

Tutkimustiedonhallinnan peruskurssi

Hannu Toivonen, Marko Salmenkivi, Inkeri Verkamo

`hannu.toivonen, marko.salmenkivi, inkeri.verkamo@cs.helsinki.fi`

Helsingin yliopisto – Tietojenkäsittelytieteen laitos

Kurssin esittely



Perustiedot

- 581264-6 Tutkimustiedonhallinnan peruskurssi (3 ov)
- tietojenkäsittelytieteen cum lauden ja laudaturin valinnainen kurssi
- soveltuu myös muiden aineiden pääaineopiskelijoille, jotka joutuvat tekemisiin tutkimustietoaineistojen kanssa
- edellyttää tietojenkäsittelytieteen approbaturin tasoisia perustietoja ja -taitoja ohjelmoinnissa sekä tiedonhallinnassa

Mitä kurssilla käsitellään

- tutkimustiedonhallinnan erityistarpeet
- datan visualisointi
- laskennallisia data-analyysimenetelmiä
- satunnaislukujen ja otosten generointi

Suhde muuhun opetukseen

- taustaa: johdatus sovellussuunnitteluun, tietokantojen perusteet, ohjelmoinnin perusteet
- sivuaa: tilastotiede, todennäköisyyslaskenta
- jatkoa:
 - tiedonhallinta, tietokannan mallinnus, tietokantarakenteet ja algoritmit, tietovarastot, paikkatieto
 - tiedon louhinnan peruskurssi, tietämyksen muodostaminen

Kurssin rakenne ja aikataulu

- Ks. [kurssin verkkosivut](#)
- pakollinen harjoitustyö: ks. [ohjeet verkossa](#)

Kurssin suoritus

- luennot
- laskuharjoitukset
- pakollinen harjoitustyö (8-20 p)
- tentti (15-40 p)
- yhteensä vähintään 30/60 p
- Huom:
 - harjoitustyön tekeminen on mahdollista vain kurssin yhteydessä
 - suoritus pelkällä loppukokeella ei ole mahdollista
 - tentin voi tarvittaessa uusia, jos harjoitustyö on ok

Kurssimateriaali

- **Z. Michalewicz (toim.): Statistical and Scientific Databases; Ellis Horwood Ltd, 1991, luvut 2, 4**
- **W. Cohen: Empirical Methods for Artificial Intelligence, luvut 2-5**
- **lisäksi erillisiä artikkeleita (ilm. myöhemmin)**
- **luentokalvot (eivät yksin kata kurssin sisältöä)**
- **materiaalin saatavuus**
 - kirjat kirjastossa, kurssilla käsiteltävät luvut kurssikansiossa
 - muu materiaali verkossa ja kansiossa
 - <http://www.cs.helsinki.fi/hannu.toivonen/teaching/tutihaK03/>

Tutkimustiedonhallinnan erityispiirteet



Erityispiirteet: osan sisältö

- Mistä tällä kurssilla on kysymys?
- Missä suhteissa tutkimustieto on erilaista kuin muu tieto?
- Millaisia erityisvaatimuksia tutkimustiedolla voi olla tiedon hallinnalle?
- Lähde: Z. Michalewicz (toim.): Statistical and Scientific Databases; luku 2

Mistä tutkimustieto on peräisin?

- tieteellinen tutkimus
 - havaintoja jostakin ilmiöstä
 - kokeellinen tutkimus
 - simulointikokeet
- toiminnan sivutuotteena syntyneet tietokokoelmat
 - tieto kerätty alun perin muuhun tarkoitukseen
 - kerättävän tiedon rakenne riippuu yleensä vahvasti alkuperäisestä käyttötarkoituksesta
- tiedon tulva: analysointi on pullonkaula
- “scientific and statistical databases”

Esimerkkejä sovellusalueista

- ympäristötutkimus
- meteorologia, seismologia
- avaruustutkimus
- arkeologia
- geenitutkimus, lääketiede
- kemia, fysiikka
- sosiologia, kielitiede
- ...

