

基于关联数据的数字人文实践

Digital Humanities Applications Based on Linked Data

上海图书馆的探索

The Exploration of Shanghai Library

上海图书馆 夏翠娟

2016-06-20

DH

Digital Humanities

数字人文

现代数字技术+工具+方法：辅助人文科学研究



Digitizing the Humanities A Future for Libraries

*Supporting Digital Humanities for Knowledge Acquisition in Modern Libraries.
Kathleen L. Sacco, etc. 2015*

从“数字图书馆”到“数字人文”

数字图书馆

数字化加工

数字化形态

文献管理和保存

文献查阅服务

数字人文

内容分析知识组织

基于知识的逻辑

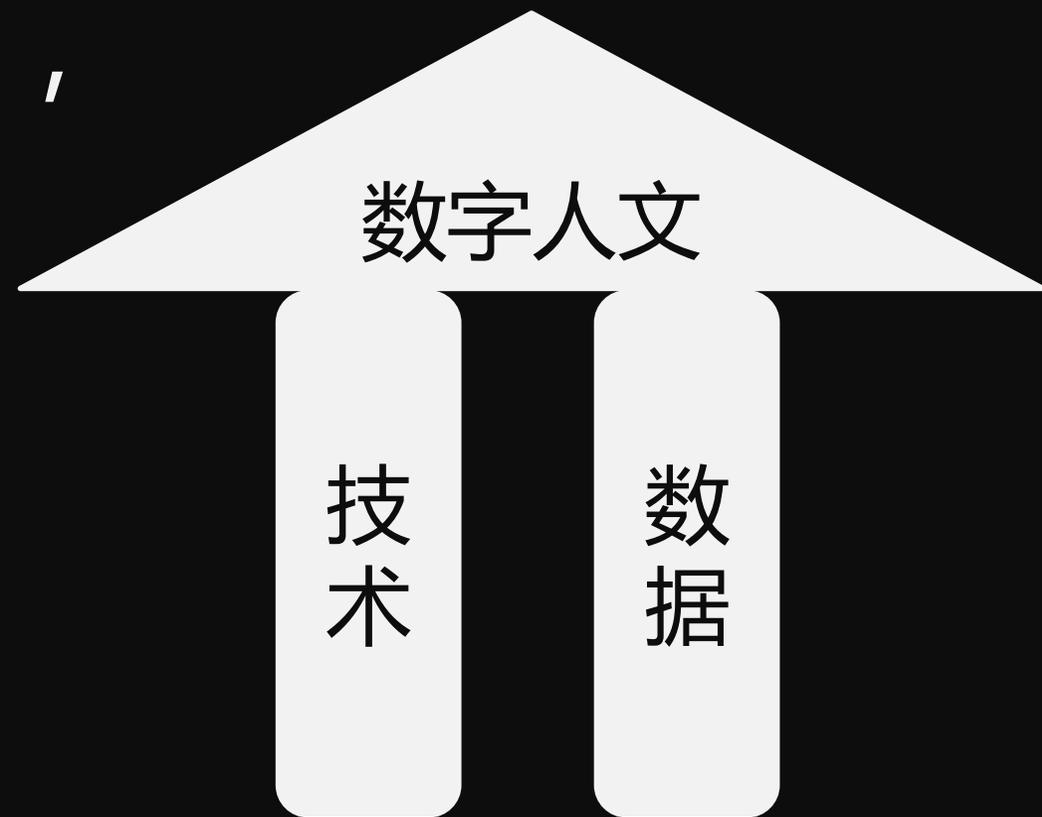
科学研究方法和工具

人文研究和知识服务

数字人文是数字图书馆建设达到一定规模后的必然发展方向。

技术：大数据，长数据，
sns，云计算，语义网...

数据：知识单元的的细
粒度化，知识组织的语义化，
知识呈现的可视化。



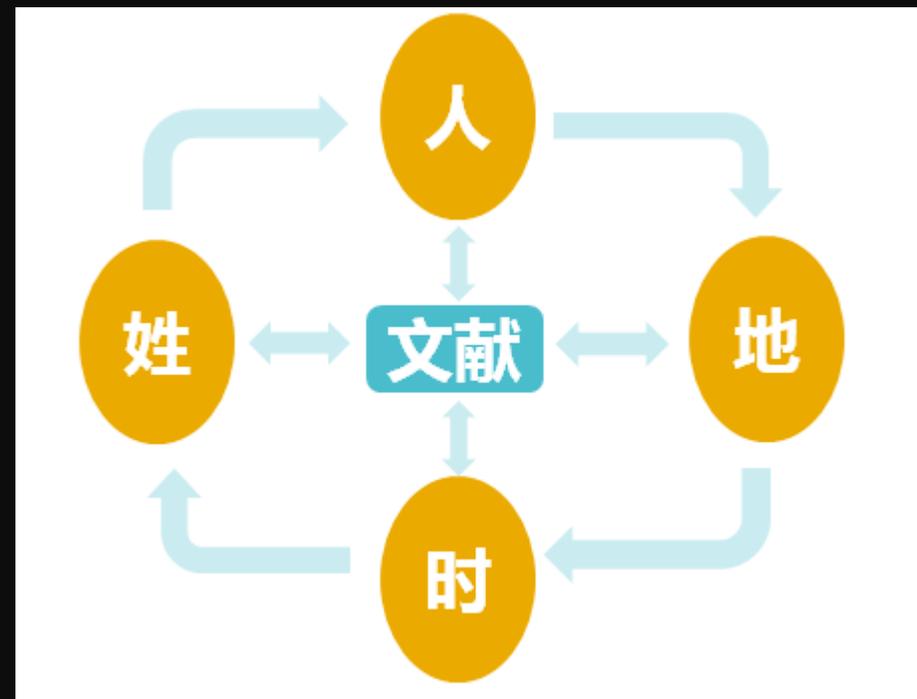
关联数据

(Linked Data)

