

## Susietieji duomenys ir dirbtinis intelektas kaip pagrindiniai techninių paslaugų transformacijos elementai

Philip E. Schreur  
Stanfordo universiteto (JAV) techninių paslaugų departamentas

<https://doi.org/10.1080/01639374.2020.1772434>

Straipsnyje kalbama apie MARC formatus naudojančių bibliotekų perėjimą prie semantinio saityno aplinkos, susietųjų duomenų ir dirbtinio intelekto paslaugų

Pateikiame straipsnio ištraukų vertimą

Per pastaruosius 50 metų technines paslaugas palankiai veikė daugelis paskatų. Perėjimas nuo kortelių katalogo prie elektroninio, nors iš pradžių bauginantis, tobulėjant technologijoms, pasirodė esąs labai naudingas bibliotekų paslaugų vartotojams ir atvėrė nuolat besiplečiantį galimybių spektrą. Skaitmeninis turinys, iš pradžių laikytas pakeisiančiu spausdintinį, leido maksimaliai naudotis gausybe papildomų šaltinių ir taikyti ištiesą spektrą analizės metodų tobulinant mokymą ir mokymąsi. Panašiai ir susietieji duomenys, tiksliau susietieji atvirieji duomenys, ir dirbtinis intelektas žada daug galimybių nenuspėjama (nors ir pavėluotais) būdais transformuoti žinių valdymą. <...> Bet kodėl bibliotekos, ir ypač jų techninės paslaugos, turėtų žengti šį milžinišką žingsnį už taip jų sėkmingai naudotų MARC formatų ribų?

Viso pasaulio bibliotekos didžiosios dalies savo bibliografinių ir autoritetinių duomenų sklaidai bei raiškai naudoja MARC besiremiančias sistemas. MARC remiasi taip pat ir pagrindinės katalogavimo priemonės (pvz., *OCLC Connexion* kliento programa). MARC formatu bibliografinius duomenis teikia pardavėjai, pvz., *OASIS*, *MARCnow*. Paieškos aplinkos, pvz., *Blacklight*, remiasi *Solr* rodyklėmis, kurios duomenis ima iš MARC laukų ir polaukių. Tačiau MARC yra ankstyvasis komunikacinis formatas, sukurtas 1960 m. siekiant mašininio būdu apdoroti bibliografinius duomenis, kurie anksčiau buvo teikiami katalogo kortelėse. Naudojant šį formatą, viename ir tame pačiame įrašė sudėtinga išreikšti įvairių duomenų elementų ryšius <...>, nes laikoma, kad žmogus analizuos įrašą kompiuterio ekrane ir pats nustatys šiuos ryšius.

MARC buvo didelis pasiekimas, leidęs bibliotekoms nebenaudoti kortelių katalogų <...>. Jis leido plėtoti integralias bibliotekų sistemas ir taupyti bibliotekų išteklių komplektavimui, katalogavimui ir paieškai skirtas lėšas. Tačiau atsiradus saitynui, kaip naujai *lingua franca*, bibliotekos pasijuto esančios vis labiau izoliuotos. Todėl bibliotekoms bandant savo metaduomenis perkelti į susietaisiais duomenimis besiremiančią aplinką, <...> ankstesnis pasiklojimas žmogaus atliekamu teisingu duomenų elementų ryšių interpretavimu tapo didele problema. Nors MARC metaduomenys gali būti konvertuojami į susietuosius duomenis, daugelis žmogaus identifikuojamų ryšių naujojoje aplinkoje nėra išreiškiami.

<...> Tarkime, kad toks perėjimas nuo izoliuotos, t. y. MARC, aplinkos prie semantinio saityno yra teisingas sprendimas. Tuomet kyla klausimas: kodėl susietieji duomenys? Pirmiausia, akivaizdu, kad nors bibliotekų paslaugų vartotojai ir naudojami specializuotomis bibliotekų katalogų sėkmingos paieškos priemonėmis, informacijos ieško taip pat ir saityne. Nors akademinuose sluoksniuose tam tikrų šaltinių (pvz., Vikipedijos) patikimumu abejojama, esama priemonių (pvz., *Google Scholar*), teikiančių visuotinę mokslinių publikacijų prieigą. <...> Kitas semantinio saityno privalumas yra ypatingas dėmesys įvairialypių išteklių sąvokų ir esybių

siečiai. Trečias privalumas yra tai, jog saitynas yra tarptautinė aplinka. Pereidamos prie susietųjų duomenų, pasaulio bibliotekos gali pasinaudoti bibliografiniais ir autoritetiniais metaduomenimis, kuriamais daugelio šalių bibliotekų ir teikiama kaip susietieji duomenys. Ir galiausiai, saitynas yra nuolat besivystanti aplinka. Neabejotina, kad susietuosius duomenis kada nors pranoks koks nors kitoks standartas. Siekdamas dalyvauti šioje transformacijoje, bibliotekos turi žengti į saityno duomenų aplinką.

