hyväksymispäivä arvosana

arvostelija

# Epätäydellisen tiedon funktionaaliset riippuvuudet

Nikita Zhuk

Helsinki 14.3.2008

Seminaariraportti

HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI


| Tiedekunta — Fakultet — Faculty | Laitos — Institution — Department |
|---|---|
| Matemaattis-luonnontieteellinen tiedekunta | Tietojenkäsittelytieteen laitos |

Tekijä — Författare — Author

Nikita Zhuk


Työn nimi — Arbetets titel — Title

Epätäydellisen tiedon funktionaaliset riippuvuudet

Oppiaine — Läroämne — Subject

Tietojenkäsittelytiede

| Työn laji — Arbetets art — Level | Aika — Datum — Month and year | Sivumäärä — Sidoantal — Number of pages |
|---|---|---|
| Seminaariraportti | 14.3.2008 | 14 sivua |


Tiivistelmä — Referat — Abstract

Funktionaaliset riippuvuudet ovat tärkeitä työkaluja relaatiotietokannan suunnittelussa, sillä ne ovat konseptuaalisesti järkevän tietokantakaavion suunnittelun taustalla. Ne on määritelty universaalirelaatiossa, jossa kaikki riippuvuuksien osana oleva tieto on tunnettu. Käytännön sovelluksissa käsityksemme reaalimaailmasta on kuitenkin puutteellinen, ja tällöin funktionaaliset riippuvuudet tulisi määritellä kontekstissa, jossa tällaista epätäydellistä tietoa esiintyy. Tämä raportti määrittelee funktionaaliset riippuvuudet epätäydellisen tiedon kontekstissa sekä esittelee funktionaalisten riippuvuuksien vahvan ja heikon voimassaolon käsitteet. Armstrongin aksioomiin perustuva päättelyjärjestelmä yleistetään epätäydellistä tietoa sisältävissä relaatioissa esiintyville funktionaalisille riippuvuuksille. Raportissa esitettyjen tietojen avulla voidaan hyödyntää laajaa funktionaalisten riippuvuuksien käyttöön perustuvaa kirjallisuutta (mm. tietokantojen normalisointi) epätäydellisen tiedon kontekstissa.

ACM Computing Classification System (CCS):
H.2.1 [Logical Design]



Avainsanat — Nyckelord — Keywords

funktionaaliset riippuvuudet, epätäydellinen tieto, null-arvo, tyhjäarvo, täydentymä

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — övriga uppgifter — Additional information

# Sisältö

# 1 Johdanto

Relaatiotietokantojen suunnitteluteoria ja metodit ovat olleet tutkimuksen kohteena 1970-luvun alun relaatiotietokantoja koskevien artikkeleiden [Cod70, Cod71] julkaisusta lähtien [Vas80]. Tietokantakaavion suunnittelussa tärkeässä asemassa ovat funktionaaliset riippuvuudet, jotka ovat formaaleja rajoitteita tiedon eri osien välillä ja jotka toimivat päätyökaluina tiedon ryhmittelyssä relaatiokaavioihin [EN99]. Riippuvuudet ovat syntaktinen keino kuvata tiedon sisällä olevia semanttisia riippuvuuksia [Vas80]. Ne ovat konseptuaalisesti järkevän tietokantakaavion suunnittelun taustalla sekä tärkeässä roolissa tietokantakaavion normalisoinnissa, eli päivitysanomalioiden kannalta edullisimman rakenteen löytämisessä.

Funktionaaliset riippuvuudet on määritelty saman relaatiokaavion kahden attribuuttijoukon välillä [Vas80]. Tästä määritelmästä seuraa oletus, että kaikki reaalimaailman mallinnettavan osan attribuutit ovat osana samaa relaatiokaaviota, jonka ilmentymää kutsutaan universaalirelaatioksi. Mikä tahansa monen relaatiokaavion tietokantakaavio, joka on tuotettu normalisointiprosessin avulla universaalirelaatiosta, voidaan nähdä kokoelmana projektioita universaalirelaatiokaaviosta. Käytännössä voi tulla vastaan tilanteita, joissa tämän universaalirelaatiokaavion ilmentymän kaikkien rivien kaikkia attribuutteja ei voida täyttää arvoilla, jolloin tätä epätäydellistä tietoa ilmaistaan ottamalla käyttöön *tyhjäarvo* (*null-arvo*). Relaatiokaavion funktionaaliset riippuvuudet on kuitenkin määritelty attribuuttijoukoille, joissa tyhjäarvoja ei esiinny [Vas80, EN99]. Jos tyhjäarvot sallitaan relaatiokaavioiden ilmentymien attribuuttien arvoiksi, funktionaalisten riippuvuuksien määritelmää sekä riippuvuuksiin liittyviä voimassaolovaatimuksia pitää laajentaa kattamaan tyhjäarvot.

