

Question Answering with Sources

Contents

- Prepare Data
- Quickstart
- The `stuff` Chain
- The `map_reduce` Chain
- The `refine` Chain
- The `map-rerank` Chain

This notebook walks through how to use LangChain for question answering with sources over a list of documents. It covers four different chain types: `stuff`, `map_reduce`, `refine`, `map-rerank`. For a more in depth explanation of what these chain types are, see [here](#).

Prepare Data

First we prepare the data. For this example we do similarity search over a vector database, but these documents could be fetched in any manner (the point of this notebook to highlight what to do AFTER you fetch the documents).

```
from langchain.embeddings.openai import OpenAIEmbeddings
from langchain.embeddings.cohere import CohereEmbeddings
from langchain.text_splitter import CharacterTextSplitter
from langchain.vectorstores.elastic_vector_search import ElasticVectorSearch
from langchain.vectorstores import Chroma
from langchain.docstore.document import Document
from langchain.prompts import PromptTemplate
```

```
with open(" /state of the union.txt") as f:
```

[Skip to main content](#)

```
texts = text_splitter.split_text(state_of_the_union)

embeddings = OpenAIEmbeddings()
```

```
docsearch = Chroma.from_texts(texts, embeddings, metadatas=[{"source": str(i)} for
i in range(len(texts))])
```

Running Chroma using direct local API.
Using DuckDB in-memory for database. Data will be transient.

```
query = "What did the president say about Justice Breyer"
docs = docsearch.similarity_search(query)
```

```
from langchain.chains.qa_with_sources import load_qa_with_sources_chain
from langchain.llms import OpenAI
```

Quickstart

If you just want to get started as quickly as possible, this is the recommended way to do it:

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="stuff")
query = "What did the president say about Justice Breyer"
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'output_text': ' The president thanked Justice Breyer for his service.\nSOURCES:
30-pl'}
```

If you want more control and understanding over what is happening, please see the information below.

The `stuff` Chain

This sections shows results of using the `stuff` Chain to do question answering with sources.

[Skip to main content](#)

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="stuff")
```

```
query = "What did the president say about Justice Breyer"
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'output_text': ' The president thanked Justice Breyer for his service.\nSOURCES:
30-pl'}
```

Custom Prompts

You can also use your own prompts with this chain. In this example, we will respond in Italian.

```
template = """Given the following extracted parts of a long document and a
question, create a final answer with references ("SOURCES").
If you don't know the answer, just say that you don't know. Don't try to make up
an answer.
ALWAYS return a "SOURCES" part in your answer.
Respond in Italian.
```

```
QUESTION: {question}
=====
{summaries}
=====
```

```
FINAL ANSWER IN ITALIAN:"""
```

```
PROMPT = PromptTemplate(template=template, input_variables=["summaries",
"question"])
```

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="stuff",
prompt=PROMPT)
query = "What did the president say about Justice Breyer"
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'output_text': '\nNon so cosa abbia detto il presidente riguardo a Justice
Breyer.\nSOURCES: 30, 31, 33'}
```

The `map_reduce` Chain

This sections shows results of using the `map_reduce` Chain to do question answering with sources.

[Skip to main content](#)

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="map_reduce")
```

```
query = "What did the president say about Justice Breyer"
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'output_text': ' The president thanked Justice Breyer for his service.\nSOURCES:
30-pl'}
```

Intermediate Steps

We can also return the intermediate steps for `map_reduce` chains, should we want to inspect them. This is done with the `return_map_steps` variable.

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="map_reduce",
return_intermediate_steps=True)
```

```
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'intermediate_steps': [' 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.',
  ' None',
  ' None',
  ' None'],
'output_text': ' The president thanked Justice Breyer for his service.\nSOURCES:
30-pl'}
```

Custom Prompts

You can also use your own prompts with this chain. In this example, we will respond in Italian.

```
question_prompt_template = """Use the following portion of a long document to see
if any of the text is relevant to answer the question.
Return any relevant text in Italian.
{context}
Question: {question}
Relevant text, if any, in Italian:"""
QUESTION_PROMPT = PromptTemplate(
```

[Skip to main content](#)

```

combine_prompt_template = """Given the following extracted parts of a long
document and a question, create a final answer with references ("SOURCES").
If you don't know the answer, just say that you don't know. Don't try to make up
an answer.
ALWAYS return a "SOURCES" part in your answer.
Respond in Italian.

