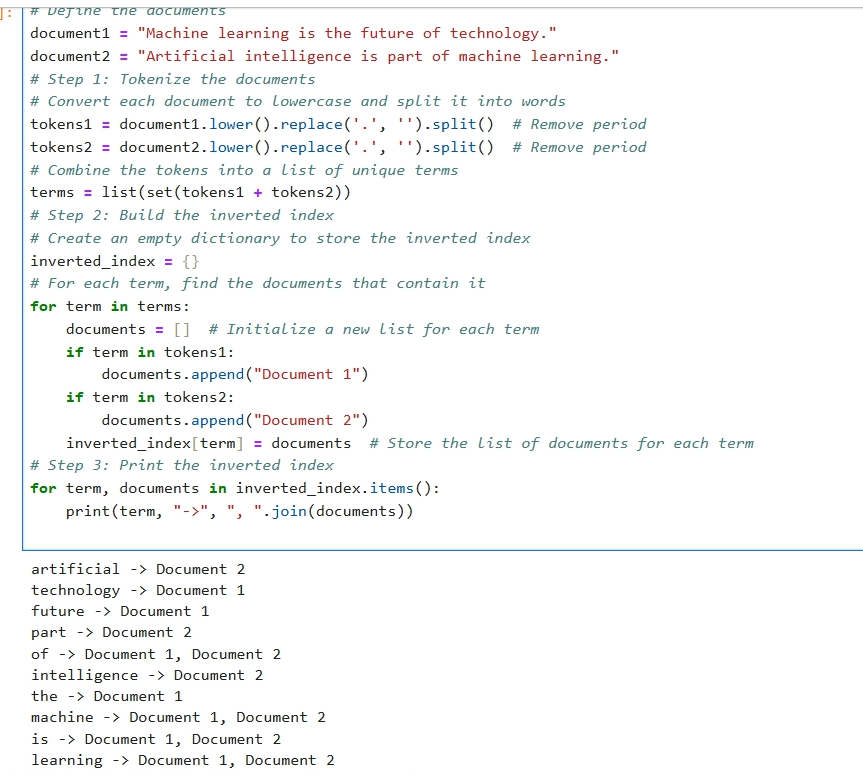
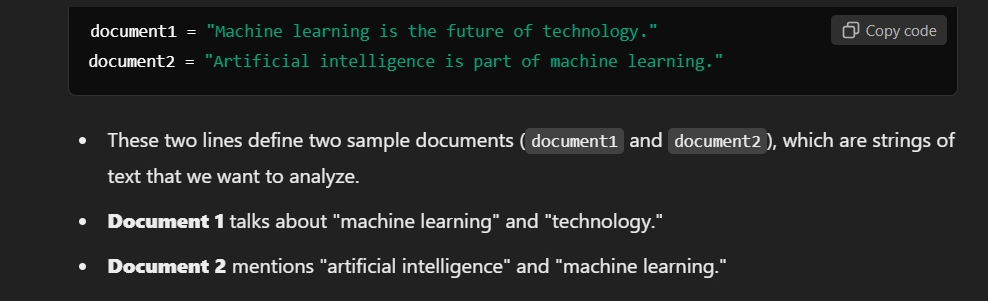
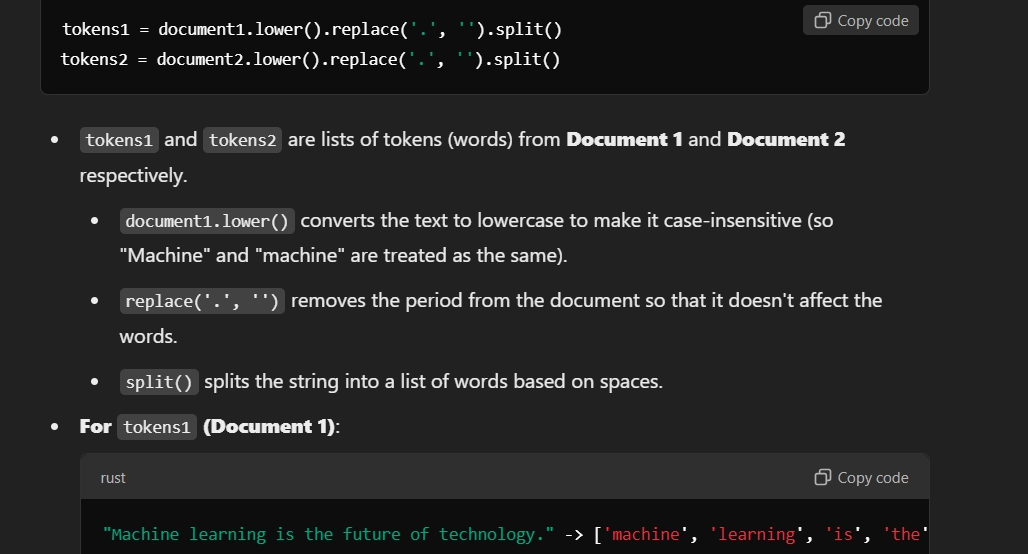
