



LIETUVOS NACIONALINĖ
MARTYNO MAŽVYDO
BIBLIOTEKA

https://www.ifla.org/files/assets/big-data/publications/a_concept_data_science_framework_for_libraries.pdf

IFLA didžiųjų duomenų speciali interesų grupė

2018 m. spalio

Bibliotekoms skirtas koncepcinis duomenų analizės modelis

Pateikiame straipsnio ištraukų vertimą

Bibliotekoms iškyla būtinybė taikyti naujus paslaugų modelius, padedančius transformuoti duomenis į informaciją. Joms būtina tobulinti savo technologinį raštingumą ypač susijusį su kodavimu ir ženklinimu, aktyviai naudotis galimybėmis, kurias teikia sparčiai besivystanti bibliotekų duomenų analizės sritis <...>. Tačiau [tai] nereiškia, kad bibliotekos pačios kuria ar valdo didžiuosius duomenis, kaip jie tradiciškai suprantami. Tiesiog pripažįstama, kad duomenų gausos pripildytame pasaulyje <...> bibliotekos būtina „nusimanyti“ apie duomenis.

Šiandien duomenų analizė suprantama kaip kompiuterinio programavimo, programinės įrangos inžinerijos ir statistikos sričių bei patirties tam tikroje kitoje srityje derinys. <...>. Dviejų nedidėles apimties tyrimų <...> buvo nustatyta duomenų analizės pareigybinių funkcijų grupė <...>. Tai yra šešios funkcijos: archyvaras, duomenų saugotojas, duomenų bibliotekininkas, duomenų analitikas, duomenų inžinierius ir duomenų žurnalistas. <...>. Duomenų bibliotekininkai susipažįsta su duomenų rinkiniais, suvokia techninius metodus ir glaudžiai bendradarbiauja su mokslininkais.

Plačiąja prasme duomenys skiriasi savo forma (kokybiniai ir kiekybiniai), struktūra (struktūriniai, nestruktūriniai, pusiau struktūriniai), šaltiniais (užfiksuoti, išvesti, <...>, trumpalaikiai), kūrėju (pirminiai, antriniai, tretiniai), tipu (identifikuojamieji, aprašomieji, metaduomenys). Paprastai tariant, jeigu kas nors sako, kad kas nors yra duomenys, reiškia tai yra duomenys. Svarbu, *kada* kas nors yra duomenys, o ne *kas* yra duomenys (t. y. duomenys yra takūs, o ne sustingę kaip knyga). Netgi klaidingi duomenys yra duomenys. Bibliotekoms tokia perspektyva teikia galimybę tolti nuo taksonomija besiremiančios koncepcijos, susijusios su artefaktų, kaip objektų, klasifikavimu.

<...> bibliotekos jau nebėra vien tik duomenų naudotojos, bet ir jų tiekėjos bei, nors ir mažesniu mastu, jų kūrėjos. <...> Nustatant patikimus duomenų šaltinius, svarbiausiais kriterijais yra duomenų integralumas ir duomenų etika. Privatumo, etikos ir nešališkos informacijos prieigos principai visada buvo svarbūs bibliotekoms – tai joms teikia galimybę būti unikaliomis mokslininkų ir tyrėjų partnerėmis bei kokybiškų duomenų šaltiniais.

Dirbtinis intelektas jau nebesuvokiamas kaip „dar tik ištiksiantis šokas“. Kaip ir kitos sritys, bibliotekos jau dabar turėtų numatyti, kaip dirbtinis intelektas paveiks informacijos profesinę sritį.

<...> bibliotekų saugomi duomenys nėra didieji duomenys per se. Antra vertus, bibliotekoms būtų naudinga į duomenis pažvelgti labiau koncepciniu požiūriu. Toks požiūris sako, kad nėra taip svarbu, ar duomenys yra didieji, ar, priešingai, „mažieji“ (t. y. duomenys, kuriuos galima apdoroti neįmantriomis priemonėmis). Nuo didžiųjų duomenų vis labiau darosi priklausoma analitika, o didieji duomenys, savo ruožtu, vis labiau priklauso nuo „mažųjų“ duomenų, iš kurių semiamasi prasmingų įžvalgų.

