

```
1 # Appendix C1 - setup.py
2
3 # Install nltk modules
4 import nltk
5 nltk.download('punkt')
6 nltk.download('stopwords')
7
8 from src.topic_importer import TopicImporter
9 from src.document_importer import DocumentImporter
10 from src.content_importer import ContentImporter
11 import os
12
13 database_name = "klassify"
14 if os.path.exists("%s.db" % database_name):
15     os.remove("%s.db" % database_name)
16
17 # Add topics and subtopics
18 print("Importing topics and subtopics:")
19 TopicImporter().run()
20
21 # Add documents and associate them subtopics
22 print("Importing documents:")
23 DocumentImporter().run()
24
25 print("Importing documents HTML:")
26 ContentImporter().import_documents_html()
27
28 print("Importing documents data:")
29 ContentImporter().extract_documents_content()
30
```