

ANALÝZA A ŠTATISTIKA MIEN A PRIEZVISK Z OTVORENÝCH ÚDAJOV SLOVENSKÝCH MIEST

ŠTEFAN TOTH (SK)

Abstrakt. Tento článok predstavuje analýzu mien a priezvisk obyvateľov slovenských miest založenú na verejne dostupných dátových sadách zo samospráv. Cieľom bolo zistiť, ktoré mužské a ženské mená sú najčastejšie v meste Žilina a v ostatných mestách, identifikovať najkratšie, najdlhšie a viacslovné krstné mená. Opisujú sa zdroje údajov, spôsob zberu a automatizovaného sťahovania dát, postup normalizácie a čistenia chybné zapísaných mien a priezvisk.

Kľúčové slová. Meno, priezvisko, štatistika, otvorené dáta, onomastika, Slovensko.

ANALYSIS AND STATISTICS OF NAMES AND SURNAMES FROM OPEN DATA OF SLOVAK CITIES

Abstract. This article presents an analysis of first and last names of residents in Slovak cities based on publicly available datasets from local municipalities. The aim was to identify the most common male and female names in the city of Žilina and in other cities, as well as the shortest, longest, and multi-word first names. The article describes the data sources, methods of data collection and automated downloading, and the process of normalizing and cleaning incorrectly recorded names and surnames.

Keywords. First name, last name, statistics, open data, onomastics, Slovakia.

Úvod

Mená a priezviská tvoria základné identifikačné znaky jednotlivcov. Ich výskyt a rozšírenie v populácii odráža kultúrne, jazykové a historické súvislosti. Výber krstného mena často závisí od tradícií, rodinného zázemia, náboženského presvedčenia alebo spoločenských trendov. Priezviská sú zase výsledkom dlhodobého jazykového vývoja a regionálnych osobitostí.

Niektoré slovenské mestá publikujú štatistiky o počtoch obyvateľov podľa krstných mien alebo priezvisk ako súčasť otvorených údajov. Tieto dáta umožňujú základnú onomastickú analýzu bez nutnosti prístupu k citlivým alebo nevelmi dostupným administratívnym registrovým zdrojom.

Tento článok sa zameriava na analýzu údajov z vybraných slovenských miest s cieľom:

- identifikovať najčastejšie mužské a ženské mená / priezviská,
- zistiť, ktoré mená a priezviská patria medzi najkratšie a najdlhšie,

- vyhľadať viacslovné alebo inak špecifické mená vyskytujúce sa v údajoch.

Súčasťou spracovania je opis spôsobu získavania údajov, čistenia, normalizácie a základných štatistických výstupov.

1. Zdroje údajov

Základným zdrojom údajov bola webová stránka mesta Žilina venovaná otvoreným dátam [2], kde je možné získať viacero štatistických výstupov o obyvateľoch, vrátane údajov typu „Počet občanov podľa krstných mien“ a „Počet občanov podľa priezvisk“. Tieto údaje sú dostupné vo formátoch PDF, XLSX, JSON, XML a XML.NET.

Na automatizované spracovanie je vhodný formát JSON alebo XML. Pre stiahnutie mien vo formáte JSON je možné použiť adresu https://egov.zilina.sk/Default.aspx?NavigationState=925:0::plac2171:_144151_5_8 a pre získanie priezvisk https://egov.zilina.sk/Default.aspx?NavigationState=935:0::plac2172:_144152_5_8.

Štruktúra týchto URL adries sa ukázala byť konzistentná aj pri iných mestách využívajúcich rovnakú platformu okrem mesta Trnava, ktoré v prípade datasetov neposkytovalo informáciu o pohlaví jednotlivých mien. Základné adresy miest boli získané prostredníctvom stránky www.pomosam.sk [4] (viď obr. 1), ktorá združuje elektronické služby viacerých samospráv.



