Semantic Search

Definition: Semantic search is a search methodology that goes beyond simple keyword matching to understand the meaning of a user's query and return relevant results.

Example: When a user searches for "planets in the solar system," the search returns information about planets like "Jupiter" and "Mars."

Related Keywords: Natural Language Processing (NLP), Search Algorithms, Data Mining

Embedding

Definition: Embedding refers to the process of transforming textual data, such as words or sentences, into low-dimensional continuous vectors. This allows computers to understand and process text effectively.

Example: The word "apple" might be represented as a vector like [0.65, -0.23, 0.17].

Related Keywords: Natural Language Processing (NLP), Vectorization, Deep Learning

Token

Definition: A token is a smaller unit obtained by splitting a piece of text. This unit can be a word, a sentence, or a phrase.

Example: The sentence "I go to school" can be tokenized into "I," "go," and "to school."

Related Keywords: Tokenization, Natural Language Processing (NLP), Parsing

Tokenizer

Definition: A tokenizer is a tool that splits textual data into tokens. It is used in natural language processing to preprocess data.

Example: The sentence "I love programming." is split into ["I", "love", "programming", "."].

Related Keywords: Tokenization, Natural Language Processing (NLP), Parsing

VectorStore

Definition: A VectorStore is a system for storing data in vector format. It is used for tasks like search, classification, and other data analysis operations.

Example: Storing word embedding vectors in a database for fast retrieval.

Related Keywords: Embedding, Database, Vectorization

SQL

Definition: SQL (Structured Query Language) is a programming language used to manage data in databases. It allows operations like querying, updating, inserting, and deleting data.

Example: SELECT \* FROM users WHERE age > 18; retrieves information about users older than 18.

Related Keywords: Database, Query, Data Management

CSV

Definition: CSV (Comma-Separated Values) is a file format used to store data, where each value is separated by a comma. It is commonly used for saving and exchanging tabular data.

Example: A CSV file with headers "Name," "Age," and "Occupation" might include data like "Hong Gil-dong, 30, Developer."

Related Keywords: Data Format, File Handling, Data Exchange

JSON

Definition: JSON (JavaScript Object Notation) is a lightweight data exchange format that uses human-readable text to represent data objects. It is widely used for data communication between systems.

Example: {"Name": "Hong Gil-dong", "Age": 30, "Occupation": "Developer"} is an example of JSON data.

Related Keywords: Data Exchange, Web Development, API

Transformer

Definition: A Transformer is a type of deep learning model used in natural language processing for tasks like translation, summarization, and text generation. It is based on the Attention mechanism.

Example: Google Translate uses Transformer models to perform translations between various languages.

Related Keywords: Deep Learning, Natural Language Processing (NLP), Attention

HuggingFace

Definition: HuggingFace is a library offering a variety of pre-trained models and tools for natural language processing, enabling researchers and developers to easily perform NLP tasks.

Example: HuggingFace's Transformers library can be used for tasks such as sentiment analysis and text generation.

Related Keywords: Natural Language Processing (NLP), Deep Learning, Library

Digital Transformation

Definition: Digital transformation is the process of leveraging technology to innovate and enhance a company’s services, culture, and operations. It focuses on improving business models and increasing competitiveness through digital technologies.

Example: A company adopting cloud computing to revolutionize data storage and processing is an example of digital transformation.

Related Keywords: Innovation, Technology, Business Model

Crawling

Definition: Crawling is the process of automatically visiting web pages to collect data. It is commonly used for search engine optimization and data analysis.

Example: Google’s search engine crawls websites on the internet to collect and index content.

Related Keywords: Data Collection, Web Scraping, Search Engine

Word2Vec

Definition: Word2Vec is a natural language processing technique that maps words to a vector space, representing semantic relationships between words. It generates vectors based on the contextual similarity of words.

Example: In a Word2Vec model, "king" and "queen" are represented as vectors that are close to each other in the vector space.

Related Keywords: Natural Language Processing (NLP), Embedding, Semantic Similarity

LLM (Large Language Model)

Definition: LLM refers to large-scale language models trained on extensive text datasets. These models are used for a variety of natural language understanding and generation tasks.

Example: OpenAI’s GPT series is a prominent example of large language models.

Related Keywords: Natural Language Processing (NLP), Deep Learning, Text Generation

FAISS (Facebook AI Similarity Search)

Definition: FAISS is a high-speed similarity search library developed by Facebook, designed to efficiently search for similar vectors in large vector datasets.

Example: FAISS can be used to quickly find similar images among millions of image vectors.

Related Keywords: Vector Search, Machine Learning, Database Optimization

Open Source

Definition: Open source refers to software whose source code is openly available for anyone to use, modify, and distribute freely. It plays a significant role in fostering collaboration and innovation.

