Travelogues

TRAVELING THROUGH SPACE AND TIME

Jan Rörden (AIT Austrian Institute of Technology)

NKOS Workshop 2018

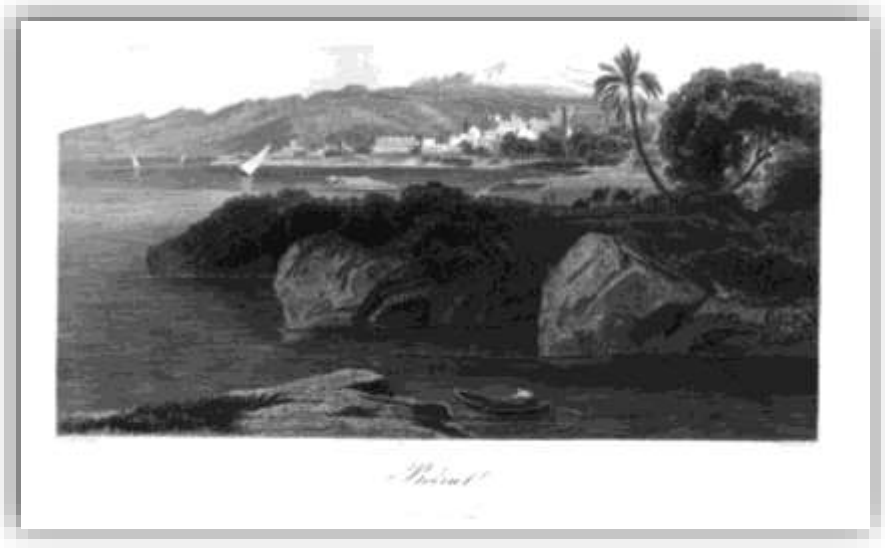
Porto, Portugal

FWF Der Wissenschaftsfonds.

Gefördert durch DFG Deutsche Forschungsgemeinschaft

OVERVIEW

- The Travelogues project
- Technical research tasks
- Preliminary results
- Next steps



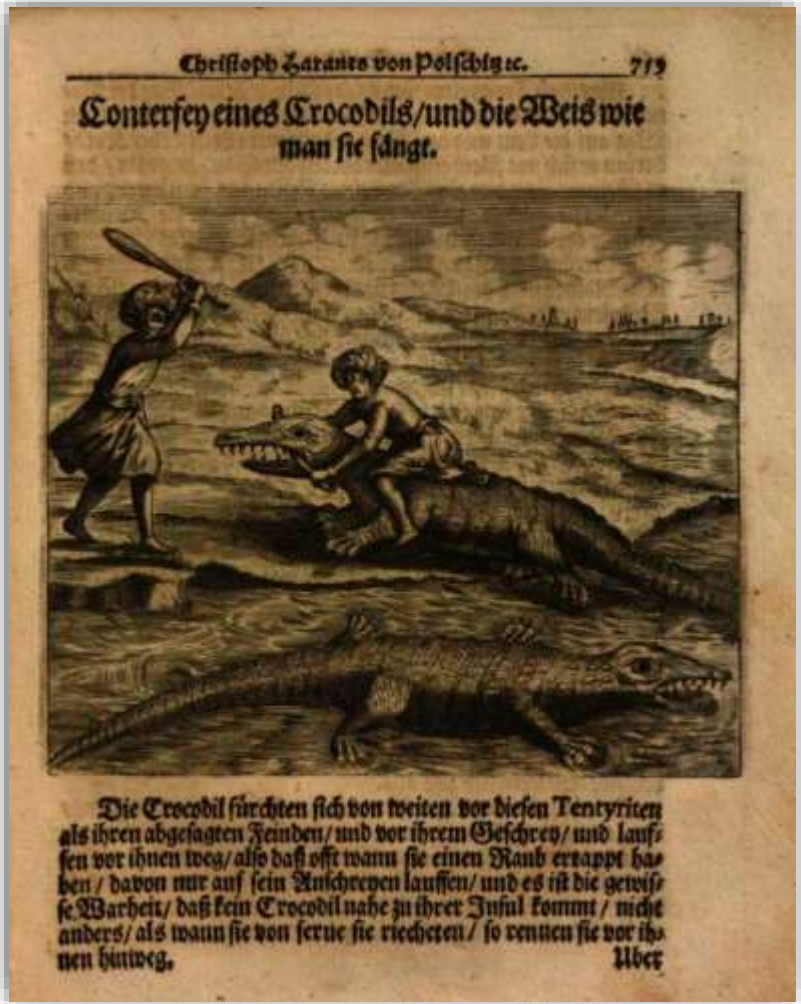
Travelogues

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

TRAVELOGUES | MOTIVATION

A relevant historical source

Large-scale digitization offers new possibilities



TRAVELOGUES | DATA SET

- 600,000 digitized books from 1500-1875
  - Estimated 1,000 travelogues
- OCR readily available
- Meta data
  - Author, publisher, title, keywords, year...

