

Everybody Lived Somewhere. Identifying Geographical names Using by AI

Simon András

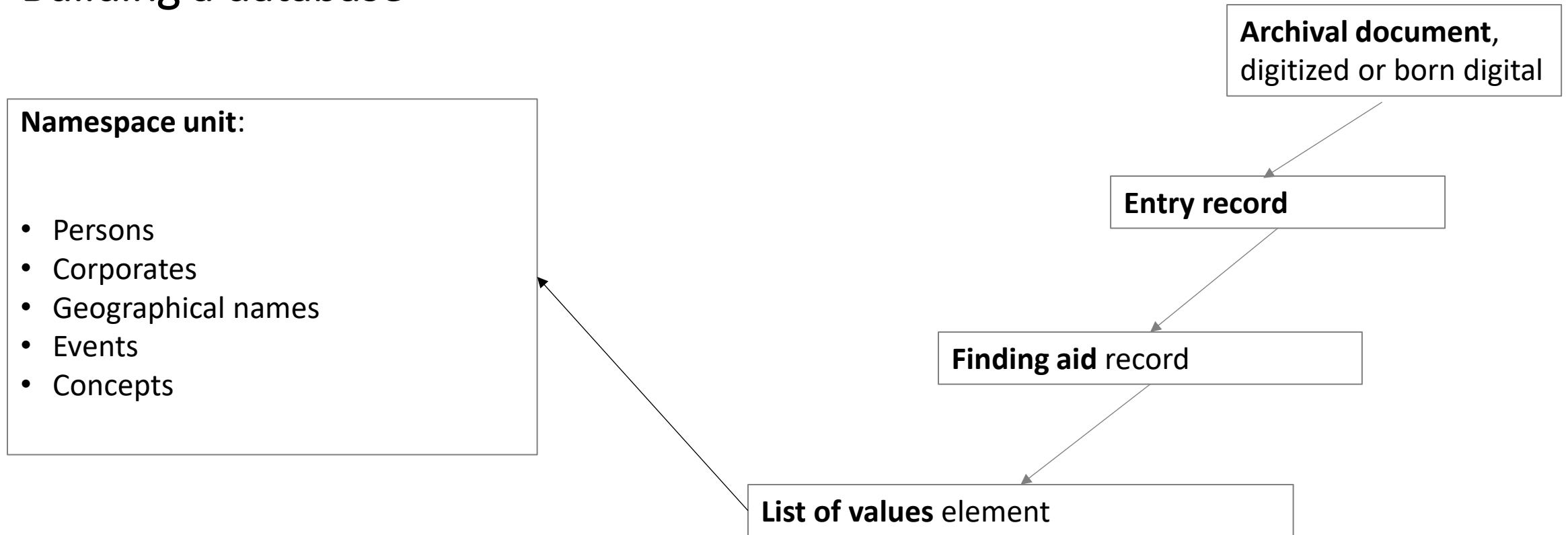
National Archives of Hungary, Digital Services Department

2026

Budapest



Building a database



Building a namespace from geographical thesauri:

Data sources –sorted by preference

1. **Geotaurusz** – Hungarian thesaurus for geographical concepts: 109008 records
2. **MNL GEO** - Geographical database of the National Archives of Hungary: 70849 records
3. **Geonames**: 12237573 records
4. Database created from „The **digital atlas of the medieval Hungary** (Engel Pál)”: 75031 records

Name formats in thesaurus (**preferred or variant**)

Basic principles and steps for building a database of entities and data enrichment by linking entities

- Final result: One concept – One database entity, coming from one datasource.
- One database entity can be attached to various terms (name variants)
- A string can **be both the preferred term** of one concept, and **the variant** of another. (Neuhaus – Selbitz / Jindřichův Hradec)
- **Rule based connection**: identical strings of names and roles, same or almost same geographical coordinates.
- **Model based connection** using AI: Similarity of strings (using various metrics e.g. soundex, jaccard), acceptable distance between geographical locations defined by coordinates.

Building a namespace from geographical thesauri: The MIREL application

- **AI based** computing
- Creating geographical name entities: **linking thesauri elements**
- **Data enrichment:** linking finding aid records and namespace entities
- Training, validating and testing **classification model**
- **Predefined datasets** for teaching AI and for human validation (size and choosing of elements of these datasets is crucial)
- An AI application creates its **own rules for prediction**
- According to the results of validation, the **model of AI can be developed, refined.**

Building a namespace: geographical record identic

Concepts, name variants, regulations

- **Name** is special or not: Freiburg – Benalmádena
- **Length of name**: Riga, Poitiers
- Possibility of **incorrect lettering** in a data source: Daugavpils
- **Changing of name** in time. Aarhus – Århus
- Possibility of making a mistake : Genève – Geneva – Genova
- Worldwide well known name is **not the official one**: Casablanca
- **Data type „role”** – settlement, river, hill – several forms in several languages
- Versions of city names: Gdańsk, Gdansk, Danzig
- Geographical **coordinates**, can be missing or incorrect
- **Relations** – e.g. part of another geographical concept

Name variants Preferred / Variant

- Freiburg / **Fribourg**
- Freiberg / **Příbor** / Pribor
- **Freiburg** im Breisgau
- Genf / **Genève** / Geneva
- **Basel** / *Bâle* / *Bázel*
- *Kronstadt* / *Brassó* / **Braşov** / Brasov



Where is Tilsit?



