Retracing the Travel Path of Marco Polo

Master’s Defence

10th May, 2021
Close reading of historical texts, take researchers a lifetime to explore and analysis... in a traditional way.
Retracing travel path from historical travelogue

- 12th century travelogue of Italian explorer Marco Polo
- Narrates his own travels through Asia and exploration of China between 1271 and 1295
- It is written by Rustichello da Pisa in Franco-Italian
- English translations used in this thesis are:
  - Hugh Murray: For text and the book Index
  - Henry Yule
Retracing travel path from historical travelogue

- 12th century travelogue of Italian explorer Marco Polo
- Narrates his own travels through Asia and exploration of China between 1271 and 1295
- It is written by Rustichello da Pisa in Franco - Italian
- English translations used in this thesis are:
  - Hugh Murray: For text and the book Index
  - Henry Yule: For the book index
Visualizing the travel path

<table>
<thead>
<tr>
<th>PART</th>
<th>Travel Route</th>
</tr>
</thead>
<tbody>
<tr>
<td>PART I</td>
<td>Within China</td>
</tr>
<tr>
<td>PART II</td>
<td>Venice -&gt; China</td>
</tr>
<tr>
<td>PART III</td>
<td>China -&gt; Venice</td>
</tr>
</tbody>
</table>
Visualizing the travel path

<table>
<thead>
<tr>
<th>PART</th>
<th>Travel Route</th>
</tr>
</thead>
<tbody>
<tr>
<td>PART I</td>
<td>Within China</td>
</tr>
<tr>
<td>PART II</td>
<td>Venice -&gt; China</td>
</tr>
<tr>
<td>PART III</td>
<td>China -&gt; Venice</td>
</tr>
</tbody>
</table>
Visualizing the travel path

<table>
<thead>
<tr>
<th>PART</th>
<th>Travel Route</th>
</tr>
</thead>
<tbody>
<tr>
<td>PART I</td>
<td>Within China</td>
</tr>
<tr>
<td>PART II</td>
<td>Venice -&gt; China</td>
</tr>
<tr>
<td>PART III</td>
<td>China -&gt; Venice</td>
</tr>
</tbody>
</table>
Visualizing the travel path

<table>
<thead>
<tr>
<th>PART</th>
<th>Travel Route</th>
</tr>
</thead>
<tbody>
<tr>
<td>PART I</td>
<td>Within China</td>
</tr>
<tr>
<td>PART II</td>
<td>Venice -&gt; China</td>
</tr>
<tr>
<td>PART III</td>
<td>China -&gt; Venice</td>
</tr>
</tbody>
</table>
Goal: Retracing the travel path of Marco Polo
General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

KAIN-DU is a western province, which was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du. Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round.
General challenges with historical texts

- OCR errors
- Different naming conventions
  - Spelling variations
  - Translation errors
  - Linguistic variations
  - Syntax structures

KAIN-DU is a western province, which was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du. Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round. 3
General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

KAIN-DU is a western province, which was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du. Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round. 3
General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures
Challenges with the travelogue of Marco Polo

- Non-existing geographical entities
  *e.g.: Greater India and Lesser India*

- Geographical renaming
  *e.g.: “Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains.”*

- Change of boundaries
  *e.g.: “Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan”*

- Unnamed and descriptive place references
  *E.g.: Plain of Bargu*
Challenges with the travelogue of Marco Polo

- **Non-existing geographical entities**
  
  *e.g.: Greater India and Lesser India*

- **Geographical renaming**
  
  *e.g.: “Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains.”*

- **Change of boundaries**
  
  *e.g.: “Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan”*

- **Unnamed and descriptive place references**
  
  *E.g.: Plain of Bargu*
Challenges with the travelogue of Marco Polo

- **Non-existing geographical entities**
  
  *e.g.: Greater India and Lesser India*

- **Geographical renaming**
  
  *e.g.: “Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains.”*

- **Change of boundaries**
  
  *e.g.: “Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan”*

- **Unnamed and descriptive place references**
  
  *E.g.: Plain of Bargu*
Challenges with the travelogue of Marco Polo

● Non-existing geographical entities
  e.g.: Greater India and Lesser India

● Geographical renaming
  e.g.: “Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains.”

