Rozpoznávač řeči adaptovaný pro generování dialektologické transkripce z audionahrávek

verze 1.0

Uživatelská příručka

a

Technická dokumentace

Jazyková paměť regionů České republiky.
Metody strojového učení pro uchování, dokumentaci a prezentaci nářečí českého jazyka.

Ministerstvo kultury ČR, NAKI III, DH23P03OVV010.
Program na podporu aplikovaného výzkumu v oblasti národní a kulturní identity 2023–2027.

Základní údaje

Název výsledku

Rozpoznávač řeči adaptovaný pro generování dialektologické transkripce z audionahrávek

Verze

1.0

Druh výsledku

R – software

Lokalizace výsledku

https://jamap.cz/prepis

Lokalizace dokumentace

https://jamap.cz/prepis/dokumentace.pdf

Vlastník

Vysoké učení technické v Brně – Ústav pro jazyk český AV ČR, v. v. i.

Stručný popis produktu

Systém na základě vstupní zvukové nahrávky automaticky generuje přepis mluveného projevu do textu včetně zachování nářečních specifik, a to i jevů diferenčních, jimiž nedisponuje spisovný jazyk. Výstup je formátován podle pravidel dialektologické transkripce, která byla definována v *Metodice pro převod strukturovaných znalostí z oboru dialektologie do strojového učení* (Šimečková a kol., 2025, s. 56–62; dostupné online). Systém je optimalizován pro záznamy tradičních dialektů češtiny z území České republiky, přičemž dokáže zachytit regionální varianty češtiny, které standardní rozpoznávače řeči ignorují.

Technické parametry produktu

Systém vychází z volně dostupného Whisper modelu od OpenAI (Radford, 2022; encoder-decoder architektura, velikost large-v3-turbo). Model byl dále trénován (specializovaný fine-tuning), a to na 16,5 hodinách nářečních nahrávek poskytnutých od ÚJČ AV ČR. Jde o nahrávky různého stáří, získávané sběry od 60. let 20. století po současnost. Nahrávky byly datově segmentovány, čištěny a lingvisticky anotovány. Pro uživatelskou jednoduchost bylo pro tento systém vytvořeno webové rozhraní, které umožňuje otestovat rozpoznávač řeči přes internetový prohlížeč jako „cloudovou službu“.

Ekonomické parametry výsledku

Software významně zrychluje a zpřesňuje přepis nahrávek, které se stanou součástí knižní monografie a otevřené fonotéky, tedy výstupů plánovaných v rámci příslušného projektu NAKI III. Veškeré přepisy musely být dosud prováděny manuálně, nyní lze proces výrazně zautomatizovat a zefektivnit. V současnosti disponuje dialektologické oddělení 2 587 nahrávkami s celkovou stopáží 1 631 hodin (ke dni 10. 9. 2025), přepisy byly dosud pořízeny

pouze u 23 % nahrávek, avšak většinově na úrovni reprezentativních ukázek o délce cca 5 minut. Ruční přepis hodinového záznamu řeči je v průměru pořizován 5 až několik desítek hodin (Šimečková a kol., 2025, s. 58), možnost automatického přepisu tak představuje výrazné zjednodušení v procesu zpracování nářečního audiálního materiálu.

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

Licence

Software slouží k interní aplikaci pro potřeby projektu a Ministerstva kultury ČR. Po domluvě je možné využití i třetími stranami, a to při dodržení zásady uvedení zdroje a zachování licence. Celé znění licence:

Autorské právo © [2024]
[Vysoké učení technické v Brně – Ústav pro jazyk český AV ČR, v. v. i.] Licencováno podle Apache License, verze 2.0 (dále „Licence“); tento soubor není možné použít jinak než v souladu s Licencí. Kopii Licence můžete získat na adrese:
http://www.apache.org/licenses/LICENSE-2.0
Pokud to nevyžaduje platný zákon nebo není uvedeno jinak, je software distribuován podle Licence na bázi jak stojí a běží, bez jakýchkoliv záruk a podmínek, ať výslovných nebo předpokládaných. Podrobnosti o Licenci a jejích omezeních jsou k dispozici v samotné Licenci.

