

КОМИСИЈИ ЗА СТУДИЈЕ II СТЕПЕНА ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА У БЕОГРАДУ

Комисија за студије II степена, Електротехничког факултета у Београду, на својој седници одржаној 10.07.2025 године именовала нас је у Комисију за преглед и оцену мастер рада дипл. инж. Маријана Христић под насловом „Систем за формирање профила корисника на основу његових активности“. Након прегледа материјала Комисија подноси следећи

ИЗВЕШТАЈ

1. Биографски подаци кандидата

Маријана Христић је рођена 23.4.2000. године у Београду. Завршила је Основну школу „Јован Јовановић Змај“ у Обреновцу 2015. године као носилац Вукове дипломе. Средње образовање наставила је у Гимназији у Обреновцу, природно-математички смер, коју је такође завршила са одличним успехом и Вуковом дипломом.

Електротехнички факултет Универзитета у Београду уписала је 2019. године, на смеру Рачунарска техника и информатика. Дипломирала је у септембру 2023. године са просечном оценом на студијама 8,16. Завршни рад на основним студијама израдила је на тему „Интернет апликација за археолошке артефакте“ под менторством проф. др Милоша Цветановића.

Мастер академске студије на Електротехничком факултету у Београду уписала је октобра 2023. године на модулу Софтверско инжењерство. Положила је све испите са просечном оценом 10,00.

2. Извештај о студијском истраживачком раду

Кандидаткиња је, као припрему за израду мастер рада, спровела преглед релевантне литературе и постојећих решења у области профилисања корисника на основу веб-активности. Посебно су анализирани: технике веб-скрејпинга и парсирања HTML садржаја, изазови обраде динамичких страница и ограничења (rate-limit), затим приступи за оптичко препознавање слова (OCR) ради екстракције текста из слика, као и савремени модели за обраду природних језика (NLP) попут BERT/DistilBERT за класификацију текстова. На основу тога дефинисан је скуп категорија интересовања и одабран је приступ који комбинује асинхроно преузимање садржаја, претпроцесирање текста и фајн-тјунинг трансформер модела.

3. Опис мастер рада

Мастер рад обухвата 39 страну, са 4 слике и 19 исечака кода. Структурно је организован у увод, четири поглавља и закључак (укупно шест поглавља), уз списак скраћеница, списак слика и литературу.

Уводно поглавље даје мотивацију и предмет истраживања, као и разлоге за избор приступа профилисању корисника на основу веб-активности.

Следеће поглавље формулише проблем и описује улазне податке, уз истицање кључних изазова као што су разноврсност и квалитет садржаја, ограничења брзине преузимања и потреба за чишћењем и дедупликацијом текста.

Треће поглавље сажето представља коришћене алате и методолошка полазишта (Kaggle, Python екосистем, парсирање HTML-а, OCR, претходна обрада текста и моделовање), дајући контекст за избор технологија и њихову међусобну комплементарност.

У четвртном поглављу систем се приказује у целини: од читавања логова и асинхроног преузимања садржаја, преко чишћења и нормализације текста, до класификације помоћу савремених трансформер-модела (DistilBERT) и агрегирања резултата у профиле интересовања по кориснику. Акценат је на репродуцибилности, модуларности и јасно дефинисаним интерфејсима између фаза, што омогућава даље проширивање и примену на већим скуповима података.

Пето поглавље доноси закључак, а шесто предлаже могућности унапређења. Рад је тиме заокружен као целина која од мотивације доследно води до функционалног система, са

прецизно означеним ограничењима и смерницама за даљи развој и примену.

4. Анализа са кључним резултатима

Рад доноси комплетан, технички коректан и репродуцибилан ток обраде од логова (IP, URL) до профила интересовања корисника. Комбинацијом HTML парсирања и оптичког препознавања текста повећана је покривеност текста у односу на класичног обиласка и дохватања података са сајтова, док примена DistilBERT омогућава семантички осетљивију класификацију садржаја у односу на једноставне статистичке методе. Излаз је јасан и употребљив: класификовани записи по интернет адреси ресурса и сумарни профили интересовања по адреси интернет протокола (топ категорије). Систем је модуларно организован (асинхронно преузимање, претпроцесирање, класификација, агрегирање), што олакшава даље проширивање и продукционизацију.

5. Закључак и предлог

Кандидаткиња је реализовала функционалан систем који аутоматизује прикупљање, обраду и класификацију веб садржаја и на тој основи формира профиле интересовања корисника. Решење је добро структурирано, прошириво и усклађено са савременом праксом у NLP-у.

На основу изложеног, Комисија предлаже Комисији за студије II степена Електротехничког факултета у Београду да рад дипл. инж. Маријане Христић прихвати као мастер рад и кандидату одобри јавну усмену одбрану.

Београд, 03.09.2025. године

Чланови комисије:

др Милош Цветановић Ванредни професор
сагласан, 01.09.2025.

др Саша Стојановић Ванредни професор
сагласан, 01.09.2025.