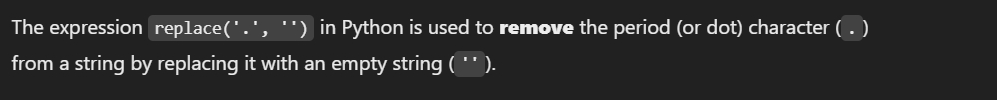
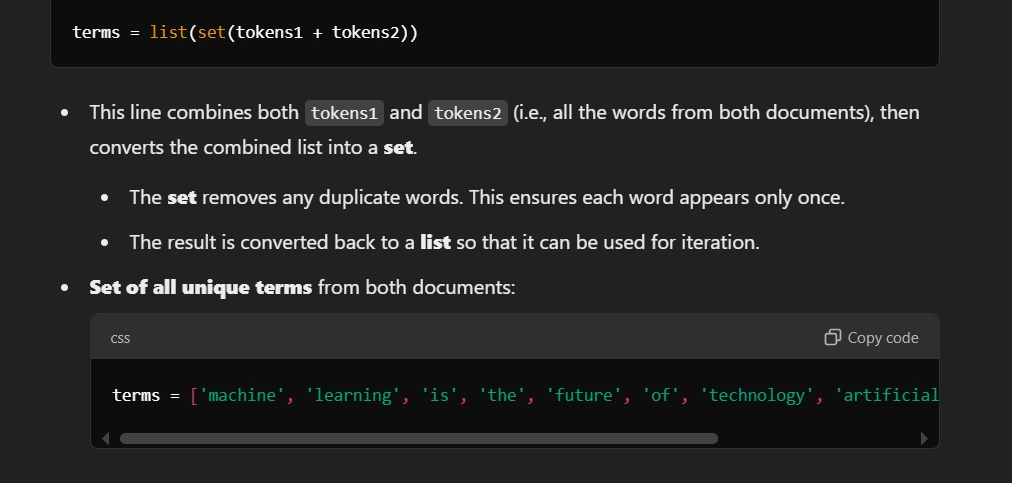
Implement a program for retrieval of documents using inverted files.

