

# ProPONAR 1.0

## Dokumentace k softwaru



**Nářečí českého jazyka interaktivně.**  
**Dokumentace a zpřístupnění mizejícího jazykového dědictví jako nedílné součásti regionálních identit.**  
DG20P02OVV029; poskytovatel podpory Ministerstvo kultury, Program NAKI II.

# Obsah

<b>Základní údaje</b>	<b>3</b>
<b>Charakteristiky výsledku</b>	<b>4</b>
K čemu ProPONAR slouží?	4
Proč je ProPONAR webová aplikace?	4
Čím je software nový a jedinečný?	5
Naplnění cílů programu NAKI II a jeho očekávaných přínosů	5
Beta testování a využití softwaru	6
<b>Technické parametry</b>	<b>7</b>
API server	7
Frontend	7
Autentizace uživatelů	8
Databázové schéma	8
Zdrojové kódy	8



# Základní údaje

**Název výsledku:**

ProPONAR

**Verze:**

1.0

**Druh výsledku:**

R - software

**Lokalizace výsledku:**

<https://proponar.ujc.cas.cz/>

**Manuál:**

<https://proponar.ujc.cas.cz/help/manual.pdf>

**Zdrojové kódy:**

<https://github.com/proponar>

**Webová stránka výsledku:**

<https://geoinformatics.upol.cz/app/nareci/vystupy/proponar/>

**Název vlastníka:**

Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i.

**Technické parametry výsledku:**

Webová aplikace pro tvorbu multimediální prostorové dialektologické databáze s API serverem implementovaným v jazyce Ruby pomocí frameworku Ruby on Rails a s frontendem v jazyce JavaScript jako SPA prostřednictvím systému React. Jako úložiště pro data je použita SQL databáze PostgreSQL.

**Ekonomické parametry výsledku:**

Během 6 měsíců užívání vlastníkem bylo prostřednictvím softwaru pořízeno 29 766 nářečních ukázek, což o řád (přesně 17,6 x) přesáhlo počty očekávané v přihlášce za celou dobu trvání projektu.

**Druh možnosti využití výsledku jiným subjektem:**

A - k využití výsledku jiným subjektem je vždy nutné nabytí licence

**Požadavek na licenční poplatek:**

N - poskytovatel licence na výsledek nepožaduje licenční poplatek

**Typ licence:**

MIT

# Charakteristiky výsledku

Software ProPONAR je první částí propojeného komplexu elektronických dialektologických nástrojů, vznikajících v rámci projektu *Nářečí českého jazyka interaktivně. Dokumentace a zpřístupnění mizejícího jazykového dědictví jako nedílné součásti regionálních identit.* (DG20P02OVV029; poskytovatel podpory Ministerstvo kultury, Program NAKI II.)

## K čemu ProPONAR slouží?

Software je určen k pořizování a organizaci multimediálních dat pro interaktivní analytické mapy (tedy mapy s interaktivními body, po jejichž rozkliknutí je možné zobrazit nebo přehrát multimediální obsah lokalizovaný do daného bodu: text, obrázek, audio, video). Software však neslouží přímo k prezentaci map (ty bude možné zobrazit na webovém geoportálu DiaMa - odevzdáván 12/2022), nýbrž k výstavbě prostorové databáze, z níž budou mapy generovány. Software se přitom specializuje na organizaci nářečních deklinačních dat, pro něž nabízí speciální funkcionality, i když je možné ho využít i pro prostorové databáze/mapy s jinými účely (např. nářeční mapy konjugační, hláskoslovné nebo lexikální, etnografické mapy reálií, mapy kulturních památek apod.)

## Proč je ProPONAR webová aplikace?

Vzhledem k tomu, že geoportál DiaMa potřebný k zobrazení mapového výsledku, bude dostupný pouze online, byl při vývoji přehodnocen původní záměr vytvořit aplikaci desktopovou, namísto toho byl ProPONAR vytvořen jako responzivní aplikace webová.

Výhody této změny:

- umožňuje práci více lidí na jedné online databázi (jako nutnost se to ukázalo i u našeho výzkumného týmu při covidových lockdownech)
- uživatel není vázán na jedno zařízení a může přidávat do databáze např. i přímo při terénním výzkumu prostřednictvím chytrého telefonu
- je uživatelsky přívětivější tím, že nenutí stahovat a instalovat si software, responzivita aplikace povede k její větší užitečnosti a rozšíření v cílových skupinách
- počítáním přístupů je možné lépe vykazovat využívání softwaru než jen počtem stažení

## Čím je software nový a jedinečný?

Inovativnost softwaru spočívá především v možnosti snadno organizovat dialektologická deklinační data. Dialektologická data obecně nejsou vhodná jako trénovací data pro strojové učení (z jedné lokality nebývají dost rozsáhlá a z více lokalit jsou jazykově heterogenní a nezobecnitelná). ProPONAR je koncipován právě pro zpracování takovýchto dat. Umožňuje poloautomatické morfologické značkování (určování, tagování) tvarů a jejich další snadnou organizaci. Po vzniku geoportálu DiaMa si bude moci každý uživatel vyfiltrovat a stáhnout data pro zvolenou oblast a zvolené slovo (formou nářečních/prostorových deklinačních typů). Na jejich základě se budou tvary zvoleného slova automaticky určovat, uživatel pak bude pouze v případě nutnosti desambiguovat (zjednotňovat) pádové homonymie. Velká část dat, která budou na geoportálu DiaMa k dispozici, je přitom výzkumným týmem zpracovávána a organizována rovněž prostřednictvím softwaru ProPONAR.

