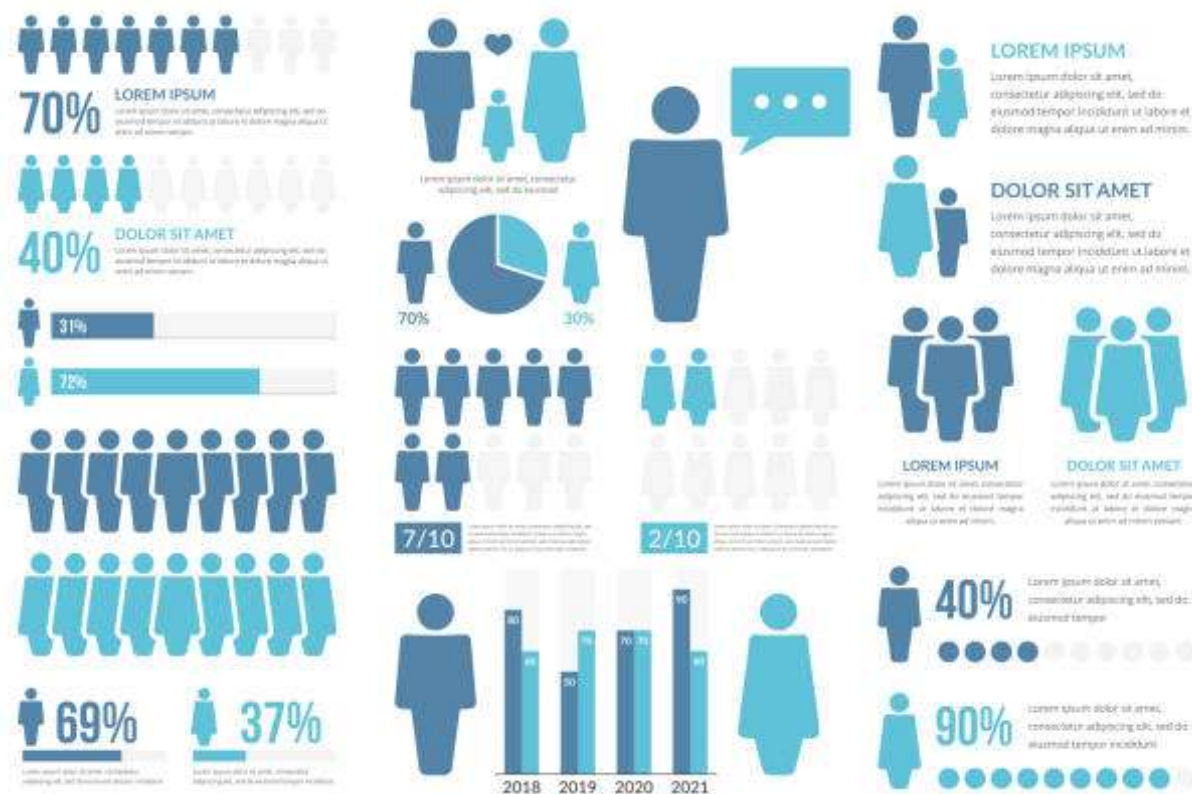


Az anyakönyvi adatstruktúrák nyelvi modell alapú elemzése nagy teljesítményű számítási infrastruktúrán

Szatucsek Zoltán

Magyar Nemzeti
Levéltár



Budapest, 2025. november 13. II. Országos HUN-REN Cloud
Találkozó

Miért nem használják a történészek számítógépet

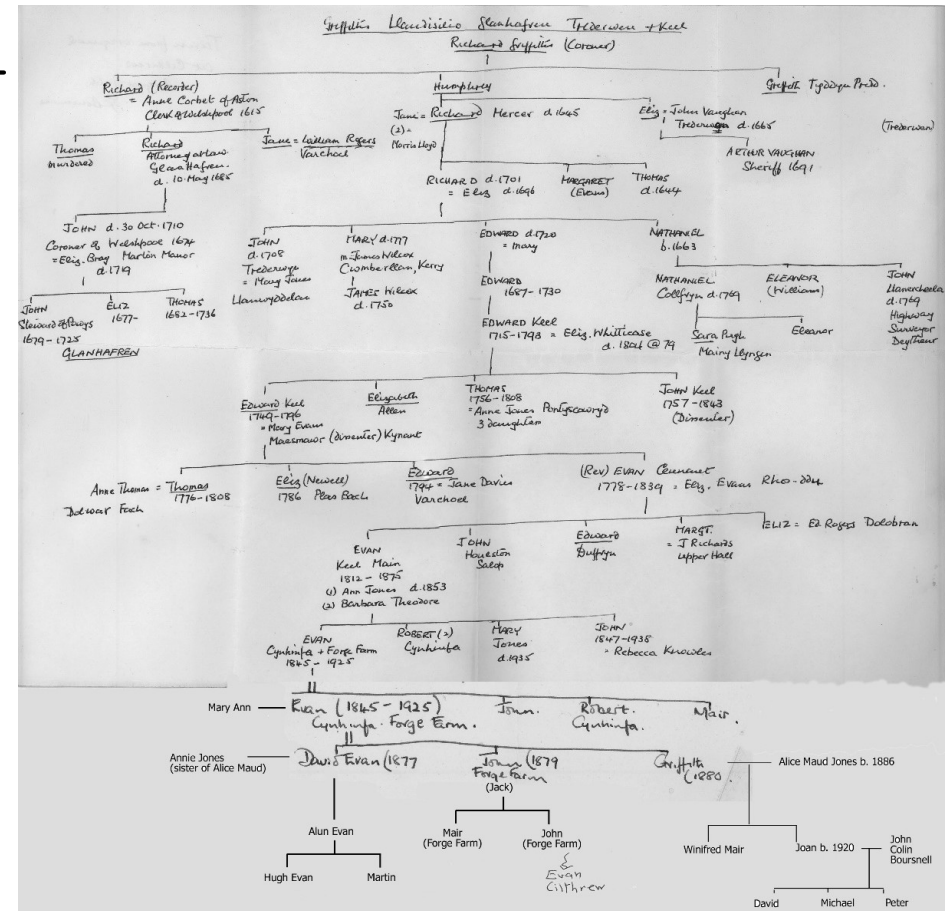
- a források korlátai
(inhomogenitás,
hiányosság,
kontextusérzékenység)
- lingvisztikai fordulat
- kelet-közép-európai
specifikumok

MI VÁLTOZOTT ÉS MIÉRT



Családrekonstitú

FLEURY, MICHEL-
HENRY, LOUIS: Des
registres
paroissiaux à
l'histoire de la
population. Paris,
1956



Történeti demográfia

- Kováts Zoltán, Dányi Dezső, Andorka Rudolf, Benda Gyula, Őri Péter

- kritikája

- korlátozott és kétséges reprezentativitás
- alkalmazói csak falvakat dolgoznak fel
- csak a hosszú ideig helyben lévőket lehet követni

A FÉRJ NEVE		A FELESÉG NEVE	
A férj apja A férj anyja		A feleség apja A feleség anyja	
A férj születési időpontja életkora a házasságkötéskor halálozási időpontja életkora a halálozáskor		A feleség születési időpontja életkora a házasságkötéskor halálozási időpontja életkora a halálozáskor	
A férj családi állapota születési helye foglalkozása vallásfelekezete		A feleség családi állapota születési helye foglalkozása vallásfelekezete	
A házasságkötés időpontja A házasság megszűnésének időpontja			
A gyermek neve	Születési időpontja	Az anya életkora szülés-kor, év	Intervallum, hónap
A gyermeke			
házas-társ és ház-as-ságának			
halálának időpontja			
1.			
2.			
3.			

