

Београд
Филолошки факултет
Универзитет у Београду
09.04.2024.

Увод у дигиталну хуманистику



САДРЖАЈ

- Да ли дигитална хуманистика и у којој мери може утицати на формирање и/или деконструкцију канона?
- Контекст COST акције Distant Reading
- SrpELTeC колекција и подржани пројекти у оквиру ње

Бренд конотација: Canon EOS series



Војна конотација: „Tsar Cannon“, Moscow, 1586



Религијска конотација



Музичка конотација

9

Musical score for measures 9-13. The score is written for four staves. The first staff (treble clef) contains the melody: measure 9 (quarter notes G4, A4, B4, C5), measure 10 (quarter notes G4, A4, B4, C5), measure 11 (quarter notes G4, A4, B4, C5), measure 12 (quarter notes G4, A4, B4, C5), and measure 13 (quarter notes G4, A4, B4, C5). The second staff (treble clef) has rests in measures 9 and 10, then quarter notes G4, A4, B4, C5 in measure 11, quarter notes G4, A4, B4, C5 in measure 12, and quarter notes G4, A4, B4, C5 in measure 13. The third and fourth staves (treble clefs) have rests in measures 9 and 10, then quarter notes G4, A4, B4, C5 in measure 11, quarter notes G4, A4, B4, C5 in measure 12, and quarter notes G4, A4, B4, C5 in measure 13.

14

Musical score for measures 14-17. The score is written for four staves. The first staff (treble clef) contains the melody: measure 14 (quarter notes G4, A4, B4, C5), measure 15 (quarter notes G4, A4, B4, C5), measure 16 (quarter notes G4, A4, B4, C5), and measure 17 (quarter notes G4, A4, B4, C5). The second staff (treble clef) has quarter notes G4, A4, B4, C5 in measure 14, quarter notes G4, A4, B4, C5 in measure 15, quarter notes G4, A4, B4, C5 in measure 16, and quarter notes G4, A4, B4, C5 in measure 17. The third staff (treble clef) has quarter notes G4, A4, B4, C5 in measure 14, quarter notes G4, A4, B4, C5 in measure 15, quarter notes G4, A4, B4, C5 in measure 16, and quarter notes G4, A4, B4, C5 in measure 17. The fourth staff (treble clef) has rests in measure 14, quarter notes G4, A4, B4, C5 in measure 15, quarter notes G4, A4, B4, C5 in measure 16, and quarter notes G4, A4, B4, C5 in measure 17.

КАНОН(ИЗАЦИЈА) ДАНАС

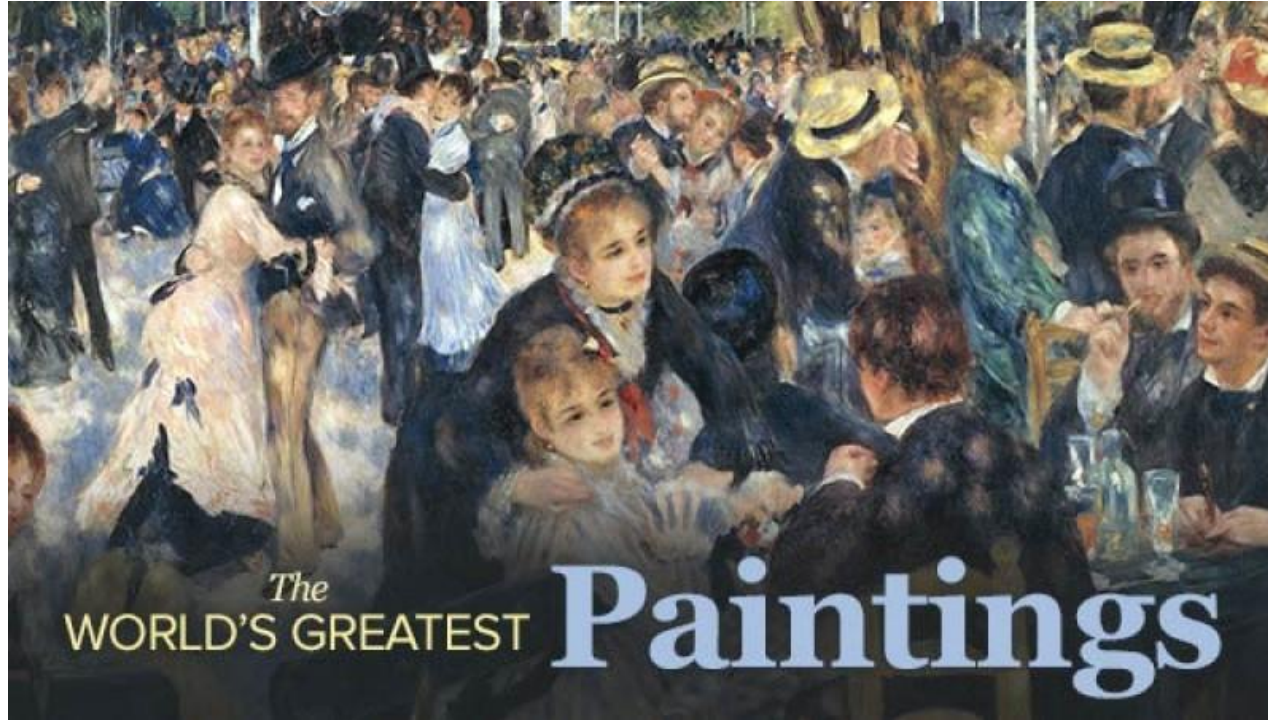
- Религија
- Култура (Уметност)

КАНОН(ИЗАЦИЈА) У РЕЛИГИЈИ

- Корпус „светих“ текстова одобрених за употребу
- Инструкције за религиозни ритуал/Службу
- Правни систем Цркве: Канонско право
- Процедуре за канонизацију свеца

КАНОН(ИЗАЦИЈА) У КУЛТУРИ (УМЕТНОСТИ)

- Корпус егземпларних културних артефаката/уметничких дела(music, literature, painting ...);
- Концепт антологија / одабира итд.



