Chinese Transactional Database: A User Manual

Brian Hazzard, Josh Loew

August 6, 2014

✧ ✧ ✧

This document is meant to be a user manual and guide for the Chinese Transactional Database. While the interface for the database is meant to be easy to use, an extra resource is often helpful to understand the functionality.

Any problems or questions not addressed in the manual may be sent to the Computer Science Systems Administrator: Tom Yanuklis at yanuklit@union.edu.

Acknowledgements

Contents

Rules and General Guidelines

This document describes the rules, guidelines, and user agreement for access to the China Transaction Trade Database ("the database") that Union College ("Union") has acquired. Described herein are restrictions and regulations on access, dissemination, and publication of works deriving from the database.

- Non-disclosure and Scope: Access to the database is limited to Union faculty, non-Union faculty working with Union faculty, and students approved by faculty (the faculty member approving a student supervise their access and use of the data). You shall not resell or distribute any of the data without written permission from World Trade Daily.

- Attribution: Union faculty, non-Union faculty working with Union faculty, and Union students shall have analytical, research, publication and co-publication rights to works derived from their access to the database; such works shall include attribution to the database.

- Minimum attribution: "General Administration of Customs. Chinese Customs Bureau Transactional Trade Database. Compiled by Robert Thompson. Beijing: Haiguan tongji bianji bu 200u. Database." Preferred attribution: "General Administration of Customs. Chinese Customs Bureau Transactional Trade Database. Beijing: Haiguan tongji bianji bu 200u. Database and Compiled by Robert Thompson, Editor-in-Chief, World Trade Daily, 2000 Mallory Lane Suite 130-244, Franklin, TN 37067 robert@worldtradedaily.com."

- Access to the servers containing the database shall be limited to the Union campus.

- Access from residential networking, guest wireless, and the public Internet is not permitted.

- Folders and files generated in your work will use applicable file and directory permissions to prevent access by other users.

- Acknowledgment: By my continued use of this interface, I agree to follow all the rules and guidelines as presented, in addition to the general application of common sense.

History of the Database

In the spring of 2013, Union College purchased a large transactional trade database from a vendor in China. The database covers a range of a substantial number of Chinese import and export data from 2000 to 2006. This data includes relevant trade data on each transaction such as the quantity of the item purchased, the company and its credentials, a unique product identification number, and several other attributes. When first received, the database was over 100 gigabytes in size as a result of poor organization and formatting. Because of its size and format, the database is difficult to traverse without sufficient hardware and an understanding of data models and database design, leading to a need for a user manual.

The goal is to make this data accessible to both students and faculty, chiefly by providing an interface and evaluation tools that would shield users from the underlying complexities of the database, and provide for tools that extract specific information.

Exploring the Interface

At the top of the page, there are six clickable tabs which organize the interface of the site.

3.1 Query

Figure 3.1: The Query Tab

The first of these tabs is the the *Query* tab, where the user inputs their search parameters to specify the information that they want from the database. Note: Some values may be input as a comma separated list.

3.2 Company Search

Figure 3.2: The Company Search Tab

The second tab is where the user may search for specific country names to input into their query. It leads them to a results page where the user can take the Country name and copy and paste it to search.

3.3 Country Search

Figure 3.3: The Country Search Tab

The third tab is where the user may search for specific company IDs to input into their query. It leads them to a results page where the user can take the company ID and copy and paste it to search.

Figure 3.4: The Queue Tab

3.4 Queue

This is where users may view the current job queue, and see roughly when their job will be run.

3.5 Downloads

Figure 3.5: The Downloads Tab

This is where the users may view completed jobs, and download them to their computer.

3.6 Help

Figure 3.6: The Help Tab

A quick reference page where users can get some tips thanks to a small FAQ.