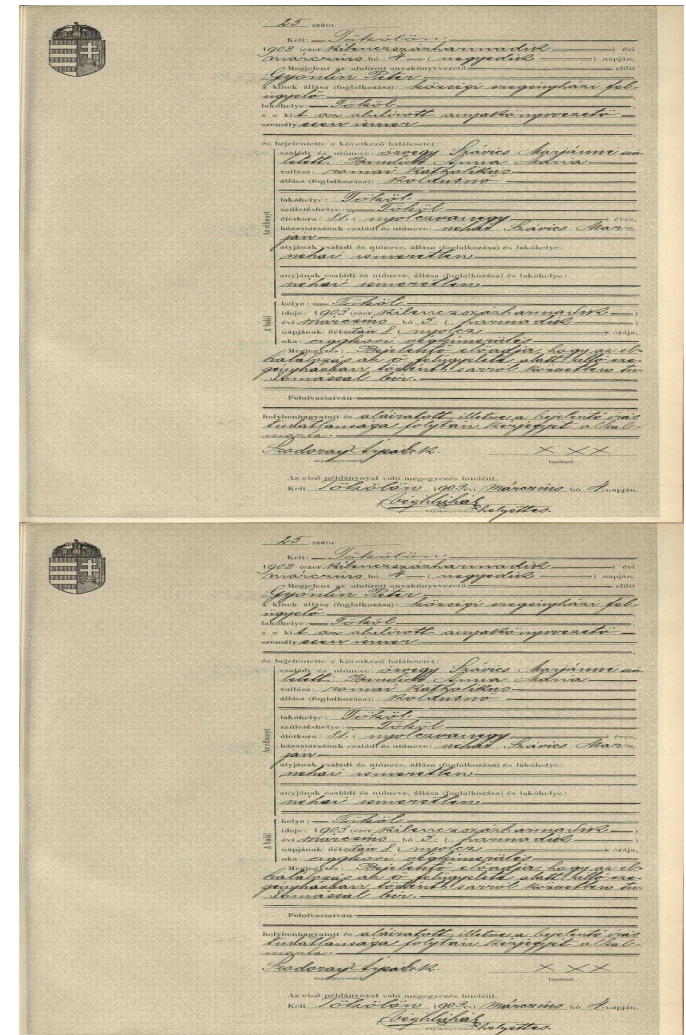
„(A) születések, házasságok és halálesetek közhitelű nyilvántartására és tanúsítására kizárólag a jelen törvény értelmében arra hivatott közegek által vezetett állami anyakönyvek szolgálnak.”

- Az 1894. évi XXXIII. törvénycikk június 21-én 4 szavazat többséggel
- az utolsó kizárólag egyházi házasságot 1895. szeptember 30-án Vaszary Béla honvéd huszárezredes és csengeri Háczy Anna között az esztergomi bazilikában Vaszary Kolos hercegprímás előtt
- első polgári házasság 1895. október 1-jén Budapesten, a Gyár utcai (ma Liszt Ferenc tér) 2. alatti anyakönyvi hivatalban Kriszhaber Károly Joakhim gyárigazgató és menyasszonya, Mezei Anna, Mezei Mór országgyűlési képviselő leánya