Obr. 1. Mapa na webovej stránke www.pomosam.sk zobrazujúca samosprávy s ich erbami implementujúce systém POMOSAM poskytujúci otvorené datasety

Ako príklad možno uviesť okrem mesta Žilina aj mesto Martin [3], kde boli údaje sprístupnené v rovnakom formáte a na priamom odkaze. V mnohých prípadoch tak bolo možné získať mená a priezviská bez potreby manuálnej navigácie v rámci portálu – k základnej adrese samosprávy stačilo pripojiť vhodný postfix URL adresy.

Nie všetky mestá zo stránky [4] však tieto datasety sprístupňujú. Z celkového počtu 59 preverovaných miest poskytuje použiteľné údaje o krstných menách 19 samospráv a o priezviskách iba 15 samospráv. Mená boli dostupné iba v nasledujúcich mestách:

- | | |
|--------------------|----------------------|
| • Banská Bystrica, | • Pezinok, |
| • Dolný Kubín, | • Poprad, |
| • Dunajská Streda, | • Považská Bystrica, |
| • Hlohovec, | • Prešov, |
| • Humenné, | • Senica, |
| • Komárno, | • Trenčín, |
| • Levice, | • Trnava, |
| • Malacky, | • Vráble, |
| • Martin, | • Žilina. |
| • Modra, | |

Priezviská boli zverejnené iba v 15 mestách, ktoré tvoria podmnožinu miest z predchádzajúceho zoznamu. Sprístupnené údaje majú jednotnú štruktúru – obsahujú zoznam objektov obsahujúcich meno (`Krstné_meno`) alebo priezvisko (`Priezvisko`), pohlavie (`Pohlavie`) a k nemu prislúchajúci počet obyvateľov (`Počet`). Počas spracovania sa zistilo, že v prípade mien mesto Trnava malo iný formát JSON súboru ako ostatných 18 miest. Názov vlastnosti bol `Meno` namiesto `Krstné_meno` a chýbala vlastnosť udávajúca pohlavie mena. Toto muselo byť ošetrené, aby sa následne mohli všetky údaje z jednotlivých JSON súborov zlúčiť a skonvertovať do jednotného formátu na ďalšie spracovanie.

2. Spracovanie údajov

Získané datasety boli spracované pomocou vlastného algoritmického riešenia v jazyku C#. Sťahovanie bolo realizované programovo, pričom boli využité známe URL štruktúry jednotlivých miest publikujúcich údaje v JSON formáte.

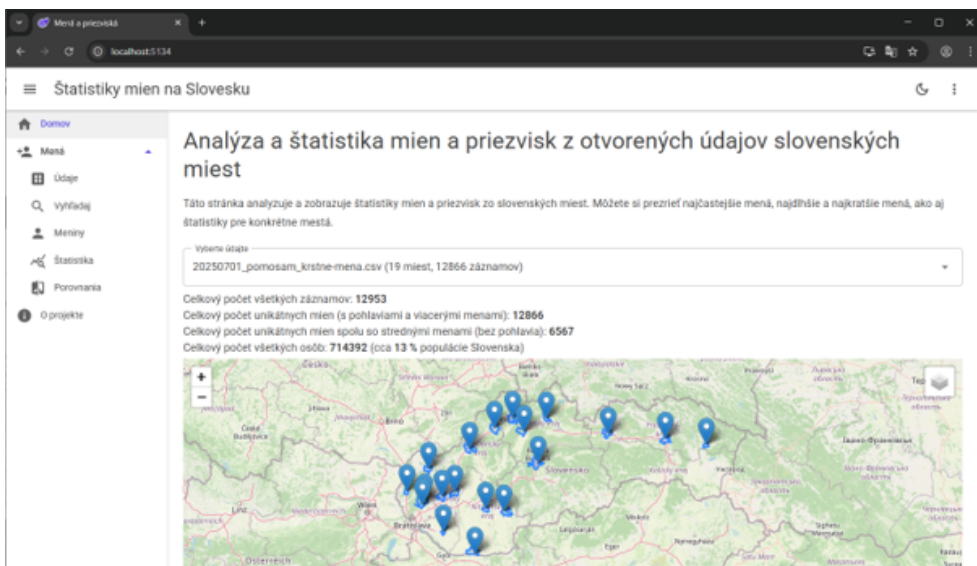
Po stiahnutí údajov nasledoval proces „čistenia“ a „normalizácie“, ktorého cieľom bolo odstrániť chyby vzniknuté zrejme nekonzistentným zápisom mien a priezvisk. Najčastejšie úpravy zahŕňali:

