

Tietokantojen perusteet

ER-mallinnus - perusteita

Tietojärjestelmän ja tietokannan suunnittelu
Käsitteellinen mallintaminen
ER-mallinnus

Tietojärjestelmän ja tietokannan suunnittelu

- Tietojärjestelmiä ja tietokantoja suunniteltaessa kartoitetaan ja analysoidaan, minkälaisia tietoja ja tietotarpeita sovellusalueella on.
 - Mistä asioista tarvitaan tietoa?
 - Miten asiat liittyvät toisiinsa? Mitä yhteyksiä asioiden välillä on?
 - Mitä tietoja asioista ja niiden välisistä yhteyksistä tarvitaan?
 - Millä tavoin tietoja jatkossa käytetään?
- Suunnittelija mallintaa sovellusaluetta sovellusalueen asiantuntijoiden antamien tietojen perusteella.
 - Sovellusalueella käytettävä ja sovellusaluetta kuvaava materiaali, asiantuntijoiden haastattelut, havainnointi

Tietojärjestelmän ja tietokannan suunnittelu

- Tietokannan tietosisällön suunnittelussa käytetään yleensä jotakin **käsitteellistä mallia**, jossa sovellusalueen tiedot kuvataan loppukäyttäjien (sovellusalueen asiantuntijoiden) käsitteitä käyttäen.
 - Puhutaan **käsitteellisestä suunnittelusta** ja **käsitteellisestä mallintamisesta**.
 - Tällä kurssilla käytetään ER-mallia käsitteelliseen mallintamiseen.
- Käsitteellisen suunnittelun jälkeen on vuorossa **rakenne- eli toteutustason suunnittelu (looginen suunnittelu)**.
 - Suunnitellaan esimerkiksi, mitä tietoja sijoitetaan mihinkin tauluihin.
- Viimeisenä vaiheena on vuorossa **fyysisen tason suunnittelu**.
 - Suunnitellaan esimerkiksi, mitä hakemistorakenteita käytetään hakujen nopeuttamiseksi.

Tietojärjestelmän ja tietokannan suunnittelu

- Tietokannan tietoja tarkastellaan siis eri abstraktiotasoilla erilaisia tietomalleja käyttäen.
 - **käsitetaso** (conceptual level)
 - Mitä tietoja käsitellään? Miten tiedot liittyvät toisiinsa?
 - esim. ER-malli
 - sisällön suunnittelu
 - **rakennetaso, toteutustaso** (structural level, implementation level)
 - Minkälaisina rakenteina tietoja käsitellään?
 - esim. SQL-tietokantojen taustalla oleva relaatiomalli
 - SQL-komentotulkin kautta tehtävät kyselyt ja päivitykset, ohjelmointi
 - **fyysinen taso** (physical level), talletustaso
 - Minkälaisina rakenteina tiedot talletetaan levyille?
 - tietokannanhallintajärjestelmän toimittajan käyttämät käsitteet
 - esim. hakuja nopeuttavat hakemistorakenteet

Tietojärjestelmän ja tietokannan suunnittelu

- Eri tasoilla olevien tietomallien avulla pystytään kuvaamaan käyttötarkoituksen kannalta olennaiset asiat.
- Tietomalli on käsitteistö, jolla kuvataan tietojen rakenne.
 - Tietojen tyyppi
 - Merkkijono, kokonaisluku, desimaaliluku, jne.
 - Tietojen väliset suhteet
 - Miten tiedot liittyvät toisiinsa
 - Tietoja koskevat rajoitteet
 - Mitkä ovat sallittuja arvoja, onko tieto pakollinen, jne.

Lisäksi tietomallit voivat sisältää operaatioita tietojen hakuun ja päivitykseen.

Tietojärjestelmän ja tietokannan suunnittelu

- Käsitteellisen, loogisen ja fyysisen suunnittelun lisäksi tarvitaan funktionaalista eli toiminnallista suunnittelua.
 - Minkälaisia sovellusohjelmia tarvitaan?
 - Minkälaisia tietokantatapahtumia sovellusohjelmissa tarvitaan?
 - Tapahtuma voi koostua esim. tietojen hausta ja päivityksestä.
 - Tietokantatapahtumiin liittyviä asioita opiskellaan TIETA7 Tietokantaohjelmointi -opintojaksolla.
 - Minkälaisia SELECT-, INSERT-, UPDATE- ja DELETE-lauseita tarvitaan?

ER-malli

- ER-malli (entity-relationship model) on käsitteellisen tason tietomalli.
 - alunperin Peter Chenin vuonna 1976 esittämä
 - artikkelissa The entity-relationship model – toward a unified view of data, ACM Transactions on Database Systems (artikkeli saatavilla kurssimateriaalissa)
- Malli on yleisesti käytössä tietokannan suunnittelussa ja tietojärjestelmien suunnittelussa.
- ER-mallissa sovellusalueen tiedot kuvataan
 - entiteetteinä (yksilöinä) (entity),
 - entiteettien välisinä suhteina (yhteyksinä) (relationship) ja
 - edellisiin liittyvinä ominaisuuksina (attribute).
- ER-mallin perusesitysmuoto on graafinen.
 - Piirretään ER-kaavioita.

