

# Metadata Practices for Chao (Yuen Ren) Papers Project

Haiqing Lin

East Asian Library, UC Berkeley

# Prof. Chao's papers



Chao Yuen Ren was an Agassiz Professor of Oriental Languages at UC Berkeley. He was credited with being the founder of modern linguistics in China. His papers that document his entire personal and professional life are varied in nature of materials and range in the origin period from early years to the time of death. The papers include personal and professional correspondence, research and teaching files, diaries, manuscripts, notes, fieldwork materials, and personal documents.

# Digitization of Chao's papers

- All the materials have been photo scanned at high resolution as an electronic collection comprising more than 110,000 digital images.

Y. R. CHAO  
1059 CREWAVANT AVE  
BERKELEY 8, CALIFORNIA  
OCT 27 1955

加禮宛先生：  
—— 敝人最近，報上看見兩次  
華人的新聞，十分欣慰，現在有  
志相見，小培培培培培培培培培  
培培培培培培培培培培培培培培  
教授對華語研究，深受愛護，  
大學題為 Teaching Assistant，  
美領事館，已正式發行，結果有  
加拿大 Alberta 大學，為化工師  
師 (Contributor)，上月已對校  
上課了，他現在這希望其美  
國較大一點地方，以便改進研  
究的機會，不知您知道有甚  
好機會，可以給他介紹一下，以  
前所辦的之學生資格，簽名未  
成，如果能有 instructor 聘書，  
Award of Diploma 似可較為辦入  
境手續，您是否為行，一不為荷。  
附上履歷一紙，備考，此致仰  
—— 程方

Map. 3 蘇澳  
22.22 24° - 34' 10" N.  
121° 30' - 40' E

30° - 32'	阿拉山 Anashi-mia (Mt) 31-5'	口子社 (K) K'ao-shia (Vg) 36-9'	羅社 (K) K'ao-shia (Vg) 37-8'
32° - 34'	塔山 Tag-shia (Mt) 33-9'	亞斯林社 (K) K'ao-shia (Vg) 37-5'	亞斯林社 (K) K'ao-shia (Vg) 37-5'
34° - 36'	加禮宛山 Kalle-nan-shia (Mt) 33-3'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	三棧 Sa-tsan 36-6'	木塔溪 Bi-tun-k'ye (Cyt) 36-0'	木塔溪 Bi-tun-k'ye (Cyt) 36-0'
	新城山 Sin-sia-shia (Mt) 36-6'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	三棧 Sa-tsan 36-6'	木塔溪 Bi-tun-k'ye (Cyt) 36-0'	木塔溪 Bi-tun-k'ye (Cyt) 36-0'
	加禮宛 Kalle-nan-shia (Mt) 33-3'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	花連溪 Hwa-lian-shia (Cyt) 36-1'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	沙連溪 Sha-lian-shia (Cyt) 35-1'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	十六股 Tsap-lak-ko 35-0'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	歸化 Kuei-hua 35-0'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	36° - 38'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'
	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'	塔山 Tag-shia (Mt) 33-9'

CHEMICAL ANALOGIES IN  
CHINESE GRAMMATICAL STRUCTURE

Yuen-Tsun Chao  
University of California, U.S.A.

Analogy in the relation between two systems which share  
certain formal features. An analogy in one of two  
analogous systems selected to illustrate the

have certain formal features in common, but differ in  
others. The relation of analogy is symmetrical: if system S<sub>1</sub>  
is analogous to system S<sub>2</sub>, then system S<sub>2</sub> is analogous  
to system S<sub>1</sub>. But when we speak of an analogy  
or analogies from system S<sub>2</sub> for the purpose of  
understanding system S<sub>1</sub>, we are far as S<sub>1</sub> shows  
certain properties of S<sub>2</sub> in common with S<sub>1</sub>, these  
the relation is no longer symmetrical. Thus  
one would likely illustrate chemical structure  
by Chinese grammar, while it is  
likely to illustrate the latter by chemical  
analogies.