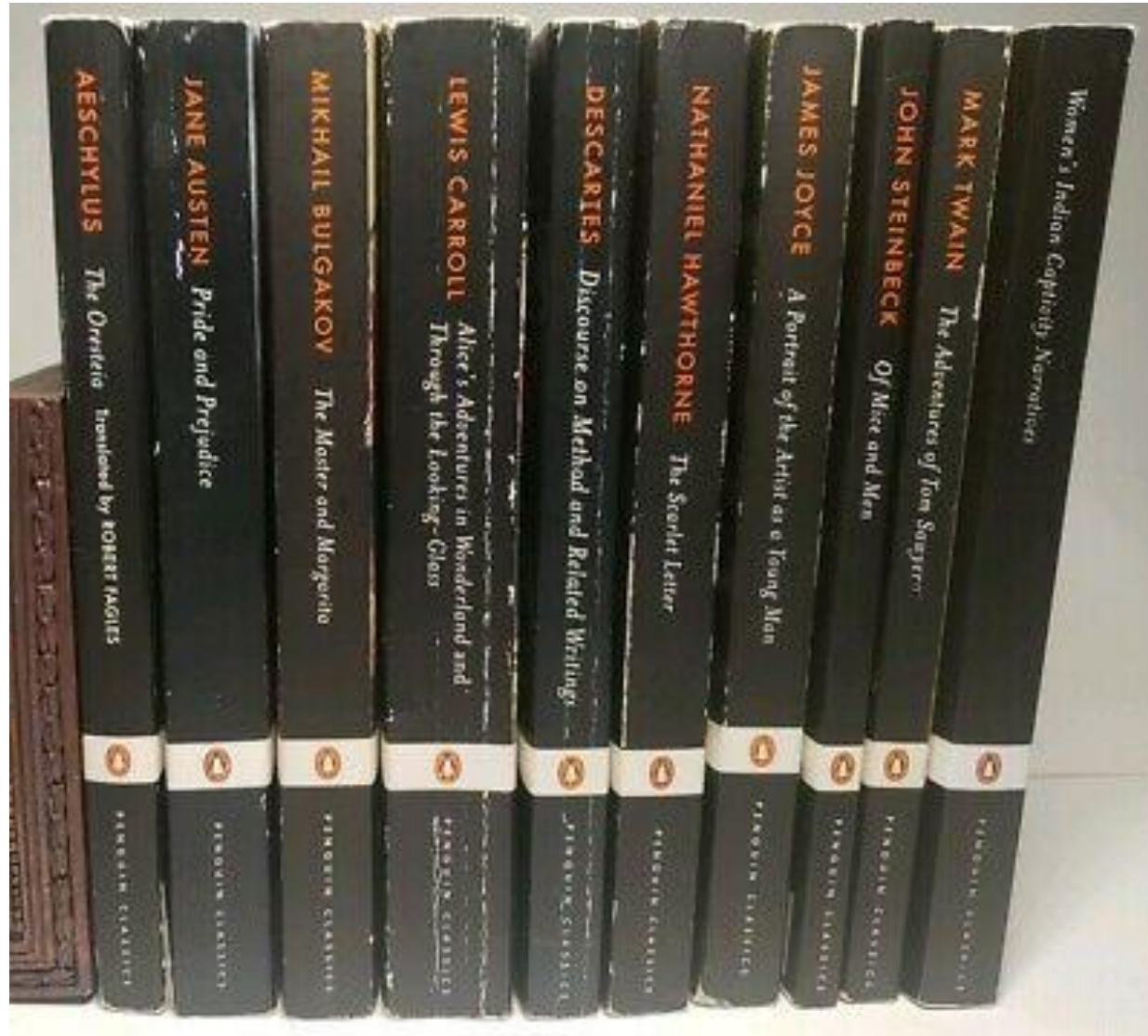
КАНОН(ИЗАЦИЈА) У КУЛТУРИ (УМЕТНОСТИ)

- Селекција егземпларних уметника

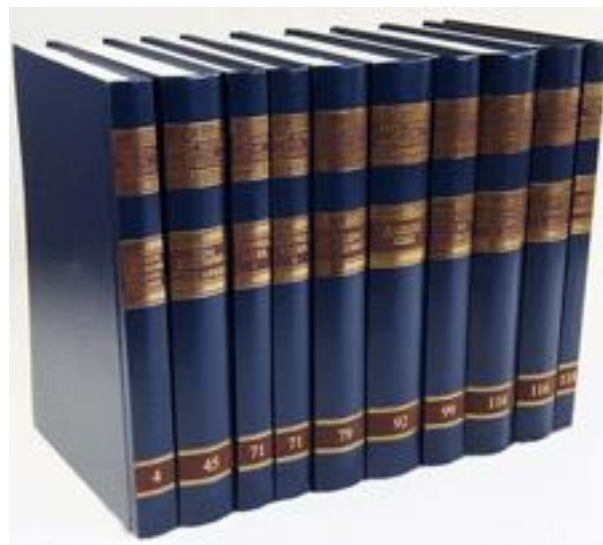
Giorgio Vasari: „Животи славних италијанских
архитекти, сликара и вајара“, 1550



Penguin, „Black Classics“ (1961-)



ДЕСЕТ ВЕКОВА СРПСКЕ КЊИЖЕВНОСТИ



МОДЕРНО ДОБА: традиција је жива!

