Bachelor Thesis Documentation

Release v1.0

Bachelor Thesis

Jul 26, 2016

CONTENTS

1 Bachelorarbeit **3**
1.1 src package . . . 4
1.1.1 Subpackages . . . 4
src.clustering package . . . 4
Submodules . . . 4
src.clustering.cluster_mappings module . . . 4
Module contents . . . 4
src.eval package . . . 4
Submodules . . . 4
src.eval.analogy module . . . 4
src.eval.concentration module . . . 4
src.eval.concept_groups module . . . 4
src.eval.eval_vectors module . . . 4
src.eval.word_similarity module . . . 4
Module contents . . . 4
src.guesser package . . . 4
Submodules . . . 4
src.guesser.svm_guesser module . . . 4
Module contents . . . 4
src.mapping package . . . 4
Submodules . . . 4
src.mapping.map_vectors module . . . 4
src.mapping.mapping_operations module . . . 4
src.mapping.mapthreading module . . . 4
Module contents . . . 4
src.prep package . . . 4
Subpackages . . . 4
Module contents . . . 7
src.trans_e package . . . 7
Submodules . . . 7
src.trans_e.add_inverse_relations module . . . 7
src.trans_e.clean_relations module . . . 8
src.trans_e.contains_entities module . . . 8
src.trans_e.convert_relations module . . . 8
src.trans_e.differentiate_datasets module . . . 8
src.trans_e.partition_data module . . . 8
Module contents . . . 9
1.1.2 Module contents . . . 9

2 Indices and tables **11**

Python Module Index 13

Index 15

Contents:

CHAPTER

ONE

BACHELORARBEIT

1.1 src package

1.1.1 Subpackages

src.clustering package

Submodules

src.clustering.cluster_mappings module

Module contents

src.eval package

Submodules

src.eval.analogy module

src.eval.concentration module

src.eval.concept_groups module

src.eval.eval_vectors module

src.eval.word_similarity module

Module contents

src.guesser package

Submodules

src.guesser.svm_guesser module

Module contents

src.mapping package

Submodules

src.mapping.map_vectors module

src.prep.corpus.convert_to_plain. **contains_tag** (*line*)

src.prep.corpus.convert_to_plain. **convert_decow_to_plain** (*decow_dir*, *out_dir*, *log_path*, *merge_nes*, *log_interval*)

src.prep.corpus.convert_to_plain. **convert_part** (*argstuple*)

src.prep.corpus.convert_to_plain. **convert_part_merging** (*argstuple*)

src.prep.corpus.convert_to_plain. **extract_named_entity** (*line*)

src.prep.corpus.convert_to_plain. **extract_sentence_id** (*tag*)

src.prep.corpus.convert_to_plain. **get_file_number** (*filename*)

src.prep.corpus.convert_to_plain. **log_time** (*logpath='log.txt'*, *interval=5*)

src.prep.corpus.convert_to_plain. **log_time_mp** (*logpath='log.txt'*, *interval=5*)

src.prep.corpus.convert_to_plain. **main** ( )

src.prep.corpus.extract_conll module

src.prep.corpus.extract_conll. **extract_conll** (*inpath*, *outpath*, *column*)

Extract information out of CoNLL files.

Parameters **inpath** (*str*) – Path to input file.

src.prep.corpus.extract_conll. **init_argparse** ( )

src.prep.corpus.extract_conll. **main** ( )

src.prep.corpus.mapper module

src.prep.corpus.prepare_corpus module

src.prep.corpus.quick_and_dirty module

src.prep.corpus.quick_and_dirty. **clean_file** (*inpath*, *outpath*)

src.prep.corpus.reducer module

Module contents

src.prep.misc package

Submodules

src.prep.misc.decorators module

src.prep.misc.decorators. **alt** (*func*)

src.prep.misc.decorators. **log_time** (*logpath='log.txt'*, *interval=5*)

src.prep.misc.decorators. **log_time_mp** (*logpath='log.txt'*, *interval=5*)

