



友万科技

www.uone-tech.cn

subinfile

处理网页源代码的利器





*cntrade: 下载股票交易数据

*chinafin: 下载上市公司财务数据

cnintraday: 下载上市公司分时交易数据

*cnstock: 下载股票代码

*chinagcode与chinaaddress: 通过百度地图API将中文地址与经纬度相互转换

*subinfile: 修改文本文件

*wordcovert: docx、doc、rtf、pdf等文件之间相互转换

psemail: 发送邮件

eventstudy: 事件研究

ttable2: 分组t检验

*结果输出: reg2docx、sum2docx、corr2docx、t2docx

即将发布: table2docx、fillup、wordcloud、mapscatter





最初目的：批量修改do文件路径

名字由来：substr()、substitute in file

- 功能：
- 1、保留含有指定字符串或匹配到正则表达式的行
 - 2、替换指定的子字符串或正则表达式匹配到的子字符串
 - 3、删除空行

意外发现：简化处理源代码，提取信息的程序

类似：file命令、rewrite





1.0版：infix版

1.1版：修正了部分bug，如运行过程中临时文件read-only的问题

2.0版：mata版





options:

- ① index(string) specifies the line which contains it will be kept. Those lines without the key string specified by index() option will be dropped.
- ② indexregex specifies that the contents you specify in index() is to be interpreted as a regular expression.
- ③ from(string) and to(string) specifies the string which is to be replaced whereas the to() option specifies the new string which will be used to replace the old one.
- ④ fromregex specifies that the contents you specify in from() is to be interpreted as a regular expression.
- ⑤ dropempty drops the empty line. If you specify both from() and dropempty, Stata will first replace the string you specify and then drop the empty line.
- ⑥ save(string) specifies the path and the file name to be saved. If you do not specify the format of the file, it will be saved as .txt by default.
- ⑦ replace permits save to overwrite an existing file which is not read-only. If you do not specify the option save(string), the original file will be replaced.

If you specify the option index(string), from(string) and dropempty in one command at the same time, the option index(string) will be executed first, then from(string), and dropempty at last.





- 1、获取页面链接
- 2、获取网页源代码：copy、curl
- 3、读入源代码：infix、import delimited、fileread()
- 4、处理源代码：→ subinfile
 - ① 保留信息所在行
 - ② 提取所需信息或删除多余内容





1、新浪财经高管任职数据:

http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpManager/stockid/600900.phtml

2、Statalist:

<https://www.statalist.org/forums/forum/general-stata-discussion/general>

3、Stata命令:

<https://ideas.repec.org/s/boc/bocode.html>

4、NBER论文信息:

<http://www.nber.org/papers/w20001>





新浪财经高管任职数据抓取（单网页）：

1、获取网页链接：

从多个链接中寻找规律，部分网页抓取需要两次爬虫，第一次抓取网页链接(深交所年报、中国土地市场网)。

新浪财经高管任职数据：

长江电力：

http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpManager/stockid/600900.phtml

万科A：

http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpManager/stockid/000002.phtml

东信和平：

http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpManager/stockid/002017.phtml





2、可行性分析:

能否从源代码中找到所需要提取的信息。

历届高管成员			
姓名	职务	起始日期	终止日期
张定明	总经理	2015-05-06	—
陈国庆	副总经理	2011-07-05	—
薛福文	副总经理	2011-07-05	—
李平诗	副总经理	2015-09-28	—
关杰林	副总经理	2015-09-28	—
谢峰	财务总监	2015-05-06	—
李绍平	董事会秘书	2016-08-30	—

[↑返回页顶↑](#)

