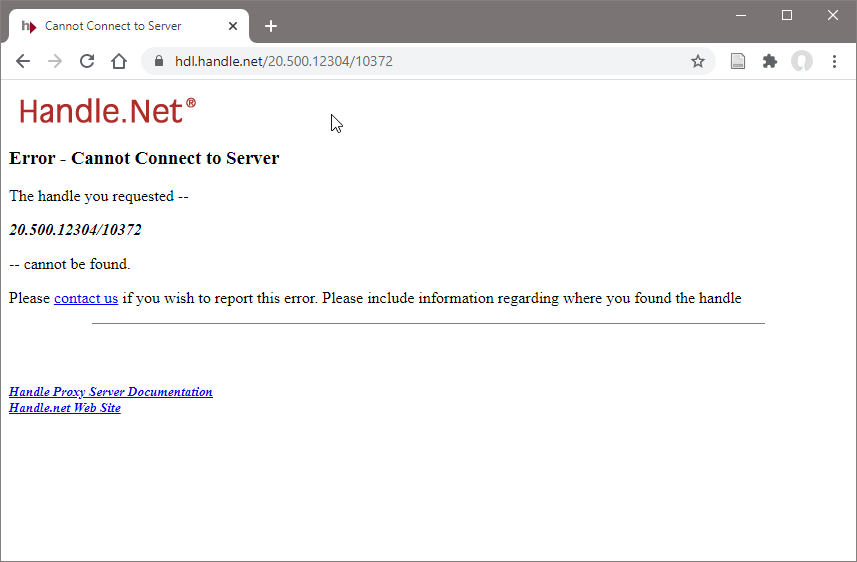
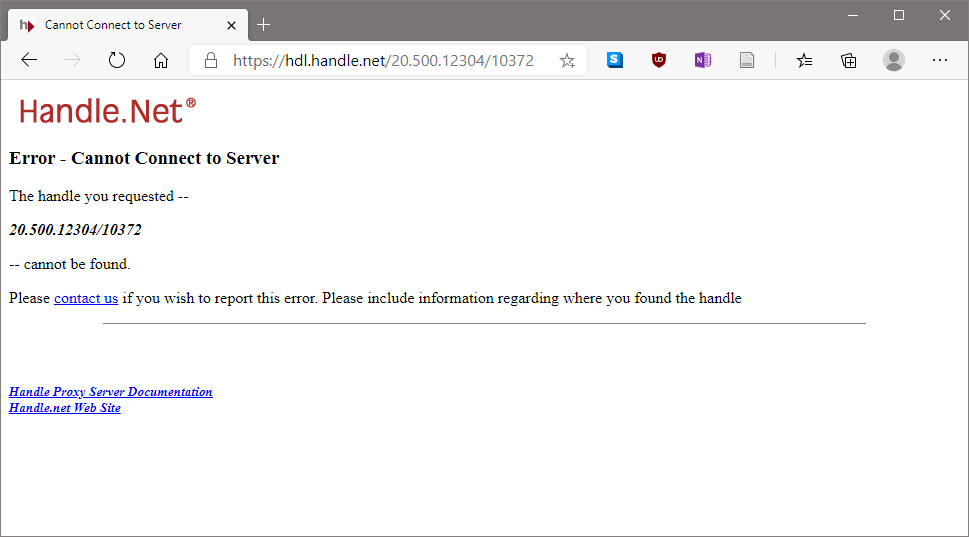
(Yang and Xu 2020)

Yang, Jiaying [杨佳颖], and Xin Xu [许鑫]. 2020. “Semantic Labeling of Advertising Images in Newspapers of the Republic of China Period: Illustrated by Shaoxing Opera Advertisement Published in Xinwen Bao” [民国报纸广告图像资源的语义标注——以《新闻报》所刊的越剧广告为例]. *Library Journal* [图书馆杂志] online first (June): 1–11. https://doi.org/10.13663/j.cnki.lj.2020.07.000.

**Questions:**

- handle working?   
<http://hdl.handle.net/20.500.12304/10372> (time out?)

Chrome Edge

- can ontology be downloaded? No sparql end point mentioned? Can it be found?

- is documented how mirador was adopted to use QGBKSY pdf files?

- can we reproduce the implementation?

- is a website available for the yue ju analysis, or anything re-usable online?

- contact authors!

Yang, Jiaying [杨佳颖], and Xin Xu [许鑫]. 2020. “Semantic Labeling of Advertising Images in Newspapers of the Republic of China Period: Illustrated by Shaoxing Opera Advertisement Published in Xinwen Bao” [民国报纸广告图像资源的语义标注——以《新闻报》所刊的越剧广告为例]. *Library Journal* [图书馆杂志] online first (June): 1–11. https://doi.org/10.13663/j.cnki.lj.2020.07.000.

