

Orange Data Mining interface showing a workflow for data analysis.

**Workflow Components:**

- Data Table** (Input)
- Line Plot** (Visualization)
- Scatter Plot** (Visualization)
- Data Info** (Information)
- Unique** (Transformation)
- Select Columns** (Transformation)
- Select Rows** (Transformation)
- Aggregate Columns** (Transformation)
- Formula** (Transformation)
- Preprocess (3)** (Transformation)
- Matching Data** (Transformation)
- Data Info (1) through (8)** (Information)
- Data Table (1)** and **Data Table (2)** (Output)

**Data Info - Orange Dialog:**

Data table properties

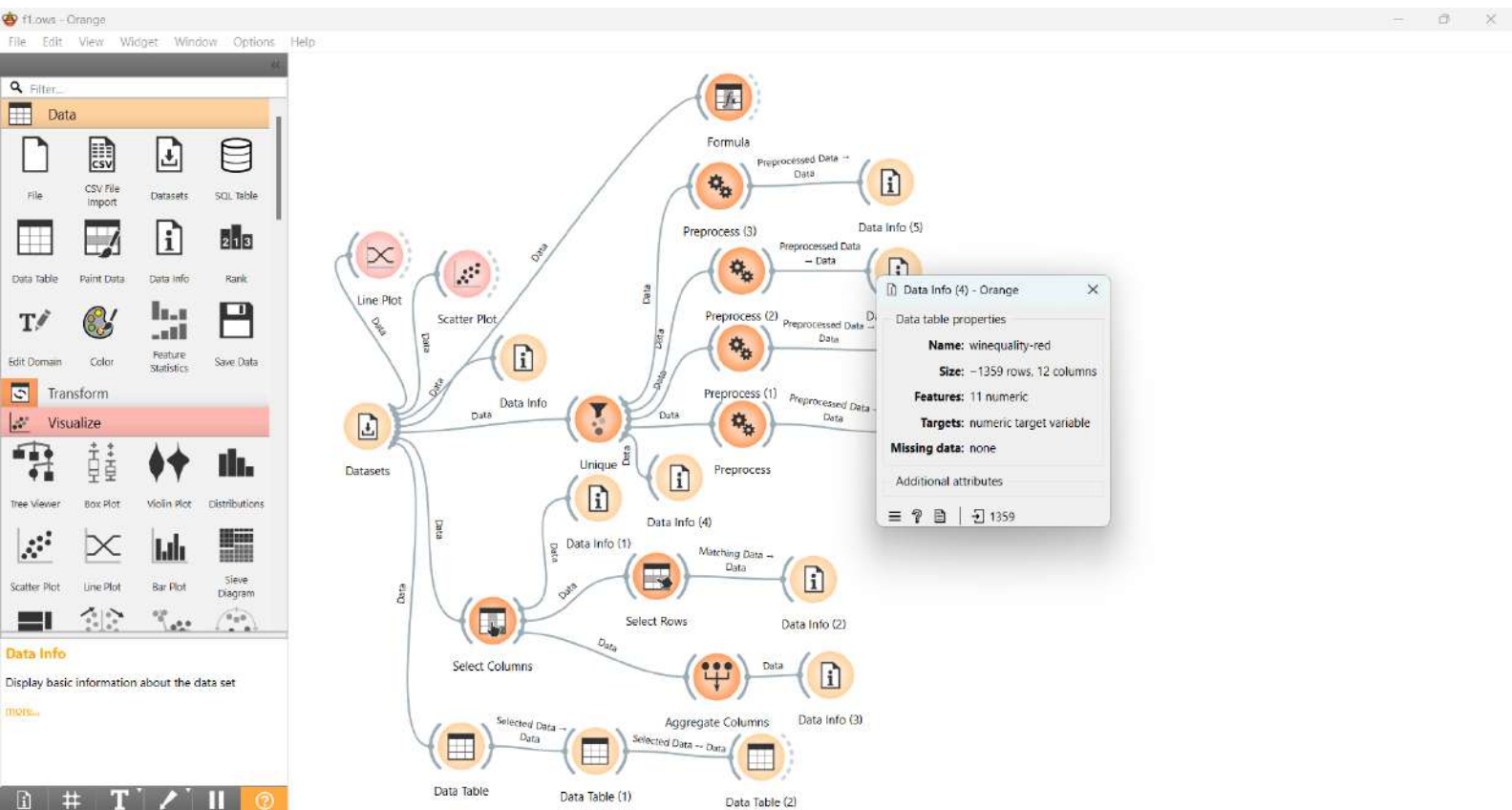
- Name:** winequality-red
- Size:** ~1599 rows, 12 columns
- Features:** 11 numeric
- Targets:** numeric target variable
- Missing data:** none
- Additional attributes:**

