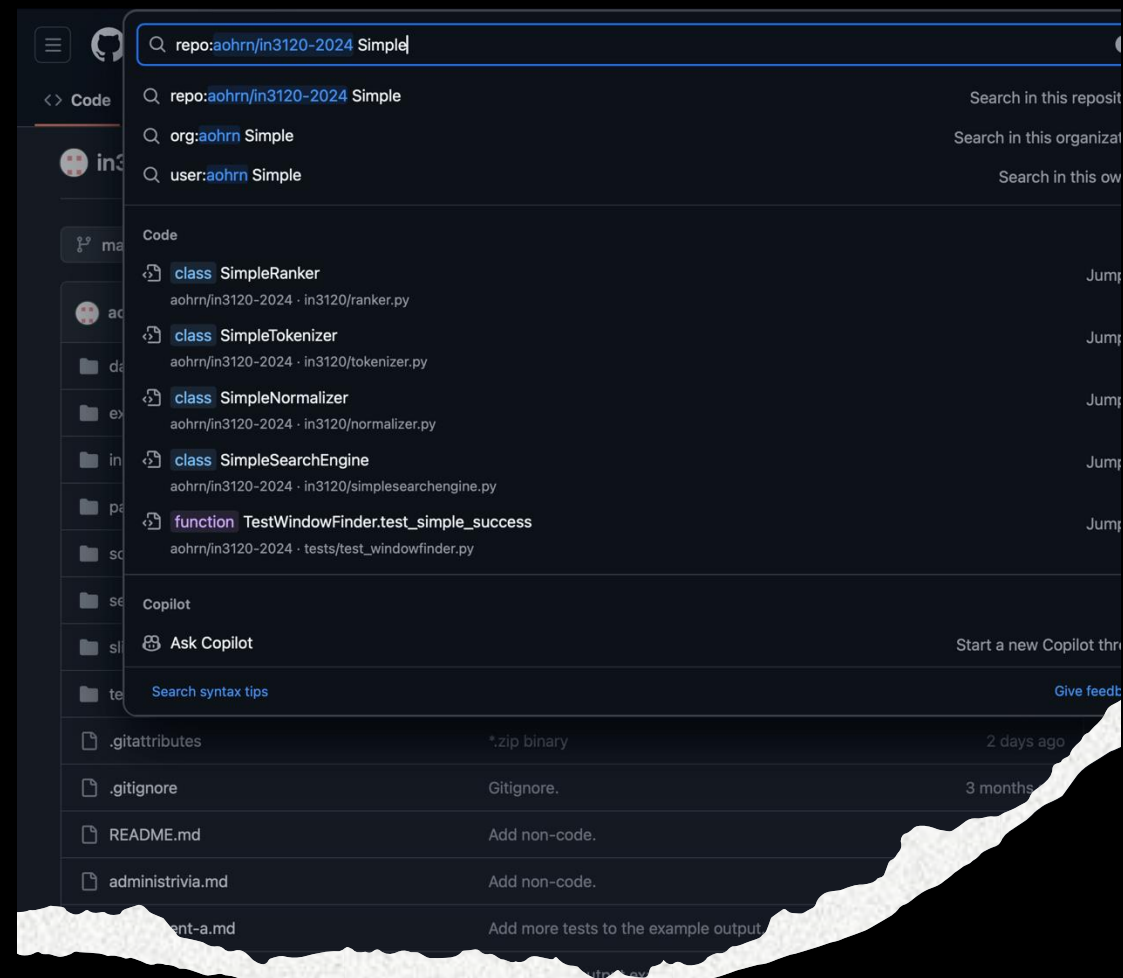


Tokenization og søk i code



Fritekstsøk vs Symbolsøk

Find in Files 101+ matches in 21+ files

File mask: *Sheet*

Simple

In Project Module Directory Scope

class SimpleSearchEngine: simplesearchengine.py 14

Realizes a simple query evaluator that efficiently performs N-of-M mat simplesearchengine.py 16

as simple-minded runtime-checkable protocols that c pip/_vendor/typing_extensions.py 787

