

特色数字人文研究环境的构建与探索

主讲人：毛京军





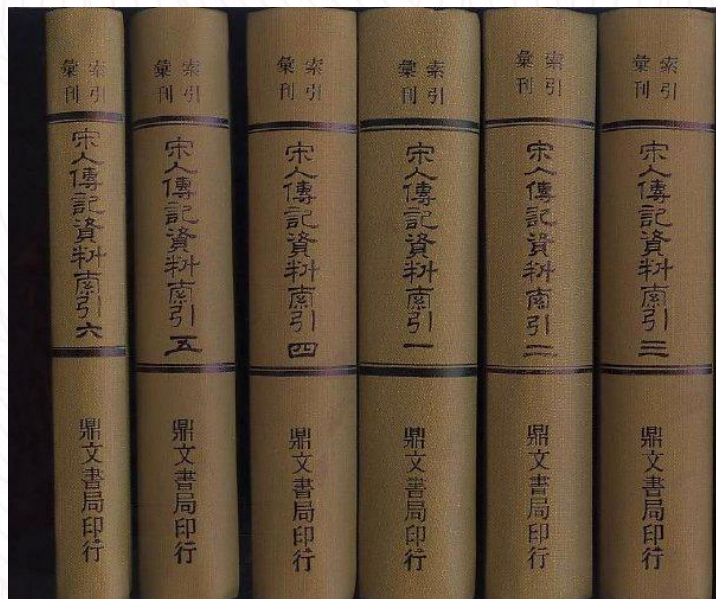
数字
人文

ABOUT

关于数字人文

关于数字人文

数字人文的研究范式和视角主要表现为**人文和计算、跨界和融合**。



数字人文——跨界融合，研究尺度相对自由

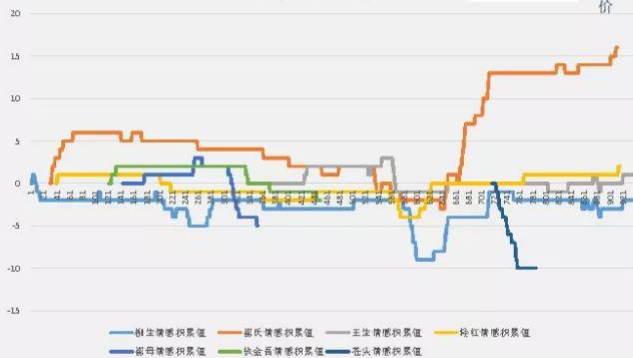
人物情感波动

Emotional fluctuation of characters



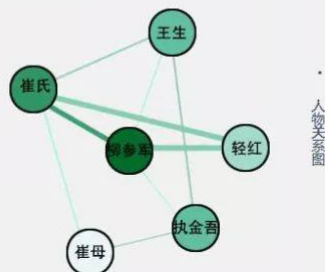
柳生的人物评价

人物情感交织图

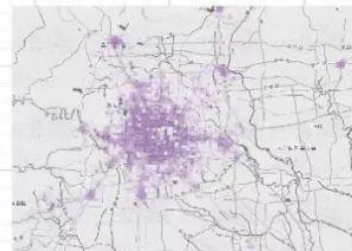


人物情感交织图

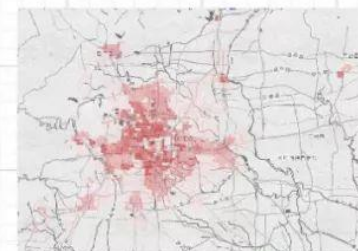
- 作为主要人物的崔氏人物情感十分积极，侧面反映了作者对其的赞扬与喜爱，而作为主人公的柳参军却是及并没有那么突出。