Vrt. perinteinen tiedonhallinta

- tutkimustiedot ja niiden tarpeet vaihtelevat paljon
- ⇒ vaihtelevia ja monipuolisia tarpeita
- 1. tietomalli
 - 2. operaatiovalikoima
 - 3. rinnakkaisuus
 - 4. tiedon eheys
 - 5. toipuminen
 - 6. fyysinen rakenne

1. Tietomalli

- yleisimmin käytetty tietomalli (ER):
 - alkiot esitetään tauluissa
 - taulujen alkioita liitetään toisiinsa vastinsarakkeiden perusteella
 - alkiot ovat joukkoja: ei järjestystä

-
- tutkimustiedon edellyttämä tietomalli:
 - merkittävä osa tutkimustiedosta on järjestettyä dataa:
 - aikasarjat
 - koesarjat
 - yksittäisten alkioiden rakenne voi olla hyvin monimutkainen:
 - koetuloksiin voi liittyä tutkittavan osan rakennekaavio
 - alkioiden väliset yhteydet

2. Operaatiovalikoima

- perinteisen tietokannan operaatiot ovat SQL-kyselyitä:
 - hae kaikki alkiot, jotka täyttävät annetun ehdon, ja tee niille haluttu operaatio

- tutkimustiedon vaatimat operaatiot:
 - sovellusaluekohtaisia operaatioita
 - operaatiot voivat sisältää monimutkaista laskentaa
 - useat operaatiot käsittelevät kaikkia tietueita
 - tyypillisiä käyttötapoja ovat otokset ja niihin kohdistuvat operaatiot
 - käsiteltävän datan valinta myös esim. visualisoinnista

3. Rinnakkaisuus

- perinteisissä sovelluksissa: rinnakkaisuus = monta samanaikaista käyttäjää jotka voivat häiritä toisiaan erityisesti tietoa päivittäessään
- samanaikaisuus tutkimustiedon yhteydessä:
 - ei yleensä paljon samanaikaisia käyttäjiä (ja he käyttävät useimmiten eri osia tietokokoelmasta)
 - tutkimustietoa ei usein päivitetä lainkaan (tai vain poikkeustilanteissa)
 - ⇒ tarvitaanko rinnakkaisuutta?
 - tietokokoelma voi olla hyvin laaja ja / tai hajautettu:
 - ⇒ edellyttääkö tehokas käyttö rinnakkaisuutta?

4. Tiedon eheys

- tietokannan sisältämän tiedon eheyttä voidaan tarkistaa esim. tietotyypin perusteella:
 - arvot hyväksytyllä välillä
 - esitysmuoto

- tutkimustiedon eheys:

- tarvitaan usein monimutkaisiakin eheystarkistuksia
- tiedon puhdistamisen yhteydessä poimitaan eroon epätavalliset arvot (outliers):
 - mahdollisesti virheelliset tiedot
 - puuttuvat tiedot
 - mielenkiintoiset havainnot?
- täydentäminen ja korjaaminen voi olla hankalaa tai mahdotonta (historia?)
- poikkeavien tietojen käsittelytapa on päätettävä tapauskohtaisesti

5. Toipuminen

- perinteisiin tietokantoihin liittyvä varmistus- ja toipumismenettely:
 - vikatilanteessa palautetaan tietokanta viimeisen onnistuneen tapahtuman jälkeiseen tilanteeseen
- tutkimustiedon edellyttämä toipuminen:
 - tyypillisesti pitkiä tapahtumia: toipuminen ei saa merkitä sitä, että paljon tehtyä työtä hylätään

6. Fyysinen rakenne

- tietokannan tyypillinen fyysinen talletus:
 - tavoitteena tehokkaat kyselyt
 - taulut talletetaan riveittäin tiedostoiksi
 - lisäksi muodostetaan hakemistoja, joiden avulla halutut kyselyt voidaan toteuttaa tehokkaammin
- tutkimustiedolle sopiva fyysinen rakenne:
 - usein alun perin sellaisenaan tiedostoissa
 - onko paras talletusrakenne:
 - tiedosto?
 - puu tai jokin muu tietorakenne?
 - tietokanta?

-
- (tutkimustiedolle sopiva fyysinen rakenne):
 - tiedon rakenne:
 - mikä on sopivin järjestys?
 - tietojen ryhmittely = paikallisuus: mitä alkioita (tai alkioiden osia) käytetään usein yhdessä?
 - esim. spatiaalinen data
 - valitun ratkaisun tehokkuus?
 - datan määrä?
 - datan kertymisnopeus?
 - harvat taulut: tilavaatimus?
 - tietojen käsittelyjärjestys: aikavaatimus?