● Change of boundaries
  e.g.: “Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan”

● Unnamed and descriptive place references
  E.g.: Plain of Bargu
Related Work

- A study by Barbaresi\(^9\) retraces the path from travel literature
  - Dataset:
    - Travel journals from China (1907)
    - Die Fackel (1899 - 1936)
  - Approach:
    - Combines coordinates, sequence and sense of time

---

Data Preparation
Text preparation

Scanned PDF
Text preparation

Scanned PDF ➔ Textual representation

OCR

Ruta Hareshbhai Bakhda
Text preparation

Scanned PDF → OCR → Textual representation → Manually → Spelling correction
Text preparation

Scanned PDF → OCR → Textual representation → Manually → Spelling correction → Section Splitting
Text preparation
Index Preparation

Three different versions:

- Index from the translation by Hugh Murray
- Index from the translation by Henry Yule
- Combined index from Murray and Yule translations
Index Preparation

Three different versions:

- Index from the translation by Hugh Murray
- Index from the translation by Henry Yule
- Combined index from Murray and Yule translations
Index Preparation

“ABASCIA (Abyssinia), kingdom of, 324. The inhabitants converted by St Thomas, 325. Its king defated the ruler of Adel (Aden), 326. Productions of the country, 327. Abraiamain (Bramins), order of, 293, 304-308.”

<table>
<thead>
<tr>
<th>Entity</th>
<th>Alternative Name</th>
<th>References</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>kingdom of</td>
<td>324</td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>The inhabitants converted by St Thomas</td>
<td>325</td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Its king defated the ruler of Adel (Aden)</td>
<td>326</td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Productions of the country</td>
<td>327</td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Abyssinia &amp; Abraiamain (Bramins), order of</td>
<td>293, 304-308</td>
</tr>
</tbody>
</table>
What is a Gazetteer?

“Gazetteers are reference list...
....that are labeled,
relevant to the task”
# Gazetteer Preparation

Index + Tag (Entity type) -> Gazetteer

<table>
<thead>
<tr>
<th>Entity</th>
<th>Alternative Name</th>
<th>References</th>
<th>Page No.</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>kingdom of</td>
<td>324</td>
<td>Location</td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>The inhabitants converted by St Thomas</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Its king defated the ruler of Adel (Aden)</td>
<td>326</td>
<td></td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Productions of the country</td>
<td>327</td>
<td></td>
</tr>
<tr>
<td>ABASCIA</td>
<td>Abyssinia</td>
<td>Abyssinia &amp; Abraiamain (Bramins), order of</td>
<td>293,304-308</td>
<td></td>
</tr>
</tbody>
</table>
Need for gazetteers

- Tasks where entity extraction is challenging
- Scarcity of proper resources for particular domain specific knowledge
- To encode additional background knowledge

In case of Marco Polo narrative…

- Absence of universal tooling for historical texts
- Lack of common standards for gazetteers
Need for gazetteers

- Tasks where entity extraction is challenging
- Scarcity of proper resources for particular domain-specific knowledge
- To encode additional background knowledge

In case of Marco Polo narrative...

- Absence of universal tooling for historical texts
- Lack of common standards for gazetteers
## Ambiguity Resolution

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alau</td>
<td>Hookalu</td>
<td>Person</td>
</tr>
</tbody>
</table>

Gazetteer from Murray translation
## Ambiguity Resolution

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alau</td>
<td>Hookalu</td>
<td>Person</td>
</tr>
</tbody>
</table>

**Gazetteer from Murray translation**

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hukalu Khan</td>
<td>Alau, Hukalu</td>
<td>Person</td>
</tr>
</tbody>
</table>

**Gazetteer from Yule translation**
## Ambiguity Resolution

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alau</td>
<td>Hookalu</td>
<td>Person</td>
</tr>
</tbody>
</table>

**Gazetteer from Murray translation**

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hukalu Khan</td>
<td>Alau, Hukalu</td>
<td>Person</td>
</tr>
</tbody>
</table>

**Gazetteer from Yule translation**

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Alternative Name</th>
<th>Entity Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alau</td>
<td>Hookalu, Hukalu, Hukalu Khan</td>
<td>Person</td>
</tr>
</tbody>
</table>

**Combined gazetteer after ambiguity resolution**

Gold Standard Setup

Travelogue

Location

Motion Events Triggers

Travelled historical Location

Travelled modern Location

Travel path on a map
Gold Standard Setup

Methodology:
- Crucial for the results evaluation
- Validation performed twice on entire travelogue
- Validation with random sampling until errors come to near zero

