**1、NoSQL**：非关系数据库，有灵活的数据模型，容易扩展，大数据量等特点。

**2、Pymongo**:

<https://www.jianshu.com/p/b8e7b1aa027c>

https://www.cnblogs.com/liangmingshen/p/9998868.html

**3、安装mongoDB:** https://blog.csdn.net/qq\_34731574/article/details/57416920

* D:\mongoDB\bin>mongod.exe --dbpath D:\mongoDB\data\db （运行mongoDB）
* C:\Program Files\MongoDB\Server\3.2\bin>mongo.exe

2020-12-02T22:02:54.513+0800 I CONTROL [main] Hotfix KB2731284 or later update

is not installed, will zero-out data files

MongoDB shell version: 3.2.4

connecting to: test

**或者：**

mongod --dbpath "D:\mongoDB\data\db" --logpath "D:\mongoDB\log\mongo.log" --install --serviceName "MongoDB"

启动服务：net start MongoDB

关闭服务：net stop MongoDB

**4、Beautiful soup**: “<https://cuiqingcai.com/1319.html>

soup = BeautifulSoup(html)

print soup.prettify()

<div class="page">

<a href="/soft/sort01/index\_2.html">

下页

</a>

<a href="/soft/sort01/index\_798.html">

尾页

</a>

</div>

<div class="intro">

<span>

本书TXT下载简介：

</span>

<p>

莫名其妙为老爹背负一身罪孽诅咒，结果又被一个莫名其妙系统选中，从此只能做一些奇奇怪怪的事情。 哎，我真的太难了啊！

</p>

TXT下载推荐地址：

<i class="intro\_info2">

http://m.uidzhx.com/Shtml85935.html

</i>

</div>

<div class="infotype">

<p>

作者：黑色暴雨

</p>

<p>

类型：玄幻奇幻

</p>

<p>

更新时间：2020-12-01T21:02:26

</p>

<p>

最新章节：

<a href="/du/83/83468/29156752.html">

第706章 冰魔

</a>

<ul class="infolink" id="btnlist">

<li>

<p>

<a href="/du/56/56814/18384815.html">

从头阅读

</a>

</p>

</li>

<li>

<p>

<a href="Javascript:void(0);" onclick="javascript:vote(56814);" rel="nofollow">

推荐本书

</a>

</p>

</li>

<li>

<p>

<a href="Javascript:void(0);" onclick="javascript:putbookcase(56814);" rel="nofollow">

加入书架

</a>

</p>

</li>

<li>

<p>

<script type="text/javascript">

get\_down\_url('http://m.uidzhx.com/Shtml56814.html','http://dzs.uidzhx.com/56/56814/灵界战雄.txt');

</script>

</p>

</li>

<div class="clear">

</div>

</ul>

</p>

</div>

**5、设置的数据库情况**

数据库db

表：

* Novels
* failed\_urls : {'url': self.url}

已记载小说篇数：玄幻小说共计360x20=7200篇

实际：6984

已保存：1088

词典：1102495

Wordtovector：(1088, 113870)