The oldest convention in Chinese states with  
a chemical analogy in that of a compound. To the  
generation of traditional grammarians any combination  
of more than one character which forms a new  
function or meaning is a compound. Not only

# Functional requirements for metadata

The following functional requirements were defined for the metadata solution for Chao's papers in order to organize this large-scale digital collection:

1. classify Chao's papers into different categories,
2. describe the core features of materials, and
3. control the workflow to organize the materials for publishing

# Categories

- 1. Correspondences
- 2. Diary
- 3. Manuscripts and research materials
  - 3.0. Phonetics and Phonology
  - 3.1. Grammar
  - 3.2. Dialects
    - 3.2.1. Dialect Studies
  - 3.3. Pragmatics
  - 3.4. Language Teaching
  - 3.5. History of Linguistics
  - 3.6. Vocabulary and Lexicography
  - 3.7. Linguistic Theory
  - 3.8. Biography
  - 3.9. Others
- 4. Lectures and teaching materials
  - 4.0. Courses (including training courses)
  - 4.1. Seminars
  - 4.2. Speech
  - 4.9. To be determined
- 5. Work files
- 6. Personal Items
- 7. Others

# Metadata Elements

- Captions for manuscripts: title, others, content descriptions
- Date
- Types pre-defined
- Person1 for Correspondences: senders; for manuscripts: authors
- Agent1 organizations
- Person2 for Correspondences: receiver; for manuscripts: other authors
- Agent2 organizations
- Place pre-defined
- Note
- Folder generated
- File Name generated
