Digital Humanities 2023, Graz

Marco Polo's Travels Revisited:
From Motion Event Detection to Optimal Path Computation in 3D Maps

12.07.2023

Andreas Niekler, Magdalena Wolska, Marvin Thiel, Matti Wiegmann, Benno Stein, and Manuel Burghard

Computational Humanities
UNIVERSITÄT LEIPZIG

Webis Group Bauhaus-Universität Weimar

OUTLINE

- **Introduction**
    - **Travel Literature**
    - **Marco Polo**

- **Natural Language Processing**
    - **Corpus**
    - **Named Entity Recognition**
    - **Motion Identification**

- **Spatial Analysis and Visualization**

- **Conclusion**

TRAVEL LITERATURE

SUBTEXT OF A LOT OF LITERATURE [2]

Source: Goodreads

− Travel is a **major motif in many literary works** [2]

− The genre of travel literature encompasses **outdoor literature, guidebooks, nature writing, and travel memoirs**. [3]

− Travel literature has its **roots in this reality** [3]

− A human being develops itself by confronting itself with reality and the "unknown" [4]

− Many examples: Che Guevara's The Motorcycle Diaries, The Travels of Marco Polo (Il Milione), Peter Mayle's A Year in Provence, …
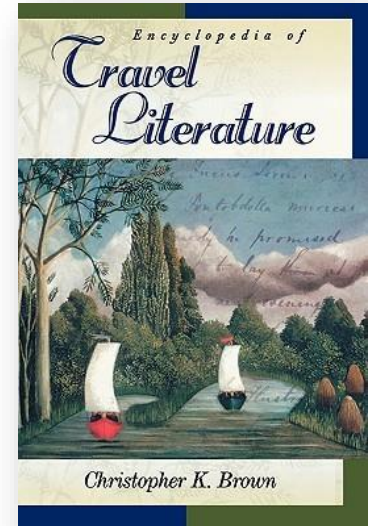
[2] Mewshaw, M. (2005). Travel, Travel Writing, and the Literature of Travel. *South Central Review.*
[3] Cuddon, J. A., & Birchwood, M. (2014). *The Penguin dictionary of literary terms and literary theory.*
[4] Brenner, Peter: *Does Travelling Matter?* The Impact of Travel Literature on European Culture.

MARCO POLO

VENETIAN MERCHANT AND EXPLORER

- **Traveled through Asia along the Silk Road** between 1271 and 1295. [4]

- *Il milione* ("The Million"), **known in English as the Travels of Marco Polo**, is a classic of travel literature. [4]

- 13th-century travelogue written down by Rustichello da Pisa from stories told by Italian explorer Marco Polo. [5]

- The book was translated into many European languages in Marco Polo's own lifetime. [5]

Source: [5]

[5] https://www.britannica.com/biography/Marco-Polo
[6] https://en.wikipedia.org/wiki/The_Travels_of_Marco_Polo

MARCO POLO'S TRAVELOGUE

IS IT REALITY OR FICTION?

– Some have questioned **whether Marco had actually travelled to China or was just repeating stories** that he had heard from other travellers:

"*Countless authors of travelogues, such as Marco Polo, presented often rather astonishing accounts seemingly unbelievable in their content for their audiences back home*" [7, p. 27]

Source: [6]