Corresponding author: Xu Xin, E-mail: [xxu@infor.ecnu.edu.cn](mailto:xxu@infor.ecnu.edu.cn)

Analysis of Yueju Shaoxing Opera 越剧 advertisements from "Xinwen Bao" 《新闻报》, 1923-1949

Data from The National Newspaper Index Database 全国报刊索引数据库

Differentiation between “digitization” 数字化 and “dataization” 数据化

**Related research:**

Ou Junyong 欧俊勇 on Chaozhou opera 潮剧 advertisements in different local materials (newspapers, theatre, fundraisers), method: (manual) counting  
[2] 欧俊勇.民国时期潮剧广告之研究[J].戏剧文学,2014(12): 109-118.

Zhang Yan 张艳 on structure of marriage advertisements   
[3] 张艳.民国征婚广告话语的文本结构探析[J].学理论, 2015(35): 80-81.

Sun Mengshi 孙梦诗 used "Shen Bao" and "Liangyou Pictorial" advertisements   
[4] 孙梦诗.民国上海广告的现代性特征探讨[J].江淮论坛, 2015(2): 181-184.

mentions reports from libraries about newspaper database infrastructure [5] and reviews of external service providers for newspaper digitization [6, 7]  
[5] 杨敏.近代中国报纸数字资源的建设和利用研究[J].图书馆工作与研究, 2014(6): 60-64.  
[6] 肖红, 槐燕.民国报纸数字化实践中的质检问题探析[J].图书馆学研究, 2017(7): 61-78, 87.  
[7] 肖红.民国报纸数字化实践中的主要问题及处理策略[J].图书馆学研究, 2017(4): 22-37.

Introduces some DH methods based on [8] and states that main research areas cover 4 levels of content: the construction of databases and data sets 数据库与数据集的建设, the development of digital humanistic aids and analysis tools 辅助与分析工具, the practice of humanistic research methods and paradigms 范式的实践, and the innovation of inheritance and communication in the humanities field 人文领域传承与传播的创新 [9]. Cites Unsworth who proposed 7 academic elements of digital humanities research: discovery, labeling, comparison, citation, sampling, interpretation and presentation 发现、标注、比较、引用、采样、阐释与呈现 [10].   
[8] Schreibman S, Siemens R, Unswarth J. A companion to digital humanities[M]. John Wiley & Sons, 2008: 251-252.  
[9] 朱本军, 聂华.跨界与融合: 全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述[J].大学图书馆学报, 2016, 34(5): 16-21.  
[10] Unsworth J. Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? [EB/OL]. (2000-05-13)[2020-01-15]. Http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html.

Introduces Semantic description 语义描述 and linked data 关联数据, triples 三元组形式的数据, retrieved through uniform resource identifier (URI) 统一资源标识符 and HTTP protocol 协议调, aiming to build a machine-readable semantic data network 机器可读的语义数据网络  
[11] Tim Berners-Lee. Linked Data[EB/OL].(2006-07-27)[2018-12-31]. https://www.w3.org/DesignIssues/LinkedData.html.

Example genealogy 家谱 ontology based on the bibliographic framework 书目框架 BIBFRAME  
[12] 夏翠娟, 刘炜, 陈涛, 等.家谱关联数据服务平台的开发实践[J].中国图书馆学报, 2016, 42(3):27-38.

Zhai Shanshan 翟姗姗 for intangible cultural heritage: development of domain ontology 领域本体 and linked data 关联数据  
[13] 翟姗姗, 刘齐进, 白阳.面向传承和传播的非遗数字资源描述与语义揭示研究综述[J].图书情报工作, 2016, 60(2): 6-13; 21.

Dong Kun 董坤 ontology model 本体模型 linked data based on the six core categories of the intangible cultural heritage project (ICH-project)  
[14] 董坤.基于关联数据的非物质文化遗产语义化组织研究[J].现代情报, 2015, 35(2): 12-17.

Tan Guoxin 谈国新 et al. added digital resources on the basis of the six core categories, and put digital humanities projects in parallel with other categories  
[15] 谈国新, 侯西龙, 庄文杰.非物质文化遗产多媒体资源语义组织研究[J].图书馆学研究, 2017(24):42-52.

