

Retsensiooni küsimuste vastused

1. Töö põhitlemiks oleva disainide kataloogi all pean silmas tervet jaotist number 4. Seda ütlen ka sissejuhatuse lõpus, kus annan ülevaate töö struktuurist. Kataloogi all pean silmas tervet jaotist, mitte vaid disaine kirjeldavaid alampunkte 4.3-4.8, sest disainide paremaks mõistmiseks ning lihtsamaks lugemiseks tuleks tutvuda eelnevalt käesolevas töös kasutatava mustri kirjelduse struktuuriga. Samuti annan selles peatükis ülevaate näitest, mida kasutan disaini tehnilise realisatsiooni kirjeldamiseks kui ka hilisemas eksperimendis. Jaotist 7 disaini kataloogi osaks ei pea, sest seal tehtud järelduste põhjal täiendasin disainide kataloogis esitatud eeliseid ja puuduseid ning lisasin viite sellele. Ehk seal esitatavad tulemused toetavad kataloogis esitatud teoreetilisi teadmisi praktiliste katsetustega.
2. Lõputööde lisades esitatakse enamasti töö teksti toetavad materjalid, nagu näiteks kellegi teise poolt tehtud kood või mudelid, millele töö kirjutamisel tuginetakse. Minu töö põhiosas välja toodud päringute puhul on tegemist töö reaalse tulemustega, sest viidi läbi analüüs leidmaks kõige parema töökiirusega päring igale disainile ja ülesande paarile. See on tingitud sellest, et SQL-andmebaasides on võimalik lahendada ühte ülesannet mitut moodi. Seetõttu arvan, et tegemist pole töö teksti toetava materjaliga ning esitan need magistritöö põhiosas.
3. Liiasus tekitab probleeme nii koodis, andmetes kui ka kirjatöös. Olen nõus, et oleksin võinud viidata antud alampeatükki kirjelduses Tabelis 1 esitatud nõutele ning öelda seejuures, et disain peaks olema võimalikult paindlik, tulemaks toime seal kirja pandud erinevate probleemidega.
4. Töö kokkuvõttes ütlen, et disaini valik on otseselt sõltuv konkreetsest olukorrast ja täidetavast ülesandest. Seega nõustun, et retsesendi poolt pakutud viis on üheks heaks võimaluseks, kuidas disainide analüüsi läbi viia ja tulemusi esitada. Tegin omaltpoolt analüüsi, kus kirjutasin iga nõudmise kohta välja - tuginedes eksperimentidele ja kirjandusele - millised oleksid selle rahuldamiseks kõige sobivamad ja ebasobivamad disainid. Siin võite näha ühte selle tabeli fragmenti, terve tabel is on nähtav GitHubi lehel, mille aadress on samuti nähtav slaidil.

	Täpsem nõue	Kõige sobivam(ad) disain(id)	Kõige vähemsobimav(ad) disain(id)
1.1	Olem peab olema kindlasti mingis seisundis	Seisundiklassifikaator	Põhiolemitüübi alamtüüpidenä Seisundite tuletamine
1.2	Olem saab olla korraga maksimaalselt ühes seisundis	Seisundiklassifikaator	Põhiolemitüübi alamtüüpidenä Seisundite tuletamine
1.3	Olem saab olla korraga mitmes seisundis	Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine Põhiolemitüübi alamtüüpidenä Seisundite tuletamine	Seisundiklassifikaator
2.1	Olemi puhul tuleb teada ainult selle hetkeseisundit	Seisundiklassifikaator Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine	Põhiolemitüübi alamtüüpidenä Seisundite tuletamine

2.2	Olemi puhul tuleb teada selle seisundite ajalugu	Temporaalsed veerud Põhiolemitüübi alamtüüpide Seisundite tuletamine	Tõeväärtustüüpi veerud Vektorkodeerimine
3.1	Olemi puhul huvitab ainult seisundi omamise fakt	Vt nõue 1.1 ja 2.1	Vt nõue 1.1 ja 2.1
3.2	Olemi puhul on vaja lisainfot seisundimuudatuse kohta	Põhiolemitüübi alamtüüpide Seisundite tuletamine	Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine
4.1	Lubatavate seisundimuudatuste kontroll ilma olekutabelita	Seisundiklassifikaator Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine	Põhiolemitüübi alamtüüpide Seisundite tuletamine
4.2	Lubatavate seisundimuudatuste kontroll olekutabeliga	Seisundiklassifikaator Vektorkodeerimine	Seisundite tuletamine
5.1	Andmebaasis ei ole vaja säilitada lisainfot seisundite kohta	Seisundiklassifikaator Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine Põhiolemitüübi alamtüüpide Seisundite tuletamine	-
5.2	Andmebaasis on vaja säilitada lisainfot seisundite kohta	Seisundiklassifikaator	Temporaalsed veerud Tõeväärtustüüpi veerud Vektorkodeerimine Põhiolemitüübi alamtüüpide Seisundite tuletamine
6.1	Tekib uus võimalik seisund, milles olem võib viibida	Seisundiklassifikaator	Põhiolemitüübi alamtüüpide Vektorkodeerimine
6.2	Kaob seisund, milles olem võib viibida	Seisundiklassifikaator	Põhiolemitüübi alamtüüpide Vektorkodeerimine