ER-malli

- ER-mallista on olemassa useita versioita. Peruskäsitteet ovat yleensä yhteisiä eri versioille, mutta esitystavassa l. notaatiossa käytettävät merkinnät ja niiden merkitys vaihtelevat.
- Näissä kalvoissa käytetään kirjan Elmasri, Navathe: Fundamentals of Database Systems esitystapaa, joka vastaa suurelta osin Peter Chenin v. 1976 ehdottamaa esitystapaa.
- Kalvoilla käydään läpi ER-mallin peruskäsitteet ja esitystapa käyttämällä esimerkkitietokantana kuvitteellista yritystä, jonka tietojärjestelmän tietokantaan tarvitaan tietoja yrityksen työntekijöistä ja heidän huollettavistaan, yrityksen osastoista ja projekteista sekä näiden välisistä yhteyksistä. (Ks. sovellusalueen kuvaus erillisestä dokumentista.)

Entiteetti ja attribuutti

- **Entiteetti** (yksilö, kohde, tavallinen entiteetti tai vahva entiteetti, entity) on sovellusalueella olemassaoleva yksilö (olio), joka on erotettavissa muista sovellusalueen yksilöistä.
- Entiteetti voi olla konkreettinen tai abstrakti.
- Kullakin entiteetillä on **attribuutteja eli ominaisuuksia**, jotka kuvaavat sitä.
 - Esim. työntekijällä työntekijänumero, etunimi, sukunimi, syntymäaika, jne.
- Kullakin entiteetillä on attribuuteille omat arvonsa.



työntekijä-entiteettejä

(88, Jukka, Susi, 1957-11-10 , Tampere, 5500.00 , 444 1234)
(33, Ville, Viima, 1975-12-08 , Nokia, 4000.50 , 444 4343)
(12, Pekka, Puro, 1985-01-09, Tampere, 3000.00,)
(98, Jenni, Joki, 1961-06-20, Lempäälä, 4300.00, 444 4488)

Entiteettityyppi

- Samankaltaiset entiteetit eli entiteetit, joilla on samat ominaisuudet, kuuluvat samaan **entiteettityyppiin** (kohdetyyppiin, yksilötyyppiin, entity type).
- Entiteettityypit kuvataan suorakaiteina, attribuutit ovaaleina.

attribuutti
(ominaisuus)



avain
(tunnisteominaisuus),
merkitty alleviivauksella

entiteettityyppi

Entiteettityyppi
määrittelee,
mitä
ominaisuuksia
siihen
kuuluvilla
entiteeteillä on.

Käsite ja ilmentymä

- ER-mallissa tehdään ero käsitteiden ja niiden ilmentymien välillä.
- On olemassa yksi käsite (esim. työntekijä) ja mahdollisesti lukuisia (tai ei yhtään) sen ilmentymiä.

käsite
työntekijä



henkilö Jenni Joki –
eräs työntekijä-
käsitteen ilmentymä



(98, Jenni, Joki, 1961-06-20, Lempäälä, 4300.00, 444 4488)

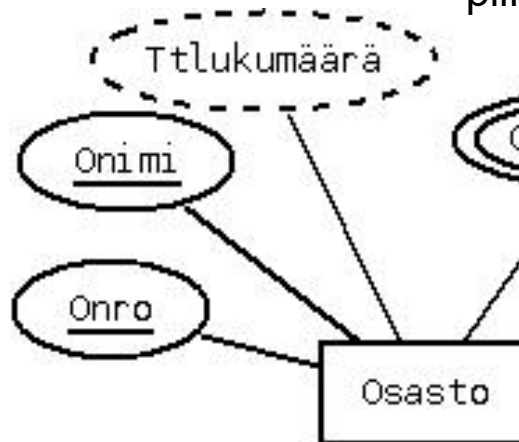
Entiteetti ja attribuutti

- Osa attribuuteista on **avaimia** eli **tunnisteominaisuuksia**.
 - Avaimen arvot ovat toisistaan poikkevia ts. kullakin entiteetillä on oma, uniikki arvonsa avaimelle
 - Käytetään entiteettien tunnistamiseen.
- Avain voi muodostua **yhdestä attribuutista** tai **usean attribuutin yhdistelmästä**. Avaimen otetaan mukaan minimimäärä attribuutteja, jotka riittävät tekemään siitä yksilöivän.
- Avaimen yksilöintikyvyn tulee olla voimassa sovellusalueen kaikissa mahdollisissa samantyyppisten entiteettien joukoissa ts. kaikissa mahdollisissa entiteettityypin ilmentymien joukoissa (ekstensioissa).
- Entiteettityypillä voi olla useita avaimia.

Attribuutit

- Attribuutti eli ominaisuus voi olla
 - yksiarvoinen – moniarvoinen (syntymäaika – kielitaito)
 - tallennettava - johdettavissa oleva (syntymäaika – ikä)
 - pysyvä – muuttuva (syntymäaika – osoite)
 - pakollinen – valinnainen (henkilötunnus – puhelinnumero)

Yksiarvoinen tallennettava attribuutti merkitään yksinkertaisella yhtenäisellä viivalla piirretyllä ovaalilla.