Dirbtinio intelekto tradiciniams bibliotekų procesams teikiamos paskatos yra kitokios. 2017 m. publikuotame straipsnyje Frey'us ir Osborn'as teigia, kad tikimybė, jog techninio bibliotekų personalo darbą pakeis kompiuterinės sistemos yra 99 proc., bibliotekinių padėjėjų – 95 proc., o bibliotekininkų – 65 proc. Ką tai reiškia iš tikrųjų? Paprasčiausiai tai, ką daro šios darbuotojų grupės, galėtų atlikti dirbtinis intelektas.

<...> Susietieji duomenys jau daro didelę įtaką techninių paslaugų procesams. 2011 m. Kongreso biblioteka pradėjo plėtoti BIBFRAME – RDF besiremiančią ontologiją, kuri turėtų pakeisti MARC formatus.

<...> Netenka abejoti, kad bibliotekų perėjimas prie susietųjų duomenų bus panašus į perėjimą nuo kortelių prie elektroninio katalogo – abiem atvejais su „ilgos uodegos“ (angl. „long-tail“) efektu. Vien tik pripažinimas, kad naujasis modelis (MARC, susietieji duomenys, dirbtinis intelektas) yra veiksmingas, perėjimo nepagreitina dėl tam tikrų probleminių dalykų, susijusių su infrastruktūra, aplinka, darbuotojais. OCLC nenustojo spausdinti katalogo kortelių iki 2015 m. spalio 1 d. – iki tol jų buvo išspausdinta beveik du milijardai. Lėtą bibliotekų perėjimą nuo popierinių katalogų prie elektroninių duomenų <...> lydėjo naujų technologijų bei duomenų ir darbo vietų praradimo baimė. Net jeigu imtume transformuoti savo metaduomenis į susietuosius duomenis, daugelis tiekėjų daugelį metų vis dar siūlytų duomenis MARC formatais ir daugelis bibliotekų partnerių naudotų MARC. Pavyzdžiui, Prancūzijos nacionalinė biblioteka artimiausiu metu vis dar naudos INTERMARC ir nemato prasmės pereiti prie susietųjų duomenų. Ji į susietuosius konvertuos tik metaduomenis, siunčiamos už bibliotekos ribų, o vidaus naudojimui skirtus duomenis teiks INTERMARC formatu.

<...> MARC naudojamas ne tik metaduomenims kurti ir jais dalintis. Daugelis bibliotekų funkcijų remiasi paslaugomis, susijusiomis su MARC laukų, polaukių ir indikatorių teikimo gudrybėmis. Viskas, pradėdant sudėtinga vidine analize ir baigiant paieška, priklauso nuo šių laukų, polaukių ir jų elementų naudojimo bendrose sistemų struktūrose. Semantinis saitynas ir jo duomenų grafai yra nepralengiama priemonė, padedanti vartotojams atlikti sėkmingą paiešką ir analizuoti. Naudojant MARC ir sąryšines duomenų bazes, gali būti paprasčiau atliekami veiksmai, nereikalaujantys sudėtingo duomenų tinklo (pvz., mokėjimai).

<...> Papildant įrašus nauja informacija, gali susijungti arba pasidalyti asmenims arba organizacijoms sudaryti pradmenys. Plečiantis monografinių leidinių serijai, plečiasi ir jai sudaromų įrašų turinys. Metaduomenys yra gyvas reiškinys; visi minėti pokyčiai atspindimi naujinant MARC. Šie pokyčiai turi atsispindėti ir po konvertavimo [saityne] pateiktuose susietuosiuose duomenyse. Svarbu tai, ar bendra sistema leidžia duomenų lygmeniu teikti kilmės informaciją, kad duomenimis besikeičiantys subjektai galėtų priimti arba nepriimti padarytų pakeitimų. Ar pakanka darbuotojų šiems pakeitimams įvertinti? Turi būti labai kruopščiai pasirengta susietųjų duomenų kontekste, ypač bendroje aplinkoje, atlikti [įprastus] esamos MARC aplinkos veiksmus (pvz., naujinimą, taisymą, šalinimą).

Konvertuojant <...> duomenis, svarbu pasirinkti ontologiją. Saitynas pagal savo pobūdį yra decentralizuota ir kintanti sistema. Bibliotekos, savo ruožtu, teikia pirmenybę nuoseklumui <...>. Šiuo metu bene geriausiai MARC formatu įrašytus duomenis atspindi Kongreso bibliotekos sukurta ontologija BIBFRAME, tačiau esama ir kitų <...> standartų (pvz., *schema.org.*, *Dublin Core*, *CIDOC CRM*).

