



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

# 题录数据中的引文缩略格式

# 01 题录数据中的引文缩略格式

## 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

## Web of science

### 将记录导出至文件

☐ 页面上的所有记录

☐ 记录来源: 1 至 2

一次不超过 500 条记录。

记录内容:

全记录与引用的参考文献

作者、标题、来源出版物

作者、标题、来源出版物、摘要

全记录

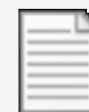
全记录与引用的参考文献

取消

导出

## 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

导出文献设置

您已选择导出 1 篇文献

选择您的导出方法

☐ MENDELEY

☐ ExLibris RefWorks

☐ RIS 格式  
EndNote、  
参考文献管理器

☒ CSV  
Excel

☐ BibTeX

☐ 纯文本  
ASCII 编码的 HTML

您想要导出什么信息?

☒ 引文信息

☐ 题录信息

☐ 摘要和关键字

☐ 资金资助详情

☐ 其他信息

☒ 作者

☒ 作者 ID

☒ 文献标题

☒ 年份

☒ EID

☒ 来源出版物名称

☒ 卷、期、页

☒ 引文计数

☒ 来源出版物和文献类型

☒ 出版年份

☐ 归属机构

☐ 连续出版物识别号 (例如 ISSN)

☐ PubMed ID

☐ 出版商

☐ 编辑

☐ 原始文献语言

☐ 通讯地址

☐ 来源出版物名称缩写

☐ 摘要

☐ 作者关键字

☐ 索引关键字

☐ 资金注册编号

☐ 资金提供机构缩写

☐ 资金提供机构

☐ 基金资助文本

☐ 商标与制造商

☐ 入藏号与化学式

☐ 会议信息

☒ 包括参考文献

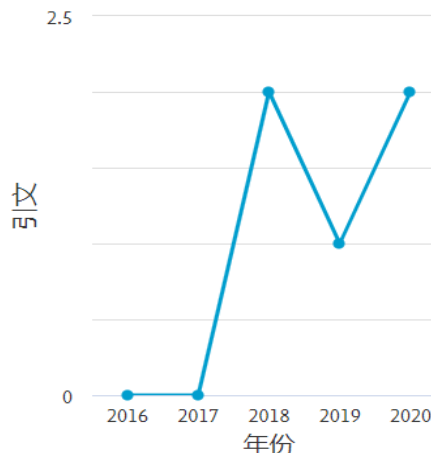
## 数据格式

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;

## 引文概览可视化

1 篇被引文献 + 添加到列表

日期范围: 2016 to 2020 ☐ 排除所有作者的自我引用 ☐ 排除书籍中的引用 更新



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
Cavaggioli 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

## Web of science期刊简写扩展成果

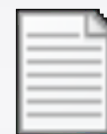
根据web of science 的【期刊简写对照表】进行缩略引文中期刊简写的扩展，使用vlookup查询函数，在对照表中查找期刊完整名称

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	TECHNOVATIO	V28	P103	10.1016/j.technovation.2007.05.006	TECHNOVATION
Bermude	2013	ADV ENG INFO	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	TECHNOVATIO	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 & STRATEC
Cozzens	2010	TECHNOL ANA	V22	P361	10.1080/09537321003647396	TECHNOLOGY ANALYSIS & STRATEC
Ernst H	1997	SMALL BUS EC	V9	P361	10.1023/A:1007921808138	SMALL BUSINESS ECONOMICS
Ernst H	1998	J ENG TECHN	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



## 方法三：自动化探索的过程

The screenshot shows a web browser window with multiple tabs open, including 'EasyConnect', 'Web of Science', 'API Article (Full)', 'dev.elsevier', '使用api下载', '如何使用API', 'python爬虫', and 'python3爬虫'. The active tab is 'python3爬虫', displaying a CSDN article page. The article title is 'python3爬虫系列20之反爬需要登录的网站三种处理方式'. The author is 'csdnzoutao' with a 'code age of 3 years'. The article has 215 likes, 1617 views, 984 comments, 2408 shares, and 113k+ followers. The article content discusses the first method for handling anti-crawling on websites requiring login: 'Cookie method: requests directly carry cookies information'. It explains that when you log in once, you can see the content you want, and you don't need to log in again. It also mentions that the server gives a Cookie to every user, and you can pass it to the server to get data directly. The article also mentions 't and ip proxy pool usage, actually this is a situation, so, if you want to write a persistent collection'.

EasyConnect x Web of Science x API Article (Full) x dev.elsevier x 使用api下载 x 如何使用API x python爬虫 x python3爬虫 x

blog.csdn.net/itbiggod/article/details/103347093

CSDN 首页 博客 学院 下载 论坛 问答 活动 专题 招聘 APP VIP会员 搜博主文章 创作中心 登录/注册

csdnzoutao  
码龄3年  
215 1617 984 2408 113万+

### python3爬虫系列20之反爬需要登录的网站三种处理方式

#### 第一招Cookie法: requests直接携带cookies信息

简单来说  
你平常在网站的时候, 你只要登录一次, 就可以一直看到你想要的内容, 过了一阵子才需要再次登录。  
或者下次打开仍然在登录状态中的?

