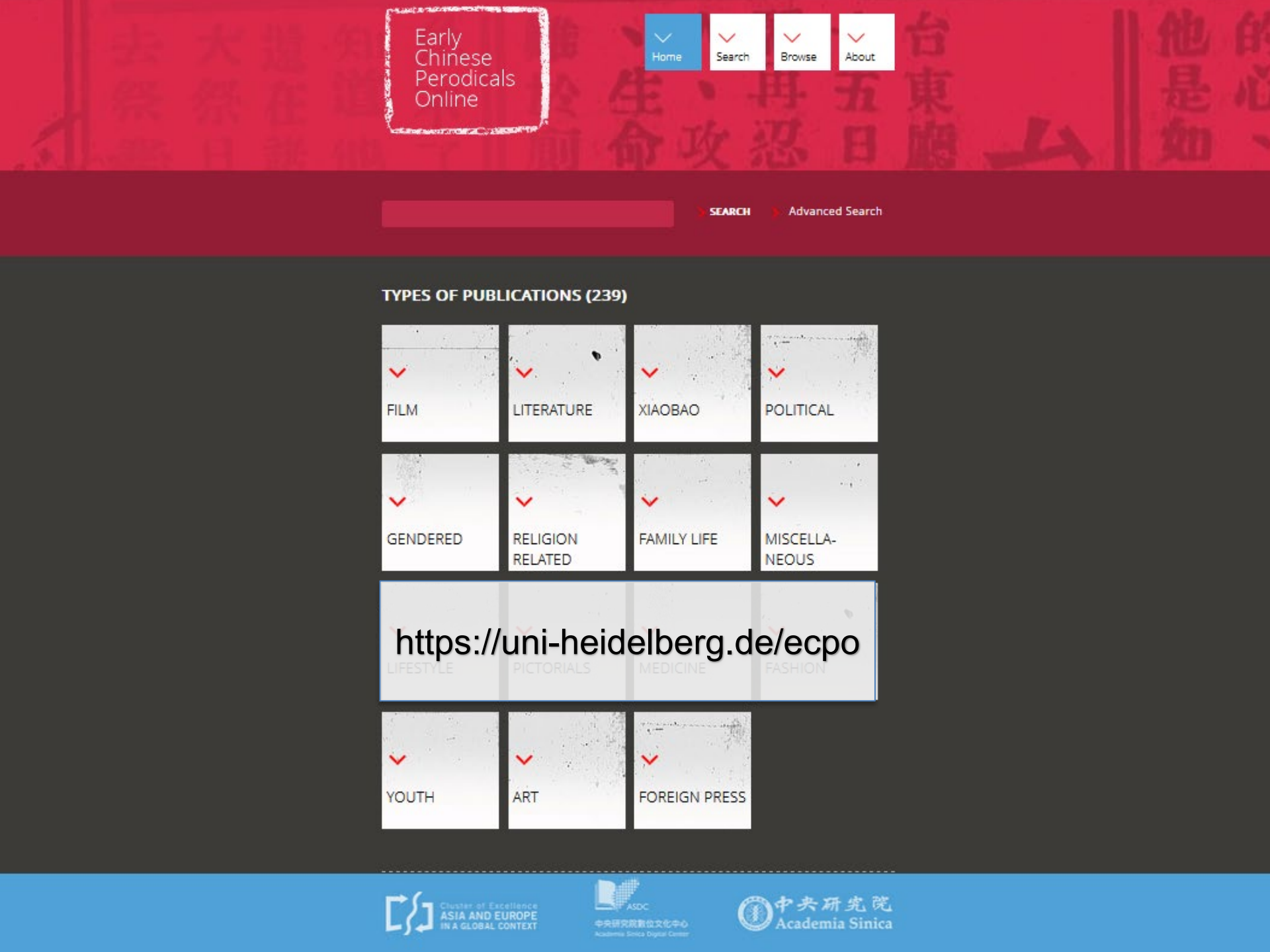
Heidelberg Centre for Transcultural Studies

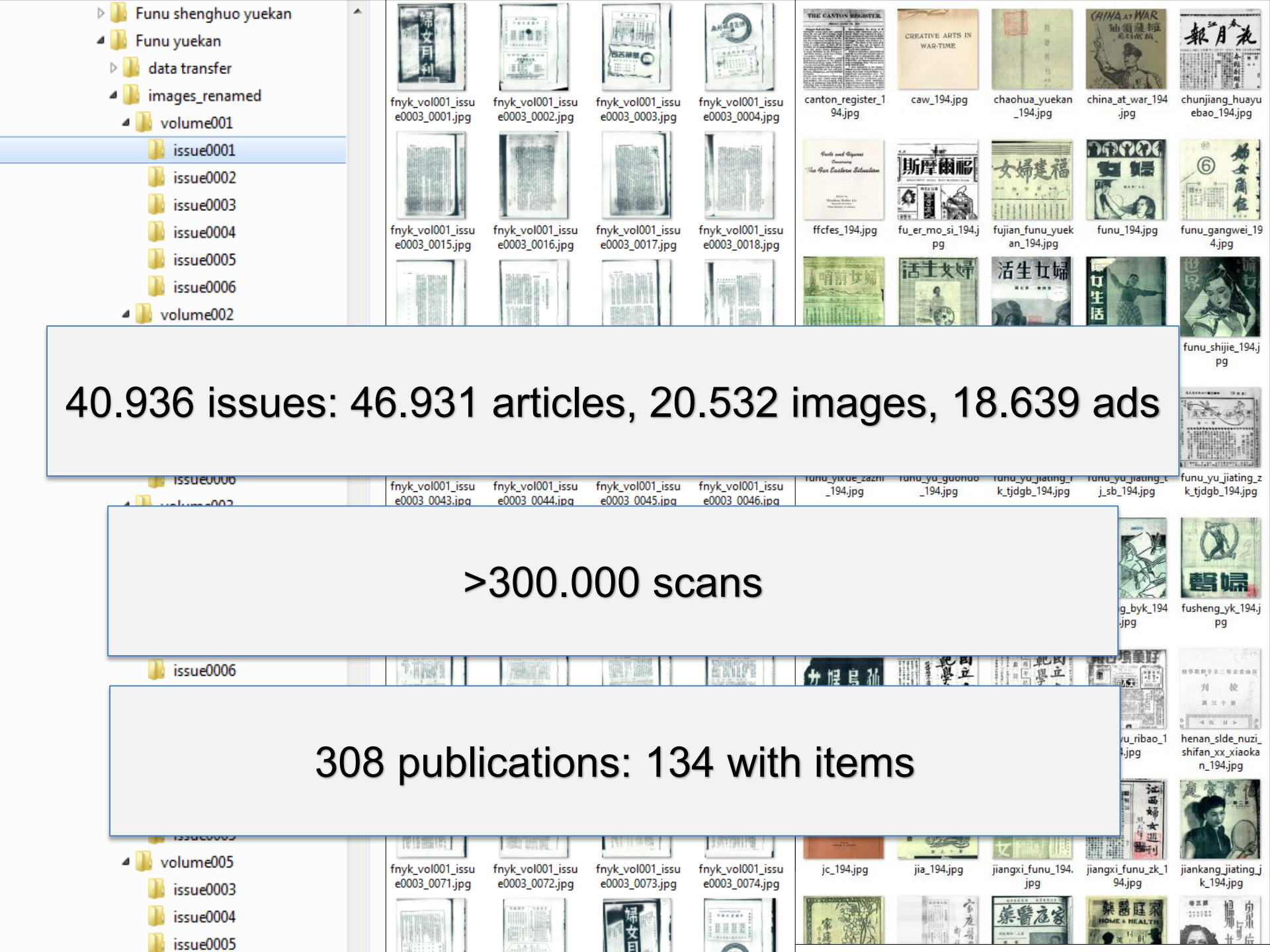
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

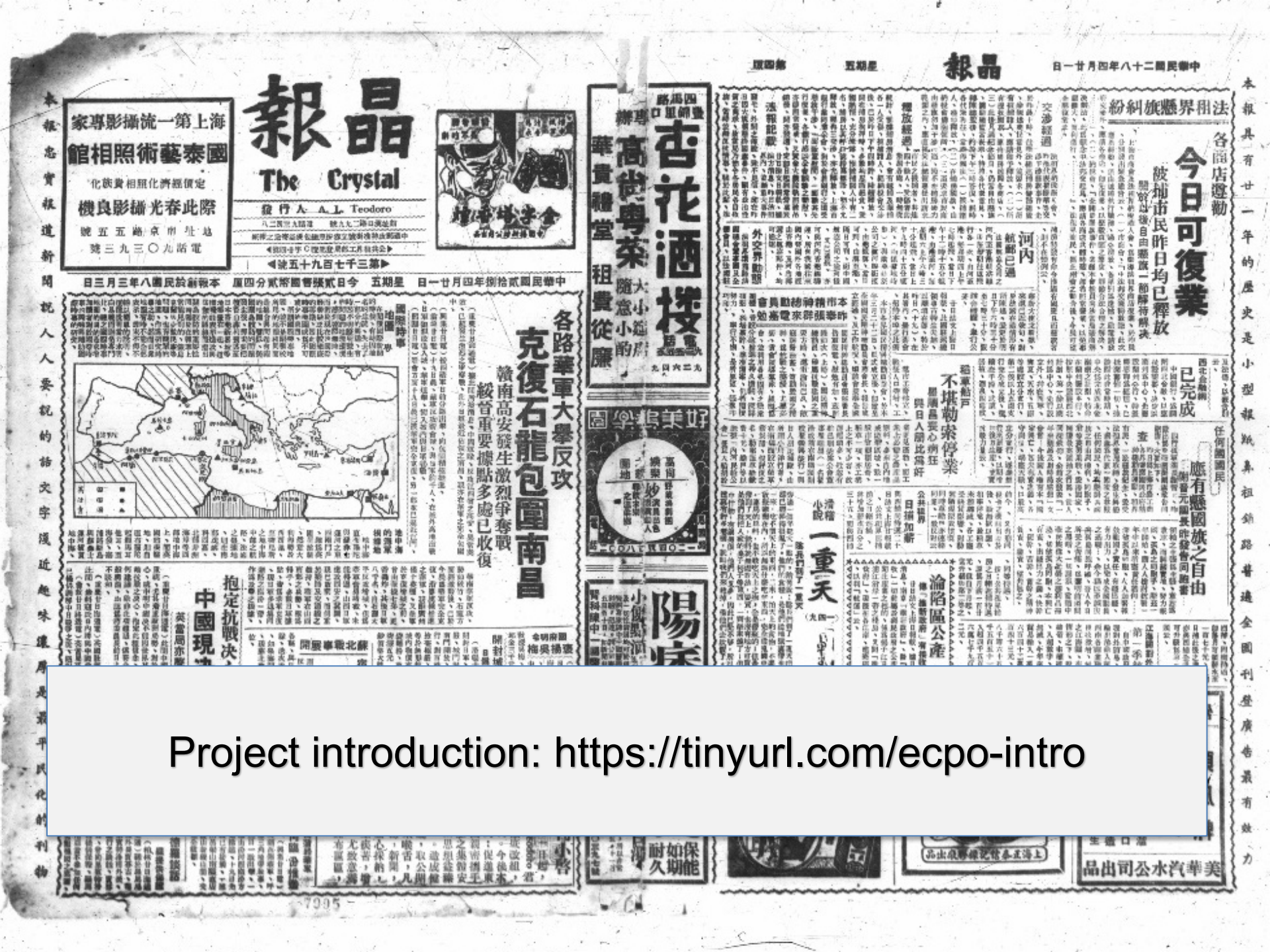
Semantic segmentation and document layout recognition: Approaches towards full text recognition of early Chinese newspapers

Matthias Arnold | Uni Heidelberg · HCTS · HRA

NewsEye International Conference “What’s Past is Prologue” | 2021-03-17


Early Chinese Perodicals Online
Home
Search
Browse
About
SEARCH
Advanced Search
TYPES OF PUBLICATIONS (239)
FILM
LITERATURE
XIAOBAO
POLITICAL
GENDERED
RELIGION RELATED
FAMILY LIFE
MISCELLA-NEOUS
https://uni-heidelberg.de/ecpo
YOUTH
ART
FOREIGN PRESS
Cluster of Excellence ASIA AND EUROPE IN A GLOBAL CONTEXT
ASDC
Academia Sinica Digital Center
中央研究院
Academia Sinica


Funu shenghuo yuekan
Funu yuekan
data transfer
images_renamed
volume001
issue0001
issue0002
issue0003
issue0004
issue0005
issue0006
volume002
issue0006
volume005
issue0003
issue0004
issue0005
40.936 issues: 46.931 articles, 20.532 images, 18.639 ads
>300.000 scans
308 publications: 134 with items



Project introduction: https://tinyurl.com/ecpo-intro

Towards full text - 1

First steps

Entertainment newspapers 小報 *xiao bao:*
Jing bao 晶報 (The Crystal) 1919-1940

Expanding data: towards fulltext

- Manual typing not feasible
- Professional double-keying very expensive
- Challenges for OCR
  - Image: secondary copies - noisy, stains, scratches, etc
  - Document: dense layout, normal segmentation fails
  - Characters: special characters (emphasis), handwriting

ABBYY® Cloud OCR SDK Demo

Customer Login
Start free trial

Home | Product Tour | Plans and Pricing | Support | SLA | Security | Demo | For Students | About us

Load source file (*.png or *.jpg): Browse... jb_0016_1919-04-18_0002+0003.jpg

I agree with the Terms of use

Or paste url to source file (*.png or *.jpg):

Recognize

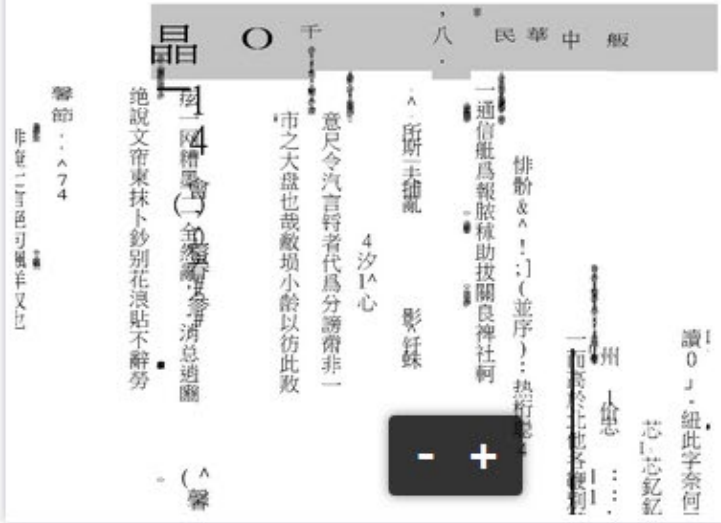
Recognize whole image

Recognition languages	Profile	Image source
Chinese Traditional ✕	DocumentConversion	Auto

More languages

Source

Result



Report bad result

ABBYY Cloud OCR SDK Demo

Customer Login
Start free trial

Home | Product Tour | Plans and Pricing | Support | SLA | Security | Demo | For Students | About us

Load source file (*.png or *.jpg): Brows... of use

Or paste url to source file (*.png or *.jpg):