Module contents

src.prep.nes package

Submodules

src.prep.nes.extractNE module

src.prep.nes.extractNE. **contains_tag** (*line*)

src.prep.nes.extractNE. **extract_named_entity** (*line*)

src.prep.nes.extractNE. **extract_sentence_id** (*tag*)

src.prep.nes.extractNE. **main** ( )

src.prep.nes.extractNE. **print_dict_in_file** (*dictionary*, *out_path*)

src.prep.nes.extractNE. **print_ids_in_file** (*dictionary*, *out_path*)

src.prep.nes.extractNE. **print_list_in_file** (*ne_list*, *out_path*)

src.prep.nes.extractNE. **process** (*inpath*, *outpath*, *logpath*)

src.prep.nes.merge module

src.prep.nes.merge. **dump_ids_dict** (*idsdict*, *outpath*)

src.prep.nes.merge. **freqWorker** (*inpath*)

src.prep.nes.merge. **idWorker** (*inpath*)

src.prep.nes.merge. **load_ids_dict** (*inpath*)

src.prep.nes.merge. **main** ( )

src.prep.nes.merge. **mergeDicts** (*dicttuple*)

src.prep.nes.merge. **merge_frequency_files** (*infiles_path*, *outpath*, *logpath*)

src.prep.nes.merge. **merge_id_dicts** (*dicttuple*)

src.prep.nes.merge. **merge_id_files** (*infiles_path*, *outpath*, *logpath*, *yaml=False*)

src.prep.nes.merge. **print_key_lengths** (*dictionary*)

src.prep.nes.merge. **rl** (*infile*)

src.prep.nes.mwe module

src.prep.nes.mwe. **create_mwe_pickle** (*inpath*, *outpath*, *logpath='./mwes.log'*)

src.prep.nes.mwe. **create_mwe_pickle2** (*inpath*, *outpath*, *logpath='./mwes.log'*)

src.prep.nes.mwe. **dump_dict_pickle** (*d*, *outpath*)

src.prep.nes.mwe. **dump_dict_pickle2** (*d*, *outpath*)

src.prep.nes.mwe. **load_dict_pickle** (*inpath*)

src.prep.nes.mwe. **load_dict_pickle2** (*inpath*)

src.prep.nes.mwe. **main** ( )

src.prep.nes.mwe. **replace_mwes** (*mwe_path*, *corpus_path*, *out_path*)

src.prep.nes.statistics module

Module contents

src.prep.relations package

Submodules

src.prep.relations.relations module

exception src.prep.relations.relations. **MissingTranslationException**
Bases: exceptions.Exception

get_id ( )

src.prep.relations.relations. **fetch_name** (*id*, *lang='en'*)

src.prep.relations.relations. **fetch_relation_triples_of_file** (*inpath*, *outpath*, *logpath*, *lang='en'*)

src.prep.relations.relations. **format_fbid** (*id*)

src.prep.relations.relations. **freebase_request** (*query*, *api_key*, *service_url*)

src.prep.relations.relations. **main** ( )

src.prep.relations.relations. **read_credentials** ( )

src.prep.relations.relations. **rl** (*infile*)

src.prep.relations.relations. **translate_name** (*name*, *lang='en'*)

src.prep.relations.relations. **translate_word2vec_question_phrases** (*inpath*, *outpath*, *lang='en'*)

Module contents

Module contents

src.trans_e package

Submodules

src.trans_e.add_inverse_relations module

src.trans_e.add_inverse_relations. **add_inverse_relations** (*relations_inpath*, *relations_outpath*, *inverse_relations*, *known_relations*)

src.trans_e.add_inverse_relations. **init_argparse** ( )

src.trans_e.add_inverse_relations. **main** ( )

src.trans_e.add_inverse_relations. **read_file_with_inverse_relations** (*inverse_inpath*)

src.trans_e.clean_relations module

src.trans_e.contains_entities module

src.trans_e.contains_entities. **contains_entities** (*entities1*, *entities2*)

src.trans_e.contains_entities. **create_new_dataset** (*entities1*, *dataset*, *outpath*)

src.trans_e.contains_entities. **extract_entities_from_relation_dataset** (*dataset_inpath*)

src.trans_e.contains_entities. **extract_entities_from_tql_file** (*tql_path*)

src.trans_e.contains_entities. **format_fbid** (*id*)

src.trans_e.contains_entities. **init_argparse** ( )

src.trans_e.contains_entities. **main** ( )

src.trans_e.convert_relations module

src.trans_e.differentiate_datasets module

src.trans_e.differentiate_datasets. **compare_entities** (*set1*, *set2*)

src.trans_e.differentiate_datasets. **init_argparse** ( )

src.trans_e.differentiate_datasets. **main** ( )

src.trans_e.differentiate_datasets. **read_dataset** (*inpath*)

src.trans_e.partition_data module

src.trans_e.partition_data. **check_data_integrity** (*data_inpath*, *remove_clones*, *outpath*)