Bibliotekos nėra atsakingos už didelių duomenų rinkinių saugojimą. <...> Bibliotekoms turėtų labiau rūpėti tai, ar didieji duomenys yra „lengvai randami“ ir „prieinami“ bei atviri analizei, o ne jų saugojimas ar apdorojimas.

Su bibliotekomis yra susiję šie duomenų tipai <...>:

- 1) Archetipiniai didieji duomenys. [Šiuos duomenis] kuria moksliniai projektai, pramonė ir valdžios įstaigos. Jie yra gausūs, sparčiai apdorojami, patikimi ir vertingi, tačiau neįvairūs. Tokius duomenis paprastai kuria prietaisai ir sistemos automatiniais procesais metu. Bibliotekose esama tokių taip kuriamų duomenų:
 - duomenų pėdsakai, esantys šalutiniu prietaiso arba sistemos pagrindinės funkcijos produktu, pvz., medžiagos srautų savikontrolė;
 - lankytojų veiklos duomenys, pvz., skolinimosi, būdų mokėjimo duomenys <...>;
 - naudojimo būdų duomenys, pvz., įrašyti sėkmingos paieškos žurnaluose;
 - vartotojų elgsenos duomenys <...>.
- 2) Sutelktiniai didieji duomenys. [Tai] duomenys, kuriuos viešai ir savanoriška teikia piliečiai, pvz., socialiniuose tinkluose, rekomenduojamojo pobūdžio sistemose, komercinėse tinklo sistemose. Tokie duomenys yra gausūs ir įvairūs, nevienodai sparčiai apdorojami, nevienodos ir neilgalaiškės vertės. Bibliotekose tai yra <...> bibliotekos *Facebook*, *Twitter* arba kt. paskyrų komentarai.
- 3) Tiesioginio vaizdo stebėjimo didieji duomenys. [Šiuos duomenis] kuria <...> patalpų stebėjimo sistemos, aplinkosauginės sistemos, asmeniniai nešiojamieji prietaisai. Jie yra gausūs, sparčiai apdorojami, nevienodos vertės, nevienodai patikimi. Bibliotekose tokie duomenys yra apsaugos vaizdo kamerų medžiaga, turniketų judėjimo informacija, patalpų temperatūros matavimo duomenys.
- 4) „Ilgauodegiai“ (angl. *long-tail*) duomenys. Tai duomenys, kuriais mokslu ir mokslinių projektų, įskaitant individualių piliečių mokslinius projektus. Iš tikrųjų, tai yra didieji duomenys: kasmet paskelbiama 2.5 mln. mokslinių publikacijų <...> ir 600 tūkst. DOI. <...> [Šie duomenys] yra gausūs, lėtai apdorojami, neįvairūs ir <...> sunkiai nustatomos vertės.

Didieji duomenys svarbūs akademinėms bibliotekoms kaip mokslinių tyrimų procesų tarpininkams ir rėmėjoms. Be tarpininkavimo teikiant mokslo duomenų prieigą per leidėjus ir duomenų kaupėjus, bibliotekos gali taip pat teikti jose saugomų panašių duomenų šaltinių prieigą.