[7] Classen, A. (2013). *East Meets West in the Middle Ages and Early Modern Times: Transcultural Experiences in the Premodern World*. De Gruyter.
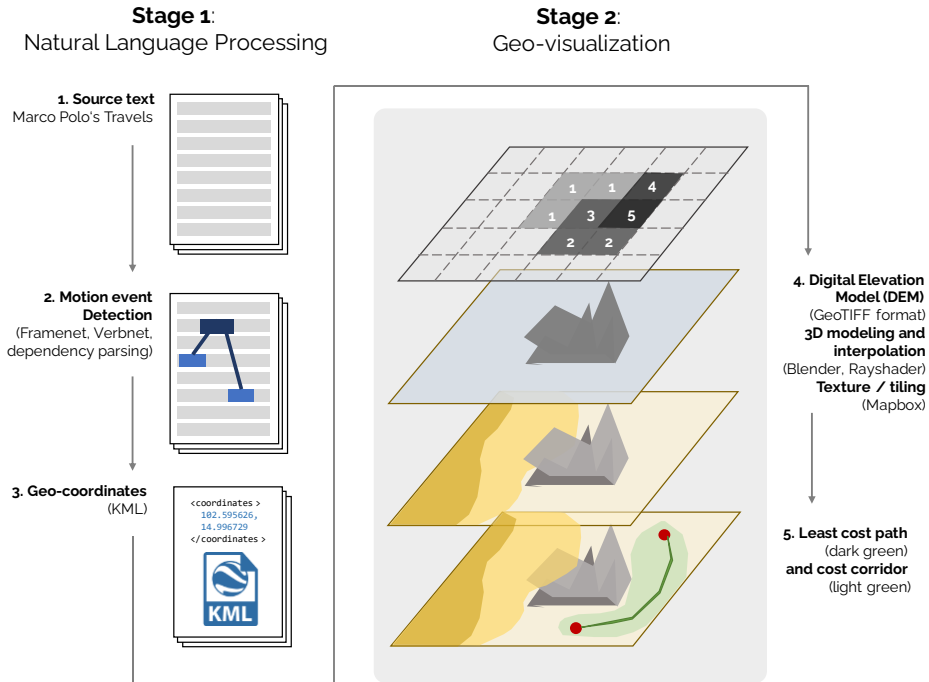
MARCO POLO'S TRAVELOGUE

Source: [5]

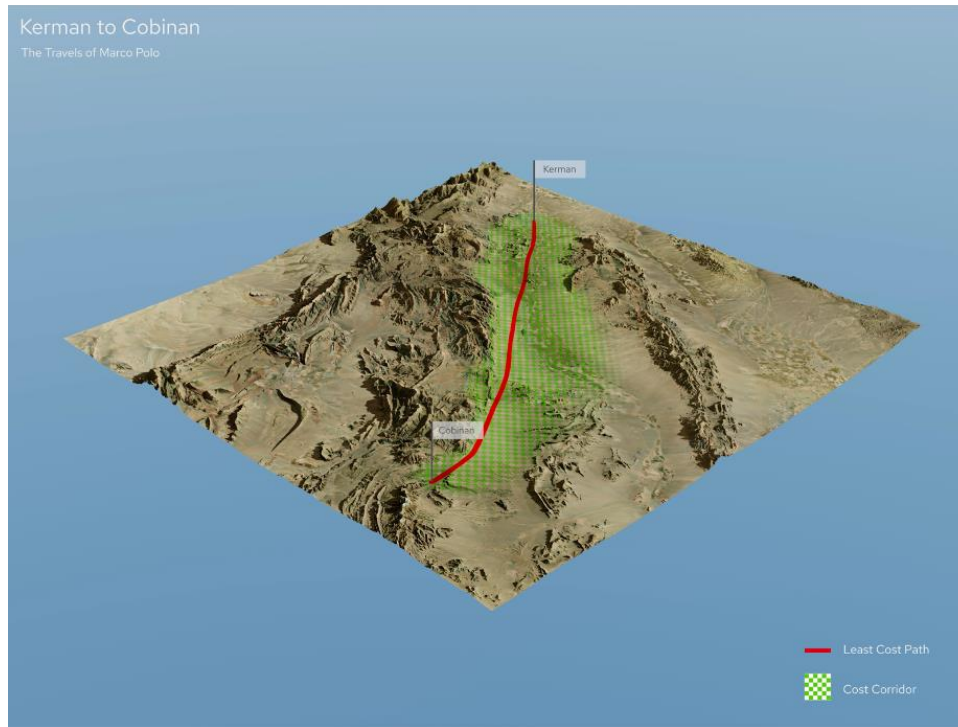# MINING TRAVEL LITERATURE

RESEARCH QUESTIONS

– How can we **process travel literature** computationally?

   – Can we **reconstruct the route** of Marco Polo by analyzing the text contents (semi-)**automatically**?

   – Can we **geo-reference location entities** from the book?

   – Which **data formats** are used in such process?

