数据分析与可视化

缩略引文的映射

秦晴　2020年5月6日

THE MAIN CONTENTS

01  题录数据中的的引文缩略格式

02  如何实现缩略引文与全文的映射

第 1 部分

题录数据中的的引文缩略格式

01 题录数据中的的引文缩略格式

各数据库题录数据引文情况对比

数据库	题录信息是否含有引文	含有的引文形式	引文分析				全文映射形式
			是否设有对引文的分析功能	引文分析功能	功能详情字段限制	分析呈现情况	是否链接全文
CNKI	否		否				是
万方	否		否				是
维普	否		否				是
Scopus	是	Aharonson, B.S., Schilling, M.A., Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution (2016) Res. Policy, 45 (1), pp. 81-96;	是	引文概览	1、期范围 2、排除作者的自我引用 3、排除书籍的引用	1、年份-引文量折线图 2、年份-引文量统计表	是
WOS	是	Aharonson BS, 2016, RES POLICY, V45, P81, DOI 10.1016/j.respol.2015.08.001	否				是
Springer link	否						

题录数据下载中有引文信息只有Web of Science和Scopus

01 题录数据中的的引文缩略格式

Web of science

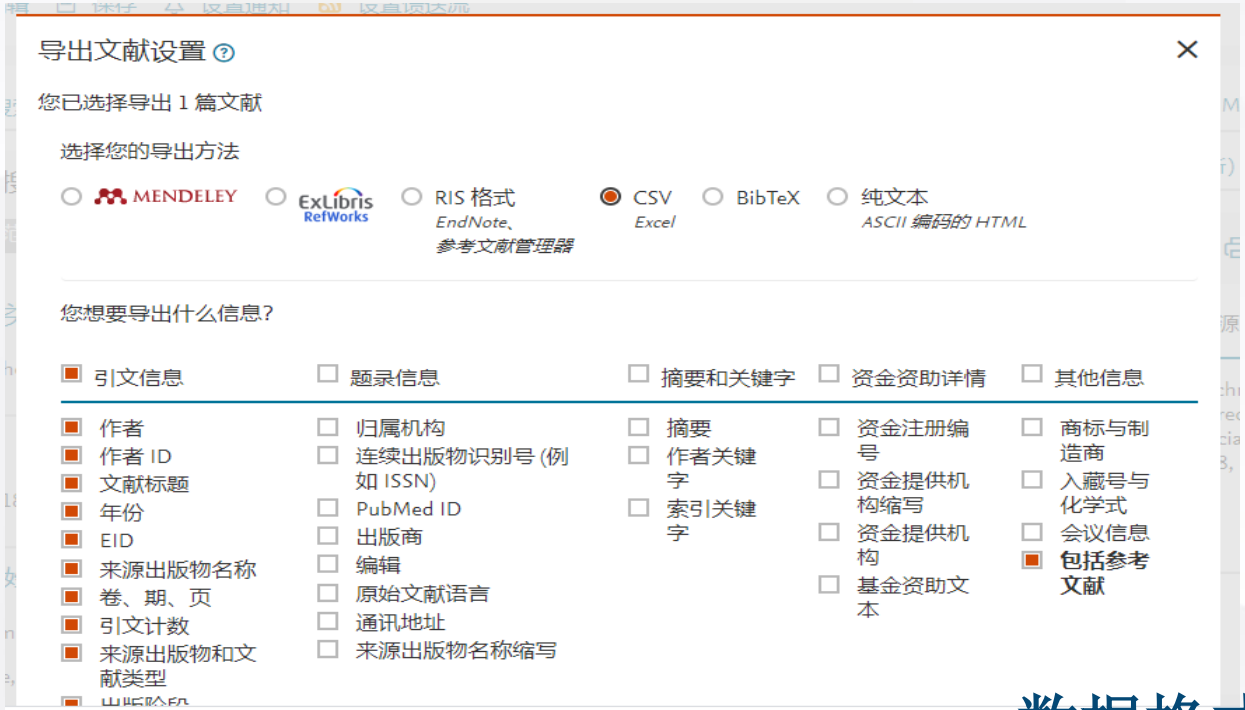
数据格式

CR Aharonson BS, 2016, RES POLICY, V45, P81, DOI 10.1016/j.respol.2015.08.001
Albino V, 2014, APPL ENERG, V135, P836, DOI 10.1016/j.apenergy.2014.08.012
Balconi M, 2004, RES POLICY, V33, P127, DOI 10.1016/S0048-7333(03)00108-2
Banuls VA, 2008, TECHNOVATION, V28, P103, DOI 10.1016/j.technovation.2007.05.006
Bermudez-Edo M, 2013, ADV ENG INFORM, V27, P358, DOI 10.1016/j.aei.2013.02.003
Bierwisch A, 2015, TECHNOL FORECAST SOC, V101, P226, DOI 10.1016/j.techfore.2015.06.014
Boccardi F, 2014, IEEE COMMUN MAG, V52, P74, DOI 10.1109/MCOM.2014.6736746

