

# 學黨史 悟思想 辦實事 開新局

以優異的成績迎接建黨一百周年

不/忘/初/心 牢/記/使/命



中國是世界四大文明古國之一，中國浩如煙海的文獻典籍記錄了歷史。然而遺憾的是，近代以來超過400萬冊中國古籍由於種種原因流散海外。

阿里巴巴達摩院、四川大學、美國加州大學伯克利分校近日公布，他們花費兩年多的時間，教會了AI識別中國古籍，將它們以數字化的形式接回故土，置於網絡公益平台，向所

有人開放。

這個項目的名字叫“漢典重光”，意為讓失落的珍貴古籍重放光芒。“漢典重光”以97.5%的準確率，識別了20萬頁古籍，覆蓋3萬多字的古籍字典。

## 遺失公告

鄉城縣興鄉建材有限公司等27個企業(單位)在甘肅州信用聯社股份有限公司鄉城信用社(現甘肅農村商業銀行股份有限公司鄉城支行)開立的開戶許可證遺失作廢，現將明細公布如下：

序號	存款人名稱	許可證編號	存款賬號
1	鄉城縣興鄉建材有限公司	J6827000015502	88330120000780041
2	鄉城縣植物油廠	J6827000015602	88330120000779700
3	鄉城縣農牧農村和科技局	Z6827000012501	88330120000748923
4	鄉城縣頤康綠色食品開發有限公司	J6827000023401	90220120000000699
5	鄉城縣正斗鄉永德村藏山羊養殖農民專業合作社	J6827000023702	90220120000000502
6	鄉城縣阿興民族民間手工藝貿易有限公司	J6827000024601	90220120000001302
7	鄉城縣然烏鄉納木村村民委員會	J6827000026002	90220120000003293
8	甘肅州亞丁紅酒莊有限公司	J6827000028601	90220120000005285
9	鄉城縣熱烏鄉克麥村高山牲畜養殖專業合作社	J6827000028902	90220120000005481
10	鄉城縣正斗鄉仁額村村民委員會	J6827000032001	90220120000010789
11	鄉城縣正斗鄉同麥村村民委員會	J6827000032601	90220120000010397
12	鄉城縣定波鄉開村村民委員會	J6827000034401	90220120000011881
13	鄉城縣正斗鄉白河村村民委員會	J6827000035501	90220120000010299
14	鄉城縣龍達建築材料有限公司	J6827000040801	90220120000014595
15	鄉城縣香巴拉格桑花文化傳播公司	J6827000041001	90220120000014399
16	鄉城縣德吉嶺文化旅游發展有限公司	J6827000041301	90220120000014791
17	鄉城縣扶貧和移民工作局	Z6827000014101	90220120000015081
18	四川康利建築工程有限公司	J6827000043101	90220120000015865
19	鄉城縣正斗鄉通麥村特色養殖專業合作社	J6827000045602	90220120000014693
20	四川松利建設工程有限公司鄉城縣分公司	J6827000053701	90220120000019359
21	道隆集團工程有限公司道隆集團工程有限公司 鄉得路TJ2標項目經理部	L6827000028201	90220120000023399
22	四川省鄉城縣桑登珠機械租賃有限公司	J6827000063101	90220120000024395
23	鄉城縣百藏坊旅遊文化產業有限公司	J6827000063201	90220120000024493
24	茂華裝飾材料專賣店	J6827000065102	90220120000024591
25	鄉城縣那拉崗村高山生態養殖專業合作社	J6827000065201	90220120000024669
26	四川科園環衛服務有限公司	L6827000029101	90220120000025097
27	鄉城縣融媒体中心機關工會	Z6827000015201	90220120000023869