- Zjednotenie veľkosti písmen, napr. `Mlchaela` → `Michaela`.
- Odstránenie nadbytočných (dvojité) medzier alebo spojovníkov, napr. `Zuleica - Nekkar` → `Zuleica-Nekkar`.

- Vloženie medzier medzi spojené mená bez oddelenia, napr. LiaEmmaJoell → Lia Emma Joell.
- Odstránenie interpunkcie alebo nepísmenových znakov, napr. Maxim, Alex → Maxim Alex.

Každá z transformácií bola realizovaná automatizovane, pričom boli následne skontrolované a potvrdené aj manuálne. V rámci celkového datasetu bolo upravených približne 70 krstných mien, ktoré obsahovali zjavné odchýlky od štandardného zápisu. Výsledné dáta boli uložené v jednotnej CSV štruktúre a pripravené na štatistickú analýzu.

Na spracovanie a získanie štatistických analýz a výsledkov bola vytvorená vlastná webová aplikácia typu ASP.NET Core – Blazor WebAssembly [5]. Aplikácia poskytuje interaktívne rozhranie s mapou (obr. 2), tabuľkami a grafmi. Využíva open source knižnicu Leaflet pre Blazor na zobrazenie miest na OpenStreetMaps mape a material design komponenty na zoradenie a filtrovanie tabuliek s vlastnými algoritmami.



Obr. 2. Ukážka webovej aplikácie na spracovanie štatistických analýz so zobrazením miest, ktoré publikujú otvorené údaje, na mape z OpenStreetMap.

3. Analýza mien

3.1. Najčastejšie mená v celom datasete

Analýza krstných mien bola založená na spočítaní výskytov každého mena naprieč všetkými mestami, ktoré poskytli dáta. Zozbierané záznamy pokrývali

19 miest s celkovým počtom 714 392 obyvateľov k dátumu 1. 7. 2025, čo predstavuje približne 13 % populácie Slovenskej republiky (na základe počtu obyvateľov podľa [6]). V celkovom rebríčku dominujú tradičné slovenské mená s historickým a náboženským pozadím [1]. Medzi mužskými menami najčastejšie figurujú

Peter, Jozef, Ján, Martin a Michal,

zatiaľ čo medzi ženskými vedú

Mária, Anna, Zuzana, Jana a Katarína.

Tabuľka 1 uvádza desať najčastejších krstných mien v celom datasete. Hodnoty predstavujú súčet výskytov daného mena v údajoch všetkých zúčastnených samospráv.

3.2. Najčastejšie mená v meste Žilina

Najčastejšie mená v meste Žilina vo veľkej miere korešpondujú s celonárodným rebríčkom, pričom možno pozorovať aj niektoré lokálne špecifiká. Tabuľka 2 zobrazuje poradie najfrekventovanejších mien v Žiline. Medzi najčastejšie mužské mená patria

Peter, Michal, Jozef, Ján a Martin,

medzi ženskými dominujú

Mária, Anna, Katarína, Zuzana a Jana.

Ide o rovnaké mená ako v predchádzajúcej tabuľke, líšiace sa však poradím.

Tabuľka 1. Top 10 najčastejších mien v 19 mestách (13 % populácie SR).

Poradie	Meno	Počet
1.	Peter	21 071
2.	Mária	20 851
3.	Jozef	15 757
4.	Anna	15 313
5.	Ján	15 084
6.	Martin	13 364
7.	Zuzana	12 068
8.	Jana	11 932
9.	Katarína	11 747
10.	Michal	11 492

Tabuľka 2. Top 10 najčastejších mien v meste Žilina.