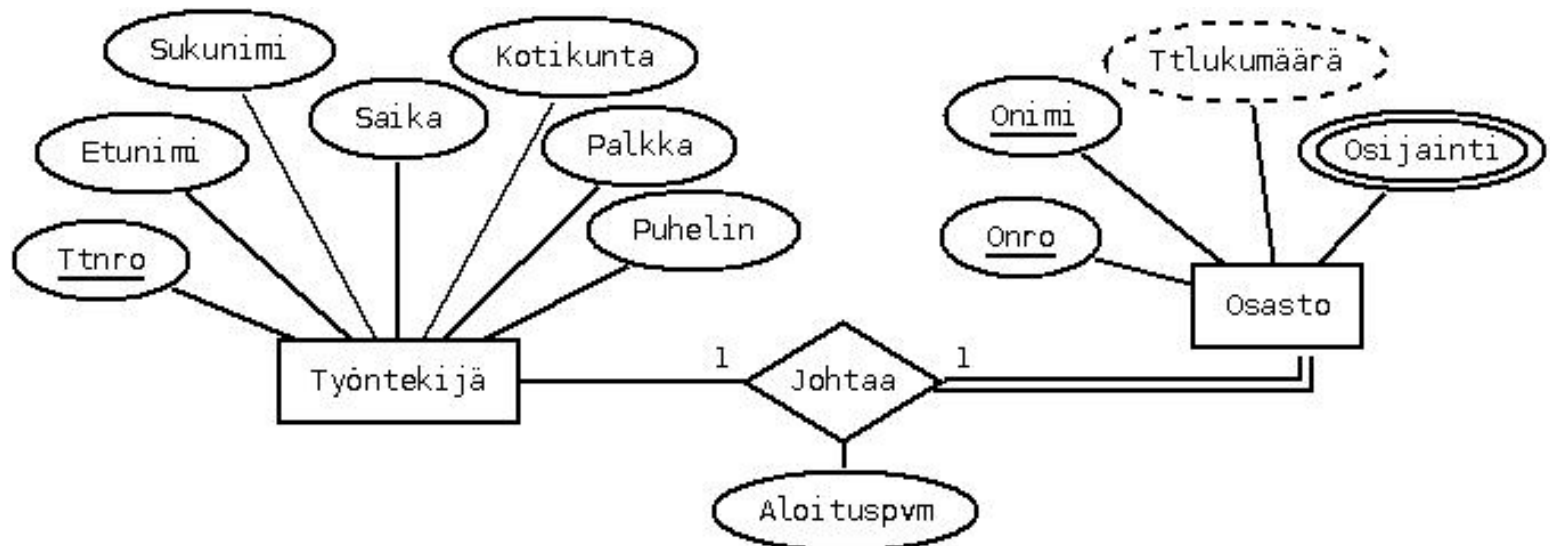
Johdettu attribuutti merkitään katkoviivalla piirretyllä ovaalilla (työntekijöiden lukumäärä)

Moniarvoinen attribuutti merkitään kaksinkertaisella viivalla piirretyllä ovaalilla.
(osastolla voi olla useita sijaintipaikkoja, toimipisteitä)

Kaksi avainta (osastot pystytään yksilöimään sekä nimen että numeron perusteella)

Suhde



- Sovellusalueen entiteettien välillä on yhteyksiä eli suhteita (relationships).
 - Esim. Jenni Joki työskentelee Hallinto-osastolla sekä johtaa sitä.
- Samankaltaiset suhteet voidaan koota **suhdetyypiksi** (relationship type), jolla voi olla attribuutteja.
- Suhdetyypit merkitään vinoneliöillä.

