GEOSPATIAL DIGITAL HUMANITIES USE CASES

Using machine learning, large language models, and generative AI

Sven Schlarb

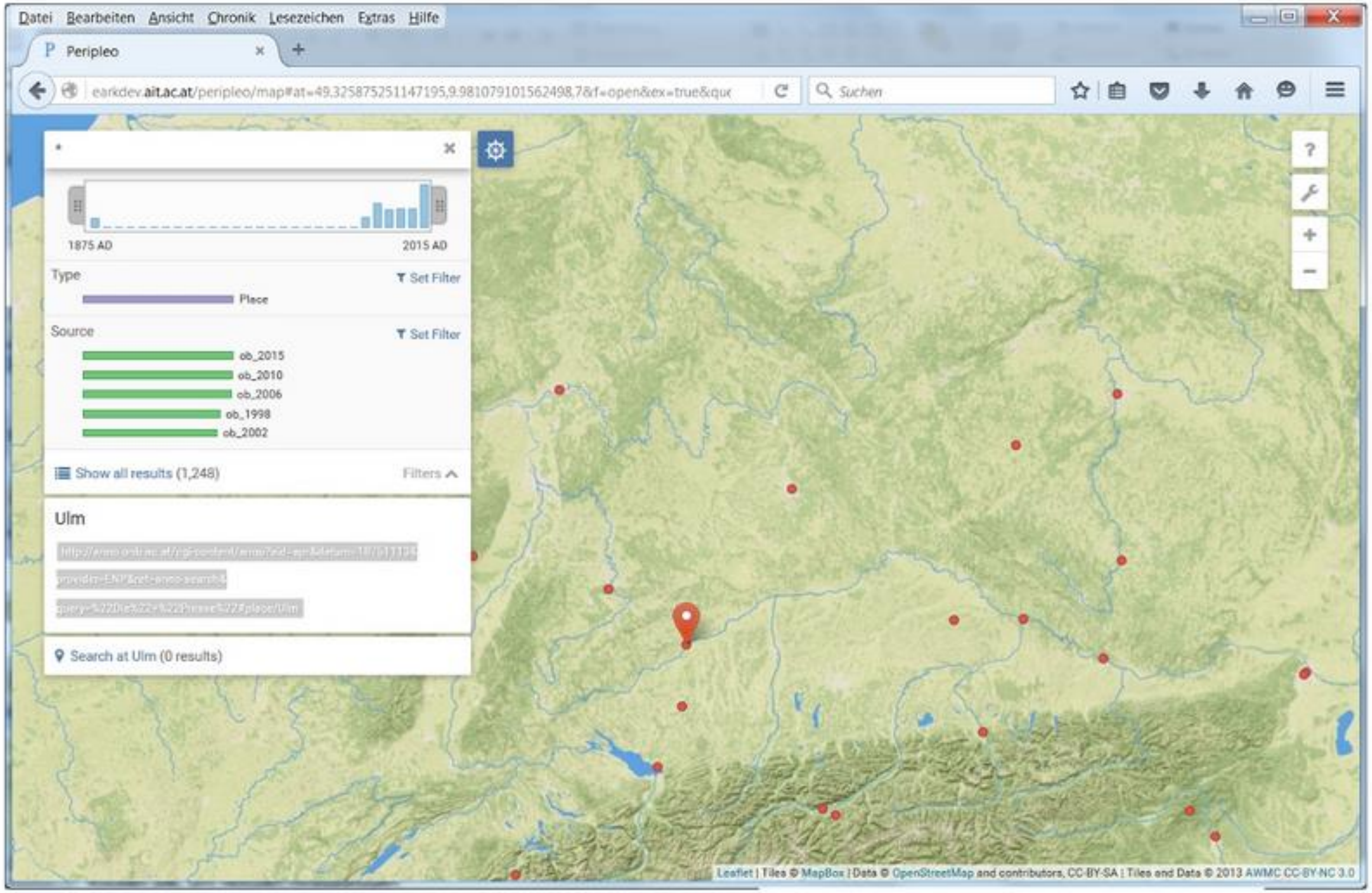
DLM Forum Members’ Meeting in Brussels 2024, Tuesday, 28th May 2024