savedrecs.txt

01 题录数据中的的引文缩略格式

scopus



引文概览可视化

数据格式

参考文献: Aharonson, B.S., Schilling, M.A., Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution (2016) Res. Policy, 45 (1), pp. 81-96;
Albino, V., Ardito, L., Dangelico, R.M., Petruzzelli, A.M., Understanding the development trends of low-carbon energy technologies: a patent analysis (2014) Appl. Energy, 135, pp. 836-854;
Balconi, M., Breschi, S., Lissoni, F., Networks of inventors and the role of academia: an exploration of Italian patent data (2004) Res. Policy, 33 (1), pp. 127-145;
Bañuls, V.A., Salmeron, J.L., Foresighting key areas in the information technology industry (2008) Technovation, 28 (3), pp. 103-111;
Bermudez-Edo, M., Noguera, M., Hurtado-Torres, N., Hurtado, M.V., Garrido, J.L., Analyzing a firm's international portfolio of technological knowledge: a declarative ontology-based OWL approach for patent documents (2013) Adv. Eng. Inform., 27 (3), pp. 358-365;
Bierwisch, A., Kayser, V., Shala, E., Emerging technologies in civil security—a scenario-based analysis (2015) Technol. Forecast. Soc., 101, pp. 226-237;

第 2 部分

如何实现缩略引文与全文的映射

02 研究结果及应用如何实现缩略引文与完整引文的映射

什么是映射

映射是指两个元素的集之间元素相互“对应”的关系。

映射的成立条件

1．定义域的遍历性：X中的每个元素x在映射的值域中都有对应对象
2．对应的唯一性：定义域中的一个元素只能与映射值域中的一个元素对应。

映射如何应用于缩略引文与完整引文

使缩略引文与引文完整的题录信息实现一一对应关系。

02 研究结果及应用如何实现缩略引文与全文的映射

实践： Web of Science

方法一：EXCEL处理：缩略引文的下载，引文格式的处理，（分列、替换...复制粘贴为文本格式）

作者	时间	期刊名	刊号	页码	doi
Aharonson BS	2016	RES POLICY	V45	P81	10.1016/j.respol.2015.08.001
Albino V	2014	APPL ENERG	V135	P836	10.1016/j.apenergy.2014.08.012
Balconi M	2004	RES POLICY	V33	P127	10.1016/S0048-7333(03)00108-2
Banuls VA	2008	TECHNOVATION	V28	P103	10.1016/j.technovation.2007.05.006
Bermudez-Edo M	2013	ADV ENG INFORM	V27	P358	10.1016/j.aei.2013.02.003
Bierwisch A	2015	TECHNOL FORECA	V101	P226	10.1016/j.techfore.2015.06.014
Boccardi F	2014	IEEE COMMUN MA	V52	P74	10.1109/MCOM.2014.6736746
Breitzman A	2015	RES POLICY	V44	P195	10.1016/j.respol.2014.06.006
Caviggioli F	2016	TECHNOVATION	V55-56	P22	10.1016/j.technovation.2016.04.003
Cho C	2016	R&D MANAGE	V46	P13	10.1111/radm.12107
Choi S	2014	TECHNOL ANAL ST	V26	10.1080/09537325.2013.850477	
Cozzens S	2010	TECHNOL ANAL ST	V22	P361	10.1080/09537321003647396
Ernst H	1997	SMALL BUS ECON	V9	P361	10.1023/A:1007921808138
Ernst H	1998	J ENG TECHNOL M	V15	P279	10.1016/S0923-4748(98)00018-6
Ernst H.	2003	World Patent Infor	V25	P233	10.1016/S0172-2190(03)00077-2