1599

**Left Panel:**

- Data:** File, CSV File Import, Datasets, SQL Table, Data Table, Paint Data, Data Info, Rank, Edit Domain, Color, Feature Statistics, Save Data.
- Transform:** Tree Viewer, Box Plot, Violin Plot, Distributions, Scatter Plot, Line Plot, Bar Plot, Sieve Diagram.
- Visualize:**

**Bottom Panel:** Data Info (Display basic information about the data set)



flows - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Preprocess

Construct a data preprocessing pipeline.

mot...

Preprocess (3) - Orange

Preprocessors

- Discretize Continuous Variables
- Continuize Discrete Variables
- Impute Missing Values
- Select Relevant Features
- Select Random Features
- Normalize Features
- Randomize
- Remove Sparse Features
- Principal Component Analysis
- CUR Matrix Decomposition

Normalize Features

- ☒ Standardize to  $\mu=0, \sigma^2=1$
- ☐ Center to  $\mu=0$
- ☐ Scale to  $\sigma^2=1$
- ☐ Normalize to interval [-1, 1]
- ☐ Normalize to interval [0, 1]

Apply Automatically

1359 1359

Select Columns

Selected Data --> Data

Aggregate Columns

Selected Data --> Data

Data Info (3)

Data Table

Data Table (1)

Data Table (2)

flows - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Preprocess

Construct a data preprocessing pipeline.

mot...

Preprocess (1) - Orange

Preprocessors

- Discretize Continuous Variables
- Continuize Discrete Variables
- Impute Missing Values
- Select Relevant Features
- Select Random Features
- Normalize Features
- Randomize
- Remove Sparse Features
- Principal Component Analysis
- CUR Matrix Decomposition

Principal Component Analysis

Components: 10

Apply Automatically

1359 1359

Select Columns

Selected Data ~ Data

Aggregate Columns

Selected Data ~ Data

Data Info (3)

Data Table

Data Table (1)

Data Table (2)

flows - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Formula

Construct new features (data columns) from a set of existing features in the input dataset.

more...

Line Plot Data

Scatter Plot Data

Datasets

Data

Formula - Orange

Variable Definitions

New X2 residual\_sugar

Remove ☐ Meta attribute Select Feature Select Function

X1 := volatile\_acidity

X2 := residual\_sugar

Send

Select Columns 1599 1599

Selected Data --> Data Table

Aggregate Columns Selected Data --> Data Table (1)

Data Info (3) Data Table (2)

The screenshot displays the Orange3 data mining software interface. On the left, a sidebar contains various widget categories: Data (File, CSV File Import, Datasets, SQL Table, Data Table, Paint Data, Data Info, Rank), Transform (Edit Domain, Color, Feature Statistics, Save Data), Visualize (Tree Viewer, Box Plot, Violin Plot, Distributions, Scatter Plot, Line Plot, Bar Plot, Sieve Diagram), and Formula. The main workspace shows a workflow starting with a 'Datasets' widget, which feeds into a 'Line Plot' and a 'Scatter Plot'. The 'Scatter Plot' is selected, and its configuration window is open. In this window, the 'Variable Definitions' section shows two new variables: 'X1 := volatile\_acidity' and 'X2 := residual\_sugar'. Below this, the 'Send' button is visible. The workflow continues from the 'Scatter Plot' to a 'Select Columns' widget, which then feeds into an 'Aggregate Columns' widget. The 'Aggregate Columns' widget outputs to 'Data Table (1)', which then feeds into 'Data Table (2)'. The 'Data Table (2)' widget is also selected, and its configuration window is open, showing the 'Data' tab with 1599 rows and 1599 columns.