TRAVELOGUES | RESEARCH GOALS

Semi-supervised creation of a corpus

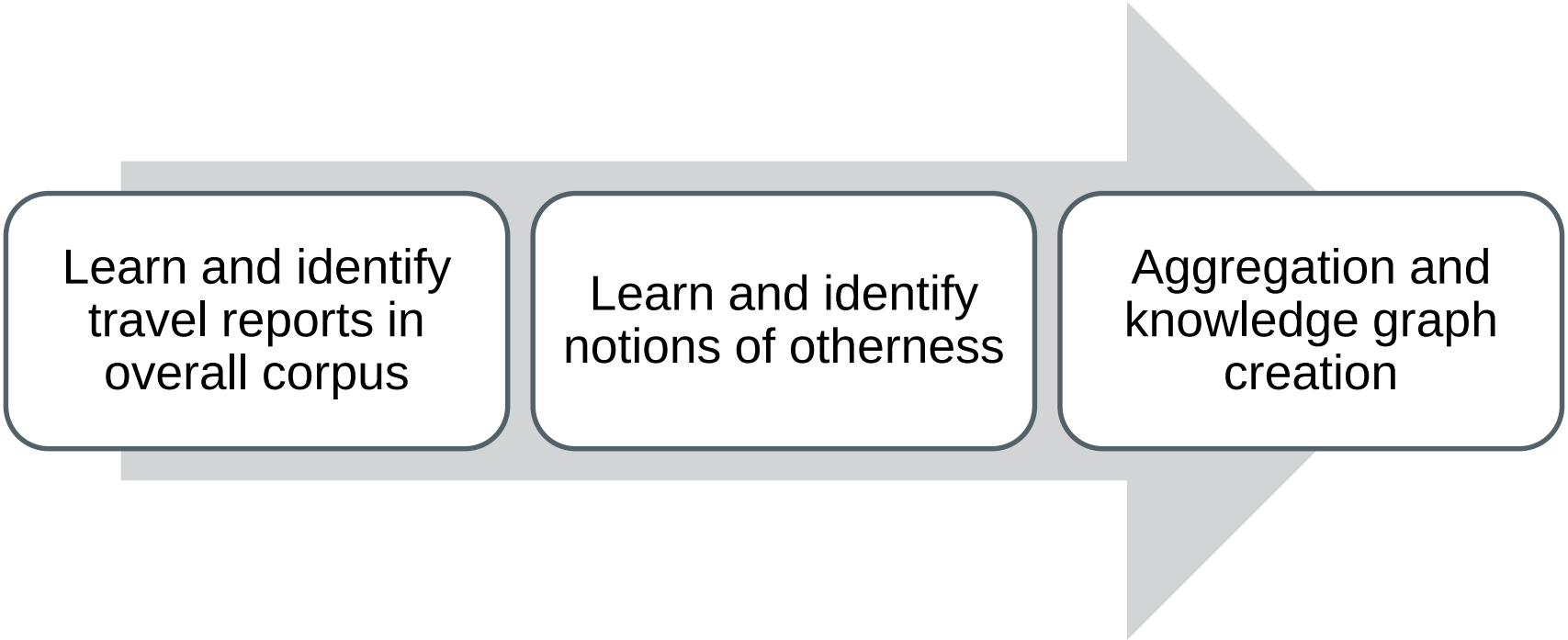
Enrichment of meta data

Perceptions of the other & discourses of the Orient

OVERVIEW

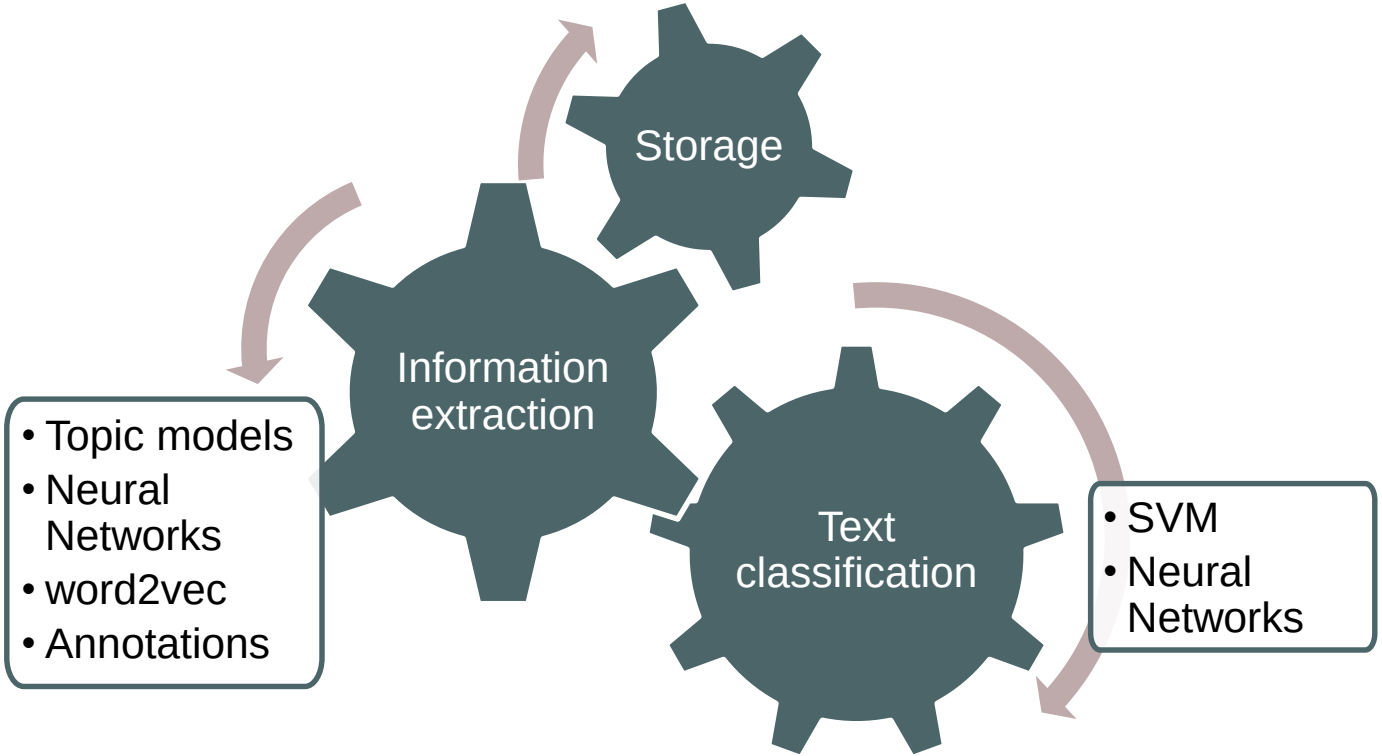
- The Travelogues project
- Technical research tasks
- Preliminary results
- Next steps

TECHNICAL ASPECT | TASKS



Hernach giengen wir durch etliche Gäßlein / in deren etliche nur die Schuster allein / in andern die Schneider allein feil haben / lauter auf den Kauff allbereit fertige Wahren / und so fortan / von andern vielen Handwercken / ja es haben auch die Köch ihre Gasse / allda sie Schöpses / Cameel / Büffelfleisch / und von anderm Vieh in grosser Menge / gesotten und gebraten / mit anderm Zugemüß / als Reiß / Erbsen / Bohnen / Hirsch / ꝛc. zugericht / item Hüner / Fisch und andere Sachen feil haben / insonderheit hab ich mit meinen Augen gesehen / daß man da Hämmel hatte mit so grossen Schwäntzen / daß daran allein so viel Fleisch war / als bey uns an dem grösten Hamelsschlägel / wie man uns dann für gewiß berichtet / daß mancher Schwantz