知识单元的的细粒度化

知识组织的语义化

数据格式的标准化的



LOD

Linked
Open
Data

LOD

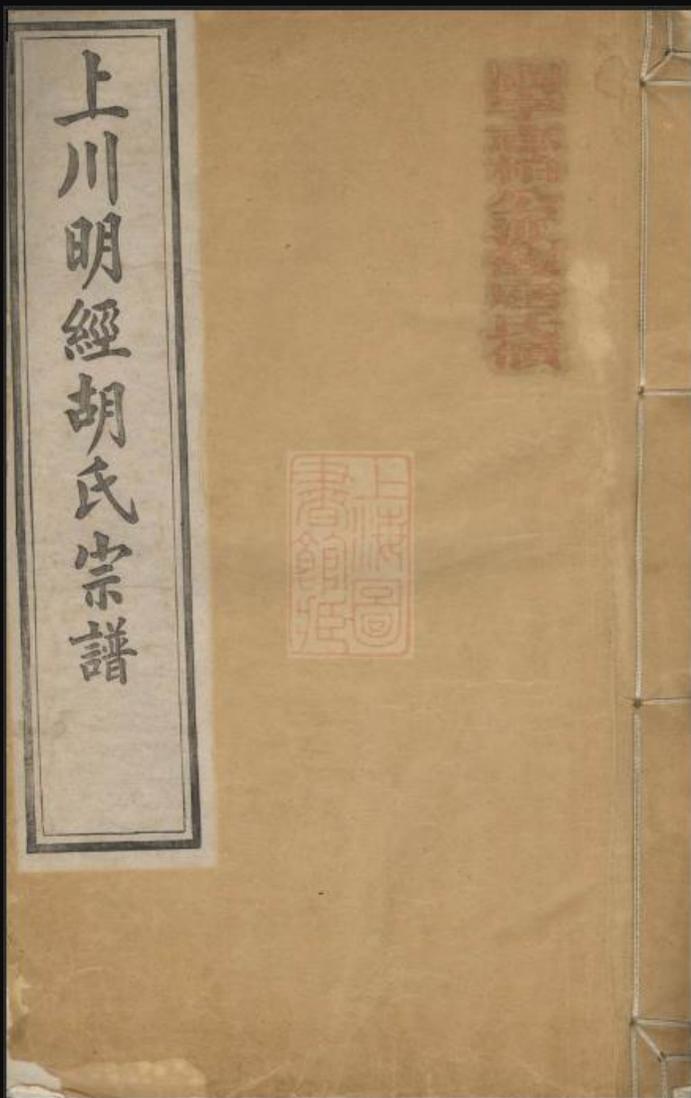
技术：LD+SNS+可视化

方法：规范控制、知识组织

数据：元数据、全文、百科/辞典/GIS、用户贡献内容.....

从元数据记录中提取知识

以胡适先生家族（上川明经胡氏）迁徙路线为例



 明经胡氏續修宗谱不分卷（江西省上
婺源

【副题名】 明經胡氏續修宗譜

【责任者】 胡璘

【纂修时间】 明嘉靖二十三年（1544）

【摘要】 始祖昌翼，谱稱本唐昭宗幼子，母何后，天祐元年生，避禍來婺源考川，遂冒義父胡三公姓。後唐同光間，登明經進士。後世子孫散居新安諸邑、遂安、淳安、樂平、鄱陽、浮梁、興安、漢陽，乃遠至貴州、雲南者，世稱明經胡氏，以別安定胡氏。是為明經胡氏統宗譜。譜存譜序、墓圖及宋胡次焱撰《論始祖》、《論姓氏》、《論過房》、《論稱呼》等。書名據版心、卷端題。

【版本】 刻本

【数量】 存三册

上图 山西家谱

馆藏地：上海图书馆
地址：上海市 徐汇区 淮海中路1555号
DOI为STJP011208：殘本（索书号：911895-97）請至上海图书馆家谱阅览室借閱

面向文献，
以元数据记录
(MARC
Record) 为
基本单位

知识单元的的细粒度化

人物：胡昌翼

时间：

唐昭宗天佑元年(公元904年)