T. S. Eliot, „What is a Classic“?, 1944/1957

„You will have anticipated the conclusion towards which I have been drawing: that those **qualities** of the classic which I have so far mentioned – **maturity of mind, maturity of manners, maturity of language and perfection of the common style** – are most nearly to be illustrated, in English literature, in the eighteenth century; and, in poetry, most in the poetry of Pope.“

Harold Bloom, *The Western Canon*, 1994

THIS BOOK studies twenty-six writers, necessarily with a certain nostalgia, since I seek to isolate the qualities that made these authors canonical, that is, authoritative in our culture. “Aesthetic value” is sometimes regarded as a suggestion of Immanuel Kant’s rather than an actuality, but that has not been my experience during a lifetime of reading. Things have however fallen apart, the center has not held, and mere anarchy is in the process of being unleashed upon what used to be called “the learned world.”

- „Изоловати квалитете“ који чине (26) аутора канонским
- Због „aesthetic value“ / против „anarchy“

Jay Stevenson, *The Complete Idiot's Guide to English Literature*, 2007

- [In] the postmodern period [...] [t]raditional literature has been found to have been written by "**dead white males**" to serve the *ideological* aims of a conservative and repressive Anglo *hegemony* [...] In an array of reactions against the **race, gender, and class biases** found to be woven into the tradition of Anglo lit, **multicultural writers and political literary theorists** have sought to expose, resist, and redress injustices and prejudices.

David Damrosch, „Frames for World Literature“, 2009

In place of this older, two-tiered model, our new system has three levels: a *hypercanon*, a *counter-canon*, and a *shadow canon*. The hypercanon is populated by the older ›major‹ authors who have held their own or even gained ground over the past twenty years. The counter-canon is composed of the subaltern and contestatory voices of writers in less-commonly-taught languages and in minor literatures within great-power languages. Many, even most, of the old ›major‹ authors coexist quite comfortably with these new arrivals to the neighborhood, very few of whom have yet accumulated anything like their fund of cultural capital. In the field of world literature today, it is the old ›minor‹ authors who fade increasingly into the background, becoming a sort of shadow canon that the older scholarly generation still knows (or, increasingly, remembers fondly from long-ago reading), but whom the younger generations of students and scholars encounter less and less. Kafka

David Damrosch, „Frames for World Literature“, 2009

- **(Традиционални) канон светске књижевности**

 - Главни канонски аутори

 - Секундарни канонски аутори

- **Канон у „постканонско“ доба**

 - Хиперканон (главни аутори који су прошли тест времена)

 - Контра-канон (корективни фактори канона, авангарда, женско писмо, мањине ...)

 - Канон сенки (аутори који временом „бледе“)

Увод у дигиталну хуманистику

- Дигиталне хуманистичке науке су интердисциплинарни скуп области које се првенствено баве коришћењем дигиталних технологија, извора и метода у оквиру истраживања у хуманистичким наукама.
- ***Diskurzivni preokret iz humanističkog računarstva u digitalnu humanistiku je obilježen diskurzom vizionarstva i dalekovidnosti, a digitalna humanistika je postala neka vrsta laboratorije i sredstva za promišljanje budućnosti ne samo tog polja nego humanistike uopšte. 2012.***

