

# Az utónevek eredetleírásának formalizálása az Utónévportálon

Sass Bálint, Raátz Judit



MAGYAR TUDOMÁNYOS AKADÉMIA  
NYELVTUDOMÁNYI INTÉZET

{sass.balint,raatz.judit}@nytud.mta.hu

## Utónévkereső<sup>béta</sup>

The screenshot shows a search interface with several filters and search criteria. On the left, there are three dropdown menus for 'tetszőleges' (arbitrary) criteria: 'alakváltozat' (variant), 'névalkotás' (naming), and 'nyelv' (language). The main search area includes fields for 'nem' (gender), 'név' (name), 'eredet' (origin), 'jelleg' (character), 'nyelv' (language), 'adott névből' (from given name), 'jelentés' (meaning), 'névnap' (name day), 'becézés' (nickname), 'szótagszám' (syllable count), 'hangrend' (sound sequence), 'magánhangzók' (vowels), 'gyakoriság' (frequency), 'össz-népszerűség' (popularity), 'újszülöttek' (newborns), and 'trend' (trend). There are also buttons for 'Mehet' (Go) and 'Törlés' (Clear).

7 találat.

**Edvárd** | ♂ | *Germán (angolszász) eredetű név* | birtok; vagyon + őrzés; védelem | „elég ritka” – „elég ritka”

**Előd** | ♂ | *Régi magyar személynév* | elsősülött; ősi | „elég ritka” – „elég ritka”

**Emil** | ♂ | *Latin eredetű név* | versengő; igyekvő | „közepesen gyakori” – „elég ritka”

**Endre** | ♂ | *Az András név alakváltozata* | „közepesen gyakori” – „elég ritka”

**Erik** | ♂ | *Germán eredetű név* | becsület (más feltevés szerint); egyedül; hatalmas; uralkodó | „közepesen gyakori” – „közepesen gyakori”

**Ernő** | ♂ | *magyar nyelvújítási alkotás - ő kicsinyítő képzővel az Erneszt - névből* | „közepesen gyakori” – „elég ritka”

**Ervin** | ♂ | *Germán eredetű név* | had(sereg) + barát; becsület; örökség; vadkan | „közepesen gyakori” – „elég ritka”

## Emil

eredet: *Latin eredetű név*

formalizálva: *Latin [ > ] ~*

jelentés: *versengő; igyekvő*

névnap: *03-10 05-22 05-28 09-11*

becézés: *Emilke Emi Emike Emici Milke*

rokon nevek: *Emília Emilián*

szótagszám: *2* – hangrend: *magas* – magánhangzók: *e-i*

gyakoriság: „közepesen gyakori” – „elég ritka”

## Az eredetleírások feldolgozása

**Cél:** „szemantikus keresés” az utónevek adataiban.

... melyek a névalkotással létrehozott neveink, ... melyek a latin eredetű kicsinyítőképzős formájú neveink?

**Nehézség:** az eredetleírás szabad szöveg → információkivonatolás szükséges.

**Megoldás:** a több mint 3500 eredetleírást szigorúan meghatározott formális alakra hozzuk, mely minden információt explicit tartalmaz, és géppel könnyen értelmezhető. Hogy az akár rejtett aspektusokra is külön-külön rákereshessünk.

**1. szakasz:** a leírások **egységesítése**, normalizálása – szabályokkal

= különféle megfogalmazású, de azonos jelentésű fordulatok → egységes alakra

~50 szabály – pl.: kisbetűsítés; „személynév” → „név”; rövidítések; névelő, „való”, „származik” stb. törlése; sorrendegységesítés; „becéző formája”, „becéző rövidülése”, „becéző alakja” → „becézője”

**2. szakasz:** a tényleges **formalizálás**

= az eredetleírásban szereplő nyelvi **elemek** + az egyik elemet a másikba alakító **műveletek** felfedése → **formalizált alak:** nyelvi elemek + műveletek

Domonkos eredetleírása	<i>A latin Dominicus név magyar formája</i>
formalizált alakja	<b>latin: Dominicus [megfelelője] magyar: ~</b>
	elem                      művelet                      elem

**Eredmény:** egyértelmű, kereshető leírások.

## Félautomatikus formalizálás

• **Egyszerűbb esetek:** automatikusan az egységesített alakból a gyakori mintázatok alapján kialakított szabályrendszer segítségével.

~50 szabály – pl.: „NYELV<sub>1</sub> eredetű NYELV<sub>2</sub> névből” → **NYELV<sub>1</sub> [ > ] NYELV<sub>2</sub> [ > ] ~**  
 „NYELV<sub>1</sub> NEVALAK NYELV<sub>2</sub> becézője” → **NYELV<sub>1</sub>:NEVALAK [megfelelője] NYELV<sub>2</sub> [becézője] ~**

⇒ eredmény: 3529-ből 2163 név készült el – **61,3%** ✓

• **Összetettebb esetek:** manuálisan, két párhuzamos annotátorral

Útmutató lényege:  
 elem = nyelv(attribútum):névalak'jelentés'definíció"  
 művelet = [megfelelője], [becézője], [formája], [női=párja], [rövidülése], [alakvált], [ > ], [névalkotás]

⇒ eredmény: mindössze 251 név esetében kaptunk két teljesen azonos leírást – **7,1%** ×

Sándor: Az Alexander név rövidülésének a megmagyarosodott alakja

Lionella: A francia 'kis oroszlán' jelentésű Leon férfinév angol Lionel változatának női párja

**Változatos hibák:** elütés, formai hibák, két elem művelet nélkül, művelettévesztés (pl.: [női=párja] helyett [megfelelője]), attribútumhiba ...

**Egyetértés: extrém alacsony.** Ilyen feladatnál, ahol nem két kategória van, hanem nagyon sokféle karaktersorozat előfordulhat értéként, nem is nagyon lehet sokkal jobbat várni?

**Ötlet:** olyan beviteli rendszer, ami eleve nem enged meg formailag rossz inputot.

## Az Utónévportál

MTA Nyelvtudományi Intézet feladata:

*az anyakönyvi bejegyzésre alkalmasnak minősített női és férfi utónevek jegyzékét kezelése és bővítése.*

Évente 80-100 új név. Új online felület: *Utónévportál*

Funkció: tájékoztatás, segítség a névválasztáshoz, névadáshoz

**Tervek:**

1. teljesség, havi frissítés → 2013. június: 3529 név

2. hiteles, egységes adatok → folyamatos javítás, **formalizált eredet** a nevek közel 70%-ánál

3. friss statisztikák, trendek → 2012-es adatok alapján

4. szabad hozzáférés → ahogy eddig is

## Példák

• női; név vége = *ó*; eredet jellege = becéző  
 → Anikó, Kató, Kleó, Magdó, Margó

• francia eredetű; a Mária névből  
 → Manon, Mariett, Marion

• 3 szótagos becéző férfinév  
*nincs* ↔ női: 57 db (amiből 29 db *-ka* végű)

•	↗	Alexander	↘	Albert
		Ármin		Andor
		Benett		Antal
		Benjamin		Arnold
		Boldizsár		Árpád
		Brendon		Béla
		Csanád		Dezső
		Domonkos		Elemér
		...		...