Package: tidyvec (via r-universe)

May 14, 2026

Title Lightweight Vector Embeddings for the Tidyverse

Version 0.1.0

Description A lightweight vector database for storing and querying embeddings within the tidyverse framework. Supports multimodal (text and image) embeddings, nearest neighbor search, and visualization, all with a tidyverse-friendly API.

License MIT + file LICENSE

URL https://flmnh-ai.github.io/tidyvec/, https://github.com/flmnh-ai/tidyvec

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Imports dplyr, tibble, purrr, rlang, ggplot2

Suggests reticulate, text2vec, Rtsne, umap, progress, qs, knitr, rmarkdown, magick, ggimage, tidyverse

Config/testthat/edition 3

VignetteBuilder knitr

Repository https://flmnh-ai.r-universe.dev

Date/Publication 2025-10-16 00:36:01 UTC

RemoteUrl https://github.com/flmnh-ai/tidyvec

RemoteRef HEAD

RemoteSha 36d0472f1f994999b64b2748d1cf1ee79500534e

Contents

`%~%` *Similarity operator*

Description

Similarity operator

Usage

```
a %~% b
```

Arguments

`a`	First vector or tidyvec object
`b`	Second vector or tidyvec object
`method`	Similarity method to use ("cosine", "euclidean", or "dot")

Value

Similarity score or tidyvec object with similarity scores

`cluster_embeddings` *Cluster embeddings using k-means*

Description

Cluster embeddings using k-means

Usage

```
cluster_embeddings(x, n_clusters = 5, cluster_column = "cluster")
```

Arguments

`x`	A tidyvec object
`n_clusters`	Number of clusters
`cluster_column`	Name for the cluster assignment column (default: "cluster")

Value

tidyvec object with cluster assignments added

`embed` *Compute embeddings for items in a tidyvec collection*

Description

Compute embeddings for items in a tidyvec collection

Usage

```
embed(x, content_column, embedding_fn = NULL, force = FALSE, ...)
```

Arguments

`x`	A tidyvec object
`content_column`	Column containing content to embed
`embedding_fn`	Embedding function to use (overrides collection's function)
`force`	Whether to overwrite existing embeddings
`...`	Additional arguments passed to the embedding function

Value

Updated tidyvec object with embeddings

`embedder_hf` *Create a HuggingFace embedding function*

Description

Create a HuggingFace embedding function

Usage

```
embedder_hf(
  model_name,
  modality = c("multimodal", "text", "image"),
  device = "cpu",
  cache_dir = NULL
)
```

Arguments

model_name	Name of the HuggingFace model
modality	Type of model ("text", "image", or "multimodal")
device	Device to use ("cpu", "cuda", or "mps")
cache_dir	Optional directory for caching model files

Value

An embedding function

---

embedder_tfidf *Create a simple TF-IDF embedding function for text*

Description

Create a simple TF-IDF embedding function for text

Usage

```
embedder_tfidf(corpus, min_freq = 2)
```

Arguments

corpus	Text corpus to build vocabulary
min_freq	Minimum term frequency

Value

An embedding function

---

inspect_collection *Print details about a tidyvec collection*

Description

Print details about a tidyvec collection

Usage

```
inspect_collection(x)
```

Arguments

x	A tidyvec object

Value

Invisibly returns the input object

nearest *Find nearest neighbors for a query in a tidyvec collection*

Description

Find nearest neighbors for a query in a tidyvec collection

Usage

```
nearest(
  x,
  query,
  n = 5,
  as_embedding = FALSE,
  method = c("cosine", "euclidean", "dot"),
  min_score = 0,
  keyword_weight = 0,
  keyword_column = NULL
)
```

Arguments

x	A tidyvec object
query	Query item (content or embedding)
n	Number of results to return
as_embedding	Whether the query is already an embedding vector
method	Similarity method ("cosine", "euclidean", "dot")
min_score	Minimum similarity score
keyword_weight	Weight for keyword matching (0-1, default 0 for pure vector search)
keyword_column	Column to search for keywords (required if keyword_weight > 0)

Value

Filtered tidyvec object with similarity scores

read_vec *Read a tidyvec collection from disk*

Description

Read a tidyvec collection from disk

Usage

```
read_vec(file)
```

Arguments

file	Path to tidyvec collection file

Value

A tidyvec object

vec *Creates a vector collection from a data frame or tibble*

Description

Creates a vector collection from a data frame or tibble

Usage

```
vec(x, embedding_column = "embedding", embedding_fn = NULL)
```

Arguments

x	A data frame or tibble
embedding_column	Name of column containing embeddings (or to be created)
embedding_fn	Function to generate embeddings (optional)

Value

A tidyvec object

`viz_embeddings` *Visualize embedding space using dimensionality reduction*

Description

Visualize embedding space using dimensionality reduction

Usage

```
viz_embeddings(
  x,
  method = c("umap", "tsne", "pca"),
  labels = NULL,
  color = NULL,
  n_neighbors = 15,
  perplexity = 30,
  images_column = NULL,
  ...
)
```

Arguments

`x`	A tidyvec object
`method`	Dimensionality reduction method ("tsne", "umap", "pca")
`labels`	Column to use for point labels
`color`	Column to use for point colors
`n_neighbors`	Number of neighbors (for UMAP)
`perplexity`	Perplexity parameter (for t-SNE)
`images_column`	Optional column containing image paths to use instead of points
`...`	Additional arguments passed to the plotting function

Value

A ggplot object

viz_images *Visualize images in a tidyvec collection using magick*

Description

Visualize images in a tidyvec collection using magick

Usage

```
viz_images(
  x,
  path_column,
  n = NULL,
  ncol = 3,
  width = 200,
  include_similarity = TRUE,
  label_columns = NULL
)
```

Arguments

x	A tidyvec object containing images
path_column	Column containing image paths or URLs
n	Number of images to display (default: all)
ncol	Number of columns in the grid (default: 3)
width	Width of each image in pixels (default: 200)
include_similarity	Whether to show similarity scores if available (default: TRUE)
label_columns	Additional columns to use as labels

Value

A magick image object

write_vec *Save a tidyvec collection to disk*

Description

Save a tidyvec collection to disk

Usage

```
write_vec(x, file)
```

Arguments

x	A tidyvec object
file	Path to save file (recommended: .qs extension)

Value

Invisibly returns the input object

Index