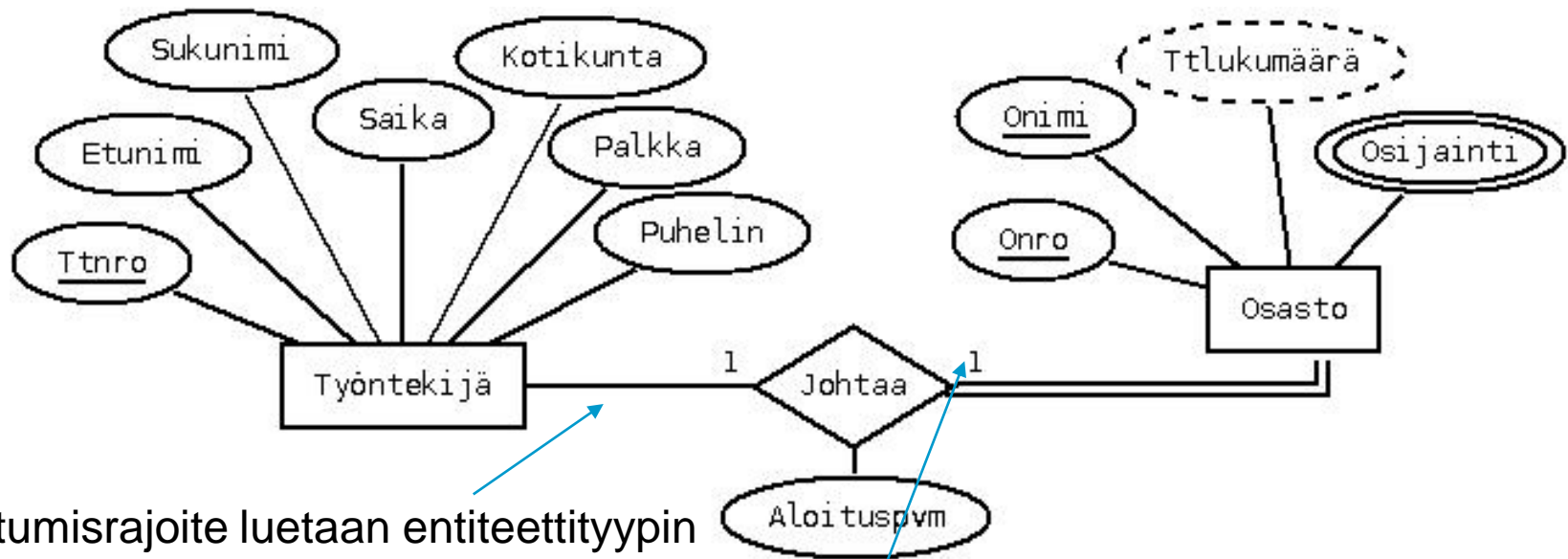


Suhdetyyppi: Rakenteelliset rajoitteet

- Sovellusalueella vallitsee suhteita koskevia rajoitteita, jotka määräävät, millä tavoin entiteetit voivat osallistua suhteisiin.
- Puhutaan **rakenteellisista rajoitteista**, joita ovat **lukumääräsuhteet ja osallistumisrajoitteet**
- **Lukumääräsuhteet** (kardinaliteetit)
 - **Kuinka moneen** tietyn suhdetyypin **suhteeseen** entiteetti voi **maksimissaan** osallistua?
 - Merkitään numeroin ja kirjaimin
 - **1:1**, yhden suhde yhteen
 - **1:N**, yhden suhde moneen
 - **M:N**, monen suhde moneen
 - **Luetaan vastakkaiselta puolelta**, suhdetyypin ”jälkeen” tulevasta numerosta tai kirjaimesta

Suhdetyyppi: Rakenteelliset rajoitteet

- **Osallistumisrajoitteet** (totaliteetit)
 - Onko entiteetin pakko osallistua tietyn suhdetyypin suhteeseen?
 - Merkitään viivoilla
 - **Osittainen osallistuminen** – yksinkertainen viiva 
Entiteetin ei ole pakko osallistua tietyn suhdetyypin suhteeseen.
 - **Täydellinen osallistuminen** – kaksinkertainen viiva 
Entiteetin on osallistuttava vähintään yhteen tietyn suhdetyypin suhteeseen. Täydellistä osallistumista kutsutaan myös **eksistenssiriippuvuudeksi**.
- Osallistumisrajoite (eli entiteetin tietyn suhdetyypin suhteiden minimimäärä) luetaan **entiteettityypistä lähtevästä viivasta**.

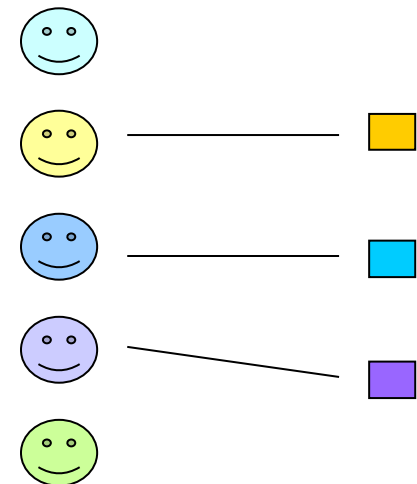


Osallistumisrajoite luetaan entiteettityypin itsensä puolelta: Työntekijöiden osallistuminen on **osittaista**, vain osa työntekijöistä johtaa osastoa (yksinkertainen viiva)

Lukumäärä luetaan entiteettityypin vastakkaiselta puolelta: työntekijä voi johtaa maksimissaan yhtä osastoa. (1)

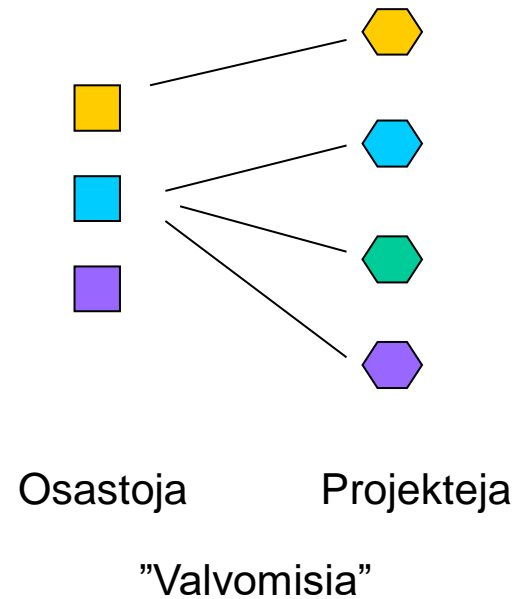
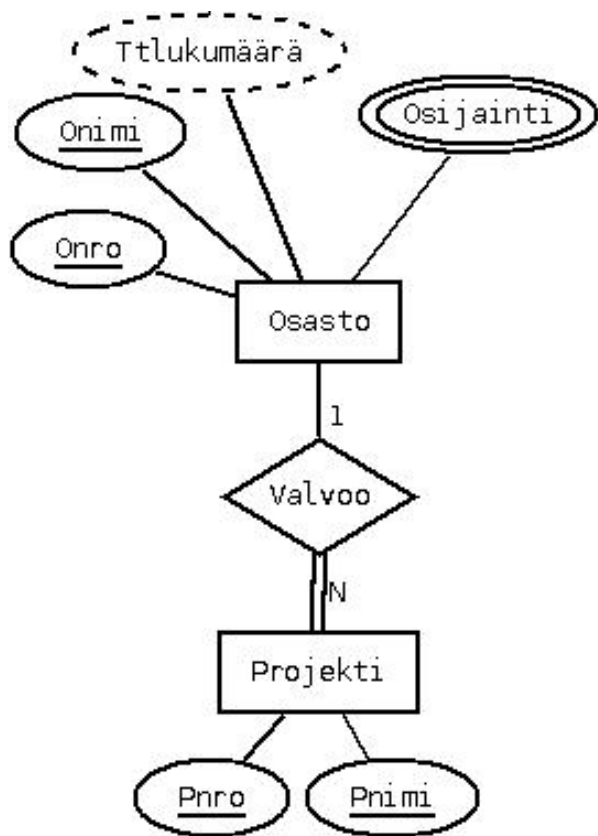
Osastojen osallistuminen on **täydellistä**: jokaisella osastolla on oltava johtaja (kaksinkertainen viiva). Osastoa johtaa maksimissaan yksi työntekijä.

