

Contextual compression

One challenge with retrieval is that usually you don't know the specific queries your document storage system will face when you ingest data into the system. This means that the information most relevant to a query may be buried in a document with a lot of irrelevant text. Passing that full document through your application can lead to more expensive LLM calls and poorer responses.

Contextual compression is meant to fix this. The idea is simple: instead of immediately returning retrieved documents as-is, you can compress them using the context of the given query, so that only the relevant information is returned. "Compressing" here refers to both compressing the contents of an individual document and filtering out documents wholesale.

To use the Contextual Compression Retriever, you'll need:

- a base Retriever
- a Document Compressor

The Contextual Compression Retriever passes queries to the base Retriever, takes the initial documents and passes them through the Document Compressor. The Document Compressor takes a list of Documents and shortens it by reducing the contents of Documents or dropping Documents altogether.

Contextual compression



Get started

Helper function for printing docs

```
def pretty_print_docs(docs):  
    print(f"\n{'-' * 100}\n".join([f"Document {i+1}:\n\n" + d.page_content for i, d in enumerate(docs)]))
```

Using a vanilla vector store retriever

Let's start by initializing a simple vector store retriever and storing the 2023 State of the Union speech (in chunks). We can see that given an example question our retriever returns one or two relevant docs and a few irrelevant docs. And even the relevant docs have a lot of irrelevant information in them.

```
from langchain.text_splitter import CharacterTextSplitter
from langchain.embeddings import OpenAIEmbeddings
from langchain.document_loaders import TextLoader
from langchain.vectorstores import FAISS

documents = TextLoader('../.../state_of_the_union.txt').load()
text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
texts = text_splitter.split_documents(documents)
retriever = FAISS.from_documents(texts, OpenAIEmbeddings()).as_retriever()

docs = retriever.get_relevant_documents("What did the president say about Ketanji Brown Jackson")
pretty_print_docs(docs)
```

Document 1:

Tonight. I call on the Senate to: Pass the Freedom to Vote Act. Pass the John Lewis Voting Rights Act. And while you're at it, pass the Disclose Act so Americans can know who is funding our elections.

Tonight, I'd like to honor someone who has dedicated his life to serve this country: Justice Stephen Breyer—an Army veteran, Constitutional scholar, and retiring Justice of the United States Supreme Court. Justice Breyer, thank you for your service.

One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence.

Document 2:

A former top litigator in private practice. A former federal public defender. And from a family of public school educators and police officers. A consensus builder. Since she's been nominated, she's received a broad range of support—from the Fraternal Order of Police to former judges appointed by Democrats and Republicans.

And if we are to advance liberty and justice, we need to secure the Border and fix the immigration system.

We can do both. At our border, we've installed new technology like cutting-edge scanners to better detect drug smuggling.

We've set up joint patrols with Mexico and Guatemala to catch more human traffickers.

We're putting in place dedicated immigration judges so families fleeing persecution and violence can have their cases heard faster.

We're securing commitments and supporting partners in South and Central America to host more refugees and secure their own borders.

Document 3:

And for our LGBTQ+ Americans, let's finally get the bipartisan Equality Act to my desk. The onslaught of state laws targeting transgender Americans and their families is wrong.

As I said last year, especially to our younger transgender Americans, I will always have your back as your President, so you can be yourself and reach your God-given potential.

While it often appears that we never agree, that isn't true. I signed 80 bipartisan bills into law last year. From preventing government shutdowns to protecting Asian-Americans from still-too-common hate crimes to reforming military justice.

And soon, we'll strengthen the Violence Against Women Act that I first wrote three decades ago. It is important for us to show the nation that we can come together and do big things.

So tonight I'm offering a Unity Agenda for the Nation. Four big things we can do together.

First, beat the opioid epidemic.

Document 4:

Tonight, I'm announcing a crackdown on these companies overcharging American businesses and consumers.

And as Wall Street firms take over more nursing homes, quality in those homes has gone down and costs have gone up.

That ends on my watch.

Medicare is going to set higher standards for nursing homes and make sure your loved ones get the care they deserve and expect.

We'll also cut costs and keep the economy going strong by giving workers a fair shot, provide more training and apprenticeships, hire them based on their skills not degrees.

Let's pass the Paycheck Fairness Act and paid leave.

Raise the minimum wage to \$15 an hour and extend the Child Tax Credit, so no one has to raise a family in poverty.

Let's increase Pell Grants and increase our historic support of HBCUs, and invest in what Jill—our First Lady who teaches full-time—calls America's best-kept secret: community colleges.

Adding contextual compression with an LLMChainExtractor

Now let's wrap our base retriever with a `ContextualCompressionRetriever`. We'll add an `LLMChainExtractor`, which will iterate over the initially returned documents and extract from each only the content that is relevant to the query.

```
from langchain.llms import OpenAI
from langchain.retrievers import ContextualCompressionRetriever
from langchain.retrievers.document_compressors import LLMChainExtractor

llm = OpenAI(temperature=0)
compressor = LLMChainExtractor.from_llm(llm)
compression_retriever = ContextualCompressionRetriever(base_compressor=compressor, base_retriever=retriever)

compressed_docs = compression_retriever.get_relevant_documents("What did the president say about Ketanji Jackson Brown")
pretty_print_docs(compressed_docs)
```

Document 1:

"One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence."

Document 2:

"A former top litigator in private practice. A former federal public defender. And from a family of public school educators and police officers. A consensus builder. Since she's been nominated, she's received

a broad range of support—from the Fraternal Order of Police to former judges appointed by Democrats and Republicans."

