

Tietosisällön kuvaaminen

- n Toteutusvälineistä riippumaton tietosisällön kuvaus
- n Entity-Relationship malliperhe
 - n Lähtökohta: Chenin malli vuodelta 1976
 - n Useita muunnelmia, pieniä eroja peruskäsitteissä ja erityisesti esitystekniikassa
 - n Mallit keskittyvät tiedon rakenteeseen ja tietojen väliin riippuvuuksiin ja sääntöihin – tiedon käsittelyäkin on mallinnettu, mutta tähän malliin pohjautuvia tietokannanhallintajärjestelmiä ei juurikaan ole tarjolla
 - n tarkoitettu suunnittelun apuvälineeksi, ihmisille ei koneille

1

Tietosisällön kuvaaminen

- n ER-mallilla laadittu tietosisällön kuvaus esitetään yleensä graafisena kaaviona
 - n kaaviotekniikoissa paljon vaihtelua
 - n tällä kurssilla käytetään graafisena esitystekniikkana UML:n luokkakaaviota

2

ER-mallin peruskäsitteet

- n Kohdetyyppi (yksilötyyppi, entity type)
 - n määritelmä joukolle samankaltaisia kohteita, joihin liittyviä tietoja halutaan tallentaa tai käsitellä
 - n kohteet voivat olla abstrakteja tai konkreettisia
 - n Esimerkkejä:
 - n opiskelija
 - n kurssi
 - n oppitunti
 - n kurssisuoritus
 - n ilmoittautuminen
 - n suunnittelija jäsentää kohdealuetta ja päättää kohdetyypit

3

ER-mallin peruskäsitteet

- n kohdetyyppi (UML kaaviossa)

4

ER-mallin peruskäsitteet

- n Attribuutti (attribute)
 - n Kohteiden ominaisuudet esitetään datana kohdetyypikohtaisten attribuuttien ja niiden arvojen avulla
 - n kohteiden ominaisuudelle annettu nimi
 - n attribuutit määritellään tyyppitasolla = kaikille kohdetyypin kohteille
 - n Esimerkkejä kohdetyypin opiskelija liittyvistä attribuuteista:
 - n nimi
 - n syntymävuosi
 - n opiskelun aloitusvuosi
 - n pääaine

5

ER-mallin peruskäsitteet

- n kohdetyyppi: opettaja
- n attribuutti: sukupuoli
- n arvojoukko: {M (=mies), N (=nainen), T (=tuntematon)}

6

ER-mallin peruskäsitteet

- Attribuutti:
 - voi olla yksiarvoinen (*opiskelija:syntymävuosi*) tai moniarvoinen (*opettaja:suoritettu tutkinto*)
 - arvo voi olla pysyvä (*opiskelija:syntymävuosi*) tai muuttuva (*opiskelija:osoite*)
 - voi olla pakollinen (*opiskelija:nimi*) tai valinnainen (*opiskelija:puhelinnumero*)
 - arvoa voidaan käyttää kohteen identifiointiin (*opiskelija:opiskelijanumero*)

7

ER-mallin peruskäsitteet

- Attribuutit UML-luokkakaaviossa

8

ER-mallin peruskäsitteet

- Yhteys (relationship)
 - mallintaa kohteiden välistä kytkentää
 - opiskelija* on saanut aikaan *kurssisuorituksen*
 - oppitunti* kuuluu *kurssille*
 - ilmoittautuminen* kohdistuu *kurssiin*
 - ilmoittautuminen* koskee *opiskelijaa*

Tieto yhteyden olemassaolosta tallennetaan datana
 *luento 31.10.07 10-12 salissa 'B123 kuuluu TIKAPE kurssiin'

9

ER-mallin peruskäsitteet

- Yhteydet määritellään kohdetyyppien välisinä ja toteutuvat yksilötasolla tyyppien kohteiden välisinä
- Yhteydelle voidaan antaa nimi, samoin osapuolten roolille yhteydessä
 - opiskelija* on saanut aikaan *kurssisuorituksen*
 - nimi voisi olla *suoritus*
 - opiskelijan rooli: suorittaja*
 - kurssisuorituksen rooli: kohde*