Querying the Database

4.1 Query

In order to retrieve data from the database, it's vital to know what exactly to input as parameters.
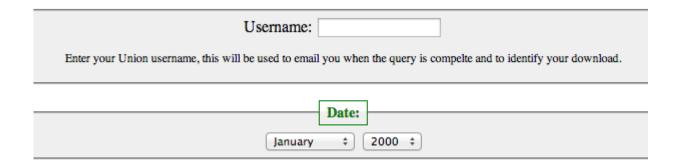

Figure 4.1: The first two fields: username and date

Here the user should put in their Union username so that the system knows who to email the results to. Please input usernames in the style of hazzardr and **NOT** hazardr@union.edu.

The date field is the month that you would like information for. Just select the month and country you wish to query.
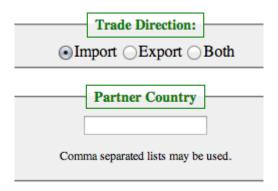

Figure 4.2: The third and fourth fields: Trade Direction and Partner Country

In the trade direction field, the user should specify whether they would like to view imports, exports, or both. If the user selects both, there will be identifiers for each element they are searching for differentiating between imports and exports.

Partner country is the country that China is trading with, but there are a few subtleties to be explained. In the result file of the query, there are two fields to note: acquiringCountry and destinationCountry. If acquiringCountry and destinationCountry are the same, then it is a simple direct trade from one country to another. However, there are cases where the two fields are not the same. If the two countries aren't the same:

- If the trade direction is exports, the "acquiringCountry" field describes a country or city that is a "middle-man" between two countries. The destinationCountry is the ultimate destination point of the exported good.

- If the trade direction is imports, the "acquiringCountry" field describes a country or city that is like a "middle-man" between two countries. The destinationCountry is the original country that China imported the good from.
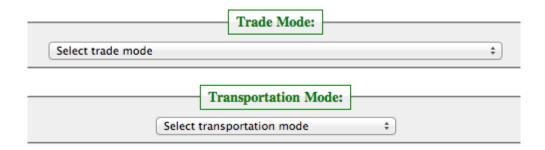

Figure 4.3: The fifth and sixth fields: Trade and Transportation Mode

Trade mode describes the context of a trade: is it a barter trade, a border trade, etc.? This field is the way that the good is traded.

Transportation mode describes the way the the product is being transported. The most common are rail, sea, or air.
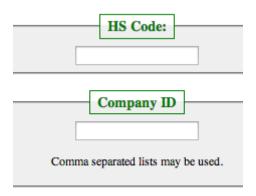

Figure 4.4: The seventh and eigth fields: HS code and Company ID

HS code stands for *Harmonized System* code, and is a specific product descriptor. They are available online. See here.

Company ID can be found using the search tab at the top of the page.

This field is the type of company, it describes whether it is a privately owned, publicly owned, or something else.
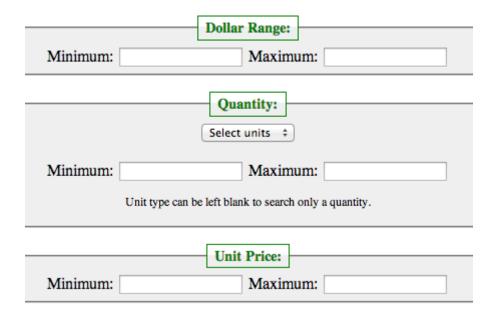
Figure 4.5: The ninth field: Company Type

Figure 4.6: The 10th, 11th, and 12th fields: Dollar Range, Quantity, and Unit Price

Dollar Range describes a range between two US dollar amounts that the user wishes to examine.

Quantity simply describes the quantity range the the user would like to examine

Unit price is a normalized way of examining a range of prices.

Figure 4.7: The last field: Limit Results

In this field, the user can input a number to limit the amount the results to so that the file size is more manageable.

4.2 Company Code

Now that we have an understanding as to what is going inside the search parameters, how do we get this information? While some of it may be the question that you're asking, there is definitely information that every user does not know off the top of their head. For example, many researchers want

to know information about a specific company or country's relationship with China. Understanding how to input these parameters into the database is vital for information retrieval. The process for retrieving both Company and Country Codes are virtually identical.

