

# Duomenų tyryba

# Turinys

1. Testas prieš pradedant nagrinėti temą
2. Duomenys, informacija, žinios
  1. Testas „Duomenys, informacija, žinios“
3. Duomenų tyryba
  1. Privalumai ir trūkumai
  2. Sąryšiai su kitomis sritimis
  3. Duomenų tyrybos taikymai (diskusija)
4. Duomenų tipai
  1. Įvairių tipų duomenų pavyzdžiai
  2. Duomenų šaltiniai
  3. Veiksmai su duomenimis
  4. Testas „Struktūrizuotų ir nestruktūrizuotų duomenų privalumai ir trūkumai“.

# Testas prieš pradedant skyrių (atlikti Moodle)

*1. Kokį vaidmenį vaidina žodis „tyryba“ duomenų tyryboje?*

- a) Tai reiškia, kad analizuojant duomenis bus taikomi moksliniai metodai.
- b) Tik specialistai, turintys atitinkamą išsilavinimą, tai supranta.
- c) Toks žodžių junginys gerai skamba.

*1. Teiginy „Mokytis duomenų tyrybos yra naudinga tik programuotojams“ yra:*

- a) Teisingas
- b) Klaidingas.

*1. Ką turėtumėte atlikti, norėdami pademonstruoti, ar krepšinio komandos žaidėjai yra yra aukštesni, negu vidutinio ūgio žmonės?*

- a) Surinkti duomenis.
- b) Turėti žinių apie statistiką ir tikimybes.