10

ER-mallin peruskäsitteet

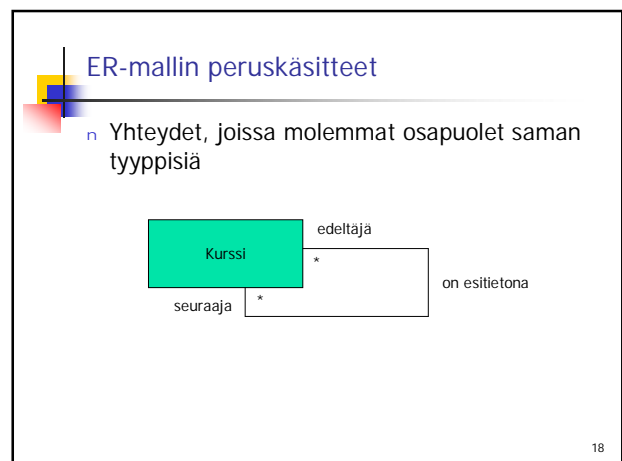
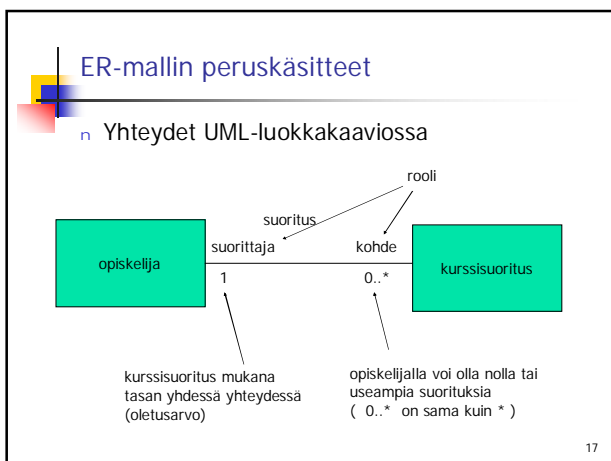
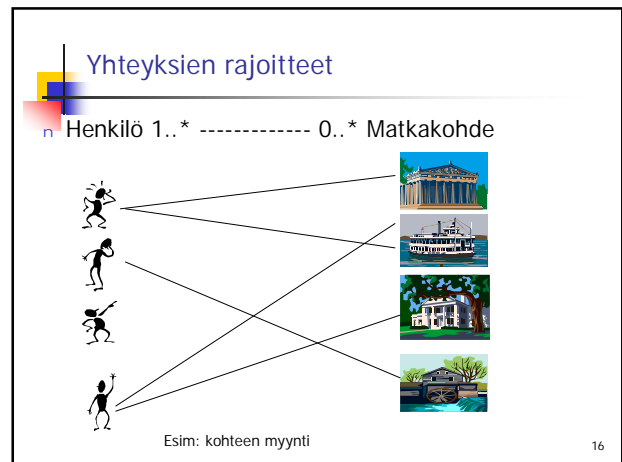
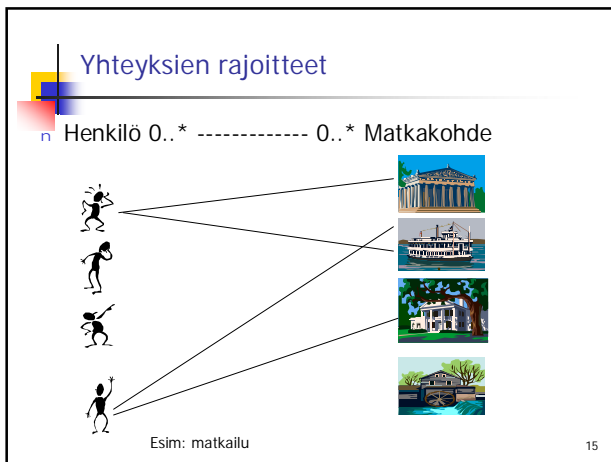
- Yhteydet
 - osallistumisrajoitteilla* kuvataan kohdetyypin ja yhteyden välistä suhdetta (rakenteellista sääntöä) – miten kohdetyypin kohteet osallistuvat yhteyksiin
 - suoritus
 - rooli suorittaja*: *opiskelija* voi olla tässä roolissa kytkeytyneet moneen suoritukseen mutta ei välttämättä yhteenkään
 - rooli kohde*: *kurssisuoritus* voi olla tässä roolissa kytkeytyneet vain yhteen opiskelijaan, eli kurssisuorituksella on vain yksi suorittaja.