State Archives of Belgium, Rue de Ruysbroeck 2, 1000 Brussels

Co-funded by
the European Union

Location names occurring in texts

AUSTRIAN INSTITUTE OF TECHNOLOGY



Peripleo - PELAGIOS Project

E-ARK datamining showcase

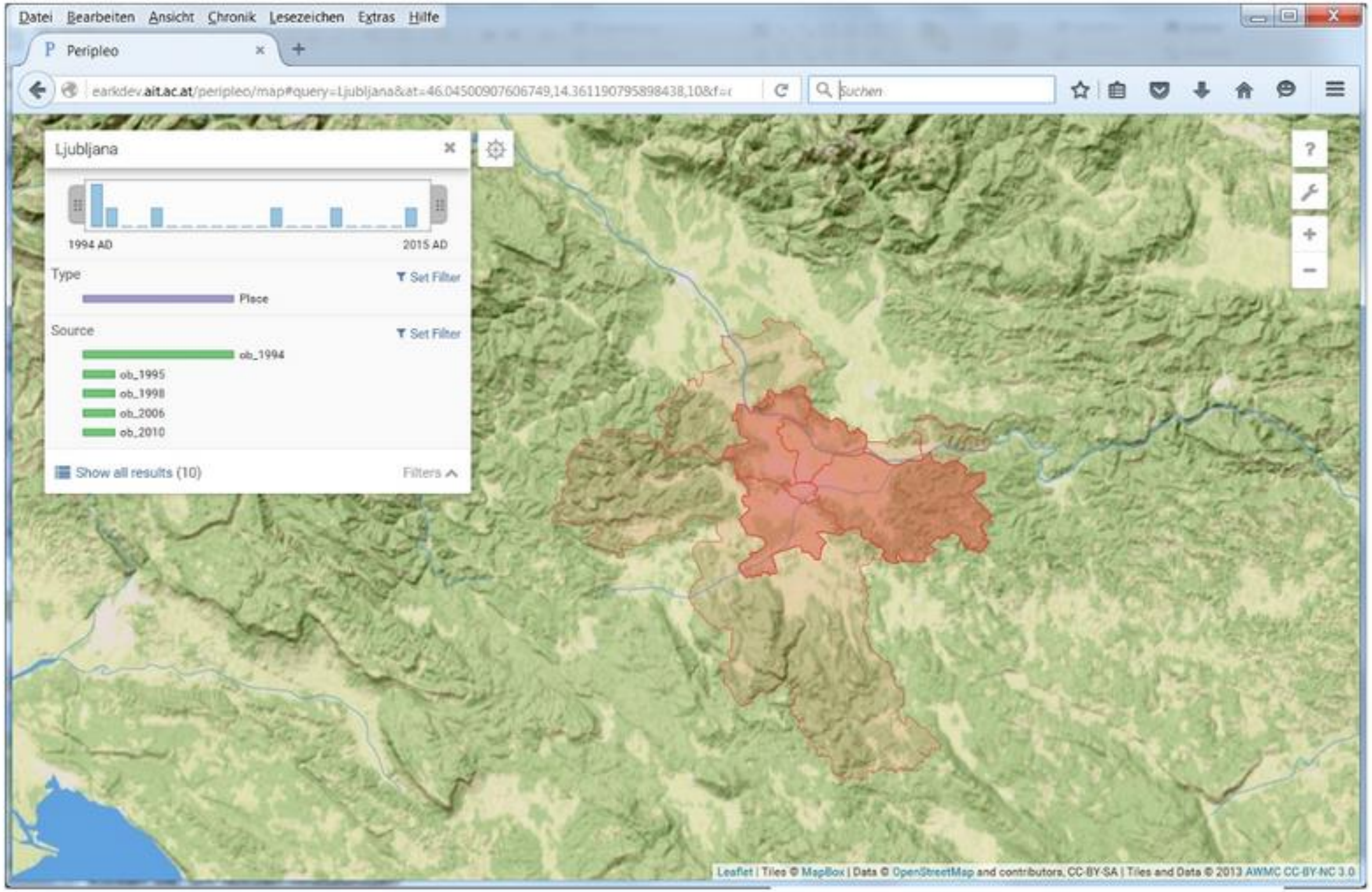
Presented at iPRES 2016

Peripleo lead developer: Rainer Simon

E-ARK

Co-funded by the European Union

Geographical/timeline search



Peripleo - PELAGIOS Project

E-ARK datamining showcase

Presented at iPRES 2016

Peripleo lead developer: Rainer Simon

Co-funded by the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY

WHAT IS A LANGUAGE MODEL

This evening I am going to the

cinema

cinema
theater
restaurant

$$\underbrace{P(\overbrace{\text{"cinema"}}^{\text{What is the probability of this token?}} \mid \overbrace{\text{"going to the"}}^{\text{Given these preceeding tokens.}})}_{\text{Conditional Probability}} = \frac{\text{Count("going to the cinema")}}{\text{Count("going to the")}} = \frac{100}{1000} = 0{,}1$$

E-ARK
Co-funded by the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY

TRAINING DATA IN A NEW DIMENSION

WWW and Social Media

Foundation Model

GPT-3 was trained on about 45 terabytes of text data; Common Crawl (60%), Web (22%), Books (16%), Wikipedia (3%)

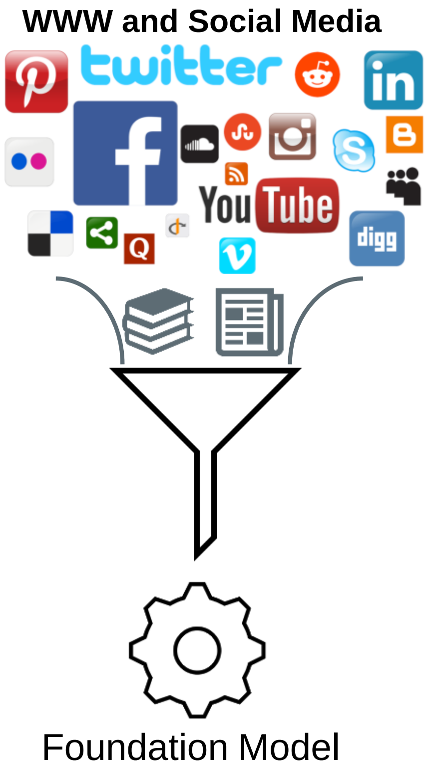
≙ 2.9 billion pages in MS Word

≈ 10 Million Books

Library of Congress in Washington, D.C., United States: 170 million items

E-ARK
Co-funded by the European Union

TRAINING DATA IN A NEW DIMENSION

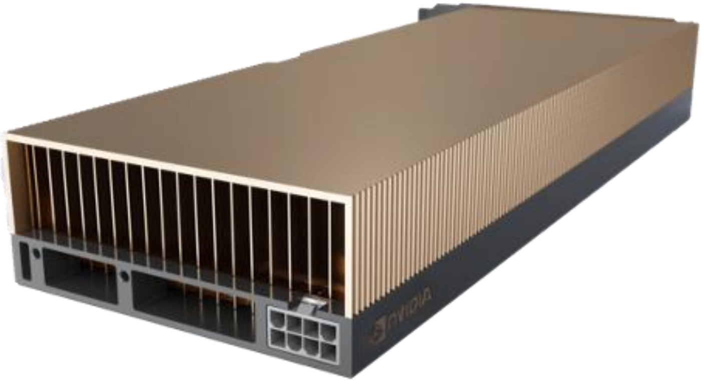


GPT-4 uses a more diverse and larger database of about 1 petabyte

Source: https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495

According to OpenAI the cost was about $100 million, and the training took 100 days, utilizing 25000 NVIDIA A100 GPUs.

Source: https://medium.com/@daniellefranca96/gpt4-all-details-leaked-48fa20f9a4a



19.058 82 €

Lieferung für 6,10 € **23. - 28. Mai**. Details

Liefern nach Österreich

Nur noch 2 auf Lager

E-ARK
Co-funded by the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY

GEOAI IS MULTIMODAL

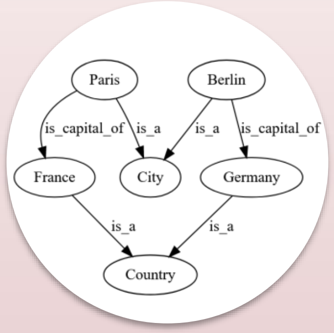
Text

Location descriptions

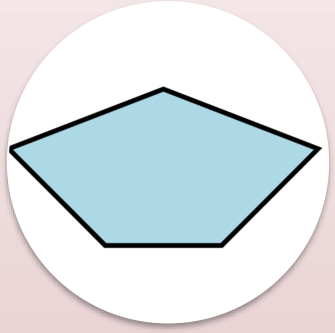
Image

Satellite, Street View

Trajectory, Route



Knowledge Graph

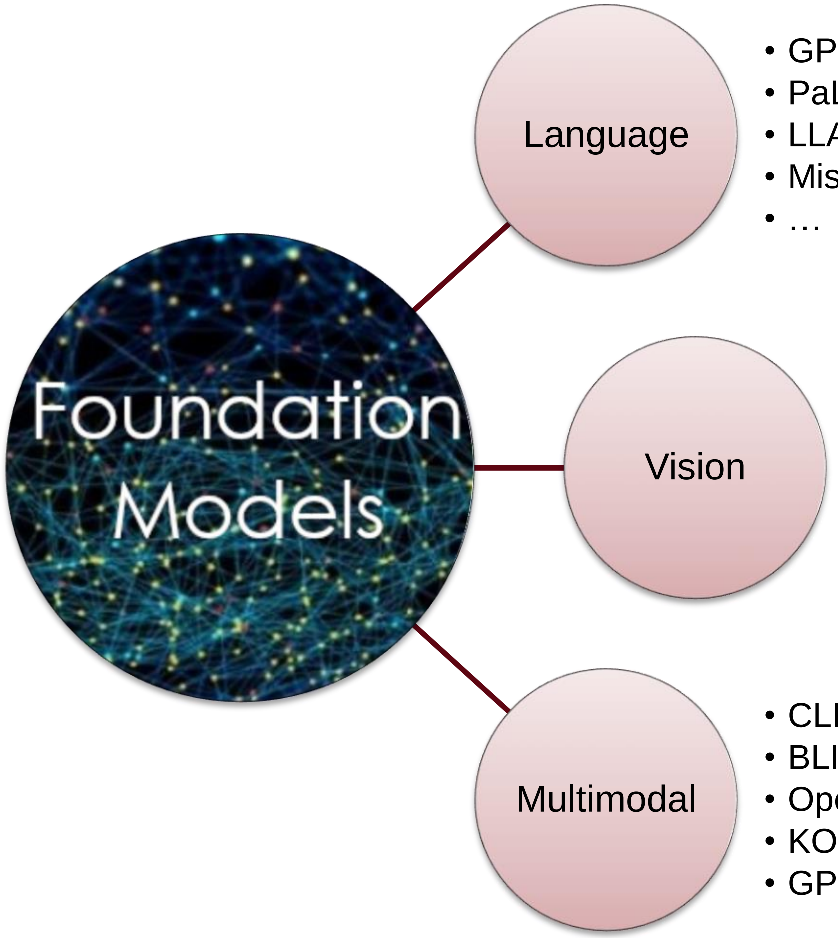


Vector Data (Poligons, Lines)