Check whether all triplets in the data are unique.

src.trans_e.partition_data. **check_set_integrity** (*indir*)

src.trans_e.partition_data. **get_stats** (*data*)

src.trans_e.partition_data. **init_argparse** ( )

src.trans_e.partition_data. **main** ( )

src.trans_e.partition_data. **partition_data** (*data*, *prts*, *outdir*, *whole=True*)

src.trans_e.partition_data. **partition_relation_wise** (*data*, *prts*)

src.trans_e.partition_data. **partition_whole** (*data*, *prts*)

src.trans_e.partition_data. **partitions_list** (*l*, *prts*)

src.trans_e.partition_data. **read_only_relations_into_set** (*inpath*)

src.trans_e.partition_data. **read_relations** (*inpath*)

src.trans_e.partition_data. **write_data_in_file** (*data*, *outfile*)

Module contents

1.1.2 Module contents

CHAPTER

TWO

INDICES AND TABLES

- genindex
- modindex
- search

S

src, 9
src.clustering, 4
src.eval, 4
src.guesser, 4
src.mapping, 4
src.prep, 7
src.prep.corpus, 5
src.prep.corpus.convert_to_plain, 4
src.prep.corpus.extract_conll, 5
src.prep.corpus.mapper, 5
src.prep.corpus.quick_and_dirty, 5
src.prep.misc, 6
src.prep.misc.decorators, 5
src.prep.nes, 7
src.prep.nes.extractNE, 6
src.prep.nes.merge, 6
src.prep.nes.mwe, 6
src.prep.relations, 7
src.prep.relations.relations, 7
src.trans_e, 9
src.trans_e.add_inverse_relations, 7
src.trans_e.contains_entities, 8
src.trans_e.differentiate_datasets, 8
src.trans_e.partition_data, 8

A

add_inverse_relations() (in module src.trans_e.add_inverse_relations), 7
alt() (in module src.prep.corpus.convert_to_plain), 4
alt() (in module src.prep.misc.decorators), 5

C

check_data_integrity() (in module src.trans_e.partition_data), 8
check_set_integrity() (in module src.trans_e.partition_data), 8
clean_file() (in module src.prep.corpus.quick_and_dirty), 5
compare_entities() (in module src.trans_e.differentiate_datasets), 8
contains_entities() (in module src.trans_e.contains_entities), 8
contains_tag() (in module src.prep.corpus.convert_to_plain), 5
contains_tag() (in module src.prep.nes.extractNE), 6
convert_decow_to_plain() (in module src.prep.corpus.convert_to_plain), 5
convert_part() (in module src.prep.corpus.convert_to_plain), 5
convert_part_merging() (in module src.prep.corpus.convert_to_plain), 5
create_mwe_pickle() (in module src.prep.nes.mwe), 6
create_mwe_pickle2() (in module src.prep.nes.mwe), 6
create_new_dataset() (in module src.trans_e.contains_entities), 8

D

dump_dict_pickle() (in module src.prep.nes.mwe), 6
dump_dict_pickle2() (in module src.prep.nes.mwe), 6
dump_ids_dict() (in module src.prep.nes.merge), 6

E

extract_conll() (in module src.prep.corpus.extract_conll), 5
extract_entities_from_relation_dataset() (in module src.trans_e.contains_entities), 8
extract_entities_from_tql_file() (in module src.trans_e.contains_entities), 8
extract_named_entity() (in module src.prep.corpus.convert_to_plain), 5
extract_named_entity() (in module src.prep.nes.extractNE), 6
extract_sentence_id() (in module src.prep.corpus.convert_to_plain), 5
extract_sentence_id() (in module src.prep.nes.extractNE), 6

F

fetch_name() (in module src.prep.relations.relations), 7
fetch_relation_triples_of_file() (in module src.prep.relations.relations), 7
format_fbid() (in module src.prep.relations.relations), 7
format_fbid() (in module src.trans_e.contains_entities), 8
freebase_request() (in module src.prep.relations.relations), 7
freqWorker() (in module src.prep.nes.merge), 6

