



File Directory

This covers how to load all documents in a directory.

Under the hood, by default this uses the [UnstructuredLoader](#)

```
from langchain.document_loaders import DirectoryLoader
```

We can use the `glob` parameter to control which files to load. Note that here it doesn't load the `.rst` file or the `.html` files.

```
loader = DirectoryLoader('../', glob="**/*.md")
```



```
docs = loader.load()
```

```
len(docs)
```

```
1
```

Show a progress bar

By default a progress bar will not be shown. To show a progress bar, install the `tqdm` library (e.g. `pip install tqdm`), and set the `show_progress` parameter to `True`.

```
loader = DirectoryLoader('../', glob="**/*.md", show_progress=True)
docs = loader.load()
```

```
Requirement already satisfied: tqdm in
/Users/jon/.pyenv/versions/3.9.16/envs/microbiome-app/lib/python3.9/site-
```

```
packages (4.65.0)
```

```
0it [00:00, ?it/s]
```

Use multithreading

By default the loading happens in one thread. In order to utilize several threads set the `use_multithreading` flag to true.

```
loader = DirectoryLoader('../', glob="**/*.md", use_multithreading=True)
docs = loader.load()
```

Change loader class

By default this uses the `UnstructuredLoader` class. However, you can change up the type of loader pretty easily.

```
from langchain.document_loaders import TextLoader
```

```
loader = DirectoryLoader('../', glob="**/*.md", loader_cls=TextLoader)
```

```
docs = loader.load()
```

```
len(docs)
```

```
1
```

If you need to load Python source code files, use the `PythonLoader`.


```

color: #bfbf7f"/data/source/langchain/langchain/document_loaders/
decoration-color: #808000; font-weight: bold">text.py</span>:
29 in
|
|
|
|
|
26
|
|
|
|
|
27
|
with
open
self.file_path, encoding=
self.encoding)
|
|
28
|
try
|
|
|
>
29
|
|
|
|
|
30
|
except
UnicodeDecodeError
as e:
|
|
31
|
if
self.autodetect_encoding:
|
|
32
|
self.detect_file_encodings()
|
|
|
|
|

```

```

color: #bfbf7f">/home/spike/.pyenv/versions/3.9.11/lib/python3.9/
```

```

codecs.py
322 in
|
|
|
|
319
def
decode
(input, f
False):
|
320
# decode input
|
|
321
data =
input
|
|
322
consumed) =
self.errors, f
|
323
# keep undecode
|
|
324
self.buf
|
325
return r
|

```

```

>
'utf-8' codec
0xca in p

```

```

text-decoration-color: #008080; font-weight: bold">0</span>: invalid continu

<span style="font-style: italic">The above exception was the direct cause of

<span style="color: #800000; text-decoration-color: #800000">|_____
style="color: #800000; text-decoration-color: #800000; font-weight: bold">Tr
#bf7f7f; text-decoration-color: #bf7f7f; font-weight: bold">(most recent cal
#800000; text-decoration-color: #800000">|_____<
<span style="color: #800000; text-decoration-color: #800000">|</span> in <sp
decoration-color: #00ff00">&lt;module&gt;</span>:<span style="color: #0000ff
#0000ff">1</span>
style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #800000">> </span>1 loader.load()
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">2 </span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #bfbf7f">/data/source/langchain/langchain/document_loaders/</span><sp
decoration-color: #808000; font-weight: bold">directory.py</span>:<span styl
color: #0000ff">84</span> in <span style="color: #00ff00; text-decoration-co
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">81 </span><span style="color: #7f7f7f; text-decor
| </span><span style="color: #0000ff; text-decoration-color: #0000ff">if</
text-decoration-color: #00ffff">self</span>.silent_errors:
style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">82 </span><span style="color: #7f7f7f; text-decor
| | </span>logger.warning(e)
decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">83 </span><span style="color: #7f7f7f; text-decor
| </span><span style="color: #0000ff; text-decoration-color: #0000ff">else
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #800000">> </span>84 <span style="color: #7f7f7f; text-decoration-col

```



```

</span><span style="color: #0000ff; text-decoration-color: #0000ff">raise</s
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">85 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">finally<
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">86 </span><span style="color: #7f7f7f; text-decor
| </span><span style="color: #0000ff; text-decoration-color: #0000ff">if</
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">87 </span><span style="color: #7f7f7f; text-decor
| | </span>pbar.update(<span style="color: #0000ff; text-decoration-colo
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #bfbf7f">/data/source/langchain/langchain/document_loaders/</span><sp
decoration-color: #808000; font-weight: bold">directory.py</span>:<span styl
color: #0000ff">78</span> in <span style="color: #00ff00; text-decoration-co
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">75 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">if</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">76 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">if</span>
style="color: #ff00ff; text-decoration-color: #ff00ff">or</span> <span style
color: #00ffff">self</span>.load_hidden: <span style="
#800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">77 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">try</spa
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #800000">> </span>78 <span style="color: #7f7f7f; text-decoration-col
</span>sub_docs = <span style="color: #00ffff; text-decoration-color: #00fff
style="color: #00ffff; text-decoration-color: #00ffff">str</span>(i), **<spa
decoration-color: #00ffff">self</span>.loader_kwargs).load() <span style
color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa

```

```

decoration-color: #7f7f7f">79 </span><span style="color: #7f7f7f; text-decor
|   </span>docs.extend(sub_docs)
decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>   <spa
decoration-color: #7f7f7f">80 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">except</
text-decoration-color: #00ffff">Exception</span> <span style="color: #0000ff
#0000ff">as</span> e:                               <span
decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>   <spa
decoration-color: #7f7f7f">81 </span><span style="color: #7f7f7f; text-decor
|   </span><span style="color: #0000ff; text-decoration-color: #0000ff">if</
text-decoration-color: #00ffff">self</span>.silent_errors:
style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #bfbf7f">/data/source/langchain/langchain/document_loaders/</span><sp
decoration-color: #808000; font-weight: bold">text.py</span>:<span style="co
#0000ff">44</span> in <span style="color: #00ff00; text-decoration-color: #0
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>   <spa
decoration-color: #7f7f7f">41 </span><span style="color: #7f7f7f; text-decor
|   </span><span style="color: #0000ff; text-decoration-color: #0000ff">exce
text-decoration-color: #00ffff">UnicodeDecodeError</span>:
style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>   <spa
decoration-color: #7f7f7f">42 </span><span style="color: #7f7f7f; text-decor
|   |   </span><span style="color: #0000ff; text-decoration-color: #0000ff">
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span>   <spa
decoration-color: #7f7f7f">43 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">else</sp
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <span
color: #800000">> </span>44 <span style="color: #7f7f7f; text-decoration-col
</span><span style="color: #0000ff; text-decoration-color: #0000ff">raise</s
decoration-color: #00ffff">RuntimeError</span>(<span style="color: #808000;
loading {</span><span style="color: #00ffff; text-decoration-color: #00ffff"
style="color: #808000; text-decoration-color: #808000">}"</span>) <span styl
color: #0000ff">from</span> <span style="color: #00ffff; text-decoration-col
underline">e</span>                               <span style="color: #800000; text-decoration-

```



```

<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">45 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">except</
text-decoration-color: #00ffff">Exception</span> <span style="color: #0000ff
#0000ff">as</span> e:
decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">46 </span><span style="color: #7f7f7f; text-decor
</span><span style="color: #0000ff; text-decoration-color: #0000ff">raise</s
decoration-color: #00ffff">RuntimeError</span>(<span style="color: #808000;
loading {</span><span style="color: #00ffff; text-decoration-color: #00ffff"
style="color: #808000; text-decoration-color: #808000">}"</span>) <span styl
color: #0000ff">from</span> <span style="color: #00ffff; text-decoration-col
underline">e</span> <span style="color: #800000; text-decorat
<span style="color: #800000; text-decoration-color: #800000">|</span> <spa
decoration-color: #7f7f7f">47 </span>
<span style="color: #800000; text-decoration-color: #800000">|</span>
<span style="color: #800000; text-decoration-color:
#800000">|
<span style="color: #ff0000; text-decoration-color: #ff0000; font-weight: bo
..<span style="color: #800080; text-decoration-color:
#800080">|/.../.../.../tests/integration_tests/examples/</span><span style="c
#ff00ff">example-non-utf8.txt</span>
</pre>

```

The file `example-non-utf8.txt` uses a different encoding the `load()` function fails with a helpful message indicating which file failed decoding.

With the default behavior of `TextLoader` any failure to load any of the documents will fail the whole loading process and no documents are loaded.

B. Silent fail

We can pass the parameter `silent_errors` to the `DirectoryLoader` to skip the files which could not be loaded and continue the load process.

```

loader = DirectoryLoader(path, glob="**/*.txt", loader_cls=TextLoader,
silent_errors=True)
docs = loader.load()

```

```
Error loading ../../../../tests/integration_tests/examples/example-
non-utf8.txt
```

```
doc_sources = [doc.metadata['source'] for doc in docs]
doc_sources
```

```
['../../../../../tests/integration_tests/examples/whatsapp_chat.txt',
'../../../../../tests/integration_tests/examples/example-utf8.txt']
```

C. Auto detect encodings

We can also ask `TextLoader` to auto detect the file encoding before failing, by passing the `autodetect_encoding` to the loader class.

```
text_loader_kwargs={'autodetect_encoding': True}
loader = DirectoryLoader(path, glob="**/*.txt", loader_cls=TextLoader,
loader_kwargs=text_loader_kwargs)
docs = loader.load()
```

```
doc_sources = [doc.metadata['source'] for doc in docs]
doc_sources
```

```
['../../../../../tests/integration_tests/examples/example-non-
utf8.txt',
'../../../../../tests/integration_tests/examples/whatsapp_chat.txt',
'../../../../../tests/integration_tests/examples/example-utf8.txt']
```