E-ARK
Co-funded by the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY

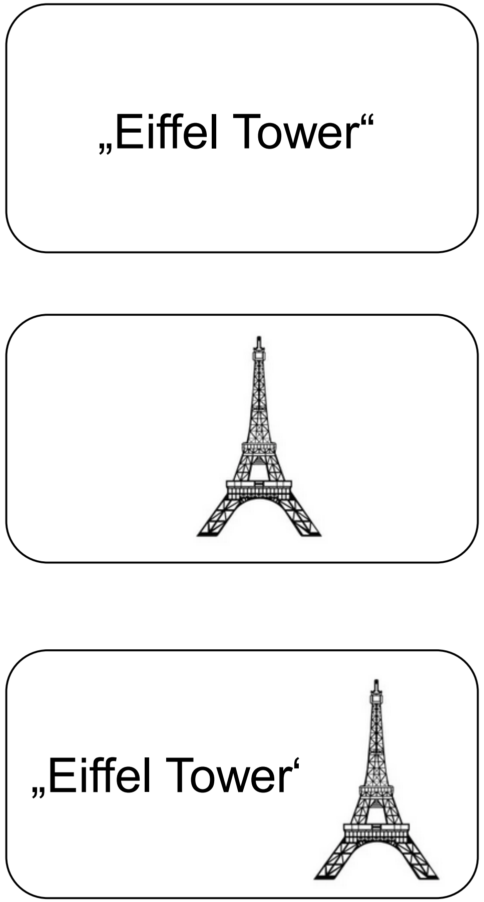
MULTIMODAL FOUNDATION MODEL



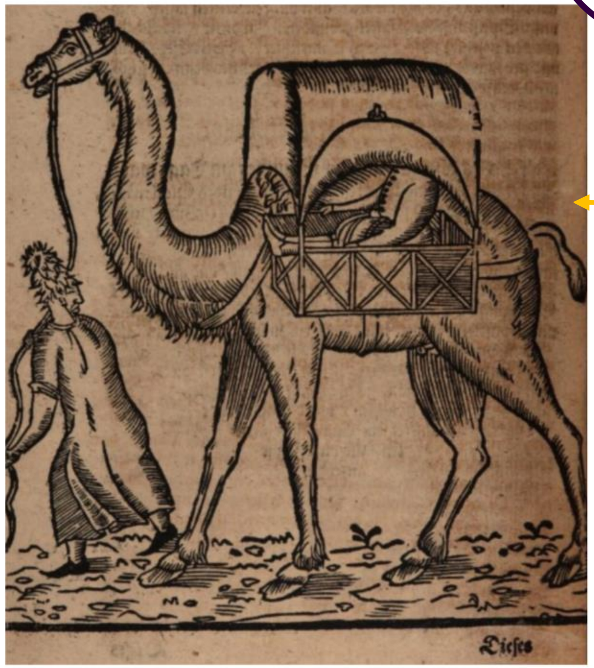
- GPT-3
- PaLM
- LLAMA
- Mistral
- …

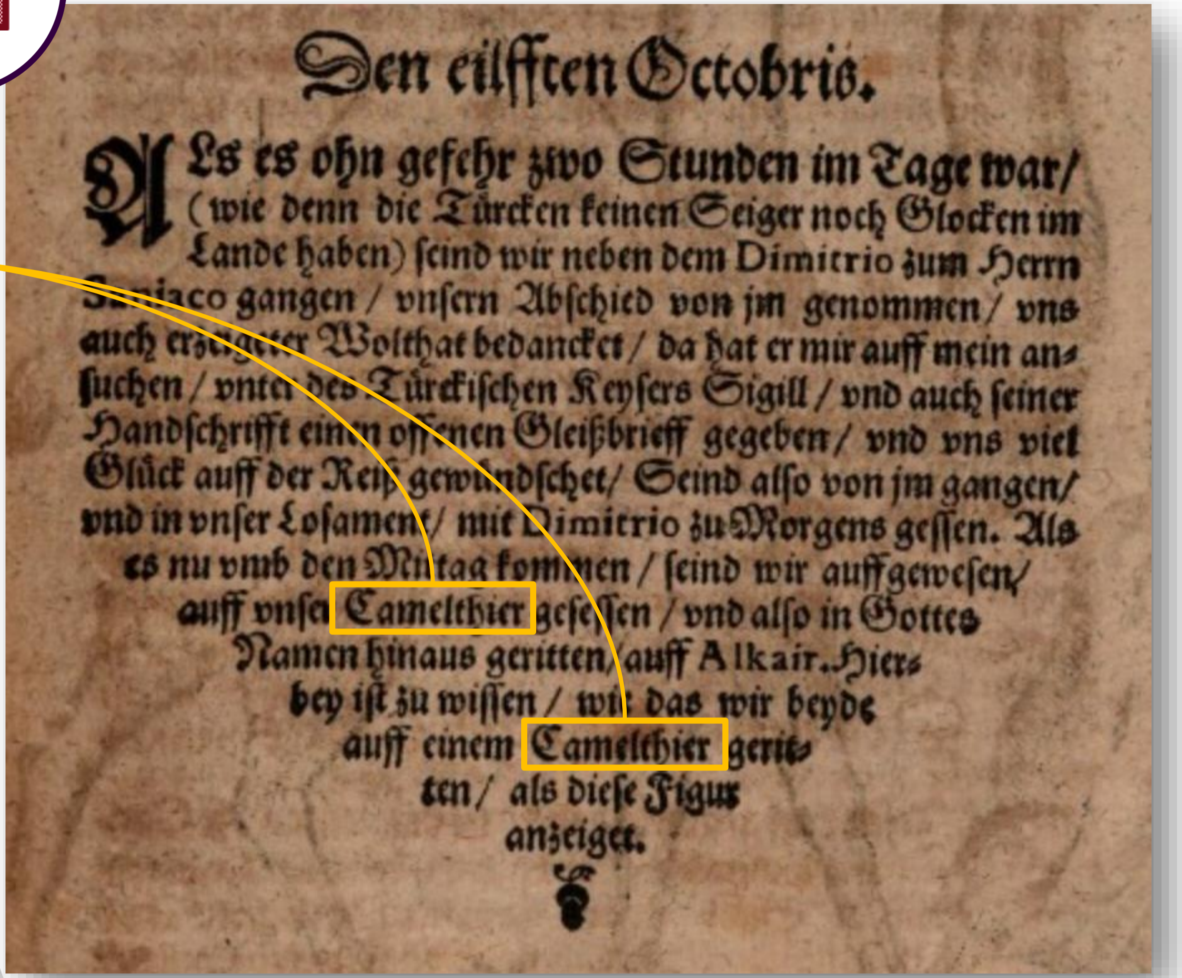
- Imagen
- StableDiffusion
- DALL-E2
- SAM
- …

- CLIP, OpenCLIP
- BLIP
- OpenFlamingo
- KOSMOS-1
- GPT-4



Images and text are encoded into a **shared latent space** where similar concepts from different modalities are close to each other.





Den eilfften Octobris.

