

# Sanjaju li androidi električne bibliotekare/ke? Primjena AI tehnologija u nacionalnim bibliotekama Evrope

Dalila Mirović

dalila.mirovic@gmail.com

**Cilj:** Cilj ovog istraživanja jeste pružiti pregled primjene tehnologija vještačke inteligencije (AI) u nacionalnim bibliotekama Evrope, s fokusom na primjenu mašinskog učenja za (polu)automatsko klasificiranje i indeksiranje, projekte digitalne humanistike te inovativne metode zaštite i prezentacije kulturnog naslijeđa. Istraživanje ima za cilj podstaknuti interesovanje stručne i znanstvene bibliotečke zajednice za mogućnosti koje nude ove tehnologije.

**Pristup/metodologija/dizajn:** Istraživanje koristi deskriptivnu metodu za predstavljanje primjene AI tehnologija u nacionalnim bibliotekama. Korištene su metode komparacije, analize i sinteze kako bi se inducirale teorijske postavke o predmetu istraživanja. Primjeri su analizirani iz prakse nacionalnih biblioteka Finske, Švedske, Norveške, Njemačke, Španije, Francuske, Nizozemske, Škotske, Luksemburga i Velike Britanije.

**Rezultati:** Rezultati istraživanja pokazuju da nacionalne biblioteke koriste AI tehnologije za automatsko i poluautomatsko klasificiranje, indeksiranje, razvoj jezičnih modela, te u projektima digitalne humanistike. Nacionalne biblioteke kao što su one u Finskoj i Njemačkoj prednjače u primjeni mašinskog učenja za automatizaciju stručne obrade građe, dok projekti poput "Živjeti s mašinama" Britanske biblioteke demonstriraju inovativne primjene AI-ja u istraživanju kulturnih promjena.

**Diskusija:** Implementacija AI tehnologija u nacionalnim bibliotekama donosi mnoge prednosti; njihove AI laboratorije (AI Labs) pružaju platformu za razvoj i testiranje novih tehnologija, dok jezični modeli koje tu razvijaju omogućavaju

dublju i bogatiju interakciju sa zbirkama. Projekti digitalne humanistike pokazuju kako AI može pomoći u očuvanju i istraživanju kulturnog naslijeđa, ali zahtijevaju interdisciplinarni pristup i saradnju različitih stručnjaka i stručnjakinja.

**Originalnost/vrijednost:** Ovo istraživanje je jedno od prvih koje ispituje primjenu AI tehnologija u nacionalnim bibliotekama Evrope, nudeći sveobuhvatan pregled trenutnog stanja i pružajući osnovu za buduća istraživanja u ovoj oblasti. Istraživanje pokazuje kako AI tehnologije mogu transformirati način na koji nacionalne biblioteke funkcionišu i kako se kulturna baština može očuvati i učiniti dostupnijom.

**Ključne riječi:** nacionalne biblioteke, nacionalne biblioteke u Evropi, AI tehnologije, AI laboratorije, automatsko klasificiranje, automatsko predmetno označavanje, Annif, digitalne biblioteke, digitalna humanistika

## Uvod

U posljednjih nekoliko godina, tehnologije vještačke inteligencije (AI) transformisale su brojne industrije, uključujući i sektor kulturne baštine. Nacionalne biblioteke, kao ključne institucije u očuvanju, zaštiti i prezentaciji kulturnog naslijeđa, sve ih češće primjenjuju kako bi unaprijedile svoje usluge i procese, a zahvaljujući svojim resursima, kapacitetima i jedinstvenoj poziciji, one zajedno s visokoškolskim bibliotekama prednjače u njihovom testiranju i implementaciji. Ove institucije koriste AI ne samo za optimizaciju svakodnevnih poslova i usluga, kakve su automatsko klasificiranje i predmetno označavanje, te chatbotovi, već i za razvoj naprednih istraživačkih alata koji omogućavaju dublju analizu i razumijevanje kulturnih fenomena, kao i za projekte iz digitalne humanistike. Primjeri iz prakse pokazuju kako AI može obogatiti interakciju korisnika/ca sa zbirkama, olakšati pristup informacijama te doprinijeti boljem očuvanju i interpretaciji kulturne baštine. Ovaj se rad koristi deskriptivnom metodom za analizu postojećih primjena AI u nacionalnim bibliotekama te metodama komparacije, analize i sinteze kako bi se došlo do teorijskih zaključaka o ulozi ovih tehnologija u budućem razvoju bibliotečke prakse. Fokus istraživanja je na specifičnim primjerima nacionalnih biblioteka iz Finske, Švedske, Norveške, Njemačke, Španije, Francuske, Nizozemske, Škotske, Luksemburga i Velike Britanije, koje su među prvima prepoznale potencijal AI tehnologija, a kako bi se podstaklo interesovanje stručne i znanstvene bibliotečke zajednice za mogućnosti koje one nude.

Iako istraživanja o primjeni vještačke inteligencije u bibliotekama, a naročito visokoškolskim bibliotekama, ne manjka, fokus na specifične primjene AI tehnologija, poput mašinskog učenja za (polu)automatsko klasificiranje ili obrade prirodnog jezika za oživljavanje fondova u nacionalnim bibliotekama, relativno je specifično i nedovoljno istraženo područje. Pa ipak, dokumenti poput IFLA-ine izjave *IFLA Statement on Libraries and Artificial Intelligence* (IFLA, 2020), u kojoj se naglašava važnost etičkih razmatranja pri korištenju AI-ja, a biblioteke podstiču da kritički procjenjuju AI alate i njihove implikacije, kako bi unaprijedili usluge bez narušavanja etičkih standarda, ili izvještaja Cordell, R. *Machine Learning + Libraries* (Cordell, 2020), u kojem je dat sveobuhvatan pregled trenutnog stanja primjene mašinskog učenja u bibliotekama i u kojem se također navodi važnost etičnog korištenja i primjene ovih tehnologija, daju važne smjernice za izučavanje AI-ja u nacionalnim bibliotekama stoga što naglašavaju potrebu da biblioteke ne trebaju tek tako prihvatiti nove tehnologije, već se i pobrinuti da njihova primjena bude u skladu s osnovnim vrijednostima bibliotekarstva, njihovom misijom i interesima korisnika/ca.

Istraživanje je strukturirano tako da se kroz nekoliko ključnih poglavlja tretiraju različiti aspekti primjene vještačke inteligencije u nacionalnim bibliotekama. Nakon uvodnog dijela, u kojem se postavlja kontekst i definira cilj istraživanja, prvo poglavlje fokusira se na AI laboratorije (*AI Labs*) unutar nacionalnih biblioteka, gdje se istražuju inicijative za testiranje i razvoj AI tehnologija unutar ovih institucija. Sljedeće poglavlje se bavi oblikovanjem i primjenom jezičnih modela koji omogućavaju razvoj naprednih funkcija poput pretraživanja i automatske transkripcije. U poglavlju “Projekti iz oblasti digitalne humanistike” navode se konkretni primjeri kako AI podržava projekte koji se bave očuvanjem i analizom kulturne baštine. Sljedeće se poglavlje bavi automatskim klasificiranjem i predmetnim označavanjem, u kojem prednjače biblioteke Njemačke i Finske. Na kraju, poglavlje “Inovativni načini pružanja usluga korisnicima/cama te ostale AI usluge” razmatra kako AI tehnologije transformiraju korisničko iskustvo u bibliotekama kroz nove usluge poput chatbotova.

## **AI laboratorije (AI Labs) unutar nacionalnih biblioteka**

Proliferacija elektronskih dokumenata i izvora te eksponencijalno širenje digitalnih biblioteka, kako digitaliziranim građom, tako i građom koja

je nastala u digitalnom obliku, nužno su stvorili potrebu za njenim što efikasnijim organizovanjem i korištenjem. Činjenica da obavezni primjerak u većini nacionalnih biblioteka danas uključuje i elektronsku građu predstavljao je dodatni izazov nacionalnim bibliotekama, naročito u stručnoj obradi, odnosno dodjeljivanju metapodataka. Primjera radi, Nacionalna biblioteka Njemačke je 2018. godine mjesečno sadržajno obrađivala između 5.000 i 10.000 publikacija (Junger, 2018), a u 2020. godini je njena zbirka porasla za oko milion novih digitalnih publikacija, kakve su elektronske knjige i časopisi (Mödden, 2022). Isto tako, u Nacionalnu biblioteku Švedske mjesečno stigne više od 50 terabajta novih podataka (Salian, 2023), dok je cilj Nacionalne biblioteke Škotske da jedna trećina njihove ukupne građe bude dostupna u digitalnom obliku do 2025. godine, a samo 2020. je godine u nju kao obavezan primjerak bilo pohranjeno više od 5,2 miliona elektronskih časopisa i knjiga, dok je, paralelno, interni program masovne digitalizacije, s osobljem koje je radilo u dvije smjene kako bi se maksimalno iskoristila kamera, povećao broj digitaliziranih izvora (i to samo onih koji ne podliježu autorskim pravima) na 201.679 u 2018. i 2019. godini. Ova biblioteka već ima oko 5 petabajta podataka pohranjenih u svojim podatkovnim centrima (Ames i Lewis, 2020).

