



## JEDAN PRISTUP ZA OTKRIVANJE IZAZOVA U KVANTNOM PROGRAMIRANJU IZ REPOZITORIJUMA OTVORENOG KODA

### AN APPROACH FOR DISCOVERING CHALLENGES IN QUANTUM PROGRAMMING FROM OPEN-SOURCE REPOSITORIES

Vladimir Filipović, Vladimir Mandić, *Fakultet tehničkih nauka, Novi Sad*

#### Oblast – IMT – INFORMACIONE TEHNOLOGIJE

**Kratak sadržaj** – U ovom radu predstavljen je pristup za analizu kvantnih softverskih repozitorijuma sa ciljem identifikacije glavnih izazova sa kojima se susreću programeri prilikom rada na kvantnim projektima. Metodologija obuhvata procese otkrivanja, kloniranja, predprocesiranja i klasifikacije repozitorijuma, kao i identifikaciju i grupisanje izazova. Korišćenjem i verifikacijom OpenAI modela u određenim koracima analize, uspešno je demonstrirano kako se veštačka inteligencija može koristiti za identifikaciju i klasifikaciju izazova u kvantnom softveru.

**Gljučne reči:** Kvantno programiranje, Softver otvorenog koda, Rudarenje podataka, Analiza komentara, OpenAI, Github

**Abstract** – This paper presents an approach for analyzing quantum software repositories with the aim of identifying the main challenges faced by developers when working on quantum projects. The methodology encompasses the processes of discovering, cloning, preprocessing, and classifying repositories, as well as identifying and grouping challenges. By utilizing and verifying the OpenAI model in certain steps of the analysis, it has been successfully demonstrated how artificial intelligence can be used to identify and classify challenges in quantum software.

**Keywords:** Quantum programming, Open source software, Data-mining, OpenAI, Github

#### 1. UVOD

Kvantno računarstvo predstavlja novu paradigmu u oblasti informacionih tehnologija koja koristi principe kvantne mehanike za obradu podataka na načine koji nisu mogući u radu sa klasičnim računarima [1]. Kvantni računari operišu sa kvantnim bitovima ili kubitima (engl. *qubit*), koji mogu da predstavljaju i nulu i jedinicu istovremeno, zahvaljujući fenomenima *superpozicije* i *sprezanja* [2]. Ove karakteristike teorijski omogućavaju kvantnim računarima da rešavaju određene probleme značajno brže nego što je to moguće sa tradicionalnim računarima, što otvara vrata napretka u raznim naučnim i industrijskim oblastima [1] [2].

#### NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Vladimir Mandić, vanr. prof.

Uprkos obećavajućim mogućnostima kvantnog računarstva, ono zbog svoje složenosti i novih paradigmi, koje se razlikuju od klasičnog softverskog inženjeringa dolazi sa sopstvenim skupom izazova [3]. Praktična primena i razvoj kvantnog softvera još uvek su u ranoj fazi razvoja, što zahteva dalja istraživanja specifičnih problema sa kojima se susreću programeri prilikom pisanja programskog koda koji se izvršavaju na kvantnom računaru [4].

Empirijske studije u kvantnom programiranju koje pokušavaju da identifikuju izazove prilikom razvijanja kvantnog softvera [4] [5] [6] su relativno retke. Najvećim delom zbog toga što je ovo polje u ranoj fazi razvoja što rezultuje u malom broju dostupnih i upotrebljivih repozitorijuma otvorenog koda [7]. Međutim, broj repozitorijuma otvorenog koda vezanih za kvantno računarstvo stalno raste poslednjih godina i danas broj repozitorijuma sa kvantnim programima prevazilazi 40000.

Cilj ovog istraživanja jeste da identifikuje najčešće i najbitnije izazove sa kojima se suočavaju softverski inženjeri koji razvijaju kvantne programe tokom razvoja svojih projekata.

Za potrebe istraživanja dizajniran je i implementiran empirijski pristup koji je baziran na analizi komentara u repozitorijumima otvorenog koda. Analizom komentara iz repozitorijuma otvorenog koda, težimo da steknemo uvid u praktične poteškoće koje se javljaju u razvoju kvantnog softvera. Ovaj proces omogućava da sistematski identifikovanje problema sa kojima se suočavaju programeri u polju kvantnog računarstva.