- c) Teisingi a) ir b) variantai.

# Duomenys, informacija, žinios (1)

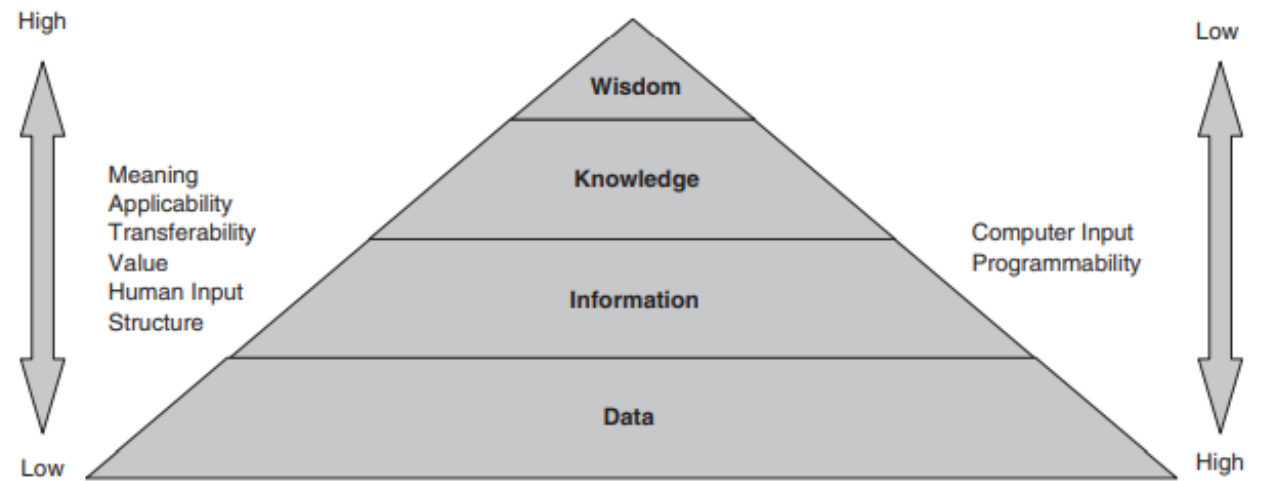
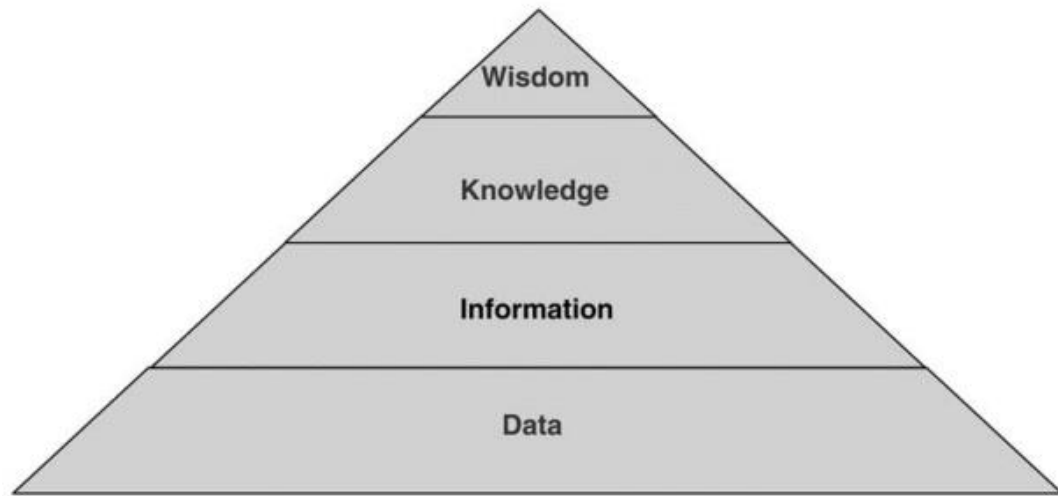
- **Duomenys** yra žaliava, iš kurios galima gauti informacijos ir žinių. Įsivaizduokite duomenis kaip vietas, vaizdus, aprašymus, atsiliepimus ir kainas, kurie sudaro informacijos, galinčios padėti planuoti, pavyzdžiui, atostogas, pagrindą.
- Duomenys tampa **informacija**, kai jiems suteikiamas kontekstas. Remiantis ankstesniu pavyzdžiu, vietos, vaizdai, aprašymai ir kainos – visa tai gali padėti pateikti informaciją, susijusią su turistų lankomomis vietomis. Duomenų rinkimas ir pateikimas padeda formuoti informaciją.
- **Žinios** yra tai, kas gaunama iš informacijos ir pritaikoma jūsų poreikiams. Žinių kaupimas – tai procesas, kurio metu informacija paverčiama pasirinkimais. Remdamiesi ankstesniu pavyzdžiu, žinodami, kad niekas iš jūsų šeimos narių nemėgsta pramogų parkų, galėsite nuspręsti, kurių vietų atostogų metu vengti ir kurios labiau tiktų jūsų šeimai.