- Viewed generated
- Sort\_Key generated
- File\_Id generated
- Ordinal generated
- Entry generated

# Data Samples

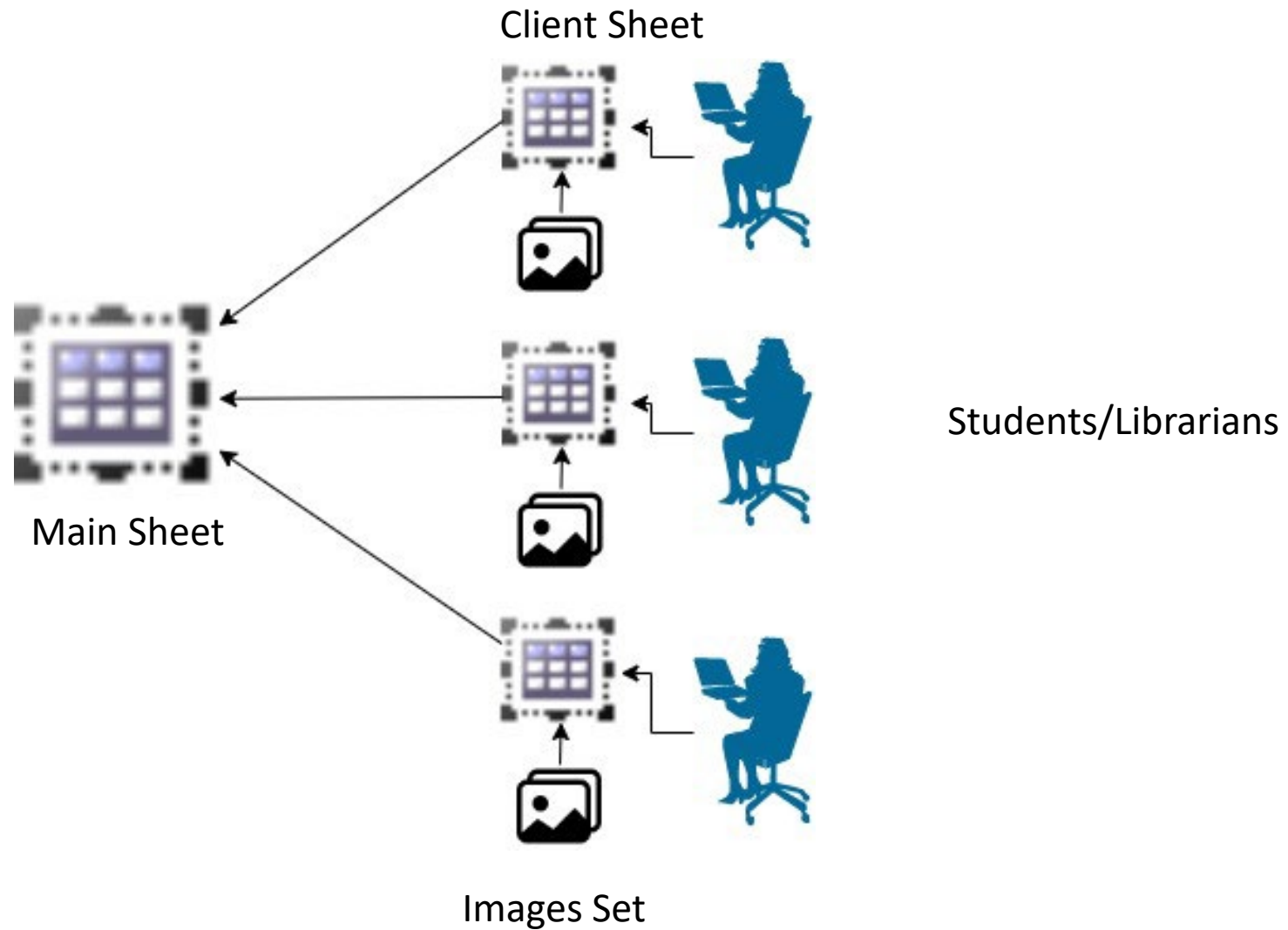
ID	Folder	File Names	Viewed	Caption	Categories	Date	File_ID	Ordinal	Note	Sort_Key	Heading	Person1	Person2	Agent1	Agent2	Place	Types	
606	Carton\01\23	01_023_0038.JPG	Viewed on 3/12/202	Ballou, Dorothy 1964-1965	1. 书信 (完整的書信)	1967.06.14	01_023_0038	0	Ballou, Dorothy 1964-1965	1	许芥昱	许芥昱	赵元任			San Francisco, CA	手稿類型--2.打字	
607	Carton\01\23	01_023_0039.JPG	Viewed on 3/11/202	Ballou, Dorothy 1964-1965	1. 书信 (完整的書信)	1967.06.14	01_023_0038	1	Ballou, Dorothy 1964-1965	1	许芥昱	许芥昱	赵元任			San Francisco, CA	手稿類型--2.打字	
608	Carton\01\23	01_023_0040.JPG	Viewed on 3/11/202	Ballou, Dorothy 1964-1965	1. 书信 (完整的書信)	1967.06.14	01_023_0038	2	Ballou, Dorothy 1964-1965	1	许芥昱	许芥昱	赵元任			San Francisco, CA	手稿類型--2.打字	
1130	Carton\01\48	01_048_0008.JPG	Viewed on 3/11/202	Bloch, Bernard 1940-1961	1. 书信 (完整的書信)	1963.01.22	01_048_0008	0	Bloch, Bernard 1940-1961	19	Bernard Bloch	Bernard Bloch	赵元任			New Haven, Ct	Indic and Far Easster	手稿類型--2.打字
1140	Carton\01\48	01_048_0016.JPG	Viewed on 3/11/202	Bloch, Bernard 1940-1961	1. 书信 (完整的書信)	1961.01.16	01_048_0016	0	Bloch, Bernard 1940-1961	19	Bernard Bloch	Bernard Bloch	赵元任			New Haven, Ct	Yale University	手稿類型--2.打字
1149	Carton\01\48	01_048_0024.JPG	Viewed on 3/11/202	Bloch, Bernard 1940-1961	1. 书信 (完整的書信)	1959.06.23	01_048_0024	0	Bloch, Bernard 1940-1961	19	Bernard Bloch	Bernard Bloch	赵元任			Kyoto, Japan		手稿類型--2.打字