Pored toga, izazov je bio i šta učiniti s građom nakon što se digitalizira, stručno obradi i pohrani – kako te ogromne količine podataka učiniti korisnim i dostupnim korisnicima i korisnicama svih profila, a naročito istraživačima i istraživačicama? Čemu digitalna i digitalizirana građa ako ona neće biti korištena, odnosno ako neće biti korištena na najoptimalniji način?

Na tu potrebu za pronalaženjem što boljih i inovativnijih načina da ove ogromne količine podataka iz biblioteka – a naročito nacionalnih biblioteka – budu iskorištene odgovorile su nove tehnologije vještačke inteligencije, konkretno mašinsko i duboko učenje, značajno poboljšavajući njihovu funkcionalnost i usluge. Prema izvještaju Cordell, R. o primjeni mašinskog učenja u bibliotekama, ove tehnologije, između ostalog, omogućavaju automatsko ekstrahovanje metapodataka, klasificiranje i napredno pretraživanje zbirke te prepoznavanje rukopisa (HTR) i optičko prepoznavanje znakova (OCR) (Cordell, 2020). Na primjer, OCR tehnologija pomaže u digitalizaciji starih rukopisa i štampanog materijala, čineći ih pretraživima, pa tako i pristupačnijim. Mašinsko učenje također podržava automatsku

izradu sažetaka i davanje preporuka, olakšavajući korisnicima i korisnicama pronalaženje relevantnih informacija unutar velikih zbirki. Osim toga, ono doprinosi njihovom očuvanju kroz prediktivnu analitiku, koja pomaže u identificiranju i sprečavanju potencijalnih oštećenja materijala. Biblioteke također koriste ove tehnologije za edukaciju osoblja i korisnika/ca o njihovoj primjeni, čime se povećava opšta digitalna pismenost i učinkovitost bibliotečkog poslovanja. Primjena mašinskog učenja, a naročito vještačkih neuronskih mreža, ne samo da unapređuje operativne procese već također transformira interakciju korisnika/ca sa zbirkama, omogućujući dublje i bogatije istraživanje kulturne baštine.



*KB\_datalab*, fizički AI laboratorij Nacionalne biblioteke Nizozemske (KBNL, s. a.)

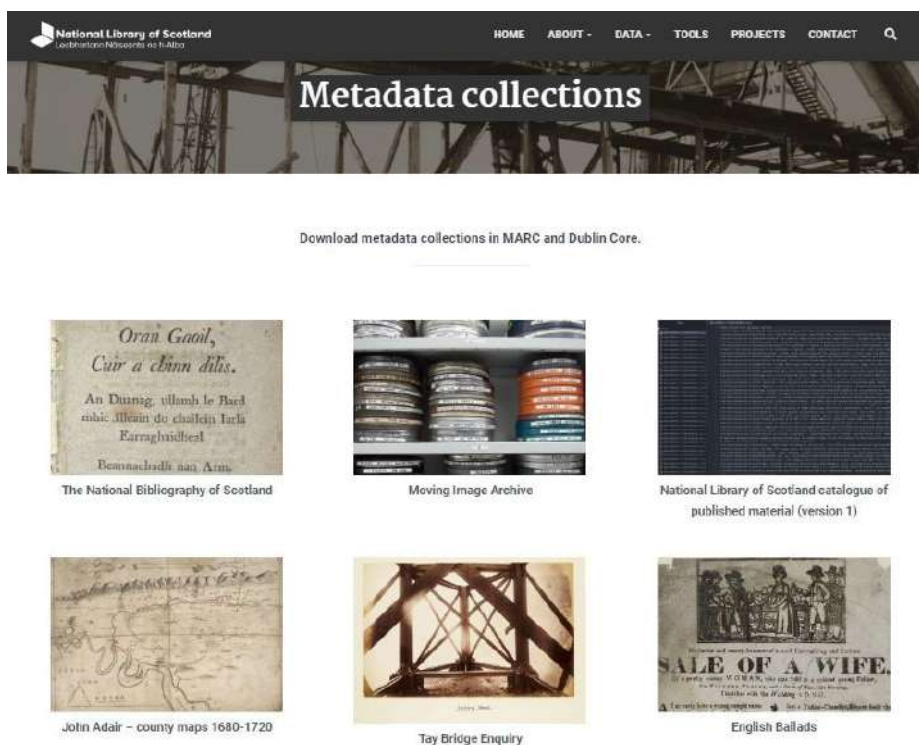
Upravo su u tu svrhu, kako navode Haffenden et al. (2021), brojne visokoškolske i nacionalne biblioteke uspostavile *AI/data laboratorije* kao mjesta za testiranje, ispitivanje i korištenje potencijala novih tehnologija, između ostalih, *BL Labs* Britanske biblioteke, *AI-Lab* Nacionalne biblioteke Norveške, *KBLab* Nacionalne biblioteke Švedske, laboratorij skoro pa istog naziva Nacionalne biblioteke Nizozemske – *KB Lab*, zatim

*BNElab* Nacionalne biblioteke Španije i *BnF datalab* Nacionalne biblioteke Francuske. Ove laboratorije prirodan su nastavak projekata digitalizacije i služe što optimalnijem korištenju zbirki iz digitalnih biblioteka kakva je *Delpher* Nacionalne biblioteke Nizozemske ili *Gallica* Nacionalne biblioteke Francuske. Isto tako, kako navode Börjeson et al. (2024), one su odgovor na sve veće zahtjeve za računalnim pristupom zbirkama, naročito u posljednjih desetak godina, te predstavljaju interne platforme “gdje se profesionalna stručnost podatkovnih znanstvenika/ca može iskoristiti za informacijsku složenost digitalizacije i facilitiranje novih oblika digitalnog istraživanja” (Börjeson et al., 2024) ili, kako to oslikavaju Haffenden et al., ti su Labovi spoj vještačke inteligencije i informacijskih nauka, mjesta gdje se “vrhunsko znanje iz podatkovne znanosti kombinira s velikim iskustvom s bibliotečkim informacijskim sistemima i radnim procesima” (Haffenden et al., 2021).

Računalni pristup zbirkama, odnosno kulturnoj baštini, posebno je zagovarao pokret *Zbirke kao podaci* (*Collections as data*). Ovaj je transformativni pristup u bibliotekama, arhivima i muzejima naglašavao važnost pretvaranja zbirki kulturnog nasljeđa u strukturirane setove podataka za računalnu analizu. On koristi digitalne zbirke za podršku novim oblicima znanstvenih istraživanja i angažmana javnosti, omogućujući rudarenje podataka i teksta, mašinsko učenje i druge računalne metode za istraživanje i tumačenje građe kulturne baštine. Proces datafikacije uključuje pretvaranje različitih vrsta sadržaja, poput teksta, slika i multimedije, u mašinski čitljive formate, čime se poboljšava pristupačnost i upotrebljivost zbirki za računalnu analizu. Ovome je pokretu izuzetno važno da su skupovi podataka otvoreno dostupni kako bi se mogli koristiti za interdisciplinarna istraživanja u područjima kakva su digitalna humanistika, društvene nauke i računalna nauka. Kako to pojašnjavaju Padilla et al. (2019), paradigma *Zbirki kao podataka* postavlja pitanje kako se digitalne i digitalizirane zbirke građe kulturne baštine mogu tretirati kao podaci, a ne samo kao surogati svojih fizičkih objekata, dakle kako da aktivno budu korištene u svom prirodnom digitalnom habitatu, a ne da budu samo statične reprezentacije. Stoga ove nacionalne biblioteke objavljuju i čine javno dostupnim razne setove podataka (*datasets*), od svojih nacionalnih bibliografija, preko urbanističkih nacrti, mapa, plakata, autorskih kataloga, do digitaliziranih zbirki, pa čak i javnih nabavki u .json, .csv i drugim formatima kako bi ih istraživači i istraživačice, ali i šira javnost slobodno mogli koristiti po vlastitim željama i potrebama.



Ovi podaci omogućili su pružanje novih usluga u bibliotekama, kakva je usluga *digitalnog istraživanja* (*digital scholarship*) Nacionalne biblioteke Škotske, koje se u članku Ames i Lewis (2020) definiše kao upotreba računalnih metoda na zbirkama ove biblioteke kako bi se omogućili novi oblici istraživanja. Također, pored toga što nacionalne biblioteke objavljuju ove setove podataka na svojim stranicama, one također pozivaju istraživače/ce i da fizički koriste njihove Labove, kao u slučaju nacionalnih biblioteka Francuske i Nizozemske.