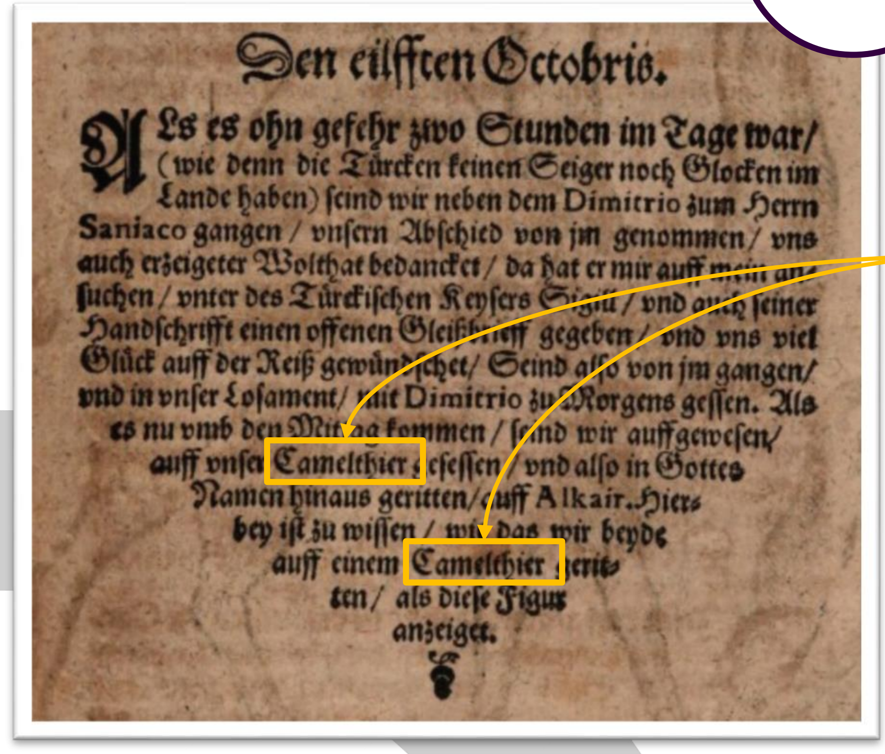
ALs es ohn gefehr zwo Stunden im Tage war/ (wie denn die Türcken keinen Seiger noch Glocken im Lande haben) seind wir neben dem Dimitrio zum Herrn Sangiaco gangen / vnsern Abschied von jm genommen / vns auch erzeigeter Wolthat bedancket / da hat er mir auff mein ansuchen / vnter des Türckischen Keysers Sigill / vnd auch seiner Handschrifft einen offenen Gleißbrieff gegeben / vnd vns viel Glück auff der Reiß gewündschet/ Seind also von jm gangen/ vnd in vnser Losament/ mit Dimitrio zu Morgens gessen. Als es nu vmb den Mittag kommen / seind wir auffgewesen/ auff vnser Camelthier gesessen / vnd also in Gottes Namen hinaus geritten/ auff Alkair. Hierbey ist zu wissen / wie das wir beyde auff einem Camelthier geritten / als diese Figur anzeiget.

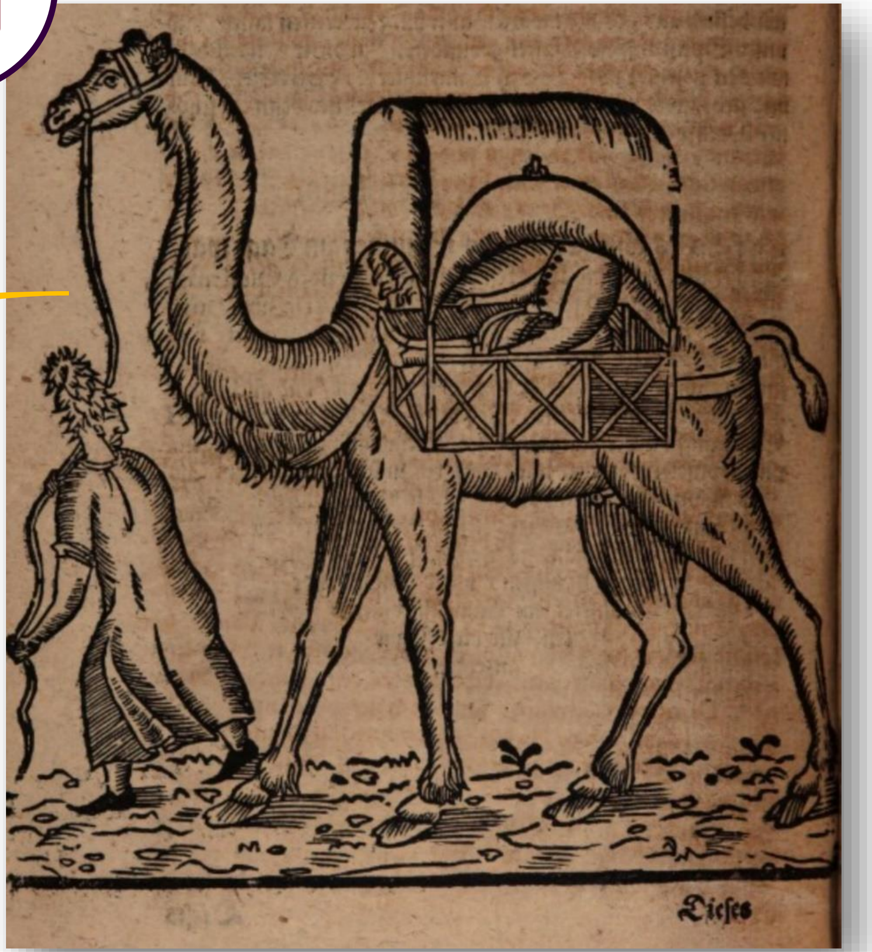
Bildquelle: http://data.onb.ac.at/rep/1032B57A, pp. 91-92

THE ONIT-PROJECT

– **Goal:** Analyse representations of "nature" in travel accounts on the Ottoman empire

– **Innovation:** Interdisciplinary method for AI-driven analysis of image-text relations in digitised historical prints


Den eilfften Octobris.
ALs es ohn gefehr zwo Stunden im Tage war/
(wie denn die Türcken keinen Seiger noch Glocken im
Lande haben) seind wir neben dem Dimitrio zum Herrn
Saniaco gangen / vnsern Abschied von jm genommen / vns
auch erzeigeter Wolthat bedancket / da hat er mir auff mein an-
suchen / vnter des Türckischen Keysers Sigel / vnd auch seiner
Handschrifft einen offenen Gleitsbrieff gegeben / vnd vns viel
Glück auff der Reiß gewünschet/ Seind also von jm gangen/
vnd in vnser Losament/ mit Dimitrio zu Morgens gessen. Als
es nu vmb den Mittag kommen / seind wir auffgewesen/
auff vnsere Camelthier gesessen / vnd also in Gottes
Namen hinaus geritten/ auff Alkair. Hier-
bey ist zu wissen / weil das wir beyde
auff einem Camelthier gerit-
ten/ als diese Figur
anzeiget.



Bildquelle: http://data.onb.ac.at/rep/1032B57A, pp. 91-92

THE ONIT-PROJECT

Corpus: Western Travelogues to Ottoman empire

- 2500 early modern book prints (1501–1850)
- 30’000 images (woodcuts, engravings, etc.)
- 4 languages (DE, FR, EN, LAT)
- **Analysis:** Role of nature representations?
- Flora, fauna, landscapes, maps

Mag. Michela VIGNOLI
AIT Austrian Institute of Technology

ONiT Similarity Explorer

Search by Text:

Search...

Number of Results:

50

Search

Clear

Select Corpora:

All	D16	D17	D18
D19	E17	E18	E19
F16	F17	F18	F19
L16	L17	L18	L19

Begin your search by entering text in the form located at the top left.
To search for similar images, click on the magnifying glass icon (Q) associated with an image.
To view more details of an image, simply click on it.