Työntekijä:Osasto, **lukumääräsuhde 1:1**



"Johtamisia"

Työntekijöitä Osastoja
ER-mallinnus osa 1: 17



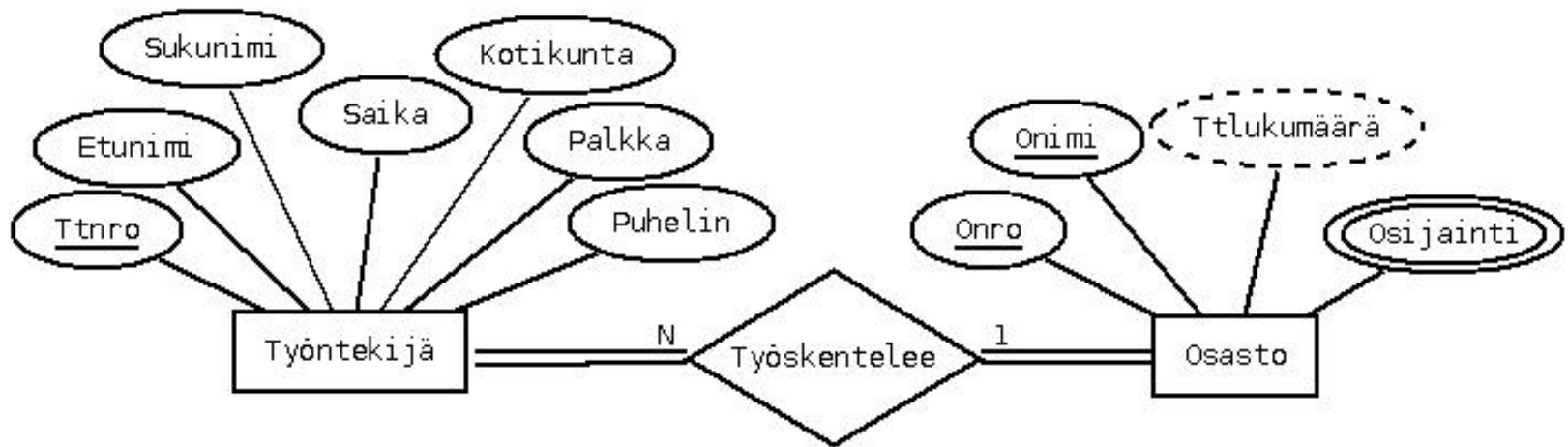
Osasto voi valvoa useaa projektia, projektia valvoo yksi osasto.

Osasto:Projekti, **lukumääräsuhte 1:N**

(Huom. ER-kaavion lukusuunta on tyypillisesti ylhäältä alaspäin, vasemmalta oikealle.)

Osasto ei välttämättä valvo ainuttakaan projektia. Osaston osallistuminen on osittaista.

Projektia valvoo vähintään yksi osasto. Projektin osallistuminen on täydellistä.

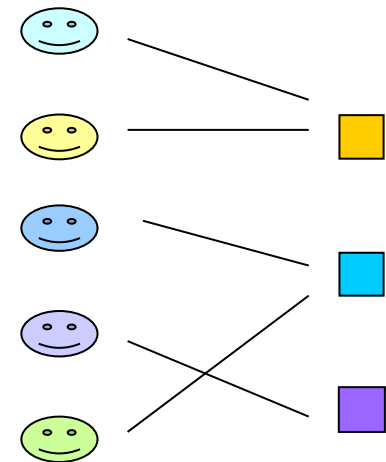


Työntekijä työskentelee enintään yhdellä osastolla. Osastolla voi työskennellä useita työntekijöitä.

Työntekijä:Osasto, **lukumääräsuhde N:1**

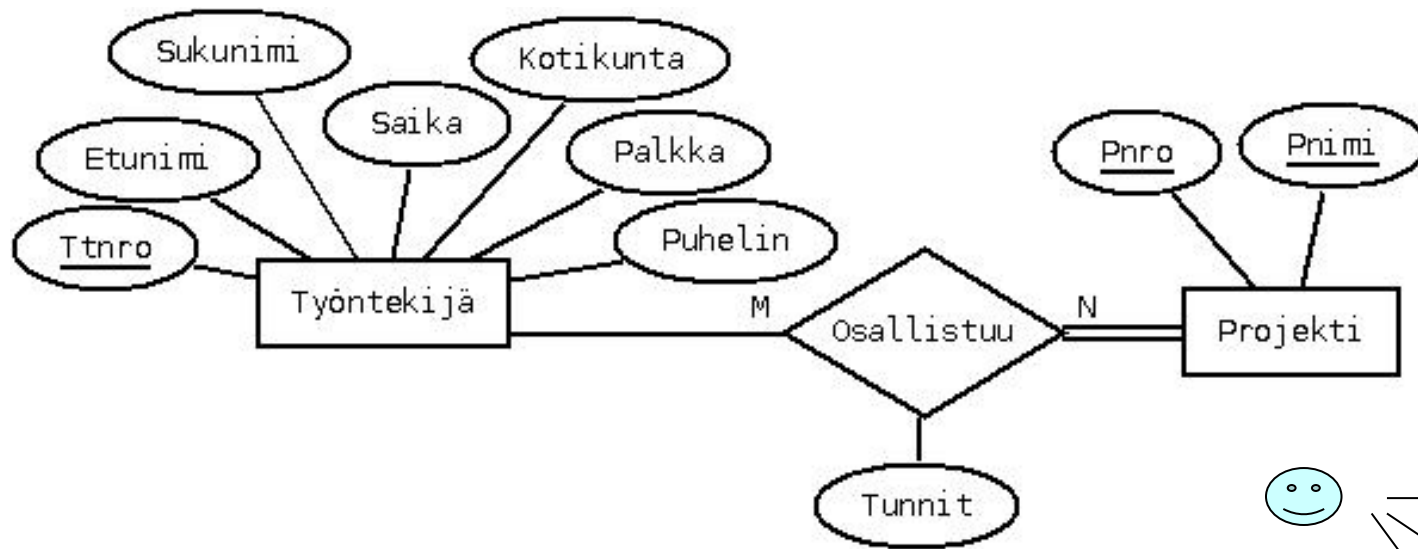
Jokainen työntekijä työskentelee vähintään yhdellä osastolla ja jokaisella osastolla työskentelee vähintään yksi työntekijä.

