

Sitemap Loader

Contents

- Filtering sitemap URLs

Extends from the `WebBaseLoader`, this will load a sitemap from a given URL, and then scrape and load all the pages in the sitemap, returning each page as a document.

The scraping is done concurrently, using `WebBaseLoader`. There are reasonable limits to concurrent requests, defaulting to 2 per second. If you aren't concerned about being a good citizen, or you control the server you are scraping and don't care about load, you can change the `requests_per_second` parameter to increase the max concurrent requests. Note, while this will speed up the scraping process, but may cause the server to block you. Be careful!

```
!pip install nest_asyncio
```

```
Requirement already satisfied: nest_asyncio in  
/Users/tasp/Code/projects/langchain/.venv/lib/python3.10/site-packages (1.5.6)  
  
[notice] A new release of pip available: 22.3.1 -> 23.0.1  
[notice] To update, run: pip install --upgrade pip
```

```
# fixes a bug with asyncio and jupyter  
import nest_asyncio  
nest_asyncio.apply()
```

```
from langchain.document_loaders.sitemap import SitemapLoader
```

```
sitemap_loader =  
SitemapLoader(web_path="https://langchain.readthedocs.io/sitemap.xml")  
  
docs = sitemap_loader.load()
```

[Skip to main content](#)

docs[0]

```
Document(page_content='\n\n\n\n\nWelcome to LangChain - 🦜🔗 LangChain  
0.0.123\n\n\n\n\nSkip to main  
content\n\n\n\nCtrl+K\n\n\n\n🦜🔗 LangChain  
0.0.123\n\nGetting Started\nQuickstart Guide\nModules\nPrompt Templates\nGetting Started\nKey Concepts\nHow-To Guides\nCreate a custom prompt template\nCreate a custom example selector\nProvide few shot examples to a prompt\nPrompt Serialization\nExample Selectors\nOutput Parsers\nReference\nPromptTemplates\nExample Selector\n\n\nLLMs\nGetting Started\nKey Concepts\nHow-To Guides\nGeneric Functionality\nCustom LLM\nFake LLM\nLLM Caching\nLLM Serialization\nToken Usage Tracking\nIntegrations\nAI21\nAleph Alpha\nAnthropic\nAzure OpenAI LLM Example\nBanana\nCerebrumAI LLM Example\nCohere\nDeepInfra LLM Example\nForefrontAI LLM Example\nGooseAI LLM Example\nHugging Face Hub\nManifest\nModal\nOpenAI\nPetals LLM Example\nPromptLayer\nOpenAI\nSageMakerEndpoint\nSelf-Hosted Models via Runhouse\nStochasticAI\nWriter\n\nAsync API for LLM\nStreaming with LLMs\nReference\nDocument Loaders\nKey Concepts\nHow To Guides\nConLLOpenAI\nAirbyte JSON\nAZLyrics\nBlackboard\nCollege Confidential\nCopy Paste\nCSV Loader\nDirectory Loader\nEmail\nEverNote\nFacebook Chat\nFigma\nGCS Directory\nGCS File Storage\nGitBook\nGoogle Drive\nGutenberg\nHacker News\nHTML\niFixit\nImages\nIMSDb\nMarkdown\nNotebook\nNotion\nObsidian\nPDF\nPowerPoint\nReadTheDocs Documentation\nRoam\ns3 Directory\ns3 File\nSubtitle Files\nTelegram\nUnstructured File Loader\nURL\nWeb Base\nWord Documents\nYouTube\n\nUtils\nKey Concepts\nGeneric Utilities\nBash\nBing Search\nGoogle Search\nGoogle Serper API\nIFTTT WebHooks\nPython REPL\nRequests\nSearxNG Search API\nSerpAPI\nWolfram Alpha\nZapier Natural Language Actions API\n\nReference\nPython REPL\nSerpAPI\nSearxNG Search\nDocstore\nText Splitter\nEmbeddings\nVectorStores\n\nIndexes\nGetting Started\nKey Concepts\nHow To Guides\nEmbeddings\nHypothetical Document Embeddings\nText Splitter\nVectorStores\nAtlasDB\nChroma\nDeeplake\nElasticSearch\nFAISS\nMilvus\nOpenSearch\nPGVector\nPinecone\nQdrant\nRedis\nWeaviate\nChatGPT Plugin Retriever\nVectorStore Retriever\nAnalyze Document\nChat Index\nGraph QA\nQuestion Answering with Sources\nQuestion Answering\nSummarization\nRetrieval Question/Answering\nRetrieval Question Answering with Sources\nVector DB Text Generation\n\nChains\nGetting Started\nHow-To Guides\nGeneric Chains\nLoading from LangChainHub\nLLM Chain\nSequential Chains\nSerialization\nTransformation Chain\nUtility Chains\nAPI Chains\nSelf-Critique Chain with Constitutional AI\nBashChain\nLLMCheckerChain\nLLMMath\nLLMRequestsChain\nLLMSummarizationCheckerChain\nModeration\nPAL\nSQLite example\n\nAsync API for Chain\n\nKey Concepts\nReference\nAgents\nGetting Started\nKey Concepts\nHow-To Guides\nAgents and Vectorstores\nAsync API for Agent\nConversation Agent (for Chat Models)\nChatGPT Plugins\nCustom Agent\nDefining Custom Tools\nHuman as a tool\nIntermediate Steps\nLoading from LangChainHub\nMax Iterations\nMulti Input Tools\nSearch Tools\nSerialization\nAdding SharedMemory to an Agent and its Tools\nCSV Agent\nJSON Agent\nOpenAPI Agent\nPandas Dataframe Agent\nPyvthon Agent\nSQL
```

