Manage knowledge bases

Beta

This product or feature is subject to the "Pre-GA Offerings Terms" in the General Service Terms section of the Service Specific Terms (/terms/service-terms#1). Pre-GA products and features are available "as is" and might have limited support. For more information, see the launch stage descriptions (/products#product-launch-stages).

A *knowledge base* represents a collection of *knowledge documents* that you provide to Dialogflow. Your knowledge documents contain information that may be useful during conversations with end-users. Some Dialogflow features use knowledge bases when looking for a response to an end-user expression. This guide describes how to create and manage knowledge bases.

A knowledge base is applied at the agent level (/dialogflow/es/docs/data-level).

Before you begin

You should do the following before reading this guide:

1. Read Dialogflow basics (/dialogflow/docs/basics).

2. Perform setup steps (/dialogflow/docs/quick/setup).

Create a knowledge base

The samples below show you how to use the Dialogflow Console, REST API (including command line), or client libraries to create a knowledge base. To use the API, call the `create` method on the `KnowledgeBase` (/dialogflow/docs/reference/common-types#knowledgebases) type.

Web UIREST (#rest)Java (#java)Node.js (#node.js)Python (#python)
(#web-ui)

Use the Dialogflow Console to create a knowledge base:

1. Go to the <u>Dialogflow ES console</u> (https://dialogflow.cloud.google.com)

2. Select an agent

3. Click **Knowledge** on the left sidebar menu

4. Click **Create Knowledge Base**

5. Enter a knowledge base name

6. Click **Save**

Add a document to the knowledge base

Your new knowledge base currently has no documents, so you should add a document to it. See <u>Supported content</u> (#supported-content) below for a description of all supported content options. You can use the <u>Cloud Storage FAQ</u> (https://cloud.google.com/storage/docs/faq) document for this example.

Note: Knowledge documents are only parsed when explicitly creating or updating them. If your source content changes after creating a knowledge document, you need to <u>update the knowledge document</u> (#update).

The samples below show you how to use the Dialogflow Console, REST API (including command line), or client libraries to create a knowledge document. To use the API, call the `create` method on the <u>Document</u> (/dialogflow/docs/reference/common-types#documents) type.

<u>Web UI</u> <u>REST</u> (#rest) <u>Java</u> (#java) <u>Node.js</u> (#node.js) <u>Python</u> (#python)
(#web-ui)

    Use the Dialogflow Console to create a knowledge document:

1. If you are not continuing from steps above, navigate to your knowledge base settings:

   a. Go to the <u>Dialogflow ES console</u> (https://dialogflow.cloud.google.com)

   b. Select an agent

   c. Click **Knowledge** on the left sidebar menu

   d. Click your knowledge base name

2. Click **New Document** or **Create the first one**

3. Enter a document name

4. Select **text/html** for **Mime Type**

5. Select **FAQ** for **Knowledge Type**

6. Select **URL** for **Data Source**

7. Enter **https://cloud.google.com/storage/docs/faq** in the **URL** field

8. Click **CREATE**

Creating a document is a long-running operation, so it may take a substantial amount of time to complete. You can poll the status of this operation to see if it has completed. Once completed, the operation contains the newly created document ID. Save this ID for future processing. For more information, see Long-running operations (/dialogflow/docs/how/long-running-operations).

Manage knowledge documents

Update knowledge document content

If you update your content referenced by a knowledge document, your knowledge document may not automatically refresh. Your content is only automatically refreshed if it is provided as a public URL and you have checked the **Enable Automatic Reload** option for the document.

To manually refresh Cloud Storage or public URL document content, call the `reload` method on the Document (/dialogflow/docs/reference/common-types#documents) type.

To manually refresh uploaded raw content, use the `delete` and `create` methods on the Document (/dialogflow/docs/reference/common-types#documents) type to re-create your document.

List knowledge documents

You can list all knowledge documents for your knowledge base. To use the API, call the `list` method on the Document (/dialogflow/docs/reference/common-types#documents) type.

Delete knowledge documents

You can delete knowledge documents for your knowledge base. To use the API, call the `delete` method on the Document (/dialogflow/docs/reference/common-types#documents) type. If you do not have the document ID, you can list the documents as described above.