Tässä raportissa laajennetaan funktionaalisten riippuvuuksien määritelmää kattamaan myös tyhjäarvot. Lisäksi määritellään funktionaalisten riippuvuuksien voimassaolo ja tutkitaan niiden päättelyjärjestelmiä tilanteessa, kun tyhjäarvot ovat sallittuja. Seuraavassa kohdassa 2 esitetään tyhjäarvon käsite sekä tyhjäarvoa käsittelevien funktioiden arvon tulkinta. Kohdassa 3 esitetään funktionaalisten riippuvuuksien perusrelaatiomallin mukainen määritelmä, joka pätee tilanteessa, jossa tyhjäarvoja ei esiinny. Kohta 4 laajentaa funktionaalisten riippuvuuksien määritelmää kattamaan tyhjäarvot. Kohdassa 5 esitellään Armstrongin aksioomiin perustuva riippuvuuksien päättelyjärjestelmä. Kohdassa 6 esitellään tapa tutkia funktionaalisten riippuvuuksien joukon voimassaoloa epätäydellisen tiedon kontekstissa. Yhteenvedossa koostetaan raportissa tehdyt tulokset.

# 2 Tyhjäarvot funktioiden attribuutteina

Tyhjäarvosta relaatiokaavion ilmentymän eli relaation attribuutin arvona käytetään tässä raportissa *puuttuvan* arvon tulkintaa [Vas79]. Puuttuvasta arvosta on kyse silloin, kun relaatiossa olevaa tyhjäarvoa vastaa reaalimaailmassa jokin olemassa oleva, todellinen arvo. Tämä arvo kuuluu relaatiokaavion attribuutin arvojoukkoon, mutta on sillä hetkellä tuntematon. Tällaisia puuttuvia arvoja voi esiintyä semanttisesti konsistentissa relaatiotietokannassa johtuen epätäydellisestä tiedosta reaalimaailman tilasta.

Tässä raportissa tyhjäarvot tullaan esittämään kaavoissa ja esimerkeissä termillä *null*. Se on käytössä tieteellisessä kirjallisuudessa [Vas79] symbolien " - " ja "@" lisäksi [IWL84, Vas79]. *Null*-termi on valittu siksi, että se on vakiintunut termi relaatiotietokantasovelluksissa. Tällä pyritään parantamaan raportin kaavojen ja esimerkkien luettavuutta.

Relaatiotietokannan käyttäjän tulee saada kyselyidensä vastaukseksi mahdollisesti epätäydellistä, mutta silti oikeellista tietoa myös silloin, kun tyhjäarvoja käytetään [IWL84]. Tietokantaan tallennettua tietoa käsittelevien funktioiden ja relaatio-operaatioiden tulee toimia semanttisesti oikein myös tyhjäarvojen yhteydessä. Tyhjäarvo pitää sisällään vähemmän tietoa kuin epätyhjä arvo, joten se approksimoi kaikkia muita arvoja [Vas80]. Tästä seuraa määritelmä: funktio *f*, joka saa parametrikseen tyhjäarvon, saa arvon $y$ jos $f$ saa arvon $y$ kaikilla parametrin arvojoukkoon kuuluvilla epätyhjillä arvoilla. Formaalisti, jos *dom(A)* on parametrin $A$ arvojoukko:

$$f(null) = \begin{cases} y, \text{ jos } f(x) = y \text{ kaikilla } x \in dom(A),\ x \neq null \\ \text{tuntematon muussa tapauksessa} \end{cases}$$

Funktio $f(null)$ ratkaistaan käymällä läpi parametrin arvo-alueen kaikki arvot ja sijoittamalla kukin arvo funktion parametriksi [Vas80]. Jos funktio saa kaikissa tapauksissa saman arvon, tyhjäarvon edustamalla puuttuvalla tiedolla ei ole väliä. Jos taas kaikki funktion saamat arvot eivät ole samoja, funktiota ei voida ratkaista ja funktion saama arvo on tällöin tuntematon.

# 3 Funktionaaliset riippuvuudet

Tässä kohdassa määritellään funktionaalisen riippuvuuden käsite. Määritelmä tehdään oletuksella, että attribuuteissa ei esiinny tyhjäarvoja. Kohdassa 4 määritelmää laajennetaan kattamaan tyhjäarvot.

Relaatiokaavion $R$ *funktionaalinen riippuvuus* on kahden attribuuttijoukon, $X$:n ja $Y$:n välinen relaatiokaavion $R$ eheysrajoite eli lauseke [EN99]:

$$R : X \rightarrow Y,$$

missä $X, Y \subseteq R$.

Funktionaalisen riippuvuuden mukaan relaatiokaavion $R$ ilmentymä $r$ voi sisältää vain sellaisia monikoita, jotka noudattavat tämän riippuvuuden määritelmää [Vas80].