[Skip to main content](#)

[illegible]

[Skip to main content](#)

Python REPLs, embeddings, search engines, and more. LangChain provides a large collection of common utils to use in your application.

Chains: Chains go beyond just a single LLM call, and are sequences of calls (whether to an LLM or a different utility). LangChain provides a standard interface for chains, lots of integrations with other tools, and end-to-end chains for common applications.

Indexes: Language models are often more powerful when combined with your own text data - this module covers best practices for doing exactly that.

Agents: Agents involve an LLM making decisions about which Actions to take, taking that Action, seeing an Observation, and repeating that until done. LangChain provides a standard interface for agents, a selection of agents to choose from, and examples of end to end agents.

Memory: Memory is the concept of persisting state between calls of a chain/agent. LangChain provides a standard interface for memory, a collection of memory implementations, and examples of chains/agents that use memory.

Chat: Chat models are a variation on Language Models that expose a different API - rather than working with raw text, they work with messages. LangChain provides a standard interface for working with them and doing all the same things as above.

Use Cases

The above modules can be used in a variety of ways. LangChain also provides guidance and assistance in this. Below are some of the common use cases LangChain supports.

Agents: Agents are systems that use a language model to interact with other tools. These can be used to do more grounded question/answering, interact with APIs, or even take actions.

Chatbots: Since language models are good at producing text, that makes them ideal for creating chatbots.

Data Augmented Generation: Data Augmented Generation involves specific types of chains that first interact with an external datasource to fetch data to use in the generation step. Examples of this include summarization of long pieces of text and question/answering over specific data sources.

Question Answering: Answering questions over specific documents, only utilizing the information in those documents to construct an answer. A type of Data Augmented Generation.

Summarization: Summarizing longer documents into shorter, more condensed chunks of information. A type of Data Augmented Generation.

Querying Tabular Data: If you want to understand how to use LLMs to query data that is stored in a tabular format (csvs, SQL, dataframes, etc) you should read this page.

Evaluation: Generative models are notoriously hard to evaluate with traditional metrics. One new way of evaluating them is using language models themselves to do the evaluation. LangChain provides some prompts/chains for assisting in this.

Generate similar examples: Generating similar examples to a given input. This is a common use case for many applications, and LangChain provides some prompts/chains for assisting in this.

