Big Data & International Trade: Creating Transparency through Information

P.G. Swartz
peter@panjiva.com

Panjiva, Inc.
www.panjiva.com

K.W. Soura
keith@panjiva.com
Overview

- Importance of global trade
- Promise of trade data
- Technical challenges and solutions
- Business applications
- Our take-aways
The Concept
Global Trade Touches All Countries
Global Trade Is Big And Growing
A Map of this Network is Very Useful

- Manufacturers/Retailers
- Logistics companies
- Governments
- Finance
The Opportunity
Valuable Information, if Extractable

- Hugely valuable information buried in a mountain of paperwork
- Complex system of privacy laws that impede acquisition or dissemination
Panjiva: Mapping the Trade Network

- **8 governments**
  - US, China, Mexico, Chile, Colombia, Panama, Peru, Uruguay
- **190 countries**
- **450 million shipping records**
- **5 million companies**

<table>
<thead>
<tr>
<th>Arrival Date</th>
<th>Raw Shipper</th>
<th>Raw Consignee</th>
<th>US Port Of Unloading</th>
<th>Container Info</th>
<th>Commodity Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-08-28</td>
<td>SAMSUNG ELECTRONICS CO., LTD 129, SAMSUNG-RO, YEONGTONG-GU SUIWON 443-742 KYONGGI-DO SOUTH KOREA</td>
<td>SAMSUNG ELECTRONICS AMERICA, INC. AMERICAN DISTRIBUTION CENTER 4921 DIRECTORS PLACE SUITE 200 SAN DIEGO, CA 92121 TB858-642-5100</td>
<td>Port of Long Beach, Long Beach, California</td>
<td>REFRIGERATOR</td>
<td></td>
</tr>
<tr>
<td>2014-08-28</td>
<td>BEAUTIFUL WINDOW ENTERPRISE CO., LTD NO.168, HU-FU-HUANG RD LIUHE TAICANG JIANGSU CHINA</td>
<td>WESTERN REFLECTIONS, LLC 1385 CHEERS ST, SUITE BROWNSVILLE, TX 78521 BROWNSVILLE TX USA</td>
<td>Port of Long Beach, Long Beach, California</td>
<td>BLINDS</td>
<td></td>
</tr>
</tbody>
</table>
Raw Data is Raw

- A range of data partners
  - Wide variety of data formats
  - Wide variety of regulations and rules
- Messy data:
  - Few unique identifiers
  - A pathological aversion to spell check
  - Records in multiple languages
Cleaning Raw Data

- Receive data
  - CSV, PDF, arcane formats
- Upload to raw data to database
  - Store efficiently
- Structure raw text
  - Geographic locations of parties
  - Ports
  - Units
Clustering Raw Data into Profiles

- Machine-learning based comparison of shipment records
- Intelligent feature extraction
- Matching to other data sets

<table>
<thead>
<tr>
<th>Notify Parties</th>
<th>Raw Shipper</th>
<th>Raw Consignee</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLANTIC TRADING AND MARKETING INC</td>
<td>PMI COMERCIO INTERNATIONAL SA DE CV</td>
<td>ATLANTIC TRADING AND MARKETING INC</td>
</tr>
<tr>
<td>5847 SAN FELIPE,STE.2100</td>
<td>AV MARINA NACIONAL NO 329</td>
<td>5847 SAN FELIPE,STE.2100</td>
</tr>
<tr>
<td>HOUSTON, TX 77057 US</td>
<td>TORRE EJECUTIVA COL.HUASTEC</td>
<td>HOUSTON, TX 77057 US</td>
</tr>
<tr>
<td>TOTAL PETROCHEMICALS &amp; REFINING USA</td>
<td>11311 MEXICO CIUDAD, MEXICO</td>
<td></td>
</tr>
<tr>
<td>INC 1201 LOUISIANA STREET SUITE 1800</td>
<td>CIUDAD 11311 MX</td>
<td></td>
</tr>
<tr>
<td>HOUSTON TX 77002 US</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Technical Take Away
Generalize from Domain Expertise

- Data Science is great
- Combined with domain expertise: even better
- Research general ways of understanding the domain
- Example: address similarity
  - Semantic understanding
  - Intelligent robust features
The Platform
Actionable Data

- Actionable data = data must fit client use profiles:
  - contact information
  - corporate data
  - financial reports
  - product information
Delivery to Clients

- Subscription website
- Customer Relationship Management tool
- Application Programming Interface (API)
- Custom Solutions
Made-to-Measure Tools
The Business Cases
Simple Case: Supply Chain

- **Problem:** I need to find / vet a new trading partner.
Zhongde Equipment Co., Ltd. Shandong