QUESTION: {question}
=====
{summaries}
=====
FINAL ANSWER IN ITALIAN:"""
COMBINE_PROMPT = PromptTemplate(
    template=combine_prompt_template, input_variables=["summaries", "question"]
)

chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="map_reduce",
    return_intermediate_steps=True, question_prompt=QUESTION_PROMPT,
    combine_prompt=COMBINE_PROMPT)
chain({"input_documents": docs, "question": query}, return_only_outputs=True)

```

```

{'intermediate_steps': ["\nStasera vorrei onorare qualcuno che ha dedicato la sua
vita a servire questo paese: il giustizista Stephen Breyer - un veterano
dell'esercito, uno studioso costituzionale e un giustizista in uscita dalla Corte
Suprema degli Stati Uniti. Giustizista Breyer, grazie per il tuo servizio.",
    ' Non pertinente.',
    ' Non rilevante.',
    " Non c'è testo pertinente."],
'output_text': ' Non conosco la risposta. SOURCES: 30, 31, 33, 20.'}

```

Batch Size

When using the `map_reduce` chain, one thing to keep in mind is the batch size you are using during the map step. If this is too high, it could cause rate limiting errors. You can control this by setting the batch size on the LLM used. Note that this only applies for LLMs with this parameter. Below is an example of doing so:

```
llm = OpenAI(batch_size=5, temperature=0)
```

The `refine` Chain

This section shows results of using the `refine` Chain to do question answering with sources.

[Skip to main content](#)

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="refine")
```

```
query = "What did the president say about Justice Breyer"
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'output_text': "\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked him for his service and praised his career as a top litigator in private practice, a former federal public defender, and a family of public school educators and police officers. He noted Justice Breyer's reputation as a consensus builder and the broad range of support he has received from the Fraternal Order of Police to former judges appointed by Democrats and Republicans. He also highlighted the importance of securing the border and fixing the immigration system in order to advance liberty and justice, and mentioned the new technology, joint patrols, dedicated immigration judges, and commitments to support partners in South and Central America that have been put in place. He also expressed his commitment to the LGBTQ+ community, noting the need for the bipartisan Equality Act and the importance of protecting transgender Americans from state laws targeting them. He also highlighted his commitment to bipartisanship, noting the 80 bipartisan bills he signed into law last year, and his plans to strengthen the Violence Against Women Act. Additionally, he announced that the Justice Department will name a chief prosecutor for pandemic fraud and his plan to lower the deficit by more than one trillion dollars in a"}
```

Intermediate Steps

We can also return the intermediate steps for `refine` chains, should we want to inspect them. This is done with the `return_intermediate_steps` variable.

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="refine",
return_intermediate_steps=True)
```

```
chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```

```
{'intermediate_steps': ['\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked Justice Breyer for his service.',
'\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked Justice Breyer for his service, noting his background as a top litigator in private practice, a former federal public defender. and a familv of public school educators and police officers. He praised
```

[Skip to main content](#)

and Republicans. He also noted that in order to advance liberty and justice, it was necessary to secure the border and fix the immigration system, and that the government was taking steps to do both. \n\nSource: 31',

'\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked Justice Breyer for his service, noting his background as a top litigator in private practice, a former federal public defender, and a family of public school educators and police officers. He praised Justice Breyer for being a consensus builder and for receiving a broad range of support from the Fraternal Order of Police to former judges appointed by Democrats and Republicans. He also noted that in order to advance liberty and justice, it was necessary to secure the border and fix the immigration system, and that the government was taking steps to do both. He also mentioned the need to pass the bipartisan Equality Act to protect LGBTQ+ Americans, and to strengthen the Violence Against Women Act that he had written three decades ago. \n\nSource: 31, 33',