Compare models: Experimenting with different prompts, models, and chains is a big part of developing the best possible application. The ModelLaboratory makes it easy to do so.

Reference Docs

All of LangChain's reference documentation, in one place. Full documentation on all methods, classes, installation methods, and integration setups for LangChain.

Reference Documentation

Guides for how other companies/products can be used with LangChain

LangChain Ecosystem

Additional Resources

Additional collection of resources we think may be useful as you develop your application!

LangChainHub: The LangChainHub is a place to share and explore other prompts, chains, and agents.

Glossary: A glossary of all related terms, papers, methods, etc. Whether implemented in LangChain or not!

Gallery: A collection of our favorite projects that use LangChain. Useful for finding inspiration or seeing how things were done in other applications.

Deployments: A collection of instructions, code snippets, and template repositories for deploying LangChain apps.

Discord: Join us on our Discord to discuss all things LangChain!

Tracing: A guide on using tracing in

[Skip to main content](#)

```
support. Please fill out this form and we'll set up a dedicated support Slack
channel.\n\n\n\n\n\n\n\n\n\n\n\nnext\nQuickstart Guide\n\n\n\n\n\n\n\n\n\n\n
Contents\n \n\n\nGetting Started\nModules\nUse Cases\nReference Docs\nLangChain
Ecosystem\nAdditional Resources\n\n\n\n\n\n\n\n\n\nBy Harrison Chase\n\n\n\n\n\n
\n      © Copyright 2023, Harrison Chase.\n      \n\n\n\n\n\n Last updated on Mar
24, 2023.\n \n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n', lookup_str='', metadata={'source':
'https://python.langchain.com/en/stable/', 'loc':
'https://python.langchain.com/en/stable/', 'lastmod': '2023-03-
24T19:30:54.647430+00:00', 'changefreq': 'weekly', 'priority': '1'},
lookup_index=0)
```

Filtering sitemap URLs

Sitemaps can be massive files, with thousands of urls. Often you don't need every single one of them. You can filter the urls by passing a list of strings or regex patterns to the `url_filter` parameter. Only urls that match one of the patterns will be loaded.

```
loader = SitemapLoader(
    "https://langchain.readthedocs.io/sitemap.xml",
    filter_urls=["https://python.langchain.com/en/latest/"]
)
documents = loader.load()
```

documents[0]

```
Document(page_content='\n\n\n\n\n\nWelcome to LangChain – 🦜🔗 LangChain  
0.0.123\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\nSkip to main  
content\n\n\n\n\n\n\n\n\n\n\n\nCtrl+K\n\n\n\n\n\n\n\n\n\n\n\n\n🦜🔗 LangChain  
0.0.123\n\n\n\n\nGetting Started\n\n\nQuickstart  
Guide\n\n\nModules\n\n\nModels\n\nLLMs\nGetting Started\nGeneric Functionality\nHow to  
use the async API for LLMs\nHow to write a custom LLM wrapper\nHow (and why) to  
use the fake LLM\nHow to cache LLM calls\nHow to serialize LLM classes\nHow to  
stream LLM responses\nHow to track token usage\n\n\nIntegrations\nAI21\nAleph  
Alpha\nAnthropic\nAzure OpenAI LLM Example\nBanana\nCerebrumAI LLM  
Example\nCohere\nDeepInfra LLM Example\nForefrontAI LLM Example\nGooseAI LLM  
Example\nHugging Face Hub\nManifest\nModal\nOpenAI\nPetals LLM  
Example\nPromptLayer OpenAI\nSageMakerEndpoint\nSelf-Hosted Models via  
Runhouse\nStochasticAI\nWriter\n\n\nReference\n\n\nChat Models\nGetting  
Started\nHow-To Guides\nHow to use few shot examples\nHow to stream  
responses\n\n\nIntegrations\nAzure\nOpenAI\nPromptLayer ChatOpenAI\n\n\n\n\nText  
Embedding Models\nAzureOpenAI\nCohere\nFake Embeddings\nHugging Face  
Hub\nInstructEmbeddings\nOpenAI\nSageMaker Endpoint Embeddings\nSelf Hosted  
Embeddings\nTensorflowHub\n\n\n\n\nPrompts\nPrompt Templates\nGetting Started\nHow-
```