始发地：长安

目的地：婺源考川

人物：胡延政

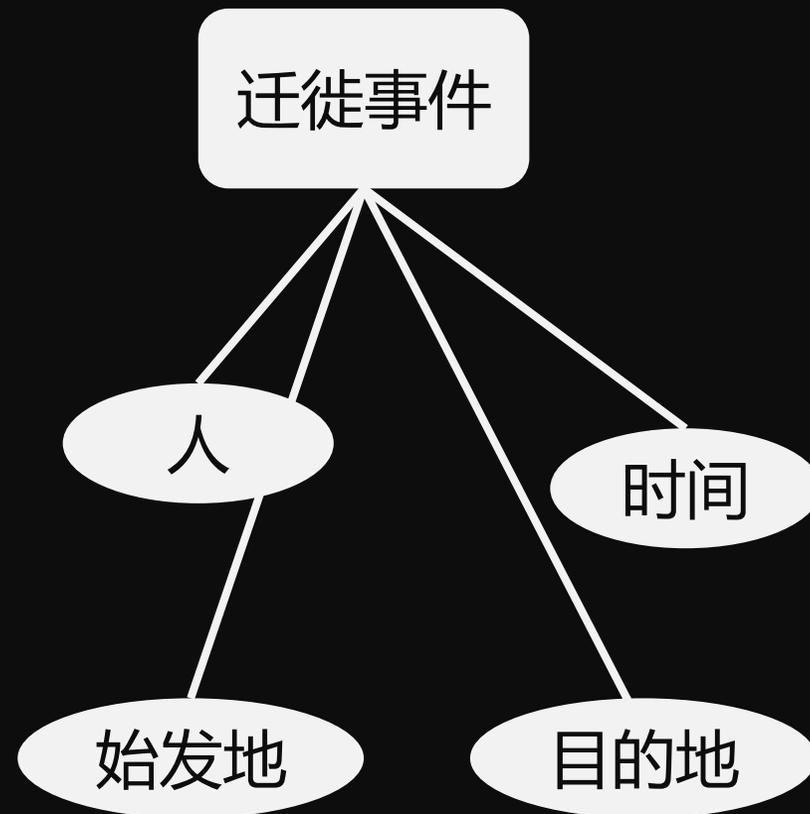
时间：

北宋开宝八年(975年)

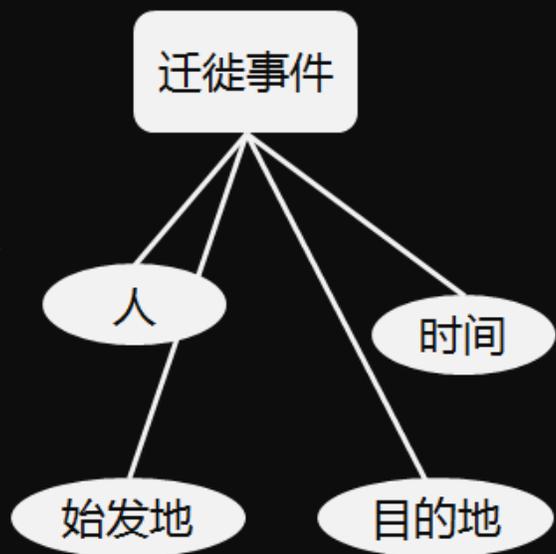
始发地：婺源考川

目的地：绩溪胡里

概念抽象



基于本体的知识组织



从概念(Concept)
到本体(Ontology)

```
shl:Migration a rdfs:Class ;  
  rdfs:label "迁徙" ;  
  rdfs:subClassOf shl:Resource ;  
  rdfs:comment "一次迁徙事件。" .  
shl:originalLocality a rdf:Property ;  
  rdfs:label "原居地" ;  
  rdfs:domain shl:Migration ;  
  rdfs:range bf:Place ;  
  rdfs:comment "迁徙之前的居地。" .  
shl:locality a rdf:Property ;  
  rdfs:label "迁居地" ;  
  rdfs:domain shl:Migration ;  
  rdfs:range bf:Place ;  
  rdfs:comment "迁徙完成后的居地。" .
```

知识组织的语义化

面向数据、事实和
知识，以RDF
Triples为基本单位

难点：实体
提取和匹配
数据错漏、
不一致、数据
单一

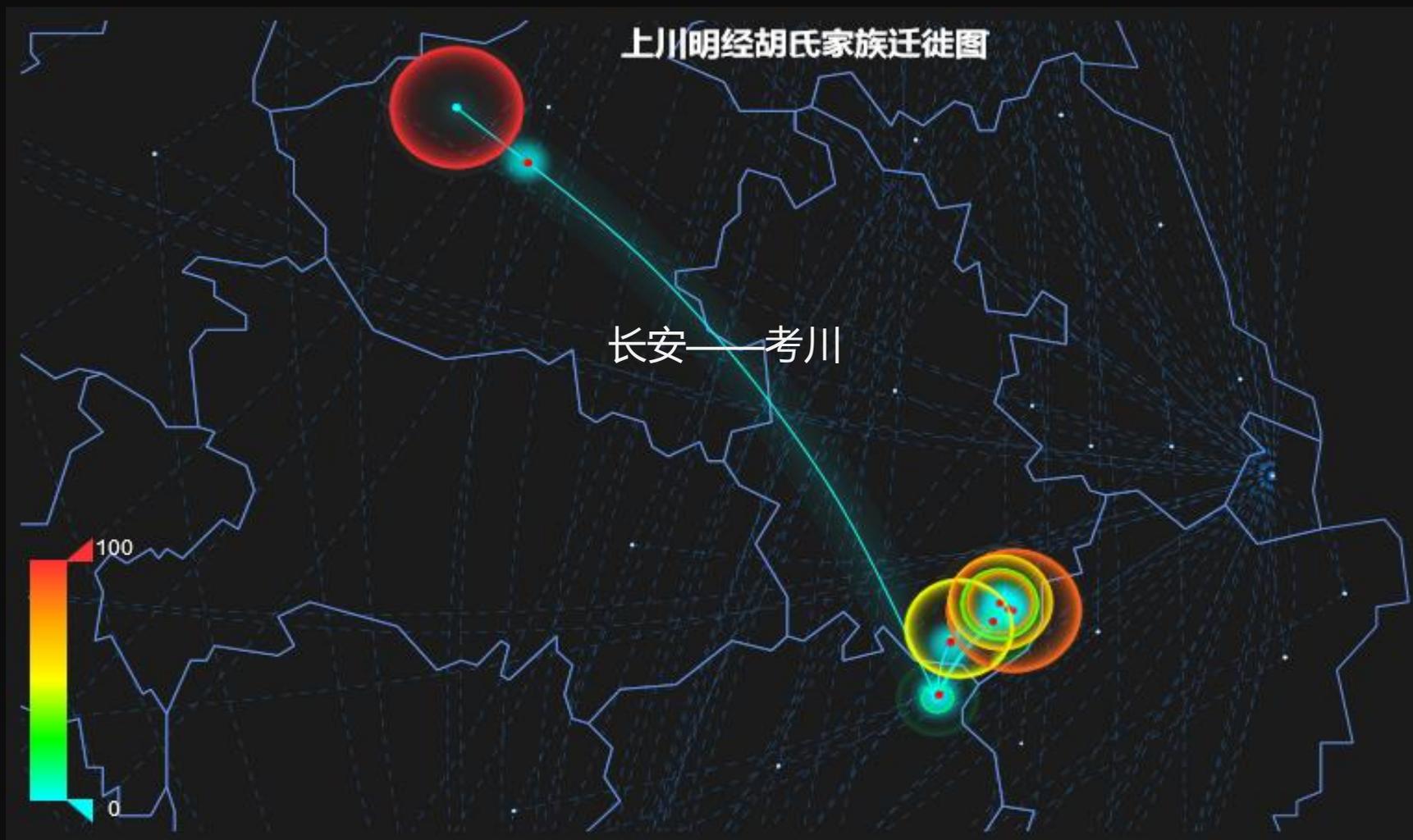
```
<http://gen.library.sh.cn/Migration/1> a shlgcn:Migration ;
    bf:eventAgent <http://gen.library.sh.cn/Person/16607> ;
    shlgcn:temporal <http://gen.library.sh.cn/Temporal/14> ;
    shlgcn:locality <http://gen.library.sh.cn/Place/1129> ;
    shlgcn:originalLocality <http://gen.library.sh.cn/Place/203> .

<http://gen.library.sh.cn/Person/16607> a shlgcn:Person ;
    foaf:name "胡昌翼" ;
    shlgcn:courtesyName "宏远";
    shlgcn:pseudonym "眉轩" ;
    shlgcn:description "唐昭宗幼子" .

<http://gen.library.sh.cn/Place/1129> a shlgcn:Place ;
    bf:label "考川" ;
    shlgcn:county "婺源县" ;
    shlgcn:city "上饶市" ;
    shlgcn:province "江西省" ;
    shlgcn:town "紫阳镇" ;
    shlgcn:village "考水村" ;
    geo:long "117.757787" ;
    geo:lat "29.27642" .

<http://gen.library.sh.cn/Temporal/14> a shlgcn:Temporal ;
    shlgcn:dynasty "唐" ;
    shlgcn:reignTitle "天佑元年" ;
    shlgcn:beginYear "904"^^xsd:int .
```