Sekä työntekijän että osaston osallistuminen on täydellistä.



"Työskentelyjä"

Työntekijöitä Osastoja

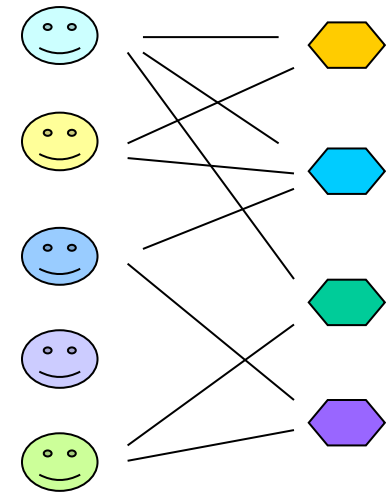


Työntekijä voi osallistua moneen projektiin.
Projektiin voi osallistua monta työntekijää.

Työntekijä:Projekti **lukumääräsuhde M:N**

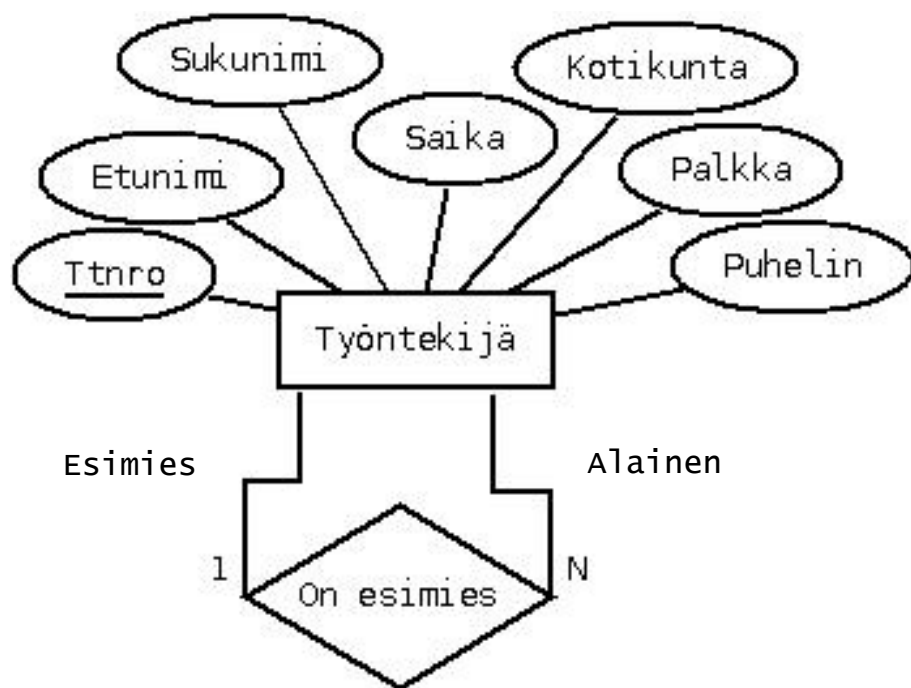
Työntekijä ei välttämättä osallistu yhteenkään projektiin. Työntekijän osallistuminen on osittaista.

Projektiin on osallistuttava vähintään yhden työntekijän. Projektin osallistuminen on täydellistä.



Työntekijöitä Projekteja
"Osallistumisia"

Roolit



Osa työntekijöistä toimii esimiehinä toisille työntekijöille: osa työntekijöistä on esimiehen **roolissa**.

Osalla työntekijöistä on esimies: osa työntekijöistä on alaisen roolissa.