[Skip to main content](#)

```

serialize prompts\n\n\nReference\nPromptTemplates\nExample Selector\n\n\n\n\nChat
Prompt Template\nExample Selectors\nHow to create a custom example
selector\nLengthBased ExampleSelector\nMaximal Marginal Relevance
ExampleSelector\nNGram Overlap ExampleSelector\nSimilarity
ExampleSelector\n\n\n\nOutput Parsers\nOutput
Parsers\nCommaSeparatedListOutputParser\nOutputFixingParser\nPydanticOutputParser\n
RetryOutputParser\nStructured Output Parser\n\n\n\n\nIndexes\nGetting
Started\nDocument Loaders\nCoNLL-U\nAirbyte JSON\nAZLyrics\nBlackboard\nCollege
Confidential\nCopy Paste\nCSV Loader\nDirectory Loader\nEmail\nEverNote\nFacebook
Chat\nFigma\nGCS Directory\nGCS File Storage\nGitBook\nGoogle
Drive\nGutenberg\nHacker
News\nHTML\niFixit\nImages\nIMSDb\nMarkdown\nNotebook\nNotion\nObsidian\nPDF\nPower
Point\nReadTheDocs Documentation\nRoam\ns3 Directory\ns3 File\nSubtitle
Files\nTelegram\nUnstructured File Loader\nURL\nWeb Base\nWord
Documents\nYouTube\n\n\n\nText Splitters\nGetting Started\nCharacter Text
Splitter\nHuggingFace Length Function\nLatex Text Splitter\nMarkdown Text
Splitter\nNLTK Text Splitter\nPython Code Text
Splitter\nRecursiveCharacterTextSplitter\nSpacy Text Splitter\ntiktoken (OpenAI)
Length Function\nTiktokenText Splitter\n\n\n\nVectorstores\nGetting
Started\nAtlasDB\nChroma\nDeep
Lake\nElasticSearch\nFAISS\nMilvus\nOpenSearch\nPGVector\nPinecone\nQdrant\nRedis\n
Weaviate\n\n\n\nRetrievers\nChatGPT Plugin Retriever\nVectorStore
Retriever\n\n\n\n\nMemory\nGetting Started\nHow-To
Guides\nConversationBufferMemory\nConversationBufferWindowMemory\nEntity
Memory\nConversation Knowledge Graph
Memory\nConversationSummaryMemory\nConversationSummaryBufferMemory\nConversationTok
enBufferMemory\nHow to add Memory to an LLMChain\nHow to add memory to a Multi-
Input Chain\nHow to add Memory to an Agent\nHow to customize conversational
memory\nHow to create a custom Memory class\nHow to use multiple memroy classes in
the same chain\n\n\n\n\nChains\nGetting Started\nHow-To Guides\nAsync API for
Chain\nLoading from LangChainHub\nLLM Chain\nSequential
Chains\nSerialization\nTransformation Chain\nAnalyze Document\nChat Index\nGraph
QA\nHypothetical Document Embeddings\nQuestion Answering with Sources\nQuestion
Answering\nSummarization\nRetrieval Question/Answering\nRetrieval Question
Answering with Sources\nVector DB Text Generation\nAPI Chains\nSelf-Critique Chain
with Constitutional AI\nBashChain\nLLMCheckerChain\nLLM
Math\nLLMRequestsChain\nLLMSummarizationCheckerChain\nModeration\nPAL\nSQLite
example\n\n\n\nReference\n\n\n\nAgents\nGetting Started\nTools\nGetting
Started\nDefining Custom Tools\nMulti Input Tools\nBash\nBing Search\nChatGPT
Plugins\nGoogle Search\nGoogle Serper API\nHuman as a tool\nIFTTT WebHooks\nPython
REPL\nRequests\nSearch Tools\nSearxNG Search API\nSerpAPI\nWolfram Alpha\nZapier
Natural Language Actions API\n\n\n\nAgents\nAgent Types\nCustom Agent\nConversation
Agent (for Chat Models)\nConversation Agent\nMRKL\nMRKL Chat\nReAct\nSelf Ask With
Search\n\n\n\nToolkits\nCSV Agent\nJSON Agent\nOpenAPI Agent\nPandas Dataframe
Agent\nPython Agent\nSQL Database Agent\nVectorstore Agent\n\n\n\nAgent
Executors\nHow to combine agents and vectorstores\nHow to use the async API for
Agents\nHow to create ChatGPT Clone\nHow to access intermediate steps\nHow to cap
the max number of iterations\nHow to add SharedMemory to an Agent and its
Tools\n\n\n\n\n\nUse Cases\n\n\nPersonal Assistants\nQuestion Answering over
Docs\nChatbots\nQuerying Tabular Data\nInteracting with
APIs\nSummarization\nExtraction\nEvaluation\nAgent Benchmarking: Search +
Calculator\nAgent VectorDB Question Answering Benchmarking\nBenchmarking
Template\nData Augmented Question Answering\nUsing Hugging Face Datasets\nLLM
Math\nQuestion Answering Benchmarking: Paul Graham Essay\nQuestion Answering