11

Yhteyksien rajoitteet

- Henkilöllä* ----- 0..1 *Matkakohdetta*

12



ER-mallin peruskäsitteet

- Yhteyksien attribuutit
 - Joissain ER-malliperheen malleissa voidaan myös yhteyksiin liittää attribuutteja
 - henkilön ja yhdistyksen väliseen jäsenyisyhteyteen voitaisiin liittää attribuutti alkamisaika

```
graph LR
    h[henkilö] ---|*| y[yhdistys]
    h -.-|*| y
    subgraph jäsenyys
        a[alkamisaika]
    end
```

19

ER-mallin peruskäsitteet

- Alisteiset kohdetyytit (weak entity type)
 - Kohteet voivat olla riippumattomia tai alisteisia jollekin toiselle kohteelle. Alisteinen kohde on olemassaoloriippuva alistavasta kohteesta. Alistettu kohde on koko olemassaolonsa ajan kytkeytyneenä yhteyden kautta samaan kohteeseen.
 - huone alisteinen rakennukselle
 - kurssisuoritus alisteinen opiskelijalle
 - kurssisuoritus alisteinen kurssille

20

ER-mallin peruskäsitteet

- Alisteiset kohdetyytit ja alistussuhteet UML-luokkakaaviossa

```
classDiagram
    rakennus "1" *-- "*" huone
    opiskelija "1" *-- "*" kurssisuoritus
    kurssi "1" *-- "*" kurssisuoritus
```

21

ER-mallin peruskäsitteet

- Sisältyvät kohdetyytit

```
classDiagram
    henkilö <|-- johtaja
    henkilö <|-- nainen
```

22

Tietosisällön analysointi

- Kokonaisvaltainen lähestymistapa:
 - pyritään löytämään kerralla koko tietosisältöä kuvaava malli
 - hankalaa, jos kohdealue on laaja
 - ensin karkea yleiskuva, sitten lisää yksityiskohtia
 - Tietokohteet ryhminä, joita sitten tarkennetaan
- Osista kokonaisuuteen:
 - jaetaan kokonaisuus osiin ja tehdään osakohtaisia malleja, jotka sitten yhdistetään kokonaisuudeksi
 - osa voisi olla jonkin käyttötapauksen tietosisältö
 - Edetään alhaalta ylös yksityiskohdista yleiskuvaan
 - Yhdistelyn ongelmat (homonymit, synonyymit, mallintaminen eri tavoin)

23

Määrittelyprosessi

- Kartoita tietokohde-ehdokkaita.
- Karsi ehdokkaita.
- Tunnista tietokohteiden väliset yhteydet.
- Täsmennä kohdekuvauksia määrittelemällä attribuutit ja niiden arvojoukot.
- Määrittele yhteyksiin liittyvät osallistumisrajoitteet.
- Tarkista onko määritelty tietosisältö tarkoitukseensa soveltuva, täydennä

24

Kohteiden kartoitus

- n Laadi luettelo tarkasteltavan ilmiön kannalta keskeisistä kohteista tai ilmiöistä, jotka voisivat tulla kyseeseen tietokohteina:
 - n toimintaan osallistujat,
 - n toiminnan kohteet,
 - n toimintaan liittyvät tapahtumat,
 - n materiaalit,
 - n tuotteet ja välituotteet,
 - n toiminnalle edellytyksiä luovat asiat.