Földrajzi név karbantartó

i Az MNL GEO forrásból származó helyek adatait módosíthatja. Új felvitelkor adja meg a földrajzi hely adatait legalább egy típusba sorolást.

Adminisztratív adatok

Azonosító 62801
Hitelesítési szint
Szolgáltató intézmény MNL (6)
Forrás MNL GEO (75028)

Entitás kapcsolatok

Entitás azonosító 10 038 342
A szabályalapú entitás képzés során keletkezett azonosító
Entitáshoz rendelés időpontja 2025.03.18. 15:53:03
Entitás név Szovjetszk
Entitás elemei
Szovjetszk [MNL GEO] - 60724
Tilsit [MNL GEO] - 62801
Szovjetszk [Geotaurusz] - 21104642
Tilsit [Geotaurusz] - 21088077
Sovetsk [geonames] - 14399406

Földrajzi név

Ha az alábbi szekció adatait szűrőként látja, akkor kitüntetett földrajzi nevet lát

Földrajzi név Tilsit
Csak olyan nevet adhat meg, amely nem szerepel kitüntetett névként (másik h

Történeti időszak kezdő - záró



1807



Rivers has no nationality



- Memel
- Nemunas
- Nyeman
- Neman
- Niemen
- Неман

Adminisztratív adatok

Azonosító	44389
Hitelesítési szint	
Szolgáltató intézmény	MNL (6)
Forrás	MNL GEO (75028)

Entitás kapcsolatok

Entitás azonosító	12 151 514
A szabályalapú entitás képzés során keletkezett azonosító	
Entitáshoz rendelés időpontja	2025.03.18. 15:52:42
Entitás név	Memel (folyó)
Entitás elemei	Memel (folyó) [MNL GEO] - 40178
	Memel (folyó) [Geotaurusz] - 21018955

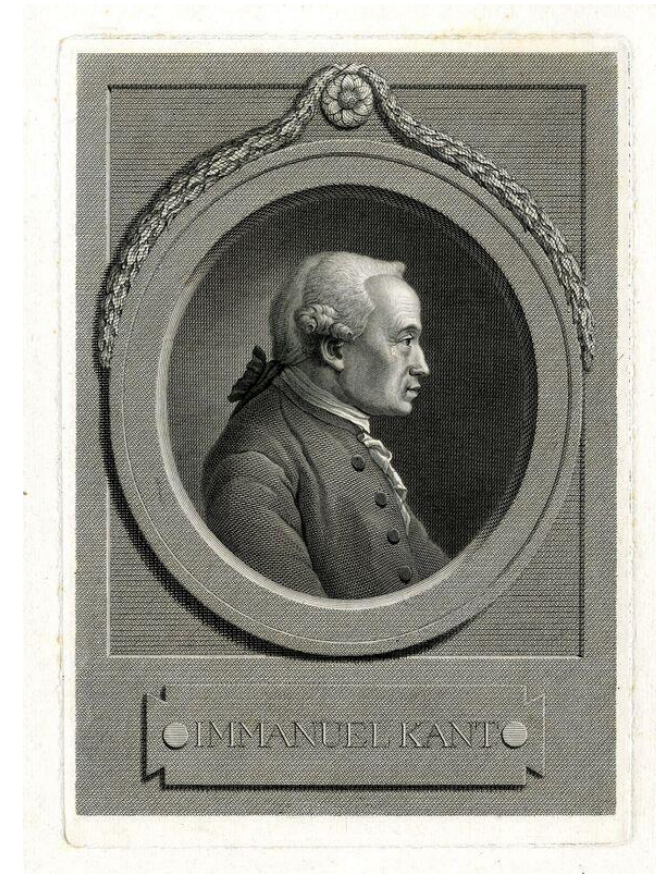
Földrajzi név

Ha az alábbi szekció adatait szűrően látja, akkor kitüntetett földrajzi nevet lát

Földrajzi név	Nemunas
Csak olyan nevet adhat meg, amely nem szerepel kitüntetett névként (másik helynél) az MNL GEO-ban	