Poradie	Meno	Počet
1.	Peter	2 536
2.	Mária	2 096
3.	Anna	1 928
4.	Michal	1 707
5.	Jozef	1 633
6.	Ján	1 628
7.	Katarína	1 594
8.	Zuzana	1 556
9.	Martin	1 545
10.	Jana	1 426

Miestne preferencie sa vyznačujú mierne vyšším zastúpením niektorých mien, no všeobecný profil najčastejších krstných mien zostáva konzistentný s ostatnými mestami.

3.3. Viacslovné mená

V analyzovanom datasete sa nachádza aj množstvo prípadov, kde jedna osoba používa dve alebo viac krstných mien. Tieto viacslovné mená predstavujú špecifickú skupinu, často spojenú s náboženskou alebo kultúrnou tradíciou, rodinným pomenovávaním alebo snahou o osobitý prejav identity.

Top 15 viacslovných mien zobrazuje tabuľka 3. Z hľadiska frekvencie dominujú ženské kombinácie, pričom najčastejším druhým menom je **Mária**. Vysoký výskyt majú aj kombinácie pozostávajúce z dvoch mužských mien, najmä **Jakub** a **Ján**.

Tabuľka 3. Top 15 viacslovných mien v datasete.

Poradie	Meno	Počet
1.	Anna Mária	97
2.	Ema Mária	17
3.	Eva Mária	15
4.	Sofia Anna	13
5.	Emma Mária	11
6.	Mária Magdaléna	11
7.	Nina Mária	11
8.	Jakub Ján	10
9.	Sofia Mária	10
10.	Anna Kristína	9
11.	Jakub Jozef	9
12.	Ján Pavol	9
13.	Samuel Ján	9
14.	Viktória Anna	9
15.	Alex Ján	8

Z celkového pohľadu zostávajú viacslovné mená v menšine, no ich štruktúra a výber komponentov odrážajú silný vplyv kultúrnych a náboženských prvkov, predovšetkým v kombináciách obsahujúcich meno **Mária** alebo **Ján**.

3.4. Najkratšie a najdlhšie mená

V zozname krstných mien sa nachádzajú aj extrémne prípady z hľadiska dĺžky zápisu. Dĺžka mena bola počítaná ako počet znakov bez medzier.

Najkratšie zaznamenané mená obsahujú len 2 znaky, pričom ide prevažne o cudzojazyčné alebo netradičné tvary. Medzi ne patria napríklad:

Ai, Ba, Ce, De, Do alebo Qi (všetky s 1–2 výskytmi).

Medzi trojznakovými menami sa už nachádzajú tradičné slovenské mená s vysokým výskytom:

Ján (15 084), Eva (11 284), Ema (2 270) a Lea (995).

Najdlhšie jednoslovné mená v celom datasete sú

Franciscoandrés (15 znakov, 1 výskyt) a Michaelangelo (13 znakov, 1 výskyt).

Meno so 14 znakmi sa v datasete nevyskytlo. Dĺžku 12 znakov má viacero mien, napríklad (všetky po 1 výskyt):

Kadardžemilj, Lchagvasuren, Mademoiselle, Massimiliano, Scheherazade.

Z mien s 11 znakmi je najfrekvencovanejšie Christopher (50 výskytov).

Medzi ďalšie patria:

Szebasztián, Szilveszter (2 výskyty),

Abdolrazagh, Abdulrahman, Abdurrahman, Bernardetta, Vlastislava (1 výskyt).

Pri dĺžke 10 znakov sa vyskytujú už známe mená s vyšším zastúpením, ako napr.:

Stanislava (1 060), Kvetoslava (509), Maximilián (263).

Ak sa do štatistiky zahrnú aj viacslovné mená, najdlhším zaznamenaným zápisom je: Viktoria Tvorimira Europa Garcia (29 znakov, 1 výskyt).

Medzi ďalšie patrili napríklad (všetky s 1 výskytom):

Christopher Antoine Constant (26 znakov), Tomáš Adriano Anirejouritse (25),

Miroslav Cristiano Ronaldo (24) a Michaela Helena Magdaléna (23).