Olkoon $R$ relaatiokaavio, $r$ sen ilmentymä ja $X$ ja $Y$ sen attribuuttijoukkoja. Funktionaalinen riippuvuus $R : X \rightarrow Y$ on voimassa $r$:ssä, jos mille tahansa kahdelle $r$:n monikolle $t$ ja $u$ pätee: jos $t[X] = u[X]$, niin $t[Y] = u[Y]$, eli attribuuttijoukon $Y$ arvot *riippuvat* attribuuttijoukon $X$ arvoista, tai attribuuttijoukon $X$ arvot *määrittävät* tai *identifioivat* attribuuttijoukon $Y$ arvot. $Y$:n sanotaan olevan *funktionaalisesti riippuvainen* $X$:stä.

Funktionaalisen riippuvuuden $R : X \rightarrow Y$ voimassaolo relaatiokaavion $R$ ilmentymässä $r$ merkitään $r \models X \rightarrow Y$ [EN99]. Funktionaalinen riippuvuus voidaan nähdä myös totuusarvoisena funktiona $f(r)$, missä $f = X \rightarrow Y$ ja $r$ on relaatiokaavion $R$ ilmentymä [Vas80]. Funktio $f$ saa arvon tosi, jos $r \models X \rightarrow Y$, ja arvon epätosi, jos $r \not\models X \rightarrow Y$.

| $r$ | leveyspiiri | pituuspiiri | paikkakunta |
|---|---|---|---|
| | 60 | 25 | Helsinki |
| | 60 | 24 | Inkoo |

Taulukko 1: Relaatiokaavion ilmentymä, jossa ei esiinny tyhjäarvoja.

**Esimerkki.** Relaatiokaavioon *R(leveyspiiri, pituuspiiri, paikkakunta)* on liitetty seuraava funktionaalinen riippuvuus:

$$f : \textit{leveyspiiri pituuspiiri} \rightarrow \textit{paikkakunta}$$

Taulukossa 1 on esitetty relaatiokaavion $R$ yksi mahdollinen ilmentymä $r$, jossa tämä riippuvuus on voimassa, siis $r \models$ *leveyspiiri pituuspiiri* $\rightarrow$ *paikkakunta*, joten $f(r) = tosi$.

# 4 Tyhjäarvot riippuvuuksissa

Seuraavaksi määritellään funktionaaliset riippuvuudet, jotka sallivat tyhjäarvot relaation attribuuttien arvoina, ja tutkitaan näiden riippuvuuksien voimassaoloa. Tässä tarkastellaan vain yhden funktionaalisen riippuvuuden voimassaoloa relaatiokaavion ilmentymässä. Useamman riippuvuuden yhtäaikaista voimassaoloa tarkastellaan tarkemmin kohdassa 6. Tyhjäarvon sisältävien monikkojen käsittelyssä tarvitaan *täydentymän* määritelmää.

Monikon $t$ *täydentymä t'* on monikko, jossa kukin attribuutti saa saman arvon kuin vastaava attribuutti $t$:ssä, jos arvo on epätyhjä. Jos attribuutin arvo $t$:ssä on tyhjäarvo, attribuutti *t'*:ssä saa yhden arvojoukkonsa arvoista [Vas80].

Määritellään joukko $AP$ relaatiokaavion $R$ mukaiselle monikolle $t$ kaikkien monikon täydentymien *t'* joukkona:

$$AP(t, R) = \{t' \mid t' \text{ on } t\text{:n täydentymä}\}$$

Relaatiokaavion ilmentymän $r$ *täydentymä r'* sisältää kaikki $r$:n monikot, joissa tyhjäarvoja ei esiinny, sekä tyhjäarvoja sisältävien monikoiden täydentymät siten, että kutakin tyhjäarvoja sisältävää $r$:n monikkoa $t$ kohden *r'* sisältää yhden $t$:n täydentymistä.

Määritellään attribuuttijoukolle $R$ projisoidun relaatiokaavion ilmentymän $r$ kaikkien täydentymien *r'* joukko:

$$AP(r, R) = \{r' \mid r \text{ on } r\text{:n täydentymä}\}$$

**Esimerkki** Relaatiokaavio *R(syntaika, yksilonumero, on_ elossa)* sallii tyhjäarvot ja $dom(on\_elossa) = \{\text{tosi}, \text{epätosi}\}$. Taulukossa 2 on esitetty relaatiokaavion $R$ yksi mahdollinen ilmentymä $r$. Taulukossa 3 on esitetty $r$:n tyhjäarvoja sisältävän monikon $t$ kaikki täydentymät. Taulukossa 4 on esitetty $r$:n kaikki täydentymät.

| *r* | syntaika | yksilonumero | on_elossa |
|---|---|---|---|
| *t* | 010170 | 101 | null |
| *u* | 010170 | 111 | tosi |