数据可视化



明经胡氏始祖的身世之谜.....

 上川明经胡氏宗谱三卷首一卷末一卷（安徽省宣城市績溪縣）  績溪

 明经胡氏續修宗谱不分卷（江西省上饒市婺源縣）  婺源

【副题名】 明經胡氏續修宗譜

【责任者】 胡璘

【纂修时间】 明嘉靖二十三年（1544）

【摘要】 始祖昌翼，谱稱本唐昭宗幼子，母何后，天祐元年生，避禍来婺源考川，遂冒義父胡三公姓。後唐同光間，登明经進士。後世子孫散居新安諸邑、遂安、淳安、樂平、鄱阳、浮梁、興安、漢阳，乃遠至貴州、雲南者，世稱明经胡氏，以别安定胡氏。是为明经胡氏统宗谱。谱存谱序、墓图及宋胡次焱撰《論始祖》、《論姓氏》、《論過房》、《論稱呼》等。书名据版心、卷端题。

【版本】 刻本

【数量】 存三册

上图 山西家谱

馆藏地：上海图书馆

地址：上海市 徐汇区 淮海中路1555号

DOI为STJP011208：殘本（索书号：911895-97）请至上海图书馆家谱阅览室借阅

家族先祖、名人

姓名	类型	朝代
胡次焱	名人	
胡昌翼	始祖	天祐元年

什么？
你说大师胡适本是
李唐皇室的后代！
图书馆员是八点档
肥皂剧的编剧教出来
的么？



从数字化全文中挖掘知识

世系表——家谱文献全文中最重要的内容

<p>兆賜公 長子</p> <p>應禎公 生康熙癸酉摘 應成次子為後</p> <p>天禮公 娶某氏生二子</p> <p>德拱公 娶某氏生一子</p> <p>錫白公 止</p>	<p>兆蔭公 次子</p> <p>應可公 遷休甯</p> <p>天蘇公 無</p> <p>德亮公 娶王氏生一子</p> <p>錫順公 止</p>	<p>兆恩公 子</p> <p>應寶公 字惟善生康熙 庚子十月十四 娶汪氏生康熙 己亥生一子</p> <p>天樂公 娶某氏生二子</p> <p>德高公 娶張氏生一子 娶王氏生二子</p> <p>錫柏公 貞為公 止</p> <p>貞開公 止</p>	<p>兆賜公 次子</p> <p>應祥公 生康熙丙戌歿 乾隆壬申娶石 氏生康熙乙未 子一子</p> <p>天信公 止</p> <p>德振公 娶某氏生一子</p> <p>錫紅公 止</p>	<p>兆賜公 三子</p> <p>應成公 生康熙庚子娶 某氏生二子以 次子天禮為應 禎後</p> <p>天燁公 娶某氏生二子</p> <p>德水公 娶某氏生一子</p> <p>錫清公 止</p>	<p>兆思公 次子</p> <p>應環公 字玉連生雍正 戊申五月二十 八娶石氏生雍 正癸丑生二子</p> <p>天仁公 生乾隆庚申六 月初八娶某氏 生一子</p> <p>根喜公 娶某氏生二子</p> <p>錫業公 名觀業娶曹氏 生二子公葬上 庵屬氏葬社土 上</p> <p>貞亮公 止</p> <p>貞奎公 止</p>
---	--	---	---	---	---