Zhou Yaolin 周耀林 drama information resources included music, stage, repertoire, communication 音乐、舞台、剧目、传播 in the ontology  
[16] 周耀林, 赵跃, 孙晶琼.非物质文化遗产信息资源组织与检索研究路径——基于本体方法的考察与设计[J].情报杂志, 2017, 36(8): 166-174.

semantic annotation of image resources: commonly used annotation methods include three major categories, namely, public annotation 大众标注, machine annotation 机器标注, and expert annotation 专家标注   
[17] 宋宁远, 王晓光.面向数字人文的图像语义标注工具调查研究[J].数字图书馆论坛, 2015(4): 7-14.

Discussion of public and machine labeling, which Zeng Ziming 曾子明 sees as ineffective, because semantic information requires professional knowledge and cannot be displayed in a structured manner   
[18] 曾子明, 周知.面向数字人文的图像语义描述模型研究[J].情报理论与实践, 2018, 41(1): 116-121.

semantic annotation models 语义标注模型 based on expert knowledge include: a three-layer semantic hierarchy model 三层语义层次模型 composed of the underlying feature layer 底层特征层, the object layer 对象层 and the semantic concept layer 语义概念层, the pyramid semantic hierarchy model 金字塔语义层次模型, and the semantic annotation model 语义标注模型 based on ontology 基于本体   
[19] 陈金菊.图像语义标注研究综述[J].图书馆学研究, 2017(18): 2-7, 20.

Wang Xiaoguang 王晓光 et al. uses a three-layer semantic hierarchy model to fine-grain 细粒度标注 the behavior and emotions of the characters in the Dunhuang murals   
[20] 王晓光, 徐雷, 李纲.敦煌壁画数字图像语义描述方法研究[J].中国图书馆学报, 2014, 40(1): 50-59.

Introduces IIIF, series of open image interaction tool resources and standards, for image browsing 浏览, integration 整合, presentation 呈现, search 搜索, authorization 授权. Examples from Stanford and Princeton   
[21] Technical Details, IIIF [EB/OL].[2019-09-1].https://iiif.io/technical-details/#beta-specifications

Current problems with Republican newspaper advertising are Image storage quality 图像存储质量, bottlenecks in professional knowledge 专业知识瓶颈, and lack of in-depth indexing 缺乏深度标引. Currently, research on the use of advertising resources in newspapers and periodicals in the Republic of China is still blank. Semantic-based image labeling can be used to solve this

(2) Introduction of Yueju history and features

(3) historical development of ads

(**3.2) Utilizing Images**

**3.2.1) Improve legibility:**

poor quality of photocopying; right-to-left word order 从右到左的语序 and vertical arrangement 竖排排列的格式, in traditional Chinese characters.

printing and typesetting technology limited, dense text, distinction was mainly based on the size and distribution of words

at present, often the resources are in full pages, and sometimes ads are cut. It’s necessary: to strengthen the construction of basic auxiliary reading tools 基础辅助阅读工具, such as marking borders 标注框线 to divide different semantic content in advertisements, and providing terminology hints提供术语提示项 to improve reading efficiency

because of repeated use of almost similar ads, functional requirement for comparing different versions of advertisements (also for repeated labeling)

**3.2.2) In-depth indexing 深度标引 to support quantitative analysis**

Currently indexed key fields are not applicable to drama advertisements, e.g. actors and the play itself

in-play repertoire information 戏内剧目信息 and out-of-play performance information 戏外演出信息

the size of the page occupied by the image

**3.2.3) Knowledge association 知识关联**

Yue Opera specific concepts 概念集, implicit 隐含 knowledge, such as genre, theater troupe, script, stage, 流派、剧团、剧本、舞台can supplement the information on the advertisement