– How can we **visualize the travel** and movement?

   – Can we use the information gained to **create immersive visualizations** to **augment the reading experience**?

   – Can we **ground reader experience** with visualizations?

MINING TRAVEL LITERATURE

Stage 1:
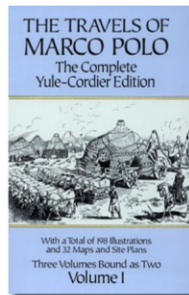Natural Language Processing

Stage 2:
Geo-visualization

1. Source text
Marco Polo's Travels

**2. Motion event
Detection**
(Framenet, Verbnet,
dependency parsing)

3. Geo-coordinates
(KML)

```
<coordinates>
  102.595626,
  14.996729
</coordinates>
```

**4. Digital Elevation
Model (DEM)**
(GeoTIFF format)
**3D modeling and
interpolation**
(Blender, Rayshader)
Texture / tiling
(Mapbox)

5. Least cost path
(dark green)
and cost corridor
(light green)

MINING TRAVEL LITERATURE

SOURCE TEXT PROCESSING

SOURCE TEXT AQUISITION

- English translation by Henry Yule, Published in 2 volumes
- Text files from Project Gutenberg

https://www.gutenberg.org/ebooks/10636, https://www.gutenberg.org/ebooks/12410

SOURCE TEXT PROCESSING

STEP 1: LOCATION ANNOTATION

− We **manually annotated all locations** in the source text following standards from **Named Entity Recognition**.

− We created a **gazetteer** based on the back matter index of the source text.
 − It lists location names and their alternative (modern) names.

Entity Name	Elobrations	Related Contents	Alternative Name	Type
Kinsay	['formerly Lin-ngan now Hang-chau fu']	['its surrender to Bayan;', 'extreme public security;', , ...	['Capital', ' Khansa', ' Khinsa', ' Khingsai', ' Khanzai', 'Cansay', 'Campsay', 'Kin-sai', 'Quin-sai', 'Kitvaal', 'Quin-sal']	GPE

− We use 3 Geographic Information Systems (GIS) to resolve modern names and (later) find the location's coordinates**.**
  − China Historical GIS, Silk Road Historical GIS, Silk Road GIS

SOURCE TEXT PROCESSING

STEP 1: COMPUTATIONAL CHALLENGES FOR NER IN TRAVELOGUES

- **Gazetteers** work well but **must be created** for every travelogue.

- **GIS** to translate historical names or find historical references **may not exist.**

- State-of-the-Art methods (we tested Flair) struggle with historical references (low recall)
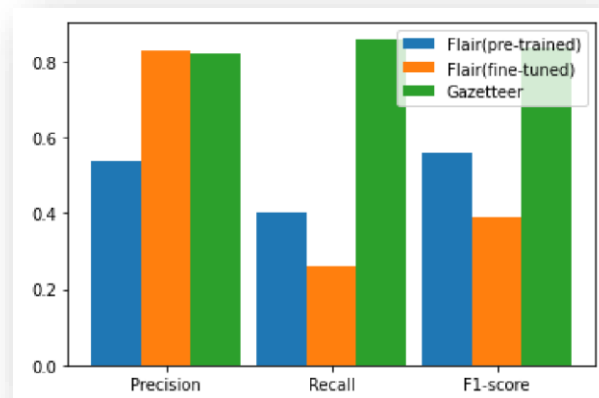
SOURCE TEXT PROCESSING

STEP 2: MOTION EVENT ANNOTATION

- We can identify **travel actions** between the location entities using **motion or movement events.**
    - These events are centered around **motion verbs** in a sentence.
    - Motion events have attributes, like source and target location of a travel:

    ...that you pass in *going* from Trebizond to Tauris...From Tauris to Persia is a journey of twelve days."

    - Attributes can also specify a travel **direction or duration.**

    Travelling through a succession of towns and villages that look like one continuous city, two days further on to the south-east, you find the great and fine city of **GHIUJU** which is under Kinsay."

- We can construct a route by combining all travel motion events.