Orange 3.10.0 - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Unique

Filter instances unique by specified key attribute(s).

more...

Line Plot

Scatter Plot

Data Info

Unique

Select Columns

Aggregate Columns

Data Table

Data Table (1)

Data Table (2)

Unique - Orange

Group by

Filter...

fixed acidity  
volatile acidity  
citric acid  
residual sugar  
chlorides  
free sulfur dioxide  
total sulfur dioxide  
density  
pH  
sulphates  
alcohol  
quality

Instance to select in each group:  
Last instance

Commit Automatically

1599 1359

flows - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Preprocess

Construct a data preprocessing pipeline.

mot...

Preprocess - Orange

Preprocessors

- Discretize Continuous Variables
- Continuize Discrete Variables
- Impute Missing Values
- Select Relevant Features
- Select Random Features
- Normalize Features
- Randomize
- Remove Sparse Features
- Principal Component Analysis
- CUR Matrix Decomposition

Impute Missing Values

- ☒ Average/Most frequent
- ☐ Replace with random value
- ☐ Remove rows with missing values.

Apply Automatically

1359 1359

Select Rows Data Info (2)

Select Columns Data

Aggregate Columns Data Info (3)

Data Table Data Table (1) Data Table (2)



flows - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Tree Viewer Box Plot Violin Plot Distributions

Scatter Plot Line Plot Bar Plot Sieve Diagram

Datasets

Load a dataset from an online repository

Search

# T / || ?

Data Table - Orange

Info

1599 instances (no missing data)  
11 features  
Numeric outcome  
No meta attributes.

Variables

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

Selection

☒ Select full rows

Restore Original Order

☒ Send Automatically

1599 | 1599 | 1599

Select Rows Data Info (2)

	quality	fixed acidity	volatile acidity	citric acid	residual sugar
1	5	7.4	0.700	0.00	1.90
2	5	7.8	0.880	0.00	2.60
3	5	7.8	0.760	0.04	2.30
4	6	11.2	0.280	0.56	1.90
5	5	7.4	0.700	0.00	1.90
6	5	7.4	0.660	0.00	1.80
7	5	7.9	0.600	0.06	1.60
8	7	7.3	0.650	0.00	1.20
9	7	7.8	0.580	0.02	2.00
10	5	7.5	0.500	0.36	6.10
11	5	6.7	0.580	0.08	1.80
12	5	7.5	0.500	0.36	6.10
13	5	5.6	0.615	0.00	1.60
14	5	7.8	0.610	0.29	1.60
15	5	8.9	0.620	0.18	3.80
16	5	8.9	0.620	0.19	3.90
17	7	8.5	0.280	0.56	1.80
18	5	8.1	0.560	0.28	1.70
19	4	7.4	0.590	0.08	4.40
20	6	7.9	0.320	0.51	1.80
21	6	8.9	0.220	0.48	1.80

Workflow Diagram:

```

graph LR
    Datasets --> LinePlot[Line Plot]
    Datasets --> ScatterPlot[Scatter Plot]
    Datasets --> DataInfo1[Data Info]
    Datasets --> Unique[Unique]
    Datasets --> SelectColumns[Select Columns]
    Datasets --> DataInfo2[Data Info (2)]
    Datasets --> AggregateColumns[Aggregate Columns]
    Datasets --> DataInfo3[Data Info (3)]
    SelectColumns -- "Selected Data --> Data" --> DataTable1[Data Table (1)]
    AggregateColumns -- "Selected Data --> Data" --> DataTable2[Data Table (2)]
    
```

untitled.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times PubMed Twitter Wikipedia

Preprocess Text Corpus to Network Bag of Words Document Embedding

Similarity Hashing Sentiment Analysis Tweet Profiler Topic Modelling

LDavis Corpus Viewer Score Documents Word Cloud

to be a to be the to be or

Concordance Document Map Word Enrichment Duplicate Detection

a+ 213a

Word List Extract Keywords Annotated Corpus Map Ontology

Semantic Viewer Collocations Statistics

Data Table

View the dataset in a spreadsheet.

more...

