

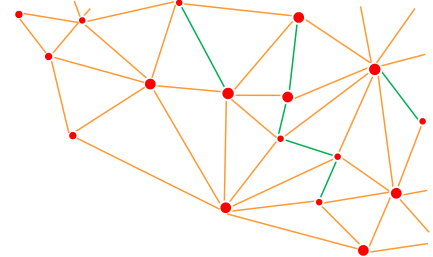
КЛАДА-БГ – ЗНАНИЯ И ТЕХНОЛОГИИ ЗА ХУМАНИТАРИСТИКАТА: ИНФОРМАЦИОННА СИСТЕМА ЗА ОПИСАНИЕ НА КОЛЕКЦИИ В МУЗЕИТЕ

Кирил Симов (kivs@bultreebank.org)

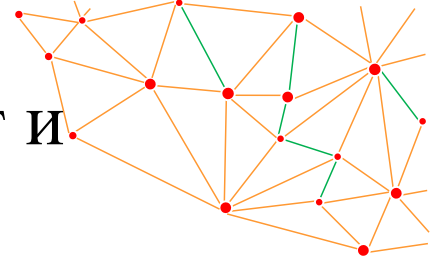
ИИКТ-БАН

Какво е КЛАДА-БГ?

- 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)
- CLaDA-BG = **CL**ARIN and **DA**RIAH in **B**ulgaria
- CLARIN EU: <https://www.clarin.eu/>
- DARIAH EU: <https://www.dariah.eu/>



Хората в КЛАДА-БГ - Изкуствен интелект и езикови технологии



Петя Осенова, Живко Ангелов, Николай Паев, Стефан Маринов, Теодор Вълчев, Ивелина Николова, Кристин Димитрова, Цветелина Александрова, Стефан Шкендеров, Десислава Нецовска, Константин Минковски, Александрина Димова, Сияна Илиева, Тихомир Алексиев, Иоана Попова, Калина Шолева, Елина Спасова, Зорница Павлова и други

Изследователски инфраструктури

Изследователските инфраструктури са сложни конструкции, включващи:

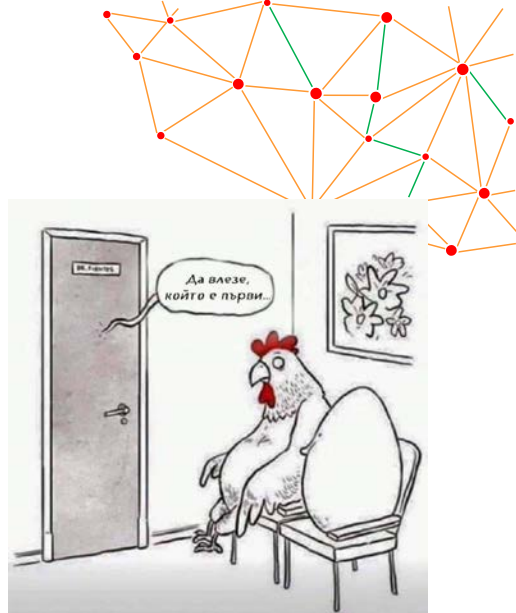
- знания,
- данни,
- хора и
- услуги

Те обединяват различни ресурси за подкрепа на **широк диапазон от потребители** и правят тези ресурси **(многократно) приложими** и **достъпни за достатъчно дълъг период**, за да **подкрепят научните изследвания** (индивидуални или съвместни) и да **споделят резултатите** от това проучване. (**FAIR principles**)

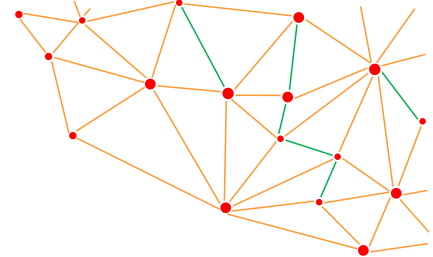
PARTHENOS Project (H2020 project of DARIAH and CLARIN)

Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

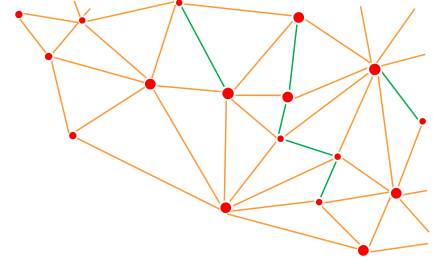


Данни и знание



- Данни – документи в различни формати и медии: текст, хипертекст, аудио, видео, изображения
- Знание – интерпретация на данните с оглед на създаване, съдържание и взаимосвързаност
- В КЛАДА-БГ приемаме, че основното количество данни са неструктурирани и не-хомогенни текстови данни: писма, протоколи, записки и други
- „Language Technology is a core Big Data technology. Growth in the volume and variety of data is mostly due to the accumulation of unstructured text data; in fact, up to 80% of all data is unstructured text data.“ <http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf> (p. 108)