УБСМ – Београд; ФИЛУМ - Крагујевац

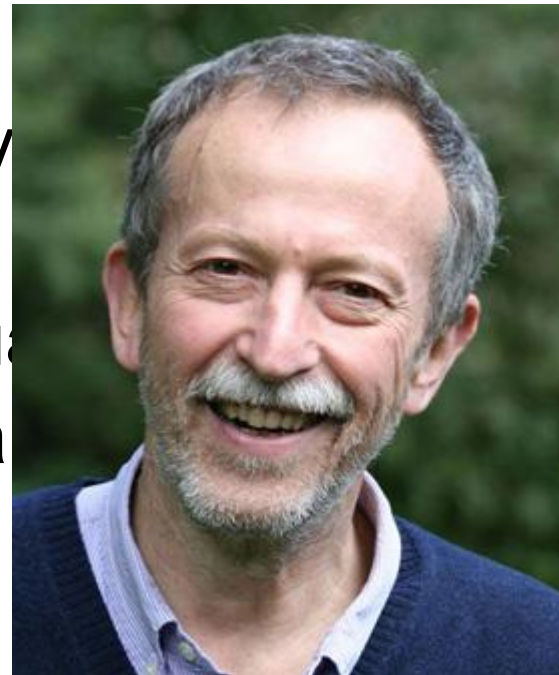


Школа Дигиталне хуманистике у Тршићу



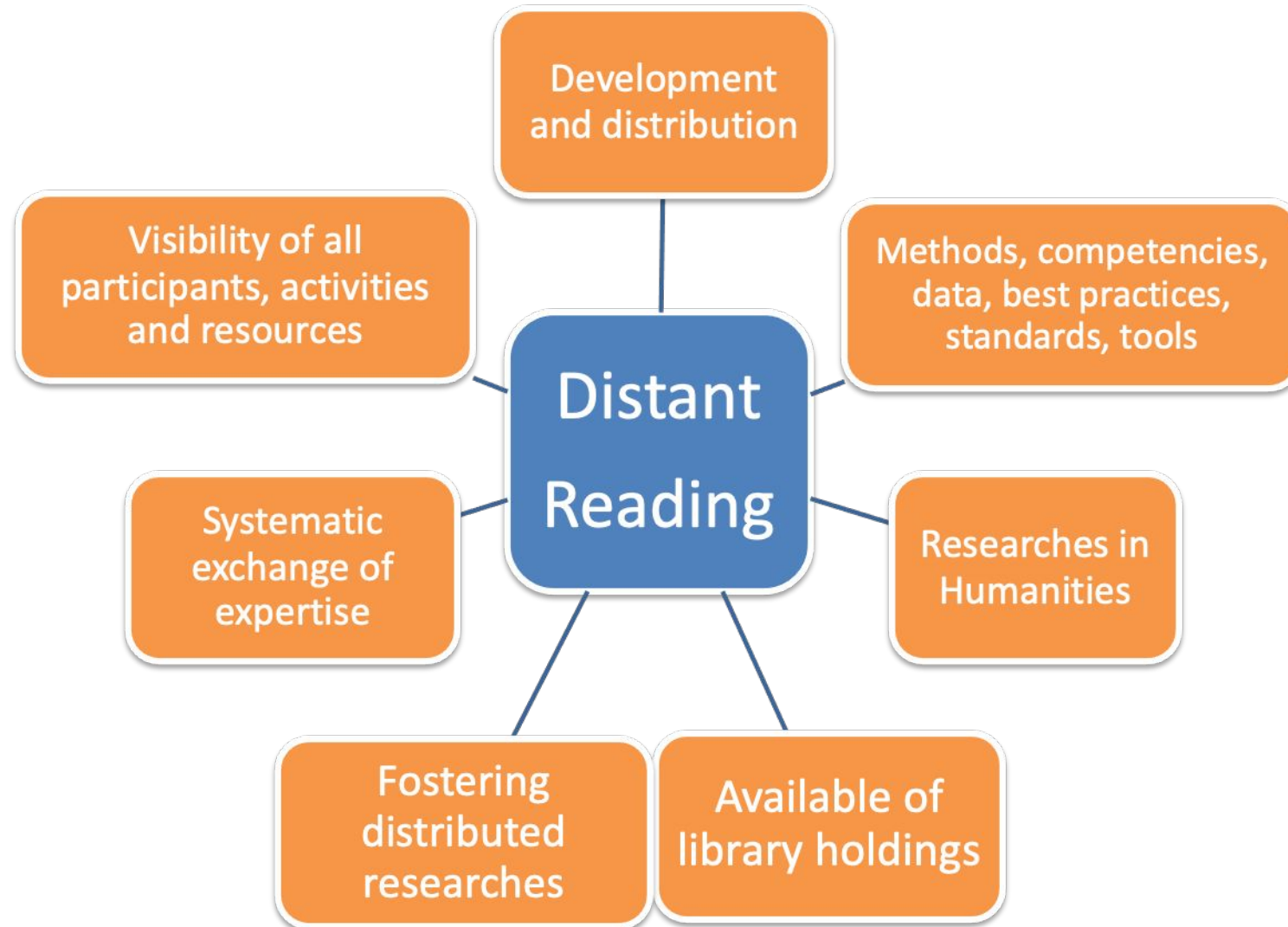
Увод у дигиталну хуманистику

- Distant Reading – Franco Moretti
- односе се на процес „разумевања литературе не проучавањем одређених текстова, већ прикупљањем и анализом огромних количина података“. Концепт је био у центру разговора о подацима у хуманистичким наукама.



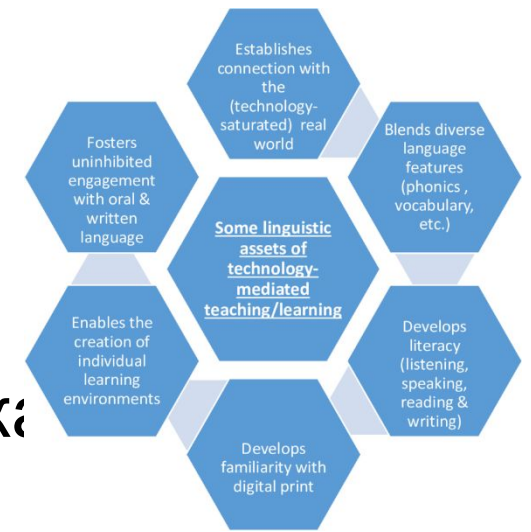
Distant Reading Paradigm - коришћење машинских (рачунарских) метода за анализу великих збирки књижевних текстова

Distant Reading Paradigm



Значај језичких технологија

- Данас су доступни различити модели због напретка у машинском и дубоком учењу, који боље могу да обрађују и трансформишу текст на природном језику.
- За „велике језике“ доступни су моћни алати и апликације (аутоматско превођење, анализа ставова и осећања, генерисање текста и одговарање на питања, мерење задовољства корисника, класификација и карактеризација текста итд).
- Српски језик, упркос заостатку, прати – у недовољно подржаном обиму – развој ресурса и алата.