A famous city: Königsberg

14581250	Kalińingrad	8499078	geonames	4888291	54,70649	20,51095
14581251	Kalíníngrad	8499078	geonames	4888291	54,70649	20,51095
14581252	Kaliningradas	8499078	geonames	4888291	54,70649	20,51095
14581258	Kalinyingrád	8499078	geonames	4888291	54,70649	20,51095
14581264	Königsberg	8499078	geonames	4888291	54,70649	20,51095
14581265	Königsberg in Preußen	8499078	geonames	4888291	54,70649	20,51095
14581266	Körigsberg	8499078	geonames	4888291	54,70649	20,51095
14581267	Królewiec	8499078	geonames	4888291	54,70649	20,51095
14581266	Körigsberg	8499078	geonames	4888291		
14581267	Królewiec	8499078	geonames	4888291		
14709787	Город Калининград	8549491	geonames	3722833	54,71667	20,41667
14709786	Gorod Kaliningrad	8549491	geonames	3722833	54,71667	20,41667
14709788	Königsberg	8549491	geonames	3722833	54,71667	20,41667
21074251	Königsberg	12377936	Geotaurusz	3722833		
33746	Königsberg	33746	MNL GEO	3722833		
14581228	Калининград	8499078	geonames	4888291	54,70649	20,51095



Földrajzi nevek



A lista sorai földrajzi nevek, rákattintással a helyhez tartozó összes névváltozatot és a többi adatokat megnézheti. Módosításra akkor van lehetősége, ha a he

A **Szűrés** eszközre kattintást követően a **?** egyetlen karaktert, a ***** tetszőleges számú karaktert helyettesít.

+ Új földrajzi hely...

[Összesen 2 találat] 1 (1-2/2) Szűrés				
Szűrési feltételek: Földrajzi név = Königsberg; Szekeszthető? = igen;				
Azonosító	Földrajzi név	Kitüntetett földrajzi név	Névváltozatok	Forrás
33 746	Königsberg	Königsberg		MNL GEO
75 225	Königsberg	Újbánya	Königsberg, Mons Regius, Nova Baňa, Nová Baňa	MNL GEO

Karbantartó

[Összesen 70 találat] < < 1 > > (1-70/70)			
Szűrési feltételek: Földrajzi név = Königsberg;			
Azonosító	Földrajzi név	Kitüntetett földrajzi név	Névváltozatok
4 753 350	Königsberg	Königsberg	Königs-Berg
4 753 353	Königsberg	Königsberg	Königs-Berg
4 753 358	Königsberg	Königsberg	Königs-Berg
4 566 590	Königsberg	Kynšperk nad Ohří	Kynšperk, Königsberg, Königsberg an der Eger, Киншперк на Охри
21 178 873	Königsberg	Királyhalma	Crihalma, Königsberg
812 076	Königsberg	Königsberg	
4 753 343	Königsberg	Königsberg	
4 753 311	Königsberg	Königsberg	
4 753 356	Königsberg	Königsberg	
4 753 316	Königsberg	Königsberg	
773 267	Königsberg	Königsberg	
4 753 318	Königsberg	Königsberg	
4 753 360	Königsberg	Königsberg	
773 266	Königsberg	Königsberg	
4 753 296	Königsberg	Königsberg	
2 324 971	Königsberg	Königsberg	
4 753 322	Königsberg	Königsberg	
792 788	Königsberg	Königsberg	
773 268	Königsberg	Königsberg	
4 753 332	Königsberg	Königsberg	
4 753 293	Königsberg	Königsberg	
4 753 297	Königsberg	Königsberg	
4 753 300	Königsberg	Königsberg	
4 753 362	Königsberg	Königsberg	
4 753 364	Königsberg	Königsberg	
4 753 365	Königsberg	Königsberg	
33 746	Königsberg	Königsberg	
4 753 348	Königsberg	Königsberg	
4 753 306	Königsberg	Königsberg	
4 753 366	Königsberg	Königsberg	
4 753 363	Königsberg	Königsberg	
4 753 330	Königsberg	Königsberg	
773 265	Königsberg	Königsberg	
773 261	Königsberg	Königsberg Horn	Königsberg
812 982	Königsberg	Königsberg	