TECHNICAL ASPECT | TOOLS



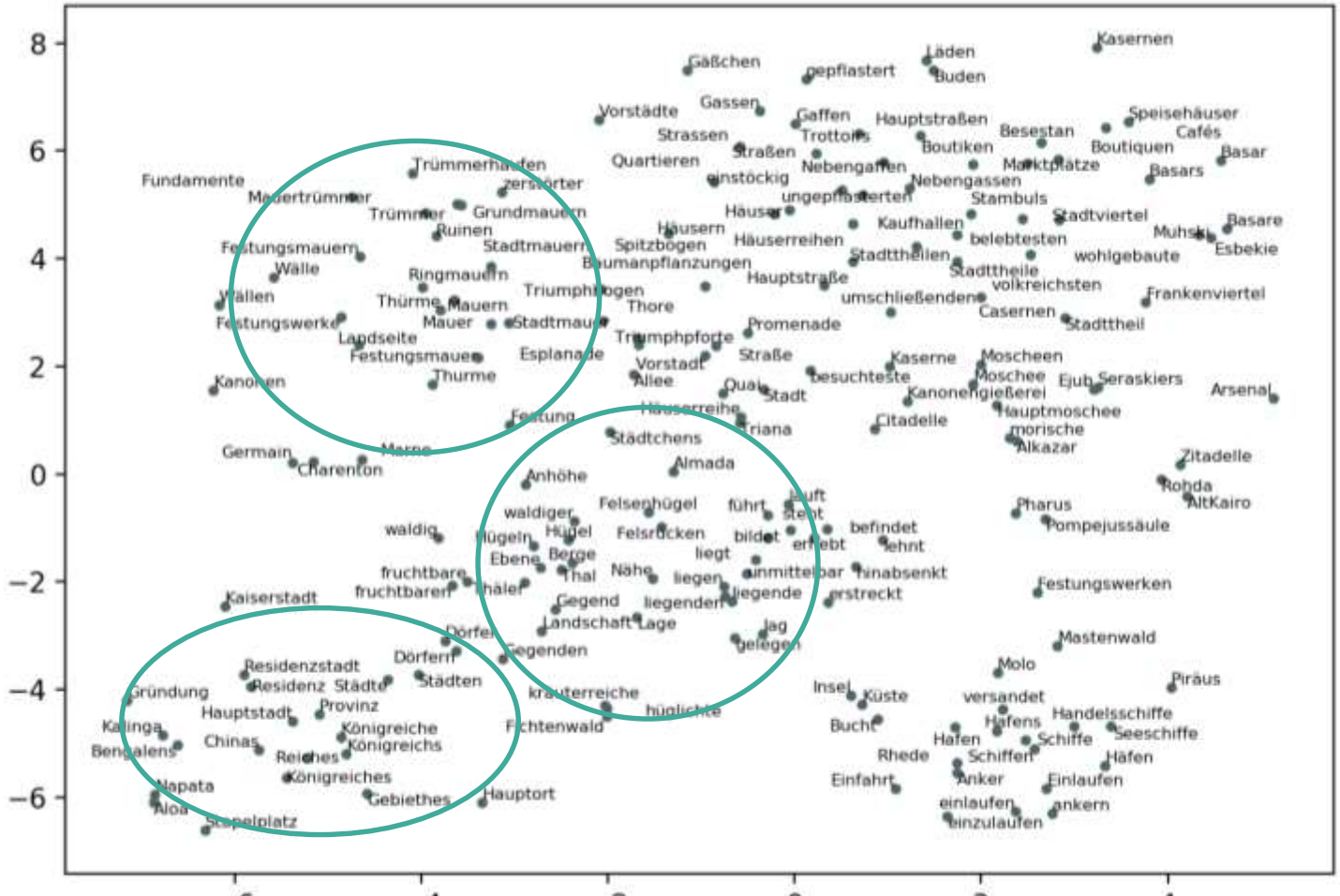
OVERVIEW

- The Travelogues project
- Technical research tasks
- Preliminary results
- Next steps

Travelogues

PRELIMINARY RESULTS | EXPLORATION

- Word2vec without pre-processing
- Classification without pre-processing
- Manual evaluation through domain experts
- Feedback/insights from domain experts