Значај језичких технологија

- Захваљујући интернету, постале су доступне огромне количине језичких ресурса и алата, као и корпуса који се могу користити за обучавање језичких модела.
- Чланову Друштва ЈеРТех већ деценијама изграђују различите ресурсе и технологије за српски језик.
- Ресурси (једнојезични и вишејезични корпуси, ел. Речници, термилошке базе), апликације (паралелни текстови, веб сервиси за аотације или проширење упита...), модели за аотацију, ресурси семантичког веба итд...

Значај језичких технологија - примери

- Једнојезични и вишејезични корпуси (No Sketch Engine <https://noske.jerteh.rs/>, Библиша <https://biblisha.jerteh.rs/>, Текстометрија ТХМ)
- електронски речници (Unitex, Leximirka)
- Корпуси **општег језика**: СрпКор2013, Срп Кор2021, ВикиКорпус, СрпELTeC (постоје и **доменски корпуси** – рударски, геолошки, математички, библиотекарски..., **двојезични корпуси** – енг,фра, нем, ита – српски, српско-српски)
- Највећи генеративни модел за српски језик ГПТ2 – ОРАО и мањи модел ГПТ2 – ВРАБАЦ (модел обучавани на Националној платформи за вештачку интелигенцију Србије)
- Практична примена: лексикографија, настава језика и књижевности, превођење итд...

Отворена наука

- Екосистем
- Укупни приходи индустрије су огромни. Пре короне, годишњи приходи су износили 19.000 милиона УСД. Тржиштем доминира 5 великих издавача: *Elsevier*, *Black & Wiley*, *Taylor & Francis*, *Springer Nature* и *SAGE*. *Elsevier* је највећи са 16% маркет шера: око 3000 часописа, са профитном маржом 40%
- У 2010-14, такозвани предаторске издавачи су узели око 75 милиона УСД и објавили скоро пола милиона чланака у око 8000 часописа
- Стога је у току процес трансформације Open Access у Open Science

Отворена наука

- Зато је тренутно на сцени једна врста удаљавања од профитних пословних модела који погоршавају неједнакост и који су у супротности са Унесковим принципима и вредностима ОС, који се у својој основи заснивају на:
 - квалитету и интегритету
 - колективном бенефиту
 - праведности и правичности
 - разноврсности и инклузивности

But the big corporations and governments need to make the significant changes!!!



Youth Power for Social Change



COST Акција

- COST Action: CA16204 – Distant Reading for European Literary History
- Општи циљ: стварање динамичне и разнолике мреже истраживача који заједнички развијају ресурсе и методе неопходне за промену начина на који се пише европска књижевна историја.
- Утемељење: Distant Reading Paradigm - коришћење машинских (рачунарских) метода за анализу великих збирки књижевних текстова

РЕЗУЛТАТ

- Заједнички теоријски и практични оквир који омогућава иновативне, софистициране, рачунарске методе анализе књижевног текста на (иницијално) најмање 10 европских језика, са богатом књижевном традицијом.
- Принцип COST Акције омогућио је и другим језицима који нису били предвиђени у иницијалном предлогу пројекта да се придруже овој активности, па су српски језик у акцији представљали Универзитет у Београду и његове чланице: Универзитетска библиотека „Светозар Марковић“ и Филолошки факултет, као и Друштво за језичке ресурсе и технологије.

Конкретни циљеви

- Изградити вишејезичну збирку европских књижевних текстова (ELTeC) која садржи око 2.500 романа у пуном тексту на најмање 10 различитих језика, што омогућава тестирање метода и упоређивање резултата у различитим националним традицијама.
- Успоставити и поделити најбоље праксе и развити иновативне рачунарске методе анализе текста прилагођене европским вишејезичним књижевним традицијама.
- Размотрите последице таквих ресурса и метода за поновно промишљање основних појмова у теорији и историји књижевности.

САДРЖАЈ



- Интегрални текстови романа (новела) од најмање 10.000 речи чије прво издање датира из 1840-1920 (како би се осигурало да нема ограничења ауторским правима). Кад год је то могуће, прва издања одабраних дела се скенирају.
- Специфичност српске књижевности – новеле (нпр. Лаза Лазаревић)
- Корпус обухвата 120 романа (5.263.071 речи, 22.700 страница, 2557 поглавља, 158.317 пасуса, 567 песама, 2972 стиха)

КАНОН?

- Да би се постигла репрезентативност корпуса:



- романи који су део корпуса обухватају читав одабрани период,
- рад су различитих аутора (узимајући у обзир заступљеност женског писања)
- различит степен каноничности (укључујући и заборављене романи који нису у књижевном канону).

КАНОН?