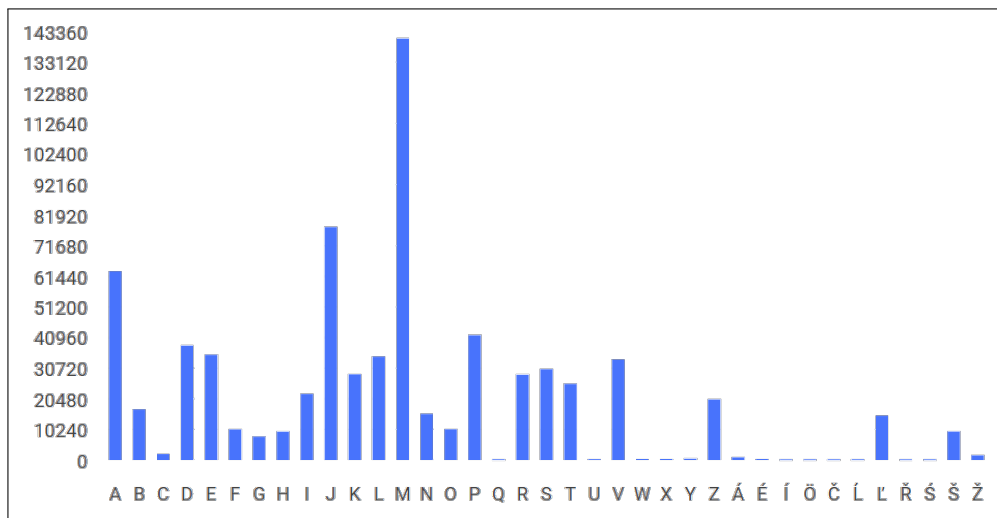
Zložené mená s tromi a viacerými časťami sa vyskytovali len výnimočne. Ide najmä o kombinácie klasických slovenských mien s cudzokrajnými alebo modernými prvkami.

Slovenská legislatíva zároveň vymedzuje maximálny počet mien, ktoré je možné zapísať do matriky. Podľa §2 ods. 1 zákona č. 300/1993 Z. z. o mene a priezvisku [7] môže mať fyzická osoba na Slovensku najviac tri mená. V praxi však možno vidieť aj prípady s viacerými menami, najmä ak ide o osoby cudzieho pôvodu alebo o prípady narodenia v zahraničí, kde matričné predpisy umožňujú širšiu voľnosť. Práve tieto záznamy sa v údajoch objavujú ako výnimočné viacslovné kombinácie so štyrmi alebo niekedy aj viacerými menami.

3.5. Začiatkové písmená mien

Frekvencia začiatkových písmen mien (obr. 3) v analyzovanom súbore naznačuje jazykové preferencie a obmedzenia slovenskej onomastiky. Najčastejšie sa mená začínajú na písmená M, J, A a P. Tieto iniciály pokrývajú veľkú časť tradičných slovenských mien, ako napríklad Mária, Ján, Anna a Peter.

Naopak, znaky ako Q, X či české Ř sa vyskytujú len ojedinele, čo súvisí s ich nízkym zastúpením v slovenskej abecede a obmedzeným výskytom v bežných krstných menách.



Obr. 3. Počet výskytov mien podľa začiatočného písmena mena.

4. Analýza priezvisk

V rámci analyzovaných údajov sa nachádzalo 78 710 výskytov priezvisk zozbieraných z 15 samospráv (cca 11 % populácie Slovenska). Najčastejšie priezviská v celom datasete zahŕňajú najmä tradičné slovenské a maďarské tvary, často odvodené od povolanií alebo všeobecne rozšírených rodových línií.

V čele rebríčka sa nachádzajú tvary *Kováč* a *Kováčová*, ktoré spolu tvoria najpočetnejšiu dvojicu mužskej a ženskej verzie jedného priezviska. Nasledujú *Horváth*, *Nagy*, *Baláz* a *Varga*, čo poukazuje aj na regionálne premiešanie slovensko-maďarského kultúrneho priestoru.