SOURCE TEXT PROCESSING

STEP 2: MOTION EVENT ANNOTATION

- We built **a list of possible motion verbs** and their attributes with the help of lexical resources.
    - VerbNet: largest online English **verb lexicon** (https://verbs.colorado.edu/verbn et/); includes verbs' predicate-argument structures
    - FrameNet: annotated **examples of words' meaning** and usage in real texts (https://framenet.icsi.berkeley.edu/fndrupal/); frame annotated.
- These verbs are then marked in the source text.
- Attributes of the for the **motion events** are annotated based on the **noun phrases and prepositional phrases belonging to the motion verb**.

SOURCE TEXT PROCESSING

STEP 2: MOTION EVENT ANNOTATION

Source: verbs.colorado.edu/verb-index/vn/run-51.3.2.php

SOURCE TEXT PROCESSING

STEP 3. GEO-COORDINATE EXTRACTION

– We extract coordinates for the location entities within the motion events from the GIS used to construct the gazetteer.

– **Result:**

   – **398 sentences which express motion of Marco Polo**

   – **and contain relative exact sources, destinations, or locations**

   – **and the locations coordinates.**

SOURCE TEXT PROCESSING

STEP 3. GEO-COORDINATE EXTRACTION

– We mark **133 locations on a Map** and link them in chronological order.
– The Result is **saved as a .kml file**.
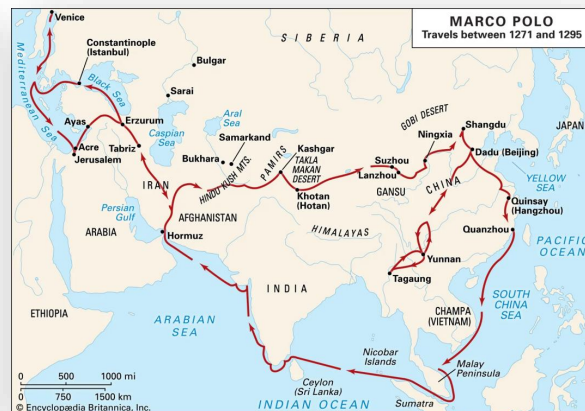
Source: [5]

SPATIAL ANALYSIS

STEP 4. DIGITAL ELEVATION MODEL

– There is **more to travel experiences than just the places** visited alone.

– Vegetation, climate, obstacles and vistas are all **details that come to mind when thinking about traveling**, especially by foot.

– Details about **landscape and the traveled roads are most often sparsely described** and days worth of travel are described using just a few sentences

"When you *depart from* this City of Cobinan, you find yourself again in a Desert of surpassing aridity, which lasts for some eight days; here are neither fruits nor trees to be seen, and what water there is bitter and bad, so that you have to carry both food and water."

SPATIAL ANALYSIS

STEP 4. DIGITAL ELEVATION MODEL

– However, when **analyzing travel writing, regular point-based maps do not normally allow for analyzing aspects of a more qualitative nature** like the experiences that were made while traveling [8].

– [9] describe an environment with the purpose
"*to create immersive geographies that link the experiential, the emotional and the symbolic elements of literary works to the nuanced, dimensional richness of places as inspired by authors and their works*".

[8] Murrieta-Flores et al., 2016
[9] Trevor M. Harris et al.

SPATIAL ANALYSIS

STEP 4. DIGITAL ELEVATION MODEL

- Create a **bounding box for a partial route**
    - Kerman: `(30.27305095, 57.0662499)`
      Cobinan (Kuhbanan): `(31.4126295, 56.28006)`
      Tonocain (Tabas): `(33.60953795, 56.9456578)`

- **Download a DEM** (Digital Elevation Model) as GTIFF
    - COP30 (Copernicus Global DEM 30m)

- **Shade the generated 3D models** (*Blender, rayshader in R*)
    - Mapbox has been chosen as the source for satellite imagery.

