如何从html文件中查找所有<a>标记链接和名称

html bash ubuntu sed

这里是一个测试文件,包含<a>标记中的链接和名称。

/tmp/test_html.txt

```
<a href="http://www.example.com/link1">example link 1</a>

<a href="http://www.example.com/link2">example link 2</a>

<a href="http://www.example.com/link3">example link 3</a>

<a href="http://www.example.com/link3">example link 3</a>

<ta href="https://www.example.com/4/0/1/40116601-1FDC-real-world-link/bar" target="_blank" class="real-world-link/bar" target="_blank" class="_blank" class="_b
```

以下命令可以找到文件中的所有链接,但无法将link和name打印在一起:

如何在Bash、grep或批处理中删除HTML文件的所有链接并将其存储在文本文件中

```
# sed -n 's/.*href="\([^"]*\).*/\1/p' /tmp/test_html.txt
```

我希望该命令可以先用name打印all links line by line,然后再使用href。

以下是预期输出:

```
# sed <...command....> /tmp/test_html.txt

example link 1 | http://www.example.com/link1
example link 2 | http://www.example.com/link2
example link 3 | http://www.example.com/link3
Real World Link | https://www.example.com/4/0/1/40116601-1FDC-real-world-link/bar
```

如何编写sed命令?

- 发布于 4 月前

/ 最佳回答:



这可能适用于您(GNU sed):

```
sed -En 's/.*href="([^{n}]^{*})"[^{>}]^{*}([^{<}]^{*})<.*/2 | ^1/p' file
```

使用-n选项筛选行,并使用{%12}选项使regexp更容易。

匹配包含href的行,后跟内部文本,并根据需要使用反向引用设置格式。

本站已为你智能检索到如下内容,以供参考:

楊 相关问答 7 个

如何链接"类"名称在相同的HTML文件的HREF?

在laravel中为标记a选择链接

如何使<a>链接标记与提交按钮一起工作?

如何分离和<a>HTML标记

我无法以html链接html文件

vuejs表路由器链接标记a

如何创建带有<a>标记的新链接?

≝ 相关教程 1 个

HTML 入门教程

易 推荐阅读 3 个

<u>Linkify是一个零依赖JavaScript插件,用于在纯文本中查找链接并将其转换为HTML<a>标记。</u>

HTML 链接

A better HTML 5 parser for PHP。

本文支持英文版本,如需查看请 <mark>点击这里!</mark>

(查看英文版本获取更加准确信息)

公众号: 1024技术圈

提供互联网知识和资讯,分享IT前沿技术,热门资源,大厂面试题