To get the company code to put into the query on the main page, it is necessary to navigate to the *Company Search* tab at the top of the page.

Figure 4.8: Company Search in the Navigation Bar

Once on the *Company Search* page, the user should type in either the name, or part of the name of the company that they wish to search for, along with the year of the data that they are searching. Then, simply hit the submit button.

Figure 4.9: The name may differ from year to year, so it is important to select the year you will be querying!

The user will then be brought to the results page where they can choose whether to select one or all of the companies by using the check boxes on the left, or the *SELECT ALL* button located above the data table. The company ID codes will be automatically generated in the box below the results table.
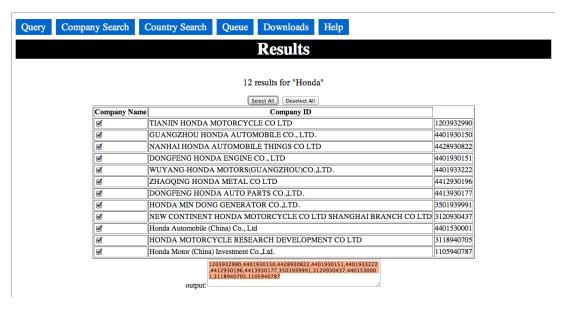

Figure 4.10: Select All button located above results taable

Once the correct company(ies) are selected, the user should hover over the text box and copy the codes to clipboard by either right clicking and selecting "Copy", or using *Control-c* for Windows users, or *Command-c* for Mac users.
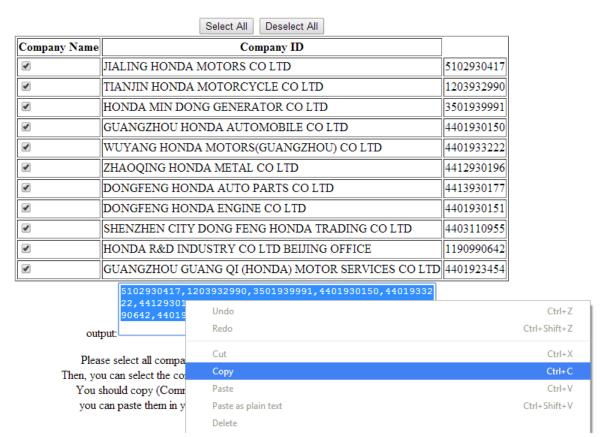
Figure 4.11: The result IDs are being highlighted

Once the information is copied to the clipboard, the user can easily paste it into the correct box for the search parameters, and submit their query!

4.3   Country Code

Finding Country Code is virtually the same process as finding Company Code. Simply repeat the previous steps, but for the Country Search tab.

Job Management and File Retrieval

5.1 Queuing

Once the query has been submitted, all running and queued jobs are listed under the *Query* tab found at the top of the page inside the navigation bar.

Figure 5.1: Query in the Navigation Bar

The figure below is a screenshot of how the queue will appear when there is one job running, and two jobs queued for completion.
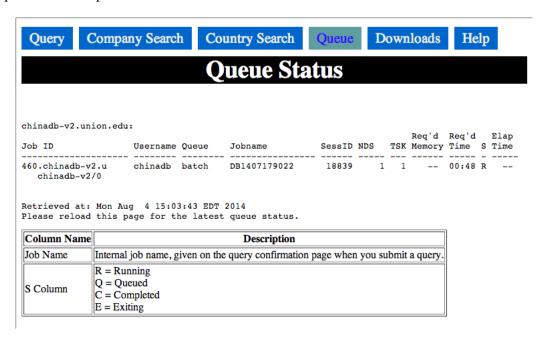