- Лаза Лазаревић „Швабица“
- Борисав Станковић „Нечиста крв“
- Стеван Сремац „Зона Замфирова“

али и:

- Пера Радловић „Без оца и мајке“ (1875)
- Коста Барунчић „Пастир краљ или Ослобођење Србије“ (1879)
- Живојин Јовичић „Робињица Злата“ (1893)

Женски аутори

- Драга Гавриловић „Из учитељичког живота“ (1884)
- Драга Гавриловић „Бабадевојка“ (1887)
- Драга Гавриловић „Девојачки роман“ (1889)

- Јелена Димитријевић „Ђул-Марикина приказња“ (1901)
- Јелена Димитријевић „Фати-султан“ (1903)
- Јелена Димитријевић „Нове“ (1912)

- Милица Јанковић „Пре среће“ (1918)

- Исидора Секулић „Ђакон Богородичине цркве“ (1919)

ДОДАТНА ВРЕДНОСТ

- Сви текстови у дигиталној верзији опремљени су:
 - Метаподацима
 - XML ознакама које описују логички и графички изглед текста
 - Појединачним структурираним елементима текста у складу са препорукама TEI (Text Encoding Initiative) пројекта.
 - Наслови, пасуси, фусноте и делови текста који су на посебан начин истакнути (нпр. делови на страном језику – 803 сегмента) експлицитно означени, као и пагинација оригиналног издања текста.

Линк до Узајамног каталога



Database ▾ New search ▾ My COBISS Profile ▾ Theme Help ▾

[Search](#) > Hit

ALL libraries (COBIB.SR union bibliographic/catalogue database)

+ Interlib. loan

Number of loans

Basic information

Details

Author – person	Комарчић, Лазар
Title	Просиоци : роман / написао Л. Комарчић
Type of material	novel
Language	Serbian
Publish date	1905
Publication and manufacture	[Београд] : Издање Задужбине Илије М. Коларца, 1905 (Београд : Штампарија Светозара Николића)
Physical description	IV, 144 стр. ; 20 cm
Series	Задужбина Илије М. Коларца ; 108
ISBN	(Broš.)
UDC	821.163.41-31
COBISS.SR-ID	130736903

Чланак ка српској Википедији



Лазар Комарчић

С Википедије, слободне енциклопедије

Лазар Комарчић (Глог^{[1][2]}, Прибој, 9/21. јануар 1839 — Београд, 9/22. јануар 1909) је био српски новинар и књижевник. Понекад је писао под псеудонимом Дегилъ.

Садржај [сакриј]

- Биографија
- Библиографија
 - Романи
 - Збирке приповедака
 - Остало
- Референце
- Спољашње везе

Биографија [уреди | уреди извор]

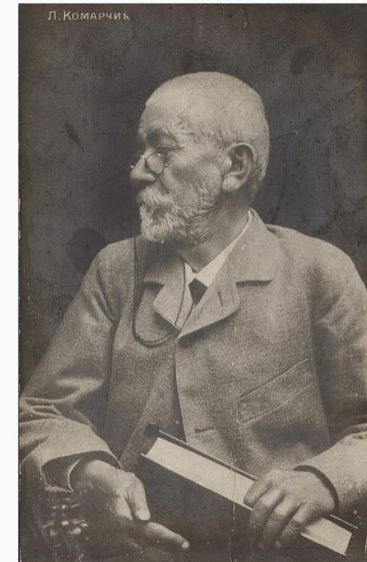
Као новинар, најпре је сарађивао у листу *Србија* око кога се седамдесетих година окупљала напредна омладина Србије, а затим и у листовима *Будућност*, *Исток*, *Нови Завет* и *Видело*. Покренуо је и сопствени лист *Збор* у Крагујевцу 1875. године, и у њему износио своје политичке погледе који су тада били под утицајем **социјалистичке** мисли **Светозара Марковића**.

Написао је 1902. први српски **научно-фантастични роман** *Једна угашена звезда*^[3], који се уз драму *Драгутина Илића После милион година* (1889) сматра за темељ модерне српске научне фантастике.^[4]

Роман *Један разорени ум* је 1893. наградила је **Српска краљевска академија**, а запажена је била и збирка приповедака *Записник једног покојника*. Написао је и романе *Драгоцен а грлица* (1880), *Два аманета* (1893), *Просиоци* (1905). Ови романи имају криминалистички заплет и спадају међу прве такве врсте на српском језику.^[5] Целокупно књижевно дело Лазара Комарчића није искључиво научно фантастично, већ има и особине романтичарске и реалистичке књижевности.^[6]

За разлику од **Жила Верна** и **Х. Џ. Велса** готово да је потпуно заборављен. **Друштво љубитеља фантастике „Лазар Комарчић“** негује научну фантастику и за име свог друштва је узело име једног од првих људи који су писали научну фантастику у српским и јужнословенским крајевима.

Лазар Комарчић



Датум рођења	21. јануар 1839.
Место рођења	 Глог Османско царство
Датум смрти	22. јануар 1909. (70 год.)

ГДЕ СЕ МОЖЕ ПРОНАЋИ САДРЖАЈ

- Платформа COST Акције:
<https://zenodo.org/communities/eltec?page=1&size=20>
- Сајт УБСМ: Удаљено читање <https://udaljenocitanje.unilib.rs/>
(паралелан преглед сканиране слике дела и ишчитаног текста, веза аутора и дела са различитим референтним базама: COBISS, VIAF (Virtual International Authority File), Википодаци, Википедија).
- Друштво за језичке ресурсе и технологије (JePTex) <http://aurora.jerteh.rs/>
(пружа увид у интегрални текст дела, речник облика свих речи употребљених у њему са фреквенцијом и комплетне конкорданце (списак употребљених речи са левим и десним контекстом) текста, веза аутора и дела са различитим референтним базама, објављивање текстова обрађених дела у ЕПУБ формату, што би омогућило комфортно читање на преносним уређајима).