Taulukko 2: Relaatiokaavion $R$ tyhjäarvoja sisältävä ilmentymä $r$

| *AP(t,R)* | syntaika | yksilonumero | on_elossa |
|---|---|---|---|
| | 010170 | 101 | tosi |
| | 010170 | 101 | epätosi |

Taulukko 3: Taulukon 2 relaation $r$ monikon $t$ täydentymäjoukko *AP(t,R)*

Kohdassa 3 esitetyn määritelmän mukaan relaatiokaavion $R$ funktionaalinen riippuvuus *f*, joka ei salli tyhjäarvoja, on voimassa relaatiokaavion ilmentymässä $r$, kun funktio *f(r)* saa arvon tosi [Vas80]. Määritelmä voidaan esittää myös sijoittamalla yksi monikoista funktion parametriksi ja sanoa, että riippuvuus on voimassa silloin, kun funktio $f$ saa arvon tosi kaikilla ilmentymän $r$ monikoilla $t$:

$$f(t,r) = \begin{cases} \text{tosi, jos } t[X] = u[X], \text{ niin } t[Y] = u[Y], \text{kaikilla } u \in r \\ \text{epätosi muussa tapauksessa} \end{cases}$$

Täydentymän määritelmän avulla voidaan esittää tyhjäarvot salliva määritelmä funktionaaliselle riippuvuudelle $R : X \rightarrow Y$ funktiona *f'* [Vas80]. Tässä määritelmässä käytetään supremumin eli joukon pienimmän ylärajan käsitettä. Funktion $f$ arvojoukko on osittain järjestetty joukko $\{\text{tosi}, \text{epätosi}, \text{tuntematon}\}$, missä tosi $<$ tuntematon ja epätosi $<$ tuntematon. Arvojen tosi ja epätosi välillä ei ole järjestystä. Tästä seuraa, että funktion $f$ saamien arvojen joukon $F$ pienin yläraja on tosi,

| *r1* | syntaika | yksilonumero | on_elossa |
|---|---|---|---|
| *t* | 010170 | 101 | tosi |
| *u* | 010170 | 111 | tosi |
| *r2* | syntaika | yksilonumero | on_elossa |
| *t* | 010170 | 101 | epätosi |
| *u* | 010170 | 111 | tosi |

Taulukko 4: Taulukon 2 relaation $r$ täydentymäjoukko *AP(r,R)* = {r1, r2}

mikäli funktio $f$ saa arvokseen arvon tosi kaikilla parametreilla. Joukon $F$ pienin yläraja on epätosi, mikäli funktio $f$ saa arvokseen arvon epätosi kaikilla parametreilla. Joukon $F$ pienin yläraja on tuntematon kaikissa muissa tapauksissa.

$$f'(t,r) = \begin{cases} f(t,r), \text{ jos null} \notin t[XY], r[XY] \\ \sup\{\ f(t',r') \mid t' \in r',\ r' \in AP(r,XY)\ \} \text{ muussa tapauksessa} \end{cases}$$

Tästä määritelmästä voidaan johtaa kuusi tapausta, joiden avulla funktionaalisen riippuvuuden voimassaolo voidaan päätellä helpommin kuin soveltamalla määritelmää suoraan [Vas80]. Seuraavissa tapauksissa oletetaan, että relaatiokaavioon $R$ on liitetty funktionaalinen riippuvuus $X \to Y$, $r$ on relaatiokaavion $R$ ilmentymä, $X \cap Y = \emptyset$, $X \cup Y = R$, $t \in r$. Lisäksi oletetaan, että monikko $t$ on ainoa monikko $r$:ssä, jossa esiintyy tyhjäarvoja. Mikäli näin ei ole, tapauksia tulee soveltaa jokaiselle $r - \{t\}$:n täydentymälle erikseen iteratiivisesti.

**Tapaus 1** *t[XY]* ei sisällä tyhjäarvoja. Jos $r$ ei sisällä monikkoa $u$, jossa $t[X] = u[X]$ ja $t[Y] \neq u[Y]$, $f'(t,r) = \text{tosi}$.

**Tapaus 2** *t[Y]* sisältää tyhjäarvon, *t[X]* ei sisällä tyhjäarvoa. Jos *t[X]* on uniikki relaatiokaavion ilmentymässä $r$, ts. $r$ ei sisällä monikkoa $u$, jossa *t[X]* = *u[X]*, $f'(t,r) = \text{tosi}$.

**Tapaus 3** *t[X]* sisältää tyhjäarvon, *t[Y]* ei sisällä tyhjäarvoa. Jos ilmentymässä $r$ ei ole monikkoa $u$, jonka projektio $X$:lle *u[X]* olisi *t[X]*:n täydentymä, $f'(t,r) = \text{tosi}$. *t[X]* on tällöin uniikki ilmentymässä $r$ riippumatta valitusta täydentymästä.

