РЕЦЕНЗИЯ

от проф. д-р Татяна Атанасова,
Институт по информационни и комуникационни технологии – БАН
на дисертационен труд за присъждане на образователната и научна степен „доктор“
по докторска програма „Информатика“,
професионално направление 4.6. Информатика и компютърни науки
на тема:
МЕТОДИ И АЛГОРИТМИ ЗА ПЕРСОНАЛИЗАЦИЯ И АДАПТИВНОСТ В СРЕДИ ЗА УПРАВЛЕНИЕ НА СЪДЪРЖАНИЕ,
представена от **Емануела Димитрова Митрева**
с ръководител проф. д-р Десислава Панева-Маринова

Със заповед № 57/17.06.2026 на Директора на ИМИ-БАН съм определена в състава на Научното жури във връзка със защита на дисертационен труд на Емануела Димитрова Митрева на тема: „Методи и алгоритми за персонализация и адаптивност в среди за управление на съдържание“, за присъждане на образователна и научна степен “доктор“ в област на висше образование 4. Природни науки, математика и информатика; професионално направление 4.6. Информатика и компютърни науки; по докторска програма „Информатика“.

1. Обща характеристика на дисертационния труд

Представеният дисертационен труд от Емануела Митрева е с обем от 167 страници, включващ увод, пет глави, приноси на дисертационния труд, списъци на авторски публикации, цитирания, докладвани резултати, 26 фигури, 15 таблици, речник на използваните термини и библиография от 202 заглавия. Работата адресира цифровата трансформация в научното и културното наследство и разглежда развитието на дигиталните библиотеки и необходимостта от внедряване на интелигентни системи за управление на огромните масиви от данни. Трудът се фокусира върху прехода от обикновено съхранение на данни към селективно и потребителски ориентирано представяне на научното и културно наследство чрез методи на Искуствен Интелект (ИИ).

2. Съдържателен анализ на представения дисертационен труд и публикациите към него

Встъпителната част на труда очертава общата постановка на задачата. Дефинирани са обектът, основната цел и предметът на изследването. В качеството на обект на дисертационния труд е дефиниран процесът по адаптация и персонализация на съдържанието в дигитални библиотеки чрез използване на методи и техники на изкуствения интелект и машинното обучение. Формулирани са пет основни изследователски задачи.

Във **втората глава** са представени теоретичните основи и съвременните концепции, свързани с персонализацията на информационно съдържание.
Отбелязана е еволюцията на съвременната дигитална библиотека от статично хранилище в гъвкава, задвижвана от ИИ екосистема, която съчетава сигурно дългосрочно архивиране с интелигентни, персонализирани и глобално достъпни услуги за потребителите. В този контекст ролята на дигиталната библиотека е преосмислена – тя не е просто архив, а динамичен център за обмен на знания, който се адаптира към нуждите на потребителите. Като високотехнологична и стабилна инфраструктура, тя стимулира академичните иновации и гарантира дълголетието на културната памет, заемайки централно място в днешната дигитална екосистема. Обосновава се необходимостта от интелигентни алгоритми за филтриране в големи масиви от електронни ресурси и използване на методи за персонализация при търсене на информация.
Представената *Таблица 2 – „Сравнение между подходите и методите за персонализация*“ демонстрира способността на автора да систематизира, класифицира и съпоставя широк спектър от технологични решения в областта на персонализацията на съдържание. Таблица 2 доказва, че докторантът е извършил задълбочен предварителен анализ на технологичния инструментариум.
Трудът систематизира развитието на концепцията за персонализация, проследявайки нейната траектория от елементарни статични инструменти (анкети) и традиционно филтриране на съдържанието до съвременни, комплексни и адаптивни подходи (хибридни модели, размити множества, машини на опорните вектори и обучение с частичен надзор).
Уместно е да се подчертае използването на термина „съвместно“, тъй като филтрирането и генерирането на препоръки се основава на споделените действия, оценки или поведение на определена потребителска група. Характеристиките на алгоритмите са дефинирани коректно от научно-приложна гледна точка (например: правилно е идентифициран проблемът със „студения старт“ при съвместното филтриране и ограничението за независимост на признаците при Наивния Бейсов класификатор).
За целите на персонализацията от съществено значение е как се представя текстовата информация и как се измерва нейната сходност. В този контекст са обобщени водещите методи за векторизация на текст и съответстващите им математически мерки за близост.
Научно-теоретичният обзор в тази глава се базира на детайлно проучване на селектирани публикации, издадени в периода 2018–2025 години. Обхванати са разнообразни емпирични, теоретични и научно-приложни изследвания. Извършеният обзор дава възможност на докторанта да аргументира убедително избора на собствен методологически подход, върху който да развие архитектурата и хибридния модел за адаптивна персонализация в следващите глави на труда.