The screenshot shows the National Library of Scotland website. At the top, there is a navigation menu with links for HOME, ABOUT, DATA, TOOLS, PROJECTS, and CONTACT. The main heading is 'Metadata collections'. Below this, there is a sub-heading: 'Download metadata collections in MARC and Dublin Core.' A grid of six data sets is displayed, each with a representative image and a caption:

- The National Bibliography of Scotland**: Image of a book cover titled 'Oran Gaol, Cuir a chinn dèilig.' with the subtitle 'An Duineag, uillamh le Bard mhòr: Mhàin do chiallaj fàrla Earragh-vhichest' and 'Bonnachadh nan Arin.'
- Moving Image Archive**: Image of several film canisters on a shelf.
- National Library of Scotland catalogue of published material (version 1)**: Image of a dark screen displaying a list of text.
- John Adair – county maps 1680-1720**: Image of an old map showing county boundaries.
- Tay Bridge Enquiry**: Image of the Tay Bridge structure.
- English Ballads**: Image of a ballad sheet titled 'SALE OF A WIFE'.

Prikaz nekih od setova podataka dostupnih na web-stranici  
Nacionalne biblioteke Škotske (NLS, s. a.)

Uz ove tehnologije, digitalne biblioteke nacionalnih, ali i drugih biblioteka, oživljavaju, a kulturna baština dobiva brojne mogućnosti i načine da bude predstavljena u novom, do sada nepoznatom svjetlu. No, također postoje i određeni izazovi što se tiče njihove implementacije u svakodnevne bibliotečke procese i usluge, a oni su prvenstveno organizacione, tehničke, etičke i finansijske prirode.



## BIBLIOGRAFÍA ESPAÑOLA

# BNElab - DATOS

<b>LICENCIAS</b> Creative Commons	<b>MATERIALES</b> Registros bibliográficos	<b>FORMATOS</b> CSV JSON MARC MARC-XML ODS TXT XML	<b>RELACIONADOS</b> • Catálogo bibliográfico	<b>CONTACTO</b> bnelab@bne.es
--------------------------------------	---	---	---	----------------------------------

Nacionalna bibliografije Španije u mašinski čitljivim formatima / Preuzeto s web-stranice *BNElab* projekta Nacionalne biblioteke Španije (BNElab, s. a.)

Jedan od glavnih tehničkih izazova je osiguranje kvalitete i dosljednosti podataka. Alati bazirani na vještačkoj inteligenciji zahtijevaju velike, visokokvalitetne setove podataka kako bi mogli učinkovito funkcionisati. Također, održavanje i nadogradnja tehnološke infrastrukture potrebne za podršku projektima vještačke inteligencije zahtijeva značajna ulaganja i ogromna finansijska sredstva te visoku stručnost osoblja. Pored toga, i etičnost je od ključne važnosti u primjeni ovih tehnologija. Biblioteke moraju uzeti u obzir i pitanja vezana za zaštitu privatnosti podataka korisnika/ca, ali i potencijalne pristrasnosti ugrađene u same algoritme (Cordell, 2020). Rješavanje ovih izazova zahtijeva sveobuhvatnu strategiju koja uključuje ulaganje u tehnologiju i infrastrukturu, kontinuiranu obuku osoblja, etičke standarde i predanost inkluzivnosti, a takva je, na primjer, njemačka državna *AI Strategija* (NJV, 2020) kroz koju se izdvajaju nemala sredstva za podršku njihovoj nacionalnoj biblioteci.

### Oblikovanje i primjena jezičnih modela

Kako je već spomenuto, svi ovi projekti digitalizacije i uvođenja novih tehnologija kroz AI Labove u nacionalnim bibliotekama nikako nisu sami sebi svrha. Cilj im je omogućiti nova istraživanja i oživjeti kulturnu baštinu. Jedan od načina na koji nacionalne biblioteke doprinose kvalitetnijem korištenju svojih zbirki – koje predstavljaju javno dobro – jeste oblikujući i primjenjujući jezične modele svojih maternjih jezika.



Haffenden et al. (2021) *jezične modele* definišu kao statističke modele koji mogu razumjeti i generisati ljudski jezik nakon obučavanja na ogromnim količinama teksta. Od oko 7.000 jezika u svijetu, najveći akcenat je stavljen na njih dvadesetak koji imaju i najviše govornika/ca, a to su ujedno i jezici koji imaju najveći korpus tekstova na kojima se mogu obučavati algoritmi za obradu prirodnog jezika. Svi jezici koji ne spadaju u ovih 20 nazivaju se *jezici s malim brojem izvora* (*low-resource languages, LRLs*) (Magueresse et al., 2020). Ti jezici nisu dovoljno komercijalno zanimljivi da bi ih podrobno razvijali tehnološki giganti poput *Googlea*, niti postoji dovoljna količina javno dostupnih podataka na njima kako bi se algoritmi mogli obučiti. Haffenden et al. (2021) važnu ulogu nacionalnih biblioteka i njihov demokatičan učinak na razvoj vještačke inteligencije vide upravo u razvijanju jezičnih modela svojih država. Nacionalne biblioteke, čija je misija da sakupljaju i čuvaju sve što je objavljeno na njihovom jeziku i koje tu građu digitaliziraju, te koje imaju bogat lingvistički materijal i veći korpus tekstova negoli ijedna druga javna ili privatna institucija/organizacija, “imat će ključnu ulogu u budućnosti obučavanja i kreiranja novih jezičnih modela” (Haffenden et al., 2021). Također, za razliku od tehnoloških giganata, nacionalne biblioteke rade na transparentan i demokatičan način, pa tako Nacionalna biblioteka Švedske svoje modele čini reprezentativnijim tako što algoritme obučava na podacima iz različitih slojeva društva te javno objavljuje i čini dostupnim svoje modele kako bi se mogli testirati (KB, 2022). Ta demokatičnost i otvorenost jesu ključni razlog da se razvoj vještačke inteligencije, pa i jezičnih modela, stavi u kontekst nacionalnih biblioteka.

Značaj razvijanja jezičnih modela jeste njihov širok raspon primjene u različitim domenama, uključujući klasifikaciju teksta, prepoznavanje imena (*name-entity recognition, NER*), označavanje dijelova govora, generisanje teksta (kao kod chatbotova i virtualnih asistenata) te sastavljanje emailova i dokumenata, zatim u prevođenju, mašinskom prevođenju i pretraživanju informacija na različitim jezicima. Također se koriste za sažimanje teksta, u personalizaciji i preporučivanju, a kao obrazovni alati pomažu u učenju jezika. Jezični modeli su u osnovi brojnih savremenih alata, poboljšavajući učinkovitost, korisničko iskustvo i pristupačnost u različitim industrijama. Pored ovako široke primjene u raznim oblastima, prema Haffenden et al. (2021), značaj jezičnih modela za biblioteke jeste u automatskom klasificiranju (o čemu će biti više riječi u narednom poglavlju), poboljšanoj mogućnosti

pretraživanja i poboljšanoj koheziji u optičkom prepoznavanju znakova (OCR). Nadalje, u bibliotekama se također jezični modeli mogu koristiti kao virtualni asistenti i chatbotovi, kakav je *CuratorBot* Nacionalne biblioteke Nizozemske, o kojem će također biti više riječi u nastavku rada. Važnost jezičnih modela za biblioteke leži u njihovoj sposobnosti da optimiziraju radne procese, poboljšaju angažman korisnika/ca i prošire pristup znanju, čime se stvara učinkovitije i korisnicima prilagođenije bibliotečko okruženje.

Nacionalne biblioteke Švedske i Norveške su 2021. godine unutar Labova razvile jezične modele za svoje jezike. Oba su bazirana na BERT-u, (*Bidirectional Encoder Representations from Transformer*), modernom modelu dubokog učenja (*deep-learning*) koji je 2018. godine razvio *Google* na korpusu od 3,3 milijarde riječi iz knjiga i Wikipedije na engleskom jeziku (Haffenden et al., 2021). Nacionalna biblioteka Švedske je razvila KB-BERT za švedski jezik, koji je obučen na još većem korpusu sastavljenom od tekstova izvučenih iz novina uz pomoć optičkog prepoznavanja znakova (OCR), elektronskih knjiga, članaka sa Wikipedije na švedskom jeziku, kao i tekstova prikupljenih s društvenih mreža, što je set od ukupno oko 3,5 milijarde riječi (Haffenden et al., 2021), dok je Nacionalna biblioteka Norveške razvila NB-BERT za svoj jezik. I jedna i druga biblioteka, pored ovih spomenutih, razvila je i druge modele, kakav je, na primjer, NB-Whisper, model za transkripciju audiozapisa na norveškom jeziku.

## The AI-lab at The National Library of Norway

### Welcome!

You have landed at the AI-lab at the National Library of Norway (NLN). If you look around here, you may find some information on what we do, fun things we have seen, and our services.

Basically this website centers around what we do and what we like – based on GLAMs as context. As such, this is on one hand our current and short-term contribution to possible applications of AI in GLAMs. The other side is the library as a contributor to the development of good AI. Understanding the latter requires the first.