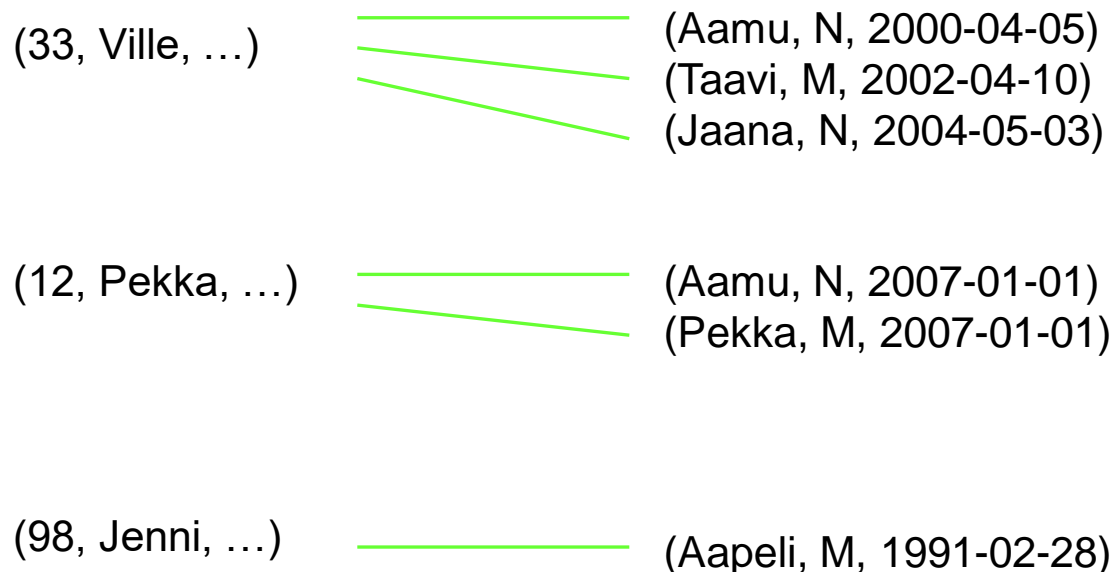
Esimies:Alainen 1:N

Heikko entiteettityyppi ja tunnistava suhdetyyppi

- Entiteettityyppiä, jolla ei ole omaa avainta, kutsutaan **heikoksi entiteettityypiksi** (weak entity type).
 - huollettava
- Heikolla entiteettityypillä on yleensä **osittainen avain**.
 - huollettavan nimi
- Heikon entiteetin tunnistamisessa käytetään toisen entiteettityypin entiteetiltä saatavaa avainta (toisten entiteettityyppien entiteeteiltä saatavia avaimia).
- Tätä toista entiteettityyppiä kutsutaan **tunnistavaksi** (identifying) tai **omistavaksi** (owner) entiteettityypiksi.
 - työntekijä
- Omistavia entiteettityyppejä voi olla useita.

Heikko entiteettityyppi ja tunnistava suhdetyyppi

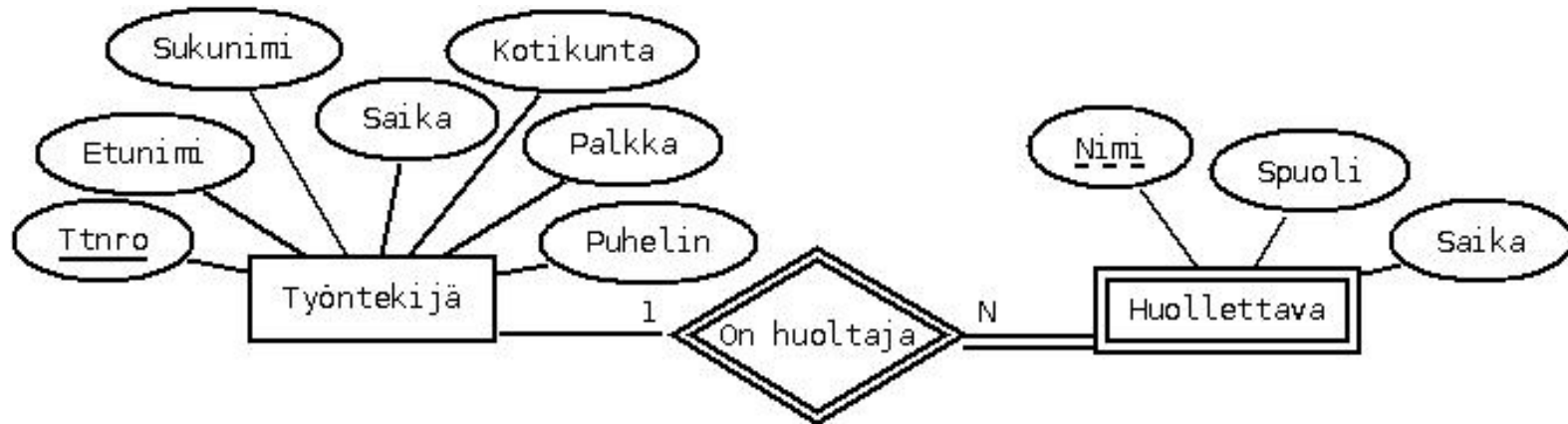
- Tunnistaminen: omistajien avaimet + mahdollinen heikon entiteetin osittainen avain
 - työntekijän ttnro ja huollettavan nimi
- Osittaisen avaimen avulla voidaan yksilöidä tiettyyn tunnistavaan entiteettiin (tunnistaviin entiteetteihin) liittyvät heikot entiteetit.