Mistä tutkimustieto koostuu

- mikrodata eli raakadata:
 - havainnot
 - simuloinnin tulokset
- makrodata eli analysoitu data = käsittelyn tuloksena syntynyt data
 - periaatteessa konstruoitavissa uudelleen, mutta voi olla työlästä
 - käsittely voi kohdistua raakadataan tai jo analysoituun dataan: esim. yhteenvedot
- metadata

Tutkimustiedon ominaisuuksia

Raakadata:

- havaintoja tai simulointituloksia
- havaintojen säännöllisyys:
 - havaintoja tehdään usein jonkin (aikaan tai paikkaan liittyvän) säännön mukaan
 - koordinaattien talletus?
- havaintojen tiheys:
 - hylätyt arvot (mittaustarkkuus)
 - havaintotiedon tiivistys

-
- ajan vaikutus koordinaatteihin:
 - havaintopisteiden koordinaatit voivat muuttua ajan funktiona

Parametridata:

- raakadataa kuvaaviin parametreihin liittyy usein luokitteluhierarkioita
 - samat tai eri operaatiot hierarkian eri tasoilla
 - yhteenvedot hierarkian eri tasoilla

Tutkimustiedon rakenne

- moniulotteisuus monessakin mielessä:
 - kymmeniä tai satoja attribuutteja
 - eri tietojoukoissa voidaan käyttää samoja attribuuttien nimiä (samoista tai samantapaisista käsitteistä)
 - attribuuttien merkitykset (luokittelukriteerit) voivat muuttua ajan kuluessa
 - datajoukosta voi olla useita eri-ikäisiä versioita yhtä aikaa käytössä

-
- rakenteellinen monimutkaisuus:
 - alkioiden välillä on monenlaisia (ja monimutkaisia) riippuvuuksia
 - alkio voi sisältää rakenteeltaan ja käsittelytavaltaan erilaisia osia
 - luokitteluhierarkiat
 - erityiset tietotyypit kuten
 - aikasarjat
 - paikkadata
 - graafinen tieto, kuvat

Erityispiirteet: yhteenveto

- tarpeet poikkeavat yleisistä kaupallis-hallinnollisista tietokannoista, mm.
 - järjestetty data
 - sovelluskohtaiset operaatiot
 - puhdistaminen ja jalostaminen
- mikrodata vs. makrodata
- moniulotteisuus

Metadata



Metadata: osan sisältö

- Mitä metadata on?
- Mihin sitä tarvitaan?
- Mitä vaikutusta metadataalla on tutkimustiedonhallinnan ratkaisuihin?
- Lähde: Z. Michalewicz (toim.): Statistical and Scientific Databases; luku 4
- Esimerkkisovellus: Pubudu Wariyapola et al.:
Ontology and Metadata Creation for the Poseidon Distributed Advances in Digital Libraries, 1999, 180–189.

Mitä metadatalalla tarkoitetaan?

- Metadata on tietoa, jota tarvitaan raakadatan ja analysoidun datan
 - hakemiseen, käsittelemiseen, jäsentämiseen, tulkitsemiseen
- Metadatan keruu ja talletus on välttämätöntä
 - muuten varsinainen data ei ole käyttökelpoista!
- tavoittena varsinaisen datan
 - ymmärrettävyys, käytettävyys, ylläpidettävyys
 - muillekin ja myöhemminkin
 - + tutkimuksen toistettavuus

Tutkimuksen metadataa

- raakadatan tuottaminen ja kerääminen
 - tutkimus- tai koejärjestelyt
 - datan keruun tai mittausprosessin kuvaus
 - kuka, missä, milloin?
- datan jalostus ja analyysi
 - suoritettut muunnokset, suodatukset, korjaukset
 - muunnosten yms. perustelut
 - tehdyt analyysit, niiden parametrit
 - sovelletut menetelmät ja työkalut

-
- tutkimustulosten käyttöohjeet
 - opastusta tiedon ymmärtämiseen, lisätiedon hankkimiseen
 - tulkinassa käytetyt mallit ja otokset
 - viittauksia ulkoisiin artikkeleihin, taksonomioihin, standardeihin
 - sovellusaluekohtainen osaaminen
 - ei välttämättä käytettävissä koneellisesti
 - äärimmillään koko tutkimushistoria rekonstruoitavissa

Metadataa on myös...