上川明经胡氏世系表扫描图片

从扫描图片到结构化数据

上川明经胡氏世系表(Excel格式)

Id	姓氏	谱名	字辈	字	号	排行	谥号	生于	卒于	父亲	母亲
5036	胡	三公	義	繼之		0		唐會昌癸亥三月初五	後唐天成丙戌五月		
5037	胡	昌翼公	昌	宏遠		1		天祐甲子三月一日	宋咸平己亥十月初三		
5038	胡	延政公	延	克修、以禮	節庵	1		後唐天成己丑十二月十五	宋大中祥符己酉	5037	詹氏
8898	胡	延賓公	延	以敬		2		後唐長興壬辰十一月初七		5037	詹氏
8897	胡	延臻公	延	以福		3		唐清泰甲午九月十五	宋大中大中祥符甲寅十月	5037	詹氏
5039	胡	百彥公	百			1				5038	汪氏、詹氏、秦氏
5040	胡	百祿公	百	廷堅	禹焦	2				5038	汪氏、詹氏、秦氏
8899	胡	百英公	百			3				5038	汪氏、詹氏、秦氏
8900	胡	百忠公	百			4				5038	汪氏、詹氏、秦氏
8901	胡	百順公	百			5				5038	汪氏、詹氏、秦氏
8904	胡	文義公	文			1				8898	
8902	胡	文昊公	文			1				8897	程氏
8903	胡	文晟公	文			2				8897	程氏
5041	胡	簡公		居敬		1				5039	汪氏、戴氏、趙氏
5043	胡	筮公				2				5039	汪氏、戴氏、趙氏
5042	胡	筮公				3				5039	汪氏、戴氏、趙氏
5044	胡	簫公				4				5039	汪氏、戴氏、趙氏
5045	胡	簧公				5				5039	汪氏、戴氏、趙氏