Preuzeto s web-stranice Nacionalne biblioteke Norveške (NB, 2024a)

Jednom oblikovani i obučeni, ovi modeli trebaju biti i primijenjeni, a Nacionalna biblioteka Norveške je za svoje modele pronašla primjenu u nekoliko aplikacija (NB, 2024b):

- *Maken*: sistem za preporučivanje, kako slika, tako i knjiga;
- *Sambib*: aplikacija koja procjenjuje da li neki izvor potencijalno može biti uključen u skandinavsku bibliografiju;
- *Prepoznavanje imena (NER)*: aplikacija koja prepoznaje imena ljudi i organizacija, ali i nazive mjesta i slično;
- *SPIFF*: fotografski asistent;
- *Prepoznavanje naslovne stranice*: aplikacija koja prepoznaje šta je naslovna strana novina, šta zadnja te kakve su stranice u njihovoj sredini.

Švedska nacionalna biblioteka je primjene svog modela pronašla u pretraživanju zbirki uz pomoć slike i teksta kao inputa, te u automatskoj transkripciji audiovizuelnih zbirki, što bi bez mašina bilo nemoguće, jer se radi o milionima sati materijala (KB, 2022). Primjene modela koje je razvila ova biblioteka mogu biti i šire, pa je moguće da će uz njih “istraživači uskoro moći kreirati hiperspecijalizirane setove podataka – na primjer, pronalazeći svaku švedsku razglednicu na kojoj se nalazi crkva, svaki tekst napisan određenim stilom ili svaki spomen neke historijske ličnosti unutar svih knjiga, novinskih članaka i TV-prenosa” (Salian, 2023).

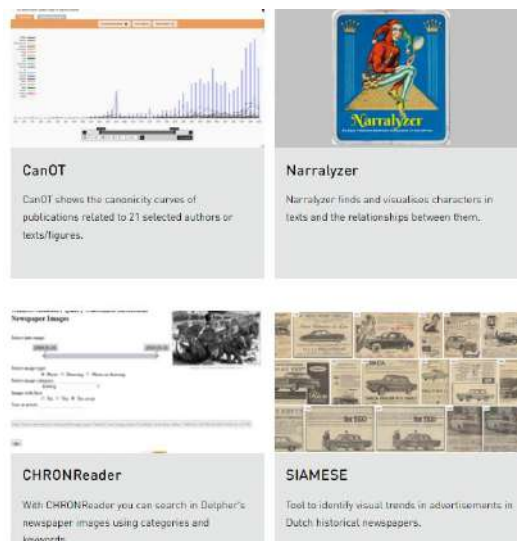


The image shows a screenshot of a news article from the National Library of Sweden. The article is titled "KB makes powerful models for language understanding available" and is dated 4 February 2020. The text of the article states: "Today, the National Library publishes three Swedish language models based on Google's 'BERT' (Bidirectional Encoder Representations from Transformers). The first tests show that KB's models outperform Google's multilingual model." Below the text is a grid of colorful book covers. In the bottom left corner of the screenshot, there is a logo for the National Library of Sweden, which consists of a circular emblem with the text "NATIONELLA BIBLIOTEKEN" around the top and "National Library of Sweden" around the bottom.

Preuzeto s web-stranice Nacionalne biblioteke Švedske (KB, 2020)

Nacionalna biblioteka Nizozemske je također pronašla vrlo zanimljive primjene svog jezičnog modela, koji nije razvijen unutar biblioteke, a neki od njih uključuju (KBNL, 2024b):

- *Canonizer*: ova aplikacija određuje da li je djelo kanonsko ili nije na osnovu njenog teksta;
- *Timelapse*: stotinu godina novina u samo jednoj minuti;
- *Annif*: automatsko dodjeljivanje ključnih riječi;
- *Word Embedding Playground*: obezbjeđuje alate za obučavanje i podešavanje modela za ugradnju riječi;
- *RDA Entity Finder*: omogućava pretraživanje kroz nekoliko bibliografskih entiteta;
- *CHRONReader*: pretraživanje novinskih slika u *Delpheru* uz pomoć kategorija i ključnih riječi;
- *SIAMESE*: alat koji vizualizira trendove u oglašavanju u nizozemskim novinama;
- *Frame generator*: alat za ekstrahovanje tema, ključnih riječi iz nizozemskog korpusa;
- *Genre classifier*: prepoznavanje žanra novinskih članaka uz pomoć običnog teksta;
- *Dictionary viewer*: vizualizira pojavljivanje određene riječi u novinskom korpusu tokom vremena.



Preuzeto s web-stranice Nacionalne biblioteke Nizozemske (KBNL, 2024b)

## Projekti iz oblasti digitalne humanistike

Kreiranje i činjenje dostupnim svih ovih jezičnih modela, setova podataka, alata i aplikacija, nije samo sebi svrha, nego služi da se vrše nova istraživanja u svim naukama, a naročito humanističkim i društvenim, koja do sada nisu bila moguća. Računalno analiziranje velikih količina podataka, mogućnost sagledavanja ogromnog broja djela istovremeno kako bi se pronašli obrasci, trendovi, veze i slično unutar velikih zbirki tekstualnih podataka danas je moguće zahvaljujući metodi *udaljenog čitanja (distant reading)*. Ovaj se pojam odnosi na analiziranje velikih korpusa teksta pomoću digitalnih alata i računalnih tehnika i suprotan je tradicionalnom, bliskom čitanju (*close reading*), koje se fokusira na detaljnu analizu pojedinačnih tekstova. Prema Tabak, E. (2017), udaljeno čitanje jeste “jedan od najprepoznatljivijih koncepata digitalne humanistike”, a kontroverza koja okružuje ovu metodu jeste nedostatak kritičizma spram načina dolaska do saznanja uz pomoć nje, jer je metoda, suštinski, kvantitativna, što je u određenoj suprotnosti sa humanističkim pristupom koji je po svojoj prirodi kvalitativan. Rezultati dobiveni mašinskim učenjem također mogu biti netransparentni, a Rebor, S. (2023) navodi da se konkretno tehnicima analize sentimenta može spočitati neadekvatnost i s teoretskog i s računalnog aspekta. Pa ipak, kako autor zaključuje, nema razloga da se ove tehnike ne koriste ukoliko im se pristupi na kritički i informisan način (Rebor, 2023).

Ove tehnike, uz pomoć Labova nacionalnih biblioteka i njihovih digitalnih zbirki, mogu pružiti brojne mogućnosti da kulturna baština bude što vidljivija široj javnosti. Svaki istraživač/ica može za vlastita istraživanja iz oblasti digitalne humanistike slobodno koristiti modele, alate i aplikacije koje su razvile nacionalne biblioteke. No, gotovo po pravilu, projekti u ovoj oblasti su interdisciplinarni i u njih su uključeni brojni akteri, pa nacionalne biblioteke, kada rade vlastite projekte iz ove oblasti, to čine u saradnji s drugim organizacijama i institucijama. Jedan od takvih jeste inovativan projekt *Živjeti s mašinama (Living with Machines)* Britanske biblioteke, koji je završen prošle godine, a trajao je od 2018. Ovaj projekt je predvodila Britanska biblioteka u saradnji s Institutom Alan Turing i brojnim drugim akademskim partnerima, a njegov cilj je bio primijeniti napredne tehnologije mašinskog učenja i vještačke inteligencije kako bi se istražile i bolje razumjele društvene i kulturne promjene koje su se desile za vrijeme industrijske revolucije u Britaniji, odnosno tokom 19. stoljeća (Institut Alan Turing, s. a.).



Preuzeto s web-stranice projekta *Živjeti s mašinama* (LWM, 2024a)

Glavni ciljevi projekta bili su digitalizacija i pretvaranje historijskih dokumenata u digitalni format, uključujući novinske članke, službene dokumente, lične dnevnike i mnoge druge vrste tekstualnih izvora, kako bi se uz primjenu mašinskog učenja ti podaci analizirali da bi se identificirali obrasci i trendovi u društvu i ekonomiji tokom industrijske revolucije. Ovaj projekt je okupio ogroman broj stručnjaka i stručnjakinja iz raznih oblasti: historičare/ke, podatkovne znanstvenike/ce (*data scientists*), geografe i geografičarke, računalne lingviste i lingvistkinje, kustose/ice te bibliotekare/ke (Institut Alan Turing, s. a.). Glavni aspekt projekta bila je primjena tehnologija vještačke inteligencije i mašinskog učenja na podatke iz starih tekstova, što je omogućilo istraživačima i istraživačicama da analiziraju velike količine podataka na način na koji bi to tradicionalnim metodama bilo nemoguće. Na primjer, jedno od istraživanja unutar projekta obuhvatalo je analizu novinskih članaka iz 19. stoljeća kako bi se identificirale promjene u radnim uslovima i životnom standardima radnika. Korištenjem metoda mašinskog učenja, istraživači/ce su uspjeli prepoznati ključne teme kao što su radnički štrajkovi, uslovi rada u fabrikama i socijalne reforme, te analizirati kako su se oni razvijali kroz vrijeme.