25

Tekstianalyysi kartoitustapana

- n Kartoituksen pohjana voi käyttää vapaamuotoista tekstikuvausta tarkasteltavasta ilmiöstä.
 - n Kuvauksesta alleviivataan tietokohde-ehdokkaita ja kerätään ne luetteloon.
 - n Ehdokkaat esiintyvät kuvauksessa usein subjekteina tai objekteina.
 - n Verbit voivat ilmaista yhteyksiä, samoin genetiivit.
 - n Alustavaa karsintaa voi tehdä sen perusteella, onko asia lainkaan oleellinen mallinnettavan ilmiön kannalta.
- n Yksinkertaista, mutta toimivaa
 - n käytännössä kuitenkin tekstikuvausta ei ole
 - n tiedot pitää kaivaa esiin haastatteleamalla ja tutustumalla erilaisiin dokumentteihin
- n >>> **Katso suunnitteluistunnon nauhoitus verkkomateriaalista**

26

Analysoi kartoituksen tulosta

- n Löydetyt ehdokkaat käydään läpi ja arvioidaan sopiiko ehdokas tietokohteeksi
 - n Liittykö kohteeseen tietosisältöä, joka on välttämätöntä järjestelmän kannalta
 - n yleensä pitäisi olla useita attribuutteja
 - n Tarvitaanko tietoa kohteiden olemassaolosta?
 - n Onko asia riittävän tärkeä kohdealueen kannalta?
 - n Eroon synonyymeistä
- n Karsintaa ja ehdokkaiden kartoitusta voidaan joutua tekemään iteratiivisesti. Ensimmäinen karsintakierros ei välttämättä tuota lopullista tulosta.

27

Määrittele yhteydet

- n Yhteyksiä kohteiden välille voi etsiä vapaamuotoisesta kuvauksesta:
 - n verbit
 - n genetiivit
 - n muut ilmaukset, jotka kuvaavat kytkentää
- n Yhteyksienkin suhteen tulisi miettiä
 - n onko yhteys oleellinen tarkasteltavan ilmiön kannalta?
 - n onko se rakenteellinen? (vallitseva asiointi)
 - n Ei pitäisi jäädä irrallisia kohteita
- n Asia pitäisi esittää vain kertaalleen:
 - n johdettavissa olevat yhteydet karsitaan tai merkitään
- n **Älä piilota yhteyksiä attribuuteiksi!**

28

Määrittele attribuutit

- n Attribuutteja saattaa löytyä vapaamuotoisesta kuvauksesta.
- n Yleensä niiden löytäminen edellyttää lisäselvityksiä kohdealueesta, esimerkiksi toiminnan osapuolten haastatteluja.
- n Attribuuttien kohdalla pitäisi myös selvittää, mihin niitä tarvitaan.
- n Ja uudelleen:
 - n Älä piilota yhteyksiä attribuutteihin!

29

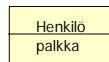
Selvitä yhteyksiin liittyvät rajoitteet

- n Osallistumisrajoitteiden avulla ilmaistaan rakenteellisia sääntöjä.
- n Säännöt eivät välttämättä tule esiin vapaamuotoisessa kuvauksessa, vaan edellyttävät tarkempaa kohdealueen analysointia.

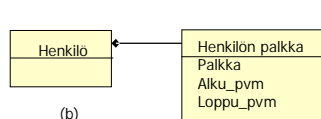
30

Eräitä mallinnusongelmia

- n Tietoon liittyvät metatiedot,
- n Esim. Työntekijän palkka
 - n Jos halutaan tietää vain nykyinen palkka voidaan palkka määritellä työntekijän attribuutiksi (a)
 - n Jos halutaan säilyttää palkkahistoria, asia tulee hankalammaksi, palkan nauttiminen onkin mallinnettava tietokohteeksi, johon liittyy attribuutteja (b)



(a)



(b)

31