# Workflows

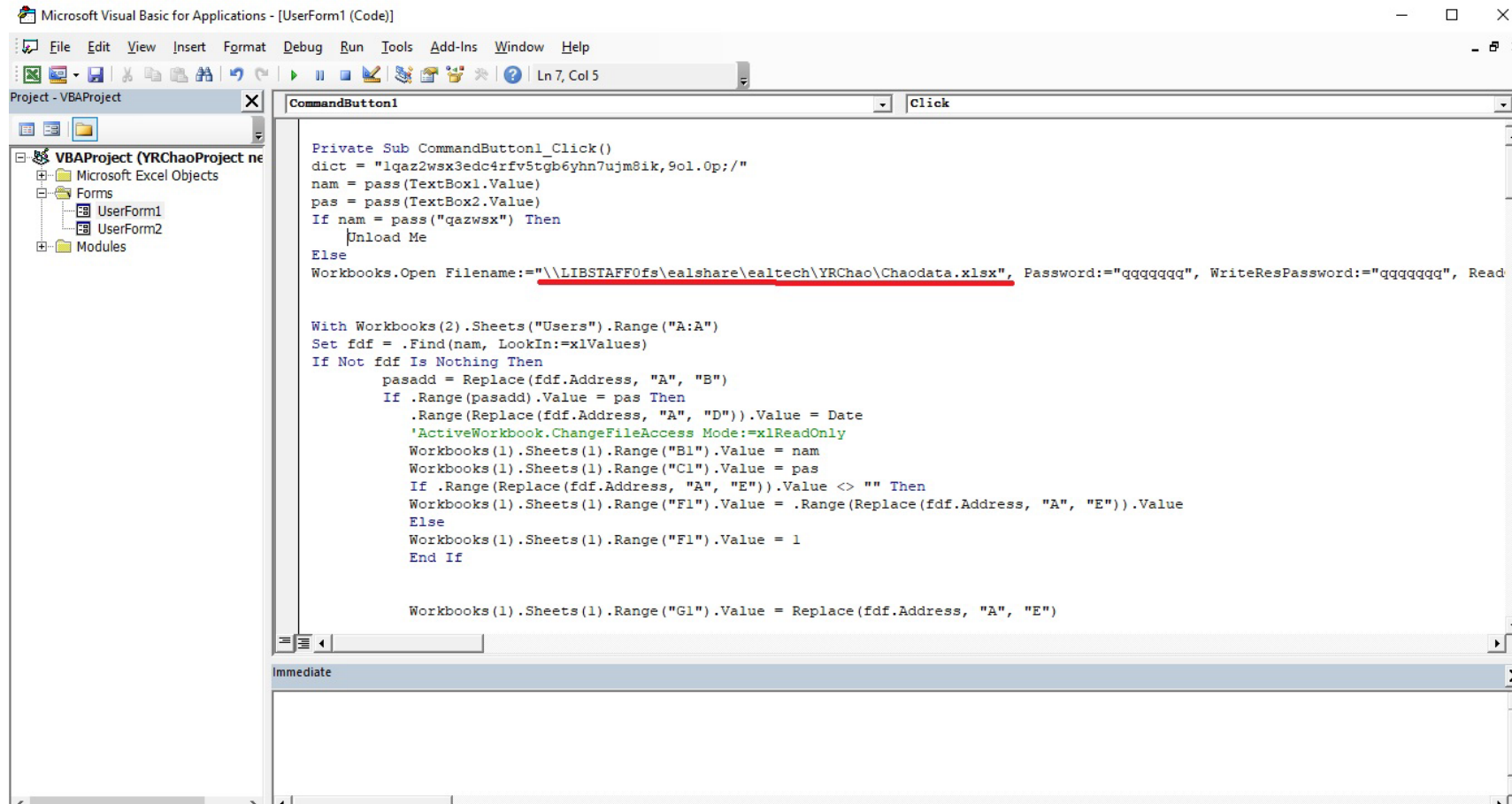
- A group of student workers was involved in extracting metadata elements from digital images. We established a work environment based on Excel spreadsheets that allows all the relevant workers to share their work and avoid duplicating efforts.
- The work environment can display the images on the side by the spreadsheets to facilitate data collection and validation. It can also generate various indexes and word documents for publishing. This environment significantly improves the efficiency of metadata generation for the project.
- Due to the security reason and size of image data set(219GB), the image data needs to be stored in local rather than online



# Server Client like Architecture



# Connect to the shared sheet



The screenshot displays the Microsoft Visual Basic for Applications (VBA) editor. The Project Explorer on the left shows a VBAProject named 'YRChaoProject' with a 'Forms' folder containing 'UserForm1' and 'UserForm2', and a 'Modules' folder. The main window shows the code for 'CommandButton1\_Click' in 'UserForm1'. The code is as follows:

```
Private Sub CommandButton1_Click()  
dict = "1qaz2wsx3edc4rfv5tgb6yhn7ujm8ik,9o1.0p:/"  
nam = pass(TextBox1.Value)  
pas = pass(TextBox2.Value)  
If nam = pass("qazwsx") Then  
    Unload Me  
Else  
    Workbooks.Open Filename:="\\LIBSTAFF0fs\ea\share\ea\tech\YRChao\Chaodata.xlsx", Password:="qqqqqq", WriteResPassword:="qqqqqq", Read  
  
    With Workbooks(2).Sheets("Users").Range("A:A")  
    Set fdf = .Find(nam, LookIn:=xlValues)  
    If Not fdf Is Nothing Then  
        pasadd = Replace(fdf.Address, "A", "B")  
        If .Range(pasadd).Value = pas Then  
            .Range(Replace(fdf.Address, "A", "D")).Value = Date  
            'ActiveWorkbook.ChangeFileAccess Mode:=xlReadOnly  
            Workbooks(1).Sheets(1).Range("B1").Value = nam  
            Workbooks(1).Sheets(1).Range("C1").Value = pas  
            If .Range(Replace(fdf.Address, "A", "E")).Value <> "" Then  
                Workbooks(1).Sheets(1).Range("F1").Value = .Range(Replace(fdf.Address, "A", "E")).Value  
            Else  
                Workbooks(1).Sheets(1).Range("F1").Value = 1  
            End If  
  
            Workbooks(1).Sheets(1).Range("G1").Value = Replace(fdf.Address, "A", "E")  
        End If  
    End If  
End Sub
```