This allows a simple-minded structural ch python3.12/site-packages/typing_extensions.py 727

"""A simple typed namespace. At runtime it is python3.12/site-packages/typing_extensions.py 1040

Simple [REPLs](https://en.wikipedia.org/wiki/Read%E2%80%93eval%E2%80%93print loop) README.md 79

The purpose of this assignment is to build a simple in-memory inverted index ar assignment-a.md 5

simplesearchengine.py in3120

```
11 from typing import TypedDict, TypedDict, TypedDict
12
13
14 class SimpleSearchEngine:
15     """
16     Realizes a simple query evaluator that efficiently performs
17     I.e., if the query contains M unique query terms, each docum
18     least N of these m terms. For example, 2-of-3 matching over
```

Open results in new tab Open in Find Window

All Classes Files Symbols Actions Text

Project Files

Simple

- simple(np.ndarray) numpy.core.tests.test_umath < Python 3.9 (assignments) >
- simple gh/.venv/lib/python3.12/.../numpy/typing/tests/data/pass < Python 3.9 (assignments) >
- simple setuptools_vendor.importlib_resources < Python 3.9 (assignments) >
- SimpleNormalizer(Normalizer) in3120.normalizer in3120-2024
- SimpleFrozenList(list) confection.util < Python 3.9 (assignments) >
- SimpleRanker(Ranker) in3120.ranker in3120-2024
- SimpleTokenizer(Tokenizer) in3120.tokenizer in3120-2024
- SimpleKey(object) srsly.ruamel.yaml.scanner < Python 3.9 (assignments) >
- SimpleLoader setuptools_vendor.importlib_resources.tests.test_custom < Python 3.9 (assignment...) >
- SimplePath(Protocol) setuptools_vendor.importlib_metadata._meta < Python 3.9 (assignments) >
- SimpleReader(abc.ABC) setuptools_vendor.importlib_resources.simple < Python 3.9 (assignments) >
- SimpleTemplate(BaseTemplate) pypm.util.bottle < Python 3.9 (assignments) >
- SimpleSearchEngine in3120.simplesearchengine in3120-2024
- SimpleFrozenDict(dict) confection.util < Python 3.9 (assignments) >
- SimpleFrozenDict(dict) spacy.util < Python 3.9 (assignments) >
- SimpleFrozenDict(dict) weasel.util.frozen < Python 3.9 (assignments) >
- SimpleFrozenList(list) spacy.util < Python 3.9 (assignments) >
- SimpleFrozenList(list) weasel.util.frozen < Python 3.9 (assignments) >
- SimpleScrapingLocator(Locator) pip._vendor.distlib.locators < Python 3.9 (assignments) >
- SimpleSerSchema(TypedDict, total=False) pydantic_core.core_schema < Python 3.9 (assignments) >
- SimpleSubClass(VerySimpleSubClass) numpy.lib.tests.test_stride_tricks < Python 3.9 (assignment...) >
- SimpleWheelCache(Cache) pip._internal.cache < Python 3.9 (assignments) >
- simplesearchengine in3120 in3120-2024
- SimpleTextIOWrapper(ObjectWrapper) tqdm.util < Python 3.9 (assignments) >
- simple_py3 gh/.venv/lib/python3.12/.../numpy/typing/tests/data/pass < Python 3.9 (assignments) >
- simpler_bdist_wheel(bdist_wheel) setuptools.tests.test_bdist_wheel < Python 3.9 (assignments) >