8
6
4
2
0
−2
−4
−6
Kasernen
Läden
Buden
Gäßchen
gepflastert
Vorstädte
Gassen
Gaffen
Hauptstraßen
Speisehäuser
Cafés
Strassen
Trottoirs
Boutiken
Besestan
Boutiquen
Basar
Straßen
Quartieren
Nebengaffen
Marktplätze
Basars
einstöckig
Nebengassen
Fundamente
Trümmerhaufen
zerstörter
Mauertrümmer
Trümmer
Grundmauern
ungepflasterten
Stambuls
Häuser
Stadtviertel
Basare
Häusern
Ruinen
Kaufhallen
belebtesten
Muhski
Esbekie
Festungsmauern
Stadtmauern
Spitzbogen
Häuserreihen
Stadttheilen
wohlgebaute
Wälle
Baumanpflanzungen
Stadttheile
Ringmauern
Hauptstraße
volkreichsten
Wällen
Triumphbogen
umschließenden
Frankenviertel
Thürme
Mauern
Thore
Casernen
Festungswerke
Mauer
Stadtmauer
Promenade
Stadttheil
Landseite
Triumphpforte
Festungsmauer
Esplanade
Straße
Kaserne
Moscheen
Vorstadt
besuchteste
Moschee
Kanonen
Allee
Quai
Stadt
Kanonengießerei
Ejub
Seraskiers
Arsenal
Häuserreihe
Hauptmoschee
Festung
Triana
Citadelle
morische
Städtchens
Alkazar
Germain
Marne
Charenton
Almada
Zitadelle
Anhöhe
Ronda
Felsenhügel
führt
läuft
Pharus
AltKairo
waldiger
steht
waldig
Hügel
Felsrücken
bildet
befindet
Pompejussäule
Hügeln
erhebt
Berge
liegt
lehnt
Ebene
fruchtbare
Thal
Nähe
liegen
unmittelbar
hinabsenkt
fruchtbaren
Thäler
liegende
erstreckt
Festungswerken
Kaiserstadt
Gegend
liegenden
Landschaft
Lage
Dörfer
Mastenwald
Gegenden
gelegen
Residenzstadt
Dörfern
Molo
Residenz
Städte
Städten
Piräus
Gründung
kräuterreiche
Insel
Küste
versandet
Hauptstadt
Provinz
Handelsschiffe
Kalinga
Königreiche
Fichtenwald
Bucht
Hafens
Chinas
Königreichs
Hafen
Schiffe
Seeschiffe
Bengalens
Reiches
Rhede
Schiffer
Häfen
Königreiches
Einfahrt
Anker
Einlaufen
Napata
Gebiethes
Hauptort
einlaufen
ankern
Aloa
Stapelplatz
einzulaufen

PRELIMINARY RESULTS | CLASSIFICATION

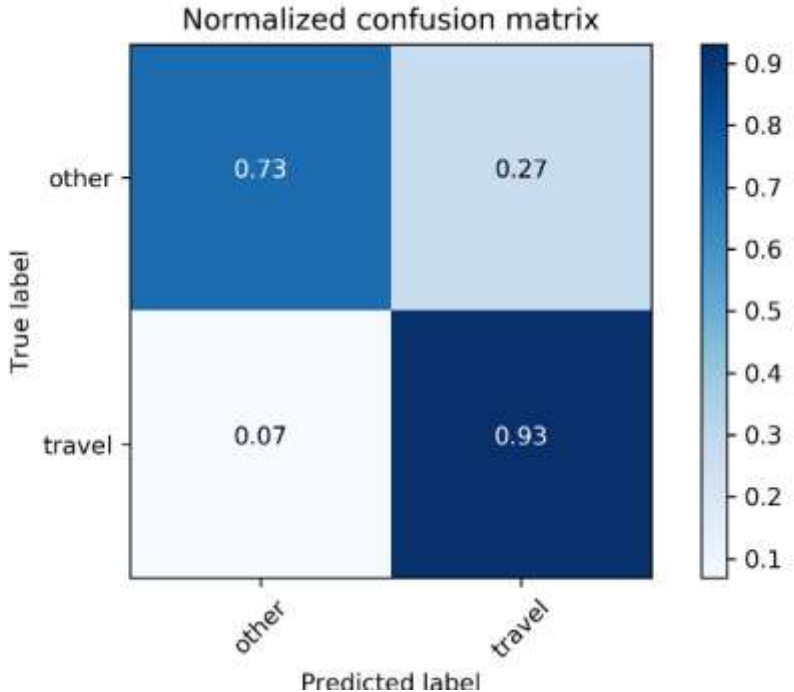
- Manually curated corpus of 30 travelogues
  - Extended to >200 books

- Support Vector Machine
  - Classifier performance evaluation through domain experts