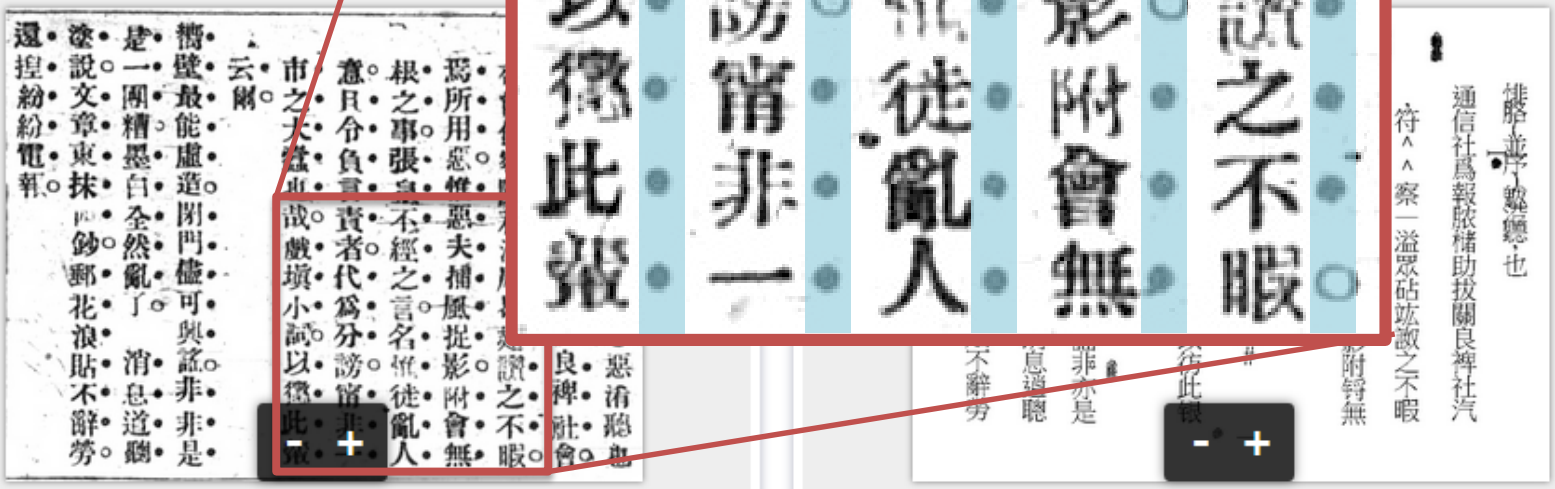
Recognize whole image

Recognition languages: Chinese Traditional

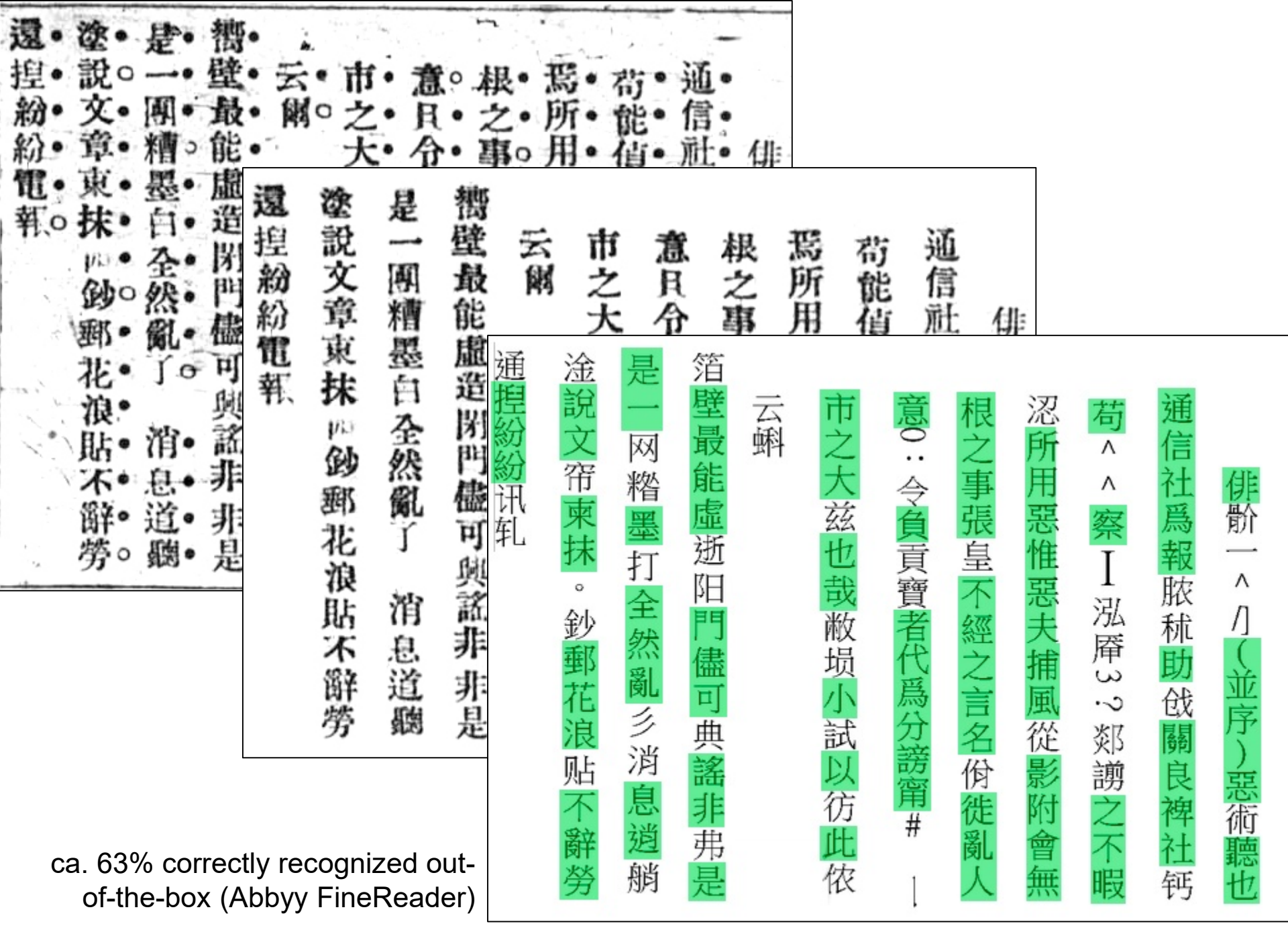
Profile: DocumentConversion

More languages

Source



Report bad result


俳骸一ㅅ刀(並序)惡術聽也
通信社爲報腋秣助餓關良裨社鈣
苟ㅅㅅ察I泓屝3.?郊謗之不暇
涊所用惡惟惡夫捕風從影附會無
根之事張皇不經之言名佾徙亂人
意○:令負貢寶者代爲分謗甯#
市之大茲也哉敝埙小試以彷此依
云蝌
箔壁最能虛造閉門儘可典謠非弗是
是一冈糌墨打全然亂彡消息逍艄
涂說文帘柬抹。鈔郵花浪貼不辭勞
通揑紛紛讯軋

ca. 63% correctly recognized out-of-the-box (Abbyy FineReader)

Towards full text - 2

Segmentation

晶報

The Crystal

發行人 A. L. Teodoro

第三千七百九十五號

中華民國貳拾捌年四月廿一日 星期五 今日貳張售國幣貳分四厘 本報創於民國八年三月三日

各路華軍大舉反攻
克復石龍包圍南昌
贛南高安發生激烈爭奪戰
綏晉重要據點多處已收復

國際時事 地圖

抱定抗戰決心
中國現決不談和
英當局亦聲明未提議調解事

蘇北戰事展開
華軍甚得手

國府明令 褒揚吳梅

開封城緊閉

本報復刊小啓

本報自四月一日起，由美商A. L. Teodoro君為發行人，繼續出版，用於今日復刊。今後本報有二大方針：（一）促進東西兩大民主國家……之集體安全……，此其一。……此其二。……請希明鑒。

法租界懸旗糾紛
各商店遵勸
今日可復業
被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

交涉經過

釋放經過

逃報記載

河內
航郵已通

外交界動態

本市精神總動員會 昨奉張群來電嘉勉

西北公路網 已完成

任何國國民 應有懸國旗之自由
明常元國長昨發書面談話

查

稻草鉛戶
不堪勒索停業
與日人關比為奸 單幫商販心病狂

滑稽小說
一重天
（一四九）
陸兵們到了一重天

日捐加衛
公共租界

淪陷區公產

工部局對外籍營業 第一季統計

-7995-

晶報

The Crystal

發行人 A. L. Teodoro

◀第三千七百九十五號▶

中華民國貳拾捌年四月廿一日 星期五 今日貳張 售國幣貳分四厘 本報創於民國八年三月三日

各路華軍大舉反攻

克復石龍包圍南昌

贛南高安發生激烈爭奪戰

綏晉重要據點多處已收復

國際時事地圖

抱定抗戰決心 中國現決不談和

英當局亦聲明未談調解事

蘇北戰事展開

華軍甚得手

本報復刊小啓

今日可復業

被捕市民昨日均已釋放

各商店遵勸

法租界懸旗糾紛

交涉經過

釋放經過

法報記載

河內

航辦已通

外交界動態

本市精神總動員會 昨舉張群來電嘉勉

不堪勒索停業

稻葉船戶

已完成

西北公路網

應有懸國旗之自由

任何國籍商

滑稽小說

一重天

一四九

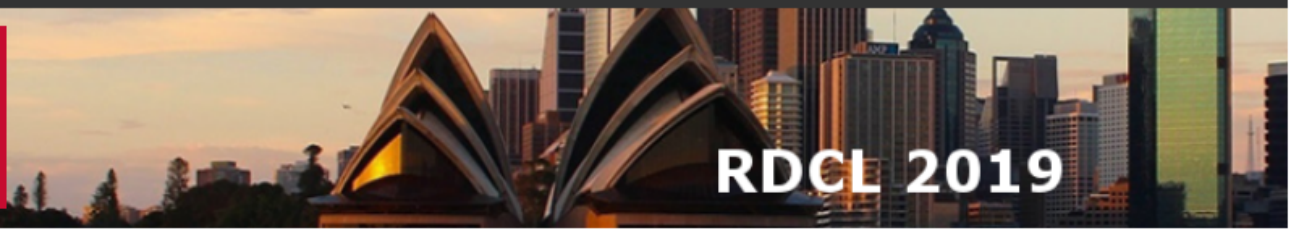
淪陷區公產

日擴加緊

第一季統計

PRImA

Pattern Recognition & Image Analysis Research Lab



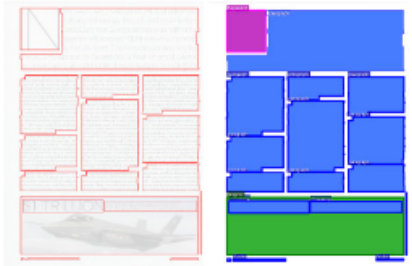
Home
Schedule
Organisers
Register
Resources / Details
Submissions
Results

RDCL2019 ICDAR Competition on Recognition of Documents with Complex Layouts

Overview

The competition presents challenges for page segmentation, region classification, and text recognition in an end-to-end scenario. The dataset contains scanned pages from contemporary magazines and technical articles. Participants will be provided with know-how and tools that aid the development or extension of their page analysis systems.

Dataset	Pages of magazines and technical articles
Main challenge	Page segmentation and region classification



https://www.primaresearch.org/RDCL2019/

Document layout recognition and page segmentation:

Page structuring elements - separators

中華民國二十八年四月廿一日　星期五　晶報　第四版

法租界懸旗糾紛

各商店遵勸 今日可復業

被捕市民昨日均已釋放

關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、爲勸導法租界商店復業、於昨晨、分別發出緊急通告云、（一）市商會、「中法邦交素稱敦睦、此次懸旗糾紛、係由越捕執行職務、過分所致、各界引爲遺憾、除電請政府交涉外、應請各商店、即先復業、以堅毅貞固之態度、靜候合法合理之解決、……」（二）納稅會、「在關於懸旗事件、業經本會推員向法當局磋商解決辦法、此爲顧念中法邦交起見、應請各商店共體時艱、即希照常營業、以免事態擴大、而利進行、……」法租界商民、經此兩會之勸告後、今晨可望全部復業云、

交涉經過

法租界納稅華人會、昨推代表晉謁法總巡捕房警務總監、交涉干涉懸旗事件、先於昨晨十時、往晤法總巡捕房警務總監、除陳述商民意見外、並要求（一）所有被捕諸人、應請即予釋放、（二）所有沒收國旗、應由捕房送回各商店、（三）此後凡遇紀念節日、仍當自由懸旗、總監當表示、待將各代表意見、轉陳總監後、約於下午三時答復、屆時各代表再往、當得辦法（一）被捕諸人、准即釋放、（二）在捕房國旗、由納稅會備函領回、（三）至要求市民自由懸旗勿加干涉一節、因不在捕房權力範圍之內、應直接向法總領事提出請求、

釋放經過

市民之被捕者、共達四十餘人、受傷者因散居各處、一時無從統計、惟據捕房消息、曾有越捕及華捕各一人受傷、被捕諸人、經納稅會交涉後、已於昨日下午四時許、陸續釋放、聞在捕房拘押時、多數均忍痛絕食、愛國熱情、充分流露、惟被捕諸人中有二名、捕房以其呼口號措辭、初不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致意慰問、一面勸其先行復業、各銀行感以公會誠摯之勸慰、亦即照常營業、又賞金大戲院執照被吊銷後、聞亦發還、定即日起照常營業云

法報記載

廿日法文日報載、據法方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、頗有政治性質之表示、故當局乃飭令各居民各自收旗、當時並無反抗情事、關於此點、法捕房特發有命令惟遇有國慶日而懸旗者、則不在禁例云、

河內航郵已通

法國航空公司之河內至香港航空線、現每星期往返飛行各一次、由河至港、每星期四上午七時起飛、當日下午十二時十五分抵港、由港至河、每星期六上午六時三十分起飛、當日上午九時四十五分抵河、（均係當地時間）、與歐亞航空公司之滇河線銜接、由港抵滇、當日可到、由滇至港、隔日可到、而中國航空公司之渝河線、又已通航、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐華新大使戈斯默、來滬已近匝旬、對於法國在華情形有所陳述、爰定於明日下午六時三十分至七時三十分、假該會禮廳、舉行公宴、

廿日法文上海日報載、比國駐滬總領事古特生夫婦、以離滬在即、特於昨日（十九）、在其署內、舉行茶會、招待本埠之比僑

本市精神總動員會 昨奉張群來電嘉勉

本市各界國民精神總動員協會、於本年三月二十二日、正式成立後、即電呈全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次（上略）、國民精神總動員、確爲抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅益甚、……

西北金融網已完成

中國銀行、爲調劑西北金融、以應抗戰需要、決定以蘭州爲中心、先籌設蘭州分行、特派鄭恩錦爲經理、已抵蘭籌備一切、據談際此非常時期、中央爲完成西北金融網之計劃、特命在蘭州籌設本行、按照中央發展西北計劃、第一除以蘭州爲中心、先行設立外、并在西寧、寧夏、天水、五原等處設立分支行、第二俟以上五處支行完全成立後、繼續在平涼、武威、張掖、酒泉四處設立分支行、逐漸推行至新疆、以期增加後方生產、充實抗戰力量云、

任何國國民 應有懸國旗之自由

謝晉元團長昨發告同胞書

稻草船戶不堪勒索停業

瞿順昌喪心病狂 與日人朋比爲奸

公共租界日捕加薪

滑稽小說 一重天

淪陷區公產

晶報 The Crystal

發行人 A. L. Teodoro
館址漢口路二九九號　電話九三四二八
中華郵政特准掛號立券按照總包優益寄送之報紙
◀公共租界工部局登記證C字廿四號▶
◀第三千七百九十五號▶
中華民國貳拾捌年四月廿一日　星期五　今日貳張售國幣貳分四厘　本報創於民國八年三月三日

本報忠實報道新聞說人人要說的話文字淺近趣味濃厚是最平民化的刊物

國際時事地圖學

地中海的海軍根據地

各路華軍大舉反攻 克復石龍包圍南昌

贛南高安發生激烈爭奪戰 綏晉重要據點多處已收復

（重慶廿日海通電）據此間所接消息、中國軍隊、反攻江西省之高安、異常奏……

（興集廿日電）綏西華軍日前分路出擊、向包圍積極挺進、……

（翁源廿日電）十九日晨、華軍反攻新會城、與日軍約千人、在城外高地激戰、日軍狼狽敗退入城、華軍追擊、衝入城內與日軍巷戰、

（韶關廿日電）新會方面十九日晨已被華軍完全克復、另一部並已逼近江門、

抱定抗戰決心 中國現決不談和

英當局亦聲明未提議調解事

蘇北戰事展開

宿遷沭陽各縣 華軍甚得手

國府明令褒揚吳梅

本報復刊小啟

本報自四月一日起、由美商A. L. Teodoro君爲發行人、徹底改組、用於今日復刊、今後本報有二大趨向：促進東西兩大民主國親密攜手、謀太平洋上之集體安全、以求民治思想之實現、此其一、造成健全正大之輿論、取公開態度、爲大衆喉舌、凡有事業、小品、新聞、趣事、靡不虛心採納、而於解除民生疾苦、增進人羣福祉、尤致意焉、此其二、謹布區區、諸希明鑒。