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Data Table - Orange

Info

950 instances  
0 features  
Target with 3 values  
4 meta attributes (5.7 % missing data)

Variables

☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection

☒ Select full rows

Restore Original Order

☒ Send Automatically

include title	medical_specialty	description	sample_name	transcription	keywords
1	Cardiovascular ...	2-D M-Mode. D...	2-D Echocardio...	2-D M-MODE ...	cardiovascular /...
2	Cardiovascular ...	2-D Echocardio...	2-D Echocardio...	1. The left vent...	cardiovascular /...
3	Cardiovascular ...	2-D Echocardio...	2-D Echocardio...	2-D ECHOCAR...	cardiovascular /...
4	Cardiovascular ...	Echocardiogra...	2-D Echocardio...	DESCRIPTION...	cardiovascular /...
5	Cardiovascular ...	Normal left ven...	2-D Doppler	2-D STUDY,1...	cardiovascular /...
6	Neurology	Cerebral Angio...	Moyamoya Dis...	CC; Confusion ...	?
7	Cardiovascular ...	Neck exploratio...	Tracheostomy	PREOPERATIVE ...	cardiovascular /...
8	Orthopedic	Youngswick ost...	Youngswick Ost...	TITLE OF OPER...	orthopedic, hall...
9	Orthopedic	Austin & Young...	Youngswick Bu...	PREOPERATIVE ...	orthopedic, hall...
10	Orthopedic	Consultation fo...	Wrist Pain	CHIEF COMPLA...	orthopedic, wrt...
11	Orthopedic	Excision of dors...	Wrist Ganglion ...	PREOPERATIVE ...	orthopedic, ori...
12	Orthopedic	Unilateral trans...	Vertebroplasty	PREOPERATIVE ...	orthopedic, tra...
13	Orthopedic	Trigger thumb r...	Trigger Thumb ...	PREOPERATIVE ...	orthopedic, trig...
14	Orthopedic	Right ulnar ner...	Ulnar Nerve Tra...	PREOPERATIVE ...	orthopedic, cub...
15	Orthopedic	Subcutaneous ...	Ulnar Nerve Tra...	PROCEDURE, ...	orthopedic, neu...
16	Orthopedic	Trigger finger r...	Trigger Finger R...	PROCEDURE, T...	orthopedic, a1 ...
17	Orthopedic	Decompression...	Ulnar Nerve De...	PREOPERATIVE ...	orthopedic, uln...
18	Orthopedic	Total knee repla...	Total Knee Repl...	PROCEDURES; ...	orthopedic, pro...
19	Orthopedic	Right total knee...	Total Knee Arth...	PREOPERATIVE ...	orthopedic, ost...
20	Orthopedic	Total left knee r...	Total Knee Repl...	PREOPERATIVE ...	orthopedic, deg...

untitled.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times Pubmed Twitter Wikipedia

Preprocess Text Corpus to Network Bag of Words Document Embedding

Similarity Hashing Sentiment Analysis Tweet Profiler Topic Modelling

LDavis Corpus Viewer Score Documents Word Cloud

Concordance Document Map Word Enrichment Duplicate Detection

Word List Extract Keywords Annotated Corpus Map Ontology

Semantic Viewer Collocations Statistics

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

1 # T / || ?

```
graph LR
    subgraph Workflow1
        Datasets --> Data
        Datasets --> Data_Table[Data Table]
        Data --> Corpus
        Corpus --> Corpus_Viewer[Corpus Viewer]
    end
    subgraph Workflow2
        Create_Corpus[Create Corpus] --> Preprocess_Text[Preprocess Text]
        Create_Corpus --> Corpus_Viewer_1[Corpus Viewer (1)]
        Create_Corpus --> Word_Cloud[Word Cloud]
        Preprocess_Text --> Word_Cloud_1[Word Cloud (1)]
    end
```

untitled.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times Pubmed Twitter Wikipedia

