

КОМИСИЈИ СТУДИЈСКОГ ПРОГРАМА МАСТЕР НИТ

Комисија студијског програма Мастер НИТ именовала нас је у Комисију за преглед и оцену мастер рада дипл. инж. Николе Бијанића под насловом „Утицај метода парчања текста на квалитет добављања информација у RAG системима“. Након прегледа материјала Комисија подноси следећи

ИЗВЕШТАЈ

1. Биографски подаци кандидата

Никола Бијанић је рођен 19.07.1993. године у Краљеву. Завршио је основну школу „Попински борци“ у Врњачкој Бањи као одличан ђак. Уписао је Врњачку гимназију у Врњачкој Бањи коју је завршио са одличним успехом. Факултет организационих наука, смер за Информационе системе и технологије уписао је 2013. године и дипломирао са просечном оценом 8,21. Дипломски рад одбранио је у септембру 2020. године са оценом 10. Дипломске академске – мастер студије на Електротехничком факултету у Београду и Факултету организационих наука у Београду, на студијском програму Напредне информационе технологије у дигиталној трансформацији, модул за Анализу података уписао је у октобру 2020. године. Положио је све испите са просечном оценом 9,6 и испунио обавезе предвиђене мастер програмом – стручна пракса и тимски стартап пројекат. Тренутно ради као *Senior Data Scientist* у фирми *EPAM*.

2. Извештај о студијском истраживачком раду

Кандидат Никола Бијанић је у оквиру припреме за израду мастер рада спровео истраживање релевантне литературе у области обраде природног језика, посебно се фокусирајући на системе за генерисање текста добављањем информација (*Retrieval-Augmented Generation - RAG*), парчање текста (*chunking*), векторске базе података, и мерење семантичке сличности. Кандидат је анализирао постојеће приступе парчању текста, укључујући рекурзивно парчање, хијерархијско парчање, парчање засновано на семантичком кластеровану, и парчање засновано на груписању семантичко сличних реченица. Истраживање је обухватило и анализу *QASPER* скупа података, који је коришћен за евалуацију. Кандидат је користио *Python*, заједно са библиотекама као што су *LangChain*, *Sentence Transformers*, *NLTK*, *scikit-learn*, *ChromaDB*, и *HuggingFace Transformers*. Детаљно је истражио и применио различите методе за генерисање ембединга (*embeddings*), као и метрике за евалуацију квалитета добављања текста и генерисаних одговора.

3. Опис мастер рада

Мастер рад обухвата 55 страна, са укупно 14 слика, 7 табела и 57 референци. Рад садржи увод, 4 поглавља, закључак, списак коришћене литературе и прилоге.

- Прво поглавље (Увод): Представља проблематику и мотивацију истраживања, циљеве рада и структуру рада. Истиче се значај RAG система и проблем оптималног парчања текста.
- Друго поглавље (Теоријска позадина): Детаљно описује RAG системе, њихову архитектуру, компоненте (индексирање, преузимање, аугментација, генерисање), ембединг моделе и њихову улогу у RAG системима.

- Треће поглавље (Методологија): Описује коришћени скуп података (QASPER), методе парчања текста (рекурзивно парчање, хијерархијско парчање, семантичко кластеровање, парчање засновано на груписању семантичко сличних реченица), као и метрике за евалуацију.
- Четврто поглавље (Резултати и анализа): Приказује резултате експерименталне евалуације различитих метода парчања текста. Анализира се број чанкова по методи, упоређују се методе парчања, разматрају се ограничења и будући рад.
- Пето поглавље (Закључак): Сумира кључне налазе истраживања, истиче допринос рада и наводи могућности за даља истраживања.
- Шесто поглавље (Литература): Списак коришћене литературе.
- Прилози: Укључују промпт шаблоне, пример *JSON* структуре *QASPER* сета података и линк ка програмском коду.

4. Анализа рада са кључним резултатима

Мастер рад кандидата Николе Бијанића представља детаљну анализу утицаја различитих метода парчања текста на квалитет добављања информација у RAG системима. Посебан допринос рада огледа се у компаративној анализи четири различита приступа парчању текста и њиховој евалуацији на *QASPER* скупу података. Резултати показују да методе засноване на семантичким репрезентацијама (парчање засновано на груписању семантичко сличних реченица, парчање засновано на семантичком кластеровању) генерално постижу боље резултате од традиционалних метода (рекурзивно парчање). Хијерархијско парчање се показало као ефикасно за добро структуриране документе. Рад такође указује на ограничења и изазове у RAG системима, као што су нејасноћа упита и ограничења ембединг модела.

5. Закључак и предлог

Кандидат Никола Бијанић је у свом мастер раду успешно истражио утицај метода парчања текста на квалитет добављања информација у RAG системима. Истраживање је спроведено систематично, уз коришћење релевантних алата и техника. Резултати рада имају практичну примену у развоју и унапређењу RAG система, посебно у домену анализе научних радова и структурираних докумената.

На основу изложеног, Комисија предлаже Комисији студијског програма Мастер НИТ да рад кандидата Николе Бијанића прихвати као мастер рад и кандидату одобри јавну усмену одбрану.

Београд, 21.03.2025. године

Чланови комисије:

Др Милош Јовановић, ван. проф.



Др Борис Делибашић, ред. проф.



Др Бошко Николић, ред. проф.