人物关系图



POI密度



POP密度



数据来源：北京城市实验室BCL

数字人文的基础是结构化数据

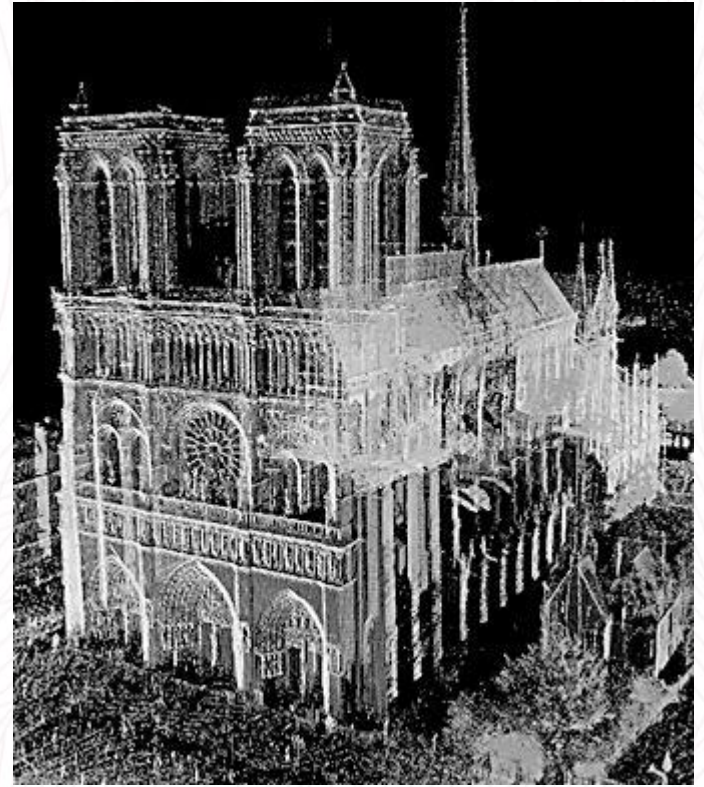


隐性知识的显性化、结构化、标签化

对在线文献中的概念和实体之间建立联系的目标应该是将人文学者从繁复的资料收集、整理和比对劳动中解放出来，自动实现不同资料、事件、记录之间的关联和集成，使人文学者的工作平台跃迁一个层次，不用再为这些基础工作费神，而将精力集中在资料分析和观点提炼上，从而极大的提升人文研究的效率。