4.1. Najčastejšie priezviská v celom datasete

Najčastejšími priezviskami v celom datasete (tabuľka 4) sú *Kováč* a *Kováčová*, nasledované menami *Horváth*, *Nagy*, *Baláz* a *Varga*.

Podobne ako pri krstných menách, aj tu je možné sledovať tradičné a historicky stabilné vzorce. V rebríčku dominujú priezviská typické pre slovenský a maďarský jazykový okruh, pričom tvar s príponou *-ová* tvorí ženský ekvivalent mužskej formy.

4.2. Najčastejšie priezviská v meste Žilina

Najčastejšími priezviskami v Žiline (tabuľka 5) sú *Baláz* a *Balázová*, nasledované priezviskami *Kováčová*, *Kováč*, *Kubová* a *Hruška*.

Rebríček priezvisk v Žiline sa čiastočne odlišuje od celonárodného poradia. Zatiaľ čo v celkovom datasete dominujú aj priezviská maďarského pôvodu ako *Nagy*,

Tabuľka 4. Top 10 najčastejších priezvisk v celom datasete.

Poradie	Priezvisko	Počet
1.	Kováč	1 341
2.	Kováčová	1 305
3.	Horváth	1 028
4.	Nagy	1 015
5.	Baláž	1 001
6.	Varga	946
7.	Balážová	926
8.	Horváthová	902
9.	Tóth	848
10.	Vargová	838

Tabuľka 5. Top 10 najčastejších priezvisk v meste Žilina.

Poradie	Priezvisko	Počet
1.	Baláž	214
2.	Balážová	168
3.	Kováčová	153
4.	Kováč	128
5.	Kubová	125
6.	Hruška	111
7.	Hrušková	96
8.	Šoška	93
9.	Šošková	91
10.	Závodský	91

Varga či Tóth, v Žiline sa tieto mená nevyskytujú medzi desiatimi najčastejšími. Namiesto toho sa objavujú regionálne špecifické priezviská ako Šoška či Závodský, ktoré môžu odrážať geografickú polohu mesta na severe Slovenska.

4.3. Najkratšie priezviská

V rámci analyzovaných údajov sa vyskytlo niekoľko priezvisk pozostávajúcich iba z dvoch znakov. Ide prevažne o zriedkavé alebo cudzojazyčné tvary, ktoré sa objavili s nízkou frekvenciou – väčšinou jeden až sedem výskytov. Medzi najkratšie zaznamenané patria napríklad:

Áč, Ag, Ág, Ji, Ič, Lý, Uj, Új, Kő alebo Qi.

Tieto priezviská môžu pochádzať z odlišných jazykových oblastí alebo predstavovať skrátené či adaptované formy zložitejších mien. V slovenskom kontexte sú výnimočné a netvorí žiadnu významnejšiu skupinu.

Ich výskyt dokrešľuje celkové rozpätie priezvisk v dátach a potvrdzuje, že okrem bežných tvarov sa v údajoch nachádzajú aj atypické a extrémne prípady, ktoré môžu byť výsledkom migrácie, naturalizácie alebo chybného zápisu.

4.4. Najdlhšie priezviská

Na opačnom konci spektra dĺžky sa v dátach vyskytli aj priezviská so značným počtom znakov. Najdlhšie zaznamenané priezviská obsahovali až 27 znakov:

Ahmed Abouelyazeid Abdelkarim a Nesvadbová Brettschneiderová.

Medzi ďalšie dlhé priezviská patria napríklad:

Edirisinghe Dissanayakalage (26 znakov),

De Sousa Gaspar Kundračiková (25), Freiherr Von Andrian Werburg (25),

Aschengeschwandtnerová (22) alebo Luna Garcia De La Infanta (21).

boli najčastejšie mená v celkovom súbore aj v meste Žilina, najkratšie a najdlhšie jednoslovné i viacslovné mená, ako aj základné formálne charakteristiky, napríklad začiatkové písmená či dĺžku zápisu.

