
CSDN Search Engine

C 目录 Contents

01

爬虫

02


搜索算法

03

前端 & 后端

04

演示



爬虫

Requests + BeautifulSoup

url : https://blog.csdn.net/diandianxiyu_geek/article/details/83657231

正则表达式匹配url

存储到mysql数据库

模拟登陆

Selenium 模拟点击 + send_keys

Cookies格式转换 + session登陆

20w+数据

扫码登录

账号登录



微信登录

帐号登录

扫码登录

☒ 下次自动登录

忘记密码

登 录

常见开源全文搜索引擎

Lucene (pyLucene)

Elasticsearch

Whoosh

Sphinx

Nutch

Solr

搜索算法： whoosh

Jieba 中文分词解析器

```
from jieba.analyse.analyzer import ChineseAnalyzer
```

多字段索引模式 schema

```
schema = Schema(url=ID(stored=True), title=TEXT(stored=True),  
                nickname=TEXT(stored=True), readcount=TEXT(stored=True)  
                , text=TEXT(stored=True, analyzer=analyzer), time=DATETIME(stored=True))
```

多字段查询词解析器 MultifieldParser 可同时匹配标题、文章内容

```
parser = MultifieldParser(['title', 'text'], schema=ix.schema)
```

搜索器 (searcher) 采用BM25F得分策略

```
with ix.searcher(weighting=scoring.BM25F()) as searcher:  
    query = parser.parse(q)  
    results = searcher.search(query, limit=100)
```

CSDN 搜索

🔍 搜索

热门搜索: [python](#) [人工智能](#) [python入门](#) [django](#) [java](#)

排序规则

[按相关度排序](#)[按时间排序](#)[按阅读量排序](#)

相关链接

[百度搜索](#)[Google](#)[CSDN搜索](#)

找到约 6066 条结果(用时0.355秒), 共约607页

[python](#) 得分: 18.17

python (英语发音: /ˈpaɪθ(ə)n/), 是一种面向对象、直译式电脑编程语言, 也是一种功能强大的通用型语言, 已经具有近二十年的发展历史, 成熟且稳定。它包含了一组完善而且容易理解的标准库, 能够轻松完成很多常见的任务。它的语法非常简捷和清晰, 与其它大多数程序设计语言使用大括号不一样, 它使... [更多](#)

作者: [ajian005](#) 阅读数: 1253 发布时间: 2012-11-21[python](#) 得分: 17.96

python python (KK 英语发音: /ˈpaɪθən/, DJ 英语发音: /ˈpaɪθən/) 是一种面向对象、直译式计算机程序设计语言, 由 Guido van Rossum 于1989年底发明, 第一个公开发行版发行于1991年。**python**语法简捷而清晰, 具有丰富和强大的类库。它常被昵... [更多](#)

作者: [iOS_David_Dai](#) 阅读数: 457 发布时间: 2012-09-25[python](#) 得分: 17.43

virtualenv 这个工具可以隔离不同版本的**python**环境, 在一些服务器上运行不同的**python**应用或脚本可能需要不同的**python**版本, 那不是能用这个东西来隔离, 待测试。。。 [更多](#)

作者: [aoerqileng](#) 阅读数: 49 发布时间: 2018-01-13[python 库](#) 得分: 17.24

python 库 0.导入 要使用**python**模块或者库,必须先导入。 **import this** 导入模块时, **python**依赖一个路径列表, 这个路径列表就是 **sys.path** 可以运行时修改。如: 当前可导入模块列表通过**sys.modules**可以获取, 他是一个字典 **1.** 标准库 **2.** 外部... [更多](#)

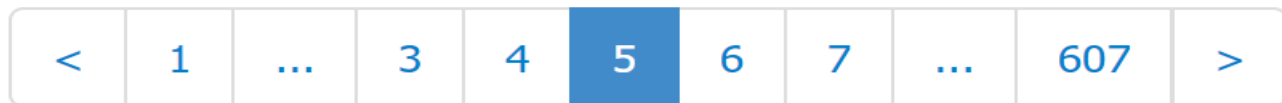
作者: [过去的我](#) 阅读数: 167 发布时间: 2016-03-09

热门搜索

[python](#)[人工智能](#)[python入门](#)[django](#)[java](#)

后端

分页



后端

热门搜索

数据库存储每次query

汇总排序

热门搜索

[python](#)

[人工智能](#)

[python入门](#)

[django](#)

[java](#)

热门搜索: [python](#) [人工智能](#) [python入门](#) [django](#) [java](#)

后端

排序方法

排序规则

按相关度排序

按时间排序

按阅读量排序

相关链接

[python 实现字符串反转的9种方法](#) 得分: 15.44

在做leetcode的试题中，做到反转整数，就涉及到字符串反转，为了尽可能可以写出更多的方法，于是写下这篇文章 样例：如 a='123456789' 反转成 a='987654321' 第一种方法：使用字符串切片 第二种方法：使用reversed() 可读性好，但速度较慢 封装使用 ... [更多](#)

作者: 陈新明 阅读数: 19 发布时间: 2018-09-30

[python 从url中提取域名和path](#) 得分: 14.48

python 从url中提取域名和path [python2代码](#) [python3代码](#) <http://www.chenxm.cc/post/720.html> [更多](#)