Третата глава се фокусира върху методологичната основа и архитектурната организация на предложеното решение за персонализирано представяне на съдържание в дигитална библиотека. Описани са архитектурата и принципите на хибридна система за персонализирани препоръки.
Предложен е съвременен модел за кодиране на данни заместващ класическа математическа редукция на размерността, която крие риск от загуба на семантични детайли. Избраната архитектура (MiniLM) генерира компактни

плътни вектори по дизайн, които успешно заменят високоразмерните разредени матрици. Този подход позволява инкрементално добавяне на нови документи без нужда от реиндексация.
В дисертацията е представен модел, съчетаващ факторите „съдържание“ и „популярност“ за справяне с проблема за „студения старт“, като акцентът е поставен върху прозрачността на архитектурата и обяснимостта на резултатите. Използва се единен индекс на сходство, който обединява поведенчески данни и съдържателни показатели. Водещото предположение в това изследване е, че този хибриден модел е по-ефективен в критични сценарии като, например, липса на история или добавяне на нови ресурси, в сравнение със системите, ползващи само един източник на данни.
Предложеният модел притежава вградена обяснимост и му позволява да се види точно какъв е приносът на всеки компонент по тематично сходство и общи „именувани същности“.
Представеният потребителско-центриран модул реализира хибриден подход за персонализирани препоръки чрез интегриране на семантична близост, претеглени поведенчески данни и вектор на глобална популярност, което гарантира устойчивост на системата при липса на потребителска история.

Четвъртата глава представя практическата имплементация и експерименталната валидация на предложената софтуерна архитектура за персонализация и на хибридния модел за препоръчване. Архитектурата разделя тежката предварителна обработка от реалното обслужване, за да гарантира бързи, точни и обясними препоръки чрез хибриден модел от съдържание и популярност. Системата за препоръки има асинхронен и интерактивен слой. Целта е тежките изчисления да не пречат на бързината на потребителските заявки. Поддържат се поетапни обновления – при добавяне или редактиране на документ се преизчисляват само неговите връзки, без да се рестартира целият процес.
Сходството между документите се изчислява чрез претеглена линейна комбинация от четири независими слоя, съчетаващи съдържанието на текстовете и информацията за наличните в тях именувани същности.
Валидирането на предложения модел се осъществява чрез двуетапна процедура, съчетаваща търсене по решетка с компонентно изключване (аблация) за измерване приноса на отделните слоеве, последвано от многокритериален анализ на структурната кохезия, разграничимост и стабилност на класирането при инкрементални данни.
Демонстрирани са предимствата на многокомпонентната оценка на документите/резултатите/препоръките.
Архитектурата е верифицирана върху синтетичен набор от данни за потребителски профили и извадка от дигитална библиотека, като са очертани „границите, в рамките на които тя е коректна и полезна“. Тази глава очертава отчетените от автора слаби страни в няколко категории: ограничения, свързани със специфичността на текстовия корпус (предимно български медии и OCR шум); ограничения на съдържателния модел (зависимост от предварително обучени модели, неоптимизирани за специфични домейни, и субективен избор на параметри); лимити при поведенческите данни, както и технически ограничения, свързани с изчислителната тежест при обработката в реално време.

Петата глава представя **заключение**, обобщава основните резултати и посочва насоките за бъдещо развитие на предложената софтуерна архитектура за персонализация в среди за управление на съдържание.

Приноси на дисертационния труд са формулирани като научни и научно-приложни, като съставляват логическо и естествено заключение на проведеното изследване. Приемам предложената от докторанта класификация и формулировка на приносите.

