

VÝZVA K PŘEDLOŽENÍ INFORMACÍ V RÁMCI PRŮZKUMU TRHU

Komplexní řešení pro poskytování plně automatizovaných i lidsky asistovaných informačních služeb typu chatbot, livechat a voicebot (zajištění kontinuity služeb, rozšíření portfolia a nové funkcionality).

(dále jen „zakázka“)

Název:	Česká republika – Český statistický úřad (ČSÚ)
Sídlo:	Na padesátém 81, 100 82 Praha 10 – Strašnice
Identifikační číslo:	000 25 593
ID datové schránky	2gfaasy
Internetová adresa profilu zadavatele:	https://nen.nipez.cz/profil/CSU
Osoba oprávněná jednat jménem či za zadavatele:	Ing. Petra Kuncová, e-mail: petra.kuncova@czso.cz , tel.: 274053104

1 FUNKČNÍ POŽADAVKY ZADAVATELE NA PŘEDMĚT PLNĚNÍ

1.1 CHARAKTERISTIKA STÁVAJÍCÍHO ŘEŠENÍ

Český statistický úřad v současné době využívá systém vyvinutý primárně pro potřeby projektu Sčítání lidu, domů a bytů 2021, který slouží k automatickému zodpovídání dotazů veřejnosti týkajících se činnosti a působnosti úřadu (chatbot běží na adrese www.czso.cz), resp. problematiky sčítání (na samostatném webu www.scitani.cz).

Komunikace s lidským operátorem livechatu byla (v podobě tlačítka „přepojit na operátora“) k dispozici pouze v době aktivního sběru dat a nyní se nevyužívá.

Architektura chatbota je postavena na automatickém rozpoznávání významu (NLU / Natural Language Understanding) položených dotazů psaných volným textem (s tlačítkovou nabídkou výchozích ani vyhledaných záměrů/„intentů“ se nepracuje). Průběžně rozšiřovaná znalostní báze má podobu otázka – odpověď. Znění konkrétní otázky se pak zobrazuje jen v případě nejednoznačného vstupu, a to v podobě seznamu 2–5 návrhů (u jedné vybrané odpovědi se vrací pouze její text).

Význam není vyhodnocován na základě fulltextu ani využití obecných (předtrénovaných) NLP modelů (např. slovních vektorů nebo také „embeddingů“), ale pomocí shody manuálně definovaných klíčových slov, jejich kombinací a ustálených spojení (mandatorního kontextu). Na 1 otázku a odpověď (při znalostní bázi o velikosti cca 400 záznamů) připadá v průměru 3,5 klíčových slov, resp. 5,5 ustálených spojení (a to včetně synonym a parafrází).

Celý systém (vč. aplikačních a DB serverů) je nyní provozován v infrastruktuře ČSÚ v režimu on-premise. Podle dat za první čtvrtletí 2022 představoval průměrný měsíční provoz chatbota celkem 2122 dotazů v rámci 1001 konverzací (uživateli). Poměr zodpovězené vs. nezodpovězené dotazy pak činil 88,4 vs. 11,6 %.

1.2 PŘEDMĚT PLNĚNÍ

Životní cyklus systému chatbota, který dnes ČSÚ využívá, se během roku 2023 naplní. Pro pokračování a rozšíření stávajících služeb zadavatel plánuje vypsát veřejnou zakázku. Cílem průzkumu trhu je proto zjistit, zda a za jakých podmínek je tento záměr realizovatelný.

Předmětem plnění zakázky je návrh komplexního řešení podle níže uvedených požadavků, z nichž část I, II a III je povinná pro všechny uchazeče, část IV pouze pro uchazeče nabízející hlasové technologie, přičemž i systém zcela bez funkcionalit voicebota vyhoví hlavnímu účelu použití.

I. Obecné požadavky

- a) plně automatizovaný robustní systém schopný obsloužit (při zpracování odpovědi max. v nižších jednotkách vteřin) řádově stovky uživatelů denně
- b) provoz stabilní cloudové platformy / on-premise varianty s plnou technickou podporou a možností samostatné uživatelské správy dat (vč. řízení aktualizací)
- c) realizace paralelního produkčního a testovací prostředí (pro ad hoc i dávkové testy)
- d) (ICT kapacity ČSÚ nezatěžující) integrace výchozího komunikačního rozhraní do webového portálu zadavatele (platforma Liferay, měsíční provoz: 200 tis. uživatelů, 886 tis. zobrazení, 343 tis. návštěv)
- e) znalostní podklady (statická, průběžně aktualizovaná báze otázek a textových odpovědí) by měly mít možnost do odpovědí dynamicky načítat aktuálně platnou hodnotu (např. míru inflace) z externího zdroje (veřejně vystavených dat např. ve formátu JSON)
- f) veškerá provozní data škálovatelně analyzovatelná jak pro účely pravidelného reportingu (počty uživatelů/konverzací, dotazů, nejčastěji žádaných otázek/odpovědí apod.), tak pro zlepšování správnosti odpovědí (přehledy nezodpovězených dotazů, historie konverzací ad.)
- g) archivace (příp. anonymizace) získaných dat odpovídající standardům GDPR (současný chatbot žádné osobní údaje nepožaduje ani nesbírá, session trvá do obnovení stránky v prohlížeči)
- h) webová rozhraní sloužící interakci s uživatelem respektující principy vstřícného UX/UI designu a vyhovující zákonným požadavkům na přístupnost
- i) řešení by mělo umožňovat rozvojové požadavky (uzpůsobení podle aktuálních potřeb); předpokládá se např. měsíční alokování/nacenení určeného objemu práce