**Tapaus 4** *t[X]* sisältää tyhjäarvon, *t[Y]* ei sisällä tyhjäarvoa. Jos ilmentymässä $r$ kaikkien monikkojen $u$ kohdalla pätee: jos projektio $X$:lle *u[X]* on *t[X]*:n täydentymä, niin *u[Y]* = *t[Y]*, $f'(t,r) = \text{tosi}$.

**Tapaus 5** *t[XY]* ei sisällä tyhjäarvoja. Jos $r$ sisältää monikon $u$, jossa $t[X] = u[X]$ ja $t[Y] \neq u[Y]$, $f'(t,r) = \text{epätosi}$.

**Tapaus 6** *t[X]* sisältää tyhjäarvon, *t[Y]* ei sisällä tyhjäarvoa. Jos ilmentymä $r$ sisältää kaikki monikon $t$ täydentymät $u$ ja jos *t[Y]* on uniikki näiden täydentymien joukossa, niin $f'(t,r) = \text{epätosi}$.

Kaikissa muissa tapauksissa funktion *f'(t,r)* arvoa ei voida päätellä yksiselitteisesti, joten se on tällöin tuntematon ja funktionaalisen riippuvuuden voimassaololle ei voida antaa yksiselitteistä totuusarvoa [Vas80].

Funktionaalisen riippuvuuden sanotaan olevan *vahvasti* voimassa, kun $f'(t,r) =$ tosi pätee kaikilla relaatiokaavion ilmentymän $r$ monikoilla $t$ [Vas80]. Funktionaalinen riippuvuus on *heikosti* voimassa, kun $f'(t,r) \neq$ epätosi pätee kaikilla relaatiokaavion ilmentymän $r$ monikoilla $t$. Epätäydellistä tietoa käsiteltäessä heikko voimassaolo on usein riittävä, koska se sallii jonkin verran epävarmuutta funktionaalisiin riippuvuuksiin. Samalla se kuitenkin takaa, ettei kannassa oleva tieto ole ristiriitaista - se ei siis riko sille asetettuja eheysrajoitteita.

# 5 Riippuvuuksien päättelyjärjestelmä

Osa relaatiokaavion riippuvuuksista voi olla implisiittisiä, kaavioon liitetyistä riippuvuuksista johdettuja riippuvuuksia, joita ei esitetä erikseen relaatiokaaviossa [EN99]. Riippuvuuksien johtamiseen käytetään ns. Armstrongin aksioomiin perustuvaa päättelyjärjestelmää, joita noudattamalla riippuvuusjoukosta voidaan johtaa uusia riippuvuuksia. Päättelyjärjestelmä koostuu yhdestä aksioomaskeemasta (*refleksiivisyys*) sekä kahdesta päättelysäännöstä, *täydennys* ja *transitiivisuus*.

- **Refleksiivisyys:** kaikki lauseet $X \to Y$, missä $Y \subseteq X \subseteq R$, ovat aksioomia.
- **Täydennys:** lauseesta $X \to Y$ voidaan päätellä lause $XZ \to YZ$, $Z \subseteq R$.
- **Transitiivisuus:** lauseista $X \to Y$ ja $Y \to Z$ voidaan päätellä lause $X \to Z$.

Olkoon relaatiokaavioon $R$ liitettynä funktionaalisten riippuvuuksien joukko $F$. Riippuvuusjoukosta $F$ voidaan johtaa riippuvuus $X_m \to Y_m$ konstruoimalla riippuvuuksien äärellinen jono $X_1 \to Y_1, X_2 \to Y_2, ..., X_m \to Y_m$, jos jokaiselle $k = 1, 2, ..., m$ pätee jokin seuraavista ehdoista [MR92]:

1. $X_k \to Y_k \in F$.
2. $X_k \to Y_k$ on aksiooma.
3. $X_k \to Y_k$ on päätelty aikaisemmista riippuvuuksista $X_i \to Y_i, 1 \leq i < k$.

Riippuvuuden $X \to Y$ sanotaan olevan $F$:n *syntaktinen seuraus* $F \vdash X \to Y$, jos riippuvuus voidaan johtaa F:stä yllä esitetyn päättelyjärjestelmän avulla. Riippuvuuden $X \to Y$ sanotaan olevan $F$:n *semanttinen seuraus* $F \models X \to Y$, jos jokaiselle $F$:n mukaiselle relaatiokaavion ilmentymälle $r$ pätee $r \models X \to Y$.

Päättelyjärjestelmän sanotaan olevan *oikeellinen*, jos kaikille riippuvuusjoukoille $F$ ja riippuvuuksille $X \to Y$ pätee:

$$\text{jos } F \vdash X \to Y, \text{ niin } F \models X \to Y$$

Oikeellisuus tarkoittaa sitä, että mikä tahansa riippuvuus, joka voidaan päätellä relaatiokaavioon $R$ liitetystä riippuvuusjoukosta $F$ käyttämällä oikeellista päättelyjärjestelmää on voimassa jokaisessa relaatiokaavion $R$ ilmentymässä $r$, jossa kaikki joukon $F$ riippuvuudet ovat voimassa [EN99].