Preprocess Text Corpus to Network Bag of Words Document Embedding

Similarity Hashing Sentiment Analysis Tweet Profiler Topic Modelling

LDavis Corpus Viewer Score Documents Word Cloud

to be a to be the to be or

Concordance Document Map Word Enrichment Duplicate Detection

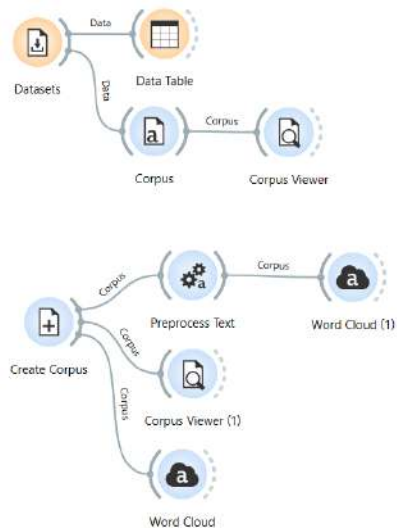
a+ Extract Keywords Annotated Corpus Map Ontology

Semantic Viewer Collocations Statistics

Create Corpus

Write/paste documents to create a corpus

more...



Create Corpus - Orange

Language: English

improves heart health, and enhances mental well-being. People who exercise regularly experience better mood and reduced stress levels. Physical activities such as running, swimming, and yoga contribute to overall fitness.

Renewable Energy Sources

Renewable energy is derived from natural sources such as sunlight, wind, and water. Solar panels convert sunlight into electricity, while wind turbines generate power using wind energy. Unlike fossil fuels, renewable energy is sustainable and environmentally friendly. Governments worldwide are investing in clean energy to reduce carbon emissions.

The Benefits of Reading

Reading enhances knowledge and improves vocabulary. It stimulates brain function and strengthens memory. Fiction books improve creativity, while non-fiction books provide valuable information. Developing a reading habit can boost concentration and reduce stress levels.

Climate Change and Its Effects

Climate change is a global concern caused by pollution and deforestation. Rising temperatures lead to melting ice caps and extreme weather conditions. Human activities, such as burning fossil fuels, increase greenhouse gas emiss

Add document ☒ Apply Automatically

4

untitled1.ows - Orange

FileEditViewWidgetWindowOptionsHelp

Filter...

NY Times

Preprocess Text

Similarity Hashing

LDavis

to be a to be the to be or

Concordance

a+

Word List

Semantic Viewer

Pubmed

Corpus to Network

Sentiment Analysis

Corpus Viewer

Document Map

213

Extract Keywords

Collocations

Twitter

Bag of Words

Tweet Profiler

Score Documents

Word Enrichment

Annotated Corpus Map

to + be be + in to + do

Statistics

Wikipedia

Document Embedding

Topic Modelling

Word Cloud

Duplicate Detection

Ontology

Corpus Viewer - Orange

Info

Tokens: n/a

Types: n/a

Matching documents: 950/950

Matches: n/a

Search features

Filter...

medical\_specialty

description

sample\_name

transcription

keywords

Display features

Filter...

medical\_specialty

description

sample\_name

transcription

keywords

Show Tokens & Tags

Auto send is on

9501 | 949 | 950

RegEx Filter:

1 2-D Echocardiogram - 1

2 2-D Echocardiogram - 2

3 2-D Echocardiogram - 3

4 2-D Echocardiogram - 4

5 2-D Doppler

6 Moyamoya Disease

7 Tracheostomy

8 Youngswick Osteotomy

9 Youngswick Bunionectomy

10 Wrist Pain

11 Wrist Ganglion Excision

12 Vertebroplasty

13 Trigger Thumb Release - 1

14 Ulnar Nerve Transposition & ...

15 Ulnar Nerve Transposition (1)

16 Trigger Finger Release

17 Ulnar Nerve Decompression

medical\_specialty:

Cardiovascular / Pulmonary

description:

2-D M-Mode. Doppler.

sample\_name:

2-D Echocardiogram - 1

transcription:

2-D M-MODE: , 1. Left atrial enlargement with left atrial diameter of 4.7 cm., 2. Normal size right and left ventricle., 3. Normal LV systolic function with left ventricular ejection fraction of 51%, 4. Normal LV diastolic function., 5. No pericardial effusion., 6. Normal morphology of aortic valve, mitral valve, tricuspid valve, and pulmonary valve., 7. PA systolic pressure is 36 mmHg. DOPPLER: , 1. Mild mitral and tricuspid regurgitation., 2. Trace aortic and pulmonary regurgitation.

keywords:

cardiovascular / pulmonary, 2-d m-mode, doppler, aortic valve, atrial enlargement, diastolic function, ejection fraction, mitral, mitral valve, pericardial effusion, pulmonary valve, regurgitation, systolic function, tricuspid, tricuspid valve, normal lv

Corpus Viewer

Display corpus contents.

more...

