How to collect data

To get enough book data, I write a crawler by python to collect those data. https://www.booktxt.net/xiaoshuodaquan/ In this website, there're around 3k books. By the benefits of multi-threading, the crawler can retrieve hundreds of books in just a few minutes.

Here's the main part of my crawler

```python
def main():
    start_time = time.time()
    books = getBooks()
    i = 0
    total_size = len(books)
    while i < 100: # 只爬前100本书，总共三千本书，实在是太多了，没必要
        if(threading.activeCount() < 50):
            thd = threading.Thread(target=getBook,args=(books[i],i))
            i += 1
            thd.start()
        time.sleep(1)
    for thd in threading.enumerate():
        if(thd.getName() != threading.currentThread().getName()):
            thd.join()
    stop_time = time.time()
    diff = stop_time - start_time
    print(diff,"seconds")
    print(diff/60,"minutes")
    print(diff/60/60,"hours")


if __name__ == '__main__':
    main()
```

How to store the content of books

We didn't store the content of books in database, instead, we store the content in some files, and store the uri location in database using a field called `location`.

All book contents will be stored under a directory named `book`. For specific book, there'e filed called `chapter_no` in our book table, which represents the total chapter number of a book. Books will be

stored in a directory with name of its `location`. And each chapter will be stored in a txt file with chapter number as file name.

Also, every book will have some introduction message stored in a file named `info.txt` and a cover image stored in a file named `cover.jpg` within the same directory.

For instance, suppose there a book with following data

```
{
    "location":"qwerty",
    "chapter_no:10
}
```

In this example, the first chapter will be stored in `book/qwerty/1.txt`. The second chapter will be stored in `book/qwerty/2.txt`. Book introduction will be stored in `book/qwerty/info.txt`. Cover image will be stored in `book/qwerty/cover.jpg`.

Where's our database

We use a server to deploy our database, which is at www.irran.top. We can login to this database through commend like this

```
mysql -h www.irran.top -u yggdrasil -p
```

How to modify book content stored in our server

Client application will issue right formatted post request. I design a few servlet written in JAVA to process those request and modify data accordingly.

Here's one of my servlets

```
@Override
public void doPost(HttpServletRequest req, HttpServletResponse resp) throws ServletEx

    try {
        String location = req.getParameter("location");
        String content = req.getParameter("content");
        String path = String.format("%s%s/info.txt", getServletContext().getRealPath(
        FileWriter fw = new FileWriter(path);
        fw.write(content);
        fw.flush();
        fw.close();
    }catch (Exception e) {
        PrintWriter out = resp.getWriter();
        out.println(e.getMessage());
    }
}
```