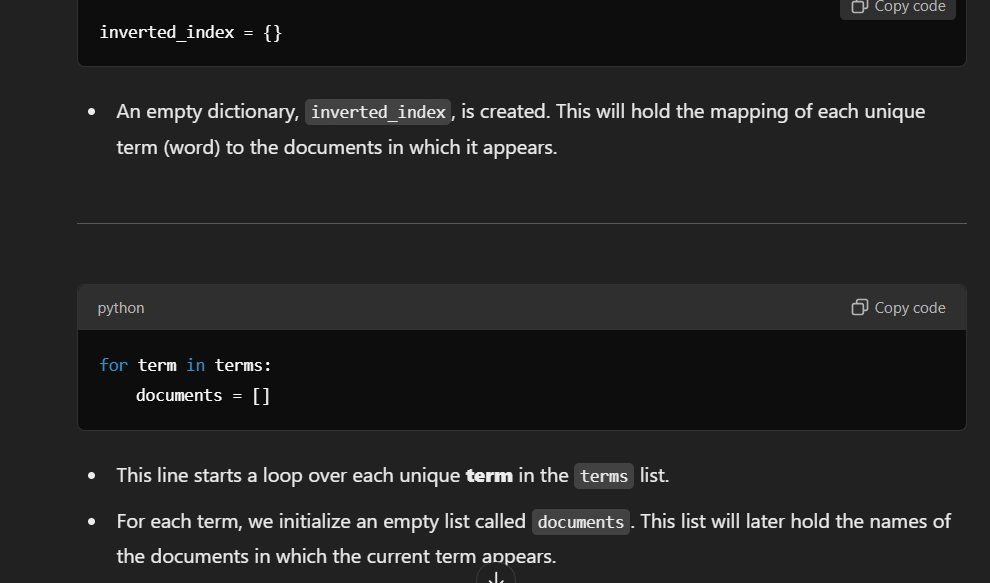


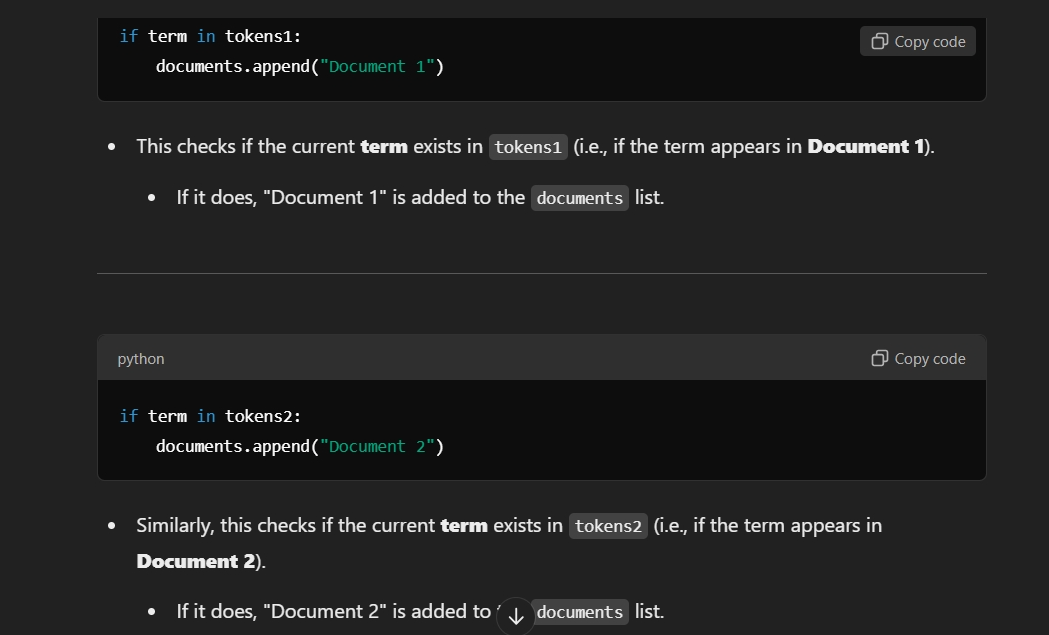


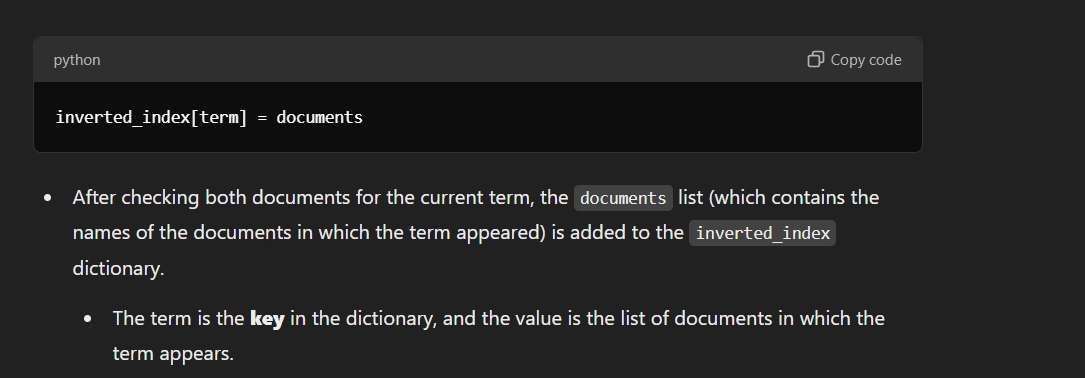


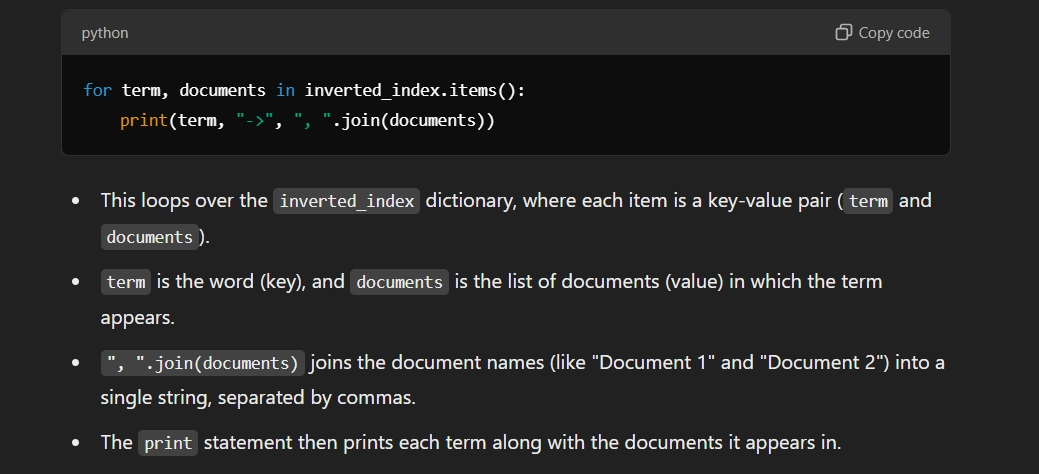


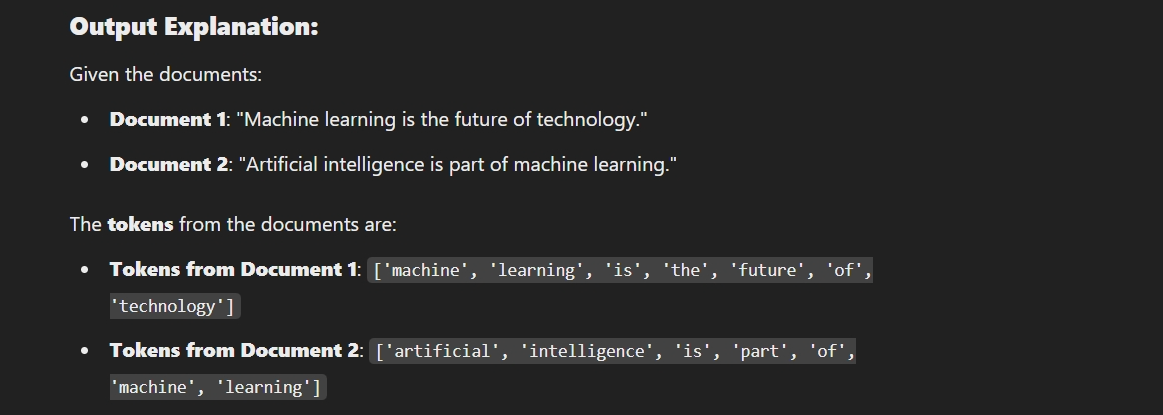


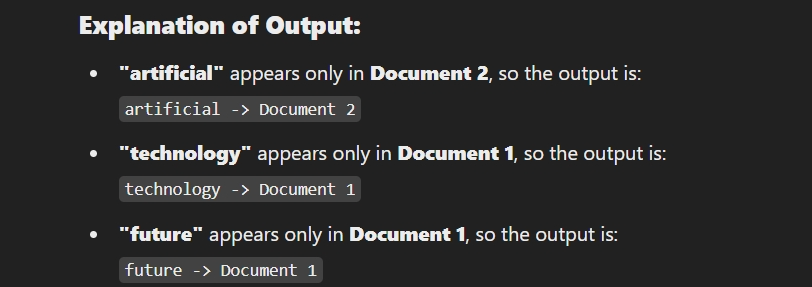


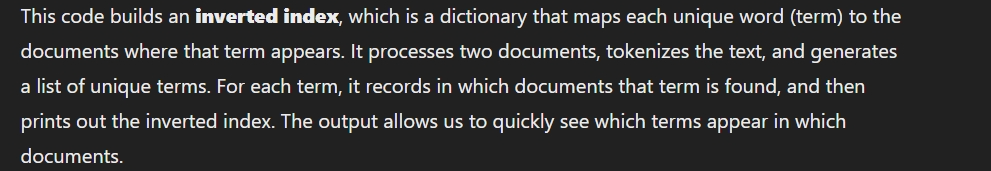


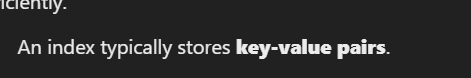