# Duomenys, informacija, žinios (2)



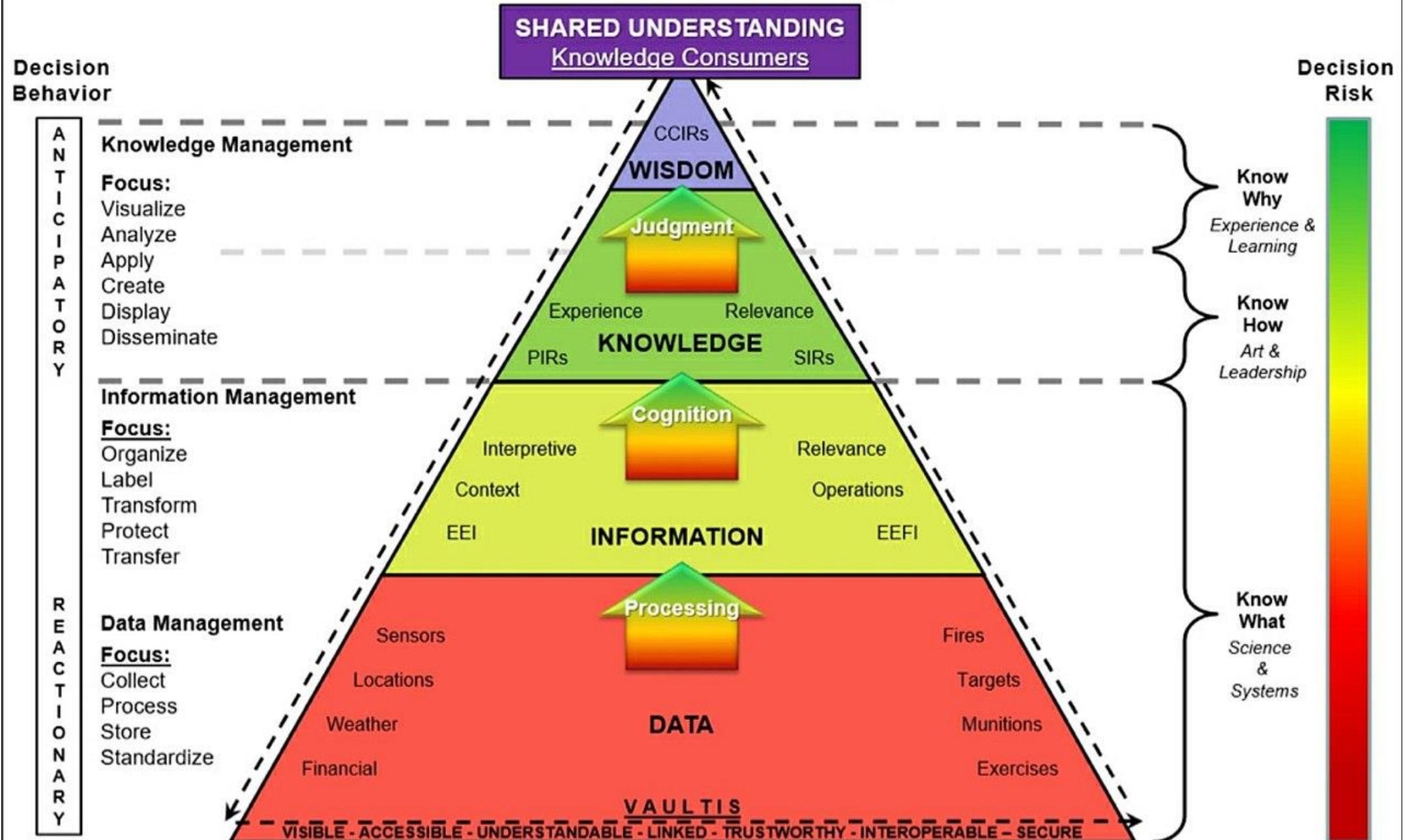
<https://internetofwater.org/valuing-data/what-are-data-information-and-knowledge/>

# DIKW piramidė



Šaltinis: Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163-180.

# The KM Cognitive Pyramid



# **Testas „Duomenys, informacija, žinios“**



# Duomenų tyryba

(šaltinis Kurasova, O. (2014). Programinės sistemos duomenų tyrybos mokymui. *Lietuvos matematikos rinkinys. Ser. B*, 55, 60-65)

- **Duomenų tyryba** – tai procesas, kurio metu, naudojant įvairius duomenų analizės įrankius, bandoma nustatyti ir atrasti „užslėptas“ duomenų struktūras ir ryšius.
- **Duomenų tyrybos tikslas** – iš įprastai didelių duomenų aibių išgauti svarbią informaciją ir atmesti nereikšmingą. Duomenų tyrybos metodai padeda „nepasiklysti“ duomenų ir informacijos gausoje.
- **Duomenų tyryba** atliekama žinių radimo duomenų aibėse procese (angl. knowledge discovery in databases), kurio metu didelių apimčių duomenų aibėse ieškoma naujos informacijos, padėsiančios įgyti žinias apie analizuojamus duomenis ir priimti tinkamiausius sprendimus.

# Duomenų tyrybos privalumai

- Duomenų tyrybos metodai leidžia organizacijoms gauti žiniomis pagrįstus duomenis.
- Duomenų tyryba leidžia organizacijoms atlikti pelningus veiklos ir gamybos pakeitimus.
- Palyginti su kitomis statistinių duomenų taikymo sritimis, duomenų tyryba yra ekonomiškai efektyvi.
- Duomenų tyryba padeda organizacijos sprendimų priėmimo procese.
- Ji palengvina automatinį paslėptų dėsningumų atradimą, taip pat tendencijų ir elgsenos prognozavimą.
- Ją galima įdiegti tiek naujoje sistemoje, tiek esamose platformose.
- Tai greitas procesas, kuris naujiems naudotojams leidžia per trumpą laiką išanalizuoti didžiulius duomenų kiekius.