II. Chatbot

- a) příprava nových dat a práce s administračním uživatelským rozhraním nesmí vyžadovat zvláštní odborné znalosti nebo dovednosti (např. na úrovni programátora nebo počítačového lingvisty)
- b) systém by měl být schopen (v přiměřeném počtu) předkládat návrhy odpovědí pro nejednoznačné nebo obecně zadané dotazy
- c) při aktualizaci dat (přidávání nových témat / otázek a odpovědí) je odpovědností systému hlídat významový překryv témat a/nebo klíčových slov či spojení (nad rámec zobrazitelných variant odpovědí)
- d) parafráze, synonyma a varianty dotazů by měly být zpracovávány v rovině obecných NLP nástrojů a technik pro češtinu, příp. angličtinu (např. předpočítaných modelů), tedy ne zadávány pouze manuálně a prostým výčtem alternujících slov/forem nebo spojení)

III. Livechat

- a) služba fungující jako komunikační „odbočka“ chatbota, nikoli samostatně (tlačítko pro zahájení konverzace s operátorem má být k dispozici až po několika interakcích s chatbotem)
- a) komunikační rozhraní disponující funkcionalitami řízení fronty požadavků (možnost přepojení na operátora jen v případě dostupnosti a dostatečné kapacity lidských sil)
- c) po ukončení konverzace dobrovolná nabídka vyjádření spokojenosti (např. tlačítkových emotikonů)

IV. Voicebot

- a) ČSÚ přijal v roce 2021 až 46 % požadavků telefonicky, na uchazeči je navrhnout možnosti uplatnění této technologie (v kontextu popisu uvedeného pod 1.1 a požadavků 1.2 I až III)
- b) představitelný je např. hlasový klon celého chatbota (přepis diktovaných dotazů a čtení odpovědí)
- c) nabízí se i podpora pouze dílčích hlasových technologií (TTS, STT) při komunikaci pomocí telefonní linky (např. jako pomoc uživatelům se specifickými potřebami)

2 DOTAZY

Otázka č. 1: Je možné (podle požadavků popsanych v části 1.2 I–IV) určit nabídkovou cenu, která obsáhne veškeré potřebné práce a činnosti, a pokud ano, v jaké výši (s rozlišením bez DPH a vč. DPH), resp. s jakým algoritmem jejího výpočtu?

Otázka č. 2: Dá se na základě poskytnutých informací sestavit předpokládaný harmonogram plnění, jenž bude reflektovat jak aspekty výstavby nového systému, tak potřeby migrace stávajících dat a zachování kontinuity poskytovaných služeb, díky němuž budeme schopni správně načasovat začátek plnění plánované zakázky?

Otázka č. 3: Jaké postupy, nástroje, techniky a technologie jste schopni – se zdůvodněním jejich přínosů i rizik při jejich nevyužití – pro řešení zadání nabídnout?

Otázka č. 4: Můžete stručně popsat své zkušenosti s realizací zakázek s obdobným předmětem plnění a u jednotlivých referencí kromě období realizace a objemu zakázky uvést také kontakty na osoby, u nichž bude možné uváděné skutečnosti ověřit?

3 LHŮTA PRO DORUČENÍ ODPOVĚDÍ

Odpovědi uchazečů budou doručeny v elektronické podobě a v českém jazyce do datové schránky ČSÚ (ID: 2gfaasy) se současným informováním kontaktní osoby zadavatele e-mailem (stepanka.zelenkova@czso.cz).

Lhůta pro doručení odpovědí uchazečů zadavateli: **pondělí 30. 5. 2022, 15.00.**

4 ZÁVĚREČNÁ USTANOVENÍ

Zadavatel nebude vracet uchazečům odpovědi na tento průzkum trhu ani jejich části. Žádný z uchazečů nemá právo na úhradu nákladů, které vynaložil v souvislosti s tímto průzkumem.

Otevírání odpovědí se uskuteční bez přítomnosti uchazečů po uplynutí lhůty pro podání odpovědí.

Zadavatel si vyhrazuje právo ověřit údaje uvedené v odpovědích uchazečů, případně si vyžádat dodatečné informace nebo další podklady.