4 753 366	Königsberg	Königsberg	
4 753 363	Königsberg	Königsberg	
4 753 330	Königsberg	Königsberg	
773 265	Königsberg	Königsberg	
773 261	Königsberg	Königsberg Horn	Königsberg
812 982	Königsberg	Königsberg	
4 753 303	Königsberg	Königsberg	
75 225	Königsberg	Újbánya	Königsberg, Mons Regius, Nova Baňa, Nová Baňa
4 753 345	Königsberg	Königsberg	
750 125	Königsberg	Königsberg	
4 753 280	Königsberg	Königsberg in Bayern	Konigsberg in Franken, Königsberg, Königsberg in Bavariya, Königsberg ü
14 075 709	Königsberg	Chojna	Czińsbarg, Hojna, Königsberg, Königsberg in Neumark, Neumark, Хойна, X
14 581 264	Königsberg	Kaliningrad	Caliningrado, Calininopolis, KGD, Kaliningradas, Kaliningrado, Kaliningradu Köriqsberg, Καλίνινγκραντ, Калининград, Калињинград, Калінінград, Ч
4 753 312	Königsberg	Königsberg	
4 753 347	Königsberg	Königsberg	
4 753 298	Königsberg	Königsberg	
15 544 841	Königsberg	Nová Baňa	Königsberg, Nova Banja, Nova Baňa, Újbánya, Нова Бања, Нова-Баня, بنا
4 753 309	Königsberg	Königsberg	
4 753 320	Königsberg	Königsberg	
4 753 328	Königsberg	Königsberg	
4 753 334	Königsberg	Königsberg	
4 753 308	Königsberg	Königsberg	
4 753 341	Königsberg	Königsberg	
4 753 326	Königsberg	Königsberg	
4 753 336	Königsberg	Königsberg	
4 753 302	Königsberg	Königsberg	
750 127	Königsberg	Königsberg	
14 709 788	Königsberg	Gorod Kaliningrad	Königsberg, Город Калининград
790 572	Königsberg	Königsberg	
4 753 324	Königsberg	Königsberg	
21 074 251	Königsberg	Königsberg	
4 753 295	Königsberg	Königsberg	
4 753 304	Königsberg	Königsberg	
4 569 226	Königsberg	Klimkovice	Königsberg
793 646	Königsberg	Königsberg	
814 862	Königsberg	Königsberg	
4 753 291	Königsberg	Königsberg	
4 753 307	Königsberg	Königsberg	
4 753 339	Königsberg	Königsberg	
814 517	Königsberg	Königsberg	
4 567 872	Königsberg	Královac	Königsberg

Földrajzi név karbantartó



Az MNL GEO forrásból származó helyek adatait módosíthatja. Új felvitelkor adja meg a földrajzi hely adatait, majd kattintson a **Mentés** legalább egy típusba sorolást.

▼ Adminisztratív adatok

Azonosító 33746

Hitelesítési szint

Szolgáltató intézmény ()

Forrás MNL GEO (75028)

▼ Entitás kapcsolatok

Entitás azonosító 3 751 275

A szabályalapú entitás képzés során keletkezett azonosító

Entitáshoz rendelés időpontja 2025.03.18. 15:52:37

Entitás név Königsberg

Entitás elemei Königsberg [MNL GEO] - 33746

Königsberg [Geotaurusz] - 21074251

Gorod Kaliningrad [geonames] - 14709786

▼ Földrajzi név

Ha az alábbi szekció adatait szűrően látja, akkor kitüntetett földrajzi nevet lát

Földrajzi név Königsberg

Csak olyan nevet adhat meg, amely nem szerepel kitüntetett névként (másik helynél) az MNL GEO-ban

Történeti időszak kezdő - záró

▼ Kitüntetett földrajzi név

Kitüntetett földrajzi név Königsberg

Csak olyan nevet adhat meg, amely nem szerepel kitüntetett névként (másik helynél) az MNL GEO-ban

Újbánya (Königsberg, Nova Banja)

15544841	Königsberg	8968251	geonames	7532769	48,42305	18,64037
21159120	Nová Baňa	12430998	Engel	7532769	48,42639	18,63766
15544834	Nová Baňa	8968251	geonames	7532769	48,42305	18,64037
15544843	Nova Banja	8968251	geonames	7532769	48,42305	18,64037
21159119	Újbánya	12430998	Engel	7532769	48,42639	18,63766
15544845	Újbánya	8968251	geonames	7532769	48,42305	18,64037
15544843	Nova Banja	8968251	geonames	7532769		
21159119	Újbánya	12430998	Engel	7532769		
15544845	Újbánya	8968251	geonames	7532769		
21021615	Újbánya	12332334	Geotaurusz	7532769		
64938	Újbánya	64938	MNL GEO	7532769		

Results: Most frequently used names of cities in archival documents

758	Brno
790	Brünn
179	Olmütz
16	Prag
17468	Prága
255	Praha

4	Basel
94	Genf
18	Nice
92	Strassburg
86	Bratislava
74754	Pozsony

306163	Bécs
18	Vienna
66	Wien
3	České Budějovice
391	Budweis
59	Split
15	Spalato

No hits: Olomuc, Pressburg, Nizza, Genève, Strasbourg