

Návrh řešení vývoje, implementace a ověřování funkčnosti

Vývoj

Systém bude mít REST API rozhraní realizované webovým serverem nginx a serverem php-fpm dostupným z něj přes FastCGI rozhraní. Php-fpm bude interně volat další subsystemy, které budou realizovat vlastní překlad a další podpůrné činnosti. Komunikace mezi webovým serverem a klienty bude prováděna zabezpečeným protokolem a bude vyžadována autentifikace uživatele.

Plánované funkcionality API

Překlad Č→N a N→Č

Překladače budou rozlišovat:

- medicínsky klíčové texty popisující stav pacienta, které mají podobu heslovitého zápisu odděleného čárkami, a které se budou překládat spolehlivě s využitím databáze překladů předdefinovaných sousloví, nebo v případě neexistence spolehlivého překladu podle odborného nebo obecného slovníku s možností desambiguace významu v původním jazyce uživatelem, a v krajním případě pomocí adaptovaného neuronového překladače;
- obecné texty skládající se z běžných vět, které budou většinou překládány adaptovaným neuronovým překladačem a ve frekventovaných případech budou použity ověřené překlady nebo překlady vytvořené pomocí pravidel a databáze překladů předdefinovaných sousloví.

Na základě rozlišení typu textu, využitých překladových technik a zarovnání mezi zdrojovým textem a překladem pak budou části textu označeny mírou spolehlivosti překladu a bude v nich vyznačena vzájemná korespondence jednotlivých částí, která bude následně využitelná při uživatelské kontrole podle slovníku.

Učení se překladače ze zpětné vazby

Překlady uživatelem potvrzené jako správné se budou zapamatovávat a následně znovu využívat. K přijetí uživatelské zpětné vazby bude k dispozici API volání.

Kontrola a opravy překládaného textu

API bude umožňovat kontrolu a vylepšování zdrojového textu už před vlastním překladem.

Včasnou opravou zdrojového textu se tak zabrání překladům nízké kvality. Konkrétně půjde o:

- Korekturu překlepů – Zdrojový text bude zkontrolován, jestli neobsahuje neznámá slova (překlepy). Pokud budou neznámá slova nalezena, pokusí se navrhnout jejich opravu na slova známá.
- Expanzi zkratk – Dostatečně jednoznačné a známé zkratky budou expandovány na jednotnou reprezentaci, která se pak stane vstupem překladu.
- Doplnění diakritiky – Do textu bez diakritiky bude na uživatelskou žádost automaticky doplněna diakritika.

- Zobrazení slovníkových hesel – K zadanému slovu nebo slovu s kontextem API najde a vrátí relevantní položky ze slovníku ve formátu vhodném k zobrazení v HTML prohlížeči.

Podpora pro prediktivní psaní

API umožní odhad pokračování psaní textu na základě prefixu a kontextu. Bude se snažit navrhnout pokračování v délce jednoho slova v případech, kdy toto pokračování bude dostatečně pravděpodobné. V případě několika pravděpodobných pokračování se bude navrhopvat několik nejpravděpodobnějších z nich. Tato funkcionality se bude průběžně učit od všech uživatelů, nicméně bude upřednostňovat možnosti charakteristické pro aktuálního uživatele.

Údržba databáze překladů předdefinovaných sousloví

API bude poskytovat rozhraní k údržbě databáze překladů předdefinovaných sousloví. Bude možné zde doplňovat nová předdefinovaná sousloví a jejich překlady v závislosti na kontextu a odstraňovat nebo měnit nevhodné překlady. Podobné API bude k dispozici i pro úpravy překladových slovníků pro kontrolu a pro učení orientačního překladače a pro úpravy databází zkratk a synonym.

Databáze překladů předdefinovaných sousloví

Databáze překladů předdefinovaných sousloví vznikne analýzou textů poskytnutých Zadavatelem, extrakcí frekventovaných sousloví z ní, a jejich překladem s využitím odborného, případně i obecného slovníku. Tyto překlady budou následně zkontrolovány člověkem.