Rad je strukturiran na sledeći način: druga sekcija pruža detaljan pregled postojeće literature i srodnih studija, postavljajući kontekst ovog istraživanja. Treća sekcija nudi sveobuhvatno objašnjenje pristupa korišćenog u empirijskoj studiji. Četvrta sekcija predstavlja rezultate studije koji su protumačeni u petoj sekciji. Poslednja sekcija rezimira nalaze rada, njihove implikacije i moguća poboljšanja.

#### 2. TEORIJSKE OSNOVE

Kvantno programiranje donosi jedinstvene izazove koji se razlikuju od klasičnog softverskog inženjeringa. Ovi izazovi uključuju ispravljanje grešaka i otpornost na kvarove, skalabilnost kvantnih sistema, koherenciju i stabilnost kubitova, kao i razvoj kvantnih algoritama [3]. Svaka od ovih oblasti zahteva specijalizovane pristupe zbog inherentne složenosti kvantne mehanike. Na primer, teškoće u

osiguravanju ispravljanja grešaka i otpornosti na otkaze proizlaze iz krhke prirode kubitata koji su podložni dekoherenciji i operativnim greškama [8]. Pored toga, skalabilnost kvantnih sistema predstavlja jedan od većih izazova, jer povećanjem broja kubitata, održavanje njihove koherencije postaje eksponencijalno teže [9]. Razumevanje i rešavanje ovih izazova od suštinskog je značaja za napredak kvantnog softverskog inženjeringa [10].

Empirijske studije u kvantnom računarstvu, iako nekada retke, sada postaju sve učestalije kako se oblast razvija i kako se povećava broj dostupnih repozitorijuma [4] [5] [6]. Kao konkretan primer možemo uzeti rad [4] u kojem je analizirano preko 700 repozitorijuma kvantnog računarstva kako bi se opisalo trenutno stanje u oblasti i kako bi se identifikovali izazovi sa kojima se suočavaju programeri. Pomenuta studija istakla je brz rast i interesovanje za kvantno računarstvo, prikazujući ekspanziju repozitorijuma. Slično tome, studija iz 2022. godine fokusirala se na izazove kvantnog softverskog inženjeringa i diskutovala o potrebi za zrelim pristupima softverskom inženjeringu prilagođenim kvantnom softveru [5].

Rudarenje softverskih repozitorijuma je dobro uspostavljen pristup u softverskom inženjeringu, posebno sa pojavom mašinskih algoritama koji mogu da obrade velike skupove podataka [11][12], [13]. Prema našim saznanjima trenutno ne postoje istraživanja koja ovu tehniku primenjuju na repozitorijumima koji se bave kvantnim programiranjem.

### 3. METODOLOGIJA

Pristup se sastoji od sledećih koraka, koji su i vizuelno predstavljeni na slici 1.

