1. database

Table: etds

add field：

ALTER TABLE etds

ADD COLUMN **pri\_identifier** VARCHAR(100) DEFAULT NULL,

ADD COLUMN **second\_identifier** VARCHAR(100) DEFAULT NULL,

ADD COLUMN **license** VARCHAR(2000) DEFAULT NULL,

ADD COLUMN **copyright** VARCHAR(2000) DEFAULT NULL,

ADD COLUMN **haspdf** TINYINT DEFAULT NULL，

ADD COLUMN **timestamp\_metadata** TIMESTAMP DEFAULT NULL，

ADD COLUMN **timestamp\_pdf** TIMESTAMP DEFAULT NULL;

A screenshot of a computer

Description automatically generated

2. About download\_PDF\_error\_log.txt

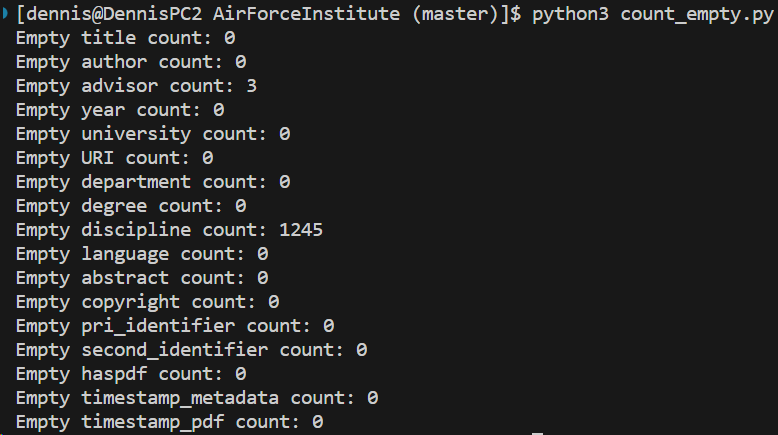
“265.pdf is not downloadable, because the downloadableUrl is None”, In general, this typically refers to the content of this webpage no longer existing, means This paper has been withdrawn. for example:

A screenshot of a computer

Description automatically generated

3. About count\_empty.py

When you run count\_empty.py, it will give you the empty fields count in the table etds:



4. About missing pdf file.

For this college, if the pdf file is non-downloadable, the HTML is often garbled, making it difficult to extract valid data. Therefore, in this school, when a PDF is not present, it automatically skips that file and does not attempt to extract data from the HTML.