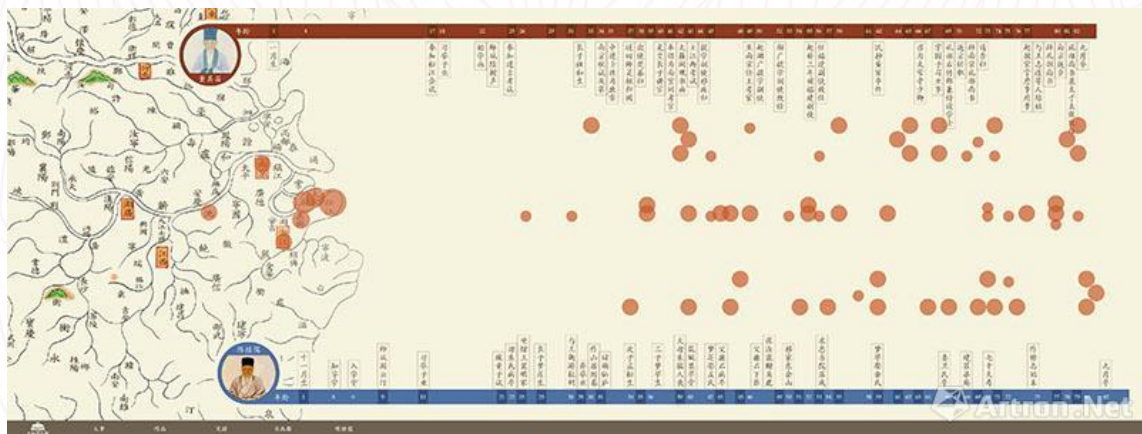
——王晓光《语义出版与数字人文》

结构化的数字人文数据是重建巴黎圣母院的基础

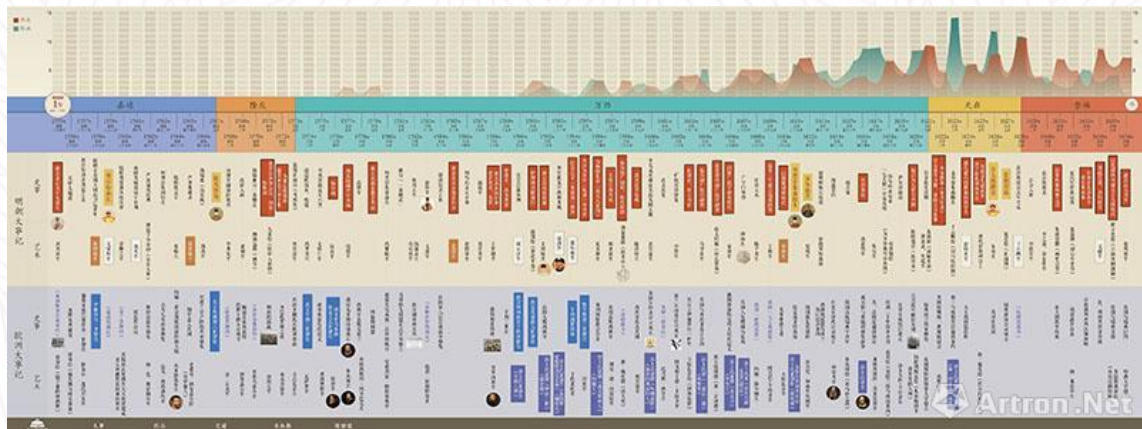


案例说明

上海董其昌博物馆数字人文展示

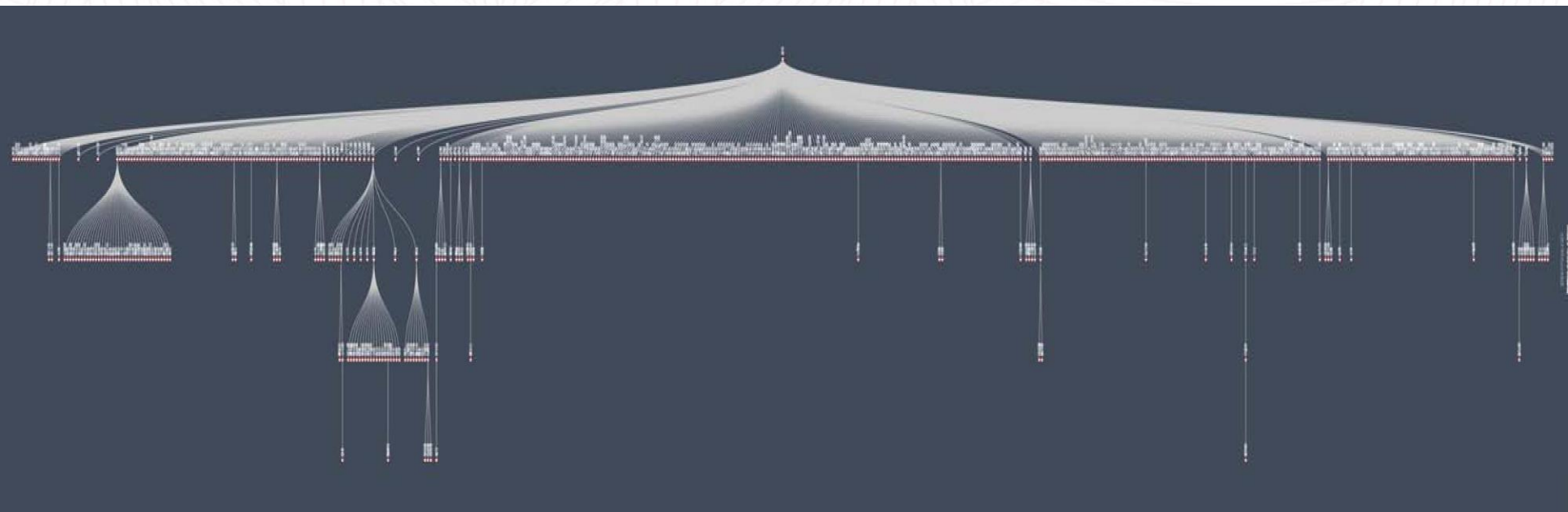


(董其昌、陈继儒关系图)



(董其昌大事年表)

