admin units and office reorganization

Ji Ma <ji_ma@fas.harvard.edu>   Mon, Jun 4, 2018 at 3:35 PM
To: Peter <peter_bol@harvard.edu>
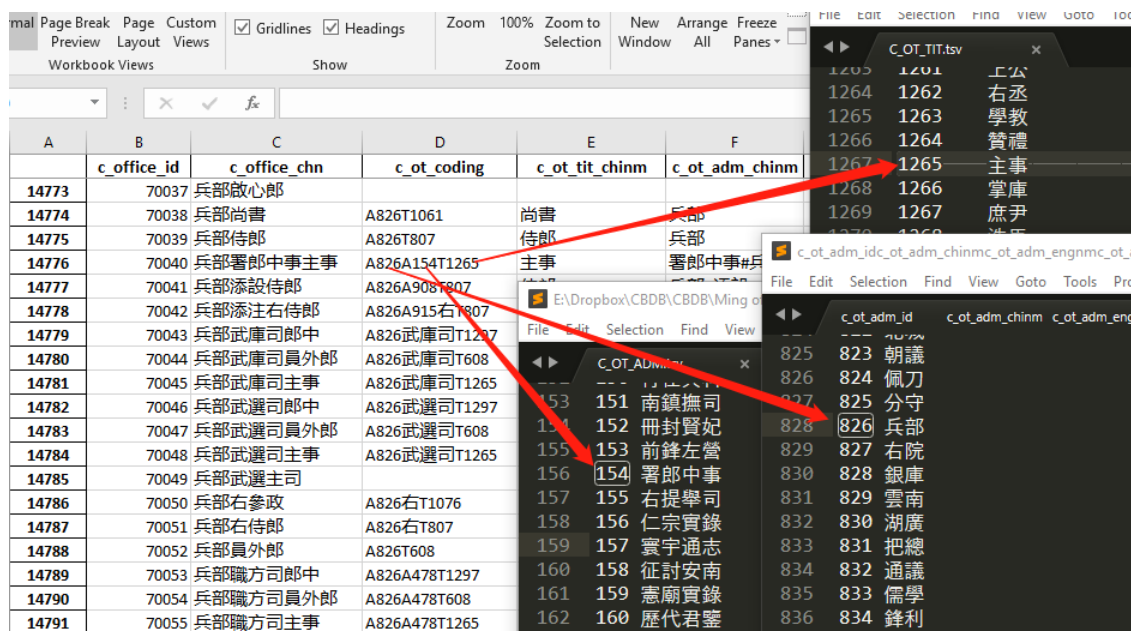Cc: "Michael A. Fuller (mafuller@uci.edu)" <mafuller@uci.edu>, "Wang, Hongsu" <hongsuwang@fas.harvard.edu>, 高策 <cegao@g.harvard.edu>, "Huang, Pengcheng" <pengchenghuang@fas.harvard.edu>

Dear Peter and Michael,

Sorry for not explaining the files and if I'm making things unnecessarily complicated. Please see my explanation below.

Columns in ming_office_title_coding.xlsx (as illustrated in the picture):

- c_office_id is from the OFFICE_CODES table.
- c_office_chn: is from the OFFICE_CODES table.
- c_ot_coding: coded values. "A" represents c_ot_adm table (C_OT_ADM.tsv file), i.e., administrative units, for example, A826==兵部. "T" represents c_ot_tit table (C_OT_TIT.tsv), i.e., titles, for example, T1265==主事.
- c_ot_tit_chinm: the coded title name, from the C_OT_TIT table.
- c_ot_adm_chinm: the coded administrative units, from the C_OT_ADM table.

An accurate coding relies on two prerequisites:

1. The title to be coded (i.e., c_office_chn) must be represented in a "standard and codable" form;
2. The dictionaries used to code the titles must be comprehensive.

More works need to be done for the two points above.

- For the first point, as Peter suggests, the "data cleaned" Excel sheet from UCI is very helpful because they have already cleaned many of, if not all of, the Ming office titles. For example, CBDB: c_office_id 70046, 兵部/武選司/郎中 --> UCI 中央中樞官署類 The Central Government, 六部門 Six Ministries, 兵部 The Ministry of War, /武選清吏司/郎中.
- For the second point, the dictionaries are still under developing. For example, 光禄寺少卿 is not successfully coded (i.e., A495少T1337) because of lacking "少卿" in the ADM dictionary/table (i.e., C_OT_ADM.tsv file).

At least for the above concerns, it should be safe to say we are on the right track. So for the following steps, if we all concur, we can do the following:

1. Use UCI's data to correct CBDB data. By saying "correct", it will be creating a new column(s) to integrate UCI's data, instead of deleting the "wrong" records in CBDB.
2. Analyze the values that are not successfully coded (the "irregularities") to see how we can further develop our dictionaries for coding.

The concern of "maintain the hierarchy and consistency": The c_ot_coding value is a structured string (like a VIN#), the office tree should be able to reconstruct from the string. We may have different opinions on this statement, but I think this point can be discussed at a later time after we have done the two steps above. How do you think?