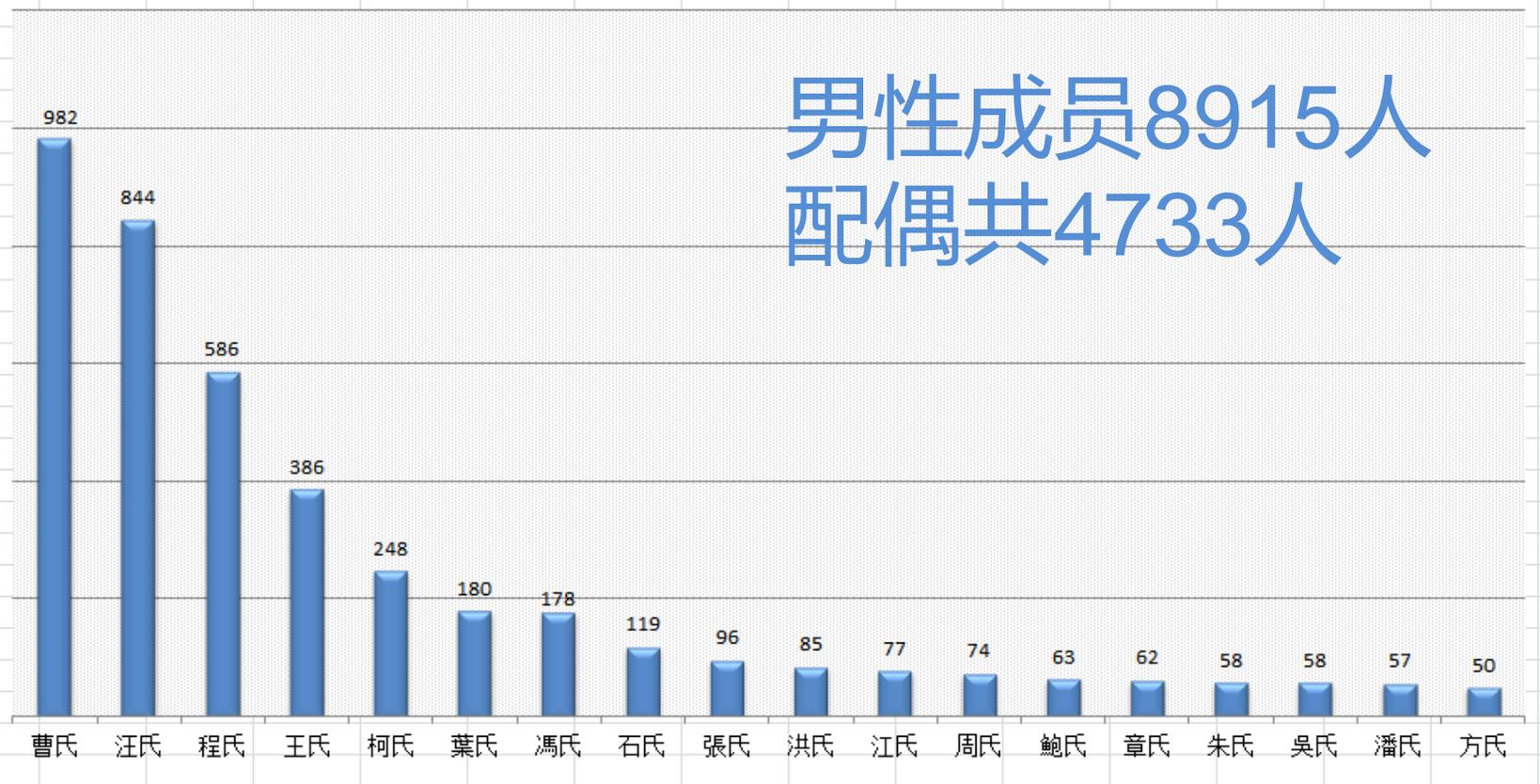
从结构化数据到机器可读的知识

描述胡适先生的RDF/Turtle数据

```
<http://data.library.sh.cn/jp/entity/person/uqwn21pa7iah9p87> a shl:Person ;
  shl:name "洪骅" ;
  foaf:name "胡适" ;
  shl:courtesyName "适之" ;
  foaf:familyName <http://data.library.sh.cn/authority/familyname/rvmgzfsec8os93mv> ;
  shl:orderOfSeniority "3" ;
  shl:birthday "光绪辛卯十一月十七未时" ;
  shl:description "考派美国留学生名适字适之事迹见学了生光绪辛卯十一月十七未时娶江氏生光绪庚寅十一月初八辰时" ;
  shl:relatedWork <http://data.library.sh.cn/jp/resource/work/jklhb5c3ga1rvxe3>;
  shl:roleOfFamily<http://data.library.sh.cn/jp/vocab/ancestor/xian-zu>;
  rel:spouseOf <http://data.library.sh.cn/jp/entity/person/kvop3qdif1okxhoa>;
  rel:childOf <http://data.library.sh.cn/jp/entity/person/mc7khkqgql3rpceb> .
```

基于RDF数据的统计和分析

胡适家族配偶姓氏出现次数统计



这些数据讲了什么样的故事？

胡李不通婚的家规是真的吗？

从唐末至民初，真的没有一人娶过姓李的姑娘耶！

如果你是上川明经胡氏的后人，娶一位姓曹的女纸吧...

保存用户头脑中的知识

用户的力量



反馈信息

[← 个人中心](#)

10 条/页

高级检索:

序号	文献题名	反馈标题	用户姓名	时间	状态	操作
1	宜兴侯庄沈氏宗谱十二卷	常熟博	沈子瞻	2016-03-18	未回复	详细
2	余氏族谱八卷	人名订正	余伟	2016-03-07	已回复	详细
3	李氏族谱四卷	家族人物补正	余伟	2016-03-07	未回复	详细
4	巫氏族谱四卷	摘要补正	余伟	2016-03-07	未回复	详细
5	余氏家谱四卷	补正	余伟	2016-03-07	未回复	详细
6	鍾氏族谱四卷	补正	余伟	2016-03-07	未回复	详细
7	颜氏族谱五卷首一卷末一卷	补正	余伟	2016-03-07	未回复	详细
8	羅氏善慶祠族谱八卷	補闕	余伟	2016-03-07	未回复	详细
9	甘氏秘谱	"谱"名	甘树桀	2016-03-01	未回复	详细
10	定陽甘氏宗谱三卷	馆藏地	甘树桀	2016-02-27	未回复	详细

显示 1 到 10 共 24 条记录

[← Previous](#) [1](#) [2](#) [3](#) [Next →](#)

反馈信息

文献题名: 宜兴侯庄沈氏宗谱十二卷

标题: 常熟博

内容: 常熟博是什么意思? 是不是在常熟博物馆有, 可以查询否?

反馈人: 沈子瞻

专家回复: * 请输入回复内容

回复人: 夏翠娟

用户贡献内容

反馈

修改 + 评审

上传家谱

全文抄录

在线修谱



盛宣怀（字:杏生 号:止叟） 生:1844 卒:1916

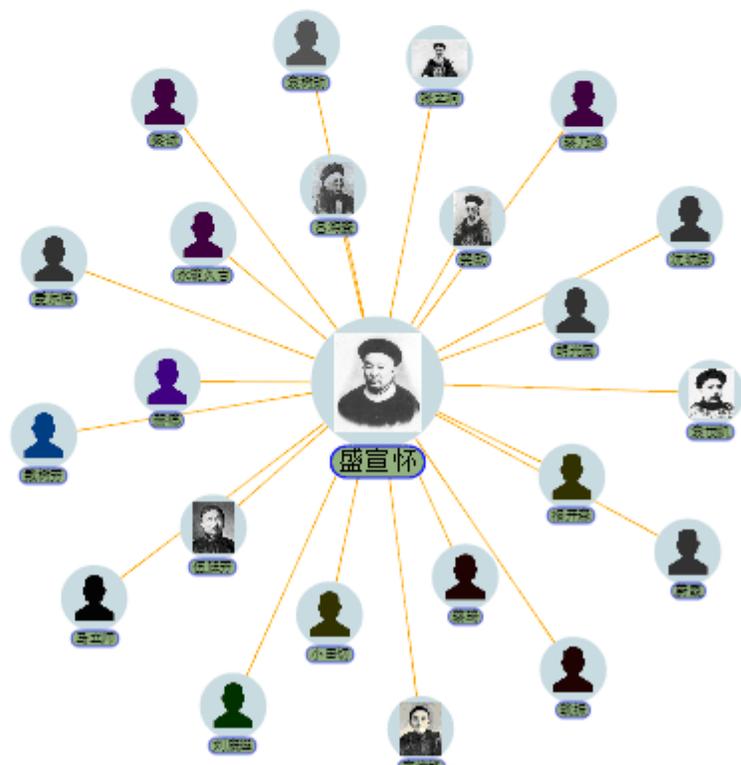
1870年入李鸿章幕。1873年任上海轮船招商局会办。1880年筹办中国电报局，任总办。1892年创办北洋大学堂。其间极力协助李鸿章举办洋务。1893年筹办华盛纺织总厂，任督办，以官督商办及官商合办形式垄断洋务企业。1895年在天津创办中西学堂。1896年接办汉阳铁厂，大冶铁矿，经办芦汉铁路。1898年兼办萍乡煤矿。同年倡办中国通商银行。1900年参与英美策划的东南互保，订立《东南互保约款》。1902年任工部左侍郎。1910年任中国红十字会会长，1911年授邮传部尚书，皇族内阁成立后任邮传部大臣。与四国银行团签订币制实业借款合同，又以铁路国有名义将已归商办的川汉、粤汉干线路权作抵押，大借外债，激起铁路风潮。武昌起义后亡命日本。1916年3月25日在上海病卒。著有《愚斋存稿》、《盛宣怀未刊信稿》，刻有《常州先哲遗书》，《经世文续编》、《林胡曾三公奏议》，另有《盛宣怀档案资料选辑》行世。