Pored tog postignuća, kroz projekt su objavljene dvije knjige, napisano je oko četrdeset radova, poglavlja knjiga i izlaganja s konferencije, održano nekoliko izložbi, ali su razvijeni i setovi podataka, softverski paketi,



istraživački i vizualizacijski alati (LWM, 2024b). Ovaj popis reflektuje multidisciplinarnu prirodu projekta *Živjeti s mašinama* i pokazuje kako integracija historijskih i računalnih metoda može otvoriti nove pristupe razumijevanju historije.



Preuzeto iz finalnog izvještaja projekta (Ahnert i Demertzi, 2023)

I Nacionalna biblioteka Španije ima nekoliko zanimljivih projekata iz oblasti digitalne humanistike, a jedan o njih je svakako *ChefBNE*, dokumentarna serija o historiji španske gastronomije, koja predstavlja “kulinarsko putovanje kroz 12 starih knjiga recepata iz zbirke Nacionalne biblioteke Španije” (ChefBNE, s. a.). U ovom su projektu istraživači/ce i kuhari/ce zajedno oživjeli historiju španske kuhinje, od toga kako se nekada kuhalo, do toga koji su sastojci korišteni. Kako stoji na web-stranici projekta, s njim se željela nanovo iskoristiti digitalna zbirka iz oblasti gastronomije ove biblioteke (ChefBNE, s. a.). Također se radi o multidisciplinarnom projektu u kojem su, pored kuhara/ica i istraživača/ica, učestvovali i historičari/ke i bibliotekari/ke, kako bi zajedno reinterpretirali i učinili vidljivim jedan dio historije svoje zemlje (ChefBNE, s. a.). Na web-stranici projekta se može pronaći 12 videa u kojima se prave i kontekstualiziraju recepti otkriveni u zbirkama ove biblioteke.



ESPAÑOL ENGLISH

**Macarrones con chocolate**

Manuel de Comida  
CEO de Chocolates y Dulces Matías López

Aran Gozoaga  
Escritora y foodstylist

Julio Valles  
Presidente Academia Castellana  
y Leonesa de Gastronomía

María Ángeles Pérez Samper  
Historiadora

**Bizcocho helado**

Fernando Sáenz  
Chef del Otrador Grate

Carlos Caballero  
Arquitecto

Julio Valles  
Presidente Academia Castellana  
y Leonesa de Gastronomía

Sonia Fernández  
Arquitecta, OBAMA

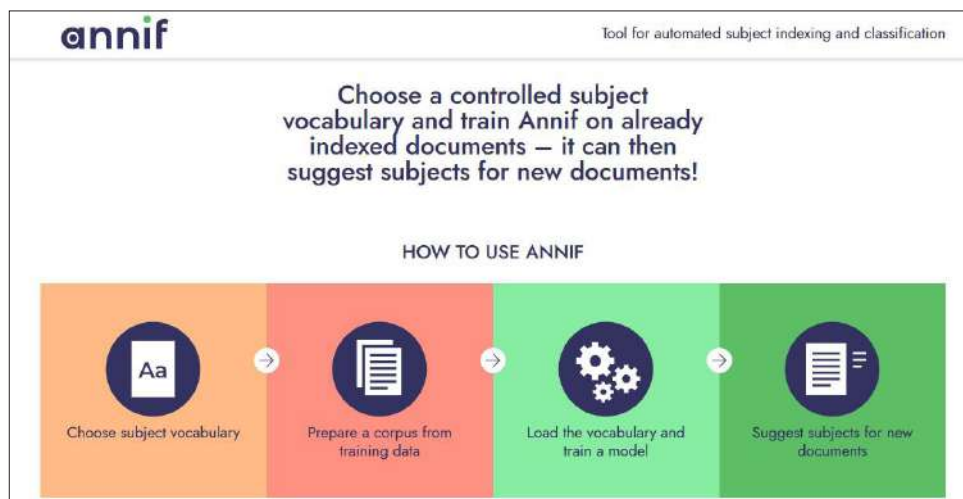
Preuzeto s web-stranice projekta (ChefBNE, s. a.)

## Automatsko klasificiranje i predmetno označavanje

Povećanje obima digitalnih biblioteka nije samo pogodovalo nastanku digitalnohumanističkih istraživanja već i potrebi za automatiziranjem stručne obrade građe koja je u digitalnim zbirkama biblioteka iz dana u dan postajala sve veća. Rad na razvijanju metoda za automatsko klasificiranje i predmetno označavanje, odnosno eksperimentisanje s automatskim dodjeljivanjem oznaka velikih bibliotečkih/bibliografskih klasifikacija poput KKB, DDK i UDK, ali i drugih klasifikacijskih sistema, posebno se intenziviralo u posljednje dvije decenije, a uveliko je olakšano pojavom modernih tehnologija vještačke inteligencije, a naročito mašinskog, odnosno dubokog učenja. U razvoju alata koji bi im u tome pomogli, ali i u njihovoj primjeni, svakako prednjače nacionalne biblioteke Finske i Njemačke.

Nacionalna biblioteka Finske je 2017. godine razvila softver za automatsko klasificiranje po imenu *Annif*, što je zapravo *Finna* unazad – *Finna.fi* je finski repozitorij online dostupnih izvora, bilo časopisa, fotografija, filmova ili mapa, u raznim formatima (Finna, 2024). Prema Suominen, O. (2019),

softveru je dato takvo *naopako* ime jer se tradicionalno indeksiranje željelo okrenuti naglavačke: “umjesto da se unese neka tema i kao odgovor dobije lista dokumenata o njoj, [uz Annif] unio bi se dokument, a kao odgovor bi bile ponuđene njegove najrelevantnije teme”. Kasnija verzija *Annifa* je “implementirana kao Python aplikacija koja koristi okvire Flask i Connexion za web-server i REST API funkcionalnost”, a predmetno indeksiranje i klasifikacija se obavljaju uz pomoć nekoliko algoritama koji se mogu koristiti sami ili kombinovati (Suominen, 2019); današnja verzija uključuje *TensorFlow*, *Omikuji*, *fastText* i *Gensim* (Annif, 2024). Upravo u tome leži prednost *Annifa*, kao i u tome što podržava tri jezika (finski, švedski i engleski) i što je softver otvorenog koda, a ponajviše što ne ovisi ni o kakvom kontrolisanom rječniku – rječnik dodaje korisnik/ca ovog alata i na njemu trenira model.

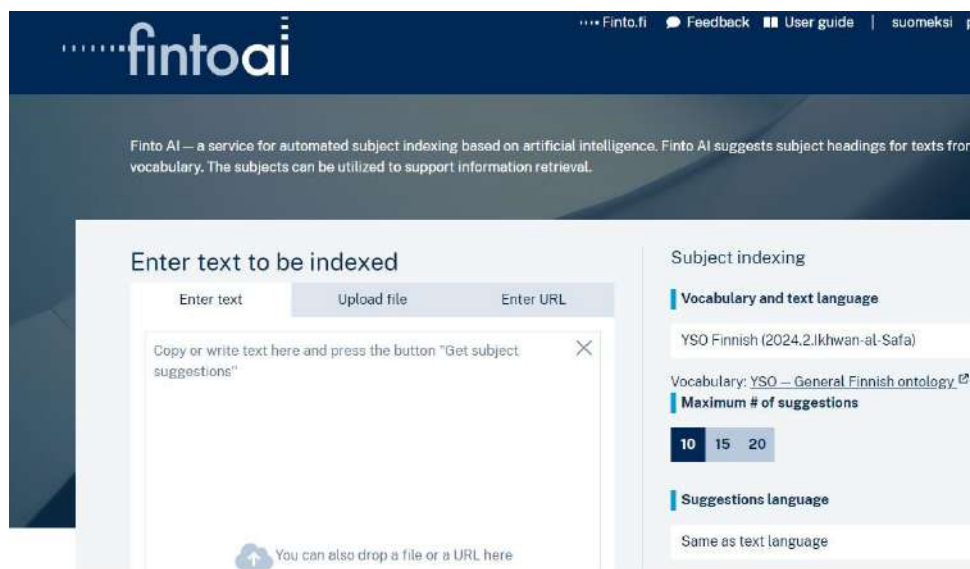


Preuzeto s *Annifove* web-stranice (Annif, 2024)

Danas ovaj vrlo obećavajući softver, pored Nacionalne biblioteke Finske, koriste i Nacionalna biblioteka Njemačke (DNB), Nacionalna biblioteka Švedske, Nacionalna biblioteka Nizozemske – kao dio većeg alata za automatsko katalogiziranje (Suominen et al., 2022) te finski izdavač *Kirjavälitys Oy* za kreiranje metapodataka za knjige koje trebaju izaći iz štampe, finska TV-kuća *Yle* za dodjeljivanje oznaka novinskim člancima, finski *Jyväskylä Univerzitet* za svoj digitalni repozitorij za predmetno indeksiranje teza i disertacija, francuski *Dissem.in* za kategorizaciju

akademskih članaka u repozitorijima s otvorenim pristupom, njemački *Informacijski centar za ekonomiju u Leibnitzu* za servis za automatsko indeksiranje *AutoSE* (Annif, 2024), kao i brojni drugi komercijalni i nekomercijalni korisnici/e.