Warm regards,
Ji

On Sun, Jun 3, 2018 at 2:26 AM Bol, Peter K. <peter_bol@harvard.edu> wrote:

> Dear Ji,
>
>
> There are a number of issues concerning office codes. At least at the moment I not persuaded by your alternative to the Fuller model, although I think we are all in agreement that we are trying both to maintain the hierarchy and consistency in coding names of offices and administrative units.
>
>
> -It would benefit others if you explained your work in the file itself.
>
> -Although I know you prefer to work in python, I think that if you make full use Excel formulas you will find that they are very powerful. Moreover it makes it easier for others to experiment with your work. In particular V-Lookup is often useful.
>
> -there are differences between how python works and CBDB standard procedures. E.g. in python counting begins with 0 as the first place but in CBDB coding 0 always means 未详。
>
> -the raw data from UCI, the dictionary, came with an explanation of fields and color coding. In the file Merged Ming Government Official Table 4 CBDB 2018-05-22_PKB_working.xlsx there is a sheet "data cleaned" which takes this information into account and should be used as the basis for creating codes.
>
>
> ming_office_title_coding.xlsx raises the issue of the difference between word division and titles. It seems to me that our office titles are not entirely consistent because the administrative units to which they belong vary.
>
> For example, 將軍 is an office title, as is 大將軍. But in some cases we would keep a longer title as the office title, even though it combines two or more words. E.g.
>
> | 征南副將軍 |
> |---|
> | 征夷副將軍 |
> | 征虜前將軍 |
> | 征虜副將軍 |
> | 征虜大將軍 |
>
> Because tshey are under (or classed as) 掛印將軍
>
>
> Similarly in

| 光祿寺卿 | A495T1337 | 卿 | 光祿寺 |
| 光祿寺少卿 | A495少T1337 | 卿 | 光祿寺 |

We would treat 卿 and 少卿 as two distinct titles. The fact that we can understand vice president as two words, does not mean the lowest level title should be president.

From: Michael Fuller [mailto:mafuller@uci.edu]
Sent: Sunday, June 03, 2018 2:03 AM
To: Ma, Ji <ji_ma@fas.harvard.edu>; Bol, Peter K. <peter_bol@harvard.edu>
Cc: Wang, Hongsu <hongsuwang@fas.harvard.edu>; Gao, Ce <cegao@g.harvard.edu>; Huang, Pengcheng <pengchenghuang@fas.harvard.edu>
Subject: RE: admin units and office reorganization

Dear Ma Ji,

I confess that I am not quite sure what I am looking at. We have the fields

c_office_id	c_office_chn		c_ot_coding	c_ot_tit_chinm	c_ot_adm_chinm
14764	70028	兵備僉事	兵備T965	僉事	
14765	70029	兵部參政	A826T1076	參政	兵部
14766	70030	兵部車駕司郎中	A826車駕司T1297	郎中	兵部
14767	70031	兵部車駕司員外郎	A826車駕司T608	員外郎	兵部
14768	70032	兵部車駕司主事	A826車駕司T1265	主事	兵部

What are the A826 and T1297 and what do they reference? What is the logic here? And what are those numbers in the far left?

Sincerely,

Michael

Sure, looking forward to your comments. Also, Michael, this must not something you exactly expect, my thought is, build something (has the potential to be right) out first, then we can have concrete examples to play with.

Best,

Ji

On Sat, Jun 2, 2018, 5:22 AM Bol, Peter K. <peter_bol@harvard.edu> wrote:

> Well，actually I am not sure this is on the right track. At least when Hongsu and I were looking at it we could not see what you were doing. There are also some historical questions about title/office division I should address. I will write more later.
>
> peter

Dear all,

We've run a quick experiment, looks like a promising direction, but still needs more work.

Path: \Ming official title Chn-Eng\data_output\ming_office_title_coding.xlsx (also contains other related files).

Best,

Ji

On Mon, May 14, 2018 at 9:55 AM, Bol, Peter K. <peter_bol@harvard.edu> wrote:

> See \Dropbox\CBDB\Ming official title Chn-Eng
> for
> Admin_category_&_Admin_unit_codes_draft_pkb

This fills in the first three tables according to the Ma-Fuller proposal.

We will soon have a list of office titles to populate the empty tables in my draft, and more admin units, which we will add to the admin unit tables. (the fact that the numbering will be out of sequence does not matter).

NOTE 1: Our principle for coding the names of offices and administrative units, and to the degree possible categories of administrative units, is that the same name should have a single code number no matter what dynasty it belongs to. Some names (e.g. the Qing banners) will be unique to a dynasty.

NOTE 2: in the UCI Dictionary the category of local military officials includes subcategories by the names of provinces and other local units. Because we register the place as part of a posting, rather than as part of an office title (thus reducing the number of office titles greatly), we have not followed the UCI Dictionary in this regard. However, for offices in the capital bureaucracy that include place names (in other words do not denote postings to local government) the place name is kept as part of the administrative unit.

NOTE 3: We need a table for Alt_names for office titles, not admin units

Peter K. Bol

Vice Provost for Advances in Learning

Wadsworth House

(Wadsworth House is the yellow colonial building in the corner of Harvard Yard, by the gate opposite the Smith Campus Center on Massachusetts Avenue. Please enter through the door marked "private" and follow the hallway to the right. Go up one flight of stairs and you will see #210 on the right).

Harvard University

617.495.8361

http://scholar.harvard.edu/pkbol/
viceprovost_learning@harvard.edu

Charles H. Carswell Professor of East Asian Languages and Civilizations

哈佛大學 東亞語言文明系 包弼德