Мисия на КЛаДА-БГ

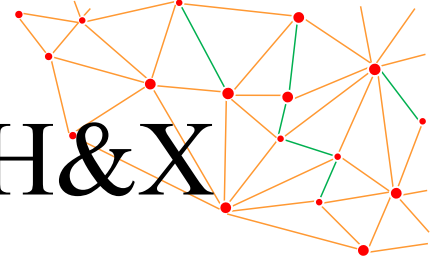


- Създаване на виртуална инфраструктура за езикови технологии, културно и историческо наследство
- Езикови технологии за крайния потребител (изследователи, учители и други)
- Семантични технологии за моделиране на артефакти
- Отворени свързани данни (Linked Open Data), езикови технологии, културно и историческо наследство
- 3D технологии за сканиране, моделиране, визуализация на културни и исторически артефакти, мултимодални данни
- **Основна цел: подкрепа на изследователската дейност в областта на социалните науки и хуманитаристиката (SH&X)**
- **Прилагане и адаптация на създадените ресурси и технологии към нови области, приложения и задачи**

Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

Изследователски процес в СН&Х

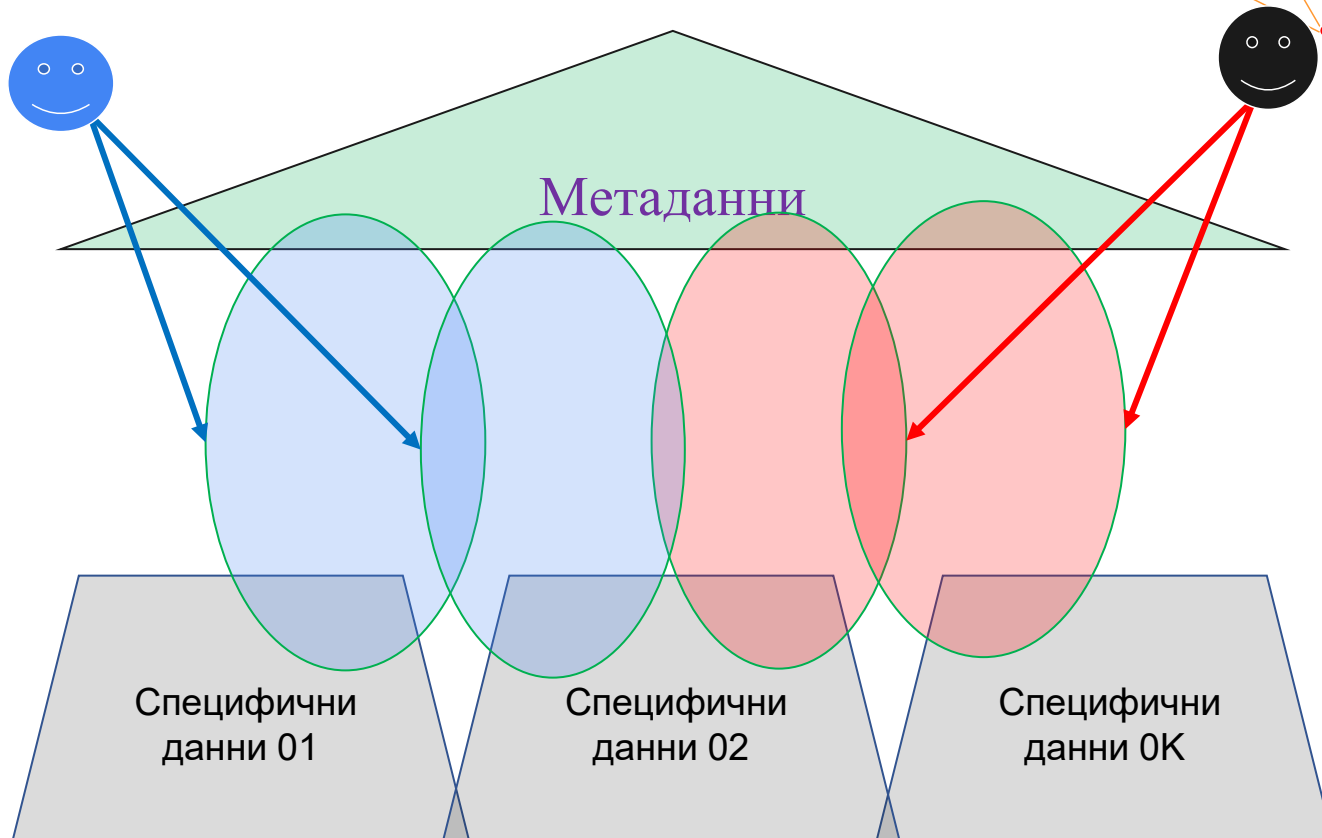


- Огромно многообразие от изследователски обекти
- Тези обекти съдържат информация за едни и същи същности, но не са свързани
- Изследователският процес в СН&Х изисква поставянето на тази информация в общ контекст
- Процесът на интеграция на данните *контекстуализация*

Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

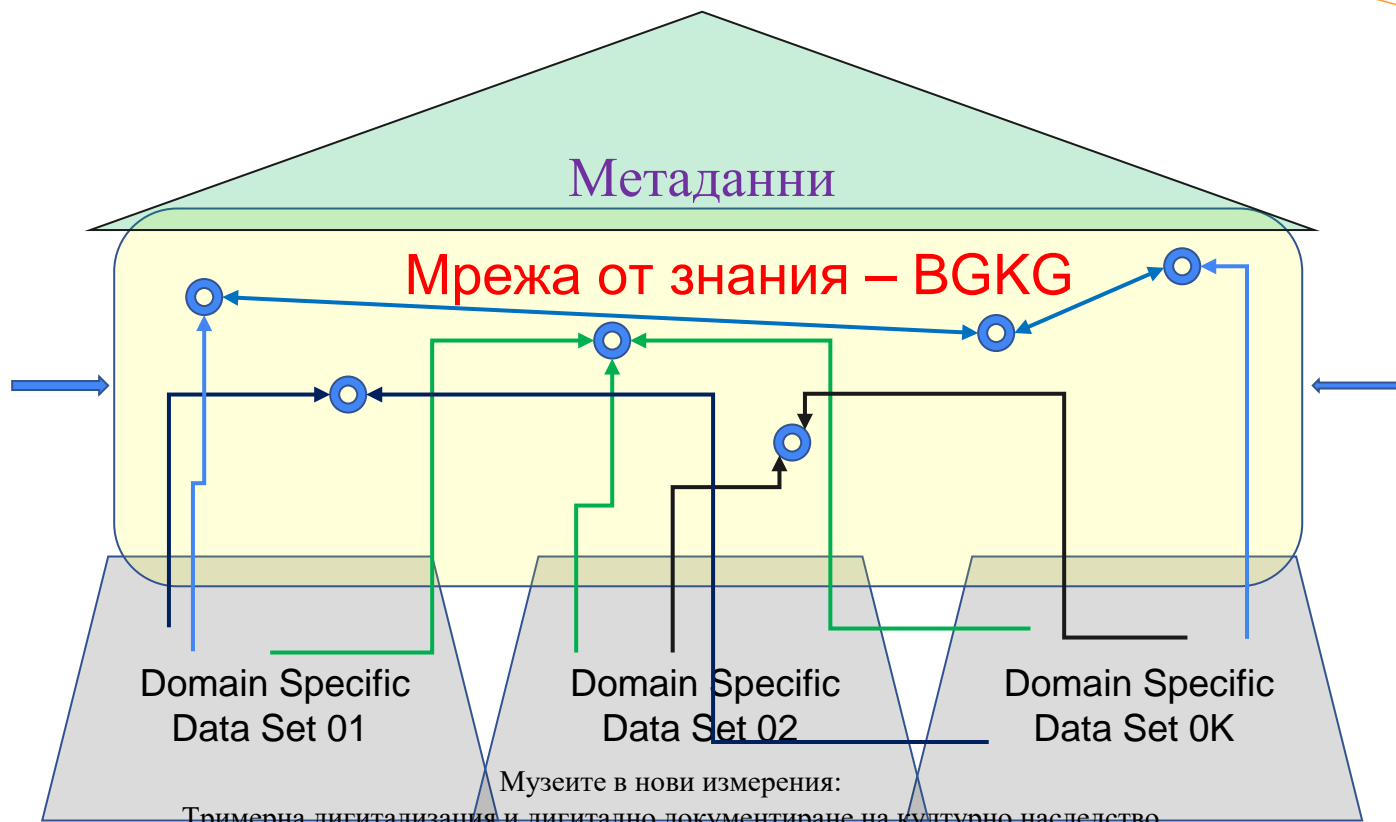
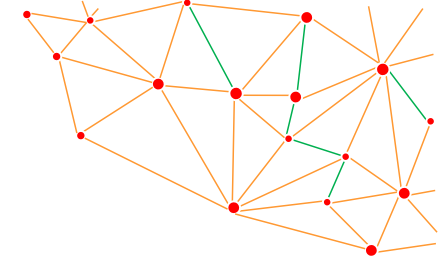
Изследователски процес в CH&X



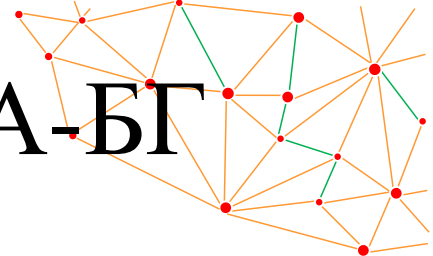
Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