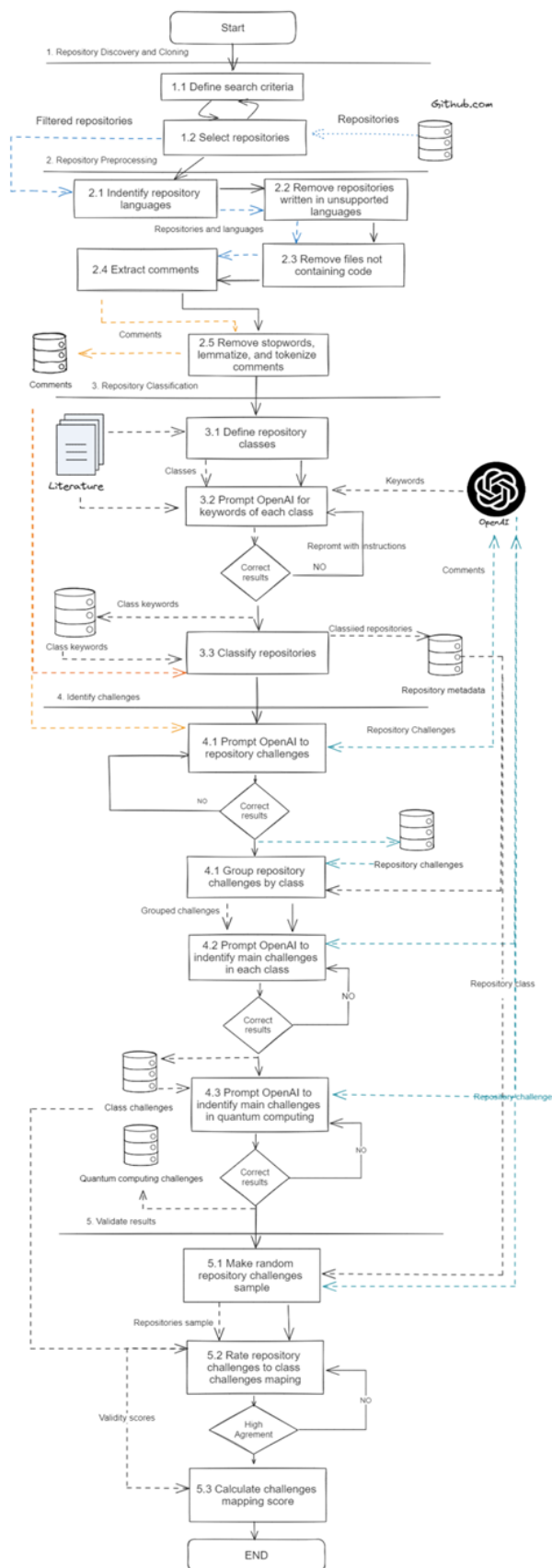
Prvi korak je (1) *identifikacija i kloniranje repozitorijuma* u kojem se otkrivaju i selektuju repozitorijuma otvorenog koda. Za potrebe pretrage, u okviru rada implementirana je jednostavna klijentska aplikacija napisana u React-u [14]. U okviru ove aplikacije korišćen je GitHub API [15] za pretragu i identifikaciju repozitorijuma vezanih za kvantno računarstvo. Prilikom pretrage korišćeni su sledeći kriterijumi:

- K1.** Relevantnost sadržaja: repozitorijum u svom nazivu ili opisu sadrži reč "Quantum",
- K2.** Aktivno održavanje: repozitorijum je ažuriran u prethodne dve godine, i
- K3.** Interesovanje zajednice: repozitorijum je označen sa barem 10 zvezdica.

Drugi korak (2) *predprocesiranje podataka* uključuje korišćenje alata pod nazivom Lynguis [16] za identifikaciju programskog jezika svakog repozitorijuma, filtriranje repozitorijuma napisanih u nepoznatim jezicima, kao i uklanjanje fajlova koji ne sadrže programski kod. Na posletku uz pomoć Pygments [17] biblioteke iz preostalih fajlova izdvojeni su komentari vezani za pojedinačne repozitorijume.

(3) *Klasifikacija repozitorijuma na osnovu komentara.* Paralelno sa korišćenjem OpenAI modela, ključne reči identifikovane su i putem pregleda literature. Iz literature su izdvojene ključne reči koje se često koriste u različitim oblastima kvantnog računarstva. Konačna lista ključnih

reči dobijena je kombinovanjem rezultata iz literature i odgovora dobijenih od OpenAI modela. Ovaj pristup omogućava dobijanje relevantnih i kontekstualno tačnih rezultata, jer su odgovori modela kontrolisani i potvrđeni putem naučnih radova [1] [2].



Slika 1. Koraci metodologije.

Kako je set podataka relativno mali i distribucija klasa nije podjednaka odlučeno je, da se u radu, klasifikacija repozitorijuma izvrši korišćenjem tehnike *fuzzy matching-a* [11].

(4) *Identifikacija izazova* je sprovedena u dva podkoraka, najpre identifikovani izazovi posebno za svaki repozitorijum koristeći OpenAI modele za analizu komentara iz izvornog koda. Ovi izazovi su potom na isti način grupisani prema izdvojenim oblastima kvantnog programiranja, a na posletku su agregirani kako bi se identifikovali opšti problemi koji nastaju prilikom razvoja kvantnih programa.

