## archive.keys from (page

## archive.get file(key)

## find bad articles (archive,

## archive.get pages():

# EXtract



### predicate):

### predicate(lines):

## region name=config.REGION NAME)

### boto3.client('s3',

find bad articles (bucket name,



### config.PREFIX):

## keys from (page

get\_pages(client,

### bucket name,

get file(<mark>client,</mark>

## bucket name,

### predicate):

### predicate(lines):