More built-in compressors: filters

LLMChainFilter

The `LLMChainFilter` is slightly simpler but more robust compressor that uses an LLM chain to decide which of the initially retrieved documents to filter out and which ones to return, without manipulating the document contents.

```
from langchain.retrievers.document_compressors import LLMChainFilter

_filter = LLMChainFilter.from_llm(llm)
compression_retriever = ContextualCompressionRetriever(base_compressor=_filter, base_retriever=retriever)

compressed_docs = compression_retriever.get_relevant_documents("What did the president say about Ketanji Jackson Brown")
pretty_print_docs(compressed_docs)
```

Document 1:

Tonight. I call on the Senate to: Pass the Freedom to Vote Act. Pass the John Lewis Voting Rights Act. And while you're at it, pass the Disclose Act so Americans can know who is funding our elections.

Tonight, I'd like to honor someone who has dedicated his life to serve this country: Justice Stephen Breyer—an Army veteran, Constitutional scholar, and retiring Justice of the United States Supreme Court. Justice Breyer, thank you for your service.

One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence.

EmbeddingsFilter

Making an extra LLM call over each retrieved document is expensive and slow. The `EmbeddingsFilter` provides a cheaper and faster option by embedding the documents and query and only returning those documents which have sufficiently similar embeddings to the query.

```
from langchain.embeddings import OpenAIEmbeddings
from langchain.retrievers.document_compressors import EmbeddingsFilter

embeddings = OpenAIEmbeddings()
embeddings_filter = EmbeddingsFilter(embeddings=embeddings, similarity_threshold=0.76)
compression_retriever = ContextualCompressionRetriever(base_compressor=embeddings_filter,
base_retriever=retriever)

compressed_docs = compression_retriever.get_relevant_documents("What did the president say about Ketanji
Jackson Brown")
pretty_print_docs(compressed_docs)
```

Document 1:

Tonight. I call on the Senate to: Pass the Freedom to Vote Act. Pass the John Lewis Voting Rights Act. And while you're at it, pass the Disclose Act so Americans can know who is funding our elections.

Tonight, I'd like to honor someone who has dedicated his life to serve this country: Justice Stephen Breyer—an Army veteran, Constitutional scholar, and retiring Justice of the United States Supreme Court. Justice Breyer, thank you for your service.

One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence.

Document 2:

A former top litigator in private practice. A former federal public defender. And from a family of public school educators and police officers. A consensus builder. Since she's been nominated, she's received a broad range of support—from the Fraternal Order of Police to former judges appointed by Democrats and Republicans.

And if we are to advance liberty and justice, we need to secure the Border and fix the immigration system.

We can do both. At our border, we've installed new technology like cutting-edge scanners to better detect drug smuggling.

We've set up joint patrols with Mexico and Guatemala to catch more human traffickers.

We're putting in place dedicated immigration judges so families fleeing persecution and violence can have their cases heard faster.

We're securing commitments and supporting partners in South and Central America to host more refugees and secure their own borders.

Document 3:

And for our LGBTQ+ Americans, let's finally get the bipartisan Equality Act to my desk. The onslaught of

state laws targeting transgender Americans and their families is wrong.

As I said last year, especially to our younger transgender Americans, I will always have your back as your President, so you can be yourself and reach your God-given potential.

While it often appears that we never agree, that isn't true. I signed 80 bipartisan bills into law last year. From preventing government shutdowns to protecting Asian-Americans from still-too-common hate crimes to reforming military justice.

And soon, we'll strengthen the Violence Against Women Act that I first wrote three decades ago. It is important for us to show the nation that we can come together and do big things.

So tonight I'm offering a Unity Agenda for the Nation. Four big things we can do together.

First, beat the opioid epidemic.

Stringing compressors and document transformers together

Using the `DocumentCompressorPipeline` we can also easily combine multiple compressors in sequence. Along with compressors we can add `BaseDocumentTransformer`s to our pipeline, which don't perform any contextual compression but simply perform some transformation on a set of documents. For example `TextSplitter`s can be used as document transformers to split documents into smaller pieces, and the `EmbeddingsRedundantFilter` can be used to filter out redundant documents based on embedding similarity between documents.

Below we create a compressor pipeline by first splitting our docs into smaller chunks, then removing redundant documents, and then filtering based on relevance to the query.

```
from langchain.document_transformers import EmbeddingsRedundantFilter
from langchain.retrievers.document_compressors import DocumentCompressorPipeline
from langchain.text_splitter import CharacterTextSplitter
```

```
splitter = CharacterTextSplitter(chunk_size=300, chunk_overlap=0, separator=". ")
redundant_filter = EmbeddingsRedundantFilter(embeddings=embeddings)
relevant_filter = EmbeddingsFilter(embeddings=embeddings, similarity_threshold=0.76)
pipeline_compressor = DocumentCompressorPipeline(
    transformers=[splitter, redundant_filter, relevant_filter]
)
```

```
compression_retriever = ContextualCompressionRetriever(base_compressor=pipeline_compressor,
base_retriever=retriever)

compressed_docs = compression_retriever.get_relevant_documents("What did the president say about Ketanji
Jackson Brown")
pretty_print_docs(compressed_docs)
```

Document 1:

One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson

Document 2:

As I said last year, especially to our younger transgender Americans, I will always have your back as your President, so you can be yourself and reach your God-given potential.

While it often appears that we never agree, that isn't true. I signed 80 bipartisan bills into law last year

Document 3:

A former top litigator in private practice. A former federal public defender. And from a family of public school educators and police officers. A consensus builder