SNA 社会网络分析 – 朱熹四代弟子图谱——传承断了？



邱勇@北京大学数字图书馆实验室
http://dh.kvlab.org/cbdb_vis/part2_dizinet.html

华东师大异构方志量化数据——明代历朝阁臣分省直与出身户类数量统计与对比

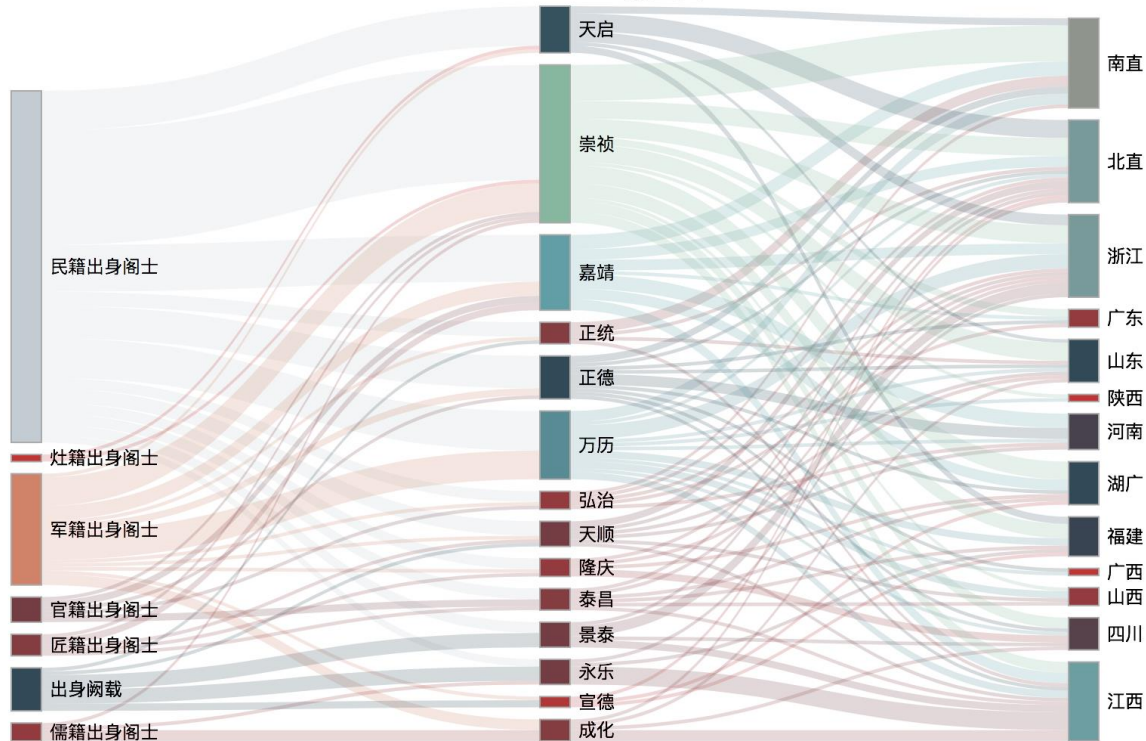
当前位置：可视化工具 > 可视化辅助工具 > 桑基图

桑基图

实例说明

上传实例数据

桑基图



发展趋势

武汉大学

数字人文研究中心

北京大学

古代史研究中心
数字图书馆实验室

南京大学

数字人文研究中心

中国人民大学

数字人文技术实验室

华中师范大学

数字方志集成平台

湖南大学

互联网信息服务研究中心

复旦大学

信息与传播研究中心

曲阜师范大学

数字人文研究中心

**国内大学开始纷纷建立
数字人文研究中心和研究平台**



数字
人文

特色数字人文构建与升级

CBDB

中文在线数字人文研究特色成果——CBDB



CBDB(中国历代人物传记资料库)之始祖为郝若贝教授 (Robert M. Hartwell) (1932–1996)。**郝若贝教授**将本资料库初版及其他财产遗赠哈佛燕京学社。

目前本资料库的开发工作系由**哈佛大学费正清中国研究中心、中央研究院历史语言研究所、北京大学中国古代史研究中心、中文在线**四方合作进行。

截至 2017年4月为止，中国历代人物传记数据库收录了约**430,000**人物传记资料。现有资料绝大多数为唐代中叶至明代中叶（即八世纪至十五世纪），其中宋代（960-1279）资料尤多。



CBDB的顾问团及技术团队

● CBDB学术顾问团

包弼德 (教授, 哈佛大学副教务长, 本项目执行委员会主席)

邓小南 (北京大学教授, 中国宋史研究会会长, 中国史学会副会长, 北京大学人文社会科学研究院院长, 北京大学中国古代史研究中心学术委员会主任)

方秀洁 (加拿大麦吉尔大学东亚学系教授)

傅君迈 (美国加利福尼亚大学尔湾分校教授)

刘铮云 (中央研究院历史语言研究所)

项洁 (台湾大学特聘教授、数位人文研究中心主任)

张晓虹 (复旦大学历史地理研究中心主任)

北京大学编辑组和哈佛大学编辑组

徐力恒、骆文、邱靖嘉、裴成国、陈师沛、王宏甦、陈克双、胡斌、董文静等120多位硕、博士成员。

● CBDB技术团队

Stuart Sheiber, Elif Yamangil, Rani Nelken, Onder Eker, Deng Ke, Peifen Hsieh, 程旭航、刘涛、张程、王明壮、谭政、翟莹等近20位成员。



特色数字人文研究环境和训练环境的基本部件组成

可视化辅助工具

提供形象直观，交互可用易用，可高度个性化定制的大量数据可视化辅助工具

文本清洗和标注

- ✓ 基于方志的文本标注程序是经莱顿大学魏希德教授授权中文在线进行二次开发而成
- ✓ 自动标记人名、地名、年号及职官数据。
- ✓ 自定义要标记的关键词，手动增加词条，自动进行全文检索标注。
- ✓ 标注后您可以边阅读文献，边参考相关文献内容，也可以使用可视化平台和工具，做进一步的分析。

文字识别

平台已经上线古籍文字识别工具供客户体验，支持竖排刻本和楷体抄本，平均识别率接近92%。同时支持横排简体中文图书的识别，精度接近99%。

基础语料库和文献资源

基于中国历代人物传记资料库 (CBDB) 的历代知名人物语料库 (全名、字号) ;
基于中国历代人物传记资料库 (CBDB) 的历代职官语料库;
基本中国历史地理信息系统 (CHGIS) 及台湾历史地名 (TWGIS) 的历代地名库
基于北京大学CBDB团队的年号和公元纪年转换库。
20亿字的文献资源

构建特色数字人文研究的网络基础设施、训练环境和研究环境

关联型基础语料库结构化

司马光 编辑

司马光（1019年11月17日—1086年10月11日），字君实，号迂叟，陕州夏县（今山西夏县）涇水乡人^[1]，世称涇水先生^[2]。北宋政治家、史学家、文学家，西晋安平献王司马孚之后。

宋仁宗宝元元年（1038年），司马光进士第，累进龙图阁直学士，宋神宗时，因反对王安石变法，离开朝廷十五年，主持编纂了中国历史上第一部编年体通史《资治通鉴》。历任仁宗、英宗、神宗、哲宗四朝，官至尚书左仆射兼门下侍郎。卒赠太师、温国公，谥文正，为人温良谨恭，刚正不阿，办事用功刻苦，勤奋，以日力不足，继之以夜自诩，其人恪守儒家教化下的规范，历来受人景仰。^[3]

司马光生平著作甚多，主要《资治通鉴》《司马公文集》、《稽古录》、《涇水记闻》、《潜虚》等。

TA说

此人博学多才刚正不阿，却偏偏保守硬把一个刚复兴的王朝推向灭亡

他自幼就熟读了左传的精髓，勤奋好学，刻苦努力，博学多才，刚正不阿，是北宋著名的政治家、史学家、文学家。但是不幸的是他却被成了北宋王朝的埋罪人，这既是他个人的性格悲剧，也是北宋王朝的悲剧。他是否呢？他是否呢？他是历史上大名鼎鼎的司马光那么问题来了，如此优秀的人物怎么会成为北宋王朝的埋罪人呢？^[4]