(5) *Validacija* je iterativno obavljena u skladu sa koracima 5.1 do 5.3 na slici 1. Validnost nasumično odabranih klasa ocenjivana je za svaki repozitorijum brojevima od 1 do 3, gde je 1 predstavljalo da izazovi datog repozitorijuma nisu odgovarajuće mapirani u izazove klase, a 3 da su izazovi dobro mapirani. Nakon ocenjivanja izračunat je *inter-rator agreement* izražen *Cohen's kappa koeficijentom*. Na kraju ovog koraka izračunata je vrednost validnosti celokupnog uzorka.

#### 4. REZULTATI ISTRAŽIVANJA

Inicijalno je pronađeno 1258 repozitorijuma, a nakon predprocesiranja broj repozitorijuma smanjen je na 764.

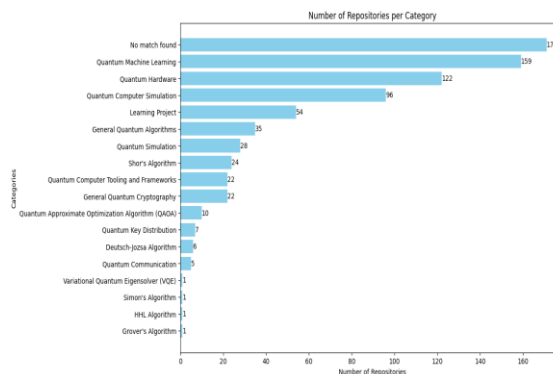
Rezultati pokazuju da su među najpopularnijim jezicima *Python* (30.6%), *Jupyter Notebook* (17.8%) (koji je od strane *Linguist* biblioteke prepoznat kao zaseban jezik od *Pythona*), *JavaScript* (9.8%), *C++* (6%), *Julia* (4.3%) i *C*(3%).

Rezultati klasifikacije kvantnih repozitorijuma prikazani su na slici 2, koja pokazuje broj repozitorijuma po kategorijama. Veliki broj repozitorijuma nije bio klasifikovan ni u jednu kategoriju zbog nedostatka ključnih reči koje bi ukazale na specifičnu oblast kvantnog računarstva.

U koraku 4.3 za svaku klasu, odnosno oblast, kvantnog inženjeringa, identifikovano je po 10 najbitnijih izazova sa kojima su se inženjeri sretali prilikom razvoja kvantnih programa u datoj oblasti. Rezultati pokazuju da svaka oblast pored izazova zajedničkih za sve oblasti kvantnog inženjeringa, sadrži i neke jedinstvene izazove.

Tabela 1 Izazovi po oblastima

Oblast	Izazovi
General Quantum Algorithms	Skalabilnost algoritama i njihova efikasnost na različitim kvantnim platformama.
Grover's Algorithm	Optimizacija performansi i smanjenje kompleksnosti implementacije.
Shor's Algorithm	Upravljanje resursima i stabilnost kubita tokom dugotrajnih proračuna.
Variational Quantum Eigensolver	Priprema kvantnog stanja i merenje sa visokim nivoom preciznosti.



Slika 2 Rezultati klasifikacije

Korišćenjem *OpenAI* modela u koraku 4.3 izazovi svih klasa agregirani su u izazove zajedničke za kvantno računarstvo. Poređani po frekvenciji javljanja identifikovana su top 3 izazova:

- (1.) korekcija grešaka i tolerancija na otkaze,
- (2.) skalabilnost kvantnih sistema,
- (3.) koherencija i stabilnost kubita,

Nakon nekoliko iteracija, vrednost *inter-rater agreement-a* kao metrike validnosti rezultata nad nasumično izabranim podskupom repozitorijuma iznosila je 0.675, a konačna ocena validnosti nasumično odabranog skupa repozitorijuma iznosila je 2.49 / 3.

