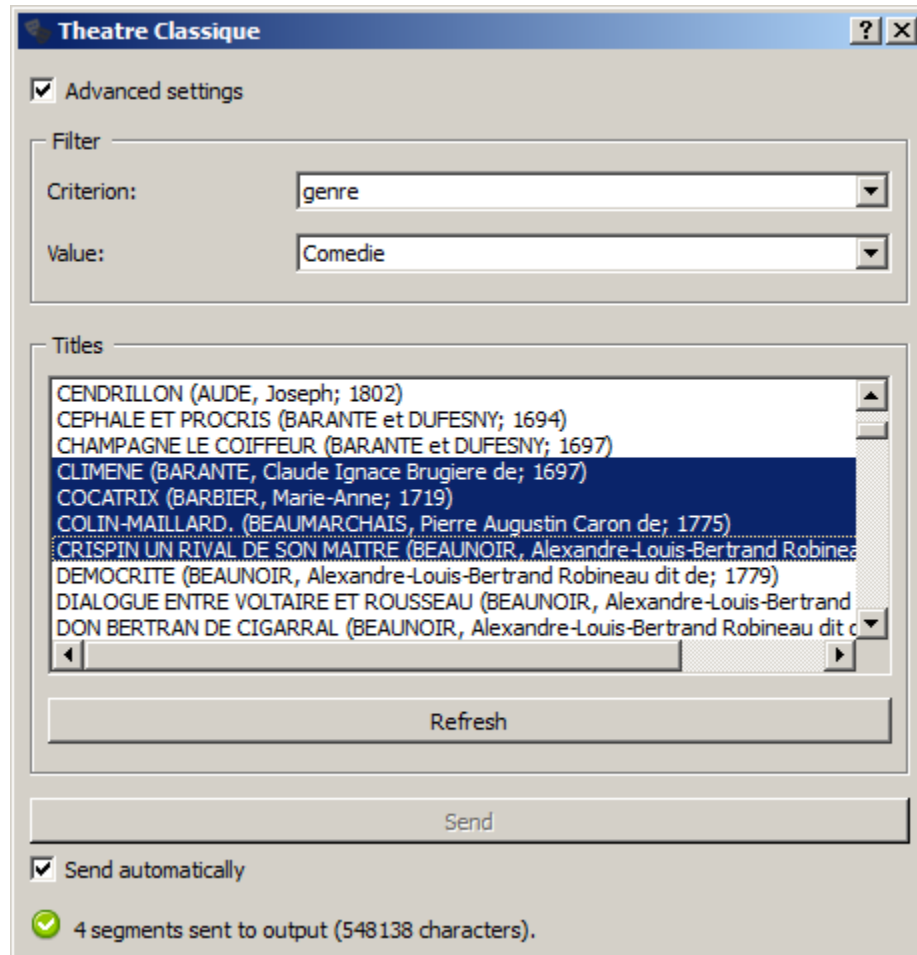
Fig. 1.2: Figure 2: **Theatre Classique** widget (advanced interface).

Messages

Information

<n> segments sent to output (<m> characters). This confirms that the widget has operated properly.

Warnings

Settings were changed, please click 'Send' when ready. Settings have changed but the **Send automatically** check-box has not been selected, so the user is prompted to click the **Send** button (or equivalently check the box) in order for computation and data emission to proceed.

Please select one or more titles. The widget instance is not able to emit data to output because no theatre play has been selected.

Errors

Couldn't download data from theatre-classique website. An error has prevented the widget to download the data from the theatre-classique (most likely related to a connection problem).

Web Search

Import Twitter, Wikipedia and Bing data.

Authors

Bassim Matar, Gregory Thonney, Cyril Nghiem, Jean Galleno, Taar Rusconi

Signals

Inputs : None

Outputs : `Text data`

>   Segmentation containing the result of a query on a web service

Description

The Web Search widget is designed for the Orange Canvas environment to generate textual data retrieved from Twitter, Wikipedia or Bing. Depending on the service, the output segmentation has the following annotations with keys :

- Twitter : date, source, search, url and author.
- Wikipedia : source and search.
- Bing : url, source, search and title.