信函

电报

相关人物

- 袁世凯
- 奕劻
- 戴树芳
- 福开森
- 马立师
- 小田切
- 爱尔德
- 刘康遐
- 蔚霞
- 沈幼彦
- 袁树勋
- 胡光圆
- 李维格
- 蔡乃煌
- 赫德
- 梁诚
- 永瀛久吉
- 吕海寰
- 蔡琦
- 张之洞
- 伍廷芳
- 载振



信函电报收发关系图

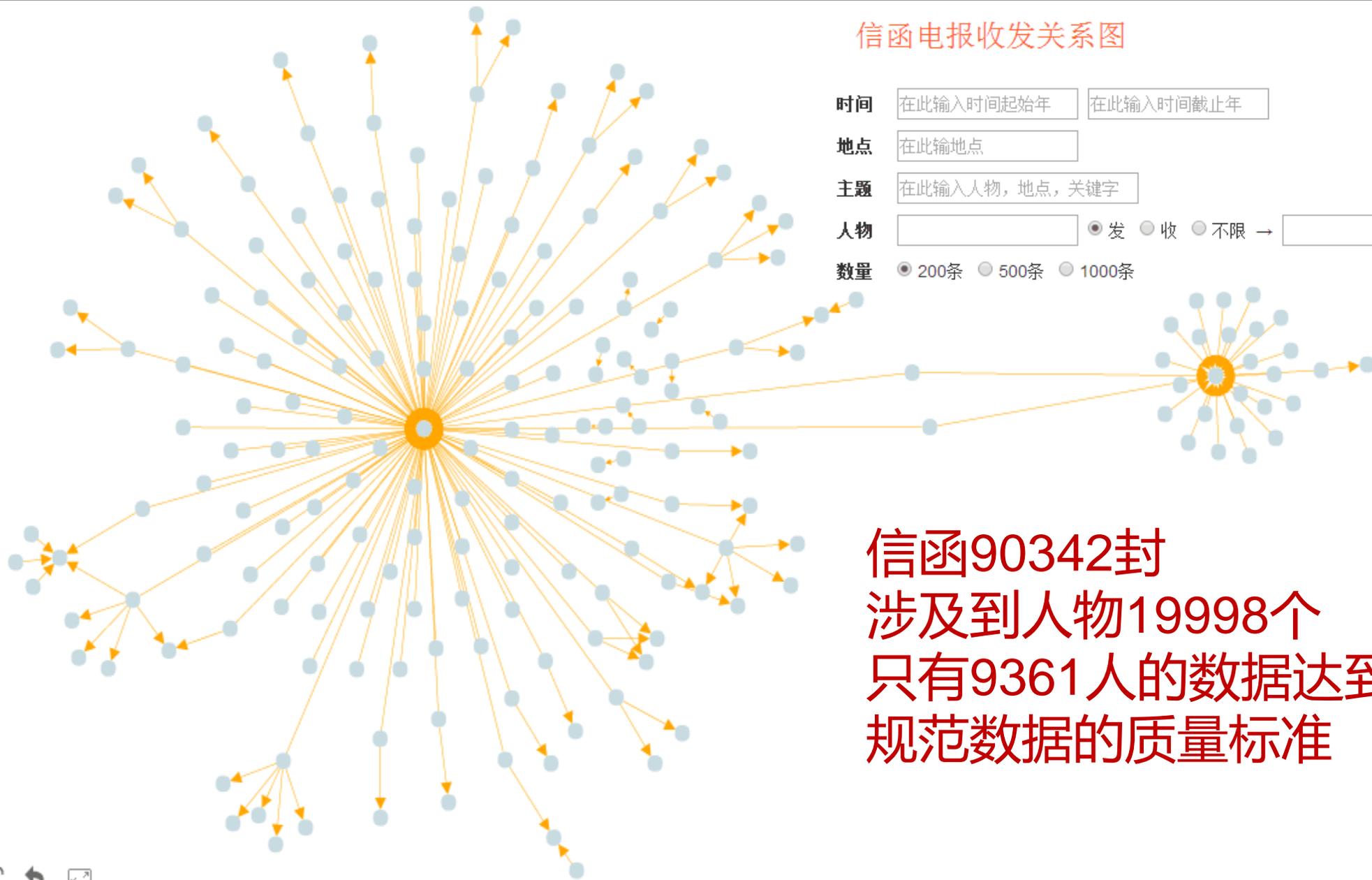
时间

地点

主题

人物 发 收 不限 →

数量 200条 500条 1000条



信函90342封
涉及到人物19998个
只有9361人的数据达到
规范数据的质量标准

盛宣怀致李鸿章函

众包

(CrowdSourcing)



截止日期: 2016年4月17日

元数据字段

全文抄录

备注

操作记录

暂存

提交

数据开放——一起来，更精彩

http://data.library.sh.cn



本网站是上海图书馆的开放数据平台，将陆续以关联开放数据（Linked Open Data）的方式向互联网公开发布各种术语词表、规范档、书目数据等。提供各种数据消费接口供开发人员调用，以促进数据的开发获取、共享和重用。

中国历史纪年表

More...

