



ПРЕДЛОЖЕНИЕ

за провеждане на докторантура в редовна форма на обучение
по реда на чл. 21, ал. 7 от ЗВО

в Института по математика и информатика при Българска академия на науките

професионално направление: 4.6. Информатика и компютърни науки
докторска програма: „Информатика“

тема на докторантурата:

**Методи и технологии за интеграция, семантично обогатяване и картографиране на
отворени данни за публично изкуство**

Предложил темата: доц. д-р Милена Добрева

Актуалност

Публичното изкуство е съществен, но недостатъчно добре документиран и систематично проучен сектор от културното наследство. То обхваща разнообразни форми – скулптури, стенописи, инсталации, временни художествени експонати, паметни плочи, които са силно обвързани с конкретния градски/селски/природен и социален контекст. Данните за публичното изкуство са разпръснати в множество хетерогенни източници : институционални регистри, научни публикации, дигитални архиви, както и в отворени платформи за представяне на знания (напр. Wikidata, OpenStreetMap), а все по-често и в социалните медии и неформални онлайн общности.

Тази фрагментация води до съществени проблеми, свързани с непълнота, несъгласуваност и липса на стандартизация на данните. В много случаи информацията е частична, дублирана или трудно съпоставима, което затруднява нейното използване и повторна употреба. Липсата на общоприети модели за описание, както и на ефективни механизми за интеграция на данни от различни източници, ограничава възможностите за изграждане на цялостна и надеждна картина на публичното изкуство като част от културната екосистема.

От информатична гледна точка, това поставя редица предизвикателства, свързани с моделиране на данни, интеграция на хетерогенни източници, семантично обогатяване и автоматизирано извличане на знания. Съществува необходимост от разработване на методи, които да позволят превръщането на разпокъсаната информация в структурирани и взаимосвързани набори от данни, съвместими с принципите на отворената. Липсата на такива решения ограничава използването на данните за публичното изкуство в различни приложения – изследователски, образователни, туристически, както и в творческите индустрии и градското планиране. Същевременно, на европейско ниво се наблюдава засилен интерес към развитието на отворени и свързани данни за културното наследство, включително в рамките на инициативи като Европейския облак за културно наследство и EOSC.

Предложената докторантура е пряко свързана с приоритетите на проекта FOCUS, по-специално с направленията „Datafication of cultural heritage“, „AI in GLAM“ и „Open knowledge infrastructures“. Чрез разработването на методи за интеграция и обогатяване на данни за публичното изкуство, тя допринася за превръщането на културните обекти в

анализируеми и повторно използваеми цифрови ресурси, като по този начин подкрепя изграждането на устойчиви изследователски и иновационни екосистеми в областта на дигиталното културно наследство.

Цел на докторантурата

Целта на докторантурата е разработването и експерименталната валидация на информатични методи и модели за интеграция, семантично обогатяване и анализ на отворени данни за публично изкуство. За постигане на тази цел се поставят следните конкретни задачи:

- **Идентифициране и систематизиране на източници на данни** за публично изкуство в България, потенциално с два сценария – столичен квартал и малък град, включително анализ на правните и лицензионни условия за тяхното използване и повторна употреба;
- **Дефиниране на концептуален и семантичен модел („паспорт“) на обекти на публичното изкуство**, както и изследване на възможностите за неговото (полу)автоматично генериране чрез интеграция на отворени и разпределени данни;
- **Разработване и експериментиране с методи за интеграция на хетерогенни данни**, включително подходи за съвместяване, обогатяване и свързване на информация от различни източници;
- **Изследване на методи за пространствен анализ, картографиране и визуализация**, приложими към данни за публично изкуство, включително комбиниране на геопропространствени и семантични представяния;
- **Систематизиране на изискванията към прототипна софтуерна среда**, която съчетава процесите по откриване, извличане, моделиране, интеграция и визуализация на данни за публичното изкуство;
- **Анализ и валидиране чрез потребителски сценарии**, насочени към различни групи – изследователи, институции (GLAM), публичен сектор и творчески индустрии.

Научни задачи и методи за изпълнение

Докторантурата ще започне с подробен анализ и оценка на наличните източници на данни за публично изкуство и липсващата информация в тях, включително идентифициране на релевантни източници, оценка на тяхното качество, пълнота и съвместимост, както и анализ на проблеми, свързани с фрагментация, дублиране и липса на стандартизация. Ще се очертаят и специфичните проблеми при курирането на този тип артефакти.

Въз основа на този анализ ще бъде разработен концептуален и семантичен модел на данните, който да позволява структурирано представяне на обектите на публичното изкуство, техните характеристики, връзки и контекст. Особено внимание ще бъде отделено на дефинирането на „паспорт“ на обектите като формализирана единица за описание, интеграция и обмен на информация.

Следващият етап включва разработване и експериментиране с методи за интеграция на хетерогенни данни от различни източници. Това ще обхваща подходи за съвместяване и свързване на обекти, дедупликация, както и хармонизация на структури и метаданни, с цел изграждане на последователни и взаимосвързани набори от данни. Ще бъдат изследвани и методите за пространствен анализ, картографиране и визуализация, позволяващи интерпретация на данните в техния географски и времеви контекст.

Ще се разгледат и подходите за комбиниране на геопропространствени и семантични представяния, както и за създаване на интерактивни визуални среди.

Очаквани резултати

В резултат от изследванията се очаква разработването на нови информатични методи и модели за интеграция, семантично обогатяване и анализ на хетерогенни данни за публично изкуство. По-конкретно, ще бъдат предложени:

- методи за интеграция и съвместяване на данни от различни отворени и институционални източници, включително подходи за разпознаване, свързване и дедупликация на обекти;
- концептуален и семантичен модел („паспорт“) на публичното изкуство, позволяващ стандартизирано описание, обмен и повторна употреба на данни;
- методи за комбиниран пространствен и семантичен анализ, които подпомагат изследването на публичното изкуство в неговия географски, времеви и социален контекст;
- описание и формализиране на потребителски сценарии за използване на интегрирани данни в различни области – научни изследвания, образование, културни политики и творчески индустрии;
- насоки за изграждането на прототипна софтуерна система, реализираща предложените методи и позволяваща експериментално валидиране с реални данни;
- научни публикации, представящи резултатите в областта на информатиката, дигиталното културно наследство и анализа на данни.

Въздействие

Очакваните резултати ще допринесат за развитието на цифровото културно наследство чрез създаване на нови подходи за структуриране, интеграция и анализ на данни за публично изкуство. Проектът ще подпомогне повторната употреба на данни и ще улесни достъпа до тях за различни групи потребители – изследователи, културни институции, публични администрации и граждани.

Разработените методи и модели ще създадат предпоставки за интеграция на българското културно наследство в европейски и международни инфраструктури, като EOSC и Europeana, и ще подпомогнат прилагането на принципите на отворената наука и Collections-as-Data.

От информатична гледна точка, проектът ще допринесе за развитието на методи за работа с хетерогенни и разпределени данни, като има потенциал за приложение и в други области извън културното наследство.

В по-широк план, резултатите ще подкрепят разработването на инструменти за културно планиране, туризъм и творчески индустрии, както и за по-добро разбиране и популяризиране на публичното изкуство като част от съвременната културна среда.

Място на зачисляване

Секция „Софтуерни технологии и информационни системи“ (ИМИ-БАН)

Използвана научна инфраструктура

КЛАДА-БГ

HEMUS supercomputer