法租界懸旗糾紛

各商店遵勸 今日可復業

被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、爲勸導法租界商店復業、於昨晨、分別發出緊急通告云、（一）市商會、「中法邦交素稱敦睦、此次懸旗糾紛、係由越捕執行職務、過分所致、各界引爲遺憾、除電請政府交涉外、應請各商店、即先復業、以堅毅貞固之態度、靜候合法合理之解決、……」、（二）納稅會、「查關於懸旗事件、業經本會推員向法當局磋商解決辦法、此爲顧念中法邦交起見、應請各商店共體時艱、即希照常營業、以免事態擴大、而利進行、……」、法租界商民、經此兩會之勸告後、今晨可望全部復業云、

交涉經過

法租界納稅華人會、昨推代表瞿振華等交涉干涉懸旗事件、先於昨晨十時、往晤法總巡捕房警務總監、除陳述商民意見外、並要求（一）所有被捕諸人、應請即予釋放、（二）所有沒收國旗、應由捕房送回各商店、（三）此後凡遇紀念節日、仍當自由懸旗、總巡總監當表示、待將各代表意見、轉陳總監後、約於下午三時答復、屆時各代表再往、當局得辦法（一）被捕諸人、准即釋放、（二）在捕房國旗、由納稅會備函領回、（三）至要求市民自由懸旗勿加干涉一節、因不在捕房權力範圍之內、應直接向法總領事提出請求、捕房特發有命令惟遇有國慶日而懸旗者、則不在禁例云、

釋放經過

市民之被捕者、共達四十餘人、受傷者因散居各處、一時無從統計、惟據捕房消息、曾有越捕及華捕各一人受傷、被捕諸人、經納稅會交涉後、已於昨日下午四時許、陸續釋放、聞在捕房拘押時、多數均忍痛絕食、愛國熱情、充分流露、惟被捕諸人中有二名、捕房以其呼口號精罵標語、初不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致意慰問、一面勸其先行復業、各銀行感以公會殷摯之勸慰、亦即照常營業、又黃金大戲院執照被吊銷後、聞亦發還、定即日起照常營業云

法報記載

廿日法文日報載、據法方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、顯有政治性質之表示、故當局乃飭令各居民各自收旗、當時並無反抗情事、關於此點、法

河內航郵已通

法國航空公司之河內至香港航空線、現每星期往返飛行各一次、由河至港、每星期四上午七時起飛、當日下午十二時十五分抵港、由港至河、每星期六上午六時三十分起飛、當日上午九時四十五分抵河、（均係當地時間）、與歐亞航空公司之滇河線銜接、由港抵滇、當日可到、由滇至港、隔日可到、而中國航空公司之渝河線、又已通航、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐華新大使戈斯默、來滬已近浹旬、對於法國在華情形有所陳述、爰定於明日下午六時三十分至七時三十分、假該會禮廳、舉行公宴、

廿日法文上海日報載、比國駐滬總領事古特生夫婦、以離滬在即、特於昨日（十九）、在其署內、舉行茶會、招待本埠之比僑及法僑、以表告別云、

西北金融網 已完成

中國銀行、爲開闢西北金融、以應抗戰需要、決定以蘭州爲中心、先籌設蘭州分行、特派鄭恩卿爲經理、已抵蘭籌備一切、據談際此非常時期、中央爲完成西北金融網之計劃、特命在蘭州籌設本行、按照中央發展西北計劃、第一除以蘭州爲中心、先行設立外、并在西寧、寧夏、天水、五原等處設立分支行、第二俟以上五處支行完全成立後、繼續在平涼、武威、張掖、酒泉四處設立分支行、逐漸推行至新疆、以期增加後方生產、充實抗戰力量云、

本市精神總動員會 昨奉張群來電嘉勉

本市各界國民精神總動員協會、於本年三月二十二日、正式成立後、即電呈全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次（上略）、國民精神總動員、確屬抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅益甚、策動愛國之精神、爲抗戰之聲援、尤屬必要、貴會領導黨紳、組織協會、並利用各界固有之職業團體、普設分途推進之持久機關、深謀碩劃、殊堪嘉許、應准備案、仍希繼續努力、奉行不懈、是所厚望、張羣叩巧印、

稻草船戶 不堪勒索停業

瞿順昌喪心病狂 與日人朋比爲奸

滬市工廠林立、戰後人口突增、工業更見蓬勃、而工業上必需之農產品原料、多來自內地或游擊區域、致一般市儈壟斷居奇、稻草一項、亦工業上之不可少者、故運滬頗多、近忽有前邑廟柴業公會董事瞿順昌（一名瞿洽成前在南碼頭開設瞿義興柴行）與日人朋比操縱、由所謂小川洋行者、在南碼頭設立柴草薪炭部、假評價之名、勒索每草船繳費十元、船戶等以既被「內河民船公會」征苴大船捐於前、今、又有形同稅卡之機關出現於後、致無款以應、相率停業、致稻草來源驟減、各工廠受缺貨影響、對勒索陋規深表憤慨、同時草船戶亦連絡同業、一致反對云

任何國國民 應有懸國旗之自由

謝晉元團長昨發告同胞書

四行孤軍團長謝晉元、昨發出爲取締懸掛國旗敬告同胞書、大要如下、

查昨日爲上海市農工商各界響應國民政府精神總動員運動、租界市民、一致懸旗紀念、爰受法租界當局取締、發生糾紛、國旗爲代表一國家之尊嚴、任何國民、均有懸掛其國旗之自由與權利、同胞乎、君等英勇果敢、愛護國家、擁護最高領袖之行爲、世人同深欽仰、余前次發表「一年來觀感及對時局感想」文會曾……今後國人唯有本「國家興亡、匹夫有責」之義、各守崗位、竭盡天職、以最高領袖之主張爲主張、意志、而腳踏實地、國、孤島之同胞……年開始、願人人……年半以來所盡責……深望孤島同胞、……效祖國之實任、……有力出力、以完……之基礎……、余今……向孤島同胞呼籲……所受之苦痛、……之黑暗、一切惡……爲、切望孤島同……華民族偉大光榮……有責、出錢出力……、硬幹、苦幹、……負責、盡責、……

公共租界 日捕加薪

與西捕同待遇

二十日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕……房之日……、原爲……元、但……當升級……

按：……、同等……

滑稽小說 一重

陸兵們到……

窮途一個年紀大一點的，陪都已經弄得很亂了。他們的一來，吃不慣天上的東西，放些毒藥在內，所以無論什帶，連馬的馬料，也不敢吃天到了天上，燃料就無處投是他們就把人家的桌子椅子中什物，有限得很，他們人還要有許多老櫃，就叫我們……

中

各路華軍大舉反攻

法租界懸旗糾紛

各商店遵勸 今日可復業

被捕市民昨日均已釋放

關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、爲勸導法租界商店復業、於昨晨、分別發出緊急通告云、(一)市商會、「中法邦交素稱敦睦、此次懸旗糾紛、係由越捕執行職務、過分所致、各界引爲遺憾、除電請政府交涉外、應請各商店、即先復業、以堅毅貞固之態度、靜候合法合理之解決、」……(二)納稅會、「查關於懸旗事件、業經本會推員向法當局磋商解決辦法、此爲顧念中法邦交起見、應請各商店共體時艱、即希照常營業、以免事態擴大、而利進行、……」、法租界商民、經此兩會之勸告後、今晨可望全部復業云、

交涉經過

法租界納稅華人會、昨推代表瞿振華等交涉干涉懸旗事件、先於昨晨十時、往晤法總巡捕房警務總監、除陳述商民意見外、並要求(一)所有被捕諸人、應請即予釋放、(二)所有沒收國旗、應由捕房送回各商店、(三)此後凡遇紀念節日、仍當自由懸旗、總監當表示、待將各代表意見、轉陳總領事後、約於下午三時答復、屆時各代表再往、當商得辦法(一)被捕諸人、准即釋放、(二)在捕房國旗、由納稅會備函領回、(三)至要求市民自由懸旗勿加干涉一節、因不在捕房權力範圍之內、應直接向法總領事提出請求、

釋放經過

市民之被捕者、共達四十餘人、受傷者因散居各處、一時無從統計、惟據捕房消息、曾有越捕及華捕各一人受傷、被捕諸人、經納稅會交涉後、已於昨日下午四時許、陸續釋放、聞在捕房拘押時、多數均忍痛絕食、愛國熱情、充分流露、惟被捕諸人中有二名、捕房以其呼口號寫標語、初不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致意慰問、一面勸其先行復業、各銀行感以公會觀摩之勸慰、亦即照常營業、又黃金大戲院執照被吊銷後、聞亦發還、定卽日起照常營業云

法報記載

廿日法文日報載、據法租界方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、顯有政治性質之表示、故當局乃飭令各居民各自收旗、當時並無反抗情事、關於此點、法捕房特發有命令惟遇有國慶日而懸旗者、則不在禁例云、

河內航郵已通

法國航空公司之河內至香港航空線、現每星期往返飛行各一次、由河至港、每星期四上午七時起飛、當日下午十二時十五分抵港、由港至河、每星期六上午六時三十分起飛、當日上午九時四十五分抵河、(均係當地時間)、與歐亞航空公司之滇河線銜接、由港抵滇、當日可到、由滇至港、隔日可到、而中國航空公司之渝河線、又已通航、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐華新大使戈斯默、來滬已近浹旬、對於法國在華情形有所陳述、爰定於明日下午六時三十分至七時三十分、假該會禮爵、舉行公宴、

廿日法文上海日報載、比國駐滬總領事古特生夫婦、以離滬在卽、特於昨日(十九)、在其署內、舉行茶會、招待本埠之比僑及法僑、以表告別云、

本市精神總動員會 昨奉張群來電嘉勉

本市各界國民精神總動員協會、於本年三月二十二日、正式成立後、即電呈全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次(上略)、國民精神總動員、確屬抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅滋甚、策勵愛國之精神、爲抗戰之聲援、尤屬必要、貴會領導羣衆、組織協會、並利用各界固有之職業團體、普設分途推進之持久機關、深謀碩劃、殊堪嘉許、應准備案、仍希繼續努力、奉行不懈、是所厚望、張羣叩巧印、

西北金融網已完成

中國銀行、爲調劑西北金融、以應抗戰需要、決定以蘭州爲中心、先籌設蘭州分行、特派鄭恩卿爲經理、已抵蘭籌備一切、據談際此非常時期、中央爲完成西北金融網之計劃、特命在蘭州籌設本行、按照中央發展西北計劃、第一除以蘭州爲中心、先行設立外、并在西寧、寧夏、天水、五原等處設立分支行、第二俟以上五處支行完全成立後、繼續在平涼、武威、張掖、酒泉四處設立分支行、逐漸推行至新疆、以期增加後方生產、充實抗戰力量云、

稻草船戶 不堪勒索停業

瞿順昌喪心病狂 與日人朋比爲奸

滬市工廠林立、戰後人口突增、工業更見蓬勃、而工業上必需之農產品原料、多來自內地或游擊區域、致一般市儈壟斷居奇、稻草一項、亦工業上之不可少者、故運滬頗多、近忽有前邑廟柴業公會董事瞿順昌(一名瞿敍發興柴行)與日人朋比操縱、由所謂小川洋行者、在南碼頭設立柴草薪炭部、假評價之名、勒索稻草船徵費十元、船戶等以旣被「內河民船公會」征且大船捐於前、今、又有形同稅卡之機關出現於後、致無款以應、相率停業、致稻草來源驟減、各工廠受缺貨影響、對勒索陋規深表憤慨、同時草船戶亦連絡同業、一致反對云

任何國國民 應有懸國旗之自由

謝晉元團長昨發告同胞書

四行孤軍團長謝晉元、昨發出爲取締懸掛國旗敬告同胞書、大要如下、昨日爲上海市農工商各界響應國民政府精神總動員運動、租界市民、一致懸旗紀念、爰受法租界當局取締、發生糾紛、國旗爲代表一國家之尊嚴、任何國民、均有懸掛其國旗之自由與權利、同胞乎、君等英勇果敢、愛護國家、擁護最高領袖之行爲、世人同深欽仰、余前次發表「一年來觀感及對時局感想」一文曾言…今後國人唯有本「國家興亡、匹夫有責」之義、各守崗位、竭盡天職、以最高

查

領袖之主張爲主張……意志、而腳踏實地……國、孤島之同胞……年開始、願人人……年半以來所盡責……深望孤島同胞、……效祖國之實任、……有力出力、以完……之基礎…、余今……向孤島同胞呼籲……所受之苦痛、……之黑暗、一切惡……華民族偉大光榮……爲、切望孤島同胞……有責、出錢出力……、硬幹、苦幹、實……負責、盡責、……

公共租界日捕加薪

與西捕同待遇

二十日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕同等、按：……房之日……、原爲……元、但……當升級……

滑稽小說 一重……

陰兵們到了……

窮途一個年紀大一點的，陪都已經弄得很亂了。他們的一來，吃不慣天上的東西，放些毒藥在內，所以無論什麼帶，連馬的馬料，也不敢吃天，到了天上，燃料就無處投，是他們就把人家的桌子椅子中什物，有限得很，他們人還被有許多老櫃，就叫我們……

淪……

消息……收」……聞此……浦江……政府……、以……

中……

各路華軍大舉反攻

法租界懸旗糾紛

交涉……

府交涉外……決辦法、……體擴大、……全部復業……於昨晨十……、除陳捷……有被捕……有沒收……三）此後……、總動……[illegible]

……人、准……納稅會備函領回……、至要求市民自由懸旗勿加干涉一節、因不在捕房權力範圍之內、應直接向法總領事提出請求

釋放經過

市民之被捕者、共達四十餘人、受傷者因散居各處、一時無從統計、惟據捕房消息、曾有越捕及華捕各一人受傷、被捕華人、經納稅會交涉後、已於昨日下午四時許、陸續釋放、聞在捕房拘押時、多數均忍痛絕食、愛國熱情、充分流露、惟被捕華人中有二名、捕房以其呼口號寫標語、初不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致意慰問、一面勸其先行復業、各銀行感以公會顧慮之勸慰、亦即照常營業、又實金大戲院執照被吊銷後、聞亦發還、定卽日起照常營業云

法報記載

廿日法文日報載、據法方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、顯有政治性質之表示、故當局乃飭令各居民各自收旗、當時並無反抗情事、關於此點、法

……午十二時十五分抵港、由港至河、每星期六上午六時三十分起飛、當日上午九時四十五分抵河、（均係當地時間）、與歐亞航空公司之滇河線銜接、由港抵滇、當日可到、由滇至港、隔日可到、而中國航空公司之渝河線、又已通航、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐……報載、比國駐滬總領事古特生夫婦、以離滬在卽、特於昨日（十九）、在其署內、舉行茶會、招待本埠之比僑、

本市精神總動員會　昨奉張群來電嘉勉

本市各界國民精神總動員協會、於本年三月二十二日、正式成立後、卽電呈全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次（上略）、國民精神總動員、確屬抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅孜孜、策動愛國之精紳、爲抗戰之聲援、尤屬必要、貴會領導群衆、組織協會、並利用各界固有之職業團體、普設分途推進之持久機關、深謀碩劃、殊堪嘉許、應准備案、仍希繼續努力、奉行不懈、是所原望、張羣叩巧印、

法文上海日……比國駐滬總……特生夫婦、……在卽、特於……十九）、在……舉行茶會……本埠之比僑……本市各界國民精神總動員協會、於本……三月二十二日、正式成立後、卽電呈……全國國民精神總動員會蔣會長、報告成……及工作狀況、頃該會接奉全國……國民精神總動員會張秘書長……滬市工廠林立、……戰後人口突增、工……不堪勒索……瞿順昌喪心……與日人朋……

不堪勒索停業

瞿順昌喪心病狂　與日人朋比爲奸

滬市工廠林立、戰後人口突增、工業更見……、而工業上必需之農產品原料、各業自內地或游擊區域、致一般市價……、亦工業……稻草一項……居奇、故……上之不……少者、運滬頗多、近忽有前邑廟柴業公會董事瞿順昌（一名瞿洽成前在南碼頭開設瞿義興柴行）與日人朋比操縱、由所謂小川洋行者、在南碼頭設立柴草薪炭部、假評價之名、勒索稻草船繳費十元、船戶等以既被「內河民船公會」征苴大船捐於……立分支行、逐漸推行至新……、以期增加後方生產、充實抗戰力量云、