地理名词表

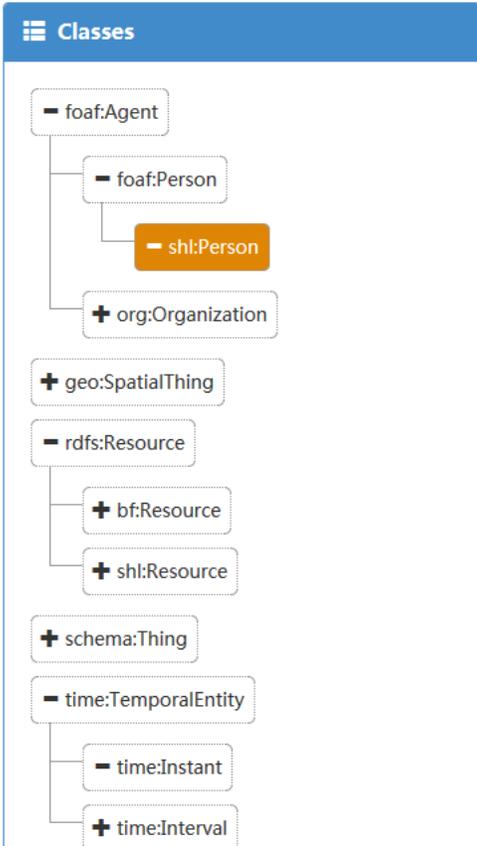
More...

机构名录

More...

http://gen.library.sh.cn:8080/ontology

Class View



shl:Person

个人。

Property	Subproperty of	Label	Used With	Expected Value
foaf:gender		性别	shl:Person	rdfs:Literal
rel:childOf		父	shl:Person	shl:Person
rel:parentOf		子女	shl:Person	shl:Person
rel:spouseOf		配偶	shl:Person	shl:Person
shl:birthday		生于	shl:Person	rdfs:Literal
shl:courtesyName		字	shl:Person	rdfs:Literal
shl:deathday		卒于	shl:Person	rdfs:Literal
shl:genealogyName		谱名	shl:Person	rdfs:Literal
shl:generationCharacter		字辈	shl:Person	rdfs:Literal
shl:migration		迁徙	shl:Person	shl:Migration
shl:orderOfSeniority		行	shl:Person	rdfs:Literal
shl:posthumousTitle		谥号	shl:Person	rdfs:Literal
shl:pseudonym		号	shl:Person	rdfs:Literal
shl:roleOfFamily		角色	shl:Person	bf:Category

http://data.library.sh.cn:8890/sparql

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#))

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

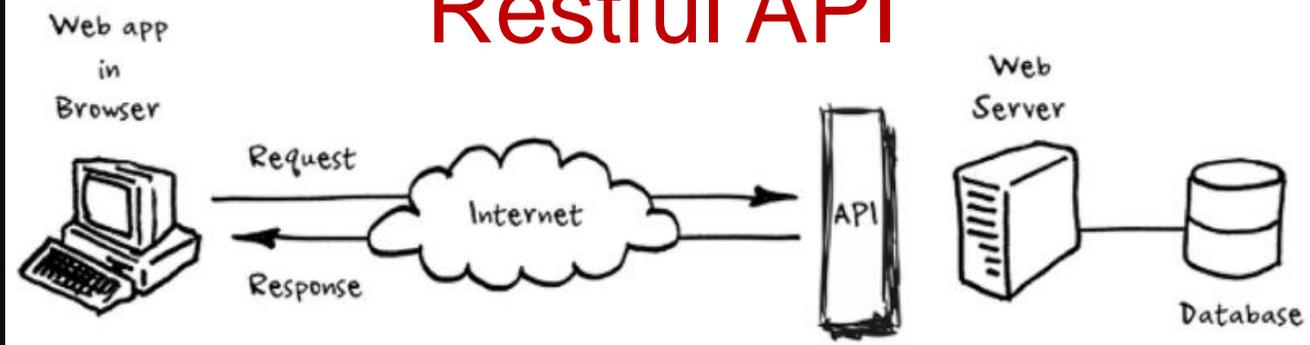
Java

C++

PHP

Python

Restful API



[http://data.library.sh.cn/jp/person/data?\[参数1\]&\[参数2\]&\[参数3\]?key=YourAPIKey](http://data.library.sh.cn/jp/person/data?[参数1]&[参数2]&[参数3]?key=YourAPIKey)

关联开放数据 (Linked Open Data)

——从本地数据到公共知识

<http://pcrc.library.sh.cn/zt/opendata/>

SHANGHAI LIBRARY

开放数据 应用开发竞赛

作品提交要求文档 (更新)

N

姓氏

608种

谱籍地名

1609种

596家收藏机构

家谱文献

53987种

先祖名人

130867人

历史朝代

880种

SPARQL
Endpoint

Restful
API

HTTP
URI

JSON
-LD

返回数据

上川明经胡氏世系
8915人及其配偶



数据消费接口

规范词表

2015~

盛档 2015~

2016 古籍

家谱 2014~

2016 翁氏笔记



家谱

- 60000种收录
- 30000种全文

名人手稿

- 70000件
- 20000人

特藏

- 老照片
- 年画
- 香烟盒
- 月份牌
-

盛宣怀档案

- 170000件

民国图书

- 213000种
- 850000册

民国期刊

- 25000种
- 1000万篇

地方志

- 5400种
- 90000册

稿抄本

- 6000种

善本

- 30000种
- 170000册

晚清期刊

- 300种
- 280000万篇

申报

- 78年
- 数十万拍

字林西报

- 101年
- 500000拍

舆图

- 14000幅

碑帖

- 150000件



感谢聆听！

Q&A!

Contact: cjxia@libnet.sh.cn