G

get_file_number() (in module src.prep.corpus.convert_to_plain), 5
get_id() (src.prep.relations.relations.MissingTranslationException method), 7
get_stats() (in module src.trans_e.partition_data), 8

I

idWorker() (in module src.prep.nes.merge), 6
init_argparse() (in module src.prep.corpus.extract_conll), 5
init_argparse() (in module src.trans_e.add_inverse_relations), 7
init_argparse() (in module src.trans_e.contains_entities), 8
init_argparse() (in module src.trans_e.differentiate_datasets), 8
init_argparse() (in module src.trans_e.partition_data), 8

L

load_dict_pickle() (in module src.prep.nes.mwe), 6

load_dict_pickle2() (in module src.prep.nes.mwe), 6
load_ids_dict() (in module src.prep.nes.merge), 6
log_time() (in module src.prep.corpus.convert_to_plain), 5
log_time() (in module src.prep.misc.decorators), 5
log_time_mp() (in module src.prep.corpus.convert_to_plain), 5
log_time_mp() (in module src.prep.misc.decorators), 5

M

main() (in module src.prep.corpus.convert_to_plain), 5
main() (in module src.prep.corpus.extract_conll), 5
main() (in module src.prep.nes.extractNE), 6
main() (in module src.prep.nes.merge), 6
main() (in module src.prep.nes.mwe), 6
main() (in module src.prep.relations.relations), 7
main() (in module src.trans_e.add_inverse_relations), 8
main() (in module src.trans_e.contains_entities), 8
main() (in module src.trans_e.differentiate_datasets), 8
main() (in module src.trans_e.partition_data), 8
merge_frequency_files() (in module src.prep.nes.merge), 6
merge_id_dicts() (in module src.prep.nes.merge), 6
merge_id_files() (in module src.prep.nes.merge), 6
mergeDicts() (in module src.prep.nes.merge), 6
MissingTranslationException, 7

P

partition_data() (in module src.trans_e.partition_data), 8
partition_relation_wise() (in module src.trans_e.partition_data), 8
partition_whole() (in module src.trans_e.partition_data), 8
partitions_list() (in module src.trans_e.partition_data), 8
print_dict_in_file() (in module src.prep.nes.extractNE), 6
print_ids_in_file() (in module src.prep.nes.extractNE), 6
print_key_lengths() (in module src.prep.nes.merge), 6
print_list_in_file() (in module src.prep.nes.extractNE), 6
process() (in module src.prep.nes.extractNE), 6

R

read_credentials() (in module src.prep.relations.relations), 7
read_dataset() (in module src.trans_e.differentiate_datasets), 8
read_file_with_inverse_relations() (in module src.trans_e.add_inverse_relations), 8
read_only_relations_into_set() (in module src.trans_e.partition_data), 8
read_relations() (in module src.trans_e.partition_data), 8
replace_mwes() (in module src.prep.nes.mwe), 7
rl() (in module src.prep.nes.merge), 6
rl() (in module src.prep.relations.relations), 7

S

src (module), 9
src.clustering (module), 4
src.eval (module), 4
src.guesser (module), 4
src.mapping (module), 4
src.prep (module), 7
src.prep.corpus (module), 5
src.prep.corpus.convert_to_plain (module), 4
src.prep.corpus.extract_conll (module), 5
src.prep.corpus.mapper (module), 5
src.prep.corpus.quick_and_dirty (module), 5
src.prep.misc (module), 6
src.prep.misc.decorators (module), 5
src.prep.nes (module), 7
src.prep.nes.extractNE (module), 6
src.prep.nes.merge (module), 6
src.prep.nes.mwe (module), 6
src.prep.relations (module), 7
src.prep.relations.relations (module), 7
src.trans_e (module), 9
src.trans_e.add_inverse_relations (module), 7
src.trans_e.contains_entities (module), 8
src.trans_e.differentiate_datasets (module), 8
src.trans_e.partition_data (module), 8

T

translate_name() (in module src.prep.relations.relations), 7
translate_word2vec_question_phrases() (in module src.prep.relations.relations), 7

W

write_data_in_file() (in module src.trans_e.partition_data), 8