K dispozici bude webové uživatelské rozhraní, ve kterém bude možné tuto databázi upravovat přidáváním předdefinovaných sousloví a jejich překladů pro jednotlivé kontexty, případně mazáním nebo měněním předdefinovaných sousloví a jejich překladů, nebo úpravou omezení kontextů. Předpokládaná iniciální velikost této databáze by měly být jednotky tisíc překladů, postupnou akumulací reálných dat v obou jazycích pak může v průběhu projektu dále růst tak, aby přírůstky odrážely skutečnou potřebu. Schopnost překládat více než 10000 medicínských termínů, kolokací či frází v každém směru bude již od počátku zajištěna integrací odborného slovníku výrazně většího rozsahu v obou směrech. Tento slovník bude také počátečním zdrojem předdefinovaných sousloví a jejich překladů. Konkrétně půjde o Německý lékařský slovník, který obsahuje 25000 hesel (jednoslovných i víceslovných) ve směru z češtiny do němčiny a 25000 hesel ve směru z němčiny do češtiny. Jako obecný slovník bude využit Německý slovník Platinum, který obsahuje 102 tisíc hesel ve směru z češtiny do němčiny a 139 tisíc hesel ve směru z němčiny do češtiny.

Dokumentace

Dokumentace API bude dostupná v českém, německém a anglickém jazyce.

Spolehlivost

Systém bude řešen jako redundantní, fungující na více serverech, které budou schopny se v případě výpadku některého serveru vzájemně zastoupit. Klienti budou znát několik přístupových bodů, v případě nedostupnosti nebo nefunkčnosti některého z nich se obrátí na další v pořadí. Databáze jednotlivých serverů se budou automaticky synchronizovat a zálohovat.

Integrace do systému Babylon

Pro integraci do systému Babylon bude poskytnut přístup k funkčnímu API, jeho dokumentace a

podle potřeby budou poskytovány mailové, telefonické nebo on-line konzultace. V případě potřeby úprav pak mohou být prováděny i úpravy nemající fundamentální vliv na architekturu systému. (Ta bude dohodnuta ve fázi analýzy.)

Ověřování funkčnosti

Funkčnost bude ověřována již od začátku přípravy systému automatickými testy dosud implementované funkcionality. Po integraci předpokládáme uživatelské testy v aplikaci Babylon a praktické ověření funkčnosti v rámci cvičení. Před nasazováním dalších verzí budou provedeny automatické testy a případné rozdíly v jejich výsledcích budou hodnoceny člověkem. Buďto systematicky, nebo alespoň namátkově v případě velkého množství rozdílů.

Pro případ nasazení verze horší, než byla předchozí, bude existovat možnost návratu k původní verzi.

Podpora a další vývoj

Dodavatel se zavazuje, že do tří měsíců po skončení projektu uzavře smlouvu o technické podpoře a zkvalitňování aplikace, týkající se minimálně:

- zkvalitňování správnosti (přesnosti a spolehlivosti) překladu, hlavně v kontextu;
- úprava a doplňování slovní zásoby ve vztahu k potřebám ZZS obou stran;
- dalších inovací, pokud povedou ke zkvalitňování překladače jako celku – tím je kromě výše zmíněného myšlena rychlost překladu, stabilita překladů a lepší implementovatelnost překladače jako takového;
- průběžnou údržbu, upgrady a update systému po SW i HW stránce u dodavatele tak, aby byla zajištěna funkčnost překladače a aby splňoval aktuální technické a bezpečnostní standardy.

Konkrétní podmínky poskytování technické podpory a zkvalitňování aplikace budou dohodnuty po skončení projektu.

Předmět veřejné zakázky

Předmětem veřejné zakázky je vývoj a implementace softwarového online česko-německého, německo-českého specializovaného překladače v oblasti zdravotnické terminologie používané při poskytování přednemocniční neodkladné péče s parametrem směru překladu. Jde o vytvoření překladače, který bude sloužit pro komunikaci mezi jednotlivými dispečinkami, personálem ZZS, ale také mezi záchranáři a pacientem. Tento překladač bude moci být využíván přímo na místě zásahu (online), ale také bude implementován do systému alarmování. Tento překladač bude disponovat tzv. "učícím se SW" a bude se sám učit na základě dennodenního využívání a to díky zpětné vazbě od personálu ZZS.