Výsledky potvrdzujú dominanciu tradičných slovenských mien a zároveň prítomnosť netypických a cudzokrajných mien ako prejav kultúrnej diverzity. Analyzovaná bola aj dĺžka krstných mien – od najkratších dvojslovných po zriedkavé viacslovné zápisy so značným počtom znakov. V prípade viacslovných mien sa najčastejšie vyskytujú kombinácie s menom Mária alebo Ján, čo naznačuje silné prepojenie na náboženskú tradíciu. Priezviská kopírujú podobný trend. Analýza dĺžky a tvaru ukázala široké spektrum – od dvojpísmenových až po takmer tridsaťznakové priezviská.

Všetky použité údaje pochádzali zo zverejnených verejných databáz v už agregovanej forme, bez väzby na konkrétne osoby. Pri ich spracovaní teda nedochádzalo k spracúvaniu osobných údajov v zmysle GDPR, a nevzniklo riziko reidentifikácie jednotlivcov [8].

Napriek dostupnosti otvorených údajov vo formáte vhodnom na automatizované spracovanie existujú viaceré obmedzenia. Z celkového počtu slovenských miest zverejňuje štruktúrované dáta o menách a priezviskách len menšia časť, pričom často chýbajú kľúčové atribúty ako napríklad pohlavie mena.

Otvorené údaje tohto typu predstavujú cenný zdroj pre jazykový, kultúrny aj demografický výskum. Ich rozširovanie a systematická podpora zo strany samospráv môže výrazne prispieť k lepšiemu porozumeniu vývoja spoločnosti a pomenúvacích trendov v slovenskom prostredí.

Podakovanie. Autor ďakuje všetkým mestám, ktoré zverejnili svoje údaje, za otvorený prístup k údajom a komunite OSSConf za inšpiratívne spätné väzby.

Literatúra

- [1] MAJTÁN, M. – POVAŽAJ, M.: *Meno pre naše diéta*, Bratislava: Obzor, 1993, ISBN 80-215-0222-3.
- [2] MESTO ŽILINA: *Open data*, Otvorené dáta mesta Žilina, <https://egov.zilina.sk/default.aspx?NavigationState=1100:0:>, [cit. 1. 7. 2025].
- [3] MESTO MARTIN: *Open data*, Otvorené dáta mesta Martin, <https://egov.martin.sk/Default.aspx?NavigationState=1100:0:>, [cit. 1. 7. 2025].
- [4] CORA GEO, s.r.o.: *POMOSAM – Portál modernej samosprávy*, <https://www.pomosam.sk/>, [cit. 1. 7. 2025].
- [5] MICROSOFT: *ASP.NET Core Blazor WebAssembly*, Oficiálna dokumentácia, <https://learn.microsoft.com/en-us/aspnet/core/blazor/>, [cit. 27. 7. 2025].
- [6] ŠTATISTICKÝ ÚRAD SR: *Počet obyvateľov Slovenskej republiky k 1. 1. 2025*, <https://www.statistics.sk>, [cit. 1. 7. 2025].
- [7] NÁRODNÁ RADA SR: *Zákon č. 300/1993 Z. z. o mene a priezvisku*, §2 ods. 1, Zbierka zákonov Slovenskej republiky, <https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/1993/300/>, [cit. 28. 7. 2025].

- [8] EURÓPSKY PARLAMENT A RADA: *Nariadenie Európskeho parlamentu a rady (EÚ) 2016/679 o ochrane fyzických osôb pri spracúvaní osobných údajov (GDPR)*, Úradný vestník EÚ L119, 4. mája 2016, <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

Kontaktná adresa

Ing. Štefan Toth, PhD., Katedra softvérových technológií, Fakulta riadenia a informatiky, Žilinská univerzita v Žiline, Univerzitná 8215/1, 010 26 Žilina, Slovensko,
E-mailová adresa: `stefan.toth@uniza.sk`