Esteban Serrano (s1885916@ed.ac.uk)
School of Informatics, University of Edinburgh

Oliver Ford (s1424408@ed.ac.uk)
School of Informatics, University of Edinburgh

Xinyu Du (s1839129@ed.ac.uk)
School of Informatics, University of Edinburgh

Anouk Lang (anouk.lang@ed.ac.uk)
School of Literatures, Languages & Cultures,
University of Edinburgh

Using Data Visualization to Explore International Trade Agreements

ollieford.github.io/DS4D-Trade-Agreement-Project/

```xml
<?xml version="1.0" encoding="UTF-8"?>
<treaty>
  <meta>
    <name>Central American Free Trade Area</name>
    <type>Free Trade Agreement</type>
    <treaty_identifier>188</treaty_identifier>
    <wto_rta_id>214</wto_rta_id>
    <status>Inactive</status>
    <notification>GATT Art. XXIV</notification>
    <date_signed>1958-06-10</date_signed>
    <date_into_force>1959-06-02</date_into_force>
    <date_notification>1956-08-22</date_notification>
    <end_implementation/>
    <date_inactive>1961-10-12</date_inactive>
    <parties_original>
      <partyisocode n="1"><CRI></partyisocode>
      <partyisocode n="2"><GTM></partyisocode>
      <partyisocode n="3"><HND></partyisocode>
      <partyisocode n="4"><NIC></partyisocode>
      <partyisocode n="5"><SLV></partyisocode>
    </parties_original>
  <parties>
</treaty>
```

This project uses **data visualization** to gain insights into the underlying **structure and relationships** of a corpus of international trade agreements gathered and structured into XML format by the **ToTA: Texts of Trade Agreements** project (Alschner et al. 2017; see github.com/mappingtreaties/tota). The dataset consists of 450 preferential international trade agreements, 424 of which are in English.

Research Questions:
1) What can be learnt about the **relationships between countries** by analysing the legal documents?
2) What might this reveal about the **landscape of international economic governance**?
3) What can be learnt about the **relationships between the trade agreements** themselves? For example, do **previous** trade agreements impact **later** ones?

Word Embeddings

Word embeddings is a machine learning technique for representing words based on the context in which they appear. It is capable of capturing a measure of similarity between words by associating those that appear frequently. In the images below, each dot represents one trade agreement: the closer two nodes are, the more similar they are. Distances between documents were calculated using an algorithm called Word Mover Distance (WMD). The model we used was pretrained on Wikipedia documents (glove-wiki-gigaword-300). Optimal 2D dot layout was achieved by using the Neato weighted springs algorithm.

Two main clusters appear clearly in the center with distinct similarities between each other: Free Trade Agreements and Free Trade Agreement and Economic Integration Agreements.
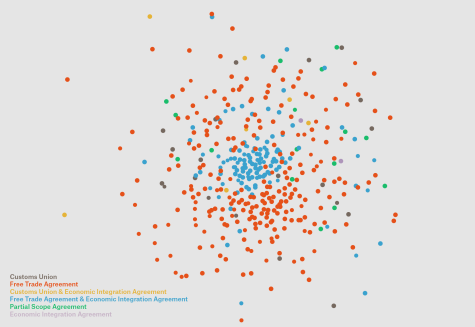
The images below shows the different types of trade agreements visualized by their relative distances (scales are consistent between the five graphs).

From these we see that there is a fair amount of overlap between individual documents in the categories of Free Trade Agreements and Free Trade Agreement and Economic Integration Agreements. The countries of North America, Asia Oceania use one set of templates to put together Free Trade Agreement and Economic Integration Agreements, while the countries of Europe use a different set of templates to draw up Free Trade Agreements, and this is captured by both the word embeddings and the topic modelling visualizations. Customs Unions, Customs Union & Economic Integration Agreements and Partial Scope Agreements documents are not so dependent on templates, something which is conveyed more clearly by the word embeddings plots above than the topic modelling plots to the right.

Directions for Future Work

As time has gone on, international trade agreements have had more chapters added to them (for example on intellectual property, e-commerce etc). Similarity and difference in these agreements therefore emerges not only in the content of chapters across agreements, but also from the fact that new chapters are gradually added. One next step, then, would be to separate the agreements into chapters in order to obtain greater granularity at the chapter level.

Scholarship on international trade is attentive to the influence of the EU and the US. But it would also be interesting to single out other countries such as China who are important actors, and to track their influence on the shape of the law in this area. The static Trade Agreements Clustered by Topic visualization, for instance, suggests that the East Asian countries, situated as they are between the long blue cluster and the long red cluster, might be conceptualized as being pulled in two directions – between the Americas and Oceania on the one hand, and Europe on the other.