УДАЉЕНО ЧИТАЊЕ

- Пројекат подржан од стране Министарства културе и информација Републике Србије, као подршка српском учешћу у COST акцији (2019).
- Захваљујући пројекту *Удаљено читање* из 2019. године, српски језик је један од девет језика укључених у прву верзију ELTeC корпуса: немачки, енглески, француски, италијански, норвешки, португалски, румунски, словеначки и српски. У овој верзији српски језик је заступљен са 30 романа.

Почетни екран на сајту пројекта



УДАЉЕНО ЧИТАЊЕ О ПРОЈЕКТУ Београд 1912 Београд 1900

Једна угашена звезда : ELTeC издање

Комарчић, Лазар (1833-1909)
Штампарија Д. Димитријевића

Једна угашена звезда : илустровани роман

Београд 1902

Покојникова жена : ELTeC издање

Станковић, Борисав (1876-1927)
Књижарница Боже О. Дачића

Покојникова жена

Београд 1907

Хаџи-Ђера : ELTeC издање

Илић, Драгутин Ј. (1858-1926)
Српска књижевна задруга

Хаџи-Ђера : приповетка

Београд 1904

Просиоци : ELTeC издање

Комарчић, Лазар (1833-1909)
Задужбина Илије М. Коларца

Просиоци

Београд 1905

Новац : роман из београдског живота : ...

Талетов, Пера С. (1875-1955)
Штампарско умјетнички завод Пахера и Кисића

Новац : роман из београдског живота

Мостар 1906

Калуђер и хајдук : приповетка о послед...

Новаковић, Стојан (1842-1915)
Задужбина И. М. Коларца

Калуђер и хајдук : приповетка о последњим данима Србије у XV веку

Београд 1919

У фронт : приповетка из живота једног б...

Борђевић, Владан (1844-1930)
Штампарија Ђ. Мунца и М. Карића

У фронт : приповетка из живота једног бившег краља

Београд 1913

Васа Решпект : ELTeC издање

Игњатовић, Јаков (1822-1889)
Српска књижевна задруга

Васа Решпект : роман

Београд 1913

Чедомир Илић : ELTeC издање

Ускоковић, Милутин (1884-1915)
Геца Кон

Чедомир Илић

Београд 1914

ЧИТАЊЕ НА ДАЉИНУ

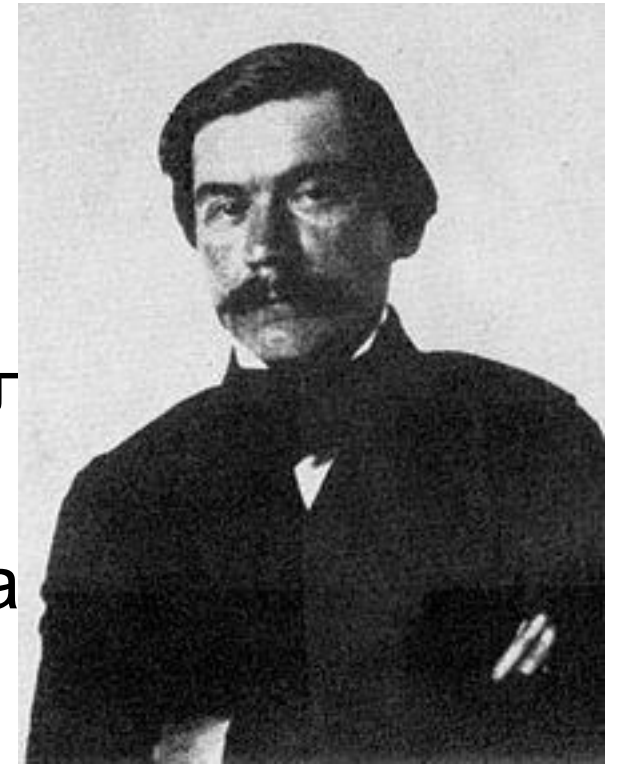
- Пројекат подржан од стране Министарства културе и информисања Републике Србије (2022)
- У склопу одрживости резултата COST Акције

Циљ:

- Наставак обраде српских романа
- Укључивање других књижевних жанрова – пре свега путописа

ПУТОПИСИ

- Чедомилъ Мијатовић „Цариградске слике и приг
- Љубомир Ненадовић „Писма из Италије“ (Нема Швајцарске)
- Јелена Димитријевић „Седам мора и три океана“ („Писма“)
- Евлија Челебија



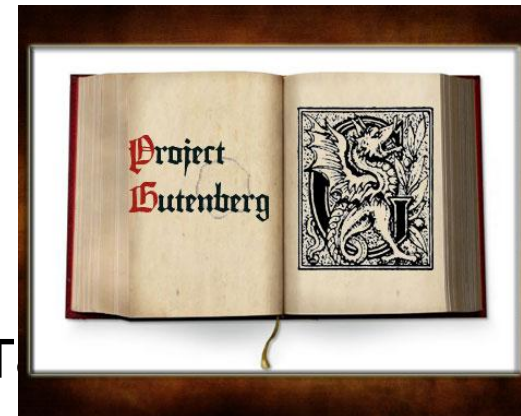
ПРОЦЕС

- Прибављање дела
- Њихова дигитализација
- Трансформација тако добијене слике у текст (OCR)
- Обављање аутоматске корекције текста
- Додатна коректура текста
- Његово снабдевање анотацијама у виду XML-етикета и метаподацима