主要题录数据为：

作者

时间

期刊名

刊号

页码

DOI

根据web of science 的【期刊简写对照表】进行缩略引文中期刊简写的扩展

02 研究结果及应用如何实现缩略引文与完整引文的映射

Web of science期刊简写扩展成果

根据web of science 的【期刊简写对照表】进行缩略引文中期刊简写的扩展，使用vlookup查询函数，在对照表中查找期刊完整名称

作者	时间	期刊名（简写）	刊号	页码	doi	完整期刊名
Aharonsc	2016	RES POLICY	V45	P81	10.1016/j.respol.2015.08.001	RESEARCH POLICY
Albino V	2014	APPL ENERG	V135	P836	10.1016/j.apenergy.2014.08.012	APPLIED ENERGY
Balconi M	2004	RES POLICY	V33	P127	10.1016/S0048-7333(03)00108-2	RESEARCH POLICY
Banuls VA	2008	TECHNOVATIC	V28	P103	10.1016/j.technovation.2007.05.006	TECHNOVATION
Bermude	2013	ADV ENG INFC	V27	P358	10.1016/j.aei.2013.02.003	ADVANCED ENGINEERING INFORM
Bierwisch	2015	TECHNOL FOR	V101	P226	10.1016/j.techfore.2015.06.014	TECHNOLOGICAL FORECASTING AN
Boccardi	2014	IEEE COMMUN	V52	P74	10.1109/MCOM.2014.6736746	IEEE COMMUNICATIONS MAGAZIN
Breitzmar	2015	RES POLICY	V44	P195	10.1016/j.respol.2014.06.006	RESEARCH POLICY
Caviggiol	2016	TECHNOVATIC	V55-56	P22	10.1016/j.technovation.2016.04.003	TECHNOVATION
Cho C	2016	R&D MANAGE	V46	P13	10.1111/radm.12107	R & D MANAGEMENT
Choi S	2014	TECHNOL ANA	V26		10.1080/09537325.2013.850477	TECHNOLOGY ANALYSIS & STRATE
Cozzens S	2010	TECHNOL ANA	V22	P361	10.1080/09537321003647396	TECHNOLOGY ANALYSIS & STRATE
Ernst H	1997	SMALL BUS EC	V9	P361	10.1023/A:1007921808138	SMALL BUSINESS ECONOMICS
Ernst H	1998	J ENG TECHNC	V15	P279	10.1016/S0923-4748(98)00018-6	JOURNAL OF ENGINEERING AND TE

02 研究结果及应用如何实现缩略引文与全文的映射

方法二：使用DOI批量下载引文的完整题录数据

DO=(10.1016/j.respol.2015.08.001 OR
10.1016/j.apenergy.2014.08.012
OR
10.1016/S0048-7333(03)00108-2
OR
10.1016/j.techfore.2009.06.001
OR
10.1016/j.ijhydene.2014.05.006
.....
10.1007/s11192-011-0543-2）

参考文献全

02 研究结果及应用如何实现缩略引文与完整引文的映射

实践：scopus

序号	缩略引文
1	Aharonson, B.S., Schilling, M.A., Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution (2016) Res. Policy, 45 (1), pp. 81-96
2	Albino, V., Ardito, L., Dangelico, R.M., Petruzzelli, A.M., Understanding the development trends of low-carbon energy technologies: a patent analysis (2014) Appl. Energy, 135, pp. 836-854
3	Balconi, M., Breschi, S., Lissoni, F., Networks of inventors and the role of academia: an exploration of Italian patent data (2004) Res. Policy, 33 (1), pp. 127-145
4	Bañuls, V.A., Salmeron, J.L., Foresighting key areas in the information technology industry (2008) Technovation, 28 (3), pp. 103-111
5	Bermudez-Edo, M., Noguera, M., Hurtado-Torres, N., Hurtado, M.V., Garrido, J.L., Analyzing a firm's international portfolio of technological knowledge: a declarative ontology-based OWL approach for patent documents (2013) Adv. Eng. Inform., 27 (3), pp. 358-365
6	Bierwisch, A., Kayser, V., Shala, E., Emerging technologies in civil security—a scenario-based analysis (2015) Technol. Forecast. Soc., 101, pp. 226-237
7	Boccardi, F., Heath, R.W., Lozano, A., Marzetta, T.L., Popovski, P., Five disruptive technology directions for 5G (2014) IEEE Commun. Mag., 52 (2), pp. 74-80
8	Breitzman, A., Thomas, P., The emerging clusters model: a tool for identifying emerging technologies across multiple patent systems (2015) Res. Policy, 44 (1), pp. 195-205
9	Caviggioli, F., Technology fusion: identification and analysis of the drivers of technology convergence using patent data (2016) Technovation, 55-56, pp. 22-32

主要题录数据为：

作者（完整）
题名（完整）
时间
期刊（缩写）
刊号
页码

数据相对完整，缺失完整期刊全称

02 研究结果及应用如何实现缩略引文与完整引文的映射

方法三：自动化探索的过程

02 研究结果及应用如何实现缩略引文与完整引文的映射

test(1).py - C:\Users\DELL\Desktop\test(1).py (3.6.5)

