

**SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN**

Aleksandar Babić

**INFORMATIVNA PLATFORMA S
INTELIGENTNIM ASISTENTOM O
ERASMUS+ ISKUSTVU U ŁÓDŹU.**

DIPLOMSKI RAD

Varaždin, 2025.

SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
V A R A Ž D I N

Aleksandar Babić

Matični broj: 0016147642

Studij: Baze podataka i baze znanja

**INFORMATIVNA PLATFORMA S INTELIGENTNIM ASISTENTOM O
ERASMUS+ ISKUSTVU U ŁÓDŽU.**

DIPLOMSKI RAD

Mentor:

doc. dr. sc. Bogdan Okreša Đurić

Varaždin, rujan 2025.

Izjava o izvornosti

Izjavljujem da je ovaj diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

Autor potvrdio prihvatanjem odredbi u sustavu FOI Radovi

Sažetak

Teoretski dio rada istražuje studiranje kao važan segment u životnom razdoblju pojedinca te se dotiče razmjene studenata u svrhu Erasmus+ studijskog boravka. Teoretski dio završava poglavljem posvećenim inteligentnim asistentima temeljenima na RAG (eng. Retrieval-Augmented Generation) tehnologiji. Praktični dio rada obuhvaća izradu web-stranice s tematikom Erasmus+ studijskog boravka i izradu inteligentnog asistenta, pri čemu će se iskoristiti postojeći jezični model. To podrazumijeva treniranje i prilagođavanje specifičnom skupu podataka, kako bi se osiguralo da inteligentni asistent odgovara upitima korisnika i pruža relevantne informacije.

Ključne riječi: Inteligentni asistent, RAG, web stranica, studiranje, Erasmus+, umjetna inteligencija.

Sadržaj

1. Uvod	1
2. Studiranje kao dio života pojedinca	2
2.1. Odluka i benefiti	2
2.2. Erasmus+ studentska razmjena	2
2.2.1. Izazovi	2
2.2.2. Kako se izboriti?	2
2.2.3. Kultura kao faktor u akademskom uspjehu	2
3. Aktualne tehnologije i alati za uspješno snalaženje i informiranje	3
3.1. Važnost inteligentnih asistenata	3
3.2. RAG tehnologija	4
3.3. Integriranje inteligentnih asistenata u aplikacije i akademske sustave	6
4. Praktični dio rada	7
4.1. Izrada web stranice	7
4.2. Izrada inteligentnog asistenta	7
5. Zaključak	8
Popis literature	9
Popis slika	10
Popis tablica	11

1. Uvod

uvod

2. Studiranje kao dio života pojedinca

2.1. Odluka i benefiti

2.2. Erasmus+ studentska razmjena

2.2.1. Izazovi

2.2.2. Kako se izboriti?

2.2.3. Kultura kao faktor u akademskom uspjehu

3. Aktualne tehnologije i alati za uspješno snalaženje i informiranje

3.1. Važnost inteligentnih asistenata

Davne 1961. godine IBM je predstavio prvi sustav koji je mogao prepoznati govor (izgovorene znamenke). Do 1990-ih kreirani su prvi komercijalni osobni asistentni koji su se aktivirali na postojanje glasa. Godine 2010. svijet je prvi put čuo za Siri, Appleovog osobnog asistenta koji je pokrenutna temelju umjetne inteligencije. U to vrijeme Siri je predstavljao revolucionarno rješenje. Petnaest godina kasnije, korisnici su zatrpani s ogromnom količinom inteligentnih asistenata. Korisničko iskustvo (eng. UX) je sada više nego ikad važno. Održavanje zadovoljstva korisnika/kupca i upravljanje njihovim očekivanjima zahtijeva stalne inovacije i poboljšanja. Korisnici žele da stranice budu intuitivne i učinkovite te da na neki način zadovolje njihove potrebe. AI asistentni i chatbotovi su postali popularni alati koji pomažu u poboljšanju korisničkog iskustva. Točnije rečeno dodaju novi način interakcije između korisnika i pružatelja usluge i samim time povećavaju angažman, zadovoljstvo ili profit. Prednosti inteligentnih asistenata uključuju [1]:

- **Brzina i jednostavnost** - AI asistentni omogućuju korisnicima da brzo i jednostavno dobiju ono što im je potrebno, smanjujući gubitak vremena. Tik Tok je dobar primjer kako usluga može neindirektno zadržati korisnika na njihovoj platformi jer su razvili mehanizam koji je naučio pretpostavljati što korisnik želi gledati.
- **Pružanje podrške 24/7** - AI asistentni su dostupni stalno, čime je omogućena podrška korisnicima kad god je to potrebno.
- **Prihvatanje od strane ljudi** - razgovor s AI asistentima se odvija kroz razne dizajne korisničkih sučelja, gdje korisnici mogu komunicirati na ljudski razumljiv način s asistentima.

