Package: RNGT (via r-universe)

July 4, 2024

Title Wrappers for 'NGT'

Version 0.0.0.9001

Description Wrappers for 'NGT' (Neighborhood Graph and Tree for indexing high-dimensional data) which performs high-speed approximate nearest neighbor searches against a large volume of data in high dimensional vector data space.

License Apache License (>= 2)

BugReports https://github.com/paithiov909/RNGT/issues

Depends R (>= 2.10)

Imports methods, R6 (>= 2.4.0), Rcpp, rlang, tibble

Suggests testthat (>= 3.0.0)

LinkingTo Rcpp

Config/testthat/edition 3

Encoding UTF-8

LazyData true

OS_type unix

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

SystemRequirements GNU make, cmake

Repository https://paithiov909.r-universe.dev

RemoteUrl https://github.com/paithiov909/RNGT

RemoteRef HEAD

RemoteSha aed056d9866ec4633e07fec2c7fa5ef598f46a72

Contents

gen7singles2018 *GloVe Model of Pokémon*

Description

A GloVe model that trained on 11,446 Pokémon parties (for single battles in 2018), while considering each party as a sentence and each Pokémon as a word, and pruning vocabularies.

Usage

```
gen7singles2018
```

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 181 rows and 66 columns.

NgtIndex *R6 Class for Graph and Tree Based Index*

Description

Graph and tree based index.

Public fields

`path` path of the index.

Methods

Public methods:

- `NgtIndex$new()`
- `NgtIndex$open()`
- `NgtIndex$create()`
- `NgtIndex$search()`
- `NgtIndex$save()`
- `NgtIndex$remove()`
- `NgtIndex$refine_anng()`
- `NgtIndex$get_object()`
- `NgtIndex$reset_defaults()`
- `NgtIndex$get_info()`
- `NgtIndex$build_index()`
- `NgtIndex$batch_insert()`
- `NgtIndex$insert()`

- NgtIndex$export_index()
- NgtIndex$import_index()
- NgtIndex$close()

Method `new()`**:** Creates a new `NgtIndex` object.

Usage:

```
NgtIndex$new(path, sub_dir = rand_name("NgtIndex"))
```

Arguments:

`path` path to the NGT index.

`sub_dir` sub-directory of the index.

Returns: a new `NgtIndex` object.

Method `open()`**:** Opens a NGT index.

Usage:

```
NgtIndex$open(read_only = FALSE, tree_disabled = FALSE, log_disabled = FALSE)
```

Arguments:

`read_only` whether the index is read only.

`tree_disabled` whether the tree is disabled.

`log_disabled` whether the log is disabled.

Method `create()`**:** Creates an empty index with the specified parameters.

Usage:

```
NgtIndex$create(
  dimension,
  edge_size_for_creation = 10,
  edge_size_for_search = 40,
 distance_type = c("l2", "l1", "normalized_l2", "hamming", "jaccard", "sparse_jaccard",
  "angle", "normalized_angle", "cosine", "normalized_cosine", "normalized_l2"),
  object_type = c("float", "byte", "float16")
)
```

Arguments:

`dimension` dimension of the vectors.

`edge_size_for_creation` number of edges for each node in the graph.

`edge_size_for_search` number of edges to search.

`distance_type` distance type.

`object_type` object type.

`sub_dir` sub directory to store the index in.

Method `search()`**:** Searches for the k approximate nearest neighbors of the specifiecd query object.

Usage:

```
NgtIndex$search(
  query,
  k = 20L,
  epsilon = 0.1,
  edge_size = -1L,
  expected_accuracy = -1,
  with_distance = TRUE
)
```

Arguments:

`query` query object.

`k` number of nearest neighbors.

`epsilon` epsilon which defines the explored range for the graph

`edge_size` number of edges for each node to explore the graph

`expected_accuracy` expected accuracy.

`with_distance` whether to return distance.

Returns: tibble.

Method `save()`**:** Saves the index.

Usage:

```
NgtIndex$save(path)
```

Arguments:

`path` path to save the index. defaults to the path of the index.

Method `remove()`**:** Removes objects from the index by their IDs.

Usage:

```
NgtIndex$remove(ids)
```

Arguments:

`ids` IDs of the objects to be removed.

Returns: integers; `ids` is returned invisibly as is.

Method `refine_anng()`**:** Refines the index with the specified parameters.

Usage:

```
NgtIndex$refine_anng(
  epsilon,
  accuracy,
  num_edges,
  num_edges_for_search,
  batch_size
)
```

Arguments:

`epsilon` epsilon which defines the explored range for the graph

`accuracy` expected accuracy.

`num_edges` number of edges for each node to explore the graph

`num_edges_for_search` number of edges to search.

`batch_size` batch size.

Method `get_object()`**:** Gets objects from the index by their IDs.

Usage:

```
NgtIndex$get_object(ids)
```

Arguments:

`ids` IDs of the objects to be retrieved.

Returns: tibble.

Method `reset_defaults()`**:** Resets the default parameters of the index.

Usage:

```
NgtIndex$reset_defaults(
  num_of_search_objects,
  search_radius,
  epsilon,
  edge_size,
  expected_accuracy
)
```

Arguments:

`num_of_search_objects` number of search objects.

`search_radius` search radius.

`epsilon` epsilon.

`edge_size` edge size.

`expected_accuracy` expected accuracy.

Method `get_info()`**:** Gets information of the index.

Usage:

```
NgtIndex$get_info()
```

Returns: named numeric vector

Method `build_index()`**:** Builds the search index.

Usage:

```
NgtIndex$build_index(num_threads = 1L, target_size_of_graph = 0L)
```

Arguments:

`num_threads` number of threads to be used for building a search index.

`target_size_of_graph` target size of the graph.

Method `batch_insert()`**:** Inserts data into the index and build a search index.

Usage:

```
NgtIndex$batch_insert(data, num_threads = 1L)
```

Arguments:

`data` data to be inserted.

`num_threads` number of threads to be used for building the index.

Returns: the IDs of the inserted objects are returned invisibly.

Method `insert()`**:** Inserts a vector to the index. To search with the index, you need to call `build_index` after call this method.

Usage:

```
NgtIndex$insert(vec)
```

Arguments:

`vec` vector to be inserted.

Returns: the ID of the inserted object is returned invisibly.

Method `export_index()`**:** Exports an index to a file.

Usage:

```
NgtIndex$export_index(path)
```

Arguments:

`path` path to save the index.

Method `import_index()`**:** Imports an index from a file.

Usage:

```
NgtIndex$import_index(path)
```

Arguments:

`path` path to load the index.

Method `close()`**:** Closes the index.

Usage:

```
NgtIndex$close()
```

Index