'\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked Justice Breyer for his service, noting his background as a top litigator in private practice, a former federal public defender, and a family of public school educators and police officers. He praised Justice Breyer for being a consensus builder and for receiving a broad range of support from the Fraternal Order of Police to former judges appointed by Democrats and Republicans. He also noted that in order to advance liberty and justice, it was necessary to secure the border and fix the immigration system, and that the government was taking steps to do both. He also mentioned the need to pass the bipartisan Equality Act to protect LGBTQ+ Americans, and to strengthen the Violence Against Women Act that he had written three decades ago. Additionally, he mentioned his plan to lower costs to give families a fair shot, lower the deficit, and go after criminals who stole billions in relief money meant for small businesses and millions of Americans. He also announced that the Justice Department will name a chief prosecutor for pandemic fraud. \n\nSource: 20, 31, 33'],

'output_text': '\n\nThe president said that he was honoring Justice Breyer for his dedication to serving the country and that he was a retiring Justice of the United States Supreme Court. He also thanked Justice Breyer for his service, noting his background as a top litigator in private practice, a former federal public defender, and a family of public school educators and police officers. He praised Justice Breyer for being a consensus builder and for receiving a broad range of support from the Fraternal Order of Police to former judges appointed by Democrats and Republicans. He also noted that in order to advance liberty and justice, it was necessary to secure the border and fix the immigration system, and that the government was taking steps to do both. He also mentioned the need to pass the bipartisan Equality Act to protect LGBTQ+ Americans, and to strengthen the Violence Against Women Act that he had written three decades ago. Additionally, he mentioned his plan to lower costs to give families a fair shot, lower the deficit, and go after criminals who stole billions in relief money meant for small businesses and millions of Americans. He also announced that the Justice Department will name a chief prosecutor for pandemic fraud. \n\nSource: 20, 31, 33'}

Custom Prompts

[Skip to main content](#)

```

refine_template = (
    "The original question is as follows: {question}\n"
    "We have provided an existing answer, including sources: {existing_answer}\n"
    "We have the opportunity to refine the existing answer"
    "(only if needed) with some more context below.\n"
    "-----\n"
    "{context_str}\n"
    "-----\n"
    "Given the new context, refine the original answer to better "
    "answer the question (in Italian)"
    "If you do update it, please update the sources as well. "
    "If the context isn't useful, return the original answer."
)
refine_prompt = PromptTemplate(
    input_variables=["question", "existing_answer", "context_str"],
    template=refine_template,
)

question_template = (
    "Context information is below. \n"
    "-----\n"
    "{context_str}"
    "\n-----\n"
    "Given the context information and not prior knowledge, "
    "answer the question in Italian: {question}\n"
)
question_prompt = PromptTemplate(
    input_variables=["context_str", "question"], template=question_template
)

```

```

chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="refine",
    return_intermediate_steps=True, question_prompt=question_prompt,
    refine_prompt=refine_prompt)
chain({"input_documents": docs, "question": query}, return_only_outputs=True)

```

```

{'intermediate_steps': ['\nIl presidente ha detto che Justice Breyer ha dedicato
la sua vita al servizio di questo paese e ha onorato la sua carriera.',
    "\n\nIl presidente ha detto che Justice Breyer ha dedicato la sua vita al
servizio di questo paese, ha onorato la sua carriera e ha contribuito a costruire
un consenso. Ha ricevuto un ampio sostegno, dall'Ordine Fraternal della Polizia a
ex giudici nominati da democratici e repubblicani. Inoltre, ha sottolineato
l'importanza di avanzare la libert  e la giustizia attraverso la sicurezza delle
frontiere e la risoluzione del sistema di immigrazione. Ha anche menzionato le
nuove tecnologie come scanner all'avanguardia per rilevare meglio il traffico di
droga, le pattuglie congiunte con Messico e Guatemala per catturare pi 
trafficienti di esseri umani, l'istituzione di giudici di immigrazione dedicati per
far s  che le famiglie che fuggono da per",
    "\n\nIl presidente ha detto che Justice Breyer ha dedicato la sua vita al

```

[Skip to main content](#)

ex giudici nominati da democratici e repubblicani. Inoltre, ha sottolineato l'importanza di avanzare la libertà e la giustizia attraverso la sicurezza delle frontiere e la risoluzione del sistema di immigrazione. Ha anche menzionato le nuove tecnologie come scanner all'avanguardia per rilevare meglio il traffico di droga, le pattuglie congiunte con Messico e Guatemala per catturare più trafficanti di esseri umani, l'istituzione di giudici di immigrazione dedicati per far sì che le famiglie che fuggono da per",