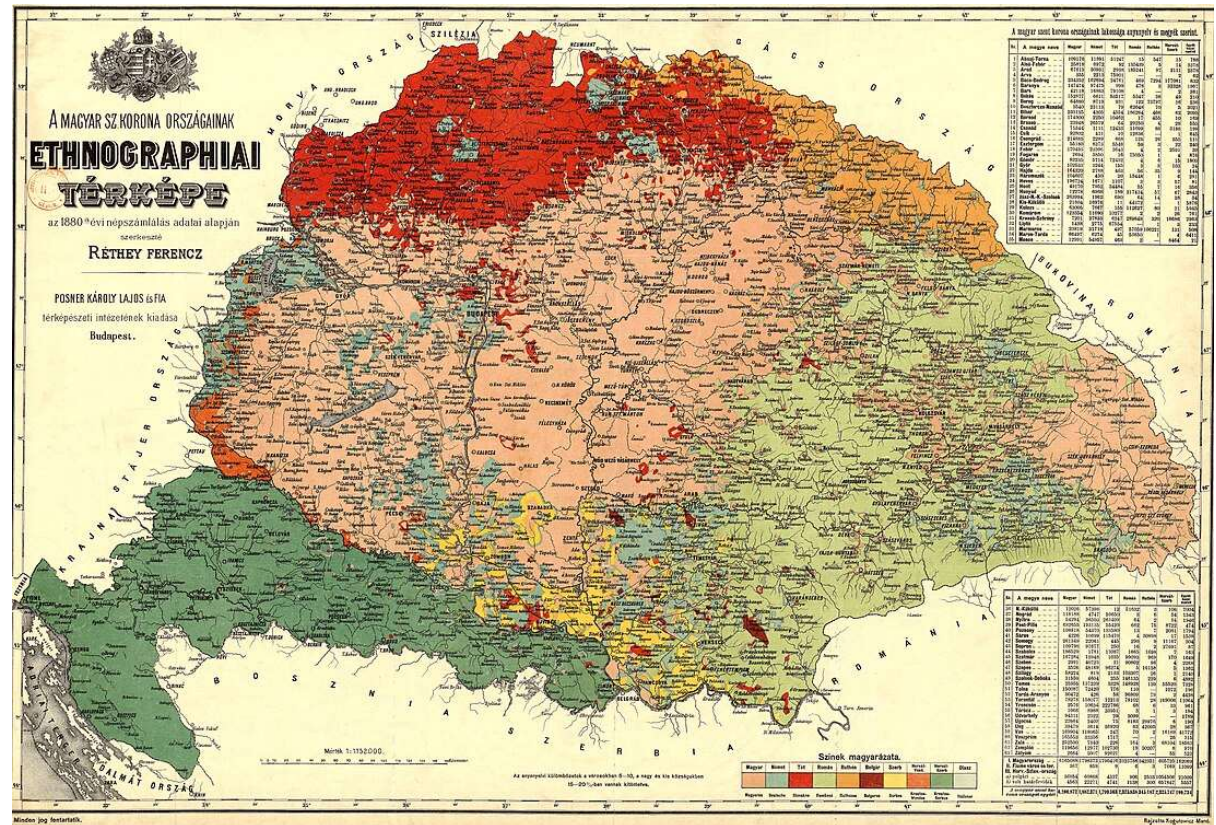


Állami anyakönyvek másodpéldányainak gyűjteménye

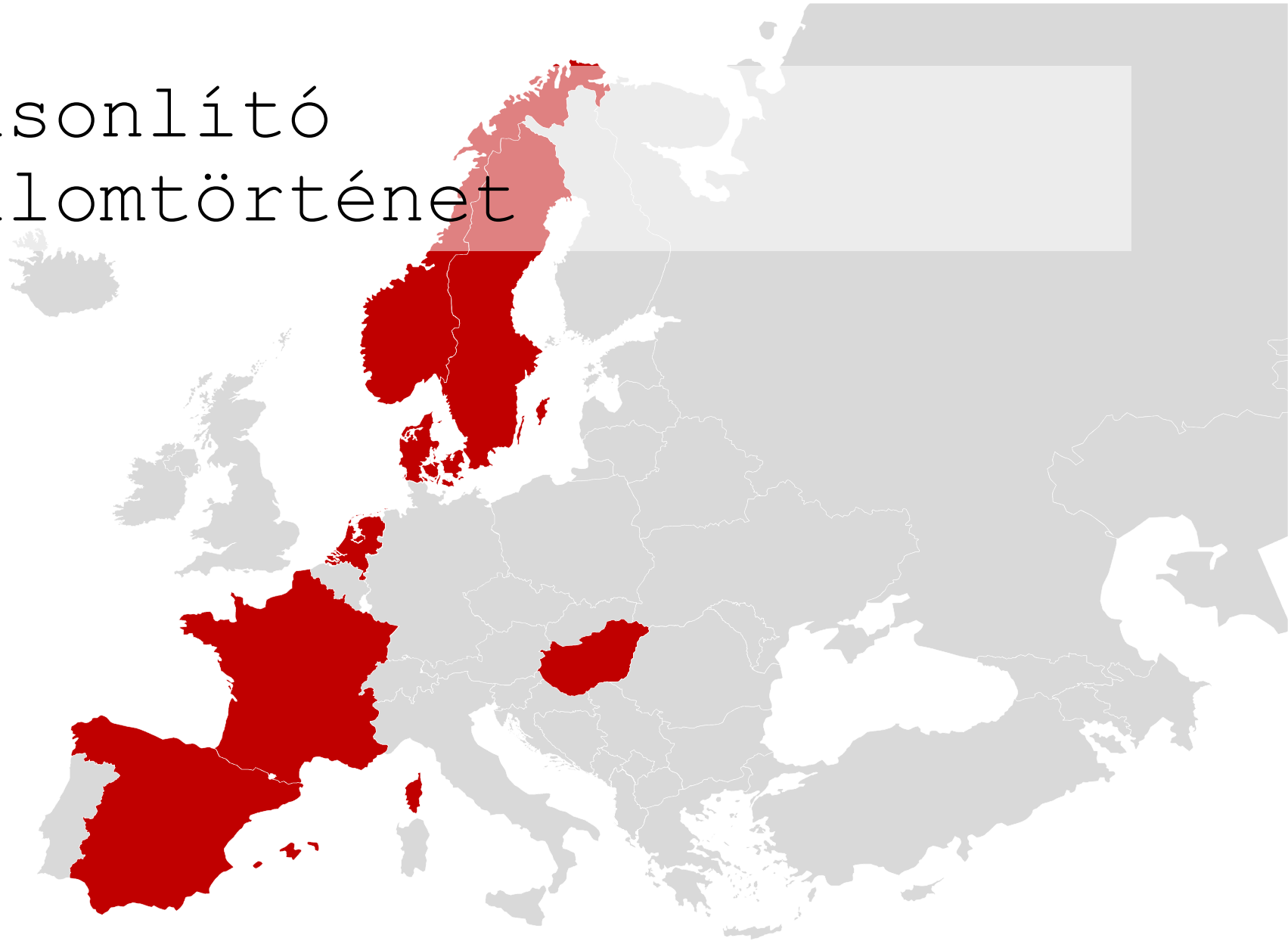
- Mennyiségek
 - 38 353 kötet
 - 22 levéltárban
 - 1918 folyóméter,
 - 23 millió oldal
 - 250m személynév
- 1895-1980
- a családkutatás mellett a történeti demográfia, családrekonstrukció, társadalomtörténet, gazdaságtörténet, orvostörténet, nőtörténet, migrációtörténet, jog- és politikatörténet forrásbázisa.



Kárpát-medencei anyakönyvi kataszter

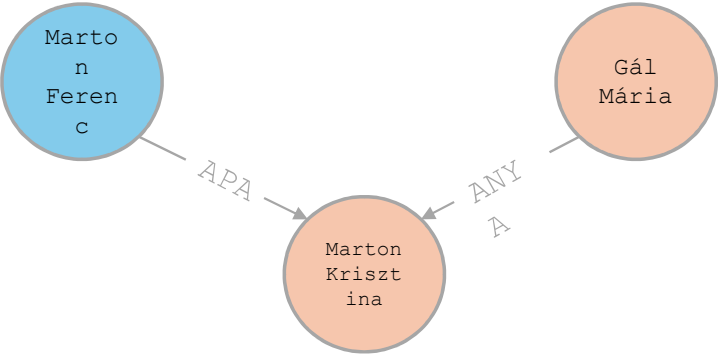


Összehasonlító társadalomtörténet



1907 szept 23	Krisztina leány err. ref.	Marton Ferenc goldműves Gál Mária Alsódabas Alsóbrád puszta
---------------------	---------------------------------	---

Marton Ferenc	apa	Alsódabas	1883
Gál Mária	anya	Alsódabas	1886
Marton Krisztina	gyermek	Alsódabas	1907



10075 Kemence

Befejezett



Képek osztályozása



Kézírásfelismerés

The image shows a handwritten family tree on a grid. The names are written in cursive, and some are highlighted with purple boxes. A digital annotation overlay is visible on the right side of the image, showing the name 'Strupka' and a list of annotations.

Strupka

Kattintson ide az annotáláshoz!

rossz szegmentáció nem olvasható áthúzott kihagyom

▼ **Mások így annotálták - 5**

- Stsupka - 2025. 05. 08. - annotálva
- Straupka - 2025. 05. 08. - annotálva
- Strapka - 2025. 05. 08. - annotálva
- Strupka - 2025. 05. 08. - annotálva
- Szrapka - 2025. 05. 08. - annotálva

▼ **Felismert - 0 hipotézis**

▼ **MENYASSZONY_SZULOK_NEVE_2 mező - 1. sora**

55/130 | 60e52729

Kézírásfelismerés

Opál 4 modell (2025.06.16)

Base modell	Microsoft: trocr-base-stage1
Tanítóanyag	abony-rev-1, allami-anyakonyv-1, allami-anyakonyv-2, nytud, kiss-jozsef, pest-100, pest-200
Benchmark készlet	benchmark-pest-megye
WER 1st hipo	27,7 %
WER best hipo	12,9 %
WER TRANS 1st hipo	26,3 %
WER TRANS best hipo	11,1 %
CER 1st hipo	7,2 %
CER best hipo	3,4 %
CER TRANS 1st hipo	5,7 %
CER TRANS best hipo	2,4 %

Rekordösszekapcsolás

születési helye: <u>Abony</u>	családi és utónév: <u>Takács</u>	születési helye: <u>Kalocsa</u>
születési ideje: <u>1855.</u> év <u>január</u> hó <u>13</u> nap	nem: <u>férfi</u> családi állapota: <u>szóltat</u> jelenleg: <u>özvegy</u>	születési ideje: <u>1889.</u> év <u>nov.</u> hó <u>25</u> nap jelenleg családjának és utónevének: <u>neke: Takács Anikó</u> <u>neke: Kalocsai Rozália</u>

Születési | apa: Haszler János - anya: Csiffári Borbála

HU-MNL-PML-XXXIII.1.a.-204.-a-4.