Travelogue → Location

Location → Motion Events Triggers

Motion Events Triggers → Travellel historical Location → Travelled modern Location → Travel path on a map
Methodology
Having told you of that Khan, I will now go to Zanghibar.
Having tell you of that Khan, I will now go to Zanghibar.
Having told you of that Khan, I will now go to Zanghibar.
Having told you of that Khan, I will now go to Zanghibar.
Having tell you of that Khan, I will now go to Zanghibar.
Having tell you of that Khan, I will now go to Zanghibar.
Having tell you of that Khan, I will now go to Zanghibar.
What is Named Entity Recognition?

“Identifying words,
classifying them into predefined categories,
such as person, location, organization, etc.”
Identifying location entities

1. Application of pre-trained NER models
   ○ that have shown promising performance for standard corpora
   ○ that are widely and commonly used
   ○ that can be generalized to any data
   ○ that are representative of different approaches
Identifying location entities

1. Application of pre-trained NER models
   ○ that have shown promising performance for standard corpora
   ○ that are widely and commonly used
   ○ that can be generalized to any data
   ○ that are representative of different approaches

<table>
<thead>
<tr>
<th>NER model</th>
<th>Data</th>
<th>Model</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLTK NER¹</td>
<td>ACE 2004</td>
<td>MaxEnt classifier</td>
<td>0.89 ± 0.11</td>
</tr>
<tr>
<td>Stanford NER²</td>
<td>CoNLL, MUC-6, MUC-7 and ACE</td>
<td>CRF Classifier</td>
<td>87.94 %</td>
</tr>
<tr>
<td>spaCy NER³</td>
<td>OntoNotes</td>
<td>Multi-task CNN</td>
<td>85.85 %</td>
</tr>
<tr>
<td>AllenNLP NER⁴</td>
<td>CoNLL</td>
<td>ELMo</td>
<td>90.87 ± 0.13</td>
</tr>
</tbody>
</table>

1. [https://nlp.stanford.edu/software/CRF-NER.shtml]
2. [https://spacy.io/]
3. [https://www.nltk.org/book/]
4. [https://allennlp.org/]
Identifying location entities

2. Gazetteer for identifying location entities

<table>
<thead>
<tr>
<th>Through</th>
<th>Persia</th>
<th>and</th>
<th>Greater</th>
<th>Armenia</th>
<th>,</th>
<th>they</th>
<th>came</th>
<th>to</th>
<th>Trebizonda</th>
</tr>
</thead>
</table>

N = 1...3
Identifying location entities

2. Gazetteer for identifying location entities

If extracted N-gram exists in Gazetteer, check the entity type from gazetteer.
Having tell you of that Khan, I will now go to Zanghibar.
What is Event Extraction?

“Information extraction task, to extract specific knowledge, related to certain incidents / events”
Extracting motion event triggers

WordNet

Seed: DEPART, TRAVEL, ARRIVE

Verb

synset

verb synset

Semantic similarity (Wu-palmer)

> t

< t

Motion verb

Non-Motion verb

VerbNet

FrameNet

5. https://wordnet.princeton.edu/
6. https://verbs.colorado.edu/verbnet/
7. https://framenet.icsi.berkeley.edu/findupal/
Extracting motion event triggers

WordNet\(^5\)

- **Seed:** DEPART, TRAVEL, ARRIVE
- **Verb synset**
- **Semantic similarity (Wu-palmer)**
  - > \(t\)
  - < \(t\)

- **Motion verb**
- **Non-Motion verb**

VerbNet\(^6\)

- **Seed:** VN Class 51
- **VN Class**
  - = seed
  - != seed

- **Motion verb**
- **Non-Motion verb**

FrameNet\(^7\)

---

5. [https://wordnet.princeton.edu/](https://wordnet.princeton.edu/)
6. [https://verbs.colorado.edu/verbnet/](https://verbs.colorado.edu/verbnet/)
7. [https://framenet.icsi.berkeley.edu/findrupal/](https://framenet.icsi.berkeley.edu/findrupal/)
Extracting motion event triggers

WordNet\(^5\)