Funded by the Austrian Science Fund (FWF: P 35245).

WHICH “VIENNA”?

1 If you’ve read many of these articles about things to see and do around a Stuckey’s, you’ll notice that nothing turned a fledgling town into a boom town quite like the railroad. Vienna was no exception. Where families worked to just to survive before the railroad, after the Cairo and Vincennes Railroad was built through Vienna around 1872, the town and its people became more prosperous as they were now able to ship goods farther across the nation.

2 Vienna is no Berlin when it comes to club culture, but there are a few places where it comes close. If you’re looking for a dance-until-dawn, hard-techno experience, Das Werk, on the banks of the Danube Canal, is the place to go (though unlike some exclusive clubs in Berlin, Das Werk’s bouncers do not care how you dress). Here, the party doesn’t start before 11 p.m., when the crowd becomes hypnotized by the D.J. — or by the light show on the walls around her.

E-ARK

Co-funded by the European Union

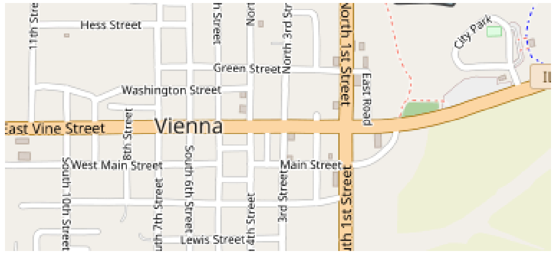
AUSTRIAN INSTITUTE OF TECHNOLOGY

KNOWLEDGE GRAPH

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX dbr: <http://dbpedia.org/resource/>
```

```
SELECT ?city
WHERE {
  ?city a dbo:City ;
        dbo:country dbr:United_States ;
        rdfs:label ?label .
  FILTER(CONTAINS(LCASE(?label), "vienna"))
}
LIMIT 1
```

https://dbpedia.org/page/**Vienna,_Illinois**



Vienna, Illinois, United States