Pored *Annifa*, Nacionalna biblioteka Finske je 2020. godine razvila i *Finto AI*, servis za automatsko predmetno označavanje baziran na *Annifu*, koji funkcionira tako što korisnik/ca na web-stranicu doda tekst ili dokument, a onda dobije predložene predmetne oznake za njega. Prijedloge daje na finskom, švedskom i engleskom jeziku i baziran je na opštoj finskoj ontologiji *YSO* (Finto AI, 2024).



Preuzeto s web-stranice Finto AI-ja (Finto AI, 2024)

Nacionalna biblioteka Njemačke (DNB) predstavlja jedinstven slučaj – jedina je koja želi u potpunosti automatizovati klasificiranje i predmetno označavanje svojih digitalnih, a vremenom i fizičkih zbirki, na čemu aktivno i radi. Ova zemlja od 2006. prikuplja svu digitalnu građu kao obavezan primjerak, pridodatu već postojećem nacionalnom korpusu koji uključuje svu tekstualnu, audio i vizualnu građu objavljenu u Njemačkoj ili na njemačkom jeziku (Junger, 2018). Kao i u ostalim bibliotekama, potreba za automatskim klasificiranjem nastala je proliferacijom digitalne građe te je DNB od 2010. godine počeo eksperimentirati s automatskim dodjeljivanjem predmetnih

odrednica, tj. automatskim indeksiranjem (koje ne vrše uz pomoć mašinskog učenja), a ujedno i s automatskim dodjeljivanjem predmetnih kategorija (opštih klasa) DDK-a (001-999) i skraćenih DDK klasifikacijskih oznaka, bez korištenja predubokih hijerarhijskih nivoa (npr. 651.532 – homeopatija) uz pomoć mašinskog učenja. Automatizacija klasifikacije se obavljala uz pomoć sofvera *Averbis*, prilagođenog potrebama DNB-a, a baziranog na SVM modelu (Junger, 2018), dok danas koriste *Annif*. Podaci na kojima se softver trenirao su bile online publikacije na engleskom i njemačkom jeziku, ali i digitalizirani sadržaji (*table of contents*) štampanih publikacija na navedenim jezicima, koji čine dvije trećine ukupnog broja podataka za obučavanje. Automatsko klasificiranje prema DDK klasifikacijskom sistemu u DNB-u ide u dva smjera:

- Automatsko dodjeljivanje predmetnih kategorija DDK: od 2012. godine svi online dokumenti (osim beletristike, poezije i dokumenata koji nisu na njemačkom ili engleskom jeziku) uključeni u nacionalnu bibliografiju Njemačke, a od 2017. i određena izdanja štampanog materijala i disertacija, dobivaju automatski dodijeljenu DDK predmetnu kategoriju (Junger, 2018). Ova biblioteka mjesečno automatski klasificira između 5.000 i 10.000 izvora;
- Automatsko dodjeljivanje DDK brojeva: od 2015. godine se potpuno automatski dodjeljuju skraćeni DDK brojevi medicinskoj grupi online građe, uz nastojanje da se ti skraćeni DDK brojevi dodjeljuju svim vrstama građe, pa i štampanoj, unutar svake oblasti. U 2017. godini skraćeni DDK se dodjeljivao “za oblast medicine, hemije, računarske znanosti, sociologije, disertacija iz medicinske oblasti i medicinske sive literature, također u štampanoj formi” (Junger, 2018). Do kraja navedene godine automatski je klasificirano 260.000 online publikacija i oko 2.500 štampanih (Junger, 2018).

Nakon što je započeo s automatskim dodjeljivanjem skraćenih DDK brojeva, DNB je prestao čak i s ručnim dodjeljivanjem dubljih oznaka, pa tako i ljudi i mašine dodjeljuju samo skraćeni DDK broj (Junger, 2018). Upravo je ovo DNB-ovo odustajanje od dodjeljivanja punih DDK brojeva i klasificiranje plitkim klasifikacijskim nivoima kritikovale EDUG (*European DDC Users Group*) stoga što oni smatraju da će DDK odigrati važnu ulogu u semantičkom webu, odnosno “u ovom prijelomnom periodu za bibliografsko katalogiziranje, kada se polako krećemo od MARC-a ka

rješenjima [koja podrazumijevaju] otvorene podatke” (Conradi, 2017). U odgovoru na otvoreno pismo koje je EDUG poslala DNB-u navedeno je da DNB nije u potpunosti odustao od DDK klasifikacije, ali da je anketa provedena među njihovim korisnicima i korisnicama pokazala da za njih DDK broj nije pretjerano relevantan u pretraživanju građe, a da su se predmetne odrednice našle na četvrtom mjestu po važnosti, te da ih je to navelo da se fokusiraju na indeksiranje uz pomoć predmetnih odrednica u *Integriranoj normativnoj bazi* (GND) (Junger, 2017) i da će, suprotno onome što EDUG smatra, upravo GND odigrati važnu ulogu u semantičkom webu “zbog podobne strukture normativnih zapisa GND-a, koji također omogućavaju povezivanje s drugim sistemima za organizaciju znanja, bilo da se radi o klasifikacijama ili standardizovanim rječnicima na njemačkom ili drugim jezicima. Uključujući povezivanje s DDK-om” (Junger, 2017). Jasno je da je DNB prilično samopouzdan u automatskom klasificiranju; u prilog im idu i rezultati kontrole kvalitete koju su vršili tokom godina. U 2022. godini, od ukupno 18% revidiranih automatski dodijeljenih DDK brojeva, 76% ih je imalo oznaku identičnu onoj ručno dodijeljenoj, a u određenim kategorijama, npr. u oblasti prava, prelazio je tačnost od 92% (Mödden, 2022). Ipak, u nekim kategorijama tačnost još uvijek nije bila na zadovoljavajućem nivou zbog nedovoljno velikog skupa podataka na kojima bi se algoritam obučavao, npr. Kategorija 980 – historija južne Amerike (Mödden, 2022), što je stalna boljka svih velikih hijerarhijskih klasifikacijskih sistema. Od 2022. godine, nakon što je ugašen softver *Averbis* kojim se DNB služio za automatsko klasificiranje, ova biblioteka svoj rad na automatskom klasificiranju implementira uz pomoć softvera *Annif* (Mödden, 2022), koji koristi sve više biblioteka, servisa i repozitorija. Jasno je da ovakvi projekti zahtijevaju uključenost stručnjaka i stučnjakinja iz raznih oblasti, jer, kako navodi Hajdarpašić, L. (2017), automatsko predmetno označavanje “može profitirati tek združivanjem znanja informatičara, lingvиста, bibliotekara i drugih strana. Tek se takvom sinergijom mogu unaprijediti metode automatskog predmetnog označavanja, osigurati prihvatljivi automatski generirani indeksni termini, to jest podržati procesi konsolidacije ogromne i rastuće količine elektronskih izvora informacija”.

U 2021. godini pokrenut je tekući projekt automatskog katalogiziranja u kojem će DNB nastojati da “istraži koje metode vještačke inteligencije mogu koristiti za mašinsku obradu i analizu tekstova na prirodnom jeziku, kako bi se dobili najpotpuniji i što tačniji kataloški podaci za opisivanje sadržaja”

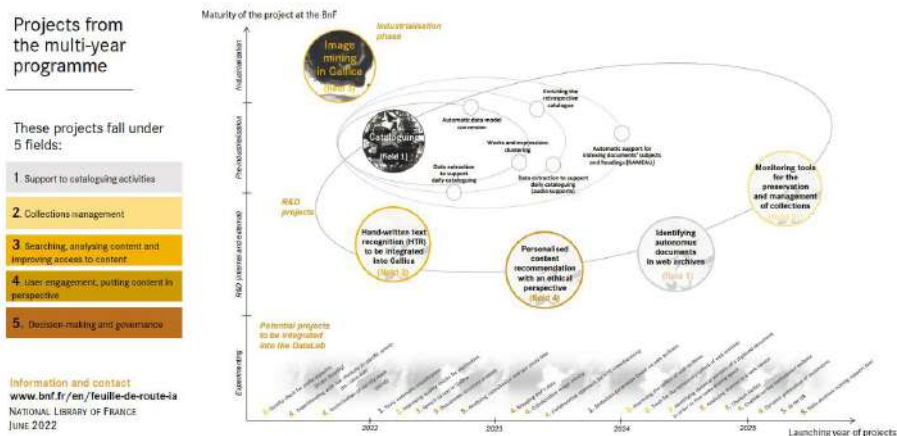


(DNB, 2022), a kako bi ostvarili kvalitetne semantičke veze između djela i predmetnih odrednica iz *Integrisane normativne baze* (GND) (DNB, 2022).



