

1、从cnbeta网站上爬数据

1)目标：选取标题(title)、内容(content)、原文链接(url)、创建时间(createtime)字段

2)技术要点：

- 使用urlopen()进行最基本的页面抓取，这时抓取的是网站的源码
- 使用read()进行文件读取
- 通过分析源码发现仅有url是使用的a标签，这样就很方便能取出url字段，使用beautifulsoup库的find_all()方法找到所有匹配结果出现的地方，需要找出所有a标签内容
- 同理可以把title、content、createtime爬出来

2、数据存储

1) 把从cnbeta网站上爬到的数据通过接口的方式传到数据库

```
1 requests.post(url, data=data)
```

2) 把从cnbeta网站上爬到的数据通过向数据库插入数据的方式传到数据库

```
1 db = MySQLdb.connect("127.0.0.1", "root", "", "test", charset='utf8' ) #连接数据库
2 print '连接数据库成功'
3 conn = db.cursor() #获取cursor指针对象，cursor对象的方法有两大类，1.执行命令,2.接收返回值
4 #使用sql语句,这里要接收的参数都用%s占位符.要注意的是,无论你要插入的数据是什么类型,占位符永远都要用%s
5 blog_sql = 'insert into blogs_blogmsg (title,content,date,author) values (%s,%s,%s,%s)'
6 param = [title, p, fdate, author] #param必须是tuple或者list
7 conn.execute(blog_sql,param) #执行sql语句
8 db.commit() #在数据库中更新数据，与之对应的rollback方法是回滚当前所有操作
9 conn.close() #关闭指针对象
10 db.close() #关闭连接对象
```

3、前端展示通过views中render函数传递数据给templates，通过模板语言来展示