## Naplnění cílů programu NAKI II a jeho očekávaných přínosů

Dialektologické analytické mapy s multimediálním obsahem jsou obecně velmi oblíbené, ale není jich mnoho. V českém prostředí nebyla před schválením tohoto projektu žádná, v r. 2020 pak vznikla v Českém národním korpusu „Mapka“ (<https://korpus.cz/mapka/>), která má však jen 13 interaktivních bodů. Téměř všechny takovéto mapy, vzniklé ve světě (a teď i u nás), mají shodné či velmi podobné formální charakteristiky. Přestože je možné těchto vlastností dosáhnout (s jistými omezeními) i pomocí komerčních aplikací, byl v naprosté většině do tvorby těchto map zapojen programátor, mnohdy i (webový) kartograf. Právě to z tvorby těchto map dělá poněkud nákladnou a výlučnou záležitost.

Software ProPONAR naproti tomu umožní tvorbu takovýchto map každému počítačově gramotnějšímu uživateli, a to prostřednictvím sestavení multimediální prostorové databáze. K jejímu vytvoření budou stačit pouze dialektologická (příp. i etnografická nebo jiná) data.

### Software ProPONAR tak naplňuje řadu očekávaných přínosů programu NAKI II:

- Umožněním vytváření exportovatelných nářečních databází a map napomůže **uchování nehmotného kulturního dědictví pro další generace** (Očekávané přínosy bod 1)
- Možnost snadného sestavení vlastní mapy má potenciál motivovat nářeční výzkum (zvl. u laiků a studentů), a tím stimulovat individuální i kolektivní paměť a přispívat k **záchraně a zlepšování hodnot kulturního dědictví** (Očekávané přínosy bod 2)
- Software umožní sestavovat intuitivní dialektologické mapy nejen výzkumnému týmu, ale komukoli, což napomůže jejich sdílení a šíření v regionálních komunitách, čímž bude zajištěno **zpřístupnění kulturního dědictví široké obci zájemců a uživatelů** (Očekávané přínosy bod 3).

- Předpokládá se využití softwaru ProPONAR při tvorbě vysokoškolských kvalifikačních prací z oboru dialektologie, čímž budou **využity výsledky výzkumu ve vzdělávání, založeném na poznání národní identity a kultury** (Očekávané přínosy bod 5).
- Software je také ze své povahy realizací **výzkumu nástrojů pro systematickou záchranu mizejících památek kulturního dědictví** (Očekávané přínosy bod 7d).

## Beta testování a využití softwaru

Software byl v beta verzi testován jednak výzkumným týmem dialektologického oddělení ÚJČ, jednak zvolenými reprezentanty předpokládaných uživatelů.

Výzkumný tým testoval beta verzi ProPONARu od prosince 2020 do konce dubna 2021, průběžně byly řešeny a opravovány nedostatky a vylepšována funkcionality. Od května 2021 pak už tým ÚJČ vytváří pomocí softwaru databázi DiDa (odevzdávána 6/2022). Software se během této doby ukázal jako mimořádně produktivní. Během 6 měsíců užívání bylo jeho prostřednictvím zpracováno 29 766 nářečních ukázek, což je o řád více než se očekávalo. Počty, kterých mělo být dle přihlášky dosaženo za celé trvání projektu, se podařilo přesáhnout 17,6x jen během těchto šesti měsíců.

V poslední fázi vývoje byl software předložen k testování týmu UPOL a bylo vybráno a osloveno 36 reprezentantů cílových skupin a budoucích uživatelů softwaru (dialektologové vně výzkumného týmu, bohemisté, počítačové lingvisté, vysokoškolští studenti jazykovědných oborů, slavisté, etnografové, sběratelé nářečí). Byť byl software v té době plně funkční, výsledkem bylo množství cenných připomínek, z nichž téměř všechny byly zapracovány.

Postupné úpravy zdrojového kódu jsou zaznamenány v githubových repozitářích: <https://github.com/proponar>.



# Technické parametry

Aplikace se skládá z těchto dvou základních komponent:

- API server.
- JavaScript Frontend.

Jako úložiště pro data je použita SQL databáze PostgreSQL. Data příloh jsou uložena na souborovém systému serveru.

API i frontend jsou zpřístupněny se serveru prostřednictvím protokolu HTTPS s certifikátem "Let's encrypt" (<https://letsencrypt.org/>).

Na serveru je provozován operační systém Ubuntu Linux.

## API server

API server je implementován v jazyce Ruby pomocí frameworku "Ruby on Rails" (<https://rubyonrails.org/>).