U sljedećem poglavlju je prikazano kako se mogu kreirati inteligentni asistentni temeljeni na RAG (eng. Retrieval-Augmented generation) tehnologiji.

3.2. RAG tehnologija

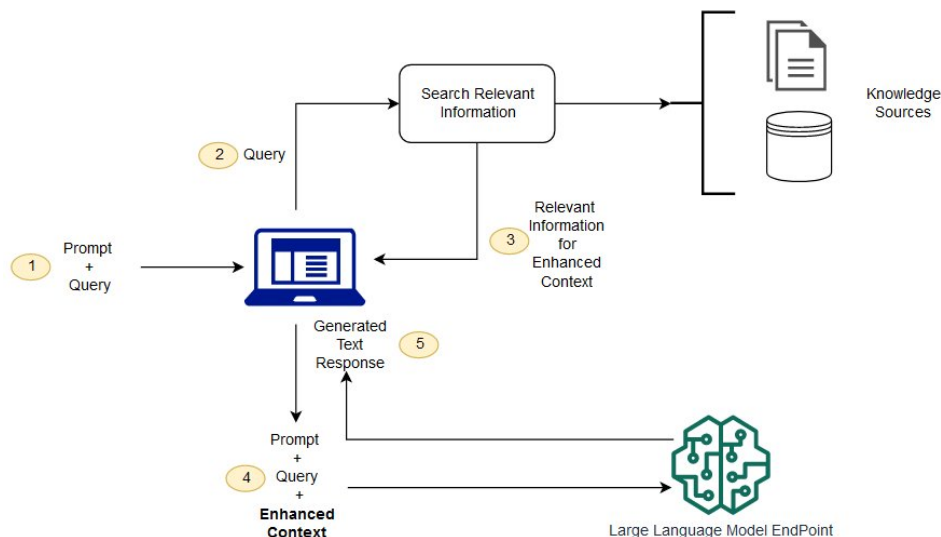
RAG je proces optimizacije izlaza velikog jezičnog modela pri čemu se referencira na autorativnu bazu znanja izvan izvora podataka za obuku prije generiranje odgovora. Veliki jezični modeli (eng. LLM) obučavaju se na ogromnim količinama podataka i koriste milijarde parametara za generiranje izlaza za zadatke poput odgovaranja pitanja, prevođenja jezika i dovršavanja rečenica. Ideja RAG-a je da proširi već moćne mogućnosti LLM-ova na određene specifične domene, a sve to bez potrebe za ponovnom obukom modela. Takav pristup poboljšava LLM-ove u smislu točnosti, korisnosti i relevantnosti [2].

LLM-ovi su ključna tehnologija umjetne inteligencije (AI) koja pokreće chatbotove. Priroda LLM-ova je takva da ponekad uvode dozu nepredvidljivosti u generiranje odgovora. Osim toga, podaci o LLM obuci su statični i imaju kranji rok za znanje koje posjeduju. Negativne strane LLM-ova uključuju [2]:

- Iznošenje lažnih informacija kada na njih nema odgovora,
- Prikazivanje zastarjelih informacija kada korisnik očekuje aktualan odgovor,
- Izrada odgovora iz neautoriziranih izvora,
- Stvaranje netočnih odgovora zbog terminološke zbrke, različiti izvori obuke koriste istu terminologiju za opisivanje različite stvari

Model velikog jezika može se zamisliti kao enuzijastičnog zaposlenika koji odbija biti informiran o aktualnim događajima, ali uvijek odgovara na pitanja s visokim samopouzdanjem. Nažalost, takav stav može negativno utjecati na povjerenje korisnika i nije nešto što bi trebalo postojati kod chatbotova. Iz takvih razloga RAG tehnologija preusmjerava LLM kako bi dohvatio relevantne informacije iz unaprijed definiranih izvora znanja [2].