SPATIAL ANALYSIS

STEP 4. DIGITAL ELEVATION MODEL
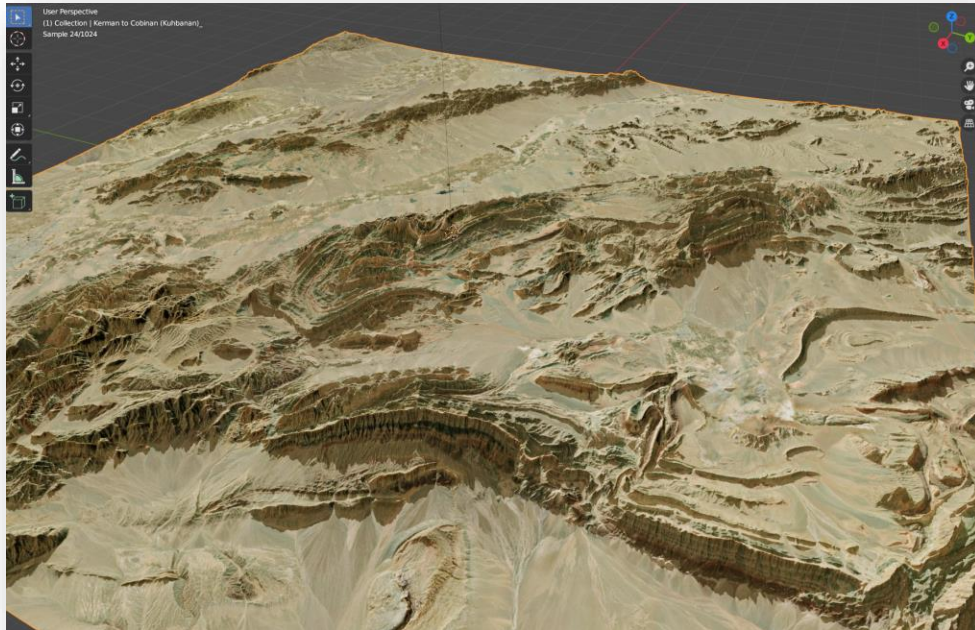
SPATIAL ANALYSIS

STEP 5. LEAST COST PATH AND COST CORRIDOR

– As suggested by [10], "**Cost-Surface Analysis (CSA) and Least-Cost-Path Analysis (LCP) can be used to facilitate more nuanced interpretations of historical works** of travel writing and topographical literature".

– Initial data used to create the map is comprised of **two coordinates**:
  – The **origin** of the route's section and the **destination**.
  – **Mountains and valleys on the map can already be used as guides to guess a possible route that is easy to travel along**.
  – In order to back the readers' intuition, the **map can still be enhanced by highlighting the areas that are most easy** to travel through.