Classifier	Precision	Recall	Accuracy	F1-score
SVM	0.82	0.93	0.84	0.87

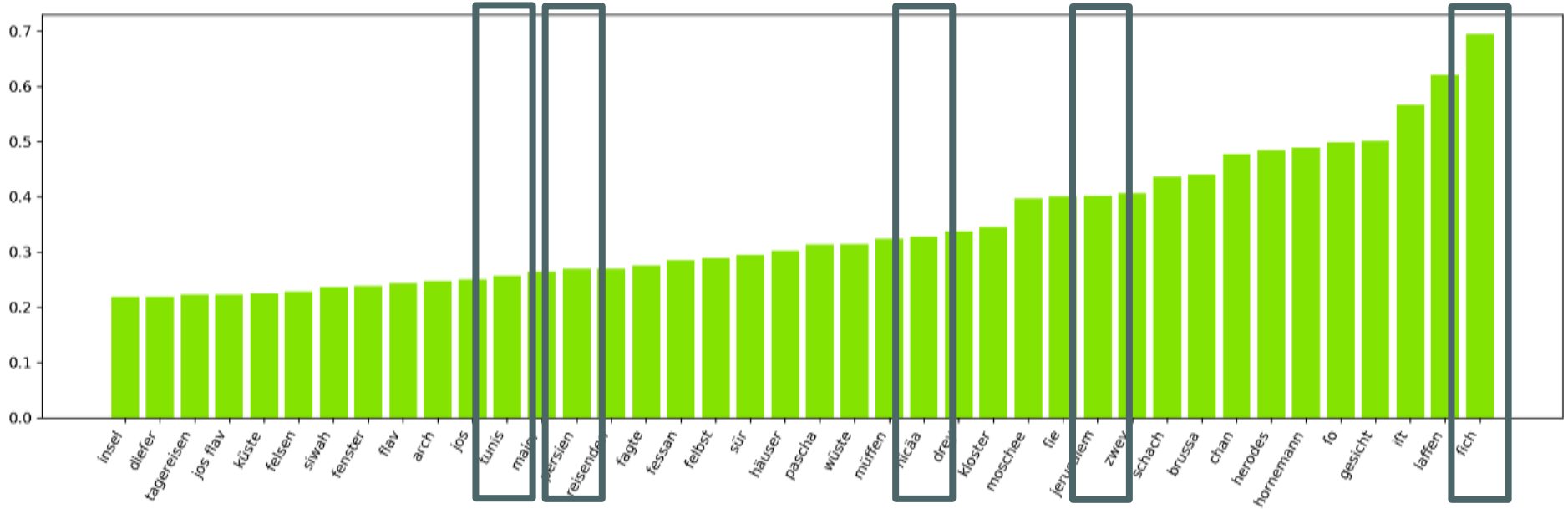
Travelogues

PRELIMINARY RESULTS | CONFUSION MATRIX



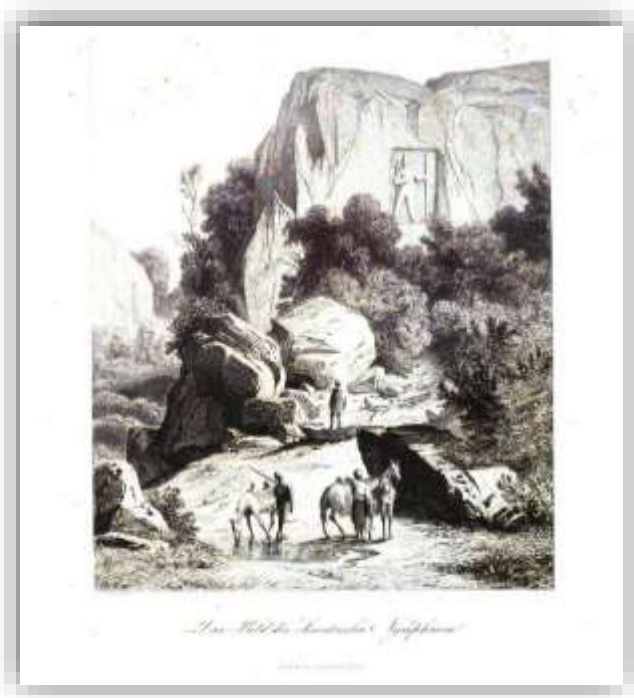
Travelogues

PRELIMINARY RESULTS | FEATURES



OVERVIEW

- The Travelogues project
- Technical research tasks
- Preliminary results
- Next steps



NEXT STEPS

- Evaluate the effect of time and OCR quality on classification performance
- Apply trained classifiers on the whole corpus
- Let experts annotate semantic structures denoting otherness
- Investigate and evaluate the use of neural networks (for classification and information extraction)
- Construct a knowledge graph on travelogues and the notion of otherness
  - Available at the ONB, to facilitate further research

Travelogues

Perceptions of the Other – A Computerized Analysis

www.travelogues-project.info

jan.roerden@ait.ac.at