# Interface

The screenshot shows an Excel spreadsheet with a UserForm and data visualizations. The spreadsheet data is as follows:

Total	Unprocessed	Processed	Changed	Viewed
116164	106041	10122	7565	2557

Year	January	February	March	April	May	June	July	August	September
2018	10122								

The UserForm (UserForm1) contains the following fields and buttons:

- User Name:
- Password:
- Login (red button)
- Sign Up (blue button)
- View Statistics (button)

The data visualizations include:

- A pie chart showing the distribution of image states: Unprocessed (106041, 84%), Processed (10122, 8%), Changed (7565, 6%), and Viewed (2557, 2%).
- A bar chart showing the number of images for each month (January, February, March, April).

The status bar at the bottom of the spreadsheet displays the date range: 1915.01.19-1915.01.20.

# Interface

The screenshot displays the Microsoft Excel interface with a scanned document image and a data entry form overlaid on it.

**Excel Interface:**

- File Name: YRChaoProject2
- Current Tab: Picture 2
- Active Cell: M1

**Scanned Document (Picture 2):**

*Carnegie Chinese Project*  
San Francisco State College

June 14, 1967

Professor Y. R. Chao  
1059 Cragmont Avenue  
Berkeley 8, California

Dear Professor Chao:

We are very happy to hear from Leo that you have agreed to let the Asian Language Publications, Inc., handle the publication and distribution of your book. "Savable Chinese". The following is a brief

*To file*

**UserForm2:**

- User Name: Inq
- Last Record No.: 2
- Row No.: 2
- File Name: G:\chao\Carron\0123\01\_023\_0038.JPG
- Caption: Ballou, Dorothy 1964-1975
- Categories: 1-书信 (完整的书信)
- Year: 1967.06.14
- From: 许芥曼
- To: 赵元任
- CC: Members of the Board of Directors Asian Language Publications, Inc
- Place: San Francisco, CA
- Institution:
- Material Types: 手稿类型-2.打字稿
- Group:  Group with previous? 01\_023\_0038 0
- Note: Problem: [dropdown]
- Note: Solution: [text area]
- Buttons: Ignore, << Previous, Next >>, Save, Rotate, Enlarge, Save to Serve, Close

UserForm1

User Name:  Last Record No.

Row No.

File Name:

Caption:

Categories:

Year:

From:

To:

CC:

Place:

Institution:

Material Types

Group  Group with previous?

Note

Problem

Solution

UserForm2

User Name:  Last Record No.

Row No.

File Name:

Caption:

Categories:

Year:

From:

To:

CC:

Place:

Institution:

Material Types

Group  Group with previous?

Note

Problem

Solution

UserForm2

User Name:  Last Record No.

Row No.

File Name:

Caption:

Categories:

Year:

From:

To:

CC:

Place:

Institution:

Material Types

Group  Group with previous?

Note

Problem

Solution

UserForm2

User Name:  Last Record No.

Row No.

File Name:

Caption:

Categories:

Year:

From:

To:

CC:

Place:

Institution:

Material Types:

Group

Group with previous?