Apa - név	Haszler János Haszler János [0.8758] Maszler János [0.8708] Hászler János [0.6533] 1x SZULOK_NEVE_ALLASA_LAKHELYE_6 PPR_ID: 65290793 JAVITVA:1 TXT: Haszler János CSALÁD: Haszler UTÓ: János DEBUG: Haszler János	
Apa - lakhely	Bambék Bambék [0.8092] Fambék [0.8012] Zambék [0.7713] Sambék [0.7188] Szambék [0.382] 1x SZULOK_NEVE_ALLASA_LAKHELYE_6 PPR_ID: 65290793 JAVITVA:1 TELEPÜLÉS: Bambék DEBUG: Bambék	
Apa - foglalkozás	napszámos napszámos [0.9998] napszállós [0.9997] napszáros [0.9995] Napszámos [0.9988] napszános [0.5155] 1x SZULOK_NEVE_ALLASA_LAKHELYE_6 PPR_ID: 65290793 JAVITVA:1 TXT: napszámos	Borbába [0.8459] Csifári [0.4015] Csiffári Borbála [0.401]
Anya - név	Csiffári Borbála Csiffári Borbála [0.8463] Csiffári Borbába [0.8459] Csifári Borbála [0.451] Cíffári Borbála [0.4015] Csiffári Borbála [0.401] 1x SZULOK_NEVE_ALLASA_LAKHELYE_6 PPR_ID: 65290793 JAVITVA:1 TXT: Csiffári Borbála CSALÁD: Csiffári UTÓ: Borbála DEBUG: Csiffári Borbála	
Anya - lakhely	Bambék Bambék [0.8092] Fambék [0.8012] Zambék [0.7713] Sambék [0.7188] Szambék [0.382] 1x SZULOK_NEVE_ALLASA_LAKHELYE_6 PPR_ID: 65290793 JAVITVA:1 TELEPÜLÉS: Bambék DEBUG: Bambék	
Anyakönyvvezető - aláírás	anyakönyvvezető anyakönyvvezető [0.999] 1x AKV_ALAIRAS_6 PPR_ID: 65290793 JAVITVA:1 DEBUG: anyakönyvvezető	
Kapcsolódó rekordok (gépi tanulás)	anya 89.877%-ban azonos a következő bejegyzésekkel: - Születési anyakönyv. Zsámbék. . anya - Születési anyakönyv. Zsámbék. . anya - Születési anyakönyv. Zsámbék. . anya - Születési anyakönyv. Zsámbék. . anya apa 65.968%-ban azonos a következő bejegyzésekkel: - Születési anyakönyv. Zsámbék. . apa - Születési anyakönyv. Zsámbék. . apa - Születési anyakönyv. Zsámbék. . apa - Születési anyakönyv. Zsámbék. . apa	Borbála [0.9931] Csifári Borbála [.] Csittári Borbála [0.7628]

KAPCSOLATOK: SZOVJET TÁBOROK MAGYAR FOGLYAI



Szovjet táborok magyar foglyai

Találatok szűrése

Születési év

1000-01-01 — 2023-12-31

1912

Év

Szűrés dátumra Összes dátum

Rendfokozat

Média

☐ Kép (2 db)

Névter használata

☒ Ne használja a névteret

☐ Személy

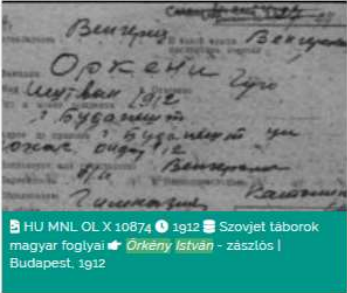
☐ Testületi

☐ Földrajzi

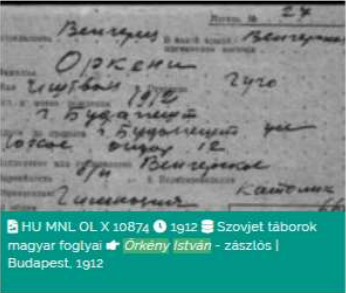
Egyszerű keresés Összetett keresés Kifejezés

Örkény István

2 dokumentumban / 2 érintett oldal (0,18g másodperc)

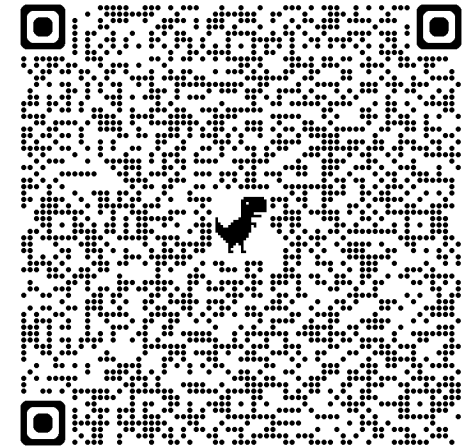


HU MNL OL X 10874 • 1912 Szovjet táborok magyar foglyai • **Örkény István** - zászlós | Budapest, 1912



HU MNL OL X 10874 • 1912 Szovjet táborok magyar foglyai • **Örkény István** - zászlós | Budapest, 1912