Päättelyjärjestelmän sanotaan olevan *täydellinen*, jos kaikille riippuvuusjoukoille $F$ ja riippuvuuksille $X \to Y$ pätee:

$$\text{jos } F \models X \to Y, \text{ niin } F \vdash X \to Y$$

Täydellisyys tarkoittaa sitä, että on mahdollista päätellä riippuvuusjoukosta $F$ kaikki mahdolliset riippuvuudet soveltaen päättelyjärjestelmän sääntöjä yhdistellen ja toistaen [EN99].

Ilman tyhjäarvoja määritelty Armstrongin aksioomiin perustuva funktionaalisten riippuvuuksien päättelyjärjestelmä on oikeellinen ja täydellinen [Vas80, EN99]. Voidaan todistaa, että tämä päättelyjärjestelmä on oikeellinen ja täydellinen myös siinä tapauksessa, jos tyhjäarvot sallitaan, kun käytetään vahvan voimassaolon määritelmää [Vas80]. Tätä väitettä ei kuitenkaan voida suoraan tehdä, mikäli kohdassa 4 määritelty heikko riippuvuuksien voimassaolo sallitaan. Seuraavassa kohdassa 6 tutkitaan päättelyjärjestelmän käyttöä tilanteessa, jossa heikko voimassaolo on sallittu.

# 6 Riippuvuuksien voimassaolo

Kohdassa 4 tarkasteltiin yksittäisen funktionaalisen riippuvuuden voimassaoloa relaatiokaavion ilmentymässä. Usein samaan relaatioon liittyy kuitenkin yksittäisen riippuvuuden sijasta funktionaalisten riippuvuuksien joukko [EN99]. Kun riippuvuuksien heikkoa voimassaoloa tutkitaan, kunkin riippuvuuden voimassaolon tarkastelu erikseen voi tuottaa eri tuloksen kuin kaikkien riippuvuuksien tarkastelu joukkona [Vas80].

| $r$ | ammattinimike | keskipalkka | elintaso |
|---|---|---|---|
| | a1 | null | e1 |
| | a1 | null | e2 |

Taulukko 5: Relaatiokaavion ilmentymä, jonka riippuvuuksia on tarkasteltava joukkona.

**Esimerkki** Relaatiokaavioon *R(ammattinimike, keskipalkka, elintaso)* on liitetty seuraavat funktionaaliset riippuvuudet:

$$f_1 : \textit{ammattinimike} \rightarrow \textit{keskipalkka}, f_2 : \textit{keskipalkka} \rightarrow \textit{elintaso}$$

Taulukossa 5 on esitetty relaatiokaavion $R$ yksi mahdollinen ilmentymä $r$. Yksitellen tarkasteltuna molemmat riippuvuudet $f_1$ ja $f_2$ ovat heikosti voimassa (molemmat saavat arvon tuntematon). Kuitenkin koko riippuvuusjoukkoa tarkasteltaessa huomataan, että keskipalkka-sarakkeen arvojen tulisi olla erisuuret, jotta riippuvuus $f_2$ olisi voimassa. Tällöin riippuvuuden voimassaolo $f_1$ olisi epätosi, eli se ei olisi voimassa heikosti, mikä on eri tulos kuin mikä saatiin riippuvuuksia yksitellen tarkastellessa.

## 6.1 Tyhjäarvoekvivalenssirajoite

Edellä nähty esimerkki osoitti, että eri riippuvuudet voivat asettaa tiettyjä rajoitteita niille arvoille, joita tyhjäarvot voivat edustaa [Vas80]. Tyhjäarvon korvaaminen millä tahansa kyseisen sarakkeen arvojoukon arvolla ei ole tällöin mahdollista. Tämä idea voidaan esittää formaalisti määrittelemällä *tyhjäarvoekvivalenssirajoite* (*Null-Equality-Constraint*, *NEC*).

Monikon $t$ attribuutin $A$, sekä monikon $u$ attribuutin $B$ välinen tyhjäarvoekvivalenssirajoite merkitään:

$$NEC : t[A] := u[B]$$

Tyhjäarvoekvivalenssirajoite tarkoittaa tilannetta, jossa kaksi tyhjäarvoa ovat arvoltaan ekvivalentteja, ts. jos toinen niistä korvataan jollakin arvolla, myös toinen tyhjäarvo on korvattava tällä samalla arvolla [Vas80]. Näin tyhjäarvot voidaan jakaa eri ekvivalenssiluokkiin.

Tyhjäarvoekvivalenssit voidaan tuoda tyhjäarvoja sisältävään relaatioon *tyhjäarvon korvaussäännön* (*Null-Substitution-Rule*, *NS-rule*) avulla [Vas80].