Hlavní činnosti:

- 1) analýza situace
 - dodavatel provede analýzu stávající situace (analýza zadavatelem poskytnutých dat a požadavků zadavatele) a za spolupráce se ZZSPK upraví vlastní návrh řešení, což bude odsouhlaseno oběma stranami
 - odsouhlasený návrh řešení bude stěžejní pro další činnosti
- 2) vývoj/úpravy aplikace
 - následně dodavatel zahájí vývoj aplikace (příp. úpravy stávající aplikace) za dohledu a konzultace se ZZSPK
 - aplikace bude vyvinuta do takové podoby, aby mohlo být přikročeno k bodu 3)
- 3) implementace do komunikačního systému Babylon, ladění a pilotní testování
 - dodavatel poskytne objednateli popis API a přístupové údaje tak, aby byl objednatel schopen implementovat aplikaci do komunikačního systému Babylon
 - po provedení implementace bude aplikace nasazena k pilotnímu testování
 - během pilotního testování bude průběžně docházet k ladění aplikace, opravě chyb a vylepšování funkcionality
- 4) předání k užívání a k využití na společných cvičeních
 - dokončená verze aplikace dle bodu 3) bude předána ke zkušebnímu provozu
 - aplikace bude nasazena na vybraných cvičeních ZZSPK, kde bude testována v „ostrých“ podmínkách
 - cvičení se bude účastnit zástupce dodavatele, který bude zaznamenávat hlášené chyby a problémy
- 5) finální odladění a akceptace díla
 - na základě výstupů z bodu 4) bude aplikace upravena a finalizována
 - finální aplikace bude předána vč. dokumentace na základě akceptačního řízení
- 6) proškolení
 - dodavatel proškolí zaměstnance (min. 2) zadavatele v otázkách stěžejních funkcionalit aplikace
 - proškolení se uskuteční v sídle zadavatele
 - součástí zaškolení bude předání podrobného popisu rozhraní aplikace

Pro každou hlavní činnost bude zřízena pracovní skupina ze zástupců dodavatele i zadavatele, která bude pravidelně (obvykle 1x měsíčně nebo dle potřeby častěji) pořádat online nebo presenční meetingy a komunikovat a koordinovat postup činností. Z každé schůzky bude dodavatelem vyhotoven zápis, kde budou zaznamenány dohodnuté postupy a odsouhlasené úpravy.

Ukončení každé hlavní činnosti bude protokolárně odsouhlaseno oběma stranami.

Technické požadavky

- Požadujeme online překladač Č/N, N/Č přes REST/API rozhraní v oblasti zdravotnické sféry a poskytování přednemocniční neodkladné péče (PNP) s parametrem směru překladu.
- Překladač musí být schopen překládat alespoň deset tisíc medicínských termínů, případně kolokací a frází v každém směru.
- Překladač poběží na serverech dodavatelské firmy na zabezpečeném protokolu dle požadavku ZZSPK (např. HTTPS) a bude vyžadovat autentifikaci uživatele.
- Řešení musí být schopno spolu s překladem vracet i jeho spolehlivost. (Rozlišovat, jestli jde o spolehlivou část překladu textu, nebo o část, která může sloužit jen pro orientaci a hrozí v ní chyby.)
- Řešení musí obsahovat co nejlepší mechanismy pro detekci úseků potenciálně problematických pro překlad ve vstupním textu. (Detekce překlepů a neznámých slov, případně jiných chyb a nekonzistencí.)
- Systém musí být schopen detekovat v textech "**předdefinovaná sousloví**" a překládat je přesně zadaným způsobem. Takto vytvořené části překladu budou pokládány za spolehlivé.
 - Množina "předdefinovaných sousloví" a jejich překladů musí být editovatelná na straně zadavatele (odebírat, přidávat i měnit). Pevné překlady "předdefinovaných sousloví" mohou být omezeny na specifické kontexty, ve kterých jsou skutečně spolehlivé.
 - Základní množina "předdefinovaných sousloví" bude vytvořena v průběhu vývoje aplikace v rámci řešení projektu a dodavatel musí v nabídce navrhnout způsob vytvoření této množiny a odhadnout její velikost.
- Dokumentace v českém, německém i anglickém jazyce.
- Systém musí být navržen a řešen tak, aby byl funkční i při výpadku jednoho ze serverů, na nichž bude provozován, nebo aby bylo zaručeno, že výpadek nebude delší než patnáct minut.