Heikko entiteettityyppi ja tunnistava suhdetyyppi

tunnistava
(omistava)
entiteettityyppi

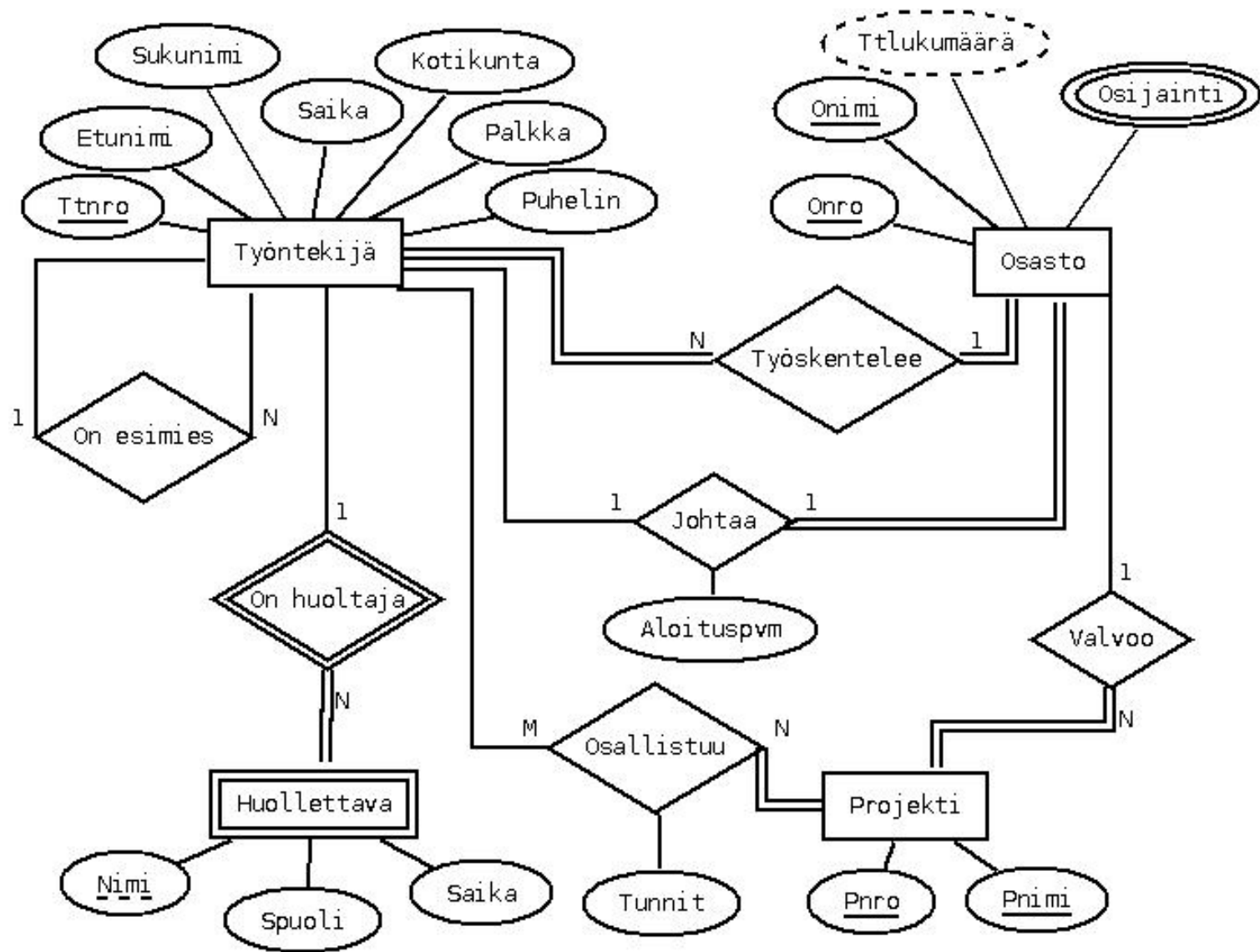
heikko
entiteettityyppi,
jolla osittainen
avain (nimi)



tunnistava
suhdetyyppi

Heikko entiteettityyppi ja tunnistava suhdetyyppi

- Heikon entiteetin on pakko osallistua tunnistavaan suhteeseen.
 - Heikon entiteettityypin osalta kyseessä on siis täydellinen osallistuminen, eksistenssiriippuvuus.
- Tunnistavien entiteettien osallistuminen on osittaista tai täydellistä.
- Tunnistaminen voi tapahtua epäsuorasti, koska tunnistava suhde voi johtaa toiseen heikkoon entiteettiin.
 - Huom! Oletetaan, että mikään entiteetti ei ole epäsuorastikaan tunnistavassa suhteessa itsensä kanssa.
- Heikko entiteettityyppi voi osallistua suhdetyyppeihin (muihin kuin tunnistava suhdetyyppi) samalla tavalla kuin tavallisetkin entiteettityypit.



ER-mallinnus: Käytännön ohjeita

- ER-kaavioon otetaan mukaan ainoastaan kaikki rakennettavan tietojärjestelmän kannalta olennaiset tiedot.
- Tunnista ensin entiteettityypit ja niiden attribuutit.
 - Mistä asioista pitää tallentaa tietoa? Entiteettityypit
 - Mitä tietoja pitää tallentaa? Entiteettityyppien attribuutit
- Tunnista suhdetyypit ja niiden attribuutit
 - Miten asiat eli entiteettityypit liittyvät toisiinsa? Mitä yhteyksiä asioiden välillä on? Suhdetyypit
 - Mitä tietoja asioiden välisistä yhteyksistä pitää tallentaa? Suhdetyyppien attribuutit
- Tarkista, että entiteetti- ja suhdetyypeillä on yksikäsitteiset nimet.

ER-mallinnus: Käytännön ohjeita

- ER-kaavion tekeminen on yleensä iteratiivinen prosessi.
 - Etsitään ensin entiteettityyppejä, attribuutti- ja suhdetyyppehdokkaita, sitten karsitaan. Etsitään uusia ehdokkaita, joita taas mahdollisesti karsitaan.
 - Attribuutit voivat muuttua entiteettityypeiksi (tai päinvastoin).
 - Attribuutit voivat muuttua suhdetyypeiksi (tai päinvastoin).
- Tekstianalyysi
 - potentiaalisia entiteettityyppejä
 - substantiivit
 - subjektit, objektit
 - potentiaalisia attribuutteja
 - substantiivit, adjektiivit
 - potentiaalisia suhdetyyppejä
 - verbit, genetiivit sekä muut ilmaukset, jotka kuvaavat kohteiden välisiä yhteyksiä
- Saman sovellusalueen voi yleensä mallintaa usealla eri tavalla.
 - Eri ratkaisuissa omat hyvät ja huonot puolensa