Seed: DEPART, TRAVEL, ARRIVE

Verb

synset

verb synset

Semantic similarity (Wu-palmer)

> t

< t

Motion verb

Non-Motion verb

VerbNet\(^6\)

Seed: VN Class 51

Verb

VN Class

= seed

!= seed

Motion verb

Non-Motion verb

FrameNet\(^7\)

Seed: Source, Goal, Path, Place, Direction

Verb

FN Frames

any in seed

! any in seed

Motion verb

Non-Motion verb

---

5. https://wordnet.princeton.edu/

6. https://verbs.colorado.edu/verbnet/

7. https://framenet.icsi.berkeley.edu/findrupal/
Same verb - different context

After climbing down a mountain, they descended on a plain area.
Same verb - different context

After climbing down a mountain, they descended on a plain area.

In this province there is a king named George, descended from that prince, and who indeed enjoy his power.
Having tell you of that Khan, I will now go to Zanghibar.
Location and Motion Event Linking

Now I will leave this topic here and go to further describe the country of Cianba.
Location and Motion Event Linking

Now I will leave this topic here and go to further describe the country of Cianba.

VP -> VBD NP* PP*  OR  VP -> VBD PP* NP*

Find all noun phrases

Find associated prepositional phrase

any

NP has location

Traveled location by Marco Polo
Results & Discussion
Evaluation Criteria for Named Entity Recognition

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>[ \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} ]</td>
<td>[ \frac{\text{True Positives}}{\text{Total Predicted Positives}} ]</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>[ \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} ]</td>
<td>[ \frac{\text{True Positives}}{\text{Total Actual Positives}} ]</td>
</tr>
<tr>
<td><strong>F1 score</strong></td>
<td>[ \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} ]</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Criteria for Named Entity Recognition

Ground truth:

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]
Evaluation Criteria for Named Entity Recognition

Ground truth:

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

NER Model:

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga
[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

<table>
<thead>
<tr>
<th>Ground Truth Entity</th>
<th>Predicted Entity</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marco Polo</td>
<td>Marco Polo</td>
<td>Exact</td>
</tr>
<tr>
<td>Armenia the Greater</td>
<td>Armenia</td>
<td>Partial</td>
</tr>
<tr>
<td>Arzinga</td>
<td>-</td>
<td>No match</td>
</tr>
</tbody>
</table>
Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

<table>
<thead>
<tr>
<th>Ground Truth Entity</th>
<th>Predicted Entity</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marco Polo</td>
<td>Marco Polo</td>
<td>Exact</td>
</tr>
<tr>
<td>Armenia the Greater</td>
<td>Armenia</td>
<td>Partial</td>
</tr>
<tr>
<td>Arzinga</td>
<td>-</td>
<td>No match</td>
</tr>
</tbody>
</table>

Matching Criteria:

- Exact Match
- Partial Match
Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

<table>
<thead>
<tr>
<th>Ground Truth Entity</th>
<th>Predicted Entity</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marco Polo</td>
<td>Marco Polo</td>
<td>Exact</td>
</tr>
<tr>
<td>Armenia the Greater</td>
<td>Armenia</td>
<td>Partial</td>
</tr>
<tr>
<td>Arzinga</td>
<td>-</td>
<td>No match</td>
</tr>
</tbody>
</table>

Matching Criteria:

- Exact Match
- Partial Match

Armenia -> correct
Greater -> Incorrect
Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

<table>
<thead>
<tr>
<th>Ground Truth Entity</th>
<th>Predicted Entity</th>
<th>Match</th>
<th>True Positive</th>
<th>False Positive</th>
<th>False Negative</th>
<th>True Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marco Polo</td>
<td>Marco Polo</td>
<td>Exact</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Armenia the Greater</td>
<td>Armenia</td>
<td>Partial</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Arzinga</td>
<td>-</td>
<td>No match</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Evaluating Gazetteers for location extraction

Gazetteer Analysis:
- Combined Index outperformed each of them individually
- Murray Gazetteer - high precision, low recall
- Yule Gazetteer - high precision, low recall
Evaluating NER for location extraction

**NER Analysis:**
- Gazetteer has best F1-score
- Stanford, spaCy, NLTK performs poorly: low precision and low recall
Results:

- Gazetteer **outperforms** all the other pre-trained NERs.
- AllenNLP performs well compared to other pre-trained NERs.
- AllenNLP has almost identical precision as the gazetteer, but low recall and hence, low F1-score.