# Duomenų tyrybos trūkumai

- Yra tikimybė, kad organizacijos gali parduoti naudingus klientų duomenis kitoms organizacijoms už pinigus. Kaip teigiama pranešime, "American Express" pardavė savo klientų kredito kortelių pirkinius kitoms organizacijoms.
- Daugeliu duomenų tyrybos analitinės programinės įrangos yra sudėtinga naudotis, o darbui su ja reikia išankstinio mokymo.
- Skirtingos duomenų tyrybos priemonės veikia skirtingai dėl jų konstrukcijoje naudojamų skirtingų algoritmų. Todėl tinkamų duomenų tyrybos priemonių pasirinkimas yra labai sudėtinga užduotis.
- Duomenų tyrybos metodai nėra tikslūs, todėl tam tikromis sąlygomis gali sukelti sunkių padarinių.

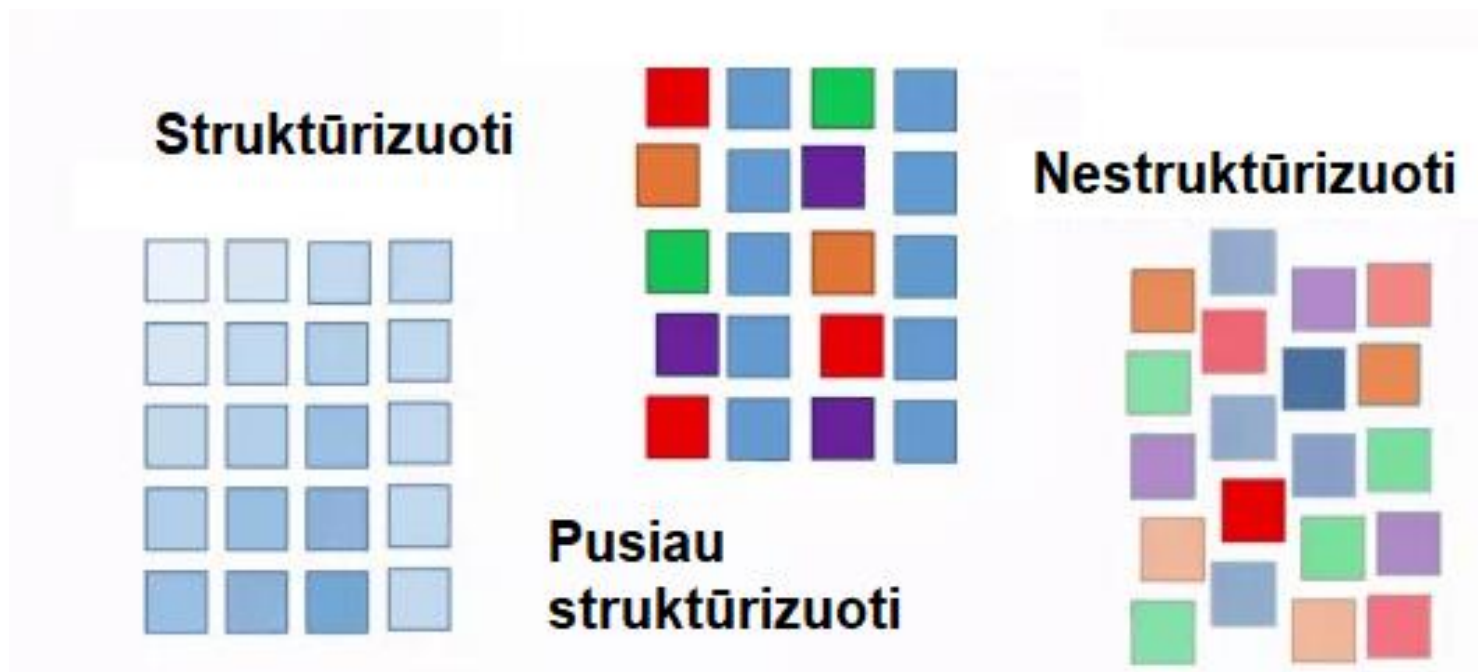
# Duomenų tyrybos sąryšiai su kitais mokslais (sritimis)