- tiedon fyysinen rakenne
 - tiedostojen nimet ja formaatit, tietotyypit
- tiedon looginen rakenne
 - tietoalkioiden nimet ja attribuutit
 - mahdolliset operaatiot

Metadatan jaottelu I

- 1. *ohjaava* (control) metadata
 - käsittelijöinä tietokannan ja järjestelmän operaatiot
 - määrämuotoinen esitystapa
- 2. *opastava* (guide) metadata
 - käsittelijöinä ihmiset
 - esitys luonnollisella kielellä
- ohjaavan metadatan käyttö suositeltavaa
 - automatisointi, kustannusten ja resurssien säästö
- opastava metadata välttämätöntä
 - datan ymmärtäminen
 - rajapinta tiedon tuottajan ja käyttäjän välillä

Metadatan jaottelu II

- 1. Käyttäjän näkökulma
 - etsintä; selailu
 - sekä raakadata että käsitelty data
- 2. Tietohakemiston näkökulma
 - laadun varmistus
 - tiedon eheys
 - sekä raakadatan että käsitellyn datan laatu
 - käsittelyoperaatiot
- 3. Ympäristön näkökulma
 - sovellusten välisen tiedonsiirron rajapinnat
 - laaja (ja laajeneva) valikoima

Metadatatietokanta

- rakenteeltaan ja käsittelytavaltaan erilaisia tietotyyppejä, kuten
 - kaavioita
 - tekstidokumentteja
 - ohjelmia
 - sääntöjä
 - taulukoita
 - viittauksia muihin lähteisiin
 - omat talletustavat ja operaatiot (työvälineet)

-
- voidaan rakentaa kuten tietokanta
 - parhaimmillaan käsitellään samassa ympäristössä kuin varsinaista dataa
 - tavoitteena on automatisoida metadatan käsittelyä
 - mahdollisia työkaluja myös XML, RDF, ontologiat
 - tietojen vaihdettavuus, yhteiset käsitteistöt

Tutkimusmetadatan piirteitä

- metadata on hyvin monimuotoista
- metadata muuttuu usein
 - kerätään uudentyyppistä dataa
 - käsitellään kerättyä dataa uusilla tavoilla
- monimuotoisuus rakenteessa ja käytössä
 - ⇒ ei keskitettyä auktoriteettia
 - ⇒ ei kaikenkattavia standardeja
 - ⇒ vain rajallinen automatisointi
- datan ja metadatan tuottavat tutkijat
 - tutkijan pitäisi päästä keskittymään sisältöön, ei rakenteeseen

Metadata: yhteenveto

- metadata kuvaa dataa
 - metadata on välttämätöntä datan käsittelemiseksi ja ymmärtämiseksi
 - tutkimustiedon tuottaminen ja jalostaminen: paljon vaihtoehtoja, täsmällisyys tärkeää
- ⇒ tarkka ja monipuolinen metadata
- miten raakadata on tuotettu ja kerätty?
 - miten dataa on käsitelty ja analysoitu?
 - kytkentä tunnettuihin taksonomioihin ja standardeihin

Tiedon mallitus ja talletus



Tiedon mallitus: osan sisältö

- Miten tietoa jäsennetään talletusta varten?
- Millaiset esitystavat ja talletusrakenteet sopivat erityisesti tutkimustiedolle?
- Lähde (esim.): Harri Laine: Johdatus sovellussuunnitteluun ja Tietokantojen perusteet (TKTL:n opetusmonisteita)

Tiedosto

- kaiken tiedon talletuksen perusta
 - monimutkaisemmat talletusrakenteet toteutetaan (yhden tai useiden) tiedostojen avulla
- hierarkkinen muistijärjestelmä
 - keskusmuisti: X ns
 - levy: X ms
 - cd-pakka: X s
- käsittelyä varten tieto tuodaan muistiin
- tehokkuuden kannalta on edullista
 - tuoda tietoa kerralla paljon
 - käsitellä muistissa oleva osa kerralla

Järjestämätön peräkkäisrakenne

- tietueet peräkkäin lisäysjärjestyksessä
- lisäys aina tiedoston loppuun
- ei hukkatilaa
- hyvä ratkaisu jos tarvitaan
 - suurta talletusnopeutta
 - staattista talletusrakennetta (= ei poistoja)
 - tehokasta tilankäyttöä
 - peräkkäiskäsittelyä aina samassa järjestyksessä (tai järjestyksellä ei väliä)
- tyypillinen talletusrakenne raakatiedon talletuksessa

Hakupuut

- tietueet (tai avaimet) on talletettu puun lehtiin
- sisäsolmun lapset osittavat sen tietueavaruuden
- puun pitäminen tasapainossa takaa tehokkuuden
 - joka lehteen yhtä pitkä matka ($\log_m(n)$)
- lukuisia muunnelmia: mm. B-, R-, LSD-, TV-puut
- erityisiä puurakenteita paikkatiedolle
 - tehokas haku 2–3-ulotteisessa datassa
 - talletuksen lokaalisuus \Rightarrow tyypillisten operaatioiden tehokkuus