3. Критични бележки

В окончателния вариант на дисертационния труд докторантът е отразил изцяло направените от мен критични бележки и препоръки по време на предварителното обсъждане на изследването.
Оценявам избора на експериментален корпус, съставен от дигитализирани броеве на многотематични периодични издания, което поставя специфични предизвикателства пред обработката на текстовете. Предлаганата софтуерна архитектура е верифицирана върху дигитализирани броеве на българския периодичен печат. Този тип многотематични периодични издания (вестници) притежават специфична структура и динамика, което придава допълнителна приложна стойност на изследването.
Постигнатите ценни резултати не изключват наличието на аспекти, които по-скоро насочват към последващи научни търсения, отколкото подлежат на критични бележки. Прекомерното разчитане на синтетично генерирани данни за доказване на стабилността на хибридния алгоритъм при сложни сценарии показва определен дефицит. За целите на поведенческото тестване докторантът използва изкуствено конструиран набор от данни, съдържащ 60 текстови документа, организирани в три клъстера. Това изисква по-нататъшно проучване на поведение на алгоритъма в контекста на реална експлоатация. Необходимо е и изследване относно добавената изчислителна сложност на семантичния слой в реална експлоатационна среда с големи обеми и натоварване.
Приложение на съвременни софтуерни инструменти за обработка на естествен език върху специфичните анализирани архивни ресурси също изисква преглед на качеството на данните на няколко нива.
Тези бележки не намаляват стойността на изследването, а по-скоро посочват възможни посоки за надграждане и разширяване на бъдещи разработки.
Декларирам, че няма доказано по законоустановения ред плагиатство в представения дисертационен труд и научни трудове по тази процедура.

4. Значимост на приносите за науката и практиката

Публикациите, отпечатани в престижни международни издания, индексирани в световни наукометрични бази данни, демонстрират способностите на докторанта да провежда научни изследвания в съвременните области на развитието на информационните технологии. Справката в Scopus за Емануела Митрева показва h-индекс=2 и 16 документа, а във WoS са индексирани 11 публикации с H-индекс=2. Забелязаните четири независими цитирания потвърждават първоначалния научен отзвук на публикуваните резултати. Отбелязвам публикацията в престижно списание Applied Sciences (MDPI) от 2026 г., посветената на хибриден подход за персонализирано и интелигентно препоръчване на съдържание в дигитални библиотеки, както и публикация в поредицата на Springer, описваща мултикомпонентна метрика за сходство.

Съгласно Правилника за прилагане на Закона за развитието на академичния състав в Република България тези публикации носят на докторанта общо 174 точки, с което е постигнато и значително надвишено изискването на минимум 30 точки по показател Г за дисертация за получаване на образователната и научна степен „доктор“.

Регистрираното присъствие и цитируемост на изследванията в световните наукометрични бази Scopus и WoS служат като обективна оценка за международното признание на докторант Емануела Митрева в избраната от нея научна област.

Емануела има и документирано участие в национални проекти: „Млади учени и постдокторанти – 2“и CLADA-BG - Национална интердисциплинарна изследователска Е-инфраструктура за ресурси и технологии за българското езиково и културно наследство, интегрирана в рамките на европейските инфраструктури CLARIN и DARIAH (КЛаДА-БГ).

Качества на автореферата

Авторефератът е в обем от 42/39 страници (български език/английски език) и отговаря на всички изисквания за изготвянето му като отразява коректно резултатите и съдържанието на дисертационния труд и научните приноси на докторанта.

5. Заключение

Дисертационният труд на Емануела Димитрова Митрева представлява актуално и технологично наситено изследване. То предлага ефективни решения чрез внедряване на модели за обработка на естествен език и алгоритми за разпознаване на именувани същности, като същевременно изгражда практически приложима и технически устойчива софтуерна архитектура за персонализиран достъп в дигиталните библиотеки. Работата напълно съответства на изискванията за придобиване на образователната и научна степен „доктор“ и адресира критични нужди на съвременното общество.

Научните приноси в труда, подкрепени от 5 публикации (включително 1 самостоятелна), удовлетворяват всички условия за присъждане на образователната и научна степен „доктор“. Постигнатите резултати надвишават минималните изисквания съгласно ЗРАСРБ и съпътстващите го нормативни документи - Правилника за прилагането му и съответните нормативни документи на БАН и ИМИ-БАН

Постигнатите резултати ми дават основание да дам **положителна оценка** и препоръчвам на почитаемото Научно жури да присъди образователната и научна степен „доктор“ на маг. **Емануела Димитрова Митрева** в професионално направление 4.6. „Информатика и компютърни науки“ по докторска програма „Информатика“.

11.08.2026 г. ……..

/ проф. д-р Татяна Атанасова /