File Edit Format Run Options Window Help

```
import requests


def get_sea_volume():
    headers = {"Content-Type": "application/json",
               "User-Agent": "Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebK
                             "Chrome/58.0.3029.110 Safari/537.36",
               'Cookie': 'bm_sz_-_.webofknowledge.com=CFB977656D772750D016BD3CA8
               }
    uri = "http://apps-webofknowledge-com.vpn.cau.edu.cn:8118/Search.do?product=
    rep = requests.get(url=uri, headers=headers)
    print("执行成功")
    print(rep.text)


if __name__ == '__main__':
    get_sea_volume()
```

Ln: 18 Col: 20

Python 3.6.5 Shell

File Edit Shell Debug Options Window Help

```
Python 3.6.5 |Anaconda, Inc.| (default, Mar 29 2018, 13:32:41) [MSC v.1900 64 b
it (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:\Users\DELL\Desktop\test.py ==========
执行成功！！！！Surprise！！！
>>>
========== RESTART: C:\Users\DELL\Desktop\test(1).py ==========
执行成功
<!DOCTYPE html>

                                                                <html lang
="en"><script src=/sf-webproxy/api/vpn-config></script><script src=/sf-webproxy
/resource/web_proxy.js?v=1.4.7></script><script src=/sf-webproxy/resource/postM
essage.js></script>
                                                                <head> <script
> (function(siteId){ (function(h,o,t,j,a,r){ h.hj=h.hj||function(){(h.hj.q=h.h
j.q||[]).push(arguments)}; h._hjSettings={hjid:siteId,hjsv:6}; a=o.getElementsB
yTagName('head')[0]; r=o.createElement('script');r.async=1; r.src=t+h._hjSettin
gs.hjid+j+h._hjSettings.hjsv; a.appendChild(r); })(window,document,'https://sta
tic.hotjar.com/c/hotjar-','.js?sv='); })("1496794"); </script><link rel="icon"
href="http://images-webofknowledge-com.vpn.cau.edu.cn:8118//WOKRS535R52/images/
wok_favicon.ico" type="image/x-icon"/> <title>Web of Science [v.5.35] - Al
l Databases Results </title><link rel="stylesheet" href="http://images-webofkn
owledge-com.vpn.cau.edu.cn:8118//WOKRS535R52/css/WoKcommon.css" type="text/css"
/><link rel="stylesheet" href="http://images-webofknowledge-com.vpn.cau.edu.cn:
8118//WOKRS535R52/css/WoKcomponents.css" type="text/css" /><link rel="styleshee
t" href="http://images-webofknowledge-com.vpn.cau.edu.cn:8118//WOKRS535R52/css/
Font-SourceSansPro/css/source-sans-pro.min.css" type="text/css" /><link rel="st
ylesheet" href="http://images-webofknowledge-com.vpn.cau.edu.cn:8118//WOKRS535R
52/css/Font-Noto-Kufi-Arabic/css/noto-kufi-arabic.css" type="text/css" /><link
rel="stylesheet" href="http://images-webofknowledge-com.vpn.cau.edu.cn:8118//WO
KRS535R52/css/Font-WOS-Custom-Icons/css/wos-custom-icons.css" type="text/css" /
><link rel="stylesheet" href="http://images-webofknowledge-com.vpn.cau.edu.cn:8
118//WOKRS535R52/css/ui-lightness/jquery-ui.css" type="text/css" /><link rel="s
tylesheet" href="http://images-webofknowledge-com.vpn.cau.edu.cn:8118//WOKRS535
R52/css/jquery-ui-wok.css" type="text/css" /><link rel="stylesheet" href="http:
//images-webofknowledge-com.vpn.cau.edu.cn:8118//WOKRS535R52/css/centralSignin.
css" type="text/css" /><script type="text/javascript" src="http://images-webofk
nowledge-com.vpn.cau.edu.cn:8118//WOKRS535R52/javascript/fa-magic.js"> </script
```

