

```
1  #!/usr/bin/python -u
2  import bd
3  import os
4  from os.path import isfile, join, splitext, basename
5
6  # this is the method to search dictionary by key
7  def findItem(obj, key):
8      if key in obj: return obj[key]
9      for k, v in obj.items():
10         if isinstance(v, dict):
11             item = findItem(v, key)
12             if item is not None:
13                 return item
14
15  # this is the method to search key with given string
16  def search(values, searchFor):
17      found = []
18      for k in values:
19         for v in values[k]:
20             if searchFor in v.lower():
21                 found.append(k)
22      return found
23
24  #parameters to access BD services
25  bds = 'https://bd-api.ncsa.illinois.edu'
26
27  # key and token from the Brown Dog API Gateway service
28  token = ""
29
30  # path to the folder to process
31  input_path = ""
32
33  # create a list of files that is not a folder
34  onlyfiles = [join(input_path, f) for f in os.listdir(input_path) if
35               isfile(join(input_path, f))]
36
37  text_store = {}
38
39  # processing the files in the given folder
40  for (input_file) in onlyfiles:
41      filename, file_extension = splitext(basename(input_file))
42      input_format = file_extension[1:]
43
44      print 'Processing file: ' + basename(input_file)
45
46      # do the extraction
47      metadata = bd.extract(bds, input_file, token, 120)
48
49      # process the metadata
50      # find the "OCR" results and store
51      for m in metadata['metadata.jsonld']:
52         txt = findItem(m, 'ocr_simple')
53         if not (txt is None):
54             text_store[basename(input_file)] = txt
55
56  # find the keyword from the ocr results
57  print search(text_store, 'information')
```