← 1 →

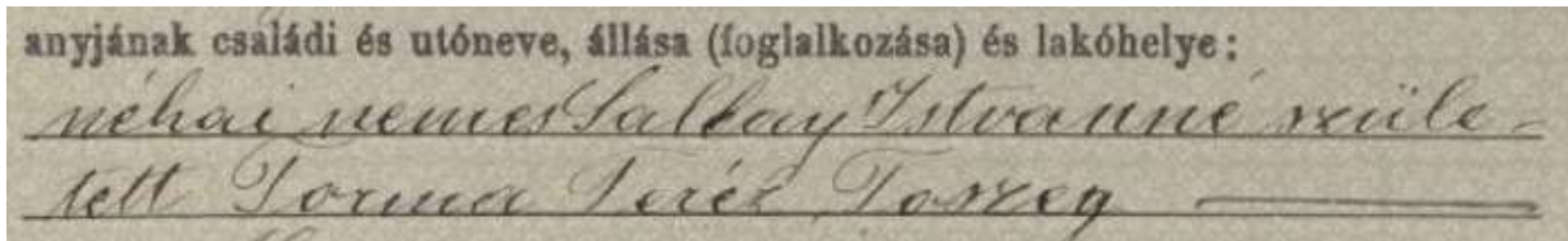


10075 Kemence

Befejezett



PPR Töltött mezők

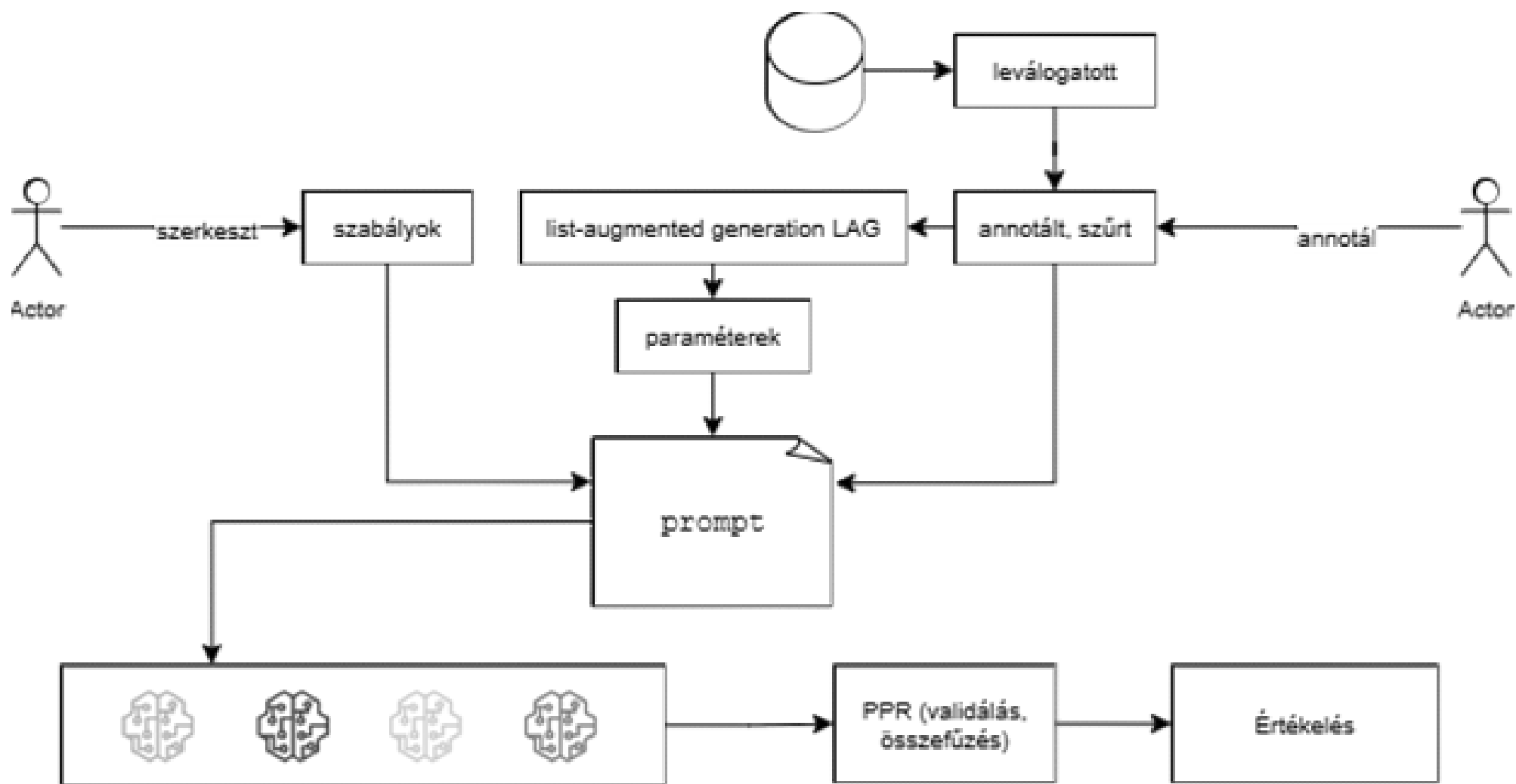


1. sor: ['néhai nemertlalkay Istvanné szüle-', 'néhai nemertlalkay Istvanné szüle-', 'néhai nemertlalkay Istvanné szüle-', 'néhai nemertlalkay Istvanné szüle-']
2. sor: ['tett Torina Teréz Tószeg', 'tett Torina Teréz Tószeg', 'tett Torina Teréz Tószeg', 'tett Torina Teréz Tószeg']

Töltött mezők

- anyja neve, foglalkozása, lakhelye
- szülők neve, foglalkozása lakhe
- hely, dátum, kelt
- menyasszony adatok
- gyermek neve, neme, vallása
- menyasszony szülei
- tanúk, név, lakhely
- elhunyt neve, foglalkozása, lakh
- vőlegény szülők neve
- vőlegény adatok
- apja neve, foglalkozása, lakhelye





Kipróbált modellek

Cloud:

OpenAi:

gpt-4o,
o1-mini,
gpt-4.5-

preview

Local:

Alibaba:

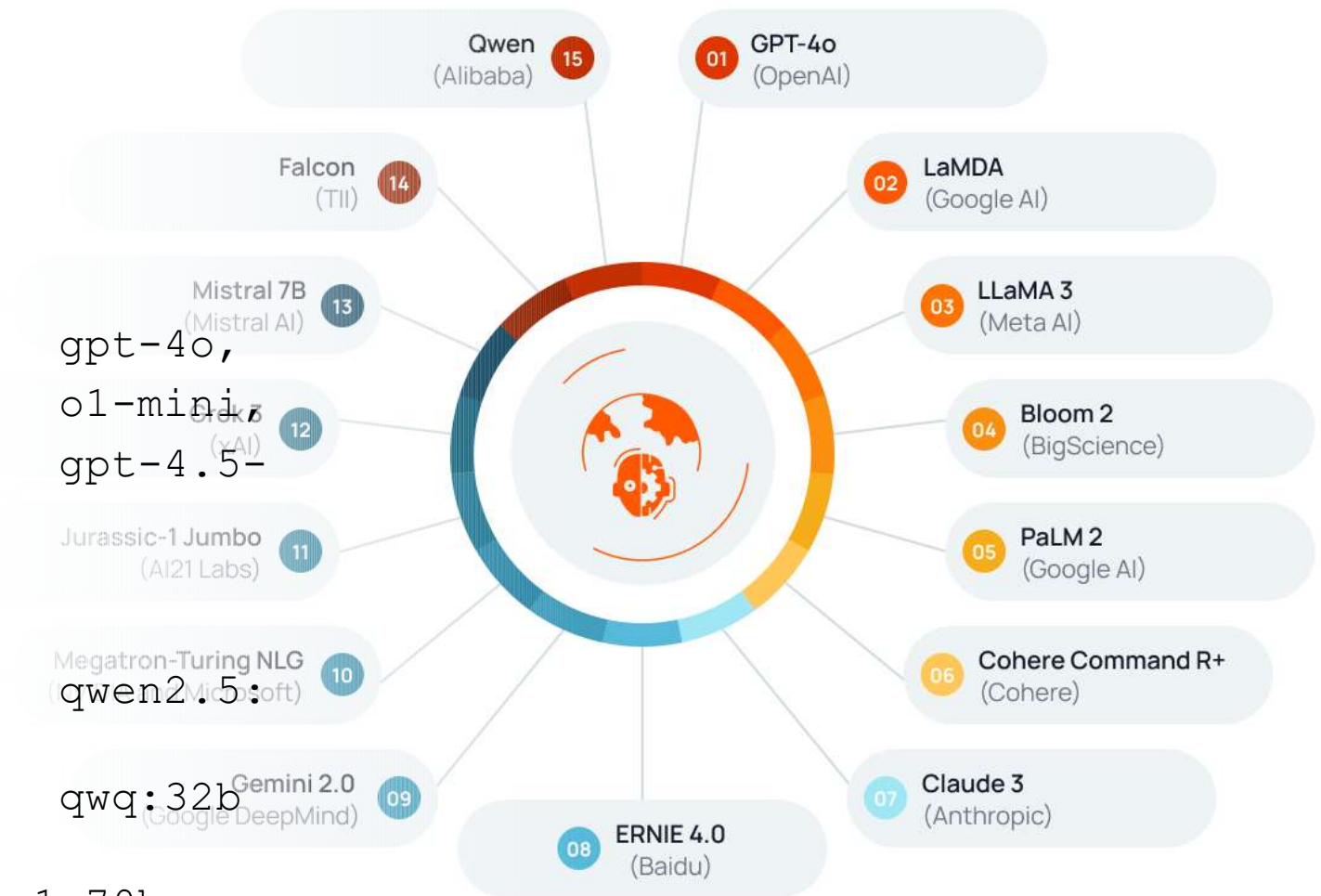
72b,

(érvelő)

DeepSeek: r1:70b

(érvelő)

Meta: LLaMA 3.2:70b





kuvasz

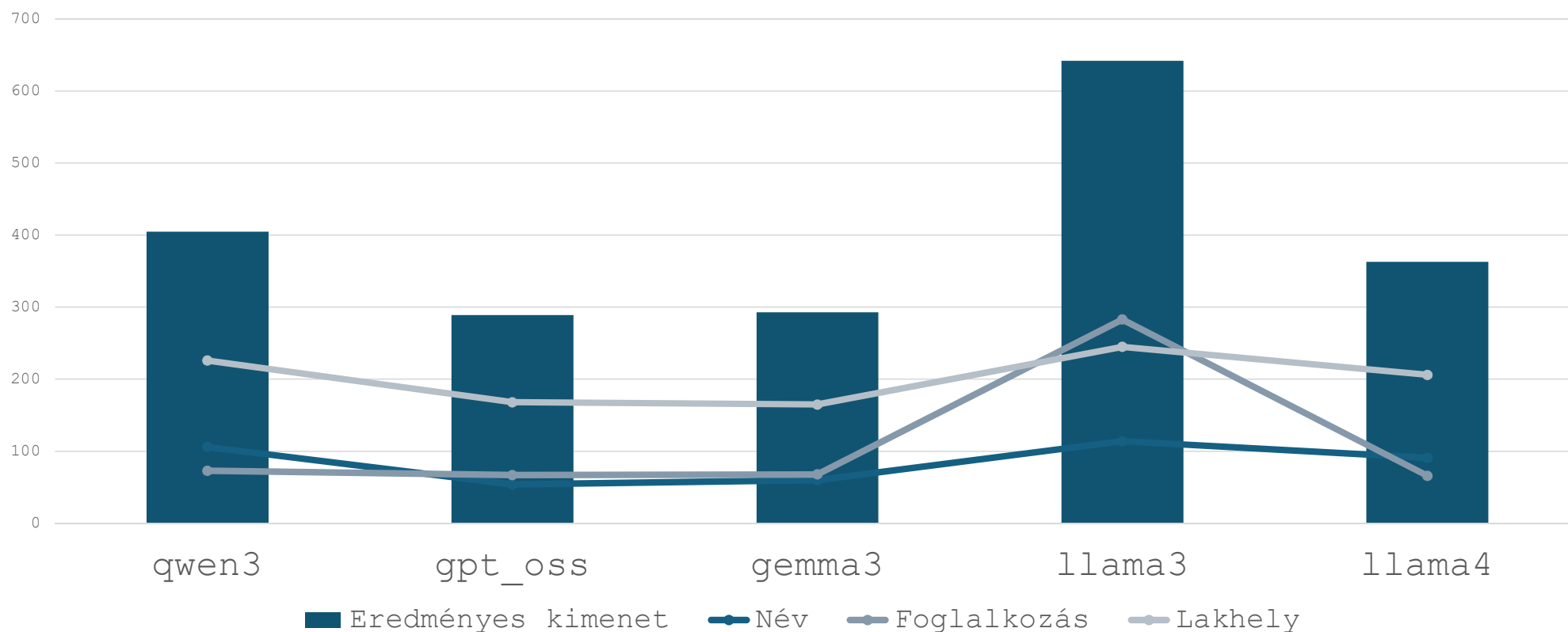
CPU: 24 mag, 32 szál
RAM: 64 GB
GPUs: 2 x [NVIDIA RTX 4090]
VRAM: 48 GB (2 x 24)
CUDA: 32.768 core (2 x
16.384)
SSD: 2 x 1 TB
HDD: 4 TB



HUN-REN Cloud

CPU: 48 mag, 96 szál
RAM: 378 GB
GPUs: 4x [NVIDIA A100-PCIE-
40GB]
VRAM: 160 GB (4 x40)
CUDA: 27.648 core (4 x
6.912)
SSD/HDD: 512 GB

Születési anyakönyv: anyja neve, foglalkozása,
lakhelye mező feldolgozásának kiértékelése
modellenként