**Entire Book NER Analysis:**

- Gazetteer has the best F1-score.
- Stanford, spaCy, NLTK perform poorly: low precision and low recall.
Results:

- Gazetteer **outperforms** all the other pre-trained NER
- AllenNLP performs well compared to other pre-trained NERs
- AllenNLP has almost identical precision as gazetteer, but low recall and hence, low F1-score

Observations:

- Pre-trained NERs do not perform well on entities with different naming convention
- Incorrectly identified entity type
  - Location -> Person
  - False negative for location
  - False positive for Person

Entire Book NER Analysis:

- Gazetteer has best F1-score
- Stanford, spaCy, NLTK performs poorly: low precision and low recall
Evaluating Motion Events Extraction

Event Extraction Analysis:
- FrameNet and WordNet has poor precision but high recall
- VerbNet has high precision but low recall
- Overall, VerbNet has highest f1 score
Results:
- Framenet and Wordnet has high precision but low recall
- VerbNet has high precision but low recall
- High recall represents that all 3 are able to identify correct motion event triggers
- But low precision shows that they also falsely identify other events as motion events

Event Extraction Analysis:
- Framenet and Wordnet has poor precision but high recall
- VerbNet has high precision but low recall
- Overall, VerbNet has highest f1 score
Results:
- FrameNet and WordNet have high precision but low recall.
- VerbNet has high precision but low recall.
- High recall indicates that all three are able to identify correct motion event triggers.
- But low precision shows that they also falsely identify other events as motion events.

Observations:
- Choice of seeds affects the results.
  - Choice of seeds can be narrowed down.
  - However, it needs in-depth domain knowledge.
  - Goal is to automate path retracing with minimal domain knowledge.

Event Extraction Analysis:
- FrameNet and WordNet has poor precision but high recall.
- WordNet has the highest recall.
- VerbNet has high precision but low recall.
Motion Event and Location Linking

Traveled Location Identification Analysis:
- Poor recall
- PART 1 has a very low recall and PART 3 has highest recall
- High precision for all three parts
- WordNet has highest F1-score
Results:

- Recall shows the identification of traveled locations which is low
- Precision shows correct identification of traveled location which is important as well and it is quite high here
- All three lexical resources have high precision but low recall
- WordNet has highest F1-score

Traveled Location Identification Analysis:
- Poor recall
- High precision
- WordNet has highest F1-score
Result: Extracted locations and interpolated travel path
Limitations

- Verbs linking

Leaving Ta-in-fu, and riding westward full seven days through very fine districts, amid numerous merchants, you find a large town, named Pi-an-fu, supported by commerce and the silk manufacture.

Riding -> find -> named -> Pi-an-fu
Limitations

● Verbs linking

"Leaving Ta-in-fu, and riding westward full seven days through very fine districts, amid numerous merchants, you find a large town, named Pi-an-fu, supported by commerce and the silk manufacture."

Riding -> find -> named -> Pi-an-fu

● Extracting traveled locations described using non-motion verbs

"Having told you all about these Tartars of East, I might go on to treat of Great Turkey."
Limitations

- Ambiguity in the narrative

“Having nothing more to tell of this island, I will go to Zanghibar.”  Not visited

“Now I will go to another city named Balk.”  visited
Limitations

● Ambiguity in the narrative

“Having nothing more to tell of this island, I will go to Zanghibar.”  Not visited

“Now I will go to another city named Balk.”  visited

● There has also been a difference of opinion between various historians about the exact travel path of Marco Polo and there has been not single agreed and verified path.
Contributions

- Corpus creation
- Generating a mapping between historical and contemporary locations in the context of Marco Polo travelogue
- Evaluating state-of-the-art NER tools for 12th century historical travelogue
- Approach for extracting traveled locations from 12th century historical texts
Future Work

- Improving state-of-the-art NER systems to avoid problems with out-of-vocabulary words
- Implementing multiple relation extractions and verbs linking
- Identification of meaning from the text
- A way to determine the order of the travel
- Finding individual route segments and connect route segments to generate a travel path