#

T

||

?

untitled1.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times Pubmed Twitter Wikipedia

Preprocess Text Corpus to Network Bag of Words Document Embedding

Similarity Hashing Sentiment Analysis Tweet Profiler Topic Modelling

LDAvis Corpus Viewer Score Documents Word Cloud

to be a to be the to be or

Concordance Document Map Word Enrichment Duplicate Detection

a+ Word List Extract Keywords Annotated Corpus Map Ontology

Semantic Viewer Collocations Statistics

Word Cloud

more...

Word Cloud - Orange

Cloud preferences

☒ Color words

Words tilt: no

Words & weights

Weight	Word
4	is
4	energy
3	such
3	as
3	to
3	wind
2	exercise
2	health
2	it
2	in
2	improves
2	enhances
2	-
2	stress

Cloud preferences

☒ Color words

Words tilt: no

Words & weights

increase information greenhouse strengthens information greenhouse by melting carbon pollution human caused clean solar regularly reduced people unlike provide developing fitness weight improves sunlight mood friendly ice worldwide change overall improve better fossil energy in gas books governments physical boost investing activities it enhances can regular derived lead concentration habit water good health and a levels heart using caps extreme natural who for stress as is such while mental memory valuable into well function from fiction exercise renewable running knowledge climate contribute reading fuels maintaining swimming concern environmentally experience management sources creativity emiss rising non power electricity convert emissions weather burning deforestation stimulates vocabulary temperatures

such|



untitled.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times Pubmed Twitter Wikipedia

Preprocess Text  
100  
1010  
01

Similarity Hashing  
A B C

LDAs

to be a  
to be the  
to be or

Concordance

a+

Word List

Semantic Viewer

Corpus to Network

Bag of Words

Document Embedding

Sentiment Analysis

Tweet Profiler

Topic Modelling

Corpus Viewer

Score Documents

Word Cloud

Document Map

Word Enrichment

Duplicate Detection

Extract Keywords

Annotated Corpus Map

Ontology

to + be  
be + in  
to + do

Collocations

Statistics

Create Corpus

Preprocess Text

Corpus Viewer (1)

Word Cloud

Preprocess Text - Orange

Preprocessors

- Transformation
- Tokenization
- Normalization
- Filtering
- N-grams Range
- POS Tagger

Preview

regular, exercise, essential, maintaining, good

Output

Tokens: 121  
Types: 101

Apply Automatically

Transformation

- ☒ Lowercase
- ☐ Remove accents
- ☐ Parse html
- ☐ Remove urls

Tokenization

- ☐ Word Punctuation
- ☐ Whitespace
- ☐ Sentence
- ☒ Regexp Pattern: \w+
- ☐ Tweet

Filtering

- ☒ Stopwords (none) New Text Docu...
- ☐ Lexicon (none) ...
- ☐ Numbers
- ☐ Includes Numbers
- ☐ Regexp

Preprocess Text

Construct a text pre-processing pipeline.

[more...](#)

1 # T / || ?



untitled.ows - Orange

File Edit View Widget Window Options Help

Filter...

NY Times Pubmed Twitter Wikipedia

Preprocess Text Corpus to Network Bag of Words Document Embedding

Similarity Hashing Sentiment Analysis Tweet Profiler Topic Modelling

LDavis Corpus Viewer Score Documents Word Cloud

to be a to be the to be or

Concordance Document Map Word Enrichment Duplicate Detection

a+ Extract Keywords Annotated Corpus Map Ontology

Semantic Viewer Collocations Statistics

Word Cloud

more...

Word Cloud (1) - Orange

Cloud preferences

☒ Color words

Words tilt: no

Words & weights

Weight	Word
4	energy
3	wind
2	exercise
2	health
2	improves
2	enhances
2	stress
2	levels
2	activities
2	renewable
2	sunlight
2	fossil
2	fuels
2	reduce

books renewable fiction enhances wind health stress energy levels improves fossil fuels reading activities