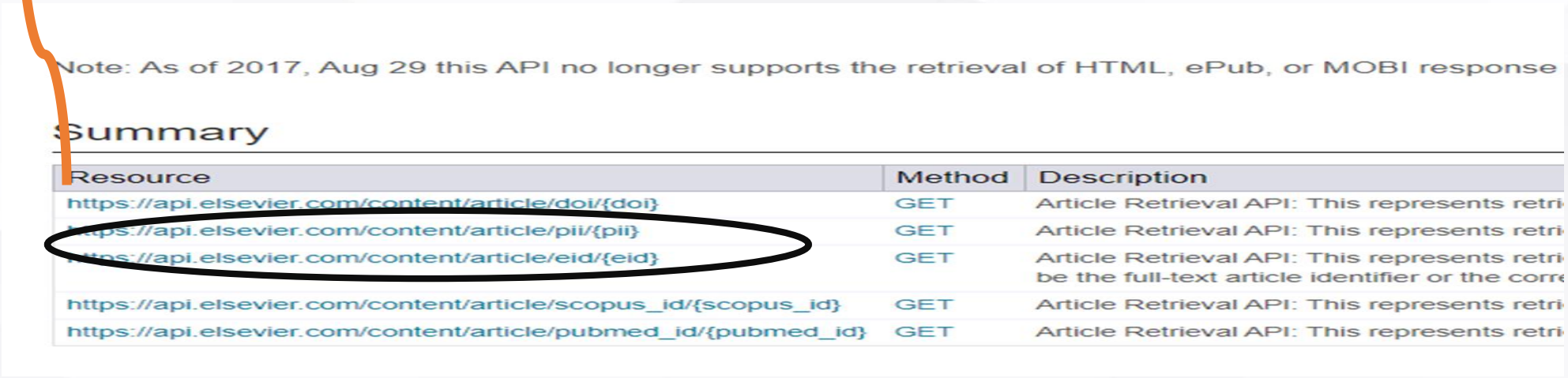
Ln: 306 Col: 39943

实践：实现自动化映射的过程

调用接口信息：提供API的调用地址(通常为URL格式)，这个地址就类似于帮助我们定位想取的包裹在储物柜的哪一行哪一列。请求获取数据：利用HTTP协议请求传递数据，通常会调用到python中request包里的get函数。

https://api.elsevier.com/content/article/doi/{doi}

文章检索API： 这表示通过DOI（文档对象标识符）检索全文文章。



Note: As of 2017, Aug 29 this API no longer supports the retrieval of HTML, ePub, or MOBI response

Summary

Resource	Method	Description
https://api.elsevier.com/content/article/doi/{doi}	GET	Article Retrieval API: This represents retri
https://api.elsevier.com/content/article/pii/{pii}	GET	Article Retrieval API: This represents retri
https://api.elsevier.com/content/article/eid/{eid}	GET	Article Retrieval API: This represents retri be the full-text article identifier or the corre
https://api.elsevier.com/content/article/scopus_id/{scopus_id}	GET	Article Retrieval API: This represents retri
https://api.elsevier.com/content/article/pubmed_id/{pubmed_id}	GET	Article Retrieval API: This represents retri

02 研究结果及应用如何实现缩略引文与完整引文的映射

Authentication API

Product Specific APIs

ScienceDirect APIs	Scopus APIs	Engineering Village APIs	Em
ScienceDirect Search V2 [*Search Tips*] Article Metadata [*Search Tips*]	Affiliation Search [*Search Tips*] Author Search [*Search Tips*] Scopus Search [*Search Tips*]	Engineering Village Search API	EM

Summary

Resource	Method	Description
https://api.elsevier.com/content/metadata/article	GET	Article Search API: This represents a field restricted search against ScienceDirect, which cor ScienceDirect is a leading full text scientific database offering journal articles and book cha 20,000 books. This API resource allows for the submission of field restricted Boolean querie relevant result metadata in user-specific text formats.

02 研究结果及应用如何实现缩略引文与完整引文的映射

02 研究结果及应用如何实现缩略引文与完整引文的映射

实践：DOI号批量执行下载命令

第一步：将题录数据的DOI号有效分列出来，形成一列。
第二步：利用python读取EXCEL的DOI列表
第三步：利用网页直接取，下载txt格式

```
# coding:utf-8
import xlrd
class ExcelUtil():
    def __init__(self, excelPath, sheetName):
        self.data = xlrd.open_workbook(excelPath)
        self.table = self.data.sheet_by_name(sheetName)
        # 获取第一行作为key值
        self.keys = self.table.row_values(0)
        # 获取总行数
        self.rowNum = self.table.nrows
        # 获取总列数
        self.colNum = self.table.ncols

    def dict_data(self):
        if self.rowNum <= 1:
            print("总行数小于1")
                return r

if __name__ == "__main__":
    filepath = "D:\\test\\web-project\\5ke\\testdata.xlsx"
    sheetName = "Sheet1"
    data = ExcelUtil(filepath, sheetName)
    print(data.dict_data())
                for x in range(self.colNum):
                    s[self.keys[x]] = values[x]
                r.append(s)
                j+=1
```

ADD YOUR SUBTITLE HERE

谢谢您的观看与聆听

秦晴

2020.5