Figure 5.2: The Queue

The output above can be daunting at first, but there are only a few things of significance. Under the *Job* column, there is a number. That number increments with each job submitted to the server to track how many queries have been submitted. The next important column is the Status (*S*) column. It denotes what the current status of all jobs is. Possible statuses include Running (*R*), Queued (*Q*), Completed (*C*), and Exiting (*E*). A job that is running is one that is currently being processed by the server. A job that is queued is one that will run once all jobs above it are out of the queue. A completed job typically will not appear in the queue, but when completed, a job is removed from this output,

and all pending jobs are moved up in the queue. Lastly, and exiting job is one that has encountered a problem and has had to exit. This is an unusual circumstance, and will rarely be seen. A job may be forced to exit if it exceeds the allocated runtime for a job (the maximum running time is 48 hours). The final column is elapsed time of the running job. It prints (in HH:MM format) the time that the current job has been running. In order to refresh the output of the queue, simply refresh the page.

There are two methods of retrieving the output of a completed job. The first one is via email. On the *Query* page, there is a field for username. When the submitted job completes, an email is sent to example@union.edu That email includes a download link to the output. Competed jobs also appear in the interface under the *Downloads* tab in the navigation bar. Both methods of retrieval link to a .zip file, which will contain a .csv (comma separated value) file, a .sql (sequel) file, and a .txt (text) file. The .csv can be opened with a spreadsheet client such as Microsoft Excel. The other two files are for debugging purposes and unless there is an issue with the output, should be ignored.

It is important to note that there are two circumstances where a completed job will be removed from the server. The first, and more common circumstance is when the job is 24+ hours old. A process searches the downloads directory every hour, and deletes files that are over 24 hours old. The second circumstance is when the server becomes short on space. Although it likely will not happen frequently, it is possible to loose your job before the 24+ hour period is up.
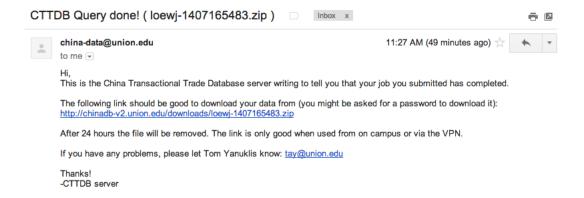

Figure 5.3: E-mail notification

Sample Query and Walk-Through

6.1 Data Selection

This chapter is intended to provide an example query to demonstrate the capability of some of the search features implemented in the system. In the demonstration, we will be searching for Honda products traded in March of 2006.

The process begins at the *Query* page. After entering a valid Union College username example@union.edu, proceed to enter the date of the desired transaction. In this case, that will be March of 2006. Next, select *Both* under the *Trade Direction* field.

✧ ✧ ✧

IMPORTANT NOTE: The only *required* fields in order to submit a query are Username, Date, and Trade Direction. All others are optional, though they may greatly help you specify the information that you want.
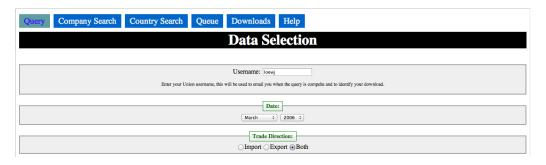
✧ ✧ ✧

Figure 6.1: Beginning of Data Selection Page

The next field to fill out is *Company ID*. In order to find Honda's company ID, navigate to the *Company Search* tab in the navigation bar. Enter *Honda* in the text box, select *2006* in the drop down menu, and press the *Submit* button.
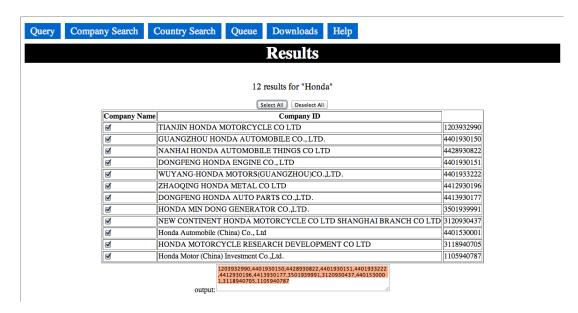
Figure 6.2: Company Search Tab