历届董事会成员			
姓名	职务	起始日期	终止日期
第4届董事会 起始日期: 2015-05-06 终止日期: 2018-05-05			
卢纯	董事长	2015-05-06	2018-05-05
卢纯	非独立董事	2015-05-06	2018-05-05
张诚	副董事长	2015-05-06	2018-05-05
张诚	非独立董事	2015-05-06	2018-05-05
杨亚	非独立董事	2015-05-06	2018-05-05
张定明	非独立董事	2015-05-06	2018-05-05
李季泽	非独立董事	2015-05-06	2018-05-05
洪文洁	非独立董事	2016-05-20	2018-05-05
宗仁怀	非独立董事	2016-05-20	2018-05-05
黄宁	非独立董事	2016-05-20	2018-05-05
周传根	非独立董事	2016-05-20	2018-05-05
赵燕	非独立董事	2016-05-20	2018-05-05
张崇久	独立董事	2015-05-06	2018-05-05
吕振勇	独立董事	2015-10-16	2018-05-05

```

<tr>
  <td class="ct" width="25%"><div align="center"><strong>姓名</strong></div></td>
  <td class="ct" width="25%"><div align="center"><strong>职务</strong></div></td>
  <td class="ct" width="25%"><div align="center"><strong>起始日期</strong></div></td>
  <td class="ct"><div align="center"><strong>终止日期</strong></div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=张定明">张定明</a></div></td>
  <td class="ccl"><div align="center">总经理</div></td>
  <td class="ccl"><div align="center">2015-05-06</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=陈国庆">陈国庆</a></div></td>
  <td class="ccl"><div align="center">副总经理</div></td>
  <td class="ccl"><div align="center">2011-07-05</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=薛福文">薛福文</a></div></td>
  <td class="ccl"><div align="center">副总经理</div></td>
  <td class="ccl"><div align="center">2011-07-05</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=李平诗">李平诗</a></div></td>
  <td class="ccl"><div align="center">副总经理</div></td>
  <td class="ccl"><div align="center">2015-09-28</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=关杰林">关杰林</a></div></td>
  <td class="ccl"><div align="center">副总经理</div></td>
  <td class="ccl"><div align="center">2015-09-28</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=谢峰">谢峰</a></div></td>
  <td class="ccl"><div align="center">财务总监</div></td>
  <td class="ccl"><div align="center">2015-05-06</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=李绍平">李绍平</a></div></td>
  <td class="ccl"><div align="center">董事会秘书</div></td>
  <td class="ccl"><div align="center">2016-08-30</div></td>
  <td class="ccl"><div align="center">--</div></td>
</tr>