API server komunikuje s SQL databází PostgreSQL a programem zajišťujícím export do formátu DOCX.

Exporty do formátu DOCX jsou implementovány samostatným programem v jazyce Python. Pomocí knihovny "docx" (<https://python-docx.readthedocs.io/en/latest/>).

## Frontend

Frontend (uživatelské rozhraní) je implementován v jazyce JavaScript jako SPA (single page application) pomocí systému "React" (<https://reactjs.org/>). Používá knihovnu komponent "Material UI" (<https://material-ui.com/>).

Pro zobrazení náhledu lokalizace je použita služba mapy.cz společnosti Seznam.cz.

Frontend je kompatibilní se všemi rozšířenými prohlížeči, které jsou podporovaný svými výrobci (Chrome, Firefox, Edge, Safari, Opera a další).

Vzhled a funkcionality frontendu je detailně popsána manuálem

(<https://proponar.ujc.cas.cz/help/manual.pdf>).

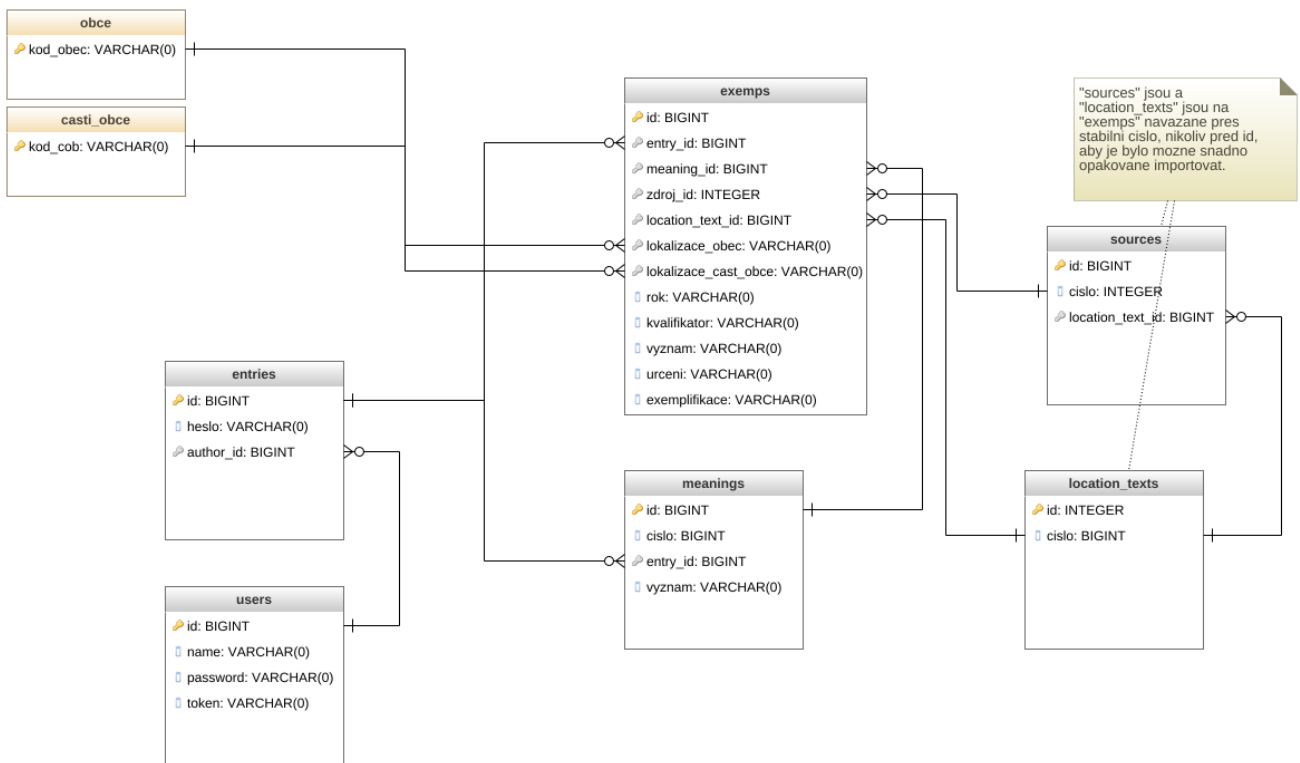
## Autentizace uživatelů

Autentizace uživatelů je prováděna proti seznamu uživatelů uloženému v databázi a pomocí protokolu OAuth2 poskytovatele identity společnosti Google

(<https://developers.google.com/identity/protocols/oauth2>).

## Databázové schéma

Databáze se skládá z tabulek, které zabezpečují lokalizaci (okresy, obce, části obce...), části pro modelování hesel a exemplifikací (hesla, významy, exemplifikace...) a dalších podpurných částí (zdroje, uživatelé...).



(<https://github.com/proponar/dida-entry/blob/master/doc/DIDA-base-schema-white.png>)

## Zdrojové kódy

Kompletní zdrojové kódy aplikace jsou hostované na platformě GitHub na adrese <https://github.com/proponar>.

Na serveru je provozován operační systém Ubuntu Linux.