č.	Funkcionalita	Jazyk	Max. bodů	Nabízeno (ANO/NE)
1	korektor překlepů ve zdrojovém textu	Č, N	celkem 10	---
1a	detekce neznámých slov	Č	3	ANO
		N	3	ANO
1b	návrhy oprav	Č, N	4	ANO
2	prediktivní psaní natrénované na základě textů dodaných Zadavatelem (hodnocen každý jazyk zvlášť)	Č, N	celkem 24	---
2a	schopnost navrhnout (dokončit) jedno slovo	Č	3	ANO
		N	3	ANO
2b	návrh sekvence více slov	Č	1	NE
		N	1	NE
2c	alternativní návrhy v případě nejistoty	Č	2	ANO
		N	2	ANO
2d	učení se během užívání od všech uživatelů	Č	3	ANO
		N	3	ANO
2e	učení se během užívání od aktuálního uživatele (pokud byly přiznány body za předchozí položku, musí být k přiznání bodů za tuto položku vstupy od aktuálního uživatele chápány jako důležitější nebo mít přednost před vstupy od ostatních):	Č	3	ANO
		N	3	ANO
3	detekce složitosti překladu (hodnoceno pro každý jazyk zvlášť) – označení částí zdrojového textu, kde je překlad	Č, N	celkem 12	---
3a	spolehlivý (takto nesmí být označena část, kde je možnost, že dojde k chybě)	Č	2	ANO
		N	2	ANO
3b	nejednoznačný (ale upřesnitelný uživatelem)	Č	2	ANO
		N	2	ANO
3c	neznámý (předpokládá se špatný překlad)	Č	1	NE
		N	1	NE
3d	nejistý = pouze orientační (předpokládá se běžná kvalita neuronového strojového překladu)	Č	1	ANO
		N	1	ANO
4	uživatelem asistovaná desambiguace vstupu (výběr správného významu víceznačného slova nebo sousloví ve zdrojovém textu uživatelem) následně využitá při překladu nebo jeho kontrole	Č	10	ANO
		N	10	ANO
5	asistované přidávání slov do slovníku korektoru překlepů včetně všech tvarů – uživatel koriguje návrhy, ale nemusí ručně zadávat všechny tvary slova, systém ale všechny tvary navrhne ke kontrole (případně úpravě) a po schválení je bude znát	Č	3	NE
		N	3	NE
6	podpora pro zobrazení zarovnaní zdrojového textu a překladu (která slova jsou překladem kterých slov)	Č, N	celkem 26	---
6a	podpora dle bodu 6, pokud je integrován i překladový slovník, kde může uživatel kontrolovat	Č -> N	10	ANO
		N -> Č	10	ANO
6b	podpora dle bodu 6, pokud slovník integrován není	Č -> N	3	NE
		N -> Č	3	NE
7	možnost přidávání slov do překladového slovníku pro orientační překlad	Č, N	celkem 15	---
7a	přidání slova a jeho překladu do druhého jazyka	Č, N	10	ANO
7b	přidání slova a jeho synonyma (například zkratky a jejího plného tvaru) ve stejném jazyce	Č, N	5	ANO
8	automatické expandování jednoznačných zkratk ve zdrojovém textu	Č	4	ANO
		N	4	ANO
9	automatická kontrola orientačního překladu podle slovníku s vyznačením částí nevyskytujících se na druhé straně	Č -> N	10	ANO
		N -> Č	10	ANO

č.	Funkcionalita	Jazyk	Max. bodů	Nabízeno (ANO/NE)
10	automatická kontrola zachování struktury věty v orientačním překladu	Č -> N	10	NE
		N -> Č	10	NE
11	možnost automatického doplnění diakritiky do zdrojového textu	Č	4	ANO
		N	4	ANO
12	přístup ke slovníku pro ověření překladu uživatelem	Č, N	celkem 50	---
12a	<i>odborná část</i>	Č -> N	celkem 15	---
		N -> Č	celkem 15	---
12a1	více než 10.000 odborných hesel v základním tvaru	Č -> N	5	ANO
		N -> Č	5	ANO
12a2	více než 20.000 odborných hesel v základním tvaru	Č -> N	5	ANO
		N -> Č	5	ANO
12a3	více než 30.000 odborných hesel v základním tvaru	Č -> N	5	NE
		N -> Č	5	NE
12b	<i>obecná část</i>	Č -> N	celkem 10	---
		N -> Č	celkem 10	---
12b1	více než 20.000 obecných hesel v základním tvaru	Č -> N	3	ANO
		N -> Č	3	ANO
12b2	více než 50.000 obecných hesel v základním tvaru	Č -> N	3	ANO
		N -> Č	3	ANO
12b3	více než 100.000 obecných hesel v základním tvaru	Č -> N	4	ANO
		N -> Č	4	ANO
13	možnost přidávání položek do slovníku pro kontrolu uživatelem	Č, N	3	ANO
14	možnost odebírání nevyhovujících překladů ze slovníku pro kontrolu uživatelem	Č, N	2	ANO
CELKEM			224	