Example: The Linux operating system is a notable open-source project.

Related Keywords: Software Development, Community, Technical Collaboration

Structured Data

Definition: Structured data is data that is organized according to a predefined format or schema, making it easy to search and analyze in systems like databases and spreadsheets.

Example: A customer information table stored in a relational database is an example of structured data.

Related Keywords: Database, Data Analysis, Data Modeling

Parser

Definition: A parser is a tool that analyzes input data (e.g., strings, files) and converts it into a structured format. It is commonly used in syntax analysis for programming languages or processing file data.

Example: Parsing an HTML document to create the DOM structure of a web page is an example of parsing.

Related Keywords: Syntax Analysis, Compiler, Data Processing

TF-IDF (Term Frequency-Inverse Document Frequency)

Definition: TF-IDF is a statistical measure used to evaluate the importance of a word in a document relative to a collection of documents. It considers both the frequency of the word in the document and its rarity across all documents.

Example: Words that appear frequently in a document but rarely in the corpus will have a high TF-IDF value.

Related Keywords: Natural Language Processing (NLP), Information Retrieval, Data Mining

Deep Learning

Definition: Deep learning is a subset of machine learning that uses artificial neural networks to solve complex problems. It focuses on learning high-level representations from data.

Example: Deep learning models are applied in tasks like image recognition, speech recognition, and natural language processing.

Related Keywords: Artificial Neural Networks, Machine Learning, Data Analysis

Schema

Definition: A schema defines the structure of a database or file, providing a blueprint for how data is stored and organized.

Example: A relational database table schema specifies column names, data types, and key constraints.

Related Keywords: Database, Data Modeling, Data Management

DataFrame

Definition: A DataFrame is a tabular data structure composed of rows and columns, commonly used for data analysis and processing.

Example: In the pandas library, a DataFrame can contain columns of various data types, making data manipulation and analysis straightforward.

Related Keywords: Data Analysis, pandas, Data Processing

Attention Mechanism

Definition: The Attention mechanism is a deep learning technique that allows models to focus on the most relevant parts of input data. It is mainly used in sequence data, such as text or time series.

Example: In translation models, the Attention mechanism helps the model focus on the crucial parts of the input sentence to produce accurate translations.

Related Keywords: Deep Learning, Natural Language Processing (NLP), Sequence Modeling

Pandas

Definition: Pandas is a Python library that provides tools for data analysis and manipulation. It enables efficient execution of data-related tasks.

Example: Using pandas, you can read a CSV file, clean the data, and perform various analyses.

Related Keywords: Data Analysis, Python, Data Processing

GPT (Generative Pretrained Transformer)

Definition: GPT is a generative language model pretrained on large datasets, designed for various text-based tasks. It generates natural language based on the provided input.

Example: A chatbot that generates detailed answers to user queries can leverage the GPT model.

Related Keywords: Natural Language Processing (NLP), Text Generation, Deep Learning

InstructGPT

Definition: InstructGPT is a GPT model optimized to perform specific tasks based on user instructions. It is designed to generate more accurate and contextually relevant results.

Example: If a user provides an instruction like "draft an email," InstructGPT generates an email based on the provided context.

Related Keywords: Artificial Intelligence (AI), Natural Language Understanding (NLU), Instruction-Based Processing

Keyword Search

Definition: Keyword search is the process of finding information based on user-entered keywords. It is the fundamental search method used in most search engines and database systems.

Example: When a user searches for "coffee shop Seoul," the search returns a list of relevant coffee shops in Seoul.

Related Keywords: Search Engine, Data Search, Information Retrieval

Page Rank

Definition: Page Rank is an algorithm that evaluates the importance of web pages, primarily used to determine the ranking of search engine results. It analyzes the link structure between web pages to assess their importance.

Example: Google’s search engine uses the Page Rank algorithm to determine the order of its search results.

Related Keywords: Search Engine Optimization (SEO), Web Analytics, Link Analysis

Data Mining

Definition: Data mining is the process of extracting useful information from large datasets by utilizing techniques such as statistics, machine learning, and pattern recognition.

Example: Analyzing customer purchase data to develop sales strategies in retail is a common example of data mining.

Related Keywords: Big Data, Pattern Recognition, Predictive Analytics

Multimodal

Definition: Multimodal refers to technologies that combine and process multiple types of data modalities, such as text, images, and audio. It is used to extract or predict richer and more accurate information through the interaction of different data formats.

Example: A system that analyzes both images and descriptive text to perform more accurate image classification is an example of multimodal technology.

Related Keywords: Data Fusion, Artificial Intelligence (AI), Deep Learning