Nõude „Olem peab olema kindlasti mingis seisundis“ puhul pean kõige enamsobivamaks disainiks Seisundiklassifikaatori tabelit, sest disainide kataloogis esitatud lahendus juba täidab antud nõuet ning muudatusi selleks tegema ei pea. Kõige vähemsobivamateks disainideks on seejuures Põhiolemitüübi alamtüüpide ja Seisundite tuletamine. Põhjuseks on see, et selliste kontrollide andmebaasi tasemel deklaratiivseks jõustamiseks puudub tänapäeva SQL-andmebaasisüsteemides võimalus, sest need ei toeta üldiste kitsenduse (ASSERTION) loomist ja alampäringuid CHECK kitsendustes. Lahenduseks on kasutada trigereid, mis tagavad, et iga olemit identifikaator on registreeritud vähemalt ühele seisundile vastavas tabelis, mis on keerukas ning suurendab oluliselt disaini realiseerimiseks vajalike lausete hulka nagu näitasid minu töö tulemused (väiksematest ligi 8 korda suuremad)

Nõude „Olemi puhul tuleb teada selle seisundite ajalugu“ puhul pean kõige enam sobivamateks disainideks Temporaalsed veerud, Põhiolemitüübi alamtüüpide ja Seisundite tuletamine. Temporaalsed veerud ja Seisundite tuletamine osutus valituks, et kataloogis esitatud lahendus juba täidab antud nõuete selle klausliga, et olem võib samas seisundis olla ainult ühe korra, Põhiolemitüübi alamtüüpide tuleb lisada vaid vastavad veerud iga seisundi kohta käivasse tabelisse. Kõige vähemsobivamateks disainideks selgusid disainid

Tõeväärtustüüpi veerud ja Vektrokodeerimine, kuna antud nõude täitmiseks tuleb kasutusele võtta kas eraldi seisundimuudatuse ajaloo tabel, mis tähendab, et andmed põhiolemitüübi võimalike seisundite kohta on esitatud mitmes tabelis või täiendada põhiolemitüübi tabelleid temporaalsete veergudega, mis omakorda viitab sellele, et olem saab olla ainult korra ühes seisundis.

Lisaks tehtud analüüsi tulemuste paremaks visualiseerimiseks lugesin iga disaini korral kokku, kui mitme nõudmise puhul oleks see disain kõige sobivamate hulgas ja kui mitme disaini puhul oleks see kõige ebasobivamate hulgas. Antud tulemused on esitatud tabelis, nagu näha siis seisundiklassifikaatori tabel on antud nõuete puhul kõige sobivam disain kasutamiseks ja Põhiolemitüübi alamtüüpidena disaini kõige vähem sobivam. Samas ei tohiks teha sellest koheselt järeldust, et Põhiolemitüübi alamtüüpidena disaini ei peaks kasutama, sest ka selle kasutamises leidub tegelikult positiivset. Nagu näiteks selliste päringute kiirus, kus on vaja teada kõiki kindlas seisundeid olevaid olemeid.

Disaini nimi	Nõuete arv (kõige enim)
Seisundiklassifikaator	10
Temporaalsed veerud	5
Tõeväärtustüüpi veerud	4
Vektorkodeerimine	5
Põhiolemitüübi alamtüüpidena	4
Seisundite tuletamine	4

Disaini nimi	Nõuete arv (kõige vähem)
Põhiolemitüübi alamtüüpidena	8
Seisundite tuletamine	7
Vektorkodeerimine	5
Tõeväärtustüüpi veerud	3
Temporaalsed veerud	2
Seisundiklassifikaator	1

5. Käesoleva töö disainid on mõeldud kasutamiseks operatiivandmete andmebaasides. Kuigi vähemalt mõnda neist võiks kasutada andmeaitades ja andmevakkades ei ole selle uurimine käesoleva töö teema. Arvan, et kuigi disaini valik on sõltuv konkreetsest olukorrast ja ülesandest, ei piira antud töös esitatud üldistused valdkondi, kus tulemusi kasutada. Kataloogis esitatud disainidele ning kasutatava näite andmebaasi tulemustele võib toetuda ka mõnda teise rakendusvaldkonda kuuluvate infosüsteemide loomisel. Nagu näiteks raamatupidamissüsteemid, mis hoiavad informatsiooni arve või laekumiste elutsükli kohta. Kui ka õppeinfosüsteemid, mis hoiavad andmeid mõne õppeaine või õpilase seisundi osas.