……前今、又有形同稅卡之機關出現於後、致無以應、相率停業、致稻草來源驟減、各工廠受缺貨影響、對勤索爾規深表憤慨、同時草船戶亦連絡同業、一致反對云

公共租界　日捕加薪

與西捕同待遇二十日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕……

應有懸國旗之自由

謝晉元團長昨發告同胞書

四行孤軍團長謝晉元、昨發出爲取締懸掛國旗敬告同胞書、大要如下、

（一）昨日爲上海市農工商各界響應國民政府精神總動員運動、租界市民、一致懸旗紀念、愛受法租界當局取締、發生糾紛、國旗爲代表一國家之尊嚴、任何國民、均有懸掛其國旗之自由與權利、同胞乎、君等英勇果敢、愛護國家、擁護最高領袖之行爲、世人同深欽仰、余前次發表「一年來觀感及對時局感想」一文、曾有「今後國人唯有本「國家興亡、匹夫有責」之義、各守崗位、竭盡天職、以最高領袖之主張爲主意志、而腳踏實……國、孤島之同胞……年開始、願人人……年半以來所盡責……深望孤島同胞……效祖國之責任、……有力出力、以完……之基礎…、余今……向孤島同胞呼籲……所受之苦痛、……之黑暗、一切惡……華民族偉大光榮……爲、切望孤島同胞……有責、出錢出力……、硬幹、苦幹、……負責、盡責、……

任何國國民

滑稽小說　一重……

窮途一個年紀大一點的、陰……都已經弄得很亂了、他們的……一來、吃不飽、穿上的東西、放些毒藥在內、所以無論什……帶、連馬的馬料、也不敢吃……是他們到了天上、人家的桌子椅子……中什物、……還有許多……[illegible]

各路華軍大舉反攻

本報……紙鼻祖銷路普……

本報具有廿一年的歷史是小型報紙鼻祖銷路普

法租界懸旗糾紛

各商店今晨開業

被……狂爲奸

上海市商會及法……、分別發出緊急通告……懸旗糾紛、係由……各商店、即……納稅會……府交涉外、應請各商店、即……一、（二）納稅會……決辦法、此爲顧念中法邦交……腰擴大、而利進行、……全部復業云、

交涉經過

法租界納稅……昨推代表……涉干涉懸旗……於昨晨十時、往晤法總巡捕……、除陳述商民意見外、並要求……有被捕諸人、應請即予釋放……有沒收國旗、應由捕房送回各……三）此後凡遇紀念節日、仍……、總副總監當表示、待將各……轉陳總監後、約於下午三時……各代表再往、當商得辦法（……人、准即釋放、（二）在捕……納稅會備函領回、（三）至……由懸旗勿加干涉、因不……範圍之內、應直接向法總領……

釋放經過

市民之被……四十餘人……散居各處……統計、惟據捕房消息、曾有……各一人受傷、被捕諸人、經……後、已於昨日下午四時許、陸續釋放、閉在捕房拘押時、多數均忍痛絕食、愛國熱情、充分流露、惟被捕諸人中有二名、捕房以其呼口號寫標語、初不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致慰問、一面勸其先行復業、各銀行感以公會懇摯之勸慰、亦即照常營業、又黃金大戲院執照被吊銷後、聞亦發還、定即日起照常營業云

法報記載

廿日法文日報載、據法方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、顯有政治性質之表示、故當局乃飭令各居民各自收旗、當時並無反抗情事、關於此點、法……

……公司之滇河線復……由港抵滇、當日可到、由滇至港、隔日可到、而中國航空公司之渝河線、又已通航、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐……

本市精神總動員會

昨奉張群來電嘉勉

……年三月二十二日、正式成立後、即……全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次（上略）、國民精神總動員、確屬抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅孜甚、策動愛國之精神、爲抗戰之聲援、尤屬必要、貴會領導羣衆、組織協會、並利用各界固有之職業團體、普設分途推進之持久機關、深謀碩劃、殊堪嘉許、應准備案、仍希繼續努力、奉行不懈、是所原望、張羣叩巧印、

……般市儈……稻草一……上之不……運滬頗……前邑廟……事報順……洽成有……設稻義興柴行）與日人朋比操縱、由所謂小川洋行者、在南碼頭設立柴草薪炭部、假評價之名、勒繳每草船繳費十元、船戶等以既被「內河民船公會」征苴大船捐於……

……物、而工……之農產品……來自內地……域、致一……斷居奇、……、亦工業……少者、故……、近忽有……業公會董……（一名狸……南碼頭開……

……前今、又有形同稅卡之機關出現於……後、致無歎以應、……相率停業、致稻草……受缺貨影響、對勒……索酾規深表憤慨、同時草船戶亦連絡同業、一致反對云

公共租界日捕加薪

與西捕同待遇二十日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕……

滑稽小說　一重

窮途一個年紀大一點的，陪都已經弄得很亂了。他們的一來，吃不慣天上的東西，放些毒藥在內，所以無論什帶，連馬的馬料，也不敢吃天到了天上，燃料也就無法是他們就把人家的桌子椅子中什物，有硬得很，他們人還沒有許多老樹，就叫我們……

陰兵們到了……

……消息……收」……聞此……浦江……收府……、以……

應有懸國旗之自……

謝晉元團長昨發告同胞……

……長謝晉元、昨……掛國旗報告同……下、上海市農工商各界國民政府精……員運動、租界……旗紀念、愛……綿、發生糾紛……一國家之尊嚴……利、同胞乎、……爲有懸掛其國……、愛護國家、……之行爲、世人……前次發表「一……時局感想」文……人唯有本「國……有責之義、各……天職、以最高……領袖之主張爲主……意志、而腳踏實地……國、孤島之同胞……年開始、願人人……年半以來所盡責……深望孤島同胞、……效祖國之實任、……有力出力、以完……之基礎…、余今……向孤島同胞呼籲……所受之苦痛、……之黑暗、一切惡……華民族偉大光榮……爲、切望孤島同……有責、出錢出力……、硬幹、苦幹、……負責、盡責、……

各路華軍大舉反攻

業、狂爲奸、物、而工、之農產品、來自內地、域、致一、斷居奇、、亦工業、少者、故、、近忽有、業公會董、（一名狸、南碼頭開
受缺貨影響、對勒索酾規深表憤慨、同時草船戶亦連絡同業、一致反對云
公共租界日捕加薪
與西捕同待遇二十日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕
滑稽小說　一重

本報具有廿一年的歷史是小型報紙鼻祖銷路普

法租界懸旗糾紛

各商店遵勸 今日可復業

被捕市民昨日均已釋放

關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、爲勸導法租界商店復業、於昨晨、分別發出緊急通告云、（一）市商會、「中法邦交素稱敦睦、此次懸旗糾紛、係由越捕執行職務、過分所致、各界引爲遺憾、除電請政府交涉外、應請各商店、即先復業、以堅毅貞固之態度、靜候合法合理之解決、……」（二）納稅會、「查關於懸旗事件、業經本會推員向法當局磋商解決辦法、此爲……擴大、而利……全部復業云」

任何國國民 應有懸國旗之自由

謝晉元團長昨發告同胞書

四行孤軍團長謝晉元、昨發出爲收縮懸掛國旗敬告同胞書、大要如下、昨日爲上海市農工商各界響應國民政府精……

各路華軍大舉反攻

本報忠實報道新聞說人人要說的話文字淺近趣味濃厚是最平民化的刊物

晶報

The Crystal

發行人 A. L. Teodoro

館址漢口路二九九號 電話九三四二八

中華郵政特准掛號立券按照總包優待寄遞之報紙

公共租界工部局登記證C字廿四號

第三千七百九十五號

中華民國貳拾捌年四月廿一日 星期五 今日貳張售國幣貳分四厘 本報創於民國八年三月三日

各路華軍大舉反攻

克復石龍包圍南昌

贛南高安發生激烈爭奪戰

綏晉重要據點多處已收復

（重慶廿日海通電）據此間所接消息、中國軍隊、反攻江西省之高安、異常奏效、已經發生激烈之爭奪戰、此外日軍最近佔領之南昌、現亦在華軍之完全包圍中、

（興集廿日電）綏西華軍日前分路出擊、向包圍積極挺進、

（翁源廿日電）十九日晨、華軍反攻新會城、與日軍約千人、在城外高地激戰、日軍狼狽退入城、華軍追擊、衝入城內與日軍巷戰、

（韶關廿日電）新會方面十九日晨已被華軍完全克復、另一部並已逼近江門、

國際時事 地圖學

地中海的海軍根據地

抱定抗戰決心

中國現決不談和

英當局亦聲明未提議調解事

（重慶廿日海通電）此間中國報章、對於歐局、現均極表重視、尤其隨重歐洲局勢、對於中國之影響、一般評論之中心、咸申明中國決不能因世界其他部份之事件、而鬆懈其抗戰之決心、抱定此種信念、中國決不顧外國方面之任何建議、除非此項建議、係符合中國人民之要求及需要、此般輿論、咸認爲蔣委員長前日之演說、亦係申明中國現時決不談和、

（倫敦廿日路透電）英當局茲鄭重聲明、日本報紙所載、英國已擬供調解中日戰爭之說、絕對不確、

蘇北戰事展開

宿遷沭灘各縣

華軍甚得手

戰局已轉入主動

（香港廿日電）

開封城緊閉

日僞鈔大跌

（洛陽廿日電）

國府明令

褒揚吳梅

桂施政計劃

（桂林廿日電）

牯嶺未失陷

（重慶廿日路透電）

綏包車不通

（五原十九日電）

曾撲皖邊區

日軍受襲擊

德羅談話

向臨汾推進

本報復刊小啓

本報自四月一日起，由美商 A.L. Teodoro 君爲發行人，徹底改組，用於今日復刊。今後本報有二大趨向：促進東西兩大民主國親密攜手，謀太平洋上之集體安全，以求民治思想徹底實現，此其一。造成健全正大之輿論，取公開態度，爲大衆喉舌，凡有專著，小品，新聞，趣事，靡不虛心採納，而於解除民生疾苦，增進人羣福祉，尤致意焉，此其二。謹布區區，諸希明鑒。

法租界懸旗糾紛

各商店遵勸

今日可復業

被捕市民昨日均已釋放

關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、爲勸導法租界商店復業、於昨晨、分別發出緊急通告云、

交涉經過

釋放經過

法報記載

河內

航郵已通

外交界動態

本市精神總動員會

昨奉張群來電嘉勉

稻草船戶

不堪勒索停業

瞿顧昌喪心病狂

與日人朋比爲奸

西北金融網

已完成

任何國國民

應有懸國旗之自由

謝晉元團長昨發告同胞書

公共租界

日捕加薪

淪陷區公產

僞「維新政府」有接收說

江海關對外貿易

第一季統計

滑稽小說

一重天

陰兵們到了一重天

（一四九）

本報具有廿一年的歷史是小型報紙鼻祖銷路普遍全國刊登廣告最有效力

Document layout recognition and page segmentation:

Page structuring elements – registers

晶報
The Crystal

發行人 A. L. Teodoro
館址漢口路二九九號 電話九三四二八
中華郵政特准掛號立券按照總包優益寄遞之報紙
◀公共租界工部局登記證C字廿四號▶
◀第三千七百九十五號▶

中華民國貳拾捌年四月廿一日 星期五 今日貳張售國幣貳分四厘 本報創於民國八年三月三日

法租界懸旗糾紛
各商店遵勸今日可復業
被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

上海市商會及法租界納稅華人會、於昨晨分別發出緊急通告云、（一）市商會「中法邦交素稱敦睦、此次懸旗糾紛、係由越捕執行職務、過分所致、各界引爲遺憾、除電請政府交涉外、應請各商店、即先復業、以堅毅貞固之態度、靜候合法合理之解決、（二）納稅會、『查關於懸旗事件、業經本會推員向法當局磋商辦法、此爲顧念中法邦交起見、應請各商店共體時艱、即希照常營業、以免事態擴大、而利進行、……」、法租界商民、經此兩會之勸告後、今晨可望全部復業云、

交涉經過

法租界納稅華人會、昨推代表瞿振華等交涉干涉懸旗事件、先於昨晨十時、往晤法總巡捕房警務總監、除陳述商民意見外、並要求（一）所有被捕諸人、應請即予釋放、（二）所有沒收國旗、應由捕房送回各商店、（三）此後凡遇紀念節日、仍當自由懸旗、總監當表示、待將各代表意見、轉陳總領事後、約於下午三時答復、屆時各代表再往、當局得辦法（一）被捕諸人、准即釋放、（二）在捕房國旗、由納稅會備函領回、（三）至要求市民自由懸旗勿加干涉一節、因不在捕房權力範圍之內、應直接向法總領事提出請求、則不在禁例云、

釋放經過

市民之被捕者、共達四十餘人、受傷者因散居各處、一時無從統計、惟據捕房消息、曾有越捕及華捕各一人受傷、被捕諸人、經納稅會交涉後、已於昨日下午四時許、陸續釋放、愛國熱情、充分流露、惟被捕諸人中有二名、捕房以其呼口號轉嚷標語、勒不允釋放、經再三交涉、亦允釋放、上海市銀行業同業公會、對於會員銀行之遭受懸旗干涉、一面致意慰問、一面勸其先行復業、各銀行感以公會懇摯之勸慰、亦即照常營業、又黃金大戲院執照被吊銷後、聞亦發還、定即日起照常營業云、

法報記載

廿日法文日報載、據法方消息、昨日法租界內、並無重大事件發生、外間之傳說、殊不足信、聞在昨日因大規模懸掛國旗事件、顯有政治性質之表示、故當局乃飭令各商民各自收旗、當時並無反抗情事、關於此點、法捕房特發有命令惟遇有國慶日而懸旗者、則不在禁例云、

河內航郵已通

法國航空公司之河內至香港航空線、現每星期往返飛行各一次、由河至港、每星期四上午七時起飛、當日下午十二時十五分抵港、由港至河、每星期六上午六時三十分起飛、當日上午九時四十五分抵河、（均係當地時間）、與歐亞航空公司之滇河線銜接、由港抵滇、當日可到、由漢至港、隔日可到、兩中國航空公司之渝河線、亦可與河內香港線聯運、所有我國往來國內暨國外各地經由香港、及河內轉遞之航空郵件、均可通達、

外交界動態

法租界環龍路法國總會董事團及全體會員、以法國駐華新大使戈斯默來滬已近彼旬、對於法國在華情形有所陳述、爰定於明日下午六時三十分至七時三十分、假該會禮廳、舉行公宴、

廿日法文上海日報載、比國駐滬總領事古特生夫婦、以離滬在即、特於昨日（十九）、在其署內、舉行茶會、招待本埠之比僑、法僑、以表告別云、

本市精神總動員會 昨奉張群來電嘉勉

本市各界國民精神總動員協會、於本年三月二十二日、正式成立後、即電呈全國國民精神總動員會蔣會長、報告成立經過、及工作狀況、頃該會接奉全國國民精神總動員會張秘書長岳軍復電、慰勉有加、茲照錄如次（上略）、國民精神總動員、確屬抗戰建國之重要方略、滬市淪陷已久、敵寇誘脅益甚、策動愛國之精神、爲抗戰之後援、尤屬必要、貴會領導羣衆、組織協會、並利用各界固有之職業團體、普設分途推進之持久機關、深謀碩劃、殊堪嘉許、應准備案、仍希繼續努力、來行不懈、是所厚望、張群叩巧印、

西北金融網已完成

中國銀行、爲調劑西北金融、以應抗戰需要、決定以蘭州爲中心、先籌設蘭州分行、特派鄭恩帥爲經理、已抵蘭籌備一切、據談際此非常時期、中央爲完成西北金融網之計劃、特命在蘭州籌設本行、按照中央發展西北計劃、第一除以蘭州爲中心、先行設立外、並在西寧、寧夏、天水、五原等處設立分支行、第二俟以上五處支行完全成立後、繼續在平涼、武威、張掖、酒泉四處設立分支行、逐漸推行至新疆、以期增加後方生產、充實抗戰力量云、