```





3、获取网页源代码

copy命令、curl

```
copy "http://vip.stock.finance.sina.com.cn/corp/go.php/vCI_CorpManager/stockid/600900.phtml" temp.txt, replace
```

```

<!doctype html>
<html>
<head>
<title>长江电力(600900)公司高管_新浪财经_新浪网</title>
<meta name="Keywords" content="长江电力公司高管,600900公司高管,新浪财经长江电力(600900)公司高管" />
<meta name="Description" content="新浪财经长江电力(600900)行情中心,为您提供长江电力(600900)公司高管信息数据查询。" />
<meta http-equiv="Content-Type" content="text/html; charset=gb2312" />

<link rel="stylesheet" type="text/css" href="http://vip.stock.finance.sina.com.cn/corp/view/style/hangqing.css" />
<link rel="stylesheet" type="text/css" href="http://vip.stock.finance.sina.com.cn/corp/view/style/dadan.css" />
<script src="http://www.sinaimg.cn/dy/js/jquery/jquery-1.7.2.min.js"></script>
<script type="text/javascript" src="http://finance.sina.com.cn/basejs/tool.js"></script>
<script type="text/javascript" src="http://finance.sina.com.cn/basejs/dataDrawer.js"></script>
<script src="http://l.sso.sina.com.cn/js/ssologin.js" type="text/javascript"></script>
<script src="http://finance.sina.com.cn/realstock/company/hotstock_daily_a.js"></script>

<script src="http://finance.sina.com.cn/realstock/company/sh600900/jsvar.js"></script>
<script type="text/javascript">
    var page_name = "公司高管";
    /* BHPznk7Cm94l1mlLT9oBbUksAQI/tgPKy65jyFVorJxi+BI093Qt4241xf9wBWP1GcpKaSbXdlW/qND10BRMwXt;HUVq5Wk1PxrRudYiHSMhE2rd+G4J8tJTsDMDuXXBGall/JHo5/+DqKHxz6WVozaqAVi0vTC008Tg== */
    //HOTSTOCK
    var hl_str_CFF_LIST="IF1309,IF1310,IF1312,IF1403";
    var bkSymbol = "-";
    var wbAppKey = '3202088101';
    var mq_msgy = 0;
    var flashURL = "http://finance.sina.com.cn/flash/cn.swf";

    //相关期货
    var RS = 0;
    RS.corr_future = [];

    /* aakt0nE90ukLEw8saxPDzCRA32ofQubzSCS/mEhxJJ8CdKwS00pRS/XEuxulvKqRNGdT/GY7cRcGaiEZz4Y8usVUHA/KsVBPk14xJZPQW708TOLuJ91W0i+jpDFU/MIG4ITQIstE2uejazG4F8ciE+d+qXI2XFf+h/InzRQvIq3PR3GmVNOnl4xTmRk6g,liu800RvbfIiCxt0I,jk9qQ== */

    //综合评级级别
    var gradeLevel = 0;
    //综合评级研究报告数量 ( TODO PHP写进页面)
    var gradeAmt = 0;
    //新股发行 增发 配股 现金分红
    var bonus=[0,0,0,0];

    /* 9P9+e1VhuYr1M5AvHjt577edAmFLWT46g9UKY8oyOFZyLqWU5U0c6WojvDUP1G/VqWML1qq8CXkzAZMYjv2dkjLmSmtNqj2EOPRA2YCA11KL/qm0X2I7Z1CYNUh8kVpccsindMuPvJqPNIT/G1Nn5v7hJaxHr9qVcnzwbwWQptDhrccow3tbUe752LxxKw9e4PspURwSf6P1rS1cc
    +ihj608r1elhxosuZ8Q== */
</script>

<!-- 环球市场滚动条, 依赖jquery, tool, dataDrawer-->
<script type="text/javascript" src="http://finance.sina.com.cn/basejs/global_index_scroller.js"></script>

<!-- 搜索建议, 无依赖-->
<script type="text/javascript" src="http://finance.sina.com.cn/basejs/suggestServer.js"></script>

<!-- 登录层, 无依赖-->

```





4、读入源代码

(1) gb2312读入Stata14要先进行转码:

```
unicode encoding set gb18030
```

```
unicode translate temp.txt, transutf8 // 文件前不能跟路径，必须在工作路径下
```

```
unicode erasebackups, badidea // 固定用法，删除备份文件
```

(2) 使用infix或者import delimited命令读入源代码:

```
infix strL v 1-100000 using temp.txt, clear
```

```
import delimited using temp.txt, clear delimiters("asgdhjbaiucbiuabconobwivquviqcboqn", asstring) encoding("utf-8")
```

注：源代码只有一行情况下直接用fileread()函数读入；源代码最后一行有所需信息时需要先使用file命令加上一个回车符:

```
tempname temp
```

```
file open `temp' using temp.txt, write text append
```

```
file write `temp' _n
```

```
file close `temp'
```





5、保留需要提取的信息所在的行:

keep if index(v, "</div></td>")