Tellimuste andmed, mida kasutati näitena antud lõputöös, on näited transaktsioonilistest andmetest. Transaktsiooniliste andmete andmemahud on palju suuremad võrreldes põhiandmetega (*master data*) ja klassifikaatoritega (*reference data*). Kaks miljonit rida transaktsiooniliste andmete tabelis on täiesti realistlik. Näiteks kaupade, teenuste või õppeainete puhul räägiksime me sadadest kuni kümnetest tuhandetest ridadest ühes tabelis.

Sellisel juhul on andmemahud sõltumata disainist üsna väikesed ja päringud täidetakse sõltumata disainist kiiresti ning seetõttu muutuvad oluliseks teised kriteeriumid nagu näiteks paindlikus ja koodi lihtsus. Töös pühendatakse ruumi ka nendele kriteeriumitele.

6. Jah, 3 punkti kasutamine Pearsoni korrelatsioonikordaja arvutamiseks on vähevõitu. Valisin selle, sest soovisin tõestada oma töö tulemusi tuginedes mõnele teaduslikule teoreemile. Kahjuks töö maht ei võimaldanud teha mõõtmisi rohkemate tulemustega.

(Pearsonit kasutasin eesmärgil hinnata kas täitmisaja kasvamise e töökiiruse vähenemise ja andmehulga suurenemise vahel on lineaarne sõltuvus või mitte)

Antud teoreemi autorid soovivad kasutada vähemalt 25-30-st punktist koosnevat andmehulka, sest vähema korral on raske märgata võimalike kõrvalekaldeid või vigu, mida tuleks uurida. Siiski ei märgi nad ära, et mis on minimaalne andmehulka kasutamaks antud teoreemi korrelatsioonikordaja leidmiseks. Wikipedia soovitab kasutada väiksemate andmehulkade korral lineaarse sõltuvuse leidmiseks Korrigeeritud korrelatsiooni koefitsienti.

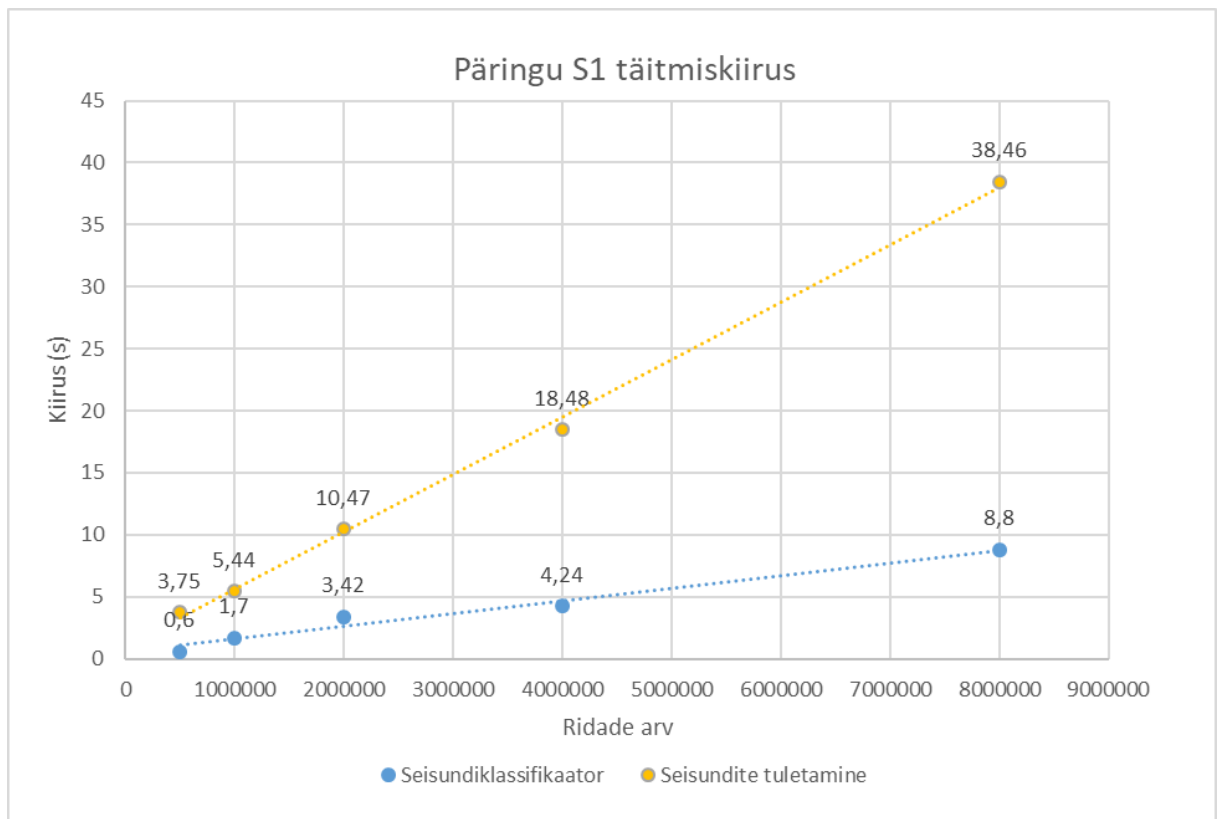
Samuti viisin läbi ka lisaeksperimenti disainidele Seisundite tuletamine ja Seisundiklassifikaatori tabel, millele mõõtsin ka päringute S1 ja S4 täitmiskiiruseid ridade arvu 4 000 000 ja 8 000 000 juures. Arvutasin saadud tulemuste põhjal uue Pearsoni korrelatsiooni kordaja ning saadud tulemuste põhjal võib öelda, et päringute töökiirus ja andmehulk on endiselt teineteisega positiivselt (kasvav seos tunnuste vahel) lineaarses sõltuvuses.