## 科技守護文明 中國古籍“回家”之路

◎ 新華社記者 童芳

### 現代科技照亮中國古籍“回家”之路

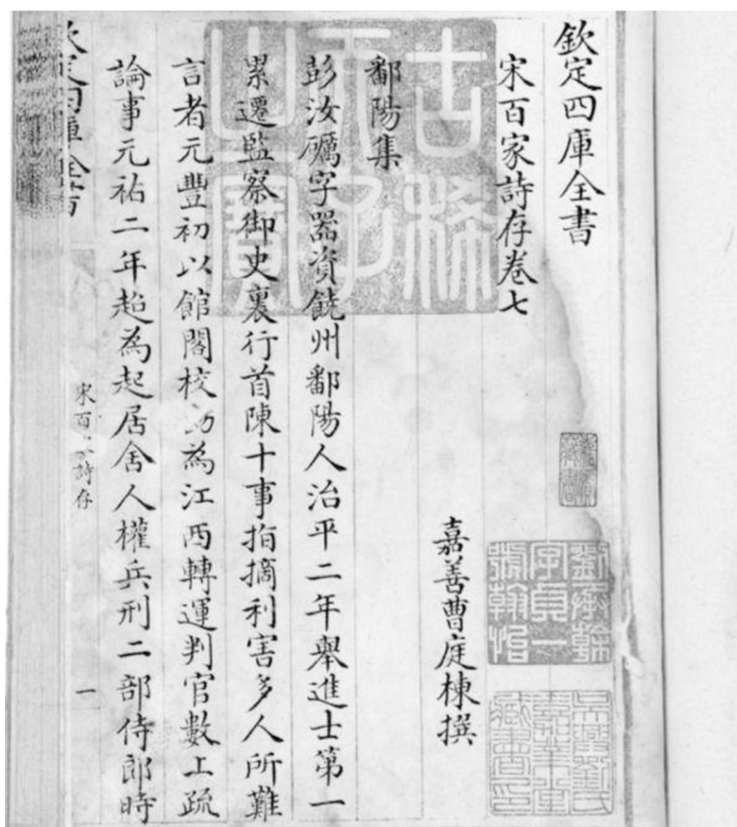
中國古籍擁有非常龐大且複雜的知識體系，包括甲骨簡牘、敦煌遺書、宋元善本、明清精槧、拓本輿圖、少數民族文獻等等。所涉及的範圍也是極其廣泛，有應對自然災害、流行疫病、經濟波動、政治鬥爭、外交危機、氣候變遷等的經驗，有戰爭、瘟疫、地震、洪澇災害、病蟲害等方面的經驗總結，還有醫療、中醫、養生、家具、服飾、飲食文化等生活經驗。

加州大學伯克利分校的東亞圖書館是全美三大東亞圖書館之一，90萬冊藏書里四成都是中文書，還有不少甲骨文和拓片。第一次到這兒的中國學者總感覺在穿越歷史的“蟲洞”，這些古籍不能運回中國，那就用數字化手段讓古籍的內容“回家”。

2019年，阿里巴巴和四川大學提出“數字化回歸”設想，四川大學歷史文化學院副院長王果與中央文獻研究室館員陳力牽線搭橋，溝通北美、歐洲、日韓等地藏書機構，最後獲得加州大學伯克利分校支持，達成共識，將伯克利東亞圖書館的中文古籍善本逐步數字化。

### 前所未有的挑戰

據王果介紹，“漢典重光”的分工非常明確——採集側把紙質書變為影印版，數字化生產側把影印版變為文字版，應用側為文字版增加檢索、字典和知識圖譜等研學系統。東亞圖書館完成第一個環節後，四川大學將和達摩院共同完成另外兩個步驟。其中，四川大學將提供一切非計算層面的專業支持，並與達摩院的機器視覺實驗室合作，共同開發古籍AI技術。首批數字化的古籍共20萬頁，



古籍四庫全書。

包含40餘種珍貴宋元善本。刚开始对这个项目很有信心，但实际上上手才知道难度有多大。

首先是中國古籍的載體很多，紙、布、竹子、木頭、甲骨、石碑……幾乎所有能用的載體都被古人留了字，不同載體上面的字識別起來差別非常大。年代久遠的紙張大多殘缺不全，上面還布滿斑點，而且排列非常複雜。古人喜歡從上到下，從右到左，還非常喜歡在上面做批注。字跡的精美也成了負擔。隸書、楷書、草書、行書都漂亮，但也真難

認。大部分字還是手寫的，不但兩個人寫的字不同，同一個人寫的字也差別很大，很多字還有不同寫法。

團隊原有的OCR(圖像文字識別)的識別準確率只有40%，這顯然是不夠的，較早涉足古籍識別的Google Books(谷歌圖書)針對的都是英文古籍，對中文也不適用。

### 大量創造性和創新性的辛勤勞動

沒有前車可鑒，只能自己來了。

“這就像教小朋友識字，確實難，但也有簡單的部分，我們就由易到難慢慢做。”王果說。

前所未有的開拓性成果，包含着中國學者大量創造性和創新性的辛勤勞動。

達摩院的古籍識別算法，用AI替代人工，大幅壓縮了專家標註工作量。在機器為主進行識別的97.5%的內容中，約有1%(1萬字左右)需要專家錄入；機器不能識別的余下2.5%(2.5萬字)的文字，全部交給專家做後期標註。相比人工專家錄入，百萬字書籍的數字化工作量從1000天降低到了35天，效率比人工專家錄入方案提升近30倍。這種古籍識別算法，為中華古籍的回歸提供了另一種可行的思路。

非計算機層面的工作也非常重要，四川大學專門組織了30多個歷史系學生來做標註，這很耗神，眼力、腦力缺一不可，但這個過程沒法省略，就像學生要學習，算法也要迭代，“只有數據夠多夠好，機器才能搞定”。

這是一件非常難但非常有意思的事情。古籍識別的挑戰巨大，但大家還是想通過技術讓古籍活起來，用科技守護文明。

據了解，達摩院、四川大學、美國加州大學伯克利分校、中國國家圖書館、浙江圖書館將繼續投入人力、物力，擴大古籍數字化回歸的數量，讓所有蒙塵的古籍重煥新生。