drop if index(v, "") | v == "</div></td>

```

...
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=谢峰">谢峰</a></div></td>
    <td class="ccl"><div align="center">财务总监</div></td>
    <td class="ccl"><div align="center">2015-05-06</div></td>
    <td class="ccl"><div align="center"></div></td>
  </tr>
<tr>
  <td class="ccl"><div align="center">
    <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=李绍平">李绍平</a></div></td>
    <td class="ccl"><div align="center">董事会秘书</div></td>
    <td class="ccl"><div align="center">2016-08-30</div></td>
    <td class="ccl"><div align="center"></div></td>
  </tr>
</tbody>
</table>
<table width="100%" border="0" align="center" cellpadding="0" cellspacing="0" class="table2">
<tr>
<td height="30" align="right" valign="middle" style="color:#009">↑<a href="#top">返回页顶</a>↑</td>
</tr>
</table>
<a name="adminTable2"></a>
<table id="comInfo1" width="100%" id="Table2">
<thead>
<tr>
<th class="ct" colspan="4">历届董事会成员</th>
</tr>
</thead>
<tbody>
<tr>
<td class="ct" width="25%"><div align="center"><strong>姓名</strong></div></td>
<td class="ct" width="25%"><div align="center"><strong>职务</strong></div></td>
<td class="ct" width="25%"><div align="center"><strong>起始日期</strong></div></td>
<td class="ct"><div align="center"><strong>终止日期</strong></div></td>
</tr>
<tr>
<td class="ct" colspan="4"><div align="center">
第4届董事会 起始日期:2015-05-06 终止日期:2018-05-05
</div></td>
</tr>
<tr>
<td class="ccl"><div align="center">
<a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=卢纯">卢纯</a></div></td>
<td class="ccl"><div align="center">董事长</div></td>
<td class="ccl"><div align="center">2015-05-06</div></td>
<td class="ccl"><div align="center">2018-05-05</div></td>
</tr>
<tr>
<td class="ccl"><div align="center">
<a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=卢纯">卢纯</a></div></td>
<td class="ccl"><div align="center">非独立董事</div></td>
<td class="ccl"><div align="center">2015-05-06</div></td>
<td class="ccl"><div align="center">2018-05-05</div></td>
</tr>
...

```



6、删除标签（尖括号内的字符），提取所需信息：

```
replace v = ustrregexra(v, "<.*?>", "")
```

```
1. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=张定明">张定明</a></div></td>
2.     <td class="ccl"><div align="center">总经理</div></td>
3.     <td class="ccl"><div align="center">2015-05-06</div></td>
4.     <td class="ccl"><div align="center">—</div></td>
5. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=陈国庆">陈国庆</a></div></td>
6.     <td class="ccl"><div align="center">副总经理</div></td>
7.     <td class="ccl"><div align="center">2011-07-05</div></td>
8.     <td class="ccl"><div align="center">—</div></td>
9. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=薛福文">薛福文</a></div></td>
10.    <td class="ccl"><div align="center">副总经理</div></td>
11.    <td class="ccl"><div align="center">2011-07-05</div></td>
12.    <td class="ccl"><div align="center">—</div></td>
13. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=李平诗">李平诗</a></div></td>
14.     <td class="ccl"><div align="center">副总经理</div></td>
15.     <td class="ccl"><div align="center">2015-09-28</div></td>
16.     <td class="ccl"><div align="center">—</div></td>
17. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=关杰林">关杰林</a></div></td>
18.     <td class="ccl"><div align="center">副总经理</div></td>
19.     <td class="ccl"><div align="center">2015-09-28</div></td>
20.     <td class="ccl"><div align="center">—</div></td>
21. <a target="_blank" href="/corp/view/vCI_CorpManagerInfo.php?stockid=600900&Name=谢峰">谢峰</a></div></td>
22.     <td class="ccl"><div align="center">财务总监</div></td>
23.     <td class="ccl"><div align="center">2015-05-06</div></td>
24.     <td class="ccl"><div align="center">—</div></td>
```



```
1. 张定明
2.  总经理
3.  2015-05-06
4.  —
5.  陈国庆
6.  副总经理
7.  2011-07-05
8.  —
9.  薛福文
10.  副总经理
11.  2011-07-05
12.  —
13.  李平诗
14.  副总经理
15.  2015-09-28
16.  —
17.  关杰林
18.  副总经理
19.  2015-09-28
20.  —
21.  谢峰
22.  财务总监
23.  2015-05-06
24.  —
```





7、将提取到的信息进行整理:

(1) post命令
















(2) 也可以用如下命令:

```
forvalues j = 1/3 {
    gen v`j' = v[_n + `j']
}
keep if mod(_n, 4) == 1
rename (v - v3) (姓名 职务 起始日期 终止日期)
```

