

4 Dokumentation

4.1 Klassen

```
class TfDPy. Trainingdata (num_words=10000, file_ending='.txt',  
folder='/my/path/', window_size=40, ratio=0.1, maxlen=81,  
chunk_size=250, gram_size=6, lang='en', chunk_scope='token', se-  
quence_scope='token', categorical_scope='sequences')
```

Parameters:	num_words : int, optional Anzahl der mfw, die für das Training verwendet wird
	file_ending : string, optional Endung der Dateien mit Trainingsdaten
	folder : string, optional Pfad zum lokalen Verzeichnis mit Trainingsdaten
	window_size : int, optional Größe des Kontextfensters für sequentielle Klassifikation
	ratio : int, optional Anteil der Daten, die für das Testen des Classifiers verwendet wird
	maxlen : int, optional Anzahl der Einheiten pro Feature, wird für Padding verwendet
	chunk_size : int, optional Größe der Teildokumente
	gram_size : int, optional Größe der n-gramme
	lang : string, optional Sprache für NLP Modelle (ISO 639-I)
	chunk_scope : string, optional Angabe auf welcher Ebene Teildokumente erstellt werden sollen
	sequence_scope : string, optional Ebene für die Erzeugung eines Wortindex
	categorical_scope : string, optional Ebene für die Umwandlung der Kategorien in integer

Methods

collect_files_from_dir(folder, fileending)	stub
collect_data_from_tsv(folder)	stub
nlp_text(lang, corpus_df)	stub
add_categories(corpus_df)	stub
add_text(lang, corpus_df)	stub
to_senteces(corpus_df)	stub
to_chunks(corpus_df, chunk_scope)	stub
to_chars(corpus_df)	stub
to_ngrams(corpus_df, scope, gram_size)	stub
generate_one_hot_matrix(corpus_df, scope)	stub
generate_sequences(corpus_df, sequences_scope, num_words)	stub
padding_sequences(corpus_df, maxlen)	stub
to_categorical_trainingdata(corpus_df, categorical_scope)	stub
to_multilabel_trainingdata(corpus_df, labels)	stub
split_training_data(x_train, x_test, y_train, y_test)	stub
load_sequential_context(corpus_df, window_size)	stub
to_sequential_trainingdata(corpus_df)	stub

collect_files_from_dir (folder, fileending)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

collect_data_from_tsv (folder)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

nlp_text (lang, corpus_df)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

add_categories (corpus_df)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

to_senteces (corpus_df)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

to_chunks (corpus_df, chunk_scope)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

to_chars (corpus_df)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

to_ngrams (corpus_df, scope, gram_size)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

generate_one_hot_matrix (corpus_df, scope)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

generate_sequences (corpus_df, sequences_scope, num_words)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

padding_sequences (corpus_df, maxlen)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten

Returns: stub

to_categorical_trainingdata (corpus_df, categorical_scope)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

to_multilabel_trainingdata (corpus_df, labels)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

split_training_data (x_train, x_test, y_train, y_test)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

load_sequential_context (corpus_df, window_size)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

to_sequential_trainingdata (corpus_df)

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
Returns:	stub

4.2 Funktionen

func TfDPy. **read_save** (f_path)

Parameters:	f_path : string Pfad zu einer Datei
-------------	---

func TfDPy. **context_grabber** (sequence, windowsize)

Parameters:	sequence : array-like Liste von Features windowsize : int Größe des Kontextfenster
-------------	---

func TfDPy. **fill_parameters** (parameter_raw)

Parameters: **parameter_raw** : dict
Dictionary aus Parametern

func TfDPy. **single_run_paramsearch** ((pipeline, classifier, sequence_scope_single, num_words_single, file_ending_single, folder_single, window_size_single, ratio_single, maxlen_single, gram_size_single, chunk_size_single, lang_single, chunk_scope_single, categorical_scope_single)

Parameters: **pipeline** : array-like
Definition der Pipeline als Liste von Arbeitsschritten
classifier : array-like
Dictionary aus Parametern passend zur Pipeline
sequence_scope_single : string
Ebene für die Erzeugung eines Wortindex
num_words_single : int
Anzahl der mfw, die für das Training verwendet wird
file_ending_single : string
Endung der Dateien mit Trainingsdaten
folder_single : string
Pfad zum lokalen Verzeichnis mit Trainingsdaten
window_size_single : int
Größe des Kontextfensters für sequentielle Klassifikation
ratio_single : int
Anteil der Daten, die für das Testen des Classifiers verwendet wird
maxlen_single : int
Anzahl der Einheiten pro Feature, wird für Padding verwendet
gram_size_single : int
Größe der n-gramme
chunk_size_single : int
Größe der Teildokumente
lang_single : string
Sprache für NLP Modelle (ISO 639-I)
chunk_scope_single : string
Angabe auf welcher Ebene Teildokumente erstellt werden sollen
categorical_scope_single : string
Ebene für die Umwandlung der Kategorien in integer

```
func TfDPy. td_paramsearch (pipeline, parameters, classifier)
```

Parameters:	pipeline : array-like Definition der Pipeline als Liste von Arbeitsschritten
	parameters : dict Dictionary aus Parametern passend zu den Arbeitsschritten
	classifier : array-like Liste mit einem oder mehreren Classifiern

5 Bibliographie