[10] Murrieta-Flores et al., 2017

SPATIAL ANALYSIS

STEP 5. LEAST COST PATH AND COST CORRIDOR
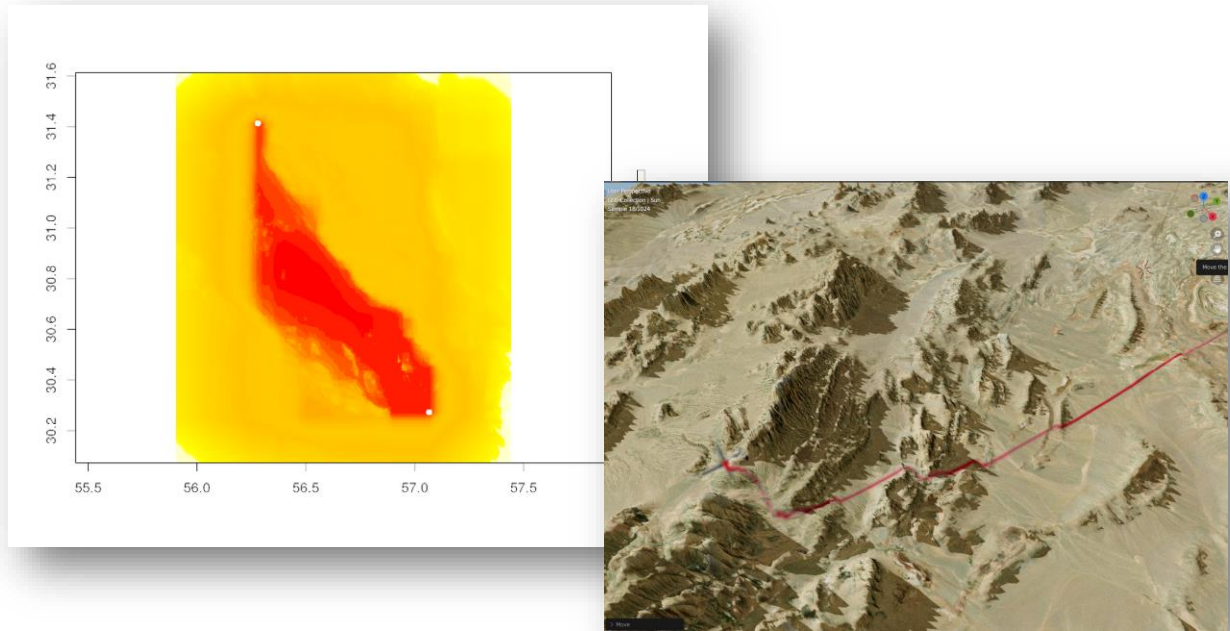
SPATIAL ANALYSIS

Cobinam to Tonocain
The Travels of Marco Polo

Cobinam

Tonocain

Route

Cost Corridor
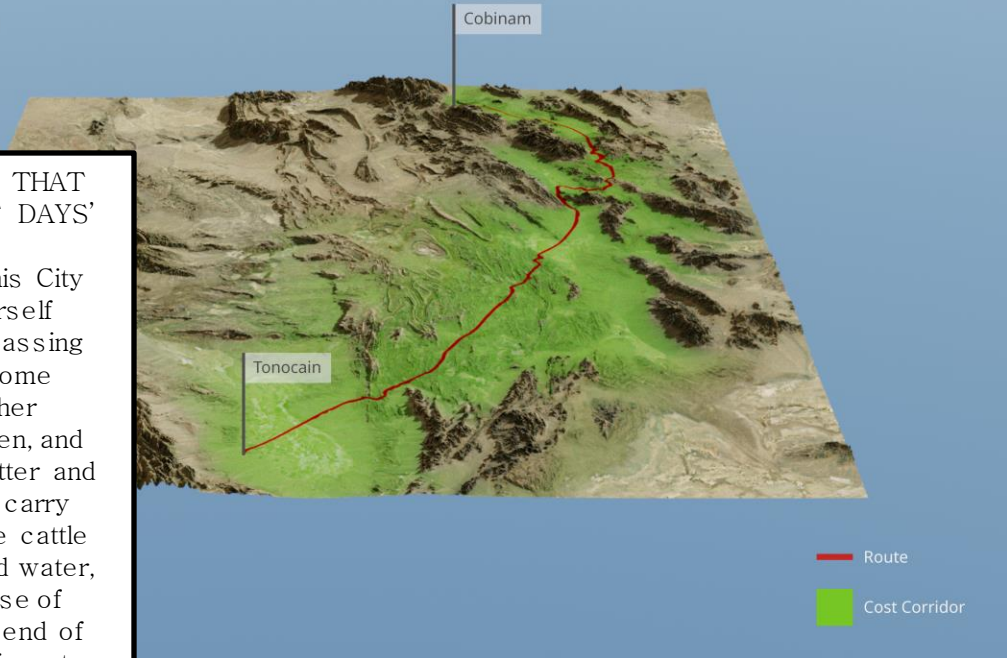
OF A CERTAIN DESERT THAT CONTINUES FOR EIGHT DAYS' JOURNEY.
When you depart from this City of Cobinan, you find yourself again in a Desert of surpassing aridity, which lasts for some eight days; here are neither fruits nor trees to be seen, and what water there is is bitter and bad, so that you have to carry both food and water. The cattle must needs drink the bad water, will they nill they, because of their great thirst. At the end of those eight days you arrive at a Province which is called TONOCAIN.