	姓名	职务	起始日期	终止日期
1.	张定明	总经理	2015-05-06	--
2.	陈国庆	副总经理	2011-07-05	--
3.	薛福文	副总经理	2011-07-05	--
4.	李平诗	副总经理	2015-09-28	--
5.	关杰林	副总经理	2015-09-28	--
6.	谢峰	财务总监	2015-05-06	--
7.	李绍平	董事会秘书	2016-08-30	--
8.	卢纯	董事长	2015-05-06	2018-05-05
9.	卢纯	非独立董事	2015-05-06	2018-05-05
10.	张诚	副董事长	2015-05-06	2018-05-05
11.	张诚	非独立董事	2015-05-06	2018-05-05
12.	杨亚	非独立董事	2015-05-06	2018-05-05
13.	张定明	非独立董事	2015-05-06	2018-05-05
14.	李季泽	非独立董事	2015-05-06	2018-05-05
15.	洪文浩	非独立董事	2016-05-20	2018-05-05
16.	宗仁怀	非独立董事	2016-05-20	2018-05-05
17.	黄宁	非独立董事	2016-05-20	2018-05-05
18.	周传根	非独立董事	2016-05-20	2018-05-05
19.	赵燕	非独立董事	2016-05-20	2018-05-05
20.	张崇久	独立董事	2015-05-06	2018-05-05
21.	吕振勇	独立董事	2015-10-16	2018-05-05
22.	曹广晶	董事	2010-06-26	2014-04-09
23.	卢纯	董事长	2014-04-25	2015-05-05
24.	卢纯	董事	2014-04-25	2015-05-05



2、Statalist

 Sticky: Stata 15 is here Started by Bill Gould (StataCorp) , 06 Jun 2017, 09:55		70	9,514	by Joseph Coveney Yesterday, 21:23 
 Multinomial Logistic Regression Started by andy macrobarty , 14 Aug 2017, 11:54		10	105	by Marcos Almeida Today, 08:23 
 How do I perform a fuzzy match between datasets using numeric values? Started by Michael Anbar , Yesterday, 14:20		7	52	by Robert Picard Today, 07:45 
 Fixed-Effect regression help Started by reece ogier , Yesterday, 10:14		5	33	by reece ogier Today, 07:38 
 Event Study with r2000 'no observations' coding issue Started by Huw Simpson , Today, 07:34		1	16	by Huw Simpson Today, 07:34 
 multiple lines graph with if conditions Started by Edwin Wong , Today, 04:37		2	16	by Nick Cox Today, 04:58 



```
<li class="current section-item">  
  <a href="https://www.statalist.org/forums/" class="h-left navbar_home">Forums</a>  
  
  <span class="channel-tabbar-divider"></span>  
  
  <span class="mobile dropdown-icon"><span class="icon h-right"></span></span>  
</li>
```

401

```
<li class="notice restore last-child" data-notice-id="1" data-notice-persistent="1">You are not  
logged in. You can browse but not post. Login or <a  
href="https://www.statalist.org/forums/register">Register</a> by clicking 'Login or Register' at the  
top-right of this page. For more information on Statalist, see the <a  
href="http://www.statalist.org/forums/help" target="_blank"><b>FAQ</b></a>. </li>
```