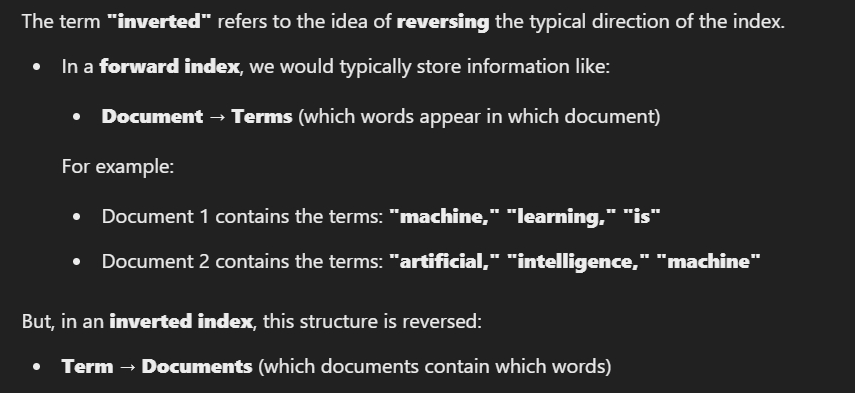












Its main parts are:

1. \*\*Words or Terms\*\*: The individual words from the documents, each listed once.

2. \*\*Document IDs\*\*: Unique numbers for each document to identify them.

3. \*\*Posting Lists\*\*: Lists of document IDs where each word appears.

4. \*\*Word Count (Optional)\*\*: Sometimes, the index also shows how many times each word appears in each document.

5. \*\*Word Position (Optional)\*\*: Sometimes, the index also notes exactly where each word appears in each document, which helps with finding exact phrases.

### Types of Inverted Indexes

1. \*\*Token-Based Inverted Index\*\*: Maps each word to its occurrences in a document