Every trade agreement is between at two parties, and sometimes two or more regions, each of which comes to the negotiating table with its own ideas. Developing a trade agreement 'fingerprint' for each country and/or region would allow researchers to discern which country had, statistically speaking, been more active in 'calling the shots' in the writing of any single agreement. How much of one country's approach had informed a particular agreement, and how much ground had they yielded to the other side in its construction?

Topic Modelling

Topic modelling is a method which analyses the statistical distribution of words in a set of documents so as to identify the latent topics or themes across those documents. Each point in the images below represents a trade agreement between one or more countries, with the colour representing the type of trade agreement. As with the word embeddings visualizations, the closer two dots are, the more similar they are (with similarity here calculated on the extent to which the documents share similar topics). The average year the treaties were signed for each cluster is shown in grey, and the regions the signatories belong to is shown in black.
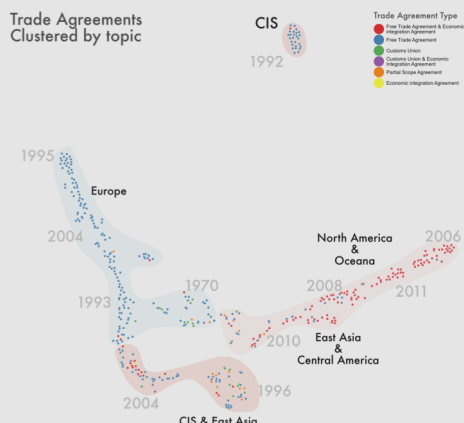
The different types of trade agreements can be separated into clusters along continental lines. European countries tend to use Free Trade Agreements, while the countries of North America and East Asia tend to use Free Trade and Economic Integration Agreements. The third cluster at the top shows how in the years following the break-up of the Soviet Union, the countries of the then-Commonwealth of Independent States set about establishing trade relations with one another using mostly Free Trade Agreements.
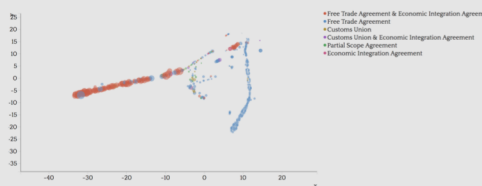
In an interactive version of this data, available at ollieford.github.io/DS4D-Trade-Agreement-Project/Topic1.html, colour again represents the agreement type, while point size represents the length of the treaty. Users can zoom in on different clusters and mouseover to obtain details of the signatories and the year, as shown in the two screenshots below. The same data as above is used, though the clusters are slightly different due to random fluctuations in the clustering algorithm.
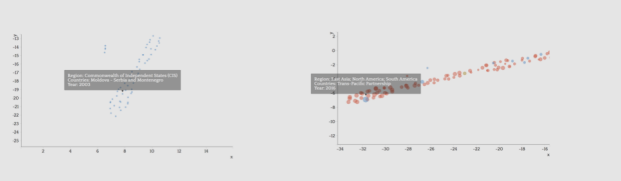
References

Alschner, Wolfgang, et al. 'Text of Trade Agreements (ToTA): A Structured Corpus for the Text-as-Data Analysis of Preferential Trade Agreements'. *Journal of Empirical Legal Studies*, vol. 15, no. 3, Sept. 2018, pp. 648–66. *Wiley Online Library*, doi:10.1111/jels.12189.
———. *Text-as-Data Analysis of Preferential Trade Agreements: Mapping the PTA Landscape*. UNCTAD Research Paper No. 5, UNCTAD/SER.RP/2017/5, *United Nations Conference on Trade and Development*, July 2017, pp. 1–29, https://unctad.org/en/PublicationsLibrary/ser_rp2017d5_en.pdf.
Kusner, Matt J., et al. 'From Word Embeddings to Document Distances'. *Proceedings of the 32nd International Conference on International Conference on Machine Learning – Volume 37*, 2015, pp. 957–966, *ACM Digital Library*, http://dl.acm.org/citation.cfm?id=3045118.3045221.
Pele, Ofir, and Michael Werman. 'A Linear Time Histogram Metric for Improved SIFT Matching'. *Computer Vision: ECCV 2008* edited by David Forsyth et al, 2008, pp. 495–508.
———. 'Fast and Robust Earth Mover's Distances'. *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 460–67. *Semantic Scholar*, doi:10.1109/ICCV.2009.5459199.