因为就是每一个使用这个网站的人, 服务器都会给他一个 Cookie, 那么下次你再请求数据的时候, 你顺带把这个 Cookie 传过去, 服务器一看有登录过直接返回数据给他。

展开

#### 的网站三种处理方式

t 和 ip代理池的使用, 实际上这是  
况, 所以, 如果要写持久型采集

授课模式: 在线直播+课后录播, 授课内容  
包含: python人工智能+python全栈  
+python自动化+爬虫

打开

#### 热门文章

Model、ModelMap和ModelAndView的  
使用详解 @ 65572

在以往的案例中, 我们都是爬那些不需要登录或者登陆要求不高的网站。  
那么当你在爬某些网站的时候, 需要你登录才可以获取数据, 怎么办?  
登录的常见方法无非是这两种

- 1、让你输入 帐号和密码登录
- 2、让你输入 帐号密码+验证码登录

今天先说第一种问题的处理办法。

举报

```

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) AppleWebKit/537.36",
               "Cookie": "bm_sz_-.webofknowledge.com=CFB977656D772750D016BD3CA8"}
    uri = "http://apps-webofknowledge-com.vpn.cau.edu.cn:8118/Search.do?product=rep"
    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="s
tylesheet" 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
<a href="https://api.elsevier.com/content/article/doi/{doi}">https://api.elsevier.com/content/article/doi/{doi}</a>	GET	Article Retrieval API: This represents retrieval of the full-text article identifier or the corresponding DOI
<a href="https://api.elsevier.com/content/article/pii/{pii}">https://api.elsevier.com/content/article/pii/{pii}</a>	GET	Article Retrieval API: This represents retrieval of the full-text article identifier or the corresponding DOI
<a href="https://api.elsevier.com/content/article/eid/{eid}">https://api.elsevier.com/content/article/eid/{eid}</a>	GET	Article Retrieval API: This represents retrieval of the full-text article identifier or the corresponding DOI
<a href="https://api.elsevier.com/content/article/scopus_id/{scopus_id}">https://api.elsevier.com/content/article/scopus_id/{scopus_id}</a>	GET	Article Retrieval API: This represents retrieval of the full-text article identifier or the corresponding DOI
<a href="https://api.elsevier.com/content/article/pubmed_id/{pubmed_id}">https://api.elsevier.com/content/article/pubmed_id/{pubmed_id}</a>	GET	Article Retrieval API: This represents retrieval of the full-text article identifier or the corresponding DOI

## Authentication API

## Product Specific APIs

## ScienceDirect APIs

ScienceDirect Search V2 [*Search Tips*]  
Article Metadata [*Search Tips*]

## Scopus APIs

Affiliation Search [*Search Tips*]  
Author Search [*Search Tips*]  
Scopus Search [*Search Tips*]

## Engineering Village APIs

Engineering Village Search API

Em

EM

## Summary

Resource	Method	Description	API	EM
<a href="https://api.elsevier.com/content/metadata/article">https://api.elsevier.com/content/metadata/article</a>	GET	Article Search API: This represents a field restricted search against ScienceDirect, which contains over 20,000 books. This API resource allows for the submission of field restricted Boolean queries and returns relevant result metadata in user-specific text formats.	API	EM



GET		Client-Token		associated with the request. This replaces the old versions of the token <i>CR-TDM-Client-Token</i> and <i>CR-Prospect-Client-Token</i> . Additional information about use of Crossref and publishers can be found at <a href="#">Crossref Text and Data Mining</a> "	
request		CR-TDM-Client-Token	xsd	responses	
header params		X-ELS-ReqId	xsd	<b>status:</b> 200	
Accept	xsd	X-ELS-ResourceVersion	opt	<b>representations</b> text/xml Article XML Response: This is XML representing the requested document. <a href="#">XML Example</a>	
query params		<b>status:</b> 200		<b>representations</b> application/json Article JSON Response: This is JSON representing the requested document. <a href="#">JSON Example</a>	
Authorization	xsd	httpAccept	xsd:string options: application/plain,	<b>status:</b> 200	
X-ELS-APIKey	xsd	access_token	xsd:string	<b>representations</b> application/pdf Article PDF Response: This is a PDF representing the requested document.	
X-ELS-Authtoken	xsd	insttoken	xsd:string	<b>status:</b> 200	
X-ELS-Insttoken	xsd	apiKey	xsd:string	<b>representations</b> application/png Article First Page PNG Response: This is a PNG image representing the first page of the requested document. <a href="#">PNG Example</a>	
CR-Clickthrough-Client-Token	xsd	reqId	xsd:string	<b>status:</b> 200	
CR-TDM-Client-Token	xsd	ver	xsd:string	<b>representations</b> application/rdf+xml Article RDF Response: This is RDF representing the requested document. <a href="#">RDF Example</a>	
X-ELS-ReqId	xsd	view	xsd:string default options:		
view	xsd	FULL, F			

## 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

```
1 # coding:utf-8
2 import xlrd
3 class ExcelUtil():
4     def __init__(self, excelPath, sheetName):
5         self.data = xlrd.open_workbook(excelPath)
6         self.table = self.data.sheet_by_name(sheetName)
7         # 获取第一行作为key值
8         self.keys = self.table.row_values(0)
9         # 获取总行数
10        self.rowNum = self.table.nrows
11        # 获取总列数
12        self.colNum = self.table.ncols
13
14    def dict_data(self):
15        if self.rowNum <= 1:
16            print("总行数小于1")
17            return r
18
19    if __name__ == "__main__":
20        filepath = "D:\\test\\web-project\\5ke\\testdata.xlsx"
21        sheetName = "Sheet1"
22        data = ExcelUtil(filepath, sheetName)
23        print(data.dict_data())
24        for x in range(self.colNum):
25            s[self.keys[x]] = values[x]
26            r.append(s)
27            j+=1
```



ADD YOUR SUBTITLE HERE

**感谢您的观看与聆听**



秦晴

2020.5