Pro uvedení zdroje doporučujeme formulaci:
Rozpoznávač řeči adaptovaný pro generování dialektologické transkripce z audionahrávek. Verze 1.0. Vysoké učení technické v Brně – Ústav pro jazyk český AV ČR, v. v. i., 2025.

Anglická verze:
Speech Recognizer Adapted for Generating Dialectological Transcriptions from Audio Recordings. Version 1.0. Brno University of Technology – Czech Language Institute, Czech Academy of Sciences, 2025.

Zdrojový kód

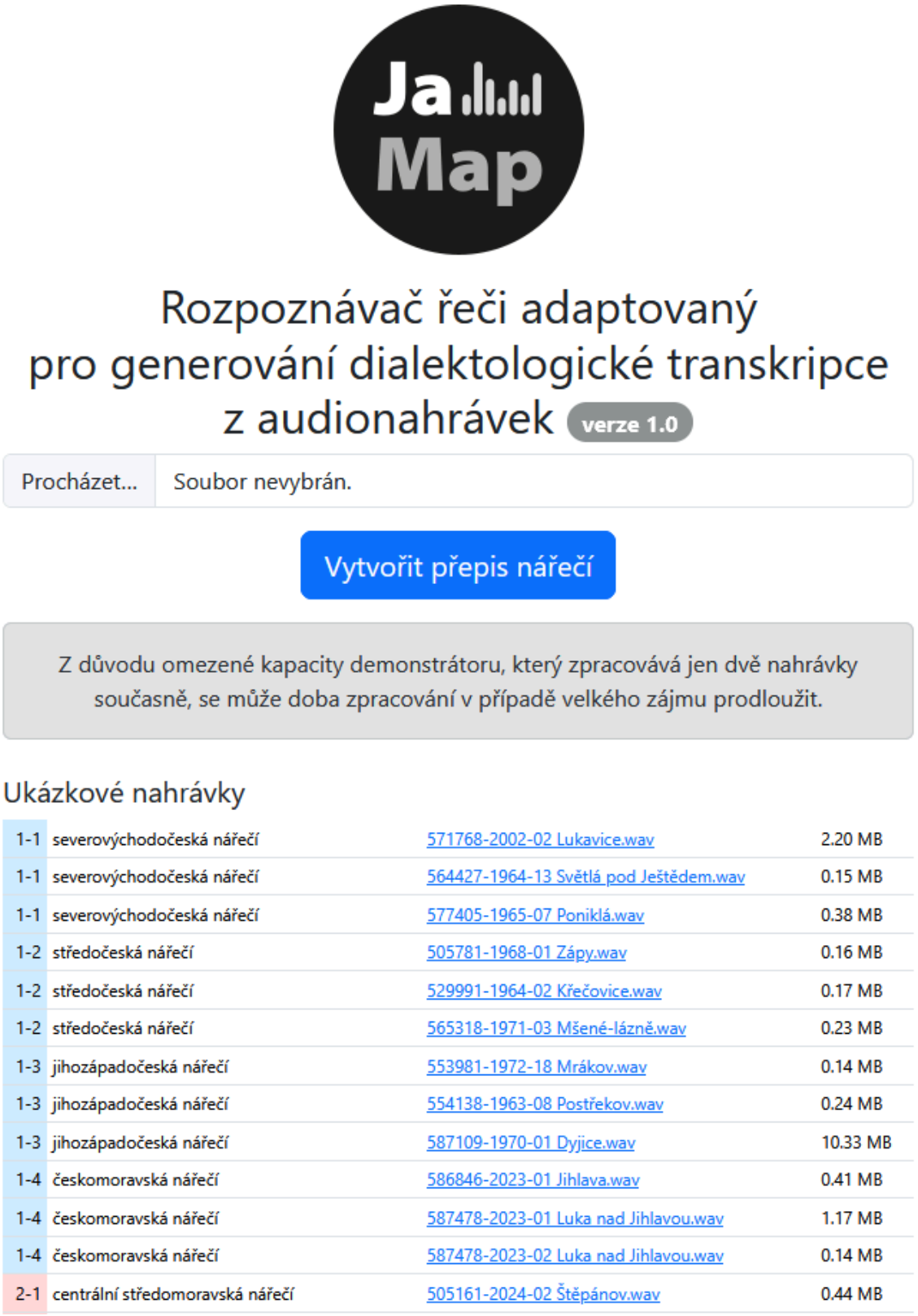
https://jamap.cz/prepis/zdrojovy-kod.tar.gz