稻草船戶不堪勒索停業
瞿順昌喪心病狂 與日人朋比爲奸

滬市工廠林立、戰後人口突增、工業上必需之農產品原料、多來自內地或游擊區域、致一般市儈藉斷居奇、稻草一項、亦工業上之不可少者、故近忽有瀝瀝頗多、前邑廟柴業公會董事瞿順昌（一名瞿洽成前在南碼頭開設瞿義興柴行）與日人朋比操縱、由所謂小川洋行者、在南碼頭設立柴草薪炭部、假評價之名、勒索每草船繳費十元、船戶等以既被「內河民船公會」征苴大船捐於前今、又有形同稅卡之機關出現於後、致無款以應、相率停業、致稻草來源驟減、各工廠受缺貨影響、對勒索陋規深表憤慨、同時草船戶亦連絡同業、一致反對云、

任何國民應有懸國旗之自由
謝晉元團長昨發告同胞書

四行孤軍團長謝晉元、昨發出爲取締懸掛國旗敬告同胞書、大要如下、昨日爲上海市農工商界響應國民政府精神總動員運動、租界市民、一致懸旗紀念、發生糾紛、法租界當局取締、爰受辱國旗爲代表一國家之尊嚴、任何國民、均有懸掛其國旗之自由與權利、同胞乎、君等英勇果敢、愛護國家、擁護最高領袖之行爲、世人同深欽仰、余前次發表「一年來觀感及對時局感想」文中曾言…今後國人唯有本「國家興亡、匹夫有責」之義、各守崗位、竭盡天職、以最高領袖之主張爲主張、意志爲意志、而腳踏實地、努力報國、孤島之同胞乎、際茲二年開始、願人人檢討抗戰一年半以來所盡責任若何…深望孤島同胞、人人加緊報效祖國之責任、有錢出錢、有力出力、以完成全國動員之基礎…、余今謹以至誠復向孤島同胞呼籲、吾人今日所受之苦痛、以黎明前間之黑暗、一切惡氛、當隨中華民族偉大光榮之勝利品掃蕩、切望孤島同胞、本興亡有責、出錢出力之義、快幹、硬幹、苦幹、實幹之精神負責、盡責、望共勉之、

公共租界日捕加薪

日法文上海日報載、公共租界巡捕房之日籍巡捕、將增加薪水百分之三十五、而與西捕同等待遇、按：公共租界捕房之日籍巡捕待遇、原爲月薪一百廿元、但自下月起、當升級與第二等之西捕、同樣待遇、即每月可領薪水至一百七十元、又聞日捕此後之進級、亦與西捕全同、并可升級爲第一等捕頭云、

淪陷區公產 傳「維新政府」有接收說

廿日法文上海日報載、據日方消息、南京「維新政府」、擬「接收」一切前屬於國民政府之公產、聞此產業之標的、係爲揚子江及黃浦江沿岸一帶之棧房、又聞「維新政府」、現擬在各江岸、建築碼頭、以利「航運」云、

江海關對外貿易第一季統計

自中國政府、管理對外貿易、促進輸出消除入超後、西南各關業務發達、稅收激增、惟上海之江海關、在特殊情形之下、稅收被截留、主權被破壞、入超數字、有增無減、今年第一季貿易輸入一萬二千六百四十二萬零五百四十三元、輸出八千零六十五萬二千五百五十一元、入超四千五百七十六萬七千九百九十二元、

滑稽小說 一重天（九四一）
陰兵們到了一重天

窮途一個年紀大一點的、陪笑說：陰兵們到了一重天把什麼都已經弄得很亂了。他們的食糧，是他們從地獄裏帶來的。一來，吃不慣天上的東西，二來，怕天上的食物，不要被人放些毒藥在內，所以無論什麼吃的東西，完全從鬼國中搬來帶到了天上，連馬的馬料，也不敢吃天上之物，全是帶來的；但是食物是他們就把人家的桌子椅子等東西，一齊拿來燒了；但是家中什物，有限得很，他們人又多，因此已覺不夠，所以看見還有許多老樹，就叫我們來砍掉，借他們去炊飯了。

國際時事地圖

各路華軍大舉反攻
克復石龍包圍南昌
贛南高安發生激烈爭奪戰
綏晉重要據點多處已收復

國府明令褒揚吳梅

（重慶廿日電）國府廿日令、褒揚故中央大學教授吳梅、並特給郵金三千元、

開封城緊閉

牯嶺未失陷

（重慶廿日路透電）據華方消息、牯嶺仍在華軍手中、日軍以四縱隊、於四月十七日沿九江蓮花洞公路、在猛烈炮火掩護之下進攻、該處戰事刻在進行中、

桂施政計劃

（桂林廿日電）桂省府頃制定廿八年度施政計劃綱要、其特別着重者、爲（一）多方促進……

蘇北戰事開展
宿遷沭灌各縣華軍甚得手
戰局已轉入主動

抱定抗戰決心 中國現決不談和
英當局亦聲明未提議調解事

綏包車不通

魯豫皖邊區日軍受襲擊

向臨汾推進

德羅談話

本報復刊小啓

本報自四月二十一日起，由美商A.L.Teodoro君爲發行人，徹底改組，用於今日復刊。今後本報有二大軒向：促進東西兩大民主國親密攜手，謀太平洋上之集體安全，以求民治思想益臻實現，此其一。造成健全正大之輿論，取公開態度，爲大衆喉舌，凡有專著，小品，新聞，趣事，靡不虛心採納，而於解除民生疾苦，增進人羣福祉，尤致意焉，此其二。謹布區區，諸希明鑒。

晶報
The Crystal

發行人 A. L. Teodoro
館址漢口路二九九號　電話九三四二八
中華郵政特准掛號立劵按照總包優益寄遞之報紙
◀公共租界工部局登記證C字廿四號▶
◀第三千七百九十五號▶

中華民國貳拾捌年四月廿一日　星期五　今日貳張售國幣貳分四厘　本報創於民國八年三月三日

各路華軍大舉反攻
克復石龍包圍南昌
贛南高安發生激烈爭奪戰
綏寧重要據點多處已收復

5 reg - (44 char)

4 reg - 35 char

國際時事地圖

1 reg - 8 char

0.5 reg - 4 char

中國現抗戰決心
抱定抗戰決心不談和

3 reg - 26 char

華軍昔得手

1 reg - 8 char

1 reg - 8 char

本報復刊小啓

本報自四月廿一日起，由美商A.L.Teodoro君為發行人，徹底改組，用於今日復刊。今後本報有二大嚮向：促進東西兩大民主國親密攜手，謀太平洋上之集體安全，以求民治思想益臻實現，此其一。造成健全正大之輿論，取公開態度，為大衆喉舌，凡有專著，小品，新聞，趣事，靡不虛心採納，而於解除民生疾苦，增進人羣福祉，尤致意焉，此其二。謹布區區，諸希明鑒。

-7995-

法租界懸旗糾紛
各商店遵勸今日可復業
被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

4 reg - 35 char

交涉經過

2 reg - 17 char

河內

1 reg - 8 char

2 reg - 17 char

已完成

1 reg - 8 char

不堪勒索停業

1.5 reg - 12 char

3 reg - (26 char)

1.5 reg - 12 char

滑稽小說 一重天

3 reg - 26 char

淪陷區公產

本報忠實報道新聞說人人要說的話文字注意趣味濃厚是最平民化的刊物

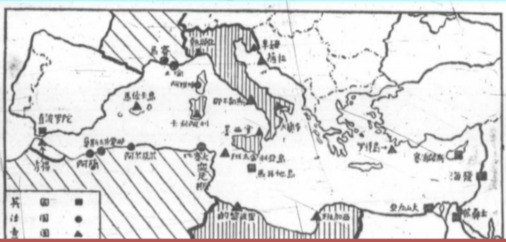
晶報 The Crystal

發行人 A. L. Teodoro

◀第三千七百九十五號▶

中華民國貳拾捌年四月廿一日 星期五 今日貳張售國幣貳分四厘 本報創刊於民國八年三月三日

1 reg - 8 char



0.5 reg - 4 char

4 reg - 35 char

各路華軍大舉反攻 克復石能包圍南昌

5 reg - (44 char)

3 reg - 26 char

中國現決不談和

1 reg - 8 char

1 reg - 8 char

本報復刊小啓

法租界懸旗糾紛 今日可復業

4 reg - 35 char

2 reg - 17 char

1 reg - 8 char

2 reg - 17 char

1 reg - 8 char

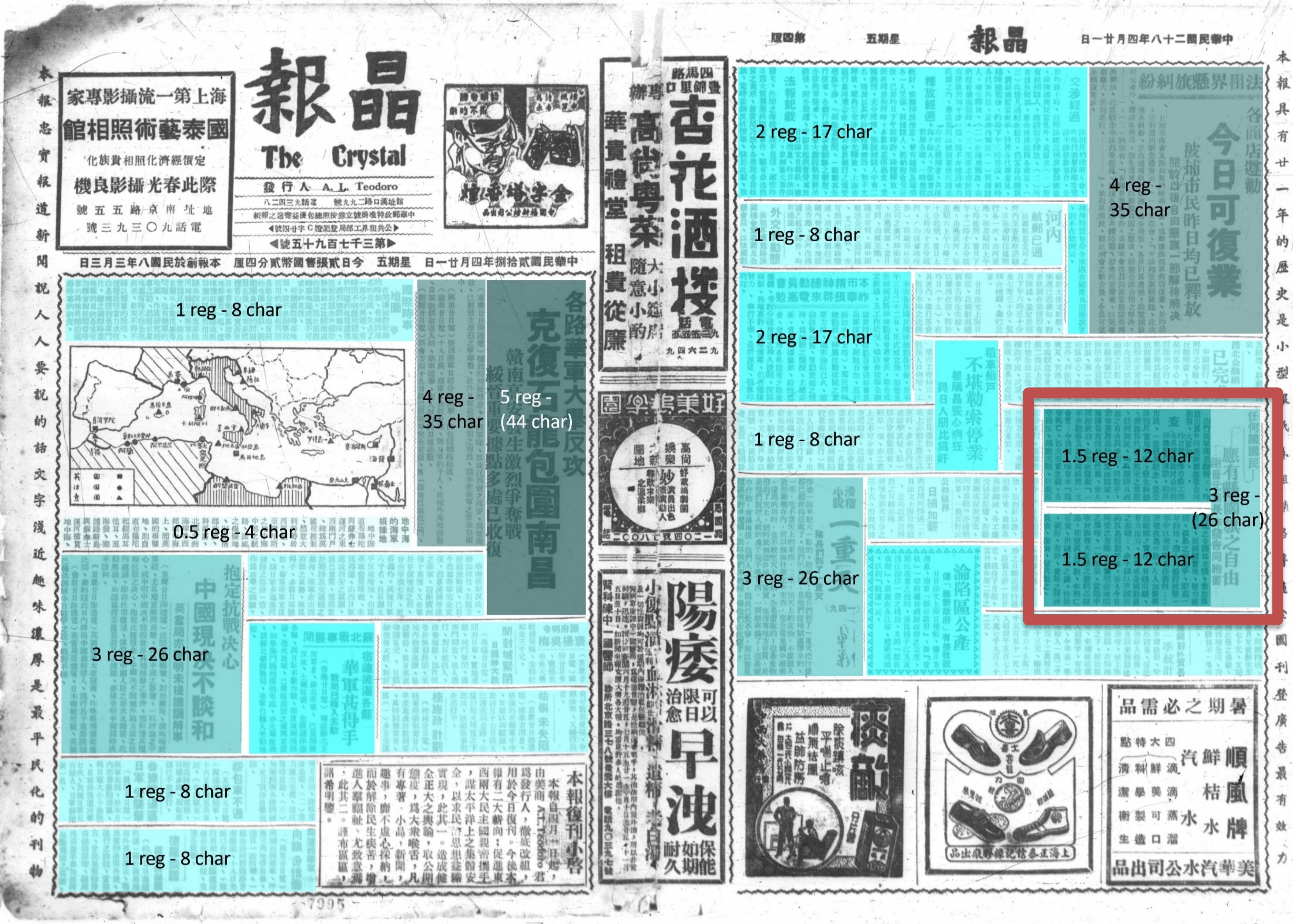
1.5 reg - 12 char

3 reg - (26 char)

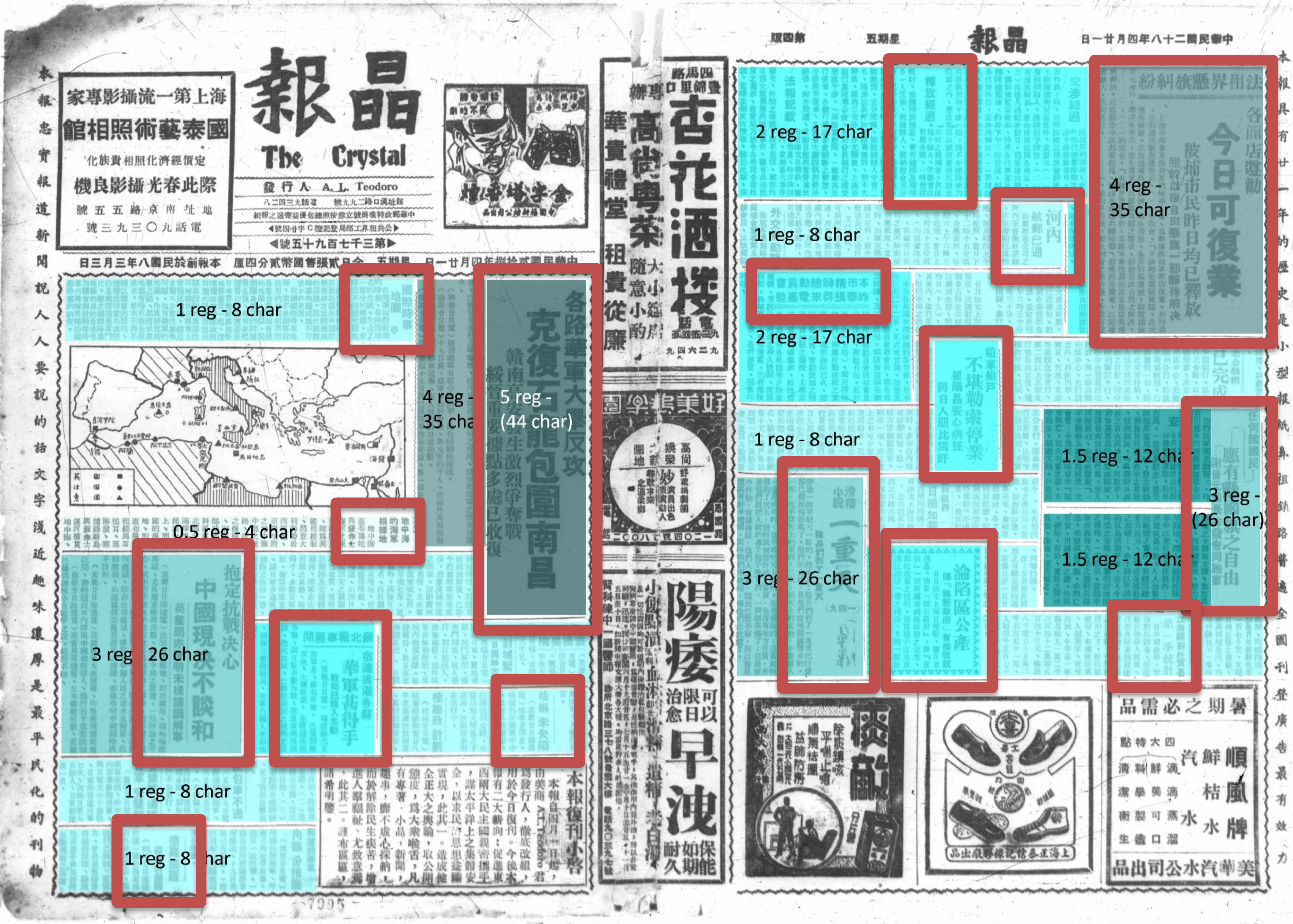
1.5 reg - 12 char

3 reg - 26 char

本報具有廿一年的歷史是小型報紙鼻祖銷路普遍全國刊登廣告最有效力



晶報
The Crystal
發行人 A. L. Teodoro
海上第一流攝影專家
國泰藝術照相館
香花酒樓
高尚粵菜
各路華軍大舉反攻
克復石能包圍南昌
中國現抗戰決心不談和
今日可復業
陽痿
早洩
暑期之必需品
順風牌
美華汽水公司出品
本報復刊小啓
1 reg - 8 char
4 reg - 35 char
5 reg - (44 char)
0.5 reg - 4 char
3 reg - 26 char
1 reg - 8 char
1 reg - 8 char
2 reg - 17 char
4 reg - 35 char
1 reg - 8 char
2 reg - 17 char
1 reg - 8 char
1.5 reg - 12 char
3 reg - (26 char)
1.5 reg - 12 char
3 reg - 26 char