Figure 6.3: Company Search Results

6.2 Data Input

The results above are all company IDs with the word *Honda* in them. These are the desired results, so press the *Select All* button on the top of the page below the navigation bar. This will check all boxes (each time a box is checked, that value is appended or pulled from the *output* field accordingly). Now, mouse over the *output* box to select all contents. Either right-click and select *Copy* or Mac users can press command-c (control-c on Windows) on their keyboard to copy the text to the clipboard. Navigate back to the *Company ID* field on the *Query* page. Right click and select *Paste* or Mac users can press command-v (control-v on Windows) on their keyboard to paste the text from the clipboard.

Figure 6.4: Company ID Field

For the demonstration, limit the results of the search to 1000 on the main interface page and press the *Query* button.

Figure 6.5: Limit Results and Submit

6.3   Data Retrieval

The job appears is the only job in the queue, and its running status is denoted by the **R** in the status column.
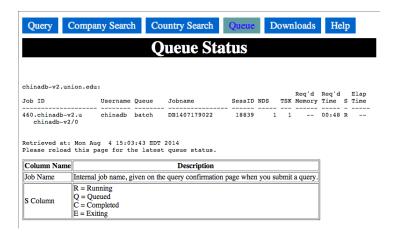

Figure 6.6: The *Queue* Tab

Once completed, an email notification will be sent. To download to the results, go to the *Downloads* page and find the desired completed job based on the username and date/time submitted.

Figure 6.7: The *Downloads* Tab

Once downloaded, unzip the file. Right-click the .csv (comma separated value) file and select "Open With." Select any spreadsheet client (such as Microsoft Excel) to view the file. Done!
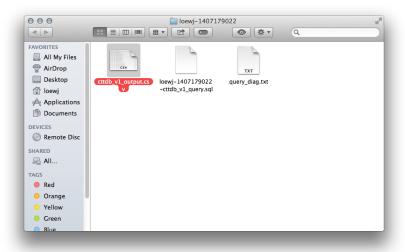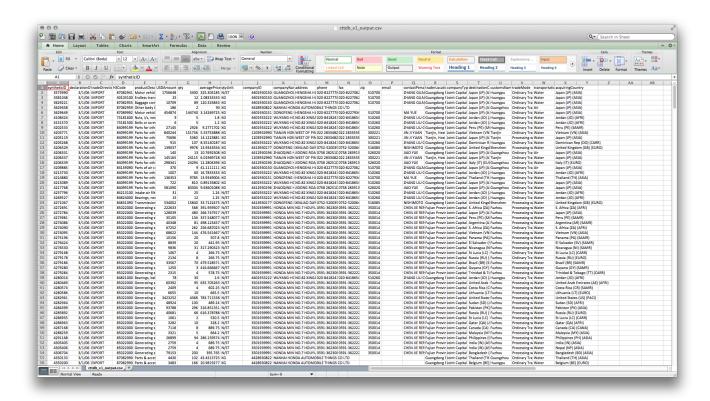
Figure 6.8: Contents of the zip file

Figure 6.9: Result File

Additional Information

If you're having problems with the system, please send an email to yanuklit@union.edu. Please try to be as descriptive as you can in your email.

7.1 Quick Help

- Queries: Only 1 query at a time runs from the run queue, but there can be many queries submitted without a problem. Please wait about 2 seconds before submitting you next query. Downloads are available for 24 hours, if the server gets full, they may be removed before 24 hours. The query result is a .zip file, containing a CSV file of your output (it also contains the SQL code generated for the query and a diagnostic file).

- Country & Company select: This searches based on partial or complete match for either. If you leave the field blank you get all the entries for countries or companies.

- Queue: shows you the current status of the batch queue of running jobs/jobs waiting to run.

- Downloads: Shows you the downloads that are currently available.