Изследователски процес в СН&Х



Контекстуализацията в КЛаДА-БГ



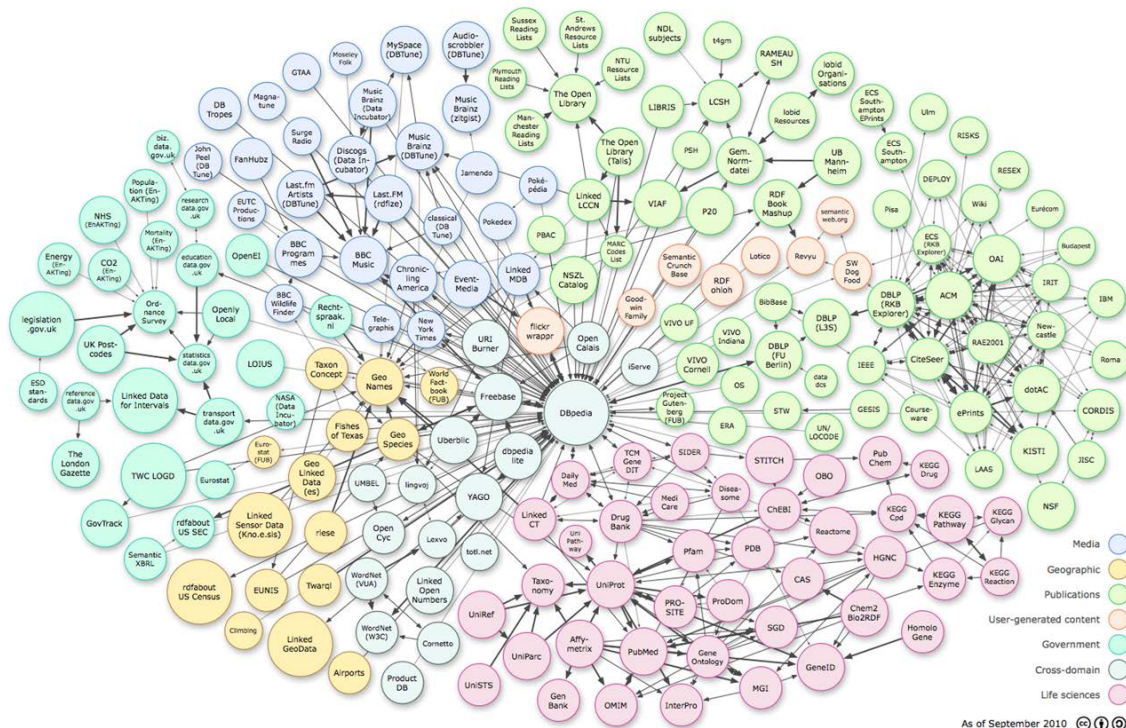
Контекстуализацията в рамките на КЛаДА-БГ се разглежда като **Българска мрежа за взаимосвързани знания (Bulgaria-centric Knowledge Graph – BGKG)** за хора, събития, обекти, документи, артефакти, автори, мнения и други

- **Хора** – техните биографии – събития в живота им, роли
- **Обекти** – географски, артефакти и други
- **Събития** – място, време и участници, връзки с други събития
- **Документи** – автори, съдържание, мнения и отношение към хора, събития, обекти – **колко надеждно е представеното знание**
- **Текстът като основен източник на информация**
- **Технология** – свързани бази от отворени данни (Linked Open Data) – мрежа знания (Knowledge graph)

Музеите в нови измерения:

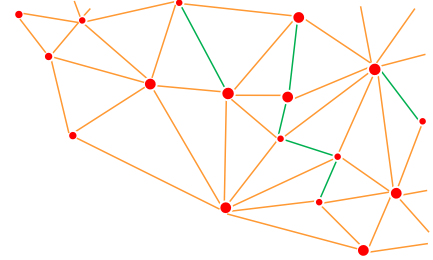
Тримерна дигитализация и дигитално документиране на културно наследство

Облак от мрежи от знания



Музете в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство



Българска мрежа от знания

Изграждането на Българската мрежа от знания (Bulgaria-centric Knowledge Graph – **BGKG**) се извършват следните дейности:

- Формализация на знанието – избор на подходящи онтологии и мрежи от отворени данни
- Формализиране на данни от партньорите в КЛАДА-БГ и други организации
- Терминология и ключови думи и изрази
- Библиографска информация
- Анотирани и описани изображения
- Основен инструмент – анализ на текст

Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

Резултати от проекти

Къща №	Домакинство №	ФАМИЛИЯ	Име	Възраст	Родство	Професия	Грамотност	Бележка	Общ брой
ЦДА,	ф. 246К, оп.	9, а. е. 524, л. 1–185. Печатно.	Ръкопис.						
1	1	МАНДИЧЕВ	Георгий	33	баща	свещеник	Да		1.
			Илина	27	майка	-	Да		1.
			Иван	9	син	ученик	Да		1.
			Асен	7	син	ученик	Да		1.
			Любица	4	дъщеря	-	Не		1.
			Йосиф	1	син	-	Не		1.
2	2	ДИМИТРОВ	Божин	33	баща	учител	Да		1.
			Траянка	29	майка	-	Да		1.
			Лазар	8	син	ученик	Не		1.
			Димитър	6	син	ученик	не		1.
			Кирил	1	син	-	-		1.
	3	ПОП ТАШОВА	Магдалина	55	майка-вдовица	уч[илищна] слугиня	не		1.
			Тома	30	син	надничар	не		1.
3	4	БЪЧВАРОВ	Димитър	45	баща	готвач	да		1.
			Султана	33	майка	-	да		1.
			Георгий	10	син	ученик	не		1.
			Марика	8	дъщеря	ученичка	да		1.

Музеите в нови измерения:

15

Публикувани книги

ДИМИТЪР ГЕНКОВ ФИЛОВ

(Роден в 1847 г. — Калофер; убит в 1887 г. — Русе)

Един от първите офицери, получили до Освобождението военно образование в Русия. Участник в Руско-турската война (1877/78) в Българското опълчение. След Освобождението на служба в Източнорумелийската милиция; участва в подготовката и провъзгласяването на Съединението; член на „Привременното правителство“. Като военен комендант на Русе взема участие в Русенския бунт (1887), когато бил убит.

Източници

За фондообразователя: Анчев, А. Българското опълчение. С., 1968, с. 16; Бендерев, Ан. История на Българското опълчение и Освобождението на България 1877—1878. С., 1930, с. 79; Дренски, Ив. Олимпий Панов. С., 1964, с. 126; Начов, Н. Калофер в миналото. С., 1927 492—493; Радев, С. Строителите... Т. 2, 1911, 696—718; НБКМ — БИА, ф. 209, арх. ед. 40 (Вж. Обзор на архивните фондове, колекции и единични постъпления, съхранявани в Български исторически архив. Кн. 4. С., 1973, 91—99).

Исторически корпус от издания в периода 1800-1950+ от Националната библиотека “Иван Вазов” – Пловдив

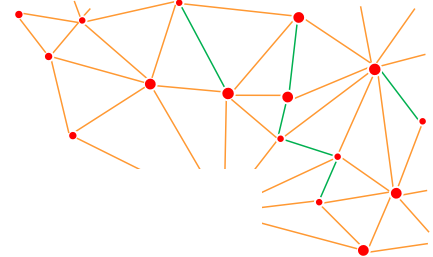
Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство



Архивни документи

АЕИМ № 859 – II



ЕТНОГРАФСКИ ИНСТИТУТ И МУЗЕЙ ПРИ БАН – СОФИЯ

ТЕМА: ТЪРГОВИЯ, ТРАНСПОРТ, ТРУДОВО-ПОМИНЪЧНИ ПЕСНИ
ОТ ПЛОВДИВСКИ ОКРЪГ
АВТОР: АНАСТАС ПРИМОВСКИ

ПРЕДАДЕНО:

СЪДЪРЖА 153 ЛИСТА

Анотирани и описани изображения

Наименование на местонахождението на произведението

Гр. Пазарджик, църквата „Св. Богородица“.

Автор и година | Школа, век

Ст. Доспевски, XIX в.

Наименование на творбата, година

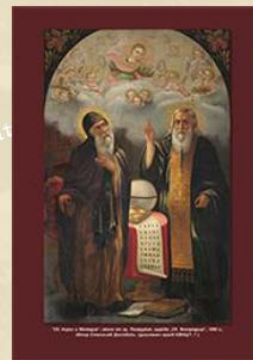
„Св. Кирил и Методий“, икона, 1860 г.

Техника, материал

Темпера върху дърво.

Размер

... / 187 см.



Подрис, година

Иконата е подрисвана в долния край, по средата, посочена е и година на написване – 1860 г. „Произведение на руското художника Г. Ставислава Доспевскогъ въ Самоковъ 1860“

Надпис, печат, стари номера

Процход

Инвентарен номер

Местонахождение на творбата в галерията
Експозиция – зала

Хранилище

Клетна

Описание

Иконата се явява в явност на църквата, на границата на църковния и гражданския живот. Експонирана е на самостоятелна поставка, богато орнаментирана. Кирил и Методий са изобразени в цял ръст, обръщати напред към зрителя. Св. Кирил е разположен в дясната част на иконата, обличен в монашеско облекло. Св. Методий е отляво, в архиерейски одежди с епитрима. С дясната си ръка посочва нагоре, където Христос Емануил е сред облаци и ангели. Кирил и Методий са разположени в интериор, между тях е изобразена книга, малка кръгла маса, покрита с червена покривка, на нея са поставени свещи с златни, сини и глауби. На пода под масата има още едни свещи и две вазни. Зад гърба на св. Кирил се виждат полупи с книги и един стол с орнаментална украса и полта. В горния край иконата завършва с арка, а в ъглите има растителен орнамент.