```
SELECT ?city
WHERE {
  ?city a dbo:City ;
        dbo:country dbr:Austria ;
        rdfs:label ?label .
  FILTER(CONTAINS(LCASE(?label), "vienna"))
}
LIMIT 1
```

https://dbpedia.org/page/**Vienna**

Vienna, Austria

https://dbpedia.org/sparql

E-ARK

Co-funded by the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY

WHICH "VIENNA"?

1 If you've read many of these articles about things to see and do around a Stuckey's, you'll notice that nothing turned a fledgling town into a boom town quite like the railroad. Vienna was no exception. Where families worked to just to survive before the railroad, after the Cairo and Vincennes Railroad was built through Vienna around 1872, the town and its people became more prosperous as they were now able to ship goods farther across the nation.

2 Vienna is no Berlin when it comes to club culture, but there are a few places where it comes close. If you're looking for a dance-until-dawn, hard-techno experience, Das Werk, on the banks of the Danube Canal, is the place to go (though unlike some exclusive clubs in Berlin, Das Werk's bouncers do not care how you dress). Here, the party doesn't start before 11 p.m., when the crowd becomes hypnotized by the D.J. — or by the light show on the walls around her.

Let's ask a well known LLM …

The text is likely referring to Vienna, a town in Illinois, United States. This conclusion is drawn from the mention of the Cairo and Vincennes Railroad, which ran through Vienna, Illinois, around 1872, facilitating the shipment of goods across the nation.

Vienna, Illinois, United States

Based on the mention of the Danube Canal and the club culture described, it's highly likely that the text is referring to Vienna, the capital city of Austria.

Vienna, Austria

E-ARK

Co-funded by the European Union

TOPONYM/LOCATION DESCRIPTION RECOGNITION

	Model	#Param	Toponym Recognition		Location Description Recognition		
			Hu2014	Ju2016	HaveyTweet2017		
			Accuracy ↓	Accuracy ↓	Precision ↓	Recall ↓	F-Score ↓
(A)	Stanford NER (nar. loc.) [30]	-	0.787	0.010	0.828	0.399	0.539
	Stanford NER (bro. loc.) [30]	-	-	0.012	0.729	0.44	0.548
	Retrained Stanford NER [30]	-	-	0.078	0.604	0.410	0.489
	Caseless Stanford NER (nar. loc.) [30]	-	-	0.460	0.803	0.320	0.458
	Caseless Stanford NER (bro. loc.) [30]	-	-	0.514	0.721	0.336	0.460
	spaCy NER (nar. loc.) [44]	-	0.681	0.000	0.575	0.024	0.046
	spaCy NER (bro. loc.) [44]	-	-	0.006	0.461	0.304	0.366
	DBpedia Spotlight[99]	-	0.688	0.447	-	-	-
(B)	Edinburgh [7]	-	0.656	0.000	-	-	-
	CLAVIN [134]	-	0.650	0.000	-	-	-
	TopoCluster [23]	-	0.794	0.158	-	-	-
(C)	CamCoder [33]	-	0.637	0.004	-	-	-
	Basic BiLSTM+CRF [77]	-	-	0.595	0.703	0.600	0.649
	DM NLP (top. rec.) [139]	-	-	0.723	0.729	0.680	0.703
	NeuroTPR [135]	-	0.675†	0.821	0.787	0.678	**0.728**
(D)	GPT2 [115]	117M	0.556	0.650	0.540	0.413	0.468
	GPT2-Medium [115]	345M	0.806	0.802	0.529	0.503	0.515
	GPT2-Large [115]	774M	0.813	0.779	0.598	0.458	0.518
	GPT2-XL [115]	1558M	0.869	**0.846**	0.492	0.470	0.481
	GPT-3 [15]	175B	**0.881**	0.811*	0.603	**0.724**	0.658
	InstructGPT [106]	175B	0.863	0.817*	0.567	0.688	0.622
	ChatGPT (Raw.) [104]	176B	0.800	0.696*	0.516	0.654	0.577
	ChatGPT (Con.) [104]	176B	0.806	0.656*	0.548	0.665	0.601

[Mai et al. 2023]

NeuroTPR	Supervised training on 599 labeled tweets and labeled sentences generated from 3000 Wikipedia articles.
GPT2-XL	Using 8 few-shot examples in the prompt.

[Mai et al. 2023, p. 6]

PROMPT

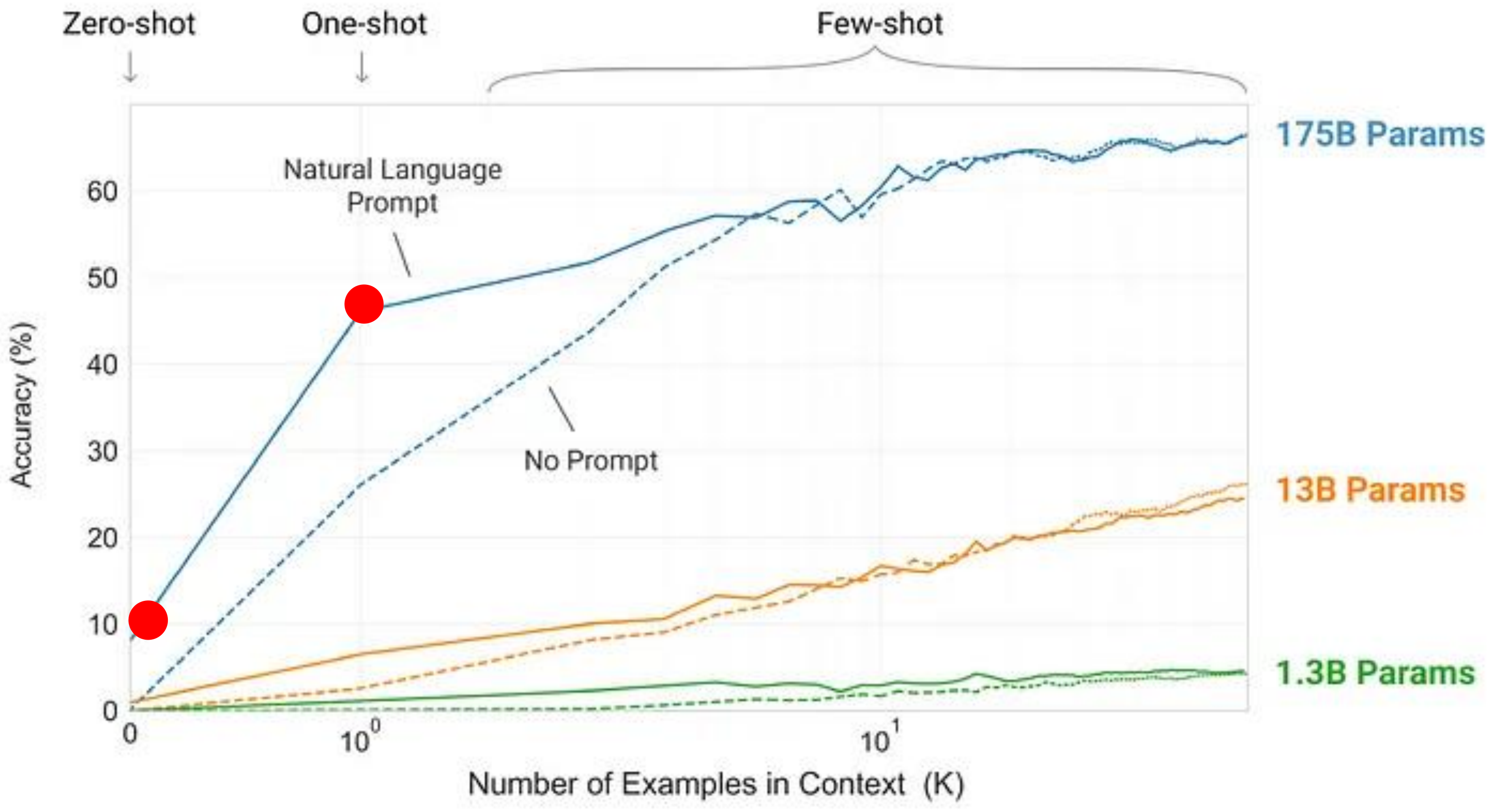
Input Text

In June 2013, heavy flooding affected Central Europe, with significant impacts on cities such as Prague in the Czech Republic and Dresden in Germany. The Elbe River overflowed its banks, reaching a peak flow at 50.1° N, 14.4° E. Many bridges in Prague were closed, and the historic city center of Dresden was submerged under several feet of water. Rescue operations were launched across affected regions, including Austria and Hungary.

Example Output

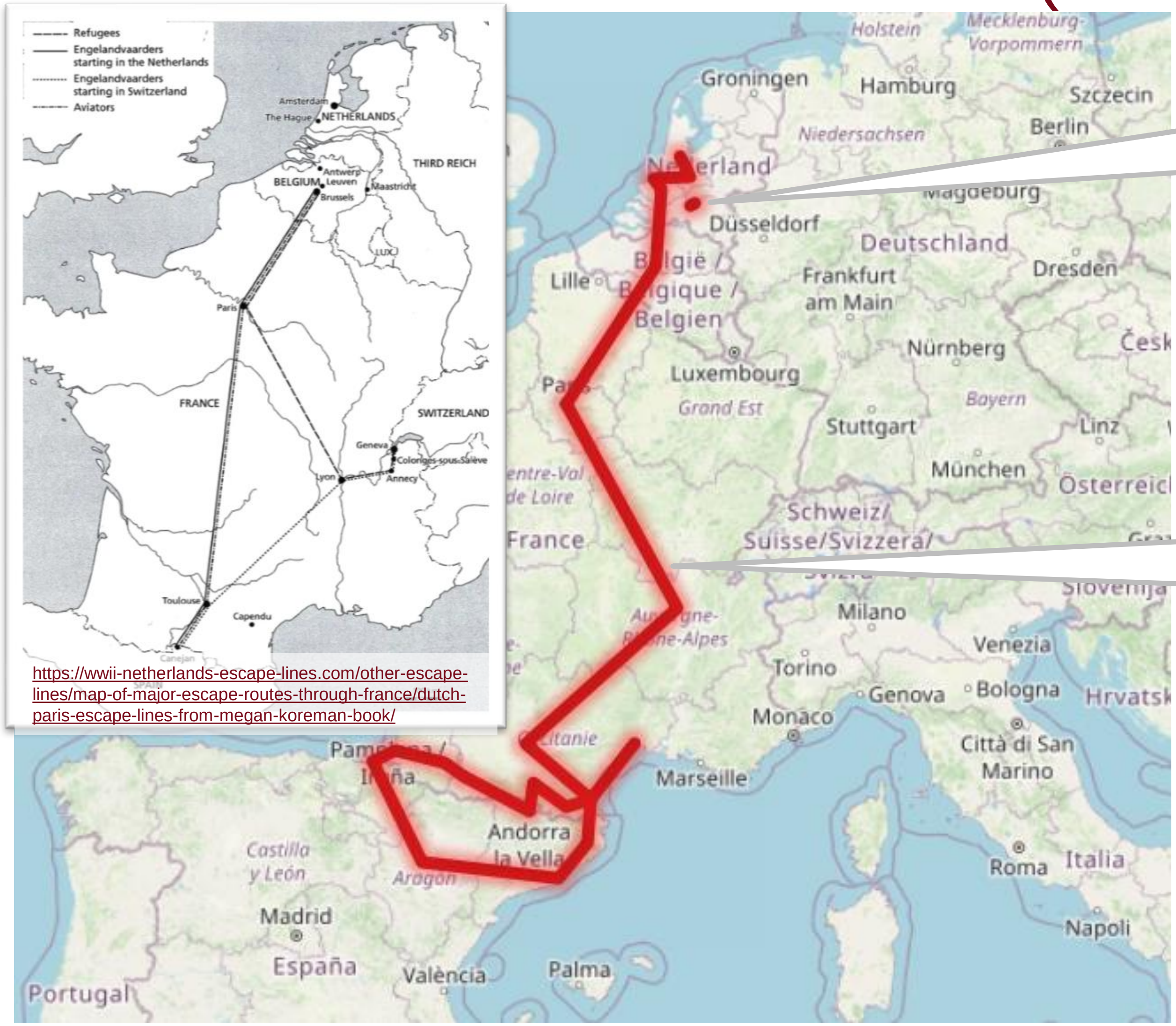
Locations	Central Europe, Prague, Czech Republic, Dresden, Germany, Elbe River, Austria, Hungary
Coordinates	50.1° N, 14.4° E
Distances and Directions	several feet of water
Geographical Features	Elbe River
Human-made Features	bridges, historic city center
Events	June 2013 heavy flooding, Elbe River overflow, bridge closures, rescue operations

CONTEXTUAL INFORMATION (FEW-SHOT)



[Tsang 2022]

DUTCH-PARIS ESCAPE LINE (WW II)



https://wwii-netherlands-escape-lines.com/other-escape-lines/map-of-major-escape-routes-through-france/dutch-paris-escape-lines-from-megan-koreman-book/

Can you give me the GeoJSON of the Dutch-Paris Escape Line?

Can you list the cities through which the Dutch-Paris Escape Line went?

Can you create a path in GeoJSON that connects all these cities?