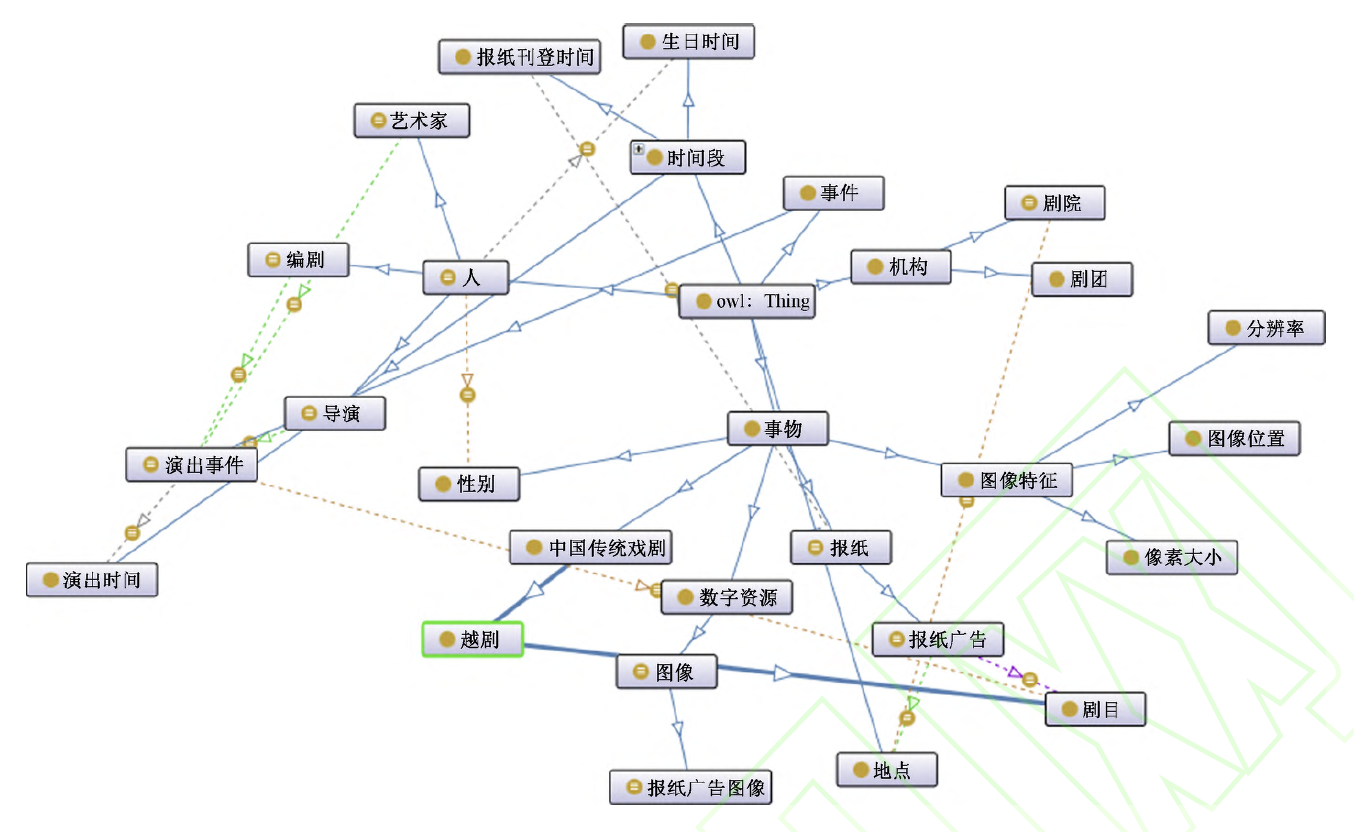
4) semantic description for Yue opera (Figure 5)

**4.2) advertising ontology 广告本体 of Yue Opera**

image resources 图像资源的实体 together with performance events 演出事件实体, refers to the intangible cultural heritage model ICHproject, including: people, places, repertoires, time periods, 人、地点、剧目、时间段

three levels of descriptive metadata, features of the digital image 数字图像底层特征的字段, the resource metadata information 资源元数据信息 of the newspaper to which the advertisement belongs, and the text and illustrations presented in the advertisement 文字与插图的信息

establish semantic associations, that newspaper advertising resources map newspaper advertisements 报纸广告资源映射报纸广告, and newspaper advertisements map performance events 报纸广告映射演出事件, and the people, repertoires, institutions, 人、剧目、机构 etc., have defined attributes 一定属性

 Figure 6: The ontology model

External Relations to standards (CBDB, CHGIS, names, geographical names), textual resources (Shanghai Yueju History" 《上海越剧志》, other primary/secondary materials), multimedia resources (old photos, images, historical videos), relationship between the characters

**5) implementation**

Using IIIF and Mirador viewer, images from QGBKSY

Workflow: image resources - semantic connotation - user annotations to preprocess the images, remove duplicates, upload them to the server, and import them in the form of URIs

Metadata based on ontology, labeling tool

URI image link via API to Mirador, extension Simple Annotation Server 简单标注服务器

In QGBKSY search for "Yueju", 3611 hits, remove duplicates, get 2818 images (pages), visualization in Mirador, there boxes and annotation

**6) outcome from sample data**

Figure 10 The distribution of Yue Opera advertisements published in "Xinwen Bao" by year

Problem: QGBKSY only provides pdf, is not friendly 并不友好 to the reading support of text-intensive 文字密集 newspaper advertisement files

Better: using the Mirador browser’s changeLayout function for comparison

**7) conclusion**

Based on six-element conceptual model of ICH-project

constructed an ontology, applicable to other traditional drama newspaper advertisements

labeling images via IIIF in Mirador, but still optimized for user experience, rather than analysis and statistics

**The data link** address of this article: <http://hdl.handle.net/20.500.12304/10372> (dead link)