- Duomenų bazės – duomenų saugojimas.
- Didieji duomenys (Big Data) – specialūs metodai ir įrankiai dideliems duomenų kiekiams apdoroti.
- Mašininis mokymasis – modelių kūrimas iš duomenų.
- Dirbtinis intelektas – nestructūrizuotų duomenų pavertimas struktūrizuotomis įžvalgomis.
- Vizualizavimas – geresniam supratimui ir įžvalgų kūrimui.

# Duomenų tyrybos taikymai: diskutuojame, siūlome variantus

- Sveikatos apsauga
- Žemėtvarka
- Energetika
- Regitra
- Švietime
- Bankų ir draudimo kompanijų
- Telekomunikacijos
- Rinkų tyrimai

# Duomenų tipai



Šaltinis: <https://www.astera.com/type/blog/structured-semi-structured-and-unstructured-data/>

# Įvairių tipų duomenų pavyzdžiai

Struktūrizuoti	Pusiaus struktūrizuoti	Nestruktūrizuoti
Žmonių sąrašas su jų telefono numeriais.	Vikipedijos puslapiai su nuorodomis	Encyclopedia Britannica tekstas
Pastarųjų 20 metų visų pastato patalpų temperatūra kiekvieną minutę.	Mokslinių straipsnių rinkinys JSON formatu su autoriais, publikacijos duomenimis ir santrauka	Failų bendrinimas su įmonių dokumentais
Visų į pastatą įeinančių žmonių amžiaus ir lyties duomenys	Interneto puslapiai	Neapdorotas vaizdo įrašas iš stebėjimo kameros

# Duomenų šaltiniai

## 1. Struktūrizuotų duomenų šaltiniai

- Daiktų internetas (IoT), įskaitant įvairių jutiklių, pavyzdžiui, temperatūros ar slėgio ir kt., duomenis, suteikia daug naudingų duomenų. Pavyzdžiui, jei namuose ar mokykloje įrengti daiktų interneto jutikliai, galime automatiškai valdyti šildymą ir apšvietimą, kad sumažintume išlaidas.
- Apklausos, kurias prašome naudotojų užpildyti po pirkimo arba apsilankius interneto svetainėje.
- Elgesio analizė gali, pavyzdžiui, padėti suprasti, kaip giliai naudotojas įeina į svetainę ir kokia yra tipinė svetainės palikimo priežastis.

## 1. Nestruktūrizuotų duomenų šaltiniai

- Tekstai gali būti turtingas įžvalgų šaltinis, pavyzdžiui, bendras nuotaikų įvertinimas arba raktinių žodžių ir semantinės reikšmės išskyrimas.
- Vaizdai arba vaizdo įrašai. Stebėjimo kameros vaizdo įrašas gali būti naudojamas siekiant įvertinti eismo intensyvumą kelyje ir informuoti žmones apie galimas spūstis.
- Interneto serverio žurnalai gali būti naudojami siekiant suprasti, kurie mūsų svetainės puslapiai lankomi dažniausiai ir kiek laiko.

## 1. Pusiau struktūrizuotų duomenų šaltiniai

- Socialinių tinklų grafikai gali būti puikus duomenų apie naudotojų asmenybes ir galimą informacijos sklaidos veiksmingumą šaltinis.
- Kai turime krūvą nuotraukų iš vakarėlio, galime pabandyti išgauti grupės dinamikos duomenis sudarydami žmonių, fotografuojančių vienas kitą, grafiką.



# Veiksmai su duomenimis

- Rinkimas
- Saugojimas
- Apdorojimas
- Vizualizavimas
- Prognostinio modelio kūrimas

**Testas „Struktūrizuotų ir nestructūrizuotų  
duomenų privalumai ir trūkumai“**