RAG (Retrieval-Augmented Generation) Results

Question: java abstract classes

Search Method: Cosine

Embedding Model: text-embedding-3-small

Embedding Endpoint: https://aihubmytesthub7294089329.openai.azure.com/

Generation Model: Llama-3.3-70B-Instruct

Generation Endpoint: https://aihubmytesthub7294089329.services.ai.azure.com/models/chat/completions?api-version=2024-05-01-preview

Generated: 2025-07-24 15:33:58

Generated Answer

Sorry, I couldn't generate an answer due to an error.

Performance Metrics

Search Time: 2.0223 seconds

Generation Time: 0.6320 seconds

Total Time: 2.6543 seconds

Source Documents

Source 1

Document: Java_9.pdf

Chunk: 5

Relevance Score: 0.5144

Content Preview: Oct. 2017 https://www.journaldev.com/12850/java -9-private -methods - interfaces

Source 2

Document: Curs2_3_SDJava.pdf

Chunk: 0

Relevance Score: 0.4680

Content Preview: Programare Java (2) prof.dr.ing. ■tefan -Gheorghe Pentiuc Modificator CLAS■ METOD■ VARIABIL■ public acces permis tuturor claselorpoate fi apelat■ din orice clas■ poate fi consultat■ / modificat■ din ...

Source 3

Document: Curs2_3_SDJava.pdf

Chunk: 10

Relevance Score: 0.4382

Content Preview: ecated. The finalization mechanism is inherently problematic. Class getClass () Returns the runtime class of this Object. int hashCode () Returns a hash code value for the object. void notify () ...

Source 4

Document: Java_9.pdf

Chunk: 4

Relevance Score: 0.4372

Content Preview: erface, enum , obiecte , instructiuni etc). jshell > int counter = 0 counter ==> 0 jshell > counter++ $6 ==> 0 jshell > counter counter ==> 1 BIBLIOGRAFIE •What's New in Oracle JDK ...

Source 5

Document: ExpresiiLambda.pdf

Chunk: 6

Relevance Score: 0.4268

Content Preview: dia, 2014 •Java 8 in Action , by Raoul -Gabriel Urma , Mario Fusco , Alan Mycroft , 2014 Manning Publications •Building Java Programs , Chapter 19, Pearson 2016

RAG Process Summary

This RAG (Retrieval-Augmented Generation) search performed the following steps:

1. **Retrieval:** Found 5 most relevant document chunks using cosine similarity with text-embedding-3-small model from https://aihubmytesthub7294089329.openai.azure.com/

2. **Augmentation:** Combined the retrieved content as context

3. **Generation:** Used Llama-3.3-70B-Instruct model from https://aihubmytesthub7294089329.services. ai.azure.com/models/chat/completions?api-version=2024-05-01-preview to generate a comprehensive answer based on the context

RAG provides more accurate and contextual answers by grounding the AI response in your specific document content. Using separate models and endpoints for embedding and generation allows for optimized performance and resource allocation.