<...> Kitas aspektas, kurį būtina iš naujo įvertinti, yra kontroliuojamųjų pradmenų (arba „autoritetinių duomenų“, kaip jie vadinami bibliotekų) kūrimas ir taikymas. Tradiciškai bibliotekos naudoja autoritetinius duomenis teikdamos kontroliuojamuosius pradmenis MARC įrašuose, tačiau autoritetinių įrašų rengimas yra sudėtingas procesas, atliekamas siauros bibliotekininkų kategorijos. Šie autoritetiniai duomenys ne vien tik identifikuoja esybę (pvz., asmenį) – jie teikia aprobuotąsias vardo formas, abipuses nuorodas, (sudėtinių kolektyvų atveju) nuorodas į ankstesnius ir vėlesnius pradmenis; tai – neįkainojamas informacijos šaltinis. Bibliotekoms transformuojant šiuos savo veiklos procesus, teks spręsti daug probleminių dalykų.

Pirmiausia, tradiciniai bibliotekų autoritetiniai duomenys susietųjų duomenų kontekste negali būti naudojami kaip identifikuojantys esybę <...>. Autoritetiniai duomenys aprašo žmones, vietas, renginius. Aprašydami įvykį, jie neidentifikuoja asmens kaip fizinės būtybės. Kadangi tik pats žmogus gali atlikti veiksmą, pavyzdžiui, būti autoriumi, prisireikia kitokio identifikatoriaus, žeminčio asmenį kaip Realaus pasaulio subjektą.

<...> Nepaisant iššūkių, išskylančių pereinant nuo bibliotekų metaduomenų prie susietųjų duomenų, ir būtinybės permąstyti pagrindinius veiklos procesus (katalogavimą, autoritetinių duomenų teikimą ir kt.), neabejotina, kad toks perėjimas būtų nepaprastai naudingas – sudarnintų bibliotekų metaduomenis ir semantinį saityną. <...> pereiti nuo kontroliuojamųjų kreipties elementų prie identifikatorių daugeliu atžvilgiu nėra sudėtinga. Ir vieni, ir kiti skirti atlikti sieties funkciją. <...>. Be to, ir vieni, ir kiti reikalauja žmogaus darbo teikiant ir tvarkant.

<...> Užtat tikimasi, kad dirbtinis intelektas sumažins žmogaus indėlį kuriant ir prižiūrint metaduomenis. Pasak Petero Fernandezo, „Bibliotekoms ne tiek yra aktualu, kokios technologijos bus paveiktos [dirbtinio intelekto], bet kokia technologija (jei iš viso tokia bus), bus nepaveikta“. Bibliotekų požiūris į tai turėtų būti ne negatyvus, o racionalus. Daugelis [bibliotekų veiklos procesų] (ypač metaduomenų priežiūros) yra šabloniški ir pasikartojantys. Dirbtinis intelektas leistų bibliotekoms išlaisvinti darbuotojų laiką ir veiksmingiau panaudoti jį nuolat besikeičiančioje aplinkoje.

<...> Be įprastos tradicinių išteklių (pvz., monografinių ir serialinių leidinių) prieigos, bibliotekoms vis dažniau tenka teikti ir netradicinių išteklių (pvz., nestruktūrinio teksto) prieigą. Tokio nestruktūrinio teksto ir netgi kaip išsamus tekstas teikiamų tradicinių išteklių (pvz., mokslinių žurnalų straipsnių) gausa daro metaduomenų kūrimą nepraktišku. Tokia lanksti priemonė kaip dirbtinis intelektas <...> gali tapti priemone šiam probleminiam aspektui išspręsti.

<...> Netradicinį skaitmeninio teksto naršymo metodą siūlo palyginti nauja dirbtiniu intelektu, kompiuterine lingvistika, tinklo teorija ir mašininis mokymusi besiremianti grafinę naudotojo sąsają turinti paieškos aplinka *Yewno*. Aplinka padeda naudotojams naršyti sąvokas ir sąsajas. <...> iš tikrųjų kuo daugiau teksto esama, tuo sėkmingiau išrenkamos sąvokos. <...> *Yewno* nepriklauso nuo žmogaus sukurtų metaduomenų.

<...> Pereidamos prie susietųjų duomenų ir semantinio saityno kalbos, bibliotekos išlaisvina savo metaduomenis iš tik joms suprantamų formatų (MARC) varžtų <...>.

<...> Paieškos ir prieigos sfera <...> transformuojasi. Konvertuodamos (katalogų) metaduomenis į susietuosius duomenis, bibliotekos gali teikti savo tradicinių išteklių prieigą semantiniame saityne. Tačiau norėdamos kokiu nors būdu teikti nuoseklią kontroliuojamąją nestruktūrinių duomenų ir gausėjančių visateksčių išteklių prieigą, bibliotekos turi, be tradicinių katalogavimo būdų, naudoti ir inovatyvias priemones (dirbtinį intelektą).