本名	司马光	出生日期	1019年11月17日
别称	司马温公、涇水先生、司马文正	主要作品	《资治通鉴》《潜虚》《司马公文集》等
字	君实、文正	主要成就	主持编纂《资治通鉴》；历任四朝，政绩卓著
所属时代	北宋、南宋	籍贯	陕州涇县涇水乡
民族	汉族	官	尚书左仆射、门下侍郎
出生时间	1019年11月17日	爵	太师、温国公
		谥	文正

目录

- 1 人物生平
 - 早年经历
 - 步入宦途
 - 涇州知州
 - 置国中央
 - 编修通鉴
 - 新法失败
 - 晚年岁月
 - 2 主要成就
- 3 个人作品
- 4 人物评价
- 5 人物轶事
- 6 家族成员
- 7 墓址
- 8 史料记载
- 9 影视形象

人物生平

早年经历

宋真宗天禧三年（1019年）十月十八日，司马光出生于光州夏县，此时其父司马池在光山任县令，所以给他起名为光。^[5]六岁时，司马池就教司马光读书，七岁时，不仅能背诵《左氏春秋》，还能讲明白书里的意思，并且做出了“破毡朋友”这一件轰动乡里的事情。

天圣九年（1031年），司马光随父亲司马池从东京出发，一路经洛阳、潼关、宝鸡，过秦岭，前往四川广元，出任州州转运使，在栈道上遇着巨蟒，他沉着冷静地手持利刃，扎进巨蟒的腹部上，使巨蟒疼痛得一震，滚下了深不可测的悬崖下边。^[6]

司马光的博学来自多方面，一方面他好学强记，另一方面他的父亲也是重视培养，他其诚实聪明，又十分懂事，深得父亲喜爱。同时，他做出过或和同僚密生交结，司马池只好把他带在身边，其满口



司马光塑像

TA说 解读那些你不知道的历史故事



司马光砸缸

赵宋之世强盛四朝，不事奢华，他私人藏用器用器在历代皇室中突出，国家安定太平，经...



本人编辑服务 上线啦!

词条统计

浏览次数: 3843791次

编辑次数: 524次历史版本

创建时间: 2018-04-27

创建者: 上海笑笑

词条贡献榜

突出贡献者:

上海笑笑

上海笑笑



结构化历史人物的“生平” / “履历” or “简历”

知某縣事	鄞縣	(1047) - (1050)	宋人傳記資料索引(電子版), 頁2363
通判	舒州	(1051) - (1054)	
群牧判官	宋朝	(1054) - (1056)	
太常寺博士		(1057) -	宋兩浙路郡守年表,
知某州軍州事	常州	(1057) - (1058)	宋兩浙路郡守年表,
三司度支判官	宋朝	(1058) - (1061)	宋人傳記資料索引(電子版), 頁1536
提點刑獄所	江南東路	(1058) - (1059)	
知制誥		(1062) - (1064)	宋兩浙路郡守年表,
丁母憂	江寧府	(1064) - (1066)	
知某州軍州事	江寧府	(1067) - (1067)	宋兩江郡守易替考,
尚書省工部工部司郎中		(1067) -	宋兩江郡守易替考,
翰林學士		(1068) -	宋人傳記資料索引(電子版), 頁1536
參知政事	宋朝	(1069) -	宋人傳記資料索引(電子版), 頁1536