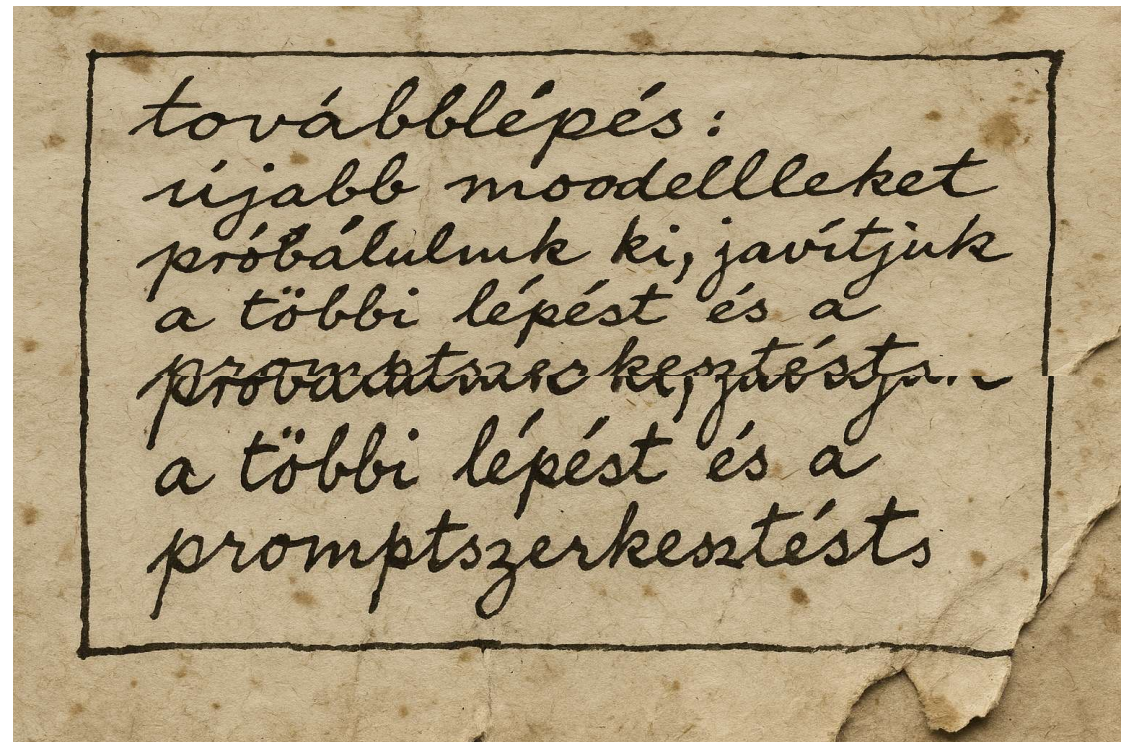
tanított LLM

- bizonytalan kimenet
- költséges tanítóanyag
- fejlesztések ritmusa

újabb modellek (llama4
scout...)

megelőző folyamat lépések
fejlesztése (szegmentáció,
HTR)

promptokat lépésekre bontjuk



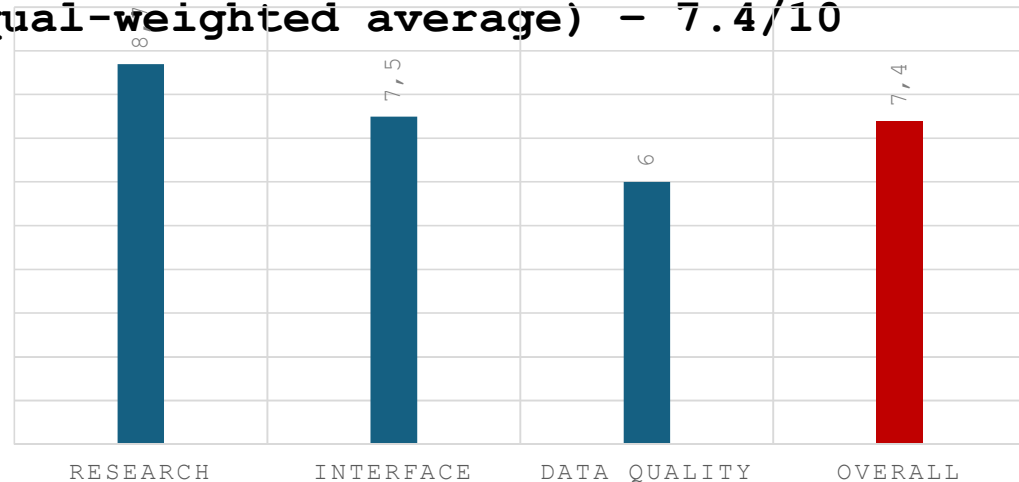
Végfelhasználói teszt (n=10)

How much can this service help your research? -
8.7/10

How clear and easy-to-use was the interface? -
7.5/10

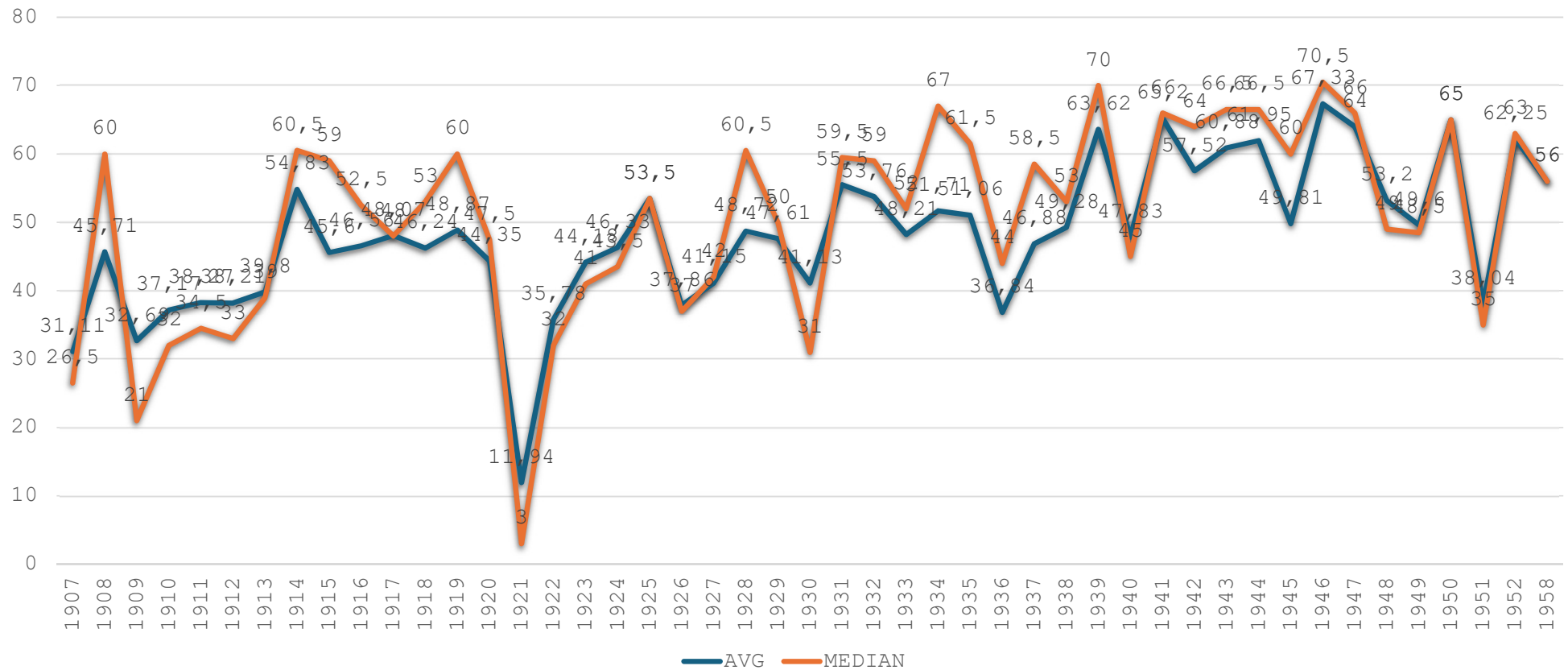
How do you rate the data quality? - 6.0/10