#### 5. DISKUSIJA

Dobijeni rezultati ukazuju na to da se programeri pored izazova koji postoje i u klasičnom softverskom inženjeringu, susreću i sa jedinstvenim izazovima u koje sa sobom nosi kvantno računarstvo.

##### 5.1 Korekcija grešaka i tolerancija na otkaze

Korekcija grešaka i tolerancija na otkaze ključni su izazovi u kvantnom računarstvu jer su kvantni sistemi veoma osetljivi na šum i dekoherenciju. Čak i male smetnje iz okoline mogu dovesti do grešaka u kvantnim bitovima (kjubitima) [8]. Npr., Šorov algoritam zahteva dugotrajne proračune u kojima se greške mogu akumulirati, dovodeći do netačnih rezultata [1].

Predložene su različite tehnike i pristupi za korekciju grešaka u kvantnom računarstvu, uključujući: *Topological quantum codes* [18], *Fault-Tolerant gates* [19] i *Surface codes* [20].

##### 5.2 Skalabilnost kvantnih sistema

Problem skalabilnosti kvantnih sistema odnosi se na sposobnost proširenja kvantnih računara na veći broj kjubita dok se održavaju njihova funkcionalnost i performanse [9].

Kako bi rešavanje ovih problema bilo moguće, potrebno je da se dati kvantni program izvršava na sistemu koji sadrži veliki broj kjubita, stoga skalabilnost određuje ili ograničava u kojoj meri je moguće primeniti kvantno programiranje za rešavanje navedenih problema. U nedostatku skalabilnih sistema, ove primene bi ostale teorijska, a ne praktična rešenja [23].

### 5.3 Koherencija i stabilnost kubita

Koherencija i stabilnost kubita od fundamentalnog su značaja za izvršavanje kvantnih operacija. Koherencija se odnosi na sposobnost kubita da održi svoje kvantno stanje tokom vremena [1]. Kubiti moraju ostati koherentni dovoljno dugo da bi omogućili izvršavanje kvantnih algoritama, ali realni kubiti gube svoju koherenciju zbog interakcija sa okolinom. Ova dekoherencija ograničava vreme tokom kojeg se kvantni proračuni mogu izvršavati, što utiče na tačnost rezultata. Razvijanje kubita sa dužim vremenom koherencije i većom stabilnosti ključno je za ostvarivanje pouzdanih kvantnih računara [21].

Ovi problemi se najvećim delom rešavaju hardverskim rešenjima. Ali moguće ih je ublažiti kroz nekoliko inženjerskih praksi kao što su: *Calderbank-Shor-Steane* (CSS) kodovi [22] kao i posebne tehnike *dizajna samih algoritama* [24].

## 6. ZAKLJUČAK

Korišćenjem metoda analiziranja komentara iz repozitorijuma otvorenog koda, identifikovani su ključni izazovi sa kojima se suočavaju programeri tokom razvoja kvantnih programa. Dobijeni rezultati u skladu su sa izazovima koji su već uočeni i diskutovani u kvantnoj računarskoj zajednici, što potvrđuje validnost pristupa korišćenog u ovoj studiji.

U tumačenju rezultata, važno je uzeti u obzir da je korišćen ograničen skup repozitorijuma, što može uticati na generalizaciju nalaza. Takođe, za bolje rezultate istraživanja bilo bi od koristi istražiti rezonovanje na osnovu kog *ChatGPT-4o* donosi odluke.

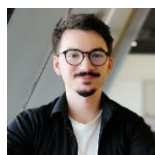
Kako bi dobijena vrednost istraživanja bila veća i relevantnija u daljim istraživanjima potrebno je uključiti i saradnju sa stručnjacima iz oblasti kvantnog inženjeringa koji bi pružili dublje razumevanje konteksta izazova i validnosti rešenja.