The interface of Web Search adapts itself according to the selected service.

Interface

As stated before, the interface is dependent of the selected service. In case Twitter is chosen, the interface looks like this :
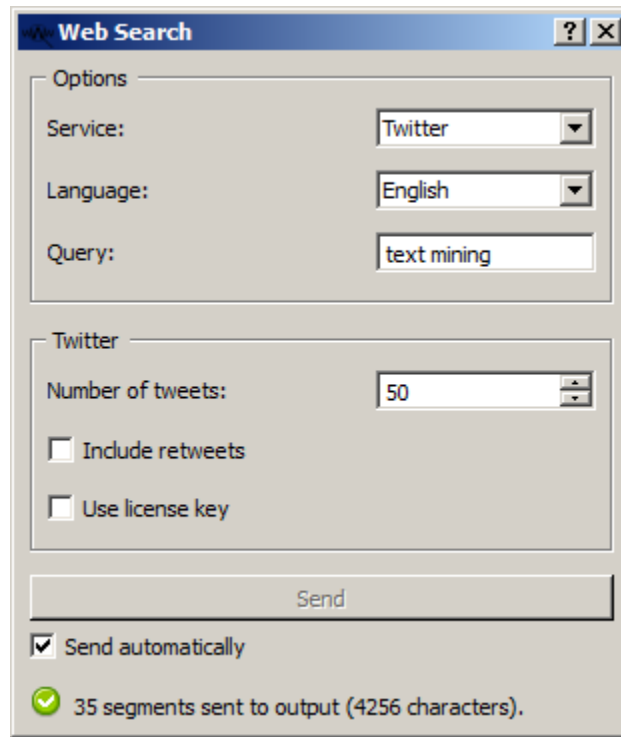

Fig. 1.3: *Figure 1 : Web Search widget with Twitter selected*

The **Service** option allows the user to select a search engine (Twitter, Wikipedia or Bing).

The **Language** option enables the user to choose the language of the retrieved data between English, French, German, Spanish, Italian and Dutch (by default : English).

The **Query** field contains the searched word(s). By default, the language is set to English.

Clicking on the **Send** button executes the request. The **Info** box above indicates the number of segments sent (in case any matches the request). For more informations about the **Info** box : read *Messages*.

Depending on which search engine is selected, different options appear on the interface. These specific aspects are stated below.

Twitter

The **Number of tweets** field allows users to retrieve up to 3000 tweets (though the actual number of results may be lower).

When **Include retweets** is checked, tweets starting with 'RT' are filtered out in order to only send "original" tweets to the output.

By default, the widget uses a public Twitter license key provided by the underlying Pattern Python library. To use a private license key, tick the **Use license key** checkbox. Four keys (which can be found under "Keys and Access Tokens" in user's account) are needed. Visit http://apps.twitter.com to learn more.

Fig. 1.4: *Figure 2 : Twitter options box with license key parameters*

Wikipedia

Fig. 1.5: *Figure 3 : Wikipedia options box*

When **Segment into sections** is checked, Wikipedia articles are divided into sections : each segment contains a section.

The **Output format** is either "Plain text" or "HTML".

Bing

Fig. 1.6: *Figure 4 : Bing options box*

The **Number of results** field allows users to retrieve up to 1000 Bing results (though the actual number of results may be lower).

Messages

Information

<n> segments sent to output (<m> characters). This confirms that the widget has operated properly.

Warnings

Settings were changed, please click 'Send' when ready. Settings have changed but the **Send automatically** check-box has not been selected, so the user is prompted to click the **Send** button (or equivalently check the box) in order for computation and data emission to proceed.

Please select one or more titles. The widget instance is not able to emit data to output because no theatre play has been selected.

No data sent to output yet. Please try to change query and settings. The query didn't retrieve any data. When confronted to this message, the user should either try to change the query or modify the settings.

Errors

Please enter valid Twitter api keys. The entered values don't match any existing key.