Tietokanta

- suuren tietomäärän pitkäaikainen säilytys
- rakenteinen tieto
 - tietoalkioiden väliset suhteet
 - tiedon looginen rakenne
- tietoon kohdistuvat kyselyt ja operaatiot
- tiedon samanaikainen käyttö
- hallintaan liittyviä tehtäviä on siirretty käyttäjältä tietokannan hallintajärjestelmälle

Tiedostot vs. tietokanta

- säilytys: OK? (varmuuskopiointi)
- rakenteisuus, eheys: käyttäjä hallitsee ja ylläpitää
- kyselyt ja operaatiot: käyttäjä ohjelmoi ja toteuttaa
- samanaikainen käyttö: käyttäjät huolehtivat itse

Tiedonhallinnan peruskäsitteitä

- eritasoiset kaaviot (schema) joilla tietokantaa kuvataan
 - käyttäjän näkymä tietokantaan
 - tietokannan looginen sisältö
 - tietokannan talletusrakenteet
- käyttäjän operaatioita varten tarvitaan
 - määrittelykieli
 - kyselykieli
 - SQL sisältää molemmat komponentit

■ työnjako

- tietokannan käyttö
- tietokannan ylläpito
- tietokannan toteutus

■ lähestymistapoja mallittamiseen

- relaatiomallit (relaatiot, attribuutit, relaatiokaaviot)
- oliiomallit (oliot, luokat, tyypit)

Tietokantajärjestelmän osat

- käyttäjä
 - tietokantaan kohdistuvat kyselyt (kyselykieli ja sovellusohjelmat)
 - tietokannan sisällön muutokset
 - tietokantakaavion muunnokset
- tietokannan hallintajärjestelmä
 - kyselyiden käsittely
 - tapahtumien käsittely
 - talletusrakenteen käsittely
- tietojen talletus
 - data + metadata

Tietokannan suunnittelun vaiheet

- 1. laaditaan malli käyttäjän näkökulmasta
 - mitä tietokannassa pitäisi olla
 - miten osat suhtautuvat toisiinsa
 - miten tietokantaa on tarkoitus käyttää
 - käyttäjän käsitteet

-
- 2. kirjoitetaan malli tietokantajärjestelmän ymmärtämällä kielellä
 - relaatorakenne, oliorakenne
 - 3. muokataan rakenne tehokkaaksi (normalisointi)
 - toistot pois, turhat rajoitukset pois
 - hakemistot

Tietomallin laatiminen

- reaalimaailman asioita, niiden ominaisuuksia ja suhteita
 - kuvaavat nimet
 - havainnollinen esitystapa: kaaviot, teksti
- malli koostuu olioista, joihin voi liittyä
 - attribuutteja
 - relaatioita
 - operaatioita (menetelmiä)
- ominaisuuksiltaan samanlaiset oliot kuuluvat samaan luokkaan

-
- terminologia vaihtelee riippuen käytettävästä esitystavasta
 - ER-malli: yksilöt, ominaisuudet, ...
 - oliomalli: luokat, oliot, menetelmät, ...

Attribuutti (ominaisuus)

- olion ominaisuus, jolla on jokin arvo
- olioon sisältyvä tieto
- attribuuttiin liittyy tyyppi (arvojoukko), esim.
 - kokonaisluku
 - merkkijono
 - (punainen, keltainen, vihreä)
- attribuutti voi olla myös rakenteinen, esim.
 - osoite: katuosoite, postinumero
- attribuutilla voi olla myös puuttuva arvo

Relaatio

- ominaisuus joka liittyy olion toiseen olioon
- olioiden välinen suhde
- relaatio voi liittää toisiinsa samaan tai eri luokkiin kuuluvia olioita
- relaatio voi olla
 - yksi yhteen
 - yksi moneen
 - moni moneen

Operaatio (menetelmä)

- toimenpide joka oliolle voidaan tehdä
- tapa jolla oliota voidaan käsitellä
- operaatiot liittyvät kiinteästi tiedon käyttötarkoitukseen
- operaatioiden kytkeminen olioihin on erityisesti oliosuunnitteluun liittyvä piirre
 - olio ja sen operaatiot muodostavat kokonaisuuden
 - olio "näky" muille olioille vain operaatioidensa välityksellä