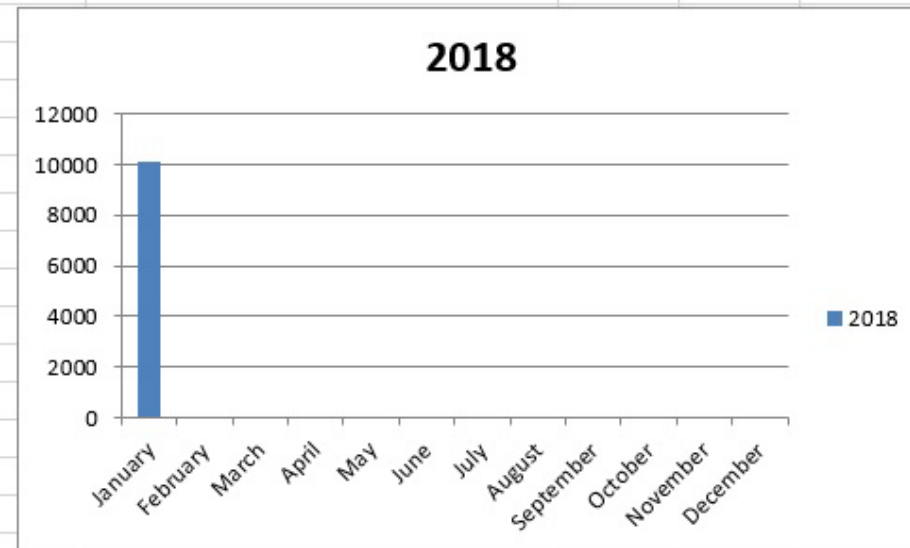
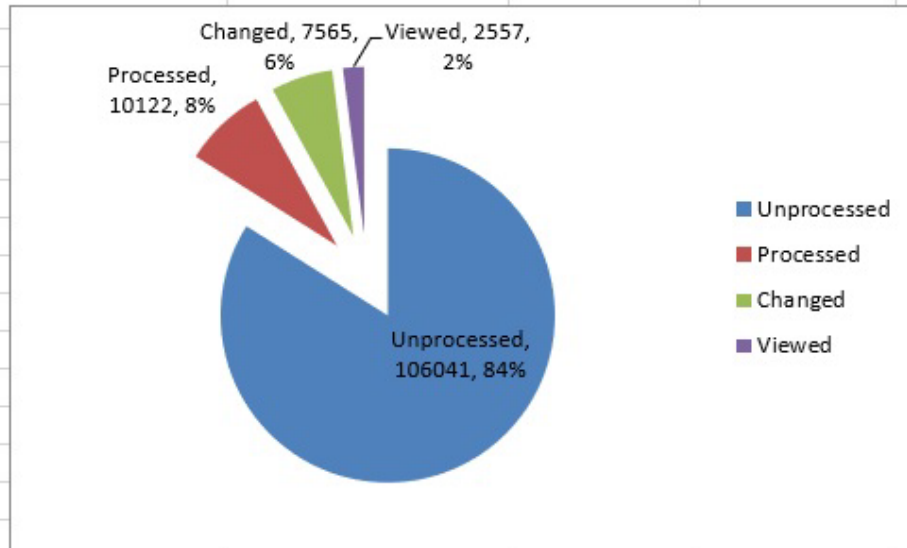
Note

Problem:

Solution:

	A	B	C	D	E	F	G	H	I
1	Categories	Names	ImageLoc	Problems	Types	Place	Institution		
2	0.待定书信	赵元任	G:\chao	无信头	手稿類型--0.手寫稿	廣東廉江	廣東廉江縣立中學		
3	1.书信 (完整的書信)	赵小中 ( Sheaujong )		incomplete	手稿類型--1.謄寫稿	廣州	國立中山大學		
4	1.书信 (完整的書信)--1.電報	<input type="checkbox"/> Brewster		不要	手稿類型--2.打字稿	Washington, D.C.	State-PB-Washington, D.C.		
5	1.书信 (完整的書信)--1.明信片	<input type="checkbox"/> Chen		不是自己的	手稿類型--3.印刷稿	Cambridge, MA	Harvard-Yenching Institute		
6	1.书信 (完整的書信)--0.家信	<input type="checkbox"/> Condry		手写批注	长沙	New Haven, CT	省立長沙女中		
7	1.书信 (完整的書信)--0.机构	<input type="checkbox"/> Cook		有關趙的著作的評論	The Journal of the C	Kowloon, Hong Kong	國民政府教育部國語統一籌備		
8	2.日记	<input type="checkbox"/> L. <input type="checkbox"/>		from whom?	University of Califor	長沙, 湖南	國民政府教育部國語推行委員		
9	3.手稿及研究資料	<input type="checkbox"/> Olive , Jack Condliffe		shawrjou Yingfang (韶州音	手稿類型--3.打字稿	北平中海	國語會		
10	3.手稿--0. 語音與音韻學	<input type="checkbox"/> R. Berol		shinhui (新會城)		北京	史語所		
11	3.手稿--1. 語法	<input type="checkbox"/> T. Davison		Meishiann		南京北極閣	Academia Sinica		
12	3.手稿--2. 方言	<input type="checkbox"/> ( 江亦雷 )		Goangjou		Nankang, Taipei, Taiwan	Acadenua Sinica		
13	3.手稿--21. 方言研究	<input type="checkbox"/>		Goangjou(廣州) 二乙		Nan-kang, Taipei, Taiwan	Dept. of Oriental Languages, U		
14	3.手稿--3. 語用學	A. Bruce Gaarder		nanning pying bain (南寧平白)		Berkeley, CA	福建省民政廳		
15	3.手稿--4. 語言教學	A. E. Zucker, Arthur W. Hummel, Y.P. Mei		Mei Shiann		福建	UC Berkeley		
16	3.手稿--5. 語言學史	A. Hoffmann ( 霍福民 )		Shawyou Keh (韶州客話)		1462 Scenic Ave, Berkeley	國立中央研究院歷史語言研		
17	3.手稿--6. 詞彙與詞典學	A. K. Warder		始興		龍頭村	Harvard University		
18	3.手稿--7. 語言學理論	A. Richard Diebold, Jr.		Guey Shiann(貴縣)		58 Nanhai Road , Taipei , Taiwan	Sino-British cultural associatio		
19	3.手稿--8. 傳記	A.C. Sutton		yneyoing Jiangjou(X平江口)		National Taiwan University, Taipei, Fromo	UNESCO		
20	3.手稿--9. 其他 ( 包括翻譯3.手	A.D.H. McLlwrick		梧州城		Taipei, Formosa	Chinese delegation of UNESCC		
21	3.手稿--9. 其他--地名表	A.H.McDannald		Shinhuey HorTsuen ( 河村 )		Taipei	Institute of Far Eastern Langua		
22	4.演讲、教學資料	A.V. Hill		Dongguan(東莞)		Seattle	The University of Wisconsin		
23	4.演讲、教學資料--0.課程 ( 包	A.W. Dorow		Guilin(桂林)		Ball State University	Stephens College		
24	4.演讲、教學資料--1.研討班	A.W. Hummel		Enyang ( 恩陽 )		Indiana	1946 UNESCO Seminar II		
25	4.演讲、教學資料--2.演講	Aaron Bar-Adon		Wenchang(文昌)		27 Walker St., Cambridge	The Masque Society		
26	4.演讲、教學資料--9.待定	Abraham Robinson		Jieyang ( 揭陽 )		644 Orange St., New Haven	Cornell University		
27	5.工作檔案 ( 機構 中文名稱--3)	Adam D. Horvath		韶州 (shawjou)		Boylston Hall, Harvard University	The National Research Institut		
28	6.私人物品	Adam G. Woyna		Lianjou(廉州)		Boylston Hall, Cambridge, Mass	University of Hong Kong, Facul		
29	6.私人物品--0.重要事件--時間	Adele Rickett		Guangjou		1059 Cragmont Ave., Berkeley	管理中英庚款董事會		
30	6.私人物品--1.通訊錄 ( 名片 )	Afia Dil		Nanjing Pyng ( 南寧平 )		27 Walker St. Cambridge, Mass	Chinese Embassy		
31	6.私人物品--2.證件, 证书	Agnes Fries		Taishan ( 台山 )		27 Walker St., Cambridge, Mass	China Film Enterprises of Ame		
32	6.私人物品--3.票據	Agnes Fries , Charles Fries		Shinhuey(新會)		126 Hotel Durant, Berkeley, CA	Bank of China		
33	6.私人物品--4.賀卡	Alan D. Corré		MeiShiann(梅縣)		216 Hotel Durant	Audio-Visual Equipment Suppl		
34	6.私人物品--5.剪報	Alan Dundes		Nanning Pyng		Boylston Hall, Harvard University, Mass	Chinese Emabssy		
35	6.私人物品--6.其他 ( 少量郵票	Albert E. Dien		TaishanCheng		27 Walker St., Cambridge 38 Mass	教育部音樂教育委員會		
36	7.存疑--1.主題不明	Albert H. Bowker		Shawjou ( 韶州 )		1462 Scenic Ave., Berkeley, CA	The Principia Press, Inc.		
37	7.存疑--2.無關緊要	Albert Storm		liehchang(樂昌墟村)		7 Harvard Place, Ann Arbor Michigan	來薰閣書店		
38	8.附件(非趙元任物品, 但和趙	Aldyth V. Morris		Wujou ( 梧州城 )		London	Queens College		