Uživatelská příručka

Krok 1
Ovládání je velice intuitivní. Na webové stránce transkriptoru uživatel vybere soubor s nářeční nahrávkou ze svého počítače, a to pomocí tlačítka Vybrat soubor. Tlačítkem Vytvořit přepis nářečí pak spustí vlastní transkriptor. Pokud si chce uživatel software pouze otestovat, aplikace nabízí 39 ukázkových nahrávek k vyzkoušení.

Poznámka: Čísla v levém sloupci určují nářeční oblast, v níž byla nahrávka pořízena a kterou prezentuje jazykový kód mluvčího. Celkem jde o 13 nářečních skupin, konkrétně:
1-1 severovýchodočeská nářečí, 1-2 středočeská nářečí, 1-3 jihozápadočeská nářečí, 1-4 českomoravská nářečí, 2-1 centrální středomoravská nářečí, 2-2 jižní středomoravská nářečí, 2-3 západní středomoravský okrajový úsek, 2-4 východní středomoravský okrajový úsek, 3-1 jižní

východomoravská nářečí, 3-2 severní východomoravská nářečí, 3-3 kopaničářská nářečí, 4-1 slezskomoravská nářečí, 4-2 slezskopolská nářečí.



Obr. 1: Náhled rozhraní

Krok 2

Výstupem jsou automaticky generované přepisy, pořízené podle současných pravidel dialektologické transkripce. Přepisy zohledňují fonetickou stránku řeči pomocí ustálené soustavy znaků, a to včetně diferenčních hláskových jevů a také rozdílů ve vokalické kvantitě, měkkostí a asimilací. Přepisy pracují se znaky české abecedy doplněnými speciálními grafy, užívanými např. pro různé *l*-ové realizace (*ł* = tvrdé *l*, *ṵ* = obalované *l*) nebo

různé kvalitativní obměny vokálů (např. *ę* = krátké otevřené/široké *e*; *ǫ* = krátké otevřené/široké *o*). Pro sluchovou kontrolu lze spustit nahraný audiosoubor, který se automaticky načte pod částí s přepisem.

Poznámka: Pro vyšší úspěšnost přepisu je nezbytné pracovat se záznamy tradičních nářečních promluv (nikoliv smíšených nebo nenářečních projevů), a to jedné osoby (nikoliv více osob), která byla v době pořízení nahrávky zástupcem starší generace, přičem splňovala i další požadavky kladené na ideálního nářečního mluvčího (k tomu Šimečková, 2024). Software je v této verzi optimalizován na záznamy pořízené starší nahrávací technikou, zejména na audiální dokumenty z 60. a 70. let 20. století nahrané cívkovým magnetofonem Tesla Sonet Duo.

Obr. 2: Náhled automaticky přepsaného textu

Charakteristiky výsledku

Rozpoznávač řeči adaptovaný pro generování dialektologické transkripce z audionahrávek je druhý elektronický výstup z řady softwarů vyvíjených v projektu Jazyková paměť regionů České republiky. Metody strojového učení pro uchování, dokumentaci a prezentaci nářečí českého jazyka. Ministerstvo kultury ČR, NAKI III, DH23P03OVV010. Software umožňuje u vstupní audionahrávky zachycující jakékoli teritoriální nářečí českého národního jazyka pořídit automatický přepis v dialektologické transkripci.