Prieš imdamiesi duomenų analizei skirtų programų, bibliotekų vadovai turi <...> atsižvelgti į keletą aspektų. <...>. Gebėjimus. Su duomenimis susijusių įgūdžių trūksta visuose bibliotekininkystės lygiuose – pradedant formaliuoju švietimu bibliotekininkystės mokslo įstaigose <...> ir baigiant bibliotekų vadovų veikla. Norint būti veiksmingu agentu duomenų valdomame pasaulyje, būtina žinoti, kur link juda bibliotekos su duomenimis susijusios paslaugos ir kaip būti duomenų analizės avangarde, gerai išmanyti duomenų analizę ir turėti jo įgūdžių. <...> Liza Federer pateikė duomenų analizės kompetencijų sąrašą: (1) suprasti „duomenų gyvavimo ciklą“; (2) mokėti parengti duomenų valdymo planą; (3) mokėti pasirinkti failų formas ir formatus; (4) mokėti sisteminti duomenis ir padaryti juos lengvai surandamus; (5) žinoti, kaip saugiai ilgą laiką saugoti duomenis; (6) propaguoti duomenų mainus skelbiant duomenų rinkinius ir suteikiant nuolatinį identifikatorių, pvz., DOI; (7) suvokti duomenis kaip mokslinės veiklos rezultata, kuris turėtų būti siejamas su pareigybės paaukštinimu.

[Du bibliotekoms svarbūs duomenų analizės kompetencijų tipai] Duomenų bibliotekininkas. Tai į duomenų saugojimą orientuotas vaidmuo, susijęs su duomenų politika, duomenų valdymo planavimu, duomenų raštingumui skirtų mokymo programų rengimu ir viešinimu <...>. Duomenų technologas. Tai yra vaidmuo, susijęs su techniniais veiklos procesais. Duomenų technologas papildo duomenų bibliotekininko žinias, susijusias su kompiuterijos įgūdžiais ir kt. <...>

Skiriama keletas archetipinių struktūrų, kurios naudojamos duomenų paslaugas teikiančiose akademinėse ir mokslinių tyrimų institucijose: (1) savarankiškai veikiantis bibliotekininkas; (2) speciali darbo grupė; (3) daigafunkcinė grupė; (4) specializuota grupė.

[Bibliotekos galėtų] padėti mokslininkams ir besimokantiesiems <...>:

- rengdamos duomenų valdymo planus ir teikdamos duomenų valdymo planų prieigą;
- padėdamos rasti duomenis (ieškant jų duomenų saugyklose);
- analizuodamos ir vizualizuodamos duomenis;

- sistemindamos duomenis; teikdamos patarimus, susijusius su pavadinimų failams suteikimu bei metaduomenimis;
- saugodamos duomenis; nukreipdamos mokslininkus į įvairias esamas duomenų saugyklas;
- dalindamosi duomenimis, teikdamos patarimus mokslininkams apie duomenų mainų platformas);
- cituodamos duomenis.

Bibliotekų perkamos arba prenumeruojamos duomenų bazės <...> galėtų būti interpretuojamos kaip didžiųjų duomenų šaltiniai, kurie galėtų būti analizuojami, vizualizuojami išgaunant paslėptąją jų vertę vieneto lygmeniu.

<...> bibliotekos veikia kaip tarpininkės, siejančios leidėjus, kaupėjus ir mokslininkus, kuriems jos teikia informacines paslaugas. Jos yra gerai pasirengusios ne tik teikti mokslo informacijos šaltinių prieigą, bet ir mokslinių duomenų prieigą sudarydamos papildomas prenumeravimo sutartis arba teisiniu lygmeniu agituodamos už autorių teisių taikymo išimtis.

Didieji duomenys taip pat kelia su metaduomenimis susijusių sunkumų. Kokybiški metaduomenys yra nepaprastai svarbūs; jie padeda suprasti duomenų turinį, jų kilmę ir kontekstą. <...> jie padeda tyrėjams „derinti“ įvairius skirtingus šaltinius. Toks poreikis darosi vis aktualesnis didžiųjų duomenų pasaulyje augant galimai naudingų duomenų kiekiui. Bibliotekos galėtų prisidėti, tam pritaikydamos savo su metaduomenimis susijusias paslaugas.

4-osios industrinės revoliucijos laikais duomenys yra vienas iš ekonomikos variklių, naujoji valiuta. Tai vienas iš esminių mašininio mokymosi ir dirbtinio intelekto kertinių akmenų. Tai teikia bibliotekoms galimybę kuruoti kokybiškus duomenis, lygiai kaip jos teikia kokybiškos informacijos prieigą.