作者: 陈新明 阅读数: 34 发布时间: 2018-09-30

[python 中easydict的使用](#) 得分: 14.13

easydict的作用：可以使得以属性的方式去访问字典的值！ Project description EasyDict allows to access dict values as attributes (works recursively). A Javascript-like prop... [更多](#)

作者: 爱上北方的冷冬 阅读数: 14 发布时间: 2018-09-30

[hanlp自然语言处理包的基本使用--python](#) 得分: 13.73

hanlp拥有：中文分词、命名实体识别、摘要关键字、依存句法分析、简繁拼音转换、智能推荐。这里主要介绍一下hanlp的中文分词、命名实体识别、依存句法分析，这里就不介绍具体的hanlp的安装了，百度教程很多，可以看这里：<http://hanlp.com/> 里面也有相关的一些介绍。我以前还... [更多](#)

作者: adnb34g 阅读数: 50 发布时间: 2018-09-28

后端

排序方法

排序规则

按相关度排序

按时间排序

按阅读量排序

相关链接

[七、python 一些基础面试题总结](#) 得分：15.86

1. [python](#)是如何进行内存管理的？ 答:从三个方面来说,一对象的引用计数机制,二垃圾回收机制,三内存池机制 一、对象的引用计数机制 [python](#)内部使用引用计数,来保持追踪内存中的对象,所有对象都有引用计数。 引用计数增加的情况: 1, 一个对象分配一个新名称 ... [更多](#)

作者: [自在逍遥](#) 阅读数: 58936 发布时间: 2014-09-12

[python urllib使用](#) 得分: 14.07

简介: [urllib2](#)是[python](#)的一个获取url (Uniform Resource Locators, 统一资源定址器) 的模块。它用[urlopen](#)函数的形式提供了一个非常简洁的接口。这使得用各种各样的协议获取url成为可能。它同时 也提供了一个稍微复杂的接口来处理常见的状况-如基本的认证, c... [更多](#)

作者: [adrianfeng](#) 阅读数: 40681 发布时间: 2010-09-05

[python 中的@符号](#) 得分: 14.64

周海汉 / 文 2010.4.11 http://blog.csdn.net/ablo_zhou [python](#) 2.4以后, 增加了@符号修饰函数对函数进行修饰, [python](#)3.0/2.6又增加了对类的修饰。 我现在使用的[python](#)版本, 支持对class的修饰: [zhouhh@zh...](#) [更多](#)

作者: [ablo_zhou](#) 阅读数: 37079 发布时间: 2010-04-11

[python IDLE 背景以及字体的修改](#) 得分: 14.03

为了保护眼睛, 决定把白色背景换掉: 1 首先, 在已经下载好的[python](#)文件目录下, 找到config-highlight.def文件, 我是在 H:\python\python3**\Lib\idlelib**文件夹下。 2. 打开文件后, 你会看到一些默认的颜色配置, 比如经典的颜色配置就是白... [更多](#)

作者: [AQ_cainiao_AQ](#) 阅读数: 26553 发布时间: 2016-06-17

找到约 99 条结果(用时0.037秒), 共约10页

[tensorflow入门, 手写数字识别](#) 得分: 34.81

最近忽然想, 学习一下tensorflow。。。 下面这段的代码是tensorflow中文文档里面超过来的, emmmmm算是入门级别的了, 挂上原来的网址: http://wiki.jikexueyuan.com/project/tensorflow-zh/tutorials/mnist_pr... [更多](#)

作者: [abc15766228491](#) 阅读数: 40 发布时间: 2018-05-18

[tensorflow入门 8](#) 得分: 25.70

tensorflow入门教程之0: BigPicture&极速入门 tensorflow入门教程之1: 基本概念以及理解 tensorflow入门教程之2: 安装和使用 tensorflow入门教程之4: 实现一个自创的CNN卷积神经网络 tensorflow入门教程之5: TensorBo... [更多](#)

作者: [a13662080711](#) 阅读数: 26 发布时间: 2018-08-09

[机器学习 人工智能 博文链接汇总](#) 得分: 24.59

115 [[入门问题](#)] [[tensorflow](#)] [[深度学习](#)] [[好玩儿的算法应用实例](#)] [[聊天机器人](#)] [[神经网络](#)] [[机器学习](#)] [[机器学习算法应用实例](#)] [[自然语言处理](#)] [[数据科学](#)] [[Python](#)] [[Java](#)] ... [更多](#)

作者: [Alice熹爱学习](#) 阅读数: 9169 发布时间: 2017-05-13

[tensorflow书推荐, 深度学习书推荐](#) 得分: 22.99

关键词: tensorflow书推荐, 深度学习书推荐 《tensorflow深度学习应用实践》 作者: 王晓华 《21个项目玩转深度学习 基于tensorflowr的实践详解》 作者: 何之源 《深度学习之tensorflow入门、原理与进阶实战》 作者: 李金洪 《零起点Ten... [更多](#)

作者: [achirandliu](#) 阅读数: 7363 发布时间: 2017-06-03

相关搜索

Query : Tensorflow

相关搜索	mnist	DL	MNIST	sklearn
	RNN	感知机	LSTM	kaldi

Query : python入门

相关搜索	Pandas	notebook	Python	sklearn
	Pycharm	scrapy	anaconda	Numpy

演 示

演示完毕 谢谢欣赏

PLEASE ADD YOUR TITLE HERE