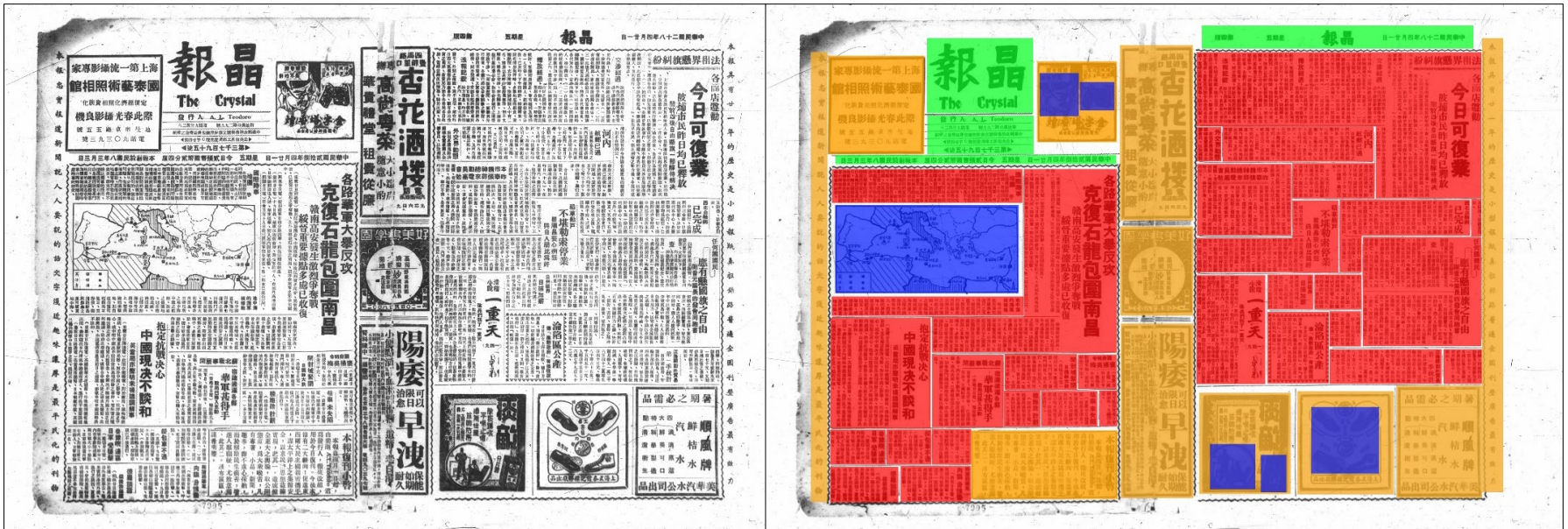

晶報
The Crystal
發行人 A. L. Teodoro
各路華軍大舉反攻
克復石龍包圍南昌
杏花酒樓
陽痿早洩
今日可復業
1 reg - 8 char
4 reg - 35 cha
5 reg - (44 char)
0.5 reg - 4 char
3 reg 26 char
1 reg - 8 char
1 reg - 8 har
2 reg - 17 char
1 reg - 8 char
2 reg - 17 char
1 reg - 8 char
3 reg - 26 char
4 reg - 35 char
1.5 reg - 12 cha
3 reg - (26 char)
1.5 reg - 12 cha

Document layout recognition and page segmentation:

Crowdsourcing

Crowdsourcing segmentation

- Pilot project with Pallas Ludens GmbH
- Let the (controlled) crowd help analyzing the pages
  - Identify and label four item types:
    - image/drawing
    - article
    - advertisement
    - additional information
  - Closely monitored and supervised
  - Crowd members NOT able to speak/read Chinese



●中華郵政特准掛號認為新聞紙類●中華郵政特准掛號立券之報紙●已在中華郵政總局掛號批准包特別優益寄送之報紙

晶報

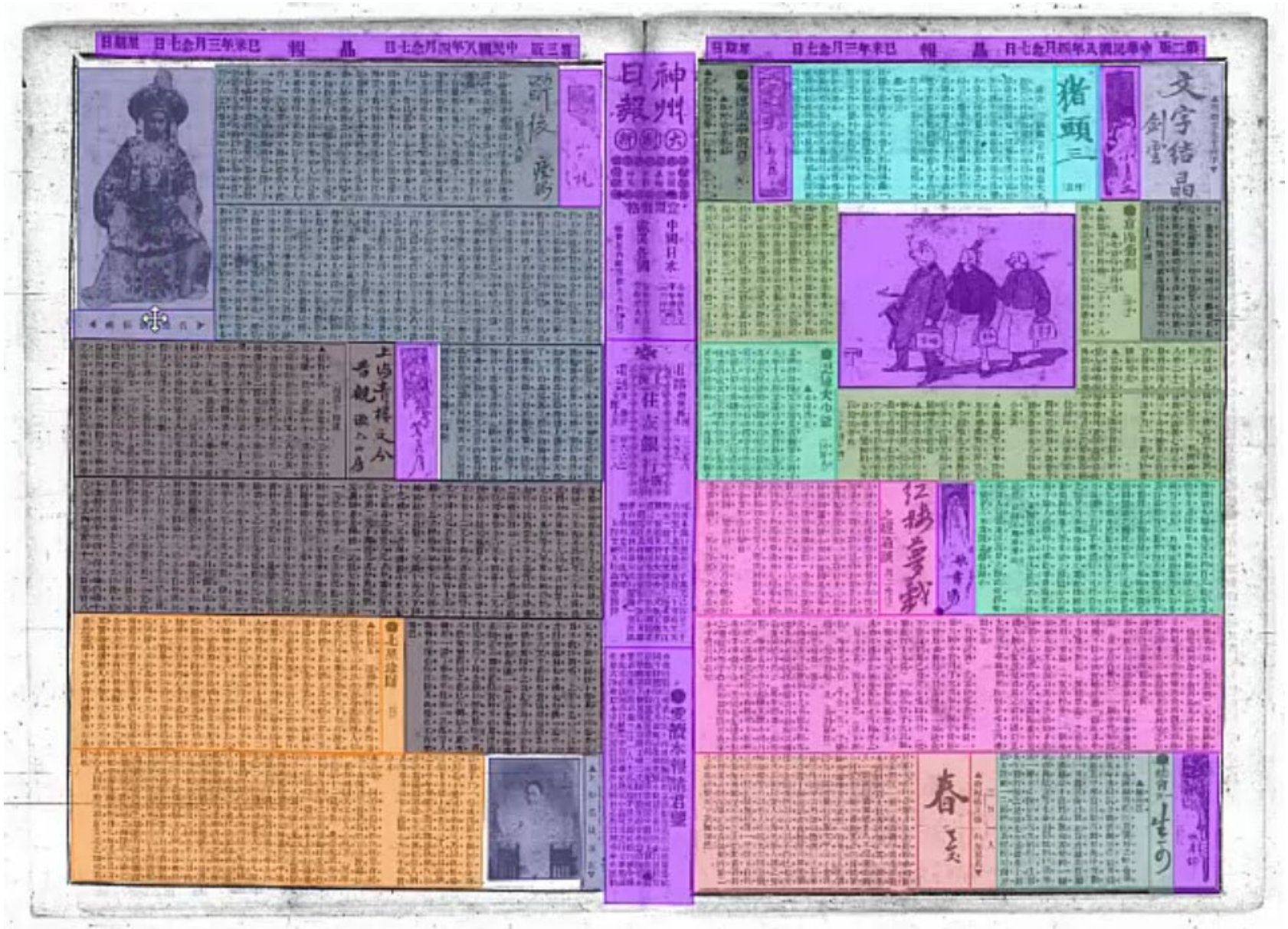
第二百十八號
今日一張　售大洋二分
定報價格
廣告例刊

寒雲鬻書例

寒雲主人好古知書深得三代漢魏之神髓愈窮而書愈工泛遊江海求書者不暇應爰擬定書例如下
楹聯　行楷每字一元
篆隸楷每字二十九　隸每字加一元
代收件處

■游戲新報第一期現已出版

Grouping semantic units



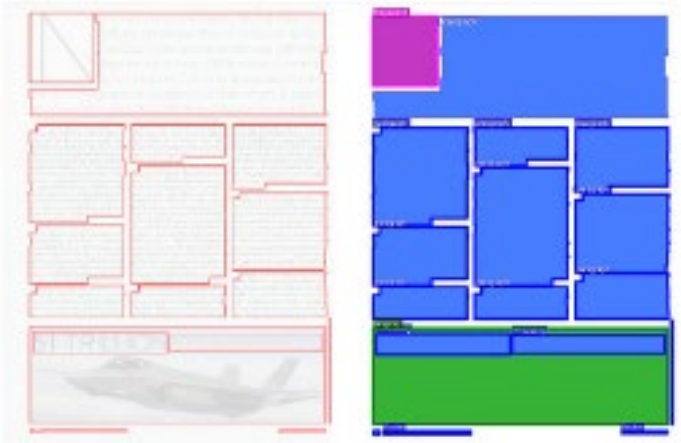
Outcome of segmentation pilot

1. Page segmentation can be outsourced to expert crowd
    - Requires supervision
    - Advanced user interfaces (high usability, efficiency)
    - For more advanced tasks: Crowd should read Chinese (semantic grouping)
2. *Jingbao* 晶報 1919-21 (927 folds) completely segmented with qualified boxes, issues of April 1919 with semantic units
3. Further processing:
    - Computational Knowledge Lab (知識計算實驗室), Dept. Engineering Science and Ocean Engineering, Taiwan National University, http://www.cklab.org/
    - Advanced Linguistic Technologies, Heidelberg
    - Seeking partners for collaboration!

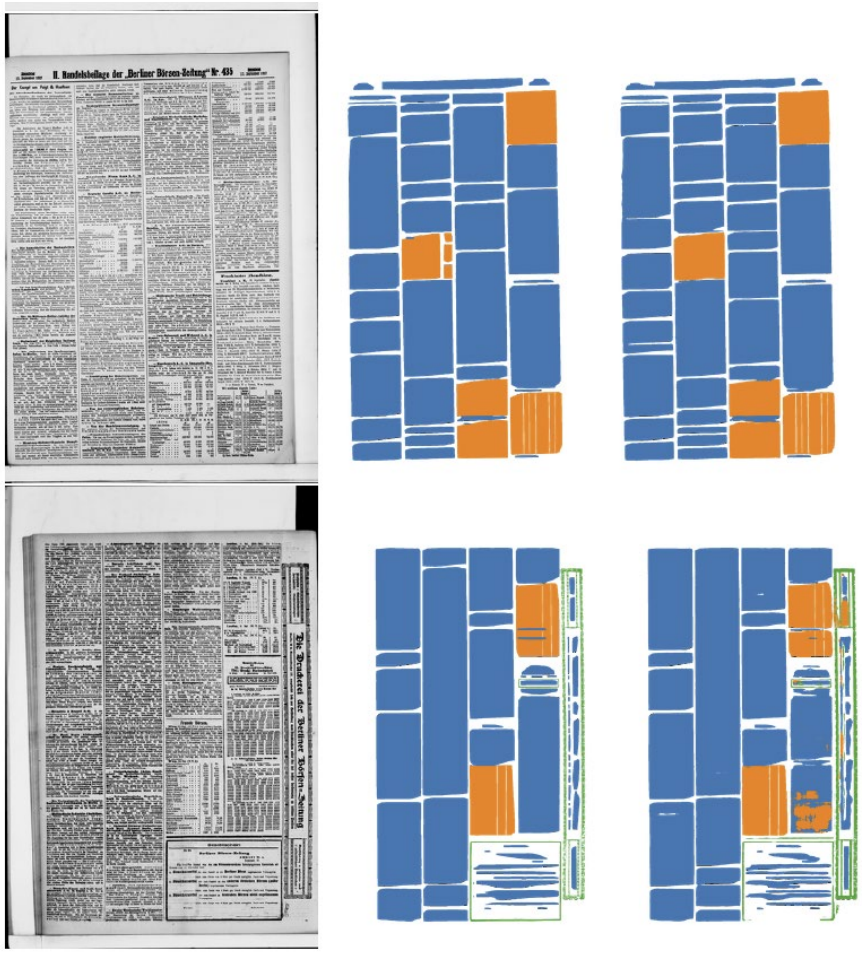
Towards full text - 3:

Machine learning

Neural networks



ICDAR 2019 – RDCL2019 challenge



Liebl and Burghardt 2020

Neural networks

Large research projects on (western, mostly Latin script) newspapers, three examples from Europe:

- Europeana newspapers http://www.europeana-newspapers.eu/
- Impresso https://impresso-project.ch/
- NewsEye https://www.newseye.eu/
- OCR-D https://ocr-d.de/

Re-usable (free, open source, open access) software, for example dhSegment - a “generic deep-learning framework for Historical Document Processing” https://github.com/dhlab-epfl/dhSegment

Our aims:

- Use for segmentation
- Train models to detect main element classes:
  Paratext (headers and marginalia), Advertisement and images, Text areas
- Requirement: Ground Truth

Towards full text - 4:

Ground truth

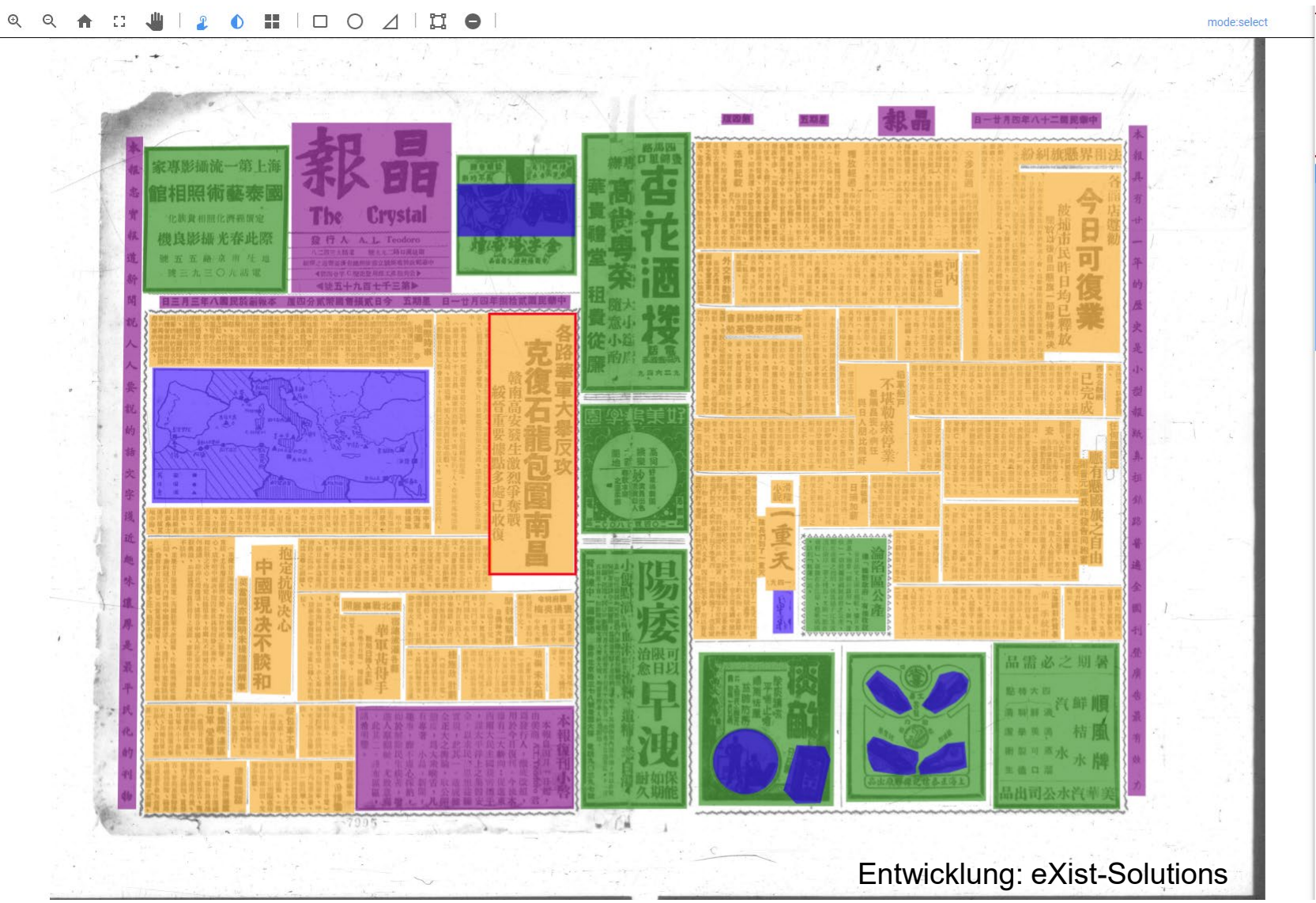
Funding for Ground Truth

Short term funding by Field of Focus 3 (ExStrat Heidelberg)