## 6.2 Tyhjäarvon korvaussääntö

Olkoon $R$ relaatiokaavio, $X \rightarrow Y$ siinä esiintyvä funktionaalinen riippuvuus, $r$ relaatiokaavion $R$ ilmentymä sekä $t$ ja $u$ ilmentymän $r$ monikoita. Riippuvuuden $X \rightarrow Y$ tyhjäarvon korvaussääntö [Vas80]:
Jos $t[X] = u[X] \neq null$ tai on asetettu rajoite $NEC : t[X] := u[X]$ :
1. Jos $t[Y] = null$ tai $u[Y] = null$, muttei molemmat, niin korvataan tyhjäarvo epätyhjällä arvolla.
2. Jos $t[Y] = u[Y] = null$, niin asetetaan rajoite $NEC : t[Y] := u[Y]$.

Kun tyhjäarvon korvaussääntöä sovelletaan iteratiivisesti relaatiokaavion $R$ ilmentymän $r$ kullekin funktionaaliselle riippuvuudelle niin pitkään, kunnes yhtään tyhjäarvon korvaussääntöä ei voida enää soveltaa, tuloksena saadaan *minimaalisesti epätäydellinen* (*minimally-incomplete*) ilmentymä $r$:stä [Vas80]. Minimaalinen epätäydellisyys tarkoittaa käytännössä mahdollisimman pientä määrää tyhjäarvoja relaatiokaavion ilmentymässä — tilannetta, jossa yhdellekään tyhjäarvolle ei voida enää johtaa epätyhjää arvoa muista epätyhjistä arvoista.

Tyhjäarvon korvaussääntöjä soveltamalla relaatiokaavion ilmentymästä voidaan kuitenkin päätyä moneen vaihtoehtoiseen minimaalisesti epätäydelliseen ilmentymään [Vas80]. Eri tilat voidaan saavuttaa vaihtamalla sitä järjestystä, missä funktionaaliset riippuvuudet käydään läpi. Jos riippuvuuksien joukossa $F$ on useampia muotoa $X \rightarrow Y$ olevia riippuvuuksia, joilla on samoja attribuutteja joukossa $Y$, niin tyhjäarvon korvaussäännön mukaan joukkoon $Y$ kuuluvat tyhjäarvon sisältävät attribuutit voivat saada eri arvoja riippuen siitä, mikä riippuvuus aiheuttaa ensimmäisenä tyhjäarvon korvauksen.

**Esimerkki** Relaatiokaavioon *R(ammattinimike, keskipalkka, elintaso)* on liitetty seuraavat funktionaaliset riippuvuudet:

$$f_1 : \ ammattinimike \rightarrow elintaso, f_2 : \ keskipalkka \rightarrow elintaso$$

| $r$ | ammattinimike | keskipalkka | elintaso |
|---|---|---|---|
| *t1* | a1 | p1 | e1 |
| *t2* | a2 | p2 | e2 |
| *t3* | a1 | p2 | null |

Taulukko 6: Relaatiokaavion ilmentymä, jossa tyhjäarvo on osana useampaa riippuvuutta.

Taulukossa 6 on esitetty relaatiokaavion $R$ yksi mahdollinen ilmentymä $r$. Jos tyhjäarvon korvaussääntöä sovelletaan ensin riippuvuudelle $f_1$, niin säännön mukaan:

$$t3[ammattinimike] = t1[ammattinimike], \text{ joten } t3[elintaso] = t1[elintaso]$$

Jos tyhjäarvon korvaussääntöä sovelletaan ensin riippuvuudelle $f_2$, niin säännön mukaan:

$$t3[keskipalkka] = t2[keskipalkka], \text{ joten } t3[elintaso] = t2[elintaso]$$

Monikon *t3* attribuutin *elintaso* arvoa ei siis voida päätellä yksiselitteisesti tyhjäarvon korvaussäännön avulla. Tämä voidaan esittää formaalisti laajentamalla tyhjäarvon korvaussäännön määritelmää *määrittelemättömän arvon* (*inconsistent element*, *nothing-value*) käsitteen avulla.

## 6.3 Laajennettu tyhjäarvon korvaussääntö

Olkoon $R$ relaatio, $X \to Y$ siinä esiintyvä funktionaalinen riippuvuus, $r$ relaatiokaavion $R$ ilmentymä sekä $t$ ja $u$ ilmentymän $r$ monikoita.
*Jos* $t[X] = u[X] \neq null$ tai on asetettu rajoite $NEC : t[X] := u[X]$ :
1. Jos $t[Y] = null$ tai $u[Y] = null$, muttei molemmat, niin korvataan tyhjäarvo epätyhjällä arvolla.
2. Jos $t[Y] = u[Y] = null$, niin asetetaan rajoite $NEC : t[Y] := u[Y]$.
3. Jos $t[Y] \neq null$, $u[Y] \neq null$, $t[Y] \neq u[Y]$, niin asetetaan $m[Y] = määrittelemätön$, kun $m[Y] = t[Y]$ tai $m[Y] = u[Y]$, kaikilla $m \in r$.