Bez RAG-a, LLM uzima korisnički unos i stvara odgovor na temelju informacija na kojima je obučen. S RAG-om se uvodi komponenta za pronalaženje informacija koja koristi korisnički unos kako bi prvo uvekla informacije iz novog izvora podataka. Nakon toga LLM kombinira novo znanje i svoje podatke za stvaranje boljih odgovora [2].



Slika 1: Konceptualni tok korištenja RAG-a s LLM-ovima

U nekoliko koraka pojašnjeno je kako funkcioniraju stvari na prikazanoj slici 1[2]:

1. **Stvaranje vanjskih podataka** - novi podaci izvan originalnog skupa podataka koji su se koristili za treniranje LLM-a se zovu vanjski podaci (eng. external data). Oni dolaze iz više izvora podataka kao što su: API-ji ili baze podataka. Podaci mogu postojati u različitim formatima. Jedna od tehnika (eng. embedding language models) pretvara podatke u numerički oblik i sprema ih u vektorsku bazu podataka (kako bi generativni modeli umjetne inteligencije to mogli razumjeti).
2. **Preuzimanje relevantnih informacija** - Korisnički upit pretvara se u vektorski prikaz i uspoređuje se s vektorskim bazama podataka. Npr. ako korisnik postavi pitanje chatbotu "Koliko dana godišnjeg odmora imam?", chatbot će morati preuzeti dokumente o politici godišnjeg odmora uz evidenciju prošlih odmora za pojedinog zaposlenika. Takvi dokumenti će biti vraćeni korisniku jer su relevantni za njegov upit.
3. **Proširivanje LLM upita** - Zatim, RAG model proširuje korisnički upit dodavanjem relevantnih dohvaćenih podataka u kontekstu. Prošireni upit omogućuje modelima velikih jezika generiranje točnih odgovora na korisničke upite.
4. **Ažuriranje vanjskih podataka** - Vrlo je bitno održati ažurne informacije i to se može ostvariti automatiziranjem procesa u stvarnom vremenu ili periodičnoj obradi.

Tablica 1: Prednosti RAG-a

Prednost	Opis
Isplativa implementacija	Razvoj chatbota započinje korištenjem temeljnog modela (eng. Foundation models). Temeljni modeli su obično dostupni putem API ključeva i obučeni su na širokom spektru generaliziranih i neoznačenih podataka. Računalni i financijski troškovi za treniranje takvih modela su jako visoki.
Trenutne informacije	Iako su izvorni podaci LLM-a relevantni za specifične potrebe, teško je održati relevantnost. RAG se može koristiti za izravno povezivanje LLM-a s feedovima društvenih medija, web stranicama s vijestima i samim time pružiti korisnicima najnovije informacije.
Povećano povjerenje korisnika	RAG omogućuje prikazivanje točnih informacija s navođenjem izvora ili citata. Korisnici također mogu pretraživati izvorne dokumente ako im je potrebno dodatno pojašnjenje.
Više kontrole za razvojne programere	Učinkovitije testiranje i poboljšavanje aplikacija. Kontroliranje i mijenjanje izvora informacija. Rješavanje i ispravljanje problema.

3.3. Integriranje inteligentnih asistenata u aplikacije i akademske sustave

4. Praktični dio rada

4.1. Izrada web stranice

4.2. Izrada inteligentnog asistenta

5. Zaključak

zaključak

Popis literature

- [1] J. Buchan. „Why you should add an AI assistant to your app or website.” Pristupljeno: 06. svibanj 2025., Zudu. *adresa:* <https://zudu.co.uk/blog/why-you-should-add-an-ai-assistant-to-your-app-or-website/>.
- [2] Amazon Web Services, Inc. „What is RAG? - Retrieval-Augmented Generation AI Explained - AWS.” Pristupljeno: 06. svibanj 2025. *adresa:* <https://aws.amazon.com/what-is/retrieval-augmented-generation/>.

Popis slika

1.	Konceptualni tok korištenja RAG-a s LLM-ovima	5
----	---	---

Popis tablica

1.	Prednosti RAG-a	6
----	---------------------------	---