~/ui/IN4120-searchtech/gh/in3120/normalizer.py

Open In Right Split

Samme på GitHub

The screenshot shows a GitHub search results page for the repository 'aohrn/in3120-2024'. The search bar at the top contains the text 'repo:aohrn/in3120-2024 Simple'. Below the search bar, there are three search filters: 'Search in this repository', 'Search in this organization', and 'Search in this owner'. The search results are displayed under the 'Code' tab, showing a list of files and functions. The files listed are 'SimpleRanker', 'SimpleTokenizer', 'SimpleNormalizer', and 'SimpleSearchEngine', all located in the 'in3120' directory. The function 'TestWindowFinder.test_simple_success' is also listed, located in the 'tests' directory. Each file entry includes a 'Jump to' link. The repository page also shows the repository name 'in3120-2024', the owner 'aohrn', and the repository description: 'anion git repository for the 2024 nester.' The repository has 7 forks and 8 stars.

repo:aohrn/in3120-2024 Simple

Search in this repository

Search in this organization

Search in this owner

Code

- `class` SimpleRanker
aohrn/in3120-2024 · in3120/ranker.py
Jump to
- `class` SimpleTokenizer
aohrn/in3120-2024 · in3120/tokenizer.py
Jump to
- `class` SimpleNormalizer
aohrn/in3120-2024 · in3120/normalizer.py
Jump to
- `class` SimpleSearchEngine
aohrn/in3120-2024 · in3120/simplesearchengine.py
Jump to
- `function` TestWindowFinder.test_simple_success
aohrn/in3120-2024 · tests/test_windowfinder.py
Jump to

Copilot

Supernaivt eksempel i in3120-2024 1/4

```
from in3120 import Tokenizer
from typing import Iterator, Tuple
import tokenize
from io import StringIO

class PythonTokenizer(Tokenizer): 3 usages
    """
    A simple tokenizer for Python files that can identify function and variable names
    """

    def __init__(self):
        pass

    def spans(self, buffer: str) → Iterator[Tuple[int, int]]: ...

    def strings(self, buffer: str) → Iterator[str]:
        tokens = tokenize.generate_tokens(StringIO(buffer).readline)
        for token in tokens:
            if token.type == tokenize.NAME:
                yield token.string
```

Supernaivt eksempel i in3120-2024 2/4

```
files = [os.path.join(directory, file) for file in os.listdir(directory)
         if os.path.isfile(os.path.join(directory, file))]
for i, file in enumerate(files):
    with open(file, "r") as f:
        content = f.read()
        corpus.add_document(in3120.InMemoryDocument(i, fields: {
            "filename": file.split("/")[-1],
            "content": content
        }))

tokenizer = in3120.PythonTokenizer()
index = in3120.InMemoryInvertedIndex(corpus, fields: ["content"], self.__normalizer, tokenizer)

engine = in3120.SimpleSearchEngine(corpus, index)
ranker = in3120.SimpleRanker()
```

Supernaivt eksempel i in3120-2024 3/4

```
matches = list(  
    engine.evaluate(  
        query: "__tokenizer __normalizer __inverted_index category",  
        options: {"match_threshold": 0.5, "hit_count": 100},  
        ranker  
    ))  
  
for match in matches:  
    print(match)
```

Supernaivt eksempel i in3120-2024 4/4

```
{ 'score': 26.0, 'document': { 'document_id': 15, 'fields': { 'filename': 'naivebayesclassifier.py',  
{ 'score': 7.0, 'document': { 'document_id': 19, 'fields': { 'filename': 'simplesearchengine.py', 'co  
{ 'score': 7.0, 'document': { 'document_id': 16, 'fields': { 'filename': 'suffixarray.py', 'content':  
{ 'score': 6.0, 'document': { 'document_id': 11, 'fields': { 'filename': 'windowfinder.py', 'content'  
{ 'score': 6.0, 'document': { 'document_id': 1, 'fields': { 'filename': 'editsearchengine.py', 'conte  
{ 'score': 5.0, 'document': { 'document_id': 7, 'fields': { 'filename': 'similaritysearchengine.py',  
{ 'score': 5.0, 'document': { 'document_id': 3, 'fields': { 'filename': 'shinglegenerator.py', 'conte  
{ 'score': 2.0, 'document': { 'document_id': 17, 'fields': { 'filename': 'stringfinder.py', 'content'
```