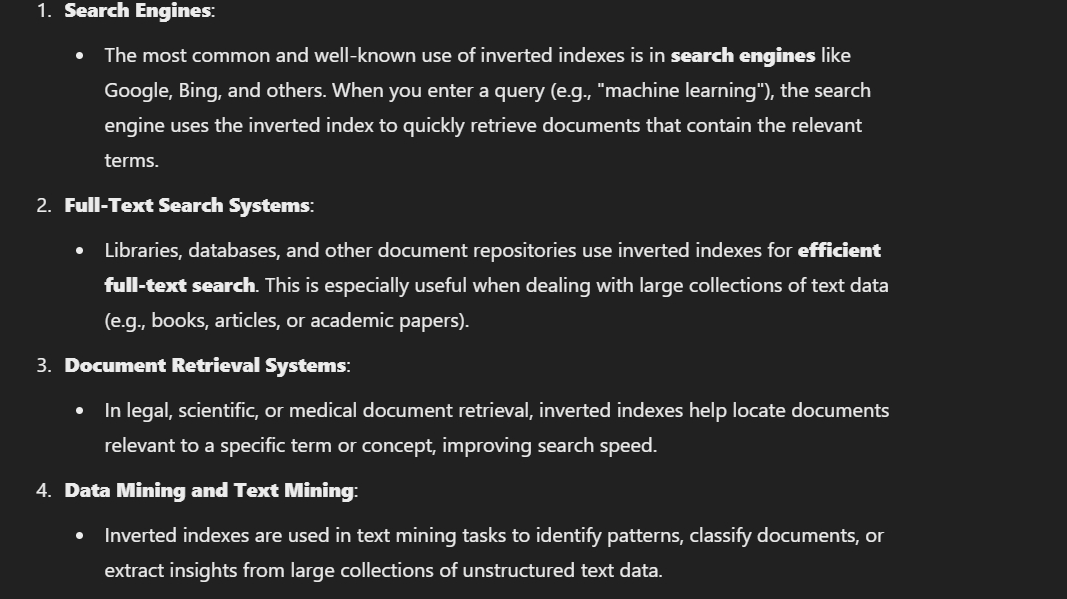
2. \*\*Field-Based Inverted Index\*\*:

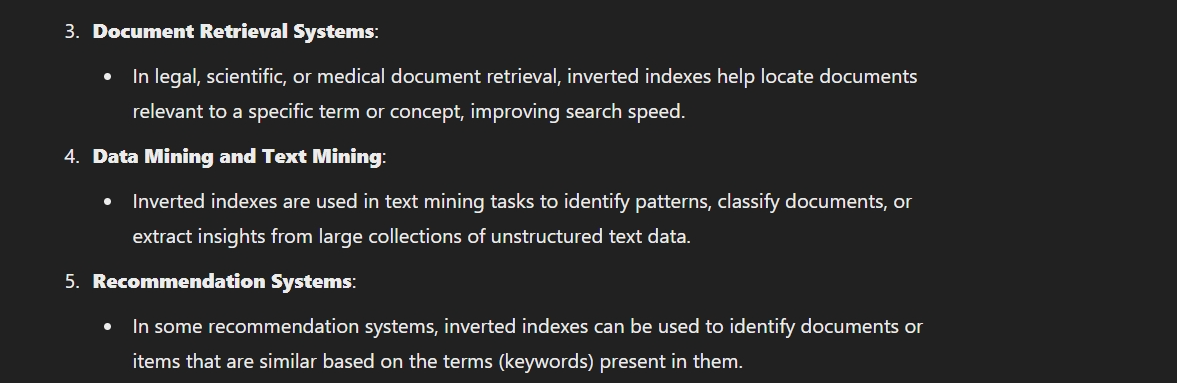
- Separates the index by different fields within a document (e.g., title, body, author

3. \*\*Positional Inverted Index\*\*:

- In addition to mapping tokens to documents, it records the positions of each token within the document.Example:The word "machine" might appear at positions 1 and 7 in Document 1.

4. \*\*Bi-Level Inverted Index\*\*:Combines token-level and positional indexes to offer enhanced search capabilities with both high-speed term lookup and position-based retrieval.





Natural Language Processing (NLP)\*\*: Supports language processing tasks like finding word patterns, synonyms, or phrase matching.

5. \*\*Spam Detection\*\*: Helps detect patterns in text that indicate spam by analyzing term frequency and distribution across documents.

6. \*\*Plagiarism Detection\*\*: Quickly identifies duplicated content across large document collections by finding exact or near-exact matches in word sequences.