ПРОЦЕС

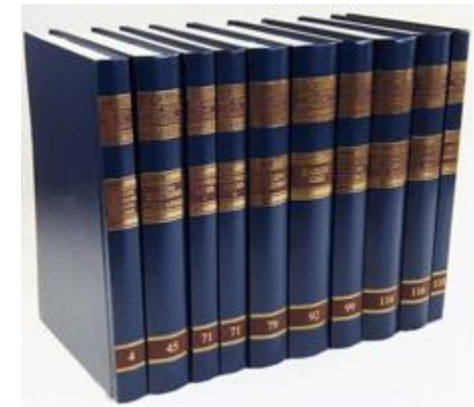
- У оквиру претходног пројекта извршени су успешни експерименти са аутоматским обележавањем текстова корпуса специфичним језичким објектима: именима људи, локација, организација и сл.
- У потпуности усклађено са препорукама COST акције *D-Reading*.

Домаћи ГУТЕНБЕРГ



- Роман 19. и 20. века изабран као адекватан почет
- Портал је замишљен тако да касније може у себе инкорпорирати и остале књижевне материјале из 19. и 20, али и других векова српске културе, који су ослобођени ауторских права.
- Својеврсни културни мост у синхронијској и дијахронијској равни српске културе, на њеном репрезентативном књижевном узорку, па би се овај портал могао посматрати као нека врста домаћег пандана сајту Гутенберг
- Пројекат се у будућности може потенцијално гранати ка лексикографији, лингвистици, али и изради речника појединих писаца који драматично недостају српској културној сцени.

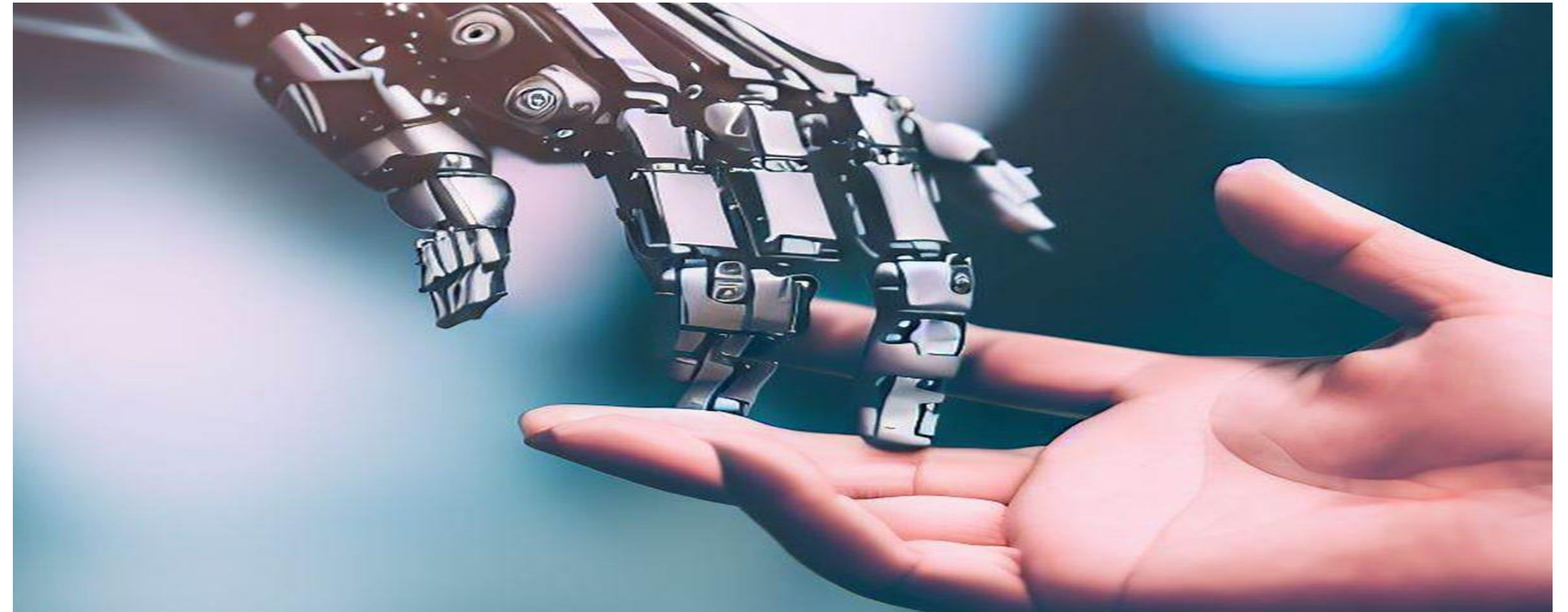
Може ли дигитална хуманистика угрозити/изменити КАНОН?



Да ли је КАНОН само теорија?



Critical Thinkink in the Age of AI Revolution



Don't ever wait to ask permission to lead, you just lead!



ХВАЛА!

milnovic@unilib.rs

<https://www.linkedin.com/in/vasilije-milnovi%C4%87-a95956184/>