Small team with Duncan Paterson and 4 RA's (typing)

- GroundTruth 1: Bounding boxes with semantic labels
    - Coordinates in json and web-annotation format
    - Jingbao April 1939 und April 1938
    - Outcome: 70 folds, 6335 shapes, Ø 90,5 shapes/fold

Annotation tool



Annotation output in json

jb_3795_1939-04-21_0002+0003.json


```
{
  "items": [
    {
      "body": [
        {
          "value": {
            "color": "purple",
            "name": "additional",
            "label": "Additional Information"
          },
          "type": "CategoryLabel"
        }
      ],
      "created": "2019-07-10T16:14:26.772Z",
      "target": [
        {
          "type": "SpecificResource",
          "source": "https://kjc-sv002.kjc.uni-heidelberg.de:8080/fcgi-bin/iipsrv.fcgi?IIIF=imageStorage/ecpo_new/jingbao/1939/04/jb_3795_1939-04-21_0002%252B0003.tif/full/full/0/default.jpg",
          "id": "s-156277526676110",
          "selector": {
            "value": "<g transform=\"matrix(1 0 0 1 3462.22464 371.24256)\"><polygon points=\"-812.0176,-74.99264 812.0176,-74.99264 812.0176,74.99264 -812.0176,74.99264 \"/></g>",
            "type": "SvgSelector"
          }
        }
      ],
      "type": "Annotation",
      "id": "s-1562775266771"
    },
    {
      "body": [
        {
          "value": {
            "color": "purple",
            "name": "additional",
            "label": "Additional Information"
          },
          "type": "CategoryLabel"
        }
      ],
      "created": "2019-07-10T16:14:48.631Z",
      "target": [
        {
          "type": "SpecificResource",
          "source": "https://kjc-sv002.kjc.uni-heidelberg.de:8080/fcgi-bin/iipsrv.fcgi?IIIF=imageStorage/ecpo_new/jingbao/1939/04/jb_3795_1939-04-21_0002%252B0003.tif/full/full/0/default.jpg",
          "id": "s-15627752886237",
          "selector": {
            "value": "<g transform=\"matrix(1 0 0 1 4331.59808 1719.37056)\"><polygon points=\"-26.18944,-1352.42848 53.00704,-1352.42848 26.23776,1352.42848 -53.00704,1352.42848 \"/></g>",
            "type": "SvgSelector"
          }
        }
      ],
      "type": "Annotation",
      "id": "s-1562775288631"
```

Funding for Ground Truth - II

Short term funding by Field of Focus 3 (ExStrat Heidelberg)

Small team with Duncan Paterson and 4 RA's (typing)

- GroundTruth 1: Bounding boxes with semantic labels
    - Coordinates in json and web-annotation format
    - Jingbao April 1939 und April 1938
    - Outcome: 70 folds, 6335 shapes, Ø 90,5 shapes/fold
- GroundTruth 2: Texts
    - Blind double-keying
    - Jingbao April 1939 - 10 issues, 40 folds
    - Outcome: ca. 245.000 characters, local XML format, Ø 6100 chars/fold (typing slow ~10h/fold)

GT full text

Local XML schema

- `<fold>` with `@xml:id` and recto/verso
- `<div>` with *page, article, image, advert, other, head* (for running head on top of page), *margin* (for marginalia)
- `<p>` (paragraph)
- `<lb/>` (line break) / `<pb/>` (page break)

Encoding in “UTF-8”:

- Characters according to original, wherever the respective Unicode code point was available
- Illegible: `&gaiji;`
- All characters entered in double-space form

Special feature in East Asian texts: different running directions

- `<div>` with `@mode` and `@dir`
- Default: `mode=“vertical-rl”`

https://github.com/exc-asia-and-europe/ecpo

jb_3795_1939-04-21_0001+0004.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://raw.githubusercontent.com/exc-asia-and-europe/ecpo/master/ecpo.rng" type="application/x
<!DOCTYPE ecpo [
    <!ENTITY gaiji "&#xe111;">
]>
<fold xml:id="jb_3795_1939-04-21_0001+0004">
    <!-- start recto page   -->
    <div type="page" page="recto" n="1" mode="vertical-rl" dir="rtl">
        <div type="head" dir="rtl" mode="horizontal-tb">
            <div>
                晶報<lb/>
            </div>
            <div dir="ltr" mode="horizontal-tb">
                The Crystal<lb/>
                發行人 A. L. Teodoro<lb/>
            </div>
            <div>
                館址漢口路二九九號　電話九三四二八<lb/>
                中華郵政特准掛號立券按照總包優益寄送之報紙<lb/>
                ◀公共租界工部局登記證C字廿四號▶<lb/>
                ◀第三千七百九十五號▶<lb/>
                中華民國貳拾捌年四月廿一日　星期五　今日兩全張售國幣貳分四厘　本報創刊於民國八年三月三日<lb/>
            </div>
        </div>
        <div type="advert">
            <div>
                煙中<lb/>
                鐵軍<lb/>
                清香<lb/>
                雋永<lb/>
            </div>
            <div mode="horizontal-tb" dir="rtl">
                &gaiji;明者&gaiji;<lb/>
                &gaiji;不吃虧<lb/>
            </div>
            <div type="image"></div>
            <div type="image"></div>
```

jb_3795_1939-04-21_0001+0004.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://raw.githubusercontent.com/exc-asia-and-europe/ecpo/master/ecpo.rng" type="application/x
<!DOCTYPE ecpo [
    <!ENTITY gaiji "&#xe111;">
]>
<fold xml:id="jb_3795_1939-04-21_0001+0004">
    <!-- start recto page   -->
    <div type="page" page="recto" n="1" mode="vertical-rl" dir="rtl">
        <div type="head" dir="rtl" mode="horizontal-tb">
            <div>
                晶報<lb/>
            </div>
            <div dir="ltr" mode="horizontal-tb">
                The Crystal<lb/>
                發行人 A. L. Teodoro<lb/>
            </div>
            <div>
                館址漢口路二九九號　電話九三四二八<lb/>
                中華郵政特准掛號立券按照總包優益寄送之報紙<lb/>
                ◀公共租界工部局登記證C字廿四號▶<lb/>
                ◀第三千七百九十五號▶<lb/>
                中華民國貳拾捌年四月廿一日　星期五　今日兩全張售國幣貳分四厘　本報創刊於民國八年三月三日<lb/>
            </div>
        </div>
        <div type="advert">
            <div>
                煙中<lb/>
                鐵軍<lb/>
                清香<lb/>
                雋永<lb/>
            </div>
            <div mode="horizontal-tb" dir="rtl">
                &gaiji;明者&gaiji;<lb/>
                &gaiji;不吃虧<lb/>
            </div>
            <div type="image"></div>
            <div type="image"></div>
```

Default setting per page:
`@mode=„vertical-rl“` and
`@dir=„rtl“` (right-to-left)