Overall rating (equal-weighted average) - 7.4/10



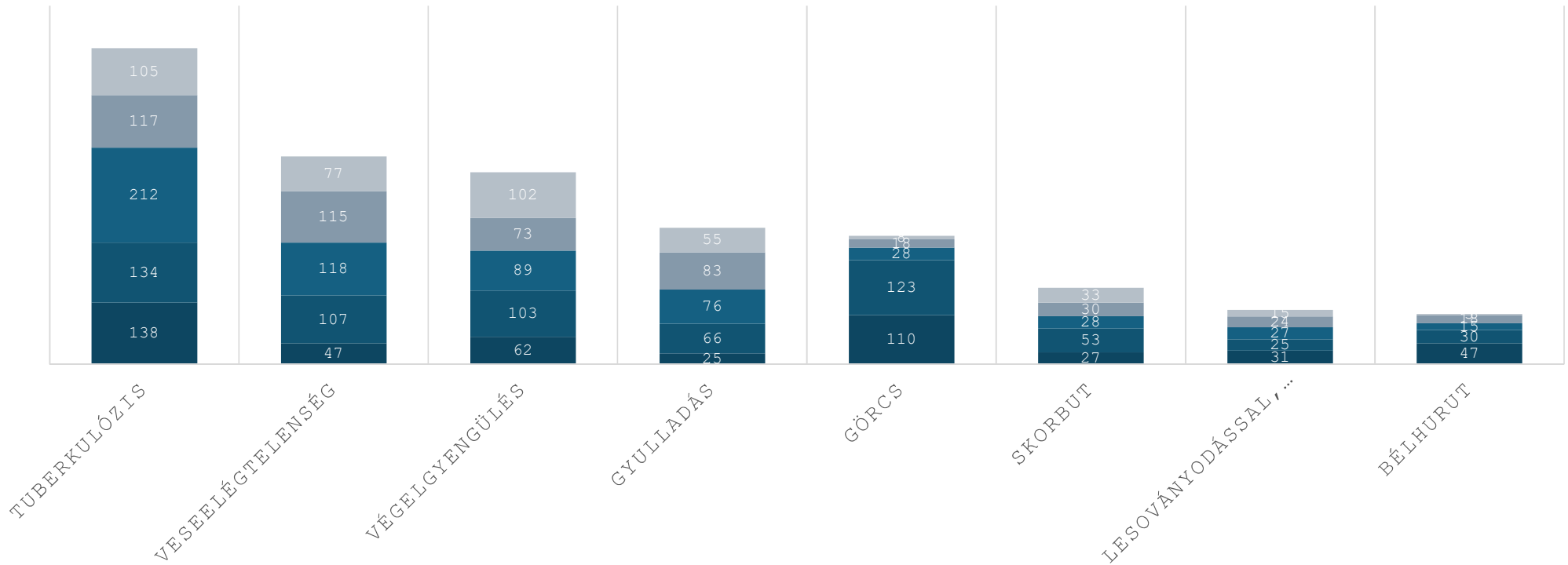
Scale: 1-10. Responses from 10 end-user testers. Overall is a simple mean of the three categories.

Elhunytak átlag- és medián életkora (Zsámbék, 1907–1958)

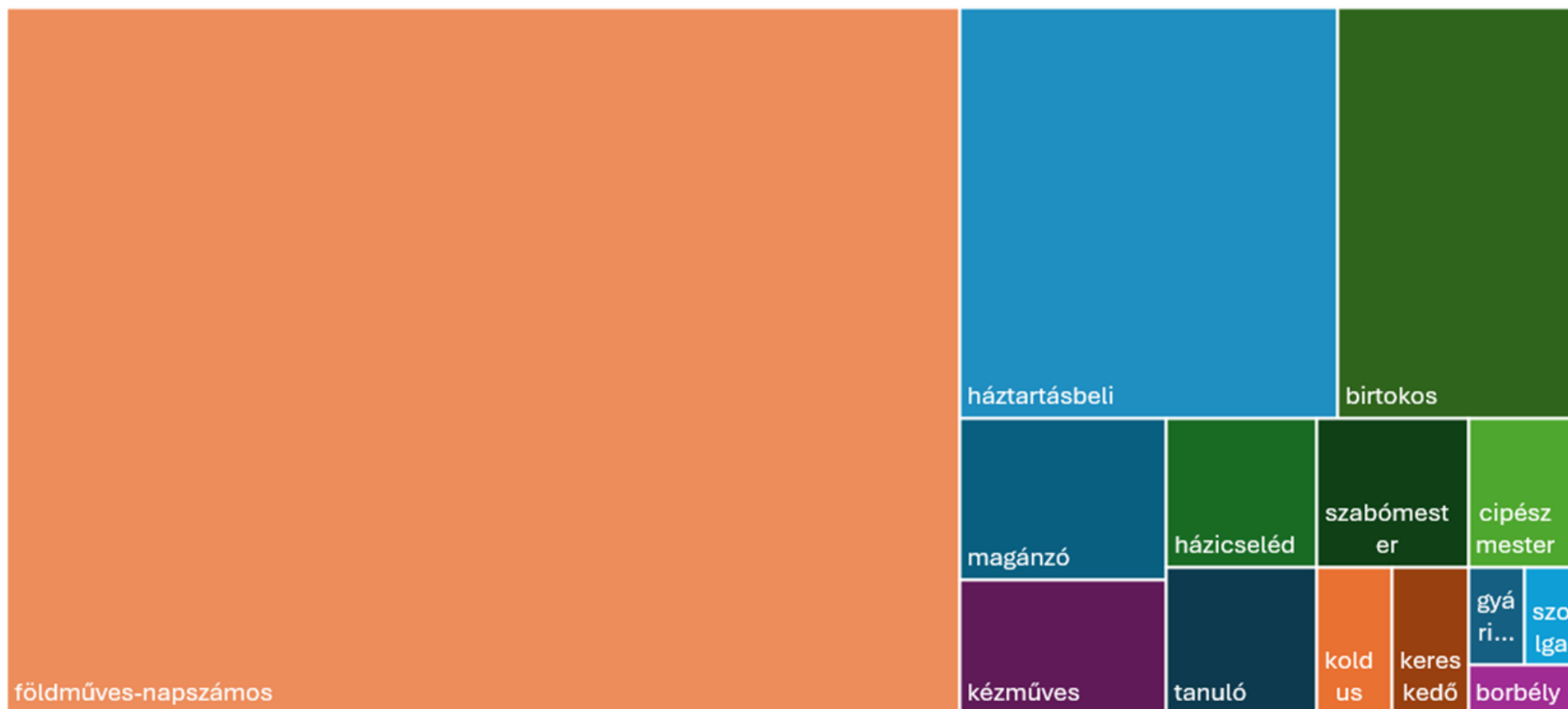


A nyolc leggyakoribb halálok (Zsámbék, 1895-1939)

■ 1890-1899 ■ 1900-1909 ■ 1910-1919 ■ 1920-1929 ■ 1930-1939



Foglalkozások a halotti anyakönyvekben (Zsámbék, 1895-1990)



Ferjhez menesete
után, 15 kőre
meg let a' fiú,
már a' férj nem
tudott sem i' vállaltat.

Magyar Nemzeti Levéltár

anyja neve:

Bánki Zsolt
Simon András
Szatucsek Zoltán
Szerényi Ildikó
Szűcs Kata Ágnes
Vadász Noémi
Varga Emese
Záros Zsolt

szatucsek.zoltan@mnl
.gov.hu