“Deo Gratias. Amen, Amen.”
When a man has left this country and travelled twenty days westward, he approaches a province on the borders of Manji named Sin-din-fu. The capital, bearing the same name, was anciently very great and noble, governed by a mighty and wealthy sovereign. He died, leaving three sons, who divided the city into three parts, and each enclosed his portion with a wall, which was within the great wall of twenty miles in circuit. They ranked still as kings, and had ample possessions; but the great khan overcame them, and took full possession of their territory. Through the city, a large river of fresh water, abounding with fish, passes and flows on to the ocean, distant eighty or a hundred days' journey; it is called Quian-su. On that current is a very great number of cities and castles, and such a multitude of ships as no one who has not seen could possibly believe. Equally wonderful is the quantity of merchandise conveyed; indeed it is so broad as to appear a sea and not a river. Within the city, it is crossed by a bridge wholly of marble, half a mile long and eight paces broad; the upper part is supported by marble columns, and richly painted; and upon it are many houses where merchants expose goods for sale; but these are set up in the morning and taken down in the evening. At one of them, larger than the others, stands the chamberlain of the khan, who receives the duty on the merchandise sold, which is worth annually a thousand golden bezants. The inhabitants are all idolaters; and from that city a man goes five days’ journey through castles, villages, and scattered houses. The people subsist by agriculture, and the tract abounds with wild beasts. There are also large manufactures of gauzes and cloth of gold. After travelling these five days, he comes to Thibet.
XLII. The Province and City of Sin-din-fu.

When a man has left this country and travelled twenty days westward, he approaches a province on the borders of Manji named Sin-din-fu. The capital, bearing the same name, was anciently very great and noble, governed by a mighty and wealthy sovereign. He died, leaving three sons, who divided the city into three parts, and each enclosed his portion with a wall, which was within the great wall of twenty miles in circuit. They ranked still as kings, and had ample possessions; but the great khan overcame them, and took full possession of their territory. Through the city, a large river of fresh water, abounding with fish, passes and flows on to the ocean, distant eighty or a hundred days’ journey; it is called Quian-su. On that current is a very great number of cities and castles, and such a multitude of ships as no one who has not seen could possibly believe. Equally wonderful is the quantity of merchandise conveyed; indeed it is so broad as to appear a sea and not a river. Within the city, it is crossed by a bridge wholly of marble, half a mile long and eight paces broad; the upper part is supported by marble columns, and richly painted; and upon it are many houses where merchants expose goods for sale; but these are set up in the morning and taken down in the evening. At one of them, larger than the others, stands the chamberlain of the khan, who receives the duty on the merchandise sold, which is worth annually a thousand golden bezants. The inhabitants are all idolaters; and from that city a man goes five days’ journey through castles, villages, and scattered houses. The people subsist by agriculture, and the tract abounds with wild beasts. There are also large manufactures of gauzes and cloth of gold. After travelling these five days, he comes to Thibet.
Identifying context using SemLink Mapping

VN class / FN frame mapping

Verb

(vn class 1, fn frame 1)
(vn class 2, fn frame 2)
(vn class N, fn frame N)
Identifying context using SemLink Mapping

1. VN class / FN frame mapping
   - (vn class 1, fn frame 1)
   - (vn class 2, fn frame 2)
   - (vn class N, fn frame N)

2. VN role / FN element mapping
   - (vn role 1, fn role 1)
   - (vn role 2, fn role 2)
   - (vn role N, fn role N)
Identifying context using SemLink Mapping

1. VN class / FN frame mapping
   - Verb
     - (vn class 1, fn frame 1)
     - (vn class 2, fn frame 2)
     - (vn class N, fn frame N)

2. VN role / FN element mapping
   - vn class, fn fame
     - (vn role 1, fn role 1)
     - (vn role 2, fn role 2)
     - (vn role N, fn role N)

VerbNet Thematic Roles
- Source
- Destination
- Location
Identifying context using SemLink Mapping

1. VN class / FN frame mapping
   - Verb
     - (vn class 1, fn frame 1)
     - (vn class 2, fn frame 2)
     - (vn class N, fn frame N)

2. VN role / FN element mapping
   - vn class, fn fame
     - (vn role 1, fn role 1)
     - (vn role 2, fn role 2)
     - (vn role N, fn role N)

VerbNet
- Thematic Roles
  - Source
  - Destination
  - Location

FrameNet
- Frame Elements
  - Source
  - Destination
  - Undesirable_location