```
"coordinates": [
    [4.895, 52.370],  // Amsterdam, Netherlands
    [5.121, 52.090],  // Utrecht, Netherlands
    [4.300, 52.067],  // The Hague (Den Haag), Netherlands
    [4.477, 51.922],  // Rotterdam, Netherlands
    [4.352, 50.850],  // Brussels, Belgium
    [2.352, 48.856],  // Paris, France
    [4.835, 45.764],  // Lyon, France
    [1.444, 43.604],  // Toulouse, France
    [2.895, 42.698],  // Perpignan, France
    [2.347, 42.573],  // Carcassonne, France
    [1.610, 42.966],  // Foix, France
    [1.521, 42.505],  // Andorra la Vella, Andorra
    [0.045, 43.076],  // Lourdes, France
    [-0.371, 43.295], // Pau, France
    [-1.474, 43.494], // Bayonne, France
    [-1.986, 43.321], // San Sebastián (Donostia), Spain
    [-1.645, 42.816], // Pamplona (Iruña), Spain
    [-0.880, 41.656], // Zaragoza, Spain
    [2.176, 41.385],  // Barcelona, Spain
    [2.824, 41.979],  // Girona, Spain
    [2.895, 42.698],  // Perpignan, France (again)
    [3.877, 43.611]   // Andorra la Vella, Andorra (again)
  ]
```

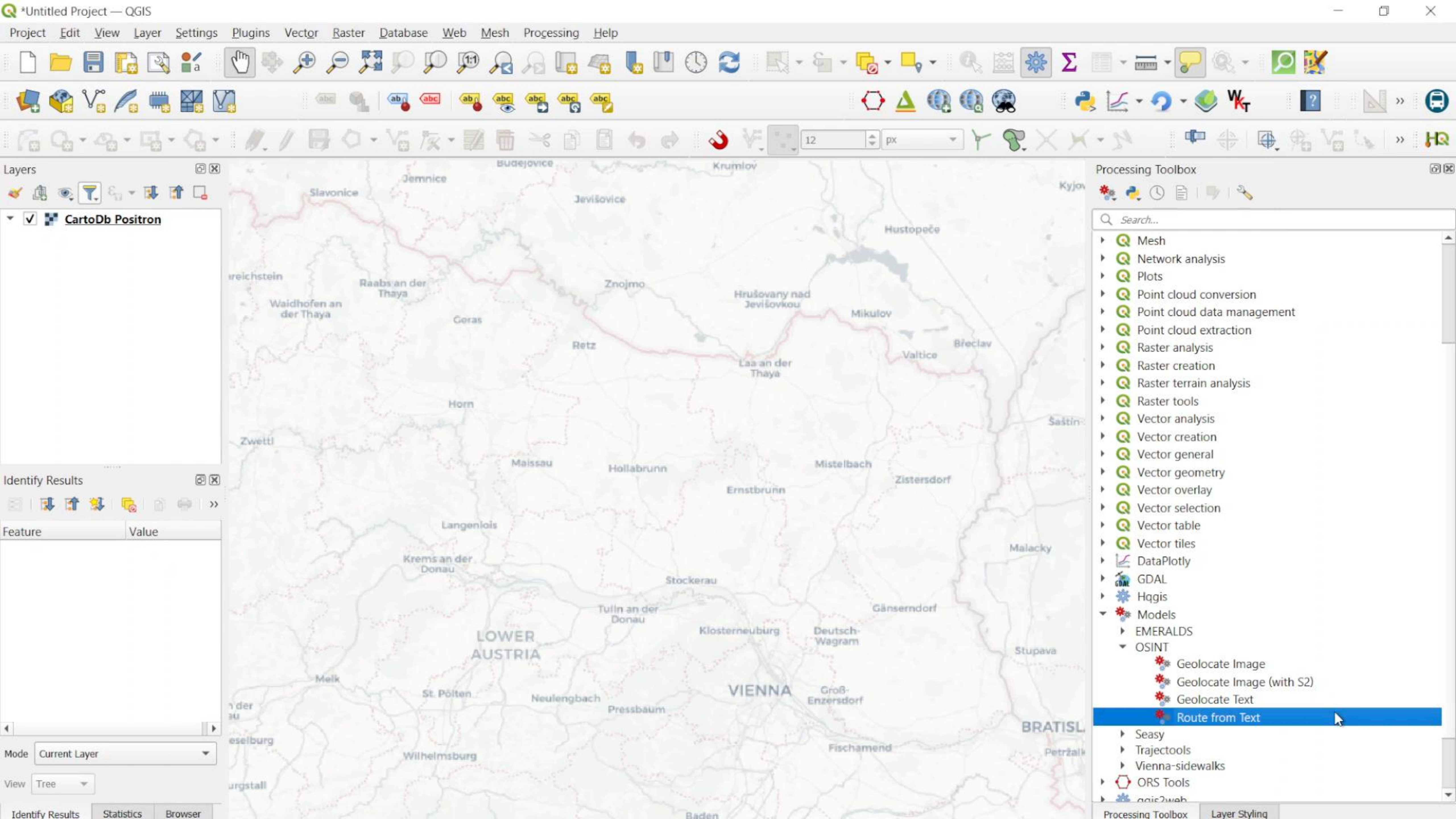
QGIS LLM ROUTE EXTRACTION

by Anita Graser

The provisional appointment date for 72 April 2024 in St Pölten has been set for those born in 2006 in the municipality of Poysdorf. On this occasion, the municipality of Poysdorf is organising a bus trip to St. Pölten and back in cooperation with the company Schiefer Personenbeförderung GmbH. The bus will be provided free of charge by the municipality.

I am born in Retz, studied in Tulln, and now I live in Vienna.

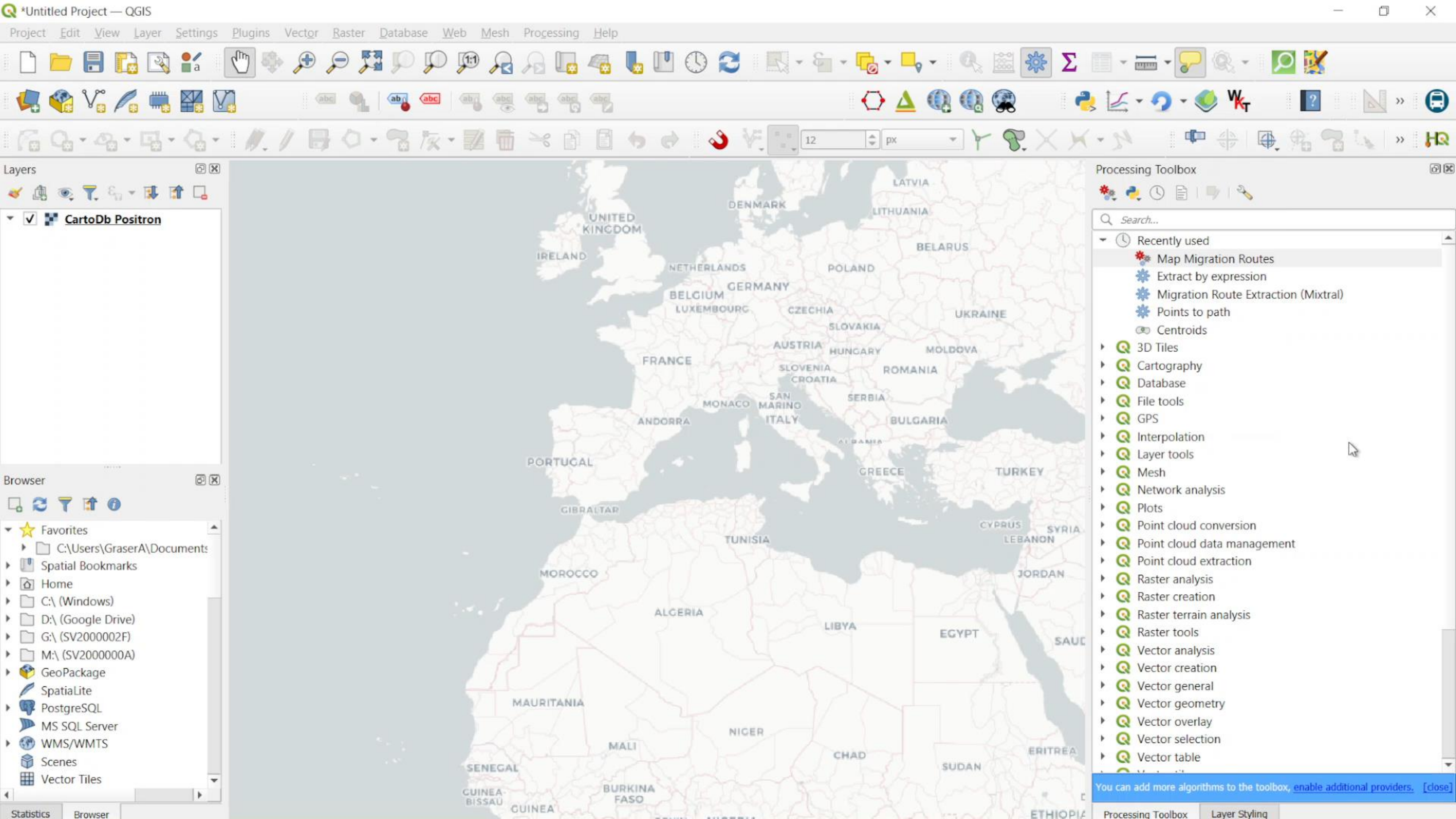
Co-funded by
the European Union


*Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
12
px
Layers
CartoDb Positron
Identify Results
Feature
Value
Mode
Current Layer
View
Tree
Identify Results
Statistics
Browser
Budejovice
Krumlov
Jemnice
Slavonice
Jevišovice
Kyjov
Hustopeče
reichstein
Raabs an der Thaya
Znojmo
Hrušovany nad Jevišovkou
Waidhofen an der Thaya
Mikulov
Geras
Retz
Břeclav
Valtice
Laa an der Thaya
Horn
Šaštín
Zwettl
Maissau
Hollabrunn
Mistelbach
Zistersdorf
Ernstbrunn
Langenlois
Malacky
Krems an der Donau
Stockerau
Gänserndorf
Tulln an der Donau
LOWER AUSTRIA
Klosterneuburg
Deutsch-Wagram
Stupava
Melk
VIENNA
St. Pölten
Neulengbach
Pressbaum
Groß-Enzersdorf
BRATISL
Fischamend
Petržalk
Wilhelmsburg
Baden
Processing Toolbox
Search...
Mesh
Network analysis
Plots
Point cloud conversion
Point cloud data management
Point cloud extraction
Raster analysis
Raster creation
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Vector selection
Vector table
Vector tiles
DataPlotly
GDAL
Hqgis
Models
EMERALDS
OSINT
Geolocate Image
Geolocate Image (with S2)
Geolocate Text
Route from Text
Seasy
Trajectools
Vienna-sidewalks
ORS Tools
qgis2web
Processing Toolbox
Layer Styling

QGIS LLM ROUTE EXTRACTION by Anita Graser