Jedinečnost a novost softwaru

Jde o první veřejně dostupný automatický přepisovač regionálně rozrůzněné češtiny, a to jak po stránce architektury softwaru, tak trénovacími daty, na nichž byl nástroj vyvinut (viz technická dokumentace níže).

Naplnění očekávaných dopadů programu NAKI III a využití softwaru

Software usnadňuje práci s nářečními nahrávkami a jejich audiální podobou, umožňuje **automaticky přepisovat nářeční zvukové projevy**, a to v podobě standardizované dialektologické transkripce. Je využitelný např. při digitalizaci fonoték a archivů mluveného slova, při zpracování záznamů z terénních výzkumů, též v běžné praxi lingvistů, etnologů, orálních historiků a dalších odborníků běžně pracujících s mluvenými projevy. Jeho využitelnost se ještě znásobí po zveřejnění dalšího nástroje vyvíjeného v rámci tohoto projektu, který bude automaticky převádět dialektologickou transkripci do hojně užívané transkripce folklorní.

Jednoduché webové rozhraní, snadné ovládání a srozumitelná prezentace výsledků rozšiřují okruh možných uživatelů od odborné veřejnosti až po studenty, pedagogy či laické zájemce o jazyk. Software má také velký potenciál ve vzdělávání, protože umožňuje prezentovat živou podobu nářečí a její transformaci do psané podoby.

Těmito charakteristikami software vyhovuje důrazu, který je v programu NAKI III kladen na směřování výzkumu k výsledkům podporujícím rozvoj inteligentních specializací ve smyslu RIS3.
Odpovídá také následujícím bodům očekávaných dopadů programu NAKI III:
3) zpřístupnění kulturního dědictví široké obci zájemců a uživatelů,
4) posílení a integrace sociálně ekonomického uplatnění kulturního dědictví ve společnosti,
6) výzkum nástrojů a jejich ověření pro:

> a) aktivní podíl na rozvíjení národní identity jako jednoho z elementů spoluvytváření evropské identity a kultury,
> b) systematickou dokumentaci a prezentaci způsobu života a kultury menšin v minulosti i současnosti,
> k) systematické a efektivní využívání prostředků technologií pro rozvoj národní kultury a kulturní identity obyvatelstva.