数字人文研究的基石——存量丰富的语料和文献

引得文献

查詢

热门搜索： 史記 孟子 禮記 荀子 孝經 說苑

引得文献 / 推薦

《增補鄭氏周易》 [東漢] 鄭玄	《周易口義》 [北宋] 胡瑗	《周易要義》 [南宋] 錢若翁
《禮記纂言》 [元] 吳澄	《禮記大全》 [明] 胡廣	《大戴禮記》 [西漢] 戴德
《古文孝經孔氏傳》 [西漢] 孔安國	《孝經指解》 [清] 玄宗	《孝經刊誤》 [南宋] 朱熹
《孝經大義》 [元] 董鼎	《孝經定本》 [元] 吳澄	《孝經述註》 [明] 項鼎
《御定孝經注》 [清] 世祖	《御纂孝經集註》 [清] 世宗胤禛	《孝經問》 [清] 毛奇齡
《史記索隱》 [唐] 司馬貞	《續史記十表》 [清] 汪鋐	《崑崙河源考》 [清] 萬斯同
《居濟一得》 [清] 張伯行	《水道提綱》 [清] 齊召南	《洛陽名園記》 [宋] 李格非
《洞霄圖志》 [宋] 鄧牧	《管平二州地名記》 [清] 顧炎武	《益部方物略記》 [宋] 宋祁
《岳陽風土記》 [宋] 范致明	《會稽三賦》 [宋] 王十朋	《桂海虞衡志》 [宋] 范成大
《嶺外代答》 [宋] 周去非	《武林舊事》 [宋] 周密	《臺海使槎錄》 [清] 黃叔敬
《宣和奉使高麗圖經》 [宋] 徐兢	《諸蕃志》 [宋] 趙汝涯	《溪蠻叢笑》 [宋] 朱輔
《宋宰輔編年錄》 [宋] 徐自明	《宣藏》 [宋] 呂本中	《百官箴》 [宋] 許月卿
《畫簾緒論》 [宋] 胡太初	《紹熙州縣釋奠儀圖》 [南宋] 朱熹	《歷代兵制》 [宋] 陳傅良
《補漢兵志》 [宋] 錢文子	《子略》 [宋] 高似孫	《直齋書錄解題》 [宋] 陳振孫
《漢藝文志考證》 [宋] 王應麟	《文淵閣書目》 [明] 楊士奇	《鋒帖平》 [宋] 姜夔
《法帖譜系》 [宋] 曹士冕	《名蹟錄》 [清] 朱珪	《石經考》 [清] 顧炎武
《唐史論斷》 [宋] 孫甫	《通鑑問疑》 [宋] 劉義仲	《兩漢筆記》 [宋] 錢時
《說苑》 [西漢] 劉向	《續孟子》 [清] 林慎思	《儒志編》 [宋] 王昶祖
《二程粹言》 [宋] 楊時	《儒言》 [宋] 晁說之	《童蒙訓》 [宋] 呂本中
《首心雜言》 [宋] 李邦獻	《袁氏世範》 [宋] 袁采	《近思錄》 [南宋] 朱熹
《明本釋》 [宋] 劉勰	《少儀外傳》 [南宋] 呂祖謙	《項氏家說》 [南宋] 項安世



查詢

引得文献 / 紀傳類 / 史記正義

《史記正義》 [作者：[唐] 張守節]

史記正義 提要

史記正義 目錄

史記正義 序

史記正義 論例證法詳

史記正義 卷一

史記正義 卷二

史記正義 卷三

史記正義 卷四

史記正義 卷五

史記正義 卷六

史記正義 卷七

史記正義 卷八

史記正義 卷九

史記正義 卷十

史記正義 卷十一

史記正義 卷十二

资源组织和文献资源发现

分卷

- 嘉慶縣江府誌序(42)
- 嘉慶縣江府誌凡例(6)
- 嘉慶縣江府誌街名(16)
- 嘉慶縣江府誌目錄(20)
- 嘉慶縣江府誌圖(133)
- 嘉慶縣江府誌卷首下(53)
- 嘉慶縣江府誌卷一(47)
- 嘉慶縣江府誌卷二(74)
- 嘉慶縣江府誌卷三(67)
- 嘉慶縣江府誌卷四(132)
- 嘉慶縣江府誌卷五(37)
- 嘉慶縣江府誌卷六(45)
- 嘉慶縣江府誌卷七(58)
- 嘉慶縣江府誌卷八(66)
- 嘉慶縣江府誌卷九(108)
- 嘉慶縣江府誌卷十(100)
- 嘉慶縣江府誌卷十一(98)
- 嘉慶縣江府誌卷十二(79)
- 嘉慶縣江府誌卷十三(34)
- 嘉慶縣江府誌卷十四(41)
- 嘉慶縣江府誌卷十五(60)
- 嘉慶縣江府誌卷十六(47)
- 嘉慶縣江府誌卷十七(76)
- 嘉慶縣江府誌卷十八(83)
- 嘉慶縣江府誌卷十九(48)
- 嘉慶縣江府誌卷二十(90)
- 嘉慶縣江府誌卷二十一(112)
- 嘉慶縣江府誌卷二十二(89)
- 嘉慶縣江府誌卷二十三(153)
- 嘉慶縣江府誌卷二十四(85)
- 嘉慶縣江府誌卷二十五(77)
- 嘉慶縣江府誌卷二十六(98)
- 嘉慶縣江府誌卷二十七(64)
- 嘉慶縣江府誌卷二十八(97)
- 嘉慶縣江府誌卷二十九(87)
- 嘉慶縣江府誌卷三十(69)
- 嘉慶縣江府誌卷三十一(60)
- 嘉慶縣江府誌卷三十二(68)
- 嘉慶縣江府誌卷三十三(67)
- 嘉慶縣江府誌卷三十四(63)
- 嘉慶縣江府誌卷三十五(62)
- 嘉慶縣江府誌卷三十六(97)
- 嘉慶縣江府誌卷三十七(74)
- 嘉慶縣江府誌卷三十八(89)
- 嘉慶縣江府誌卷三十九(92)
- 嘉慶縣江府誌卷四十(56)
- 嘉慶縣江府誌卷四十一(93)
- 嘉慶縣江府誌卷四十二(62)
- 嘉慶縣江府誌卷四十三(62)