## 7. LITERATURA

- [1] Shor, P. W. (1998). Quantum computing. *Documenta Mathematica*, 1(1000), 467-486.
- [2] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010
- [3] Almudever, C. G., Lao, L., Fu, X., Khammassi, N., Ashraf, I., Iorga, D., ... & Bertels, K. (2017, March). The engineering challenges in Quantum programming
- [4] De Stefano, M., Pecorelli, F., Di Nucci, D., Palomba, F., & De Lucia, A. (2022). Software engineering for quantum programming: How far are we?. *Journal of Systems and Software*, 190, 111326.
- [5] Openja, M., Morovati, M. M., An, L., Khomh, F., & Abidi, M. (2022). Technical debts and faults in open-source quantum software systems: An empirical study. *Journal of Systems and Software*, 193, 111458.
- [6] Li, H., Khomh, F., & Openja, M. (2021, September). Understanding quantum software engineering challenges an empirical study on stack exchange forums and github issues. In *2021 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 343-354). IEEE.
- [7] Fingerhuth, M., Babej, T., & Wittek, P. (2018). Open source software in quantum computing. *PloS one*, 13(12), e0208561.

- [8] Gottesman, D. (2002). An introduction to quantum error correction. In *Proceedings of Symposia in Applied Mathematics* (Vol. 58, pp. 221-236).
- [9] Van Meter, R., & Devitt, S. J. (2016). The path to scalable distributed quantum computing. *Computer*, 49(9), 31-42.
- [10] Murillo, J. M., Garcia-Alonso, J., Moguel, E., Barzen, J., Leymann, F., Ali, S., ... & Wimmer, M. (2024). Challenges of Quantum Software Engineering for the Next Decade: The Road Ahead. arXiv preprint arXiv:2404.06825.
- [11] Lorensuhewa, S., Pham, B., & Geva, S. (2002). Keyword-based Text Matching Approach for Design Style Recognition. In *Proceedings First International Workshop on Knowledge Discovery in Multimedia and Complex Data (KDMCD'2002)* (pp. 51-58). <http://db.cs.ualberta.ca/kdmcd02/>.
- [12] McCallum, A., & Nigam, K. (1999). Text classification by bootstrapping with keywords, EM and shrinkage. In *Unsupervised learning in natural language processing*.
- [13] Meqdadi, O., Alhindawi, N., Alsakran, J., Saifan, A., & Migdadi, H. (2019). Mining software repositories for adaptive change commits using machine learning techniques. *Information and Software Technology*, 109, 80-91.
- [14] React documentation. Dostupno: <https://react.dev/> [Pristup: 5.5.2024]
- [15] Github API documentation. Dostupno: <https://docs.github.com/en/rest> [Pristup: 5.5.2024]
- [16] Linguist documentation. Dostupno: <https://github.com/github-linguist/linguist/blob/master/docs/how-linguist-works.md> [Pristup: 1.6.2024]
- [17] Pygments documentation. Dostupno: <https://pygments.org/docs/> [Pristup:3.6.2024]
- [18] Kitaev, A. Y. (2003). Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1), 2-30.
- [19] Knill, E. (2005). Quantum computing with realistically noisy devices. *Nature*, 434(7029), 39-44. Link
- [20] Fowler, A. G., Mariantoni, M., Martinis, J. M., & Cleland, A. N. (2012). Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3), 032324
- [21] Franklin, D., & Chong, F. T. (2004). Challenges in reliable quantum computing. *Nano, quantum and molecular computing: implications to high level design and validation*,
- [22] Harris, R. J., McMahon, N. A., Brennen, G. K., & Stace, T. M. (2018). Calderbank-Shor-Steane holographic quantum error-correcting codes. *Physical Review A*, 98(5), 052301.
- [23] Hoefler, T., Häner, T., & Troyer, M. (2023). Disentangling hype from practicality: On realistically achieving quantum advantage. *Communications of the ACM*, 66(5), 82-87.
- [24] Berberich, J., Fink, D., & Holm, C. (2024). Robustness of quantum algorithms against coherent control errors. *Physical Review A*, 109(1), 012417.

### Kratka biografija:



Vladimir Filipović rođen je 23.septembra 1998. godine u Petrovcu na Mlavi. Dobitnik je stipendije za izuzetno nadarene studente Ministarstva prosvete, nauke i tehnološkog razvoja. Osnovne akademske studije završava 2021. godine sa opštim uspehom 9,62.

kontakt: vladimirfilipovic@uns.ac.rs