Състояние

Отлично.

Реставрация

Библиография

Божков, Ат. Изобразителна на Кирил и Методий през вековете. С., 1989, 186, обр. 128.
Василев, Ас. Български възрожденски майстори. С. 1965, 405.
Василев, Ас. Обрътите на Кирил и Методий в България. С., 1970, 17, 114.
Григорова, Т., НЕМУ ДА БЪДЕ СЛАВА, ЧЕСТ И ПОЧУТ. – (Изложба и част на 1150-годишнината от Успението на св. Константин-Кирил Философ. В: Рилският парк. Старобългаристика. кн. 4, 2019, 119-131.
Енциклопедия на изобразителните изкуства в България. Т. 1. С., 1980, 262-263.
Мавродиев, Н. Изкуството на българското Възраждане. С., 1957, 304.
Чешмеджиев, Д. Чешмеджиев, Д. Изобразителни святски Кирила и Мефодия в изкустве епохи Българското възрождение. В: В началото было слово... Москва-Санкт-Петербург, 2013, 57.

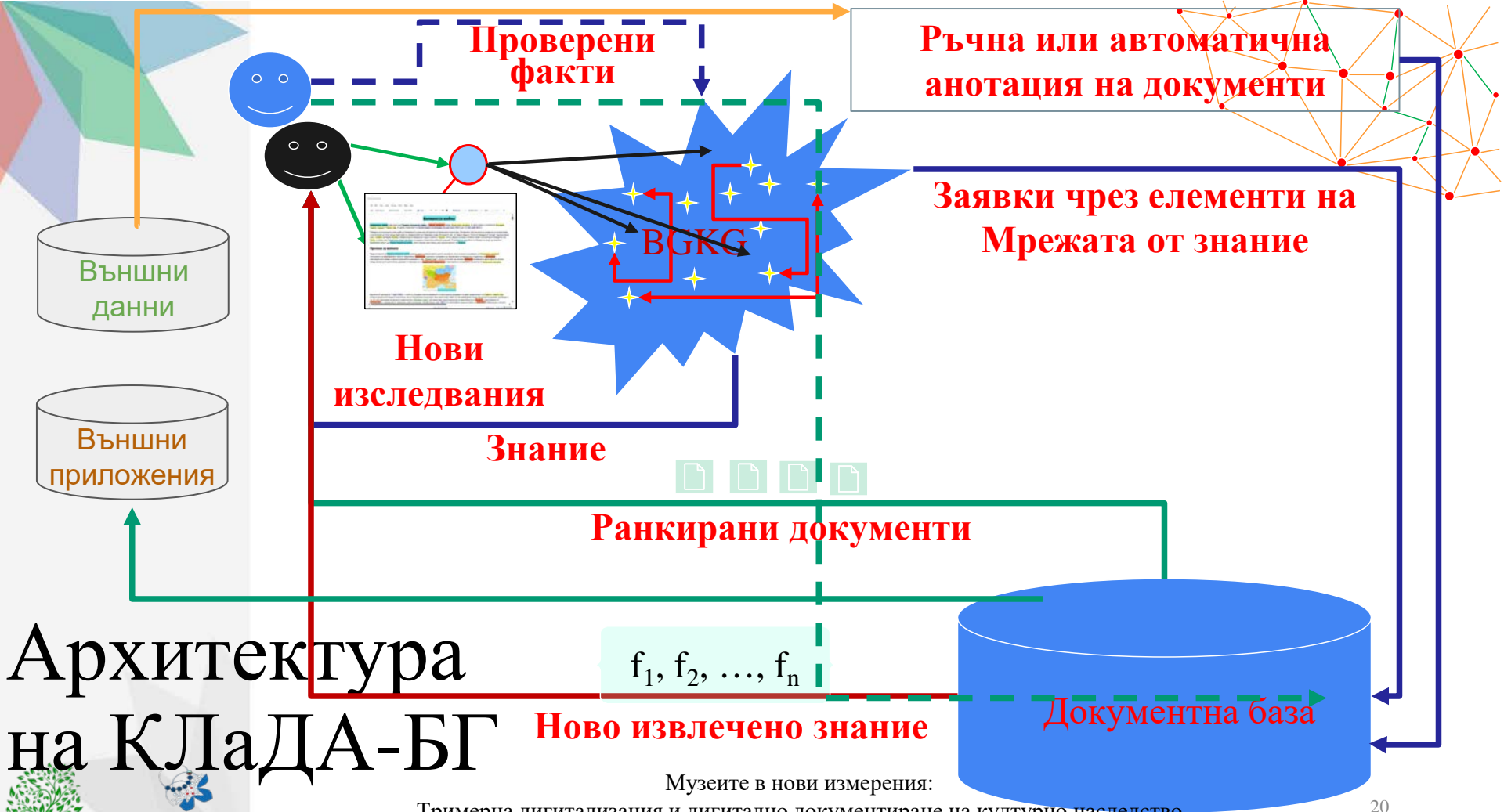
Полца

Дата на актуализация

18 г.