第 2 共 100 页

添加 删除

案是時吳興累年失收人頗流移吳郡十城亦不全
 熟劫盜屢起征戌未歸強丁疏少太子慮吏爲人憲
 農蕪失業善人從役抄盜彌增因疏請權停此功武帝優詔論焉

唐元和五年王仲舒治蘇隄松江爲路唐書王仲舒傳

案蘇州府志云時松陵鎮南北西俱水鄉抵郡無陸路至是始通今吳江縣城北三里橋北行至長洲縣界七里橋曰古塘自觀瀾鋪至漵浦鋪十里曰石塘自漵浦至平望三十里曰官塘自平望南行至秀水縣王江涇曰土塘曰荻塘據此是隄應在蘇州府存之以備考

吳越天寶八年置都水營使以主水事募卒爲都號曰撩淺軍治河築隄一路徑下吳松江一路自急水港下澗

山湖入海十國春秋吳越武肅王世家

寶正二年潘柘湖及新涇塘由小官浦入海十國春秋

內容編輯

保存 撤銷

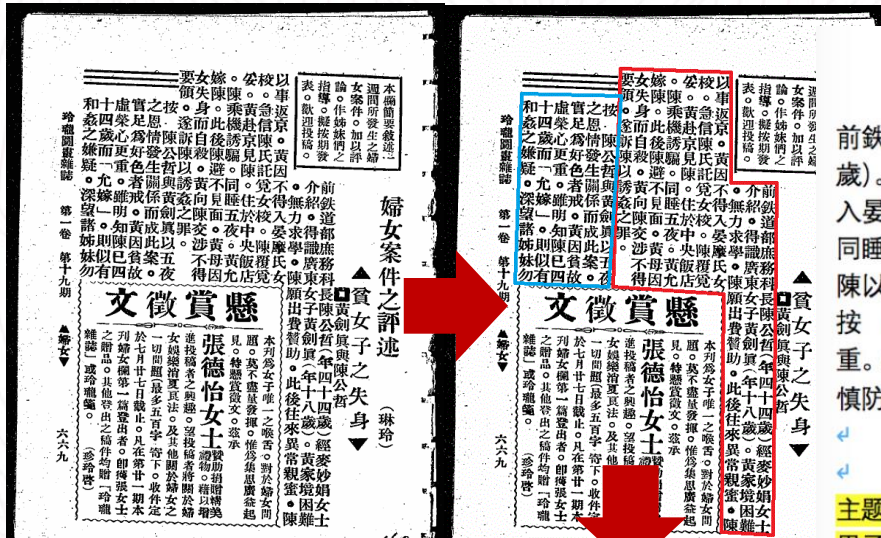
《案是時吳興累年失收人頗流移吳郡十城亦不全/熟劫盜屢起征戌未歸強丁疏少太子慮吏爲人憲》
 《農蕪失業善人從役抄盜彌增因/疏請權停此功武帝優詔論焉》
 唐元和五年王仲舒治蘇隄松江爲路《唐書王仲舒傳》
 《案蘇州府志云時松陵鎮南北西俱水鄉抵郡無陸路至是始通今吳江縣城北三里橋北行至長洲縣界七里橋曰古塘自觀瀾鋪至漵浦鋪十里曰石塘/自漵浦至平望三十里曰官塘據此/是隄應在蘇州府存之以備考》
 吳越天寶八年置都水營使以主水事募卒爲都號曰撩淺軍治河築隄一路徑下吳松江一路自急水港下澗
 山湖入海《十國春秋吳越/武肅王世家》
 寶正二年潘柘湖及新涇塘由小官浦入海《十國/春秋》

插入 删除

- 1 《案是時吳興累年失收人頗流移吳郡十城亦不全/熟劫盜
- 2 《農蕪失業善人從役抄盜彌增因/疏請權停此功武帝優詔
- 3 唐元和五年王仲舒治蘇隄松江爲路《唐書王仲舒傳》
- 4 《案蘇州府志云時松陵鎮南北西俱水鄉抵郡無陸路至是
- 5 《界七里橋曰古塘自觀瀾鋪至漵浦鋪十里曰石塘/自漵浦
- 6 《縣王江涇曰土塘曰荻塘據此/是隄應在蘇州府存之以備
- 7 吳越天寶八年置都水營使以主水事募卒爲都號曰撩
- 8 淺軍治河築隄一路徑下吳松江一路自急水港下澗
- 9 山湖入海《十國春秋吳越/武肅王世家》
- 10 寶正二年潘柘湖及新涇塘由小官浦入海《十國/春秋》

深度学习技术能为数字人文做什么？





貧女子之失身 黃劍真與陳公哲

前鐵道部庶務科長陳公哲（年四十四歲）經麥妙娟女士介紹。得識廣東女子黃劍真（年十八歲）。黃家境困難。無力求學。陳願出費贊助。此後往來異常親密。陳以事返京。黃因不得入晏摩氏女校。急信陳氏託覓女校。陳覆覓妥。黃赴京見陳。住於中央飯店。陳乘機誘騙。同睡五夜。黃允嫁陳。此後陳避不見面。黃母因女失身而自殺。黃向陳交涉不得要領。遂訴陳以誘姦之罪。

按 陳公哲與黃劍真以五夜之恩情發生關係而成此案。實足為好色者戒。黃因貧故虛榮心更重。雖明知陳已四十四歲而“允嫁”。則似有和姦之嫌疑。深望諸姊妹勿為虛榮所誤。更應慎防如陳之男子。以破壞女之貞操為能事者。