3、Stata命令

```
230 <LI class="list-group-item downfree"><B>S458393 <A HREF="/c/boc/bocode/s458393.html">CODEBOOK_RIPPER: Stata module  
to convert metadata managed in a spreadsheet into do files with renaming, notes as well as variable and value  
labelling</A></B><BR><I>by</I> Niels Henrik Bruun  
231  
232 <LI class="list-group-item downfree"><B>S458392 <A HREF="/c/boc/bocode/s458392.html">CPRDCRIBS: Stata module  
providing template do-files for inputting CPRD datasets into Stata</A></B><BR><I>by</I> Roger Newson  
233  
234 <LI class="list-group-item downfree"><B>S458391 <A HREF="/c/boc/bocode/s458391.html">AGGIND: Stata module to  
aggregate indicators among units within a specified radius</A></B><BR><I>by</I> Andreas Hartung  
235  
236 <LI class="list-group-item downfree"><B>S458390 <A HREF="/c/boc/bocode/s458390.html">THSEARCH: Stata module to  
evaluate threshold search model for non-linear models based on information criterion</A></B><BR><I>by</I> Ho Fai  
Chan & Brenda Gannon & David Harris & Mark Harris  
237  
238 <LI class="list-group-item downfree"><B>S458389 <A HREF="/c/boc/bocode/s458389.html">STD_BETA: Stata module to  
calculate centered and standardized coefficients after estimation</A></B><BR><I>by</I> Doug Hemken  
239  
240 <LI class="list-group-item downfree"><B>S458388 <A HREF="/c/boc/bocode/s458388.html">WALD_MSE: Stata module to  
calculate the maximum mean square error (MSE) of a point estimator of the mean</A></B><BR><I>by</I> Chuck Manski &  
Max Tabord-Meehan  
241  
242 <LI class="list-group-item downfree"><B>S458387 <A HREF="/c/boc/bocode/s458387.html">FLOWCHART: Stata module to  
generate subject disposition flow diagram figures in LaTeX using PGF/TikZ to include in manuscripts</A></B><BR>  
<I>by</I> Isaac M. E. Dodd  
243  
244 <LI class="list-group-item downfree"><B>S458386 <A HREF="/c/boc/bocode/s458386.html">ENCODEFROM: Stata module to  
rename and encode variables using an external crosswalk</A></B><BR><I>by</I> Sally Hudson
```



NBER论文信息抓取（单个网页）：

1、获取网页链接：

<http://www.nber.org/papers/w20001>

<http://www.nber.org/papers/w20002>

<http://www.nber.org/papers/w20003>

每个网页只有最后的编号不同。



3、获取网页源代码：

```
copy "http://www.nber.org/papers/w20001" temp.txt, replace
```

4、读入网页源代码：

```
infix strL v 1-100000 using "temp.txt", clear
```



5、保留提取信息所在的行:

```
keep if ustrregexm(v, `"(</h1>)|(</b><br>)|(</small></a></p>)"`) ///
      | index(v[_n - 1], "</h1>") ///
      | index(v[_n - 1], `"<p style="margin-left: 40px; margin-right: 40px; text-align: justify">"`)
```

v

```
1. <h1 class='title'>Option Value of Work, Health Status, and Retirement Decisions in Japan: Evidence from the Japanese Study on Aging and Retirement (JSTAR)</h1>
2. <h2 class='bibtop' style='text-align:center'><a href="/people/satoshi_shimizutani">Satoshi Shimizutani</a>, <a href="/people/takashi_oshio">Takashi Oshio</a>, <a href="/people/fujiil">Mayu Fujii</..
3. <b>NBER Working Paper No. 20001</b><br>
4. <b>Issued in March 2014</b><br>
5. This study examined the factors that affect the retirement decisions of the middle-aged and elderly in Japan, focusing especially on their earnings, public pension benefits, and health status. Usi..
6. <td align="center" valign="middle"><p style="line-height: 8px;"><a href="http://www.nber.org/papers/w20001.pdf" target="_blank">", "")
```

v

1. Option Value of Work, Health Status, and Retirement Decisions in Japan: Evidence from the Japanese Study on Aging and Retirement (JSTAR)
2. Satoshi Shimizutani, Takashi Oshio, Mayu Fujii
3. NBER Working Paper No. 20001
4. Issued in March 2014
5. This study examined the factors that affect the retirement decisions of the middle-aged and elderly in Japan, focusing especially on their earnings, public pension benefits, and health status. Usi..
6. <http://www.nber.org/papers/w20001.pdf>



7、整理抓取的信息：

```
sxpose, clear
```

```
rename (_var1 - _var6) (Title Author NBER_No IssuedTime Abstract URL)
```



- 1、“利器”主要体现在处理源代码中每行一条所需信息时
- 2、保留所需信息特征不在该行时无法处理



- 1、from()选项中的文本替换成不同的内容，多个from对应多个to(新版功能)
- 2、正则表达式的扩充(pcre)
- 3、删除掉对应的行(新版功能)