Музеите в нови измерения:

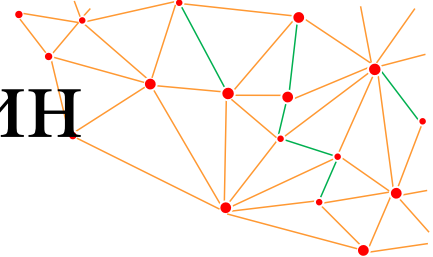
Примерна дигитализация и дигитално документиране на културно наследство



Архитектура на КЛаДА-БГ

Музеите в нови измерения:
 Примерна дигитализация и дигитално документиране на културно наследство

Приложение в рамките на един музей



- Създаване на колекции от артефакти
- Създаване на архиви с документи
- Взаимосвързаност на създадените колекции в рамките на музея и извън него
- Взаимосвързаност на документите и други информационни обекти
- Документиране на колекцията
- Разпространение, използване и разширения

Създаване на колекции от артефакти

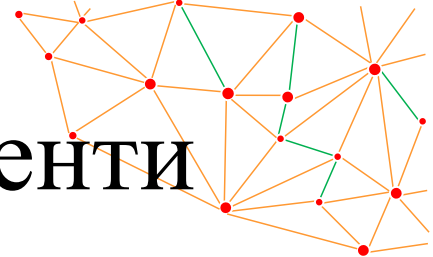
Дигитализация на артефактите:

- създаване на цифрови версии на съответните предмети - двумерно и тримерно заснемане, ако колекцията е с артефакти от материалното културно наследство. Създаване на подходящ аудио- и/или визуален запис
- документиране на артефактите в текстови описания

Създаване на колекцията:

- концептуализация на артефактите - формализиране на терминологията, използвана за класификация на артефактите
 - класифициране на артефактите спрямо концептуализацията
 - свързване с външни обекти - материали на музея; елементи на Мрежата знание на КЛАДА-БГ; материали на други музеи, библиотеки и други
 - записване в съответен софтуер и публикуване чрез този софтуер или по друг начин
- Тримерна дигитализация и дигитално документиране на културно наследство

Създаване на архиви с документи



Дигитализация на документите:

- двумерно сканиране на документите
- OCR а сканираните документи - проверка на качеството
- описание на документите в метаданни и съдържание

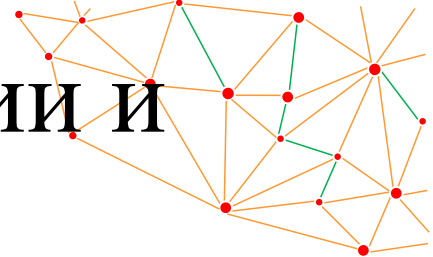
Създаване на колекцията:

- форматиране на съдържанието на документа
- аотиране на документите с лингвистична и фактологична информация
- зареждане в различни (унифицирани) търсачки над (анотиран) текст
- документиране на архива
- публикуване на архива

Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

Взаимосвързаност на колекции и архиви



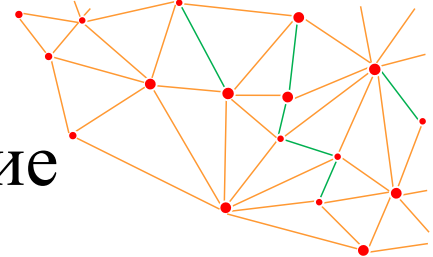
Свързаността на колекциите и документите с други информационни обекти като Мрежи от думи, Мрежи от знания, други колекции и архиви подобрява стойността (качеството) на самите колекции и архиви

Много от информацията, която е необходима за разбирането на представените обекти (артефакти и документи), вече е добре описана на друго място и може да се използва наготово

Свързаността може да се добави ръчно по време на създаването на цифровия вариант на дадения обект или автоматично с помощта на големи езикови модели (наричани още изкуствен интелект)

Основни елементи на свързването е разпознаване на наименовани същности и терминология

Езикови технологии за извличане на знание



КЛАДА-БГ трябва да предостави езикови технологии за:

- Анотиране и свързване на понятия/ключови думи (пее)е)
- Анотиране и свързване на наименовани обекти (Лили Иванова)
- Анотиране и свързване на събития (фактът: кой къде е пял)