Preuzeto s web-strance DNB-a (DNB, 2022)

Što se drugih nacionalnih biblioteka u Evropi tiče, Nacionalna biblioteka Francuske ima namjeru automatizovati katalogiziranje uz pomoć vještačke inteligencije (BnF, 2024), a Nacionalna biblioteka Španije također planira u okviru *BNElaba* istražiti mogućnost automatskog klasificiranja i katalogiziranja (BNElab, 2024). Također, Nacionalna biblioteka Nizozemske u svom *Planu istraživanja 2023–2026.* navodi da će se, između ostalog, fokusirati na automatsko dodjeljivanje metapodataka na osnovu naslova publikacije (KBNL, 2023b).



Preuzeto s web-stranice Nacionalne biblioteke Francuske (BnF, s. a.)

## Inovativni načini pružanja usluga korisnicima/cama te ostale AI usluge



Preuzeto s postera (Love et al., 2023)

Pored navedenih načina upotrebe AI-ja u nacionalnim bibliotekama, postoje i razni drugi. Jedan od njih jeste pružanje usluga korisnicima i korisnicima uz pomoć chatbotova. Jedan takav značajan projekt jeste *CuratorBot* – chatbot koji je razvila Nacionalna biblioteka Nizozemske u saradnji s *TU Delft*, a koji sa posjetiocima razgovara o *Knjizi o ribama (Visboeck)* Adriaena Coenena, rukopisu s kraja 16. stoljeća (KBNL, 2024a). Kako navode na stranici Nacionalne biblioteke Nizozemske, ova knjiga sa brojnim ilustracijama morskih stvorenja i morskih čudovišta mnogima nije lako čitljiva i ostavlja otvorena pitanja o motivima autora (KBNL, 2024a). Međutim, uz *CuratorBot* koji se “hrani” raznolikim informacijama o ovom rukopisu, od onoga što se nalazi na webu do ekspertnog znanja stručnjaka i stručnjakinja iz Nacionalne biblioteke, on postaje mnogo jasniji (KBNL, 2024a) i posjetiteljima i posjetiteljicama se lakše upoznati s tekстом.

I Nacionalna biblioteka Luksemburga je krajem prošle godine počela koristiti chatbot, i to OpenAI-jev ChatGPT koji razumije francuski, njemački i engleski jezik, a internetskim korisnicima i korisnicama pomaže

u istraživanjima luksemburške historije, dajući odgovore na osnovu starih novinskih članaka (BnL, 2023). Kako se navodi u izvještaju ove biblioteke za 2023. godinu, “stručnjaci i stručnjakinje iz BnL-a su indeksirali digitalizirane luksemburške dokumente i pripremili moćnu bazu podataka koja omogućava semantičko pretraživanje” (BnL, 2023).

Pored korištenja chatbotova za pružanje usluga korisnicima i korisnicama, brojne nacionalne biblioteke, između ostalih Francuske, Španije i Švedske, kao dio svojih usluga nude mogućnost personaliziranih sistema za preporučivanje građe. Također, ono što je zanimljivo jeste da Nacionalna biblioteka Francuske tematski pobire i arhivira svu građu na webu vezanu za AI kao dio svoje misije obaveznog primjerka. Na primjer, u ljeto 2021. alati za indeksiranje ove biblioteke obradili su više od 700 web-stranica i Twitter-naloga, po odabiru osoblja, što iznosi više od 10 miliona URL-ova (BnF, 2024).

## Umjesto zaključka

Ovi projekti ilustriraju raznolike mogućnosti primjene AI, od automatskog dodjeljivanja metapodataka i digitalizacije dokumenata do poboljšanja korisničkog iskustva i podrške u istraživanju. Nacionalne biblioteke Njemačke i Finske uistinu prednjače u primjeni mašinskog učenja u automatskom klasificiranju i predmetnom označavanju, dok one Švedske i Norveške pronalaze brojne primjene jezičnih modela svojih maternjih jezika koje su same razvile. Također, Nacionalna biblioteka Španije svoju nacionalnu bibliografiju objavljuje u mašinski čitljivim formatima, a Britanska biblioteka je nedavno okončala ogroman digitalnohumanistički projekt *Živjeti s mašinama* koji pokazuje kako AI može pomoći u analizi historijskih podataka i omogućiti dublje razumijevanje kulturnih i društvenih promjena. Upotreba AI-ja za tekstualno i rudarenje podataka u Nacionalnoj biblioteci Francuske ili kao chatbotova u nacionalnim bibliotekama Luksemburga i Nizozemske dodatno naglašava važnost ovih tehnologija u modernom istraživanju. Međutim, ovi projekti sa sobom nose i izazove kao što su pitanje etičnosti, potreba za velikim količinama podataka, kompleksnost njihove implementacije, a svakako i sredstva kojima se oni podržavaju. Pa ipak, uprkos tim izazovima, dosadašnje uspješne integracije AI-ja u nacionalnim bibliotekama dokaz su potencijala ovih tehnologija da transformiraju način na koji pristupamo i istražujemo našu kulturnu baštinu.

U budućnosti, kako se AI tehnologije nastavljaju razvijati, možemo očekivati još inovativniju primjenu ovih tehnologija u bibliotekama, što će dodatno obogatiti našu sposobnost da očuvamo, istražujemo i cijenimo kulturnu baštinu. Nacionalne biblioteke imaju potencijal da postanu vodeće institucije u primjeni AI u očuvanju i prezentaciji kulturnog nasljeđa, stvarajući pritom nove standarde i modele za ostale kulturne institucije širom svijeta. Nastavak istraživanja i razvoj u ovom polju ključni su za osiguranje da ove tehnologije budu korištene na način koji maksimizira njihove prednosti, dok se istovremeno minimiziraju potencijalni rizici i izazovi.

## Literatura

- Ahnert, R. i Demertzi, L. (2023). *Living with Machines Final Report*. London: The Alan Turing Institute. <https://bl.iro.bl.uk/concern/reports/3a9af031-1ee1-4299-afef-12513e2ee3e4>
- Ames, S. i Lewis, S. (2020). Disrupting the library: Digital scholarship and Big Data at the National Library of Scotland. *Big Data & Society* 7(2). <https://doi.org/10.1177/2053951720970576>
- Annif*. (2024). <https://annif.org>
- BL – British Library. (s. a.). *Knowledge Matters: The British Library 2023–2030*. <https://www.bl.uk/about-us/Knowledge-Matters-British-Library-Strategy-2023-30.pdf>
- BNElab – El proyecto de reutilización digital de la Biblioteca Nacional de España. (2024). *Inteligencia Artificial*. <https://bnelab.bne.es/proyecto/inteligencia-artificial/>
- BNElab – El proyecto de reutilización digital de la Biblioteca Nacional de España. (s. a.). *Bibliografía Española*. <https://bnelab.bne.es/dato/bibliografia-espanola/>
- BnF – Bibliothèque nationale de France. (2024). *BnF and Artificial Intelligence*. <https://www.bnf.fr/en/artificial-intelligence-bnf#bnf-resources>
- BnF – Bibliothèque nationale de France. (s. a.). *BnF Roadmap on AI, 2021–2026*. [https://www.bnf.fr/sites/default/files/2022-01/Poster\\_AI%20Roadmap\\_BnF\\_202112.pdf](https://www.bnf.fr/sites/default/files/2022-01/Poster_AI%20Roadmap_BnF_202112.pdf)
- BnL – Bibliothèque Nationale du Luxembourg. (2023). *Rapport d'activité 2023 Bibliothèque Nationale du Luxembourg*. <https://bnl.public.lu/dam-assets/publications/rapports/2023-rapport-annuel.pdf>