```

[Skip to main content](#)

Chinook\n/n\n/nReference\n/nInstallation\n/nIntegrations\n/nAPI
References\n/nPrompts\n/nPromptTemplates\n/nExample Selector\n/n\n/nUtilities\n/nPython
REPL\n/nSerpAPI\n/nSearxNG Search\n/nDocstore\n/nText
Splitter\n/nEmbeddings\n/nVectorStores\n/n\n/n\n/nChains\n/nAgents\n/n\n/n\n/nEcosystem\n/n\n/nLangChain Ecosystem\n/nAI21

Labs\n/nAtlasDB\n/nBanana\n/nCerebrumAI\n/nChroma\n/nCohere\n/nDeepInfra\n/nDeep Lake\n/nForefrontAI\n/nGoogle Search Wrapper\n/nGoogle Serper
Wrapper\n/nGooseAI\n/nGraphsignal\n/nHazy Research\n/nHelicone\n/nHugging Face\n/nMilvus\n/nModal\n/nNLPCloud\n/nOpenAI\n/nOpenSearch\n/nPetals\n/nPGVector\n/nPinecone\n/nPromptLayer\n/nQdrant\n/nRunhouse\n/nSearxNG Search
API\n/nSerpAPI\n/nStochasticAI\n/nUnstructured\n/nWeights & Biases\n/nWeaviate\n/nWolfram Alpha Wrapper\n/nWriter\n/n\n/n\n/nAdditional
Resources\n/n\n/nLangChainHub\n/nGlossary\n/nLangChain
Gallery\n/nDeployments\n/nTracing\n/nDiscord\n/nProduction
Support\n/n.rst\n/n\n/n\n/n\n/n\n/n.pdf\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/n\n/nWelcome to LangChain\n/n\n/n\n/n\n/n Contents \n/n\n/n\n/nGetting Started\n/nModules\n/nUse Cases\n/nReference Docs\n/nLangChain Ecosystem\n/nAdditional
Resources\n/n\n/n\n/n\n/n\n/n\n/n\n/nWelcome to LangChain#\n/nLangChain is a framework for developing applications powered by language models. We believe that the most powerful and differentiated applications will not only call out to a language model via an API, but will also:\n/nBe data-aware: connect a language model to other sources of data\n/nBe agentic: allow a language model to interact with its environment\n/nThe LangChain framework is designed with the above principles in mind.\n/nThis is the Python specific portion of the documentation. For a purely conceptual guide to LangChain, see here. For the JavaScript documentation, see here.\n/nGetting Started#\n/nCheckout the below guide for a walkthrough of how to get started using LangChain to create an Language Model application.\n/nGetting Started Documentation\n/n\n/n\n/n\n/n\n/nModules#\n/nThere are several main modules that LangChain provides support for.\n/nFor each module we provide some examples to get started, how-to guides, reference docs, and conceptual guides.\n/nThese modules are, in increasing order of complexity:\n/nModels: The various model types and model integrations LangChain supports.\n/nPrompts: This includes prompt management, prompt optimization, and prompt serialization.\n/nMemory: Memory is the concept of persisting state between calls of a chain/agent. LangChain provides a standard interface for memory, a collection of memory implementations, and examples of chains/agents that use memory.\n/nIndexes: Language models are often more powerful when combined with your own text data - this module covers best practices for doing exactly that.\n/nChains: Chains go beyond just a single LLM call, and are sequences of calls (whether to an LLM or a different utility). LangChain provides a standard interface for chains, lots of integrations with other tools, and end-to-end chains for common applications.\n/nAgents: Agents involve an LLM making decisions about which Actions to take, taking that Action, seeing an Observation, and repeating that until done. LangChain provides a standard interface for agents, a selection of agents to choose from, and examples of end to end agents.\n/n\n/n\n/n\n/n\n/nUse Cases#\n/nThe above modules can be used in a variety of ways. LangChain also provides guidance and assistance in this. Below are some of the common use cases LangChain supports.\n/nPersonal Assistants: The main LangChain use case. Personal assistants need to take actions, remember interactions, and have knowledge about your data.\n/nQuestion Answering: The second big LangChain use case. Answering questions over specific documents, only utilizing the information in those documents to construct an answer.\n/nChatbots: Since language models are good at producing text, that makes them ideal for creating chatbots.\n/nQuerying Tabular Data: If you want to understand how to use LLMs to query data that is