Disain ja ridade arv	S1	S4
Seisundiklassifikaator 500 000	0,60	0,00068
Seisundiklassifikaator 1 000 000	1,70	0,00084
Seisundiklassifikaator 2 000 000	3,42	0,00098
Seisundiklassifikaator 4 000 000	4,24	0,00111
Seisundiklassifikaator 8 000 000	8,80	0,00125

Disain ja ridade arv	S1	S4
Seisundite tuletamine 500 000	3,75	0,37
Seisundite tuletamine 1 000 000	5,44	0,90
Seisundite tuletamine 2 000 000	10,47	1,39
Seisundite tuletamine 4 000 000	18,48	1,98
Seisundite tuletamine 8 000 000	38,46	3,21

Andmehulk	1 S1	1 S4	6 S1	6 S4
500000	0,6	0,00068	3,75	0,37
1000000	1,7	0,00084	5,44	0,9
2000000	3,42	0,00098	10,47	1,39
4000000	4,24	0,00111	18,48	1,98
8000000	8,8	0,00125	38,46	3,21
1 S1 Pearson	0,986535			
1 S4 Pearson	0,923804			
6 S1 Pearson	0,999049			
6 S4 Pearson	0,984747			

Täpselt seda sama näitab ka slaidil esitatud hajuvusdiagramm, mis on üheks võimaluseks leidmaks kahe või enama mõõtmistulemuse vastastikust seost. Kui hajuvusdiagrammi peal asuv punktirühm on sirgjoonega hästi kokkuvõetav, on korrelatsioonikordaja absoluutväärtus kõrde. Nagu eelnevalt sai öeldud, siis töö teoreemi autorid soovivad kasutada suuremat andmehulka, sest jäävad märkamata võimalikud vead või kõrvalekõlled, mida tuleks uurida. Seega võib öelda, et jah Pearsoni korrelatsioonikoefitsiendi arvutamine 3 punkti põhjal on vähene põhjusel, sest nagu näha jooniselt, siis mõlema disaini puhul tuleks uurida lähemalt mõningaid tulemusi, mis asuvad trendi joonest kaugemal. Siiski ei muutunud töös leitud korrelatsioonikordajad oluliselt ning nende põhjal tehtud järeldusi otseselt valeks ei pea.



7. Tunnistan tehtud kirjavigu, seoses „mina“ ja „teie“ vormide kasutamisega soovisin tähelepanu juhtida oma isikule panusele oma magistritöös. Samuti ei täpsusta ka lõputöö koostamise ja vormindamise juhend, millist kirjastiili tuleks kasutada magistritöös. Mitmed erinevad allikad, mis annavad soovitusi keelekasutuse kohta akadeemilistes töödes, ütlevad, et kaudset kõnet hakati kasutama peale 1920ndat aastat teadustööde koostamisel eemaldamaks lugejas tunnet, et tegemist on vaid sinu isiklike arvamustega, mis ei põhine teaduslikul tõestusel. Samas arvavad mõned inimesed, et sõna „Mina“ vältimine on üsna silmakirjalik, sest lugejale jääda mulje nagu autor varjaks oma arvamust. Mis omakorda tekitab küsimusi töö teaduslikus täpsuses kui esitatud väitele ei järgne autori heaks kiitvat selgitust või seda ümber lükkavat argumenti. Seega jõutake erinevates allikatest järeldusele, et kui töö autor kirjeldab oma tegevusi praktilise kallakuga uuringutes, siis otsese kõne kasutamine on vastuvõetavam.