"\n\nIl presidente ha detto che Justice Breyer ha dedicato la sua vita al servizio di questo paese, ha onorato la sua carriera e ha contribuito a costruire un consenso. Ha ricevuto un ampio sostegno, dall'Ordine Fraternal della Polizia a ex giudici nominati da democratici e repubblicani. Inoltre, ha sottolineato l'importanza di avanzare la libertà e la giustizia attraverso la sicurezza delle frontiere e la risoluzione del sistema di immigrazione. Ha anche menzionato le nuove tecnologie come scanner all'avanguardia per rilevare meglio il traffico di droga, le pattuglie congiunte con Messico e Guatemala per catturare più trafficanti di esseri umani, l'istituzione di giudici di immigrazione dedicati per far sì che le famiglie che fuggono da per"]

'output_text': "\n\nIl presidente ha detto che Justice Breyer ha dedicato la sua vita al servizio di questo paese, ha onorato la sua carriera e ha contribuito a costruire un consenso. Ha ricevuto un ampio sostegno, dall'Ordine Fraternal della Polizia a ex giudici nominati da democratici e repubblicani. Inoltre, ha sottolineato l'importanza di avanzare la libertà e la giustizia attraverso la sicurezza delle frontiere e la risoluzione del sistema di immigrazione. Ha anche menzionato le nuove tecnologie come scanner all'avanguardia per rilevare meglio il traffico di droga, le pattuglie congiunte con Messico e Guatemala per catturare più trafficanti di esseri umani, l'istituzione di giudici di immigrazione dedicati per far sì che le famiglie che fuggono da per"}]

The `map-rerank` Chain

This section shows results of using the `map-rerank` Chain to do question answering with sources.

```
chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="map_rerank",
metadata_keys=['source'], return_intermediate_steps=True)
```

```
query = "What did the president say about Justice Breyer"
result = chain({"input_documents": docs, "question": query},
return_only_outputs=True)
```

```
result["output_text"]
```

[Skip to main content](#)

dedicating his life to serve the country.'

```
result["intermediate_steps"]
```

```
[{'answer': ' The President thanked Justice Breyer for his service and honored him
for dedicating his life to serve the country.',
  'score': '100'},
 {'answer': ' This document does not answer the question', 'score': '0'},
 {'answer': ' This document does not answer the question', 'score': '0'},
 {'answer': ' This document does not answer the question', 'score': '0'}]
```

Custom Prompts

You can also use your own prompts with this chain. In this example, we will respond in Italian.

```
from langchain.output_parsers import RegexParser

output_parser = RegexParser(
    regex=r"(.*?)\nScore: (.*)",
    output_keys=["answer", "score"],
)

prompt_template = """Use the following pieces of context to answer the question at
the end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered the
user's question. This should be in the following format:

Question: [question here]
Helpful Answer In Italian: [answer here]
Score: [score between 0 and 100]

Begin!

Context:
-----
{context}
-----

Question: {question}
Helpful Answer In Italian: """
PROMPT = PromptTemplate(
    template=prompt_template,
    input_variables=["context", "question"],
    output_parser=output_parser,
)

chain = load_qa_with_sources_chain(OpenAI(temperature=0), chain_type="map_rerank",
```

[Skip to main content](#)

```
result = chain({"input_documents": docs, "question": query},  
              return_only_outputs=True)
```

```
result
```

```
{'source': 30,  
 'intermediate_steps': [{'answer': ' Il presidente ha detto che Justice Breyer ha  
dedicato la sua vita a servire questo paese e ha onorato la sua carriera.',  
  'score': '100'},  
 {'answer': ' Il presidente non ha detto nulla sulla Giustizia Breyer.',  
  'score': '100'},  
 {'answer': ' Non so.', 'score': '0'},  
 {'answer': ' Il presidente non ha detto nulla sulla giustizia Breyer.',  
  'score': '100'}],  
 'output_text': ' Il presidente ha detto che Justice Breyer ha dedicato la sua  
vita a servire questo paese e ha onorato la sua carriera.'}
```