Summary | Products | Shipments | Chinese Exports | Customers | Corporate | Contact

Products

See more product information

Shipments

126
Shipments sent

TOP PRODUCT TERMS

- beer
- brewing
- beer brewing
- brewing equipment
- beer brewing equipment
Problem: I need to optimize shipment flows between ports / trade lanes.
<table>
<thead>
<tr>
<th>Foreign Ports</th>
<th>Shipments</th>
<th>Port</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trendline</td>
<td>3,316,390</td>
<td>Shanghai</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>2,724,605</td>
<td>Yantian</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>1,984,695</td>
<td>Kaohsiung</td>
<td>Taiwan</td>
</tr>
<tr>
<td></td>
<td>1,544,130</td>
<td>Hong Kong</td>
<td>Hong Kong</td>
</tr>
<tr>
<td></td>
<td>1,427,792</td>
<td>Singapore</td>
<td>Singapore</td>
</tr>
<tr>
<td></td>
<td>1,192,303</td>
<td>Ningpo</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>1,189,623</td>
<td>Busan</td>
<td>South Korea</td>
</tr>
<tr>
<td></td>
<td>754,459</td>
<td>Tsingtao</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>388,212</td>
<td>Keelung</td>
<td>Taiwan</td>
</tr>
<tr>
<td></td>
<td>360,967</td>
<td>Laem Chabang</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>339,182</td>
<td>Yung Tau</td>
<td>Vietnam</td>
</tr>
<tr>
<td></td>
<td>308,192</td>
<td>Xiamen</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>249,702</td>
<td>Tokyo</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>237,383</td>
<td>Chiwan</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>221,676</td>
<td>Kobe</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>217,875</td>
<td>Dagu/Tanggu</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>212,822</td>
<td>Yokohama</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>171,629</td>
<td>Nagoya</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>144,759</td>
<td>Manzanillo</td>
<td>Panama</td>
</tr>
<tr>
<td></td>
<td>115,710</td>
<td>Bremerhaven</td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td>96,529</td>
<td>Taipei</td>
<td>Taiwan</td>
</tr>
<tr>
<td></td>
<td>92,429</td>
<td>Tanjung Pelepas</td>
<td>Malaysia</td>
</tr>
<tr>
<td></td>
<td>72,144</td>
<td>Rotterdam</td>
<td>Netherlands</td>
</tr>
<tr>
<td></td>
<td>71,829</td>
<td>Lazaro Cardenas</td>
<td>Mexico</td>
</tr>
<tr>
<td></td>
<td>68,267</td>
<td>All Other Peoples Republic Of China Ports</td>
<td>China</td>
</tr>
<tr>
<td>Carrier SCAC</td>
<td>Address</td>
<td>Phone</td>
<td>Matching Shipments</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------</td>
<td>----------</td>
<td>--------------------</td>
</tr>
<tr>
<td>APLU - Apil Co Pte Ltd Noi Group</td>
<td>456 ALEXANDRIA RD NOL BLDO #06-00 Singapore</td>
<td></td>
<td>1,621,600 shipments</td>
</tr>
<tr>
<td>MAEU - Maersk Line</td>
<td>9300 ARROWPOINT BLVD CHARLOTTE North Carolina United States 28273</td>
<td>7045712608</td>
<td>1,236,886 shipments 0.46%</td>
</tr>
<tr>
<td>EGLV - Evergreen Line</td>
<td>EVERGREEN SHIPPING AGENCY (AMERICA) CORP. JERSEY CITY New Jersey United States 07302</td>
<td>2037912901</td>
<td>1,090,291 shipments 7.11%</td>
</tr>
</tbody>
</table>
Advanced Case: Governments

- **Problem:** I need to find anomalies in shipment data at the local, state, regional, port, or national level.
The Business Take Away
Generalize from Domain Expertise

- Understand the exact nature of needs
- Understand these needs as examples of a class
- Construct front-end and back-end infrastructure to solve as many needs as possible
The Future
What’s Next?

● More governments
● Higher quality data
● More tools for the wide range of clients
● Custom(s) services
Want Additional Information?

- API access
- Custom services
- Other questions
- Contact us!
  - Peter G. Swartz, Data Scientist
    - peter@panjiva.com
  - Keith W. Soura, Technical Account Manager
    - keith@panjiva.com
Big Data & International Trade: Creating Transparency through Information

For API Access, send an e-mail to: keith@panjiva.com

P.G. Swartz
peter@panjiva.com

Panjiva, Inc.
www.panjiva.com

K.W. Soura
keith@panjiva.com