主題詞：誘奸——兩性關係——虛榮

男子形象：偽君子

男子行徑：始亂終棄，玩弄女性

女子形象：學生 虛榮 受騙 失身 自殺（母）

擇偶觀：謹慎 心態 年齡

勸導指迷與改良生活：內在修養，慎防此類男子

前鐵道部庶務科長陳公哲（年四十四歲）經麥妙娟女士介紹。得識廣東女子黃劍真（年十八歲）。黃家境困難。無力求學。陳願出費贊助。此後往來異常親密。陳以事返京。黃因不得入晏摩氏女校。急信陳氏託覓女校。陳覆覓妥。黃赴京見陳。住於中央飯店。陳乘機誘騙。同睡五夜。黃允嫁陳。此後陳避不見面。黃母因女失身而自殺。黃向陳交涉不得要領。遂訴陳以誘姦之罪。

按 陳公哲與黃劍真以五夜之恩情發生關係而成此案。實足為好色者戒。黃因貧故虛榮心更重。雖明知陳已四十四歲而“允嫁”。則似有和姦之嫌疑。深望諸姊妹勿

文本清洗和标注+自助学科服务



基于中国历代人物传记资料库 (CBDB) 的历代知名人物语料库 (全名、字号) ;

基于北京大学CBDB团队的年号和公元纪年转换库。

基本中国历史地理信息系统 (CHGIS) 及台湾历史地名 (TWGIS) 的历代地名库

基于中国历代人物传记资料库 (CBDB) 的历代职官语料库;

项目内容	语料条目
历代知名人物语料库	422600
历代地名语料库	11200
年号和公元纪年语料库	590
历代职官语料库	28370

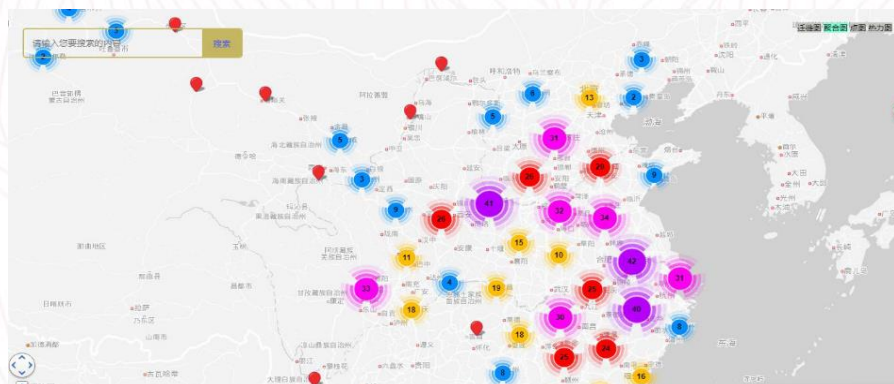
民国《玲珑》杂志文本清洗和标注

The screenshot shows a web-based interface for text annotation. On the left, there is a sidebar with a search bar and several buttons: '载入关键词/正规表达式', '关键词助手', '词夹子', and a row of colored buttons for '姓名', '别名', '时间', '地名', and '官名'. Below these is a '关键词' input field. The main area on the right is titled 'B: 关键词标记' and displays a text passage with several words highlighted in blue boxes. The highlighted words are: 失身, 誘騙, 誘殺, 誘姦, 虛榮, 雀戰, 多心, 婚約糾紛, 雀戰, 雀戰, 女學生, 嗜好, 愛情, and 雀戰. The text discusses a woman's life and her relationship with a man, mentioning various events and emotions.

- 文本的标注、挖掘与分析是数字人文研究的起点和最重要的方法之一。
- 随着分析方法的多元化，文本信息的属性、关键词随着可视化能力的增强、GIS技术的普及，逐渐成为历史文本语义表达中的热点。

数字人文与空间人文的结合——同源年谱信息的不同表示

社会活动与地理信息系统的关系



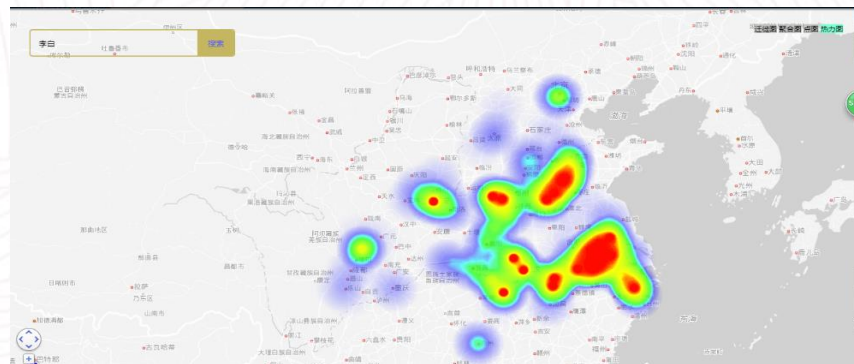
聚合图



迁徙图



点图



热力图



感谢聆听