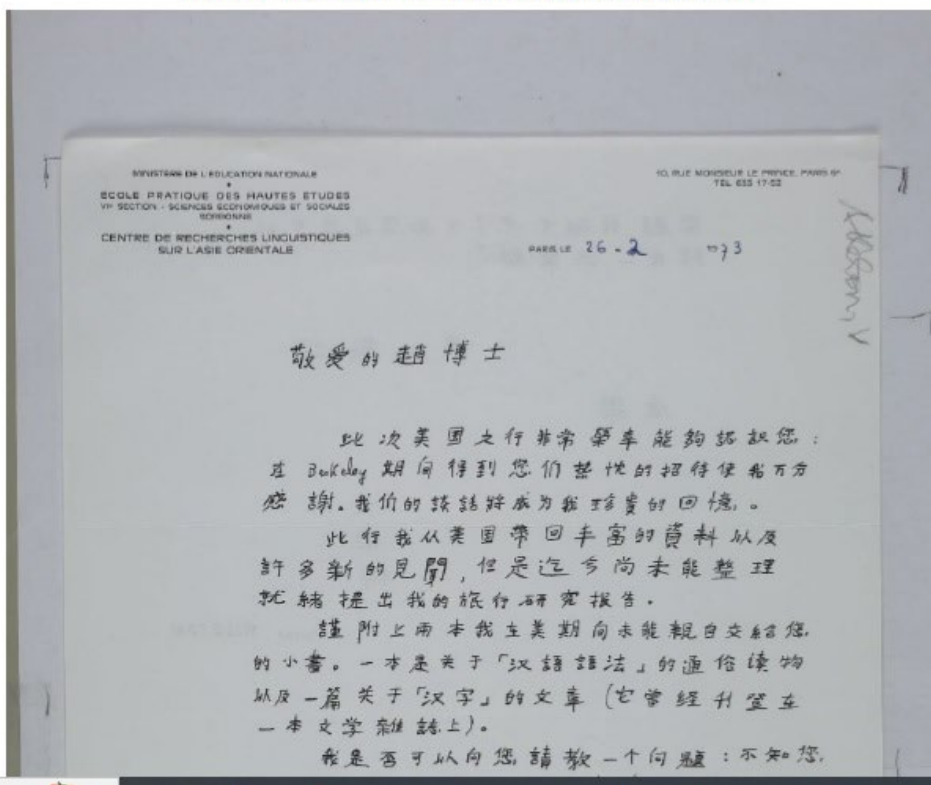
# Statistics



# Generate Word and PDF

Alleton, Viviane (艾乐桐) 【信件数: 1 件】

1973 年 02 月 26 日 Viviane Alleton 致赵元任



## 目录

Abercrombie, David 【信件数: 1 件】 .....	1
1936 年 03 月 18 日 David Abercrombie 致赵元任.....	1
Abrams, Samuel J. 【信件数: 1 件】 .....	2
1943 年 09 月 17 日 Samuel J. Abrams, Rabbi 致赵元任.....	2
Ace, Goodman 【信件数: 1 件】 .....	3
1967 年 02 月 29 日 赵元任致 Goodman Ace (The Saturday Review).....	3
Adams, John E. 【信件数: 4 件】 .....	4
1966 年 11 月 14 日 John E. Adams 致赵元任.....	4
1975 年 04 月 13 日 John E. Adams 致赵元任.....	5
日期不详 John E. Adams 致赵元任, 杨步伟.....	6
日期不详 John E. Adams 致赵元任, 杨步伟.....	7
Akiyasu, Todo (藤堂明保) 【信件数: 2 件】 .....	8
1959 年 03 月 26 日 Todo Akiyasu 致赵元任.....	8
1959 年 04 月 02 日 赵元任致 Todo Akiyasu (藤堂明保) .....	11



# Generate Index

Abercrombie, David 【信件数：1件】	1	2(01)001
1936年03月18日 David Abercrombie致赵元任	1	2(01)001
Abrams, Samuel J. 【信件数：1件】	2	2(01)002
1943年09月17日 Samuel J. Abrams, Rabbi致赵元任	2	2(01)002
Ace, Goodman 【信件数：1件】	3	2(01)003
1967年02月29日 赵元任致Goodman Ace (The Saturday	3	2(01)003
Adams, John E. 【信件数：4件】	4	2(01)004
1966年11月14日 John E. Adams致赵元任	4	2(01)004
1975年04月13日 John E. Adams致赵元任	5	2(01)005
日期不详 John E. Adams致赵元任, 杨步伟	6	2(01)006
日期不详 John E. Adams致赵元任, 杨步伟	7	2(01)007
Akiyasu, Todo (藤堂明保) 【信件数：2件】	8	2(01)008
1959年03月26日 Todo Akiyasu致赵元任	8	2(01)008
1959年04月02日 赵元任致Todo Akiyasu (藤堂明保)	11	2(01)011
Alatis, James E. 【信件数：1件】	12	2(01)012
日期不详 赵元任致James E. Alatis	12	2(01)012
Albright, Robert W. 【信件数：1件】	13	2(01)013
1955年02月05日 Robert W. Albright致赵元任	13	2(01)013
Albright, William Foxwell 【信件数：1件】	14	2(01)014
1958年02月23日 赵元任致William Foxwell Albright	14	2(01)014
Alderson-Smith, Gael 【信件数：2件】	15	2(01)015

# Next Generation

2.朱文方印  
2.朱文方印  
2.朱文方印  
2.朱文方印  
2.朱文方印  
0.9278664  
0.89638245  
0.85844046  
0.7954252  
0.76070446



- Thank you!