

Инфраструктурният проект КЛАДА-БГ: мрежа от свързани отворени знания

Кирил Симов

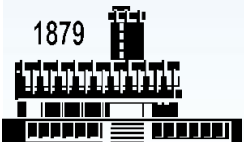
Институт по информационни и комуникационни технологии – БАН

Димитър Минев

Народна библиотека „Иван Вазов“ – Пловдив

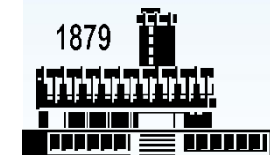
XXX Национална конференция на ББИА

„Библиотеките в дигиталната епоха“, София, 15 – 16 октомври 2020 г.



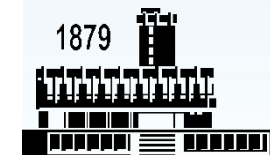
Съдържание

- Какво е КЛАДА-БГ?
- Дефиниция на изследователска инфраструктура
- Българска мрежа от знания
- Езикови ресурси за извличане на знание от текст
- Мрежата – среща в библиотеките
- Заключение



Какво е КЛаДА-БГ?

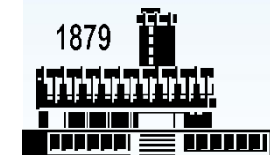
- Национална интердисциплинарна изследователска Е-инфраструктура за ресурси и технологии за българското езиково и културно наследство, интегрирана в рамките на европейските инфраструктури CLARIN и DARIAH ([КЛаДА-БГ](#))
- CLaDA = **CL**ARIN and **D**ARIAH in **B**ulgaria
- CLARIN EU: <https://www.clarin.eu/>
- DARIAH EU: <https://www.dariah.eu/>



Консорциум на КЛАДА-БГ

Технологични партньори

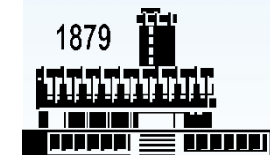
- Институт по информационни и комуникационни технологии – БАН (ИИКТ-БАН)
- Институт по математика и информатика – БАН (ИМИ-БАН)
- Софийски университет „Св. Климент Охридски“ (СУ)
- Нов български университет (НБУ)
- Шуменски университет „Епископ Константин Преславски“ (ШУ)
- Сдружение „Булгариана“ (БУЛГ)
- Онтотекст АД (ОНТО)



Консорциум на КЛАДА-БГ

Партньори, предоставящи съдържание и експертиза

- Кирило-Методиевски научен център – БАН (КМНЦ–БАН)
- Институт по балканистика с център по тракология – БАН (ИБЦТ–БАН)
- Институт по етнология и фолклор с етнографски музей – БАН (ИЕФЕМ–БАН)
- Народна библиотека „Иван Вазов“ – Пловдив (НБИВ-Пловдив)
- Регионален исторически музей – София (РИМС)
- Югозападен университет „Неофит Рилски“ (ЮЗУ)
- Бургаски свободен университет (БСУ)
- Сирма медия (СМ)



Изследователски инфраструктури

Изследователските инфраструктури са сложни системи, включващи:

- знания
- данни
- хора
- 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 тях.

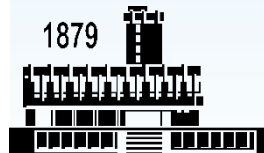
Ключов елемент е **контекстуализацията** на елементите на инфраструктурата.

Контекстуализация в КЛаДА-БГ

Контекстуализацията в рамките на КЛаДА-БГ се разглежда като мрежа за взаимосвързани описания на 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
- Текстът като основен източник на информация
- Технология – свързани бази от отворени данни (**Linked Open Data**) – граф знания (**Knowledge graph**)

Цел: създаване на Мрежа от знания за България



Езикови ресурси за извличане на знание

КЛаДА-БГ трябва да предостави езикови технологии за:

- Извличане и свързване на понятия/ключови думи
- Извличане и свързване на наименовани обекти
- Извличане и свързване на събития

Свързването е идентификация на едни и същи същности в различни текстове и между текст и знание в мрежата от знания



Лили Иванова

пя на

амфитеатъра в Пловдив



Мрежата от знания и библиотеките

- Библиотеките в своята същност представляват мрежи от знания
- КЛАДА-БГ може да предостави средствата тези мрежи да се развиват в дигиталния свят като се формализират не само метаданните за притежанията и свързването им с описание извън рамките на конкретната библиотечна колекция, **НО**
- да бъдат свързани на базата на **съдържанието** на библиотечните документи.

Напр: **Царят участва в празника на Съединението в Пловдив.**

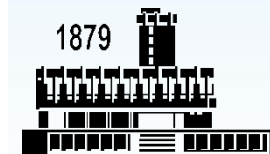
Фердинанд I отпразнува Съединението в града на седемте тепета.

или

Свързване на едно и също **събитие**, **човек** и **място**, назависимо, че компонентите присъства в различни документални източници и в различна корелация помежду си.

Заклучение

- КЛаДА-БГ има за цел да създаде среда за изследвания в областта на социалните науки и хуманитаристиката.
- Семантичното представяне под формата на онтологии и мрежа от знания.
- Езикови технологии като механизъм за извличане на знание и достъп до него.
- Услугите на КЛаДА-БГ ще подпомогнат да се експлицира знанието, съхранявано в библиотеки като НБИВ-Пловдив, и неговото интегриране със знание от други източници.



Благодаря за вниманието.