jb_3795_1939-04-21_0001+0004.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://raw.githubusercontent.com/exc-asia-and-europe/ecpo/master/ecpo.rng" type="application/x
<!DOCTYPE ecpo [
    <!ENTITY gaiji "&#xe111;">
]>
<fold xml:id="jb_3795_1939-04-21_0001+0004">
    <!-- start recto page   -->
    <div type="page" page="recto" n="1" mode="vertical-rl" dir="rtl">
        <div type="head" dir="rtl" mode="horizontal-tb">
            <div>
                晶報<lb/>
            </div>
            <div dir="ltr" mode="horizontal-tb">
                The Crystal<lb/>
                發行人 A. L. Teodoro<lb/>
            </div>
            <div>
                館址漢口路二九九號　電話九三四二八<lb/>
                中華郵政特准掛號立券按照總包優益寄送之報紙<lb/>
                ◀公共租界工部局登記證C字廿四號▶<lb/>
                ◀第三千七百九十五號▶<lb/>
                中華民國貳拾捌年四月廿一日　星期五　今日兩全張售國幣貳分四厘　本報創刊於民國八年三月三日<lb/>
            </div>
        </div>
        <div type="advert">
            <div>
                煙中<lb/>
                鐵軍<lb/>
                清香<lb/>
                雋永<lb/>
            </div>
            <div mode="horizontal-tb" dir="rtl">
                &gaiji;明者&gaiji;<lb/>
                &gaiji;不吃虧<lb/>
            </div>
            <div type="image"></div>
            <div type="image"></div>
```

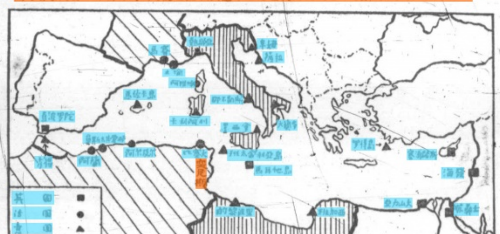
If needed, we encode changes in sub-level `<div>`'s
After `</div>` all changes reset to default

本報忠實報道新聞說人人要說的話文字淺近趣味濃厚是最平民化的刊物

晶報
The Crystal
發行人 A. L. Teodoro
館址漢口路二九九號 電話九三四二八
中華郵政特准掛號立券按照總包價寄遞之報紙
公共租界工部局登記證C字廿四號
第三千七百九十五號
中華民國貳拾捌年四月廿一日 星期五 今日貳張售國幣貳分四厘 本報創於民國八年三月三日

各路華軍大舉反攻
克復石龍包圍南昌
贛南高安發生激烈爭奪戰
綏晉重要據點多處已收復

國際時事地圖

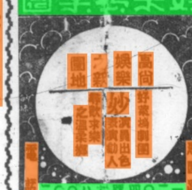


中國現決不談和
抱定抗戰決心

蘇北戰事開展
華軍甚得手

國府明令褒揚吳佩孚

本報復刊小啟



法租界懸旗糾紛
今日可復業
各商店遵勸
被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

交涉經過

釋放經過

法報記載

河內航郵已通

外交界動態

本市精神總動員會 昨奉張群來電嘉勉

稻草船戶不堪勒索停業

應有懸國旗之自由

滑稽小說 一重天

淪陷區公產

本報具有廿一年的歷史是小型報紙鼻祖銷路普遍全國刊登廣告最有效力

晶報 The Crystal

發行人 A. L. Teodoro

館址漢口路二九九號　電話九三四二八

中華郵政特准掛號立劵照按總包優待寄送之報紙

◀公共租界工部局登記證C字廿四號▶

◀第三千七百九十五號▶

中華民國貳拾捌年四月廿一日　星期五　今日貳張售國幣貳分四厘　本報創於民國八年三月三日

各路華軍大舉反攻
克復石龍包圍南昌
贛南高安發生激烈爭奪戰
綏晉重要據點多處已收復

國際時事 地圖

地中海的海軍根據地

抱定抗戰決心
中國現決不談和
英當局亦聲明未提議調解事

蘇北宿遷沭灌各縣
華軍甚得手

開封城緊閉
日僞鈔大跌

枯嶺未失陷

綏包軍不通

皖邊區

本報復刊小啓

本報自四月廿一日起，由美商 Teodoro 君為發行人，徹底改組，用於今日復刊。今後本報有大前提：促進東西兩大民主國親密攜手，謀太平洋上之集體安全，以求民治思想益臻實現，此其一。造成健全正大之輿論，取公開態度，為大衆喉舌，凡有專著、小品、新聞、趣事，靡不虛心採納，而於解除民生疾苦，增進人羣福祉，尤致意焉，此其二。謹布區區，諸希明鑒。

法租界懸旗糾紛
各商店遵勸
今日可復業
被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

交涉經過

釋放經過

法報記載

河內

外僑動態

本市精神總動員會
昨張群來電嘉勉

西北金融網已完成

不堪勒索停業

任何國國民
有懸國旗之自由
晉元團長昨發告同胞書

小說　一重天
陰兵們到了一重天

淪陷區公產

江海關對外貿易

公共租界

Towards full text - 5:

First results

晶報 The Crystal

發行人 L. A. Teodoro
第三千七百九十五號
中華民國貳拾貳捌年四月廿一日 星期五 今日貳張 每份國幣貳分四厘 本報創於民國八年三月三日

各路華軍大舉反攻 克復石龍包圍南昌

贛南高安發生激烈爭奪戰
綏晉重要據點多處已收復

國際時事地圖

抱定抗戰決心 中國現決不談和

英當局亦聲明未提議調解

蘇北戰事展開
宿遷淮陰各縣 華軍甚得手

本報復刊小啓

法租界懸旗糾紛 各商店遵勸 今日可復業

被捕市民昨日均已釋放
關於以後自由懸旗一節靜待解決

交涉經過

釋放經過

法報記載

河內航郵已通

外交界動態

本市精神總動員會 昨舉張群來電嘉勉

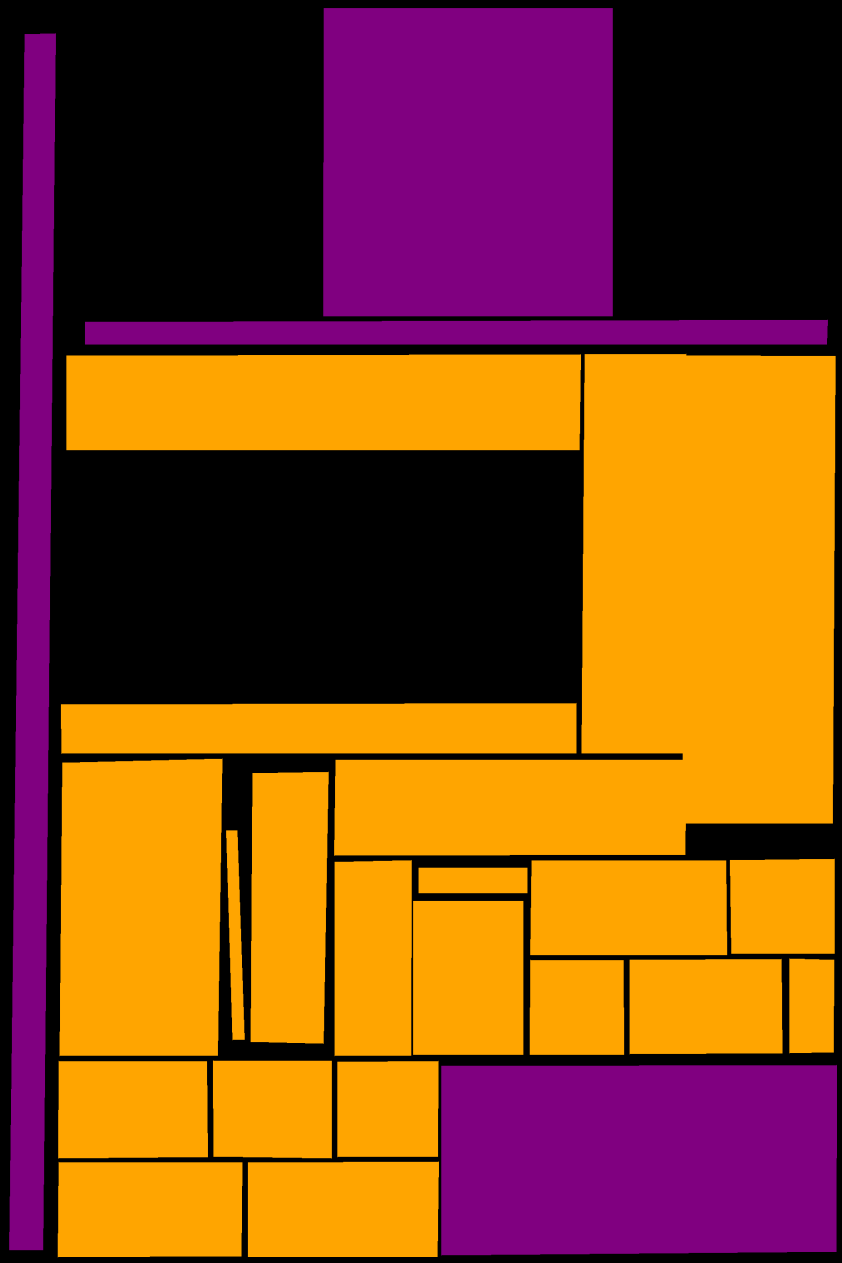
西北合作社已完成

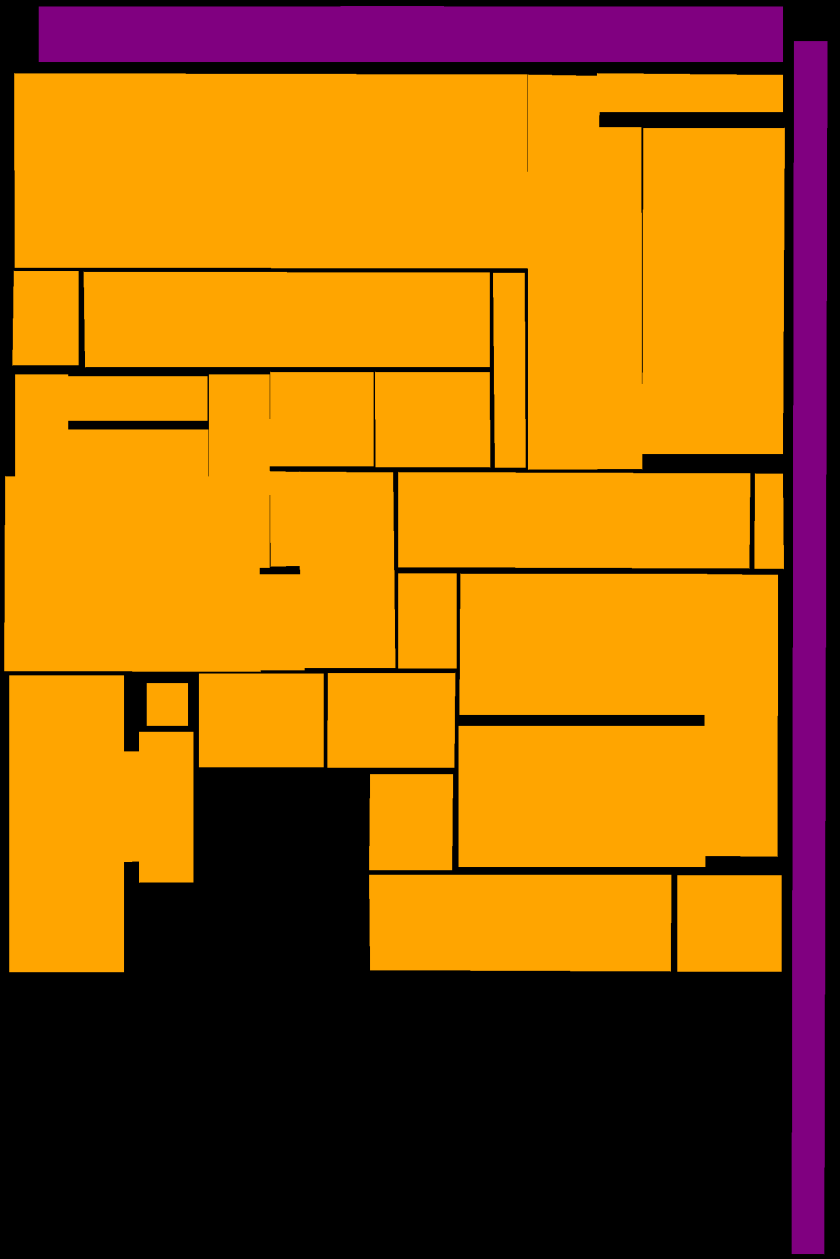
任何國國民 應有懸國旗之自由

稻草航戶 不堪勒索停業

滑稽小說 一重天

淪陷區公產





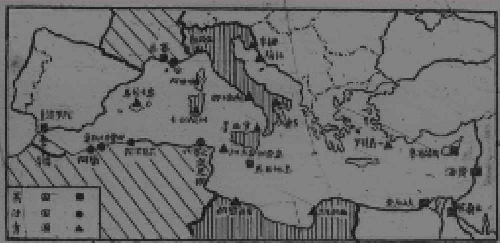
晶報 The Crystal

發行人 A. L. Teodoro

第三千七百九十五號

中華民國廿捌年四月廿一日 星期五 今日貳張 每份國幣貳分四厘 本報創於民國八年三月三日

國際時事地圖



各路華軍大舉反攻 克復石龍包圍南昌

贛南高安發生激烈爭奪戰 綏晉重要據點多處已收復

抱定抗戰決心 中國現決不談和

蘇北戰事緊張

華軍甚得手

本報復刊小啓

本報自四月二十日起，由美商A. L. Teodoro君爲發行人，徹底改組，用於今日復刊。今後本報有二大轉向：促進東西兩大民主國親密携手，謀太平洋上之集體安全，以求民主思想之實現，此其一。造成健全正大之輿論，取公開態度，爲大衆喉舌，凡有專著、小品、新聞趣事，靡不虛心採納，而於解除民生痛苦，增進人羣福利，尤致意焉，此其二。謹布區區，諸希明鑒。

法租界懸旗糾紛

各商店遵勸 今日可復業

被捕市民昨日均已釋放

交涉經過

河內

外交界動態

稻草船戶 不堪勒索停業

應有懸國旗之自由

已完成

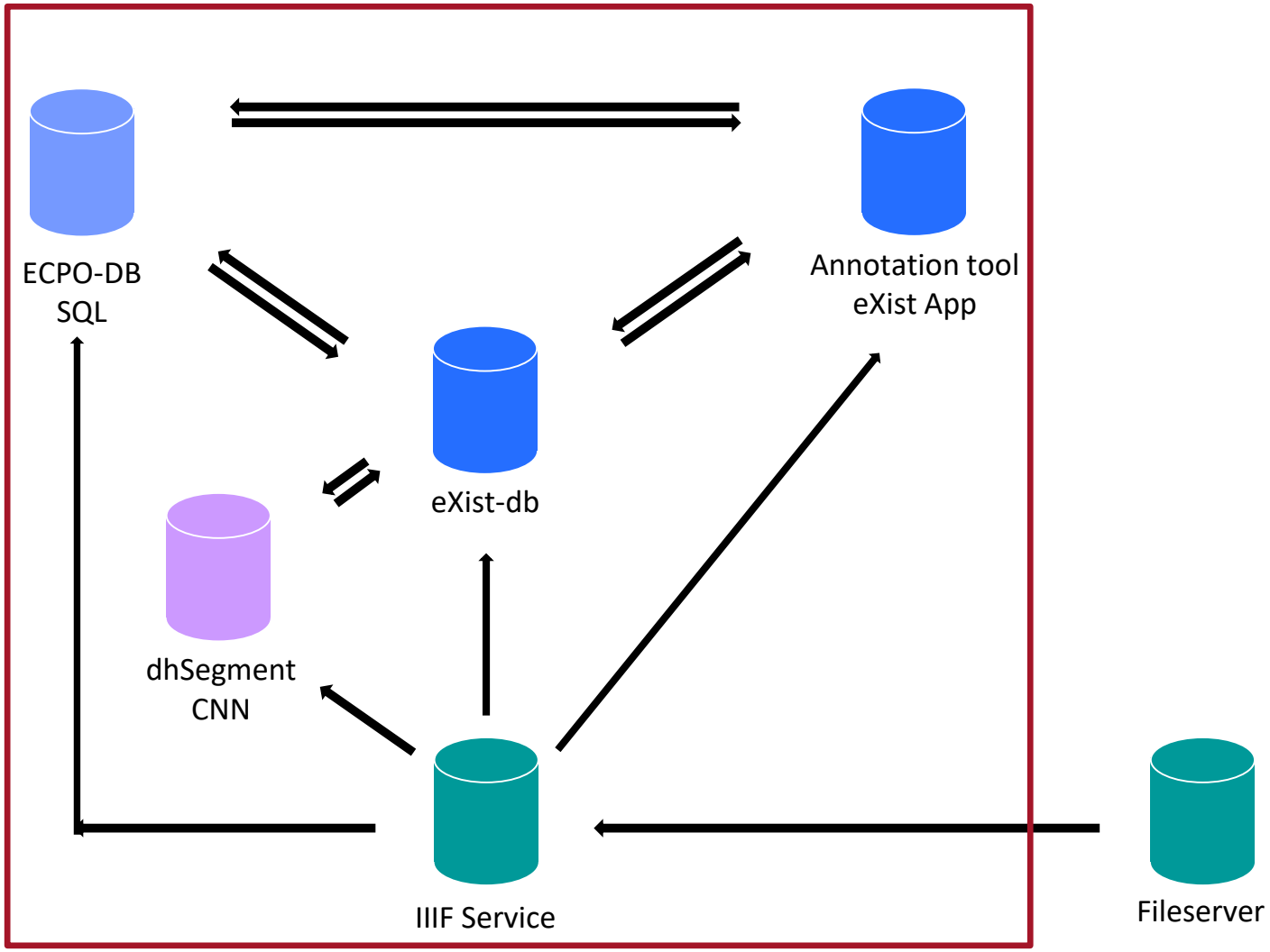
淪陷區公產

一重天（滑稽小說）

Towards full text - 6:

Infrastructure

ECPO – Expanding the infrastructure



Closing remarks

(more challenges)

Challenges - I

Scalable infrastructure to accomodate additional modules, and (potentially) more data/traffic

- Solution: re-design of infrastructure, containerisation, XML-db (texts)

Neural networks require strong machines (e.g. GPU server)

- Solution: bwForCluster MLS&WISO (High Performance Cluster)

Looking for partners to collaborate

- Our data is available for re-use, if you feel challenged by Chinese newspapers from the Republican era – do go ahead
- If you are interested in any kind of partnership – do let me know

Challenges - II

Fighting the Latin-script bias

Algorithms working for European medieval manuscripts or 19th ct. Newspapers can NOT just be re-used

Especially non-Latin / non-alphabetic script material shows characteristics, existing software solutions may still be unaware of

Solutions?
publish data-sets, make these characteristics more commonly known, explain specific challenges to the expert-communities

Challenges - III

Fighting biased opinion - „Aren‘t the Chinese already doing it, anyway?“

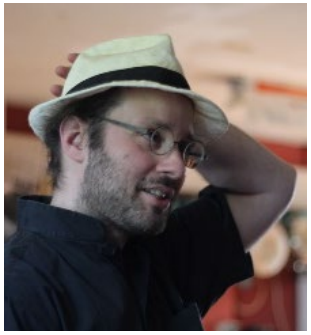
- Little is known about Chinese research on processing historical material
- Research about Republican newspapers is no “hot topic“, like, for example, “handwritten text recognition”
- Academic projects using Neural networks for research are scarce
- Chinese Libraries usually “just” offer author-title indexes
- Access to academic research in China usually behind paywall and/or limited from outside China

=> Systematic literature overview (in progress)

- Where can relevant information be found
- Which projects use neural networks for their research
- Which approaches were chosen, what worked, what did not
- Can results be reproduced, are models and datasets shared
- Are we able to re-use their algorithms, models, data sets

- Our report will be submitted to a Special Issue on Digital Humanities and East Asian Studies, IJDH (due May 2021)

Our Team



Duncan Paterson

Jia Xie 谢佳

Matthias Arnold

Part of the larger research group

Selected References

Women and the Periodical Press in China's Global Twentieth Century: A Space of Their Own? Ed. by Joan Judge, Barbara Mittler and Michel Hockx, Cambridge University Press, 2018.

Oliveira, Sofia Ares, Benoit Seguin, and Frederic Kaplan. "DhSegment: A Generic Deep-Learning Approach for Document Segmentation." In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 7–12. IEEE, 2018. https://doi.org/10.1109/ICFHR-2018.2018.00011.

Arnold, Matthias, and Lena Hessel. "Transforming Data Silos into Knowledge: Early Chinese Periodicals Online (ECPO)." In *Heuveline, Vincent, Gebhart, Fabian und Mohammadianbisheh, Nina (Hrsg.): E-Science-Tage 2019: Data to Knowledge*, 95–109. Heidelberg: heiBOOKS, 2020. https://doi.org/10.11588/heibooks.598.c8420.

Sung, Doris, Liying Sun, and Matthias Arnold. "The Birth of a Database of Historical Periodicals: Chinese Women's Magazines in the Late Qing and Early Republican Period." *Tulsa Studies in Women's Literature* 33 (2): 227–37, 2014. https://doi.org/10.1353/tsw.2014.0004.

Arnold, Matthias. "Multilingual Research Projects: Challenges for Making Use of Standards, Authority Files, and Character Recognition." *Digital Studies / Le champ numérique*, special issue "Towards Multilingualism in Digital Humanities", ed. Martin Lee und Cosima Wagner (manuscript accepted by Special issue editors, currently under review by journal).

Wick, Christoph, and Frank Puppe. "Fully Convolutional Neural Networks for Page Segmentation of Historical Document Images." *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, April 2018, 287–92. doi:10.1109/DAS.2018.39

Liebl, Bernhard, and Manuel Burghardt. "An Evaluation of DNN Architectures for Page Segmentation of Historical Newspapers," 15 April 2020. https://arxiv.org/abs/2004.07317v1

Early Chinese Periodicals Online (ECPO) – Project introduction. AAS2021. https://tinyurl.com/ecpo-intro

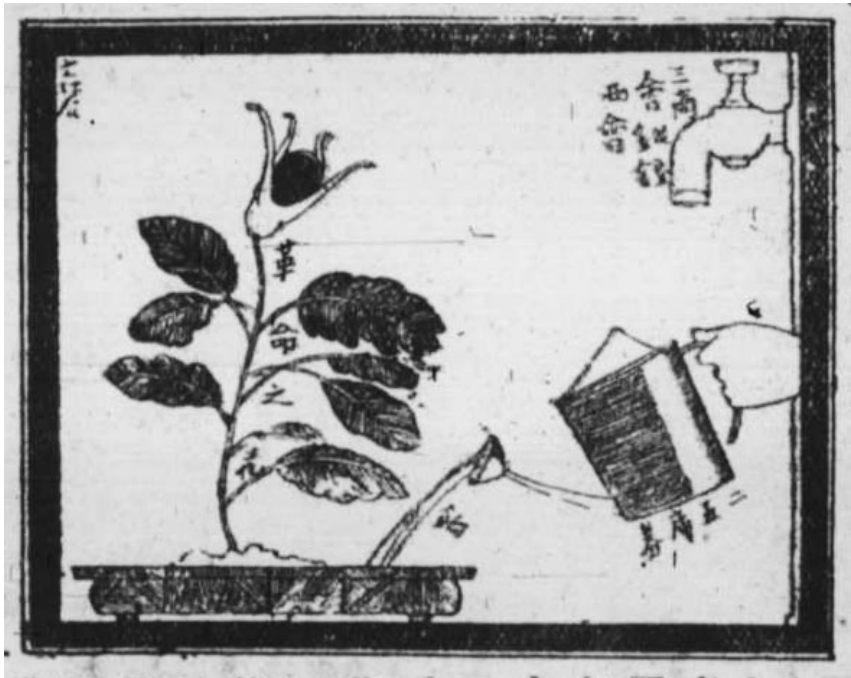
Contact

Matthias Arnold

matthias.arnold@uni-hd.de
http://tinyurl.com/matthias-arnold

http://uni-heidelberg.de/ecpo

https://github.com/exc-asia-and-europe/ecpo



Heidelberg Centre for Transcultural Studies

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Thank you!