Technická dokumentace

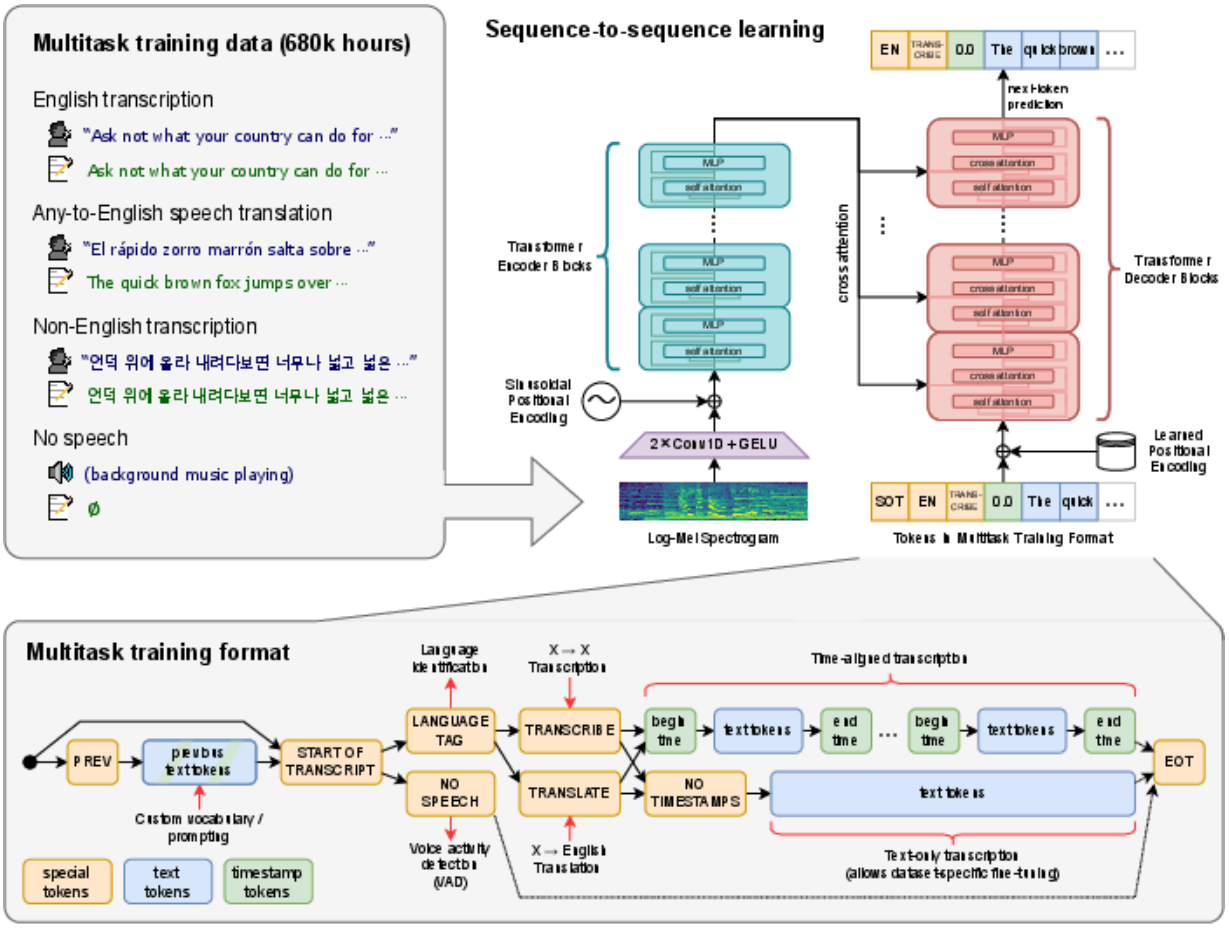
Architektura

Základ tvoří Whisper encoder-decoder architektura (viz Radford, 2022)**.** Použit byl princip transfer learningu: předtrénování na velkých obecně dostupných datech spolu se specializovaným fine-tuningem na českých nářečních datech (s celkovou stopáží 16,5 h).

Trénovací data:

Výstup:
- Systém vrací čistý textový přepis.
- Pro účely výzkumu je možné doplnit i časové značky (timestamps).
- Přepis je optimalizován pro starší historické nahrávky, ale dokáže pracovat i s novými záznamy.

Architektura systému je zobrazena na obr. 3.



Obr. 3: Architektura Whisper systému

Trénovací data

Systém byl trénován na českých nářečních datech, poskytnutých dialektologickým oddělením ÚJČ AV ČR, v. v. i. Jedná se o unikátní data získávaná terénním výzkumem od 50. let 20. století do současnosti, která jsou uložena v Archivu zvukových záznamů nářečních promluv a nyní převáděna do budované Databáze nářečních promluv pro odbornou veřejnost (k těmto zdrojům viz Šimečková, 2024). Data byla segmentována, vyčištěna a anotována lingvisty.
Pro trénování systému bylo použito 16,5 hodin nahrávek z různých nářečních skupin.

Popis demonstrátoru

Automatický přepis nářečí je implementován jako nástroj v příkazové řádce v jazyce Python, daná forma je tudíž primárně vhodná pro použití softwaru v rámci vědeckovýzkumné činnosti, nikoliv pro testování méně technicky zdatnými uživateli. Z toho důvodu bylo implementováno i webové rozhraní a „cloudové“ zapouzdření rozpoznávače. Zjednodušené schéma demonstrátoru zachycuje obrázek 4.

