



Skupni zapisnik strokovnega dialoga –Virtualni pomočnik na VDT ki je potekal 16. in 17. 5. 2024, na lokaciji Vrhovnega državnega tožilstva RS

Zakon o javnem naročanju (ZJN-3, Ur.l. RS št. 91/15, v nadaljevanju ZJN-3) v 64. členu omogoča naročniku, da pred začetkom postopka javnega naročanja z gospodarskimi subjekti izvede strokovni dialog in v okviru tega zaprosi ali upošteva nasvete, ki jih bo lahko uporabil pri pripravi razpisne dokumentacije, pod pogojem, da taki nasveti oziroma priporočila ne preprečujejo ali omejujejo konkurence, ter ne pomenijo kršenja načela enakopravne obravnave ponudnikov in načela transparentnosti javnega naročanja.

Ministrstvo za pravosodje je kot naročnik 9. 5. 2024 na svoji spletni strani objavilo povabilo zainteresiranim gospodarskim subjektom za sodelovanje pri strokovnem dialogu zvezi s pripravo javnega naročila za razvoj informacijskega sistema Virtualni pomočnik.

Interes za sodelovanje v strokovnem dialogu je izkazalo sedem subjektov, eden je naknadno odstopil od sodelovanja, in sicer:

- Margis d.o.o., Brnčičeva ul. 41G, 1000 Ljubljana,
- Telekom d.d., Cigaletova ulica 15, Ljubljana, 1000 Ljubljana,
- SRC d.o.o., Tržaška cesta 116, 1001 Ljubljana,
- Pareto AI rešitve d.o.o., Videmska ulica 3, 1000 Ljubljana,
- Comitto d.o.o., Ulica Franca Kostanjevca 8, 2204 Miklavž na Dravskem polju, - Comland d.o.o., Litostrojska cesta 58C, 1000 Ljubljana,
- Vitasis d.o.o., Partizanska cesta 8A, 1381 Rakek – ODSTOPIL.

Naročnik se je po pregledu prejetih prijav odločil, skupaj z uporabnikom, da z vsemi zainteresiranimi gospodarskimi subjekti izvede individualne sestanke (varovanje poslovne skrivnosti) v dolžini cca. 1h, na katerih bodo gospodarski subjekti predstavili svoje poglede ter odgovarjali na naročnikova oz. uporabnikova vprašanja. Strokovni dialog se je izvedel v času od 16. do 17. maja 2024.

Individualni sestanki so se izvedli po sledečem urniku:

- četrtek, 16. 5. ob 9:00 uri, MARGIS d.o.o.
- četrtek, 16. 5. ob 11:00 uri SRC d.o.o.
- četrtek, 16. 5. ob 13:30 uri Pareto AI rešitve
- petek, 17. 5. ob 9:00 uri Telekom d.o.o.
- petek, 17. 5. ob 11:00 uri Comitto d.o.o.
- petek, 17. 5. ob 13:30 uri Comland d.o.o.

Dnevni red sestankov:

- 1) Kratka predstavitev vsebine projekta „Virtualni pomočnik“
- 2) Vprašanja in odgovori (Q&A) – razprava

V nadaljevanju podajamo povzetek razprave in zapis izpostavljenih ugotovitev v sumarni obliki. Naročnik je vsakokrat na kratko predstavil vsebino projekta in vsem udeležencem postavil ista vprašanja. Ob



vsakem vprašanju se je razvila tudi krajša razprava, kjer so udeleženci pojasnili svoje vidike in poznavanja, lahko pa so postavili tudi vprašanja z namenom boljšega razumevanja vsebine.

Vsem zainteresiranim so bila predstavljena spodaj navedena vprašanja z obrazložitvijo:

1. Državno tožilstvo ima v svojem informacijskem sistemu oziroma na datotečnem strežniku shranjene občutljive oziroma posebne osebne podatke. Neupravičeno razkritje le teh bi lahko imelo vpliv na potek postopka, ki je še v teku. Zato bo razvijalec svoje delo opravljal na lokaciji Vrhovnega državnega tožilstva, kar je praksa tudi v tujini. Torej zaposleni izvajalca bo sedel na VDT, strojna oprema, ki jo izvajalec potrebuje, ter osebni računalnik zaposlenega bosta vključena v mrežo VDT. Prav tako bo na lokaciji VDT nameščen strežnik oziroma, na katerem bo potekal razvoj, testiranje in uporaba razvitega orodja. Dostop do delovnega prostora na VDT bo zaposleni pri izvajalcu imel vsak delovni dan od 7h do 17h. Delovna oprema, ki jo bo izvajalec prinesel s sabo, in strežnik ne bosta smela zapustiti delovnega prostora. Kaj bi izvajalec potreboval, da bo zaposleni, ki pripravlja bazo za učenje in razvija programsko rešitev, lahko čim boljše opravljal svoje delo na lokaciji VDT? Kakšni morajo biti tehnični pogoji za delovanje strežnika oziroma strežnikov na naši lokaciji? Kakšna bi bila namestitev strežnikov (možnost virtualizacije, zahtevana procesorska moč, zahtevane grafične kartice, zahtevan RAM, zahtevan diskovni prostor, potrebne povezave – LAN, proxy, baze podatkov). Kakšna je vaša ocena glede potrebnih virov, kapacitet potrebnih strežnikov, potrebnih licenc ter finančnih stroškov za vzpostavitev končnega delujočega strežniškega okolja za namen projekta Virtualni pomočnik?
2. Glede na to, da smo kot organizacija, znotraj širše državne uprave, zavezani slediti strategiji digitalizacije, želimo uporabljati odprtokodne rešitve. Zato bi želeli, da se model razvije na odprtokodni rešitvi, ki se bo izkazala za najbolj uporabno v slovenskem jeziku in za naše potrebe. V razvoju pa je slovenski veliki jezikovni model. Zanima nas, kako bi potekala morebitna prilagoditev modela na slovenski veliki jezikovni model, ko bo ta dokončan in na razpolago. Koliko časa bi potrebovali za učenje modela na novem velikem jezikovnem modelu? Kaj bi za to potrebovali?
3. Na tožilstvu imamo meta podatke in dokumente hranjene v dveh informacijskih sistemih. Dokumenti v dokumentnem sistemu, ostali podatki so hranjeni v SQL bazi. Povezava med njima je enolična identifikacijska številka. Ali lahko zelen program razvijemo na način, da ne uporabljamo meta podatkov iz informacijskega sistema državnega tožilstva, torej le na osnovi dokumentov? Ali je to mogoče z vključitvijo semantičnega modela oziroma na kak način? Ali je nujno potrebno, da uporabimo tudi meta podatke iz informacijskega sistema? Posamezen modul namreč ne bo smel uporabljati vseh podatkov oziroma dokumentov, ampak bo njegovo delovanje moralo biti omejeno na le določen sklop podatkov (zaradi pravic dostopa do dokumentov).
4. Dokumenti se v IS tožilstva preslikajo in naredi se OCR. Kakšna je priporočena resolucija skeniranja in kakšna barvna globina, da je OCR najbolj optimalen? Kakšni so minimalni pogoji, da je OCR sprejemljiv?
5. Kako se pripravijo podatki za učenje (ročno/strojno)? Ali je potrebna kakšna selekcija dokumentov oziroma podatkov pri učenju? Koliko sodelovanja z uporabnikom je pri tem potrebno?
6. Ker želimo omogočiti tudi dostope do dokumentov v izvorni obliki, nas zanima ali lahko UI rekonstruira izvorno obliko dokumenta (ki je 100% enaka)? Želeli bi se izogniti kopiranju izvornih dokumentov. Torej ali si v procesu indeksacije zapomni tudi izvorno obliko dokumenta?



7. Kako se umetna inteligenca samostojno uči oziroma kako lahko nadzorujemo sprotno učenje modela in spremljamo njegov napredek? Ali se lahko v sistem vključi evalvacija ponujenega dokumenta? Uporabnik zavrne posamezne dele in vpiše popravek, s čimer bi sistem dobil povratno informacijo o boljšem rezultatu. Ali bo potrebno naknadno večje prilagajanje modela, na primer ročno, če se sam ne nauči dobro ali se model prilagaja samodejno?
8. Kako vaše podjetje skrbi za varovanje osebnih podatkov?
9. Koliko ur dela je, po vaši oceni, potrebnega za razvoj takšne aplikacije? Kako hitro bi lahko vzpostavili učno in testno okolje ter kaj bi za to potrebovali? Kdaj bi lahko potem lahko pričakovali prvi prototip? Kolikšna bi bila ocenjena vrednost razvoja predstavljene aplikacije?
10. Ali bi izpostavili kakšen pomemben vidik, ki ni bil omenjen oziroma menite, da bi nas morali na kaj opozoriti?

V razpravi smo ugotovili, da vsi zainteresirani skrbijo za varstvo podatkov na način, da jih čim manj obdelujejo oziroma, da pridejo v stik z osebnimi podatki v čim manjši možni meri. Izkazalo se je, da anonimizacija nima vpliva na učenje modela, zato bomo premislili v kolikšni meri bodo podatki anonimizirani.

Vsi udeleženci so izpostavili težave, ki bi jih imeli ob zahtevi, da se dela izvajajo na lokaciji uporabnika, saj je zelo pogosto potrebno timsko delo, kar bi zahtevalo prisotnost več strokovnjakov na lokaciji. Uporabnik bo zato ponovno razmislil o zahtevi.

Glede na počasen razvoj slovenskega velikega jezikovnega modela so bili udeleženci precej skeptični, da bi lahko enostavno prilagodili slovenski jezikovni model na potrebe uporabnika ob koncu projekta. Poleg tega je bila omenjena tudi težava porabe procesorske in grafične moči ter samega učenja modela. Zato je uporabnik ponovno pretehtal morebitno vključenost zahteve v specifikacije.

Vsi udeleženci so povedali, da je pri razvoju Virtualnega pomočnika smiselno uporabiti vse podatke, ki so na voljo, saj potem lahko uporabijo več tehnik umetne inteligence za zagotovitev rezultatov, kot jih zahteva uporabnik. Poleg tega je potrebno imeti dostop do dokumentov, če jih želimo prikazati, saj rekonstrukcija iz morebitne vektorske baze ni mogoča.

Kvaliteta preslikovanja dokumentov (OCR) nima odločilnega vpliva na rezultate Virtualnega pomočnika, zato je standardna kvaliteta dovolj za prepoznavo besedila v dokumentu. Poleg tega obstajajo tudi druge tehnike, s katerimi se lahko prepozna besedilo.

Udeleženci prav tako niso videli težav pri zagotavljanju razložljivosti rezultatov Virtualnega pomočnika. Obstajajo različni načini, kako bi lahko zagotovili, da Virtualni pomočnik beleži vir informacij in ga predstavi uporabniku.

Na vprašanje glede ocene potrebnih človeških virov je noben udeleženec ni želel podati. Povedali so, da je težko oceniti potrebne človeške vire in bi lahko podali popolnoma napačne ocene, zato ne bi špekulirali.

Individualni sestanki in strokovni dialog so se zaključili 17. 5. 2024. Ponovna srečanja v okviru strokovnega dialoga niso bila potrebna.

Skupni zapisnik se objavi na spletni strani www.gov.si