Jos laajennettua tyhjäarvon korvaussääntöä sovelletaan taulukossa 6 esitettyyn relaatiokaavion ilmentymään, saadaan tulokseksi taulukossa 7 esitetty minimaalisesti epätäydellinen ilmentymä, jossa esiintyy määrittelemättömiä arvoja.

| *r* | ammattinimike | keskipalkka | elintaso |
|---|---|---|---|
| *t1* | a1 | p1 | määrittelemätön |
| *t2* | a2 | p2 | määrittelemätön |
| *t3* | a1 | p2 | määrittelemätön |

Taulukko 7: Relaatiokaavion minimaalisesti epätäydellinen ilmentymä, jossa esiintyy määrittelemättömiä arvoja.

Laajennetun tyhjäarvon korvaussäännön avulla voidaan päätellä funktionaalisten riippuvuuksien joukon $F$ heikko voimassaolo relaatiokaavion ilmentymässä $r$ [Vas80]. $F$ on heikosti voimassa ilmentymässä $r$ jos ja vain jos sen minimaalisesti epätäydellisessä ilmentymässä ei ole yhtään määrittelemätöntä arvoa. Armstrongin aksioomiin perustuva riippuvuusjoukolle $F$ sovellettu päättelyjärjestelmä on oikeellinen ja täydellinen tällaisessa minimaalisesti epätäydellisessä ilmentymässä.

# 7 Yhteenveto

Raportissa tutkittiin funktionaalisia riippuvuuksia epätäydellisen tiedon kontekstissa. Tärkeimmässä osassa oli funktionaalisten riippuvuuksien voimassaolon tarkastelu, sillä voimassa olevien riippuvuuksien perusteella voidaan relaatiotietokanta normalisoida ja vähentää näin tietokannan päivitysanomalioita.

Epätäydellistä tietoa esittävä tyhjäarvo edustaa olemassa olevaa, mutta sillä hetkellä tuntematonta arvoa, jolloin se approksimoi kaikkia muita saman arvojoukon arvoja. Tällöin tyhjäarvon attribuuttinaan saava funktio palauttaa tietyn arvon vain, jos tämä funktio palauttaa saman arvon kaikilla saman arvojoukon arvoilla. Mikäli näin ei ole, funktion arvoa ei voida yksiselitteisesti päätellä.

Tämän periaatteen perusteella määriteltiin funktionaalisen riippuvuuden vahva voimassaolo, jossa riippuvuus on voimassa relaatiokaavion ilmentymässä. Vahvan riippuvuuden voimassaolon lisäksi määriteltiin heikko riippuvuuden voimassaolo, jossa riippuvuuden voimassaolo on epävarmaa, mutta sitä ei voida osoittaa epätodeksi. Tällaisen epävarmuuden salliminen on perusteltua, sillä kuvamme reaalimaailmasta on usein epätäydellinen ja hieman löysemmät eheysrajoitteet voivat helpottaa käytännön sovelluksia.

Raportissa myös tuotiin esille se, että funktionaalisten riippuvuuksien yhteydessä käytetty Armstrongin aksioomiin perustuva päättelyjärjestelmä on oikeellinen ja täydellinen myös siinä tapauksessa, kun relaatiokaavioiden ilmentymissä esiintyy tyhjäarvoja ja jos käytetään vahvan voimassaolon vaatimusta. Sen sijaan heikon voimassaolon vaatimuksen tapauksessa relaatiokaavion ilmentymästä tulee ensin johtaa minimaalisesti epätäydellinen ilmentymä ilman määrittelemättömiä arvoja, jolloin päättelyjärjestelmää voi soveltaa tähän ilmentymään.

# Lähteet


Cod70 Codd, E. F., A relational model of data for large shared data banks. *Commun. ACM*, 13,6(1970), sivut 377–387.

Cod71 Codd, E. F., Further normalization of the data base relational model. IBM Research Report, San Jose, California, RJ909.

EN99 Elmasri, R. A. ja Navathe, S. B., *Fundamentals of Database Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

IWL84 Imieliński, T. ja Witold Lipski, J., Incomplete information in relational databases. *J. ACM*, 31,4(1984), sivut 761–791.

MR92 Mannila, H. ja Räihä, K.-J., *The Design of Relational Databases*. Addison-Wesley Publishing Company, 1992.

Vas79 Vassiliou, Y., Null values in data base management a denotational semantics approach. *SIGMOD '79: Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 1979, ACM, sivut 162–169.

Vas80 Vassiliou, Y., Functional dependencies and incomplete information. *VLDB '1980: Proceedings of the Sixth International Conference on Very Large Data Bases*, Montreal, Quebec, Canada, 1980, VLDB Endowment, sivut 260–269.