Luokkahierarkia — periytyminen

- oliolla voi olla samat ominaisuudet kuin jonkin luokan olioilla, mutta lisäksi joukko erityisominaisuuksia ⇒ aliluokka
- aliluokka perii yliluokkansa ominaisuudet
 - attribuutit
 - relaatiot
 - operaatiot
- luokat muodostavat hierarkian
- moniperintä: luokalla on monta yliluokkaa ja se perii kaikkien niiden ominaisuudet

Avain

- attribuutti tai joukko attribuutteja
 - avaimen oltava yksikäsitteinen, eli kahden eri olion avaimet eivät voi olla samat
- ⇒ avain yksilöi olion
- usein jokin attribuutti on erityisesti valittu avaimeksi
 - esim. henkilötunnus
- avaineheys: avaimella ei voi olla puuttuvaa arvoa
- oliomallissa jokaisella oliolla on yksikäsitteinen identiteetti: oliotunniste (OID)
- oliomallissa kahden eri olion kaikki attribuutit voivat olla samat

Suunnitteluperiaatteita

- mallin pitää vastata todellisuutta
 - mallin pitää olla (mahdollisimman) yksinkertainen
- ⇒ sovellus- ja tapauskohtainen kompromissi
- minimaalisuus
 - samoja asioita ei pidä toistaa
 - luokkahierarkian hyödyntäminen
 - rajoitukset
 - avaineheys, viite-eheys
 - attribuutin arvojoukkoa koskevat rajoitukset

Relaatiomalli

- relaatiotietokantojen perustana oleva yksinkertaiseen malli
- kaikki tiedot esitetään taulukkoina (relaatioina)
- attribuutit = relaatiotaulun sarakkeet
- relaatiokaavio = taulun otsikkorivi, esim.
 - mittaus(järvi, laji, lkm)
- relaatiokaaviot muodostavat yhdessä relaatiotietokantakaavion
- taulun rivi = monikko (tietue): sisältää kutakin attribuuttia vastaavan komponentin, esim.
 - (Tsuolbmajavri, Heterodissocladius maeari, 43)

-
- kaavion attribuutit esitetään aina samassa järjestyksessä
 - attribuutit ovat yksinkertaisia eivätkä koosteisia
 - relaation monikot muodostavat joukon
 - kaksi eri monikkoa eivät voi olla identtisiä
 - monikoiden käsittely perustuu niiden attribuuttien arvoihin, ei niiden "sijaintiin taulussa"
 - relaatiokaavio määrittelee rakenteen, joka on (yleensä melko) staattinen
 - relaatiotaulun sisältö (= relaation ilmentymä) voi muuttua
 - monikoiden lisäykset, poistot, päivitykset

Tietomallista relaatiomalliin

- olioluokka \Rightarrow relaatiotaulu
- yksinkertainen attribuutti \Rightarrow relaatiotaulun attribuutti
- rakenteinen attribuutti
 - esim. joukko relaatiotaulun attribuutteja
- olioiden välinen relaatio (yhteys)
 - attribuutti, joka sisältää viitattavan olion avaimen
 - hyvä ratkaisu, jos viitattavia olioita on vain yksi
 - relaatiotaulu, jonka attribuutteina ovat molempien olioiden avaimet
 - hyvä ratkaisu, jos relaatio on monesta moneen

Aliluokka relaatiomallissa

- kutakin aliluokkaa vastaa relaatiotaulu, joka sisältää kaikki attribuutit (sekä omat että perityt):
 - kaikki olion ominaisuudet löytyvät yhdestä taulusta
 - yhden attribuutin arvoja monessa taulussa

tai

- kunkin aliluokan relaatiotaulussa on vain ko. aliluokan omat ominaisuudet (ja avain)
 - olion ominaisuuksia monessa taulussa
 - samoja attribuutteja ei esitetä eri tauluissa

tai

- kaikki aliluokat esitetään yhtenä relaatiotauluna
 - komponentti saa puuttuvan arvon, jos attribuutti ei koske tätä monikkoa

Tiedon mallitus: yhteenveto

- talletusvaihtoehtoja: tiedostot, hakupuut ja muut erikoisrakenteet, tietokanta
- tietomallin käsitteitä:
 - attribuutti, relaatio, operaatio, luokkahierarkia
- tavoitteena yksikertainen mutta riittävän todenmukainen malli
- relaatiomalli, oliomalli