SPATIAL ANALYSIS
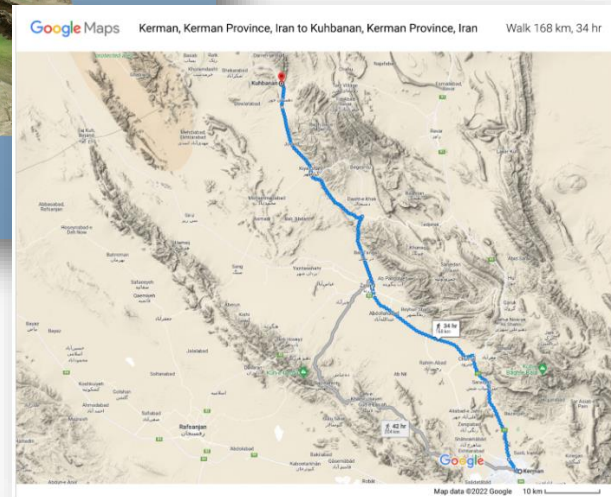
PIPELINE FOR TRAVEL LITERATURE

RESULTS

– Detailed annotation and linking of location entities and travel events in "***Travels of Marco Polo"***.

– Method to computationally extract the route from Travelouges.

– Generated a route map and compared to Google Maps.
  – While Google Maps provides a good overview of the route and the terrain's topography, the **renderings provide a more photo-realistic and immersive representation** of the route.

– Paths and Cost Corridors
  – Least Cost Path and **cost corridors show an optimal way to travel** but not necessarily a comfortable one.
  – Readers can **align their expectations with the descriptions** in the book

LITERATURE

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Bol, P. K. (2005). China Historical GIS. Historical Geography, 150–152.

Brenner, P. J. (2011). El viaje y la percepción del otro: Viajeros por la Península Ibérica y sus descripciones (siglos XVIII y XIX) (R. Musser, Hrsg.; Bd. 141, S. 11–22). Iberoamericana / Vervuert.

Classen, A. (2013). East Meets West in the Middle Ages and Early Modern Times: Transcul-tural Experiences in the Premodern World. De Gruyter. https://books.google.de/books?id=snPnBQAAQBAJ

Harris T. M., Lafone F. & Bonenberger D. (2016). From Mapping Text in Space to Experienc-ing Text in Place: Exploring Literary Virtual Geographies. In: Cooper, D., Donaldson, C., & Murrieta-Flores, P. (Eds.). Literary Mapping in the Digital Age (1st ed.). Routledge. https://doi.org/10.4324/9781315592596

Cuddon, J. A., & Birchwood, M. (2014). The Penguin dictionary of literary terms and literary theory (R. Habib, Hrsg.; 5. ed., publ. in paperback). Penguin Books.

European Space Agency & Airbus. (2021). Copernicus DEM [Data set]. European Space Agency. https://doi.org/10.5270/ESA-c5d3d65

Halder, K., Akbik, A., Krapac, J., & Vollgraf, R. (2020). Task Aware Representation of Sen-tences for Generic Text Classification. COLING 2020, 28th International Conference on Com-putational Linguistics.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. Association for Computational Linguistics (ACL) System Demonstrations, 55–60.

Mewshaw, M. (2005). Travel, Travel Writing, and the Literature of Travel. South Central Re-view, 22(2), 2–10. https://doi.org/10.1353/scr.2005.0042

Miller, G.A. (1998). WordNet: An electronic lexical database. MIT Press.

Murrieta-Flores, P., Donaldson, C., & Gregory, I. (2017). GIS and Literary History: Advancing Digital Humanities Research throug h the Spatial Analysis of Historical Travel Writing and Topographical Literature. Digital Humanities Quarterly. http://hdl.handle.net/10034/620256

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., & Scheffczyk, J. (2016). FrameNet II: Extended theory and practice. International Computer Science Institute.

Schuler, K. K., Korhonen, A., & Brown, S. (2009). VerbNet Overview, Extensions, Mappings and Applications. Proceedings of Human Language Technologies: The 2009 Annual Confer-ence of the North American Chapter of the Association for Computational Linguistics, Com-panion Volume: Tutorial Abstracts, 13–14.

Stowe, K., Preciado, J., Conger, Kathryn, Brown, S. W., Kazeminejad, G., & Palmer, M. (2021). SemLink 2.0: Chasing Lexical Resources. 30th Annual Meeting of the Association for Computational Linguistics, 222–227.