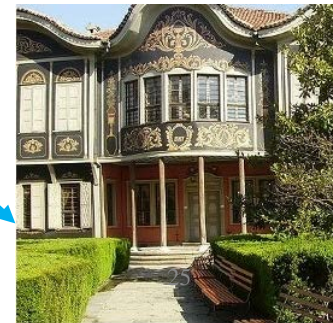
Свързването е идентификация на едни и същи същности в различни текстове и между текст и знание в мрежата от знания - разширяване на съдържанието на всеки информационен обект



Лили Иванова

пя на

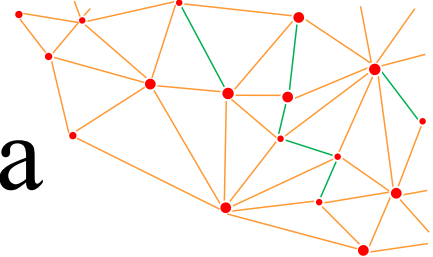
амфитеатъра в Пловдив



Музеите в нови измерения:

Тримерна дигитализация и дигитално документиране на културно наследство

Документиране на колекцията

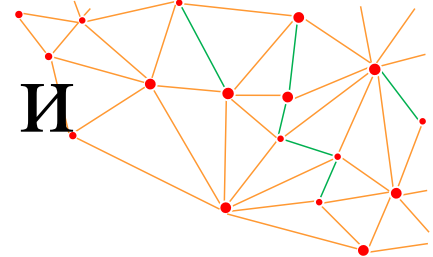


Метаданни за самата колекция или архив - какъв е произходът на първичните обекти; как е направена самата колекция или архив

Какви права за достъп са дефинирани за дадената колекция или архив: всеобщ достъп с права за използване в други колекции, архиви или приложения; ограничен достъп до група от потребители; ограничен достъп според вида на използване

Описание на добри практики на използване

Разпространение, използване и разширения

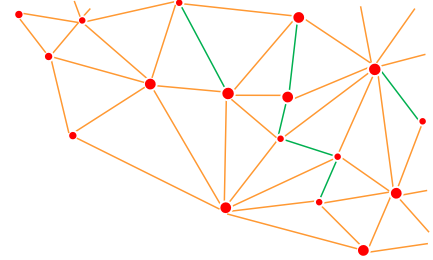


КЛАДА-БГ предлага разпространение на информацията за дадена колекция или архив в рамките на самата инфраструктура (включително и в чужбина) с основно приложение при извършване на изследователска дейност

Ако притежателят на колекцията/архива допусне, може да се даде възможност за изтегляне на колекцията/архива, след свързване на потребителя с притежателя

КЛАДА-БГ предоставя механизми за достъп до отделните елементи на колекцията/архива с цел изграждане на уеб сайтове

КЛАДА-БГ предоставя единен механизъм за цитирания на колекциите/архивите



Приложения в даден музей

КЛАДА-БГ е отворена за съвместни дейности за подпомагане на дейностите по дигитализация на музейните фондове

Основната цел би била да се построи Мрежа от знания на музея по отношение на неговите фондове, архиви и човешки потенциал

Една такава Мрежа от знания би отразявала местната история, култура, етнография и документи, свързани с мрежата от знания на КЛАДА-БГ

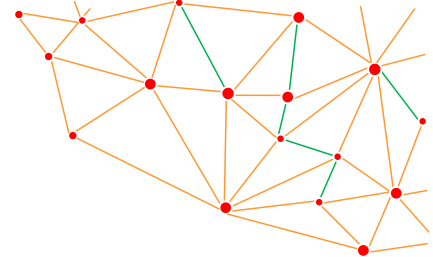
Това би допринесло за видимостта на музея по отношение на по-широк кръг от артефакти и документи и по-голямо използване в изследванията в тези области

Използване на изкуствения интелект

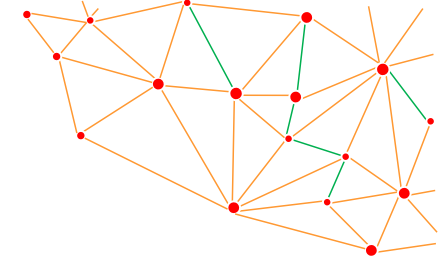
Освен за анализ на текст, изкуственият интелект в КЛАДА-БГ ще бъде използван в следните дейности:

- подобряване на резултата от OCR и извличане на текст от PDF документи
- автоматичен преход от стар към нов правопис
- грешкоуловител
- търсене в колекции и архиви
- отговори на въпроси
- диалогови системи

Заключение



- Представена беше мотивацията и елементите на системата за създаване, управление и използване на колекции от документи, описания на артефакти и натрупването на знание
- Тази система позволява създаването, управлението и използването на колекции и архиви в музеи
- В следващата презентация ще бъдат представени някои съществуващи приложения на системата



Благодаря за вниманието!