Supported content

The following knowledge document types are supported:

- **FAQ**: The document content contains question and answer pairs as either HTML or CSV. Typical FAQ HTML formats are parsed accurately, but unusual formats may fail to be parsed. CSV must have questions in the first column and answers in the second, with no header. Because of this explicit format, they are always parsed accurately.

- **Extractive QA**: Documents for which unstructured text is extracted and used for question answering.

The following table shows the supported MIME types by *Knowledge Type* and *Source*.

Knowledge Type \ Source	Uploaded file (Document.content) (*NOT* recommended)	Uploaded file (Document.raw_content) (recommended)	File from Cloud Storage (Document.contentUri)	File from public URL (Document.contentUri)
FAQ	text/csv	text/csv	text/csv	text/html
Extractive QA	text/plain, text/html	text/plain, text/html, application/pdf	text/plain, text/html, application/pdf	N/A

Document content has the following known issues, limitations, and best practices:

General:

- Files from public URLs must have been crawled by the Google search indexer, so that they exist in the search index. You can check this with the Google Search Console (https://search.google.com/search-console). Note that the indexer does not keep your content fresh. You must explicitly update your knowledge document when the source content changes.

- CSV files must use commas as delimiters.

- Confidence scores are not yet calibrated between FAQs and Knowledge Base Articles, so if you use both FAQ and Knowledge Base Articles, the best result may not always be the

highest.

- Dialogflow removes HTML tags from content when creating responses. Because of this, it's best to avoid HTML tags and use plain text when possible.

- Google Assistant responses have a 640 character limit per chat bubble, so long answers are truncated when integrating with Google Assistant.

- The maximum document size is 50 MB.

- When using Cloud Storage files, you should either use public URIs or private URIs that your user account or service account has access to.

Specific to FAQ:

- CSV must have questions in the first column and answers in the second, with no header.

- Use CSV whenever possible, because CSV is parsed most accurately.

- Public HTML content with a single QA pair is not supported.

- The number of QA pairs in one document should not exceed 2000.

- Duplicate questions with different answers is not supported.

- You can use any FAQ document; the FAQ parser is capable of handling most FAQ formats.

Specific to Extractive QA:

- Extractive QA is currently experimental. It is based on similar technologies that have been tried and tested at Google in products like Search and Assistant. Send us your feedback (/dialogflow/docs/support) on how well it works for Dialogflow.

- Content with dense text works best. Avoid content with many single sentence paragraphs.

- Tables and lists are not supported.

- The number of paragraphs in one document should not exceed 2000.

- If an article is long (> 1000 words), try to break it down into multiple, smaller articles. If the article covers multiple issues, it can be broken into shorter articles covering the individual issues. If the article only covers one issue, then focus the article on the issue description and keep the issue resolution short.

- Ideally, only the core content of an article should be provided (issue description and resolution). Additional content like author name, modification history, related links, and ads are not important.

- Try to include a description for the issues an article can help with and/or sample queries that this article can answer.

Using Cloud Storage

If your content is not public, storing your content in Cloud Storage (/storage/docs) is the recommended option. When creating knowledge documents, you provide the URLs for your Cloud Storage objects.

Creating Cloud Storage buckets and objects

When creating the Cloud Storage bucket:

- Be sure that you have selected the GCP project you use for Dialogflow.

- Ensure that the user account or service account you normally use to access the Dialogflow API has read permissions to the bucket objects (/storage/docs/access-control).

- Use the Standard Storage class (/storage/docs/storage-classes#standard).

- Set the bucket location (/storage/docs/locations) to a location nearest to your location. You will need the location ID (for example, `us-west1`) for some API calls, so take note of your choice.

Follow the Cloud Storage quickstart (/storage/docs/quickstart-console) instructions to create a bucket and upload files.

Supplying a Cloud Storage object to a knowledge base document

To supply your content:

- Create a knowledge base as described above (#create-kb).

- Create a knowledge document as described above (#create-kd). When calling the `create` method on the Document (/dialogflow/docs/reference/common-types#documents) type, set the

`contentUri` field to the URL of your Cloud Storage document. The format of this URL is `gs://bucket-name/object-name`.

Previous

← **Manage intents with the API** (/dialogflow/es/docs/how/manage-intents)

Next

Update data with a FieldMask (/dialogflow/es/docs/how/field-mask) →