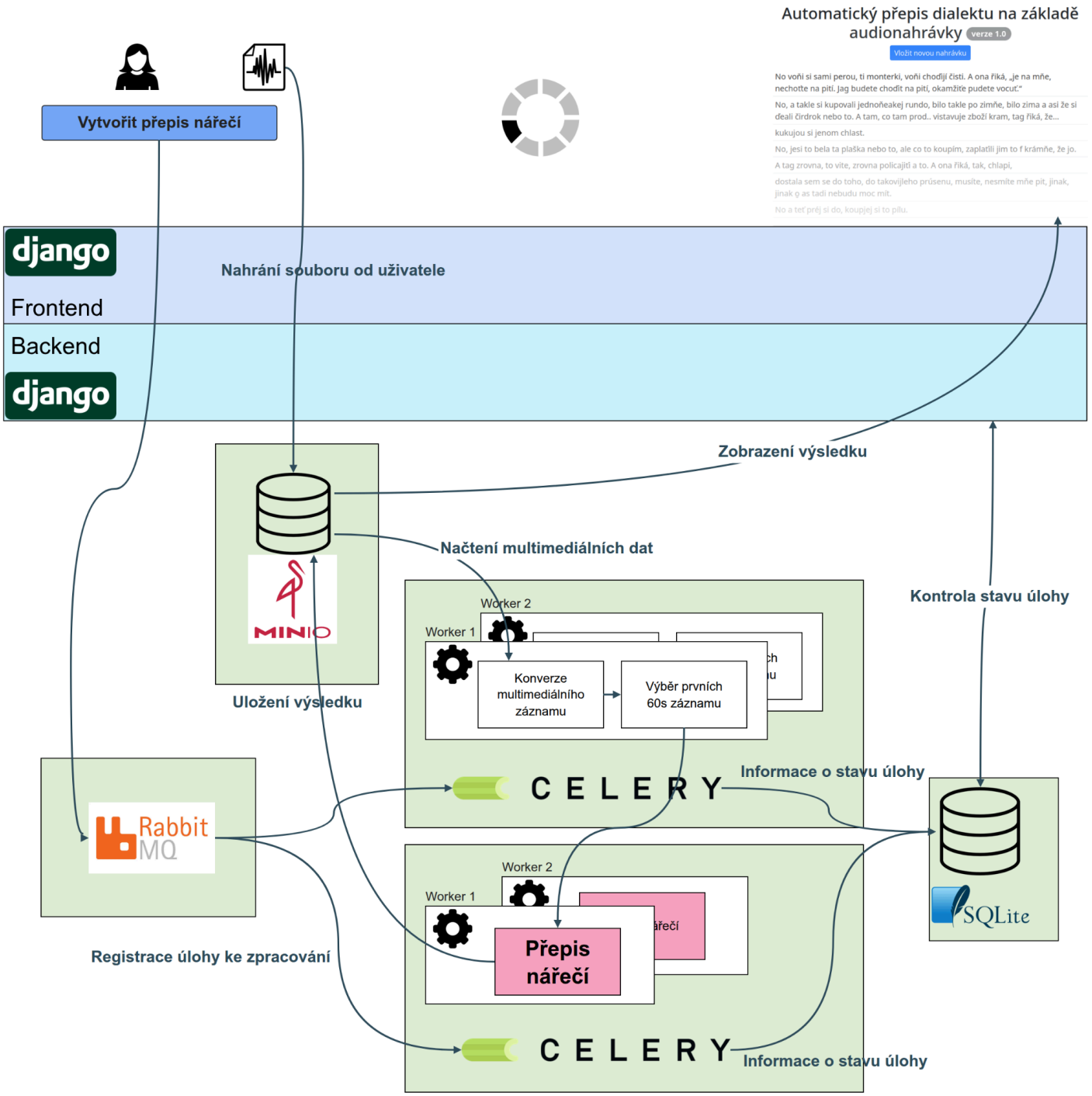
Demonstrátor se skládá z následujících komponent:

- Minio = datové úložiště
- RabbitMQ = obsluha fronty zpracovávaných úloh
- Celery = distribuovaný systém zpracování úloh
- Django = backend pro Celery; webový frontend

Výše uvedené komponenty běží jako služby zapouzdřené do Docker kontejnerů na dedikovaném serveru na Fakultě informačních technologií Vysokého učení technického v Brně. Demonstrátor je přístupný z adresy https://jamap.cz/prepis. Po vložení nahrávky uživatelem a stisku tlačítka Vytvořit přepis nářečí (viz obr. 1) je tato nahrávka odeslána a uložena do Minio úložiště. Dále je webovým frontendem vytvořena úloha Přepis nářečí z dané nahrávky. Uživateli se zobrazí rotující ikona, která znázorňuje, že se úloha zpracovává. Frontend se průběžně dotazuje backendu na výsledek zpracování.
Úloha je přes frontu RabbitMQ poslána ke zpracování. Pokud jsou k dispozici volné prostředky pro zpracování (worker), tak Celery přijme úlohu a je zahájeno zpracování.
V současné době je možné zpracovávat maximálně 2 úlohy současně. Pokud uživatelé vloží v krátkém sledu více úloh, musí počkat, než se dostanou na řadu. Vzhledem k tomu, že je úloha automatického přepisu náročná na výpočetní zdroje, byla zapouzdřena do samostatného kontejneru. Tím lze tuto úlohu pouštět na fyzicky jiném hardware.
Úloha je rozdělena do 3 po sobě jdoucích kroků:

- Konverze vstupních dat na požadovaný formát `PCM mono s16le 16kHz`. Vstupem tedy může být „libovolný“ multimediální soubor včetně videozáznamu.
- Extrakce prvních 60 vteřin z nahrávky. K tomuto kroku se přistupuje z důvodu omezení paměťových nároků demonstrátoru rozpoznávače a také zkrácení čekací doby na výstup.
- Automatický přepis nářečí z 60vteřinové nahrávky a uložení výsledku do databáze backendu včetně indikace úspěšného dokončení úlohy.

V momentě, kdy frontend detekuje úspěšné dokončení úlohy, zobrazí výsledky uživateli (viz obr. 2).
Upozornění: úloha automatického přepisu nářečí vždy před zpracováním zvukové nahrávky načítá modely. Tento krok trvá cca 15 vteřin. Velmi krátká nahrávka se tedy zpracovává cca 20 vteřin, minutová pak cca 40 vteřin (pokud je prázdná fronta).



Obr. 4: Zjednodušené schéma demonstrátoru přepisu nářečí. Komponenta automatického přepisu nářečí je zvýrazněna červeně. Zelené boxy indikují použití Docker kontejneru

Literatura

ŠIMEČKOVÁ, M. – STUPŇÁNEK, B. – KARAFIÁT, M. – VOŽENÍLEK, V. – VONDRÁKOVÁ, A. – NÉTEK, R. (2025). *Metodika pro převod strukturovaných znalostí z oboru dialektologie do strojového učení*. Brno – Olomouc: Ústav pro jazyk český Akademie věd České republiky – Vysoké učení technické v Brně – Univerzita Palackého v Olomouci, 2025. Dostupné z: https://www.jamap.cz/metodika.

ŠIMEČKOVÁ, M. (2024). *Archiv zvukových záznamů nářečních promluv*. Praha: Academia. Též online. Dostupné z: http://www.vedakolemnas.cz/sys/galerie-download/VKN-132_.pdf.

RADFORD, A. et al. (2022)**.** *Robust Speech Recognition via Large-Scale Weak Supervision.* arXiv: 2212.04356 [cs.CL]. [online]. Dostupné z: https://arxiv.org/abs/2212.04356.