[Skip to main content](#)

powerful in order to give them more up-to-date information and allow them to take actions.

Extraction: Extract structured information from text.

Summarization: Summarizing longer documents into shorter, more condensed chunks of information. A type of Data Augmented Generation.

Evaluation: Generative models are notoriously hard to evaluate with traditional metrics. One new way of evaluating them is using language models themselves to do the evaluation. LangChain provides some prompts/chains for assisting in this.

Reference Docs

All of LangChain's reference documentation, in one place. Full documentation on all methods, classes, installation methods, and integration setups for LangChain.

Reference Documentation

LangChain Ecosystem

Guides for how other companies/products can be used with LangChain

LangChain Ecosystem

Additional Resources

Additional collection of resources we think may be useful as you develop your application!

LangChainHub: The LangChainHub is a place to share and explore other prompts, chains, and agents.

Glossary: A glossary of all related terms, papers, methods, etc. Whether implemented in LangChain or not!

Gallery: A collection of our favorite projects that use LangChain. Useful for finding inspiration or seeing how things were done in other applications.

Deployments: A collection of instructions, code snippets, and template repositories for deploying LangChain apps.

Tracing: A guide on using tracing in LangChain to visualize the execution of chains and agents.

Model Laboratory: Experimenting with different prompts, models, and chains is a big part of developing the best possible application. The ModelLaboratory makes it easy to do so.

Discord: Join us on our Discord to discuss all things LangChain!

Production Support: As you move your LangChains into production, we'd love to offer more comprehensive support. Please fill out this form and we'll set up a dedicated support Slack channel.

next

Quickstart Guide

Contents

Getting Started

Modules

Use Cases

Reference Docs

LangChain Ecosystem

Additional Resources

By Harrison Chase

© Copyright 2023, Harrison Chase.

Last updated on Mar 27, 2023.

'https://python.langchain.com/en/latest/', 'loc': 'https://python.langchain.com/en/latest/', 'lastmod': '2023-03-27T22:50:49.790324+00:00', 'changefreq': 'daily', 'priority': '0.9'}, lookup_index=0)