- Börjesson, L. et al. (2024). *Transfiguring the Library as Digital Research Infrastructure: Making KBLab at the National Library of Sweden*. <https://www.diva-portal.org/smash/get/diva2:1745880/FULLTEXT01.pdf>
- Brygfjeld, S. A. et al. (2018). Machine Learning for Production of Dewey Decimal. U: *IFLA WLIC 2018*. Kuala Lumpur, Malezija. <https://library.ifla.org/id/eprint/2216/1/115-brygfjeld-en.pdf>
- CILIP – Chartered Institute of Library and Information Professionals. (2021). *Research report: The impact of AI, machine learning, automation and robotics on the information professions*. <https://www.cilip.org.uk/page/researchreport>
- ChefBNE. (s. a.). <https://chefbne.bne.es>
- Conradi, E. (2017). *DDC and automatic classification*. <https://edug.pansoft.de/tiki-index.php?page=DDC+and+automatic+classification>
- Cordell, R. (2020). *Machine Learning + Libraries: A Report on the State of the Field*. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf?loclr=blogsig>
- DNB – Deutsche Nationalbibliothek. (2021). *Changes to subject cataloguing procedures at the German National Library from 1 July 2019*. [https://www.dnb.de/SharedDocs/Downloads/EN/Professionell/Erschliessen/veraenderungenInhaltsererschliessungDnbJuli2019.pdf?\\_\\_blob=publicationFile&v=7](https://www.dnb.de/SharedDocs/Downloads/EN/Professionell/Erschliessen/veraenderungenInhaltsererschliessungDnbJuli2019.pdf?__blob=publicationFile&v=7)
- DNB – Deutsche Nationalbibliothek. (2022). *Automatic Cataloguing System*. <https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/KI/KI.html>
- DNB – Deutsche Nationalbibliothek. (2023). *Netzwerk maschinelle Verfahren in der Erschliessung*. <https://wiki.dnb.de/display/FNMVE/Netzwerk+maschinelle+Verfahren+in+der+Erschliessung>
- Finna. (2024). <https://finna.fi>
- Finto AI. (2024). <https://ai.finto.fi/?locale=en>
- Hajdarpašić, L. (2017). Razvoj i izazovi metoda automatskog predmetnog označavanja. *Bosniaca* 18(18), 12-18. <https://bosniaca.nub.ba/index.php/bosniaca/article/view/29/29>
- Haffenden, C. et al. (2021). Making and Using AI in the Library: Creating a BERT Model at the National Library of Sweden. *SocArXiv*, January 12. <https://osf.io/preprints/socarxiv/k9duq>
- IFLA – International Federation of Library Associations and Institutions. (2020). *IFLA Statement on Libraries and Artificial Intelligence*.

- <https://repository.ifla.org/server/api/core/bitstreams/8d21df6c-e61b-41a5-b089-0353a5552dbd/content>
- Institut Alan Turing. (s. a.). *Living with Machines: Transforming the Way We Study the Past*. <https://www.turing.ac.uk/research/research-projects/living-machines>
- Junger, U. (2017). *EDUG statement regarding future subject cataloguing at the DNB*. <https://edug.pansoft.de/tiki-index.php?page=DDC+and+automatic+classification>
- Junger, U. (2018). Automation first – the subject cataloguing policy of the Deutsche Nationalbibliothek. U: *IFLA WLIC 2018*. Kuala Lumpur, Malezija. <https://library.ifla.org/id/eprint/2213/1/115-junger-en.pdf>
- KB – Kungliga biblioteket. (2020). *Tillgängliggör kraftfulla modeller för språkförståelse*. <https://www.kb.se/samverkan-och-utveckling/nytt-fran-kb/nyheter-samverkan-och-utveckling/2020-02-04-kb-tillgangliggor-kraftfulla-modeller-for-sprakforstaelse.html>
- KB – Kungliga biblioteket. (2022). *Digital humaniora eller humanistisk datavetenskap?* <https://www.kb.se/hitta-och-bestall/samlingsbloggen/blogginlagg/2022-04-14-digital-humaniora-eller-humanistisk-datavetenskap.html>
- KBNL – De nationale bibliotheek van Nederland. (2023a). *10.3 million euros to unlock cultural heritage with AI*. <https://www.kb.nl/en/news/103-million-euros-unlock-cultural-heritage-ai>
- KBNL – De nationale bibliotheek van Nederland. (2023b). *Research Agenda 2023–2026*. [https://www.kb.nl/sites/default/files/documents/onderzoeksagenda\\_2023-2026\\_in\\_opmaak\\_eng.pdf](https://www.kb.nl/sites/default/files/documents/onderzoeksagenda_2023-2026_in_opmaak_eng.pdf)
- KBNL – De nationale bibliotheek van Nederland. (2024a). *Artificial intelligence: find out how the KB handles AI*. <https://www.kb.nl/en/over-ons/expertises/artificial-intelligence>
- KBNL – De nationale bibliotheek van Nederland. (2024b). *Tools*. [https://lab.kb.nl/products/product\\_type/tool](https://lab.kb.nl/products/product_type/tool)
- KBNL – De nationale bibliotheek van Nederland. (s. a.). *Opening of the KB Datalab*. <https://lab.kb.nl/news/opening-kbdatalab>
- Kummervold, P. E. et al. (2021). *Operationalizing a National Digital Library: The Case for a Norwegian Transformer Model*. <https://arxiv.org/pdf/2104.09617>
- Love, J. et al. (2023). *The Visboeck Curatorbot: Discussing Heritage Collections with Machines*. <https://repository.ifla.org/handle/123456789/3058>
- LWM – Living With Machines. (2024a). <https://livingwithmachines.ac.uk>



- LWM – Living With Machines. (2024b). *Final Report And Case Study Videos*. <https://livingwithmachines.ac.uk/achievements/>
- Magueresse, A. et al. (2020). *Low-resource Languages: A Review of Past Work and Future Challenges*. <https://arxiv.org/pdf/2006.07264>
- Mödden, E. (2022). Artificial intelligence, machine learning and bibliographic control. DDC Short Numbers – Towards machine-based classifying. *Journal of Library and Information Science* 13(1), 256-264. <https://www.torrossa.com/it/resources/an/5111773>
- NB – Nasjonalbibliotekets. (2024a). *AI Lab*. <https://ai.nb.no>
- NB – Nasjonalbibliotekets. (2024b). *Applications*. <https://ai.nb.no/applications/>
- NLS – National Library of Scotland. (s. a.). *Metadata Collection*. <https://data.nls.uk/data/metadata-collections/>
- NJV – Njemačka vlada. (2020). *Artificial Intelligence Strategy of the German Federal Government: 2020 Update*. [https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung\\_KI-Strategie\\_engl.pdf](https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf)
- Padilla, T. (2018). Collections as data: Implications for enclosure. *C&RL News* 79(6), 296-300. <https://crln.acrl.org/index.php/crlnews/article/view/17003/18740>
- Padilla, T. et al. (2019). *Final Report – Always Already Computational: Collections as Data*. <https://zenodo.org/records/3152935>
- Rebora, S. (2023). Sentiment Analysis in Literary Studies. A Critical Survey. *DHQ* 17(2). <https://www.digitalhumanities.org/dhq/vol/17/2/000691/000691.html>
- Salian, I. (2023). Booked for Brilliance: Sweden’s National Library Turns Page to AI to Parse Centuries of Data. *Nvidia*, 23. 1. 2023. <https://blogs.nvidia.com/blog/sweden-library-ai-open-source/>
- Suominen, O. (2019). Annif: DIY Automated Subject Indexing Using Multiple Algorithms. *Liber Quarterly* 29. <https://liberquarterly.eu/article/view/10732/11613>
- Suominen, O. et al. (2022). Annif and Finto AI: Developing and Implementing Automated Subject Indexing. *JLIS.it* 13(1). <https://www.jlis.it/index.php/jlis/article/view/437/430>
- Tabak, E. (2017). *Digitalna humanistika: debate, aplikacije, izazovi*. Zenica: Filozofski fakultet Univerziteta u Zenici. [https://unze.ba/digitalna\\_humanistika/digitalna\\_humanistika.pdf](https://unze.ba/digitalna_humanistika/digitalna_humanistika.pdf)

# Do Androids Dream Of Electronic Librarians? Application of AI Technologies in the National Libraries Across Europe

Dalila Mirović

[dalila.mirovic@gmail.com](mailto:dalila.mirovic@gmail.com)

**Goal:** The goal of this research is to provide an overview of the application of artificial intelligence (AI) technologies in national libraries across Europe, with a focus on the use of machine learning for (semi)automated classification and indexing, digital humanities projects, and innovative methods for the preservation and presentation of cultural heritage. The study aims to encourage interest among the professional and academic librarian communities in the possibilities offered by these technologies.

**Design/methodology/approach:** This research employs a descriptive method to present the application of AI technologies in national libraries. Comparative, analytical, and synthetic methods were used to inductively develop theoretical frameworks related to the subject of the research. Examples were analyzed from the practices of national libraries in Finland, Sweden, Norway, Germany, Spain, France, Netherlands, Scotland, Luxembourg and Great Britain.

**Results:** Findings of the research indicate that national libraries are utilizing AI technologies for automated classification, indexing, language model development, and digital humanities projects. National libraries in Finland and Germany are leading in the implementation of machine learning for automatic classification and subject indexing, while projects like the British Library's "Living with Machines" showcase innovative AI applications in researching cultural change.

**Discussion:** The implementation of AI technologies in national libraries has numerous advantages; their AI labs provide a platform for the development and testing of new technologies, while language models they develop enable deeper and

richer interaction with collections. Digital humanities projects demonstrate how AI can aid in the preservation and exploration of cultural heritage, but they require an interdisciplinary approach and collaboration among various experts.

**Originality/value:** This research is among the first to examine the application of AI technologies in national libraries across Europe, offering a comprehensive overview of the current state and providing a foundation for future research in this field. The study demonstrates how AI technologies can transform the operations of national libraries and enhance the preservation and accessibility of cultural heritage.

**Keywords:** National Libraries, European National Libraries, AI technologies, AI Labs, Automatic Classification, Automatic Subject Indexing, Annif, Digital Libraries, Digital Humanities