The western Mediterranean route consists of two major sub-routes. One leads across the westernmost part of the Mediterranean or the Strait of Gibraltar from Morocco and Algeria towards the Spanish mainland. Migrants also occasionally land on the Balearic Islands. This sub-route is primarily used by migrants of Algerian and Moroccan nationality. The second sub-route leads from Morocco, Western Sahara and Mauritania to the Canary Islands. It is primarily Moroccan nationals and citizens of various West African countries who land here. Migration from North Africa to Spain (directly, via Ceuta, via Melilla or via the Canary Islands) fell significantly compared to the previous year. A total of around 31,800 people travelled to Europe in this way (previous year 43,200). The proportionate distribution among the sub-routes largely corresponds to that of the previous year. Around 15,700 people arrived in the Canary Islands (previous year 23,050) and 14,200 in mainland Spain and the Balearic Islands (previous year 18,930). 1,900 people arrived in the two exclaves of Ceuta and Melilla, 650 more than in the previous year. Spanish-Moroccan relations improved significantly over the course of 2022. This is likely to be one of the reasons for the decline in Uber journeys to the Canary Islands and, in particular, the slump in landings in November and December. This route is of secondary importance for migration to Switzerland.

E-ARK
Co-funded by the European Union


*Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Layers
CartoDb Positron
Browser
Favorites
C:\Users\GraserA\Documents
Spatial Bookmarks
Home
C:\ (Windows)
D:\ (Google Drive)
G:\ (SV2000002F)
M:\ (SV2000000A)
GeoPackage
SpatiaLite
PostgreSQL
MS SQL Server
WMS/WMTS
Scenes
Vector Tiles
Statistics
Browser
Processing Toolbox
Search...
Recently used
Map Migration Routes
Extract by expression
Migration Route Extraction (Mixtral)
Points to path
Centroids
3D Tiles
Cartography
Database
File tools
GPS
Interpolation
Layer tools
Mesh
Network analysis
Plots
Point cloud conversion
Point cloud data management
Point cloud extraction
Raster analysis
Raster creation
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Vector selection
Vector table
You can add more algorithms to the toolbox, enable additional providers. [close]
Processing Toolbox
Layer Styling

THANKS TO MY COLLEAGUES AT AIT

MICHELA VIGNOLI
Scientist
Data Science & Artificial Intelligence
Center for Digital Safety & Security

AIT Austrian Institute of Technology GmbH
Giefinggasse 4 | 1210 Vienna | Austria
michela.vignoli@ait.ac.at | www.ait.ac.at

THE ONIT-PROJECT

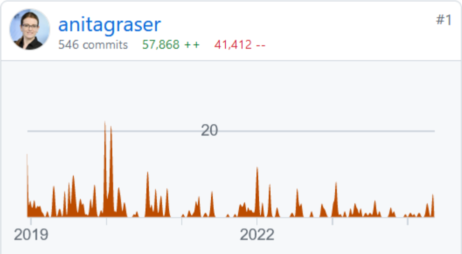
ONiT

ANITA GRASER
Senior Scientist
Data Science & Artificial Intelligence
Center for Digital Safety & Security

AIT Austrian Institute of Technology GmbH
Giefinggasse 4 | 1210 Vienna | Austria
anita.graser@ait.ac.at | www.ait.ac.at

QGIS LLM EXPERIMENTS

movingpandas



E-ARK

Co-funded by the European Union

THANK YOU!

SVEN SCHLARB
Scientist
Data Science & Artificial Intelligence
Center for Digital Safety & Security

AIT Austrian Institute of Technology GmbH
Giefinggasse 4 | 1210 Vienna | Austria
T +43 50550-4179 | M +43 664 8251379 | F +43 50550-4150
sven.schlarb@ait.ac.at | www.ait.ac.at

Co-funded by the European Union

BIBLIOGRAPHIC REFERENCES

- [Mai et al. 2023] Mai et al., "On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence." 2023.
- [Tsang 2022] Sik-Ho Tsang, „Review — GPT-3: Language Models are Few-Shot Learners", 2022.