

Wikipedia

Fetching data from [MediaWiki RESTful web service API](#).

Inputs

- None

Outputs

- Corpus: A collection of documents from the Wikipedia.

Wikipedia widget is used to retrieve texts from Wikipedia API and it is useful mostly for teaching and demonstration.

1. Query parameters:
 - Query word list, where each query is listed in a new line.
 - Language of the query. English is set by default.
 - Number of articles to retrieve per query (range 1-25). Please note that querying is done recursively and that disambiguations are also retrieved, sometimes resulting in a larger number of queries than set on the slider.
2. Select which features to include as text features.
3. Information on the output.
4. Produce a report.
5. Run query.

Example

This is a simple example, where we use **Wikipedia** and retrieve the articles on ‘Slovenia’ and ‘Germany’. Then we simply apply default preprocessing with **Preprocess Text** and observe the most frequent words in those articles with **Word Cloud**.

The screenshot shows the Orange data mining software interface. A workflow is visible with three widgets: Wikipedia, Preprocess Text, and Word Cloud. The Word Cloud widget is active, displaying a word cloud and a table of words and weights.

Word Cloud Widget Details:

- Info:** 0 words in a topic, 20 documents with 8232 words.
- Cloud preferences:**
 - ☒ Color words
 - Words tilt: [slider]
 - Regenerate word cloud button
- Words & weights table:**

Weight	Word
864	germany
439	german
334	slovenia
221	world
156	war
143	european
140	national
133	team
132	women
126	hitler
121	country
- Save Image button**

Wikipedia Widget Details:

- Query:** Slovenia, Germany
- Language:** English
- Articles per query:** 10
- Text includes:**
 - ☒ Title
 - ☒ Content
 - ☒ Summary
 - ☒ Url
- Info:** Articles count 20
- Buttons:** Report, Search

Wikipedia works just like any other corpus widget (NY Times, Twitter) and can be used accordingly.